

**Mémoire présenté devant l'Université de Paris Dauphine  
pour l'obtention du diplôme du Master d'Actuariat  
et l'admission à l'Institut des Actuaires**

**le**

Par : Sibylle Lot

Titre : Définition des critères de choix du modèle de provisionnement et illustration avec des données d'arrêt de travail

Confidentialité :  Non  Oui (Durée :  1 an  2 ans)

---

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

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

*Entreprise :*

Nom : Prim'Act

**Prim'Act**

Signature : Mme Inès Burr

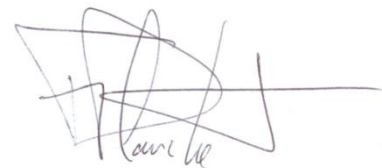


*Membres présents du Jury du Master  
Actuariat de Paris Dauphine*

*Directeur de Mémoire en entreprise :*

Nom : M. Frédéric Planchet

Signature :



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

*Secrétariat :*

*Signature du responsable d'entreprise*

*Bibliothèque :*

*Signature du candidat*



## Résumé

---

Le montant des provisions représente une partie importante du passif de l'assureur dans son bilan. Son calcul est donc un élément clé. Un large choix de méthodes est possible. Comme le note Hans Bühlmann dans la postface du tome 2 de DENUIT et CHARPENTIER [2005]) :

*« le provisionnement n'est pas un problème de modélisation complexe et sophistiqué, mais plutôt un exercice de choix de modèle » .*

L'objectif de ce mémoire est de donner des critères de choix de méthode de provisionnement et du niveau de segmentation des données à appliquer.

Nous développons dans un premier temps une présentation générale du provisionnement. Les problématiques liées au provisionnement et les différentes provisions sont exposées. Chaque catégorie de risques et de types de données est décrit ainsi que les méthodes traditionnellement utilisées pour chacun. Cette analyse permet de définir des critères de choix de méthodes en fonction du risque et/ou des données disponibles.

Afin d'illustrer nos propos, une comparaison de plusieurs méthodes de provisionnement, sur des données d'arrêt de travail, est faite dans la suite. La base de données étudiée a la particularité d'être volumineuse ce qui permet de travailler à différents niveaux de granularités. Ainsi, une méthode agrégée (Chain Ladder), une méthode ligne à ligne (table de maintien en incapacité) et une méthode ligne à ligne sur une population segmentée avec le modèle de Cox sont confrontées.

---

*Mots clés* : Provisionnement, arrêt de travail, Kaplan-Meier, Chain-Ladder, Whittaker-Henderson, table d'expérience, incapacité, modèle de Cox



## Abstract

---

The amount of provisions represents a significant portion of the insurer's liabilities on its balance sheet. Its calculation is therefore a key element. A wide choice of methods is possible. As Hans Bühlmann notes in the afterword to volume 2 of DENUIT and CHARPENTIER[2005]):

"Provisioning is not a complex and sophisticated modelling problem, but rather an exercise in model choice."

The goal of this memorandum is to give criteria for the choice of provisioning method and the level of data segmentation to be applied.

In the first part, we develop a general presentation of provisioning. The issues related to provisioning and the various provisions are explained. Each category of risks and data types is described, as well as the methods traditionally used for each. This analysis makes it possible to define criteria for choosing the type of methods according to the risk and/or available data.

To illustrate our comments, a comparison of several provisioning methods, based on work stoppage data, is presented below. The database studied has the particularity of being voluminous, which makes it possible to work at different levels of granularity. Thus, an aggregated method (Chain Ladder), a line-by-line method (disability maintenance table) and a line-by-line method on a segmented population with the Cox model are compared.

---

*Key words* : Reserving, work stoppage, Kaplan-Meier, Chain-Ladder, Whittaker-Henderson, experience table, incapacity, Cox model.



# Note de Synthèse

## Contexte et Problématique

L'assurance est caractérisée par un cycle de production inversé : l'assureur reçoit les primes de ses assurés avant de connaître le montant final exact du coût des sinistres. L'assureur doit donc constituer des provisions afin de pouvoir couvrir les sinistres de ses assurés. Le montant de ces provisions représente une part importante du passif au bilan de l'assureur. Il est donc important de déterminer ces provisions au plus juste.

Les provisions peuvent être définies comme la somme des espérances des coûts futurs et d'une marge de prudence. Selon la directive Solvabilité II, les provisions techniques correspondent à la somme des meilleures estimations et de la marge de risque.

*« La meilleure estimation correspond à la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent (valeur actuelle attendue des flux de trésorerie futurs), estimée sur la base de la courbe des taux sans risque pertinent. Le calcul de la meilleure estimation est fondé sur des informations actualisées et crédibles et des hypothèses réalistes et il fait appel à des méthodes actuarielles et statistiques adéquates, applicables et pertinentes ». (Article 77 de la Directive 2009/138/CE).*

Il n'existe pas une méthode universelle pour calculer les provisions, chaque méthode ayant ses avantages et ses inconvénients. La notion de précision d'une méthode par rapport à une autre est définie dans ce mémoire comme la comparaison du montant de provisions calculé avec le montant réellement indemnisé. La méthode qui a la différence la plus faible est dite la plus prudente.

Quel que soit le niveau de provision calculé par l'assureur, cela ne change pas le montant payé par ce dernier à l'assuré ou ayant droit à terme. Cela permet seulement à l'assureur de voir plus ou moins vite son résultat. En effet, si l'assureur sous-provisionne, il réalisera un résultat positif les premières années puis des mali les années suivantes. Si au contraire, la provision est trop élevée, le résultat sera négatif au départ puis des boni seront réalisés.

Hans Bühlmann note dans la postface du tome 2 de DENUIT et CHARPENTIER [2005], « *le provisionnement n'est pas un problème de modélisation complexe et sophistiqué, mais plutôt un exercice de choix de modèle* » : quelle méthode doit utiliser un actuair en fonction des données qu'il a.

Ce mémoire présente deux objectifs :

- Présenter les raisons et les enjeux du provisionnement ainsi que les méthodes utilisées sur le marché selon les risques et les données.
- Comparer différentes méthodes de provisionnement sur des données d'arrêt de travail afin d'illustrer les critères de choix présentés.

## Méthodologie

Le choix de la méthode de provisionnement est un exercice délicat et dépend du type de données et du type de sinistres. Pour cela, il est nécessaire de distinguer les sinistres graves et attritionnels. En général, ces sinistres sont provisionnés séparément pour que les sinistres graves ne soient pas sous-provisionnés et n'introduisent pas un biais dans le calcul des provisions des sinistres attritionnels.

Les sinistres peuvent être également séparés entre les sinistres courts et longs ainsi qu'entre les sinistres connus et les IBNR.

Ces caractéristiques qui définissent les sinistres et les risques ont un impact sur le choix des méthodes de provisionnement.

Les méthodes utilisées, agrégées ou individuelles, ont certains avantages et inconvénients. Elles sont basées sur un principe commun : les évolutions du passé donnent des informations sur les évolutions futures. Les données passées permettent de paramétrer les modèles. En ligne à ligne, les données passées sont utilisées pour construire des tables qui sont ensuite appliquées sur les données individuellement. Pour les méthodes agrégées, les données du passé sont regroupées afin de déterminer des cadences de déroulement. Les évolutions futures sont calculées de manière agrégée à l'aide de ces cadences. Quel que soit la méthode, un historique et des données suffisantes sont nécessaires. Dans le cas contraire, l'assureur peut utiliser la méthode de Loss Ratio ou S/P. Cette méthode nécessite une connaissance précise du marché.

L'étude des caractéristiques des différents types de sinistres (assurance de Personne, IARD, ...) permet de distinguer les grandes problématiques et critères déterminant ainsi le choix d'une méthode plutôt qu'une autre. Par exemple en Assurance de Personnes, la distinction majeure est la durée de vie du sinistre. Un modèle individuel, qui est au plus proche des données et qui prend en compte la particularité de chaque sinistre, semble plus adapté pour les sinistres longs pour lesquels l'enjeu du provisionnement est plus important étant donné leurs caractéristiques. Cependant, les données sur les sinistres sont obtenues au fur et à mesure de la vie du sinistre, ce qui peut rendre l'application de ces méthodes difficile à mettre en œuvre. A l'inverse, l'utilisation d'un modèle ligne à ligne pour des sinistres courts n'apporte généralement pas d'éléments complémentaires aux méthodes agrégées.

En IARD, le critère essentiel est celui de la distinction entre les sinistres graves et attritionnels. Les enjeux et les caractéristiques des sinistres graves semblent nécessiter une méthode ligne à ligne, afin de modéliser au mieux chaque sinistre et ses particularités. Au contraire, les sinistres attritionnels sont fréquents et avec des montants plus standards, ainsi une méthode agrégée type Chain Ladder semble le plus adapté.

Ainsi ces différents éléments permettent de définir certains critères de choix de méthode.

		Vie du sinistre	
Type d'assurance	Caractéristiques	IBNR	Sinistres connus
Assurance de Personne	Court Terme	Méthode de triangles	Méthodes de triangles
	Long Terme		Méthodes ligne à ligne
IARD	Graves		Méthodes ligne à ligne
	Attritionnels		Méthodes de triangles

Tableau de synthèse des méthodes



Une fois la méthode de provisionnement choisie, il faut également déterminer si elle est appliquée à l'ensemble de la population ou si une segmentation est réalisée. Généralement, les données sont, au minimum, classées par types d'activités ou peuvent être classifiées de manière très fine afin d'obtenir des ensembles homogènes de risques et avoir un modèle représentant aux mieux la réalité. La segmentation se fait selon le comportement des individus de la population assurée, les cadences de règlements, les caractéristiques des produits ou la proportion de tardifs.

Cependant, en segmentant, la problématique du manque de données pour obtenir des tables ou des triangles fiables peut se poser. Il est donc nécessaire d'arbitrer entre le niveau de granularités des classifications, la qualité et la quantité de données à disposition.

Afin d'illustrer ces différents critères, une comparaison de plusieurs méthodes de provisionnement sur des données d'arrêt de travail est réalisée par la suite. Ce type de sinistres est traditionnellement provisionné en ligne à ligne, ainsi tous les types de méthodes pourront être appliqués à nos données.

La base de données utilisée à la particularité d'être volumineuse et comprend de nombreuses variables qualitatives (CSP, Sexe, Mode de souscription et Type d'arrêt) et quantitatives (date de début et de fin d'incapacité, date de fin de franchise,...). Une segmentation de la population est donc possible en fonction des variables qualitatives. L'ensemble des données sont comprises, après retraitement, entre le 01/01/2005 et le 31/12/2007 ; l'année 2005 est utilisée comme année d'apprentissage et les années 2006 et 2007 comme années de test. Le pas temporel choisi est mensuel.

Les montants d'indemnisation étant absents de la base de données, on considère qu'un euro est versé par nombre de jours indemnisés, ce qui revient à comparer le nombre de jours indemnisés et celui estimé.

Les projections des nombres de jours indemnisés sont effectuées par exercice civil. Pour les arrêts en cours au 31/12/N-1, les durées d'arrêt résiduelles pour l'année N sont estimées et sont comparées avec la durée effective des prestations sur cette même année.

La durée d'indemnisation effective des arrêts en cours au 31/12/N-1 sur l'année N est estimée comme le nombre de jours, plafonnée à 1 an, séparant la date de fin des indemnisations et le 31/12/N-1.

Ce mécanisme de calcul est reconduit successivement pour les années 2006 à 2007. Par exemple, on estime pour les arrêts en cours au 31/12/2005, la durée d'arrêt résiduelle sur 2006, et ainsi de suite. Au final, on obtient pour les années 2006 à 2007 une estimation annuelle des durées d'indemnisation cumulées pour les arrêts en cours au début de chaque année considérée.

Il convient donc de définir ce qu'est un sinistre « en cours » à une date donnée. A la date d'arrêt 31/12/N, les sinistres en cours sont ceux dont la date de fin de franchise est antérieure au 31/12/N et la date de sortie de l'état d'incapacité est strictement supérieure au 31/12/N.

Les différentes méthodes de provisionnement utilisées sont les suivantes :

- la méthode agrégée de Chain Ladder-Mack,
- la méthode ligne à ligne de table de maintien en incapacité à partir de l'estimation de Kaplan-Meier et un lissage de Whittaker-Henderson sur la population globale,
- la méthode ligne à ligne de table de maintien en incapacité à partir de l'estimation de Kaplan-Meier et un lissage de Whittaker-Henderson sur la population segmentée avec le modèle de Cox.

La méthode de Chain Ladder est la méthode la plus utilisée car simple à mettre en pratique et à comprendre. Elle est basée sur des triangles de liquidation de montants de sinistre ou de nombres, qui agrègent les données individuelles des sinistres par année de survenance et année de développement.

La méthode Kaplan-Meier permet d'obtenir une estimation de la fonction de survie en fonction de l'âge à l'entrée et de l'ancienneté dans le sinistre. Le lissage des taux obtenus est fait avec une méthode non-paramétrique (Whittaker-Henderson) au vu de la quantité importante de données.

Le volume et les données qualitatives de la base permettent d'appliquer le modèle de Cox. On obtient ainsi des coefficients par classe de population. Une table d'expérience est réalisée avec les données de la classe la plus importante. Dans cette étude, elle est composée des femmes, non-cadres, avec un mode de souscription de convention collective et avec un arrêt dû à la maladie. Pour obtenir les coefficients de maintien des autres classes, les coefficients obtenus par le modèle de Cox sont appliqués aux taux de maintiens de la table principale.

Ces trois méthodes permettent de comparer les résultats obtenus avec une méthode agrégée et une méthode ligne à ligne, mais également d'observer l'impact de la segmentation sur le montant de provisionnement calculé.

## Résultats

On définit comme Boni/Mali la différence en pourcentage entre la durée provisionnée et la durée réelle.

L'intervalle de confiance à 95% est déterminé par les bornes Min et Max.

Les calculs à l'aide des triangles de développement donnent les résultats suivants :

Indemnisation	Année	Réel	Estimé	Ecart en %	Volatilité	Borne Min	Borne Max
<b>Nombre total des jours futurs indemnisés sur un an</b>	2007	399 924	466 979	16,8%	8 142	451 021	482 938

### Résultats de la méthode de Chain Ladder

Ces résultats prennent en compte les IBNR ; c'est-à-dire les sinistres survenus en 2005 mais indemnisés à partir de 2006.

On constate un boni de liquidation de 16,8%.

Cependant, en fonction du nombre de mois pris pour le calcul des coefficients, on obtient des résultats variables : avec 6 mois d'historiques, on obtient un boni de 22,1% et un boni de 16,5% avec 4 mois d'historique. Les résultats obtenus avec la méthode de Chain Ladder sont donc volatiles en fonction du nombre de mois d'ancienneté pris en compte pour le calcul des coefficients de passage. Cela rend cette méthode peu fiable avec les données utilisées.

Les résultats obtenus avec la table de maintien en incapacité basée sur la population globale sont les suivants :

Indemnisation	Année	Nombre de sinistres	Réel	Estimé	Ecart en %	Volatilité	Borne Min	Borne Max
<b>Nombre total des jours futurs indemnisés sur un an</b>	2007	1893	293 425	266 483	- 9,18	10 782	245 350	287 616

### Résultats avec la table construite sur la population totale

Avec une table de maintien sur une population segmentée, les résultats sont les suivants :

Indemnisation	Année	Nombre de sinistres	Réel	Estimé	Ecart en %	Volatilité	Borne Min	Borne Max
<b>Nombre total des jours futurs indemnisés sur un an</b>	2007	1893	293 425	310 658	5,87%	11 457	288 203	333 113

### Résultats avec la table construite sur la population segmentée

On constate un boni de liquidation de 5,9% pour l'année 2007 avec une population segmentée et un mali de 9,2% avec la population globale.

## Conclusion

Les données étudiées ont la particularité d'être en volume important, ce qui a permis en plus de la méthode de Chain Ladder et de tables couramment utilisées, d'appliquer des tables avec segmentation de la population selon des variables explicatives.

Les résultats obtenus montrent que la méthode de Chain Ladder ne semble pas adaptée à ce type de sinistres. Les résultats étant relativement volatiles en fonction du nombre de mois d'ancienneté pris pour obtenir les coefficients de passage. La méthode ligne à ligne avec une table sur la population globale obtient de meilleurs résultats que Chain Ladder. Cependant les malis obtenus sont encore relativement importants. Finalement, la méthode la plus concluante est celle ligne à ligne avec une population segmentée. Donc, plus la maille d'observation et d'apprentissage des données est fine plus le montant de provisions est précis.

Cependant, tous les portefeuilles ne permettent pas de segmenter comme nous avons pu le faire, par manque de volume ou manque de variables explicatives. Les modèles ligne à ligne ne sont pas encore développés ou généralisés pour tous les types de sinistres.

Il serait intéressant pour compléter cette étude de calculer la durée d'indemnisation par sinistres directement à l'aide d'un modèle de régression (GLM ou boosting). En effet, les données étant très peu censurées, nous disposons d'informations suffisantes pour développer un modèle de régression.



# Synthesis note

## Problem and subject

Insurance is characterized by a reverse production cycle: the insurer receives the premiums from its insureds before knowing the exact final amount of the cost of claims. The insurer must therefore set aside reserves in order to cover the claims of its insureds. The amount of these provisions represents a significant portion of the insurer's balance sheet liabilities. It is therefore important to determine these provisions with accuracy.

Provisions can be defined as the sum of expectations of future costs and a margin of prudence. According to the Solvency II Directive, technical provisions are the sum of best estimates and risk margin.

*"The best estimate is the probability weighted average of future cash flows, taking into account the time value of money (expected present value of future cash flows), estimated on the basis of the relevant risk-free yield curve. The calculation of the best estimate is based on current and credible information and realistic assumptions and uses appropriate, applicable and relevant actuarial and statistical methods."* (Article 77 of Directive 2009/138/EC)

There is no universal method for calculating provisions; indeed each method has its pros and cons. The notion of precision of one method in relation to another is defined in this memorandum as the comparison of the amount of provisions calculated with the amount actually compensated. The method with the smallest difference is called the most conservative.

Whatever the level of provision calculated by the insurer, this does not change the amount paid by the insurer to the insured or entitled in the future. This only allows the insurer to see his result more or less quickly. Indeed, if the insurer underfunds, it will achieve a positive result in the first years and then mali in the following years. If, on the other hand, the provision is too high, the result will be negative at the beginning and then bonuses will be realized.

Hans Bühlmann notes in the afterword to Volume 2 of DENUIT and CHARPENTIER[2005], "provisioning is not a complex and sophisticated modelling problem, but rather an exercise in model choice": what method should an actuary use based on the data he has.

This brief has two objectives:

- Present the reasons and issues for funding and the methods used in the market based on risks and data.
- Compare different provisioning methods on work stoppage data to illustrate the selection criteria presented.

## Process

The choice of the provisioning method is a difficult task and it depends on the type of data and the type of claims. To do this, it is necessary to identify both serious and attritional losses. Serious losses are expensive but with a low frequency while attritional losses are cheaper but frequent. In general, these claims are provisioned separately so that serious claims are not underfunded and do not introduce a bias in the calculation of attritional claims reserves.

Claims can also be separated between short and long claims as well as between known claims and IBNRs.

These characteristics that define claims and risks have an impact on the choice of funding methods.

The methods used, whether aggregated or individual, have certain advantages and disadvantages. They are based on a common principle: past developments provide information on future developments. The past data are used to configure the models. Online to line, past data is used to build tables that are then applied to the data individually. For aggregated methods, past data is grouped together to determine flow rates. Whatever the method, sufficient history and data are required. Otherwise, the insurer may use the Loss Ratio or S/P method. This method requires a precise knowledge of the market.

The study of the characteristics of the various types of claims (personal insurance, property and casualty insurance, etc.) makes it possible to distinguish the major problems and criteria that determine the choice of one method over another. For example, in Life and Health Insurance, the major distinction is the life of the claim. An individual model that is as close as possible to the data and that takes into account the specificity of each claim seems more appropriate for long claims for which the issue of provisioning is more important given their characteristics. However, claims data are obtained over the life of the claim, which can make the method application - whatever method is at stake - difficult to implement. Conversely, the use of a line-by-line model for short claims does not generally provide additional elements to the aggregated methods.

In P&C, the essential criterion is the distinction between serious and attritional losses. The stakes and characteristics of serious claims seem to require a line-by-line method, in order to best model each claim and its particularities. On the contrary, attritional claims are provisioned using several triangles methods used simultaneously.

Thus these different elements make it possible to define certain criteria for the choice of method.

		Vie du sinistre			
		Type of insurance	Caractéristiques	IBNR	Known claims
Nature du sinistre	Life and health insurance	Short-term	Triangles Method	Triangles Method	Triangles Method
		Long term			Line-by-line Method
	Property and casualty insurance	Serious			Line-by-line Method
		Attritionals			Triangles Method

Summary table of methods

Once the provisioning method has been chosen, it must also be determined whether it is applied to the entire population or whether segmentation is carried out. Generally, the data are, at a minimum, classified by type of activity or can be classified in a very detailed way in order to obtain homogeneous sets of risks and to have a model that best represents reality. Segmentation is based on the behaviour of individuals in the insured population, payment rates, product characteristics or the proportion of late payments.

However, when segmenting, the problem of lack of data to obtain reliable tables or triangles may arise. It is therefore necessary to decide between the level of granularity of classifications, the quality and quantity of data available.

To illustrate these different criteria, a comparison of several provisioning methods on work stoppage data is then carried out. This type of claim is traditionally provisioned online, so all types of methods can be applied to our data.

The database used is particularly voluminous and includes many qualitative variables (CSP, Sex, Underwriting method and Type of termination) and quantitative variables (start and end date of disability, end date of deductible,...). It is therefore possible to segment the population according to qualitative variables. All data are included, after restatement, between 01/01/2005 and 31/12/2007; 2005 is used as the learning year and 2006 and 2007 as the test years. The time step chosen is monthly.

As the amounts of compensation are not included in the database, it is assumed that one euro is paid per number of days compensated, which is equivalent to comparing the number of days compensated and the estimated number.

Projections of the number of days compensated are made by calendar year. For shutdowns in progress at 31/12/N-1, the remaining shutdown periods for year N are estimated and compared with the actual duration of the services over the same year.

The effective compensation period for judgments in progress at 31/12/N-1 in year N is estimated as the number of days, capped at 1 year, between the end date of the compensation and 31/12/N-1.

This calculation mechanism is successively renewed for the years 2006 to 2007. For example, for shutdowns in progress at 31/12/2005, the remaining shutdown period for 2006, and so on, is estimated. In the end, an annual estimate is obtained for the years 2006 and 2007 of the cumulative compensation periods for the judgments in progress at the beginning of each year targeted.

It is therefore necessary to define what is an "ongoing" claim at a given date. At the 31/12/N stop date, claims in progress are those for which the end date of the deductible is before 31/12/N and the date of exit from the state of incapacity is strictly higher than 31/12/N.

The different provisioning methods used are as follows:

- the Chain Ladder-Mack aggregate method,
- the line-by-line method of maintaining incapacity based on the Kaplan-Meier estimate and a Whittaker-Henderson smoothing on the overall population,
- the line-by-line disability maintenance table line method based on the Kaplan-Meier estimate and a Whittaker-Henderson smoothing on the population segmented with the Cox model.

The Chain Ladder method is the most widely used method because it is simple to apply and understand. It is based on triangles for liquidating claim amounts or numbers, which aggregate individual claim data by accident year and development year.

The Kaplan-Meier method provides an estimate of the survival function as a function of age at entry and seniority in the claim. The smoothing of the obtained rates is done with a non-parametric method (Whittaker-Henderson) due to the large amount of data.

The volume and qualitative data in the database allow the Cox model to be applied. This results in coefficients by population class. With the same method as for the overall population, a maintenance table is calculated for the main class. Thus, Cox coefficients can be applied on this table according to the classes.

These three methods make it possible to compare the results obtained with an aggregated method and a line-by-line method, but also to observe the impact of segmentation on the calculated provisioning amount.

## Results

The percentage difference between the provisioned term and the actual term is defined as Boni/Mali.

The 95% confidence interval is determined by the Min and Max terminals.

Calculations using the development triangles give the following results:

Indemnisation	Year	Real	Estimated	Difference in %	Volatility	Borne Min	Borne Max
Nombre total des jours futurs indemnisés sur un an	2007	399 924 €	466 979 €	16,8%	8 142	451 021	482 938

### Results of the Chain Ladder method

These results take into account IBNRs; i.e. claims that occurred in 2005 but are compensated from 2006 onwards.

There is a liquidation surplus. However, depending on the number of months to calculate the coefficients, variable results are obtained: a bonus of 22.1% is a result of 6 months of historical data and a bonus of 16,5% is a result of 4 months of historical data.

To be noted : the results obtained with the Chain Ladder method are volatile depending on the number of months of seniority taken into account for the calculation of the transition coefficients. This makes this method unreliable with the data used.

The results obtained with the disability maintenance table based on the overall population are as follows:

Indemnisation	Year	Number of claims	Real	Estimated	Difference in %	Volatility	Borne Min	Borne Max
Nombre total des jours futurs indemnisés sur un an	2007	1893	293 425 €	266 483 €	-9,18%	10 782	245 350	287 616

### Results with the table built on the total population

With a maintenance table on a segmented population, the results are as follows:

Indemnisation	Year	Number of claims	Real	Estimated	Difference in %	Volatility	Borne Min	Borne Max
Nombre total des jours futurs indemnisés sur un an	2007	1893	293 425 €	310 658 €	5,87%	11 457	288 203	333 113

### Results with the table built on the segmented population

There is a liquidation surplus of 5.9% for 2007 with a segmented population and a deficit of 9.2% with the overall population.



## Conclusion

The data studied are particularly voluminous, which made it possible. In addition to the Chain Ladder method and tables, it was therefore possible to apply a method with tables with population segmentation according to explanatory variables.

The results obtained show that the Chain Ladder method does not seem to be suitable for this type of claim. The results are relatively volatile depending on the number of months of seniority taken to obtain the transition coefficients. The line-by-line method with a table on the global population achieves better results than Chain Ladder. However, the malis obtained are still relatively high. Finally, the most conclusive method is the line-by-line method with a segmented population. Therefore, the finer the observation and learning mesh of the data, the more precise the amount of provisions is.

However, not all portfolios allow us to segment as we have been able to do, due to lack of volume or explanatory variables. Line-by-line models are not yet developed or generalized for all types of claims.

To complete this study, it would be interesting to calculate the duration of compensation per claim directly using a regression model (GLM or boosting). Indeed, since the data are very little censored, we have sufficient information to develop a regression model.



## Remerciements

Avant toute chose, je tiens à remercier mon référent d'entreprise, Madame Inès BURR, consultante en actuariat chez Prim'Act, pour son suivi, ses conseils avisés et son soutien tout au long de la réalisation de mon mémoire. Je la remercie également pour son encadrement et le partage de son expérience professionnelle au cours des diverses missions que j'ai eu l'occasion de réaliser au sein de Prim'Act.

Par ailleurs, je tiens à exprimer ma grande gratitude à Monsieur Frédéric PLANCHET, associé chez Prim'Act, pour sa disponibilité, son expertise et ses corrections régulières tout au long de l'élaboration de ce mémoire. À cela s'ajoute des remerciements pour ses précieux conseils et ses nombreuses références en la matière.

Mes remerciements s'adressent également à mon référent universitaire, Madame Angéline ROCHE, pour sa disponibilité, son suivi continu et ses multiples conseils, ainsi qu'aux équipes pédagogiques du département de Mathématiques de l'Université Paris Dauphine.

J'adresse des remerciements particuliers à toute l'équipe Prim'Act, ainsi qu'à leurs dirigeants pour le partage de leurs connaissances tout au long de ce stage et pour leur accueil chaleureux.

Enfin je remercie toutes les personnes ayant contribué de près ou de loin à l'élaboration de ce mémoire, notamment toutes celles qui ont relu ce mémoire.

# Table des matières

Note de Synthèse.....	7
Synthesis note.....	13
Remerciements.....	19
Table des matières.....	20
Introduction.....	23
Chapitre 1 : .....	25
<b>Pourquoi et comment provisionner ? Panorama des méthodes de provisionnement</b> .....	25
1.1    Description de la vie d'un sinistre.....	25
1.2    Les provisions .....	27
1.2.1    Définition d'une provision.....	27
1.2.2    Les différentes provisions à calculer .....	30
1.3    Choix de la méthode de provisionnement .....	32
1.3.1    Le type de données .....	32
1.3.2    Le type de sinistres.....	33
1.3.3    Comparaison des méthodes et notion de précision .....	34
1.4    Les méthodes pour l'Assurance Non-Vie .....	35
1.4.1    Assurances de personnes et Responsabilité Civile Corporelle (RCC).....	36
1.4.2    Assurances de choses ou de biens et Responsabilité Civile Matérielle.....	39
1.4.3    Assurance construction : RC décennale et dommages ouvrages .....	40
1.5    Évolution des méthodes .....	41
1.5.1    Limites et solutions des méthodes agrégées .....	41
1.5.2    Les évolutions académiques .....	42
Chapitre 2 : .....	45
<b>Méthode de provisionnement en prévoyance</b> .....	45
2.1    La prévoyance.....	45
2.2    Présentation des données .....	46
2.2.1    La base de données.....	46
2.2.2    Le traitement des données .....	46
2.2.3    Statistiques descriptives.....	49
2.3    Table d'expérience .....	63
2.3.1    Éléments de base des modèles de survie.....	63
2.3.2    Estimation de la loi de survie par l'estimateur de Kaplan-Meier.....	64
2.3.3    Lissage des taux.....	65
2.3.4    Segmentation .....	69
2.4    Méthodes utilisant des triangles.....	73
2.4.1    Construction du triangle de développement.....	73
2.4.2    Méthode de Chain Ladder .....	74
2.4.3    Méthode de Mack.....	75
Chapitre 3 : .....	77

<b>Analyses et Comparaisons</b> .....	77
<b>3.1 Résultats avec la méthode Chain Ladder</b> .....	77
<b>3.1.1 Principe</b> .....	77
<b>3.1.2 Résultats</b> .....	78
<b>3.2 Résultats avec des tables d'expérience</b> .....	80
<b>3.2.1 Principe</b> .....	80
<b>3.2.2 Calcul pour un euro de prestation</b> .....	81
<b>3.2.3 Résultats avec la population globale</b> .....	82
<b>3.2.4 Analyse de l'hétérogénéité de la population</b> .....	82
<b>3.2.5 Résultats avec la population segmentée</b> .....	84
<b>Conclusion</b> .....	87
<b>Références</b> .....	88
<b>Lexique</b> .....	90
<b>Annexes</b> .....	91



# Introduction

L'assurance, indispensable et très développée de nos jours, existait déjà, d'une certaine manière, à l'antiquité. La première mise en place la plus marquante a été à l'époque des grandes expéditions en transport maritime. Des sommes importantes d'argent étaient nécessaires pour armer les navires, les armateurs faisaient des prêts auprès de banquiers. Le risque pris était grand mais pouvant rapporter d'importants intérêts : en cas de disparition (nauffrage ou perte) de la marchandise ou du bateau, les montants prêtés n'étaient pas remboursés et en cas de réussite de la mission la rémunération était importante, jusqu'à plus de 30% d'intérêt. D'autres initiatives étaient similaires à de l'assurance : des coopératives de corps de métiers s'organisaient pour permettre aux familles de subvenir à leurs besoins en cas de disparition du chef de famille.

Les provisions sont une part importante du passif des assureurs : 1 843,9 milliards d'euros en 2017 pour les assureurs Vie et Mixtes et 157.5 milliards d'euros pour les assureurs Non-Vie d'après la Fédération Française des Assurances. Ces montants augmentent chaque année. Le calcul des provisions est une des problématiques majeures des assureurs.

Au fur et à mesure, les développements des outils mathématiques et statistiques ont permis d'enrichir les méthodes et les techniques des actuaires pour tarifier et gérer les contrats. Aujourd'hui, les actuaires disposent d'un panel important de méthodes à disposition pour calculer les provisions : agrégées, ligne à ligne, déterministes, stochastiques. Cependant, il n'existe pas une méthode idéale, un choix qui peut s'avérer compliquer, est donc nécessaire.

L'objectif de ce mémoire est de donner des critères de choix pour les méthodes de provisionnement. La prise en compte de l'hétérogénéité sur la qualité de la prévision, en comparant des approches agrégées et ligne à ligne sur des données globales ou segmentées sera étudiée. Afin d'illustrer ces éléments, des données d'arrêt de travail seront utilisées. La comparaison des méthodes sera effectuée sur les durées d'indemnisation prévues et effectivement observées.

Dans la première partie, nous ferons une synthèse des méthodes de provisionnement en Assurance Non-Vie. Le but de cette partie est d'expliquer les raisons pour lesquelles un assureur provisionne, quelles sont ses contraintes et quelles méthodes sont généralement utilisées en fonction de la catégorie de risques.

Dans la suite du mémoire, nous présenterons les données utilisées pour l'étude. L'étude se base sur des données d'arrêt de travail qui ont la particularité d'être en volume conséquent. Le traitement de données effectuées sera exposé. Afin de connaître au mieux la population étudiée, des analyses statistiques seront conduites.

Dans la troisième partie, nous exposerons les différentes méthodes de provisionnement permettant de réaliser notre étude. Des méthodes ligne à ligne à l'aide de tables d'expériences globale ou segmentée ainsi que des méthodes agrégées reposant sur des triangles seront proposées.

Enfin dans la dernière partie, une application des méthodes sera effectuée avec les données présentées en seconde partie. Les résultats obtenus nous permettront de faire une analyse et une comparaison des durées d'indemnisation prévues par les différents modèles avec les valeurs réellement observées.



## Chapitre 1 :

# Pourquoi et comment provisionner ? Panorama des méthodes de provisionnement

L'assurance, contrairement à l'industrie classique, a un cycle de production inversé (cf. MARQUETTY [2012]). Les assurés payent une prime à l'assureur qui, en échange, s'engage à régler les sinistres qui ont lieu pendant la période de couverture. L'assureur reçoit les primes de ses assurés avant de connaître le montant final exact du coût des sinistres. De plus, les paiements des sinistres peuvent se faire en une ou plusieurs fois. Des provisions, représentant l'équivalent en capital des sommes restant à payer à l'assuré, doivent donc être constituées (cf. PLANCHET et al. [2006]). L'assureur doit être en capacité de déterminer au plus juste ses provisions pour pouvoir honorer ses engagements envers ses assurés.

### 1.1 Description de la vie d'un sinistre

Les différentes étapes de la vie d'un sinistre et de son règlement peuvent être schématiquement décrites de la manière suivante :

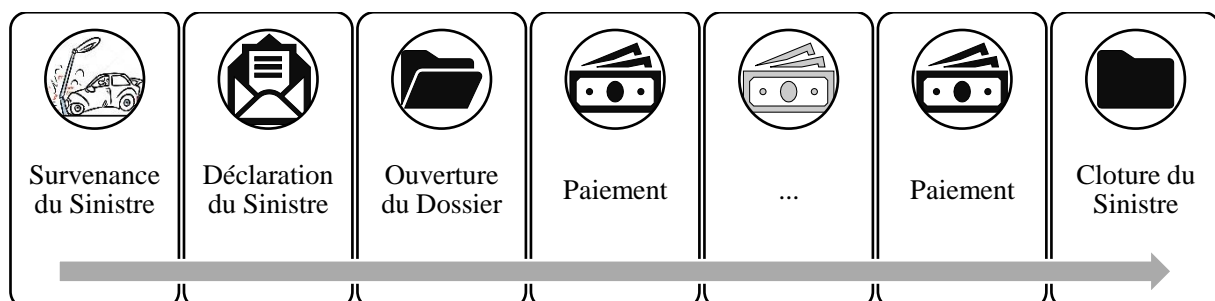


Figure 1: Vie d'un sinistre

- **La survenance du sinistre** : cet élément essentiel permet à l'assureur de connaître la validité de la garantie au moment du sinistre et la date de début de l'indemnité (avec plus ou moins un délai de carence). La date de survenance n'est pas toujours facile à définir. Par exemple pour un dégât des eaux, quelle date prendre entre celle de début de fuite du tuyau ou celle des premières constatations de dégâts ?
- **La déclaration du sinistre auprès de l'assureur** : l'assuré se doit de faire une déclaration dans un délai imparti auprès de son assureur pour que la garantie soit activée. Les délais de déclaration sont indiqués dans l'article L 113-2 du Code des Assurances.
- **L'ouverture du dossier par le gestionnaire** : la déclaration du sinistre permet à l'assureur de récupérer auprès de l'assuré certaines informations et de vérifier que la demande correspond

bien aux critères de la garantie (régularité du contrat, circonstances autorisées et type de sinistres).

- **L'expertise, estimation du montant de règlement** : à partir des informations obtenues, le gestionnaire détermine une estimation de la charge totale du sinistre : coût du sinistre nommée Provision Dossier en IARD ou la charge des prestations futures en assurance de personne. A cette occasion, le gestionnaire peut faire appel à l'avis d'un expert (médecin en cas de dommages corporels, expert conformité pour vérifier le nombre de pièces d'une assurance habitation, ...).
- **Le règlement de l'indemnité / Calcul de provisions** : l'assureur a une durée déterminée contractuellement pour indemniser (en une ou plusieurs fois) l'assuré, une fois que le montant est fixé. Le délai de règlement peut être allongé en raison de désaccord entre les parties pouvant aller jusqu'au recours en justice.
- **La clôture du dossier** : cette dernière étape intervient une fois que l'assuré ou le tiers a été entièrement indemnisé. L'assureur a le droit de réclamer auprès des personnes responsables ou de leurs assureurs, le montant des sommes réglées. Enfin une réouverture du dossier est possible (relativement rare) et correspond notamment à des aggravations médicales et des rechutes.

Différents éléments font que l'assureur peut ne pas connaître directement le coût final d'un sinistre au moment de la déclaration du sinistre : l'estimation de la charge ultime peut évoluer dans le temps, le dossier du sinistre peut être réouvert, l'expertise du coût du sinistre peut durer longtemps, la durée de versement d'une prestation peut être incertaine... Tous ces éléments peuvent entraîner plusieurs flux de règlements et rendent plus compliqué l'estimation du coût total du sinistre et le calcul des provisions. Ces dernières représentent l'espérance de la fraction du coût total qui reste à régler au moment de l'inventaire à laquelle s'ajoute une marge pour risque<sup>1</sup>.

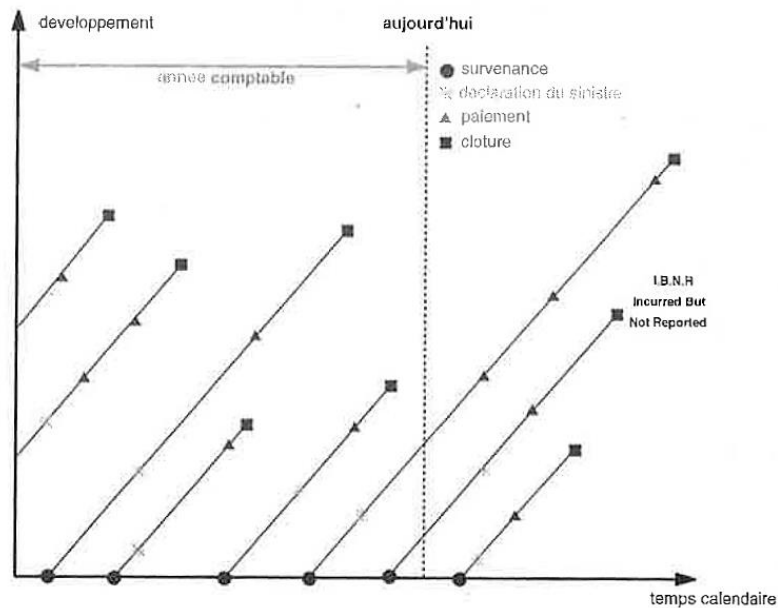
Les incertitudes dans le calcul des provisions proviennent de la non connaissance du montant final payé, et des dates de paiement. Des éléments extérieurs ont également une influence sur le montant final provisionné : des changements de politique de souscription, des modifications de la gestion des sinistres, l'inflation, des évolutions réglementaires...

Selon la description de la vie d'un sinistre et la date d'inventaire, on peut proposer de différencier trois catégories de sinistres, qui vont nécessiter des approches différentes du fait d'une information disponible spécifique à chaque catégorie :

- **Les sinistres survenus et déclarés** : la date de déclaration est antérieure à la date d'inventaire.
- **Les sinistres survenus mais non déclarés** : la date de d'inventaire est entre la date de survenance et la date de déclaration (IBNR : *Incurred But Not Reported*).
- **Les sinistres non encore survenus** : la date d'inventaire est antérieure à la date de survenance, l'assureur ayant par ailleurs déjà encaissé la prime pour la période correspondante.

---

<sup>1</sup> La marge pour risque est explicite dans le dispositif prudentiel Solvabilité 2 et implicite, au travers du choix d'hypothèses prudentes pour les comptes sociaux.



**Figure 2: Aspect dynamique de la gestion des sinistres**  
**Source : CHARPENTIER et DENUIT [2005] Mathématiques de l'Assurance Non Vie**

Sur le graphique ci-dessus appelé « Diagramme de Lexis », on distingue cinq sinistres clos (clôturés avant la ligne verticale « aujourd'hui »), un sinistre connu et déclaré, un IBNR et un sinistre non encore survenu.

Le montant de la provision correspond à la différence entre le montant ultime et les prestations déjà réglées. Il intègre l'ensemble des charges relatives aux sinistres (les prestations proprement dites ainsi que l'ensemble des frais de gestion, d'expertise, etc.).

Les problématiques de provisionnement sont les suivantes : quelle méthode de provisionnement choisir pour obtenir un niveau suffisant de provisions ? Comment prendre en compte les événements particuliers (sinistres exceptionnellement graves, les recours, ...) ? Comment être sûr d'avoir assez provisionné ?

## 1.2 Les provisions

### 1.2.1 Définition d'une provision

Les provisions représentent une part importante du passif au bilan de l'assureur. Elles sont la contrepartie en capital des engagements de prestations futures envers les assurés. On peut mesurer l'activité d'un assureur avec son chiffre d'affaires mais aussi avec le volume de provisions au bilan (particulièrement pour les branches longues). Les prestations futures ont un caractère aléatoire et il n'existe pas une méthode universelle de détermination du montant des provisions.

On peut définir formellement les provisions comme la somme des espérances des coûts futurs et d'une marge de prudence. La marge peut être incluse ou non dans le montant de provisions et permet de couvrir les engagements envers les assurés, les années comprenant des sinistres extrêmes.

Les provisions sont calculées brutes de réassurance et nettes de recours (Art. R. 331-1 du Code des Assurances).

Les provisions ne sont pas définies de la même manière selon la réglementation prise en compte : Solvabilité II ou comptabilité sociale française.

- **Solvabilité II**

Selon la directive Solvabilité II, **les provisions techniques correspondent à la somme des meilleures estimations (BE) et de la marge de risque.**

$$Provision = BE + Marge\ de\ risque$$

ACTIF	PASSIF
Actifs	Fonds Propres
	Marge pour risque
	Best Estimate

**Figure 3: Bilan simplifié d'une compagnie d'Assurance**

« La meilleure estimation correspond à la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent (valeur actuelle attendue des flux de trésorerie futurs), estimée sur la base de la courbe des taux sans risque pertinent. Le calcul de la meilleure estimation est fondé sur des informations actualisées et crédibles et des hypothèses réalistes et il fait appel à des méthodes actuarielles et statistiques adéquates, applicables et pertinentes ». (Article 77 de la Directive 2009/138/CE)

$$BE = E \left( \sum_k \frac{CF_k}{(1 + i_k)^k} \middle| I \right),$$

avec  $CF_k$  les flux futurs,  $i_k$  les taux d'actualisation et  $I$  les informations (flux passés, informations sur les assurés et hypothèses) connues à la date de calcul.

Le calcul de l'espérance est fait sous la probabilité risque neutre (cf. LAURENT et al. [2014]). L'espérance permet de prendre en compte l'ensemble des scénarios possibles et l'actualisation intègre la valeur temporelle de l'argent. L'actualisation nécessite d'avoir une cadence de règlement des flux futurs que la méthode soit agrégée ou ligne à ligne.

BLUM et al. [1998] présentent l'importance de la définition de la notion de *Best Estimate* (meilleure estimation) : quand on parle de « meilleure estimation », on cherche à obtenir une sorte de valeur cible du montant des flux futurs, à une date donnée. On pourrait choisir différentes notions mathématiques :

- **Le point médian** entre les points extrêmes de la distribution statistique. Mais cette valeur ne donne pas d'informations sur la distribution globale.
- **Le mode de la distribution**, qui correspond au point avec la probabilité la plus élevée. Cependant, si la distribution est multimodale, ce point n'est plus unique. Le mode peut être que légèrement plus élevé que les autres points, il n'est alors que légèrement plus probable que les autres points.
- **La médiane** est autant susceptible d'être probablement trop élevée ou trop basse par rapport à l'estimation du montant réel à payer. Il est complexe de travailler avec, car le montant des provisions implique de sommer les provisions par types de garanties. A moins que ces

catégories soient identiquement distribuées, la somme de leur médiane ne correspond pas à la médiane de la distribution globale.

- **La moyenne** correspond à la moyenne pondérée en fonction de la probabilité de tous les résultats possibles pour l'ensemble des sinistres à payer. Cette valeur est mathématiquement pratique. La moyenne d'une distribution globale sera égale à la somme des moyennes des catégories (identiquement distribuées ou non).

Le *Best Estimate* est défini comme la meilleure estimation possible de la valeur attendue en moyenne.

La marge de risque « est calculée de manière à garantir que la valeur des provisions techniques est équivalente au montant que les entreprises d'assurance et de réassurance demanderaient pour reprendre et honorer les engagements d'assurance et de réassurance » (Article 77 de la Directive 2009/138/CE).

La marge de risque permet d'ajouter une marge de prudence au *Best Estimate*. Elle peut être interprétée comme le coût en capital pour une société en run-off pour couvrir l'ensemble de ses engagements avec une probabilité de 99,5%.

« les entreprises d'assurance et de réassurance procèdent à une évaluation séparée de la meilleure estimation et de la marge de risque » (Article 77 de la Directive 2009/138/CE).

- **Comptabilité sociale**

Dans les comptes sociaux des compagnies d'assurance en France, les principes suivants sont à respecter :

- Principe de prudence ;
- Principe de comptabilité en coût historique ;
- Principe de continuité de l'activité.

D'après le Code des Assurances, les provisions doivent être « suffisantes » pour permettre le règlement des engagements vis-à-vis des assurés :

« [...] Les provisions techniques suffisantes pour le règlement intégral de leurs engagements vis-à-vis des assurés ou bénéficiaires de contrats [...] calculées, sans déduction des réassurances cédées [...] » (Art. R. 331-1 du Code des Assurances).

Seules les provisions mathématiques de rentes sont actualisées. La provision doit être suffisante et intègre donc implicitement une marge de prudence contrairement au calcul de *Best Estimate* en Solvabilité II où la marge est calculée séparément. On aura donc un montant de provision qui sera supérieur ou égal à l'espérance des flux futurs.

« En pratique, on attend des entreprises d'assurance une évaluation de leur engagement à un niveau qui se révèle en moyenne supérieure au coût total final des sinistres » LUSTMAN et al [2001].

On peut noter l'absence d'information dans les textes réglementaires sur les méthodes à utiliser. Cela peut entraîner des interprétations différentes des textes.

La charge de prestations futures peut être considérée comme une variable aléatoire dont on cherche à déterminer la loi. Une fois cette loi connue, comment peut-elle être exploitée pour obtenir la provision ?

### 1.2.2 Les différentes provisions à calculer

Les provisions sur les sinistres comportent les montants à mettre en réserve pour couvrir les prestations mais également tous les éléments qui sont liés aux sinistres comme les frais de gestion, les sinistres exceptionnels et les recours. Les provisions sont calculées au moins à chaque inventaire.

L'ensemble des provisions que doit calculer un assureur dans son bilan sont décrites dans le Code des Assurances (R 331-6 du Code des Assurances.).

- **Provision pour Sinistres à Payer (PSAP)**

La PSAP est la provision principale d'un assureur Non-Vie, elle représente le montant des dépenses nécessaires (frais et montants payés à l'assuré) aux règlements complets des sinistres connus et restant à payer à la date de clôture.

Cette provision est décrite dans le code des Assurances aux articles R331-6, R331-15 à 18 et R331-26.

La PSAP correspond à la différence entre le montant ultime calculé par année de survenance moins le montant de prestations déjà réglé.

$$PSAP = Charge\ ultime - Prestations\ réglées,$$

En regroupant les données par année de survenance (en ligne) et année de développement (en colonne), on lit les prestations connues cumulées dans la partie supérieure et les charges à estimer dans la partie inférieure :

	1	2	...	j	...	n-1	n
1	$C_{1,1}$	$C_{1,2}$	...	...	...	$C_{1,n-1}$	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$	...	...	...	$C_{2,n-1}$	
...	...	...	...	...	...	...	
i	...	...	...	$C_{i,j}$	...	...	
...	...	...	...	...	...	...	
n-1	$C_{n-1,1}$	$C_{n-1,2}$	...	...	...		
n	$C_{n,1}$						

Cette diagonale comprend tous les règlements cumulés déjà effectués à la date d'arrêt comptable.

	1	2	...	j	...	n-1	n
1	$C_{1,1}$	$C_{1,2}$	...	...	...	$C_{1,n-1}$	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$	...	...	...	$C_{2,n-1}$	$\hat{C}_{2,n}$
...	...	...	...	...	...	...	...
i	...	...	...	$C_{i,j}$	...	...	...
...	...	...	...	...	...	...	...
n-1	$C_{n-1,1}$	$C_{n-1,2}$	...	...	...	$\hat{C}_{n-1,n-1}$	$\hat{C}_{n-1,n}$
n	$C_{n,1}$	$\hat{C}_{n,2}$	...	...	...	$\hat{C}_{n,n-1}$	$\hat{C}_{n,n}$

Cette dernière colonne correspond à la charge ultime estimée pour chaque année de survenance.

Le montant de charges restant à payer peut être séparé entre le montant pour les sinistres connus et le montant pour les sinistres tardifs (IBNR).

$$PSAP = \text{Provision Sinistres Connus} + IBNR$$

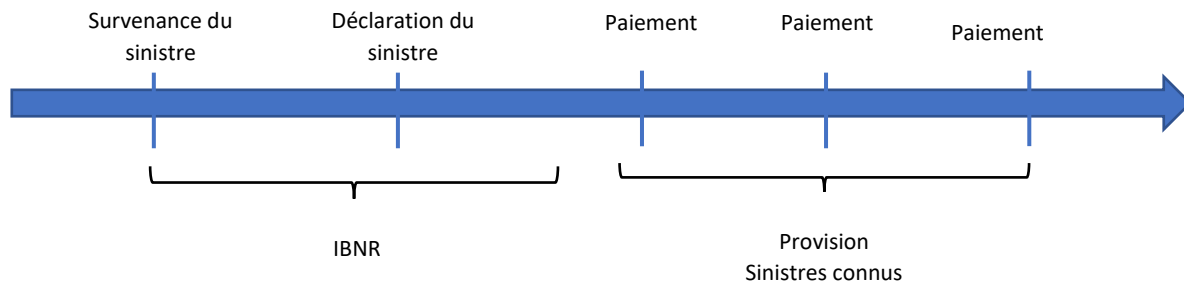


Figure 4: Schéma des provisions selon le moment de la vie du sinistre

La grande différence entre ces deux types de sinistres est l'absence de données dans le cas des IBNR et la connaissance de données pour les sinistres connus. Cela peut entraîner une différence de méthodologie et un calcul séparé des sinistres connus et des IBNR.

Pour les **sinistres déclarés**, les dates de survenance et de déclaration ainsi que des montants de prestations déjà réglés et l'estimation du montant des prestations futures, sont connues. L'aléa réside dans l'incertitude du coût total et dans les dates de règlements.

On peut distinguer les provisions pour rentes. Elles correspondent à la valeur actuelle probable de l'ensemble des rentes. Par exemple, les rentes invalidité dues à un accident de voiture en RC auto. La particularité de ces provisions est que le montant de rente est généralement connu, l'aléa se trouve donc essentiellement sur la durée du versement de la rente.

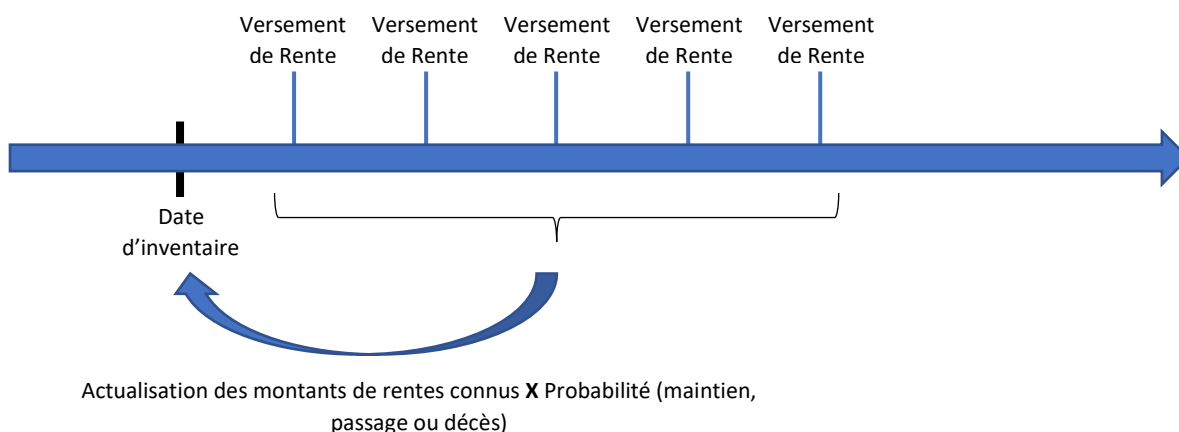


Figure 5: Schéma du provisionnement des rentes.

Dans le cas des **sinistres non encore déclarés**, toutes les informations sont inconnues. Plusieurs méthodes existent pour estimer le montant d'IBNR, il peut être :

- déduit de la différence entre le montant ultime et le montant de provisions des sinistres connus ;
- calculé comme le produit du coût moyen de sinistres et un taux de sinistres tardifs ;
- déterminé en reconstituant des sinistres à l'aide de l'historique, puis en calculant le montant de provision comme les sinistres connus.

- **Provision de recours :**

Les provisions de recours correspondent aux montants de remboursement d'une indemnité en attente d'actions en justice ou de réclamations à un tiers.

**Exemple :**

Un accident implique deux véhicules : le véhicule de A passe au feu rouge et percute un véhicule de B. L'assuré B va être remboursé par son assurance qui va ensuite exercer un recours contre l'assureur de A. Ce dernier devra rembourser à l'assureur de B l'indemnisation versée.

- **Provision pour frais de gestion :**

La provision pour frais de gestion couvre les frais des gestions des contrats quand ils ne sont pas couverts par d'autres revenus.

- **Provision pour Egalisation (PE) :**

Cette provision permet de faire face aux sinistres exceptionnels (risques atomiques, événements naturels, risques d'attentat, de terrorisme, pollution, ...). Le Code Général des Impôts donne la formule du calcul.

## 1.3 Choix de la méthode de provisionnement

Il existe de nombreuses méthodes de provisionnement basées sur des modèles de données agrégées ou ligne à ligne. Selon le type de risque, de compagnie, de données et de gestion de sinistres, l'actuaire choisit une méthode adaptée au contexte. De plus, comme noté précédemment, le provisionnement est contraint par la réglementation.

Il est nécessaire de distinguer le calcul des provisions et la mesure d'incertitude autour de cette estimation. Les méthodes, comme celle de Mack, qui permettent de déterminer l'incertitude ou la volatilité du montant de provision calculé, donnent des informations complémentaires sur la provision, mais ne changent pas le montant de provision établi.

Quel que soit le niveau de provision calculé par l'assureur, cela ne change pas le montant payé par ce dernier à l'assuré ou ayant droit à terme. Cela permet seulement à l'assureur de voir plus ou moins vite son résultat. En effet, si l'assureur sous-provisionne, il réalisera un résultat positif les premières années puis des mali les années suivantes. Si au contraire, la provision est trop élevée, le résultat sera négatif au départ puis des boni seront réalisés.

Hans Bühlmann note dans la postface du tome 2 de DENUIT et al. [2005], « *le provisionnement n'est pas un problème de modélisation complexe et sophistiqué, mais plutôt un exercice de choix de modèle* » : quelle méthode doit utiliser un actuaire en fonction des données qu'il a.

### 1.3.1 Le type de données

Le choix de la méthode de provisionnement dépend du type de risque et des données disponibles : une analyse des données est donc nécessaire. Cette étape nécessite une attention particulière. Elle permet de trouver les anomalies (valeurs incohérentes, doublons, ...) et de définir les regroupements homogènes. Ces derniers permettent de distinguer les sinistres attritionnels ou ordinaires, qui sont



fréquents et peu coûteux, et les sinistres graves, qui ont une fréquence plus faible mais dont le coût est bien plus élevé.

Fréquence/Coût	Faible	Fort
Faible		Sinistres graves
Fort	Sinistres Attritionnels	

**Figure 6: Tableau des types de sinistres par fréquence et coût.**

Les sinistres avec une faible fréquence et un faible coût rentrent dans la catégorie des sinistres attritionnels.

Les sinistres à forte fréquence et coût élevé<sup>2</sup> ne sont *a priori* pas assurables.

La séparation des sinistres graves et attritionnels nécessite la détermination d'un seuil au-delà duquel les sinistres sont considérés comme graves. On peut utiliser pour cela, la méthode de l'estimateur de Hill, l'estimateur du maximum de vraisemblance, la pente de l'espérance résiduelle ou une méthode graphique de la fonction moyenne des excès (ces différentes méthodes sont présentées dans les mémoires de BOYER CHAMMARD [2007-2008] et ITTAH [2016]).

Les caractéristiques des sinistres attritionnels font qu'ils sont particulièrement appropriés aux modèles agrégés. Le coût faible et la fréquence élevée permettent d'avoir des historiques statistiquement fiables.

À l'inverse, les sinistres graves ont une fréquence et des montants volatiles, ce qui rend les méthodes agrégées plus compliquées à appliquer.

Le provisionnement de ces deux types de sinistres peut se faire selon différentes méthodes.

Par exemple, en assurance IARD, les méthodes utilisées pour chacun des types de sinistres (ITTAH [2016]) sont souvent basées sur des triangles :

- Pour les sinistres attritionnels, les triangles des règlements permettent de calculer la charge agrégée.
- Pour les sinistres graves, la fréquence est calculée à partir du triangle des nombres et le montant utilisé correspond au coût moyen.

Néanmoins, on peut également faire du provisionnement ligne à ligne sur les sinistres graves tout en gardant la méthode des triangles sur les sinistres attritionnels (VU [2015]).

### 1.3.2 Le type de sinistres

Le choix de la méthode de provisionnement dépend également du type de sinistre considéré : sinistres connus ou IBNR. L'absence de certaines informations pour les IBNR empêche l'application des méthodes utilisées pour des sinistres connus.

Il est nécessaire de déterminer la manière dont le provisionnement des sinistres survenus mais non encore déclarés (IBNR) va être réalisé. En fonction des méthodes de provisionnement choisies, le montant d'IBNR est intégré à celui des sinistres connus. Il est intéressant d'analyser la proportion de sinistres tardifs selon le type de risques. En branches dommages ou frais de santé, il y a en général peu

<sup>2</sup> La notion de « coût élevé » s'entend au regard du montant de la prime payée ex-ante et n'a pas de sens absolu.

de tardifs, les sinistres sont connus et réglés rapidement (environ moins de deux ans), contrairement aux branches de Responsabilité Civile qui ont une grande proportion de tardifs.

Les sinistres peuvent également être différenciés en fonction de leur durée d'écoulement, soit le temps entre la survenance et la fin des règlements du sinistre. On distingue les sinistres à développement court comme les frais de Santé et ceux à développement longs comme la RC Décennale. Les enjeux de provisionnement ne sont pas les mêmes.

Pour les sinistres courts, le montant de prestations est en majorité réglé dès la première année. Les montants de provisions sont donc relativement faibles proportionnellement au montant à l'ultime.

A l'inverse, pour les sinistres longs, l'enjeu du provisionnement est bien plus important à cause de la répartition des montants indemnisés dans le temps et de la proportion de prestations restantes à payer. De plus, sur les branches longues la proportion de sinistres tardifs est importante ce qui augmente l'aléa sur l'estimation de la provision ultime.

### 1.3.3 Comparaison des méthodes et notion de précision

Les méthodes utilisées par les assureurs pour calculer les provisions peuvent être des évaluations individuelles ou des méthodes statistiques basées sur des données historiques. Elles peuvent être déterministes ou stochastiques.

La méthode la plus courante est Chain Ladder (cf Chapitre 2). Ce qui rend cette méthode si populaire est sa simplicité d'utilisation. Cette méthode se base sur la cadence de règlements à l'aide d'un triangle de développement.

Dans les années 1930, des livres de droit, comme ASTESAN [1938], font déjà référence à ce type de méthode : « *pour une société d'assurance donnée, il existe un certain rythme dans le règlement des sinistres survenus au cours de chaque exercice [...] ce rythme est appelé la cadence des règlements* ». « *Les entreprises d'assurances, lors de l'inventaire, établissent une ventilation par exercice des règlements de sinistres effectués au cours de l'année [...], la cadence des règlements varie avec les branches* ».

Cependant, cette méthode se base sur des hypothèses fortes :

- des données en nombre important pour que les méthodes statistiques soient suffisamment fiables et précises ;
- des cadences de règlement régulières ;
- des fréquences peu volatiles ;
- un futur et un présent structurellement semblables au passé.

D'autres méthodes ont été développées pour compléter ou compenser les limites de Chain-Ladder : la méthode de Bornhuetter Ferguson intègre une donnée exogène au triangle (avis d'expert) et le modèle de Mack donne le même résultat que Chain-Ladder, mais permet d'obtenir une estimation de la volatilité en plus de l'espérance.

Le tableau ci-après présente les statistiques réalisées sur le marché français de l'Assurance Non Vie par ASTIN [2016]. On peut noter que les méthodes les plus utilisées sont des méthodes déterministes basées sur des triangles d'écoulement.

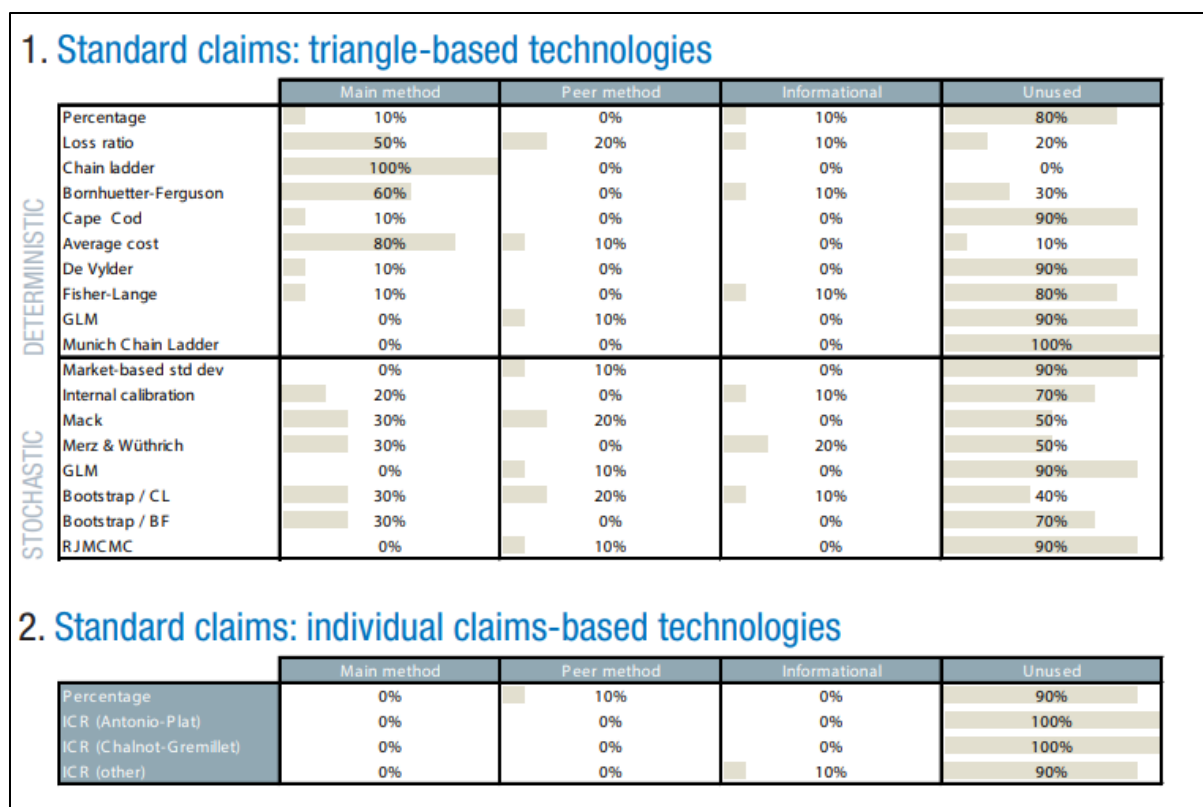


Figure 7: Statistiques (ASTIN [2016])

L'actuaire dispose de plusieurs méthodes pour déterminer le montant de provision, mais il n'existe pas une méthode universelle. Chaque méthode a ses avantages et ses inconvénients et est plus ou moins adapté selon le type de portefeuille. Toutes les méthodes ne donnent pas le même montant de réserves. Une méthode peut donner un montant de provisions plus élevé qu'une autre, elle sera donc plus prudente.

Afin de choisir entre plusieurs méthodes de provisionnement, on peut effectuer un test rétro-actif ou *Backtesting*. Cela consiste à calculer un montant de provisions à partir de données passées à une date d'inventaire donnée avec plusieurs méthodes. Puis, ce montant est comparé au montant de prestations observé dans la réalité. La méthode qui a la différence, en valeur absolue, la plus faible avec le montant réel, est la plus précise étant donné le portefeuille. La notion de précision d'une méthode par rapport à un autre peut faire référence à cette comparaison avec le montant réel sur des données passées.

La directive Solvabilité II insiste sur le risque de réserve et pousse les assureurs Non-Vie à réduire l'incertitude sur le calcul des provisions. La détermination de la méthode qui donnera les règlements futurs au plus juste avec une faible incertitude, est essentielle.

Dans la réalité, les actuaires utilisent plusieurs méthodes complémentaires, afin d'avoir une vision au plus juste du risque et des règlements des flux futurs.

## 1.4 Les méthodes pour l'Assurance Non-Vie

Cette partie est une synthèse des méthodes utilisées sur le marché de l'Assurance Non Vie par type de sinistres et de garanties. Nous chercherons à expliquer le choix de l'utilisation des méthodes selon les données et le risque.

Les opérations d'assurance sont classifiées par branche d'activité dans le droit français ou par risques dans Solvabilité II (voir Annexe C).

Cependant, on peut séparer l'assurance non-vie en trois grandes catégories :

- **Les assurances de choses ou de biens** : les biens de l'assuré ou son patrimoine (branches 3 à 9, 16 et 17) ont subi un préjudice : accidents, vols, incendie et autres dommages. Les intérêts économiques ou financiers de l'assuré peuvent également être couverts comme pour la perte d'exploitation ou de loyers. Pour ce type d'assurance, le bénéficiaire correspond généralement au souscripteur.
- **Les assurances de responsabilité** (branches 10 à 13): l'assuré subi un préjudice financier suite à des dommages causés à un tiers et dont il est responsable juridiquement. Les victimes sont dédommées par l'assureur à la place de l'assuré. Le bénéficiaire est donc toujours un tiers.
- **Les assurances de personnes** (branches 1 et 2) : la personne physique de l'assuré a subi un préjudice suite à un accident corporel ou à une maladie.

Les assurances de biens et de responsabilité ainsi que les frais de santé de l'assurance de personnes sont basées sur le principe d'indemnisation. Pour l'assurance de personnes, le principe forfaitaire est appliqué. Le montant de réparation du préjudice est fixé contractuellement.

#### 1.4.1 Assurances de personnes et Responsabilité Civile Corporelle (RCC)

Dans notre étude, on regroupe les garanties corporelles et RCC. Ces garanties donnent lieu à des prestations similaires : remboursements des frais de santé, rentes invalidité, indemnités journalières, ...

- **Arrêt de Travail : Rentes Invalidité et Indemnités Journalières**

Pour l'arrêt de travail, l'assureur indemnise l'assuré, ou un tiers dans le cas de la RC, d'une perte de salaire due à une incapacité ou à une invalidité qui l'empêche d'exercer une activité professionnelle. En France, la plupart des gens sont couverts par la Sécurité Sociale, mais des contrats de complémentaire santé peuvent être souscrits.

Les données nécessaires au provisionnement sont la date de survenance du sinistre et les informations du sinistré (âge, sexe, ancienneté, montant de rémunération, ...). Ces branches de provisionnement sont des branches longues, c'est-à-dire que le déroulement dépasse généralement 5-10 ans.

Le provisionnement des sinistres survenus et connus se fait tête par tête à l'aide de lois biométriques (loi de maintien et loi de passage). Ces lois sont des tables réglementaires élaborées par le BCAC ou des tables d'expérience certifiées et propres à l'entreprise. Elles définissent selon l'âge de l'assuré et l'ancienneté du sinistre, la probabilité de se maintenir dans l'état (incapacité ou invalidité) ou de passer en invalidité. Il existe aussi des tables de mortalité pour provisionner les rentes de conjoint et d'éducation (semblables à de l'assurance vie).

**Exemple :**

Assuré	Date de naissance	Date de survenance	Date d'entrée en Inval	Date de fin de franchise Incap	Date de fin de franchise Inval	Date de fin Incap	Age retraite	Type de garantie	Montant II	Montant de rente Inval	PM Incap	PM Inval en attente	PM Inval
1	21/10/1978	30/10/2017	-	29/11/2017	27/01/2021	28/01/2018	62	Incap-Inval	61 €	8 320 €	1555,09	12210,41	
2	26/04/1958	06/04/2007	04/07/2010	06/05/2007	04/07/2010	05/07/2007	62	Inval		8316			18726,342

**Figure 8: Base de données Incapacité-Invalidité**

Age	Mois d'ancienneté dans le sinistre						
	0	1	...	11	...	22	23
23 ou -	10000	2842		187		87	84
24	10000	2931		183		84	82
25	10000	3080		216		102	97
...				...			
43	10000	4473	...	892	...	488	464
...				...			
63	10000	5542		1494		940	894
64	10000	5565		1465		940	895
65	10000	5588		1437		940	895
66	10000	5611		1408		940	896

**Figure 9: Table BCAC de maintien en Incapacité (nombre de personnes encore présentes)**

L'assuré n°1 est en incapacité. Sachant qu'il est né le 21/10/1978 et que la date de survenance de son sinistre est le 30/10/2017, son âge au moment du sinistre est de 39 ans.

Sachant que la date de fin de franchise en incapacité (29/11/2017) est antérieure à la date d'arrêt, la date de début de provisionnement est le 31/12/2017. À cette date, l'assuré a donc 2 mois d'ancienneté. Ces informations permettent de lire la probabilité de maintien (ou de passage selon le risque) dans le tableau.

Le montant de provision est calculé en multipliant le nombre de jour restant à indemniser (date de fin d'incapacité moins la date de clôture), le montant d'IJ et la probabilité de maintien.

Pour l'assuré n°2, la probabilité est obtenue de la même manière en prenant en compte la date d'entrée en invalidité.

Les différentes dates (survenance, naissance, début et fin d'incapacité) permettent de déterminer l'âge à la date de survenance, l'ancienneté dans le sinistre et la période restant à couvrir. Cela permet de lire dans les tables BCAC, la probabilité de rester en incapacité ou en invalidité pour les PM d'incapacité ou d'invalidité et la probabilité de passage pour les PM d'invalidité en attente. Cette probabilité est ensuite multipliée au montant de rente attribué à l'assuré.

Afin de déterminer les IBNR, des triangles de provisionnement sont réalisés pour obtenir le nombre ou le montant ultime des sinistres.

$$IBNR = Provision \text{ à l'ultime} - Provision \text{ sinistres connus} - Règlements$$

**Exemple :**

Le nombre de sinistres à l'ultime est calculé à l'aide d'un triangle de nombres.

Triangle nombre de sinistres projeté		0	1	2	3	4	5	6
2016	2	196	464	498	506	508	508	508
2016	3	218	458	474	484	486	488	488
2016	4	238	476	502	504	508	510	510
2017	1	238	464	480	482	484	486	486
2017	2	270	488	496	502	505	506	506
2017	3	272	434	455	461	463	464	464
2017	4	310	606	635	643	646	648	648

Coefficients	1,955	1,048	1,013	1,005	1,003	1,000
--------------	-------	-------	-------	-------	-------	-------

**Figure 10: Triangle de nombres de sinistres en Arrêt de Travail**

Les coefficients de passage peuvent être calculés de différentes manières selon la méthode choisie. Dans l'exemple présent, les coefficients sont obtenus avec la méthode de Chain Ladder (cf. partie 3.2.2).

Le triangle de nombres est complété à l'aide des coefficients de déroulement.

Année	Trimestre	Ultimes (1)	Connus (2)	IBNR (3)=(1)-(2)	Coût Moyen (4)	Provision IBNR (5)=(4)*(3)
2016	2	508,00	508,00	-	4 944	-
2016	3	488,20	488,00	0,20	4 944	1 006
2016	4	509,76	508,00	1,76	4 944	8 679
2017	1	485,85	482,00	3,85	4 944	19 048
2017	2	506,30	496,00	10,30	4 944	50 934
2017	3	464,24	434,00	30,24	4 944	149 497
2017	4	648,28	310,00	338,28	4 944	1 672 570
<b>Total</b>		<b>3 610,63</b>	<b>3 226,00</b>	<b>384,63</b>	<b>4 944</b>	<b>1 901 735</b>

**Figure 11: Tableau de calcul de la provision d'IBNR**

On obtient ici le montant de provision d'IBNR en appliquant la formule suivante :

$$\text{Prov IBNR} = \text{Coût Moyen} * (\text{Nb Ultime} - \text{Nb Connus})$$

Pour le trimestre 2 de l'année 2017, le nombre de sinistres à l'ultime est de 506,30 et de connus de 496 ce qui donne un nombre d'IBNR de  $506,30 - 496 = 10,30$ .

Pour obtenir le montant d'IBNR, le nombre de sinistres est multiplié au coût moyen :  $10,30 * 4 944 = 50 934$ .

On peut noter que le montant de charge ultime étant déterminé par une méthode agrégée de triangle de règlements, l'intérêt du calcul ligne à ligne des sinistres en cours est réduit.

Afin de garder les avantages du calcul ligne à ligne, il serait possible de reconstituer des lignes pour les IBNR à l'aide de triangles de nombres et de *Model Point*. Ainsi la méthode ligne à ligne peut être utilisée sur l'ensemble du calcul de provisions.

- **Frais de Santé**

Les frais de santé correspondent au paiement des frais engagés par l'assuré pour des frais médicaux (remboursement ou frais forfaitaire) en complément de ce qui est remboursé par la Sécurité Sociale.

Les données nécessaires pour le provisionnement des frais de santé sont les données sur le sinistre (date de survenance, type) et les données sur l'assuré (âge, ancienneté, ...). Ce type de garantie a d'ordinaire une durée de déroulement court.

Le provisionnement est communément fait à l'aide de triangles d'écoulement. Ce type de sinistres est généralement rapidement connu et réglé, ce qui minimise l'incertitude due au calcul des IBNR. En effet, comme le montre l'exemple suivant, la grande majorité de la charge finale à payer est connue dès la première année de règlement. Le montant à provisionner est relativement faible par rapport au montant payé à l'ultime.

**Exemple : Données de règlements**

Les données utilisées sont l'année de survenance, les dates et les montants des paiements par assuré. Elles sont ensuite sommées et agrégées dans ce type de tableau.

Surv/Dev		Année de developpement				Total
		1	2	3	4	
Année de survenance	2014	91 611 352	584 032	43 268	1 721	92 238 652
	2015	89 157 342	599 249	18 005		89 774 596
	2016	98 506 756	472 472			98 979 228
	2017	75 742 469				75 742 469

**Figure 12: Tableau de règlements (non cumulés) par année de survenance et charge à l'ultime**

Les sinistres survenus en 2014 sont entièrement réglés à la fin de la quatrième année de développement pour un montant final global de 92 238 652 €. On peut remarquer que plus de 99% du règlement final est payé la première année.

Un provisionnement ligne à ligne serait possible sur ce type de sinistres, les assureurs ayant généralement les données. Néanmoins, la rapidité de règlement de ces sinistres rend ce type de méthodes sans intérêt complémentaire par rapport aux méthodes basées sur des triangles.

Cependant, l'une des problématiques importantes du provisionnement des frais de santé est la prise en compte de l'inflation. Les frais médicaux sont un domaine dans lequel l'inflation est plus importante que l'inflation nationale.

### 1.4.2 Assurances de choses ou de biens et Responsabilité Civile Matérielle

L'assurance de choses ou de biens englobe les corps de véhicules, les marchandises transportées, l'incendie et éléments naturels, les autres dommages aux biens. Cela concerne les biens ou la responsabilité civile de particuliers mais également d'entreprise.

Les méthodes de provisionnement classiquement utilisées pour provisionner ce type de garantie sont des méthodes utilisant les cadences appliquées aux nombres, charges et paiements. Les sinistres exceptionnels et les événements Catastrophes Naturelles étant traités à part.

Cependant, les actuaires sont confrontés à certaines problématiques comme l'accélération des cadences de règlements dues à l'amélioration des systèmes de gestion, ou le fait d'avoir un marché très concurrentiel avec une volatilité importante d'une année à l'autre dans la branche Transport. Ces éléments peuvent être difficiles à intégrer dans certains modèles.

## Le cas des Catastrophes Naturelles

En France, la loi du 13 juin 1982 permet à toute personne ayant une assurance dommage d'être couverte en cas de dommages dus à une Catastrophe Naturelle en contrepartie d'une prime complémentaire obligatoire à son assurance dommage.

Les Catastrophes Naturelles étant des événements rares, il est compliqué de se baser sur les cadences pour l'évaluation des provisions. Les Catastrophes Naturelles sont donc fréquemment provisionnées avec une méthode fréquence/coût.

### 1.4.3 Assurance construction : RC décennale et dommages ouvrages

L'assurance construction couvre les sinistres dus à la construction et cela sur 10 ans pour le cas de la RC Décennale.

D'après le code des assurances, les assureurs doivent, en complément des PSAP, calculer des provisions pour sinistres non encore manifestés (PSNEM). Cette provision couvre les sinistres des contrats déjà souscrits mais qui n'ont pas encore eu lieu ; cette provision est constituée dès la date d'ouverture du chantier (DROC).

Comme expliqué dans BOURRY et al. [2018], le calcul des provisions prend en compte trois dimensions : survenance, développement et DROC, ce qui rend plus complexe l'évaluation des provisions.

En général, le provisionnement des sinistres passés et futurs est évalué à l'aide de triangles indépendants. Le calcul des PSAP se fait à l'aide de triangles par année de survenance et année de développement puis par l'application de méthodes classiques (Chain Ladder, Bornhuetter-Ferguson,...). Cependant, des approches différentes peuvent être utilisées, comme par exemple en utilisant le coût moyen pour renforcer le poids des avis d'experts.

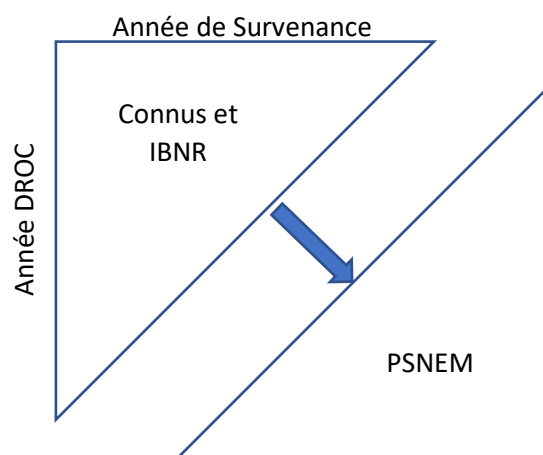


Figure 13: Triangles par année de DROC et de survenance

Un triangle par DROC et années de survenance à partir des montants à l'ultime répartis par DROC permet de déterminer la PSNEM comme le total de la projection du triangle inférieur.

Une des particularités de ces garanties est la longueur de couverture (environ 10 ans). Il est donc nécessaire de prendre en compte l'inflation des années futures et de l'intégrer dans le calcul des provisions. De plus, le très grand nombre de recours présent en garantie Construction complexifie l'évaluation de la provision. Il est donc nécessaire d'étudier les taux de recours.



Un provisionnement individuel de ces garanties construction, longues et atypiques, compléterait la vision des méthodes classiques. Cela permettrait d'intégrer des informations extérieures aux sinistres comme l'environnement économique, des informations sur le contrat et les circonstances du sinistre.

« *Le rapport du groupe de travail sur le calcul du Best Estimate en Assurance Dommage* » de l'ACPR de 2007 (cf. ACPR [2007]) présente les techniques les plus adaptées et les méthodes approchées possibles par catégories de risques. Les problématiques dues aux données et aux méthodes utilisées sont également décrites. Le tableau en annexe récapitule les éléments de cette étude.

Comme nous avons pu le remarquer dans notre étude, les méthodes basées sur les triangles et les cadences de développement des sinistres ou basées sur les fréquences et le coût moyen sont majoritairement utilisées sur le marché. Elles sont tout particulièrement appréciées par ces utilisateurs pour leur simplicité de mise en œuvre et leur rapide compréhension. Cependant, il devient important de réussir à intégrer aux modèles les avis d'experts, les tendances, les changements de gestion ou de réassurance, les données propres à chaque sinistre... Une vision individuelle des sinistres peut être une solution à l'insertion de ces éléments.

## 1.5 Évolution des méthodes

Avec le besoin accru de modèles de provisionnement plus précis, tirer parti des informations intégrées dans les données des sinistres individuels est une alternative prometteuse par rapport aux méthodes traditionnelles par triangle.

### 1.5.1 Limites et solutions des méthodes agrégées

Cependant, on peut noter depuis quelques années une prise de conscience des limites des modèles agrégés pour obtenir des évaluations robustes. Ces limites sont présentées dans BOUMEZOUED [2017].

Dans l'ensemble, ces limites sont dues à une perte d'informations lors de l'agrégation des données des sinistres individuels (par exemple, date de survenance, retard de déclaration, et date et montants des paiements).

A cause de ces pertes d'information, il est, par exemple, très compliqué d'appliquer un traité de réassurance non proportionnel sur des données agrégées.

En raison du manque de données sur les dernières périodes de développement, les erreurs d'estimation sont plus importantes.

De plus, ces méthodes ne permettent pas de modéliser le montant d'un sinistre d'un assuré en particulier, ses données personnelles ne sont donc pas prises en compte.

Enfin, une des hypothèses fortes de ces méthodes est l'indépendance des générations, ce qui n'est pas souvent vérifié dans la réalité.

Les développements récents dans les techniques de collecte, de stockage et d'analyses de données permettent le provisionnement individuel. Il est désormais devenu important de créer des modèles plus souples pour les utilisateurs opérationnels (gestionnaire de sinistres, souscripteur, réassureur, ...) pour prendre en compte les éléments suivants :

- La définition de développements spécifiques des sinistres : fréquence, flux ;
- Les changements possibles dans les produits, le contexte légal ou le traitement des déclarations ;
- L'évaluation et le suivi des risques (détecter les changements de tendances) ;
- Le calcul à part IBNR ;
- Les variables explicatives (éléments extérieurs).

Dans le domaine du provisionnement, de nombreux articles ont été écrits pour essayer de faire progresser les modèles existants et donner un cadre mathématique aux modèles.

### 1.5.2 Les évolutions académiques

L'article de MACK [1993] est un des articles fondateurs dans le cadre du provisionnement. Il fut le premier à donner une justification probabiliste de l'utilisation des modèles agrégés sous certaines conditions. Il a complété ces méthodes en déterminant un intervalle de confiance de la provision et permettant ainsi d'avoir un niveau d'incertitude.

Pour les modèles basés sur des triangles, le calcul de l'erreur d'estimation est devenu quelque chose de standard. Cependant, l'étude de l'incertitude des paramètres pour les modèles individuels est moins étudiée. Cette question est abordée dans les articles de HAASTRUP et al.[1996] et de ANTONIO et al. [2014].

L'article de BOUMEZOUED [2017] présente un modèle de provisionnement individuel qui permet de calculer séparément les sinistres connus et les IBNR. Le modèle distingue la modélisation de la survenance et de la déclaration du sinistre et des paiements futurs. Dans le cas pratique étudié, l'estimation de l'erreur de prédiction du modèle est réduite en comparaison des résultats obtenus avec le modèle de Mack.

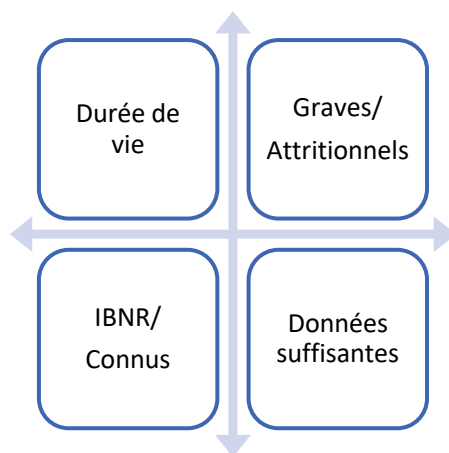
Des articles ont également été écrits sur les modèles ligne à ligne, on peut notamment citer les articles d'ARJAS [1989], JEWELL [1989], NORBERG [1993]. Ils présentent un cadre probabiliste au provisionnement ligne à ligne en introduisant la théorie des martingales et des processus ponctuels.

Plus récemment, ANTONIO et al. [2014] complètent encore les connaissances dans ce domaine en ajoutant au cadre théorique une application pratique. Ils appliquent le cadre de modélisation développé par NORBERG [1993] à un portefeuille d'assurance responsabilité civile utilisant l'estimateur du maximum de vraisemblance. Il est montré que le provisionnement individuel donne une meilleure précision que les modèles agrégés utilisés dans l'étude.

Ces articles ont enrichi les connaissances dans le domaine du provisionnement et les modèles utilisés donnant de nouvelles idées aux utilisateurs.

### Synthèse

Cette présentation générale du provisionnement permet d'avoir une vision d'ensemble des problématiques et des méthodes de provisionnement des assureurs. Le choix de la méthode de provisionnement est un exercice délicat prenant en compte de nombreux paramètres.



**Figure 14: Eléments caractéristiques des sinistres**

Les caractéristiques qui définissent les sinistres et les risques ont un impact sur le choix des méthodes de provisionnement.

		Vie du sinistre		
		IBNR	Sinistres connus	
Nature du sinistre	Type d'assurance			
	Assurance de Personne	Court Terme	Méthode de triangles	Méthodes de triangles
		Long Terme		Méthodes ligne à ligne
	IARD	Graves		Méthodes ligne à ligne
Attritionnels		Méthodes de triangles		

**Figure 15: Tableau de synthèse des méthodes**

Il semblerait que certaines méthodes soient préférées à d'autres selon ces critères :

Les méthodes agrégées et individuelles sont basées sur un principe commun : les évolutions du passé donnent des informations sur les évolutions futures. Les données passées permettent de paramétrer les modèles. En ligne à ligne, les données passées sont utilisées pour construire des tables qui sont ensuite appliquées sur les données individuellement. Pour les méthodes agrégées, les données du passé sont regroupées afin de déterminer des cadences de déroulement. Les évolutions futures sont calculées de manière agrégées à l'aide de ces cadences.

En absence d'historique suffisant sur les données, l'assureur peut utiliser la méthode de Loss Ratio ou S/P<sup>3</sup>. Cette méthode nécessite une connaissance précise du marché.

L'ensemble des méthodes demande un historique et des données sur les sinistres suffisants. Des données précises par assuré sont indispensables pour calculer les provisions avec un modèle individuel. Ces données peuvent toujours être agrégées afin d'appliquer des modèles avec des

<sup>3</sup> S/P : Sinistre/Prime

triangles. Cependant l'inverse n'est pas vrai, les données disponibles pour une méthode agrégée ne permettent pas toujours d'appliquer un modèle ligne à ligne.

En Assurance de Personnes, la distinction majeure est la durée de vie du sinistre. Un modèle individuel qui est au plus proche des données et qui prend en compte la particularité de chaque sinistre, semble plus adapté pour les sinistres longs pour lesquels l'enjeu du provisionnement est plus important étant donné leurs caractéristiques. Cependant, les données sur les sinistres sont obtenues au fur et à mesure de la vie du sinistre, ce qui peut rendre l'application de ces méthodes difficile à mettre en œuvre. A l'inverse, l'utilisation d'un modèle ligne à ligne pour des sinistres courts n'apporte généralement pas d'éléments complémentaires aux méthodes agrégées.

En IARD, le critère essentiel est celui de la distinction entre les sinistres graves et attritionnels. Les enjeux et les caractéristiques des sinistres graves semblent nécessiter une méthode ligne à ligne, afin de modéliser au mieux chaque sinistre et ses particularités. Au contraire, les sinistres attritionnels sont provisionnés à l'aide de plusieurs méthodes de triangles utilisées simultanément.

Une fois la méthode de provisionnement choisie, il faut également déterminer si elle est appliquée à l'ensemble de la population ou si une segmentation est réalisée. Généralement, les données sont, au minimum, classées par types d'activités ou peuvent être classifiées de manière très fine afin d'obtenir des ensembles homogènes de risques et avoir un modèle représentant au mieux la réalité. La segmentation se fait selon le comportement des individus de la population assurée, les cadences de règlements, les caractéristiques des produits ou la proportion de tardifs.

Cependant, en segmentant la problématique du manque de données pour obtenir des tables ou des triangles fiables peut se poser. Il est donc nécessaire d'arbitrer entre le niveau de granularités des classifications, la qualité et la quantité de données à disposition.

Dans la suite, nous chercherons à présenter des critères de choix de méthodes de provisionnement. Nous étudierons notamment les avantages et les inconvénients des méthodes ligne à ligne et agrégés avec des données globales ou segmentées. Afin d'illustrer nos propos, nous utiliserons des données d'arrêt de travail traditionnellement provisionnées en ligne à ligne, ainsi l'ensemble des méthodes pourront être appliquées à nos données.

## Chapitre 2 :

# Méthode de provisionnement en prévoyance

Les données utilisées pour notre étude sont des données d'incapacité.

### 2.1 La prévoyance

Les contrats de prévoyance couvrent les personnes en cas de maladie, de décès, d'incapacité ou d'invalidité. Ces risques entraînent une impossibilité temporaire ou définitive d'exercer une activité professionnelle.

On peut séparer en deux grandes catégories :

- Les garanties en cas de décès ;
- Les garanties en cas d'accident ou de maladie.

Le contrat de prévoyance peut être souscrit par un individu, une entreprise ou une association. Il faut donc différencier la prévoyance collective et la prévoyance individuelle.

L'accident ou la maladie d'un assuré provoque une interruption totale ou partielle de l'activité professionnelle. Cela induit une baisse des revenus de l'assuré qui est compensée par l'assureur partiellement ou dans sa globalité.

Un contrat de prévoyance donne lieu aux garanties suivantes :

- Le remboursement des frais de santé en complément du montant pris en charge par la Sécurité Sociale ;
- Le versement d'une rente incapacité, invalidité ou dépendance ;
- Le versement d'un capital ou d'une rente de conjoint ou d'éducation en cas de décès.

En cas d'arrêt de travail, il faut distinguer deux états : l'incapacité et l'invalidité. Ces derniers sont définis selon des critères de la Sécurité Sociale. Dans les deux cas, cela correspond à une impossibilité de travailler dû à des raisons médicales pour un temps donné ou définitivement. L'incapacité dure au maximum 3 ans avec possibilité de revenir à un état valide. L'invalidité est un état définitivement acquis et généralement plus grave que l'incapacité. Dans certains cas, une activité partielle est possible.

Pour plus d'information sur la prévoyance, se référer à PLANCHET et al. [2011].

Dans ce mémoire nous nous intéresserons uniquement à la garantie d'**incapacité**.

## 2.2 Présentation des données

### 2.2.1 La base de données

Les données utilisées pour cette étude proviennent d'une source publique<sup>4</sup> (site « Ressources Actuarielles »). Cependant, certaines spécificités sur les contrats nous sont inconnues car aucune information complémentaire accompagne les données. Certains choix de retraitements de données ont donc été faits en se basant sur le comportement d'un sinistre d'incapacité classiquement observé et certaines particularités des données n'ont pu être expliquées.

La base des données est constituée des variables suivantes :

VARIABLES	DESCRIPTIF
Sexe	Sexe de l'assuré: Homme (H) / Femme (F)
CSP	Catégorie socio-professionnelle de l'assuré: Cadres (CADRES)/ Non-Cadres (NONCAD) / Ensemble du personnel (ENSPER)
ModeSouscription	Mode de souscription du contrat : Standard (STD) / Spécifique (SPE) / Convention Collective (CNV)
TypeArret	Nature du sinistre: Maladie/ Accident du Travail
DateNaissance	Date de naissance de l'assuré
DateSurvenance	Date de survenance du sinistre
DateEntree	Date de début d'indemnisation (Date du premier versement)
DateSortie	Date de fin d'indemnisation
MotifReprise	Statut du sinistre/ Motif de sortie de l'incapacité

Figure 16: Liste des champs

### 2.2.2 Le traitement des données

Le traitement des données est une étape indispensable qui permet de détecter les anomalies et vérifier la cohérence. Pour construire la base, nous avons suivi le processus suivant :

#### *Etape 1 : Sélection des sinistres par année de survenance*

Les données couvrent la période de 2002 à 2008. L'incapacité dure au maximum 1095 jours, soit trois ans. Notre période d'observation permet donc d'observer des sinistres dans leur intégralité. Une partie des données (60%) seront utilisées pour construire les différents modèles d'estimation de provisions et l'autre partie permettront de comparer les prévisions des modèles avec les données observées dans la réalité.

#### *Etape 2 : Ajout de l'âge de l'assuré à la date de survenance du sinistre*

L'âge de l'assuré à la date de survenance est indispensable pour pouvoir utiliser les tables réglementaires (variable en année) :

$$\text{Age à la Survenance} = (\text{Date de survenance} - \text{Date de Naissance})/365,25$$

#### *Etape 3 : Ajout des informations sur le sinistre :*

L'ancienneté à l'entrée correspond à la période de franchise (variables en jours) :

$$\text{Ancienneté à l'entrée} = \text{Date d'Entrée} - \text{Date de survenance} .$$

L'ancienneté à la sortie représente le nombre de jours passé en Incapacité (y compris la franchise) pour l'assuré (variables en jours).

<sup>4</sup> <http://www.ressources-actuarielles.net/C1256F13006585B2/0/73F372CA4B4A9B62C1257A8800686046>

$$\text{Ancienneté à la sortie} = \text{Date de Sortie} - \text{Date de survenance} .$$

#### **Etape 4 : Ajout de l'indicateur de censure :**

Cette variable est une indicatrice qui est égale à 0 si le sinistre est en cours et 1 si le sinistre est clos.

#### **Etape 5 : Suppression des données incohérentes**

Certaines anomalies peuvent exister dans les données, des tests ont été réalisés pour supprimer les lignes avec des erreurs :

- ✓ L'ancienneté à l'entrée doit être supérieure à 0 ;
- ✓ L'ancienneté à l'entrée doit être inférieure à l'ancienneté à la sortie ;
- ✓ L'ancienneté à la sortie doit être inférieure à 36 mois (1095 jours). Pour des questions d'arrondi la date de fin théorique (date de survenance plus ans) est également comparée à la date de sortie. Si cette dernière est supérieure à la date de fin théorique, le sinistre est supprimé.

#### **Etape 6 : Sélection des sinistres par âge à la date de survenance**

Une sélection est réalisée de manière à écarter les assurés qui ont un âge à la date de survenance du sinistre strictement inférieur à 23 ans (âge moyen de début de carrière) et supérieur à 62 ans (âge de fin des tables du BCAC 2010).

	Nombres de lignes
Base initiale	560 725
Base retraitée: Anomalie sur les anciennetés	542 619
Base retraitée: Anomalie sur les dates de sortie	542 261
Base retraitée: Anomalie sur l'âge	520 186
<b>Base retraitée</b>	<b>520 186</b>

**Figure 17: Evolution de la volumétrie de la base**

A la suite des différents retraitements de données, 7% de la population initiale a été supprimé.

La base nettoyée va nous permettre de faire une étude statistique des données afin d'obtenir les caractéristiques du portefeuille étudié.

#### **Ajout de variables**

Pour les différentes méthodes utilisées dans la suite de l'étude, il est nécessaire d'ajouter certaines variables :

- Nombre de jours Indemnisés :

$$Nb\text{JoursIndem} = \text{DateSortie} - \text{DateEntree},$$

Une autre formule pourrait être également utilisée :

$$Nb\text{JoursIndem} = \text{DateSortie} - \text{DateEntree} + 1.$$

Ainsi, quand les deux dates sont égales, un jour est indemnisé.

- Classe d'âge

Moins de 35 ans :  $Age\text{Survenance} \leq 30$ ,

Entre 35 et 45 ans :  $30 < Age\text{Survenance} \leq 45$ ,

Entre 45 et 55 ans :  $45 < Age\text{Survenance} \leq 55$ ,

Plus de 55 ans :  $55 < AgeSurvenance$ .

- Classe de durée

Moins de 1 an :  $AncSortie/365,25 \leq 12$ ,

Entre 1 et 2 ans :  $12 < AncSortie/365,25 \leq 24$ ,

Entre 2 et 3 ans :  $24 < AncSortie/365,25 \leq 36$ ,

Plus de 3 ans :  $36 < AncSortie/365,25$ ,

- Classe de franchise

Pas de franchise :  $AncEntree = 0$ ,

]0 ;30] :  $0 < AncEntree \leq 30$ ,

]30 ;60] :  $30 < AncEntree \leq 60$ ,

]60 ;90] :  $60 < AncEntree \leq 90$ ,

]90 ;365] :  $90 < AncEntree \leq 365$ ,

]365 ;max[ :  $365 < AncEntree$ .

On obtient donc une table de données avec les variables suivantes :

VARIABLES	DESCRIPTIF
Sexe	Sexe de l'assuré: Homme (H) / Femme (F)
CSP	Catégorie socio-professionnelle de l'assuré: Cadres (CADRES)/ Non-Cadres (NONCAD) / Ensemble du personnel (ENSPER)
ModeSouscription	Mode de souscription du contrat : Standard (STD) / Spécifique (SPE) / Convention Collective (CNV)
TypeArret	Nature du sinistre: Maladie/ Accident du travail
DateNaissance	Date de naissance de l'assuré
DateSurvenance	Date de survenance du sinistre
DateEntree	Date de début d'indemnisation (Date du premier versement)
DateSortie	Date de fin d'indemnisation
MotifReprise	Statut du sinistre/ Motif de sortie de l'incapacité
AgeSurvenance	Age de l'assuré à la date de survenance du sinistre
AncEntree	Période de franchise
AncSortie	Ancienneté à la sortie de l'incapacité
NbJoursIndem	Nombre de jours indemnisés
non_censure	Indicatrice de censure
classeAge	Classe pour l'âge de survenance
classeDuree	Classe pour l'ancienneté à la sortie
classeFranchise	Classe pour la période de franchise
AnneeSurvenance	Année de la date de survenance

**Figure 18: Récapitulatif des variables**

Les variables essentielles à la construction de la table sont :

- La date de naissance de l'assuré ;
- La date de survenance du sinistre ;
- La date de sortie.

Les variables dites explicatives, pouvant avoir un impact sur le maintien en incapacité :

- Le sexe ;



- La CSP ;
- Le mode de souscription ;
- Le type d'arrêt ;

Des statistiques sur l'ensemble de ces variables vont permettre de déterminer l'influence qu'elles ont sur le taux de maintien et de détecter les anomalies.

### 2.2.3 Statistiques descriptives

Des statistiques descriptives sur l'ensemble de la population retraitée sont présentées ci-après afin de décrire plus précisément le contenu de la base.

#### Variables numériques et dates

Champs	Minimum	Moyenne	Maximum
Date de Sinistre	03/01/2002	08/11/2006	30/12/2008
Premier Jour Indemnisé	04/01/2003	04/01/2007	30/12/2008
Fin du règlement	02/01/2005	30/01/2007	31/12/2008
Date de naissance	22/04/1940	25/03/1964	26/12/1985
Nombre de jour en Incapacité	1	83	1096
Nombre de jour indemnisé	1	26	1095
Âge à l'arrêt	23,00	42,62	62,00

Figure 19: Variables numériques et dates

#### Statistiques descriptives par année de survenance

La ventilation des sinistres par année de survenance montre un volume significatif :

Année de survenance	Nombre	%	Âge moyen	Nombre moyen de Jours Indemnisés
2002	1 760	0,3%	48,2	33,8
2003	4 361	0,8%	47,9	39,0
2004	18 906	3,6%	44,6	42,6
2005	122 912	23,6%	42,3	26,2
2006	119 795	23,0%	42,4	27,6
2007	126 212	24,3%	42,5	25,6
2008	126 240	24,3%	42,6	20,9
<b>Total</b>	<b>520 186</b>	<b>100,0%</b>	<b>42,6</b>	<b>25,8</b>

Figure 20: Statistiques par année de survenance

On constate que le nombre de survenance se stabilise à partir de 2005 (environ 24%). Les années antérieures sont peut représentatives. Les analyses suivantes permettront de statuer sur la conservation de ces années.

Années	Année du premier règlement								Total
	2002	2003	2004	2005	2006	2007	2008		
2002	-	-	-	1 760	-	-	-	-	<b>1 760</b>
2003	-	3	-	2 850	1 508	-	-	-	<b>4 361</b>
2004	-	-	29	15 090	2 569	1 218	-	-	<b>18 906</b>
2005	-	-	-	106 944	12 750	2 201	1 017	-	<b>122 912</b>
2006	-	-	-	-	103 493	12 968	3 334	-	<b>119 795</b>
2007	-	-	-	-	-	109 239	16 973	-	<b>126 212</b>
2008	-	-	-	-	-	-	126 240	-	<b>126 240</b>

Figure 21: Nombres de sinistres par année de survenance et année du premier règlement

On peut noter certaines particularités de la base : les sinistres survenus sur les années 2002 à 2004 sont peu nombreux et ont une durée de franchise très importante liée à une anomalie potentielle dans les données (franchise supérieure à 90 jours).

### Ancienneté des arrêts

Les arrêts ont été ventilés par ancienneté. La durée d'arrêt est majoritairement inférieure à 1 an au sein de ce portefeuille (94%).

Ancienneté de l'arrêt	Nombre	%
Moins de 1 an	490 078	94,2%
Entre 1 an et 2 ans	19 004	3,7%
Entre 2 an et 3 ans	11 104	2,1%
Plus de 3 ans	0	0,0%
<b>Total</b>	<b>520 186</b>	<b>100,0%</b>

**Figure 22: Statistiques sur l'ancienneté des arrêts**

Années/Durée d'indemnisation		Moins de 1 an	Entre 1 an et 2 ans	Entre 2 an et 3 ans	Plus de 3 ans	Total
Année de survenance	2002	0	2	1 758	0	<b>1 760</b>
	2003	12	1 789	2 560	0	<b>4 361</b>
	2004	12 766	3 841	2 299	0	<b>18 906</b>
	2005	116 915	3 752	2 245	0	<b>122 912</b>
	2006	113 628	3 925	2 242	0	<b>119 795</b>
	2007	120 517	5 695	0	0	<b>126 212</b>
	2008	126 240	0	0	0	<b>126 240</b>

**Figure 23: Nombres de sinistres par année de survenance et durées de règlement**

Les sinistres des années 2002 et 2003 durent plus d'un an, contrairement aux autres années durant majoritairement moins d'un an. Ces deux années semblent atypiques. Le nombre de sinistres d'une durée de plus de deux ans est cependant relativement constant sur l'ensemble des années de survenance.

### Répartition des franchises

La plupart des sinistres sont associés à une franchise de moins de 30 jours, cependant, les franchises longues ne sont pas négligeables :

Franchise	Nombre	%
]0,30]	55 554	10,7%
]30,60]	32 646	6,3%
]60,90]	22 161	4,3%
]90,365]	47 242	9,1%
>365	25 401	4,9%
Pas de Franchise	337 182	64,8%
<b>Total</b>	<b>520 186</b>	<b>100,0%</b>

**Figure 24: Statistiques par classes de franchise**

Les sinistres avec des franchises de plus de 90 jours sont des anomalies.

Années / Franchise		Franchise						Total
		]0,30]	]30,60]	]60,90]	]90,365]	>365	Pas de Franchise	
Année de survenance	2002	0	0	0	0	1 760	0	<b>1 760</b>
	2003	1	0	0	0	4 358	2	<b>4 361</b>
	2004	2 678	1 907	1 629	6 976	5 706	10	<b>18 906</b>
	2005	13 299	7 314	4 943	10 190	4 688	82 478	<b>122 912</b>
	2006	13 144	7 562	5 112	10 176	4 729	79 072	<b>119 795</b>
	2007	13 891	7 910	5 403	10 926	4 160	83 922	<b>126 212</b>
	2008	12 541	7 953	5 074	8 974	0	91 698	<b>126 240</b>

**Figure 25: Nombres de sinistres par année de survenance et classe de franchise**

On peut noter une stabilisation du nombre de sinistres par classe à partir de l'année 2005. Cet élément renforce le fait que les années 2002 à 2004 soient atypiques.

### Statistiques descriptives par âge à la date de survenance

La ventilation des sinistres par tranche d'âge montre que la population est répartie de manière homogène sur les trois premières classes d'âges.

Age à la date de survenance	Nombre	%
Moins de 35 ans	135 437	26,0%
Entre 35 et 45 ans	153 721	29,6%
Entre 45 et 55 ans	167 964	32,3%
Plus de 55 ans	63 064	12,1%
<b>Total</b>	<b>520 186</b>	<b>100,0%</b>

**Figure 26: Statistiques selon l'âge à la date de survenance**

Années/Âge à la survenance		Moins de 35 ans	% population de l'année	Entre 35 et 45 ans	% population de l'année	Entre 45 et 55 ans	% population de l'année	Plus de 55 ans	% population de l'année	Total
Année de survenance	2002	138	7,84%	411	23,35%	821	46,65%	390	22,16%	<b>1 760</b>
	2003	381	8,74%	987	22,63%	2 070	47,47%	923	21,16%	<b>4 361</b>
	2004	3 827	20,24%	5 043	26,67%	7 005	37,05%	3 031	16,03%	<b>18 906</b>
	2005	33 249	27,05%	36 677	29,84%	38 790	31,56%	14 196	11,55%	<b>122 912</b>
	2006	31 891	26,62%	35 832	29,91%	38 038	31,75%	14 034	11,72%	<b>119 795</b>
	2007	33 207	26,31%	37 386	29,62%	40 339	31,96%	15 280	12,11%	<b>126 212</b>
	2008	32 744	25,94%	37 385	29,61%	40 901	32,40%	15 210	12,05%	<b>126 240</b>

**Figure 27: Répartition de la population par année de survenance et classe d'âge à la survenance**

Les années de survenance 2005 à 2008 semblent similaires dans la répartition de leur population par tranche d'âge. L'année 2004 ressemble relativement à ces années dans une proportion plus faible.

A l'inverse, les trois premières années sont assez différentes.

A la suite de ces analyses, les années 2002 à 2004 sont atypiques aussi bien en nombre qu'en répartition comparativement aux années postérieures. Ces lignes vont être supprimées dans suite de l'étude. Il en sera de même pour les franchises de plus de 90 jours.

Evolution de la base de données	Nombres de lignes
Base initiale	560 725
Base retraitée	520 186
Base retraitée: Anomalie sur la franchise	447 543
Base retraitée: Selection des années de survenance 2005 à 2008	441 316
<b>Base finale</b>	<b>441 316</b>

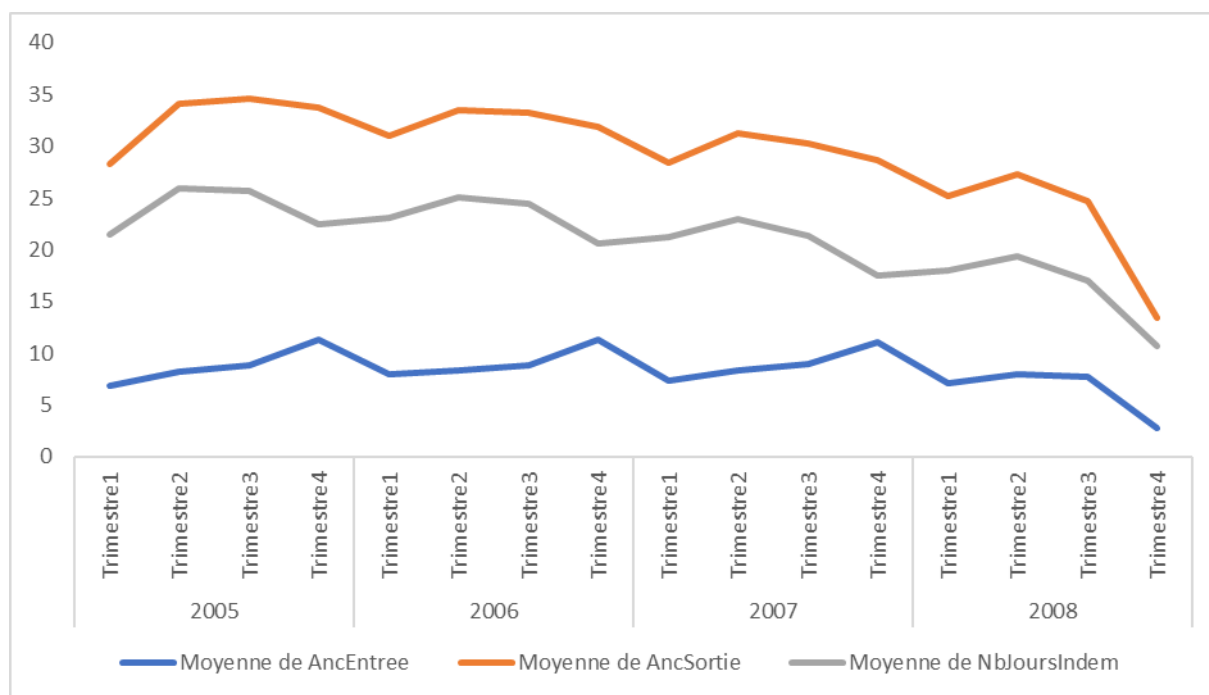
**Figure 28: Evolution de la volumétrie de la base**

À la suite de ce second retraitement, 20 % des sinistres de la base initiale ont été supprimés. On note que les années de survenance 2002 à 2004 et les anomalies sur les franchises auraient biaisé les résultats de l'étude si elles avaient été gardées.

### Etude des périodes d'indemnisation par année

Afin d'avoir une vision plus précise des caractéristiques des sinistres, l'ancienneté en début et fin d'indemnisation ainsi que le nombre de jours indemnisés sont représentés. Sachant que les sinistres provisionnés dans les parties suivantes seront uniquement les sinistres clos, l'analyse suivante ne prend en compte que ces sinistres.

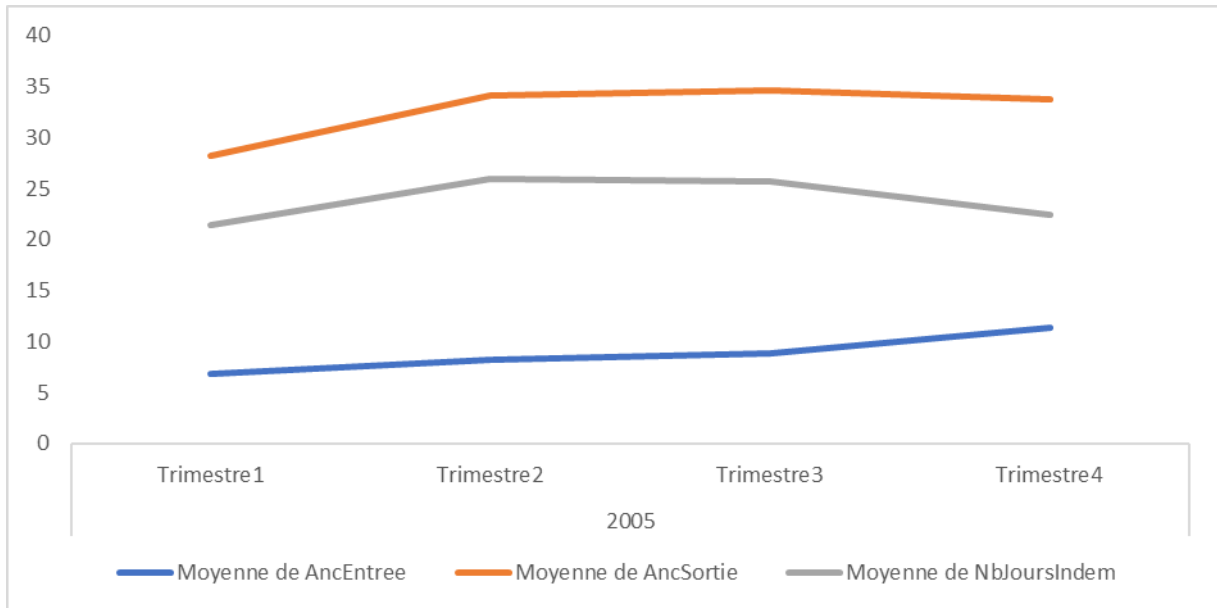
**Figure 29: Représentation de l'ancienneté moyenne à la fin de la franchise, de sortie et du nombre de jours moyen indemnisés par mois de survenance (2005 à 2008)**



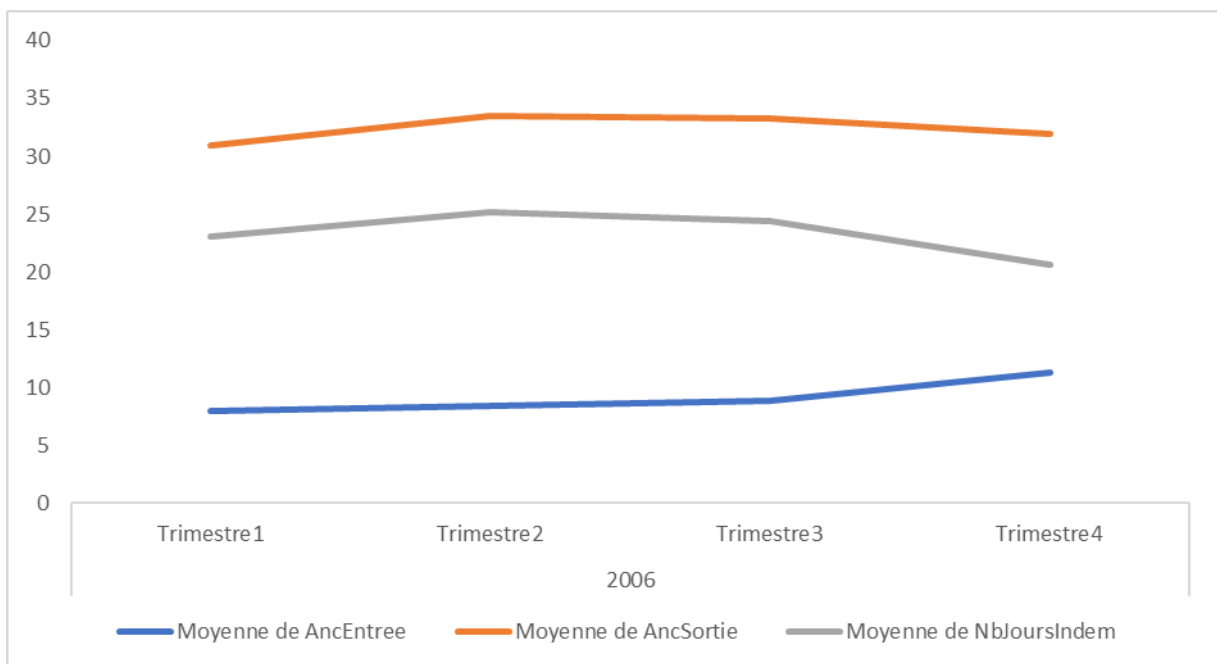
On peut noter à la lecture de ce graphique que les années 2005 à 2007 semblent relativement similaires avec une durée moyenne de nombres de jours indemnisés entre 20 et 25 jours. On remarque que l'année 2008 se distingue par une baisse importante sur les deux derniers trimestres de l'ancienneté à la sortie et d'une baisse moins importante de l'ancienneté à l'entrée : cela entraîne une baisse du nombre de jours indemnisés baissant en-dessous des 15 jours.

Nous pourrions observer avec plus de précision les années séparément avec les graphiques suivants.

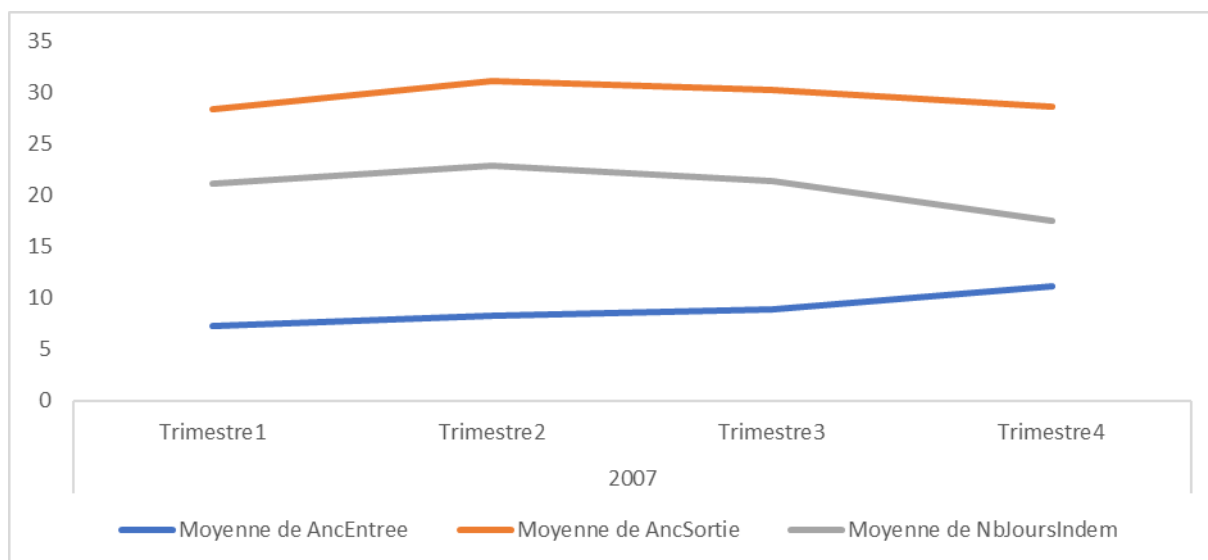
**Figure 30 : Représentation de l'ancienneté moyenne à la fin de la franchise, de sortie et du nombre de jours moyen indemnisés par mois de survenance pour l'année 2005.**



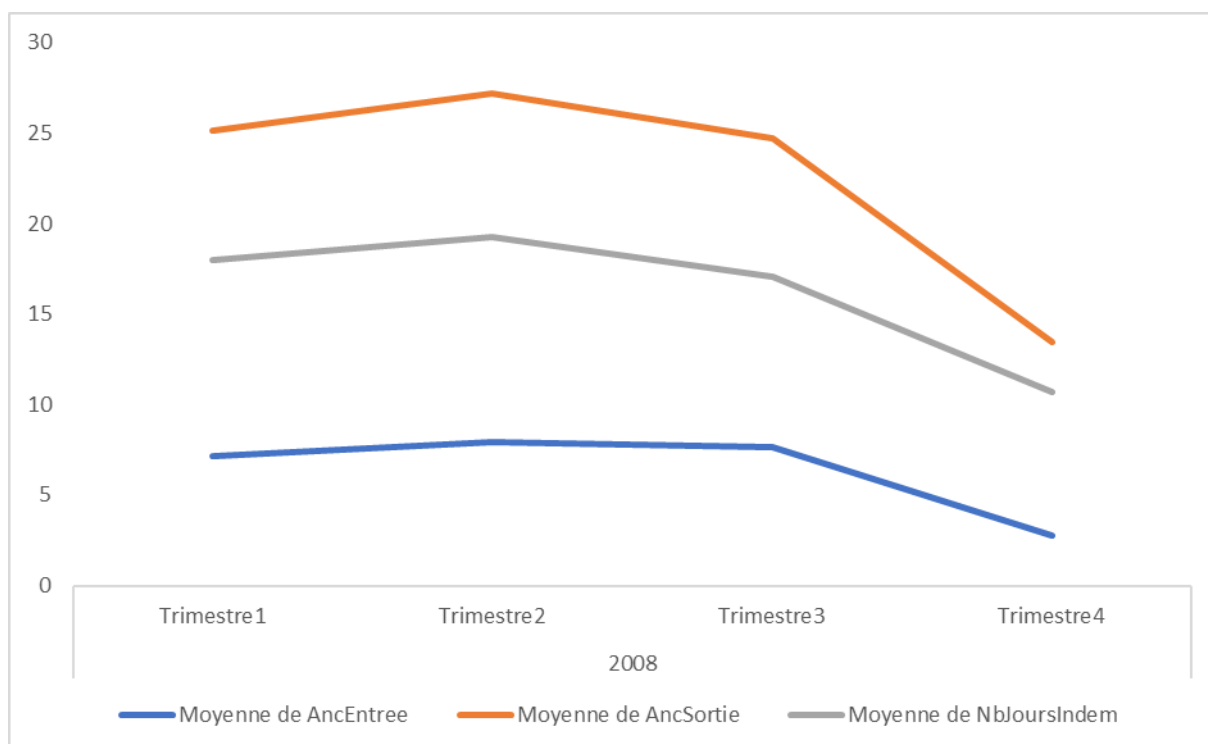
**Figure 31 : Représentation de l'ancienneté moyenne à la fin de la franchise, de sortie et du nombre de jours moyen indemnisés par mois de survenance pour l'année 2006.**



**Figure 32 : Représentation de l'ancienneté moyenne à la fin de la franchise, de sortie et du nombre de jours moyen indemnisés par mois de survenance pour l'année 2007.**



**Figure 33 : Représentation de l'ancienneté moyenne à la fin de la franchise, de sortie et du nombre de jours moyen indemnisés par mois de survenance pour l'année 2008.**



Le graphique ci-dessous permet de bien observer la baisse précédemment observée du nombre de jours indemnisés notamment sur le dernier trimestre.

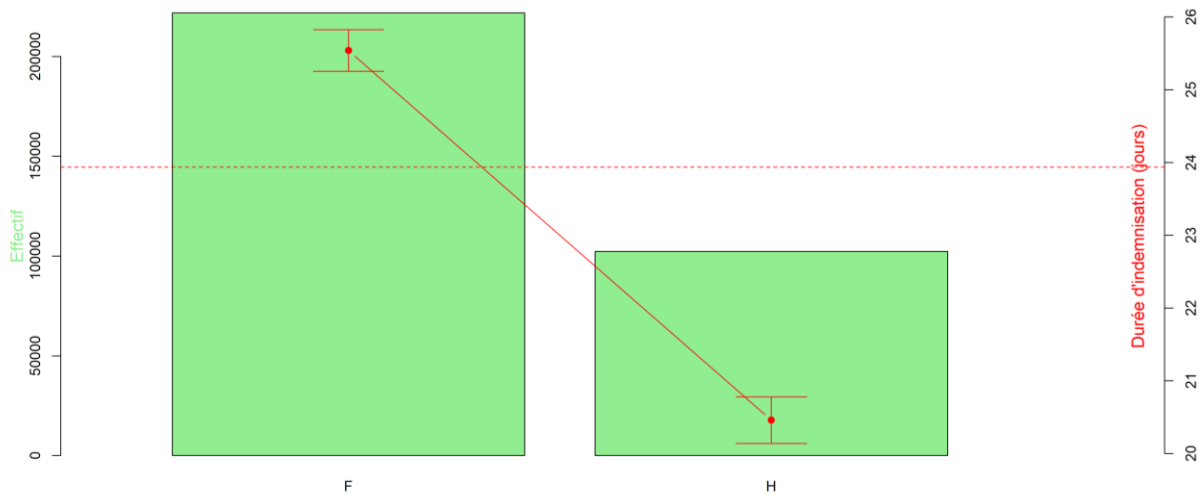
L'analyse faite sur le graphique représentant les quatre années observées est renforcée par l'étude des années séparément. L'année 2008 se distingue bien des années antérieures.

A l'aide de ses éléments, on peut donc conclure que l'année 2008 diffère des années 2005 à 2007. Afin d'avoir une étude d'année similaire permettant de comparer les résultats obtenus, il est décidé de supprimer cette année 2008.

## Analyse de la base de 2005 à 2007

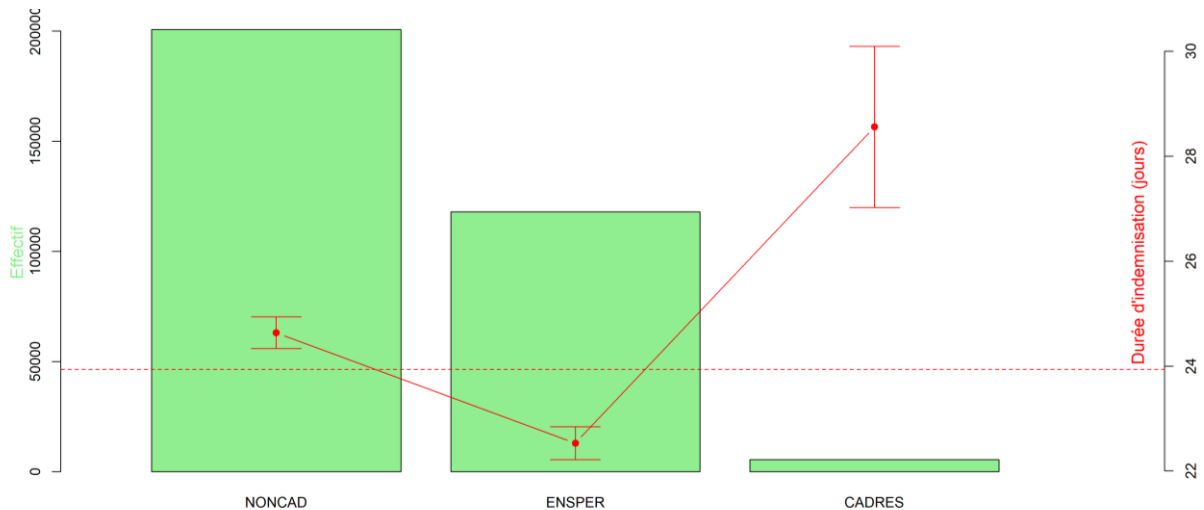
Ces analyses de volumétrie sont complétées par quelques mesures, dans un cadre univarié, de l'impact des variables de segmentation disponibles sur la durée de maintien en incapacité. Cette étude est menée sur la base retraitée finale (324 050 lignes). L'effectif est représenté en vert et la durée d'indemnisation en rouge (point moyen et intervalle de confiance à 95%).

**Figure 34: Impact du sexe sur la durée de maintien**



Le sexe impacte la durée de maintien : Les femmes restent plus longtemps en incapacité. Cependant, l'écart de nombre de jours entre les deux classes est au maximum de quatre jours. Cela reste relativement faible par rapport aux 1 096 jours d'arrêt possibles.

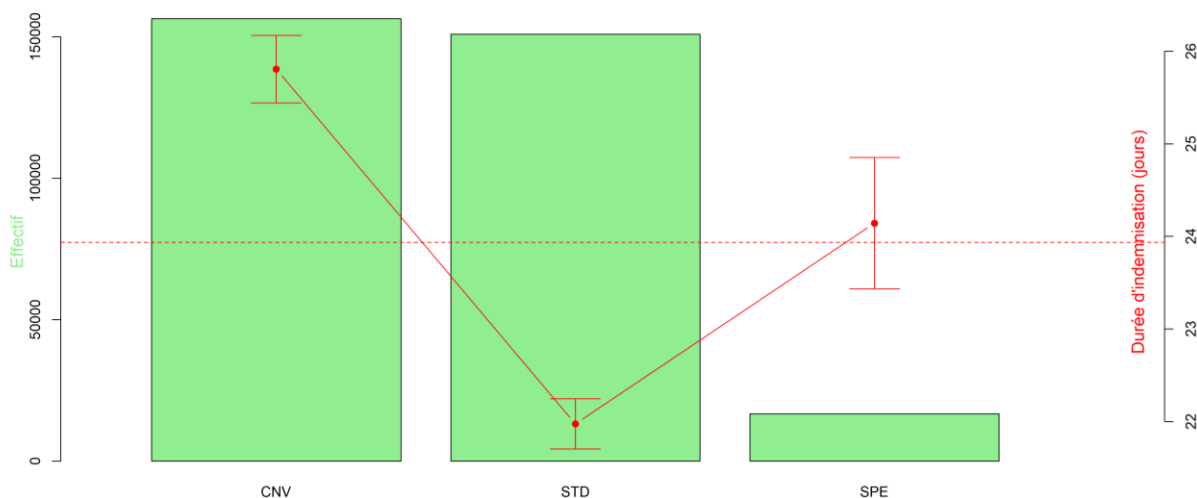
**Figure 35: Impact de la CSP sur la durée de maintien<sup>5</sup>**



Les cadres sont en proportion moins importante que les deux autres CSP représentées, mais ont une durée d'indemnisation plus longue. On note que la durée d'indemnisation reste relativement similaire sur les trois classes (variation de 24 à 30 jours).

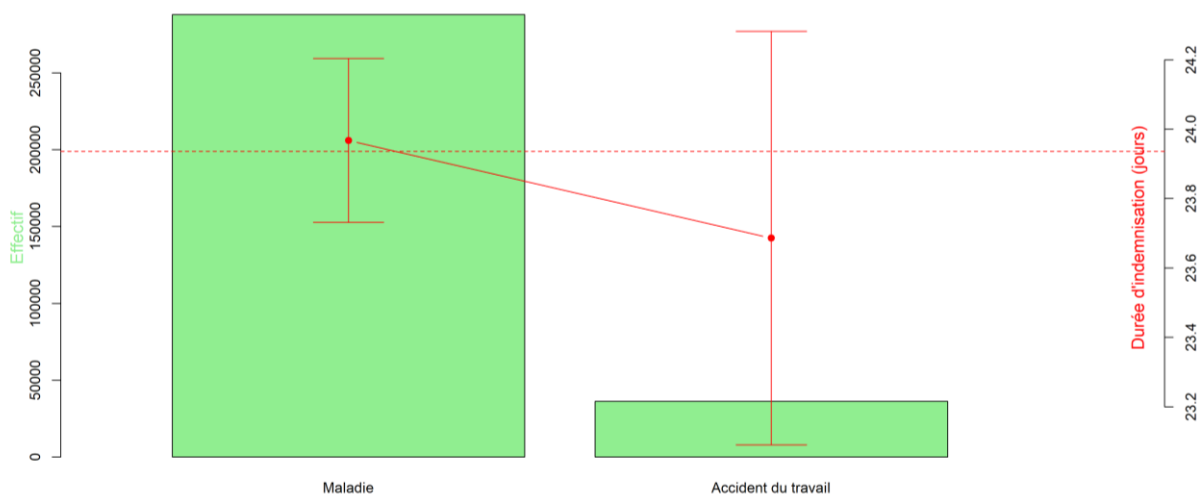
<sup>5</sup> ENSPER : Ensemble du Personnel

**Figure 36: Impact du Mode de Souscription sur la durée de maintien**



La majorité de l'effectif global est répartie sur les modes de souscription standard<sup>6</sup> et de convention collective avec une durée d'indemnisation moyenne de 25 jours. On note qu'il y a moins de 4 jours d'écart entre les durées moyennes de chacun de ces deux modes de souscription.

**Figure 37: Impact du type d'arrêt sur le maintien en incapacité**



Les arrêts pour maladie représentent la grande majorité des arrêts. En moyenne, la durée d'indemnisation est de 25 jours.

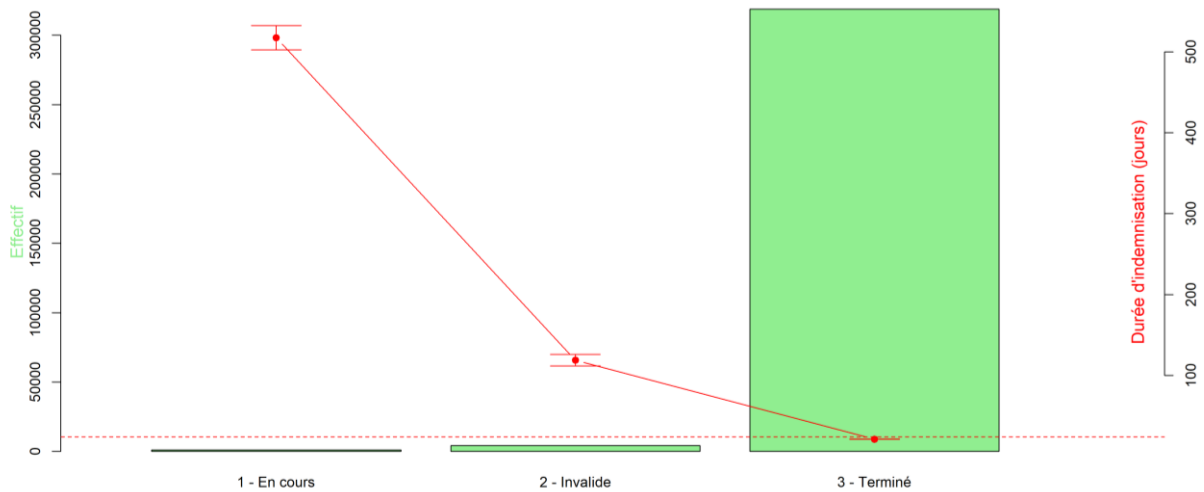
<sup>6</sup> STD : Standard (garanties prises au niveau de l'entreprise)

SPE : Spécifique (garanties prises au niveau de l'entreprise mais adaptées : pour grande entreprise)

CNV : Convention Collective (garanties prises au niveau de la branche de la profession)



**Figure 38: Impact du motif de reprise sur le maintien en incapacité**



La durée de maintien dépend du motif de reprise (statut, motif de sortie). La grande majorité des sinistres de la base sont clos avec une durée de maintien en incapacité d'environ 15-20 jours. Les classes « 1 - en cours » et « 2 - Invalide » sont peu volumineuses, la durée d'indemnisation est donc plus volatile (plus de 100 jours).

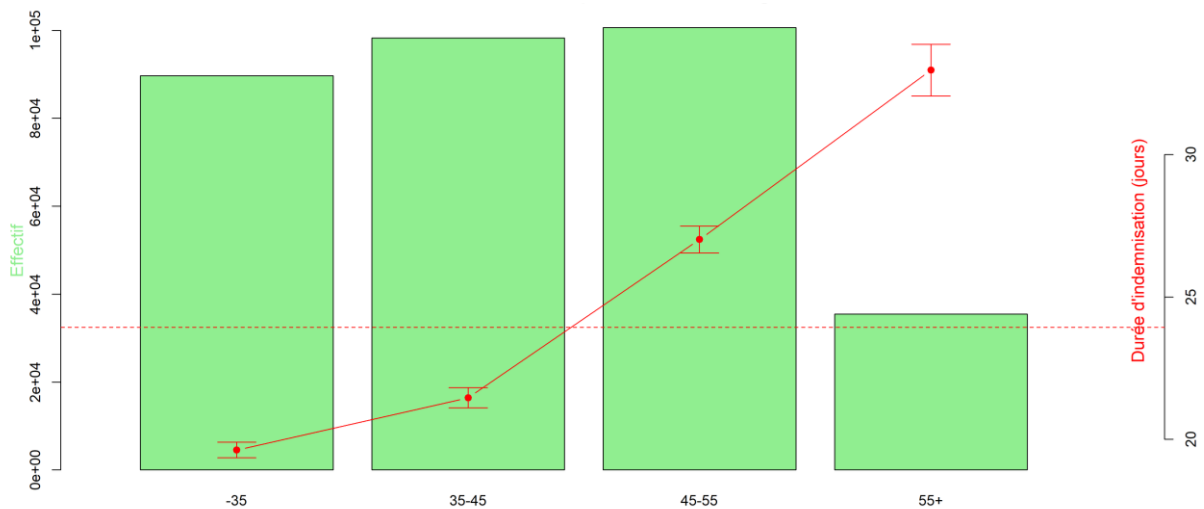
La base de données étudiée a la particularité d'avoir l'essentiel des sinistres clos. Cela est dû au fait que la population est vue avec suffisamment de recul. Ainsi, on obtient peu de données censurées.

Années/Durée d'indemnisation		1 - En cours	Clos	Total	% ouvert par année
Année de survenance	2005	2	108 032	<b>108 034</b>	0,00%
	2006	227	104 663	<b>104 890</b>	0,22%
	2007	744	110 382	<b>111 126</b>	0,67%

**Figure 39: Tableau des sinistres ouverts par année de survenance**

L'essentiel des sinistres « 1 – En cours » au 31/12/2007 sont survenus en 2007 (76%). Pour toutes les années de survenance, les sinistres « 1 – En cours » représentent moins de 1% des sinistres.

**Figure 40: Impact de l'âge à l'entrée sur la durée de maintien en incapacité**



Les durées de maintien sont logiquement croissantes avec l'âge d'entrée en incapacité.

D'après ces statistiques, on peut définir une population type, dont les caractéristiques sont les suivantes :

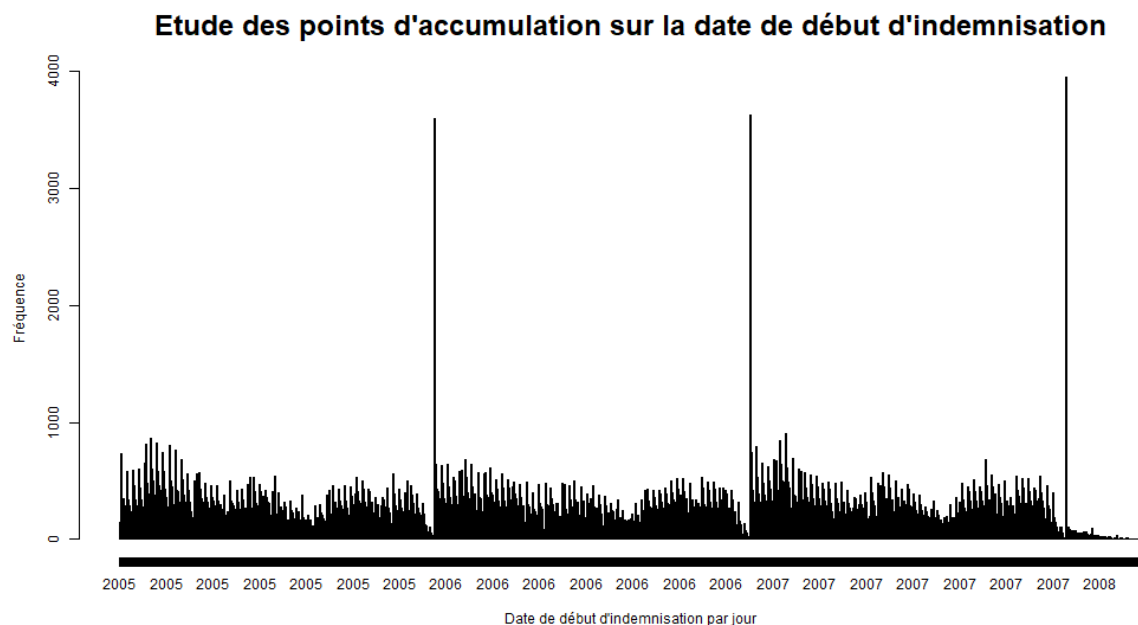
- Sexe : Femme ;
- CSP : Non cadre ;
- Type d'arrêt : Maladie ;
- Mode de souscription : Convention Collective.

Le motif de reprise d'un sinistre correspond à son motif de sortie, qui n'est donc pas connu lors de la provision d'un sinistre. Cette caractéristique du sinistre n'est donc pas utilisée comme variable de segmentation de la population.

### Points d'accumulation

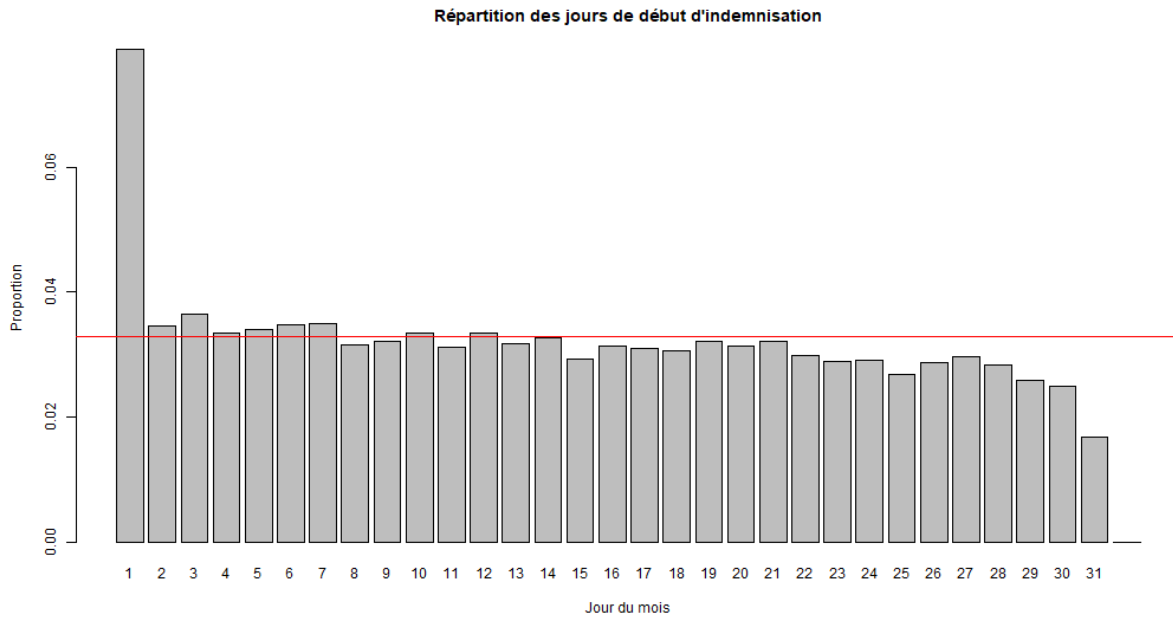
Cette partie a pour but de détecter des anomalies sur la qualité des données et notamment des pics d'accumulation.

- Dates de débuts d'indemnisation

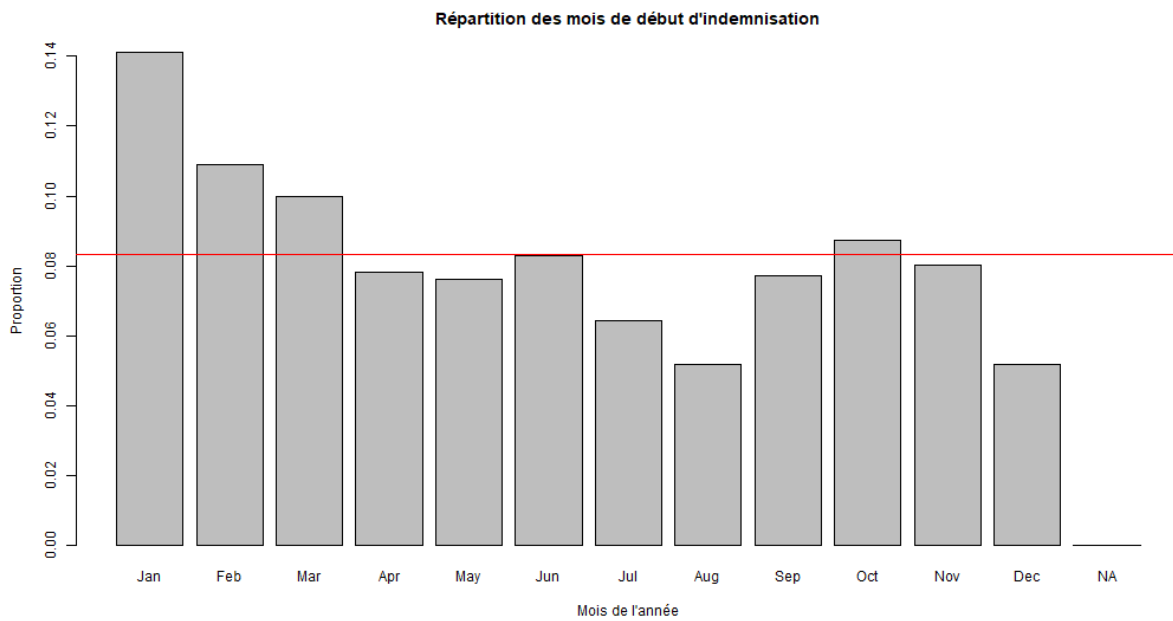


On peut noter des pics importants au 1<sup>er</sup> janvier de chaque année. Cette anomalie peut être expliquée par un problème de gestion : les dates d'entrée étant forcées au 01/01 quand la date n'est pas renseignée dans le système. Soit le nombre de jours indemnisés est bon et la date simplement décalée, soit les autres dates relatives au sinistres sont justes (date de survenance et date de sortie) et la date d'entrée est fautive ce qui voudrait dire que la durée d'indemnisation observée est fautive.

L'observation de ces anomalies donne lieu à un retraitement décrit par la suite.



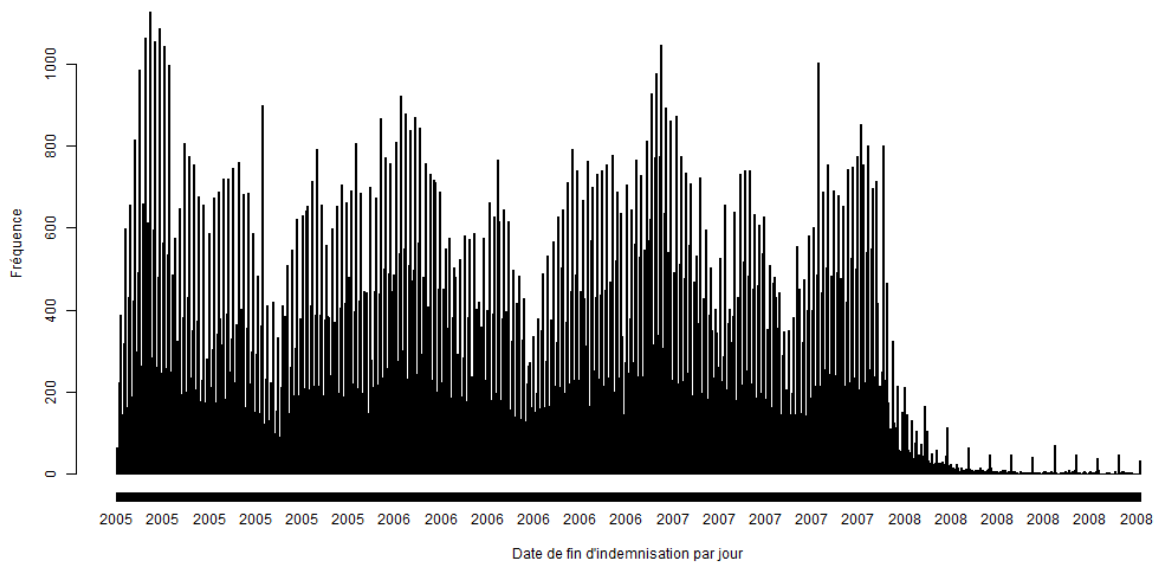
Le pic du premier jour du mois s'explique par les pics du 1<sup>er</sup> janvier de chaque année.



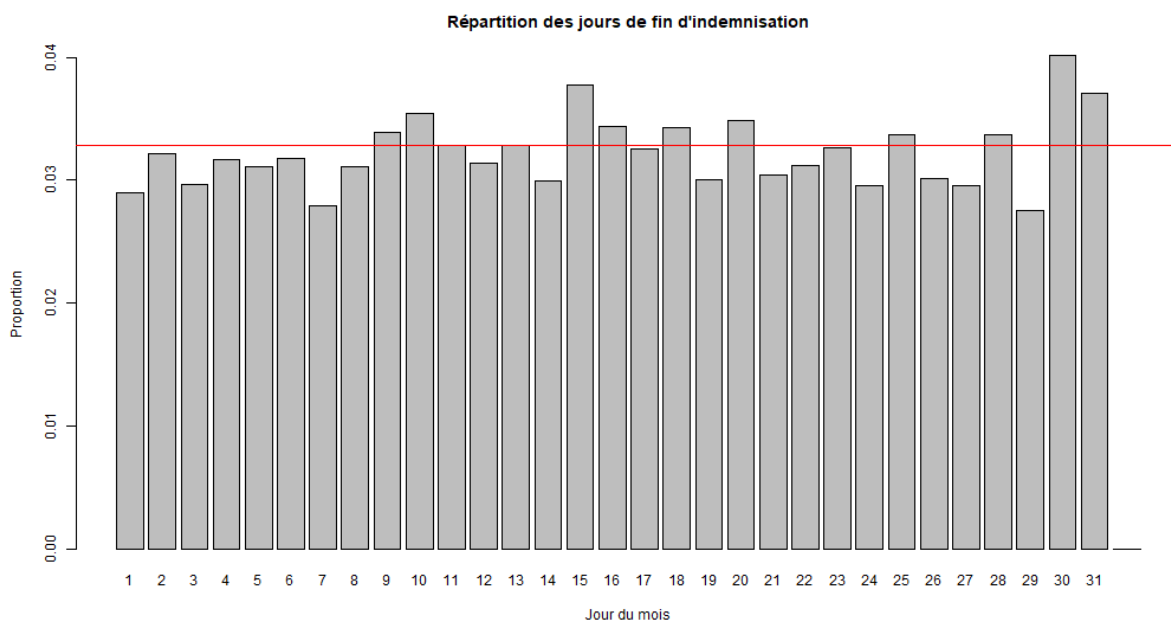
On peut noter que beaucoup de sinistres ont un début d'indemnisation dans les premiers mois de l'année. Cela reste cependant relativement bien reparti sur les différents mois de l'année.

- Dates de fin d'indemnisation

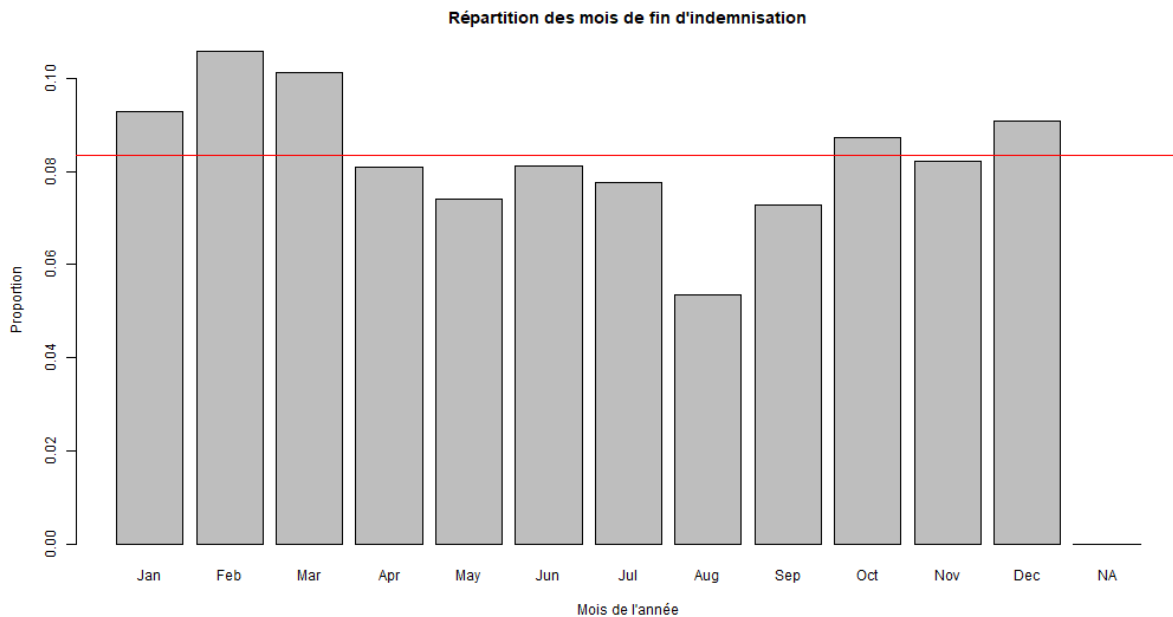
### Etude des points d'accumulation sur la date de fin d'indemnisation



L'étude des points d'accumulation des dates de fin d'indemnisation par jour permet d'observer une répartition homogène sur les différentes années. On peut observer un creux par année dans les environs du mois d'août. Cette observation est à vérifier avec les graphiques suivants.

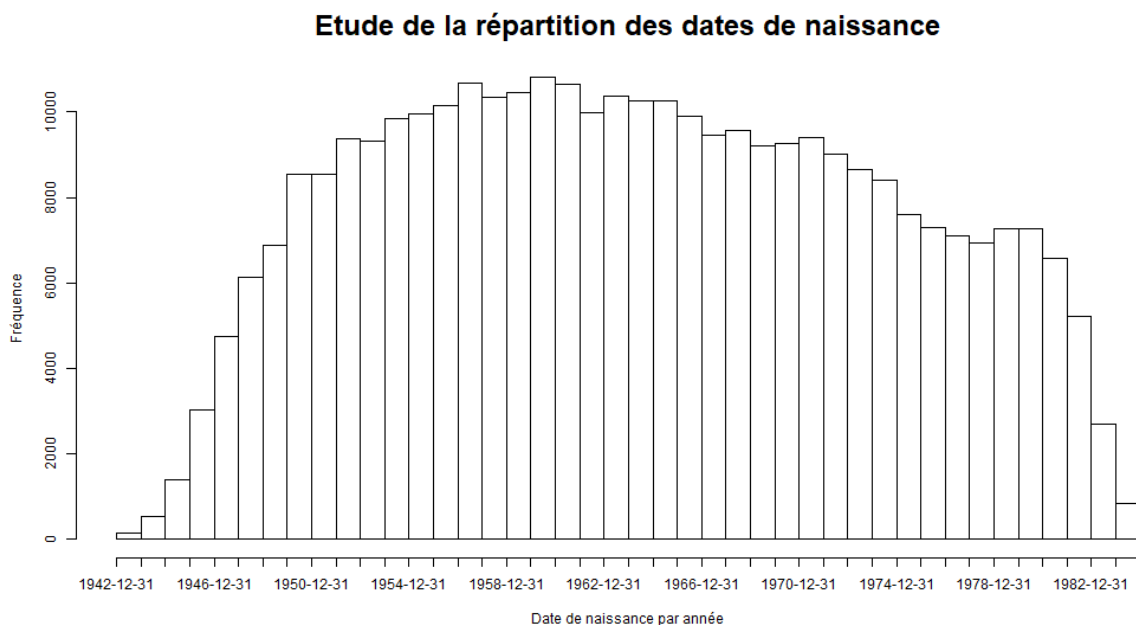


On note une répartition des jours de fin d'indemnisation sur les jours du mois homogène.

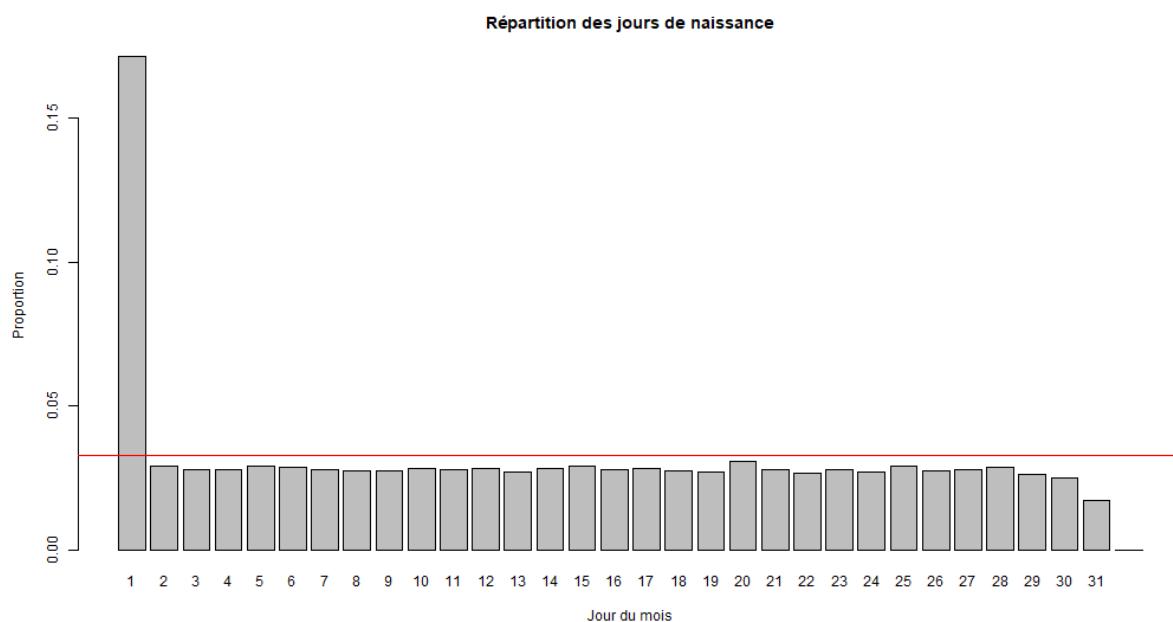


Comme souligné avec le graphique de l'accumulation des jours de fin d'indemnisation sur les différentes années observées. On remarque que le mois d'août est effectivement le mois avec la plus petite proportion de sinistres se terminant.

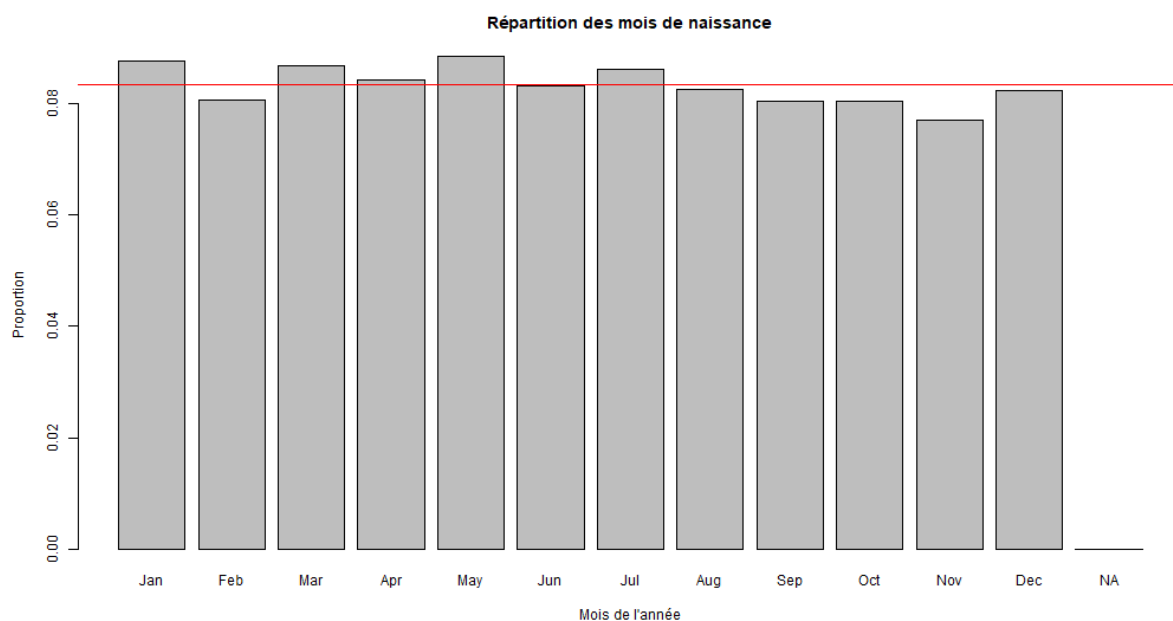
- Dates de naissance



L'ensemble des dates de naissance est bien reparti dans le temps, ce qui permet d'avoir une population avec tous les âges représentés.



On peut noter un pic important le premier jour du mois pour les dates de naissance. Quand seuls le mois et l'année de naissance sont connus par l'assureur, ce dernier met le 1<sup>er</sup> comme jour du mois.



Les mois de naissances des assurés de la base observées sont uniformément répartis.

### Points d'accumulation et retraitement

L'anomalie principale remarquée dans l'étude de ces précédents graphiques, est celle des pics des dates d'entrée au 1<sup>er</sup> Janvier des années 2006 à 2008.

Plusieurs retraitements de ces anomalies sont possibles :

- Un décalage aléatoire entre 0 et 6 mois avant ou après des sinistres. Les dates de survenance, d'entrée et de sortie étant maintenues afin de garder les durées d'indemnisation.

- Une suppression de ces sinistres.

L'élément important est de savoir si la date d'entrée est fautive mais les autres dates justes, ce qui impliquerait que les durées d'indemnisation soient fausses. Dans ce cas, la suppression semble plus juste.

Si les durées sont justes mais l'ensemble des dates des sinistres décalées, la répartition des sinistres de manière aléatoire sur les 6 mois précédents et suivants est plus adaptée.

Dans la suite de l'étude, l'ensemble des résultats présentés sont ceux de la première démarche. Les résultats de la seconde démarche seront confrontés à ceux de la première dans la dernière partie.

## Conclusion

L'étude statistique de la base de données a permis d'identifier de nombreuses anomalies. De nombreux retraitements de données ont été nécessaires afin d'obtenir un portefeuille homogène et pour pouvoir faire les calculs des parties suivantes. Il est indispensable d'avoir une base relativement propre pour ensuite appliquer les méthodes, car même si les méthodes sont justes, si les données ne le sont pas, les résultats ne seront pas justes.

Les modèles présentés dans cette partie sont construits à l'aide de la population retraitée avec une date de survenance antérieure au 31/12/2005 ; c'est ce qu'on appelle la population d'apprentissage. Les sinistres avec une date de survenance postérieure au 31/12/2005 serviront à faire le test rétroactif (*backtesting*) des méthodes. Ainsi, les sinistres survenus en 2006 permettent d'obtenir la liste des sinistres à provisionner au 31/12/2006, et l'année 2007 donne le recul nécessaire pour comparer le nombre de jours provisionnés et le nombre de jours réellement indemnisés.

## 2.3 Table d'expérience

Avant de présenter l'estimateur de Kaplan-Meier pour déterminer les taux de sortie bruts, nous allons rappeler quelques éléments mathématiques pour les modèles de survie.

### 2.3.1 Eléments de base des modèles de survie

Soit  $T_x$  une variable aléatoire dans  $R_+$ , représentant le temps écoulé dans l'état d'incapacité depuis son entrée pour un individu, d'âge  $x$ .

Soit  $f_x(t)$  sa fonction de densité et  $F_x(t)$  sa fonction de répartition.

$$f_x(t) = \frac{dF_x(t)}{dt},$$

$$F_x(t) = P(T_x \leq t).$$

**Définition 1 (Fonction de survie)** Soit  $S_x(t)$  la probabilité de survie, définie de la manière suivante :

$$S_x(t) = 1 - F_x(t) = P(T_x > t).$$

La fonction  $S$  est décroissante entre 1 et 0.

**Définition 2 (Espérance de survie)** L'espérance de survie ou durée moyenne de survie, notée  $E(T)$  est définie par :

$$E(T_x) = \int_0^{\infty} t dF_x(t) = - \int_0^{\infty} t dS_x(t) = \int_0^{\infty} S_x(t) dt.$$

**Définition 3 (Fonction de hasard)** La fonction de hasard  $h_x(t)$ , aussi nommée taux de décès ou risque instantané, représente la probabilité instantanée de quitter l'état d'incapacité.

$$h_x(t) = \frac{f_x(t)}{S_x(t)} = -\frac{d}{dt} \ln(S_x(t)).$$

La fonction de hasard détermine la loi de  $T$ , on peut l'écrire  $S_x(t) = \exp\left(-\int_0^t h_x(s) ds\right)$ .

**Définition 4 (Fonction de hasard cumulé)** La fonction de hasard cumulé est définie par :

$$H_x(t) = \int_0^t h_x(s) ds.$$

On a donc les relations suivantes :

$$S_x(t) = \exp(-H_x(t)),$$

$$H_x(t) = -\ln(S_x(t)).$$

On pose également  $L_{x,t}$  le nombre de personnes entrées dans l'état à l'âge  $x$  et encore présentes au bout de  $t$  mois.

### Notion de censure et de troncature

Les données étudiées sont censurées à droite. C'est-à-dire que certains individus sont toujours en incapacité au 31/12/2008. On ne connaît pas la date de sortie de ces individus.

Les données dites avec troncature correspondent à des individus qui sont déjà en incapacité au 01/01/2002. La date de début de l'incapacité est donc inconnue. Les données de l'étude ne présentent pas de troncature car toutes les dates de survenance sont postérieures à 2002.

Les notions de censure et de troncature sont prises en compte dans le calcul de l'estimateur de Kaplan-Meier.

### 2.3.2 Estimation de la loi de survie par l'estimateur de Kaplan-Meier

L'estimation de la loi de maintien en incapacité peut se faire à l'aide de différentes catégories d'estimateurs paramétriques ou non-paramétriques. Les estimateurs paramétriques sont fondés sur le choix *a priori* de la fonction de survie. Les estimateurs non-paramétriques permettent de calculer directement la fonction à l'aide des données.

Au vu du grand nombre de données à notre disposition, nous utilisons un estimateur non-paramétrique. L'estimateur de Kaplan-Meier en est un et permet de tenir compte des censures et troncatures.

L'estimateur de Kaplan-Meier s'inspire d'un constat simple : la probabilité d'être encore en incapacité à l'instant  $t_2$  nécessite d'avoir dépassé l'instant  $t_1$  (avec  $t_2 > t_1$ ). On peut donc écrire :

$$S_x(t_2) = P(T_x > t_2) = P(T_x > t_2 \cap T_x > t_1) = P(T_x > t_2 | T_x > t_1) P(T_x > t_1).$$

On peut renouveler le conditionnement sur les instants précédents.

Soient  $t_1 < \dots < t_i < \dots < t_n$  une suite d'instant, on pose  $p_i = P(T_x > t_i | T_x > t_{i-1})$  la probabilité d'être encore en incapacité entre  $t_{i-1}$  et  $t_i$ .

On obtient :



$$S_x(t) = \prod_{i, t_i < t} p_i.$$

Afin d'obtenir une estimation de la fonction de survie, il faut estimer  $p_i$ . On peut utiliser pour cela la méthode du maximum de vraisemblance (démonstration cf. Annexe) et l'on obtient :

$$\hat{p}_i = 1 - \frac{d_i}{r_i},$$

avec,

- $d_i$  le nombre de sortie de l'état à la date  $t_i$  ;
- $r_i$  la population à risque juste avant l'instant  $t_i$ .

La fonction de survie est donc estimée par :

$$\hat{S}_x(t) = \prod_{i, t_i \leq t} \left(1 - \frac{d_i}{r_i}\right).$$

## Résultats

Le graphique ci-après représente les taux bruts de sortie estimés avec la méthode de Kaplan-Meier au pas mensuel à partir de nos données.

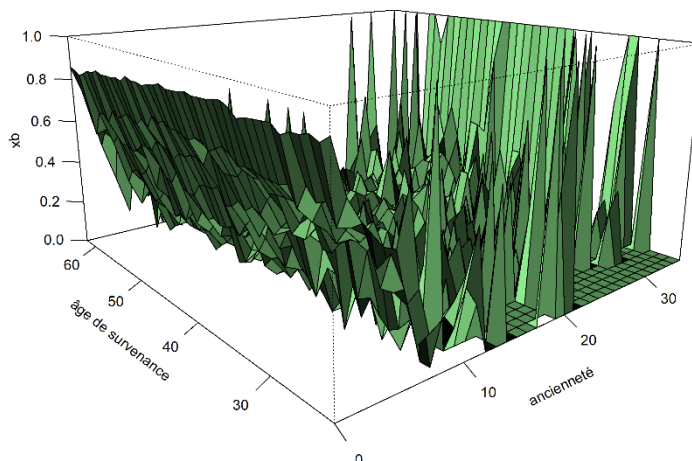


Figure 41: Taux de sortie bruts Kaplan-Meier (pas de temps mensuel)

On peut noter que les taux sont irréguliers pour l'ensemble des âges. Cette loi doit être lissée pour obtenir des résultats qui reflètent le phénomène observé.

### 2.3.3 Lissage des taux

La méthode de Kaplan-Meier permet d'obtenir une estimation  $\hat{q}_x$  du taux de sortie réel  $q_x$ . Une erreur d'estimation  $e_x$  est définie par :

$$e_x = \hat{q}_x - q_x.$$

L'ajustement permet de diminuer ce niveau d'erreur et d'obtenir une courbe plus régulière. Sachant que les taux sont construits par âge et ancienneté, il faut lisser les taux dans deux dimensions.

Différentes méthodes de lissage existent : paramétrique ou non-paramétrique. La quantité importante de données à disposition permet d'utiliser la méthode non-paramétrique de Whittaker-Henderson.

## Méthode de Whittaker-Henderson

Les éléments d'explication de la méthode sont repris de PLANCHET et al. [2006].

Cette méthode conjugue un critère de fidélité et un critère de régularité. Pour obtenir le lissage des données, la somme de ces critères est minimisée.

Soit  $\hat{q} = (\hat{q}_{i,j})_{1 \leq i \leq p, 1 \leq j \leq q}$  l'estimateur des taux ( $\hat{q}_{i,j}$  estimateur à l'âge  $i$  et ancienneté  $j$ ).

- Le critère de fidélité

Le critère de fidélité est un critère des moindres carrés ordinaires pondérés. La qualité de l'ajustement est quantifiée à l'aide de ce critère. On le note

$$F = \sum_{i=1}^p \sum_{j=1}^q w_{ij} (q_{ij} - \hat{q}_{ij})^2,$$

Avec :

- $q_{ij}$  : les taux ajustés avec Whittaker-Henderson à chercher,
- $\hat{q}_{ij}$  : les taux estimés avec Kaplan-Meier,
- $w_{ij}$  : les poids.

Dans notre étude,  $w$  correspond au taux d'exposition au risque.

- Le critère de régularité

Le critère de régularité est décomposé en un critère horizontal et un critère vertical.

L'opérateur « différence avant » composé  $n$  fois est défini par :

$$\begin{cases} \Delta u(x) = u(x+1) - u(x), \\ \Delta^n u(x) = \underbrace{\Delta \dots \Delta}_{n \text{ fois}} u(x), \end{cases}$$

On démontre par récurrence (cf. Annexe) que  $\Delta^n u(x)$  s'écrit avec des coefficients binomiaux.

$$\Delta^n u(x) = \sum_{j=0}^n C_n^j (-1)^{n-j} u(x+j).$$

La régularité se mesure avec les critères verticaux et horizontaux.

- Le critère vertical est la somme des opérateurs différences verticaux :

$$S_v = \sum_{j=1}^q \sum_{i=1}^{p-z} (\Delta_v^z q_{ij})^2,$$

Avec  $\Delta_v^z$  l'opérateur pour la régularité verticale qui prend  $q_{ij}$  comme une série indicée par  $i$  avec  $j$  fixé.

- Le critère horizontal est la somme des opérateurs différences horizontaux :

$$S_h = \sum_{j=1}^q \sum_{i=1}^{p-z} (\Delta_h^z q_{ij})^2,$$

Avec  $\Delta^y_h$  l'opérateur pour la régularité horizontale qui prend  $q_{ij}$  comme une série indicée par  $j$  avec  $i$  fixé.

La fonction à minimiser est donc :

$$M(q) = F(q) + \alpha S_v(q) + \beta S_h(q),$$

ce qui revient à résoudre :

$$\frac{\partial M}{\partial q} = 0.$$

Une résolution matricielle est utile pour résoudre ce problème d'optimisation. Certains éléments sont à changer ou à construire pour cela.

- Réorganisation des  $\hat{q}_{ij}$  en un vecteur colonne  $u$  de taille  $p \times q$  tel que :

$$u_{q(i-1)+j} = \hat{q}_{i,j}.$$

- La matrice des poids est construite en mettant sur les diagonales les lignes de  $(w_{ij})$  :

$$w_{q(i-1)+j, q(i-1)+j}^* = w_{ij}.$$

- Construction de la matrice  $K_v^z$  de dimension  $(q(p-z), m)$  tel que :

$$S_v = {}^t(qK_v^z) (K_v^z q).$$

- Construction de la matrice  $K_h^y$  de dimension  $(p(q-y), m)$

$$S_h = {}^t(qK_h^y) (K_h^y q).$$

Les taux lissés sont calculés de la manière suivante :

$$q^* = (w^* + \alpha {}^t K_v^z K_v^z + \beta {}^t K_h^y K_h^y)^{-1} w^* u.$$

(Démonstration cf Annexe)

### Application :

Afin de lisser la table obtenue par la méthode Kaplan-Meier, il faut déterminer les paramètres :

- les poids : dans notre étude, les poids sont les taux d'exposition au risque.
- les ordres verticaux et horizontaux : ces paramètres varient entre 1 et 3 et correspondent au degré de régularité : l'ordre est défini à 2 dans notre étude.
- les critères de fidélité  $\alpha$  et  $\beta$  : ces paramètres sont des poids de régularité verticale et horizontale variant entre 1 et 100 (dans notre étude, différentes valeurs ont été testées et les valeurs retenues sont 100 pour les deux critères).

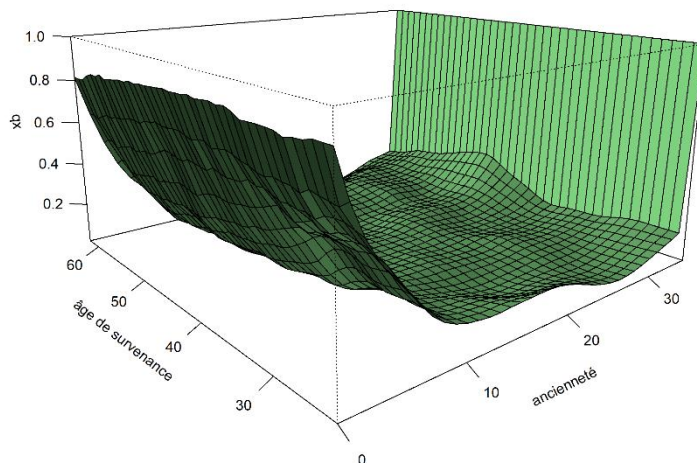


Figure 42: Taux de sortie lissé (WH)

### Validation

La validation d’une table d’expérience lissée peut s’effectuer en confrontant les taux lissés et les taux bruts par âge et par ancienneté. Un intervalle de confiance à 95% est calculé de la manière suivante :

$$IC = \left[ esp - 1.96 * \sqrt{qAjust * (1 - qAjust) * expo}; esp + 1.96 * \sqrt{qAjust * (1 - qAjust) * expo} \right]$$

avec,

*esp* : Somme selon l’ancienneté ou l’âge des sorties modèles,

*expo* : Exposition de la population observée,

*qAjust* : Somme selon l’ancienneté ou l’âge des sorties modèles divisés par *expo*.

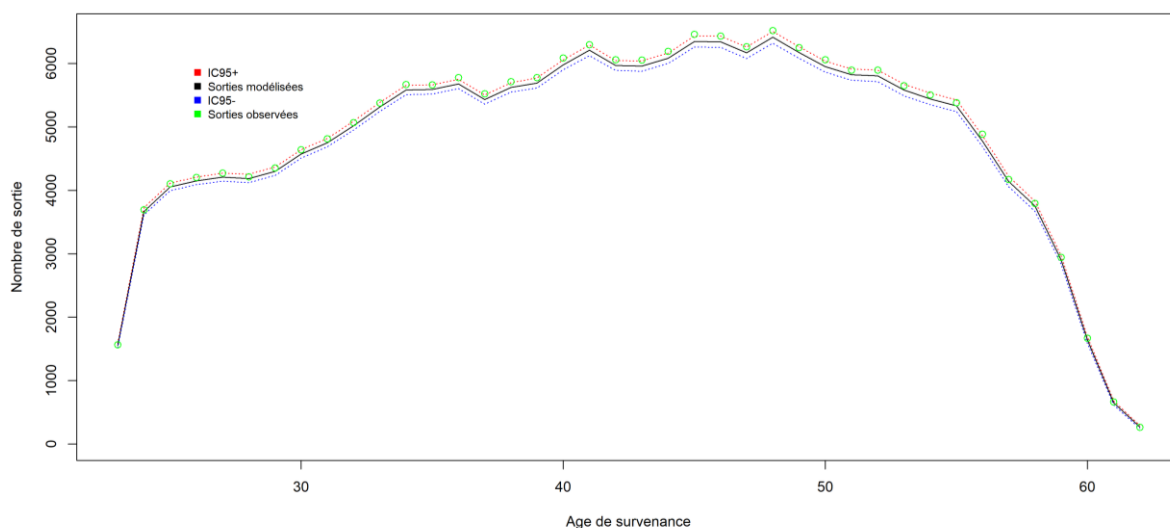
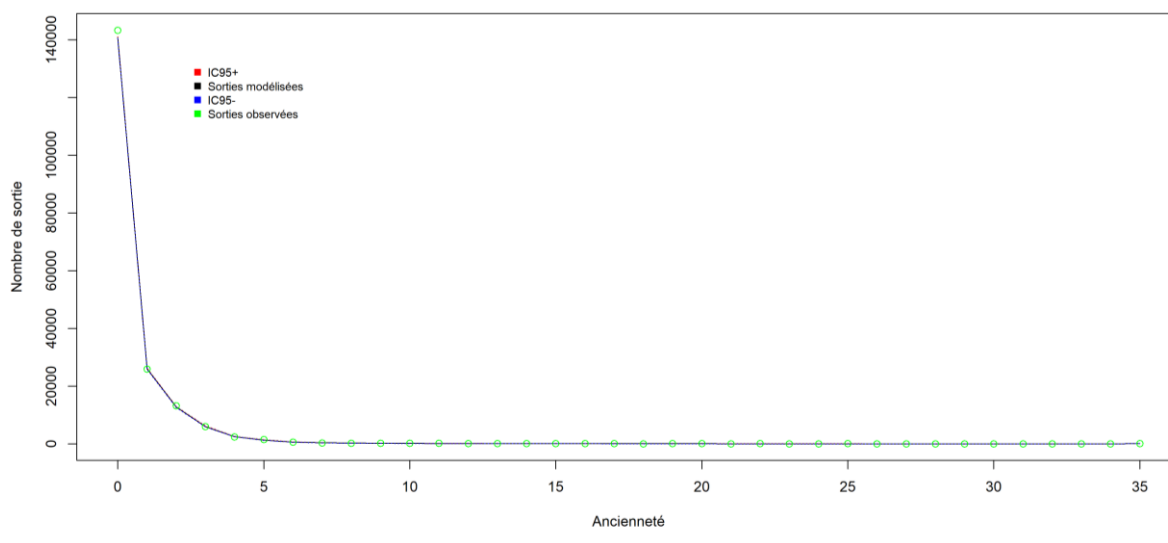


Figure 43: Comparaison modèle/ réalisation par âge de survénance.

On note que les sorties observées sont globalement positionnées sur la borne supérieure de l'intervalle de confiance ponctuel à 95%, signe que le modèle sous-estime légèrement les taux de sortie. Il a donc par conséquent tendance à légèrement sur-estimer les durées de maintien.



**Figure 44: Comparaison modèle/réalisation par ancienneté.**

On observe que les sorties observées par mois d'ancienneté se situent sur la courbe des sorties modélisées, donc le modèle représente bien les données selon l'ancienneté.

### 2.3.4 Segmentation

Afin de représenter au mieux la réalité et le risque représenté, une segmentation de la population et l'estimation d'une table d'expérience par segment pourraient être réalisées. Cependant, on se retrouve souvent confronté au problème du manque de données pour construire les tables. Les outils des modèles de durée avec des variables explicatives permettent de contourner ce problème. La population est observée dans sa globalité et l'effet de chaque variable explicative sur la fonction de hasard est calculé. Le modèle le plus utilisé est le modèle de Cox.

#### Modèle à hasard proportionnel

On considère une fonction de survie de base  $S_0(t)$  et on suppose que la fonction de survie de nos données observées est de la forme  $S(t) = S_0(t)^\theta$ , avec  $\theta > 0$  un paramètre inconnu.

On obtient :  $f(t) = \theta S(t)^{\theta-1} f_0(t)$ ,

La fonction de hasard associée à  $S(t)$  est donc :  $h(t) = \frac{f(t)}{S(t)} = \theta \frac{f_0(t)}{S_0(t)} = \theta h_0(t)$ .

Le nom de « modèle à hasard proportionnel » vient donc de la proportionnalité entre la fonction de hasard et celle de base.

#### Modèle de Cox

Le modèle de Cox est un modèle à hasard proportionnel qui inclut des variables explicatives pour déterminer le paramètre  $\theta > 0$ . Ce dernier s'écrit donc :

$$\theta = e^{z'\beta}.$$

On obtient,

$$h(t) = h_0(t)e^{z'\beta}.$$

Grâce à la relation entre la fonction de survie et de hasard, on peut réécrire le modèle ainsi :

$$S(t) = S_0(t)^{\exp(z'\beta)},$$

avec

- $z = (z_1, \dots, z_p)'$  le vecteur de variables explicatives,
- $\beta = (\beta_1, \dots, \beta_p)'$  le vecteur de paramètres. L'effet des covariables sur le risque instantané est représenté par ce vecteur.

On définit la fonction de hasard conditionnellement aux variables explicatives par  $h(t|Z)$ .

Le modèle peut être réécrit comme un modèle de régression linéaire :

$$\ln(h(t|Z = z)) = \ln(h_0(t)) + \sum_{i=1}^p z_i \beta_i.$$

Les coefficients sont estimés par la méthode du maximum de vraisemblance partielle.

### Vraisemblance partielle de Cox

La vraisemblance du modèle s'écrit :

$$L(t|\beta) = \prod_{i=1}^n h_0(t_i) \exp\left(\sum_{k=1}^p z_{ik} \beta_k\right) S_0(t_i)^{d_i \exp(z_{ik} \beta_k)}.$$

Cette formule dépend de la fonction de hasard non connue. COX [1972] propose une vraisemblance partielle qui ne dépend pas de la fonction de hasard. Pour cela, on s'appuie sur les données réelles non censurées.

$$P(\text{indiv. } i \text{ sort à } T_j | \text{une sortie en } T_k) = \frac{P(\text{indiv. } i \text{ sort à } T_j)}{P(\text{une sortie en } T_k)}.$$

La participation d'un individu  $i$  à la fonction de vraisemblance est donc

$$\frac{h(t|z_i)}{\sum_{k \in \text{expo}(t)} h(t, z_k)} = \frac{\exp(\sum_{j=1}^p z_{ij} \beta_j)}{\sum_{k \in \text{expo}(t)} \exp(\beta' z_k)},$$

Avec  $\text{expo}(t)$  les individus ayant un risque de sortie au temps  $t$ .

La fonction de vraisemblance partielle s'écrit donc :

$$L_{\text{Cox}}(t|\beta) = \prod_{i=1}^n \frac{\exp(\sum_{j=1}^p z_{ij} \beta_j)}{\sum_{k \in \text{expo}(t)} \exp(\beta' z_k)}.$$

### Détermination des coefficients

En maximisant cette fonction ou la log-vraisemblance, on détermine les coefficients  $\beta$ .

La log-vraisemblance s'écrit :

$$\log(L_{\text{Cox}}(t|\beta)) = \sum_{i=1}^n \left\{ \sum_{j=1}^p z_{ij} \beta_j - \log \left( \sum_{k \in \text{expo}(t)} \exp(z_{kj} \beta_j) \right) \right\}.$$

En dérivant par rapport à  $\beta_j$  avec  $j = 1, \dots, p$

$$\frac{\partial(\log(L_{\text{Cox}}(t|\beta)))}{\partial \beta_j} = \sum_{i=1}^n z_{ij} - \sum_{i=1}^n \frac{\sum_{k \in \text{expo}(t)} z_{jk} \exp(\beta' z_k)}{\sum_{k \in \text{expo}(t)} \exp(\beta' z_k)}.$$

On obtient la valeur des  $\beta_j$  en résolvant le système d'équation égal à 0. Pour cela, on utilise l'algorithme de Newton-Raphson.

### Estimation des probabilités de sortie

Une fois la courbe de la population principale connue ainsi que les coefficients  $\beta_j$ , on peut déterminer les probabilités de sortie par âge :

$$S(t) = S_0(t) \exp(z' \beta).$$

On rappelle que la fonction de survie peut s'écrire :

$$S(t) = \exp\left(-\int_0^t h(s, x) ds\right).$$

Soit  ${}_a q_t$  la probabilité pour un individu entré en incapacité à l'âge  $x$  de quitter l'incapacité entre  $t$  et  $t+a$  et  ${}_a p_t$  la probabilité pour un individu entré en incapacité à l'âge  $x$  de rester en incapacité entre  $t$  et  $t+a$  :

$$\begin{aligned} {}_a q_t &= 1 - {}_a p_t = 1 - \frac{S(t+a)}{S(t)}, \\ &= 1 - \frac{\exp\left(-\int_0^{t+a} h(s) ds\right)}{\exp\left(-\int_0^t h(s) ds\right)}, \\ &= 1 - \exp\left(-\int_t^{t+a} h(s) ds\right). \end{aligned} \quad (1)$$

Or, on sait que :

$$h(t) = h_0(t) e^{z' \beta}.$$

On remplace dans (1),

$$\begin{aligned} {}_a q_t &= 1 - \exp\left(-\int_t^{t+a} h_0(s) \exp(\beta' z) ds\right), \\ {}_a q_t &= 1 - \exp\left(-\int_t^{t+a} h_0(s) ds\right)^{\exp(\beta' z)}. \end{aligned} \quad (2)$$

Or, on peut appliquer (1) à la fonction de hasard de base.

On note  ${}_a q_{0,t}$  la probabilité  ${}_a q_t$  avec comme fonction de survie  $S_0$ .

$${}_a q_{0,t} = 1 - \exp\left(-\int_t^{t+a} h_0(s) ds\right).$$

On remplace dans (2) et on obtient,

$${}_a q_t = 1 - (1 - {}_a q_{0,t})^{\exp(\beta'z)}.$$

On peut redéfinir  ${}_a q_t$  et  ${}_a q_{0,t}$  par  $q(x|z; \beta)$  et  $q_0(x)$  respectivement. Les autres courbes sont donc obtenues par la formule suivante :

$$q(x|z; \beta) = 1 - (1 - q_0(x))^{\exp(-z'\beta)}.$$

Pour appliquer le modèle de Cox les hypothèses suivantes doivent être vérifiées :

**(H1) L'hypothèse de risques proportionnels** : le rapport des fonctions de hasard de deux modalités d'une variable explicative est indépendant du temps.

En effet dans le modèle de Cox, si on suppose que le modèle ne possède qu'une variable explicative, on obtient :

$$h(t) = h_0(t) \exp(Z_1 \beta_1).$$

Soit une population à  $n$  individus,  $Z_1 = (z_{11}, \dots, z_{1k}, \dots, z_{1n})'$  avec  $z_{1k}$  la modalité de la variable  $Z_1$  du  $k^{\text{ième}}$  individu.

Pour deux individus distincts  $i$  et  $k$ , on obtient :

$$\frac{h(t, z_{1i})}{h(t, z_{1k})} = \exp(\beta_1(z_{1i} - z_{1k})) \Rightarrow \ln\left(\frac{h(t, z_{1i})}{h(t, z_{1k})}\right) = \beta_1(z_{1i} - z_{1k}).$$

On obtient bien un rapport indépendant du temps.

### Validation de l'hypothèse de risques proportionnels

Certains tests permettent de valider l'utilisation du modèle :

- Test graphique : tracer la fonction  $\log(-\log(S))$  par classe. (Explication par ELIAS [2013])
- Test du résidu de Schonefeld. (Explication en détail par THERNEAU et al. [2000])

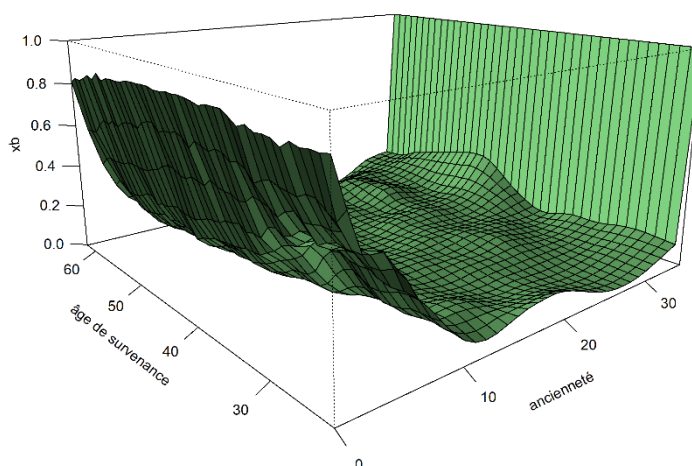
**(H2) L'hypothèse de log-linéarité** : le logarithme de la fonction de hasard est linéaire par rapport aux variables explicatives.

### Application et résultats

Les statistiques de la partie précédente permettent de déterminer les caractéristiques de la population de base : soit une femme, non cadre avec un mode de souscription de convention collective, et avec un arrêt dû à la maladie.

Pour cette population, nous calculons une table de maintien à l'aide de l'estimateur de Kaplan-Meier. Cette table est ensuite lissée avec la méthode de Whittaker-Henderson.





**Figure 45: Taux de sortie lissés de la population de référence**

La fonction de hasard du modèle dépend du sexe, de la CSP, du mode de souscription, du type d'arrêt.

Variable	Catégorie	Coef	Exp(coef)	Exp(-coef)
CSP	Ensemble du Personnel	- 0,042	0,959	1,043
	Cadres	- 0,246	0,782	1,279
Mode de souscription	Standard	- 0,007	0,993	1,007
	Spécial	- 0,113	0,893	1,119
Type d'arrêt	Accident du Travail	0,015	1,015	0,985
Sexe	Homme	0,154	1,167	0,857

**Figure 46: Tableau des coefficients de Cox**

On observe que les probabilités conditionnelles de sortie sont logiquement fortement impactées par la nature de l'arrêt et les caractéristiques des assurés.

Les durées de maintien en incapacité sont en particulier plus longues pour les cadres et l'ensemble du personnel, par rapport aux non-cadres.

Les taux de sortie des modes de souscription Standard ou Spécifique sont plus faibles qu'en Convention Collective.

Les probabilités de sortie des hommes sont plus élevées que celles des femmes.

## 2.4 Méthodes utilisant des triangles

Les méthodes vues précédemment sont des méthodes dites ligne à ligne. Dans cette partie nous allons présenter des méthodes agrégées basées sur des triangles de développement.

### 2.4.1 Construction du triangle de développement

Les triangles de développement sont une manière d'agréger les sinistres en fonction de la date de survenance et un pas d'écoulement du sinistre, afin d'obtenir une vision de la dynamique des sinistres. Les triangles peuvent comporter différentes informations : règlements, nombre de sinistres, charges, ... Les montants de prestations n'étant pas disponibles dans la base de données, nous considérerons que 1€ de rente est versé par jour indemnisé. Cela revient à faire un triangle sur les durées d'indemnisation. Généralement, les triangles sont construits avec un pas annuel. Dans le cadre de cette étude, les triangles sont réalisés au pas mensuel, pour plus de précision.

Les notations utilisées dans cette partie sont les suivantes :

- $i$  désigne le mois de survenance du sinistre,  $i = 0, \dots, n$  ;
- $j$  désigne le mois de développement,  $j = 0, \dots, n$  ;
- $X_{ij}$  le montant de prestations des sinistres survenus en mois  $i$  et payé après  $j$  mois de développement.
- $C_{ij}$  le montant cumulé des prestations de sinistres survenus le mois  $i$ , au jème mois de développement, on a donc :  $C_{i,j} = \sum_{k=1}^j X_{ik}$

	1	2	...	$j$	...	$n-1$	$n$
1	$C_{1,1}$	$C_{1,2}$	...	...	...	$C_{1,n-1}$	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$	...	...	...	$C_{2,n-1}$	
...	...	...	...	...	...		
$i$	...	...	...	$C_{i,j}$	...		
...	...	...	...	...	...		
$n-1$	$C_{n-1,1}$	$C_{n-1,2}$					
$n$	$C_{n,1}$						

Figure 47: Triangle de règlements

Le triangle peut se lire de différentes façons :

- En colonne : pour un mois de développement donné, on observe le montant payé indépendamment de l'année de survenance ;
- En ligne : pour un mois de survenance donné, on observe la cadence de règlement ;
- En diagonale : celle-ci correspond à un mois comptable et permet d'observer l'inflation d'un mois sur l'autre.

L'information connue correspond au triangle supérieur, soit :

$$I = \{(C_{i,j}) | 0 \leq i + j \leq n\}.$$

L'objectif est d'estimer la partie inférieure du triangle par les différentes méthodes présentées dans les parties suivantes,  $\{(\hat{C}_{i,j}) | i + j > n, i \leq n, j \leq n\}$ .

	1	2	...	$j$	...	$n-1$	$n$
1	$C_{1,1}$	$C_{1,2}$	...	...	...	$C_{1,n-1}$	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$	...	...	...	$C_{2,n-1}$	$\hat{C}_{2,n}$
...	...	...	...	...	...	...	...
$i$	...	...	...	$C_{i,j}$	...	...	...
...	...	...	...	...	...	...	...
$n-1$	$C_{n-1,1}$	$C_{n-1,2}$	...	...	...	$\hat{C}_{n-1,n-1}$	$\hat{C}_{n-1,n}$
$n$	$C_{n,1}$	$\hat{C}_{n,2}$	...	...	...	$\hat{C}_{n,n-1}$	$\hat{C}_{n,n}$

Figure 48: Triangle de règlements complété

## 2.4.2 Méthode de Chain Ladder

La méthode de Chain Ladder est basée sur l'hypothèse de stabilité des données et du délai écoulé entre la date de survenance du sinistre et de la cadence de développement. Le comportement du futur est similaire à celui du passé.

Les hypothèses du modèle sont les suivantes :

**(H1)** les années de survénance sont indépendantes entre elles,

**(H2)** les années de développement sont les variables explicatives du comportement des sinistres futurs.

A l'aide des données historiques, on détermine des coefficients de passage qui permettent de passer d'une année de développement à la suivante :  $C_{i,j+1} = \lambda_j C_{i,j}$ .

Les coefficients de passage ou facteurs de développement  $\lambda_j$  sont estimés de la manière suivante :

$$\widehat{\lambda}_j = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}} \text{ pour } j = 1, \dots, n-1,$$

On obtient donc, avec  $i + j \geq n + 1$

$$\widehat{C}_{i,j} = [\widehat{\lambda}_{n+1-i} \cdots \widehat{\lambda}_{j-1}] C_{i,n+1-i}.$$

### 2.4.3 Méthode de Mack

Le modèle de Mack (1993) se base sur la méthode de Chain Ladder. L'objectif est d'estimer les erreurs commises lors de la détermination des provisions.

Les hypothèses du modèle de Mack sont les suivantes :

**(H1)**  $(C_{i,j})_{j=1,\dots,n}$  et  $(C_{i',j})_{j=1,\dots,n}$  sont indépendants pour tout  $i, i' = 1, \dots, n$ .

**(H2)** Il existe  $\lambda_j$  tel que  $E(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = \lambda_j C_{i,j}$  pour  $0 \leq i \leq n, 0 \leq j \leq n$

**(H3)** Il existe  $\sigma_j^2$  tel que  $Var(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = \sigma_j^2 C_{i,j}$  pour  $0 \leq i \leq n, 0 \leq j \leq n$

Notons, on a :

$$\bullet \quad \widehat{\lambda}_j = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}} \text{ pour } j = 1, \dots, n-1,$$

$$\bullet \quad \begin{cases} \widehat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=1}^{n-j-1} C_{i,j} \left( \frac{C_{i,j+1}}{C_{i,j}} - \widehat{\lambda}_j \right)^2, & 0 \leq j \leq n-2, \\ \widehat{\sigma}_{n-1}^2 = \min \left( \frac{\widehat{\sigma}_{n-2}^4}{\widehat{\sigma}_{n-3}^2}, \min(\widehat{\sigma}_{n-3}^2, \widehat{\sigma}_{n-2}^2) \right). \end{cases}$$

Sous les hypothèses (H1), (H2) et (H3), ces estimateurs sont des estimateurs sans biais de  $\lambda_j$  et  $\sigma_j^2$

Les coefficients de passage ( $\lambda_j$ ) sont estimés par les mêmes facteurs de Chain-Ladder ( $\widehat{\lambda}_j$ ).

Soit  $\widehat{R}_i$  l'estimateur du montant de provision par année de survénance  $R_i$  et  $\widehat{R}$  l'estimateur de la provision globale  $R$

L'écart quadratique moyen (MSEP : Mean Square Error) peut ainsi être calculé par années d'exercice de la manière suivante :

$$MSEP(\widehat{R}_i) = E \left( (\widehat{R}_i - R_i)^2 | C_{i,j}, i + j \leq n \right).$$

On peut l'estimer sous les hypothèses précédentes par :

$$\widehat{MSEP}(\widehat{R}_i) = \widehat{C}_{i,n}^2 \sum_{j=n-i}^{n-1} \frac{\widehat{\sigma}_j^2}{\widehat{\lambda}_j^2} \left( \frac{1}{\widehat{C}_{i,j}} + \frac{1}{\sum_{k=1}^{n-j} C_{i,k}} \right) \text{ avec } i=1, \dots, n,$$

L'estimateur de la provision totale est donné par la somme des erreurs quadratiques par année d'exercice :

$$\widehat{MSEP}(\widehat{R}) = \sum_{i=1}^n \left\{ \widehat{MSEP}(\widehat{R}_i) + \widehat{C}_{i,n} \left( \sum_{k=i+1}^n \widehat{C}_{k,n} \right) \sum_{j=n-i}^{n-1} \frac{2\widehat{\sigma}_j^2 / \widehat{\lambda}_j^2}{\sum_{h=1}^{n-j} \widehat{C}_{h,j}} \right\}.$$

Pour la démonstration des résultats  $\widehat{\sigma}_j^2$ ,  $\widehat{MSEP}(\widehat{R}_i)$  et  $\widehat{MSEP}(\widehat{R})$ , nous renvoyons à MACK [1993]. On démontre que l'estimateur de  $\lambda_j$  est sans biais en Annexe.

## Chapitre 3 :

### Analyses et Comparaisons

#### 3.1 Résultats avec la méthode Chain Ladder

##### 3.1.1 Principe

L'objectif est de comparer le nombre de jours effectivement indemnisés avec l'estimation du nombre de jours qui auraient été payés à partir des triangles de Chain Ladder. Les estimations de la durée indemnisée sont réalisées à l'aide de la cadence de règlements.

Nous disposons des arrêts dont les dates de survenance sont comprises entre le 01/01/2005 et le 31/12/2007.

Les projections des nombres de jours indemnisés sont effectuées à pas mensuel par année comptable. Pour les arrêts en cours au 31/12/N-1, les durées d'arrêt résiduelles pour l'année N sont estimées et sont comparées avec la durée effective des prestations sur cette même année.

Avec la méthode des triangles, on obtient sur les sinistres de 2005 à 2007 des triangles complétés comme ci-dessous :

	1	2	...	j	...	n-1	n
1	$C_{1,1}$	$C_{1,2}$	...	...	...	$C_{1,n-1}$	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$	...	...	...	$C_{2,n-1}$	$\hat{C}_{2,n}$
...	...	...	...	...	...	...	...
i	...	...	...	$C_{i,j}$	...	...	...
...	...	...	...	...	...	...	...
n-1	$C_{n-1,1}$	$C_{n-1,2}$	...	...	...	$\hat{C}_{n-1,n-1}$	$\hat{C}_{n-1,n}$
n	$C_{n,1}$	$\hat{C}_{n,2}$	...	...	...	$\hat{C}_{n,n-1}$	$\hat{C}_{n,n}$

Figure 49: Triangle de Chain Ladder

Afin d'obtenir le nombre de jours estimés au cours de l'année suivante, pour chaque mois de survenance, le nombre de jours indemnisés au 31/12/N-1 est soustrait au nombre de jours estimés au 31/12/N.

Cela revient à calculer (dans le cadre du pas mensuel) :

$$Provision = \text{Douzième Diagonale} - \text{Diagonale Principale} ,$$

La diagonale principale est la dernière diagonale du triangle cumulé (cases grises et nombres en blanc). La douzième diagonale correspond à la diagonale principale décalée de 12 mois (cases grises et nombres en rouge). Pour les mois de survenance pour lesquels la projection ne va pas jusqu'à 12 mois, la dernière colonne est prise.

Date de surv/ Mois	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
Janv-05	81260	142244	174725	197725	211522	219536	225353	230112	234106	237843	241081	244128	246727	248874	250956	252724	254265	255745	257075	258254	259541	260750	261909	263035	264135
Fév-05	79060	137106	164246	183423	194719	202595	208077	212235	215844	218907	221755	224261	226367	228473	230393	232315	234117	235948	237664	239174	240659	242101	243525	244945	246345
Mars-05	78036	133563	162106	182214	194245	202134	207484	211997	215794	219232	222142	224603	227269	229712	232098	234357	236524	238631	240580	242459	244196	245827	247388	248983	250456
Avr-05	61851	113627	140261	160793	174117	182619	189170	194223	198763	202699	205975	209337	212333	215222	217723	220161	222490	224579	226509	228295	230035	231738	233417	235044	236644
Mai-05	59006	114727	144505	164463	175777	183179	188306	192406	195928	198846	201844	204462	206933	209106	211083	212846	214497	216047	217512	218951	220341	221570	222649	223662	224650
Juin-05	62246	116381	139165	153910	162036	167180	171036	174153	176585	179015	181136	183045	184813	186589	188301	189786	191160	192377	193542	194673	195774	196746	197618	198388	198979
Juil-05	47326	80826	97385	108904	115554	120109	123202	125646	128068	130174	132164	133916	135635	137214	138589	139824	140902	141973	142968	143924	144838	145645	146355	146979	147526
Août-05	35949	76100	97660	110754	118776	123276	125804	128397	131615	134033	136219	138364	140479	142405	144296	145947	147564	148674	149716	150718	151675	152521	153264	153923	154512
Sept-05	60518	114162	137765	151944	159055	164150	168674	172300	175593	178439	181135	183701	185892	187849	189645	191302	192907	194359	195721	197031	198282	199483	200638	201750	202820
Oct-05	63232	104819	119413	137862	147919	155787	161259	166015	170008	173691	177075	180054	182822	185080	187150	188829	190415	191846	193192	194484	195719	196810	197769	198612	199365
Nov-05	50885	83459	109038	126505	137745	144946	150337	154353	158036	161525	164500	167357	169772	172000	173746	175305	176775	178107	179355	180556	181702	182715	183605	184394	185145
Déc-05	32326	72436	91069	109389	119580	126727	131743	135901	139680	142986	145888	148414	150633	152511	154036	155419	156729	157902	159009	160074	161089	161988	162776	163476	164117
Janv-06	77993	138760	173580	198193	213427	223146	229991	235610	240297	244374	247781	250500	252462	254681	256981	259450	261778	263974	265961	267852	269618	271330	272843	274171	275366
Fév-06	64405	119378	146107	165157	176648	184904	189570	193456	196856	199647	202225	205009	207591	209894	211994	213896	215690	217314	218837	220301	221708	222937	224022	224985	225857
Mars-06	69728	127358	158515	181380	195796	205579	212088	217160	221399	225221	228615	231763	234682	237286	239659	241809	243838	245673	247395	249051	250633	252036	253257	254366	255366
Avr-06	53811	100307	125064	144398	156265	163791	169274	173456	177121	180164	182879	185396	187732	189814	191713	193433	195056	196524	197902	199226	200491	201609	202591	203462	204244
Mai-06	56542	109620	136576	154468	164214	170580	174736	177981	181504	184622	187404	189984	192377	194516	196457	198219	199883	201387	202798	204156	205452	206698	207804	208796	209695
Juin-06	59560	112173	136969	153249	163240	169355	173476	177477	180989	184097	186871	189445	191830	193959	195891	197658	199315	200815	202228	203576	204869	206014	207014	207905	208705
Juil-06	48285	84067	100813	112389	118590	122521	126128	129038	131591	133851	135865	137739	139474	141021	142432	143709	144915	146006	147029	148017	148953	149783	150513	151160	151767
Août-06	36554	77811	99324	113607	121651	126852	130588	133605	136243	138584	140672	142609	144405	146007	147467	148791	150039	151168	152228	153241	154201	155080	155835	156505	157100
Sept-06	61915	118464	145261	162110	173845	180798	186122	190415	194183	197518	200499	203252	205816	208098	210180	212066	213845	215459	216955	218417	219804	221029	222106	223069	223909
Oct-06	66941	112917	137333	146613	156807	163517	168276	172203	175617	178634	181325	183822	186137	188202	189988	191790	193400	194855	196213	197534	198789	199997	200870	201734	202494
Nov-06	54919	97333	109733	124293	132934	138617	142696	145911	148379	151137	153535	155798	157988	159988	161844	163595	165188	166645	168057	169424	170749	172038	173201	174244	175174
Déc-06	32210	58682,89	71956,4	81489,07	87153,3	90880,06	93556,17	95714,14	97607,98	99284,59	100780,6	102168	103455,1	104602,8	105649	106597	107491,3	108300,4	109059,4	109789,3	110486,5	111102,6	111643,6	112123,6	112595,6

Figure 50: Exemple de triangle de Chain Ladder complété

### 3.1.2 Résultats

Les calculs à l'aide des triangles de développement donnent le nombre total des jours futurs indemnisés sur un an pour les sinistres survenus et ouverts au 31/12/2006. Ainsi, on obtient le nombre de jours indemnisés au cours de l'année 2007 au titre de la survenance 2006.

Indemnisation	Année	Réel	Estimé	Ecart en %	Volatilité	Borne Min	Borne Max
Nombre total des jours futurs indemnisés sur un an	2007	399 924	466 979	16,8%	8 142	451 021	482 938

Tableau 51: Résultats de la méthode de Chain Ladder

Ces résultats prennent en compte les IBNR ; c'est-à-dire les sinistres survenus en 2006 mais indemnisés à partir de 2007.

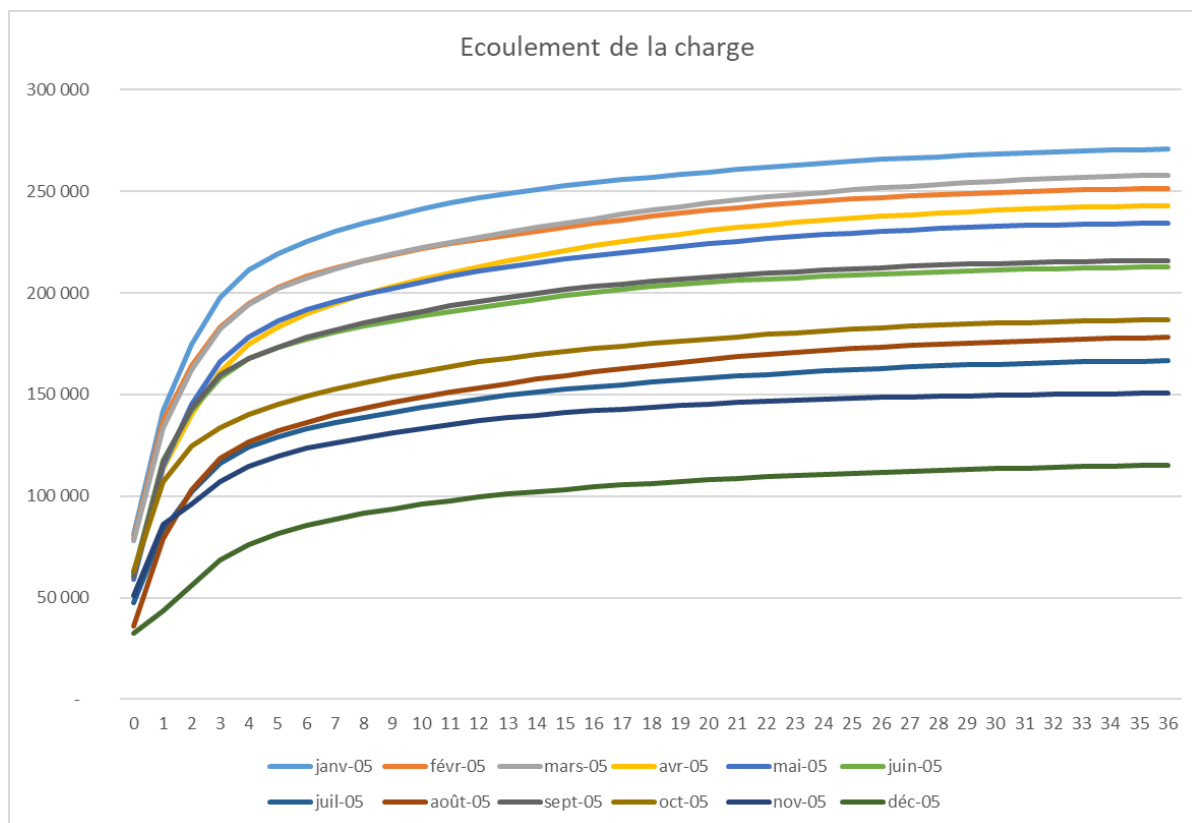
On constate un boni de liquidation 16,8%. Cependant, en fonction du nombre de mois pris pour le calcul des coefficients, on obtient des résultats variables : avec 6 mois d'historique, on obtient un boni de 22,1% et un boni de 16,5% avec 4 mois d'historique.

	2007	Reel à 1 an	Prov à 1 an	Ecart absolu	Ecart relatifs
12 mois		399 924,00	466 979,43	67 055,43	16,77%
9 mois		399 924,00	486 455,28	86 531,28	21,64%
7 mois		399 924,00	474 442,06	74 518,06	18,63%
6 mois		399 924,00	488 259,48	88 335,48	22,09%
5 mois		399 924,00	488 915,00	88 991,00	22,25%
4 mois		399 924,00	465 753,87	65 829,87	16,46%

Tableau 52: Résultats de Chain Ladder en fonction du nombre de mois d'historique pour le calcul des coefficients de passage

On note que les résultats obtenus avec la méthode de Chain Ladder sont volatiles en fonction du nombre de mois d'ancienneté pris en compte pour le calcul des coefficients de passage. Cela rend cette méthode peu fiable avec les données utilisées.

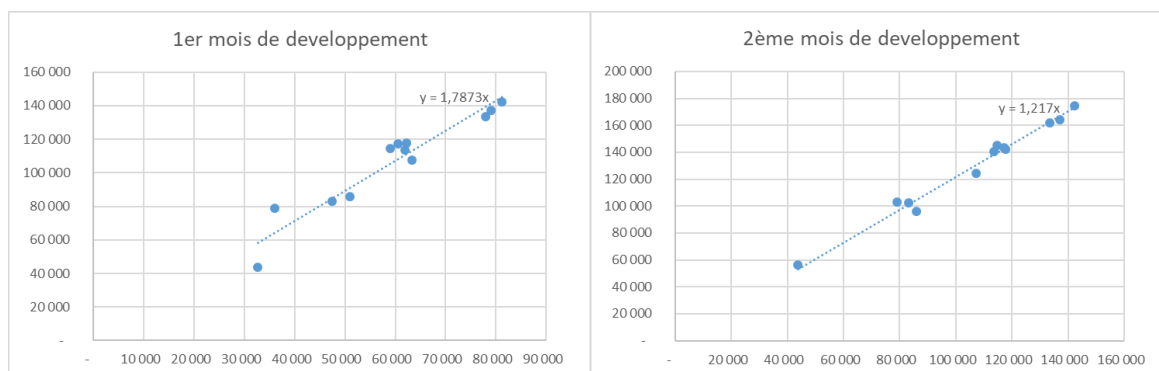
**Figure 53: Courbe représentative du triangle d'écoulement**

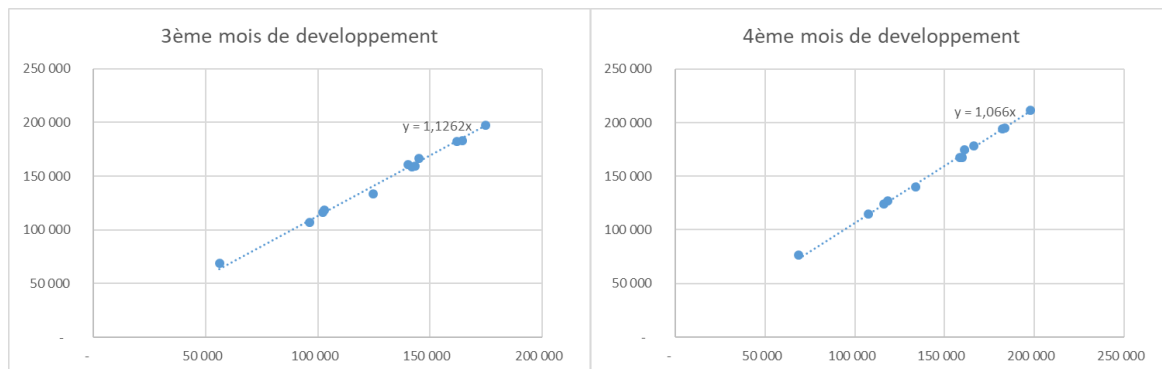


Ainsi, on observe que les sinistres sont réglés au bout de 24 mois.

Les graphiques suivants permettent de valider l'utilisation de la méthode de Chain Ladder avec les données utilisées.

**Tableaux : Validation des hypothèses de Chain Ladder**





Ces graphiques représentent les éléments de l'année  $i+1$  de développement du triangle selon l'année  $i$ .

On peut noter que l'ensemble des points est relativement bien aligné sur une droite passant par l'origine. Ces éléments valident bien l'utilisation de la méthode de Chain Ladder avec les données présentes.

Cependant, la méthode de Chain Ladder donne un boni important ; la méthode avec une table d'expérience est donc testée dans la partie suivante afin d'essayer d'obtenir des résultats plus proches du nombre de jours réellement indemnisés.

## 3.2 Résultats avec des tables d'expérience

### 3.2.1 Principe

L'objectif est de comparer le nombre de jours effectivement indemnisés avec l'estimation du nombre de jours qui auraient été payé à partir des tables d'expérience. Les estimations de la durée indemnisée sont réalisées sur la base de la table d'expérience.

Nous disposons des arrêts dont les dates de survenance sont comprises entre le 01/01/2005 et le 31/12/2007.

Les projections des nombres de jours indemnisés sont effectuées par exercice civil. Pour les arrêts en cours au 31/12/ $N-1$ , les durées d'arrêt résiduelles pour l'année  $N$  sont estimées et sont comparées avec la durée effective des prestations sur cette même année.

La durée d'indemnisation effective des arrêts en cours au 31/12/ $N-1$  sur l'année  $N$  est estimée comme le nombre de jours, plafonnée à 1 an, séparant la date de fin des indemnisations et le 31/12/ $N-1$ .

Ce mécanisme de calcul est reconduit successivement pour les années 2006 à 2007. Par exemple, on estime pour les arrêts en cours au 31/12/2005, la durée d'arrêt résiduelle sur 2006, et ainsi de suite. Au final, on obtient pour les années 2006 à 2007 une estimation annuelle des durées d'indemnisation cumulées pour les arrêts en cours au début de chaque année considérée.

Il est donc possible de comparer pour les arrêts en cours au 31/12 des années 2006 à 2007:

- Le nombre de jours estimés, calculé sur la base des tables d'expérience,
- le nombre de jours réellement indemnisés.

$$\text{Boni/Mali} = (\text{Durée Provisionnée} - \text{Durée Réelle}) / \text{Durée Réelle}.$$

Il convient donc de définir ce qu'est un sinistre « en cours » à une date donnée. A la date d'arrêt 31/12/ $N$ , les sinistres en cours sont sélectionnés de la manière suivante :



- Date d'Entrée  $\leq 31/12/N$ ,
- Date de Sortie  $> 31/12/N$

Les formules de calcul sont décrites ci-après. Les formules fournies concernent des durées sur un exercice comptable.

### 3.2.2 Calcul pour un euro de prestation

Les règlements espérés de l'année  $N$  d'un assuré, entré en incapacité à l'âge  $x$  dont l'ancienneté dans l'arrêt est  $a$  sont calculés à l'aide de la formule suivante :

$$R(x, a) = \frac{1}{2} \sum_{k=a}^{\min(35, a+1)} \left[ \frac{l_k^x + l_{k+1}^x}{l_a^x} \right].$$

On prend le minimum entre 35 et  $a+1$  (pour la projection à 1 an), car la table va de 0 à 35 mois d'ancienneté (soit les 3 ans d'incapacité).

Cette formule est en accord avec celle du BCAC, elle permet d'estimer le nombre de jours d'indemnisation espéré sur un an à partir de la loi de maintien en incapacité  $(l_k^x)_{x,k}$  ( $x$  représente l'âge à l'entrée et  $k$  l'ancienneté).

Afin d'intégrer les incertitudes liées aux fluctuations d'échantillonnage, on construit un intervalle de confiance autour du montant des règlements espérés. Cet intervalle de confiance est construit à partir de l'estimation de la volatilité du montant global des règlements espérés<sup>7</sup>. Le calcul s'effectue en trois étapes :

- première étape : estimation des volatilités individuelles des règlements espérés de l'année  $N$  au moyen de l'expression suivante :

$$\sigma_{x,a} = \sqrt{\sum_{k=a}^{\min(35, a+1)} \left( k - a + \frac{1}{2} \right) \left[ \frac{l_k^x + l_{k+1}^x}{l_a^x} \right] - R(x, a)^2},$$

- deuxième étape : estimation de la volatilité de la provision globale (sous l'hypothèse d'indépendance des arrêts entre les assurés) :

$$\sigma_{\text{provision}} = \sqrt{\sum_{i=1}^{I_N} \sigma_{x_i, a_i}^2},$$

- troisième étape : on déduit des deux premières étapes l'intervalle de confiance à 95 % suivant :

$$IC_{\text{provision}} = \left[ P - 1,96 \times \sigma_{\text{provision}}, P + 1,96 \times \sigma_{\text{provision}} \right],$$

avec  $P = \sum_{i=1}^{I_N} R(x_i, a_i)$  le montant estimé des règlements de l'exercice  $N$ . La probabilité que le

montant des prestations réglées observé se situe dans l'intervalle de confiance  $IC_{\text{provision}}$  est donc approximativement égale à 95 %.

<sup>7</sup> L'intervalle de confiance est estimé en approchant la vraie loi du montant global estimé par une loi normale.

### 3.2.3 Résultats avec la population globale

La comparaison des prestations estimées et des prestations réelles conduit au tableau suivant. Pour chaque année  $N$  de 2006 à 2007, on compare sur le périmètre des sinistres ayant fait l'objet d'un règlement en  $N+1$  et en cours au 31/12/ $N-1$ , le nombre de jours effectivement réglés en  $N+1$  et la durée estimée par la table d'expérience.

Indemnisation	Année	Nombre de sinistres	Réel	Estimé	Ecart en %	Volatilité	Borne Min	Borne Max
Nombre total des jours futurs indemnisés sur un an	2007	1893	293 425	266 483	- 9,18	10 782	245 350	287 616

**Tableau 54: Résultats avec la table construite sur la population totale**

On constate un mali de liquidation de 9,2%.

Nous avons observé dans la partie 2 de ce mémoire, des pics de fréquence au 01/01 pour les années 2006 à 2008 pour les dates de début d'indemnisation. Les résultats précédemment présentés sont obtenus en décalant de 6 mois avant et après les dates des sinistres ayant cette particularité.

Nous aurions pu également supprimer ces sinistres atypiques. Cette démarche donne les résultats suivants :

Indemnisation	Année	Nombre de sinistres	Réel	Estimé	Ecart en %	Volatilité	Borne Min	Borne Max
Nombre total des jours futurs indemnisés sur un an	2007	1599	261 010	230 284	- 11,77	10 114	210 461	250 107

**Tableau 55: Résultats sur la population globale avec les sinistres ayant une date de début d'indemnisation au 01/01 supprimés.**

On peut noter que les résultats selon chaque méthode sont relativement proches et dans les mêmes ordres de grandeur.

Les résultats obtenus nous mènent à étudier plus précisément l'hétérogénéité de la population.

### 3.2.4 Analyse de l'hétérogénéité de la population

Une étude des durées moyennes d'indemnisation et des anciennetés à l'entrée et à la sortie des sinistres ouverts au 31/12 de chaque année est réalisée.

Figure 56: Graphique des durées par classe de population au 31/12/2005

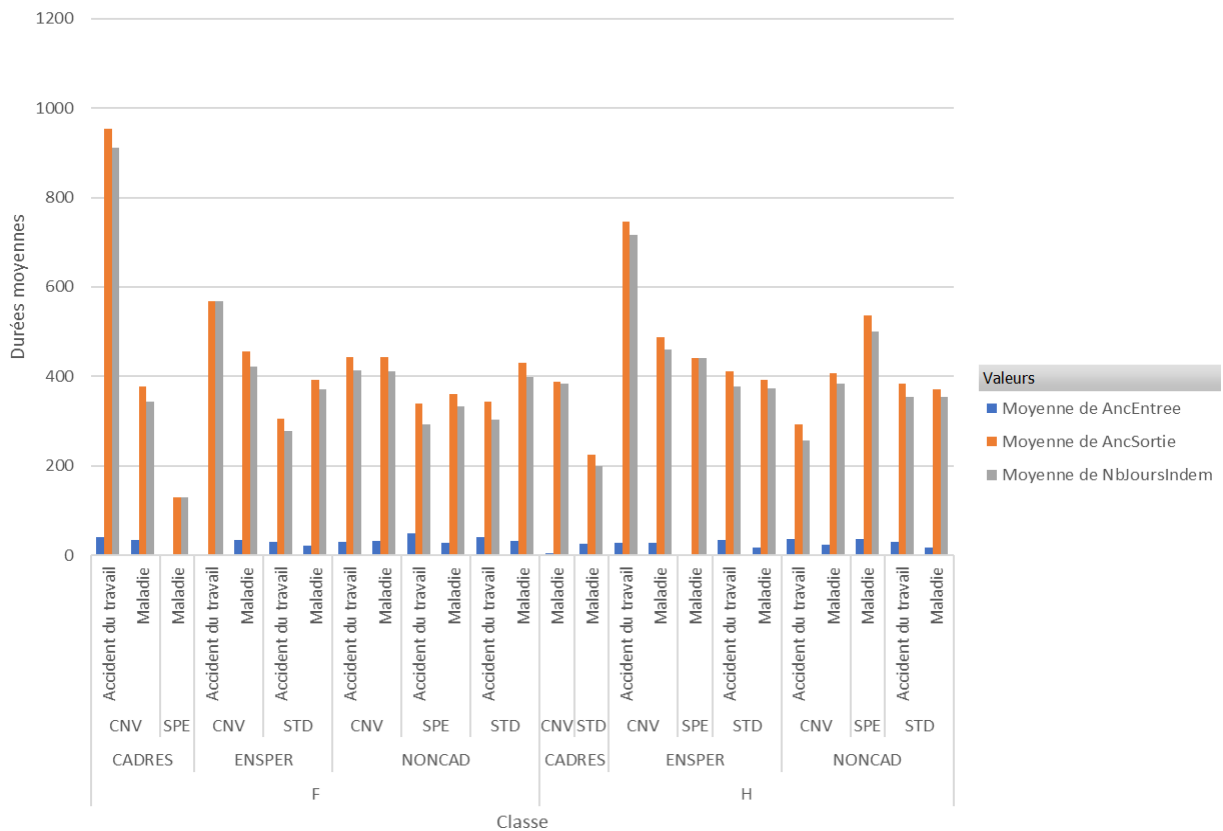
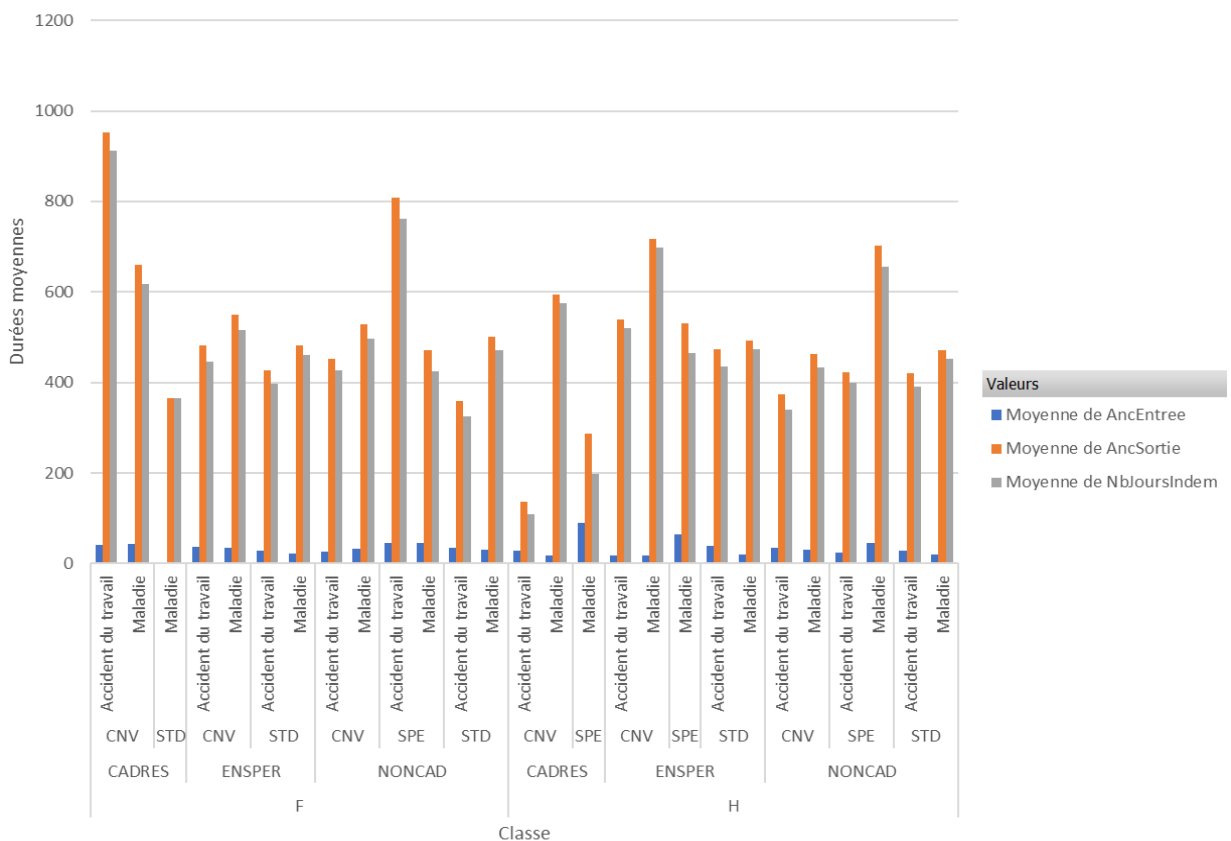


Figure 57: Graphique des durées par classe de population au 31/12/2006



On observe que les années 2005 (année d'apprentissage) et 2006 (année de test) sont relativement similaires. Cependant, on peut noter que selon les classes les durées moyennes d'indemnisation varient de manière importante. Certaines classes possèdent peu d'éléments ce qui rend l'interprétation des moyennes plus compliquées.

Afin d'avoir un modèle plus proche des données observées, la population est segmentée à l'aide du modèle de Cox. Des tables de maintien en incapacité sur cette population segmentée sont utilisées dans la partie suivante pour calculer le montant de provision à un an.

### 3.2.5 Résultats avec la population segmentée

Le calcul du nombre de jours indemnisés provisionné est réalisé à partir d'une table construite sur la population la plus importante et les coefficients de Cox. Les durées provisionnées sont ensuite comparées aux durées réelles selon la même méthode que pour la population globale.

Pour rappel, on obtient les coefficients de Cox suivants :

Variable	Catégorie	Coef	Exp(coef)	Exp(-coef)
CSP	Ensemble du Personnel	- 0,042	0,959	1,043
	Cadres	- 0,246	0,782	1,279
Mode de souscription	Standard	- 0,007	0,993	1,007
	Spécial	- 0,113	0,893	1,119
Type d'arrêt	Accident du Travail	0,015	1,015	0,985
Sexe	Homme	0,154	1,167	0,857

Figure 58: Tableau des coefficients de Cox

La comparaison des prestations estimées et des prestations réelles conduit au tableau suivant. Pour l'année  $N = 2007$ , on compare, sur le périmètre des sinistres ayant fait l'objet d'un règlement en  $N+1$  et en cours au 31/12/ $N-1$ , le nombre de jours effectivement réglés en  $N+1$  et la durée estimée par la table d'expérience :

Indemnisation	Année	Nombre de sinistres	Réel	Estimé	Ecart en %	Volatilité	Borne Min	Borne Max
Nombre total des jours futurs indemnisés sur un an	2007	1893	293 425	310 658	5,87%	11 457	288 203	333 113

Tableau 59: Résultats avec la table construite sur la population segmentée

On constate un boni de liquidation de 5,9% pour l'année 2007.

Concernant la volatilité des prestations, on observe que les nombres de jours indemnisés réels sont dans l'intervalle de confiance à 95 % (déterminé par les bornes Min et Max), signe que la table représente correctement les données.

De même que pour la population globale, nous avons calculé le montant de provision sans les sinistres ayant une date de début d'indemnisation au 01/01. On obtient des résultats similaires à ceux précédemment observés.

Indemnisation	Année	Nombre de sinistres	Réel	Estimé	Ecart en %	Volatilité	Borne Min	Borne Max
Nombre total des jours futurs indemnisés sur un an	2007	1599	261 010	267 991	2,67%	10 739	246 942	289 040

Tableau 60: Résultats avec la table construite sur la population segmentée sans les sinistres ayant une date de début d'indemnisation le 01/01

On observe avec cette seconde méthode un boni de 2,7%. On obtient avec cette méthode des résultats plus précis, dans le sens plus proche du montant réellement indemnisé, qu'avec la méthode de la répartition des sinistres atypiques.

## Synthèse

Dans cette partie, l'évaluation de la provision à un an et sa comparaison au montant réellement indemnisé ont été réalisées à l'aide de plusieurs méthodes différentes afin de trouver celle la plus adaptée aux données et au type de sinistre étudié.

La première méthode appliquée est la plus traditionnelle, celle de Chain Ladder. La volatilité des résultats obtenus, en fonction du nombre de mois pris en compte pour la détermination des coefficients de passage, ne permet pas de conclure que cette méthode est fiable pour le calcul du montant de provision dans le cadre des données observées.

La seconde méthode utilisée est celle d'une table de maintien construite à partir des données. Cette méthode est celle habituellement utilisée dans le cas de données d'arrêt de travail ou de décès. Les résultats obtenus sont plus probants que ceux de Chain Ladder mais sous-estiment le montant à provisionner.

Nous noterons que malgré un nettoyage de la base pour obtenir une base homogène et étudiable, nos données semblent être relativement atypiques. Afin de s'adapter au mieux à ces données, la méthode ligne à ligne est appliquée à la population segmentée grâce au modèle de Cox. Un boni global est observé. Cette dernière méthode est la plus adaptée pour provisionner les données étudiées.

On peut donc conclure que la méthode la plus précise avec les données utilisées pour le calcul des provisions est une méthode ligne à ligne avec segmentation. On observe que plus la méthode est construite sur une maille fine des données, plus le résultat obtenu est proche de celui attendu.

Cependant, la base étudiée a la particularité d'avoir un volume important. Cela permet de faire une segmentation de la population. Or ceci n'est pas toujours faisable lorsque les données disponibles ont un volume moins important.



## Conclusion

Le provisionnement des sinistres en cours est un élément essentiel du passif des assureurs. De nombreuses méthodes sont utilisées sur le marché.

Notre étude permet d'identifier les méthodes à appliquer selon le type de sinistres et de risques. Une application de choix de méthodes a été réalisée à l'aide de données d'arrêt de travail. Traditionnellement, des tables d'incapacité réglementaires ou certifiées sont utilisées.

Les données étudiées ont la particularité d'être en volume important, ce qui a permis en plus de la méthode de Chain Ladder et de tables couramment utilisées, d'appliquer des tables avec segmentation de la population selon des variables explicatives.

Les résultats obtenus montrent que la méthode de Chain Ladder ne semble pas adaptée à ce type de sinistres. Les résultats étant relativement volatiles en fonction du nombre de mois d'ancienneté pris pour obtenir les coefficients de passage.

La méthode ligne à ligne avec une table sur la population globale obtient de meilleur résultat que Chain Ladder. Cependant les malis obtenus sont encore relativement importants.

Finalement, la méthode la plus concluante est celle ligne à ligne avec une population segmentée. Nous obtenons donc pour nos données que : plus la maille d'observation et d'apprentissage des données est fine plus le montant de provisions est précis.

Cependant, tous les portefeuilles ne permettent pas de segmenter comme nous avons pu le faire par manque de volume ou manque de variables explicatives. Les modèles ligne à ligne ne sont pas encore développés ou généralisés pour tous les types de sinistres.

Il serait intéressant pour compléter cette étude de calculer la durée d'indemnisation par sinistres directement à l'aide d'un modèle de régression (GLM ou boosting). En effet, les données étant très peu censurées, nous disposons d'informations suffisantes pour développer un modèle de régression.

## Références

- ACPR - *Le rapport du groupe de travail sur le calcul du Best Estimate en Assurance Dommage*
- ANTONIO K., PLAT R. [2012] *Micro-level stochastic loss reserving for general*
- ARJAS E., [1989]. *The claims reserving problem in non-life insurance: Some structural ideas*. ASTIN
- ASTIN [2016] *Non-Life reserving practices*, Report 2016, ASTIN
- BLUM K-A., OTTO D-J. [1998] *Best Estimate loss reserving: an actuarial perspective* Casualty Actuarial Society Forum Fall 1998, 55-101.
- BOUMEZOUED A., DEVINEAU L. [2017], *Individual claims reserving: a survey*. <hal01643929>
- BOUMEZOUED A., DEVINEAU L., TAILLIEU F. [2017], *The future of (individual) reserving*, Milliman White Paper
- BOURRY C., TAILLIEU F., WEISSLINGER R., [2018] *Assurance Construction* , Milliman White Paper
- CHARPENTIER A., DENUIT M. ; [2005] *Mathématiques de l'assurance non-vie, tome 2, Tarification et provisionnement*, Economica
- CHARPENTIER A., DEVINEAU L., NESSI JM. [2010] *Mesurer le risque lors du calcul des provisions pour sinistres à payer*.
- Code des Assurances
- Elias C., [2013] *Construction d'une table d'expérience sur le maintien en invalidité*
- JEWELL W-S., [1989] *Predicting IBNYR events and delays: I. continuous time*, ASTIN
- Journal officiel de l'Union Européenne [2009], *Directive 2009/138/CE du parlement européen et du conseil du 25 novembre 2009*.
- Journal officiel de l'Union Européenne [2014], *Règlement délégué (UE) 2015/35 de la commission du 10 octobre 2014*.
- LAURENT J.P., NORBERG R., PLANCHET F. [2016] *Modelling in life insurance – a management perspective*, EAA Series, Springer.
- LUSTMAN F., LE VALOIS F., LEFLAIVE V., PEQUEUX O., [2001] *Contrôle prudentiel en situation de crise*, revue Risques, n°58
- LUZI M. [2006] *Assurance IARD*, Economica
- MACK T., [1993] *Distribution-free calculation of the standard error of chain ladder reserve estimates*, ASTIN
- NORBERG R., [1993] *Prediction of outstanding liabilities in non-life insurance* , ASTIN
- PATRAT C. ; BESSON J-L., [2005] *Assurance Non-Vie - Modélisation, Simulation*, Economica



PARTRAT C., LECOEUR E., NISIPASU E., NESSI J-M., REIZ O. [2007] *Provisionnement technique en Assurance Non-vie*, Economica

PLANCHET F., THÉRON P., [2006] *Modèles de durée Application actuarielles*, Economica

PLANCHET F., WINTER J. [2006] *Les provisions techniques des contrats de prévoyance collective - détermination et pilotage*, Paris : Economica

RAHMOUNI A., [2016] *Construction de lois expérimentées en incapacité temporaire de travail*

THERNEAU T., GRAMBSCH P. [2000] *Modeling Survival Data: Extending the Cox Model*, Springer

# Lexique

ACPR : Autorité de Contrôle Prudentiel et de Résolution

BCAC : Bureau Commun d'Assurances Collectives

CSP : Catégories Sociaux-Professionnelles

DROC : Date Réglementaire d'Ouverture de Chantier

IBNR : Incurred but Not Reported (Sinistres survenus et non encore déclarés)

PE : Provision d'Egalisation

PM : Provision Mathématique

PPNA : Provision pour Primes Non Acquises

PRC : Provision pour Risques Croissants

PRE : Provision pour Risque d'Exigibilité

PREC : Provision pour Risque En Cours

PSAP : Provision pour Sinistres à payer

PSEM : Provision pour Sinistres Non Encore Manifestés

RC : Responsabilité Civile

SCR : Solvency Capital Requirement

## **Annexes**

## Annexe A

# Classification par Branches

### A.1 Code des Assurances

En droit français : article R32-1 du Code des Assurances :

- 1- Accidents
- 2- Maladie
- 3- Corps de véhicules terrestres (autres que ferroviaires)
- 4- Corps de véhicules ferroviaires
- 5- Corps de véhicules aériens
- 6- Corps de véhicules maritimes, lacustres et fluviaux
- 7- Marchandises transportées
- 8- Incendie et éléments naturels
- 9- Autres dommages aux biens
- 10- Responsabilité civile véhicules terrestres automoteurs
- 11- Responsabilité civile véhicules aériens
- 12- Responsabilité civile véhicules maritimes, lacustres et fluviaux
- 13- Responsabilité civile générale
- 14- Crédit
- 15- Caution
- 16- Pertes pécuniaires diverses
- 17- Protection juridique
- 18- Assistance
- 20- Vie-Décès
- 21- Nuptialité-Natalité
- 22- Assurances liées à des fonds d'investissement
- 23- Opérations tontinières
- 24- Capitalisation
- 25- Gestion de fonds collectifs

## A.2 Solvabilité II

### Classification par risques

- Dépenses médicales
- Protection du revenu
- Indemnités de travail
- Responsabilité automobile, tiers
- Automobile, autres classes
- Marine, aviation et transport
- Incendie et autres dommages de biens
- Responsabilité civile, tiers
- Crédit et assurance caution
- Juridique
- Assistance
- Assurance non-vie divers.

## Annexe B

### Récapitulatif étude ACPR

Récapitulatif des éléments de l'étude : « *Rapport du groupe de travail sur le calcul du Best Estimate en Assurance Dommage* » de l'ACPR de 2007 (cf. ACPR [2007]) présente les techniques les plus adaptées et les méthodes approchées possibles par catégories de risques.

Risques	Déroulement	Problématiques	Méthodes
<b>RC matérielle et dommages automobile</b>	Court	- Augmentation des cadences de règlements due à l'amélioration du processus de gestion	- Méthodes avec des cadences appliquées aux paiements, recours, nombres de charges de sinistres D/D
<b>RC Corporelle Auto</b>	Long	- Changement dans les méthodes de calcul des dossiers sinistres D/D - Séparation des arrérages et les PM de rentes en service pour les données - Déroulement long qui nécessite un historique important et des données séparées Graves et hors graves	- Méthodes cadences appliquées aux paiements, recours, nombres de charges de sinistres D/D - Méthodes d'exposition appliquées aux paiements - Méthode avec des loss ratios ultimes en fonction de la sinistralité déjà survenue et de la politique de tarification
<b>Transport</b>	Moyen (Corps et facultés) Long (RC)	- Marché très concurrentiel ce qui entraine une volatilité importante d'une année sur l'autre - Nombre de sinistres tardifs importants à cause du mode de distribution et du mode de rattachement	- Méthodes cadences appliquées aux paiements, recours, nombres de charges de sinistres D/D - Méthodes d'exposition appliquées aux paiements - Méthode avec des loss ratios ultimes en fonction de la sinistralité déjà survenue et de la politique de tarification
<b>Incendie/Dommages aux biens</b>	Long (court hors sinistre lourd)	- Augmentation des cadences de règlements due à l'amélioration du processus de gestion	- Méthodes cadences appliquées aux paiements, recours, nbres de charges de sinistres D/D
<b>RC générale</b>	Très long		- Méthodes cadences appliquées aux paiements, recours, nbres de charges de sinistres D/D
<b>Assistance (Auto, Voyage, Santé, Assistance à domicile)</b>	Inférieur ou égale à 2 ans		- Dossier/Dossier, Cadences (Nombres charges, recours, clôtures) - Les cas de garantie inférieur à 1 an , les méthodes de cadences ne sont pas adaptées sauf si on a une base mensuelle ou trimestrielle)
<b>Protection juridique</b>	5-10 an minimum		- Dossier/Dossier - Cout moyen (IBNR), - Cadences de nombres de sinistres
<b>Construction</b>	15 ans (par survenance), 30 ans (par DROC)	- Analyse des PSAP et des PSEM - DO: Analyse indispensable du taux de recours - PSEM: besoin d'un historique important	- Méthodes de cadences (de liquidation, de manifestation) - Pour les DROC récentes et pour les sinistres graves : méthodes d'exposition
<b>Evènements naturels</b>		- Les événements étant rares, l'utilisation des cadences peut être compliqué. Les événements peuvent être trié par date de survenance. On peut utiliser des méthodes basées sur des couts moyens et le nombre de sinistres déclarés par événements.	- Méthodes Fréquence/Coût

Risques	Déroulement	Problématique	Méthodes
<b>Crédit/Cautio</b>	Variable de quelques mois à plusieurs années selon le type de garantie	<ul style="list-style-type: none"> <li>- Faible nombres de sinistres (risque de sévérité plutôt qu'un risque de fréquence) sur certains type de garantie</li> <li>- Détermination difficile de la date de survenance du sinistre (surtout en cas de caution simple)</li> <li>- Besoin d'un historique important.</li> <li>- Changement dans les méthodes de calcul des dossiers sinistres D/D (notamment les forfaits)</li> </ul>	<ul style="list-style-type: none"> <li>- Méthode des cadences appliquées aux charges brutes de recours. (avec, si la quantité de données le permet, le retraitement des sinistres graves)</li> <li>- Etude de la cadence des paiements et des coûts à l'ultime</li> <li>- Pour les écoulements longs : modèles de durée</li> </ul>
<b>Acceptation de réassurance</b>	Variable en fonction des branches sous-jacentes (moyen à long)	<ul style="list-style-type: none"> <li>- Opérations multi-devises - cours unique pour tous les exercices ou application d'un cours par année d'origine.</li> <li>- Retard dans la réception des données / manque d'informations détaillées chronique (ex.: décomposition par chef de préjudice des sinistres souvent non disponible en XL)</li> <li>- Comptes complets / incomplets</li> <li>- Traitements comptables (survenance, clean cut...)</li> </ul>	<ul style="list-style-type: none"> <li>- Méthodes de cadence appliquées aux règlements et aux charges</li> <li>- Méthodes d'expositions sur primes</li> <li>- Coûts moyens sur les évènements exceptionnels</li> </ul>

Source : ACPR - Le rapport du groupe de travail sur le calcul du Best Estimate en Assurance Domage (2007)



## Annexe C

### Démonstrations

#### Estimateur de $p_i = P(T_x > t_i | T_x > t_{i-1})$ (Kaplan-Meier)

La probabilité d'observer  $d_i$  décès à l'instant  $t_i$  est donnée par une loi binomiale

$$B(r_i, (1 - p_i)),$$

$r_i$  correspond au nombre de personnes à risque au temps  $t_i$ , c'est-à-dire au nombre de personnes à risque en  $t_{i-1}$  moins les décès et les individus censurés du temps précédent.

$$r_i = r_{i-1} - d_{i-1} - c_{i-1}.$$

La fonction de vraisemblance sur  $]t_{i-1}, t_i]$ ,

$$L(d_i, p_i) = \binom{d_i}{r_i} (1 - p_i)^{d_i} (p_i)^{r_i - d_i}.$$

La log-vraisemblance s'écrit,

$$\text{Log}L(d_i, p_i) = \ln \left( \binom{d_i}{r_i} (1 - p_i)^{d_i} (p_i)^{r_i - d_i} \right),$$

$$\text{Log}L(d_i, p_i) = \ln \left( \binom{d_i}{r_i} \right) + d_i \ln(1 - p_i) + (r_i - d_i) \ln p_i.$$

La dérivée par rapport à  $p_i$  donne :

$$\frac{\partial \text{Log}L(d_i, p_i)}{\partial p_i} = \frac{r_i - d_i}{p_i} - \frac{d_i}{1 - p_i}.$$

En annulant la dérivée, on obtient :

$$\frac{\partial \text{Log}L(d_i, p_i)}{\partial p_i} = \frac{r_i - d_i}{p_i} - \frac{d_i}{1 - p_i} = \frac{r_i - d_i - p_i r_i}{p_i(1 - p_i)} = 0,$$

et

$$p_i = 1 - \frac{d_i}{r_i}.$$

**Démonstration de la formule :**  $\Delta^n u(x) = \sum_{j=0}^n C_n^j (-1)^{n-j} u(x+j)$ .

L'opérateur différence avant est défini par :

$$\begin{cases} \Delta u(x) = u(x+1) - u(x), \\ \Delta^n u(x) = \Delta \left( \Delta \left( \dots \left( \Delta u(x) \right) \right) \right), \end{cases}$$

On prouve la formule (1) suivante, par récurrence,

$$\Delta^n u(x) = \sum_{j=0}^n C_n^j (-1)^{n-j} u(x+j) \quad (1),$$

Pour  $k=2$ ,

$$\begin{aligned} \Delta^2 u(x) &= \Delta(\Delta u(x)) = \Delta(u(x+1) - u(x)), \\ \Delta^2 u(x) &= [u(x+2) - u(x+1)] - [u(x+1) - u(x)], \\ \Delta^2 u(x) &= u(x+2) - 2 * u(x+1) + u(x). \end{aligned}$$

Or,

$$C_2^0 (-1)^{2-0} u(x) + C_2^1 (-1)^{2-1} u(x+1) + C_2^2 (-1)^{2-2} u(x+2) = u(x) - 2 * u(x+1) + u(x+2).$$

Donc, la proposition (1) est vraie pour  $n=2$ .

On suppose que pour tout  $k \geq 2$ , on a  $\Delta^k u(x) = \sum_{j=0}^k C_k^j (-1)^{k-j} u(x+j)$ .

Pour le rang  $k+1$ , on a :

$$\begin{aligned} \Delta^{k+1} u(x) &= \Delta \left( \Delta^k u(x) \right), \\ \Delta^{k+1} u(x) &= \Delta \left( \sum_{j=0}^k C_k^j (-1)^{k-j} u(x+j) \right), \\ \Delta^{k+1} u(x) &= \sum_{j=0}^k C_k^j (-1)^{k-j} u(x+j+1) - \sum_{j=0}^k C_k^j (-1)^{k-j} u(x+j), \end{aligned}$$

$$\begin{aligned} \Delta^{k+1} u(x) &= \left( C_k^k (-1)^{k-k} u(x+k+1) \right) + \left( C_k^{k-1} (-1)^{k-(k-1)} - C_k^k (-1)^{k-k} \right) u(x+k) + \dots \\ &\quad + \left( C_k^0 (-1)^{k-0} - C_k^1 (-1)^{k-1} \right) u(x+1) - C_k^1 (-1)^{k-0} u(x). \end{aligned}$$

Or,

$$C_k^{j+1} (-1)^{k-(j+1)} - C_k^j (-1)^{k-j} = -(-1)^{k-j} \left( \frac{k!}{(j+1)!(k-j-1)!} + \frac{k!}{j!(k-j)!} \right),$$

$$C_k^{j+1} (-1)^{k-(j+1)} - C_k^j (-1)^{k-j} = -(-1)^{k-j} \frac{k!}{j!(k-j)!} \left( \frac{k-j}{j+1} + 1 \right),$$

$$C_k^{j+1} (-1)^{k-(j+1)} - C_k^j (-1)^{k-j} = -(-1)^{k-j} \frac{k!}{j!(k-j)!} \frac{k+1}{j+1},$$

$$C_k^{j+1} (-1)^{k-(j+1)} - C_k^j (-1)^{k-j} = -(-1)^{k-j} \frac{(k+1)!}{(j+1)!(k-j)!},$$

$$C_k^{j+1}(-1)^{k-(j+1)} - C_k^j(-1)^{k-j} = (-1)^{k-j+1}C_{k+1}^{j+1}.$$

On obtient bien :

$$\Delta^{k+1}u(x) = \sum_{j=0}^{k+1} C_{k+1}^j (-1)^{k+1-j} u(x+j).$$

### Démonstration formule WH

On rappelle que :

$$F = \sum_{i=1}^p \sum_{j=1}^q w_{ij} (q_{ij} - \hat{q}_{ij})^2,$$

avec :

- $q_{ij}$  : les taux ajustés avec Whittaker-Henderson à chercher,
- $\hat{q}_{ij}$  : les taux estimés avec Kaplan-Meier,
- $w_{ij}$  : les poids.

Si on remplace dans la formule précédemment démontrée  $k+1$  par  $n$ , on obtient :

$$\Delta^n u(x) = \sum_{j=0}^n C_n^j (-1)^{n-j} u(x+j).$$

Nous rappelons également que la régularité se mesure avec les critères verticaux et horizontaux :

- Le critère vertical est la somme des opérateurs différences verticaux :

$$S_v = \sum_{j=1}^q \sum_{i=1}^{p-z} (\Delta^z_v q_{ij})^2,$$

Avec  $\Delta^z_v$  l'opérateur pour la régularité verticale qui prend  $q_{ij}$  comme une série indexée par  $i$  avec  $j$  fixé.

- Le critère horizontal est la somme des opérateurs différences horizontaux :

$$S_h = \sum_{j=1}^p \sum_{i=1}^{q-z} (\Delta^z_h q_{ij})^2,$$

Avec  $\Delta^z_h$  l'opérateur pour la régularité horizontale qui prend  $q_{ij}$  comme une série indexée par  $j$  avec  $i$  fixé.

La fonction à minimiser est donc :

$$M(q) = F(q) + \alpha S_v(q) + \beta S_h(q).$$

Ce qui revient à résoudre :

$$\frac{\partial M}{\partial q} = 0.$$

Une résolution matricielle est utile pour résoudre ce problème d'optimisation. Certains éléments sont à changer ou à construire pour cela.

- Réorganisation des  $\hat{q}_{ij}$  en un vecteur colonne  $u$  de taille  $p \times q$  tel que

$$u_{q(i-1)+j} = \hat{q}_{i,j}.$$

- La matrice des poids est construite en mettant sur les diagonales les lignes de  $(w_{ij})$  :

$$W_{q(i-1)+j, q(i-1)+j}^* = w_{ij}.$$

- Construction de la matrice  $K_v^z$  de dimension  $(n(p-z), m)$  tel que

$$S_v = {}^t(qK_v^z)(K_v^z q).$$

- Construction de la matrice  $K_h^y$  de dimension  $(p(n-y), m)$

$$S_h = {}^t(qK_h^y)(K_h^y q).$$

Les taux lissés sont calculés de la manière suivante :

$$q^* = (w^* + \alpha^t K_v^z K_v^z + \beta^t K_h^y K_h^y)^{-1} w^* u.$$

### Démonstration :

On peut réécrire :

$$F = {}^t(q-u)w^*(q-u),$$

$$S_v = {}^t(qK_v^z)(K_v^z q),$$

$$S_h = {}^t(qK_h^y)(K_h^y q).$$

Ainsi, on a :

$$M(q) = {}^t(q-u)w^*(q-u) + \alpha^t (qK_v^z)(K_v^z q) + \beta^t (qK_h^y)(K_h^y q).$$

En dérivant par rapport à  $q$ , on obtient :

$$\frac{\partial M(q)}{\partial q} = w^*(q-u) + \alpha^t K_v^z K_v^z q + \beta^t K_h^y K_h^y q = 0,$$

$$(w^* + \alpha^t K_v^z K_v^z + \beta^t K_h^y K_h^y)q = w^* u,$$

$$q = (w^* + \alpha^t K_v^z K_v^z + \beta^t K_h^y K_h^y)^{-1} w^* u.$$

# Annexe D

## Tables d'expérience

Table d'expérience construite sur la population totale

Age à la survenance/1000 (x)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35				
23	10 000	964	263	106	53	31	20	14	11	8	7	5	4	4	3	3	2	2	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0			
24	10 000	1162	356	141	66	36	23	15	11	9	7	6	5	4	4	3	3	2	2	2	2	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0			
25	10 000	1070	309	115	51	26	16	10	7	6	5	4	3	3	3	2	2	2	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0			
26	10 000	1142	327	122	55	28	16	10	7	6	5	4	4	3	3	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
27	10 000	1191	346	136	64	33	19	12	9	7	6	5	4	4	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
28	10 000	1319	422	176	85	45	26	17	12	10	8	7	6	5	4	4	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
29	10 000	1103	329	122	63	34	20	13	9	7	6	5	5	4	4	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
30	10 000	1099	327	131	62	33	19	13	9	7	6	5	5	4	4	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
31	10 000	1136	341	140	68	37	22	14	10	8	7	6	5	5	4	4	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2		
32	10 000	1188	359	149	75	41	25	16	12	10	8	7	6	6	5	5	4	4	4	4	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2		
33	10 000	1254	373	152	74	41	25	17	13	11	9	8	7	7	6	6	5	5	4	4	4	4	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2		
34	10 000	1004	289	113	53	30	19	14	11	9	8	7	6	6	5	5	4	4	4	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2		
35	10 000	851	239	92	45	26	17	13	10	9	8	7	6	6	5	5	4	4	4	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2		
36	10 000	879	236	91	44	26	17	14	12	10	9	8	7	6	5	5	4	4	4	4	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2		
37	10 000	948	252	101	53	32	23	17	14	13	11	10	9	8	7	7	6	6	5	5	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3	3		
38	10 000	972	269	111	59	38	26	21	17	15	14	12	11	10	9	8	7	7	6	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3		
39	10 000	989	281	116	62	40	29	23	20	17	15	14	12	11	10	9	8	7	6	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3		
40	10 000	921	248	99	53	35	26	21	18	16	14	12	11	10	9	8	7	7	6	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3		
41	10 000	995	265	103	56	37	27	22	19	17	15	13	12	11	10	9	8	7	6	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3	3		
42	10 000	1036	288	117	62	41	31	25	21	19	17	15	13	12	11	10	9	8	7	7	6	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3	3	3		
43	10 000	1037	285	116	62	41	31	25	22	19	17	16	14	13	12	11	11	10	9	8	8	7	6	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3		
44	10 000	933	251	103	56	37	28	23	20	18	16	15	14	13	12	11	11	10	9	9	8	7	7	6	5	5	4	4	4	4	3	3	3	3	3	3	3	3		
45	10 000	1029	279	116	64	43	33	28	25	23	21	19	18	17	16	15	14	13	13	12	11	10	10	9	8	7	6	6	5	5	4	4	4	4	3	3	3	3	3	
46	10 000	1083	289	122	70	48	38	32	29	26	24	23	21	20	19	18	17	16	15	15	14	13	12	11	10	9	8	7	7	6	6	5	5	4	4	3	3	3	3	
47	10 000	1186	323	139	81	57	45	39	35	32	30	28	26	24	23	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	7	6	6	5	5	4	4	3	3	
48	10 000	1143	308	130	75	53	43	37	32	29	27	25	24	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	6	5	5	4	4	3	3	3	3	
49	10 000	1144	330	139	78	55	44	37	33	30	27	25	24	22	21	20	18	17	17	16	15	14	13	12	12	11	10	9	8	7	7	6	6	5	5	4	4	3	3	
50	10 000	1078	309	131	76	53	42	35	31	28	26	24	22	21	19	18	17	16	15	14	13	12	11	10	9	8	7	6	6	5	5	4	4	3	3	3	3	3	3	
51	10 000	1277	385	170	99	70	54	46	41	37	34	32	29	27	25	23	21	20	19	18	17	16	15	14	13	12	10	9	8	7	7	6	6	5	5	4	4	3	3	
52	10 000	1245	373	167	97	69	54	46	41	38	35	32	30	27	25	23	21	20	18	17	16	14	13	12	11	10	9	8	7	6	6	5	5	4	4	3	3	3	3	
53	10 000	1291	383	173	104	73	57	49	44	40	37	35	32	29	27	25	23	21	19	18	17	15	14	12	11	10	9	8	7	6	6	5	5	4	4	3	3	3	3	
54	10 000	1300	404	189	111	78	62	53	48	44	41	39	36	34	31	28	26	23	22	20	19	17	16	14	13	11	10	9	8	7	7	6	6	5	5	4	4	3	3	
55	10 000	1316	416	199	120	87	70	60	54	50	47	44	42	39	36	33	30	27	25	23	22	20	18	16	15	13	12	10	9	8	8	7	6	6	5	5	4	4	3	
56	10 000	1246	393	188	116	85	70	61	55	51	48	46	43	39	36	33	30	27	25	24	22	20	18	17	15	13	12	11	9	8	7	6	6	5	5	4	4	3	3	
57	10 000	1356	453	220	136	100	82	72	66	62	57	53	49	45	41	37	34	31	28	26	24	22	20	18	17	15	13	12	10	9	8	7	6	6	5	5	4	4	3	3
58	10 000	1397	463	224	140	103	84	74	68	62	56	51	46	41	37	34	30	28	25	23	21	19	18	16	15	13	12	11	9	8	7	6	6	5	5	4	4	3	3	
59	10 000	1560	531	262	165	122	98	84	74	66	59	52	46	41	36	32	28	25	23	20	19	17	16	15	13	12	11	10	9	8	7	6	6	5	5	4	4	3	3	
60	10 000	1717	615	299	182	129	101	84	72	64	56	49	42	37	32	28	25	22	20	18	16	15	14	13	12	11	10	9	8	7	7	6	6	5	5	4	4	3	3	
61	10 000	1480	576	276	164	113	85	69	59	51	44	38	33	29	25	22	19	17	15	13	12	11	10	10	9	8	8	7	7	6	6	5	5	4	4	3	3	3		
62	10 000	1603	616	243	143	97	72	56	47	40	34	30	26	22	19	17	15	13	11	10	9	8	7	7	6	6	5	5	5	4	4	4	3	3	3	3	3	3	3	

## Table d'expérience construite sur la population : Femme/ Non-Cadre/ Maladie/Convention Collective

Age à la survenance / Mois d'ancienneté	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		
23	10000	940	298	144	83	51	33	23	16	12	10	9	8	7	7	6	5	5	4	4	4	3	3	2	2	2	2	2	1	1	1	1	1	1	1	1	1	
24	10000	1306	472	219	117	68	42	28	21	16	14	12	11	10	9	8	8	7	6	6	5	5	4	4	3	3	3	3	2	2	2	2	2	1	1	1	1	
25	10000	1191	382	158	79	43	26	17	13	11	9	8	7	7	6	6	5	5	4	4	4	3	3	3	2	2	2	2	2	2	1	1	1	1	1	1	1	
26	10000	1189	328	127	63	34	21	14	11	9	8	7	6	6	5	5	4	4	4	3	3	3	3	2	2	2	2	2	2	2	1	1	1	1	1	1	1	
27	10000	1105	311	121	59	32	20	14	11	9	8	7	7	6	6	5	5	4	4	4	3	3	3	3	2	2	2	2	2	2	2	2	2	1	1	1	1	
28	10000	1419	450	185	89	49	30	21	17	14	13	12	11	10	9	8	8	7	7	6	6	6	5	5	5	5	4	4	4	4	3	3	3	3	3	2	2	
29	10000	1361	426	175	86	48	29	20	16	14	12	11	11	10	9	8	8	7	7	6	6	6	5	5	5	5	4	4	4	4	4	4	3	3	3	3	2	
30	10000	1255	409	177	92	51	31	21	16	14	13	12	11	10	9	9	8	7	7	6	6	6	5	5	5	5	5	4	4	4	4	4	3	3	3	3	2	
31	10000	1233	401	187	101	57	35	23	18	15	13	12	11	11	10	9	9	8	8	7	7	6	6	6	5	5	5	5	4	4	4	4	4	4	3	3	3	3
32	10000	1105	359	175	99	59	37	25	20	17	15	14	13	11	11	10	9	9	8	8	7	7	6	6	6	5	5	5	5	5	4	4	4	4	4	3	3	3
33	10000	1309	440	198	110	69	46	33	27	23	21	19	17	16	14	13	12	11	11	10	10	9	9	8	8	7	7	7	6	6	6	5	5	5	5	4	4	4
34	10000	1004	294	123	65	41	30	23	20	17	16	14	13	12	11	10	9	8	8	7	7	6	6	6	5	5	5	5	4	4	4	4	4	3	3	3	3	3
35	10000	785	220	92	50	33	24	20	17	15	14	13	12	11	10	9	8	7	7	6	6	6	5	5	5	5	4	4	4	4	3	3	3	3	3	3	3	3
36	10000	872	257	117	71	48	36	29	26	23	21	20	18	16	15	13	12	11	11	10	9	9	8	8	7	7	6	6	6	5	5	5	5	5	4	4	4	4
37	10000	1003	277	129	79	56	42	34	30	27	25	22	21	19	17	15	14	13	12	11	11	10	10	9	9	8	8	7	7	6	6	6	5	5	5	5	4	4
38	10000	983	266	126	79	57	44	36	32	28	26	23	21	19	18	16	15	14	13	12	11	11	10	9	9	8	8	7	7	6	6	6	5	5	5	5	4	4
39	10000	910	273	134	84	62	49	42	36	32	29	26	24	22	20	19	17	16	15	14	13	12	11	10	9	9	8	8	7	7	6	6	6	5	5	5	5	4
40	10000	810	233	105	64	48	39	33	29	26	23	21	19	18	16	15	14	13	12	12	11	10	10	9	8	8	7	7	6	6	6	5	5	5	5	4	4	4
41	10000	772	205	91	56	42	34	29	26	23	21	19	17	16	15	14	13	12	11	11	10	9	9	8	7	7	6	6	6	5	5	5	5	4	4	4	4	3
42	10000	989	325	159	101	76	61	52	45	41	37	34	31	28	26	25	23	22	21	19	18	17	16	14	13	12	11	11	10	9	8	8	7	6	6	5	5	4
43	10000	960	290	140	90	68	55	47	42	38	35	32	30	28	26	25	23	22	21	20	18	17	16	14	13	12	11	10	10	9	8	8	7	6	6	5	5	4
44	10000	904	242	111	71	54	44	39	36	33	30	28	26	25	23	22	21	20	19	18	17	16	14	13	12	11	10	9	9	8	8	7	6	6	5	5	4	3
45	10000	975	268	128	84	64	54	48	44	41	38	36	34	32	30	29	27	26	25	24	22	21	19	17	15	14	13	12	12	11	10	9	8	7	6	6	5	4
46	10000	920	255	122	82	64	54	48	45	42	40	38	36	34	32	30	28	27	26	25	23	22	20	18	17	16	14	14	13	12	11	10	9	8	7	6	6	4
47	10000	1039	289	135	92	73	62	56	52	49	47	45	43	41	38	36	34	32	31	29	28	26	24	23	21	20	18	17	16	15	14	13	11	10	9	8	7	5
48	10000	959	243	106	69	55	47	42	39	36	35	33	32	31	29	27	25	24	23	22	21	20	19	18	17	16	14	13	12	12	11	10	9	8	7	6	6	4
49	10000	907	244	112	71	54	46	41	37	34	33	31	30	29	27	25	24	22	22	21	20	19	18	17	16	15	13	12	11	10	10	9	8	7	6	5	4	3
50	10000	885	242	110	70	53	44	39	36	33	31	30	28	27	25	24	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	3	3
51	10000	1093	329	155	101	78	66	58	53	50	46	44	42	39	37	35	33	31	29	27	26	24	22	21	19	17	16	14	12	11	10	9	8	7	6	5	4	4
52	10000	1162	341	166	111	88	75	67	61	57	53	50	47	45	42	39	36	34	32	30	27	25	24	22	20	18	16	14	12	11	10	9	8	6	5	4	4	3
53	10000	1051	297	152	106	85	72	65	59	55	52	49	46	43	41	38	35	33	30	28	26	24	22	20	18	16	14	13	11	10	9	8	7	6	5	4	3	3
54	10000	1117	327	161	108	85	73	65	60	56	53	51	48	46	43	40	37	34	32	29	27	25	23	21	19	16	14	13	11	10	9	8	7	6	5	4	3	3
55	10000	1279	416	213	145	118	103	92	85	80	76	73	70	66	62	57	53	48	45	41	39	36	33	30	26	23	20	18	16	14	12	11	9	8	7	6	5	4
56	10000	1026	336	178	125	102	91	83	77	72	67	63	60	57	53	49	45	42	39	36	33	31	29	26	23	20	17	15	13	12	10	9	8	7	6	5	4	4
57	10000	1373	484	262	184	152	136	126	117	107	98	91	85	79	73	67	62	58	53	49	46	42	39	35	32	28	25	22	19	17	15	13	11	9	8	6	6	
58	10000	1199	411	225	161	133	119	109	100	89	80	72	65	59	54	49	46	42	39	36	33	31	29	26	24	21	19	17	15	13	11	10	9	8	7	6	6	6
59	10000	1342	485	270	198	163	142	126	112	99	86	74	64	57	51	46	42	39	36	33	31	29	27	25	23	21	19	17	15	13	12	11	9	8	7	6	6	
60	10000	1567	611	338	237	187	157	135	117	102	87	74	63	55	49	44	40	37	34	32	30	28	26	24	22	21	19	18	16	15	13	12	10	9	8	8	8	
61	10000	1426	515	272	178	134	109	92	79	68	59	50	43	37	33	29	27	25	23	22	20	19	17	16	15	14	13	12	11	10	9	8	7	6	6	6	6	
62	10000	1262	411	206	131	95	74	62	52	45	39	34	30	26	22	20	18	17	16	15	14	13	12	11	10	9	9	8	8	7	6	6	6	6	6	5	5	