1. **Introduction**

Une police d'assurance est une promesse par l'assureur au preneur d'assurance pour payer les réclamations futures pour une prime reçue à l'avance. En conséquence, les assureurs ne connaissent pas le coût initial de leur service, mais ils s'appuient sur l'analyse et le jugement des données historiques pour prévoir un prix durable pour leur offre. En général, la plupart des polices sont exécutées pour une période de 12 mois. Cependant, le processus de paiement des réclamations peut prendre des années, voire des décennies.

Ainsi, il ne devrait pas être une surprise que le poste le plus important dans le passif du bilan d'un assureur est souvent la provision ou les réserves pour les paiements de sinistres futurs. Ces provisions peuvent être décomposées en pertes déclarées à la compagnie d'assurance et des pertes encourues mais non déclarées.

Dans ce projet, à l’étude exploratoire de données de règlement des sinistres couverts sous une police d’assurance auto et on va essayer d’étendre cette étude en exploitant les données historiques pour en déduire des informations qui pourraient servir la compagnie d’assurance dans la mise en place de sa stratégie.

1. **Description des données**

Ci-après un extrait du tableau de données à étudier du fichier « TriangleGlobalAutoRC V3bis » :

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **2004** | **2005** | **2006** | **2007** | **2008** |
| **2008** | 166 930 | 171 000 | 155 943 | 134 107 | 84 556 |
|  | 151 397 | 136 695 | 230 559 | 230 899 | 203 088 |
|  | 688 905 | 824 350 | 608 156 | 715 712 | 580 882 |
| **2009** | 166 790 | 187 783 | 188 427 | 161 657 | 140 280 |
|  | 84 620 | 166 452 | 251 583 | 262 130 | 199 241 |
|  | 705 713 | 792 316 | 645 181 | 706 423 | 676 987 |
| **2010** | 175 669 | 206 996 | 208 167 | 189 066 | 200 898 |
|  | 114 924 | 172 517 | 273 300 | 262 722 | 213 312 |
|  | 665 472 | 826 246 | 631 119 | 742 216 | 718 491 |

Les données consistent en fait en un tableau qui prend :

-en lignes les périodes pendant lesquelles les règlements de sinistres étaient réglés

-en colonnes les périodes d’occurrence des sinistres

Les montants de règlements ont été comptabilisés par 4 mois.

L’intervalle de développement de paiement de sinistres s’étale de 2008 (début d’année) à 2015 (fin d’année) 🡪 pour une période de 8 ans.

Les années d’occurrence des sinistres s’étale de 2004 à 2015 🡪 une période de 12 ans.

L’ordre de grandeur des montants de règlement des sinistres n’a pas été communiqué, on va supposer que les montants sont représentés en €.

On remarque que quelques données sont manquantes. Ces données peuvent être catégorisées en deux types :

-des données manquantes qui peuvent êtres dues à un manque de méthodes de comptabilisation ou de modifications de la façon dont les données ont été enregistrées dans la base de données. (Ce sont généralement les données les plus historiques)

-des données manquantes dues à une limitation dans l’horizon de développent des règlements de sinistres (Si on se situe à 2015 –comme c’est bien le cas dans notre exemple- les données seront normalement tronquées à partir de cette année là).

1. **Préparation des données**

Les données historiques d'assurance sont souvent présentées sous la forme d'une structure triangulaire, montrant l'évolution des sinistres dans le temps pour chaque période d'exposition (origine du sinistre).

Afin de pouvoir visualiser les données sous une forme triangulaire, des manipulations doivent être faites sur le set de données pour :

-Inverser les colonnes et les lignes puisque dans le set de données d’origine les colonnes sont les dates d’origine et les lignes sont les périodes de développement.

-Renommer les lignes et les colonnes (les colonnes doivent représenter des périodes et non des années)

-Regrouper les montants de développement des règlements des sinistres par années (les périodes de développement étaient chaque 4 mois)

-Supprimer les lignes de développement pour lesquelles les données historiques sont manquantes (données en haut à gauche de notre triangle)

-Représenter les données de règlement des sinistres en cumulant les montants d’une période de développement à une autre

On trouve ainsi le résultat, en triangle, suivant :

1 2 3 4 5 6 7 8

2008 868526.3 1885034 3017734 4091070 5173490 6196024 7170345 8171773

2009 694191.4 1620526 2620527 3638420 4667380 5659283 6592866 NA

2010 752782.5 1613640 2521089 3455853 4410063 5277889 NA NA

2011 906791.4 1903626 2990364 4078264 5131938 NA NA NA

2012 838580.9 1783640 2745767 3696008 NA NA NA NA

2013 732526.8 1479867 2216140 NA NA NA NA NA

2014 761331.0 1616497 NA NA NA NA NA NA

2015 938917.7 NA NA NA NA NA NA NA

Ce triangle montre les valeurs connues de la perte de chaque année d'origine et des évaluations annuelles par la suite. Par exemple, les valeurs connues de perte provenant de la période d'exposition de 2014 sont respectivement de 761’331 1’616’497 à la fin des années 2014 et 2015.

La dernière diagonale, c'est-à-dire le vecteur [8’171’773, . . ., 938’917.7] en haut à droite en bas à gauche elle affiche l'évaluation la plus récente disponible.

Les en-têtes des colonnes [1, 2, . . ., 8] correspondent aux âges (en années) des observations dans la colonne par rapport au début de la période d'exposition.

Une autre représentation du triangle pourrait se faire par incrémentation des règlements en comparant ces derniers d’une année à l’année qui la précède. Le triangle ci-dessous est le résultat de cette méthode :

1 2 3 4 5 6 7 8

2008 868526.3 1016507.6 1132700.6 1073336.0 1082419.4 1022534.3 974320.3 1001428

2009 694191.4 926335.0 1000000.6 1017892.7 1028960.3 991903.0 933582.5 NA

2010 752782.5 860858.0 907448.8 934763.5 954209.7 867826.5 NA NA

2011 906791.4 996835.0 1086738.1 1087899.5 1053674.1 NA NA NA

2012 838580.9 945059.0 962127.0 950240.7 NA NA NA NA

2013 732526.8 747340.1 736272.8 NA NA NA NA NA

2014 761331.0 855166.4 NA NA NA NA NA NA

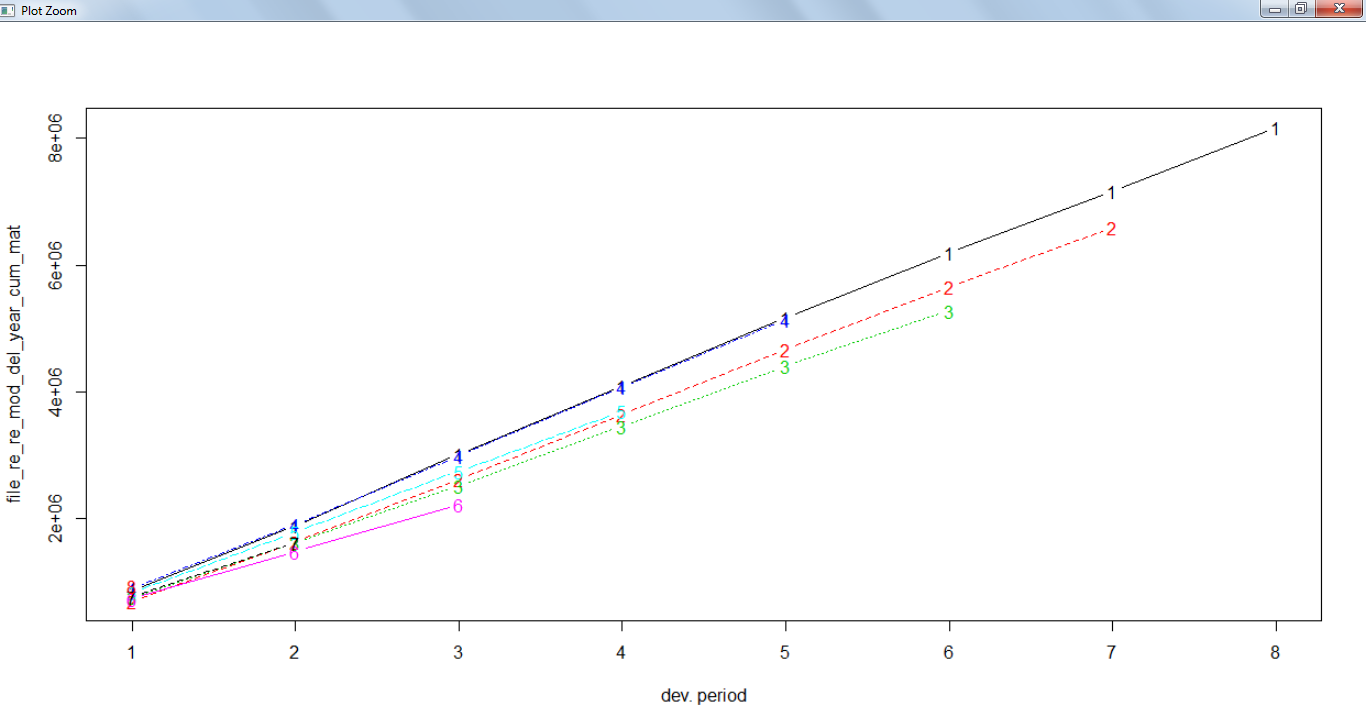
2015 938917.7 NA NA NA NA NA NA NA

Le tableau ci-dessus montre bien que d’une année de développement à l’autre les incréments sont positifs 🡪 Cette vue des données nous donne une idée sur la concentration des règlements des sinistres ainsi que de leur évolution par année de développement.

1. **Visualisation des données**

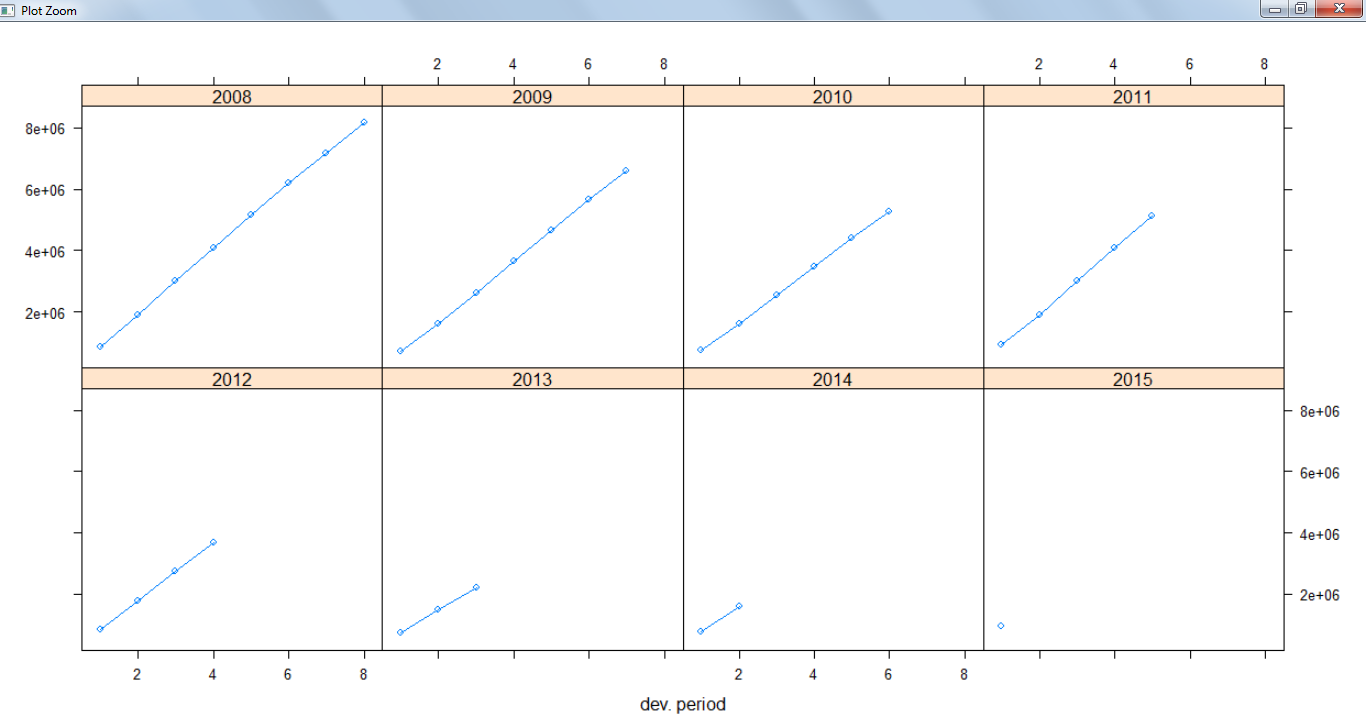
Dans cette partie on va visualiser les développements des règlements des sinistres en superposant les graphes pour chaque ligne (origine des sinistres).

On obtient ainsi le graphe suivant :

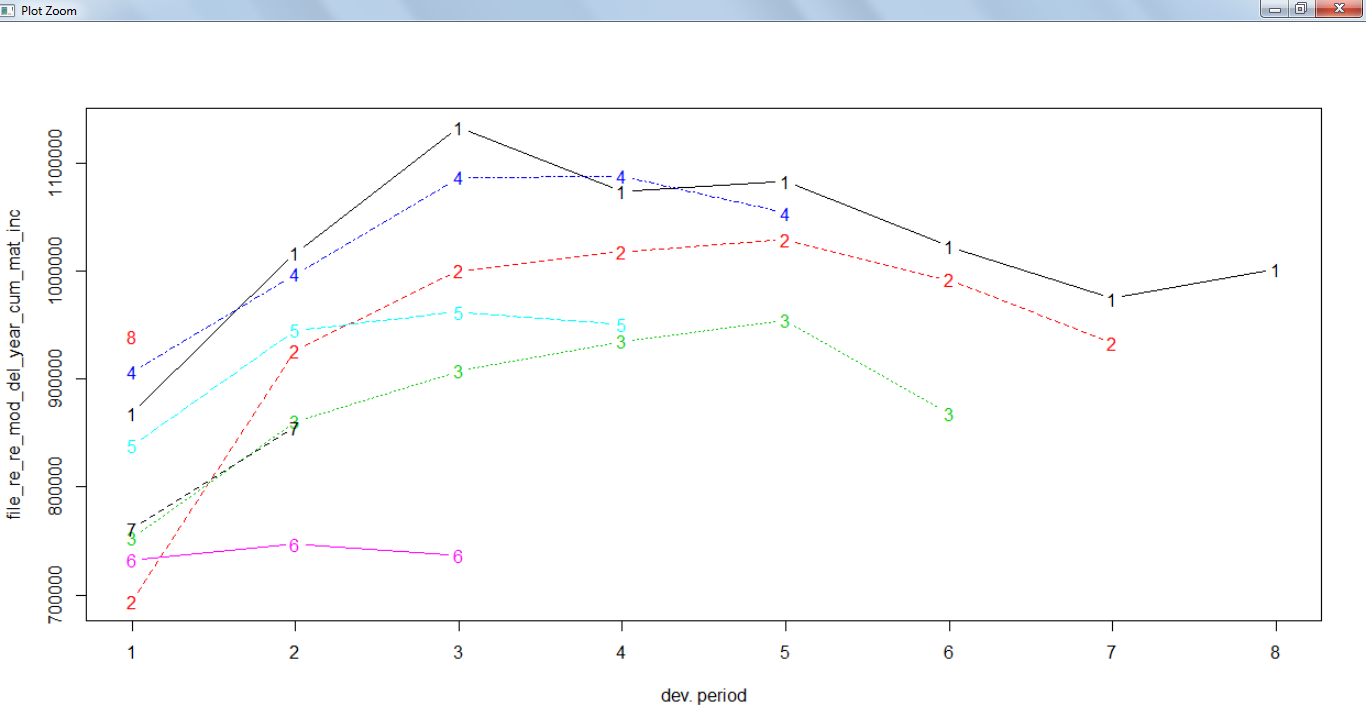


Le graphe ci-dessus ne montre pas une tendance particulière des montants de règlements cumulés. On remarque que les montants de règlements au terme de la première année à partir de chaque année d’origine sont tous du même ordre de grandeur.

On va représenter individuellement les graphes qui ont été superposés dans le graphe précédent :

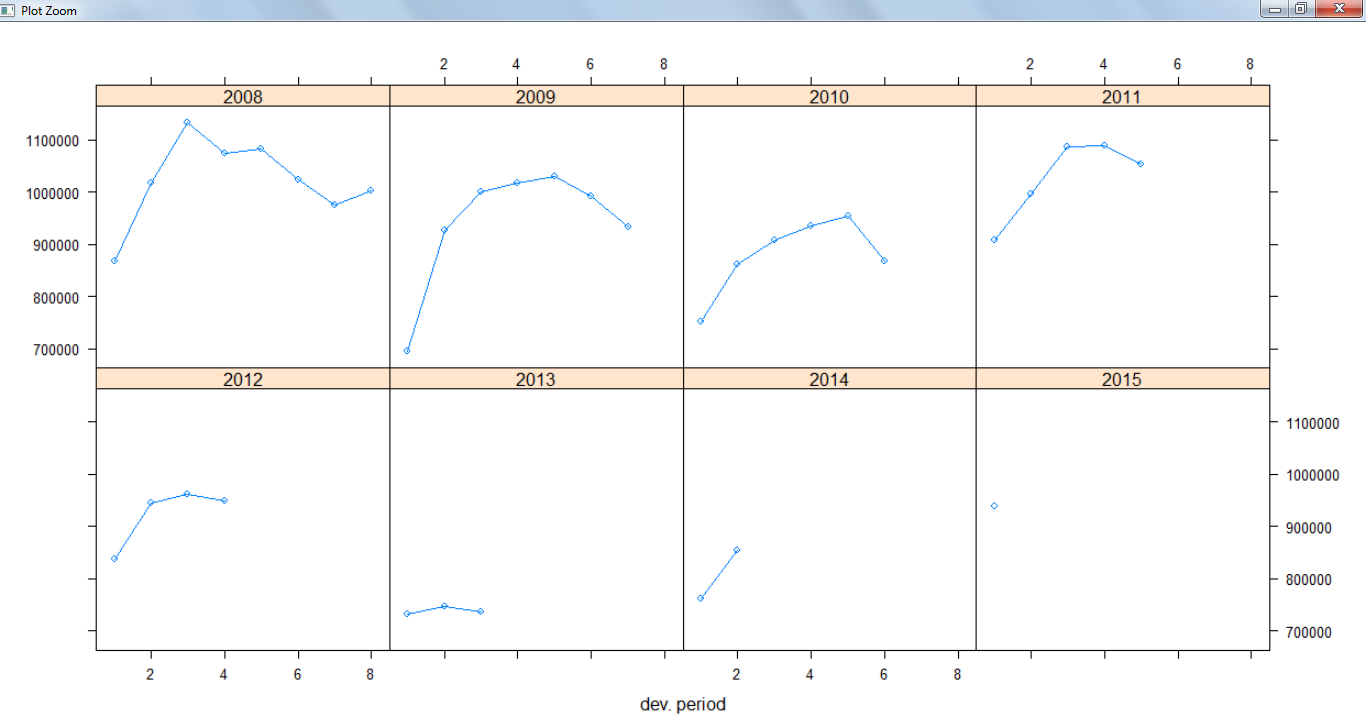


En ce qui concerne la vue des données par incrémentation, en représentant les données on retrouve le graphe suivant :



Ce qu’on peut déduire du graphe précédent on remarque que les charges de règlements de sinistres ont une forme convexe ce qui signifie que les charges annuelles ont tendance à décroître en fonction des années de développement.

Une autre représentation en dissociant les graphes donne une vision plus claire sur l’évolution des charges.



1. **Estimations des provisionnements**

L'objectif de cette partie de l’étude exploratoire des données est de prévoir l'évolution future des sinistres dans le coin inférieur droit du triangle précédemment construit et les développements potentiels au-delà de l'âge de 8 ans.

Pour ce faire, on va utiliser comme première méthode celle de Chain Ladder qui est un algorithme déterministe pour prévoir les règlements de sinistres basées sur des données historiques. Cette méthode suppose que l'évolution proportionnelle des sinistres d'une période de développement à l'autre est la même pour toutes les années d'origines.

Les rapports de liaisons entre les années de développement sont calculés comme les ratios moyens de développement pondérés en fonction du volume d'un triangle de développement cumulatif de pertes d'une période de développement à la suivante Cik, i, k = 1,. . . N :

Les résultats de cette méthode sont les suivants pour les rapports de liaisons entre les années de développements :

2.142828 1.566313 1.364445 1.269875 1.202251 1.160932 1.139663

Ces rapports donnent ainsi un résultat de prévision des charges futures en se basant sur le triangle des données historiques comme suit :

1 2 3 4 5 6 7 8

2008 868526.3 1885034 3017734 4091070 5173490 6196024 7170345 8171773

2009 694191.4 1620526 2620527 3638420 4667380 5659283 6592866 7513642

2010 752782.5 1613640 2521089 3455853 4410063 5277889 6127272 6983023

2011 906791.4 1903626 2990364 4078264 5131938 6169877 7162810 8163186

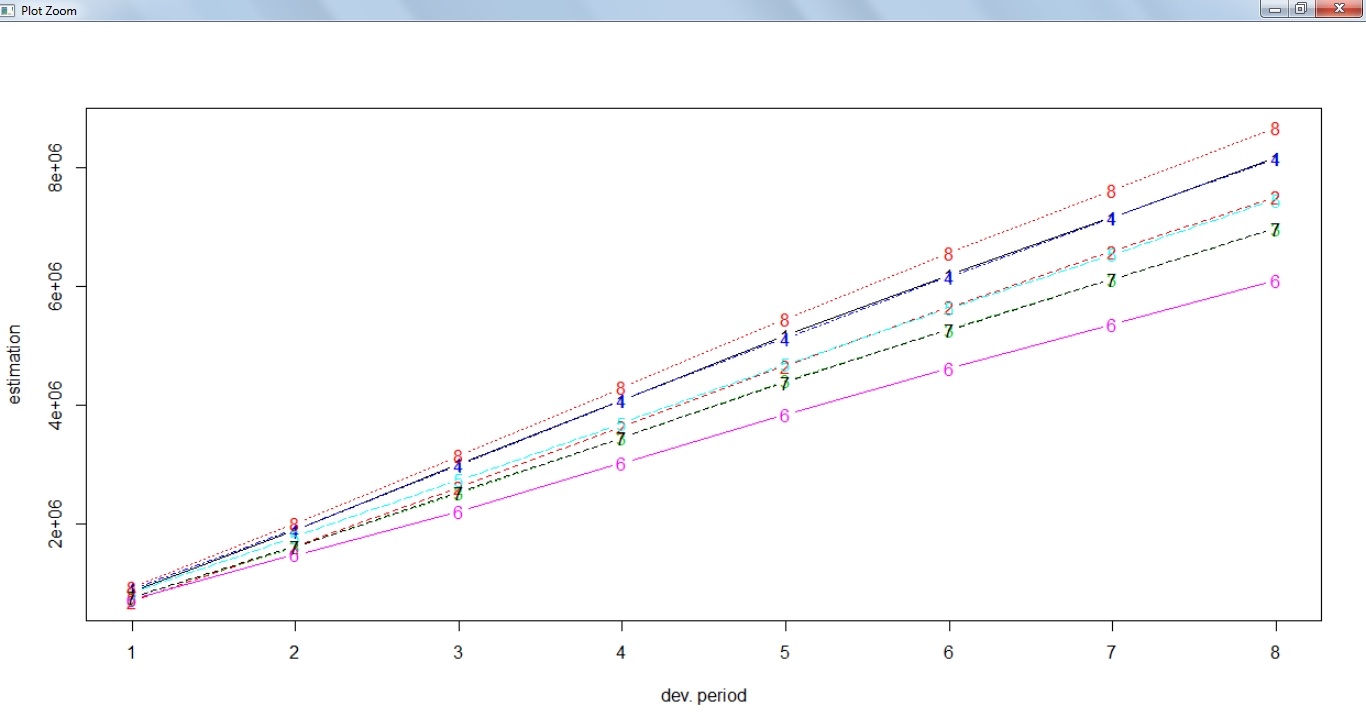
2012 838580.9 1783640 2745767 3696008 4693467 5642725 6550822 7465726

2013 732526.8 1479867 2216140 3023800 3839847 4616460 5359397 6107904

2014 761331.0 1616497 2531941 3454694 4387029 5274309 6123116 6978286

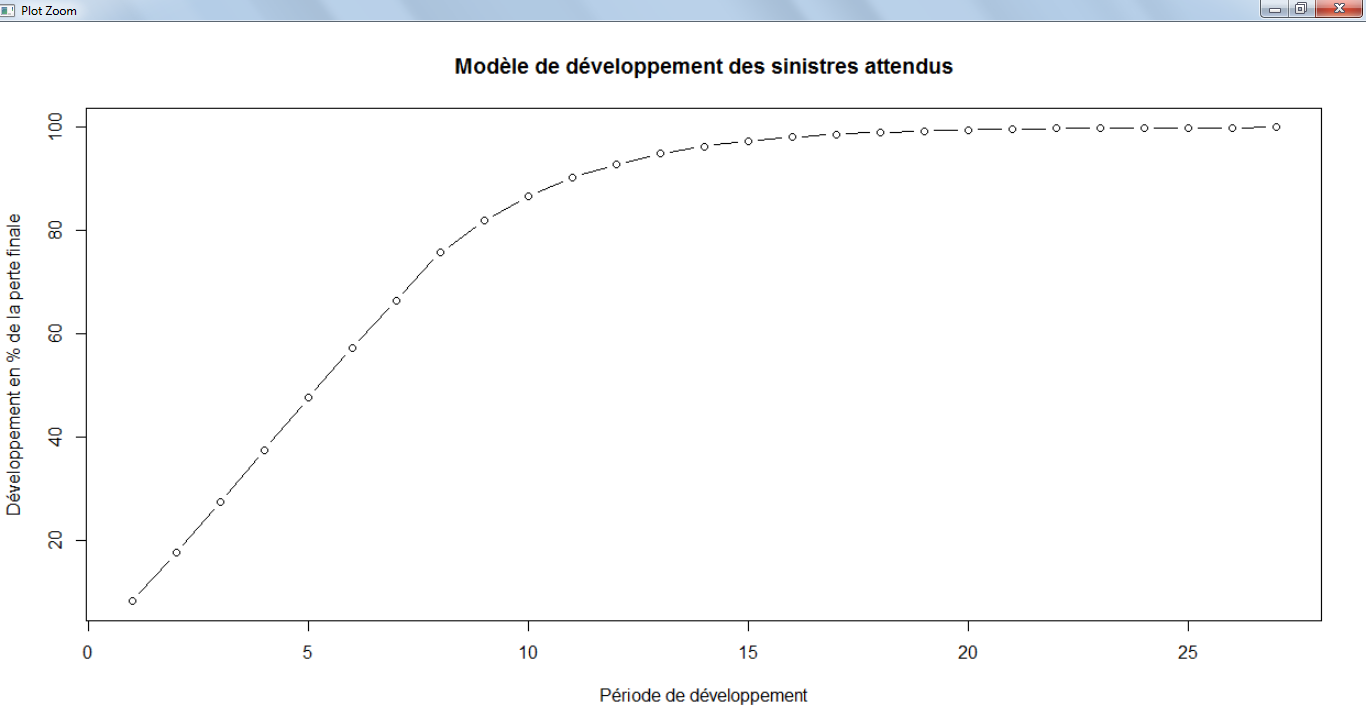
2015 938917.7 2011939 3151327 4299811 5460222 6564556 7621006 8685375

Ces données historiques et données de prévision donnent les représentations suivantes développées toutes au même âge de 8 ans :



Souvent, il n'est pas approprié de supposer que l'année d'origine la plus ancienne est pleinement développée. Une approche typique consiste à extrapoler les rapports de développement en supposant un modèle log-linéaire.

Les resultants d’une telle method d’extrapollation nous indique le modèle de développementdes sinistres attendus et ce sur la base des facteurs d’évolution des charges déduits par la méthode de Chain Ladder, ci-après le graphe correspondant :



Les facteurs d’évolution sont alors appliqués au dernier montant connu des sinistres cumulés pour prévoir la prochaine période de développement. On obtient ainsi le triangle de développement suivant :

1 2 3 4 5 6 7 8 final

2008 868526 1885034 3017734 4091070 5173490 6196024 7170345 8171773 10797572

2009 694191 1620526 2620527 3638420 4667380 5659283 6592866 7513642 9927967

2010 752782 1613640 2521089 3455853 4410063 5277889 6127272 6983023 9226846

2011 906791 1903626 2990364 4078264 5131938 6169877 7162810 8163186 10786226

2012 838581 1783640 2745767 3696008 4693467 5642725 6550822 7465726 9864655

2013 732527 1479867 2216140 3023800 3839847 4616460 5359397 6107904 8070530

2014 761331 1616497 2531941 3454694 4387029 5274309 6123116 6978286 9220587

2015 938918 2011939 3151327 4299811 5460222 6564556 7621006 8685375 11476207

Le total des charges estimées non encore payées est de : 45’728'562 €