



المندوبية السامية للتخطيط
HAUT-COMMISSARIAT AU PLAN

ROYAUME DU MAROC
..*.*
HAUT COMMISSARIAT AU PLAN
..*.*.*.*

INSTITUT NATIONAL
DE STATISTIQUE ET D'ECONOMIE APPLIQUEE



Projet de Fin d'Etudes

Élaboration d'un modèle actuariel de l'assurance maladie obligatoire du secteur privé au Maroc

Préparé par : *M. Mohammed TIOUR*

Sous la direction de : *M. Abderrahim OULIDI (INSEA)*
Mlle Nabila BOUZEGGAOUI (ANAM)

Soutenu publiquement comme exigence partielle en vue de l'obtention du

Diplôme d'Ingénieur d'Etat

Option : ACTUARIAT-FINANCE

Devant le jury composé de :

- *M. Jelloul EL MABROUK (INSEA)*
- *M. Abderrahim OULIDI (INSEA)*
- *Mlle Nabila BOUZEGGAOUI (ANAM)*

Résumé

L'assurance maladie obligatoire est l'un des principaux piliers de la sécurité sociale. La maîtrise des différents soubassements du secteur est donc d'une importance primordiale, particulièrement que ce secteur est en pleine croissance au Maroc et ne date que de l'année 2005.

L'objet de cette étude est d'élaborer un modèle actuariel de l'assurance maladie obligatoire au Maroc, c'est-à-dire modéliser les ressources et les dépenses du secteur pour pouvoir juger de sa viabilité. Cette étude se focalisera sur la population des salariés du secteur privé, dont la couverture est assurée par la Caisse Nationale de la Sécurité Sociale.

Les dépenses et les ressources d'un régime d'assurance sont étroitement liées à la taille et à la structure de sa population. C'est pour cela, que la première partie de ce travail fait l'objet de la modélisation de la population assurée par la Caisse Nationale de la Sécurité Sociale. Le modèle, choisi est un modèle empirique basé sur le bilan des entrées et sorties que connaît la population étudiée. Cette étape sera suivie de la modélisation des remboursements dont bénéficie cette population. Pour ce faire, nous allons confronter les différents modèles économétriques suivants :

- La régression linéaire multiple
- Les modèles linéaires généralisés
- Les modèles linéaires à densité mélange

Ensuite, nous allons étudier les assiettes de la population, qui servent à calculer les ressources du régime moyennant les taux de cotisations, pour finalement, consolider les résultats des trois volets (population, remboursements et contributions) pour dresser le compte des produits et charges de l'assurance maladie obligatoire du secteur privé.

Mots clés : assurance maladie obligatoire, régression linéaire multiple, modèles linéaires généralisés, modèles linéaires à densité mélange, modèle actuariel, projections démographiques.

Abstract

The compulsory health insurance is one of the main pillars of social security. Mastering different bases of the sector is therefore of paramount importance, especially as this sector is still growing in Morocco and dates only from 2005.

The purpose of this study is to develop an actuarial model of compulsory health insurance in Morocco, that is to say, modeling resources and expenditures of the sector in order to assess its viability. This study focuses on the population of private sector employees, whose coverage is provided by the Caisse Nationale de la Sécurité Sociale.

The expenditures and the resources of the insurance system are closely linked to the size and structure of its population. For this, the aim of the first part of this work is to model the population covered by the Caisse Nationale de la Sécurité Sociale. The model chosen is an empirical model based on the balance of the entries and exits of the studied population. This step will be followed by modeling of the refunds that this population benefits from. To do this, we will compare the following different econometric models:

- Multiple Linear Regression
- Generalized Linear Models
- Finite Linear Mixture Models

Then we will study the contribution basis of the population used to calculate the resources of the system, using the contribution rates, to finally consolidate the results of the three components (population, refunds and contributions) to produce a profit and loss account of the private sector's compulsory health insurance system.

Keywords: compulsory health insurance, Multiple Linear Regression, Generalized Linear Models, Finite Linear Mixture Models, actuarial model, Population projection.

Dédicace

Je dédie ce travail à mes parents qui m'ont soutenu tout au long de mon parcours scolaire et à tous ceux qui ont contribué à la réalisation de ce travail.

Mohammed TIOUR

Remerciement

Je tiens tout d'abord à exprimer mes sincères remerciements à mon encadrant interne M. Abderrahim OULIDI, pour l'aide et les conseils pertinents qu'il m'a apportés tout au long de mon projet de fin d'études.

Mes sincères remerciements s'adressent également à mon encadrante Mme Nabila BOUZEGGAOUI, chef du service Monitoring des Paramètres Financiers, pour ses efforts, son écoute et son encouragement qui ont fait de mon stage une expérience très enrichissante.

J'ai également bénéficié pendant ces derniers mois de précieux conseils de Monsieur M. Abdeljabbar BOUGHALEB, chef de Division Décisionnel et Analyse de données. Ses sensibilités professionnelles lui ont permis de m'apprendre les intérêts pratiques dans mes recherches en assurance santé.

Je remercie tout particulièrement et témoigne ma reconnaissance à tout le personnel du département Études Économiques et Actuarielles pour leur soutien et leur disponibilité durant toute la période de mon stage.

Enfin, je remercie le corps professoral de l'INSEA, pour ses efforts continus afin de nous offrir une meilleure formation, ainsi que toute autre personne ayant participé de différentes façons à la réussite de mon projet de fin d'études.

Table des matières

RÉSUMÉ	- 3 -
ABSTRACT	- 4 -
DEDICACE	- 5 -
REMERCIEMENT	- 6 -
TABLE DES MATIERES	- 7 -
LISTE DES ABREVIATIONS	- 9 -
LISTE DES TABLEAUX ET FIGURES	- 10 -
INTRODUCTION	- 12 -
CHAPITRE 1 : CONTEXTE GENERAL	- 13 -
I. L'AMO AU MAROC	- 13 -
I.1. <i>Population couverte</i>	- 13 -
I.2. <i>Organismes gestionnaires</i>	- 14 -
II. PRESENTATION DE L'ANAM	- 14 -
III. L'AMO DU SECTEUR PRIVE	- 15 -
III.3. <i>La CNSS</i>	- 15 -
III.4. <i>Financement</i>	- 15 -
III.5. <i>Ouverture des droits</i>	- 16 -
III.6. <i>ALD</i>	- 16 -
III.7. <i>Historique de l'AMO CNSS</i> :	- 16 -
CHAPITRE 2 : POPULATION	- 18 -
I. DESCRIPTION DE LA POPULATION ASSUREE PAR LA CNSS	- 18 -
I.1. <i>Structure des données</i> :	- 18 -
I.2. <i>Population selon inputs</i>	- 19 -
I.2.1. Effectif par âge :	- 19 -
I.2.2. Type du bénéficiaire :	- 20 -
I.2.3. Type d'assuré :	- 21 -
I.2.4. Selon le sexe du bénéficiaire :	- 21 -
I.2.5. Ouverture de droit :	- 22 -
I.2.6. Etat (ALD vs Non-ALD) :	- 22 -
I.2.7. Tranche de revenu :	- 23 -
II. MODELISATION DE LA POPULATION :	- 23 -
II.3. <i>Paramètres et notations</i>	- 24 -
II.4. <i>Mise en équation des flux démographiques</i>	- 26 -
II.4.1. Assuré actifs	- 27 -
II.4.2. Retraités	- 27 -
II.4.3. Conjoints	- 27 -
II.4.4. Les veufs(ves)	- 29 -
II.4.5. Les enfants	- 29 -
II.4.6. Les orphelins	- 30 -
II.4.7. Estimation des paramètres	- 30 -
II.4.8. Test de sensibilité	- 31 -
CHAPITRE 3 : CONSOMMATION	- 32 -
I. ANALYSE DESCRIPTIF :	- 32 -
I.1. <i>Structure des données</i> :	- 32 -
I.2. <i>Exploration des données</i>	- 33 -
I.2.1. Grandeurs globales :	- 33 -
I.2.2. Analyse descriptive :	- 33 -
1) Consommation médicale par âge :	- 33 -
2) Consommation médicale selon le type du bénéficiaire	- 35 -
3) Consommation médicale par sexe :	- 36 -
4) Consommation médicale selon la variable ALD	- 37 -

5)	Consommation médicale par secteur d'établissement de soins :	- 38 -
I.3.	<i>Dépendances entre inputs</i> :	- 39 -
I.3.1.	Variables catégorielles deux-à-deux	- 39 -
1)	Dépendance entre la variable ALD et le sexe	- 39 -
2)	Décomposition de la statistique de chi-deux. (TYPE_BEN, SEXE)	- 40 -
I.3.2.	Dépendance avec la variable TR_AGE :	- 42 -
II.	MODELE DE LA SINISTRALITE	- 44 -
II.4.	<i>Sinistralité globale</i>	- 44 -
II.5.	<i>Montant moyen remboursé (Sévérité)</i>	- 46 -
II.5.1.	Rappel sur la régression linéaire multiple	- 46 -
1)	Principe :	- 46 -
2)	Écriture matricielle :	- 47 -
3)	Pondération	- 48 -
4)	Variables qualitatives	- 48 -
II.5.2.	Application	- 49 -
II.5.3.	Rappel sur les GLM :	- 55 -
1)	La famille exponentielle	- 55 -
2)	La fonction de lien canonique	- 56 -
3)	Estimation	- 56 -
II.5.4.	Application	- 57 -
1)	Remboursement sur la consommation ambulatoire (médicaments)	- 57 -
2)	Remboursements sur la consommation hospitalière	- 59 -
II.5.5.	Finite Mixture Models	- 63 -
II.5.6.	Application	- 64 -
II.6.	<i>Effectif sinistré (Fréquence)</i>	- 64 -
II.6.1.	Application de la régression de Poisson	- 65 -
II.6.2.	Application de la régression binomiale négative	- 67 -
II.6.3.	Application du modèle ZINB	- 68 -
III.	CONCLUSION	- 69 -
CHAPITRE 4 : ASSIETTES DE COTISATION ET PROJECTIONS		- 70 -
I.	ASSIETTE DE COTISATION	- 70 -
I.1.	<i>Etude descriptive des assiettes de cotisation</i>	- 70 -
I.1.1.	Description de la base de donnée	- 70 -
I.1.2.	Assiette selon le type d'assuré	- 71 -
I.1.3.	Assiette selon la tranche d'âge	- 72 -
I.1.4.	Assiette selon état de santé	- 73 -
I.1.5.	Assiette selon le sexe	- 74 -
I.1.6.	Assiette selon la tranche de revenu	- 75 -
I.2.	<i>Modèle de l'assiette de cotisation</i>	- 75 -
I.2.1.	Choix du modèle	- 75 -
I.2.2.	Interprétation	- 81 -
II.	PROJECTIONS	- 85 -
II.3.	<i>Résultats</i>	- 85 -
II.4.	<i>Conclusion</i>	- 87 -
CONCLUSION GENERALE		- 88 -
BIBLIOGRAPHIE		- 90 -
ANNEXES		- 91 -

Liste des abréviations

ANAM : Agence Nationale de l'Assurance Maladie

AMO : Assurance Maladie Obligatoire

CNOPS : Caisse Nationale des Organismes de Prévoyance Sociale

CNSS : Caisse Nationale de Sécurité Sociale

ALD : Affection de Longue Durée

Liste des tableaux et figures

LISTE DES TABLEAUX

TABLEAU 1 : STRUCTURE DE LA BASE DE DONNEES DE LA POPULATION AMO-CNSS DE 2012.....	- 19 -
TABLEAU 2 : TEST D'INDEPENDENCE ENTRE LES VARIABLES DE CONSOMMATION CATEGORIEL	- 39 -
TABLEAU 3 : EFFECTIF SINISTRE DE LA POPULATION AMO-CNSS SELON LE SEXE ET L'ETAT DE SANTE DU BENEFICIAIRE.....	- 40 -
TABLEAU 4 : EFFECTIF SINISTRE DE LA POPULATION AMO-CNSS SELON LE SEXE ET LE TYPE DU BENEFICIAIRE	- 40 -
TABLEAU 5 : ESTIMATION DES PARAMETRES DU MODELE DE LA REGRESSION DE POISSON APPLIQUE AUX EFFECTIFS SINISTRES	- 65 -
TABLEAU 6 : ESTIMATION DES PARAMETRES DU MODELE DE LA REGRESSION BINOMIALE NEGATIVE APPLIQUE AUX EFFECTIFS SINISTRES	- 67 -
TABLEAU 7 : TEST D'ANOVA DE L'AJUSTEMENT DE L'ASSIETTE MOYENNE AU MODELE SIMPLE DE REGRESSION LINEAIRE.....	- 75 -
TABLEAU 8 : TEST D'ANOVA DE L'AJUSTEMENT DE L'ASSIETTE MOYENNE AU MODELE LOGARITHMIQUE DE REGRESSION LINEAIRE DE L'ASSIETTE MOYENNE DE COTISATION	- 77 -
TABLEAU 9 : ESTIMATION DES PARAMETRES DU MODELE LOGARITHMIQUE DE REGRESSION LINEAIRE DE L'ASSIETTE DE COTISATION MOYENNE	- 79 -
TABLEAU 10 : ESTIMATION DES PARAMETRES DU MODELE LOGARITHMIQUE DE REGRESSION LINEAIRE DE L'ASSIETTE DE COTISATION MOYENNE APRES ELIMINATION DES MODALITES NON-SIGNIFICATIVES.....	- 80 -
TABLEAU 11 : TABLEAU 10 : ESTIMATION DES PARAMETRES DU MODELE LOGARITHMIQUE DE REGRESSION LINEAIRE DE L'ASSIETTE DE COTISATION MOYENNE APRES ELIMINATION DES MODALITES NON-SIGNIFICATIVES ET INTRODUCTION DES VARIABLES D'INTERACTION.....	- 81 -
TABLEAU 12 : EXPONENTIELS DES PARAMETRES DU MODELE LOGARITHMIQUE DE REGRESSION LINEAIRE DE L'ASSIETTE DE COTISATION MOYENNE	- 84 -
TABLEAU 13 : COMPTE TECHNIQUE DES PRODUITS ET CHARGE DE L'AMO DU SECTEUR PRIVE SUR LA PERIODE 2013-2017 EN DH.....	- 85 -
TABLEAU 14 : RESULTAT TECHNIQUE DU CPC DE L'AMO DU SECTEUR PRIVE SUR LA PERIODE 2012-2017 EN MILLIARDS DE DH	- 86 -
TABLEAU 15 : RESULTAT EXCEDENT DU BILAN DE L'AMO DU SECTEUR PRIVE SUR LA PERIODE 2007-2011 EN MILLIARDS DE DH	- 86 -

LISTE DES FIGURES

FIGURE 1 : EFFECTIFS DE LA POPULATION AMO-CNSS DE 2012 SELON L'AGE	- 19 -
FIGURE 2 : PYRAMIDE DES AGES DE LA POPULATION AMO-CNSS DE 2012.....	- 20 -
FIGURE 3 : EFFECTIFS DE LA POPULATION AMO-CNSS 2012 SELON LE TYPE DE BENEFICIAIRE	- 20 -
FIGURE 4 : EFFECTIFS DE LA POPULATION AMO-CNSS SELON LE TYPE D'ASSURE	- 21 -
FIGURE 5 : REPARTITION DE LA POPULATION AMO-CNSS 2012 SELON LE SEXE.....	- 21 -
FIGURE 6 : REPARTITION DE LA POPULATION AMO-CNSS SELON LE STATUT DU DROIT	- 22 -
FIGURE 7 : REPARTITION DE LA POPULATION AMO-CNSS SELON L'ATTEINTE D'AFFECTION DE LONGUE DUREE.....	- 22 -
FIGURE 8 : EFFECTIFS DE LA POPULATION AMO-CNSS 2012 SELON LA TRANCHE DE REVENU.....	- 23 -
FIGURE 9 : EFFECTIFS GLOBAUX PROJETES DE LA POPULATION AMO-CNSS.....	- 30 -
FIGURE 10 : CONSOMMATION MEDICALE MOYENNE D'UN BENEFICIAIRE DE L'AMO-CNSS PAR TRANCHE D'AGE.....	- 34 -
FIGURE 11 : EFFECTIF SINISTRE SELON L'AGE DES BENEFICIAIRES.....	- 34 -
FIGURE 12 : TAUX DE SINISTRALITE DE LA POPULATION AMO-CNSS SELON L'AGE	- 35 -
FIGURE 13 : CONSOMMATION MEDICALE MOYENNE D'UN BENEFICIAIRE DE L'AMO-CNSS SELON SON TYPE.....	- 35 -
FIGURE 14 : CONSOMMATION MEDICALE MOYENNE D'UN BENEFICIAIRE DE L'AMO-CNSS SELON LE SEXE	- 36 -
FIGURE 15 : TAUX DE SINISTRALITE DE LA POPULATION AMO-CNSS SELON LE SEXE.....	- 36 -
FIGURE 16 : CONSOMMATION MEDICALE MOYENNE D'UN BENEFICIAIRE DE L'AMO-CNSS SELON SON ETAT DE SANTE.....	- 37 -
FIGURE 17 : TAUX DE SINISTRALITE DE LA POPULATION AMO-CNSS SELON L'ETAT DE SANTE DES BENEFICIAIRES	- 37 -
FIGURE 18 : REPARTITION DES SOINS HOSPITALIERS ADMINISTRES A LA POPULATION AMO-CNSS SELON LE SECTEUR DE L'ETABLISSEMENT DE SOINS.....	- 38 -
FIGURE 19 : TAUX DE REMBOURSEMENT SELON LE SECTEUR DE L'ETABLISSEMENT DE SOINS.....	- 38 -
FIGURE 20 : SCHEMATISATION DE LA PROCEDURE DE DECOMPOSITION DE LA STATISTIQUE DE CHI-DEUX	- 42 -
FIGURE 21 : TAUX DE SINISTRALITE DES BENEFICIAIRES ALD SELON L'AGE	- 43 -
FIGURE 22 : SORTIE DE LA REGRESSION LINEAIRE PONDEREE DU MODELE SIMPLE DES REMBOURSEMENTS	- 50 -
FIGURE 23 : SORTIE DE LA REGRESSION LINEAIRE PONDEREE DU MODELE LOGARITHMIQUE DES REMBOURSEMENTS.....	- 51 -
FIGURE 24 : QQ-PLOT DES RESIDUS DU MODELE LOGARITHMIQUE ET DE LA LOI NORMALE	- 51 -
FIGURE 25 : RESIDUS DU MODELE SIMPLE DE REGRESSION LINEAIRE PONDEREE.....	- 52 -
FIGURE 26 : RESIDUS DU MODELE SIMPLE DE REGRESSION LINEAIRE PONDEREE SELON LE SECTEUR D'ETABLISSEMENT	- 53 -
FIGURE 27 : RESIDUS DU MODELE SIMPLE DE REGRESSION LINEAIRE PONDEREE SELON L'ETAT DE SANTE DU BENEFICIAIRE	- 53 -
FIGURE 28 : RESIDUS DU MODELE SIMPLE DE REGRESSION LINEAIRE PONDEREE SELON LE TYPE DE SOINS	- 54 -
FIGURE 29 : QQ-PLOT DE L'AJUSTEMENT DES REMBOURSEMENTS DES SOINS AMBULATOIRES RELATIFS AUX BENEFICIAIRES NON-ALD A LA LOI GAMMA	- 58 -
FIGURE 30 : QQ-PLOT DE L'AJUSTEMENT DES REMBOURSEMENTS DES SOINS AMBULATOIRES RELATIFS AUX BENEFICIAIRES NON-ALD A LA LOI LOG-NORMALE	- 58 -
FIGURE 31 : QQ-PLOT DE L'AJUSTEMENT DU REMBOURSEMENT DES SOINS HOSPITALIERS RELATIFS AUX BENEFICIAIRES NON-ALD A LA LOI GAMMA	- 59 -
FIGURE 32 : QQ-PLOT DE L'AJUSTEMENT DU REMBOURSEMENT DES SOINS HOSPITALIERS RELATIFS AUX BENEFICIAIRES NON-ALD A LA LOI LOG-NORMALE	- 60 -
FIGURE 33 : HISTOGRAMME DES REMBOURSEMENTS DES SOINS HOSPITALIERS RELATIFS AUX BENEFICIAIRES NON-ALD.....	- 60 -
FIGURE 34 : HISTOGRAMME DES REMBOURSEMENTS DES SOINS HOSPITALIERS RELATIFS AUX BENEFICIAIRES ALD.....	- 61 -
FIGURE 35 : HISTOGRAMME AJUSTE DES REMBOURSEMENTS DES SOINS HOSPITALIERS RELATIFS AUX BENEFICIAIRES NON-ALD	- 62 -
FIGURE 36 : HISTOGRAMME AJUSTE DES REMBOURSEMENTS DES SOINS HOSPITALIERS RELATIFS AUX BENEFICIAIRES ALD.....	- 62 -
FIGURE 37 : CONTRIBUTION A L'ASSIETTE GLOBALE DE CHAQUE TYPE D'ASSURE	- 71 -
FIGURE 38 : ASSIETTE DE COTISATION MOYENNE DE LA POPULATION AMO-CNSS SELON LE TYPE D'ASSURE	- 71 -
FIGURE 39 : ASSIETTE DE COTISATION MOYENNE DE LA POPULATION AMO-CNSS SELON LA TRANCHE D'AGE	- 72 -

FIGURE 40 : ASSIETTE DE COTISATION MOYENNE DE LA POPULATION AMO-CNSS SELON L'ETAT DE SANTE DU BENEFICIAIRE	- 73 -
FIGURE 41 : ASSIETTE DE COTISATION MOYENNE DE LA POPULATION AMO-CNSS SELON LE SEXE	- 74 -
FIGURE 42 : CONTRIBUTION A L'ASSIETTE GLOBALE DE LA POPULATION AMO-CNSS SELON LE SEXE.....	- 74 -
FIGURE 43 : QQ-PLOT DE L'AJUSTEMENT DES RESIDUS DU MODELE SIMPLE DE REGRESSION LINEAIRE DE L'ASSIETTE MOYENNE DE COTISATION A LA LOI NORMALE	- 76 -
FIGURE 44 : HISTOGRAMME DES RESIDUS DU MODELE SIMPLE DE REGRESSION LINEAIRE DE L'ASSIETTE MOYENNE DE COTISATION AJUSTES A LA LOI NORMALE	- 76 -
FIGURE 45 : DIAGRAMME DES RESIDUS DU MODELE SIMPLE DE REGRESSION LINEAIRE DE L'ASSIETTE MOYENNE DE COTISATION	- 77 -
FIGURE 46 : QQ-PLOT DE L'AJUSTEMENT DES RESIDUS DU MODELE LOGARITHMIQUE DE REGRESSION LINEAIRE DE L'ASSIETTE DE COTISATION MOYENNE A LA LOI NORMALE	- 78 -
FIGURE 47 : HISTOGRAMME DES RESIDUS DU MODELE LOGARITHMIQUE DE REGRESSION LINEAIRE DE L'ASSIETTE DE COTISATION MOYENNE AJUSTES A LA LOI NORMALE	- 78 -

INTRODUCTION

L'assurance maladie obligatoire est l'une des composantes principales de la sécurité sociale servant à la couverture des risques et frais de soins de santé. Instaurer un tel système est considéré comme étant un devoir de l'État selon la Déclaration universelle des droits de l'homme de 1948.

Au Maroc, la consommation médicale prend une place de plus en plus importante dans les dépenses des ménages. En effet, ces dépenses représentaient 4.0% du PIB en 1999, 5.3% en 2002 et 6.4 % en 2012. C'est pour cela qu'assurer la pérennité du système d'assurance maladie s'inscrit au cœur des préoccupations de l'État. Ainsi, une maîtrise de l'influence des différents facteurs socio-économiques sur le comportement du consommateur marocain face aux soins de santé est d'une importance primordiale dans la gestion du système de l'assurance maladie obligatoire.

C'est dans ce contexte qu'est née l'Agence National de l'Assurance Maladie (ANAM). Qui œuvre pour la viabilité de l'assurance maladie obligatoire par sa qualité de régulateur. L'ANAM a donc pour mission d'encadrer les organismes gestionnaires de l'assurance maladie obligatoire et cela en se basant sur les études économiques et actuarielles qu'elle réalise.

Le présent rapport a pour objectif d'étudier la situation économique actuelle du secteur de l'assurance maladie obligatoire au Maroc et de dresser un aperçu sur sa situation future. Cette étude se focalisera sur les salariés du secteur privé, qui sont assurés par la Caisse Nationale de la Sécurité Sociale, et se répartira en quatre chapitres :

- Un premier chapitre donnant un aperçu sur les différents acteurs et les multiples composants du secteur de l'assurance maladie obligatoire marocain.
- Un deuxième chapitre faisant l'analyse et les projections de la population assurée par la CNSS.
- Un troisième chapitre étudiant les remboursements dont bénéficie la population assurée par la CNSS.
- Un quatrième chapitre traitant des assiettes de cotisations et consolidant les résultats de l'étude.

Chapitre 1 : Contexte général

L'objet de ce chapitre est de donner un aperçu général sur le secteur de l'AMO au Maroc. Pour ce faire, nous allons commencer par introduire quelques généralités sur l'AMO, puis ensuite présenter l'Agence Nationale de l'Assurance maladie pour ensuite détailler quelques soubassements de l'AMO relative au secteur privé.

I. L'AMO au Maroc

Le premier pas du Maroc vers l'instauration du système de l'AMO actuelle date de la publication de loi 65-00, portant code de la couverture de médicale de base, en Novembre 2002. Un extrait du préambule de ce texte est le suivant :

« Afin de concrétiser l'engagement de l'Etat, qui consacre le principe du droit à la santé tel que prévu par les conventions internationales, la présente loi constitue le parachèvement de l'expérience du Maroc en matière de couverture médicale et consolide les droits acquis par les citoyens marocains bénéficiant d'une assurance maladie ¹ »

Les principes suivants sont clairement exprimés par la loi :

- Le fondement du financement de la santé sur les principes de solidarité et d'équité.
- La prise en charge collective et solidaire des dépenses de santé.
- La généralisation à terme de la couverture médicale de base.
- L'atteinte à terme de l'accès universel aux soins.
- L'interdiction de toute discrimination due à l'âge, au sexe, à la nature de l'activité, au niveau et à la nature du revenu ou aux antécédents pathologiques.

Ce n'est qu'en Septembre 2005, après la publication des décrets de mise en application de la loi 65-00 (Août 2005), que le Maroc a connu l'entrée en vigueur de l'AMO.

I.1. Population couverte

Les assurés couverts par l'AMO au Maroc sont constitués des personnes exerçant une activité lucrative, des titulaires de pension, des anciens résistants et membres de l'armée de libération et des étudiants ainsi que de leurs ayants droits (enfants et conjoints). Cette couverture est maintenue :

- Pendant 6 mois en cas de cessation d'activité de l'assuré
- Pendant 12 mois en cas de dissolution du lien de mariage du conjoint
- Pendant 24 mois pour le conjoint survivant et les enfants en cas de décès de l'assuré.

La limite d'âge concernant la couverture des enfants est de 21 ans, prolongeable à 26 ans pour les enfants non mariés et poursuivant leurs études. Cette limite n'est pas imposée aux enfants handicapés.

¹ Dahir n° 1-02-296 du 25 rejev 1423 (3 octobre 2002) portant promulgation de la loi n° 65-00 portant code de la couverture médicale de base. Bulletin officiel n° 5058 du 16 ramadan 1423 (21 novembre 2002)

I.2. Organismes gestionnaires

Les deux principaux organismes gestionnaires de l'AMO au Maroc sont la Caisse National de Sécurité Sociale (CNSS) et la Caisse Nationale des Organismes de Prévoyance Sociale (CNOPS). Ils assurent deux catégories différentes de population. La couverture de la CNOPS concerne les salariés et les titulaires de pensions du secteur public, tandis que celle de la CNSS concerne les salariés et les titulaires de pensions du secteur privé.

Cependant, il existe une tranche des salariés du secteur privé dont l'AMO est assurée par des compagnies d'assurance privées. En effet, à la date de la mise en vigueur de l'AMO au Maroc, un ensemble de sociétés privées souscrivaient à une couverture de base pour leurs salariés au moyen de contrats de groupes chez des compagnies d'assurance privées. Dans le cadre de la loi 65-00, plus précisément l'article 114, un délai de cinq ans renouvelable a été accordé à ces sociétés pour organiser le transfert de la couverture vers la CNSS. Jusqu'à aujourd'hui, cette transition n'est pas encore achevée. Cette catégorie d'assurés sera dite « population 114 ».

II. Présentation de l'ANAM

Outre les organismes gestionnaires déjà cités, l'Agence Nationale de l'Assurance Maladie est l'un des principaux acteurs du secteur de l'AMO au Maroc.

L'ANAM a été créé le 26 mai 2005 en tant qu'établissement public à caractère administratif, doté de la personnalité morale et de l'autonomie financière, et placé sous la tutelle de l'État. La création de l'ANAM vient en vertu de l'article 54 de la loi 65-00, dont un extrait est le suivant :

« Les organismes gestionnaires sont soumis au contrôle technique de l'Etat, qui a pour objet de veiller au respect par ces organismes des dispositions de la présente loi et des textes pris pour son application. ¹ »

Missions de l'ANAM

Les missions de l'ANAM sont clairement énoncées aux articles 59 et 60 de la loi 65-00, et se résument à la veille sur l'équilibre financier du système obligatoire de couverture médicale de base à travers l'encadrement technique et la mise en place d'outils de régulation du système. L'ANAM veille, donc, sur le bon fonctionnement de l'AMO et au respect de la loi 65-00 par les organismes gestionnaires. Un extrait de l'article 59 de la loi 65-00 est le suivant :

« [...] elle est chargée (l'ANAM) de :

- *s'assurer, de concert avec l'administration, de l'adéquation entre le fonctionnement de l'assurance maladie obligatoire de base et les objectifs de l'Etat en matière de santé;*
- *[...]*
- *proposer à l'administration les mesures nécessaires à la régulation du système d'assurance maladie obligatoire de base et, en particulier, les mécanismes appropriés de maîtrise des coûts de l'assurance maladie obligatoire de base et veiller à leur respect;*

¹ Dahir n° 1-02-296 du 25 rejeb 1423 (3 octobre 2002) portant promulgation de la loi n° 65-00 portant code de la couverture médicale de base. Bulletin officiel n° 5058 du 16 ramadan 1423 (21 novembre 2002)

- [...]
- *veiller à l'équilibre global entre les ressources et les dépenses pour chaque régime d'assurance maladie obligatoire de base;*
- *apporter l'appui technique aux organismes gestionnaires pour la mise en place d'un dispositif permanent d'évaluation des soins dispensés aux bénéficiaires de l'assurance maladie obligatoire de base dans les conditions et selon les formes édictées par l'administration;*
- [...]¹»

L'organigramme de l'ANAM figure dans l'annexe IV.

III.L'AMO du secteur privé

Étant donné que notre étude se focalise sur l'assurance maladie obligatoire du secteur privé au Maroc, nous allons présenter quelques détails sur son fonctionnement.

III.3. La CNSS

Comme spécifié plus tard, hormis la population 114, la couverture des salariés et des pensionnés du secteur privé est assurée par la CNSS.

La CNSS est un établissement public doté de la personnalité civile et de l'autonomie administrative et financière par le Dahir n°1-59-148 du 03-12-1959 sous la tutelle Ministère des finance et du Ministère de l'emploi, ce régime a été modifié par le Dahir n°1-79-148 du 27-07-1972. Elle a pour mission de gérer la sécurité sociale relative aux travailleurs du secteur privé contre toute perte éventuelle de leurs moyens de subsistance de façon provisoire ou .

Depuis l'entrée en vigueur de l'assurance maladie obligatoire au Maroc, la CNSS est chargée de la gestion de l'assurance maladie obligatoire pour les salariés assujettis au régime marocain de sécurité sociale qui ne disposaient pas d'une couverture d'assurance maladie facultative au moment de l'entrée en vigueur de la loi et les titulaires d'une pension d'un montant minimum de 500 DH/mois, et cela, concernant les actes hospitaliers. Ce n'est qu'en 2010 que la CNSS a étendu ses remboursements aux soins ambulatoires (médicaments).

III.4. Financement

La participation des salariés au financement de l'AMO est assurée par une contribution patronale et une contribution salariale assise sur les salaires des cotisants. Ce financement se fait à part égal entre l'employeur (2%) et l'affilié (2%). Cependant l'employeur verse en plus une *contribution de solidarité* de l'ordre de 1,5% contre laquelle il bénéficie d'une diminution de un point de la cotisation patronale sur les allocations familiales.

En ce qui concerne les retraités, ils supportent la totalité de la cotisation qui est de l'ordre de 4%, et ceci, pour une pension supérieur ou égale à 500 DH. Les assurés volontaires quant à eux sont soumis à une cotisation au taux de 4 % du montant de leur rémunération mensuelle. Cette contribution de solidarité est aussi supportée par les employeurs des salariés de la population 114.

¹ Dahir n° 1-02-296 du 25 rejev 1423 (3 octobre 2002) portant promulgation de la loi n° 65-00 portant code de la couverture médicale de base. Bulletin officiel n° 5058 du 16 ramadan 1423 (21 novembre 2002)

III.5. Ouverture des droits

Un bénéficiaire de la couverture AMO de la CNSS sinistré n'a le droit au remboursement (ouverture de droit) que si l'assuré qui le prend en charge a passé une période de cotisation, dite *période de stage*, de 54 jours ouvrables successifs ou non pendant les 6 mois précédant sa déclaration.

En cas d'interruption du travail, l'assuré ou les ayants droit (conjoint, enfant à charge de moins de 21 ans ou 26 ans si étudiant et sans limite d'âge si handicapé) bénéficient du maintien de leur droit aux prestations pendant une période maximum de six mois.

En cas de dissolution du mariage, l'ex-conjoint d'un assuré qui ne bénéficie pas d'un régime d'assurance maladie obligatoire de base, continue à bénéficier des prestations de l'AMO pendant un an.

Les ayants droit de l'assuré décédé qui n'ont aucun régime d'assurance maladie obligatoire de base continuent de bénéficier des prestations de l'AMO pendant une période de deux années.

III.6. ALD

Les affections de longue durée (ALD) sont définies comme des maladies chroniques, dont la thérapie est coûteuse et pour laquelle l'Assurance Maladie Obligatoire assure une prise en charge pour tous les traitements nécessaires. En vertu de l'article 9 de la loi 65-00, les frais des soins relatifs à ces maladies sont remboursés par la CNSS à la hauteur de 90% si ces soins sont conduits dans un hôpital public et 70% sinon.

La liste des ALD donnant droit à l'exonération a été fixée par arrêté du ministre de la santé n°2518-05. Elle répertorie plus de 140 maladies dont les critères de choix sont :

- la fréquence de la maladie (prévalence ou incidence)
- la gravité de la maladie surtout en termes d'incapacité et d'invalidité
- la chronicité de la maladie
- la charge de morbidité
- le coût de la prise en charge

III.7. Historique de l'AMO CNSS :

Le système de l'AMO du secteur privé a connu beaucoup de faits marquants depuis sa création. Une liste des événements les plus importants de l'AMO du secteur privé est la suivante :

- Novembre 2002 : Publication de la loi 65-00, portant code de la couverture médicale de base
- Août 2005 : Publication des décrets de la mise en application de la loi 65-00
- Septembre 2005 – Mars 2006 : Entrée en vigueur de l'AMO et lancement de la collecte des cotisations
- Mars 2006 : Remboursement et prise en charge des soins au profit des bénéficiaires
- Mars 2006 – décembre 2011 : Révision de la liste des médicaments
- Juillet 2007 : Extension de la couverture à une nouvelle population de pensionnés suite à la baisse du seuil de pension ouvrant droit à l'AMO
- Juillet 2008 – décembre 2009 : Exonération partielle ou totale du ticket modérateur relatif à 108 maladies.
- Février 2010 : Extension de l'AMO aux soins ambulatoires.
- Septembre 2011 : Extension de la couverture aux veuf(ves) et aux orphelins.

- Avril 2013 : Entrée en vigueur de la mesure relative à la prise en charge des médicaments coûteux en mode tiers payant.
- Mai 2013 : Exonération du ticket modérateur relatif à 53 nouvelles maladies

Chapitre 2 : Population

Ce chapitre est consacré à la population bénéficiaire de l'AMO du secteur privé. Nous allons donc, décrire cette population selon les différentes caractéristiques des bénéficiaires, pour ensuite élaborer un modèle qui servira aux projections démographiques. Ces projections seront exploitées pour projeter les dépenses de la CNSS ainsi que ces ressources.

I. Description de la population assurée par la CNSS

Cette partie fait l'objet d'une analyse descriptive de la population du secteur privé assurée par la CNSS en 2012 au titre l'AMO. Dans ce qui suit cette population sera dite 'population AMO-CNSS'.

Avant d'exposer les particularités de ladite population, nous allons tout d'abord présenter la structure de la base de données mise à notre disposition par l'agence, ensuite nous allons décrire la répartition de l'effectif de cette population selon les inputs d'intérêt.

I.1. Structure des données :

La base de données de la population AMO-CNSS mise à notre disposition comporte les variables décrites dans le tableau ci-dessous :

Nom	Désignation	Modalités
TR_AGE	Tranche d'âge	1 : [0,1[, 2 : [1,5[, 3 : [5,10[, . . . , 15 : [65,70[, 16 : 70 et plus, 17 : non renseigné
TYPE_BEN	Type du bénéficiaire	A : Assuré, CA : Conjoint de l'assuré, EA : Enfant de l'assuré, CM : Conjoint ayant les droits maintenus, EM : Enfant ayant les droits maintenus
TYPE_ASS	Type de l'assuré	A : Actif, I : Inactif, P1 : Retraité et invalide, P2 : conjoint survivant, P3 : Orphelin survivant
SEXE	Sexe du bénéficiaire	M : Masculin, F : Féminin
OUVERT_DROIT	Indique si les droits du bénéficiaire sont ouverts	O : Oui, N : Non
Ville	Ville de résidence du bénéficiaire.	A chaque ville du royaume et même aux grandes villes d'Europe est affectée une modalité.
ALD	Indique si le bénéficiaire est atteint d'une maladie de longue durée	O : Oui, N : Non
SECT_ACTIV	Secteur d'activité de l'assuré	Une lettre est associée à chaque secteur d'activité.
TR_REV	Tanche de revenu de l'assuré	1 : Moins de 1400 dhs 2 : [1400,2100[3 : [2 100,2 800[4 : [2 800,3 500[5 : [3 500,4 250[6 : [4 250,5 000[7 : [5 000,6 000[8 : [6 000,7 000[9 : [7 000,8 250[10 : [8 250,10 000[11 : [10 000,12 000[12 : [12 000,16 000[13 : Plus de 16 000 dhs

Tableau 1 : Structure de la base de données de la population AMO-CNSS de 2012

Pour chaque combinaison de modalités de ces variables (inputs), la base de données fournit l’effectif des bénéficiaires ayant ces caractéristiques. Ces effectifs figurent dans la colonne **EFFECTIF** (output).

Dans ce qui suit, une combinaison de modalités des inputs sera dite ‘classe de population’, chaque classe correspond donc à une ligne de la base de données.

Une colonne REGION, représentant la région de résidence du bénéficiaire, a été ajoutée à la base de données. Ceci a été fait à l’aide du langage SQL en hébergeant les données dans un serveur MySQL.

Notons que la base de données totalise 534 160 lignes, représentant les 4 629 348 bénéficiaires de l’AMO CNSS de à fin 2012. En moyenne chaque classe contient 8.78 bénéficiaire, avec un minimum de 1 bénéficiaire par classe et un maximum de 10122 bénéficiaires. Nous constatons aussi que : 50% des classes contiennent moins de 2 classes, 75% contiennent moins de 5 et 3% des classes contiennent plus de 50% de la population.

I.2. Population selon inputs

Dans cette partie nous allons présenter la répartition de l’effectif de la population AMO CNSS selon les inputs d’intérêt.

I.2.1. Effectif par âge :

La répartition de la population AMO CNSS par âge est représentée par le graphique suivant :

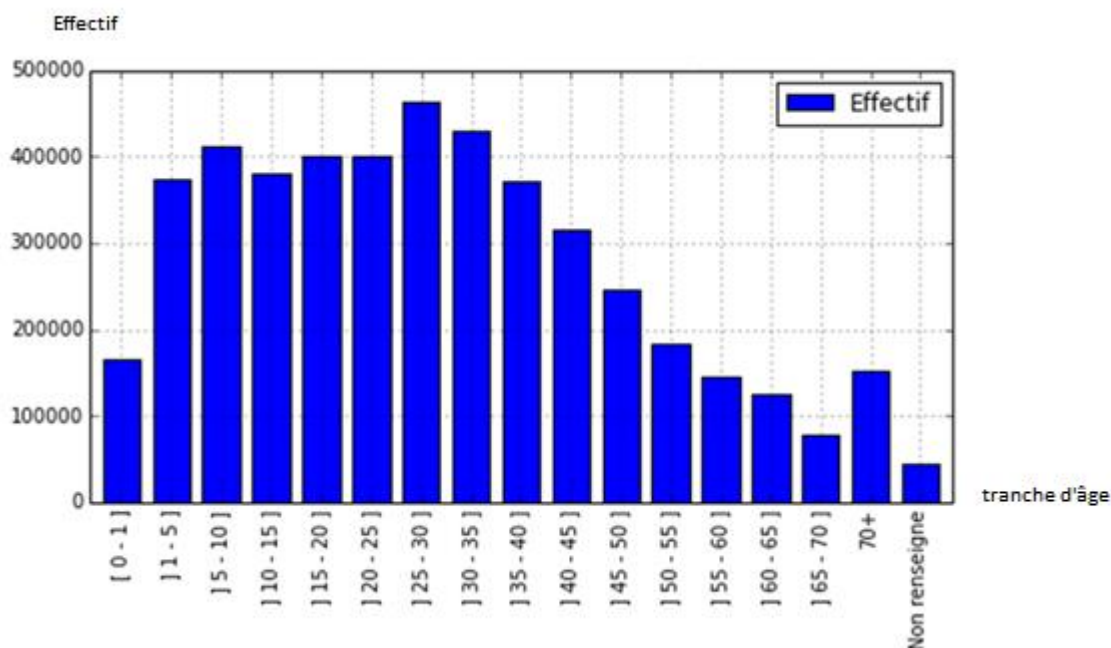


Figure 1 : Effectifs de la population AMO-CNSS de 2012 selon l'âge

Nous constatons que la tranche d’âge qui présente le pic en termes d’effectif est celle de 25–30 ans. Nous remarquons également une tendance baissière à partir de cette tranche, excepté pour la tranche ‘70+’, ce qui peut être dû au fait que l’amplitude de cet intervalle n’est pas bien déterminée. Nous observons aussi que l’effectif de la population présente une certaine stabilité pour les jeunes (5-25 ans).

En faisant distinction entre les sexes, nous retrouvons la pyramide des âges suivante :

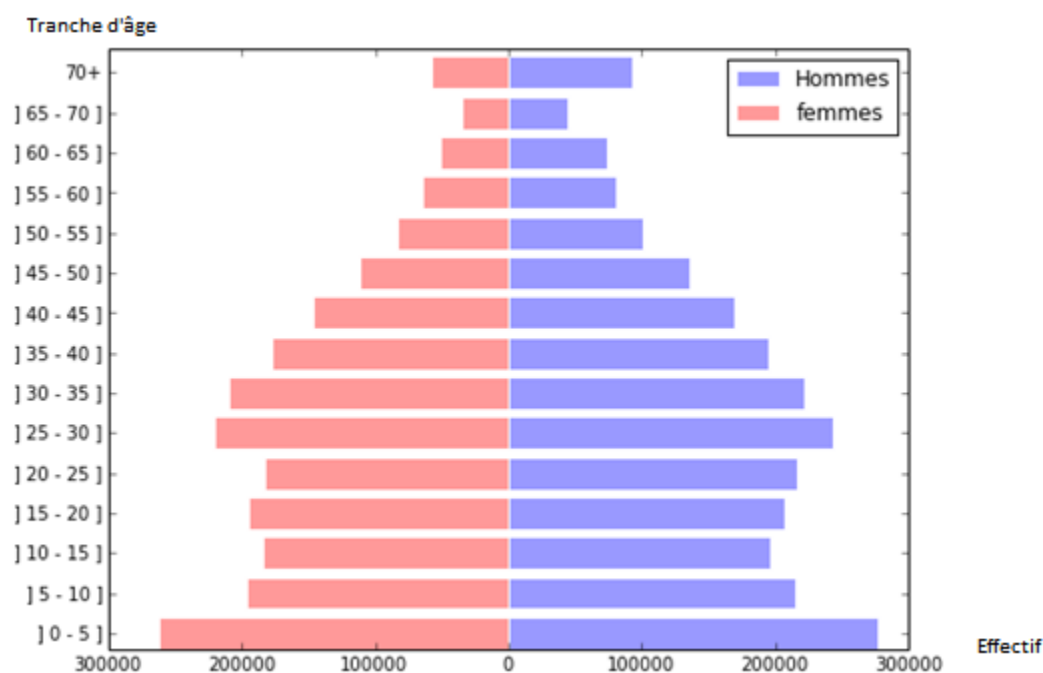


Figure 2 : Pyramide des âges de la population AMO-CNSS de 2012

Nous remarquons que la pyramide des âges est globalement symétrique, et qu'elle a une forme en 'As de pique', traduisant une reprise de natalité et donc un rajeunissement de la population.

1.2.2. Type du bénéficiaire :

Le graphique en 'camembert' suivant représente l'effectif de la population AMO CNSS par type de bénéficiaire :

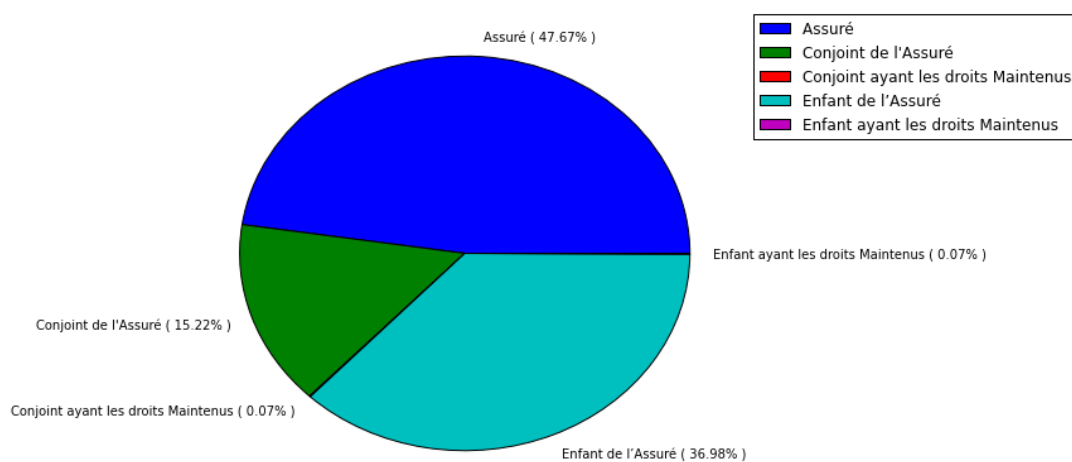


Figure 3 : Effectifs de la population AMO-CNSS 2012 selon le type de bénéficiaire

On constate que près de la moitié de la population AMO CNSS est constituée d'assurés, tandis que les ayant droits maintenus ne représentent qu'une minorité avec 0.14% de la population globale

I.2.3. Type d'assuré :

Le graphique suivant retrace la répartition de l'effectif des bénéficiaires de l'AMO-CNSS selon le type de l'assuré :

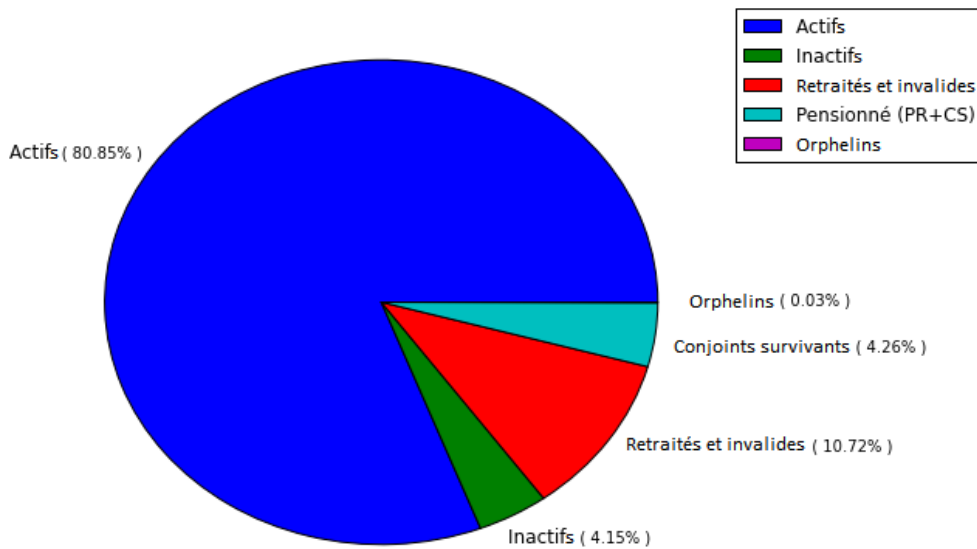


Figure 4 : Effectifs de la population AMO-CNSS selon le type d'assuré

Nous notons que la majorité des bénéficiaires de l'AMO-CNSS (80.85%) sont des assurés actifs et leurs ayants droit. Les pensionnés (retraités et invalides) et leurs ayants droit représentent près de 11% de la population globale, tandis que l'effectif des orphelins est négligeable.

I.2.4. Selon le sexe du bénéficiaire :

La figure suivante montre que la population AMO CNSS est presque uniformément répartie selon le sexe du bénéficiaire, avec un sex-ratio de 1.11 homme par femme :

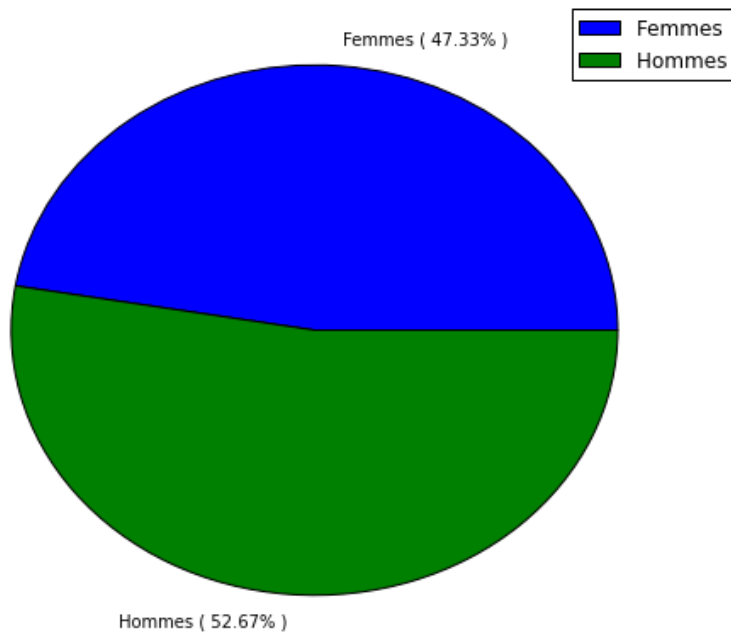


Figure 5 : Répartition de la population AMO-CNSS 2012 selon le sexe

Pour tester si la différence entre l'effectif des deux sexes est significative on recourt au test de Chi-deux. Ce test indique à quelle mesure la distribution de l'effectif selon les sexes est différente d'une distribution uniforme (50%-50%). La valeur de la statistique du test est $\chi^2 = 0.28595$, elle suit une loi de Chi-deux à un degré de liberté. On retrouve alors une p-value de 0.59282 ce qui traduit que la différence d'effectif entre les hommes et les femmes n'est donc pas significative.

1.2.5. Ouverture de droit :

En ce qui concerne la variable *OUVERT_DROIT*, la répartition de la population AMO-CNSS est représentée par le graphique suivant :

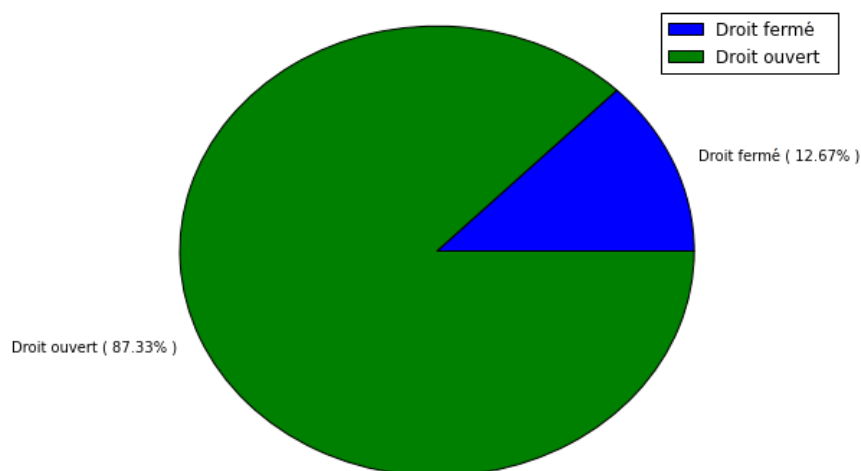


Figure 6 : Répartition de la population AMO-CNSS selon le statut du droit

Nous constatons que la population AMO CNSS est majoritairement ouverte de droit, cependant la proportion des bénéficiaires ayant des droits fermés n'est pas négligeable.

1.2.6. Etat (ALD vs Non-ALD) :

Il importe de noter que le caractère ALD ou non-ALD des bénéficiaires est d'une importance primordiale, car il impacte significativement leur consommation médicale.

Le graphique suivant dresse la répartition de la population AMO CNSS selon la variable ALD :

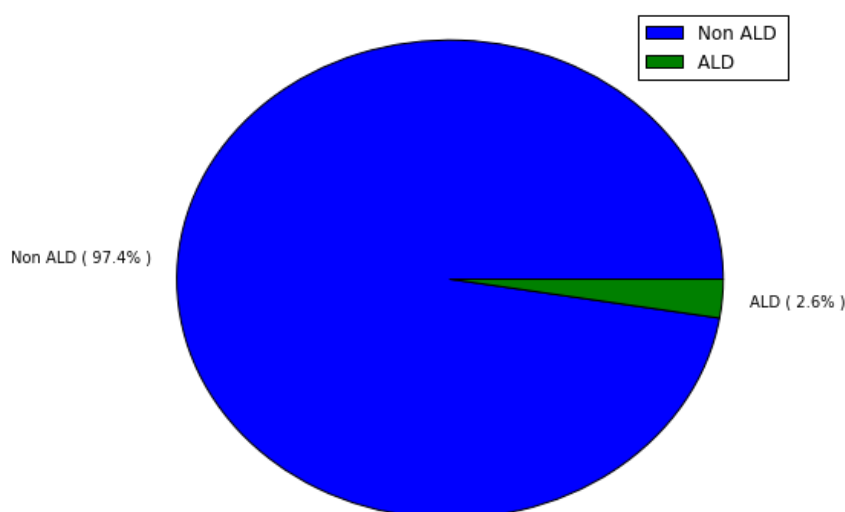


Figure 7 : Répartition de la population AMO-CNSS selon l'atteinte d'une affection de longue durée

Nous remarquons que la population atteinte d'ALD ne représente que 2.6% de la population AMO-CNSS. Toutefois, nous allons voir, dans le chapitre correspondant à la consommation, qu'elle s'accapare plus de 50% des montants remboursés.

1.2.7. Tranche de revenu :

La répartition de la population AMO CNSS selon la tranche de revenu de l'assuré est présentée par le graphique suivant :

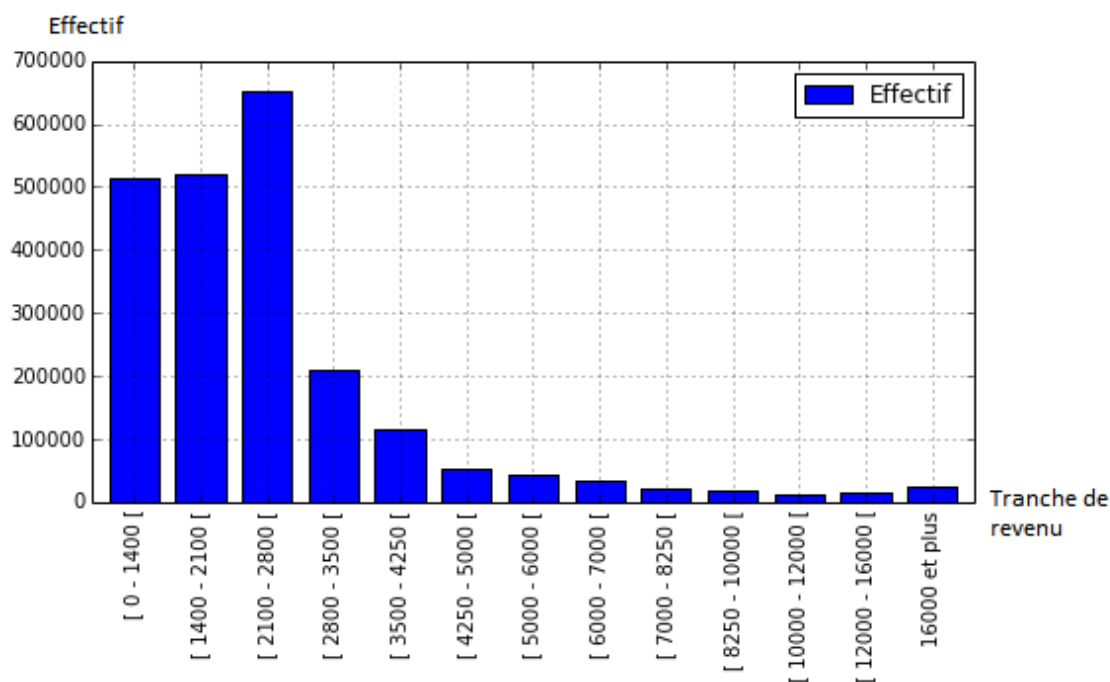


Figure 8 : Effectifs de la population AMO-CNSS 2012 selon la tranche de revenu

Nous constatons que les effectifs des trois premières tranches de salaires sont les plus importants. En effet, ils représentent près de 75 % de la population globale.

Dans ce qui suit nous allons décrire le modèle démographique qu'on adoptera pour projeter la population AMO CNSS avant de tester sa sensibilité.

II. Modélisation de la population :

La population AMO CNSS connaît différents flux démographiques. L'objet de cette partie est de décrire et de mettre en œuvre un modèle démographique pour modéliser ces différents flux.

Vu que les données mises à notre disposition sont des données groupées (sous forme de cube), et peuvent connaître des changements dans le futur, nous proposons d'adopter un modèle empirique. En effet, l'avantage d'un tel modèle est la flexibilité de ses hypothèses et donc son adaptabilité face aux différentes formes de données.

Avant d'entamer la mise en équation, nous allons définir les paramètres et décrire le système de notations que nous allons adopter.

II.3. Paramètres et notations

On distingue entre les différents types de bénéficiaires suivants :

- Actifs (A,A)
- Retraités (vieillesse et invalidité) (R,A)
- Conjoints d'actifs (A,C)
- Conjoints de retraités (R,C)
- Enfants d'actifs (A,E)
- Enfants de retraités (R,E)
- Veufs(ves) (V,A)
- Orphelins (O,A)

Pour chacun de ces types nous disposons des effectifs selon les variables suivantes :

- Age, notée x
- Sexe, notée s
- ALD (1 : atteint d'ALD, 0 : non atteint)
- Région de résidence, notée rr

Les effectifs seront notés E_n . En guise d'exemple, l'effectif des assurés, retraités, de sexe féminin, d'âge 30, résidant dans la région 01 et non ALD ($ALD = 0$) sera noté :

$$E_n(A, R, F, 30, 01, 0)$$

Remarque : Un signe « . » est marqué à la place d'une variable sur laquelle on veut appliquer une sommation. Par exemple, l'effectif des assurés retraités d'âge 30 et non ALD est noté :

$$E_n(A, R, ., 30, ., 0)$$

La liste suivante est celle des différents paramètres qui serviront à la modélisation des différents flux démographiques correspondants aux divers types de bénéficiaires définis antérieurement.

Paramètres relatives aux Actifs :

- $Adhes(n)$: nombre de nouvelles adhésions de l'année n , estimé à partir de l'effectif initial moyennant un *taux moyen d'entrée TME*. ($Adhes(n) = TME * \sum E_n(A, A, *, *, *, *)$)
- $TR(type_ass, x, sexe, rr)$: Taux de répartition des nouvelles recrues par âge, sexe et région de résidence)
- $Ts(x)$: taux de sortie par âge des actifs pour une raison autre que le décès ou l'invalidité.
- r : âge de la retraite.
- min : Age minimum des actifs.

Paramètre relatives aux ALD

- $PoidsALD(x)$: Poids des ALD par âge, c'est-à-dire la proportion d'ALD pour un âge donné.
- $TAGpoids_ALD$: taux d'aggravation de $Poids_ALD$

- T_{inv} : probabilité de tomber invalide avec un taux d'invalidité ne permettant plus de pouvoir continuer à travailler, mais plutôt de bénéficier d'une pension de retraite pour motif d'invalidité.

Paramètres Matrimoniaux

- $Tac(TYPE_ASS, s)$: taux d'adhésion des conjoints par sexe, il sert au calcul des nouveaux conjoints issus, soit des mariages des adhérent, soit de la migration de la population bénéficiant des modalités de l'article 114)
- d : Différence d'âge moyenne entre un homme et sa femme.

Paramètres relatives aux enfants

- $PSe(x)$: Probabilité de scolarisation des enfants à l'âge $x > 21$
- $Tfec$: Taux de fécondité
- $PE(s)$: Poids des enfants par sexe
- Nme : Nombre moyen d'enfants par assuré
- e : Différence d'âge moyenne entre l'assuré et un enfant à sa charge

Autres paramètres

- $q(x)$: Probabilité de décéder entre l'âge x et $x+1$
- $D_n(Type_ass, type_ben, s, x, rr)$: Décès de l'année n .
- $INV_n(Inv, A, s, x, rr)$: Nouveaux invalides de l'année n selon le sexe, l'âge et la région de résidence.
- $N_n(type_ass, type_ben, s, x, rr)$: Nouvelles adhésions des assurés selon le type d'assuré, le type du bénéficiaire, le sexe, l'âge et la région de résidence.
- $S_n(A, A, s, x, rr)$: Nouvelles sorties des actifs pour une raison autre que le décès.
- $NV_n(A, C, s, x, rr)$: Nouveaux veufs
- $NO_n(*, E, *, x, *)$: Nombre des nouveaux orphelins suite au décès d'actif ou de retraité.
- $ENR_n(A, E, *, x, *)$: représente les enfants des nouveaux retraités au cours de l'année $n - 1$.

Équations préliminaires

Les équations suivantes sont des équations préliminaires qui définissent les liens entre quelques paramètres énoncés ci-dessus.

Remarque : Le signe « * » sera marqué à la place d'une variable lorsqu'une formule est valable indépendamment de la modalité prise par cette variable.

Pour obtenir l'effectif des ALD d'un groupe de bénéficiaire à partir de l'effectif totale de ce groupe, on utilise la formule suivante :

$$E_n(*, *, *, x, 1, *) = PoidsALD(x) * (1 + TAGpoids_ALD)^n * E_n(*, *, *, x, ., *)$$

Pour les non-ALD :

$$E_n(*, *, *, x, 0, *) = (1 - PoidsALD(x)) * (1 + TAGpoids_ALD)^n * E_n(*, *, *, x, ., *)$$

Ces deux équations supposent que le caractère ALD d'un bénéficiaire ne dépend que de son âge. Le sexe peut aussi influencer la proportion d'ALD par catégorie, cependant le graphe suivant montre que cette influence, que l'on néglige, est minime.

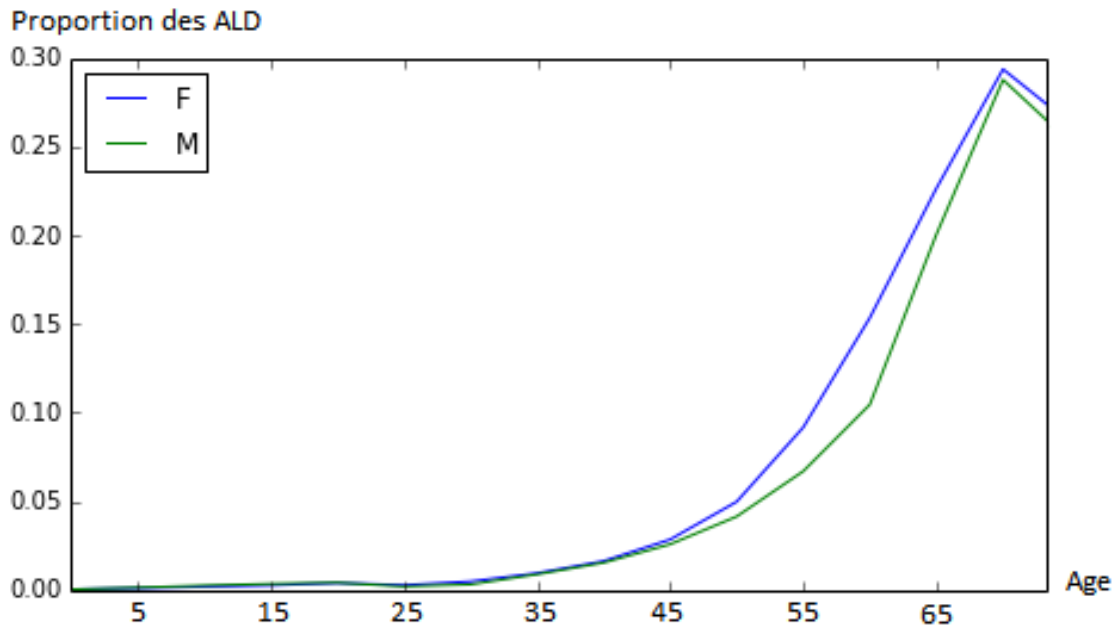


Figure 9 : Proportion des ALD par âge de la population AMO-CNSS selon le sexe

En ce qui concerne les décès de chaque année on retient la formule suivante :

$$D_n(*,*,*,x,*) = q(x) * E_n(*,*,*,x,*,.)$$

Les probabilités de décès proviendront de la table de mortalité TV 88-90. Cette table présente une espérance de vie de 81 ans supérieure à celle publiée par le Haut-Commissariat au Plan (74 ans), et donc permet de prendre une position de prudence face à la vieillesse de la population.

Les nouvelles adhésions d'assurés par catégorie seront déduites à partir de l'adhésion globale de chaque année moyennant le taux de répartition des nouvelles recrues selon la formule suivante :

$$N_n(*,A,*,x,*) = Adhes(n) * TR(*,x,*,*)$$

Ce taux comprend trois composantes par type d'assuré:

$$TR(activ,x,s,rr) = PoidsA(x) * PoidsA(s) * PoidsA(rr)$$

Et :

$$TR(retraité,x,s,rr) = PoidsR(x) * PoidsR(s) * PoidsR(rr)$$

Et finalement les sorties autre qu'en cas de décès ou d'invalidité seront calculées moyennant le taux de sortie pour autre cause que le décès comme suit :

$$S_n(A,A,*,x,*,*) = TS * E_n(A,A,*,x,*,*)$$

II.4. Mise en équation des flux démographiques

Dans cette partie, nous allons définir les équations qui permettront d'obtenir les effectifs, selon les variables d'intérêt, d'une année n à partir de ceux de l'année $n - 1$. Ces équations sont basées sur un bilan des entrées et sorties et seront catégorisées part type de bénéficiaire.

II.4.1. Assurés actifs

Pour $min \leq x < r$, les effectifs des actifs de l'année n s'écrivent :

$$E_n(A, A, *, x, *, .) = E_{n-1}(A, A, *, x - 1, *, .) + N_{n-1}(A, A, *, x - 1, *) \\ - N_{inv_{n-1}}(Inv, A, *, x - 1, *) - D_{n-1}(A, A, *, x - 1, *) \\ - S_{n-1}(A, A, *, x - 1, *)$$

Vue que la proportion des actifs ayants un âge supérieur à 60 est de 0.53 % de la population des actifs AMO CNSS, nous allons négliger cette catégorie de bénéficiaires.

II.4.2. Retraités

S'agissant des retraités, il y'a lieu de distinguer entre les anciens retraités ($x > r$), et les nouveaux ($x = r$).

Pour $x > r$:

$$E_n(R, A, *, x, *, .) = E_{n-1}(R, A, *, x - 1, *, .) + N_{n-1}(R, A, *, x - 1, *) - D_{n-1}(R, A, *, x - 1, *) \\ + N_{inv_{n-1}}(Inv, A, *, x - 1, *)$$

Pour $x = r$:

$$E_n(R, A, *, x, *, .) = E_{n-1}(A, A, *, x - 1, *, .) + E_{n-1}(R, A, *, x - 1, *, .) + N_{n-1}(A, A, *, x - 1, *) \\ - D_{n-1}(R, A, *, x - 1, *, .)$$

Dans l'équation ci-dessus, le terme « $E_{n-1}(R, A, *, x - 1, *, .)$ » représente les gens ayant bénéficié d'une retraite anticipée.

II.4.3. Conjoints

On s'intéresse ici aux conjoints des gens étant affilié à l'AMO-CNSS. L'effectif des conjoints d'actifs est le suivant :

$$E_n(A, C, *, x, *, .) = E_{n-1}(A, C, *, x - 1, *, .) - D_{n-1}(A, C, *, x - 1, *) + N_{n-1}(A, C, *, x - 1, *) \\ - CNR_{n-1}(A, C, *, x - 1, *) - CN_{inv_{n-1}}(A, C, *, x - 1, *) \\ - S_{n-1}(A, C, *, x - 1, *) - NV_{n-1}(A, C, *, x - 1, *)$$

En ce qui concerne les conjoints de retraités :

$$E_n(R, C, *, x, *, .) = E_{n-1}(R, C, *, x, *, .) - D_{n-1}(R, C, *, x - 1, *, .) + N_{n-1}(R, C, *, x - 1, *) \\ + CNR_{n-1}(A, C, *, x - 1, *) + CN_{inv_{n-1}}(A, C, *, x - 1, *) \\ - NV_{n-1}(R, C, *, x - 1, *)$$

Le terme « $N_{n-1}(*, C, *, x - 1, *)$ » représente les nouvelles adhésions des conjoints, d'âge $x - 1$ au cours de l'année $n - 1$. Ce qui englobe les mariages des adhérents au régime et les conjoints des nouveaux adhérents provenant de la population 114. Selon le sexe du conjoint, cette grandeur se calcule comme suit :

- Conjoint de sexe féminin :

$$N_{n-1}(*, C, F, x - 1, *) = Tac(*, F) * E_{n-1}(*, A, H, x + d, *)$$

- Conjoint de sexe Masculin :

$$N_{n-1}(*, C, H, x - 1, *) = Tac(*, H) * E_{n-1}(*, A, F, x - d, *)$$

Le terme « $NV_{n-1}(*, C, *, x - 1, *)$ » représente les nouveaux veufs(ves) d'adhérents de l'année $n - 1$.

Pour calculer cet effectif, nous définissons le poids des conjoints à charge selon le sexe comme suit :

$$P_Conj_n(*, F, x, *) = \frac{E_n(*, C, F, x, *, .)}{E_n(*, A, H, x + d, *, .)}$$

Et :

$$P_Conj_n(*, H, x, *) = \frac{E_n(*, C, H, x, *, .)}{E_n(*, A, F, x - d, *, .)}$$

Et donc nous avons :

$$NV_{n-1}(*, C, F, x, *) = P_Conj_{n-1}(*, F, x, *) * D_{n-1}(*, A, H, x + d, *)$$

Et :

$$NV_{n-1}(*, C, H, x, *) = P_Conj_{n-1}(*, H, x, *) * D_{n-1}(*, A, F, x - d, *)$$

Nous allons calculer de la même manière les « $S_{n-1}(A, C, *, x - 1, *)$ », qui représente le départ des conjoints suite au départ de l'adhérent pour autre raison que le décès, selon les deux équations suivantes :

$$S_n(A, C, F, x, *) = P_Conj_n(*, F, x, *) * S_n(A, A, H, x + d, *)$$

Et :

$$S_n(A, C, H, x, *) = P_Conj_n(*, H, x, *) * S_n(A, A, F, x + d, *)$$

Le terme « $CNR_{n-1}(*, C, *, x - 1, *)$ » représente les conjoint des adhérent sortant à la retraite à l'année n . On le calcule, selon le sexe comme suit :

$$CNR_{n-1}(*, C, M, x - 1, *) = \begin{cases} E_{n-1}(A, C, M, x - 1, *, .) & \text{si } x - 1 - d + 1 = r \\ 0 & \text{sinon} \end{cases}$$

Et :

$$CNR_{n-1}(*, C, F, x - 1, *) = \begin{cases} E_{n-1}(A, C, F, x - 1, *, .) & \text{si } x - 1 + d + 1 = r \\ 0 & \text{sinon} \end{cases}$$

Finalement, le terme « $CNInv_{n-1}(*, C, *, x - 1, *)$ » représente les conjoints des nouveaux invalides et ce calcule selon le sexe comme suit :

$$CNInv_n(A, C, F, x, *) = P_Conj_n(*, F, x, *) * Ninv_n(A, A, H, x + d, *)$$

Et :

$$CNInv_n(A, C, H, x, *) = P_Conj_n(*, H, x, *) * Ninv_n(A, A, F, x + d, *)$$

II.4.4. Les veufs(ves)

Pour une année donnée n , l'effectif des veufs selon les variables d'intérêt est le suivant :

$$E_n(V, A, *, x, *, .) = E_{n-1}(V, A, *, x - 1, *, .) - D_{n-1}(V, A, *, x - 1, *) + NV_{n-1}(A, C, *, x - 1, *) + EV_{n-1}(*, x - 1, *) + NV_{n-1}(R, C, *, x - 1, *)$$

Le terme « $EV_{n-1}(*, x - 1, *)$ » représente les nouvelles entrées de veufs suite à la migration de la population 114. On le calcule à partir de l'effectif globale des veufs de la population 114 multiplié par des poids représentant les variable d'intérêt, c'est-à-dire que :

$$EV_n(s, x - 1, rr) = EV_n(., ., ., ., .) * PoidseV_n(x - 1) * PoidseV_n(s) * PoidseV_n(rr) * Poidse114_n$$

Où « $Poidse114_n$ » la proportion des 114 qui vas basculer vers L'AMO CNSS à l'année n .

II.4.5. Les enfants

En ce qui concerne les enfants des adhérents, on distingue entre les enfants d'actifs et ceux des retraités. Selon ces deux catégories, on calcule l'effectif des enfants comme suit :

Pour $x > 0$:

$$E_n(A, E, *, x, *, .) = PSe(x) * [E_{n-1}(A, E, *, x - 1, *, .) + EE_{n-1}(A, E, *, x - 1, *) - D_{n-1}(A, E, *, x - 1, *) - ENR_{n-1}(A, E, *, x - 1, *) - ENInv_{n-1}(A, E, *, x - 1, *) - S_{n-1}(A, E, *, x - 1, *) - NO_{n-1}(A, E, *, x - 1, *)]$$

Et :

$$E_n(R, E, *, x, *, .) = PSe(x) * [E_{n-1}(R, E, *, x - 1, *, .) + EE_{n-1}(R, E, *, x - 1, *) - D_{n-1}(R, E, *, x - 1, *) + ENR_{n-1}(A, E, *, x - 1, *) + ENInv_{n-1}(A, E, *, x - 1, *) - NO_{n-1}(R, E, *, x - 1, *)]$$

Si $x = 0$:

$$E_n(*, E, s, x, *, .) = Tfec * PE(s) * \sum_{k=15}^{49} [E_{n-1}(*, A, F, k, *, .) + E_{n-1}(*, C, F, k, *, .)]$$

Avec :

- $EE_{n-1}(A, E, *, x - 1, *, .)$: l'effectif des enfants d'actifs issus de la population 114, on le calcule de la même façon que l'effectif des veufs issus de la population des 114.
- $NO_{n-1}(*, E, *, x - 1, *)$: Nombre des nouveaux orphelins suite aux décès d'actifs ou de retraités, ce calcule comme suit :

$$NO_{n-1}(*, E, s, x - 1, *) = Nme * D_{n-1}(*, A, *, x - 1 + e, *) * PE(s)$$

- $S_{n-1}(A, E, *, x - 1, *)$: représente les sorties des enfants suite à la sortie des parents pour autres raisons que le décès, ce calcule comme suit :

$$S_{n-1}(A, E, s, x - 1, *) = Nme * S_{n-1}(A, A, *, x - 1 + e, *) * PE(s)$$

- $ENR_{n-1}(A, E, *, x - 1, *)$: représente les enfants des nouveaux retraités au cours de l'année $n - 1$, se calcule comme suit :

$$ENR_{n-1}(A, E, *, x - 1, *) = \begin{cases} E_{n-1}(A, E, *, x - 1, *) & \text{si } x - 1 - e + 1 = r \\ 0 & \text{sinon} \end{cases}$$

- $ENInv_{n-1}(A, E, *, x - 1, *)$: représente les enfants des nouveaux invalides au cours de l'année $n - 1$, ce calcule comme suit :

$$ENInv_{n-1}(A, E, s, x - 1, *) = Nme * NInv_{n-1}(Inv, A, *, x - 1 + e, *) * PE(s)$$

II.4.6. Les orphelins

En ce qui concerne les orphelins de la population AMO CNSS leur fonction d'effectif s'écrit comme suit :

$$E_n(O, A, *, x, *, .) = E_{n-1}(O, A, *, x, *, .) + EO_{n-1}(O, A, *, x - 1, *) - D_{n-1}(O, A, *, x - 1, *) + NO_{n-1}(A, E, *, x - 1, *) + NO_{n-1}(R, E, *, x - 1, *)$$

Le terme « $EO_{n-1}(O, A, *, x - 1, *)$ » représente les orphelins issus de la migration de la population 114 et ce calcul de la même façon que celui des veufs issus de la même population.

Nous venons d'énoncer les équations qui serviront plus tard aux projections. Dans la partie qui suivra, nous allons estimer les différents paramètres définis ci-dessus. Ces estimations proviendront des bases de données relatives à la population AMO CNSS.

II.4.7. Estimation des paramètres

Pour estimer les différents paramètres du modèle démographique proposé, l'ANAM préconise de se baser sur les données de la période 2009-2012. Pour des soucis de confidentialité, les tables utilisées et les valeurs retenues des paramètres ne figureront pas dans ce rapport.

Les projections démographiques ont été programmées à l'aide du langage Python. Le code de la fonction de projection figure dans l'annexe III.

La figure suivante représente l'évolution de la population globale selon notre modèle :

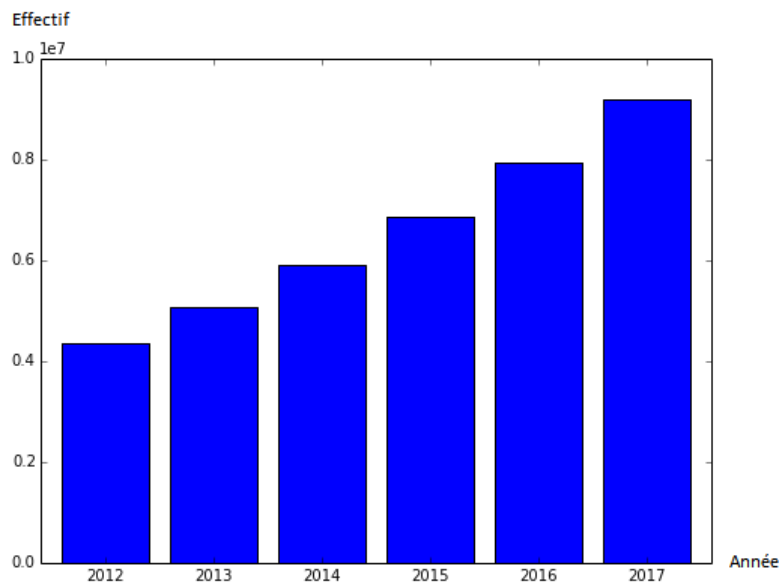


Figure 10 : Effectifs globaux projetés de la population AMO-CNSS

Nous constatons que l'allure de l'évolution de la population ne présente pas d'anomalies. Cette même allure est retrouvée en ce qui concerne les projections des effectifs de tous les types de bénéficiaires.

II.4.8. Test de sensibilité

Avant d'utiliser les prévisions obtenues, il est primordial de tester la sensibilité du modèle face aux variations des paramètres. Nous allons nous intéresser à la sensibilité des différentes sorties du modèle pour des variations locales des divers paramètres.

L'étude de la sensibilité locale nous permettra d'identifier les paramètres qui devront être estimés avec le plus de précision. En effet si le modèle est très sensible à l'un des paramètres, l'intervalle de confiance de ce dernier devrait être le plus petit possible, sinon la sortie du modèle risque d'être entachée d'un niveau élevé d'incertitude.

Sensibilité au taux d'adhésion

Le taux d'adhésion est une grandeur qui impacte globalement la population. Nous choisirons comme output du modèle l'effectif global de la population.

En faisant varier le taux d'adhésion de 1% nous constatons que l'effectif global de la population projetée varie de 0.1789 %, tandis que variation de 10% de ce taux fait varier l'effectif total de 1.79%. L'impact de la variation du taux d'adhésion sur la sortie du modèle n'est donc pas alarmant.

Sensibilité au taux de sortie pour autre que le décès ou la retraite

Le taux de sortie pour autre que le décès ou la retraite est directement lié à l'effectif des actifs de la population, cet effectif est retenu comme output du test.

Une variation de 1%, respectivement 10%, de ce taux fait varier l'effectif des actifs de -0.001%, respectivement -0.014%. Le modèle n'est donc pas très sensible à ce paramètre.

Sensibilité au taux de fécondité

Pour tester la sensibilité du modèle au taux de fécondité, nous retenons le nombre d'enfant comme output du modèle.

Une variation de la valeur du taux de fécondité de 1% induit une variation du nombre d'enfants de 0.075 %, tandis qu'une variation du taux de 10% induit une augmentation de 0.748% du nombre d'enfant. Cependant, pour chaque variation du taux de fécondité, le nombre de nouveaux-nés varie de la même proportion. Donc, l'estimation du taux de fécondité doit être la plus précise possible.

Sensibilité au taux d'invalidité

L'effectif des invalides est calculé à partir de celui des actifs. Nous allons donc nous intéresser à la variation du nombre d'actif face à celle du taux d'invalidité.

Une augmentation de 1% du taux d'invalidité entraîne une diminution du nombre d'actifs de 0.00075%. Le modèle n'est donc pas très sensible au taux d'invalidité.

Chapitre 3 : Consommation

Ce chapitre est consacré à la consommation relative à l'AMO, c'est-à-dire aux remboursements des frais des soins.

Nous exploiterons les données mises à notre disposition pour décrire les remboursements de l'année 2012 selon les caractéristiques des soins s'y relatant et ceux du bénéficiaire. Pour ensuite modéliser ces remboursements à l'aide des variables d'intérêt, qui sont ceux retenus lors de la modélisation démographique.

I. Analyse descriptive :

Dans cette partie, nous allons tout d'abord présenter la structure des données disponibles sur la consommation en assurance maladie, ensuite dégager les statistiques d'intérêt, pour enfin exposer les résultats de l'analyse des données.

I.1. Structure des données :

Les données mises à notre disposition sont organisés sous forme de cube (ou hypercube) selon les axes (inputs ou variables descriptives) suivants :

- *EXER_SURV* : L'exercice de survenance du groupe de sinistres. Nous nous intéresserons aux sinistres survenus en 2012.
- *CODE_ACTE* : Un code représentant les actes médicaux assujettis à l'assurance maladie.
- *VILLE* : Ville de résidence du bénéficiaire.
- *TYPE_DOSS* : indique si le dossier est de nature tiers payant ou déboursement direct.
- *TYPE_ASS* : Type d'assuré, identique à celle présentée au niveau de la base de donnée de a population.
- *TYPE_BEN* : Type du bénéficiaire, identique à celle présentée au niveau de la base de donnée de a population.
- *TR_AGE* : Tranche d'âge du bénéficiaire du soin médicale.
- *SEXE* : Sexe du bénéficiaire.
- *SECTEUR_ETABL* : Secteur de l'établissement relatif au soin médical (public ou privé)
- *ALD* : Indique si le bénéficiaires est atteint ou non d'ALD.
- *CODE_ALD* : Indique la nature de l'affection de longue durée.
- *TYPE_SOINS* : Type de soins, ambulatoire ou bien hospitalisation.

Les outputs du cube sont les suivants :

- *FRAIS_ENG* : Frais engagés globaux de la classe de population (remboursés ou non)
- *EFFECT_SINISTR* : L'effectif sinistré selon la classe de population.
- *NBRE_DOSS* : Nombre de dossiers liquidés de la classe de population.
- *MONT_REM* : Montant globale liquidé de la classe de population.

Parmi ces outputs, la variable *MONT_REM* est d'un intérêt particulier, car elle constitue les dépenses du régime en question et permet donc de juger de sa viabilité. Néanmoins, cette grandeur est influencée par l'effectif sinistré de la tranche, qui lui-même est l'un des outputs de la base de donnée.

Pour ne pas fausser l'analyse, nous allons calculer le montant moyen remboursé de la tranche, qui sera nommé *MONT_REM_M* et qui vérifie :

$$MONT_REM_M = \frac{MONT_REM}{EFFECT_SINISTR}$$

Une correspondance entre les villes et les régions a été établie pour dresser la colonne *REGION*, contenant la région de résidence du bénéficiaire.

Dans ce qui suit, le terme « classe de consommation » fera référence à une ligne de la base de données.

I.2. Exploration des données

I.2.1. Grandeurs globales :

Les grandeurs approximatives suivantes sont relatives à l'année 2012 et concernent la population globale des assurés :

- Frais de santé de la population globale : 2.15 milliards de DH
- Montant des remboursements de l'année : 1.49 milliards de DH
- Taux de remboursement global de l'année : 68%
- Effectif sinistré de l'année : 650 000

I.2.2. Analyse descriptive :

Nous allons nous intéresser à la consommation médicale des bénéficiaires de l'AMO CNSS, selon quelques variables descriptives et à travers les trois grandeurs : l'effectif sinistré, le montant moyen remboursé et les frais moyens engagés par personne.

1) Consommation médicale par âge :

Le graphe suivant représente les montants moyens remboursés, ainsi que les frais moyens engagés par bénéficiaire selon l'âge :

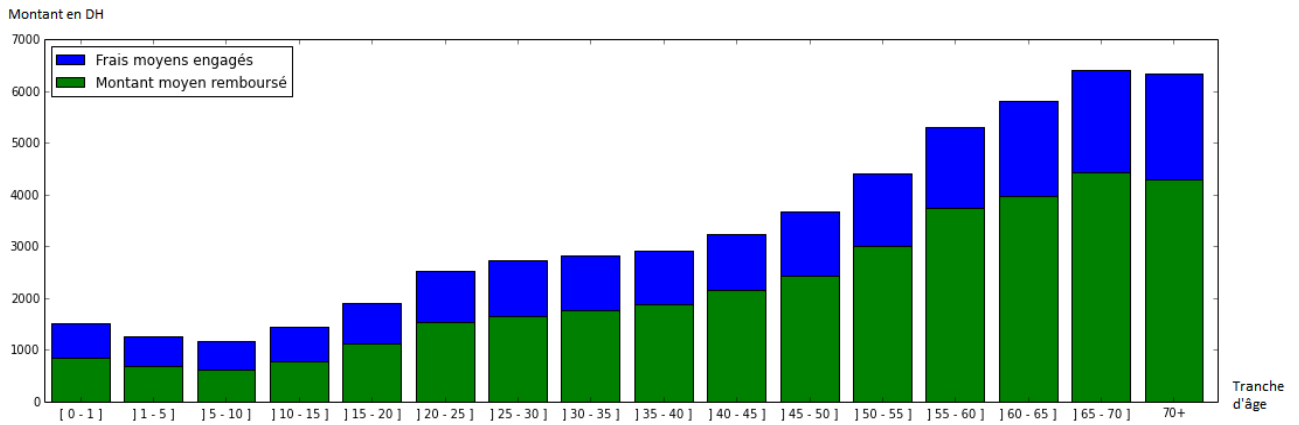


Figure 11 : Consommation médicale moyenne d'un bénéficiaire de l'AMO-CNSS par tranche d'âge

On remarque qu'à partir de 5 ans, les montants moyens engagés augmentent avec l'âge, ce qui n'est pas le cas des moins de 5 ans, ceci pourrait être dû à la fragilité des enfants.

Les effectifs sinistrés de chaque âge sont les suivants :

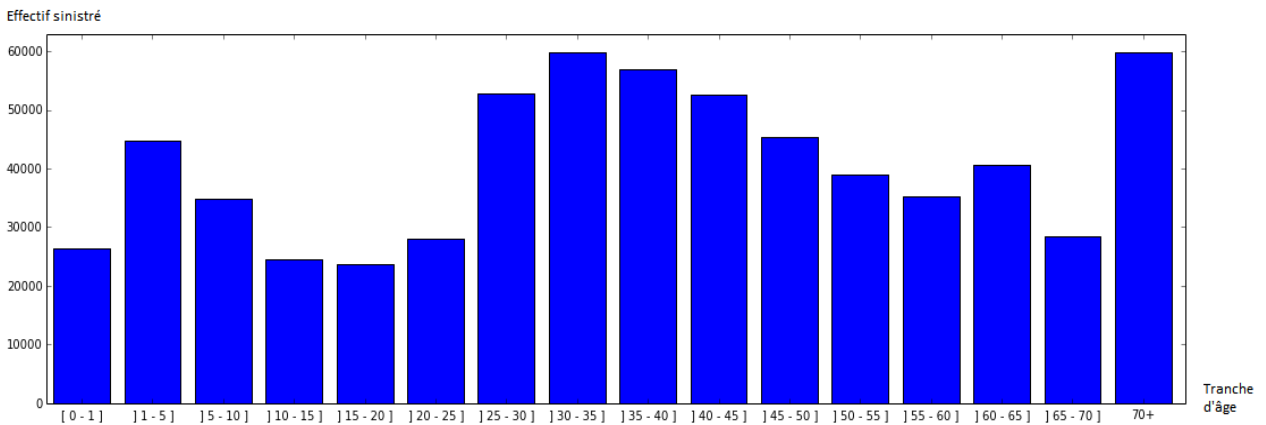


Figure 12 : Effectif sinistré selon l'âge des bénéficiaires

Les effectifs sinistrés varient significativement selon l'âge, cependant on ne peut conclure que l'une des tranches d'âge est plus risquée que les autres, vu que le graphique représente l'effectifs sinistré absolu et non le taux de sinistralité (nombre de sinistre par personne) de chaque tranche.

En effet, l'effectif sinistré de deux classes ayant le même niveau de risque sera proportionnel à la taille de chaque tranche.

Nous allons donc nous intéresser, dans ce qui suit, au taux de sinistralité de chaque tranche au lieu de l'effectif sinistré. Pour ce faire nous devons réaliser une jointure entre la base de données de la consommation et celle relative à la population. La fonction *merge* du package python Pandas permet de réaliser une telle opération.

Les taux de sinistralité selon les tranches d'âge sont les suivants :

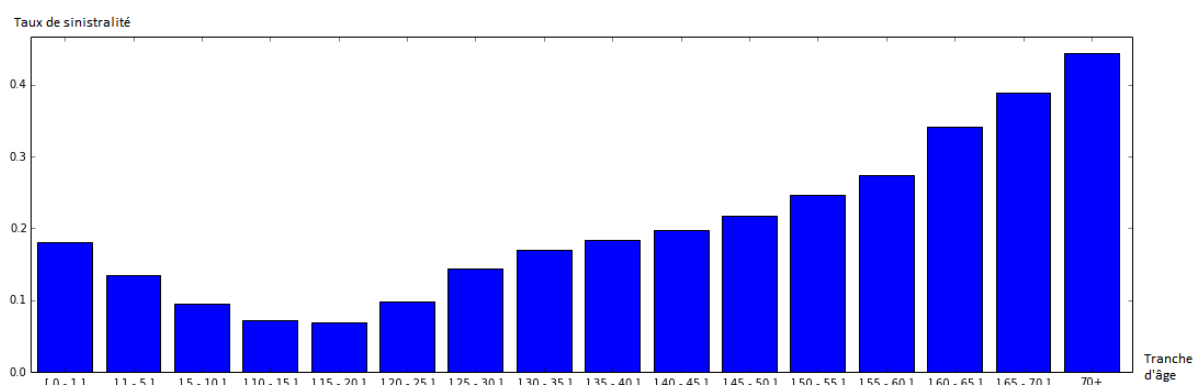


Figure 13 : Taux de sinistralité de la population AMO-CNSS selon l'âge

Nous constatons que la tendance de la courbe du taux de sinistralité par âge rejoint celle des frais engagés par âge. On peut donc conclure qu'à partir de 10-15 ans, les bénéficiaires sont sujets à des sinistres de plus en plus fréquents et coûteux. D'un point de vu risque, les bénéficiaires deviennent de plus en plus risqués avec le vieillissement dès qu'ils dépassent l'âge de 15 ans et inversement.

2) Consommation médicale selon le type du bénéficiaire

La figure suivante représente les montants moyens remboursés, ainsi que les frais moyens engagés selon le type de bénéficiaire:

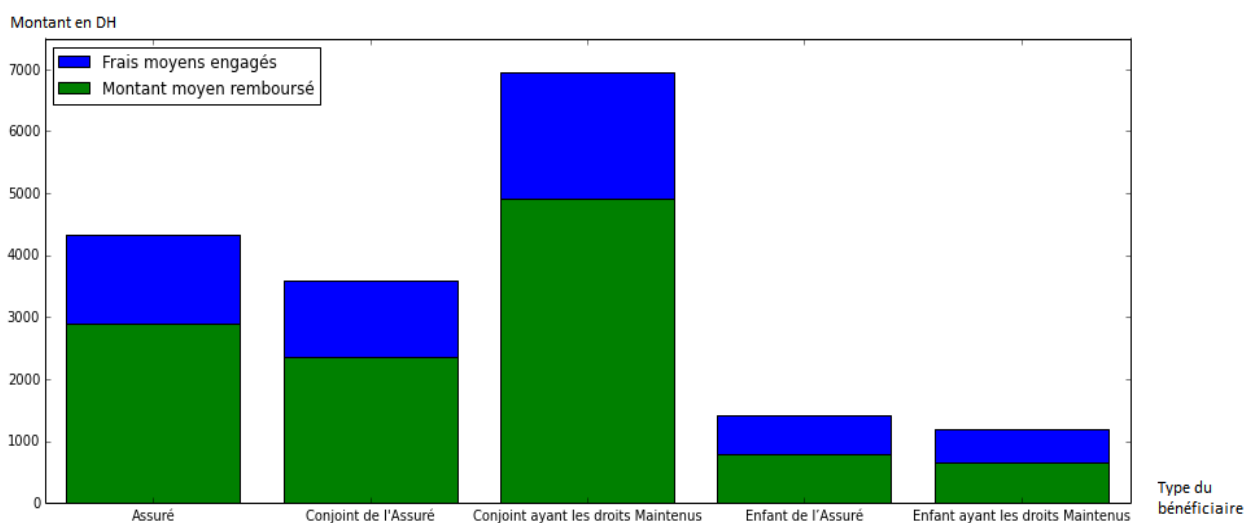


Figure 14 : Consommation médicale moyenne d'un bénéficiaire de l'AMO-CNSS selon son type

Nous remarquons que les frais moyens engagés par les conjoints ayants les droits maintenus sont significativement plus importants que ceux des autres types de bénéficiaires. Cela pourrait être dû au fait que cette catégorie est constituée majoritairement de conjoints d'assurés décédés de causes naturelles, et donc de bénéficiaires relativement âgés.

3) Consommation médicale par sexe :

Les frais moyens engagés et le montant moyen remboursé selon le sexe sont illustrés dans le graphique suivant :

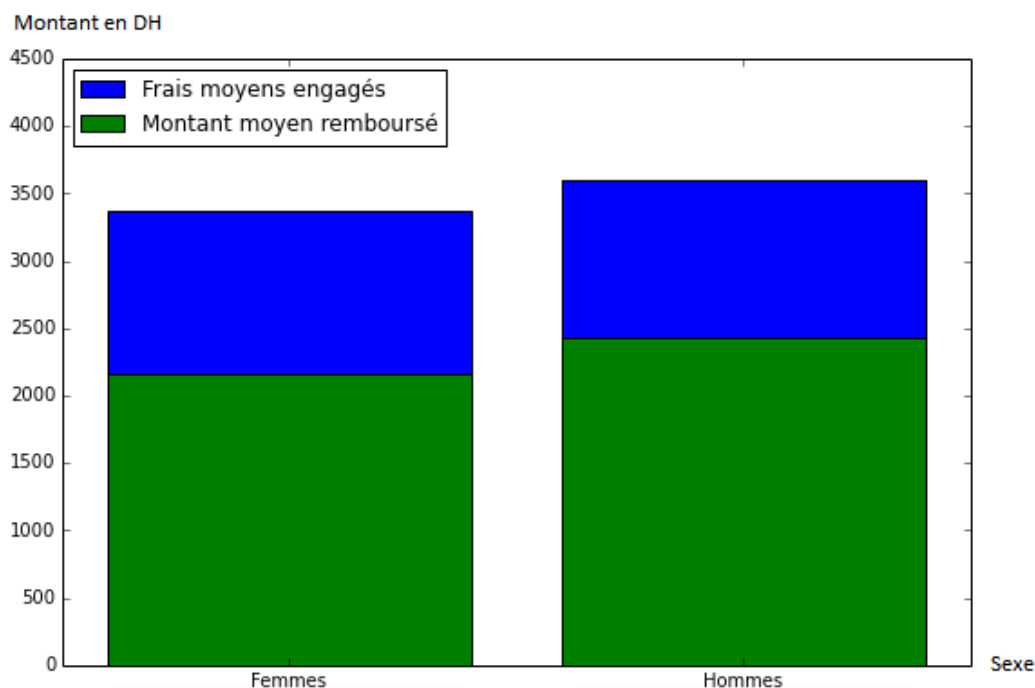


Figure 15 : Consommation médicale moyenne d'un bénéficiaire de l'AMO-CNSS selon le sexe

Il existe une légère différence de près de 200DH entre le coût moyen des soins des hommes et ceux des femmes. Cependant, comme le montre le graphe suivant, le taux de sinistralité des femmes est plus important que celui des hommes.

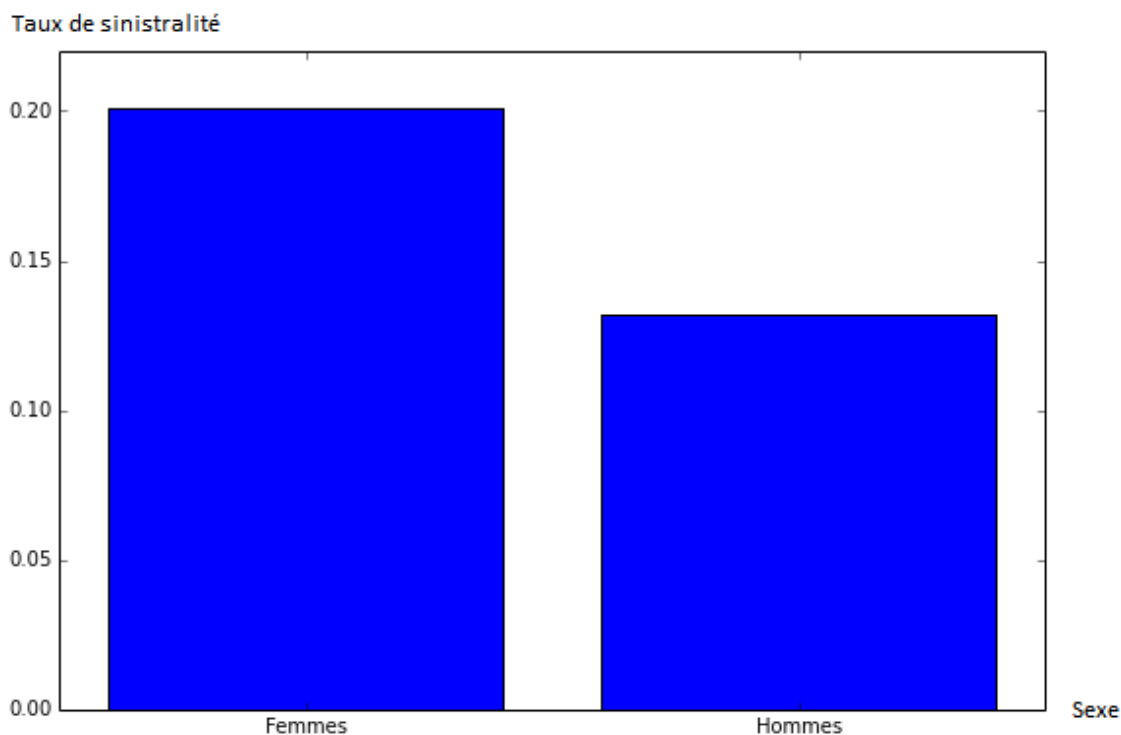


Figure 16 : Taux de sinistralité de la population AMO-CNSS selon le sexe

4) Consommation médicale selon la variable ALD

Nous nous intéressons aux remboursements et aux frais moyens engagés par les bénéficiaires selon que ceux-là soient atteints de maladies de longues durées (ALD) ou non. Ces deux grandeurs sont illustrées par le graphique suivant :

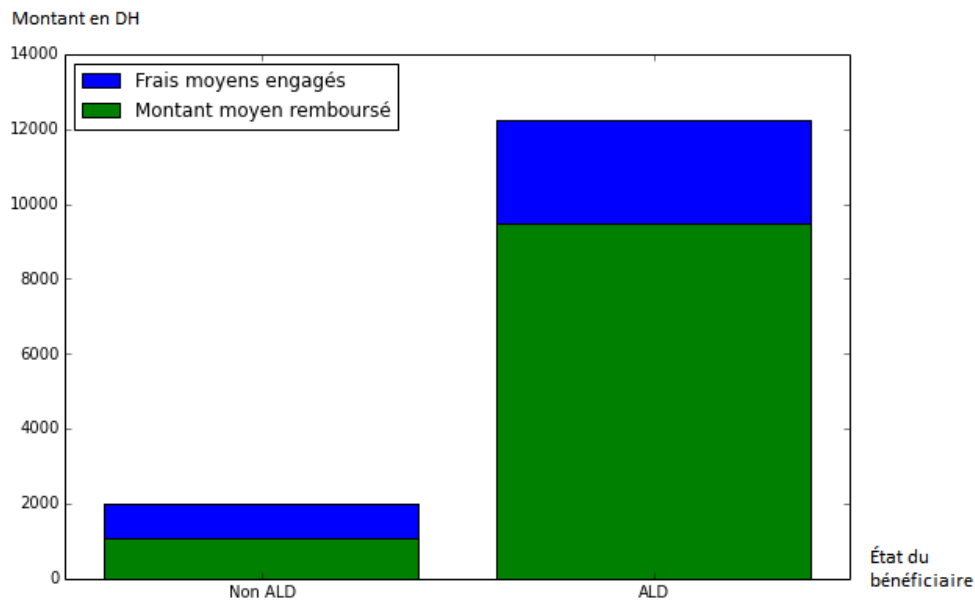


Figure 17 : Consommation médicale moyenne d'un bénéficiaire de l'AMO-CNSS selon son état de santé

Nous constatons que les frais moyens engagés par une personne atteinte d'ALD représentent plus de six fois ceux d'une personne non atteinte, ce qui est naturel vu que les soins médicaux relatifs aux maladies de longue durée sont plus onéreux.

On s'attend aussi à un taux de sinistralité plus élevé chez les ALD, ce qui est illustré par le graphique suivant :

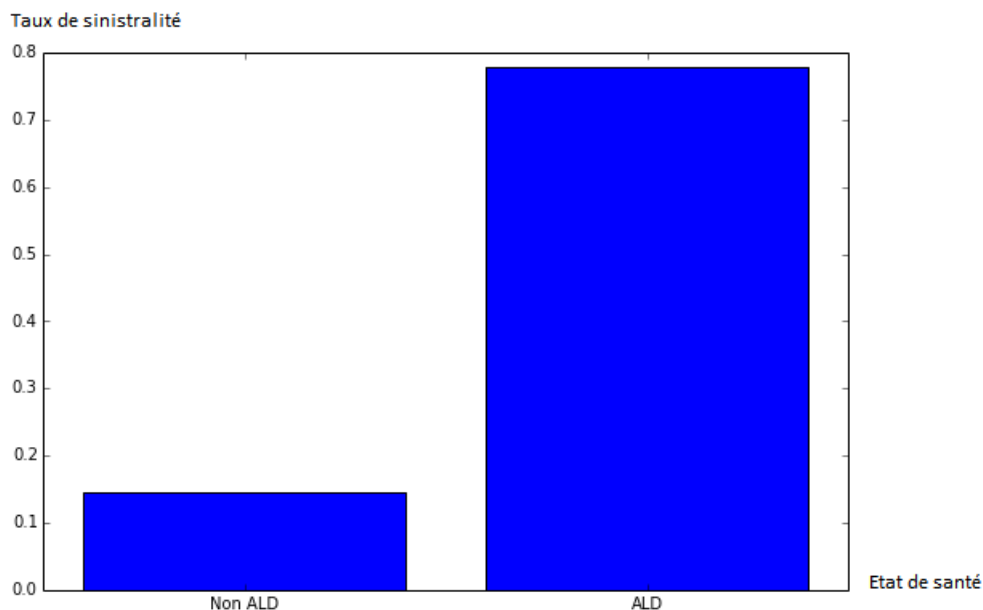


Figure 18 : Taux de sinistralité de la population AMO-CNSS selon l'état de santé des bénéficiaires

5) Consommation médicale par secteur d'établissement de soins :

Il est intéressant de comparer les frais et les remboursements des soins administrés par les hôpitaux du secteur public et ceux du secteur privé.

On filtre d'abord sur la variable *TYPE_SOINS* pour ne prendre en compte que les soins hospitaliers et on regroupe la base de données selon la variable *SECTEUR_ETABLE*.

Le diagramme en camembert suivant représente les proportions des sinistres relatives à chaque secteur d'établissement :

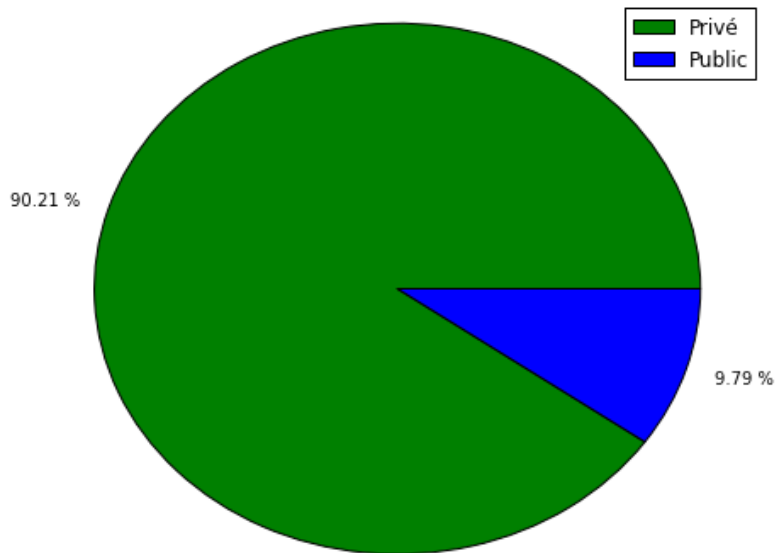


Figure 19 : Répartition des soins hospitaliers administrés à la population AMO-CNSS selon le secteur de l'établissement de soins

On remarque que la majorité des hospitalisations sont effectuées dans le secteur privé. Il existe aussi une différence significative entre le coût moyen d'hospitalisation du secteur privé (13 577 DH) et celui du secteur public (9 141 DH). Cependant la CNSS ne fait pas de discrimination entre les bénéficiaires hospitalisés au secteur public et ceux hospitalisés au secteur privé vu que le taux de remboursement relatif aux soins hospitaliers, représenté par le graphique ci-dessous, est presque le même pour les deux secteurs.

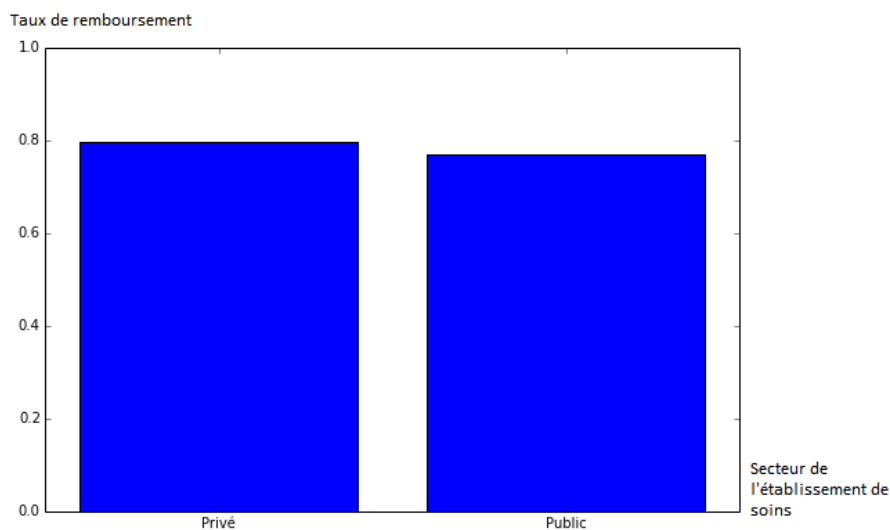


Figure 20 : Taux de remboursement selon le secteur de l'établissement de soins

I.3. Dépendances entre inputs :

Dans cette partie nous allons tester l'existence de dépendance entre les variables d'intérêt. Nous n'allons pas croiser toutes les variables deux à deux vu que l'étude du degré de dépendance entre quelques variables (Exemple : Sexe et Région) n'apporte pas d'information pertinente à notre étude.

I.3.1. Variables catégorielles deux-à-deux

Pour tester la dépendance entre variables qualitatives nous utiliserons le test de Chi-deux.

La statistique de Chi-deux traite les modalités des variables comme étant indépendantes, cependant la variable TR_AGE est une variable quantitative (variable intervalle), ses modalités sont donc ordonnées, d'autres approches sont donc préconisées.

La librairie *STATSMODELS* contient la fonction *chi2_contingency()* qui permet d'effectuer le test de Chi-deux sur une table de contingence à deux critères.

On définit une fonction *get_contingency2w()* qui permet d'extraire une table de contingence à deux critères à partir de la base de données et on effectue le test de Chi-deux sur toutes les combinaisons pertinentes de variables catégorielles.

Les sorties des tests sont les suivantes :

ALD VS TYPE_SOINS X ² = 7131.73527576 p-value = 0.0 DDL = 1	TYPE_BEN VS SEXE X ² = 439139.519152 p-value = 0.0 DDL = 4	SEXES VS SECTEUR_ETABL X ² = 30.7951637741 p-value = 2.8675e-08 DDL = 1
SECTEUR_ETABL VS ALD X ² = 7530.54289936 p-value = 0.0 DDL = 1	SEXES VS TYPE_SOINS X ² = 230.423142397 p-value = 4.8202e-52 DDL = 1	SEXES VS ALD X ² = 2781.97701522 p-value = 0.0 DDL = 1
SECTEUR_ETABL VS TYPE_SOINS X ² = 1434.66557949 p-value = 0.0 DDL = 1		

Tableau 2 : Test d'indépendance entre les variables de consommation catégoriel

Toutes ses combinaisons présentent donc une dépendance. Cependant les tests d'indépendance ne donnent pas d'idée détaillée sur la nature de dépendance entre les variables. Il existe plusieurs méthodes pour pouvoir comprendre le lien entre celle-ci.

Dans le cas des variables catégorielles on peut recourir à la décomposition de la statistique de Chi-deux, ou, quand le nombre de modalités de chaque variable est de deux, examiner le rapport des cotes de la table (Odds Ratio).

1) Dépendance entre la variable ALD et le sexe

Nous cherchons à répondre à la question suivante : Est-ce que l'atteinte d'ALD impacte la sinistralité des hommes et de femmes de la même façon ?

Chaque variable n'a que deux modalités, donc on examine le rapport des cotes de la table de contingence suivante :

	Non-ALD	ALD
Femmes	986327	235405
Hommes	638857	183405

Tableau 3 : Effectif sinistré de la population AMO-CNSS selon le sexe et l'état de santé du bénéficiaire

En notant p (q) la probabilité qu'un homme (femme) sinistré(e) soit atteint(e) d'une affection de longue durée, nous avons :

$$p = \frac{183405}{183405 + 638857} = 0.2230 \quad \text{et} \quad q = \frac{235405}{235405 + 986327} = 0.1927$$

Le rapport des côtes, ou le risque relatif, est défini comme :

$$odd = \frac{p/(1-p)}{q/(1-q)} = 1.2 > 1$$

Ce rapport, étant supérieur à un, indique que l'atteinte d'ALD impacte positivement la sinistralité des hommes plus que celle des femmes.

2) Décomposition de la statistique de chi-deux. (TYPE_BEN, SEXE)

On s'intéresse aux deux variables TYPE_BEN et SEXE. La table de contingence représentant ces deux variables est la suivante :

	A	CA	CM	EA	EM	profile lignes
Femmes	163771	167309	737	58687	27	390531
Hommes	194428	3765	8	64608	29	262838
Profile colonnes	358199	171074	745	123295	56	653369

Tableau 4 : Effectif sinistré de la population AMO-CNSS selon le sexe et le type du bénéficiaire

Avec :

- A : Assuré
- CA : Conjoint d'assuré
- EA : Enfant d'assuré
- CM : Conjoints ayant les droits maintenus
- EM : Enfant ayant les droits maintenus

◆ Rappel sur le test de Chi-deux :

Soit une table de contingence bidimensionnelle à i lignes et j colonnes. On considère les notations suivantes :

- n_{ij} : L'effectif figurant dans la cellule (i, j) .
- n_{i*} : L'effectif totale de la ligne i .
- n_{*j} : L'effectif totale de la ligne j .
- n : Effectif globale de la table de contingence.
- $\mu_{ij} = \frac{n_{i*} \times n_{*j}}{n}$: L'effectif espéré de la cellule (i, j) en cas d'indépendance

La formule de la statistique de Chi-deux est la suivante :

$$\chi^2 = \sum_{(i,j)} \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}}$$

Cette statistique suit une loi de chi-deux à $(i - 1)(j - 1)$ degrés de libertés.

L'hypothèse nulle du test est : H_0 : Les deux variables sont indépendantes. On rejette H_0 quand la statistique dépasse le quantile de la loi de chi-deux au seuil fixé (5% généralement).

Remarque : Un autre test asymptotiquement équivalent au test de Chi-deux est le test du maximum de vraisemblance. Ce test repose sur la statistique G^2 qui elle aussi suit une loi de chi-deux à $(i - 1)(j - 1)$ degrés de libertés et qui se calcule selon la formule suivante :

$$G^2 = 2 \sum_{(i,j)} n_{ij} \log \left(\frac{n_{ij}}{\mu_{ij}} \right)$$

Nous avons donc :

$$\chi^2 = 140372.651 \quad \text{et} \quad G^2 = 179713.802$$

Le test sur ces deux variables affirme qu'elles ne sont pas indépendantes. Cependant nous pouvons soupçonner que les enfants des assurés sont répartis uniformément selon le sexe et que la dépendance entre les deux variables ne provient que du fait que la majorité des conjoints (modalités CA et CM) sont de sexe féminin. Pour tester cela, on peut subdiviser la table de contingence en sous-tables de façon à faire ressortir les catégories d'individus d'intérêt. Toutefois, la subdivision doit suivre les trois règles suivantes :

- a. La somme des degrés de liberté des sous-tables doit être égale à celle de la table d'origine.
- b. Une cellule provenant de la table d'origine doit figurer dans une seule sous-table.
- c. Les sommes marginales (selon les lignes ou les colonnes) de la table d'origine doivent correspondre aux sommes marginales d'une seule des sous-table.

Une subdivision de la table $TYPE_BEN \times SEXE$ respectant les règles précitées est la suivante :

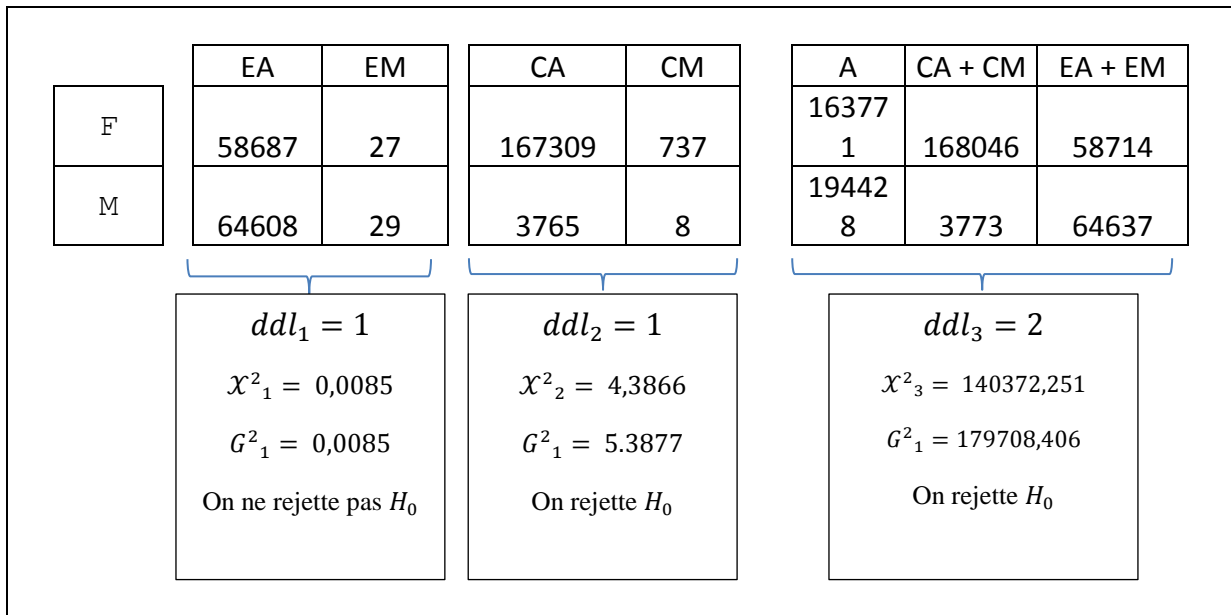


Figure 21 : Schématisation de la procédure de décomposition de la statistique de Chi-deux

En effet, nous concluons que les enfants des assurés sont répartis uniformément selon le sexe.

Remarque : Pour savoir si la subdivision effectuée est juste, la somme des statistiques du maximum de vraisemblance des sous-tables doit être égale à celle de la table d'origine. Ce qui est le cas de la subdivision effectuée.

1.3.2. Dépendance avec la variable TR_AGE :

On ce qui concerne la dépendance entre une variable quantitative (TR_AGE) et une ou plusieurs variables catégorielles, le test adéquat est le test d'ANOVA.

Nous devons tout d'abord recoder la variable TR_AGE de façon à ce que chaque modalité soit remplacée par le centre de l'intervalle qu'elle représente.

Encore une fois, étudier la dépendance entre la variable TR_AGE avec quelques une des variables descriptives n'apporte pas d'information pertinente. Par exemple il est naturel que les enfants soient moins âgés que les conjoints et les assurés.

Les tableaux des outputs des tests jugés pertinent sont les suivants :

- ANOVA : TR_AGE et SEXE

	df	sum_sq	mean_sq	F	PR(>F)
SEXE	1	59.387623	59.387623	3.557618	0.059274
Residual	342587	5718834.391909	16.693086	NaN	NaN

- ANOVA : TR_AGE et ALD

	df	sum_sq	mean_sq	F	PR(>F)
ALD	1	920558.345689	920558.345689	65725.285087	0
Residual	342589	4798353.672791	14.006152	NaN	NaN

- ANOVA : TR_AGE et TYPE_SOINS

	df	sum_sq	mean_sq	F	PR(>F)
TYPE_SOINS	1	0.888267	0.888267	0.053211	0.817567
Residual	342587	5718868.582569	16.693186	NaN	NaN

Nous concluons donc, que la variable *TR_AGE* ne dépend pas des deux variables *SEXE* et *TYPE_SOINS*. C'est-à-dire que l'âge impacte de la même façon la sinistralité des deux sexes, et qu'il a la même influence sur l'effectif des consommations en soins ambulatoires et celui des soins hospitaliers.

La dépendance entre ALD et *TR_AGE* peut être due au fait que la santé des personnes âgées est généralement plus faible que celle des moins âgées. Ceci est illustré par la courbe suivante représentant la proportion des ALD sinistrés par âge :

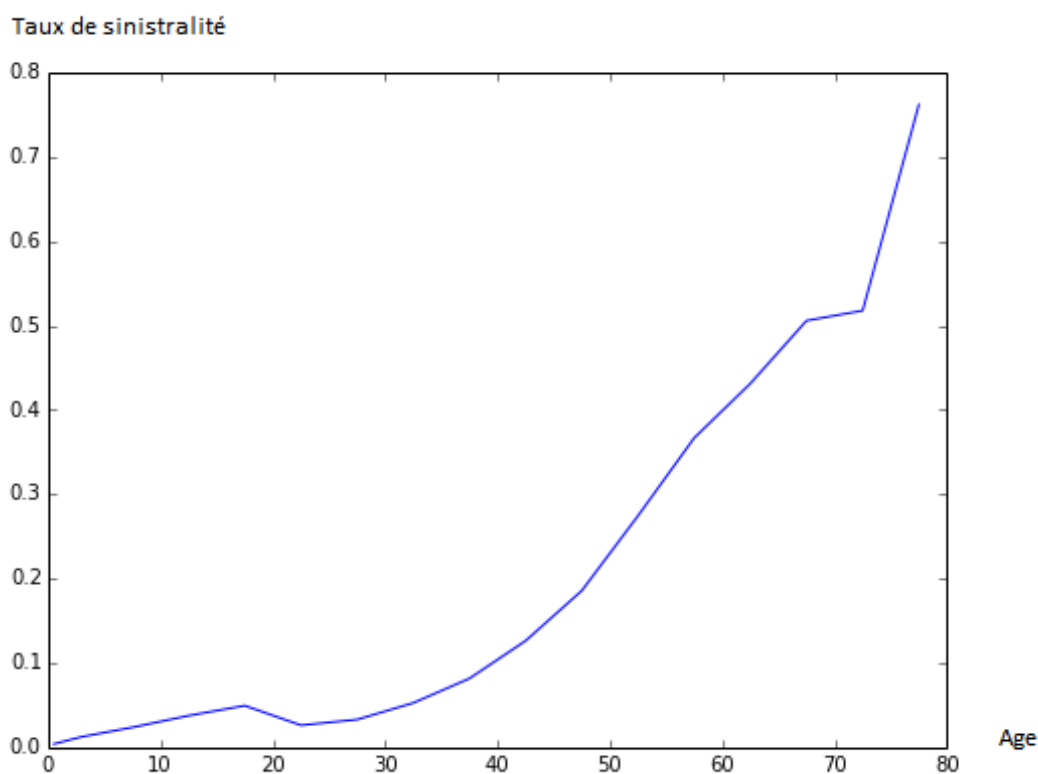


Figure 22 : Taux de sinistralité des bénéficiaires ALD selon l'âge

Remarque : Le test de chi-deux n'est pas approprié dans ce cas pour deux raisons :

1. Il ne prend pas en considération le caractère ordinal des variables.
2. Il ne prend pas en compte la distance entre les modalités des variables intervalles.

Une application du test de Chi-deux pour ce cas aurait donné les résultats ci-dessous qui contredisent les résultats de l'ANOVA:

TR_AGE VS TYPE_SOINS $\chi^2 = 12403.3203141$ p-value = 0.0 DDL = 16	TR_AGE VS SEXE $\chi^2 = 105740.515526$ p-value = 0.0 DDL = 16
---	---

II. Modèle de la sinistralité

Comme nous l'avons vu, nous disposons de données groupées selon des variables relatives au consommateur ou bien au dossier. Nous ne disposons donc pas de données tête-par-tête.

Notre objectif est de pouvoir estimer les montants qui seront remboursés par la CNSS à ces assurés durant les années futures. Ces montants représentent les charges des sinistres à payer, nous allons donc suivre une approche Fréquence/Sévérité, où les fréquences de survenance des sinistres et leurs charges seront étudiés séparément.

II.4. Sinistralité globale

Pour modéliser les sinistres à charge de l'assureur, la littérature ¹ propose deux approches différentes.

La première est celle du Risque Collectif, où la charge globale des sinistres (S) est considérée comme étant la somme des charges de chaque sinistre. Ce modèle est représenté par l'équation suivante :

$$S = X_1 + X_2 + \dots + X_N, \quad N = 0, 1, 2, \dots$$

Où les X_s représentent les charges des sinistres et N représente le nombre global de sinistres de l'année. Ce modèle suppose que la distribution de N ne dépend pas de celle des X_s et que pour $N = n$, les X_s sont indépendants et leur distribution ne dépend pas de n .

Le deuxième modèle est celui du Risque Individuel, qui représente la somme des montants des sinistres à charge de l'assureur comme étant la somme des charges générées par contrat. Le nombre des contrats (n) par année est supposé non-aléatoire. Ce modèle est représenté par l'équation suivante :

$$S = X_1 + X_2 + \dots + X_n$$

Les X_s , supposés indépendants et pas nécessairement identiquement distribués, représentent les charges engendrées par le contrat numéro s durant l'année. Cette charge peut en effet être nulle, ce qui est vrai en réalité pour une grande proportion de contrats d'une compagnie

¹ AGRESTI, Alan. *Categorical Data Analysis*. 2nd ed. Gainesville, Florida : John Wiley & Sons, 2002. ISBN : 978-0-471-36093-3.

d'assurance et donc la fonction de distribution des X_s peut présenter un pic en zéro et nécessitera par conséquent des traitements particuliers.

Par leur spécifications, ces deux modèles nécessitent soit la donnée des charges des sinistres par contrat soit la donnée de la charge de chaque sinistre séparément. Or les données mises à notre disposition par l'organisme sont des données sur la sinistralité de groupes de population.

Nous proposons le modèle suivant qui est une adaptation des deux précédents aux données disponibles :

$$S = a_1X_1 + a_2X_2 + \dots + a_nX_n$$

Où n est le nombre de catégories ou tranches de la population et a_i l'effectif sinistré de la tranche i . Le terme X_i représente ici la sinistralité moyenne d'un individu de la tranche i .

Le modèle considéré renvoi à celui du Risque Collectif par le fait que les a_i sont des variables aléatoires et que leur somme peut être analogue au N , et renvoi aussi au modèle du risque individuel par le fait que l'on s'intéresse à la sinistralité d'un nombre fixé de groupes.

Les X_s sont supposés indépendants ainsi que les a_s et la distribution des X_s est supposée indépendante de celle des effectifs sinistrés. Cette indépendance permet d'établir la relation suivante :

$$\bar{S} = \bar{a}_1\bar{X}_1 + \bar{a}_2\bar{X}_2 + \dots + \bar{a}_n\bar{X}_n$$

Une reformulation de notre objectif est de pouvoir projeter les charges moyennes des sinistres de chaque groupe (\bar{X}_s) ainsi que l'effectif moyen sinistré (\bar{a}_s) de ce groupe et ceci pour pouvoir estimer le montant total des sinistres à charge de la CNSS.

Cependant pour pouvoir faire de telles projections nous devons disposer de données sur plusieurs années (données de panel) ce qui n'est malheureusement pas le cas. Nous allons donc considérer que :

- La proportion sinistrée d'une tranche de la population ne dépend pas du temps (cependant l'effectif de la tranche dépend du temps)
- Les remboursements ne dépendent pas du temps, vu que les remboursements des actes d'hospitalisation s'effectuent sur la base du tarif national de référence qui est relativement stable et que les remboursements relatifs au médicament connaissent une désinflation (-14.56% en 2011 et -1.28% en 2012)

Ces hypothèses sont très simplistes et ne reflètent pas la réalité des affiliés de l'AMO. En effet ce secteur est en plein développement au Maroc. De nouvelles catégories d'affiliés pourraient être intégrées au régime AMO dans le future proche (les travailleurs indépendants à titre d'exemple), ce qui impactera fort probablement le comportement global des affiliés de l'AMO. De plus les actes médicaux et les médicaments sujets aux remboursements connaissent un constant élargissement, ce qui pourrait modifier le comportement de l'assuré face à la consommation médicale (exemple : l'effet d'appel). Ceci étant, pour minimiser l'impact de ces phénomènes, nous ne dépasserons pas les cinq années lors des prévisions (prévisions à moyen terme).

II.5. Montant moyen remboursé (Sévérité)

Le montant moyen remboursé d'une tranche de la population est calculé comme étant le remboursement global de la classe divisé par l'effectif sinistré.

Pour les catégories de la population où les effectifs sinistrés sont assez grands (supérieurs à 30), ces moyennes peuvent être directement extraites de la base de données et être considérées comme des estimations du montant remboursé à un individu de la catégorie et cela en vertu de la loi des grands nombres. Cependant, pour les catégories où la sinistralité est faible ou pour les catégories n'ayant pas présenté de sinistralité (ou pas encore intégrées par la couverture médicale) une estimation directe n'est pas possible. Nous proposons donc d'ajuster des modèles économétriques aux montants moyens remboursés des catégories à grande sinistralité et utiliser les modèles les plus appropriés pour expliquer les remboursements relatifs aux autres catégories.

Les modèles économétriques considérés sont les suivants :

- La régression linéaire multiple.
- Les modèles linéaires généralisés (GLM).
- Les modèles linéaires à densité mélange.

II.5.1. Rappel sur la régression linéaire multiple

1) Principe :

La régression linéaire multiple est une méthode statistique qui cherche à expliquer les valeurs prises par une variable Y , dite endogène à l'aide d'un nombre p de variables X_j , ($j = 1, \dots, p$), dites exogènes.

L'équation de régression s'écrit :

$$y_i = a_0 + a_1x_{i,1} + \dots + a_px_{i,p} + \epsilon_i$$

Avec :

- $i = 1, \dots, n$: indice représentant la $i^{\text{ème}}$ observation, avec $n \geq p$;
- y_i : la $i^{\text{ème}}$ observation de la variable Y ;
- $x_{i,j}$: la $i^{\text{ème}}$ observation de la variable X_j .
- ϵ_i : composante aléatoire représentant l'information omise par le modèle linéaire.

La régression linéaire repose sur les hypothèses suivantes :

- H1.** Les X_j sont non aléatoires ;
- H2.** $E[\epsilon_i] = 0$, l'espérance de l'erreur est nulle ;
- H3.** $E[\epsilon_i^2] = \sigma_\epsilon^2$, la variance de l'erreur est constante ;
- H4.** $COV(\epsilon_i, x_{i,j}) = 0$, l'hypothèse de l'homoscédasticité.
- H5.** $COV(\epsilon_i, \epsilon_{i'}) = 0$ pour $i \neq i'$, l'hypothèse de la non-autocorrélation des résidus ;

- H6.** $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon)$, l'hypothèse de la normalité des erreurs ;
- H7.** La matrice $X'X$ est régulière;
- H8.** $\frac{X'X}{n}$ tend vers une matrice finie non singulière lorsque n tends vers l'infini ;
- H9.** $n > p + 1$, le nombre d'observations est strictement supérieur au nombre des paramètres à estimer;

La méthode d'estimation des paramètres la plus utilisée est celle des Moindres Carrée Ordinaire (MCO). Le principe est de trouver les paramètres (a_0, a_1, \dots, a_p) qui minimisent la quantité suivante :

$$S = \sum_{i=1}^n (y_i - a_0 + a_1 x_{i,1} + \dots + a_p x_{i,p})^2 = \sum_{i=1}^n \epsilon_i^2$$

2) Écriture matricielle :

En adoptant une écriture matricielle, l'équation de régression devient :

$$Y = Xa + \epsilon$$

Avec :

$$\begin{aligned} \bullet \quad X &= \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,p} \end{pmatrix} & \bullet \quad a &= \begin{pmatrix} a_0 \\ \vdots \\ a_p \end{pmatrix} \\ \bullet \quad Y &= \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} & \bullet \quad \epsilon &= \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \end{aligned}$$

Et donc nous avons :

$$S = \epsilon' \epsilon = (Y - Xa)'(Y - Xa)$$

En annulant la dérivée matricielle de S par rapport à a , on retrouve :

$$\frac{\partial S}{\partial a} = -2(X'Y) + 2(X'X)a = 0$$

D'où :

$$(X'X)a = X'Y$$

L'estimateur des moindres carrés ordinaires est donc :

$$\hat{a} = (X'X)^{-1}X'Y$$

3) Pondération

La régression linéaire multiple considère que toutes les observations ont la même importance, ce qui n'est pas toujours le cas notamment quand :

- Chaque observation est retrouvée plusieurs fois.
- Les observations des variables exogènes sont des moyennes sur des populations de tailles différentes.

Pour remédier à ces situations, on affecte une pondération (ω_i) à chaque observation. On parle alors de Régression Linéaire Pondérée.

Dans ce cas on cherche à minimiser la grandeur suivante :

$$S = \sum_{i=1}^n \omega_i (y_i - a_0 + a_1 x_{i,1} + \dots + a_p x_{i,p})^2 = (Y - Xa)'W(Y - Xa)$$

Avec :

$$\bullet \quad W = \begin{pmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & \omega_n \end{pmatrix}$$

On peut remarquer que :

$$S = (W^{1/2}Y - W^{1/2}Xa)'(W^{1/2}Y - W^{1/2}Xa)$$

En posant $W^{1/2}Y = Y_*$ et $W^{1/2}X = X_*$ nous avons :

$$S = (Y_* - X_*a)'(Y_* - X_*a)$$

Comme établi plus tard, l'estimateur des moindres carrés est :

$$\hat{a} = (X_*'X_*)^{-1}X_*'Y_* = (X'WX)^{-1}X'WY$$

4) Variables qualitatives

Dans une régression linéaire, les variables exogènes peuvent être catégorielles. Dans ce cas on doit recoder ces variables à l'aide de variables binaires.

On procède généralement comme suit : Soit C une variable qualitative à k modalités, on définit $k - 1$ variables binaires, chacune relative à une des modalités de C . Pour chaque observation, si C prend une modalité représentée par l'une des variables binaires, cette variable sera la seule à être non-nulle, sinon toutes les variables binaires seront nulles.

Autrement dit, une variable qualitative à k modalités est représentée par un système de $k - 1$ variables binaires prenant k états.

◆ Exemple :

Soit C une variable catégorielle à trois modalités : (M_1, M_2, M_3) . Nous disposons des trois observations suivantes :

Observation	C
1	M_2
2	M_3
3	M_1

Ces observations peuvent être recodées comme suit :

Observation	δ^1	δ^2
1	0	1
2	0	0
3	1	0

Où δ^1 représente la modalité M_1 et δ^2 représente la modalité M_2 .

Notation : Soit x et y deux variables quantitatives et f une fonction réelle. L'équation du modèle de régression linéaire multiple expliquant $z = f(y)$ à l'aide des variables x et C est la suivante :

$$f(y_i) = a_0 + a_1\delta^1 + a_2\delta^2 + a_3x + \epsilon_i$$

Ce modèle sera noté dans ce qui suit par l'équation symbolique :

$$f(y) = C + x$$

II.5.2. Application

Appliquons un modèle de régression linéaire pondérée pour expliquer le montant moyen remboursé à l'aide des inputs de la base de données.

L'outil de travail est le module *Statsmodels* qui contient la fonction « *wls()* », abréviation de *Weighted Least Squares* qui signifie *Moindres Carré Pondérée*. A cette fonction sont spécifiés : la base de données, les variables exogènes, la variable endogène et la variable représentant les poids de chaque observation.

Nous allons tout d'abord commencer par ajuster le modèle le plus naturel, celui représenté par l'équation symbolique suivante :

$$MONT_REM_M = SEXE + SECTEUR_ETABL + ALD + TYPE_SOINS + TR_AGE + TYPE_ASS + TYPE_BEN + REGION$$

La sortie fournie par la fonction « *wls()* » est la suivante :

WLS Regression Results						
Dep. Variable:	MONT_REM_M		R-squared:	0.367		
Model:	WLS		Adj. R-squared:	0.366		
Method:	Least Squares		F-statistic:	222.6		
Date:	Fri, 06 Jun 2014		Prob (F-statistic):	0.00		
Time:	23:55:24		Log-Likelihood:	-1.2932e+05		
No. Observations:	10758		AIC:	2.587e+05		
Df Residuals:	10729		BIC:	2.589e+05		
Df Model:		28				
	coef	std err	t	P> t	[95.0% Conf. Int.]	
Intercept	624.3010	1447.925	0.431	0.666	-2213.900	3462.502
SEXE[T.M]	128.7949	113.519	1.135	0.257	-93.724	351.314
SECTEUR_ETABL[T.PU]	-1134.9152	145.272	-7.812	0.000	-1419.675	-850.155
ALD[T.O]	5913.6324	140.417	42.115	0.000	5638.389	6188.876
TYPE_SOINS[T.H]	9207.6900	152.070	60.549	0.000	8909.604	9505.776
TYPE_ASS[T.I]	-272.3051	660.331	-0.412	0.680	-1566.677	1022.067
TYPE_ASS[T.P1]	-191.6470	175.956	-1.089	0.276	-536.553	153.259
TYPE_ASS[T.P2]	-499.4967	252.480	-1.978	0.048	-994.405	-4.588
TYPE_ASS[T.P3]	-1024.9809	4982.083	-0.206	0.837	-1.08e+04	8740.824
TYPE_BEN[T.CA]	-366.1020	132.526	-2.762	0.006	-625.877	-106.327
TYPE_BEN[T.CM]	426.7751	1581.470	0.270	0.787	-2673.198	3526.748
TYPE_BEN[T.EA]	-181.2768	223.232	-0.812	0.417	-618.852	256.299
TYPE_BEN[T.EM]	-211.4671	6811.356	-0.031	0.975	-1.36e+04	1.31e+04
REGION[T.02]	-111.9938	1613.695	-0.069	0.945	-3275.135	3051.147
REGION[T.03]	-268.7762	1633.593	-0.165	0.869	-3470.920	2933.368
REGION[T.04]	-504.1214	1432.868	-0.352	0.725	-3312.808	2304.565
REGION[T.05]	-670.3098	1440.193	-0.465	0.642	-3493.354	2152.735
REGION[T.06]	-505.0530	1444.690	-0.350	0.727	-3336.913	2326.807
REGION[T.07]	-387.9078	1437.092	-0.270	0.787	-3204.875	2429.059
REGION[T.08]	-572.7889	1439.623	-0.398	0.691	-3394.717	2249.139
REGION[T.09]	-469.5606	1427.016	-0.329	0.742	-3266.775	2327.654
REGION[T.10]	-434.5702	1431.466	-0.304	0.761	-3240.509	2371.368
REGION[T.11]	-457.9788	1441.468	-0.318	0.751	-3283.522	2367.564
REGION[T.12]	-273.7021	1473.224	-0.186	0.853	-3161.493	2614.089
REGION[T.13]	-473.6297	1440.919	-0.329	0.742	-3298.097	2350.837
REGION[T.14]	-533.9511	1440.360	-0.371	0.711	-3357.323	2289.421
REGION[T.15]	-370.3900	1473.484	-0.251	0.802	-3258.692	2517.912
REGION[T.16]	-604.7900	1434.948	-0.421	0.673	-3417.554	2207.974
TR_AGE	18.7762	26.850	0.699	0.484	-33.854	71.406
Omnibus:	10267.263		Durbin-Watson:	1.809		
Prob(Omnibus):	0.000		Jarque-Bera (JB):	884534.978		
Skew:	4.402		Prob(JB):	0.00		
Kurtosis:	46.541		Cond. No.	1.51e+03		

Figure 23 : Sortie de la régression linéaire pondérée du modèle simple des remboursements

L'ajustement est faible ($R^2 = 36,7\%$) ce modèle n'est donc pas adapté.

Nous pensons donc à inclure les variables d'interaction. Néanmoins, nous constatons que l'ajustement du modèle est toujours faible ($R^2 = 43.9\%$).

Vu que toutes les variables exogènes sont corrélées entre elles et que le montant moyen remboursé par classe est une grandeur positive, on peut penser à un modèle log-linéaire, où le logarithme de la variable endogène est une fonction linéaire des variables exogènes.

L'équation symbolique de ce modèle est la suivante :

$$\ln(MONT_REM_M) = SEXE + SECTEUR_ETABL + ALD + TYPE_SOINS + TR_AGE + TYPE_ASS + TYPE_BEN + REGION$$

La sortie relative à ce modèle est la suivante :

WLS Regression Results						
Dep. Variable:	log(MONT_REM_M)	R-squared:	0.931			
Model:	WLS	Adj. R-squared:	0.931			
Method:	Least Squares	F-statistic:	5194.			
Date:	Sat, 07 Jun 2014	Prob (F-statistic):	0.00			
Time:	00:03:27	Log-Likelihood:	-23614.			
No. Observations:	10758	AIC:	4.729e+04			
Df Residuals:	10729	BIC:	4.750e+04			
Df Model:		28				
	coef	std err	t	P> t	[95.0% Conf. Int.]	
Intercept	5.8978	0.078	75.370	0.000	5.744	6.051
SEXE[T.M]	-0.0882	0.006	-14.373	0.000	-0.100	-0.076
SECTEUR_ETABL[T.PU]	-0.6293	0.008	-80.151	0.000	-0.645	-0.614
ALD[T.O]	1.4775	0.008	194.702	0.000	1.463	1.492
TYPE_SOINS[T.H]	2.0218	0.008	246.002	0.000	2.006	2.038
TYPE_ASS[T.I]	-0.1936	0.036	-5.424	0.000	-0.264	-0.124
TYPE_ASS[T.P1]	-0.0038	0.010	-0.404	0.686	-0.022	0.015
TYPE_ASS[T.P2]	-0.1095	0.014	-8.022	0.000	-0.136	-0.083
TYPE_ASS[T.P3]	-0.5187	0.269	-1.927	0.054	-1.047	0.009
TYPE_BEN[T.CA]	-0.1301	0.007	-18.159	0.000	-0.144	-0.116
TYPE_BEN[T.CM]	-0.0763	0.085	-0.892	0.372	-0.244	0.091
TYPE_BEN[T.EA]	-0.1879	0.012	-15.574	0.000	-0.212	-0.164
TYPE_BEN[T.EM]	-0.3988	0.368	-1.083	0.279	-1.120	0.323
REGION[T.02]	0.0436	0.087	0.500	0.617	-0.127	0.215
REGION[T.03]	0.0236	0.088	0.267	0.790	-0.149	0.197
REGION[T.04]	0.1318	0.077	1.702	0.089	-0.020	0.284
REGION[T.05]	0.1383	0.078	1.776	0.076	-0.014	0.291
REGION[T.06]	0.1148	0.078	1.470	0.142	-0.038	0.268
REGION[T.07]	0.2052	0.078	2.642	0.008	0.053	0.357
REGION[T.08]	0.1763	0.078	2.265	0.024	0.024	0.329
REGION[T.09]	0.3421	0.077	4.436	0.000	0.191	0.493
REGION[T.10]	0.2718	0.077	3.514	0.000	0.120	0.423
REGION[T.11]	0.1489	0.078	1.912	0.056	-0.004	0.302
REGION[T.12]	0.1035	0.080	1.300	0.194	-0.053	0.260
REGION[T.13]	0.1592	0.078	2.044	0.041	0.007	0.312
REGION[T.14]	0.2565	0.078	3.295	0.001	0.104	0.409
REGION[T.15]	0.1361	0.080	1.709	0.087	-0.020	0.292
REGION[T.16]	0.1296	0.078	1.671	0.095	-0.022	0.282
TR_AGE	0.0399	0.001	27.518	0.000	0.037	0.043
Omnibus:	1565.523	Durbin-Watson:	1.854			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	17569.518			
Skew:	0.317	Prob(JB):	0.00			
Kurtosis:	9.228	Cond. No.:	1.51e+03			

Figure 24 : Sortie de la régression linéaire pondérée du modèle logarithmique des remboursements

Pour le modèle logarithmique, l'ajustement est bon ($R^2 = 93.1\%$). Cependant les p-values des coefficients de ce modèle n'ont pas de sens, car le test de Jarque-Bera de normalité indique que les résidus ne proviennent pas d'une loi normale. Nous devons donc examiner les résidus du modèle plus en détail.

Le diagramme quantile-quantile (QQplot) comparant les résidus à une distribution normale est le suivant :

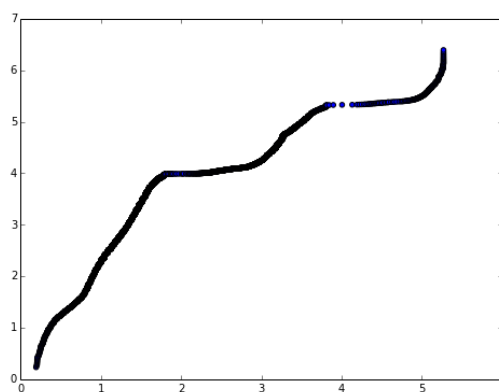


Figure 25 : QQ-plot des résidus du modèle logarithmique et de la loi normale

Nous constatons à partir du diagramme ci-dessus que les résidus ne s'ajustent pas à la loi normale.

Avant de conclure que le modèle de régression linéaire est inapproprié à la distribution du montant moyen remboursé, nous remarquons que le diagramme des résidus ci-dessous, concernant le modèle sans transformation logarithmique, présente une forme atypique.

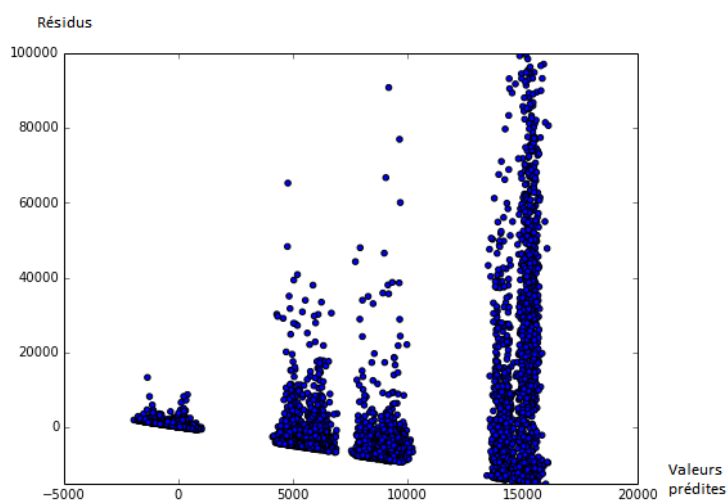


Figure 26 : Résidus du modèle simple de régression linéaire pondérée

Les résidus présentent donc une certaine discontinuité dans leur distribution. Ceci est généralement lié à la non-homogénéité des observations. Une solution à ce problème est d'ajuster autant de modèles que de sous-populations, quatre modèles sont donc nécessaires dans notre cas.

Il est donc nécessaire d'identifier les quatre sous-ensembles d'observations. Pour cela nous allons utiliser une méthode graphique qui consiste à donner des couleurs différentes aux résidus selon les modalités de l'une des variables explicatives et voir si une des couleurs coïncide avec l'une des catégories.

Le graphe suivant représente les résidus en affectant des couleurs différentes pour chaque modalité de la variable *SECTEUR_ETABL* :

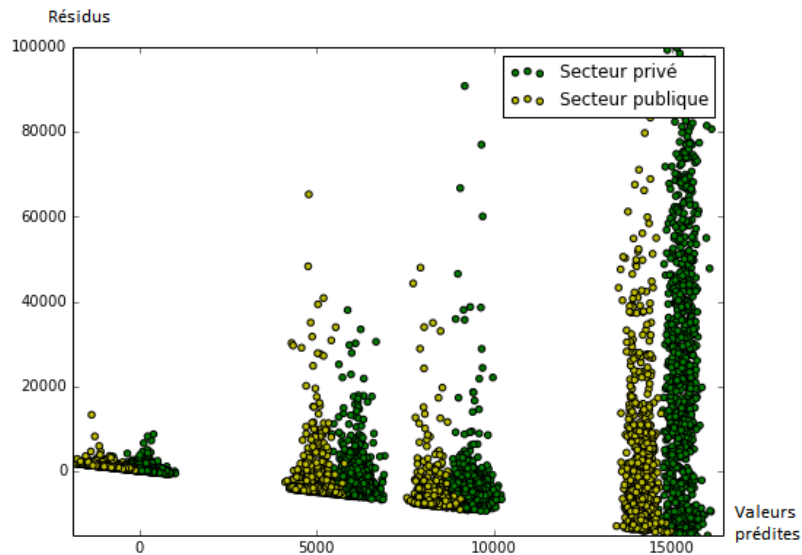


Figure 27 : Résidus du modèle simple de régression linéaire pondérée selon le secteur d'établissement

Nous constatons que l'appartenance à l'une des quatre catégories n'est pas influencée par le secteur de l'établissement de soins.

On reconduit la même démarche concernant la variable « ALD », nous retrouvons le graphique suivant :

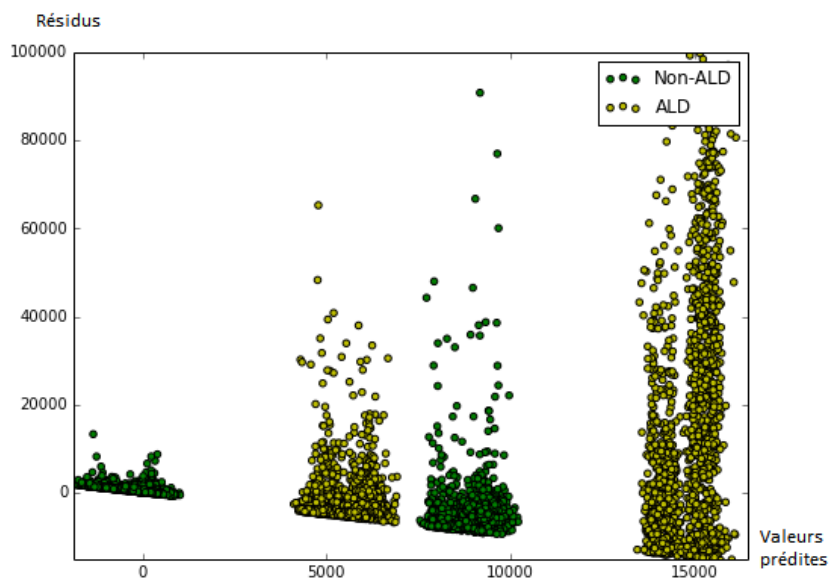


Figure 28 : Résidus du modèle simple de régression linéaire pondérée selon l'état de santé du bénéficiaire

Il est clair, à travers la figure ci-dessus, que la variable ALD est l'une des sources d'hétérogénéité de la population. En effet, à partir de la gauche, le premier et troisième groupe de sous-observations concernent les bénéficiaires atteints d'affections de longues durées.

N'ayant pas complètement distingué entre les sous-groupes d'observations, nous redressons le diagramme des résidus en distinguant entre les observations relatives aux hospitalisations et ceux relatives aux soins ambulatoires. La figure obtenue est la suivante :

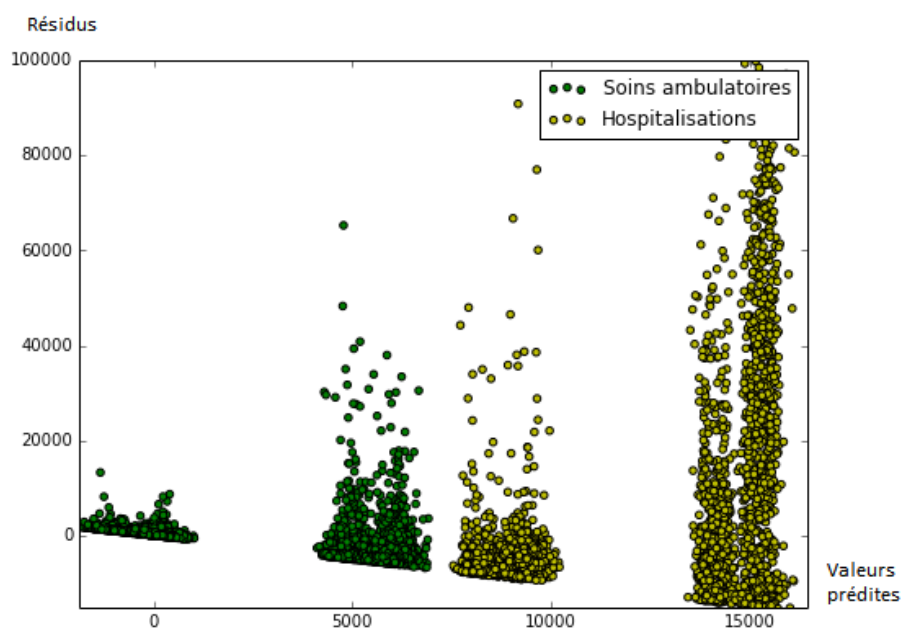


Figure 29 : Résidus du modèle simple de régression linéaire pondérée selon le type de soins

Nous nous apercevons donc, à partir de la gauche, que les deux premiers sous-ensembles d'observations sont relatives aux soins ambulatoires. Nous concluons alors que les quatre sous-populations non-homogènes sont :

- Les remboursements des frais de soins ambulatoires relatifs aux bénéficiaires non-ALD
- Les remboursements des frais de soins ambulatoires relatifs aux bénéficiaires ALD
- Les remboursements des frais de soins hospitaliers relatifs aux bénéficiaires non-ALD
- Les remboursements des frais de soins hospitaliers relatifs aux bénéficiaires ALD

Les sous-populations étant identifiées, nous ajustons les modèles définis antérieurement pour chacune d'elle. Les diagrammes des résidus de ces ajustements, ne présentent pas de discontinuité. Cependant, l'hypothèse de normalité des résidus n'est vérifiée pour aucun des ajustements (Les sorties figurent en annexe).

◆ Conclusion

Plusieurs difficultés ont été rencontrées en ce qui concerne l'application du modèle de régression linéaire pondérée au montant moyen remboursé par tête, notamment :

- Non-linéarité de la variable endogène par rapport aux variables exogènes. Ce qui a été contourné par une transformation logarithmique de la variable endogène.
- Non-homogénéité de la population (discontinuité des résidus). Ce qui a été contourné par l'identification des sous-populations hétérogènes.
- Non-normalité des résidus du modèle de régression linéaire appliqué aux sous-populations identifiés.

Face à l'inadéquation des données au modèle de la régression linéaire, il est naturel de se tourner vers un modèle moins restrictif en termes d'hypothèses et plus approprié à la distribution des données.

Dans ce qui suit nous allons présenter le cadre théorique des Modèles Linéaires Généralisés (GLM), une généralisation souple de la régression linéaire, puis essayer d'appliquer de tels modèles aux montants moyens remboursés.

II.5.3. Rappel sur les GLM :

La dénomination « Modèles Linéaires Généralisés » (à ne pas confondre avec le modèle linéaire général) revient à Nelder and Wedderburn (1972) et à McCullagh and Nelder (1989) qui ont montré que les démarches de modélisation de la moyenne μ , d'une variable aléatoire Y , à l'aide d'une combinaison linéaire de variables endogènes $\mathbf{X}'\beta$, sont similaires quand la distribution de Y appartient à la famille exponentielle. Ainsi, ils ont regroupé ces démarches sous le nom des Modèles Linéaires Généralisés.

L'objectif est de modéliser la moyenne $\mu = E(Y|\mathbf{X})$ à l'aide d'une combinaison de variables endogènes $\mathbf{X}'\beta$.

Les GLM n'imposent pas de relation linéaire directe entre μ et \mathbf{X} , mais laissent introduire une fonction de *lien* G telle que :

$$\eta = G(\mu) = \mathbf{X}'\beta$$

Pour chacune des distributions de la famille exponentielle il existe au moins une fonction lien dite *canonique* qui simplifie la procédure d'estimation.

1) La famille exponentielle

On dit qu'une loi de probabilité appartient à la famille exponentielle si sa distribution s'écrit sous la forme :

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

Les fonctions a , b et c varient d'une distribution à une autre.

Notre paramètre d'intérêt est θ , vu que, pour la famille exponentielle, il existe un lien direct entre θ et μ .

Considérons l'exemple de la loi de Bernoulli dont la densité, pour $y \in \{0,1\}$, s'écrit :

$$f(y) = \mu^y (1 - \mu)^{1-y}$$

Sachant que $\mu^y = e^{y \ln(\mu)}$, on peut écrire :

$$f(y) = \exp \left(y \times \ln \left(\frac{\mu}{1 - \mu} \right) + \ln(1 - \mu) \right)$$

Par identification :

$$\theta = \ln \left(\frac{\mu}{1 - \mu} \right) \quad ; \quad a(\psi) = 1 \quad ; \quad c(y, \psi) = \ln(1 - \mu)$$

Donc la loi de Bernoulli appartient à la famille exponentielle.

2) La fonction de lien canonique

Pour une loi de probabilité appartenant à la famille exponentielle, la fonction de lien canonique G est celle qui permet d'écrire :

$$\eta = G(\mu) = \theta$$

Cette fonction de lien permet de simplifier la procédure d'estimation.

En guise d'exemple, pour la loi de Bernoulli, la fonction de lien canonique est la fonction Logit : $x \mapsto \ln\left(\frac{x}{1-x}\right)$.

3) Estimation

L'appartenance de la distribution de Y à la famille exponentielle permet de calculer les estimateurs de maximum de vraisemblance des paramètres des modèles linéaires généralisés.

On admet les deux résultats suivants relatifs aux lois de la famille exponentielle :

$$E(Y) = \mu = b'(\theta)$$

Et :

$$Var(Y) = V(\mu)a(\psi) = b''(\theta)a(\psi)$$

Comme leur nom l'indique, les estimateurs de maximum de vraisemblance reposent sur la maximisation de la fonction de vraisemblance ou celle du log-vraisemblance de l'échantillon étudié.

Pour des observations (Y_i, \mathbf{X}_i) , où $i = 1, \dots, n$, la fonction de log-vraisemblance s'écrit :

$$l(\mathbf{Y}, \boldsymbol{\mu}, \psi) = \sum_{i=1}^n l(Y_i, \mu_i, \psi) = \sum_{i=1}^n \left\{ \frac{Y_i \theta_i - b(\theta_i)}{a(\psi)} + c(Y_i, \psi) \right\}$$

Maximiser cette grandeur revient à maximiser :

$$\sum_{i=1}^n \{Y_i \theta_i - b(\theta_i)\}$$

On définit la déviance comme étant :

$$D = -2 \sum_{i=1}^n \{Y_i \theta_i - b(\theta_i)\}$$

La littérature propose généralement de minimiser la déviance, ce qui est équivalent à la maximisation de la vraisemblance

Pour minimiser D nous introduisons son gradient comme suit :

$$\nabla(\beta) = \frac{\partial}{\partial \beta} \left[-2 \sum_{i=1}^n \{Y_i \theta_i - b(\theta_i)\} \right] = -2 \sum_{i=1}^n \{Y_i - b'(\theta_i)\} \frac{\partial}{\partial \beta} \theta_i$$

Notre objectif est de résoudre :

$$\nabla(\beta) = 0$$

Généralement, résoudre ce problème revient à résoudre un système d'équations non-linéaires en β . Une approche itérative est donc nécessaire. La méthode généralement utilisée est la méthode IRWLS (Iteratively reweighted least squares) qui repose sur l'algorithme de Newton-Raphson. Il est à noter que l'IRWLS est implémentée dans la plus part des logiciels statistiques, et donc il n'est pas nécessaire de la programmer étape-par-étape.

II.5.4. Application

Lors de l'application des GLMs, nous allons faire la distinction entre les données relatives aux différentes sous-populations hétérogènes identifiées antérieurement.

1) Remboursement sur la consommation ambulatoire (médicaments)

Comme mentionné dans ce qui précède, il est nécessaire de spécifier la distribution de la variable endogène.

Plusieurs outils informatiques permettent d'ajuster des données aux différentes lois de probabilité et de juger de la pertinence de l'ajustement par des diagnostics graphiques où bien des tests d'hypothèse. Cependant il nous a été impossible de trouver un tel outil adapté aux données groupées, c'est-à-dire un outil qui permet d'ajuster les données pondérées.

On pourrait suggérer de dupliquer chaque ligne de la base de donnée selon le nombre de sinistres s'y relatant, cependant vu le nombre de sinistres, cette démarche est lourde en termes du temps de calcul (plus de 5 heures pour les données relatives aux sinistres hospitaliers représentant moins 6% de l'effectif globale des sinistres). Pour contourner un tel obstacle, nous proposons de calculer les quantiles de ces données groupées et d'utiliser le package « riskDistributions » du logiciel statistique **R** qui permet d'ajuster une série de données à une loi de probabilité selon ces quantiles.

Pour juger de la pertinence de l'ajustement nous nous baserons sur le qq-plot (diagramme quantiles-quantiles) des données réelles contre ceux de la distribution théorique.

Un récapitulatif de la démarche que nous allons suivre est le suivant :

- 1- Définir le sous ensemble de données d'intérêt
- 2- Calculer des quantiles de la variable *MONT_REM_M* s'y relatant (en prenant en compte la pondération des données)
- 3- Ajuster une distribution à ces quantiles à l'aide du package « riskDistribution »
- 4- Dresser le qq-plot de la distribution réelle contre la distribution théorique
- 5- Conclure quant à l'ajustement

Les montants moyens remboursés sont positifs et nous devons les ajuster à l'aide de l'une des lois exponentielle, les lois candidates sont donc : Gamma, Exponentielle, Weibull et Log-normale.

Si les données ne s'ajustent à aucune de ces lois, le modèle GLM n'est donc pas utilisable.

En considérant l'ensemble des montants remboursés des frais des consommations des soins ambulatoires, les données ne s'ajustent à aucune des lois candidates. Ceci peut être dû à l'hétérogénéité de la population décrite lors de l'ajustement au modèle de régression linéaire. Nous allons donc distinguer entre le remboursement relatifs aux soins ambulatoires concernant les bénéficiaires ALD et ceux relatif aux non-ALD.

En conduisant la démarche décrite antérieurement, nous concluons que les données relatives aux personnes non-ALD s’ajustent à une loi gamma, tandis que ceux relatives aux personnes ALD s’ajustent à une loi log-normale. Les qq-plots de ces deux ajustements sont les suivants :

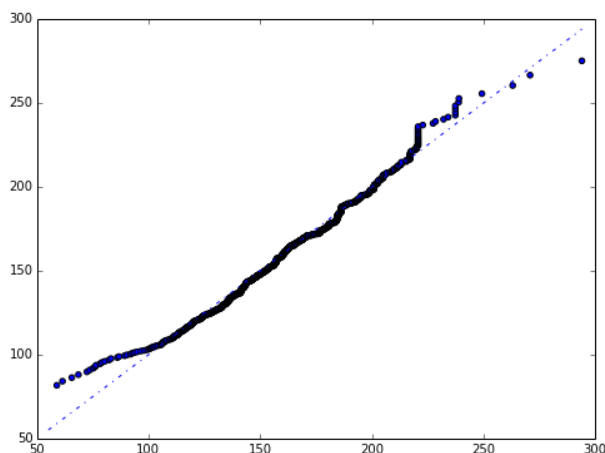


Figure 30 : QQ-plot de l’ajustement des remboursements des soins ambulatoires relatifs aux bénéficiaires non-ALD à la loi gamma

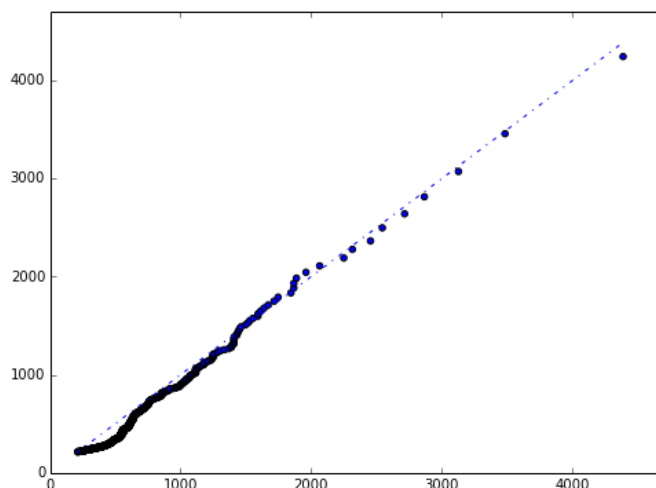


Figure 31 : QQ-plot de l’ajustement des remboursements des soins ambulatoires relatifs aux bénéficiaires non-ALD à la loi log-normale

Après avoir identifié la loi adéquate à la variable endogène, et ce pour chacune des sous populations susdites, nous allons passer à l’estimation des coefficients du modèle GLM.

Pour ajuster les deux modèles de la consommation médicale ambulatoire nous utiliserons le logiciel SAS, premièrement parce que la librairie *STATSMODELS* de *Python* ne permet pas d’ajuster un modèle GLM en pondérant les observations et deuxièmement parce que le logiciel de statistique *R* ne permet pas d’utiliser la loi log-normale comme distribution de la variable endogène.

Nous utiliserons les mêmes variables retenus par le modèle démographique et nous introduirons à chaque fois la variable d’interaction minimisant l’AIC (critère d’information d’Akaike).

Etant donné que les sorties obtenues par SAS sont volumineuses, vu le nombre de modalités des variables qualitatives, nous allons les exposer en annexe.

◆ Ajustement des consommations ambulatoires des non-ALD :

Comme précisé précédemment, nous commençons par ajuster le modèle suivant :

$$MONT_REM_M = TYPE_ASS + TYPE_BEN + SEXE + TR_AGE + REGION$$

Nous retrouvons un AIC de 16506.

En introduisant les variables d'interaction, nous retenons le modèle avec un AIC de 16446 et dont l'équation symbolique est la suivante :

$$MONT_REM_M = TYPE_ASS + TYPE_BEN + SEXE + TR_AGE + REGION + TR_AGE * SEXE + TYPE_ASS * TYPE_BEN + TYPE_BEN * SEXE$$

◆ Ajustement des consommations ambulatoires des ALD :

En reconduisant la même démarche pour les consommations ambulatoires des bénéficiaires atteints d'affections de longues durées, on retient le modèle suivant :

$$MONT_REM_M = TYPE_ASS + TYPE_BEN + SEXE + TR_AGE + REGION + TR_AGE * SEXE$$

L'AIC de ce modèle est de 720 contre 752 pour le modèle sans interactions.

2) Remboursements sur la consommation hospitalière

En essayant de reconduire la même démarche pour les montants remboursés concernant les hospitalisations des bénéficiaires non-ALD, les qq-plot obtenus des différents ajustements aux lois positives de la famille exponentielle sont loin d'être linéaires, comme le montre les qq-plot suivants relatives, respectivement, à la loi log-normale et à la loi gamma :

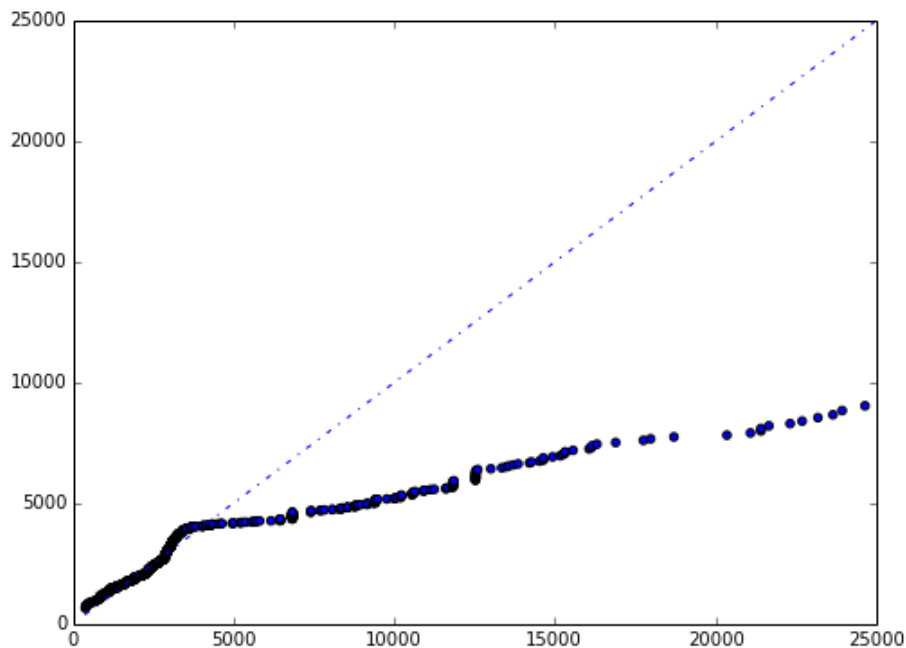


Figure 32 : QQ-plot de l'ajustement du remboursement des soins hospitaliers relatifs aux bénéficiaires non-ALD à la loi gamma

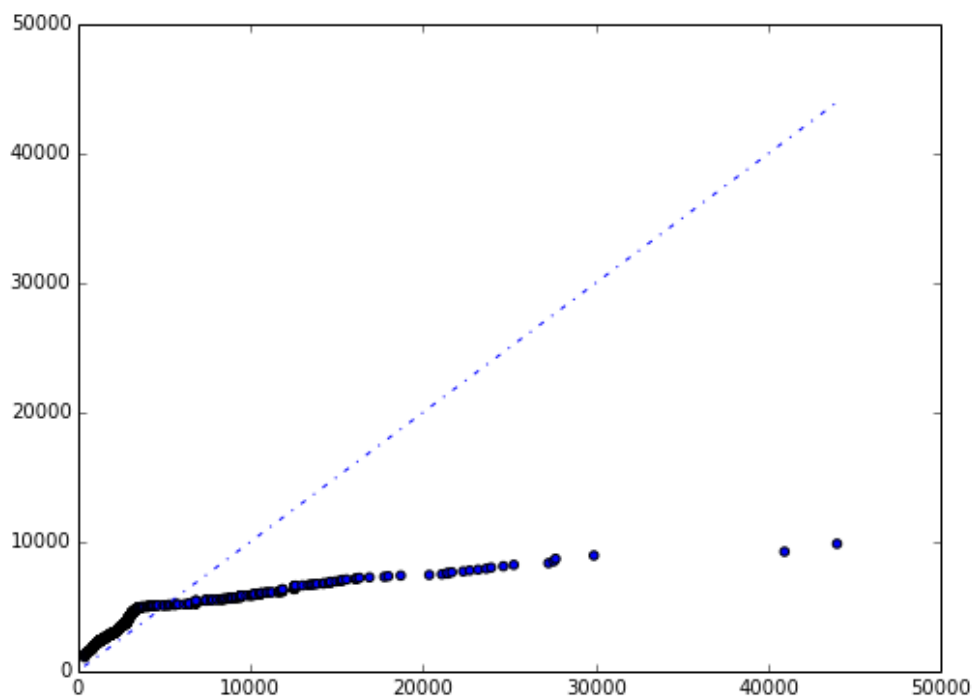


Figure 33 : QQ-plot de l'ajustement du remboursement des soins hospitaliers relatifs aux bénéficiaires non-ALD à la loi log-normale

Nous examinons alors les histogrammes des deux groupes, ALD et non ALD pour se faire une idée sur la forme de la distribution à ajuster. Les histogrammes sont les suivants :

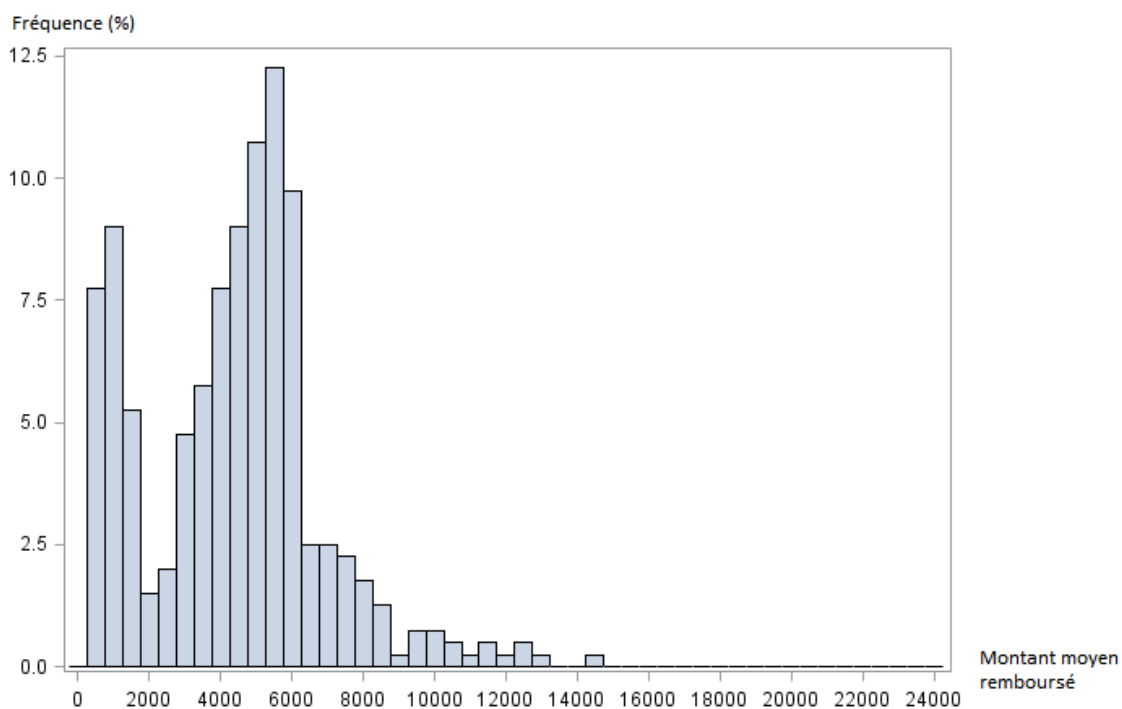


Figure 34 : Histogramme des remboursements des soins hospitaliers relatifs aux bénéficiaires non-ALD

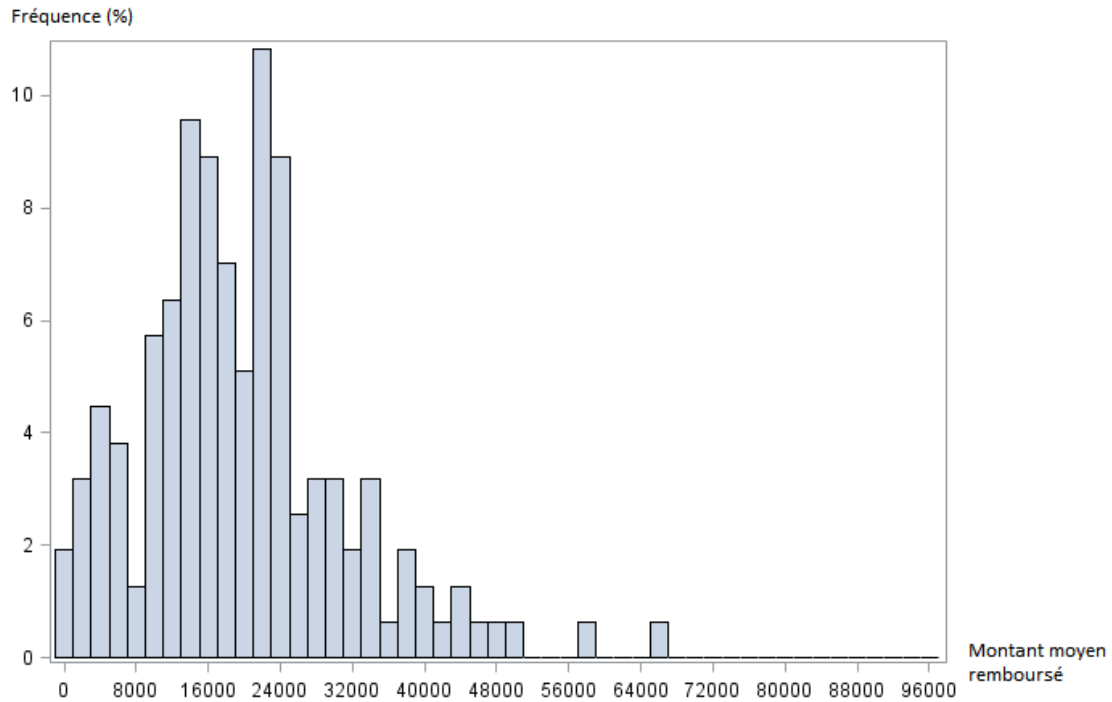


Figure 35 : Histogramme des remboursements des soins hospitaliers relatifs aux bénéficiaires ALD

Ces histogrammes sont largement différents, donc la distinction entre ALD et non ALD est pertinente. De plus nous remarquons que ces deux graphes présentent plus d'un seul pic, ce qui n'est pas le cas des distributions positives de la famille exponentielle.

Les densités appropriés à cette forme de distribution sont connus sous le nom de **densité mélange** ou **loi mélange**, définis comme étant une combinaison convexe de plusieurs fonctions de densité. La construction d'une telle loi se fait comme suit : En prenant une densité $f(x, \theta)$ d'une variable aléatoire X paramétrée par θ , si on considère une famille de paramètres $\theta_1, \theta_2, \dots, \theta_k$ et une famille de scalaires a_1, a_2, \dots, a_k tels que $\sum_{i=1}^k a_i = 1$ alors la fonction définie par $g(x, \theta_1, \theta_2, \dots, \theta_k) = \sum_{i=1}^k a_i f(x, \theta_i)$ est la fonction densité d'une loi mélange à k composantes.

Le logiciel SAS fournit la procédure **fm** qui permet d'ajuster un ensemble de données à de telles lois. En appliquant la procédure à nos données, nous trouvons que les montants remboursés aux bénéficiaires non-ALD proviennent du mélange de cinq variables aléatoires de lois gamma schématisé comme suit :

$$0.0748 \times G(8.6654 ; 2.3506) + 0.0967 \times G(8.6654 ; 2.3506) + 0.3729 \times G(8.6654 ; 2.3506) + 0.2573 \times G(8.6654 ; 2.3506) + 0.1983 \times G(8.6654 ; 2.3506)$$

Où « $G(a, b)$ » représente une variable aléatoire de loi gamma et de paramètres a et b .

Nous obtenons aussi la loi du montant remboursé aux bénéficiaires ALD qui est la suivante :

$$0.0609 \times G(8.3414 ; 0.07771) + 0.1259 \times G(10.0140 ; 6.8705) + 0.1717 \times G(10.3756 ; 0.2167) + 0.1058 \times G(9.3878 ; 0.4869) + 0.1994 \times G(9.7779 ; 0.3417) + 0.3363 \times G(9.8646 ; 0.02119)$$

Les densités de ces deux ajustements, accompagnés par les histogrammes des données réelles sont les suivants :

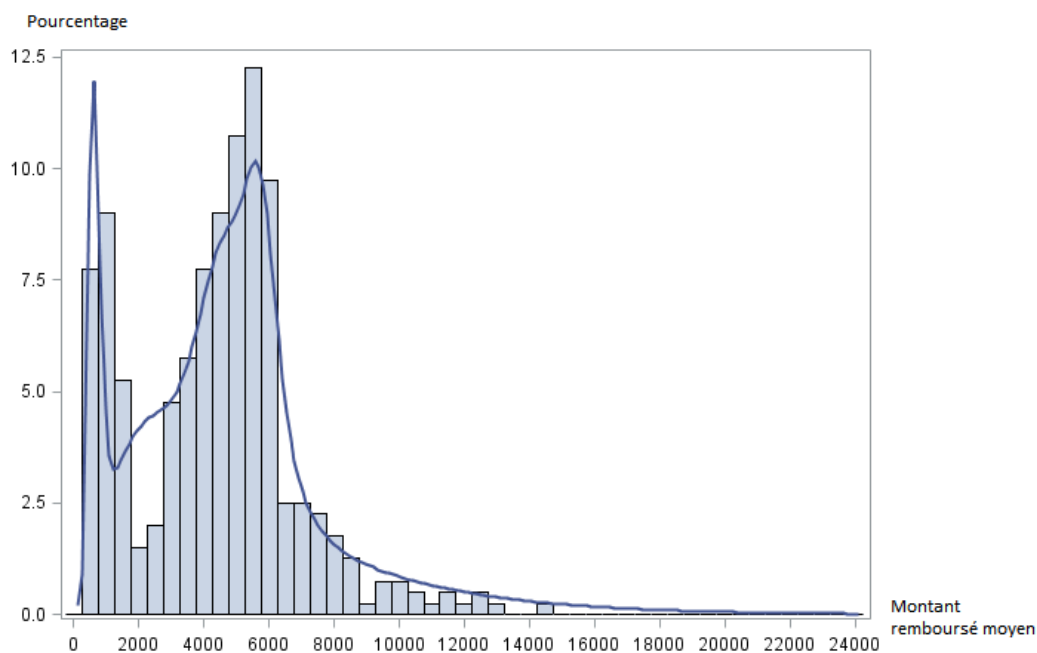


Figure 36 : Histogramme ajusté des remboursements des soins hospitaliers relatifs aux bénéficiaires non-ALD

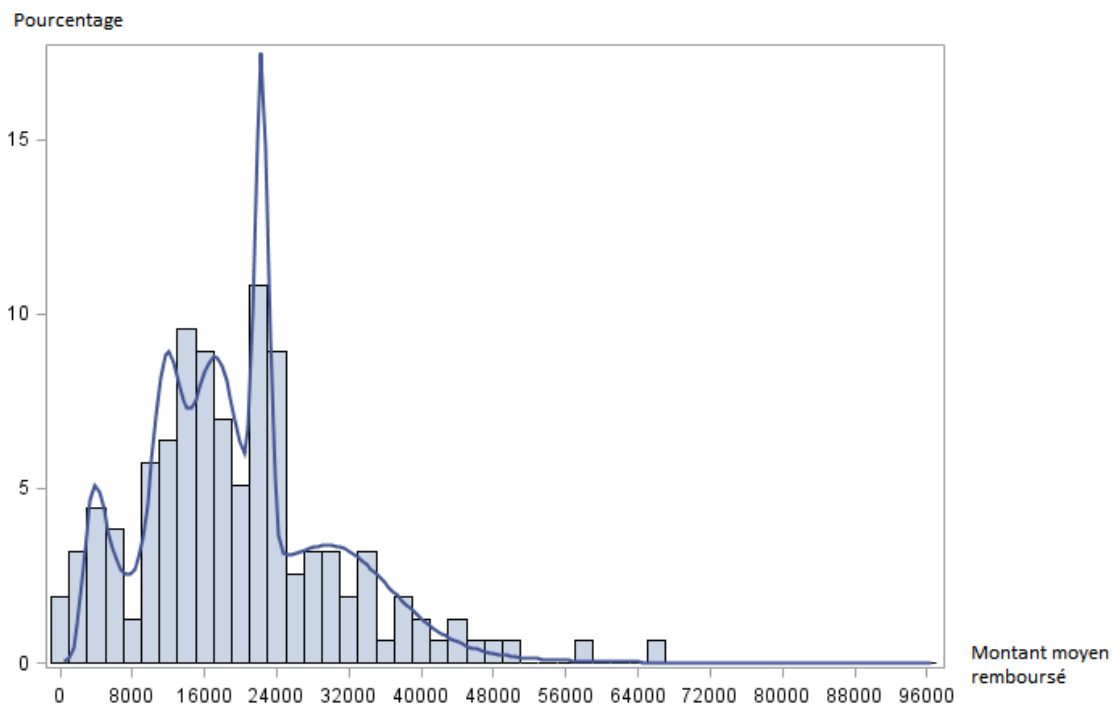


Figure 37 : Histogramme ajusté des remboursements des soins hospitaliers relatifs aux bénéficiaires ALD

On remarque que ces nouvelles distributions réussissent à capter la multitude de pics que présente la distribution de nos données. Cependant le modèle GLM n'est pas adapté à de telles lois de probabilité car elles n'appartiennent pas à la famille exponentielle.

Les modèles appropriés pour les variables endogènes ayant une densité mélange sont connus sous le nom de *Mixture Models*, ce qui peut être traduit par « Modèles à densité mélange ». Ces modèles permettent de capter l'hétérogénéité que peut présenter une population sans pour autant avoir besoin d'identifier sa source.

Dans ce qui suit, nous nous intéresserons au cas où le mélange de loi est fini. Ces modèles portent le nom de *Finite Mixture Models* qu'on notera « FMM ». Nous allons tout d'abord introduire brièvement le cadre théorique des modèles à densité mélange.

II.5.5. Finite Mixture Models

Les modèles à densité mélange (FMM) étendent le cadre des modèles linéaires généralisé au cas où la densité de la variable endogène est une densité mélange de lois exponentielles. Ces densités mélange apparaissent lorsque la population étudiée englobe plusieurs sous-populations hétérogènes.

Supposons que nous avons identifié m sous-populations avec des proportions $\pi_1, \pi_2, \dots, \pi_m$, la variable endogène suit donc une loi à densité mélange à m composantes.

La fonction de densité de Y_j sachant qu'elle appartient à la classe i (supposée de la famille exponentielle) est de la forme suivante :

$$f_j(y_j|\theta_{ij}, \psi_i) = \exp \left\{ \frac{y_j \theta_{ij} - b(\theta_{ij})}{a(\psi_i)} + c(y_j, \psi_i) \right\}$$

Les observations relatives à une sous-population sont considérées indépendantes et le paramètre ψ_i , dit de dispersion, est supposé constant pour chaque sous-population.

Comme dans le cas des modèles linéaires généralisés, nous cherchons à prédire la moyenne (μ_{ij}) de Y_j sachant qu'elle appartient à la sous-population i .

Nous introduisons donc une fonction de lien G telle que :

$$\eta_{ij} = G(\mu_{ij}) = G(b'(\theta_{ij})) = \mathbf{X}'_j \beta_i$$

La probabilité que Y_j appartienne à la classe i étant égale à π_i , la densité de Y_j s'écrit :

$$f_j(y_j|\omega) = \sum_{i=1}^m \pi_i \times f_j(y_j|\theta_{ij}, \psi_i) = \sum_{i=1}^m \pi_i \times \exp \left\{ \frac{y_j \theta_{ij} - b(\theta_{ij})}{a(\psi_i)} + c(y_j, \psi_i) \right\}$$

Où ω représente les paramètres du modèle à estimer.

$$\omega = \begin{pmatrix} \pi_1 & \beta_1 & \psi_1 \\ \vdots & \vdots & \vdots \\ \pi_m & \beta_m & \psi_m \end{pmatrix}$$

L'estimation de ω s'effectue en maximisant la vraisemblance de l'échantillon étudié (de taille n), qui s'écrit :

$$L(y, \omega) = \prod_{j=1}^n f_j(y_j|\omega)$$

Ici encore une approche itérative est nécessaire. L'algorithme de maximisation généralement utilisé est nommé « Algorithme espérance-maximisation » (Expectation-maximisation algorithm) introduit par Dempster, Laird et Rubin (1977), sur lequel se base la procédure **fmm** de SAS.

II.5.6. Application

Après avoir ajusté nos données à des lois de probabilités mélange à l'aide la procédure FMM de SAS, on utilise cette même procédure pour estimer les paramètres de nos modèles. Les sorties complètes figurent en annexe.

Rappelons que les variables exogènes sont : *TYPE_ASS*, *TYPE_BEN*, *SEXE*, *REGION* et *TR_AGE*.

Pour chacune des deux sous populations (ALD et non-ALD) nous commençons par ajuster le modèle simple. En reconduisant la même démarche utilisée pour la sélection des modèles relatives aux consommations ambulatoires (minimisation de l'AIC), nous retenons le modèle suivant pour les ALD :

$$MONT_REM_M = TYPE_ASS + TYPE_BEN + TYPE_ASS * TYPE_BEN + SEXE + TR_AGE + SEXE * TR_AGE$$

Et le modèle suivant pour les non-ALD :

$$MONT_REM_M = TYPE_ASS + TYPE_BEN + TYPE_ASS * TYPE_BEN + SEXE + TR_AGE + REGION$$

Les modèles étant choisis et leurs paramètres estimés, on peut maintenant prédire les remboursements pour les classes d'affiliés qui ne figurent pas dans la base de données et d'avoir des estimations de ces montants pour les classes où l'effectif sinistré est faible, et donc pour lesquelles l'estimation de cette grandeur par la moyenne n'est pas fiable.

La sévérité des sinistres de l'AMO du secteur privé étant modélisée, nous nous intéresserons dans ce qui suit à la fréquence de survenance de ces sinistres

II.6. Effectif sinistré (Fréquence)

Comme indiqué plus tard, la base de données sur la consommation médicale contient la colonne *EFFECTIF_SINISTR* qui représente le nombre de sinistres relatifs à chaque classe de la population.

L'objectif de cette partie est d'étudier la relation entre le nombre de sinistres (variable endogène) et les variables descriptives (variables exogènes) de chaque catégorie et cela pour lisser ses effectifs, c'est-à-dire d'éliminer l'effet aléatoire que contient la variable endogène.

La variable endogène est une variable de comptage. Les modèles linéaires généralisés permettent l'étude de telles variables. En effet, la loi de probabilité de Poisson et la loi binomiale négative sont des lois adaptées aux phénomènes de comptage et appartiennent à la famille exponentielle.

Nous allons, dans ce qui suit, appliquer ces deux modèles linéaires généralisés à l'effectif sinistré.

Remarque : Les modèles linéaires généralisés adaptés aux données de comptage introduisent la notion de variable d'*offset* ou d'*exposition*. L'introduction d'une telle variable revient à

rapporter la variable endogène sur l'offset. Par exemple, si l'on cherche à modéliser les accidents de travail annuels d'un ensemble d'ouvriers et que le nombre de jours de travail (exposition) vari d'un ouvrier à un autre, il est donc plus judicieux de modéliser le nombre d'accidents par jour de travail. Ici, la variable offset est le nombre de jours de travail. Dans notre cas nous allons considérer le nombre d'assuré de chaque tranche comme offset. Nous modélisons donc en réalité le nombre de sinistre par personne.

II.6.1. Application de la régression de Poisson

Nous utilisons la procédure GENMOD du logiciel SAS pour ajuster les modèles de Poisson. Nous commençons par considérer le modèle suivant :

$$EFFECT_SINISTR = TR_AGE + TYPE_ASS + TYPE_BEN + SEXE + ALD + REGION + SEXE * TR_AGE + ALD * TR_AGE$$

Nous sommes en réalité intéressé par le taux de sinistralité de chaque tranche et non pas le nombre de sinistres absolu. Nous utilisons donc la variable EFFECTIF comme offset. Cette variable représente le nombre de bénéficiaires de chaque tranche, elle a été extraite de la base de données sur la population en faisant une jointure entre les deux bases de données.

Nous retrouvons la sortie suivante :

Parameter		Estimate	Pr > ChiSq	Parameter		Estimate	Pr > ChiSq
Intercept		0.2132	<.0001	REGION	4	-0.1960	<.0001
TR_AGE		0.0117	<.0001	REGION	5	0.0272	<.0001
TYPE_ASS	A	0.2416	<.0001	REGION	6	0.1577	<.0001
TYPE_ASS	I	-1.7491	<.0001	REGION	7	-0.1745	<.0001
TYPE_ASS	O	-0.7028	<.0001	REGION	8	0.2021	<.0001
TYPE_ASS	R	0.4901	<.0001	REGION	9	0.4308	<.0001
TYPE_ASS	V	0.0000	.	REGION	10	0.2873	<.0001
TYPE_BEN	A	0.2700	<.0001	REGION	11	-0.0596	<.0001
TYPE_BEN	C	0.3440	<.0001	REGION	12	-0.3090	<.0001
TYPE_BEN	E	0.0000	.	REGION	13	-0.0714	<.0001
SEXE	F	0.4976	<.0001	REGION	14	0.1181	<.0001
SEXE	M	0.0000	.	REGION	15	-0.1631	<.0001
ALD	N	-2.2926	<.0001	REGION	16	0.0000	.
ALD	O	0.0000	.	TR_AGE*SEXE	F	-0.0050	<.0001
REGION	1	-0.4502	<.0001	TR_AGE*SEXE	M	0.0000	.
REGION	2	-0.1082	<.0001	TR_AGE*ALD	N	0.0588	<.0001
REGION	3	-0.4079	<.0001	TR_AGE*ALD	O	0.0000	.

Tableau 5 : Estimation des paramètres du modèle de la régression de poisson appliqué aux effectifs sinistrés

Nous remarquons donc que tous les paramètres sont significatifs. Cependant nous devons vérifier que notre modèle ne présente pas de sur-dispersions.

La sur-dispersion se traduit par le fait d'avoir une variance observée supérieure à celle estimée par le modèle. Ce phénomène provient du fait que la loi de Poisson n'a qu'un seul paramètre, ce qui ne permet pas d'ajuster la variance indépendamment de la moyenne.

Pour diagnostiquer un tel phénomène nous devons calculer le paramètre de dispersion du modèle (φ) qui est égal à la statistique de Pearson du modèle divisée par ses degrés de liberté ($ddl = nbr\ d'observations - nbr\ de\ paramètres - 1$).

Pour notre cas nous avons :

$$\varphi = \frac{\chi^2}{ddl} = \frac{529044.3084}{5898} = 89.6989$$

Notre modèle présente donc une sur-dispersion. Pour remédier à une telle situation, on peut utiliser la loi binomiale négative, qui a deux paramètres, au lieu de la loi Poisson.

La loi binomiale négative peut même être vue comme une loi de Poisson dans le paramètre suit une loi gamma. En effet, soit une variable aléatoire X qui suit une loi de Poisson $\mathcal{P}(\lambda)$ dont le paramètre λ suit une loi gamma, X suit donc une loi mélange et la probabilité qu'elle soit égale à k , entier, s'écrit :

$$\mathbb{P}(X = k) = \int_0^{+\infty} \mathbb{P}(X_\lambda = k) \times f(\lambda, r, \theta) d\lambda$$

Où X_λ suit une loi de Poisson de paramètre fixe λ et $f(\lambda, r, \theta)$ est la densité de la loi Gamma de paramètres r et θ .

On notant Γ la fonction gamma, nous avons :

$$\mathbb{P}(X = k) = \int_0^{+\infty} \frac{\lambda^k e^{-\lambda}}{k!} \times \frac{\lambda^{r-1} e^{-\lambda/\theta}}{\Gamma(r)\theta^r} d\lambda$$

Donc :

$$\mathbb{P}(X = k) = \int_0^{+\infty} \frac{\lambda^{k+r-1} e^{-\lambda(1+\theta)/\theta}}{k! \Gamma(r)\theta^r} d\lambda$$

On effectue le changement de variable $t = \lambda \frac{1+\theta}{\theta}$ nous obtenons :

$$\mathbb{P}(X = k) = \left(\frac{\theta}{1+\theta}\right)^{r+k} \frac{1}{\Gamma(r)k!\theta^r} \int_0^{+\infty} t^{k+r-1} e^{-t} dt$$

D'où :

$$\mathbb{P}(X = k) = \left(\frac{1}{1+\theta}\right)^r \left(\frac{\theta}{1+\theta}\right)^k \frac{\Gamma(k+r)}{\Gamma(r)k!}$$

En posant $p = \left(\frac{1}{1+\theta}\right)$ et $q = \left(\frac{\theta}{1+\theta}\right)$ et en remarquant que $p + q = 1$ nous obtenons :

$$\mathbb{P}(X = k) = \frac{\Gamma(k+r)}{\Gamma(r)k!} p^r q^q$$

Sachant que $\frac{\Gamma(k+r)}{\Gamma(r)} = \frac{(k+n-1)!}{(n-1)!}$, nous avons :

$$\mathbb{P}(X = k) = \frac{(k+n-1)!}{(n-1)!k!} p^r q^q$$

Cette expression est celle de la fonction de densité de la loi binomiale négative.

II.6.2. Application de la régression binomiale négative

Nous utilisons toujours la procédure GENMON de SAS. Après l'ajustement nous constatons que quelques modalités ne sont pas significatives.

En pratique, pour obtenir le modèle le plus parcimonieux, nous devons regrouper, pour chaque variable, les deux modalités les moins significatives, puis reconduire l'estimation et répéter cette procédure, jusqu'à ne plus avoir de modalité non significatives. Cependant une telle démarche va complexer le calcul des prévisions, vu que toutes les modalités ont été retenues dans la projection démographique.

la sortie de l'ajustement est la suivante :

Parameter		Estimate	Pr > ChiSq	Parameter		Estimate	Pr > ChiSq
Intercept		1.0223	<.0001	REGION	14	0.0195	0.7567
TR_AGE		-0.0553	<.0001	REGION	15	-0.0926	0.1789
TYPE_ASS	A	0.2550	<.0001	REGION	16	0.0000	.
TYPE_ASS	I	-1.6598	<.0001	TYPE_ASS*TYPE_BEN	A	-0.1901	0.0156
TYPE_ASS	O	-0.0052	0.9782	TYPE_ASS*TYPE_BEN	A	-0.3292	0.0010
TYPE_ASS	R	0.2870	<.0001	TYPE_ASS*TYPE_BEN	A	0.0000	.
TYPE_ASS	V	0.0000	.	TYPE_ASS*TYPE_BEN	I	-0.1873	0.0659
TYPE_BEN	A	0.8076	<.0001	TYPE_ASS*TYPE_BEN	I	0.3483	0.0046
TYPE_BEN	C	0.8652	<.0001	TYPE_ASS*TYPE_BEN	I	0.0000	.
TYPE_BEN	E	0.0000	.	TYPE_ASS*TYPE_BEN	O	-1.1492	<.0001
SEXE	F	-0.3930	<.0001	TYPE_ASS*TYPE_BEN	O	0.0000	.
SEXE	M	0.0000	.	TYPE_ASS*TYPE_BEN	R	0.3513	<.0001
ALD	N	-2.8887	<.0001	TYPE_ASS*TYPE_BEN	R	-0.0035	0.9729
ALD	O	0.0000	.	TYPE_ASS*TYPE_BEN	R	0.0000	.
REGION	1	-0.1838	0.0453	TYPE_ASS*TYPE_BEN	V	0.0000	.
REGION	2	-0.0699	0.3786	TYPE_ASS*TYPE_BEN	V	0.0000	.
REGION	3	-0.2127	0.0072	TYPE_ASS*TYPE_BEN	V	0.0000	.
REGION	4	-0.1749	0.0049	TR_AGE*SEXE	F	0.0535	<.0001
REGION	5	0.1487	0.0191	TR_AGE*SEXE	M	0.0000	.
REGION	6	0.1584	0.0138	TR_AGE*ALD	N	0.0399	<.0001
REGION	7	-0.0825	0.1914	TR_AGE*ALD	O	0.0000	.
REGION	8	0.1669	0.0084	SEXE*ALD	F	0.5420	<.0001
REGION	9	0.3752	<.0001	SEXE*ALD	F	0.0000	.
REGION	10	0.2048	0.0007	SEXE*ALD	M	0.0000	.
REGION	11	-0.0191	0.7629	SEXE*ALD	M	0.0000	.
REGION	12	-0.1527	0.0284	REGION	14	0.0195	0.7567
REGION	13	0.0016	0.9804	REGION	15	-0.0926	0.1789

Tableau 6 : Estimation des paramètres du modèle de la régression binomiale négative appliqué aux effectifs sinistrés

Lorsque l'on ajuste des données de comptage à l'un des modèles linéaires généralisés, il est commun d'avoir des zéros en excès. C'est généralement le cas en assurance maladie quand la variable à modéliser est le nombre de sinistre par contrat et non pas par classe de la population, car la majorité des contrats ne présente pas de sinistralité durant l'année.

La proportion des classes dans notre cas ne présentent pas de sinistralité et de 28.15 %, on peut donc suspecter d'être dans la situation d'excès de zéro. Les deux modèles présentés ci-dessus ne prennent pas en considération ce phénomène. Cependant il existent des modèles prenant en compte l'excès de zéros, ZIP (Zero-Inflated Poisson) et ZINB (Zero-Inflated Negative Binomial), qui appartiennent aux modèles à densité mélange précédemment exposés.

Le modèle ZINB suppose que chaque observation Y_i de la variable endogène suit une loi de la forme suivante :

$$Y_i \sim \begin{cases} 0 & \text{avec une probabilité } p_i \\ NB(n_i, q_i) & \text{avec une probabilité } 1 - p_i \end{cases}$$

Ce qui est équivalent à :

$$Y_i \sim \begin{cases} 0 & \text{avec une probabilité } p_i + (1 - q_i) \times q_i^{n_i} \\ k & \text{avec une probabilité } (1 - p_i) \times C_{k+n_i-1}^k \times q_i^{n_i} (1 - q_i)^k \end{cases}$$

Où p_i est la probabilité de nullité de Y_i et NB représente une variable aléatoire de loi binomiale négative.

L'ajustement du modèle se fait donc en deux temps :

- Un modèle linéaire généralisé pour p_i , qui suit une loi de Bernoulli, avec une fonction de lien Logit.
- Une régression binomiale négative.

Et donc, en estimant le modèle par un logiciel de statistique, nous devons spécifier les variables des deux sous-modèles.

II.6.3. Application du modèle ZINB

La procédure GENMOD de SAS permet d'ajuster des données de comptage au modèle ZINB. La procédure d'estimation ne converge pas quand on spécifie trop de termes d'interaction. Pour remédier à ceci nous allons estimer le modèle sans interaction et introduire les interactions une à une, pour finalement retenir le modèle avec le moindre AIC.

Les termes d'interaction retenus en procédant comme décrit sont : $TYPE_ASS * TYPE_BEN$, $ALD * SEXE$, $SEXE * TR_AGE$ et $TR_AGE * ALD$. Cependant on remarque qu'en éliminant le terme $ALD * SEXE$ de la partie Logit du modèle, le AIC passe de 45900.5542 à 45853.3849.

En notant P la de nullité, le modèle ZINB retenu est :

$$EFFECT_SINISTR = TYPE_ASS + TYPE_BEN + SEXE + ALD + REGION + TYPE_BEN * TYPE_ASS + SEXE * TR_AGE + SEXE * ALD$$

Et :

$$P = TYPE_ASS + TYPE_BEN + SEXE + ALD + REGION + TYPE_BEN * TYPE_ASS + SEXE * TR_AGE$$

Maintenant nous allons comparer le modèle ZINB à celui de la régression binomiale négative pour confirmer ou bien infirmer l'excès de zéros.

La comparaison entre les deux modèles peut être réalisée à l'aide du test du rapport de vraisemblance. En effet le modèle de régression binomiale négative peut être vu comme un modèle ZINB avec des contraintes de nullité sur les paramètres relatives à la partie Logit, et donc le premier est un modèle contraint du deuxième.

La statistique du test du rapport de vraisemblance est la suivante :

$$D = -2 \times \ln \left(\frac{\text{vraisemblance du modèle contraint}}{\text{vraisemblance du modèle non contraint}} \right)$$

Dans notre cas :

$$D = -2 \times ((-27600.4232) - (-22839.6925)) = 9531.4614$$

L'hypothèse nulle du test du rapport de vraisemblance est que les contraintes du modèle contraint sont vérifiées.

La statistique D suit asymptotiquement une loi de χ^2 de degrés de libertés (ddl) égale à la différence entre les degrés de libertés du modèle contraint et ceux du modèle non contraint. Dans notre cas :

$$ddl = 5891 - 5839 = 52$$

Nous avons :

$$D \gg ddl$$

Et donc nous rejetons H_0 , c'est-à-dire que nous retenons le modèle non-contraint : ZINB. Ce choix est confirmé par la comparaison des AIC (45853.3849 pour le modèle choisi et 55270.8465 pour le modèle rejeté).

III. Conclusion

A ce stade, nous avons ajusté les remboursements à l'aide de différents modèles économétriques et cela pour différentes sous-populations non-homogènes. De par ces ajustements, nous disposons d'estimations des remboursements moyens annuels selon les différentes variables descriptives, ce qui est équivalent à l'estimation de la sévérité des sinistres de l'AMO. De plus, nous avons ajusté les effectifs des sinistres par classe de population aux modèles adéquats aux données de comptage pour avoir une estimation de la fréquence de sinistralité.

En conjuguant les résultats de ce chapitre et ceux des projections démographiques, nous pouvons prédire les dépenses de la CNSS en assurance maladie obligatoire.

Dans ce qui suit, nous nous intéresserons aux recettes de la CNSS relatives à l'AMO, et ce en se basant sur les assiettes de cotisation de la population AMO CNSS de 2012.

Chapitre 4 : Assiettes de cotisation et Projections

Ce chapitre s'articule sur deux axes. Un premier axe relatif à l'assiette de cotisation des bénéficiaires de l'AMO-CNSS, où cette grandeur sera décrite selon les différentes catégories de cotisants ainsi que leurs différentes caractéristiques. Puis un deuxième axe concernant les projections, où les résultats de l'étude seront consolidés sous forme d'un compte de produits et charges (CPC).

I. Assiette de cotisation

L'assiette de cotisation d'une classe de population active, au titre de l'année n , représente le montant global duquel a été prélevée la contribution de l'assuré et la contribution de l'employeur au cours de cette année.

Dans ce qui suit, nous allons tout d'abord effectuer une étude descriptive des données relatives aux assiettes de la population AMO-CNSS selon chacune des variables inputs de la base de données, pour ensuite étudier l'influence de toutes ces variables à la foi, et ceci en se basant sur un modèle économétrique.

I.1. Etude descriptive des assiettes de cotisation

Cette partie sera réservée à la présentation d'une étude descriptive des assiettes de cotisation de la population AMO-CNSS de l'année 2012

I.1.1. Description de la base de donnée

La base de données des assiettes de cotisation de la population AMO CNSS de l'année 2012 fournit les assiettes de cotisations (*ASS_COTIS*) de chaque classe de population. Ces classes étant définies à l'aide des variables descriptives suivantes :

- *TYPE_ASS* : Type de l'assuré
- *TR_AGE* : Tranche d'âge de l'assuré
- *SEXE* : Sexe de l'assuré
- *OUVERT_DROIT* : 'N' pour les assurés en période de stage, 'O' sinon.
- *VILLE* : Ville de résidence de l'assuré
- *ALD* : 'O' si l'assuré est atteint d'une maladie de longue durée, 'N' sinon.
- *TR_REV* : Tranche de revenu de l'assuré.

Pour éliminer l'effet de la taille de chaque classe, nous allons calculer l'assiette moyenne de chaque classe, et ce, en faisant une jointure avec la base de donnée de la population des assurés uniquement. Cette jointure permet d'introduire la colonne *EFFECTIF* représentant l'effectif de chaque classe.

Nous recodons alors une nouvelle variable *ASS_COTIS_M* comme suit :

$$ASS_COTIS_M = \frac{ASS_COTIS}{EFFECTIF}$$

Nous allons nous intéresser maintenant à l'assiette de cotisation selon chacune des variables descriptives.

I.1.2. Assiette selon le type d'assuré

En regroupant la base de données selon la variable *ASS_COTIS*, nous dressons le diagramme en camembert suivant, représentant la part de l'assiette de chaque type d'assuré dans l'assiette globale.

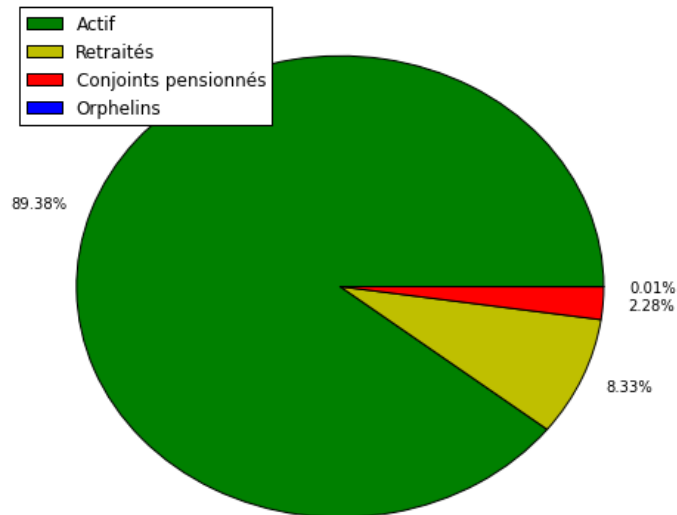


Figure 38 : Contribution à l'assiette globale de chaque type d'assuré

Nous remarquons que les actifs représentent près de 90% de l'assiette globale de la CNSS, suivis par les retraités avec seulement 8.33% tandis que les veufs et les orphelins ne dépassent pas les 3%.

Les assiettes moyennes de ces catégories sont représentées par la figure suivante :

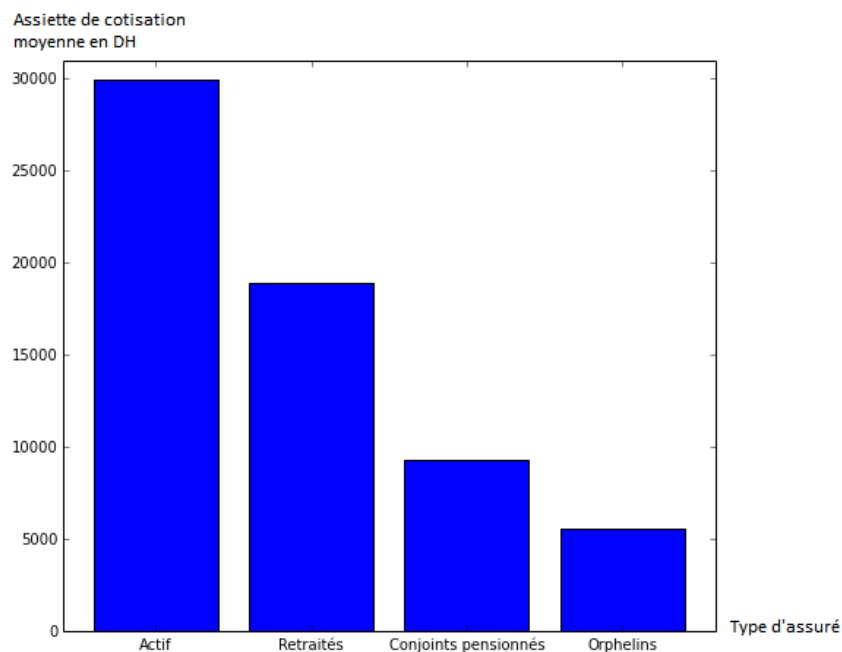


Figure 39 : Assiette de cotisation moyenne de la population AMO-CNSS selon le type d'assuré

Nous remarquons que les actifs représentent la catégorie de contribution la plus importante en ce qui concerne l'AMO du secteur privé, suivis de retraités puis des autres pensionnés.

I.1.3. Assiette selon la tranche d'âge

La figure suivante représente l'assiette moyenne de chaque tranche d'âge :

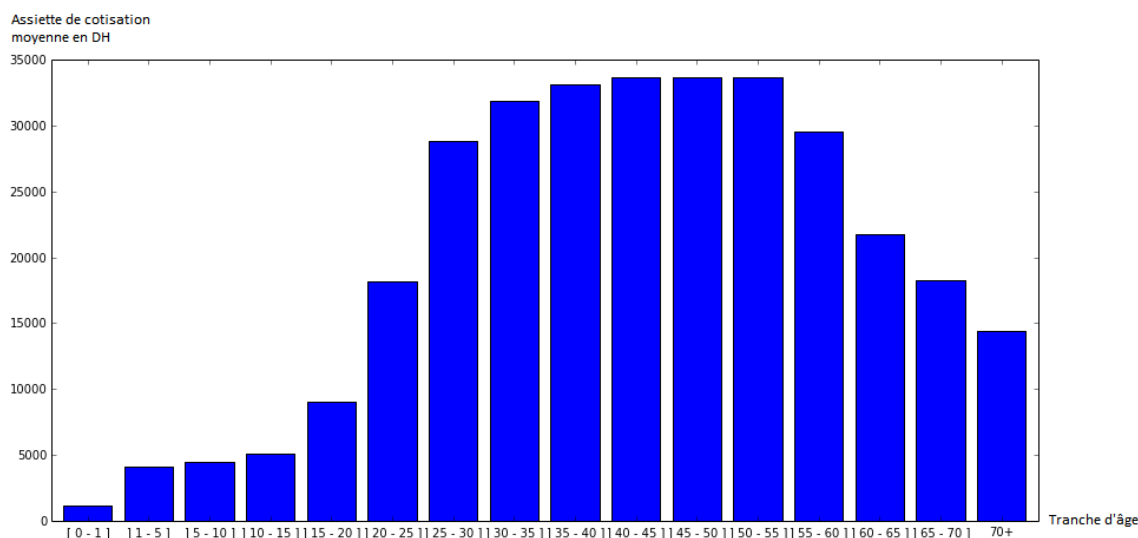


Figure 40 : Assiette de cotisation moyenne de la population AMO-CNSS selon la tranche d'âge

Ce graphique peut-être subdivisé en trois parties : La première partie, qui représente majoritairement les orphelins est celle des moins de 15 ans, présente une assiette moyenne faible. La deuxième partie est celle des actifs (entre 20 et 55 ans) où l'assiette moyenne connaît une croissance logarithmique. La troisième partie est celle des retraités, où l'assiette moyenne chute avec l'âge (vu que la pension de retraite est calculée sur la base des salaires de la période d'activité et que le niveau global des salaires connaît une hausse continue).

Les tranche d'âge]15-20] et]55-60] sont des tranches hybrides, la première est départagée entre les orphelins et les actifs, tandis que la deuxième est départagée entre les actifs et les bénéficiaires de la retraite anticipée.

Vu que les actifs sont la catégorie d'assurés qui contribuent le plus aux recettes de la CNSS, nous nous intéressons à l'évolution de leurs assiettes de cotisation moyennes selon l'âge. Cette évolution est illustrée par le graphique suivant :

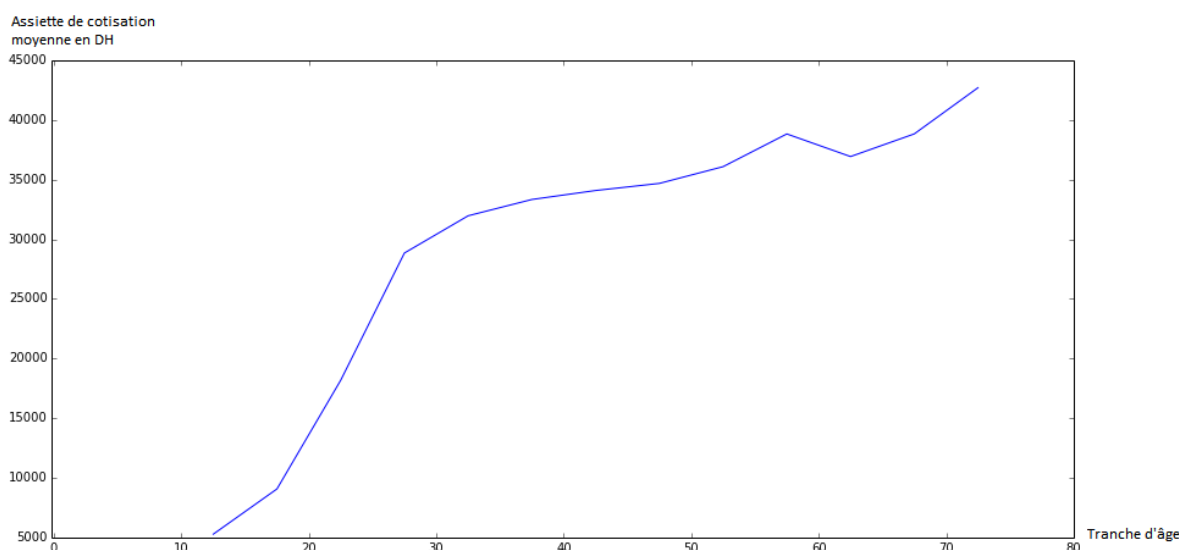


Figure 41 : Évolution de l'assiette moyenne de cotisation de la population AMO-CNSS selon l'âge

Nous remarquons une tendance générale haussière, vu que les salaires augmentent généralement avec l'âge. Cependant la pente de la courbe diminue considérablement après les vingtaines, ce qui illustre le fait que la croissance du salaire avec l'âge est plus faible en fin de carrière qu'au début de carrière.

Néanmoins, après l'âge de 55 ans, la courbe présente une allure en 'N'. Cette allure est rencontrée généralement chez les organismes de prévoyance. Elle est interprétée comme suit : Aux environs de 55 ans, le salaire moyen des actifs commence à diminuer vu que la capacité de travail de ces derniers devient généralement faible. Vers 60 ans, les actifs dont la capacité à travailler a diminué (le salaire aussi) sont soit obligés de sortir à la retraite, soit de sortir de plein gré pour ne pas voir leur rente de retraite diminuer (spécialement si celle-ci est calculée à base du dernier salaire). Les actifs qui restent dans le régime après 60 ans sont généralement les grands cadres et les directeurs, dont le travail ne nécessite pas de force physique et dont l'efficacité n'est pas fortement impactée par l'âge. Ces derniers ont des salaires élevés, donc la moyenne des salaires de la population active augmente, ce qui justifie le redressement final de la courbe.

I.1.4. Assiette selon état de santé

Le caractère ALD d'un bénéficiaire impacte dramatiquement sa consommation, cependant son assiette de cotisation moyenne n'est que légèrement affectée, comme le montre la figure suivante, représentant l'assiette moyenne des ALD et celle des non-ALD :

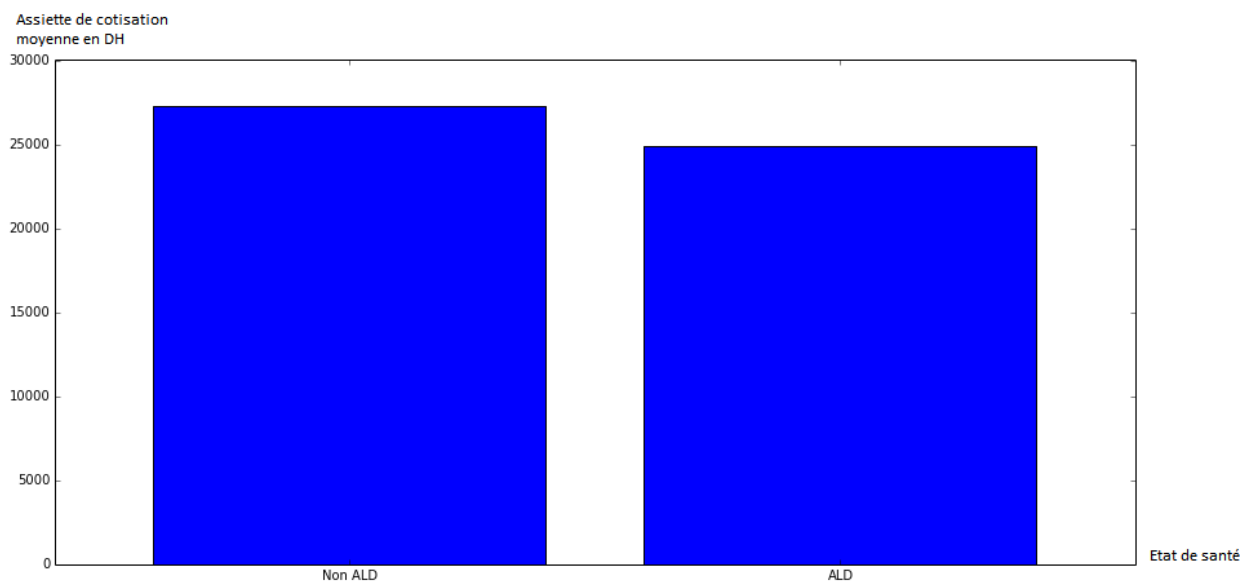


Figure 42 : Assiette de cotisation moyenne de la population AMO-CNSS selon l'état de santé du bénéficiaire

I.1.5. Assiette selon le sexe

Les assiettes de cotisation moyennes selon les sexes sont représentées par le graphique suivant :

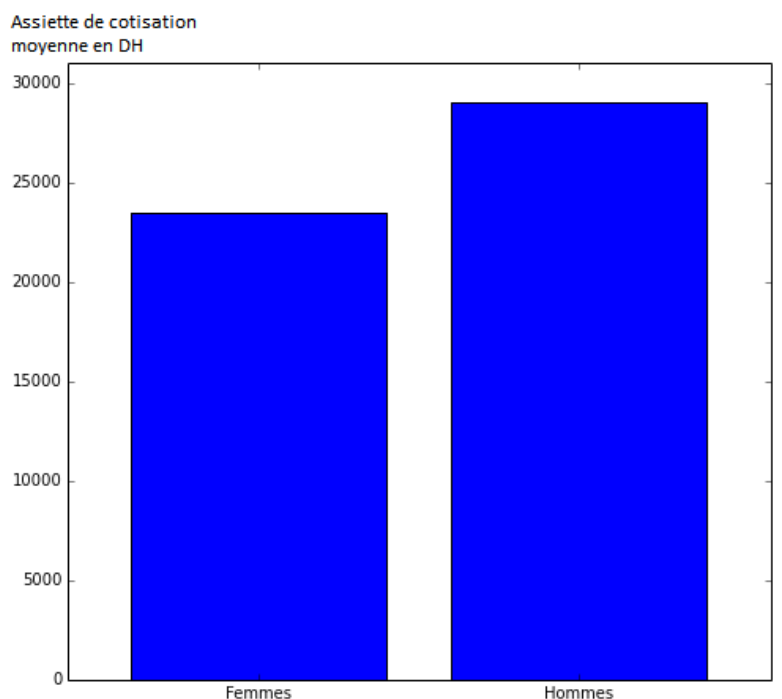


Figure 43 : Assiette de cotisation moyenne de la population AMO-CNSS selon le sexe

Nous remarquons que la différence entre les assiettes moyennes des deux sexes est significative (plus de 9500 DH).

Il importe de noter que les actifs et les retraités de l'AMO-CNSS sont majoritairement de sexe masculin, et donc que la contribution des hommes à l'assiette globale est largement supérieure à celle des femmes. Le graphique suivant illustre ceci.

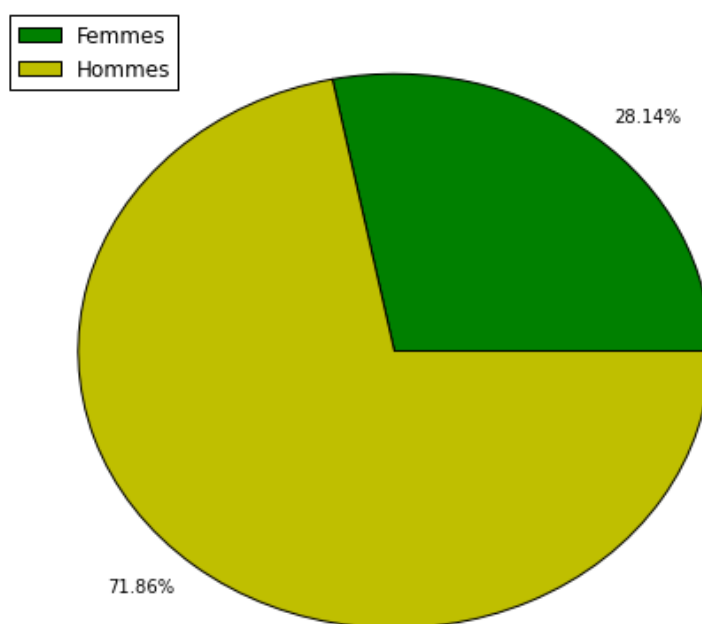
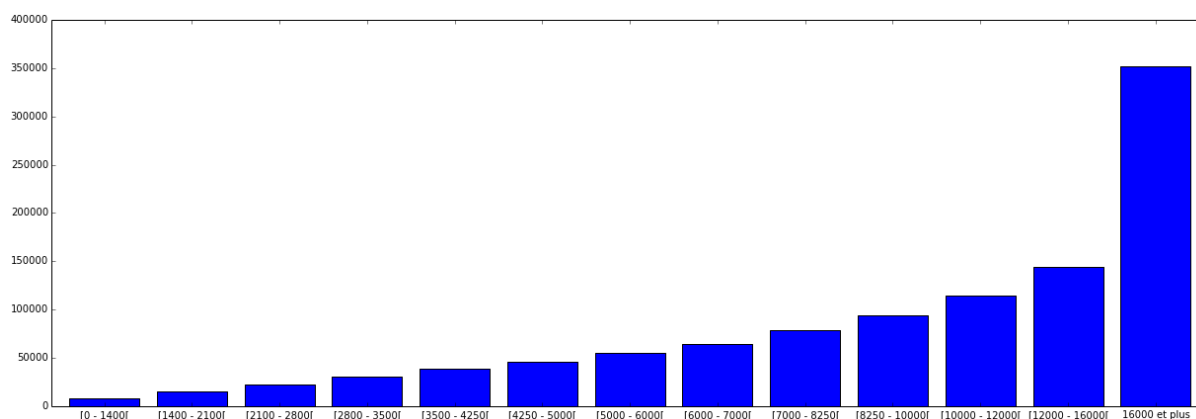


Figure 44 : Contribution à l'assiette globale de la population AMO-CNSS selon le sexe

I.1.6. Assiette selon la tranche de revenu

Le revenu et l'assiette sont deux notions différentes. L'assiette d'un individu ayant un revenu annuel de 12 000 DH et qui n'est déclaré que pendant 10 mois de l'année est de 10 000 DH.

Nous constatons que la relation entre revenu et l'assiette moyenne est globalement positive comme le montre la figure suivante :



Vu que la dernière tranche de revenu trahi la tendance du graphique, elle devrait être détaillée davantage.

Dans la partie qui suit nous allons étudier la dépendance globale entre assiette moyenne et les variables descriptive. Pour ce faire nous allons ajuster les assiettes moyennes des classe aux différentes variables descriptives.

I.2. Modèle de l'assiette de cotisation

Dans cette partie, nous allons choisir le modèle économétrique le plus adéquat aux assiettes moyennes de la population AMO-CNSS, pour ensuite interpréter les résultats de l'ajustement.

I.2.1. Choix du modèle

Nous commençons par ajuster le modèle de régression linéaire multiple suivant :

$$ASS_COTIS_M = AGE + TYPE_ASS + OUVERT_DROIT + ALD + REGION + SEXE$$

Le tableau de l'analyse de la variance du modèle est le suivant :

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	2.910789E14	1.455394E13	330.26	<.0001
Error	1300	5.72887E13	44068228448		
Corrected Total	1320	3.483676E14			

Tableau 7 : Test d'ANOVA de l'ajustement de l'assiette moyenne au modèle simple de régression linéaire

La liaison entre la variable endogène et la variable exogène est donc significative, comme l'indique aussi le coefficient d'ajustement $R^2 = 0.8356$. Mais avant de pouvoir utiliser les coefficients de ce modèle, nous devons nous assurer de la normalité des résidus.

Les logiciels à notre disposition ne permettent pas d'effectuer de test d'ajustement pour des données groupés, nous utiliserons les méthodes graphiques pour le diagnostic de la normalité.

Le diagramme quantile-quantile fourni par la sortie est le suivant :

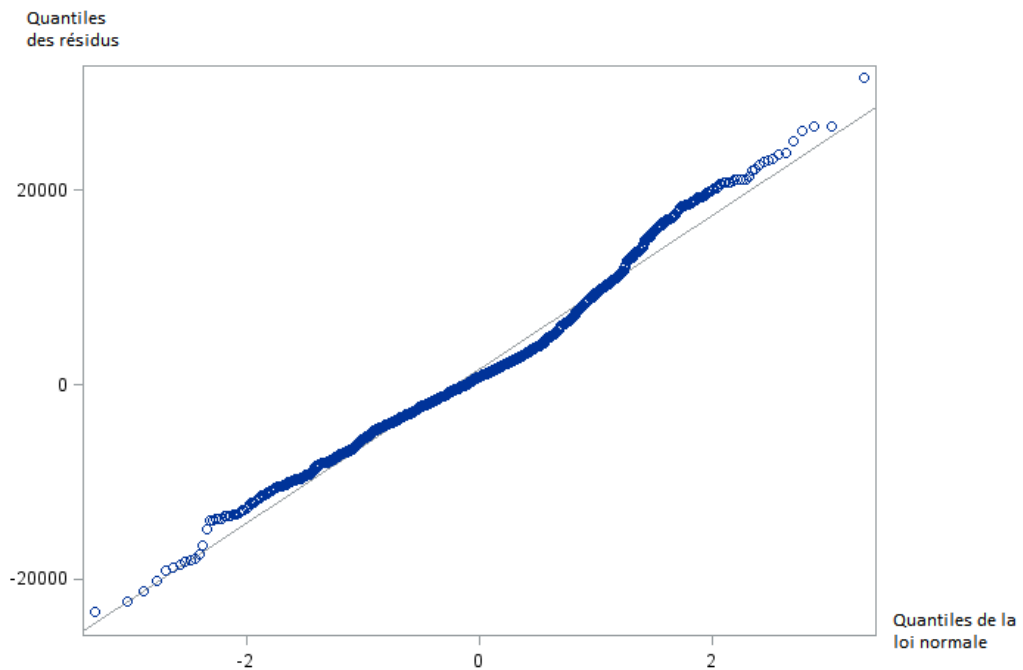


Figure 45 : QQ-plot de l'ajustement des résidus du modèle simple de régression linéaire de l'assiette moyenne de cotisation à la loi normale

Le diagramme est un peu proche de la première bissectrice, cependant les fluctuations ne sont pas négligeables.

Pour s'assurer de la qualité de l'ajustement, nous dessinons l'histogramme des résidus suivant :

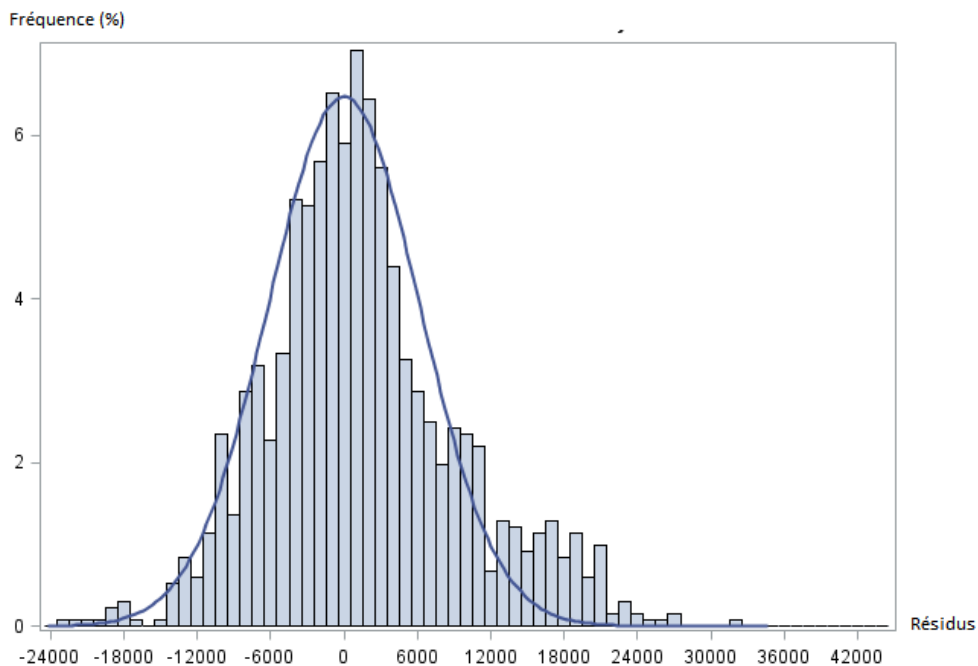


Figure 46 : Histogramme des résidus du modèle simple de régression linéaire de l'assiette moyenne de cotisation ajustés à la loi normale

L’histogramme est bien asymétrique, et la densité de l’ajustement de la loi normale diffère significativement de la densité empirique des données. La normalité des résidus n’est donc pas raisonnable comme hypothèse. De plus, le diagramme des résidus montre l’existence d’autocorrélation des erreurs et d’hétéroscédasticité.

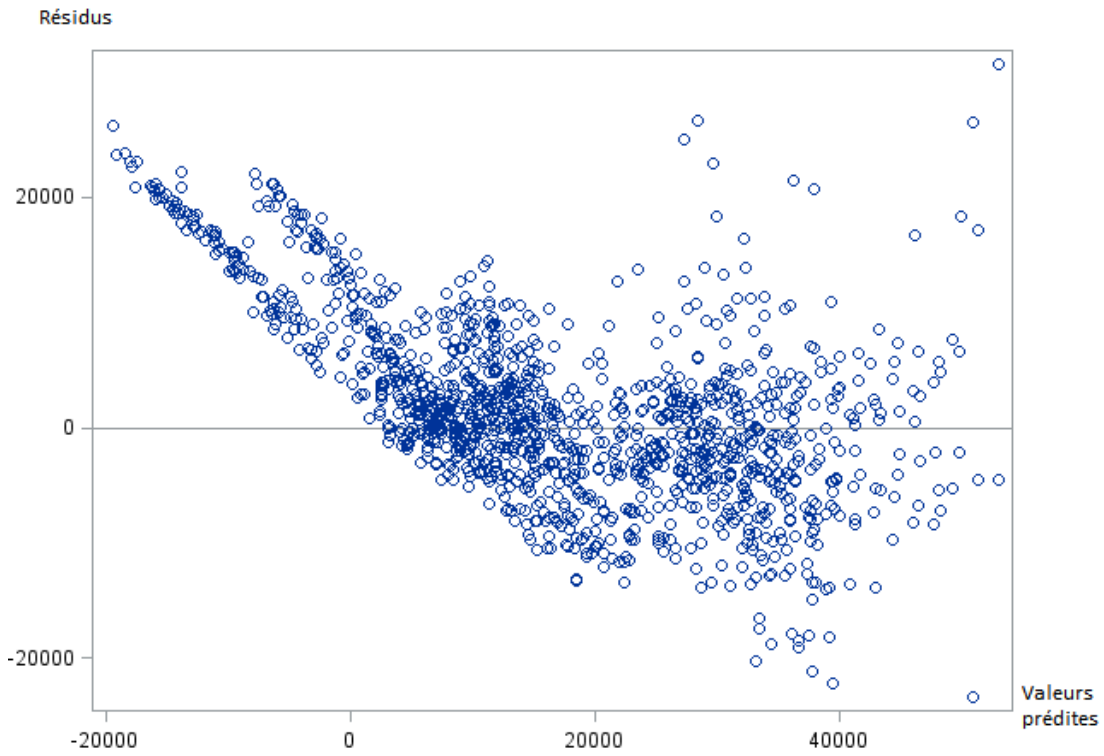


Figure 47 : Diagramme des résidus du modèle simple de régression linéaire de l’assiette moyenne de cotisation

Nous remarquons aussi l’existence de valeurs prédites négatives. Pour remédier à une telle situation, nous considérerons le logarithme de l’assiette moyenne de cotisation comme variable endogène.

Nous commençons par estimer le modèle sans interactions suivant :

$$\log(\text{ASS_COTIS_M}) = \text{AGE} + \text{TYPE_ASS} + \text{OUVERT_DROIT} + \text{ALD} + \text{REGION} + \text{SEXE}$$

La sortie du test ANOVA du modèle est la suivante :

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	654138	32707	344.17	<.0001
Error	1300	123542	95.03259		
Corrected Total	1320	777680			

Tableau 8 : Test d’ANOVA de l’ajustement de l’assiette moyenne au modèle logarithmique de régression linéaire de l’assiette moyenne de cotisation

L’ajustement est donc toujours bon, avec un coefficient d’ajustement de $R^2 = 0.8411$.

L'étape suivante est le diagnostic de la normalité, qu'on effectue à l'aide du digramme quantile-quantile suivant :

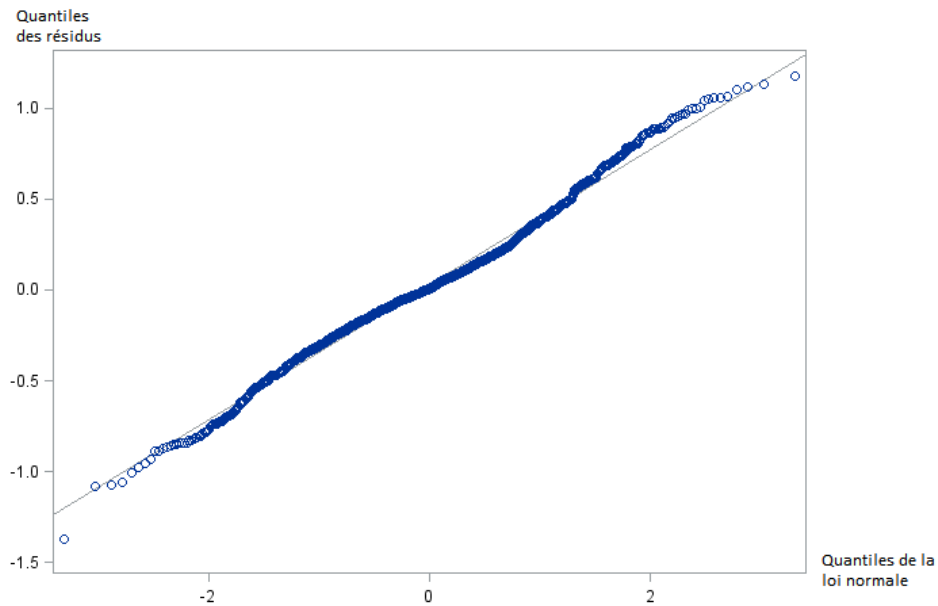


Figure 48 : QQ-plot de l'ajustement des résidus du modèle logarithmique de régression linéaire de l'assiette de cotisation moyenne à la loi normale

L'alignement de ce qq-plot est légèrement meilleur que celui du modèle précédent. Pour s'assurer de l'ajustement, nous dressons l'histogramme des résidus et la densité de l'ajustement à la loi normale :

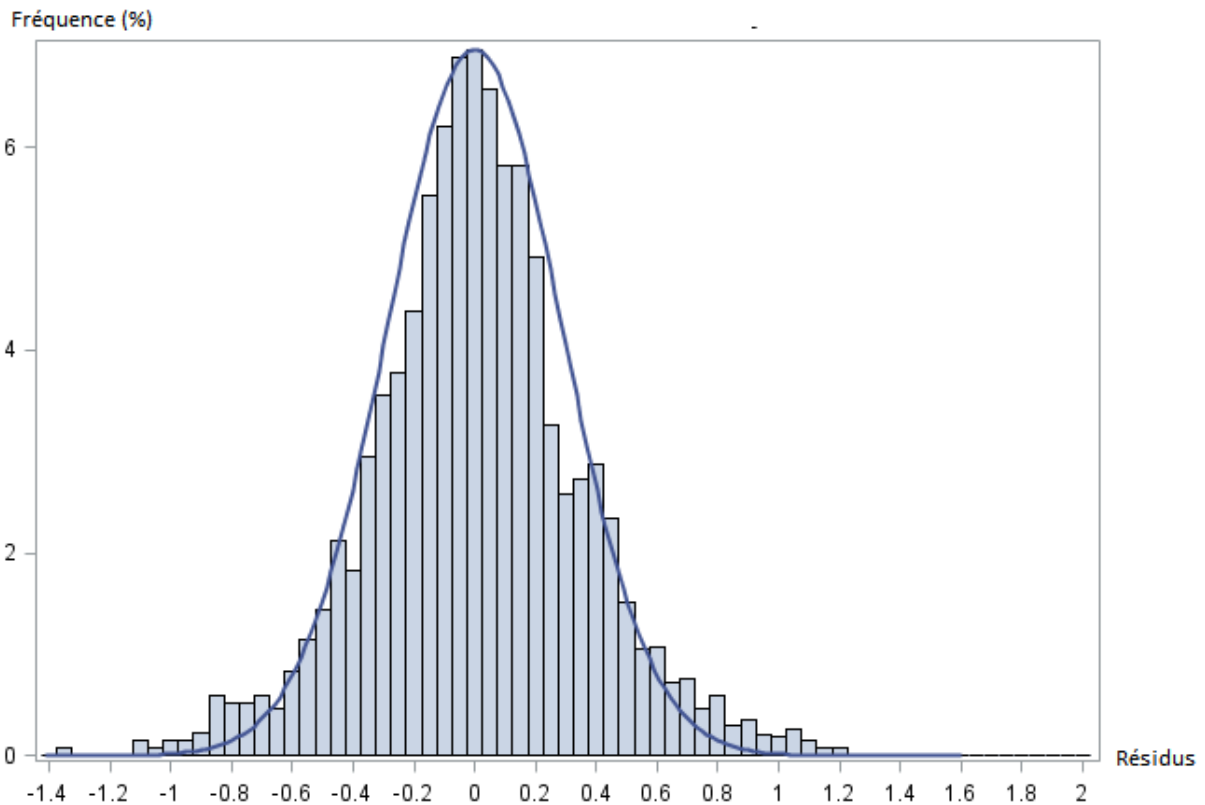


Figure 49 : Histogramme des résidus du modèle logarithmique de régression linéaire de l'assiette de cotisation moyenne ajustés à la loi normale

Nous constatons que l’histogramme en forme de cloche est assez symétrique et que la densité théorique est très proche de la densité empirique. Nous concluons donc que l’hypothèse de normalité des résidus est vérifiée.

Les coefficients de l’ajustement du modèle logarithmique sont les suivants :

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	7.92006	0.71584	11.06	<.0001
TR_AGE	1	0.01286	0.00067193	19.13	<.0001
TYPE_ASS_A	1	1.18715	0.71585	1.66	0.0975
TYPE_ASS_R	1	0.16289	0.71667	0.23	0.8202
TYPE_ASS_V	1	-0.33819	0.71694	-0.47	0.6372
SEXE_M	1	0.14586	0.01555	9.38	<.0001
OUVERT_DROIT_O	1	1.02843	0.01797	57.23	<.0001
ALD_O	1	0.01570	0.03590	0.44	0.6619
REGION__10	1	-0.09140	0.02539	-3.60	0.0003
REGION__11	1	-0.48834	0.03109	-15.71	<.0001
REGION__16	1	-0.33638	0.02626	-12.81	<.0001
REGION__04	1	-0.51721	0.02329	-22.21	<.0001
REGION__07	1	-0.32649	0.02629	-12.42	<.0001
REGION__08	1	-0.49976	0.03485	-14.34	<.0001
REGION__05	1	-0.57210	0.03086	-18.54	<.0001
REGION__14	1	-0.37254	0.03308	-11.26	<.0001
REGION__13	1	-0.49622	0.03121	-15.90	<.0001
REGION__06	1	-0.44964	0.03707	-12.13	<.0001
REGION__15	1	-0.66319	0.04730	-14.02	<.0001
REGION__12	1	-0.61150	0.04362	-14.02	<.0001
REGION__02	1	-0.48171	0.09480	-5.08	<.0001
REGION__03	1	-0.59015	0.08579	-6.88	<.0001
REGION__01	1	-0.53391	0.15421	-3.46	0.0006

Tableau 9 : Estimation des paramètres du modèle logarithmique de régression linéaire de l’assiette de cotisation moyenne

On élimine pas-à-pas les variables non-significatives. Si une variable présente plus de deux modalités, on regroupe les modalités les moins significatives entre elles.

Nous retrouvons alors un modèle d'un coefficient d'ajustement de $R^2 = 0.8512$ dont les paramètres sont les suivants :

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	7.58354	0.05312	142.76	<.0001
TR_AGE	1	0.01286	0.00066895	19.23	<.0001
TYPE_ASS_A	1	1.52382	0.03589	42.46	<.0001
TYPE_ASS_R	1	0.50174	0.03480	14.42	<.0001
SEXE_M	1	0.14586	0.01553	9.39	<.0001
OUVERT_DROIT_O	1	1.02869	0.01795	57.32	<.0001
REGION__10	1	-0.09178	0.02537	-3.62	0.0003
REGION__11	1	-0.48900	0.03104	-15.75	<.0001
REGION__16	1	-0.33693	0.02622	-12.85	<.0001
REGION__04	1	-0.51778	0.02325	-22.27	<.0001
REGION__07	1	-0.32712	0.02625	-12.46	<.0001
REGION__08	1	-0.50030	0.03482	-14.37	<.0001
REGION__05	1	-0.57265	0.03083	-18.58	<.0001
REGION__14	1	-0.37307	0.03304	-11.29	<.0001
REGION__13	1	-0.49681	0.03117	-15.94	<.0001
REGION__06	1	-0.45024	0.03703	-12.16	<.0001
REGION__15	1	-0.66384	0.04726	-14.05	<.0001
REGION__12	1	-0.61222	0.04357	-14.05	<.0001
REGION__02	1	-0.48240	0.09473	-5.09	<.0001
REGION__03	1	-0.59088	0.08572	-6.89	<.0001
REGION__01	1	-0.53451	0.15411	-3.47	0.0005

Tableau 10 : Estimation des paramètres du modèle logarithmique de régression linéaire de l'assiette de cotisation moyenne après élimination des modalités non-significatives.

L'étape suivante est d'introduire les variables d'interactions un à un on ne retenant que les variables significatives.

Les deux variables d'interaction retenues sont celle qui représentent l'interaction entre le sexe et l'âge, notée « SEXE__M__TR_AGE », et celle qui indique si le bénéficiaire est un actif de sexe masculin ou non, notée « SEXE__M__TYPE_ASS_A ».

Le coefficient d'ajustement du modèle retenu est de 86.02 %, et ses coefficients sont les suivants :

Parameter Estimates					
Variable	D F	Paramete r Estimate	Standar d Error	t Valu e	Pr > t
Intercept	1	8.14527	0.08027	101.47	<.0001
TR_AGE	1	0.00450	0.00112	4.01	<.0001
TYPE_ASS_A	1	1.25353	0.04581	27.36	<.0001
TYPE_ASS_R	1	0.51232	0.05422	9.45	<.0001
SEXE_M	1	-0.69756	0.10474	-6.66	<.0001
SEXE_M__TYPE_ASS__ A	1	0.40976	0.06952	5.89	<.0001
SEXE_M__TR_AGE	1	0.01248	0.00137	9.12	<.0001
OUVERT_DROIT_O	1	1.02158	0.01743	58.61	<.0001
REGION__10	1	-0.09355	0.02461	-3.80	0.0002
REGION__11	1	-0.48471	0.03015	-16.08	<.0001
REGION__16	1	-0.34178	0.02544	-13.43	<.0001
REGION__04	1	-0.51435	0.02256	-22.80	<.0001
REGION__07	1	-0.32391	0.02547	-12.72	<.0001
REGION__08	1	-0.49968	0.03378	-14.79	<.0001
REGION__05	1	-0.57698	0.02991	-19.29	<.0001
REGION__14	1	-0.37678	0.03206	-11.75	<.0001
REGION__13	1	-0.49627	0.03024	-16.41	<.0001
REGION__06	1	-0.44951	0.03592	-12.51	<.0001
REGION__15	1	-0.66349	0.04585	-14.47	<.0001
REGION__12	1	-0.60462	0.04228	-14.30	<.0001
REGION__02	1	-0.47796	0.09191	-5.20	<.0001
REGION__03	1	-0.58724	0.08316	-7.06	<.0001
REGION__01	1	-0.52577	0.14951	-3.52	0.0005

Tableau 11 : Tableau 10 : Estimation des paramètres du modèle logarithmique de régression linéaire de l'assiette de cotisation moyenne après élimination des modalités non-significatives et introduction des variables d'interaction.

I.2.2. Interprétation

Nous commençons par interpréter les signes des coefficients. Si le signe d'une modalité est positif, cela veut dire que, toutes choses égales par ailleurs, les actifs de cette catégorie ont une assiette plus élevée que celle de ceux appartenant à la catégorie de référence.

On tire donc du tableau les constatations suivantes :

- Les ouverts de droit cotisent donc plus que les actifs ayant les droits fermés.
- Les assurés et les retraités cotisent plus que les veufs et les orphelins, ces derniers cotisent de la même façon (vu que la modalité représentant les veufs n'est pas significative).
- Toutes les régions cotisent moins que la région de référence, celle de Casablanca.
- L'âge impacte positivement l'assiette des cotisants.

Le coefficient de la variable $SEXE_M$, indiquant si le bénéficiaire est de sexe masculin, est négatif. On ne peut conclure que les cotisants de sexe masculin cotisent généralement moins que les actifs de sexe féminin, car les termes d'interaction relatifs à cette variable ont un signe contraire. On peut dire que :

- Pour les non actifs, toutes choses égales par ailleurs, et tant que l'âge du cotisant ne dépasse pas $\frac{0.69756}{0.01248} = 55.89$ ans, les femmes cotisent plus que les hommes.
- Pour les actifs, toute choses égales, et tant que l'âge du cotisant ne dépasse pas $\frac{0.69756-0.40976}{0.01248} = 23.06$ ans, les femmes cotisent plus que les hommes.

En effet, sachant que le terme d'interaction est une multiplication entre deux variables, nous avons :

$$\frac{\partial \ln(ASS_COTIS_M)}{\partial SEXE_M} = -0.69756 + 0.01248 \times TR_AGE + 0.40976 \times TYPE_ASS_A$$

Et donc :

$$\frac{\partial \ln(ASS_COTIS_M)}{\partial SEXE_M} \geq 0 \Leftrightarrow \begin{cases} TYPE_ASS_A = 1 \text{ et } TR_AGE \geq \frac{0.69756 - 0.40976}{0.01248} \\ \text{ou} \\ TYPE_ASS_A = 0 \text{ et } TR_AGE \geq \frac{0.69756}{0.01248} \end{cases}$$

D'où le résultat.

Après avoir interprété les signes des coefficients, nous passons à l'interprétation des leurs valeurs.

La difficulté réside dans le fait que la variable d'intérêt, ASS_COTIS_M , a connu une transformation logarithmique. Dans ce cas, l'interprétation se fait en raisonnant à l'aide des moyennes géométriques et non des moyennes arithmétiques.

En guise d'illustration, considérons le modèle ne contenant que la constante :

$$y = \ln(ASS_COTIS_M) = b$$

L'ajustement de ce modèle donne :

$$\hat{b} = 10.05893$$

Nous savons que :

$$\hat{b} = \bar{y} = \frac{1}{n} \sum y_i$$

Où n est le nombre d'observations.

Donc :

$$\exp(\hat{b}) = \exp\left(\frac{1}{n} \sum y_i\right) = \left(\prod \exp(y_i)\right)^{\frac{1}{n}}$$

D'où :

$$\exp(\hat{b}) = \left(\prod \text{ASS_COTIS_M}_i\right)^{\frac{1}{n}} = \overline{\overline{\text{ASS_COTIS_M}}}$$

Avec $\overline{\overline{\text{ASS_COTIS_M}}}$, la moyenne géométrique de ASS_COTIS_M .

Passons maintenant au modèle simple suivant:

$$\ln(\text{ASS_COTIS_M}) = a \times \text{SEXE_M} + b$$

L'ajustement du modèle donne :

$$\hat{a} = 0.27937 \quad \text{et} \quad \hat{b} = 9.87045$$

En suivant le même cheminement, nous pouvons établir que « $\exp(\hat{b})$ » est la moyenne géométrique de ASS_COTIS_M des classes de sexe féminin et « $\exp(\hat{a} + \hat{b})$ » est celle des classes de sexe masculin. Donc :

$$\exp(\hat{a}) = \frac{\exp(\hat{a} + \hat{b})}{\exp(\hat{b})} = \frac{\overline{\overline{\text{ASS_COTIS_M}}}_{\text{sexe masculin}}}{\overline{\overline{\text{ASS_COTIS_M}}}_{\text{sexe féminin}}}$$

Ainsi, l'exponentiel du coefficient \hat{a} est interprétable comme étant de combien la moyenne géométrique de ASS_COTIS_M chez les classes de sexe masculin est supérieur à celle des classes de sexe féminin.

Vu que $\exp(\hat{a}) = 1.32$, on peut dire que globalement, la moyenne géométrique de la cotisation moyenne d'une classe de sexe masculin est supérieure à celle de sexe féminin de 32%.

Sachant que l'objectif de l'interprétation des résultats d'un modèle étant la simplification de la réalité, est que l'interprétation des coefficients des termes d'interaction étant doublement complexe pour le modèle logarithmique, nous nous intéressons au modèle sans interactions.

On dresse le tableau suivant des exponentiels des coefficients du modèle :

Variable	Coefficients	exp(Coefficient)
Constante	7,58354	1965,574799
TR_AGE	0,01286	1,012943045
TYPE_ASS_A	1,52382	4,589724498
TYPE_ASS_R	0,50174	1,651592543
SEXE_M	0,14586	1,157034192
OUVERT_DROIT_O	1,02869	2,797398841
REGION_10	-0,09178	0,912305835
REGION_11	-0,489	0,613239327
REGION_16	-0,33693	0,713958815
REGION_04	-0,51778	0,59584185
REGION_07	-0,32712	0,720997218
REGION_08	-0,5003	0,606348728
REGION_05	-0,57265	0,56402878
REGION_14	-0,37307	0,688617028
REGION_13	-0,49681	0,608468582
REGION_06	-0,45024	0,637475139
REGION_15	-0,66384	0,514870431
REGION_12	-0,61222	0,542145968
REGION_02	-0,4824	0,617300092
REGION_03	-0,59088	0,553839691
REGION_01	-0,53451	0,585956338

Tableau 12 : Exponentiels des paramètres du modèle logarithmique de régression linéaire de l'assiette de cotisation moyenne

Vu que pour une classe déterminée, l'assiette de cotisation moyenne étant fixe, la moyenne géométrique de cette assiette moyenne est égale à l'assiette moyenne elle-même. On conclut donc que, toutes choses égales :

- Les assiettes moyenne des cotisants d'un certain âge sont supérieurs de 1.29 % à celles de de l'âge inférieur.
- Les assiettes moyennes des actifs et des retraités sont égales à respectivement, 4.59 et 1.65 fois celle des orphelins et des veufs.
- Les assiettes moyennes des hommes sont supérieures à celles des femmes de 15.70 %.
- L'assiette moyenne d'un cotisant résidant dans la région de Casablanca est près du double de celle d'un cotisant résidant dans la région d'Al Hoceima (*REGION_15*)

Remarque : Face à l'absence d'une base de données des assiettes de cotisation de la CNSS pour les années antérieures à 2012 (données de panel), nous ne pouvons pas effectuer des prévisions pour les années futures. Pour ce, nous allons nous contenter d'appliquer le taux d'évolution moyen des salaires de la CNSS, calculé sur la période 2009-2012, aux assiettes de l'année 2012.

II. Projections

Ayant élaboré un modèle démographique de la population AMO-CNSS et estimé les remboursements moyen par personne, et sous l'hypothèse de l'évolution constante des assiettes, nous allons dans cette partie consolider les résultats obtenus afin de dresser les comptes de produits et charges pour les cinq années avenir.

II.3. Résultats

Le CPC élaboré est un CPC technique qui ne concerne que les comptes des prestations et ceux des contributions. En effet, le modèle du CPC publié par l'ANAM dans ces rapports annuels contient plusieurs comptes dont la modélisation n'a pas été traitée par notre étude, telles les comptes : *Produits de placement, Pénalités sur cotisations, Réserves de sécurité, etc..*

Le CPC concernant la population globale est le suivant :

Libellé	Années		
	2013	2014	2015
Cotisations actifs - part salariale	1 295 294 819	1 656 073 262	2 091 236 669
Cotisation actifs - part patronale	1 295 294 819	1 656 073 262	2 091 236 669
Cotisation actifs éligibles - solidarité	971 471 114	1 242 054 947	1 568 427 501
Cotisation retraités	223 324 857	277 323 876	346 544 043
Cotisations pensionnés autre que retraités	43 352 312	51 012 659	60 624 012
Total cotisations	3 828 737 922	4 882 538 006	6 158 068 894
Prestations aux assurés	1 191 002 336	1 434 112 967	1 726 316 331
Prestations aux conjoints des assurés	411 535 900	477 531 458	558 971 386
Prestations aux enfants des assurés	167 299 010	179 161 288	192 948 146
Total prestations	1 769 837 246	2 090 805 712	2 478 235 863
Résultat technique	2 058 900 675	2 791 732 294	3 679 833 031
Libellé	Années		
	2016	2017	
Cotisations actifs - part salariale	2 616 587 809	3 238 106 371	
Cotisation actifs - part patronale	2 616 587 809	3 238 106 371	
Cotisation actifs éligibles - solidarité	1 962 440 856	2 428 579 778	
Cotisation retraités	433 077 943	547 500 823	
Cotisations pensionnés autre que retraités	72 377 385	86 768 730	
Total cotisations	7 701 071 802	9 539 062 073	
Prestations aux assurés	2 072 744 177	2 482 136 896	
Prestations aux conjoints des assurés	661 131 139	785 455 584	
Prestations aux enfants des assurés	208 338 031	226 540 540	
Total prestations	2 942 213 346	3 494 133 020	
Résultat technique	4 758 858 455	6 044 929 054	

Tableau 13 : Compte technique des produits et charge de l'AMO du secteur privé sur la période 2013-2017 en DH

Nous constatons donc que le résultat net est en constante augmentation. L'allure de son évolution est représentée par la courbe suivante :

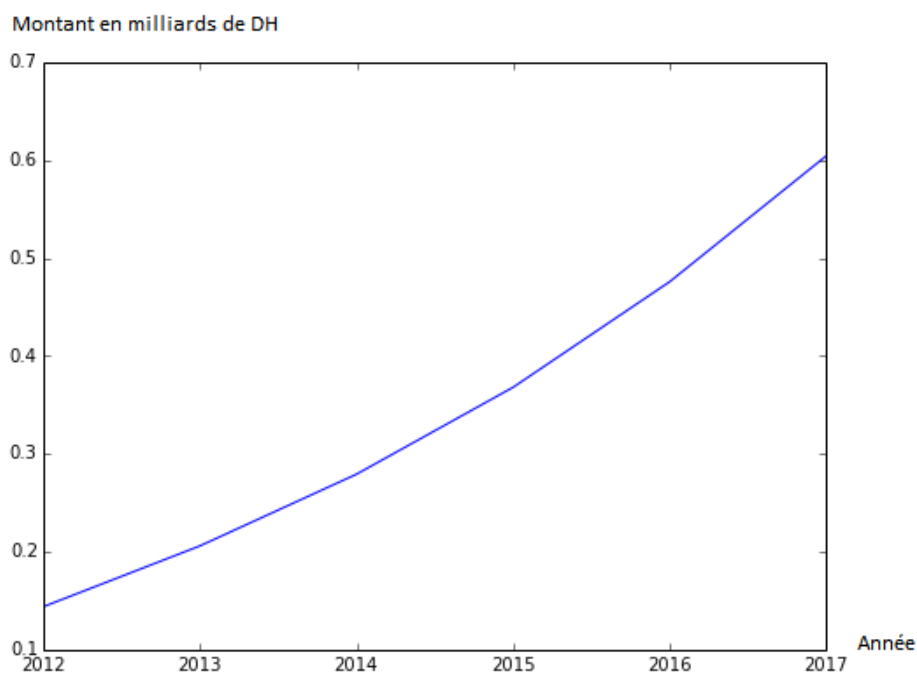


Tableau 14 : Résultat technique du CPC de l'AMO du secteur privé sur la période 2012-2017 en milliards de DH

Cette forte évolution n'est pas alarmante, vu qu'elle rejoint celle de l'excédent du bilan de l'AMO du secteur privé sur la période quinquennale 2007-2011, qui est illustré par la figure suivante :

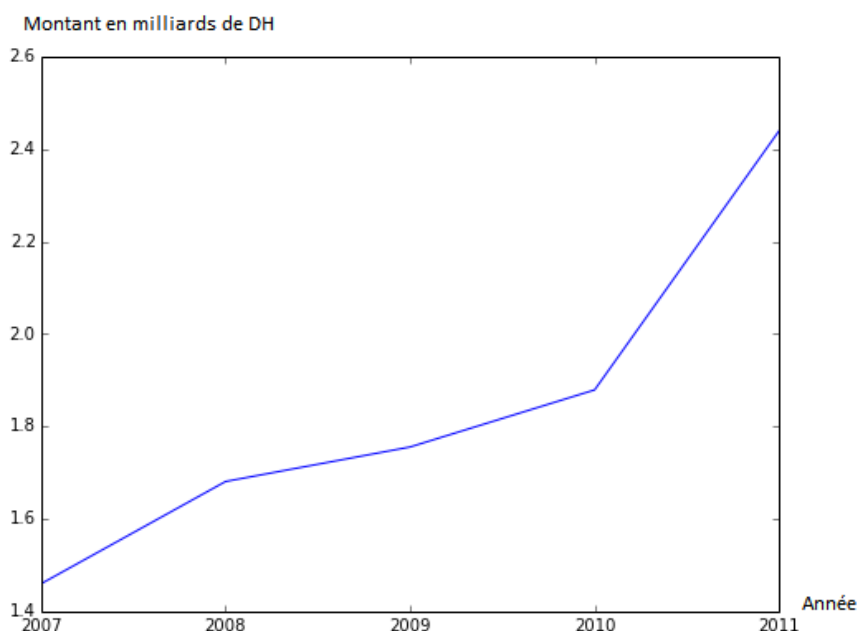


Tableau 15 : Résultat excédent du bilan de l'AMO du secteur privé sur la période 2007-2011 en milliards de DH

II.4. Conclusion

Vu les résultats du compte des produits et charges élaborés, nous constatons que l'équilibre financier du régime de l'assurance maladie obligatoire du secteur privé sera maintenu dans les cinq prochaines années.

Néanmoins, notre étude suppose que les soins sujets au remboursement, ainsi que la catégorie de la population couverte, ne connaîtrons pas de changement. Ce qui n'est pas représentatif de la réalité, vu que le secteur AMO au Maroc est en pleine extension.

Ceci étant, le résultat technique à court terme retrouvé selon les hypothèses de l'étude est important. Ce qui indique que l'AMO du secteur privé est prêt à prendre plus de risque, soit en élargissant le panier des soins remboursables, la liste des affections de longues durées ou la tranche de population couverte.

Conclusion générale

L'objet de ce rapport était d'établir un modèle actuariel de l'assurance maladie obligatoire du secteur privé, c'est-à-dire, modéliser les ressources et les dépenses du secteur pour fournir un aperçu sur sa performance. Il avait pour but de dresser un compte de produits et charges technique, dont le résultat net est un indicateur de la viabilité du régime.

Après avoir décrit le contexte dans lequel le besoin d'une telle étude s'est fait ressentir, une première étape a consisté à une étude de la population assurée par la CNSS. A l'issue de ce volet, nous avons élaboré un modèle empirique servant à la projection de la population. Ce modèle a été basé sur le bilan des flux, internes et externes, que connaît la population en question. L'estimation des différents paramètres du modèle, qui représente une grande partie de l'élaboration du modèle, a été omise dans ce rapport vu le caractère confidentiel des résultats dégagés.

Étant en mesure de projeter la population assurée auprès de la CNSS au titre de l'AMO, nous nous sommes intéressés à la consommation de cette dernière, c'est-à-dire les remboursements dont elle bénéficie. Lors de la modélisation des remboursements, une distinction entre les soins hospitaliers et les soins ambulatoires s'est avérée nécessaire. Les modèles GLM se sont avérés adéquats à la modélisation des remboursements à titre de soins ambulatoires, cependant l'hétérogénéité des données concernant les soins hospitaliers a constitué un obstacle à l'application de tels modèles. Pour remédier à ceci, nous nous sommes tournés vers les modèles à *densité mélange*, ces modèles considèrent que la distribution de la variable étudiée est constituée d'un mélange de lois de probabilité. Ce qui permet de prendre en considération une éventuelle hétérogénéité de la population étudiée sans identifier sa source.

Finalement, une étude descriptive de l'assiette de cotisation de la population assurée a été menée. Cette étude s'est déroulée en deux volets. Un premier volet étudiant l'effet de chaque variable descriptive sur l'assiette, puis un deuxième volet, étudiant l'effet joint de toutes ces variables en ajustant un modèle de régression linéaire multiple aux données. Dans cette partie du rapport ont été consolidés les résultats des chapitres précédents afin de dresser le compte des produits et charges technique de l'AMO du secteur privé. Ce CPC a fait dégager un résultat net positif et croissant à moyen terme, ce qui indique la viabilité du système de l'AMO du secteur privé.

Toutefois, il convient d'émettre certaines réserves quant aux résultats de notre étude. D'une part, le contexte économique et politique de l'assurance maladie au Maroc est en constante évolution. Cette évolution est illustrée par l'extension de la couverture à de nouvelles catégories socio-professionnelles, ainsi que l'élargissement du panier des soins remboursables. D'autre part, les différentes projections ont été entravées par le fait que les données sur l'assiette et la consommation mises à notre disposition ne sont relatifs qu'à l'année 2012.

Bibliographie

- Dahir n° 1-02-296 du 25 rejev 1423 (3 octobre 2002) portant promulgation de la loi n° 65-00 portant code de la couverture médicale de base. Bulletin officiel n° 5058 du 16 ramadan 1423 (21 novembre 2002)
- A.KLUGMAN, Stuart, H.PANJER, Harry, E.WILLMOT, Gordon. Aggregate Loss Models. In *Loss Models: From Data to Decisions*. 4th ed. Hoboken, New Jersey : John Wiley & Sons, 2012, pp.173-182. ISBN : 978-1-118-31532-3.
- CAISSE NATIONALE DE SECURITE SOCIALE. *AMO - Assurance Maladie Obligatoire*. [en ligne]. Mis à jour en 2014. Disponible à l'adresse :<<http://www.cnss.ma/fr/content/amo-assurance-maladie-obligatoire>>
- AGRESTI , Alan. *Categorical Data Analysis*. 2nd ed. Gainesville, Florida : John Wiley & Sons, 2002. ISBN : 978-0-471-36093-3.
- MÜLLER, Marlene. Generalized Linear Models [en ligne]. 6 janvier 2004. [Consulté le 06.01.2014]. Disponible à l'adresse : <<http://marlenemueller.de/publications/HandbookCS.pdf>>
- Grasland, Claude. *Initiation Aux Methodes Statistiques En Sciences Sociales* [en ligne]. Mis à jour en octobre 2006. Disponible à l'adresse : <<http://grasland.script.univ-paris-diderot.fr/STAT98/STAT98.htm>>
- CENTRE DES LIAISONS EUROPEENNES ET INTERNATIONALES DE SECURITE SOCIALE. *Le régime marocain de sécurité sociale* [en ligne]. Mis à jour en octobre 2011. Disponible à l'adresse : <http://www.cleiss.fr/docs/regimes/regime_maroc.html>
- GESTEIRA COSTA FILHO, Ivan. *Mixture Models for the Analysis of Gene Expression: Integration of Multiple Experiments and Cluster Validation*. Th. doct. : Biologie. Berlin : Freien Universitat, 2008.

Annexes

I. Sorties Python

I.1. Régression linéaire des montants moyens remboursés relatifs aux soins ambulatoires et aux bénéficiaires non-ALD

WLS Regression Results						
=====						
Dep. Variable:	MONT_REM_M	R-squared:	0.701			
Model:	WLS	Adj. R-squared:	0.699			
Method:	Least Squares	F-statistic:	377.8			
Date:	Mon, 09 Jun 2014	Prob (F-statistic):	0.00			
Time:	01:11:37	Log-Likelihood:	-36128.			
No. Observations:	4217	AIC:	7.231e+04			
Df Residuals:	4190	BIC:	7.248e+04			
	Df Model:		26			
=====						
	coef	std err	t	P> t	[95.0% Conf. Int.]	

Intercept	367.6648	49.916	7.366	0.000	269.803	465.526
SEXE[T.M]	-86.5339	4.013	-21.561	0.000	-94.402	-78.666
SECTEUR_ETABL[T.PU]	-232.6106	5.487	-42.391	0.000	-243.368	-221.853
TYPE_ASS[T.I]	-20.5397	23.989	-0.856	0.392	-67.570	26.491
TYPE_ASS[T.P1]	44.0952	6.553	6.729	0.000	31.248	56.943
TYPE_ASS[T.P2]	1.0600	9.823	0.108	0.914	-18.199	20.319
TYPE_ASS[T.P3]	6.5291	195.062	0.033	0.973	-375.896	388.954
TYPE_BEN[T.CA]	-91.9592	4.853	-18.948	0.000	-101.474	-82.444
TYPE_BEN[T.CM]	98.9237	71.696	1.380	0.168	-41.639	239.486
TYPE_BEN[T.EA]	-52.5610	7.872	-6.677	0.000	-67.995	-37.127
TYPE_BEN[T.EM]	-39.9316	225.788	-0.177	0.860	-482.596	402.733
REGION[T.02]	-3.1901	55.299	-0.058	0.954	-111.606	105.226
REGION[T.03]	-6.2383	55.956	-0.111	0.911	-115.941	103.464
REGION[T.04]	-4.5944	49.354	-0.093	0.926	-101.355	92.166
REGION[T.05]	23.3124	49.630	0.470	0.639	-73.989	120.614
REGION[T.06]	5.0682	49.762	0.102	0.919	-92.491	102.628
REGION[T.07]	55.1249	49.517	1.113	0.266	-41.954	152.204
REGION[T.08]	41.3349	49.622	0.833	0.405	-55.950	138.620
REGION[T.09]	138.3794	49.141	2.816	0.005	42.036	234.723
REGION[T.10]	87.6095	49.313	1.777	0.076	-9.070	184.289
REGION[T.11]	35.9256	49.667	0.723	0.470	-61.449	133.300
REGION[T.12]	19.5455	50.795	0.385	0.700	-80.040	119.131
REGION[T.13]	26.1613	49.653	0.527	0.598	-71.185	123.508
REGION[T.14]	90.3753	49.640	1.821	0.069	-6.946	187.696
REGION[T.15]	17.7546	50.818	0.349	0.727	-81.875	117.385
REGION[T.16]	17.0352	49.441	0.345	0.730	-79.895	113.965
TR_AGE	25.7445	0.963	26.738	0.000	23.857	27.632
=====						
Omnibus:	5211.926	Durbin-Watson:	1.932			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2273818.660			
Skew:	6.193	Prob(JB):	0.00			

I.3. Régression linéaire des montants moyens remboursés relatifs aux soins ambulatoires et aux bénéficiaires ALD

WLS Regression Results

```

=====
Dep. Variable:          MONT REM M    R-squared:                0.123
Model:                  WLS           Adj. R-squared:           0.115
Method:                 Least Squares F-statistic:              15.45
Date:                   Mon, 09 Jun 2014 Prob (F-statistic):       4.61e-64
Time:                   01:13:56     Log-Likelihood:          -31301.
No. Observations:      2902          AIC:                     6.266e+04
Df Residuals:          2875          BIC:                     6.282e+04

                                Df Model:                26
=====

```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
Intercept	5464.3730	1664.678	3.283	0.001	2200.290	8728.456
SEXE[T.M]	295.6799	103.306	2.862	0.004	93.118	498.242
SECTEUR ETABL[T.PU]	-1332.0033	92.875	-14.342	0.000	-1514.112	-1149.895
TYPE_ASS[T.I]	-1000.7485	672.985	-1.487	0.137	-2320.330	318.833
TYPE_ASS[T.P1]	170.1171	124.388	1.368	0.172	-73.781	414.015
TYPE_ASS[T.P2]	53.1528	169.884	0.313	0.754	-279.955	386.260
TYPE_ASS[T.P3]	-902.8025	2599.882	-0.347	0.728	-6000.624	4195.019
TYPE_BEN[T.CA]	-544.5428	111.843	-4.869	0.000	-763.844	-325.242
TYPE_BEN[T.CM]	486.7693	727.831	0.669	0.504	-940.354	1913.893
TYPE_BEN[T.EA]	-1910.1323	261.592	-7.302	0.000	-2423.060	-1397.205
TYPE_BEN[T.EM]	-840.0885	1.18e+04	-0.071	0.943	-2.39e+04	2.22e+04
REGION[T.02]	-126.9246	1895.336	-0.067	0.947	-3843.279	3589.430
REGION[T.03]	128.4114	1941.524	0.066	0.947	-3678.508	3935.330
REGION[T.04]	-194.5031	1651.147	-0.118	0.906	-3432.054	3043.048
REGION[T.05]	-485.6520	1654.198	-0.294	0.769	-3729.186	2757.883
REGION[T.06]	-576.9936	1660.472	-0.347	0.728	-3832.829	2678.842
REGION[T.07]	-422.9330	1653.866	-0.256	0.798	-3665.817	2819.951
REGION[T.08]	-397.0972	1653.673	-0.240	0.810	-3639.602	2845.408
REGION[T.09]	-27.8606	1646.674	-0.017	0.987	-3256.642	3200.921
REGION[T.10]	-88.1108	1649.067	-0.053	0.957	-3321.583	3145.362
REGION[T.11]	-760.7794	1655.859	-0.459	0.646	-4007.569	2486.011
REGION[T.12]	-1087.0733	1683.687	-0.646	0.519	-4388.429	2214.282
REGION[T.13]	-78.8302	1655.598	-0.048	0.962	-3325.109	3167.449
REGION[T.14]	-256.1224	1654.442	-0.155	0.877	-3500.134	2987.889
REGION[T.15]	254.6702	1680.267	0.152	0.880	-3039.979	3549.319
REGION[T.16]	-332.3524	1652.737	-0.201	0.841	-3573.021	2908.316
TR_AGE	-152.7423	23.423	-6.521	0.000	-198.669	-106.815

```

=====
Omnibus:                5122.368    Durbin-Watson:           1.947
Prob(Omnibus):          0.000     Jarque-Bera (JB):       10707400.800
Skew:                   12.050     Prob(JB):                0.00
Kurtosis:               299.599     Cond. No.                4.50e+03
=====

```

I.4. Régression linéaire des montants moyens remboursés relatifs aux soins hospitaliers et aux bénéficiaires ALD

WLS Regression Results

```

=====
Dep. Variable:          MONT REM M      R-squared:                0.272
Model:                  WLS             Adj. R-squared:           0.260
Method:                 Least Squares   F-statistic:             22.80
Date:                   Mon, 09 Jun 2014 Prob (F-statistic):       2.70e-87
Time:                   01:15:53       Log-Likelihood:          -19302.
No. Observations:      1550           AIC:                     3.866e+04
Df Residuals:          1524           BIC:                     3.879e+04

                                Df Model:                25
=====

```

	coef	std err	t	P> t	[95.0% Conf. Int.]	
Intercept	1.02e+05		2.11e+04	4.841	0.000	6.07e+04 1.43e+05
SEXE[T.M]	5379.6966		1415.538	3.800	0.000	2603.087 8156.306
SECTEUR ETABL[T.PU]	-2.008e+04		1533.952	-13.093	0.000	-2.31e+04 -1.71e+04
TYPE_ASS[T.I]	-9249.3571		7634.830	-1.211	0.226	-2.42e+04 5726.529
TYPE_ASS[T.P1]	1853.7622		1677.281	1.105	0.269	-1436.261 5143.786
TYPE_ASS[T.P2]	3971.3881		2449.984	1.621	0.105	-834.310 8777.086
TYPE_ASS[T.P3]	-1484.7670		4.45e+04	-0.033	0.973	-8.88e+04 8.59e+04
TYPE_BEN[T.CA]	1507.9980		1541.566	0.978	0.328	-1515.818 4531.814
TYPE_BEN[T.CM]	-1048.6335		9169.637	-0.114	0.909	-1.9e+04 1.69e+04
TYPE_BEN[T.EA]	-2.516e+04		4051.001	-6.212	0.000	-3.31e+04 -1.72e+04
REGION[T.02]	-9981.3505		2.49e+04	-0.401	0.688	-5.88e+04 3.88e+04
REGION[T.03]	-1.737e+04		2.6e+04	-0.667	0.505	-6.84e+04 3.37e+04
REGION[T.04]	-1.903e+04		2.09e+04	-0.911	0.363	-6e+04 2.2e+04
REGION[T.05]	-2.375e+04		2.1e+04	-1.133	0.257	-6.49e+04 1.74e+04
REGION[T.06]	-2.169e+04		2.11e+04	-1.029	0.304	-6.3e+04 1.97e+04
REGION[T.07]	-1.002e+04		2.09e+04	-0.478	0.632	-5.11e+04 3.11e+04
REGION[T.08]	-1.918e+04		2.1e+04	-0.915	0.360	-6.03e+04 2.19e+04
REGION[T.09]	-2.134e+04		2.08e+04	-1.024	0.306	-6.22e+04 1.95e+04
REGION[T.10]	-1.116e+04		2.09e+04	-0.534	0.593	-5.21e+04 2.98e+04
REGION[T.11]	-1.147e+04		2.1e+04	-0.546	0.585	-5.27e+04 2.97e+04
REGION[T.12]	-1230.2786		2.13e+04	-0.058	0.954	-4.31e+04 4.06e+04
REGION[T.13]	-1.736e+04		2.1e+04	-0.827	0.408	-5.85e+04 2.38e+04
REGION[T.14]	-1.778e+04		2.1e+04	-0.848	0.397	-5.89e+04 2.34e+04
REGION[T.15]	-1.562e+04		2.14e+04	-0.730	0.465	-5.76e+04 2.63e+04
REGION[T.16]	-1.26e+04		2.09e+04	-0.602	0.547	-5.36e+04 2.84e+04
TR AGE	-3862.4297		302.075	-12.786	0.000	-4454.957 -3269.902

```

=====
Omnibus:                1044.902      Durbin-Watson:            1.864
Prob(Omnibus):          0.000        Jarque-Bera (JB):         44808.962
Skew:                   2.574        Prob(JB):                 0.00
Kurtosis:               28.833       Cond. No.:                2.31e+03
=====

```

II. Sorties SAS

II.5. Modèle simple des remboursements des soins ambulatoires relatifs aux non-ALD

Model Information	
Data Set	GLM_AMBU.AMBU_NON_ALD
Distribution	Gamma
Link Function	Log
Dependent Variable	MONT_REM_M
Scale Weight Variable	EFFECT_SINISTR

Number of Observations Read	930
Number of Observations Used	930
Sum of Weights	533086

Class Level Information		
Class	Levels	Values
TYPE_ASS	3	A I P
TYPE_BEN	3	A C E
SEXE	2	F M
REGION	16	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	908	45132.8366	49.7058
Scaled Deviance	908	984.7882	1.0846
Pearson Chi-Square	908	68733.5585	75.6978
Scaled Pearson X2	908	1499.7505	1.6517
Log Likelihood		-8229.9030	
Full Log Likelihood		-8229.9030	
AIC (smaller is better)		16505.8060	
AICC (smaller is better)		16507.0246	
BIC (smaller is better)		16617.0153	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	7.8466	0.0635	7.7220	7.9711	15246.8	<.0001
TYPE_ASS	A	1	0.0417	0.0386	-0.0340	0.1173	1.17	0.2802
TYPE_ASS	I	1	-0.4640	0.2810	-1.0147	0.0867	2.73	0.0987
TYPE_ASS	P	0	0.0000	0.0000	0.0000	0.0000	.	.
TYPE_BEN	A	1	0.2557	0.0452	0.1671	0.3443	31.99	<.0001
TYPE_BEN	C	1	0.2206	0.0449	0.1326	0.3087	24.14	<.0001
TYPE_BEN	E	0	0.0000	0.0000	0.0000	0.0000	.	.
SEXE	F	1	0.0485	0.0216	0.0061	0.0909	5.02	0.0251
SEXE	M	0	0.0000	0.0000	0.0000	0.0000	.	.
TR_AGE		1	0.0430	0.0057	0.0318	0.0542	56.39	<.0001
REGION	1	1	-2.1188	0.5130	-3.1243	-1.1133	17.06	<.0001
REGION	2	1	-1.2790	0.1601	-1.5927	-0.9652	63.82	<.0001
REGION	3	1	-0.4364	0.1670	-0.7638	-0.1090	6.82	0.0090
REGION	4	1	0.6217	0.0456	0.5323	0.7110	186.04	<.0001
REGION	5	1	-0.0180	0.0539	-0.1236	0.0877	0.11	0.7391
REGION	6	1	0.2370	0.0578	0.1237	0.3502	16.81	<.0001
REGION	7	1	0.0924	0.0506	-0.0067	0.1916	3.34	0.0677
REGION	8	1	0.5694	0.0538	0.4639	0.6749	111.87	<.0001
REGION	9	1	-0.7630	0.0375	-0.8364	-0.6895	414.75	<.0001
REGION	10	1	0.0828	0.0448	-0.0050	0.1705	3.42	0.0646
REGION	11	1	0.2309	0.0548	0.1236	0.3382	17.78	<.0001
REGION	12	1	0.0494	0.0815	-0.1103	0.2092	0.37	0.5442
REGION	13	1	0.4832	0.0552	0.3750	0.5914	76.59	<.0001
REGION	14	1	-0.2877	0.0548	-0.3950	-0.1804	27.61	<.0001
REGION	15	1	-0.5692	0.0840	-0.7338	-0.4046	45.94	<.0001
REGION	16	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	0.0218	0.0010	0.0200	0.0238		

Note: The scale parameter was estimated by maximum likelihood.

II.6. Modèle retenu des remboursements des soins ambulatoires relatifs aux non-ALD

Model Information	
Data Set	GLM_AMBU.AMBU_NON_ALD
Distribution	Gamma
Link Function	Log
Dependent Variable	MONT_REM_M
Scale Weight Variable	EFFECT_SINISTR

Number of Observations Read	930
Number of Observations Used	930
Sum of Weights	533086

Class Level Information		
Class	Levels	Values
TYPE_ASS	3	A I P
TYPE_BEN	3	A C E
SEXE	2	F M
REGION	16	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Criteria For Assessing Goodness Of Fit			
Criterion	DF	Value	Value/DF
Deviance	901	41847.7300	46.4459
Scaled Deviance	901	981.3175	1.0891
Pearson Chi-Square	901	64531.5587	71.6222
Scaled Pearson X2	901	1513.2469	1.6795
Log Likelihood		-8192.7577	
Full Log Likelihood		-8192.7577	
AIC (smaller is better)		16445.5155	
AICC (smaller is better)		16447.5845	
BIC (smaller is better)		16590.5710	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept			1	7.4476	0.0847	7.2816	7.6136	7731.60	<.0001
TYPE_ASS	A		1	0.5020	0.0733	0.3583	0.6456	46.92	<.0001
TYPE_ASS	I		1	-0.3302	0.7383	-1.7773	1.1169	0.20	0.6547
TYPE_ASS	P		0	0.0000	0.0000	0.0000	0.0000	.	.
TYPE_BEN	A		1	0.6359	0.1116	0.4173	0.8546	32.49	<.0001
TYPE_BEN	C		1	-0.0121	0.1847	-0.3741	0.3500	0.00	0.9479
TYPE_BEN	E		0	0.0000	0.0000	0.0000	0.0000	.	.
SEXE	F		1	-0.0231	0.0407	-0.1029	0.0567	0.32	0.5703
SEXE	M		0	0.0000	0.0000	0.0000	0.0000	.	.
TR_AGE			1	0.0465	0.0076	0.0316	0.0614	37.36	<.0001
REGION	1		1	-2.1147	0.4949	-3.0847	-1.1446	18.26	<.0001
REGION	2		1	-1.2745	0.1545	-1.5772	-0.9717	68.08	<.0001
REGION	3		1	-0.4350	0.1612	-0.7509	-0.1192	7.29	0.0069
REGION	4		1	0.6265	0.0440	0.5402	0.7127	202.59	<.0001
REGION	5		1	-0.0137	0.0520	-0.1156	0.0883	0.07	0.7929
REGION	6		1	0.2436	0.0558	0.1343	0.3530	19.07	<.0001
REGION	7		1	0.0996	0.0488	0.0039	0.1953	4.16	0.0414
REGION	8		1	0.5783	0.0520	0.4764	0.6802	123.82	<.0001
REGION	9		1	-0.7587	0.0362	-0.8296	-0.6878	440.31	<.0001
REGION	10		1	0.0863	0.0432	0.0017	0.1710	3.99	0.0457
REGION	11		1	0.2333	0.0528	0.1298	0.3369	19.50	<.0001
REGION	12		1	0.0517	0.0787	-0.1024	0.2059	0.43	0.5106
REGION	13		1	0.4860	0.0533	0.3815	0.5904	83.18	<.0001
REGION	14		1	-0.2878	0.0528	-0.3913	-0.1842	29.66	<.0001
REGION	15		1	-0.5667	0.0810	-0.7255	-0.4078	48.90	<.0001
REGION	16		0	0.0000	0.0000	0.0000	0.0000	.	.
TYPE_ASS*TYPE_BEN	A	A	1	-0.5297	0.0819	-0.6903	-0.3692	41.82	<.0001
TYPE_ASS*TYPE_BEN	A	C	1	-0.5240	0.0864	-0.6935	-0.3546	36.75	<.0001
TYPE_ASS*TYPE_BEN	A	E	0	0.0000	0.0000	0.0000	0.0000	.	.
TYPE_ASS*TYPE_BEN	I	A	1	-0.5245	0.8936	-2.2759	1.2269	0.34	0.5572
TYPE_ASS*TYPE_BEN	I	C	1	-0.0440	0.8186	-1.6484	1.5604	0.00	0.9571
TYPE_ASS*TYPE_BEN	I	E	0	0.0000	0.0000	0.0000	0.0000	.	.
TYPE_ASS*TYPE_BEN	P	A	0	0.0000	0.0000	0.0000	0.0000	.	.
TYPE_ASS*TYPE_BEN	P	C	0	0.0000	0.0000	0.0000	0.0000	.	.
TYPE_ASS*TYPE_BEN	P	E	0	0.0000	0.0000	0.0000	0.0000	.	.
TR_AGE*SEXE	F		1	-0.0134	0.0079	-0.0288	0.0019	2.93	0.0869
TR_AGE*SEXE	M		0	0.0000	0.0000	0.0000	0.0000	.	.
TYPE_BEN*SEXE	A	F	1	0.2704	0.0762	0.1211	0.4198	12.60	0.0004
TYPE_BEN*SEXE	A	M	0	0.0000	0.0000	0.0000	0.0000	.	.
TYPE_BEN*SEXE	C	F	1	0.8484	0.1656	0.5239	1.1730	26.25	<.0001
TYPE_BEN*SEXE	C	M	0	0.0000	0.0000	0.0000	0.0000	.	.
TYPE_BEN*SEXE	E	F	0	0.0000	0.0000	0.0000	0.0000	.	.

Analysis Of Maximum Likelihood Parameter Estimates									
Parameter			DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
TYPE_BEN*SEXE	E	M	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale			1	0.0234	0.0010	0.0215	0.0256		

Note: The scale parameter was estimated by maximum likelihood.

II.7. Modèle simple des remboursements des soins ambulatoires relatifs aux ALD

Model Information	
Data Set	GLM_AMBU.AMBU_ALD
Response Variable	MONT_REM_M
Response Distribution	Lognormal
Link Function	Identity
Variance Function	Default
Weight Variable	EFFECT_SINISTR
Variance Matrix	Diagonal
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Residual

Class Level Information		
Class	Levels	Values
TYPE_ASS	2	A P
TYPE_BEN	3	A C E
SEXE	2	F M
REGION	13	4 5 6 7 8 9 10 11 12 13 14 15 16

Number of Observations Read	409
Number of Observations Used	409

Dimensions	
Covariance Parameters	1
Columns in X	22
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	409

Optimization Information	
Optimization Technique	None
Parameters	19
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Not Profiled

Fit Statistics	
-2 Res Log Likelihood	713.57
AIC (smaller is better)	751.57
AICC (smaller is better)	753.62
BIC (smaller is better)	826.98
CAIC (smaller is better)	845.98
HQIC (smaller is better)	781.46
Pearson Chi-Square	12439.85
Pearson Chi-Square / DF	31.82

Parameter Estimates									
Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	DF	t Value	Pr > t
Intercept					11.4004	0.2182	391	52.25	<.0001
TYPE_ASS	A				-0.04728	0.07360	391	-0.64	0.5210
TYPE_ASS	P				0
TYPE_BEN		A			0.2651	0.1956	391	1.36	0.1762
TYPE_BEN		C			0.2852	0.1950	391	1.46	0.1443
TYPE_BEN		E			0
SEXE			F		-0.4378	0.04677	391	-9.36	<.0001
SEXE			M		0
TR_AGE					0.04389	0.01443	391	3.04	0.0025
REGION				4	0.4227	0.1119	391	3.78	0.0002
REGION				5	-0.1138	0.1251	391	-0.91	0.3636
REGION				6	-0.03710	0.1566	391	-0.24	0.8128
REGION				7	-0.4738	0.1234	391	-3.84	0.0001
REGION				8	0.1642	0.1232	391	1.33	0.1832
REGION				9	0.2681	0.08862	391	3.03	0.0026
REGION				10	0.3525	0.1050	391	3.36	0.0009
REGION				11	-0.2881	0.1322	391	-2.18	0.0300
REGION				12	-0.8989	0.2657	391	-3.38	0.0008
REGION				13	-0.07617	0.1356	391	-0.56	0.5746
REGION				14	-0.5626	0.1306	391	-4.31	<.0001
REGION				15	-0.7787	0.2905	391	-2.68	0.0077
REGION				16	0
Scale					31.8155	2.2754	.	.	.

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
TYPE_ASS	1	391	0.41	0.5210
TYPE_BEN	2	391	1.09	0.3373
SEXE	1	391	87.61	<.0001

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
TR_AGE	1	391	9.25	0.0025
REGION	12	391	17.00	<.0001

II.8. Modèle retenu des remboursements des soins ambulatoires relatifs aux ALD

Model Information	
Data Set	GLM_AMBU.AMBU_ALD
Response Variable	MONT_REM_M
Response Distribution	Lognormal
Link Function	Identity
Variance Function	Default
Weight Variable	EFFECT_SINISTR
Variance Matrix	Diagonal
Estimation Technique	Restricted Maximum Likelihood
Degrees of Freedom Method	Residual

Class Level Information		
Class	Levels	Values
TYPE_ASS	2	A P
TYPE_BEN	3	A C E
SEXE	2	F M
REGION	13	4 5 6 7 8 9 10 11 12 13 14 15 16

Number of Observations Read	409
Number of Observations Used	409

Dimensions	
Covariance Parameters	1
Columns in X	24
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	409

Optimization Information	
Optimization Technique	None
Parameters	20
Lower Boundaries	1
Upper Boundaries	0
Fixed Effects	Not Profiled

Fit Statistics	
-2 Res Log Likelihood	720.06
AIC (smaller is better)	760.06
AICC (smaller is better)	762.33
BIC (smaller is better)	839.38
CAIC (smaller is better)	859.38
HQIC (smaller is better)	791.50
Pearson Chi-Square	12439.60
Pearson Chi-Square / DF	31.90

Parameter Estimates									
Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	DF	t Value	Pr > t
Intercept					11.4097	0.2427	390	47.00	<.0001
TYPE_ASS	A				-0.04773	0.07387	390	-0.65	0.5186
TYPE_ASS	P				0
TYPE_BEN		A			0.2678	0.1982	390	1.35	0.1775
TYPE_BEN		C			0.2880	0.1977	390	1.46	0.1460
TYPE_BEN		E			0
SEXE			F		-0.4565	0.2180	390	-2.09	0.0369
SEXE			M		0
TR_AGE					0.04303	0.01739	390	2.47	0.0138
REGION				4	0.4226	0.1121	390	3.77	0.0002
REGION				5	-0.1137	0.1252	390	-0.91	0.3644
REGION				6	-0.03702	0.1568	390	-0.24	0.8135
REGION				7	-0.4738	0.1235	390	-3.84	0.0001
REGION				8	0.1642	0.1233	390	1.33	0.1838
REGION				9	0.2682	0.08873	390	3.02	0.0027
REGION				10	0.3526	0.1052	390	3.35	0.0009
REGION				11	-0.2881	0.1324	390	-2.18	0.0302
REGION				12	-0.8986	0.2661	390	-3.38	0.0008
REGION				13	-0.07613	0.1358	390	-0.56	0.5753
REGION				14	-0.5625	0.1307	390	-4.30	<.0001
REGION				15	-0.7785	0.2909	390	-2.68	0.0078
REGION				16	0
TR_AGE*SEXE			F		0.001367	0.01552	390	0.09	0.9299
TR_AGE*SEXE			M		0
Scale					31.8964	2.2841	.	.	.

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
TYPE_ASS	1	390	0.42	0.5186
TYPE_BEN	2	390	1.08	0.3401
SEXE	1	390	4.39	0.0369
TR_AGE	1	390	9.00	0.0029
REGION	12	390	16.96	<.0001
TR_AGE*SEXE	1	390	0.01	0.9299

II.9. Modèle simple des remboursements des soins hospitaliers relatifs aux non-ALD

Model Information	
Data Set	FMMODEL.HOSPIT_NON_ALD
Response Variable	MONT_REM_M
Weight Variable	EFFECT_SINISTR
Type of Model	Homogeneous Regression Mixture
Distribution	Gamma
Min Components	1
Max Components	6
Link Function	Log
Estimation Method	Maximum Likelihood

Class Level Information		
Class	Levels	Values
TYPE_ASS	3	A I P
TYPE_BEN	3	A C E
SEXE	2	F M
REGION	13	10 11 12 13 14 15 16 4 5 6 7 8 9

Number of Observations Read	379
Number of Observations Used	379

Component Description for Mixture Models	
Model ID	Gamma
1	1
2	2
3	3
4	4
5	5
6	6

Component Evaluation for Mixture Models											
Model ID	Number of				-2 Log L	AIC	AICC	BIC	Pearson	Max Gradient	
	Components		Parameters								
	Total	Eff.	Total	Eff.							
1	1	1	20	20	8319.85	8359.85	8362.20	8438.60	534.37	2.099E-7	
2	2	2	41	41						27.946	
3	3	3	62	62	8024.46	8148.46	8173.18	8392.58	778.60	0.000076	
4	4	4	83	83						2.257	
5	5	5	104	104						3.098	
6	6	6	125	125						1050	

The model with 3 components (ID=3) was selected as 'best' based on the AIC statistic. Models that failed to converge are indicated with blank values for the likelihood-based criteria.

Convergence criterion (GCONV=1E-8) satisfied.

Fit Statistics	
-2 Log Likelihood	8024.5
AIC (smaller is better)	8148.5
AICC (smaller is better)	8173.2
BIC (smaller is better)	8392.6
Pearson Statistic	778.6
Effective Parameters	62
Effective Components	3

Parameter Estimates for 'Gamma' Model									
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value	Pr > z
1	Intercept					9.1289	0.01034	882.67	<.0001
1	TYPE_ASS	A				-0.09007	0.000980	-91.90	<.0001
1	TYPE_ASS	I				-0.6076	0	.	.
1	TYPE_ASS	P				0	.	.	.
1	TYPE_BEN		A			0.09064	0	.	.
1	TYPE_BEN		C			0.4413	0	.	.
1	TYPE_BEN		E			0	.	.	.
1	SEXE			F		-0.1380	0	.	.
1	SEXE			M		0	.	.	.
1	TR_AGE					0.04616	0.001388	33.27	<.0001
1	REGION				10	0.5413	0	.	.
1	REGION				11	0.8146	0	.	.
1	REGION				12	0.5829	0	.	.
1	REGION				13	0.9209	0	.	.
1	REGION				14	-0.1711	0	.	.

Parameter Estimates for 'Gamma' Model									
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value	Pr > z
1	REGION				15	0.04481	0	.	.
1	REGION				16	0.7969	0	.	.
1	REGION				4	1.4140	0	.	.
1	REGION				5	0.6208	0	.	.
1	REGION				6	0.9707	0	.	.
1	REGION				7	0.5311	0	.	.
1	REGION				8	0.9626	0	.	.
1	REGION				9	0	.	.	.
2	Intercept					8.5045	0.1101	77.24	<.0001
2	TYPE_ASS	A				0.01756	0.07668	0.23	0.8189
2	TYPE_ASS	I				-0.01685	0.5827	-0.03	0.9769
2	TYPE_ASS	P				0	.	.	.
2	TYPE_BEN		A			0.3472	0.09061	3.83	0.0001
2	TYPE_BEN		C			0.4627	0.09181	5.04	<.0001
2	TYPE_BEN		E			0	.	.	.
2	SEXE			F		-0.06672	0.04916	-1.36	0.1747
2	SEXE			M		0	.	.	.
2	TR_AGE					0.03822	0.01145	3.34	0.0008
2	REGION				10	0.8355	0.05205	16.05	<.0001
2	REGION				11	1.3770	0.07901	17.43	<.0001
2	REGION				12	1.1020	0.1545	7.13	<.0001
2	REGION				13	1.1442	0.09486	12.06	<.0001
2	REGION				14	0.3013	0.08505	3.54	0.0004
2	REGION				15	0.4840	0.1228	3.94	<.0001
2	REGION				16	1.0642	0.08286	12.84	<.0001
2	REGION				4	1.7690	0.06203	28.52	<.0001
2	REGION				5	1.0061	0.07873	12.78	<.0001
2	REGION				6	1.4229	0.08054	17.67	<.0001
2	REGION				7	0.7671	0.06867	11.17	<.0001
2	REGION				8	1.3087	0.06868	19.06	<.0001
2	REGION				9	0	.	.	.
3	Intercept					9.9859	0.3394	29.42	<.0001
3	TYPE_ASS	A				-0.1671	0.2598	-0.64	0.5200
3	TYPE_ASS	I				-1.0383	3.8291	-0.27	0.7863
3	TYPE_ASS	P				0	.	.	.
3	TYPE_BEN		A			-0.4341	0.2678	-1.62	0.1050
3	TYPE_BEN		C			0.1480	0.2512	0.59	0.5559
3	TYPE_BEN		E			0	.	.	.
3	SEXE			F		-0.3755	0.1251	-3.00	0.0027
3	SEXE			M		0	.	.	.
3	TR_AGE					0.06023	0.03331	1.81	0.0705
3	REGION				10	0.6238	0.1654	3.77	0.0002

Parameter Estimates for 'Gamma' Model									
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value	Pr > z
3	REGION				11	0.06178	0.3924	0.16	0.8749
3	REGION				12	0.1678	0.7390	0.23	0.8204
3	REGION				13	0.7081	0.1864	3.80	0.0001
3	REGION				14	-0.4318	0.2166	-1.99	0.0462
3	REGION				15	-0.2972	0.7473	-0.40	0.6908
3	REGION				16	0.7188	0.1618	4.44	<.0001
3	REGION				4	1.2373	0.1432	8.64	<.0001
3	REGION				5	0.4656	0.2495	1.87	0.0620
3	REGION				6	0.6213	0.4118	1.51	0.1313
3	REGION				7	0.5436	0.1550	3.51	0.0005
3	REGION				8	0.9697	0.2760	3.51	0.0004
3	REGION				9	0	.	.	.
1	Scale Parameter					38520	0		
2	Scale Parameter					0.2159	0.03306		
3	Scale Parameter					0.03092	0.004401		

Parameter Estimates for Mixing Probabilities						
Component	Effect	Linked Scale				Probability
		Estimate	Standard Error	z Value	Pr > z	
1	Intercept	-3.8840	0.7396	-5.25	<.0001	0.0063
2	Intercept	0.8065	0.2036	3.96	<.0001	0.6870

II.10. Modèle retenu des remboursements des soins hospitaliers relatifs aux non-ALD

Model Information	
Data Set	FMMODEL.HOSPIT_NON_ALD
Response Variable	MONT_REM_M
Weight Variable	EFFECT_SINISTR
Type of Model	Homogeneous Regression Mixture
Distribution	Gamma
Min Components	1
Max Components	6
Link Function	Log
Estimation Method	Maximum Likelihood

Class Level Information		
Class	Levels	Values
TYPE_ASS	3	A I P
TYPE_BEN	3	A C E
SEXE	2	F M
REGION	13	10 11 12 13 14 15 16 4 5 6 7 8 9

Number of Observations Read	379
Number of Observations Used	379

Component Description for Mixture Models	
Model ID	Gamma
1	1
2	2
3	3
4	4
5	5
6	6

Component Evaluation for Mixture Models										
Model ID	Number of				-2 Log L	AIC	AICC	BIC	Pearson	Max Gradient
	Components		Parameters							
	Total	Eff.	Total	Eff.						
1	1	1	22	22	8280.59	8324.59	8327.43	8411.22	531.29	0.000090
2	2	2	45	45						0.086
3	3	3	68	68						22.238

Component Evaluation for Mixture Models											
Model ID	Number of				-2 Log L	AIC	AICC	BIC	Pearson	Max Gradient	
	Components		Parameters								
	Total	Eff.	Total	Eff.							
4	4	4	91	91	7910.62	8092.62	8150.96	8450.93	914.32	5.520	
5	5	5	114	114						2.785	
6	6	6	137	137						27.862	

The model with 4 components (ID=4) was selected as 'best' based on the AIC statistic. Models that failed to converge are indicated with blank values for the likelihood-based criteria.

Convergence criterion (GCONV=1E-8) satisfied.

Fit Statistics	
-2 Log Likelihood	7910.6
AIC (smaller is better)	8092.6
AICC (smaller is better)	8151.0
BIC (smaller is better)	8450.9
Pearson Statistic	914.3
Effective Parameters	91
Effective Components	4

Parameter Estimates for 'Gamma' Model								
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
1	Intercept					8.5927	17378	0.00
1	TYPE_ASS	A				0.8301	0	.
1	TYPE_ASS	I				-3.8411	0	.
1	TYPE_ASS	P				0	.	.
1	TYPE_BEN		A			1.0880	0	.
1	TYPE_BEN		C			0.1323	0	.
1	TYPE_BEN		E			0	.	.
1	SEXE			F		0.5959	0	.
1	SEXE			M		0	.	.
1	TR_AGE					0.01710	2896.31	0.00
1	REGION				10	0.7161	0	.
1	REGION				11	-0.1156	0	.
1	REGION				12	-1.1509	0	.
1	REGION				13	0.2159	0	.
1	REGION				14	-1.0341	0	.
1	REGION				15	-1.6742	0	.
1	REGION				16	0.1473	0	.
1	REGION				4	0.7799	0	.

Parameter Estimates for 'Gamma' Model								
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
1	REGION				5	-0.2705	0	.
1	REGION				6	1.3843	0	.
1	REGION				7	-0.1843	0	.
1	REGION				8	0.1462	0	.
1	REGION				9	0	.	.
1	TYPE_ASS*TYPE_BEN	A	A			-0.5480	0	.
1	TYPE_ASS*TYPE_BEN	A	C			0.01269	0	.
1	TYPE_ASS*TYPE_BEN	A	E			0	.	.
1	TYPE_ASS*TYPE_BEN	I	C			0	.	.
1	TYPE_ASS*TYPE_BEN	P	A			0	.	.
1	TYPE_ASS*TYPE_BEN	P	C			0	.	.
1	TYPE_ASS*TYPE_BEN	P	E			0	.	.
2	Intercept					8.4603	0.3276	25.82
2	TYPE_ASS	A				0.1880	0.3291	0.57
2	TYPE_ASS	I				-0.01796	0.2647	-0.07
2	TYPE_ASS	P				0	.	.
2	TYPE_BEN		A			1.0248	0.3594	2.85
2	TYPE_BEN		C			0.8163	0.3683	2.22
2	TYPE_BEN		E			0	.	.
2	SEXE			F		-0.06988	0.04223	-1.65
2	SEXE			M		0	.	.
2	TR_AGE					-0.00276	0	.
2	REGION				10	0.7409	0.08420	8.80
2	REGION				11	0.8019	0.06357	12.61
2	REGION				12	0.7371	0.1367	5.39
2	REGION				13	1.1319	0.08339	13.57
2	REGION				14	-0.06831	0.1402	-0.49
2	REGION				15	0.06557	0.1739	0.38
2	REGION				16	0.6311	0.04831	13.06
2	REGION				4	1.5094	0.04978	30.32
2	REGION				5	0.9950	0.07065	14.08
2	REGION				6	0.9839	0.08526	11.54
2	REGION				7	0.6927	0.08107	8.54
2	REGION				8	1.2113	0.06854	17.67
2	REGION				9	0	.	.
2	TYPE_ASS*TYPE_BEN	A	A			-0.5274	0.3452	-1.53
2	TYPE_ASS*TYPE_BEN	A	C			0.4027	0.3451	1.17
2	TYPE_ASS*TYPE_BEN	A	E			0	.	.
2	TYPE_ASS*TYPE_BEN	I	C			0	.	.
2	TYPE_ASS*TYPE_BEN	P	A			0	.	.
2	TYPE_ASS*TYPE_BEN	P	C			0	.	.
2	TYPE_ASS*TYPE_BEN	P	E			0	.	.

Parameter Estimates for 'Gamma' Model								
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
3	Intercept					8.8727	0.2646	33.53
3	TYPE_ASS	A				0.1330	0.2674	0.50
3	TYPE_ASS	I				0.008515	0.2745	0.03
3	TYPE_ASS	P				0	.	.
3	TYPE_BEN		A			0.5024	0.2791	1.80
3	TYPE_BEN		C			0.3018	0.2918	1.03
3	TYPE_BEN		E			0	.	.
3	SEXE			F		-0.2496	0.04692	-5.32
3	SEXE			M		0	.	.
3	TR_AGE					0.03153	0	.
3	REGION				10	0.7193	0.05332	13.49
3	REGION				11	1.1449	0.08075	14.18
3	REGION				12	1.2293	0.2057	5.98
3	REGION				13	0.9056	0.1133	7.99
3	REGION				14	0.04419	0.08335	0.53
3	REGION				15	0.3707	0.1296	2.86
3	REGION				16	0.7922	0.04656	17.01
3	REGION				4	1.7920	0.04612	38.85
3	REGION				5	0.8625	0.08236	10.47
3	REGION				6	1.3933	0.06655	20.94
3	REGION				7	0.7902	0.04909	16.10
3	REGION				8	1.2746	0.04919	25.91
3	REGION				9	0	.	.
3	TYPE_ASS*TYPE_BEN	A	A			-0.2822	0.2715	-1.04
3	TYPE_ASS*TYPE_BEN	A	C			-0.07741	0.2844	-0.27
3	TYPE_ASS*TYPE_BEN	A	E			0	.	.
3	TYPE_ASS*TYPE_BEN	I	C			0	.	.
3	TYPE_ASS*TYPE_BEN	P	A			0	.	.
3	TYPE_ASS*TYPE_BEN	P	C			0	.	.
3	TYPE_ASS*TYPE_BEN	P	E			0	.	.
4	Intercept					9.0741	2.4112	3.76
4	TYPE_ASS	A				0.8524	2.4097	0.35
4	TYPE_ASS	I				-0.2496	2.2654	-0.11
4	TYPE_ASS	P				0	.	.
4	TYPE_BEN		A			0.5467	2.4836	0.22
4	TYPE_BEN		C			0.1310	2.4784	0.05
4	TYPE_BEN		E			0	.	.
4	SEXE			F		-0.3167	0.1554	-2.04
4	SEXE			M		0	.	.
4	TR_AGE					0.06831	0.05433	1.26
4	REGION				10	0.3522	0.1591	2.21
4	REGION				11	0.3626	0.3779	0.96

Parameter Estimates for 'Gamma' Model								
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
4	REGION				12	-0.3998	0.5940	-0.67
4	REGION				13	0.3133	0.2799	1.12
4	REGION				14	-0.6586	0.2440	-2.70
4	REGION				15	-0.5781	0.7988	-0.72
4	REGION				16	0.7441	0.1707	4.36
4	REGION				4	1.0635	0.2381	4.47
4	REGION				5	-0.06548	0.3220	-0.20
4	REGION				6	0.4701	0.4857	0.97
4	REGION				7	0.3768	0.3512	1.07
4	REGION				8	1.1098	0.4515	2.46
4	REGION				9	0	.	.
4	TYPE_ASS*TYPE_BEN	A	A			-1.0223	2.4220	-0.42
4	TYPE_ASS*TYPE_BEN	A	C			0.1483	2.4309	0.06
4	TYPE_ASS*TYPE_BEN	A	E			0	.	.
4	TYPE_ASS*TYPE_BEN	I	C			0	.	.
4	TYPE_ASS*TYPE_BEN	P	A			0	.	.
4	TYPE_ASS*TYPE_BEN	P	C			0	.	.
4	TYPE_ASS*TYPE_BEN	P	E			0	.	.
1	Scale Parameter					38520	0	
2	Scale Parameter					0.5356	0.2559	
3	Scale Parameter					0.4566	0.08384	
4	Scale Parameter					0.03807	0	

Parameter Estimates for 'Gamma' Model						
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
1	Intercept					0.9996
1	TYPE_ASS	A				.
1	TYPE_ASS	I				.
1	TYPE_ASS	P				.
1	TYPE_BEN		A			.
1	TYPE_BEN		C			.
1	TYPE_BEN		E			.
1	SEXE			F		.
1	SEXE			M		.
1	TR_AGE					1.0000
1	REGION				10	.
1	REGION				11	.
1	REGION				12	.
1	REGION				13	.
1	REGION				14	.
1	REGION				15	.
1	REGION				16	.

Parameter Estimates for 'Gamma' Model						
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
1	REGION				4	.
1	REGION				5	.
1	REGION				6	.
1	REGION				7	.
1	REGION				8	.
1	REGION				9	.
1	TYPE_ASS*TYPE_BEN	A	A			.
1	TYPE_ASS*TYPE_BEN	A	C			.
1	TYPE_ASS*TYPE_BEN	A	E			.
1	TYPE_ASS*TYPE_BEN	I	C			.
1	TYPE_ASS*TYPE_BEN	P	A			.
1	TYPE_ASS*TYPE_BEN	P	C			.
1	TYPE_ASS*TYPE_BEN	P	E			.
2	Intercept					<.0001
2	TYPE_ASS	A				0.5679
2	TYPE_ASS	I				0.9459
2	TYPE_ASS	P				.
2	TYPE_BEN		A			0.0043
2	TYPE_BEN		C			0.0267
2	TYPE_BEN		E			.
2	SEXE			F		0.0980
2	SEXE			M		.
2	TR_AGE					.
2	REGION				10	<.0001
2	REGION				11	<.0001
2	REGION				12	<.0001
2	REGION				13	<.0001
2	REGION				14	0.6260
2	REGION				15	0.7062
2	REGION				16	<.0001
2	REGION				4	<.0001
2	REGION				5	<.0001
2	REGION				6	<.0001
2	REGION				7	<.0001
2	REGION				8	<.0001
2	REGION				9	.
2	TYPE_ASS*TYPE_BEN	A	A			0.1266
2	TYPE_ASS*TYPE_BEN	A	C			0.2433
2	TYPE_ASS*TYPE_BEN	A	E			.
2	TYPE_ASS*TYPE_BEN	I	C			.
2	TYPE_ASS*TYPE_BEN	P	A			.
2	TYPE_ASS*TYPE_BEN	P	C			.
2	TYPE_ASS*TYPE_BEN	P	E			.

Parameter Estimates for 'Gamma' Model						
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
3	Intercept					<.0001
3	TYPE_ASS	A				0.6189
3	TYPE_ASS	I				0.9752
3	TYPE_ASS	P				.
3	TYPE_BEN		A			0.0718
3	TYPE_BEN		C			0.3010
3	TYPE_BEN		E			.
3	SEXE			F		<.0001
3	SEXE			M		.
3	TR_AGE					.
3	REGION				10	<.0001
3	REGION				11	<.0001
3	REGION				12	<.0001
3	REGION				13	<.0001
3	REGION				14	0.5960
3	REGION				15	0.0042
3	REGION				16	<.0001
3	REGION				4	<.0001
3	REGION				5	<.0001
3	REGION				6	<.0001
3	REGION				7	<.0001
3	REGION				8	<.0001
3	REGION				9	.
3	TYPE_ASS*TYPE_BEN	A	A			0.2985
3	TYPE_ASS*TYPE_BEN	A	C			0.7855
3	TYPE_ASS*TYPE_BEN	A	E			.
3	TYPE_ASS*TYPE_BEN	I	C			.
3	TYPE_ASS*TYPE_BEN	P	A			.
3	TYPE_ASS*TYPE_BEN	P	C			.
3	TYPE_ASS*TYPE_BEN	P	E			.
4	Intercept					0.0002
4	TYPE_ASS	A				0.7235
4	TYPE_ASS	I				0.9123
4	TYPE_ASS	P				.
4	TYPE_BEN		A			0.8258
4	TYPE_BEN		C			0.9578
4	TYPE_BEN		E			.
4	SEXE			F		0.0415
4	SEXE			M		.
4	TR_AGE					0.2086
4	REGION				10	0.0269
4	REGION				11	0.3373
4	REGION				12	0.5009

Parameter Estimates for 'Gamma' Model						
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
4	REGION				13	0.2630
4	REGION				14	0.0069
4	REGION				15	0.4693
4	REGION				16	<.0001
4	REGION				4	<.0001
4	REGION				5	0.8388
4	REGION				6	0.3332
4	REGION				7	0.2833
4	REGION				8	0.0140
4	REGION				9	.
4	TYPE_ASS*TYPE_BEN	A	A			0.6730
4	TYPE_ASS*TYPE_BEN	A	C			0.9514
4	TYPE_ASS*TYPE_BEN	A	E			.
4	TYPE_ASS*TYPE_BEN	I	C			.
4	TYPE_ASS*TYPE_BEN	P	A			.
4	TYPE_ASS*TYPE_BEN	P	C			.
4	TYPE_ASS*TYPE_BEN	P	E			.
1	Scale Parameter					
2	Scale Parameter					
3	Scale Parameter					
4	Scale Parameter					

Parameter Estimates for Mixing Probabilities						
Component	Effect	Linked Scale				Probability
		Estimate	Standard Error	z Value	Pr > z	
1	Intercept	-2.8360	0.4573	-6.20	<.0001	0.0133
2	Intercept	0.4510	0	.	.	0.3568
3	Intercept	0.5719	0.2027	2.82	0.0048	0.4026

II.11. Modèle simple des remboursements des soins hospitaliers relatifs aux ALD

Model Information	
Data Set	FMMODEL.HOSPIT_NON_ALD
Response Variable	MONT_REM_M
Weight Variable	EFFECT_SINISTR
Type of Model	Homogeneous Regression Mixture
Distribution	Gamma
Min Components	1
Max Components	6
Link Function	Log
Estimation Method	Maximum Likelihood

Class Level Information		
Class	Levels	Values
TYPE_ASS	3	A I P
TYPE_BEN	3	A C E
SEXE	2	F M
REGION	13	10 11 12 13 14 15 16 4 5 6 7 8 9

Number of Observations Read	379
Number of Observations Used	379

Component Description for Mixture Models	
Model ID	Gamma
1	1
2	2
3	3
4	4
5	5
6	6

Component Evaluation for Mixture Models										
Model ID	Number of				-2 Log L	AIC	AICC	BIC	Pearson	Max Gradient
	Components		Parameters							
	Total	Eff.	Total	Eff.						
1	1	1	20	20	8319.85	8359.85	8362.20	8438.60	534.37	2.099E-7
2	2	2	41	41						27.946
3	3	3	62	62	8024.46	8148.46	8173.18	8392.58	778.60	0.000076
4	4	4	83	83						2.257
5	5	5	104	104						3.098
6	6	6	125	125						1050

The model with 3 components (ID=3) was selected as 'best' based on the AIC statistic. Models that failed to converge are indicated with blank values for the likelihood-based criteria.

Convergence criterion (GCONV=1E-8) satisfied.

Fit Statistics	
-2 Log Likelihood	8024.5
AIC (smaller is better)	8148.5
AICC (smaller is better)	8173.2
BIC (smaller is better)	8392.6
Pearson Statistic	778.6
Effective Parameters	62
Effective Components	3

Parameter Estimates for 'Gamma' Model									
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value	Pr > z
1	Intercept					9.1289	0.01034	882.67	<.0001
1	TYPE_ASS	A				-0.09007	0.000980	-91.90	<.0001
1	TYPE_ASS	I				-0.6076	0	.	.
1	TYPE_ASS	P				0	.	.	.
1	TYPE_BEN		A			0.09064	0	.	.
1	TYPE_BEN		C			0.4413	0	.	.
1	TYPE_BEN		E			0	.	.	.
1	SEXE			F		-0.1380	0	.	.
1	SEXE			M		0	.	.	.
1	TR_AGE					0.04616	0.001388	33.27	<.0001
1	REGION				10	0.5413	0	.	.
1	REGION				11	0.8146	0	.	.

Parameter Estimates for 'Gamma' Model

Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value	Pr > z
1	REGION				12	0.5829	0	.	.
1	REGION				13	0.9209	0	.	.
1	REGION				14	-0.1711	0	.	.
1	REGION				15	0.04481	0	.	.
1	REGION				16	0.7969	0	.	.
1	REGION				4	1.4140	0	.	.
1	REGION				5	0.6208	0	.	.
1	REGION				6	0.9707	0	.	.
1	REGION				7	0.5311	0	.	.
1	REGION				8	0.9626	0	.	.
1	REGION				9	0	.	.	.
2	Intercept					8.5045	0.1101	77.24	<.0001
2	TYPE_ASS	A				0.01756	0.07668	0.23	0.8189
2	TYPE_ASS	I				-0.01685	0.5827	-0.03	0.9769
2	TYPE_ASS	P				0	.	.	.
2	TYPE_BEN		A			0.3472	0.09061	3.83	0.0001
2	TYPE_BEN		C			0.4627	0.09181	5.04	<.0001
2	TYPE_BEN		E			0	.	.	.
2	SEXE			F		-0.06672	0.04916	-1.36	0.1747
2	SEXE			M		0	.	.	.
2	TR_AGE					0.03822	0.01145	3.34	0.0008
2	REGION				10	0.8355	0.05205	16.05	<.0001
2	REGION				11	1.3770	0.07901	17.43	<.0001
2	REGION				12	1.1020	0.1545	7.13	<.0001
2	REGION				13	1.1442	0.09486	12.06	<.0001
2	REGION				14	0.3013	0.08505	3.54	0.0004
2	REGION				15	0.4840	0.1228	3.94	<.0001
2	REGION				16	1.0642	0.08286	12.84	<.0001
2	REGION				4	1.7690	0.06203	28.52	<.0001
2	REGION				5	1.0061	0.07873	12.78	<.0001
2	REGION				6	1.4229	0.08054	17.67	<.0001
2	REGION				7	0.7671	0.06867	11.17	<.0001
2	REGION				8	1.3087	0.06868	19.06	<.0001
2	REGION				9	0	.	.	.
3	Intercept					9.9859	0.3394	29.42	<.0001
3	TYPE_ASS	A				-0.1671	0.2598	-0.64	0.5200
3	TYPE_ASS	I				-1.0383	3.8291	-0.27	0.7863
3	TYPE_ASS	P				0	.	.	.
3	TYPE_BEN		A			-0.4341	0.2678	-1.62	0.1050

Parameter Estimates for 'Gamma' Model

Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value	Pr > z
3	TYPE_BEN		C			0.1480	0.2512	0.59	0.5559
3	TYPE_BEN		E			0	.	.	.
3	SEXE			F		-0.3755	0.1251	-3.00	0.0027
3	SEXE			M		0	.	.	.
3	TR_AGE					0.06023	0.03331	1.81	0.0705
3	REGION				10	0.6238	0.1654	3.77	0.0002
3	REGION				11	0.06178	0.3924	0.16	0.8749
3	REGION				12	0.1678	0.7390	0.23	0.8204
3	REGION				13	0.7081	0.1864	3.80	0.0001
3	REGION				14	-0.4318	0.2166	-1.99	0.0462
3	REGION				15	-0.2972	0.7473	-0.40	0.6908
3	REGION				16	0.7188	0.1618	4.44	<.0001
3	REGION				4	1.2373	0.1432	8.64	<.0001
3	REGION				5	0.4656	0.2495	1.87	0.0620
3	REGION				6	0.6213	0.4118	1.51	0.1313
3	REGION				7	0.5436	0.1550	3.51	0.0005
3	REGION				8	0.9697	0.2760	3.51	0.0004
3	REGION				9	0	.	.	.
1	Scale Parameter					38520	0		
2	Scale Parameter					0.2159	0.03306		
3	Scale Parameter					0.03092	0.004401		

Parameter Estimates for Mixing Probabilities

Component	Effect	Linked Scale				Probability
		Estimate	Standard Error	z Value	Pr > z	
1	Intercept	-3.8840	0.7396	-5.25	<.0001	0.0063
2	Intercept	0.8065	0.2036	3.96	<.0001	0.6870

II.12. Modèle retenu des remboursements des soins hospitaliers relatifs aux ALD

Model Information	
Data Set	FMMODEL.HOSPIT_NON_ALD
Response Variable	MONT_REM_M
Weight Variable	EFFECT_SINISTR
Type of Model	Homogeneous Regression Mixture
Distribution	Gamma
Min Components	1
Max Components	6
Link Function	Log
Estimation Method	Maximum Likelihood

Class Level Information		
Class	Levels	Values
TYPE_ASS	3	A I P
TYPE_BEN	3	A C E
SEXE	2	F M
REGION	13	10 11 12 13 14 15 16 4 5 6 7 8 9

Number of Observations Read	379
Number of Observations Used	379

Component Description for Mixture Models	
Model ID	Gamma
1	1
2	2
3	3
4	4
5	5
6	6

Component Evaluation for Mixture Models											
Model ID	Number of				-2 Log L	AIC	AICC	BIC	Pearson	Max Gradient	
	Components		Parameters								
	Total	Eff.	Total	Eff.							
1	1	1	22	22	8280.59	8324.59	8327.43	8411.22	531.29	0.000090	
2	2	2	45	45						0.086	
3	3	3	68	68						22.238	
4	4	4	91	91	7910.62	8092.62	8150.96	8450.93	914.32	5.520	
5	5	5	114	114						2.785	
6	6	6	137	137						27.862	

The model with 4 components (ID=4) was selected as 'best' based on the AIC statistic. Models that failed to converge are indicated with blank values for the likelihood-based criteria.

Convergence criterion (GCONV=1E-8) satisfied.

Fit Statistics	
-2 Log Likelihood	7910.6
AIC (smaller is better)	8092.6
AICC (smaller is better)	8151.0
BIC (smaller is better)	8450.9
Pearson Statistic	914.3
Effective Parameters	91
Effective Components	4

Parameter Estimates for 'Gamma' Model								
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
1	Intercept					8.5927	17378	0.00
1	TYPE_ASS	A				0.8301	0	.
1	TYPE_ASS	I				-3.8411	0	.
1	TYPE_ASS	P				0	.	.
1	TYPE_BEN		A			1.0880	0	.
1	TYPE_BEN		C			0.1323	0	.
1	TYPE_BEN		E			0	.	.
1	SEXE			F		0.5959	0	.
1	SEXE			M		0	.	.
1	TR_AGE					0.01710	2896.31	0.00
1	REGION				10	0.7161	0	.
1	REGION				11	-0.1156	0	.
1	REGION				12	-1.1509	0	.
1	REGION				13	0.2159	0	.

Parameter Estimates for 'Gamma' Model

Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
1	REGION				14	-1.0341	0	.
1	REGION				15	-1.6742	0	.
1	REGION				16	0.1473	0	.
1	REGION				4	0.7799	0	.
1	REGION				5	-0.2705	0	.
1	REGION				6	1.3843	0	.
1	REGION				7	-0.1843	0	.
1	REGION				8	0.1462	0	.
1	REGION				9	0	.	.
1	TYPE_ASS*TYPE_BEN	A	A			-0.5480	0	.
1	TYPE_ASS*TYPE_BEN	A	C			0.01269	0	.
1	TYPE_ASS*TYPE_BEN	A	E			0	.	.
1	TYPE_ASS*TYPE_BEN	I	C			0	.	.
1	TYPE_ASS*TYPE_BEN	P	A			0	.	.
1	TYPE_ASS*TYPE_BEN	P	C			0	.	.
1	TYPE_ASS*TYPE_BEN	P	E			0	.	.
2	Intercept					8.4603	0.3276	25.82
2	TYPE_ASS	A				0.1880	0.3291	0.57
2	TYPE_ASS	I				-0.01796	0.2647	-0.07
2	TYPE_ASS	P				0	.	.
2	TYPE_BEN		A			1.0248	0.3594	2.85
2	TYPE_BEN		C			0.8163	0.3683	2.22
2	TYPE_BEN		E			0	.	.
2	SEXE			F		-0.06988	0.04223	-1.65
2	SEXE			M		0	.	.
2	TR_AGE					-0.00276	0	.
2	REGION				10	0.7409	0.08420	8.80
2	REGION				11	0.8019	0.06357	12.61
2	REGION				12	0.7371	0.1367	5.39
2	REGION				13	1.1319	0.08339	13.57
2	REGION				14	-0.06831	0.1402	-0.49
2	REGION				15	0.06557	0.1739	0.38
2	REGION				16	0.6311	0.04831	13.06
2	REGION				4	1.5094	0.04978	30.32
2	REGION				5	0.9950	0.07065	14.08
2	REGION				6	0.9839	0.08526	11.54
2	REGION				7	0.6927	0.08107	8.54
2	REGION				8	1.2113	0.06854	17.67
2	REGION				9	0	.	.
2	TYPE_ASS*TYPE_BEN	A	A			-0.5274	0.3452	-1.53
2	TYPE_ASS*TYPE_BEN	A	C			0.4027	0.3451	1.17
2	TYPE_ASS*TYPE_BEN	A	E			0	.	.

Parameter Estimates for 'Gamma' Model

Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
2	TYPE_ASS*TYPE_BEN	I	C			0	.	.
2	TYPE_ASS*TYPE_BEN	P	A			0	.	.
2	TYPE_ASS*TYPE_BEN	P	C			0	.	.
2	TYPE_ASS*TYPE_BEN	P	E			0	.	.
3	Intercept					8.8727	0.2646	33.53
3	TYPE_ASS	A				0.1330	0.2674	0.50
3	TYPE_ASS	I				0.008515	0.2745	0.03
3	TYPE_ASS	P				0	.	.
3	TYPE_BEN		A			0.5024	0.2791	1.80
3	TYPE_BEN		C			0.3018	0.2918	1.03
3	TYPE_BEN		E			0	.	.
3	SEXE			F		-0.2496	0.04692	-5.32
3	SEXE			M		0	.	.
3	TR_AGE					0.03153	0	.
3	REGION				10	0.7193	0.05332	13.49
3	REGION				11	1.1449	0.08075	14.18
3	REGION				12	1.2293	0.2057	5.98
3	REGION				13	0.9056	0.1133	7.99
3	REGION				14	0.04419	0.08335	0.53
3	REGION				15	0.3707	0.1296	2.86
3	REGION				16	0.7922	0.04656	17.01
3	REGION				4	1.7920	0.04612	38.85
3	REGION				5	0.8625	0.08236	10.47
3	REGION				6	1.3933	0.06655	20.94
3	REGION				7	0.7902	0.04909	16.10
3	REGION				8	1.2746	0.04919	25.91
3	REGION				9	0	.	.
3	TYPE_ASS*TYPE_BEN	A	A			-0.2822	0.2715	-1.04
3	TYPE_ASS*TYPE_BEN	A	C			-0.07741	0.2844	-0.27
3	TYPE_ASS*TYPE_BEN	A	E			0	.	.
3	TYPE_ASS*TYPE_BEN	I	C			0	.	.
3	TYPE_ASS*TYPE_BEN	P	A			0	.	.
3	TYPE_ASS*TYPE_BEN	P	C			0	.	.
3	TYPE_ASS*TYPE_BEN	P	E			0	.	.
4	Intercept					9.0741	2.4112	3.76
4	TYPE_ASS	A				0.8524	2.4097	0.35
4	TYPE_ASS	I				-0.2496	2.2654	-0.11
4	TYPE_ASS	P				0	.	.
4	TYPE_BEN		A			0.5467	2.4836	0.22
4	TYPE_BEN		C			0.1310	2.4784	0.05
4	TYPE_BEN		E			0	.	.
4	SEXE			F		-0.3167	0.1554	-2.04

Parameter Estimates for 'Gamma' Model

Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
4	SEXE			M		0	.	.
4	TR_AGE					0.06831	0.05433	1.26
4	REGION				10	0.3522	0.1591	2.21
4	REGION				11	0.3626	0.3779	0.96
4	REGION				12	-0.3998	0.5940	-0.67
4	REGION				13	0.3133	0.2799	1.12
4	REGION				14	-0.6586	0.2440	-2.70
4	REGION				15	-0.5781	0.7988	-0.72
4	REGION				16	0.7441	0.1707	4.36
4	REGION				4	1.0635	0.2381	4.47
4	REGION				5	-0.06548	0.3220	-0.20
4	REGION				6	0.4701	0.4857	0.97
4	REGION				7	0.3768	0.3512	1.07
4	REGION				8	1.1098	0.4515	2.46
4	REGION				9	0	.	.
4	TYPE_ASS*TYPE_BEN	A	A			-1.0223	2.4220	-0.42
4	TYPE_ASS*TYPE_BEN	A	C			0.1483	2.4309	0.06
4	TYPE_ASS*TYPE_BEN	A	E			0	.	.
4	TYPE_ASS*TYPE_BEN	I	C			0	.	.
4	TYPE_ASS*TYPE_BEN	P	A			0	.	.
4	TYPE_ASS*TYPE_BEN	P	C			0	.	.
4	TYPE_ASS*TYPE_BEN	P	E			0	.	.
1	Scale Parameter					38520	0	
2	Scale Parameter					0.5356	0.2559	
3	Scale Parameter					0.4566	0.08384	
4	Scale Parameter					0.03807	0	

Parameter Estimates for 'Gamma' Model

Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
1	Intercept					0.9996
1	TYPE_ASS	A				.
1	TYPE_ASS	I				.
1	TYPE_ASS	P				.
1	TYPE_BEN		A			.
1	TYPE_BEN		C			.
1	TYPE_BEN		E			.
1	SEXE			F		.
1	SEXE			M		.
1	TR_AGE					1.0000
1	REGION				10	.
1	REGION				11	.
1	REGION				12	.

Parameter Estimates for 'Gamma' Model						
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
1	REGION				13	.
1	REGION				14	.
1	REGION				15	.
1	REGION				16	.
1	REGION				4	.
1	REGION				5	.
1	REGION				6	.
1	REGION				7	.
1	REGION				8	.
1	REGION				9	.
1	TYPE_ASS*TYPE_BEN	A	A			.
1	TYPE_ASS*TYPE_BEN	A	C			.
1	TYPE_ASS*TYPE_BEN	A	E			.
1	TYPE_ASS*TYPE_BEN	I	C			.
1	TYPE_ASS*TYPE_BEN	P	A			.
1	TYPE_ASS*TYPE_BEN	P	C			.
1	TYPE_ASS*TYPE_BEN	P	E			.
2	Intercept					<.0001
2	TYPE_ASS	A				0.5679
2	TYPE_ASS	I				0.9459
2	TYPE_ASS	P				.
2	TYPE_BEN		A			0.0043
2	TYPE_BEN		C			0.0267
2	TYPE_BEN		E			.
2	SEXE			F		0.0980
2	SEXE			M		.
2	TR_AGE					.
2	REGION				10	<.0001
2	REGION				11	<.0001
2	REGION				12	<.0001
2	REGION				13	<.0001
2	REGION				14	0.6260
2	REGION				15	0.7062
2	REGION				16	<.0001
2	REGION				4	<.0001
2	REGION				5	<.0001
2	REGION				6	<.0001
2	REGION				7	<.0001
2	REGION				8	<.0001
2	REGION				9	.
2	TYPE_ASS*TYPE_BEN	A	A			0.1266
2	TYPE_ASS*TYPE_BEN	A	C			0.2433
2	TYPE_ASS*TYPE_BEN	A	E			.

Parameter Estimates for 'Gamma' Model						
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
2	TYPE_ASS*TYPE_BEN	I	C			.
2	TYPE_ASS*TYPE_BEN	P	A			.
2	TYPE_ASS*TYPE_BEN	P	C			.
2	TYPE_ASS*TYPE_BEN	P	E			.
3	Intercept					<.0001
3	TYPE_ASS	A				0.6189
3	TYPE_ASS	I				0.9752
3	TYPE_ASS	P				.
3	TYPE_BEN		A			0.0718
3	TYPE_BEN		C			0.3010
3	TYPE_BEN		E			.
3	SEXE			F		<.0001
3	SEXE			M		.
3	TR_AGE					.
3	REGION				10	<.0001
3	REGION				11	<.0001
3	REGION				12	<.0001
3	REGION				13	<.0001
3	REGION				14	0.5960
3	REGION				15	0.0042
3	REGION				16	<.0001
3	REGION				4	<.0001
3	REGION				5	<.0001
3	REGION				6	<.0001
3	REGION				7	<.0001
3	REGION				8	<.0001
3	REGION				9	.
3	TYPE_ASS*TYPE_BEN	A	A			0.2985
3	TYPE_ASS*TYPE_BEN	A	C			0.7855
3	TYPE_ASS*TYPE_BEN	A	E			.
3	TYPE_ASS*TYPE_BEN	I	C			.
3	TYPE_ASS*TYPE_BEN	P	A			.
3	TYPE_ASS*TYPE_BEN	P	C			.
3	TYPE_ASS*TYPE_BEN	P	E			.
4	Intercept					0.0002
4	TYPE_ASS	A				0.7235
4	TYPE_ASS	I				0.9123
4	TYPE_ASS	P				.
4	TYPE_BEN		A			0.8258
4	TYPE_BEN		C			0.9578
4	TYPE_BEN		E			.
4	SEXE			F		0.0415
4	SEXE			M		.

Parameter Estimates for 'Gamma' Model						
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
4	TR_AGE					0.2086
4	REGION				10	0.0269
4	REGION				11	0.3373
4	REGION				12	0.5009
4	REGION				13	0.2630
4	REGION				14	0.0069
4	REGION				15	0.4693
4	REGION				16	<.0001
4	REGION				4	<.0001
4	REGION				5	0.8388
4	REGION				6	0.3332
4	REGION				7	0.2833
4	REGION				8	0.0140
4	REGION				9	.
4	TYPE_ASS*TYPE_BEN	A	A			0.6730
4	TYPE_ASS*TYPE_BEN	A	C			0.9514
4	TYPE_ASS*TYPE_BEN	A	E			.
4	TYPE_ASS*TYPE_BEN	I	C			.
4	TYPE_ASS*TYPE_BEN	P	A			.
4	TYPE_ASS*TYPE_BEN	P	C			.
4	TYPE_ASS*TYPE_BEN	P	E			.
1	Scale Parameter					
2	Scale Parameter					
3	Scale Parameter					
4	Scale Parameter					

Parameter Estimates for Mixing Probabilities						
Component	Effect	Linked Scale				Probability
		Estimate	Standard Error	z Value	Pr > z	
1	Intercept	-2.8360	0.4573	-6.20	<.0001	0.0133
2	Intercept	0.4510	0	.	.	0.3568
3	Intercept	0.5719	0.2027	2.82	0.0048	0.4026

II.13. Modèle simple des remboursements des soins hospitaliers relatifs aux ALD

Model Information	
Data Set	FMMODEL.HOSPIT_ALD
Response Variable	MONT_REM_M
Weight Variable	EFFECT_SINISTR
Type of Model	Homogeneous Regression Mixture
Distribution	Gamma
Min Components	1
Max Components	6
Link Function	Log
Estimation Method	Maximum Likelihood

Class Level Information		
Class	Levels	Values
TYPE_ASS	2	A P
TYPE_BEN	2	A C
SEXE	2	F M
REGION	12	10 11 12 13 14 16 4 5 6 7 8 9

Number of Observations Read	93
Number of Observations Used	93

Component Description for Mixture Models	
Model ID	Gamma
1	1
2	2
3	3
4	4
5	5
6	6

Component Evaluation for Mixture Models										
Model ID	Number of				-2 Log L	AIC	AICC	BIC	Pearson	Max Gradient
	Components		Parameters							
	Total	Eff.	Total	Eff.						
1	1	1	17	17	1.1848E9	1.1848E9	1.1848E9	1.1848E9	1.1716E9	234434
2	2	1	35	35						4955
3	3	3	53	53						1.608
4	4	4	71	71						1.709
5	5	5	89	89						10.500
6	6	6	107	107						4.659

The single-component model (ID=1) was selected as 'best' based on the AIC statistic. Models that failed to converge are indicated with blank values for the likelihood-based criteria.

Convergence criterion (GCONV=1E-8) satisfied.

Fit Statistics	
-2 Log Likelihood	1.1848E9
AIC (smaller is better)	1.1848E9
AICC (smaller is better)	1.1848E9
BIC (smaller is better)	1.1848E9
Pearson Statistic	1.1716E9
Unscaled Pearson Chi-Square	528.3
Effective Parameters	17
Effective Components	1

Parameter Estimates for 'Gamma' Model								
Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value	Pr > z
Intercept					13.4603	0.000089	151996	<.0001
TYPE_ASS	A				-0.1566	0.000031	-5045.0	<.0001
TYPE_ASS	P				0	.	.	.
TYPE_BEN		A			0.02955	0.000021	1421.83	<.0001
TYPE_BEN		C			0	.	.	.
REGION				10	0.01511	0.000024	635.99	<.0001
REGION				11	-0.08961	0.000065	-1388.7	<.0001
REGION				12	-0.5530	0.000123	-4487.8	<.0001
REGION				13	-0.3836	0.000052	-7309.2	<.0001
REGION				14	-0.7650	0.000053	-14529	<.0001
REGION				16	-0.05707	0.000037	-1558.9	<.0001
REGION				4	0.08326	0.000029	2832.01	<.0001
REGION				5	-0.07716	0.000038	-2012.5	<.0001

Parameter Estimates for 'Gamma' Model								
Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value	Pr > z
REGION				6	0.09720	0.000089	1091.69	<.0001
REGION				7	-0.4119	0.000045	-9060.6	<.0001
REGION				8	-0.08494	0.000040	-2117.3	<.0001
REGION				9	0	.	.	.
SEXE			F		-0.3468	0.000017	-20261	<.0001
SEXE			M		0	.	.	.
TR_AGE					0.006461	5.848E-6	1104.84	<.0001
Scale Parameter					2217507	0		

II.14. Modèle retenu des soins hospitaliers relatifs aux ALD

Model Information	
Data Set	FMMODEL.HOSPIT_ALD
Response Variable	MONT_REM_M
Weight Variable	EFFECT_SINISTR
Type of Model	Homogeneous Regression Mixture
Distribution	Gamma
Min Components	1
Max Components	6
Link Function	Log
Estimation Method	Maximum Likelihood

Class Level Information		
Class	Levels	Values
TYPE_ASS	2	A P
TYPE_BEN	2	A C
SEXE	2	F M
REGION	12	10 11 12 13 14 16 4 5 6 7 8 9

Number of Observations Read	93
Number of Observations Used	93

Component Description for Mixture Models	
Model ID	Gamma
1	1
2	2
3	3
4	4
5	5
6	6

Component Evaluation for Mixture Models										
Model ID	Number of				-2 Log L	AIC	AICC	BIC	Pearson	Max Gradient
	Components		Parameters							
	Total	Eff.	Total	Eff.						
1	1	1	19	19	1.1774E9	1.1774E9	1.1774E9	1.1774E9	1.1599E9	431.122
2	2	2	39	39	9.1851E8	9.1851E8	9.1851E8	9.1851E8	9757.66	184.735
3	3	2	59	59	1.1895E9	1.1895E9	1.1895E9	1.1895E9	54572.01	265.480
4	4	1	79	79	2490.16	2648.16	3620.47	2848.24	93.10	2.901E-6
5	5	2	99	99						0
6	6	6	119	119						47.493

The model with 4 components (ID=4) was selected as 'best' based on the AIC statistic. Models that failed to converge are indicated with blank values for the likelihood-based criteria.

Convergence criterion (GCONV=1E-8) satisfied.

Fit Statistics	
-2 Log Likelihood	2490.2
AIC (smaller is better)	2648.2
AICC (smaller is better)	3620.5
BIC (smaller is better)	2848.2
Pearson Statistic	93.1024
Effective Parameters	79
Effective Components	1

Parameter Estimates for 'Gamma' Model								
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
1	Intercept					13.2948	0.3912	33.99
1	TYPE_ASS	A				-0.2149	0.1688	-1.27
1	TYPE_ASS	P				0	.	.
1	TYPE_BEN		A			0.01318	0.08067	0.16
1	TYPE_BEN		C			0	.	.
1	REGION				10	0.01320	0.08440	0.16
1	REGION				11	-0.09974	0.2283	-0.44
1	REGION				12	-0.5669	0.4354	-1.30
1	REGION				13	-0.3921	0.1857	-2.11
1	REGION				14	-0.7745	0.1864	-4.15
1	REGION				16	-0.05696	0.1293	-0.44
1	REGION				4	0.08704	0.1049	0.83
1	REGION				5	-0.07778	0.1353	-0.57
1	REGION				6	0.08327	0.3149	0.26

Parameter Estimates for 'Gamma' Model

Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
1	REGION				7	-0.4120	0.1605	-2.57
1	REGION				8	-0.08702	0.1416	-0.61
1	REGION				9	0	.	.
1	SEXE			F		-0.08933	0.3633	-0.25
1	SEXE			M		0	.	.
1	TR_AGE					0.01870	0.02650	0.71
1	TR_AGE*SEXE			F		-0.01821	0.02520	-0.72
1	TR_AGE*SEXE			M		0	.	.
1	TYPE_ASS*TYPE_BEN	A	A			0.08172	0.1672	0.49
1	TYPE_ASS*TYPE_BEN	A	C			0	.	.
1	TYPE_ASS*TYPE_BEN	P	A			0	.	.
1	TYPE_ASS*TYPE_BEN	P	C			0	.	.
2	Intercept					3.2684	0	.
2	TYPE_ASS	A				0.007253	0	.
2	TYPE_ASS	P				0	.	.
2	TYPE_BEN		A			0.07233	0	.
2	TYPE_BEN		C			0	.	.
2	REGION				10	0.01422	0	.
2	REGION				11	0.006477	0	.
2	REGION				12	0.000397	0	.
2	REGION				13	0.005022	0	.
2	REGION				14	-0.00127	0	.
2	REGION				16	-0.01289	0	.
2	REGION				4	-0.01973	0	.
2	REGION				5	-0.01097	0	.
2	REGION				6	-0.00262	0	.
2	REGION				7	-0.00940	0	.
2	REGION				8	-0.01112	0	.
2	REGION				9	0	.	.
2	SEXE			F		-0.1884	0	.
2	SEXE			M		0	.	.
2	TR_AGE					0.01444	0	.
2	TR_AGE*SEXE			F		-9.5616	0	.
2	TR_AGE*SEXE			M		0	.	.
2	TYPE_ASS*TYPE_BEN	A	A			0.02907	0	.
2	TYPE_ASS*TYPE_BEN	A	C			0	.	.
2	TYPE_ASS*TYPE_BEN	P	A			0	.	.
2	TYPE_ASS*TYPE_BEN	P	C			0	.	.
3	Intercept					3.2684	0	.
3	TYPE_ASS	A				0.007253	0	.
3	TYPE_ASS	P				0	.	.
3	TYPE_BEN		A			0.07233	0	.

Parameter Estimates for 'Gamma' Model

Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
3	TYPE_BEN		C			0	.	.
3	REGION				10	0.01422	0	.
3	REGION				11	0.006477	0	.
3	REGION				12	0.000397	0	.
3	REGION				13	0.005022	0	.
3	REGION				14	-0.00127	0	.
3	REGION				16	-0.01289	0	.
3	REGION				4	-0.01973	0	.
3	REGION				5	-0.01097	0	.
3	REGION				6	-0.00262	0	.
3	REGION				7	-0.00940	0	.
3	REGION				8	-0.01112	0	.
3	REGION				9	0	.	.
3	SEXE			F		-0.1884	0	.
3	SEXE			M		0	.	.
3	TR_AGE					0.01444	0	.
3	TR_AGE*SEXE			F		-9.5616	0	.
3	TR_AGE*SEXE			M		0	.	.
3	TYPE_ASS*TYPE_BEN	A	A			0.02907	0	.
3	TYPE_ASS*TYPE_BEN	A	C			0	.	.
3	TYPE_ASS*TYPE_BEN	P	A			0	.	.
3	TYPE_ASS*TYPE_BEN	P	C			0	.	.
4	Intercept					3.2684	0	.
4	TYPE_ASS	A				0.007253	0	.
4	TYPE_ASS	P				0	.	.
4	TYPE_BEN		A			0.07233	0	.
4	TYPE_BEN		C			0	.	.
4	REGION				10	0.01422	0	.
4	REGION				11	0.006477	0	.
4	REGION				12	0.000397	0	.
4	REGION				13	0.005022	0	.
4	REGION				14	-0.00127	0	.
4	REGION				16	-0.01289	0	.
4	REGION				4	-0.01973	0	.
4	REGION				5	-0.01097	0	.
4	REGION				6	-0.00262	0	.
4	REGION				7	-0.00940	0	.
4	REGION				8	-0.01112	0	.
4	REGION				9	0	.	.
4	SEXE			F		-0.1884	0	.
4	SEXE			M		0	.	.
4	TR_AGE					0.01444	0	.

Parameter Estimates for 'Gamma' Model

Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Estimate	Standard Error	z Value
4	TR_AGE*SEXE			F		-9.5616	0	.
4	TR_AGE*SEXE			M		0	.	.
4	TYPE_ASS*TYPE_BEN	A	A			0.02907	0	.
4	TYPE_ASS*TYPE_BEN	A	C			0	.	.
4	TYPE_ASS*TYPE_BEN	P	A			0	.	.
4	TYPE_ASS*TYPE_BEN	P	C			0	.	.
1	Scale Parameter					0.1780	0.02569	
2	Scale Parameter					2217793	0	
3	Scale Parameter					2217793	0	
4	Scale Parameter					2217793	0	

Parameter Estimates for 'Gamma' Model

Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
1	Intercept					<.0001
1	TYPE_ASS	A				0.2029
1	TYPE_ASS	P				.
1	TYPE_BEN		A			0.8702
1	TYPE_BEN		C			.
1	REGION				10	0.8757
1	REGION				11	0.6622
1	REGION				12	0.1929
1	REGION				13	0.0348
1	REGION				14	<.0001
1	REGION				16	0.6596
1	REGION				4	0.4067
1	REGION				5	0.5655
1	REGION				6	0.7915
1	REGION				7	0.0103
1	REGION				8	0.5390
1	REGION				9	.
1	SEXE			F		0.8058
1	SEXE			M		.
1	TR_AGE					0.4805
1	TR_AGE*SEXE			F		0.4698
1	TR_AGE*SEXE			M		.
1	TYPE_ASS*TYPE_BEN	A	A			0.6250
1	TYPE_ASS*TYPE_BEN	A	C			.
1	TYPE_ASS*TYPE_BEN	P	A			.
1	TYPE_ASS*TYPE_BEN	P	C			.
2	Intercept					.
2	TYPE_ASS	A				.
2	TYPE_ASS	P				.

Parameter Estimates for 'Gamma' Model						
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
2	TYPE_BEN		A			.
2	TYPE_BEN		C			.
2	REGION				10	.
2	REGION				11	.
2	REGION				12	.
2	REGION				13	.
2	REGION				14	.
2	REGION				16	.
2	REGION				4	.
2	REGION				5	.
2	REGION				6	.
2	REGION				7	.
2	REGION				8	.
2	REGION				9	.
2	SEXE			F		.
2	SEXE			M		.
2	TR_AGE					.
2	TR_AGE*SEXE			F		.
2	TR_AGE*SEXE			M		.
2	TYPE_ASS*TYPE_BEN	A	A			.
2	TYPE_ASS*TYPE_BEN	A	C			.
2	TYPE_ASS*TYPE_BEN	P	A			.
2	TYPE_ASS*TYPE_BEN	P	C			.
3	Intercept					.
3	TYPE_ASS	A				.
3	TYPE_ASS	P				.
3	TYPE_BEN		A			.
3	TYPE_BEN		C			.
3	REGION				10	.
3	REGION				11	.
3	REGION				12	.
3	REGION				13	.
3	REGION				14	.
3	REGION				16	.
3	REGION				4	.
3	REGION				5	.
3	REGION				6	.
3	REGION				7	.
3	REGION				8	.
3	REGION				9	.
3	SEXE			F		.
3	SEXE			M		.
3	TR_AGE					.

Parameter Estimates for 'Gamma' Model						
Component	Effect	TYPE_ASS	TYPE_BEN	SEXE	REGION	Pr > z
3	TR_AGE*SEXE			F		.
3	TR_AGE*SEXE			M		.
3	TYPE_ASS*TYPE_BEN	A	A			.
3	TYPE_ASS*TYPE_BEN	A	C			.
3	TYPE_ASS*TYPE_BEN	P	A			.
3	TYPE_ASS*TYPE_BEN	P	C			.
4	Intercept					.
4	TYPE_ASS	A				.
4	TYPE_ASS	P				.
4	TYPE_BEN		A			.
4	TYPE_BEN		C			.
4	REGION				10	.
4	REGION				11	.
4	REGION				12	.
4	REGION				13	.
4	REGION				14	.
4	REGION				16	.
4	REGION				4	.
4	REGION				5	.
4	REGION				6	.
4	REGION				7	.
4	REGION				8	.
4	REGION				9	.
4	SEXE			F		.
4	SEXE			M		.
4	TR_AGE					.
4	TR_AGE*SEXE			F		.
4	TR_AGE*SEXE			M		.
4	TYPE_ASS*TYPE_BEN	A	A			.
4	TYPE_ASS*TYPE_BEN	A	C			.
4	TYPE_ASS*TYPE_BEN	P	A			.
4	TYPE_ASS*TYPE_BEN	P	C			.
1	Scale Parameter					
2	Scale Parameter					
3	Scale Parameter					
4	Scale Parameter					

Parameter Estimates for Mixing Probabilities						
Component	Effect	Linked Scale				Probability
		Estimate	Standard Error	z Value	Pr > z	
1	Intercept	111.61	0	.	.	1.0000
2	Intercept	-0.6952	0	.	.	17E-50
3	Intercept	-0.2897	0	.	.	25E-50

III.Code python de la fonction de projection démographique

```

#paramètres
Tfec = ??
PE = pd.DataFrame([[??, ??]], columns = ['M', 'F'])
Nme = pd.DataFrame([[??, ??]], columns = ['A', 'R'])
#différence d'age moyen entre homme et femme
d = ??

#taux d'adhésion des conjoints
tac = pd.DataFrame([[??, ??], [??, ??]], columns = ['F', 'M'], index = ['A', 'R'])

#Taux de dsorie pour autre que le décès (non disponible)
Ts = ??
#taux d'invalidité
Tinv = ??

#On définit les adhésions au cours de l'année 2012
tem = ??

#proportion des 114 parmi les entrées
p114 = ??

#Fonction de projection
def project(pop_a_projetter):

    adhes_n = pop_a_projetter.EFFECTIF.sum()*tem

    pop_a_projetter = pd.merge(pop_a_projetter, poids_rep, how = 'outer')
    poids114 = pd.read_csv(r'C:\PFE\WORK\BD\CNSS\POPULATION\poids114', )
    poids114.drop('Unnamed: 0', axis = 1, inplace = True)
    pop_a_projetter = pd.merge(pop_a_projetter, poids114, how = 'outer')
    pop_a_projetter.TR_114.fillna(0, inplace = True)
    pop_a_projetter.EFFECTIF.fillna(0, inplace = True)
    pop_a_projetter['TR'] = p114 * pop_a_projetter.TR_114 + (1 - p114) *
pop_a_projetter.TR_non_114

    # Actifs
    pop_actifs = pop_a_projetter[np.logical_and(pop_a_projetter.TYPE_ASS == 'A',
pop_a_projetter.TYPE_BEN == 'A')].copy()
    pop_actifs = pop_actifs[pop_actifs.TR_AGE < 60]
    pop_actifs = add_origin_category(pop_actifs, 'TR_AGE', 'EFFECTIF')
    pop_actifs = pop_actifs.groupby(['REGION', 'SEXE', 'TR_AGE', 'TYPE_ASS',
'TYPE_BEN']).sum().reset_index()

    pop_actifs['q_DECES'] = 10

    for i in table_mortalite.age:
        pop_actifs['q_DECES'][pop_actifs.TR_AGE == i] = table_mortalite.q.loc[i]

    pop_actifs['D_n'] = pop_actifs.EFFECTIF * pop_actifs.q_DECES

#On répartit les nouvelles entrées
pop_actifs['N_n'] = adhes_n * pop_actifs.TR

pop_actifs['Ninv_n'] = Tinv * pop_actifs.EFFECTIF

pop_actifs['S_n'] = Ts * pop_actifs.EFFECTIF
pop_actifs['TR_AGE_n_plus_un'] = pop_actifs.TR_AGE + 1

```

```

pop_actifs['E_n_plus_un'] = pop_actifs.EFFECTIF + pop_actifs.N_n - pop_actifs.Ninv_n -
pop_actifs.D_n - pop_actifs.S_n

#Nouveaux retraités
pop_actifs['TYPE_ASS'][pop_actifs.TR_AGE_n_plus_un >= 60] = 'R'

#colonnes d'intérêt
foo = list(pop_actifs.columns)
colones = foo[0:2] + foo[3:5] + ['TR_AGE_n_plus_un', 'E_n_plus_un']

pop_actifs_n_plus_un = pop_actifs[colones].copy(deep = True)

# Retraités

pop_retraite = pop_a_projetter[np.logical_and(pop_a_projetter.TYPE_ASS == 'R',
pop_a_projetter.TYPE_BEN == 'A')].copy()
pop_retraite = add_origin_category(pop_retraite, 'TR_AGE', 'EFFECTIF')
pop_retraite = pop_retraite.groupby(['REGION', 'SEXE', 'TR_AGE', 'TYPE_ASS',
'TYPE_BEN']).sum().reset_index()

pop_retraite['q_DECES'] = 10

for i in table_mortalite.age:
    pop_retraite['q_DECES'][pop_retraite.TR_AGE == i] = table_mortalite.q.loc[i]

#décès
pop_retraite['D_n'] = pop_retraite.EFFECTIF * pop_retraite.q_DECES

#adhésions
pop_retraite['N_n'] = adhes_n * pop_retraite.TR

#entrées pour les retraites anticipées
pop_retraite['N_n'][pop_retraite.TR_AGE < 60] = adhes_n * pop_retraite.TR_non_114
pop_retraite['TR_AGE_n_plus_un'] = pop_retraite.TR_AGE + 1
pop_retraite['E_n_plus_un'] = pop_retraite.EFFECTIF + pop_retraite.N_n - pop_retraite.D_n
pop_retraite_n_plus_un = pop_retraite[colones].copy(deep = True)

#La table des actifs contient quelque retraité, on les dissocie
foo = pd.concat([pop_retraite_n_plus_un, pop_actifs_n_plus_un])

pop_retraite_n_plus_un = foo[np.logical_or(foo.TR_AGE_n_plus_un >= 60, foo.TYPE_ASS ==
'R')].copy(deep = True)
pop_retraite_n_plus_un['TYPE_ASS'] = 'R'

pop_actifs_n_plus_un = foo[np.logical_and(foo.TR_AGE_n_plus_un < 60, foo.TYPE_ASS !=
'R')].copy(deep = True)
pop_actifs_n_plus_un['TYPE_ASS'] = 'A'

# Conjoints

pop_conjoints_d_actifs = pop_a_projetter[np.logical_and(pop_a_projetter.TYPE_ASS == 'A',
pop_a_projetter.TYPE_BEN == 'C')].copy()
pop_conjoints_d_actifs = add_origin_category(pop_conjoints_d_actifs, 'TR_AGE', 'EFFECTIF')
pop_conjoints_d_actifs = pop_conjoints_d_actifs.groupby(['REGION', 'SEXE', 'TR_AGE',
'TYPE_ASS', 'TYPE_BEN']).sum().reset_index()

# Conjoints d'actifs

pop_conjoints_d_actifs['q_DECES'] = 10

for i in table_mortalite.age:
    pop_conjoints_d_actifs['q_DECES'][pop_conjoints_d_actifs.TR_AGE == i] =
table_mortalite.q.loc[i]

#décès
pop_conjoints_d_actifs['D_n'] = pop_conjoints_d_actifs.EFFECTIF *
pop_conjoints_d_actifs.q_DECES

```

```

pop_actifs['Sexe_conjoints'] = 'Z'

pop_actifs['Sexe_conjoints'][pop_actifs.SEXE == 'M'] = 'F'
pop_actifs['Sexe_conjoints'][pop_actifs.SEXE == 'F'] = 'M'

#on ajout les tacs à la table des actifs de n

pop_actifs['Tac'] = 10
for s in pop_actifs.Sexe_conjoints.value_counts().index:
    pop_actifs['Tac'][pop_actifs.SEXE == s] = tac[s].loc['A']

#Puis on calcule les nouveaux conjoints de ces actifs

pop_actifs['Conjoints'] = pop_actifs.EFFECTIF * pop_actifs.Tac

pop_actifs['Age_Conjoint'] = 1000

pop_actifs['Age_Conjoint'][pop_actifs.SEXE == 'M'] = pop_actifs.TR_AGE - d
pop_actifs['Age_Conjoint'][pop_actifs.SEXE == 'F'] = pop_actifs.TR_AGE + d

#conjoint sortant pour sortie d'actif autre que décès ou retraite
pop_actifs['Sc_n'] = Ts * pop_actifs.EFFECTIF * pop_actifs.Tac

#on ajoute les poids des conjoints a la table pop_actifs
pop_actifs['P_CONJ'] = 0

for i in pop_actifs.index:
    pop_actifs['P_CONJ'].loc[i] = P_CONJ(pop_actifs.TYPE_ASS.loc[i],
pop_actifs.Sexe_conjoints.loc[i],
pop_actifs.Age_Conjoint.loc[i], pop_actifs.REGION.loc[i],
pop_a_projetter )

#on ajoute les nouvelles veufs

pop_actifs['NV_n'] = pop_actifs.D_n*pop_actifs.P_CONJ

#conjoint sortant pour sortie d'actif autre que décès ou retraite
pop_actifs['Sc_n'] = Ts * pop_actifs.EFFECTIF * pop_actifs.P_CONJ

#région, sexe du conjoint, Age
foo = pop_actifs[['REGION', 'Sexe_conjoints', 'Age_Conjoint', 'Sc_n', 'NV_n', 'Conjoints'
]]
N_conjoints_d_actif = pd.DataFrame(array(foo), columns = ['REGION', 'SEXE', 'TR_AGE',
'Sc_n', 'NV_n', 'N_n' ])

N_conjoints_d_actif = N_conjoints_d_actif.groupby(['REGION', 'SEXE',
'TR_AGE']).sum().reset_index()

#On fait une jointure entre la table de conjoints et les conjoints issus des mariages des
actifs (ou 114)

pop_conjoints_d_actifs = pd.merge(pop_conjoints_d_actifs, N_conjoints_d_actif, how =
'outer', on = ['REGION', 'SEXE', 'TR_AGE'] )
pop_conjoints_d_actifs.fillna(0, inplace = True)

pop_conjoints_d_actifs.TYPE_ASS = 'A'
pop_conjoints_d_actifs.TYPE_BEN = 'C'

#conjoints des nouveaux invalides
pop_conjoints_d_actifs['CNinv_n'] = pop_conjoints_d_actifs.N_n * Tinv

#les conjoints des nouveaux retraités seront identifiées comme ceux ayant un age de 58
pop_conjoints_d_actifs['TYPE_ASS'][pop_conjoints_d_actifs.TR_AGE == 58] = 'R'

#On calcule l'Effectif de n plus un
pop_conjoints_d_actifs['E_n_plus_un'] = pop_conjoints_d_actifs.EFFECTIF -
pop_conjoints_d_actifs.D_n + pop_conjoints_d_actifs.N_n - pop_conjoints_d_actifs.CNinv_n -
pop_conjoints_d_actifs.Sc_n - pop_conjoints_d_actifs.NV_n

pop_conjoints_d_actifs['TR_AGE_n_plus_un'] = pop_conjoints_d_actifs.TR_AGE + 1

```

```

pop_conjoints_d_actifs_n_plus_un = pop_conjoints_d_actifs[colones].copy(deep = True)

# Conjoints de retraité

pop_conjoints_de_retraite = pop_a_projetter[np.logical_and(pop_a_projetter.TYPE_ASS ==
'R', pop_a_projetter.TYPE_BEN == 'C')].copy()
pop_conjoints_de_retraite = add_origin_category(pop_conjoints_de_retraite, 'TR_AGE',
'EFFECTIF')
pop_conjoints_de_retraite = pop_conjoints_de_retraite.groupby(['REGION', 'SEXE', 'TR_AGE',
'TYPE_ASS', 'TYPE_BEN']).sum().reset_index()

pop_conjoints_de_retraite['q_DECES'] = 10

for i in table_mortalite.age:

    pop_conjoints_de_retraite['q_DECES'][pop_conjoints_de_retraite.TR_AGE == i] =
table_mortalite.q.loc[i]

#décès
pop_conjoints_de_retraite['D_n'] = pop_conjoints_de_retraite.EFFECTIF *
pop_conjoints_de_retraite.q_DECES

#on ajout les tacs à la table des actifs de n

pop_retraite['Tac'] = 10
for s in pop_retraite.SEXE.value_counts().index:
    pop_retraite['Tac'][pop_retraite.SEXE == s] = tac[s].loc['R']

#Puis on calcule les nouveaux conjoints de ces actifs

pop_retraite['Conjoints'] = pop_retraite.EFFECTIF * pop_retraite.Tac

pop_retraite['Sexe_conjoints'] = 'Z'
pop_retraite['Age_Conjoint'] = 1000

pop_retraite['Sexe_conjoints'][pop_retraite.SEXE == 'M'] = 'F'
pop_retraite['Sexe_conjoints'][pop_retraite.SEXE == 'F'] = 'M'

pop_retraite['Age_Conjoint'][pop_retraite.SEXE == 'M'] = pop_retraite.TR_AGE - d
pop_retraite['Age_Conjoint'][pop_retraite.SEXE == 'F'] = pop_retraite.TR_AGE + d

#on ajoute les poids des conjoints a la table pop_actifs
pop_retraite['P_CONJ'] = 0

for i in pop_retraite.index:
    pop_retraite['P_CONJ'].loc[i] = P_CONJ(pop_retraite.TYPE_ASS.loc[i],
pop_retraite.Sexe_conjoints.loc[i],
pop_retraite.Age_Conjoint.loc[i],
pop_retraite.REGION.loc[i], pop_a_projetter )

#on ajoute les nouvelles veufs
pop_retraite['NV_n'] = pop_retraite.D_n*pop_retraite.P_CONJ

#région, sexe du conjoint, Age
foo = pop_retraite[['REGION', 'Sexe_conjoints', 'Age_Conjoint', 'NV_n', 'Conjoints' ]]
N_conjoints_de_retraite = pd.DataFrame(array(foo), columns = ['REGION', 'SEXE', 'TR_AGE',
'NV_n', 'N_n' ])

pop_conjoints_de_retraite.EFFECTIF.sum()

N_conjoints_de_retraite = N_conjoints_de_retraite.groupby(['REGION', 'SEXE',
'TR_AGE']).sum().reset_index()

#On fait une jointure entre la table de conjoints et les conjoints issus des mariages des
actifs (ou 114)

pop_conjoints_de_retraite = pd.merge(pop_conjoints_de_retraite, N_conjoints_de_retraite,
how = 'outer', on = ['REGION', 'SEXE', 'TR_AGE' ])
pop_conjoints_de_retraite.EFFECTIF.fillna(0, inplace = True)
pop_conjoints_de_retraite.TYPE_ASS = 'R'

#conjoints des nouveaux invalides
pop_conjoints_de_retraite['CNInv_n'] = pop_conjoints_de_retraite.N_n * Tin

```

```

#On calcule l'Effectif de n_plus_un
pop_conjoints_de_retraite['E_n_plus_un'] = pop_conjoints_de_retraite.EFFECTIF -
pop_conjoints_de_retraite.D_n + pop_conjoints_de_retraite.N_n +
pop_conjoints_de_retraite.CNinv_n + pop_conjoints_de_retraite.NV_n

pop_conjoints_de_retraite['TR_AGE_n_plus_un'] = pop_conjoints_de_retraite.TR_AGE + 1

pop_conjoints_de_retraite.TYPE_ASS.replace('A', 'R', inplace = True)

pop_conjoints_de_retraite_n_plus_un = pop_conjoints_de_retraite[colones].copy(deep = True)

# Veufs

pop_veufs = pop_a_projetter[np.logical_and(pop_a_projetter.TYPE_ASS == 'V',
pop_a_projetter.TYPE_BEN == 'A')].copy()
pop_veufs = add_origin_category(pop_veufs, 'TR_AGE', 'EFFECTIF')
pop_veufs = pop_veufs.groupby(['REGION', 'SEXE', 'TR_AGE', 'TYPE_ASS',
'TYPE_BEN']).sum().reset_index()

#On extrait les nouveaux veufs des actifs
foo = pop_actifs[['REGION', 'Sexe_conjoints', 'Age_Conjoint', 'NV_n']]
nouv_veufs_actifs = pd.DataFrame(array(foo), columns = ['REGION', 'SEXE', 'TR_AGE',
'NV_n'])

nouv_veufs_actifs = nouv_veufs_actifs.groupby(['REGION', 'SEXE',
'TR_AGE']).sum().reset_index()

#on les insère dans la table des veufs avec un préfixe '_A'
pop_veufs = pd.merge(pop_veufs, nouv_veufs_actifs, how = 'outer', on = ['REGION', 'SEXE',
'TR_AGE']).fillna(0)
pop_veufs.rename( columns = { 'NV_n' : 'NV_n_A'}, inplace = True )

#On extrait les nouveaux veufs des retraités
foo = pop_retraite[['REGION', 'Sexe_conjoints', 'Age_Conjoint', 'NV_n']]
nouv_veufs_retraite = pd.DataFrame(array(foo), columns = ['REGION', 'SEXE', 'TR_AGE',
'NV_n'])

nouv_veufs_retraite = nouv_veufs_retraite.groupby(['REGION', 'SEXE',
'TR_AGE']).sum().reset_index()

#on les insère dans la table des veufs avec un préfixe '_R'
pop_veufs = pd.merge(pop_veufs, nouv_veufs_retraite, how = 'outer', on = ['REGION',
'SEXE', 'TR_AGE']).fillna(0)
pop_veufs.rename( columns = { 'NV_n' : 'NV_n_R'}, inplace = True )

pop_veufs['q_DECES'] = 10

for i in table_mortalite.age:

    pop_veufs['q_DECES'][pop_veufs.TR_AGE == i] = table_mortalite.q.loc[i]

#décès
pop_veufs['D_n'] = pop_veufs.EFFECTIF * pop_veufs.q_DECES

pop_veufs['EV_n'] = adhes_n*pop_veufs.TR

pop_veufs['E_n_plus_un'] = pop_veufs.EFFECTIF - pop_veufs.D_n + pop_veufs.NV_n_A +
pop_veufs.NV_n_R + pop_veufs.EV_n

pop_veufs['TR_AGE_n_plus_un'] = pop_veufs.TR_AGE + 1

pop_veufs_n_plus_un = pop_veufs[colones].copy(deep = True)

pop_veufs_n_plus_un.loc[:, 'TYPE_ASS'] = 'V'
pop_veufs_n_plus_un.loc[:, 'TYPE_BEN'] = 'A'

# Enfants

# Enfants d'actifs x!=0

pop_enfants_actif = pop_a_projetter[np.logical_and(pop_a_projetter.TYPE_ASS == 'A',
pop_a_projetter.TYPE_BEN == 'E')].copy()
pop_enfants_actif = add_origin_category(pop_enfants_actif, 'TR_AGE', 'EFFECTIF')

```

```

pop_enfants_actif = pop_enfants_actif.groupby(['REGION', 'SEXE', 'TR_AGE', 'TYPE_ASS',
'TYPE_BEN']).sum().reset_index()

pop_enfants_actif.EFFECTIF.sum()

#probabilité de scolarisation

_p = [1]*22 + [1 - 0.15*i for i in range(1,7,1)] + [0]*60
Pse = pd.DataFrame({'p' : _p }, index = array(range(0, len(_p) , 1)) )

pop_enfants_actif['Pse'] = 10

for i in Pse.index:

    pop_enfants_actif['Pse'][pop_enfants_actif.TR_AGE == i] = Pse.p.loc[i]

pop_enfants_actif['q_DECES'] = 10

for i in table_mortalite.age:

    pop_enfants_actif['q_DECES'][pop_enfants_actif.TR_AGE == i] = table_mortalite.q.loc[i]

pop_enfants_actif['D_n'] = pop_enfants_actif.EFFECTIF * pop_enfants_actif.q_DECES

#nombre de nouveaux orphelin

#Nouvelles entrées
pop_actifs['Age_enfant'] = pop_actifs.TR_AGE - 30
pop_actifs['EE_n'] = 0
pop_actifs['EE_n'][pop_actifs.TR_AGE >30] = pop_actifs.N_n * Nme.A[0]

#Nouveaux orphelins
pop_actifs['NO_n'] = 0
pop_actifs['NO_n'][pop_actifs.TR_AGE >30] = pop_actifs.D_n * Nme.A[0]

#enfants sortant pour autre que décès du parent
pop_actifs['Se_n'] = 0
pop_actifs['Se_n'][pop_actifs.TR_AGE >30] = pop_actifs.S_n * Nme.A[0]

#On extrait les nouveaux orphelins des retraités
foo = pop_actifs[['REGION', 'Age_enfant', 'NO_n', 'Se_n', 'EE_n']][pop_actifs.Age_enfant
> 0]
nouv_orphelins_sorties_actif = pd.DataFrame(array(foo), columns = ['REGION', 'TR_AGE',
'NO_n', 'Se_n', 'EE_n',])

#Répartir selon le poids des sexes

nouv_orphelins_sorties_actif_M = nouv_orphelins_sorties_actif.copy()
nouv_orphelins_sorties_actif_M[['NO_n', 'Se_n', 'EE_n']] =
nouv_orphelins_sorties_actif_M[['NO_n', 'Se_n', 'EE_n']]*PE.M[0]
nouv_orphelins_sorties_actif_M['SEXE'] = 'M'
nouv_orphelins_sorties_actif_F = nouv_orphelins_sorties_actif.copy()
nouv_orphelins_sorties_actif_F[['NO_n', 'Se_n', 'EE_n']] =
nouv_orphelins_sorties_actif_F[['NO_n', 'Se_n', 'EE_n']]*PE.F[0]
nouv_orphelins_sorties_actif_F['SEXE'] = 'F'

nouv_orphelins_sorties_actif = pd.concat([nouv_orphelins_sorties_actif_M,
nouv_orphelins_sorties_actif_F] , ignore_index = True)

nouv_orphelins_sorties_actif = nouv_orphelins_sorties_actif.groupby(['REGION', 'TR_AGE',
'SEXE']).sum().reset_index()

#Nous devons joindre nouv_orphelins_actif et pop_enfants_actif

pop_enfants_actif = pd.merge(pop_enfants_actif, nouv_orphelins_sorties_actif, how =
'outer', on = ['REGION', 'SEXE', 'TR_AGE'] )
pop_enfants_actif.EFFECTIF.fillna(0, inplace = True)
pop_enfants_actif.TYPE_ASS.fillna('A', inplace = True)

pop_enfants_actif['E_n_plus_un'] = pop_enfants_actif.Pse*(pop_enfants_actif.EFFECTIF +

```

```

pop_enfants_actif.EE_n
pop_enfants_actif.Se_n
pop_enfants_actif.D_n
pop_enfants_actif.NO_n

pop_enfants_actif['TR_AGE_n_plus_un'] = pop_enfants_actif.TR_AGE + 1
# Enfants d'actifs x = 0
#Nouvelles naissancee (concerne les femmes en âge de procréation)
#issus des actifs
ages_procreation = np.logical_and(pop_actifs.TR_AGE >15, pop_actifs.TR_AGE < 50)
pop_actifs['NNE_n'] = 0
pop_actifs['NNE_n'][np.logical_and(pop_actifs.SEXE == 'F', ages_procreation)] = Tfec *
pop_actifs.EFFECTIF
#issus des conjoint d'actifs
ages_procreation = np.logical_and(pop_conjoints_d_actifs.TR_AGE >15,
pop_conjoints_d_actifs.TR_AGE < 50)
pop_conjoints_d_actifs['NNE_n'] = 0
pop_conjoints_d_actifs['NNE_n'][np.logical_and(pop_conjoints_d_actifs.SEXE == 'F',
ages_procreation)] = Tfec * pop_conjoints_d_actifs.EFFECTIF
nouveaux_nes_actifs = pd.concat([pop_actifs[['REGION', 'TYPE_ASS', 'TYPE_BEN', 'NNE_n']],
pop_conjoints_d_actifs[['REGION', 'TYPE_ASS', 'TYPE_BEN', 'NNE_n']]], ignore_index = 1)
nouveaux_nes_actifs = nouveaux_nes_actifs.groupby(['REGION', 'TYPE_ASS',
'TYPE_BEN']).sum().reset_index()
nouveaux_nes_actifs_M = nouveaux_nes_actifs.copy()
nouveaux_nes_actifs_M['NNE_n'] = nouveaux_nes_actifs_M['NNE_n']*PE.M[0]
nouveaux_nes_actifs_M['SEXE'] = 'M'
nouveaux_nes_actifs_F = nouveaux_nes_actifs.copy()
nouveaux_nes_actifs_F['NNE_n'] = nouveaux_nes_actifs_F['NNE_n']*PE.F[0]
nouveaux_nes_actifs_F['SEXE'] = 'F'
nouveaux_nes_actifs = pd.concat([nouveaux_nes_actifs_M, nouveaux_nes_actifs_F],
ignore_index = True)
nouveaux_nes_actifs['TR_AGE_n_plus_un'] = 0
nouveaux_nes_actifs.rename( columns = { 'NNE_n' : 'E_n_plus_un'}, inplace = True )
nouveaux_nes_actifs['TYPE_BEN'] = 'E'
#Les enfants issu d'actif(n) contiennent les enfants des nouveaux retraités
pop_enfants_actif_n_plus_un = pd.concat([pop_enfants_actif[colones].copy(deep = True),
nouveaux_nes_actifs])
# Enfants de retraités
#
pop_enfants_retraite = pop_a_projetter[np.logical_and(pop_a_projetter.TYPE_ASS == 'R',
pop_a_projetter.TYPE_BEN == 'E')].copy()
pop_enfants_retraite = add_origin_category(pop_enfants_retraite, 'TR_AGE', 'EFFECTIF')
pop_enfants_retraite = pop_enfants_retraite.groupby(['REGION', 'SEXE', 'TR_AGE',
'TYPE_ASS', 'TYPE_BEN']).sum().reset_index()
pop_enfants_retraite['Pse'] = 10
for i in Pse.index:
pop_enfants_retraite['Pse'][pop_enfants_retraite.TR_AGE == i] = Pse.p.loc[i]
pop_enfants_retraite['q_DECES'] = 10

```

```

for i in table_mortalite.age:
    pop_enfants_retraite['q_DECES'][pop_enfants_retraite.TR_AGE == i] =
table_mortalite.q.loc[i]

    pop_enfants_retraite['D_n'] = pop_enfants_retraite.EFFECTIF *
pop_enfants_retraite.q_DECES

#Nouvelles entrées
pop_retraite['Age_enfant'] = pop_retraite.TR_AGE - 30
pop_retraite['EE_n'] = 0
pop_retraite['EE_n'][pop_retraite.TR_AGE >30] = pop_retraite.N_n * Nme.A[0]

pop_enfants_retraite.EFFECTIF.fillna(0, inplace = True)
pop_enfants_retraite.TYPE_ASS.fillna('R', inplace = True)

pop_enfants_retraite['E_n_plus_un'] =
pop_enfants_retraite.Pse*(pop_enfants_retraite.EFFECTIF - pop_enfants_retraite.D_n)

pop_enfants_retraite['TR_AGE_n_plus_un'] = pop_enfants_retraite.TR_AGE + 1
pop_enfants_retraite.TYPE_BEN.fillna('E', inplace = True)

pop_enfants_retraite_n_plus_un = pop_enfants_retraite[colones].copy(deep = True)

#La table des actifs contient quelque retraité, on les dissocie
foo = pd.concat([pop_enfants_retraite_n_plus_un, pop_enfants_actif_n_plus_un])

pop_enfants_retraite_n_plus_un = foo[foo.TYPE_ASS == 'R'].copy(deep = True)

pop_enfants_actif_n_plus_un = foo[foo.TYPE_ASS != 'R'].copy(deep = True)
pop_enfants_actif_n_plus_un.loc[:, 'TYPE_ASS'] = 'A'

# Orphelins

pop_orphelins = pop_a_projetter[np.logical_and(pop_a_projetter.TYPE_ASS == 'O',
pop_a_projetter.TYPE_BEN == 'A')].copy()
pop_orphelins = add_origin_category(pop_orphelins, 'TR_AGE', 'EFFECTIF')
pop_orphelins = pop_orphelins.groupby(['REGION', 'SEXE', 'TR_AGE', 'TYPE_ASS',
'TYPE_BEN']).sum().reset_index()

pop_orphelins['q_DECES'] = 10

for i in table_mortalite.age:
    pop_orphelins['q_DECES'][pop_orphelins.TR_AGE == i] = table_mortalite.q.loc[i]

pop_orphelins['D_n'] = pop_orphelins.EFFECTIF * pop_orphelins.q_DECES

#Les entrées des orphelins issus des 114
pop_orphelins['EO_n'] = adhes_n * pop_orphelins.TR

#Les nouveaux orphelins issus des actifs figurent déjà dans la table des actifs

#les nouveaux orpheins issus des retraités
pop_retraite['NO_n'] = 0
pop_retraite['NO_n'] = pop_retraite.D_n * Nme.R[0]*0

#On ajoute l'age de l'enfant
pop_retraite['Age_enfant'] = 0
pop_retraite['Age_enfant'] = pop_retraite.TR_AGE - 30

#ces orphelins doivent être répartis selon le sexe
foo = pop_retraite[['REGION', 'Age_enfant', 'NO_n']][pop_retraite.Age_enfant > 0]
nouv_orphelins_et_114_retraite = pd.DataFrame(array(foo), columns = ['REGION', 'TR_AGE',
'NO_n', ])

nouv_orphelins_et_114_retraite_M = nouv_orphelins_et_114_retraite.copy()
nouv_orphelins_et_114_retraite_M[['NO_n', ]] = nouv_orphelins_et_114_retraite_M[['NO_n',
]]*PE.M[0]
nouv_orphelins_et_114_retraite_M['SEXE'] = 'M'

nouv_orphelins_et_114_retraite_F = nouv_orphelins_et_114_retraite.copy()
nouv_orphelins_et_114_retraite_F[['NO_n']] =

```

```

nouv_orphelins_et_114_retraite_F[['NO_n']]*PE.F[0]
nouv_orphelins_et_114_retraite_F['SEXE'] = 'M'

nouv_orphelins_et_114_retraite = pd.concat([nouv_orphelins_et_114_retraite_M,
nouv_orphelins_et_114_retraite_F], ignore_index = True)
nouv_orphelins_et_114_actifs = nouv_orphelins_sorties_actif[['REGION', 'TR_AGE', 'SEXE',
'NO_n']]
nouv_orphelins_et_114_actifs =
nouv_orphelins_et_114_actifs[nouv_orphelins_et_114_actifs.TR_AGE < 18]

nouv_orphelins_et_114_retraite = nouv_orphelins_et_114_retraite.groupby(['REGION',
'TR_AGE', 'SEXE']).sum().reset_index()
nouv_orphelins_et_114_actifs = nouv_orphelins_et_114_actifs.groupby(['REGION', 'TR_AGE',
'SEXE']).sum().reset_index()

nouv_orphelins_et_114 = pd.merge(nouv_orphelins_et_114_retraite,
nouv_orphelins_et_114_actifs, how = 'outer', on = ['REGION', 'SEXE', 'TR_AGE'], suffixes =
('_R', '_A'))
nouv_orphelins_et_114.fillna(0, inplace = True)

nouv_orphelins_et_114 = nouv_orphelins_et_114.groupby(['REGION', 'SEXE',
'TR_AGE']).sum().reset_index()

#Le tableaux "nouv_orphelins_et_114" contient les nouveau orphelins d'actifs et de
rertaités
#Nous le joignons a "pop_orphelins"
pop_orphelins = pd.merge(pop_orphelins, nouv_orphelins_et_114, how = 'outer', on =
['REGION', 'SEXE', 'TR_AGE'], )
pop_orphelins.EFFECTIF.fillna(0, inplace = True)

pop_orphelins['TR_AGE_n_plus_un'] = pop_orphelins.TR_AGE - 1

pop_orphelins['E_n_plus_un'] = (pop_orphelins.EFFECTIF + pop_orphelins.EO_n -
pop_orphelins.D_n
+ pop_orphelins.NO_n_A + pop_orphelins.NO_n_R )

pop_orphelins_n_plus_un = pop_orphelins[colones].copy(deep = True)

pop_orphelins_n_plus_un.TYPE_ASS = 'O'

#
# Les tables obtenues sont les suivantes :
#
#     pop_actifs_n_plus_un
#     pop_retraite_n_plus_un
#     pop_conjoints d actifs n plus un
#     pop conjoints de retraite_n_plus_un
#     pop_veufs_n_plus_un
#     pop_enfants_actif_n_plus_un
#     pop_enfants_retraite_n_plus_un
#     pop_orphelins_n_plus_un
#

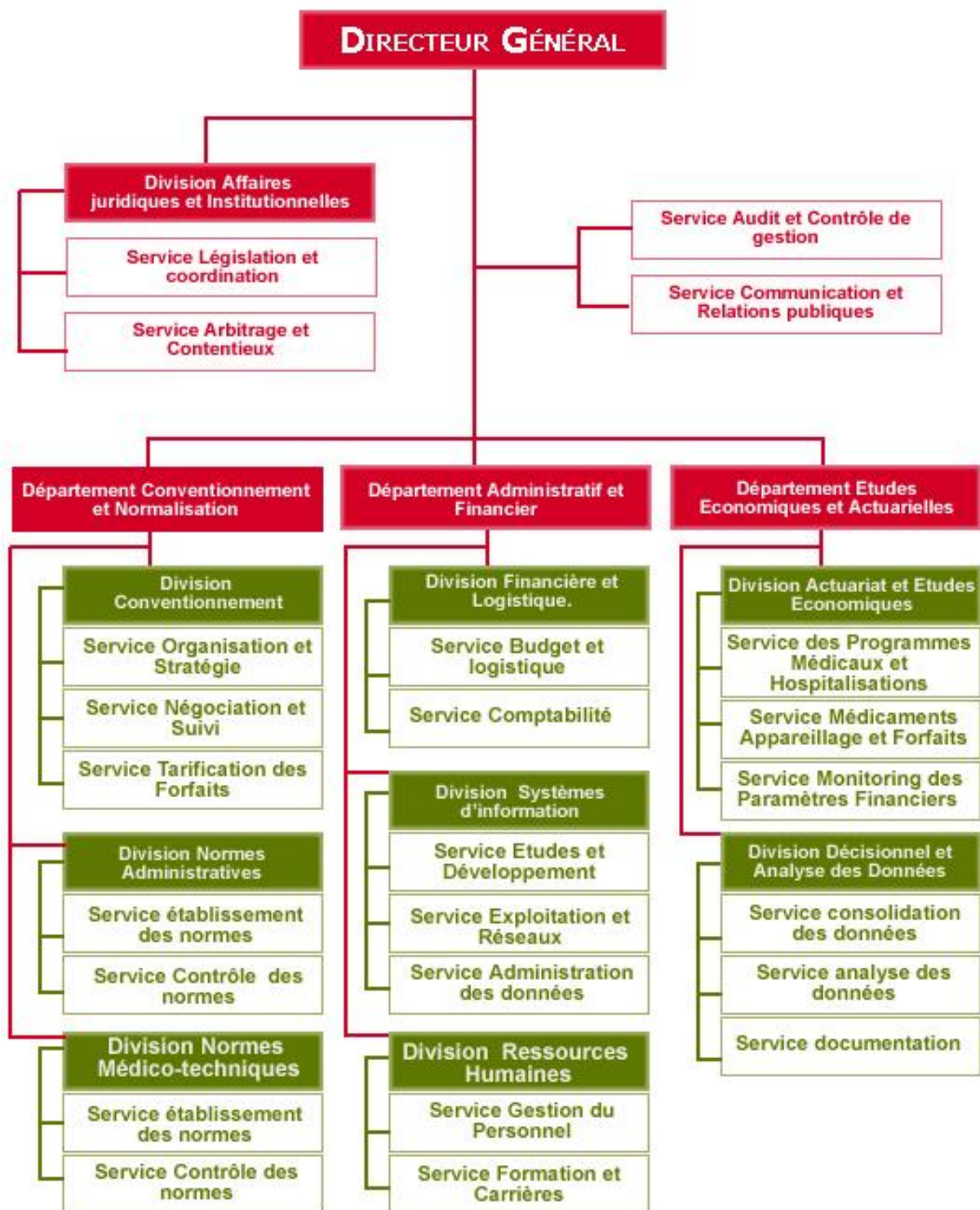
pop_n_plus_un_list = [pop_actifs_n_plus_un, pop_retraite_n_plus_un,
pop_conjoints_d_actifs_n_plus_un, pop_conjoints_de_retraite_n_plus_un,
pop_veufs_n_plus_un, pop_enfants_actif_n_plus_un,
pop_enfants_retraite_n_plus_un, pop_orphelins_n_plus_un]

pop_n_plus_un = pd.concat(pop_n_plus_un_list, ignore_index = True)
pop_n_plus_un = pop_n_plus_un.groupby(['REGION', 'SEXE', 'TR_AGE_n_plus_un', 'TYPE_ASS',
'TYPE_BEN']).sum().reset_index()

return pop_n_plus_un

```

IV. Organigramme de l'ANAM



V. Liste des ALD

1- Accident vasculaire cérébral ou médullaire ischémique ou hémorragique
2- Affections malignes du tissu lymphatique ou hématopoïétique
3- Anémies hémolytiques chroniques sévères
4- Aplasies médullaires sévères
5- Artériopathies chroniques
6- Asthme sévère
7- Cardiopathies congénitales
8- Cirrhoses du foie
9- Diabète insulino-dépendant et diabète non insulino-dépendant
10- Epilepsie grave
11- Etat de déficit mental
12- Formes graves des affections neurologiques et neuromusculaires
13- Glaucome chronique
14- Hypertension artérielle sévère
15- Insuffisance cardiaque
16- Insuffisance rénale aiguë
17- Insuffisance rénale chronique terminale
18- Insuffisance respiratoire chronique grave
19- Lupus érythémateux aigu disséminé
20- Maladie coronaire
21- Maladie de Crohn évolutive
22- Maladie de Parkinson
23- Maladies chroniques actives du foie (hépatites B et C)
24- Myélodysplasies sévères
25- Néphropathies graves
26- Polyarthrite rhumatoïde évolutive grave
27- Psychoses
28- Rectocolite hémorragique évolutive
29- Rétinopathie diabétique
30- Sclérodémie généralisée évolutive
31- Sclérose en plaques
32- Spondylarthrite ankylosante grave
33- Syndrome d'immunodéficience acquise (SIDA)
34- Syndromes néphrotiques
35- Troubles graves de la personnalité
36- Troubles héréditaires de l'hémostase
37- Troubles mentaux et/ou de personnalité dus à une lésion, à un dysfonctionnement cérébral ou à une lésion physique
38- Troubles permanents du rythme et de la conductivité
39- Tumeurs malignes
40- Valvulopathies rhumatismales
41- Vascularites