

CONFERENCE INTERAFRICAINNE DES MARCHES D'ASSURANCES



MEMOIRE DE FIN DE FORMATION DIPLOME D'ETUDES SUPERIEURES SPECIALISEES EN ASSURANCES

THEME

**ANALYSE STATISTIQUE ET MODELISATION DE
LA MORTALITE DU CONTRAT SIKA RETRAITE
DE NSIA VIE**

ENCADREUR :

M^{me} Yvette AKOUA
*Directeur technique et de
l'actuariat de la Holding du
Groupe NSIA*

REALISE PAR :

Djagana OUATTARA
*Etudiant en DESSA
à l'IIA de Yaoundé*

17^e Promotion DESSA

AVANT-PROPOS

L'Institut International des Assurances de Yaoundé (IIA) est l'organe de la Conférence Interafricaine des Marchés d'Assurances (CIMA), chargé de la formation des cadres en assurance dans les pays membres. Créé en 1972, l'Institut a formé un nombre assez important de cadres qui exercent leurs fonctions aussi bien dans les Directions Nationales d'Assurances que dans les sociétés privées d'assurance et de réassurance.

« Learning by Doing »-Apprendre par la Pratique-, est la méthode idéale d'apprentissage à laquelle l'Institut n'est pas resté en marge. Ainsi après une formation théorique de qualité, un stage pratique de six (6) mois est prévu en entreprise afin de livrer aux étudiants les aspects pratiques de l'assurance.

C'est dans ce cadre que nous nous sommes retrouvés à Nouvelle Société Interafricaine d'Assurances (NSIA Vie) de Côte d'Ivoire. NSIA Vie a racheté en 1996 Assurances Générales de Côte d'Ivoire (AGCI-Vie) qui avait été créée en 1988. Aujourd'hui NSIA Vie fait partie du Groupe NSIA qui s'est implanté dans la plupart des pays membres de la CIMA.

Notre étude a consisté à modéliser la mortalité du portefeuille **Sika Retraite**. Ce thème est d'un grand intérêt en ce sens qu'il peut constituer un outil de contrôle et de sélection des risques à la souscription.

Que cet écrit soit l'occasion pour nous d'adresser nos remerciements au Directeur Général de l'IIA, Monsieur Jean-Grattien ZANOVI pour sa contribution à la formation des cadres assureurs en Afrique.

Nous remercions aussi Monsieur Momath NDAO, Directeur des études à l'IIA, pour sa compétence et l'effort fourni pour former des assureurs de plus en plus performants.

Nous remercions également la Direction de NSIA Vie de Côte d'Ivoire qui a bien voulu nous recevoir dans ses locaux et nous livrer des connaissances pratiques en assurance.

Nous remercions également toute notre famille pour le soutien permanent au cours de notre cursus scolaire.

Nos remerciements s'adressent enfin à tous ceux qui de près ou de loin nous ont aidé à la réussite de ce travail.

Nous ne pouvons espérer que ce mémoire soit dépourvu de toute obscurité, nous le soumettons pour autant à votre appréciation.

SOMMAIRE

	Page
AVANT-PROPOS	1
RESUME	3
ABSTRACT	4
INTRODUCTION	5
PARTIE I : ETUDE STATISTIQUE DU PORTEFEUILLE « SIKA RETRAITE »	8
CHAPITRE I : STATISTIQUE DESCRIPTIVE DU PORTEFEUILLE « SIKA RETRAITE »	9
I. CARACTERISTIQUES DU CONTRAT « SIKA RETRAITE »	9
A. Catégorie du contrat	9
B. Principes de fonctionnement	9
II. STATISTIQUES DU PORTEFEUILLE	11
A. Population des assurés « Sika Retraite »	11
B. Population des décès intervenus	13
CHAPITRE II : STATISTIQUE EXPLORATOIRE DU PORTEFEUILLE « SIKA RETRAITE »	16
I. METHODES D'ANALYSE FACTORIELLE DES DONNEES	17
A. Analyse factorielle : Géométrie de l'information	17
B. Différentes méthodes d'analyse factorielle	17
II. PRESENTATION DES RESULTATS DE L'ANALYSE	19
A. Choix de la méthode	19
B. Résultats de l'analyse	20
PARTIE II : MODELISATION DE LA MORTALITE DU PORTEFEUILLE « SIKA RETRAITE »	26
CHAPITRE I: CLASSIFICATION AUTOMATIQUE DES ASSURES	27
I. METHODE DE CLASSIFICATION	27
A. Objet	27
B. Raisons de la classification	28
II. RESULTATS	28
A. Choix du nombre de classes	28
B. Description des classes	29
CHAPITRE II : MODELISATION DE LA PROBABILITE DE DECES	33
I. MODELISATION DECISIONNELLE	33
A. Analyse discriminante	33
B. Résultats de l'analyse	34
II. REGRESSION LOGISTIQUE	36
A. Modèle dichotomique	36
B. Résultats du modèle	38
CONCLUSION	46
ANNEXES	48
BIBLIOGRAPHIE	52

RESUME

La tarification d'un contrat d'assurance exige en plus des éléments techniques, la connaissance de certaines caractéristiques de l'assuré. C'est une tâche importante car la solvabilité de l'assuré en dépend fortement. Pourtant cette tâche n'est pas facile quand on sait que ce qui vient de l'assuré comme renseignements n'est pas garanti de fiabilité, empêchant ainsi l'assureur de mieux apprécier le risque. Dans cette perspective, l'assureur est condamné de se munir d'assez de moyens de contrôle complémentaires. Ainsi, notre objectif dans la présente étude est de fournir un outil performant permettant d'apprécier à la souscription du contrat le risque que souhaite couvrir l'assureur : le décès de l'assuré.

L'assurance, et en particulier l'assurance vie a la propriété de n'attirer en majorité que des personnes dont la réalisation du risque est quasi certain. Par exemple, les personnes malades ont tendance à souscrire les contrats en cas de décès. Pour cela, l'étude a été menée de manière à pronostiquer le décès ou la survie d'un assuré lorsque l'assureur dispose des caractéristiques de celui-là. Ces caractéristiques sont: *l'âge à la souscription du contrat, le sexe, le secteur d'activité et le lieu de résidence de l'assuré.*

Un modèle dit dichotomique a été élaboré permettant d'évaluer la probabilité de décès pendant la durée du contrat connaissant les caractéristiques ci-dessus de l'assuré. Pour l'atteinte de cet objectif, une conduite méthodique a été élaborée. Une étude statistique préalable a été faite et a permis d'explorer le portefeuille, toute chose ayant permis de conduire avec efficacité la modélisation.

ABSTRACT

The pricing of an insurance contract requires in addition to the technical elements, the knowledge of some feature that the insured arranges. It is an important task because the insurer's solvency depends strongly on it. Yet this task is not easy when we know that what comes from the insured as information is not guaranteed of reliability, stopping the insurer thus to appreciate the risk more better. In this perspective, the insurer is condemned to carry enough means of controls complementary. Thus, our objective in the present survey is to provide an effective tool permitting to appreciate at the subscription of the contract the risk that wishes to cover the insurer: the insured's death.

Insurance, and in particular life assurance has the property to attract in majority people of which the realization of the risk is quasi certain. For example, sick people have tendency to subscribe the contracts in case of death. For it, the survey has been led in order to forecast the death or an insured's survival when the insurer has features of that one. These features are: *the age at the subscription of the contract, the sex, the business stream and the place of residence of the insured.*

A model said dichotomic has been elaborated permitting to value the probability of death during the length of the contract when the insured's features seen above are known. To reach this objective, a methodical conduct has been elaborated. A previous statistical study has been made and has permitted to explore the wallet that permits to lead the modelling with efficiency.

INTRODUCTION

La création des richesses couplée avec l'augmentation du patrimoine détenu par les agents économiques, entreprises et personnes physiques, suscite le besoin incessant de se couvrir contre les aléas. La production d'une entreprise peut être interrompue par la survenance d'un incendie. Des aléas naturels aux aléas économiques en passant par ceux d'ordre humain, les entreprises sont perpétuellement soumises au dictat de ces phénomènes, au point où leurs survies dépendent fortement de la fréquence de survenance et de la gravité de tels phénomènes. De même la disparition précoce d'un père ou d'une mère, chef de ménage, peut porter préjudice grave à sa progéniture.

L'une des méthodes traditionnelles de prévention des risques est la réduction c'est à dire la mise oeuvre de moyens pour diminuer la probabilité de survenance des événements. Aussi puissants que soient les moyens de prévention et de protection contre les risques, aucune garantie sûre ne s'offre au détenteur d'un patrimoine. Ce qui explique le transfert du risque à un professionnel de gestion de risques, l'Assureur, qui se charge de couvrir le risque. D'où la définition de l'assurance suivante : *l'assurance est une opération par laquelle une partie, l'assureur s'engage, moyennant le paiement d'une somme d'argent appelée cotisation ou prime, à verser une prestation à une autre partie, l'assuré, en cas de réalisation du risque.*

Dans cette logique, deux volontés contradictoires sont en présence ; l'assureur ayant pour souhait de ne garantir que des risques moins graves tandis que le transférant du risque ne met à la charge de l'assureur que ce qu'il ne peut supporter, donc les risques lourds.

En conséquence, l'un des soucis de l'assureur est la fixation du juste prix du transfert du risque de manière à pouvoir faire face au coût du risque en cas de survenance de sinistre. En effet l'assureur fixe le prix de vente de ses produits avant de connaître plus tard le prix d'achat contrairement à ce qui se passe dans l'industrie traditionnelle : c'est le phénomène de l'inversion du cycle de production dans l'industrie de l'assurance. L'assureur est donc en perpétuel contrôle de son portefeuille. Et ce contrôle est plus que nécessaire pour l'assureur vie africain en raison du fait que le marché est encore jeune et en pleine croissance.

La cotation d'un contrat exige la connaissance des statistiques anciennes et récentes portant sur le portefeuille du contrat en question. Dans cette perspective, l'assureur est en permanence soumis au risque d'insolvabilité lié à la non maîtrise de ces statistiques.

En assurance de personnes, si la cotation des contrats en cas de vie échappe à cette exigence de connaissance des statistiques, les contrats en cas de décès ont plus que besoin des

statistiques du passé. La justification est que la provision mathématique constituée au cours du contrat n'est pas suffisante pour faire face au paiement du capital en cas de survenance du décès. Il revient à l'assureur de recourir au principe de mutualisation.

Le décès est une fatalité. Il intervient à un moment donné et met fin à la vie d'un individu. Les démographes se sont donnés pour tâche de modéliser la survenance des décès dans une population par des tables dites de Mortalité. Les assureurs vendent à la population des contrats portant sur la vie et utilisent les tables de mortalité pour la tarification des contrats. Les contrats en cas de décès garantissent le paiement d'un capital au décès de l'assuré. Différentes options de garanties sont offertes ; le caractère aléatoire du contrat demeurant parfois dans le nombre de primes à payer par le souscripteur ou dans la survenance du décès lui même.

En assurance en cas de décès, du fait de l'insuffisance de la provision mathématique constituée, une mutualisation est faite par l'assureur qui se propose dans cette tâche de répartir l'ensemble des primes d'une année donnée entre les bénéficiaires désignés pour les décès intervenus. Les primes des personnes vivantes servent donc par conséquent au paiement des capitaux des assurés décédés.

Pour ne pas faire peser sur la mutualité les personnes à haut risque, l'assureur se charge de soumettre les assurés à un questionnaire médical. Parfois le questionnaire médical n'est pas un bon moyen de sélection des risques. Pour cela, nous nous proposons d'étudier le profil des clients à « *haut risque* » pour la Compagnie. Sera considéré comme client à « *haut risque* » tout assuré dont le décès intervient avant la cinquième année du contrat.

Il est à noter que la tarification des contrats en cas de décès ne se fait pas exactement comme en assurance dommage. En assurance dommage, les statistiques des années précédentes servent à tarifer les contrats de l'année d'assurance. La masse des sinistres estimée est répartie sur l'ensemble des assurés. En revanche, en assurance en cas de décès, la prime de chaque individu est déterminée en tenant compte de sa probabilité de décès donnée par les tables de mortalité, et des capitaux garantis pour une année d'assurance. Et bien pourtant dans la gestion du portefeuille l'assureur procède à un système de mutualisation. La probabilité de décès est une valeur numérique. Aussi faible soit-elle, cela ne signifie pas que l'assuré ne décèdera pas. Il se pose donc le problème de mesure et la situation réelle du portefeuille. C'est la raison pour laquelle nous orientons notre étude dans l'objectif d'évaluer la probabilité de décès de l'assuré avant la cinquième année du contrat. D'où le thème : **«Analyse statistique et modélisation de la mortalité du contrat Sika Retraite».**

Pour mener à bien cette tâche, nous avons décidé de traiter le sujet en deux parties. La première partie, traitant de la description statistique, comprend deux chapitres où il sera question dans l'un de faire une simple description statistique et dans l'autre une analyse statistique exploratoire du portefeuille **Sika Retraite**.

La deuxième partie, traitant de la modélisation du portefeuille du contrat en question, se subdivise aussi en deux chapitres. Le premier se charge de classer les assurés et le deuxième de modéliser la probabilité de survenance de décès d'un assuré avant la cinquième année de contrat.

PARTIE I : ETUDE STATISTIQUE DU PORTEFEUILLE « SIKA RETRAITE »

CHAPITRE I : STATISTIQUE DESCRIPTIVE DU PORTEFEUILLE « SIKA RETRAITE »

La statistique descriptive donne une présentation des données sous la forme de tableaux et graphiques. Elle fait ressortir également les caractéristiques de tendances centrales (la moyenne, le mode...), les caractéristiques de dispersion (variance, écart moyen, quartiles...) et les caractéristiques de concentration.

L'intérêt de ce chapitre est de nous fournir les outils en vue de la modélisation de la probabilité de décès. En présentant les données, nous aurons l'avantage d'émettre les hypothèses réalistes sur les variables influençant la probabilité de décès.

Avant de rechercher les hypothèses, nous présentons le contrat faisant l'objet de notre étude.

I. CARACTERISTIQUES DU CONTRAT « SIKA RETRAITE »

A. Catégorie du contrat

1. Objet du contrat

Le contrat individuel d'assurance **SIKA RETRAITE** offre deux garanties :

- La garantie **retraite** qui a pour objet de permettre au contractant de constituer une épargne ;

- La garantie complémentaire et optionnelle **décès** qui paie le capital souscrit en cas de décès de l'assuré avant le terme de la garantie. Cette garantie est d'une durée de cinq (5) ans renouvelable et n'est acquise qu'après six (6) mois effectifs d'assurance sauf cas de décès accidentel.

2. Type du contrat

Les contrats d'assurance vie se combinent le plus souvent pour donner des produits ayant à la fois la garantie en cas de décès et/ou la garantie en cas de vie et une composante épargne. **NSIA Vie** a élaboré un contrat garantissant un capital en cas de décès de l'assuré et le paiement de l'épargne constituée par le client. C'est un produit Vie épargne avec une option décès.

B. Principes de fonctionnement

1. Population assurable

Pour souscrire au contrat **Sika retraite** le proposant doit être âgé d'au moins 12 ans et d'au plus 65 ans. Un questionnaire médical est mis en place pour servir de contrôle de la santé

des proposant. Son rôle est de ne pas inclure dans le portefeuille des individus dont l'état de santé est hors norme ou à défaut de faire payer une surprime. En fait, il s'agit pour l'assureur non seulement de protéger le portefeuille contre les risques de surmortalité mais également de ne pas faire peser sur la mutualité des personnes dont la probabilité de décès est largement au dessus de la moyenne.

2. Intervenants au contrat

2.1. Obligations de l'assuré

La cotisation est fixée librement par le contractant. Le minimum mensuel requis est de FCFA 10 000. Cette cotisation est composée d'une prime décès et d'une prime épargne. L'assuré fixe à la souscription le montant du capital décès. La prime décès est donc déterminée en fonction de ce capital et dont le paiement reste obligatoire pour l'assuré. La prime épargne quant à elle n'est pas obligatoire compte tenu du caractère non obligatoire de l'épargne. On ne peut forcer un individu à épargner.

2.2. Obligations de NSIA Vie

Au titre de la **retraite**, le contrat garantit un capital minimum auquel s'ajoutent les participations aux bénéfices et une prime de fidélité.

Le capital minimum est égal aux conditions nettes de frais, capitalisé aux taux d'intérêt de 3,5%. Les frais sont les suivants :

	1 ^{ère} Année	2 ^{ème} à la 10 ^{ème} Année	A partir de la 11 ^{ème} Année
Charges annuelles fixes	12 000	12 000	0
Frais d'opération sur cotisations	60%	8%	5%

Les participations aux bénéfices sont distribuées à hauteur de 85% des résultats financiers et 90% des résultats techniques.

Les primes de fidélité sont versées après **la cinquième, dixième et quinzième année d'assurance** aux contractants qui ont payé régulièrement leurs cotisations et qui n'ont procédé à aucune opération d'avance ou de rachat partiel. Elles sont cumulatives et égales respectivement à **20%, 40% et 60% de la cotisation de la première année d'assurance**.

La prime de fidélité n'est acquise que si le contrat arrive à son terme contractuel.

Au titre de la garantie **décès**, le capital est plafonné à au plus dix (10) fois la cotisation annuelle avec un maximum de **10 000 000 Francs CFA**.

II. STATISTIQUES DU PORTEFEUILLE¹

A. Population des assurés « Sika Retraite »

1. Composition du portefeuille

La base qui a été mise à notre disposition comporte 11.994 assurés. Les entrées en portefeuille de ces assurés se sont fait au fil du temps depuis le lancement du contrat en 1998. On a assisté aussi à des sorties dont les causes sont les rachats totaux, les décès et les contrats arrivés à terme. Du fait que le contrat **Sika retraite** a une composante *épargne* le rachat est possible. Mais dans notre analyse nous nous intéressons seulement à la composante *décès*. En conséquence, dans la suite lorsque nous parlerons de **Sika Retraite** nous réduisons ce contrat à la composante *décès*.

2. Structure du portefeuille

2.1. Répartition de la population assurée en fonction du sexe

Le portefeuille des assurés se compose de 78% d'hommes et de 22% de femmes. Les études démographiques montrent que dans une population, l'on a en moyenne 49% d'hommes et 51% de femmes. En vertu de cette loi nous pouvons affirmer que les hommes ont une propension plus élevée à souscrire le contrat que les femmes. Il est à signaler qu'en tenant compte de la proportion de femmes exerçant un métier hors activité ménagère on constate que la proportion est dans les limites normales car peu de femmes exerce une activité en dehors du ménage en Côte d'Ivoire.

Tableau 1 : Répartition de la population assurée en fonction du sexe de l'assuré.

	Effectif	Fréquence
Hommes	9412	78%
Femmes	2582	22%
Total	11994	100%

Source : Nos calculs

2.2. Répartition de la population assurée en fonction du lieu de résidence

En prenant comme clé de répartition le *lieu de résidence*, on constate qu'un sur deux assurés habite la ville d'Abidjan. Ce qui signifie que le contrat est un peu partout commercialisé sur le territoire ivoirien. Cependant les zones du Nord ne sont pas concernées par la vente en raison de la situation politique du moment dans le pays.

¹ Les proportions interprétées dans cette partie ont été testées. Nous prions le lecteur de bien vouloir se référer aux annexes pour plus d'informations sur le test de proportion.

Tableau 2 : Répartition de la population assurée en fonction du lieu de résidence de l'assuré

	Effectif	Fréquence
Abidjan	6042	50%
Hors Abidjan	5952	50%
Total	11994	100%

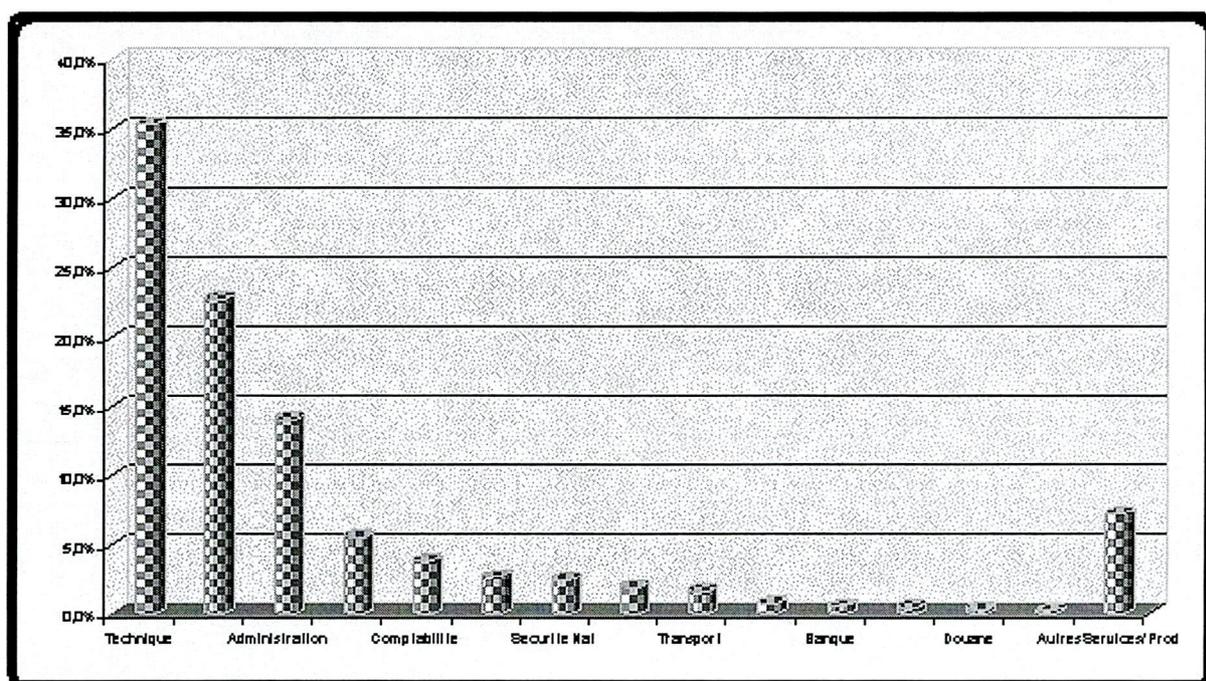
Source : Nos calculs

2.3. Répartition de la population assurée en fonction du secteur d'activité

La population des assurés ayant souscrit la garantie décès est composée de 35,5 % de personnes exerçant une activité relevant du domaine technique. Ce sont les ingénieurs techniques, les techniciens, les informaticiens, les employés des autres industries techniques. Les personnes en charge de l'éducation, les instituteurs, les professeurs, les directeurs d'écoles, les chercheurs, représentent 22,8% de l'ensemble du portefeuille. Viennent ensuite ceux exerçant une activité dans l'administration avec 14,1%. Il s'agit entre autres des directeurs administratifs, les secrétaires de Direction, etc.

Cette composition du portefeuille nous indique que les personnes exerçant une activité informelle s'assurent moins. Cela pourrait s'expliquer par le fait que ces personnes ignorent l'existence de l'assurance. Si tel n'est pas le cas, le manque de moyens financiers pour ces personnes ne leur permet pas d'épargner et encore moins de souscrire un contrat d'assurance.

Graphique 1 : Répartition de la population assurée en fonction du secteur d'activité de l'assuré



Source : Notre étude

B. Population des décès intervenus**1. Composition du portefeuille décès**

De la naissance du contrat au jour de la réception des données, 169 décès sont intervenus dans le portefeuille en question. Dans la suite nous ferons une décomposition plus détaillée de ces décès.

2. Structure du portefeuille décès**2.1. Répartition de la population assurée décédée en fonction de l'âge à la souscription et du sexe**

Le portefeuille décès du contrat se répartit de 84% d'hommes et 16% de femmes. En comparaison au portefeuille des assurés nous remarquons que moins de femmes décèdent que d'hommes puisque les proportions étaient de 78% pour les hommes contre 22% pour les femmes. L'on pourrait en conséquence signifier que la mortalité du portefeuille des clients du contrat est liée au sexe.

Les assurés du contrat décédés ont en moyenne, l'âge de 38 ans à la souscription. Les hommes comme les femmes ont le même âge moyen à la souscription de 38 ans. L'âge moyen à la souscription du contrat de l'ensemble des clients du portefeuille est de 36 ans. Ce qui nous amène à conclure que l'âge à la souscription est une variable déterminante qui influe sur la probabilité de décès d'un assuré.

Tableau 3 : Répartition des assurés décédés en fonction du sexe et de l'âge.

	Sexe			Total
	Hommes	Femmes		
Groupe d'âge (ans)	20-24	2%	0%	2%
	25-29	9%	16%	10%
	30-34	22%	26%	22%
	35-39	21%	11%	19%
	40-44	25%	16%	24%
	45-49	19%	21%	19%
	50-54	2%	5%	2%
	55-59	1%	5%	2%
	Total	100%	100%	100%

Source : Nos calculs

2.2. Répartition de la population assurée décédée en fonction du lieu de résidence

Alors qu'un sur deux assurés habitent hors de la ville d'Abidjan, plus des deux tiers des assurés, soit 69 %, résidant hors d'Abidjan sont décédés en cours de contrat.

Tableau 4 : Répartition de la population assurée décédée en fonction du sexe de l'assuré.

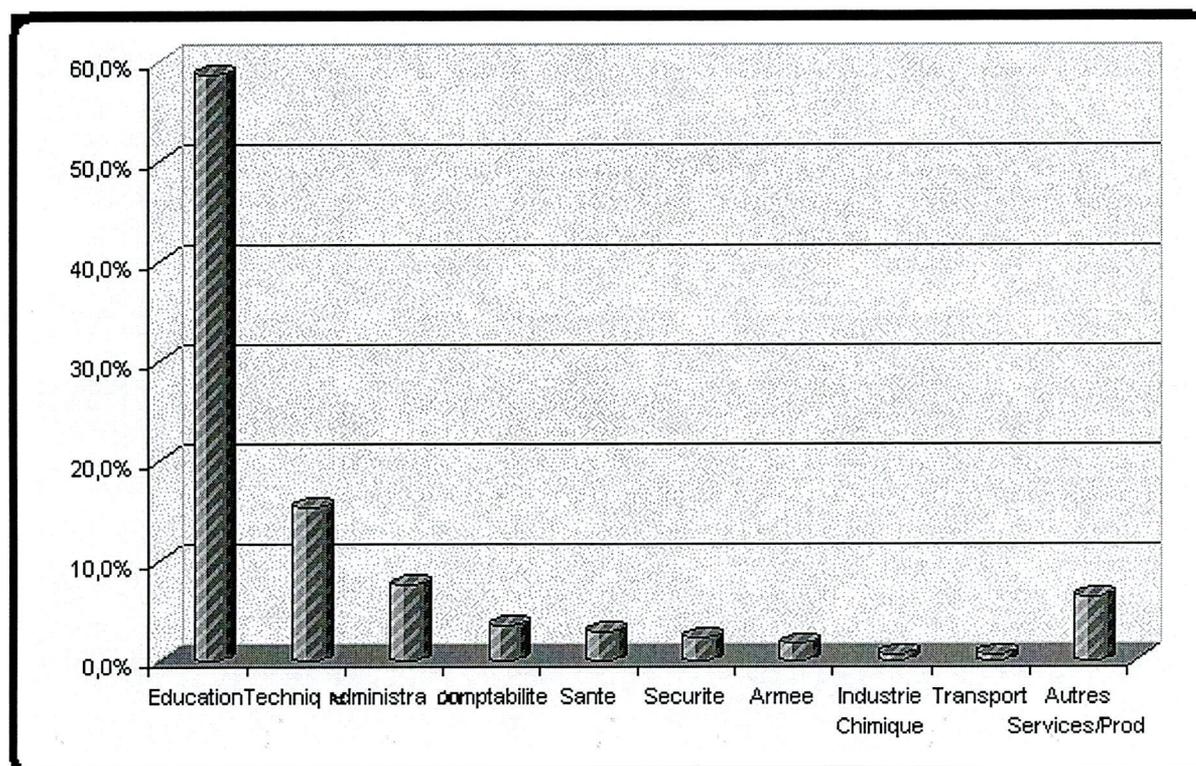
	Effectif	Fréquence
Abidjan	53	31%
Hors Abidjan	116	69%
Total	169	100%

Source : Nos calculs

2.3. Répartition de la population assurée décédée en fonction du secteur d'activité

Alors que les personnes exerçant une activité relevant du domaine de l'éducation représentent 22,8% de la population assurée, le nombre de décès parmi ces personnes représente plus de la moitié des décès intervenus, soit 58,6%. La remarque à faire est que près de neuf personnes décédées sur dix dans le domaine de l'éducation sont des instituteurs. Cela confirme l'hypothèse que la profession influence la mortalité, hypothèse que nous chercherons à vérifier dans la suite.

Graphique 2 : Répartition de l'ensemble des décès intervenus en fonction du secteur d'activité



Source : Notre étude

Il ressort de ce qui précède que plus de la moitié des décès intervenus concernent des assurés qui exerçaient leur activité dans le domaine de l'éducation. Pourtant dans la composition du portefeuille, la proportion des assurés du secteur de l'éducation est moins du tiers de l'ensemble.

Cette analyse étant faite, nous nous chargerons dans le chapitre suivant de présenter les caractéristiques des assurés susceptibles d'influencer leurs probabilités de décès.

CHAPITRE II : STATISTIQUE EXPLORATOIRE DU PORTEFEUILLE « SIKA RETRAITE »

L'analyse des données désigne l'ensemble des méthodes à partir desquelles on collecte, organise, résume, présente et étudie des données pour permettre d'en tirer des conclusions et de prendre des décisions. La collecte et l'étude des données sont importantes depuis longtemps pour bon nombre d'activités ou de professions. Les économistes aident les gouvernants à prendre des décisions sur la base des informations qu'ils traitent : taux d'inflation, taux de chômage, valeurs des monnaies, indices des prix, etc. Les médecins établissent leurs diagnostics en analysant les données de leurs patients. Les chefs d'entreprises font reposer leurs décisions sur des études de marchés et sur l'opinion des consommateurs. Et bien d'autres disciplines de base, telles que la physique, la chimie, la géologie, l'histoire, la sociologie, utilisent des faits numériques qu'on appelle *données*.

Si l'analyse des données n'est pas une science nouvelle, il n'en est pas moins vrai qu'elle a profondément changé depuis les années soixante grâce à l'évolution des ordinateurs et des technologies de l'information et de la communication. On est passé ainsi en quelques années du *calcul statistique* à la *visualisation interactive des données*.

Dans ce chapitre, nous nous sommes donnés pour tâche de faire usage de cette discipline pour expliquer au mieux les résultats du portefeuille *Sika Retraite* commercialisé par NSIA Vie. Pour passer des données à l'information et de l'information à la décision, il faut une démarche rigoureuse c'est-à-dire de la méthode. Ainsi, avant de présenter les résultats de l'analyse nous allons faire un bref aperçu sur les différentes méthodes d'analyse des données tout en faisant abstraction des formules mathématiques qui enrichissent la théorie de l'analyse exploratoire des données.

Aussi il sera question de spécifier les variables susceptibles d'influencer la probabilité de décès d'un assuré compte tenu de certaines caractéristiques socio-démographiques de l'assuré avant de procéder à la modélisation de cette probabilité dans la deuxième partie.

I. METHODES D'ANALYSE FACTORIELLE DES DONNEES²

A. Analyse factorielle : Géométrie de l'information

1. Objectif

L'objectif de l'analyse factorielle est d'élaborer et de faire figurer géométriquement dans un espace euclidien de faible dimension les informations les plus diverses consignées dans des tableaux numériques à double entrée, lesquelles, compte tenu de la taille et de la complexité de ces tableaux, ne peuvent être facilement synthétisées.

La représentation géométrique fournit aisément de l'ensemble des données numériques une image qui permet d'un seul coup d'œil, de saisir rapidement l'ensemble des éléments présentés et de mettre en évidence les faits essentiels. L'adjectif *factorielle* vient du fait qu'à partir des variables initiales de la base de données, de nouvelles variables appelées *facteurs* sont obtenues par combinaison linéaire des anciennes. Ces facteurs forment un espace euclidien de dimension plus petite que celle de l'espace formé par les variables initiales, permettant de résumer au mieux l'information.

2. Données et choix des méthodes

Sur une population statistique, des données diverses peuvent être recueillies pour étudier le comportement de cette population. On est souvent amené à définir des variables permettant de recueillir ces données. On se trouve le plus souvent en face de deux types de variables : *les variables quantitatives*³ et *les variables qualitatives*. En face de chaque type de variable, l'analyste est amené à faire usage de méthodes appropriées pour son analyse afin de livrer la meilleure information contenue dans les données. Des outils performants d'analyse factorielle des données sont développés pour passer des données à l'information selon que les variables sont quantitatives ou qualitatives.

B. Différentes méthodes d'analyse factorielle

1. Données quantitatives : ACP

1.1. Objectif

L'analyse en composante principale ou ACP a pour objectif de représenter graphiquement les relations entre les variables quantitatives (ou assimilées à des variables quantitatives) et de visualiser en même temps les individus qui sont dans la relation avec ces

² L'analyse factorielle des données fait appel à des formules mathématiques utiles pour comprendre la théorie de cette discipline. Nous recommandons à tout lecteur intéressé par cette discipline de bien vouloir se référer à la bibliographie en fin de document.

³ Une variable est dite quantitative lorsque la moyenne calculée sur cette variable a un sens. En conséquence, une nette différence doit être faite avec les variables dont les modalités sont représentées par des codes numériques. Ainsi *le capital assuré* par un client, *son âge*, *la prime versée* sont des variables quantitatives. En revanche, *son sexe* dont les modalités peuvent prendre des valeurs 1=masculin et 2=Féminin, est une variable qualitative.

variables. A partir de cette photographie géométrique l'analyste pourra interpréter les résultats en fonction de sa connaissance du sujet étudié.

L'ACP est une méthode factorielle linéaire parce qu'elle n'est efficace que si la liaison entre les variables est mesurée par le coefficient de corrélation linéaire. Elle ne prend donc pas en compte les liaisons non linéaires.

1.2. Principes d'analyse

L'ACP étudie des corrélations entre plusieurs variables quantitatives. Elle permet également de repérer une structuration des individus en groupes homogènes : qui ressemble à qui et de quel point de vue ?

Elle étudie les variables en terme de corrélation et les individus en terme de ressemblance. Cette méthode basée sur le choix des axes pertinents pour l'analyse, développe plusieurs critères. Le choix des axes se fera soit selon le *critère du coude*, soit selon le *critère de Kaiser* (Valeur propre supérieure à 1) soit selon le *critère empirique* (basé sur les pourcentages). Ainsi, si un nombre p d'axes parmi les n axes est choisi, on fait des graphiques (nuage des points individus et nuage des points variables) en croisant ces p axes deux à deux.

En analyse des données, il existe des éléments qui permettent d'interpréter avec plus de rigueur et de prudence le nuage des points individus et celui des points variables. Ces éléments sont nommés des aides à l'interprétation. Ce sont le *taux d'inertie*, le Cos^2 et la *contribution à l'inertie*.

Le *taux d'inertie* est un critère global qui permet de savoir globalement si le nuage a été plus ou moins bien projeté.

Le Cos^2 d'un point sur un axe factoriel mesure le cosinus carré de l'angle que fait ce point avec cet axe. Il sert à apprécier la qualité du positionnement des points en représentation factorielle comparée à leur configuration réelle.

Les *contributions à l'inertie* du nuage de points permettent d'identifier les points très influents pouvant déterminer à eux seuls l'orientation de certains axes. Les points les plus contributifs au positionnement d'un axe sont ceux qui ont les coordonnées les plus fortes.

2. Données qualitatives : AFC et ACM

2.1. Objectifs

L'objectif de l'analyse factorielle des correspondances (AFC) dite binaire consiste en la visualisation de données de fréquences issues du croisement de deux variables qualitatives.

Elle est une méthode d'analyse adaptée aux tableaux de contingences et permet d'étudier les éventuelles relations existant entre deux variables nominales.

Quant à l'analyse des correspondances multiples (ACM), c'est l'étude des liaisons entre plusieurs variables qualitatives. C'est une analyse des correspondances d'un tableau disjonctif complet.

2.2. Principes d'analyse

L'AFC et l'ACM sont des méthodes factorielles non linéaires. Elles déterminent les modalités de variables qui interviennent dans la liaison et aussi la manière dont elles interviennent en termes d'attraction, de répulsion et d'indépendance. Ainsi, comme en ACP, elles utilisent les mêmes outils d'aides à interprétation et les règles d'interprétations sont sensiblement les mêmes.

Compte tenu des distances entre les éléments du tableau disjonctif complet, on exprime :

- la proximité entre individus en terme de ressemblance. Deux individus se ressemblent s'ils ont choisi globalement les mêmes modalités ;
- la proximité entre modalités de variables différentes en terme d'association ;
- la proximité entre deux modalités d'une même variable en terme de ressemblance.

II. PRESENTATION DES RESULTATS DE L'ANALYSE

A. Choix de la méthode

1. Justification du choix

Il résulte des développements précédents que la méthode appropriée à utiliser dans le cas de notre analyse est l'analyse des correspondances multiples et ce en raison du caractère qualitatif des variables d'études. La variable *âge de l'assuré à la souscription* est une variable numérique mais pour l'adapter à notre analyse elle a été transformée en une variable qualitative que nous avons nommée *groupe d'âge de l'individu* qui donne le groupe d'âge auquel appartient l'assuré. En se basant sur les analyses préliminaires fournies par les logiciels utilisés nous avons choisi trois classes d'âge pour mieux discriminer les individus:

- Groupe 1=les assurés âgés de moins de 30 ans
- Groupe 2=les assurés âgés entre 31 et 45 ans
- Groupe 3=les assurés âgés de plus de 45 ans.

2. Choix des variables

2.1. Variables actives

Lorsqu'il s'agit de faire une analyse des données, quelle que soit la méthode utilisée, l'on se doit de définir les variables d'intérêts désignés sous le nom de *variables actives*. Dans notre analyse, les variables actives sont :

- le groupe d'âge auquel appartient l'assuré lors de la souscription du contrat ;
- le sexe de l'assuré ;
- le secteur d'activité de l'assuré ;
- le lieu de résidence de l'assuré.

2.2. Variables illustratives

Les variables illustratives en analyse des données ont pour intérêt d'enrichir l'analyse. Elles ne participent pas à la fixation des facteurs et permettent d'apporter des détails plus intéressants. Dans notre cas la variable illustrative est *la probabilité de décès de l'assuré avant la cinquième année du contrat*.

B. Résultats de l'analyse

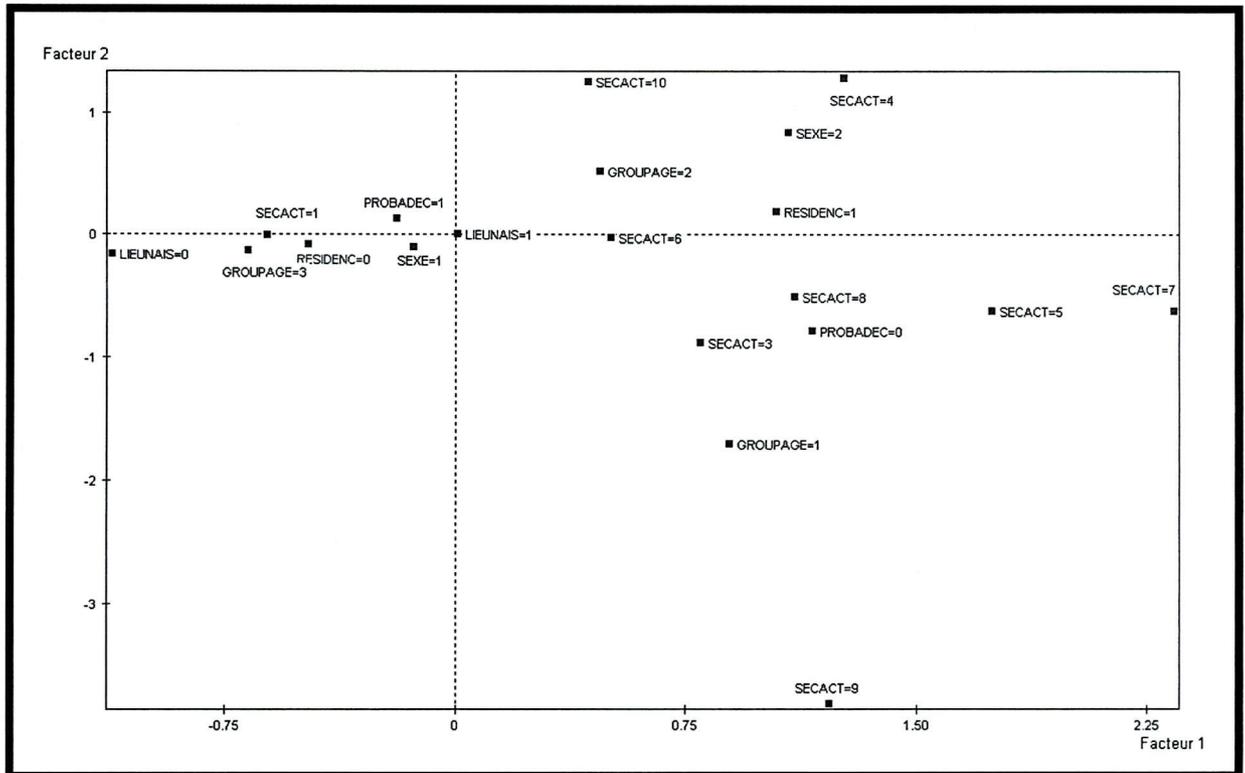
1. Nuage des points variables

Les facteurs 1, 2, 3 fournissent plus de 80 % de l'information sur la base. C'est la raison pour laquelle, nous présenterons successivement les plans factoriels (1, 2), (1, 3) et (2,3).

Le plan factoriel (1, 2) permet de constater que les assurés de sexe masculin, habitant hors d'Abidjan, exerçant dans le domaine de l'éducation et ayant un âge supérieur à 45 ans sont décédés avant la cinquième année du contrat. Par conséquent, le fait de disposer de ces caractéristiques indique qu'on a de fortes chances que le décès de l'assuré intervienne avant la cinquième année du contrat.

En revanche, les personnes âgées de moins de 30 ans et exerçant dans les domaines d'activité du transport, de l'armée et de la santé ont survécu après cinq (5) années de contrat. Ainsi les personnes ayant ces caractéristiques ont une forte probabilité de survie cinq (5) ans après la souscription du contrat.

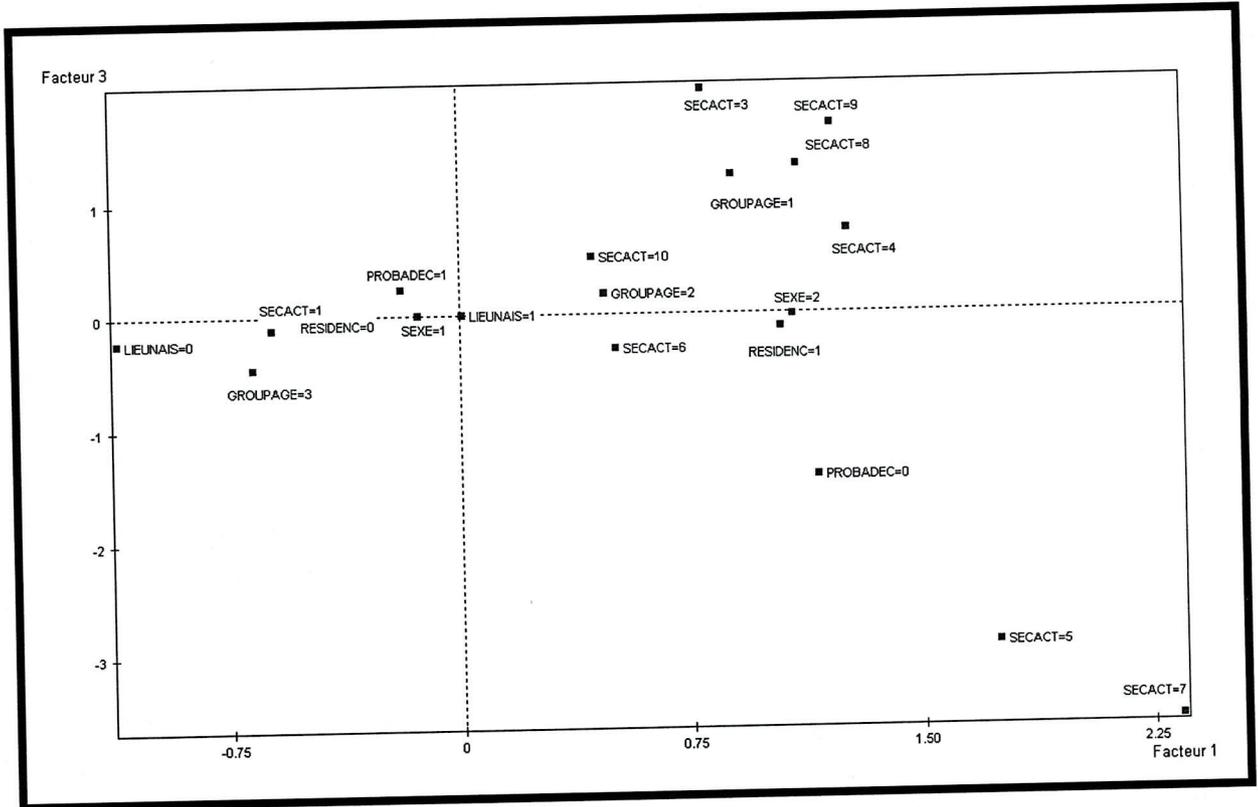
Graphique 3 : Représentation des variables dans le plan factoriel (1 , 2)



Source : Notre étude

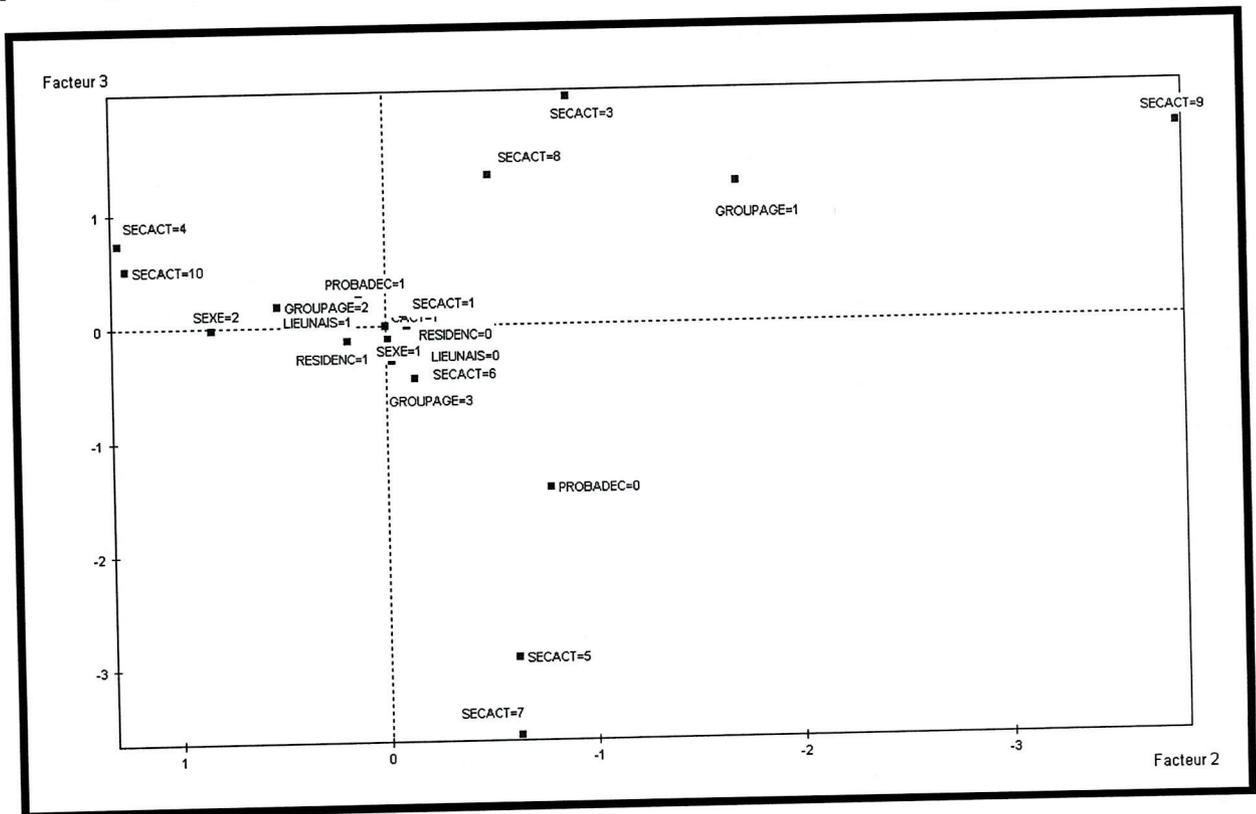
Les renseignements fournis par le plan (1 , 3) indiquent que les assurés exerçant dans les autres secteurs d'activités ont un comportement identique en matière de mortalité. Cela témoigne de l'importance qu'il faut accorder à la proposition de subdiviser la population en plusieurs catégories afin de vendre des contrats plus spécifiques. Au vu des résultats de l'analyse, on remarque que les assurés du domaine de l'éducation coûtent cher à la mutualité. Un contrat sur mesure, répondant aux besoins des assurés de l'éducation, permettrait d'ajuster la prime de cette catégorie et d'avoir une gestion plus adaptée.

Graphique 4 : Représentation des variables dans le plan factoriel (1 , 3)



Source : Notre étude

Graphique 5 : Représentation des variables dans le plan factoriel (2 , 3)



Source : Notre étude

2. Nuage des points individus

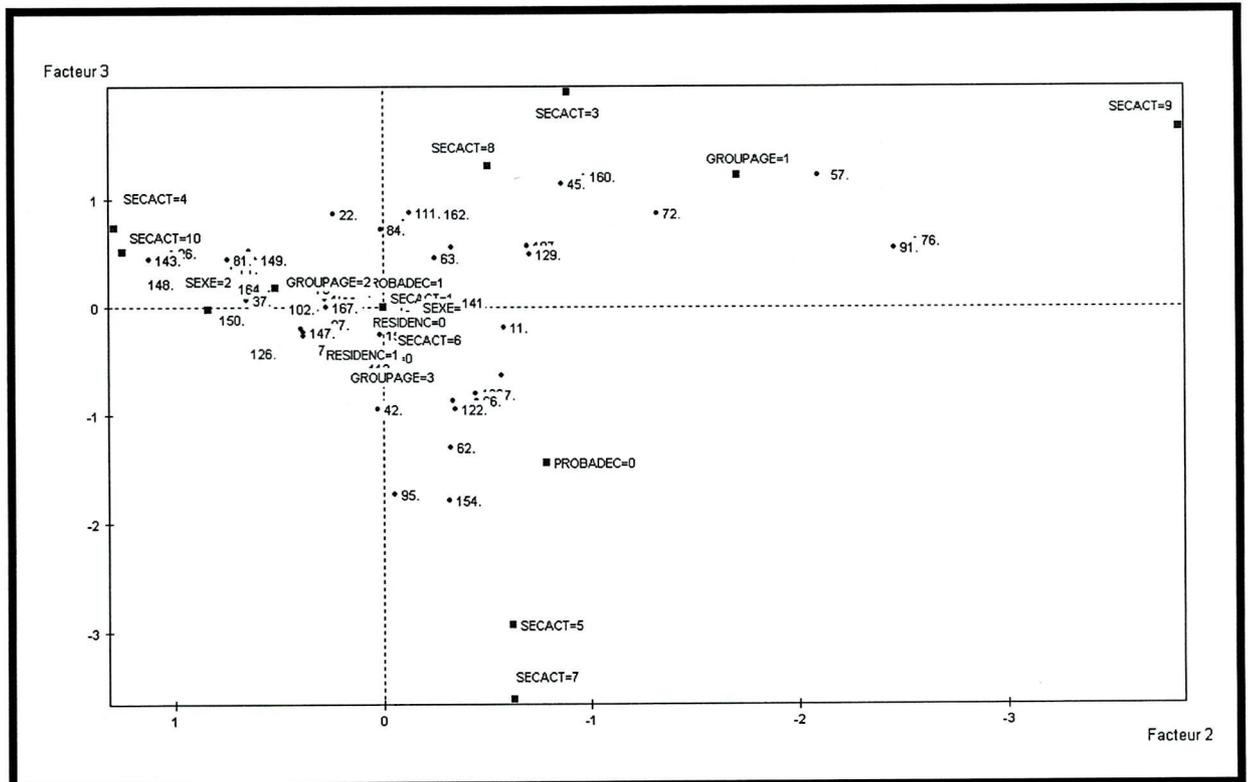
Au vu des trois représentations suivantes, on retient qu'il existe trois (3) groupes d'individus dans le portefeuille décès du contrat étudié :

- les personnes exerçant dans le domaine de l'éducation, résidant hors d'Abidjan et âgées de plus de 45 ans ont une probabilité de décès avant la cinquième année du contrat élevée;

- les personnes exerçant dans les domaines du transport, de l'armée ou de la santé et âgées de moins de 30 ans ont une probabilité de décès avant la cinquième année du contrat faible;

- les personnes exerçant dans les autres domaines d'activité, dont la probabilité de décès n'est liée à aucune des caractéristiques étudiées.

Graphique 8 : Représentation des variables et des individus dans le plan factoriel (1, 3)



Source : Notre étude

L'analyse statistique du portefeuille **Sika Retraite** a permis de structurer le portefeuille et de retenir que le fait que l'assuré exerce une certaine fonction, le fait qu'il ait un certain âge et le fait d'être de sexe masculin et non de sexe féminin, sont autant de caractéristiques qui influencent la mortalité de l'assuré et qui sont susceptibles d'augmenter la probabilité de décès de celui-ci.

Cette analyse, qui est un préalable important à la modélisation de la mortalité, nous a permis de pouvoir guider et de formuler des hypothèses de travail qui constituent la base de travail de la seconde partie.

Il sera donc question d'évaluer la probabilité de décès d'un assuré à partir des informations que sont le sexe, le groupe d'âge à la souscription, le secteur d'activité et le lieu de résidence de l'assuré. A cet effet, divers outils seront utilisés afin de s'assurer de la pertinence des résultats.

**PARTIE II : MODELISATION DE LA MORTALITE DU PORTEFEUILLE « SIKA
RETRAITE »**

CHAPITRE I: CLASSIFICATION AUTOMATIQUE DES ASSURES

On désigne par classification automatique, un ensemble de méthodes très diverses et très empiriques permettant de structurer un ensemble d'observations multidimensionnelles en sous-ensembles cohérents.

On emploie souvent la classification automatique en complément de l'analyse factorielle vue au deuxième chapitre de la première partie, et qui permet de photographier ces observations de manière efficace.

Ce chapitre nous permettra d'assurer un meilleur regroupement des assurés afin de confirmer le travail fait avec l'analyse des correspondances multiples, utile pour la modélisation dans le chapitre suivant.

I. METHODE DE CLASSIFICATION

A. Objet

1. Objectif

Les méthodes de classifications visent à construire des classes d'individus décrites par un ensemble de variables qualitatives ou quantitatives, ou des classes de variables observées sur des individus statistiques. Dans les deux cas, on met en jeu un tableau de données Individus -Variables mais l'objectif de la classification est différent selon que l'on classe les individus ou les variables.

2. Différentes méthodes de classification⁴

Pour classer les individus, plusieurs méthodes sont utilisées. Il s'agit des méthodes de partition, des méthodes de segmentation, les méthodes de discrimination et les méthodes de classification hiérarchique. Nous ferons usage dans notre analyse de la dernière méthode en raison de la manière dont nous voulons conduire et présenter notre étude.

L'édification d'une classification hiérarchique peut se faire de deux manières : pour la première, à partir des éléments, on forme des petites classes ne comportant que des individus très semblables, puis, à partir de celles-ci, on construit des classes de moins en moins homogènes, jusqu'à obtenir la classe toute entière. Ce mode de construction s'appelle la *classification ascendante hiérarchique*. Pour la seconde, on procède par dichotomies

⁴ Il existe plusieurs méthodes de classification. Le lecteur pourra se référer à la bibliographie pour davantage d'informations sur la question. L'objectif étant ici de présenter l'intérêt de la classification nous avons expressément ignoré cette présentation afin de ne pas perdre de vue qu'il ne s'agit pas d'une analyse purement statistique mais une aide à la décision.

successives à partir de l'ensemble total, jusqu'aux éléments qui composent cet ensemble ; ce mode s'appelle la *classification descendante hiérarchique*. Nous opterons dans la suite pour la classification ascendante.

B. Raisons de la classification

1. Classification des individus

Une entreprise peut arbitrairement décider de segmenter ses clients selon leurs critères sociologiques, selon leurs achats ou selon les types d'achats. Pour une compagnie d'assurance, le Directeur technique peut s'intéresser à classer les individus selon la rentabilité de leur contrat. Chaque classification des individus sera différente selon le point de vue adopté, le critère de classification, et chaque classification sera insatisfaisante parce qu'un seul critère sera arrêté pour ce faire.

L'enjeu de la classification sera de construire une partition, ou un système de classes emboîtées ou pas, telle qu'elle soit bâtie sur l'ensemble des critères jugés pertinents pour l'entreprise et non pas seulement quelques-uns pris séparément les uns des autres. Une telle classification permet d'avoir une vision multidimensionnelle des clients. Les méthodes de classification vont donc permettre à l'entreprise de construire des classes d'individus tout en gardant leurs caractères multidimensionnels. C'est la raison pour laquelle les méthodes de classification s'inscrivent dans les méthodes d'analyse des données et complètent les méthodes d'analyse factorielle, les classes jouant en classification le rôle des axes factoriels.

2. Classification des variables

Il arrive très souvent que des phénomènes de la nature soient liés les uns aux autres ; et en conséquence que les variables définies pour décrire un phénomène soient redondantes. Il peut alors être utile de classer les variables pour réduire le nombre et pour sélectionner les plus représentatives. Muni de ce nouvel ensemble de variables, on pourra alors classer les individus.

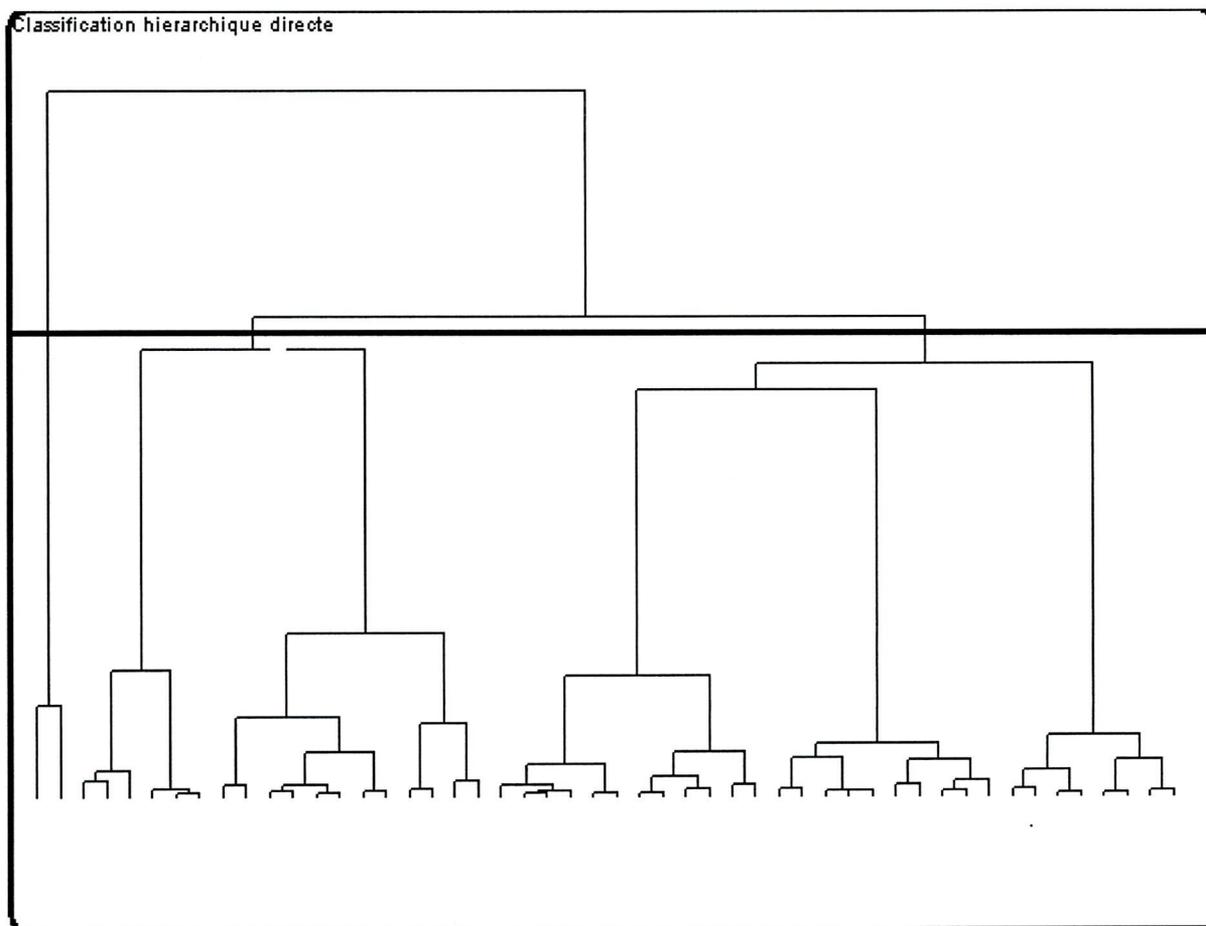
II. RESULTATS

A. Choix du nombre de classes

1. Arbre de classification

La méthode de classification est la classification ascendante dont l'arbre est présenté dans le graphique 9.

Graphique 9 : Arbre de classification des assurés du portefeuille



Source : Notre étude

2. Nombre de classes à retenir

D'après le graphique 9, les partitions les plus intéressantes à priori sont en 2 classes, 3 classes et 6 classes. Plus il y a de classes plus la partition est détaillée. Nous retiendrons la partition en 3 classes conformément à ce qui a été fourni par les résultats de l'analyse factorielle.

B. Description des classes

1. Description des partitions

Le tableau 5 fournit une présentation en 3 classes dont la première contient 46,75 % des individus, la deuxième 17,16% et la dernière 36,09%.

Tableau 5 : Caractérisation des classes retenues

CARACTERISATION PAR LES MODALITES DES CLASSES OU MODALITES									
CLASSE 1 / 3									
V. TEST PROB	POURCENTAGES			MODALITE			IDEN POIDS		
	CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES DES VARIABLES					
46.75 CLASSE 1 / 3									aa1a 79
9.86	0.0000	76.77	96.20	58.58	SECACT=1	Secteur d activité	AD_1	99	
8.03	0.0000	66.38	97.47	68.64	RESIDENC=0	Lieu de résidence	AC_1	116	
5.98	0.0000	72.54	69.62	44.97	GROUPAGE=3	Groupe d âge	AF_3	76	
5.32	0.0000	54.48	100.00	85.80	PROBADEC=1	Probabilité de décès avant cinq ans	AA_2	145	
4.58	0.0000	52.67	100.00	88.76	SEXE=1	Sexe de l assuré	AB_1	150	
-4,40	0.0000	27.63	26.58	44.97	GROUPAGE=2	Groupe d âge	AF_2	76	
-4,41	0.0000	7.69	2.53	15.38	SECACT=6	Secteur d activité	AD_5	26	
-4,58	0.0000	0.00	0.00	11.24	SEXE=2	Sexe de l assuré	AB_2	19	
CLASSE 2 / 3									
V. TEST PROB	POURCENTAGES			MODALITE			IDEN POIDS		
	CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES DES VARIABLES					
17.16 CLASSE 2 / 3									aa2a 29
7.87	0.0000	83.33	68.97	14.20	PROBADEC=0	Probabilité de décès avant cinq ans	AA_1	24	
3.70	0.0000	100.00	17.24	2.96	SECACT=5	Secteur d activité	AD_4	5	
3.70	0.0000	100.00	17.24	2.96	SECACT=9	Secteur d activité	AD_8	5	
3.37	0.0000	52.94	31.03	10.06	GROUPAGE=1	Groupe d âge	AF_1	17	
-3,09	0.0001	9.09	31.03	58.58	SECACT=1	Secteur d activité	AD_1	99	
-7,87	0.0000	6.21	31.03	85.80	PROBADEC=1	Probabilité de décès avant cinq ans	AA_2	145	
CLASSE 3 / 3									
V. TEST PROB	POURCENTAGES			MODALITE			IDEN POIDS		
	CLA/MOD	MOD/CLA	GLOBAL	CARACTERISTIQUES DES VARIABLES					
36.09 CLASSE 3 / 3									aa3a 61
6.33	0.0000	71.70	62.30	31.36	RESIDENC=1	Lieu de résidence	AC_2	53	
6.24	0.0000	61.84	77.05	44.97	GROUPAGE=2	Groupe d âge	AF_2	76	
4.82	0.0000	100.00	21.31	7.69	SECACT=4	Secteur d activité	AD_3	13	
3.30	0.0000	73.68	22.95	11.24	SEXE=2	Sexe de l assuré	AB_2	19	
2.90	0.0002	100.00	9.84	3.55	SECACT=8	Secteur d activité	AD_7	6	
2.89	0.0002	81.82	14.75	6.51	SECACT=10	Secteur d activité	AD_9	11	
2.67	0.0004	61.54	26.23	15.38	SECACT=6	Secteur d activité	AD_5	26	
-3,30	0.0000	31.33	77.05	88.76	SEXE=1	Sexe de l assuré	AB_1	150	

Source : Notre étude

2. Caractéristiques des classes⁵

Il serait encore plus intéressant que la partition soit faite en un nombre de classes plus élevé afin d'avoir une meilleure discrimination. Mais compte tenu du faible nombre de variables dépendantes nous avons choisi parmi les partitions les plus intéressantes, celles en trois classes.

Ainsi, au regard du tableau 5, la classe 1 est celle des enseignants résidant dans les villes de l'intérieur de la Côte d'Ivoire. Dans ce groupe, ce qui est pertinent à retenir, est que 54,48% de l'ensemble des décès intervenus avant la cinquième année du contrat sont des enseignants. Aussi, parmi les personnes âgées de plus de 45 ans, 72,54% d'entre eux sont dans ce groupe. Aucun assuré de sexe féminin ne se retrouve dans ce groupe. De plus, les personnes de ce groupe sont tous décédés avant le terme du contrat c'est-à-dire avant la cinquième année suivant la souscription du contrat.

La deuxième classe quant à elle est celle de l'armée et de la santé. Tous les assurés de ces secteurs d'activité se retrouvent tous dans cette classe. 52,94% des assurés de moins de 30 ans se retrouvent dans cette classe et 83,33% des assurés dont le décès n'est pas intervenu avant la date terme du contrat sont de ce groupe. Ce groupe s'oppose donc au premier groupe qui se caractérise par une probabilité de décès élevée.

Le troisième groupe est celui de la technique et de l'administration générale. Tous les assurés de ces secteurs se retrouvent également dans la classe. Ce sont des personnes âgées de plus de 30 ans et de moins de 45 ans. Ce groupe a la particularité de présenter une proportion égale des assurés dont le décès est intervenu avant le terme du contrat et ceux dont le décès n'est pas intervenu. D'autres variables non incluses dans notre étude pourront permettre de mieux les discriminer.

A l'issue de la classification faite sur les assurés, nous retenons 3 grands groupes. Ce sont :

- le groupe des enseignants caractérisés par une forte mortalité ;
- le groupe de l'armée et de la santé caractérisé par une faible mortalité ;

⁵ Dans le tableau 5 la lecture est la suivante :

CLA/MOD = Nombre d'individus possédant la modalité dans la classe / nombre total d'individus possédant la modalité.

= 100% indique que tous les individus possédant cette modalité sont dans la classe. La connaissance de cette seule modalité suffit donc au classement.

MOD/CLA = Nombre d'individus possédant la modalité dans la classe / effectif de la classe.

= 100% indique que tous les individus de la classe possèdent la modalité.

- le groupe des techniques et du personnel de l'administration générale caractérisé par une mortalité moyenne.

La classification a enrichi les classes obtenues au moyen de l'analyse factorielle. Dans le prochain chapitre, cette analyse préliminaire aidera à mieux conduire la modélisation et aussi à prendre quelques précautions pour conduire à bien cette méthodologie.

CHAPITRE II : MODELISATION DE LA PROBABILITE DE DECES

Pour mener à bien une étude, il est souvent utile de se munir d'assez d'outils afin de contrôler la véracité et la cohérence des résultats. Dans ce chapitre, notre objectif est de donner des pronostics sur les assurés lorsqu'ils entrent nouvellement dans le portefeuille.

Nous allons pour cela employer comme méthode préliminaire, celle du Scoring afin de se fixer une idée sur la régression logistique qui nous permettra d'enrichir davantage le processus de modélisation.

I. MODELISATION DECISIONNELLE

A. Analyse discriminante

1. Présentation

Considérons une population d'individus notée I , partitionné a priori en k classes et décrites par p variables quantitatives. Ce partitionnement peut être établi à partir d'une variable qualitative en k classes. Les méthodes de discrimination ont pour objet d'expliquer la variable qualitative qui a fourni la partition au moyen des variables explicatives, donc d'étudier en quoi se différencient les classes de la partition.

Deux objectifs fondamentaux sont recherchés quand une analyse discriminante est faite.

2. Objectifs

2.1. Objectif descriptif

L'analyse discriminante vise à chercher les combinaisons linéaires des variables qui permettent de séparer au mieux les k catégories d'individus, et de donner une représentation graphique adéquate qui rend compte de cette séparation.

2.2. Objectif prévisionnel

L'objectif fondamental de l'analyse discriminante est de prévoir la classe d'affectation d'un nouvel individu décrit par un ensemble de variables. L'assureur peut se donner la tâche de prédire le comportement d'un nouvel assuré. C'est d'ailleurs cet objectif que nous recherchons dans cette partie. Les résultats de l'analyse du portefeuille en question dans la partie suivante seront fournis de façon méthodique afin de permettre à tout lecteur non initié à l'analyse discriminante, de comprendre le processus de décision.

B. Résultats de l'analyse

1. Tableau d'affectation des individus

SPAD est un logiciel d'analyse des données dont nous avons fait usage pour notre analyse. Les informations que nous tirons à partir des analyses faites sous ce logiciel, nous indiquent que 75% des individus sont bien classés et 25% sont mal classés.

En effet, au moyen de notre base d'analyse, les individus ont été classés en deux groupes :

- Ceux dont le décès est intervenu avant la cinquième année du contrat (date de fin du contrat) que nous avons choisis de qualifier « **haut risque** » ;
- Ceux dont le décès n'intervient pas avant la cinquième année du contrat (date de fin du contrat), que nous avons choisis de qualifier « **sans risque** ».

Le logiciel, au vu des observations, affecte chaque individu dans l'une des classes spécifiées. Et nous constatons que dans la classe des individus que nous avons qualifiés « **haut risque** » 115 sont classés dans ce groupe par le logiciel. Et dans le groupe des assurés « **sans risque** », 12 sont classés par le logiciel dans ce groupe. Soit un total de 127 assurés bien classés; ce qui représente une proportion de 75% de l'ensemble des assurés.

En revanche, le logiciel affecte 12 des assurés « **sans risque** » dans le groupe des assurés « **haut risque** » et 30 des assurés « **haut risque** » dans le groupe des assurés « **sans risque** ». Soit un total de 42 assurés mal classés ; soit une proportion de 25% de l'ensemble des assurés.

Tableau 6 : Classification des individus sous SPAD.

		Bien classés	Mal classés	Total
Groupe d'origine	AA_1	12 50%	12 50%	24 100%
	AA_2	115 79%	30 21%	145 100%
	Total	127 75%	42 25%	169 100%

Source : Notre étude

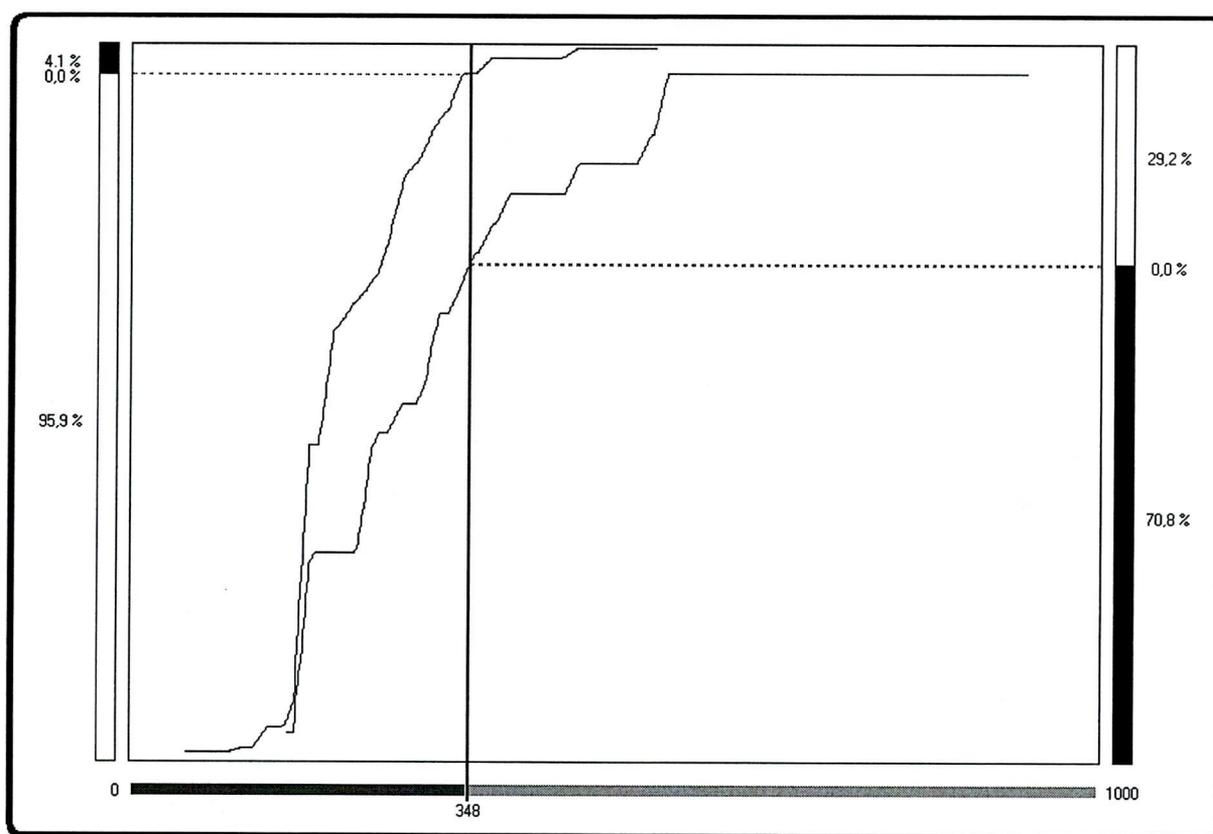
Les résultats montrent qu'il y a une erreur qu'on commet en classant les assurés. Notre objectif étant d'arriver à la meilleure discrimination, nous allons nous servir de la fonction score pour diminuer le risque d'affectation des individus.

2. Fonction score⁶

L'affectation d'un individu dans une classe précise est faite au moyen de ce qu'a été défini et appelé fonction Score. Le principe est qu'une distance est calculée pour chaque individu par rapport à l'individu moyen de chaque sous groupe. L'individu est par conséquent affecté au groupe pour lequel la distance calculée est la plus faible.

Le graphique 10 correspond au tracé de la fonction score. Une mauvaise classification des assurés dont le décès est intervenu avant le terme du contrat signifie que le pronostic à faire sur un nouvel assuré sera erroné. Ainsi nous avons fixé le seuil minimum de score à 348 et nous obtenons 95,9% des assurés à « haut risque » et seulement 4,1% d'entre eux sont mal classés. A ce seuil nous faisons un taux de bon classement des assurés « sans risque » de 29,2% et un taux de non classement de 70,8%. Ceci ne semble pas être mauvais dans la mesure où plus de 85% des assurés sont décédés avant la cinquième année du contrat.

Graphique 10 : Graphe de la fonction score



Source. Notre étude

La règle de décision avec cette classification au moyen de la fonction score, pour un nouvel assuré est la suivante :

⁶ La théorie sur la fonction score est très enrichissante mais avec des calculs plus complexes. Pour des raisons de simplicité et de clarté nous avons expressément écarté de cette partie les développements théoriques.

- Si son score est inférieur ou égal à 348, il est très probable que l'assuré ne soit pas vivant cinq (5) ans après la souscription du contrat ;
- Si son score est supérieur à 348, on a de bonnes raisons de penser qu'il ne décède pas avant la fin du contrat.

Cette méthode constitue un outil de modélisation moins satisfaisante et c'est pour cette raison que nous introduisons la régression pour davantage d'informations sur la modélisation de la mortalité de notre portefeuille.

II. REGRESSION LOGISTIQUE⁷

Il s'agit ici d'une application de la régression logistique de sorte que nous n'insisterons pas sur les détails mathématiques, mais plutôt sur les concepts fondamentaux. Néanmoins, la statistique est avant tout une discipline faisant appel aux mathématiques et même si le programme statistique prend en charge tous les aspects formels, un minimum de formalisme est nécessaire pour bien illustrer les concepts.

Rappelons que dans ce chapitre notre but est de proposer le modèle qui s'ajustera le mieux aux données observées ; ceci dans l'intention de mettre en exergue les facteurs décrivant de façon significative la mortalité. Le but final étant de dire au moyen de ses caractéristiques si le nouveau client est à risque ou non.

A. Modèle dichotomique

1. Justification du modèle

L'analyse des comportements individuels ne repose pas toujours sur des variables continues comme le revenu, la consommation, l'investissement... ; elle peut porter sur des pratiques à caractères qualitatifs comme, par exemple, le fait de consommer un certain bien, le fait d'exercer une activité parmi un ensemble, le fait d'adhérer volontairement à une convention d'assurance, le fait de décéder avant une certaine date.

Dans ces cas, la méthode traditionnelle des Moindres Carrées Ordinaires ne semble plus adaptée car elle doit tenir compte de l'absence de continuité de la variable endogène et souvent de l'absence d'un ordre naturel entre les modalités de cette variable. En général, on a recours à deux types de modèles selon l'hypothèse faite sur la distribution des erreurs : le

⁷ Comme tout modèle de régression, le modèle de régression logistique s'applique après vérification de certains critères assurant sa robustesse. On peut entre autre citer les problèmes de sur-dispersion, de sous-dispersion, de colinéarités entre variables explicatives, de liaison entre co-variables et variable réponse. Notre objectif n'étant pas de faire la théorie de la régression, nous invitons le lecteur désireux d'approfondir ses connaissances sur cette méthode à bien vouloir se référer à la bibliographie.

modèle Logit et le modèle Probit. Dans la suite, nous choisirons le modèle Logit en faisant l'hypothèse que les erreurs ont une distribution logistique.

2. Formalisation du modèle

La variable expliquée Y prend deux états. Dans notre analyse il s'agit d'étudier la mortalité des assurés du contrat **Sika retraite**. La variable S s'énonce de manière suivante :

$S_i = 1$ si l'assuré i est décédé au cours des cinq (5) ans de contrat⁸

$S_i = 0$ sinon

On dispose de cinq (5) variables exogènes :

-Age de l'assuré à la souscription du contrat (*AGESOUS*)

-Lieu de naissance de l'assuré (*LIEUNAIS*)

-Sexe de l'assuré (*SEXE*)

-Lieu de résidence de l'assuré (*LIEURES*)

-Secteur d'activité de l'assuré (*SECACTI*)

On cherche à savoir quelles sont les variables qui interviennent de façon significative dans le pronostic final des assurés à propos de la survenance du décès avant le terme du contrat et à évaluer l'impact de chacune de ces variables sur la probabilité de décès avant les cinq ans de contrat.

S_i tel que défini est une variable de Bernoulli avec la probabilité de succès p de décès avant la cinquième année du contrat, et la probabilité d'échec $1-p$, être en vie cinq ans après la naissance du contrat.

Il s'agit d'évaluer la probabilité p sachant l'âge, le sexe, la profession, le lieu de naissance et le lieu de résidence de l'assuré en utilisant le modèle Logit. Toutes ces variables explicatives ne rentreront pas dans le modèle. Il est question de retenir le meilleur modèle afin d'évaluer la probabilité en question.

$$\text{Prob}(S_i=1/\text{agesous}, \text{sexe}, \text{secacti}, \text{lieures}, \text{lieunais}) = \frac{e^\theta}{1+e^\theta}$$

Où θ est l'expression du modèle estimé et e^θ désigne la fonction exponentielle de θ .

⁸ Nous rappelons que la garantie décès est souscrite pour cinq ans renouvelables. En prenant par exemple comme référence « le décès intervient la première, la deuxième, la troisième ou la quatrième année de contrat » on obtient des résultats dont la discrimination n'est pas bonne.

B. Résultats du modèle**1. Interprétation des résultats****1.1. Significativité des coefficients et qualité de l'ajustement**

STATA⁹ est le logiciel utilisé pour la modélisation de la mortalité. Plusieurs modèles ont été faits en incluant toutes les variables d'intérêt, mais celui qui semble le plus pertinent est celui présenté dans le tableau suivant :

Tableau 7 : Estimation du modèle

probadec	Odds Ratio	Std. Err.	z	P>z	[95% Conf. Interval]
groupage	0.7213027	0.2815233	-0.84	0.003	0.3356565 1.55003
sexe	1.32839	0.9395491	0.40	0.088	0.332112 5.313331
residenc	0.2258214	0.1150703	-2.92	0.003	0.083181 0.6130646
secact	0.9149745	0.0706663	-1.15	0.050	0.7864444 1.06451
Log likelihood=		-62.251895		Pseudo R² = 0.7985	

Source : Notre étude

Les résultats nous montrent que le décès d'un assuré avant la date terme du contrat est bien expliqué par son sexe, le groupe d'âge auquel il appartient, le lieu de résidence et le secteur d'activité auquel il appartient. Les résultats sont significatifs au seuil d'erreur de 10%. En d'autres termes, étant donné, un individu décédé de par ses caractéristiques, on a 90% de chance que le modèle fasse un pronostic exact sur l'individu. Le lieu de naissance n'apporte pas une grande information dans le décès d'un individu avant la fin du contrat.

Ainsi l'expression analytique de θ dans la formule précédente est :

$$\theta = (0,72130)1_{\text{groupage}} + (1,32839)1_{\text{sexe}} + (0,22582)1_{\text{residenc}} + (0,914974)1_{\text{secact}}$$

Où

$1_{\text{groupage}} = 1$ si l'assuré est âgé de plus de 45 ans, 0 sinon

$1_{\text{sexe}} = 1$ si l'assuré est de sexe masculin, 0 sinon

$1_{\text{residenc}} = 1$ si l'assuré réside à Abidjan, 0 si hors Abidjan

$1_{\text{secact}} = 1$ si l'assuré est du secteur de l'éducation, 0 sinon.

L'évaluation de la probabilité p de décès avant la fin du contrat est donc :

$$Pr ob (Si = 1 / \text{agesous}, \text{sexe}, \text{sec acti}, \text{lieures}) = p = \frac{e^{\theta}}{1 + e^{\theta}}$$

⁹ Nous exposerons en annexe les commandes utilisées afin de permettre à tout utilisateur désireux de connaître la régression logistique de mieux s'orienter.

$$\text{Et } 1 - \text{Prob}(Si=1/agesous, sexe, sec acti, lieures) = 1 - p = \frac{1}{1 + e^{\theta}}$$

Ainsi on a $\frac{p}{1-p} = e^{\theta}$ et on détermine p connaissant la valeur de θ .

1.2. Evaluation du pouvoir prédictif du modèle

On utilise le modèle logistique pour modéliser la probabilité des attributs 0/1 de la variable dépendante y en fonction des co-variables x_1, x_2, \dots, x_p . A partir des probabilités estimées on décidera en fixant un seuil S_0 , par exemple à 0,5, de classer l'individu dans la catégorie $y = 1$ si sa probabilité de décès est supérieure au seuil et dans la catégorie $y = 0$ sinon. Il s'agit d'une règle de classement :

- Si la probabilité estimée est supérieure à un seuil S_0 fixé, on classe le client comme « **haut risque** » c'est à dire que le décès intervient avant la fin du contrat;

- Si au contraire la probabilité est inférieure ou égale à S_0 , le client est classé « **sans risque** » c'est à dire que le décès n'intervient pas avant la fin du contrat.

Il est intéressant de déterminer la performance du classement et la manière dont celui-ci dépend du seuil (ou de la règle) choisi. Pour cela nous allons illustrer les notions de *sensibilité* et de *spécificité*.

La *sensibilité* est définie comme la probabilité de classer l'individu dans la catégorie $y=1$ étant donné qu'il est effectivement observé dans celle-ci :

$$\text{Sensibilité} = \text{Probabilité (le client est classé « haut risque sachant que probadec=1)}$$

La *spécificité* est définie comme la probabilité de classer l'individu dans la catégorie $y=0$ étant donné qu'il est effectivement observé dans celle-ci :

$$\text{Spécificité} = \text{Probabilité (le client est classé « sans risque » sachant que probadec=0)}$$

La qualité de la méthode de classification est généralement mesurée par ces deux indicateurs (sensibilité et spécificité) au moyen de la courbe ROC (Receiver Operating Characteristic curve) qui est la courbe représentative de la sensibilité en fonction de la spécificité. Ainsi, l'aire au-dessous de la courbe ROC nous permet de mesurer globalement la capacité du modèle à affecter correctement les sujets à leurs classes respectives

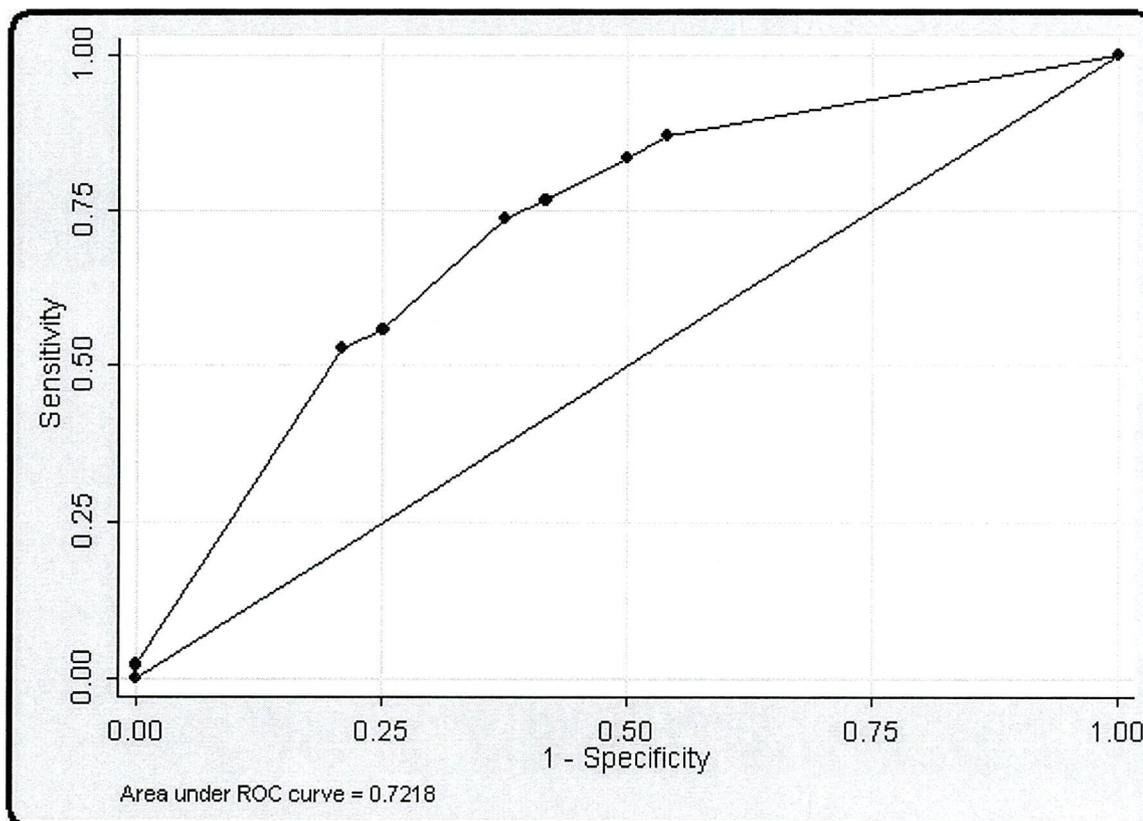
On retiendra comme règle de décision, la règle suivante :

- Si aire ROC = 0,5 il n'y a pas de discrimination ;

- Si aire $0,7 \leq \text{ROC} < 0,8$ la discrimination est acceptable ;
- Si aire $0,8 \leq \text{ROC} < 0,9$ la discrimination est excellente ;
- Si aire $\text{ROC} \geq 0,9$ la discrimination est exceptionnelle ;

Le graphique 11 nous donne la courbe ROC de notre modèle. Avec une valeur ROC = 0,7218. Ce qui signifie que la discrimination opérée entre les individus est acceptable au seuil de $S_0=5\%$.

Graphique 11 : Courbe ROC du modèle



Source : Notre étude

2. Prédiction

2.1. Erreur de prédiction.

Tableau 8 : Pouvoir de prédiction du modèle

Dependent Variable: PROBADEC						
Method: ML - Binary Logit (Quadratic hill climbing)						
Date: 08/31/06 Time: 20:03						
Sample: 1 169						
Included observations: 169						
Prediction Evaluation (success cutoff C = 0.86)						
	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	15	41	56	24	145	169
P(Dep=1)>C	9	104	113	0	0	0
Total	24	145	169	24	145	169
Correct	15	104	119	24	0	24
% Correct	62.50	71.72	70.41	100.00	0.00	14.20
% Incorrect	37.50	28.28	29.59	0.00	100.00	85.80
Total Gain*	-37.50	71.72	56.21			
Percent Gain**	NA	71.72	65.52			
	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	5.50	18.50	24.00	3.41	20.59	24.00
E(# of Dep=1)	18.50	126.50	145.00	20.59	124.41	145.00
Total	24.00	145.00	169.00	24.00	145.00	169.00
Correct	5.50	126.50	132.00	3.41	124.41	127.82
% Correct	22.91	87.24	78.11	14.20	85.80	75.63
% Incorrect	77.09	12.76	21.89	85.80	14.20	24.37
Total Gain*	8.71	1.44	2.47			
Percent Gain**	10.16	10.16	10.16			
*Change in "% Correct" from default (constant probability) specification						
**Percent of incorrect (default) prediction corrected by equation						

Source : Notre étude

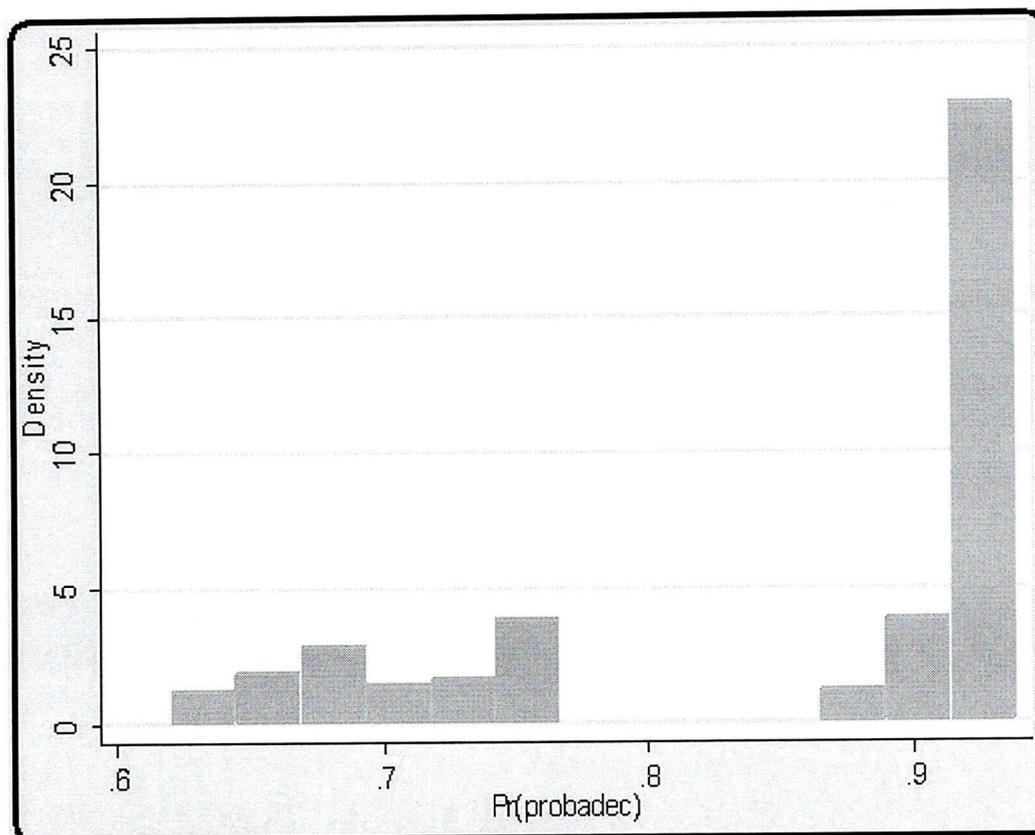
La table de classification donne les classes réelles (colonne Dep=0 et Dep=1) et les classes d'affectation (ligne P(Dep=0)<=C et P(Dep=1)>C. Le Seuil C est égal à 0,86. On lit dans la ligne Total que l'échantillon se répartit entre 24 individus ayant la valeur 0 pour la variable dépendante - en d'autres termes il y a 24 assurés dont le décès est intervenu après la cinquième année de contrat c'est à dire après le premier renouvellement du contrat-et 145 individus ayant la valeur 1-les décès avant la cinquième année de contrat.

La ligne Correct indique que 15 des 24 décès intervenus après la cinquième année du contrat ont bien été classés par la règle, ce qui fait un pourcentage de 62,50%. De même, on lit que 104 des 145 décès intervenus avant la cinquième année du contrat sont bien classés. Ce qui donne un pourcentage de 71,72%. La règle classe mieux les assurés « haut risque » que les assurés « sans risque ». Un total de 119 sur 169 des assurés décédés ont donc été bien

classés soit un pourcentage de 70,41% au seuil de 0,86, soit un taux d'erreur de prédiction de 29,59%. Cette valeur est très grande si on veut éviter les situations dans lesquelles l'assuré ne peut être classé. Ainsi les probabilités estimées et leur représentation nous permettrons de fixer un seuil où l'erreur de prédiction sera acceptable.

2.2. Choix de la probabilité seuil et Applications du modèle

Graphique 12 : Histogramme des probabilités estimées



Source : Notre étude

L'histogramme du graphique 12 ci-dessus montre qu'une forte proportion des individus de notre base, ont des probabilités estimées de décès avant la cinquième année du contrat de plus de 0,9. Mais la question qui se pose est celle de savoir à partir de quelle probabilité estimée peut-t-on classer un assuré « *haut risque* ». Ce seuil S_0 ne se fixe pas de façon arbitraire mais en construisant un test d'hypothèses statistiques dont la formulation est la suivante :

(H_0) : Le client n'est pas « à haut risque »¹⁰

Contre

(H_1) : Le client est « à haut risque »

¹⁰ Dans la théorie des tests, H_0 est l'hypothèse dont l'erreur de première espèce est la plus grave.

Fixons comme risque de première espèce du test $\alpha \in]0,1[$; c'est à dire que la probabilité pour le modèle de prédire à tort les « sans risque » vaut α .

Dans cette étude nous modélisons la probabilité d'être client à haut risque. De ce fait, nous pouvons prendre comme statistique de test, les probabilités estimées par le modèle. Soit Z cette statistique de test ; et z une réalisation de la variable aléatoire Z .

Notre test d'hypothèses (H_0) contre (H_1) se formule comme suit :

Si $z < S_\alpha$, on accepte (H_0)

Si $z \geq S_\alpha$, on rejette (H_0)

Avec S_α choisi de sorte que la probabilité pour le modèle de mal prédire le client « sans risque » c'est-à-dire. $\Pr_{H_0}(Z \geq S_\alpha) = \alpha$

D'où $F_Z^0(S_\alpha) = 1 - \alpha$, F_Z^0 étant la fonction de répartition de la variable aléatoire Z sous l'hypothèse (H_0).

La loi de Z est inconnue et de ce fait on se propose de déterminer le seuil S_α de façon empirique. Pour cela on se servira des probabilités estimées dans chaque sous population.

Tableau 9 : Fréquences des Probabilités estimées par le modèle sous (H_0)

Probabilités estimées sous H_0	Fréquences (%)	Fréquences cumulées (%)
0,62	4,2	4,2
0,65	4,2	8,3
0,66	4,2	12,5
0,67	8,3	20,8
0,68	12,5	33,3
0,71	12,5	45,8
0,72	4,2	50,0
0,73	4,2	54,2
0,75	4,2	58,3
0,77	4,2	62,5
0,90	8,3	70,8
0,92	8,3	79,2
0,93	20,8	100
Total	100	

Source : Nos calculs

Si on se fixe un seuil d'erreur α de 10%, on trouve comme valeur estimative de S_α de 0,93.

Pour un risque de première espèce α fixé, on définit la puissance π_α du test associé par : $\pi_\alpha = 1 - \beta$ où β est la probabilité pour le modèle de faire une mauvaise prédiction pour le

client « sans risque » ; c'est à dire $\beta = \Pr_{H_1}(Z < S_\alpha)$ β est encore appelé risque de deuxième espèce.

De ce fait nous pouvons définir la puissance π_α du test comme étant la probabilité pour le modèle de bien prédire les assurés à « haut risque ». On se sert donc des informations contenues dans le tableau 10. On évalue la valeur de β à 45,5% et donc $\pi_\alpha = 54,5\%$. Il convient de remarquer que la puissance n'est pas suffisamment satisfaisante compte tenu du fait que dans la population des « *haut risque* », plusieurs d'entre eux ont une probabilité estimée exactement égale au seuil S_α de 0,93.

Tableau 10 : Fréquences des Probabilités estimées par le modèle sous (H_1)

Probabilités estimées sous H_1	Fréquences (%)	Fréquences cumulées (%)
0,63	1,4	1,4
0,64	2,1	3,4
0,66	2,8	6,2
0,67	1,4	7,6
0,68	2,1	9,7
0,69	2,8	12,4
0,70	0,7	13,1
0,72	1,4	14,5
0,73	2,8	17,2
0,74	0,7	17,9
0,75	4,8	22,8
0,76	3,4	26,2
0,88	2,8	29,0
0,89	2,1	31,0
0,90	2,8	33,8
0,91	6,2	40,0
0,92	5,5	45,5
0,93	51,7	97,2
0,94	2,8	100
Total	100	

Source : Nos calculs

La règle de décision lorsque nous sommes en face d'un nouvel assuré est la suivante :

Si $z < 0,93$; on a de bonnes raisons de penser que le client n'est pas à « haut risque » ;

Si $z \geq 0,93$; on a de bonnes raisons de penser que le client est à « haut risque ».

z étant une réalisation de la variable aléatoire Z représentant les probabilités estimées par le modèle.

Ainsi nous présentons dans le tableau 11 les pronostics des assurés nouveaux entrant dans le portefeuille.

Agessous : Age de l'assuré à la souscription du contrat

Sexe : Sexe de l'assuré

Résidence : Lieu de résidence de l'assuré

Sectact : Secteur d'activité de l'assuré

Proba : Probabilité de décès estimée par le modèle

Pronostic : Pronostic fait sur l'assuré sur la durée des cinq (5) ans de contrats.

Tableau 11 : Pronostics sur les assurés à la souscription du contrat

Agesous	Sexe	Résidence	Secact	Proba	Pronostic
1	1	1	1	0,96	DECES
1	1	1	0	0,91	SURVIE
1	1	0	1	0,95	DECES
1	1	0	0	0,89	SURVIE
1	0	1	1	0,87	SURVIE
1	0	1	0	0,72	SURVIE
1	0	0	1	0,84	SURVIE
1	0	0	0	0,67	SURVIE
0	1	1	1	0,92	SURVIE
0	1	1	0	0,83	SURVIE
0	1	0	1	0,90	SURVIE
0	1	0	0	0,79	SURVIE
0	0	1	1	0,76	SURVIE
0	0	1	0	0,56	SURVIE
0	0	0	1	0,71	SURVIE
0	0	0	0	0,50	SURVIE

Source : Nos calculs

Remarque :

-*Agesous* =1 si l'assuré est âgé de plus de 45 ans, 0 sinon ;

-*Sexe*=1 si l'assuré est de sexe masculin, 0 sinon ;

-*Résidence* =1 si l'assuré réside à Abidjan, 0 si hors Abidjan ;

-*Secact* =1 si l'assuré est du secteur de l'éducation, 0 sinon.

CONCLUSION

L'objectif recherché à travers cette analyse était de déterminer les caractéristiques des assurés du contrat **Sika Retraite** commercialisé par NSIA Vie qui influent sur la mortalité. L'étude a été orientée dans le but de classer les clients en deux classes : les uns qualifiés « haut risque » et les autres « sans risque ». Les clients à « haut risque » désignent les assurés dont le décès intervient avant la cinquième année du contrat. Il était donc question de donner le profil des assurés « haut risque », ce qui devait permettre de prendre toutes les précautions lors de la tarification lorsque la Compagnie se trouve en face d'un nouvel assuré.

Pour conduire avec succès cette opération, il a fallu procéder avec méthode. C'est pourquoi nous avons procédé à une analyse statistique du portefeuille, toute chose qui a permis d'orienter l'étude et de pouvoir définir les variables dépendantes qui influencent la mortalité des assurés. A l'issue de cette étude statistique, il ressort que la mortalité des assurés du portefeuille est influencée par les variables telles que : *l'âge à la souscription, le sexe, le secteur d'activité et le lieu de résidence de l'assuré*. Le monde de l'éducation en Côte d'Ivoire conformément aux analyses faites, semble le plus touché par la mortalité. En particulier, les enseignants de l'éducation primaire sont ceux qui ont enregistré le plus de décès.

La procédure méthodique adoptée devant conduire à la modélisation de la mortalité, il a donc été question au terme de notre analyse d'évaluer la probabilité de décès d'un assuré dans les cinq (5) ans suivant la souscription du contrat. Ainsi, tout assuré dont la probabilité de décès est supérieure ou égale à 0,93 selon le modèle, sera qualifié d'assuré à « haut risque ».

Nous nous sommes rendu compte de certaines limites de notre étude qu'il convient de signaler ici. Les variables dépendantes, même si elles sont pertinentes ne sont pas suffisantes. La base de données mise à notre disposition ne comportait qu'une liste limitée de variables. Pour mieux apprécier et bien expliquer la mortalité, nous pensons que, sans être exhaustif, d'autres variables comme *le nombre d'enfants à charges, le statut matrimonial, le niveau scolaire atteint, le revenu annuel* peuvent être incluses dans l'analyse. Voilà pourquoi nous préconisons que désormais, les propositions d'assurances devront contenir des questions qui permettent de prendre en compte ces informations et par suite, faire l'objet d'une analyse dans le temps. Car une étude rigoureuse exige que soient réunies les exigences de quantité suffisante du portefeuille et du comportement du portefeuille dans le temps.

Concevoir une stratégie de contrôle des assurés à la souscription, afin d'éviter l'antiselection, telle doit être l'une des préoccupations de l'assureur car les questionnaires médicaux élaborés pour filtrer l'entrée à l'assurance présentent des limites du point de vue fiabilité des réponses des assurés. C'est pourquoi nous nous sommes permis de donner le profil des assurés présentant un risque majeur pour la compagnie d'assurance. Cet outil devait alors permettre de corriger les primes à la souscription de certains assurés sinon de penser à l'élaboration de contrats adaptés aux besoins de certains assurés et ce dans le but de ne pas faire prendre en charge par la mutualité des assurés dont la probabilité de décès est élevée. Nous pensons notamment aux assurés du domaine de l'éducation, qui ont une forte probabilité de décès avant le terme du contrat, conduisant en conséquence à une mortalité réelle supérieure à la mortalité théorique. Un contrat spécifique pour ce groupe devrait donc permettre d'ajuster les primes et de ce fait, chacun des assurés paierait la juste prime.

Le modèle fourni est un outil devant permettre de prendre des mesures de précautions dans la tarification des risques assurés. Toutefois, en fonction des objectifs fixés par la compagnie d'assurance, elle peut expressément décider de faire de l'antiselection. Alors le modèle est donc adapté à la situation de tout assureur vie qui désire se procurer de moyens divers et suffisamment complémentaires pour vendre au mieux ses contrats d'assurance, car nous pensons que l'art de vaincre la fraude à l'assurance devrait avoir pour corollaire l'art du contrôle permanent des différents portefeuilles détenus par les assureurs.

ANNEXES

Annexe 1 : Commandes utilisées sur STATA au cours de l'analyse

-Pour la comparaison des proportions dans un tableau de contingence Var1 en fonction de Var2.

marasc Var1 Var2

-Pour régression logistique de la variable dichotomique dépendante Var1 en fonction des variables explicatives Var2 Var3...

logistic Var1Var2 Var3 ...

-Pour l'histogramme de la variable Var1

Histogram Var1

-Pour déterminer la sensibilité et la spécificité:

Lstat

lsens

-Pour la courbe ROC:

lroc

-Pour calculer les probabilités prédites par le modèle:

predict p

Annexe 2 : Répartition de la population des assurés décédés en fonction du Secteur d'activité et du sexe

SECTEUR D'ACTIVITE	RESIDENCE		TOTAL
	Abidjan	Hors Abidjan	
ADMINISTRATION	46%	54%	100%
ARMEE	33%	67%	100%
TECHNIQUE	54%	46%	100%
COMPTABILITE	83%	17%	100%
EDUCATION	17%	83%	100%
INDUSTRIE CHIMIQUE	0%	100%	100%
SANTE	40%	60%	100%
SECURITE NAT	50%	50%	100%
TRANSPORT	100%	0%	100%
AUTRES SERVICES/PROD	45%	55%	100%
TOTAL	31%	69%	100%

Source : Notre étude

Annexe 3 : Mois moyen de survénance des décès par secteur d'activité

SECTEUR D'ACTIVITE	Mois moyen de survénance du décès
ADMINISTRATION	5,7
ARMEE	6,0
TECHNIQUE	6,2
COMPTABILITE	5,8
EDUCATION	6,0
INDUSTRIE CHIMIQUE	11,0
SANTE	7,0
SECURITE NAT	4,0
TRANSPORT	7,0
AUTRES SERVICES/PROD	5,3
TOTAL	6,0

Source : Notre étude

Annexe 4 : Délai de survénance des décès par secteur d'activité

SECTEUR D'ACTIVITE	*DELAISIN= Délai de survénance du décès en années	TOTAL
ADMINISTRATION	Moyenne de DELAISIN	2,5
	Max de DELAISIN	6,0
	Min de DELAISIN	0,0
ARMEE	Moyenne de DELAISIN	2,0
	Max de DELAISIN	3,0
	Min de DELAISIN	1,0
TECHNIQUE	Moyenne de DELAISIN	2,8
	Max de DELAISIN	5,0
	Min de DELAISIN	0,0
COMPTABILITE	Moyenne de DELAISIN	2,3
	Max de DELAISIN	5,0
	Min de DELAISIN	0,0
EDUCATION	Moyenne de DELAISIN	1,7
	Max de DELAISIN	6,0
	Min de DELAISIN	0,0
INDUSTRIE CHIMIQUE	Moyenne de DELAISIN	1,0
	Max de DELAISIN	1,0
	Min de DELAISIN	1,0
SANTE	Moyenne de DELAISIN	4,0
	Max de DELAISIN	6,0
	Min de DELAISIN	1,0
SECURITE NAT	Moyenne de DELAISIN	3,3
	Max de DELAISIN	6,0
	Min de DELAISIN	0,0
TRANSPORT	Moyenne de DELAISIN	7,0
	Max de DELAISIN	7,0
	Min de DELAISIN	7,0
AUTRES SERVICES/PROD	Moyenne de DELAISIN	2,6
	Max de DELAISIN	8,0
	Min de DELAISIN	0,0
TOTAL	Moyenne de DELAISIN	2,2
	Max de DELAISIN	8,0
	Min de DELAISIN	0,0

Source : Notre étude

Annexe 5: Capitaux totaux et primes versées par les assurés décédés par secteur d'activité

SECTEUR D'ACTIVITE		TOTAL
ADMINISTRATION	Somme des Capitaux	15 726 785
	Moyenne des Capitaux	1 209 753
	Somme des Primes	144 000
	Moyenne des Primes	11 077
ARMEE	Somme des Capitaux	3 324 463
	Moyenne des Capitaux	1 108 154
	Somme des Primes	30 000
	Moyenne des Primes	10 000
TECHNIQUE	Somme des Capitaux	55 061 245
	Moyenne des Capitaux	2 117 740
	Somme des Primes	260 000
	Moyenne des Primes	10 000
COMPTABILITE	Somme des Capitaux	12 600 000
	Moyenne des Capitaux	2 100 000
	Somme des Primes	60 000
	Moyenne des Primes	10 000
EDUCATION	Somme des Capitaux	166 062 377
	Moyenne des Capitaux	1 677 398
	Somme des Primes	1 042 000
	Moyenne des Primes	10 525
INDUSTRIE CHIMIQUE	Somme des Capitaux	1 800 000
	Moyenne des Capitaux	1 800 000
	Somme des Primes	10 000
	Moyenne des Primes	10 000
SANTE	Somme des Capitaux	8 021 896
	Moyenne des Capitaux	1 604 379
	Somme des Primes	50 000
	Moyenne des Primes	10 000
SECURITE NAT	Somme des Capitaux	6 226 802
	Moyenne des Capitaux	1 556 701
	Somme des Primes	40 000
	Moyenne des Primes	10 000
TRANSPORT	Somme des Capitaux	1 680 000
	Moyenne des Capitaux	1 680 000
	Somme des Primes	10 000
	Moyenne des Primes	10 000
AUTRES SERVICES/PROD	Somme des Capitaux	14 944 762
	Moyenne des Capitaux	1 358 615
	Somme des Primes	115 000
	Moyenne des Primes	10 455
TOTAL	Somme des Capitaux	285 448 330
	Moyenne des Capitaux	1 689 043
	Somme des Primes	1 761 000
	Moyenne des Primes	10 420

Source : Notre étude

Annexe 6 : Caractéristiques de l'âge par secteur d'activité et par âge

		SEXE			
		HOMMES	FEMMES	TOTAL	
SECTEUR D'ACTIVITE	ADMINISTRATION	Moyenne de AGESOUS	35,8	34,5	35,4
		Max de AGESOUS	42,0	41,0	42,0
		Min de AGESOUS	31,0	28,0	28,0
	ARMEE	Moyenne de AGESOUS	26,3		26,3
		Max de AGESOUS	29,0		29,0
		Min de AGESOUS	24,0		24,0
	AUTRES SERVICES/PROD	Moyenne de AGESOUS	38,1	44,3	39,8
		Max de AGESOUS	46,0	47,0	47,0
		Min de AGESOUS	31,0	40,0	31,0
	COMPTABILITE	Moyenne de AGESOUS	38,5		38,5
		Max de AGESOUS	49,0		49,0
		Min de AGESOUS	28,0		28,0
	EDUCATION	Moyenne de AGESOUS	45,9	38,8	45,4
		Max de AGESOUS	54,0	55,0	55,0
		Min de AGESOUS	26,0	25,0	25,0
	INDUSTRIE CHIMIQUE	Moyenne de AGESOUS	40,0		40,0
		Max de AGESOUS	40,0		40,0
		Min de AGESOUS	40,0		40,0
	SANTE	Moyenne de AGESOUS	37,7	30,0	34,6
		Max de AGESOUS	49,0	33,0	49,0
		Min de AGESOUS	32,0	27,0	27,0
SECURITE NAT	Moyenne de AGESOUS	29,0		29,0	
	Max de AGESOUS	33,0		33,0	
	Min de AGESOUS	25,0		25,0	
TECHNIQUE	Moyenne de AGESOUS	41,1	49,0	41,7	
	Max de AGESOUS	56,0	50,0	56,0	
	Min de AGESOUS	20,0	48,0	20,0	
TRANSPORT	Moyenne de AGESOUS	45,0		45,0	
	Max de AGESOUS	45,0		45,0	
	Min de AGESOUS	45,0		45,0	
TOTAL	Moyenne de AGESOUS	42,8	38,9	42,3	
	Max de AGESOUS	56,0	55,0	56,0	
	Min de AGESOUS	20,0	25,0	20,0	

Source : Notre étude

BIBLIOGRAPHIE

1. **BAKAYOKO (S)**, Cours d'assurance de personnes, IIA DESSA, 2005
2. **BRY (X)**, Introduction à l'analyse factorielle et à la classification, ENEA Dakar, les cahiers de la statistique et de l'économie appliquée, 1997, p. 59
3. **BRY (X)**, Introduction à l'analyse factorielle des correspondances (simples et multiples), ENEA Dakar, les cahiers de la statistique et de l'économie appliquée, 1997, p. 45
4. **DAVID (R.A) et al.**, Statistiques pour l'économie et la gestion, De Boeck, Université sa, Nouveaux horizons, p.779, ISBN 2-7445-0121-2
5. **JAMBU (M)**, Méthodes de base de l'analyse des données, Paris, éd. Eyrolles, Janvier 1999, p. 400, ISBN 2-212-05256-1
6. **KANKUMO (T.O)**, Analyse statistique du profil des clients à haut risque du portefeuille automobile d'une compagnie d'assurance, Université de Yaoundé I, Soutenance de mémoire 2005, p.55
7. **KEHO (Y)**, Modèles dichotomiques logit et probit ; de la théorie à la pratique, ENSEA ITS, 2004
8. **PETAUTON (P)**, Théorie et pratique de l'assurance vie, Bordas, Paris, 1991, p. 199, ISBN 2-04-019862-8
9. **PIRON (M) et al.**, Statistique exploratoire multidimensionnelle, Paris, Dunod, 1995, p.433, ISBN 2-10-002886-3

TABLE DES MATIERES

	Page
AVANT-PROPOS	1
RESUME	3
ABSTRACT	4
INTRODUCTION	5
PARTIE I : ETUDE STATISTIQUE DU PORTEFEUILLE « SIKA RETRAITE »	8
CHAPITRE I : STATISTIQUE DESCRIPTIVE DU PORTEFEUILLE « SIKA RETRAITE »	9
I. CARACTERISTIQUES DU CONTRAT « SIKA RETRAITE »	9
A. Catégorie du contrat	9
1. Objet du contrat	9
2. Type du contrat	9
B. Principes de fonctionnement	9
1. Population assurable	9
2. Intervenants au contrat	10
2.1. Obligations de l'assuré	10
2.2. Obligations de NSIA Vie	10
II. STATISTIQUES DU PORTEFEUILLE	11
A. Population des assurés « Sika Retraite »	11
1. Composition du portefeuille	11
2. Structure du portefeuille	11
2.1. Répartition de la population assurée en fonction du sexe	11
2.2. Répartition de la population assurée en fonction du lieu de résidence	11
2.3. Répartition de la population assurée en fonction du secteur d'activité	12
B. Population des décès intervenus	13
1. Composition du portefeuille décès	13
2. Structure du portefeuille décès	13
2.1. Répartition de la population assurée décédée en fonction de l'âge à la souscription et du sexe	13
2.2. Répartition de la population assurée décédée en fonction du lieu de résidence	14
2.3. Répartition de la population assurée décédée en fonction du secteur d'activité	14
CHAPITRE II : STATISTIQUE EXPLORATOIRE DU PORTEFEUILLE « SIKA RETRAITE »	16
I. METHODES D'ANALYSE FACTORIELLE DES DONNEES	17
A. Analyse factorielle : Géométrie de l'information	17
1. Objectif	17
2. Données et choix des méthodes	17
B. Différentes méthodes d'analyse factorielle	17
1. Données quantitatives : ACP	17
1.1. Objectif	17
1.2. Principes d'analyse	18
2. Données qualitatives : AFC et ACM	18
2.1. Objectifs	18
2.2. Principes d'analyse	19

II. PRESENTATION DES RESULTATS DE L'ANALYSE	19
A. Choix de la méthode	19
1. Justification du choix	19
2. Choix des variables	20
2.1. Variables actives	20
2.2. Variables illustratives	20
B. Résultats de l'analyse	20
1. Nuage des points variables	20
2. Nuage des points individus	23
PARTIE II : MODELISATION DE LA MORTALITE DU PORTEFEUILLE « SIKA RETRAITE »	26
CHAPITRE I: CLASSIFICATION AUTOMATIQUE DES ASSURES	27
I. METHODE DE CLASSIFICATION	27
A. Objet	27
1. Objectif	27
2. Différentes méthodes de classification	27
B. Raisons de la classification	28
1. Classification des individus	28
2. Classification des variables	28
II. RESULTATS	28
A. Choix du nombre de classes	28
1. Arbre de classification	28
2. Nombre de classes à retenir	29
B. Description des classes	29
1. Description des partitions	29
2. Caractéristiques des classes	31
CHAPITRE II : MODELISATION DE LA PROBABILITE DE DECES	33
I. MODELISATION DECISIONNELLE	33
A. Analyse discriminante	33
1. Présentation	33
2. Objectifs	33
2.1. Objectif descriptif	33
2.2. Objectif prévisionnel	33
B. Résultats de l'analyse	34
1. Tableau d'affectation des individus	34
2. Fonction score	35
II. REGRESSION LOGISTIQUE	36
A. Modèle dichotomique	36
1. Justification du modèle	36
2. Formalisation du modèle	37
B. Résultats du modèle	38
1. Interprétation des résultats	38
1.1. Significativité des coefficients et qualité de l'ajustement	38
1.2. Evaluation du pouvoir prédictif du modèle	39
2. Prédiction	41
2.1. Erreur de prédiction.	41
2.2. Choix de la probabilité seuil et Applications du modèle	42
CONCLUSION	46
ANNEXES	48
BIBLIOGRAPHIE	52

**LISTE DES TABLEAUX ET
GRAPHIQUES**

Tableaux

Tableau 1 : Répartition de la population assurée en fonction du sexe de l'assuré.....	11
Tableau 2 : Répartition de la population assurée en fonction du lieu de résidence de l'assuré	12
Tableau 3 : Répartition des assurés en fonction du sexe et de l'âge.....	13
Tableau 4 : Répartition de la population assurée décédée en fonction du sexe de l'assuré.....	14
Tableau 5 : Caractérisation des classes retenues.....	30
Tableau 6 : Classification des individus sous SPAD.....	34
Tableau 7 : Estimation du modèle.....	38
Tableau 8 : Pouvoir de prédiction du modèle.....	41
Tableau 9 : Fréquences des Probabilités estimées par le modèle sous (H_0).....	43
Tableau 10 : Fréquences des Probabilités estimées par le modèle sous (H_1).....	44
Tableau 11 : Pronostics sur les assurés à la souscription du contrat.....	45

Graphiques

Graphique 1 : Répartition de la population assurée en fonction du secteur d'activité de l'assuré	12
Graphique 2 : Répartition de l'ensemble des décès intervenus en fonction du secteur d'activité	14
Graphique 3 : Représentation des variables dans le plan factoriel (1 , 2)	21
Graphique 4 : Représentation des variables dans le plan factoriel (1 , 3)	22
Graphique 5 : Représentation des variables dans le plan factoriel (2 , 3)	22
Graphique 6 : Représentation des variables et des individus dans le plan factoriel (1 , 2)	24
Graphique 7 : Représentation des variables et des individus dans le plan factoriel (2 , 3)	24
Graphique 8 : Représentation des variables et des individus dans le plan factoriel (1 , 3)	25
Graphique 9 : Arbre de classification des assurés du portefeuille	29
Graphique 10 : Graphe de la fonction score	35
Graphique 11 : Courbe ROC du modèle	40
Graphique 12 : Histogramme des probabilités estimées	42

Annexes

Annexe 1 : Commandes utilisées sur STATA au cours de l'analyse.....	48
Annexe 2 : Répartition de la population des assurés décédés en fonction du Secteur d'activité et du sexe.....	48
Annexe 3 : Mois moyen de survenance des décès par secteur d'activité.....	49
Annexe 4 : Délai de survenance des décès par secteur d'activité.....	49
Annexe 5 : Capitaux totaux et primes versées par les assurés décédés par secteur d'activité..	50
Annexe 6 : Caractéristiques de l'âge par secteur d'activité et par âge.....	51

RESUME

La tarification d'un contrat d'assurance exige en plus des éléments techniques, la connaissance de certaines caractéristiques que dispose l'assuré. C'est une tâche importante car la solvabilité de l'assuré en dépend fortement. Pourtant cette tâche n'est pas facile quand on sait que ce qui vient de l'assuré comme renseignements n'est pas garanti de fiabilité, empêchant ainsi l'assureur de mieux apprécier le risque. Dans cette perspective, l'assureur est condamné de se munir d'assez de moyens de contrôle complémentaires. Ainsi, notre objectif dans la présente étude est de fournir un outil performant permettant d'apprécier à la souscription du contrat le risque que souhaite couvrir l'assureur : le décès de l'assuré.

L'assurance, et en particulier l'assurance vie a la propriété de n'attirer en majorité que des personnes dont la réalisation du risque est quasi certain. Par exemple, les personnes malades ont tendance à souscrire les contrats en cas de décès. Pour cela, l'étude a été menée de manière à pronostiquer le décès ou la survie d'un assuré lorsque l'assureur dispose des caractéristiques de celui-là. Ces caractéristiques sont: *l'âge à la souscription du contrat, le sexe, le secteur d'activité et le lieu de résidence de l'assuré.*

Un modèle dit dichotomique a été élaboré permettant d'évaluer la probabilité de décès durant la durée du contrat connaissant les caractéristiques de l'assuré ci dessus. Pour l'atteinte de cet objectif, une conduite méthodique a été élaborée. Une étude statistique préalable a été faite et a permis d'explorer le portefeuille, toute chose ayant permis de conduire avec efficacité la modélisation.