

**CLASSIFICATION DES
CATEGORIES
SOCIO-ECONOMIQUES AU
CAMEROUN.**

Par :

Jerôme DONGMO AZEBAZE
Maître ès Sciences

Dirigé par:

Pr. Henri GWET

Chef de Département de Mathématiques et Sciences
Physiques à l'ENSP de Yaoundé.

Octobre 2007

Dédicaces

Je dédie ce modeste travail à :

- mon père ZEBAZE Bernard, qu'il trouve dans ce travail le fruit de ses nombreux sacrifices ;
- A tous les membres de ma famille, pour leur hospitalité et soutien.

Remerciements

Au pr. Henri GWET, Maître de conférences à l' ENSP, qui malgré ses multiples occupations à su coordonner avec tact les enseignements du ce Master, et pour avoir accepté de diriger ce travail.

A M .TCHAMDA Claude, Ingénieur statisticien, pour m'avoir permis d'effectuer ce stage dans son service à la sous direction des Statistiques Démographiques et Sociales de l'INS.

A M. NANTCHOUANG Azer, ingénieur statisticien économiste, mon encadreur de stage, qui malgré ses multiples occupations, m'a initié dans le domaine de l'économie, spécialement dans les analyses de la pauvreté.

A tous mes enseignants de Master de statistiques appliquées pour leur encadrement, en particulier au Dr Eugène- Patrice NDONG NGUEMA qui n'a ménagé aucun effort pou notre formation.

A tous les membres de ma famille pour leur soutien.

A tous mes camarades de promotion pour leur solidarité. Je pense ici spécialement à TCHAKOUTE Romain, TAYOU Juliette.

A M .NGOGANG Léandre, TSAGUIM Elvis, DONGMO Gislain , cadres à l'INS pour leur aide.

Que tous ceux qui, de près ou de loin, ont contribué à l'achèvement de ce travail et qui n'ont été cité trouvent ici mes remerciements les plus profonds.

Liste des tableaux

Table des figures

Abréviations

CDH : Classification Descendante Hiérarchique

DSRP : Document de Stratégies de Réduction de la Pauvreté

INS : Institut Nationale de la Statistique

PPTE: Pays Pauvre Très Endetté

SPAD : Système Portable d'Analyse des données

ICP : Indicateur Composite de Pauvreté.

Avant propos

Le stage pratique obligatoire de fin de formation des étudiants du Master de Statistique Appliquée, de l'École Nationale Supérieure Polytechnique (ENSP), est un élément essentiel dans leur formation. En effet les connaissances théoriques accumulées y sont appliquées. C'est aussi l'occasion de se familiariser avec les différents outils statistiques. En outre, les expériences acquises par l'étudiant lui donnent la possibilité d'identifier et de s'intéresser à des voies d'approfondissement de la formation théorique. En somme, c'est un appoint considérable à des futurs chercheurs et cadres qui auront à charge les travaux de conception, de collecte, d'analyse, et de diffusion de l'information .

Ce stage qui permet, en outre un contact avec les réalités du monde du travail, a duré 3 mois. Le nôtre s'inscrivant en conformité avec les objectifs préalablement cités, s'est déroulé à la sous direction de la Statistique Démographique et Sociale de l'Institut National de la Statistique du Cameroun du 1er Juillet au 31 Septembre. Le thème d'étude qui a été retenu est la classification des catégories socio-économique au Cameroun »

PRESENTATION DE LA STRUCTURE D'ACCUEIL

L'Institut National de la Statistique est un établissement public et administratif créé en 2001 par décret présidentiel Numéro 2001 /100 du 20 avril 2001. Il remplace la Direction de la Statistique et de la Comptabilité Nationale (DSCN) qui jusqu'alors était sous le Ministère de la Planification, de la Programmation du Développement et de l'Aménagement du Territoire (MINPLAPDAT).

Les missions assignées à l'INS sont de divers ordres. Il est chargé :

- de rendre disponibles les données et les indicateurs statistiques nécessaires à la gestion économique et sociale ;
- d'assurer la conservation des fichiers de recensements et enquêtes réalisés par les administrations publiques et les organismes subventionnés ou contrôlés par l'État ;
- de favoriser le développement des sciences statistiques et les recherches économiques relevant de sa compétence, de promouvoir la formation du personnel spécialisé pour le fonctionnement du système national d'information statistique.

L'organigramme de l'Institut National de Statistique n'étant pas encore adopté, celui de l'ancienne structure (la DSCN), dont les sous-directions sont les suivantes, reste en application :

- la Sous- Direction des Etudes et de la Normalisation Statistique (ENS) ;
- la Sous- Direction de la Comptabilité Nationale et des Synthèses Statistiques (CSS) ;
- la Sous -Direction des Statistiques Démographiques et Sociales (SDS) ;
- la Sous- Direction des Statistiques d'Entreprises (SSE) ;
- la Sous-Direction de l'Informatique et des Banques de Données (IBD) ;
- la Division des Enquêtes et Etudes Statistiques auprès des Ménages (DEM) ;
- le Secrétariat Permanent du Plan Comptable (SPC) ;
- le Service d'Ordre (SO).

Résumé

La plus part des pays africains au sud du sahara sont perpétuellement engagés dans la lutte contre la pauvreté. Pour bien mener cette politique, la connaissance de la structure de la population s'avère nécessaire. La classification est l'un de ses outils pouvant contribuer à la résolution de ce problème. En effet la classification a pour de grouper un ensemble d'objet en classes de telles sorte que deux objets appartenant à une même classe soient plus similaires entre eux que deux objets appartenant à des classes différentes, et ceux relativement à un critère fixé comme mesure de proximité. On distingue les méthodes de classification hiérachiques et les méthodes de partitionnements. Après avoir présenté la méthode de Classification Ascendante Hiérachique, nous appliquons à la base de données Enquête Camerounaise Auprès des Ménages(ECAMII). Grace à la regression logistique nous avons éssayer d'expliquer les classes obtenues après la classification.

Abstract

More the share of the African countries in the south of the Sahara are perpetually engaged in the fight against poverty. For following this policy well, the knowledge of the structure of the population proves to be necessary. Classification is one of its tools being able to contribute to the resolution of this problem. Indeed classification has for grouping a whole of object in classes of such kind that two objects belonging to the same class are more similar between them than two objects belonging to different classes, and those relative with a criterion fixed like measures proximity.

One distinguishes the hierachic methods of classification and the methods of partitionings. After having presents the method of Ascending Classification Hiérachique, we apply to the Cameronian data base Enquête Near Households (ECAMII). Thanks to the logistic regression we have to try to explain the classes obtained after classification

Introduction

Contexte

Un Problème commun à bon nombre de pays africains au sud du Sahara depuis la fin des années 80 est la grande pauvreté. Selon les estimations de la Banque Mondiale (2000), de 217,2 millions de personnes vivant avec moins de 1 dollars par jour en Afrique subsaharien en 1987, l'on est passée à 290,9 millions en 1998, soit un accroissement de 34%. En particulier, au Cameroun, la décennie de 1985 à 1994 constitue une période de grave crise économique, période marquée par de fortes turbulences et qui s'est traduite par des déséquilibres macro-économiques et financiers importants. L'on a alors assisté à une dégradation de l'activité économique en raison notamment de la baisse des principaux produits d'exportation (pétrole, cacao, café, coton). L'on a également assisté durant cette période à une réduction de l'offre publique des services de base, à la montée du chômage, à la baisse des revenus et à la détérioration des conditions et cadre de vie des ménages.

Depuis 1995, le Cameroun a renoué avec la croissance économique grâce aux réformes macroéconomiques et budgétaires engagées par les autorités. Dès lors la reprise économique observée depuis 1995 a amélioré les revenus d'exportation et généralement les finances publiques mais n'a pas encore permis de sortir les populations de leur état de pauvreté. Face à ce constat, les autorités nationales appuyées par des institutions financières internationales (Banque Mondiale, Fond Monétaire Internationale, Programme des Nations Unies pour le Développement), se sont engagées dans une nouvelle génération de politiques économiques sociales mises en perspective dans le Document de Stratégies de Réduction de la Pauvreté (DSRP) préparé dans le cadre de l'initiative renforcée d'allègement de la dette des Pays Pauvres Très Endettés (PPTE). A travers ces politiques, les autorités poursuivent entre autres objectifs le renforcement de la sphère sociale. La lutte contre la pauvreté se situe ainsi au coeur des préoccupations des nouvelles politiques au Cameroun. Cependant, ces politiques ne peuvent être plus rendables et plus efficaces que si elles sont basées sur des études menées.

Dès lors, pour contribuer alors à l'élaboration des stratégies permettant de mieux orien-

ter les actions de planification conduites au niveau nationale et même internationale, et pour assurer également un meilleur accès des ménages aux services sociaux de base, il est nécessaire de bien connaître la structure de la population, c'est-à-dire déceler les groupes plus ou moins homogènes caractériser par les critères spécifiques : c'est la catégorisation de la société. Elle a essentiellement pour but d'identifier les différents groupes de la population qui sont, d'une part, assez homogènes, et d'autre part, suffisamment distincts par leur comportement des autres groupes, et serviront à établir un rapport entre la situation des personnes et un certain nombre de phénomènes démographiques, sociaux, économiques et culturels. La classification de la population devient donc une nécessité pour tout organisme qui a pour mission l'étude des disparités et inégalités en vue de la réduction de la pauvreté

Problématique

«Le niveau de vie d'un individu est un concept pluridimensionnel basé sur des normes socio culturelles et reflétant une interaction entre ses capacités, sa dotation en ressources et les opportunités offertes par sa société » (voir [2]). Le concept de niveau de vie donne donc un critère fondamental d'évaluation de la performance de tout système socioéconomique, c'est pourquoi l'amélioration des niveaux de vie de la population constitue l'un des défis majeurs des différentes politiques économiques menées au Cameroun. Cette amélioration passe par la disposition d'éléments permettant de décrire la situation sociale et économique du pays.

L'identification des divers regroupements socio-économiques est l'un de ces éléments indispensables à une bonne description de l'environnement socio-économique en général, et l'identification de la classe des pauvres en particulier.

Vue que la pauvreté est multiforme et recouvre diverses dimensions, l'on se pose la question de savoir comment cerner ce phénomène. Dubois (2001) [5] dans un article consacré à une réflexion sur la définition de la pauvreté distingue de façon fonctionnelle trois formes de pauvreté :

- la pauvreté monétaire « ou de revenu » qui exprime un aspect du niveau de vie et résulte d'une insuffisance de ressources engendrant une consommation insuffisante ;
- la pauvreté des conditions de vie ou « d'existence » qui traduit le fait qu'il devient difficile de satisfaire un grand nombre de besoins fondamentaux et que l'on se trouve en situation de manque dans des domaines relatifs à l'alimentation (déséquilibre nutritionnel), à la santé (non accès aux soins primaires), à l'éducation (non scolarisation), au logement, etc.
- et la pauvreté de potentialités ou de « capacités » qui fait que l'on n'a pas réussi à constituer le capital minimal qui permet de vivre, de fonctionner correctement et

de créer, dans une société donnée, cela engendrant une insuffisante mise en valeur des capacités individuelles.

Au Cameroun, de nombreuses sources d'informations, notamment les deux grandes enquêtes ECAM I (1996) et ECAM II (2001), nous donnent une idée sur les manifestations et les facteurs de la pauvreté. Toute fois la littérature sur la pauvreté au Cameroun accorde une part très importante à l'approche monétaire dans le ciblage des pauvres. Des travaux conduits dans le cadre du DSRP soulignent que, du point de vue des populations, les éléments liés au revenu constituent la plus importante classe de caractéristique de la condition de la pauvreté au Cameroun.

Bien que la politique de lutte contre la pauvreté basée sur le revenu domine sur les autres, elle est limitée en ce sens que, l'accroissement du revenu des ménages pauvres n'est pas la meilleure façon d'accroître la satisfaction des besoins de base. (Voir [1])

Quelques unes des raisons évoquées sont les suivantes :

-les besoins en éducation, santé, eau et en hygiènes sont plus facilement satisfaits par des services publics que par des revenus accrus.

-les individus n'utilisent pas toujours leurs revenus pour accroître leurs nutriments et leur santé;

-il y a souvent une distribution inéquitable des ressources à l'intérieur des ménages.

De ce point de vue, dans un environnement comme celui du Cameroun où l'offre de certains biens de base est limitée (eau potable, électricité, moyens de communications, etc.) il est d'une nécessité indéniable que le cadre de vie des ménages au sens large (logement, équipements, accessibilité à certaines infrastructures de base) nécessite une étude particulière.

Dès lors, une étude de la pauvreté par l'approche condition de vie s'avère nécessaire car elle favorise des politiques ciblées. Cette approche a déjà fait l'objet

d'étude menée par certains auteurs à l'instar de B. FOKO TAGNE, F. NDEM , R. TCHAKOTE , Dans leurs études, ils utilisent certaines méthodes statistiques (ACM, AFM) pour construire l'indicateur composite de pauvreté.

Dans notre étude en particulier, nous utiliserons comme technique d'analyse la classification en prenant pour variables actives celles relatives aux conditions de vie. Cette méthode a l'avantage de détecter non seulement la classe des personnes jugées pauvre, mais aussi d'autres classes se retrouvant dans la société. Ainsi, ce travail se voit attribuer les objectifs suivants : -construire des ensembles homogènes d'individus liés entre eux par un modèle de vie et de comportement. -chercher une règle d'affectation aux différentes classes sociales qui se dégageront. La concrétisation de ces objectifs permettra de définir les critères qui justifient les regroupements au sein de la société camerounaise, de mieux appréhender les inégalités et d'identifier les classes sociales pauvres.

Plan de travail

Pour atteindre les objectifs fixés, nous avons divisées notre travail en trois chapitres. Le premier chapitre est réservé aux concepts de base. Nous faisons ici dans un premier temps une revue théorique des notions de classes sociales et de la pauvreté ,et ensuite nous présentons la méthode de classification Ascendante Hiérachique

Au chapitre 3 nous proposons une méthode de catégorisation de la population camerounaise en classes. Par la suite , nous utilisons le logiciel SPAD pour appliquer cette méthode aux données ECAM II. AU chapitre 4, nous allons utiliser la regression logistique pour batir modèle d'affectation des individus dans les classes sociales.

Résumé exécutif

Sujet

Classification des catégories socio économiques et pauvreté

Problème

: Avec les données de l'enquête ECAMII , l'Institut National de la Statistique a mis sur pied un indicateur de niveau de vie basé sur l'approche monétaire .Pour sa construction, cet indicateur de vie a été approché par la consommation finale annuelle des ménages qui dans une certaine mesure, reflète mieux le niveau de vie des ménages Comparativement aux revenus qui est difficile à mesurer. Avec cet indicateur on distingue les personnes et non pauvres. Le problème ici est 'étudier le regroupement de la population camerounaise suivant les critères liés à l'emploi et aux conditions de vie d'une part et d'autre part retrouvé parmi ces classes celle des personnes pauvres. Il est aussi question d'établir une règle d'affectation ou d'expliquer l'appartenance à un groupe.

Méthodologie

Pour déterminer le regroupement des ménages en classe, nous avons utilisé l'ACM afin d'utiliser ses coordonnées factoriels dans la classification. ensuite nous avons utilisé la méthode de Classification Ascendante Hiérarchique qui consiste à fusionner deux objets (ou individus) au sens d'une mesure de proximité de sorte que deux objets groupés à une étape le reste jusqu'au terme du processus de classification. Il s'agit ici, à partir des éléments terminaux, de former de petites classes ne comportant que des individus les plus semblables, et à partir de celle-ci, de construire des classes de moins en moins homogènes jusqu'à obtenir la classe toute entière qui réunit tous les éléments.

Les classes étant constituées, nous utilisons la regression logistique pour expliquer la probabilité d'appartenance à chaque classe.

Résultats

Nous présentons ici les principaux résultats de notre travail : Le nombre de classe obtenu est de trois et les statistiques sont resumées dans le tableau suivant :

Classe	Effectif	Pourcentage (%)
Classe1	4736	43.27
Classe2	4847	44.30
Classe3	1360	12.43
Total	10943	100

Nous avons constaté que : Les ménages de la classe1 sont caractérisés par les conditions de vie précaire. les chefs de familles travaillent le plus souvent dans leur propre compte et exerçant les entreprises agricoles informelles. Un tableau croisé de la variable classe1 avec le milieu de residence montre que près de 70% des ménages ici sont dans le milieu rural. Les ménages de la classe2 sont caractérisés par des conditions de vie moyennes. Les ménages de la classe 3 sont caractérisés ici par de très bonnes conditions de vie . En effet les chefs de ménages ici possèdent les possèdent presque les équipement de hautes valeur,logent dans les villa, ect. Ils travaillent dans les administration publiques et entreprises privées informelles. Ce sont des cadres en ce qui concerne le statut professionnel.

CONCEPTS DE BASE

Dans ce chapitre, nous présentons quelques aspects de notions de classes sociales et de pauvreté.

1.1 Concepts de classes sociales

La notion de classe sociale fait le plus souvent référence aux inégalités sociales. Ce concept a longtemps constitué et constitue encore l'un des thèmes les plus discutés dans la littérature sociologique. Plusieurs théories, parfois contradictoire sont souvent développées par des chercheurs. L'une de ces théories découle des travaux de Karl Marx, et l'autre, de ceux de Max Weber, qui remet en question la pensée marxiste.

Selon la pensée marxiste, les classes sociales se définissent en fonction de la propriété des moyens de production et de la force de travail. Les marxistes reconnaissent trois classes principales : la petite bourgeoisie, dont les membres possèdent des entreprises (moyens de production), travaillent à leur compte et n'ont pas d'employés; le prolétariat ou la classe ouvrière, dont les membres ne détiennent aucun moyen de production et échangent leur force de travail contre un salaire; et la bourgeoisie ou la classe capitaliste, regroupant les détenteurs des moyens de production, qui achètent la force de travail, vivent bien et s'enrichissent grâce à la plus-value provenant du travail des ouvriers. (La plus-value correspond à la différence entre la valeur totale des marchandises lorsqu'elles sont échangées sur le marché et la valeur de la force de travail et des moyens de production).

Les non-marxistes affirment que les classes sociales peuvent se définir selon les inégalités de revenu, d'instruction, de pouvoir et de prestige professionnel, mais ils étudient souvent ces formes d'inégalités sociales sans tenir compte de la définition marxiste de classe sociale. Ils identifient par exemple diverses classes sociales selon le facteur d'inégalité sociale faisant l'objet d'une étude donnée. Ils classent et ordonnent les groupes selon leurs critères établis (revenu, instruction, pouvoir, prestige professionnel).

D'après Max WEBER, On peut parler de «classes» lorsqu' un certain nombre d'individus ont en commun une composante causale spécifique quant à leurs chances de

vie, composante se manifestant d'une part exclusivement par des intérêts économiques, concernant la possession de biens et les possibilités d'obtention de revenus, et se manifestant d'autre part, sur le marché des biens ou sur le marché du travail. (Voir [6])

1.2 Concepts de pauvreté

La notion de pauvreté est une catégorie fondamentalement descriptive : est pauvre celui qui, comparé avec d'autres individus de sa société, est placé, d'après une série de traits retenus comme principes de catégorisation, dans les plus bas niveaux. Ainsi la notion de pauvreté renvoie à l'idée de certaines carences en biens et en services qui, dans une société déterminée, sont considérées comme indispensables à tous ses membres (Véronique Petit, 2006).

Pour évaluer le bien être individuel, il existe plusieurs approches conceptuelles qui se distinguent par l'importance qu'attache l'analyste à la manière dont l'individu juge son bien être et la gamme des facteurs pris en compte qu'on se fonde ou non sur le revenu (Ravallion, 1996).

Approche welfarist

C'est une approche fondée sur le bien-être économique. Les welfarist ramènent le concept de bien-être soit directement au concept d'utilité commun en économie, soit indirectement via le terme bien-être économique compris comme l'utilité générée par la consommation totale. L'utilité elle-même est conçue comme un état mental, tel que le bonheur, le plaisir ou la satisfaction du désir procuré à une personne par la consommation (ou la possession) de biens et services (voir [1]). Cette approche préconise des politiques axées sur l'augmentation de la productivité, de l'emploi, etc., et donc du revenu, pour alléger la pauvreté. Cette approche était vue jusqu'à récemment comme l'unique façon d'étudier la pauvreté et elle domine sur toutes les autres. Selon Asselin et Dauphin. La Banque Mondiale est l'un des leaders parmi les organismes de développement qui prônent fortement ce concept.

Approche par les conditions de vie ou besoins de base

Selon cette approche, les pauvres sont ceux qui sont privés d'un certain ensemble de commodités de base perçues comme un préalable à l'atteinte d'une certaine qualité de vie (Asselin et Dauphin, 2000). Ces besoins incluent une alimentation adéquate, une bonne santé, savoir lire et écrire, disposer d'un logement adéquat, d'un bon habille-

ment. L'avantage de cette approche est qu'elle favorise des politiques ciblées dans la lutte contre la pauvreté.

B. FOKO TAGNE, F. NDEM et R. TCHAKOTE (2007) ont mené une étude de la pauvreté au Cameroun en utilisant cette approche. Ils ont construit un indicateur composite de pauvreté en utilisant la méthode factorielle.

L'un des principaux problèmes auquel se confronte cette approche, est la détermination même des besoins de base. Ces dernières sont souvent déterminées de manière exogène par le planificateur, l'analyste ou les experts (les nutritionnistes, physiologistes) indépendamment des perceptions des populations. Un autre problème est souvent la subjectivité dans le choix du seuil de pauvreté.

Approche par les capacités ou « capabilities »

Cette approche n'est basée ni sur l'utilité ni sur la satisfaction de besoins de base, mais sur des habiletés ou capacités humaines.

Selon Amartya Sen, le principal maître d'oeuvre de cette approche, la valeur de la vie d'une personne dépend en fait d'un ensemble de façons d'être et de faire (« being and doing »), qu'il regroupe sous le terme général de fonctionnements. Les fonctionnements sont donc des accomplissements, dont avoir de l'utilité est un exemple, alors que les capacités se réfèrent à la liberté de choisir parmi les divers fonctionnements. Est considéré comme pauvre par cette approche, une personne qui n'a pas les capacités d'atteindre un certain sous-ensemble de fonctionnements. C'est ainsi que Sen (1970) a toujours considéré la notion de liberté comme étant le principal vecteur de développement grâce, notamment, à la latitude offerte aux individus de choisir les modes de vie auxquels ils aspirent. La description des caractéristiques des biens d'un individu est donc souvent insuffisante pour évaluer le niveau de bien-être, car il faut aussi savoir ce que l'individu parvient à faire et être avec les biens et les caractéristiques dont il dispose, ce que Sen désigne par fonctionnements voir[7]. L'approche par les capacités se heurte à de nombreux problèmes de mesure qui ne sont par encore résolus. S'il est possible de mesurer certains fonctionnements déjà accomplis, il est beaucoup plus difficile d'apprécier l'étendue des fonctionnements non accomplis, mais que la personne peut éventuellement mettre en oeuvre si elle le décide.

1.3 Revue théorique de la classification

Concernant l'analyse multidimensionnelle, les analyses factorielles (Analyse en composantes principales, Analyse en composantes multiples) se révèlent impuissantes pour former des groupes lorsque le nombre d'unité statistique ou de variables devient très

important. Les représentations graphiques sur les plans principaux sont inutilisables, alors les méthodes de classification deviennent indispensables pour effectuer les regroupements homogènes.

Dans cette partie, nous présentons comme méthodes de classification la classification Ascendante hiérarchique

Classification Ascendante Hiérarchique

C'est une méthode de classification qui consiste à fusionner deux objets (ou individus) au sens d'une mesure de proximité de sorte que deux objets groupés à une étape le reste jusqu'au terme du processus de classification. Il s'agit ici, à partir des éléments terminaux, de former de petites classes ne comportant que des individus les plus semblables, et on continue le processus jusqu'à l'obtention d'une seule classe formée de tous les éléments.

Définition 1.2: Soit E un ensemble, E . On appelle dissimilarité sur E , toute application $d : E \times E \rightarrow \mathbb{R}_+$ vérifiant :

$$(i) : \forall x, y \in E, d(x, y) = 0 \Rightarrow x = y$$

$$(ii) : \forall x, y \in E, d(x, y) = d(y, x)$$

Remarque 1.3.1. Avant de lancer une quelconque classification, une fois les individus à classer sélectionnés et les variables qui serviront de critère de classification spécifiées, l'on devrait :

* définir au préalable une distance ou un indice d'écart entre paire d'individus (dissimilarité);

* Choisir une règle de calcul pour les distances entre les classes

* Déterminer un critère d'agrégation des individus dans les classes.

Critère d'agrégation

Pour définir le critère d'agrégation, on a souvent fait recours aux méthodes suivantes, étant donné $d(i, j)$ la distance ou dissimilarité entre deux individus.

a) La méthode du saut simple (single linkage) : c'est l'algorithme obtenu en utilisant l'indice d'agrégation. $D_1(A, B) = \min_{i \in A, j \in B} d(i, j)$

b) Méthode de diamètre (complète linkage) Cela revient à l'utilisation de l'indice d'agrégation $D_2(A,B) = \max d(i,j) \mid i \in A, j \in B$

c) Méthode de la moyenne L'on utilise ici l'indice d'agrégation

d) Méthode de Ward

L'indice d'agrégation utilisée est :

$$D_3(A,B) = \frac{P_A * P_B}{P_A + P_B} d^2(g_A, g_B)$$

où P_A et P_B sont les masses respectives de A et B; g_A et g_B sont les centres de gravités respectifs de A et B .

c) Algorithme de CAH

Les données étant l'ensemble des n objets (ou individus) à classer, $x_i, i = 1, \dots, n$ et les distances (ou dissimilarités) entre les objets $d(x_i, x_j)$, l'algorithme agglomératif est alors le suivant :

1) Initialisation avec n classe C_i formées chacune de 1 objet : $C_i = \{x_i\}$ et poser

$$D(C_i, C_j) = d(x_i, x_j)$$

2) Fusionner les deux classes C_p et C_q les plus proches pour former une nouvelle classe $C_k = C_p \cup C_q$ tels que $D(C_p, C_q) = \min D(C_i, C_j) \forall i, j$

3) calculer les distances entre la nouvelle classe C_k et les autres $D(C_k, C_i) \mid i \neq p, q$.

4) Itération: répéter n-1 fois les étapes 2 et 3 jusqu'à l'obtention d'une seule classe regroupant tous les objets.

5) STOP

A la fin de l'algorithme, on a construit une hiérarchie qui est définie par la fusion de 2 classes à chaque étape. Les indices de la hiérarchie sont donnés par la valeur $D(C_p, C_q)$ quand on fusionne C_p, C_q pour obtenir C_k .

d) Lecture et coupure d'un arbre hiérarchique

L'arbre hiérarchique ou Dendrogramme est une représentation graphique , sous forme d'arbre binaire des agrégations successives jusqu' à la réunion en une seule classe en une seule classe des individus. La hauteur d'une branche est proportionnelle à l'indice de dissemblance ou entre les deux objets regroupés .Dans le cas du saut de Ward , c'est la perte de la variance inter classe.

Le niveau d'un noeuds est définit comme la valeur de l'indice d'agrégation obtenu en agrégeant deux classes. Les agrégations successives des classes donnent les niveaux des noeuds, ceux-ci forment une suite croissante du fait qu'on agrège d'abord les classes ou

les individus les plus proches. On représente les niveaux de noeuds par un diagramme en bâton appelé histogramme des indices de niveaux.

La lecture d'un arbre hiérarchique se fait de la racine jusqu'aux éléments terminaux. Une coupure d'une branche de l'arbre donne les partitions, mais toutes les partitions ne sont pas bonnes. Pour cela, on utilise le test du coude sur l'histogramme des indices de niveau pour avoir la meilleure partition. Selon le test de coude, les partitions les plus intéressantes sont obtenues en coupant les noeuds de l'arbre dont les niveaux forment un saut brutal dans l'histogramme des indices de niveaux. L'application du critère du coude se justifie par le fait que le niveau du noeud que l'on coupe représente le coût de passage à la partition suivante :

d) Limites de l'algorithme CAH Les algorithmes de CAH présentent

une certaine complexité pratique. En effet, il faut à chaque étape balayer les $\frac{n(n-1)}{2}$ éléments de la matrice des distances afin d'en rechercher l'élément minimal (puisqu'on agrège les individus ou les classes les plus proches donc de distance minimale). Si plusieurs paires d'éléments réalisent le minimum, il doit falloir opérer un choix pour agréger les classes.

ELABORATION DES CLASSES SOCIO-ÉCONOMIQUES

Dans ce chapitre, après avoir présenté l'enquête ainsi que les variables d'étude, nous ferons une classification afin de constituer les catégories socio-économiques.

2.1 Source des données

Les données utilisées dans cette étude proviennent de l'enquête ECAM II réalisée par l'INS entre Septembre et Décembre 2001.

Présentation de ECAM II.

Objectif principal

L'objectif principal de l'enquête est de mettre en place les bases d'un dispositif permanent de suivi et d'évaluation des conditions de vie des ménages en général et du programme de réduction de la pauvreté en particulier. Cette enquête qui permet d'établir la situation de référence a été préparée dans cette perspective ; en même temps, elle permet d'évaluer les effets des programmes et politiques mis en oeuvre au Cameroun au cours des cinq dernières années

Structure des questionnaires de ECAMII

Le questionnaire de l'enquête s'articule autour d'un certain nombre de modules correspondant aux domaines d'étude qui sont eux-mêmes fonction des objectifs retenus. Le questionnaire de l'ECAMII est composé de 16 sections à savoir :

- (i) les renseignements généraux sur le ménage
- (ii) la composition et les caractéristiques de membres du ménages, (iii) la santé des membres du ménage

- (iv) le niveau d'instruction des membres du ménage âgés d'au moins 5 ans, (v) l'activité des membres du ménage
- (vi) la natalité, la mortalité et la fécondité
- (vii) l'anthropométrie et la couverture vaccinale pour les enfants âgés de moins de 35 mois, (viii), le logement et l'équipement du ménage
- (ix) les migrations du ménage, (x) l'accessibilité des infrastructures de base
- (xi) la perception des conditions de vie
- (xii) les entreprises familiales non agricoles
- (xiii) le patrimoine matériel et financier, (xiv) l'agriculture et les activités du monde rural
- (xv) les dépenses rétrospectives non alimentaire
- (xvi) et le relevé de compte des dépenses et acquisitions journalières du ménage.

Champ de l'enquête

Le champ de l'enquête est sur le plan géographique le territoire national. L'opération concerne l'ensemble des ménages ordinaires (par opposition aux ménages collectifs : internats, casernes, hôpitaux, couvents, etc.) résidant sur l'ensemble du territoire national à l'exclusion des membres du corps diplomatique et de leurs ménages.

L'unité statistique est le ménage ordinaire, défini comme un ensemble composé d'une ou de plusieurs personnes (unité socio-économique), ayant un lien de sang, de mariage ou non, vivant dans un ou plusieurs logements de la même concession (cet ensemble de logements constituant une unité d'habitation), mettant en commun tout ou partie de leurs ressources, pour subvenir aux dépenses courantes, prenant le plus souvent leurs repas en commun, et reconnaissant l'autorité d'une seule personne appelée chef de ménage (ou personne de référence).

Les unités d'observation sont en même temps le ménage (logement, habitat, dépenses Indivisibles du ménage, etc.) et les individus (caractéristiques démographiques, dépenses Individuelles, etc.). Les analyses vont porter essentiellement sur les ménages typés selon Certaines de leurs caractéristiques (lieu de résidence, situation d'activité du chef, composition, etc.), mais également sur les individus.

Base de sondage

La base de sondage utilisée à l'ECAMII est constituée des zones de dénombrement (ZD) du recensement de la population qui date d'avril 1987. Elle est donc très vieille,

particulièrement pour le milieu urbain qui change très rapidement. Pour cela, une mise à jour des zones de dénombrement a été faite.

Plan d'échantillonnage

Le plan d'échantillonnage de cette enquête dérive d'un sondage stratifié. Douala et Yaoundé ont été définies comme des strates à part. Dans chacune des dix provinces, on distingue deux strates, à savoir une strate rurale et une strate urbaine. L'enquête a donc travaillé sur un total de 22 strates dont 10 rurales et 12 urbaines. Les strates urbaines sont à leur tour divisées en sous strates, à savoir les villes de 50 000 habitants au moins (strates urbaines) et celles de 10 000 à moins de 50 000 habitants (strates semi urbaines).

2.1.1 Coefficient d'extrapolation

Les coefficients d'extrapolation dérivent directement du mode du tirage. Le coefficient d'extrapolation d'une zone de dénombrement (ZD) k est l'inverse de la probabilité d'un ménage de la ZD concernée. On a :

$$\star \text{ Pour les villes de Douala et Yaoundé: } C_{1k} = \frac{N_{ij}d_{ijk}}{n_{ij}e_{ijk}}$$

$$\star \text{ Pour les zones urbaines (grandes villes): } C_{2k} = \frac{Z_{ik}d_{ik}}{z_{ik}e_{ik}}$$

$$\star \text{ Dans les zones semi-urbaines et rurales: } C_{3k} = \frac{a_i M_{ij} Z_{ijk} d_{ijk}}{m_{ij} z_{ijk} e_{ijk}}$$

où

N_{ij} = nombre total de ZD dans l'arrondissement j de la ville i ;

n_{ij} = nombre de ZD tirées dans l'arrondissement j de la ville i ;

e_{ijk} = nombre de ménages effectivement tirés et enquêtés dans la ZD k de l'arrondissement j de la strate i ;

e_{ij} = nombre de ménages effectivement tirés et enquêtés dans la ZD urbaine k ;

d_{ijk} = nombre de ménages dénombrés dans la ZD k de l'arrondissement j de la strate i ;

d_{ik} = nombre total de ménages dénombrés dans la ZD urbaine k ;

z_{ijk} = nombre de ZD tirées dans la ville j de la province i ; Z_{ijk} = nombre total de ZD dans la ville j de la province i , $i = 1$ à 10 ;

z_{ij} = nombre de ZD urbaines tirées dans la province i ; Z_{ij} = nombre total de ZD urbaines dans la province i , $i = 1$ à 10 ;

a_i = nombre d'arrondissements sélectionnés pour l'enquête dans la strate i d'après le découpage administratif en vigueur en 1987 ;

m_{ij} = nombre de ménages dénombrés dans l'arrondissement (ville) j de la strate i en 1987 ;

M_{ij} = nombre total de ménages dénombrés dans l'ensemble des arrondissements (villes) de la strate i en 1987.

Présentation des données

La base de donnée est sous forme d'un fichier de type SPSS. Le fichier d'étude est le fichier ménage de l'ECAMII qui comporte les caractéristiques de 10992 ménages. La personne de référence dans le ménage est le chef de ménage. Le tableau de donnée comporte 10992 lignes (correspondant aux ménages) et 228 variables.

Pour limiter l'arbitraire dans le choix des attributs dans notre catégorisation, et compte tenu du fait qu'un accent est mis ici sur les conditions de vie et la pauvreté, nous avons tenu compte.

- Des perceptions de la population synthétisées dans le DSRP car il y est mentionné (p.24) que les populations estiment que l'indisponibilité des services essentiels les maintient dans un état de pauvreté. Pour définir le bien-être, la population camerounaise fait référence principalement à la satisfaction des besoins de base tels que le logement, l'éducation, l'habillement, la nourriture, la santé, l'hygiène et l'assainissement. Donc ces critères jouent un rôle inconsiderable dans la discrimination des individus.
- des choix opérées dans d'autres études par d'autres auteurs. En effet, dans le cadre de l'étude de la pauvreté relative aux conditions de vie en France, Madior Fall et Daniel Verger(2005) déclare que ce qu'il faut regarder, dans l'étude de la population et de la pauvreté, c'est la proportion des individus qui, au travers de questions d'opinion, se déclarent dans l'impossibilité, pour des raisons pécuniaires et non pour des raisons de goûts de s'assurer une consommation décente en matière d'alimentation, de logement, d'habillement, soit dans les registres usuellement qualifiés de fondamentaux. Donc la possession des biens privés est un élément à prendre en compte pour comprendre le niveau de bien-être (confort du logement, possession des biens durables, etc.,)

Compte tenu de ce qui précède, les variables retenues pour l'analyse sont issues des sections du questionnaire relatives à la santé, au niveau d'instruction, au logement et équipement du ménage, à l'accessibilité aux infrastructures de base. Sont concernées également dans l'étude les variables démographiques issues des sections renseignements généraux du ménage, composition et caractéristiques des membres du ménage, les variables liées à l'emploi. Pour notre étude, nous avons alors retenu 31 variables. Le tableau présenté à l'annexe nous donne un récapitulatif des variables et leurs modalités.

2.2 Caractéristiques de la pauvreté monétaire

Dans cette partie nous utiliserons un indicateur de niveau de vie construit par l'INS avec la base de donnée ECAMII pour mesurer la pauvreté monétaire . Avec cet indicateur, nous faisons une description de vos variables retenues dans l'étude, dans le but de pouvoir donner un sens à divers groupes obtenus après la classification, et aussi dans l'identification de la classe des personnes jugées pauvres. Pour sa construction, cet indicateur de vie a été approché par la consommation finale annuelle des ménages qui dans une certaine mesure, reflète mieux le niveau de vie des ménages

comparativement aux revenus qui est difficile à mesurer. La consommation finale du ménage construite à cet effet inclut quatre éléments distincts dont on peut citer la consommation monétaire, l'autoconsommation, les transferts en nature reçus d'autres ménages, le loyer imputé des ménages propriétaires de leur logement ou logés gratuitement.

Cet indicateur de niveau de vie a débouché sur un seuil de pauvreté minimum 232 547 francs et un seuil maximum de 345.535 francs CFA par an. Ainsi, les ménages dont la consommation annuelle par équivalent adulte est inférieure à 232 547 sont considérés comme « pauvre ». Ceux dont les dépenses sont comprises entre 232.547 FCFA et 345.535 FCFA sont considérés comme « Intermédiaires ». Enfin, ceux dont les dépenses sont supérieures ou égales à 345.535 FCFA sont « Non pauvres » . Les personnes pauvres sont celles qui vivent dans les ménages pauvres. Les non pauvres dans notre étude regroupent intermédiaires et riches.

Caractéristiques socio-démographiques et pauvreté monétaire

D'importantes inégalités selon le sexe, l'âge, le statut matrimonial, le niveau d'instruction et le statut de pauvreté selon l'approche monétaires sont observés.

La plupart des ménages ont à leur tête un homme. En effet, 76% des ménages ont pour chef un homme. Parmi les ménages pauvres , 21.3% sont dirigés par les femmes. Par contre chez les ménages non pauvres , 25,5% ont pour chef une femme. Par rapport au niveau de vie , la pauvreté semble plus touchée les ménages dirigés par les hommes que ceux dirigés par les femmes.

Par rapport au statut matrimonial, La proportion des ménages dont le chef est marié monogame est la plus importante(44,7%). L'on dénombre aussi ici 17.4% de célibataire, 15.7% de polygame, 11.1% de veuf ou veuve, 4.5% de divorcé ou séparé et 6.6% de chef de ménage vivant en union libre .Le tableau I nous montre que parmi les ménages pauvres 49,9% sont dirigées par les monogames. L'on remarque ici que, par rapport au niveau, les ménages dirigés par les célibataires semblent les moins touchés par la pauvreté .

Age de chef de ménage : la moyenne est 43 ans. Pour avoir une bonne lisibilité des

		niveau de vie			
		pauvres		non pauvres	
		Effectif	% colonne	Effectif	% colonne
sexe	masculin	2174	78,7%	6099	74,5%
	feminin	587	21,3%	2083	25,5%
age	moins de 29 ans	367	13,3%	1894	23,1%
	30 à 44 ans	948	34,3%	3189	39,0%
	45 à 59 ans	840	30,4%	1890	23,1%
	plus de 60 ans	606	21,9%	1209	14,8%
statutm	celibataire	236	8,5%	1672	20,4%
	marie mono-game	1365	49,4%	3526	43,1%
	marie poly-game	603	21,8%	1117	13,7%
	veuf/veuve	324	11,7%	892	10,9%
	divorce/separe	94	3,4%	393	4,8%
	en union libre	139	5,0%	582	7,1%
instruct	non scolarise	1149	41,6%	1793	21,9%
	primaire	1075	38,9%	2383	29,1%
	secondaire 1er cycle	395	14,3%	1870	22,9%
	secondaire 2nd cycle	112	4,1%	1296	15,8%
	superieur	30	1,1%	840	10,3%
Ensemble		2761	100,0%	8182	100,0%

TAB. 2.1 – Caractéristiques socio-démographiques et niveau de vie.

groupes suivant l'âge, les groupes d'âge ont été formés à savoir age moins de 29 ans, 30 à 44 ans, 45 à 59 ans et 60 ans et plus. La majorité des chefs de ménages ont pour age compris entre 30 et 44 ans (37.8%). Les personnes âgés (60 ans et plus) sont les plus touchées par la pauvreté contrairement au ménages jeunes où la plus part des chef de ménages sont non pauvres

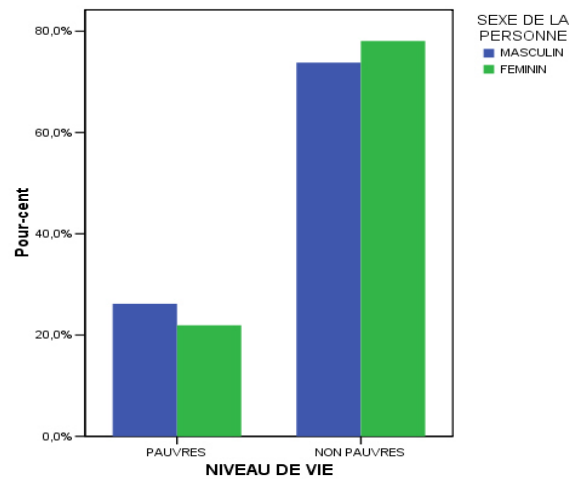


FIG. 2.1 – Niveau de vie et sexe du chef de ménage.

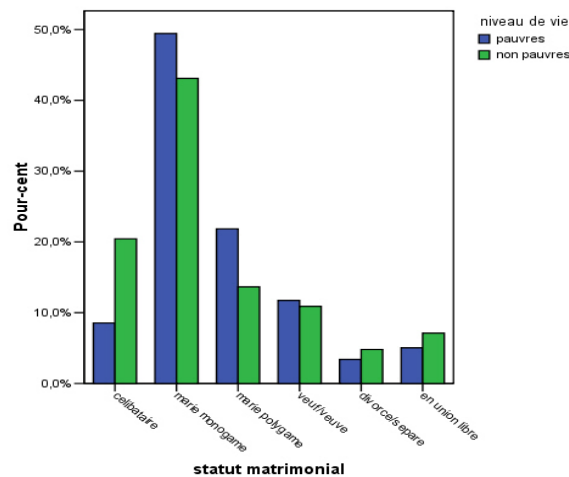


FIG. 2.2 – Niveau de vie et statut matrimonial.

Par rapport au niveau d'instruction, les chefs de ménages ayant le niveau d'instruction primaire sont en majorité (31.6%), suivie du niveau non scolarisé(26.9%), du niveau secondaire 1er cycle (20.7), du secondaire 2e cycle(12.9%) et enfin du supérieur(8%). A l'observation du tableau TAB.2.1 et de la figure FIG.2.3, il ressort que la pauvreté sévit beaucoup plus dans les ménages où le chef est non scolarisé ou a le niveau de primaire. Par contre, les chefs de ménages ayant le niveau du secondaire et du supérieur sont en général les non pauvres.

Disparités régionales de la pauvreté monétaire

L'enquête ECAMII, compte tenu de son objectif principal, à savoir mesurer la pauvreté monétaire et les conditions de vie, s'est déroulée dans toutes les régions du pays en distinguant également les types de milieu .Le tableau 2.2 nous donne quelques

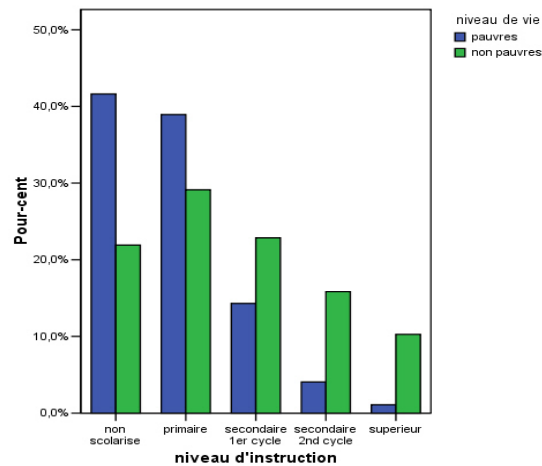


FIG. 2.3 – Niveau d'instruction et niveau de vie.

caractéristiques des ménages suivant les régions et le milieu de résidence. Lorsque l'on répartit les ménages de chaque région selon la variable niveau de vie l'on observe que la pauvreté selon l'approche monétaire épargne relativement les deux grandes métropoles camerounaises. Seulement 3,5% et 3,1% des ménages vivant respectivement à Douala et Yaoundé sont pauvres. Aussi les provinces du Centre (Yaoundé non compris), Littoral (Douala non compris) et le Sud abritent des ménages un peu moins pauvres. A l'opposé parmi les ménages pauvres, ceux se trouvant dans la région de l'Extrême-Nord sont le plus nombreux (18,4%).

Le tableau 2.2 nous montre également que la pauvreté monétaire affecte beaucoup plus les ménages de la zone rurale. En effet, 57,8% des ménages pauvres s'y retrouvent contre 24,4% et 17,8% respectivement pour les milieux urbains et semi-urbains.

caractéristiques de ménages suivant l'habitat

Type de logement

Le type de logement dans lequel vivent les ménages a été relevé. Il ressort que 45,8% des ménages au niveau national logent dans une maison isolée, définie comme une maison d'un seul tenant dans une cour. Ce type de logement est surtout fréquent en zone rurale où les espaces existent concomitamment avec une densité relativement faible de la population. Au niveau régional, on les rencontre plus fréquemment dans la partie Sud du pays, notamment au Centre, au Littoral, au Sud, au Nord-Ouest, à l'Ouest et à l'Est. Par rapport au niveau de vie, les pauvres sont relativement plus nombreux que les non pauvres à occuper ce type de logement

Un ménage sur trois vit dans une maison à plusieurs logements, définie comme une grande maison divisée en plusieurs chambres, pièces ou logements indépendants. Ce type d'habitat est surtout fréquent en ville où la rareté des logements et la cherté des

		niveau de vie			
		pauvres		non pauvres	
		Effectif	% colonne	Effectif	% colonne
douala/yaounde/ province	douala	96	3,5%	1016	12,4%
	yaounde	86	3,1%	1009	12,3%
	adamaoua	285	10,3%	462	5,6%
	centre	233	8,4%	631	7,7%
	est	181	6,6%	557	6,8%
	extreme-nord	508	18,4%	811	9,9%
	littoral	183	6,6%	549	6,7%
	nord	267	9,7%	587	7,2%
	nord-ouest	325	11,8%	556	6,8%
	ouest	255	9,2%	819	10,0%
	sud	147	5,3%	614	7,5%
	sud-ouest	195	7,1%	571	7,0%
milieu de resi- dence	urbain	673	24,4%	4276	52,3%
	semi-urbain	491	17,8%	1637	20,0%
	rural	1597	57,8%	2269	27,7%
Total		2761	100,0%	8182	100,0%

TAB. 2.2 – Région, milieu de résidence et niveau de vie.

loyers favorisent cette sorte de regroupement. Ces dernières abritent également plus de ménages non pauvres.

Un autre type de logement dans lequel une frange importante de ménages habitent est la concession ou le saré. C'est un ensemble de bâtiments ou de constructions séparés dans l'espace et disposés généralement derrière ou autour d'une construction principale, à usage d'habitation et abritant un ménage ou une famille. Près de 16.7% de ménages au niveau national logent dans ce type d'habitat qu'on rencontre surtout dans la partie septentrionale du pays et quelque peu à l'Est et au Nord-Ouest. Ce mode de construction est attaché à la culture des peuples du Grand Nord d'où le nom local de saré qui lui est associé et qui est devenu une terminologie nationale. Les concessions ou saré se rencontrent surtout en milieu rural et sont beaucoup plus habitées par des ménages pauvres.

Les villas et les immeubles à appartements sont le dernier type de logement en-

registré au cours de l'enquête ; ils ont la particularité d'être bâtis exclusivement en matériaux définitifs modernes et sont donc d'un haut standing. Dans l'ensemble du pays, ces types d'habitat sont habités 2.6% de ménages. On les rencontre surtout dans le milieu urbain et tout particulièrement dans les deux grandes métropoles (Yaoundé et Douala) rapport au niveau de vie, les ménages s'y abritant sont surtout non pauvres.

Type des lieux d'aisance

Il a été demandé aux ménages au cours de l'enquête de dire quel type de lieu d'aisance ils utilisent. Les résultats révèlent sans surprise que c'est à Douala et à Yaoundé que l'on trouve le plus les WC avec chasse eau, environ 22% de ménages de ces deux villes en disposent. Dans les autres villes, seulement un ménage sur 10 possède un WC avec chasse eau. Par ailleurs, les provinces du Nord et de l'Extrême Nord sont celles qui ont les proportions les plus élevées de ménages qui n'ont pas du tout de WC.

Lorsque l'on considère le statut de pauvreté, environ deux ménages pauvres sur trois n'ont comme lieu d'aisance que des latrines non aménagées. C'est en zone rurale que l'on rencontre les lieux d'aisance les plus précaires. Chez les non pauvres par contre, 12.5% de ménages ont des WC avec chasse eau, 47.8% ont des latrines aménagées et 36.3% ont des latrines non aménagées

Les autres caractéristiques des ménages à savoir les matériaux du sol, les matériaux du murs le mode d'approvisionnement en eau de boisson sont observées dans le TABLEAU VII à l'annexe.

caractéristiques des ménages suivant l'activité du chef

Le tableau 2.3 nous donne certaines caractéristiques du chef de ménages suivant son statut professionnel et suivant le statut de pauvreté monétaire ou niveau. On ici que la moitié des chefs de ménages exercent dans les entreprises informelles agricoles (35.1%) et dans les entreprises informelles non agricoles (24.8%). C'est dans ces secteurs institutionnels où l'on retrouve la plus des ménages pauvres. En effet parmi les ménages pauvres, ceux entreprises informelles agricoles representes 58.8%. A l'opposé les ménages du secteur publique et des entreprises privées formelles sont les moins touchés par la pauvreté.

2.3 Catégorisation des classes socio-économiques

Il est question pour nous de former les groupes homogènes des personnes à partir des variables sélectionnées pour l'étude. Dès lors la classification devient utile. Elle a pour objet la description d'un ensemble d'individus ou d'objets caractérisés par un ensemble d'attributs, à l'aide de leur regroupement en classe. Ces classes sont établies de telle

		niveau de vie			
		pauvres		non pauvres	
		Effectif	% colonne	Effectif	% colonne
statut d'occupation	cadres	45	1,6%	879	10,7%
	employés qualif,	96	3,5%	901	11,0%
	autres salariés	277	10,0%	1155	14,1%
	patrons	46	1,7%	235	2,9%
	propre compte	1951	70,7%	3760	46,0%
	chomeurs/autr, n, sal,	346	12,5%	1252	15,3%
secteur institutionnel	adm, publique	80	2,9%	1017	12,4%
	entse publique	45	1,6%	190	2,3%
	entse privée form,	157	5,7%	1433	17,5%
	entse inf, agric,	1612	58,4%	2229	27,2%
	entse inf, n, agric	548	19,8%	2163	26,4%
	chomeur	319	11,6%	1150	14,1%
Total		2761	100,0%	8182	100,0%

TAB. 2.3 – Niveau de vie et statut professionnel.

sorte que les objets appartenant à la même classe soient les plus semblables possibles, et que les objets appartenant à deux classes différentes soient les plus dissemblables possibles.

2.3.1 Présentation de la méthode

La méthode de classification utilisée est la Classification Hiérarchique Ascendante . Le logiciel utilisé à ce niveau est le logiciel SPAD. Chaque individu est caractérisé par les valeurs prises sur un ensemble de variables. Ces variables sont le plus souvent de nature différente. On distinguera, en premier lieu, les variables selon le rôle qu'elles jouent dans la démarche de la classification: on distinguera les variables actives et les variables passives ou illustratives de la classification. Les variables actives aussi appelées critères de classification servent à constituer les groupes alors que les variables passives vont servir, dans un deuxième temps, à expliquer ces groupes . Ici nous ferons notre classification avec les variables actives liées aux conditions de vie et à l'emploi. Les différentes classes seront expliquées plus tard avec la régression . En second lieu, les

variables diffèrent selon la nature de la mesure utilisée. Elles peuvent être qualitatives ou quantitatives.

Les variables (actives) utilisées dans la classification sont : **Type de logement**

Type des lieux d'aisance

Matériaux du sol

Matériaux du mur

Mode d'approvisionnement en eau de boisson

source d'énergie pour la cuisine

matériau du toit

source d'éclairage du logement

statut d'occupation (professionnel)

Victime d'un handicap

Appréciation de son état de santé

secteur institutionnel

possession d'appareils (poste radio, réfrigérateur, congélateur, ventilateur, cuisinière, » réchaud a gaz, réchaud a pétrole, véhicule ,téléviseur, fer a repasser).

Avec la base de donnée nous appliquons la méthode d'Analyse des Correspondance Multiples (ACM). Les coordonnées factorielles créées par cette analyse nous permettent de procéder à la classification hiérarchique. Les classifications issues des Analyses de Correspondance multiples aboutissent à des typologies dont les classes sont très hétérogènes entre elles et leurs éléments sont très homogènes.

La méthode utilisée est celle de WARD. Cette méthode repose sur le calcul de l'inertie entre deux classes ou individus. En effet, lorsqu'on agrège deux individus ou deux classes, on perd de l'information. On cherche à obtenir à chaque pas un minimum local de l'inertie intra classe ou un maximum de l'inertie interclasse. En d'autres termes, on agrègera de préférence les individus ou les classes qui font le moins varier l'inertie intra classe.

Pour mesurer statistiquement la dépendance ou la liaison entre les variables et les classes nous utiliserons ici la valeur test (V-test) . Cette valeur test permet de trier par ordre d'importance les modalités et les variables les plus caractéristiques d'une classe. Une variable ou une modalité sera d'autant plus intéressante pour une classe que la valeur test sera significative (c'est à dire supérieur à 2 en valeur absolue pour le seuil usuel de 5%). Un V-test grand et positif signifie modalité sur représenté, par contre un V-test grand et négatif signifie modalité sous représenté. Pour la caractérisation des classes nous avons aussi besoin de connaître la signification de CLA/MOD et MOD/CLA :

- CLA/MOD : c'est le nombre d'individu possédant la modalité dans la classe divisé

par le nombre total d'individu possédant la modalité. En d'autre terme c'est le pourcentage de ceux qui possède la modalité et qui sont la classe. Si $CLA/MOD = 100$ (si un individu possède la ,on sait qu'il est dans la classe).

- MOD/CLA : c'est le nombre d'individus possédant la modalité dans la classe divisé par l'effectif de la classe. $MOD/CLA = 100\%$ signifie que tous les individus de la classe possèdent la modalité

2.3.2 Choix du nombre de classes

Le choix du nombre de classes se fait à l'aide du test du coude. En effet, à travers l'observation de la courbe des indices de niveau et du dendrogramme, on détecte le nœud ayant le plus grand saut.

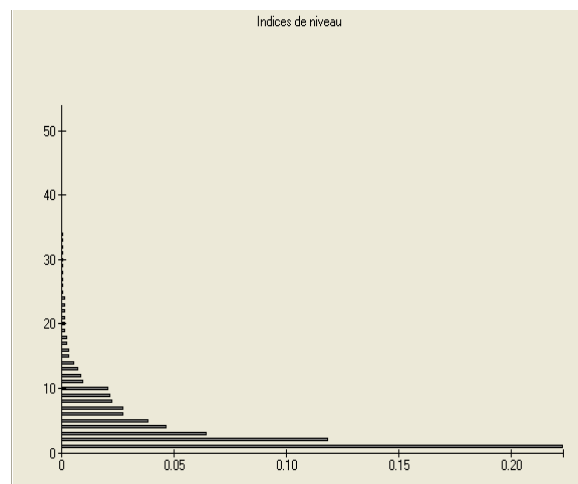


FIG. 2.4 – Courbe des indices de niveau.

Le dendrogramme nous indique l'ordre dans lequel les agrégations successives ont été opérées. Il nous indique également la valeur de l'indice d'agrégation à chaque niveau d'agrégation. Il est généralement pertinent d'effectuer la coupure après les agrégations correspondant à des valeurs peu élevées de l'indice et avant les agrégations correspondant à des valeurs élevées. En coupant l'arbre au niveau d'un saut important de cet indice, on peut espérer obtenir une partition de bonne qualité car les individus regroupés en-dessous de la coupure étaient proches, et ceux regroupés après la coupure sont éloignés. L'histogramme des indices de niveau et l'arbre de classification ou Dendrogramme obtenu nous suggèrent un regroupement optimal en 3 classes Car on note une forte rupture entre le 2e indice le plus fort et le 3e.

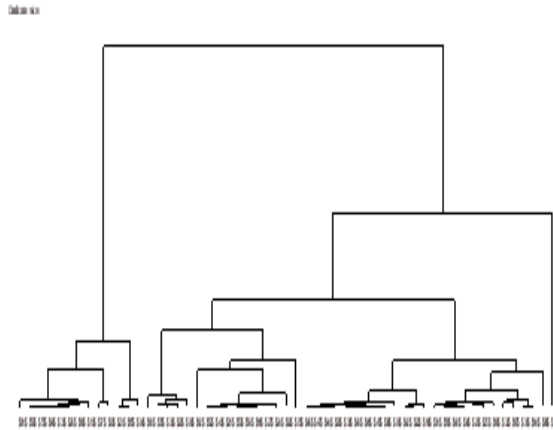


FIG. 2.5 – Dendrogramme.

2.3.3 Description des classes

Ayant fait la classification en 3 classes comme suggéré précédemment, à l'aide des résultats présentés à l'annexe (TABLEAU VI: résultats de la classification) , nous constituons le tableau ci-dessous donnant la répartition des classes en effectif.

Classe	Effectif	Pourcentage (%)
Classe1	4736	43.27
Classe2	4847	44.30
Classe3	1360	12.43
Total	10943	100

TAB. 2.4 – Effectif des classes.

Source: Elaborer par nous à partir de ECAMII

Description de la classe1

Cette classe représente 43.27% de notre échantillon soit un effectif de 4736 ménages Ici, les chefs de ménages travaillent en majorité dans les entreprises informelles agricoles. En effet 93.58% des personnes exerçant dans ce secteur s'y retrouve . C'est ainsi que parmi les individus de cette , 79.13% exercent dans ce domaine .

Dans cette classe, nous notons une précarité de matériaux des habitats. En effet, 83.43% des ménages ont le bois ou terre comme matériaux du sol, 99.78% des toits en nattes ou chaumes s'y retrouvent, 81.47% des maison construites en terre ou brique simple se retrouvent également dans cette classe. Dans cette classe les type de logement les courant sont la concession (ou saré) et la maison isolée .

On note également ici une rareté d'appareil d'équipement . En effet, dans cette classe 99,67% de ménages n'ont pas de téléviseur, 99.66% n'ont pas de ventilateur, 84.50% ne possède pas de fer à repasser, 63.11% n'ont pas un poste radio. Les équipements comme le réfrigérateur, le téléphone, la cuisinière, le véhicule, le congélateur sont presque inexistantes. Les ménages utilisent le pétrole et le bois respectivement comme source d'éclairage du logement (79.5%) et source d'énergie pour la cuisine (98.19%). Leur principal mode d'approvisionnement en eau de boisson est la rivière ou le marigot. Dans cette classe 71.65% ont pour type de lieux d'aisance les latrines non aménagées. Au vu des caractéristiques des ménages de la classe1, caractéristiques liées aux mauvaises conditions de vie, l'on peut assimiler cette classe comme étant celle des ménages pauvres.

La classe1 étant formée , nous allons expliquer l'appartenance à cette classe à partir des variables qui n'ont pas servies à la mise sur pied de cet indicateur, par les variables socio-démographiques.

Le tableau 2.5 ci-dessous explique l'appartenance à la classe1 suivant le milieu de residence. L'on peut remarquer ici que les ménages de la classe1 vivent pour la plupart dans une zone rurale.

		milieu			Total
		Urbain	semi-urbain	rural	
classe1	non	3450	1013	1633	6096
	oui	416	1115	3316	4847
Total		3866	2128	4949	10943

TAB. 2.5 – Tableau croisant classe1 et milieu de residence.

		niveau d'instruction				
		non scolaire	primaire	secondaire 1er cycle	secondaire 2er cycle	supérieur
classe1	non	713	1698	1673	1271	852
	oui	2229	1760	592	137	18
Total		2942	3458	2265	1408	870

TAB. 2.6 – Tableau croisant classe1 et niveau d'instruction du chef de ménage.

Description de la classe2

Cette classe représente 44.30% de l'effectif total. Contrairement à la classe1, les matériaux de maisons sont le plus souvent en matériaux définitifs. Dans cette classe 90.89% des sols sont en ciment ou carreau. Presque toutes les toitures sont en tole ou en

		statut matrimonial					
		celibataire	marie mo- nogame	marie poly- game	veuf/ veuve	divorce/ séparé	en union libre
classe1	non	1433	2754	815	467	235	503
	oui	475	2137	905	749	252	218
Total		1908	4891	1720	1216	487	721

TAB. 2.7 – Tableau croisant classe1 et le statut matrimonial.

		catholique	protestant	autres chrétien	musulmans	animistes	autres
classe1	non	2821	1845	237	1059	102	143
	oui	1578	1169	169	1307	439	74
Total		4399	3014	406	2366	541	217

TAB. 2.8 – Tableau croisant classe1 et religion.

		sexe		Total
		masculin	féminin	
classe1	non	4737	1470	6207
	oui	3536	1200	4736
Total		8273	2670	10943

TAB. 2.9 – Tableau croisant classe1 et le sexe du chef de ménage.

		taille ménage			
		au plus 2 enfants	de 3 à 5 enfants	de 6 à 8 enfants	9 enfants et plus
classe1	non	1581	2091	1564	971
	oui	1228	1747	1131	630
Total		2809	3838	2695	1601

TAB. 2.10 – Tableau croisant classe1 et taille du ménage.

tuile. Le principal mode d'approvisionnement en eau est le robinet commun. Les latrines aménagées sont le type des lieux d'aisance le plus utilisé. On note ici la possession de certains équipements tels que le poste radio, le réchaud à gaz, le fer à repasser, le ventilateur, le téléviseur bien que les équipements tels que la cuisinière, le véhicule le téléphone ne sont presque pas à leur portée. Les chefs de ménages exercent beaucoup plus dans le secteur d'entreprise formelle non agricole et d'entre privée formelle. Ici les ménages occupent les maisons à plusieurs logements.(53.69%), ont l'électricité comme principale source d'éclairage, 89.66% des ménages utilisent le pétrole comme source

d'énergie pour la cuisine. Il est également à noter que 64.71%. Il est également à noter que 64.71% des maisons ayant la planche ou le carabot comme matériaux des murs se retrouvent aussi dans cette Classe. L'on peut les ménages qui s'y retrouvent ont un niveau de vie moyen.

A la suite de la description de cette classe, nous présentons des tableaux ci-dessous dans lesquels la variable classe est expliquée d'autres variables n'ayant pas participer à l création de classe.

		niveau d'instruction				
		non scola- risé	primaire	secondaire 1er cycle	secondaire 2er cycle	supérieur
classe2	non	2273	1931	863	510	519
	oui	669	1527	1402	898	351
Total		2942	3458	2265	1408	870

TAB. 2.11 – Tableau croisant classe2 et niveau d'instruction du chef de ménage.

		statut matrimonial					
		celibataire	marie mo- nogame	marie po- lygame	veuf/ veuve	divorce/ séparé	en union libre
classe2	non	651	2918	1109	843	300	275
	oui	1257	1973	611	373	187	446
Total		1908	4891	1720	1216	487	721

TAB. 2.12 – Tableau croisant classe2 et le statut matrimonial.

		religion					
		catholique	protestant	autres chrétien	musulmans	animistes	autres
classe2	non	2212	1645	228	1448	454	109
	oui	2187	1369	178	918	87	108
Total		4399	3014	406	2366	541	217

TAB. 2.13 – Tableau croisant classe2 et religion.

Description de la classe3

La classe3 représente 12.44% de notre échantillon. Comme dans la classe2, les matériaux d'habitat sont en matériaux définitifs. Les murs des maison sont en béton, parpaing ou pierre de taille (86.72%), le sol est en ciment carreau(98.40%), le toit

		sexe		Total
		masculin	féminin	
classe2	non	4609	1487	6096
	oui	3664	1183	4847
Total		8273	2670	10943

TAB. 2.14 – Tableau croisant classe2 et taille du ménage.

		age				Total
		moins de 29 ans	30 à 44 ans	45 à 59 ans	60 et plus	
classe2	non	904	2126	1723	1343	6096
	oui	1357	2011	1007	472	4847
Total		2261	4137	2730	1815	10943

TAB. 2.15 – Tableau croisant classe2 et age.

est tôle ou en tuile (89.95%) Il est à noter que 82.70% des ménages utilisant le WC avec chasse eau se retrouvent dans cette classe. Ici également, presque tous les ménages possèdent tous les équipement. Contrairement à la classe2 on note ici la présence d'équipement de grande valeur : se retrouvent dans cette classe 74.59% des ménages possédant une cuisinière, 71.62% possédant un congélateur, 68.62% possédant un téléphone, 81.40% possédant une voiture. La principale source d'énergie pour la cuisine est le gaz ou l'électricité. Les ménages utilisent comme source d'éclairage du logement l'électricité abonnée à la SONEL.. Les chefs de ménages sont le plus souvent des cadres en ce qui concerne le statut professionnel et exercent dans l'administration publique et entreprise privée formelle. Au vu des caractéristiques de cette classe l'on peut dire que les ménages de classes ont un très bon niveau de vie. On peut parler de classe des riches.

		milieu			Total
		Urbain	semi-urbain	rural	
classe3	non	3853	1896	3834	9583
	oui	1096	232	32	1360
Total		4949	2128	3866	10943

TAB. 2.16 – Tableau croisant classe3 et milieude residence.

Nous donnons ici le graphique des individus sur le plan factoriel des deux premiers axes. Ce graphique nous montre bien le regroupement des individus suivant trois groupes.

		niveau d'instruction					Total
		non scolarisé	primaire	secondaire 1er cycle	secondaire 2er cycle	supérieur	
classe3	non	2898	3287	1994	1035	369	9583
	oui	44	171	271	373	501	1360
Total		2942	3458	2265	1408	870	10943

TAB. 2.17 – Tableau croisant classe3 et niveau d'instruction du chef de ménage.

		statut matrimonial						Total
		celibataire	marie mo- nogame	marie po- lygame	veuf/veuve	divorce/séparé	célibataire libre	
classe3	non	1732	4110	1516	1122	439	664	9583
	oui	176	781	204	94	48	57	1360
Total		1908	4891	1720	1216	487	721	10943

TAB. 2.18 – Tableau croisant classe3 et le statut matrimonial.

		religion					Total	
		catholique	protestant	autres chrétien	musulmans	animistes		autres
classe3	non	3765	2538	347	2225	526	182	9583
	oui	634	476	59	141	15	35	1360
Total		4399	3014	406	2366	541	217	10943

TAB. 2.19 – Tableau croisant classe3 et l'âge du chef de ménage.

		sexe		Total
		masculin	féminin	
classe3	non	7200	2383	9583
	oui	1073	287	1360
Total		8273	2670	10943

TAB. 2.20 – Tableau croisant classe3 et le sexe du chef de ménage.

		taille ménage				Total
		au plus 2 enfants	de 3 à 5 enfants	de 6 à 8 enfants	9 enfants et plus	
classe3	non	2658	3380	2237	1308	9583
	oui	151	458	458	293	1360
Total		2809	3838	2695	1601	10943

TAB. 2.21 – Tableau croisant classe3 et l'âge du chef de ménage.

		age				Total
		moins de 29 ans	30 à 44 ans	45 à 59 ans	60 et plus	
classe3	non	2137	3463	2303	1680	9583
	oui	124	674	427	135	1360
Total		2261	4137	2730	1815	10943

TAB. 2.22 – Tableau croisant classe3 et l'age du chef de ménage.

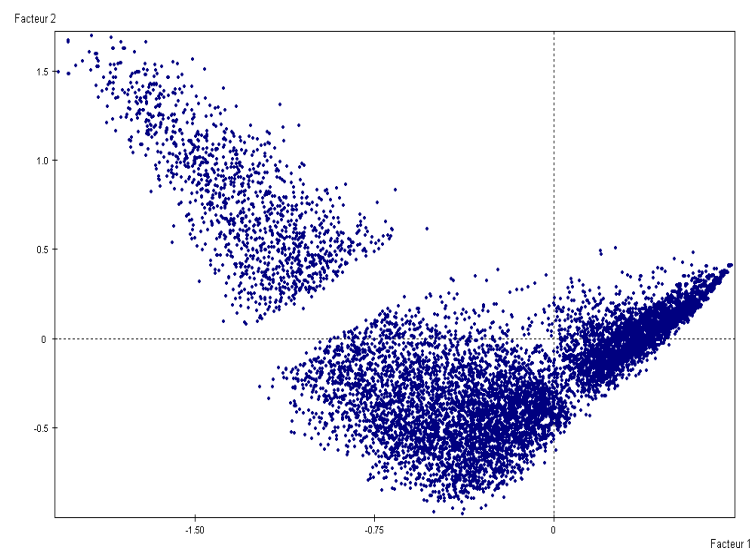


FIG. 2.6 – Repartition des individus sur le 1er plan factoriel .

MODÉLISATION

Dans ce chapitre, nous ferons une régression logistique afin d'expliquer l'appartenance d'un individu à une classe donnée.

3.1 Etude théorique

3.1.1 Régression logistique

La régression est une méthode à partir de laquelle on cherche à modéliser l'information contenue dans une base de données au moyen d'une équation algébrique. L'on cherche ainsi à faire passer une courbe mathématique par un ensemble de points expérimentaux afin d'appréhender l'évolution du phénomène étudié. L'évolution de la variable observée est alors expliquée à partir d'un ensemble de variables dites explicatives.

Dans le cas d'une régression linéaire multiple, on cherche une relation entre une variable continue (dépendante, expliquée ou endogène) notée Y en fonction des observations (variables indépendantes, explicatives, exogènes) notées X_i , $i = 1, \dots, k$ et du type quantitatif soit:

$$Y = a_0 + AX + \varepsilon = a_0 + \sum_{i=1}^k a_i X_i + \varepsilon,$$

avec les notations $A = (a_1, \dots, a_k)$; $X = (X_1, \dots, X_k)^T$: où a_0 et A sont des paramètres à estimer et ε le terme d'erreur aléatoire, qui doit suivre une loi normale $\mathcal{N}(0, \sigma^2)$. La régression linéaire cherche alors à faire passer la « meilleurs » droite par un ensemble de points en minimisant l'erreur des moindres carrés.

Cependant, dans le cas où la variable explicative est qualitative, la spécification d'un modèle linéaire est incorrecte. Selon la nature de la variable d'intérêt (dichotomique, polytomique) l'on utilisera le modèle logistique binaire décrit alors la modélisation d'un variable qualitative Y à 2 modalités: (par exemple 1 ou 0, succès ou échec appartenance à un groupe ou pas, présence ou absence, ...). L'objectif est adapté à cette situation

en cherchant à expliquer les probabilités

$$\pi = P(Y = 1) \quad \text{ou} \quad 1 - \pi = P(Y = 0),$$

où plutôt une transformation de ces probabilités par l'observation conjointe des variables explicatives. Lorsque

$$Y = \begin{cases} 1 & \text{si l'évènement s'est réalisée} \\ 0 & \text{si l'évènement ne s'est pas réalisé} \end{cases}$$

La probabilité de survenue de l'évènement peut être définie comme l'espérance de la variable codée Y_i , puisque $E(Y_i) = P(Y_i = 1) \times 1 + P(Y_i = 0) \times 0 = \pi_i$.

3.1.2 Spécification du modèle logistique

Au lieu d'estimer Y_i , l'objectif est alors ici d'estimer $P_i = P(Y_i = 1)$ la probabilité de succès de l'évènement, avec P_i compris entre 0 et 1. le problème posé est qu'une mesure de probabilité est bornée à droite et à gauche. Il convient donc de trouver un moyen pour supprimer ces bornes.

En divisant P par $(P - 1)$, la borne à gauche est annulée car lorsque P tend vers 1, $\frac{P}{P-1}$ tend vers $+\infty$. Si P tend vers 0, alors $\frac{P}{1-P}$ tend vers 0. On applique une seconde transformation de type logarithmique de telle manière à ce que lorsque P tend vers 1, la transformation tend vers $-\infty$.

On définit la fonction logit comme étant l'application $\text{logit} :]0,1[\rightarrow]-\infty, +\infty[$ $P \mapsto \text{logit}(P) = \ln\left(\frac{P}{1-P}\right)$. Elle est continue, strictement croissante sur $]0,1[$ et à valeurs dans $]-\infty, +\infty[$. La fonction logit admet de ce fait une bijection réciproque notée F où F est l'application $F :]-\infty, +\infty[\rightarrow]0,1[$ $t \mapsto \frac{1}{1+e^{-t}} = \frac{e^t}{1+e^t}$. F est appelée fonction logistique tandis que logit porte le nom de fonction de lien. Le modèle qu'on va chercher à estimer va donc prendre la forme suivante:

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta X + \mu,$$

où β_0 est une constante $\beta = (\beta_1, \dots, \beta_k)$ un vecteur de paramètres à estimer et μ le terme erreur.

Ainsi, à partir de la fonction logistique F , il est plus facile de retrouver la probabilité de succès P avec

$$P = P(Y = 1/X) = F[\text{logit}(P)] = \frac{e^{\beta_0 + \beta X}}{1 + e^{\beta_0 + \beta X}} \quad (3.1)$$

et où β_0 , β et X sont connus.

estimation des paramètres du modèle

Le modèle une fois construite, le problème posé est d'estimer les paramètres. La première difficulté qui surgit est celle de l'inconnue P . La seule information que l'on dispose à son sujet est que l'évènement s'est réalisé ou pas, suivant les observations. En reconsidérant le modèle (3.1), l'on peut préciser que $P = P(Y = 1/X)$ représente en fait une moyenne: c'est la proportion des individus qui sont dans le groupe (par exemple) et qui possèdent un vecteur commun de caractéristiques. Un moyen simple d'estimer le modèle serait alors de regrouper les individus qui ont un même profil, de calculer cette proportion, et d'utiliser celle-ci comme valeur de la probabilité. Ensuite, de calculer le logit appliquer une méthode des moindres carrés ordinaires (MCO).

Au lieu de raisonner ainsi en deux étapes, on peut utiliser une autre technique d'estimation appelée maximum de vraisemblance.

Méthode du maximum de vraisemblance.

considérons un échantillon d'apprentissage $\{(X_i = x_i, Y_i = y_i)\}_{i=1}^N$ de N individus, avec $y_i \in \{0,1\}$ tel que:

$$y_i = \begin{cases} 1 & \text{si } P_i = F(\beta_0 + \beta x_i) \\ 0 & \text{si } P_i = 1 - F(\beta_0 + \beta x_i) \end{cases}$$

où $x_i = (x_i^1, \dots, x_i^k)_i^T \forall i \in \{1, \dots, N\}$, est un vecteur de caractéristiques observables sur l'individu i ; β_0 un paramètre inconnu, $\beta = (\beta_1, \dots, \beta_k) \in \mathbb{R}^k$ est un vecteur de paramètres inconnus.

Les valeurs observées y_i sont alors considérées comme les réalisations d'un processus binomial avec une probabilité de $F(\beta_0 + \beta x_i)$. Ainsi la vraisemblance associée à l'observation y_i s'écrit sous la forme: $L(y_i, \beta) = P_i(1 - P_i)^{1-y_i}$.

Dès lors, la vraisemblance associée à l'échantillon de taille N notée $Y = (y_1, \dots, y_N)$ s'écrit de la façon suivante:

$$L(Y, \beta) = \prod_{i=1}^N P_i^{y_i} (1 - P_i)^{1-y_i} = \prod_{i=1}^N [F(\beta_0 + \beta x_i)]^{y_i} [1 - F(\beta_0 + \beta x_i)]^{1-y_i}$$

où

$$F(\beta_0 + \beta x_i) = \frac{e^{\beta_0 + \beta x_i}}{1 + e^{\beta_0 + \beta x_i}}.$$

De cette définition de la vraisemblance, on en déduit la log vraisemblance qui est sous la forme:

$$\log L(Y, \beta_0, \beta) = \sum_{i=1}^N y_i \log[F(\beta_0 + \beta x_i)] + \sum_{i=1}^N (1 - y_i) \log[1 - F(\beta_0 + \beta x_i)].$$

L'estimateur du maximum de vraisemblance des paramètres β_0 et β est obtenu en maximisant soit la fonction de vraisemblance soit la log vraisemblance.

Les paramètres $(\widehat{\beta}_0, \widehat{\beta})$ se trouvent en résolvant l'équation $G(\widehat{\beta}_0, \widehat{\beta}) = 0$ où $G(\widehat{\beta}_0, \widehat{\beta})$ désigne la gradient de $\log L(Y, \beta_0, \beta)$.

3.1.3 Evaluation et validation du modèle

Vraisemblance

Un bon modèle est un modèle dont la vraisemblance est grande, c'est-à-dire qui tend vers 1. En pratique, on utilise la «moins double log vraisemblance»: $-2LL$, de sorte que lorsque la vraisemblance tend vers 1, alors $-2LL$ tend vers 0.

Pour tester l'adéquation d'un modèle, on utilise l'hypothèse nulle suivante: $H_0 : -2LL = 0 = 0$. La $-2LL$ se distribuant comme le Khi 2, avec $N - P$ degré de liberté (N est la taille de l'échantillon et P est le nombre de paramètres). Si la probabilité est inférieure au seuil de risque, on est amené à rejeter l'hypothèse nulle que le modèle convient. (Si cete probabilité est supérieure, on ne peut rejeter l'hypothèse nulle: on conviendra que le modèle est correct).

L'évaluation du modèle logistique peut se faire en utilisant plusieurs techniques employées le pus souvent de manière concourante. La pertinence d'un modèle se fonde en général sur la significativité des coefficients, sur la significativité globale du modèle au moyen des tests de vraisemblances. L'analyse des résidus et la puissance des discriminances permettent également de jauger la pertinence du modèle.

Tests d'hypothèses sur les paramètres

Test de Wald

On considère le test $H_0 : \beta_j = a$ où β_j désigne la j^{eme} composante du vecteur de paramètres $\beta = (\beta_1, \dots, \beta_k)' \in \mathbb{R}^k$ d'un modèle (dichotomique). L'idée du test de Wald est d'accepter l'hypothèse nulle si l'estimateur $\widehat{\beta}_j$ est proche de a ; la statistique de ce test est une mesure de proximité de $\beta_j - a$ à zéro. $\widehat{\beta}_j$ étant la valeur estimée de β_j sous cette hypothèse nulle $(\widehat{\beta}_j - a)$ suit asymptôtiquement laloi normale $\mathcal{N}(0, \widehat{\sigma}_j^2)$, où $\mathcal{N}(0, \widehat{\sigma}_j^2)$ est l'estimateur de la variance de $\widehat{\beta}_j$. Pour N assez grand, le rapport $\frac{\widehat{\beta}_j - a}{\widehat{\sigma}_j^2}$ suit asymptotiquement une distribution normale centrée et réduite.

La statistique de Wald se définit par:

$$\text{Wald}(\beta_j) = \frac{(\widehat{\beta}_j - a)^2}{\text{var} \widehat{\beta}_j} = \left(\frac{\widehat{\beta}_j - a}{\widehat{\sigma}^2} \right)^2 = Z_j.$$

La statistique Z_j suit un Khi-deux à un degré de liberté \mathcal{X}_1^2 . Ainsi, si l'on note $\mathcal{X}_{95\%}^2(1)$ le quantile à 95% de la loi $\mathcal{X}^2(1)$, le test de Wald au seuil de 5% de l'hypothèse H_0 consiste à accepter H_0 si $Z_j < \mathcal{X}_{5\%}^2$, et refuser si $Z_j > \mathcal{X}_{95\%}^2(1)$. on parlera de la

significativité du coefficient β_j si $P(\chi_1^2 > Z)$ (qui est la p-value) est inférieure au seuil du test.

Remarque 3.1.1. *La plus part des logiciels (sauf SAS) ne proposent pas cette statistique de Wald, mais une statistique Z_j définie comme la racine carré de $\frac{(\hat{\beta}_j - a)^2}{\text{Var}(\hat{\beta}_j)}$ définie précédemment. Compte tenu des liens entre la loi normale centrée réduite et la loi du Khi-deux à un degré de liberté, on a immédiatement sous H_0 :*

$$Z_j = \frac{\hat{\beta}_j - a}{\hat{\sigma}_j} \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

et en particulier pour un test de nullité $H_0 : \beta_0 = 0$, on retrouve $Z_j = \frac{\hat{\beta}_j}{\hat{\sigma}_j} \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0,1)$.

Test du rapport des maxima de vraisemblance [2]

Dans le cas des modèles dichotomiques, on peut appliquer sans difficulté particulière la logique du test du rapport des maxima de vraisemblance. Ainsi, on estime le modèle non contraint (il s'agit ici du modèle considéré sans la variable du coefficient à tester): Soient $\hat{\beta}_j$ et $\hat{\beta}_j^c$ les deux estimations ainsi obtenues. La statistique du *LRT* correspond alors tout simplement à l'écart des log-vraisemblances. On a le résultat suivant:

Définition 3.1.1. *La statistique LRT_j du test du rapport des maxima de vraisemblance associé au test unidirectionnel $H_0 : \beta_j = a$ contre $H_1 : \beta_j \neq a$ admet la loi suivante sous H_0*

$$LRT_j = -2 \left[\log L(y, \hat{\beta}_j) - \log L(y, \hat{\beta}_j^c) \right] \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \chi^2(1),$$

où $\hat{\beta}_j$ et $\hat{\beta}_j^c$ désignent respectivement les estimateurs non contraint de β_j .

Ainsi, si l'on note $\chi_{95\%}^2(1)$ le quantile à 95% de la loi du $\chi^2(1)$, le test des maxima de vraisemblance au seuil 5% de l'hypothèse H_0 consiste à accepter H_0 si $LRT_j < \chi_{95\%}^2(1)$, et à refuser H_0 si $LRT_j > \chi_{95\%}^2(1)$.

Remarque 3.1.2. *Cette procédure est asymptotiquement équivalente à celle d'un test de Wald.*

3.1.4 Evaluation du pouvoir discriminant du modèle

Reclassement

La régression logistique est utilisée pour modéliser la probabilité des attributs 0 et 1 de la variable dépendante y en fonction des variables x_1, x_2, \dots, x_p . A partir des probabilités estimées, on décidera en fixant un seuil, par exemple à 0.05 de classer l'individu dans la catégorie $y = 1$ si la probabilité est supérieure au seuil et dans la catégorie $y = 0$ sinon. On définit ainsi une règle de classement: si

$$\hat{P}_j = \hat{P}(y = 1/x_j) \geq \text{seuil alors } \hat{y} = 1 \text{ et } 0 \text{ sinon.}$$

Une fois le classement des individus fait, il est intéressant de déterminer la performance. Pour cela, l'on peut utiliser les notions de sensibilité et de spécificité.

Définition 3.1.2. *La sensibilité est définie comme la probabilité de classer l'individu dans la catégorie $y = 1$ (on dit que le test est positif): $\text{sensibilité} = P(\text{test} + / y = 1)$.*

Définition 3.1.3. *La spécificité est la probabilité de classer l'individu dans la catégorie $y = 0$ (on dit que le test est négatif): $\text{spécificité} = P(\text{test} - / y = 0)$.*

Remarque 3.1.3. *Lorsqu'on varie le seuil, la sensibilité et la spécificité changent, puisque la règle de classement est modifiée. Afin de représenter les valeurs de toutes les possibilités du seuil, on dessine sur un graphe les courbes de sensibilité et de spécificité.*

3.2 Discussion et interprétation des résultats

Après la classification, une variable que nous avons nommé classe a été générée grâce au logiciel SPAD. Ici l'on a utilisé les méthodes d'archivage et d'exportation des données incorporées dans ce logiciel. Ses modalités correspondant aux différents groupes formés sont classe1, classe2, classe3. Nous avons vu précédemment que les ménages se trouvant dans la classe1 vivent dans de mauvaises conditions, ceux de la classe2 dans de conditions de vie moyenne, et les ménages de la troisième classe classe3 dans de bonnes conditions de vie. Les classes étant formées, nous construisons ainsi un model logistique afin d'expliquer ces classes suivant les variables socio-démographiques.

Méthodologie de l'analyse

On fera d'abord une étude bivariable univariée. On étudiera à l'aide du test de Khi2 la liaison entre la variable dépendante et les différentes variables explicatives d'une part, et d'autres part les variables explicatives entres elles. Si une variables n'est pas liée avec la variable explicatives, on l'élimine de l'étude. Si deux variables explicatives sont liées entre elle, on retient une seule pour l'analyse. Une fois les variables explicatives sélectionnées, nous procéderons à une analyse multivariable qui se fera à l'aide de la regression logistique. Ceci permettra de faire une étude collective des facteurs décrivant l'appartenance à une classe donnée connaissant les caractéristiques de l'individu (ou ménage).

3.2.1 Modèle discriminant de la classe1

Analyses bivariées

Les variables candidates à la régression sont surtout les variables socio-démographiques. Car il est question pour d'expliquer la probabilité d'appartenance à une classe connaissant

les caractéristiques du ménage. On a alors les variables suivantes :

- **milieu** : désigne le milieu de vie. Ses modalités sont urbain, semi-urbain et rural.
- **religion** : désigne la religion du chef de ménage. ses modalités sont : catholique, protestant, autres chrétiens, musulmans, animistes, autres
- **sexe** : désigne le sexe du chef de ménage .ses modalités sont masculin , féminin
- **age** : désigne l'âge du chef de ménage .Initialement quantitative, on l'a rendu en variables qualitatives. Ses modalités sont : moins de 29 ans, 30 à 44 ans, 45 à 59 ans, 60 ans et plus
- **statutm** : désigne le statut matrimonial du chef de ménage . Ses modalités sont : célibataire, marié monogamie, marié polygamie, veuf/veuve, divorcé/séparé, en union libre.
- **hand** : désigne si le chef de ménage est victime d'un handicap ou pas. Ses modalités sont : oui, non.
- **Santé** : donne l'appréciation de l'état de santé du chef de ménage. Ses modalités sont : bon, assez bon , passable, mauvais.
- **instruct** : désigne le niveau d'instruction du chef de ménage . Ses modalités sont : non scolaire, primaire, secondaire 1er cycle, secondaire 2e cycle, supérieur.

Avant d'analyser nos données au moyen d'un modèle de régression logistique multivariées, il est nécessaire de procéder à des analyses bivariées qui nous permettront d'appréhender les facteurs de risques potentiellement associés au résultat. Sur la base de ses résultats, l'on procédera à un tri préalable de ses facteurs selon leur degré d'évidence afin d'éviter les risques de colinéarité ou la difficulté d'interprétation. Nous avons fait l'analyse bivariée à l'aide du test d'indépendance de Chi-deux. Les résultats sont présentés dans le tableau ci-après :

variables	P-Value
instruct	0.000 ***
tailm	0.027 *
hand	0.000 ***
statutm	0.000 ***
age	0.000 ***
sexe	0.987 NS
religion	0.000 ***
milieu	0.000 ***

TAB. 3.1 – Analyses bivariées des facteurs potentiels associés à classe1.

Signification des codes : NS :non significatif ; * :significatif ; ** * : très significatif

Au vu du tableau ci-dessus, le test de Khideux pour les variables statutm et sexe avec la variable explicative classe1 ne sont significatif. On doit alors les retirer du modèle. Donc les variables qui vont participer à la regression logistique sont les suivantes : instruct, tailm, hand, age, religion, milieu.

Analyses multivariables

Nous allons à présent étudier l'effet conjoint de plusieurs variables sur la probabilité d'appartenir à la classe1 ;Ici l'on estimera la contribution pour chaque facteur à l'explication de cette classe . il est à rappeler que la classe1 est assimiler à la classe des personnes jugées pauvres.

Estimation du modèle

La méthode utilisée ici est la méthode pas à pas. On commence par le modèle qui contient toutes les variables. On élimine la variables qui la plus grande p-value. Le premier modele obtenu est le suivant :

```
Call: glm(formula = classe1 ~ milieu + religion + age + hand +
statutm + tailm + instruct, family = binomial, data = tabpq)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1663	-0.7570	-0.3036	0.8272	2.7092

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.307553	0.156177	8.372	< 2e-16 ***
milieu2	-0.625147	0.063968	-9.773	< 2e-16 ***
milieu3	-2.888714	0.074267	-38.89	< 2e-16 ***
religion2	-0.034869	0.064515	-0.540	0.588865
religion3	-0.136705	0.141376	-0.967	0.333563
religion4	-0.158122	0.080029	-1.976	0.048177 *
religion5	-0.797352	0.162774	-4.899	9.66e-07 ***
religion6	-0.069028	0.188550	-0.366	0.714291
age2	-0.404193	0.079939	-5.056	4.28e-07 ***
age3	-0.713688	0.093280	-7.651	1.99e-14 ***
age4	-0.793232	0.110014	-7.210	5.58e-13 ***
hand2	0.011750	0.104506	0.112	0.910477
statutm2	-0.359026	0.089545	-4.009	6.09e-05 ***
statutm3	-0.414391	0.111952	-3.702	0.000214 ***
statutm4	-0.514915	0.120000	-4.291	1.78e-05 ***
statutm5	-0.567014	0.145224	-3.904	9.45e-05 ***
statutm6	0.001575	0.125595	0.013	0.989995
tailm2	-0.235336	0.078009	-3.017	0.002555 **
tailm3	-0.243780	0.087310	-2.792	0.005236 **
tailm4	-0.201443	0.100061	-2.013	0.044093 *
instruct2	0.642153	0.083450	7.695	1.41e-14 ***
instruct3	0.925030	0.095840	9.652	< 2e-16 ***
instruct4	0.575868	0.106915	5.386	7.19e-08 ***
instruct5	-0.630807	0.116621	-5.409	6.34e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Le critère AIC ici est 8738. On remarque que le coefficient de la variable n'est pas significatif au seuil de 5%. On l'élimine du modèle et on procède à une autre estimation du modèle.

```
Call: glm(formula = classe1 ~ milieu + religion + age + statutm +
tailm + instruct, family = binomial, data = tabpq)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.1660	-0.7570	-0.3037	0.8274	2.7098

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.318586	0.121514	10.851	< 2e-16 ***
milieu2	-0.624904	0.063931	-9.775	< 2e-16 ***
milieu3	-2.888485	0.074238	-38.909	< 2e-16 ***
religion2	-0.034943	0.064511	-0.542	0.588051
religion3	-0.136951	0.141356	-0.969	0.332627
religion4	-0.157923	0.080009	-1.974	0.048403 *
religion5	-0.797065	0.162748	-4.898	9.70e-07 ***
religion6	-0.069180	0.188550	-0.367	0.713690
age2	-0.404370	0.079923	-5.060	4.20e-07 ***
age3	-0.714530	0.092980	-7.685	1.53e-14 ***
age4	-0.794869	0.109048	-7.289	3.12e-13 ***
statutm2	-0.358758	0.089512	-4.008	6.12e-05 ***
statutm3	-0.414215	0.111941	-3.700	0.000215 ***
statutm4	-0.514701	0.119986	-4.290	1.79e-05 ***
statutm5	-0.566696	0.145194	-3.903	9.50e-05 ***
statutm6	0.001595	0.125593	0.013	0.989865
tailm2	-0.235162	0.077993	-3.015	0.002568 **
tailm3	-0.243551	0.087285	-2.790	0.005266 **
tailm4	-0.201258	0.100048	-2.012	0.044261 *
instruct2	0.642018	0.083441	7.694	1.42e-14 ***
instruct3	0.924999	0.095839	9.652	< 2e-16 ***
instruct4	0.575654	0.106897	5.385	7.24e-08 ***
instruct5	-0.630638	0.116609	-5.408	6.37e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Null deviance: 12023.4 on 8753 degrees of freedom Residual
deviance: 8686.2 on 8730 degrees of freedom AIC: 8734.2

On note ici que le critère AIC qui est égale à 8734,6 qui est inférieur à celui du modèle précédent. Donc le deuxième modèle est meilleur que le premier.

3.2.2 Modèle discriminant de la classe2

Analyse bivariée A l'aide du test de Le test de Khi-deux ,on teste le degré de liaison entre la variable classe2 et les autres.les résultats sont donnés dans le tableau suivant suivant:

variables	P-Value
instruct	0.000 ***
tailm	0.000 ***
hand	0.374 NS
statutm	0.000 ***
age	0.000 ***
sexe	0.9955 NS
religion	0.000 ***
milieu	0.000 ***

Le tableau ci-dessus nous montre que le test de Khi-deux pour les variables sexe et hand ne sont pas significatifs au seuil de 5%. Ces variables n'expliquent pas donc la probabilité d'appartenir à la classe 2.

Estimation du modèle

Ayant supprimé les variables sexe et age de la regression, on obtient le modèle suivant :

Call: glm(formula = classe2 ~ milieu + religion + age + tailm +
instruct + statutm, family = binomial, data = tabpq)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.2159	-0.7594	-0.3072	0.8461	2.7197

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.48857	0.12127	12.275	< 2e-16 ***
milieu2	-0.58917	0.06395	-9.213	< 2e-16 ***
milieu3	-2.88405	0.07444	-38.741	< 2e-16 ***
religion2	-0.07744	0.06485	-1.194	0.232400
religion3	-0.06312	0.14067	-0.449	0.653655
religion4	-0.20396	0.07903	-2.581	0.009856 **
religion5	-0.94431	0.16566	-5.700	1.20e-08 ***
religion6	-0.19366	0.19354	-1.001	0.317000
age2	-0.41984	0.07983	-5.259	1.44e-07 ***
age3	-0.67739	0.09297	-7.286	3.20e-13 ***
age4	-0.78769	0.10838	-7.268	3.65e-13 ***
tailm2	-0.27629	0.07767	-3.557	0.000375 ***
tailm3	-0.34643	0.08681	-3.990	6.60e-05 ***
tailm4	-0.20355	0.09963	-2.043	0.041047 *
instruct2	0.62833	0.08289	7.580	3.45e-14 ***
instruct3	0.84997	0.09567	8.884	< 2e-16 ***
instruct4	0.54339	0.10614	5.119	3.06e-07 ***
instruct5	-0.68373	0.11660	-5.864	4.52e-09 ***
statutm2	-0.49274	0.08872	-5.554	2.80e-08 ***
statutm3	-0.52258	0.11120	-4.699	2.61e-06 ***
statutm4	-0.60005	0.12041	-4.983	6.25e-07 ***
statutm5	-0.58251	0.14522	-4.011	6.04e-05 ***
statutm6	0.02677	0.12778	0.210	0.834045

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 12017 on 8753 degrees of freedom Residual
deviance: 8661 on 8731 degrees of freedom AIC: 8707

Modèle discriminant de la classe3

variables	P-Value
instruct	0.000 ***
tailm	0.000 ***
hand	0.408 NS
statutm	0.000 ***
age	0.000 ***
sexe	0.0027 ***
religion	0.000 ***
milieu	0.000 ***

3.2.3 Estimation du modele de la classe3

Call: glm(formula = classe3 ~ milieu + religion + age + tailm +
instruct + statutm, family = binomial, data = tabpq)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.9898	-0.4092	-0.1694	-0.0586	3.5201

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.8366699	0.2584892	-22.580	< 2e-16 ***
milieu2	-0.7066793	0.0996420	-7.092	1.32e-12 ***
milieu3	-3.0257262	0.2254412	-13.421	< 2e-16 ***
religion2	0.2733817	0.0909215	3.007	0.00264 **
religion3	0.0371199	0.2044986	0.182	0.85596
religion4	-0.0004308	0.1356259	-0.003	0.99747
religion5	-0.7490025	0.3728252	-2.009	0.04454 *
religion6	0.4088275	0.2546158	1.606	0.10835
age2	0.9360608	0.1391473	6.727	1.73e-11 ***
age3	1.2643342	0.1589777	7.953	1.82e-15 ***
age4	1.6931061	0.1985841	8.526	< 2e-16 ***
tailm2	0.8299213	0.1403279	5.914	3.34e-09 ***
tailm3	1.1727408	0.1506147	7.786	6.90e-15***
tailm4	1.3486608	0.1642036	8.213	< 2e-16***
instruct2	1.2052173	0.2088231	5.771	7.86e-09 ***
instruct3	2.2602195	0.2139739	10.563	< 2e-16 ***
instruct4	3.2838586	0.2189028	15.001	< 2e-16 ***
instruct5	4.5094013	0.2255434	19.994	< 2e-16 ***
statutm2	0.2718425	0.1386839	1.960	0.04998 *
statutm3	0.3572386	0.1729195	2.066	0.03884*
statutm4	0.5297769	0.2026458	2.614	0.00894 **
statutm5	0.6042330	0.2378496	2.540	0.01107 *
statutm6	-0.5448826	0.2041886	-2.669	0.00762 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1) Null
deviance: 6540.8 on 8753 degrees of freedom Residual deviance:
4153.9 on 8731 degrees of freedom AIC: 4199.9

CONCLUSION

Au terme de notre travail qui a consisté à trouver une typologie de 10992 ménages selon les caractéristiques socio-économiques du chef de ménages, étant donné que le but de cette typologie est l'amélioration des conditions de vie afin de réduire la pauvreté, nous avons d'abord eu à caractériser ces ménages suivant l'approche pauvreté monétaire. Après avoir présenté la méthode de classification Ascendante hiérarchique, nous avons procédé à une catégorisation des ménages en implémentant sous le logiciel SPAD la classification Ascendante Hiérarchique. A la suite de cette classification, trois groupes distingués et homogènes se sont distingués. Un premier groupe représentant 43.27% des ménages considéré comme le groupe des pauvres vu les conditions précaires de vie de ces ménages. Ces ménages se retrouvent essentiellement dans le milieu rural. Un deuxième groupe représentant 44.30% des ménages considéré comme celui des ménages à niveau de vie moyen. Un troisième groupe représentant 12.43% considéré comme celui des ménages riches. Après avoir décrit ces groupes notre travail a consisté à la recherche d'un modèle expliquant la probabilité d'appartenance des ménages dans les différentes classes connaissant les variables démographiques. Il est à noter ici que l'étude approfondie de la classe 1 (classe pauvre) peut aider les autorités publiques dans l'amélioration des conditions de vie des populations et la lutte contre la pauvreté.

ANNEXE

code	Nom de la variables	modalités
région	douala/yacounde/province	1=Douala;2=Yacoundé 3=Adamaoua ; 4=Centre 5=Est ; 6=Extrême-Nord 7=Littoral ; 8=Nord ; 9=Nord-Ouest ;10=Ouest 11=Sud ;12=Sud-Ouest
milieu	milieu de residence	1=Urbain 2=Semi-Urbain 3=Rural
religion	religion du chef de menage	1=Catholique ;2=protestant 3=autres chretiens 4=musulmans ;5=animistes 6=autres
sexe	sexe de la personne	1=masculin 2=féminin
age	age de la personne	1=moins de 29 ans 2=30 à 44 ans 3=45 à 59 ans 4=60 ans et plus
statutm	statut matrimonial	1= célibataire 2=marie monogame 3=marie polygame 4=veuf/veuve 5=divorcé/séparé 5= en union libre
hand tailm	victime d'un handicap Taille ménage	1=oui ; 2= non 1=au plus 2 enfants 2=de 3 à 5 enfants 3=de 6 à 8 enfants 4=9 enfants et plus
santé	appreciation de son etat de sante	1=bon ; 2= assez bon 3= passable ; 4= mauvais
instruct	niveau d'instruction	1= non scolarise 2=primaire 3= secondaire 1er cycle 4=secondaire 2° cycle 5= supérieur
logement	type de logement	1=maison isolee 2=maison à plusieurs logements 3=villa moderne 4=immeuble a appartements 5=concession'saire

aisance	type des lieux d'aisance	1=wc avec chasse eau 2= latrine aménagée 3= latrine non aménagée 4= pas de wc ; 5= autres
radio	possession d'un poste radio	1=oui ; 2=non
refrig	possession d'un réfrigérateur	1=oui ; 2=non
cong	possession d'un congélateur	1=oui ; 2=non
vent	possession d'un ventilateur	1=oui ; 2=non
cuisi	possession d'une cuisinière	1=oui ; 2=non
rechg	possession d'un réchaud à gaz	1=oui ; 2=non
rechp	possession d'un réchaud à pétrole	1=oui ; 2=non
vehi	possession d'un véhicule	1=oui ; 2=non
Fer_rep	possession d'un fer à repasser	1=oui ; 2=non
Tv	possession d'un téléviseur	1=oui ; 2=non
teleph	possession d'un téléphone	1=oui ; 2=non
Sol	matériaux du sol	1=ciment ou carreaux 2=bois ou terre ; 3=autres
murs	matériaux du murs	1=béton/parpaing/pierre de taille 2=planche/carabot 3=têne/brique simple 4=pise/terre battue 5=nattes/chaumes ; 6=autres
eau	mode d'approvisionnement en eau de boisson	1=robinet individuel 2=robinet commun 3= forage/borne fontaine 4=puits/source aménagée 5=rivière/maigot ; 6=autres
So_en_cui	source d'énergie pour la cuisine	1=bois ; 2=gaz/électricité 3=pétrole ; 4= charbon/sciure 5=autres
toit	matériau du toit	1=ciment ; 2=tole/tuile 3=nattes/chaumes ; 4=autres
eclairage	source d'éclairage du logement	1=pétrole 2=électricité abonnée à la SONEL 3=électricité non abonnée à SONEL 4=groupe électrogène 5= gaz ; 6= Autres
Stat_prof	statut professionnel	1=cadres ; 2=employés qualif. 3=autres salariés ; 4=patrons 5=propre compte ; 6=chômeurs/autr. n. sal.
sectirs	secteur institutionnel	1=adm. publique ; 2=entse publique 3=entse privée form ; 4=entse inf. agric. 5=entse inf. n. agric ; 6=chômeur

Bibliographie

- [1] L. ASSELIN et A. DAUPHIN. *Mesure de la pauvreté: un cadre conceptuel*. CENTRE CANADIEN D'ETUDE ET DE COOPERATION INTERNATIONALE.
- [2] JAMES E. CURTIS. *classes sociales*.
<http://www.thecanadianencyclopedia.com/index.cfm?PgNm=AboutThisSite&Params=F1>
- [3] ESSAM-NSSAH Boniface. *inégalité et bien être social*. 1re édition, de Boeck Université, Bruxelles.
- [4] - B. FOKO TAGNE , F. NDEM et R. TCHAKOTE. *pauvreté et inégalité de vie au Cameroun: une approche micro multidimensionnelle*:cahier de recherche PMMA 2007-02.
- [5] Dr Eugène- Patrice NDONG NGUEMA. *cours de Data Mining non supervisé*, 2007.
- [6] Véronique PETIT. *introduction, définition et mesures de la pauvreté, programme de recherche de la pauvreté en Guinée*.
- [7] MAX WEBER. *Class, status, party, Extrait de Gerth and Mills*. from Max Weber, 1958, Oxford University Press, pp. 181-186.
- [8] SILBER J., 2001. Amartya Sen. *La mesure de la croissance économique et du développement social*. Revue d'économie du développement, 3/2001, pp. 107-135. (pdf-selecte-Droy).
- [9] ECAM II. *Document de méthodologie*.

Table des matières

Dédicaces	i
Remerciements	ii
Liste des tableaux	iii
Table des figures	iv
Abréviations	v
Avant propos	vi
Résumé	viii
Abstract	ix
Introduction générale	1
Résumé exécutif	5
1 Concepts de base	7
1.1 Concepts de classes sociales	7
1.2 Concepts de pauvreté	8
1.3 Revue théorique de la classification	9
2 Elaboration des classes socio-économiques	13
2.1 Source des données	13
2.1.1 Coefficient d'extrapolation	15
2.2 Caractéristiques de la pauvreté monétaire	17
2.3 Catégorisation des classes socio-économiques	22
2.3.1 Présentation de la méthode	23
2.3.2 Choix du nombre de classes	25

2.3.3	Description des classes	26
3	Modélisation	33
3.1	Etude théorique	33
3.1.1	Régression logistique	33
3.1.2	Spécification du modèle logistique	34
3.1.3	Evaluation et validation du modèle	36
3.1.4	Evaluation du pouvoir discriminant du modèle	37
3.2	Discussion et interprétation des résultats	38
3.2.1	Modèle discriminant de la classe1	38
3.2.2	Modèle discriminant de la classe2	43
3.2.3	Estimation du modele de la classe3	46
4	Conclusion	47
5	Annexe	48
	Bibliographie	50

Table des figures

2.1	Niveau de vie et sexe du chef de ménage.	19
2.2	Niveau de vie et statut matrimonial.	19
2.3	Niveau d'instruction et niveau de vie.	20
2.4	Courbe des indices de niveau.	25
2.5	Dendrogramme.	26
2.6	Repartition des individus sur le 1er plan factoriel	32

Liste des tableaux

2.1	Caractéristiques socio-démographiques et niveau de vie.	18
2.2	Région,milieu de résidence et niveau de vie.	21
2.3	Niveau de vie et statut professionnel.	23
2.4	Effectif des classes.	26
2.5	Tableau croisant classe1 et milieu de residence.	27
2.6	Tableau croisant classe1 et niveau d'instruction du chef de ménage.	27
2.7	Tableau croisant classe1 et le statut matrimonial.	28
2.8	Tableau croisant classe1 et religion.	28
2.9	Tableau croisant classe1 et le sexe du chef de ménage.	28
2.10	Tableau croisant classe1 et taille du ménage.	28
2.11	Tableau croisant classe2 et niveau d'instruction du chef de ménage.	29
2.12	Tableau croisant classe2 et le statut matrimonial.	29
2.13	Tableau croisant classe2 et religion.	29
2.14	Tableau croisant classe2 et taille du ménage.	30
2.15	Tableau croisant classe2 et age.	30
2.16	Tableau croisant classe3 et milieude residence.	30
2.17	Tableau croisant classe3 et niveau d'instruction du chef de ménage.	31
2.18	Tableau croisant classe3 et le statut matrimonial.	31
2.19	Tableau croisant classe3 et l'age du chef de ménage.	31
2.20	Tableau croisant classe3 et le sexe du chef de ménage.	31
2.21	Tableau croisant classe3 et l'age du chef de ménage.	31
2.22	Tableau croisant classe3 et l'age du chef de ménage.	32
3.1	Analyses bivariabes des facteurs potentiels associées à classe1.	39