

# CARACTÉRISTIQUES ET DÉTERMINANTS DU SOUS-EMPLOI AU CAMEROUN

Par :

**HYEFOUAIS NGNIODEM Achille Stéphane**

hyefouais2001@voila.fr, stachylle@yahoo.fr

Sous la direction de :

**Dr. Eugène-Patrice NDONG NGUEMA**

Chargé de Cours, ENSP de Yaoundé

Sous l'encadrement de :

**M. ANJUENNEYA NJOYA Arouna**

Ingénieur Statisticien Économiste, INS de Yaoundé

Sous la supervision de :

**Pr. Henri GWÉT**

Maître de Conférences, ENSP de Yaoundé.

**Octobre 2006**

---

---

# Dédicaces

---

A mon père et à ma mère  
Monsieur **Roger NGNIODEM**,  
Madame **NGNIODEM née Christine KENNE** ;  
et à tous les membres de ma famille.

Pour toute l'attention et le soin qu'ils ne cessent d'accorder à mon éducation et à ma santé.

---

# Remerciements

---

Au **Pr. Henri GWÉT**, Maitre de Conférences à l'ENSP, pour avoir accepté de superviser ce travail malgré ses multiples occupations, et surtout pour avoir su coordonner avec tact les enseignements de ce Master.

Au **Dr. Eugène-Patrice NDONG NGUEMA**, Chargé de Cours à l'ENSP, pour avoir dirigé ce mémoire avec délicatesse. Par la rigueur qu'il a exigé à mon endroit et l'intérêt qu'ils portent à mes travaux, il m'a permis d'approfondir les connaissances acquises au cours de ma formation.

A **M. ANJUENNEYA NJOYA Arouna**, mon encadreur de stage à l'Institut National de Statistique, pour m'avoir proposé ce thème de recherche, et pour m'avoir initié dans le domaine de l'économie, spécialement dans les analyses des problèmes de pauvreté, d'emploi et du secteur informel.

A tous les enseignants du Master de Statistique Appliquée, en particulier, au **Pr. Danielle FLORENS** de l'Université Paris-Dauphine, pour tout le soin qu'ils ont accordé à notre formation de Master.

Je tiens également à exprimer ma gratitude à Monsieur **Barnabé OKOUDA**, sous-directeur des Etudes et de la Normalisation Statistique à l'INS, pour nous avoir reçu dans sa sous-direction et pour tout le matériel de travail qu'il a mis à notre disposition lors du stage.

Et à Monsieur **Ambroise ABANDA**.

A mes **parents, mes frères et sœurs**, pour le soutien sans faille et la patience qu'ils ont su me montrer. Je pense ainsi à la famille TSAYIM, à la famille NGNIODEM, et la famille TUAYO que je ne peux oublier à la fin d'un exercice pareil.

A ma sœur Nadine ASSONGWA NGNIODEM qui ne cesse de m'épauler sans hésitation, et pour qui j'ai beaucoup de d'admiration.

Aux amis et camarades : M. BONSO KAHOU Achille, M. Ngoko Augustin, M. WOUBOU Hugues,... Je remercie aussi M. Patrice Takam, M. Patrick TOTOUM, M. Aurelien MANFOUO N., M. Bertrand KAMGA F. et mlle Arlette WAFO.

---

# Sigles et abréviations

---

- BIT : Bureau International du travail
- BMO : Bureau de la Main d' Oeuvre
- CIST : Conférence Internationale des Statisticiens du Travail
- DIAL : Développement et Insertion Internationale
- ECAM : Enquête Camerounaise Auprès des Ménages
- EESI : Enquête sur l'Emploi et le Secteur Informel
- ENS : Sous-Direction des Études et de la Normalisation Statistiques
- ENSP : École Nationale Supérieure Polytechnique
- FNE : Fonds National de l'Emploi
- INS : Institut National de la Statistique
- INSEE : Institut National de la Statistique et des Etudes Économiques de France
- MINPLAPDAT : Ministère de la Planification, de la Programmation du développement et de l'Aménagement du Territoire
- MCO : Moindres carrés ordinaires
- MMV : Méthode du Maximum de Vraisemblance
- RGPH3 : Troisième Recensement Général de la Population et de l'Habitat
- SMIG : Salaire Minimum Interprofessionnel Garanti
- ZD : Zone de dénombrement

---

# Résumé

---

Le sous-emploi est l'une des variantes du marché de travail qui est en pleine effervescence au Cameroun. Ce problème affecte particulièrement les femmes, les jeunes et, s'agissant du milieu institutionnel, le secteur informel qui à lui seul abrite plus de 90% des actifs de plus de 10 ans, n'est pas de reste. Les institutions internationales comme le Bureau International du Travail (BIT) ou l'Organisation Internationale du Travail (OIT) ont distingué trois sortes de sous-emploi : le sous-emploi visible, c'est-à-dire lié au temps de travail ; le sous-emploi invisible lié au revenu ; et le sous-emploi global qui mesure toutes les distorsions du marché de travail dans un système économique donné et qui englobe les deux premières.

Dans ce mémoire, nous avons principalement étudié les sous-emplois visible et invisible, et avons choisi de regarder le sous-emploi global comme un résumé des deux autres, avec le chômage. Le choix porté sur le modèle logistique est dû à sa facilité à pouvoir régresser une variable dépendante du type dichotomique à partir des variables de types divers ; ainsi permet-il de classer les actifs occupés en situation de sous-emploi ou pas, afin d'établir le profil d'un sous-employé au Cameroun.

La première analyse économétrique a porté sur le sous emploi invisible. Nous avons obtenu deux modèles (un de treize variables et l'autre de douze variables) que nous avons confrontés à l'aide d'une batterie de tests, pour finalement retenir celui de treize variables dont le pouvoir de discrimination s'estime à près de 84%. Les principaux déterminants qui en découlent sont : le sexe, l'âge, le niveau d'instruction, le temps de travail, la catégorie socioprofessionnelle, le secteur institutionnel et la région de résidence.

Le deuxième modèle a servi à expliquer le sous-emploi visible. Sa construction s'est avérée un peu délicate<sup>1</sup> et a retenu dix variables, pour un taux de discernement égale à 72%. Ici, les caractéristiques les plus en vue sont : le nombre hebdomadaire de jours de travail, le revenu d'activité principale et secondaire, la promotion et le secteur institutionnel.

On peut déjà, à partir des éléments obtenus ci-dessus, proposer le sexe, l'âge, le revenu, le temps alloué au travail, le secteur d'activité, la région de résidence et le niveau d'instruction comme

---

<sup>1</sup>La délicatesse de l'étude peut trouver son explication dans la rareté du sous-emploi visible, son indicateur étant de 12.1% de la population active occupée.

déterminants du sous emploi global. Toutefois, si les interprétations qui découlent de ces modèles sont assez conformes avec celles de l'analyse empirique, c'est certainement parce qu'il s'agit de phénomènes liés à la conjoncture, et donc facilement observables à partir du rythme de vie des populations.

---

# Abstract

---

Underemployment can be presented as one of the variants of the job market which is in full affluence in Cameroon. This problem particularly affects women and youths, and dealing with institutional milieu, the informal sector which knows almost 90% of the active population for more than 10 years is as well concerned. The international organizations like the International Labour Organization (ILO) have distinguish three forms of it : the visible underemployment that is being related to working hours, the invisible underemployment related to the working revenu and the global underemployment which measures all the distortions of the job market in a given economic system and which consists of the first two.

In this dissertation, we have tried to analyse visible and invisible underemployment, and we have tried to look at the global underemployment as the summary of the two others, with the unemployment. We used the logistic model due to its facility to regress a dependent variable which is dichotomic from variables of various natures; as such it permits to classify the active group occupied in an underemployment situation or not, such as to establish a profile of an underemployed worker in Cameroon. The review of literature considerably contributed in the choice of candidate modelling variables, giving that the EESI 2005 data consisted of 234 variables.

The first econometric analysis was made up of invisible underemployment. We obtained two models (one of thirteen variables and the other of twelve variables) which we then confronted with a help of a series of tests, to finally retain that of thirteen variables which has a discriminative level estimated of approximately 84%. According to this model, the main characteristics of the invisible underemployment are : sex, age, the educational level, the time of work, social and occupational group, institutional sector and the residence region.

The second model served to explain the visible underemployment. Its construction revealed a bit delicate<sup>2</sup> and retained ten variables with a discernement rate of about 72%. This model allowed us to know that the visible underemployment depends highly on the weekly number of work days, the working income, the appointment to a post, the size of company and the institutional sector.

---

<sup>2</sup>The delicacy of this research can be found in the explanation of the scarcity of visible underemployment, its indicator being 12.1% of the active occupied population.

Consequently, we suggested to look the sex, the age, the educational level, the income, the time of work, the institutional sector and the residence region as determining factors of the global underemployment. However, if the interpretations derived from these models are in conformity with the empirical analysis, it is certainly because it consists of phenomena related to the conjuncture, and then easily perceived from the life style of the population.

---

# Résumé Exécutif

---

Dans ce mémoire, nous nous sommes intéressés à lister les principaux facteurs qui déterminent le sous-emploi au Cameroun, et à préciser la manière dont ils influencent ce phénomène. Notre attention portée à ce sujet vient du fait que le taux de chômage BIT au Cameroun (4,46%) donne l'illusion que le marché du travail se porte bien. Autrement dit, il traduit l'abondance des emplois, mais ne peut ressortir le fait que 7 travailleurs camerounais sur 10 ont un salaire mensuel principal inférieur au SMIG. Dès lors, le sous-emploi apparaît comme l'un des indicateurs BIT capables de traduire la réalité du marché du travail camerounais.

Au préalable, nous avons présenté les données de l'enquête EESI sur lesquelles nous basons notre étude et ce, à l'état brute afin de saisir la qualité de celles-ci (type, unités, dispersions,...), et la manière dont elles ont été recueillies. Ceci a permis de remarquer un taux très faible, sinon négligeable, de données manquantes, ce qui justifie la suppression des individus associés pendant l'étude. C'est aussi le moment de déceler des transformations à faire sur les variables avant utilisation de celles-ci.

## **Présentation des données sur le sous-emploi**

Pour la population active occupée, l'âge moyen est de 33,59%. Cependant, cette moyenne est très instable car l'écart type ici est de 14,6. On peut également remarquer que la population devient dense entre 20 et 40 ans. Concernant le genre, les hommes prennent une proportion de 54,3%, laissant 45,7% aux femmes. Le statut matrimonial exprime une densité élevée de célibataires (38,7%) et de monogames (37%) au Cameroun en 2005. Il révèle aussi que la polygamie (5,5%) et l'union libre (6,8%) prennent encore de l'ampleur dans ce pays. Sur le plan éducatif, on peut appréhender deux faits : 17,2% des travailleurs sont non scolarisés, pendant que seulement 5,8% d'entre-eux ont eu accès au supérieur. Par ailleurs, le secteur institutionnel se décompose en 4 : le public (7,7%), le privé formel (7,6%), l'informel agricole (33,9%) et l'informel non agricole (50,8%); le tout réparti en 3 zones : le milieu rural (24,1%), le milieu semi-urbain (23,5%) et le milieu urbain (52,4%). Sur le plan typiquement professionnel, on relève 2,7% de hauts cadres et ingénieurs assignés, 4,3% de cadres moyens, 6,3% d'employés qualifiés, 6,5% d'ouvriers semi-qualifiés, 8,1% de manœuvres, 3,2% d'employeurs, 48,4% d'auto-employés, 17,1% de ceux qui pratiquent l'aide-familial, 3,2% d'apprentis

et enfin 0,1% d'inclassables. Parmi ces employés, moins d'un sur 10 ont été recyclés, et 5% seulement ont connu une promotion au sein de l'entreprise.

Il est nécessaire de construire les indicateurs clés du sous-emploi au Cameroun (taux de chômage BIT, taux de sous-visible, invisible et global). Ceci permet de dégager l'ampleur du problème, et ce, suivant les facteurs économiques et sociaux comme le sexe, les classes d'âge, le niveau d'instruction, le revenu, le secteur d'activité...

### **Indicateurs clés du sous-emploi au Cameroun en 2005**

1. Le taux de sous-emploi visible en 2005 est de 12,12%. Il est un peu plus élevé chez les femmes (12,36%) que chez les hommes (11,89%), et plus élevé en milieu urbain (14,68%) qu'en milieu rural (10,48%). L'informel non agricole enregistre le chiffre le plus haut (19,85%) pendant que le privé formel s'empare du plus bas (6,18%). Même si ce phénomène semble rare, il reste fortement lié à la conjoncture : 1 actif occupé sur deux est sous-employé visible à cause des mauvaises conditions (financières) de travail.
2. Le taux de sous-emploi invisible s'évalue à 69,3% au Cameroun en 2005, 81% pour les femmes et 57,21% pour les hommes. Ce taux cache des disparités régionales. En effet, il est plus élevé en zone rurale (83,39%) qu'en zone urbaine (41,68%), et dans l'informel agricole (88,66%) et non agricole (58,49%) que dans le public (8,19%). Ce phénomène tend à disparaître, au fur à mesure que le niveau d'instruction augmente : 85,64% pour les non scolarisés, 32% pour les secondaires 2<sup>nd</sup> cycle et, 12,09% pour les employés ayant fait au moins une expérience dans l'enseignement supérieur.

Une étude plus approfondie permet, au moyen d'une modélisation logistique, d'énumérer les principaux déterminants du sous-emploi au Cameroun. Mais, avant toute chose, nous avons tenu à apurer les données en les recodant systématiquement pour les rendre simples et interprétables.

### **Étude économétrique et résultats**

Dans un premier temps, nous avons modélisé le sous-emploi invisible, avec une variable dépendante binaire qui classe un actif occupé en 1 si son revenu mensuel principal est inférieur au SMIG, et 0 sinon. Comme déterminants, on a pu retenir :

- Le sexe : les femmes sont bien plus exposées que les hommes.
- L'âge : les adolescents semblent plus sensibles. Bien plus, le sous-emploi diminue lorsque l'âge augmente.
- L'instruction : le modèle construit donne un grand avantage aux hommes instruits dans l'acquisition des emplois bien rémunérés.
- Le rang professionnel peut aussi influencer la situation d'emploi adéquat liée au revenu. Par

rapport aux hauts cadres, la vulnérabilité est 10 fois plus importante chez les manœuvres et les apprentis, et 2 fois plus élevée chez les ouvriers qualifiés.

- Le nombre de mois de travail, ou le nombre de jours et d’heures par semaine sont tout aussi significatifs pour mesurer le sous-emploi invisible et le respect scrupuleux de la législation dans ce domaine peut contribuer à la chute du phénomène au Cameroun.
- Pour le secteur d’activité, la présente étude montre que l’informel reste au centre du problème du sous-emploi invisible.

Le second modèle porte sur le sous-emploi visible interpellant ceux des actifs occupés qui, malgré eux, travaillent pendant moins de 35 heures par semaine dans l’activité principale. Il a été construit sur la base des mêmes techniques économétriques que le précédent, et permet d’identifier comme déterminants du sous-emploi visible les facteurs suivants :

- Contre toute attente, le sexe ne détermine pas le fait de travailler pendant moins de 35 heures par semaine, même si l’on sait que les femmes sont les plus soumises aux contraintes familiales et sanitaires.
- L’expérience professionnelle joue aussi un rôle non négligeable : les novistes (moins de 5 ans) étant plus exposés que les experts (plus de 10 ans).
- La faiblesse du revenu (moins de 23 500 FCFA) est aussi un gage pour l’évaluation du sous-emploi visible au Cameroun.
- Le temps mis au travail ne sera pas en reste : il est probable qu’à plus de 9 mois de travail par an, et à plus de 5 jours de travail par semaine, on ait tendance à éviter le phénomène.
- Également, le recyclage apporte beaucoup de moyens aux actifs occupés pour ne pas vivre ce type de sous-emploi.
- Le facteur âge révèle, après l’analyse, que le sous-emploi visible semble une affaire de jeunes et de vieux, avec la tendance à épargner les adultes.
- Sur le plan religieux, les musulmans sont les plus exposés, par rapport aux catholiques.

Comme réserve à l’étude, nous pouvons dire que la difficulté à construire le second modèle peut être due à une insuffisance des données liées au sous-emploi visible, comme les conditions d’accès aux soins de santé par exemple. Il est tout aussi vrai que ceux qui vivent dans des zones de couverture médicale faible peuvent être en risque de faire moins de temps dans leur activité principale.

### **Recommandations**

Ainsi, à partir des éléments tirés de l’analyse du sous-emploi au Cameroun, on peut faire quelques recommandations. Il serait nécessaire de faciliter l’accès à l’éducation et de fixer un seuil d’éducation, pour rendre les jeunes capables à intégrer le monde du travail. Il faut aussi améliorer la qualification

des travailleurs par les stages de formation et de mise à niveau. Pour réduire la prépondérance des emplois précaires, nous suggérons une revalorisation et la pratique effective du Salaire Minimum Interprofessionnel Garanti. Enfin, on peut déclarer qu'en fin d'année 2005, il naît le besoin de planifier le secteur informel et créer les conditions nécessaires pour encourager les investissements productifs et accroître le revenu de l'informel.

# INTRODUCTION GÉNÉRALE

---

Généralement, lorsqu'on s'intéresse à l'évolution du marché du travail dans une économie donnée, l'indicateur le plus utilisé est le taux de chômage. Il mesure le déséquilibre entre l'offre et la demande d'emploi. Concrètement, le taux de chômage indique la proportion de la population active dépourvue d'emploi, mais qui reste disponible et cherche activement du travail.

Cet indicateur, défini par les institutions internationales et adopté par bon nombre de pays, ne permet cependant pas une description complète du marché de travail. En effet, le taux de chômage à lui seul ne sied pas à la comparaison de l'efficacité de deux types de marché du travail, encore moins à l'appréciation du degré d'exclusion qu'engendrent deux systèmes sociaux aussi différents l'un de l'autre que la France et les États-Unis. Par exemple, l'article "**Les chiffres trompeurs du chômage**" [14], révèle qu'en 1995, le taux de chômage des jeunes en France est de 25.9%, et de 12.1% aux USA. De prime à bord, on est tenté de croire que le chômage est plus élevé en France qu'au USA à cette date. Pourtant, un jeune chômeur n'est pas perçu de la même manière dans les deux systèmes : en France, le taux officiel de chômage exclut les jeunes mères célibataires à cause de leur indisponibilité, contrairement aux États-Unis malgré l'absence des crèches publiques. Par ailleurs, en 1979 en Grande Bretagne, les statisticiens ont éprouvé des difficultés d'interprétation et de mesure du chômage. Suivant les normes classiques du Bureau International du Travail (BIT), ce taux avoisinait 6%. Mais en tenant compte des réalités et des exigences nationales, le taux réel de chômage s'élevait à plus de 12%, soit à peu près le double. Ces derniers chiffres tirés de l'essai "**Pour une nouvelle mesure du sous-emploi**" d'Alain MASSOT [13], exhibe là la difficulté pour le chômage BIT à épouser les circonstances et les réalités nationales.

Lorsqu'on examine la situation du marché du travail dans les pays sous-développés, particulièrement en Afrique sub-saharienne, on relève de manière générale un taux de chômage BIT relativement faible. Ceci ne traduit cependant pas l'efficacité du marché de travail comme on l'aurait pensé dans un pays développé. Un exemple illustratif est celui du Cameroun où, selon le Rapport Principal de l'Enquête sur l'Emploi et le Secteur Informel (EESI) réalisée en 2005 par l'INS, le taux de chômage est en baisse au cours de ces dernières années et a été estimé à 4,46% en 2005 ; ce qui est ambigu lorsque les chiffres de la même enquête relèvent un taux de 69,3% d'actifs gagnant moins que le

Salaire Minimum fixé au Cameroun à 23 500 FCFA.

Les faibles taux du chômage dans les pays en développement résultent de l'abondance des emplois traditionnels, où le gain insuffisant des travailleurs engendre une production moins importante que celle qu'ils auraient fournie s'ils travaillaient sous une bonne rémunération. En outre, ces pays ne possèdent pas pour la plupart de programmes d'aide aux chômeurs, et une bonne partie des travailleurs sans emploi se livrent à une activité de subsistance, même si celle-ci ne cadre pas avec leur formation ou n'assure pas une vie décente ; dérobant ainsi ces travailleurs du plein emploi, état idéal pour une bonne description du marché du travail.

Fort de tout ce qui précède, on peut intuitivement remarquer que, malgré l'objectivité de sa définition, le taux de chômage BIT ne traduit pas totalement la situation de l'économie. Certes il permet de contrôler l'offre et la demande de travail dans un système économique aussi complexe soit-il, mais il ne représente que la partie visible de "l'iceberg" et, de ce fait, nécessite l'appui d'autres indicateurs complémentaires pour décrire efficacement le marché du travail.

Face à cette situation, les statisticiens du travail ont introduit en 1925, lors de leur 2<sup>ème</sup> conférence internationale, un autre indicateur prenant en compte la durée de travail, les compétences du travailleur et le revenu de ce dernier. Ce nouvel indicateur, appelé "taux de sous-emploi", est destiné à accompagner le taux de chômage dans une interprétation intelligente des situations du marché du travail.

Étant donné qu'au Cameroun, l'enquête EESI de 2005 a fourni un taux de sous-emploi global supérieur à 75% de la population active, il est tout à fait légitime de s'interroger sur les contours de ce phénomène. Il serait donc judicieux de mener l'étude du marché du travail au Cameroun plus sous l'angle du sous-emploi que sous celui du chômage.

## 1.1 Objectifs de l'analyse du sous-emploi

### 1.1.1 Objectif principal

L'objectif principal de ce mémoire est de contribuer à l'amélioration de l'analyse des problèmes de l'emploi au Cameroun, en se focalisant sur les caractéristiques et les déterminants du sous-emploi, et l'élaboration des stratégies et mesures à court et à long terme, dans l'optique de promouvoir le plein emploi, productif et librement choisi.

### 1.1.2 Objectifs spécifiques

Spécifiquement, cette étude permettra de :

1. établir le profil des personnes en situation de sous-emploi ;
2. identifier les déterminants du sous-emploi visible, invisible et global, dans le but d'en tirer des recommandations pour les politiques en faveur de l'emploi ;
3. proposer des axes d'amélioration.

Ce mémoire s'organise en sept chapitres. Le premier chapitre est introductif, et indique le taux sous-emploi comme indicateur de relais pour le taux de chômage. Le chapitre 2 décrit la structure de l'Institut National de Statistique en insistant bien plus sur les tâches assignées. Dans le troisième chapitre, nous définissons les concepts et présentons les différentes formes de sous-emploi. La revue de la littérature y est aussi. Le quatrième chapitre passe en revue l'enquête EESI 2005, notamment à travers la technique de recueil et la présentation des données. Le chapitre 5 porte sur la méthodologie de l'analyse du sous-emploi, en exposant le matériel statistique approprié. Le chapitre 6 est l'analyse proprement dite, où l'on détermine les éléments essentiels expliquant le sous-emploi au Cameroun. C'est ici qu'intervient la modélisation. Compte tenu de la nature du phénomène à étudier, nous construisons deux modèles logistiques (un pour le sous-emploi invisible et l'autre pour le sous-emploi visible) destinés à classer un individu en état de sous-emploi ou pas. Il s'achève par l'interprétation des résultats. Le septième chapitre porte essentiellement sur la conclusion et les recommandations. A partir des aspects dégagés dans le chapitre 6, nous proposons une démarche à suivre pour aller du sous-emploi vers le plein emploi librement choisi. La dernière partie dénommée annexe, concerne les tableaux de construction des indicateurs et les tableaux croisés.

---

# PRÉSENTATION DE LA STRUCTURE D'ACCUEIL

---

Dans le cadre de la formation pratique qui s'inscrit parmi les activités du Master 2 de Statistique Appliquée, de l'École Nationale Supérieure Polytechnique (ENSP), nous avons été amené à effectuer un stage académique s'étendant sur une période de quatre mois ( du 22 mai 2006 au 15 septembre 2006 ) à l'Institut National de Statistique (INS). Ce stage vise un double objectif : celui de la mise en pratique des connaissances acquises à l'école, mais aussi de nous accommoder à la réalité du milieu professionnel.

L'objet du présent chapitre est de décrire l'INS et ses différentes structures ; plus précisément, la Sous-Direction des Etudes et de la Normalisation Statistique (ENS).

## 2.1 Description de la structure d'accueil

L'Institut National de Statistique est un établissement public et administratif créé en 2001 par décret présidentiel N° 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 :

- d'assurer la coordination des activités du système national d'information statistique ;
- 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).

Notre stage s'est déroulé à de l'ENS. Placé sous l'autorité d'un Sous-Directeur, cette Sous-Direction a pour missions :

- le développement des méthodologies, des enquêtes et autres études statistiques ;
- les études relatives à l'adéquation de la production nationale des statistiques aux besoins des différents catégories d'utilisateurs ;
- l'actualisation de la réglementation statistique ;
- la centralisation de l'information sur les programmes et projets statistiques des administrations et organismes publics ;
- l'instruction des dossiers de demandes de visas statistiques ;
- la préparation des dossiers et de tous les éléments concourant au bon déroulement des sessions du conseil national de la statistique ;
- la centralisation des informations sur la conjoncture ;
- la publication et la diffusion des statistiques de la direction ;
- la gestion des fonds documentaires et des archives de la direction.

Pour faciliter son bon fonctionnement, l'ENS est organisée en trois services : le Service des Etudes Statistiques, le Service de la Normalisation Statistique et le Service des Publications, de la Diffusion, de la Documentation et des Archives. Chaque service est placé sous l'autorité d'un chef service.

### 2.1.1 Le Service des Etudes Statistiques

Il comprend deux Bureaux : le Bureau des études et méthodes, et le Bureau des programmes.

Les missions de ce service sont les suivantes :

- développement des méthodes applicables aux travaux statistiques, et ce suivant les services concernés ;
- préparation des projets de plans de développement à moyen terme de la statistique en fonction des orientations générales arrêtées en la matière, et l'établissement des bilans d'exécution de ces plans ;

- préparation, avec les autres services de l'INS, des projets de programmes annuels et pluriannuels de travaux statistiques de l'ensemble des administrations et organismes publics et du suivi de leur réalisation ;
- l'analyse des éventuelles insuffisances des systèmes d'information statistiques, notamment sur le plan de la méthodologie ;
- l'instruction des dossiers de demandes de visas statistiques.

### 2.1.2 Le Service de la Normalisation Statistique

Cet autre Service comprend aussi deux Bureaux : le Bureau des nomenclatures et le Bureau de l'organisation des fichiers. Ses missions s'organisent autour de deux points principaux :

- l'élaboration et l'adaptation des nomenclatures et des classifications internationales aux besoins et aux spécificités du pays ;
- le suivi, en liaison avec les services administratifs concernés, de l'évolution des fichiers administratifs, la proposition des éléments d'organisation utiles à leur exploitation statistique en veillant à la sauvegarde du secret statistique.

### 2.1.3 Le Service des Publications, de la Diffusion, de la Documentation et des Archives

Ce dernier Service est composé de quatre Bureaux :

- ▲ le Bureau de publication ;
- ▲ le Bureau de la diffusion ;
- ▲ le Bureau de la documentation ;
- ▲ le Bureau des archives,

qui lui permettent de pouvoir assumer ses multiples obligations. En fait, il est chargé :

- de la confection et de la diffusion des publications de l'INS ;
- de la confection de l'annuaire statistique et des bulletins de statistiques générales ;
- de la participation au développement de la diffusion des données statistiques ;
- de la gestion du dépôt légal des données issues des enquêtes ayant bénéficié du visa statistique ;
- de la gestion des fonds documentaires de l'INS ;
- de l'organisation et de la gestion des archives de l'INS.

---

# HISTORIQUE ET DÉFINITION DES CONCEPTS DU SOUS-EMPLOI

---

Depuis son introduction en 1925 parmi les indicateurs du marché de travail, le sous-emploi est un concept dont le sens évolue au fil des années. Cette évolution est due à la diversité des éléments visant l'intégralité de sa définition et qui sont susceptibles d'empêcher le plein emploi. Nous nous référons ici à la revue internationale du travail intitulée "*PERSPECTIVES, du nouveau pour les statistiques du travail*", rédigée par Patrick Bollé en 1999.

## 3.1 Historique

Le problème de la définition et de la mesure du sous-emploi a figuré à l'ordre du jour de plusieurs *Conférences Internationales des Statisticiens du Travail* (CIST). Les statisticiens du travail ont posé le problème du sous-emploi lors de la 2<sup>ème</sup> CIST, en 1925. La nécessité même de la mesure du sous-emploi est plutôt mentionnée explicitement à la 6<sup>ème</sup> CIST en 1947, mais jusque là, aucune définition ni méthode de mesure n'est arrêtée à ce sujet. C'est en 1954, pendant la 8<sup>ème</sup> CIST qu'est proposée la toute première définition du sous-emploi : *le sous-emploi visible* faisant état d'une période de référence de courte durée, d'horaires de travail réduits et de la volonté du travailleur d'effectuer un plus grand nombre d'heures de travail. Toutefois, ce n'est qu'en 1957 que la 9<sup>ème</sup> CIST a adopté la première définition statistique internationale du sous-emploi et a jeté les bases des directives internationales actuelles. Celles-ci figurent dans deux résolutions :

- la résolution concernant la mesure et l'analyse du sous-emploi et la sous-utilisation des ressources de main-d'œuvre, adoptée par la 11<sup>ème</sup> CIST en 1966 ;
- la résolution concernant les statistiques de la population active, de l'emploi, du chômage et du sous-emploi, qui modifiait certaines parties de la résolution précédente et qui a été adoptée par la 13<sup>ème</sup> CIST en 1982.

Il faut remarquer qu'en dehors du sous-emploi visible, il existe d'autres formes de sous-emploi que l'on regroupe sous l'appellation *sous-emploi invisible* et qui reflète une répartition inadéquate des

ressources en main d'œuvre ou un déséquilibre fondamental entre la main d'œuvre et les autres facteurs de production.

Le sous-emploi dans un pays, quelle que soit sa forme, dépend fortement des conditions de son marché du travail et, par conséquent, les circonstances nationales influencent nécessairement la décision de sa mesure.

Jusqu'en 1997, la définition du sous-emploi reste applicable uniquement suivant la seule dimension de sous-emploi visible et manque même de précision notamment au niveau des comparaisons internationales. Face à cela, une réunion d'experts est organisée visant un double objectif : remettre à jour la définition du sous-emploi visible et adopter une définition concertée des autres formes de sous-emploi, remplaçant celle du sous-emploi invisible et fondée sur des critères se prêtant à une mesure dans les enquêtes auprès des ménages. Ceci va faciliter la tâche à la Commission du sous-emploi de la 16<sup>ème</sup> CIST à classer le sous-emploi dans le cadre de la main d'œuvre (parce que les personnes sous-employées subissent généralement les mêmes contraintes que les chômeurs et tendent à se comporter comme eux) et de le scinder en deux grands groupes tels que stipulés dans le rapport de la 16<sup>ème</sup> CIST : le sous-emploi lié à la durée du travail et les situations d'emploi inadéquat.

## 3.2 Définitions des concepts et critères d'éligibilité

Clairement, la définition du sous-emploi préconisée par le BIT englobe toutes les personnes pourvues d'un emploi, salarié ou non, qu'elles soient au travail ou absentes du travail, qui travaillent involontairement moins que la durée normale du travail dans leur activité et qui étaient à la recherche d'un travail supplémentaire ou disponibles pour un tel travail durant la période de référence (OIT, 1998).

### 3.2.1 Sous-emploi lié à la durée du travail

On dit qu'une personne employée est victime du sous-emploi lié à la durée du travail lorsque sa durée de travail est insuffisante par rapport à une autre situation d'emploi possible que cette personne est disposée à occuper et disponible pour le faire.

#### Critères d'éligibilité :

Une personne pourvue d'un emploi est concernée par le sous-emploi lié à la durée du travail si celle-ci vérifie simultanément les trois conditions suivantes :

- elle est *disposée à faire davantage d'heures* : c'est-à-dire qu'elle souhaite (en recherchant activement) : soit accroître son volume horaire de travail dans son emploi actuel, soit occuper un ou plusieurs autres emplois en plus de son emploi actuel en vue d'effectuer davantage d'heures de travail, soit encore substituer à son emploi actuel un ou plusieurs autres emplois qui lui permettraient de travailler plus ;
- elle est *disponible pour faire davantage d'heures* dans une *période ultérieure*. Même si la période ultérieure n'est pas spécifiée, elle devrait tenir compte, selon les circonstances nationales, de la période (préavis) dont ont généralement besoin les travailleurs pour quitter un emploi et en commencer un autre ;
- elle a effectué, pendant la période de référence, un *volume horaire de travail*, tous emplois confondus, *inférieur à un seuil relatif à la durée du travail effectué* à choisir selon les circonstances locales. Ce seuil peut être défini, par exemple, par rapport à la distinction entre emploi à plein temps et emploi à temps partiel, aux valeurs médianes, moyennes, ou aux normes relatives aux heures de travail telles que spécifiées par la législation. Dans la pratique, on utilisera en priorité les conventions nationales (législation, pratique courante, normes conventionnelles).

Parmi les méthodes de mesure du sous-emploi visible, *le temps partiel subi* est utilisé par bon nombre de pays, et se caractérise par la volonté du travailleur à temps partiel d'accroître son volume horaire de travail.

### 3.2.2 Situations d'emploi inadéquat

Il s'agit de situations de travail qui réduisent les compétences, les aptitudes et le bien-être des travailleurs par rapport à un autre emploi.

#### Différentes formes d'emploi inadéquat :

Même si la 16<sup>ème</sup> CIST reconnaît que les méthodes statistiques nécessaires pour décrire de telles situations demandent encore à être développées, elle reconnaît tout de même trois types de situations particulières :

- l'emploi inadéquat lié aux *qualifications* est caractérisé par une utilisation inadéquate ou insuffisante des qualifications professionnelles. Les personnes se trouvant dans cette forme d'emploi inadéquat sont celles qui, durant la période de référence, désirent ou cherchent à changer leur emploi actuel pour un autre leur permettant d'utiliser plus pleinement leurs qualifications ;
- l'emploi inadéquat lié au *revenu* (on parle dans ce cas de sous-emploi invisible) comprend les personnes gagnant un revenu inférieur à un seuil fixé par les circonstances nationales, et qui

désireraient ou chercheraient à changer leur emploi actuel pour un autre en vue d'accroître leur revenu ;

- l'emploi inadéquat lié à un *volume horaire de travail trop élevé* se réfère à une situation où une personne désirerait ou chercherait à faire moins d'heures de travail (sans aller en dessous d'un plancher fixé selon les circonstances nationales) qu'elle n'en avait faites pendant la période de référence, soit dans le même emploi, soit dans un autre emploi avec une réduction correspondante du revenu.

### 3.3 Définitions de quelques indicateurs du marché de travail utilisés

Ici, nous présenterons toutes les définitions (du point de vue de la mesure) des indicateurs BIT du marché de travail, telles que formulées dans le Rapport de la 16<sup>me</sup> CIST en 1998.

Pour la *population en âge de travailler*, le BIT propose celle dont les âges sont supérieurs à 15 pendant la période de référence, tout en donnant une marge de liberté aux différents pays afin de pouvoir l'adapter à leur propre contexte juridique ou socio-économique. Dans le cadre conceptuel de notre étude, nous la prendrons à partir de 10 ans pour la simple raison que peu d'enfants travaillent à cet âge.

La *population active* va représenter non seulement les actifs occupés, mais aussi les chômeurs au sens du BIT.

Un *chômeur* est entendu au sens de chômeur BIT comme une personne ayant dépassé l'âge de 10 ans (population en âge de travailler), qui, pendant de la période de référence, était à la fois :

- sans travail, c'est-à-dire n'était pas pourvue d'un emploi, salarié ou non salarié, au cours de la période de référence (une semaine) ;
- disponible pour travailler dans un emploi, salarié ou non, durant la période de référence (deux semaines) ;
- à la recherche d'un travail, c'est-à-dire avait pris des dispositions spécifiques au cours d'une période récente spécifiée (quatre dernières semaines ou douze derniers mois) pour chercher un emploi salarié ou non.

Le *taux d'activité au sens du BIT* est le rapport de la population active sur la population en âge de travailler.

Le *taux de chômage BIT* est le rapport du nombre de chômeurs BIT sur la population active.

Le *taux de sous-emploi invisible* est le rapport du nombre d'actifs occupés dont le revenu mensuel

est inférieur au salaire minimum, sur la population active occupée.

Le *taux de sous emploi visible* est le rapport du nombre d'actifs occupés travaillant involontairement moins de 35 heures par semaine (on admet qu'ils doivent travailler 7 heures par jour ouvrable) sur la population active occupée.

Le *taux de sous-emploi global* désigne le rapport du nombre de chômeurs BIT, et d'actifs occupée en situation de sous-emploi (visible comme invisible) sur la population active.

### 3.4 Revue de la littérature

La mesure du sous-emploi n'est naturellement pas régulée de la même manière d'un pays à l'autre. En France par exemple, sur la base de la définition du BIT, les résolutions de l'Institut National de la Statistique et des Études Économiques (INSEE) répartissent les personnes en sous-emploi en trois groupes : celles qui travaillent à temps partiel tout en recherchant un emploi pour travailler davantage, celles qui travaillent à temps partiel sans rechercher un autre emploi mais en souhaitant travailler davantage et en restant disponibles pour le faire, et enfin celles travaillant à temps complet ayant involontairement travaillé moins que d'habitude, c'est-à-dire ayant subi une situation de chômage technique, partiel ou de grève.

Cette dernière catégorie est souvent négligée, au profit des deux premières (personnes travaillant à temps partiel) et qui se résument sous le nom de sous-emploi lié au *temps partiel subi* (dans la mesure où il est imposé). Selon la publication **INSEE PREMIERE** [5], ce type de sous-emploi a été estimé en 2005, à près de 95% de l'ensemble des actifs occupés, travaillant sous le régime à temps partiel. Il touche en majorité les femmes car elles pratiquent plus le temps partiel que les hommes (30% contre 5% pour les hommes). Dans la tranche d'âge 25-49 ans, le sous-emploi lié au temps partiel subi affecte près de 40% des hommes parce que ces derniers ne trouvent pas toujours un travail à temps plein ; pourtant, il est d'environ 50% chez les femmes. Effectivement, qu'elles souhaitent ou non travailler davantage, les femmes à temps partiel sont globalement moins disponibles pour le faire (elles doivent s'occuper de leurs enfants ou d'un membre de leur famille), condition nécessaire à la caractérisation d'une situation de sous-emploi. Sommairement, le sous-emploi s'observe plus chez les femmes (80%), les jeunes de moins de 25 ans, les non-diplômés et les étrangers en France.

Également, le sous-emploi lié au temps partiel est aussi la statistique de mesure du sous-emploi en Wallonie, région méridionale de la Belgique. Selon la 7ème publication du *Service des Etudes Statistiques* de 1999 [9], le sous-emploi féminin est très fortement lié au temps partiel subi (85% en 1986, par rapport au temps partiel proprement dit) et représente pratiquement le double du sous-

emploi masculin. Bref dans la période de 1981 à 1998, le taux de sous-emploi wallon n'excède pas 11% et admet pour déterminants la gent féminine, le travail intérimaire.

Au Cameroun, les Enquêtes Camerounaises Auprès des Ménages (ECAM) I et II réalisées en 1996 et 2001 respectivement n'étaient pas compatibles avec une mesure du sous-emploi selon le BIT pour des raisons de questionnaires non adaptés. Mais, pour présenter entièrement tous les indicateurs du sous-emploi, les experts de l'Institut National de Statistique (INS) ont dû faire une hypothèse simple et vraisemblable pour classer les individus en sous-emploi. Elle se fonde sur le fait que les individus aspirent généralement à une meilleure situation professionnelle. Puisque la rémunération ici est jugée insuffisante aussi bien dans le public que dans le privé, les individus sont obligés d'exercer plusieurs activités. La législation camerounaise est d'ailleurs souple à ce sujet, notamment pour les fonctionnaires. On se rapproche ainsi du sous-emploi BIT. Cependant, ces fonctionnaires, bien que n'étant pas satisfaits de leurs conditions, ils sont moins tentés, aversion pour le risque oblige, par une expérience dans le privé. Ainsi est-il difficile de classer les fonctionnaires dans le sous-emploi. Suivant cette méthode de mesure, le sous-emploi au Cameroun a été estimé à plus d'un tiers de la population active, selon une publication de l'INS [7], de décembre 2004 .

La littérature la plus récente sur le sous-emploi au Cameroun est celle de l'enquête EESI de 2005. De par la nature du questionnaire qui a servi à la collecte, celle-ci présente les données beaucoup plus fiables sur l'emploi et le secteur informel, par rapport aux données d'ECAM I et II. Ici, le taux de sous-emploi visible correspond à la proportion des actifs occupés de plus de 10 ans, travaillant moins de 35 heures par semaine, et s'estime à près de 12,12%, s'il faut se ramener au Rapport Principal de l'EESI [2]. Le sous-emploi invisible (ou situation d'emploi inadéquat, lié au revenu) quand à lui, se mesure en terme de pourcentage des actifs occupés de plus de 10 ans, qui gagnent moins de 23 500 FCFA dans leur activité principale. Le même Rapport Principal indique 69,30% comme taux de cet indicateur. Le sous-emploi invisible frappe davantage les femmes que les hommes, et se manifeste dans le secteur informel agricole. Les jeunes et les plus âgés gardent les taux les plus élevés (plus de 70%). Le niveau d'instruction est une caractéristique importante du sous-emploi invisible, et s'illustre en majorité par les non scolarisés, qui demeurent les plus exposés.

# TECHNIQUES DE RECUEIL ET PRÉSENTATION DES DONNÉES UTILISÉES

---

## 4.1 Justification de l'enquête

Avant l'enquête EESI, l'Institut National de la Statistique (INS) avait déjà réalisé les Enquêtes Camerounaises Auprès des Ménages (ECAM I en 1996 et ECAM II en 2001) à l'échelle nationale, mais celles-ci ne portaient que sur quelques aspects de l'emploi. La première enquête véritable sur l'emploi, le secteur informel et la consommation des ménages faite au Cameroun remonte à l'an 1993 et n'a couvert que la ville de Yaoundé. Donc il était tout à fait indiqué d'organiser une opération sur tout le triangle national en vue d'explorer davantage le marché du travail et les conditions d'activité dans le secteur informel, identifié par ailleurs comme une tanière de pauvres. Tel est l'opportunité qu'offre l'enquête EESI 2005 et dont les résultats vont permettre d'affiner le diagnostic des implications des questions de l'emploi sur la pauvreté et de poser les bases du suivi et de l'évaluation des programmes et politiques arrêtés dans ce domaine.

L'enquête EESI est une variante du système d'enquête 1-2-3 (outil statistique développé par l'organisme DIAL permettant de mener une étude sur la pauvreté. Ce système d'enquête a été utilisé pour la première fois en Afrique dans la ville de Yaoundé en 1993), enquête pour laquelle la phase 3 n'a pas été réalisée.

## 4.2 Techniques de recueil des données

L'enquête EESI a été réalisée en 2005 sur toute l'étendue du territoire national, précisément au niveau des départements du pays. Elle a été faite en deux phases : la phase 1 qui a recueilli de manière aléatoire un échantillon de 8 536 ménages en s'intéressant aux individus de 10 ans et plus. Elle a utilisé deux formulaires, dont un qui indexait le ménage tout entier et ses caractéristiques sociodémographiques, et l'autre qui se focalisait sur chacun des individus du ménage, leurs conditions d'activité et les données liées à l'insertion sur le marché du travail. La phase 2 qui concerne le Secteur

Informel.

L'échantillon est issue de 12 régions : les 10 provinces administratives du Cameroun et les deux métropoles Yaoundé et Douala mises à part à cause de leur importance. Chaque région est divisée en trois différentes strates (urbain, semi-urbain et rural) en prenant en compte les degrés de variabilité des phénomènes de l'emploi et du comportement du secteur informel. Les deux grandes villes citées sont dorénavant prises comme urbaines. Ce qui fait un bilan de 32 strates dont 10 rurales, 10 semi-urbaines et 12 urbaines

La nécessité d'explorer le milieu rural vient du fait qu'il abrite plus de pauvres que les milieux urbain et semi-urbain ; et favorise l'accès aux activités informelles non agricoles comme l'élevage, l'artisanat

Notre travail n'exige que les données de la phase 1.

#### 4.2.1 Unité d'échantillonnage et d'observation

A la première phase, l'enquête EESI retient comme unité statistique d'investigation le ménage ordinaire et ses différents membres âgés de 10 ans ou plus. Mais dans le cadre de l'étude du sous-emploi, l'unité statistique sera un membre du ménage âgé de 10 ans ou plus, sur qui on a récolté des informations socio-économiques.

#### 4.2.2 Méthodologie de l'enquête

##### Base de sondage

Elle est formée des zones de dénombrement (ZD) issue des travaux cartographiques menés en 2003 au Cameroun, pour le Troisième Recensement Général de la Population et de l'Habitat (RGPH3). Une ZD est entendue ici comme une portion du territoire délimitée par des détails visibles et renfermant en principe entre 700 et 1100 habitants, soit entre 140 et 220 ménages en moyenne. Le territoire national se trouve ainsi divisé en 17 800 ZD qui constituent les unités de la base.

##### Taille de l'échantillon

Pour se donner la taille de l'échantillon, les intérêts ont été plus portés sur les niveaux des analyses attendues que sur la précision des résultats souhaités. C'est la raison pour laquelle on a eu recours aux coefficients d'extrapolation pour calculer les différents indicateurs du marché de travail. Pour une strate donnée, on se propose de disposer d'un minimum de 130 ménages. Toutefois, afin d'éviter les cas de refus ou toute autre forme d'échec, au moins 150 ménages ont été enquêtés par strate.

Enfin, dans le souci de mieux saisir la diversité des emplois, 600 ménages ont été enquêtés par province, 1200 ménages à Yaoundé et 1400 ménages à Douala ; ce qui fait un cumul d'environ 8 600 ménages enquêtés pour la phase 1.

### Plan de sondage

Il est stratifié en deux ou trois degrés selon le milieu de résidence. Dans le milieu urbain constitué des localités de 50 000 habitants ou plus, on a fait un tirage aléatoire et à probabilités égales au premier degré des unités primaires, et au second degré des ménages après un dénombrement exhaustif de ces unités primaires. Dans le milieu rural constitué des localités de moins de 50 000 habitants, le tirage s'est fait à trois degrés. D'abord on a tiré au premier degré à probabilités inégales les localités d'enquête, puis au deuxième degré les unités primaires et au troisième degré les ménages.

Le détail du calcul des probabilités de tirage suivant les différents milieux est disponible dans le *Document de méthodologie de l'EESI* [1]. Le calcul des coefficients d'extrapolation de la zone de dénombrement  $K$  (inverse des probabilités de tirage) est donné par :

- pour les strates Yaoundé et Douala :  $C_{1k} = Z_{ij}E_{ijk}/z_{ij}e_{ijk}$  ;
- pour les strates urbaines (grandes villes) :  $C_{2k} = Z_iE_{ik}/z_ie_{ik}$  ;
- pour les strates semi-urbaines (petites villes) :  $C_{3k} = M_iZ_{ij}E_{ijk}/m_{ij}z_{ij}e_{ijk}$  ;
- pour les strates rurales :  $C_{4k} = M_iZ_{ij}E_{ijk}/m_{ij}z_{ij}e_{ijk}$  ;

avec les notations suivantes :

- $m_{ij}$  : nombre de ménages dénombrés dans la partie rurale de l'arrondissement  $j$  de la province  $i$  ;
- $M_i$  : nombre total de ménages de toutes les localités rurales dans la province  $i$  ;
- $z_{ij}$  : nombre de ZD tirées dans la partie rurale de l'arrondissement  $j$  de la province  $i$  ;
- $Z_{ij}$  : nombre total de ZD dans la partie rurale de l'arrondissement  $j$  de la province  $i$  ;
- $e_{ijk}$  : nombre de ménages effectivement tirés et enquêtés dans la ZD  $k$  d'arrondissement  $j$  dans la province  $i$  ;
- $E_{ijk}$  : nombre total de ménages dénombrés dans la ZD  $k$  de l'arrondissement  $j$  dans la province  $i$ .

## 4.3 Présentation des données

La base de données se présente sous forme d'un tableau rectangulaire de 38599 individus contre 234 variables. Les variables sont de plusieurs types : qualitatif, quantitatif, catégoriel. Ce tableau ne

comporte pratiquement pas de données manquantes.

Pour la phase 1 de l'enquête, nous disposons de données regroupées par modules, avec leurs objectifs et les variables correspondantes comme le montre la structure du questionnaire, en fin de ce document. L'étude du sous-emploi exclut nécessairement certaines de ces modules tels que le module Chômage (le chômage au sens du BIT tient compte, à priori, de ceux qui sont sans travail), du module TP (Trajectoire et Perspectives) qui traite des problèmes de la mobilité professionnelle et permet de recueillir des informations sur l'emploi antérieur du père de l'enquêté et des perspectives de l'emploi futur des membres du ménage, en âge de travailler. Les modules RHE (Revenu Hors Emploi) et EH (Emploi Habituel) ne sont pas présents dans la base de données, et le module SE (Situation d'emploi). L'absence du module SE est due au fait qu'elle avait pour rôle de classer les individus par module, au cours de l'interrogatoire.

Concrètement, les variables que nous présenterons seront issues, en plus des caractéristiques sociodémographiques, des modules Activité Principale, Activité Secondaire et Recherche d'emploi (cet autre module s'intéresse à ceux qui ont déjà un emploi).

### 4.3.1 Analyse descriptive

La population totale de la base est de 38599 enquêtés. Les données que nous allons présenter sont relatives aux actifs occupés qui représentent une quinzaine de milliers parmi ces enquêtés.

#### Caractéristiques sociodémographiques :

- Le **sexe** : En 2005 au Cameroun, les actifs occupés sont presque également répartis suivant le sexe, avec une proportion légèrement plus élevée pour les hommes (54,3% pour les hommes et 45,7% pour les femmes).
- L'**âge** : La population active occupée est jeune, avec un âge moyen de 33,59 ans. Mais cette moyenne présente une très forte variabilité (l'écart type de la variable âge est de 14,6). De plus, l'âge semble se distribuer suivant une gaussienne, comme le montre la figure *FIG.4.1*

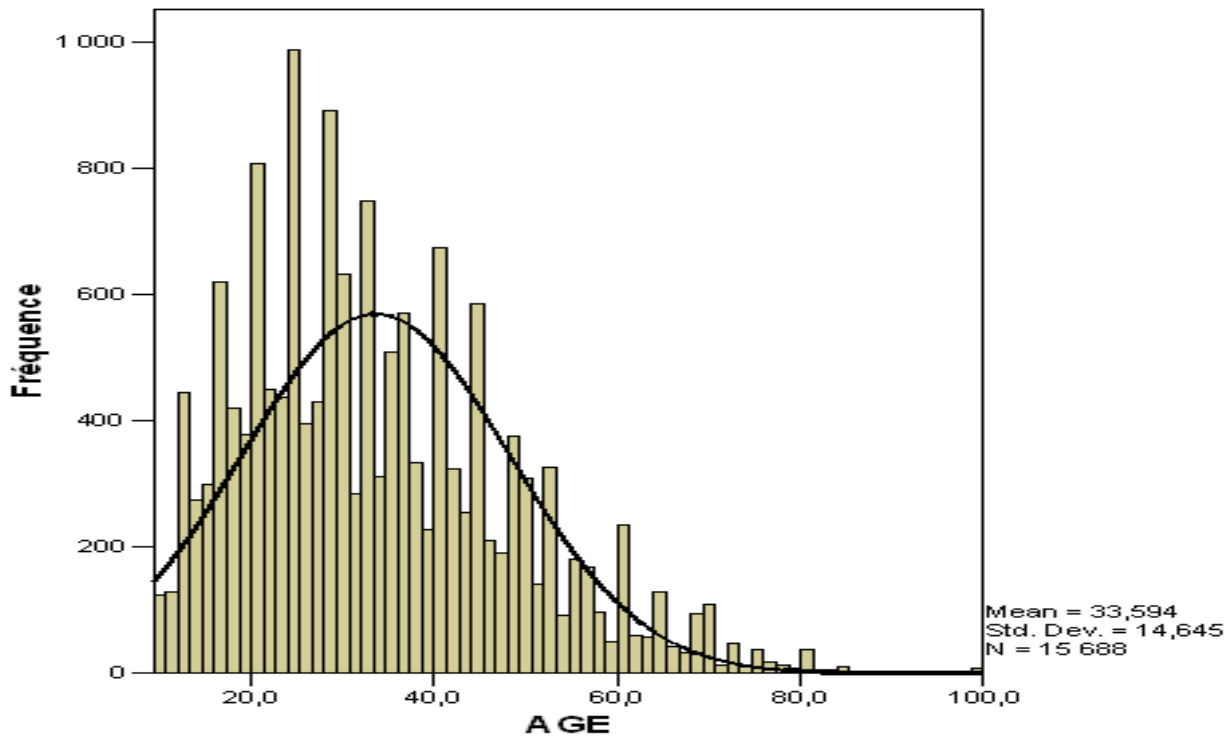


FIG. 4.1 – Distribution de l'âge

- Le statut matrimonial est largement dominé par les modalités "célibataire" (38,7%) et "monogame" (37%). Tout de même, on a 8,9% de polygames, 5,5% de veufs, 3% de divorcés. Le concubinage s'élève ici à 6,8% des actifs occupés.
- Le **niveau d'instruction**

TAB. 4.1 – Répartition des actifs occupés par niveau d'instruction.

Niveau d'instruction	Pourcentage
Non scolarisé	17,2
Primaire	39,5
Secondaire 1 cycle	25,0
Secondaire 2 cycle	12,4
Supérieur	5,8
Total	100,0

- La **région** : Le diagramme en bâtons ci-dessous (*FIG.4.2*) montre que le plan de sondage et la définition des zones de dénombrement ont dans l'ensemble atteint leurs objectifs, qui était de permettre un échantillonnage de même taille pour chaque ville, en donnant un peu plus de poids aux deux villes Yaoundé et Douala et aux provinces de l'extrême-Nord et de l'Ouest.

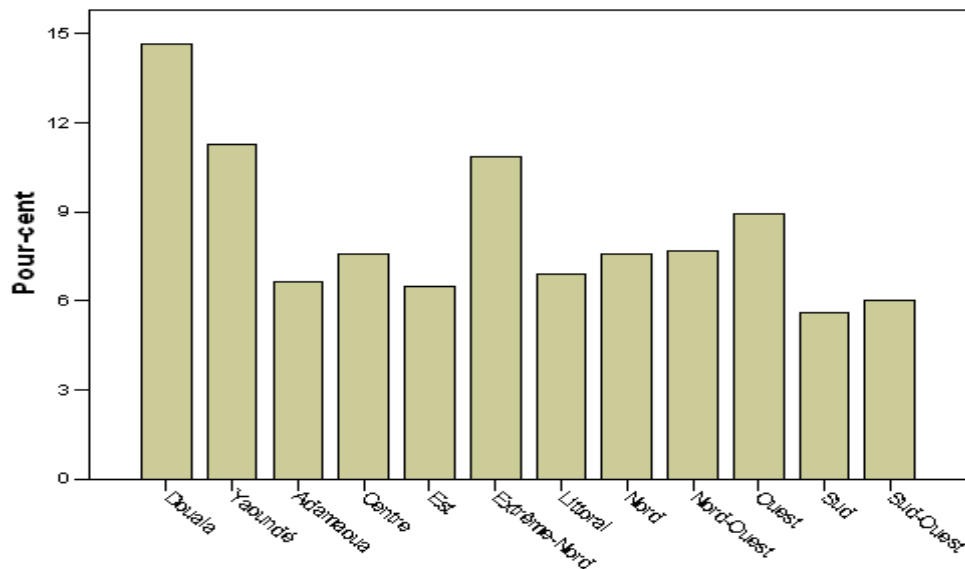


FIG. 4.2 – Répartition des actifs occupés par région d'enquête.

– **Organisation administrative :**

TAB. 4.2 – Répartition des actifs occupés suivant l'Organisation administrative.

Organisation administrative	Pourcentage
Chef-lieu de province	18,96
Chef-lieu de département	19,8
Chef-lieu d'arrondissement	22,52
Village	38,0
Données manquantes	0,71
Total	100,0

- **Religion :** Dans la population considérée, les religions catholique (40,8%), protestante (28,4%) et l'islam (20,2%) sont les plus en vue, contrairement aux autres chrétiens, animistes et païens qui n'y sont qu'à l'état de traces (moins de 2,3% chacun).

**Activités :**

- Le **zone d'activité** : l'activité au Cameroun est à moitié rurale (52,4%), l'autre moitié étant partagée presque équitablement entre le milieu semi-urbain (23,5%) et le milieu urbain (24,1%).
- Le **secteur institutionnel** : cette même activité est largement dominée par le secteur informel qui à lui seul occupe plus de 90% de l'ensemble. Plus précisément, 5 actifs sur 10 exercent dans l'informel non agricole et 4 sur 10 dans l'informel agricole. Cependant, le public et le privé formel représente environ chacun les 5% de l'activité.

- La **catégorie socioprofessionnelle** : ici, le travail à propre compte (48,4%) se définit comme mode de la distribution, suivi des aides familiales (17,1%). Les autres rangs professionnels (cadre supérieur, ingénieur et assimilé, cadre moyen, employé/ouvrier qualifié et semi-qualifié, employeur, manœuvre, apprenti, et inclassable) y sont faiblement représentés (moins de 8%).
- Le **taille de l'entreprise** : les structures d'une personne occupent 31,5% de l'activité, et confirment ce que prévoit la distribution de la catégorie socioprofessionnelle. De plus, on a 18,7% pour les entreprises de 2 personnes, 24,% pour les entreprises de 3 à 5 personnes, enfin les entreprises de plus de 6 personnes s'illustrent avec moins de 17%.
- Les variables comme **revenu d'activité principale**, **revenu d'activités secondaires**, et même **revenu d'activités principale et secondaires** sont de nature quantitative. Leurs histogrammes n'étant pas assez expressifs, il est plus utile de les observer dans le sommaire en annexe, dans le tableau *Tab.8.11*, avec respectivement les dénominations "revactp", "revacts" et "revact".
- **Recherche d'emploi** : de tous les actifs occupés au Cameroun en 2005, seuls 13% veulent changer de métier pour multiples raisons.
- **Motifs de la recherche d'emploi** : Effectivement, près de 58% souhaitent accroître leurs revenus, 26% recherchent un travail plus intéressant et environ 11% préfèrent un emploi sécurisant. Toutefois, seulement moins de 27% de ceux-ci sont prêts à augmenter leurs heures de travail dans le nouvel emploi désiré.
- **Obtention** : les moyens d'obtention d'un emploi qui s'avèrent les plus concrets sont : l'initiative personnelle (45,2%), les relations personnelles (40%), par contact direct avec l'employeur (7,6%), et par concours (5,5%). Chercher du travail à partir des petites annonces, par les structures appropriées (Fonds National de l'Emploi, le Bureau de la Main d'Oeuvre, les agences et associations privées) semblent très peu fructueuses (moins de 1,6%).
- **Recyclage** : La base de données de l'EESI montre que plus de 9 employés sur 10 ne bénéficient pas d'une formation ou d'un stage de recyclage financé par l'entreprise. Pour ainsi dire, l'activité au Cameroun s'appuie en majorité sur l'instruction ou sur la formation professionnelle académique. Ici, nous avons 3 données manquantes sur une quinzaine de milliers.
- **Promotion** : les personnes ayant reçu une promotion au sein de l'entreprise sont très peu nombreuses (5%). Ici, on a 4 données manquantes.
- **Mois de travail par an** : Dans leurs activités principales, un peu plus de 60% des employés travaillent pendant les 12 mois de l'année. Ce même chiffre avoisine 35% dans l'activité secondaire.

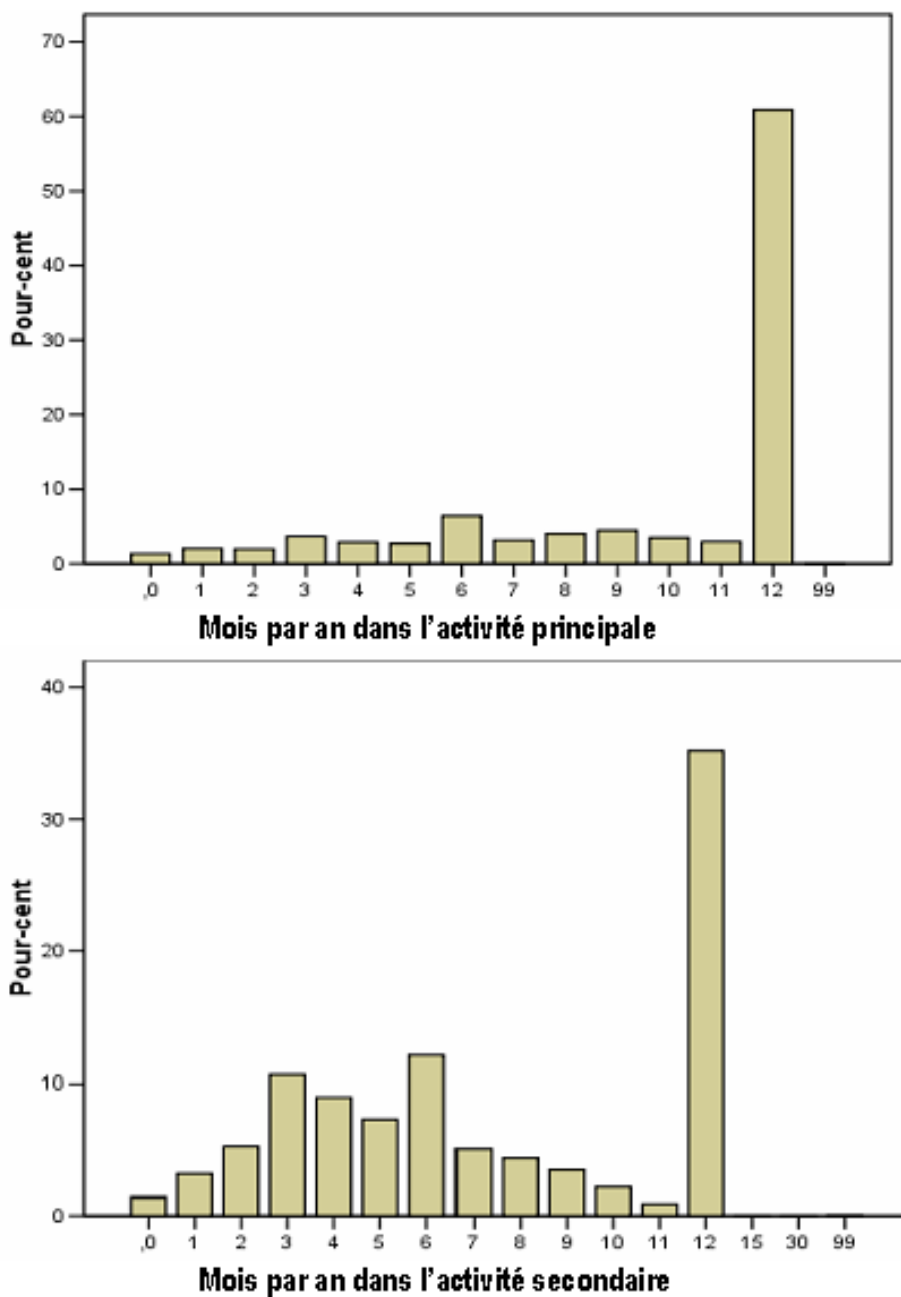


FIG. 4.3 – Nombre de mois de travail par an dans l'activité principale, et secondaire.

– **Jours de travail par semaine :**

TAB. 4.3 – Nombre de jours de travail par semaine, dans l'activité principale.

Nombre de Jours par semaine	0	1	2	3	4	5	6	7	Total
Pourcentage	0,6	2,7	4,7	6,3	8,6	19,12	35,5	22,3	100,0

– **Heures de travail par semaine :** Cette variable est quantitative, nous renvoyons sa description par la variable "ap10c" dans le sommaire de l'annexe au tableau *Tab.8.11*.

– **Durée supplémentaire :**

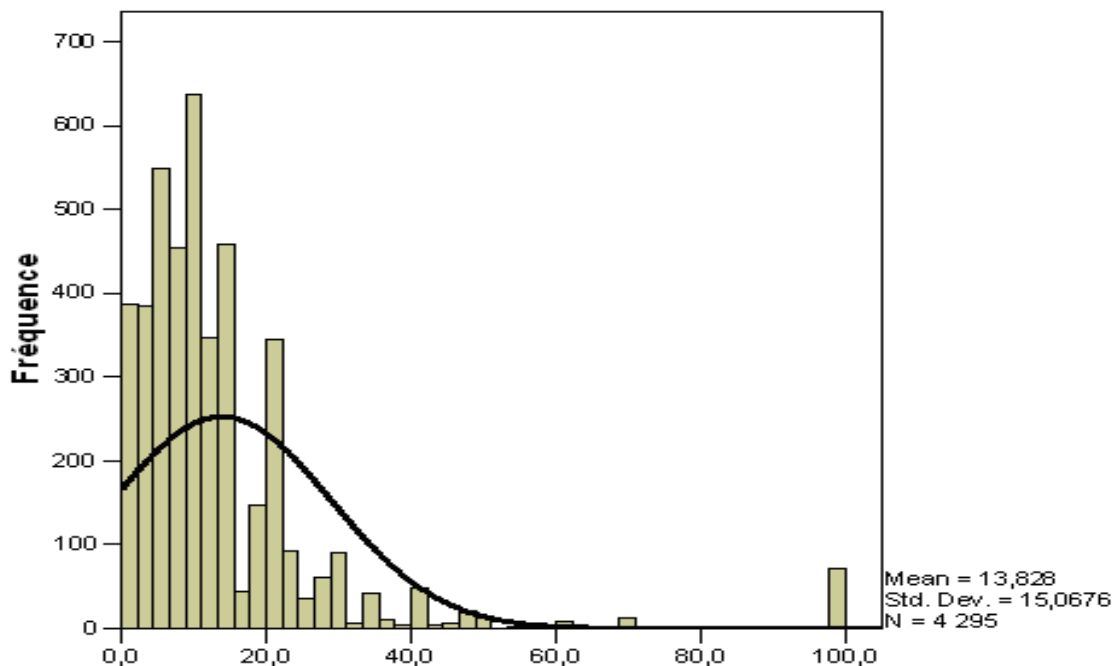


FIG. 4.4 – Distribution des heures supplémentaires.

### 4.3.2 Quelques chiffres du sous-emploi au Cameroun

Les résultats empiriques de la mesure du sous-emploi donnent une première idée de la courbe du sous-emploi suivant le milieu d'activité, le lieu de résidence, le sexe, l'âge... Par exemple, le sous-emploi invisible se fait plus ressentir dans le milieu rural qu'urbain. Une analyse plus détaillée suivant les régions montre que l'Extrême-Nord est la première victime du phénomène, avec 88 personnes sur 100 en situation de sous-emploi. A l'opposé, les deux métropoles Yaoundé et Douala enregistrent les taux de sous-emploi les plus bas : 29 et 31%.

D'une manière générale, le sous-emploi invisible se manifeste moins, d'une part chez les individus instruits que chez les non scolarisés, et d'autre part chez les acteurs du secteur formel (public et

privé) que chez ceux de l'informel (agricole et non agricole). Les chiffres sont bien présentés dans les tableaux *Tab 8.3*, *Tab 8.4*, *Tab 8.5* de la partie annexe. De prime abord, une question se pose sur la significativité de ces proportions. Autrement dit, est-ce que cet empirisme ne cache pas des aspects et des éléments nécessaires à l'appréciation du sous-emploi invisible ? Tel est l'objectif de l'étude qui est d'analyser les facteurs socio-économiques qui influencent le sous-emploi invisible au Cameroun. Ces facteurs sont analysés du côté du revenu mensuel et sur le plan micro-économique (au niveau individuel).

# FORMULATION LOGISTIQUE DU SOUS-EMPLOI

La régression est un procédé généralement utilisé en statistique pour modéliser l'information contenue dans une base de données chiffrées, au moyen d'une équation algébrique. Dans le cas d'une régression linéaire multiple, on recherche une relation entre une variable continue (dépendante, expliquée ou endogène) notée  $Y$  en fonction des observations (variables indépendantes, explicatives, exogènes ou régresseurs) notées  $X_i$ ,  $i = 1, \dots, k$  et du type quantitatif, soit :

$$Y = a_0 + AX + \varepsilon = a_0 + \sum_{i=1}^k a_i X_i + \varepsilon \quad (5.1)$$

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  $\varepsilon$  le terme erreur aléatoire, qui doit suivre une loi normale  $\mathcal{N}(0, \sigma^2)$ .

Cependant, dans de nombreux cas, la spécification du modèle linéaire est incorrecte, notamment dans le cas des modèles de probabilité où la variable dépendante est dichotomique (modèle logistique binaire *logit* ou *probit*) ou polytomique (modèle multinomial).

La régression logistique binaire permet d'étudier la relation entre une variable réponse binaire et plusieurs variables explicatives qui peuvent être quantitatives, nominales, catégorielles,...

Plus généralement, le résultat d'une observation binaire  $Y$  est "*succès*" lorsque  $Y = 1$  ou "*échec*" lorsque  $Y = 0$ , et exprime la survenue d'un événement. La régression logistique est très utilisée en science de la santé pour prédire l'occurrence des pathologies (cancer,...) et ne nécessite pas de postulats sur la distribution des données des prédicteurs ou la linéarité des relations.

## 5.1 Spécification du modèle logistique binaire

Au lieu d'estimer  $Y$ , l'objectif est d'estimer plutôt  $P(Y = 1/X) = p$ , la probabilité de succès de l'événement sachant  $X$ .

Puisqu'on cherche à prédire la valeur de  $Y$  connaissant celle de  $X$ , on va l'estimer par sa moyenne théorique conditionnelle  $E(Y/X)$ . Ainsi :

$$\begin{aligned} E(Y/X) &= 0 \times [1 - P(Y = 1/X)] + 1 \times P(Y = 1/X) \\ &= P(Y = 1/X) = p. \end{aligned} \quad (5.2)$$

Il se pose alors un problème important : une mesure de probabilité est toujours bornée à gauche par 0 et à droite par 1 ; donc le modèle linéaire (5.1) perd son sens. De plus, la régression linéaire simple suppose à priori que l'ensemble d'apprentissage  $\{(X_i = x_i ; Y_i = y_i)\}_{i=1}^n$  est indépendant et les  $X_i$  continues, et exige l'homoscédasticité de la variance des résidus. Ce qui est incompatible avec la nature de  $Y$  (qualitatif binaire) et celle des  $X_i$ .

Pour contourner la difficulté, il convient de trouver un modèle qui sied à une régression linéaire multiple et qui supprime les bornes de  $p$  : le modèle logistique.

### 5.1.1 La transformation *logit*

La nature binaire de la variable  $Y$  permet d'établir que, conditionnellement à  $X$ ,  $Y$  suit une loi binomiale de paramètre  $p$ , et admet la fonction  $h(y, p) = p^y(1-p)^{1-y}$ , comme fonction de densité. Cette fonction fait partir de la famille exponentielle. En effet,

$$\begin{aligned} h(y, p) &= p^y(1-p)^{1-y} \\ &= \exp\{y \ln(p) + (1-y) \ln(1-p)\} \\ &= \exp\left\{y \ln\left(\frac{p}{1-p}\right) + \ln(1-p)\right\} \\ &= (1-p) \exp\left\{y \ln\left(\frac{p}{1-p}\right)\right\}. \end{aligned}$$

Si l'écriture générale de la densité de probabilité d'une loi exponentielle se présente sous la forme :  $(y, p) \mapsto a(p)b(y)\exp\{yQ(p)\}$ , avec  $Q(p)$  son paramètre naturel, alors la quantité  $\ln\left(\frac{p}{1-p}\right)$  est le paramètre naturel de la loi binomiale.

Intuitivement, on définit la fonction *logit* comme étant l'application :

$$\begin{aligned} \text{logit} &: ]0, 1[ \longrightarrow ]-\infty, +\infty[ \\ p &\longmapsto \text{logit}(p) = \ln\left(\frac{p}{1-p}\right). \end{aligned} \quad (5.3)$$

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  $F$  :

$$\begin{aligned} F &: ]-\infty, +\infty[ \longrightarrow ]0, 1[ \\ t &\longmapsto F(t) = \frac{1}{1 + e^{-t}} = \frac{e^t}{1 + e^t}, \end{aligned} \quad (5.4)$$

$F$  est appelée *fonction logistique*, tandis que *logit* porte le nom de *fonction de lien*.

En s'appuyant sur cette transformation, nous proposons la construction du modèle suivant :

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta X + u ; \quad (5.5)$$

où  $\beta_0$  est une constante,  $\beta = (\beta_1, \dots, \beta_k)$  un vecteur de paramètres à estimer et  $u$  le terme erreur. Ainsi, à partir de la fonction logistique  $F$ , il est plus facile de retrouver la probabilité de succès  $p$  :

$$p = P(Y = 1/X) = F(\text{logit}(p)) = \frac{e^{\beta_0 + \beta X}}{1 + e^{\beta_0 + \beta X}}, \quad (5.6)$$

et ce en connaissant  $\beta_0$ ,  $\beta$  et  $X$ .

## 5.2 Estimation des paramètres du modèle

La première difficulté qui surgit est celle de l'inconnue  $p$ . La seule information que l'on dispose à son sujet est que l'événement s'est réalisé ou pas, suivant les valeurs des observations. En observant la formule (5.2), on remarque que  $p$  est une moyenne : c'est la proportion des individus malades (par exemple) et qui possèdent un vecteur commun de caractéristiques  $X$ . Une idée assez simple serait alors de regrouper les individus qui ont un même profil, de calculer cette proportion et de l'utiliser comme valeur de la probabilité. Ensuite, de régresser le logit par une méthode de Moindres Carrés Ordinaires (MCO).

Mais nous allons procéder autrement. Plutôt que de raisonner en deux étapes, on va appliquer la Méthode du Maximum de Vraisemblance (MMV).

### 5.2.1 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 :

$$y_i = \begin{cases} 1 & \text{si } p_i = F(\beta_0 + \beta x_i) \\ 0 & \text{si } p_i = 1 - F(\beta_0 + \beta x_i) \end{cases} \quad (5.7)$$

où  $x_i = (x_i^1, \dots, x_i^k)^T$ ,  $\forall i \in \{1, \dots, N\}$ , est un vecteur de caractéristiques observables sur l'individu  $i$ ;

$\beta_0$  un paramètre inconnu ;

$\beta = (\beta_1, \dots, \beta_k) \in \mathbb{R}^k$  est un vecteur de paramètres inconnus.

### Fonction de vraisemblance conditionnelle

Manifestement, conditionnellement aux  $x_i$ , les valeurs observées de  $y_i$  sont les réalisations d'un processus binomial avec une probabilité  $F(\beta_0 + \beta x_i) = p_i$ , compte tenu de la relation (5.7). La vraisemblance des échantillons associés à notre modèle logistique s'écrit comme celle des échantillons

associés aux modèles binomiaux ; à la seule différence que les probabilités  $p_i$  varient d'un individu à l'autre ( $p_i$  est fonction de la caractéristique  $x_i$ ).

Pour ainsi dire, la vraisemblance conditionnelle associée à l'observation  $i$  s'écrit :

$$L(y_i, \beta_0, \beta) = p_i^{y_i} (1 - p_i)^{1-y_i}.$$

Et donc la vraisemblance conditionnelle de l'échantillon  $y = (y_1, \dots, y_N)$  de taille  $N$  s'écrit sous la forme :

$$\begin{aligned} L(y, \beta_0, \beta) &= \prod_{i=1}^N p_i^{y_i} (1 - p_i)^{1-y_i} \\ &= \prod_{i=1}^N [F(\beta_0 + \beta x_i)]^{y_i} [1 - F(\beta_0 + \beta x_i)]^{1-y_i}. \end{aligned} \quad (5.8)$$

Ainsi, si les estimations des probabilités  $p_i$  sont en accord avec les observations, la vraisemblance sera maximisée. Il revient donc de chercher les valeurs de  $\beta_0$  et de  $\beta$  qui maximise  $L(y, \beta_0, \beta)$ .

Dans la pratique, il est plus loisible de se servir de la Log-vraisemblance notée  $\log L(y, \beta_0, \beta)$  :

$$\log L(y, \beta_0, \beta) = \sum_{i=1}^N y_i \log [F(\beta_0 + \beta x_i)] + \sum_{i=1}^N (1 - y_i) \log [1 - F(\beta_0 + \beta x_i)]. \quad (5.9)$$

La fonction logarithme étant continue et strictement croissante, la Log-vraisemblance se maximise avec les valeurs  $\widehat{\beta}_0$  et  $\widehat{\beta}$ , en même temps que la vraisemblance.

Pour avoir le maximum  $(\widehat{\beta}_0, \widehat{\beta})$ , il ne reste plus qu'à annuler le gradient de la fonction  $\log L(y, \beta_0, \beta)$ . Mais du point de vue pratique, à cause de la macroforme de  $\log L(y, \beta_0, \beta)$  et de la grandeur de la taille  $N$ , on utilise des méthodes numériques d'optimisation telles que la méthode de descente de gradient, pour obtenir les valeurs estimées de  $\beta_0$  et de  $\beta$ .

## 5.3 Evaluation et validation du modèle

La pertinence statistique d'un modèle se fonde en général sur la significativité individuelle des coefficients, sur la significativité globale du modèle au moyen des tests de vraisemblance (test de Wald) d'une part, et sur la puissance de discriminance et l'analyse des résidus d'autre part.

### 5.3.1 Test des paramètres, significativité des coefficients $\beta_0$ et $\beta$

#### a. Le test de Wald

Soit  $\widehat{\beta}_j$ , la valeur estimée du coefficient  $\beta_j$  ; soit  $\text{var}(\widehat{\beta}_j)$  la variance de  $\beta_j$ , estimée par  $\widehat{\sigma}_j^2$ , pour  $j$  allant de 1 à  $k$ .

Pour ce test, l'hypothèse nulle est  $\beta_j = 0$ . Sous cette hypothèse,  $\widehat{\beta}_j$  suit asymptotiquement une loi normale  $\mathcal{N}(0, \widehat{\sigma}_j^2)$ , pour  $N$  assez grand. Ainsi, le rapport  $\frac{\widehat{\beta}_j}{\widehat{\sigma}_{\beta_j}}$  obéit à peu près à une distribution normale centrée et réduite.

La statistique de Wald se définit par :

$$Wald(\beta_j) = \frac{\widehat{\beta}_j^2}{var(\widehat{\beta}_j)} = \left( \frac{\widehat{\beta}_j}{\widehat{\sigma}_{\beta_j}} \right)^2 = z. \quad (5.10)$$

La statistique  $z$  suit un Khi-deux à un degré de liberté  $\chi_1^2$ . La significativité de  $\widehat{\beta}_j$  notée  $sign(\widehat{\beta}_j) = P(\chi_1^2 > z)$  (il s'agit de la  $p$ -value) sera réalisée si elle est inférieure au seuil du test.

Attention : Ce test est très pratique pour porter un jugement sur l'importance statistique d'une variable dichotomique ou continue. On doit cependant s'en méfier lorsqu'il s'agira d'une variable polytomique, qui requiert l'introduction de plusieurs variables indicatrices (et donc plusieurs autres coefficients à tester) dans le modèle. Dans ce second cas, on conseille un test du Khi-deux.

En effet, pour le paramètre  $\beta = (\beta_1, \dots, \beta_k)$ , on teste l'hypothèse :

$H_0 : \beta_1 = \dots = \beta_k = 0$ , contre l'hypothèse

$H_1 : \exists M \subseteq \{1, \dots, k\} / \forall i \in M, \beta_j \neq 0$  et  $\forall j \in \{1, \dots, k\} \setminus M, \beta_j = 0$ ,

avec pour statistique de test la quantité

$$Wald(\beta) = \widehat{\beta} \left( \widehat{V}(\widehat{\beta}) \right)^{-1} \widehat{\beta}^T,$$

où  $\widehat{V}(\widehat{\beta})$  est la matrice de variance-covariance estimée des  $\widehat{\beta}_j$ . Sous l'hypothèse  $H_0$ ,  $Wald(\beta)$  converge en loi vers un  $\chi_k^2$ .

Une autre manière de faire est de procéder par le test du rapport de vraisemblance.

## b. Le test du rapport de vraisemblance

Le principe est le suivant : on calcule la -2Log-vraisemblance du modèle à  $k + 1$  paramètres et celle du modèle à  $k$  paramètres (en excluant la variable associée au coefficient à tester), puis on teste la différence

$$\Delta_{Khi-deux} = -2 [\log L(y, \beta_0, \beta_1, \dots, \beta_j, \dots, \beta_k) - \log L(y, \beta_0, \beta_1, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_k)], \quad (5.11)$$

en admettant que  $\Delta_{Khi-deux}$  suit asymptotiquement un Khi-deux à un degré de liberté.

### 5.3.2 Test de significativité globale du modèle

L'indépendance entre les variables explicatives doit toujours être vérifiée dans toute régression. Il existe des tests d'indépendance bien appropriés à l'instar du test de KHI-deux, mais dans cer-

tains logiciels comme STATA, c'est après la régression qu'un message d'erreur est généré en cas de dépendance.

### a. Le pseudo- $\varphi^2$ de Mac Fadden

Il est défini par :  $\text{pseudo-}\varphi^2 = 1 - \frac{L}{L_0}$  ; où  $L$  est la Log-vraisemblance du modèle avec les régresseurs (modèle complet) et  $L_0$  la Log-vraisemblance du modèle sans régresseurs (modèle nul). Ce nombre mesure le gain d'information apporté par les variables exogènes par rapport à la seule connaissance des probabilités des événements. Un pseudo- $\varphi^2$  assez proche de 0 signifie que le comportement des régresseurs ne change pratiquement rien sur la vraisemblance du modèle. Pour ainsi dire, un modèle est d'autant meilleur que son pseudo- $\varphi^2$  est grand.

### b. Les mesures de vraisemblance du modèle (goodness of fit)

Il s'agit de la fonction de vraisemblance finale (la 2log-vraisemblance) et le Khi-deux du modèle.

**La Log-vraisemblance :** Un modèle est d'autant bon que sa vraisemblance est grande. Mais en pratique, on utilise la "*moins double vraisemblance*" :  $-2LL$ . La valeur du -2 Log likelihood n'indique rien en soi, mais sa diminution indique également que le modèle est amélioré après l'introduction de la nouvelle variable.

Pour donc tester l'adéquation d'un modèle, l'hypothèse nulle est :  $H_0 : -2LL = 0$ . D'après christophe HURLIN [4], la  $-2LL$  se distribue sous  $H_0$  comme un Khi-deux, avec  $N - r$  degré de liberté ( $r$  étant le nombre de paramètres). Ainsi, si la probabilité critique se situe en-dessous du seuil de test, alors on rejette  $H_0$  qui dit que le modèle convient. Mais si la dite probabilité est supérieure au seuil, on ne rejette pas  $H_0$ , et on conclut que le modèle est acceptable.

**La Déviance et la Statistique de Pearson :** La fonction de vraisemblance résume l'information que les données apportent sur les paramètres inconnus d'un modèle. Lorsque les paramètres inconnus sont égaux à leurs estimations au sens du maximum de vraisemblance, la valeur de la vraisemblance est une très bonne mesure de la qualité d'ajustement des données par le modèle : c'est la vraisemblance maximale, que nous notons  $L(y, \beta_0, \beta)$  pour notre modèle. Cette grandeur cependant dépend de la taille de l'échantillon. C'est pourquoi elle est comparée à une grandeur calculée sur les mêmes données, mais pour un autre modèle appelé modèle saturé. (Le modèle saturé comporte autant de paramètres qu'il y a d'observations dans l'échantillon. Dans ces conditions, les valeurs de  $p_i = P(Y_i = 1/X)$  sont égales à celles de  $Y_i$ . Si on note  $L(s)$  la vraisemblance du modèle saturé

alors la déviance se définit par :

$$\begin{aligned} D &= -2 \log \frac{L(y, \beta_0, \beta)}{L(s)} \\ &= -2 \sum_{i=1}^N \left[ y_i \log \left( \frac{p_i}{y_i} \right) + (1 - y_i) \log \left( \frac{1 - p_i}{1 - y_i} \right) \right]. \end{aligned} \quad (5.12)$$

La déviance correspond en régression linéaire à la somme des carrés résiduelles, permettant au modèle de mieux s'ajuster aux données au fur et à mesure que sa valeur diminue. Cependant, la formule ci-dessus présente des réserves ; son calcul n'utilise que les  $p_i$ . C'est sans doute la raison qui a poussé Pearson à proposer la statistique suivante :

$$Z = \sum_{i=1}^N \frac{(y_i - p_i)^2}{p_i (1 - p_i)}. \quad (5.13)$$

$Z$  suit asymptotiquement une loi du Khi-deux  $\chi_{N-k-1}^2$  et tient compte de  $p_i$  et de  $y_i$  simultanément. Cette statistique également évolue inversement avec la qualité de l'ajustement, mais son inconvénient est de dépendre fortement de la taille de l'échantillon.

#### Le test de HOSMER et LEMESHOW

C'est un test d'adéquation qui s'articule autour de cinq points. Après avoir

(1) estimé le modèle, il faut

(2) calculer et ordonner les  $\hat{p}_i$  ( $p_i$  estimé par le modèle) dans l'ordre croissant, ensuite

(3) les partager en  $G$  groupes (quintile, deciles, percentiles)

(5) Soit  $n_g$  le nombre d'observations dans le groupe  $g$ ;  $g = 1, 2, \dots, G$ . Soient  $\bar{y}_g = \sum_{i=1}^G \frac{y_i}{n_g}$  et

$\bar{\pi}_g = \sum_{i=1}^G \frac{\hat{p}_i}{n_g}$ . La statistique de HOSMER et LEMESHOW s'écrit :

$$HL = \sum_{g=1}^G \frac{(n_g \bar{y}_g - n_g \bar{\pi}_g)^2}{n_g \bar{\pi}_g (1 - \bar{\pi}_g)} \quad (5.14)$$

Cette statistique est approximativement distribuée selon un Khi-deux à  $G - 2$  degrés de liberté.

Pour ce test, on regarde la  $p$ -value :  $p\text{-value} = P(HL > \chi^2)$ , où  $\chi^2 = \sum_{g=1}^G \frac{(O-E)^2}{E}$ .

Si la  $p\text{-value} > \alpha = 0.05$ , alors le modèle est dit ajustable aux données ; sinon l'ajustement sera dit douteux. Notons que cette statistique dépend du nombre de groupes choisis. C'est pourquoi Hosmer et Lemeshow la recommandent non pas comme un test à proprement parler, mais comme un indicateur d'adéquation, même si la précision augmente avec le nombre de groupes. Ainsi, pour dire qu'un modèle est adéquat, il faudrait aussi associer à  $HL$  des critères d'information tels que le *AKAIKE* (AIC) le *critère Bayésien* BIC et la *courbe Roc*.

**Le reclassement :** Si l'on construit un modèle de prédiction, alors il est nécessaire d'examiner son pouvoir de discernement. Les *outcomes* se définissent comme les valeurs  $\hat{p}_i$  de  $p_i$ , prédites par le modèle. Ici, on s'intéresse à évaluer la capacité du modèle à discriminer les *outcomes* positifs  $Y = 1$  des *outcomes* négatifs  $Y = 0$ , à partir d'un seuil bien pensé en relation avec le phénomène étudié. L'intérêt est tout d'abord porté sur deux notions : la sensibilité et à la spécificité.

La sensibilité se définit comme la probabilité pour le modèle de classer un individu en *succès* (auquel cas le test est dit positif) lorsqu'il est réellement en *succès* :

$$\text{sensibilité} = P(\text{Test} + / Y = 1)$$

La spécificité quant à elle est la probabilité pour le modèle de classer un individu en *échec* (auquel cas le test est dit négatif) lorsqu'il est réellement en *échec* :

$$\text{spécificité} = P(\text{Test} - / Y = 0)$$

Ensuite, on évalue le taux de bon classement. L'appréciation de l'erreur de prédiction est d'une importance capitale, et dépend fortement de la nature du phénomène<sup>1</sup>.

### Receiver Operating Curve (Courbe Roc)

Les seules sensibilité et spécificité du modèle ne suffisent pas, l'aire sous la courbe roc *Airoc* permet aussi d'apprécier la puissance de discrimination. La courbe Roc est la courbe des caractéristiques d'efficacité. Elle sert à l'étude des variations de la spécificité et de la sensibilité d'un test, et ce suivant des valeurs données d'un seuil de discrimination. La courbe Roc se construit en répartissant la quantité  $1 - \text{spécificité}$  (taux de faux mauvais sous-employés dans la sous-population des bons sous-employés) en abscisse, et *sensibilité* en ordonnées. Le seuil du test est l'abscisse du point où la spécificité et la sensibilité coïncident. Le pouvoir discriminant du modèle se déduit à partir de l'aire sous la courbe Roc et suivant les références ci dessous :

- si  $Airoc = 0.5$ , alors il n'y a pas de discrimination ;
- si  $0.5 \leq Airoc < 0.7$ , la discrimination est insuffisante ;
- si  $0.7 \leq Airoc < 0.8$ , la discrimination est acceptable ;
- si  $0.8 \leq Airoc < 0.9$ , la discrimination est excellente ;
- si  $Airoc \geq 0.9$ , la discrimination est exceptionnelle.

Les critères d'information **AIC** et **BIC** définis par :

$$AIC = \frac{-2\text{Log}L(y, \hat{\beta}_0, \hat{\beta}) + 2(k+1)}{N}, \text{ et } BIC = \frac{-2\text{Log}L(y, \hat{\beta}_0, \hat{\beta}) + (k+1)\text{Log}(N)}{N}, \quad (5.15)$$

<sup>1</sup>Une erreur de 20% n'est pas perçue de la même manière en assurance, qu'en santé

interviennent aussi dans l'appréciation des modèles, et sont d'ailleurs préférables à la Déviance, en ce sens que le calcul du  $AIC$  et du  $BIC$  tient compte du nombre de paramètres du modèle. Toutefois, leur importance se perçoit lors de la comparaison de plusieurs modèles.

### 5.3.3 L'analyse des résidus

Lorsqu'on parle de **résidus de Pearson**, il s'agit de la quantité :

$$r_i = \frac{y_i - p_i}{\sqrt{p_i(1 - p_i)}}.$$

Son interprétation est un peu délicate en ce sens qu'elle dépend de l'amplitude  $y_i - p_i$  et de la variance  $p_i(1 - p_i)$ . Pour y remédier, on se sert des **résidus standardisés de Pearson** définis par :

$$r_i^{std} = \frac{r_i}{\sqrt{1 - h_{ii}}}, \forall i = 1, \dots, k \quad (5.16)$$

avec  $h_{ii} = p_i(1 - p_i) X_i^T \text{Var}(\beta) X_i$ ,  $i^{\text{eme}}$  élément de la diagonale principale de la matrice

$$H = V^T X^T (XVX^T)^{-1} VX;$$

où  $V$  est la matrice de variance-covariance du modèle tel que

$$\text{Var}(Y_i) = p_i(1 - p_i) \quad \text{et} \quad \text{Cov}(Y_i, Y_j) = -p_i p_j, \quad \text{pour } i \neq j.$$

L'analyse des résidus consiste à déterminer les *outliers* (points ou observations mal modélisés) et les *influential points* (observations affectant significativement les observations). Pour une régression presque parfaite, 95% des résidus standardisés doivent être coincés dans l'intervalle  $[-2; 2]$ . C'est pourquoi en pratique, on tente dans la mesure du possible, d'exclure les individus dont la valeur absolue des résidus standardisés est supérieure à 2.

## 5.4 Comparaison de deux ou plusieurs modèles

### 5.4.1 Le Test du rapport des vraisemblances

Ce test permet uniquement de comparer les modèles imbriqués, c'est-à-dire emboîtés les uns dans les autres, à travers les variables ; car un des modèles peut être obtenu en contraignant certains paramètres de l'autre. Dans le cas de deux modèles imbriqués, le premier modèle ( $m_1$ ) possédant  $k$  paramètres libres et le second ( $m_2$ )  $k + m$  paramètres, le rapport de la vraisemblance s'écrit :

$$\begin{aligned} RV &= -2 \left[ \text{Log} \left( \frac{L(m_1)}{L(m_2)} \right) \right] \\ &= -2 [\text{Log} L(m_1) - \text{Log} L(m_2)]. \end{aligned} \quad (5.17)$$

La loi du Khi-deux permet une fois de plus de dire que  $RV$  suit asymptotiquement un khi-deux  $\chi^2(m)$ .

Concernant les modèles non imbriqués, il est conseillé de garder le modèle qui vérifiera le maximum de critères de sélection et de validation. Dans la confrontation, un modèle gagne un score lorsqu'il a le plus petit chiffre de la  $-2\log$  vraisemblance, des critères d'information AIC et BIC, des degrés de liberté; et le plus grand chiffre de pseudo- $\varphi^2$ , de taux de classification, et de l'aire sous la courbe Roc. En cas de parité égale des scores, se servir de l'analyse des résidus pour choisir le meilleur modèle.

### 5.4.2 Interprétation des coefficients de la régression

Dans le modèle de régression (5.5), apparaissent deux coefficients  $\beta_0$  et  $\beta$  encore appelés respectivement constante et coefficient de régression. L'interprétation du coefficient ne se fonde pas sur lui-même, mais sur le *Odds Ratio*. C'est une notion très utilisée en épidémiologie et qui mesure le risque associé à une variable explicative du modèle. Soit deux individus  $a$  et  $b$  qui se distinguent par la  $j^{\text{ème}}$  caractéristique :

$$\begin{aligned} \text{logit}(p_a) &= \beta_0 + \beta_1 X_1 + \dots + \beta_j (X_j + \delta) + \dots + \beta_k X_k \\ \text{logit}(p_b) &= \beta_0 + \beta_1 X_1 + \dots + \beta_j X_j + \dots + \beta_k X_k. \end{aligned}$$

Alors, on a d'après la formule (5.5) :

$$\begin{aligned} \beta_j \delta &= \text{logit}(p_a) - \text{logit}(p_b) \\ &= \log \left[ \frac{p_a / (1 - p_a)}{p_b / (1 - p_b)} \right]. \end{aligned}$$

Ce qui implique que :

$$e^{\beta_j \delta} = \frac{p_a / (1 - p_a)}{p_b / (1 - p_b)}. \quad (5.18)$$

La quantité  $\frac{p_a / (1 - p_a)}{p_b / (1 - p_b)}$  désigne en probabilité le rapport de chances (ou de risques) de succès sur les chances de l'échec; autrement dit; c'est le Odds Ratio :

$$OR(p_a, p_b) = e^{\beta_j \delta}. \quad (5.19)$$

Il mesure l'effet d'une variable explicative sur la variable dépendante. En effet, si le Odds ratio entre le régresseur  $X_j$  et la variable dépendante  $Y$  est plus petit que 1 (avec 1 n'appartenant pas à son intervalle de confiance), alors la présence de la variable explicative  $X_j$  dans le modèle augmente considérablement la probabilité pour que l'événement  $Y = 1$  se réalise.

### a. Cas d'une variable explicative dichotomique

Si on code la variable  $X_j$  en "0" et "1", la quantité  $\delta$  prend la valeur 1 de sorte que

$$e^{\beta_j} = OR(p_1, p_0) = \frac{e^{\beta_0 + \beta_j}}{e^{\beta_0}} = \frac{e^{\text{logit}P(Y=1/X_j=1)}}{e^{\text{logit}P(Y=1/X_j=0)}} \quad (5.20)$$

Donc, dans un modèle logistique, l'exponentiel du coefficient de la régression s'interprète comme le Odds Ratio que lui est associé. Si  $X_j$  représente par exemple la race ("0"=Noir et "1"=Blanc), un  $OR(p_1, p_0) < 1$  entraîne qu'il y a  $e^{\beta_j}$  fois moins de risques pour un Blanc de s'exposer au phénomène par rapport à un Noir, toutes choses étant égales par ailleurs.

### b. Cas d'une variable explicative polytomique

Il suffit de transformer chaque catégorie de la variable en une nouvelle variable dichotomique. Concrètement, chaque modalité de la variable polytomique doit être prise comme une variable binaire. L'interprétation dans ce cas se fait à partir d'une d'entre elles, prise comme référence.

### c. Cas d'une variable explicative continue

$\beta_j$  correspond à la variation du logit. La formule (5.18) montre qu'à une augmentation d'une unité de  $X_j$ , correspond une multiplication de risque par  $e^{\beta_j}$ . Toutefois, on peut recoder la variable continue en classes pour en faire une du type polytomique afin de mieux enrichir son interprétation.

---

# MODÉLISATION DU SOUS-EMPLOI ET INTERPRÉTATION DES RÉSULTATS

---

## 6.1 Apurement des données et importation dans STATA

Le fichier de travail est du type SPSS. Nous avons sollicité le logiciel STATA pour mener notre analyse parce qu'il semble plus approprié pour une telle opération et surtout pour le détail résiduel dont il fait état et la facilité qu'il offre dans l'interprétation des résultats. Pour transférer les données vers STATA, on se sert de STAT-TRANSFER, un logiciel qui permet la conversion des fichiers d'un logiciel à l'autre. Nous travaillons avec un seuil de 5%.

Le fichier étant à l'état brut, il faut gérer le peu de données manquantes et recoder les variables pour mieux les utiliser en tant que de besoin. Expressément, nous travaillons avec les variables polytomique (ou dichotomique), ce qui nous laisse le choix de recoder les modalités des variables quantitatives en classes. Ainsi :

- la variable **ap10a** désignant le nombre annuel de mois de travail a été recodé en **mtrav\_an** du type catégoriel dont les modalités sont : un semestre, de 7 à 11 mois et toute l'année entière ;
- la variable **ap10b** le nombre hebdomadaire de jours de travail recodé en **jtrav\_sem** dont les niveaux sont : mi-semaine, les jours ouvrables, et jusqu'au week-end ;
- le nombre d'heures de travail par semaine **ap10c** sera représenté dans le modèle par la variable **htrav\_sem** décrite par : moins que le quota horaire (35 heures), de 35 à 48 heures, et enfin plus de 48 heures ;
- **ap8a2**, l'ancienneté dans l'entreprise en termes d'années, est codée **ancienneté** et définie à quatre niveaux : moins de 4 ans, de 5 à 10 ans, de 11 à 20 ans, et plus de 20 ans ;
- **ap3**, la catégorie socioprofessionnelle, remplacée par **catsospr** dont les catégories sont les suivantes : hauts cadres, ouvriers qualifiés, employeurs, et manœuvres ;
- **m10b** représente les chefs-lieux des divisions administratives et se code en **divadm**. Ses catégories sont : chef lieu de province, chef lieu de département, chef lieu d'arrondissement, village ;

- **ap8b**, le mode d’obtention du travail est codé **obtention** dont les modalités sont : relations personnelles, contact direct avec l’employeur, médias et structures appropriées, et initiative personnelle ;
- **ap8a3** renseigne sur l’obtention d’une promotion au sein de l’entreprise et se code **promotion** ; elle est de type dichotomique ;
- **Ap8a42** renseigne sur le **recyclage** du travailleur, recyclage financé par l’entreprise ;
- le secteur institutionnel **siind** a été recodé en **sectinst** comportant le public, le privé formel, l’informel non agricole et l’informel agricole ;
- le statut matrimonial **m6** sera représenté par **statutmat** dont les modalités sont : célibataire, monogame, polygame+union libre, veuf+divorcé ;
- le revenu mensuel d’activité principale et secondaire **revact** est transformée en **revact2** ayant deux modalités : moins que le SMIG et supérieure ou égal à 23.500 ;
- **clage2** désigne l’âge de l’enquêté codé en classes : [0 , 20[ , [20 , 30[ , [30 , 45[ et plus de 45 *ans.* ;
- **ap5** mesure la taille de l’entreprise, nous l’avons codé en **tail\_entrp** une personne, deux personnes de 3 à 5 personnes, plus de 5 personnes ;
- **r1** permet de savoir si un actif occupé souhaite changer de métier ou pas. Il sera représenté dans le modèle par la variable **rech\_empl**, du type dichotomique.
- **r2** recense les raisons pour lesquelles un actifs occupés peut vouloir changer d’emploi. Après apurement, nous ’avons recodé en **motif\_rech**, dont les modalités sont : accroître le revenu, améliorer les conditions de travail, trouver un travail plus intéressant, et autres.
- **r3a** permet de savoir si un travailleur est près à faire des heures supplémentaires dans son emploi principal. Cette variable est remplacée par **hsupp**, du type dichotomique.
- **r3b** mesure la durée supplémentaire souhaitée (en heures) et a été recodée en **durésuppl**. Ses catégories sont : [0; 5[; [5; 10[; [10; 15[; [15; 20[; [20; 60[.
- **religion** renseigne sur les convictions religieuses du travailleur et possède quatre modalités : catholiques, protestants, musulmans, et autres chrétiens.

Il est à noter que ces codifications prennent en compte les significations des codes et veillent à ce que les modalités des nouvelles variables aient la même importance.

Les syntaxes en STATA de cette étude ont été renvoyés au paragraphe 10.2.1. Entre autres, il y a des variables à générer comme le **sexe** (**m4**) et la **région** d’enquête (**g1**). Étant donné qu’en sous-emploi, on travaille uniquement avec les actifs occupés, nous avons tenu à les filtrer avant la conversion en fichier STATA.

## 6.2 Modélisation du sous-emploi invisible

Par définition, ce sous-emploi concerne les actifs occupés dont le revenu mensuel d'activité principale (**revactp** exprimé en milliers de francs CFA) est inférieure au  $SMIG = 23\,500$ . On adopte le modèle de la forme (5.5), avec :

$$y_{inv} = \begin{cases} 1 & \text{si revactp} < 23.500 \\ 0 & \text{si revactp} \geq 23.500. \end{cases}$$

Initialement, nous procédons par des régressions univariées. Autrement dit, on effectue une régression de chacune des variables explicatives de la base de données avec la variable dépendante. Celles qui sont retenues comme pouvant faire partir du modèle du sous-emploi invisible sont :

sexe, âge, nivinst, divadm, m17, catsospr, ap8a1, ancienneté, region, ap8a5, ap8a41, recyclage, ap8a43, obtention, mtrav\_an, jtrav\_sem, htrav\_sem, hsupp, durésuppl, rech\_empl, motif\_rech, clage2, revact2, statutmat, as1c, as6, as9a, as9b, tail\_entrp, religion, sectinst, siind, sactiv ;

car les  $p$ -value associées à leurs coefficients dans les modèles univariés sont inférieures ou égales à 0,2. Les autres variables ont dans ce cas, très peu de chances de faire partir du modèle. Tout de même, elles seront réévaluées dans le modèle final, ensemble avec les variables significatives.

Il est nécessaire de calculer une matrice de corrélation afin de vérifier la multicollinéarité des variables. Ainsi, si deux variables sont fortement corrélées entre elles, elles ne peuvent toutes les deux être incluses dans le modèle, étant donné que la variation de l'une entraîne celle de l'autre. Il faudra alors procéder à une sélection selon la pertinence au regard des objectifs de l'étude.

Ayant choisi 0,75 comme seuil de corrélation, les variables corrélées sont : l'expérience professionnelle, ancienneté dans l'entreprise d'une part, le secteur d'activité, et le secteur institutionnel d'autre part. A la place de l'indépendance entre les variables, dans STATA, on s'intéresse plutôt à la *multicolinéarité* dont le résultat est généré automatiquement après une régression. De plus, il y a la possibilité de savoir si les variables introduites dans la régression sont pertinentes pour le phénomène à expliquer et si les probabilités prédites par le modèle sont assez proches de celles observées. Ceci s'obtient avec la commande *linktest* ; et s'exécute juste après la régression.

Nous avons obtenu deux modèles imbriqués, l'un de treize variables et l'autre de douze variables, et qui diffèrent d'un seul variable : **divadm**.

### 6.2.1 Etude du modèle 1

La modélisation de  $y_{inv}$  avec les variables citées ci-dessus, donne le tableau suivant :

TAB. 6.1 – Tableau de régression du modèle 1, sous-emploi invisible.

yinv	coef.	Std. Err.	z	p-value	[95% Conf. Interval]
sexe	.9209897	.0479373	19.21	$< 10^{-3}$	.8270343   1.014945
nivinst	-.4030249	.0263697	-15.28	$< 10^{-3}$	-.4547085   .3513413
catsospr	.7310701	.0404733	18.06	$< 10^{-3}$	.6517439   .8103964
ancienneté	-.0876921	.0289502	-3.03	0.002	-.1444334   -.0309509
region	.061002	.0066751	9.14	$< 10^{-3}$	.047919   .0740851
recyclage	.3218003	.0657288	4.90	$< 10^{-3}$	.1929742   .4506263
mtrav_an	-.2945223	.0290265	-10.15	$< 10^{-3}$	-.3514133   -.2376314
jtrav_sem	-.1964947	.0388228	-5.06	$< 10^{-3}$	-.2725859   -.1204034
htrav_sem	-.4403721	.0339904	-12.96	$< 10^{-3}$	-.5069921   -.373752
clage2	-.4724536	.0315713	-14.96	$< 10^{-3}$	-.5343322   -.410575
statutmat	-.0747872	.0285997	-2.61	0.009	-.1308416   -.0187327
sectinst	1.044771	.0442299	23.62	$< 10^{-3}$	.9580823   1.13146
divadm	.1121869	.0218304	5.14	$< 10^{-3}$	.0694   .1549737
const	-3.184209	.2646186	-12.03	$< 10^{-3}$	-3.702852   -2.665566

Certains indicateurs de la robustesse du modèle 1 sont rangés dans le tableau ci-dessous :

TAB. 6.2 – Indicateurs de robustesse du modèle 1.

Nombre obs	LR chi2(13)	p-value	Déviance	Pseudo $R^2$
14391	7392.82	$< 10^{-4}$	-6114.2897	0.3768

Le *linktest* et le test de multicolinéarité se sont avérés significatifs. En répartissant les 14391 individus en 30 groupes, le test de Hosmer-Lemeshow donne une  $p$ -value de 0.4981 ; preuve que le modèle 1 s'ajuste bien aux données<sup>1</sup>. Les probabilités critiques d'estimation des coefficients  $\beta_j$ ,  $j = 0, \dots, 13$ , montre bien que le test de Wald associé à chacun d'eux est significatif ; il en est de même pour la significativité globale du modèle 1 ( $p$ -value  $< 10^{-4}$ ). Pour la classification, la sensibilité et la spécificité coïncident à une césure de 0.6, pour ainsi définir un pouvoir discriminant de 79.31%. La discriminance est ensuite jugée excellente avec l'aire sous la courbe Roc estimée à 0.8794.

<sup>1</sup>voir la formulation du test de Hosmer-Lemeshow au paragraphe 5.3.2

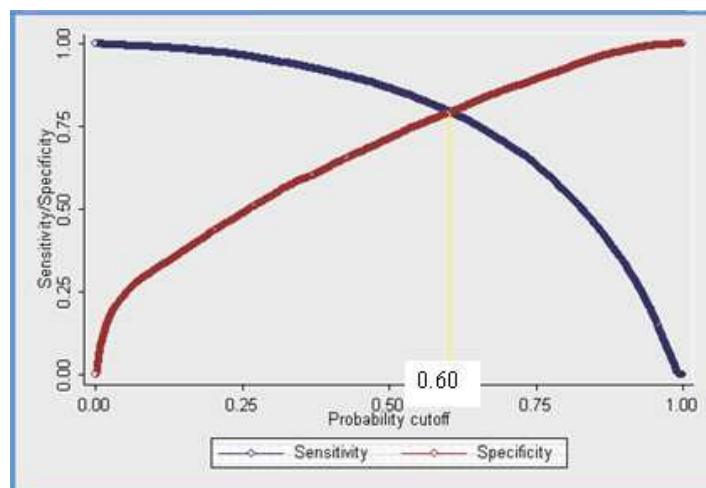


FIG. 6.1 – Césure de classement du sous-emploi invisible.

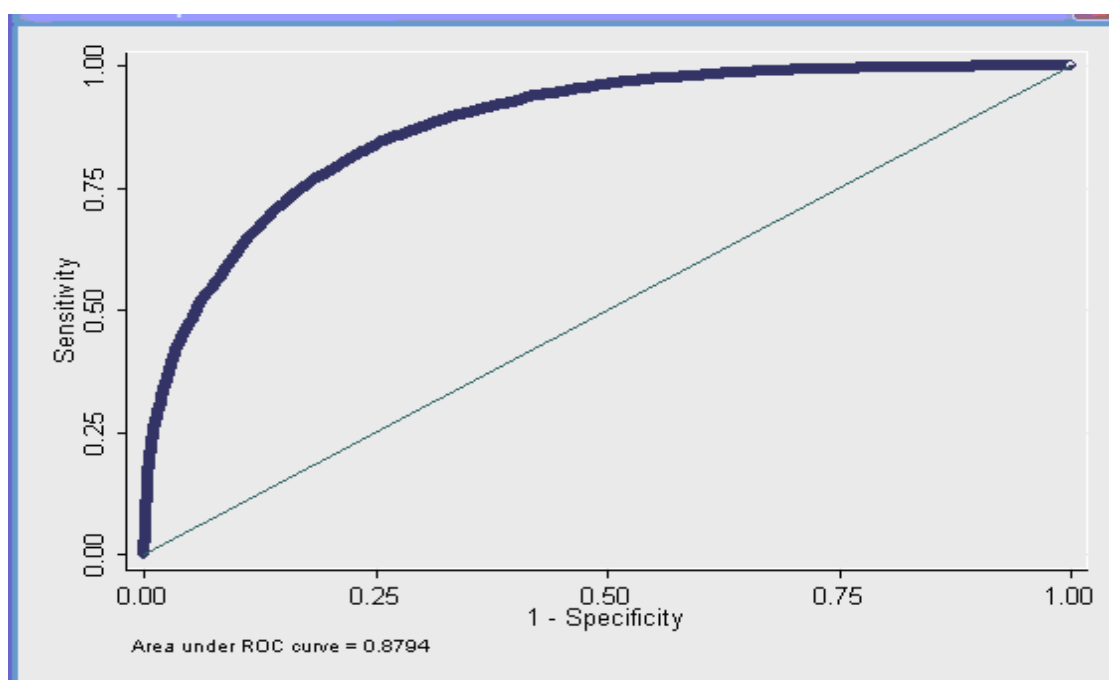


FIG. 6.2 – Courbe Roc du modèle 1, sous-emploi invisible.

### Analyse des résidus

Le nuage des résidus standardisés de Pearson en fonction des individus pour le modèle 1 révèle dans la figure (Fig 6.3) qu'une partie des résidus a une valeur absolue  $> 2$ , les individus concernés sont appelés les *outliers* ou individus mal classés par le modèle 1. Il serait intéressant de les identifier, les extraire et de réévaluer la regression.

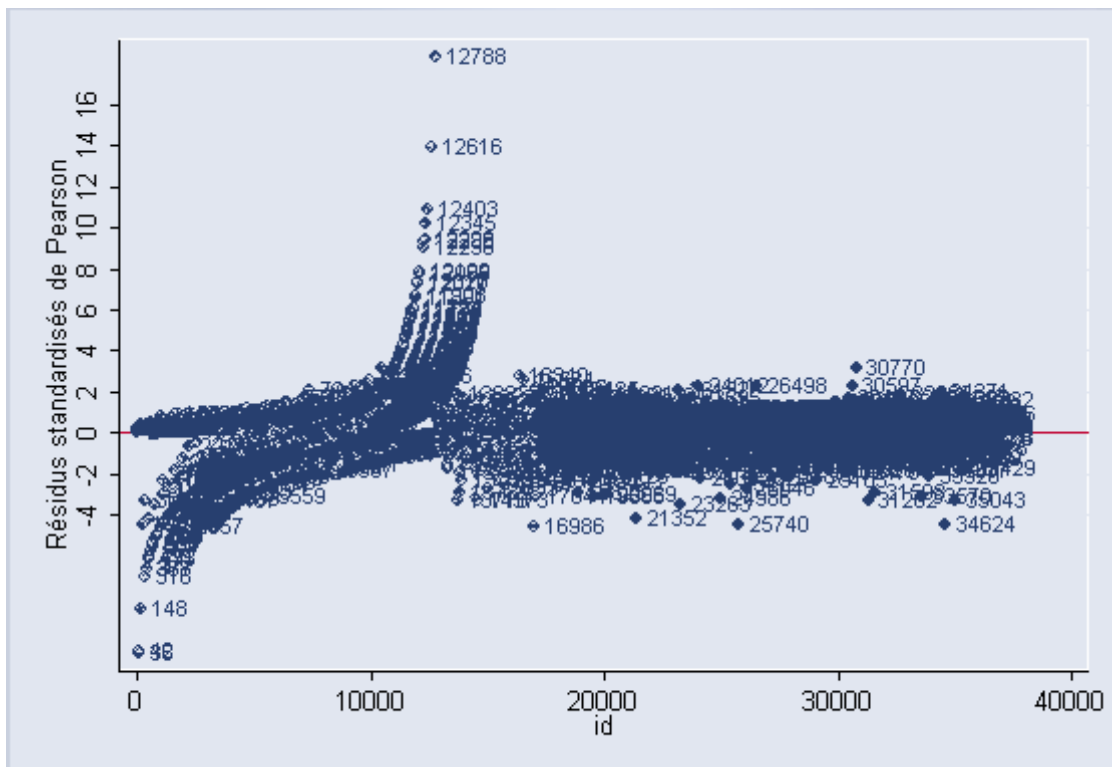


FIG. 6.3 – Résidus standardisés du modèle1.

Ces outliers sont au nombre de 664, ce qui représente à peu près 4,7% de la population étudiée .  
 En estimant à nouveau le modèle 1 après leur extraction , nous avons le résultat suivant :

TAB. 6.3 – Régression du modèle 1 modifié, sous-emploi invisible.

yinv	coef.	Std. Err.	z	p-value	[95% Conf. Interval]
sexe	1.429709	.0588339	24.30	$< 10^{-3}$	1.314397   1.545022
nivinst	-.5958891	.0317948	-18.74	$< 10^{-3}$	-.6582059   -.5335724
divadm	.1454898	.025926	5.61	$< 10^{-3}$	.0946757   .1963038
catsospr	1.024673	.0505182	20.28	$< 10^{-3}$	.9256588   1.123687
ancienneté	-.1897867	.0350636	-5.41	$< 10^{-3}$	-.2585102   -.1210633
region	.0825581	.0078935	10.46	$< 10^{-3}$	.0670872   .0980291
recyclage	.4385092	.0766422	5.72	$< 10^{-3}$	.2882932   .5887251
mtrav_an	-.4511974	.0352432	-12.80	$< 10^{-3}$	-.5202728   -.3821221
jtrav_sem	-.3559668	.0477343	-7.46	$< 10^{-3}$	-.4495244   -.2624093
htrav_sem	-.6305613	.0401152	-15.72	$< 10^{-3}$	-.7091856   -.551937
clage2	-.7094632	.0386322	-18.36	$< 10^{-3}$	-.785181   -.6337453
statutmat	-.1084755	.033731	-3.22	0.001	-.1745871   -.0423639
sectinst	1.858697	.0635636	29.24	$< 10^{-3}$	1.734115   1.98328
const	-5.179645	.3326089	-15.57	$< 10^{-3}$	-5.831547   -4.527744

TAB. 6.4 – Indicateurs de robustesse du modèle 1 modifié.

Nombre obs	LR chi2(13)	p-value	Déviante	Pseudo $R^2$
13727	9741.65	$< 10^{-4}$	-4433.2939	0.5235

L'accroissement du *pseudo  $R^2$*  indique que le modèle 1 s'est encore éloigné du modèle nul, après la perte des 664 individus. Le test de Hosmer-Lemeshow reste concluant et l'aire sous la courbe Roc augmente légèrement (0.9302), pour apprécier une classification de 83.54%.

## 6.2.2 Etude du modèle 2

Le modèle 2 a déjà été formulé plus haut et son implementation dans STATA génère des résultats aussi bien intéressants que ceux du modèle 1. Après une analyse résiduelle, on extrait 714 résidus de valeur absolue supérieure à 2. Le tableau du modèle 2 modifié est le suivant :

TAB. 6.5 – Régression du modèle 2 modifié, sous-emploi invisible.

yinv	coef.	Std. Err.	z	p-value	[95% Conf. Interval]
sexe	1.386215	.0572927	24.20	$< 10^{-3}$	1.273923 1.498506
nivinst	-.5996334	.030712	-19.52	$< 10^{-3}$	-.6598279   -.5394389
catsospr	1.048862	.049837	21.05	$< 10^{-3}$	.9511837   1.146541
ancienneté	-.1819621	.0341734	-5.32	$< 10^{-3}$	-.2489407   -.1149836
region	.0890582	.0076344	11.67	$< 10^{-3}$	.0740951   .1040213
recyclage	.4705679	.0752377	6.25	$< 10^{-3}$	.3231046   .6180311
mtrav_an	-.4394397	.0344857	-12.74	$< 10^{-3}$	-.5070304 -.371849
jtrav_sem	-.3387217	.0465943	-7.27	$< 10^{-3}$	-.4300449   -.2473985
htrav_sem	-.5955807	.0391866	-15.20	$< 10^{-3}$	-.6723851 -.5187763
clage2	-.6608847	.0372058	-17.76	$< 10^{-3}$	-.7338067 -.5879627
statutmat	-.0867713	.0330563	-2.62	0.009	-.1515605 -.021982
sectinst	1.840181	.0611896	30.07	$< 10^{-3}$	1.720251 1.96011
const	-5.166405	.3254866	-15.87	$< 10^{-3}$	-5.804347 -4.528463

TAB. 6.6 – Indicateurs de robustesse du modèle 2 modifié.

Nombre obs	LR chi2(12)	p-value	Déviante	Pseudo R2
14011	9778.37	$< 10^{-4}$	-4621.9406	0.5140

Le présent modèle 2 (modifié) s'ajuste correctement aux données, avec le test de Hosmer-Lemeshow qui fournit une  $p$ -value égale à 0.4305. Le taux de discrimination s'élève à 83.45% à partir de la même césure 0.6. L'aire sous la courbe Roc étant 0.9273 conduit à une classification exceptionnelle.

### 6.2.3 Sélection du modèle adéquat pour le sous-emploi invisible

Nous sommes en présence de deux modèles imbriqués en compétition. Le test du rapport de vraisemblances paraît approprié pour orienter le choix. Dans la syntaxe du logiciel STATA, il se réalise avec la fonction *lrtest*<sup>2</sup>. Cependant, plutôt que de se focaliser sur un seul élément de comparaison, on se sert de la commande *fitstat* pour avoir un listing de tous les points de comparaison, y compris le *lrtest*. Pour chaque modèle, elle s'exécute après la régression. Nous avons choisi d'opposer les deux modèles dans un tableau, celui qui suit :

<sup>2</sup>ou test de rapport de vraisemblances

TAB. 6.7 – Comparaison du modèle 1 modifié et du modèle 2 modifié.

	McFadden's R2	LR chi2(12)	Count $R^2$	df	Déviante	AIC	BIC
modèle 1	0.524	9741.653	0.839	13	-4433.2939	0.648	-9617.800
modèle 2	0.514	9778.369	0.838	12	-4621.9406	0.662	-9663.798

Suivant les critères de selection résumés au paragraphe 5.4.1, le modèle 1 sera retenu comme celui le plus capable d'expliquer pourquoi près de 69.3% d'actifs occupés au Cameroun gagnent mensuellement moins de 23 500 Fcfa; et dont l'équation mathématique est la suivante :

$$\text{logit}(p) = y_{inv} = -5.18 + 1.43\text{sexe} - 0.6n\text{vinst} + 1.02\text{catsospr} - 0.19\text{ancienneté} + 0.82\text{region} + 0.44\text{recyclage} - 0.45\text{mtrav\_an} - 0.35\text{jtrav\_sem} - 0.63\text{htrav\_sem} - 0.71\text{clage2} - 0.10\text{statutmat} + 1.85\text{sectinst} + 1.45\text{divadm}.$$

Cependant, l'interprétation du modèle 1 ne se fera pas directement sur le tableau *Tab 6.3*, puisque les variables ici sont pour la plupart polytomiques. On convient de transformer leurs modalités en variables dichotomiques, rendant ainsi possible l'interprétation. On obtient alors le tableau suivant, qui se construit directement avec les Odds ratio à la place des coefficients de la régression :

TAB. 6.8 – Modèle 1 "dichotomisé", sous-emploi invisible.

yinv	Odds Ratio	Std. Err.	z	p-value	[95% Conf. Interval]
sexe/Homme	ref	ref	ref		
Femme	4.848901	.3081385	24.84	$< 10^{-3}$	4.281057   5.492063
nivinst/Non scolarisé	ref	ref	ref		
Primaire	.5605134	.0509759	-6.37	$< 10^{-3}$	.4690007   .6698825
Sec 1 <sup>e</sup> cycle	.3744831	.0388714	-9.46	$< 10^{-3}$	.3055465   .4589729
Sec 2 <sup>nd</sup> cycle	.2440165	.0309679	-11.11	$< 10^{-3}$	.1902803   .312928
Supérieure	.1773055	.0397063	-7.72	$< 10^{-3}$	.1143146   .2750062
divadm/Chef-lieu prov	ref	ref	ref		
Chef-lieu dept	.9330819	.0822005	-0.79	0.432	.7851137   1.108937
Chef-lieu Arrond	1.109093	.0964746	1.19	0.234	.9352468   1.315255
Village	1.384466	.1197788	3.76	$< 10^{-3}$	1.16853   1.640307
catsospr/Haut cadre	ref	ref	ref		
Ouvrier qualifié	6.508315	3.209005	3.80	$< 10^{-3}$	2.476128   17.10661
Employeur	15.08777	7.388828	5.54	$< 10^{-3}$	5.777925   39.39835
Manœuvre	46.12041	22.71998	7.78	$< 10^{-3}$	17.5619   121.1197
ancienneté/Moins de 4 ans	ref	ref	ref		
De 4 à 10 ans	.6413796	.0470485	-6.05	$< 10^{-3}$	.5554886   .7405513
De 11 à 20 ans	.6502897	.0641789	-4.36	$< 10^{-3}$	.5359193   .7890678
Plus de 20 ans	.4700535	.0573709	-6.19	$< 10^{-3}$	.3700468   .5970874
region /Douala	ref	ref	ref		
Yaoundé	.9713309	.1020798	-0.28	0.782	.7905186   1.1935
Adamaoua	2.649934	.3611915	7.15	$< 10^{-3}$	2.028685   3.461431
Centre	1.519576	.1841269	3.45	0.001	1.198346   1.926914
Est	2.758692	.3662624	7.64	$< 10^{-3}$	2.126631   3.578611
Extrême-Nord	2.466339	.2944965	7.56	$< 10^{-3}$	1.951704   3.116676
Littoral	1.937157	.2495448	5.13	$< 10^{-3}$	1.504918   2.493542
Nord	1.955083	.259465	5.05	$< 10^{-3}$	1.5073   2.535892
Nord-Ouest	7.007503	1.014653	13.45	0.000	5.276103   9.307076
Ouest	3.415317	.4224763	9.93	$< 10^{-3}$	2.680013   4.352362
Sud	1.442614	.1886944	2.80	0.005	1.116382   1.864179
Sud-Ouest	2.392334	.3224596	6.47	$< 10^{-3}$	1.836917   3.11569

yinv	Odds Ratio	Std. Err.	z	p-value	[95% Conf. Interval]
recyclage/Oui	ref	ref	ref		
Non	1.508774	.1222163	5.08	$< 10^{-3}$	1.287282   1.768376
mtrav_an/6 mois	ref	ref	ref		
De 7 à 11 mois	.7197528	.0689293	-3.43	0.001	.5965756   .8683627
12 mois	.4675392	.0361327	-9.84	$< 10^{-3}$	.4018231   .5440029
jtrav_sem/3 jours	ref	ref	ref		
Jours ouvrables	.5553577	.0634425	-5.15	$< 10^{-3}$	.4439499   .694723
Jusqu'au week-end	.5205166	.0587937	-5.78	$< 10^{-3}$	.4171471   .6495013
htrav_sem/Moins de 25 hrs	ref	ref	ref		
De 35 à 48 hrs	.4451226	.0358765	-10.04	$< 10^{-3}$	.3800787   .5212976
Plus de 48 hrs	.2794092	.0238345	-14.95	$< 10^{-3}$	.2363908   .3302561
clage2/Moins de 20 ans	ref	ref	ref		
De 20 à 30 ans	.2777729	.0294803	-12.07	$< 10^{-3}$	.2256062   .3420021
De 31 à 45 ans	.1261131	.0151667	-17.22	$< 10^{-3}$	.0996305   .1596349
Plus de 45 ans	.1260418	.0173148	-15.08	$< 10^{-3}$	.0962902   .164986
statutmat/Célibataire	ref	ref	ref		
Monogame	.7656548	.0602787	-3.39	0.001	.6561745   .8934015
Polygame et union libre	.7471221	.0687714	-3.17	0.002	.623792   .8948358
Veuf et divorcé	.7330624	.0862377	-2.64	0.008	.5821103   .9231592
sectinst/Public	ref	ref	ref		
Privé formel	2.237353	.791922	2.28	0.023	1.118015   4.477354
Informel non agricole	20.96315	6.635064	9.61	$< 10^{-3}$	11.27306   38.98262
Informel agricole	133.0725	43.00443	15.13	$< 10^{-3}$	70.63333   250.7075

### 6.3 Modélisation du sous-emploi visible

Être en état de sous-emploi visible, c'est faire involontairement moins de 35 heures de travail par semaine. La mise en œuvre de la modélisation de ce phénomène dépend aussi bien du motif de cet absence au travail que du nombre hebdomadaire d'heures de travail. Ainsi, la variable dépendante  $y_{viz}$  se construit avec deux variables : **ap10c** pour le nombre d'heures et **ap11a1** pour le motif. **ap11a1** permet de juger si l'acte est involontaire ou pas. C'est une variable catégorielle à cinq catégories : 1) ne veut pas travailler, 2) horaires fixée par l'employeur, 3) moins de travail dû à la mauvaise conjoncture, 4) problèmes personnels (santé, travaux domestiques, ...) et 5) autre. On a alors :

$$y_{viz} = \begin{cases} 1 & \text{si } ((ap10c < 35) \text{ et } (ap11a1 = 2, 3 \text{ ou } 5)) \\ 0 & \text{si } ((ap10c \geq 35) \text{ ou } (ap11a1 = 1 \text{ ou } 4)) \end{cases} \quad (6.1)$$

La procédure est la même que pour le sous-emploi invisible. Après les régressions univariées, le modèle économétrique pour le sous-emploi visible a retenu dix variables.

Nous avons l'assurance que ces dix variables ne sont pas corrélées entre elles, qu'elles sont pertinentes pour le sous-emploi visible avec des probabilités théoriques assez proches des probabilités observées, et qu'il n'y a pas de *multicolinéarité* dans le modèle. Le tableau suivant permet d'en dire plus, notamment sur la significativité des coefficients.

TAB. 6.9 – Régression logistique du sous-emploi visible.

yviz	coef.	Std. Err.	z	p-value	[95% Conf. Interval]
jtrav_sem	-1.20462	.0358806	-33.57	< 10 <sup>-3</sup>	-1.274945   -1.134295
ancienneté	-.1016165	.0399132	-2.55	0.011	-.1798449   -.0233882
revact2	-.387998	.0686328	-5.65	< 10 <sup>-3</sup>	-.5225157   -.2534802
religion	.0848799	.0278381	3.05	0.002	.0303182   .1394417
sectinst	-.5457549	.0520409	-10.49	< 10 <sup>-3</sup>	-.6477532   -.4437567
mtrav_an	-.0748787	.0344062	-2.18	0.030	-.1423136   -.0074438
tail_entrp	-.1622585	.0267193	-6.07	< 10 <sup>-3</sup>	-.2146273   -.1098896
clage2	-.1160909	.035129	-3.30	0.001	-.1849426   -.0472393
hsupp	-1.339282	.0578569	-23.15	< 10 <sup>-3</sup>	-1.452679   -1.225884
promotion	.9851913	.2329973	4.23	< 10 <sup>-3</sup>	.5285249   1.441858
const	4.146734	.5267007	7.87	< 10 <sup>-3</sup>	3.114419   5.179048

TAB. 6.10 – Indicateurs de robustesse du modèle du sous-emploi visible.

Nombre obs	LR chi2(10)	$p$ -value	Déviante	Pseudo $R^2$
13376	2374.77	$< 10^{-4}$	-4175.3258	0.2214

Le test de HosmerLemeshow est positif et avec la valeur 0.13 comme césure, on obtient un taux de classement de 74.77%. On peut croire que cette classification est excellente dès lors que l'aire sous la courbe roc est égale à 0.8218.

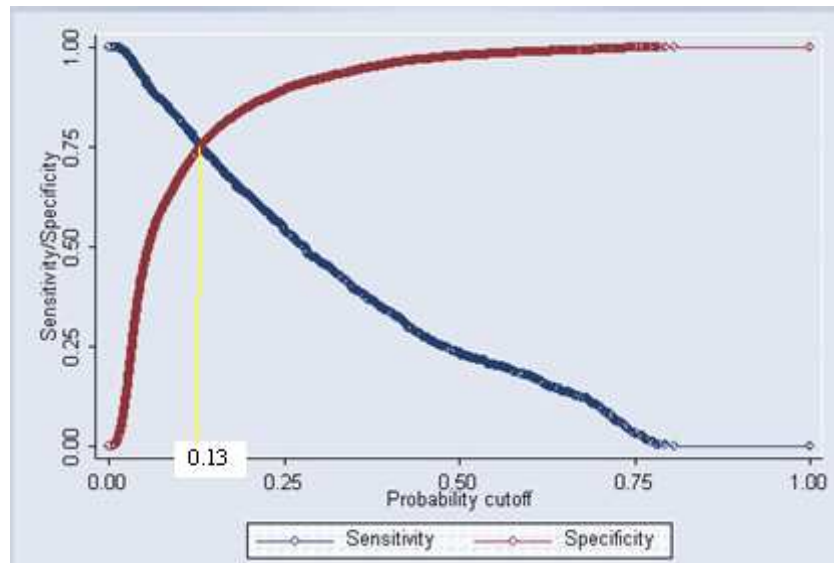


FIG. 6.4 – Césure de classement du sous-emploi visible

Que dire des résidus pour un tel modèle ?



TAB. 6.11 – Modèle modifié du sous-emploi visible.

yviz	coef.	Std. Err.	z	p-value	[95% Conf. Interval]
jtrav_sem	-1.556708	.0444365	-35.03	< 10 <sup>-3</sup>	-1.643801   -1.469614
ancienneté	-.1880401	.049312	-3.81	< 10 <sup>-3</sup>	-.2846898   -.0913905
revact2	-.5054769	.0818727	-6.17	< 10 <sup>-3</sup>	-.6659445   -.3450094
religion	.063514	.0334924	1.90	0.048	-.0021298   .1291578
sectinst	-.8431434	.0627634	-13.43	< 10 <sup>-3</sup>	-.9661575   -.7201294
mtrav_an	-.0165511	.0407729	-0.41	0.015	-.0964645   .0633623
tail_entrp	-.2934967	.0320499	-9.16	< 10 <sup>-3</sup>	-.3563134   -.23068
clage2	-.1855474	.0417535	-4.44	< 10 <sup>-3</sup>	-.2673828   -.103712
hsupp	-1.641947	.0691651	-23.74	< 10 <sup>-3</sup>	-1.777509   -1.506386
promotion	2.175276	.43595	4.99	< 10 <sup>-3</sup>	1.320829   3.029722
const	4.205561	.9188915	4.58	< 10 <sup>-3</sup>	2.404566   6.006555

TAB. 6.12 – Indicateurs de robustesse du modèle modifié du sous-emploi visible.

Nombre obs	LR chi2(10)	p-value	Déviance	Pseudo R <sup>2</sup>
12588	2851.80	< 10 <sup>-4</sup>	-3007.3158	0.3216

Le nouveau modèle est plus robuste que le premier car le pseudo  $R^2$  s'accroît, le taux de classement passe à 82.29% avec la même césure 0.13. Une fois de plus, la courbe Roc confirme que le classement est excellent. Nous allons garder ce dernier modèle comme modèle final du sous-emploi visible dont l'équation mathématique est :

$$\text{logit}(p) = y_{viz} = 4.20 - 1.56jtrav\_sem - 0.19ancienneté - 0.50revact2 + 0.06religion - 0.84sectinst - 0.01mtrav\_an - 0.29tail\_entrp - 0.18clage2 - 1.64hsupp + 2.17promotion.$$

Le modèle interprétable du sous-emploi visible avec les ratios de chance est formé par le tableau qui suit :

TAB. 6.13 – Modèle interprétable du sous-emploi visible.

yviz	Odds ratio	Std. Err.	z	p-value	[95% Conf. Interval]
jtrav_sem/3jours	ref	ref	ref	ref	
Jours ouvrables	.1880326	.0158656	-19.81	$< 10^{-3}$	.1593717   .2218477
Jusqu'au week-end	.0440964	.0039505	-34.84	$< 10^{-3}$	.0369953   .0525606
ancienneté/moins de 4 ans	ref	ref	ref	ref	
de 5 à 10 ans	.8652192	.076304	-1.64	0.101	.7278776   1.028475
de 11 à 20 ans	.6776087	.0955339	-2.76	0.006	.5140091   .893279
Plus de 20 ans	.6180971	.1328127	-1.89	0.039	.4808118   .793577
revact2/Moins de 23 500 Frs CFA	ref	ref	ref	ref	
23 500 Frs CFA et plus	.6394568	.0530162	-5.39	$< 10^{-3}$	.5435501   .7522857
religion/Catholiques	ref	ref	ref	ref	
Protestants	1.075136	.0906473	0.86	0.390	.9113738   1.268324
Musulmans	1.28903	.1163484	2.81	0.005	1.080024   1.538484
Autres chrétiens	.9957781	.1241111	-0.03	0.973	.7799578   1.271317
sectinst/Public	ref	ref	ref	ref	
Privé formel	1.289703	.6015143	0.55	0.585	.5170009   3.217274
Informel non agricole	1.805436	.825347	1.29	0.196	.7369857   4.422879
Informel agricole	.4991191	.230789	-1.50	0.133	.2016577   1.23536
mtrav_an/6 mois	ref	ref	ref	ref	
de 7 à 11 mois	.813135	.0874515	-1.92	0.045	.658593   1.003941
12 mois	.9554785	.0798149	-0.55	0.586	.8111789   1.125447
tail_entrp/une personne	ref	ref	ref	ref	
2 personnes	.7451563	.0706044	-3.10	0.002	.6188638   .8972216
de 3 à 5 personnes	.6557301	.0604591	-4.58	$< 10^{-3}$	.5473226   .7856098
Plus de 5 personnes	.7230202	.0898181	-2.61	0.009	.5667727   .9223421
clage2/moins de 2 ans	ref	ref	ref	ref	
de 20 à 30 ans	.7375073	.0702041	-3.20	0.001	.6119833   .8887777
de 31 à 45 ans	.7707415	.083729	-2.40	0.017	.6229297   .9536269
Plus de 45 ans	.5979657	.0826202	-3.72	$< 10^{-3}$	.456107   .7839454
hsupp/ Non	ref	ref	ref	ref	
Oui	.1951406	.0138739	-22.98	$< 10^{-3}$	.1697578   .2243187
promotion/Oui	ref	ref	ref	ref	
Non	6.34825	2.789659	4.21	$< 10^{-3}$	2.682899   15.02117

## 6.4 Interprétation des résultats

Ici, on se servira d'une notion beaucoup plus utilisée en épidémiologie : le Odds Ratio ou rapport de chances.

### 6.4.1 Cas du sous-emploi invisible

Le modèle 1 (dans le tableau *Tab 6.8*) que nous avons construit et jugé robuste pour l'analyse a retenu 13 facteurs essentiels pouvant expliquer le fait que certains travailleurs gagnent moins que le Salaire Minimum Interprofessionnel Garanti : le genre, l'âge, le statut matrimonial, le niveau d'instruction, la catégorie socioprofessionnelle, l'ancienneté dans l'entreprise ou l'expérience professionnelle, le recyclage du travailleur, financé par l'entreprise, l'appartenance régionale, la zone de résidence, le secteur institutionnel, et le temps de travail. L'analyse des résultats de l'estimation économétrique a permis de tirer les conclusions suivantes sur les déterminants du sous-emploi invisible au Cameroun :

◇ Au niveau du sexe, le modèle 1 donne une propension cinq fois plus élevée aux femmes à être sous-employées que les hommes (l'Odds ratio correspondant au sexe féminin est égale à 5 significativement à 5%). Ceci n'est guère surprenant car les résultats de l'analyse descriptive qui se dégage des tableaux *Tab 8.3*, *Tab 8.4* et *Tab 8.5* de l'annexe, confirment cet avantage qu'ont les hommes.

◇ Lors de l'étude, le facteur âge a été évalué en quatre classes. On note un de risque plus élevé de se retrouver en situation de sous-emploi invisible chez les moins de 20 ans. Non seulement les Odds ratio associés sont inférieures 1 mais ont tendance à décroître au fur et à mesure que l'âge augmente ; laissant entrevoir ainsi le fait que plus l'individu s'éloigne de ses 20 ans, moins il s'expose au phénomène.

◇ Concernant le statut matrimonial, seuls les célibataires sont plus en risque de vivre une situation de sous-emploi invisible, principalement les jeunes adultes de moins de 25 ans dont la vie active est encore instable.

◇ La présente étude fait en outre ressortir les bienfaits de l'éducation, donnant globalement plus d'avantages aux Hommes instruits dans l'acquisition d'un emploi décent et bien rémunéré. Ceci va étroitement avec le tableau *Tab 8.5* qui montre que cet avantage augmente avec le niveau d'instruction.

◇ Il va de soi que le rang occupé par un travailleur dans l'entreprise influence significativement le niveau de revenu de ce dernier. La catégorie socioprofessionnelle laisse une vulnérabilité au sous-

emploi invisible 46 fois plus élevée chez les manoeuvres et apprentis et six fois plus élevée chez les ouvriers qualifiés, par rapport aux hauts cadres. Ceci explique le recours perpétuel aux séminaires de formation et aux stages de perfectionnement régulièrement observés dans les structures d'emploi au Cameroun, donnant droit au reclassement. Plus précisément, se faire recycler revient à se créer des opportunités de promotion, et se donner plus de chances (avec un rapport de chances supérieure 1.5) pour éviter le sous emploi invisible.

◇ Le nombre de mois de travail par an est une variable déterminante du sous-emploi invisible, avec une contribution négative à son évolution. En effet, le simple fait d'atteindre un quota de 7 mois d'exercice par an apporte plus de chances d'éviter une rémunération relativement faible. Toutefois, cette prévention a tendance à se renforcer au fur et à mesure que ce quota augmente. Intuitivement, le phénomène se veut plus influençable par la distribution hebdomadaire du temps de travail, étant donné que ceux qui font plus de 35 heures de travail par semaine s'exposent moins au sous-emploi invisible, à condition de franchir le seuil de 7 mois par an.

◇ L'une des particularités de ce modèle est de pouvoir mesurer les distances entre les 12 régions du pays suivant le sous-emploi lié au revenu : en se référant à la ville de Douala, l'analyse économétrique prévoit en moyenne deux fois plus de risques aux autres régions (la ville de Yaoundé exclue) de s'exposer au sous-emploi invisible. Essentiellement, la région du Nord-Ouest présente des prédispositions environ sept fois plus remarquables pour entretenir ce type de sous-emploi. Ce constat appuie le résultat fourni par les statistiques élémentaires qui classent le Nord-Ouest comme mode de la distribution du sous emploi invisible suivant les grandes régions du Cameroun, Douala étant la ville la moins touchée (cf. tableau *Tab 8.4*). Puisque la ville de Yaoundé ne réagit pas sur la question du sous-emploi invisible par rapport à celle de Douala, on en déduit que les deux villes vivent la même réalité du phénomène.

◇ Le secteur institutionnel est lui aussi classé par notre modèle économétrique comme facteur pouvant influencer le sous-emploi invisible. Par rapport au secteur public, ce sous-emploi tend à gagner du terrain 21 fois plus vite dans l'informel non agricole et 133 fois plus vite dans l'informel agricole ; aussi on reconnaît là le secteur informel comme la véritable " tanière des pauvres " au Cameroun. A ce niveau aussi, le modèle évolue en droite ligne avec les observations empiriques enregistrées dans le tableau *Tab 8.9*.

## 6.4.2 Cas du sous-emploi visible

Le modèle (*Tab 6.13*) a permis avec un pouvoir discriminant de 82.29%, d'énumérer quelques points saillants du phénomène :

○ A l'échelle annuelle, le nombre de mois de travail a un effet sensible sur le sous-emploi visible en ce sens que l'avantage revient à ceux qui parviennent à faire sept mois, par rapport à ceux qui par an travaillent pendant moins de six mois. En ramenant l'analyse au niveau de la semaine, un employé qui travaille pendant moins de trois jours par semaine est plus en risque de sous-emploi que celui qui le fait tous les jours ouvrables. La situation est nettement plus claire pour un autre qui travaille jusqu'au week-end.

○ Nous avons initialement noté que l'ancienneté dans l'entreprise et l'expérience professionnelle sont deux variables corrélées. Vu sous cet angle, le sous-emploi visible évolue de manière décroissante en fonction du temps alloué à la profession. L'avantage est accordé à ceux qui auront une expérience professionnelle de plus de onze ans, par rapport à ceux qui y travaillent depuis moins de 5 ans.

○ Un autre aspect de ce phénomène se vérifie au niveau du revenu du travailleur. En effet, ceux qui mensuellement gagnent moins que le salaire minimum sont plus en risque de faire involontairement moins de 35 heures par semaine dans leur emploi principal. Ceci peut traduire l'abandon **volontaire** de l'emploi principal par les travailleurs pour s'acquitter des tâches secondaires afin de pouvoir arrondir les fins des mois.

○ A travers cette étude, on peut dire que le phénomène de recyclage du personnel apporte plus de possibilités aux travailleurs de s'occuper à plein temps. Les avancements et les nominations donnent aux travailleurs six fois plus de chance de s'occuper à plein temps. Cette propension plus élevée chez les travailleurs qui connaissent des promotions s'explique par le fait que la nomination augmente la précision de leur tâche dans l'entreprise et leur attribuent des responsabilités et des obligations. Pour la taille de l'entreprise, le présent modèle laisse comprendre que les Hommes qui s'auto emploient sont plus en risque de sous-emploi que ceux qui ensemble, forment des entreprises de 2, 3, 4 ou 5 personnes.

○ Le facteur âge n'est pas à négliger. Le bilan économétrique issu du dernier modèle révèle que le sous-emploi visible n'est pas seulement une affaire de jeunes et de vieux comme on l'aurait pensé, mais dans sa course il entraîne aussi les adultes d'entre 30 à 45 ans.

○ Les convictions religieuses influencent aussi le temps de travail de l'individu. Par rapport au catholiques, seuls les musulmans sont menacés par le sous-emploi visible en raison probable de leur programme de prière et de louanges à Allah, qui semble plus chargé et plus contraignant que pour les autres religions.

○ Contrairement à ce qu'on aurait pu penser, le sexe ne contribue aucunement à l'explication du sous-emploi lié au temps de travail. Pas même le côté maternel de la femme, ni ses responsabilités conjugales ne donne le poids au facteur sexe d'influencer le sous-emploi visible au Cameroun.

Que les résultats issus des modèles logistiques du sous-emploi visibles et invisible coïncident avec ceux de la statistique descriptive n'est pas un fait surprenant. Nous avons remarqué que le seuil de la césure de classification du modèle visible est de 0.13 et celui du modèle invisible est 0.6. Manifestement, ces césures se trouvent dans un voisinage assez petit des indicateurs correspondant, l'indicateur du sous'emploi visible étant de 12.2% et celui du sous-emploi étant 69.3%. Ceci est bien observable dans les figures *Fig 6.1 et Fig 6.4* .

---

## CONCLUSION ET RECOMMANDATIONS

---

A côté du chômage, la plupart des pays d'Afrique connaissent un problème aussi grave, sinon plus, un vaste sous-emploi à cause des activités peu productives et peu rémunératrices. En fin d'année 2005, le Cameroun peut se féliciter d'avoir réalisé un grand objectif, le recul du taux de chômage qui explique l'amélioration sensible du marché du travail <sup>1</sup>. Cependant, les politiques de stabilisation, de redressement économique et financier, mises en œuvre à cet effet, n'ont pas pu combattre le sous-emploi ; particulièrement celui lié au revenu, qui s'élève en 2005 à 69.3% de la population active occupée et frappe davantage les femmes que les hommes. Certes le niveau d'instruction de la femme est dans l'ensemble inférieur à celui de l'homme, mais il n'en est pas très éloigné. Le gouvernement se doit alors de mettre un accent sur la qualité de l'emploi, particulièrement pour les femmes.

En outre, la tranche d'âge la plus vulnérable au sous-emploi est celle des jeunes de moins de 20 ans. Il s'agit des non scolarisés et de ceux qui sortent précocement du système scolaire pour diverses raisons. Il serait souhaitable de faciliter l'accès à l'éducation et de fixer un seuil d'éducation, afin de s'assurer que le niveau de chaque jeune travailleur lui permette de pouvoir suivre une éventuelle formation professionnelle.

Un autre facteur essentiel du sous-emploi invisible est la formation. Il est indispensable d'améliorer la qualification des travailleurs. Un haut niveau de formation augmente leurs capacités d'adaptation et d'innovation et les rend plus entreprenants et plus performants. C'est ici le lieu de réclamer la revalorisation et la pratique effective du Salaire Minimum Inter Garanti ; ainsi peut-on faire réduire sensiblement la prépondérance des emplois précaires.

La faiblesse du taux de chômage au Cameroun implique l'abondance des emplois, mais des emplois concentrés pour la plupart dans le secteur informel où l'insuffisance du revenu maintient de nombreux travailleurs dans l'impossibilité de pourvoir à leurs besoins fondamentaux et à ceux de leur famille. Un intérêt particulier doit être porté sur la planification du secteur informel (agricole et non agricole), car il importe de créer les conditions nécessaires pour encourager les investissements

---

<sup>1</sup>Le Rapport d'ECAM2 [8] de 2001 indique un taux de chômage de 7.5% et un taux d'activité de 72.2% au Cameroun, tandis que le Rapport Principal de l'EESI [2] indique 4.46% comme taux de chômage et près de 75% comme participation au marché de travail en 2005.

productifs et accroître le revenu de l'informel. On observe que le respect du temps de travail a aussi une incidence positive, favorisant ainsi la productivité et le développement de l'économie.

Le taux de sous-emploi visible (12.2% d'actifs occupés) en 2005, montre que le marché de travail au Cameroun est devenu plus objectif ces dernières années, en fournissant aux actifs des emplois pour lesquels le respect des normes horaires de travail est assuré. Cependant, l'assiduité et la constance au travail doivent demeurer une préoccupation des pouvoirs publics qui gagneraient à réévaluer et à améliorer les conditions de travail<sup>2</sup>, dans le souci de maintenir un taux de sous-emploi visible toujours plus faible.

---

<sup>2</sup>De tous ceux qui travaillent involontairement pendant moins de 35 heures par semaine, plus de 50% le font pour des problèmes de conjoncture ; voir tableau **tab 8.10**

## ANNEXE

Ces tableaux ont tous une même source, ils proviennent de l'enquête EESI réalisée par l'INS en 2005. Ils déterminent pour la plupart les indicateurs du marché de travail en rapport avec le sous-emploi.

TAB. 8.1 – Taux de chômage BIT par sexe et par milieu de résidence.

Milieu de résidence	Masculin	Féminin	Total
Urbain	8,90	13,48	10,90
Semi-urbain	4,61	2,79	3,75
Rural	1,48	1,34	1,40
Total	4,23	4,69	4,46

TAB. 8.2 – Taux d'activité au sens du BIT par sexe et par milieu de résidence, pour les plus de 15 ans.

Milieu de résidence	Masculin	Féminin	Total
Urbain	75,38	58,41	66,95
Semi-urbain	81,15	66,37	73,43
Rural	88,77	85,40	86,97
Total	82,81	73,97	78,22

TAB. 8.3 – Taux de sous-emploi invisible par sexe et par milieu de résidence.

Milieu de résidence	Masculin	Féminin	Total
Urbain	29,98	57,69	41,68
Semi-urbain	55,45	82,04	68,09
Rural	77,03	89,19	83,39
Total	57,21	80,99	69,30

TAB. 8.4 – Taux de sous-emploi invisible par sexe et par région d'enquête.

Région d'enquête	Masculin	Féminin	Total
Douala	19,84	47,51	31,10
Yaoundé	20,51	41,54	29,02
Adamaoua	66,45	79,69	72,08
Centre	50,84	70,67	61,31
Est	72,10	89,18	80,28
Extrême-Nord	82,52	92,65	87,58
Littoral	46,75	78,77	64,25
Nord	71,75	93,97	82,35
Nord-Ouest	75,16	89,34	83,03
Ouest	69,31	87,54	79,47
Sud	49,58	70,44	59,16
Sud-Ouest	53,93	77,03	64,68
Total	57,21	80,99	69,30

TAB. 8.5 – Taux de sous-emploi invisible par milieu de résidence et par niveau d'instruction.

Niveau d'instruction	Urbain	Semi-urbain	Rural	Total
Non scolarisé	62,84	74,31	89,12	85,64
Primaire	53,61	72,87	85,45	77,37
Secondaire 1 cycle	43,35	68,33	74,33	60,29
Secondaire 2 cycle	24,19	41,93	50,21	32,76
Supérieur	9,03	17,98	25,10	12,09
Total	40,68	68,09	83,39	69,31

TAB. 8.6 – Taux de sous-emploi visible par sexe et par Secteur institutionnel.

Secteur institutionnel	Masculin	Féminin	Total
Public	8,19	8,79	8,36
Privé formel	5,54	8,91	6,18
Informel non agricole	17,26	22,46	19,85
Informel agricole	9,67	6,59	8,03
Total	11,89	12,36	12,12

TAB. 8.7 – Taux de sous-emploi visible par sexe et par milieu de résidence.

Milieu de résidence	Masculin	Féminin	Total
Urbain	12,10	18,20	14,68
Semi-urbain	15,60	17,14	16,33
Rural	11,29	9,74	10,48
Total	11,89	12,36	12,12

TAB. 8.8 – Taux de scolarisation par sexe et par milieu de résidence.

Milieu de résidence	Masculin	Féminin	Total
Urbain	85,07	83,06	84,07
Semi-urbain	84,20	74,48	79,30
Rural	80,56	66,52	73,34
Total	84,04	77,88	80,95

TAB. 8.9 – Taux de sous-emploi invisible par sexe et par secteur institutionnel.

institutionnel	Masculin	Féminin	Total
Public	8,47	7,52	8,19
Privé formel	9,08	14,02	10,01
Informel non agricole	46,17	70,93	58,49
Informel agricole	84,17	92,66	88,66
Total	57.21	80,99	69.30

TAB. 8.10 – Tableau croisé heures de travail contre motifs.

htrav_sem vs Pourquoi moins de 35h	Moins de 35 h	De 35h à 48h	Plus de 48h	total
Ne veut pas travailler	1476	3	1	1480
Horaire fixé par l'employeur	439	57	1	497
Mauvaise conjoncture	1010, soit 50.80%	3	2	1015
Problèmes personnels	1431	1	.	1432
autres	539	3	.	542
total	4895	67	4	4966

TAB. 8.11 – Résumé statistique de quelques variables.

Variabes	Observations	Mean	Std. Dev.	Min	Max
m4	35439	1.501086	.5000059	1	2
m5	35439	22.02449	17.61664	0	99
m6	35439	1.648833	1.213772	1	6
m8	35439	2.264821	1.444514	1	9
m10b	34959	2.529992	1.217052	1	9

Variables	Observations	Mean	Std. Dev.	Min	Max
m10c	34959	24.58077	17.7261	1	99
m15	35439	1.425012	1.481268	0	6
m17	25553	2.011584	1.391035	0	9
ap1	14741	4881.809	2925.707	10	9999
ap2	14740	203.4708	157.1319	11	430
ap3	14740	6.217639	1.921377	1	10
ap5	13465	2.575938	1.773338	0	9
ap8a1	14740	7.427137	9.753376	0	99
ap8a2	14740	7.771235	9.964094	0	99
ap8a3	14740	1.951289	.2480709	1	9
ap8a41	14740	1.918318	.3003553	1	9
ap8a5	14740	10.26906	11.74679	0	99
ap8b	14737	4.650539	3.379019	0	9
ap10a	14740	9.777544	3.852596	0	99
ap10b	14740	5.30156	1.588012	0	9
ap10c	14740	40.52734	21.4131	0	999
ap11a	5002	2.855058	1.497042	1	9
ap11b	4444	2.193969	1.243013	1	9
as4	3922	6.885518	1.84641	1	99
as6	3847	2.106836	1.225391	0	9
as9a	3922	7.514023	4.431511	0	99
as9b	3922	19.40184	16.46294	0	99
r1	14740	1.888195	.4897862	1	9
r2	1941	1.992272	1.673148	1	9
r3a	14739	1.750933	.5919734	1	9
r3b	4019	13.81264	15.0251	0	99
clage	35439	5.024098	3.540616	1	18
siind	35439	4.213324	1.079901	1	5
nivinst	35439	2.222805	1.069548	1	5
revactp	14740	38.78573	69.97566	0	1009.486
revacts	14740	4.629876	21.20752	0	1220
revact	14738	43.4215	76.31499	0	1820

---

# Bibliographie

---

- [1] Enquête sur l'Emploi et le Secteur Informel (EESI) : DOCUMENT DE METHODOLOGIE, INS, mars 2005 ;
- [2] Enquête sur l'Emploi et le Secteur Informel au Cameroun 2005 (Le Rapport Principal de l'Institut National de Statistique) ;
- [3] Cours de Régression Logistique Appliquée, Patrick Taffé, PhD Institut Universitaire de Médecine Sociale et Préventive (IUMSP) et Centre d'épidémiologie Clinique (CepiC) Lausanne, Août 2004 ;
- [4] Économétrie des Variables Qualitatives, Polycopié de Cours de Christophe HURLIN, Janvier 2003 ;
- [5] Revue INSEE PREMIERE N° 1046 octobre 2005 ;
- [6] Les mécomptes du chômage, MARUANI Margaret (2002), Paris, Bayard, 2002 ;
- [7] Les Statistiques sur l'Emploi et le Marché de Travail au Cameroun, INS, décembre 2004 ;
- [8] Pauvreté et Marché du Travail au Cameroun en 2001 ;
- [9] Revue SES N° 7 Tendances Économiques, Service des Etudes Statistiques de Wallonie, novembre 1999 ;
- [10] PERSPECTIVES, du nouveau pour les statistiques du travail, par Patrick Bollé en 1999 ;
- [11] La Mesure du sous-emploi, Bureau International Du Travail GENÈVE, 6-15 octobre 1998 ;
- [12] Rapport de la Seizième Conférence Internationale des Statisticiens du travail Genève, 6-15 octobre 1998, BIT 1998 ;
- [13] Pour une nouvelle mesure du sous-emploi d'Alain MASSOT, le 08 janvier 1998 par Les Presses de l'Université Laval ;
- [14] Les chiffres trompeurs du chômage, par Francisco Vergara dans le journal *Le Monde diplomatique*, page 17, janvier 1997 ;
- [15] Actes du Séminaire sur le Secteur Informel et la Politique Économique en Afrique Sub-saharienne, Bamako, du 10 au 14 mars 1997, AFRISTAT 1997 ;
- [16] Statistical Package for Social Science (SPSS), version 12.0 ;
- [17] STAT-TRANSFER-6 ;
- [18] STATA, version 8.0, avec les packages "stbplus" et "updates".

---

# Table des matières

---

Dédicaces	i
Remerciements	ii
Sigles et abréviations	1
Résumé	2
Abstract	4
Résumé Exécutif	6
<b>1 Introduction Générale</b>	<b>10</b>
1.1 Objectifs de l'analyse du sous-emploi . . . . .	11
1.1.1 Objectif principal . . . . .	11
1.1.2 Objectifs spécifiques . . . . .	11
<b>2 Présentation de la structure d'accueil</b>	<b>13</b>
2.1 Description de la structure d'accueil . . . . .	13
2.1.1 Le Service des Etudes Statistiques . . . . .	14
2.1.2 Le Service de la Normalisation Statistique . . . . .	15
2.1.3 Le Service des Publications, de la Diffusion, de la Documentation et des Archives	15
<b>3 Historique et définition des concepts du sous-emploi</b>	<b>16</b>
3.1 Historique . . . . .	16
3.2 Définitions des concepts et critères d'éligibilité . . . . .	17
3.2.1 Sous-emploi lié à la durée du travail . . . . .	17
3.2.2 Situations d'emploi inadéquat . . . . .	18
3.3 Définitions de quelques indicateurs du marché de travail utilisés . . . . .	19
3.4 Revue de la littérature . . . . .	20

<b>4</b>	<b>Techniques de recueil et présentation des données utilisées</b>	<b>22</b>
4.1	Justification de l'enquête . . . . .	22
4.2	Techniques de recueil des données . . . . .	22
4.2.1	Unité d'échantillonnage et d'observation . . . . .	23
4.2.2	Méthodologie de l'enquête . . . . .	23
4.3	Présentation des données . . . . .	24
4.3.1	Analyse descriptive . . . . .	25
4.3.2	Quelques chiffres du sous-emploi au Cameroun . . . . .	30
<b>5</b>	<b>Formulation logistique du sous-emploi</b>	<b>32</b>
5.1	Spécification du modèle logistique binaire . . . . .	32
5.1.1	La transformation <i>logit</i> . . . . .	33
5.2	Estimation des paramètres du modèle . . . . .	34
5.2.1	Méthode du Maximum de Vraisemblance . . . . .	34
5.3	Evaluation et validation du modèle . . . . .	35
5.3.1	Test des paramètres, significativité des coefficients $\beta_0$ et $\beta$ . . . . .	35
5.3.2	Test de significativité globale du modèle . . . . .	36
5.3.3	L'analyse des résidus . . . . .	40
5.4	Comparaison de deux ou plusieurs modèles . . . . .	40
5.4.1	Le Test du rapport des vraisemblances . . . . .	40
5.4.2	Interprétation des coefficients de la régression . . . . .	41
<b>6</b>	<b>Modélisation du sous-emploi et interprétation des résultats</b>	<b>43</b>
6.1	Apurement des données et importation dans STATA . . . . .	43
6.2	Modélisation du sous-emploi invisible . . . . .	45
6.2.1	Etude du modèle 1 . . . . .	45
6.2.2	Etude du modèle 2 . . . . .	49
6.2.3	Sélection du modèle adéquat pour le sous-emploi invisible . . . . .	50
6.3	Modélisation du sous-emploi visible . . . . .	54
6.4	Interprétation des résultats . . . . .	59
6.4.1	Cas du sous-emploi invisible . . . . .	59
6.4.2	Cas du sous-emploi visible . . . . .	60
<b>7</b>	<b>Conclusion et Recommandations</b>	<b>63</b>

8 Annexe

65

---

# Liste des tableaux

---

4.1	Répartition des actifs occupés par niveau d'instruction. . . . .	26
4.2	Répartition des actifs occupés suivant l'Organisation administrative. . . . .	27
4.3	Nombre de jours de travail par semaine, dans l'activité principale. . . . .	30
6.1	Tableau de régression du modèle 1, sous-emploi invisible. . . . .	46
6.2	Indicateurs de robustesse du modèle 1. . . . .	46
6.3	Régression du modèle 1 modifié, sous-emploi invisible. . . . .	49
6.4	Indicateurs de robustesse du modèle 1 modifié. . . . .	49
6.5	Régression du modèle 2 modifié, sous-emploi invisible. . . . .	50
6.6	Indicateurs de robustesse du modèle 2 modifié. . . . .	50
6.7	Comparaison du modèle 1 modifié et du modèle 2 modifié. . . . .	51
6.8	Modèle 1 " <i>dichotomisé</i> ", sous-emploi invisible. . . . .	52
6.9	Régression logistique du sous-emploi visible. . . . .	54
6.10	Indicateurs de robustesse du modèle du sous-emploi visible. . . . .	55
6.11	Modèle modifié du sous-emploi visible. . . . .	57
6.12	Indicateurs de robustesse du modèle modifié du sous-emploi visible. . . . .	57
6.13	Modèle interprétable du sous-emploi visible. . . . .	58
8.1	Taux de chômage BIT par sexe et par milieu de résidence. . . . .	65
8.2	Taux d'activité au sens du BIT par sexe et par milieu de résidence, pour les plus de 15 ans. . . . .	65
8.3	Taux de sous-emploi invisible par sexe et par milieu de résidence. . . . .	65
8.4	Taux de sous-emploi invisible par sexe et par région d'enquête. . . . .	66
8.5	Taux de sous-emploi invisible par milieu de résidence et par niveau d'instruction. . . . .	66
8.6	Taux de sous-emploi visible par sexe et par Secteur institutionnel. . . . .	67
8.7	Taux de sous-emploi visible par sexe et par milieu de résidence. . . . .	67
8.8	Taux de scolarisation par sexe et par milieu de résidence. . . . .	67
8.9	Taux de sous-emploi invisible par sexe et par secteur institutionnel. . . . .	68

---

8.10	Tableau croisé heures de travail contre motifs. . . . .	68
8.11	Résumé statistique de quelques variables. . . . .	68

---

# Table des figures

---

4.1	Distribution de l'âge . . . . .	26
4.2	Répartition des actifs occupés par région d'enquête. . . . .	27
4.3	Nombre de mois de travail par an dans l'activité principale, et secondaire. . . . .	29
4.4	Distribution des heures supplémentaires. . . . .	30
6.1	Césure de classement du sous-emploi invisible. . . . .	47
6.2	Courbe Roc du modèle 1, sous-emploi invisible. . . . .	47
6.3	Résidus standardisés du modèle1. . . . .	48
6.4	Césure de classement du sous-emploi visible . . . . .	55
6.5	Courbe Roc du modèle du sous-emploi visible . . . . .	56
6.6	Résidus standardisés du modèle, sous-emploi visible. . . . .	56