
M.E.S., Numéro 131, Vol. 3, novembre – décembre 2023

<https://www.mesrids.org>

Dépôt légal : MR 3.02103.57117

N°ISSN (en ligne) : 2790-3109

N°ISSN (impr.) : 2790-3095

Mise en ligne le 11 décembre 2023



Revue Internationale des Dynamiques Sociales
Mouvements et Enjeux Sociaux
Kinshasa, novembre - décembre 2023

AGGREGATION DE L'INDICE DE PAUVRETE EN RD CONGO :*Approche par la régression*

par

Serge BOKUMA ONSITI*Chef de Travaux à l'ISS-Kinshasa et Directeur
des Statistiques Démographiques et Sociales à l'INS***Doris LOFANDJO BONGENGE***Assistant chercheur***Dieudonné MUSONI SALABWE***Doctorant***Patrick MPIA WEETI***Assistant chercheur***Gabriel KATUNDA MFUMU MANU***Chef de travaux**(Tous) Doctorants, Faculté des Sciences Economiques et de Gestion
Université de Kinshasa***Résumé**

Les enquêtes sociodémographiques sont utilisées pour mesurer le niveau de vie de la population, mais elles sont souvent limitées dans les pays en développement en raison de contraintes budget-consommations. Les méthodes traditionnelles utilisent des variables sociodémographiques pour construire un indice composite de niveau de vie, souvent basés sur les caractéristiques et les actifs des ménages. Cet article présente une nouvelle agrégation de l'indice du niveau de vie basée sur l'utilisation des coefficients de régression logistique binaire multiple modélisant l'utilisation de la contraception par les descripteurs de logement et de son confort, basée sur les données de l'enquête MICS2010-RDC de 2010. Les résultats montrent que les catégories très pauvres et pauvres constituent 67% de la population du Congo en 2010.

Mots-clés : Niveau de vie, méthodes non monétaires, analyse factorielle, utilisation de la contraception, régression logistique binaire, scores, indice.

Abstract

Sociodemographic surveys are used to measure the population's standard of living, but they are often limited in developing countries due to budget-consumption constraints. Traditional methods use sociodemographic variables to construct a composite standard of living index, often based on household characteristics and assets. This article presents a new aggregation of the standard of living index based on the use of multiple binary logistic regression coefficients modeling the use of contraception by the descriptors of housing and its comfort, based on data from the MICS2010-DRC survey of 2010. The results show that the very poor and poor categories constitute 67% of the population of Congo in 2010.

Keywords : Standard of living, non-monetary methods, factor analysis, contraceptive use, binary logistic regression, scores, index.

INTRODUCTION

En 2000, 193 Etats membres de l'Organisation des Nations Unies (ONU) et au moins 23 Organisations internationales avaient inscrit la réduction de l'extrême pauvreté comme un des huit objectifs, qualifiés d'Objectifs du Millénaire pour le Développement (OMD), à atteindre par l'humanité à l'horizon 2015. Les résultats mitigés de ceux-ci ont conduit les Nations Unies à prolonger pour quinze ans supplémentaires ce programme au travers les Objectifs du Développement Durable (ODD) à atteindre à l'horizon 2030.

De ce fait, le bien-être des citoyens passe pour être la préoccupation primordiale de tout gouvernement. Réduire la pauvreté suppose disposer des outils nécessaires pour mieux la connaître. Plusieurs indicateurs calculés selon différents procédés permettent de mesurer le niveau de vie de la population : aux approches monétaires basées sur la consommation des ménages, s'ajoutent les approches non monétaires utilisant les données sur le bien-être de la population. Toutes ces approches aboutissent à la détermination du niveau de vie de la population et permettent d'orienter les politiques de développement dans les pays.

Cet article s'inspire de plusieurs travaux qui ont été réalisés dans ce domaine et propose une autre approche d'agrégation de l'indice de niveau de vie des individus aussi appelé indice de pauvreté par l'utilisation des coefficients de régression logistique binaire multiple pour déterminer les scores (pondérations) des variables entrant dans le calcul de cet indice. L'indice ainsi obtenu est un indice composite.

I. CONSTRUCTION DES INDICES DE PAUVRETE PAR L'APPROCHE NON MONETAIRE

L'objectif de ce travail étant de construire un indice pour mesurer la pauvreté, il est indispensable de commencer par définir ce concept. Au-delà de son caractère subjectif et multidimensionnel, la pauvreté se définit comme le manque ou l'insuffisance des ressources ne permettant pas à un individu l'accès à une alimentation correcte et suffisante, l'accès à l'éducation et aux soins de santé de base de qualité, l'accès au logement décent et hygiénique (Noumbissi et Sanderson, 1998).

L'approche classique de mesure de pauvreté est celle qui consiste à recourir aux dépenses de consommation en vue de mesurer la pauvreté ou le niveau de vie des ménages. Cette approche dite monétaire définit la pauvreté par une insuffisance de revenu (Duspays, 2008). Cependant, mesurer le niveau de vie des ménages selon cette unique approche n'est pas aisé surtout dans les pays en développement où les individus ont parfois du mal à connaître leur revenu et/ou à quantifier (convertir en valeur monétaire) leur consommation. Aussi, dans ces pays, les questions relatives au revenu ou à la consommation des ménages sont dans la plupart des cas interprétées comme une intrusion dans la vie privée des gens. Par conséquent, soit les ménages n'arrivent pas à connaître leur revenu ou le niveau de leur consommation, soit encore ils ne sont pas disposés à fournir des informations correctes quant à leur revenu ou leurs dépenses de consommation.

Compte tenu de ces difficultés, d'autres techniques permettant de construire un proxy de niveau de vie ont été développées. Ces méthodes se servent généralement des informations sur les caractéristiques de l'habitat et des biens possédés par les ménages. Les approches ainsi développées sont dites non monétaires ou indirectes et sont basées sur les conditions de vie des ménages (Dackam, 2000). Plusieurs travaux ont été réalisés ces dernières décennies sur l'approche dite non monétaire ou indirecte. Certains auteurs prennent en compte aussi bien les caractéristiques de l'habitat que les actifs de ménages (Shapiro et Tambashe, 1996 ; Hewett et Montgomery, 2001, Ambapour et Moussana, 2007, etc.), d'autres ajoutent à ces variables de logement, l'instruction de la femme, de l'enfant, l'accès aux services de santé, etc. (Noumbissi et Sanderson, 1998 et Mabika, 2010) ; et d'autres encore se limitant aux caractéristiques de l'habitat (Kobiane, 1998 et 2006).

Quelle que soit l'approche, la mesure de la pauvreté suppose généralement la connaissance de deux éléments : l'*identifiant*, c'est-à-dire la variable (ou les variables) à utiliser pour mesurer la pauvreté et la *méthode d'agrégation*, c'est-à-dire la manière dont ces variables sont synthétisées dans la construction de cet indice (Kobiane, 2003).

Méthodes d'agrégation de l'indice

Un indice (ou indicateur) composite est la combinaison linéaire des variables (somme pondérée ou somme simple) permettant de résumer un ensemble de variables mesurant un même concept en une seule variable. L'indice composite représente a priori un meilleur indicateur du concept étudié que chaque variable prise séparément (SCHOUAKER, 2009). Dans la littérature, différentes méthodes d'agrégation sont utilisées pour construire l'indice de pauvreté. Ces méthodes se basent souvent sur les analyses factorielles (Analyse en Composante Principale ou Analyse en Correspondance Multiple), parfois couplées avec les méthodes de classification (Ascendante hiérarchique, K-means, etc.).

Cet article utilise l'approche non monétaire basée sur les caractéristiques de l'habitat (logement) et la possession de certains actifs par les ménages pour appréhender le niveau de vie des ménages. Cependant, quant à l'agrégation de l'indice, à la place des méthodes factorielles habituellement utilisées, les pondérations (poids) des variables entrant dans la construction de l'indice sont déterminées à partir des coefficients issus de la modélisation d'un indicateur reflétant le niveau de vie ou le bien-être des individus.

II. CONSTRUCTION DE L'INDICE PAR LA METHODE DE REGRESSION

2.1. Choix de la réponse

Utiliser la modélisation statistique pour déterminer les scores des variables suppose le ciblage préalable de la réponse. Puisque les descripteurs permettent d'appréhender le niveau de vie d'un ménage ou des individus, la variable réponse doit tout aussi l'être, c'est-à-dire celle-ci doit traduire la même réalité que les descripteurs car elle n'est rien d'autre que la combinaison linéaire de ceux-ci. Le choix de la réponse est porté dans cet article à la variable « *Pratique contraceptive* » dans la mesure où elle est l'un des indicateurs

démographiques et de santé reproductive reflétant le niveau de vie des individus. En effet, plusieurs études ont montré que l'accès à la contraception influence la fécondité et traduit le niveau de pauvreté et des inégalités dans un pays. L'accès à la contraception, étant faible dans les pays en développement, fait que la fécondité soit élevée chez les pauvres¹ (Rapport Banque mondiale, 1985, Schoumaker et Tabutin, 1999 et Vimard, 2010). Elle est d'ailleurs considérée par les Nations Unies comme un indicateur de développement.

3.2. Estimation du modèle et pertinence des variables

La variable réponse étant binaire et les descripteurs multiples, le modèle le mieux adapté pour l'estimation des paramètres est la régression logistique binaire multiple (Tenenhaus, 2007).

3.2.1. Présentation du modèle logistique binaire multiple

Soit la réponse $Y_i = 1$ si la femme utilise une méthode de contraception et $Y_i = 0$ sinon que l'on cherche à expliquer par J descripteurs $x_{i1}, x_{i2}, \dots, x_{ij}$. Notons, $\pi(x) = \text{Prob}(Y_i=1|x_{ij})$ la probabilité que Y_i prennent la valeur 1 pour un X_{ij} fixé à x_{ij} . La variable aléatoire Y_i est une loi de Bernoulli de paramètre $\pi(x)$, pour un X_{ij} fixé à x_{ij} , de moyenne de $E(Y_i | X_{ij}) = \pi(x)$ et de variance $\text{Var}(Y_i | X_{ij}) = \pi(x)[1-\pi(x)]$.

Le modèle de régression logistique binaire multiple peut s'écrire (BRESSOUX P., 2008) :

$$\text{logit}[\pi(x)] \equiv \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \sum_{j=1}^J \beta_j x_{ij} \quad (1)$$

Ou, en termes de probabilité prédite :

$$\pi(x) = P(Y_i=1|x_{ij}) = \frac{\exp(\beta_0 + \sum_{j=1}^J \beta_j x_{ij})}{1 + \exp(\beta_0 + \sum_{j=1}^J \beta_j x_{ij})} \quad (2)$$

L'estimation du modèle (1) se fait avec la méthode de maximum de vraisemblance car la relation entre la moyenne de Y et les paramètres β n'est pas linéaire² (HAFNER, 2011).

Certains descripteurs de cette étude étant catégorielles nominales à plus de deux modalités, c'est le cas des caractéristiques du logement (*nature de mur, nature de sol, nature de toit, combustible de cuisson, etc.*), il a été jugé utile de les introduire dans le modèle en les transformant en des variables binaires. Cette transformation s'avère nécessaire dans le souci d'avoir le ratio *nombre de variables - nombre d'individus* assez équilibré. Ainsi, une variable explicative nominale à K modalités ($K > 2$) est transformée en $K-1$ variables binaires et introduite dans la régression logistique en considérant la $K^{\text{ième}}$ modalité comme référence à partir de laquelle les autres modalités s'interprètent.

III. PERTINENCE DES VARIABLES

Cette section présente le test permettant d'évaluer la significativité des coefficients et donc la pertinence des variables dans un modèle logistique. Il s'appuie sur le modèle probabiliste issu de la maximisation de vraisemblance.

L'hypothèse générale de base est :

$$H_0 : L\beta = 0 \quad \text{Où } L \text{ est la matrice à } n \text{ lignes et } J+1 \text{ colonnes et de rang } n.$$

¹ Dans le rapport de la Banque Mondiale de 1985 portant sur les changements démographiques et le développement économique, il est écrit « ce sont les pauvres, avec un faible niveau d'instruction et des services sanitaires et de planification familiale déficients qui ont beaucoup d'enfants ».

² La combinaison linéaire de $E(Y_i | X_{ij}) = \text{Prob}(Y_i=1|x_{ij})$ varie entre $-\infty$ et $+\infty$ et n'est pas une probabilité ; aussi, dans l'échantillon nous disposons de Y_i mais pas de $\text{Prob}(y_i|x_{ij})$. Dans ces cas, les hypothèses de moindre carré ordinaire, notamment celles relatives à l'homoscédasticité et la normalité des résidus posent problèmes.

- Le test individuel³ ($H_0 : \beta_j = 0$) pour les coefficients se fait généralement par la statistique de **Wald** ; cette statistique est utilisée par exemple dans le logiciel SPSS et sa valeur vaut :

$$W_j = \frac{\hat{\beta}_j^2}{\hat{\sigma}_{(\hat{\beta}_j)}^2} \square \chi_{(1)}^2 \tag{3}$$

Elle suit asymptotiquement une loi de Chi-deux à un degré de liberté. Pour un $\alpha = 0,05$, H_0 est rejetée lorsque sa valeur dépasse 3,84 (valeur critique de Chi-deux).

Comme on peut le remarquer, la statistique de Wald n'est rien d'autre que le carré du *t-student*. Certains logiciels comme R préfèrent utiliser directement le *t-student*⁴.

IV. EVALUATION DE LA QUALITÉ DE L'AJUSTEMENT

En régression logistique binaire, la qualité de l'ajustement du modèle doit être analysée avant d'envisager l'interprétation des résultats et leur utilisation à d'autres fins. Trois types de mesures sont généralement utilisés pour faire le bilan du modèle :

- d'abord la mesure de la liaison entre la réponse et les prédicteurs. Ce procédé utilise une statistique similaire au coefficient de détermination R^2 de la régression linéaire⁵,
- ensuite l'indicateur de l'adéquation entre le modèle estimé et les données à partir du test de Hosmer et Lemeshow et
- enfin la mesure du pouvoir prédictif du modèle utilisant deux indicateurs : la **proportion des prédictions correctes** (*ppm*) ou la **courbe ROC**⁶ (Kleinbaum. et KLEIN, 2010).

C'est cette dernière mesure que nous allons utiliser pour évaluer la qualité du modèle estimé. La *ppm* et la *courbe ROC* sont des mesures de la performance d'un classifieur binaire, c'est-à-dire d'un système qui a pour objectif de mesurer la capacité du modèle à prédire correctement le comportement des individus en se basant sur leurs caractéristiques. Pour cette étude, nous nous limitons à présenter la *ppm*, les lecteurs intéressés par la chose peuvent consulter le livre de base écrit par Hosmer et LEMESHOW ou lire Chitou B., 2006 ou Tenenhaus M., 2007. Pour comprendre cette notion (*ppm*), représentons sur une table de contingence les individus selon la probabilité $\hat{\pi}_i$ au seuil θ :

Tableau I – Table de classification basée des individus au seuil θ

Affectation	Valeurs prédites		Total
	$\hat{Y} = 1$	$\hat{Y} = 0$	

³ Le test individuel des coefficients peut se généraliser à un test de comparaison de J coefficients (test de significativité globale). Dans ce cas, $W_J = \hat{\beta}_J' \left[\hat{\sigma}_{(\hat{\beta}_J)}^2 \right]^{-1} \hat{\beta}_J \square \chi_{(J)}^2$; avec $\hat{\sigma}_{(\hat{\beta}_J)}^2 = [X' \hat{V} X]^{-1}$ et \hat{V} est la matrice diagonale des J coefficients à évaluer $\hat{\pi}(x)[1-\hat{\pi}(x)]$ et $\hat{\beta}_J$ le vecteur des J coefficients.

⁴ Un autre test très utilisé pour évaluer la pertinence des variables est le **Test du rapport de vraisemblance** (LR-test). Celui-ci peut jouer deux rôles : tester la significativité d'un coefficient β_j donné ($H_0 : \beta_j = 0$) et/ou déterminer la significativité globale du modèle. Dans ce dernier cas, il est en quelque sorte l'équivalent du test F en régression linéaire. $G = [-2L(\text{modèle réduit})] - [-2L(\text{modèle complet})]$

Où L(modèle réduit) est la log-vraisemblance du modèle à p variables ($p < J$) à (n-p-1) ddl et L(modèle complet) la log-vraisemblance du modèle complet à J variables à (n-J-1) ddl

⁵ Cette statistique ne s'interprète pas dans la régression logistique car ne donne pas de bons résultats.

⁶ *Receiver Operating Characteristic* fut inventée pendant la seconde guerre mondiale pour montrer la séparation entre les signaux radar et le bruit de fond.

Valeurs observées	Y = 1	$n_{11}(\theta)$	$n_{10}(\theta)$	n_1
	Y = 0	$n_{01}(\theta)$	$n_{00}(\theta)$	n_0
Total		$n_1(\theta)$	$n_0(\theta)$	n

Source : Inspiré de TENENHAUS, 2007

Dans cette table, la quantité n désigne la taille de l'échantillon ; $n_{11}(\theta)$ représente le nombre d'individus choisissant $Y = 1$ pour lequel le modèle a bien prédit $Y = 1$ et $n_{00}(\theta)$ le nombre d'individus choisissant $Y = 0$ pour lequel notre modèle a prédit $Y = 0$. Cependant, les nombres $n_{10}(\theta)$ et $n_{01}(\theta)$ mesurent les erreurs de classification. Un modèle bien ajusté aura donc des valeurs $n_{11}(\theta)$ et $n_{00}(\theta)$ maximales, c'est-à-dire le *ppm* égale à 1.

$$ppm = \frac{n_{11}(\theta) + n_{00}(\theta)}{n} \quad (5)$$

4.1. Construction de l'indice

Principe général de construction de l'indice

La construction de l'indice repose sur le principe suivant :

- d'abord sélectionner des variables mesurant un concept (caractéristiques du logement et son confort) et les transformer pour les rendre comparables (si nécessaire)
- ensuite construire un nouvel indicateur par combinaison (souvent linéaire) des variables (somme pondérée ou somme simple) en vue d'obtenir une variable d'intervalle ou "quasi-intervalle" et
- enfin éventuellement regrouper les valeurs du nouvel indicateur en catégories ordonnées (variable ordinale).

Sur la base des résultats obtenus de la modélisation, l'indice de pauvreté se construit comme suit : Soit $V_i(x)$, l'indice composite à construire, représentant ici le niveau de vie des Congolais du point de vue de leurs conditions de vie, son expression est donnée par :

$$V_i(x) = \sum \alpha_j x_{ij} \quad (6)$$

où les x_{ij} sont les variables caractéristiques du logement et son confort pour un individu i , et α_j le poids de chaque variable x_j dans l'indice obtenu des paramètres β_j de la modélisation de la pratique contraceptive.

L'indice ainsi obtenu est une variable continue d'échelle⁷.

4.2. Standardisation de l'indice

Une variable standardisée, aussi appelée variable centrée-réduite, est une variable transformée par une transformation linéaire en une nouvelle variable $V_i'(x)$ telle que cette nouvelle variable ait une moyenne nulle et soit de variance unité :

$$V_i'(x) = \frac{V_i(x) - \mu}{\sigma} \quad (7)$$

⁷ Pour les besoins de l'analyse, elle a été standardisée (centrée réduite).

Après la standardisation, la distribution de $V_i'(x)$ ne dépend plus de $\mu = \text{moyenne}(V(x))$ et de $\sigma = \text{écart-type}(V(x))$, ce qui permet de construire des tests et intervalles de confiance à partir des valeurs de ces paramètres⁸. La variable qu'on obtient peut-être discrétisée en variable catégorielle ordonnée (variable ordinale).

V. DONNEES ET VARIABLES ANALYSEES

Les données analysées dans cette étude sont tirées de la troisième Enquête par Grappes à Indicateurs Multiples (MICS) réalisée en République Démocratique du Congo en 2010.

Le choix de la méthode de régression logistique binaire multiple est dicté par la nature des variables qu'on analyse. Dans cette étude, la variable cible⁹ (autrement appelé variable d'intérêt, variable endogène, variable réponse, variable dépendante ou simplement variable à expliquer) est binaire (dichotomique ou booléenne) du fait que lors de cette enquête la femme devait dire si *oui* ou *non* elle a utilisé une méthode de contraception moderne dans les 12 mois ayant précédé l'enquête. Les autres variables dites explicatives (descripteurs, exogènes, facteurs¹⁰, prédicteurs, régresseurs, etc.) sont : (i) pour les caractéristiques du logement : *Approvisionnement en eau, Mode d'aisance, Moyen énergétique de cuisson et d'éclairage, Nature du sol, du toit et des murs* et (ii) pour les actifs essentiels du ménages : *Radio, Télévision, Téléphone fixe, Réfrigérateur, Groupe électrogène, Lit, Lampe à pétrole, Houe, Montre, Gsm, Bicyclette, Motocyclette, Charrette, Voiture ou Camion, Bateau à moteur, Baleinière ou Pirogue motorisée et Ordinateur*.

VI. RESULTATS

6.1. Détermination des scores des variables

La modélisation de la pratique contraceptive a donné les résultats suivants (Tableau II).

Tableau II - Paramètres du logement et son confort¹¹

Variable	B	Stand. Error	exp(β)
Constante	-1,965***	0,092	0,140
Matériel du sol - fini (Réf. Naturel)	0,257*	0,096	1,235
Matériel de mur - fini (Réf. Naturel)	0,473***	0,101	1,605
Télévision (Réf. Non)	0,337***	0,093	1,401
Lampe (Réf. Non)	0,176**	0,064	1,193
Houe (Réf. Non)	-0,231***	0,063	0,794
Gsm (Réf. Non)	0,331***	0,074	1,392

Signif. codes: **** 0,001 ; *** 0,01 ; ** 0,05

Source : Nos calculs, données MICS2010-RDC

Les coefficients issus de cette modélisation sont considérés comme des scores de chaque variable dans l'indice de pauvreté. Par cette méthode, tous les descripteurs n'ont pas la même importance, seules les variables statistiquement significatives sont prises en compte dans la construction de l'indice.

6.2. Validation du modèle

Le Tableau 3 suivant donne la proportion des individus dont le modèle a correctement prédit le comportement. Le nombre 7943 désigne la taille de l'échantillon ; 181 représente le nombre d'individus choisissant $Y = 1$ (ayant utilisé la contraception) pour lequel le modèle a bien classifié comme ayant effectivement utilisé la contraception ($\hat{Y} = 1$) et 6156 est le nombre d'individus choisissant $Y = 0$ (n'ayant pas utilisé la contraception) et pour lequel le modèle a prédit $\hat{Y} = 0$. Cependant, les nombres 1492 et 114 mesurent les erreurs de classification.

⁸ En général μ et σ ne sont pas connues, elles sont remplacées dans l'équation (7) par leurs estimateurs, ce qui permet de calculer $V_i'(x)$.

⁹ Nous utiliserons plus souvent le terme variable réponse ou simplement la réponse.

¹⁰ Le terme descripteurs sera plus utilisé dans cette étude.

¹¹ Seuls les paramètres significatifs sont présentés.

Tableau III - Bonne prédiction des individus au seuil de classification $\theta = 0,5$

Affectation		Valeurs prédites		Proportion de bien classés
		$\hat{Y} = 1$	$\hat{Y} = 0$	
Valeurs observées	Y = 1	181	1492	Sensibilité = 10,8%
	Y = 0	114	6156	Spécificité = 98,2%
Proportion de prédiction correcte = $(181+6156)/7943$				79.8%

Source : Nos calculs, données MICS2010-RDC

$$ppm = \frac{(181 + 6156)}{7943} = 79.8\%$$

Le modèle estimé donne près de 80% de prédiction correcte. Fort de ces résultats, le modèle ainsi prédit peut être considéré comme bon et les paramètres qui ont été estimés sont utilisables pour la construction de l'indice de pauvreté.

6.3. Indice composite de niveau de vie

L'indice est obtenu en remplaçant la valeur de chaque paramètre significatif β_j du modèle estimé (Tableau 2) dans l'équation (6) :

$$V_i(x) = \sum \alpha_j x_{ij} = 0,257x_1 + 0,473x_2 + 0,337x_3 + 0,176x_4 - 0,231x_5 + 0,331x_6$$

Sa standardisation donne :

$$V_i'(x) = \frac{V_i(x) - 0,4522}{0,8104}$$

Tableau IV - Points de séparation (césure) de l'indice en trois groupes

Statistiques indice_standardisé	
Minimum	-1,653608247423
Maximum	2,179381443299
Moyenne	0,452169744012
Ecart-type	0,810442562812
Centiles	
33,33333333	-0,56701030927835
66,66666667	0,27216494845361

Source : Nos calculs, données MICS2010-RDC

6.4. Répartition des Congolais selon leur niveau de vie

L'indice de niveau de vie obtenu peut être discrétisée en trois catégories distinctes¹² : *Individus riches*, *Individus intermédiaires* et *Individus pauvres*. L'individu est affecté dans une catégorie en fonction du niveau de son indice. Un indice faible traduit un niveau de précarité élevé et dans ce cas l'individu est considéré comme pauvre ; un indice moyen correspond à un niveau de vie intermédiaire et un indice élevé correspond à un niveau de vie élevé.

On note que, comme dit dans les paragraphes précédents, la pauvreté perçue du point de vue des conditions de vie des ménages est une pauvreté relative.

¹² Le choix de trois catégories est fonction de ce qu'on cherche, souvent il est arbitraire. On pouvait tout aussi prendre cinq classes. Néanmoins, il est généralement conseillé de former des groupes impairs pour mieux séparer les individus.

Tableau V – Répartition des congolais selon leur niveau de vie :

	Effectifs	Pourcentage
Pauvre	2796	35,2
Moyen ou intermédiaire	2515	31,7
Riche	2632	33,1
Total	7943	100

Source : Nos calculs, données MICS2010-RDC

La répartition en trois classes donne les résultats suivants : 35% des Congolais sont considérés comme pauvres, 32% comme moyens ou intermédiaires et 33% comme riches (ou non pauvres). En réalité les deux premières classes représentent les « extrêmement » pauvres et les pauvres, ce qui nous donne près de 67% des individus vivant dans la pauvreté. Bien que subjectif, la pauvreté définie ici correspond à peu près aux résultats obtenus dans d'autres études utilisant des approches différentes. En 2010, Moumami a montré, à partir des données de l'Enquête 1-2-3 de 2005, que la pauvreté en République Démocratique du Congo touchait près de 70% des individus (Moumami, 2010). Ces chiffres sont également proches de ceux du premier Document de Stratégie pour la Croissance et la Réduction de la Pauvreté (DSCR 1) du Ministère du Plan de 2006. On note cependant que les études récentes ont montré que le taux de pauvreté monétaire a baissé et se situerait autour de 56,2% (OCDD et INS, 2021).

CONCLUSION

Cet article vient de montrer qu'il est possible d'utiliser les coefficients de régression de façon générale et en particulier ceux de la modélisation logistique binaire multiple comme pondération des variables dans l'agrégation d'un indice de niveau de vie des individus. Ce procédé est rapide et facile à mettre en œuvre pourvu qu'un choix judicieux soit opéré au niveau de la variable réponse qui, comme cela a été dit, doit traduire le même concept de pauvreté ou de bien-être que ses descripteurs.

En utilisant les données de l'enquête MICS2010-RDC organisée en République Démocratique du Congo en 2010, essentiellement le module contraception où a été tiré la variable réponse et le module condition de vie des ménages, cette approche a montré que près de 67% des Congolais vivaient dans la pauvreté. Ces résultats sont très proches de ceux présentés par d'autres auteurs ayant utilisé les analyses factorielles (ACP, ACM, Classification hiérarchique, etc.) couramment admis dans ces types d'analyse. Cependant, puisqu'il s'agit d'une nouvelle (autre) approche, les lecteurs intéressés à l'utiliser sont appelés à comparer les résultats présentés dans cette approche avec ceux utilisant les approches factorielles avec les mêmes données pour mieux évaluer son efficacité. Aussi, dans la procédure de régression logistique binaire, un modèle n'est bon que s'il donne également les meilleurs résultats avec les données externes c'est-à-dire autres que celles ayant été utilisées pour sa mise en œuvre. Ainsi, les variables significatives retenues ici peuvent être utilisées avec les données récentes de l'Enquête MICS de 2018 ou d'autres enquêtes similaires pour voir si le même modèle est aussi bon. La difficulté pour appliquer cette approche réside dans le choix de la variable réponse. Il n'est pas garanti que le choix d'une variable réponse autre que la pratique contraceptive donne les mêmes résultats. De même, en choisissant la pratique contraceptive comme réponse, il faut s'assurer que les données à analyser n'ont pas été collectées dans un pays où cette pratique est obligatoire (la Chine, Inde, etc.) ou interdite, car dans le cas de la Chine par exemple, l'utilisation de la contraception ne relève pas du libre choix des individus mais plutôt de la contrainte de la loi et, de ce fait, ne reflète pas le niveau de vie des habitants.

BIBLIOGRAPHIE

- Ambapour S. et Moussana H., 2007, *Pauvreté et fécondité, Pauvreté et fécondité au Congo*, Document de travail, n°14/2007, BAMSI, Brazzaville (document internet).
- Bokuma Onsite S., 2011, *Pauvreté et indicateurs sociaux de santé maternelle en République Démocratique du Congo*, Mémoire de Master en statistique, Institut de Statistique Biostatistique et sciences actuarielles (ISBA), Université Catholique de Louvain, Louvain la neuve.
- Bressoux P., 2008, *Modélisation statistique appliquée aux sciences sociales*, De boeck, Bruxelles.
- Chitou B., 2006, *Modèles logistiques appliqués*, Ensea, Abidjan.
- Dackam Ngatchou R., 2004, *Recensement Général de la Population et de l'Habitat en Afrique (RGPH)*, UNFPA, Equipe régionale d'appui technique-1, Dakar.

- Dupays S., 2008, *Déchiffrer les statistiques économiques et sociales*, Dunod, Paris.
- Escofier B. et Pagès J., 2008, *L'analyse factorielle simple et multiple*, Dunod, Paris.
- Evina A. et Ngoy K., 1998, *Fécondité, planification familiale et crise en milieu urbain camerounais : le cas d'une ville moyenne Edéa*, in *Crises, pauvreté et changement démographiques dans les pays du sud*, Edition ESTEM, Paris (document internet).
- Hafner C., 2011, *Econométrie*, notes de cours, Université Catholique de Louvain, Louvain-la-neuve.
- Hewet P. et Montgomery M-R., 2001, *Poverty and public services in developing-country cities*, Working papers, Population Council (document internet).
- Husson F., et al., 2009, *Analyse des données avec R*, PUR, Rennes.
- Kleinbaum D. G. et Klein M., 2010, *Logistic regression: a self-learning text*, Springer, New York.
- Kobiane J-F., 1998, *Essai de construction d'un profil des ménages à Ouagadougou à partir des caractéristiques de l'habitat*, in *Crise, pauvreté et changement démographiques dans les pays du sud*, Estem, Paris (document internet).
- Kobiané J-F., 2003, *Habitat et bien d'équipement comme indicateurs de niveau de vie des ménages : diversités des méthodologies et implications en matière des politiques sociales*, UEPA, 4ème Conférence Africaine sur la Population, « Population et pauvreté en Afrique : réagir aux défis du 21ème siècle », Tunis (document internet).
- Kobiane J-F., 2006, *Ménages et scolarisation des enfants au Burkina Faso : à la recherche des déterminants de la demande scolaire*, thèse de doctorat en démographie, Académia-Bruylant, Louvain-la-neuve.
- Mabika C., 2010, *Pauvreté et accès des enfants aux services de base en R.D. Congo*, Editions Universitaires Européennes, Saarbrücken.
- Moumami A., 2010, *Analyse de la pauvreté en République Démocratique du Congo*, BAD Working papers series (document internet).
- Noumbissi A. et Sanderson J-P., 1998, *Pauvreté et comportements démographiques au Cameroun : à la recherche d'un indicateur de pauvreté*, in *Crises, pauvreté et changements démographiques dans les pays du sud*, Edition ESTEM, Paris (document internet).
- Observatoire Congolais du Développement Durable et Institut National de la Statistique, 2021, *Enquête par Grappes à Indicateurs des ODD, 2020 : Rapport final*. Kinshasa, République Démocratique du Congo.
- Schoumaker B. et Tabutin D., 1999, *Relation entre pauvreté et fécondité dans les pays du sud : connaissances, méthodologies et illustrations*, Institut de démographie-UCL, Louvain-la-neuve.
- Schoumaker B., 2009, *Analyse quantitative en sciences sociales II : construction d'échelles et indicateurs composites*, notes de cours, Institut de démographie, Université Catholique de Louvain, Louvain-la-neuve.
- Shapiro D. et Tambashe O., 1996, *Gender, poverty and school enrollment in Kinshasa-Zaire*, Working papers, Department of Economics, Univ. Park, Pennsylvanie (document internet).
- Tenenhaus M., 2007, *Statistique : méthodes pour décrire, expliquer et prévoir*, Dunod, Paris.
- Vandeschrick C. et Wautelet J-M., 2003, *De la statistique descriptive aux mesures des inégalités*, Académia Bruylant, Louvain-la-neuve.
- Vimard P., 2010, *Politique de la planification familiale, régulation de la fécondité et pratique contraceptive : une Afrique aux multiples visages*, in Fassassi R., Vignikin K. et Vimard P., *La régulation de la fécondité en Afrique : transformations et différenciations au tournant du XXIè siècle*, Académia-Bruylant, Louvain-la-neuve.