

Présentation 4



DÉMARCHE MÉTHODOLOGIQUE POUR L'ORGANISATION DE COLLECTE DES DONNÉES SUR LES REVENUS ET LES DÉPENSES DE CONSOMMATION

SORO Dognimon, dognimon@yahoo.fr

COTONOU, DU 20 AU 22 MARS 2018



Plan de présentation

- Objectifs principaux d'une méthodologie de collecte de données
- Etat des lieux sur les statistiques de revenus et des dépenses de consommation
- Principales sources des données
- Sources possibles d'erreurs dans une enquête
- Principes de base pour la conception des questions
- Echantillonnage
- Caractéristiques d'un échantillon
- Traitement des données à problèmes et imputations
- Traitement des loyers fictifs
- Traitement des dépenses des fêtes et cérémonies
- Traitement des transferts en nature et en espèces
- Traitement des biens durables
- Prise en compte de la structure et de la composition
- Prise en compte de différences régionales du niveau de vie du ménage



Objectifs principaux d'une méthodologie de collecte de données

- ▶ Collecter les informations requises en respectant les critères
 - ▶ d'exactitude
 - ▶ d'efficacité
 - ▶ de simplicité
- ▶ Une conception qui vise la qualité et la pertinence
 - ▶ en vue de réduire les erreurs de mesure
 - ▶ et d'assurer un meilleur rapport qualité/coût

Etat des lieux sur les statistiques de revenus et des dépenses de consommation

- Faible déclaration au fisc et faible disponibilité de données sur les revenus et la consommation
- données incomplètes (souvent sur l'une de la composante)
- Absence de série (la conséquence est que l'analyse sur la question est insuffisante)
- la quasi-totalité des activités économiques se situe dans l'informel
- l'exploitation des données existante est encore faible, etc.

Causes principales

Les statistiques bénéficient peu de financement

Le soutien des partenaires au développement s'oriente vers les domaines les plus prioritaires (niveau de pauvreté et condition de vie des ménages)



Principales sources des données

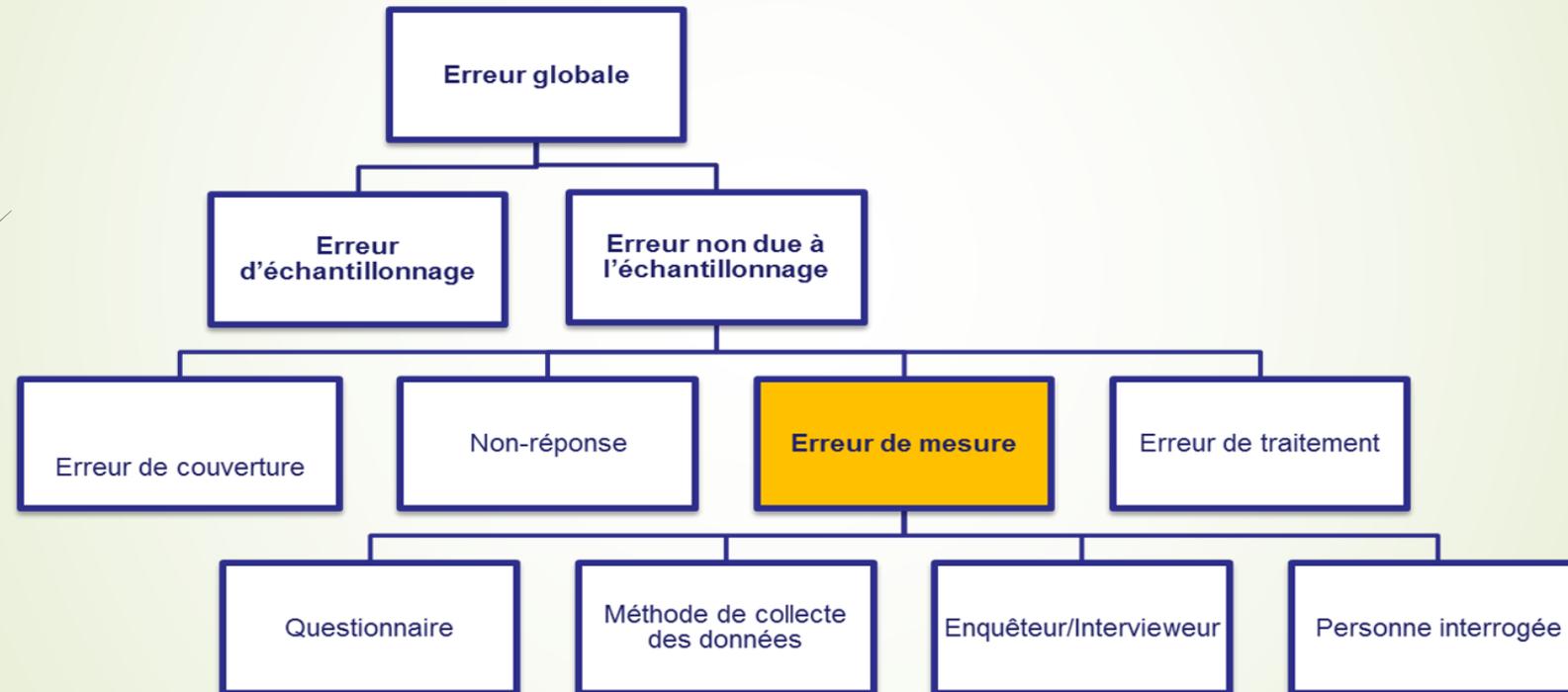
❖ *Sources primaires*

- Enquêtes main d'œuvre (EMO)
- Enquête spécifique auprès des Unités de Production Informelle (UPI)
- Enquêtes auprès des entreprises
- Enquête sur la pauvreté et le niveau de vie des ménages (ENV)
- Enquête spécifique auprès des ménages sur les revenus et les dépenses des ménages

❖ *Sources administratives*

- Fichiers du service public de la paie (direction de la solde)
- Fichier de l'agence de sécurité sociale
- Autres sources

Sources possibles d'erreurs dans une enquête



Causes d'erreurs de mesure

- **La méthode de collecte des données**
 - ❖ Entrevues directes , par téléphone, ..
- **Questionnaire**
 - ❖ Problèmes de spécification (mesure inadéquate du concept choisi)
 - ❖ Effets du plan d'échantillonnage – formulation des questions, longueur, ordre, présentation du questionnaire
- **Enquêteur/Intervieweur**
 - ❖ Variabilité dans la tenue de l'entrevue, erreurs d'interprétation/d'enregistrement, manipulation délibérée, rôles culturels et sensibilité des sujets
- **Personne interrogée**
 - ❖ Malentendu, problèmes de mémoire (omissions, télescopage), désirabilité sociale, répondant substitut

Principes de base pour la conception des questions

❑ Essentiel pour la qualité des données

➤ Concepts et définitions dans les statistiques sur la MO sont généralement complexes

- Comprennent les éléments essentiels de la définition / critère
- Langage clair et simple (couramment utilisé)
- Langage neutre
- Un concept ou une question à la fois
- Contexte clair, période de référence



Principes de base : administration des questions

- Essentielle à la qualité des données
- Les concepts et les définitions statistiques relatifs aux statistiques sur la main d'œuvre ont tendance à être complexes
 - Lire textuellement
 - Inclure les éléments essentiels de la définition/du critère
 - Employer des mots courts et simples (couramment utilisés)
 - Langage neutre
 - Un concept/une question à la fois
 - Contexte clair, période de référence

Principes de base : choix du type de questions

➤ Fermées

- ❖ Lorsque tous les choix de réponses ou la plupart sont connus
- ❖ Lorsqu'il est possible de classer les réponses dans un ensemble de choix gérables

➤ Ouvertes

- ❖ Nouveaux sujets où l'éventail de réponses n'est pas connu
- ❖ Lorsque les réponses sont très variées (liste des biens de consommation)
- ❖ Pour une enquête en profondeur

➤ Diverses

- ❖ Fermées plus «autres (préciser)»

➤ Fermées (les plus fréquentes dans les enquêtes sur la main-d'œuvre)

- ❖ Catégories mutuellement exclusives et exhaustives
- ❖ Eviter les catégories de réponses trop nombreuses (< 10-12)

Principes de base : mise à l'essai ou pilote

Mise à l'essai sur le terrain:

- Essais initiaux du questionnaire sur le terrain
- Peuvent couvrir l'ensemble du questionnaire ou de nouvelles sections/sections réexaminées
- Il s'agit d'améliorer la conception du questionnaire
- Pas nécessaire d'avoir un échantillon représentatif, mais le questionnaire doit faire l'objet d'essais à plusieurs endroits qui représentent la population visée

Utile pour

- Vérifier la pertinence globale de questions prévues pour divers groupes de population
- Évaluer la compréhension cognitive des questions/réponses
- Déterminer l'éventail des possibilités de réponse à une question
- Mettre à l'essai d'autres questions/séquences d'un questionnaire
- **Essentielle pour réduire/contrôler les erreurs de mesure**



Principes de base : formation des agents

Objectifs:

- Veiller à l'uniformité dans les procédures d'entrevue de l'enquête
- Assurer la qualité de l'information recueillie

Contenu:

- But de l'enquête et fins auxquelles les résultats serviront
- Examen des concepts et définitions utilisés dans le questionnaire
- Protocoles régissant les contacts avec les ménages sélectionnés, la planification des visites
- Critères d'admissibilité et instructions sur la manière de mener à bien l'entrevue, de remplir le questionnaire, de vérifier la cohérence des réponses et la qualité des données
- Stratégies relatives à l'aversion pour les refus et la conversion des refus

Principes de base : formation des agents

Méthodes :

- Explication des concepts et des définitions contenus dans le questionnaire
- Examen du questionnaire, point par point
- Etudes de cas, simulations d'entrevue, démonstration d'entrevue, entrevue devant la classe
- Expérience pratique relative à l'examen des procédures de collecte de données; par la suite, tenue d'une session d'évaluation

Durée:

- Prévoir au moins une semaine de formation pour un nouveau questionnaire et les enquêteurs nouvellement recrutés
- Organiser des sessions d'évaluation dans les premières semaines suivant la collecte des données
- Dispenser des formations de recyclage pour les périodes de collecte de données de plus longue durée

Echantillonnage

L'échantillonnage de l'enquête décrit le processus de sélection d'un échantillon d'éléments à partir d'une population cible

Types d'échantillonnage

► Échantillons non probabilistes

La probabilité de sélection d'unités dans l'échantillon est inconnue.

- ❖ Les méthodes ne peuvent être évaluées que de manière subjective
- ❖ Les estimations de l'enquête seront biaisées, et l'ampleur et la direction inconnues
- ❖ Peuvent être utilisées pour des échantillons très réduits lorsqu'on n'essaie pas d'assurer la précision des estimations, par exemple, les tests cognitifs du questionnaire d'enquête

► Échantillons probabilistes

Méthode de sélection selon laquelle chaque unité de la population (personne, ménage, etc.) a des chances égales et connues d'inclusion dans l'échantillon.

- ❖ Avantages : Les résultats peuvent être généralisés et les erreurs d'échantillonnage (précision des estimations d'échantillonnage) peuvent être mesurées.
- ❖ L'accent sera mis sur les échantillons aléatoires dans la présentation

Caractéristiques d'un échantillon

Le plan d'échantillonnage nécessite un équilibre entre quatre critères principaux :

1. Orientation des objectifs : La sélection des échantillons et l'estimation doivent tenir compte des objectifs de recherche alignés sur le plan d'échantillonnage et les conditions générales.

2. Mesurabilité : Les plans de d'échantillonnage qui permettent l'estimation de la variabilité d'échantillonnage, mesurée en erreurs types.

L'inférence statistique dépend des échantillons probabilistes.

3. Caractère pratique : Le modèle de sélection de l'échantillon doit être traduit en instructions simples, claires, pratiques et complètes pour le bureau et le personnel de terrain.

4. Économie : Veiller à ce que les objectifs soient atteints à un coût minimal avec un degré de succès dans la réalisation de ces objectifs



Population cible

Population destinée à être couverte par l'enquête (aussi appelée « univers de couverture »). La taille de l'échantillon dépend également de la population cible que devra couvrir l'enquête. Comme dans le cas des indicateurs, une enquête sur les ménages porte fréquemment sur plusieurs populations cibles (sous-groupes) comme : sexe, situation d'occupation, statut dans la profession, etc. Pour calculer la taille de l'échantillon, il faut par conséquent prendre en considération chacune des populations cibles.

Précision et confiance statistique

La taille de l'échantillon dépend directement du degré de précision que doivent présenter les indicateurs. Plus les estimations doivent être précises ou fiables, et plus l'échantillon devra être nombreux, et ce de plusieurs ordres de grandeur. Une grande précision a nécessairement un impact sur la taille de l'échantillon et par conséquent sur le coût de l'enquête.

Pour garantir l'efficacité du processus d'échantillonnage, un certain nombre de techniques sont utilisées notamment la stratification, l'échantillonnage en grappes et l'échantillonnage par phases.

La stratification

La stratification consiste à subdiviser la population en sous-populations (strates) distincts sur la base d'informations auxiliaires préalables avant de l'interroger. Les éléments de l'échantillon sont alors sélectionnés indépendamment, dans chaque strate, à la lumière des caractéristiques à mesurer.

Dans le cadre d'enquêtes main d'œuvre, les variables de stratification peuvent être : le district ou région, le milieu de résidence (urbain/rural), groupe ethnique, etc.

L'échantillonnage en grappe

L'échantillonnage en grappes consiste à sélectionner les échantillons de manière que tous les membres d'un groupe soient interrogés, les groupes eux-mêmes étant définis comme étant les grappes. Par exemple, **tous les habitants dans d'un ilot** seraient enquêtés lorsqu'une **ZD** est sélectionnée.

Ces dernières années, cette méthode est utilisée de manière plus large pour désigner, d'une façon plus générale, les enquêtes comportant une avant-dernière étape d'échantillonnage qui sélectionné (et définit) les grappes, comme villages, zones d'énumération du recensement. Lors de la dernière étape de l'échantillonnage, un sous-échantillon de ménages de chaque grappe sélectionnée sont interrogés, et non pas tous dans le but de réduire les coûts.

L'échantillonnage par phases

La sélection d'un échantillon en plusieurs phases a des avantages pratiques pour le processus de sélection lui-même. Elle permet d'isoler, par phases successives, les zones géographiques où sera menée l'enquête et en particulier d'établir une liste des ménages et de mener les entrevues. Avec un échantillonnage en grappes, la procédure de sélection comporte généralement au minimum deux phases, premièrement la sélection des grappes et, deuxièmement, la sélection des ménages.

Avantages

- La conception de l'échantillon est simplifiée, ce qui réduit le risque d'erreurs autres que des erreurs d'échantillonnage lors de l'application de l'échantillon;
- Réduit le coût de l'enquête;
- Lorsque la base de sondage est obsolète : une étape de sélection peut être introduite pour limiter la taille de la zone à lister

Caractéristiques d'un échantillon

Taille de l'échantillon

Selon la théorie statistique, la taille des échantillons dépend surtout des erreurs systématiques estimées et du degré de précision recherché. La taille de l'échantillon tient également compte des restrictions budgétaires.

Précision

Il importe toujours que la **précision estimée soit acceptable**, notamment pour les enquêtes sur les revenus, la consommation si l'on veut obtenir des estimations (nationales, régionales, groupes, ou une ventilation plutôt qu'une simple moyenne ou un simple total).

D'une manière générale, la formule d'estimation de la taille de l'échantillon

$$n_h, \text{ est } n_h = (z^2) (r) (1-r) (f) (k)/(p) (n^*) (e^2)$$

n_h : le paramètre calculé et la taille de l'échantillon

z : est la statistique qui définit le niveau de confiance requis

r : est une estimation de l'un des indicateurs à mesurer lors de l'enquête

f : est l'effet imputable à la conception de l'échantillon (deff), suppose comme étant de 2,0 (valeur par défaut)

k : est le multiplicateur visant à tenir compte du taux prévu de non-réponse

p : est la proportion de la population totale représentée par la population cible sur laquelle est fondé le paramètre r

n^* : est la taille moyenne (nombre de personnes par ménage)

e : la marge d'erreur à ne pas dépasser

Certaines valeurs sont recommandées pour obtenir une bonne taille d'échantillon

z : La statistique z à utiliser devrait être de **1,96** pour un degré de confiance de **95 %**, cette valeur est considérée comme la norme pour déterminer le niveau de confiance à atteindre pour évaluer la marge d'erreur dans une enquête sur les ménages. Sa valeur de **1,645** pour un degré de confiance de **90 %**.

f : la valeur par défaut de l'effet imputable à la conception de l'échantillon **f** ou **deff** est habituellement considérée comme étant de **2,0**.

k : le multiplicateur de non-réponse, **k**, est suivant l'expérience acquise habituellement inférieur à **10 %** dans les pays en développement. Une valeur de **1,1** pour **k** serait par conséquent un choix prudent

e : Pour la marge d'erreur, **e**, il est recommandé de fixer le niveau de précision à **10 %** de **r**; par conséquent, **e = 0,10r**

Enfin, une version réduite de la formule peut être utilisée lorsque les valeurs des paramètres recommandés sont utilisés et donnent :

$$n_h = (84,5) * (1-r) / (r) * (p) \quad \text{où :}$$

- **n_h** est le paramètre à calculer et est la taille d'échantillon en termes de nombre de ménages à sélectionner ;
- **r** est une estimation d'un indicateur clé qui doit être mesuré par l'enquête
- **p** est la proportion de la population totale représentée par la population cible et sur laquelle le paramètre **r** est basé

Le traitement des données à problèmes

Les données **nulles**, **invalides** (hors champ) et **les non-réponses** influencent négativement la qualité des résultats d'enquête.

Il existe une multitude de méthodes pour les atténuer, et l'idéal est de les traiter au cas par cas dans le cas du redressement des données.

Quelques méthodes d'imputations

- **La méthode déductive** (pendant l'enquête, la saisie, la vérification)
- **La méthode du hot deck** recourt à d'autres enregistrements (donneurs) de la même enquête pour répondre à la question
- **La méthode de la substitution** repose sur la disponibilité de données comparables (information sur le même sujet lors une opération précédente)

Quelques méthodes d'imputations

- **La méthode de l'estimateur** (une information est tirée d'une autre question ou d'une autre réponse d'un enregistrement)
- **La méthode d'imputation moyenne** (l'insertion dans le champ vierge d'une valeur moyenne tirée des unités déclarantes de même caractéristiques)
- **La méthode du cold deck** (ici, le donneur ne provient pas de la même enquête. Il peut venir d'enquêtes antérieures)
- **La méthode du plus proche voisin** (la valeur attribuée est celle du voisin immédiat)



Traitement des loyers fictifs

(source AFRISTAT)

Le loyer est une somme versée par un locataire en contrepartie de la jouissance d'un bien. Le loyer que le propriétaire du bien ne paie pas correspond à un revenu fictif. Le terme « loyer fictif », ou « loyer imputé », désigne le service de location que les propriétaires de logement se rendent à eux-mêmes.

En d'autres termes, ce sont les loyers que les propriétaires auraient dû payer s'ils étaient locataires du logement dans lequel ils habitent

Traitement des loyers fictifs

(source AFRISTAT)

Trois principaux modes d'estimation

- ❖ Loyer collecté directement auprès des enquêtés (locataire, non locataire, logé gratuitement, autoconsommation);
- ❖ Loyers fictifs estimés par modèle de régression Log linéaire (réaliser la régression par région);
- ❖ Loyers fictifs estimés suivant le modèle de sélection de Heckman.



Traitement des dépenses des fêtes et cérémonies (source AFRISTAT)

En général, le traitement de ces dépenses n'intervient pas dans la consommation des ménages, et cela dans presque tous les Etats membres qui ont réalisé au moins une enquête auprès des ménages sauf en Côte d'Ivoire.

Dans le cas ivoirien, les dépenses des fêtes et cérémonie sont intégrées dans une rubrique « transfert » et sont prises en compte dans l'estimation de la dépense de consommation du ménage.



Traitement des transferts en nature et en espèces (source AFRISTAT)

L'optique consommation adoptée pour la mesure du niveau de vie des ménages recommande la prise en compte des consommations courantes de toute origine.

C'est dans ce cadre que les transferts en nature sont généralement intégrés dans la consommation des ménages dans toutes les études sur la pauvreté dans les Etats membres d'AFRISTAT.

Traitement des biens durables (source AFRISTAT)

La consommation de certains biens s'étale sur une période dépassant largement la période d'observation (l'année). Il est donc nécessaire de ne retenir que la part de la consommation effectivement imputable à l'année d'observation de la consommation.

Cette approche s'applique le plus souvent lorsque l'on convient de retenir au titre de la consommation des ménages, tous les biens acquis sur la période d'observation considérée. Dans ce cas, on ne considère que les biens acquis durant cette période.

Ne pas prendre en compte certains biens **de très grandes valeurs** (moyens de déplacement, certaines dépenses en matériels d'équipement) qui tendrait à surestimer le niveau de consommation du ménage. De plus, ces dépenses sont effectuées de façon très irrégulière.

Traitement des biens durables (source AFRISTAT)

Trois approches de calcul

Intégration globale

Cette approche s'applique le plus souvent lorsque l'on convient de retenir au titre de la consommation des ménages, tous les biens acquis sur la période d'observation considérée. Dans ce cas, on ne considère que les biens acquis durant cette période.

Intégration partielle

Ne pas prendre en compte certains biens de très grandes valeurs (moyens de déplacement, certaines dépenses en matériels d'équipement) qui tendrait à surestimer le niveau de consommation du ménage. De plus, ces dépenses sont effectuées de façon très irrégulière.

Traitement des biens durables (source AFRISTAT)

Valeur d'usage

La consommation de certains biens s'étale sur une période dépassant largement la période d'observation (l'année). Il est donc nécessaire de **ne retenir que la part de la consommation effectivement imputable à l'année d'observation** de la consommation.

Le coût d'usage U_0 , au cours de la période 0 , du bien de consommation durable s'écrit :
$$U_0 = P_0 - \frac{P_{1s}}{(1 + R_0)}$$

- Avec :
- P_0 le prix d'achat du bien durable à la période 0 .
 - P_{1s} le prix de revente du bien durable à la période $1s$
 - $1+R_0$ l'actualisation du terme P_{1s} (valeur équivalente en début de période)
 - R_0 est le taux d'intérêt nominal appliqué au consommateur au début de la période 0 .

Le taux d'intérêt nominal est en général fixé par les pouvoirs publics, l'INS ou un Conseil des normes comptables des pays.

La prise en compte de la structure et de la composition du ménage (échelle d'équivalence) (source AFRISTAT)

Elle stipule ainsi que l'unité de consommation des membres du ménage varie suivant leur groupe d'âge et leur sexe (tableau).

Cette échelle est plus précise sur la composition du ménage en prenant en compte un paramètre supplémentaire relatif au sexe et aux groupes d'âges.

(la Côte d'Ivoire ne le calcul pas pour le moment)

Echelle d'Oxford : Cette échelle affecte le coefficient **1** au premier adulte, le coefficient **0,7** à tout adulte supplémentaire (NA-1) et le coefficient **0,5** à tous les individus du ménages âgés de moins de 14 ans, c'est-à-dire aux enfants (NE).

$$Uc_{\text{oxf}} = 1 + 0,7(NA - 1) + 0,5NE$$

La prise en compte de la structure et de la composition du ménage (source AFRISTAT)

Echelle d'Oxford amendé (Echelle OCDE) : Plus adaptée aux réalités européennes, cette échelle affecte une unité de consommation au premier adulte du ménage, **0,5** unité de consommation à tous les autres adultes (NA-1) entendus comme individus de 14 ans et plus et **0,3** à tous les individus de moins de 14 ans (enfants).

$$Uc_{ocde} = 1 + 0,5(NA - 1) + 0,3NE$$

La prise en compte de la structure et de la composition du ménage (source AFRISTAT)

Echelle de la FAO : Cette échelle est plus précise sur la composition du ménage en prenant en compte un paramètre supplémentaire relatif au sexe et aux groupes d'âges. Elle stipule ainsi que l'unité de consommation des membres du ménage varie suivant leur groupe d'âge et leur sexe (tableau).

Tableau : Echelle des unités de consommation en équivalent adulte

	Échelle Équivalence adulte			Échelle Équivalence adulte	
	Homme	Femme		Homme	Femme
0 – 1 an	0,27	0,27	13 – 15 ans	0,96	0,83
1 – 3 ans	0,45	0,45	16 – 19 ans	1,02	0,77
4 – 6 ans	0,61	0,61	20 – 50 ans	1,00	0,77
7 – 9 ans	0,73	0,73	51 ans et plus	0,86	0,79
10 – 12 ans	0,86	0,73			

Source : FAO

La prise en compte de différences régionales du niveau de vie (source AFRISTAT)

Les différences régionales du niveau des prix jouent un rôle important dans le calcul du seuil de pauvreté car les prix varient parfois fortement d'une région à l'autre, affectant la composition du panier de biens consommés. Il est bon d'en tenir compte pour l'estimation du seuil de pauvreté, qu'il soit absolu ou relatif. En général, deux alternatives sont observées dans les pays.

Dans tous les cas de figure, les déflateurs estimés contribuent à l'harmonisation spatiale des différences de niveau de vie. En divisant l'indicateur du niveau de vie par la valeur de ce déflateur, on obtient une information sur le niveau de vie des ménages capable de garantir la comparabilité des résultats à produire.

La prise en compte de différences régionales du niveau de vie (source AFRISTAT)

Estimation des pseudos déflateurs du coût de la vie

La démarche vise à estimer un seuil de pauvreté spécifique pour chaque région qui ne tient compte que de la structure de consommation et des prix de la région. Sur la base des seuils estimés, on estime le pseudo déflateur comme étant le rapport du seuil de la région considérée sur le seuil de la région qui a été retenue comme référence (en général il s'agit de la principale agglomération ou la capitale économique ou politique).

Ce cas s'applique quand le projet d'étude n'a pas spécialement prévu un dispositif de collecte systématique des prix dans toutes les régions. Elle n'est souhaitable que quand l'information sur les prix n'est pas fiable.



La prise en compte de différences régionales du niveau de vie (source AFRISTAT)

Estimation des pseudos déflateurs du coût de la vie

Cette approche s'applique lorsque l'on a estimé un seuil unique de pauvreté par rapport à une région de référence (en général il s'agit de la capitale ou de la principale agglomération du pays). Elle nécessite donc au préalable de calculer un indice spatial du coût de la vie qui sert de déflateur régional pour les dépenses des ménages. Cette méthode permet d'avoir un seuil national et autorise de ce fait une comparabilité interrégionale. Elle s'applique aussi bien à la pauvreté relative qu'à la pauvreté absolue.



Je vous remercie de votre attention!