



Organisation des Nations Unies
pour l'alimentation et l'agriculture

BASE DE SONDAGE PRINCIPALE (BSP) POUR LES STATISTIQUES AGRICOLES

Module 3 – Séance 4:

**Plan de sondage et estimation lorsque la BSP
est une base de sondage multiple**

Objectifs de la présentation

- Présenter le concept de l'échantillonnage à bases multiples
- Aborder les principes de l'échantillonnage à bases multiples
- Régler les questions découlant de l'utilisation de l'échantillonnage à bases multiples
- Discuter de l'estimation dans le cadre des bases multiples

Plan

- Les principes de l'échantillonnage à bases multiples
- Problèmes dans l'application des enquêtes à bases multiples
- ESTIMATEUR À BASE DE SONDAGE DOUBLE
 - Hartley et l'estimateur de triage
 - L'estimateur de Fuller-Burmeister
 - L'estimateur de Skinner-Rao
 - L'estimateur à base de sondage unique
 - Choisir parmi les estimateurs à base de sondage double
- Estimation des paramètres du domaine
- Utiliser des informations auxiliaires
- Répartition de la taille de l'échantillon entre les bases de sondage

Introduction

- Principe de base: Problème de couverture (unités dans le champ de l'enquête manquantes):
 - Difficulté de fournir une liste de l'intégralité des unités de la population (cible) avec une base unique.
 - En d'autres termes: aucune des bases de sondage unique ne peut couvrir l'intégralité de la population cible.
- Une des solutions consiste à utiliser la base multiple:
 - Dans certaines situations, la couverture peut être améliorée en utilisant plus d'une base d'échantillonnage
 - en effet, plusieurs bases ensemble ont une meilleure couverture qu'une base unique
 - L'avantage est que l'on peut exploiter les points forts et compenser les faiblesses de chaque type de base de sondage.



QU'EST-CE
QU'UNE BASE
MULTIPLE?

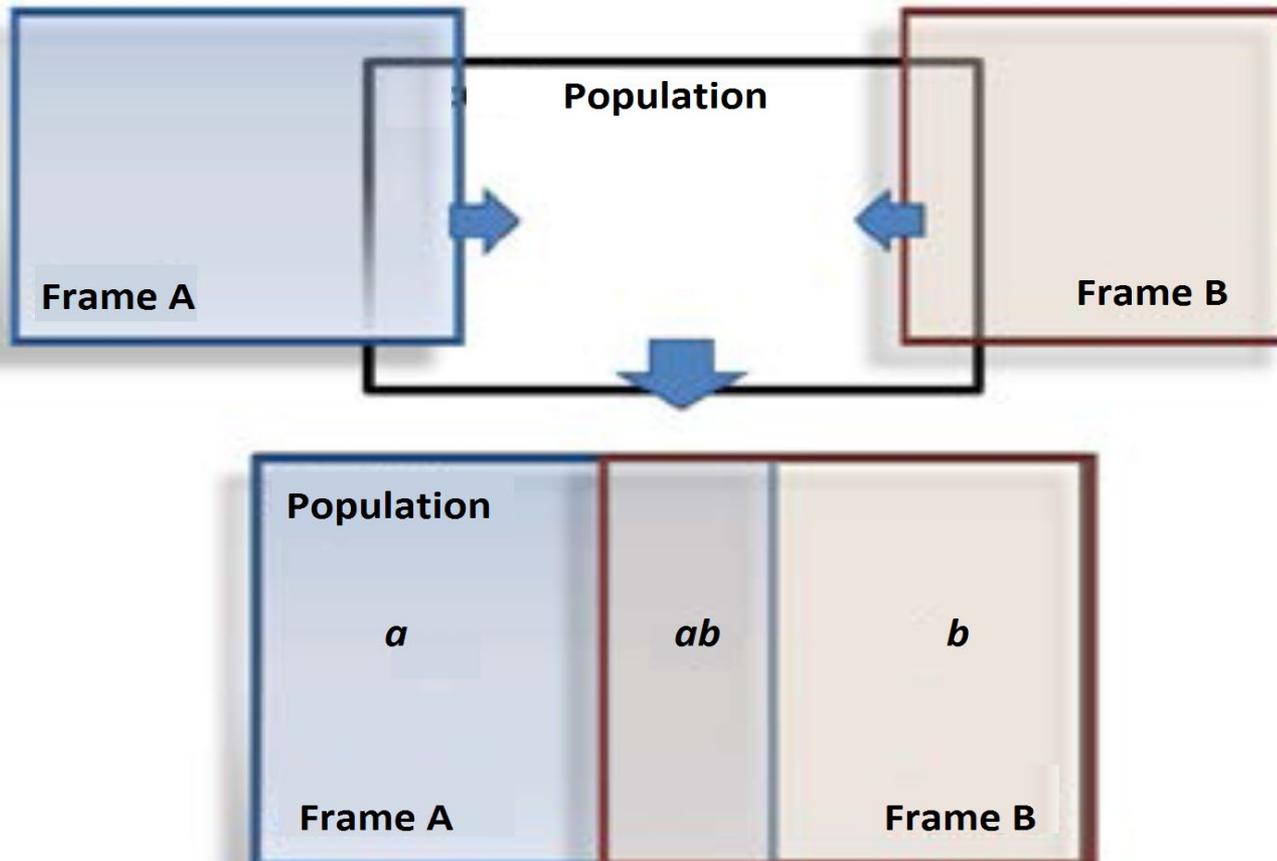
1.1 Définition d'une Base multiple?

- L'échantillonnage à bases multiples implique l'utilisation conjointe de deux ou plusieurs bases de sondage.
- Pour des fins agricoles, cela implique généralement l'utilisation conjointe d'une base de sondage aréolaire et d'une base de sondage de type liste.
- Exemple: Secteur agricole:
 - Utilisation de plusieurs bases de sondage de type liste
 - Liste des producteurs de cultures vivrières, liste des producteurs de cultures commerciales...
 - Utilisation de bases de sondage de type liste et aréolaire

1.2 Exemples de Bases multiples?

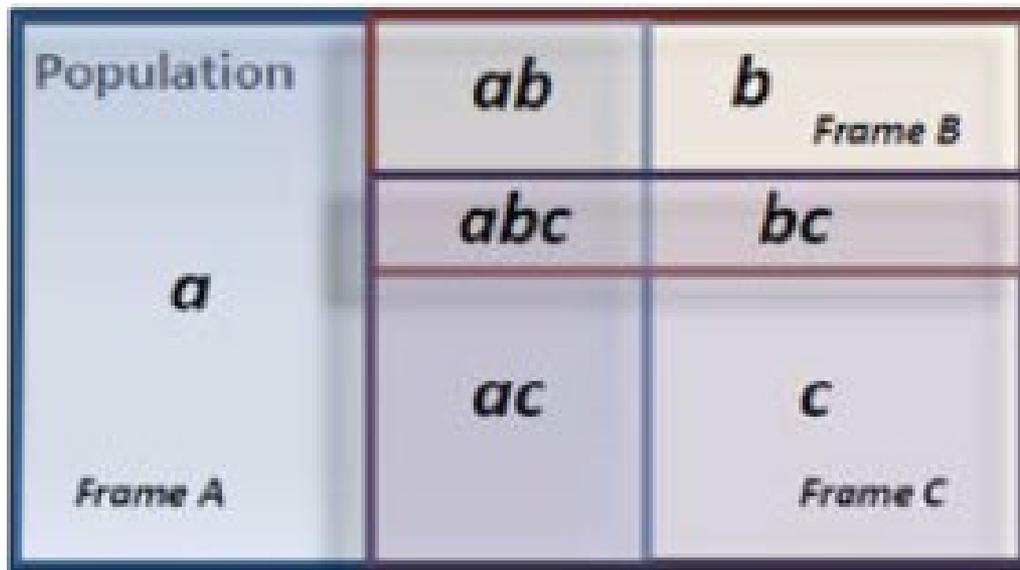
- Lachan et Dennis (1993): échantillonnage de la population des sans-abri à Washington DC (voir aussi Lohr 1999 p.402).
 - Trois bases de sondage ont été considérées:
 - Les refuges pour sans-abris
 - Les soupes populaires et
 - Les campements et les rues

1.2 Exemple de schéma: base de sondage double



1.2 Exemple de schéma

- Trois bases de sondage



1.3 Deux principales exigences de l'utilisation d'une Base multiple

- Il y a deux exigences importantes
 - Complétude, et
 - Identifiabilité
- La **complétude**: L'union des deux bases de sondage offre une couverture complète de la population cible. De cette manière, chaque élément est répertorié dans au moins une des bases de sondage (*reportez-vous à la diapositive suivante*)
- L'**identifiabilité**: Pour tout élément échantillonné, il est possible de comprendre s'il appartient ou pas à une des bases.

1.3 Deux principales exigences de l'utilisation d'une Base multiple (suite)

- **Complétude:**

- Chaque unité de la population appartient à au moins une base de sondage.
- Le concept de «complétude» comporte deux aspects: couverture et informations fournies pour chaque unité de la base de sondage.
- La base de sondage aréolaire est utilisée pour garantir la complétude de la base de sondage principale, car elle est capable de couvrir toutes les exploitations agricoles et leurs terres.

1.3 Deux principales exigences de l'utilisation d'une Base multiple (suite)

- **Identifiabilité:**

- Pour toute unité d'échantillonnage d'une base de sondage quelconque, il est possible de déterminer si l'unité déclarante appartient à une autre base de sondage.
- La condition d'identifiabilité est satisfaite en déterminant quelles unités déclarantes de la base de sondage aréolaire peuvent également être sélectionnées à partir de la base de sondage de type liste



Avantages et problèmes dans l'application des Enquêtes à bases multiples

2.1 Avantages de la Base de sondage multiple

- S'appuient sur les points forts de chaque base et minimisent leurs faiblesses.
- Dans le cas d'une base aréolaire combinée à une base de type liste:
 - La BM (base multiple) permet de créer facilement et à moindre coût des listes d'exploitations agricoles uniquement dans les zones sélectionnées, au lieu de le faire dans l'ensemble du pays.
 - La collecte des données peut s'avérer peu coûteuse parce que les unités d'échantillonnage sont réunies dans les zones sélectionnées, au lieu d'être réparties dans tout le territoire du pays.
- La variabilité peut être contrôlée et mesurée.
- Permet l'étude de produits spéciaux ou rares

2.2 Problèmes dans l'application des enquêtes à bases multiples

- **Les enquêtes à bases multiples comprennent toutes les difficultés des enquêtes à une seule base**
 - Toutes les exploitations agricoles de la **base de sondage de type liste doivent être parfaitement identifiées** par le nom, l'adresse et toute autre forme de nom qui permet d'identifier l'unité agricole.
 - La nécessité de faire correspondre des noms de la base de sondage aréolaire aux noms de la liste des unités statistiques complique le processus de l'enquête et peut comporter des erreurs non liées à l'échantillonnage.
 - **Faire correspondre l'unité d'échantillonnage de la base de sondage aréolaire avec une unité déclarante**, et avec le nom de la base de sondage de type liste. Il est essentiel qu'un nom soit associé au lopin de terre.
 - Choix parmi de nombreuses listes disponibles. Par exemple, une liste de noms peut provenir du recensement agricole et une autre d'une source administrative. Choix d'une liste ou les combiner?

2.2 Problèmes dans l'application des enquêtes à bases multiples

- **Quelques conseils pratiques**
 - **Le besoin d'identifier tous les domaines lorsqu'il y a deux ou plusieurs bases de sondage de type liste** complique grandement l'enquête et le processus d'estimation. Pour cette raison, il est plus pratique de combiner toutes les listes et d'éliminer les doublons avant l'échantillonnage.
 - **Un problème courant est la tentation de faire une liste aussi grande** que possible pour éviter la présence de valeurs aberrantes. Toutefois, plus la liste est grande, plus elle est sujette aux doublons. Plus la liste est petite, plus il est facile d'éviter les doublons et de déterminer le domaine sans chevauchement.
 - **Un autre problème courant est la tentation d'ajouter des noms trouvés dans l'enquête à base de sondage aréolaire** à la base de sondage de type liste. Ces ajouts introduisent un biais vers le bas, car les probabilités d'estimation ont été modifiées (réduites) au moment de l'ajout à la liste.

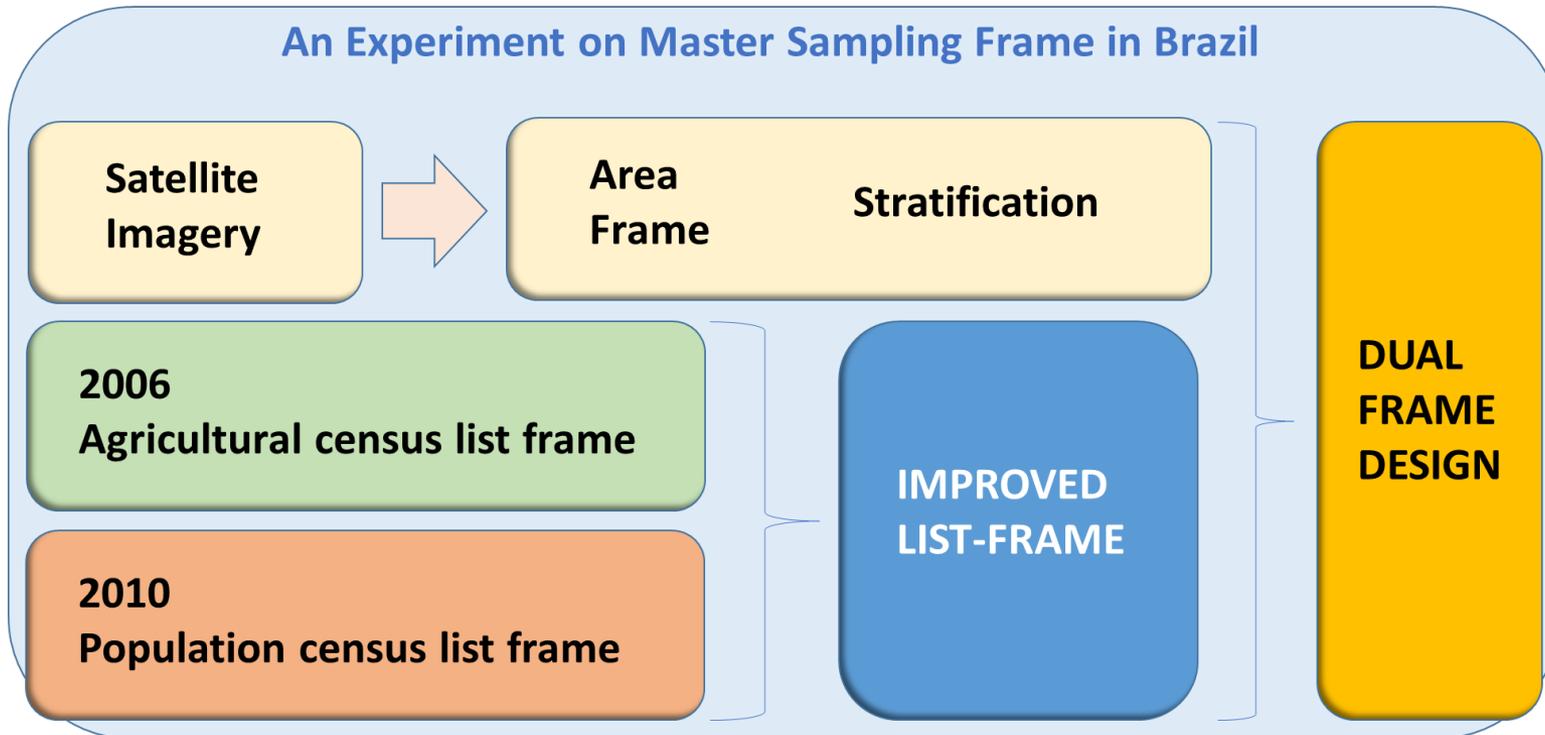
2.2 Problèmes dans l'application des enquêtes à bases multiples Quelques conseils pratiques

- Identifier deux domaines pour la base de sondage aréolaire:
 - 1^{er} Domaine: Celles qui ne sont pas sur la liste (non-chevauchant)
 - 2^{ème} Domaine: Celles qui ne sont pas sur la liste (chevauchant).
- Hypothèse d'identification:
 - Chaque exploitation agricole ou ménage inclus dans la Base de type liste (BL) a une chance d'être sélectionné(e) dans la base de sondage aréolaire (BA).
 - L'identification de domaines doit être basée sur les unités échantillonnées de la base de sondage aréolaire, et non pas sur l'ensemble de la base de sondage aréolaire.
 - Si un nom de la BA ne correspond pas exactement à un nom de la BL, mais est proche, alors il est possible de comparer les adresses ou les noms secondaires associés aux unités déclarantes.

2.3 Échantillonnage à bases multiples: Brésil

- La BSP est utilisée pour le Système d'enquêtes intégrées auprès des ménages, dans lequel toutes les enquêtes individuelles auprès des ménages utilisent la même base de sondage.
- La base de sondage aréolaire était constituée de strates de l'utilisation des sols, établie selon le taux de terres cultivées ou la prédominance des cultures.
- Plusieurs listes d'exploitations qui regroupent un pourcentage important de la variable totale ont été élaborées en se basant sur les informations du recensement agricole de 1985.
- Ces listes, qui comprennent un nombre relativement faible d'exploitations, sont appelées Listes des exploitations spéciales et sont mises à jour annuellement.
- Pour une variable donnée, l'estimateur à bases multiples est la somme des estimateurs des deux échantillons, l'estimateur de l'échantillon aréolaire et l'estimateur de l'échantillon dans le sondage de type liste fondé sur la base de sondage de type liste des exploitations spéciales.

2.3 Échantillonnage à bases multiples Brésil



2.4 Résumé

- Ce module présente les principes de l'échantillonnage à bases multiples.
- Il donne un aperçu des problèmes courants dans l'application de l'échantillonnage à bases multiples du fait du besoin d'identifier le chevauchement entre les bases de sondage.
- Si le terme «échantillonnage à bases multiples» implique que plus de deux bases peuvent être utilisées, la complexité à déterminer le chevauchement entre les bases de sondage semble encourager la limite du choix à deux bases de sondage.
- La conclusion générale est que la base de sondage de type liste contenant principalement de grandes fermes commerciales et des fermes produisant des articles importants mais rares doit être maintenue aussi petite que possible.



Estimateurs à base de sondage double

Principaux estimateurs à base de sondage double

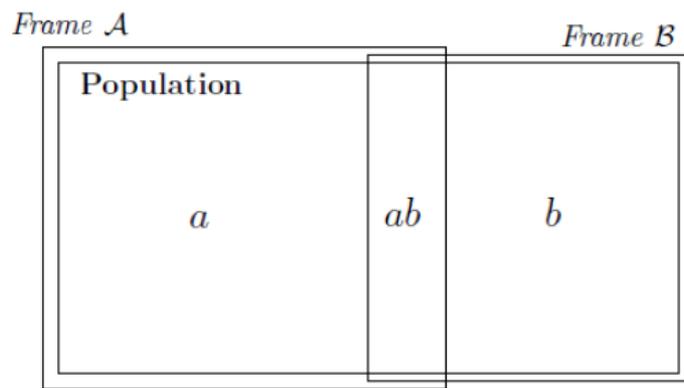
- Hartley et l'estimateur de triage
- L'estimateur de Fuller-Burmeister
- L'estimateur de Skinner-Rao
- L'estimateur à base de sondage unique

3.1 ESTIMATEUR À BASE DE SONDAGE DOUBLE cadre général

- Hartley (1962) a été le premier à définir la méthodologie de la base multiple.
- Des recherches complémentaires sur ce travail ont été entreprises par Cochran (1965), Lund (1968), Fuller and Burmeister (1972), Hartley (1974), et Bosecker et Ford (1976).
- Exemple d'utilisation d'une base multiple: l'échantillonnage de la population des sans-abri à Washington DC (Lachan et dennis 1993, voir aussi Lohr 1999 p.402).
 - Trois bases ont été prises en considération:
 - Les refuges pour sans-abris
 - Les soupes populaires
 - Les campements et les rues

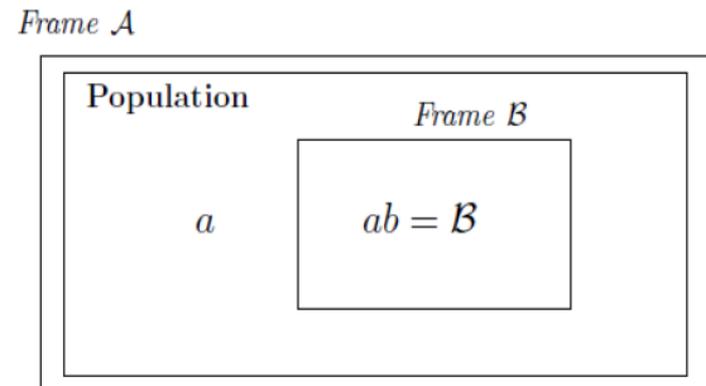
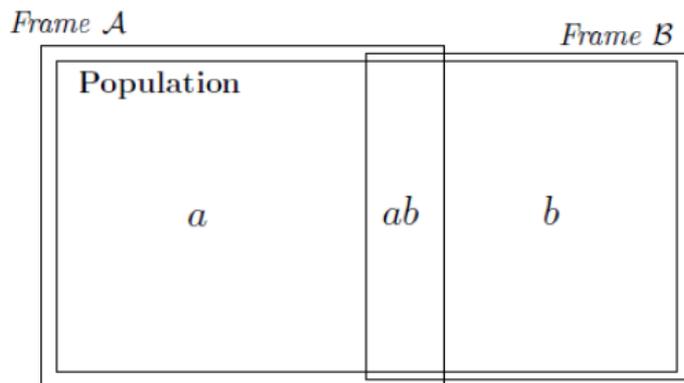
3.1 Estimateur à base de sondage double: contexte (suite)

- La théorie de base de l'échantillonnage à bases multiples (Hartley, 1962; Kott et Vogel, 1995) commence par diviser la population en domaines mutuellement exclusifs.
- L'inférence dans les enquêtes à base prend en considération le scénario général illustré par la figure ci-dessous, dans lequel il est possible d'identifier trois domaines : a , b et ab .



3.1 Estimateur à base de sondage double: contexte (suite)

- La méthode de la base de sondage double combine (par exemple) des échantillons indépendants à partir d'une liste incomplète et une base aréolaire qui est supposée être complète.
- Les deux situations ci-dessous peuvent être envisagées



3.1 Estimateur à base de sondage double: contexte (suite)

- Examinons la formule ci-dessous:
 - π_k^A et π_k^B les probabilités d'inclusion de premier ordre pour les éléments de chacune des deux bases
 - y_k Soit la valeur de la variable d'intérêt pour $k \in U$
 - U_{ab} désigne l'ensemble des éléments de la population appartenant au domaine ab
 - Alors que S_{ab}^A désigne l'ensemble d'échantillons d'éléments de U_{ab} sélectionné à partir de la Base A
 - $N_A = N_a + N_{ab}$ et $N_B = N_b + N_{ab}$

3.2 Estimateurs à base de sondage double: Estimateur de Hartley

- Hartley (1962): peut être exprimé comme une moyenne pondérée entre les estimateurs appropriés d'Horvitz-Thompson (Horvitz et Thompson, 1952) appliquée à chaque domaine de la base de sondage double:

- $$\hat{t}_H = \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + \sum_{k \in S_b} \frac{y_k}{\pi_k^B} + \sum_{k \in S_{ab}} y_k^*$$

- $$y_k^* = \begin{cases} p \frac{y_k}{\pi_k^A}, & \text{si } k \in S_{ab}^A \\ (1 - p) \frac{y_k}{\pi_k^B}, & \text{si } k \in S_{ab}^B \end{cases}$$

- Et p est une pondération constante ($0 < p < 1$) choisie pour minimiser la variance de l'estimateur

3.2 Estimateurs à base de sondage double: l'estimateur de triage

- Si l'on considère une enquête agricole avec Une Base A (base aréolaire) et Base B (base liste), et que B est incluse dans A ($B = A \cap B$), l'estimateur de triage d'Hartley est:

- $\hat{Y} = \hat{Y}_a + \hat{Y}_{ab} = \hat{Y}_a + \hat{Y}_L$, (expression generale)

- $\hat{t}_H = \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + \sum_{k \in S_{ab}} y_k^*$

- $\hat{t}_H = \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + \sum_{k \in S_b} \frac{y_k}{\pi_k^B}$

- $\hat{t}_H = \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + p \sum_{k \in S_{ab}} y_k^* + (1 - p) \sum_{k \in S_b} \frac{y_k}{\pi_k^B}$

3.2 Estimateurs à base de sondage double: l'estimateur de triage

- En supposant qu'un échantillonnage aléatoire simple est appliqué à chaque base, et en désignant les variances de la population du domaine par σ_a^2 et σ_b^2 se rapproche de la variance pour des échantillons stratifiés avec une allocation proportionnelle aux tailles du domaine (ignorant les facteurs de correction de la population finie)

$$\text{Var}(\hat{t}_H) = \frac{N_A^2}{n_A} \left[\sigma_a^2 \left(1 - \frac{N_{ab}}{N_A} \right) + \phi^2 \sigma_{ab}^2 \frac{N_{ab}}{N_A} \right] + \frac{N_B^2}{n_B} \left[\sigma_b^2 \left(1 - \frac{N_{ab}}{N_B} \right) + (1 - \phi)^2 \sigma_{ab}^2 \frac{N_{ab}}{N_B} \right].$$

$$\hat{N}_{ab}^A = \sum_{k \in S_{ab}^A} \frac{1}{\pi_k^A}, \quad \hat{N}_{ab}^B = \sum_{k \in S_{ab}^B} \frac{1}{\pi_k^B}$$

3.3 L'estimateur de Fuller-Burmeister

- Fuller et Burmeister (1972) proposent un estimateur qui utilise des informations de l'échantillon des bases pour estimer N_{ab}
- Par conséquent, l'estimateur de Fuller-Burmeister est donné par:
- $$\hat{t}_{FB} = \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + \sum_{k \in S_b} \frac{y_k}{\pi_k^B} + \sum_{k \in S_{ab}} y_k^* + p_2 (\hat{N}_{ab}^A - \hat{N}_{ab}^B)$$
- $$y_k^* = \begin{cases} p_1 \frac{y_k}{\pi_k^A}, & \text{si } k \in S_{ab}^A \\ (1 - p_1) \frac{y_k}{\pi_k^B}, & \text{si } k \in S_{ab}^B \end{cases}$$
- p_1 et p_2 sont des constantes de pondération ($0 < p_1, p_2 < 1$) choisies pour minimiser la variance de l'estimateur \hat{t}_{FB}

$$\hat{N}_{ab}^A = \sum_{k \in S_{ab}^A} \frac{y_k}{\pi_k^A} \text{ and } \hat{N}_{ab}^B = \sum_{k \in S_{ab}^B} \frac{y_k}{\pi_k^B}$$

3.4 Estimateurs à base de sondage unique

- Auteurs: Bankier (1986) et Kalton et Anderson (1986)
- Cet estimateur s'appuie sur un ensemble de poids d'échantillonnage qui permettent à l'estimateur d'être écrit comme la somme de seulement deux estimateurs de Horvitz-Thompson, chacun couvrant les données de l'échantillon à partir d'une des bases. Il est écrit comme illustré ci-dessous :

- $$\hat{t}_B = \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + \sum_{k \in S_b} \frac{y_k}{\pi_k^B} + \sum_{k \in S_{ab}} \frac{y_k}{\pi_k^A + \pi_k^B}$$

3.5 L'estimateur de Skinner-Rao

- Auteurs: Skinner et Rao (1996)
- C'est un estimateur du pseudo-maximum de vraisemblance (PMV) qui utilise un ensemble unique de poids dans chaque base. Cet estimateur peut être exprimé comme suit:

$$\bullet \hat{t}_{FB} = \frac{N_A - \hat{N}_{ab,SR}}{N} \sum_{k \in S_a} \frac{y_k}{\pi_k^A} + \frac{N_B - \hat{N}_{ab,SR}}{N} \sum_{k \in S_b} \frac{y_k}{\pi_k^B} + \sum_{k \in S_{ab}} \gamma y_k^*$$

$$\bullet y_k^* = \begin{cases} \frac{y_k}{\pi_k^A}, & \text{si } k \in S_{ab}^A \\ \frac{y_k}{\pi_k^B}, & \text{si } k \in S_{ab}^B \end{cases}, \quad \gamma = \frac{\hat{N}_{ab,SR}}{\pi_k^A \hat{N}_{ab}^A + \pi_k^B \hat{N}_{ab}^B}$$

3.6 Choisir parmi les estimateurs à base de sondage double

- Les estimateurs présentés ci-dessus affichent différents niveaux de complexité en fonction de la façon dont ils fournissent des estimations pour le domaine chevauchant *Yab*.
- Il est recommandé que l'estimateur soit choisi en fonction de sa simplicité.
- Les estimateurs de triage sont les plus simples à comprendre et à appliquer dans la pratique, laissant de côté les aspects plus complexes de la collecte de données et du processus de mise en correspondance.

3.6 Choisir parmi les estimateurs à base de sondage double

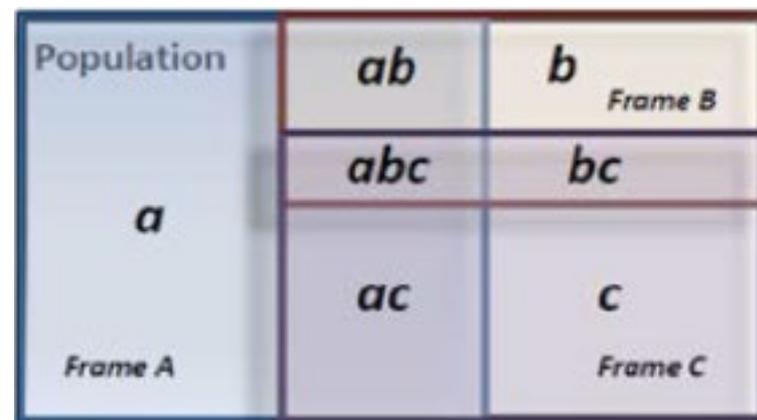
- La précision des estimations peut être améliorée en étudiant la faisabilité et l'efficacité d'autres estimateurs sur la base d'études de simulation à un stade ultérieur
- Ces études cherchant à comparer les performances statistiques des estimateurs à base de sondage double doivent prendre en compte les particularités de chaque pays et des plans d'échantillonnage probabiliste spécifiques.
- Afin d'améliorer la précision, il ne faut pas seulement rechercher des estimateurs à base de sondage double performants, mais également déterminer comment intégrer des variables auxiliaires – qui peuvent être obtenues à partir d'au moins une des bases de sondage – dans le processus d'inférence.



Estimation des paramètres du domaine

4. Estimation des paramètres du domaine

- Par exemple, si trois bases de sondage chevauchantes sont utilisées simultanément dans une approche à bases multiples, sept (2^{3-1}) domaines seront créés,
- En général, pour n base de sondage chevauchante, cela serait 2^{n-1} domaines
- Des domaines seraient indentifiés, et faire des inférences pour chacun d'entre eux est stratégique pour la production d'estimations selon la méthode de l'échantillonnage à bases multiples.



4. Estimation des paramètres du domaine (suite)

- Prenons le Domaine α pour illustrer les estimateurs.
- Supposons qu'un échantillon aléatoire a été sélectionné de la base de sondage A , avec:
- $\pi_k^A = P(k \in S_A)$, comme probabilité que l'unité k soit incluse dans l'échantillon (probabilité d'inclusion de premier ordre)
- $\pi_{kl}^A = P(k, l \in S_A)$, comme probabilité que les unités k et l soient incluses dans le même échantillon (probabilité d'inclusion de deuxième ordre).

4.1 Estimateur de domaine de Horvitz-Thompson

- Si la taille du domaine N_a est inconnue, alors l'estimateur suivant de type Horvitz-Thompson est recommandé (Sarndal, Swensson et Wretmann, 1992; page 390):
- Pour le total du domaine: $\hat{Y}_a = \sum_{k \in S_A} \frac{y_k}{\pi_k^A} \delta_{ak}$
- Pour la moyenne du domaine: $\bar{Y}_a = \frac{\hat{Y}_a}{\hat{N}_a}$
- Où $\delta_{ak} = \begin{cases} 1, & \text{if } k \in a \\ 0, & \text{if not} \end{cases}$ et $\hat{N}_a = \sum_{k \in S_A} \frac{\delta_{ak}}{\pi_k^A}$
- Sa variance et son estimateur de variance peuvent être écrits comme suit:
- $V(\hat{Y}_a) = \sum_{k \in A} \sum_{l \in A} (\pi_{kl}^A - \pi_k^A \pi_l^A) \frac{y_k}{\pi_k^A} \frac{y_l}{\pi_l^A} \delta_{ak} \delta_{al}$
- $\widehat{V(\hat{Y}_a)} = \sum_{k \in S_A} \sum_{l \in S_A} (\pi_{kl}^A - \pi_k^A \pi_l^A) \frac{y_k}{\pi_k^A} \frac{y_l}{\pi_l^A} \frac{\delta_{ak} \delta_{al}}{\pi_{kl}^A}$

4.2 L'estimateur de domaine pondéré π pour le total:

- D'un autre côté, si la taille du domaine N_a est connue, alors l'estimateur pondéré π est réputé fournir une meilleure performance statistique (Sarndal, Swensson et Wretmann, 1992: p. 390):

- $\hat{Y}_a = N_a \bar{Y}_a$

- La variance de l'estimateur de domaine pondéré π peut être approximée par:

- $$V(\hat{Y}_a) = \sum_{k \in A} \sum_{l \in A} (\pi_{kl}^A - \pi_k^A \pi_l^A) \frac{y_k - y_a}{\pi_k^A} \frac{y_l - y_a}{\pi_l^A} \delta_{ak} \delta_{al}$$

- $$\widehat{V(\hat{Y}_a)} = \sum_{k \in S_A} \sum_{l \in S_A} (\pi_{kl}^A - \pi_k^A \pi_l^A) \frac{y_k - \bar{Y}_a}{\pi_k^A} \frac{y_l - \bar{Y}_a}{\pi_l^A} \frac{\delta_{ak} \delta_{al}}{\pi_{kl}^A}$$

- $$\bar{Y}_a = \sum_{k \in A} \frac{y_k}{N_a} \delta_{ak}$$



Utiliser des informations auxiliaires

5.1 Utiliser des informations auxiliaires

- Qu'est-ce qu'une information ou des données auxiliaire(s)?
 - Tout type d'information pertinente, précise et utile qui peut être utilisée pour améliorer la précision des décisions ou des estimations
- Importance: utiles pour améliorer le plan de sondage par exemple:
 - Stratification
 - Plan d'échantillonnage
 - Répartition de la taille de l'échantillon (p. ex. par strates)
 - Guide de procédure d'échantillonnage (PPT ou PPTM)
 - Choix de la méthode de sélection

5.1 Utiliser des informations auxiliaires

- Source de données auxiliaires
 - Autres sources de données
 - Recensement de la population ou de l'agriculture,
 - Toute donnée pertinente de l'enquête
 - Données des systèmes de rapports administratifs
 - Rapport annuel
 - Rapport d'activité
 - Annuaire statistique
- Cela est très utile quand une base contient des informations auxiliaires pour améliorer le plan d'échantillonnage ou les estimateurs

5.2 Exemples d'informations auxiliaires

- Taille totale de la population agricole par ZD
- Superficie totale cultivée d'une région spécifique
- Production totale d'une récolte spécifique dans une région donnée
- Écart-type pour une variable spécifique
- Moyenne de la variable (âge, production, superficie,...)

5.2 Utilisation d'informations auxiliaires:

- Tiré du Manuel sur les BSP, 2015:
 - «Compte tenu de la disponibilité d'informations auxiliaires (externes) auprès d'au moins une des bases de sondage, il est souhaitable d'en profiter en intégrant ces informations dans le schéma d'échantillonnage, à l'aide de plans d'échantillonnage PPT ou PMPT, ou de les utiliser pour aider à la construction d'un estimateur par régression plus approprié.»
 - «Une liste de noms sans informations d'identification sur les caractéristiques de la ferme et de sa taille relative en termes de superficie et de nombres de têtes de bétail ne sera statistiquement pas plus efficace qu'une liste des segments ou des points dans la base de sondage aréolaire. La raison principale d'adopter une base de sondage de type liste est l'utilisation de ses informations auxiliaires aux fins d'échantillonnage.»
 - «Les UPE sont sélectionnées avec PPT, la variable auxiliaire étant l'intensité de culture (Delincé, 2015).»

5.2 Utilisation d'informations auxiliaires:

- Tiré du **Technical Report Series GO-08-2015: Lier les bases aréolaire et de type liste dans les enquêtes agricoles**
 - «Le choix du schéma d'échantillonnage dépend de la nature des unités constitutives de la base de sondage en cours d'utilisation, ainsi que de la nature de toute information auxiliaire qui peut être disponible»
 - «Des Informations auxiliaires peuvent être disponibles dans des bases de sondage de type liste, ce qui permettrait l'utilisation de schémas d'échantillonnage efficaces tels que l'échantillonnage stratifié, l'échantillonnage avec probabilité proportionnelle à la taille ou même les deux, ainsi que d'estimateurs par calage et par régression.»
 - «Identifier des méthodes pour améliorer la mise en correspondance de plusieurs sources de données au niveau des unités constitutives de la base de sondage peut appuyer la faisabilité de bases de sondage de bonne qualité, non seulement en ce qui concerne le taux de couverture mais également pour ce qui est de l'accès à des informations auxiliaires.»



Répartition de la taille de l'échantillon entre les bases de sondage

*Comment répartir la taille d'échantillonnage
entre la base de sondage aréolaire et la base de
sondage de type liste?*

Répartition de la taille de l'échantillon entre les bases de sondage:

- Dans une enquête à base de sondage double, compte tenu de la taille estimée d'un échantillon, le problème de la répartition de la taille de l'échantillon entre les bases de sondage doit être traité.
- Supposons que l'estimateur de triage est utilisé et qu'un échantillon aléatoire simple (sans remplacement) est prélevé dans les deux bases de sondage A et L . Alors,
- $\hat{Y}_S = \hat{Y}_\alpha + \hat{Y}_L^*$

Répartition de la taille de l'échantillon entre les bases de sondage

- Maintenant, le problème de répartition avec une base de sondage double est le même que le problème de détermination des valeurs de n_A et n_L qui réduiront au minimum $Var(\hat{Y}_S)$ soumis à des contraintes de coûts.

$$Var(\hat{Y}_S) = Var(\hat{Y}_A) + Var(\hat{Y}_L) = \left(\frac{N_A^2 \sigma_{ay}^{2*}}{n_A} - N_A \sigma_{ay}^{2*} \right) + \left(\frac{N_L^2 \sigma_{Ly}^2}{n_L} - N_L \sigma_{Ly}^2 \right)$$

Supposons que le coût total C impliqué dans l'enquête est tel que $C = c_0 + n_A c_A + n_L c_L$

Où c_0 sont les frais généraux fixes et c_A et c_L représentent le coût d'échantillonnage et d'observation d'un élément des bases A et L respectivement.

$$\sigma_{ay}^{2*} = P_a (\sigma_{ay}^2 + Q_a \mu_{ay}^{2+}), P_a = \frac{N_a}{N}, Q_a = 1 - P_a$$

Répartition de la taille de l'échantillon entre les bases de sondage

- Dans ces conditions, on sait que la répartition optimale est

$$n_A = n \frac{\sqrt{\frac{N_\alpha^2 \sigma_{\alpha y}^{2*}}{c_A}}}{\sqrt{\frac{N_\alpha^2 \sigma_{\alpha y}^{2+}}{c_A}} + \sqrt{\frac{N_L^2 \sigma_{Ly}^2}{c_L}}}$$

$$n_L = n \frac{\sqrt{\frac{N_L^2 \sigma_{Ly}^2}{c_L}}}{\sqrt{\frac{N_\alpha^2 \sigma_{\alpha y}^{2*}}{c_A}} + \sqrt{\frac{N_L^2 \sigma_{Ly}^2}{c_L}}}$$

$$\sigma_{\alpha y}^{2*} = P_a (\sigma_{\alpha y}^2 + Q_a \mu_{\alpha y}^{2+}), P_a = \frac{N_a}{N_A}, Q_a = 1 - P_a$$

Résumé

- L'échantillonnage à bases multiples implique l'utilisation conjointe de deux ou plusieurs bases de sondage.
- Pour des fins agricoles, cela implique généralement l'utilisation conjointe d'une base de sondage aréolaire et d'une base de sondage de type liste.
- Deux exigences importantes (la complétude et l'identifiabilité) sont nécessaires pour l'utilisation d'une base multiple.

Résumé (suite)

- La base multiple tire parti des points forts de chaque base de sondage et minimisent leurs faiblesses.
- Plusieurs méthodes d'estimation avec différents niveaux de complexité sont élaborées dans le cadre des bases de sondage multiples, mais il est recommandé que l'estimateur soit choisi sur la base de sa simplicité. Les estimateurs de triage sont dans la pratique les plus simples à comprendre et à appliquer.
- En outre, lors de la répartition de la taille d'échantillon entre la base de sondage aréolaire et la base de type liste, celle qui sera définie doit réduire au minimum la variance $Var(\hat{Y}_s)$ soumis à des contraintes de coûts.

Références

- Publications et livres
 - FAO, 2015. Programme mondial pour le recensement de l'agriculture 2020 Collection FAO: Développement statistique Séries 15., Vol. 1. Rome
 - Stratégie mondiale pour l'amélioration des statistiques agricoles et rurales., 2016. *Manuel sur les Statistiques des coûts de production agricoles: Guide pour la collecte, la compilation et la diffusion des données.*, Rome, Italie
 - Stratégie mondiale pour l'amélioration des statistiques agricoles et rurales., 2015. *Manuel sur les bases de sondage principale pour les statistiques agricoles: Développement de la base de sondage, conception du plan d'échantillonnage et estimation*, Rome, Italie
 - Stratégie mondiale pour l'amélioration des statistiques agricoles et rurales, 2012. Plan d'action, Rome, Italie
 - Pascal Ardilly, 2006, Les techniques de sondage. Paris: Editions TECHNIP, 675 p
- Document de travail et références web
 - Stratégie Mondiale, 2015. Lier les bases aréolaire et de type liste dans les enquêtes agricoles Série de Rapports techniques GO-08-2015.
 - Cristiano, F., Leite, A., Ospina, R., et al. 2016. Atelier sur les BSP pour les enquêtes agricoles, Harare, Zimbabwe. UNECA
 - Banque mondiale, 2010, Intégration de l'agriculture dans le système statistique national, Atelier

Merci