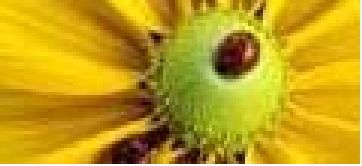




Structure d'échantillonnage

par

Araar Abdelkrim



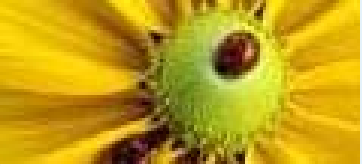
Données d'échantillon

Les données des enquêtes présentent habituellement quatre caractéristiques importantes:

1. ils viennent avec des poids d'échantillonnage (SW), également appelés poids de probabilité inverse;
2. ils sont stratifiés;
3. ils sont regroupés;
4. les observations d'échantillons fournissent des informations agrégées sur les ménages, telles que:

Total des dépenses des ménages;

Nombre d'unités statistiques du ménage (nombre d'individus par exemple).



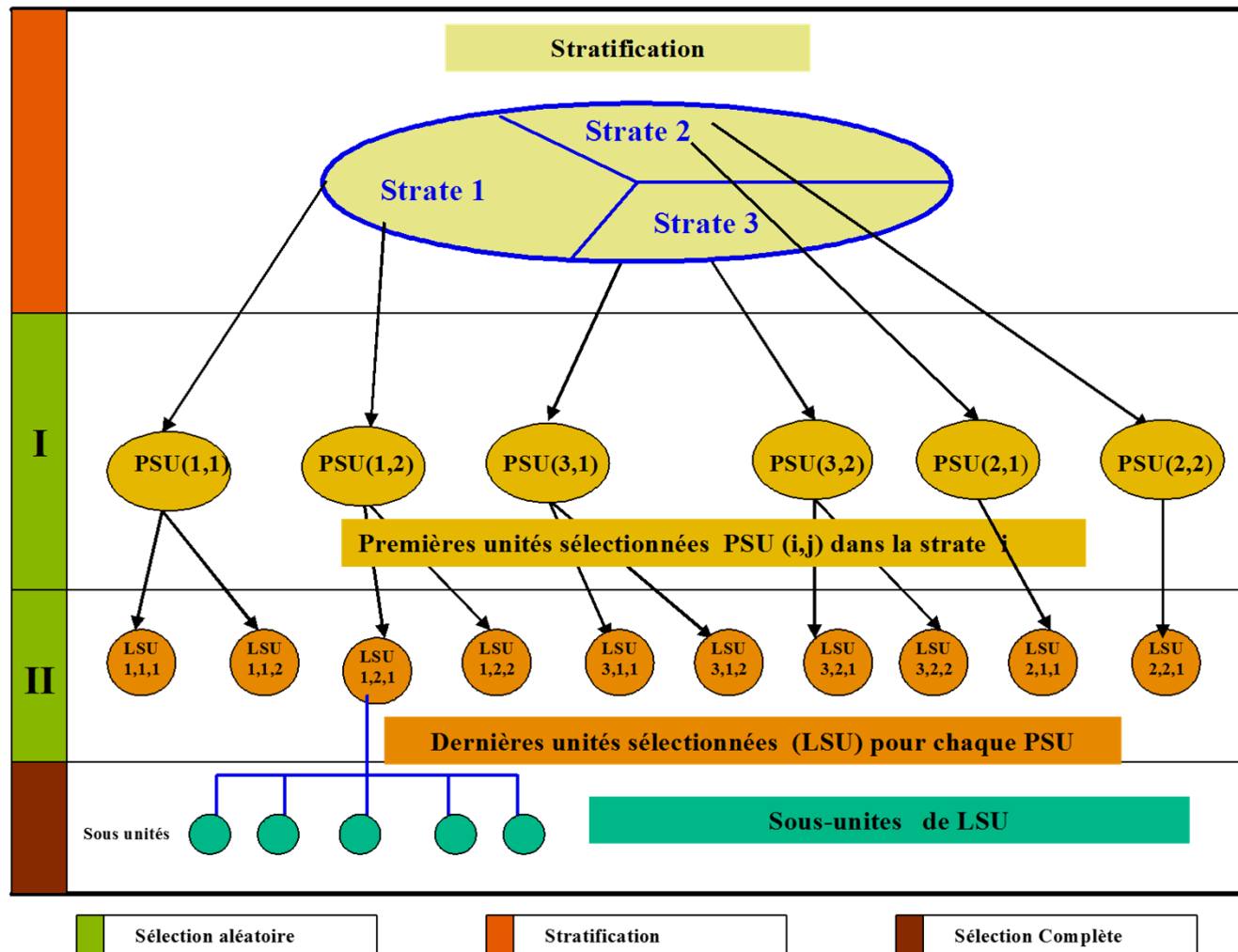
Données d'échantillon

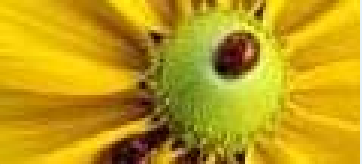
- SRS (échantillonnage aléatoire simple) est rarement utilisé pour générer des enquêtes auprès des ménages.
- Au lieu de cela, un pays est d'abord divisé en zones géographiques ou administratives, appelées strates.
- Un premier niveau de sélection aléatoire est fait en général pour sélectionner des Unités primaires d'échantillonnage (désignées comme PSU) pour chaque strate.
- Les PSU sont souvent des provinces, des départements, des villages, etc. À l'intérieur de chaque PSU, il peut y avoir d'autres niveaux de sélection aléatoire.
- Par exemple, dans chaque province, un certain nombre de villages peuvent être choisis aléatoirement et, dans chaque village sélectionné, un certain nombre de ménages peuvent être choisis aléatoirement.
- Les dernières observations de l'échantillon constituent les dernières unités d'échantillonnage (LSU). Chaque observation de l'échantillon peut ensuite fournir des renseignements agrégés sur l'ensemble des individus qui se trouvent dans ce LSU (tels que les dépenses totales du ménage, la taille du ménage, etc.) .



Conception d'échantillonnage avec échantillonnage en deux étapes

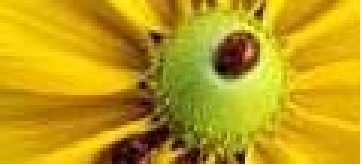
Conception d'échantillonnage à deux niveaux de sélection aléatoire





Stratification et variance des estimateurs

- Les variables d'intérêt, telles que le revenu du ménage, tendent à être moins variables au sein d'une strate que dans l'ensemble de la population.
- Les ménages d'une même strate partagent généralement les mêmes caractéristiques socio-économiques telles que l'emplacement géographique, les conditions climatiques et les caractéristiques démographiques. Ces caractéristiques déterminent le niveau de vie de ces ménages.
- La stratification diminue généralement la variance d'échantillonnage des estimateurs en s'assurant que l'information est recueillie dans toutes les strates.



Impact de groupement (ou tirage multiniveaux)

- En général, les variables d'intérêt (telles que le revenu des ménages, par exemple) varient plus d'une grappe à une autre. Par conséquent, la sélection en plusieurs étapes réduit la «quantité» d'informations générées par l'échantillonnage, car elle ne capture pas autant de grappes que le SRS.
- L'impact des observations d'échantillonnage en grappes tend donc à diminuer la précision des estimateurs et à augmenter ainsi leur variance d'échantillonnage.

Structure de l'échantillonnage et Stata

Dans la fenêtre de commande de Stata, tapez la commande *db svyset* pour ouvrir la boîte de dialogue suivante:

The image displays two side-by-side screenshots of the 'svyset - Survey data settings' dialog box in Stata.

Left Screenshot (Main tab):

- Number of stages: 1
- Sampling units: psu
- Strata: strata
- Finite pop. correction: (empty dropdown)
- Clear settings button
- Note: Empty or "_n" in "Sampling units" above indicates sampling of observations.

Right Screenshot (Weights tab):

- Weight type: Sampling weight variable (selected)
- Weight variable: weight
- Importance weight variable (rare): (empty dropdown)
- Balanced repeated replicate (BRR) weight variables: (empty dropdown)
- Fay's adjustment: (empty text box)
- Jackknife replicate weight variables: (empty dropdown)
- Help Weights... button

Both windows have 'OK', 'Cancel', and 'Submit' buttons at the bottom.



Structure de l'échantillonnage et Stata

Pour afficher les informations la structure de l'échantillonnage dans Stata, utilisez la commande:

svydes

Survey: Describing stage 1 sampling units

pweight: **weight**

VCE: **linearized**

Strata 1: **strata**

SI 1: **psu**

FPC 1: **<zero>**

Stratum	#Units	#Obs	#Obs per Unit		
			min	mean	max
1	42	838	19	20.0	20
2	37	733	17	19.8	20
3	98	1959	19	20.0	20
4	55	1093	19	19.9	20
5	66	1286	13	19.5	20
6	41	779	1	19.0	20
7	97	1937	19	20.0	20
7	436	8625	1	19.8	20



Principales variables de la structure d'échantillonnage

- **Strate** Fournit le nom de la variable (dans un format entier) qui contient les identificateurs Stratum.

PSU Donne le nom de la variable (dans un format entier) qui contient les identifiants des *Primary Sampling Units*.

POIDS DE L'ÉCHANTILLON Spécifie le nom de la variable Poids d'échantillonnage.