

ECHANTILLONNAGE

Pr Ag LEYE

OBJECTIFS

- Définir l'échantillonnage
- Décrire les différentes méthodes empiriques d'échantillonnage
- Décrire les différentes méthodes aléatoires d'échantillonnage
- Estimer la taille de l'échantillon dans les études d'observation avec échantillonnage aléatoire simple

Echantillonnage

- ***Sondage*** = échantillonnage = sélection d'un échantillon à partir d'une population
- ***Echantillon*** = sous- ensemble d'une population obtenu par sondage.
- ***Population*** = ensemble d'***unités statistiques*** ayant une caractéristique commune et susceptibles d'observation.

- **A partir des valeurs observées sur un échantillon, on cherche à obtenir des renseignements sur la population entière, par **estimation** de paramètres**
- **Echantillon **représentatif** : obtenu par **tirage aléatoire** ou procédure équivalente.**

Méthodes d'échantillonnage

Les méthodes empiriques

- **Sondages par choix raisonné : unités statistiques choisies suivant des critères prédéfinis**

La méthode des quota

La méthode des unités- types

- **La méthode des itinéraires**

Méthodes aléatoires

- **Sondage élémentaire**
- **Sondage systématique**
- **Sondage à plusieurs degrés**
- **Sondage en grappes**
- **Sondage stratifié**

Méthodes empiriques

1. Méthode des quotas

- **Principe** = si des individus sont, par certaines caractéristiques, à l'image de la population générale, ils le seront aussi pour les facteurs étudiés.
- On se fixe le nombre de personnes à enquêter pour **chaque catégorie**, pour atteindre, autant que faire se peut, toutes les catégories de population à couvrir par l'enquête.

2. Méthode des unités - types

- Population subdivisée en sous-ensembles homogènes et échantillon constitué d'unités-types capables de résumer chacun de ces sous-ensembles.
- Ex: population du Sénégal résumée par un certain nombre de quartiers représentant grandes agglomérations, certaines communes les villes d'importance moyenne, certains villages le monde rural.

3. Méthode des itinéraires

- Limiter l'initiative laissée aux enquêteurs pour le choix des sujets à interroger, en leur **imposant un itinéraire** précisant les coordonnées exactes des logements sélectionnés.
- Permet d'échapper à la subjectivité de l'enquêteur, mais souvent compliquée.

Méthodes aléatoires

Principes des méthodes aléatoires(1)

- tirer au sort un certain nombre d'individus, qui, s'ils sont en nombre suffisant, pourront être considérés comme représentatifs de la population dont ils sont issus.
- données recueillies à partir d'un échantillon ne peuvent fournir qu'une **estimation de la valeur** réelle du paramètre dans la population dont il est issu.

Principes des méthodes aléatoires (2)

- Valeurs réelles: *fluctuent* en fait dans un certain intervalle autour de la moyenne ou de la proportion observées sur l'échantillon.
- Sachant que l'on accepte un certain risque de conclure à tort, on pourra dire qu'il y a x % de chances que le vrai pourcentage ou la vraie moyenne soient situés dans cet intervalle, appelé *intervalle de confiance*.

Les méthodes de sondage aléatoire nécessitent en général de posséder une liste de la population à étudier : c'est la *base de sondage* (recensement démographique, fichiers administratifs, annuaire téléphonique, liste d'abonnés, liste de logements, etc.).

1. Sondage élémentaire

- Consiste à effectuer tirage au sort directement parmi l'ensemble des unités constituant la population étudiée. Chaque unité a une *probabilité identique* d'être tirée.
- *Unités statistiques* (unités au niveau desquelles sont exprimées les résultats) sont confondues avec les *unités de sondage* (unités à partir desquelles est réalisé l'échantillonnage).

2. Sondage systématique

- C'est un procédé permettant de simplifier la réalisation de la méthode de sondage élémentaire.
- Si l'on désire tirer 150 individus dans une population de 3000, on peut effectuer 150 tirages au sort en procédant ainsi :

on détermine d'abord un **pas de sondage** égal ici à $3000/150=20$

puis on se fixe un point de départ en tirant de manière aléatoire un nombre entre 1 et 20, soit 11.

Le premier individu sélectionné est le 11^{ème} sur la liste, le second le 31^{ème} ($11+20=31$), le troisième le 51^{ème} ($31+20=51$), ainsi de suite.

3. Sondage à plusieurs degrés

- **consiste à effectuer plusieurs tirages successifs.**
- **Tirage à 2 degrés: premier tirage effectué dans les unités primaires. Pour chaque unité primaire, tirage d'unités secondaire effectué.**
- **Ex : tirage au sort d'entreprises, puis, dans chaque entreprise, tirage au sort de sujets interrogés.**

4. Sondage en grappes

- Cas particulier de sondage à plusieurs degrés dans lequel **l'ensemble** des unités secondaires dans chacune des unités primaires tirées sont sondées.
- Ainsi, dans l'exemple ci-dessus, tous les individus travaillant dans les entreprises tirées seraient enquêtés.

5. Sondage stratifié

- moyen d'accroître la précision d'un sondage
- principe : limiter fluctuations dues au hasard en effectuant non plus un tirage au sort unique dans une population hétérogène, mais plusieurs tirages, chacun étant réalisé dans une sous-population **plus homogène** quand au facteur étudié.

population préalablement découpée en différentes classes ou strates, qui, pour apporter un gain de précision, devront être établies à partir d'un critère présentant une forte liaison avec la variable étudiée.

Ex : enquête sur recours aux soins:

stratification efficace : subdiviser la ville concernée en 3 secteurs : un secteur aisé, un secteur de classe moyenne et un secteur de « démunis ».

Tirage au sort des unités de sondage effectué indépendamment à l'intérieur de chacun des 3.

Calcul de la taille de l'échantillon

Détermination nombre de sujets nécessaire

Enquête d'observation avec échantillonnage simple, nombre de sujets à inclure dépend de

- % attendu de sujets présentant le caractère recherché (p)
- précision souhaitée pour ce % (i)
- risque de 1^{ère} espèce consenti (**risque α**)
représentant le risque que le % réel dans la population s'écarte du % observé sur l'échantillon (de plus de 2 écarts-types, pour $\alpha = 0,05$).

nombre de sujets $n = (\varepsilon_{\alpha}^2 * p * q) / i^2$

ε est l'écart- réduit correspondant au risque α consenti (si $\alpha = 5\%$, $\varepsilon = 1,96 \approx 2$).

p est la fréquence approximative du facteur étudié

q est le complément de p ($q = 1-p$).

- **Si sondage en grappes, tenir compte de l'effet de grappe.** Ex : enquête de couverture vaccinale réalisée à partir de 30 grappes : effet de grappe = 2 : 2 fois plus de sujets).

En fait, on est parfois obligé d'adopter la démarche inverse : estimer d'abord le nombre de sujets que l'on pense pouvoir enquêter en fonction des moyens disponibles et déterminer à partir de là la précision escomptée: $= \varepsilon_{\alpha} * \sqrt{(p * q) / n}$.

CONCLUSION

Méthode de référence : échantillon aléatoire

Représentativité

Précisions des estimations grâce à la statistique

Reproductibilité

Comparabilité

Mais

- **Nécessite une méthodologie stricte**
- **Problème des biais de sélection (non répondants)**