Echantillonnage

Pr Mrabet M

Introduction

- **Echantillon:**
  - Taille de l’échantillon: n
    - **Taux de sondage: f**
  - Rapport entre n et population(N)
- **Hétérogénéité:** éviter les biais de sélection
- Échantillon « **satisfaisant** »: **représentativité** (pluralité), **taille**
- **PLAN DE SONDAGE:**
  - **Populations**
  - **Base de sondage**
  - **Taille de l’échantillon**
  - **Types de sondage**

Populations

- **Définir population cible**

  - Exemple:

  On souhaite réaliser une enquête auprès de la pp marocaine âgé de 18 à 65 ans et organiser celle-ci par tel. La population joignable doit don avoir un tel. L'échantillon est constitué à partir de la liste téléphonique (base de sondage). La pp définie par la base de sondage de la liste téléphonique est donc la population source de l'enquête (dans laquelle est puisée l'échantillon)

Base de sondage

- **Outil** (aucun doublon):
  - **liste** (identification des individus en plus de l'age, sexe des individus)
    - Exemple liste téléphonique, liste des logements d'une ville, liste électorale

  - **Découpage géographique**:
    - **Exemple: E**nquête de couverture vaccinale OMS; la liste des enfants de certaines régions géographiques n'existe pas . On utilise alors des bases aréolaires: la méthode consiste en une sélection de zone géographique puis un nouveau tirage au sort d'enfants dans les zones préalablement sélectionnées
- Dès lors la ppp est bien définie, il faut poser les questions suivantes
  - **Pp est elle listée quelque part?**
  - **Peut on obtenir directement ou faut il faire appel à plusieurs listes à combiner?**
    - Ex:

    pour lister l'ensemble des médecins d'une région, il faut tout d'abord obtenir la liste des établissements de santé puis la liste des médecins par établissement

Type de sondage

- **Sondage non aléatoire**
  - **De convenance**: sélectionner les individus qui se présentent à un endroit donné à un moment donné
  - **De volontaires**: les individus se sélectionnent eux- mêmes
  - **Par quotas:** « les caractères décrivant une pp ne sont pas indépendants » variables de contrôle exemple: quotas de sexe

- **Sondage aléatoire:**

Sondage aléatoire

Chaque unité de la population à une **probabilité** connue différente de zéro d'être retenue lors du tirage au sort: ne dépend pas du choix humain, seul le **hasard** régit l'inclusion ou non

Procédure de sélection: il faut:

1. **Obtenir la liste de tous les éléments de la pp source (base de sondage)**
2. **numéroter les éléments de cette liste de 1 à N**
3. **Définir la taille de l'échantillon n**
4. **Tirer au sort n nombres parmi N** (logiciels: fonction alea Excel, table de nombres aléatoires …)

Tirage du premier élément puis sélection des suivants: sondage systématique

- **Définir la fraction d'échantillonnage n/N**
- **Calculer un pas de sondage k= N/n =1/f**
- **Choisir un nombre d au hasard entre 1 et k**
- **L'échantillon est formé des unités d, d+k, D+ (n-1)k**
- **Exemples:**
  - **on souhaite réaliser un échantillon de 20 personnes parmi la population de la clientèle d'un médecin (environ 1000 patients de 20 ans). Il existe une baes de sondage, triée par ordre alphabétique. Le pas de sondage est de 1000/20= 50; un nombre entre 1 et 50 est tiré au sort: supposons que ce nombre soit égal à 13 ; seront inclus dans l'échantillon les individus ayant comme numéros 13, 63, 113….**
  - **On souhaite réaliser une enquête auprès d'une clientèle d'une pharmacie mais aucune liste permettant de construire une base de sondage n'est disponible; on décide donc de réaliser un sondage systématique. La fréquentation de cette pharmacie est d'environ 300 personnes dans la journée et on souhaite sélectionner 60 sujets. On tire au sort un nombre entre 1 et 5 , par ex 4 et on sélectionne donc le 4ème client puis le 9ème et ainsi de suite tous les 5**

Sondage aléatoire simple

- Lorsque la base de sondage existe
- Chaque individu de la pp source peut être associé à un nombre entre 1 et N
- On sélectionne n individus, en tirant au sort n nombres parmi les N

Chaque individu de la pp a la même probabilité d'être sélectionné

Sondage stratifié

Lorsque les bases de sondage possèdent des informations annexes. Il est intéressant de construire l'échantillon en sélectionnant aléatoirement des individus en tenant compte de ces informations. Il faut décomposer la pp source en sous–populations, puis tirer au sort un échantillon dans chaque sous-pp

Exemple:

La liste des médecins d'un établissement de santé comprend la variable sexe de la personne. Il est intéressant de construire l'échantillon en sélectionnant aléatoirement des individus chez les hommes d'une part et chez les femmes d'autre part. on scindera la pp source en une pp de femmes et pp d'homme, puis on tire au sort un échantillon dans chacune de ces deux sous populations.

Ces sous- pp sont appelées **strates**

Une strate donnée est plus homogène que ne l'est la pp dans sa globalité

Les variables **critères de stratification** sont d'autant plus pertinents qu'ils sont en corrélation étroite avec les variables étudiées dans l'enquête: meilleure représentativité

Sondage à plusieurs degrés

Plusieurs niveaux de sondage:

Exemple: Pour réaliser un échantillon auprès des infirmiers hospitaliers d'une région, on peut sélectionner un certain nombre de structures hospitalières à partir de la liste de l'ensemble des établissements de santé de la région. Ensuite, pour chaque établissement sélectionné, on peut demander à l'administration la liste des services, et de nouveau réaliser un sondage aléatoire d'un certain nombre de services. Enfin on peut demander à chaque service de fournir la liste des infirmiers du service et de nouveau réaliser un sondage aléatoire…

Il y'a 3 niveaux de sélection: sondage à 3 degrés

- **unités primaires: les établissements (premier niveau de sélection)**
- **Unités secondaires: les services**
- **Unités tertiaires: les infirmiers**

Sondage en grappes

Il s'agit d'un cas particulier de tirages à plusieurs degrés

Dans le cas d'un sondage à deux degrés, l'ensemble des unités secondaires contenues dans les unités primaires sélectionnées sont intégrées à l'échantillon: les unités primaires sont des « grappes » d'unités secondaires

Exemple: infirmiers dans les établissements de santé. Si on réalise un sondage en grappes, tous les infirmiers des établissement sélectionnés en unités primaires seront inclus dans l'échantillon

Ce type de sondage est très utilisé en raison des commodités qu'il présente:

- **il est inutile d'avoir initialement la liste des unités secondaires pour préparer l'échantillonnage: la liste des unités primaires suffit**
- **il permet de concentrer géographiquement l'enquête et donc de réduire le cout**

Dans l'exemple précédant, il est moins difficile d'interroger tous les infirmiers de quelques établissements que d'interroger quelques infirmiers dans tous les établissements

Exercice

Une enquête doit être réalisée auprès des enfants inscrits en école élémentaire dans une région, pour mesurer la prévalence des troubles de la vue dans cette tranche d'âge. Pour cela, les enfants inclus dans l'enquête devront se soumettre à un test d'aptitude visuelle. Le nombre d'enfants inscrits en école élémentaire est supposé d'environ

50 000 sur **300 écoles**. Les écoles sont de taille très différente et le nombre de classes varie d'une école à l'autre. Deux types d'enquêtes sont envisagées par l'organisateur de cette enquête.

1er type d'enquête : l'organisateur envisage de sélectionner aléatoirement une classe d'enfants dans chaque établissement du département puis de faire passer le test visuel à tous les enfants de cette classe.

S'agit-il :

A. d'un sondage aléatoire simple
B. d'un sondage à deux degrés avec stratification au deuxième degré
C. d'un sondage en grappe
D. d'un sondage par quotas
E. d'un sondage en grappe stratifié

Chaque enfant de la population source a-t-il une probabilité non nulle d'être sélectionné ?

A. Oui
B. Non
C. On ne peut rien dire

Chaque enfant de la population source a-t-il la même probabilité d'être sélectionné ?

A. Oui
B. Non
C. On ne peut rien dire

Quels sont les biais possibles dus à l'échantillonnage ?

A. Mauvaise répartition de l'âge
B. Sur-représentation des enfants à troubles visuels
C. Biais de sélection dû à l'organisateur
D. Sur représentation des enfants des petites écoles

2ème type d’enquête :

l’organisateur envisage de sélectionner aléatoirement 20 établissements, puis de tirer au sort 3 enfants dans chaque classe des établissements sélectionnés.

S’agit-il :

A. d’un sondage aléatoire simple
B. d’un sondage à deux degrés avec stratification au deuxième degré
C. d’un sondage en grappe
D. d’un sondage par quotas
E. d’un sondage en grappe stratifié

Chaque enfant de la population source a-t-il la même probabilité d’être sélectionné ?

A. Oui
B. Non
C. On ne peut rien dire

Chaque enfant de la population source a-t-il la même probabilité d’être sélectionné ?

A. Oui
B. Non
C. On ne peut rien dire

- **Réponse:**

  E- d’un sondage en grappe stratifié

- **Commentaires :**

Il s’agit d’un sondage en grappe stratifié (réponse E) : les grappes d’enfants sont les classes sélectionnées, et les strates sont les écoles. Chaque strate de la population est bien représentée puisqu’une classe est sélectionnée dans chaque strate. Cependant, il faut remarquer que la structure de l’âge n’est pas réellement prise en compte et que ce sondage pourrait amener à une sélection uniquement d’enfants de la classe d’âge 8-9 ans par exemple.

Chaque enfant de la population source a-t-il une probabilité non nulle d’être sélectionné ?

A. Oui

B. Non

C. On ne peut rien dire

Commentaires :

Chaque enfant a une probabilité non nulle d’être sélectionné (réponse n° A) : en effet, les écoles sont toutes représentées puisqu' il y a un tirage aléatoire de la classe dans chacune des écoles.

Chaque enfant de la population source a-t-il la même probabilité d'être sélectionné ?

A. Oui
B. Non
C. On ne peut rien dire

Commentaires :

Chaque enfant n'a pas la même probabilité d'être sélectionné (réponse n° 2) : en effet, les écoles sont de taille différente, donc sont constituées de plus ou moins de classes ; un enfant appartenant à une école avec de nombreuses classes aura une probabilité plus faible que sa classe soit sélectionnée qu'un enfant appartenant à une école ayant peu de classes.

Quels sont les biais possibles dus à l'échantillonnage ?

A. Mauvaise répartition de l’âge
B. Sur-représentation des enfants à troubles visuels
C. Biais de sélection dû à l’organisateur
D. Sur représentation des enfants des petites écoles

Commentaires :

Il y a un risque de :

- mauvaise répartition des tranches d'âge (réponse A), puisque ce critère n'est pas contrôlé à l'inclusion.

- sur-représentation des enfants des petites écoles car un enfant d’une petite école a une probabilité plus grande d’être sélectionné (réponse D).

2ème type d'enquête :

l'organisateur envisage de sélectionner aléatoirement 20 établissements, puis de tirer au sort 3 enfants dans chaque classe des établissements sélectionnés.

S'agit-il :

A. d'un sondage aléatoire simple
B. d'un sondage à deux degrés avec stratification au deuxième degré
C. d'un sondage en grappe
D. d'un sondage par quotas
E. d'un sondage en grappe stratifié

Commentaires :

Il s'agit d'un sondage à deux degrés stratifié (réponse B) : il y a en réalité deux niveaux de sélection, les unités primaires sont les établissements (premier tirage aléatoire), puis il y a sélection aléatoire des enfants (les unités secondaires) dans chaque classe, la seconde sélection étant réalisée dans chaque strate constituée par les classes.

Chaque enfant de la population source a-t-il une probabilité non nulle d'être sélectionné ?

A. Oui
B. Non
C. On ne peut rien dire

Commentaires :

Chaque enfant a une probabilité non nulle d'être sélectionné (réponse A) : en effet, les sélections se font par tirage aléatoire.

Chaque enfant de la population source a-t-il la même probabilité d'être sélectionné ?

A. Oui
B. Non
C. On ne peut rien dire

Commentaires :

Chaque enfant n'a pas la même probabilité d'être sélectionné (réponse B) : en effet, les écoles sont de taille différente et les classes sont également de taille différente, or 3 élèves par classe sont sélectionnés, donc les élèves d'une petite classe ont plus de chances d'être inclus dans l'enquête.