Programme du cours SONDAGES

Chapitre 1 : Eléments de la théorie des Sondages

Chapitre 2 : Sondage aléatoire simple

Chapitre 3 : Sondage stratifié

Chapitre 4 : Sondage par grappes

1. : Eléments de la théorie des Sondages

La collecte des données primaires se fait le plus souvent à l’aide d’un sondage. La conception d’un sondage se fait en trois étapes :

* Détermination de la population
* Choix de la méthode d’échantillonnage
* Détermination de la taille de l’échantillon

# Concepts de Base et Notations

# Population et Individu

On appelle population un ensemble d’éléments possédant les informations désirées pour répondre aux objectifs de l’étude.

On note une population de taille N (connue ou inconnue) tel que

Où est appelé individu (ou unité statistique) d’identifiant .

# Base de sondage

On appelle base de sondage une liste qui répertorie tous les individus d’une population.

# Caractère

Un caractère est une qualité que l’on étudie chez les individus d’une population. On le note par C’est aussi la variable d’intérêt qui prend la valeur pour l’individu .

On s’intéresse à obtenir l’information sur le paramètre (inconnu), qui est une fonction de , i.e.,

On distingue deux cas :

**1er cas :**

Si la variable est quantitative alors peut représenter :

* Le total de dans la population , noté
* La moyenne de , notée (c’est la valeur centrale de ).
* Le quantile ou la médiane de.
* La variance ou l’écart-type de L’écart-type corrigé de la population est le réel :

(Il mesure la dispersion de autour de ).

Dans ce cas, on peut calculer, par exemple :

* Le revenu total ou moyen
* Le nombre de chômeurs, …

**2er cas :**

Si la variable est qualitative alors peut représenter des pourcentages d’individus de la population dont la variable prend telle ou telle modalité.

Dans ce cas, on peut calculer par exemple, la proportion d’individus votant pour Mr A.

Pour calculer ou évaluer les paramètres de la population, deux méthodes sont possibles : « *Recensement »*  et « *Sondage »*

# Recensement

C’est la collecte d’informations sur la totalité des individus. On a accès à tous les individus de la population et on peut mesurer les valeurs de pour chacun d’entre eux.

Toutefois, cela n’est pas toujours possible pour des raisons de coût, de temps ou autres contraintes.

# Sondage

On étudie les valeurs de sur un ensemble d’individus issus de la population.

# Echantillon

On appelle échantillon un ensemble d’individus issus d’une population, sélectionnée pour être interrogés.

Soit le nombre d’individus dans un échantillon .

# Méthodes d’échantillonnage

On peut utiliser deux types de méthodes de sélection : probabilistes ou non probabilistes.

* Méthodes probabilistes (ou aléatoires) :

Les individus sont sélectionnés selon un procédé probabiliste , où chaque individu a une probabilité donnée (connue d’avance) d’appartenir à l’échantillon .

* Méthodes non probabilistes (empiriques) :

La sélection repose sur un choix raisonné. C’est le sondage par quotas (coût moins élevé).

# Plan de sondage

On appelle plan de sondage, une procédure probabiliste permettant de sélectionner un échantillon dans une population.

Un plan de sondage est dit :

* *Plan de Sondage Aléatoire*, si chaque individu de la population a une probabilité connue de se retrouver dans l’échantillon.
* *Plan de Sondage Simple*, si chaque individu a la même probabilité d’être sélectionné ; les probabilités sont égales (PE).
* *Plan de Sondage Sans Remise* (SR), si un même individu ne peut apparaître qu’une seule fois dans l’échantillon.
* *Plan de Sondage Avec Remise* (AR), si un même individu peut apparaître plusieurs fois dans l’échantillon et si l’ordre est important pour cet individu.

**Remarque**

En statistique inférentielle, un échantillon est obtenu par le tirage des individus avec remise (AR). Ainsi, un échantillon de individus est la liste des individus obtenus par prélèvements indépendants. Un individu peut être donc prélevé plusieurs fois.

**Définition**

Un plan de sondage peut être définit comme une loi de probabilité sur l’ensemble des parties non vides de la population , noté .

On note la variable aléatoire représentant l’échantillon aléatoire dont la réalisation est

et la dont la loi de probabilité est tel que

.

**Propriétés d’un plan de sondage :**

1. Etant une loi de probabilité alors on a :
2. Le sondeur c’est lui qui décide sur le plan de sondage à utiliser (contrairement à la statistique classique).
3. est fixé a priori mais pas forcément connu.

**Remarque :**

On supposera dans ce coursNe dépend pas de la variable d’intérêt : On dit dans ce cas, que le plan est non-informatif.

On notera la taille de l’échantillon . Elle peut être la même ou non pour tous les échantillons.

**Exemple :**

Considérons la population . L’ensemble des parties non vides de est :

Alors, les différents échantillons (ou encore ) sont :

* Sont des échantillons de taille 1.
* Sont des échantillons de taille 2.
* est toute la population car

Dans un échantillon, chaque individu est interrogé, et on note la valeur de pour cet individu.

Les valeurs sont utilisées pour construire un estimateur de ф.

On cherche à inférer les résultats de l’échantillon ω à la population U.

D’où, il faut obtenir la précision de .