

CHAPITRE II

Échantillon

1. Introduction

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

2. Population

Population

Ensemble des éléments possédant les informations désirées pour répondre aux objectifs de l'étude.

Échantillon

Ensemble des individus sélectionnés dans la population pour être enquêtés.

Base de sondage

Représentation des éléments de la population : liste interne, annuaire...

3. Méthodes d'échantillonnage

L'échantillonnage consiste à sélectionner les individus dans la population.

Deux grandes familles de méthodes

- méthodes probabilistes (ou aléatoires) : tous les individus ont la même chance d'être sondés.
- méthodes non probabilistes (ou empiriques) : la sélection repose sur un choix raisonné.

Le choix peut dépendre de l'existence de base de sondage.

Méthodes probabilistes

- *Échantillon aléatoire simple* : les éléments sont tirés de manière aléatoire dans la base de sondage.
- *Échantillon systématique* : les éléments sont tirés de manière régulière dans la base de sondage.
- *Échantillon stratifié* : La population est découpée en sous-populations (strates) homogènes. Les éléments sont sélectionnés de manière aléatoire dans chaque strate.
- *Échantillon en grappes* : La population est découpée en sous-populations (grappes) hétérogènes. On sélectionne ensuite de manière aléatoire les grappes à retenir.

Méthodes non probabilistes

- *Échantillon de convenance* : le choix des individus est basé sur des critères pratiques.
- *Échantillon par jugement (ou a priori)* : le choix des individus est basé sur les compétences ou la représentativité de l'individu.
- *Échantillon « boule de neige »* : L'échantillon est construit progressivement sur proposition des individus sondés.
- *Échantillon par quotas* : La population est décrite par quelques caractéristiques descriptives (quotas) puis on construit un échantillon ayant les mêmes caractéristiques.

Exemple d'échantillon par la méthode des quotas

	Population	Echantillon
Hommes	51.61 %	516
Femmes	48.39 %	484
Total	100 %	1000

Exemple d'échantillon par la méthode des quotas croisés

Population	20-29 ans	30-49 ans	50-65 ans	Total
Hommes	15.25 %	20.37 %	12.77 %	48.39 %
Femmes	20.84 %	17.54 %	13.23 %	51.61 %
Total	36.09 %	37.91 %	26.00 %	100 %

Echantillon	20-29 ans	30-49 ans	50-65 ans	Total
Hommes	152	204	128	484
Femmes	208	175	133	516
Total	360	379	261	1000

4. Taille de l'échantillon

Deux facteurs antagonistes

- Précision des résultats
- Coût (et durée) de collecte.

Objectif

- Taille minimale
- Pour une précision fixée *a priori*.

Estimation statistique

- généralisation à la population entière des résultats observés sur l'échantillon
- nécessite un échantillon probabiliste
- de taille $n \geq 50$ et $\leq 1/10$ de la population

Erreur d'estimation

$$\text{erreur} = |\text{grandeur échantillon} - \text{grandeur population}|$$

Cas d'une moyenne (variable quantitative)

$$\text{erreur} = |m_{\text{échantillon}} - m_{\text{population}}| \leq 1.96 \frac{\sigma}{\sqrt{n}}$$

Cas d'une proportion (variable qualitative)

$$\text{erreur} = |p_{\text{échantillon}} - p_{\text{population}}| \leq 1.96 \sqrt{\frac{p(1-p)}{n}}$$

Taille de l'échantillon

- Pour une moyenne (m) :

$$n \geq (1.96 \sigma / \text{précision})^2$$

- Pour une proportion (p) :

$$n \geq (1.96 \times 0.5 / \text{précision})^2 \approx 1 / \text{précision}^2$$

Exemple pour une moyenne

- Précision : 5 €

- Ecart-type : $\sigma = 55 \text{ €}$
- Taille de l'échantillon :

$$n \geq (1.96 \times 55/5)^2 \approx 465$$

- les femmes seront pondérés par $1.17 = 53/45$
- les hommes seront pondérés par $0.85 = 47/55$

Exemple pour une proportion

- Précision : $2 \% = 0.02$
- Taille de l'échantillon

$$n \geq (1.96 \times 0.5/0.02)^2 \approx 2400$$

Ajustement de la taille

Dans la détermination de la taille de l'échantillon, il faut tenir compte du

- *Taux d'incidence* (T_i) : % de répondants correspondants aux critères de l'enquête.
- *Taux d'achèvement* (T_a) : % de répondants qui iront jusqu'au terme du questionnaire.

$$\text{Taille nécessaire} = \frac{\text{Taille souhaitée}}{T_i \times T_a}$$

5. Redressement de l'échantillon

Le redressement a pour but de corriger les défauts de représentativité de l'échantillon :

- sur-représentation de certaines catégories d'individus ;
- quotas non respectés ;
- collecte empirique.

Trois méthodes sont possibles :

- extraction a posteriori d'un sous-échantillon ;
- suppression de questionnaires ;
- pondération.

Redressement par extraction a posteriori

Lorsqu'on possède un échantillon important mais biaisé, il est possible d'en extraire un sous-échantillon aléatoire et/ou respectant des quotas.

Exemple

Sur un échantillon de 3000 personnes, on peut extraire un sous-échantillon aléatoire de 1000 personnes respectant les quotas de sexe et d'âge de la population.

Redressement par suppression de questionnaires

On supprime des questionnaires (au hasard) dans les catégories sur-représentées.

Exemple

La population mère est composée de 53 % de femmes mais l'échantillon comporte que 45 femmes pour 100 questionnaires.

On supprime 15 questionnaires d'hommes (au hasard).

La proportion de femmes est alors de $45/85 = 52.9 \%$

Redressement par pondération

Cela consiste à attribuer une importance différente à chaque observation par un coefficient (poids) positif.

Un poids égal à 1 représente un individu non pondéré.

Exemple

Une population mère est composée de 53 % de femmes mais l'échantillon ne comporte que 45 de femmes pour 100 questionnaires.