

RÉSUMÉ DES CONCEPTS DE STATISTIQUE ET D'ANALYSE DE DONNÉES

La statistique

- ▶ Branche des mathématiques qui a trait à la collecte des données, au classement, à l'analyse et à l'interprétation des données afin d'en tirer des conclusions et de faire des prévisions.

Population et échantillon

- ▶ L'ensemble des personnes, des objets, faits ou événements visé par une étude constitue ce qu'on appelle **population**.
- ▶ Ce sur quoi porte précisément l'étude est appelé **caractère** ou **variable**. C'est le sujet de la question ou de l'étude.
- ▶ Les éléments de la population sont appelés **individus** ou **unités statistiques**.
- ▶ Un **échantillon** est une partie, un sous-ensemble d'une population. Il doit être représentatif de la population.
- ▶ La **taille (n)** de l'échantillon est le nombre d'éléments qu'il contient.
- ▶ Dans une **population homogène**, l'ensemble des individus se ressemblent par rapport au caractère étudié.
- ▶ Dans une **population hétérogène**, les individus sont très différents par rapport au caractère étudié.

Choisir la sorte d'étude à réaliser

• Recensement

- ▶ Étude qui porte sur **toute** la population. La population entière est interrogée;
- ▶ Degré de certitude très élevé;
- ▶ Beaucoup de ressources nécessaires (\$\$\$).

• Sondage

- ▶ Étude qui porte sur une partie de la population. On interroge seulement un échantillon de la population;
- ▶ Un certain degré d'incertitude;
- ▶ Utilisé quand la population est trop nombreuse ou inaccessible;
- ▶ Peu coûteux.

Procédés pour recueillir des données dans l'échantillon choisi

- ▶ Questionnaire écrit;
- ▶ Entrevue téléphonique ou en personne;
- ▶ Observation directe (avec une grille d'observation);
- ▶ Observation documentaire (bases de données);
- ▶ Utilisation d'instruments mécaniques ou électroniques;
- ▶ Internet.

Choisir un échantillon représentatif de la population avec une méthode d'échantillonnage

• Aléatoire

- ▶ Échantillon pris au hasard dans la population;
- ▶ Avec population homogène;
- ▶ On peut identifier tous les membres de la population.
 - Ex: Parmi 30 personnes d'un groupe, piger un des noms dans un sac.

• Systématique

- ▶ Échantillon formé à partir d'un procédé que l'on applique systématiquement;
- ▶ Avec individus de la population classés (liste de clients, bottin, base de données).
 - Ex: D'une page du bottin téléphonique, sélectionner une personne à tous les 25 noms.

• Stratifié

- ▶ Échantillon divisé en sous-groupes proportionnels aux sous-groupes (strates) de la population;
- ▶ Avec une population hétérogène.

• Par grappes

- ▶ Échantillon choisi dans une population homogène étant divisée en groupes (grappes) représentatifs;
- ▶ Échantillon formé de tous les individus des grappes choisies;
- ▶ Grappes petites et hétérogènes.

Les sources de biais empêchant un échantillon d'être représentatif (ce qui peut causer des erreurs dans les résultats)

- ▶ Échantillon non représentatif de la population;
- ▶ Mauvaise formulation de la question;
- ▶ Échantillon trop petit;
- ▶ Collecte de données incomplète;
- ▶ Traitement des indécis incorrect;
- ▶ Implication et intérêt des personnes interrogées;
- ▶ Erreur de mesure ou lecture imprécise;
- ▶ Taux de réponse trop faible;
- ▶ Mauvaises conditions environnementales pour la cueillette des données;
- ▶ Influence ou l'attitude de l'enquêteur;
- ▶ Rejet d'une trop grande partie de l'échantillon;
- ▶ Présentation inadéquate des résultats obtenus (graphiques inadéquats, titres accrocheurs, conclusions irréelles, etc.);
- ▶ Choix de réponses incorrect.

Comparaison d'échantillons

- ▶ Le plus grand échantillon n'est pas toujours le plus représentatif.
- ▶ La précision d'un sondage augmente si on accroît la taille de l'échantillon, mais pas de façon proportionnelle.

Il existe 2 types de variable dans une étude statistique

• Qualitative

- ▶ Les données recueillies sont des mots, des qualités.

• Quantitative

- ▶ Les données recueillies sont des nombres.

▶ Quantitative discrète

- Prend des valeurs numériques isolées. Souvent des entiers.

▶ Quantitative continue

- Prend des valeurs nombreuses incluses dans une classe ou un intervalle. Souvent des nombres décimaux.

• La moyenne (\bar{x})

- ▶ C'est le centre d'équilibre de la distribution.
- ▶ C'est la somme des données divisée par le nb de données.
- ▶ Se calcule avec la formule

$$\bar{x} = \frac{\text{Somme des données}}{\text{Nombre de données}}$$

