

DETERMINATION DE L'ECHANTILLON

GASMI Abdelkader

October 14, 2024

POSITION DU PROBLEME

La détermination d'un échantillon nécessite la détermination:

- ▶ de sa taille : nombre d'éléments à extraire de la population;
- ▶ de la qualité de ses éléments : nature des éléments à extraire (caractéristiques démographiques, comportements, etc.).

Les objectifs essentiels que doit remplir un échantillon sont de 2 sortes:

- ▶ Il doit être représentatif de la population mère, c'est-à-dire qu'il doit donner des informations sur cette population avec la précision exigée ;
- ▶ Il doit avoir un coût compatible avec le budget disponible.

DETERMINATION DE LA TAILLE DE L'ECHANTILLON.

La détermination de la taille d'un échantillon dépend essentiellement de deux facteurs :

- ▶ La précision souhaitée : plus on souhaite des résultats précis, plus l'échantillon nécessaire est important.
- ▶ Le budget disponible : plus on augmente la taille de l'échantillon, plus le coût de l'enquête s'accroît.

La taille de l'échantillon doit être celle qui permet d'atteindre le meilleur équilibre entre le risque d'erreurs d'échantillonnage, le coût induit par ces erreurs, et le coût de l'échantillonnage lui-même.

Cas où la loi de probabilité de la variable aléatoire n'est pas connue.

Dans ce cas on utilise l'inégalité de Bienaymé Tchebycheff, elle aboutit, en général, à des échantillons de tailles élevées.

Taille d'échantillon pour estimer une moyenne.

- ▶ La taille de l'échantillon dépend de la précision souhaitée pour la généralisation des résultats.
- ▶ La précision (ou erreur d'échantillonnage) s'exprime en valeur absolue ou relative. Elle représente la largeur de l'intervalle de confiance de la moyenne. Soit ϵ la moitié de cette largeur.

Pour obtenir un maximum de fiabilité dans les résultats, on commence par se fixer une marge d'erreur ϵ que l'on accepte ; on se fixe ensuite un seuil de confiance $(1 - \alpha)$, qui représente la probabilité minimale pour que la moyenne calculée à partir de l'échantillon ne s'écarte pas de la moyenne de la population de plus de ϵ . Ceci s'écrit :

$$P(|\bar{X} - m| < \epsilon) \geq 1 - \alpha$$

Ce qui donne:

$$n = \frac{\sigma^2}{\epsilon^2 \cdot \alpha}$$

avec :

- ▶ n : Taille de l'échantillon ;
- ▶ ϵ : Précision souhaitée ;
- ▶ α : Définit le seuil de confiance $(1 - \alpha)$;
- ▶ σ : Ecart- type d'échantillon, il est souvent inconnu, il faut avoir des informations antérieures ou mener une étude pilote.

Taille d'échantillon pour estimer une proportion.

- ▶ La taille de l'échantillon dépend de la précision souhaitée pour la généralisation des résultats.
- ▶ La précision (ou erreur d'échantillonnage) s'exprime en valeur absolue ou relative. Elle représente la largeur de l'intervalle de confiance de la proportion. Soit ϵ la moitié de cette largeur.

Pour obtenir un maximum de fiabilité dans les résultats, on commence par se fixer une marge d'erreur ϵ que l'on accepte ; on se fixe ensuite un seuil de confiance $(1 - \alpha)$, qui représente la probabilité minimale pour que la fréquence calculée à partir de l'échantillon ne s'écarte pas de la proportion dans la population de plus de ϵ . Ceci s'écrit :

$$P(|f_n - p| < \epsilon) \geq 1 - \alpha$$

Ce qui donne:

$$n = \frac{p \cdot q}{\epsilon^2 \cdot \alpha}$$

avec :

- ▶ n : Taille de l'échantillon ;
- ▶ ϵ : Précision souhaitée ;
- ▶ α : Définit le seuil de confiance ($1 - \alpha$);
- ▶ p : Proportion dans la population ($q = 1 - p$). Elle est souvent inconnue, il faut avoir des informations antérieures ou mener une étude pilote, sinon on utilise une proportion de 50%.

Cas où la loi de probabilité de la v.a est une loi normale

Dans ce cas, on utilise la loi normale.

Taille d'échantillon pour estimer une moyenne

Cas des prélèvements dans une population finie avec remise ou dans une population infinie sans remise

Pour obtenir un maximum de fiabilité dans les résultats, on commence par se fixer une marge d'erreur ϵ que l'on accepte ; on se fixe ensuite un seuil de confiance $(1 - \alpha)$, qui représente la probabilité minimale pour que la moyenne calculée à partir de l'échantillon ne s'écarte pas de la moyenne de la population de plus de ϵ . Ceci s'écrit :

$$P(|\bar{X} - m| < \epsilon) \geq 1 - \alpha$$

Ce qui donne :

$$n = Z^2_{1-\frac{\alpha}{2}} \frac{\sigma^2}{\epsilon^2}$$

Cas des prélèvements dans une population finie sans remise

De la même manière, on trouve :

$$n = \frac{Z^2 \frac{\alpha \sigma^2 N}{1 - \frac{\alpha}{2}}}{\epsilon^2 N + Z^2 \frac{\alpha \sigma^2}{1 - \frac{\alpha}{2}}}$$

Taille d'échantillon pour estimer une proportion

Pour obtenir un maximum de fiabilité dans les résultats, on commence par se fixer une marge d'erreur ϵ que l'on accepte ; on se fixe ensuite un seuil de confiance $(1 - \alpha)$, qui représente la probabilité minimale pour que la fréquence calculée à partir de l'échantillon ne s'écarte pas de la proportion dans la population de plus de ϵ . Ceci s'écrit :

$$P(|f_n - p| < \epsilon) \geq 1 - \alpha$$

Cas des prélèvements dans une population finie avec remise ou dans une population infinie sans remise.

On trouve :

$$n = Z^2_{1-\frac{\alpha}{2}} \frac{pq}{\epsilon^2}$$

avec :

- ▶ n : Taille de l'échantillon ;
- ▶ ϵ : Précision souhaitée ;
- ▶ p : Proportion dans la population ($q = 1 - p$). Elle est souvent inconnue, il faut avoir des informations antérieures ou mener une étude pilote, sinon on utilise une proportion de 50%.

Cas des prélèvements dans une population finie sans remise

On trouve :

$$n = \frac{Z^2_{\alpha} pqN}{\epsilon^2 N + Z^2_{\alpha} pq}$$