

Fiche synthèse ➡ Échantillonnage

I. Intervalle de fluctuation

Dans quel contexte utilise-t-on un intervalle de fluctuation ?

On considère un caractère C dont la probabilité p d'apparition est connue ou supposée connue. On extrait un échantillon de taille n et on souhaite avoir une estimation de la fréquence d'apparition de caractère C dans cet échantillon.

On utilise donc un intervalle de fluctuation lorsqu'on connaît la probabilité p , ou on émet une hypothèse sur p et on veut une estimation de la fréquence f .

Définition

Soit $\alpha \in]0 ; 1[$. Soit Z la variable aléatoire qui suit une loi normale centrée réduite $\mathcal{N}(0 ; 1)$.

On appelle u_α l'unique réel vérifiant $p(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$.

Un intervalle de fluctuation asymptotique de la fréquence f au seuil de $1 - \alpha$, est :

$$I_n = \left[p - u_\alpha \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

En particulier :

- un intervalle de fluctuation asymptotique au seuil de 95% est donné par :

$$I_n = \left[p - 1,96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$$

Interprétation

Cela signifie que pour n assez grand, la fréquence f a environ 95% de chances d'être dans l'intervalle I_n .

Remarques

- La borne de gauche est approchée par défaut, la borne à droite est approchée par excès.
- Un intervalle de fluctuation asymptotique au seuil de 99% est donné par : $I_n = \left[p - 2,58 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 2,58 \times \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$

En pratique

- on vérifie que p est connue ou supposée connue
- on identifie la valeur de n et on vérifie les trois conditions suivantes : $n \geq 30$; $np \geq 5$; $n(1-p) \geq 5$
- on calcule l'intervalle de fluctuation asymptotique (le plus souvent, au seuil de 95% demandé).
- Dans le cas où l'on teste une hypothèse émise sur p :
 - on calcule la fréquence observée dans l'échantillon donné et :
 - si $f \notin I_n$ alors on rejette l'hypothèse que la proportion est p ;
 - si $f \in I_n$ alors l'hypothèse émise sur p est validée avec un risque d'erreur de 5%.

II. Intervalle de confiance

Dans quel contexte utilise-t-on un intervalle de confiance ?

On considère une population dans laquelle on souhaite estimer une proportion p inconnue d'apparition d'un caractère C dans une population. Pour cela, on prélève un échantillon de taille n dans cette population et on appelle f la fréquence d'apparition de ce caractère dans cet échantillon.

Définition

L'intervalle $I_c = \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]$ est appelé intervalle de confiance de la proportion p au niveau de confiance de 95%.

En pratique

- on identifie la valeur de n et on calcule la valeur de f dans l'échantillon.
- on vérifie les trois conditions suivantes : $n \geq 30$; $nf \geq 5$; $n(1-f) \geq 5$
- on calcule l'intervalle de confiance qui donne une estimation de la proportion p .

Remarque

Il est bon de noter que l'amplitude de l'intervalle de confiance est $\frac{2}{\sqrt{n}}$.

I. Intervalle oui mais lequel ? Questions en vrac...

Situation 1

Dans un pays, 10% des plages étaient atteintes par des algues toxiques. On a modifié le processus de rejets chimiques : on admet que le nouveau processus de rejet, très différent du précédent, pourrait modifier cette proportion.

On prend un échantillon aléatoire de 150 prélèvements, on constate que 18 présentent des traces d'algues toxiques.

Peut-on penser que le nouveau traitement a un impact sur le pourcentage de plages polluées ?

Situation 2

Un industriel fabrique des smartphones. Pour contrôler la qualité de la production, il en teste 200 : 92% fonctionnent correctement.

Qu'en conclure sur l'ensemble de la production ?

Situation 3

Le gérant d'un service de transports urbains annonce fièrement : « il y a suffisamment de bus en circulation pour que les usagers soient assis 95% du temps ». Sébastiana pense que ce n'est pas possible. Elle fait un relevé sur 120 trajets et note qu'elle n'a pas pu s'asseoir durant 27 trajets.

Que peut-elle penser de l'affirmation du gérant ?

Situation 4

D'après une réglementation sur les cultures OGM, le seuil acceptable de contamination des cultures biologiques par des cultures OGM pour être qualifiées de « sans OGM » est de 0,9%.

Après une analyse faite par un laboratoire, Bernard, producteur de produits biologiques, a été informé que sur les 1000 plants de maïs testés dans son champ, 35 provenaient d'une culture OGM.

Préparer sa défense.

Situation 5

Le pourcentage de personnes obèses en France est de 12,4%. On cherche à déterminer un intervalle de fluctuation asymptotique de la fréquence des personnes obèses dans une entreprise de 1400 salariés :

- au seuil de 80% d'abord
- au seuil de 99% ensuite.

On arrondira les bornes de l'intervalle à 10^{-3} près.