Échantillonnage

⋇Échantillon:

On appelle échantillon de taille n un ensemble de résultat obtenus après la répétition indépendante d'une expérience aléatoire n fois.

Exemple 1:

- Le lancer d'un dé n fois.
- Le sondage de n consommateurs sur un produit.
- La production hebdomadaire d'une usine est de 10000 unités. Parmi les unités produites, on prélève 1000 unités. On dit dans ce cas qu'on a prélevé un échantillon de taille $n = 1000$.

⋇Intervalle de fluctuation (ou de pari):

Il permet d'avoir une idée sur les écarts que peut avoir une proportion théorique dans un échantillon donné.

Dans cet intervalle, on devrait trouver la fréquence du caractère observé avec une certitude de 95%.

Autrement dit:

L'intervalle de fluctuation au seuil de 95% (au risque de 5%) d'une fréquence dans un échantillon de taille n, est un intervalle dont le centre est une proportion théorique P, dans lequel la fréquence observée doit y appartenir avec une probabilité de 0.95

Pour $0.2 \leq P \leq 0.8$ et $n \geq 25$. L'intervalle de fluctuation au seuil de 95% de f est : $[P - \frac{1}{\sqrt{n}} \,;\, P + \frac{1}{\sqrt{n}}]$

Remarques :

▶On utilise l'intervalle de fluctuation lorsque la proportion théorique P est connue.

▶Plus n est grand, plus l'intervalle de fluctuation est petit, en effet :

- $n = 25$ l'intervalle de fluctuation est : $[P - 0.2 \,;\, P + 0.2]$ càd $P \pm 20\%$
- $n = 10000$ l'intervalle de fluctuation est : $[P - 0.01 \,;\, P + 0.01]$ càd $P \pm 1\%$

Propriété (Loi des grands nombres):

Plus la taille de l'échantillon n est grande, plus la fréquence observée tend vers la probabilité.

Exemple2:

Un fournisseur de pommes annonce 75% de 1er choix dans sa livraison. On décide de tester cette affirmation. Sur un échantillon de 150 pommes, on y trouve 107 pommes de 1er choix.

Q: L'affirmation du fournisseur est-elle vraie au seuil de 95% ?

Solution :

$P = 0.75 \longleftarrow \textit{Proportion théorique}$

$n = 150 \longleftarrow \textit{Echantillon}$

$f = \frac{107}{150} = 0.71 \longleftarrow \textit{Fréquence observée}$

Pour vérifier si l'affirmation du fournisseur est vraie, on détermine d'abord un intervalle de fluctuation au seuil de 95%; en ayant pour proportion théorique 75% sur un échantillon de taille $n = 150$.

On a bien : $0.2 \leq P \leq 0.8$ et $n \geq 25$ alors :

$IF = [P - \frac{1}{\sqrt{n}}\,;\, P + \frac{1}{\sqrt{n}}]$

$\Longleftrightarrow IF = [0.75 - \frac{1}{\sqrt{150}}\,;\, 0.75 + \frac{1}{\sqrt{150}}]$

$\Longleftrightarrow IF = [0.668\,;\, 0.831]$

Comme la fréquence observée $f = 0.71$ appartient à l'intervalle de fluctuation, donc on en conclut que le fournisseur a raison au seuil de 95%. Autrement dit : Sur 95% des échantillons possibles de taille $n = 150$, la fréquence des pommes de 1er choix est comprise entre 0.668 *et* 0.831.

Exemple3:

À Isfahan, il fait beau un jour sur deux. En 2019 il a fait beau, 125jours sur 225 observés.

Q: Est-ce normal ?

Solution:

À Isfahan, il fait beau un jour sur deux $\Longleftrightarrow$La proportion théorique est : $P = \frac{1}{2}$

il a plu 125jours sur 225 observés $\Longleftrightarrow \begin{cases} n = 225 \\ f = \frac{125}{225} \simeq 0.55 \end{cases}$

Pour vérifier si l'observation est normale, on détermine d'abord un intervalle de fluctuation au seuil de 95%; en ayant pour proportion théorique $\frac{1}{2}$, sur un échantillon de taille $n = 225$.
On a bien : $0.2 \leq P \leq 0.8$ et $n \geq 25$ alors :
$IF = [P - \frac{1}{\sqrt{n}}\,;\, P + \frac{1}{\sqrt{n}}]$
$\Longleftrightarrow IF = [0.5 - \frac{1}{\sqrt{225}}\,;\, 0.5 + \frac{1}{\sqrt{225}}]$
$\Longleftrightarrow IF = [0.43\,;\, 0.57]$
Comme la fréquence observée $f = 0.55$ appartient à l'intervalle de fluctuation, donc on en conclut que l'observation est normale au seuil de 95%.

⋇Intervalle de confiance:
Dans le cas où la proportion théorique est inconnue, il va être utile de déterminer un intervalle de confiance, afin d'estimer un encadrement à P.
On dit dans ce cas qu'on estime une proportion inconnue à partir d'une fréquence observée.

Exemple4: Soit une urne avec des boules de différentes couleurs dont la couleur rouge. Le but est d'avoir une estimation sur la proportion des boules rouges dans l'urne.

$P_{rouge} = \frac{N_{rouge}}{N_{Total}}$

Méthode:
On tire un échantillon de taille $n = 100$.
Parmi les boules on trouve 40 rouges.

On ne peut pas conclure directement que la proportion des boules rouges soit de 0.4 car dans un autre échantillon de même taille, on n'est pas certain d'obtenir 40 boules rouges.

Pour estimer la proportion des boules rouges on détermine un intervalle de confiance au seuil de 95%.

$f = 0.4$ donc $f \in [0.2; 0.8]$, et $n = 100$ donc $n \geq 25$.

$IC = [f - \frac{1}{\sqrt{n}}\,;\, f + \frac{1}{\sqrt{n}}]$
$\Longleftrightarrow IC = [0.4 - \frac{1}{\sqrt{100}}\,;\, 0.4 + \frac{1}{\sqrt{100}}]$
$\Longleftrightarrow IC = [0.3\,;\, 0.5]$
On en conclut que pour 95% des échantillons possibles de taille n=100, la proportion des boules rouges est comprise entre 30% et 50%.

Remarque: Si l'écart de l'intervalle est important, il faudra augmenter la taille de l'échantillon.

Intervalle de confiance:

Soit f la fréquence observée, et n la taille de l'échantillon:
Pour $0.2 \leq f \leq 0.8$ ET $n \geq 25$ on a :
L'intervalle de confiance IC de la proportion théorique P est :
$IC = [f - \frac{1}{\sqrt{n}}\,;\, f + \frac{1}{\sqrt{n}}]$ au seuil de 95% (ou au risque de 5%)

Exemples5:

Une société spécialisée dans le sondage d'opinions relatifs à une élection, effectue une étude d'intention de vote sur 1052 personnes. Parmi les interrogés 614 déclarent vouloir voter pour le candidat du parti des travailleurs.
Q: Est-ce-que ce candidat a raison de penser qu'il va gagner l'élection?

Solution:
En sa basant sur la fréquence observée on va estimer si le candidat sera élu ou pas.
Rmq : Afin que le candidat soit élu il faut qu'il obtienne au moins 50% de voix.

Dans cette perspective, on utilise l'intervalle de confiance, qui a pour centre la fréquence observée. On dispose des données :
$f = \frac{614}{1052} = 0.583$; et l'échantillon $n = 1052$.
Ainsi, on va estimer la proportion théorique grâce à la fréquence observée.

$IC = [f - \frac{1}{\sqrt{n}}\,;\, f + \frac{1}{\sqrt{n}}]$
$\Longleftrightarrow IC = [0.583 - \frac{1}{\sqrt{1052}}\,;\, 0.583 + \frac{1}{\sqrt{1052}}]$
$\Longleftrightarrow IC = [0.55\,;\, 0.61]$.

On en déduit que la proportion théorique varie de 55% à 61% au seuil de 95%. C'est à dire que dans 95% des échantillons possibles de taille $n = 1052$, le candidat peut espérer obtenir entre 55% et 61% des voix.
Ainsi le candidat du parti des travailleurs a raison de penser qu'il va gagner.