Variables aléatoires
Problème de recherche : l'enquête de satisfaction

Une enseigne de magasins affiche dans ses publicités : « 70 % de nos clients sont satisfaits. » Une association de consommateurs interroge au hasard 100 clients de l'enseigne : 58 se déclarent satisfaits.

La mission

1. Un responsable de l'enseigne répond : « Sur 100 personnes, on ne peut pas tomber exactement sur 70. L'écart est dû au hasard. » A-t-il raison de dire qu'on ne trouve pas toujours 70 % ?
2. On suppose que l'affirmation de l'enseigne est vraie. Écrire un programme Python qui simule l'interrogation de 100 clients et renvoie la fréquence de clients satisfaits. L'exécuter plusieurs fois.
3. Compléter le programme pour simuler 1 000 enquêtes de 100 clients, et compter la proportion d'enquêtes qui donnent une fréquence inférieure ou égale à 0,58.
4. Que peut-on en conclure sur la publicité de l'enseigne ?

Pour aller plus loin

Dans les simulations, entre quelles valeurs se trouvent environ 95 % des fréquences observées ? Comparer avec l'intervalle $\left[0{,}7 - \frac{1}{\sqrt{100}} ; 0{,}7 + \frac{1}{\sqrt{100}}\right]$.

Coups de pouce

Bloqué ? Des indices progressifs sont disponibles en ligne. N'en ouvrir qu'un à la fois, et seulement après avoir cherché !

Éléments de correction

1. Le hasard de l'échantillon

Oui : la fréquence observée sur un échantillon varie d'un échantillon à l'autre (c'est la **fluctuation d'échantillonnage**). Même si 70 % des clients sont satisfaits, une enquête sur 100 personnes peut donner 66, 73… La question est de savoir si 58 est un écart *normal*.

2 et 3. Simulation

```
from random import random
def enquete():
    c = 0
    for i in range(100):
        if random() < 0.7:
            c = c + 1
    return c / 100
n = 0
for k in range(1000):
    if enquete() <= 0.58:
        n = n + 1
print(n / 1000)
```

Chaque appel de `enquete()` donne une fréquence proche de 0,7 (par exemple 0,68, 0,73, 0,71…). Sur 1 000 enquêtes, la proportion de fréquences inférieures ou égales à 0,58 est très faible : on obtient en général entre 0,3 % et 1,2 % (la valeur théorique est environ 0,7 %).

4. Conclusion

Si 70 % des clients étaient vraiment satisfaits, une enquête aussi défavorable n'arriverait qu'environ une fois sur 140. Ce n'est pas impossible, mais c'est très peu probable : la publicité de l'enseigne est **douteuse**. L'association a de bonnes raisons de la contester.

Pour aller plus loin

En triant les 1 000 fréquences simulées, on constate qu'environ 95 % d'entre elles sont comprises entre 0,61 et 0,79 environ. L'intervalle proposé est [0,6 ; 0,8] : il contient environ 98 % des fréquences, donc au moins 95 %. La fréquence 0,58 est en dehors de cet intervalle.

Ce qu'on a utilisé du chapitre

— Simulation d'échantillons avec Python.
— Fluctuation d'échantillonnage.
— Loi de Bernoulli.