STATISTIQUES
FLUCTUATION, ESTIMATION

I. PRISE DE DÉCISION : INTERVALLE DE FLUCTUATION AU SEUIL DES 95%

Si l'on effectue un tirage au sort (avec remise) dans une urne contenant une proportion $p = 0,4$ de boules blanches, il y a de grande chances de tirer environ 40 de boules blanches.

Ca n'est généralement vrai, que si l'on effectue un nombre n significativement important de tirages.

Il est possible de prouver que dans des conditions particulières, la proportion correspondra environ à 40 dans 95 des cas. C'est ce qu'exprime l'intervalle de fluctuation asymptotique.

THÉORÈME

Soient p la proportion connue, n le nombre de tirage, et f la fréquence effective (observée).

Si $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, alors :

f est dans l'intervalle $[p - 1,96\,\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,, p + 1,96\,\frac{\sqrt{p(1-p)}}{\sqrt{n}}]$ avec une probabilité de $0,95$

EXERCICE 1 - VÉRIFICATION D'UNE HYPOTHÈSE

La proportion p d'enfants prématurés en France est de 6%. Des chercheurs font l'hypothèse qu'une femme ayant eu un travail pénible pendant sa grossesse a plus de chance d'avoir un enfant prématuré.

Ils réalisent une enquête sur $n = 400$ femmes ayant eu un travail pénible pendant leur grossesse.

Dans cet échantillon, 50 femmes ont eu des prématurés.

1. Déterminer à 10^{-4} près, l'intervalle de fluctuation I au seuil de 95%.
2. Calculer la fréquence effective f de prématurés sur l'échantillon.
3. On valide l'hypothèse si f est plus grande que la borne supérieure de l'intervalle I. Conclure.

CORRECTION

1

$I = [p - 1,96\,\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,, p + 1,96\,\frac{\sqrt{p(1-p)}}{\sqrt{n}}] \simeq [0,037; 0,083]$

Les conditions $n = 400 \geq 30$, $np = 24 \geq 5$ et $n(1-p) = 376 \geq 5$ sont satisfaites, donc l'échantillon est significatif.

2. $f = \frac{50}{400} = 0,125$

3. $f > 0,083$, donc l'hypothèse selon laquelle un travail pénible durant la grossesse acroît les risques de naissance prématurée est validée.

II. INTERVALLE DE CONFIANCE 0,95

Lorsqu'une étude statistique déclare que 35% des français préfèrent la montagne à la plage, il s'agit d'une fréquence ($p = 0,35$) estimée, mais que l'on ne connaît pas réellement (il est impossible d'interroger tous les français).

On peut prouver mathématiquement que cette statistique est fiable à 95% si on a interrogé un échantillon de n français suffisemment grand.

Cet outil mathématique est l'intervalle de confiance 0,95.

THÉORÈME

Soient f la fréquence observée, n la taille d'un échantillon et p la fréquence réelle (inconnue).

Si $n \geq 30$, $nf \geq 5$ et $n(1-f) \geq 5$, alors :

p est dans l'intervalle $[f - \frac{1}{\sqrt{n}} , f + \frac{1}{\sqrt{n}}]$ avec une probabilité de $0,95$

EXERCICE 2 - AIDE À LA DÉCISION

Un médecin veut tester un traitement. Il expérimente sur deux groupes de malades ayant la même pathologie.

- Le groupe A (150 malades) reçoit le traitement. La maladie est réduite pour 121 patients.
- Le groupe B (150 malades) reçoit un placebo. La maladie est réduite pour 94 patients.

1. Déterminer à 10^{-2} près les fréquences f_A et f_B des patients en rémission dans chaque groupe.
2. Déterminer à 10^{-2} près l'intervalle de confiance 0,95 pour chaque groupe.
3. On considère que le traitement fonctionne si les deux intervalles sont disjoints. Conclure.

CORRECTION

1. $f_A = \frac{121}{150} \simeq 0,81$ et $f_B = \frac{94}{150} \simeq 0,63$
2.

Pour le groupe A l'intervalle est $[f_A - \frac{1}{\sqrt{n}} , f_A + \frac{1}{\sqrt{n}}] \simeq [0,73; 0,89]$

Pour le groupe B l'intervalle est $[f_B - \frac{1}{\sqrt{n}} , f_B + \frac{1}{\sqrt{n}}] \simeq [0,55; 0,71]$

3. Les deux intervalles de confiance 0,95 sont bien disjoints, on peut considérer le traitement efficace.