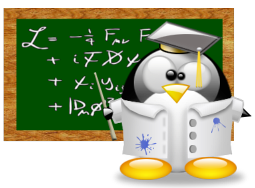
Échantillonnage

Stéphane PASQUET, 19 mai 2016

Sommaire



Prérequis

- Intervalle de fluctuation (classe de Seconde)
- Loi binomiale

Définition du problème

On dispose d'un dé cubique. On souhaite vérifier si ce dé est pipé ou non.

Définition

Intervalle de fluctuation (rappel)

On considère un échantillon de taille $n \geqslant 25$ dans lequel on observe un caractère précis dont la probabilité observée sur toute une population est p $(0,2 \leqslant p \leqslant 0,8)$.
La fréquence f du caractère observé est telle que :

$$f \in \left[p - \frac{1}{\sqrt{n}} \,; 1 + \frac{1}{\sqrt{n}}\right].$$

Ainsi, en lançant notre dé 100 fois (par exemple) et en observant la fréquence d'apparition de la face « 1 » (par exemple), si cette fréquence est dans l'intervalle de fluctuation, nous pouvons considérer que le dé n'est pas pipé.

Exemple

Supposons que la fréquence d'apparition de la face « 1 » est $f = \dfrac{21}{100} = 0,21$.
L'intervalle de fluctuation est :

$$I = \left[\frac{1}{6} - \frac{1}{\sqrt{100}} \,; \frac{1}{6} + \frac{1}{\sqrt{100}}\right] \approx [0,067\,; 0,267].$$

$f \in I$ donc nous pouvons considérer que le dé n'est pas pipé.

Définition

Intervalle de fluctuation de fréquences aléatoires

Soit X une variable aléatoire suivant la loi binomiale $\mathscr{B}(n\,;p)$. On considère alors la variable aléatoire $\mathrm{F} = \dfrac{\mathrm{X}}{n}$ représentant la fréquence (aléatoire) du succès.

Un `intervalle de fluctuation de F` au seuil de 95 % est un intervalle de la forme $\left[\dfrac{a}{n}\,; \dfrac{b}{n}\right]$ tel que $\mathrm{P}\left(\dfrac{a}{n} \leqslant \mathrm{F} \leqslant \dfrac{b}{n}\right) \geqslant 0,95$, ce qui équivaut à $\mathrm{P}(a \leqslant \mathrm{X} \leqslant b) \geqslant 0,95$, avec $0 \leqslant a \leqslant n$ et $0 \leqslant b \leqslant n$.

Méthode

Trouver un intervalle convenable

Dans la pratique, on cherchera a et b tels que :

- a est le plus petit entier tel que $\mathrm{P}(\mathrm{X} \leqslant a) > 2,5\,\%$;
- b est le plus petit entier tel que $\mathrm{P}(\mathrm{X} \leqslant b) \geqslant 97,5\,\%$.

Définition

- Si la fréquence du caractère observé n'est pas dans l'intervalle de fluctuation au seuil de 95 % de la fréquence aléatoire correspondante, on dit que l'on `rejette l'hypothèse` ;
- Si la fréquence du caractère observé est dans l'intervalle de fluctuation au seuil de 95 % de la fréquence aléatoire correspondante, on dit que l'on `ne rejette pas l'hypothèse`.

Exemple

Dans une usine de composants électroniques, les statistiques permettent de dire que la probabilité qu'un certain circuit imprimé soit défaillant est égale à 0,012.
Pour vérifier la production, le chef de production décide de prendre au hasard 100 circuits imprimés à l'issue d'une chaîne de montage.
Notons X la variable aléatoire représentant le nombre de circuits imprimés défaillants. X suit bien une loi binomiale de paramètres $n = 100$ et $p = 0{,}012$.

À l'aide d'un logiciel, on construit la table suivante :

k	$P(X \leqslant k)$
0	0.299016021496
1	0.662193375541
2	0.880540894065
3	0.967172703034
4	0.992688752033
5	0.998639053337
6	0.99978334205
7	0.999969975605
8	0.999996327205
9	0.999999598929
10	0.99999996054

On voit ci-contre que :

- $\mathrm{P}(\mathrm{X} \leqslant 0) > 0,025$. Par conséquent, $a = 0$;
- $\mathrm{P}(\mathrm{X} \leqslant 3) < 0,975$ et $\mathrm{P}(\mathrm{X} \leqslant 4) \geqslant 0,975$ donc $b = 4$.

Par conséquent, un intervalle de fluctuation de la fréquence aléatoire au seuil de 95 % est :

$$I = \left[\frac{0}{100}\,;\frac{4}{100}\right] = [0\,;0,04]\ .$$

Ainsi, s'il y a plus de 4 % de circuits imprimés défaillants parmi ceux pris au hasard, le chef de production pourra conclure que la chaîne de production devra être stoppée.

Remarque

Dans un tableau, pour construire une table comme dans l'exemple ci-dessus, on commencera par entrer « 0 » dans la cellule `A1` puis la formule suivante dans `B1` afin de la copier dans la colonne `B` :

```
=LOI.BINOMIALE(A1;<valeur de n>;<valeur de p>;1)
```