ECHANTILLONNAGE

Le principe :

On considère par exemple l'expérience suivante consistant à lancer plusieurs fois un dé et à noter si la face supérieure affichée est un 4 ou un autre nombre.

La valeur supposée et théorique de la probabilité d'obtenir un 4 est $\frac{1}{6}$.

La mise en défaut ou non de cette expérience, nous permettra d'affirmer s'il est raisonnable de penser que le dé est pipé ou ne l'est pas.

En réalisant l'expérience un certain nombre de fois (échantillon), on mesure la fréquence d'apparition du 4. Si la fréquence observée et la proportion théorique sont trop "éloignées" (dépassent un seuil fixé) alors on peut rejeter la valeur théorique et considérer que le dé est pipé. Dans le cas inverse, on considère qu'il ne l'est pas.

I. Echantillon

On fait l'hypothèse que 55% des électeurs ont voté pour le candidat A.
La proportion théorique est donc : p = 0,55.
Pour savoir si l'on peut accepter cette hypothèse, on interrogera au hasard à la sortie des urnes 50 personnes pour calculer la fréquence observée (paragraphe III).
Si la fréquence observée et la proportion théorique ne sont pas trop "éloignées" alors on pourra accepter l'hypothèse.

Soit X la variable aléatoire qui compte le nombre de personnes qui ont voté pour le candidat A.
X suit une loi binomiale de paramètre n = 50 et p = 0,55.
On a ainsi :

k	17	18	19	20	21	22	23	24	25	26	27
P(X=k)	0,001	0,003	0,006	0,012	0,021	0,034	0,05	0,069	0,087	0,102	0,112

k	28	29	30	31	32	33	34	35	36	37	38
P(X=k)	0,112	0,104	0,089	0,07	0,051	0,034	0,021	0,012	0,006	0,003	0,001

Pour k<17 et k>38, les probabilités sont inférieures à 10^{-3} et peuvent être considérées comme négligeables.

II. Intervalle de fluctuation

Dire que la fréquence observée et la proportion théorique ne sont pas trop "éloignées" signifie que la fréquence observée appartient à un intervalle à définir appelé **intervalle de fluctuation au seuil de 95%**.

Pour un échantillon de taille n = 50, l'intervalle de fluctuation au seuil de 95%

est $I=\left[\frac{a}{50};\frac{b}{50}\right]$ où :

- a est le plus petit entier tel que : P($X \leq a$) > 2,5%
- b est le plus petit entier tel que : P($X \leq b$) ≥ 97,5%

On reprend l'exemple du paragraphe précédent en cumulant les probabilités :

k	17	18	19	20	21	22	23	24	25	26	27
P(X=k)	0,001	0,003	0,006	0,012	0,021	0,034	0,05	0,069	0,087	0,102	0,112
P(X≤k)	0,001	0,004	0,010	0,022	**0,043**	0,077	0,127	0,196	0,283	0,385	0,497

On dépasse 0,025 à partir de k = 21.
Le plus petit a tel que P(X≤a) soit supérieur à 0,025 est a = 21

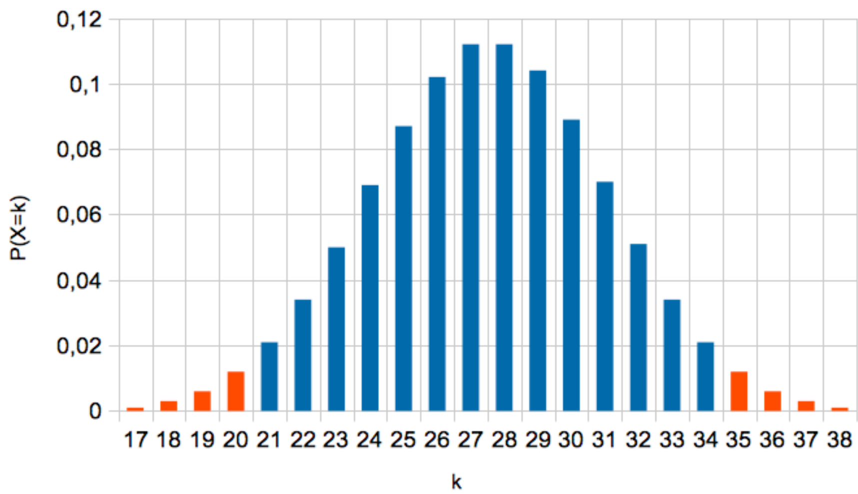
k	28	29	30	31	32	33	34	35	36	37	38
P(X=k)	0,112	0,104	0,089	0,07	0,051	0,034	0,021	0,012	0,006	0,003	0,001
P(X≤k)	0,609	0,713	0,802	0,872	0,923	0,957	**0,978**				

On dépasse 0,975 à partir de k = 34.
Le plus petit b tel que P(X≤b) soit supérieur à 0,975 est b = 34

L'intervalle de fluctuation au seuil de 95 % est $I=\left[\frac{21}{50};\frac{34}{50}\right]=\left[0,42\,;0,68\right]$.

Pour un échantillon de 50 personnes, il y a au moins 95% de chance qu'il y ait entre 42 % et 68 % des électeurs qui votent pour le candidat A.

Sur le graphique, la plus petite somme des probabilités supérieure ou égale à 95 % est représentée en bleue.



Méthode : Déterminer un intervalle de fluctuation au seuil de 95%

Vidéo https://youtu.be/o6bHRO9vHjc

X est une variable aléatoire qui suit une loi binomiale de paramètres n = 6 et p = 0,3.
Déterminer l'intervalle de fluctuation au seuil de 95% de la proportion théorique p.

A l'aide de la calculatrice, on obtient le tableau de valeurs P($X \leq k$) pour k entier compris entre 0 et 6.

X	Y1
0	.11765
1	.42018
2	.74431
3	.92953
4	.98907
5	.99927
6	1

Le plus petit entier a tel que : P($X \leq a$) soit supérieur à 0,025 est a = 0.
Le plus petit entier b tel que : P($X \leq b$) soit supérieur à 0,975 est b = 4.

L'intervalle de fluctuation au seuil de 95 % est $I = \left[\frac{0}{6};\frac{4}{6}\right] = \left[0;\frac{2}{3}\right]$.

Il y a 95% de chance que X soit inférieure à 66,7%.

III. Prise de décision

Règle de décision : Si la fréquence observée appartient à l'intervalle de fluctuation, on accepte l'hypothèse au seuil de 95 %. Dans le cas contraire, on la rejette.

Vidéo https://youtu.be/cxMdYBvywK0

On reprend l'exemple du paragraphe précédent.
On interroge 50 électeurs à la sortie des urnes. Parmi ceux-là, 22 affirment avoir voté pour le candidat A.
Peut-on accepter l'hypothèse que 55% des électeurs ont voté pour le candidat A ?

L'intervalle de fluctuation au seuil de 95 % est $I = \left[0,42;0,68\right]$.

Règle de décision :
Si la fréquence observée appartient à l'intervalle $I = \left[0,42;0,68\right]$ alors on accepte l'hypothèse que 55% des électeurs ont voté pour le candidat A. Dans le cas contraire, on la rejette.

La fréquence observée est égale à $f = \frac{22}{50} = 0,44$.

Comme 0,44 appartient à l'intervalle de fluctuation $I = [0{,}42\,;0{,}68]$, on en déduit qu'on peut accepter l'hypothèse.

On ne peut cependant pas affirmer être certain que l'hypothèse est vraie. En effet, la probabilité de se tromper n'est pas nulle mais égale à 0,05 = 5%.

Les étapes de résolution :

1) Donner les paramètres de loi binomiale suivie par X :
   Proportion théorique p et taille de l'échantillon n.
2) Calculer a et b pour déterminer l'intervalle de fluctuation I.
3) Calculer la fréquence observée f.
4) Appliquer la règle de décision : $f \in I$?