Chapitre 10 : 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, on produit ainsi un échantillon sur lequel on mesure la fréquence d'apparition du 4. Si la fréquence et la valeur théorique sont trop "éloignées" - *c'est-à-dire qu'elles dépassent un seuil fixé* - alors on peut rejeter la valeur théorique et considérer que le dé est pipé. Dans le cas contraire, on considère qu'il ne l'est pas.

I. Echantillon

Définitions : On obtient un **échantillon** de taille n en prélevant n éléments d'une population au hasard, successivement et avec remise.
Un **échantillonnage** est le prélèvement d'un échantillon dans une population.

Exemples :

- On lance plusieurs fois une pièce de monnaie, on note si elle tombe sur "pile" ou "face".
- On tire au hasard une boule dans une urne qui ne contient que des boules rouges et des boules blanches. On s'intéresse au nombre de boules rouges tirées.
- Si l'effectif est grand, on peut considérer qu'un prélèvement d'un échantillon sans remise est assimilé à un prélèvement avec remise. C'est le cas, par exemple, d'un sondage "sortie des urnes" à une élection.

Propriété : On considère une population dont une proportion p des individus possède un caractère donné.
On prélève dans cette population un échantillon de taille n.
La variable aléatoire qui associe le nombre d'individus possédant ce caractère suit une loi binomiale de paramètres n et p.

Exemple :
On fait l'hypothèse que 55% des électeurs ont voté pour le candidat A. On interroge au hasard à la sortie des urnes 50 personnes.
X est la variable aléatoire qui compte le nombre k de personnes qui ont voté pour le candidat A.

La loi de probabilité de X est 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<……. et k>……. les probabilités sont inférieures à 10^{-3} et peuvent être considérées comme négligeables.

<u>Avec le tableur :</u>
Il est possible d'obtenir la loi de probabilité.

B2 =LOI.BINOMIALE(A2;50;0,55;0)

	A	B	C	D	E
1	k	P(X=k)			
2	0	4,6E-018			
3	1	2,8E-016			
4	2	8,4E-015			
5	3	1,6E-013			
6	4	2,4E-012			
7	5	2,6E-011			
8	6	2,4E-010			
9	7	1,9E-009			

Avec la loi binomiale *B*(50 ; 0,55) :
Pour calculer P(X = 20), il faut saisir : =LOI.BINOMIALE(20;50;0,55;0)
Pour calculer P(X ≤ 20), il faut saisir : =LOI.BINOMIALE(20;50;0,55;1)

II. Intervalle de fluctuation

<u>Définition :</u> On considère une population dont une proportion p des individus possède un caractère donné.
On prélève dans cette population un échantillon de taille n.
Soit X la variable aléatoire associée au nombre d'individus possédant ce caractère.
L'<u>intervalle de fluctuation au seuil de 95 %</u> associée à la variable aléatoire X est : $\left[\frac{a}{n};\frac{b}{n}\right]$ où : - a est le plus petit entier tel que : $P(X \leq a) > 2{,}5\%$

- b est le plus petit entier tel que : $P(X \leq b) \geq 97{,}5\%$

<u>Remarque :</u>

Dire que l'intervalle de fluctuation au seuil de 95 % est $\left[\frac{a}{n};\frac{b}{n}\right]$ signifie que pour un échantillon de n personnes, il y a au moins 95 % de chance qu'il y ait entre $100 \text{x} \frac{a}{n}$ %

et $100 \text{x} \frac{b}{n}$ % des individus qui possèdent le caractère donné.

Exemple :

Vidéo https://youtu.be/o6bHRO9vHjc

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)	0,002	0,005	0,01	0,023	0,044	0,077	0,127	0,196	0,283	0,386	0,498

→ → → …

$P(X \leq a) > 0{,}025$
Le plus petit est a = …….

K	28	29	30	31	32	33	34	35	36	37	38
P(X£k)	0,61	0,713	0,802	0,872	0,923	0,957	0,978	0,989	0,995	0,998	0,999

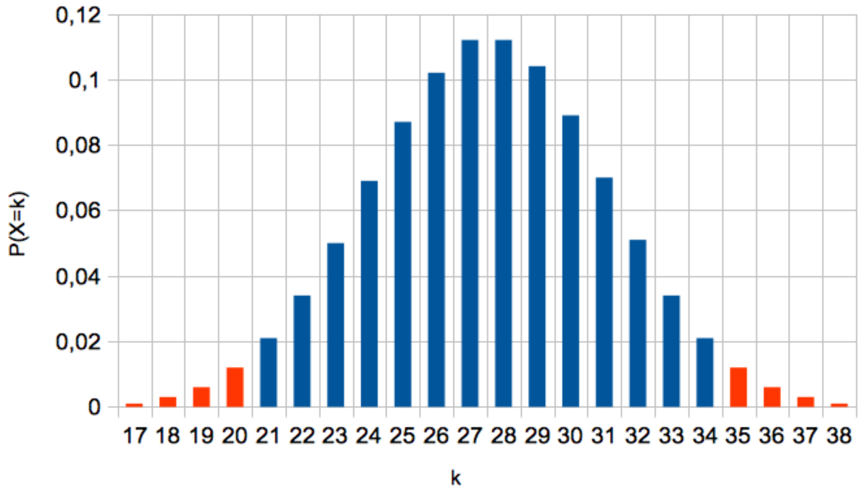
→ → → …

$P(X \leq b) \geq 0{,}975$
Le plus petit est b = ……

L'intervalle de fluctuation au seuil de 95 % est : ………………………………………..

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.



Avec le tableur :
Il est possible d'obtenir les probabilités cumulées.

C2 =LOI.BINOMIALE(A2;50;0,55;1)

	A	B	C	D	E
1	k	P(X=k)	P(X<=k)		
2	0	4,6E-018	4,6E-018		
3	1	2,8E-016	2,8E-016		
4	2	8,4E-015	8,7E-015		
5	3	1,6E-013	1,7E-013		
6	4	2,4E-012	2,5E-012		

Propriété : Pour un échantillon de taille supérieure à 30 l'intervalle $\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$ est une bonne approximation de l'intervalle de fluctuation au seuil de 95%.

Exemple :

Dans l'exemple précédent, on trouve : $\left[0,55-\frac{1}{\sqrt{50}};0,55+\frac{1}{\sqrt{50}}\right]\approx\left[0,41;0,69\right]$

III. Prise de décision

Critère 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.

Exemple :

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 ………………………

La fréquence observée est égale à……………

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 inférieure ou égale à 0,05.