CHAPITRE 10 Echantillonnage

1. Constituer un échantillon par simulation

Rappels

- La formule =ENT(ALEA()+p) simule le tirage d'un élément dans une population où un caractère est présent dans la proportion p. Si le résultat est 1, le caractère est présent pour l'élément tiré. Si le résultat est 0, il n'est pas présent.
- Quand on extrait un échantillon d'une population où un caractère est présent avec la proportion p, on n'observe en général pas exactement cette proportion dans l'échantillon. Cela s'appelle la fluctuation d'échantillonnage. Mais si l'échantillon est suffisamment grand, la fréquence observée ne s'éloigne pas beaucoup de p.

1. Dans un mélange de graines de fleurs hautes, il y a 15 % de graines de lupin.
On s'intéresse à la fréquence de graines de lupin dans des sachets contenant cent graines. La quantité de graines est assez importante pour qu'on puisse considérer un sachet de cent graines comme un échantillon de la production, obtenu par 100 tirages successifs avec remise.

a. Expliquer comment la formule =ENT(ALEA()+0,15) permet de simuler l'étude du nombre de graines de lupin dans un tel échantillon avec le tableur.

b. Calculer la fréquence des graines de lupin dans l'échantillon ci- dessous :

	A	B	C	D	E	F	G	H	I	J
1	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0
3	0	0	0	1	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0
5	1	0	1	0	0	0	0	0	0	0
6	0	0	1	1	0	0	0	0	0	0
7	1	0	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	0	0	0	1
9	0	0	0	0	0	0	0	0	1	1
10	0	0	0	0	0	0	0	0	0	0

c. Expliquer pourquoi la fréquence observée ici n'est pas égale à la proportion dans la population des graines.

2. On lance 100 fois une pièce supposée truquée. La probabilité d'obtenir face est 0,55.

a. Donner une formule qui permet de simuler le lancer de cette pièce.

b. À l'aide du tableur, simuler 100 lancers de cette pièce en préparant un tableau comme ci-dessous :

	A	B	C	D	E	F	G	H	I	J	K	L
1	0	1	1	1	1	1	1	0	1	1		
2	0	1	0	0	1	0	0	1	1	0		Fréquence de face
3	1	0	0	1	1	0	1	1	1	1		0,57
4	0	1	0	1	1	0	1	0	1	0		
5	1	1	1	0	1	1	0	1	0	0		
6	1	0	0	0	1	1	0	1	0	1		
7	1	0	1	0	1	0	1	0	0	1		
8	1	0	1	0	1	0	1	1	1	0		
9	0	1	0	1	0	1	0	1	1	1		
10	1	0	0	1	1	1	0	1	0	0		

c. Que faut-il saisir en L3 pour obtenir la fréquence de face ?

d. Expliquer pourquoi la fréquence que vous trouvez est sans doute différente de celle proposées ici.

2. Utiliser une formule pour déterminer un intervalle de fluctuation

Rappels

- Si la proportion théorique p dans une population est comprise entre 0,2 et 0,8 et si la taille n des échantillons considérés est supérieure à 25, un intervalle de fluctuation au seuil de 95 % de la fréquence observée dans les échantillons de taille n est l'intervalle :

$$I=\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right].$$

- Ceci signifie que dans environ 95 % des échantillons, la fréquence observée est dans cet intervalle.

3. On reprend la situation de l'exercice 1.
Dans ce mélange de graines, il y a 20 % de delphinium et 30 % de lin bleu.
Donner, en utilisant la formule ci-dessus, un intervalle de fluctuation au seuil de 95 % de la proportion de delphinium d'une part et de lin bleu d'autre part pour des échantillons de taille 100.

4. a. Quelle est la probabilité d'obtenir 6 lorsqu'on lance un dé cubique non pipé ?
b. Donner, en utilisant la formule ci-dessus, un intervalle de fluctuation au seuil de 95 % pour des échantillons de taille 50.
On arrondira au millième.

3. Prendre une décision à partir d'un échantillon

Rappels

On sait qu'une population présente un certain caractère dans une proportion p. On suppose que les conditions sont réunies pour utiliser l'intervalle de fluctuation au seuil de 95% rappelé ci-dessus.
On s'intéresse à un échantillon issu d'une sous-population et on se demande si la proportion dans cet échantillon est p.
On observe que le caractère est présent dans l'échantillon avec une fréquence f.
La règle de décision est alors la suivante :

- Si f appartient à I, on ne rejette pas l'hypothèse que la proportion dans la sous-population est p.
- Si f n'appartient pas à I, on rejette l'hypothèse que la proportion dans la sous-population est p.

5. Environ 61 % des français possèdent un Smartphone.
a. Déterminer avec la formule rappelée ci-dessus un intervalle de fluctuation au seuil de 95 % de la fréquence de français possédant un Smartphone dans un échantillon de taille 200.
b. Dans une commune considérée comme défavorisée, le maire a effectué un sondage sur 200 de ses administrés choisis au hasard car il pense que ce pourcentage ne s'applique pas à sa commune.
108 possèdent un Smartphone.
Que peut-on conclure ?

6. Le taux de recyclage des emballages ménagers est d'environ 67 % en France.

a. Donner un intervalle de fluctuation au seuil de 95 % de la fréquence de personnes qui recyclent les emballages ménagers dans un échantillon de taille 500.

b. Une communauté de commune qui a consacré beaucoup d'efforts pour informer les ménages se demande si ses efforts ont payé. Dans cette communauté, sur un échantillon de 500 personnes choisies au hasard, 365 recyclent correctement. Peut-on considérer que l'information a été efficace ?

4. Connaître la loi binomiale

Rappels

- On considère une expérience à deux issues.

Une est appelée succès et a pour probabilité p.
On répète cette expérience n fois de façon identique et indépendante.
On appelle X la variable aléatoire qui donne le nombre de succès.
X suit la loi binomiale de paramètres n et p. On la note $B(n\ ;\ p)$.

- L'espérance de X est égale à $n \times p$.
- $P(X = k)$ et $P(X \leq k)$, pour tout k compris entre 0 et n, s'obtiennent à la calculatrice (voir rabat d du manuel).

7. Dans une usine de fabrication d'assiettes en faïence, on considère que 5 % des assiettes présentent un défaut.
On prélève 10 assiettes au hasard. La production est suffisamment importante pour que l'on puisse considérer qu'il s'agit de tirages avec remise.
On note X la variable aléatoire qui donne le nombre d'assiettes présentant un défaut parmi les dix.
a. Justifier que X suit une loi binomiale dont on précisera les paramètres.
b. À l'aide de la calculatrice, calculer $P(X = 3)$ et $P(X \leq 1)$.
On arrondira au dix-millième.
c. Interpréter ces probabilités.

8. En France, environ 4 % des personnes sont daltoniennes, c'est-à-dire ne perçoivent pas correctement les couleurs.
On choisit au hasard 50 français. Le nombre de français est suffisamment grand pour que l'on puisse considérer qu'il s'agit d'un tirage avec remise.
On note X la variable aléatoire qui donne le nombre de personnes daltoniennes dans l'échantillon.
a. Justifier que X suit une loi binomiale dont on précisera les paramètres.
b. À l'aide de la calculatrice, calculer $P(X = 3)$ et $P(X \geq 4)$.
On arrondira au dix-millième.
c. Interpréter ces probabilités.

Réponses aux exercices complémentaires

1. a. La fonction ALEA() renvoie un nombre aléatoire X tel que
$0 \leq X < 1$. ALEA()+0,15 renvoie donc
$X + 0{,}15$, et $0{,}15 \leq X + 0{,}15 < 1{,}15$.
Par conséquent quand on prend la partie entière, on obtient 1 dans 15 % des cas et 0 dans 85 % des cas.
Il suffira de compter les 1 et de diviser par 100 pour obtenir la proportion de graine de lupins dans l'échantillon.
b. Il y a douze 1, donc la fréquence des graines de lupin est ici 12 %.
c. Ceci est dû aux fluctuations d'échantillonnage.

2. a. la formule est = ENT(ALEA()+0,55.
b. En L3, il faut saisir =SOMME(A1:J10)/100.
c. Ceci est dû aux fluctuations d'échantillonnage. On peut obtenir de nouveaux échantillons en appuyant sur la touche F9.

3. I = *0,2−1 100 ; 0,2+1 100 = 0,1;0,3* pour les graines de delphinium.
I' = *0,3−1 100;0,3+1 100= 0,2 ; 0,4* pour les graines de lin bleu.

4. a. la probabilité d'obtenir 6 est *16*.
b. I = *16−150 ; 16+150*
I ≈*0,025 ; 0,308*

5. a. I = *0,61−1 200 ; 0,61+1 200*
I ≈*0,539 0,681*
b. *108 200=0,54*
Cette fréquence appartient à l'intervalle de fluctuation, donc il n'y a pas lieu de penser que dans cette commune 61 % des habitants possèdent un Smartphone.

6. a. I = *0,67−1 500 ; 0,67+1 500*
I ≈*0,625 ; 0,715*.
b. *365 500=0,73*.
Cette fréquence n'appartient pas à l'intervalle de fluctuation, donc on peut penser que l'information a été efficace.

7. a. Les tirages sont considérés comme avec remise, donc on répète 10 fois la même expérience à deux issues de façon indépendante. Par conséquent X suit la loi binomiale de paramètres 10 et 0,05.
b. $P(X = 3) \approx 0{,}010\,5$
$P(X \leq 1) \approx 0{,}913\,9$
On a calculé la probabilité qu'il y ait une assiette présentant un défaut, puis la probabilité qu'il y ait au plus une assiette présentant un défaut dans un échantillon de 10 assiettes.

8. a. Le choix des 50 français est considéré comme avec remise, donc on répète 50 fois la même expérience à deux issues de façon indépendante. Par conséquent X suit la loi binomiale de paramètres 50 et 0,04.
b. $P(X = 3) \approx 0{,}184\,2$
$P(X \geq 4) = 1 - P(X \leq 3) \approx 1 - 0{,}860\,9$
$P(X \geq 4) \approx 0{,}139\,1$.
On a calculé la probabilité qu'il y ait trois Français daltoniens, puis au moins 4 Français daltoniens, dans un échantillon de 50 Français.