

Chapitre 10 ~ Fluctuation d'échantillonnage & simulations

I – Fluctuation d'échantillonnage

1. Tableau d'effectifs

Définition

On appelle **échantillon de taille n** du caractère X une liste de n résultats obtenus chacun en choisissant un individu dans la population, puis en enregistrant la valeur du caractère X pour cet individu et en le remettant dans la population.

On appelle **échantillon exhaustif de taille n** du caractère X une liste de n résultats obtenus comme précédemment sans remettre l'individu dans la population.



Remarque

Les formules qui suivent ne fonctionnent qu'avec la première méthode, mais on admet qu'elles fonctionnent aussi avec la seconde dès que l'effectif dépasse 30. Dans le langage courant, l'échantillon désigne aussi la population sur laquelle l'étude a porté...

Pourquoi travailler sur des échantillons ?

1. L'étude détruit l'individu : lorsqu'on s'intéresse à la durée de vie des ampoules, on ne peut pas toutes les tester sinon il ne restera plus d'ampoules à vendre...
2. L'effectif est trop grand : lorsqu'on souhaite par exemple s'intéresser aux élections présidentielles, on ne peut pas demander à tous les français leur intention de vote...

Exercices :

lancer 70 fois un dé et construire un tableau d'effectifs et de fréquences (≈ 5 minutes)

Bilan de l'exercice : la probabilité d'obtenir 6 est $\frac{1}{6} \approx 16,6\%$. A-t-on trouvé ces résultats ?

→ Ce phénomène s'appelle la **fluctuation d'échantillonnage**. Quand on compare des échantillons de même taille, plus ils fluctuent, moins ils sont fiables et donc moins on peut généraliser les résultats de l'un des échantillons à l'ensemble de la population. Il n'est donc pas très étonnant que les fréquences obtenues par chacun d'entre nous n'aient pas été très proches des 16,6 % théoriques !

Qu'en est-il lorsqu'on met en commun tous les résultats de la classe : s'approche-t-on davantage de la fréquence théorique ??

2. Intervalle de fluctuation d'une fréquence

Définition

On appelle **échantillon de taille n** du caractère X une liste de n résultats obtenus chacun en choisissant un individu dans la population, puis en enregistrant la valeur du caractère X pour cet individu et en le remettant dans la population.

On appelle **échantillon exhaustif de taille n** du caractère X une liste de n résultats obtenus comme précédemment sans remettre l'individu dans la population.

Exercice : 26 % des Français se déclarent allergiques au pollen. On étudie la fréquence f des personnes allergiques dans un échantillon de taille 400.

1. Au seuil de 95 %, dans quel intervalle f varie-t-elle ?
2. Sur un échantillon de taille 400, on relève 120 personnes allergiques : peut-on dire que le taux d'allergies dans cet échantillon est anormalement élevé ?

Solution :

1. D'après la formule précédente, que l'on peut appliquer car $p = 0,26$ est bien compris entre 0,2 et 0,8, et $n = 400$ est bien supérieur à 30, on a :

$$f \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right] = \left[0,26 - \frac{1}{\sqrt{400}} ; 0,26 + \frac{1}{\sqrt{400}} \right] = \left[0,26 - \frac{1}{20} ; 0,26 + \frac{1}{20} \right] = [0,26 - 0,05 ; 0,26 + 0,05] = [0,21 ; 0,31].$$

95 % des échantillons auront une fréquence observée dans l'intervalle $I = [0,21 ; 0,31]$.

2. Puisque $f \in [0,21 ; 0,31]$, une multiplication par 400 donne que dans 95 % des échantillons, il y a entre $0,21 \times 400 = 84$ et $0,31 \times 400 = 124$ personnes allergiques. On ne peut donc pas dire que le nombre de personnes allergiques dans cet échantillon soit anormalement élevé.

	En classe : 30, 31 p. 282	Exercices : 32 p. 282 + 33 p. 283
--	------------------------------	--------------------------------------

3. Intervalle de confiance d'une proportion

Soient f et p deux réels compris entre 0 et 1, et $n \in \mathbb{N}^*$. Alors

$$\begin{aligned} f \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right] &\Leftrightarrow p - \frac{1}{\sqrt{n}} \leq f \leq p + \frac{1}{\sqrt{n}} \\ &\Leftrightarrow p \leq f + \frac{1}{\sqrt{n}} \leq p + \frac{2}{\sqrt{n}} \quad \text{et} \quad p - \frac{2}{\sqrt{n}} \leq f - \frac{1}{\sqrt{n}} \leq p \\ &\Rightarrow f - \frac{1}{\sqrt{n}} \leq p \leq f + \frac{1}{\sqrt{n}} \quad \Leftrightarrow p \in \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]. \end{aligned}$$

L'implication inverse se montre de la même manière.



Définition

Parmi tous les échantillons de taille n possibles ayant comme fréquence observée f , 95 % des intervalles associés $\left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right]$ contiennent la proportion effective p . Cet intervalle est appelé **intervalle de confiance** de p au seuil de confiance 95 % (ou encore au risque d'erreur de 5 %).

Exercice : Lors d'une élection, un sondage portant sur un échantillon aléatoire de 1000 personnes donne 400 votants en faveur d'un candidat L. Au risque d'erreur de 5 %, quelle information peut-on obtenir sur la proportion réelle d'électeurs envisageant de voter pour L ?

Solution : La définition ci-dessus donne

$$p \in \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}} \right] = \left[\frac{400}{1000} - \frac{1}{\sqrt{1000}} ; \frac{400}{1000} + \frac{1}{\sqrt{1000}} \right] = [0,4 - 0,032 ; 0,4 + 0,032] = [0,368 ; 0,432].$$

La proportion réelle d'électeurs envisageant de voter pour le candidat L est donc comprise entre $\frac{368}{1000}$ et $\frac{432}{1000}$.

Interrogation orale :	En classe : 36 p. 283	Exercices : 38 p. 283
-----------------------	--------------------------	--------------------------

II – Simulation

1. Le tableur

Pour étudier ce qui se passe sur des échantillons encore plus grands (ou en plus grand nombre), nous allons utiliser la puissance de calcul d'un ordinateur afin de simuler le lancer des dés. En quelques fractions de seconde, nous pourrions obtenir un échantillon de plusieurs milliers de lancers !!!

a) La fonction ALEA()

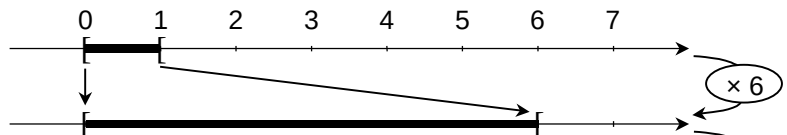
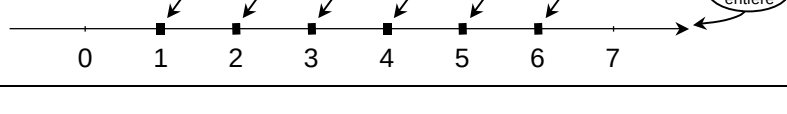
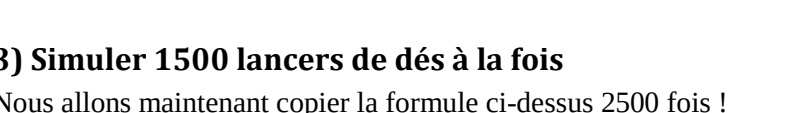
- Lancez le tableur puis taper dans la cellule A1 la formule : `=ALEA()`
- Appuyez sur la touche Entrée pour valider : Le tableur affiche alors un nombre aléatoire x tel que $0 < x < 1$.
- Appuyez plusieurs fois sur la touche : Le contenu de la cellule A1 est mis à jour de façon aléatoire.

b) Simuler le lancer d'un dé

Nous allons devoir maintenant modifier par étapes successives le contenu de cette cellule A1 de façon à obtenir à la fin un entier entre 1 et 6.

Pour chacune des étapes ci-dessous :

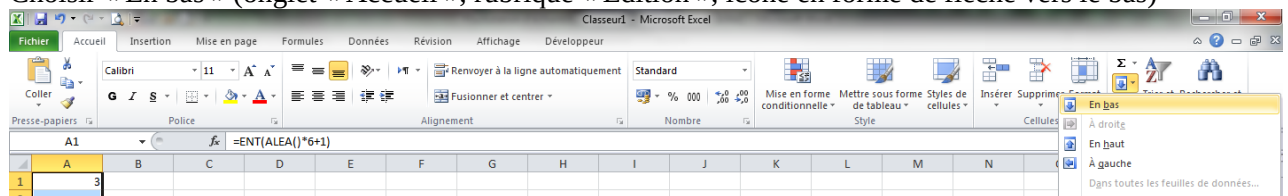
- Tapez la formule dans A1
- Complétez la 3^{ème} colonne du tableau ci-dessous
- Appuyez plusieurs fois sur la touche F9 pour vérifier le résultat

	Formule écrite en A1	Le résultat est alors
	=ALEA ()	compris entre :
	=6*ALEA ()	compris entre :
	=6*ALEA () + 1	compris entre :
	=ENT (6*ALEA () + 1)	soit :

3) Simuler 1500 lancers de dés à la fois

Nous allons maintenant copier la formule ci-dessus 2500 fois !

- Sélectionner en haut : , et taper A1:A2500 pour sélectionner d'un seul coup les 2500 premières cellules de la colonne A
- Choisir « En bas » (onglet « Accueil », rubrique « Édition », icône en forme de flèche vers le bas)



pour copier la formule qui est en A1 dans tout le reste de la sélection

- Appuyez plusieurs fois sur la touche  pour vérifier le résultat

4) Préparation du tableau dans lequel on va reporter les résultats :

Remplir les cellules C1 à E1 puis C2 à C7 comme ci-dessous

	A	B	C	D	E
1	5		face	effectif	fréquence
2	2		1		
3	6		2		
4	6		3		
5	5		4		
6	1		5		
7	4		6		

5) Déterminer les effectifs :

Pour compter le nombre de 1 qu'il y a dans la colonne A, nous allons utiliser la fonction appelée NB.SI(plage;critère). Cette fonction compte le nombre de cellules qui répondent à un critère donné à l'intérieur d'une plage de cellules donnée. Ici la plage de cellules correspond aux 1500 lancers (donc « A1:A2500 ») de dés et le critère est d'être égal à 1.

- En D2, entrez la formule : =NB.SI(A\$1:A\$2500;C2) (rem: le contenu de C2 est justement le nbre 1 !)
- Sélectionnez les cellules de D2 à D7 puis choisir "Edition/Remplir/En bas"
- Appuyez plusieurs fois sur la touche  pour vérifier que les effectifs oscillent bien autour de 2500/6 ≈ 417.

6) Déterminer les fréquences :

- En E2, entrez la formule : $=D2/2500$
- Sélectionnez les cellules de E2 à E7 puis choisir "Edition/Remplir/En bas"
- Appuyez plusieurs fois sur la touche  pour vérifier que les fréquences oscillent bien autour de $1/6 \approx 0,167$.

TP : 2 p. 273		
------------------	--	--

2. Bilan

Même si ce résultat n'a pas été mathématiquement montré, nous pouvons affirmer que notre hypothèse de départ est bel et bien juste.

En réunissant dans le I les résultats de tous les élèves de la classe, nous avons obtenu un échantillon dont la taille était proche des 1500 lancers de dés de la simulation. Les fréquences obtenues présentaient-elles un écart par rapport au 17% théorique analogue à celui que nous avons observé par simulation ?

L'avantage de la simulation avec le tableur a été double : non seulement nous avons pu travailler avec un gros échantillon pour réduire les phénomènes de fluctuation mais en plus, nous avons pu recommencer la simulation plusieurs fois pour vérifier que cette fluctuation était faible et avoir une idée de l'erreur commise.