

Fluctuations d'échantillonnage

ACTIVITÉ (Intervalle de fluctuation).

Dans la classe de Seconde 14, il y a 9 garçons et 28 filles, ce qui paraît disproportionné.

On peut se demander toutefois si, lorsqu'on choisit 37 élèves au hasard dans une population constituée d'une moitié de filles et d'une moitié de garçons, cette distribution est rare.

1. Quelle est la fréquence des filles dans la classe de Seconde 14 ?
2. Expliquer comment simuler le choix de 37 élèves au hasard dans une population d'une moitié de filles et d'une moitié de garçons à l'aide de la fonction *random* de la calculatrice.
3. Procéder à cette simulation en notant le nombre de filles et de garçons obtenus et calculer la fréquence des filles dans votre simulation (arrondie au centième).
4. Écrire cette fréquence au tableau et noter les résultats des simulations de la classe dans le tableau ci-dessous :

5. Déterminer, pour cette série statistique :
 - (a) les valeurs extrêmes, les premier et troisième quartiles, les premier et neuvième déciles, la médiane et la moyenne ;
 - (b) représenter le diagramme en boîte correspondant ;
 - (c) déterminer l'intervalle interquartile et interpréter le résultat ;
 - (d) déterminer l'intervalle interdécile et interpréter le résultat.
6. D'après ces résultats, peut-il arriver que le hasard produise une distribution comparable à celle de la Seconde 14 ? Si oui, est-ce fréquent ?
7. Les résultats obtenus par la classe peuvent très bien être eux aussi exceptionnels, aussi a-t-on besoin d'une règle plus objective. Nous utiliserons la propriété suivante, qu'on admettra :

PROPRIÉTÉ. Dans une population, la proportion d'un caractère est p . On produit un échantillon de taille n de cette population et on détermine la fréquence f du caractère dans cet échantillon. Si p est compris entre 0,2 et 0,8 et si n est supérieur ou égal à 25, alors, dans 95 % des cas au moins, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}} \right]$, que l'on appelle intervalle de fluctuation au seuil de 95 % (ou au risque de 5 %)

Dans notre exemple :

- Si c'est le cas, on peut avancer, au risque de 5 % de se tromper, que l'échantillon (la classe) est représentatif d'une population (le lycée) comportant une moitié de filles et d'une moitié de garçons.
- Si ce n'est pas le cas, on peut avancer, au risque de 5 % de se tromper, que l'échantillon (la classe) n'est représentatif pas d'une population (le lycée) comportant une moitié de filles et d'une moitié de garçons.

Les raisons peuvent être nombreuses :

- la classe a été volontairement constituée d'une majorité de filles ;
- la classe a été constituée en fonction des options et les filles choisissent plus souvent telle option ;
- on est dans les 5 % exceptionnels et donc on se trompe en s'avançant ainsi ;
- le modèle choisit n'est pas le bon (la population de référence, le lycée, ne comporte pas une moitié de filles et une moitié de garçons) ;
- etc.

- (a) Dans notre population de référence, quelle est la valeur de p qu'on a supposée ?
- (b) Quelle est la valeur de n ?
- (c) Déterminer alors l'intervalle de fluctuation correspondant à cette expérience.
- (d) Quel pourcentage des fréquences obtenues par la classe appartient à cet intervalle ?
- (e) La fréquence des filles de la Seconde 14 appartient-elle à cet intervalle ?

8. Et si notre supposition était fautive ?

À l'administration du lycée, on peut obtenir l'information suivante : « Au Lycée Dupuy de Lôme, pour l'année scolaire 2010–2011, il y a en Seconde 524 élèves, dont 329 filles et 195 garçons ».

- (a) Déterminer l'intervalle de fluctuation (toujours pour un échantillon de taille 37).
- (b) La fréquence des filles de la Seconde 14 appartient-elle à cet intervalle ? Qu'en conclure ?

Bilan et compléments

Lorsqu'on étudie un caractère d'une population, la connaissance de la population entière n'est en général pas envisageable (pour des raisons de temps, d'organisation ou de coût). On doit se contenter de la connaissance d'un échantillon de la population. Pour prendre des décisions fondées sur des théories mathématiques, il est indispensable que l'échantillon soit prélevé au hasard.

Définition On appelle *échantillon de taille n* une liste de n résultats obtenus par n répétitions indépendantes d'une même *expérience aléatoire*.

Définition Pour une population donnée, des échantillons aléatoires produits selon le même protocole peuvent avoir des compositions différentes. On dit qu'il y a *fluctuation d'échantillonnage*.

Propriété Dans une population, la proportion d'un caractère est p . On produit un échantillon de taille n de cette population et on détermine la fréquence f du caractère dans cet échantillon. Si p est compris entre 0,2 et 0,8 et si n est supérieur ou égal à 25, alors, dans 95 % des cas au moins, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$, que l'on appelle *intervalle de fluctuation au seuil de 95 % (ou au risque de 5 %)*

Exercices

EXERCICE 1.

Une urne contient 10 boules : **cinq** rouges, **trois** noires et **deux** blanches. On tire une boule et on note sa couleur et on la remet dans l'urne.

1. Avec la table de nombres aléatoires entiers de 0 à 9 donnée ci-dessous, simuler 25 tirages en expliquant votre méthode.

7	1	3	3	1	0	5	3	4	6	9	6	3	0	8	1	8	5	8	2
4	5	7	4	8	2	3	1	5	4	6	3	4	6	2	6	7	8	9	3
0	2	7	9	8	4	8	5	6	2	7	4	0	6	8	8	0	0	7	1
9	2	8	7	3	0	1	6	6	5	1	5	9	1	5	1	4	4	2	8
5	4	2	0	7	4	8	0	6	5	7	5	7	8	1	8	9	6	7	5

2. Calculer les fréquences obtenues pour chaque couleur.
3. Déterminer pour chacune des couleurs l'intervalle de fluctuation pour un échantillon de taille 25. Vos fréquences sont-elles dans ces intervalles? Conclure.

EXERCICE 2.

On lance deux dés et on note **la somme des deux nombres obtenus**.

1. Quels sont les résultats possibles?
2. Avec la table de nombres aléatoires entiers de 1 à 6 donnée ci-dessous, simuler 25 lancers en expliquant votre façon de procéder.

5	4	4	4	3	1	4	2	6	1	3	5	3	1	2	4	4	4	5	6
5	3	2	4	3	1	6	5	3	3	6	1	4	4	2	1	4	2	6	2
6	3	6	3	3	5	3	6	2	1	2	6	2	5	3	6	5	5	3	5
5	5	2	4	3	1	3	3	2	3	2	1	6	3	5	2	5	2	2	6
2	4	3	6	1	3	6	2	4	4	1	1	2	4	3	2	6	1	5	6

3. Donner la suite des 25 résultats obtenus.
4. Calculer les fréquences obtenues pour chaque résultat possible.
5. Norbert a procédé lui aussi à une simulation de 25 lancers, avec une autre table de nombres aléatoires, et il a obtenu les résultats suivants :

Face	2	3	4	5	6	7	8	9	10	11	12
Effectif	1	1	2	2	1	4	4	4	3	2	1

Comparer ces résultats à ceux de votre simulation.

6. Une simulation à l'ordinateur a donné les résultats suivants :

Face	2	3	4	5	6	7	8	9	10	11	12
Effectif	24	49	86	103	145	178	139	114	77	55	30

Comparer ces résultats à ceux de votre simulation.

7. Norbert pense que toutes les sommes devraient être présentes dans une même proportion.

- Si Norbert a raison, quel est l'intervalle de fluctuation pour un échantillon de taille 1000 ?
- Les fréquences des résultats obtenus à l'ordinateur sont-elles à 95 % dans cet intervalle ?
- Qu'en conclure ?

EXERCICE 3.

L'ensemble des faits évoqués ci-dessous est réel.

En novembre 1976 dans un comté du sud du Texas, RODRIGO PARTIDA était condamné à huit ans de prison pour cambriolage d'une résidence et tentative de viol.

Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que 79,1 % de la population de comté était d'origine mexicaine, sur les 870 personnes convoquées pour être jurés lors d'une certaine période de référence, il n'y eût que 339 personnes d'origine mexicaine.

- Déterminer l'intervalle de fluctuation correspondant à la proportion d'origine mexicaine pour un échantillon de taille 870.
- La fréquence des personnes d'origine mexicaine dans les personnes convoquées est-elle dans cet intervalle ?
- Qu'en conclure ?

EXERCICE 4.

À Dupuy de Lôme, pour la session 2009 du baccalauréat général, il y a eu 290 reçus pour 320 candidats se présentant à l'épreuve. Les fréquences des reçus en Série L, ES et S étaient, respectivement, 0,766, 0,896 et 0,963.

Déterminer si les différences de réussite entre les filières peuvent être dues aux fluctuations d'échantillonnage.

EXERCICE 5.

Dans le village chinois de Xicun en 2000, il est né 20 enfants dont 16 garçons. On suppose que la proportion de garçons et de filles est la même à la naissance dans toute l'espèce humaine.

Déterminer si la fréquence des naissances de garçons dans le village de Xicun en 2009 peut être due aux fluctuations d'échantillonnage.

EXERCICE 6.

Avez-vous vérifié que toutes les conditions étaient remplies pour appliquer les intervalles de fluctuation dans les deux exercices précédents ?

EXERCICE 7.

Au premier tour de l'élection présidentielle française de mai 2007, parmi les suffrages exprimés, les proportions, en pourcentage, pour les candidats ayant obtenu pour de 2 % des suffrages, étaient les suivantes :

Bayrou	Besancenot	De Villiers	Le Pen	Royal	Sarkozy
18,57	4,08	2,23	10,44	25,87	31,18

Cinq mois plus tôt, le 13 décembre 2006, l'institut de sondage BVA faisait paraître un sondage effectué sur un échantillon de 797 personnes dont voici les résultats, en pourcentage, concernant les candidats précédemment cités :

Bayrou	Besancenot	De Villiers	Le Pen	Royal	Sarkozy
7	4	2	10	34	32

- Pour quels candidats peut-on appliquer les intervalles de fluctuation parmi ceux présents au premier tour ?
- Pour ces candidats déterminer les intervalles de fluctuation pour un échantillon de taille 797.
- Les résultats du sondage donnent-ils des fréquences appartenant à ces intervalles ?
- Qu'en conclure ?

EXERCICE 8.

Les résultats seront donnés au centième.

Les données du tableau ci-dessous sont celles de l'année scolaire pour les Premières générales à Beaussier pour l'année scolaire 2004–2005 :

	1 ES	1 S	1 L	Total
Filles	76	92	50	218
Garçons	43	76	13	132
Total	119	168	63	350

- On s'intéresse d'abord à la proportion de garçons et de filles dans l'établissement.
 - Déterminer les proportions de garçons et de filles dans le lycée cette année là.
Peut-on utiliser les intervalles de fluctuations dans le cas des filles et des garçons ?
 - Déterminer les intervalles de fluctuations pour des échantillons de tailles respectives 119, 168 et 63.
 - Calculer les fréquences de garçons et de filles dans chacune des trois filières.
 - Dans quelles filières peut-on dire, au seuil de 95 %, que la fréquence des filles et des garçons peut être due aux fluctuations d'échantillonnage ?
- En vous inspirant de la question précédente, déterminer pour chaque sexe si l'on peut dire, au seuil de 95 %, que la fréquence des ES, S et L peut être due aux fluctuations d'échantillonnage.

EXERCICE 1 : Le jeu de pile ou face

Estimer - Décider

- Pour une expérience de PILE ou FACE donnée, dans le modèle défini par une probabilité de $p = 0,5$; écrire un algorithme puis un programme sur la calculatrice permettant de calculer la fréquence f des « FACE » lors de 100 lancers.
- A l'aide de l'intervalle de fluctuation au seuil de 95 %, que peut-on dire, avant d'effectuer une simulation, de la valeur de la fréquence f ?
 - Calculer 20 valeurs de f avec ce programme, parmi ces valeurs combien sont en dehors de l'intervalle de fluctuation ? Ce résultat est-il conforme à la définition de l'intervalle de fluctuation ?
- Adapter le programme pour N lancers, quelle est l'évolution des fréquences f lorsque N augmente jusqu'à 1000 ?

EXERCICE 2 : Sondage

A et B sont candidats à l'élection présidentielle. La population semble partagée entre les deux candidats. Un journal décide de réaliser un sondage sur un échantillon de 900 personnes.

Dire entre quelles valeurs on peut s'attendre à trouver le pourcentage de personnes favorables à A dans un tel échantillon. Le journal effectue le sondage, et trouve que 52% de l'échantillon est favorable à A.

Que devrait-il dire à ses lecteurs ?

EXERCICE 3 : Prise de décision ou preuve statistique

La parité dans le recrutement en entreprise, c'est quoi ?

Extrait du doc ressource probabilités et statistiques.

« La parité, cela signifie que l'identité sexuelle n'intervient pas au niveau du recrutement ; c'est à dire qu'au niveau du caractère « homme ou femme », les résultats observés pourraient être obtenus par choix au hasard des individus dans la population. »

Deux entreprises A et B recrutent dans un bassin d'emploi où il y a autant d'hommes que de femmes, avec la contrainte du respect de la parité.

- Dans l'entreprise A, il y a 100 employés dont 43 femmes.
- Dans l'entreprise B, il y a 2500 employés dont 1150 femmes.

Quelle entreprise respecte la parité ?

EXERCICE 4 : Prise de décision

Dans une usine, le responsable de la fabrication affirme que le nombre de produits fabriqués présentant un défaut est égal à 7%.

Sur la chaîne de fabrication, on a prélevé 38 produits, et on a relevé 6 produits présentant un défaut.

Qu'en pensez-vous ?