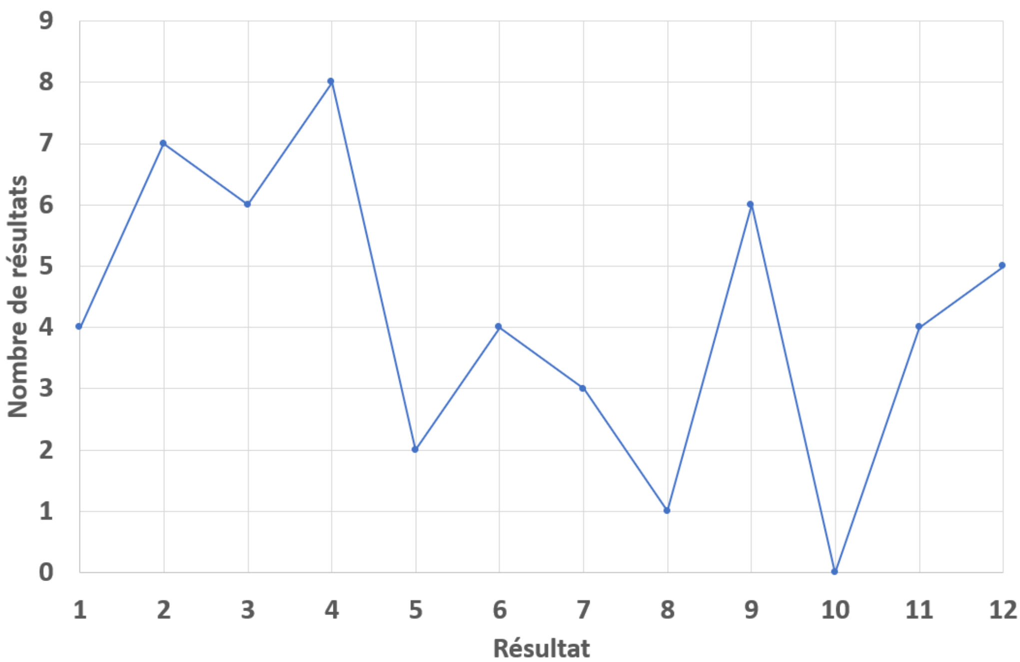
Fluctuation d'échantillonnage

1er Gestion - Administration

Exercice 1

Pour la fête du lycée, Keylane organise une loterie avec une roue permettant de tirer un numéro de 1 à 12.

1. Pour tester la roue, il la lance un certain nombre de fois et reporte les résultats dans le graphique ci-dessous.



a) Déterminer la taille de l'échantillon ainsi réalisé.
Il faut lire pour chaque résultat le nombre de résultats. On a alors $n = 4+7+6+8+2+4+3+1+6+0+4+5 = 50$

b) Calculer la fréquence de sortie du nombre 12.
Pour le nombre 12 il y a eu 5 tirages. Alors la fréquence est $f = 5/50 = 0.1$.

2. On simule 50 lancers de la roue et on obtient les résultats suivants :

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

Le 12 apparait 2 fois.

b) Calculer la fréquence d'apparition du nombre 12.
Il y a eu 50 lancers. La fréquence d'apparition du nombre 12 est alors $f = 2/50 = 0.04$.

3. On refait 8 fois la simulation pour 50 lancers et on obtient la fréquence du nombre 12 pour chacune des simulations :

Échantillon	1	2	3	4	5	6	7	8
Fréquence	0.1	0.1	0.08	0.04	0.1	0.06	0.08	0.04

Calculer la moyenne des fréquences obtenues.
On fait la moyenne de fréquence. $\overline{f} = (0.1+0.1+0.08+0.04+0.1+0.06+0.08+0.04)/8 \approx 0.08$

4. Si la roue est équilibrée, la probabilité de sortie du nombre 12 est $p = \dfrac{1}{12}$.
Comparer cette probabilité avec la moyenne des fréquences.
La probabilité est $p = 1/12 \approx 0.8$. On remarque que la probabilité est presque égale à la moyenne des fréquences obtenues. C'est normal, car pour un nombre important d'expérience la valeur de la fréquence d'apparition d'un évènement s'approche de la valeur de la probabilité de réalisation de ce même évènement.

Exercice 2

Le tableau suivant indique, pour chaque président élu aux États-Unis depuis 1936, la proportion p en faveur de l'élu le jour du vote et la proportion f prévue par le sondage Gallup avec l'élection (source : www.galup.com).

Année	Président élu	Vote p	Sondage f	Ecart
2008	Obama	0.55	0.53	0.02
2004	Bush	0.49	0.507	0.017
2000	Bush	0.48	0.479	0.001
1996	Clinton	0.52	0.492	0.028

Année	Président élu	Vote p	Sondage f	Ecart
1992	Clinton	0.49	0.433	0.057
1988	Bush	0.56	0.53	0.03
1984	Reagan	0.59	0.592	0.002
1980	Reagan	0.47	0.508	0.038
1976	Carter	0.48	0.501	0.021
1972	Nixon	0.62	0.618	0.002
1968	Nixon	0.43	0.435	0.005
1964	Johnson	0.64	0.613	0.027
1960	Kennedy	0.505	0.501	0.004
1956	Eisenhower	0.595	0.578	0.017
1952	Eisenhower	0.51	0.554	0.044
1948	Truman	0.445	0.495	0.05
1944	Roosevelt	0.515	0.538	0.023
1940	Roosevelt	0.52	0.55	0.03
1936	Roosevelt	0.557	0.625	0.068

1. Comment la dernière colonne est-elle calculée ?
La dernière colonne du tableau correspond à l'écart entre la valeur p et la valeur f.

2. Donner les deux années où l'écart entre le sondage et le résultat du vote a été le plus grand.
Les deux années où l'écart entre le sondage et le résultat du vote a été le plus grand sont les années 1936 et 1992.

3. Pour les quatre premiers sondages (1936 à 1948), Gallup n'utilisait pas de tirage aléatoire (méthode des "quotas"), alors que depuis 1952 cet institut utilise un échantillonnage aléatoire.
a) Calculer la moyenne des écarts de 1936 à 1948.
La moyenne des écarts est $\overline{f_1} = (0.068 + 0.03 + 0.023 + 0.05)/4 \approx 0.043$.

b) Vérifier que l'écart moyen de 1952 à 2008 est plus de deux fois inférieur au précédent.
En procédant de la même façon, on fait la moyenne de la fréquence et on trouve $\overline{f_2} \approx 0.021$.
On remarque bien que 0.021 est plus de deux fois inférieur à 0.043 .

c) Que peut-on conclure pour la méthode aléatoire ?
On peut alors en conclure que la méthode aléatoire est plus précise, car elle provoque un écart moyen moindre.

4. Si l'on compare les sondages Gallup à des échantillons aléatoire de taille 1000, la fluctuation d'échantillonnage prévoit, dans environ 95% des cas, un écart inférieur à $\frac{1}{\sqrt{1000}}$.

a) Donner une valeur approchée, à 0.001 près, de $\frac{1}{\sqrt{1000}}$.

On trouve $\frac{1}{\sqrt{1000}} \approx 0.032$.

b) Combien de sondages ont-ils un écart supérieur à la valeur précédente ?
Il y a 5 sondages supérieurs à la valeur précédente.

Exercice 3

Partie A : au milieu du XXe siècle

Vers les années 1950, face à l'augmentation de la consommation de tabac et du nombre de cancers du poumon, débutent des études statistiques sur le sujet. À l'hôpital Bellevue, en 1952, la fréquence des "grands fumeurs" (plus de 15 cigarettes par jour) parmi les malades est 44%. Parmi les 1357 malades soignés pour un cancer du poumon, 806 sont de grands fumeurs (source : Zeiseil et Kaye, *Prove it with Figure*).

1. Calculer, à 0.01 près, les bornes $p - \frac{1}{\sqrt{n}}$ et $p + \frac{1}{\sqrt{n}}$ de l'intervalle de fluctuation de 95% des fréquences des échantillons aléatoires de tailles $n = 1357$, lorsqu'on suppose que la proportion de grands fumeurs est $p = 0.44$.
Pour $n = 1357$ et $p = 0.44$ on trouve les bornes $[0.41; 0.47]$.

2. Calculer la fréquence f des grands fumeurs parmi les 1357 malades atteints de cancer du poumon.
Sur les 1357 malades il y a 806 qui sont de grands fumeurs. Alors on trouve $f = 806/1357 = 0.59$.

3. Est-il raisonnable de penser que la différence entre f et p est uniquement due au hasard ? **Expliquer**.
La valeur de la fréquence ne fait pas partie de l'intervalle [0.41 ;0.47]. La différence est significative et ce n'est donc pas du au hasard.

Partie B : à la fin du XXe siècle

Lors d'un sondage aléatoire effectué aux États-Unis en 1995, sur 737 fumeurs quotidiens, seuls 295 estimèrent courir un risque de cancer supérieur à celui des non-fumeurs de leur âge (source : *Journal of the American Medical Association*, 1999).

1. Calculer la fréquence f des fumeurs interrogés pensant prendre un risque.
La fréquence est $f = 295/737 = 0.40$.

2. Si l'on suppose que 50% des fumeurs aux États-Unis pensent prendre un risque, **calculer** l'intervalle de fluctuation $\left[0.5 - \frac{1}{\sqrt{737}}; 0.5 + \frac{1}{\sqrt{737}}\right]$ des fréquences de plus de 95% des échantillons de taille 737.
On trouve $[0.46; 0.54]$.

3. D'après le sondage effectué, peut-on estimer que moins de 50% des fumeurs aux États-Unis pensent prendre un risque ? **Expliquer**.
La fréquence ne fait pas partie de l'intervalle $[0.46; 0.54]$. Pour que la fréquence fasse partie de l'intervalle il faudrait une probabilité d'au moins $p \approx 0.43$ soit 43%. On peut alors estimer que moins de 50% des fumeurs aux États-Unis pensent prendre un risque.

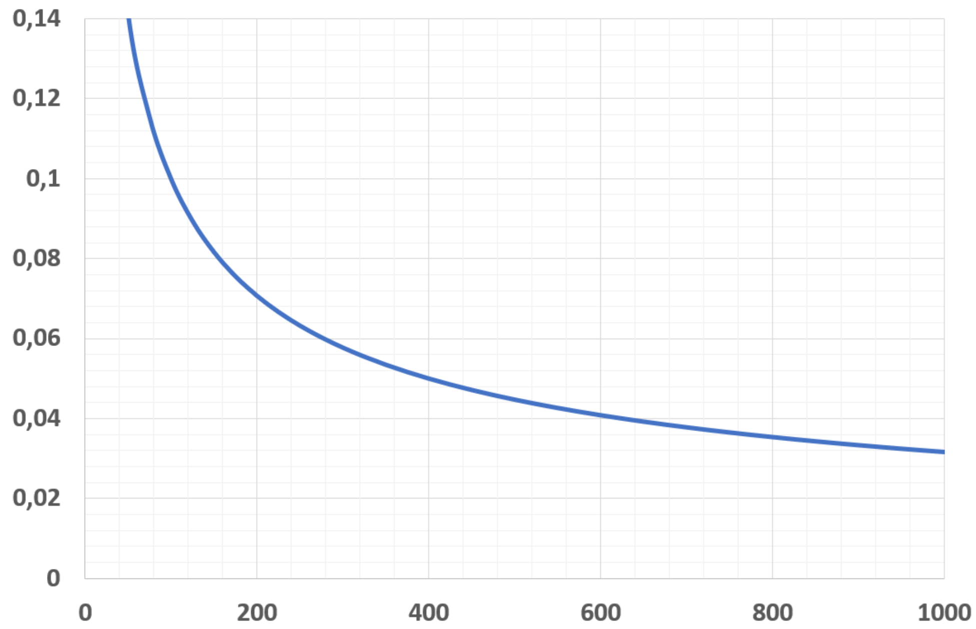
Exercice 4

Pour décider la construction d'un grand stade, la municipalité d'une ville importante voudrait sonder la population pour estimer si plus de 50% des électeurs sont favorables. Si cette proportion est $p = 0.5$, plus de 95% des sondages aléatoires de taille n fournissent une fréquence dans l'intervalle $I = \left[0.5 - \frac{1}{\sqrt{n}}; 0.5 + \frac{1}{\sqrt{n}}\right]$. La municipalité ne décidera de construire le stade que si la fréquence obtenue lors du sondage est supérieur à $0.5 + \frac{1}{\sqrt{n}}$.

1. Calculer les bornes de l'intervalle I pour $n = 100$.
On trouve $[0.4; 0.6]$.

2. Sur 100 électeurs interrogés au hasard, 54 sont favorables à la construction du stade. La municipalité décide-t-elle de construire le stade ?
La fréquence correspondant aux électeurs favorables à la construction du stade est $f = 54/100 = 0.54$. Cette valeur est inférieur à $0.5 + \frac{1}{\sqrt{100}} = 0.6$. Alors la mairie décide de ne pas construire le stade.

3. La courbe ci-dessous a pour équation $f(x) = \frac{1}{\sqrt{x}}$



Lire approximativement la plus petite valeur de x pour laquelle on a $\frac{1}{\sqrt{x}} \leq 0.04$.
La plus petit valeur, lisible sur le graphique, de x pour laquelle on a $\frac{1}{\sqrt{x}} \leq 0.04$ est $x \approx 600$.

4. Sur 650 électeurs interrogés au hasard, 54% sont favorables à la construction du stade. La municipalité décide-t-elle de construire le stade ?
On a cette fois-ci $n = 650$ car il y a 650 électeurs. Les bornes deviennent $[0.461; 0.539]$. La fréquence est de 0.54. Cette valeur est supérieur à $0.5 + \frac{1}{\sqrt{650}} = 0.539$. Alors dans ce cas là la municipalité décide de construire le stade.