

Echantillonnage et statistiques

Exercice 1 :

- 1) Dans une population de truites de rivière, on compte autant de truites mâles que de truites femelles. Certaines pollutions peuvent entraîner une augmentation de truites femelles. Sur un prélèvement de 400 truites d'une rivière, on a relevé 225 femelles. Ce résultat doit-il conduire à suspecter une pollution ?



- 2) De la même façon, certaines espèces sont plus sensibles que d'autres à la pollution. C'est le cas de l'anguille européenne. Habituellement cette espèce représente 8% des poissons d'une rivière, et sur un prélèvement de 900 poissons on en a trouvé 39. Pouvez-vous conclure à une pollution ?

Exercice 2 : La bonne proportion de noisettes.

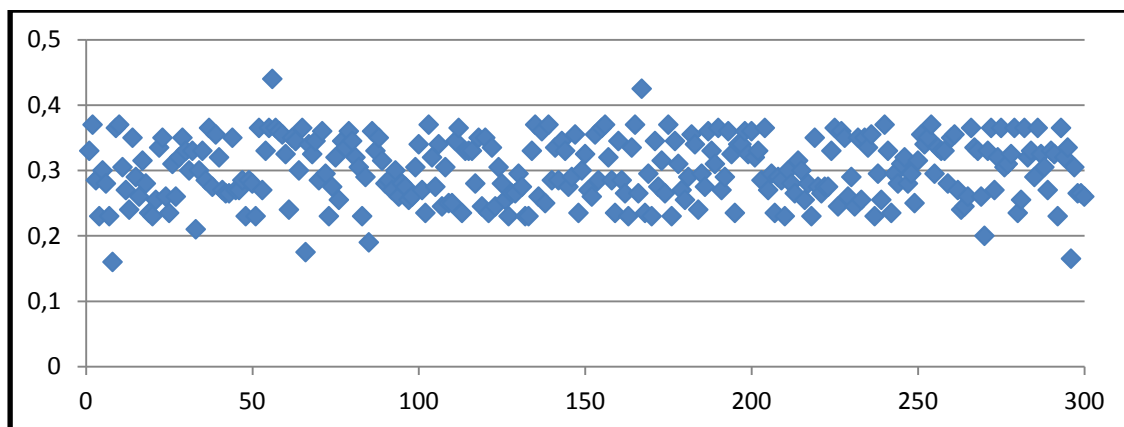


Un chocolatier artisanal fabrique des tablettes de chocolat au lait avec des morceaux de noisettes.

Il est prévu que les morceaux de noisettes représentent 30 % de la masse de chaque tablette.

Pour vérifier que la machine est réglée correctement, on prélève chaque jour, une tablette de 200g considérée comme un échantillon aléatoire de taille 200.

- 1) Calculer les bornes de l'intervalle de fluctuation au seuil de 95 %.
- 2) En 2015, on a relevé les masses de noisettes, puis on a calculé les fréquences que l'on a reportées sur la carte de contrôle ci-dessous.



Tracer **sur cette carte**, avec la précision permise par le graphique, les droites horizontales correspondant à la fréquence théorique et aux bornes de l'intervalle de fluctuation.

- 3) Face au succès de ses produits, le chocolatier ouvre une succursale dans une autre ville. Lors de la visite inopinée d'un ingénieur contrôle-qualité, un relevé sur une plaque de chocolat de 200g, prise au hasard, est effectué. Le contrôleur trouve une masse de 44 g de noisettes.
 - a) Calculer la fréquence correspondant à cette masse et la reporter sur la carte de contrôle ci-dessus.
 - b) Quelle conclusion pourra en tirer le contrôleur sur le réglage de la machine de cette succursale?

- 4) Donner un encadrement de la masse nécessaire de noisettes dans une plaque de 200g de chocolat pour que la proportion de 30 % soit respectée.
La réponse sera donnée au gramme près.

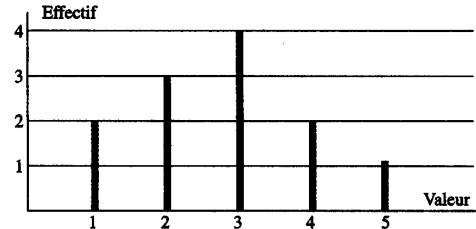
Exercice 3 : VRAI/ FAUX

| | | | | | |
|------------------|----|-----|------|----|-------|
| Valeur | 5 | 9 | 12 | 16 | Total |
| Effectif | 80 | 260 | | 50 | 500 |
| Fréquence | | | 22 % | | |

1/ On considère la série statistique :

- La fréquence de la valeur 5 est 16 %. vrai faux
- La fréquence de l'événement : « la valeur est supérieure à 10 » est 27 %. vrai faux

2/ On considère le diagramme en bâtons :



Pour la série représentée,

- l'étendue est 4 vrai faux
- la moyenne est égale à 3 vrai faux

Exercice 4 :

Le tableau ci-dessous donne la capacité des disques durs, en Go, des ordinateurs d'un magasin.

| | | | | | | | | |
|----------|----|-----|-----|-----|-----|-----|------|------|
| Go | 80 | 160 | 250 | 320 | 500 | 800 | 1000 | 1150 |
| effectif | 2 | 9 | 11 | 7 | 5 | 2 | 4 | 3 |

- 1) Calculer la moyenne de cette série.
- 2) Calculer la médiane de cette série.
- 3) Calculer le troisième quartile de cette série.
- 4) Pour chacune des affirmations, dire si elle est vraie ou fausse, en justifiant votre réponse.
 - a) Environ la moitié des disques durs a une capacité de moins 422 Go.
 - b) Environ un quart des disques durs a une capacité de plus de 500 Go.

Exercice 5 :

Une machine fabrique des fers cylindriques pour le béton armé. On contrôle le fonctionnement de la machine en prélevant un échantillon de 100 pièces au hasard.

La mesure de leurs diamètres, à 0,1 mm près, a donné :

| | | | | | | | | | | | | | | | |
|------------------|------|------|------|------|------|------|------|----|------|------|------|------|------|------|-----|
| Diamètres | 24,3 | 24,4 | 24,5 | 24,6 | 24,7 | 24,8 | 24,9 | 25 | 25,1 | 25,2 | 25,3 | 25,4 | 25,5 | 25,6 | 25, |
| Effectifs | 2 | 4 | 8 | 7 | 13 | 16 | 11 | 8 | 6 | 9 | 5 | 4 | 4 | 2 | 1 |

1. **En utilisant une calculatrice**, donnez la moyenne \bar{x} , la médiane, le premier quartile Q_1 et le troisième quartile Q_3 de cet échantillon.
2. On estime que la machine fonctionne correctement si :
 - $Q_3 - Q_1$ est inférieur à 2% de la moyenne
 - L'écart entre la moyenne et la médiane est inférieur à 0,1
 - Au moins 90% des diamètres sont dans l'intervalle $[\bar{x} - 0,5 ; \bar{x} + 0,5]$

Les critères donnés ci-dessus sont-ils vérifiés ? Justifiez votre réponse. Peut-on alors considérer que cette machine fonctionne correctement ?

Eléments de correction

Exercice 1 :

- 1) Dans une population de truites de rivière, on compte autant de truites mâles que de truites femelles. Certaines pollutions peuvent entraîner une augmentation de truites femelles. Sur un prélèvement de 400 truites d'une rivière, on a relevé 225 femelles. Ce résultat doit-il conduire à suspecter une pollution ?

$$p = 0,5 \text{ d'où } 0,2 \leq p \leq 0,8$$

$$n = 400 \text{ et donc } n \geq 25$$

On peut donc déterminer l'intervalle de fluctuation au seuil de 95% :

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

$$\text{Or, } p - \frac{1}{\sqrt{n}} = 0,5 - \frac{1}{\sqrt{400}} = 0,45$$

$$p + \frac{1}{\sqrt{n}} = 0,5 + \frac{1}{\sqrt{400}} = 0,55 \quad I = [0,45; 0,55]$$

La fréquence observée des truites femelles est: $f = \frac{225}{400} = 0,5625$

Ainsi f n'appartient pas à l'intervalle de fluctuation I , on peut donc dire, avec risque de 5% que l'échantillon n'est pas représentatif, ce qui signifie que l'on peut penser à une pollution.

- 2) Ici $p = 0,5$ donc p n'est pas compris entre 0,2 et 0,8. Le cours de seconde ne permet donc pas de conclure

Exercice 2 : La bonne proportion de noisettes.



Un chocolatier artisanal fabrique des tablettes de chocolat au lait avec des morceaux de noisettes.

Il est prévu que les morceaux de noisettes représentent 30 % de la masse de chaque tablette.

Pour vérifier que la machine est réglée correctement, on prélève chaque jour, une tablette de 200g considérée comme un échantillon aléatoire de taille 200.

- 1) Calculer les bornes de l'intervalle de fluctuation au seuil de 95 %.

$$p = 0,3 \text{ d'où } 0,2 \leq p \leq 0,8$$

$$n = 200 \text{ et donc } n \geq 25$$

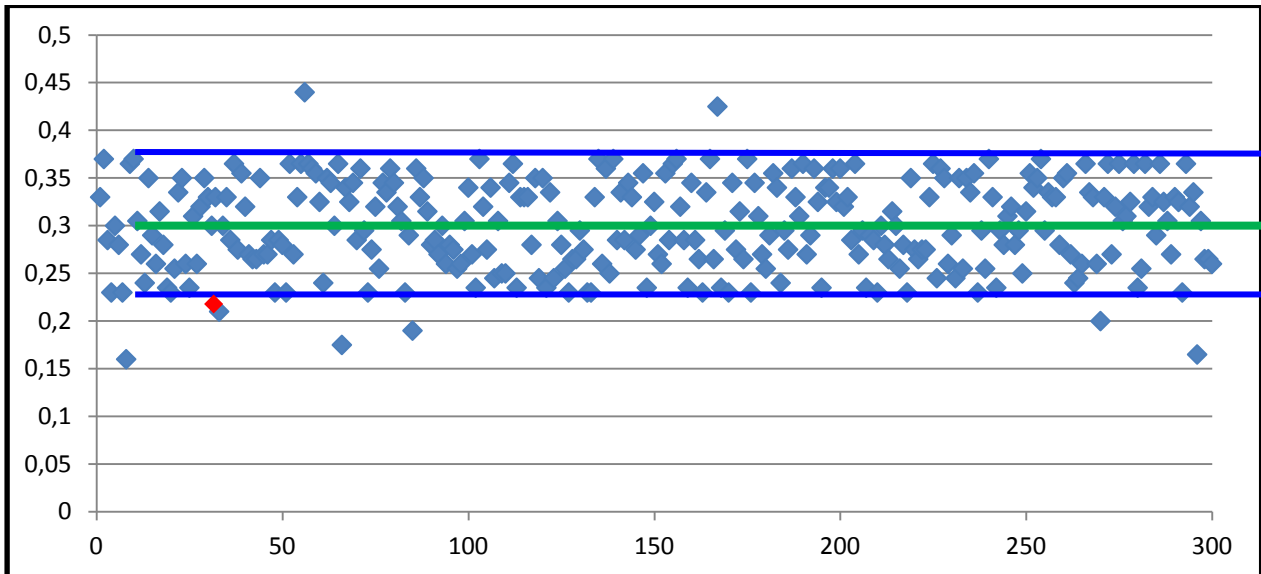
On peut donc déterminer l'intervalle de fluctuation au seuil de 95% :

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

$$\text{Or, } p - \frac{1}{\sqrt{n}} = 0,3 - \frac{1}{\sqrt{200}} \approx 0,229$$

$$p + \frac{1}{\sqrt{n}} = 0,3 + \frac{1}{\sqrt{200}} \approx 0,371 \quad I = [0,229; 0,371]$$

- 2) En 2015, on a relevé les masses de noisettes, puis on a calculé les fréquences que l'on a reportées sur la carte de contrôle ci-dessous.



Tracer **sur cette carte**, avec la précision permise par le graphique, les droites horizontales correspondant à la fréquence théorique et aux bornes de l'intervalle de fluctuation.

- 3) Face au succès de ses produits, le chocolatier ouvre une succursale dans une autre ville. Lors de la visite inopinée d'un ingénieur contrôle-qualité, un relevé sur une plaque de chocolat de 200g, prise au hasard, est effectué.

Le contrôleur trouve une masse de 44 g de noisettes.

- a) Calculer la fréquence correspondant à cette masse et la reporter sur la carte de contrôle ci-dessus.

$$f = \frac{44}{200} = 0,22$$

- b) Quelle conclusion pourra en tirer le contrôleur sur le réglage de la machine de cette succursale?

$f \notin I$. Cette fréquence n'appartenant pas à l'intervalle de fluctuation au seuil de 95%.

On peut donc dire, avec un risque d'erreur de 5%, que le hasard seul ne peut pas expliquer l'écart entre p et f , et donc qu'il faudrait revoir le réglage de la machine dans cette succursale

- 4) Donner un encadrement de la masse nécessaire de noisettes dans une plaque de 200g de chocolat pour que la proportion de 30 % soit respectée.

La réponse sera donnée au gramme près.

Pour que la proportion de 30 % soit respectée, il suffit que la fréquence f appartienne à l'intervalle de fluctuation I avec $I = [0,229; 0,371]$

Or cette fréquence est égale à $\frac{m}{200}$ où m est la masse (en gramme) de noisettes dans une plaque de 200g de chocolat

On doit donc résoudre : $0,229 \leq \frac{m}{200} \leq 0,371$

$$0,229 \times 200 \leq m \leq 0,371 \times 200$$

$$45,8 \leq m \leq 74,2$$

Donc au gramme près : m doit être compris entre 46g et 74 g

Exercice 3 : VRAI/ FAUX

| Valeur | 5 | 9 | 12 | 16 | Total |
|-----------|----|-----|------|----|-------|
| Effectif | 80 | 260 | | 50 | 500 |
| Fréquence | | | 22 % | | |

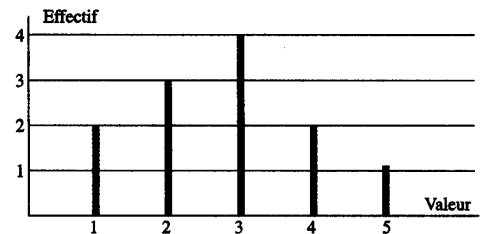
1/ On considère la série statistique :

- La fréquence de la valeur 5 est 16 %. **VRAI** : $\frac{80}{500} = 0,16 = \frac{16}{100}$
- La fréquence de l'événement : « la valeur est supérieure à 10 » est 27 %. **FAUX**
La fréquence de la valeur 16 est $\frac{50}{500} = 0,1$, donc il y a 10% de valeurs 16. Les valeurs supérieures à 10 représentent donc une fréquence de 32% (22+ 10)

2/ On considère le diagramme en bâtons :

Pour la série représentée,

- l'étendue est 4 **VRAI** $5-1 = 4$
- la moyenne est égale à 3 **FAUX**



$$\frac{2 \times 1 + 3 \times 2 + 4 \times 3 + 2 \times 4 + 1 \times 5}{12} = \frac{33}{12} = 2,75$$

Exercice 4 :

1) Déterminer la moyenne de cette série.

$$\frac{80 \times 2 + 160 \times 9 + 250 \times 11 + 320 \times 7 + 500 \times 5 + 800 \times 2 + 1000 \times 4 + 1150 \times 3}{43}$$

$$= \frac{18140}{43} \approx 421,86 \text{ (ou } 421,9 \text{ ou } 422)$$

2) Déterminer la médiane de cette série.

$$\frac{43}{2} = 21,5 \text{ donc la médiane est la } 22^{\text{ème}} \text{ valeur c'est-à-dire } 250.$$

3) Déterminer le troisième quartile de cette série.

$$43 \times \frac{3}{4} = 32,25 \text{ donc le troisième quartile est la } 33^{\text{ème}} \text{ valeur, c'est-à-dire } 500.$$

4) a) **Faux : pour répondre à cette question il faut utiliser la médiane et pas la moyenne.**

La moitié des disques durs a une capacité de moins 250 Go

b) Vrai : $Q_3 = 500$ donc les trois quart des disques durs a une capacité de moins de 500 Go.

Exercice 5 :

1) A la calculatrice on trouve :

$$\bar{x} = 24,906 ; \text{ Me } 24,85 ; Q_1 = 24,7 \text{ et } Q_3 = 25,1$$

2) Le premier critère est vérifié. En effet :

$$Q_3 - Q_1 = 25,1 - 24,7 = 0,4$$

ce qui est bien inférieur à 2% de la moyenne car $\frac{2}{100} \times 24,906 = 0,49812$

Le deuxième critère est vérifié. En effet :

L'écart entre la moyenne et la médiane est :

$$24,906 - 24,85 = 0,056 \text{ ce qui est inférieur à } 0,1$$

Le troisième critère n'est pas vérifié. En effet :

$$[\bar{x} - 0,5 ; \bar{x} + 0,5] = [24,906 - 0,5 ; 24,906 + 0,5] = [24,406 ; 25,406]$$

Cet intervalle contient donc les valeurs du tableau de 24,5 à 25,4.

| | | | | | | | | | | | | | | | |
|------------------|------|------|------|------|------|------|------|----|------|------|------|------|------|------|------|
| Diamètres | 24,3 | 24,4 | 24,5 | 24,6 | 24,7 | 24,8 | 24,9 | 25 | 25,1 | 25,2 | 25,3 | 25,4 | 25,5 | 25,6 | 25,7 |
| Effectifs | 2 | 4 | 8 | 7 | 13 | 16 | 11 | 8 | 6 | 9 | 5 | 4 | 4 | 2 | 1 |

Cet intervalle contient donc 87 valeurs sur un total de 99 valeurs, soit en pourcentage :

$$\frac{87}{99} \times 100 \approx 87,9 \text{ ce qui est inférieur à } 90.$$

On peut donc considérer que cette machine ne fonctionne pas correctement.

