

STATISTIQUES DESCRIPTIVES A UNE ET A DEUX DIMENSIONS

B. BIHIN, F. DREZE, V. COQUETTE, A.-C. WAUTHY

Mesurez la taille et la masse d'un échantillon de dix cailloux prélevés aléatoirement. A partir de vos mesures, déterminez les paramètres de tendance centrale et de dispersion de votre échantillon. Répondez ensuite aux questions qui vous sont posées.

Vos mesures (groupe n°.....)

Masse des 10 cailloux

Caillou	Masse (g)
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

Mesures de tendance centrale

Moyenne:

Mode:

Médiane:

Mesures de dispersion

Amplitude:

SCE:

Variance:

Ecart-type:

Coefficient de variation:

Taille des 10 cailloux

Caillou	Taille (cm)			
	T1	T2	T3	moy
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				

Mesures de tendance centrale

Moyenne:

Mode:

Médiane:

Mesures de dispersion

Amplitude:

SCE:

Variance:

Ecart-type:

Coefficient de variation:

Statistiques descriptives à une dimension

Décrire une série de données

Calculer et interpréter les paramètres de tendance centrale

- Q1. Certains échantillons de dix cailloux ne présentent pas de mode. Dans ce cas, comment pouvez-vous faire pour obtenir quand même une "information modale" ?
- Q2. Si Excel ne l'avait pas fait pour vous, comment auriez-vous déterminé la médiane de votre échantillon de dix cailloux ?
- Q2.1. Quelle est l'étape préalable indispensable pour calculer la médiane d'un échantillon ?
- Q2.2. Comment déterminer la médiane d'un nombre impair de données. Par exemple, quelle est la médiane de la variable "taille" (cm) de cet échantillon de 5 cailloux: 8,9 – 6,9 – 6,4 – 7,6 – 7,0 ?
- Q2.3. Comment déterminer la médiane d'un nombre pair de données. Par exemple, quelle est la médiane de la variable "taille" (cm) de cet échantillon de 6 cailloux: 8,9 – 6,9 – 6,4 – 7,6 – 7,0 – 9,1 ?

Calculer et interpréter les paramètres de dispersion

- Q3. D'après vous, quelle variable varie le plus: la taille ou la masse ? Pourquoi ?
- Q4. Observez les variances calculées pour des échantillons de dix cailloux. Comment expliquez-vous la différence d'ordre de grandeur de cette variance entre la masse et la taille ?
- Q4.1. Quelle est l'unité de mesure de la variance de la variable "taille" ?
- Q4.2. Quelle est l'unité de mesure de la variance de la variable "masse" ?
- Q4.3. Comment l'expression de la masse en kilogrammes (plutôt qu'en grammes) influencerait-elle l'ordre de grandeur de la variance ?
- Q4.4. Comment l'expression de la taille en millimètres (plutôt qu'en centimètres) influencerait-elle l'ordre de grandeur de la variance ?
- Q5. Observez les écarts-types calculés pour des échantillons de dix cailloux. Comment expliquez-vous la différence d'ordre de grandeur de cet écart-type entre la masse et la taille ?
- Q5.1. Quelle est l'unité de mesure de l'écart-type de la variable "taille" ?
- Q5.2. Quelle est l'unité de mesure de l'écart-type de la variable "masse" ?
- Q6. Quel est l'intérêt du coefficient de variation par rapport à l'écart-type ?
- Q6.1. Quelle est l'unité de mesure du coefficient de variation ?
- Q6.2. Quelle est la valeur minimale possible pour le coefficient de variation ? Que signifie-t-elle ?

Etude de la symétrie ou de l'asymétrie

Q7. Parmi les histogrammes construits en Duplo, choisissez-en un qui représente une distribution symétrique. Justifiez votre choix.

Q8. Parmi les histogrammes construits en Duplo, choisissez-en un qui représente une distribution asymétrique. Justifiez votre choix.

Q9. Observez les paramètres de mesure de la tendance centrale de votre échantillon de dix cailloux.

Q9.1. Quelle est la moyenne de la taille de vos dix cailloux ? Et la médiane ? Et le mode ?

Q9.2. D'après les valeurs de tendance centrale de la variable "taille" pour votre échantillon de dix cailloux, la distribution de cette variable vous semble-t-elle plutôt symétrique ou asymétrique ?

Q9.3. Quelle est la moyenne de la masse de vos dix cailloux ? Et la médiane ? Et le mode ?

Q9.4. D'après les valeurs de tendance centrale de la variable "masse" pour votre échantillon de dix cailloux, la distribution de cette variable vous semble-t-elle plutôt symétrique ou asymétrique ?

Q9.5. Dans le cas d'une distribution asymétrique, dans quel ordre se trouveront les paramètres de tendance centrale ?

Q10. Observez les paramètres de tendance centrale calculés sur l'ensemble des observations.

Q10.1. D'après ces paramètres, la variable "taille" suit-elle une distribution plutôt symétrique ou asymétrique ?

Q10.2. La conclusion à laquelle vous arrivez en observant les paramètres de tendance centrale (symétrie ou asymétrie) correspond-t-elle à ce que vous observez sur l'histogramme construit en Duplo ?

Q10.3. D'après ces paramètres, la variable "masse" suit-elle une distribution plutôt symétrique ou asymétrique ?

Q10.4. La conclusion à laquelle vous arrivez en observant les paramètres de tendance centrale (symétrie ou asymétrie) correspond-t-elle à ce que vous observez sur l'histogramme construit en Duplo ?

Construire un histogramme

Q11. Dans les histogrammes que vous avez construits, que représente la hauteur d'une colonne ?

Q12. Que faire lorsqu'une observation est exactement égale à la borne (inférieure ou supérieure) de l'intervalle de classe ? Par exemple, dans quelle classe allez-vous placer une observation de 8,0 cm si vous avez choisi de construire un histogramme de la variable "taille" dont l'intervalle de classes est de 2 cm ?

Q13. Observez les histogrammes représentant les distributions de la taille de dix cailloux.

Q13.1. Quels sont les intervalles de classes des histogrammes (un intervalle de classe par histogramme observé) ?

Q13.2. Pour chaque histogramme observé, combien dénombrez-vous de classes non-vides ?

Q13.3. Si vous deviez présenter les résultats de vos mesures de la taille de 10 cailloux dans un travail, quel histogramme (parmi ceux observés) choisiriez-vous ? Justifiez votre choix.

Q14. Observez les histogrammes représentant les distributions de la masse de trente cailloux.

Q14.1. Quels sont les intervalles de classes des histogrammes (un intervalle de classe par histogramme observé) ?

Q14.2. Pour chaque histogramme observé, combien dénombrez-vous de classes non-vides ?

Q14.3. Si vous deviez présenter les résultats des mesures de la masse de 30 cailloux dans un travail, quel histogramme (parmi ceux observés) choisiriez-vous ? Justifiez votre choix.

Comparer plusieurs séries de données

Q15. Observez l'histogramme représentant la distribution de la variable "taille" pour un échantillon de dix cailloux avec un intervalle de classe de 2 cm. Observez également l'histogramme représentant la distribution de la variable "taille" pour un échantillon de trente cailloux avec un intervalle de classe de 2 cm.

Q15.1. Le nombre d'observations représentées est-il le même sur les deux histogrammes ?

Q15.2. L'intervalle de classe est-il le même sur les deux histogrammes ?

Q15.3. L'allure générale des deux histogrammes est-elle la même ?

Q15.4. Quelle est l'importance d'une observation parmi l'ensemble des mesures ?

Q15.5. Comment pourriez-vous comparer ces deux histogrammes ?

Q16. Observez l'histogramme représentant la distribution de la variable "masse" pour un échantillon de dix cailloux avec un intervalle de classe de 50 g. Observez également l'histogramme représentant la distribution de la variable "masse" pour un échantillon de dix cailloux avec un intervalle de classe de 200 g.

Q16.1. Le nombre d'observations représentées est-il le même sur les deux histogrammes ?

Q16.2. L'intervalle de classe est-il le même sur les deux histogrammes ?

Q16.3. L'allure générale des deux histogrammes est-elle la même ?

Q16.4. Dans chacun des histogrammes, combien d'observation se trouvent entre 0 et 200 g ?

Q16.5. Comment pourriez-vous comparer ces deux histogrammes ?

Q17. Observez l'histogramme représentant la distribution de la variable "masse" pour un échantillon de dix cailloux avec un intervalle de classe de 50 g. Observez également l'histogramme représentant la distribution de la variable "masse" pour un échantillon de trente cailloux avec un intervalle de classe de 200 g.

Q17.1. Le nombre d'observations représentées est-il le même sur les deux histogrammes ?

Q17.2. L'intervalle de classe est-il le même sur les deux histogrammes ?

Q17.3. L'allure générale des deux histogrammes est-elle la même ?

Q17.4. Comment pourriez-vous comparer ces deux histogrammes ?

Statistiques descriptives à deux dimensions

Décrire la relation entre deux variables

Construire un graphique représentant la relation entre deux variables

Q18. Observez les graphiques que vous avez construits en Duplo.

Q18.1. A quel endroit disposez-vous d'une représentation de la relation entre la taille et la masse des cailloux ?

Q18.2. Comment s'appelle ce type de graphique ?

Q18.3. Que représente une colonne de Duplo dans ce type de graphique ?

Q18.4. Ce type de graphique vous semble-t-il facile à représenter sur papier ? Pourquoi ?

Q19. Observez le diagramme de dispersion représenté au tableau.

Q19.1. Que représente l'axe des abscisses ?

Q19.2. Que représente l'axe des ordonnées ?

Q19.3. Que représente chaque point du graphique ?

Q19.4. Que représente chaque triangle du graphique ?

Calculer et interpréter la dispersion

Q20. A partir du diagramme de dispersion représentant les observations de taille et de masse de tous les cailloux, quantifier la dispersion des points autour de la moyenne générale. Choisissez un point du graphique.

Q20.1. Quel est l'écart entre ce point et la moyenne de toutes les tailles ?

Q20.2. Quel est l'écart entre ce point et la moyenne de toutes les masses ?

Q20.3. A partir des écarts selon la taille et selon la masse, comment pouvez-vous obtenir une information de surface ? Souvenez-vous, dans les statistiques à une dimension, chaque écart était mis au carré.

Q21. Lorsqu'on étudie la relation entre la taille et la masse des cailloux, la SPE sera-t-elle positive ou négative ? Justifiez.

Q21.1. Que signifie, au niveau de l'allure générale d'un diagramme de dispersion, une SPE positive ?

Q21.2. Que signifie, au niveau de l'allure générale d'un diagramme de dispersion, une SPE négative ?

Q22. La SPE possède un inconvénient majeur. Lequel ?

Q22.1. Comment l'augmentation du nombre de mesures va influencer la SPE ?

Q22.2. Comment résoudre ce problème ?

Variabilité d'une distribution d'échantillonnage

Q23. Observez la répartition des moyennes dans le diagramme de dispersion dessiné au tableau.

Q23.1. L'allure (forme, orientation) du nuage des moyennes est-elle semblable à celle du nuage de mesures individuelles ?

Q23.2. La dispersion des points du nuage des moyennes est-elle semblable et celle du nuage des mesures individuelles ?