
La qualité de la rédaction sera prise en compte dans la note. Les réponses devront être justifiées.

Documents interdits, calculatrice UPPA autorisée.

On donnera les résultats à 10^{-2} près.

Le barème est donné à titre indicatif.

Exercice n°1 Les données ci-dessous représentent les durées (exprimées en minutes) de 400 consultations d'un site web :

Durée de consultation	$]0; 8]$	$]8; 16]$	$]16; 24]$	$]24; t]$	$]t; 44]$	$]44; 60]$	$]60; 84]$
Effectif	24	a	b	68	56	44	12

Certaines données sont illisibles et ont été remplacées par t , a et b .

- Déterminer a et b en sachant que le quatrième décile est égal à 19 minutes.
- Déterminer t sachant que la durée moyenne de consultation est égale à 26 minutes.
- Tracer l'histogramme de la distribution.
- Tracer la fonction des fréquences cumulées. Situer sur ce graphique la médiane et la calculer. Calculer les autres quartiles.
- Représenter les différentes caractéristiques numériques à l'aide d'une boîte à moustaches (box-plot).

Exercice n°2 On a mesuré la teneur minérale des vertèbres lombaires de quarante femmes. Les mesures (en g/dm^2) sont données dans la liste ci-dessous : (on notera x cette série statistique)

60 61 63 64 66 67 69 70
71 71 72 74 75 75 76 76
77 77 78 79 79 80 81 81
81 82 82 83 84 84 85 86
87 88 88 89 92 94 95 97

- Préciser la population et la variable étudiées. Quel est le type de cette variable ?
- On décide de regrouper les données de manière discrète.
 - Combien observe-t-on de modalités ?
 - Représenter les données regroupées par un graphique adapté.
 - Calculer la moyenne (on précisera la formule qui aura été utilisée).
 - Calculer la variance (on précisera la formule qui aura été utilisée), puis l'écart-type.
 - Calculer la médiane (on précisera la formule qui aura été utilisée).
- Vu le nombre de modalités, on décide de regrouper les données selon les classes suivantes : $[60; 69[$, $[69; 76[$, $[76; 80[$, $[80; 84[$, $[84; 91[$ et $[91; 98[$.

- Donner la série des données regroupées.
 - Représenter les données regroupées par un graphique adapté.
 - Déterminer la classe modale.
 - Calculer la moyenne (on précisera la formule qui aura été utilisée).
 - Calculer la variance (on précisera la formule qui aura été utilisée) puis l'écart-type.
 - Calculer la médiane (on précisera la formule qui aura été utilisée).
 - Commenter les différences constatées entre les différentes caractéristiques étudiées ci-dessus.
4. On observe dix individus supplémentaires : (on notera y cette série statistique)

61 71 71 77 78 79 80 81 86 89

Sur cet échantillon, on obtient une moyenne égale à 77,3 et une médiane égale à 78,5.

- Peut-on calculer rapidement la moyenne sur les cinquante individus en utilisant les réponses de la question 2 ? Si oui, calculer cette nouvelle moyenne.
- Peut-on calculer rapidement la médiane sur les cinquante individus en utilisant les réponses de la question 2 ? Si oui, calculer cette nouvelle médiane.

Exercice n°3 On s'intéresse à la production annuelle de cônes de mélèzes situés à 1200 mètres d'altitude dans la vallée d'Ayes (Briançonnais, Alpes du Sud). La production de cônes a été estimée grossièrement en comptant le nombre de cônes sur des branches longues d'un mètre de plus de cent arbres choisis au hasard. L'intensité de la production a été mesurée sur une échelle graduelle allant de 1 (pas de cône) à 6 (récolte importante). Dans le tableau ci-dessous, on donne les observations effectuées entre 1975 et 2014 :

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

- Quelle est la population étudiée ? Quel est l'effectif total ?
- Préciser la variable étudiée et indiquer précisément sa nature. Justifier votre réponse avec soin.
- Quelle est la caractéristique qui a un sens pour ces données ? La préciser.
- Quelles sont les représentations graphiques possibles de ces données ? Effectuez-en une de votre choix.

Exercice n°4 Le centre de transfusion sanguine de Pau a observé la répartition suivante sur 10000 donneurs en 2014.

Facteur/Groupe	O	A	B	AB
Rhésus +	4387	2657	587	194
Rhésus -	989	996	232	73

- Quelle est la population étudiée ? Quelles sont les variables étudiées ? Préciser leur nature.
- Calculer le coefficient Φ^2 de Pearson. Peut-on affirmer que les deux variables sont indépendantes ?

Exercice n°5

Dans une étude¹ sur $n = 820$ accidents de voiture, trois variables ont été prises en compte : "éjection du conducteur" (modalités : "oui" ou "non"), "gravité des blessures" (modalités : "grave" ou "peu grave") et "type d'accident" (modalités : "collision" ou "retournement"). Cela a conduit au tableau de contingence à trois dimensions donné ci-dessous :

	Collision		Retournement	
	Grave	Peu grave	Grave	Peu grave
Ejection	350	150	60	112
Non-éjection	26	23	19	80

1. Quelle est la population ? Quelle est la nature des variables étudiées ?
2. Pour l'ensemble des accidents, quel est la fréquence de la modalité "grave" (pour la variable "gravité des blessures") ?
3. Construire le tableau des fréquences croisant les variables "gravité des blessures" et "type d'accident". Les deux variables peuvent-elles être considérées comme indépendantes ?
4. Calculer la fréquence conditionnelle de la modalité "grave" (pour la variable "gravité des blessures") sachant qu'il y a eu collision. Calculer la fréquence conditionnelle de la modalité "grave" (pour la variable "gravité des blessures") sachant qu'il y a eu retournement. Comparer.
5. Calculer et comparer les quatre fréquences conditionnelles suivantes de la modalité "grave" :
 - (a) sachant qu'il y a eu éjection et collision ;
 - (b) sachant qu'il y a eu éjection et retournement ;
 - (c) sachant qu'il y a eu non-éjection et collision ;
 - (d) sachant qu'il y a eu non-éjection et retournement.

Exercice n°6 Le tableau donne la répartition de $n = 2000$ individus selon l'âge et le sport principal pratiqué.

Age \ Sport	Equitation	Rugby	Golf	Natation	Pelote basque
Moins de 20 ans	50	140	20	140	150
Entre 20 et 30 ans	80	150	50	170	250
Entre 30 et 40 ans	80	50	70	100	200
Plus de 40 ans	30	20	60	90	100

1. Pour chacune des variables étudiées, préciser leur nature.
2. Déterminer les distributions marginales des deux variables.
3. Déterminer la distribution conditionnelle de la variable "âge" sachant que le sport principal pratiqué est le golf.
4. Déterminer la distribution conditionnelle de la variable "sport principal" sachant que les individus ont entre 30 et 40 ans.
5. Calculer le coefficient Φ^2 de Pearson. Peut-on affirmer que les deux variables sont indépendantes ? Justifier avec soin votre réponse.

1. R. Christensen, *Log-linear models and logistic regression*, Springer Texts in Statistics, 1997.