

Statistique descriptive

Statistique univariée

Exercice 1

Dans un établissement scolaire, on demande à un échantillon de 50 élèves combien de fois ils mangent à la cantine dans la semaine.

Les données sont représentées dans le tableau ci-dessous :

Nombre de repas	0	1	2	3	4	5
Nombre de personnes	13	8	10	6	3	10

1. A la main, déterminer le mode, la médiane, le premier quartile et le troisième quartile de la série, la moyenne et l'écart-type de la série.
2. Vérifier les résultats précédents avec la calculatrice.
3. Proposer une démarche Python permettant d'obtenir la liste des effectifs cumulés.

Exercice 2

Une expérience a pour but de comparer le poids moyen d'animaux âgés de un an et qui ont reçu l'un ou l'autre des deux régimes alimentaires A et B depuis leur naissance. L'observation a porté sur 24 sujets de poids sensiblement égaux à la naissance. Les tableaux de relevé sont les suivants :

Masse (en gramme)	Groupe A	Groupe B
[1250 ; 1350[0	3
[1350 ; 1450[1	3
[1450 ; 1550[2	3
[1550 ; 1650[3	1
[1650 ; 1750[4	0
[1750 ; 1850[2	1
[1850 ; 1950[0	1

1. Tracer les deux histogrammes.
2. A la calculatrice, déterminer tous les indicateurs associés au groupe A et au groupe B.
3. Question ouverte : selon vous, y a-t-il un régime meilleur que l'autre ?

Exercice 3

On s'intéresse à la teneur en fer de deux aliments : les épinards et les lentilles. Pour savoir lequel de ces deux aliments apporte le plus de fer, on procède à des analyses dont les résultats sont rapportés dans le tableau suivant :

Echantillon	1	2	3	4	5	6	7	8	9	10
Teneur de fer en mg pour 100 g de produit frais pour les épinards	2,64	2,75	2,82	2,72	2,66	2,59	2,83	2,70	2,67	2,62
Teneur de fer en mg pour 100 g de produit frais pour les lentilles	9,02	9,08	8,82	8,94	8,95	9,11	9,14	9,02	9,04	8,85

1. A la calculatrice, déterminer tous les indicateurs statistiques associés aux échantillons d'épinards et aux échantillons de lentilles.
2. Conclure.

Exercice 4

Une série statistique x est représentée par classes dans le tableau suivant :

x	[0, 2[[2, 6[[6, 7[[7, 9[
Effectifs	5	30	15	10

1. Représenter graphiquement les effectifs cumulés.
On utilisera les bornes supérieures de chaque classe et on reliera les points par une ligne polygonale.
2. Proposer une estimation graphique de la médiane de la série en expliquant la démarche.
3. Proposer une démarche Python permettant d'obtenir la liste des fréquences cumulées.

Exercice 5

Le tableau suivant contient les données récentes sur le nombre de cas de varicelle signalés au Canada de 2005 à 2008, par groupe d'âge (en années) :

Classe d'âge	[0, 1[[1, 5[[5, 10[[10, 15[[15, 20[[20, 25[[25, 30[[30, 40[[40, 60[[60, 100]
2005	99	421	729	190	48	115	22	45	43	44
2006	72	226	352	100	39	25	36	63	43	47
2007	61	172	307	93	31	27	21	45	28	47
2008	75	254	455	139	43	16	35	41	58	35

1. Quel est le nombre moyen de cas déclarés par an ?
2. Dresser l'histogramme correspondant à l'année 2008.
3. Représenter les fréquences cumulées croissantes pour 2008 et en déduire graphiquement une approximation du premier quartile, de la médiane et du troisième quartile.

Statistique bivariée**Exercice 6**

Dans un chenil, on essaye dans un but de dressage de tester les qualités des chiens à des appels sonores et à des signaux visuels. Après une période d'apprentissage suffisante pour tous, on mesure pour chacun des types de signaux les temps de réflexe T_s au son et T_l à la lumière (temps séparant le déclenchement du signal et la réaction de l'animal). Le but de cet expérimentateur est de voir si le temps de réflexe à la lumière est plus ou moins rapide par rapport au temps de réflexe au son.

Pour cela, il dispose d'un échantillon de 10 chiens (A, B, C, D, E, F, G, H, I, J). Les résultats sont les suivants :

Chien	T_s (en centième de s)	T_l (en centième de s)
A	108	112
B	100	99
C	78	82
D	81	91
E	69	82
F	87	94
G	88	85
H	81	81
I	87	88
J	94	99

1. Calculer la moyenne et l'écart-type de l'échantillon pour le test du son.
2. Calculer la moyenne et l'écart-type de l'échantillon pour le test de la lumière.
3. Tracer le nuage de points, calculer le coefficient de corrélation entre le test sonore et le test lumineux. Conclure.

Exercice 7

Le tableau ci-dessous donne les résultats obtenus à partir de 10 essais de laboratoire concernant la charge de rupture d'un acier en fonction de sa teneur en carbone :

x (teneur en carbone en %)	0,60	0,61	0,62	0,64	0,65	0,66	0,68	0,70	0,71	0,74
y (charge de rupture en kg)	71	70	75	74	80	79	79	86	87	95

1. Tracer le nuage de points.
2. Un ajustement affine semble-t-il adapté dans cette situation ? Calculer le coefficient de corrélation pour confirmer.
3. Déterminer une équation de la droite de régression et la représenter sur le graphique.
4. Un acier a une teneur en carbone de 0,77%.
Donner une estimation de sa charge de rupture.

Exercice 8

On cherche à étudier la relation entre le nombre d'enfants d'un couple et son salaire mensuel en euros. On dispose de la série bidimensionnelle suivante :

Salaire en euros (x)	510	590	900	1420	2000	600	850	1300	2200
Nombre d'enfants (y)	4	3	2	1	0	5	6	7	8

Source : *imagination de l'auteur.*

1. Représenter le nuage de points et calculer le coefficient de corrélation linéaire de x et y .
2. Ces deux séries statistiques sont-elles indépendantes ?

Exercice 9

On considère la série statistique bivariée suivante :

x	1,4	4	5	6	9	12	14	15
y	27	44	54	68	119	214	334	406

1. (a) Tracer le nuage de points associé à cette série.
(b) Un ajustement affine semble-t-il adapté ?
2. On procède au changement de variable $Y = \ln(y)$.

(a) Compléter le tableau :

x	1,4	4	5	6	9	12	14	15
y	27	44	54	68	119	214	334	406
$Y = \ln(y)$								

- (b) Réaliser l'ajustement affine de Y en x .
- (c) En déduire un ajustement de y en x .

Exercice 10

Une étude portant sur les dépenses mensuelles moyennes, en euros, a été effectuée sur des ménages français composés d'un couple et d'un enfant mineur.

Les données récoltées sont résumées dans le tableau suivant :

Revenus mensuels moyens (x)	790	1198	1577	1986	3007	4989	7906
Dépenses alimentaires (y)	614	738	828	908	1094	1406	1730
Dépenses habillement-loisirs (z)	40	80	131	199	406	951	2030

1. Représenter sur une même figure le nuage des points x et y et le nuage des points x et z .
Quels ajustements peut-on prévoir ?
2. (a) A l'aide de la calculatrice, calculer le coefficient de corrélation linéaire entre $\ln(x)$ et $\ln(y)$ et celui entre $\ln(x)$ et $\ln(z)$.
(b) En déduire des relations de la forme : $y = kx^\alpha$ et $z = \frac{h}{10^5}x^\beta$.
On arrondira k , h , α et β à 10^{-2} près.
3. (a) Déterminer les dépenses alimentaires et les dépenses d'habillement-loisirs d'un ménage dont les revenus mensuels moyens sont de 4000 euros.
(b) Déterminer le revenu mensuel d'un ménage qui dépense autant pour la nourriture que pour l'habillement-loisirs

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$