

Statistiques descriptives

Exercice 1 Dans une salle, 9 personnes sont assises, leur moyenne d'âge est de 25 ans. Dans une autre salle, 11 personnes sont réunies, leur moyenne d'âge est de 45 ans. On rassemble les deux groupes de personnes. Calculer la moyenne d'âge du groupe ainsi constitué.

Exercice 2 Dans un groupe, 25% des familles ont un enfant, 40% ont deux enfants, 20% ont trois enfants et 15% ont quatre enfants. Quel est le nombre moyen d'enfants dans ce groupe ?

Exercice 3 Dans une entreprise de 360 employés, les cadres gagnent en moyenne 2 000€ et les ouvriers 1200€. L'ensemble des employés gagnent en moyenne 1400€.

1. Quel est le nombre de cadres ?
2. Si on augmente tous les ouvriers de 5%,
 - (a) Quelle est la nouvelle moyenne des salaires des employés de l'entreprise ?
 - (b) L'étendue des salaires augmente-t-elle ?
 - (c) Le salaire médian augmente-t-il ?
3. Mêmes questions si au lieu d'augmenter les ouvriers, on augmente tous les cadres de 5%.

Exercice 4 A la question, "Les statistiques permettent de mentir avec assurance : Quelle est votre opinion ?", 80 personnes interrogées ont ainsi répondu :

Pas du tout d'accord	10
Plutôt d'accord	15
Indifférente	12
Plutôt en accord	18
Tout à fait d'accord	25

Soit X la variable associée à cette enquête.

1. Quelles sont les modalités de X ? Quel est son type ? Quel est son mode ?
2. Établir la distribution en fréquence de cette variable et représenter là par un diagramme rectangulaire, un diagramme en bâtons ou un diagramme semi-circulaire.
3. Quelle est la proportion de sujets n'ayant pas d'opinion extrêmement tranchée sur la question ? Ayant une opinion extrêmement tranchée sur la question ?
4. Quelle est la proportion de sujets qui répondent négativement à la question ? Qui répondent positivement à la question ?

Exercice 5 Voici les 72 résultats d'un examen d'anglais :

12 8 15 11 4 7 13 2 9 10 17 13 14 3 6 6 8 12 9 16 12 9 4 15
 5 3 13 2 18 5 6 11 10 14 6 8 17 10 14 11 16 10 8 10 9 11 10 14
 7 13 19 14 10 15 12 13 6 12 11 9 13 16 15 13 5 10 7 16 10 8 16 11

- Variable quantitative discrète :

1. Remplir le tableau suivant :

Notes	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
Effectif																						

2. Déterminer le mode, la moyenne et l'écart type de la série statistique suivant ce mode de stockage des données.

• Variable quantitative continue :

1. Remplir le tableau suivant :

Intervalle de notes	[0;4[[4;8[[8;12[[12;16[[16;20]
Effectif					

2. Déterminer le mode, la moyenne et l'écart type de la série statistique suivant ce mode de stockage des données.

Exercice 6 Une population de ménages a été répartie en fonction du nombre de parts familiales permettant le calcul de l'impôt sur le revenu.

Nombre de Parts	1	1,5	2	2,5	3	3,5	4	4,5
Nombre de Ménages	48	58	136	184	210	122	62	12

1. Quel est le caractère étudié et quelles sont les modalités ?
2. Déterminer l'étendu et le mode de cette série statistique.
3. Donner une représentation graphique de cette population.
4. Représenter la courbe des fréquences cumulées croissantes.
5. Calculer la médiane, la moyenne et l'écart-type de cette variable.

Exercice 7 Répartition des revenus mensuels dans une population de 4000 personnes en euro :

Salaires	[0;1000[[1000;2000[[2000;3000[[3000;4000[[4000;5000[[5000;6000[
Effectifs	1431	685	1180	468	204	32

1. Quel est le caractère étudié et quelles sont les modalités ?
2. Déterminer l'étendu et le mode de cette série statistique.
3. Après avoir calculé les fréquences, représenter graphiquement la série statistique précédente.
4. Calculer la moyenne ainsi que l'écart type de cette série.
5. Tracer le diagramme des fréquences cumulées croissantes.
6. En déduire la médiane, les quartiles et tracer la boîte à moustache correspondante.
7. Parmi les salaires de moins de 1000 euros, les 2/3 sont inférieurs à 500 euros. Quel nombre de salariés cela représente-t-il ?

Exercice 8 Deux tireurs X et Y s'affrontent en vue d'une sélection lors d'une épreuve comportant 25 tirs sur cible. Les résultats obtenus ont été consignés dans le tableau suivant :

Points	50	30	20	10	0
X	5	7	6	5	2
Y	7	4	6	4	4

1. La moyenne par tir permet-elle de départager les deux concurrents ? Que se passe-t-il si on élimine les cinq plus mauvais tirs de chacun.
2. Calculer la médiane des deux série. Peut-on départager les deux concurrents ?
3. Représenter par deux histogrammes, l'un sous l'autre les séries obtenues par X et Y . Quel est d'après vous le tireur le plus régulier.

4. Calculer les écarts types des deux séries.
5. Tracer les boîtes à moustache correspondant à chacune des deux séries.

Exercice 9 A l'aide des données récoltées sur la page suivante :

http://www.insee.fr/fr/ffc/docs_ffc/psdc.htm

répondre à la question suivante : Quand une commune française n'a pas beaucoup d'habitants, ça veut dire qu'elle en a combien ?

Exercice 10 Le tableau suivant 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.

Teneur en carbone : x_i	70	60	68	64	66	64	62	70	74	62
Charge de rupture (en kg) : y_i	87	71	79	74	79	80	75	86	95	70

1. Représenter graphiquement le nuage de points de coordonnées (x_i, y_i) . On prendra en abscisse 1cm pou une unité en représentant les abscisses à partir de la valeur 60. On prendra en ordonnée à partir de 70.
2. Calculer les coordonnées du point moyen de ce nuage.
3. Déterminer la valeur approchée arrondie à 10^{-3} du coefficient de corrélation linéaire de la série statistique de variables x et y . Interpréter le résultat.
4. Déterminer une équation de la forme $y = ax + b$ de la droite \mathcal{D} de régression de y en x par la méthode des moindres carrés. On donnera les valeurs approchées des coefficients a et b . Tracer la droite \mathcal{D} sur le graphique de la question 1.
5. Déterminer une équation de la forme $x = a'y + b'$ de la droite \mathcal{D}' de régression de x en y par la méthode des moindres carrés. On donnera les valeurs approchées des coefficients a' et b' . Tracer la droite \mathcal{D}' sur le graphique de la question 1.
6. Un acier a une teneur en carbone de 77. Donner une estimation de sa charge de rupture.

Exercice 11 Un mobile est propulsé à très grande vitesse sur un axe, puis il est ralenti. On s'intéresse à la vitesse de ce mobile durant le freinage. Dans tout l'exercice, les distances sont exprimées en mètres, les temps en seconde et donc les vitesses en mètre par seconde. Les résultats seront arrondis au dixième. On a relevé les vitesses instantanées v_i de ce mobile aux instants t_i , pour i variant de 0 à 7.

t_i en s	0	1	2	3	4	5	6	7
v_i en $m.s^{-1}$	215	140	85	57	36	29	27	22

1. Dessiner le nuage de points de cette série statistique et expliquer pourquoi on n'envisage pas un ajustement affine de ce nuage.
2. On pose $n_i = \ln(v_i - 15)$ pour i variant de 0 à 7. Dresser le tableau de la série (t_i, n_i) .
3. Donner une équation de la droite de régression de n en t par la méthode des moindres carrés.
4. En déduire une expression de la vitesse v en fonction du temps t sous la forme $v = \alpha e^{\beta t} + \gamma$, où α , β et γ sont des réels à déterminer et tracer la courbe obtenu sur le graphique de la question 1.

Statistiques descriptives (Méthodes)

☞ Comment calculer la moyenne d'une série statistique à une variable ?

- i. Si on dispose un tableau de données ponctuelles :
 1. totaliser toutes les valeurs observées,
 2. diviser le total par le nombre de valeurs observées (effectif total de la série).
- ii. Si on dispose d'un tableau d'effectifs :
 1. multiplier chaque valeur par son effectif,
 2. totaliser tous ses produits, soit S cette somme,
 3. totaliser tous les effectifs ce qui donne un effectif total n de la série,
 4. diviser la somme S par l'effectif n .
- iii. Si on dispose d'un tableau de fréquences,
 1. multiplier chaque valeur par sa fréquence,
 2. totaliser tous ces produits.
- iv. Si on dispose d'un tableau d'effectifs (ou de fréquences) de classes :
 1. déterminer d'abord les centres des classes,
 2. se ramener au cas ii. ou iii. en remplaçant les valeurs par les centres des classes.
- v. Si on sait que la série est obtenue par agrégation de plusieurs séries statistiques d'effectif et de moyenne connus : se ramener à la définition de la moyenne.

☞ Comment calculer la médiane d'une série statistique à une variable ?

- i. Si on dispose un tableau de données ponctuelles :
 1. ranger les n valeurs observées x_i dans l'ordre croissant au sens large ; noter $(x_{(i)})$ la suite de valeurs ainsi ordonnées,
 2. si n est impair, la médiane est $x_{(\frac{n+1}{2})}$; si n la médiane est $Me = \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)})$.
- ii. Si on dispose d'un tableau d'effectifs ou de fréquence :
 1. calculer les fréquences cumulées croissantes,
 2. c'est la première classe pour laquelle on dépasse 0.5.

Pour déterminer le premier ou troisième quartile on procède de manière identique en remplaçant 0.5 respectivement par 0.25 et 0.75.

☞ Comment calculer l'écart type d'une série statistique à une variable ?

1. On calcule la moyenne arithmétique \bar{x} .
2. Sur le tableau d'effectif ou de fréquence, on rajoute une ligne correspondant à la valeur $(x_i - \bar{x})^2$ où x_i sont les valeurs présent par la série statistique.
3. On calcule la moyenne suivant cette nouvelle valeur.

☞ **Comment calculer une droite de régression par la méthode des moindres carrés ?**

Étant donné une série statistique à deux variables (X, Y) représenté par les points M_i de coordonnées (x_i, y_i) .

Pour calculer la droite de régression de Y en X :

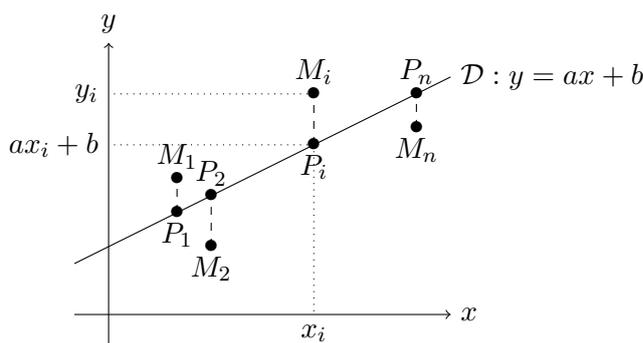
1. Calculer \bar{x} , \bar{y} et $V(X)$.
2. Calculer

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

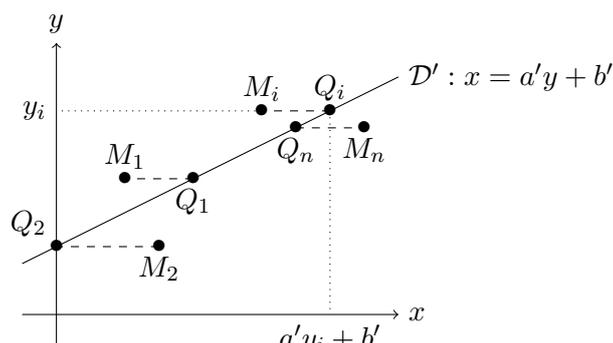
3. Calculer $a = \frac{\text{Cov}(X, Y)}{V(X)}$ et $b = \bar{y} - a\bar{x}$.
4. La droite de régression \mathcal{D} par la méthode des moindres carrés a pour équation $y = ax + b$, elle vérifie les propriétés suivantes :
 - elle passe par le point moyen (\bar{x}, \bar{y}) ;
 - elle minimise les écarts aux carrés $\sum_i P_i M_i^2$ des points à la droites suivant l'axe des ordonnées.

Pour calculer la droite de régression de X en Y :

1. Calculer en plus $V(Y)$.
2. Calculer $a' = \frac{\text{Cov}(X, Y)}{V(Y)}$ et $b' = \bar{y} - a'\bar{x}$.
3. La droite de régression \mathcal{D}' par la méthode des moindres carrés a pour équation $x = a'y + b'$ (ce qui est équivalent à $y = \frac{1}{a'}x - \frac{b'}{a'}$), elle vérifie les propriétés suivantes :
 - elle passe par le point moyen (\bar{x}, \bar{y}) ;
 - elle minimise les écarts aux carrés $\sum_i Q_i M_i^2$ des points à la droites suivant l'axe des ordonnées.



Régression de Y en X



Régression de X en Y

☞ **Comment calculer le coefficient de corrélation linéaire ?** Étant donné une série statistique à deux variables (X, Y) pour calculer le coefficient de corrélation linéaire, on organise les calculs de la manière suivante :

1. Calculer \bar{x} , \bar{y} , $\sigma(X)$ et $\sigma(Y)$.
2. Calculer $\text{Cov}(X, Y)$.
3. Le coefficient de corrélation est $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$.

☞ **Comment interpréter le coefficient de corrélation linéaire ?** Soit θ l'angle formé par la droite \mathcal{D} de régression de Y en X et la droite \mathcal{D}' de régression de X en Y . Le coefficient de corrélation linéaire ρ correspond à $\rho = \cos(\theta)$.

- Si $|\rho| = 1$ alors $a = \frac{1}{a'}$ les droites \mathcal{D} et \mathcal{D}' sont confondues.
- Si $\frac{\sqrt{3}}{2} \leq |\rho| \leq 1$ il y a une bonne corrélation linéaire, \mathcal{D} et \mathcal{D}' sont presque confondues.
- Si $\frac{1}{2} \leq |\rho| \leq \frac{\sqrt{3}}{2}$ il y a une corrélation linéaire médiocre, \mathcal{D} et \mathcal{D}' forment un angle important.
- Si $0 \leq |\rho| \leq \frac{1}{2}$ il y a une mauvaise corrélation linéaire, \mathcal{D} et \mathcal{D}' sont presque orthogonales.