

Exercice 1

A une série $X = \{x_1, x_2, \dots, x_n\}$, on associe la fonction $f : \mathbb{R} \rightarrow \mathbb{R}_+$ définie par :

$$f(x) = \sum_{i=1}^n (x - x_i)^2.$$

- 1) Démontrer que la fonction f admet un minimum sur $\mathbb{R}$ et indiquer en quelle valeur de x il est atteint.
- 2) Que représente, d'un point de vue statistique, la valeur de ce x trouvée à la question précédente ?
- 3) On note $\bar{x}$ la moyenne de la série et σ son écart-type. Soit p le nombre d'éléments de la série statistique compris entre $\bar{x} - 2\sigma$ et $\bar{x} + 2\sigma$. Montrer que

$$\sum_{i=1}^n (\bar{x} - x_i)^2 \geq 4(n - p)\sigma^2.$$

- 4) En déduire qu'au moins les trois quarts des éléments de la série statistique sont compris dans l'intervalle $[\bar{x} - 2\sigma, \bar{x} + 2\sigma]$.

Exercice 2

Soit $(x_1, y_1), \dots, (x_n, y_n)$ une série statistique à deux dimensions. On se propose d'ajuster sur ces données une relation linéaire $y = ax$.

- 1) Déterminer a tel que $S(a) = \sum_{i=1}^n (y_i - ax_i)^2$ soit minimum.
- 2) a étant égal à la valeur obtenue dans la première question, vérifier qu'après ajustement on a :

$$\left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i^2 \right) = \left(\sum_{i=1}^n x_i y_i \right)^2 + \left(\sum_{i=1}^n (y_i - ax_i)^2 \right) \left(\sum_{i=1}^n x_i^2 \right).$$

- 3) Pour mesurer la qualité de l'ajustement linéaire, on pose $d = \frac{\left(\sum_{i=1}^n x_i y_i \right)^2}{\left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n y_i^2 \right)}.$

- a) Démontrer que $d \in [0, 1]$.
- b) Démontrer que tous les points sont alignés si et seulement si $d = 1$.
- 4) **Application** : Les rayons γ pur émis par une substance radioactive sont en partie absorbés par les écrans de plomb. Si N désigne le nombre d'atomes radioactifs exprimé dans une unité qui correspond à la mesure au compteur Geiger, on a la loi théorique : $N = N_0 e^{-ax}$ où N_0 correspond à l'absence d'écran, x désigne l'épaisseur des écrans et a une constante. On a obtenu les mesures ci-dessous ; où n est le nombre d'écrans de 2 mm d'épaisseur.

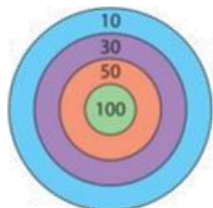
n	0	1	2	3	4	5	6	7
N	8623	7333	6273	5347	4679	4031	3384	2956

- a) Déterminer la valeur de la constante a en utilisant un ajustement linéaire.

- b) Calculer la valeur du coefficient d .

Exercice 3

Deux tireurs à l'arc A et B ont chacun réalisé 25 tirs sur cette cible. Un tir rapporte le nombre de points indiqué sur la zone et zéro point lorsque la cible est manquée. Ils ont obtenus les résultats suivants :



	0	10	30	50	100
A	1	6	2	11	5
B	2	8	3	4	8

- 1) a) Calculer, pour chaque concurrent, l'étendue des scores et la moyenne des tirs.
b) Ces indicateurs permettent-ils de départager les tireurs A et B ?
c) Quel tireur vous semble, a priori, sans aucun calcul, le plus régulier ? Pourquoi ?
- 2) Déterminer la médiane et l'écart inter-quartile des points marqués par chaque concurrent. Cela confirme-t-il votre impression ?
- 3) a) Déterminer l'écart-type de la série des points marqués par chaque joueur.
b) En déduire le tireur le plus régulier. Justifier.

Exercice 4

Une étude théorique de l'évolution d'une population en extinction conduit à penser que le nombre d'individus N de cette population varie avec le temps suivant une loi du type $N(t) = a \exp(-kt)$; où a et k sont des constantes strictement positives. On veut déterminer expérimentalement la valeur de la constante k . Pour cela, on observe pendant 8 mois un échantillon composé initialement de 200 individus et on note 'a la fin de chaque mois le nombre de survivants. Les résultats sont les suivants :

t (en mois)	1	2	3	4	5	6	7	8
survivants après le t ième jour	180	154	140	120	112	97	84	76

- 1) En déduire une valeur approchée de k par la méthode des moindres carrés.
- 2) Quel sera 'a votre avis, le nombre de survivants de cet échantillon
 - a) à la fin de l'année en cours ?
 - b) à la fin de l'année suivante ?

Exercice 5

N.B : Les parties I)) et II) de cet exercice sont indépendantes.

- I) Soient la série double (x_i, y_i) ; $i = 1, \dots, n$, et a, b, a' et b' des nombres réels. On rappelle que

$$Cov(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad V(x) = \bar{x}^2 - \bar{x}^2.$$

- 1) Montrer que $Cov(x, y) = \frac{1}{n} \sum x_i y_i - \bar{x} \bar{y}$
- 2) On pose $z = ax + b$ et $w = a'y + b'$.
 - a) Calculer $V(z)$ et $V(x + y)$.
 - b) calculer $cov(x, x)$ et $cov(z, w)$
- II) On considère que x_i désigne le rang ou le numéro du mois et y_i le niveau de pollution relevé dans une ville de Janvier 1994 à Décembre 1998. On donne

$$\sum y_i = 6000, \quad \sum y_i^2 = 1500000 \quad \text{et} \quad \sum x_i y_i = 300000$$

- 1) Calculer et commenter $Cov(x, y)$ ainsi que le coefficient de corrélation linéaire.
- 2) Déterminer l'équation de la droite de régression de Y en X en utilisant la méthode des moindres carrés.
- 3) A combien peut-on estimer le niveau de pollution pour le mois de Janvier 1999.

Exercice 6

Dans un magasin de pièces détachées, sur un lot de 100 pièces vendues en une année; les prix s'échelonnent entre 200 FCFA et 800 FCFA selon la répartition suivante :

Prix en FCFA	[200,300[	[300,450[	[450,550[	[550,600[	[600,800[
Nombre de pièces vendues	15	35	25	10	15

1. Calculer les effectifs cumulés croissantes et les fréquences cumulées croissantes de cette série statistique.
2. Déterminer le mode M_0 de cette série, graphiquement et par calcul.
3. Déterminer la médiane M_e de cette série en explicitant vos calculs. Donner son interprétation.
4. Déterminer les quartiles Q_1 et Q_3 puis calculer l'inter-quartile relatif. Conclure.

Exercice 7

Le tableau suivant est relatif aux valeurs des exportations (X) et des importations (Y) de 9 pays :

X \ Y	[25, 500[	[500, 1500[	[1500, 2500[	Total
[10, 500[	4	n_{12}	n_{13}	$n_{1.}$
[500, 1500[	n_{21}	1	n_{23}	$n_{2.}$
[1500, 2500[	n_{31}	n_{32}	1	$n_{3.}$
Total	5	2	2	$n_{..}$

Compléter le tableau de contingence relatif à cette double série sachant que les moyennes conditionnelles de X sont : $X_1 = 404$, $X_2 = 1500$, $X_3 = 1127,5$.

Exercice 8

On a administré une drogue à 40 rats. Pour 10 rats du lot, la dose d était de 1 unité ($d_1 = 1$), pour 10 autres de 2 unités ($d_2 = 2$), pour 10 autres de 4 unités ($d_3 = 4$) et pour les 10 derniers de 8 unités ($d_4 = 8$). On a mesuré la performance Y de ces rats sous l'influence de la drogue dans une certaine tâche et obtenu les résultats moyens suivants :

Dose d_i	d_1	d_2	d_3	d_4
Performance y_i	9	12,1	21,2	22,8

On pose $x_i = 2 \frac{\ln d_i}{\ln 2} - 3$ pour $i = 1, 2, 3, 4$.

- 1) Représenter dans le plan le nuage de point (x_i, y_i) pour $i = 1, 2, 3, 4$.
- 2) Déterminer l'équation de la droite de régression de Y par rapport à X par la méthode des moindres carrés.
- 3) Calculer le coefficient de corrélation linéaire entre X et Y .
- 4) Donner une estimation de la performance pour une dose de 3 unités.

Exercice 9

Dans une société de gardiennage, le salaire mensuel moyen est de 18 euros. Les salaires mensuels moyens des employés hommes et femmes sont respectivement de 20 euros et 12 euros. Déterminer le nombre d'hommes et de femmes sachant que la société emploie 20 personnes.

Exercice 10

Soit la série à deux valeurs isolées x_1, x_2 positives. Montrer que $H \leq G \leq \bar{X}$.