

Feuille de TD n° 1. Elaboration des données expérimentales

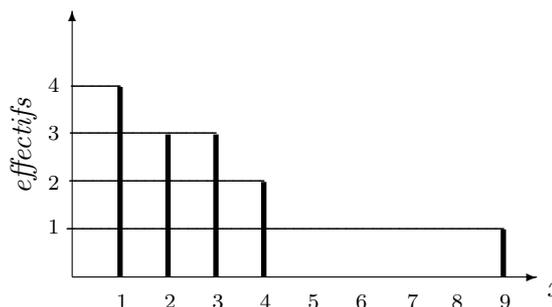
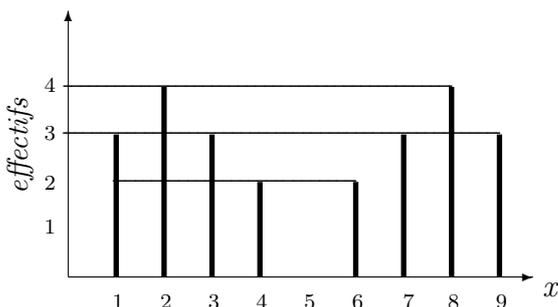
Exercice 1 Pour tester son efficacité, un somnifère est administré à 14 sujets insomniaques. Les résultats sont les suivants, en heures de sommeil :

Patient n^0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x =heures de sommeil	5	4	4	5	9	6	8	5	3	6	6	7	8	5

- Organiser l'échantillon de la variable x dans un tableau d'effectifs et de fréquences.
- Réprésenter les effectifs de l'échantillon de la variable x par un diagramme en bâtons.
- On écrit les données dans l'échantillon sous la forme x_1, x_2, \dots, x_N (donc $N = 14$ et $x_1 = 5$, $x_2 = 4$, etc). Déterminer l'effectif cumulé et la fréquence cumulée de x_6 et de x_7 .
- Tracer le diagramme des fréquences cumulées.
- Déterminer la moyenne, le mode, la médiane et les quartiles de l'échantillon.
- Calculer l'étendue, l'étendue interquartile, l'écart absolu moyen, la variance et l'écart-type de l'échantillon.

Exercice 2 Les graphiques suivants sont les diagramme en bâtons des effectifs de deux échantillons différents de la variable x .

- Tracer sur les graphiques la moyenne, le mode, la médiane et les quartiles.
- Calculer les paramètres de dispersion pour chacun des échantillons.



Exercice 3 Dans une étude sur le parasitisme des châtaignes, on compte le nombre de parasites présents dans chaque fruit d'un échantillon. Les résultats sont les suivants :

n^0 parasites par fruit	0	1	2	3	4	5	6
fréquence	0,6	0,25	0,1	0,02	0,015	0,01	0,005

- Tracer le diagramme en bâtons des fréquences.
- Calculer la moyenne et l'écart-type pour les données ci-dessus.
- Peut-on déterminer la taille de l'échantillon considéré ?

Exercice 4 Des études récentes sur la dynamique du carbone organique total (TOC) en Méditerranée ont analysé les concentrations de TOC dans des échantillons d'eau prélevés dans une même zone. Les données relevées pendant 12 semaines successives sont les suivantes (en mmol C m²) :

8300 8378 8480 8643 9230 8948 9023 8833 8878 8705 8430 8400

- Construire le graphique chronologique des données.

- (b) Déterminer la moyenne \bar{x} des données et la tracer sur le graphique de (a).
(c) On note par x_k la k -ème donnée dans l'échantillon des concentrations de TOC. Calculer $\sum_{k=1}^{12} (x_k - \bar{x})$.
(d) Calculer $\sum_{k=1}^N (x_k - \bar{x})$ pour un échantillon arbitraire x_1, \dots, x_N de la variable x .

Exercice 5 Le botaniste américain Edgar Anderson collecta en 1935 les données suivantes. Elles donnent les mesures en centimètres de la longueur du sépale pour trois espèces d'iris : *Iris setosa*, *I. versicolor* et *I. virginica*.

Iris setosa

7,9 7,7 7,7 7,7 7,7 7,6 7,4 7,3 7,2 7,2 7,2 7,1 7,0 6,9 6,9 6,9 6,9
6,8 6,8 6,8 6,7 6,7 6,7 6,7 6,7 6,7 6,7 6,7 6,6 6,6 6,5 6,5 6,5 6,5
6,5 6,4 6,4 6,4 6,4 6,4 6,4 6,4 6,3 6,3 6,3 6,3 6,3 6,3 6,3 6,3

I. versicolor

6,3 6,2 6,2 6,2 6,2 6,1 6,1 6,1 6,1 6,1 6,1 6,0 6,0 6,0 6,0 6,0 6,0
5,9 5,9 5,9 5,8 5,8 5,8 5,8 5,8 5,8 5,8 5,7 5,7 5,7 5,7 5,7 5,7 5,7
5,7 5,6 5,6 5,6 5,6 5,6 5,6 5,5 5,5 5,5 5,5 5,5 5,5 5,5 5,4 5,4

I. virginica

5,4 5,4 5,4 5,4 5,3 5,2 5,2 5,2 5,2 5,1 5,1 5,1 5,1 5,1 5,1 5,1 5,1
5,1 5,0 5,0 5,0 5,0 5,0 5,0 5,0 5,0 5,0 5,0 4,9 4,9 4,9 4,9 4,9 4,9
4,8 4,8 4,8 4,8 4,8 4,7 4,7 4,6 4,6 4,6 4,6 4,5 4,4 4,4 4,4 4,3

Pour chacun des échantillons :

- (a) Construire le tableau d'effectifs et des fréquences.
(b) Grouper les données par classes d'amplitude constant 0,4 et calculer les effectifs et les fréquences des classes.
(c) Tracer les histogrammes des fréquences.

Pour une classe $[a, b[$ (ou $[a, b]$) on appelle $(a + b)/2$ le centre de la classe. Soient I_1, \dots, I_h les classes d'un échantillon de la variable x . On indique par X_k le centre de la classe I_k et par N_k son effectif. La moyenne de l'échantillon peut être calculée approximativement par la formule

$$\bar{x}_{\text{appr}} = \frac{1}{N} \sum_{k=1}^h N_k X_k,$$

où N est la taille de l'échantillon. De la même façon, la variance et l'écart-type de l'échantillon peuvent être calculés approximativement par les formules

$$(v_x)_{\text{appr}} = \frac{1}{N} \sum_{k=1}^h N_k (X_k - \bar{x}_{\text{appr}})^2 \quad \text{et} \quad \sigma_{\text{appr}} = \sqrt{(v_x)_{\text{appr}}}.$$

- (d) Calculer la moyenne approximative et l'écart-type approximatif des trois espèces d'iris.
(e) Comparer la moyenne approximative \bar{x}_{appr} et l'écart-type approximatif σ_{appr} des données classées d'iris *setosa* avec respectivement la moyenne et l'écart-type calculés à partir des données.

Exercice 6 (Manipulation du symbole \sum) On désigne par \mathbb{N} l'ensemble des nombres naturels, c'est-à-dire $\mathbb{N} = \{0, 1, 2, 3, \dots\}$.

- (a) Pour tout $k \in \mathbb{N}$ on pose $x_k := k$. Evaluer :
(1) $\sum_{k=3}^5 x_k$ et $\sum_{k=1}^7 x_k$;
(2) $\sum_{k=1}^N x_k$, où $N \in \mathbb{N}$ est un nombre naturel fixé.
(b) Soit $x_1 = 3$, $x_2 = 2$ et $x_3 = 1$. Evaluer $\sum_{j=1}^3 (x_j \sum_{k=1}^j x_k)$.

(c) Soit x_1, x_2, \dots, x_N un échantillon de nombres réels. Vérifier si les égalités suivantes sont vraies ou fausses :

- (1) $\sum_{k=1}^N (x_k - 1) = \sum_{k=1}^N x_k - 1$;
- (2) $\sum_{k=1}^N (x_{k+1} - x_k) = x_{N+1} - x_1$;
- (3) $\sum_{k=1}^N a \cdot x_k = a \sum_{k=1}^N x_k$, où a est un nombre réel fixé.
- (4) $\sum_{k=1}^N x_k^2 = \left(\sum_{k=1}^N x_k \right)^2$.

Exercice 7 Soit x_1, \dots, x_N un échantillon de taille N de la variable x . On suppose que x_1, \dots, x_h sont les valeurs distinctes de x dans l'échantillon. Soit N_k l'effectif de x_k et F_k la fréquence de x_k dans l'échantillon.

(a) Vérifier les *formules de Koenig* pour la variance de x :

$$v_x = \frac{1}{N} \sum_{k=1}^N x_k^2 - \bar{x}^2 \quad (1)$$

$$v_x = \frac{1}{N} \sum_{k=1}^h N_k x_k^2 - \bar{x}^2 \quad (2)$$

$$v_x = \sum_{k=1}^h F_k x_k^2 - \bar{x}^2. \quad (3)$$

(b) Vérifier la *formule de Koenig* pour la covariance des variables x et y .

$$c_{x,y} = \frac{1}{N} \sum_{k=1}^N x_k y_k - \bar{x} \bar{y}.$$

Exercice 8 Le rendement en céréales est relié linéairement à la quantité d'engrais utilisée. Sept observations donnent les valeurs suivantes pour l'engrais x (en kg/ha) et le rendement correspondant y (en quintaux/ha).

x (en kg/ha)	100	200	300	400	500	600	700
y (en q/ha)	40	50	50	70	65	65	80

- (a) Représenter graphiquement ces points, et tracer approximativement la droite qui passe le mieux à travers ces points.
- (b) Déterminer le coefficient de corrélation linéaire entre x et y dans échantillon donné.
- (c) Déterminer la droite d'ajustement au moyen de la méthode des moindres carrés.
- (d) Représenter graphiquement les points et la droite d'ajustement trouvée.
- (e) Utiliser l'équation d'ajustement pour prévoir le rendement si on utilise 450 kg/ha.

Exercice 9 Une enquête a été réalisée sur l'ensemble des 943 naissances uniques des 3 maternités d'une même ville pendant l'année 1999. Les femmes ont été interrogées sur leur consommation d'alcool pendant la grossesse et ont été classées en 4 catégories selon leur consommation journalière moyenne d'alcool (variable x , en g/jour). Les résultats concernant le poids de naissance des nouveau-nés (variable y , en kg) sont présentés dans le tableau ci-dessous. Sur la deuxième ligne du tableau apparaissent les effectifs de chaque catégorie et sur la troisième ligne (resp. sur la dernière ligne) sont indiquées les sommes des valeurs de la variable y (resp. de la variable y^2) pour les nouveau-nés correspondants à chaque catégorie.

consommation moyenne x (en g/jour)	0	5	15	25
effectif	593	224	87	39
$\sum y$	1962,9	716,8	265,4	111,1
$\sum y^2$	6632,6	2342,3	837,1	327,1

- (a) Déterminer la droite de régression de y en fonction de x au moyen de la méthode des moindres carrés.
- (b) Calculer le coefficient de corrélation linéaire de x et y .

Exercice 10 (extrait de l'examen de janvier 2004) La vitesse d'une réaction dans laquelle une substance S donne un produit P sous l'effet d'une enzyme E est donnée par l'équation de Michaelis-Menten :

$$v = V_{\max} \frac{[S]}{[S] + K}$$

où $[S]$ est la concentration de S , V_{\max} est la vitesse maximum possible et K est la constante de Michaelis-Menten.

- (a) Ecrire $1/v$ comme fonction de $1/[S]$ (représentation de Lineweaver-Burk).

Expérimentalement, on fait agir une enzyme E à concentration fixe sur un substrat S à diverses concentrations; en mesurant la concentration $[S]$ (exprimée en 10^{-5} mol/l) à différents instants, on évalue la vitesse initiale (exprimée en $\mu\text{mol}/\text{min}$). Dans le tableau suivant, on a reporté les valeurs correspondantes de $1/[S]$ et $1/v$.

$x = 1/[S]$	2	1	0,8	0,7	0,5	0,4	0,3	0,2	0,1
$y = 1/v$	0,105	0,052	0,045	0,04	0,03	0,025	0,02	0,015	0,01

- (b) Déterminer la médiane et les 1^{er} et 3^{ème} quartiles pour la variable $1/[S]$ dans l'échantillon considéré.
- (c) On peut vérifier que les moyennes de $1/[S]$ et $1/v$ dans l'échantillon sont

$$\bar{x} := \text{moyenne de } 1/[S] = 2/3 \quad \text{et} \quad \bar{y} := \text{moyenne de } 1/v = 0,0380.$$

Peut-on déduire que les moyennes de $[S]$ et de v sont respectivement $3/2$ et $1/0,038$? Justifier votre réponse.

- (d) On note respectivement x_k et y_k les k -èmes données dans les échantillons de valeurs de $1/[S]$ et de $1/v$ dans le tableau. En utilisant la formule de Koenig, déterminer la variance de la variable $1/[S]$ et donc la covariance de $1/[S]$ et de $1/v$ pour l'échantillon donné en sachant que

$$\sum_{k=1}^9 x_k^2 = 6,68 \quad \text{et} \quad \sum_{k=1}^9 x_k y_k = 0,361.$$

- (e) Déduire, au moyen de la méthode des moindres carrés, les valeurs des paramètres V_{\max} et K à partir des valeurs de $1/[S]$ et $1/v$ données dans le tableau.