

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE

UNIVERSITE ABOUBEKR BELKAID-TLEMSEN

FACULTE DES SCIENCES DE L'INGENIEUR

Recueil d'exercices
(avec corrigés)

Préparé par
Dr Djawad Zendagui

TD n°1

Exercice n°1

Le tableau ci après donne la concentration d'azote total (mg/L) dans la rivière Yamaska pour 19 stations différentes.

1. Identifier la population, le caractère et les modalités
2. En utilisant la méthode de Sturge établir le tableau statistique
3. En utilisant la méthode de Yule établir le tableau statistique
4. Tracer l'histogramme de cette variable

Echantillon	Concentration (mg/L)
1	1.37
2	1.29
3	0.87
4	2.14
5	1.1
6	1.12
7	2.43
8	0.85
9	1.35
10	1.13
11	0.48
12	1.17
13	1.55
14	1.04
15	1.01
16	0.6
17	2.02
18	0.93
19	1.98

Exercice n°2

Le tableau suivant donne les consommations mondiales par source d'énergie et par zone (1992)¹ en **Mtep** (millions de tonnes équivalent pétrole) (fossile = pétrole + gaz naturel + charbon) :

	Fossile	Nucléaire	Hydraulique
Amérique du nord	1 930,4	188,5	50,5
Amérique latine	341,3	3,3	33,6
Europe O.C.D.E.	1 172,1	197,5	38,1
Europe hors O.C.D.E.	1 431,0	58,3	23,0
Moyen-Orient	252,8		1,2
Afrique	208,6	2,4	6,2
Asie et Australasie	1 737,4	81,9	36,0

1. Identifier les deux caractères étudiés
2. Représenter graphiquement ces données.
3. Mettre en évidence, graphiquement, la partie du monde consommant, en proportion, le plus d'énergie d'origine hydraulique.

NB : La représentation de ces données doit prendre en compte le type de source d'énergie et la distribution géographique.

¹ (Source : Encyclopædia Universalis, **Les Chiffres du monde 1994**, p.544).

Corrigé du TD n°1

Exercice n°1 :

1.

<u>Population</u>	Les points de mesure (les 19 stations)
<u>Individu</u>	Une station parmi ces stations
<u>Caractère</u>	La concentrations d'azote total (mg/L)
<u>Modalités</u>	Les valeurs de cette concentration [0.48, 2.43] mg/L

2. Méthode de Sturge

Le nombre d'individus $n = 19$.

Les valeurs observées du caractère ont pour valeur maximale et minimale

$$V_{\max} = 2.43 \text{ mg/L et } V_{\min} = 0.48 \text{ mg/L}$$

- Détermination du nombre de pas k :

$$k = 1 + 3.3 \log_{10}(19) = 5.22 \text{ soit } k = 5$$

- Calcul du pas

$$Pas = \frac{2.43 - 0.48}{5} = 0.39 \text{ mg/L}$$

Limite des classes (mg/L)		Centre de classe (mg/L)	Effectif	Fréquence
Limite inférieure	Limite supérieure			
0.48	0.87	0.675	3	0.16
0.87	1.26	1.065	8	0.42
1.26	1.65	1.455	4	0.21
1.65	2.04	1.845	2	0.11
2.04	2.43	2.235	1	0.05
2.43	2.82	2.625	1	0.05
		Total	19	1

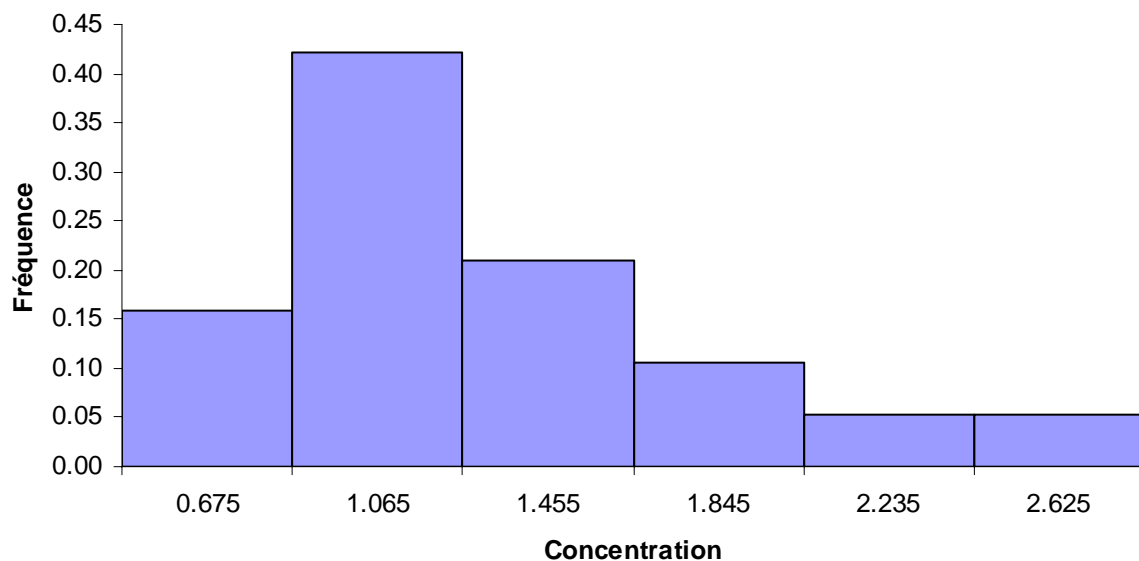
3. Méthode de Yule

$n = 19$, $V_{\max} = 2.43 \text{ mg/L}$ et $V_{\min} = 0.48 \text{ mg/L}$

- Détermination du nombre de pas k :

$k = 2.5 \sqrt[4]{19} = 5.21$ soit $k = 5$ (même résultat que celui obtenu par la méthode de Sturge).

4. Le tracé de l'histogramme



Exercice n°2 :

1. La question est de savoir, considérant une consommation d'énergie : *quelle est son origine et dans quelle partie du monde elle est consommée*. Il y a donc deux variables statistiques *qualitatives* à étudier.
2. **Représentation graphique** : Les effectifs (consommations) des modalités de ces deux variables statistiques s'obtiennent en faisant les sommes par ligne ou par colonne du tableau de l'énoncé.

Origine	fossile	nucléaire	hydraulique	Mtep
Amérique du nord	1 930,4	188,5	50,5	2 169,4
Amérique latine	341,3	3,3	33,6	378,2
Europe O.C.D.E.	1 172,1	197,5	38,1	1 407,7
Europe hors O.C.D.E.	1 431,0	58,3	23,0	1 512,3
Moyen-Orient	252,8		1,2	254,0
Afrique	208,6	2,4	6,2	217,2
Asie et Australasie	1 737,4	81,9	36,0	1 855,3
Mtep	7 073,6	531,9	188,6	7 794,1

Variable "**Origine de l'énergie consommée**" :

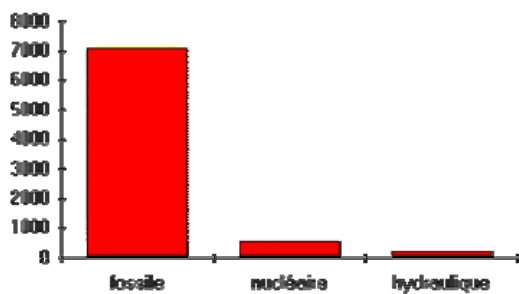
Origine	fossile	nucléaire	hydraulique
Mtep	7 073,6	531,9	188,6

Variable "**Partie du monde**" :

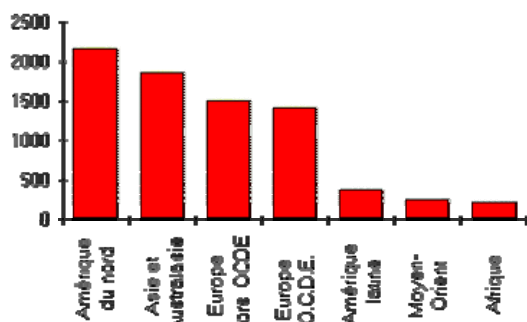
	Mtep
Amérique du nord	2 169,4
Amérique latine	378,2
Europe O.C.D.E.	1 407,7
Europe hors O.C.D.E.	1 512,3
Moyen-Orient	254,0
Afrique	217,2
Asie et Australasie	1 855,3

a) Histogramme.

Les comparaisons sont plus faciles lorsque les données sont classées, par ordre croissant ou par ordre décroissant.

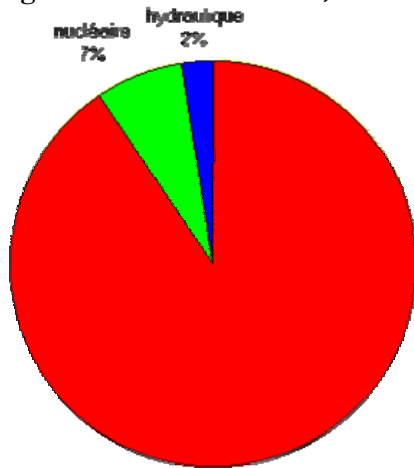


Consommation d'énergie suivant la source

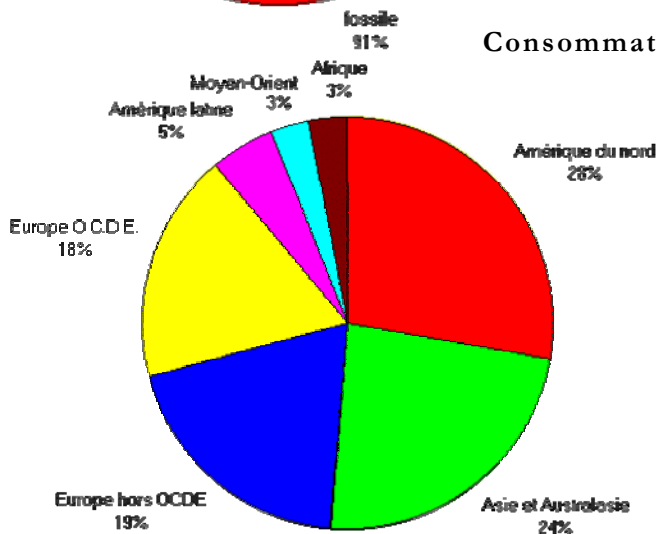


Consommation d'énergie suivant la zone

b) Diagrammes circulaires, ou secteurs.



Consommation d'énergie suivant la source

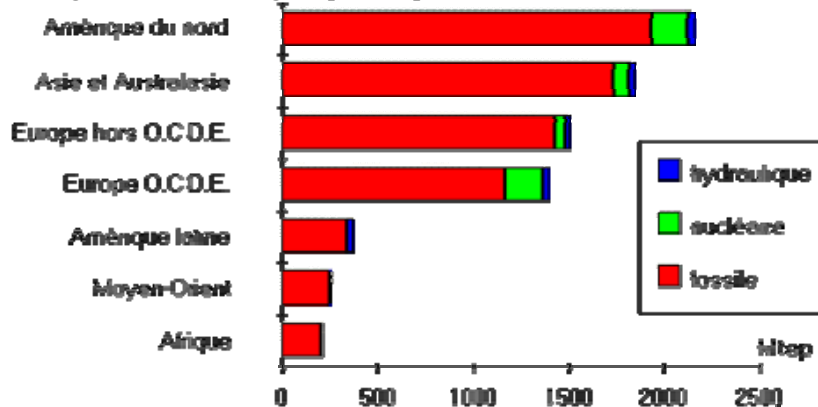


Consommation d'énergie suivant la zone

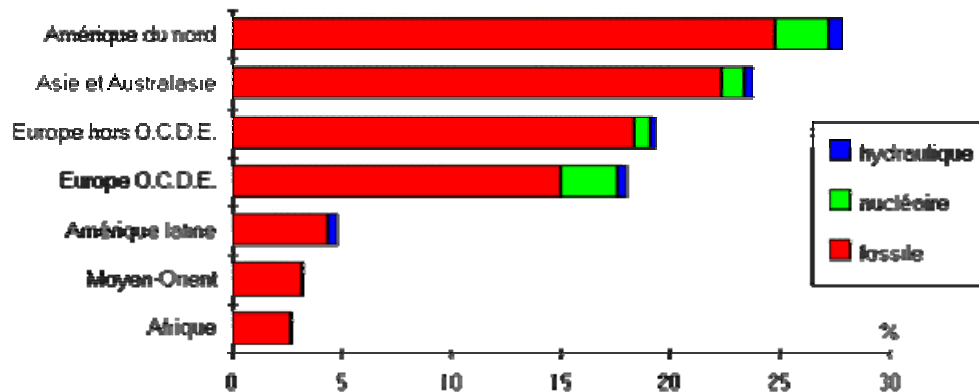
Le diagramme en histogramme et le diagramme circulaire montrent l'écrasante prépondérance des énergies fossiles dans la consommation mondiale d'énergie.

c) Barres.

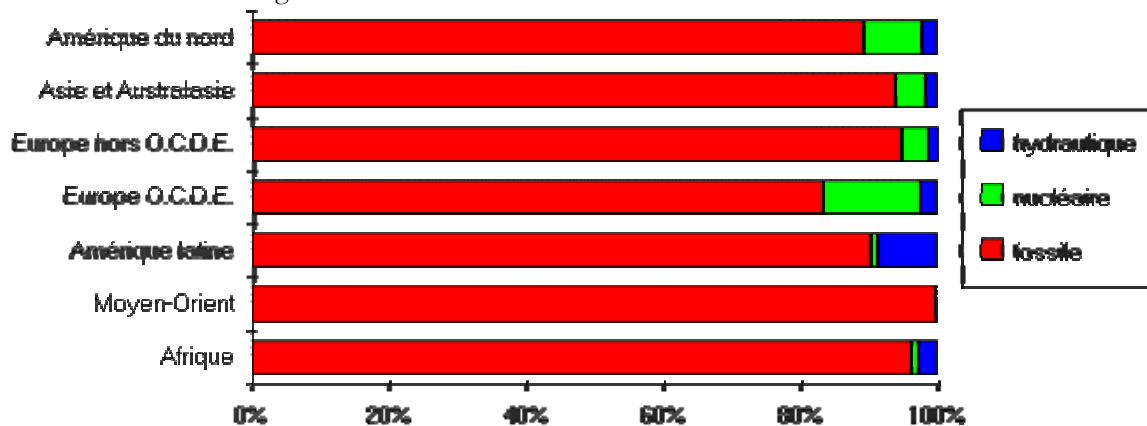
Les histogrammes peuvent être horizontaux : on obtient ce qu'on appelle un **diagramme en barres**. La représentation graphique des deux variables peut se faire par un diagramme en barres, avec une barre horizontale pour chaque partie du monde. Chaque barre horizontale a une longueur proportionnelle à la consommation de la partie du monde, et la barre est divisée en trois parties proportionnellement à la répartition des origines de l'énergie que la partie du monde consomme.



Cette représentation absolue devient une représentation relative lorsqu'on divise chaque énergie consommée par l'énergie totale consommée : la forme générale de la représentation graphique reste inchangée, seule l'échelle horizontale est différente, elle est relative au lieu d'être absolue.



La représentation en barres permet aussi de comparer la part des diverses sources d'énergie consommée à l'intérieur de la consommation de chaque zone. Dans cette représentation, chaque barre horizontale a la même longueur et elle est divisée en trois parties proportionnelles aux pourcentages des diverses sources à l'intérieur de la consommation d'énergie d'une même zone.



Dans le graphique précédent, on voit que :

- l'**Amérique latine**, qui a la partie **hydraulique** bleue la plus importante, est la partie du monde qui, en proportion, consomme le plus d'énergie d'origine hydraulique ;
- l'**Europe O.C.D.E.**, qui a la partie **nucléaire** verte la plus importante, est la partie du monde qui, en proportion, consomme le plus d'énergie d'origine nucléaire ;
- le **Moyen-Orient**, qui a la partie **fossile** rouge la plus importante, est la partie du monde qui, en proportion, consomme le plus d'énergie d'origine fossile.

Exercice : 1

La Figure 1 donne l'histogramme ainsi que la fonction cumulée de la variable statistique X (m).

1. Déterminer graphiquement le mode
2. Déterminer graphiquement la médiane
3. Déterminer la moyenne

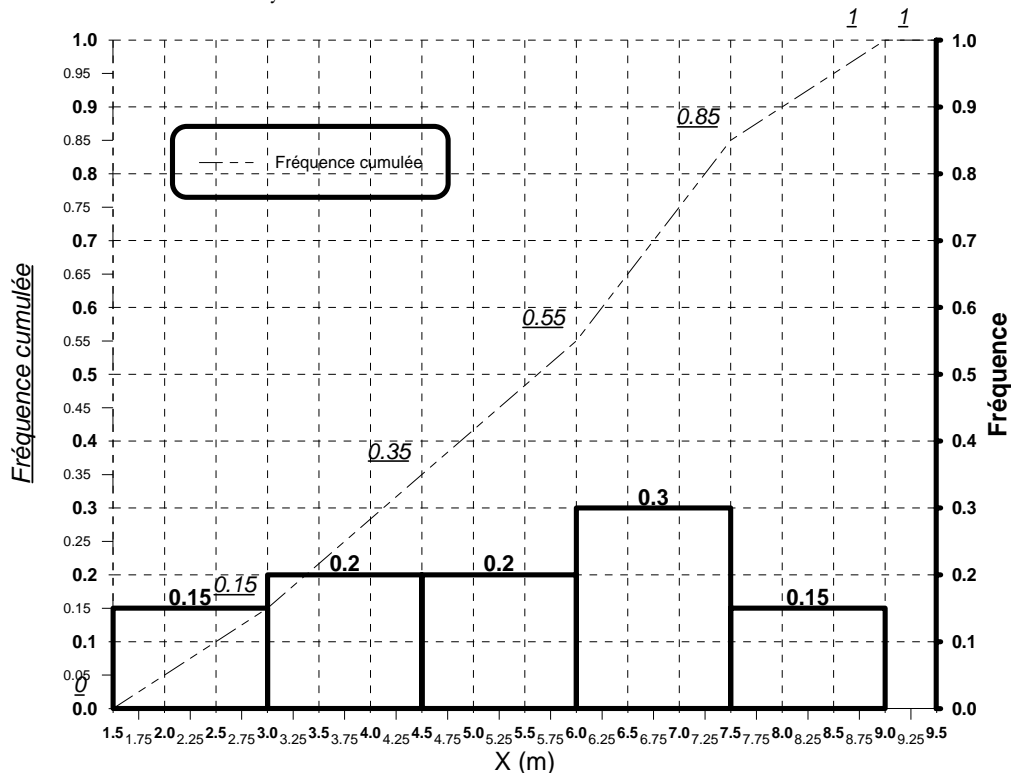


Figure 1

Exercice 2 : Principe du juste-à-temps

Quelques informations à lire avant de faire l'exercice

La gestion de la production.

- Années 45-75 : offre << demande → production de masse
- Années 75-80 : offre = demande → concurrence
- Recherche de critères de différenciation et de performance (diversité, qualité, rapidité)
- Années 80-aujourd'hui : offre > demande → diversité
- Concurrence accrue et Mondialisation

Un exemple de la gestion de la production est : **le principe du juste-à-temps**.

L'origine du juste-à-temps est le secteur automobile japonais :

- Période après 45
- Marché fermé sur l'extérieur et de petite taille
- Pas de matière première
- Manque de place, organisation optimisée des ateliers
- Peu de flexibilité de main d'œuvre

Le juste-à-temps avance le concept des 7 zéros : Accident, conflit, défaut, délais, pannes, papier, stock.

Une des difficultés de ce principe est justement la maîtrise de la chaîne de production. Le présent exercice a pour but d'analyser une série de données sur la maîtrise de cette chaîne.

Maintenant vous pouvez faire l'exercice

Une pièce mécanique est fabriquée en deux temps. Elle passe en premier lieu dans un appareil (noté A) puis passe au niveau d'un autre appareil (noté B). Le temps d'attente de la pièce, pour être traitée par l'appareil B, doit être fixe et maîtrisable. Ce qui est n'est pas le cas. Pour ce faire, un traitement statistique sur ce temps d'attente est effectué pour savoir s'il faut rajouter un autre appareil B

Un opérateur chronomètre le temps que met cette pièce pour être traitée par l'appareil B. Il doit donc pour chaque pièce relever le temps d'attente (en minutes). Le tableau 1 donne les valeurs de ces temps d'attente² :

1. Définir la population, l'individu, le caractère et les modalités de cette étude.
2. Etablir le tableau statistique
3. Tracer l'histogramme et la fonction cumulative (Figure 2).
4. Déterminer graphiquement la médiane (Figure 2)
5. Calculer les moyennes arithmétique, géométrique, harmonique et quadratique
6. Déterminer graphiquement (Figure 2) et analytiquement le mode
7. Classer les moyennes, le mode et la médiane. Conclusion
8. Si la moyenne du temps d'attente dépasse 2 min alors un autre appareil de type B doit être rajouté. Pour le présent est il obligatoire de rajouter un autre appareil de type B ?

Tableau N°1

N° de l'observation	1	2	3	4	5	6	7	8	9	10	11	12
Valeur du temps d'attente (min décimale)	2,2	2,1	1,78	1,42	1,4	1,82	2	1	2,9	2,3	1,7	1,6

Exercice n°3

Dans une wilaya, on a relevé la superficie des exploitations agricoles. On a obtenu les résultats suivants :

Tableau N°2

Taille (hectares)	Nombre
0 - 30	15
30 - 60	40
60 - 90	30
90 - 120	10
120 - 150	5

Exemple : 30 exploitations ont une superficie comprise entre [60-90[hectares.

1. Tracer l'histogramme de cette distribution (Figure 3)
2. Tracer la fonction cumulative (Figure 4)
3. Déterminer graphiquement le mode (Figure 3) et la médiane (Figure 4)
4. Sachant que la moyenne nationale est d'environ 50 hectares, pensez-vous que cette wilaya est une wilaya de petites exploitations ?
5. Calculer la moyenne et l'écart type.

² La notation décimale est adoptée, c'est-à-dire qu'une conversion est pratiquée du système normal vers le système décimal

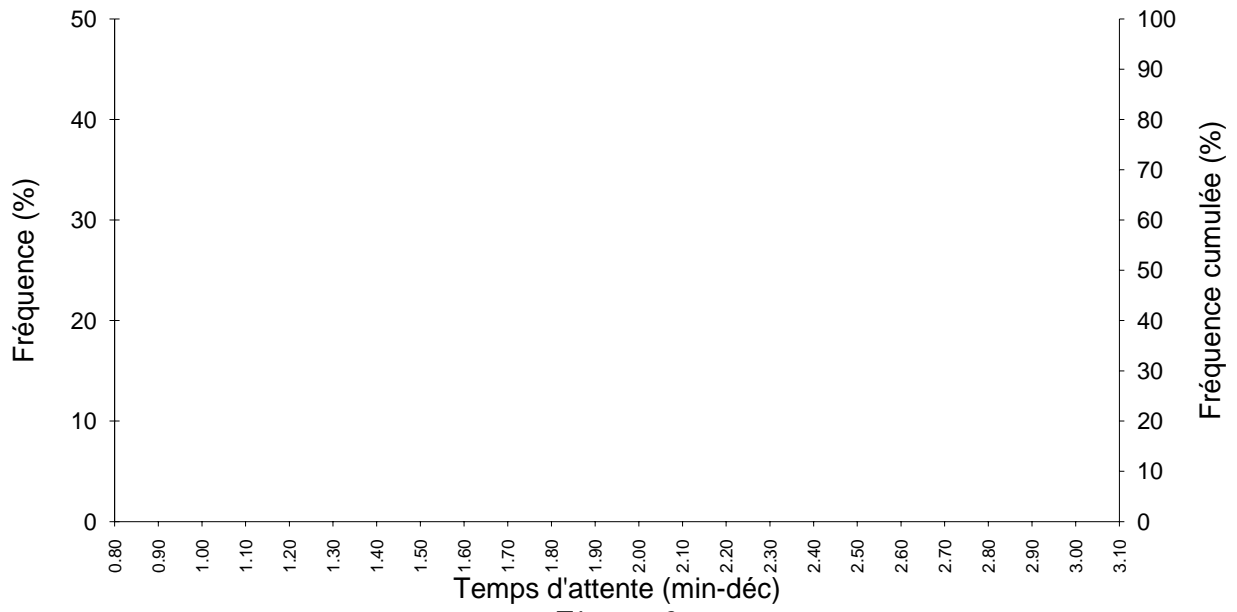


Figure 2

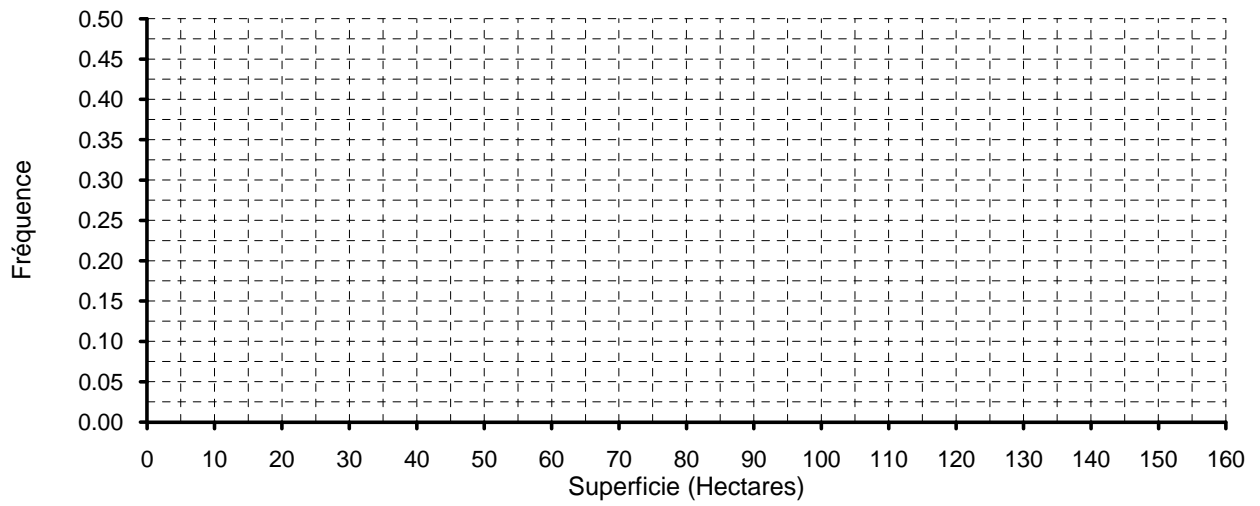


Figure 3

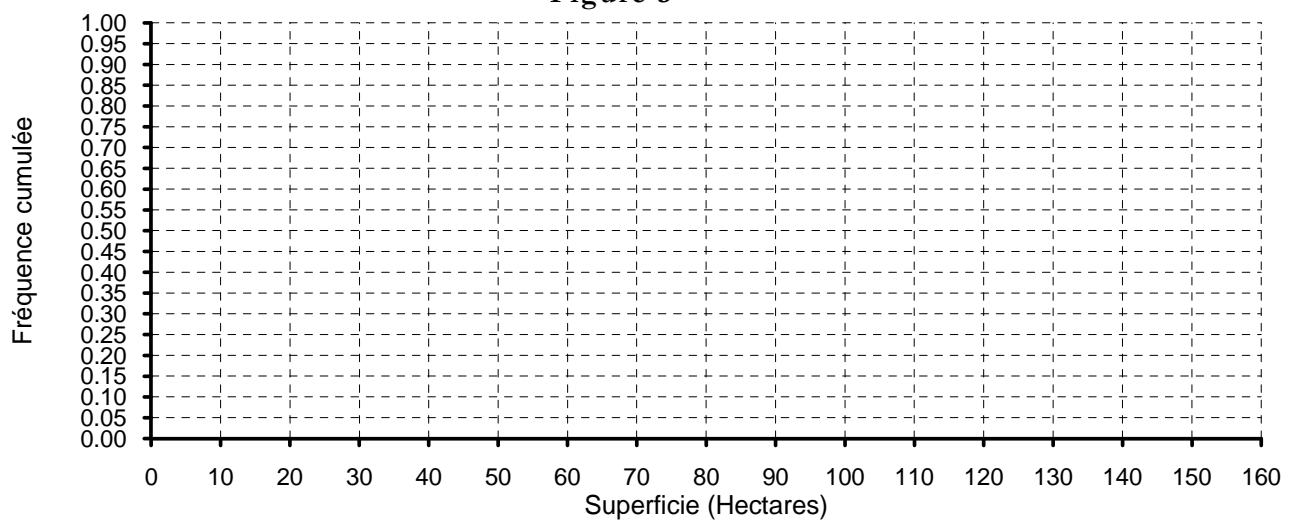


Figure 4

Corrigé du TD n°2

Exercice n°1

1°) Détermination graphique du mode. Il y'a deux méthodes pour déterminer graphiquement le mode

Méthode 1 : Centre de la classe modale (Figure 1)

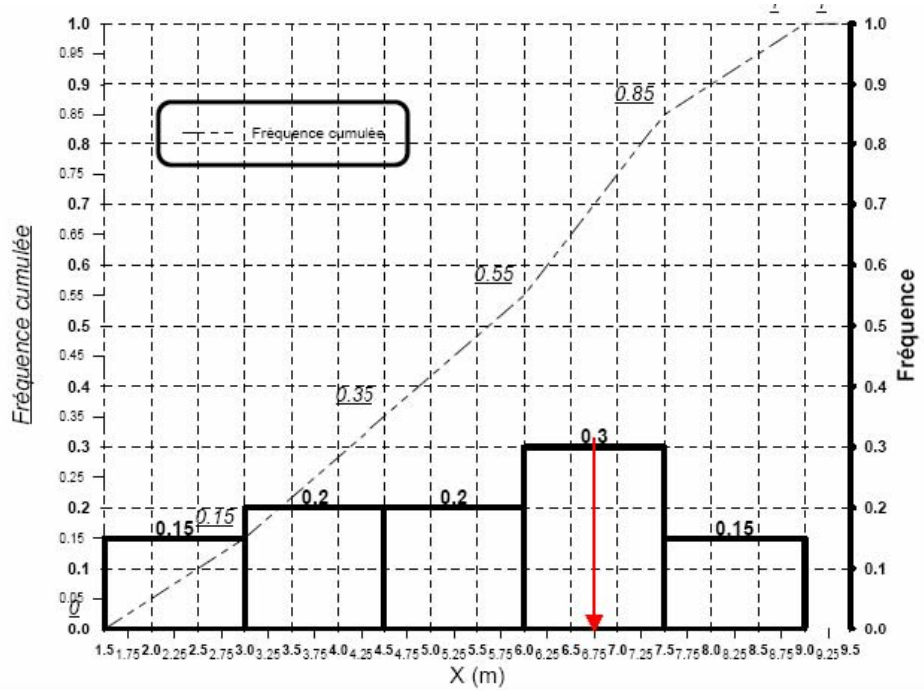


Figure 1

La classe modale est $[6.0, 7.5[$ (m), c'est à dire la classe n°4. La première méthode pour le calcul du mode est de prendre le centre de la classe n°4, de ce fait $M_0 = 6.75$ m

Méthode 2 : En prenant en compte les classes adjacentes (Figure 2)

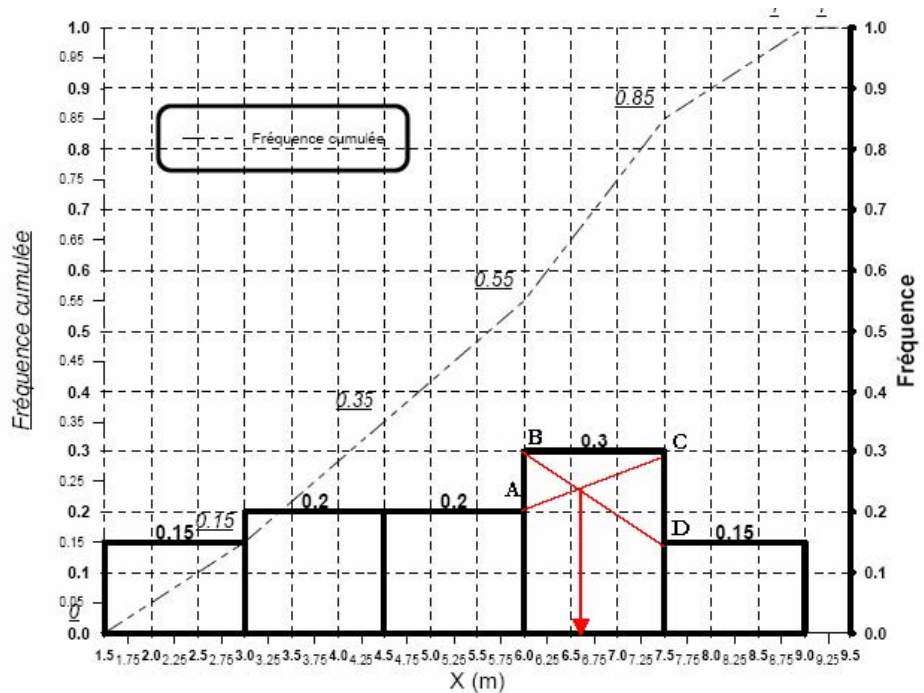


Figure 2

Dans ce cas, il faut joindre les droites (AC) et (BD) et déterminer le point d'intersection de ces deux droites. L'abscisse de ce point donne le mode. Dans le présent cas, $M_o=6.60$ m

2°) Détermination graphique de la médiane

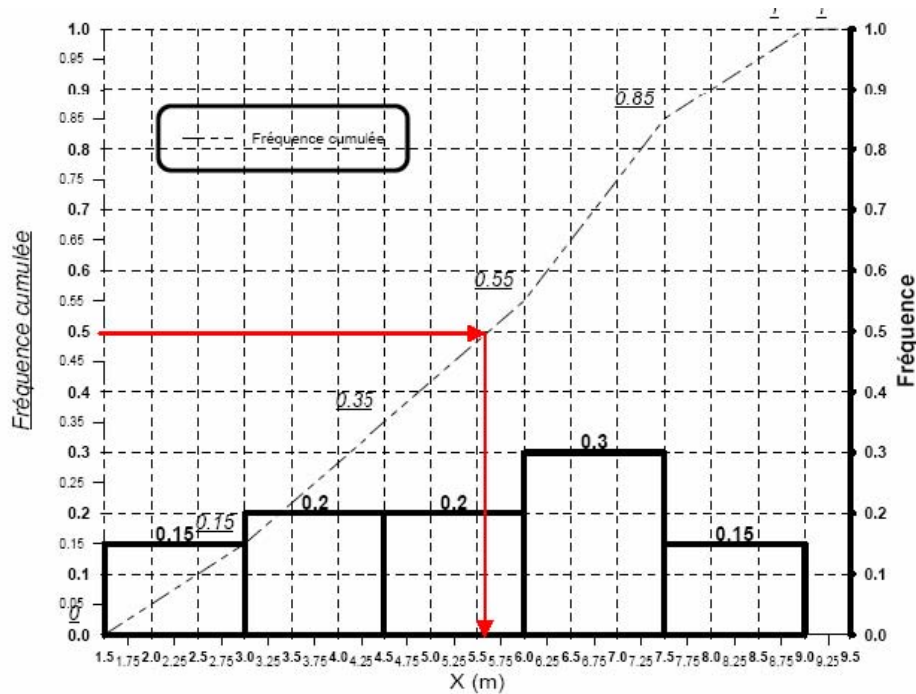


Figure 3

La médiane est l'abscisse de la fonction cumulative pour une fréquence cumulée de 50%. D'après le graphe (Figure 3) $M_e=5.6$ m

3°) Détermination de la moyenne

Pour le calcul de la moyenne, il faut établir le tableau statistique

(1)	(2)	(3)	(4)	(5)	(6)
N° de la classe	Valeur min de la classe (m)	Valeur max de la classe (m)	Fréquence f_k	Centre de classe $X_k=[(2)+(3)]/2$	(4) X (5)
1	1,5	3	0,15	2,25	0,3375
2	3	4,5	0,2	3,75	0,75
3	4,5	6	0,2	5,25	1,05
4	6	7,5	0,3	6,75	2,025
5	7,5	9	0,15	8,25	1,2375
		Somme des fréquences=	1	Moyenne=	5,4

La moyenne $\bar{X} = 5,4$ m.

Récapitulatif :

$M_o=6.6$ m, $M_e=5.6$ m et $\bar{X} = 5,4$ m.

Exercice2 : Principe du juste-à-temps

1. Définir la population, l'individu, le caractère et les modalités de cette étude.

Population 12 Pièces mécaniques
Individu Une pièce parmi les 12 pièces
Caractère Temps d'attente mesurée en min-dec. C'est le temps qui s'écoule entre la sortie de la pièce mécanique de l'appareil A et l'entrée de la pièce mécanique dans l'appareil B. Ce caractère sera noté ci-après **T**.
Modalités Valeurs de ce temps d'attente comprises entre 1 et 2.9 min-dec.
 NB : 2.90 min-dec équivaut à 2minute et 0.90x60s=54 s soit 2 minutes et 54 secondes.

2. Etablir le tableau statistique

- Le nombre d'individus $n=12$. Les valeurs observées du caractère ont pour valeurs maximale et minimale $T_{min}=1$ min-dec, $T_{max}=2.9$ min-dec
- Détermination du nombre de pas k par la méthode de Sturge :

$$k = 1 + 3.3 \log_{10}(12) = 4.56$$
 soit $k=5$
- Calcul du pas

$$Pas > \frac{T_{max} - T_{min}}{k} = \frac{2.9 - 1.0}{5} > 0.38$$
 min-dec

On peut prendre $Pas=0.39$ min-dec ou $Pas=0.40$ min-dec. Considérons un pas de 0.40 min-dec.

- Vérification de la valeur maximale

$$T_{min} + Pas * k > T_{max}, 1 + 0.40 * 5 = 3 > 2.9 \rightarrow \text{C'est vérifié}$$

Classe (min-dec)		Centre de classe (min-dec)	Effectif	Fréquence (en %)	N° de la classe (j=1,..,5)
V_{min}	V_{max}	T_j	n_j	f_j	j
1	1,4	1,2	1	08,33	1
1,4	1,8	1,6	5	41,67	2
1,8	2,2	2	3	25,00	3
2,2	2,6	2,4	2	16,67	4
2,6	3	2,8	1	08,33	5

3. Tracer l'histogramme et la fonction cumulative.

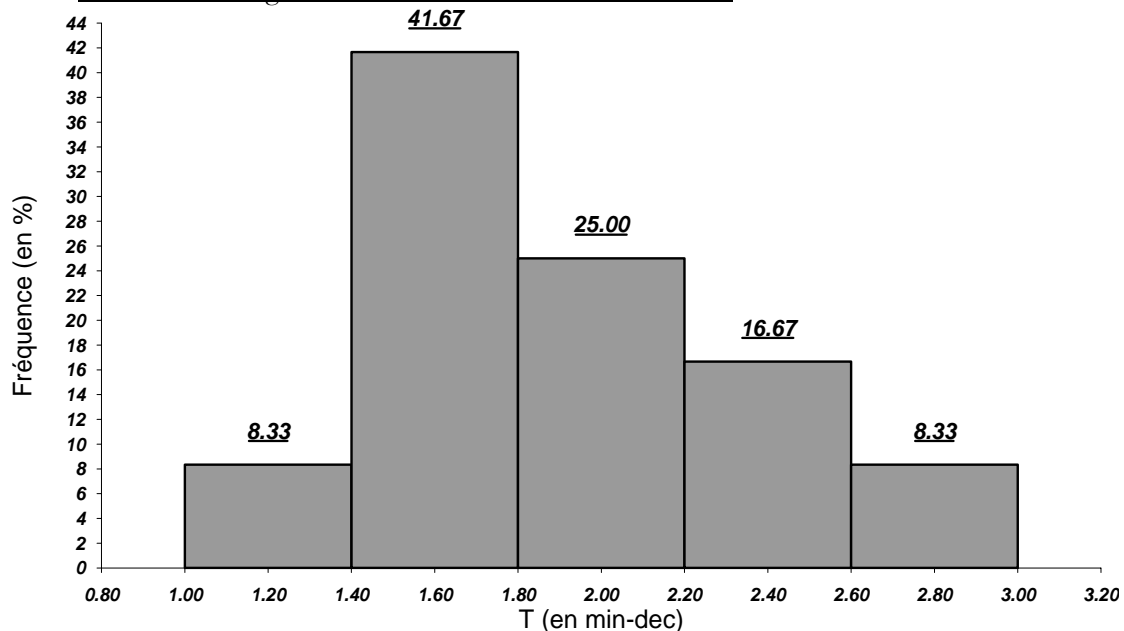


Figure 1 Histogramme

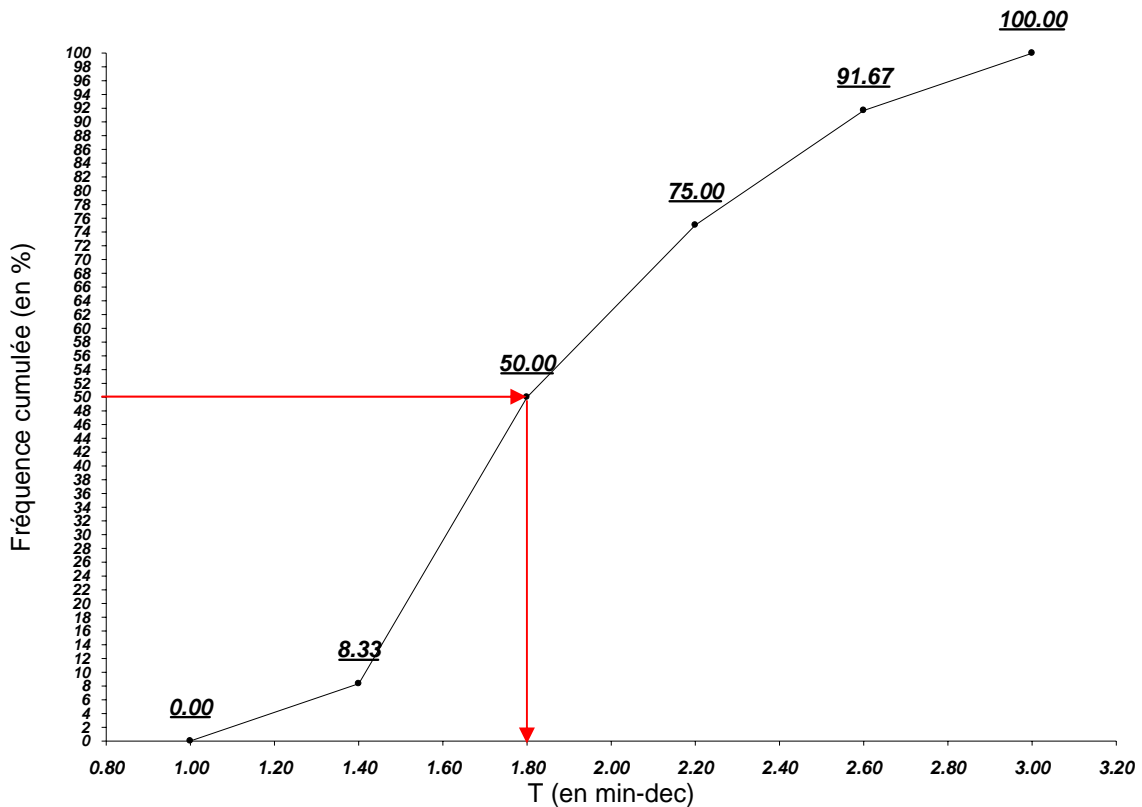


Figure 2 *Fonction cumulative*

4. Déterminer graphiquement la médiane (Figure 2)

Sur la Figure 2, la fonction cumulative a été tracée. La médiane, notée M_e , correspond à la fréquence cumulée de 50%. C'est-à-dire qu'il y a 50% des individus (ce sont les pièces) dont la modalité est inférieure à M_e .

Graphiquement on remarque que

$$M_e = 1.80 \text{ min-dec}$$

5. Calculer les moyennes arithmétique, géométrique, harmonique et quadratique

5.1 Moyenne harmonique : $\bar{T}_{-1} = \left(\sum [f_k (T_k)^{-1}] \right)^{-1}$ soit $\bar{T}_{-1} = \frac{1}{\sum \left[f_k \frac{1}{T_k} \right]}$

Vmin	Vmax	Centre de classe T_k	Fréquence f_k	$f_k \frac{1}{T_k}$
(1)	(2)	(3)	(4)	(5)=(4)/(3)
1	1,4	1,2	0,08	0,06944444
1,4	1,8	1,6	0,42	0,26041667
1,8	2,2	2	0,25	0,125
2,2	2,6	2,4	0,17	0,06944444
2,6	3	2,8	0,08	0,0297619
Somme = $\sum \left[f_k \frac{1}{T_k} \right] =$				0,55406746
Moyenne harmonique = 1/Somme				1,80483438

$$\bar{T}_{-1} = 1.80 \text{ min-dec}$$

5.2 Moyenne géométrique : $\bar{T}_0 = \left(\sum [f_k (T_k)^0] \right)^{\frac{1}{p}}$, soit $\log(\bar{T}_0) = \frac{\sum [f_k \log(T_k)]}{n}$

Vmin	Vmax	Centre de classe T_k	Fréquence f_k	$f_k \log(T_k)$
(1)	(2)	(3)	(4)	(5)=(4)*log[(3)]
1	1,4	1,2	0,08	0,01519346
1,4	1,8	1,6	0,42	0,19583485
1,8	2,2	2	0,25	0,1732868
2,2	2,6	2,4	0,17	0,14591146
2,6	3	2,8	0,08	0,08580162
Somme= $\sum [f_k \log(T_k)]$				0,61602818
Moyenne géométrique=exp(Somme)				1,85155935

$$\bar{T}_0 = 1.86 \text{ min-dec}$$

5.3 Moyenne arithmétique $\bar{T}_1 = \left(\sum [f_k (T_k)^1] \right)^{\frac{1}{n}}$, soit $\bar{T}_1 = \frac{\sum [f_k T_k]}{n}$

Vmin	Vmax	Centre de classe T_k	Fréquence f_k	$f_k T_k$
(1)	(2)	(3)	(4)	(5)=(4)*(3)
1	1,4	1,2	0,08	0,1
1,4	1,8	1,6	0,42	0,66666667
1,8	2,2	2	0,25	0,5
2,2	2,6	2,4	0,17	0,4
2,6	3	2,8	0,08	0,23333333
Moyenne arithmétique= $\frac{\sum [f_k T_k]}{n}$				1,9

$$\bar{T}_1 = 1.90 \text{ min-dec}$$

5.4 Moyenne quadratique $\bar{T}_2 = \left(\frac{\sum [f_k (T_k)^2]}{n} \right)^{\frac{1}{2}}$, soit $\bar{T}_2 = \sqrt{\frac{\sum [f_k (T_k)^2]}{n}}$

Vmin	Vmax	Centre de classe T_k	Fréquence f_k	$f_k (T_k)^2$
(1)	(2)	(3)	(4)	(5)=(4)*(3)*(3)
1	1,4	1,2	0,08	0,12
1,4	1,8	1,6	0,42	1,06666667
1,8	2,2	2	0,25	1
2,2	2,6	2,4	0,17	0,96
2,6	3	2,8	0,08	0,65333333
Somme= $\sum [f_k (T_k)^2]$				3,8
Moyenne arithmétique= $\sqrt{\frac{\text{Somme}}{n}}$				1,94935887

$$\bar{T}_2 = 1.95 \text{ min-dec}$$

6. Déterminer graphiquement (Figure 2) et analytiquement le mode

Graphiquement :

Sur la figure 3, l'histogramme de la distribution est schématisé. Le mode est l'intersection des droites AD et BC. Graphiquement, on note que le mode est égale à $\hat{\lambda} = 1.65 \text{ min-dec}$. Toutefois, ceci n'est qu'une appréciation graphique.

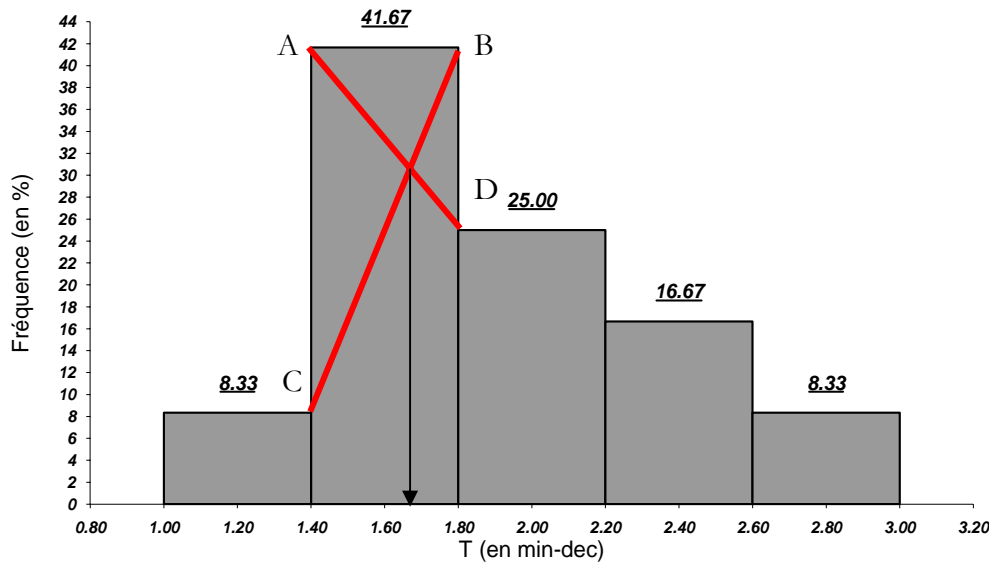


Figure 3 Détermination graphique du mode

Analytiquement le mode est défini comme suit :

$$M_o = x_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} (x_{i+1} - x_i), \text{ Avec } \Delta_1 = 41.67\% - 8.33\% = 33.34\%, \Delta_2 = 41.67\% -$$

$$25.00\% = 16.67\%$$

$$x_i \equiv x_1 = 1.40 \text{ min-dec et } x_{i+1} \equiv x_2 = 1.80 \text{ min-dec.}$$

$$\text{Soit } M_o = 1.40 + \frac{33.34}{33.34 + 16.67} (1.80 - 1.40) = 1.67 \text{ min-dec}$$

$$M_o = 1.67$$

min-dec

7. Classer les moyennes, le mode et la médiane. Conclusion

Mode	<	Moyenne harmonique et médiane	<	Moyenne géométrique et moyenne arithmétique tirée de la série statistique	<	Moyenne arithmétique	<	Moyenne quadratique
M_o	<	\bar{T}_{-1} et M_e	<	\bar{T}_0 et \bar{T}	<	\bar{T}_1	<	\bar{T}_2
1.67	<	1.80	<	1.86	<	1.90	<	1.95

Noter que $\bar{T} = (2.2 + 2.1 + 1.78 + 1.42 + 1.40 + 1.82 + 2 + 1 + 2.9 + 2.3 + 1.7 + 1.6) / 12 = 1.86$ min-dec. \bar{T} est la moyenne arithmétique tirée de la série statistique.

Globalement on remarque que le mode et la moyenne quadratique sont très éloignés des autres caractéristiques de tendance centrale. De plus on note que les moyennes \bar{T}_0 et \bar{T} d'une part et \bar{T}_1 d'autre part sont très proches.

8. Si la moyenne du temps d'attente dépasse 2 min alors un autre appareil de type B doit être rajouté. Pour le présent est il obligatoire de rajouter un autre appareil de type B ?

La moyenne arithmétique étant égale à 1.90 min-dec soit 1 min et 54 secondes. Cette moyenne étant inférieure à 2 minutes, de ce fait, il n'est pas obligatoire de rajouter un autre appareil de type B.

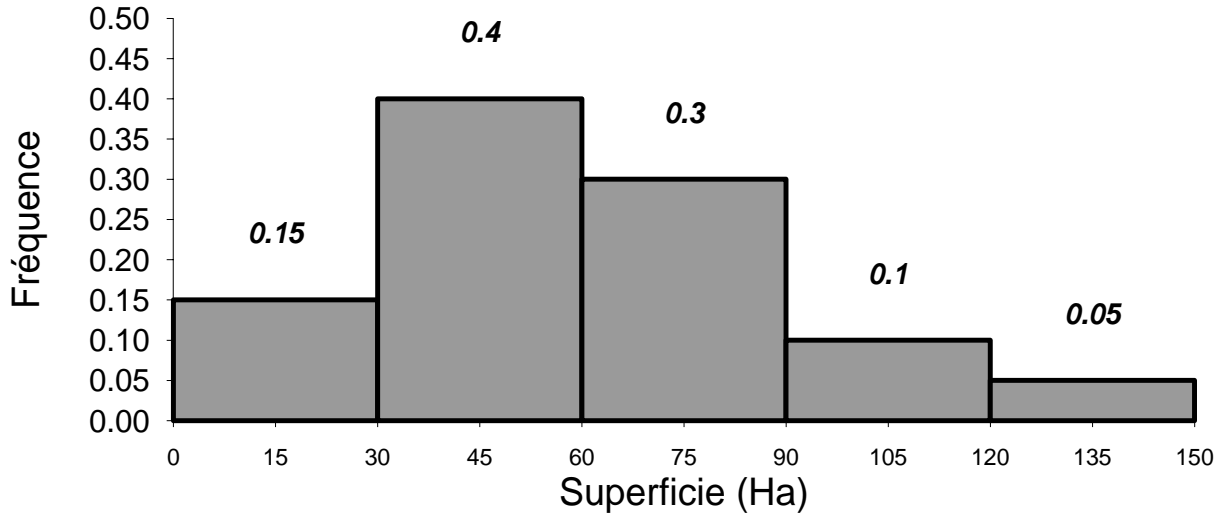
Exercice3 :

1. *Tracer l'histogramme de cette distribution (Figure 3)*

Pour tracer l'histogramme, il faut tout d'abord établir le tableau statistique

N° de la classe j	Intervalle des classes		Centre de classe s_j (en Ha)	Effectif n_j	Fréquence f_j
	Taille min (en Ha)	Taille max (en Ha)			
1	0	30	15	15	0,15
2	30	60	45	40	0,4
3	60	90	75	30	0,3
4	90	120	105	10	0,1
5	120	150	135	5	0,05

Le nombre de classe est de $k=5$



2. *Tracer la fonction cumulative (Figure 4)*

Pour tracer la fonction cumulative, il faut tout d'abord calculer les fréquences cumulées (voir tableau ci-dessous)

Valeurs (Ha)	Fréquence cumulée
0	0
30	0.15
60	0.55
90	0.85
120	0.95
150	1.00

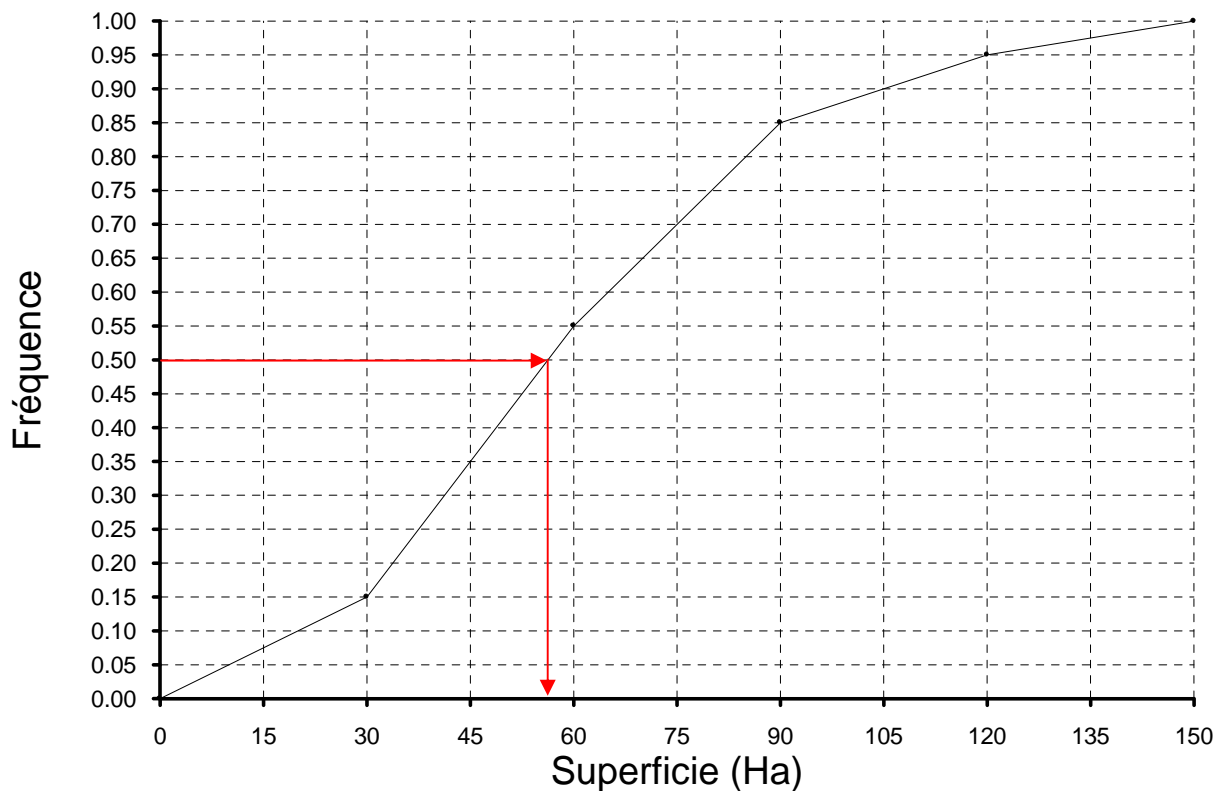


Figure 4 Fréquence cumulée

3. Déterminer graphiquement le mode (Figure 3) et la médiane (Figure 4)

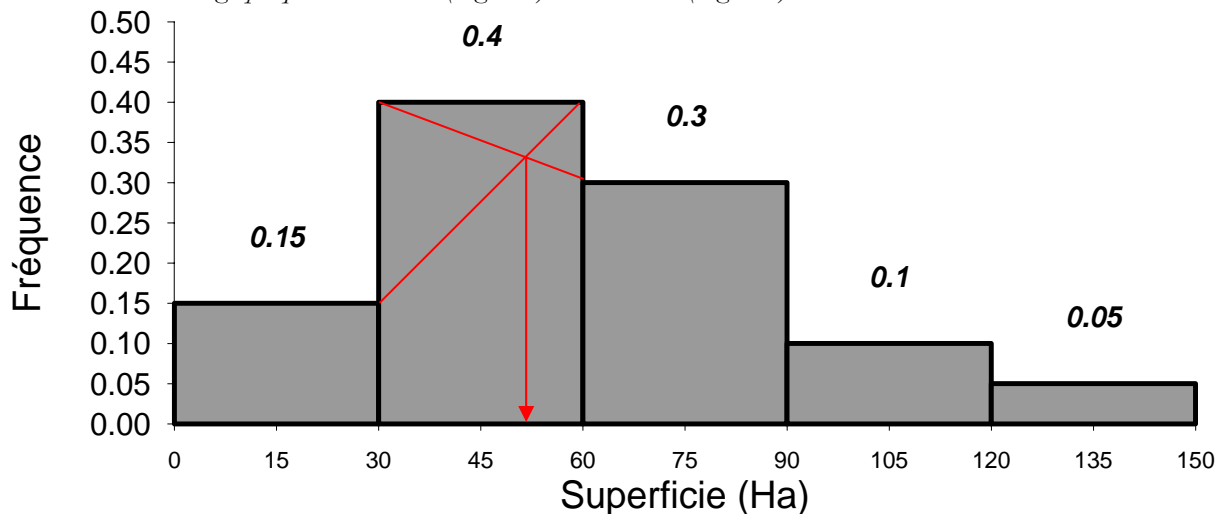


Figure 3' Histogramme de la distribution

Sur la Figure 3', on remarque le mode est compris entre 45 et 60 Ha, il est pratiquement légèrement inférieur à 52.5 Ha (milieu entre 45 et 60 Ha). On peut donc considérer que **Mo=52 Ha**.

Sur la Figure 4, on remarque que la médiane est comprise entre 45 et 60 HA. Graphiquement la médiane est égale à **Me=56 Ha**

NB : Ces deux résultats ne sont que l'interprétation visuelle des deux figures. L'erreur est donc possible.

4. Sachant que la moyenne nationale est d'environ 50 hectares, pensez-vous que cette wilaya est une wilaya de petites exploitations ?

La médiane et le mode sont des caractéristiques de tendance centrale, ils représentent donc l'ensemble des valeurs de la série ou du tableau statistique et ont un poids similaire à celui de la moyenne.

La moyenne nationale, de 50 Ha, étant inférieure au mode ou à la médiane, donc on peut dire que cette wilaya n'est pas une wilaya de petite exploitation.

5. Calculer la moyenne et l'écart type.

La moyenne tirée du tableau statistique est calculée par $\bar{S} = \sum_{j=1}^k f_j s_j$,

où $j=1, \dots, k$

f_j est la fréquence de la classe j et s_j est le centre de la classe j

L'écart type est calculé par
$$\sigma_S = \sqrt{\sum_{j=1}^k f_j (s_j - \bar{S})^2} \quad (\text{Equation 1})$$

ou bien par
$$\sigma_S = \sqrt{\sum_{j=1}^k f_j (s_j)^2 - \bar{S}^2} \quad (\text{Equation 2})$$

Pour tracer l'histogramme, il faut tout d'abord établir le tableau statistique

N° de la classe j	Intervalle des classes		Centre de classe s_j (en Ha)	Effectif n_j	Fréquence f_j	$f_j s_j$ (1)	$f_j (s_j - \bar{S})^2$ (2)	$f_j (s_j)^2$ (3)
	Taille min (en Ha)	Taille max (en Ha)						
1	0	30	15	15	0,15	2,25	303,75	33,75
2	30	60	45	40	0,4	18	90	810
3	60	90	75	30	0,3	22,5	67,5	1687,5
4	90	120	105	10	0,1	10,5	202,5	1102,5
5	120	150	135	5	0,05	6,75	281,25	911,25

Pour le calcul de la moyenne il faut juste sommer les valeurs de la colonne (1), on trouve $\bar{S} = 60$ Ha.

Pour le calcul de l'écart type, on peut utiliser l'une des deux formules décrites plus haut :

1^{ère} méthode : Il faut sommer les valeurs de la colonne (2) pour trouver que

$$\sum_{j=1}^k f_j (s_j - \bar{S})^2 = 945 \text{ (Ha)}^2, \text{ donc } \sigma_S = \sqrt{945} \quad \sigma_S = 30.74 \text{ Ha}$$

2^{ème} méthode : Il faut sommer les valeurs de la colonne (3) pour trouver que

$$\sum_{j=1}^k f_j (s_j)^2 = 4545 \text{ (Ha)}^2, \text{ donc } \sigma_S = \sqrt{4545 - 60^2} = \sqrt{945} \quad \sigma_S = 30.74 \text{ Ha}$$

Les deux méthodes donnent le même résultat

TD n°3

Exercice : 1

La prévision météorologique est une science en pleine évolution. Elle a pour objectif de prédire un ensemble de paramètres comme la température, la pression, la pluviométrie etc... L'objectif du présent exercice est la recherche d'une corrélation entre deux paramètres à savoir la température (notée T et mesurée en degré Celsius) et la pluviométrie (notée P et mesurée en mm). Le tableau ci-dessous donne les moyennes par mois de ces deux paramètres.

Mois	Septembre	Octobre	Novembre	Décembre	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août
Pluviométrie (mm)	13	23	49	49	50	64	79	48	40	10	5	6
Température (°C)	23	17	14	10	10	11	13	15	17	23	27	28

1. Identifier Population, individu, caractères et modalités
2. Calculer la moyenne et l'écart type de P et T
3. En supposant que la relation entre P et T est de type linéaire déterminer la droite de régression entre ces deux paramètres.
4. Calculer le coefficient de corrélation. Conclusion
5. Refaire les questions (2), (3) et (4) en utilisant le tableau statistique.

Corrigé du TD n°3

Exercice1 :

Corrigé avec la série statistique

1. Identifier Population, individu, caractères et modalités

Population Les 12 mois de l'année
Individu Un mois parmi ces mois
Caractère Il y a deux caractères : Pluviométrie (mesurée en mm et noté P) et la Température (mesurée en °C et noté T)
Modalités Pour le caractère Pluviométrie : Les modalités varient de 5 à 79 mm
 Pour la caractère Température : Les modalités varient de 10 à 28°C

2. Calculer la moyenne et l'écart type de P et T

Les moyennes et écart types de P et T vont être directement calculés à partir de la série statistique

Individu	P (mm) (1)	T (°C) (2)	$(P_i - \bar{P})^2$ (3)	$(T_i - \bar{T})^2$ (4)
1	13	23	544,4444444	32,11111111
2	23	17	177,7777778	0,11111111
3	49	14	160,4444444	11,11111111
4	49	10	160,4444444	53,77777778
5	50	10	186,7777778	53,77777778
6	64	11	765,4444444	40,11111111
7	79	13	1820,444444	18,77777778
8	48	15	136,1111111	5,444444444
9	40	17	13,44444444	0,11111111
10	10	23	693,4444444	32,11111111
11	5	27	981,7777778	93,44444444
12	6	28	920,1111111	113,7777778
	$\bar{P} = 36,33$ mm	$\bar{T} = 17,33$ °C	$546,7222222$ (mm) ²	$37,88888889$ (°C) ²
			$\sigma_P = 23,38209191$ mm	$\sigma_T = 6,155395104$ °C

3. En supposant que la relation entre P et T est de type linéaire déterminer la droite de régression entre ces deux paramètres.

L'équation de la droite de régression entre les deux variables P et T s'écrit comme suit :

$$T = \bar{T} - \frac{\text{cov}(P, T)}{\sigma_P^2} \bar{P} + \frac{\text{cov}(P, T)}{\sigma_P^2} P$$

avec $\text{cov}(P, T) = \frac{1}{n} \sum_{i=1}^n (P_i - \bar{P})(T_i - \bar{T})$ qui peut s'écrire aussi $\text{cov}(P, T) = \frac{1}{n} \sum_{i=1}^n P_i T_i - \bar{T} \cdot \bar{P}$

Noter que le point (\bar{P}, \bar{T}) appartient à cette droite.

Individu	P (mm) (1)	T (°C) (2)	$(P_i - \bar{P})(T_i - \bar{T})$ (3)	$P_i T_i$ (4)
1	13	23	-132,22	299
2	23	17	4,44	391
3	49	14	-42,22	686
4	49	10	-92,89	490
5	50	10	-100,22	500
6	64	11	-175,22	704
7	79	13	-184,89	1027
8	48	15	-27,22	720
9	40	17	-1,22	680
10	10	23	-149,22	230
11	5	27	-302,89	135
12	6	28	-323,56	168
	$\bar{P} = 36,33$ mm	$\bar{T} = 17,33$ °C	$\sum_{i=1}^n (P_i - \bar{P})(T_i - \bar{T}) = -1527,33$	$\sum_{i=1}^n P_i T_i = 60305$
			$\text{cov}(P, T) = -1527,33/12$ $\text{cov}(P, T) = -127,28$ mm.°C	$\text{cov}(P, T) = (1/12)*60305 - 17,33*36,33$ $\text{cov}(P, T) = -127,28$ mm.°C

La covariance $\text{cov}(P, T)$ peut être calculée soit par

L'Equation
$$\text{cov}(P, T) = \frac{1}{n} \sum_{i=1}^n (P_i - \bar{P})(T_i - \bar{T})$$
 (Voir Colonne N°3 du tableau ci-dessus)

Ou bien par

L'Equation
$$\text{cov}(P, T) = \frac{1}{n} \sum_{i=1}^n P_i T_i - \bar{T} \cdot \bar{P}$$
 (Voir Colonne N°4 du tableau ci-dessus)

$$T = \bar{T} - \frac{\text{cov}(P, T)}{\sigma_P^2} \bar{P} + \frac{\text{cov}(P, T)}{\sigma_P^2} P = 17,33 - \frac{(-127,28)}{546,73} 36,33 + \frac{(-127,28)}{546,73} P$$

Soit : $T = 25,78 - 0,23P$

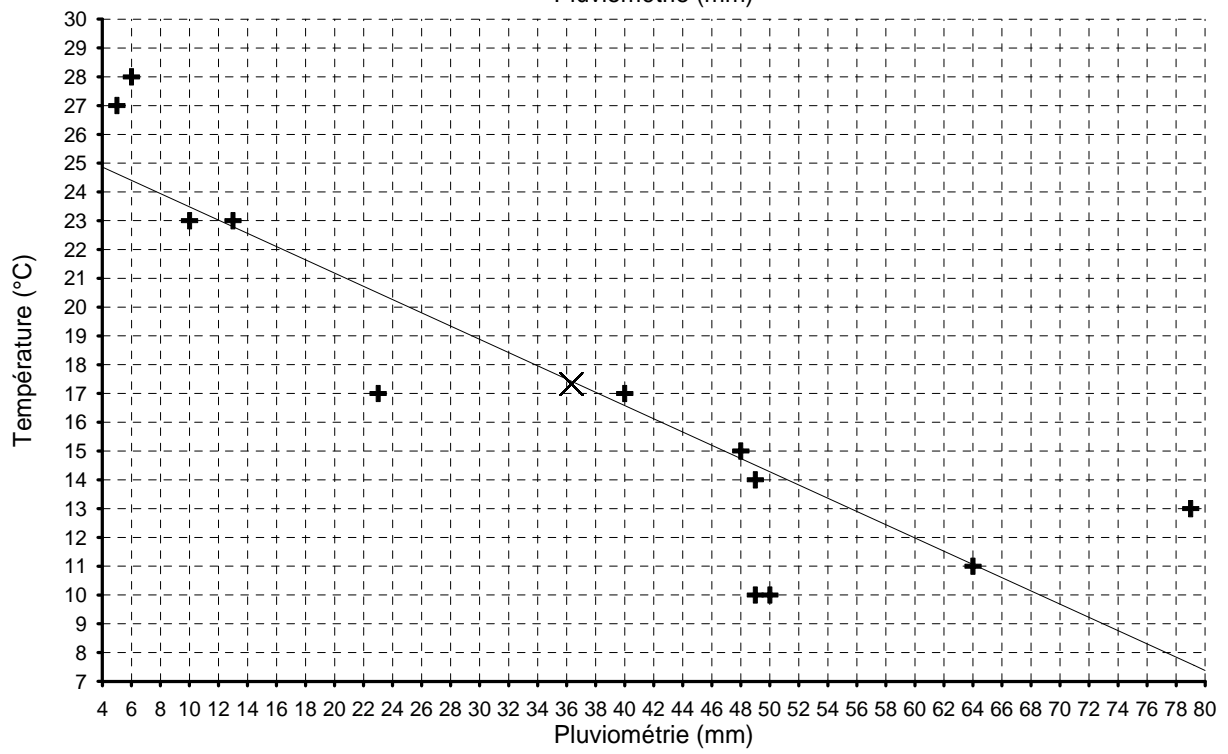
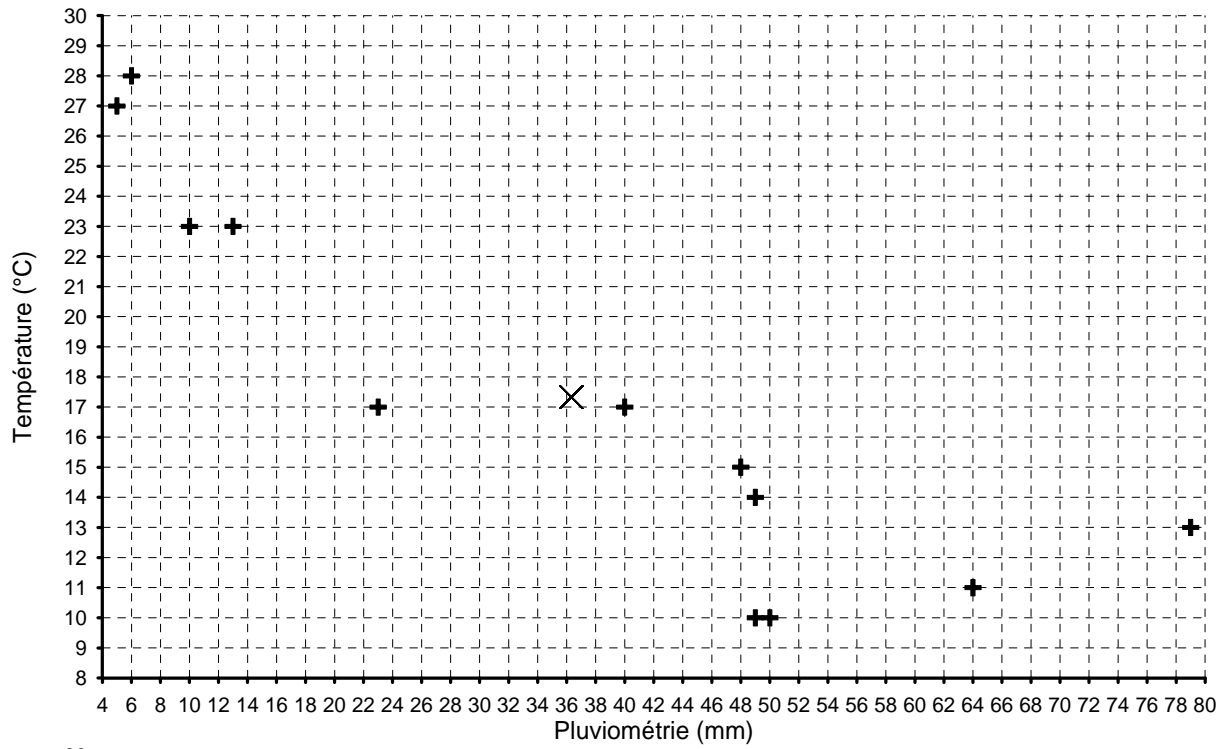
En particulier lorsque pour $P = \bar{P} = 36,33$ mm on trouve $P = 17,42$ °C qui est très proche de 17,33°C

4. Calculer le coefficient de corrélation. Conclusion

Le coefficient de corrélation permet de mesurer le degré de corrélation entre deux variables.

$$r = \frac{\text{cov}(P, T)}{\sigma_P \sigma_T} = \frac{-127,28}{23,38 \cdot 6,15} = -0,88$$

Du moment que $|r| = 0,88 > 0,7$ alors on peut considérer que la relation trouvée entre P et T est bonne.



Corrigé du TD n°3

Exercice 1 :

Corrigé avec le tableau statistique

1. Identifier Population, individu, caractères et modalités

- Population** Les 12 mois de l'année
Individu Un mois parmi ces mois
Caractère Il y a deux caractères : Pluviométrie (mesurée en mm et noté P) et la Température (mesurée en °C et noté T)
Modalités Pour le caractère Pluviométrie : Les modalités varient de 5 à 79 mm
 Pour la caractère Température : Les modalités varient de 10 à 28°C

Afin de répondre aux questions (2), (3) et (4), il faut établir le tableau statistique

2. Etablir le tableau statistique pour le caractère T et P :

Caractère P

- Le nombre d'individus $n=12$. Les valeurs observées du caractère P ont pour valeurs maximale et minimale $P_{min}=5$ mm, $P_{max}=79$ mm
- Détermination du nombre de pas k par la méthode de Sturge :
 $k=1+3.3\log_{10}(12)=4.56$ soit $k=5$
- Calcul du pas $Pas > \frac{P_{max} - P_{min}}{k} = \frac{79-5}{5} = 14.8$, soit Pas=15 mm
- Vérification de la valeur maximale
 $P_{min} + Pas * k > P_{max}, 5 + 15 * 5 = 80 > 79 \rightarrow$ C'est vérifié

Caractère T

- Le nombre d'individus $n=12$. Les valeurs observées du caractère ont pour valeurs maximale et minimale $T_{min}=10$ °C, $T_{max}=28$ °C
- Détermination du nombre de pas k par la méthode de Sturge :
 $k=1+3.3\log_{10}(12)=4.56$ soit $k=5$
- Calcul du pas $Pas > \frac{T_{max} - T_{min}}{k} = \frac{28-10}{5} = 3.6$ soit Pas=4 °C
- Vérification de la valeur maximale
 $T_{min} + Pas * k > T_{max}, 10 + 4 * 5 = 30 > 28 \rightarrow$ C'est vérifié

Mois	Septembre	Octobre	Novembre	Décembre	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août
Pluviométrie (mm)	13	23	49	49	50	64	79	48	40	10	5	6
Température (°C)	23	17	14	10	10	11	13	15	17	23	27	28

T (°C) \ P (mm)		[10,14[[14,18[[18,22[[22,26[[26,30[
		12	16	20	24	28	
[5,20[12.5	0	0	0	2	2	4
[20,35[27.5	0	1	0	0	0	1
[35,50[42.5	1	3	0	0	0	4
[50,65[57.5	2	0	0	0	0	2
[65,80[72.5	1	0	0	0	0	1
		4	4	0	2	2	12

Ensuite nous allons calculer les fréquences conjointes f_{ij} : où $(i,j)=1,2,\dots,5$. Par exemple, f_{11} est la fréquence des individus qui ont la modalité 1 (c'est-à-dire [5,20]) de P et la modalité 1 (c'est-à-dire [10,14]) de T $f_{11}=0$. Autre exemple $f_{14} = \frac{2}{12}$ soit 0.17.

T °C \ P mm		12	16	20	24	28	
		12.5	0	0	0	0.17	0.17
27.5	0	0.08	0	0	0	0.08	
42.5	0.08	0.25	0	0	0	0.33	
57.5	0.17	0	0	0	0	0.17	
72.5	0.08	0	0	0	0	0.08	
	0.33	0.33	0	0.17	0.17	1	

3. Calculer la moyenne et l'écart type de P et T

$$\bar{T}_0 = \sum_{j=1}^5 f_j T_j = 0.33*12 + 0.33*16 + 0.17*24 + 0.17*28 = 18,08 \text{ } ^\circ\text{C}$$

$$\bar{P}_0 = \sum_{i=1}^5 f_i P_i = 0.34*12.5 + 0.08*27.5 + 0.33*42.5 + 0.17*57.5 + 0.08*72.5 = 36.05 \text{ mm}$$

$$\sigma_T = \sqrt{\sum_{j=1}^5 f_j (T_j - \bar{T}_0)^2} = \sqrt{\sum_{j=1}^5 f_j T_j^2 - \bar{T}_0^2}$$

$$\sigma_T^2 = (0.33*12^2 + 0.33*16^2 + 0.17*24^2 + 0.17*28^2) - 18.08^2 = 36.3136 \text{ } (^\circ\text{C})^2 \text{ soit } \sigma_T = 6,03 \text{ } ^\circ\text{C}$$

$$\sigma_P = \sqrt{\sum_{i=1}^5 f_i (P_i - \bar{P}_0)^2} = \sqrt{\sum_{i=1}^5 f_i P_i^2 - \bar{P}_0^2}$$

$$\sigma_P^2 = (0.34*12.5^2 + 0.08*27.5^2 + 0.33*42.5^2 + 0.17*57.5^2 + 0.08*72.5^2) - 36.05^2 = 392,6475 \text{ (mm)}^2 \text{ soit } \sigma_P = 19,82 \text{ mm.}$$

4. En supposant que la relation entre P et T est de type linéaire déterminer la droite de régression entre ces deux paramètres.

L'équation de la droite de régression entre les deux variables P et T s'écrit comme suit :

$$T = \bar{T} - \frac{\text{cov}(P,T)}{\sigma_P^2} \bar{P} + \frac{\text{cov}(P,T)}{\sigma_P^2} P$$

$$\text{Avec } \text{Cov}(P,T) = \sum_{i=1}^{i=k} \sum_{j=1}^{j=k} f_{ij} (P_i - \bar{P}_0)(T_j - \bar{T}_0) \text{ ou bien } \text{Cov}(P,T) = \sum_{i=1}^{i=k} \sum_{j=1}^{j=k} f_{ij} P_i T_j - \bar{P}_0 \bar{T}_0$$

D'après le tableau statistique on obtient :

$$\text{Cov}(P,T) = (0.17*12.5*24 + 0.17*12.5*28 + 0.08*27.5*16 + 0.08*42.5*12 + 0.25*42.5*16 + 0.17*57.5*12 + 0.08*72.5*12) - 18.08*36.05 = -108,384 \text{ } ^\circ\text{C.mm}$$

Donc la droite de régression s'écrit :

$$T = \bar{T} - \frac{\text{cov}(P,T)}{\sigma_P^2} \bar{P} + \frac{\text{cov}(P,T)}{\sigma_P^2} P = 18.08 - \frac{-108.384}{19,82^2} 36.05 + \frac{-108.384}{19,82^2} P$$

$$\text{Soit } T = 28.03 - 0.276P$$

En remplaçant P par $\bar{P}_0 = 36.05 \text{ mm}$, on trouve $T = 18.0802 \text{ } ^\circ\text{C}$ soit à peu près $\bar{T}_0 = 18.08 \text{ } ^\circ\text{C}$

5. Calculer le coefficient de corrélation. Conclusion

Le coefficient de corrélation permet de mesurer le degré de corrélation entre deux variables.

$$r = \frac{\text{cov}(P,T)}{\sigma_P \sigma_T} = \frac{-108.384}{19.82 * 6.03} = -0.906. \text{ Du moment que } |r| = 0.906 > 0.7 \text{ alors on peut}$$

considérer que la relation trouvée entre P et T est bonne.

Il a été observé que les séismes majeurs, donc destructeurs, sont précédés par un certain nombre de phénomènes, dits événements précurseurs, tels que le comportement anormal des animaux, les anomalies géophysiques, variation du niveau de la nappe phréatique. Il est donc, théoriquement, possible de prédire l'occurrence des séismes en utilisant ces informations et ce grâce à la théorie de probabilité. En terme de probabilité, les différents cas de figure s'expriment comme suit :

$$P_{1|1} = \text{Probabilité [séisme se produit sachant qu'il a été prédit]} \quad (1)$$

$$P_{0|1} = \text{Probabilité [séisme ne se produit pas sachant qu'il a été prédit]} \quad (2)$$

$$P_{0|0} = \text{Probabilité [séisme ne se produit pas sachant qu'il n'a été prédit]} \quad (3)$$

$$P_{1|0} = \text{Probabilité [séisme se produit sachant qu'il n'a été prédit]} \quad (4)$$

Il a été tout à fait clair que

$$P_{1|1} = 1 - P_{0|1} \quad (5)$$

$$P_{0|0} = 1 - P_{1|0} \quad (6)$$

Le défi majeur dans l'utilisation de cette approche est la fiabilité du modèle mis en place dans le sens où si le modèle prédit l'occurrence du séisme alors ce séisme doit se produire et inversement s'il ne le prédit pas alors obligatoirement le séisme ne doit pas avoir lieu. En d'autres termes, ce système de prédiction est considéré comme fiable si :

$$P_{1|1} = 1 \quad (7)$$

$$P_{0|0} = 1 \quad (8)$$

Si $P_{0|1}$ est grande alors il y aura de fausses alertes et personne ne va vous croire car vous êtes entrain de prédire un séisme mais il ne se produit pas (à la limite vous allez être viré de votre job). Par contre si $P_{1|0}$ est grande alors un autre problème se pose, le séisme majeur s'est produit et a probablement tout détruit, peut être qu'il n'y aura personne pour vous juger car votre système n'a rien prédit sachant que le séisme a eu lieu.

Les Equations (7) et (8) sont généralement très difficile à obtenir. Considérons les événements suivants :

E = séisme se produit dans une journée

E^c = séisme ne se produit pas dans une journée³.

Les deux événements sont considérés comme étant ME et CE. Considérons plusieurs événements précurseurs et définissons ces événements comme suit :

A, B, C, ... = événements précurseurs qui se sont produits

A^c, B^c, C^c, \dots = événements précurseurs qui ne sont pas produits

A titre d'exemple :

A = comportement anormal des animaux,

A^c = comportement normal des animaux.

³ L'exposant c est introduit pour l'événement complémentaire

Pour des raisons de traitement pratique, on considère que l'observation de ces événements est effectuée 24 heures avant la possible occurrence du séisme. C'est-à-dire que si l'événement précurseur est observé le samedi, alors probablement il aura lieu le dimanche et que l'occurrence du séisme le lundi dépendra uniquement de l'information qui sera observée le dimanche et non pas celle du samedi et dimanche.

Supposons qu'un séisme majeur se produit une fois tous les 300 ans (soit à peu près 1 jour sur 100000 jours) alors la probabilité d'occurrence du séisme est de $1/100000$ soit :

$$P[E]=10^{-5} \quad (9)$$

Essayons maintenant d'introduire l'information concernant un événement précurseur (par exemple A). On suppose que :

$$P[A/E]=0.1 \quad (10)$$

(Probabilité : Événement précurseur A se produit sachant que le séisme a eu lieu).

$$P[A/E^c]=0.001 \quad (11)$$

(Probabilité : Événement précurseur A se produit sachant que le séisme n'a pas eu lieu).

Appliquons le théorème de Bayes:

$$P_{1|1} = P[E | A] = P[E] P[A | E] / P[A] \quad (12)$$

D'après le théorème de la probabilité Totale (TPT):

$$\begin{aligned} P[A] &= P[A \cap E] + P[A \cap E^c] = P[A/E] \times P[E] + P[A/E^c] \times P[E^c] \\ &= (0.1) \times (10^{-5}) + (0.001) \times (1 - 10^{-5}) = 0.0010099 \end{aligned}$$

Ainsi :

$$P_{1|1} = P[E | A] = 10^{-5} (0.1) / \{ (0.1) \times (10^{-5}) + (0.001) \times (1 - 10^{-5}) \}$$

$$P_{1|1} = 0.001 \quad (13)$$

L'équation (9) montre qu'il y a une chance sur 100000 pour que le séisme majeur se produise. L'équation (13) montre qu'il y a une chance sur 1000 pour que le séisme majeur se produise (Nette amélioration). Si votre système d'alerte se base sur l'équation (13) alors à chaque fois que l'événement précurseur A se produit, vous allez déclencher l'alerte. MAIS QUI PEUT VOUS CROIRE ? En fait, pour 1000 alertes seulement une est bonne, les autres (999) ça sera de fausses alertes. Donc il faut améliorer le modèle. COMMENT ? En introduisant un autre événement précurseur. Considérons par exemple un autre événement précurseur B. On suppose que :

$$P[B/E]=0.1 \quad (14)$$

(Probabilité : Événement précurseur B se produit sachant que le séisme a eu lieu).

$$P[B/E^c]=0.001 \quad (15)$$

(Probabilité : Événement précurseur B se produit sachant que le séisme n'a pas eu lieu).

Appliquons le théorème de Bayes:

$$P_{1|1} = \text{Probabilité [séisme se produit sachant qu'il a été prédit]} \quad (16)$$

Que veut dire prédit ? Tout simplement que les deux événements précurseurs A et B se sont produits (non pas A ou B mais A et B)

$$P_{1|1} = P[E | A \cap B] = P[E] P[A \cap B | E] / P[A \cap B] \quad (17)$$

Calculons $P[A \cap B | E]$ et $P[A \cap B]$.

$$P[A \cap B | E] = ???$$

Une hypothèse simplificatrice est introduite (vous avez le droit de la contredire), on suppose que l'occurrence que l'événement A n'implique pas nécessairement l'occurrence de l'événement B. Cette hypothèse bien sûr n'est pas tout à fait exacte, pourquoi ? Considérons par exemple :

A=Événement : comportement anormal des animaux

B=Événement : Remontée de la nappe phréatique due au séisme

Dire que les événements A et B sont indépendants implique que si le comportement des animaux est anormal alors la nappe phréatique peut remonter comme elle peut ne pas remonter. En tous cas cette hypothèse va nous permettre d'écrire $P[A \cap B | E]$ comme suit :

$$P[A \cap B | E] = P[A | E] \times P[B | E] = 0.1 \times 0.1 = 0.01 \quad (18)$$

$$P[A \cap B] = ???$$

En appliquant le théorème de la probabilité totale (TPT):

$$\begin{aligned} P[A \cap B] &= P[\{A \cap B\} \cap E] + P[\{A \cap B\} \cap E^c] \\ &= P[A \cap B | E] \times P[E] + P[A \cap B | E^c] \times P[E^c] \end{aligned} \quad (19)$$

Au niveau de l'équation (19) on remarque que $P[A \cap B | E^c]$ n'est pas connue. Cette expression sera calculée comme suit (voir Equation 18):

$$P[A \cap B | E^c] = P[A | E^c] \times P[B | E^c] = 0.001 \times 0.001 = 10^{-6} \quad (20)$$

Par conséquent :

$$P[A \cap B] = 0.01 \times 10^{-5} + 10^{-6} \times (1 - 10^{-5}) = 1.09999 \cdot 10^{-6} \quad (21)$$

Ainsi :

$$P_{1|1} = 10^{-5} \times 0.01 / \{1.09999 \cdot 10^{-6}\} = 0.1 \quad (22)$$

Résumons :

1. Sans système de prédiction

Probabilité d'occurrence du séisme = 10^{-5} (1 journée sur 100000).

2. Avec un système de prédiction :

a) Un seul événement précurseur : A

Probabilité d'occurrence du séisme sachant que l'événement précurseur A s'est produit = 0.001 (1 journée sur 1000 journées). En fait lorsque A se produit alors vous allez déclencher l'alerte mais vous avez 1 chance sur 1000 pour qu'un séisme se produit, donc 999 fausses alertes.

b) Deux événements précurseurs : A et B

Probabilité d'occurrence du séisme sachant que les événements précurseurs se sont

produits=0.01 (1 journée sur 10 journées). En fait lorsque A et B se réalisent alors vous allez déclencher l'alerte mais vous avez 1 chance sur 10 pour qu'un séisme se produit, donc 9 fausses alertes.

Rappelez vous le modèle doit vous donner du 100% (ou presque).

Une petite information :

La probabilité de décrocher le gros lot à la loterie est d'environ 10^{-6} (0,000001, c'est-à-dire vous avez une chance sur un million pour être le prochain gagnant) et pourtant les gens continuent à jouer en espérant de gagner.

Exercice n°1 : Prédiction des séismes

Considérons quatre (04) événements précurseurs (A, B, C et D). On suppose que les quatre événements ont la même probabilité d'occurrence sachant que le séisme s'est produit ou non :

$$P[A/E] = P[B/E] = P[C/E] = P[D/E], \quad P[A/E^c] = P[B/E^c] = P[C/E^c] = P[D/E^c].$$

Les événements A, B, C et D sont indépendants sachant que le séisme s'est produit ou non. --

Question :

1. Evaluer votre système d'alerte (c'est-à-dire calculer $P_{1|1}$) pour les cas suivants :
 - 1.1 $P[\text{Evénement précurseur se produit}/E]=0.1$ et $P[\text{Evénement précurseur se produit}/E^c]=0.001$
 - 1.2 $P[\text{Evénement précurseur se produit}/E]=0.1$ et $P[\text{Evénement précurseur se produit}/E^c]=0.00001$
 - 1.3 $P[\text{Evénement précurseur se produit}/E]=1$ et $P[\text{Evénement précurseur se produit}/E^c]=0.01$
2. Comparer les résultats obtenus.
3. Pouvez indiquer vos idées concernant l'amélioration de la prédiction.

Exercice n°2 : Système d'alarme d'un appareil mécanique

Un appareil mécanique est relié à un système d'alarme qui se déclenche lorsque des conditions anormales sont relevées. Pour une journée particulière, la probabilité de déclenchement de ce système, lorsque des conditions anormales sont observées, est de 0.95. Cette probabilité est de 0.005 lorsque les conditions normales sont observées. Par ailleurs, la probabilité d'occurrence de conditions anormales en une journée particulière est de 0.005.

1. Quelle est la probabilité de fausses alertes, c'est-à-dire lorsque les conditions sont normales alors que l'alarme s'est déclenchée.
2. Quelle est la probabilité d'événements inidentifiables, c'est-à-dire la probabilité que les conditions soient anormales alors que le système d'alarme ne s'est pas déclenché.
3. Combien de fausses alertes et combien d'événements inidentifiables peuvent être observés sur une période de 10 ans.

Corrigé du TD n°4

Exercice n°1 : Prédiction des séismes

Sans système de prédiction

Probabilité d'occurrence du séisme = 10^{-5}

(1 journée sur 100000). Donc 99999 fausses alertes

Avec un système de prédiction

- Un seul événement précurseur : A

Probabilité d'occurrence du séisme sachant que l'événement précurseur A s'est produit = 0.001 (1 journée sur 1000 journées). En fait lorsque A se produit alors vous allez déclencher l'alerte mais vous avez 1 chance sur 1000 pour qu'un séisme se produit, donc 999 fausses alertes.

(100 journées sur 100000). Donc 99900 fausses alertes

- Deux événements précurseurs : A et B

Probabilité d'occurrence du séisme sachant que les événements précurseurs se sont produits = 0.01 (1 journée sur 10 journées). En fait lorsque A et B se réalisent alors vous allez déclencher l'alerte mais vous avez 1 chance sur 10 pour qu'un séisme se produit, donc 9 fausses alertes.

(10000 journées sur 100000). Donc 90000 fausses alertes

<u>Nombre de précurseur</u>	<u>Nombre de fausses alertes</u>
<u>0</u>	<u>99999</u>
<u>1</u>	<u>99900</u>
<u>2</u>	<u>90000</u>

L'exercice n°1 consiste à évaluer le système de prédiction avec quatre (04) événements précurseurs (A, B, C et D). On suppose que les quatre événements ont la même probabilité d'occurrence sachant que le séisme s'est produit ou non :

$$P[A/E] = P[B/E] = P[C/E] = P[D/E],$$

$$P[A/E^c] = P[B/E^c] = P[C/E^c] = P[D/E^c].$$

Que veut dire prédit ? Tout simplement que les événements précurseurs A, B, C et D se sont produits (non pas A ou B ou C ou D mais A, B, C et D)

$$P_{1/1} = P\left[\frac{E}{\{A \cap B \cap C \cap D\}}\right] = \frac{P[E \cap \{A \cap B \cap C \cap D\}]}{P[A \cap B \cap C \cap D]} \frac{P[\{A \cap B \cap C \cap D\} \cap E]}{P[A \cap B \cap C \cap D]}$$

Soit
$$P_{1/1} = \frac{P[E]P\left[\frac{\{A \cap B \cap C \cap D\}}{E}\right]}{P[A \cap B \cap C \cap D]}$$

Pour calculer $P_{1/1}$ il faut donc calculer :

$P[E]$ (Déjà calculée = 10^{-5}), $P\left[\frac{\{A \cap B \cap C \cap D\}}{E}\right]$ et $P[A \cap B \cap C \cap D]$

$P\left[\frac{A \cap B \cap C \cap D}{E}\right] = P\left[\frac{A}{E}\right]P\left[\frac{B}{E}\right]P\left[\frac{C}{E}\right]P\left[\frac{D}{E}\right]$ car les évènements A, B, C et D sont indépendants.

$$P[A \cap B \cap C \cap D] = P[\{A \cap B \cap C \cap D\} \cap E] + P[\{A \cap B \cap C \cap D\} \cap E^c]$$

$$P[A \cap B \cap C \cap D] = P\left[\frac{A \cap B \cap C \cap D}{E}\right]P[E] + P\left[\frac{A \cap B \cap C \cap D}{E^c}\right]P[E^c]$$

$$P\left[\frac{A \cap B \cap C \cap D}{E}\right] = P\left[\frac{A}{E}\right]P\left[\frac{B}{E}\right]P\left[\frac{C}{E}\right]P\left[\frac{D}{E}\right]$$

$$P\left[\frac{A \cap B \cap C \cap D}{E^c}\right] = P\left[\frac{A}{E^c}\right]P\left[\frac{B}{E^c}\right]P\left[\frac{C}{E^c}\right]P\left[\frac{D}{E^c}\right]$$

1°) Ayant mis en place la méthodologie pour l'évaluation du système de prédiction, on passe maintenant vers l'application numérique :

1.1 P[Evènement précurseur se produit/E]=0.1 et P[Evènement précurseur se produit/E^c]=0.001

• P[Evènement précurseur se produit/E]=0.1

$$P\left[\frac{A}{E}\right] = P\left[\frac{B}{E}\right] = P\left[\frac{C}{E}\right] = P\left[\frac{D}{E}\right] = 0.1$$

• P[Evènement précurseur se produit/E^c]=0.001

$$P\left[\frac{A}{E^c}\right] = P\left[\frac{B}{E^c}\right] = P\left[\frac{C}{E^c}\right] = P\left[\frac{D}{E^c}\right] = 0.001$$

$$P[E] = 10^{-5}$$

$$P[E^c] = 1 - 10^{-5}$$

$$P_{1/1} = 0.99900101 \approx 0.999$$

• Quatre évènements précurseurs : A, B, C et D

Probabilité d'occurrence du séisme sachant que les évènements précurseurs se sont produits=0.999 (999 journées sur 1000 journées). En fait lorsque A, B, C et D se réalisent alors vous allez déclencher l'alerte mais vous avez 1 chance sur 10 pour qu'un séisme se produit, donc 9 fausses alertes.

(99900journées sur 100000). Donc 100 fausses alertes

1.2 P[Evènement précurseur se produit/E]=0.1 et P[Evènement précurseur se produit/E^c]=0.00001

• P[Evènement précurseur se produit/E]=0.1

$$P\left[\frac{A}{E}\right] = P\left[\frac{B}{E}\right] = P\left[\frac{C}{E}\right] = P\left[\frac{D}{E}\right] = 0.1$$

• P[Evènement précurseur se produit/E^c]=0.00001

$$P\left[\frac{A}{E^c}\right] = P\left[\frac{B}{E^c}\right] = P\left[\frac{C}{E^c}\right] = P\left[\frac{D}{E^c}\right] = 0.00001$$

$$P[E] = 10^{-5}$$

$$P[E^c] = 1 - 10^{-5}$$

$$P_{1/1} = 0.9999999999$$

• Quatre évènements précurseurs : A, B, C et D

Probabilité d'occurrence du séisme sachant que les évènements précurseurs se sont produits=0.9999999999 (9999999999 journées sur 1000 milliards de

journées). En fait lorsque A , B , C et D se réalisent alors vous allez déclencher l'alerte mais vous avez 9999999999 chances sur 1000 milliards pour qu'un séisme se produise, donc 1 fausse alerte pour 1000 milliards .

Donc 0,000001 fausses alertes pour 100000 alertes

1.3 $P[\text{Evénement précurseur se produit}/E]=1$ et $P[\text{Evénement précurseur se produit}/E^c]=0.01$

$\rightarrow P[\text{Evénement précurseur se produit}/E]=1$

$$P[A/E]=P[B/E]=P[C/E]=P[D/E]=1$$

$\rightarrow P[\text{Evénement précurseur se produit}/E^c]=0.01$

$$P[A/E^c]=P[B/E^c]=P[C/E^c]=P[D/E^c]=0.01$$

$$P[E]=10^{-5}$$

$$P[E^c]=1-10^{-5}$$

$$P_{1/1} = 0.9990010089810290000000$$

$$\approx 0.999$$

- Quatre événements précurseurs : A , B , C et D

Probabilité d'occurrence du séisme sachant que les événements précurseurs se sont produits = 0.999 (999 journées sur 1000 journées). En fait lorsque A , B , C et D se réalisent alors vous allez déclencher l'alerte mais vous avez 1 chance sur 10 pour qu'un séisme se produise, donc 9 fausses alertes.

(99900journées sur 100000). Donc 100 fausses alertes

2°) Comparer les résultats:

Pour comparer les résultats, il faut utiliser les mêmes probabilités :

Nombre de précurseur Nombre de fausses alertes

0 99999

1 99900

2 90000

4 100

3°) Comment améliorer la prédiction

- Augmenter le nombre de précurseurs
- Choisir des précurseurs qui présentent les caractéristiques suivantes :
 - $P[A/E]$ doit tendre vers 1
 - $P[A/E^c]$ doit tendre vers 0

Exercice n°2 : Système d'alarme d'un appareil mécanique

Il faut tout d'abord définir les évènements:

A: Alarme se déclenche

A^c : Alarme ne se déclenche pas

N: Conditions normales

N^c : Conditions anormales

Pour une journée particulière,

La probabilité de déclenchement de ce système, lorsque des conditions anormales sont observées, est de 0.95. Donc:

$$P[A/N^c] = 0.95$$

Cette probabilité est de 0.005 lorsque les conditions normales sont observées.

$$P[A/N] = 0.005$$

La probabilité d'occurrence de conditions anormales en une journée particulière est de 0.005

$$P[N^c] = 0.005$$

1) Quelle est la probabilité de fausses alertes, c'est-à-dire lorsque les conditions sont normales alors que l'alarme s'est déclenchée.

$$P[N/A] = \text{????}$$

$$P[N/A] = \frac{P[N \cap A]}{P[A]} = \frac{P[A \cap N]}{P[A]}, \quad P[N/A] = \frac{P[A/N]P[N]}{P[A]}, \quad P[A] = P[A \cap N] + P[A \cap N^c]$$

$$P[A] = P[A/N]P[N] + P[A/N^c]P[N^c], \quad P[N/A] = \frac{P[A/N]P[N]}{P[A/N]P[N] + P[A/N^c]P[N^c]}$$

$$P[N/A] = \frac{0.005 \times (1 - 0.005)}{0.005 \times (1 - 0.005) + 0.95 \times 0.005} = 0.5116$$

2) Quelle est la probabilité d'événements inidentifiables, c'est-à-dire la probabilité que les conditions soient anormales alors que le système d'alarme ne s'est pas déclenché.

$$P[N^c/A^c]$$

$$P[N^c/A^c] = \frac{P[A^c/N^c]P[N^c]}{P[A^c/N]P[N] + P[A^c/N^c]P[N^c]}$$

$$P[N^c/A^c] = \frac{(1 - 0.95) \times 0.005}{(1 - 0.005) \times (1 - 0.005) + (1 - 0.95) \times 0.005} = 0.0002525$$

3) Combien de fausses alertes et combien d'événements inidentifiables peuvent être observés sur une période de 10 ans.

$$NB = P[N \cap A] \times 365 \times 10 = P[N/A]P[A] \times 365 \times 10$$

$$P[A] = 0.005 \times (1 - 0.005) + 0.95 \times 0.005 = 0.009725$$

$$P[N/A] = 0.5116$$

$$NB = 18$$

Exercice n°1 :

Calculer la Fonction de distribution cumulative de la VA X définie par :

$$f_X(x) = \begin{cases} 0.1, & 90 \leq x \leq 100 \\ 0, & \text{sinon} \end{cases}$$

$$f_X(x) = \begin{cases} 2(1-x), & 0 \leq x \leq 1 \\ 0, & \text{sinon} \end{cases}$$

Exercice n°2 :

Soit X : VA durée de vie d'un composant électronique (mesurée en heure).

$$f_X(x) = \begin{cases} 0, & x < 100 \\ \frac{100}{x^2}, & x \geq 100 \end{cases}$$

Calculer la probabilité pour que ce composant ait une durée de vie inférieure à 150 heures

Exercice n°3 :

Deux personnes pensent se voir entre 9h00 et 10h00. Chaque personne ne peut attendre plus de 10 minutes. C'est-à-dire qu'au-delà de 10 minutes elle partira si la deuxième personne ne vient pas. On suppose les deux personnes peuvent arriver à n'importe quel moment entre 9h00 et 10h00 et que leurs temps d'arrivée ne sont pas connus.

Considérons la variable aléatoire

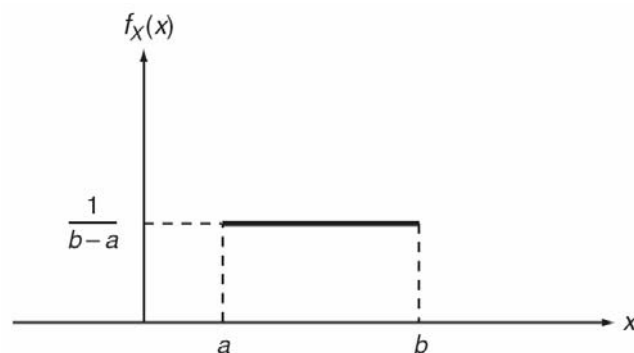
X décrite par : le nombre de minutes écoulé jusqu'à l'arrivée de la personne A

Y décrite par : le nombre de minutes écoulé jusqu'à l'arrivée de la personne B

Les deux VA présentent des valeurs qui varient entre 0 et 60 minutes.

La loi de probabilité de la VA X (ou Y) est décrite par la loi uniforme :

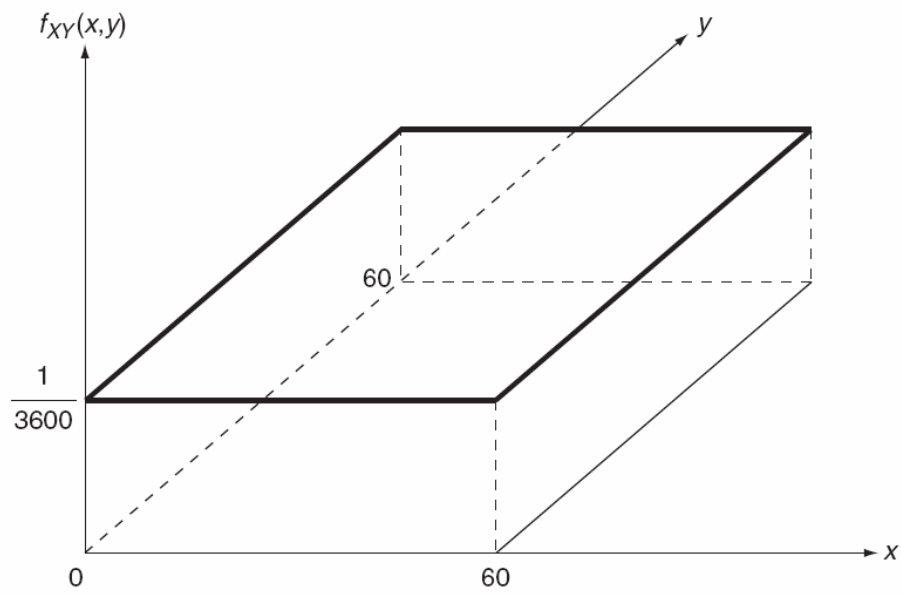
$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sinon} \end{cases}$$



La FDPC⁴ est définie par :

$$f_{XY}(x, y) = \begin{cases} \frac{1}{3600}, & 0 \leq x \leq 60 \text{ et } 0 \leq y \leq 60 \\ 0 & \text{sinon} \end{cases}$$

⁴ FDPC : Fonction de densité de probabilité conjointe



Question :

Il est recherché la probabilité pour que les deux personnes se rencontrent entre 9h00 et 10h00.

Corrigé du TD n°5

Exercice n°1 :

$$\bullet f_X(x) = \begin{cases} 0.1, & 90 \leq x \leq 100 \\ 0, & \text{sinon} \end{cases}$$

La Fonction de distribution cumulative de la VA X est définie par :

$$F_X(x) = \int_{-\infty}^x f_X(u) du$$

Si $x < 90$ alors

$$F_X(x) = \int_{-\infty}^x f_X(u) du = \int_{-\infty}^x 0 du = 0, \text{ donc } F_X(x) = 0$$

Si $90 \leq x \leq 100$ alors

$$F_X(x) = \int_{-\infty}^x f_X(u) du = \int_{-\infty}^{90} 0 du + \int_{90}^x 0.1 du = [0.1u]_{90}^x$$

$$F_X(x) = \frac{x-90}{10}$$

Si $x > 100$ alors

$$F_X(x) = \int_{-\infty}^x f_X(u) du = \int_{-\infty}^{90} 0 du + \int_{90}^{100} 0.1 du + \int_{100}^x 0 du = 1$$

$$F_X(x) = 1$$

$$\bullet f_X(x) = \begin{cases} 2(1-x), & 0 \leq x \leq 1 \\ 0, & \text{sinon} \end{cases}$$

La Fonction de distribution cumulative de la VA X est définie par :

$$F_X(x) = \int_{-\infty}^x f_X(u) du$$

Si $x < 0$ alors

$$F_X(x) = \int_{-\infty}^x f_X(u) du = \int_{-\infty}^x 0 du = 0, \text{ donc } F_X(x) = 0$$

Si $0 \leq x \leq 1$ alors

$$F_X(x) = \int_{-\infty}^x f_X(u) du = \int_{-\infty}^0 0 du + \int_0^x 2(1-u) du = [(2u - u^2)]_0^x$$

$$F_X(x) = 2x - x^2$$

Si $x > 1$ alors

$$F_X(x) = \int_{-\infty}^x f_X(u) du = \int_{-\infty}^0 0 du + \int_0^1 2(1-u) du + \int_1^x 0 du = 1$$

$$F_X(x) = 1$$

Exercice n°2 :

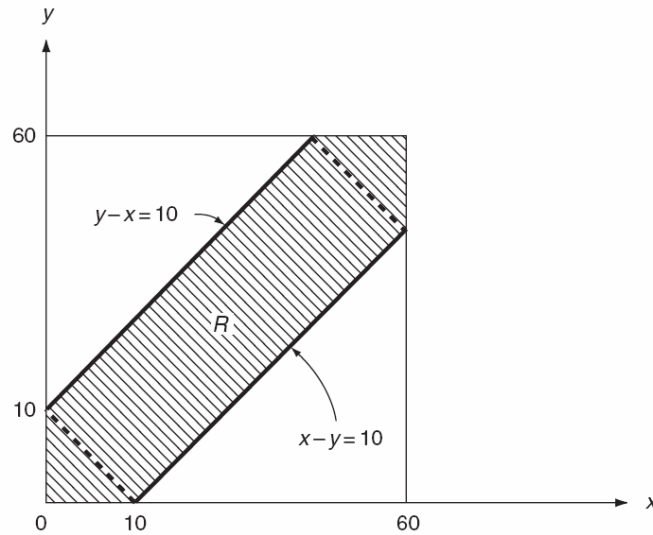
Soit A l'évènement identifié par :

Le composant a une durée de vie inférieure à 150 heures

$$\text{Prob}(A) = P[X \leq 150] = \int_{-\infty}^{150} f_X(x) dx = \int_{-\infty}^{100} 0 dx + \int_{100}^{150} \frac{100}{x^2} dx = \left[-\frac{100}{x} \right]_{100}^{150} = 0,33$$

Exercice n°3 :

Probabilité que les deux personnes se rencontrent = $P[|X - Y| \leq 10]$



$$\begin{aligned} P[|X - Y| \leq 10] &= \\ &= [2(5)(10) + 10\sqrt{2}(50\sqrt{2})]/3600 = \frac{11}{36} \end{aligned}$$

Exercice n°1 (06 points)

Soit X la surface (مساحة) d'une maison (mesurée en m²). Le traitement de l'information relatif à 100 maisons a permis de dresser l'histogramme de la variable statistique X (voir Figure 1):

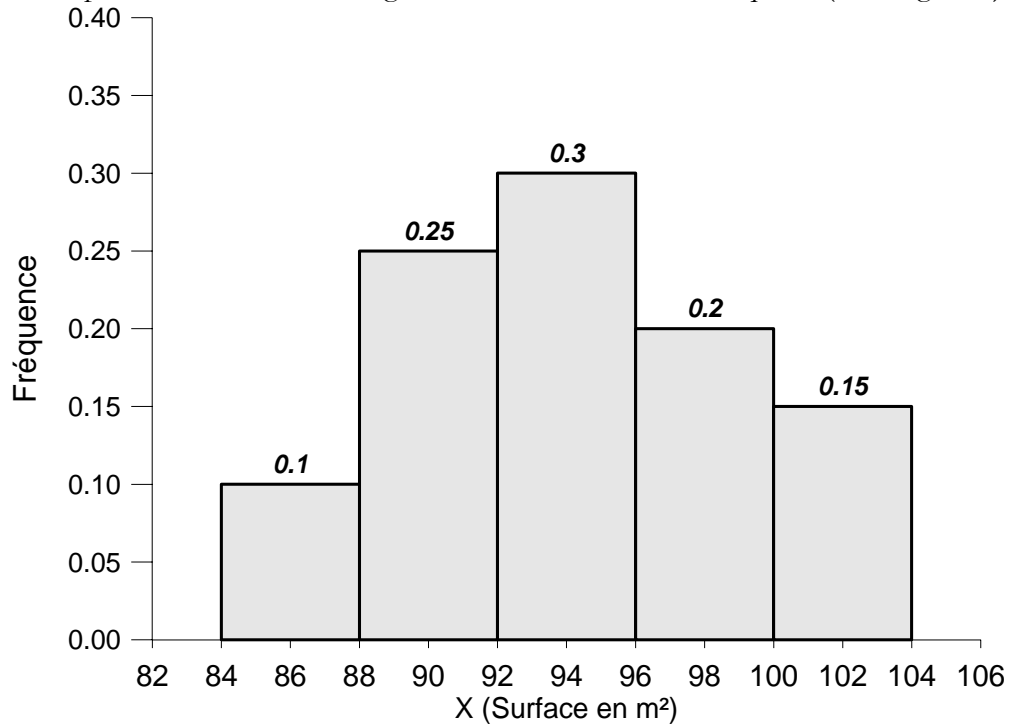


Figure 1 Histogramme de la variable statistique X Surface d'une maison (mesurée en m²)

- 1) Déterminer la moyenne de la variable X (02 points)
- 2) Déterminer l'écart-type de la variable X (01 point)
- 3) Tracer la fonction cumulative (01 point)

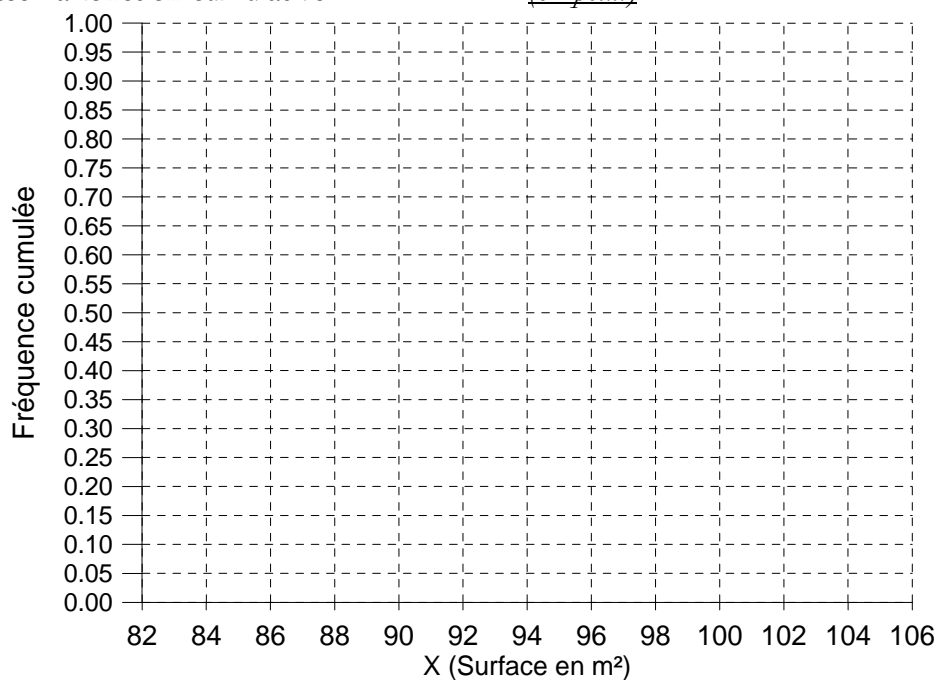


Figure 2 Fonction cumulative de la variable statistique X Surface d'une maison (mesurée en m²)

- 4) Donner la définition de la médiane (01 point)

- 5) Déduire graphiquement la médiane (01 point)

Exercice n°2 (06 points)

Une usine produit des pièces d'une machine. Pour chaque pièce (individu), on dispose du coût de sa production (DA) et du temps nécessaire pour sa réalisation (en heures). Le tableau ci-après (série statistique) donne cette répartition :

<i>Individu</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
Temps (X) mesurée en heures	2	3	5	2	4
Coût (Y) mesurée en Dinars	10	16	23	12	18

Droite de corrélation linéaire :
$$Y = \bar{Y} - \frac{\text{cov}(X, Y)}{\sigma_X^2} \bar{X} + \frac{\text{cov}(X, Y)}{\sigma_X^2} X$$

Coefficient de corrélation
$$R = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Covariance
$$\text{cov}(X, Y) = \sum_{i=1}^n \frac{1}{n} (x_i - \bar{X})(y_i - \bar{Y})$$

Les calculs qui vont suivre doivent être effectués en utilisant la série statistique (voir tableau plus haut)

- 1) Calculer la moyenne de la variable statistique X (0.5 point)
- 2) Calculer la moyenne de la variable statistique Y (0.5 point)
- 3) Calculer l'écart-type de la variable statistique X (0.5 point)
- 4) Calculer l'écart-type de la variable statistique Y (0.5 point)
- 5) Calculer la covariance des variables statistiques X et Y (01 point)
- 6) En supposant qu'il existe une corrélation linéaire entre X et Y, déterminer cette droite de corrélation (01 point)
- 7) Calculer le coefficient de corrélation. Conclusion ? (01 point)
- 8) Une nouvelle pièce est réalisée en 6 heures. Estimer le coût de production de cette pièce en utilisant la droite de corrélation établie (Question 6). (01 point)

Exercice n°3 (04 points)

Un magasin reçoit des machines de deux usines (US-1) et (US-2) avec les proportions (ou probabilité) suivantes :

- 40% des machines viennent de l'usine US-1-
- 60% des machines viennent de l'usine US-2-

Par ailleurs, la probabilité pour qu'une machine soit défectueuse sachant qu'elle provienne de l'usine US-1 est de 4%. Enfin, la probabilité pour qu'une machine soit défectueuse sachant qu'elle provienne de l'usine US-2 est de 3%.

Machine défectueuse: آلة فيها عطب أو عيب

- 1) On achète au hasard une machine de ce magasin, quelle est la probabilité qu'elle ne soit pas défectueuse (1.5 points)
- 2) Sur 1000 machines vendues au magasin, quelle est le nombre de machines défectueuses (01 point)
- 3) Sachant que la machine achetée au magasin est défectueuse quelle est la probabilité qu'elle provienne de l'usine US-1- (1.5 point)

Exercice n°4 (04 points)

La durée de vie d'une télévision est une variable aléatoire X (mesurée en ans). Cette variable aléatoire suit une loi de probabilité définie par la fonction de densité de probabilité suivante :

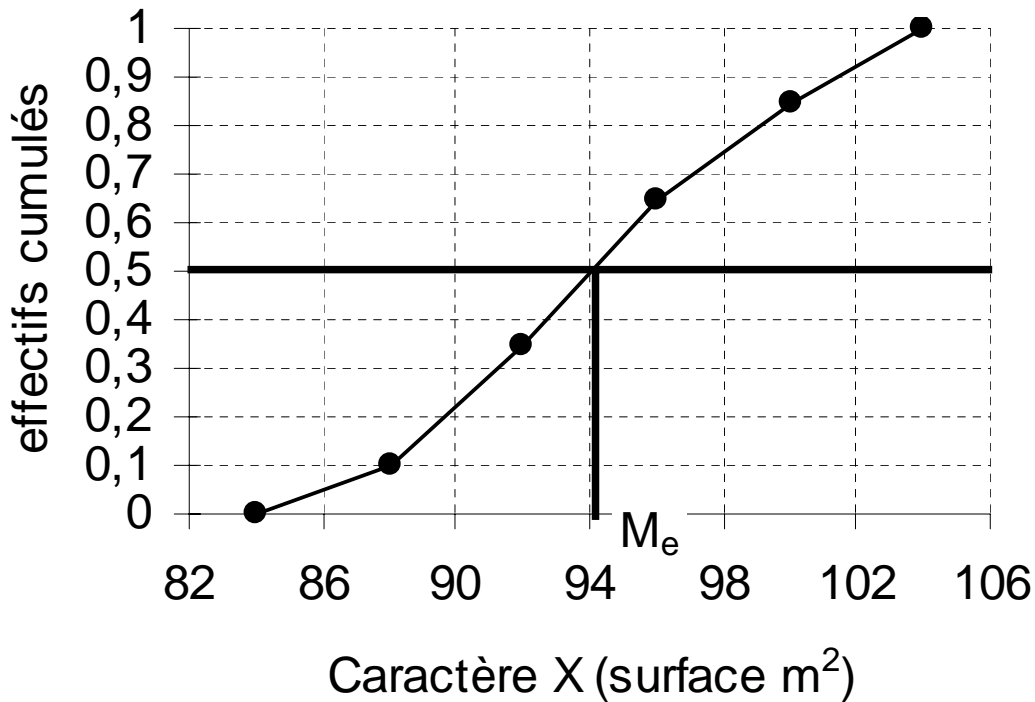
$$f_X(x) = \begin{cases} \frac{1}{8} e^{-\frac{x}{8}} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

1. Vérifier que cette fonction est une FDP (Fonction de densité de probabilité) (01 point)
2. Calculer la probabilité que la télévision que vous venez d'acheter ait une durée de vie supérieure à 8 ans. (01 point)
3. Calculer l'espérance mathématique de la variable aléatoire X. (01 point)
4. Vous possédez une telle télévision depuis 2 ans. Quelle est la probabilité que sa durée de vie soit encore de 8 ans à partir de maintenant
C'est-à-dire il faut calculer $P[X > 10 / X > 2]$ (01 point)

Corrigé de l'Examen 1

Exercice n°1

1. La moyenne arithmétique est donnée par : $\bar{X} = \frac{1}{N} \sum_{i=1}^5 n_i x_i = \sum_{i=1}^5 f_i x_i$, dans l'exercice il est donné les fréquences relatives, f_i . x_i représente le centre des classes. On trouve que $\bar{X} = 94,2 \text{ m}^2$.
2. L'écart type, σ est donnée par : $\sigma^2 = V = \sum_{i=1}^5 f_i (x_i - \bar{X})^2 = 23,16 \text{ (m}^4\text{)}$
3. Le tracé de la courbe cumulative



Fonction cumulative de la variable statistique X surface d'une maison (en m²)

4. Définition de la médiane : La médiane, M_e , est le caractère qui partage l'effectif en deux parts égales. 50% 'd'individus' ont un caractère inférieur à M_e , et 50% ont caractère supérieur à M_e .
- 5.
6. Graphiquement, la médiane correspond à l'abscisse du point d'intersection de la courbe cumulative et l'horizontale qui correspond à 50% de l'effectif, voir graphe.

Exercice n°2

Pour répondre aux questions de 1 à 6 il est pratique de remplir d'abord le tableau suivant :

Individu	1	2	3	4	5	somme
Temps X (H)	2	3	5	2	4	16
coût Y (Dinars)	10	16	23	12	18	79
X²	4	9	25	4	16	58
Y²	100	256	529	144	324	1353
X.Y	20	48	115	24	72	279

(notation V.S. variable statistique ou caractère)

1. La moyenne arithmétique de la V. S. X est donnée par la formule : $\bar{X} = \frac{1}{5} \sum_{i=1}^5 x_i = \frac{16}{5} = 3,2$ (heures et centième d'heure).

2. La moyenne arithmétique de la variable statistique Y est donnée par la formule : $\bar{Y} = \frac{1}{5} \sum_{i=1}^5 y_i = \frac{79}{5} = 15,8$ (Dinars).

3. L'écart type de la V.S. X est donné par : $\sigma_X = \sqrt{\frac{1}{5} \sum_{i=1}^5 (x_i - \bar{X})^2} = \sqrt{X^2 - \bar{X}^2} = \sqrt{\frac{58}{5} - (3,2)^2} = 1,166$

4. L'écart type de la V.S. Y est donné par : $\sigma_Y = \sqrt{\frac{1}{5} \sum_{i=1}^5 (y_i - \bar{Y})^2} = \sqrt{Y^2 - \bar{Y}^2} = \sqrt{\frac{1353}{5} - (15,8)^2} = 4,578$

5. La covariance des V.S. X et Y est donnée par : $\text{Cov}(X, Y) = \frac{1}{5} \sum_{i=1}^5 (x_i - \bar{X})(y_i - \bar{Y}) = \frac{1}{5} \sum_{i=1}^5 (x_i y_i) - \bar{X} \cdot \bar{Y} = 5,24$

6. Les paramètres a et b de la droite de régression linéaire $Y = aX + b$ sont donnés par : $a = \frac{\text{Cov}(X, Y)}{\sigma_X^2} = \frac{5,24}{1,36} = 3,8529$ et $b = \bar{Y} - a\bar{X} = 15,8 - (3,8529)(3,2) = 3,47$. La droite de corrélation s'écrit donc : $Y = 3,85X + 3,47$

7. Voir les hypothèses, le coefficient de corrélation R est donnée par : $R = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = 0,98$.

Le coefficient de corrélation est proche de 1, on peut conclure que les deux variables sont bien corrélées. Une pièce est réalisée en $X=6$ heure, le coût peut être estimé en utilisant la droite de corrélation : $Y = 3,85(6) + 3,47 = 26,59$ (dinars)

Exercice n°3

On commence par interpréter les hypothèses :

- M1 est l'événement qu'une machine provient de l'US-1 ; $\text{Prob}(M1) = 0,4$
- M2 est l'événement qu'une machine provient de l'US-2 : $\text{Prob}(M2) = 0,6$
- D désigne une machine défectueuse, D^c désigne machine non défectueuse.
- La probabilité qu'une machine soit défectueuse sachant qu'elle provient de l'US-1 est : $\text{Prob}(D/M1) = 0,04$
- La probabilité qu'une machine soit défectueuse sachant qu'elle provient de l'US-2 est : $\text{Prob}(D/M2) = 0,03$

1. La probabilité qu'une machine ne soit pas défectueuse est donnée par :

$$\text{Prob}(D^c) = 1 - \text{Prob}(D)$$

$$\text{Prob}(D) = \text{Prob}(D \cap M1) + \text{Prob}(D \cap M2) = \text{Prob}(D/M1) \cdot \text{Prob}(M1) + \text{Prob}(D/M2) \cdot \text{Prob}(M2)$$

$$\text{Prob}(D^c) = 1 - (0,04)(0,4) + (0,03)(0,6) = 1 - 0,034 = 0,966$$

2. Sur 1000 machines vendues au magasin le nombre de machines défectueuses est de : $1000 \cdot \text{Prob}(D) = 1000(0,034) = 34$ machines

3. La probabilité qu'une machine provient du magasin US-1 sachant qu'elle est défectueuse est donnée par :

$$\text{Prob}(M1/D) = \frac{\text{Prob}(D/M1) \cdot \text{P}(M1)}{\text{Prob}(D/M1) \cdot \text{P}(M1) + \text{Prob}(D/M2) \cdot \text{P}(M2)} = \frac{(0,04) \cdot 0,4}{(0,04)(0,4) + (0,03)(0,6)} = \frac{0,016}{0,034} = 47\%$$

Exercice n°4

1. La fonction f_X est une FDP si : $f_X(x) \geq 0$ pour tout x réel et $\int_{-\infty}^{+\infty} f_X(t) dt = 1$

- La fonction exponentielle est positive donc $f_X(x) \geq 0$

$$- \int_{-\infty}^{+\infty} f_X(t) dt = \int_0^{+\infty} \frac{1}{8} e^{-\frac{t}{8}} dt = \left[-e^{-\frac{t}{8}} \right]_0^{+\infty} = 1.$$

Donc f_X est bien une FDP.

2. L'événement « la télévision ait une durée de vie supérieure à 8 » peut être exprimé par « $X \geq 8$ »

$$\text{Prob}(X \geq 8) = \int_8^{+\infty} \frac{1}{8} e^{-\frac{t}{8}} dt = \left[-e^{-\frac{t}{8}} \right]_8^{+\infty} = e^{-1}$$

3. L'espérance mathématique est donnée par : $E(X) = \int_{-\infty}^{+\infty} t \cdot f_X(t) dt$

Le calcul de $E(X)$ peut se faire par intégration par partie, si on pose :

$$u = t \quad du = dt$$

$$dv = \frac{1}{8} e^{-\frac{t}{8}} \quad v = -e^{-\frac{t}{8}}$$

$$\text{On a alors : } E(X) = \int_0^{+\infty} t \cdot \frac{1}{8} e^{-\frac{t}{8}} dt = \left[-t \cdot e^{-\frac{t}{8}} \right]_0^{+\infty} - \int_0^{+\infty} \left(-e^{-\frac{t}{8}} \right) du = \int_0^{+\infty} \left(e^{-\frac{t}{8}} \right) du = \left[8e^{-\frac{t}{8}} \right]_0^{+\infty} = 8$$

4.

$$\text{Prob}(X \geq 10 / X \geq 2) = \frac{\text{Prob}[(X \geq 10) \cap (X \geq 2)]}{\text{Prob}(X \geq 2)} = \frac{\text{Prob}(X \geq 10)}{\text{Prob}(X \geq 2)} =$$

$$\left[\int_{10}^{+\infty} \frac{1}{8} e^{-\frac{t}{8}} dt \right] / \left[\int_2^{+\infty} \frac{1}{8} e^{-\frac{t}{8}} dt \right] = \left[-e^{-\frac{t}{8}} \right]_{10}^{+\infty} / \left[-e^{-\frac{t}{8}} \right]_2^{+\infty} = \frac{e^{-\frac{10}{8}}}{e^{-\frac{2}{8}}} = e^{-\frac{10}{8} + \frac{2}{8}} = e^{-1}$$

Exercice n°1 (07 points)

Soit X la variable statistique (mesurée en m). Le traitement de l'information relative à cette variable statistique a permis de dresser l'histogramme de la variable statistique X (voir Figure 1):

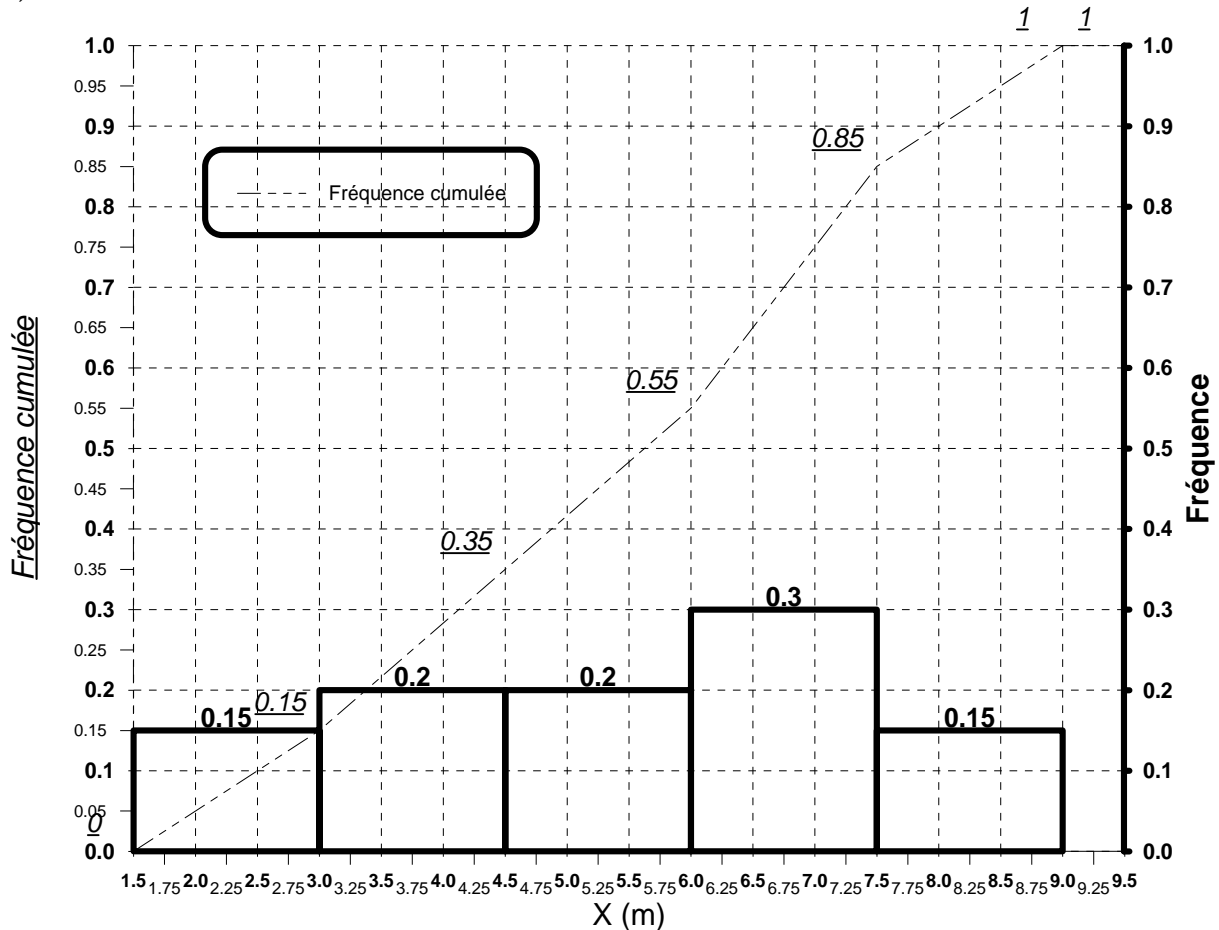


Figure 1

Histogramme de la variable statistique X (m)

- 1) Déterminer la moyenne de la variable X (02 points)
- 2) Déterminer l'écart-type de la variable X (02 points)
- 3) Déterminer graphiquement la médiane (01 point)
- 4) Déterminer graphiquement le mode (01 point)
- 5) Classifier le mode, la médiane et la moyenne et formuler la conclusion qui s'impose (01 point)

Exercice n°2 (05 points)

Soit X et Y deux variables statistiques mesurées sur un même individu. Par exemple, pour l'individu n°3, X=2 et Y=8.

Individu	1	2	3	4	5
X	3	4	2	5	3
Y	12	14	8	19	11

Droite de corrélation linéaire :
$$Y = \bar{Y} - \frac{\text{cov}(X, Y)}{\sigma_X^2} \bar{X} + \frac{\text{cov}(X, Y)}{\sigma_X^2} X$$

Coefficient de corrélation
$$R = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Covariance

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

Les calculs qui vont suivre doivent être effectués en utilisant la série statistique (voir tableau plus haut)

- 1) Calculer la moyenne de la variable statistique X (0.5 point)
- 2) Calculer la moyenne de la variable statistique Y (0.5 point)
- 3) Calculer l'écart-type de la variable statistique X (0.5 point)
- 4) Calculer l'écart-type de la variable statistique Y (0.5 point)
- 5) Calculer la covariance des variables statistiques X et Y (01 point)
- 6) En supposant qu'il existe une corrélation linéaire entre X et Y, déterminer cette droite de corrélation (01 point)
- 7) Calculer le coefficient de corrélation. Conclusion ? (01 point)

Exercice n°3 (04 points)

Un magasin reçoit des machines de quatre usines (US-1), (US-2), (US-3) et (US-4) avec des proportions (ou probabilité) identiques. Par ailleurs,

- La probabilité pour qu'une machine soit défectueuse sachant qu'elle provienne de l'usine US-1 est de 4%
- La probabilité pour qu'une machine soit défectueuse sachant qu'elle provienne de l'usine US-2 est de 3%
- La probabilité pour qu'une machine soit défectueuse sachant qu'elle provienne de l'usine US-3 est de 2.5%
- La probabilité pour qu'une machine soit défectueuse sachant qu'elle provienne de l'usine US-4 est de 5%

Machine défectueuse:

آلة فيها عطب أو عيب

- 1) On achète au hasard une machine de ce magasin, quelle est la probabilité qu'elle ne soit pas défectueuse (1.5 points)
- 2) Sur 1000 machines vendues au magasin, quelle est le nombre de machines défectueuses (01 point)
- 3) Sachant que la machine achetée au magasin est défectueuse quelle est la probabilité qu'elle provienne de l'usine US-2- (1.5 point)

Exercice n°4 (04 points)

Soit $f_X(x)$ la fonction de densité de probabilité de la variable aléatoire X. Cette fonction est définie comme suit :

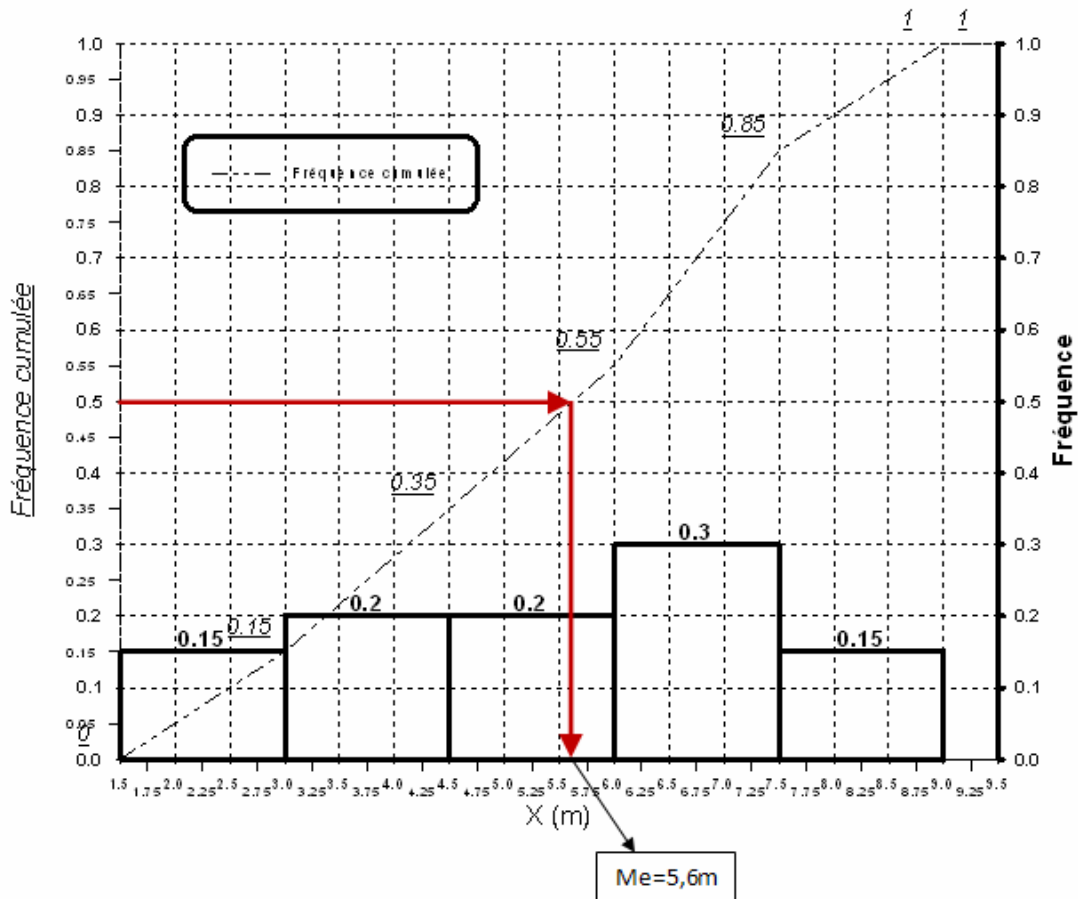
$$f_X(x) = \begin{cases} 0.5 & \text{si } 0 \leq x \leq 2 \\ 0 & \text{sinon} \end{cases}$$

1. Vérifier que cette fonction est une FDP (Fonction de densité de probabilité) (01 point)
2. Calculer la probabilité d'avoir X supérieure ou égale à 1 (01 point)
3. Calculer l'espérance mathématique de la variable aléatoire X. (01 point)
4. Calculer l'écart type de la variable aléatoire X. (01 point)

Corrigé de l'examen

Exercice n°1

1. La moyenne arithmétique est donnée par : $\bar{X} = \frac{1}{N} \sum_{i=1}^5 n_i x_i = \sum_{i=1}^5 f_i x_i$, dans l'exercice il est donné les fréquences relatives, f_i . x_i représente le centre des classes. On trouve que $\bar{X} = 5,4 \text{ m}$.
2. L'écart type, σ est donnée par : $\sigma^2 = V = \sum_{i=1}^5 f_i (x_i - \bar{X})^2 = 3,80 \text{ (m}^2\text{)}$ soit $\sigma = 1,95 \text{ m}$
3. Détermination graphique de la médiane

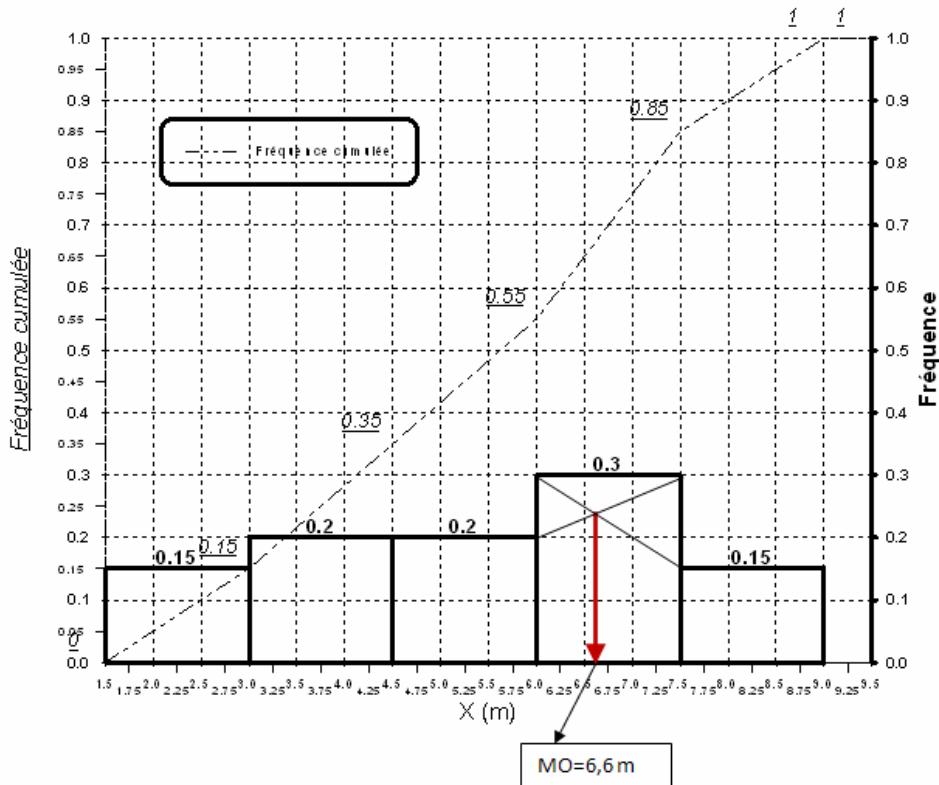


4. Détermination graphique du mode

La détermination graphique du mode peut se faire par deux méthodes :

Centre de la classe modale [6, 7,5], Soit $M_O = 6,75 \text{ m}$

La méthode des droites (voir figure) Soit $M_O = 6,6 \text{ m}$



5. Classifier le mode, la médiane et la moyenne et formuler la conclusion qui s'impose)
 $\bar{X} = 5,4 \text{ m} < Me = 5,6 \text{ m} < M_O \text{ (méthode des droites)} = 6,6 \text{ m} < M_O \text{ (méthode Centre de classe)} = 6,75 \text{ m}$

Exercice n°2

Soit X et Y deux variables statistiques mesurées sur un même individu. Par exemple, pour l'individu n°3, X=2 et Y=8.

<i>Individu</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
X	3	4	2	5	3
Y	12	14	8	19	11

Droite de corrélation linéaire :
$$Y = \bar{Y} - \frac{\text{cov}(X,Y)}{\sigma_X^2} \bar{X} + \frac{\text{cov}(X,Y)}{\sigma_X^2} X$$

Coefficient de corrélation
$$R = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Covariance
$$\text{cov}(X,Y) = \sum_{i=1}^n \frac{1}{n} (x_i - \bar{X})(y_i - \bar{Y})$$

Pour répondre aux questions de 1 à 6 il est pratique de remplir d'abord le tableau suivant :

Individu	1	2	3	4	5	Somme
X	3	4	2	5	3	17
Y	12	14	8	19	11	64
X²	9	16	4	25	9	63
Y²	144	196	64	361	121	886
X.Y	36	56	16	95	33	236

(notation V.S. variable statistique ou caractère)

1. La moyenne arithmétique de la V. S. X est donnée par la formule : $\bar{X} = \frac{1}{5} \sum_{i=1}^5 x_i = \frac{17}{5} = 3,4$

2. La moyenne arithmétique de la variable statistique Y est donnée par la formule :

$$\bar{Y} = \frac{1}{5} \sum_{i=1}^5 y_i = \frac{64}{5} = 12,8.$$

3. L'écart type de la V.S. X est donné par :

$$\sigma_X = \sqrt{\frac{1}{5} \sum_{i=1}^5 (x_i - \bar{X})^2} = \sqrt{X^2 - \bar{X}^2} = \sqrt{\frac{63}{5} - (3,4)^2} = 1,02$$

4. L'écart type de la V.S. Y est donné par :

$$\sigma_Y = \sqrt{\frac{1}{5} \sum_{i=1}^5 (y_i - \bar{Y})^2} = \sqrt{Y^2 - \bar{Y}^2} = \sqrt{\frac{886}{5} - (12,8)^2} = 3,66$$

5. La covariance des V.S. X et Y est donnée par :

$$Cov(X, Y) = \frac{1}{5} \sum_{i=1}^5 (x_i - \bar{X})(y_i - \bar{Y}) = \frac{1}{5} \sum_{i=1}^5 (x_i y_i) - \bar{X} \cdot \bar{Y} = 3,68$$

6. Les paramètres a et b de la droite de régression linéaire $Y = aX + b$ sont donnés par :

$$a = \frac{Cov(X, Y)}{\sigma_X^2} = \frac{3,68}{1,04} = 3,54 \quad \text{et} \quad b = \bar{Y} - a\bar{X} = 12,8 - 3,54 * 3,4 = 0,764.$$

La droite de corrélation s'écrit donc : $Y = 3,54 X + 0,764$

7. Voir les hypothèses, le coefficient de corrélation R est donnée par : $R = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = 0,98.$

Le coefficient de corrélation est proche de 1, on peut conclure que les deux variables sont bien corrélées.

Exercice n°3

On commence par interpréter les hypothèses :

- M1 est l'événement qu'une machine provient de l'US-1 ;
- M2 est l'événement qu'une machine provient de l'US-2 ;
- M3 est l'événement qu'une machine provient de l'US-3
- M4 est l'événement qu'une machine provient de l'US-4

Du moment que le magasin reçoit les machines des quatre usines avec des proportions (ou probabilité) identiques alors :

$$\text{Prob}(M1) = \text{Prob}(M2) = \text{Prob}(M3) = \text{Prob}(M4) = 0,25$$

- D désigne une machine défectueuse, D^c désigne machine non défectueuse.

- ✓ La probabilité qu'une machine soit défectueuse sachant qu'elle provient de l'US-1 est : $\text{Prob}(D/M1) = 0,04$
- ✓ La probabilité qu'une machine soit défectueuse sachant qu'elle provient de l'US-2 est : $\text{Prob}(D/M2) = 0,03$
- ✓ La probabilité qu'une machine soit défectueuse sachant qu'elle provient de l'US-3 est : $\text{Prob}(D/M3) = 0,025$
- ✓ La probabilité qu'une machine soit défectueuse sachant qu'elle provient de l'US-4 est : $\text{Prob}(D/M4) = 0,05$

1. La probabilité qu'une machine ne soit pas défectueuse est donnée par :

$$\text{Prob}(D^c) = 1 - \text{Prob}(D)$$

Or

$$\text{Prob}(D) = \text{Prob}(D \cap M1) + \text{Prob}(D \cap M2) + \text{Prob}(D \cap M3) + \text{Prob}(D \cap M4)$$

Par conséquent :

$$\begin{aligned} \text{Prob}(D) &= \text{Prob}(D/M1) \cdot \text{Prob}(M1) + \text{Prob}(D/M2) \cdot \text{Prob}(M2) + \\ &\quad + \text{Prob}(D/M3) \cdot \text{Prob}(M3) + \text{Prob}(D/M4) \cdot \text{Prob}(M4) \end{aligned}$$

$$\begin{aligned}\text{Prob}(D) &= 0,04 * 0,25 + 0,03 * 0,25 + 0,025 * 0,25 + 0,05 * 0,25 \\ &= 0,03625\end{aligned}$$

$$\text{Soit } \text{Prob}(D^c) = 1 - \text{Prob}(D) = 1 - 0,03625 = 0,96375$$

2. Sur 100000 machines vendues au magasin le nombre de machines défectueuses est de :
 $100000 \cdot \text{Prob}(D) = 100000(0,03625) = 3625$ machines

3. La probabilité qu'une machine provient du magasin US-2 sachant qu'elle est défectueuse est donnée par

$$\text{Prob}(M2/D) = \frac{\text{Prob}(D/M2) \cdot P(M2)}{\text{Prob}(D)} = \frac{(0,03) \cdot (0,25)}{0,03625} = 0,21$$

Exercice n°4

1. La fonction $f_X(x)$ est une FDP si : $f_X(x) \geq 0$ pour tout x réel et $\int_{-\infty}^{+\infty} f_X(t) dt = 1$

- La fonction égale à 0,5 ou 0 est donc toujours positif $f_X(x) \geq 0$

$$- \int_{-\infty}^{+\infty} f_X(t) dt = \int_0^2 0,5 dt = [0,5t]_0^2 = 1.$$

Donc f_X est bien une FDP.

2. L'événement est donc « $X \geq 1$ »

$$\text{Prob}(X \geq 1) = \int_1^{+\infty} f_X(t) dt = \int_1^2 0,5 dt + \int_2^{+\infty} 0 dt = 0,5$$

3. L'espérance mathématique est donnée par : $E(X) = \int_{-\infty}^{+\infty} t \cdot f_X(t) dt$

$$E(X) = \int_0^2 t \cdot \frac{1}{2} dt = \left[\frac{t^2}{2} \right]_0^2 = 1$$

4. L'écart type s'exprime par $\sigma_X = \sqrt{\int_{-\infty}^{+\infty} (t - E(X))^2 \cdot f_X(t) dt}$

$$\sigma_X = \sqrt{\int_0^2 \frac{(t-1)^2}{2} dt} = \sqrt{\left[\frac{(t-1)^3}{6} \right]_0^2} = \frac{1}{\sqrt{3}}$$