

Chapitre 1

LES DISTRIBUTIONS A UN CARACTERE

I. GENERALITES : VOCABULAIRE

Qu'est-ce que la Statistique ?

La statistique est une science qui en globe un ensemble de méthodes et de théories appliquées à l'analyse de phénomènes et de données dont le comportement ne peut être décrit avec exactitude, mais plutôt être analysé dans un contexte d'incertitude. Le but ultime de ces méthodes et modèles est d'arriver à des conclusions pratiques pour éventuellement proposer des recommandations et des mesures correctives s'il y a lieu ou encore pour mieux comprendre ces phénomènes et éventuellement en prévoir le comportement.

La statistique est l'ensemble des instruments et de recherches mathématiques permettant de déterminer les caractéristiques d'un ensemble de données (généralement vaste).

On désigne aussi sous le vocable statistique un nombre calculé à partir d'observation (en usant de méthodes statistiques). Les statistiques sont de ce fait, le produit des analyses reposant sur l'usage de la statistique.

1. Population

On appelle population un ensemble d'éléments soumis à une étude statistique. C'est l'ensemble de références sur lequel porte les observations. Il peut s'agir d'un ensemble de personnes, d'animaux ou de choses.

Prenons comme ensemble d'étude statistique, les machines à dactylographier stockées dans une salle. L'ensemble des machines à dactylographier de cette salle est une population : c'est ici la population cible

2. Individu

Un individu est un élément de la population. C'est l'unité de la population statistique.

Une machine de la salle à dactylographier est un individu.

3. Caractère

Un caractère est une information dont on recueille (ou observe ou mesure) la **valeur** sur chaque individu. C'est en quelque sorte l'objet d'étude sur la population.

Sur les machines à dactylographier de la salle on peut étudier : la marque, le nombre de touches, la durée de vie... qui sont des exemples de caractères. Le caractère est aussi appelé variable parce que la valeur de l'information n'est pas la même d'un individu à l'autre.

4. Modalité

Une modalité est une valeur du caractère.

Par rapport à un caractère donné, une modalité est la situation dans laquelle se trouve un individu. Une machine de la salle peut porter la marque IBM. Si le caractère étudié est la marque, alors IBM est une modalité de ce caractère.

La modalité est généralement notée x_i .

5. Effectif d'une modalité

Statistiques descriptives

L'effectif d'une modalité est le nombre de fois qu'apparaît cette modalité. Si trois machines et trois seulement de la salle portent la marque IBM, alors l'effectif de la modalité IBM est 3 machines. L'effectif est aussi appelé fréquence absolue.

6. Effectif total

L'effectif « **N** » d'une population est le nombre total d'individus de cette population. Pour notre exemple, c'est le nombre total de machines dans la salle.

Si $n_1, n_2, n_3, \dots, n_k$ sont les effectifs des différentes modalités d'un caractère étudié sur une population, alors $N = n_1 + n_2 + n_3 + \dots + n_k$.

On écrit plus simplement : $N = \sum_{i=1}^k n_i$

7. Fréquence relative d'une modalité

C'est le rapport de l'effectif n_i de la modalité i par l'effectif total **N**.

Elle est notée f_i : $f_i = \frac{n_i}{N}$

On montre que $f_1 + f_2 + f_3 + \dots + f_k = 1$

NB : la fréquence relative peut être donnée en pourcentage.

Dans ce cas : $f_i = \frac{n_i}{N} \times 100$ et $f_1 + f_2 + f_3 + \dots + f_k = 100$.

8. Caractère qualitatif

Un caractère est dit qualitatif lorsque ses modalités ne sont pas quantifiables (mesurables).

Le caractère « marque » par exemple est un caractère qualitatif parce que ses modalités (IBM, Samsung, HP...) ne sont pas mesurables.

9. Caractère quantitatif

Un caractère quantitatif est un caractère dont chaque modalité est une valeur mesurable.

Les caractères nombre de touches et durée de vie sont des caractères quantitatifs.

Un caractère quantitatif peut être **discret** ou **continu**.

10. Caractère quantitatif discret ou variable discrète

Une variable discrète est une variable dont les modalités ne peuvent prendre que des valeurs bien déterminées ou isolées.

Le caractère nombre de touches est un caractère discret. En effet le nombre de touches d'une machine est un nombre bien déterminé : 30 touches, 20 touches, 45 touches...

11. Variable continue

C'est une variable dont les valeurs des modalités décrivent un intervalle de $\mathbb{R}$. En d'autres termes, les modalités d'un caractère quantitatif continu sont susceptibles de prendre toutes les valeurs d'un intervalle de $\mathbb{R}$.

Le caractère « durée de vie » d'une machine à dactylographier est une variable continue.

Quand l'intervalle des modalités est trop grand, on le découpe en plusieurs petits intervalles appelés classes. Le regroupement d'une série statistique en classe n'est pas unique. Il peut différencier d'un statisticien à l'autre.

12. Classe

Une classe statistique est un intervalle de modalités.

Pour l'exemple du caractère « durée de vie », on peut avoir la classe « de 2 à 4 ans ». Dans un tableau statistique, on choisira très souvent d'écrire $[2; 4[$ ou $[2; 4]$.

Statistiques descriptives

De même qu'on parle d'effectif d'une modalité, on parle aussi d'effectif d'une classe qui en fait n'est rien d'autre que le nombre d'individus dont les modalités sont contenues dans cette classe. On calcul alors la fréquence relative d'une classe de la même manière qu'en 7).

13. Centre de classe

Le centre c_i d'une classe $[b_i; b_{i+1}[$ ou $]b_i; b_{i+1}]$ ou $]b_i; b_{i+1}[$ d'origine b_i et d'extrémité b_{i+1} où i est un entier naturel est la demi-somme des bornes :

$$c_i = \frac{b_i + b_{i+1}}{2}$$

14. Amplitude

L'amplitude a_i d'une classe d'origine b_i et d'extrémité b_{i+1} est la différence entre la borne supérieure et la borne inférieure: $a_i = b_{i+1} - b_i$

15. Enquête

Par rapport à un caractère donné, une enquête est une collecte d'informations. L'enquête peut être exhaustive ou partielle.

16. Recensement

Un recensement est une enquête qui porte sur tous les individus de la population. On l'appelle aussi enquête exhaustive.

17. Sondage

Le sondage ou enquête partielle porte sur un échantillon ou une portion de la population.

18. Dépouillement

Le dépouillement d'une enquête sur une population est une opération de décompte des individus en fonction des modalités ou des classes.

Exemple de dépouillement

On a interrogé 20 élèves d'une classe de BT secrétariat du CBCG de Cocody au sujet du nombre de cours auxquels chacun a pu assister avant les congés de Noël 2012. Les résultats obtenus sont les suivants :

2 ; 2 ; 1 ; 3 ; 5 ; 1 ; 5 ; 3 ; 4 ; 4 ; 2 ; 7 ; 3 ; 5 ; 5 ; 8 ; 9 ; 5 ; 5 ; 10.

Tableau de dépouillement

Nombre de cours (x_i)	Dépouillement	Effectif d'élèves (n_i)
2		3
1		2
3		3
5		6
4		2
7		1
8		1
9		1
10		1

19. Tableau statistique

Un tableau statistique est un tableau récapitulatif de d'une enquête. On distingue deux types de tableaux :

Statistiques descriptives

🌟 Tableau à simple entrée

Répartition de 20 élèves d'une classe de BT secrétariat du CBCG de Cocody selon le nombre de cours auxquels chacun a pu assister avant les congés de Noël 2012.

Modalités	1	2	3	4	5	6	7	8	9	10	Total
Effectifs	2	3	3	2	6	0	1	1	1	1	20

🌟 Tableau à double entrée

Répartition de 60 élèves selon le sexe et le nombre de stylos que possède chaque individu.

Nombre de stylos \ Sexe	Sexe		
	Masculin	Féminin	Total
1	6	3	09
2	7	5	12
3	9	7	16
4	13	10	23
Total	35	25	60

II. SERIES STATISTIQUES A UNE VARIABLE

1. Tableaux statistiques

Exemple 1 : Cas d'une variable discrète :

Tableau donnant la répartition de 40 ouvriers de la société D. A. D. selon le nombre d'enfants scolarisés.

i	Nombre d'enfants scolarisé x_i	Effectifs n_i (nombre d'ouvriers)	Fréquences f_i (en %)	Effectifs cumulés		Fréquences cumulées	
				Croissants	Décroissant	Croissantes	Décroissantes
1	0	3	7,5	3	40	7,5	100
2	1	5	12,5	8	37	20	92,5
3	2	16	40	24	32	60	80
4	3	8	20	32	16	80	40
5	4	6	15	38	8	95	20
6	5	2	5	40	2	100	5
	Total	40	100				

Interprétation de quelques éléments du tableau.

Questions :

- Quelle est la proportion d'ouvriers ayant moins de 3 enfants scolarisés ?
- Quelle est la proportion d'ouvriers ayant au plus 1 enfant scolarisé ?
- Quelle est la proportion d'ouvriers ayant au moins 4 enfants scolarisés ?
- Quelle est la proportion d'ouvriers ayant plus de 3 enfants scolarisés ?
- Combien d'ouvriers scolarisent plus de 2 enfants ? Au plus 2 enfants ?
- Combien d'ouvriers scolarisent 5 enfants ?

Réponses :

- La proportion d'ouvriers ayant moins de 3 enfants scolarisés est de 60%.
- La proportion d'ouvriers ayant au plus 1 enfant scolarisé est de 20%.
- La proportion d'ouvriers ayant au moins 4 enfants scolarisés est de 95%.
- La proportion d'ouvriers ayant plus de 3 enfants scolarisés est de 20%.

Statistiques descriptives

16 ouvriers scolarisent plus de 3 enfants. 24 ouvriers scolarisent au plus 2 enfants.
2 ouvriers scolarisent 5 enfants.

Exemple 2 : cas d'une variable continue

La répartition de candidats à un test de présélection suivant le total des points obtenus.

i	Total obtenu	Centre de classe c_i	Effectifs n_i	Effectifs cumulés		Fréquences cumulées (%)	
				Croissants	Décroissants	Croissantes	Décroissantes
1	[20;30[	25	27	27	160	16,875	100,000
2	[30;40[	35	43	70	133	43,750	83,125
3	[40;50[	45	38	108	90	67,500	56,250
4	[50;60[	55	28	136	52	85,000	32,500
5	[60;70[	65	21	157	24	98,125	15,000
6	[70;80[	75	3	160	3	100,000	1,875
	Total		160				

Questions

Quel est le nombre de candidats dont le total des points obtenu est :

- a) Moins de 40 ?
- b) Au moins 60 ?
- c) Entre 50 et 60 ?

Réponses :

- a) 70 candidats ont moins de 40 points au total.
- b) 24 candidats ont obtenu au moins 60 points.
- c) 28 candidats ont obtenu un total de points entre 50 et 60.

Exercice

80 employés d'une agence de tourisme, répartis suivant le nombre d'enfants à charge sont soumis à la fin de l'année à une rétribution de prime à raisons de 2 500F par enfants à charge.

Le tableau incomplet de la répartition se présente comme suit :

i	Nombre d'enfants à charge	Effectifs (n_i)	Fréquence (f_i)	Fréquence cumulées décroissantes	Fréquence cumulées croissantes
1	0				
2	1		0,1875	0,95	
3	2	29			
4	3		0,225		
5	4	10			
6	5		0,0375		
7	6			0,0125	

Questions :

- 1) Complétez le tableau.
- 2) Quel est le nombre d'employés ayant :
 - a- 0 enfant à charge ?

Statistiques descriptives

- b- Moins de 4 enfants à charge ?
- c- Au moins 3 enfants à charge ?
- 3) Quel est le montant global des primes perçues par les employés ayant :
 - a- Entre 2 et 5 enfants inclus
 - b- Au plus 2 enfants ?
 - c- Plus de 4 enfants ?

Réponses :

- 1) Utiliser le tableau précédent.
- 2) le nombre d'employés ayant :
 - a-
 - b-
 - c-
- 3) le montant global des primes perçues par les employés ayant :
 - a-
 - b-
 - c-

2. Représentations graphiques

Un graphique est une synthèse visuelle de toutes les informations recueillies dans un tableau statistique. Quoique moins précis qu'un tableau de valeurs statistiques, le graphique offre une vue d'ensemble de la répartition spatiale ou chronologique de la population. Suivant la nature du caractère étudié, on utilise différents modes de représentation.

a) Diagramme à bandes

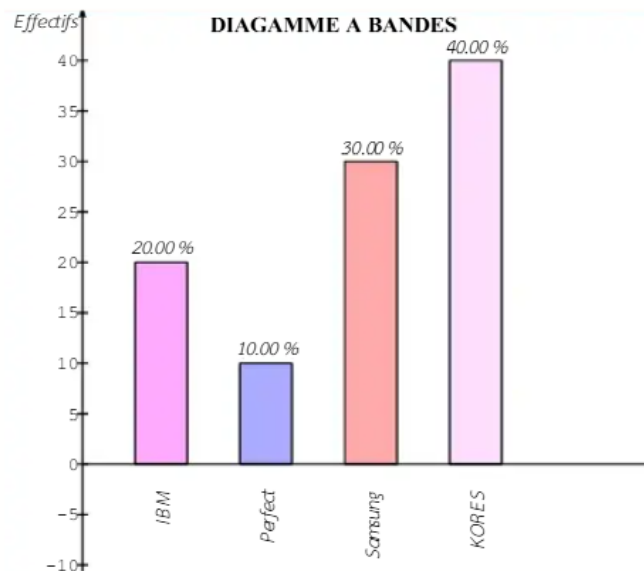
Le diagramme à bande est utilisé pour les caractères qualitatifs. Il se construit de la façon suivante :

- ✓ On construit deux axes perpendiculaires (abscisse et ordonnée)
- ✓ Sur l'axe des abscisses on place les différentes modalités
- ✓ Sur l'axe des ordonnées on place les effectifs n_i ou les fréquences relatives f_i .
- ✓ On associe à chaque modalité une bande verticale dont la hauteur est proportionnelle à l'effectif de la modalité.

Exemple

Dans un magasin est stockées 50 machines à dactylographier. La répartition de ces machines selon leur marque est consignée dans le tableau suivant :

Marques	Effectif
IBM	10
Perfect	5
Samsung	15
KORES	20
Total	50



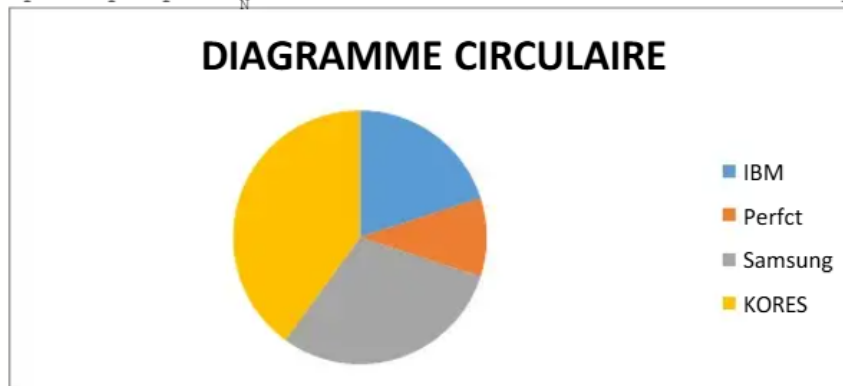
Statistiques descriptives

b) Diagramme à secteurs

Le diagramme circulaire ou à secteur est aussi utilisé pour les caractères qualitatifs. Il se construit de la façon suivante :

L'effectif total est représenté par un cercle ou par un demi-cercle. Chaque modalité est représentée par un secteur circulaire que l'on peut différencier par des couleurs ou des nuances différentes et dont la surface est proportionnelle à l'effectif correspondant. Ainsi, l'angle au centre θ_i associé à la modalité i est égal au produit de la fréquence f_i par 360° ou 180° .

$\theta_i = 360f_i \Leftrightarrow \theta_i = 360 \frac{n_i}{N}$ ou bien dans le cas d'un demi-cercle $\theta_i = 180f_i \Leftrightarrow \theta_i = 180 \frac{n_i}{N}$.



c) Diagramme en bande

Le diagramme en bande se construit de la façon suivante :

On construit une barre verticale de hauteur H . Chaque modalité est représentée par une portion de la barre dont la surface est proportionnelle à l'effectif.



d) Diagramme en bâtons

Le diagramme en bâtons est utilisé pour les caractères quantitatifs discrets.

On marque sur l'axe des abscisses les modalités et sur l'axe des ordonnées, les effectifs ou fréquences. En chaque point de coordonnées $(x_i; n_i)$ ou $(x_i; f_i)$, obtenu, on construit un segment (bâton) parallèle à l'axe des ordonnées jusqu'à l'axe des abscisses. On obtient ainsi un diagramme en bâtons.

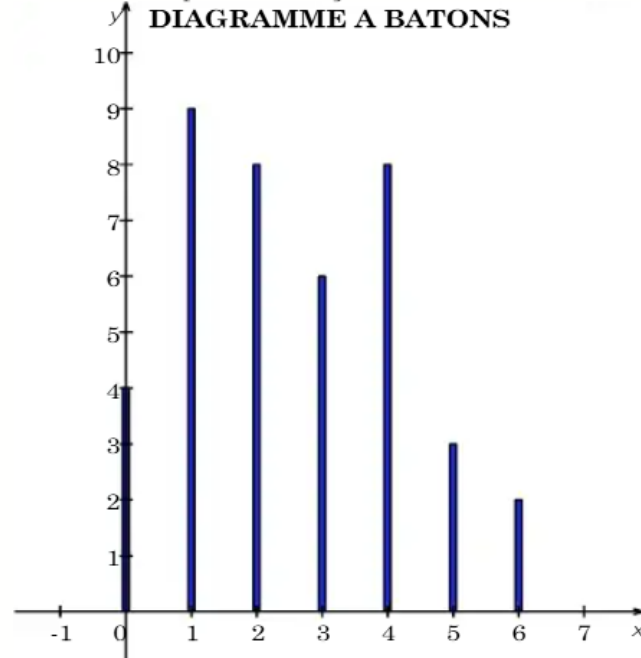
Exemple :

Soit la distribution statistique suivante :

Statistiques descriptives

Modalités x_i	0	1	2	3	4	5	6	Total
Effectifs n_i	4	9	8	6	8	3	2	40

Représentons cette série par un diagramme à bâtons



3. Paramètres de position

Les paramètres de position permettent de voir les modalités autour desquelles sont regroupés les individus. Ils sont souvent appelés caractéristiques de tendance centrale.

a) Le mode

• Définition

Le mode m_0 (ou dominante) d'une série statistique est la valeur du caractère qui a l'effectif (ou la fréquence) maximal(e).

Il peut exister plusieurs modes dans une même série.

Lorsque la variable est continue, on définit la classe modale comme étant la classe dont l'effectif (ou l'effectif corrigé en cas d'amplitude inégales) est le plus élevé. On admet alors que le mode est le centre de cette classe.

• Détermination

✓ Cas d'une variable discrète

Un mode se lit aisément dans le tableau des valeurs (ou sur le diagramme à bâtons)

Exemple : cas de la répartition de 80 employés d'une agence de tourisme selon le nombre d'enfants à charge.

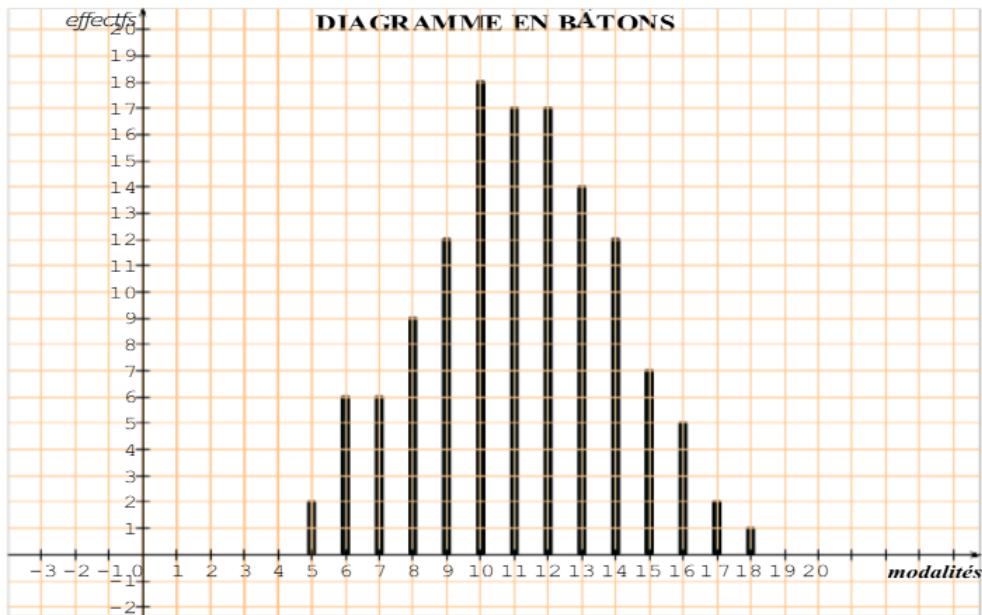
Nombre d'enfants à charge	0	1	2	3	4	5	6
Effectifs d'employés	4	15	29	18	10	3	1

Ici, l'effectif le plus élevé est **29**. Il correspond à la modalité **2 enfants à charge**. Donc le mode de cette série est 2 enfants à charge.

Exercice

Soit le graphique suivant donnant la distribution de 128 étudiants selon la note obtenue à l'issu d'un devoir commun de statistiques.

Statistiques descriptives



Quel est le mode cette distribution ?

✓ Cas d'une variable continue

Exemple 1 : Tableau de distribution de 25 élèves selon la taille.

Taille en mètre	[1,40 ; 1,50[	[1,50 ; 1,60[	[1,60 ; 1,70[	[1,70 ; 1,80[	Total
Effectifs	3	7	10	5	25

On remarque ici que toutes les classes ont la même amplitude $a=0,1\text{ m}$.

La classe ayant l'effectif le plus élevé est la classe $[1,60 ; 1,70[$. Le mode

m_0 est donc : $m_0 = \frac{1,60 + 1,70}{2}$ soit $m_0 = 1,65\text{ m}$

Exemple 2 : Tableau de distribution de 184 véhicules endommagés suivant la distance parcourue.

Classe (distance en 10^3 Km)	[80;85[	[85;90[	[90;100[	[100;110[	[110;125[	[125;150[	[150;200[	Total
Effectifs (n_i)	10	14	60	44	20	28	8	184
Amplitudes (a_i)	5	5	10	10	15	25	50	
Intervalle unitaires (UI)	1	1	2	2	3	5	10	
Effectifs corrigés $n_i' = \frac{n_i}{UI}$	10	14	30	22	6,6	5,6	0,8	

NB : l'effectif corrigé n_i' est donné par : $n_i' = \frac{n_i}{UI} = \frac{n_i}{a_i} \times 5$ où $UI = \frac{a_i}{5}$

Statistiques descriptives

Ici les classes n'ont pas les mêmes amplitudes. La classe ayant l'effectif corrigé le plus élevé 30, est la classe $[90;100[$. Le mode est donc

$m_o = 95\,000\text{ km}$.

b) La moyenne arithmétique

• Définition

Considérons une série statistique de modalités $x_1, x_2, x_3, \dots, x_p$ dont les effectifs respectifs sont : $n_1, n_2, n_3, \dots, n_p$.

La moyenne arithmétique de cette série est le nombre réel noté $\bar{X}$ tel que :

$$\bar{X} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n_1 + n_2 + \dots + n_p} = \frac{1}{N} \sum_{i=1}^p n_i x_i$$

NB : Lorsque l'on ne fait pas mention des effectifs dans un exercice, on suppose que $n_1 = n_2 = n_3 = \dots = n_p = 1$. Dans ce cas si la population totale est

N , on calcule $\bar{X}$ en faisant : $\bar{X} = \frac{x_1 + x_2 + \dots + x_p}{n_1 + n_2 + \dots + n_p} = \frac{1}{N} \sum_{i=1}^p x_i$

• Détermination

Cas de la répartition de 80 employés d'une agence selon le nombre d'enfants à charge.

La moyenne arithmétique de la série correspondante est :

$$\bar{X} = \frac{1}{80} \sum_{i=1}^7 n_i x_i \Rightarrow \bar{X} = \frac{1}{80} (4 \times 0 + 15 \times 1 + 29 \times 2 + 18 \times 3 + 10 \times 4 + 3 \times 5 + 1 \times 6) \Rightarrow \bar{X} = 2,35$$

A noter : Lorsque les modalités sont regroupées en classes, on calcule la moyenne en utilisant les centres de classe.

Exercices :

- Les notes obtenues par 128 élèves à un devoir sont consignées dans le tableau suivant :

Modalités	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total
Effectifs	2	6	6	9	12	18	17	17	14	12	7	5	2	1	128

Calculer la note moyenne des élèves ayant participé au devoir.

- Calculer la moyenne correspondante à la série donnant la répartition de 184 véhicules endommagés selon la distance parcourue.
- Même exercice pour le tableau donnant la répartition de 25 élèves selon la taille.

Interprétation de la moyenne

La moyenne est la valeur qu'auraient toutes les observations (ou modalités) si elles étaient égales. Elle correspond à un partage du total des modalités en parties égales.

Pour l'exercice a) précédent, une moyenne de 9,36 signifie si tous les élèves avaient eu la même note, cette note serait 9,36.

c) La médiane

• Définition

Soit une série statistique dont les modalités sont rangées en ordre croissant ou décroissant.

On appelle médiane de cette série la modalité qui la partage en deux parties de même effectif.

• Détermination.

✓ Cas d'un caractère continu

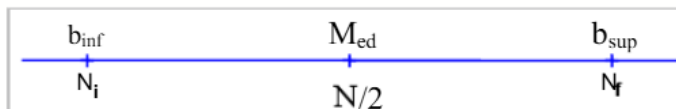
Statistiques descriptives

On détermine à l'aide des effectifs (ou fréquences) cumulé(e)s croissant(e) ou décroissant(e) la classe contenant la médiane. Par la suite on calcule la médiane par interpolation linéaire.

Exemple : cas de la distribution de 184 véhicules endommagés suivant la distance parcourue.

Classe (distance en 10^3 Km)	[80;85[	[85;90[	[90;100[	[100;110[	[110;125[	[125;150[	[150;200[	Total
Effectifs	10	14	60	44	20	28	8	184
Effectifs cumulé croissants	10	24	84	128	148	176	184	

$N=184 \Rightarrow \frac{N}{2}=92$. 92 est compris entre 84 et 128. On suppose alors que la médiane est contenue dans la classe correspondante à l'effectif cumulé 128. Soit [100;110[on admet que $F(100)=84$ et $F(110)=128$.



$$\frac{M_{ed} - b_{inf}}{b_{sup} - b_{inf}} = \frac{\frac{N}{2} - N_i}{N_f - N_i}$$

$$M_{ed} = \frac{\frac{N}{2} - N_i}{N_f - N_i} (b_{sup} - b_{inf}) + b_{inf}$$

Application numérique

$$M_{ed} = \frac{92 - 84}{128 - 84} (110 - 100) + 100 \Rightarrow M_{ed} = 101,82 \cdot 10^3 \text{ km}$$

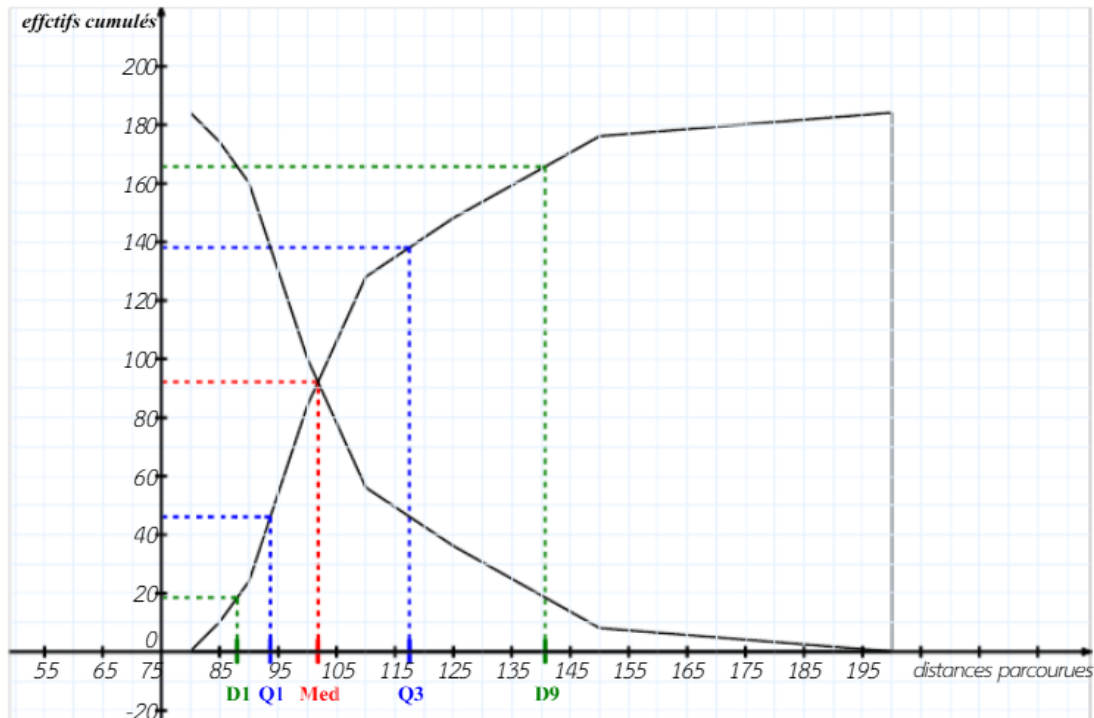
Remarque :

La détermination de la médiane se fait aussi à l'aide des polygones des effectifs (ou des fréquences) cumulé (e)s croissant(e)s ou décroissant(e)s.

C'est l'abscisse du point d'ordonnées $\frac{N}{2}$ en effectif cumulé (ou 50 % en fréquence cumulée) en considérant chaque polygone. C'est aussi le point d'intersection des deux polygones.

En guise d'exemple de déterminer graphiquement la médiane de la série relative à la distribution de 184 véhicules endommagés suivant la distance parcourue.

Statistiques descriptives



• Cas d'une variable discrète

On suppose que les modalités sont rangé par ordre croissant (ou décroissant).

Cas (m_1) : Si l'effectif total N est impair, alors la médiane est la

$\left(\frac{N+1}{2}\right)^{\text{ème}}$ modalité.

Exemple : La médiane de cet ensemble de 9 modalités {2 ; 3 ; 4 ; 6 ; 8 ;

9 ; 11 ; 14 ; 15} est 8. Car 8 est la $\left(\frac{9+1}{2}\right)^{\text{ème}}$ (c'est-à-dire la 5^{ème})

modalité.

Cas (m_2) : Si l'effectif total N est pair, alors la médiane est entre la

$\left(\frac{N}{2}\right)^{\text{ème}}$ et la $\left(\frac{N+1}{2}\right)^{\text{ème}}$ modalité.

Lorsque les modalités son pondérées par des effectifs, on attribue la médiane à la modalité correspondante à l'effectif cumulé croissant le plus proche de $\frac{N}{2}$ mais supérieur à $\frac{N}{2}$.

Exemple : cas de la répartition de 80 employés d'une agence de tourisme selon le nombre d'enfants à charge.

Nombre d'enfants à charge	0	1	2	3	4	5	6	Total
Effectifs d'employés	4	15	29	18	10	3	1	80
Effectifs cumulés croissants	4	19	48	66	76	79	80	

$\frac{N}{2} = \frac{80}{2} = 40$ soit $19 < \frac{N}{2} < 48$. On prendra la modalité correspondante à

l'effectif cumulé 48 comme médiane. Soit $M_{ed} = 2$ enfants à charge .

Statistiques descriptives

NB : Si $\frac{N}{2}$ coïncide avec un effectif cumulé, alors on retrouve dans le cas (m_2).

Interprétation de la médiane

Exemple de la distribution de 184 véhicules endommagés suivant la distance parcourue.

$M_{ed} = 101,82 \cdot 10^3$ km le nombre de véhicules ayant parcouru moins de $101,82 \cdot 10^3$ km est égal au nombre de véhicules ayant parcouru au moins $101,82 \cdot 10^3$ km.

d) Les quantiles

On parle de quantiles lorsqu'une série statistique est partagée en k parties de même effectif.

- Lorsque $k=4$, les quantiles sont appelés quartiles. Les quartiles sont justement les trois valeurs de la série qui partagent les observations rangées par valeur croissante ou décroissantes, en quatre groupes de même effectif.

On les note généralement Q_1, Q_2 , et Q_3 . Se sont les modalités qui

correspondent respectivement à $\frac{N}{4}, \frac{2N}{4}$ et $\frac{3N}{4}$.

- Lorsque $k=10$, les quantiles sont appelés déciles. Les déciles sont les neuf valeurs de la série qui partagent les observations rangées par valeur croissante ou décroissantes, en dix groupes de même effectif.

On les note généralement D_1, D_2, D_3, D_9 et sont les modalités qui

correspondent respectivement à $\frac{N}{10}, \frac{2N}{10}, \frac{3N}{10}, \dots, \frac{9N}{10}$.

- Lorsque $k=100$, les quantiles sont appelés centiles. Les centiles sont les 99 valeurs de la série qui partagent les observations rangées par valeur croissante ou décroissantes, en cent groupes de même effectif.

On les note généralement $C_1, C_2, C_3, \dots, C_{99}$. Se sont les modalités qui

correspondent respectivement à $\frac{N}{100}, \frac{2N}{100}, \frac{3N}{100}, \dots, \frac{99N}{100}$.

NB : les quantiles se déterminent de la même manière que la médiane.

Exercice :

Pour la série donnant la répartition de 128 élèves selon les notes... déterminer les quartiles et les déciles puis interpréter ces paramètres.

4. Paramètres de dispersion

Les paramètres de position d'une série (mode, moyenne, médiane, quantiles) donnent une idée sommaire de la distribution des observations, mais ils ne suffisent pas à la caractériser.

Pour avoir une parfaite connaissance de cette distribution, il faut voir comment les valeurs de la série sont réparties autour des paramètres de position. Ce sont des paramètres de dispersion qui permettent une telle connaissance.

a. L'étendue

On appelle étendue d'une série statistique, la différence entre la plus grande et la plus petite valeur que peut prendre le caractère. Elle est généralement notée e

Exemple 1 :

Cas de la répartition de 80 employés d'une agence de tourisme selon le nombre d'enfants à charge (Sériel).

Statistiques descriptives

Nombre d'enfants à charge	0	1	2	3	4	5	6	Total
Effectifs d'employés	4	15	29	18	10	3	1	80

L'étendue de cette série est : $e = 6 - 0 \Rightarrow e = 6$. Soit 6 enfants à charge.

Exemple 2 :

Tableau de distribution de 25 élèves selon la taille (Série 2).

Taille en mètre	[1,40 ; 1,50[	[1,50 ; 1,60[	[1,60 ; 1,70[	[1,70 ; 1,80[	Total
Effectifs	3	7	10	5	25

L'étendue de cette série est : $e = 1,80 - 1,40 = 0,40$. Soit 0,40 m.

b. L'écart absolu moyen

On appelle écart absolu moyen d'une série statistique, la moyenne des valeurs absolues des écarts à la moyenne arithmétique $\bar{X}$; On le note e_m .

On considère la série dont les observations (ou les centres de classe)

$x_i = \{x_1, x_2, x_3, \dots, x_k\}$ ont pour effectifs $n_i = \{n_1, n_2, n_3, \dots, n_k\}$. L'écart absolu moyen de cette série est alors :

$$e_m = \frac{1}{N} \sum_{i=1}^k n_i |x_i - \bar{X}| \quad N \text{ est l'effectif total.}$$

$$e_m = \sum_{i=1}^k f_i |x_i - \bar{X}| \quad \text{Avec } f_i = \frac{n_i}{N}$$

Interprétation :

Plus l'écart absolu moyen est élevé, plus la série est dispersée.

Exemple 1 :

Calcul de l'écart absolu moyen de la série 1 ($\bar{X} = 2,35$)

x_i	0	1	2	3	4	5	6	Total
n_i	4	15	29	18	10	3	1	80
$ x_i - \bar{X} $	2,35	1,35	0,35	0,65	1,65	2,65	3,65	
$n_i x_i - \bar{X} $	9,4	20,25	10,15	11,7	16,5	7,95	3,65	76,6

$$e_{m1} = \frac{1}{80} \sum_{i=1}^7 n_i |x_i - \bar{X}| = \frac{1}{80} (76,6) = 0,9575 \text{ avec } \bar{X} = 2,35$$

Exemple 2 :

Calcul de l'écart absolu moyen de la série 2

Taille en mètre	[1,40 ; 1,50[	[1,50 ; 1,60[	[1,60 ; 1,70[	[1,70 ; 1,80[	Total
Centre de classe x_i	1,45	1,55	1,65	1,75	
Effectifs n_i	3	7	10	5	25
$n_i x_i$	4,35	10,85	16,50	8,75	40,45
$n_i x_i - \bar{X} $	0,504	0,476	0,320	0,660	1,96

$$\bar{X} = \frac{1}{25} (40,45) \Rightarrow \bar{X} = 1,618$$

$$e_{m2} = \frac{1}{25} (1,96) \Rightarrow e_{m2} = 0,0784. \text{ Soit environ } 0,08 \text{ m.}$$

c. La Variance :

La variance est la moyenne arithmétique des carrés des écarts des valeurs de la variable à leur moyenne arithmétique.

Pour une série dont les observations (ou les centres de classe)

$x_i = \{x_1, x_2, x_3, \dots, x_k\}$ ont pour effectifs $n_i = \{n_1, n_2, n_3, \dots, n_k\}$ La variance est :

Statistiques descriptives

$$V(X) = \frac{1}{N} \sum_{i=1}^k n_i (x_i - \bar{X})^2 \quad \text{ou} \quad V(X) = \sum_{i=1}^k f_i (x_i - \bar{X})^2$$

On démontre la formule de **Koenigs** suivante:

$$V(X) = \left(\frac{1}{N} \sum_{i=1}^k n_i x_i^2 \right) - \bar{X}^2 \quad \text{ou} \quad V(X) = \left(\sum_{i=1}^k f_i x_i^2 \right) - \bar{X}^2$$

Exemple1 :

Calcul de la variance de la série 1 ($\bar{X}=2,35$)

x_i	0	1	2	3	4	5	6	Total
n_i	4	15	29	18	10	3	1	80
$(x_i - \bar{X})^2$	5,5225	1,8225	0,1225	0,4225	2,7225	7,0225	13,3225	
$n_i(x_i - \bar{X})^2$	22,09	27,3375	3,5525	7,605	27,225	21,0675	13,3225	122,2
$n_i x_i^2$	0	15	116	162	160	75	36	564

$$V(X) = \frac{1}{80} \sum n_i (x_i - \bar{X})^2 = \frac{1}{80} (122,2) = 1,5275 \text{ avec } \bar{X} = 2,35.$$

Ou selon la formule de Koenigs :

$$V(X) = \left(\frac{1}{80} \sum n_i x_i^2 \right) - \bar{X}^2 = \frac{1}{80} (564) - 2,35^2 = 1,5275.$$

Exemple2 :

Calcul de la variance de la série 2 ($\bar{X}=1,618$)

Taille en mètre	[1,40 ; 1,50[	[1,50 ; 1,60[	[1,60 ; 1,70[	[1,70 ; 1,80[	Total
Centre de classe x_i	1,45	1,55	1,65	1,75	
Effectifs n_i	3	7	10	5	25
$n_i x_i$	6,3075	16,8175	27,225	15,3125	65,6625

$$V(X) = \left(\frac{1}{25} \sum n_i x_i^2 \right) - \bar{X}^2. \text{ Soit } V(X) = \left(\frac{1}{25} \cdot 65,6625 \right) - 1,618^2 \Rightarrow V(X) = 0,008576.$$

d. L'écart type

L'écart type est la racine carrée de la variance. Il est noté σ_x pour la variable X et on a : $\sigma_x = \sqrt{V(X)}$

Interprétation :

Plus l'écart type est élevé, plus la série est dispersée.

Exemple

L'écart type de la série 2 est : $\sigma_x = \sqrt{0,008576}$. Soit $\sigma_x = 0,0926$

e. Ecart Interquartile

L'écart interquartile est la différence entre les deux quantiles extrêmes d'une série.

Ainsi l'écart interquartile est la différence entre le 3^{ème} quartile et le 1^{er} quartile.

Il vaut : $Q_3 - Q_1$. Il mesure l'amplitude de l'intervalle $[Q_3 ; Q_1]$ qui contient 50% des informations.

Statistiques descriptives

Plus $Q_3 - Q_1$ est petite, moins les valeurs de la série sont dispersées (plus les valeurs de la série sont concentrées) ; plus $Q_3 - Q_1$ est grand, plus les valeurs de la série sont dispersées.

Exemple

Pour la série 2, admettons que le premier quartile est $Q_1 = 1,55$ et le troisième est $Q_3 = 1,69$. L'intervalle interquartile est alors : $Q_3 - Q_1 = 1,69 - 1,55$, soit 0,14 m.

(les valeur de la série sont moins dispersées car l'écart interquartile est beaucoup plus faible que l'étendue qui est 0,40 m)

NB : on peut comparer l'écart interquartile et la moitié de l'étendue. En effet

$Q_3 - Q_1 < \frac{e}{2}$ nous fait pensée à une série moins dispersée.

f. Coefficient de Variation

Considérons la moyenne arithmétique $\bar{X}$ et l'écart type σ_x d'une population statistique.

On appelle coefficient de variation notée C_v , le rapport de l'écart type σ_x par la moyenne arithmétique $\bar{X}$: $C_v = \frac{\sigma_x}{\bar{X}}$

L'intérêt du coefficient de variation réside dans la possibilité de comparaison de deux séries statistiques énoncées dans des unités de mesure différentes.

Exemple

Le coefficient de variation de la série 2 est : $C_v = \frac{0,0926}{1,618} = 0,0572$.

En général les paramètres de dispersion trouvent leur importance dans la comparaison de plusieurs séries.

Chapitre 2

LES DISTRIBUTIONS A DEUX CARACTERES**Définition :**

P est une population de N individus, sur laquelle sont définis deux caractères X et Y. Sur chaque individu i de P on observe une modalité X_i du caractère X et une modalité Y_i du caractère Y. On appelle **série statistique double** de caractères (X ; Y) l'ensemble des couples de modalités $(X_i ; Y_i)$.

I. NUAGE DE POINTS**1. Définition**

Considérons une population statistique à n éléments pour laquelle on se propose d'étudier simultanément deux caractères X et Y.

Population n	1 ^{er}	2 ^{ème}	3 ^{ème}	...	$i^{\text{ème}}$	...	$n^{\text{ème}}$
Caractère X	X_1	X_2	X_3	...	X_i	...	X_n
Caractère Y	Y_1	Y_2	Y_3	...	Y_i	...	Y_n

Dans le plan muni d'un repère orthogonal (O, I, J), on associe à chaque couple $(x_i; y_i)$ un point M_i de coordonnées $(x_i; y_i)$ représentant le $i^{\text{ème}}$ individu de la population.

On appelle nuage de points, l'ensemble des n points M_i dont les coordonnées $(x_i; y_i)$ sont des mesures correspondantes des deux variables statistiques X et Y.

2. Représentation

La représentation graphique d'un nuage de points se fait comme dans l'exemple suivant :

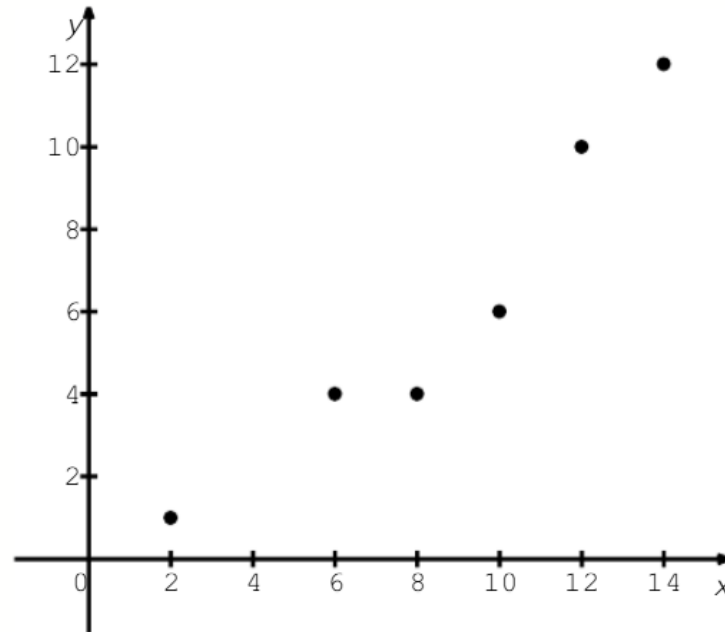
Exemple :

Soit la série statistique suivant :

X	2	6	8	10	12	14
Y	1	4	4	6	10	12

Représenter le nuage de points associé à cette série

Statistiques descriptives



3. Point Moyen

On appelle point moyen d'un nuage de n points de coordonnées $(x_i; y_i)$, le point $G(\bar{X}; \bar{Y})$ tel que $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$ et $\bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i$

Exemple : déterminer les coordonnées du point moyen associé à la série précédente. Le placer dans le même repère que le nuage de points.

Eléments de réponse : $\bar{X} = 8,667$ et $\bar{Y} = 6,167$

II. AJUSTEMENT AFFINE PAR LA METHODE DE MAYER

1. Principe de la méthode de Mayer

On suppose que les modalités X_i sont rangées dans l'ordre croissant.

Soit n l'effectif total de la population.

- cas d'un effectif pair $n = 2p$, $p \in \mathbb{N}^*$

On désigne par E_1 l'ensemble des p premiers points et par G_1 son point moyen.

On désigne par E_2 l'ensemble des p derniers points et par G_2 son point moyen

- cas d'un effectif impair $n = 2p+1$, $p \in \mathbb{N}^*$

L'un des sous-nuages comportera un point de plus que l'autre.

2. Équation de la droite de Mayer

Dans les deux cas, la droite de Mayer est la droite (G_1G_2) d'équation réduite $y = ax + b$

Statistiques descriptives

$$\text{Avec } a = \frac{y_{G_2} - y_{G_1}}{x_{G_2} - x_{G_1}} \text{ et } b = y_{G_1} - a x_{G_1}$$

Application

On a interrogé 4 élèves au sujet de leur argent de poche X (en FCFA) par jour et de leur temps de navigation Y(en minutes) sur Internet, par semaine. Les informations obtenues sont consignées dans le tableau suivant:

x_i	600	800	1000	1200
y_i	120	280	540	660

- Partager cette série en deux sous-séries.
- Soit G_1 le point moyen associé à la sous-série (S_1) et G_2 le point moyen associé à la sous-série(S_2). Calculez les coordonnées de G_1 et de G_2 , et déterminez par la méthode de Mayer, l'équation de la droite d'ajustement du nuage de points associé à la série étudiée.

Eléments de réponse : G_1 (700 ; 200), G_2 (1100 ; 600), $y = x - 500$.

III. AJUSTEMENT PAR LA METHODE DES MOINDRES CARRES

1. Définition

L'ajustement par la méthode des moindres carrés est l'ajustement obtenu en minimisant la somme des carrés des écarts entre les valeurs calculées d'une variable et ses valeurs observées.

2. Coefficient de corrélation linéaire

Il n'est pas toujours facile de dire qu'un ajustement linéaire est justifié ou non en observant tout simplement un nuage de points. D'ailleurs notre vue peut nous tromper si nous n'avons pas d'élément scientifique pour confirmer ce que nous constatons sur le graphique. D'où l'utilité et l'importance du coefficient de corrélation linéaire.

• La covariance :

Soit une série statistique à deux variables (X,Y) .

On considère le cas où l'effectif de chaque couple de modalité $(x_i; y_i)$ est égal à 1.

$\bar{X}$ est la moyenne des x_i et $\bar{Y}$ celle des y_i .

La covariance de la série (X,Y) est le nombre réel noté $\text{Cov}(X,Y)$ tel que

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

• Le coefficient de corrélation linéaire

Le coefficient de corrélation linéaire d'une série statistique à deux

variables (X,Y) est le nombre réel noté r défini par $r = \frac{\text{Cov}(X,Y)}{\sqrt{v(X) \cdot v(Y)}}$

On démontre que : $-1 \leq r \leq 1$

Interprétation

Statistiques descriptives

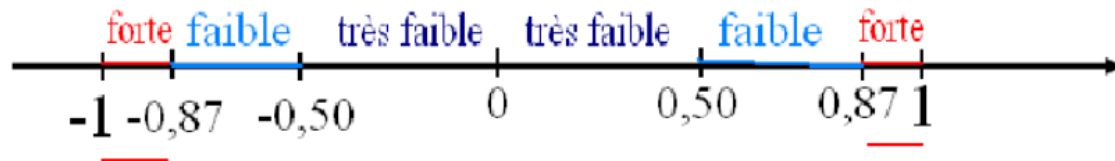
Si $r \in [-1; -0,87] \cup [0,87; 1] \Leftrightarrow 0,87 \leq |r| \leq 1$, on dit qu'il y a une **forte corrélation linéaire** entre les deux variables. On dit alors qu'un ajustement linéaire est justifié.

Si $r \in]-0,87; -0,50] \cup [0,50; 0,87[\Leftrightarrow 0,50 \leq |r| < 0,87$, on dit qu'il y a une **faible corrélation linéaire** entre les deux variables.

Si $r \in]-0,60; 0] \cup [0; 0,60[\Leftrightarrow 0 \leq |r| < 0,6$, on dit qu'il y a une **très faible corrélation linéaire** entre les deux variables. Dans ce cas, un ajustement linéaire est dépourvu de sens.

En particulier, si $r = 1$ ou $r = -1$, on dit qu'il y a une **parfaite corrélation linéaire** entre les deux variables.

Si $r = 0$, on dit qu'il n'y a **pas de corrélation linéaire** entre les deux variables.



Attention !

Il n'y a pas de corrélation linéaire ne signifie pas que les deux variables ne sont pas liées : les deux variables peuvent être liées d'un lien de nature autre que linéaire.

Exemple

x_i	4	5	6	7	8
y_i	5	2	1	2	5

$$y = (x - 6)^2 + 1$$

3. Equation des droites de régression

a) Droite de régression de Y en X

Elle se note $(D_{Y/X})$ et l'une de ses équations peut s'écrire sous la forme :

$$y = ax + b \text{ avec } a = \frac{\text{Cov}(X, Y)}{v(X)} \text{ et } b = \bar{Y} - a\bar{X}.$$

b) Droite de régression de X en Y

Elle se note $(D_{X/Y})$ et l'une de ses équations peut s'écrire sous la forme :

$$x = a'y + b' \text{ avec } a' = \frac{\text{Cov}(X, Y)}{v(Y)} \text{ et } b' = \bar{X} - a'\bar{Y}.$$

Application (voir exo 2 sur la fiche)

N.B : L'équation $y = ax + b$ donne une estimation de Y lorsque X est donné ou connu.

Exemple d'estimation de y pour $X = \dots$

Exemple :

Il a été relevé chez huit planteurs de coton, la superficie X (en ha) et la production Y (en tonnes) dans le tableau ci-dessous :

x_i	2	4	6	8	10	12	14	16
y_i	10	15	12	20	22	30	32	35

Statistiques descriptives

- a) Représenter dans un repère orthonormé (O, I, J) le nuage de points de la série statistique double (X, Y) . On prendra :
En abscisses 1 cm pour 1 ha et en ordonnées 1 cm pour 2 tonnes.
- b) Calculer le coefficient de corrélation linéaire. Un ajustement linéaire pourrait-il être justifié ?
- c) Déterminer l'équation de la droite de régression de y en x . Quel serait le niveau de la production pour un champ de coton d'une superficie de 20 hectares ?
- d) Déterminer l'équation de la droite de régression de x en y puis estimer la superficie d'un champ de coton qui permettrait d'avoir une production de 40 tonnes.

Elément de réponse :

$$r = 0,97 \text{ et } y = 1,88x + 5,07$$

TRAVAUX DIRIGES

Séries statistiques à une variable

Exercice 1

On donne la distribution de 25 élèves d'une classe selon le nombre de leurs frères et sœurs.

Nombre de frères et sœurs	0	1	2	3	4	5
Nombre d'élèves (n_i)	2	5	x	4	3	1

1. Quelle est la population étudiée ?
2. Quel est le caractère étudié et quelle est sa nature ?
3. Calculer x .
4. Reproduire le tableau en le complétant avec les effectifs cumulés croissants.
5. Quel est le nombre d'élèves ayant au moins 3 frères et sœurs ?
6. Quel est le nombre d'élèves ayant moins de 3 frères et sœurs ?
7. Combien de frères et sœurs a la majorité des élèves ?
8. Déterminer la médiane de la série étudiée.

Exercice 2

Dans une entreprise on a demandé à chacun des 35 ouvriers quelle était la durée du trajet de son domicile à son service. Ces durées arrondies aux multiples de 5 supérieurs, sont :

45 60 90 90 60 100 100 105 100 60

75 60 60 100 90 75 85 105 110 60

30 70 75 90 60 90 100 35 75 100

110 125 120 100 90.

1. Ranger ces durées par ordre croissant.
2. Faire un dépouillement en constituant des classes statistiques d'amplitude constante égale à 15 minutes.
3. Dresser le tableau statistique comportant la colonne:
 - des classes
 - des centres de classe
 - des effectifs
 - des fréquences relatives en valeurs décimales
 - des fréquences relatives cumulées croissantes et décroissantes en pourcentages.
4. Donner la classe modale.
5. Déterminer la moyenne, puis la variance.

Statistiques descriptives

6. Déterminer la classe médiane.
7. Quel est le pourcentage d'ouvriers faisant :
 - a. Moins d'une heure de trajet
 - b. Au moins une heure 30 minutes de trajet.

Exercice 3

Dans une localité donnée, 50 habitations ont été recensées suivant le nombre de pièces de chacune d'elles. Le résultat du recensement se présente comme suit :

5 ; 3 ; 4 ; 3 ; 2 ; 1 ; 3 ; 3 ; 3 ; 2 ; 4 ; 3 ; 2 ; 3 ; 1 ; 4 ;
5 ; 1 ; 3 ; 4 ; 3 ; 2 ; 3 ; 2 ; 3 ; 4 ; 2 ; 3 ; 4 ; 3 ; 2 ; 3 ; 1 ;
5 ; 2 ; 3 ; 2 ; 3 ; 5 ; 3 ; 2 ; 3 ; 4 ; 2 ; 3 ; 4 ; 3 ; 1 ; 2 ; 3.

1. a- Quelle est la population étudiée ?
b- Quel est le caractère étudié ; préciser la nature de ce caractère (qualitatif, quantitatif discret, quantitatif continu)
c- Déterminer les modalités de ce caractère.
2. Elaborer un tableau de dépouillement.
En déduire un tableau statistique comportant les fréquences cumulées croissantes et cumulées décroissantes en pourcentage
3. Quel est le pourcentage d'habitations ayant :
 - a-trois pièces ?
 - b- moins de quatre pièces ?
 - c- au moins trois pièces ?
4. Si une société immobilière souhaite construire des habitations à louer dans cette localité, quel type d'habitations réaliserait-elle ? Pourquoi ?
5. Calculer la moyenne, la variance de cette série statistique.
6. Déterminer le mode et la médiane

Exercice 4

Le tableau ci-dessous présente un extrait des résultats de notes obtenues par des élèves à une interrogation :

Note sur 10	Effectifs	Fréquence en %	Effectifs cumulés croissants	Effectifs cumulés décroissants
[2 ; 4[			3	
[4 ; 5[			9	
[5 ; 8[			14	
[8 ; 10[			20	
Total				

1. Reproduire et compléter le tableau ci-dessus.
2. Calculer la note médiane
3. a) Construire sur le même graphique les polygones des effectifs cumulés croissants et décroissants.
b) En déduire graphiquement la valeur de la médiane.

Echelle : Abscisses : 1cm → 1
Ordonnées : 1cm → 2 élèves

Séries statistiques à deux variables

Exercice 1

Chaque mois, le comptable d'une société de la place fait le bilan de ses dépenses X_i

Effectuée, ainsi que celui des bénéfices Y_i réalisés. Au terme de huit mois d'activités, Voici le tableau qu'il a réalisé en exprimant X_i et Y_i en millions de francs.

Numéro du mois	1	2	3	4	5	6	7	8
----------------	---	---	---	---	---	---	---	---

Statistiques descriptives

Dépenses X_i (en million de F)	4,5	6,6	5,4	7,8	10,2	9	8,5	10,8
Bénéfice Y_i (en million de F)	1,2	3	1,8	3,4	5	4,5	4,2	6

1. Représenter le nuage de points associé à cette série statistique double (X_i, Y_i) , dans un Repère orthonormé (O, I, J) . (Unité graphique : 1cm pour un million de francs sur l'axe des abscisses et 2cm pour 1million de francs sur l'axe des ordonnées).
2. Calculer le coefficient de corrélation linéaire entre les deux variables de la série.
Que peut-on en déduire ?
3. Déterminer l'équation de la droite de Mayer relative à cette série.
4. Estimer le bénéfice si les dépenses s'élèvent à 15 000 000 F.
5. Quel pourrait être le niveau des dépenses si l'entreprise veut avoir un bénéfice de 10 000 000 F?

Exercice 2:

Un professeur d'anglais a relevé les moyennes x_i en français et les moyennes y_i en mathématiques de 9 de ses élèves. Voici le tableau qu'il a obtenu.

X_i	4	6	8	9	10	11	12	14	16
Y_i	6	8	9	10	10	12	13	14	15

1. a) Représenter le nuage de points associé à cette série statistique double, dans le plan muni d'un repère orthonormé (O, I, J) . (Unités graphiques : 0,5 cm sur chaque axe)
b) Calculer les coordonnées du point moyen M et le placer dans le repère (O, I, J) .
- 2- a) En utilisant la méthode de MAYER, déterminer une équation de la droite d'ajustement linéaire (D) du nuage de points construit précédemment. Tracer (D).
b) Quelle moyenne obtiendrait au deuxième semestre un élève qui, au premier semestre a obtenu la moyenne 18 ?
- 3) Etait-ce judicieux de procéder à un ajustement linéaire ?
- 4) Donner l'équation de la droite de régression de y en x.

Exercice 4

A la fin du premier semestre de l'année académique 2005 - 2006, il a été relevé pour 5 élèves d'une classe de première année en transit transport, le nombre d'absence X et la moyenne Y en mathématique. Les résultats obtenus sont consignés dans le tableau suivant :

X_i	0	5	6	14	15
y_i	17	14	12	9	8

Déterminer l'équation de la droite d'ajustement (par la méthode de Mayer) de cette série et estimer la moyenne en maths d'un élève qui aurait 20 h d'absence.

Exercice 5

Statistiques descriptives

Le tableau suivant donne pour six mois de l'année 2004, les montants X_i des frais de publicité de l'entreprise de presse « BON JOURNAL » et les montants Y_i de ses chiffres d'affaires, exprimés en millions de francs.

mois	Jan.	Fév.	Mar.	Avr.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
X_i	5,8	4	6,4	4,6	5,2	7						
Y_i	127	102	137	116	117	142						

1. Représenter le nuage de points associé à cette série.

Echelle : Abscisses : 2,5 cm pour 1 million à partir de 4 millions
Ordonnées : 2,5 cm pour 10 millions à partir de 100 millions.

2. Est-il possible d'envisager un ajustement linéaire ? Pourquoi ?
3. Calculer le coefficient de corrélation linéaire pour confirmer votre réponse en 2)
4. Déterminer l'équation de la droite de régression de y en x.
5. En supposant que les frais de publicité augmente chaque mois de 500 000 F à partir du mois de juin, et que les chiffres d'affaires évoluent suivant l'équation de la droite de régression de y en x, compléter le tableau en déterminant :
 - a. les frais de publicité des six derniers mois,
 - b. les chiffres d'affaires prévisionnels de ces mois.
- 6) Etablir l'équation de la droite de régression de x en y.
À quelle date « BON JOURNAL » pourrait-il atteindre 223,51 millions de chiffre d'affaires ?

NB : On donnera les résultats des calculs au 1000ème près.

Chapitre 8:

LES INDICES STATISTIQUES

I. INTRODUCTION :

Supposons que nous suivions l'évolution, au fil des années, du prix d'une bouteille de sucrerie vendue dans une cave coopérative donnée. Le 20 novembre de chaque année, la connaissance du prix nous permet, dès la 2^{ème} année, d'établir une comparaison, la 1^{ère} année étant prise comme année de base ou année de référence.

Par exemple: $P_{20-11-90} = 275F$

$P_{20-11-91} = 285F$

$P_{20-11-92} = 270F$

Le rapport $\frac{P_{20-11-91}}{P_{20-11-90}} = \frac{285}{275} = 1,04$ définit un indice statistique, dans la pratique multiplié par 100 pour en faciliter la lecture et l'utilisation,

soit $I_{\%} = \frac{P_1}{P_0} \times 100 = 104$, indice du prix de la bouteille de sucrerie à

l'époque 1 (20-11-91) par rapport à l'époque 0 (20-11-90), époque de base ou de référence.

Dire que « la bouteille de sucrerie est à l'indice 104 », c'est constater une augmentation de prix de 4% par rapport à l'année de base.

De même, l'indice $I_{\%} = \frac{P_2}{P_0} \times 100 = \frac{270}{275} = 98$ indique, en 1992, une diminution de

prix de 2% par rapport au prix de 1990 (remarquons que l'indice d'une

Statistiques descriptives

période déterminée par rapport à la même période $I_i = \frac{P_i}{P_0} \times 100 = 100$ est

toujours égal à 100; l'indice de base est donc 100).

L'indice $I_{91} = \frac{270}{285} \times 100 = 95$ constate une baisse de prix de 5% par rapport à

l'année 1991 prise comme année de référence base 100.

De tels indices, se rapportant à un seul élément, sont des indices simples (ou indices particuliers ou indices élémentaires).

La grandeur mesurable n'est pas obligatoirement un prix : on peut comparer des quantités ou des volumes d'un même produit (manufacturé, exporté, importé, ...).

Si plus ambitieux, nous souhaitons étudier la variation, à chaque rentrée scolaire, du coût total des fournitures pour un écolier d'une classe déterminée, la construction d'un indice devient laborieuse.

En effet, il est nécessaire de préciser la nature des fournitures dont cette catégorie d'écoliers a besoin (cahiers, crayons, ... faut-il tenir compte des vêtements, des chaussures ?....), et leur quantité respective, tâche d'autant plus complexe et délicate que la pondération de chaque élément varie d'un enfant à l'autre.

Après avoir défini chaque élément constitutif et sa pondération, nous pourrions établir un indice à chaque rentrée scolaire. Un tel indice est appelé **indice synthétique**.

II. INDICES SIMPLES

La définition d'un indice n'offre aucune difficulté : c'est un nombre unique qui caractérise la variation relative d'une grandeur mesurable simple (prix, quantité, volume, etc.) entre deux périodes déterminées (période de référence et période actuelle, dite période courante) ou bien, au moment, au même moment, en deux lieux différents. Ce nombre est obtenu en multipliant par 100 le rapport des valeurs à deux époques (ou en deux places) différentes dont une est choisie comme base.

$$I_{\%} = \frac{P_n}{P_0} \times 100$$

Indice simple $I_{\%}$ de la grandeur à la date n , calculé sur la base 100 à la date 0 (date de base).

1. Choix de la période de base

On choisit généralement l'année ou le mois.

Mais il peut arriver que l'on prenne comme base non pas une année, mais une période de plusieurs années afin de réaliser un équilibre entre années très disparates (par exemple, la construction d'indices agricoles doit tenir compte du fait qu'une année aux récoltes abondantes peut être suivie d'une année catastrophique). Dans ce cas, toutefois exceptionnel, la base n'est plus constituée par l'ensemble des observations à une époque donnée, mais une moyenne des observations sur une période donnée.

Application 1

Les prix moyens de vente au détail d'un produit pendant les années 1987 - 1992 sont les suivants :

Année	1987	1988	1989	1990	1991	1992
Prix moyen au détail (en francs par Kg)	5260	5420	4750	5040	5330	5410

Statistiques descriptives

Années	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
Production	340239	263336	339914	299110	350181	357485	399650	325873	634315	421942

a) 1987 étant l'année de référence, nous obtenons les indices simples suivants :

Année	1987	1988	1989	1990	1991	1992
Indice simple (base 100 en 1987)						

b) 1989 étant l'année de référence, nous obtenons les indices simples suivants :

Année	1987	1988	1989	1990	1991	1992
Indice simple (base 100 en 1989)						

c) les années 1988 à 1990 constituent la période de référence, nous obtenons les indices simples suivants :

Année	1987	1988	1989	1990	1991	1992
Indice simple (base 1988 à 1990)						

Notons que la moyenne arithmétique des indices correspondant à la période de référence

1988 à 1990 est.....= 100, comme pour toute période de base.

2. Propriétés des indices simples

a. Réversibilité

Considérons les indices simples I_{j_0} , I_{j_1} .

Un indice est réversible si :

$$\frac{I_{j_0}}{100} = \frac{1}{\frac{I_{j_1}}{100}} \quad \text{ou} \quad \frac{I_{j_0}}{100} \times \frac{I_{j_1}}{100} = 1$$

b. Transférabilité

Considérons les indices simples I_{j_0} , I_{j_1} , I_{j_2} .

Un indice est transférable si :

$$\frac{I_{j_0}}{100} = \frac{I_{j_1}}{100} \times \frac{I_{j_2}}{100} \quad \text{ou} \quad I_{j_0} = I_{j_1} \times \frac{I_{j_2}}{100}$$

Application 2

L'évolution de la production de cognac (en hectolitres d'alcool pur) de 1963 à 1972 est résumée par le tableau suivant :

Calculer l'indice simple de quantité correspondant à chacune des années de la période 1963-1972, en prenant successivement comme période de référence :

- 1963
- 1968
- La période 1966 et 1967, soit deux années.

Statistiques descriptives

SOLUTION :

a)

Années	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
Indice simple (base 100 en)										

b)

Années	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
Indice simple (base 100 en)										

c)

Années	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
Indice simple (base 100 en)										

III. LES INDICES SYNTHETIQUES OU COMPOSEES

1. Activité

Une entreprise de restauration rapide a enregistré ces dernières années le coût de revient du menu type le plus demandé.

Les résultats donnés dans le tableau ci-dessous précisent pour les principaux constituants du menu :

- les quantités nécessaires à la réalisation du menu ;
- le coût unitaire de chacun des constituants du menu.

Menu	1999		2001	
	Q_0	P_0	Q_1	P_1
Hamburger	120 Kg	840 F/Kg	130 Kg	930 F/Kg
Frites	110 Kg	385 F/Kg	90 Kg	415 F/Kg
Boisson gazeuse	30L	170 F/L	50L	200 F/L

L'indice du coût de revient de ce menu type est **un indice composée**.

Son calcul prend en compte l'évolution des quantités Q_0 et Q_1 et des prix P_0 et P_1 des trois produits.

1. Calculer l'indice du coût de revient selon la méthode de **LASPEYRES** : on ne tient compte que des quantités Q_0 .

- Coût de revient d'un menu en 1999 :

$C_{99} =$

- Coût de revient d'un menu en 2001 :

$C_{01} =$

- En prenant comme base 100 en 1999, l'indice de **LASPEYRES** est :

$I_{0/99} =$

Statistiques descriptives

Le calcul d'un indice composé par la méthode de LASPEYRES se fait toujours avec les quantités fixées à l'époque de référence.

2. Calculer l'indice du coût de revient selon la méthode de **PAASCHE** : on ne tient compte que des quantités Q_1 .

Coût de revient d'un menu en 1999 et 2001 en utilisant les quantités consommées en 2001 :

$C_{99} = \dots\dots\dots$

$C_{01} = \dots\dots\dots$

L'indice de **PAASCHE** est :

$I_{0/99} = \dots\dots\dots$

Le calcul de l'indice de PAASCHE se fait avec les quantités consommées en 2001.

2. Remarques :

- Les indices calculés.....etsont légèrement différents : il faut donc préciser la méthode utilisée dans le calcul d'un indice composé.
- Le calcul par la méthode de LASPEYRES est le plus utilisé (on utilise toujours les mêmes coefficients : ceux de l'époque 0)
- Les indices composés sont également appelés : « indices synthétiques » ou « indices pondérés ».

3. Indices composés

- L'indice composé est un nombre sans unité, qui traduit l'évolution dans le temps de plusieurs grandeurs.
- Le calcul de l'indice composé peut se faire par deux méthodes différentes :

a. Méthode de LASPEYRES :

Les coefficients de pondération sont les quantités consommées à l'époque de référence :

$$I_{0/99} = \frac{\sum q_0 p_1}{\sum q_0 p_0} \times 100$$

p_0 : prix à l'époque 0
 q_0 : quantité
 p_1 : prix à l'époque 1
 q_1 : quantité

b) Méthode de PAASCHE :

Les coefficients de pondération sont les quantités consommées à l'époque 1 :

$$I_{0/99} = \frac{\sum q_1 p_1}{\sum q_1 p_0} \times 100$$

p_0 : prix à l'époque 0
 q_0 : quantité
 p_1 : prix à l'époque 1
 q_1 : quantité

Exemple

On donne le tableau suivant :

Produits	Epoque 0		Epoque 1	
	Prix unitaire	quantité	Prix unitaire	Quantité
A	15	2	18,5	5
B	6	10	4,5	8
C	3	25	4	32

Calculer à 0,1 près, l'indice composé des prix :

Statistiques descriptives

1. par la méthode de LASPEYRES.
2. par la méthode de PAASCHE.

Statistiques descriptives

TRAVAUX DIRIGES

EXERCICE 1 :

Un commerçant
7 020F en

vend un magnétoscope de
novembre 1987. En

novembre 1986, le prix du magnétoscope était de 6 500F.

En prenant pour base 100 en novembre 1986, calculer l'indice du prix du magnétoscope en novembre 1987 par rapport à novembre 1986.

EXERCICE 2 :

En avril 1985, le prix d'un kilogramme de café de marque et de qualité déterminées est 384F. Le prix de ce kilogramme de café est passé à 470,4F en juin 1987.

Quel est l'indice du prix de vente du café en juin 1987 en prenant pour base 100 le prix en avril 1985 ?

EXERCICE 3 :

Compléter le tableau suivant :

	P_0	P_t	$I_{t/0}$
Article A	164	188,6	...
Article B	237		94
Article C	...	98,49	117,25

EXERCICE 4 :

Le prix d'un vélomoteur est de 41 860F en avril 1987. L'indice du prix de ce vélomoteur, à cette date, par rapport au prix de février 1984 est 115.

Quel était le prix de ce vélomoteur en février 1984 ?

EXERCICE 5 :

Quantités pêchées en côte d'ivoire (en milliers de tonnes) :

1970	1980	1981	1982	1983
764,4	765,4	721,0	738,0	780,0

Calculer les indices simples pour les années 1980 à 1983, base 100 en 1970

(résultats au $\frac{1}{10}$ par excès) des quantités pêchées.

EXERCICE 6 :

Le prix d'un litre d'un produit déterminé était aux différentes époques considérées :

Epoque	Prix (FCFA)
0	36
1	39,06
2	40,68
3	43,74
4	50,85

1. calculer les indices $I_{1/0}$; $I_{2/0}$; $I_{3/0}$; $I_{4/0}$ du prix du produit par rapport à l'époque d'origine.

2. Calculer l'indice $I_{4/2}$ (indice du produit à l'époque 4 par rapport à l'époque 2).

3. Vérifier la relation : $I_{4/0} = \frac{1}{100}(I_{4/2} \times I_{2/0})$

EXERCICE 7 :

En 1984, un aspirateur est vendu 15 520F. Un modèle équivalent de la même marque valait 9 700F en 1980.

1. Calculer l'indice $I_{84/80}$ du prix de l'aspirateur en 1984, base 100 en 1980.

2. En 1982, l'indice $I_{82/80}$ correspondant était de 128 (base 100 en 1980). Quel était le prix de l'aspirateur en 1982 ?

3. Calculer $I_{84/82}$. Ensuite, vérifier la relation : $I_{84/80} = \frac{1}{100}(I_{84/82} \times I_{82/80})$.

4. Quel a été le pourcentage d'augmentation entre 1980 et 1984 (par rapport au prix de 1980) ?

Statistiques descriptives

EXERCICE 8:

Prix

Article	Epoque	
	0	t
A	52	59
B	61	63
C	48	47

Quantités

Article	Epoque	
	0	t
A	52	59
B	61	63
C	48	47

1. Calculer l'indice synthétique des prix $I_{t/0}$ pour les 3 articles par la méthode de Laspeyres.
2. Compter l'indice synthétique des prix $I_{t/0}$ pour les 3 articles par la méthode de Paasche.

Résultats au $\frac{1}{10}$ par excès.