Partie 2

Chapitre 7

Statistiques

Plan du chapitre

❊	**Activités préliminaires**	
❊	**Cours**	
	❖	Séries statistiques à deux caractères quantitatifs.
	❖	Ajustement affine d'une série statistique double.
	❖	Exemples d'ajustement non affine.
	❖	Exercices résolus.
❊	**Résumé du cours**	
❊	**Avec L'ordinateur**	
❊	**Exercices et Problèmes**	
❊	**Aperçu Historique**	

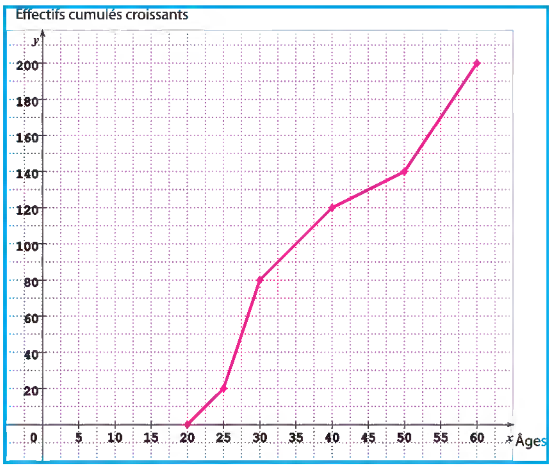
ACTIVITÉS PRÉLIMINAIRES

1 Une entreprise emploie 36% d'hommes et 96 femmes .La répartition des horaires du travail hebdomadaire est consignée dans le tableau ci-dessous :

Horaire	[20, 25 [	[25, 30 [	[30, 35 [	[35, 40 [	Total
Femmes	15	22			96
Hommes	2			41	
Total		29		64	

1) a) Compléter le tableau.
   b) Quel est le pourcentage des employés travaillant moins de 35 heures par semaine ?
   c) Quel est le pourcentage des femmes travaillant plus de 30 heures par semaine ?
   d) Quel est le pourcentage des hommes travaillant plus de 35 heures par semaine ?
2) Le tableau précèdent représente trois séries statistiques (horaire, femmes); (horaire, hommes); (horaire, employés).
   a) Représenter chacune de ces séries par un histogramme.
   b) Pour chacune de ces séries, déterminer le temps moyen de travail hebdomadaire d'un employé et l'écart- type de cette série.

2 L'âge des employés d'une usine varie entre 20 et 60 ans.
Le diagramme ci-contre représente le polygône des effectifs cumulés croissants de la répartition des âges.
1) Déterminer les classes et leurs effectifs.
2) Déterminer l'âge médian.
3) Calculer l'âge moyen et l'écart-type de cette série statistique.



- La moyenne d'une série statistique à caractère quantitatif continu est donnée par la formule : $\bar{x} = \frac{\sum_{i=1}^{p} n_i c_i}{n}$ où c_i est le centre de la classe $[x_i\ ,\ x_{i+1}[$; n_i l'effectif de cette classe et n est l'effectif total de la série.
- L'écart-type de cette série est $\sigma=\sqrt{V}$ avec $V = \frac{1}{n}\sum_{i=1}^{p} n_i c_i^2 - \left(\bar{x}\right)^2 = \overline{x^2} - \left(\bar{x}\right)^2$.
- La médiane est la valeur M_e du caractère qui correspond à un effectif cumulé croissant égal à $\frac{n}{2}$ si n est pair, $\left(\text{ou } \frac{n+1}{2} \text{ si n est impair}\right)$.

3 Un relevé statistique des tailles X (en cm) et des poids Y (en kg) d'un échantillon de 100 élèves a permis de construire le tableau suivant :

X \ Y	[40, 45[	[45, 50[	[50, 55[	[55, 60[
[150, 155[	18	10	2	0
[155, 160[	3	16	5	1
[160, 165[	0	5	13	5
[165, 170[	0	2	6	14

1) Donner la distribution marginale de X et la distribution marginale de Y.

2) Calculer $\overline{X}$; $\overline{Y}$; $V(X)$; $V(Y)$; $\sigma(X)$ et $\sigma(Y)$.

Retenons

Soient deux variables statistiques quantitatives X et Y définies sur une même population E d'effectif total N. Soient $x_1, x_2, ..., x_p$ les valeurs de X et $y_1, y_2, ..., y_q$ celles de Y (ou les centres des classes).

On note : n_{ij} l'effectif correspondant à $(X = x_i$ et $Y = y_j)$; $n_{i.} = n_{i1} + n_{i2} + ... + n_{iq}$ et $n_{.j} = n_{1j} + n_{2j} + ... + n_{pj}$. On a le tableau à double entrée suivant :

Y / X	y_1	...	y_j	...	y_q	Totaux
x_1	n_{11}	...	n_{1j}	...	n_{1q}	$n_{1.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	...	n_{ij}	...	n_{iq}	$n_{i.}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_p	n_{p1}	...	n_{pj}	...	n_{pq}	$n_{p.}$
Totaux	$n_{.1}$	...	$n_{.j}$	...	$n_{.q}$	N

La distribution marginale de X est donnée par le tableau suivant :

Valeurs de X	x_1	. . .	x_i	. . .	x_p
Effectif marginal	$n_{1.}$	. . .	$n_{i.}$	. . .	$n_{p.}$

* La moyenne arithmétique de X est $\overline{X} = \frac{1}{N}\sum_{i=1}^{p} n_{i.} \times x_i = \sum_{i=1}^{p} f_{i.} \times x_i$ où $f_{i.} = \frac{n_{i.}}{N}$ est la fréquence marginale associée à la modalité x_i de X.

* La variance marginale de X est $V(X) = \frac{1}{N}\sum_{i=1}^{p} n_{i.} \times x_i^2 - \left(\overline{X}\right)^2$.

* L'écart-type de X est $\sigma(X) = \sqrt{V(X)}$.

COURS

Séries statistiques à deux caractères quantitatifs

Activité 1

On teste la distance d de freinage (en mètres) d'un véhicule en fonction de sa vitesse v (en km/h) sur route humide. Les résultats obtenus sont consignés dans le tableau ci-dessous :

Vitesse (v)	40	50	60	70	80	90	100
Distance de freinage (d)	21,5	28	37,5	48	60	75	92,5

1) a) Construire le nuage de points sur une feuille de papier millimétré (choisir le point de cordonnées (30,10) comme intersection des axes orthogonaux du repère et prendre 1cm pour 5km/h sur l'axe des abscisses et 1cm pour 5m sur l'axe des ordonnées).
   b) Placer le point moyen G du nuage.
   c) Sur quelle courbe semblent être situés les points du nuage ?
2) On pose $y = \sqrt{d}$. Compléter le tableau suivant :

v	40	50	60	70	80	90	100
y							

   a) Construire le nouveau nuage de points.
   b) Placer le point moyen G' de ce nuage.
   c) Sur quelle ligne semblent être situés les points du nuage ?

Activité 2

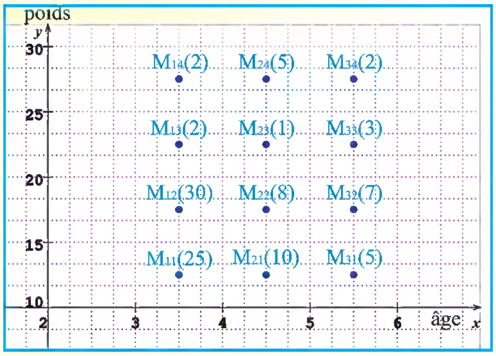
La répartition de 100 enfants selon leurs âges (en années) et leur poids (en kg) est consignée dans le tableau suivant :

Âges / Poids	**[3,4[**	**[4,5[**	**[5,6[**
[10,15[	25	10	5
[15,20[	30	8	7
[20,25[	2	1	3
[25,30[	2	5	2
Totaux			

1) Recopier et compléter le tableau ci-dessus.
2) Construire, dans un repère orthogonal, le nuage de points représentant cette série statistique double.
3) Déterminer les coordonnées du point moyen G et le placer dans le même repère.

Le plan est muni d'un repère orthogonal $(O,\vec{u},\vec{v})$.
Soit une série statistique à deux caractères quantitatifs X et Y.
On désigne par $x_1, x_2, ..., x_p$ les valeurs de X et par $y_1, y_2, ..., y_q$ celles de Y (ou les centres des classes). On note n_{ij} l'effectif correspondant à ($X = x_i$ et $Y = y_j$)

* On appelle **nuage de points** de la série considérée, l'ensemble des points $M_{ij}(x_i,y_j)$ tels que $1 \leq i \leq p$ et $1 \leq j \leq q$.
Chacun d'eux est affecté par l'effectif n_{ij} correspondant.
* On appelle **point moyen** du nuage de points, le point $G(\overline{X},\overline{Y})$ où $\overline{X}$ et $\overline{Y}$ désignent respectivement les moyennes arithmétiques des distributions de X et de Y.



Exemple de nuage de points

Activité 3

Dans le tableau suivant, le service financier d'une entreprise a consigné les frais de publicité x_i et le chiffre d'affaires y_i (exprimés en milliers de dinars) correspondant à 8 années consécutives:

Rang de l'année (i)	1	2	3	4	5	6	7	8
Frais de publicité (x_i)	1,1	1,4	1,6	2	2,2	2,3	2,2	2,4
Chiffre d'affaire (y_i)	59	68	67	71	71	69	67	72

1) a) Construire le nuage de points $M(x_i, y_i)$ sur une feuille de papier millimètré. On placera l'intersection des axes au point de coordonnées (1, 58) et en prendra sur l'axe des abscisses 1cm pour 0,2 mille dinars et sur l'axe des ordonnées 1cm pour 2 mille dinars.
b) Calculer le montant moyen $\overline{X}$ de frais de publicité sur cette période.
c) Calculer le chiffre d'affaires moyen $\overline{Y}$ sur cette période.
d) Placer le point moyen $G(\overline{X}, \overline{Y})$.

2) a) Donner une équation de la droite D =(M_1M_8) et tracer D.
   Le point G est -il sur D ?
   b) Donner une équation de la droite Δ = (GM_1) et tracer Δ.
   c) Laquelle des droites D et Δ représente le mieux le nuage des points M_i ?
3) Au cours d'une année, les frais de publicité étaient de 2,5 mille dinars.
   Peut-on prévoir le chiffre d'affaires correspondant ?

Activité 4

Le nombre de passagers (exprimé en milliers) dans un aéroport est relevé au cours des mois de janvier et de juillet pendant dix années consécutives. Les résultats sont donnés dans le tableau ci-dessous :

Rang de l'année i	1	2	3	4	5	6	7	8	9	10
Nombre de passagers en janvier (x_i)	350	360	410	420	440	480	510	530	580	630
Nombre de passagers en juillet (y_i)	620	625	630	690	790	760	940	920	1030	1040

1) Construire le nuage de points $M_i(x_i,y_i)$. On placera l'intersection des axes au point de coordonnées (300,600).
2) a) Calculer les cordonnées $G(\bar{x},\bar{y})$ du point moyen G de ce nuage.
   b) Calculer les cordonnées des points moyens G_1 et G_2 des nuages formés respectivement des cinq premiers points $\{M_i \;/\; i \leq 5\}$ et des cinq derniers points $\{M_i \;/\; 6 \leq i \leq 10\}$.
   c) Donner une équation de la droite D = (G_1G_2) puis tracer D.
3) En janvier d'une année il y a eu 700 mille passagers .Quel nombre de passagers peut-on prévoir pour le mois de juillet de la même année ?

Ajustement affine d'une série statistique double

Lorsque le nuage des points, représentant graphiquement une série statistique à deux caractères X et Y, a une forme allongée, on peut approcher la relation entre les deux variables X et Y par une relation affine définie par : Y= aX + b ou X= a'Y + b'.
On appelle **ajustement affine** toute méthode permettant la détermination d'une telle relation. Dans ce qui suit, nous présenterons deux méthodes : la méthode de Mayer et la méthode des moindres carrés.

Méthode de Mayer :

Activité 1

Une banque a enregistré les nombres de retraits opérés dans un guichet automatique pendant une journée. Le tableau suivant donne les montants (en DT) des retraits et leurs effectifs :

Montant en DT : x_i	40	35	30	25	20	15	10	5
Effectifs de retraits : y_i	19	20	17	11	13	6	7	2

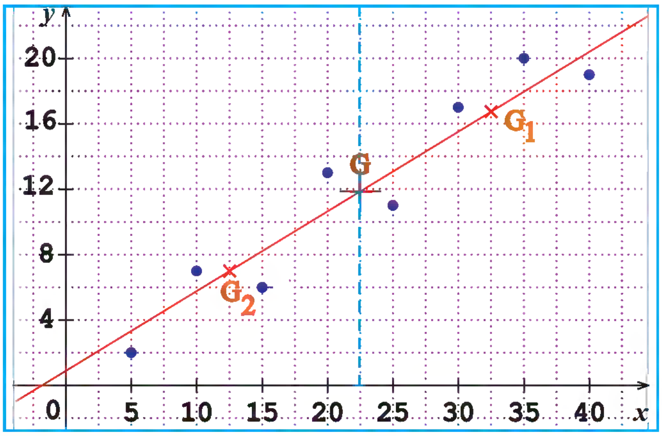
1) a) Construire, dans un repère orthogonal, le nuage des points représentant cette série statistique.
b) Quelle particularité peut -on remarquer au sujet de la forme du nuage ?
c) Déterminer, les cordonnées du point moyen G de ce nuage. Placer G.

2) On partage l'ensemble des points du nuage en deux parties. La première partie P_1 correspond aux retraits inférieurs ou égaux à 25 DT et la deuxième partie P_2 correspond aux autres retraits.
a) Déterminer les coordonnées des points moyens G_1 et G_2 respectifs des parties P_1 et P_2 . Placer G_1 et G_2 dans le même repère.
b) Donner une équation cartésienne de la droite (G_1G_2) et vérifier que cette droite passe par le point G.

3) Quel nombre de retraits de 50 DT peut -on prévoir en une journée ?

Commentaires :

- Le nuage de points de la série considérée a une forme allongée.
- La droite (G_1G_2) passe près des points du nuage. Elle est appelée droite de Mayer.
- Nous disons qu'on a réalisé un ajustement affine du caractère y en fonction de la variable x à l'aide de la méthode de Mayer.



La méthode de Mayer consiste à :
a) Partager le nuage de points en deux parties P_1 et P_2 satisfaisant à :
* P_1 et P_2 sont situées de part et d'autre par rapport à une droite parallèle à l'axe des ordonnées. ... / ...

* cardinal(P_1) = cardinal(P_2) si le nombre des points du nuage est pair.
* cardinal(P_1) = cardinal(P_2) +1 ou cardinal(P_1) = cardinal(P_2) -1 si le nombre des points du nuage est impair.

b) Déterminer les points moyens respectifs G_1 et G_2 des parties P_1 et P_2.

c) La droite (G_1G_2) est alors la droite d'ajustement affine du nuage de points représentant la série considérée. Cette droite est la droite de Mayer associée au nuage de points. Elle passe par le point moyen G du nuage.

Activité 2

L'évolution du pourcentage des frais de santé par rapport au budget des ménages d'un pays au cours des six dernières années est consignée dans le tableau suivant :

Rang de l'année : X	1	2	3	4	5	6
Pourcentage : Y	7,2	9,7	11,9	12,5	14	16

1) Construire le nuage de points représentant cette série double dans un repère orthogonal du plan.
2) Déterminer et placer le point moyen G de ce nuage.
3) Par la méthode de Mayer, Déterminer une équation cartésienne de la droite d'ajustement affine de Y en X.
4) Quel pourcentage de frais de santé par rapport au budget des ménages peut-on prévoir pour l'année de rang 10 ?

Activité 3

Dans le tableau suivant on a consigné les moyennes annuelles en mathématiques obtenues par dix élèves de classe terminale et leurs notes de mathématiques au Baccalauréat :

Moyennes annuelles : X	7	10	6	15	16	18	15	10	7	10
Notes de Math au Baccalauréat : Y	7	8	5	11	17	18	13	11	9	10

1) Construire le nuage de points de la série statistique double (X, Y) ainsi définie.
2) Préciser le point moyen G du nuage.
3) Réaliser un ajustement affine par la méthode de Mayer.
4) Peut-on prévoir une estimation de la note de mathématiques au baccalauréat d'un élève ayant obtenu une moyenne annuelle en mathématiques égale à 13 ?

Méthode des Moindres carrés :

Soit une série statistique à deux caractères quantitatifs X et Y.
On peut reconnaître la relation affine éventuelle entre les deux variables X et Y à l'aide d'un moyen non graphique et en faisant intervenir deux nouveaux paramètres statistiques à savoir : la covariance et le cœfficient de corrélation linéaire.

1) Covariance

Définition

Soit une série statistique à deux caractères quantitatifs X et Y.
On désigne par $x_1, x_2, ..., x_n$ les valeurs de X et par y_1, y_2,..., y_n celles de Y (ou les centres des classes).On note n l'effectif total de la population observée.
On appelle covariance de la série double considérée le réel,noté Cov(X,Y), et défini par : $\text{Cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{X}.\overline{Y}$ où $\overline{X}$ et $\overline{Y}$ sont les moyennes arithmétiques respectives des distributions $(x_i)_{1 \le i \le n}$ et $(y_i)_{1 \le i \le n}$ de X et Y.

Remarque :

On a : $\text{Cov}(X, Y) = \overline{XY} - \overline{X}.\overline{Y}$; où $\overline{XY}$ désigne la moyenne arithmétique de la distribution $(x_i y_i)_{1 \le i \le n}$ du caractère statistique produit .

Activité 1

Le tableau suivant donne les nombres d'accidents de la circulation enregistrés au cours des cinq dernières années dans une ville :

Rang de l'année : X	1	2	3	4	5
Nombre d'accidents : Y	82	90	117	96	102

1) a) Calculer les moyennes arithmétiques $\overline{X}$ et $\overline{Y}$.
   b) Calculer les variances V(X) et V(Y).
2) a) Calculer Cov(X,Y).
b) Comparer $\text{Cov}^2(X,Y)$ au produit V(X).V(Y).
3) Construire le nuage de points de la série double. Existe-t-il une relation affine entre les variables X et Y ?

Commentaires :

Soit une série statistique a deux caractères quantitatifs X et Y.
Lorsque $\text{Cov}^2(X,Y)$ est assez voisin de V(X).V(Y), on peut dire que la relation entre X et Y est approximativement affine.
On peut alors trouver deux réels a et b tels que Y est proche de aX + b.
Cette approximation pourrait être très utile pour décrire un phénomène aléatoire et prévoir son évolution.

Activité 2

Une étude a été réalisée dans une entreprise pour connaître l'effet des longues durées de travail sur la qualité de la production. Le tableau suivant donne les résultats de cette étude :

Durée en heure du travail quotidien : X	8	9	10	12	14	15
Nombre d'erreurs commises : Y	6	7	10	14	20	25

1) a) Calculer, les écarts-types, $\sigma(X)$ et $\sigma(Y)$.
   b) Calculer Cov(X,Y).
   c) Comparer Cov(X,Y) au produit $\sigma(X).\sigma(Y)$.
   d) Que peut -on en déduire ?
2) Utiliser la méthode de Mayer pour déterminer un ajustement affine de Y en X.
3) Peut-on prévoir le nombre d'erreurs commises par un ouvrier qui a travaillé 13 heures de suite en une journée ?

2) Cœfficient de corrélation linéaire

Définition

Soit une série statistique à deux caractères quantitatifs X et Y non constants observés dans une population donnée d'effectif total n.
On appelle **cœfficient de corrélation linéaire** de la série double (X, Y), le réel r défini par : $r = \dfrac{\mathrm{Cov}(X, Y)}{\sigma(X).\sigma(Y)}$ où $\sigma(X)$ et $\sigma(Y)$ sont les écarts-types respectifs des variables statistiques X et Y.

Activité 1

Le tableau suivant donne le nombre de résistances électriques vendues, dans un magasin de composantes électriques, suivant leurs puissances (en W) en un mois :

Puissances (en W) : X	100	150	200	300	500
Nombre de résistances : Y	50	70	100	150	250

1) a) Calculer Cov(X,Y) , $\sigma(X)$ et $\sigma(Y)$.
   b) Peut-on déduire qu'il existe une relation du type affine entre X et Y ?
2) a) Calculer le cœfficient de corrélation linéaire r du couple (X, Y).
   b) Comparer r à 1.
3) a) Tracer les droites D_1 et D_2 d'équations respectives : y = ax + b et x =a'y +b'

$$D_1 : y = ax + b \text{ avec } a = \frac{\mathrm{cov}(X, Y)}{V(X)} \text{ et } b = \overline{Y} - a\overline{X}$$

$$D_2 : x = a'y + b' \text{ avec } a' = \frac{\mathrm{cov}(X, Y)}{V(Y)} \text{ et } b' = \overline{X} - a'\overline{Y}$$

b) Construire le nuage des points de la série double considérée.

c) Comment sont situées les droites D_1 et D_2 par rapport au nuage de points ?

d) Montrer que le point moyen $G(\overline{X},\overline{Y})$ appartient à D_1 et D_2 .

Commentaires :

Soit une série statistique à deux caractères quantitatifs X et Y.

Lorsque le cœfficient de corrélation linéaire r du couple (X, Y) est proche, en valeur absolue, de 1 ($0.75 \leq |r| \leq 1$), le nuage de points de la série considérée a une forme allongée et il est possible d'approcher la liaison entre X et Y par deux relations affines représentées graphiquement par deux droites D_1 et D_2 passant par le point moyen $G(\overline{X},\overline{Y})$ du nuage.

* La droite D_1: appelée droite de régression de Y en X et ayant pour équation :

$y = ax + b$ avec $a = \dfrac{\text{cov}(X, Y)}{V(X)}$ et $b = \overline{Y} - a\overline{X}$.

* La droite D_2: appelée droite de régression de X en Y et ayant pour équation :

$x = a'y + b'$ avec $a' = \dfrac{\text{cov}(X, Y)}{V(Y)}$ et $b' = \overline{X} - a'\overline{Y}$.

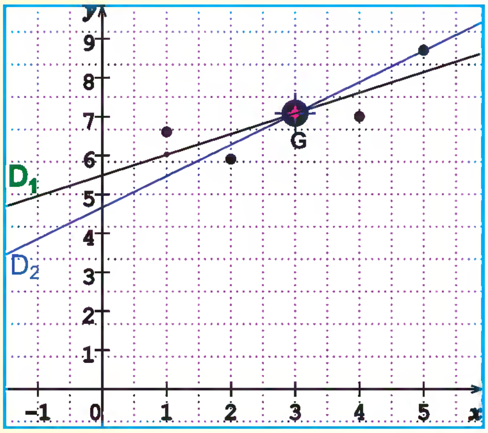
Exemple :

$r \simeq 0,81$;

D_1: $y = 0,53\,x + 5,49$

D_2: $x = 1,24\,y - 5,79$

G étant le point moyen du nuage de points ci-contre correspondant à la série double (X,Y).



Activité 2

Le tableau suivant donne l'âge X et la tension artérielle Y de 10 personnes :

X	58	40	74	34	65	49	53	51	36	40
Y	16,7	13,1	17,2	11,6	15,5	15,1	14,2	14,4	13,0	14,2

1) Construire le nuage de points de cette série statistique.

2) Déterminer la moyenne et la variance de chacune des variables X et Y.

3) Déterminer le cœfficient de corrélation linéaire r des variables X et Y.

4) Déterminer une équation de la droite de régression de Y en X.

5) Estimer la tension artérielle d'une personne âgée de 45 ans.

Activité 3

Dans une grande surface, le prix de vente promotionnel d'un produit (en DT) est affiché en fonction de son poids (en g) dans le tableau suivant :

Prix du produit: X (en D.T)	0,650	0,900	1,100	1,300	1,500	2,600
Poids du produit : Y (en g)	100	150	200	250	300	500

1) a) Calculer le cœfficient de corrélation linéaire r de la série double (X, Y).
   b) Un ajustement affine est-il justifié ?
2) Donner les équations des droites de régression relativement à un repère orthogonal du plan.
3) Quel prix peut-on prévoir pour un produit de poids 1 kg ?

Exemples d'ajustement non affine

Activité 1

Les frais de publicité (en DT) engagés par une entreprise durant les dix dernières années sont consignés dans le tableau suivant :

Rang de l'année(X)	1	2	3	4	5	6	7	8	9	10
Frais de publicité (Y)	1015	1310	1690	1977	2803	3606	4640	5960	7400	9000

1) Compléter le tableau suivant :

Rang de l'année(X)	1	2	3	4	5	6	7	8	9	10
Z = ln Y										

2) a) Représenter, dans un repère orthogonal, par un nuage de points, la série statistique double définie par (X, Z).
   b) Déterminer les coordonnées du point moyen G de ce nuage. Placer G dans le même repère.
3) a) Tracer une droite D passant par G et assez près des points du nuage.
   b) Déterminer une équation cartésienne de D.
   c) En déduire une relation entre X et Z.
4) Trouver, alors, une relation entre X et Y.

Activité 2

Une entreprise fabrique un produit A.

1) Le prix de revient de fabrication est composé de a (en DT) de frais fixes et b (enDT) par unité produite. Pour n unités produites ce prix de revient est alors a + n.b.
Montrer que le prix de revient d'une unité produite est $y = \frac{a}{n} + b$.

2) On se propose de trouver les cœfficients a et b.
a) Dans le tableau suivant, on a enregistré le prix unitaire pour chaque nombre d'unités produites :

Nombre d'unités : n	1	2	3	4	5	6	7	8	9	10
Prix de revient unitaire : y	150	81	54	44	38	32	29	26	23	21

Représenter le nuage de points de la série double (n, y). Un ajustement affine vous parait-il justifié ?

b) On pose $x = \frac{1}{n}$.
- Calculer les valeurs de x à 10^{-2} près, par défaut.
- Représenter le nuage de points de la série double (x, y). Un ajustement affine vous parait-il justifié ?
- Déterminer une équation de la droite de régression de y en x.
- En déduire une relation entre y et n.

c) Quelle estimation peut-on donner du prix de revient d'une unité de produit A si on en fabrique 20 ?

Activité 3

Le tableau suivant donne la tension u (en volts) aux bornes d'un condensateur à l'instant t (en secondes) :

t (en s.)	30	60	90	120	150	180	210	240
u (en V)	2,35	1,40	0,80	0,50	0,30	0,20	0,15	0,10

On pose y = ln u

1) Dresser le tableau de valeurs de la série double (t, y).

2) a) Calculer le cœfficients de corrélation entre t et y .
b) Un ajustement affine est-il justifié ?

3) a) Donner une équation de la droite de régression de y en t.
b) En déduire une relation entre t et u.

4) Déterminer l'instant T en lequel u vaut 1 volt.

Exercices résolus

Exercice 1

Le tableau suivant donne la distance de freinage d (en mètres) d'une voiture en fonction de sa vitesse v (en km/h) :

v (km/h)	30	40	50	60	70	80
d (en mètres)	42	60	80	90	95	110

1) Calculer $\bar{v}$, $\bar{d}$, V(v), V(d) et cov(v,d) .
2) a) Calculer le cœfficient de corrélation linéaire entre v et d.
   b) Y- a- t- il forte corrélation affine entre v et d ? Justifier.
3) Soit Δ la droite de régression de d en v. On considère qu'une équation cartésienne de Δ est : d = 1,3.v + 8. Calculer la distance de freinage lorsque la voiture roule à 100 km/h.
4) La vitesse de la voiture est de 140km/h, lorsque le conducteur, roulant suivant une ligne droite aperçoit un obstacle situé à une distance de 200 mètres.
   Pourrait-il, alors, éviter cet obstacle sachant qu'il met une seconde pour appuyer sur les freins ?

(Bac Tunisien : juin 2005)

Solution :

1) $\bar{v} = 55$; $\bar{d} = 79,5$; $V(v) = 291,66$; $V(d) = 511,25$ et $cov(v,d) = \overline{v.d} - \bar{v}.\bar{d} = 379,166$.

2) a) $r = r(v,d) = \dfrac{cov(v,d)}{\sigma(v).\sigma(d)} \simeq 0,981$.

   b) Ce coefficient r est proche de 1.Donc il y a une forte corrélation affine entre v et d et un ajustement affine est justifié.

3) d = 1,3.v + 8. Si v = 100 km/h alors d = 138 m.

4) Si v = 140 km/h alors d = 190m.
   En une seconde, la voiture parcourt la distance $d_0 = \dfrac{140.000}{3600} = 38,8\,m$.

   La distance totale parcourue serait : $d + d_0 = 228,8\,m > 200\,m$ .Par conséquent, le conducteur ne pourra pas éviter l'obstacle.

Exercice 2

Le rendement R d'une variété de blé (en quintaux par hectare) et la quantité E d'engrais azotés (en kg/ha) utilisée pendant la culture sont indiqués dans le tableau suivant:

E (kg / ha)	50	60	70	80	90
R (q / ha)	35,7	41,4	45,7	47,2	50,8

1) a) Calculer le coefficient de corrélation linéaire du couple (E, R).
   b) Que peut-on en déduire ?
2) Déterminer une équation de la droite de régression de R en E relativement à un repère orthogonal du plan.
3) Quel rendement peut-on prévoir pour une culture utilisant une quantité d'engrais azotés E = 100kg/ha ?

Solutions :

1) a) $\overline{E} = 70$; $\overline{R} = 44{,}16$; $r(E,R) = \dfrac{\text{cov}(E,R)}{\sigma(E).\sigma(R)} \simeq 0{,}98$.

b) Le coefficient de corrélation est très voisin de 1, on en déduit que l'une des variables peut être approchée par une fonction affine de l'autre variable.

2) $R = a.E + b$ avec $a = \dfrac{\text{cov}(E,R)}{V(E)} = 0{,}36$ et $b = \overline{R} - a.\overline{E} = 18{,}92$.

L'équation demandée est alors : R = 0,36.E +18,92.

3) Pour E=100kg / ha, la formule précédente donne:

$R = 0{,}36 \times 100 + 18{,}92 = 54{,}92$ (q / ha) .

Exercice 3

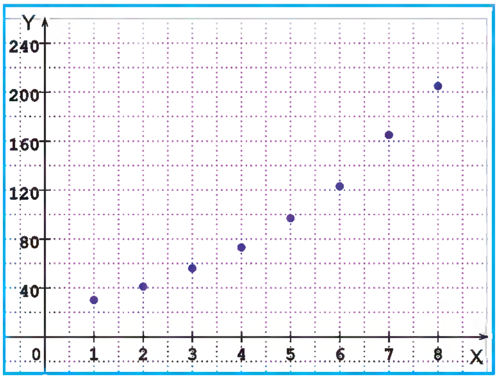
Le tableau suivant indique l'évolution de la consommation d'énergie électrique dans un pays au cours de 8 années successives :

Rang de l'année : X	1	2	3	4	5	6	7	8
Consommation (en Twh) : Y	30	41	56	73	97	123	165	205

1) a)Représenter le nuage de points de la série double (X, Y) dans un repère orthogonal du plan.
   b) Calculer le coefficient r_1 de corrélation linéaire du couple (X, Y).
   c) Déterminer par la méthode des moindres carrés une équation de la droite D de régression de Y en X.
2) On suppose que la relation entre X et Y est du type exponentiel : $Y = k.e^{aX}$ et on pose V =ln Y.
   a) Représenter le nuage de points de la série double (X, V) dans un repère orthogonal du plan et calculer le coefficient r_2 de corrélation linéaire du couple (X, V).
   b) En déduire qu'il y'a une forte corrélation linéaire entre V et X puis construire la droite D' de régression de V en X.
   d) En écrivant V =a.X + b où b = ln k , trouver alors les réels a et b.

Solution :

1) a) Représentation du nuage de points de la série double (X, Y) ; voir figure ci-contre :



b) On a :

$\overline{X} = 4,5$; $\overline{Y} = 98,75$; $\sigma(X) = 2,29$;$\sigma(Y) = 57,9$

$\mathrm{cov}(X,Y) = 129,37$; $r_1 = r(X,Y) = \dfrac{\mathrm{cov}(X,Y)}{\sigma(X).\sigma(Y)} = 0,97$

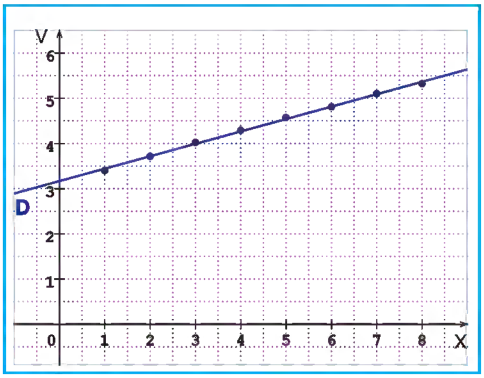
c) Une équation de la droite D de régression de X en Y est : y = 9,62.x + 29,71.

2) V = ln Y; On a aors, le tableau suivant :

X	1	2	3	4	5	6	7	8
V	3,401	3,713	4,025	4,290	4,574	4,812	5,105	5,323

On a : $\overline{X} = 4,5$; $\overline{V} = 4,40$; $\sigma(X) = 2,29$;$\sigma(V) = 0,62$;

$\mathrm{cov}(X,V) \approx 1,441$; $r_2 = r(X,V) = \dfrac{\mathrm{cov}(X,V)}{\sigma(X).\sigma(V)} \approx 0,99$;



On a r(X, V) est voisin de 1 donc il y a forte corrélation linéaire entre V et X et un ajustement affine est donc, justifié.

c) Une équation de la droite Δ de régression de X en V est : v = 0,27.x + 3,17 . Ainsi,on a:

a = 0,27 et b = 3,17 ce qui donne :

Y = $e^{3,17}.e^{0,27.X} = e^{0,27.X+3,17}$.

3) Soit f(x) = $e^{0,27.X+3,17}$. Etudier et réprésenter graphiquement f dans le même repére que le nuage de points de la premiére question.

RÉSUMÉ DU COURS

*** Nuage de points d'une série statistique double :**

Le plan est muni d'un repère orthogonal (O,u,v) .

Soit une série statistique à deux caractères quantitatifs X et Y. On désigne par $x_1, x_2, \dots, x_p$ les valeurs de X et par $y_1, y_2, \dots, y_q$ celles de Y (ou les centres des classes). On note n_{ij} l'effectif correspondant à $(X = x_i$ et $Y = y_j)$.

- On appelle nuage de points de la série considérée, l'ensemble des points $M_{ij}(x_i, y_j)$ tels que $1 \leq i \leq p$ et $1 \leq j \leq q$ .Chacun d'eux est affecté par l'effectif n_{ij} correspondant.

- Le point moyen du nuage de points est $G(\overline{X}, \overline{Y})$ où $\overline{X}$ et $\overline{Y}$ désignent respectivement les moyennes arithmétiques de X et Y.

*** Ajustement affine d'une série statistique double :**

Lorsque le nuage des points, représentant graphiquement une série statistique à deux caractères X et Y, a une forme allongée, on peut approcher la relation entre les deux variables X et Y par une relation affine définie par : $Y = a.X + b$ ou $X = a'.Y + b'$. Toute méthode permettant la détermination d'une telle relation, s'appelle un ajustement affine du couple (X, Y).

Méthode de Mayer :

La méthode de Mayer consiste à :

a) Partager le nuage de points en deux parties P_1 et P_2 satisfaisant à :

* P_1 et P_2 sont situées de part et d'autre par rapport à une droite parallèle à l'axe des ordonnées.

* $card(P_1) = card(P_2)$ si le nombre des points du nuage est pair.

* $card(P_1) = card(P_2) + 1$ ou $card(P_1) = card(P_2) - 1$ si le nombre des points du nuage est impair.

b) Déterminer les points moyens respectifs G_1 et G_2 des parties P_1 et P_2 .

c) La droite (G_1G_2) est alors la droite d'ajustement affine du nuage de points représentant la série considérée. Cette droite s'appelle la droite de Mayer.

Méthode des moindres carrés :

Soit une série statistique à deux caractères quantitatifs X et Y.

On désigne par $x_1, x_2, \dots, x_n$ les valeurs de X et par $y_1, y_2, \dots, y_n$ celles du caractère Y (ou les centres des classes).On note n l'effectif total de la population observée. Soit r le cœfficient de corrélation linéaire du couple (X, Y) :

$$r = \frac{Cov(X, Y)}{\sigma(X).\sigma(Y)}$$

où $\sigma(X)$ et $\sigma(Y)$ sont les écarts-types respectifs des variables statistiques X et Y;

et $Cov(X, Y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{X}.\overline{Y}$ est la covariance du couple (X, Y).

* Lorsque le cœfficient de corrélation linéaire du couple (X, Y) est proche, en valeur absolue, de 1 $(0.75 \leq |r| \leq 1)$, il est possible d'approcher la liaison entre X et Y par deux relations affines représentées graphiquement par deux droites passant par le point moyen G du nuage. La droite D_1 : appelée droite de régression de Y en X et ayant pour équation :

$$y = ax + b \text{ avec } a = \frac{\text{cov}(X, Y)}{V(X)} \text{ et } b = \overline{Y} - a\overline{X}$$

La droite D_2 : appelée droite de régression de X en Y et ayant pour équation :

$$x = a'y + b' \text{ avec } a' = \frac{\text{cov}(X, Y)}{V(Y)} \text{ et } b' = \overline{X} - a'\overline{Y}$$

AVEC L'ORDINATEUR

Activité 1

On se propose de résoudre, à l'aide du logiciel **Sine qua non** :(Ressource sur Internet: http://perso.orange.fr/patrice.rabiller/SineQuaNon/menusqn.htm), le problème ci-dessous proposé par une entreprise qui fabrique et vend des lots de circuits électriques.
Ces circuits sont garantis un an. L'entreprise envisage d'augmenter la durée de la garantie et désire connaître l'incidence de cette mesure sur les bénéfices.
Le tableau suivant indique le pourcentage y de circuits d'un lot qui ont une panne au cours de x semestres d'utilisation :

x (semestres)	1	2	3	4	5	6	7	8	9	10
y (pourcentage)	2	3	4	7	9	11	16	20	23	31

1) Pour construire le nuage des points $M_i\,(x_i,\, y_i)$ à l'aide du logiciel :
* Définir, tout d'abord, le repère **(menu Définir, repère),** tout en choisissant l'origine, la graduation, la forme de la grille, etc. Par exemple : on prendra 1cm pour 1 semestre et 0,5 cm pour 1%.
* **Définir** successivement les points $M_i(x_i,y_i)$ (choisir **série statistique double**,dans le menu Définir, et remplir le tableau de données,…) ,etc. puis valider par OK.
Question : Peut-on envisager un ajustement affine résumant le nuage obtenu ?

2) Pour visualiser la droite de régression de y en x :
* Ouvrir le logiciel et taper sur l'icône correspondant à la définition d'une série statistique double.
* Faire entrer les données du tableau précédent, cocher la nature de régression (linéaire : y en fonction de x). Vous obtenez alors des renseignements sur les paramètres statistiques de cette série double, etc.
* Valider par OK pour obtenir la représentation graphique de la droite de régression de y en x.

3) Le logiciel consigne que : cov(x, y) = 25,5 ; le cœfficient de corrélation linéaire r = 0,97 et l'équation de la droite de régression de y en x est y = 3,09 x - 4,4 .
Justifier à votre tour, par le calcul, ces résultats consignés.

4) Pour un lot, la part de bénéfice z, exprimée en pourcentage du prix de vente du lot, est liée à la variable y précédente par la relation : z = 40 - 1,2 y .
En utilisant la relation obtenue dans la question 3) entre x et y, déterminer le nombre maximum de semestres de garantie que l'entreprise peut accorder à ses clients tout en conservant un bénéfice z , exprimé en pourcentage du prix de vente d'un lot, supérieur ou égal à 25.

Activité 2

Pour résoudre une activité se rapportant à une série statistique double, en utilisant une calculatrice Sharp, il faut choisir tout d'abord le mode statistique double :

Mode Statistique Double :

Sharp Mode (0 -2)	Sharp Mode (0 -1)	Sharp Mode (0 - 3)
2ndf , Mode , 2 ==> Stat xy	2ndf , Mode , 1 , 1 ==> Stat 1	2ndf , Mode , 3 , 1 ==> Stat 1

Exemple : Dans une population de 30 logements, on a observé les deux variables statistiques suivantes : X : nombre d'enfants dans chaque logement. Y : nombre de pièces du logement.

*** Les résultats individuels sont consignés dans le tableau suivant :**

x_i	0	3	0	3	1	3	1	2	3	2	1	1	1	0	3	0	1	2	3	1	2 1	3	1	2	2	3	2	1	2
y_i	1	3	2	2	1	3	2	2	2	1	3	2	3	1	3	2	2	3	3	1	3 2	2	3	3	1	2	3	2	2

*** Ces résultats sont résumés dans le tableau à double entrée suivant:**

Y \ X	1	2	3	Totaux
0	2	2	0	**4**
1	2	5	3	**10**
2	2	2	4	**8**
3	0	4	4	**8**
Totaux	**6**	**13**	**11**	**30**

Saisie du tableau à double entrée :	La série marginale de X :	La série marginale de Y :
0 STO 1 STO 2 M+	La moyenne arithmétique de X :	La moyenne arithmétique de Y :
0 STO 2 STO 2 M+	RCL 4 → $\overline{X}$ = 1.66	RCL 7 → $\overline{Y}$ = 2.16
1 STO 1 STO 2 M+	La variance V(X) :	La variance V(Y) :
1 STO 2 STO 5 M+	RCL 6 x^2 = → V(X) = 1,02	RCL 9 x^2 = → V(X) = 0.53
1 STO 3 STO 3 M+	L'écart-type de σ(X) :	L'écart-type de σ(Y) :
2 STO 1 STO 2 M+	RCL 6 → σ(X) = 1,01	RCL 9 → σ(Y) = 0.73
2 STO 2 STO 2 M+		
2 STO 3 STO 4 M+		
3 STO 2 STO 4 M+		
3 STO 3 STO 4 M+		

Le coefficient de corrélation linéaire r : RCL ÷ → 0.38 (corrélation faible).

La covariance Cov(X, Y) : RCL ÷ x RCL 6 x RCL 9 = → Cov(X, Y) = 0.28

Droite de régression de Y en X : Y = α X + β ,

α : RCL) → α = 0.28 ; β : RCL (→ β = 1.69

Droite de régression de X en Y : X = α' Y + β',

α' : RCL ÷ x^2 ÷ RCL) = → α' = 0.04 ; β' : $\overline{X}$ – α' $\overline{Y}$ → 1.58.

EXERCICES ET PROBLÈMES

1 On a relevé, dans le tableau ci-dessous, pour 100 véhicules la distance parcourue (en milliers de km) en 3 ans :

Distance parcourue	De75 à 85	De85 à 95	De95 à 105	De105 à 115	De115 à 125
Effectif	10	18	40	20	12

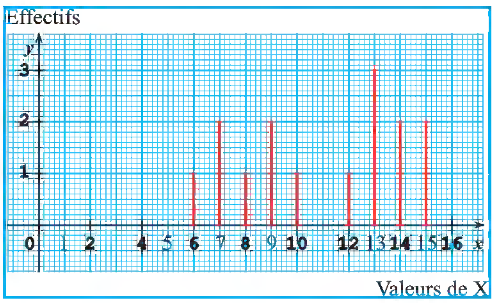
1) Construire l'histogramme correspondant à cette série statistique.
2) Calculer la distance moyenne parcourue par un véhicule en une année.

2 Dans deux entreprises A et B, les salaires horaires ont été classés de la manière suivante :

Salaire (en dinars)	De 1 à 1,5	De1,5 à 2	De2 à 2,5	De2,5 à 3	De3 à 3,5
Effectif de A	10	18	40	20	12
Effectif de B	10	18	40	20	12

1) Établir le tableau associé aux fréquences des deux séries statistiques.
2) a) Construire dans un même repère orthogonal les histogrammes des fréquences.
b) Comparer, globalement, les deux entreprises.
3) Calculer, pour chaque série, le salaire horaire moyen ; la médiane et lécart-type.

3 Le diagramme ci-dessous, représente un caractère quantitatif discret X.



1) Déterminer la moyenne, la médiane et le mode de la série statistique associée à X.
2) Soit le caractère quantitatif Y dont les valeurs sont : 2 ; 3 ; 4 ; 5 ; 9 ; 10 ; 10 ; 11 ; 12 ; 12 ; 12 ; 12 ; 15 ; 16 ; 17.
a) Représenter graphiquement Y.
b) Montrer que X et Y ont même moyenne, même médiane et même mode.
3) a) Calculer pour chaque caractère : l'étendue, l'écart moyen et l'écart-type.
b) Interpréter les résultats obtenus.

4 Le tableau ci-dessous donne la taille et le poids de cinq enfants :

Taille X (en cm)	95	103	109	122	133
Poids Y(en kg)	15	16	19	24	30

1) a) Construire le nuage de points de cette série double dans un repère orthogonal du plan.
b) Déterminer et placer le point moyen G de ce nuage.
2) Utiliser la méthode de Mayer pour déterminer un ajustement affine de la série considérée.
3) Utiliser les résultats de la question précédente pour déterminer une valeur approchée du poids d'une enfant dont la taille est 115 cm.

5 Les dépenses annuelles d'un pays pour la promotion des nouvelles technologies de l'information et de la communication, au cours des six dernières années (en milliers de D.US) sont consignées dans le tableau suivant :

Rang de l'année : X	1	2	3	4	5	6
Dépenses (en M.D.US) : Y	445	474	504	530	568	588

1) Représenter graphiquement la série par un nuage de points et placer le point moyen G dans un repère orthogonal du plan.
2) On considère les points du nuage correspondant aux années 1, 2 et 3 et on désigne par G_1 le point moyen de ces trois points. Soit G_2 le point moyen des trois points correspondant aux années 4, 5 et 6.
   a) Placer les points G_1 et G_2 et tracer (G_1G_2).
   b) Déterminer une équation de la droite (G_1G_2).
   c) Que représente G pour le segment $[G_1G_2]$?
3) Déterminer une valeur approchée de la dépense au cours de l'année de rang 10 ?

6 Le tableau ci-dessous illustre l'influence de la pluviométrie sur le rendement d'une variété de blé en tonnes par hectare au cours d'une année :

Pluviosité (en mm) : X	165	130	120	80	40	30
Rendement (en t/ha) : Y	0,37	0,30	0,29	0,25	0,20	0,12

1) Calculer les coordonnées du point moyen.
2) Représenter, dans un repère orthogonal, cette série double par un nuage de points.
3) Peut-on envisager un ajustement affine résumant le nuage ? Si oui donner une équation de la droite de régression de Y en X.

7 Les hauteurs barométriques (liées à la pression atmosphérique) varient en fonction de l'altitude conformément au tableau suivant :

Altitude (en km) : X	0	1	2	4	6	10
Hauteur barométrique (en cm) : Y	76	67	59	46	35	20

1) a) Construire, dans un repère orthogonal, le nuage de points.
   b) Peut-on justifier un ajustement affine de la série considérée ?
2) Utiliser la méthode de Mayer pour donner une équation cartésienne d'une droite résumant le nuage de points.
3) En déduire l'altitude d'un lieu où la hauteur barométrique est de 60 cm.

8 Le tableau suivant représente une série chronologique donnant le chiffre d'affaires Y (en millions de D.T) d'une entreprise au cours des dix dernières années. (X est le rang de l'année).

X	1	2	3	4	5	6	7	8	9	10
Y	8	12	12	15	17	22	27	32	37	42

1) a) Calculer le coefficient de corrélation linéaire de la série double (X,Y) ainsi définie.
   b) Que peut-on en déduire ?
2) Déterminer deux équations des droites de régression D_1 et D_2 dans un repère orthogonal du plan.
3) En supposant que l'évolution du chiffre d'affaires reste la même, donner le rang de l'année à partir duquel le chiffre d'affaires dépassera 50 millions de dinars tunisiens.

9 On donne les résultats obtenus à partir de 10 essais de laboratoire concernant « la charge y, en kg, de rupture d'un acier » en fonction de « sa teneur x, pour 10.000 unités, en carbone ».

X	72	60	68	66	64	62	64	70	62	74
Y	90	70	80	80	75	75	80	85	70	100

1) Représenter le nuage de points de cette série statistique double.
2) Calculer $\overline{X}$, $\overline{Y}$, $V(X)$, $V(Y)$ et $cov(X,Y)$.
3) Est-il possible d'envisager une liaison affine entre X et Y ? Justifier la réponse à l'aide d'un calcul.
4) Déterminer et construire la droite de régression de Y en X.
5) Quelle pourrait -être la charge de rupture d'un acier ayant une teneur en carbone de 65 pour 10.000 unités?

10 Les nombres de litres d'essence achetés à une pompe d'essence de 8 véhicules selon leurs puissances (en chevaux) sont consignés dans le tableau suivant :

Puissance (en chevaux) : x	4	5	6	8	9	10	11	12
Nombre de litres d'essence : y	10	20	25	30	40	32	25	45

1) a) Calculer $\sigma(x)$, $\sigma(y)$ et $cov(x,y)$
b) Calculer le coefficient de corrélation linéaire r du couple (x, y).
c) Quelle conclusion peut-on en tirer ?
2) Déterminer les équations des droites de régression correspondant à cette série statistique double.
3) Quel volume d'essence acheté peut-on prévoir pour un véhicule de puissance 7 chevaux ?

11 Un relevé statistique des tailles X (en cm) et des poids Y (en kg) d'un échantillon de 100 enfants a permis de construire le tableau suivant :

X \ Y	[20,25[	[25,30[	[30,35[	[35,40[
[100,110[	18	10	2	0
[110,120[	3	16	5	1
[120,130[	0	5	13	5
[130,140[	0	2	6	14

1) Construire, dans un repère orthogonal du plan, le nuage de points associé à cette série statistique double.
2) Déterminer $\overline{X}$, $\overline{Y}$, $V(X)$, $V(Y)$.
3) Calculer $cov(X,Y)$ en utilisant la formule : $cov(X,Y) = \dfrac{1}{100} \sum_{\substack{1 \le i \le 4 \\ 1 \le j \le 4}} n_{ij}.c_i c'_j$

où les c_i sont les centres des classes de la variable X et les c'_j sont les centres des classes de la variable Y ; n_{ij} désigne l'effectif correspondant à $(X = x_i \text{ et } Y = y_i)$.
4) Déterminer le coefficient de corrélation linéaire r du couple (X, Y) et dire si on peut estimer un ajustement affine entre X et Y.
5) Déterminer la droite de régression de Y en X.

12 Les notes de 15 élèves d'une classe de 1ère année secondaire dans le devoir de synthèse numéro 2 : X en Français, Y en Mathématiques et Z en sciences physique sont consignés dans le tableau suivant :

N° de l'élève	X = notes de Français	Y = notes de Maths	Z = notes de Sc.Physique
1	6.5	12	7,5
2	6.5	10	7
3	6	8	5
4	8	7.5	4.5
5	9.5	14.5	8
6	4.5	8.5	5.5
7	12	12	6

8	6.5	12	7,5
9	6.5	10	7
10	6	8	5
11	8	7.5	4.5
12	9.5	14.5	8
13	4.5	8.5	5.5
14	12	12	6
15	12	12	7
moyenne	....	....	....
médiane	....	....	....
mode	....	....	....
écart-type	....	....	....

A) Compléter le tableau (on pourra utiliser une feuille de calcul Excel ou une calculatrice).

B) On se propose d'étudier la relation entre les variables statistiques X et Y.

1) a) Construire, dans un repère orthogonal, le nuage de points de la série double (X, Y)

b) Peut-on justifier un ajustement affine de la série considérée ?

c) Donner une équation cartésienne de la droite de Mayer associée au nuage de points.

2) a) Calculer le coefficient de corrélation linéaire r du couple (X, Y).

b) Quelle conclusion peut-on en trier ?

3) Déterminer les équations des droites de régression correspondantes.

C) Etudier de même, les séries doubles : (X, Z) et (Y, Z).

13 Dans le tableau statistique suivant : X désigne la température moyenne extérieure en 24 heures et Y désigne la consommation de pétrole de chauffage pour les mêmes 24 heures et pour une famille donnée.

X en degrés	-2	0	4	8	10
Y en litres	40	30	20	15	10

1) Calculer le coefficient de corrélation linéaire entre X et Y et vérifier qu'il y a une forte corrélation linéaire entre ces deux variables.

2) Déterminer une équation cartésienne de la droite de régression de Y en X.

3) Quelle prévision (en litres) sur sa consommation de pétrole peut faire la famille considérée, si une vague de froid persiste pendant 48 heures avec une température moyenne extérieure de -4° ?.

14 L'évolution du taux de chômage dans un pays au cours de 7 années successives est la suivante :

T	1	2	3	4	5	6	7
C	6,6	5,9	7,2	7	8,7	8	9,3

T : temps et C : taux de chômage (en%)

1) Construire, dans un repère orthogonal, le nuage de points de la série double (T, C).

2) Peut-on justifier un ajustement affine de la série double (T, C) ?

3) Trouver par la méthode des moindres carrés l'équation c = at + b.

15 Le tableau suivant indique la distribution de 50 logements en fonction de leur nombre X de pièces principales et de leur surface Y en m^2.

X \ Y	30	50	70	90	120
1	1				
2	1	2	2		
3		1	6	6	
4			2	16	5
5				4	4

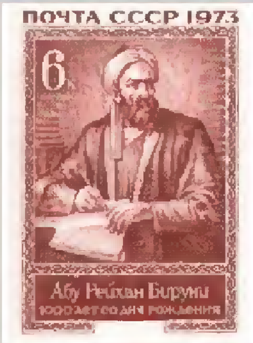
1) a) Déterminer les distributions marginales associées à X et à Y.
b) Calculer X, Y, V(X) ; V(Y) ; $\sigma(X)$ et $\sigma(Y)$.
2) Construire le nuage des points, représentant la série statistique double donnée.
3) a) Calculer Cov(X, Y) et le coefficient de corrélation linéaire r du couple (X, Y).
b) Déterminer une équation cartésienne de la droite D de régression de Y en X.

16 Soient deux séries d'observations relatives à la quantité consommée, X, de parfum en dl et le revenu annuel,Y, des ménages en milliers de D.T:

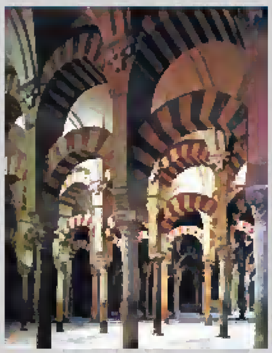
xi	2	4	7	10	11	14
yi	0.5	0.9	9	25	62	110
xi	18	20	21	22		
yi	180	290	680	1800		

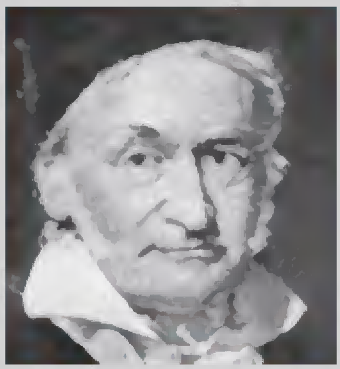
1) a) Construire, dans un repère orthogonal, le nuage de points de la série double (X, Y)
b) Peut-on justifier un ajustement affine de la série considérée ?
c) Calculer le coefficient de corrélation linéaire du couple (X, Y) et conclure.
2) On pose U = ln (X) et V = ln (Y).
a) Construire, dans un repère orthogonal, le nuage de points de la série double (U, V).
b) Peut-on justifier un ajustement affine de la série double (U, V) ?
c) Calculer le coefficient de corrélation linéaire du couple (U, V) et conclure.
d) Prouver que : $v = 3{,}31.u - 3{,}91$.
e) En déduire que : $y = 0{,}02e^{3{,}31.\ln x}$.
f) Quel revenu annuel peut-on prévoir pour une quantité de 13 dl de parfum consommé ?

APERÇU HISTORIQUE



Al Birouni
(973 - 1048)




Carl Friedrich Gauss
(1777 / 1855)

Depuis le 20ème siècle av. J.C, les chinois et les Egyptiens étudiaient l'évolution de leurs productions agricoles et organisaient des recensements de leurs populations.

Dans l'Antiquité, les Grecs et les Romains traitaient des données chiffrées relatives à leurs populations et à la richesse de leurs territoires.

Du 8$^{\text{ème}}$ siècle jusqu'au 15$^{\text{ème}}$ siècle, les mathématiciens arabes, influencés par les cultures babylonienne, grecque et indienne, furent particulièrement novateurs en calcul numérique, en algèbre et en trigonométrie. Ils ont introduit l'analyse combinatoire, l'analyse numérique et l'interpolation et ont ainsi contribué au développement des techniques permettant le traitement des informations chiffrées qu'ils ont mis en œuvre dans la résolution des problèmes de la vie courante et en astronomie. Citons entre autres : Al Khawarizmi, Thabet Ibn Quorra, Abou Kamil, Al Birouni, Omar Al Khayam, Al battani, Abou Al Wafa, Ibn Al Haythem et Al Kashi.

Au 18$^{\text{ème}}$ siècle, le développement des statistiques a permis l'instauration des bases solides aux prévisions et aux décisions relatives à la vie courante.

Au 19$^{\text{ème}}$ siècle, le mathématicien Laplace (Laplace Pierre Simon (1749-1827)) découvrit que la théorie des probabilités pouvait constituer une aide précieuse aux méthodes statistiques.

Pour étudier et préciser les formes de la terre (ellipsoïde aplatie), les mouvements oscillatoires de la Lune ainsi que ceux de Saturne, Euler (1748) et Mayer (1750) établissent la méthode des moyennes. Avant les travaux de ces deux mathématiciens, la statistique n'était qu'une observation des faits et des classements de données .

Gauss donnera les premiers algorithmes qui vont permettre de rationaliser les études statistiques, il utilise la méthode des moindres carrés.

Les travaux de Boscovich (1755) permettent de minimiser la somme des valeurs absolues des écarts, facilitant des calculs statistiques souvent trop longs et source d'erreurs.

De nos jours, la statistique est considérée comme outil efficace et fiable pour analyser des phénomènes, interpréter des données, faire des prévisions et prendre des décisions dans les divers domaines économiques, politiques, sociaux, biologiques ou physiques.