

I- Séries Statistiques Doubles :

◆ **Exemple** : Le tableau suivant donne le poids en Kg et la taille en cm d'un groupe de 10 enfants :

P_i	25	27	23	30	27	23	25	30	32	28
T_i	90	92	85	99	93	88	92	98	99	90

► Le couple $(P_1, T_1) = (25, 90)$ veut dire que l'enfant N°1 pèse 25 Kg et mesure 90 cm.

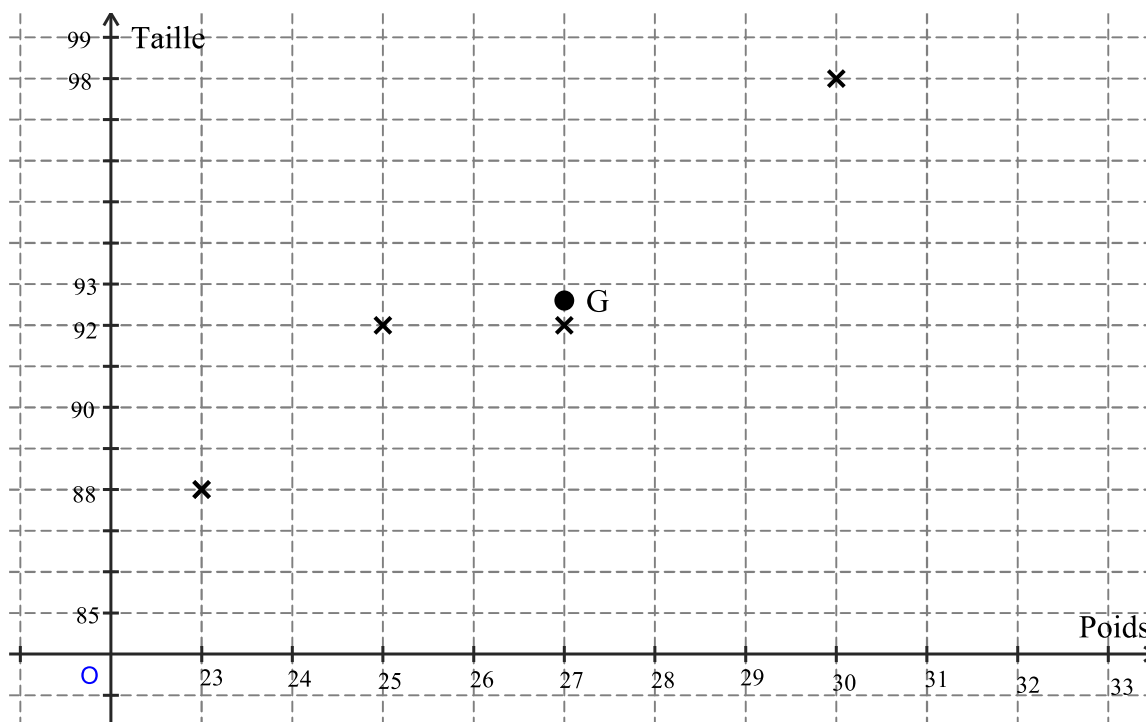
► On a donc une population de 10 enfants sur laquelle on a observé simultanément les deux variables P et T.

◆ **Définition** : On dit qu'un couple (X, Y) de variables statistiques définit une série double si les deux variables X et Y sont observés simultanément sur une même population.

► La moyenne arithmétique des poids est : $\bar{P} = \dots\dots\dots$

► La moyenne arithmétique des Tailles est : $\bar{T} = \dots\dots\dots$

► Placer dans un repère orthogonal l'ensemble des points $M_i(P_i, T_i)$:



◆ **Définition** : Soit une série statistique définie par deux variables 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 de Y. Le plan étant rapporté à un repère orthogonal.

L'ensemble des points $M_i(x_i, y_i)$; $i \in \{1, 2, \dots, n\}$ est appelé Nuage De Points.

Le point $G(\bar{x}, \bar{y})$ est appelé point moyen du nuage.

◆ **Distributions marginales :**

Soit le tableau statistique suivant : X : note en mathématiques ; Y : nombre de frères et sœurs. N = 100

X \ Y	0	1	2	3	4	5	6	Totaux
[0,4[1	0	1	1	0	1	1	
[4,8[2	2	4	3	3	4	2	20
[8,12[5	5	10	7	6	4	3	
[12,16[2	3	5	5	4	4	2	25
[16,20[1	1	2	3	2	1	0	
Totaux	11		22		15		8	100

- Les totaux inscrits en marge de chaque tableau à double entrée définissent deux distributions marginales, l'une associée à la première variable statistique et l'autre associée à la deuxième variable statistique.

X : Note en Maths	[0,4[[4,8[[8,12[[12,16[[16,20[Total
$\sum_{j=1}^7 n_{ij}$		20		25	10	100

Distribution marginale de X

Y : nombre de frères et sœurs	0	1	2	3	4	5	6	Total
Effectif $n_{\bullet j}$	11		22		15		8	100

Distribution marginale de Y

- Calcul de la **moyenne** (\bar{X}) ; la **variance** ($V(X)$) et l'**écart-type** ($\sigma(X)$)

► $\bar{X} = \frac{\sum_{i=1}^p x_i n_i}{N} = \frac{(2 \times 5) + (6 \times 20) + (10 \times 40) + (14 \times 25) + (18 \times 10)}{100} = \dots\dots\dots$

► $V(X) = \frac{\sum_{i=1}^p x_i^2 \cdot n_i}{N} - \bar{X}^2 = \frac{(2^2 \times 5) + (6^2 \times 20) + (10^2 \times 40) + (14^2 \times 25) + (18^2 \times 10)}{100} - \bar{X}^2 = \dots\dots\dots$

► $\sigma(X) = \sqrt{V(X)} \approx \dots\dots\dots$

► $\bar{Y} = \frac{\sum_{i=1}^q y_i n_i}{N} = \frac{(0 \times 11) + (1 \times 11) + (2 \times 22) + (3 \times 19) + (4 \times 15) + (5 \times 14) + (6 \times 8)}{100} = \dots\dots\dots$

► $V(Y) = \frac{\sum_{i=1}^q y_i^2 \cdot n_i}{N} - \bar{Y}^2 = \frac{(0^2 \times 11) + (1^2 \times 11) + (2^2 \times 22) + \dots\dots\dots + (6^2 \times 8)}{100} - \bar{Y}^2 = \dots\dots\dots$

► $\sigma(Y) = \sqrt{V(Y)} \approx \dots\dots\dots$

II- 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.

1) Méthode de Mayer :

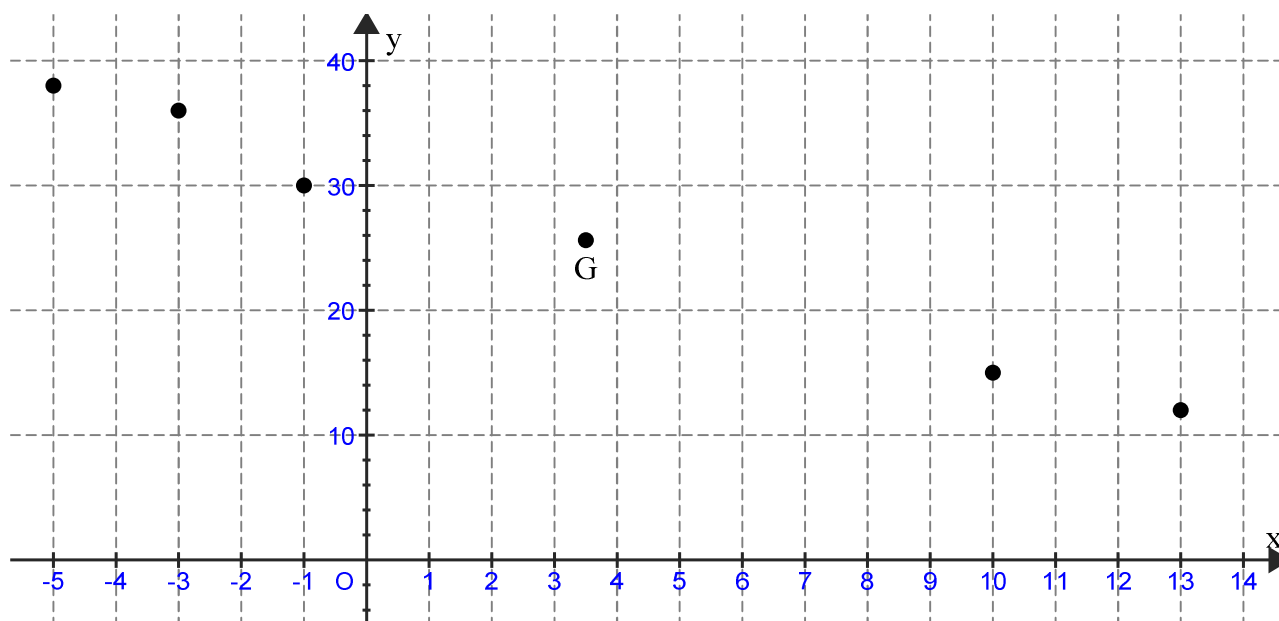
La méthode de Mayer consiste à :

- Partager le nuage de points en deux parties P_1 et P_2 situées de part et d'autre par rapport à une droite parallèle à l'axe des ordonnées et contenant à peu près le même nombre de points.
- Déterminer les points moyens respectifs G_1 et G_2 des parties P_1 et P_2 .
- La droite (G_1G_2) est alors la droite d'ajustement affine du nuage de points représentant la série.
- La droite (G_1G_2) est appelée droite de Mayer et passe par le point moyen G du nuage global.

◆ Exemple :

Le tableau ci-dessous présente la consommation de fuel d'une habitation en fonction de la température.

Température x_i en °C	-5	-3	-1	2	5	7	10	13
Consommation y_i de fuel /24h en L	38	36	30	29	25	20	15	12



- 1) Compléter le nuage de points $M(x_i, y_i)$ dans le repère ci-dessus.
- 2) Fractionner le nuage de points en deux parties égales.
- 3) Calculer les coordonnées du point moyen G_1 de la première partie du nuage.

$$G_1 \left(\frac{-5-3-1+2}{4} ; \frac{38+36+30+29}{4} \right) \text{ alors } G_1 (\dots ; \dots)$$

4) Calculer les coordonnées du point moyen G_2 de la deuxième partie du nuage.

$$G_2 (\dots ; \dots)$$

5) Tracer la droite $(G_1 G_2)$.

6) Calculer les coordonnées du point moyen G du nuage. $G (\dots ; \dots)$

7) Déterminer l'équation réduite de la droite $(G_1 G_2)$. $(y = ax + b)$.

$$a = \frac{y_{G_1} - y_{G_2}}{x_{G_1} - x_{G_2}} = \dots \approx -1,45$$

$$b = y_{G_1} - a \cdot x_{G_1} = \dots \approx 30,71 \text{ donc } (G_1 G_2) : y = \dots$$

8) A partir de l'équation de la droite, donner une estimation de la consommation de fuel pour une température de -10°C .

9) Déterminer graphiquement, à l'aide de la droite d'ajustement, la température pour une consommation de 22L.

10) Retrouver le résultat précédent par le calcul à partir de l'équation de $(G_1 G_2)$.

2) Méthode des Moindres carrés :

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 paramètres statistiques à savoir : la covariance $\text{Cov}(X, Y)$ et le coefficient de corrélation linéaire r .

◆ **Covariance** : Soit une série statistique (X, Y) double définie par $\{x_1, \dots, x_n\}$ et $\{y_1, \dots, y_n\}$ observée sur une population de n individus. On appelle covariance du couple (X, Y) le réel :

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

◆ **Exemple** : Soit la série statistique double définie par le tableau suivant, Compléter le tableau :

x_i	2	5	3	1	1	4	2	3	$\sum x_i = \dots$	$\bar{X} = \dots$
y_i	25	40	10	5	0	15	50	12	$\sum y_i = \dots$	$\bar{Y} = \dots$
$x_i y_i$	50		30		0		100		$\sum x_i y_i = \dots$	$\overline{XY} = \dots$

$$\text{Cov}(X, Y) = \overline{XY} - \bar{X} \bar{Y} = \dots$$

◆ **Exercice** : Calculer la covariance de la série statistique double (X, Y) définie par :

x_i	1	2	2	2	5
y_i	7	8	9	5	8

$\bar{X} =$

$\bar{Y} =$

$\overline{XY} =$

$\text{Cov}(X, Y) = \overline{XY} - \bar{X} \bar{Y} =$

◆ **Coefficient de corrélation linéaire** :

On appelle coefficient de corrélation linéaire le réel r défini par : $r = \frac{\text{Cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}$; $r \in [-1, 1]$

◆ **Exercice** :

Calculer le coefficient de corrélation linéaire r pour la série statistique suivante :

x_i	1	2	3	4	5	6	7	8	9	10
y_i	200	205	211	216	220	225	240	260	280	300

• $\bar{X} =$

• $\bar{Y} =$

• $\overline{XY} =$

• $V(X) =$

• $V(Y) =$

• $\text{Cov}(X, Y) =$ • $\sigma(X) =$ • $\sigma(Y) =$

• $r =$

◆ **Théorème** :

X et Y deux variables statistiques observées sur une population d'effectif N.

- Si $0,75 \leq |r| \leq 1$ alors il y a une relation linéaire entre X et Y ; ($Y = aX + b$; $X = a'Y + b'$)

représentées graphiquement par deux droites passant par $G(\bar{X}, \bar{Y})$.

- $Y = aX + b$: Droite de régression de Y en X avec $a = \frac{\text{Cov}(X, Y)}{V(X)}$ et $b = \bar{Y} - a\bar{X}$.

- $X = a'Y + b'$: Droite de régression de X en Y avec $a' = \frac{\text{Cov}(X, Y)}{V(Y)}$ et $b' = \bar{X} - a'\bar{Y}$.

◆ **Exercice :**

Soit la série statistique suivante :

x_i	100	150	200	300	500
y_i	0,7	1	1,2	1,6	2,3

Donner les résultats arrondi à 10^{-4} près si nécessaire.

- 1) Calculer le coefficient de corrélation linéaire r .
- 2) Existe-t-il une relation de type affine entre X et Y .
- 3) Déterminer une équation de la droite de régression de Y en X .
- 4) Pour $x = 650$ que peut-on prévoir pour y ?.

◆ **Solution :**

- 1) • $\bar{X} = \frac{1}{5}(100 + 150 + 200 + 300 + 500) = 250$;
• $V(X) = \frac{1}{5}(100^2 + 150^2 + 200^2 + 300^2 + 500^2) - 250^2 = 20000$
• $\sigma(X) = \sqrt{20000} \approx 141,4214$
• $\bar{Y} = \frac{1}{5}(0,7 + 1 + 1,2 + 1,6 + 2,3) = 1,36$;
• $V(Y) = \frac{1}{5}(0,7^2 + 1^2 + 1,2^2 + 1,6^2 + 2,3^2) - 1,36^2 \approx 0,3064$
• $\sigma(Y) = \sqrt{0,3064} \approx 0,5535$
• $\overline{XY} = \frac{1}{5}((100 \times 0,7) + (150 \times 1) + (200 \times 1,2) + (300 \times 1,6) + (500 \times 2,3)) = 418$
• $\text{Cov}(X, Y) = \overline{XY} - \bar{X}\bar{Y} = 418 - 250 \times 1,36 = 78$
• $r = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \approx 0,9965$
- 2) $r \approx 0,9965$ donc $0,75 \leq r \leq 1$ alors il existe une relation de type affine entre X et Y .
- 3) Equation de la droite de régression de Y en X : $Y = aX + b$
avec $a = \frac{\text{Cov}(X, Y)}{V(X)} = 0,0039$ et $b = \bar{Y} - a\bar{X} = 0,385$
- 4) Si $x = 650$ alors $y = 0,0039 \times 650 + 0,385 = 2,92$

◆ **Utilisation de la calculatrice :** (Exemple pour une calculatrice Sharp)

Le tableau suivant donne la distance de freinage d (en mètres) d'une voiture, en fonction de sa vitesse v (en kilomètres par heure)

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

1) Calculer \bar{v} , \bar{d} , $V(v)$, $V(d)$, $\sigma(v)$, $\sigma(d)$, $Cov(v,d)$ et le coefficient de corrélation linéaire entre v et d .

2) Déterminer une équation de la droite de régression de d en v ($d = a v + b$).

- Choisir le mode statistique double : 2ndF DRG 2

- Entrer les données en tapant

30	STO	42	M+
40	STO	60	M+
50	STO	80	M+
60	STO	90	M+
70	STO	95	M+
80	STO	110	M+

- \bar{v} : RCL 4 → $\bar{v} = 55$
- \bar{d} : RCL 7 → $\bar{d} = 79,5$
- $V(v)$: RCL 6 x² = → $V(v) = 291,66$
- $V(d)$: RCL 9 x² = → $V(d) = 511,25$
- $\sigma(v)$: RCL 6 → $\sigma(v) \approx 17,07$
- $\sigma(d)$: RCL 9 → $\sigma(d) \approx 22,61$
- $Cov(v,d)$: RCL ÷ × RCL 6 × RCL 9 = → $Cov(v,d) = 379,166$.
- r : RCL ÷ → $r \approx 0,981$
- a : RCL) → $a = 1,3$
- b : RCL (→ $b = 8$
- Donc : $d = 1,3v + 8$