I- Séries Statistiques Doubles :

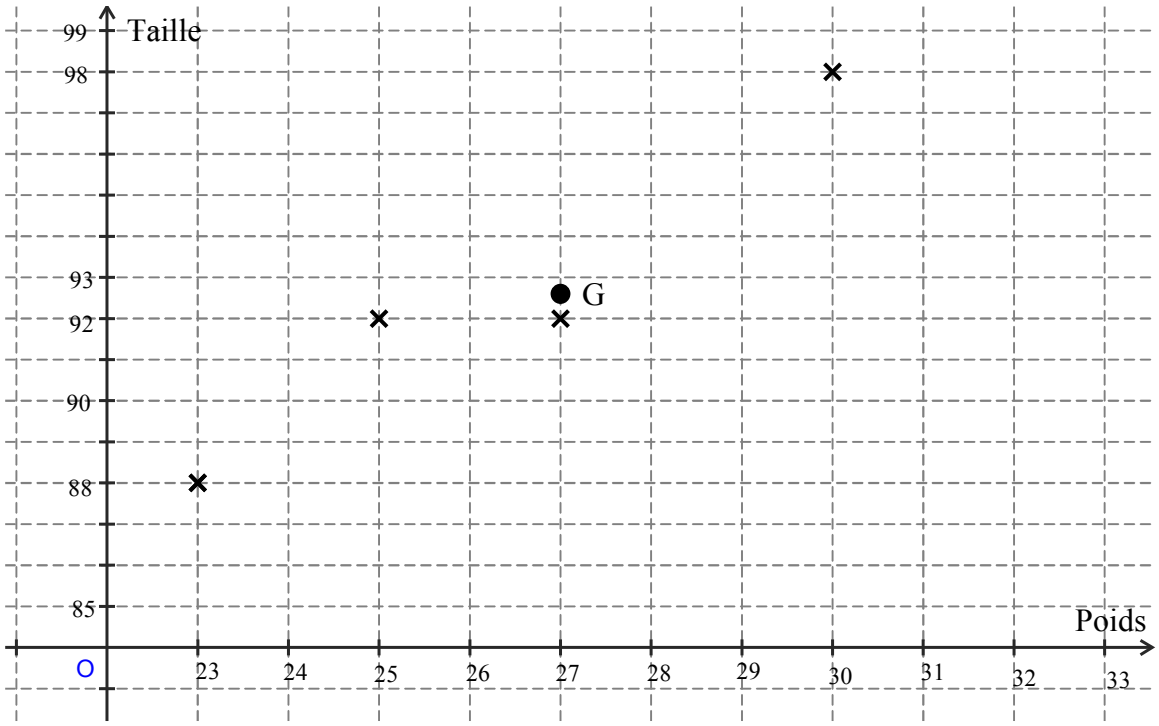
◆ **Exemple** : Le tableau suivant donne le poids en Kg et la taille en cm d'un groupe de 10 enfants :

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 définies une série double si les deux variables X et Y sont observés simultanément sur une même population.

- La moyenne arithmétique des poids est : $\overline{P} =$ .....
- La moyenne arithmétique des Tailles est : $\overline{T} =$ .....
- 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 $x_1, x_2, ..., x_n$ les valeurs de X et par $y_1, y_2, ..., y_n$ celles de Y. Le plan étant rapporté à un repère orthogonal.

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

Le point $G(\overline{x}, \overline{y})$ 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 soeurs	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 ($\overline{X}$) ; la variance** $(V(X))$ **et l'écart-type** $(\sigma(X))$

▸ $\overline{X} = \dfrac{\sum_{i=1}^{p} x_i n_i}{N} = \dfrac{(2\times 5)+(6\times 20)+(10\times 40)+(14\times 25)+(18\times 10)}{100} =$ .....

▸ $V(X) = \dfrac{\sum_{i=1}^{p} x_i^{\,2} \cdot n_i}{N} - \overline{X}^2 = \dfrac{(2^2\times 5)+(6^2\times 20)+(10^2\times 40)+(14^2\times 25)+(18^2\times 10)}{100} - \overline{X}^2 =$ .....

▸ $\sigma(X) = \sqrt{V(X)} \approx$ .....

▸ $\overline{Y} = \dfrac{\sum_{i=1}^{q} y_i n_i}{N} = \dfrac{(0\times 11)+(1\times 11)+(2\times 22)+(3\times 19)+(4\times 15)+(5\times 14)+(6\times 8)}{100} =$ .....

▸ $V(Y) = \dfrac{\sum_{i=1}^{q} y_i^{\,2} \cdot n_i}{N} - \overline{Y}^2 = \dfrac{(0^2\times 11)+(1^2\times 11)+(2^2\times 22)+\ldots\ldots\ldots+(6^2\times 8)}{100} - \overline{Y}^2 =$ .....

▸ $\sigma(Y) = \sqrt{V(Y)} \approx$ .....

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 :

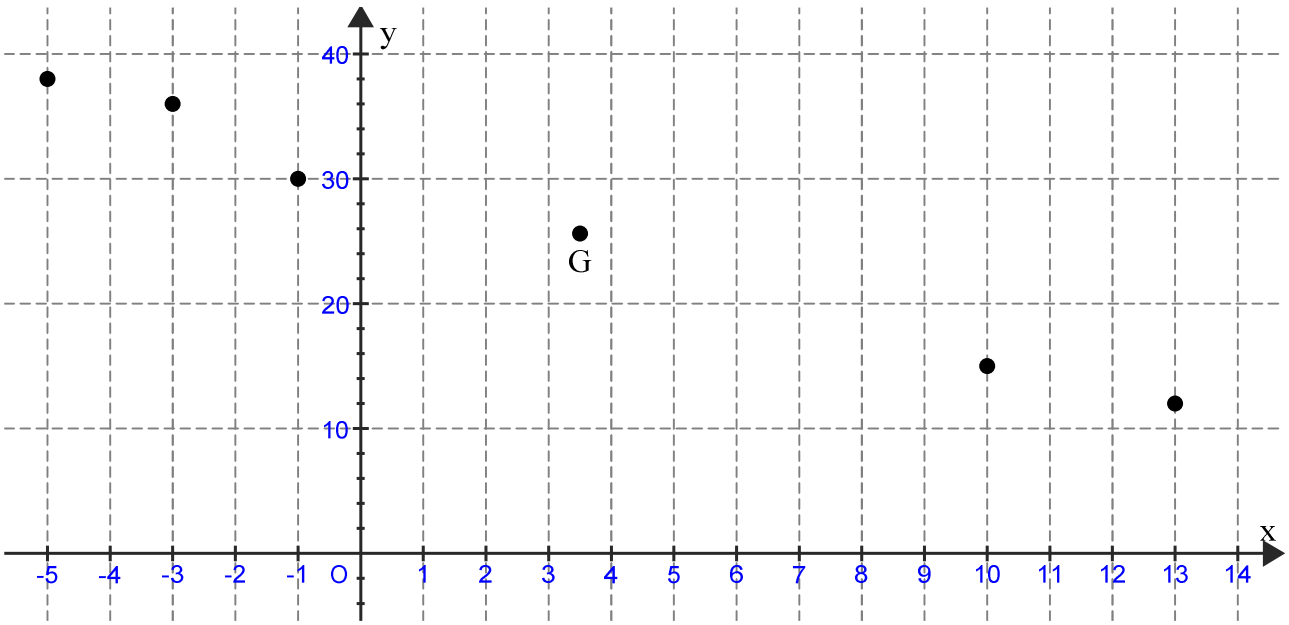
La méthode de Mayer consiste à :

- 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 prés 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.

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

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

$G_2(\ldots\ldots\ldots\ldots \ ; \ \ldots\ldots\ldots\ldots)$

5) Tracer la droite (G_1G_2).

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

7) Déterminer l'équation réduite de la droite (G_1G_2). $(y = ax + b)$.

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

$b = y_{G_1} - a \cdot x_{G_1} = \ldots\ldots\ldots\ldots \approx 30,71$ donc $(G_1G_2) : y = \ldots\ldots\ldots\ldots$

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_1G_2).

2) <u>Méthode des Moindres carrés</u> :

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 $Cov(X,Y)$ et le coefficient de corrélation linéaire r.

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

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

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

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

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

◆ **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

$\overline{X}$ = ...

$\overline{Y}$ = ...

$\overline{XY}$ = ...

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

◆ **Coefficient de corrélation linéaire** :

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

◆ **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

- $\overline{X}$ = ... • $\overline{Y}$ = ...
- $\overline{XY}$ = ...
- $V(X)$ = ...
- $V(Y)$ = ...
- $Cov(X, Y)$ = ... • $\sigma(X)$ = ... • $\sigma(Y)$ = ...
- 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 ; $(Y = a\ X + b \ ; \ X = a'Y + b')$ représentées graphiquement par deux droites passant par $G(\overline{X}, \overline{Y})$.
- $Y = a\ X + b$: Droite de régression de Y en X avec $a = \dfrac{Cov(X, Y)}{V(X)}$ et $b = \overline{Y} - a\overline{X}$.
- $X = a'Y + b'$:Droite de régression de X en Y avec $a' = \dfrac{Cov(X, Y)}{V(Y)}$ et $b' = \overline{X} - a'\overline{Y}$.

◆ **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)
- $\overline{X} = \frac{1}{5}(100+150+200+300+500) = 250$;
- $V(X) = \frac{1}{5}\left(100^2+150^2+200^2+300^2+500^2\right) - 250^2 = 20000$
- $\sigma(X) = \sqrt{20000} \approx 141,4214$
- $\overline{Y} = \frac{1}{5}(0,7+1+1,2+1,6+2,3) = 1,36$;
- $V(Y) = \frac{1}{5}\left(0,7^2+1^2+1,2^2+1,6^2+2,3^2\right) - 1,36^2 \approx 0,3064$
- $\sigma(Y) = \sqrt{0,3064} \approx 0,5535$
- $\overline{XY} = \frac{1}{5}\left((100\times0,7)+(150\times1)+(200\times1,2)+(300\times1,6)+(500\times2,3)\right) = 418$
- $Cov(X,Y) = \overline{XY} - \overline{X}\,\overline{Y} = 418 - 250\times1,36 = 78$
- $r = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} \approx 0,9965$

2) $r \approx 0.9965$ donc $0,75 \le r \le 1$ 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 $a = \frac{Cov(X,Y)}{V(X)} = 0,0039$ et $b = \overline{Y} - a\overline{X} = 0,385$

4) Si $x = 650$ alors $y = 0,0039\times650 + 0,385 = 2,92$

◆ **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] $\rightarrow \bar{v} = 55$
- $\bar{d}$: [RCL] [7] $\rightarrow \bar{d} = 79,5$
- $V(v)$: [RCL] [6] [x^2] [=] $\rightarrow V(v) = 291,66$
- $V(d)$: [RCL] [9] [x^2] [=] $\rightarrow V(d) = 511,25$
- $\sigma(v)$: [RCL] [6] $\rightarrow \sigma(v) \approx 17,07$
- $\sigma(d)$: [RCL] [9] $\rightarrow \sigma(d) \approx 22,61$
- $Cov(v,d)$: [RCL] [÷] [×] [RCL] [6] [×] [RCL] [9] [=] $\rightarrow Cov(v,d) = 379,166$.
- r : [RCL] [÷] $\rightarrow r \approx 0,981$
- a : [RCL] [)] $\rightarrow a = 1,3$
- b : [RCL] [(] $\rightarrow b = 8$
- Donc : $d = 1,3v + 8$