I) **Rappels** :

◆**Exemple** : Dans un groupe de dix élèves, voici les notes à un devoir : 12, 4, 16, 16, 10, 7, 9, 12, 9, 12.

- Cette série de note est une série statistique $X=\{x_1, x_2, x_3, x_4, x_5, x_6\}$ avec x_i les valeurs prises par X, à chaque valeur x_i on associe un effectif n_i égale au nombre d'élèves ayant obtenu la note x_i.
- La fréquence de la valeur d'une série est : $f_i=\frac{n_i}{N}$ avec N l'effectif total $(N=\sum n_i)$.

x_i (Notes)	4	7	9	10	12	16	
n_i (Effectif)			2				$\sum n_i =$
f_i (Fréquence)		0,1				0,2	$\sum f_i =$
$x_i\ f_i$				1	3,6		$E(X)=$
${x_i}^2\ f_i$	1,6					51,2	$E(X^2)=$

- La moyenne de ce groupe est :.....
- La variance de X est : $V(X)=E(X^2)-(E(X))^2=$.....
- L'écart-type de X est : $\sigma(X)=\sqrt{V(X)}=$.....
- Le tableau suivant donne la durée (en heures) passée chaque jour devant la télévision pour 1000 personnes.

Durées (en heures)	[0;2[	[2;4[	[4;6[	[6;8[	[8;10[	
C_i (centre de classe)			5			
n_i (Effectif)	100	200	500	150	50	$\sum n_i =$
f_i (Fréquence)						$\sum f_i =$
$C_i\ f_i$						$E(X)=$
${C_i}^2\ f_i$						$E(X^2)=$

- $V(X)=E(X^2)-(E(X))^2=$…… - $\sigma(X)=\sqrt{V(X)}=$……

◆ **Définitions** :

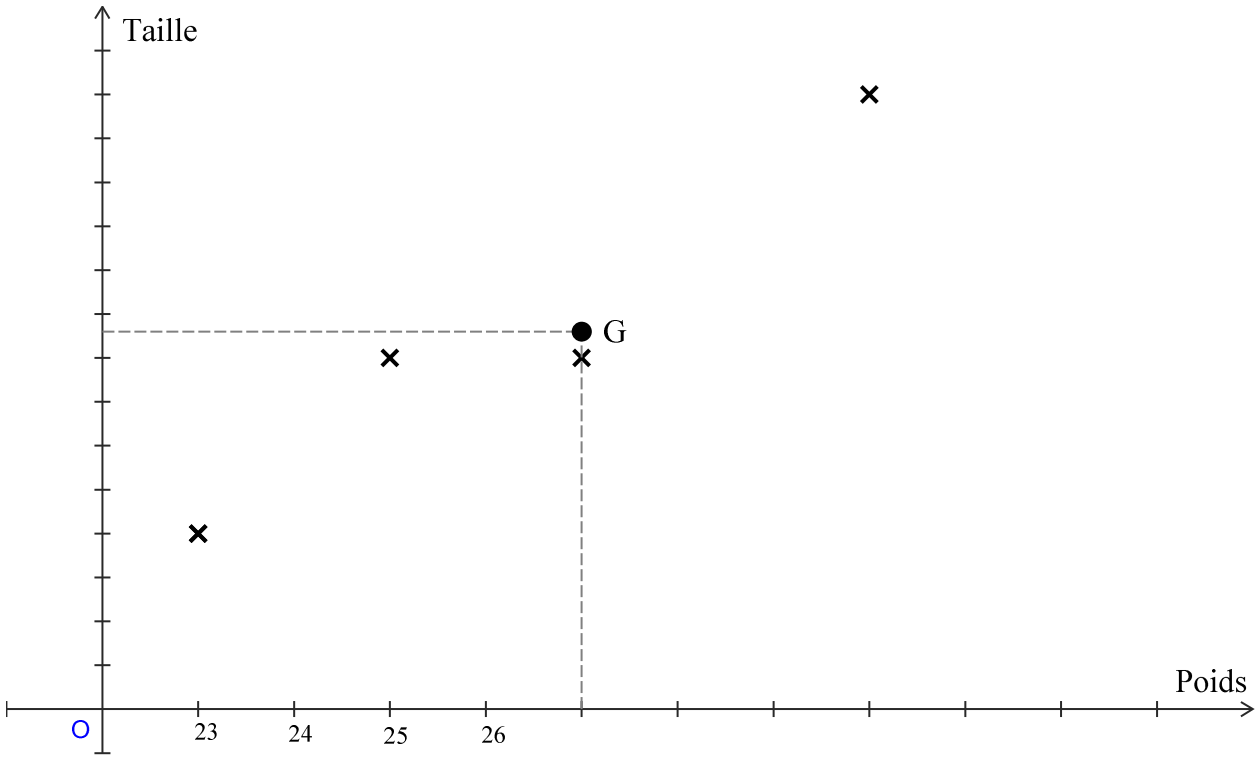
- ❖ La première série est une série quantitative discrète.
- ❖ La deuxième série est une série quantitative continue.
- ❖ On appelle Mode ou classe Modale la valeur de X correspondant à l'effectif le plus haut.
  - Pour la série quantitative discrète le Mode est :……..
  - Pour la série quantitative continue la classe Modale est :……..
- ❖ On appelle Médiane d'une série statistique d'effectif total N la valeur de la variable qui correspond à l'effectif cumulé $\frac{N}{2}$ si N est pair et $\frac{N+1}{2}$ si N est impair.

x_i (Notes)	4	7	9	10	12	16
n_i (Effectif)	1	1	2	1	3	2
E_i (Eff.Cum.↗)						

$\frac{N}{2}=5$ alors $M_e=$…….

I_i (Durées)	[0,2[	[2,4[	[4,6[	[6,8[	[8,10[
n_i (Effectif)	100	200	500	150	50
E_i (Eff.Cum.↗)					

$\frac{N}{2}=500$; • ⟨500⟨ • alors • ⟨M_e⟨ • alors $\frac{M_e-4}{500-300}=\frac{6-4}{800-300}$ D'où $M_e=$…


Taille
G
Poids
O
23
24
25
26

▸ Compléter le tableau à double entrée :

X \ Y	7	8	9	12	13	Totaux
1	4	1	3	0	0	8
2	0					
3	0		0			
Totaux	4				3	

◆ **Définition** : Soient X et Y deux variables statistiques définies sur une population E d'effectif total N.
On désigne par $(x_i)_{1 \le i \le p}$ les modalités de X et par $(y_j)_{1 \le j \le q}$ celles de Y.

▸ Le nombre d'individus de la population E vérifiant simultanément $X = x_i$ et $Y = y_j$ est noté n_{ij} et est appelé l'effectif associé au couple (x_i, y_j).

▸ Le réel $f_{ij} = \frac{n_{ij}}{N}$ est appelé la fréquence associée au couple (x_i, y_j).

▸ Le réel $p_{ij} = 100\, f_{ij}$ est appelé la fréquence en pourcentage associée au couple (x_i, y_j).

◆ **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				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			10	

Distribution marginale de X

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

Distribution marginale de Y

▸ Calculer $\overline{X}$; V(X) et σ(X)

▸ Calculer $\overline{Y}$; V(Y) et σ(Y)

III) Liaison entre deux caractères :

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

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

◆ **Remarque** : On développant la formule précédente, on obtient : $Cov(X,Y)=\frac{1}{n}\sum_{i=1}^{n}x_iy_i-\overline{X}\ \overline{Y}=\overline{XY}-\overline{X}\ \overline{Y}$

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

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

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

◆ **Remarque** : Dans le cas où la série statistique est définie par un tableau à double entrée :

$X=(x_i)_{i=1}^{p}$ et $Y=(y_j)_{j=1}^{q}$ on a $\overline{X}=\frac{1}{N}\sum_{i=1}^{p}n_{i\bullet}\cdot x_i=\sum_{i=1}^{p}f_{i\bullet}\cdot x_i$ et $\overline{Y}=\frac{1}{N}\sum_{j=1}^{q}n_{\bullet j}\cdot y_j=\sum_{j=1}^{q}f_{\bullet j}\cdot y_j$

$$\overline{XY}=\frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{q}n_{ij}\ x_i\ y_j=\sum_{i=1}^{p}\sum_{j=1}^{q}f_{ij}\ x_i\ y_j$$

◆ **Exemple** : Compléter le tableau suivant et en déduire $Cov(X,Y)$

x_i \ y_j	3	5	7	9	Totaux	$x_i\ n_{i\bullet}$
30	$n_{11}=20$; $n_{11}x_1y_1=1800$	15 ; 2250	5 ; 1050	0 ; 0	$n_{1\bullet}=40$; 5100	1200
40	10 ; 1200	5	4 ; 1120	1		
50	7 ; 1050	5	2 ; 700	1	15 ; 3450	750
60	1 ; 180	4	3	2 ; 1080	10 ; 3720	
70	2 ; 420	3	5 ; 2450	5		1050
Totaux	$n_{\bullet 1}=40$; 4650		19		100 ; $\sum_{i=1}^{p}\sum_{j=1}^{q}n_{ij}\ x_i\ y_j=$	$\sum_{i=1}^{p}n_{i\bullet}\cdot x_i=$ ....
$y_j\ n_{\bullet j}$	120		133		$\sum_{j=1}^{q}n_{\bullet j}\cdot y_j=$	

$\overline{X} = \frac{1}{N}\sum_{i=1}^{p} n_{i\bullet} \cdot x_i = \ldots\ldots$; $\overline{Y} = \frac{1}{N}\sum_{j=1}^{q} n_{\bullet j} \cdot y_j = \ldots\ldots$; $\overline{XY} = \frac{1}{N}\sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij}\, x_i\, y_j = \ldots\ldots$

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

◆ **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

$\overline{X} = \ldots\ldots$; $\overline{Y} = \ldots\ldots$; $\overline{XY} = \ldots\ldots$

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

◆ **Application de la covariance** :

Soient X et Y deux variables quantitatives observées dans une population d'effectif total n.
On suppose qu'il existe deux réels a et b tels que $Y = a\,X + b$.
On montre que : ▸ $V(Y) = a^2\ V(X)$.

▸ $Cov^2(X,Y) = V(X)\cdot V(Y)$ donc $\frac{Cov^2(X,Y)}{V(X)\cdot V(Y)} = 1$ alors $\frac{|Cov(X,Y)|}{\sigma(X)\cdot\sigma(Y)} = 1$.

▸ On remarque que s'il y a une relation affine entre X et Y alors $\frac{|Cov(X,Y)|}{\sigma(X)\cdot\sigma(Y)} = 1$

◆ **Définition** :

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

◆ **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

$\overline{X} =$ | $\overline{Y} =$

$V(X) =$ | $V(Y) =$

$\sigma(X) =$ | $\sigma(Y) =$

$\overline{XY} =$

$Cov(X,Y) =$

$r =$

◆ **Ajustement linéaire** :

On a vu que si $Y = a\ X + b$ alors $|r| = 1$,donc si $|r|$ est proche de 1 il existe une corrélation linéaire entre X et Y

▸ **Théorème** :

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

- Si $0,75 \leq |r| \leq 1$ ou le nuage de points a une forme allongée 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	2,5

1) Calculer le coefficient de corrélation linéaire r .

2) Existe-t-il une relation de type affine entre X et Y.

3) Déterminer l'équation de la droite de régression de Y par rapport à X .

4) Pour $x = 700$ que peut-on prévoir pour y ?.

◆ **Exercice :** (Bac 2005)

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 $\overline{v}$, $\overline{d}$, $V(v)$, $V(d)$ et $Cov(v,d)$.

2) a) Calculer le coefficient de corrélation entre v et d.

b) Y-a-t-il forte corrélation 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 140 km/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 ?