II- PARAMETRES DE DISPERSION.

On a décrit les distributions statistiques dans les chapitres précédents en utilisant:

- Les représentations graphiques qui font apparaître l'allure générale de la distribution (manque de précision).
- Les paramètres de tendance centrale qui décrivent la zone centrale de la distribution, mais sont incapables de donner une description sur la structure interne de la distribution et par conséquent comment sont dispersées les valeurs de la série statistique autour d'une caractéristique de position

Citons dans un premier temps un exemple concret :
1 ère série (notes sur 2O) :   4   20
2 ème série (notes sur 2O) : 11 13

On constate que les deux séries ont la même moyenne ($\overline{x}$ = 12) , mais la 2 ème série est moins dispersée que la 1 ère.

Donc il est nécessaire de disposer d'autres paramètres qui étudient la dispersion (structure interne).
Le paramètre le plus efficace pour rendre compte de la dispersion d'une série d'observations est **la variance** ou l'écart-type.

I- La variance.

C'est une distance moyenne des observations à la moyenne arithmétique qui constitue une mesure de dispersion, plus précisément c'est la moyenne arithmétique des carrés des écarts à la moyenne arithmétique

Cas des données non groupées :

On définit la variance S^2 d'un ensemble de n valeurs $x_1, x_2, \ldots, x_n$ par :

$$S^2 = (1/n) \sum (X_i - \overline{X})^2$$

Cas des données groupées :

Quand x1, x2, x3,.........xN ont respectivement des effectifs d'apparition n1, n2, n3 ....nN, la variance prend la forme :

$$\mathbf{S^2 = (1/n)\ \Sigma ni\ (Xi - \bar{X})^2}$$

II- L’écart-type :

La variance est une quantité élevée au carré. Si la variable x représente un poids exprimé en kg, alors la variance sera exprimée en kg², unité qui n'a pas de sens. On utilise l'écart- type qui est la racine carrée de la variance afin d'obtenir une unité homogène avec celle de la variable statistique x.

- Plus l'écart-type est faible, plus les valeurs de la série sont concentrées (moins dispersées) autour de leur moyenne.
- La moyenne et l'écart-type sont souvent utilisés dans les tests statistiques, c'est la raison pour laquelle ils doivent être bien calculés

Exemples :

1) Données non groupées :

Soit la série : 5, 7, 1O, 14.

La moyenne $\overline{X} = (1/n) \sum^{n} xi = (1/4) \sum^{4} xi = 36 / 4 = 9.$

Variance

$S^2=(1/n) \sum (Xi- \overline{X})^2 =(1/4) [(5-9)^2+(7-9)^2+(1O-9)^2+(14-9)^2] = 11,5$

L'écart-type $S = \sqrt{11,5} = 3,39$

$$S^2 = (1/n)\ \Sigma n_i (X_i - \overline{X})^2$$

$$\overline{X} = (1/n)\ \Sigma n_i x_i$$

2) **Données groupées :**

Le même calcul pour la variable discontinue et continue.
On prend les données de poids des 50 nouveaux nés :

classes	**centres de classes xi**	**ni**	**ni.xi**	**$(x_i-\overline{x})^2$**	**$n_i(x_i-\overline{x})^2$**
2,0-2,5	2,25	2	4,5	1,90	3,80
2,5-3,0	2,75	4	11	0,77	3,77
3,0-3,5	3,25	6	19,5	0,14	0,87
3,5-4,0	3,75	30	112,5	0,014	0,43
4,0-4,5	4,25	8	34	0,38	3,07
		Σ ni = 50	**Σ ni.xi =181,5**		**$\Sigma n_i.(x_i-\overline{x})^2$ =11,26**

la variance $S^2 = (1/n)\ \Sigma\ ni\ (Xi-X)^2 = (1/50)\ \Sigma\ ni\ (xi-3.63)^2$.

$= 11,26/5O = O,22$

L'écart-type $S = \sqrt{O,22} = O,47$ Kg.

CALCUL PRATIQUE DE L'ECART-TYPE ET DE LA VARIANCE

En développant la formule de définition :

$$S^2 = (1/n) \sum_{1}^{N} ni\,(Xi - \overline{X})^2$$ données groupées

et

$$S^2 = (1/n) \sum^{n} (Xi - \overline{X})^2$$ données non groupées.

On obtient une expression adaptée pour les calculs numériques :

$$S^2 = (1/n) \sum_{1}^{N} n_i X_i^2 - [\bar{x}]^2$$ données groupées

et

$$S^2 = (1/n) \sum_{1}^{n} X_i^2 - [\bar{x}]^2$$ données non groupées

Exemple : les données du tableau 2

Classes en (kg)	Centres de classes xi	Effectifs ni	ni.xi	ni.xi²
2,O - 2,5	2,25	2	4,5	10,125
2,5 - 3,O	2,75	4	11	30,225
3,O - 3,5	3,25	6	19,5	63,375
3,5 - 4,O	3,75	30	112,5	421,875
4,O - 4,5	4,25	8	34	144,5
		∑ni=50	**∑ni.xi=18 1.5**	**∑ni.xi²= 670,125**

L'écart-type S =$\sqrt{O,22}$ = O,47 Kg

Le coefficient de variation (coefficient de dispersion)

C'est une caractéristique de dispersion relative qui sert à comparer les dispersions des distributions qui ne sont pas de même nature. Il est défini par le rapport de l'écart- type à la moyenne.

$CV=S/\overline{X}$

Il est souvent exprimé sous forme de pourcentage, et est indépendant du choix des unités. Il permet de comparer les dispersions de deux séries différentes soit par les unités, soit par leur nature, exemple : une série de salaires en dinars et une série de salaires en dollars, ou encore une série de tailles à une série de poids.

En général dans la pratique quand :

CV > O,33 dispersion importante

CV < O,33 dispersion faible.