Module : Statistiques

Chapitre: les paramètres de dispersion

Nous avons vu que les valeurs centrales sont nécessaires pour caractériser une série statistique, cependant elles ne sont pas en général suffisantes pour donner une conclusion sur la répartition d'un caractère, ainsi par exemple on peut rencontrer deux populations ayant même valeur centrale mais elles n'ont pas une même dispersion autour du centre.

Soit par exemple les notes de deux étudiants dans cinq matières :

	L'étudiant N° 1	L'étudiant N° 2
Philosophie	6	2
Histoire	8	6
Mathématiques	10	10
Anglais	12	14
Français	14	18

Les deux étudiants ont même moyenne et même médiane : 10, mais nous remarquons que leurs notes ne sont pas réparties de la même façon autour de la valeur central. On dit donc que les deux séries n'ont pas la même répartition, et plus précisément la série 2 (notes de l'étudiant N° 2) est plus dispersée que la série 1 (note de l'étudiant N° 1).

1- L'étendue de la série :

L'étendue, appelée aussi amplitude ou intervalle maximale d'une série est égale à la différence des valeurs extrêmes de la série. Elle est notée E; Ainsi dans l'exemple précédent des notes des deux étudiants N°1 et N°2 la série {6, 8, 10, 12, 14} a pour étendue $E = 14 - 6 = 8$. Et la série {2, 6, 10, 14, 18} a pour étendue $E = 18 - 2 = 16$.

2- Les quantiles :

Définition :

On appelle quantile d'ordre $\alpha\%$, on dit aussi fractiles, et qu'on note Q_α, la valeur x_i du caractère telle que $\alpha\%$ des valeurs observées soient inférieures strictement à χ_i.

Autrement dit un quantile d'ordre n est une série ayant n groupes d'effectifs égaux. Ainsi la médiane est un quantile d'ordre 2.

Si F désigne la fonction « fréquence cumulé croissante », alors :

$$F(Q_\alpha) = \frac{\alpha}{100}.$$

Les quantiles les plus fréquemment utilisées dans la statistique descriptive sont :

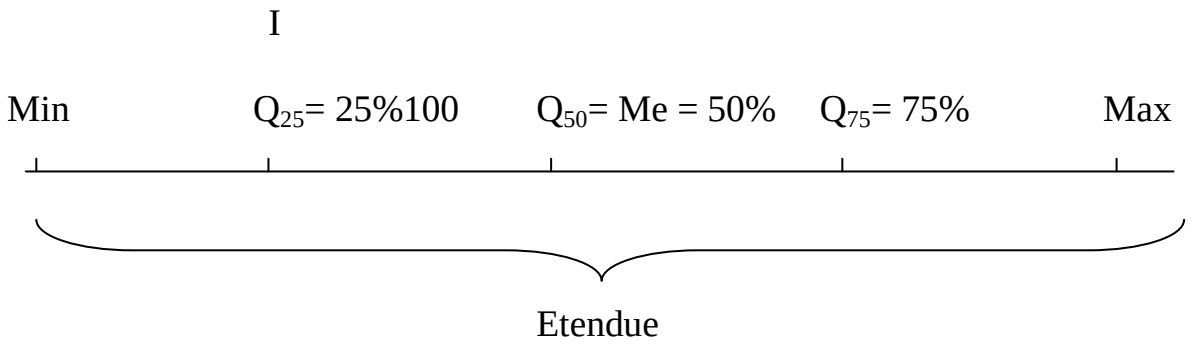
-La médiane Me : est le quantile d'ordre 50%. Elle partage la série des valeurs observées en deux séries de même taille.

-Les quartiles, c'est-à-dire les quantiles d'ordre 4, notée Q_1, Q_2, Q_3, elle partage la série en 4 séries de même taille :

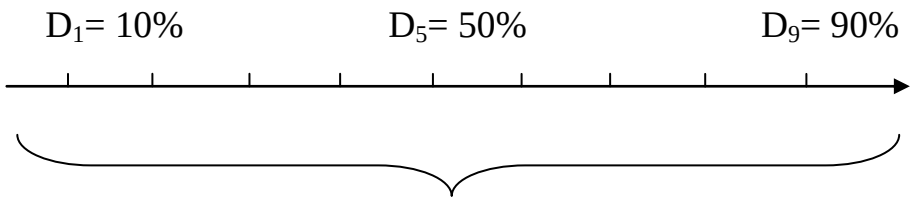
25% des observations sont inférieures au 1er quartile Q_{25} :

50% des observations sont inférieures au 2ème quartile Q_{50} :

75% des observations sont inférieures au 3ème quartile Q_{75} :



-Les déciles ; ce sont les quartiles d'ordre 10, ils partagent l'effectif total en dix groupes égaux. Il y en a neuf déciles notés D_1, D_2, D_3……..D_9.



-Les centiles ; ce sont les quartiles d'ordre 1, ils partagent l'effectif total en quatre-vingt-dix groupes égaux. Elles sont notées C_1, C_2, C_3……..C_{99}.

3-Les intervalles interquantiles :

Un intervalle interquantile d'ordre α contient α% des valeurs observées centrales, en laissant $\frac{(100 - \alpha)}{2}$ % des plus petits et $\frac{(100 - \alpha)}{2}$ % des plus grands. On note :

$I_\alpha = \left[\frac{Q_{100} - \alpha}{2}, \frac{Q_{100} + \alpha}{2} \right]$. ou $I.I_Q$, ou simplement I_Q pour designer un intervalle interquartile.

Ainsi, un intervalle interquartile est la différence entre Q_{75} et Q_{25}, et on a : $I_Q = Q_{75} - Q_{25}$, cet intervalle contient 50% des observations centrales.

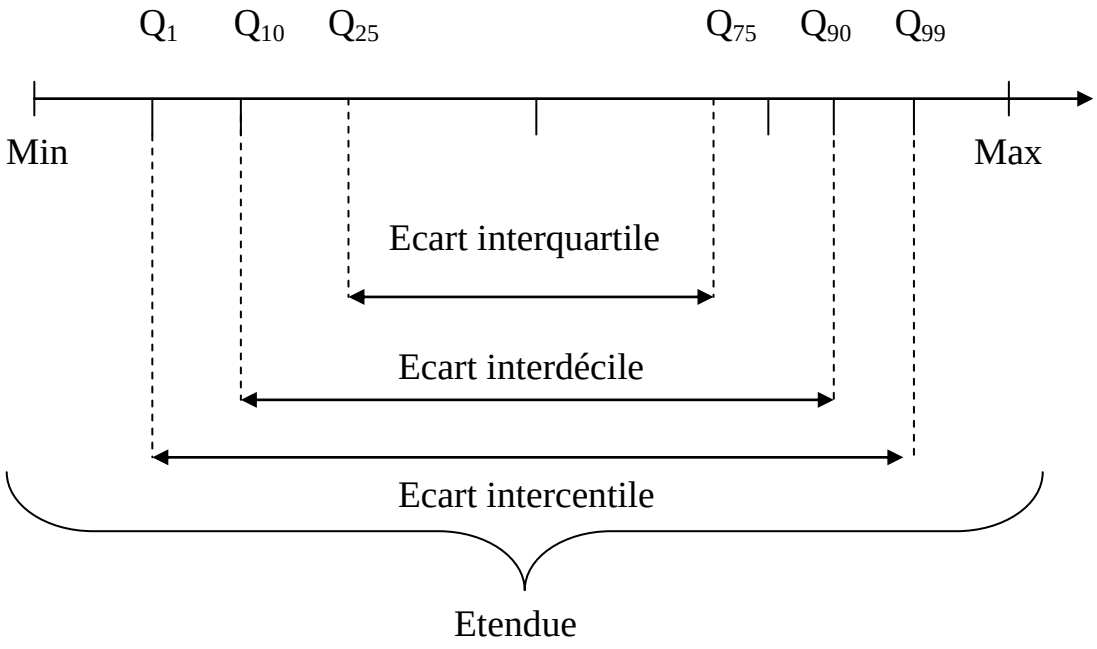
La différence $Q_{75} - Q_{25}$ s'appelle écart interquartile.

L'intervalle interdécile noté I_D ou I.Id est la différence entre Q_{90} et Q_{10}, et on a : $I_D = Q_{90} - Q_{10}$, cet intervalle contient 80% des observations.

La différence $Q_{90} - Q_{10}$ s'appelle écart interdécile.

L'intervalle interdécile noté I_C ou $I.I_C$ est la différence entre Q_{99} et Q_1, et on a : $I_C = Q_{99} - Q_1$, cet intervalle contient 98% des observations.

La différence $Q_{99} - Q_1$ s'appelle écart intercentile.



Tous ces écarts permettent de mesurer la dispersion de la série autour de la médiane.

Exemple :

On donne la distribution des revenus annuels des 460 cadres d'un groupe bancaire.

Classe de revenus (en dirhams)	Effectifs
de 50 000 à moins de 70 000	24
de 70 000 à moins de 80 000	32
de 80 000 à moins de 90 000	51
de 90 000 à moins de 100 000	70
de 100 000 à moins de 105 000	47
de 105 000 à moins de 110 000	41
de 110 000 à moins de 120 000	70
de 120 000 à moins de 130 000	58
de 130 000 à moins de 140 000	40
de 140 000 à moins de 160 000	24
de 160 000 à moins de 190 000	3

Déterminer l'étendre de la série, et les intervalles interquartile et interdécile.

En effet,

Classe de revenus (en milliers de dirhams)	Effectifs	Effectifs cumulés croissants 'ECC'
[50 , 70 [	24	24
[70 , 80 [	32	56
[80 , 90 [	51	107
[90 , 100 [	70	177
[100 , 105 [	47	224
[105 , 110 [	41	265
[110 , 120 [	70	335
[120 , 130 [	58	393
[130 , 140 [	40	433
[140 , 160 [	24	457
[160 , 190 [	3	460

L'étendue de la série est E = 190 – 50 = 140 milliers de dirhams.

Le quartile Q_1 correspond au caractère du 460/4 = 115^{ième} individu. Il appartient à la classe 90 – 100 :

$$Q_1 = 90 + 8 \times \frac{10}{70} = 91{,}14 \text{ milliers de dirhams.}$$

Le quartile Q_3 correspond au caractère du (3/4) x 460 = 345^{ième} individu. Il appartient à la classe 110 – 120 :

$$Q_3 = 120 + 10 \times \frac{10}{58} = 121{,}72 \text{ milliers de dirhams.}$$

L'intervalle interquartile est $I_Q = Q_{75} - Q_{25} = 121,72 - 95,6 = 26,12$.

Soit 26 120 DH.

Le décile D_1 correspond au caractère du $460/10 = 46^{ième}$ individu. Il appartient à la classe 70 – 80 :

$$D_1 = 70 + 10 \times \frac{22}{32} = 76,875.$$

De même Le décile D_9 correspond au caractère du $9 \times 460/10 = 414^{ième}$ individu. Il appartient à la classe 70 – 80 :

$$D_9 = 130 + 10 \times \frac{21}{40} = 135,525.$$

L'intervalle interdécile est $I_d = D_9 - D_1 = 135,525 - 76,975 = 58,55$

Soit 58 550 DH.

4- L'écart absolu moyen :

4-1- Ecart absolu moyen par rapport à la moyenne :

L'écart absolu moyen par rapport à la moyenne noté e_χ, encore appelé écart arithmétique, est la moyenne arithmétique des valeurs absolues des écarts de tous les termes de la série à leur moyenne arithmétique.

Si nous disposons d'une série de k observations distinctes χ_1, χ_2, ...χ_k affectés des coefficients n_1, n_2, …n_k, et dont la moyenne est $\bar{\chi}$, avec

$n = \sum_{i=1}^{i=k} n_i$ (n est l'effectif total).

L'écart absolu moyenne s'écrit alors $e_{\bar{x}} = \frac{\sum_{i=1}^{i=k} n_i \left| x_i - \bar{x} \right|}{\sum_{i=1}^{i=k} n_i} = \sum_{i=1}^{i=k} f_i \left| x_i - \bar{x} \right|$.

Remarquons que cet écart absolu moyen est nul si les observations sont égales.

4-2- Ecart absolu moyen par rapport à la médiane :

L'écart absolu moyen par rapport à la médiane noté e_{Me}, est la moyenne arithmétique des valeurs absolues des écarts de tous les termes de la série à leur médiane.

$$e_{Me} = \frac{\sum_{i=1}^{i=k} n_i |x_i - Me|}{\sum_{i=1}^{i=k} n_i} = \sum_{i=1}^{i=k} f_i |x_i - Me| .$$

C'est le plus petit des écarts absolus moyens.

Exemple :

On considère la distribution suivante :

Classes	[15,25[	[25,35[	[35,45[	[45,55[	[55,65[	[65,75[	[75,85[
Effectifs	9	15	22	29	17	6	2

a) calculer l'écart absolu moyen par rapport à la moyenne.
b) calculer l'écart absolu moyen par rapport à la médiane.

En effet,

Classes	x_i	n_i	$n_i x_i$	$\chi_i - \bar{\chi}$	$n_i \lvert\chi_i - \bar{\chi}\rvert$	F_i(%)	$\lvert\chi_i - M_e\rvert$	$n_i \lvert\chi_i - M_e\rvert$
[15,25[	20	9	180	25,6	230,4	9	26,38	237,42
[25,35[	30	15	450	15,6	234,0	24	16,38	245,70
[35,45[	40	22	880	5,6	123,2	46	6,38	140,36
[45,55[	50	29	1 450	4,4	127,6	75	3,62	104,98
[55,65[	60	17	1 020	14,4	244,8	92	13,62	231,54
[65,75[	70	6	420	24,4	146,4	98	23,62	141,72
[75,85[	80	2	160	34,4	68,8	100	33,62	67,24
$\sum$		100	4 560		1 175,2			1 168,96
$\sum/n$			45,60		1 175,2			1 168,96

a) $\bar{\chi} = 45,6$ et $e_{\bar{\chi}} = 11,75$

b) $Me = 45 + 10.(50 - 46)/29 \approx 46,38$ et $e_{Me} \approx 11,69$.

5- Variance et écart-type

Les paramètres de dispersion sont des nombres qui mesurent la dispersion des valeurs autour d'un paramètre de position (χ, Me, …). Ils s'expriment dans la même unité que les observations et permettent de comparer des séries statistiques de même nature.

5-1- Variance :

Variance d'une série non pondérée

Soit une série de n observation $\chi_1, \chi_2, ...\chi_p$ de moyenne arithmétique $\overline{\chi}$.
Si X est une variable discrète, la variance est le nombre réel positif :

$$Var(X) = \frac{1}{n}\sum_{i=1}^{i=p}(x_i - \overline{x})^2 .$$

Variance d'une série pondérée

$$Var(X) = \frac{1}{n}\sum_{i=1}^{i=p} n_i(x_i - \overline{x})^2 = \sum_{i=1}^{i=p} f_i(x_i - \overline{x})^2$$

Si X est une variable continue, $Var(X) = \sum_{i=1}^{i=p} f_i(c_i - \overline{x})^2$.

où c_i est le centre de la classe $\left(c_i = \frac{\alpha_i + \alpha_{i+1}}{2}\right)$ et α_i, α_{i+1} sont les bornes de l'intervalle $[\alpha_i, \alpha_{i+1}[$.

5-2- Ecart- type :

On appelle écart-type la racine carrée de la variance. On note

$$\sigma_X = \sqrt{Var(X)}.$$

Écart-type d'une série non pondérée,

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{i=n}(x_i - \overline{x})^2}{n}} .$$

Ecart-type d'une série pondérée

Le calcul de l'écart-type pour les séries pondérées n'est pas fondamentalement différent de celui effectué précédemment : les écarts à la moyenne doivent être comptés autant de fois qu'ils apparaissent dans la série.

Si la série est à caractère continu, comme pour le calcul de moyenne, on prend pour valeur du caractère le centre de chaque classe.

Soit k observations distinctes $\chi_1, \chi_2, ...\chi_k$, d'effectifs respectifs $n_1, n_2, ...n_p$ et de moyenne $\overline{\chi}$:

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{i=k} n_i (x_i - \bar{x})^2}{\sum_{i=1}^{i=k} n_i}}.$$

Théorème de Kœnigs

On a $Var(X) = \frac{\sum_{i=1}^{i=p} n_i (x_i - \bar{x})^2}{\sum_{i=1}^{i=p} n_i}$ peut aussi s'écrire:

$$Var(X) = \frac{\sum_{i=1}^{i=p} n_i x_i^2}{\sum_{i=1}^{i=p} n_i} - \bar{x}^2 = \sum_{i=1}^{i=p} f_i x_i^2 - \bar{x}^2 .$$

6- Coefficient de variation :

Pour faciliter les comparaisons entre séries, on utilise une mesure de dispersion relative appelée coefficient de variation.

Le coefficient de variation noté Cv est un nombre sans dimension et indépendant des unités choisies. On peut le considérer comme un indice de dispersion « absolu ». Le coefficient de variation est le rapport de l'écart-type à la moyenne arithmétique $\bar{\chi}$:

$$Cv = \frac{\sigma}{\chi}.$$

Exemple :

Soit à calculer l'écart-type et le coefficient de variation des observations suivantes :

χ_i	$\chi_i - \bar{\chi}$	$(\chi_i - \bar{\chi})^2$
3	-6	36
5	-4	16
8	-1	1
11	2	4
12	3	9
15	6	36
$\sum = 54$		$\sum = 102$

La série n'est pas pondérée, on utilise donc les formules de l'écart-type et de la moyenne arithmétique suivantes :

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{i=6}(x_i - \bar{x})^2}{n}}, \bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n}$$

Soient donc $\bar{x} = \frac{54}{6} = 9, \sigma_X = \sqrt{\frac{102}{6}} = 4{,}12$

Le coefficient de variation est $C_V = \frac{4{,}12}{9} = 0{,}45 = 45\%.$

Remarque :

Dans certaines cas, il est préférable d'utiliser la formule de Kœnigs pour le calcul de l'écart-type, elle s'apprête plus commode.

$$Var(X) = \frac{\sum_{i=1}^{i=p} n_i x_i^2}{\sum_{i=1}^{i=p} n_i} - \bar{x}^2, \text{ et } \sigma_X = \sqrt{\frac{\sum_{i=1}^{i=p} x_i^2}{n} - \bar{x}^2}.$$

Exemple :

Cas d'une série non pondérée

Soit 4 observations, telles que $\chi_1 = 4$, $\chi_2 = 7$, $\chi_3 = 11$, $\chi_4 = 12$.

$$\bar{\chi} = \frac{4 + 7 + 11 + 12}{4} = 8{,}5.$$

χ_i	χ_i^2
4	16
7	49
11	121
12	144
$\sum$=	330

$\sigma_\chi = \sqrt{10{,}25} = 3{,}2.$

Remarque :

Pour toute observation statistique, on a :

$e_{Me} \leq e_\chi \leq \sigma_\chi$