Paramètre de position et de dispersion

Sommaire

Objectif : On veut obtenir des paramètres pour condenser l'information contenue dans la série statistique étudiée.

- Paramètres de position : donnent un ordre de grandeur des mesures et l'existence de valeurs centrales autour desquelles se groupent les mesures.
- Paramètre de dispersion : donnent la dispersion des valeurs de la série autour de la valeur centrale.

I. Paramètre de Position

On considère une série statistique prenant n valeurs $x_1, x_2, \dots, x_n$

1. Moyenne arithmétique

DEFINITION :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- *Série à caractère discret :*

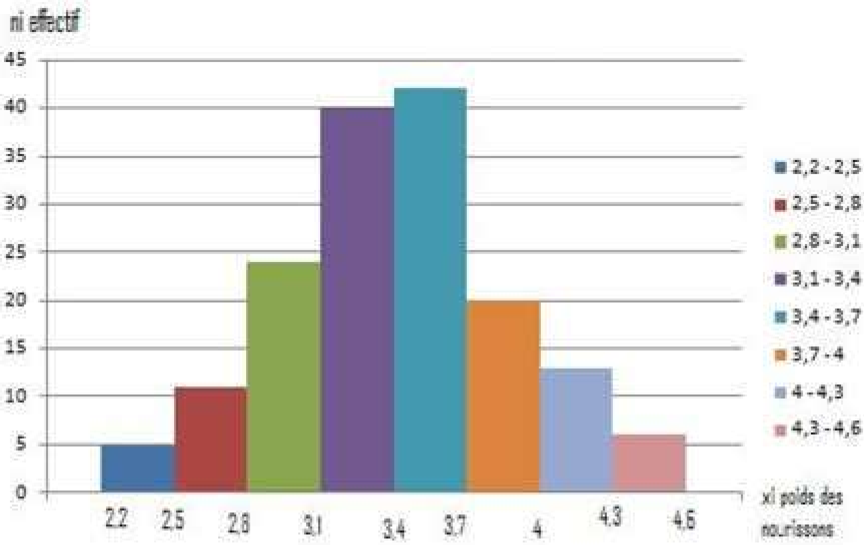
On va avoir p valeurs $x_1, x_2, \dots x_p$ est l'effectif de x_i

(f_i sera la fréquence relative de x_i)

alors $\bar{x} = \frac{n_1x_1+n_2x_2+\cdots+n_px_p}{n} = \frac{1}{n}\sum_{i=1}^{p} n_ix_i$

ou encore $\bar{x} = \sum_{i=1}^{p} x_i = \sum_{i=1}^{p} f_ix_i$

d'où l'histogramme :



On veut calculer la moyenne arithmétique

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{x_1 + x_2 + \ldots + x_{100}}{100}$$

$= \frac{0+2+3+2+4+2+3+\cdots}{100}$ $\{100\ valeurs$ ➔ assez long

$$\frac{n_1x_1 + n_2x_2 + \ldots n_px_p}{100} = \frac{n_1x_1 + n_2x_2 + n_3x_3 + n_4x_4 + n_5x_5}{100}$$

$\frac{7*0+20*1+43*2+25*3+5*4}{100} = \frac{1}{100}(20 + 86 + 75 + 20) = \frac{201}{100} = 2{,}01$

REMARQUE : On parle de moyenne arithmétique pondérée par les fréquences relative f_i par la relation

$$\bar{x} = \sum_{i=1}^{P} f_i x_i$$

2. Série à caractère continu

On a une série statistique avec p classes de centre x_i et d'effectif n_i.

Alors : $\bar{x} = \frac{n_1x_1+n_2x_2+\ldots+n_px_p}{n}$ avec x_p le centre des classes

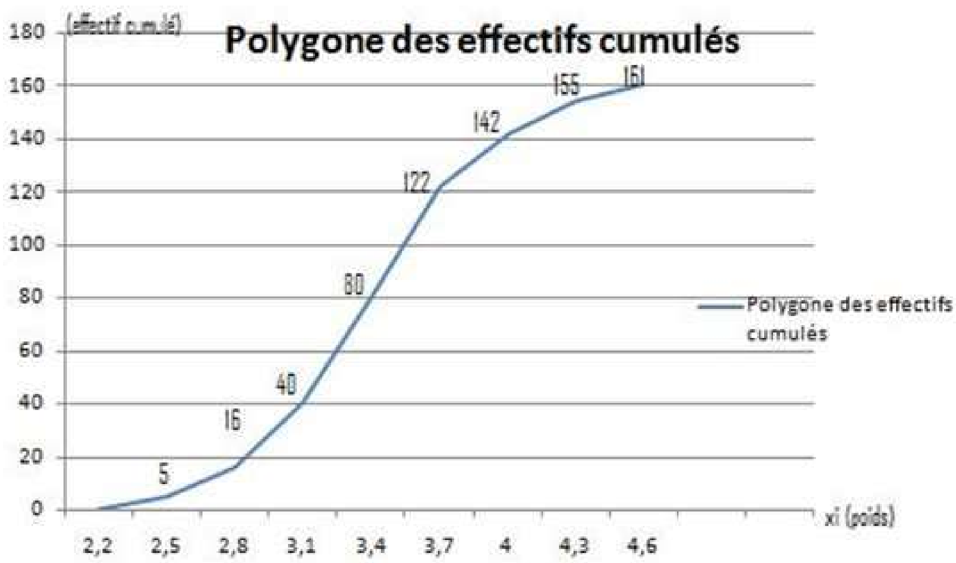
ou encore :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i = \sum_{i=1}^{p} f_i x_i$$

Ligne brisée joignant les points (x_i, n_i) avec x_i centre des classes

3. Polygône des effectifs cumulés :

Exemple des nourrissons :



Exemple des nourrissons :

Classes	1	2	3	4	5	6	7	8
Centre	2,35	2,65	2,95	3,25	3,55	3,85	4,15	4,45
Effectif <- ni	5	11	24	40	42	20	13	6
$f_i(\%)$	3,1	6,8	14,9	24,8	26,1	12,4	8,1	3,7

$\boxed{\frac{n_i}{n}}$

$\boxed{\rightarrow n=161}$

D'où $\bar{x} = \sum_{i=1}^{8} f_i x_i = \frac{3,1}{100} * 2,35 + \frac{6,8}{100} * 2,65 + \cdots = 3,406\ kg$

x_i est le centre des classes

REMARQUE : Calcul plus rapide

Sur l'exemple des nourrissons.

On va utiliser la moyenne sur le ''n° de la classe''

On remarque que : $x_i = M + az_i$

Avec M le poids des nourrissons

Dans cet exemple :

$$M = 2,05kg$$

$$a = 0,3kg$$

ou $x_1 = 2,05 + (0,3 * 1) = 2,35\ kg$

$x_2 = 2,05 + 0,3 * z_2 = 2,05 + 0,3 * 2 = 2,65$

…

alors :

$$\bar{x} = \sum_{i=1}^{8} n_i x_i = \frac{1}{n}\sum_i n_i(M + az_i)$$

$$= \frac{1}{n}(\sum_i n_i M) + \frac{1}{n}(\sum_i n_i a z_i) = \frac{M}{n}\sum_i n_i + \frac{a}{n}\sum_i n_i z_i$$

Exemple : famille de 4 enfants, x_i = nombre de garçons

x_i	0	1	2	3	4
$effectif\ ni$	7	20	43	25	5
$f.relative\ fi$	7%	20%	43%	25%	5%

$= \frac{n_i}{n}$

Série statistique étudiée : 100 familles

D'où $\bar{x} = 2,05 + 0,3 * \frac{728}{161} = 3,406$

4. Mediane

On va ordonner les valeurs de la série statistique et on va prendre la valeur du milieu.
On va ainsi trouver la médiane notée M_e

SERIES A CARACTERE DISCRET :

Exemples : série 3,3,4,5,7,7,9

Mediane $M_e = 5,5$

Série : 3,3,4,$\boxed{5,6}$,7,7,9

$M_e = \frac{5+6}{2} = 5,5$

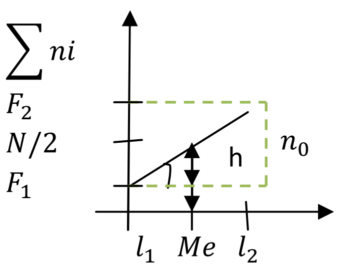
SERIES A CARACTERE CONTINU :

On va considérer que la série est déjà ordonnée par classes et la médiane M_e va tomber dans une certaine classe $[l_1, l_2[$.

Cette classe correspond à la moitie de l'effectif total (n).

On suppose les n_0 valeurs de la classe $[l_1, l_2]$ uniformément réparties dans cette classe.

Si on reprend le diagramme cumulé :



L'angle s'appelle α.

On veut calculer M_e =?

- détermination graphique par rapport au diagramme cumulatif

- On a $\tan \alpha = \frac{h}{M_e - l_1} = \frac{n_0}{l_2 - l_1}$

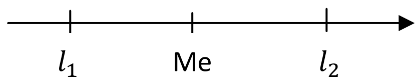
Or $\frac{n}{2} = F_1 + h = F_1 + \frac{F_2 - F_1}{l_2 - l_1}(M_e - l_1)$

D'où $\frac{n}{2} - F_1 = \frac{F_2 - F_1}{l_2 - l_1}(M_e - l_1)$

$\Leftrightarrow M_e - l_1 = \left(\frac{n}{2} - F_1\right)\frac{l_2 - l_1}{F_2 - F_1}$

$\Leftrightarrow$

$$M_e = l_1 + \left(\frac{n}{2} - F_1\right)\frac{l_2 - l_1}{F_2 - F_1}$$



où $\begin{cases} F_1 \text{ effectif cumulé en } l_1 \\ F_2 \text{ effectif cumulé en } l_2 \\ n \text{ effectif total} \\ n_0 = F_2 - F_1 = \text{effectif de la classe} \end{cases}$

<u>Exemple</u> (nourrissons)

Classes	2,2 à 2,5	2,5 à 2,8	2,8 à 3,1	3,1 à 3,4	3,4 à 3,7	3,7 à 4,0	4,0 à 4,3	4,3 à 4,6	Total
n_i	5	11	24	40	42	20	13	6	161
$\sum n_i$	5	16	40	80	122	142	155	161	

Dans quelle classe se trouve la médiane ?

Classe 3,4 à 3,7 où se situe Me

Si $n = 161 \Rightarrow \frac{n}{2} = 80{,}5$

Donc on a : $l_1 = 3{,}4$ $l_2 = 3{,}7$

$$F_1 = 80$$
$$F_2 = 122$$

$$\frac{n}{2} = 80{,}5$$

D'où $Me = 3{,}4 + (80{,}5 - 80) * \frac{3{,}7-3{,}4}{122-80}$

$$= 3{,}4 + 0{,}5 * \frac{0{,}3}{42} = 3{,}403\ kg$$

$$\boxed{Me = 3{,}403\ kg}$$

<u>Remarque :</u> Graphiquement sur le polygone des fréquences relatives cumulées, la médiane correspond à l'abscisse qui a ½ en ordonnée.

QUARTILES

- La médiane sépare la série en deux groupes de même effectif.
- Les quartiles séparent la série en 4 groupes de même effectifs.
- Cas des séries à caractères discret :

<u>Exemples :</u> 4, 5, 6, 11, 13, 14, 16

Q_1 Me Q_2

A gauche de Q_1: $\frac{7}{4} = 1{,}75$ termes

On prend <u>1.</u>

A droite de Q_3, on a aussi un seul terme

$= Q_2$

<u>Exemple</u> : 4,5,6,7,11,1: , , ,

Q_1 $Me = Q_2$ Q_3

A gauche de Q : $\frac{9}{4}$ termes = 2,25

Donc on va prendre 2 termes

Pour Q_3, on prend 2 termes au dessus.

Cas des séries à caractère continu :

On peut se servir du polygone des fréquences relatives cumulées.
Pour la médiane : On regarde la valeur à 0,5

Pour les quartiles : On regarde à 0,25 et 0,75

Ou on peut utiliser les formules :

$$Q_1 = l_1 + \left(\frac{n}{4} - F_1\right)\frac{l_2 - l_1}{F_2 - F_1}$$

Ici la classe $[l_1, l_2[$ est celle où se situe le premier quantile.

Pour Q_3 on va prendre $\frac{3n}{4}$

<u>Remarque :</u> on peut définir aussi des déciles (10) et dans quantiles/centiles (100).

5. Mode ou valeur dominante

Le mode est la valeur la plus fréquente (effectif le plus important).

- <u>Cas des séries à caractère discret</u> :

<u>Exemple</u> : 2,2,5,7,9,9,9,10,10,11

Valeur dominante ou mode.

Notation : M_o

<u>Remarque :</u> on peut avaoir plusieurs modes possibles

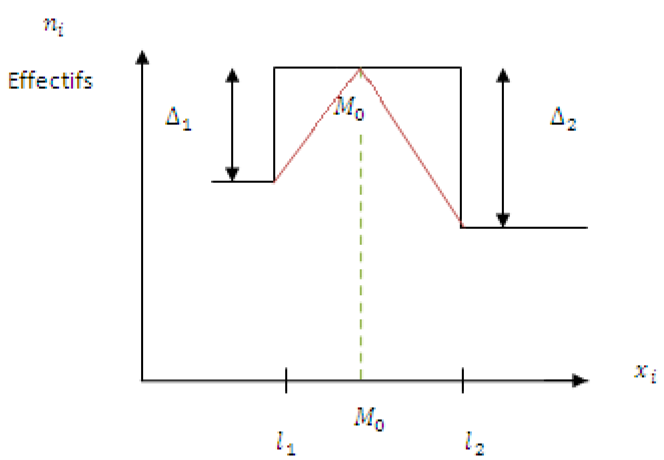
<u>Cas des séries à caractère continue :</u>

On va parler de classe modale ou dominante. Il s'agit de la classe avec l'effectif le plus important.

On veut calculer la valeur du mode.

Si on trace l'histogramme et on représente.

La classe modale (classe où se trouve le mode).



Le mode est défini de la façon suivante : on considère que les deux segments de droite en rouge ont la même pente en valeur absolue. Si la pente est identique, alors : $\left|\frac{\Delta_1}{M_0-l_1}\right| = \left|\frac{\Delta_2}{l_2-M_0}\right|$

On va calculer $M_o$en fonction de l_1, l_2, Δ_1 et Δ_2.

A partir de la relation précédente, on a :

$$\Delta_1(l_2 - M_0) = \Delta_2(M_0 - l_1)$$

$$\Leftrightarrow \Delta_1 l_2 - \Delta_1 M_0 = \Delta_2 M_0 - \Delta_2 l_1$$

$\Leftrightarrow \Delta_1 l_2 + \Delta_2 l_1 = M_o(\Delta_1 + \Delta_2)$

$\Leftrightarrow M_o = \frac{\Delta_1 l_2 + \Delta_2 l_1}{\Delta_1 + \Delta_2}$

$\Leftrightarrow M_o = \frac{(\Delta_1 l_2 + \Delta_2 l_1 + \Delta_1 l_1 - \Delta_1 l_1)}{\Delta_1 + \Delta_2} = \frac{l_1(\Delta_1 + \Delta_2)}{\Delta_1 + \Delta_2} + \frac{\Delta_1 l_2 - \Delta_1 l_1}{\Delta_1 + \Delta_2}$

$Mo = l_1 + \dfrac{\Delta_1}{\Delta_1 + \Delta_2}(l_2 - l_1)$

D'où $M_0 = l_1 + \frac{\Delta 1}{\Delta 1 + \Delta 2}\ (l_2 - l_1)$

Avec $[l_1, l_2[$ classe modale ou dominante

$\Delta 1 =$ excédent d' effectif de la classe modale par rapport à la classe inférieur

$\Delta_2 =$ excédent d''effectif de la classe modale par rapport à la classe supérieure.

<u>Exemple</u> (nourrissons)

Classe modale : $[3{,}4 - 3{,}7[, n_5 = 42$

$$\Delta_1 = 42 - 40 = 2$$

$$\Delta_2 = 42 - 20 = 22$$

D'où $M_0 = 3{,}4 + \frac{2}{2+22}(3{,}7 - 3{,}4)$

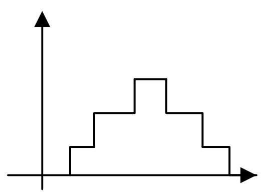
$$3{,}4 + \frac{2}{24} * 0{,}3$$

$$= 3{,}425$$

<u>Remarque</u> :

- <u>Série unimodale</u> :

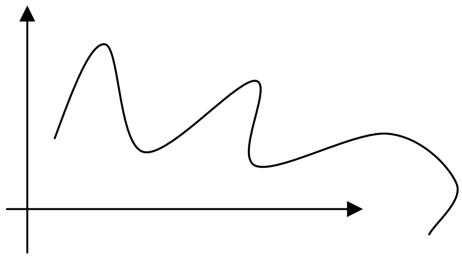
Histogramme :



Une seule classe dominante

donc série uni modale.

Série plurimodale :



Si série symétrique :

On peut montrer que $\overline{x} = M_e = M_o$

II. Paramètres de dispertion

1. Etendue

Etendue : écart entre la plus grande valeur $x_{\max}$ et la plus petite valeur $x_{\min}$ de la série.

Etendue $= x_{\max} - x_{\min}$

2. Ecart moyen

On considère une série statistique de valeurs x_i , i variant de 1 à n (n assez grand) , de moyenne $\bar{x}$

Ecart de x_i : $e_i = |x_i - \bar{x}|$

- <u>Ecart moyen de la série</u> :

C'est la moyenne arithmétique des écarts.

On a $\bar{e} = \frac{e_1+e_2+e_3+\ldots+e_n}{n} = \frac{1}{n}\sum_{i=1}^{n} e_i$

<u>cas discret :</u>

Valeurs x_i, i de 1 à p(p petit), effectif n_i

Fréquence relative, $f_i = \frac{n_i}{n}$

<u>cas continu :</u> classes, x_i centre des classes, i variant de 1 à p (p=nombre de classe)

Alors on peut écrire : $\bar{e} = \frac{1}{n}\sum_{i=1}^{p} n_i e_i = \sum_{i=1}^{p} f_i e_i$

<u>Exemple :</u> nombre de garçon dans les familles de 4 enfants

On a vu que $\bar{x} = 2{,}01 \approx 2$

Alors $\bar{e} = \frac{7}{100}|0-2| + \frac{20}{100}|1-2|` + \frac{43}{100}|2-2| + \frac{25}{100}|3-2| + \frac{5}{100}|4-2|$

D'où $\bar{e} = \frac{14+20+25+10}{100} = 0{,}69$

3. Variance et écart type

- La variance est la moyenne arithmétique des carrés des écarts des x_i

Variance notée : ${\sigma_x}^2$

Donc $\boxed{{\sigma_x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

- Ecart type ou écart quadratique, noté σ_x , est la racine carrée de la variance.
  <u>Remarque</u> :
  - Cas discret : x_i(i variant de 1 à p), f_i
  - Cas continu : x_i(centre), f_i

alors

$$\boxed{\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{p} n_i(x_i-\bar{x})^2 = \sum_{i=1}^{p} f_i(x_i-\bar{x})^2}$$

Formule plus simple de calcul :

$$\boxed{{\sigma_x}^2 = \frac{1}{n}\sum_{i=1}^{p} n_i{x_i}^2 - (\bar{x})^2}$$

Démonstration

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \sum_{i=1}^{n}\left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) = \sum_{i=1}^{n}{x_i}^2 - \sum_{i=1}^{n}2\bar{x}x_i + \sum_{i=1}^{n}\bar{x}^2 = \sum_{i=1}^{n}{x_i}^2 - 2\bar{x}\sum_{i=1}^{n}x_i + \bar{x}^2\sum_{i=1}^{n}1$$

Donc $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}{x_i}^2 - 2\bar{x} * n\bar{x} + n\bar{x}^2$

On a $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i \Rightarrow \sum_{i=1}^{n}x_i = n\,\bar{x}$

D'où $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}{x_i}^2 - 2n(\bar{x})^2 + n(\bar{x})^2$

$= \sum_{i=1}^{n}{x_i}^2 - n(\bar{x})^2$

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

Ou encore $\boxed{{\sigma_x}^2 = \sum_{i=1}^{p} f_i{x_i}^2 - \bar{x}^2}$

- Méthode de calcul pour la variance

1. On va calculer la moyenne $\bar{x}$
2. On va calculer ${\sigma_x}^2$ par la formule la plus simple.

   <u>Exemple :</u> on avait $\bar{x}$=2,01

   d'où $\sigma_x^2 = \frac{7}{100} * 0^2 + \frac{20}{100} * 1^2 + \frac{43}{100} * 2^2 + \frac{25}{100} * 3^2 + \frac{5}{100} * 4^2 - (2{,}01)^2$

   $\sigma_x^2 = \frac{20 + 172 + 225 + 80}{100} - 4{,}04$

   $= 4{,}97 - 4{,}04 = \boxed{0{,}93} \rightarrow \tau_x = \sqrt{0{,}93} = \boxed{0{,}96}$

Propriété de la variance

Si $x_i = M + az_i$ avec z_i le numéro de la classe
(Valable pour les séries à caractère continu)
Alors $\boxed{\sigma_x = a\sigma Z}$

Démonstration

$$\bar{x} = M + a\bar{z}$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(\cancel{M} + az_i - \cancel{M} - a\bar{z})^2 = \sum_{i=1}^{n}(az_i - a\bar{z})^2$$

$$= a^2 \sum_{i=1}^{n}(z_i - \bar{z})^2$$

D'où $\frac{1}{n} \sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{a^2}{n} \sum_{i=1}^{n}(z_i - \bar{z})^2$

$$\sigma_X{}^2 = a^2\sigma_Z^2 \Rightarrow \sigma_X = a\,\sigma_Z$$

Exemple :

8 classes, centre 2,35 ; 2,65 ; …. $\begin{cases} M = 2{,}05 \\ a = 0{,}3 \end{cases}$

$$\bar{x} = M + a\bar{z}$$

On a vu que $\bar{x} = 3{,}406$ (cf cours précédent)

Avec x $\bar{z} = 4{,}522$

On veut calculer σ_x ?
On a vu que $\sigma_x = a\,\sigma_z$
Or $\sigma_z^2 = \sum_{i=1}^{8} f_i z_i^2 - (\bar{z})^2$ <- Formule pratique pour calculer la variance
D'où $\sigma_Z{}^2 = \frac{5*1^2+11*2^2+\cdots+6*8^2}{161} - (4{,}522)^2 = \frac{3696}{161} - (4{,}522)^2 = 22{,}956 - 20{,}448 = 2{,}508$

D'où $\sigma_Z = \sqrt{2{,}508} = 1{,}583$

Donc $\sigma_x = 0{,}3 * 1{,}583 = 0{,}475$

4. Ecart interquartile

DEFINITION : L'écart interquartile est représenté par $Q_3 - Q_1$ et il contient 50% des valeurs de la série

5. Coefficient de variation

$$CdV = \frac{\sigma_X}{\bar{x}}$$

Intérêt de ce coefficient de variation :

- ⇨ Valeur sans unité, cela permet donc de comparer les dispersions de série qui n'ont pas la même unié.

Remarque : ce calcul est difficile si la valeur moyenne est proche de 0.