Paramètres de dispersion

Les paramètres de dispersion précisent le degré de variation ou fluctuation des différentes valeurs d'une série autour d'une valeur centrale.

Les âges au mariage pour deux groupes d'individus sont les suivants

Groupe1	21	22	23	24	30	30	36	37	38	39
Groupe2	28	29	29	30	30	30	30	31	31	32

Les caractéristiques de tendance centrale (mode, médiane et moyenne) des deux séries sont

$$\begin{array}{ll} \underline{Groupe 1} & \underline{Groupe\ 2} \\ M_0 \ = 30 & M_0 \ = 30 \\ Me \ = 30 & Me \ = 30 \\ \bar{x} \ = 30 & \bar{x} \ = 30. \end{array}$$

Cependant les 2 séries correspondent à des observations qui se distribuent très différemment.

Dans le groupe 1, les valeurs s'écartent notoirement de la valeur 30, ce qui n'est pas le cas dans le groupe 2.

1. Etendue

L'étendue e d'une série est donnée par

$$e = X_{max} - X_{min}.$$

Dans le cas d'une série regroupée dans des classes, l'étendue est estimée par

e =Centre de la dernière classe-Centre de la première classe.

L'étendue est un paramètre de dispersion rapide à déterminer. C'est un indicateur instable de la dispersion, dépend des valeurs extrêmes.

2. Ecarts interquantiles

Pour éviter d'effectuer des calculs sur des valeurs extrêmes (souvent aberrantes), on choisit souvent de les écarter de la série.

Intervalle	Signification	Ecart
Intervalle interquartile $[Q_1, Q_3]$	contient 50% des observations	Ecart interquartile $\boldsymbol{Q_3 - Q_1}$
Intervalle interdécile $[D_1, D_9]$	contient 80 des observations	Ecart interdécile $\boldsymbol{D_9 - D_1}$
Intervalle intercentile $[C_1, C_{99}]$	contient 98% des observations	Ecart intercentile $\boldsymbol{C_{99} - C_1}$

3. Ecart absolu moyen (Ecart arithmétique)

C'est une mesure de la dispersion qui prend en compte toutes les données. Il s'agit d'une moyenne arithmétique d'écarts par rapport à une valeur centrale et non plus d'un écart entre deux valeurs.

- Ecart absolu moyen par rapport à la moyenne ($\bar{e}_{\bar{x}}$)
  - Pour une série groupée

$$\bar{e}_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{k} n_i|x_i - \bar{x}|.$$

  - Pour une série brute $X_1, \dots, X_n$,

$$\bar{e}_{\bar{x}} = \frac{1}{n}\sum |X_i - \bar{x}|.$$

- Ecart absolu moyen par rapport à la médiane ($\bar{e}_{Me}$)
  - Pour une série groupée

$$\bar{e}_{Me} = \frac{1}{n}\sum n_i|x_i - Me|.$$

  - Pour une série brute $X_1, \dots, X_n$,

$$\bar{e}_{\bar{x}} = \frac{1}{n}\sum |X_i - Me|.$$

Plus généralement, un écart absolu moyen $\bar{e}_a$ par rapport à une valeur a et donné par

$$\bar{e}_a = \frac{1}{n}\sum n_i|x_i - a|.$$

Notons que $\bar{e}_{Me}$ est le plus petit des écarts absolus.

4. Variance et écart type

L'usage des écarts absolus moyens est peu commode du fait des valeurs absolues. On leur préfère l'écart type ou son carré la variance.

- **Variance** :
  - On appelle variance de la série $X_1, \dots, X_n$ le nombre noté $V(X), Var(X)$ ou σ_X^2 défini par

$$V(X) = \frac{1}{n}\sum (X_i - \bar{x})^2.$$

  - Dans le cas d'une distribution statistique $(x_i, n_i)_{i=\overline{1,k}}$ d'un caractère discret

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

  - Dans le cas d'une distribution $([e_{i-1}, e_i[\ ,\ n_i)\ _{i=\overline{1,k}}$ d'un caractère continu

$V(X) = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2$ où x_i est le centre de la $i$ème classe.

Formule développée de la variance : Le calcul de la variance par la formule de définition est toujours plus long que par la formule dite développée

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

En effet :

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

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

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

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

$$= \frac{1}{n}\sum n_i x_i^2 - \bar{x}^2.$$

- **Ecart type** : L'écart type, noté σ_X, n'est autre que la racine de la variance

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

Propriété

Si $z_i = ax_i + b$ alors $\sigma_Z^2 = a^2\sigma_X^2$ et $\sigma_Z = |a|\sigma_X \quad (= a\sigma_X \;\; si\ a \geq 0)$.

$$\sigma_Z^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (z_i - \bar{z})^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (ax_i + b - (a\bar{x} + b))^2$$

$$= \frac{1}{n}\sum_{i=1}^{k} n_i (ax_i - a\bar{x})^2$$

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

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

$$= a^2\sigma_X^2.$$

Remarque :

L'écart type (appelé aussi écart quadratique moyen) et la variance sont des indicateurs de dispersion de même nature. L'écart type est mesuré dans la même unité que la variable. La variance est mesurée dans l'unité au carré.

Théorème de König

Pour tout $a \in \mathbb{R}$, on a

$$\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - a)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2 + (\bar{x} - a)^2 = V(X) + (\bar{x} - a)^2$$

ou de manière équivalente

$$V(X) = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - a)^2 - (\bar{x} - a)^2.$$

Si $a = 0$, on obtient la formule développée de $V(X)$.

Preuve

$$\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - a)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x} + \bar{x} - a)^2$$

$$= \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2 + \frac{1}{n}\sum_{i=1}^{k} n_i(\bar{x} - a)^2 + \frac{2}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})(\bar{x} - a)$$

$$= \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2 + \frac{1}{n}(\bar{x} - a)^2\sum_{i=1}^{k} n_i + \frac{2}{n}(\bar{x} - a)\sum_{i=1}^{k} n_i(x_i - \bar{x})$$

Comme $\sum_{i=1}^{k} n_i = n$ et $\sum_{i=1}^{k} n_i(x_i - \bar{x}) = 0$ alors

$$\frac{1}{n}\sum_{i=1}^{k} n_i(x_i - a)^2 = V(X) + (\bar{x} - a)^2.$$

Exemple

Soit la distribution statistique (d'une variable continue X) suivante

Classes	Effectifs n_i
[3,4[	26
[4,5[	33
[5,6[	64
[6,7[	7
[7,8[	10
	$n = 140$

Calculer la variance par la formule de définition et par la formule développée.

- **Formule de définition**

Classes	Effectifs n_i	Centre de classe x_i	$n_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$n_i(x_i - \bar{x})^2$
[3,4[	26	3.5	91	-1.58	2.496	64.896
[4,5[	33	4.5	148.5	-0.58	0.336	11.088
[5,6[	64	5.5	352	0.41	0.168	10.752
[6,7[	7	6.5	45.5	1.42	2.016	14.112
[7,8[	10	7.5	75	2.41	5.808	58.08
$\sum$	$n = 140$		712			158.928

$$\bar{x} = \frac{712}{140} = 5.0857$$

$$V(X) = \frac{158.928}{140} = 1.14$$

- **Formule développée**

Classes	Effectifs n_i	Centre de classe x_i	$n_i x_i$	$n_i {x_i}^2$
[3,4[	26	3.5	91	318.5
[4,5[	33	4.5	148.5	668.25
[5,6[	64	5.5	352	1936
[6,7[	7	6.5	45.5	295.75
[7,8[	10	7.5	75	562.5
$\sum$	$n = 140$		712	3781

$$V(X) = \frac{3781}{140} - 5.0857^2 = 1.14$$

5. Coefficient de variation (dispersion relative)

Il permet de comparer la dispersion de plusieurs distributions. Il est indépendant de l'unité de mesure de la variable, du nombre et de l'ordre de grandeur des observations. Pour supprimer à la fois l'effet d'unité et d'ordre de grandeur, on divise l'écart type par la moyenne de la série.

Le coefficient de variation CV est donné par

$$CV = \frac{\sigma_X}{\bar{x}}.$$

.