

Chapitre 5. Statistiques descriptives : Paramètres de dispersion

Mathématiques et statistiques appliquées
Département TC1-IUT de Sceaux

Damien THOMINE

Objectifs

- Savoir calculer l'écart-type d'une série statistique.
- Interpréter les paramètres de dispersion : écart inter-quartile, écart-type.

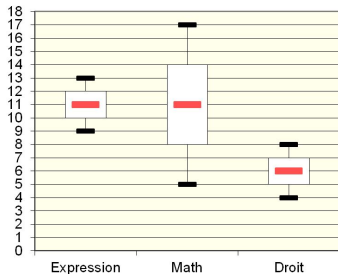
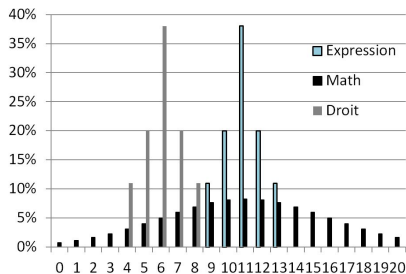
Plan du cours

- 1 Écart interquartile
- 2 Variance
- 3 Écart-type

Qu'est-ce que la dispersion ?

Les **paramètres de dispersion** mesurent l'étalement des observations autour de la valeur centrale

Exemple. Les diagrammes ci-dessous montrent les résultats à 3 examens.



- Dispersion Expression < Dispersion Math
- Dispersion Expression = Dispersion Droit

Section 1

Écart interquartile

Écart interquartile

L'**intervalle interquartile** $[Q_1, Q_3]$ couvre au moins la moitié centrale des observations.

L'**écart interquartile** $= Q_3 - Q_1$ mesure la taille de cet intervalle.

Exemple.

Expression $Q_1=10$ et $Q_3=12$:

écart interquartile $= 12 - 10 = 2$

Math $Q_1=8$ et $Q_3=14$:

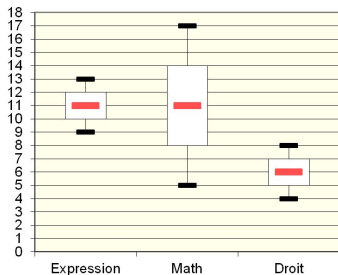
écart interquartile $= 14 - 8 = 6$

Droit $Q_1=5$ et $Q_3=7$:

écart interquartile $= 7 - 5 = 2$

Avantages : Intuitif, robuste.

Désavantage : Ne dépend pas du tout des valeurs extrêmes ; ne voit pas s'il y a quelques valeurs très éloignées.



Section 2

Variance

Variance et écart-type : introduction

La **variance** et l'**écart-type** mesurent l'écart des données par rapport à la moyenne. Ils sont les paramètres de dispersion les plus importants.

Avantages :

- Ils intègrent les informations sur toutes les données.
- Ils ont des très bonnes propriétés mathématiques.
- Ils peuvent être généralisés à l'étude de l'interaction de deux variables statistiques (cf. covariance étudiée en S2).
- On peut les estimer facilement à partir d'un échantillon (cf. S3).

Désavantages :

- Intuitivement plus difficiles à saisir.
- Assez compliqués à calculer.

La variance et l'écart-type sont liés :

$$\sigma := \mathbf{\acute{e}cart\text{-}type} = \sqrt{\mathbf{variance}} \quad \text{et} \quad \sigma^2 := \mathbf{variance} = \mathbf{\acute{e}cart\text{-}type}^2$$

Variance

La **variance** d'une série de n observation x_1, \dots, x_n de moyenne \bar{x} est :

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}.$$

C'est la *moyenne du carré des écarts à la moyenne* \bar{x} .

Exemple. Voici les âges d'un groupe de 6 enfants :

$$x_1 = 6, x_2 = 6, x_3 = 7, x_4 = 7, x_5 = 7, x_6 = 8$$

L'âge moyen est $\bar{x} = 6,8$ ans.

La variance est :

$$\begin{aligned} \sigma^2 &= \frac{(6 - 6,8)^2 + (6 - 6,8)^2 + (7 - 6,8)^2 + (7 - 6,8)^2 + (7 - 6,8)^2 + (8 - 6,8)^2}{6} \\ &= \frac{(-0,8)^2 + (-0,8)^2 + (0,2)^2 + (0,2)^2 + (0,2)^2 + (1,2)^2}{6} \\ &= \frac{0,64 + 0,64 + 0,04 + 0,04 + 0,04 + 1,44}{6} = 0,47 \end{aligned}$$

Variance à partir des effectifs

On peut calculer la moyenne à partir de modalités y_1, \dots, y_p et des effectifs n_i associés :

$$\sigma^2 = \frac{(y_1 - \bar{x})^2 \times n_1 + \dots + (y_p - \bar{x})^2 \times n_p}{n}$$

Exemple. Voici les âges d'un groupe de 6 enfants :

Âge y_i	6	7	8	Total
Effectif n_i	2	3	1	6

$$\begin{aligned} \sigma^2 &= \frac{(6 - 6,8)^2 \times 2 + (7 - 6,8)^2 \times 3 + (8 - 6,8)^2 \times 1}{6} \\ &= \frac{(-0,8)^2 \times 2 + (0,2)^2 \times 3 + (1,2)^2 \times 1}{6} \\ &= \frac{0,64 \times 2 + 0,04 \times 3 + 1,44 \times 1}{6} = 0,47 \end{aligned}$$

Test

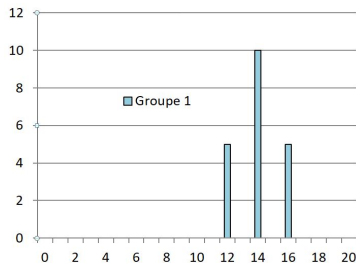
Âges du Groupe 1

Âge y_i	12	14	16	Total
Effectif n_i	5	10	5	20

● On s'attend à ce que l'âge moyen soit de 14 ans.

● Calculer la moyenne et la variance de l'âge du groupe 1

$$\begin{aligned} \bar{x} &= \frac{12 \times 5 + 14 \times 10 + 16 \times 5}{20} = 14 \\ \sigma^2 &= \frac{(12-14)^2 \times 5 + (14-14)^2 \times 10 + (16-14)^2 \times 5}{20} \\ &= \frac{(-2)^2 \times 5 + 0^2 \times 10 + 2^2 \times 5}{20} \\ &= \frac{4 \times 5 + 0 + 4 \times 5}{20} = 2 \end{aligned}$$



Variance à partir des fréquences

On peut calculer la moyenne à partir de modalités y_1, \dots, y_p et des fréquences f_1 associés.

$$\sigma^2 = (y_1 - \bar{x})^2 \times f_1 + \dots + (y_p - \bar{x})^2 \times f_p$$

Exemple. Voici les âges d'un groupe de 6 enfants :

Âge y_i	6	7	8	Total
Fréquence f_i	33,3%	50%	16,7%	100%

$$\sigma^2 = (6 - 6,8)^2 \times 0,333 + (7 - 6,8)^2 \times 0,5 + (8 - 6,8)^2 \times 0,167 \cong 0,47$$

Test

Âges du Groupe 2

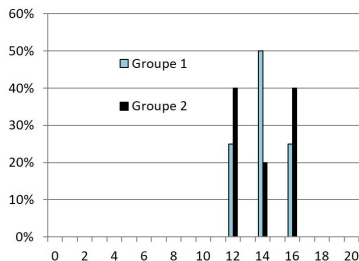
Âge y_i	12	14	16	Total
Fréquence f_i	40%	20%	40%	100%

• On s'attend à ce que la variance du groupe 2 soit plus **grande** que la variance du groupe 1.

• Calculer la moyenne et la variance de l'âge du groupe 2

$$\bar{x} = 12 \times 0,4 + 14 \times 0,2 + 16 \times 0,4 = 14$$

$$\begin{aligned} \sigma^2 &= (12 - 14)^2 \times 0,4 + (14 - 14)^2 \times 0,2 + (16 - 14)^2 \times 0,4 \\ &= 4 \times 0,4 + 0 + 4 \times 0,4 = 3,2 \end{aligned}$$



Variance : autres formules

Il existe d'**autres formules** équivalentes, parfois mieux adaptées, pour calculer la variance :

$$\sigma^2 = \frac{x_1^2 + \dots + x_n^2}{n} - \bar{x}^2 = \frac{y_1^2 \times n_1 + \dots + y_p^2 \times n_p}{n} - \bar{x}^2 = y_1^2 \times f_1 + \dots + y_p^2 \times f_p - \bar{x}^2$$

c'est-à-dire la moyenne des carrés moins le carré de la moyenne.

Exemple.

Âge y_i	6	7	8	Total
Effectif n_i	2	3	1	6
Fréquence f_i	33,3%	50%	16,7%	100%

$$\bar{x} = 6,8$$

$$\sigma^2 = \frac{6^2 + 6^2 + 7^2 + 7^2 + 7^2 + 8^2}{6} - 6,8^2 = 0,47$$

$$\sigma^2 = \frac{6^2 \times 2 + 7^2 \times 3 + 8^2 \times 1}{6} - 6,8^2 = 0,47$$

$$\sigma^2 = 6^2 \times 0,333 + 7^2 \times 0,5 + 8^2 \times 0,167 - 6,8^2 = 0,47$$

Données groupées par classes

Quand les données sont groupées par classes, on ne dispose plus des valeurs exactes. On peut remplacer les classes par leur valeur central (comme pour le calcul de la moyenne), quitte à rajouter un terme correctif.

On note :

- c_i les centres des classes ;
- A_i les amplitudes des classes ;
- f_i les fréquences des classes.

La **moyenne** est $\bar{x} = f_1 c_1 + \dots + f_p c_p$.

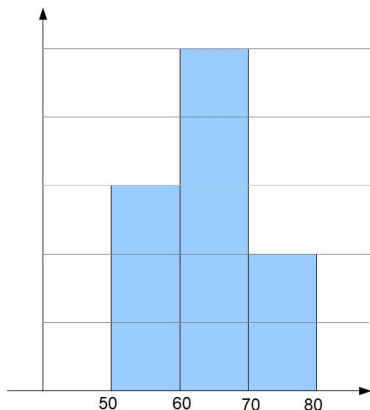
La variance est donnée par :

$$\sigma^2 = f_1(c_1 - \bar{x})^2 + \dots + f_p(c_p - \bar{x})^2 + \frac{1}{12} (f_1 \times A_1^2 + \dots + f_p \times A_p^2).$$

Test : Variance par classes

Poids en grammes des oeufs d'un élevage

Poids	[50,60[[60,70[[70,80[
Fréquence f_i	30%	50%	20%
centre c_i	55	65	75



- Calculer la moyenne et la variance du poids des oeufs

$$\bar{x} = 0,3 \times 55 + 0,5 \times 65 + 0,2 \times 75 = 64 \text{ grammes}$$

$$\sigma^2 = \frac{0,3 \times (55 - 64)^2 + 0,5 \times (65 - 64)^2 + 0,2 \times (75 - 64)^2}{12} \simeq 49 + 8 = 57 \text{ grammes}^2$$

Section 3

Écart-type

Écart-type

La variance n'est pas facilement comparable aux données ; notamment elle n'a pas la même unité de mesure. On préfère donc utiliser l'**écart-type** :

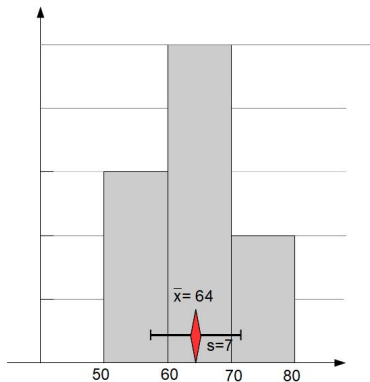
$$\sigma = \sqrt{\text{variance}} = \sqrt{\sigma^2}$$

Exemple. Poids en grammes des oeufs.

Variance : $\sigma^2 \simeq 57 \text{ grammes}^2$

Écart-type : $\sigma \simeq \sqrt{57} \simeq 7,6 \text{ grammes}$

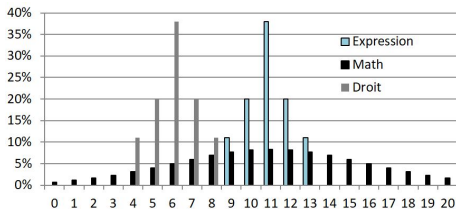
Exemple. On peut penser approximativement que le poids d'un oeuf s'éloigne typiquement de 7,6g du poids moyen de 64g.



Test

Le diagramme ci-dessous montre les résultats à 3 examens.

Compléter les (in)égalités suivantes avec ">" "<" ou "=" :



- moyenne Droit 6
- moyenne Expr 11
- moyenne Math 15
- moyenne Math 7

On calcule : moyenne Math=10,8

- écart-type Math écart-type Expr
- écart-type Droit écart-type Expr
- écart-type Expr 2
- écart-type Math 2

On a écart-type Expr=écart-type Droit=1,13 et écart-type Math=4,48

Récapitulatif des formules : variance

Données exactes

Données	formules
Non-groupées	$\sigma^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{x_1^2 + \dots + x_n^2}{n} - \bar{x}^2$
Groupées effectif	$\sigma^2 = \frac{(y_1 - \bar{x})^2 \times n_1 + \dots + (y_p - \bar{x})^2 \times n_p}{n} = \frac{y_1^2 \times n_1 + \dots + y_p^2 \times n_p}{n} - \bar{x}^2$
Groupées fréquence	$\sigma^2 = (y_1 - \bar{x})^2 \times f_1 + \dots + (y_p - \bar{x})^2 \times f_p = y_1^2 \times f_1 + \dots + y_p^2 \times f_p - \bar{x}^2$

Données groupées par classes

Appliquer une des formules précédentes, en utilisant pour y_i le centre des classes.

Si les classes sont larges, ajouter le terme correctif

$$\frac{1}{12} \left(f_1 \times A_1^2 + \dots + f_p \times A_p^2 \right)$$

Récapitulatif des formules : écart-type

Écart-type= σ

$$\sigma = \sqrt{\text{variance}} = \sqrt{\sigma^2}$$

Remarques finales

Unités

Si la variables statistique a une unité (ans, mètres, clients...), alors les paramètres de positions (moyenne, médiane, quartiles, déciles), l'écart inter-quartile et l'écart-type ont la même unité.

Intepétation

Comme on l'a vu au chapitre précédent, la moyenne et la médiane ont des interprétations légèrement différentes. La moyenne prend en compte toutes les données, mais est tirée par les valeurs extrêmes (exemple du salaire). La même différence existe entre l'écart inter-quartile et l'écart-type : l'écart-type prend en compte toutes les données, mais est tiré vers le haut s'il y a quelques valeurs extrêmes très éloignées.