Statistique Descriptive
Chapitre III : Résumés numériques des variables quantitatives

Anas KNEFATI

Université Rennes 2

UNIVERSITÉ RENNES ueb

Plan

1. Introduction
2. Mesures de tendance centrale ou de position
   - Moyenne
   - Médiane
   - Mode
   - Quartiles et déciles
3. Mesures de dispersion
   - Étendue - Écart moyen absolu - Écart médian absolu
   - Ecart inter-quartiles - Box-plot (Boite à moustaches)
   - Variance et écart-type

Plan

Mesures

- Les informations contenues dans les variables quantitatives peuvent être résumées au moyen d'indicateurs numériques.
- Type d'indicateurs

  Mesures de tendance centrale ou de position fournissent un ordre de grandeur des valeurs de la série statistique et la position où se situent ces valeurs

    - Moyenne
    - Médiane
    - Mode
    - Quartiles et déciles

Mesures

- Les informations contenues dans les variables quantitatives peuvent être résumées au moyen d'indicateurs numériques.
- Type d'indicateurs

  Mesures de tendance centrale ou de position fournissent un ordre de grandeur des valeurs de la série statistique et la position où se situent ces valeurs
    - Moyenne
    - Médiane
    - Mode
    - Quartiles et déciles

  Mesures de dispersion rendent en compte de l'éparpillement des données autour des valeurs centrales
    - Écart moyen absolu et écart médian absolu
    - Écart inter-quartile
    - Variance et écart-type

Plan

1. Introduction

2. Mesures de tendance centrale ou de position
   - Moyenne
   - Médiane
   - Mode
   - Quartiles et déciles

3. Mesures de dispersion
   - Étendue - Écart moyen absolu - Écart médian absolu
   - Ecart inter-quartiles - Box-plot (Boite à moustaches)
   - Variance et écart-type

Moyenne

Définition

- x : Variable statistique, $x_1, \cdots, x_n$ observations.
- Moyenne = La somme des valeurs divisées par leur nombre.

Moyenne

Définition

- x : Variable statistique, $x_1, \cdots, x_n$ observations.
- Moyenne = La somme des valeurs divisées par leur nombre.

Calcul

- **Données brutes** : $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Moyenne

Définition

- x : Variable statistique, $x_1, \cdots, x_n$ observations.
- Moyenne = La somme des valeurs divisées par leur nombre.

Calcul

- **Données brutes** : $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- **Données agrégées d'une variable discrète à K modalités** :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{K} n_i x_i = \sum_{i=1}^{K} f_i x_i \quad \text{où} \quad \left\{ \begin{array}{ll} n_i & \text{est l'effectif de la } i^{\text{ème}} \text{ classe,} \\ f_i & \text{est la fréquence de la même classe.} \\ n & \text{effectif total} \end{array} \right.$$

Moyenne

Définition

- x : Variable statistique, $x_1, \cdots, x_n$ observations.
- Moyenne = La somme des valeurs divisées par leur nombre.

Calcul

- **Données brutes** : $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
- **Données agrégées d'une variable discrète à K modalités** :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{K} n_i x_i = \sum_{i=1}^{K} f_i x_i \qquad \text{où} \quad \left\{ \begin{array}{ll} n_i & \text{est l'effectif de la } i^{\text{ème}} \text{ classe,} \\ f_i & \text{est la fréquence de la même classe.} \\ n & \text{effectif total} \end{array} \right.$$

- **Données agrégées d'une variable continue de K classes** :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{K} n_i c_i = \sum_{i=1}^{K} f_i c_i \qquad \text{où} \quad c_i \quad \text{est le centre de la} \quad i^{\text{ème}} \text{ classe.}$$

Moyenne - Exemples

Variable discrète : Nb de pièces des appatements d'un immeuble

Modalités x_i	Effectifs n_i	$n_i x_i$
1	10	10
2	7	14
3	12	36
4	16	64
Total	45	124

Alors $\bar{x} = \frac{124}{45} \approx 2.76$

Moyenne - Exemples

Variable discrète : Nb de pièces des appatements d'un immeuble

Modalités x_i	Effectifs n_i	$n_i x_i$
1	10	10
2	7	14
3	12	36
4	16	64
Total	45	124

Alors $\bar{x} = \frac{124}{45} \approx 2.76$

Variable continue : Notes des étudiants

Classe	Effectifs n_i	Centres c_i	$n_i c_i$	Fréquences fi	$f_i c_i$
[0; 5[	6	2.5	15	0.1	0.25
[5; 8[	21	6.5	136.5	0.36	2.34
[8; 12[	8	10	80	0.14	1.4
[12; 15[	10	13.5	135	0.17	2.3
[15; 20[	14	17.5	245	0.24	4.2
Total	59		611.5	1	10.5

$\bar{x} = \frac{611.5}{59} \approx 10.4$

Médiane

Définition

- C'est la valeur possible de la variable (observée ou non) qui sépare la série d'observations en deux ensembles d'effectifs égaux.
- 50% des observations sont inférieures à la médiane et 50% sont supérieures à elle.
- C'est le point au milieu de la série ordonnée.

Médiane

Définition

- C'est la valeur possible de la variable (observée ou non) qui sépare la série d'observations en deux ensembles d'effectifs égaux.
- 50% des observations sont inférieures à la médiane et 50% sont supérieures à elle.
- C'est le point au milieu de la série ordonnée.

Calcul - Variable discrète

- On ordonne les observations en liste croissante

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$$

- Si le nombre d'observations n est impair, il y a une seule valeur au milieu de la série ordonnée qui sera la médiane : $M = x_{(\frac{n+1}{2})}$
- Sinon, il y a deux valeurs au milieu de la série et donc la médiane sera leur moyenne : $M = \frac{x_{(\frac{n}{2})}+x_{(\frac{n}{2}+1)}}{2}$

Médiane - exemples

Répartition des hôtels à Paris en 2010 en fonction de nombre d'étoiles

-

Classement d'hôtel	0	1	2	3	4	5
Effectifs	30	100	500	590	148	62
Effectifs cumulées	30	130	630	1220	1368	1430

Médiane - exemples

Répartition des hôtels à Paris en 2010 en fonction de nombre d'étoiles

-

Classement d'hôtel	0	1	2	3	4	5
Effectifs	30	100	500	590	148	62
Effectifs cumulées	30	130	630	1220	1368	1430

- $n = 1430$ pair, alors $M = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{x_{(715)} + x_{(716)}}{2} = \frac{3+3}{2} = 3$

Médiane - exemples

Répartition des hôtels à Paris en 2010 en fonction de nombre d'étoiles

-

Classement d'hôtel	0	1	2	3	4	5
Effectifs	30	100	500	590	148	62
Effectifs cumulées	30	130	630	1220	1368	1430

- $n = 1430$ pair, alors $M = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{x_{(715)} + x_{(716)}}{2} = \frac{3+3}{2} = 3$

Nb d'enfants en millier de 0 à 18 ans par famille en France (2012).

-

Nombre d'enfants	1	2	3	4 et plus	Total
Effectifs	3 479,9	2 978	987,2	277,9	7 723
Effectifs cumulés	3 479,9	6 457.9	7 445.1	7 723	

Médiane - exemples

Répartition des hôtels à Paris en 2010 en fonction de nombre d'étoiles

-

Classement d'hôtel	0	1	2	3	4	5
Effectifs	30	100	500	590	148	62
Effectifs cumulées	30	130	630	1220	1368	1430

- $n = 1430$ pair, alors $M = \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{x_{(715)} + x_{(716)}}{2} = \frac{3+3}{2} = 3$

Nb d'enfants en millier de 0 à 18 ans par famille en France (2012).

-

Nombre d'enfants	1	2	3	4 et plus	Total
Effectifs	3 479,9	2 978	987,2	277,9	7 723
Effectifs cumulés	3 479,9	6 457.9	7 445.1	7 723	

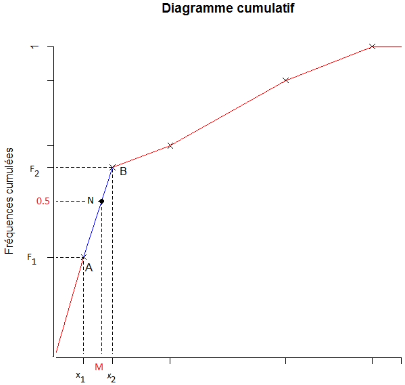
- $n = 7\ 723$ impair, alors $M = x_{(\frac{n+1}{2})} = x_{(3862)} = 2$.

Médiane - Variable continue (interpolation linéaire)

- Classe médiane : $M \in [x_1; x_2[$
- $A(x_1; F_1)$, $B(x_2; F_2)$ et $N(M; 0,5)$
- Hypothèse : La répartition à l'intérieur de chaque classe est uniforme.
- A, N et P sont alignés ;
- pente (AN)= pente (AB)

$$\frac{0,5 - F_1}{M - x_1} = \frac{F_2 - F_1}{x_2 - x_1}$$

- Donc $M = x_1 + (x_2 - x_1)\left(\frac{0,5 - F_1}{F_2 - F_1}\right)$
- M est une valeur approchée de la médiane



Médiane - Variable continue (exemple)

	Notes	Fi
	$[0;5[$	0.1
	$[5;8[$	0.46
		0.5
$M \in$	$[8;12[$	0.6
	$[12;15[$	0.77
	$[15;20[$	1

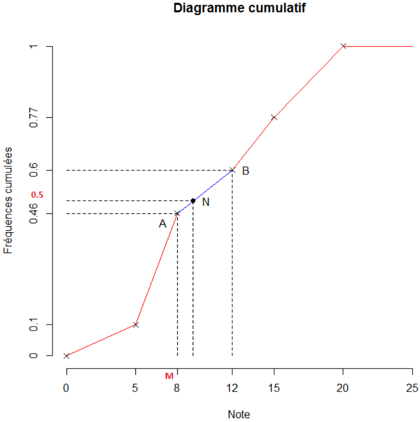
- Classe médiane : $[8;12[$
- $A(8;0,46)$, $B(12;0,6)$ et $N(M;0,5)$

$$\frac{0,5-0,46}{M-8} = \frac{0,6-0,46}{12-8}$$

- Donc $M \approx 9,14$

Médiane - Variable continue (exemple)

	Notes	Fi
	$[0;5[$	0.1
	$[5;8[$	0.46
		0.5
$M \in$	$[8;12[$	0.6
	$[12;15[$	0.77
	$[15;20[$	1



- Classe médiane : $[8;12[$
- $A(8;0,46)$, $B(12;0,6)$ et $N(M;0,5)$

$$\frac{0,5-0,46}{M-8} = \frac{0,6-0,46}{12-8}$$

- Donc $M \approx 9,14$

Comparaison entre la moyenne et la médiane

- Soit $x_1, ..., x_n$ une série statistique, alors
  - La moyenne réalise le minimum de la fonction suivante en x
$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \sum_{i=1}^{n} (x_i - x)^2$$
.
  - La médiane réalise le minimum de la fonction suivante en x
$$g : \mathbb{R} \to \mathbb{R}, \quad g(x) = \sum_{i=1}^{n} |x_i - x|$$
.
- La moyenne est très affectée par les valeurs extrêmes, et ce n'est pas le cas pour la médiane qui est plus robuste que la moyenne.

Mode

- Le mode rend compte de l'endroit où les données sont les plus concentrées
- Le mode d'une variable discrète est la(les) valeur(s) la(les) plus fréquente(s).
- Pour une variable continue, on parle d'une classe modale qui est celle de plus grande densité.
- Lorsque la médiane égale à la moyenne et au mode, la distribution des fréquences est dit symétrique. Sinon, elle est dit asymétrique.

Quartiles et déciles

Quartiles

- Quartiles : Ce sont les trois valeurs qui divisent la série statistique en quatre groupes d'effectifs égaux.
- variable discrète : d'abord, on ordonne la série statistique puis
  - ordre de Q_1 est le premier entier $\geq \frac{n}{4}$.
  - $Q_2 = M$
  - ordre de Q_3 est le premier entier $\geq \frac{3n}{4}$
- Variable continue : Méthode d'interpolation linéaire avec $F(Q_1) = 0.25$ et $F(Q_3) = 0.75$.

Quartiles et déciles

Quartiles

- Quartiles : Ce sont les trois valeurs qui divisent la série statistique en quatre groupes d'effectifs égaux.
- variable discrète : d'abord, on ordonne la série statistique puis
  - ordre de Q_1 est le premier entier $\geq \frac{n}{4}$.
  - $Q_2 = M$
  - ordre de Q_3 est le premier entier $\geq \frac{3n}{4}$
- Variable continue : Méthode d'interpolation linéaire avec $F(Q_1) = 0.25$ et $F(Q_3) = 0.75$.

Déciles

- Déciles : ce sont les neuf valeurs qui divisent la série statistique en dix groupes d'effectifs égaux.
- Le calcul des déciles est similaire au calcul des quartiles.

Quartiles et déciles

Quartiles

- Quartiles : Ce sont les trois valeurs qui divisent la série statistique en quatre groupes d'effectifs égaux.
- variable discrète : d'abord, on ordonne la série statistique puis
  - ordre de Q_1 est le premier entier $\geq \frac{n}{4}$.
  - $Q_2 = M$
  - ordre de Q_3 est le premier entier $\geq \frac{3n}{4}$
- Variable continue : Méthode d'interpolation linéaire avec $F(Q_1) = 0.25$ et $F(Q_3) = 0.75$.

Déciles

- Déciles : ce sont les neuf valeurs qui divisent la série statistique en dix groupes d'effectifs égaux.
- Le calcul des déciles est similaire au calcul des quartiles.

Quantiles

Le quantile d'ordre τ de la variable Y est la valeur q dont la fréquence cumulée égale à τ

Quartiles - Variable continue (exemple)

ClasseNotes	Fi
[0; 5[	0.1
[5; 8[	0.46
[8; 12[	0.6
[12; 15[	0.77
[15; 20[	1

Quartiles - Variable continue (exemple)

ClasseNotes	Fi
$[0;5[$	0.1
$[5;8[$	0.46
$[8;12[$	0.6
$[12;15[$	0.77
$[15;20[$	1

- $Q_1 \in [5;8[$
- $\frac{0.25-0.1}{Q_1-5} = \frac{0.46-0.1}{8-5}$ Donc $Q_1 \approx 6.25$

Quartiles - Variable continue (exemple)

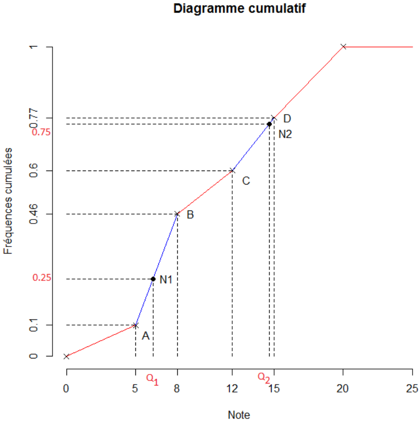
ClasseNotes	Fi
[0; 5[	0.1
[5; 8[	0.46
[8; 12[	0.6
[12; 15[	0.77
[15; 20[	1

- $Q_1 \in [5; 8[$
- $\frac{0.25-0.1}{Q_1-5} = \frac{0.46-0.1}{8-5}$ Donc $Q_1 \approx 6.25$
- $Q_3 \in [12; 15[$
- $\frac{0.75-0.6}{Q_3-12} = \frac{0.77-0.6}{15-12}$ Donc $Q_3 = 14.65$

Quartiles - Variable continue (exemple)

ClasseNotes	Fi
[0; 5[	0.1
[5; 8[	0.46
[8; 12[	0.6
[12; 15[	0.77
[15; 20[	1

- $Q_1 \in [5; 8[$
- $\frac{0.25-0.1}{Q_1-5} = \frac{0.46-0.1}{8-5}$ Donc $Q_1 \approx 6.25$
- $Q_3 \in [12; 15[$
- $\frac{0.75-0.6}{Q_3-12} = \frac{0.77-0.6}{15-12}$ Donc $Q_3 = 14.65$



Déciles - Variable continue (exemple)

ClasseNotes	Fi
$[0; 5[$	0.1
$[5; 8[$	0.46
$[8; 12[$	0.6
$[12; 15[$	0.77
$[15; 20[$	1

Déciles - Variable continue (exemple)

ClasseNotes	Fi
$[0;5[$	0.1
$[5;8[$	0.46
$[8;12[$	0.6
$[12;15[$	0.77
$[15;20[$	1

- $D_1 = 5$

Déciles - Variable continue (exemple)

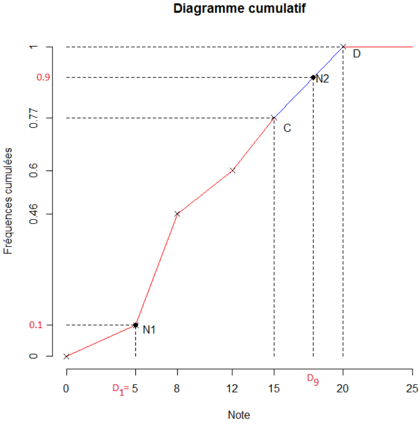
ClasseNotes	Fi
$[0;5[$	0.1
$[5;8[$	0.46
$[8;12[$	0.6
$[12;15[$	0.77
$[15;20[$	1

- $D_1 = 5$
- $D_9 \in [15;20[$
- $\frac{0{,}9-0.77}{D_9-15} = \frac{1-0.77}{20-15}$, donc $D_9 \approx 17.83$

Déciles - Variable continue (exemple)

ClasseNotes	Fi
[0; 5[	0.1
[5; 8[	0.46
[8; 12[	0.6
[12; 15[	0.77
[15; 20[	1

- $D_1 = 5$
- $D_9 \in [15; 20[$
- $\frac{0{,}9-0.77}{D_9-15} = \frac{1-0.77}{20-15}$, donc $D_9 \approx 17.83$



Plan

Étendue - Écart moyen absolu - Écart médian absolu

Étendue

- C'est la différence entre la plus grande et la plus petite valeur observée
- Elle n'est fonction que des deux valeurs extrêmes

Étendue - Écart moyen absolu - Écart médian absolu

Étendue

- C'est la différence entre la plus grande et la plus petite valeur observée
- Elle n'est fonction que des deux valeurs extrêmes

Écart moyen absolu

- C'est la moyenne de la valeur absolue des écarts à la moyenne
- c-à-d : $E_m = \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|$

Étendue - Écart moyen absolu - Écart médian absolu

Étendue

- C'est la différence entre la plus grande et la plus petite valeur observée
- Elle n'est fonction que des deux valeurs extrêmes

Écart moyen absolu

- C'est la moyenne de la valeur absolue des écarts à la moyenne
- c-à-d : $E_m = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$

Écart médian absolu

- C'est la moyenne de la valeur absolue des écarts à la médiane
- c-à-d : $E_M = \frac{1}{n}\sum_{i=1}^{n}|x_i - M|$

Étendue - Écart moyen absolu - Écart médian absolu

Étendue

- C'est la différence entre la plus grande et la plus petite valeur observée
- Elle n'est fonction que des deux valeurs extrêmes

Écart moyen absolu

- C'est la moyenne de la valeur absolue des écarts à la moyenne
- c-à-d : $E_m = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$

Écart médian absolu

- C'est la moyenne de la valeur absolue des écarts à la médiane
- c-à-d : $E_M = \frac{1}{n} \sum_{i=1}^{n} |x_i - M|$

NB : Si les données sont agrégées, on multiple la valeur absolue de chaque modalité par l'effectif associé

Exemple

Étendue, E_m et E_M

Classe	Effectifs	Centres	$\|c_i - \bar{x}\|$	$n_i\|c_i - \bar{x}\|$	$\|c_i - M\|$	$n_i\|c_i - M\|$
[0; 5[	6.00	2.50	7.99	47.94	6.64	39.84
[5; 8[	21.00	6.50	3.99	83.79	2.64	55.44
[8; 12[	8.00	10.00	0.49	3.92	0.86	6.88
[12; 15[	10.00	13.50	3.01	30.10	4.36	43.60
[15; 20[	14.00	17.50	7.01	98.14	8.36	117.04
Total	59			263.89		262.8

Exemple

Étendue, E_m et E_M

Classe	Effectifs	Centres	$\|c_i - \bar{x}\|$	$n_i\|c_i - \bar{x}\|$	$\|c_i - M\|$	$n_i\|c_i - M\|$
[0; 5[	6.00	2.50	7.99	47.94	6.64	39.84
[5; 8[	21.00	6.50	3.99	83.79	2.64	55.44
[8; 12[	8.00	10.00	0.49	3.92	0.86	6.88
[12; 15[	10.00	13.50	3.01	30.10	4.36	43.60
[15; 20[	14.00	17.50	7.01	98.14	8.36	117.04
Total	59			263.89		262.8

- $\bar{x} \approx 10.49$ et $M \approx 9.14$
- Étendue =20-0=20
- $E_m = 263.89/59 \approx 4.47$
- $E_M = \frac{262.8}{59} \approx 4.45$

Ecart inter-quartiles - Box-plot (Boite à moustaches)

Ecart inter-quartiles

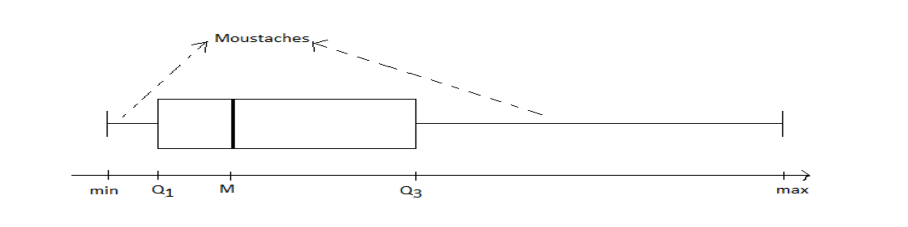
- Au moins, 50% des observations $\in [Q_1; Q_3]$.
- Ecart inter-quartiles $=$ la longueur de l'intervalle $[Q_1; Q_3]$
  $IQ = Q_3 - Q_1$

Ecart inter-quartiles - Box-plot (Boite à moustaches)

Ecart inter-quartiles

- Au moins, 50% des observations $\in [Q_1; Q_3]$.
- Ecart inter-quartiles = la longueur de l'intervalle $[Q_1; Q_3]$
  $IQ = Q_3 - Q_1$

Box-plot (Boite à moustaches)



On peut aussi se limiter par D_1 et D_9 au lieu de *max* et *min*

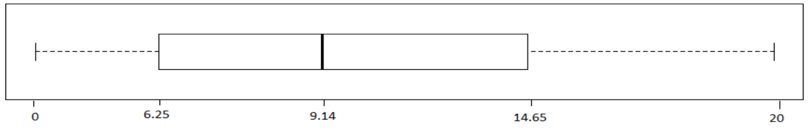
Exemple : Notes des étudiants

- Au moins, 50% des notes $\in [6.25; 14.65]$
- $IQ = Q_3 - Q_1 \approx 8.4$

Exemple : Notes des étudiants

- Au moins, 50% des notes $\in [6.25; 14.65]$
- $IQ = Q_3 - Q_1 \approx 8.4$

Boxplot des notes



Variance et écart-type

- **Variance** = la moyenne des carrés des écarts à la moyenne $V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Variance et écart-type

Variance et écart-type

- **Variance** = la moyenne des carrés des écarts à la moyenne $V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- **Propriété :** $V = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2$ (La moyenne des carrés - le carré de la moyenne)

Variance et écart-type

Variance et écart-type

- **Variance** = la moyenne des carrés des écarts à la moyenne $V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- **Propriété :** $V = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2$ (La moyenne des carrés - le carré de la moyenne)
- **Données agrégées d'une variable discrète de K modalités** :

$$
\begin{aligned}
V &= \frac{1}{n}\sum_{i=1}^{K} n_i(x_i - \bar{x})^2 &&= \sum_{i=1}^{K} f_i(x_i - \bar{x})^2 \\
&= \left(\frac{1}{n}\sum_{i=1}^{K} n_i x_i^2\right) - \bar{x}^2 &&= \left(\sum_{i=1}^{K} f_i x_i^2\right) - \bar{x}^2
\end{aligned}
$$

Variance et écart-type

- **Variance** = la moyenne des carrés des écarts à la moyenne $V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- **Propriété :** $V = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2$ (La moyenne des carrés - le carré de la moyenne)
- **Données agrégées d'une variable discrète de K modalités** :

$$\begin{aligned} V &= \frac{1}{n}\sum_{i=1}^{K} n_i(x_i - \bar{x})^2 &&= \sum_{i=1}^{K} f_i(x_i - \bar{x})^2 \\ &= \left(\frac{1}{n}\sum_{i=1}^{K} n_i x_i^2\right) - \bar{x}^2 &&= \left(\sum_{i=1}^{K} f_i x_i^2\right) - \bar{x}^2 \end{aligned}$$

- **Données agrégées d'une variable continue de K classes** :

$$\begin{aligned} V &= \frac{1}{n}\sum_{i=1}^{K} n_i(c_i - \bar{x})^2 &&= \sum_{i=1}^{K} f_i(c_i - \bar{x})^2. \\ &= \left(\frac{1}{n}\sum_{i=1}^{K} n_i c_i^2\right) - \bar{x}^2 &&= \left(\sum_{i=1}^{K} f_i c_i^2\right) - \bar{x}^2 \end{aligned}$$

où $\begin{cases} n_i & \text{est l'effectif de la } i^{\text{ème}} \text{ classe, et } n \text{ est l'effectif total,} \\ f_i & \text{est la fréquence de la même classe.} \\ c_i & \text{est le centre de la } i^{\text{ème}} \text{ classe} \end{cases}$

Variance et écart-type

Variance et écart-type

- **Variance** = la moyenne des carrés des écarts à la moyenne $V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- **Propriété :** $V = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2$ (La moyenne des carrés - le carré de la moyenne)
- **Données agrégées d'une variable discrète de K modalités** :

$$\begin{aligned} V &= \frac{1}{n}\sum_{i=1}^{K} n_i(x_i - \bar{x})^2 &&= \sum_{i=1}^{K} f_i(x_i - \bar{x})^2 \\ &= \left(\frac{1}{n}\sum_{i=1}^{K} n_i x_i^2\right) - \bar{x}^2 &&= \left(\sum_{i=1}^{K} f_i x_i^2\right) - \bar{x}^2 \end{aligned}$$

- **Données agrégées d'une variable continue de K classes** :

$$\begin{aligned} V &= \frac{1}{n}\sum_{i=1}^{K} n_i(c_i - \bar{x})^2 &&= \sum_{i=1}^{K} f_i(c_i - \bar{x})^2. \\ &= \left(\frac{1}{n}\sum_{i=1}^{K} n_i c_i^2\right) - \bar{x}^2 &&= \left(\sum_{i=1}^{K} f_i c_i^2\right) - \bar{x}^2 \end{aligned}$$

où $\begin{cases} n_i & \text{est l'effectif de la } i^{\text{ème}} \text{ classe, et } n \text{ est l'effectif total,} \\ f_i & \text{est la fréquence de la même classe.} \\ c_i & \text{est le centre de la } i^{\text{ème}} \text{ classe} \end{cases}$

- **Écart-type** : C'est la racine carrée de la variance : $\sigma_x = \sqrt{v}$

Variance et écart-type

- La variance indique de quelle manière la série statistique se disperse autour de sa moyenne.
- La variance est toujours positive ou nulle
- Une variance de zéro signale que toutes les valeurs sont identiques
- Une petite variance est signe que les valeurs sont proches les unes des autres
- Une variance élevée est signe que celles-ci sont très écartées
- La variance n'a pas la même unité que les observations
- l'écart-type est la racine carrée de la variance, cet indicateur s'exprimant dans la même unité que les observations.

Notes

Classe	Effectifs	Centres	$c_i - \bar{x}$	$(ci - \bar{x})^2$	$n_i(ci - \bar{x})^2$	f_i	$f_i(ci - \bar{x})^2$
[0; 5[	6.00	2.50	-7.99	63.84	383.04	0.10	6.38
[5; 8[	21.00	6.50	-3.99	15.92	334.32	0.36	5.73
[8; 12[	8.00	10.00	-0.49	0.24	1.92	0.14	0.03
[12; 15[	10.00	13.50	3.01	9.06	90.60	0.17	1.54
[15; 20[	14.00	17.50	7.01	49.14	687.96	0.24	11.79
Total	59		0		1497.85	1	25.5

Variance et écart-type - Exemple

Notes

Classe	Effectifs	Centres	$c_i - \bar{x}$	$(ci - \bar{x})^2$	$n_i(ci - \bar{x})^2$	f_i	$f_i(ci - \bar{x})^2$
[0; 5[	6.00	2.50	-7.99	63.84	383.04	0.10	6.38
[5; 8[	21.00	6.50	-3.99	15.92	334.32	0.36	5.73
[8; 12[	8.00	10.00	-0.49	0.24	1.92	0.14	0.03
[12; 15[	10.00	13.50	3.01	9.06	90.60	0.17	1.54
[15; 20[	14.00	17.50	7.01	49.14	687.96	0.24	11.79
Total	59		0		1497.85	1	25.5

Calcul

- $\bar{x} \approx 10.49$
- $V = \frac{1497.85}{59} \approx 25.4$
- $\sigma = \sqrt{25.5} \approx 5$