Université Abdelmalek Essaadi
ENCG - TANGER

Semestre 2

chap4 : Les caractéristiques de dispersion

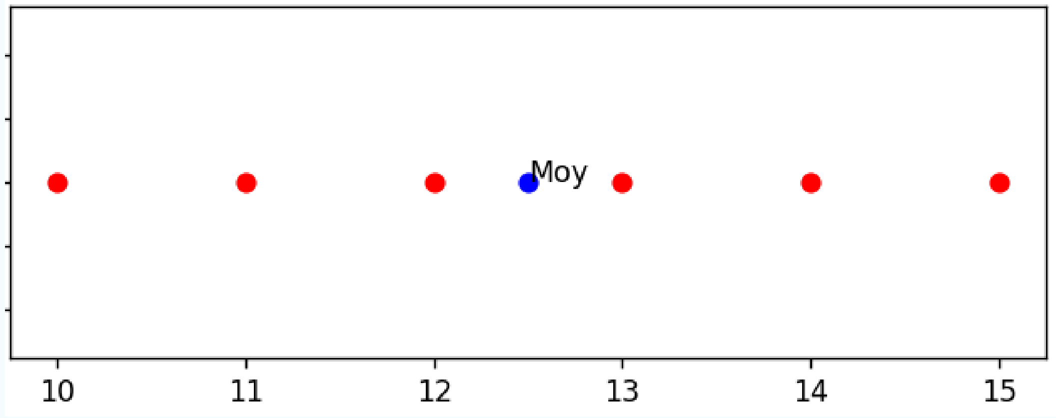
Prof: Abdelali ZBAKH <a.zbakh@uae.ac.ma>

Caractéristiques de dispersion:

Elles servent à mesurer la variabilité de la variable statistique et de juger de la pertinence (représentativité) de la caractéristique de tendance centrale.

Exemple : on considère les deux séries suivantes:

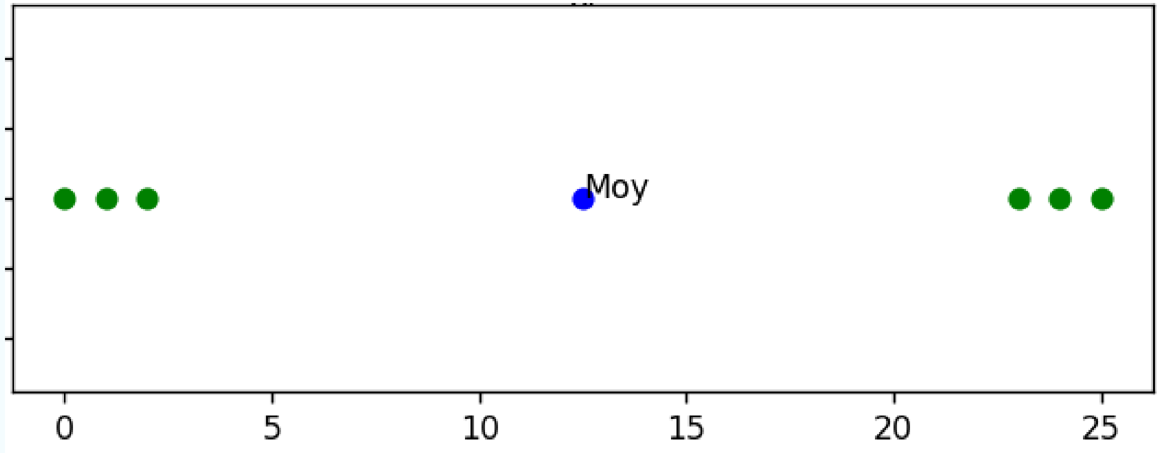
- ✓ x = 10,11, 12, 13, 14, 15.
- ✓ La moyenne et la médiane sont égales à **12.5**.



Statistique Descriptive

Caractéristiques de dispersion:

✓ x =0, 1, 2, 23, 24, 25 a la même moyenne et la même médiane que la première.



Elle en est pourtant très différente : les observations sont beaucoup plus écartées les unes des autres.

La dispersion : on évalue la distance entre les précédentes caractéristiques ($\bar{x}$ et Me) et la série des observations xi

Caractéristiques de dispersion:

- ✓ Etendue (intervalle de variation)
- ✓ Ecarts interquantiles
- ✓ Ecart absolu
- ✓ Variance
- ✓ Ecart-type

chap4 : Les caractéristiques de dispersion

L'étendue:

C'est la différence (écart) entre la plus grande et la plus petite valeur observée

Interprétation :

- Plus l'étendue d'une série est grande, plus la série est hétérogène.
- Plus l'étendue est petite, plus la série est homogène.

$$E = \max(x_i) - \min(x_i)$$

Exemple: Le tableau suivant donne la répartition des notes de 31 étudiants.

- $\overline{x}$**=10,97**
- **L'étendue de la série:**

E = 18-5 =13

Notes	Effectif	Eff cumulé Croissant
5	1	1
8	2	3
9	6	9
10	7	**16**
11	5	21
12	4	25
14	3	28
16	2	30
18	1	31

La connaissance de l'étendue permet de mieux cerner la **dispersion autour des valeurs de position**. Ainsi, une étendue élevée par rapport à la moyenne arithmétique renseigne sur une importante dispersion

Application 1:

les notes obtenues par 3 étudiants sont représentées par les séries suivantes:

- ✓ Samir: 15 ; 9 ; 14 ; 13 ; 10 ; 12 ; 12 ; 11 ; 10
- ✓ Karima: 4 ; 6 ; 18 ; 7 ; 17 ; 12 ; 12 ; 18
- ✓ Rachid: 13 ; 13 ; 12 ; 10 ; 12 ; 3 ; 14 ; 12 ; 14 ; 15

1. Pour chaque série, calculer la moyenne, la médiane et l'étendue
2. Interpréter les résultats.

Application 1: Solution

- ✓ Samir: 15 ; 9 ; 14 ; 13 ; 10 ; 12 ; 12 ; 11 ; 10

Moyenne= $(15+9+14+13+10+12+12+11+10)\div 9\approx 11{,}78$

$\underbrace{9 ; 10 ; 10 ; 11}_{\text{4 notes}} ; \mathbf{12} ; \underbrace{12 ; 13 ; 14 ; 15}_{\text{4 notes}}$

médiane

$E=15-9=6$

- ✓ Karima: 4 ; 6 ; 18 ; 7 ; 17 ; 12 ; 12 ; 18

Moyenne= $(4+6+18+7+17+12+12+18)\div 8=11{,}75$

$\underbrace{4 ; 6 ; 7 ; 12}_{\text{4 notes}} // \underbrace{12 ; 17 ; 18 ; 18}_{\text{4 notes}}$ Me=12

$E= 18-4=14$

- ✓ Rachid: 13 ; 13 ; 12 ; 10 ; 12 ; 3 ; 14 ; 12 ; 14 ; 15

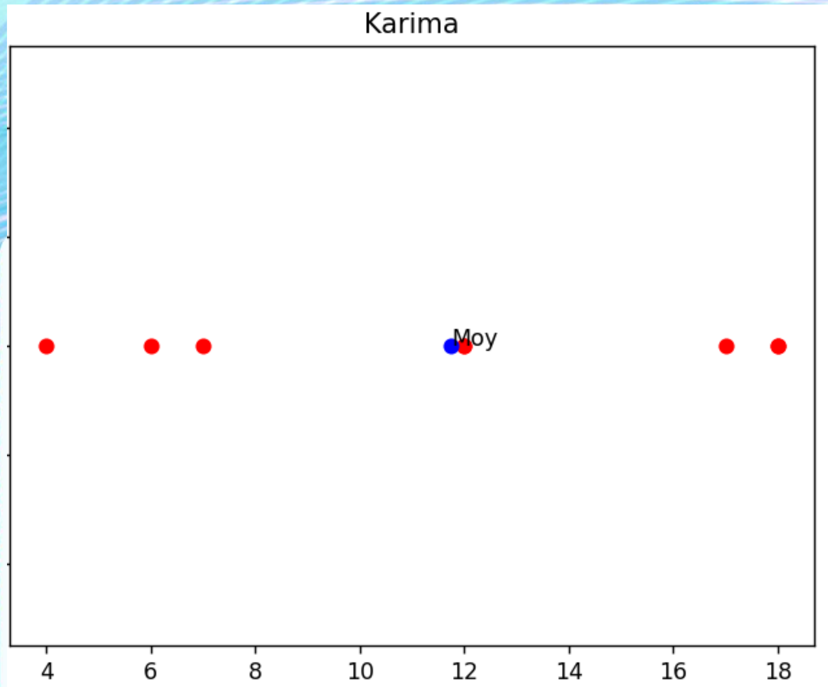
Moyenne= $(13+13+12+10+12+3+14+12+14+15)\div 10=11{,}8$

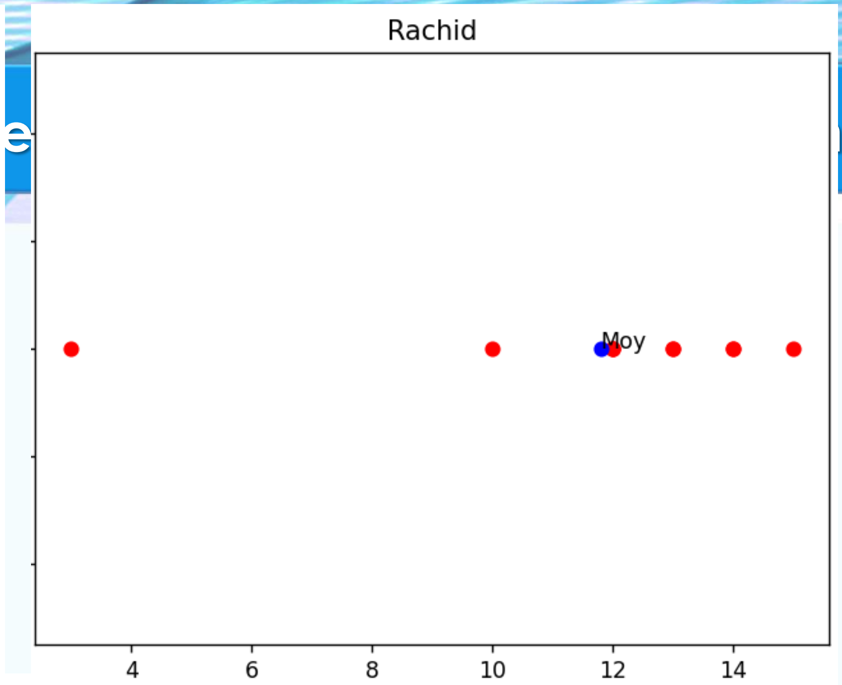
$\underbrace{3 ; 10 ; 12 ; 12 ; 12}_{\text{5 notes}} // \underbrace{13 ; 13 ; 14 ; 14 ; 15}_{\text{5 notes}}$ Me=12,5

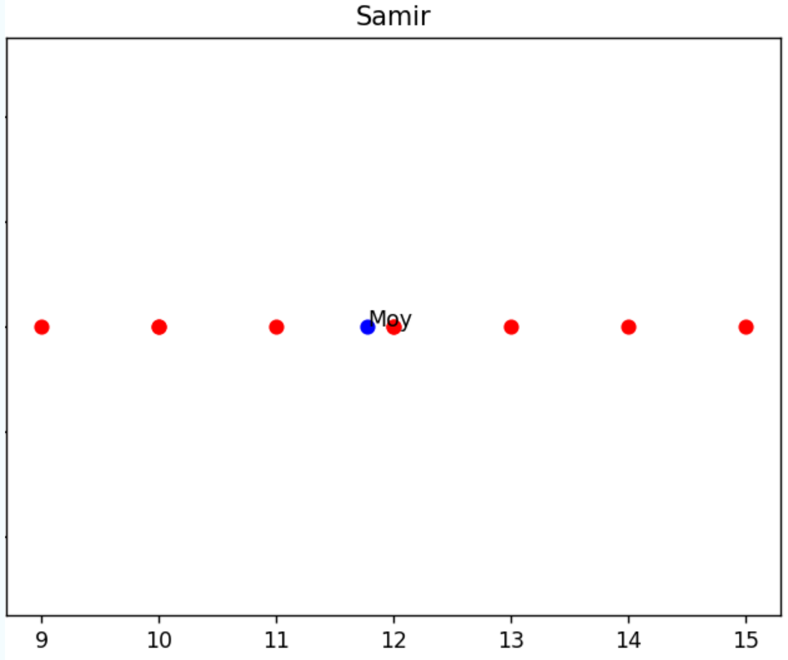
$E= 15-3=12$

Interprétation :

Les étendues de Karima et Rachid sont plus grandes, donc leurs notes sont plus hétérogènes (plus irrégulières, plus dispersées) que celles de Samir.







<u>Interprétation :</u>

Les étendues de Karima et Rachid sont plus grandes, donc leurs notes sont plus hétérogènes (plus irrégulières, plus dispersées) que celles de Samir.

Statistique Descriptive

Application 2:

Soit le relevé des tailles des élèves d'une classe de seconde :

Tailles	[150 , 160 [	[160 , 165 [	[165 , 170 [	[170 , 175 [	[175 , 180 [	[180 , 190 [
Effectifs	3	1	10	4	7	4

- Calculer l'étendue.

Les intervalles interquantiles:

- C'est la différence entre le dernier et le premier quantile, ainsi :
  - ✓ **l'intervalle interquartile** : **$IQ = Q_3 - Q_1$** contenant 50% des observations ;
  - ✓ **l'intervalle interdécile** : **$ID = D_9 - D_1$** contenant 80% des observations ;
  - ✓ **l'intervalle intercentile** : **$IC = C_{99} - C_1$** contenant 98% des observations.

- L' écart interquartile mesure la **dispersion** des valeurs autour de la médiane ; plus **l'écart** est **petit**, plus les valeurs de la série appartenant à l'intervalle interquartile sont **concentrées** autour de la **médiane**.

- Contrairement à l'étendue qui mesure l'écart entre la plus grande et la plus petite valeur, **l'écart interquartile élimine les valeurs extrêmes** qui peuvent être douteuses (des valeurs rares ou aberrantes), cependant il ne tient compte que de 50% de l'effectif.

Q

Exemple 1: Le tableau suivant donne la répartition des notes de 31 étudiants.

Notes	Effectif	Eff cumulé Croissant
5	1	1
8	2	3
9	6	9
10	7	**16**
11	5	21
12	4	25
14	3	28
16	2	30
18	1	31

On a

- N=31
- N/4 = 7.75 donc Q1 = 9
- N/2=15.5 donc Q2=Me = 10
- 3N/4 = 23.25 donc Q3 = 12
- Intervalle interquartile = **[9 , 12]**
- L'écart interquartile = **IQ = 12-9 = 3**

Exemple 2: On souhaite comparer les salaires mensuels, en euros, des employés de deux entreprises A et B. On dispose des données suivantes.

	Plus petit salaire	Q_1	Médiane	Q_3	Plus grand salaire
Entreprise A	1 300	1 600	1 800	2 400	6 500
Entreprise B	1 350	1 400	1 800	3 100	6 800

- Le salaire médian est le même dans les deux entreprises et les valeurs extrêmes sont sensiblement égales.
- Dans l'entreprise A, l'intervalle [1 600 ; 2 400] contient environ 50 % des salaires les plus centraux. La différence 2 400 – 1 600 = **800** est appelée **écart interquartile**.
- Dans l'entreprise B, l'intervalle [1 400 ; 3 100] contient environ 50 % des salaires les plus centraux. L'écart interquartile est 3 100 – 1 400 = **1 700**.
- La comparaison des écarts interquartiles des deux entreprises permet de comparer la **dispersion** des deux séries : les 50 % des salaires les plus centraux sont plus étendus dans l'entreprise B.
- En conclusion, les salaires sont plus dispersés autour du salaire médian dans l'entreprise B.

Ecart absolu moyen:

C'est la moyenne des écarts absolus entre chaque observation et la moyenne.

- Cas simple : $\boldsymbol{EM = \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{X}|}$
- Cas avec pondération : $\boldsymbol{EM = \frac{1}{N}\sum_{i=1}^{k} n_i \times |x_i - \bar{X}|}$

Exemple: on considère la série de 6 observations : X=10, 11,12, 13, 14, 15

- La moyenne est égale à 12.5.
- L'écart absolu moyen est égal à :

EM= $\frac{1}{6}\left(|10 - 12.5| + |11 - 12.5| + |12 - 12.5| + |13 - 12.5| + |14 - 12.5| + |15 - 12.5|\right)$=**1.5**

Ecart absolu moyen:

- **L'écart absolu moyen** mesure la dispersion des valeurs observées d'une variable statistique autour d'une valeur centrale.

- Une valeur **faible** de l'écart absolu moyen traduit une **faible dispersion** des valeurs autour de la valeur centrale.

- Cependant la comparaison de cette caractéristique pour deux séries est difficile car sa valeur dépend de **l'ordre de grandeur** (échelle ou unité de mesure) des observations.

Statistique Descriptive

La variance :

- La variance est **la moyenne** des écarts (élevés au carré) des valeurs observées par rapport à la moyenne arithmétique de la série.
- On la note $V(X)$ pour une variable notée X.

- Cas simple: $V(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{X})^2$
- Cas avec pondération : $V(X) = \frac{1}{N}\sum_{i=1}^{k} n_i \times (x_i - \bar{X})^2$

En pratique, On utilise les formules:

$$V(x) = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right) - \bar{x}^2$$

$$V(x) = \left(\sum_{i=1}^{k} f_i x_i^2\right) - \bar{x}^2$$

Application 3:

- Dans une maternité, on a relevé les poids, en kilogrammes, de 30 filles à l'âge de 9 mois :

Poids (en kg)	7,2	7,7	7,9	8,3	8,5	8,7	8,8	9,0	9,1	9,5	10,2
Effectifs	1	3	1	4	3	6	3	2	2	2	3

On a fait de même pour 30 garçons :

Poids (en kg)	7,5	8,0	8,4	8,6	8,8	9,0	9,4	9,6	9,9	10,0	10,9	11,6
Effectifs	1	2	2	4	4	3	3	1	3	3	2	2

- Le poids moyen des filles est 8,71 kg et celui des garçons 9,31 kg
- Calculer la dispersion des valeurs autour de la moyenne?

Filles

Poids (en kg)	7,2	7,7	7,9	8,3	8,5	8,7	8,8	9,0	9,1	9,5	10,2
Effectifs	1	3	1	4	3	6	3	2	2	2	3

Garçons

Poids (en kg)	7,5	8,0	8,4	8,6	8,8	9,0	9,4	9,6	9,9	10,0	10,9	11,6
Effectifs	1	2	2	4	4	3	3	1	3	3	2	2

Application3:Solution

Pour mesurer la **dispersion** des valeurs autour de la moyenne, on considère chaque valeur de la série, puis on calcule la **variance**.

✓ Pour la série des poids des filles:

$$V = \frac{1\times(7{,}2-8{,}71)^2+3\times(7{,}7-8{,}71)^2+\cdots+3\times(10{,}2-8{,}71)^2}{30} = \frac{15{,}207}{30} = 0{,}5069 \ \text{kg}^2$$

✓ Pour la série des poids des garçons:

$$V' \approx 1 \text{ kg}^2.$$

- On constate donc que $V < V'$: cela signifie que, en moyenne, les valeurs de la série des poids des filles sont plus proches de leur moyenne que ne le sont celles des garçons.
- On dit que la série des poids des filles est **moins dispersée**.

L'écart type :

- Afin d'obtenir un indicateur de dispersion dans la même unité que les valeurs de la série, on définit **l'écart-type** comme la racine carrée de la variance.
- On le note $\sigma(X)$ ou σ_X , Sa formule est :

$$\boldsymbol{\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{X})^2}}$$

Ou

$$\boldsymbol{\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i \times (x_i - \bar{X})^2}}$$

Exemple:

- les valeurs de la série des poids des filles s'expriment en kg.
- La moyenne s'exprime donc également en kg : c'est 8,71 kg.
- L'écart entre chaque valeur et la moyenne s'exprime en kg. Le carré de cet écart s'exprime donc en kg^2.
- Afin d'obtenir un indicateur de dispersion dans la même unité que les valeurs de la série, on définit **l'écart-type** comme la racine carrée de la variance :

$$\sigma(X) = \sqrt{V(X)} = \sqrt{0{,}5069} = 0{,}71 \text{ kg}.$$

Filles

Poids (en kg)	7,2	7,7	7,9	8,3	8,5	8,7	8,8	9,0	9,1	9,5	10,2
Effectifs	1	3	1	4	3	6	3	2	2	2	3

- L'**écart-type** et la **variance** mesurent la dispersion de la variable autour de la moyenne.
- Des valeurs **élevées** (respectivement faibles) de ces caractéristiques traduisent une **grande** (respectivement faible) **dispersion** des valeurs autour de la moyenne.

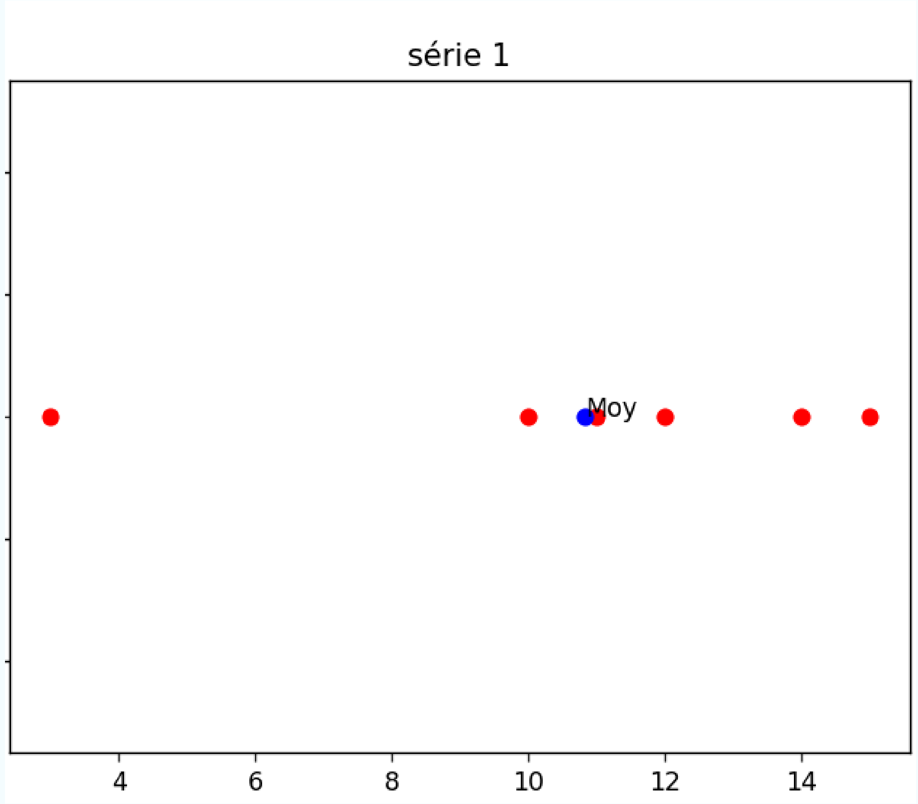
Exemple:

✓ $x_1 = 10$, $x_2 = 11$, $x_3 = 12$, $x_4 = 13$, $x_5 = 14$, $x_6 = 15$.

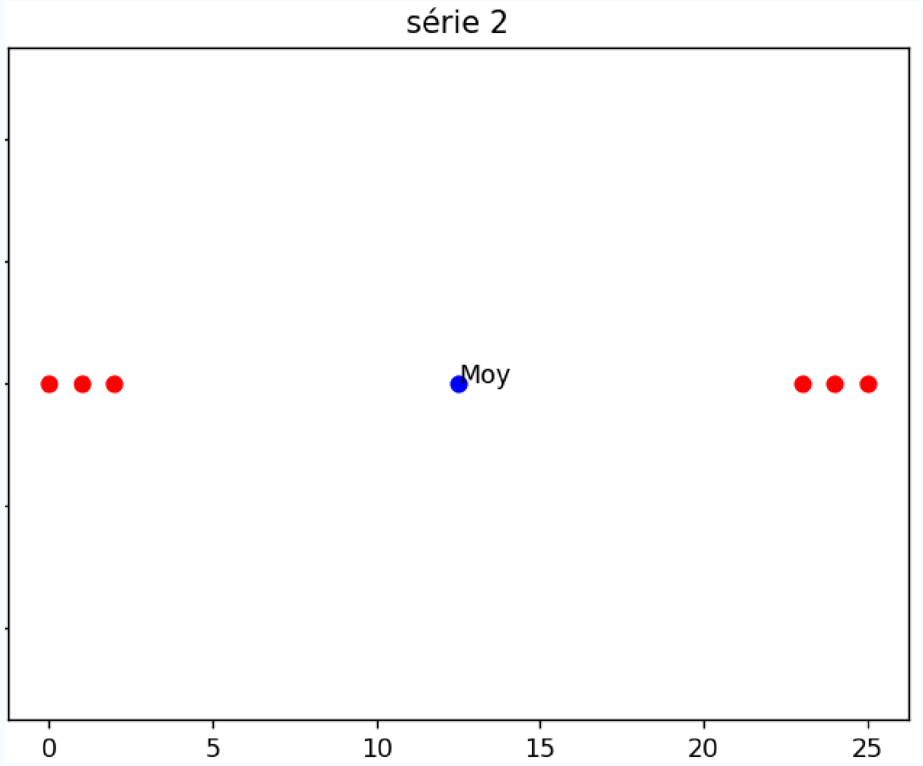
La moyenne est égale à 12.5, L'écart absolu moyen est égal à 1.5, et et l'écart-type 1,71

✓ $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 23$, $x_5 = 24$, $x_6 = 25$

- a la même moyenne et la même médiane que la précédente.
- Elle en est pourtant très différente :l'écart absolu moyen (11.5) et l'écart-type (11.53) sont beaucoup plus grands que les précédents.



l'écart-type = 1,71



l'écart-type= 11.53

Statistique Descriptive

Application 4:

- Le tableau suivant présente les recettes exécutées par les communes de Y entre 2007 et 2013 (en millions de DH).

Y	2007	2008	2009	2010	2011	2012	2013
Recettes	228,5	263,3	434,3	802,3	791,9	1027,6	982,9

- Calculer la **variance** et **l'écart type**

- **Recette moyenne $\overline{X}$**

$= \frac{1}{7}$ (228,5 + 263,3 + 434,3 + 802,3 + 791,9 + 1027,6 + 982,9)

= 647,26

- **La variance**

Var (X)= $\frac{1}{N}\sum_{i=1}^{N} X_i^2 - (\overline{X})^2$

$= \frac{1}{7}(228{,}5^2 + 263{,}3^2 + 434{,}3^2 + 802{,}3^2 + 791{,}9^2 + 1027{,}6^2 + 982{,}9^2) - 647{,}26^2$

=95 768,88

- **Ecart type:** $\sqrt{Var(X)}$

$\sigma(X) = 309{,}47$

Statistique Descriptive

Application 5:

Les salaires mensuels en (10 Dhs) de 36 employés d'une entreprise industrielle se résument comme suite dans le tableau:

Salaires mensuels(X_i)	109	120	123	105	110	590	112	
Nombre d'employés (N_i)	6	4	2	3	5	8	8	36

- Calculer la **variance** et **l'écart type**

- **Moyenne pondérée $\overline{X}$**
- $\overline{X} = \frac{1}{36}$(6x109+ 4x120+ 2x123+ 3x105+ 5x110 + 8x590 + 8x112)

= 218,4

- **La variance pondérée:**
- = Var(X) = $\frac{1}{36}$(6x109²+ 4x120²+ 2x123²+ 3x105²+ 5x110² + 8x590² + 8x112²)− $218{,}4^2$

=39 481, 5

- **Ecart type:** $\sqrt{Var(X)}$

$\sigma(X) = 169{,}69$

Coefficient de variation : l'écart-type/ l'intervalle interquartile

- Le coefficient de variation est un nombre sans dimension qui permet de comparer deux variables statistiques de natures différentes. On l'exprime souvent en pourcentage.

- Le coefficient de variation de l'écart-type est le rapport entre l'écart-type et la moyenne de la distribution. On le note CV_σ

$$\boldsymbol{CV_\sigma = \frac{\sigma(X)}{\bar{X}} = \frac{Ecart\ type}{Moyenne}}$$

- De façon analogue, on définit le coefficient de variation de l'intervalle interquartile par :

$$\boldsymbol{CV_Q = \frac{I_Q}{M_e} = \frac{Q_3 - Q_1}{Q_2}}$$

✓ Le **coefficient de variation** présente l'avantage de ne pas être sensible à l'ordre de grandeur de la variable mais seulement à la dispersion des valeurs autour de la moyenne

✓ Plus CV est grand plus la dispersion est importante

Statistique Descriptive

Coefficient de variation : l'écart-type/ l'intervalle interquartile

Exemple:

- Les deux séries suivantes correspondent aux salaires mensuels perçus par les ouvriers d'une entreprise exprimés une première fois en **DH** et une seconde fois en **centimes**.

✓ *Série 1:*1 150, 1 200, 1 600, 1 850, 2 150, 2 200, 2 350, 2 400, 3 000.

✓ *Série 2:*115 000, 120 000, 160 000, 185 000, 215 000, 220 000, 235 000, 240 000, 300 000.

- Nous trouvons **σ_1=566,55 DH** et **σ_2=56 655 centimes**
- Le fait que σ_2 est plus élevée que σ_1 ne révèle pas une grande dispersion de la second par rapport à la première : c'est seulement **l'effet des unités de mesure** (1 dh = 100 centimes).

Coefficient de variation : l'écart-type/ l'intervalle interquartile

Exemple:

- Les **coefficients de variation** (CV) associés à chaque série :

$$CV_1 = \frac{\sigma_1}{\overline{x}_1} = \frac{566,55}{1988,88} = 0,28 \quad \textbf{et} \quad CV_2 = \frac{\sigma_2}{\overline{x}_2} = \frac{56655}{198888} = 0,28.$$

Conclusion: Les deux séries sont bien de même dispersion.

Application 6:(A faire)

Soit la distribution suivante :

xi	ni
1	120
2	52
3	27
4	11
5	2

Travail à faire:

- Calculer la moyenne
- Calculer l'étendue
- Calculer l'intervalle interquartile
- Calculer l'écart absolu moyen
- Calculer la variance
- Calculer l'écart type

Application 7:(A faire)

Une enquête réalisée par la société « Soleil », spécialisée dans la vente de produits de beauté, sur la répartition de ses clients en fonction de leur salaire a donné les résultats suivants :

Salaires	Centres x_i	n_i
[0 – 1 000[	500	30
[1 000 – 1 500[	1250	25
[1 500 – 2 000[	1750	14
[2 000 – 2 500[	2250	9
[2 500 – 3 000[	2750	12
[3 000 – 3 500[	3250	6
[3 500 – 4 000[	3750	4
Totaux		100

Travail à faire:

- Calculer la moyenne
- Calculer l'étendue
- Calculer l'intervalle interquartile
- Calculer l'écart absolu moyen
- Calculer la variance
- Calculer l'écart type