Introduction aux statistiques pour les archéologues

Josef Wilczek

Masarykova univerzita, Brno

L‘univesité de Bourgogne, Dijon

josef.wilczek@hotmail.com

Statistiques descriptives

1) Tables des fréquences
2) Méthodes graphiques
3) **Les tendances centrales**
4) Les tendances de la variabilité (dispersion)
5) Distributions

Les tendances centrales

- Il s’agit des valeurs autour lesquels les autres valeurs se concentrent

- **Moyenne** – pour les donnés quantitatives
- **Médiane** – pour les données quantitatives et qualitatives sur une échelle ordinale
- **Mode** – pour les données quantitatives et qualitatives

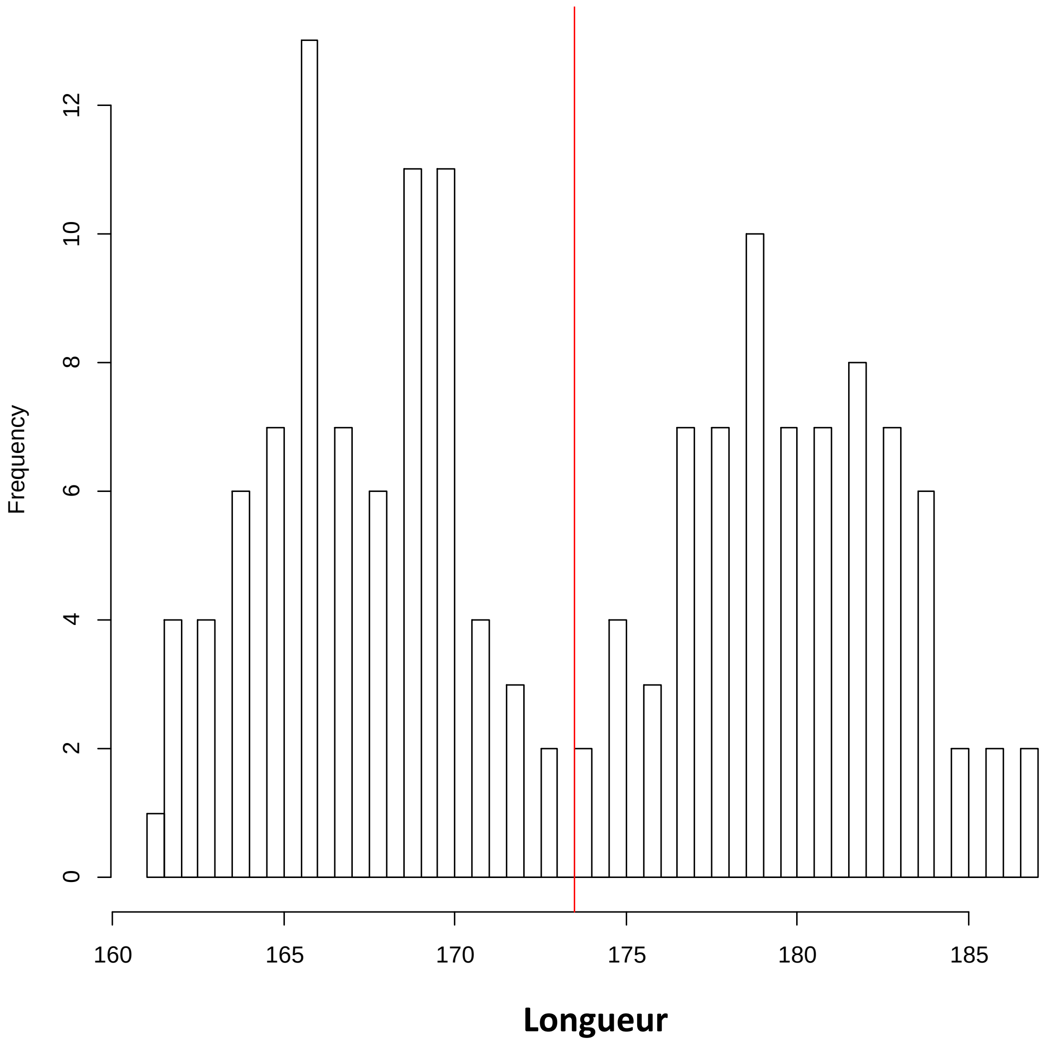
Moyenne (variables continues)

- Plusieurs moyennes
- La moyenne est plus sensible aux valeurs extrêmes (« outliers »)

- **Moyenne arithmétique**

- **Moyenne arithmétique pondérée**

Histogramme des longueurs des haches (en cm)



Moyenne (variables continues)

- Plusieurs moyennes
- La moyenne est plus sensible aux valeurs extrêmes (« outliers »)

- **Moyenne arithmétique**
  - $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

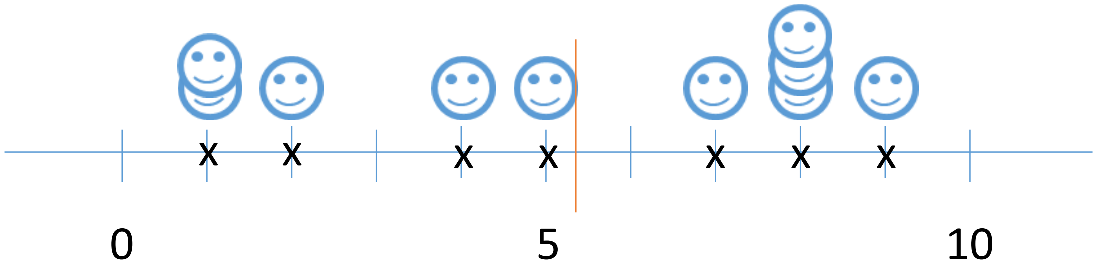
Moyenne (variables continues)

Quelle est la moyenne des bières bu par 10 étudiants ?

- Plusieurs moyennes
- La moyenne est plus sensible aux valeurs extrêmes (« outliers »)

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$n = 10$$

- **Moyenne arithmétique**
  - $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$



$$\bar{x} = \frac{1+1+2+4+5+7+8+8+8+9}{10} = 5.3$$

Moyenne (variables continues)

- Plusieurs moyennes
- La moyenne est plus sensible aux valeurs extrêmes (« outliers »)

- **Moyenne arithmétique**
  - $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Quelle est la moyenne des bières bu par 10 H et 7 F?

$H = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$

$n(H) = 10$

$F = 1, 1, 1, 2, 2, 3, 5$

$n(F) = 7$

$$\bar{x} = \frac{53 + 15}{10 + 7} = \frac{68}{17} = 4$$

Moyenne (variables continues)

- Plusieurs moyennes
- La moyenne est plus sensible aux valeurs extrêmes (« outliers »)

- **Moyenne arithmétique**
  - $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

?

?

$H =$

$F =$

$n(H) = 10$

$n(F) = 7$

$\overline{x(H)} = 5.3$

$\overline{x(F)} = 2.14$

Mais comment on va calculer la moyenne si le lendemain, personne ne se souvient de rien?

1) On sait seulement que 10H ont bu en moyenne 5.3 et 7F 2.14
2) Quelle est la moyenne par une personne ?

$$5.3 + 2.14 \neq 4$$

Moyenne (variables continues)

- Plusieurs moyennes
- La moyenne est plus sensible aux valeurs extrêmes (« outliers »)

- **Moyenne arithmétique**
  - $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

- **Moyenne arithmétique pondérée**
  - Dans le cas qu'on veut calculer la moyenne des deux échantillons
  - Dans le cas que les valeur dans l'échantillon ont l'importance (le poids) différents
  - $\bar{x} = \frac{\sum_{i=1}^{d} w_i x_i}{\sum_{i=1}^{d} w_i}$

?

?

$H =$

$F =$

$n(H) = 10$

$n(F) = 7$

$\overline{x(H)} = 5.3$

$\overline{x(F)} = 2.14$

Mais comment on va calculer la moyenne si le lendemain, personne ne se souvient de rien?

1) On sait seulement que 10H ont bu en moyenne 5.3 et 7F 2.14
2) Quelle est la moyenne par une personne ?

$$5.3 + 2.14 \neq 4$$

Moyenne (variables continues)

- Plusieurs moyennes
- La moyenne est plus sensible aux valeurs extrêmes (« outliers »)

- **Moyenne arithmétique**
  - $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

- **Moyenne arithmétique pondérée**
  - Dans le cas qu'on veut calculer la moyenne des deux échantillons
  - Dans le cas que les valeur dans l'échantillon ont l'importance (le poids) différents
  - $\bar{x} = \frac{\sum_{i=1}^{d} w_i x_i}{\sum_{i=1}^{d} w_i}$

?

?

$H =$

$F =$

$n(H) = 10$

$n(F) = 7$

$\overline{x(H)} = 5.3$

$\overline{x(F)} = 2.14$

Mais comment on va calculer la moyenne si le lendemain, personne ne se souvient de rien?

1) On sait seulement que 10H ont bu en moyenne 5.3 et 7F 2.14
2) Quelle est la moyenne par une personne ?

$$5.3 + 2.14 \neq 4$$

On va utiliser le nombre des H et F comme des poids :

$$w_H = 10$$
$$w_F = 7$$

$$\bar{x} = \frac{10 * 5.3 + 7 * 2.14}{10 + 7} = \frac{53 + 14.98}{17} = \frac{67,98}{17} = 4$$

Moyenne (variables quantitatives discrètes)

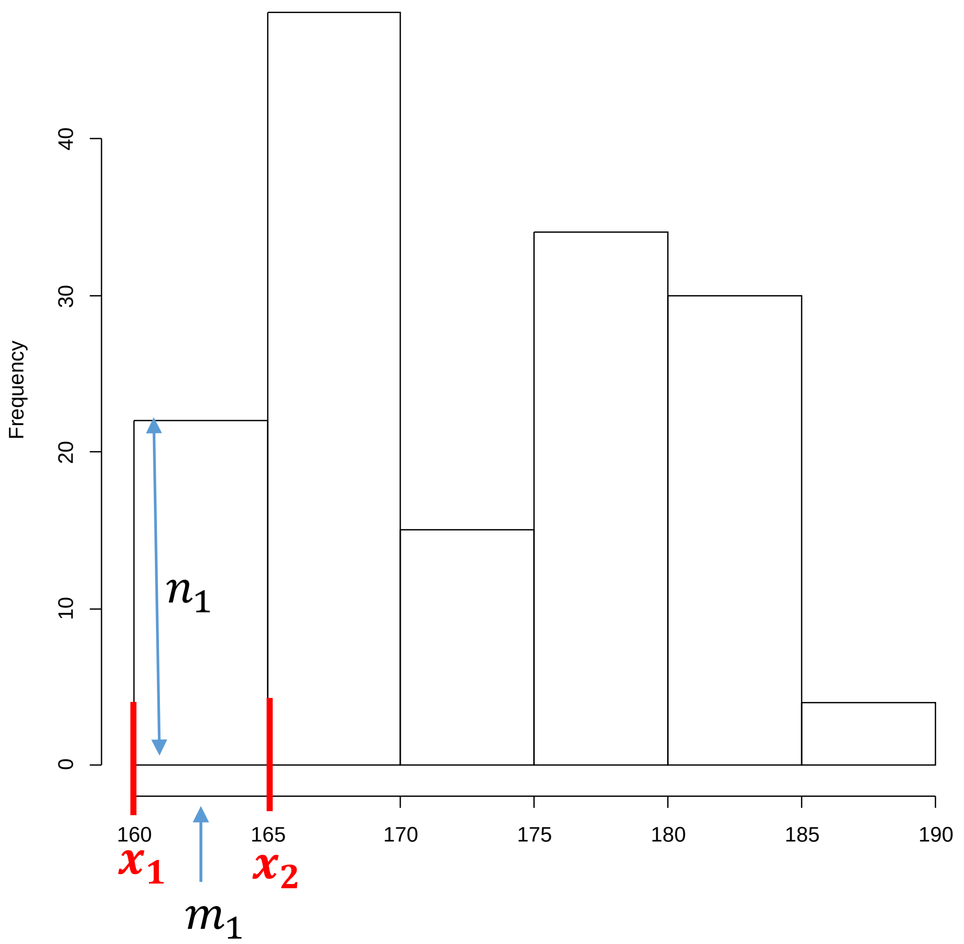
- Pour les variables en classes
- *Ex : On veut définir la taille des haches qui sont dans l'intervalles de 5 cm*
- $\bar{x} = \frac{\sum_{i=1}^{p} n_i m_i}{\sum_{i=1}^{p} n_i} = \frac{Somme\ (Frequence\ de\ chaque\ class\ *centre\ de\ classe)}{Somme\ des\ individues}$
  - $m_i = \frac{x_i + x_{i+1}}{2}$ - centre de classe i = la moyenne de deux limites de la classe i

Comment

1) Pour chaque classe
   1) On calcule la fréquence et centre
   2) On multiplie ces deux valeurs
2) On fait la somme de ces valeurs
3) … et on la divise par la somme totale des fréquences

Taille	Fréquence	Centre	Fréquence * Centre
160-165	22	162,5	
165-170	48	167,5	
170-175	15	172,5	
175-180	34	177,5	
180-185	30	182,5	
185-190	4	187,5	
Total	153		

Histogramme des tailles des haches (en cm)



Moyenne (variables quantitatives discrètes)

- Pour les variables en classes
- *Ex : On veut définir la taille des haches qui sont dans l'intervalles de 5 cm*
- $\bar{x} = \frac{\sum_{i=1}^{p} n_i m_i}{\sum_{i=1}^{p} n_i} = \frac{Somme\ (Frequence\ de\ chaque\ class\ *centre\ de\ classe)}{Somme\ des\ individues}$
    - $m_i = \frac{x_i + x_{i+1}}{2}$ - centre de classe i = la moyenne de deux limites de la classe i

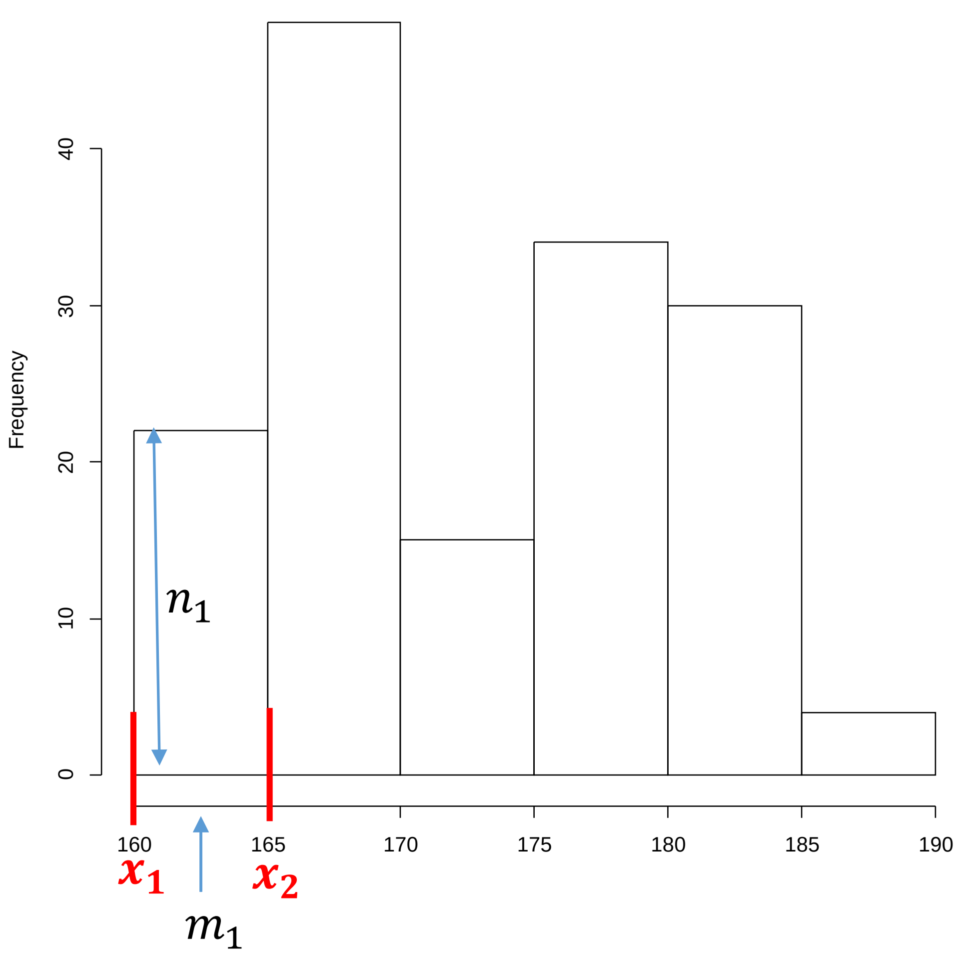
Comment

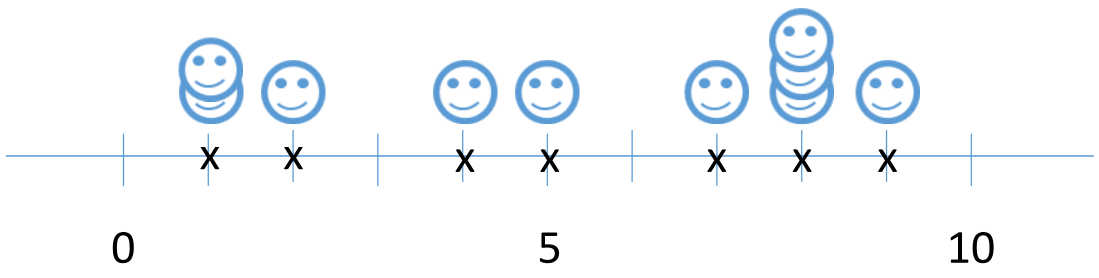
1) Pour chaque classe
    1) On calcule la fréquence et centre
    2) On multiplie ces deux valeurs
2) On fait la somme de ces valeurs
3) … et on la divise par la somme totale des fréquences

Taille	Fréquence	Centre	Fréquence * Centre
160-165	22	162,5	3575
165-170	48	167,5	8040
170-175	15	172,5	2587,5
175-180	34	177,5	6035
180-185	30	182,5	5475
185-190	4	187,5	750
Total	153		26462,5

$$\bar{x} = \frac{26462,5}{153} = 172,96$$

Histogramme des tailles des haches (en cm)




x x x x x x x
0
5
10

Médiane

- La valeur qui divise l'échantillon en deux partis
- La valeur pour laquelle la fréquence cumulative est 0.50
- Moins sensitive aux valeurs extrêmes

- **Comment calculer ?**

1) On ordonne les valeurs de la plus petit a la plus grande
2) On trouve médiane
   - Pour n impair – c'est la valeur au milieu
     - $med = x_{\left(\frac{n+1}{2}\right)}$

Q: Quelle est le médiane de ces valeurs :

$$x = 5, 3, 1, 6, 1, 3, 4$$
$$n = 7$$

$$x(ordonées) = 1, 1, 3, 3, 4, 5, 6$$

$$med(x) = x_{\left(\frac{7+1}{2}\right)} = x_{(4)} = 3$$

Médiane

- La valeur qui divise l'échantillon en deux partis
- La valeur pour laquelle la fréquence cumulative est 0.50
- Moins sensitive aux valeurs extrêmes

- **Comment calculer ?**

1) On ordonne les valeurs de la plus petit a la plus grande
2) On trouve médiane
   - Pour n impair – c'est la valeur au milieu
     - $med = x_{\left(\frac{n+1}{2}\right)}$
   - Pour n pair – c'est la moyenne arithmétique des deux valeurs qui sont le plus au milieu
     - $med = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$

Q: Quelle est le médiane de ces valeurs :

$$x = 5, 3, 1, 6, 1, 3, 4$$
$$n = 7$$

$$x(ordonées) = 1, 1, 3, 3, 4, 5, 6$$

$$med(x) = x_{\left(\frac{7+1}{2}\right)} = x_{(4)} = 3$$

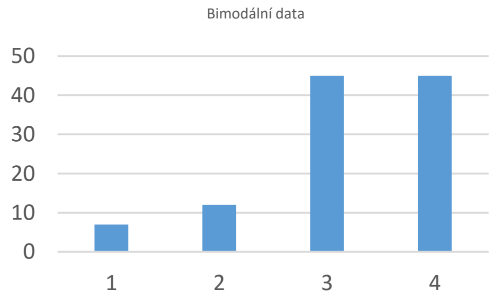
Q: Quelle est le médiane de ces valeurs :

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$n = 10$$

$$med(x) = \frac{x_{\left(\frac{10}{2}\right)} + x_{\left(\frac{10}{2}+1\right)}}{2} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{5+7}{2} = 6$$

Mode

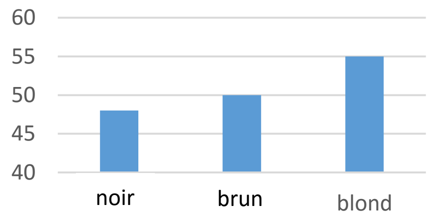
- La valeur/classe la plus fréquente
- Une distribution peut avoir un ou plusieurs modes
  - Uni, Bi, Tri-modale



- Avantages
  - pas sensible aux valeurs extrêmes
  - bon indicateur des populations hétérogènes

- Désavantage
  - Il n’est pas toujours la valeur centrale d’une distribution
  - Il peut changer si on change l’intervalle de la classe

Classe	Fréquence absolue
Blond	55
Noir	48
Brun	50

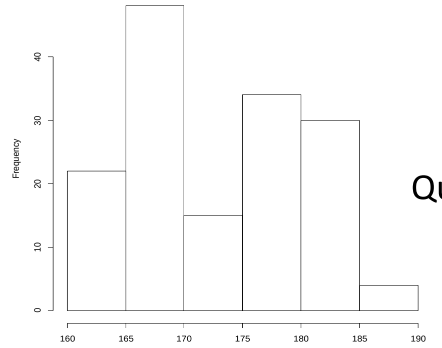


$mode = blond$

Qualitatives

Taille	Fréquence absolue
160-165	22
165-170	48
170-175	15
175-180	34
180-185	30
185-190	4

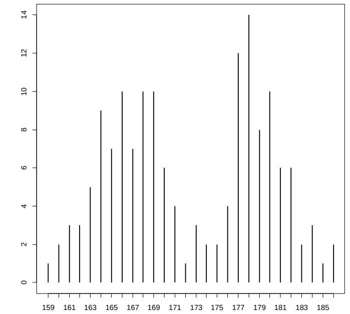
Histogramme des tailles des haches



$mode =$ 165-170

Quantitatives discrètes

Individu	Taille
Hache 1	174
Hache 2	168
Hache 3	158
…	…
Hache 152	179
Hache 153	173



$mode = 178$

Continuels

Statistiques descriptives

1) Tables des fréquences
2) Méthodes graphiques
3) Les tendances centrales
4) **Les tendances de la variabilité (dispersion)**
5) Distributions

Variabilité

- Montre comment les valeurs sont différents
  - un par apport des autres
  - d'après une certain valeur (surtout de la moyenne et médiane)

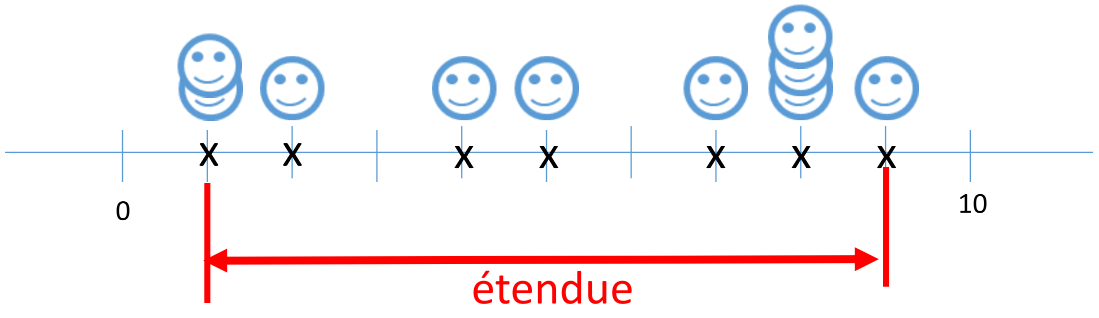
- Variabilité est exprimé par
  - Etendue
  - Ecart de la moyenne
  - La variance
  - Ecarte type
  - Quantiles

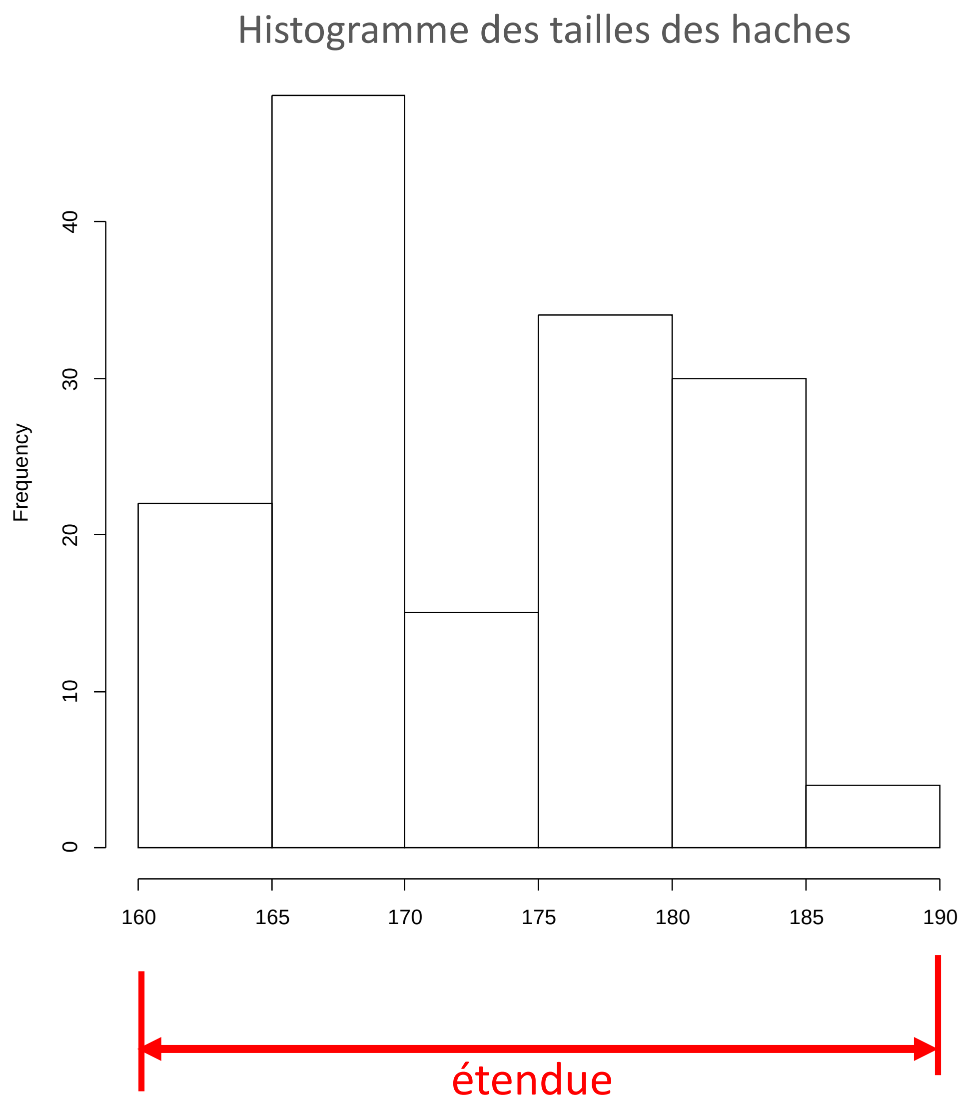
Etendue

- La différence entre la valeur la plus grande et la plus petite

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

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$





Ecart moyen et Ecart médian

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

- $Ecart\ moyen = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n}|x_i - med|}{n}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

Ecart moyen et Ecart médian

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

- $Ecart\ moyen = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n} |x_i - med|}{n}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

Ecart moyen et Ecart médian

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

$$EM = \frac{|1-5.3| + |1-5.3| + |2-5.3| + \cdots + |9-5.3|}{10} = \frac{27}{10} = 2.7$$

- $Ecart\ moyen = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n} |x_i - med|}{n}$

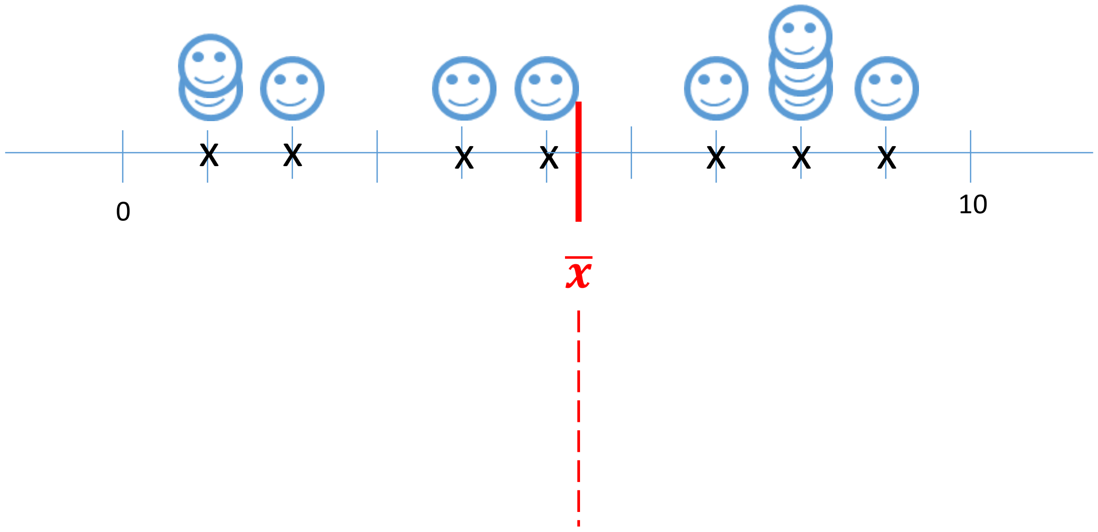
Ecart moyen et Ecart médian

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

- $Ecart\ moyen = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n}|x_i - med|}{n}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$$EM = \frac{|1-5.3| + |1-5.3| + |2-5.3| + \cdots + |9-5.3|}{10} = \frac{27}{10} = 2.7$$



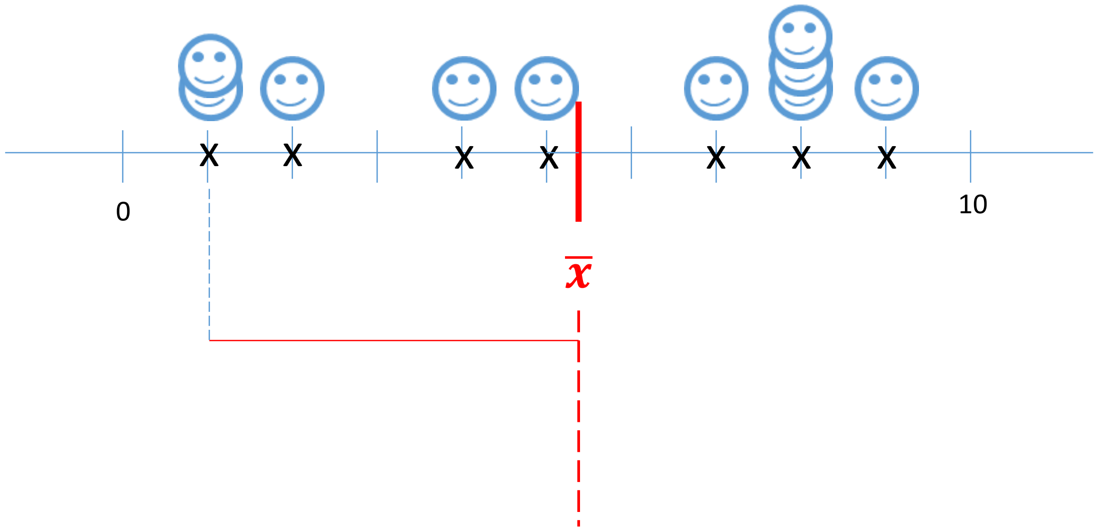
Ecart moyen et Ecart médian

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

- $Ecart\ moyen = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n}|x_i - med|}{n}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$$EM = \frac{|1-5.3| + |1-5.3| + |2-5.3| + \cdots + |9-5.3|}{10} = \frac{27}{10} = 2.7$$



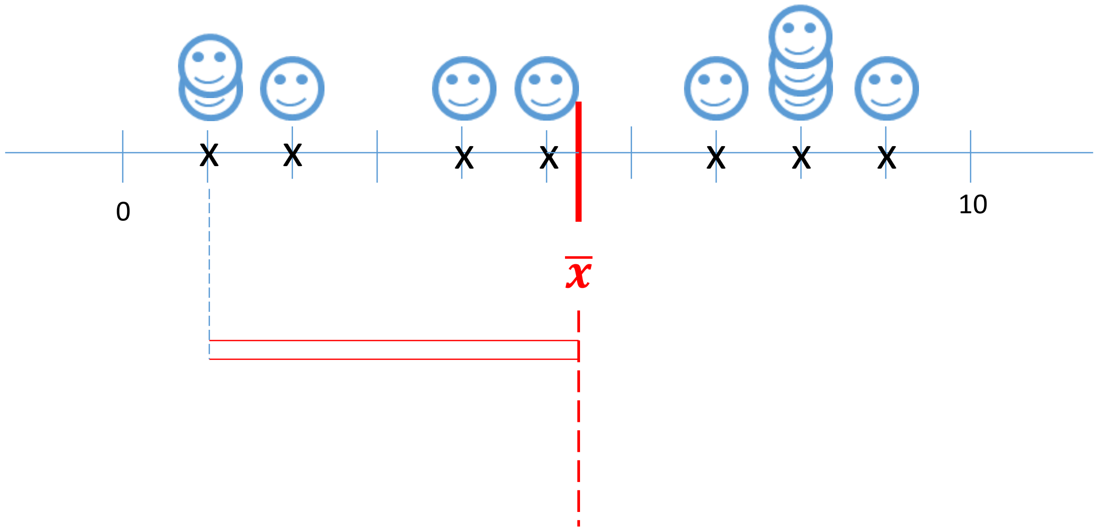
Ecart moyen et Ecart médian

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

- $Ecart\ moyen = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n}|x_i - med|}{n}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$$EM = \frac{|1-5.3| + |1-5.3| + |2-5.3| + \cdots + |9-5.3|}{10} = \frac{27}{10} = 2.7$$



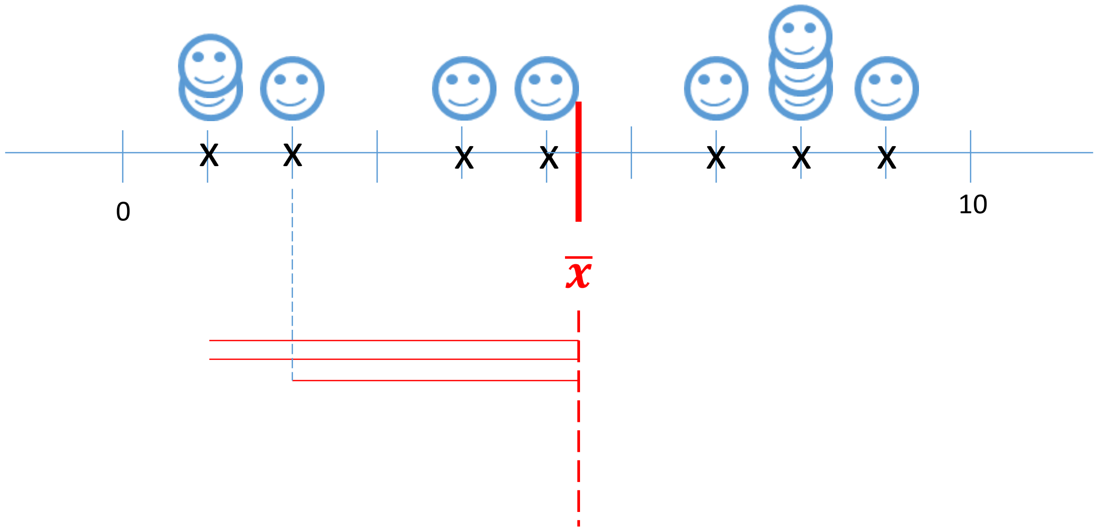
Ecart moyen et Ecart médian

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

- $Ecart\ moyen = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n} |x_i - med|}{n}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$$EM = \frac{|1-5.3| + |1-5.3| + |2-5.3| + \cdots + |9-5.3|}{10} = \frac{27}{10} = 2.7$$



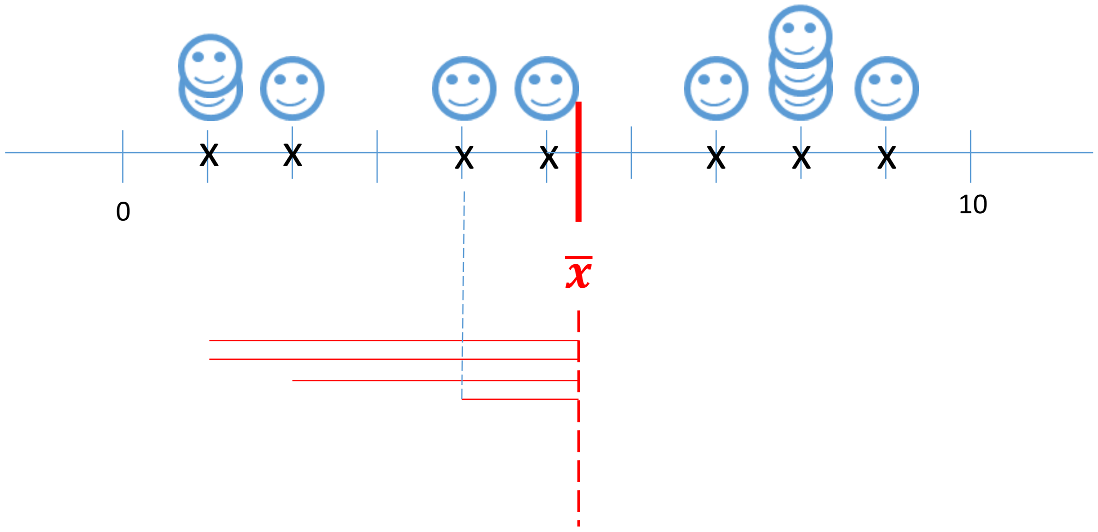
Ecart moyen et Ecart médian

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

- $Ecart\ moyen = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n}|x_i - med|}{n}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$$EM = \frac{|1-5.3| + |1-5.3| + |2-5.3| + \cdots + |9-5.3|}{10} = \frac{27}{10} = 2.7$$



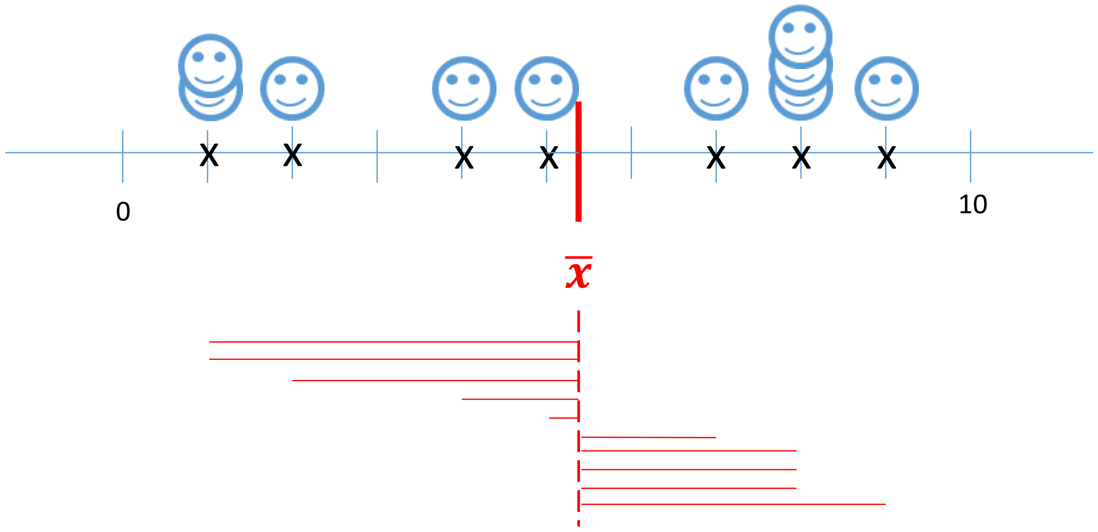
Ecart moyen et Ecart médian

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

- $Ecart\ moyen = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n} |x_i - med|}{n}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$$EM = \frac{|1 - 5.3| + |1 - 5.3| + |2 - 5.3| + \cdots + |9 - 5.3|}{10} = \frac{27}{10} = 2.7$$



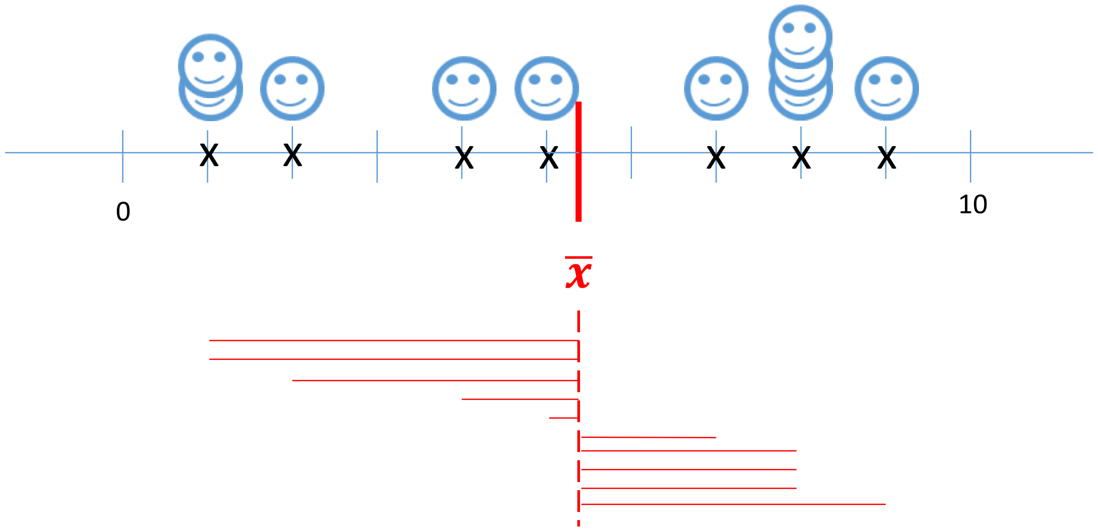
Ecart moyen et Ecart médian

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

- $Ecart\ moyen = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n}|x_i - med|}{n}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$$EM = \frac{|1-5.3| + |1-5.3| + |2-5.3| + \cdots + |9-5.3|}{10} = \frac{27}{10} = 2.7$$



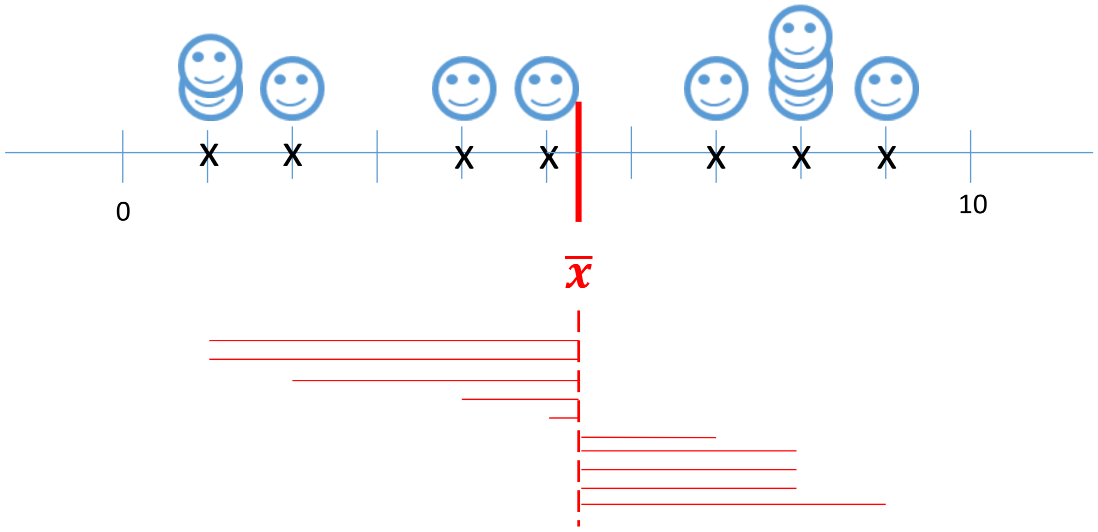
Ecart moyen et Ecart médian

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

$$EM = \frac{|1-5.3| + |1-5.3| + |2-5.3| + \cdots + |9-5.3|}{10} = \frac{27}{10} = 2.7$$

- $Ecart\ moyen = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n}|x_i - med|}{n}$



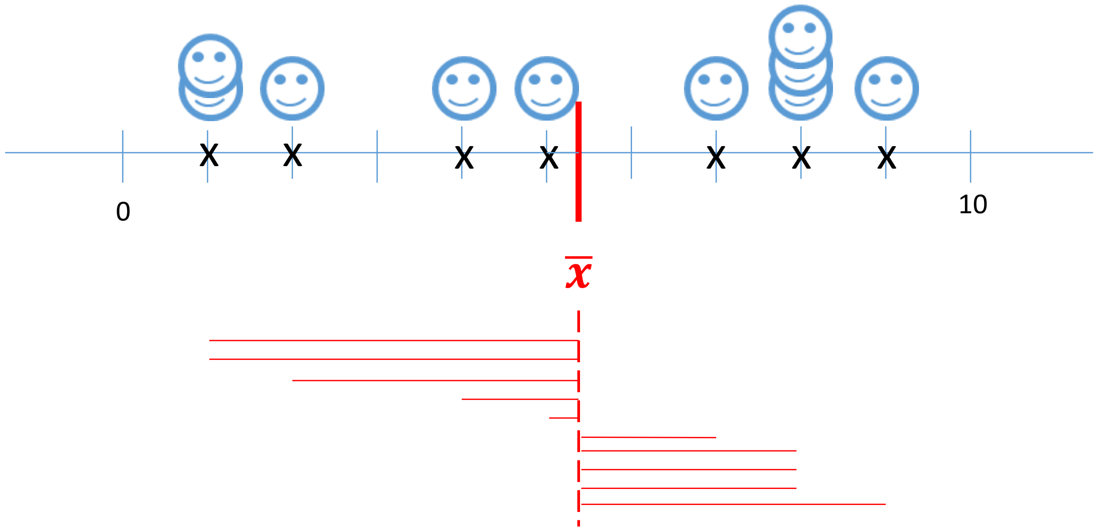
Ecart moyen et Ecart médian

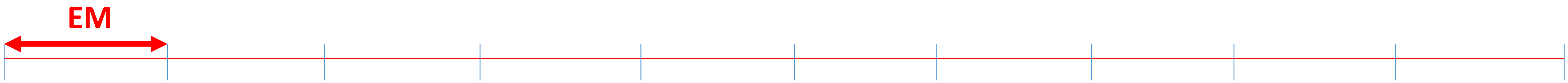
- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

- $Ecart\ moyen = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n}|x_i - med|}{n}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$$EM = \frac{|1-5.3| + |1-5.3| + |2-5.3| + \cdots + |9-5.3|}{10} = \frac{27}{10} = 2.7$$





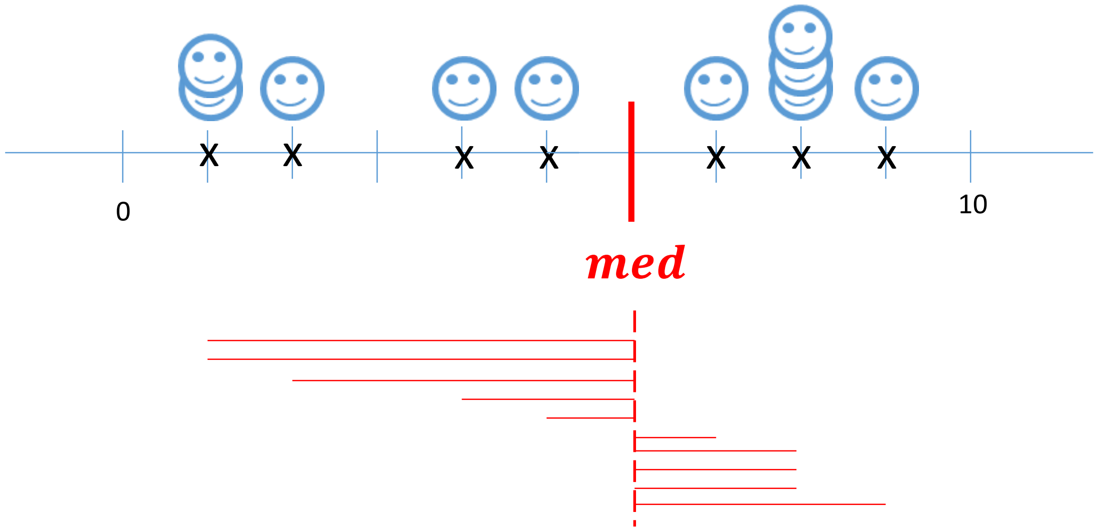
Ecart moyen et Ecart médian

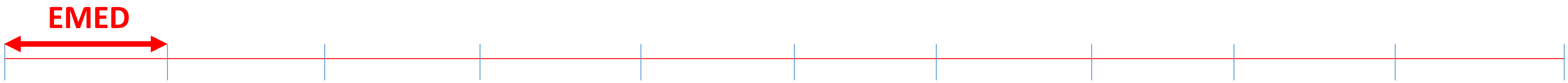
$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$med = 6$$
$$n = 10$$

- Le plus simple a comprendre
- Reflètent les distances moyens des valeurs de la moyenne/médiane

$$EMED = \frac{|1-6| + |1-6| + |2-6| + \cdots + |9-6|}{10} = \frac{27}{10} = 2.7$$

- $Ecart\ moyen = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$
- $Ecart\ médian = \frac{\sum_{i=1}^{n} |x_i - med|}{n}$





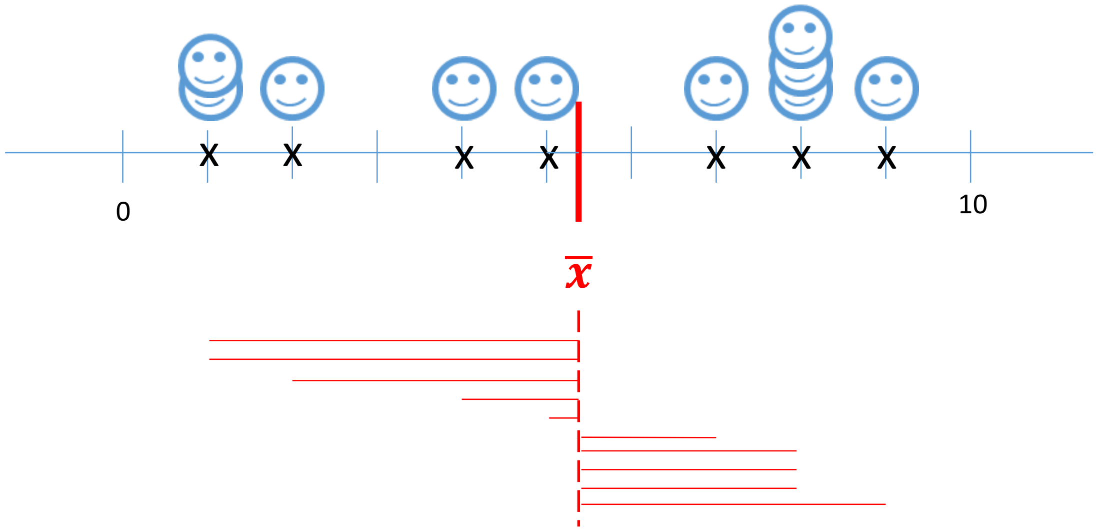
Variance

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

$$s_x{}^2 = \frac{(1-5.3)^2 + (1-5.3)^2 + (2-5.3)^2 + \cdots + (9-5.3)^2}{10-1} = \frac{88.1}{9} = 9.79$$

- **Pour population** :
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N}$
- **Pour l'échantillon** :
  - $s_x{}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$



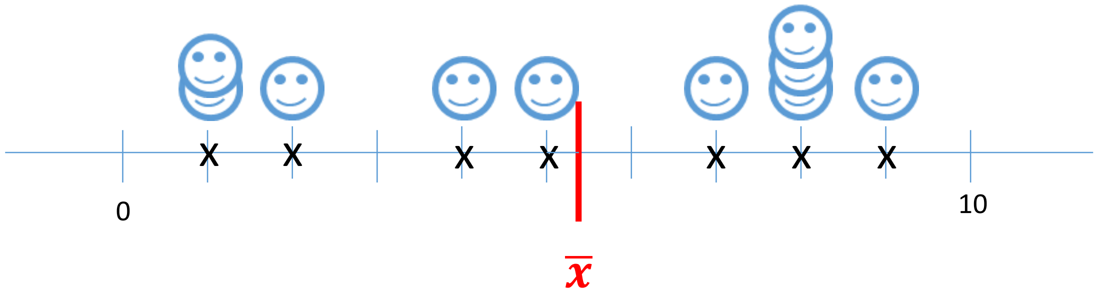
Variance

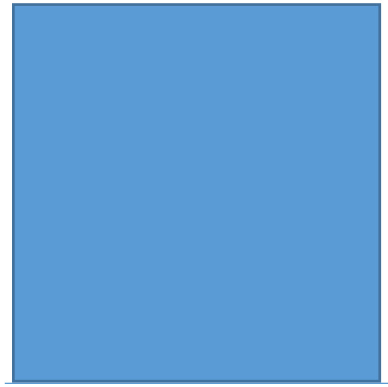
- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

- **Pour population** :
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N}$
- **Pour l'échantillon** :
  - ${s_x}^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$

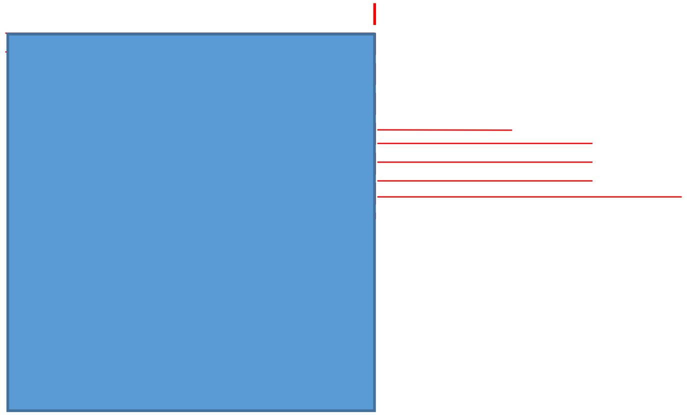
$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$${s_x}^2 = \frac{(1-5.3)^2 + (1-5.3)^2 + (2-5.3)^2 + \cdots + (9-5.3)^2}{10-1} = \frac{88.1}{9} = 9.79$$





Aire d'un carré est la valeur d'une parenthèse en carré



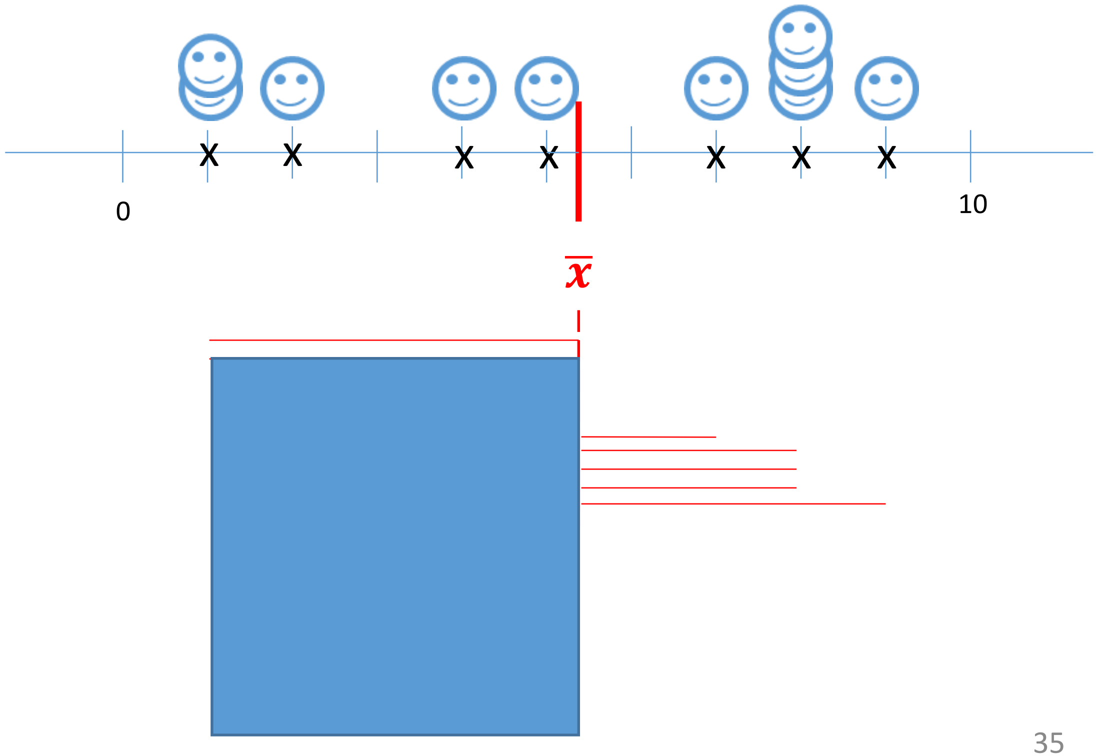
Variance

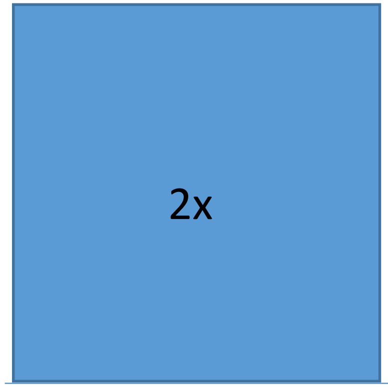
$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

$$s_x{}^2 = \frac{(1-5.3)^2 + (1-5.3)^2 + (2-5.3)^2 + \cdots + (9-5.3)^2}{10-1} = \frac{88.1}{9} = 9.79$$

- **Pour population** :
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N}$
- **Pour l'échantillon** :
  - $s_x{}^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$





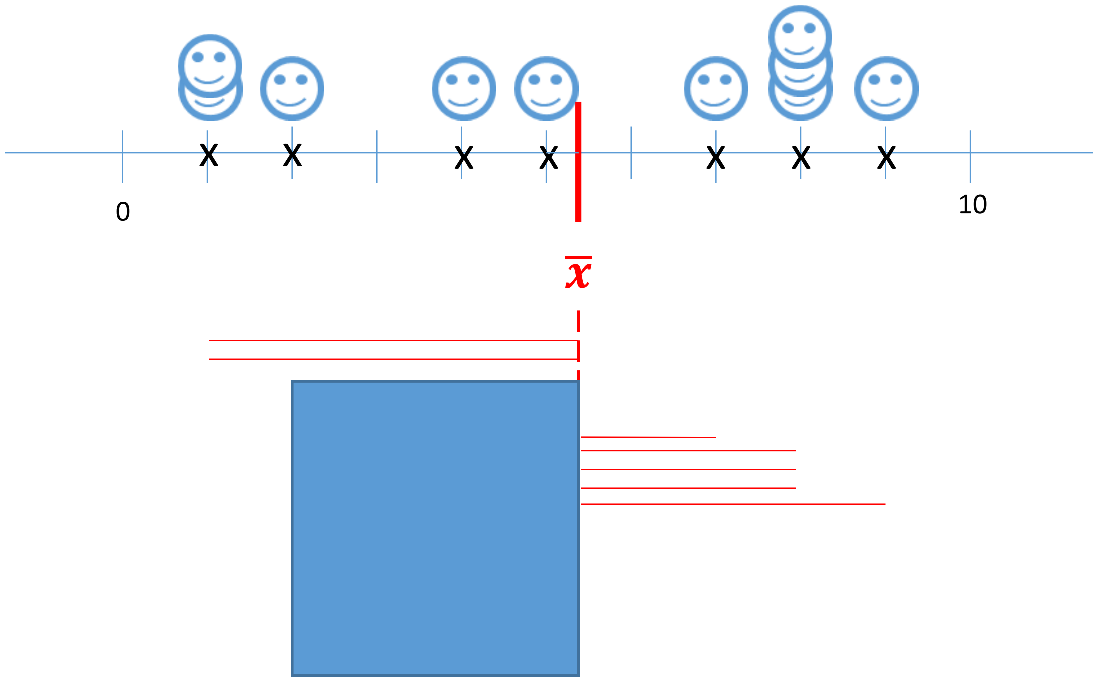
Variance

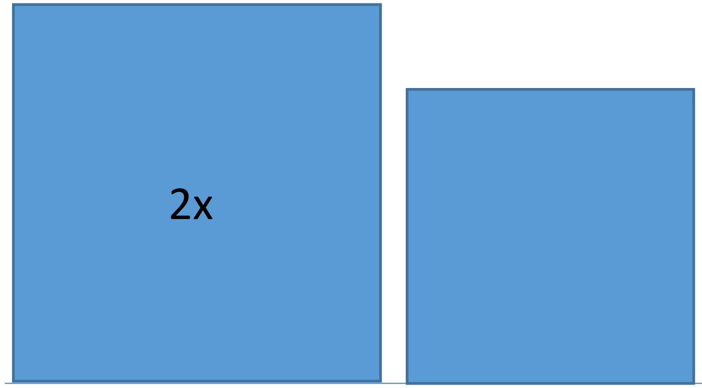
- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

- **Pour population** :
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N}$
- **Pour l'échantillon** :
  - ${s_x}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$${s_x}^2 = \frac{(1-5.3)^2 + (1-5.3)^2 + (2-5.3)^2 + \cdots + (9-5.3)^2}{10-1} = \frac{88.1}{9} = 9.79$$





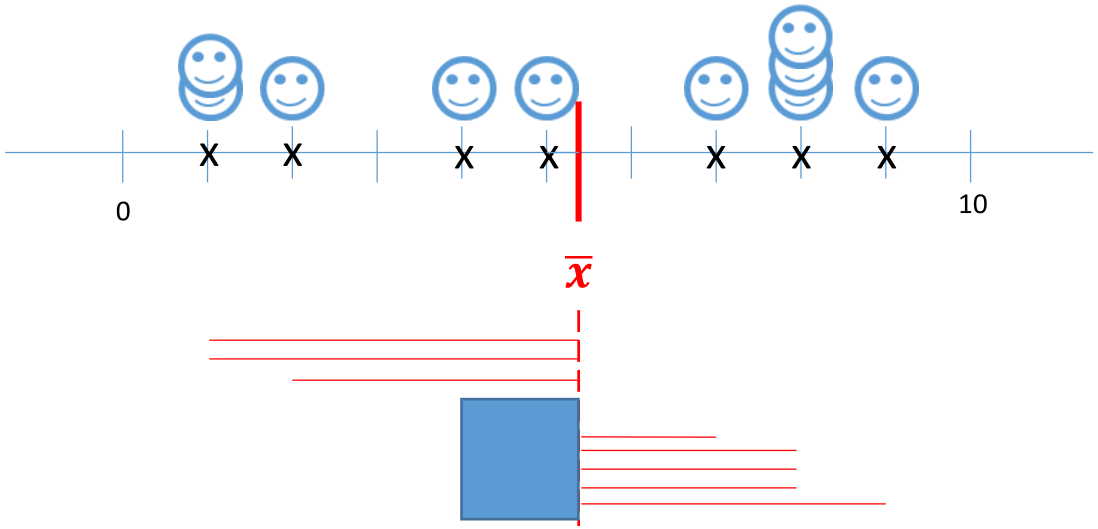
Variance

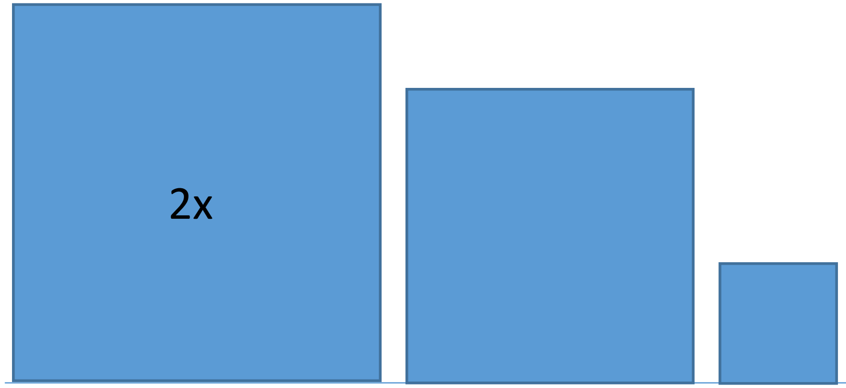
$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

$$s_x{}^2 = \frac{(1-5.3)^2 + (1-5.3)^2 + (2-5.3)^2 + \cdots + (9-5.3)^2}{10-1} = \frac{88.1}{9} = 9.79$$

- **Pour population :**
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N}$
- **Pour l'échantillon :**
  - $s_x{}^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$





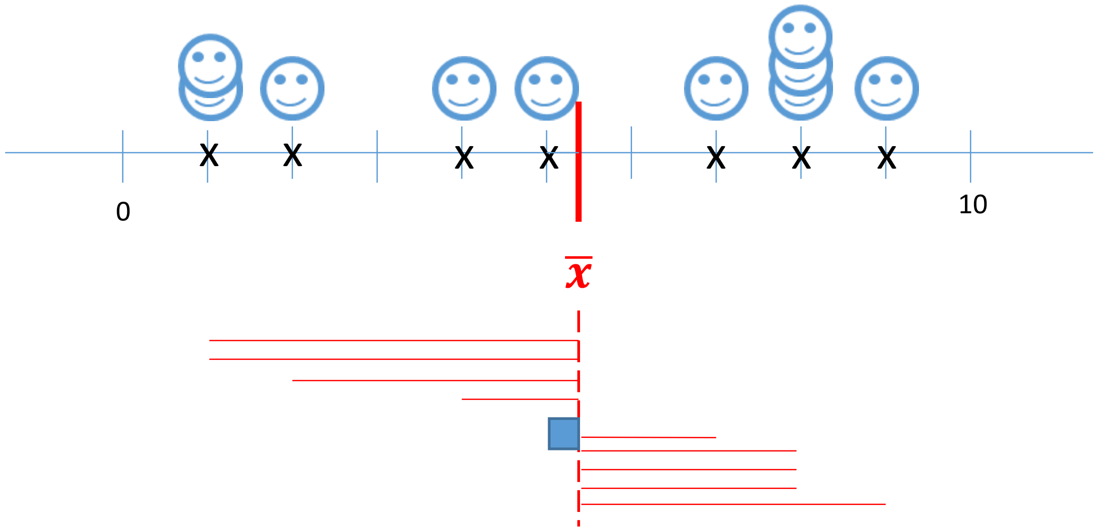
Variance

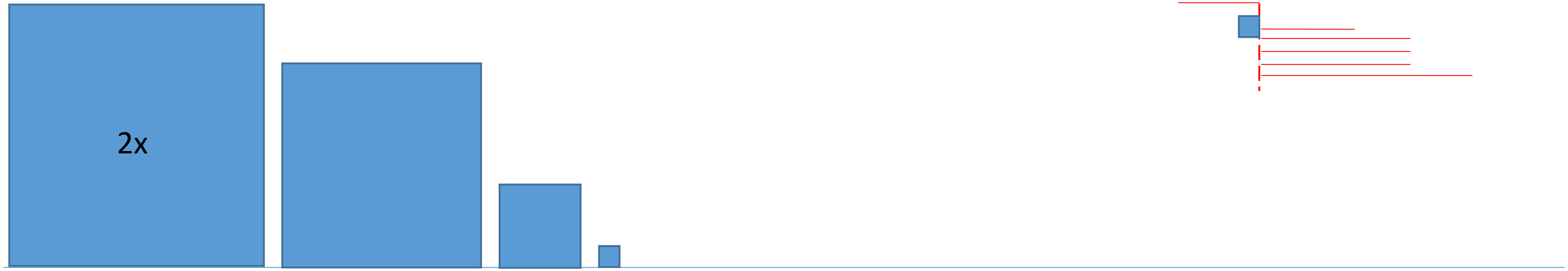
- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

- **Pour population** :
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N}$
- **Pour l'échantillon** :
  - ${s_x}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$${s_x}^2 = \frac{(1-5.3)^2 + (1-5.3)^2 + (2-5.3)^2 + \cdots + (9-5.3)^2}{10-1} = \frac{88.1}{9} = 9.79$$





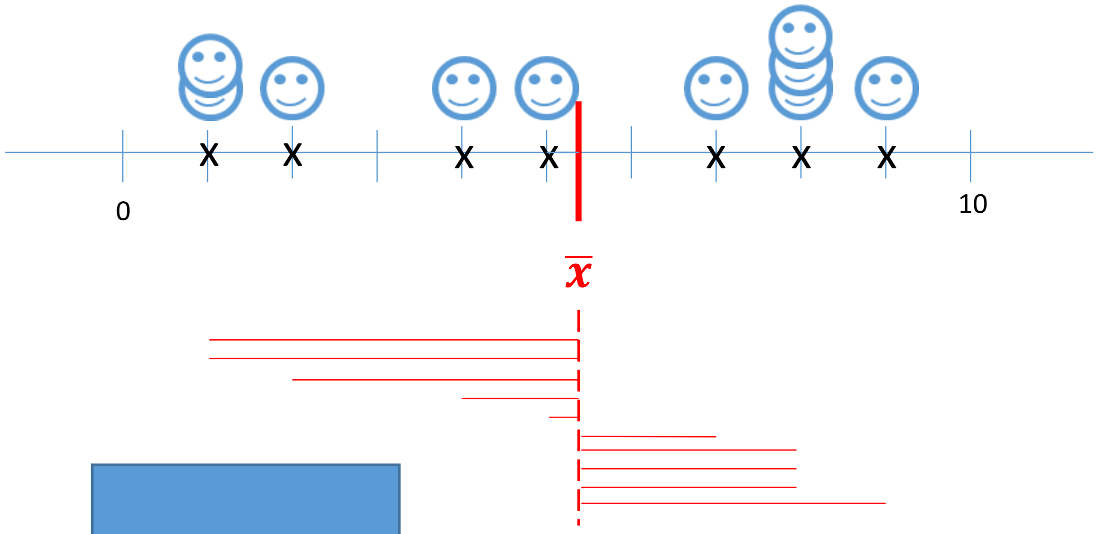
Variance

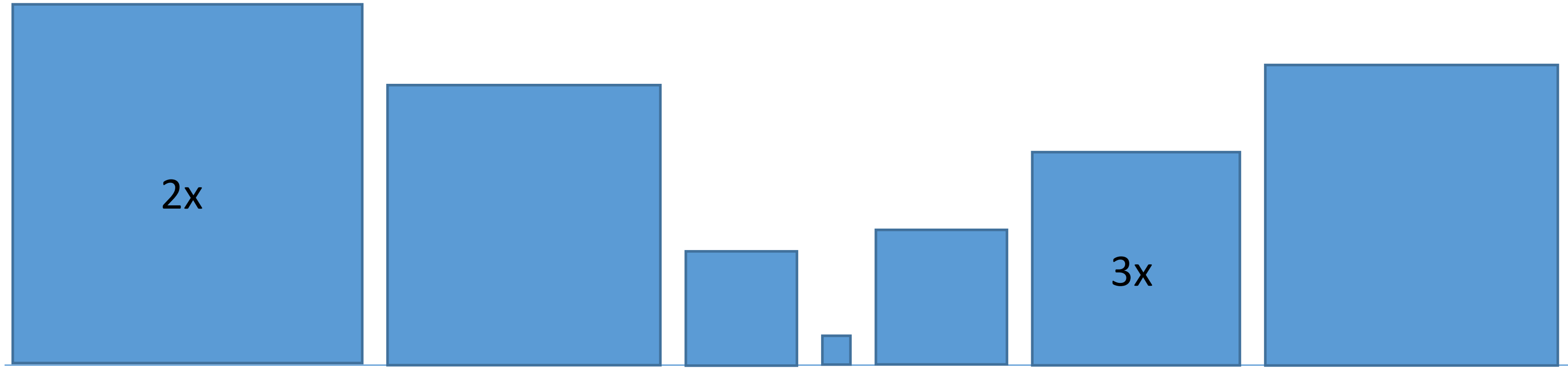
$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

$$s_x{}^2 = \frac{(1-5.3)^2 + (1-5.3)^2 + (2-5.3)^2 + \cdots + (9-5.3)^2}{10-1} = \frac{88.1}{9} = 9.79$$

- **Pour population** :
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N}$
- **Pour l'échantillon** :
  - $s_x{}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$





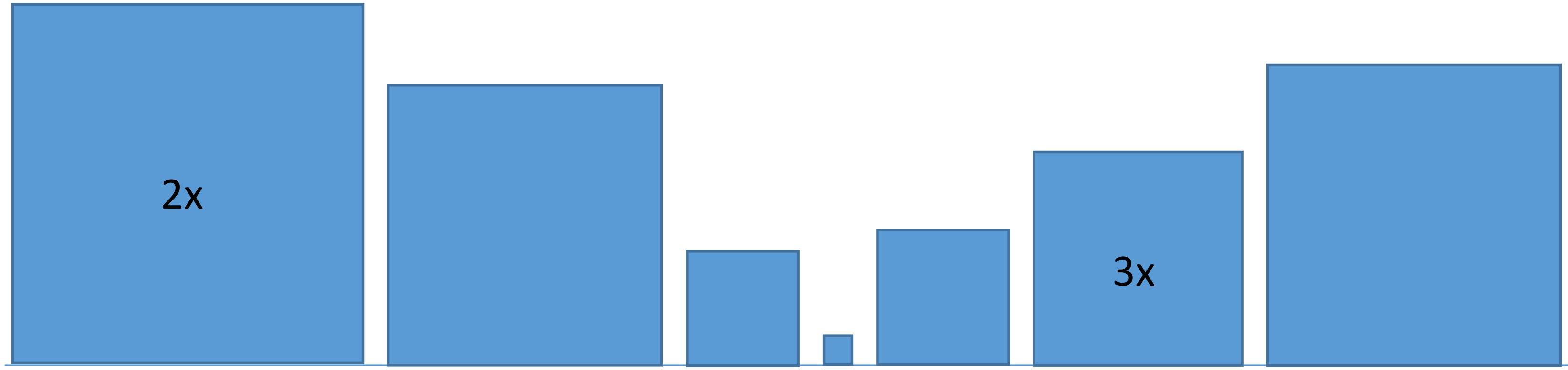
$x = 1\ 1\ 2\ 4\ 5\ 7\ 8\ 8\ 8\ 9$

Variance

- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

$s_x^{\ 2} = (1 -$ $= 9.79$

- **Pour population** :
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N}$
- **Pour l'échantillon** :
  - $s_x^{\ 2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

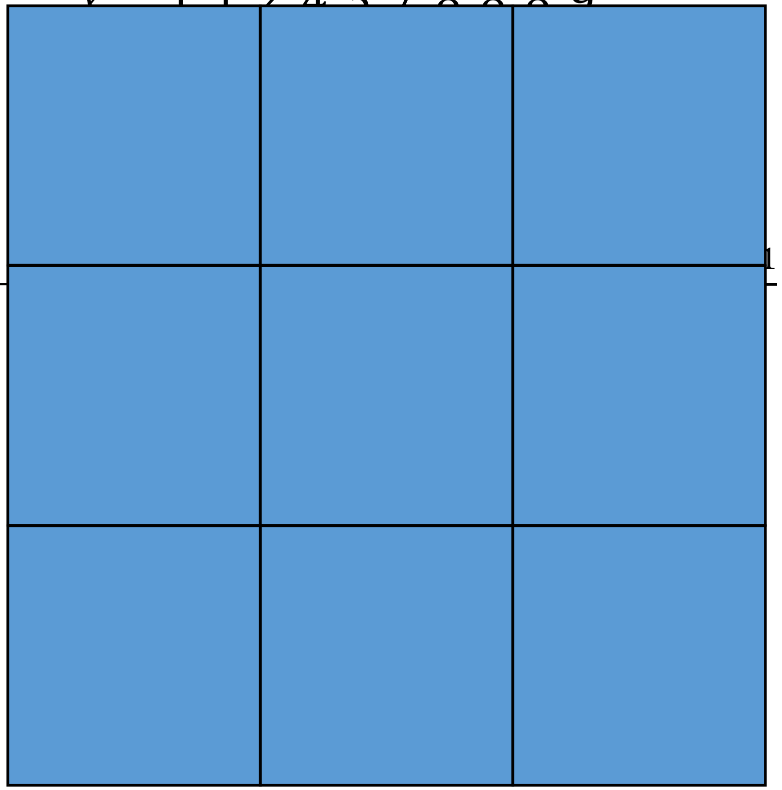


Variance

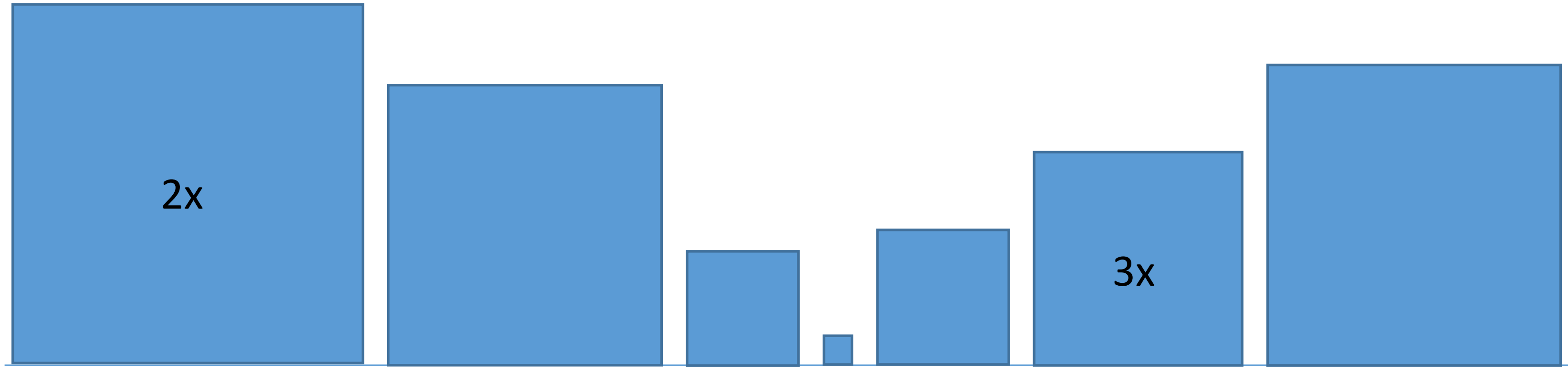
- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

- **Pour population** :
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N}$
- **Pour l'échantillon** :
  - $s_x{}^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$

$x = 1\ 1\ 2\ 4\ 5\ 7\ 8\ 8\ 8\ 9$



$s_x{}^2 = \frac{(1 - \ldots}{\ldots} = 9.79$



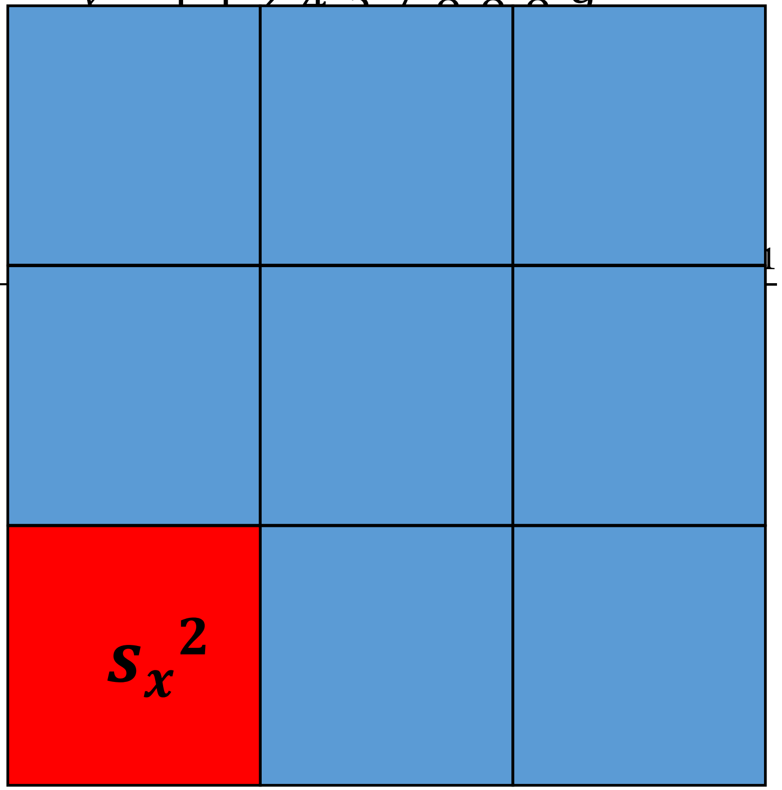
Variance

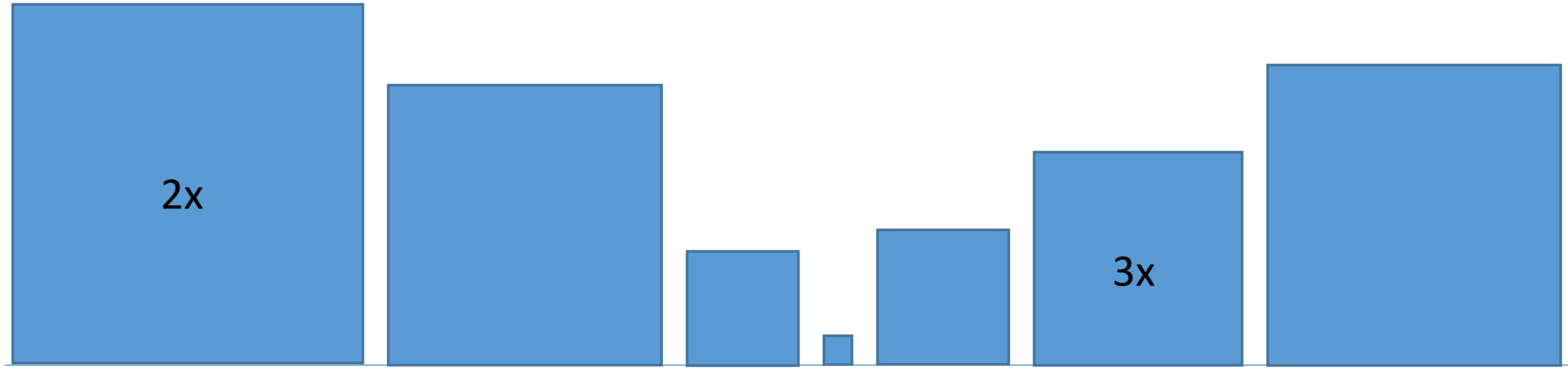
- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

- **Pour population** :
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N}$
- **Pour l'échantillon** :
  - $s_x{}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

$x = 1\ 1\ 2\ 4\ 5\ 7\ 8\ 8\ 8\ 9$

$s_x{}^2 = \frac{(1 - \ldots}{\ldots} = 9.79$





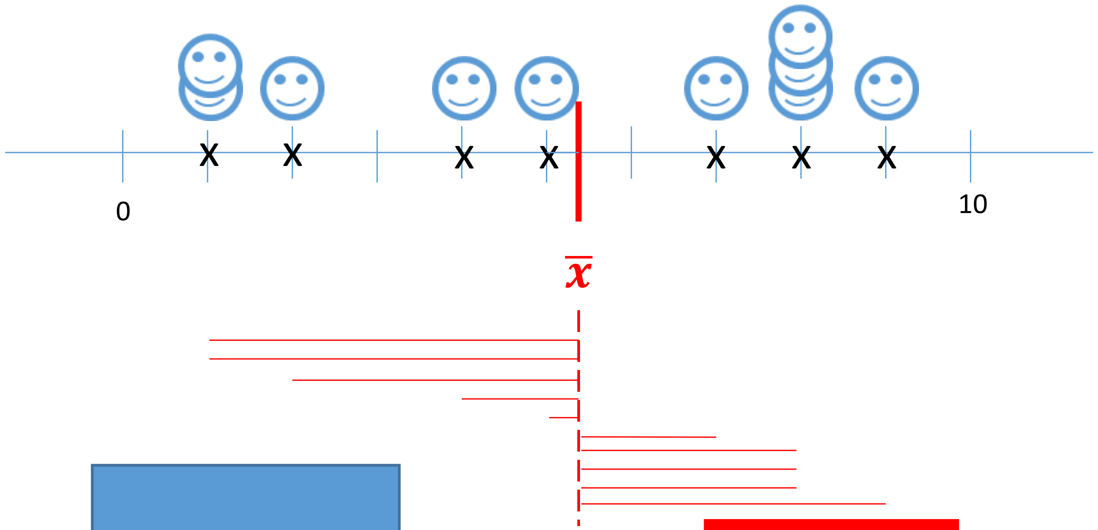
Variance

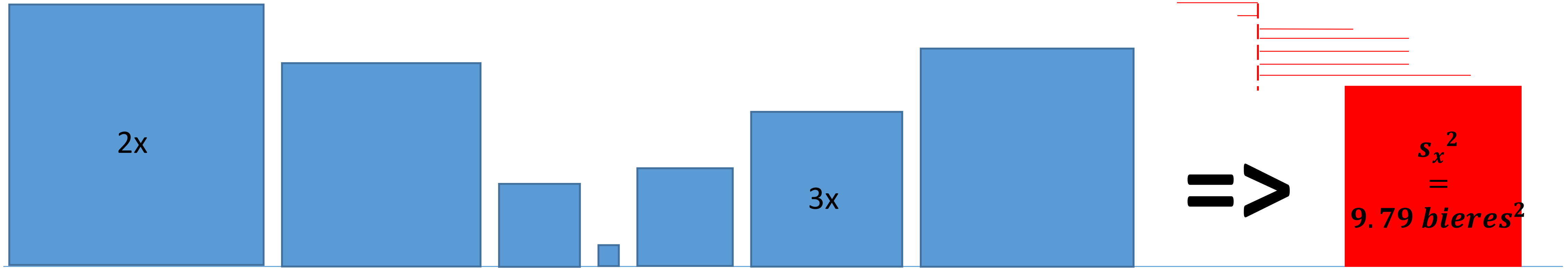
$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

- Presque la même chose que EM, mais au lieu de calculer les distances, on calcule les distances carrées
- On peut la imaginer comme le carré moyen

$$s_x{}^2 = \frac{(1-5.3)^2 + (1-5.3)^2 + (2-5.3)^2 + \cdots + (9-5.3)^2}{10-1} = \frac{88.1}{9} = 9.79$$

- **Pour population** :
  - $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{N}$
- **Pour l'échantillon** :
  - $s_x{}^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$





Ecart type

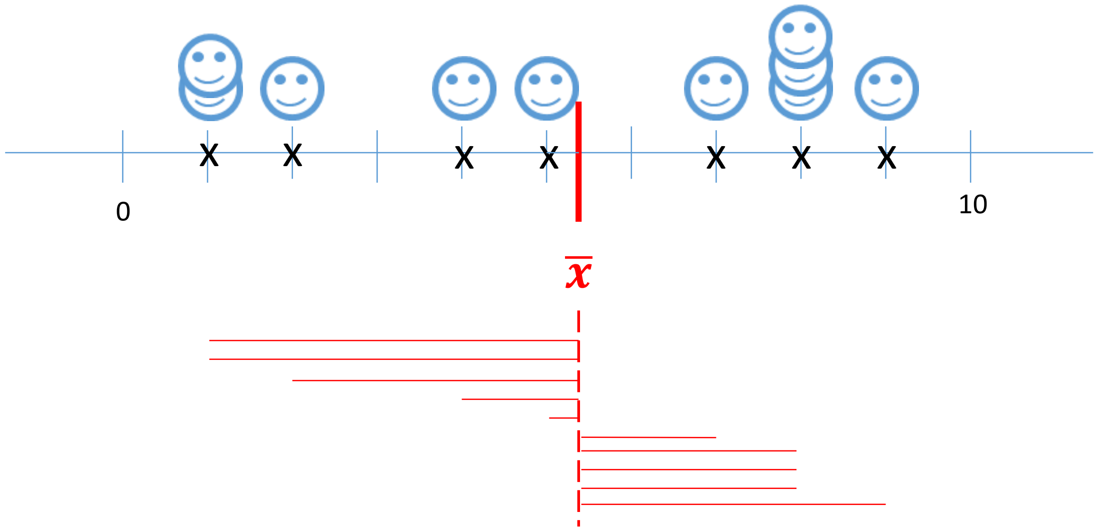
- Exprime la variabilité moyenne des donnés
- Égale a la racine carré de la variance => un coté d'un « carré moyen »

- Pour population: $\sigma = \sqrt{\sigma^2}$
- Pour échantillon: $s_x = \sqrt{{s_x}^2}$

- Plus l'écart type est grand, plus les valeurs sont éloignés (dispersées) autour de la moyenne
- Les unités sont le même que pour les données

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$${s_x}^2 = 9.79$$
$$s_x = \sqrt{9.79} = 3.13$$



Ecart type

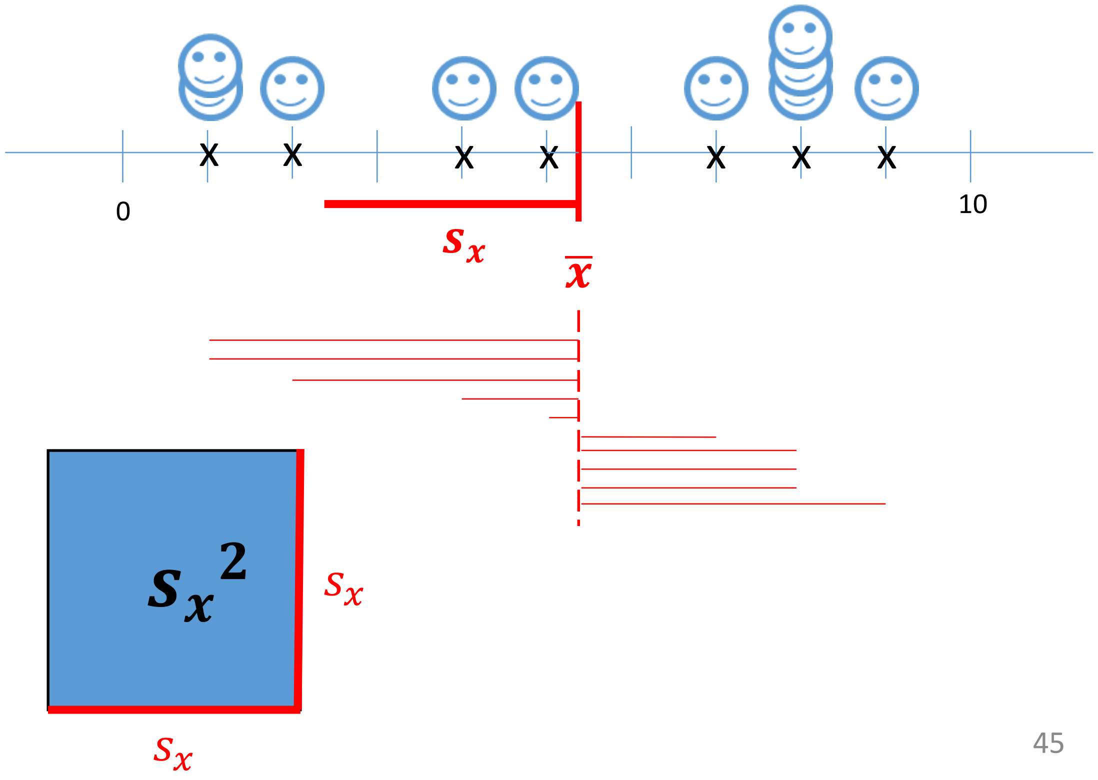
- Exprime la variabilité moyenne des donnés
- Égale a la racine carré de la variance => un coté d'un « carré moyen »

- Pour population: $\sigma = \sqrt{\sigma^2}$
- Pour échantillon: $s_x = \sqrt{{s_x}^2}$

- Plus l'écart type est grand, plus les valeurs sont éloignés (dispersées) autour de la moyenne
- Les unités sont le même que pour les données

$$x = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$\bar{x} = 5.3$$
$$n = 10$$

$${s_x}^2 = 9.79$$
$$s_x = \sqrt{9.79} = 3.13$$



Le coefficient de variation

- Pour les variables continues
- Pourquoi ? – Pour comparer les donnés
  - qui ont les moyennes très différents
  - qui sont dans les unités différents

- *Ex: Pour comparer la variance entre la taille des haches et des fibules*
- *Ex: Pour comparer la variance entre le poids des bracelets en verre et la quantité du bronze dans les épingles*

- Comment ?
- $c_v = \frac{s_x}{\overline{|x|}}$ Ratio entre l'écart type et la valeur absolue de la moyenne
- Si on veut avoir la valeur du CV en pourcentage, on multiplie par 100
- $c_v = \frac{100 * s_x}{\overline{|x|}}$ (%)

Comparaison de la variabilité des bières et la taille des haches.

Le coefficient de variation

- Pour les variables continues
- Pourquoi ? – Pour comparer les donnés
  - qui ont les moyennes très différents
  - qui sont dans les unités différents

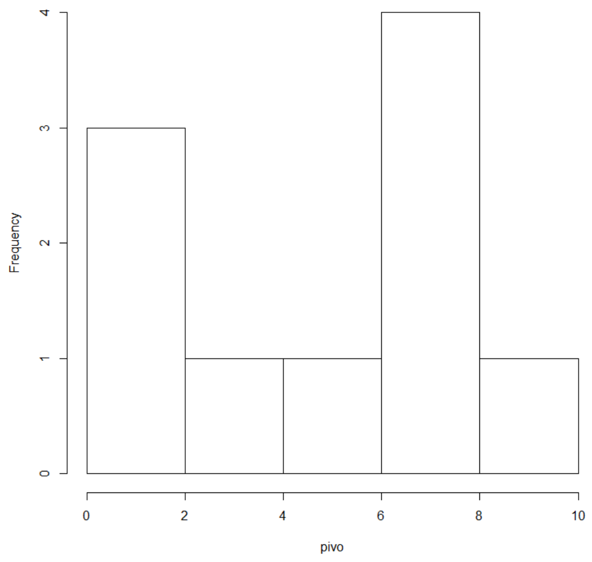
- *Ex: Pour comparer la variance entre la taille des haches et des fibules*
- *Ex: Pour comparer la variance entre le poids des bracelets en verre et la quantité du bronze dans les épingles*

- Comment ?
- $c_v = \frac{s_x}{|x|}$ Ratio entre l'écart type et la valeur absolue de la moyenne
- Si on veut avoir la valeur du CV en pourcentage, on multiplie par 100
- $c_v = \frac{100 * s_x}{|x|}$ (%)

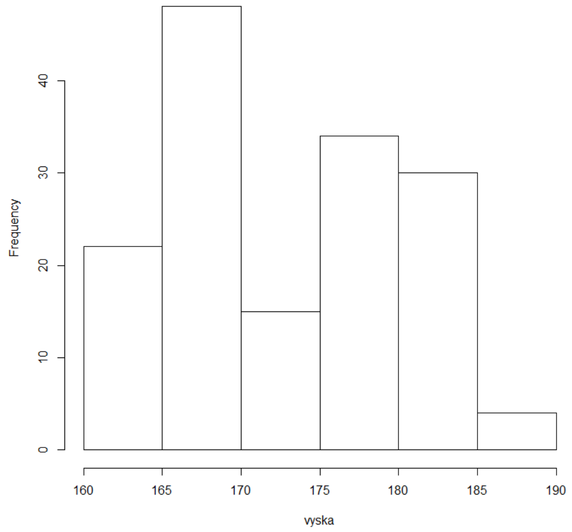
$$x(biere) = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$n = 10$$
$$\bar{x} = 5.3\ biere$$
$$s_x(biere) = 3.13\ biere$$

$$x(taille) = 164, 169, 168, 164, 167, \ldots, 169$$
$$n = 153$$
$$\bar{x} = 173.5\ cm$$
$$s_x(taille) = 51.5\ cm$$

Histogramme des bières



Histogramme des tailles des haches



Le coefficient de variation

- Pour les variables continues
- Pourquoi ? – Pour comparer les donnés
  - qui ont les moyennes très différents
  - qui sont dans les unités différents

- *Ex: Pour comparer la variance entre la taille des haches et des fibules*
- *Ex: Pour comparer la variance entre le poids des bracelets en verre et la quantité du bronze dans les épingles*

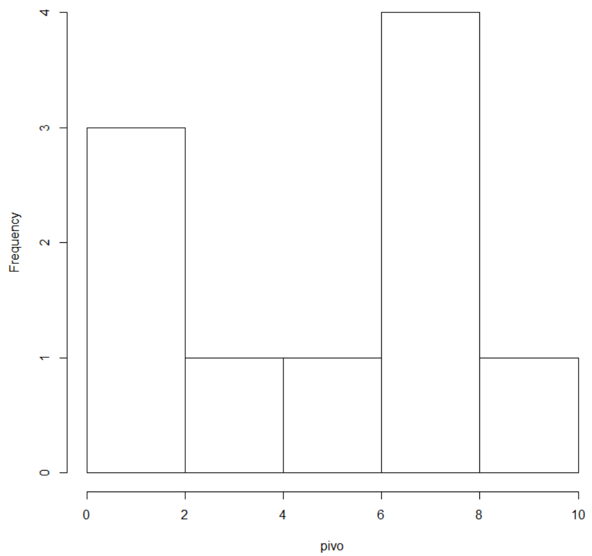
- Comment ?
- $c_v = \frac{s_x}{|x|}$ Ratio entre l’écart type et la valeur absolue de la moyenne
- Si on veut avoir la valeur du CV en pourcentage, on multiplie par 100
- $c_v = \frac{100 * s_x}{|\bar{x}|}$ (%)

Comparaison de la variabilité des bières et la taille des haches.

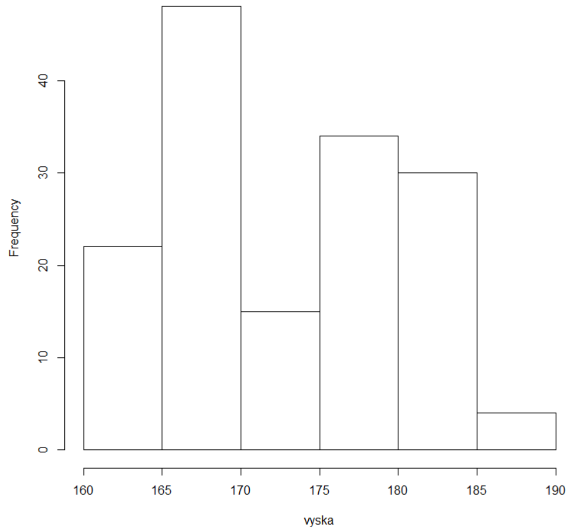
$$x(biere) = 1, 1, 2, 4, 5, 7, 8, 8, 8, 9$$
$$n = 10$$
$$\bar{x} = 5.3\ biere$$
$$s_x(biere) = 3.13\ biere$$

$$x(taille) = 164, 169, 168, 164, 167, \ldots, 169$$
$$n = 153$$
$$\bar{x} = 173.5\ cm$$
$$s_x(taille) = 51.5\ cm$$

Histogramme des bières



Histogramme des tailles des haches



$$c_v(biere) = \frac{100 * 3.13}{5.3} = 59.1\%$$

$$c_v(taille) = \frac{100 * 51.5}{153} = 29.7\%$$

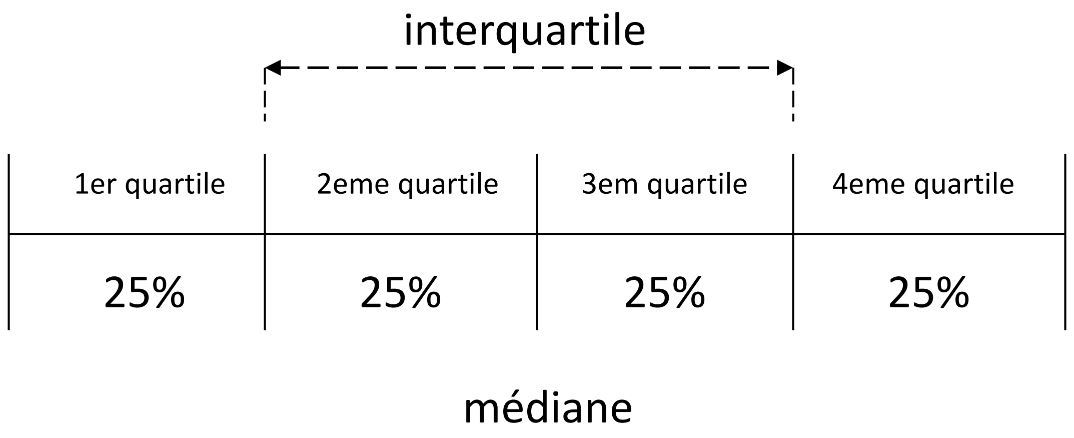
La variabilité des bières est plus élevé que celle de la taille des haches.

Quantiles

- Valeurs, qui séparent les donnés sur quelques parties qui sont plus ou moins égales
- Pour variables quantitatives continus et discrètes
- Pourquoi ?
    - On peut voir où les données se concentrent
    - Pour trouver intervalle comprenant 50 % des observations les plus au centre de la distribution

- *Ex: On peut trouver que 50% des volumes des tonneaux le plus caractéristiques se trouvent entre 1.3 et 1.7 L.*
- *Ex: La plupart des fibules ont les spirales avec le diamètre entre 1 et 2 cm.*

- D’après le nombre des parties, on peut distinguer :
    - Quartiles : Séparent la distribution sur 4 parties égales
    - Quintiles : 5
    - Déciles : 10
    - Centiles : 100
    - Médiane : 2



Quantiles pour les variables continus

- **Comment?**
- Ordonner les valeurs de plus petit au plus grand
- Appliquer : $\frac{n \cdot p}{100} \leq \boldsymbol{z} \leq \left(\frac{n \cdot p}{100} + 1\right)$
  - où le nombre $\boldsymbol{z}$ donne l'indice du quantile avec $\boldsymbol{p} -$ pourcents

Quantiles pour les variables continus

Trouvez les quartiles pour la taille des haches.

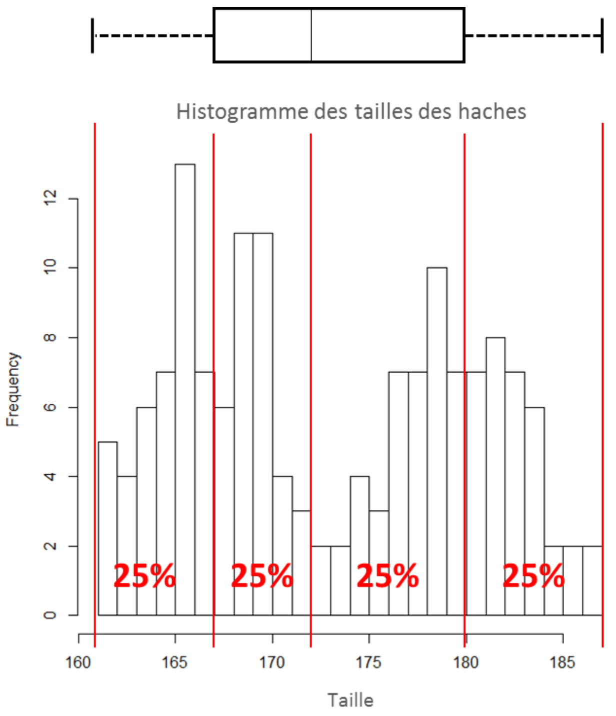
$$x = 164, 169, 168, 164, 167, \dots, 169$$
$$n = 153$$

C:\Users\josef\Desktop\L3 - BDD – 2016\Haches.xlsx

- **Comment?**
- Ordonner les valeurs de plus petit au plus grand
- Appliquer : $\frac{n \cdot p}{100} \leq \boldsymbol{z} \leq \left(\frac{n \cdot p}{100} + 1\right)$
  - où le nombre $\boldsymbol{z}$ donne l'indice du quantile avec $\boldsymbol{p}$ − pourcents

Quantiles pour les variables continus

- **Comment?**
- Ordonner les valeurs de plus petit au plus grand
- Appliquer : $\frac{n \cdot p}{100} \leq \boldsymbol{z} \leq \left(\frac{n \cdot p}{100} + 1\right)$
  - où le nombre $\boldsymbol{z}$ donne l'indice du quantile avec $\boldsymbol{p}$ – pourcents



Trouvez les quartiles pour la taille des haches.

$$x = 164, 169, 168, 164, 167, \ldots, 169$$
$$n = 153$$

C:\Users\josef\Desktop\L3 - BDD – 2016\Haches.xlsx

$$x(ord) = 159, 160, 160, 161, \ldots, 186$$

$$\frac{n \cdot p}{100} \leq \boldsymbol{z} \leq \left(\frac{n \cdot p}{100} + 1\right)$$

$\boldsymbol{p = 25\%}$	$\boldsymbol{p = 50\%}$	$\boldsymbol{p = 75\%}$
$\frac{153 * 25}{100} \leq z \leq \frac{153 * 25}{100} + 1$	$\frac{153 * 50}{100} \leq z \leq \frac{153 * 50}{100} + 1$	$\frac{153 * 75}{100} \leq z \leq \frac{153 * 75}{100} + 1$
$38.25 \leq z \leq 39.25$	$76.5 \leq z \leq 77.5$	$114.75 \leq z \leq 115.75$
$z = 39$	$z = 77$	$z = 115$
$x[39] = 167$	$x[77] = 170$	$x[75] = 179$