

**La moyenne (arithmétique) = "mean"**

**Exemple: 7 données:** 99,8 101,7 102,2 104,0 97,4 96,5 102,2

**Moyenne = somme des données divisée par leur nombre**

**$n$  données notées  $x_1 = 99,8, x_2 = 101,7, \dots, x_n = 102,2$**

$$\bar{x} = \frac{99,8 + 101,7 + 102,2 + 104,0 + 97,4 + 96,5 + 102,2}{7} = 100,54$$

**définition de la moyenne:**

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

### **Propriétés:**

**1.**  $n\bar{x} = x_1 + x_2 + \dots + x_n$

**2. la somme des écarts à la moyenne vaut 0:**  $\sum_{i=1}^n (x_i - \bar{x}) = 0$

**Calcul: changer d'origine (100) et d'unité (0,1)**

$$\bar{x} = 100 + 0,1 \frac{(-2) + (17) + (22) + (40) + (-26) + (-35) + (22)}{7}$$

$$= 100 + 0,1 \times 5,4 = 100,54$$

**Inconvénient: la moyenne est affectée par des données aberrantes**

**Donnée aberrante 202,2 au lieu de 102,2  $\Rightarrow$  moyenne = 114,8**

**Version pondérée: si poids  $w_i$ , alors**

$$\bar{x} = \frac{w_1x_1 + w_2x_2 + \dots + w_nx_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum_{i=1}^n w_ix_i}{\sum_{i=1}^n w_i}$$

**Exemple: moyenne de moyennes  $\Rightarrow$  pondérer par les effectifs**

## La médiane ("median")

**A déterminer à partir de la suite *ordonnée* des  $n$  données**

**Exemple:** 99,8 101,7 102,2 104,0 97,4 96,5 102,2

**Tri**  $\Rightarrow$  96,5 97,4 99,8 101,7 102,2 102,2 104,0

**Si  $n$  est impair, prendre la donnée centrale (autant à gauche qu'à droite)**

**Si  $n$  est pair, prendre la moyenne des deux données centrales**

**7 = impair  $\Rightarrow$  4e donnée = 101,7**

96,5	97,4	99,8	101,7	102,2	102,2	104,0
------	------	------	-------	-------	-------	-------

**Variante si 6 données:** 99,8 101,7 102,2 104,0 97,4 96,5

**Tri**  $\Rightarrow$  96,5 97,4 99,8 101,7 102,2 104,0

**6 = pair  $\Rightarrow$  moyenne des 3e et 4e données =**

$$\frac{99,8 + 101,7}{2} = 100,08$$

96,5      **97,4**      99,8      101,7      102,2      104,0

**Peu sensible aux données aberrantes**

**Exemple: Donnée aberrante 202,2 au lieu de 102,2  $\Rightarrow$**

99,8   101,7   102,2   104,0   97,4   96,5   202,2

**même médiane (moyenne = 114,8)**

## **Quantiles (fractiles) d'ordre $p$ ou $100p$ %**

**Partir de la fonction de distribution**

**En ordonnée:  $p$**

**En abscisse, on trouve  $x_p$ , le quantile d'ordre  $p$**

**Si  $p = 50$  % : médiane**

**Courbe lisse  $\Rightarrow$  pas de problème, 1 solution**

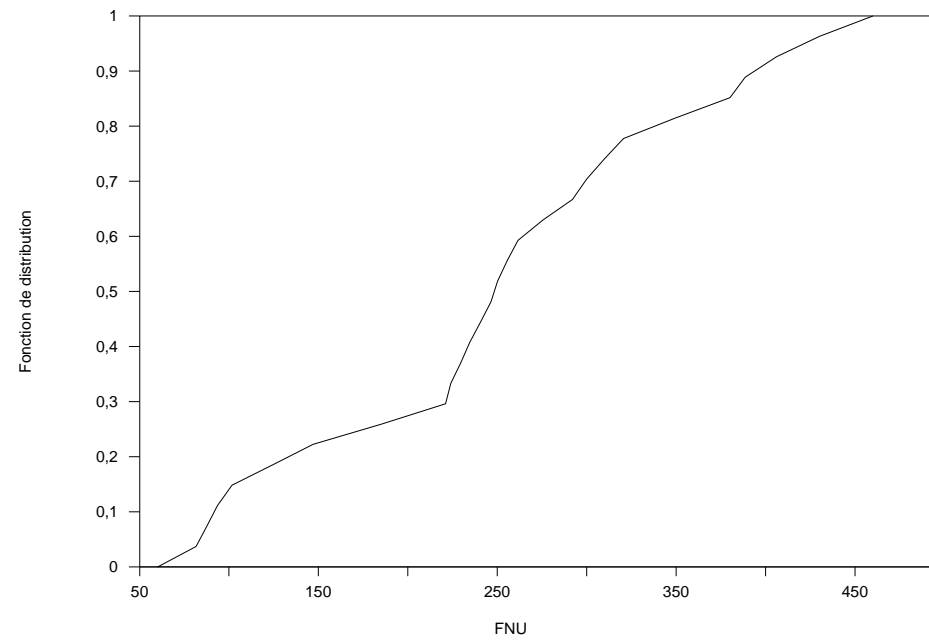
**Courbe en escalier:**

**si contremarche, descendre**

**si marche, prendre le milieu de la marche et descendre**

**Si  $p = 25\%$ : 1er quartile**

**Si  $p = 75\%$ : 3e quartile**



**Propriété: soient  $p_{\text{inf}} < p_{\text{sup}}$**

**on détermine  $x_{p_{\text{inf}}} < x_{p_{\text{sup}}}$**

**l'intervalle de  $x_{p_{\text{inf}}}$  à  $x_{p_{\text{sup}}}$  contient la proportion  $p_{\text{sup}} - p_{\text{inf}}$  des données**

**Exemple: intervalle à 95 %: basé sur  $x_{0,025}$  et  $x_{0,975}$**



## Variance et écart-type ("standard deviation")

**Exemple:** 99,8 101,7 102,2 104,0 97,4 96,5 102,2  $\Rightarrow \bar{x} = 100,54$

**Calculer les écarts à la moyenne**

Donnée $x_i$	Ecart à $\bar{x}$ $x_i - \bar{x}$	Ecart au carré $(x_i - \bar{x})^2$	Ecart à 100 $u_i = x_i - 100$	Ecart à 100 au carré $u_i^2 = (x_i - 100)^2$
99,80	-0,74	0,5476	-0,2	0,04
101,70	1,16	1,3456	1,7	2,89
102,20	1,66	2,7556	2,2	4,84
104,00	3,46	11,9716	4,0	16,00
97,40	-3,14	9,8596	-2,6	6,76
96,50	-4,04	16,3216	-3,5	12,25
102,20	1,66	2,7556	2,2	4,84
<b>Somme</b>	<b>0,02</b>	<b>45,5572</b>	<b>3,8</b>	<b>47,62</b>

## Variance

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \Rightarrow s^2 = \frac{45,5572}{7} = 6,51$$

$\Rightarrow$

**Autrement (changement d'origine ici à 100)**

$$s^2 = \frac{\sum_{i=1}^n u_i^2}{n} - \bar{u}^2 = \frac{47,62}{7} - \left(3, \frac{8}{7}\right)^2 = 6,803 - (0,543)^2 = 6,803 - 0,295 = 6,51$$

**Ecart-type:  $s$  = racine carrée de la variance =  $\sqrt{6,51} = 2,55$**

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

**Avantage: exprimé dans les mêmes unités que les données, donc que la moyenne**

**Idee de dispersion**

**mais pénalisation des grands écarts**

**Cas particulier: si toutes les données égalent la moyenne  $\Rightarrow s^2 = 0$  et  $s = 0$**

## Remarques:

1.  $s^2 < 0$  est impossible
2.  $s$  est indiqué par  $\sigma$  ou  $\sigma_n$  sur les calculatrices [c'est la variance "population"]
3. On voit aussi (anglo-saxon)  $s$  calculé différemment, indiqué par  $\sigma_{n-1}$ , noté ici  $S$

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{45,5572}{6} = 7,59 \Rightarrow S = 2,75$$

c'est la variance "échantillon" (explication plus tard)

## 4. Le calcul de la variance n'est pas toujours correct:

**exemple:** 10 000 000, 10 000 001, 10 000 002  $\Rightarrow \bar{x} = 10000001$  et

$$s^2 = \frac{(-1)^2 + (0)^2 + (1)^2}{3} = \frac{2}{3} = ? \frac{(10000000)^2 + (10000001)^2 + (10000002)^2}{3} - (10000001)^2$$

**En général NON** à cause du nombre de chiffres (sauf logiciels bien faits)

## **Autres caractéristiques d'une distribution**

### ***Autres mesures de dispersion:***

**L'étendue ("range"): maximum - minimum [très sensible aux données aberrantes]**

**L'étendue interquartile *EI* ("interquantile range"):**

**3e quartile - 1er quartile [couvre 50 % de la distribution]**

***Asymétrie* ("skewness"): mesuré par skewness [très sensible aux données aberrantes]**

**symétrie  $\Rightarrow 0$**

**dissymétrie à gauche  $\Rightarrow >0$**

**dissymétrie à droite  $\Rightarrow <0$**

**dissymétrie autre**

***Aplatissement* ("kurtosis"): mesuré par kurtosis [extrêmement sensible aux données aberrantes]**

**référence à la loi normale**

**plus effilée à queues épaisses  $\Rightarrow >0$**

**aplatie  $\Rightarrow <0$**

## Exemple de SAS Univariate

```

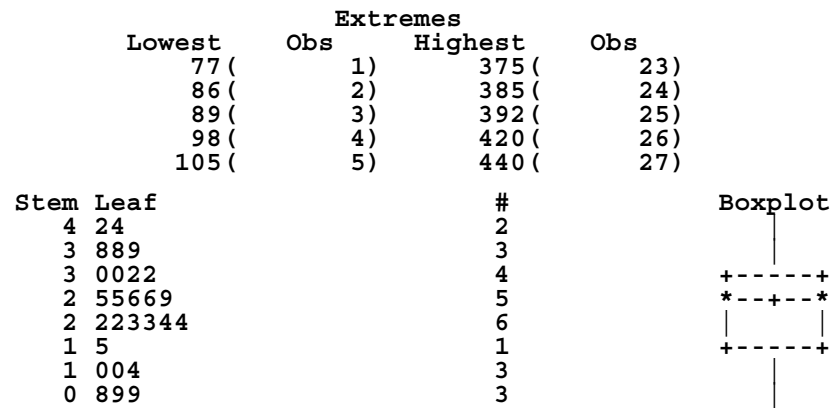
----- ENZYMES DANS LA POLYARTHRITE -----
(...) GROUP=1 -----
----- GROUP=2 -----
Univariate Procedure
Variable=FNU

              Moments
N              27  Sum Wgts              27
Mean          247.9259  Sum              6694
Std Dev       102.9951  Variance       10607.99
Skewness      -0.02813  Kurtosis      -0.50856
USS           1946032  CSS           286415.9
CV            41.5427  Std Mean      19.82142
T:Mean=0      12.50798  Prob>|T|          0.0001
Num ^ = 0      27      Num > 0          27
M(Sign)        13.5    Prob>|M|          0.0001
Sgn Rank       189    Prob>|S|          0.0001
W:Normal       0.950164  Prob<W           0.2364

              Quantiles (Def=5)
100% Max       440      99%           440
 75% Q3        317      95%           420
 50% Med       248      90%           392
 25% Q1        150      10%           89
  0% Min       77       5%           86
                   1%           77

Range          363
Q3-Q1          167
Mode           77

```



-----+  
 Multiply Stem.Leaf by 10\*\*+2

