1 Les caractéristiques de tendance centrale

Les mesures de tendance centrale sont des valeurs autour desquelles gravitent les différentes valeurs de la série. Elles nous donnent une idée sur l'ordre de grandeur des données

1.1 La moyenne arithmétique

La moyenne arithmétique, que l'on note $\bar{x}$, est la somme des valeurs prises par la variable statistique, divisée par le nombre d'observations. Cette moyenne est dite simple par opposition à la moyenne pondérée par les effectifs correspondant à chaque valeur de la variable statistique

$$\bar{x} = \frac{\sum_{i=1}^{n} xini}{n}$$

1.1.1 La moyenne arithmétique d'une variable quantitative discrète

$$\bar{x} = \frac{\sum_{i=1}^{k} xini}{n}$$

xi : esl la i^{éme} valeur distincle

ni est l'effectif de la i^{éme} valeur distincte

k est l'effectif total

n est le nombre total de valeurs distinctes

Exemple

Soit la distribution suivante, qui représente la répartition de 50 portes de souris selon le nombre de nouveau-nés

Nombre de nouveu-nés (xi)	Nombre de portées (ni)	$xi\ ni$
0 1 2 3 4 5 6 7	5 8 11 7 9 5 4 1	0 8 22 21 36 25 24 7
	n =50	$\sum_{i=1}^{k} xini = 143$

$\bar{x}$ =143/50= 2.86 souris blanches

1.1.2 La moyenne arithmétique d'une variable quantitative continue

$$\bar{x} = \frac{\sum_{i=1}^{k} xini}{n}$$

xi : est le centre de la i^{éme} classe

ni est l'effectif de la i^{éme} classe

k est le nombre de classe

n est l'effectif total

Exemple

Le tableau ci-dessous représente la répartition des 219 salariés selon leurs salaires bruts mensuels

Salaire brut (10^3 Da)	Effectif (ni)	Centre de classe (xi)	$xini$
4-5	12	4.5	54
5-6	23	5.5	126.5
6-7	42	6.5	273
7-9	56	8.0	448
9-11	34	10.0	340
11-15	32	13.0	416
15-23	16	19.0	304
23-31	4	27 .0	108
	$n = 219$		$\sum_{i=1}^{k} xini = 2069.5$

$\bar{x}$ =2069.5/219= 9.45^{x} 10^3 Da

Quelques remarques sur la moyenne arithmétique

- La moyenne arithmétique ne divise pas nécessairement la distribution en deux parties d'effectifs égaux. De plus, la présence des valeurs extrêmes aberrantes peut avoir une influence marquée sur la valeur de la moyenne
- La moyenne arithmétique est une mesure de tendance centrale la plus représentative d'une serie statistique sauf dans deux cas :
  - Lorsque la distribution est bimodale ou multimodale
  - Lorsque la série comporte des valeurs extrêmes aberrantes. La présence de quelques valeurs exceptionnellement grandes ou petites dans la série statistique à un effet sur la moyenn

1.2 La médiane

La médiane, que l'on note Me, est la valeur qui partage la série statistique ordonnée en deux parties égales, chacune comprenant le même nombre d'observations

1.2.1 La médiane d'une variable quantitative discrète

Pour calculer la médiane d'une variable quantitative discrete, on procéde comme suit :

- Préciser les fréquences cumulées croissantes ou décroissantes
- Déterminer l'observation qui divise la série statistique en deux parties égales
- Chercher dans le tableau de distribution à quelle valeur de la variable, cette observation appartient elle

Exemple

Les résultats d'une enquête donnent la répartition de 500 familles selon le nombre d'enfants

Nombre d'enfants	effectifs	FCAC	FCAD	FCRC	FCRD
0	50	50	500	0.1	1
1	120	170	450	0.34	0.90
2	130	290	330	0.60	0.66
3	110	410	200	0.82	0.40
4	90	500	90	1	0.1
TOTAL	500				

La médiane de cette distribution est :

- n/2= 500/2= 250^{éme} observation, donc la 250^{éme} observation elle divise la distribution en deux parties égales
- dans la colonne des FCAC, on trouve la 250^{éme} observation appartient à la valeur 2 enfants
- donc Me= 2 enfants par famille

Il y'a autant de famille ayant moins de 2 enfants que de famille ayant plus de 2 enfants

1.2.2 La médiane d'une variable quantitative continue

On obtient la médiane d'une variable quantitative continue en procédant comme suit :

1- trouver la classe médiane
2- on associe ensuite la médiane à la valeur obtenue par l'application de la formule suivante

$$\text{Me}= li + C\left(\frac{\frac{n}{2}-\sum fi}{fmed}\right)$$

li : La limite inférieure de la classe médiane

C : Amplitude de la classe médiane

n : L'effectif total

$\sum fi$: La somme des effectifs de toutes les classes qui précédent la classe médiane

$fmed$: L'effectif de la classe médiane

Exemple

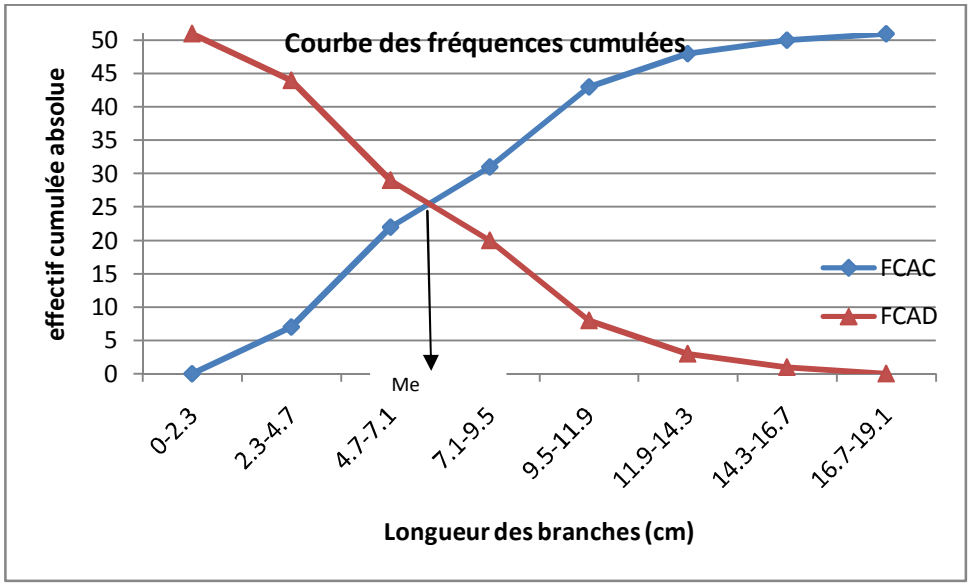
Soit la distribution suivante :

Classes (x_i)	Effectif (ni)	Fréquence relative	Fréquence cumulée absolue croissante	Fréquence cumulée absolue décroissante
[2.3 ; 4.7 [	7	0.14 (7÷51)	7	51
[4.7 ; 7.1 [	15	0.29 (15÷51)	22 (7+15)	44 (51-7)
[7.1 ; 9.5 [	9	0.18 (9÷51)	31 (22+9)	29 (44-15)
[9.5 ; 11.9 [	12	0.23 (12÷51)	43 (31+12)	20 (29-9)
[11.9 ; 14.3 [	5	0.10 (5÷51)	48 (43+5)	8 (20-12)
[14.3; 16.7 [	2	0.04 (2÷51)	50 (48+2)	3 (8- 5)
[16.7 ; 19.1 [	1	0.02 (1÷51)	51 (50+1)	1 (3-2)
TOTAL	51	1		

1- n/2= 51/2=25.5 = 26 ^{éme} observation
2- selon la colonne des FCAC, on la 26 ^{éme} observation se trouve dans la classe [7.1 ; 9.5[
3- donc, on applique la formule de la médiane

Me= $7.1+ 2.4\left(\frac{26-22}{9}\right)$ =8.16 cm

La détermination graphique de la médiane



Quelques remarques sur la médiane

- La médiane est facile à calculer et son interprétation est simple. La médiane ne dépend pas de la valeur numérique des observations, elle dépend plutôt de l'ordre de ces observations. Cette propriété lui permet de ne pas être influencée par les observations aberrantes contrairement à la moyenne arithmétique.

« On préfère la médiane à la moyenne arithmétique comme mesure représentative d'une série statistique si cette dernière comporte des valeurs exceptionnellement grandes ou petites ».

- La médiane se prête mal au calcul algébrique.

1.3 Le Mode

De toutes les mesures de tendance centrale, c'est le mode qui se détermine le plus facilement. En effet , le mode dénoté Mo est la valeur de la variable qui présente l'effectif le plus élevé.

1.3.1 Le mode d'une variable quantitative discontinue

Exemple

La répartition suivante représente la distribution de 56 accidents selon le nombre de victimes

Nombre de victimes	1	2	3	4	5	TOTAL
Nombre d'accidents	32	13	5	3	3	56
Pourcentage	57.1%	23.2%	8.9%	5.4%	5.4%	100%

Le mode de cette distribution est Mo=1 victime, le plus grande nombres d'accidents ont fait 1 seule victime

1.3.2 Le mode d'une variable quantitative continue

Pour déterminer le mode d'une variable quantitative continue, on identifie dans un premier temps la classe modale, puis on associe le mode à la valeur obtenue par l'application de la formule suivante :

$$\text{Mo}= li + C\left(\frac{\Delta 1}{\Delta 1+\Delta 2}\right)$$

li:La limite inférieure de la classe modale

C : Amplitude de la classe modale

$\Delta 1$: La différence entre les effectifs de la classe modale et les effectifs de la classe précédente

$\Delta 2$: la différence entre les effectifs de la classe modale et les effectifs de la classe suivante

Exemple

Soit la distribution suivante

La taille (cm)	Nombre d'étudiants
150-160	10
160-170	30
170-180	60
180-190	5
TOTAL	105

1- La classe modale est 170-180
2- Donc
   li: 170
   C : 10 (180-170=10)
   $\Delta 1$: 60-30=30
   $\Delta 2$: 60-5=55

$$Mo= 170 + 10\left(\frac{30}{30+55}\right) = 173.5 \text{ cm}$$

NB : dans le cas d'une distribution à classes inégales, on détermine le mode après corrections des effectifs

Exemple

La série suivante représente la répartition des exploitations agricoles selon leurs superficies évaluées en ha

Superficies (xi)	Effectifs (ni)	Amplitude de chaque classe	Effectif corrigé
0-1	35	1 (1-0=1)	35
1-2	30	1	30
2-3	25	1	25
3-5	54	2	54/2=27
5-10	100	5	100/5=20
10-20	80	10	80/10=8
20-22	6	2	6/2=3
	330		

1- Choisir l'amplitude de base :

Dans notre exemple l'amplitude base est de 1

2- Correction des effectifs

On prend par exemple la classe 5-10, l'amplitude de cette classe est de 5 (10-5=5), elle est de 5 fois plus grande que l'amplitude de base 1, donc on divise son effectif sur 5, de cette façon on obtient l'effectif corrigé

3- Calcul du mode

Après correction des effectifs, il apparait que la classe modale est 0-1, on applique alors notre formule :

li: 0
C : 1 (1-0=1)
$\Delta 1$: 35-0=35
$\Delta 2$: 350-5=5

$$Mo= 0 + 1\left(\frac{35}{35+5}\right) = 0.875 \text{ ha}$$

Quelques remarques sur le mode

Le mode tire ses avantages du fait que :

- Qu'il est facile à calculer,
- d'avoir une signification assez immédiate
- la présence de deux ou plusieurs modes, dans des échantillons provenant d'une même population, est un indice sérieux de la présence d'un facteur d'hétérogénéité influençant la distribution de la variable
- dans l'étude d'un variable qualitative, le mode, identifié à la modalité dont l'effectif est le plus grand, demeure la seule mesure de tendance centrale applicable

Parmi les inconvénients du mode :

- le mode ne se prête pas au calcul algébrique
- le mode est sensible aux fluctuations d'échantillonnage
- le mode est très sensible au découpage des classes

2 Les caractéristiques de position

Les caractéristiques de position servent à situer une donnée d'une série statistique par rapport aux autres données.

Les quantiles (fractiles), sont des valeurs qui partagent une distribution en un certain nombre de parties égales. Les plus utilisés sont :

Les quartiles (Q), les quintiles (V), les déciles (D), les centiles (C)

2.1 Les principales caractéristiques de position

2.1.1 Quartiles (Q)

Les quartiles divisent l'effectif de la série , préalablement ordonnée par ordre croissant, en quatre parties égales

- Le 1^{ére} quartile $Q_{1=}P_{25}$, est tel que 25% des observations lui sont inférieures et 75% lui sont supérieures
- Le 2^{éme} quartile $Q_{2=}P_{50}$, est tel que 50% des observations lui sont inférieures et 50% lui sont supérieures, c'est la médiane
- Le 3^{éme} quartile $Q_{3=}P_{75}$, est tel que 75% des observations lui sont inférieures et 25% lui sont supérieures

2.1.2 Déciles (D)

Les déciles, divisent la s'effectif de la série statistique en 10 parties égales

- Le 1^{ere} décile $D_{1=}P_{10}$, est la valeur de la variable, telle que 10% des observations lui sont inférieures et 90% lui sont supérieures
- L e 5éme décile $D_{5=}P_{50}$, est la valeur de la variable, telle que 50% des observations lui sont inférieures et 50% lui sont supérieures
- Le 9^{éme} décile $D_9=P_{90}$, est la valeur de la variable, telle que 90% des observations lui sont inférieures et 10% lui sont supérieures

2.1.3 Centiles (C)

Les centiles divisent l'effectif classé dans un ordre croissant en 100 parties égales

- Le 1^{ere} centile $C_{1=}P_1$, est la valeur de la variable, telle que 1% des observations lui sont inférieures et 99 % lui sont supérieures
- Le 50eme centile $C_{50=}P_{50}$, est la valeur de la variable, telle que 50% des observations lui sont inférieures et 50 % lui sont supérieures
- L e 99^{éme} centile $C_{99} =P_{99}$, est la valeur de la variable, telle que 99% des observations lui sont inférieures et 1 % lui sont supérieures

2.1.3.1 Les quantiles d'une variable quantitative discrète

La série suivante représente la distribution de 500 familles selon le nombre d'enfants par famille

Nombre d'enfants	effectifs	FCAC	FCAD	FCRC	FCRD
0	50	50	500	0.1	1
1	120	170	450	0.34	0.90
2	130	290	330	0.60	0.66
3	110	410	200	0.82	0.40
4	90	500	90	1	0.1
TOTAL	500				

Calculer : Q1 , Q2 ,Q3 ,D7

a- Calcul Q_1

Q1= $\frac{n}{4}$, 500/4= 125^{éme} observations, qui correspond dans la distribution des fréquences cumulées à la valeur de la variable 1 enfant, donc Q1= 1 enfant

Q2= $\frac{2n}{4}$, 1000/4= 250^{éme} observations, qui correspond dans la distribution des fréquences cumulées à la valeur de la variable 2 enfant, donc Q2= 2 enfants

Q3= $\frac{3n}{4}$, 3*500/4= 375^{éme} observations, qui correspond dans la distribution des fréquences cumulées à la valeur de la variable 3 enfant, donc Q3= 3 enfants

D_7=$\frac{7n}{10}$=7*500/10=350^{éme} observations, qui correspond dans la distribution des fréquences cumulées à la valeur de la variable 3 enfant, donc D_7= 3 enfants

2.1.3.2 Les quantiles d'une variable quantitative continue

Exemple

Reprenons l'exemple de la répartition de 51 branches d'arbre selon leurs longueurs

Classes (x_i)	Effectif (ni)	Fréquence relative	Fréquence cumulée absolue croissante	Fréquence cumulée absolue décroissante
[2.3 ; 4.7 [	7	0.14	7	51
[4.7 ; 7.1 [	15	0.29	22	44
[7.1 ; 9.5 [	9	0.18	31	29
[9.5 ; 11.9 [	12	0.23	43	20
[11.9 ; 14.3 [	5	0.10	48	8
[14.3; 16.7 [	2	0.04	50	3
[16.7 ; 19.1 [	1	0.02	51	1
TOTAL	51	1		

Calculer Q1 , Q2 ,Q3 ,D9

Calcul : Q1 , Q2 ,Q3 ,D9

Déterminer la classe Q1

Q1= $\frac{n}{4} =$ 51/4= 12.75= 13 branches, qui correspond dans la distribution des fréquences cumulées à la classe [4.7 ; 7.1 [

$$Q1 = li + C\left(\frac{\frac{n}{4} - \Sigma fiQ1}{fQ1}\right) = Me = 4.7 + 2.4\left(\frac{13-7}{15}\right) = 5.66cm$$

Déterminer la classe Q2 :

Q2= $\frac{2n}{4} =$ 2*51/4= 25.5= 26 branches, qui correspond dans la distribution des fréquences cumulées à la classe classe [7.1 ; 9.5[

$$Me = 7.1 + 2.4\left(\frac{26-22}{9}\right) = 8.16 \text{ cm}$$

Déterminer la classe Q3 :

Q3= $\frac{3n}{4} =$ 3*51/4= 39^{éme} branches, qui correspond dans la distribution des fréquences cumulées à la classe classe [9.5 ; 11.9 [

$$Q3 = li + C\left(\frac{\frac{3n}{4} - \Sigma fiQ3}{fQ3}\right) = 9.5 + 2.4\left(\frac{39-31}{12}\right) = 11.09cm$$

Déterminer la classe D7 :

D7= $\frac{7n}{10} =$ 9*51/10= 45.9=46^{éme} branches, qui correspond dans la distribution des fréquences cumulées à la classe classe [11.9 ; 14.3 [

$$D9 = li + C\left(\frac{\frac{9n}{10} - \Sigma fiD9}{fD9}\right) = 11.9 + 2.4\left(\frac{46-43}{5}\right) = 13.34cm$$

3 Les caractéristiques de dispersion

Une caractéristique de dispersion est un nombre, qui permet d'estimer dans quelle mesure des observations s'écartent les unes des autres.

Deux distributions peuvent avoir la même moyenne arithmétique, le même mode et la même médiane, et être assez différentes l'une de l'autre

Exemple

Soit deux groupes de notes :

A= 0 ; 7 ; 7 ; 7 ; 7 ; 7 ; 7 ; 7 ; 10 ; 11 $\longrightarrow \bar{x} = Mo = Me = 7$

B=3 ; 4 ; 5 ; 6 ; 7 ; 7 ; 8 ; 9 ; 10 ; 11 $\longrightarrow \bar{x} = Mo = Me = 7$

Les deux groupes ont les mêmes caractéristiques de valeur centrale ; on ne peut pas conclure que les deux groupes ont les mêmes notes, dans le groupe B les notes sont plus dispersées par rapport au groupe A. il devient alors nécessaire d'ajouter, aux indications fournies par les mesures de tendance centrale d'autres renseignements permettant de quantifier cette dispersion.

3.1 L'étendu

L'étendu d'une série statistique est égale à la différence entre la valeur maximale et la valeur minimale de la série

3.2 L'intervalle interquartile

L'intervalle interquartile est l'intervalle qui contient 50% des observations, on laissant 25% des observations de part et d'autre de l'intervalle. L'intervalle interquartile représente la moitié centrale des effectifs observés

IQ=Q3-Q1

3.3 L'intervalle interdecile

Cet intervalle contient 80% des observations, en laissant respectivement 10% des observations à droite et à gauche de l'intervalle

ID=D9-D1

3.4 La boite à moustaches

Très synthétiques, ces diagrammes sont utiles pour résumer et mettre en relief les caractéristiques des séries statistiques ou les comparer.

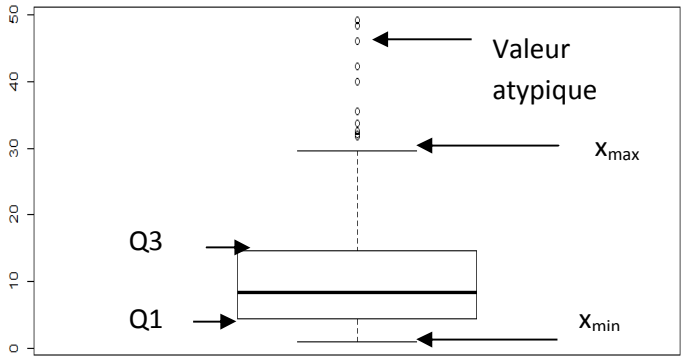
Sur un axe gradué sont représentées :

- Les valeurs extrêmes (minimum et maximum)
- Les 3 quartiles (Q1,Q2,Q3)
- L'intervalle interquartile
- Les valeurs atypiques à l'extérieur de l'intervalle des valeurs extrémes

Certains auteurs comme Tukey, préfèrent de choisir comme extrémités du diagramme les valeurs :

Q1- 1.5 (Q3-Q1) et Q3-1.5 (Q3-Q1)

La boite à moustaches



3.5 La variance et l'ecart-type

La variance est la moyenne arithmétique des carrés des écarts des valeurs de la variable à leur moyenne arithmétique, on la désigne par σ^2

$$\sigma^2 = \frac{\sum_{i=1}^{n}(xi-\bar{x})^2\ ni}{n}$$

Après simplification de la formule, la variance devient :

$$\sigma^2 = \frac{\sum_{i=1}^{n} nixi^2}{\sum ni} - \bar{x}$$

L'écart-type est la racine carrée de la variance

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(xi-\bar{x})^2\ ni}{n}}$$

3.5.1 Calcul de la variance d'une variable quantitative discrète

La série suivante représente le nombre d'absences d'un échantillon de 10 étudiants , calculer la variance de cette distribution

Nombre d'absences (xi)	Nombre d'etudiants (ni)	ni xi^2
1	2	2
2	1	4
4	1	16
6	1	36
7	4	196
9	1	81
TOTAL	10	335

σ^2=335/10 – 5.1^2 = 7.49

$\sigma = \sqrt{7.49}$ =2.73

3.5.2 Calcul de la variance d'une variable quantitative continue

Le tableau ci-dessous représente la répartition des 219 salariés selon leurs salaires bruts mensuels, Calculer la variance te l'écart-type

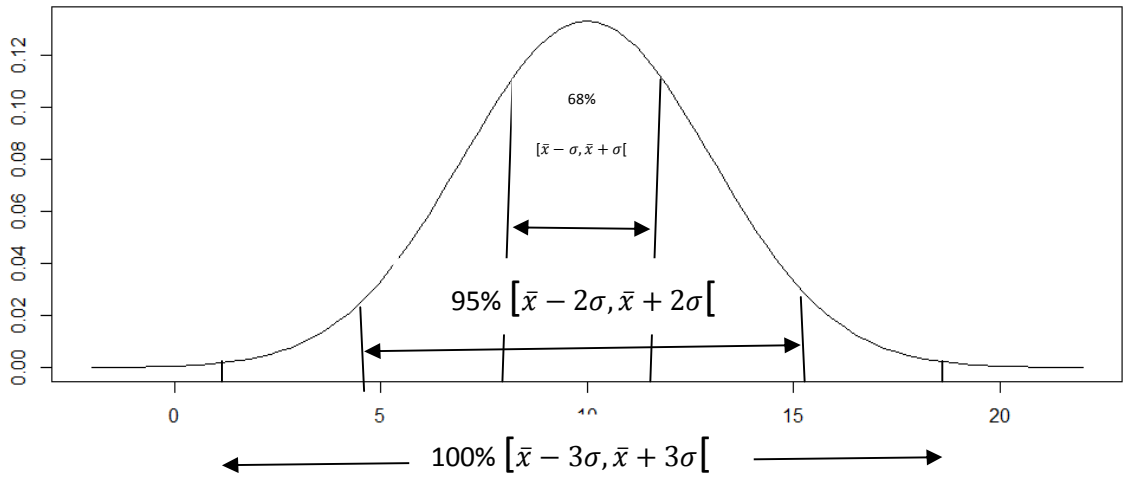
Salaire brut (10^3 Da)	Effectif (ni)	Centre de classe (xi)	$nixi^2$
4-5	12	4.5	243
5-6	23	5.5	695.75
6-7	42	6.5	1774.50
7-9	56	8.0	3584
9-11	34	10.0	3400
11-15	32	13.0	5408
15-23	16	19.0	5776
23-31	4	27 .0	2916
	$n = 219$		$\sum_{i=1}^{k} nixi^2 = 23797.25$

σ^2=23797.25/219 – 9.45^2 = 19.3651x 10^3 , donc $\sigma = 4.4$ x10^3

3.5.3 Relation entre l'écart-type et la moyenne arithmétique

La moyenne et l'écart-type sont les deux valeurs les plus utilisées, pour caractérisées une distribution. Ce qui rend l'écart-type particulièrement intéressant comme mesure de dispersion, c'est la constance des propriétés suivantes, que l'on peut observer à chaque fois que la distribution traitée est unimodale et passablement symétrique :

- Environ 68% de toutes les observations sont comprises dans l'intervalle $[\bar{x} - \sigma, \bar{x} + \sigma[$
- Environ 95% de toutes les observations sont comprises dans l'intervalle $\left[\bar{x} - 2\sigma, \bar{x} + 2\sigma\right[$
- Environ 100% de toutes les observations sont comprises dans l'intervalle $\left[\bar{x} - 3\sigma, \bar{x} + 3\sigma\right[$



3.6 Le coefficient de variation

Soit X une variable quantitative, mesurée dans un échantillon donnée, de moyenne $\bar{x}$ et d'écart-type s alors, le coefficient de variation de X, dénoté CV, est défini ainsi :

$$CV=\frac{S}{\bar{x}} \times 100$$

Il s'agit d'une grandeur sans dimension (sans unité), donc indépendante du choix de l'unité de mesure, et s'exprime en pourcentage. Le coefficient de variation CV est un indice de l'hétérogénéité d'un échantillon

- Un CV à 15% ou moins ⟹ échantillon homogène
- Un CV supérieur à 15% ⟹ échantillon hétérogène

3.7 La cote Z, ou la cote standard

Etant donné une série numérique dont la moyenne et l'écart-type de la variable x sont respectivement $\bar{x}$ et s , la cote standard associée à une valeur x_i quelconque appartenant à cette série numérique est dénotée Z_i , et est de la façon suivante :

$$Z_i=\frac{xi-\bar{x}}{s}$$

- Il s'agit d'une valeur sans dimension ;
- Le signe Z_i donne la position de la valeur xi par rapport à $\bar{x}$
- La valeur Z_i se situe généralement dans l'intervalle [-3,+3[

La cote Z permet :

1- de comparer les valeurs d'une variable appartenant à plusieurs échantillons, en les ramenant sur la même échelle ;
2- de mettre en lumière (en évidence) le caractère d'exceptionnalité d'une valeur