

LES PARAMETRES DE REDUCTION

Dr . N Fermas
Médecin Epidémiologiste
CHU de Sétif
13/01/2021

1. INTRODUCTION

Paramètres de réduction = Statistiques de réduction :
VARIABLE QUANTITATIVE

Valeurs numériques permettant de **résumer** et de donner les **principales caractéristiques** d'une distribution statistique

Echantillon :

Généralisation des résultats de l'échantillon à l'ensemble de la population

3 Groupes de paramètres :

- *Les paramètres de tendance centrale*
- *Les paramètres de position*
- *Les paramètres de dispersion*

2. LES PARAMETRES DE TENDANCE CENTRALE

Centre de la série statistique

2.1. LA MOYENNE

2.1.1. La moyenne arithmétique

La moyenne arithmétique simple :

$$m = \frac{\sum x_i}{n}$$

Soit la série suivante :

3,3,3,3,4,4,5,5,5,5,7,9

$$m = \frac{3+3+3+3+4+4+5+5+5+5+7+9}{12} = \frac{56}{12} = 4,67$$

La moyenne arithmétique pondérée :

Observations répétées

Calcul plus simple

$$m = \frac{\sum n_i x_i}{n}$$

$$m = \frac{(4 \times 3) + (2 \times 4) + (4 \times 5) + 7 + 9}{12} = \frac{56}{12} = 4,67$$

Recours aux fréquences relatives :

$$f_i = \frac{n_i}{n}$$

$$m = \sum f_i x_i$$

$$\sum f_i = f_1 + f_2 + f_3 + \dots + f_n = 1$$

$$m = \frac{4}{12} \times 3 + \frac{2}{12} \times 4 + \frac{4}{12} \times 5 + \frac{1}{12} \times 7 + \frac{1}{12} \times 9$$

$$= 1 + 0,67 + 1,67 + 0,58 + 0,75 = 4,67$$

Poids de 19 étudiants (kg)

76,34 52,99
 60,4 57,74
 68,28 59,79
 57,74 60,4
 64,99 61,82
 83,45 61,82
 79,65 64,1
 64,1 64,99
 72,88 65,45
 69,12 66,33
 59,79 68,28
 61,82 69,12
 61,82 70,13
 76,36 70,56
 66,33 72,88
 52,99 76,34
 70,56 76,36
 70,13 79,65
 65,45 83,45

7 classes de 5 kg

Indice de classe	Classe	X_i	N_i	$N_i X_i$
1	50 – 55	52,5	1	52,5
2	55 – 60	57,5	2	115
3	60 – 65	62,5	5	312,5
4	65 – 70	67,5	4	270
5	70 – 75	72,5	3	217,5
6	75 – 80	77,5	3	232,5
7	80 - 85	82,5	1	82,5
Σ			19	1282,5

$\Sigma N_i X_i = 1282,5 \text{ kg}$
 $m = 1282,5 / 19$
 $m = 67,5 \text{ Kg}$

$\Sigma X_i = 1282,2 \text{ kg}$
 $m = 1282,2 / 19$
 $m = 67,48 \text{ Kg}$

Propriétés de la moyenne arithmétique :

□ $\Sigma (X_i - m) = 0$

□ $\Sigma (X_i - m)^2 =$ Quantité minimale

□ Echantillon constitué de 2 sous-échantillons A et B

$$m = \frac{n_A m_A + n_B m_B}{n_A + n_B}$$

□ Généralisation à k sous-échantillons

$$m = \frac{\Sigma n_k m_k}{\Sigma n_k}$$

2.2. LE MODE

Modalité de la variable correspondant à l'effectif absolu le plus élevé

*Distribution du nombre annuel d'épisodes de syndrome grippal
chez une population de 77 sujets*

X_i	N_i
0	14
1	16
2	18
3	17
4	12
ΣN_i	77



Mode m_o

Série statistique groupée

Taille de 307 footballeurs algériens

Indice de classe	Classe	Contre de classe X_i	Effectif N_i
1	159,5 – 161,5	160,5	7
2	161,5 – 163,5	162,5	4
3	163,5 – 165,5	164,5	10
4	165,5 – 167,5	166,5	23
5	167,5 – 169,5	168,5	19
6	169,5 – 171,5	170,5	39
7	171,5 – 173,5	172,5	55
8	173,5 – 175,5	174,5	48
9	175,5 – 177,5	176,5	35
10	177,5 – 179,5	178,5	31
11	179,5 – 181,5	180,5	16
12	181,5 – 183,5	182,5	9
13	183,5 – 185,5	184,5	5
14	185,5 – 187,5	186,5	3
15	187,5 – 189,5	188,5	1
16	189,5 – 191,5	190,5	2

$\sum N_i = 307$

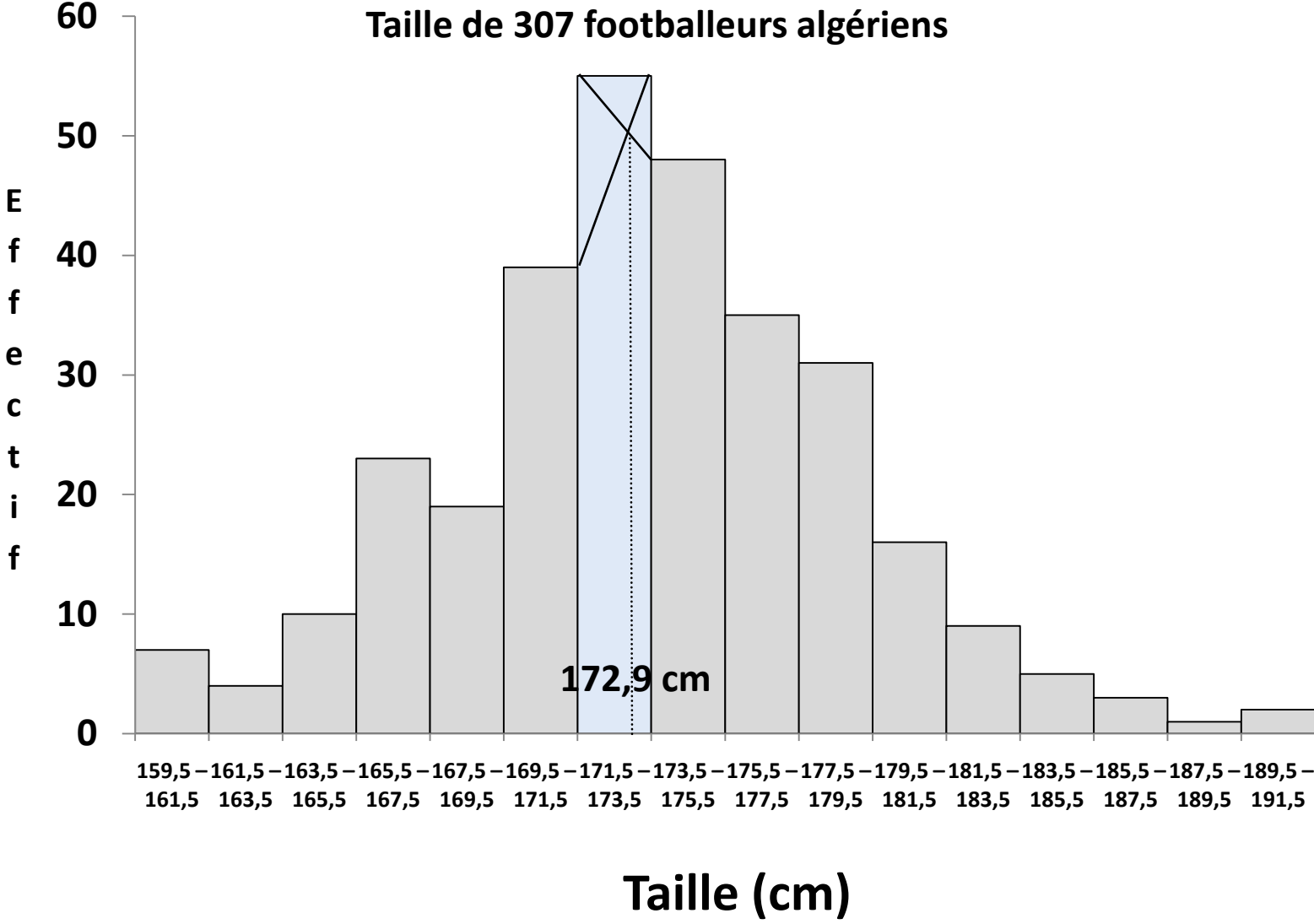
Classe modale



Mode m_o



Détermination graphique du mode



Détermination du mode par les calculs :

$$m_o = x_i + \frac{D_1}{D_1 + D_2} \times a$$

$$m_o = 171,5 + \frac{16}{16 + 7} \times 2$$

$$m_o = 172,9 \text{ cm}$$

Séries monomodales

- bimodales
- multimodales

$$m_o = x_i + \frac{D_1}{D_1 + D_2} \times a$$

Taille de 307 footballeurs algériens

Indice de classe	Classe	Contre de classe x_i	Effectif N_i
1	159,5 – 161,5	160,5	7
2	161,5 – 163,5	162,5	4
3	163,5 – 165,5	164,5	10
4	165,5 – 167,5	166,5	23
5	167,5 – 169,5	168,5	19
6	169,5 – 171,5	170,5	39
7	171,5 – 173,5	172,5	55
8	173,5 – 175,5	174,5	48
9	175,5 – 177,5	176,5	35
10	177,5 – 179,5	178,5	31
11	179,5 – 181,5	180,5	16
12	181,5 – 183,5	182,5	9
13	183,5 – 185,5	184,5	5
14	185,5 – 187,5	186,5	3
15	187,5 – 189,5	188,5	1
16	189,5 – 191,5	190,5	2
$\Sigma N_i = 307$			

x_i = Limite inférieure de la classe modale

$$D_1 = 55 - 39 = 16$$

$$D_2 = 55 - 48 = 7$$

a = Amplitude de classe = 2

$$m_o = 171,5 + \frac{16}{16 + 7} \times 2$$

2.3. LA MEDIANE

Valeur de la variable qui se trouve au milieu de la série statistique

Après classement :

Observations inférieures à la médiane

=

Observations supérieures à la médiane

Données non groupées :

Petit effectif :

Ordonner les observations

Pendre celle du milieu

Poids de 19 étudiants (kg)

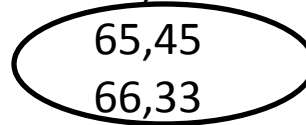
52,99
57,74
59,79
60,4
61,82
61,82
64,1
64,99
65,45
66,33
68,28
69,12
70,13
70,56
72,88
76,34
76,36
79,65
83,45

Nombre d'observations pair = Intervalle médian

52,99
57,74
59,79
60,4
61,82
61,82
64,1
64,99
65,45
66,33
68,28
69,12
70,13
70,56
72,88
76,36
79,65
83,45



Médiane



Intervalle médian

Modalités peu nombreuses d'une série discontinue :

Médiane correspond au 1^{er} effectif cumulé « moins de » $\geq N/2$
 $N/2 = 77 / 2 = 38,5$

*Distribution du nombre annuel d'épisodes de syndrome grippal
chez une population de 77 sujets*

Médiane



X_i	N_i	$N_i (-)$
0	14	14
1	16	30
2	18	48
3	17	65
4	12	77
ΣN_i	77	

Données groupées :

- Classe médiane
- Centre de la classe médiane

Méthode d'interpolation :

$$m_e = x_i + \frac{N/2 - S}{N_i} \times a$$

$$m_e = 171,5 + \frac{153,5 - 102}{55} \times 2 = 173,4 \text{ cm}$$

$$m_e = x_i + \frac{N/2 - S}{N_i} \times a$$

$$N/2 = 307 / 2 = 153,5$$

Taille de 307 footballeurs algériens

Indice de classe	Classe	Contre de classe X_i	Effectif N_i	Effectif Cumulé « moins de » $N_i(-)$
1	159,5 – 161,5	160,5	7	7
2	161,5 – 163,5	162,5	4	11
3	163,5 – 165,5	164,5	10	21
4	165,5 – 167,5	166,5	23	44
5	167,5 – 169,5	168,5	19	63
6	169,5 – 171,5	170,5	39	102
7	171,5 – 173,5	172,5	55	157
8	173,5 – 175,5	174,5	48	205
9	175,5 – 177,5	176,5	35	240
10	177,5 – 179,5	178,5	31	271
11	179,5 – 181,5	180,5	16	287
12	181,5 – 183,5	182,5	9	296
13	183,5 – 185,5	184,5	5	301
14	185,5 – 187,5	186,5	3	304
15	187,5 – 189,5	188,5	1	305
16	189,5 – 191,5	190,5	2	307
$\Sigma N_i = 307$				

1^{er} effectif cumulé
« moins de » $\geq N/2 = 157$

x_i = Limite inférieure
de la classe médiane

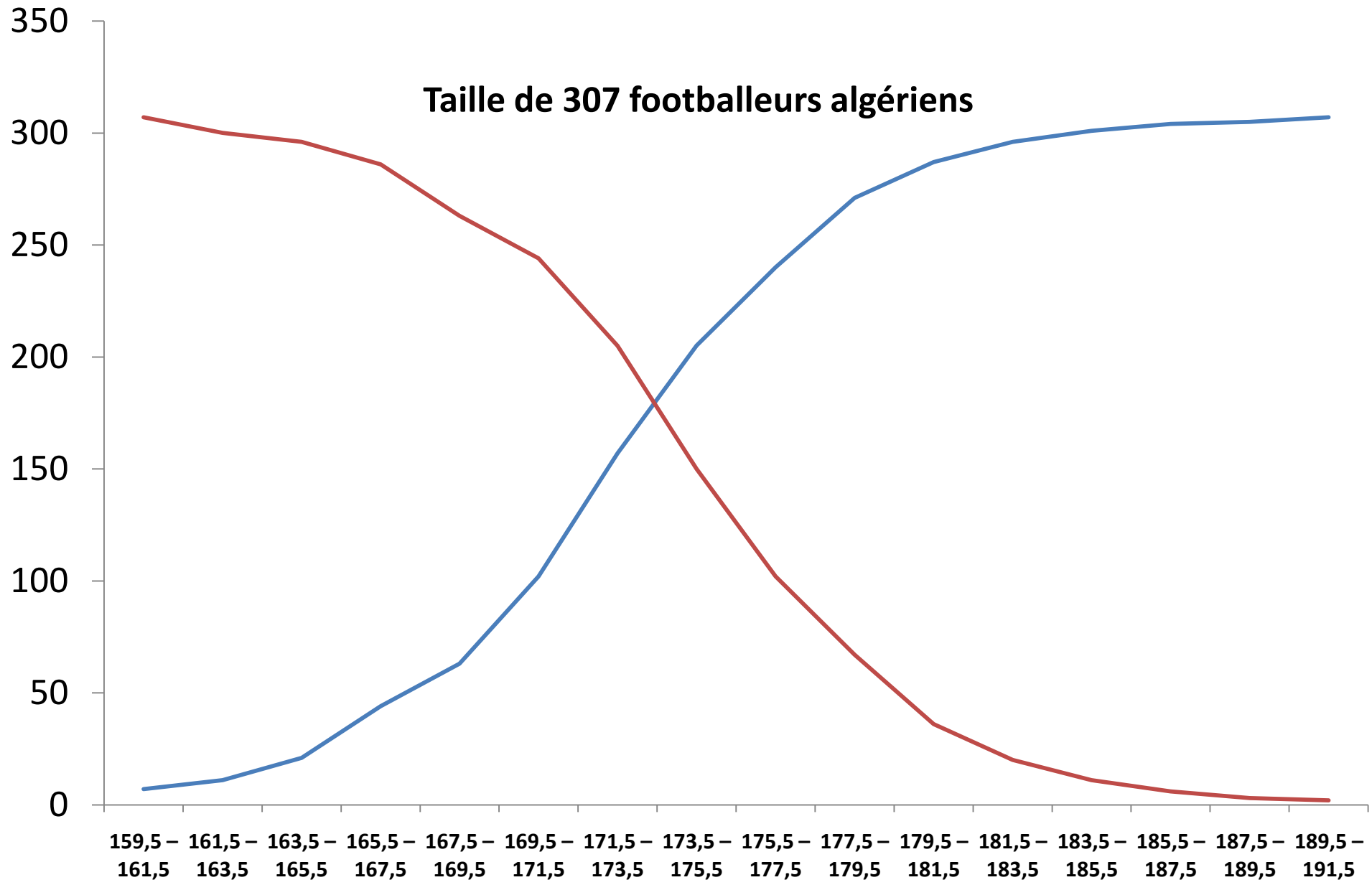
S = Effectif cumulé « moins de »
jusqu' à la classe médiane
non comprise

N_i = Effectif absolu de la classe médiane

a = Amplitude de classe = 2

$$m_e = 171,5 + \frac{153,5 - 102}{55} \times 2 = 173,4 \text{ cm}$$

Taille de 307 footballeurs algériens



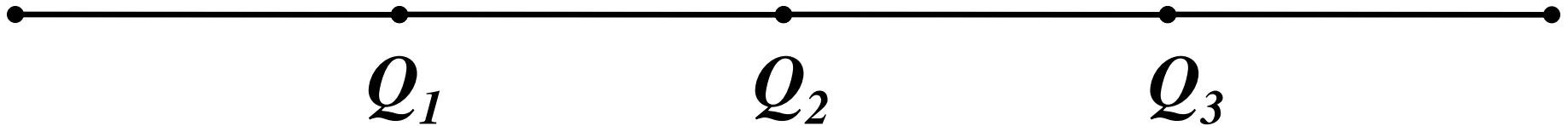
3. LES PARAMETRES DE POSITION

SERIE ORDONNEE :

- Divisible en parties égales (2, 4, 10 ou 100)
- Paramètres occupant un certain rang
- Médiane = paramètre de position :
divise la série en 2 parties égales

3.1. LES QUARTILES

Divisent la série statistique en **4 parties égales** comprenant le **même nombre de sujets**



$Q_1 = 25^{\text{ème}}$ valeur sur 100

$Q_2 = 50^{\text{ème}}$ valeur sur 100 = *Médiane*

$Q_3 = 75^{\text{ème}}$ valeur sur 100

Taille de 307 footballeurs algériens

Indice de classe	Classe	Contre de classe X_i	Effectif N_i	Effectif Cumulé « moins de » $N_i(-)$	Effectif Cumulé « plus de » $N_i(+)$
1	159,5 – 161,5	160,5	7	7	307
2	161,5 – 163,5	162,5	4	11	300
3	163,5 – 165,5	164,5	10	21	296
4	165,5 – 167,5	166,5	23	44	286
5	167,5 – 169,5	168,5	19	63	263
6	169,5 – 171,5	170,5	← 39	102	244
7	171,5 – 173,5	172,5	← 55	157	205
8	173,5 – 175,5	174,5		205	150
9	175,5 – 177,5	176,5	← 35	240	102
10	177,5 – 179,5	178,5	31	271	67
11	179,5 – 181,5	180,5	16	287	36
12	181,5 – 183,5	182,5	9	296	20
13	183,5 – 185,5	184,5	5	301	11
14	185,5 – 187,5	186,5	3	304	6
15	187,5 – 189,5	188,5	1	305	3
16	189,5 – 191,5	190,5	2	307	2
			$\Sigma N_i = 307$		

Q_1 correspond à la **76,75^{ème}** valeur (25 % de 307) : 170,5 cm

Q_2 correspond à la **153,5^{ème}** valeur (50 % de 307) : 172,5 cm

Q_3 correspond à la **230,25^{ème}** valeur (75 % de 307) : 176,5 cm

Méthode d'interpolation :

Tenir compte du rang recherche

$$Q_1 = 169,5 + \frac{(307 / 4) - 63}{39} \times 2 = 170,2 \text{ cm}$$

$$Q_3 = 175,5 + \frac{(307 \times \frac{3}{4}) - 205}{35} \times 2 = 176,9 \text{ cm}$$

Possibilité de déterminer les quartiles graphiquement

Q_1 correspond à la 76,75^{ème} valeur (25 % de 307)

1^{er} effectif cumulé « moins de » $\geq N/4 = 102$

x_i = Limite inférieure de la classe correspondante

S = Effectif cumulé « moins de » jusqu' à la classe en question non comprise

Taille de 307 footballeurs algériens

Indice de classe	Classe	Contre de classe X_i	Effectif N_i	Effectif Cumulé « moins de » $N_i(-)$
1	159,5 – 161,5	160,5	7	7
2	161,5 – 163,5	162,5	4	11
3	163,5 – 165,5	164,5	10	21
4	165,5 – 167,5	166,5	23	44
5	167,5 – 169,5	168,5	19	63
6	169,5 – 171,5	170,5	39	102
7	171,5 – 173,5	172,5	55	157
8	173,5 – 175,5	174,5	48	205
9	175,5 – 177,5	176,5	35	240
10	177,5 – 179,5	178,5	31	271
11	179,5 – 181,5	180,5	16	287
12	181,5 – 183,5	182,5	9	296
13	183,5 – 185,5	184,5	5	301
14	185,5 – 187,5	186,5	3	304
15	187,5 – 189,5	188,5	1	305
16	189,5 – 191,5	190,5	2	307

$\Sigma N_i = 307$

N_i = Effectif absolu de la classe en question

a = Amplitude de classe = 2

$$Q_1 = 169,5 + \frac{(307/4) - 63}{39} \times 2 = 170,2 \text{ cm}$$

Q_3 correspond à la **230,25^{ème}** valeur (75 % de 307)

1^{er} effectif cumulé « moins de » $\geq N \times \frac{3}{4} = 240$

x_i = Limite inférieure de la classe correspondante

S = Effectif cumulé « moins de » jusqu' à la classe en question non comprise

Taille de 307 footballeurs algériens

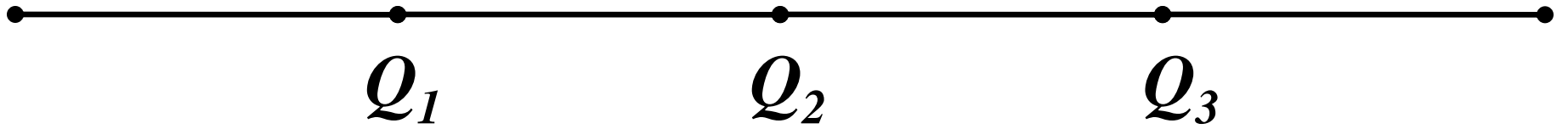
Indice de classe	Classe	Centre de classe x_i	Effectif N_i	Effectif Cumulé « moins de » $N_i(-)$
1	159,5 – 161,5	160,5	7	7
2	161,5 – 163,5	162,5	4	11
3	163,5 – 165,5	164,5	10	21
4	165,5 – 167,5	166,5	23	44
5	167,5 – 169,5	168,5	19	63
6	169,5 – 171,5	170,5	39	102
7	171,5 – 173,5	172,5	55	157
8	173,5 – 175,5	174,5	48	205
9	175,5 – 177,5	176,5	35	240
10	177,5 – 179,5	178,5	31	271
11	179,5 – 181,5	180,5	16	287
12	181,5 – 183,5	182,5	9	296
13	183,5 – 185,5	184,5	5	301
14	185,5 – 187,5	186,5	3	304
15	187,5 – 189,5	188,5	1	305
16	189,5 – 191,5	190,5	2	307

N_i = Effectif absolu de la classe en question

a = Amplitude de classe = 2

$$Q_3 = 175,5 + \frac{(307 \times \frac{3}{4}) - 205}{35} \times 2 = 176,9 \text{ cm}$$

$\Sigma N_i = 307$



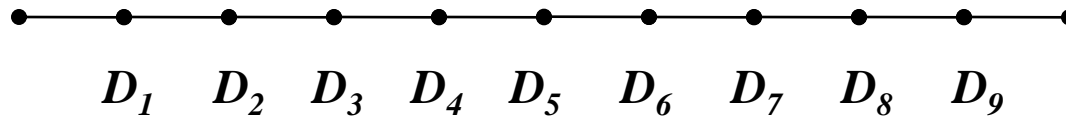
Intervalle inter-quartile

$$= Q_3 - Q_1$$

Englobe 50 % de sujets

3.2. LES DECILES

Divisent la série statistique en **10 parties égales** comprenant le **même nombre de sujets**



$D_1 = 10^{\text{ème}}$ valeur sur 100

$D_2 = 20^{\text{ème}}$ valeur sur 100

$D_5 = 50^{\text{ème}}$ valeur sur 100 = *Médiane*

.....

3.3. LES PERCENTILES

Divisent la série statistique en **100 parties égales**
comprenant le **même nombre de sujets**

$N \geq 1000$

$P_1 = 10^{\text{ème}}$ valeur sur 1000

$P_{42} = 420^{\text{ème}}$ valeur sur 1000

$P_{50} = 500^{\text{ème}}$ valeur sur 1000 = *Médiane*

$P_{10} = 100^{\text{ème}}$ valeur sur 1000

$P_{20} = D_2$

$P_{25} = Q_1$

$P_{75} = Q_3$

4. LES PARAMETRES DE DISPERSION

2 Séries statistiques :

- $S_1 = 15, 20, 25, 30, 35$
- $S_2 = 5, 15, 25, 35, 45$

$$m1 = m2 = 25$$

Estimation de la **variabilité** de la série

Mesures d'étalement des observations autour des paramètres de tendance centrale

4.1. L'ÉTENDUE (La marge)

Différence entre les deux valeurs extrêmes de la série

Taille de 307 footballeurs algériens

Minimum 159,5 cm

Maximum 191,5 cm

Marge = $191,5 - 159,5 = 32$ cm

- $S_1 : 35 - 15 = 20$

- $S_2 : 45 - 5 = 40$

4.2. L'INTERVALLE INTER-QUARTILE

Q1 - Q3 : *50 % des observations*

4.3. LE 10-90 PERCENTILE RANGE

Élimine 10 % des valeurs à chaque extrémité

Garde 80 % des valeurs

Taille de 307 footballeurs algériens : **166,3 – 180,2 cm**

4.4. L'ECART ABSOLU MOYEN

$$EAM = \frac{\sum n_i | x_i - m |}{n}$$

- $S_1 : 6$
- $S_2 : 12$

S1

x_i	$ x_i - m $	$(x_i - m)^2$
15	10	100
20	5	25
25	0	0
30	5	25
35	10	100
$\sum x_i = 125$	$\sum x_i - m = 30$	$\sum (x_i - m)^2 = 250$

S2

x_i	$ x_i - m $	$(x_i - m)^2$
5	20	400
15	10	100
25	0	0
35	10	100
45	20	400
$\sum x_i = 125$	$\sum x_i - m = 60$	$\sum (x_i - m)^2 = 1000$

4.5. LA VARIANCE ET L'ECART-TYPE

4.5.1. Formules :

Variance : S_x^2

Données non groupées :

$$S_x^2 = \frac{\sum (x_i - m)^2}{n}$$

Données groupées :

$$S_x^2 = \frac{\sum n_i (x_i - m)^2}{n}$$

Ecart - type : S_x

Données non groupées :

$$S_x = \sqrt{\frac{\sum (x_i - m)^2}{n}}$$

Données groupées :

$$S_x = \sqrt{\frac{\sum n_i (x_i - m)^2}{n}}$$

❖ S1

$$S_x^2 = \frac{250}{5} = 50$$

$$S_x = \sqrt{50} = 7.07$$

❖ S2

$$S_x^2 = \frac{1000}{5} = 200$$

$$S_x = \sqrt{200} = 14,14$$

4.5.2. Variance d'une série à faible effectif :

$N < 30$

Remplacer le dénominateur par : $(n - 1)$

4.5.3. Variance d'un échantillon constitué de sous-échantillons :

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2} + \frac{n_1 (m_1 - m)^2 + n_2 (m_2 - m)^2}{n_1 + n_2}$$



Variance pondérée des 2
variances = variance résiduelle
= Variance intra échantillon



Variance des moyennes

Même moyenne : Variance totale = Variance résiduelle

Exemple : Etude de la masse grasse en Kg de filles scolarisées âgées de 9 ans de 4 communes

Commune	C1	C2	C3	C4
Effectif	42	62	32	38
Moyenne	4,6	4,1	5	4,3
Ecart-type	1,3	0,9	1,4	0,9

$$m = \frac{(42 \times 4,6) + (62 \times 4,1) + (32 \times 5) + (38 \times 4,3)}{42 + 62 + 32 + 38} = \frac{770,8}{174} = 4,4 \text{ Kg}$$

$$S^2 = \frac{(42 \times 1,7) + (62 \times 0,8) + (32 \times 2) + (38 \times 0,8)}{42 + 62 + 32 + 38}$$

$$+ \frac{42 \times (4,6 - 4,4)^2 + 62 \times (4,1 - 4,4)^2 + 32 \times (5 - 4,4)^2 + 38 \times (4,3 - 4,4)^2}{42 + 62 + 32 + 38}$$

$$S^2 = 1,24 + 0,11 = 1,35 \text{ kg}^2$$

$$S = \sqrt{1,35} = 1,16 \text{ kg}$$

4.5.3. Variance d'une variable $x = y + z$:

2 séries de variables y et z de même taille échantillonnale n telles que :

$$x_i = y_i + z_i$$

$$S_x^2 = \frac{\sum (y - m_y)^2}{n} + \frac{\sum (z - m_z)^2}{n} + \frac{2\sum (y - m_y)(z - m_z)}{n}$$

Variance totale = variance de y + Variance de z + covariance (y,z)

4.6. LE COEFFICIENT DE VARIATION

$$CV = \frac{s}{m} \quad (\%)$$

- ✓ Degré d'homogénéité (hétérogénéité) d'une série statistique
- ✓ Degré de dispersion de 2 variables d'une même série statistique
- ✓ Proportion de dispersion des variables de deux séries statistiques

X_1	X_2	Etendue	Variance	Ecart-type	CV (%)
4	4	0	0	0	0
5	3	2	1	1	25
6	2	4	4	2	50
7	1	6	9	3	75
8	0	8	16	4	100

2 groupes :

- **Groupe 1** : $m_1 = 115$ $s_1 = 4$ $CV = 3,48 \%$

- **Groupe 2** : $m_2 = 145$ $s_2 = 4$ $CV = 2,76 \%$