Correction des exercices sur le TD1 :
« Résumé d'une série de mesures »

Solution de l'exercice 1 :

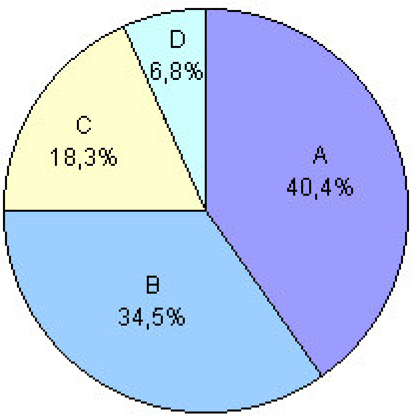
L'effectif total est : 51 210 + 43 821 + 23 212 + 8597 = 126 840

On peut alors calculer les différentes fréquences :

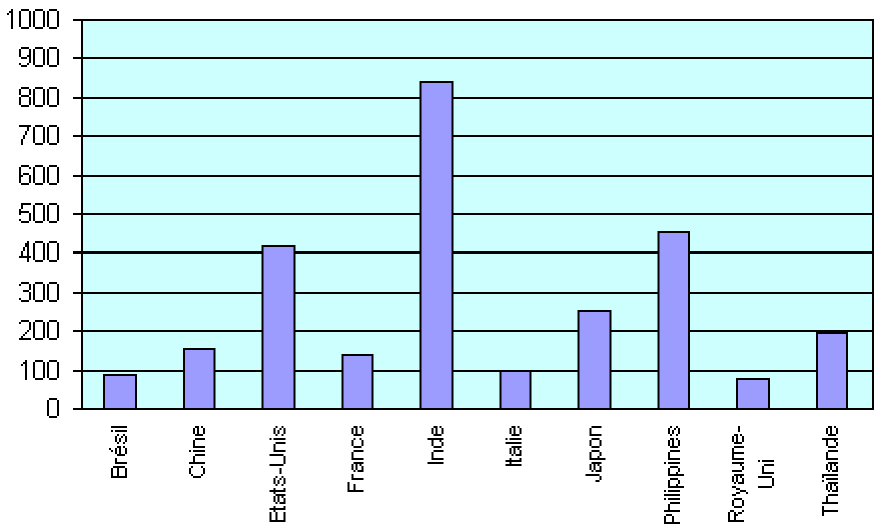
$\frac{51\ 210}{126\ 840} \approx 0,404$; $\frac{43\ 821}{126\ 840} \approx 0,345$; $\frac{23\ 212}{126\ 840} \approx 0,183$; $\frac{8\ 597}{126\ 840} \approx 0,068$

Candidat	A	B	C	D
Nombre de voix obtenues	51 210	43 821	23 212	8 597
Fréquence (en pourcentage)	40,4%	34,5%	18,3%	6,8%

Dans un diagramme circulaire, la mesure angulaire de chaque secteur correspond à la fréquence.
Le cercle complet correspond à un angle au centre de 360°,
le secteur du candidat A aura pour mesure angulaire 360° x 40,4% ≈ 145°
le secteur du candidat B aura pour mesure angulaire 360° x 34,5% ≈ 124°
le secteur du candidat C aura pour mesure angulaire 360° x 18,3% ≈ 66°
le secteur du candidat D aura pour mesure angulaire 360° x 6,8% ≈ 25°
On obtient alors le diagramme circulaire ci dessous :



Solution de l'exercice 2 :



Solution de l'exercice 3 :

Il faut, pour calculer le prix de vente moyen d'un CD, tenir compte du nombre d'exemplaires vendus (effectif) à chacun des prix relevés.
Le prix de vente moyen d'un CD est :

$$\frac{97 \times 15 + 34 \times 16 + 43 \times 17 + 20 \times 18 + 6 \times 19}{97 + 34 + 43 + 20 + 6} = \frac{3\,204}{200} = 16{,}02$$

Solution de l'exercice 4 :

1°) Pour la classe A :

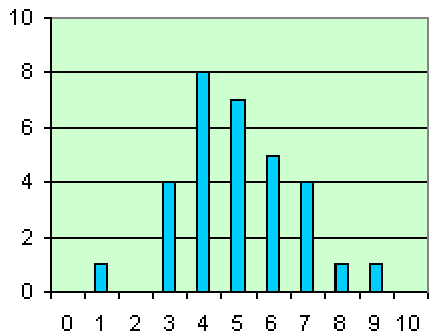
Note : x_i	0	1	2	3	4	5	6	7	8	9	10
Effectif : n_i	0	1	0	4	8	7	5	4	1	1	0

Pour la classe B :

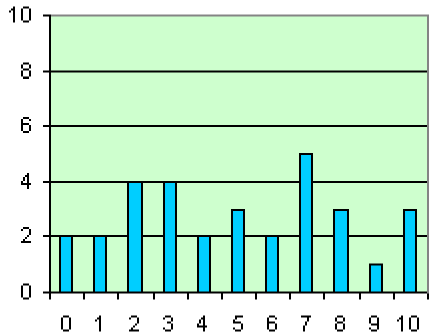
Note : x_i	0	1	2	3	4	5	6	7	8	9	10
Effectif : n_i	2	2	4	4	2	3	2	5	3	1	3

2°) On obtient pour chacune des classes le diagramme à barres ci-dessous :

Classe A



Classe B



3°) La moyenne de A est donnée par $\overline{x_A} = \frac{1 \times 1 + 4 \times 3 + 8 \times 4 + 7 \times 5 + 5 \times 6 + 4 \times 7 + 1 \times 8 + 1 \times 9}{1 + 4 + 8 + 7 + 5 + 4 + 1 + 1}$

La moyenne de la classe A est : $\overline{x_A} = \frac{155}{31} = 5$

La moyenne de B est $\overline{x_B} = \frac{2 \times 0 + 2 \times 1 + 4 \times 2 + 4 \times 3 + 2 \times 4 + 3 \times 5 + 2 \times 6 + 5 \times 7 + 3 \times 8 + 1 \times 9 + 3 \times 10}{2 + 2 + 4 + 4 + 2 + 3 + 2 + 5 + 3 + 1 + 3}$

La moyenne de la classe B est : $\overline{x_B} = \frac{155}{31} = 5$

4°)

Pour la classe A :

Note (x_i)	0	1	2	3	4	5	6	7	8	9	10
Effectif (n_i)	0	1	0	4	8	7	5	4	1	1	0
$(x_i - \overline{x})^2$	25	16	9	4	1	0	1	4	9	16	25
$n_i (x_i - \overline{x})^2$	0	16	0	16	8	0	5	16	9	16	0

Pour la classe B :

Note	0	1	2	3	4	5	6	7	8	9	10
Effectif	2	2	4	4	2	3	2	5	3	1	3
$(x_i - \overline{x})^2$	25	16	9	4	1	0	1	4	9	16	25
$n_i (x_i - \overline{x})^2$	50	32	36	16	2	0	2	20	27	16	75

Pour la classe A, la variance est donnée par : $V_A = \frac{16 + 16 + 8 + 5 + 16 + 9 + 16}{31} \approx 2{,}77$

Pour la classe A, $V_A \approx 2{,}77$, et $\sigma_A \approx 1{,}67$

Pour la classe B, la variance est : $V_B == \frac{50 + 32 + 36 + 16 + 2 + 2 + 20 + 27 + 16 + 75}{31} \approx 8{,}90$

Pour la classe B, $V_B \approx 8{,}90$ et $\sigma_B \approx 2{,}98$

5°) La classe A et la classe B ont la même moyenne, mais au vu de l'écart-type de chaque série, on peut dire que la classe A est une classe homogène alors que la classe B est une classe hétérogène.

6°) Pour la classe A, la note du 16ème élève est : 5
Pour la classe B, la moyenne du 16ème élève est : 5

Remarque : chaque classe ayant 31 élèves, la note du note du 16ème élève (c'est-à-dire de l'élève ayant 15 élèves avant lui et 15 élèves après) est appelée la médiane de la série.
La médiane est comme la moyenne une mesure de tendance centrale, on peut la trouver sans faire de calculs compliqués, et on peut remarquer que le fait de changer la note du meilleur élève de la classe (ou du moins bon) modifiera la moyenne, mais ne modifiera pas la médiane.

Solution de l'exercice 5 :

1°) La lecture du graphique permet de compléter la ligne des effectifs.
On calcule ensuite les effectifs cumulés.

Taille	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111
Effectif	1	0	0	1	0	1	0	2	2	6	6	3	13	15	22	22
Effectif cumulé	1	1	1	2	2	3	3	5	7	13	19	22	35	50	72	94

Taille	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	
Effectif	21	21	19	14	16	18	19	15	3	3	8	5	1	1	1	
Effectif cumulé	115	136	155	169	185	203	222	237	240	243	251	256	257	258	259	

2°) On trouve l'effectif total de la série dans la dernière colonne des effectifs cumulés.
L'effectif total de la série est : 259

3°) La série a un nombre impair de termes : 259. On a $\frac{259}{2} = 129{,}5$.

La médiane est donc la valeur du 130ème terme.
On peut observer à partir de la ligne des effectifs cumulés que le 130ème terme de la série correspond à la valeur 113.
La médiane de cette série est donc 113.

4°) On a $259 \times 0{,}25 = 64{,}75$. Le 1er quartile est donc la valeur du 65ème terme de la série.

On peut observer à partir des effectifs cumulés que le 1er quartile est 110.

On a $259 \times 0{,}75 = 194{,}25$. Le 3ème quartile est donc la valeur du 195ème terme de la série.

On peut observer à partir des effectifs cumulés que le 3ème quartile est 117.

L'intervalle interquartile est donc l'intervalle $[110 ; 117]$.

L'écart interquartile est $117 - 110 = 7$.

5°) On a $259 \times 0{,}1 = 25{,}9$. Le 1er décile est donc la valeur du 26ème terme de la série.

On peut observer à partir des effectifs cumulés que le 1er décile est 108.

On a $259 \times 0{,}9 = 233{,}1$. Le 9ème décile est donc la valeur du 234ème terme de la série.

On peut observer à partir des effectifs cumulés que le 9ème décile est 119.

L'intervalle interdécile est donc l'intervalle $[108 ; 119]$.

L'écart interdécile est $119 - 108 = 11$.

Remarques

Il est beaucoup plus simple de déterminer la médiane, les quartiles et les déciles que de déterminer la moyenne, la variance et l'écart-type.

Dans toutes les séries :
L'intervalle interquartile contient environ 50% des données.
L'intervalle interdécile contient environ 80% des données.

On peut définir de façon semblable les notions de 1^{er} centile et $99^{ème}$ centile qui sont parfois utilisées.

Solution de l'exercice 6 :

1°) En organisant les données, on obtient le tableau suivant :

Cholestérol	130	140	150	160	170	180	190	200	210	220	230	240
Effectif	1	1	2	2	4	5	8	12	10	17	15	17
Effectif cumulé	1	2	4	6	10	15	23	35	45	62	77	94

Cholestérol	250	260	270	280	290	300	310	320	330	340	350	360
Effectif	21	12	19	10	7	13	6	5	6	1	4	2
Effectif cumulé	115	127	146	156	163	176	182	187	193	194	198	200

2°) La série a un nombre pair de terme : 200 et on a $\frac{200}{2} = 100$.

La médiane est donc la demi-somme des valeurs du $100^{ème}$ et du $101^{ème}$ terme .
On peut observer à partir des effectifs cumulés que ces deux termes correspondent à la valeur 250.
La médiane de cette série est donc 250.

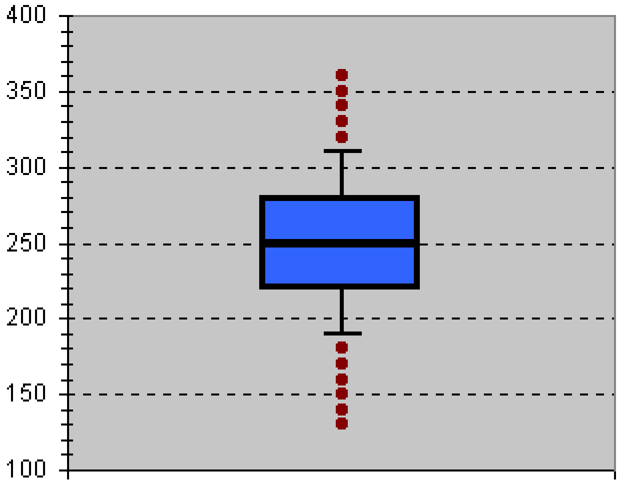
On a $200 \times 0,25 = 50$. Le 1^{er} quartile est donc la valeur du $50^{ème}$ terme de la série c'est-à-dire 220.

On a $200 \times 0,75 = 150$. Le $3^{ème}$ quartile est donc la valeur du $150^{ème}$ terme de la série c'est-à-dire 280.

On a $200 \times 0,1 = 20$. Le 1^{er} décile est donc la valeur du $20^{ème}$ terme de la série c'est-à-dire 190.

On a $200 \times 0,9 = 180$. Le $9^{ème}$ décile est donc la valeur du $180^{ème}$ terme de la série c'est-à-dire 310.

3°) On peut alors tracer le diagramme en boîte correspondant :



Solution de l'exercice 7:

2) $\overline{x} = \dfrac{55 \times 15 + 65 \times 18 + 75 \times 15 + 85 \times 23 + 95 \times 34 + 105 \times 21 + 115 \times 15 + 125 \times 9}{15+18+15+23+34+21+15+9}$

$= \dfrac{13360}{150} \approx \mathbf{89{,}07\ cm}$

Variance = $\dfrac{1249350}{150} - 89{,}07^2 \approx 395{,}54$

$\boldsymbol{\sigma} = \sqrt{395{,}54} \approx \mathbf{19{,}89\ cm}$

3) $\dfrac{150}{2} = 75$ donc méd $\in$ [90 ; 100[. On applique alors une interpolation linéaire à l'aide du théorème de Thalès, en partant du polygone des effectifs cumulés croissants. On obtient alors :

$\dfrac{Me - 90}{100 - 90} = \dfrac{75 - 71}{105 - 71} = \dfrac{Me - 90}{10} = \dfrac{4}{34}$

Me $= 90 + \dfrac{40}{34} \approx \mathbf{91{,}176\ cm}$

$\dfrac{150}{4} = 37{,}5$ donc Q1 $\in$ [70 ; 80[. Par un même raisonnement que pour la médiane, nous avons :

$\dfrac{Q_1 - 70}{70 - 60} = \dfrac{37{,}5 - 33}{48 - 33}$ donc $\mathbf{Q_1 \approx 73\ cm}$

$\dfrac{3 \times 150}{4} = 112{,}5$ donc $Q_3 \in$ [100 ; 110[. On a : $\dfrac{Q_3 - 100}{110-100} = \dfrac{112{,}5 - 105}{21}$ donc $Q_3 = 100 + \dfrac{75}{21}$

$\mathbf{Q_3 \approx 103{,}57\ cm}$

4) La plage de normalité est $[\ \overline{x} - 2\sigma\ ;\ \overline{x} + 2\sigma\ ] = [49{,}29\ ;\ 128{,}85] \approx [50\ ;\ 130]$ ce qui représente environ 100 % de la population étudiée.

Solution de l'exercice 8 :

2) $\overline{x} = \dfrac{5 \times 175 + 15 \times 185 + 22 \times 195 + 23 \times 205 + 20 \times 215 + 14 \times 225 + 12 \times 235 + 9 \times 245}{120} = \dfrac{25130}{120}$

$\overline{\mathbf{x}} \approx \mathbf{209{,}42\ cm}$

$\sigma^2 = \dfrac{5 \times 175^2 + 15 \times 185^2 + 22 \times 195^2 + 23 \times 205^2 + 20 \times 215^2 + 14 \times 225^2 + 12 \times 235^2 + 9 \times 245^2}{120} - \dfrac{25130^2}{120^2}$

$= \dfrac{5305800}{120} - \dfrac{631516900}{14400} = \dfrac{636696000-631516900}{14400} = \dfrac{5179100}{14400}$

$\sigma = \sqrt{\dfrac{5179100}{14400}}$

$\boldsymbol{\sigma} \approx \mathbf{18{,}96\ cm}$

3) $\dfrac{120}{2} = 60$ donc Me $\in$ [200 ; 210[. On applique alors une interpolation linéaire à l'aide du théorème de Thalès, en partant du polygone des effectifs cumulés croissants. On obtient alors :

$\dfrac{Me - 200}{210 - 200} = \dfrac{60-42}{65-42} = \dfrac{18}{23}$

Donc Me $= 10 \times \dfrac{18}{23} + 200$

Me $= \dfrac{180 + 4600}{23}$

Me $= \dfrac{4780}{23}$

Me $\approx 207{,}83$ cm

- $\frac{120}{4} = 30$ donc $Q_1 \in [190 ; 200[$

$\frac{Q_1 - 190}{200 - 190} = \frac{30\text{-}20}{42\text{-}20} = \frac{10}{22}$

$Q_1 = 190 + 10 \times \frac{10}{22}$

$Q1 = \frac{4180 + 100}{22} = \frac{4280}{22}$

$\mathbf{Q_1 \approx 194{,}55}$ **cm**

- $120 \times \frac{3}{4} = 90$ donc $Q_3 \in [220 ; 230[$

$\frac{Q_3 - 220}{230\text{-}220} = \frac{90 - 85}{99 - 85} = \frac{5}{14}$

$Q_3 = 10 \times \frac{5}{14} + 220$

$Q_3 = \frac{50 + 3080}{14}$

$Q_3 = \frac{3130}{14}$

$\mathbf{Q_3 \approx 223{,}57}$ **cm**

4) $[\,\overline{x} - \sigma ; \overline{x} + \sigma[= [190{,}46 ; 228{,}38[\approx [190 ; 230]$

$\frac{100 \times 79}{120} = 65{,}8\ \%$

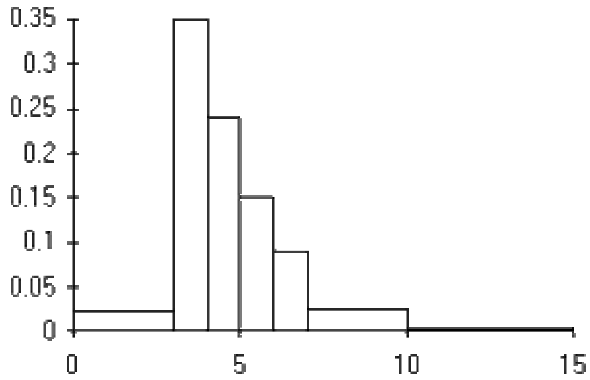
65,8 % de la population a une performance comprise entre 190 cm et 230 cm.

Solution de l'exercice 9:

1°/ Histogramme des densités de fréquences.

La densité de fréquence d_i de la i^e classe est, par définition, la fréquence f_i de la classe divisée par l'amplitude de la classe :

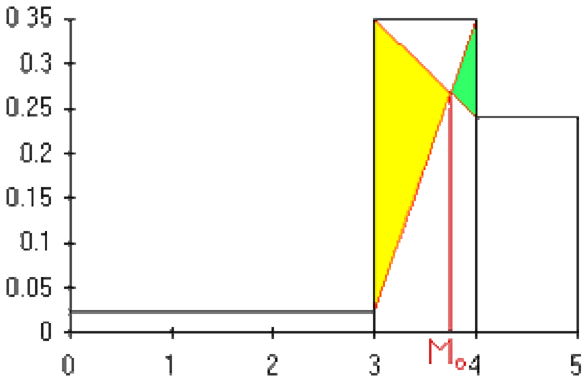
X (milliers de F)	%	f_i	d_i	$\sum f_i$
[0 ; 3 [	7	0,07	0,023	0,07
[3 ; 4 [	35	0,35	0,350	0,42
[4 ; 5 [	24	0,24	0,240	0,66
[5 ; 6 [	15	0,15	0,150	0,81
[6 ; 7 [	9	0,09	0,009	0,90
[7 ; 10 [	8	0,08	0,027	0,98
[10 ; 15 [	2	0,02	0,004	1,00



2°/ Détermination du mode

La classe modale est la classe qui possède la plus forte densité de fréquence : sur l'histogramme précédent, on voit que c'est la classe [3 ; 4[.
Lorsqu'on suppose que la répartition des salaires à l'intérieur d'une classe est uniforme, le mode se détermine en tenant compte des classes adjacentes.

Dans les triangles semblables jaune et vert, le rapport des hauteurs est égal au rapport des grands côtés :



$$\frac{M_o-3}{4-M_o}=\frac{0{,}350-\frac{0{,}07}{3}}{0{,}350-0{,}240}$$

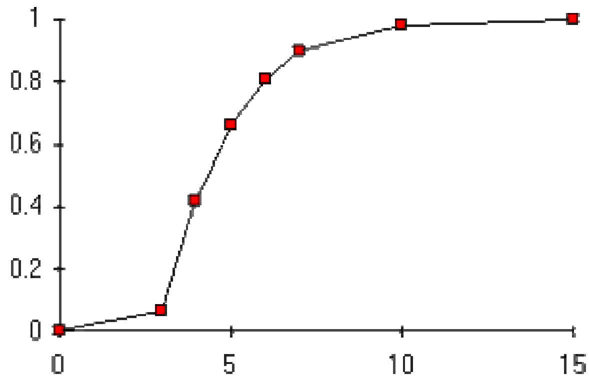
$$\frac{M_o-3}{0{,}250-\frac{0{,}07}{3}}=\frac{4-M_o}{0{,}350-0{,}240}=\frac{1}{0{,}700-0{,}240-\frac{0{,}07}{3}}$$

$$M_o=3+\frac{1{,}05\text{-}0{,}07}{1{,}38\text{-}0{,}07}=3+\frac{98}{131}\approx 3{,}75$$

3°/ Fonction de répartition empirique

La fonction de répartition empirique s'obtient en joignant, par des segments de droites, les points ayant, pour abscisses, les bornes supérieures des classes et, pour ordonnées, les fréquences cumulées des classes.
C'est ce qu'on appelle aussi le **diagramme cumulatif des fréquences**.

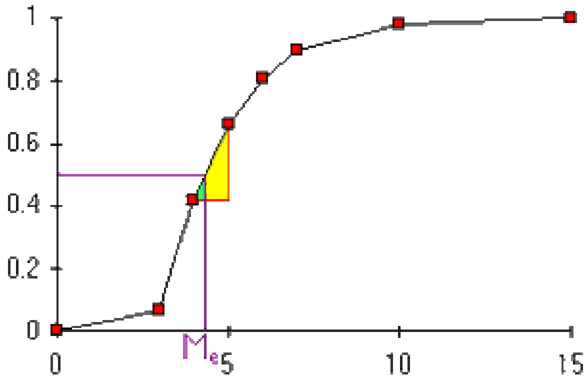
X (milliers de F)	$\sum f_i$
0	0,00
3	0,07
4	0,42
5	0,66
6	0,81
7	0,90
10	0,98
15	1,00



4°/ Médiane

La médiane est la valeur de salaire qui partage en deux parties de même cardinal l'ensemble des agents civils de l'état.
On l'obtient en prenant l'abscisse du point d'ordonnée 0,500 sur le diagramme cumulatif des fréquences précédent.

La valeur de la médiane s'obtient, dans la classe médiane [4 ; 5[, par interpolation linéaire entre les points (4 ; 0,42) et (5 ; 0,66) :



$$\frac{M_e - 4}{5 - 4} = \frac{0{,}50 - 0{,}42}{0{,}66 - 0{,}42}$$

$$M_e = 4 + \frac{8}{24} = 4 + \frac{1}{3} \approx 4{,}33$$

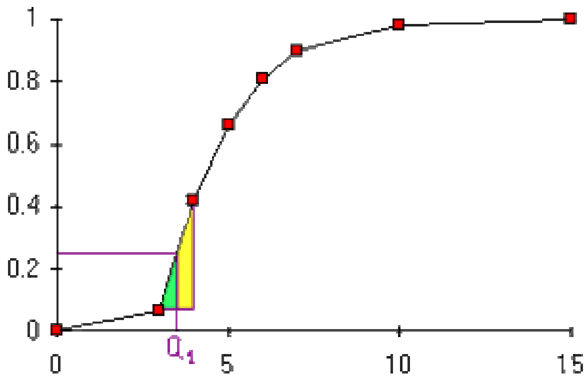
La médiane est aussi le deuxième quartile :

$$Q_2 = M_e = 4{,}33$$

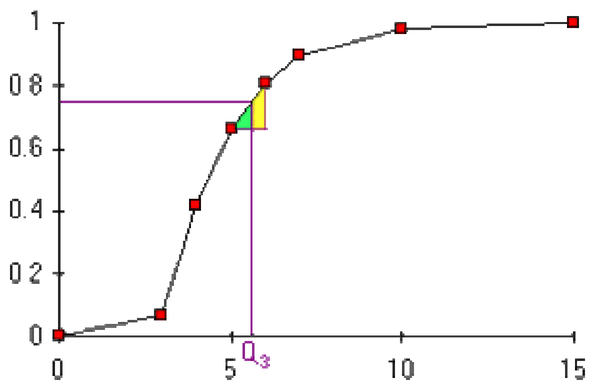
5°/ Intervalle interquartile

Le premier quartile Q_1 a, dans le diagramme précédent, pour ordonnée, 0,25.
Q_1 appartient donc à la classe [3 ; 4[et, dans cette classe, on l'obtient par interpolation linéaire entre les points (3 ; 0,07) et (4 ; 0,42).

$$Q_1 = 3 + 0{,}25 - \frac{0{,}07}{0{,}42\text{-}0{,}07} = 3 + \frac{18}{35} \approx 3{,}51$$



Le troisième quartile Q_3 a, dans le diagramme précédent, pour ordonnée, 0,75.
Q_3 appartient donc à la classe [5 ; 6[et, dans cette classe, on l'obtient par interpolation linéaire entre les points (5 ; 0,66) et (6 ; 0,81).



$$Q_3 = 5 + \frac{0{,}75\text{-}0{,}66}{0{,}81\text{-}0{,}66} = 5 + \frac{9}{15} = 5 + \frac{3}{5} = 5{,}60$$

L'intervalle interquartile est la différence entre Q_3 et Q_1 :

$$Q_3 - Q_1 = 5{,}60 - 3{,}51 = 2{,}09$$

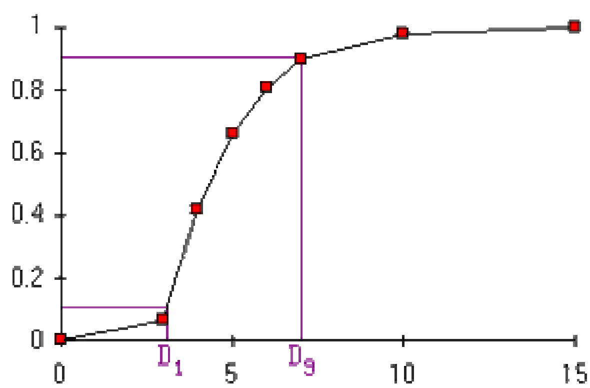
6°/ Rapport des déciles

Le premier décile D_1 a, pour ordonnée dans le diagramme cumulatif des fréquences, 0,10.
D_1 appartient à la classe [3 ; 4[et, dans cette classe, on l'obtient par interpolation linéaire entre les points (3 ; 0,07) et (4 ; 0,42).

$$D_1 = 3 + \frac{0{,}10\text{-}0{,}07}{0{,}42\text{-}0{,}07} = 3 + \frac{3}{35} \approx 3{,}09$$

Le neuvième décile D_9 a, pour ordonnée dans le diagramme cumulatif des fréquences, 0,90.

$D_9 = 7$



Le rapport des déciles est $\frac{\text{D9}}{\text{D1}} = \frac{245}{108} \approx 2{,}27$.

Le rapport des déciles est proportionnel à D_9 et inversement proportionnel à D_1 : il augmente si la proportion de gens touchant un salaire élevé augmente ou si la proportion de gens touchant un salaire faible diminue.
Ce rapport est donc une mesure des écarts de salaires chez les agents civils de l'état.

Sachant que ce rapport valait 2,4 en 1976, on voit qu'il a diminué en 4 ans : cela signifie que les écarts de salaires ont diminué dans cette période.

Solution de l'exercice 10:

1) $\mathbf{\overline{x} = 1230}$ **€** et **$\boldsymbol{\sigma}$ = 248,79 €**

2) $\overline{x'}$ **= 1,03 $\overline{\mathbf{x}}$ + 20 = 1286,90 €**

$\boldsymbol{\sigma}$' = 1,03$\boldsymbol{\sigma}$ = 1,03 ×248,79 = 256,25 €

3) On multiplie par $\frac{1}{\sigma} = \frac{1}{248,79} \approx 0,004$

puis on ajoute $\frac{-\overline{x}}{\sigma} = \frac{-1230}{248,79} \approx -4,94$.

Solution de l'exercice 11:

1) Dressons d'abord un tableau avec les notes concentrées réduites :

	Poids	Zpoids	Hauteur	Zhauteur	60m	Z60m
A	6.50 m	0.75	1.10 m	-1	9.4s	-0.66
B	5.20 m	-2.5	1.25 m	2	8.7s	1.66

On en déduit que pour A, la meilleure performance est en poids et sa moins bonne est en hauteur. On peut déduire de la même manière pour B que sa meilleure performance est en hauteur et sa moins bonne en poids.
Remarque : Quand il s'agit de temps, on ne calcule pas Ztemps en faisant x- M mais en faisant M – x.

2) Meilleure performance de B : 2 en hauteur

$$\frac{XA_{poids} - M_{poids}}{\sigma_{poids}} = \frac{1,25 - 1,15}{0,05}$$

ZApoids = ZBhauteur

$$XApoids \frac{-6,20}{0,40} = \frac{1,25 - 1,15}{0,05}$$

XApoids = 7 m

L'élève A doit réaliser 7 m au poids pour égaler la meilleure performance de l'élève B.