Index

1 page 180 Diagramme en boîte

Rappel des définitions : médiane et quartiles

Définition du premier quartile :

Le premier quartile Q_1 est la plus petite valeur de la série telle que 25 % au moins des effectifs ont une valeur inférieure ou égale à Q_1.

Important : Dans cette définition, le quartile Q_1 est une valeur de la série

(*il existe une autre définition où Q_1 n'est pas nécessairement une valeur de la série et que certains logiciels utilisent).*

En pratique : méthode de détermination de Q_1 :

on range la série dans l'ordre des valeurs croissantes.

On divise l'effectif N par 4 (25% de l'effectif). ($q_1 = \frac{N}{4}$).

Si q_1 est un entier, Q_1 est la valeur du $q_1{}^{\text{ième}}$ élément.

Si q_1 n'est pas un entier, on prend l'entier q_1' juste au-dessus, et Q_1 est la valeur du $q_1'{}^{\text{ième}}$ élément.

Définition du troisième quartile :

Le troisième quartile Q_3 est la plus petite valeur de la série telle que 75 % au moins des effectifs ont une valeur inférieure ou égale à Q_3.

Important :Dans cette définition, le quartile Q_3 est une valeur de la série

En pratique : méthode de détermination de Q_3 :

on range la série dans l'ordre des valeurs croissantes.

On multiplie l'effectif N par $\frac{3}{4}$ (75% de l'effectif). ($q_3 = \frac{3N}{4}$).

Si q_3 est un entier, Q_3 est la valeur du $q_3{}^{\text{ième}}$ élément.

Si q_3 n'est pas un entier, on prend l'entier q_3' juste au-dessus, et Q_3 est la valeur du $q_3'{}^{\text{ième}}$ élément.

Définition de la médiane Me :

La médiane est le nombre réel tel que la série ordonnée est partagée en deux séries d'effectif égal.

50% des valeurs de la série sont inférieures ou égales à la médiane Me, 50% des valeurs de la série sont supérieures ou égales à la médiane Me.

Remarque : la médiane Me n'est pas nécessairement une valeur de la série.

En pratique : méthode de détermination de Me :

on range la série dans l'ordre des valeurs croissantes.

Si l'effectif N est impair, la médiane Me est la $\frac{N+1}{2}$ ième valeur de la série.

Si l'effectif N est pair, la médiane Me est la moyenne des $\frac{N}{2}$ et $\frac{N}{2}$ +1 valeurs de la série.

Quelques remarques à propos des entiers naturels :

L'ensemble des entiers naturels est noté $\mathbb{N}$.

Les entiers naturels pairs sont les multiples de 2.

Si N est un entier naturel pair, il existe un entier naturel p tel que $N = 2p$.

Les entiers naturels impairs ne sont pas divisibles par 2.

Si N est un entier naturel impair, il existe un entier naturel p tel que $N = 2p + 1$.

Si N et N + 1 sont deux entiers consécutifs, l'un est pair, l'autre est impair.

Tout nombre réel x est encadré par deux entiers consécutifs.

Soit n un entier tel que $n \leqslant x < n + 1$ (*bien noter les inégalités large à gauche, stricte à droite*)

n est la partie entière de x notée E(x).

Exemple : E(π) = 3, E(15,7) = 15, E(–5,8) = –6, $E\left(\frac{15}{4}\right) = 3$

Dans les notations utilisées supra pour les quartiles, $q'_1 = E\left(\frac{N}{4}\right) + 1$ lorsque $\frac{N}{4} \notin \mathbb{N}$.

$q'_3 = E\left(\frac{3N}{4}\right) + 1$ lorsque $\frac{3N}{4} \notin \mathbb{N}$.

Diagramme en boîte ou à moustaches :

2) ***Diagramme en boîte ou à moustaches :***

Soit une série statistique où les valeurs sont classées dans l'ordre croissant.

La valeur minimale est notée Min, la valeur maximale est notée Max.

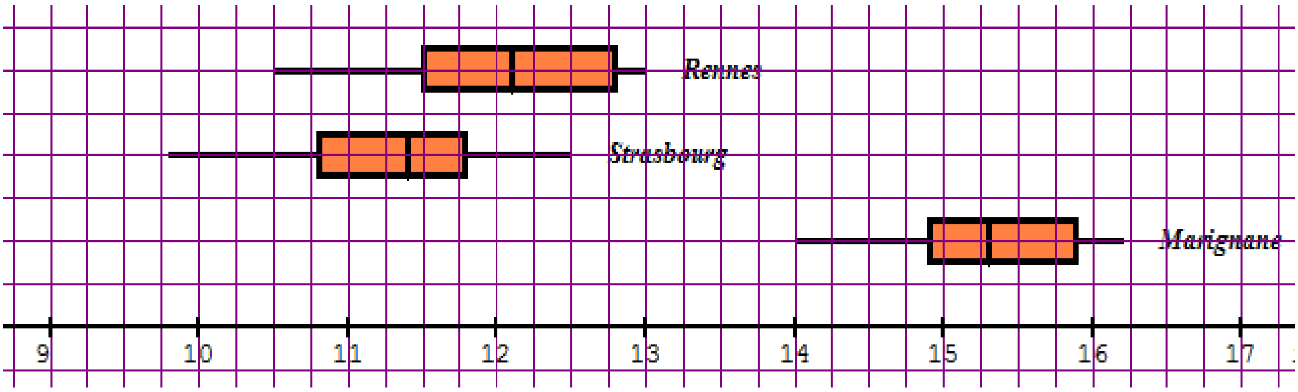
D'après la définition des quartiles Q_1 et Q_3 et de la médiane Me, on a :

$$\overbrace{Min \quad \ldots Q_1}^{25\%} \overbrace{\ldots Me \ldots}^{50\%} \overbrace{Q_3 \ldots Max}^{25\%}$$

L'intervalle interquartile [Q_1 ; Q_3] contient 50 % des effectifs. Cet intervalle est représentée par un rectangle (contenant la médiane Me).

Les intervalles [Min ; Q_1] et [Q_3 ; Max] sont représentés par des segments (moustaches).

Les diagrammes suivants sont construits avec le logiciel Sinequanon.



Villes	Minimum	Q_1	Me	Q_3	Maximum
Rennes	10,5	11,5	12,1	12,8	13
Strasbourg	9,8	10,8	11,4	11,8	12,5
Marignane	14	14,9	15,3	15,9	16,2

Les calculs pour la série de Marignane :

Après l'avoir classé dans l'ordre croissant :

14	14,3	14,4	14,5	14,8	14,9	14,9	14,9	15	15,2	15,3	15,3
15,4	15,4	15,6	15,6	15,7	15,9	15,9	16	16,1	16,2	16,2	

N = 23 ; Min = 14 ; Max = 16,2

$\frac{23}{4} = 5{,}825$ La 6ième valeur est 14,9 $Q_1 = 14{,}9$

23 est impair, on a donc la 12ième valeur pour médiane : Me = 15,3

$\frac{3\times 23}{4} = 17{,}475$ La 18ième valeur est 15,9 $Q_3 = 15{,}9$

---

2 page 181 Mesurer une dispersion

Rappel : moyenne

Soit une série statistique d'effectif total N.

Les valeurs de la série sont notées x_i et les effectifs de chaque valeur sont notés n_i .

On peut écrire sous forme d'un tableau :

Valeurs	x_1	…	x_i		x_p
Effectifs	n_1	…	n_i		n_p

$n_1 + \ldots + n_p$ = N.

La fréquence de la valeur x_i est $f_i = \frac{n_i}{N}$.

La somme des fréquences est égale à 1.

La moyenne de la série notée $\bar{x} = \dfrac{n_1 x_1 + \ldots + n_p x_p}{N} = f_1 x_1 + \ldots + f_p x_p$.

$$= \frac{1}{N}\sum_{i=1}^{i=p} n_i x_1 = \sum_{i=1}^{i=p} f_i x_1$$

Quand une série Σ est divisée en deux sous-séries Σ_1 et Σ_2 d'effectifs respectifs N_1 et N_2,

on peut calculer les moyennes $\overline{x_1}$ et $\overline{x_2}$ des sous-séries Σ_1 et Σ_2 et, on a : $\bar{x} = \dfrac{N_1\overline{x_1} + N_2\overline{x_2}}{N_1 + N_2}$

Correction de l'activité 2.

A	7	7	7	8	8	9	10	11	11	11	12	12	12	12	13
B	2	3	4	4	5	6	7	10	12	12	13	13	16	17	18	18

1) **Indicateurs de position : moyenne et médiane**

moyenne de la série A : $\overline{x_A} = \dfrac{3\times 7 + 2\times 8 + 9 + 10 + 3\times 11 + 4\times 12 + 13}{15} = 10$

moyenne de la série B : $\overline{x_B} = \dfrac{2 + 3 + 2\times 4 + 5 + 6 + 7 + 10 + 2\times 12 + 2\times 13 + 16 + 17 + 2\times 18}{16} = 10$

Médiane de la série A : $M_A = 11$ (c'est la $8^{\text{ième}}$ valeur)

Médiane de la série B : $M_B = 11$ (c'est la moyenne des $8^{\text{ième}}$ et $9^{\text{ième}}$ valeurs).

2) **Indicateurs de dispersion déjà connus : étendue, écart interquartile.**

a) Série A : étendue $e_A = 13 – 7 = 6$,

$Q_1 = 8$ (c'est la $4^{\text{ième}}$ valeur) et $Q_3 = 12$ (c'est la $12^{\text{ième}}$ valeur), d'où, $i_A = 12 – 8 = 4$

Série B : étendue $e_B = 18 – 2 = 16$,

$Q_1 = 4$ (c'est la $4^{\text{ième}}$ valeur) et $Q_3 = 13$ (c'est la $12^{\text{ième}}$ valeur), d'où, $i_B = 13 – 4 = 9$.

b) Si on remplace la valeur minimale 7 par 2 et la valeur maximale 13 par 18 dans la série A,

la moyenne n'est pas modifiée car on a enlevé 5 points et ajouté 5 points, la somme totale n'est pas modifiée.

(De façon générale, la moyenne est modifiée par le changement d'une valeur).

La médiane n'est pas modifiée car la valeur centrale dépend de l'effectif et non des valeurs extrêmes.

L'étendue est modifiée, elle vaut maintenant : 18 – 2 = 16

L'écart interquartile n'est pas modifié (même raison que pour la médiane, les $4^{\text{ième}}$ et $12^{\text{ième}}$ valeurs n'ont pas changé).

3) **Un nouvel indicateur de dispersion : écart-type.**

Pour mesurer la dispersion, on s'intéresse aux écarts à la moyenne.

Rappel : pour chaque série la moyenne est 10.

A	7	7	7	8	8	9	10	11	11	11	12	12	12	12	13	
écart relatif	–3	–3	–3	–2	–2	–1	0	1	1	1	2	2	2	2	3	
écart absolu	3	3	3	2	2	1	0	1	1	1	2	2	2	2	3	
carré des écarts	9	9	9	4	4	1	0	1	1	1	4	4	4	4	9	
B	2	3	4	4	5	6	7	10	12	12	13	13	16	17	18	18
écart relatif	–8	–7	–6	–6	–5	–4	–3	0	2	2	3	3	6	7	8	8
écart absolu	8	7	6	6	5	4	3	0	2	2	3	3	6	7	8	8
carré des écarts	64	49	36	36	25	16	9	0	4	4	9	9	36	49	64	64

a) la moyenne des écarts relatifs vaut 0 (par définition de la moyenne).

b) la moyenne des écarts absolus, l'écart absolu moyen :

Série A : $\dfrac{3\times3+2\times2+1+0+3\times1+4\times2+3}{15}=\dfrac{28}{15}$.

Série B : $\dfrac{8+7+2\times6+5+4+3+0+2\times2+2\times3+6+7+2\times8}{16}=\dfrac{78}{16}=\dfrac{39}{8}$.

Cet écart est peu utilisé en pratique.

c) La moyenne des carrés des écarts est appelée la **variance**.

Série A : $V_A=\dfrac{3\times9+2\times4+1+0+3\times1+4\times4+9}{15}=\dfrac{64}{15}$.

Série B : $V_B=\dfrac{64+49+2\times36+25+16+9+0+2\times4+2\times9+36+49+2\times64}{16}=\dfrac{474}{16}$

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

$$s_A=\sqrt{\frac{64}{15}}=\frac{8\sqrt{15}}{15}\approx 2{,}06$$

$$s_B=\frac{\sqrt{474}}{4}\approx 5{,}44$$

Remarquer :

L'écart-type est mesuré dans la même unité que la série et la moyenne.

La variance est mesurée par l'unité au carré.

Exemple : on a effectué un prélèvement d'objets de longueurs mesurées en cm.

La variance est en cm², la moyenne et l'écart-type sont en cm.

3 page 181 minimiser une dispersion

le choix de la variance (et de l'écart type) pour mesurer la dispersion n'est pas anodin ... en effet :

Valeur x_k	–2	0	3	5
Effectif n_k	1	2	4	3

1) la moyenne m : $m = \dfrac{-2+2\times0+4\times3+3\times5}{10} = 2{,}5$

La variance V : $V = \dfrac{(-2-2{,}5)^2+2\times(0-2{,}5)^2+4\times(3-2{,}5)^2+3\times(5-2{,}5)^2}{10} = \dfrac{52{,}5}{10} = 5{,}25$.

2) t est un réel, $f(t)$ est la moyenne des carrés des écarts des termes x_k de la série à ce réel t.

a) $f(t) = \dfrac{(-2-t)^2+2\times(0-t)^2+4\times(3-t)^2+3\times(5-t)^2}{10}$

$= \dfrac{1}{10}\left(4+4t+t^2+2t^2+36-24t+4t^2+75-30t+3t^2\right)$

$= t^2 - 5t + 11{,}5 = (t - 2{,}5)^2 - 6{,}25 + 11{,}5$

$= (t - 2{,}5)^2 + 5{,}25$ (forme canonique d'un polynôme du second degré)

b) Lorsque $t = 2{,}5$, f atteint son minimum qui vaut 5,25.

La fonction f atteint son minimum en $m = 2{,}5$ et la valeur minimale est $V = 5{,}25$.

Cette démarche se généralise à toute série : la moyenne est la valeur qui minimise la fonction f et cette valeur minimale est la variance.

Commentaires :

$(x_i - t)^2$ est la distance au carré entre deux nombres.

On a une série dont les valeurs sont les distances $(x_i - t)^2$.

On calcule la moyenne $f(t)$ de cette série qui est fonction de t.

La moyenne $\bar{x}$ est la valeur de t où la fonction f atteint son minimum.

La variance est la valeur minimale, soit : $V = f(\bar{x})$

---

4 page 181 Le symbole Σ.

$S = a_1 + a_2 + \ldots + a_n = \sum_{i=1}^{n} a_i$

1) $A = \sum_{i=1}^{2} \dfrac{1}{i} = \dfrac{1}{1} + \dfrac{1}{2} = \dfrac{3}{2}$.

$B = \sum_{j=0}^{3} j^3 = 0^3 + 1^3 + 2^3 + 3^3 = 36$

$C = \sum_{k=2}^{6} k(k+1) = 2\times3 + 3\times4 + 4\times5 + 5\times6 + 6\times7 = 110$

2) $S = 1 + \dfrac{1}{2} + \ldots + \dfrac{1}{10} = \sum_{i=1}^{10} \dfrac{1}{i}$ $\qquad T = (2-10)^2 + (3-10)^2 + \ldots + (15-10)^2 = \sum_{i=2}^{15} (i-10)^2$.

3) $U = \sum_{k=0}^{3} (k^2 - 3k + 4)$

Somme de 4 termes : $U = (0^2 - 3\times0 + 4) + (1^2 - 3\times1 + 4) + (2^2 - 3\times2 + 4) + (3^2 - 3\times3 + 4)$

On peut commuter les termes, factoriser les facteurs communs, associer les termes …

U = (0² + 1² + 2² +3²) –3(0 + 1 + 2 + 3) + (4 + 4 + 4 + 4)

$$U = \sum_{k=0}^{3} k^2 - 3\sum_{k=0}^{3} k + \sum_{k=0}^{3} 4 = V$$

TP3 page 189 Effet de structure

A- Comparaison de deux entreprises

Entreprise de Monsieur Dupré

Salaire S / Sexe	$1 \leqslant S < 2$	$2 \leqslant S < 3$	$3 \leqslant S < 4$	Total
Femmes	150	40	30	220
hommes	10	20	20	50
Total	160	60	50	270

Entreprise de Monsieur Fortin

Salaire S / Sexe	$1 \leqslant S < 2$	$2 \leqslant S < 3$	$3 \leqslant S < 4$	Total
Femmes	15	3	2	20
hommes	130	70	50	250
Total	145	73	50	270

Calcul des moyennes de salaire :

$$\overline{S_{\text{Dupré}}} = \frac{160\times1{,}5+60\times2{,}5+50\times3{,}5}{160+60+50} \approx 2{,}09 \text{ (en milliers d'euros).}$$

$$\overline{S_{\text{Fortin}}} = \frac{145\times1{,}5+73\times2{,}5+52\times3{,}5}{145+73+52} \approx 2{,}15 \text{ (en milliers d'euros).}$$

Calcul des moyennes de salaire par sexe :

Femmes : $$\overline{S_{\text{Femmes}_{\text{Dupré}}}} = \frac{150\times1{,}5+40\times2{,}5+30\times3{,}5}{220} \approx 1{,}95 \text{ (en milliers d'euros).}$$

$$\overline{S_{\text{Femmes}_{\text{Fortin}}}} = \frac{15\times1{,}5+3\times2{,}5+2\times3{,}5}{20} = 1{,}85 \text{ (en milliers d'euros).}$$

Hommes : $$\overline{S_{\text{Hommes}_{\text{Dupré}}}} = \frac{10\times1{,}5+20\times2{,}5+20\times3{,}5}{50} = 2{,}7 \text{ (en milliers d'euros).}$$

$$\overline{S_{\text{Hommes}_{\text{Fortin}}}} = \frac{130\times1{,}5+70\times2{,}5+50\times3{,}5}{250} = 2{,}18 \text{ (en milliers d'euros).}$$

Les salaires hommes sont plus élevés que les salaires femmes et comme l'effectif des hommes est beaucoup plus grand chez " Fortin " que chez " Dupré ", la moyenne générale est entraînée vers le haut chez " Fortin ".

Répartition en pourcentage :

Entreprise de Monsieur Dupré

Salaire S / Sexe	$1 \leqslant S < 2$	$2 \leqslant S < 3$	$3 \leqslant S < 4$	Total
Femmes	55,56	14,81	11,11	81,48
hommes	3,7	7,41	7,41	18,52
Total	59,26	22,22	18,52	100

Entreprise de Monsieur Fortin

Salaire S / Sexe	$1 \leqslant S < 2$	$2 \leqslant S < 3$	$3 \leqslant S < 4$	Total
Femmes	5,56	1,11	0,74	7,41
hommes	48,15	25,93	18,52	92,59
Total	53,7	27,04	19,26	100

Retenir la propriété suivante :

Soit une population d'effectif total N répartie en deux sous-groupes disjoints d'effectifs N_1 et N_2 ($N_1 + N_2 = N$)

Soit $\overline{x_1}$, $\overline{x_2}$, $\overline{x}$ les moyennes respectives des sous-groupes et la moyenne générale.

On a : $N\,\overline{x} = N_1\,\overline{x_1} + N_2\,\overline{x_2}$, ou encore : $\overline{x} = \dfrac{N_1 \times x_1 + N_2 \times x_2}{N}$

Preuve : Par définition de la moyenne, la somme de toutes les valeurs de la série est $N\,\overline{x}$

la somme de toutes les valeurs du premier sous-groupe est $N_1\,\overline{x_1}$

la somme de toutes les valeurs du second sous-groupe est $N_2\,\overline{x_2}$

Comme les deux sous-groupes disjoints et que leur réunion donne la population entière : $N\,\overline{x} = N_1\,\overline{x_1} + N_2\,\overline{x_2}$

B- Comparaison de deux lycées.

Lycée A	Pourcentage de candidats	Taux de réussite	Pourcentage brut de reçus
Groupe G_1	30	0,5	15
Groupe G_2	40	0,59	23,6
Groupe G_3	20	0,82	16,4
Groupe G_4	10	0,95	9,5
Effectif en %	100	XXXXXXXXX	64,5

1) Sur 100 élèves du lycée A, le nombre de reçus est : 15 + 23,6 + 16,4 + 9,5 = 64,5

le pourcentage de reçus : 64,5%.

Cellule D2 : =B2*C2 et « tirer » vers le bas en D3, D4, D5.

2) lycée B des répartitions possibles

	A	B	C	D
1		Pourcentage de candidats	Taux de réussite	Pourcentage de reçus
2	G'1	10	0,48	4,8
3	G'2	20	0,58	11,6
4	G'3	30	0,81	24,3
5	G'4	40	0,93	37,2
6	Total	100	XXXXXX	77,9

	A	B	C	D
1		Pourcentage de candidats	Taux de réussite	Pourcentage de reçus
2	G'1	5	0,48	2,4
3	G'2	20	0,58	11,6
4	G'3	30	0,81	24,3
5	G'4	45	0,93	41,85
6	Total	100	XXXXXX	80,15

	A	B	C	D
1		Pourcentage de candidats	Taux de réussite	Pourcentage de reçus
2	G'1	25	0,48	12
3	G'2	25	0,58	14,5
4	G'3	25	0,81	20,25
5	G'4	25	0,93	23,25
6	Total	100	XXXXXX	70

Cellule D6 : =SOMME(D2:D5)

Cellule B5 := B6-B2-B3-B4
(en B6 : 100, en modifiant la donnée dans une des cellules B2, B3, B4, on est certain d'avoir la somme égale à 100)

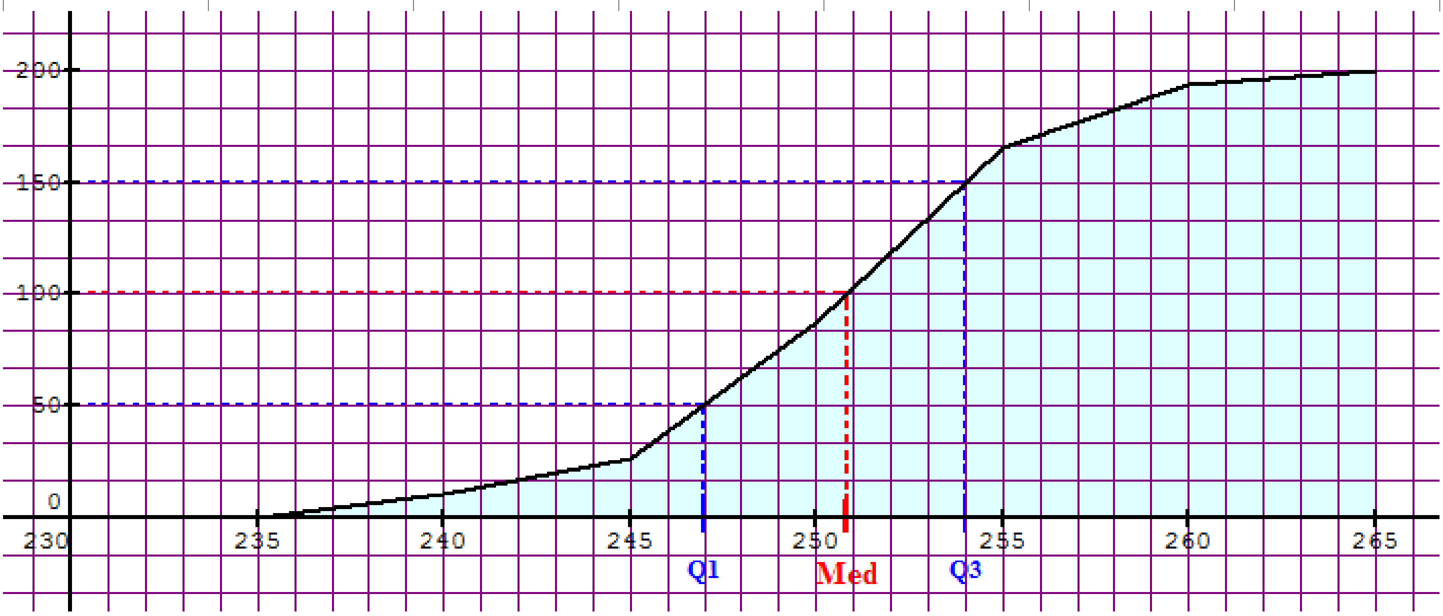
Lycée B	Pourcentage de candidats	Taux de réussite	Pourcentage brut de reçus
Groupe G'_1	10	0,48	4,8
Groupe G'_2	20	0,58	11,6
Groupe G'_3	30	0,81	24,3
Groupe G'_4	40	0,93	37,2
Effectif en %	100	XXXXXXXXX	77,9

Pour que le taux de réussite global soit plus fort dans le lycée B que dans le lycée A bien que le taux de réussite soit plus faible par jurys, il faut que le nombre de candidats dans les jurys aux taux de réussite élevés soit nettement plus important que dans les jurys aux taux de réussite faible.

10 page 191

Données de la série

Masse en g	**[235 ; 240[**	**[240 ; 245[**	**[245 ; 250[**	**[250 ; 255[**	**[255 ; 260[**	**[260 ; 265]**
Effectifs	10	16	61	79	28	6
eff. cumulés	10	26	87	166	194	200

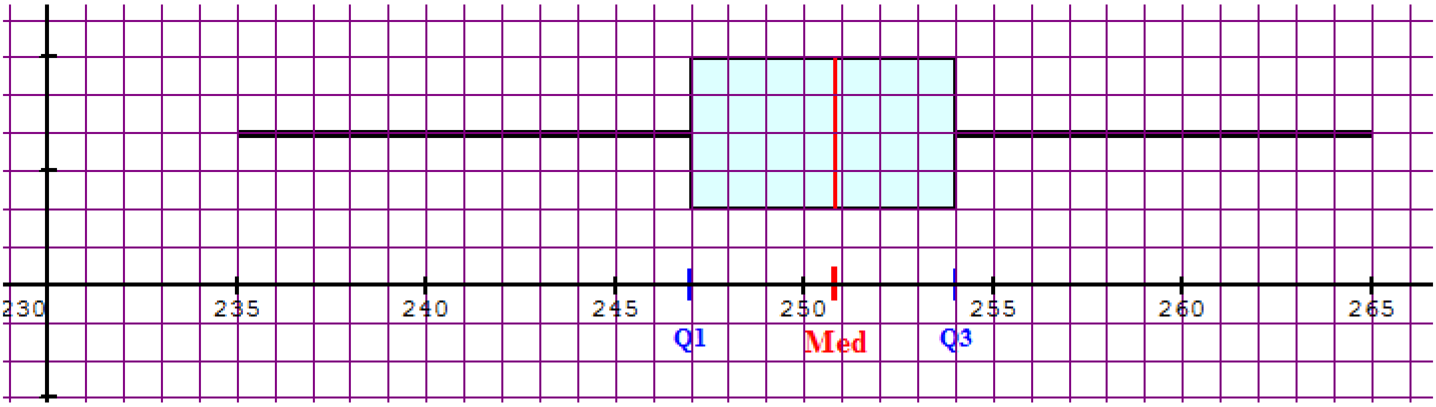


a) La médiane est la masse de la 100ième boîte : ≈ 251 g

b) le premier quartile est la masse de la 50ième boîte : ≈ 247 g

c) e premier quartile est la masse de la 150ième boîte : ≈ 254 g

d) l'écart interquartile vaut : 7 g.



La moyenne : 250,425

l'écart-type : 5,3916

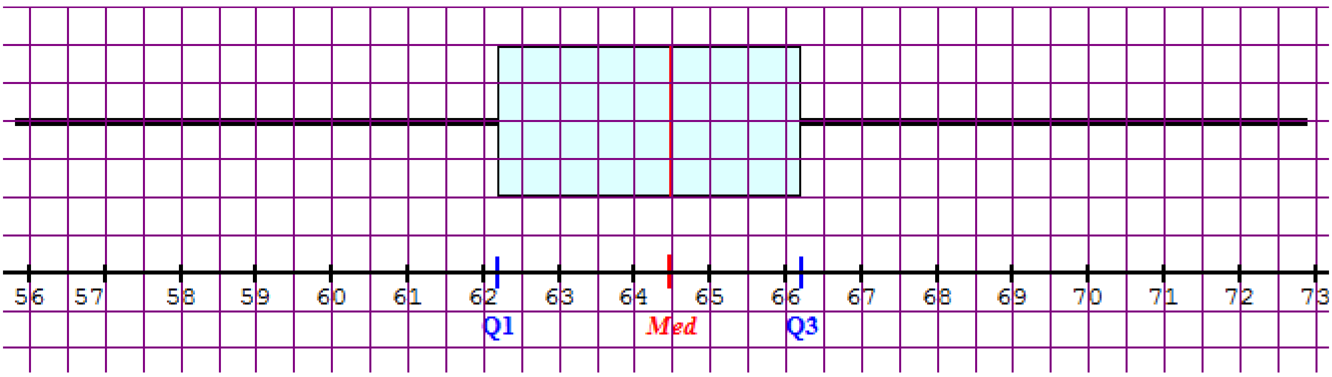
---

16 page 192 Contrôles comparatifs

Le relevé ordonné des 40 mesures :

55,8	58,7	59,7	60,3	60,7	61,2	61,2	61,4	62,1	62,2
62,3	62,4	63,1	63,4	63,4	63,6	64,1	64,4	64,4	64,5
64,5	64,6	64,8	65	65,3	65,5	65,6	65,9	66,1	66,2
66,4	67	67,1	67,6	68,7	68,8	69,7	69,8	71	72,9

1 a) Min =55,8 Q1 = 62,2 Me = $\frac{64{,}5+64{,}5}{2}$ = 64,5 Q3 = 66,2 Max= 72,9



b) Les valeurs aberrantes sont les valeurs qui n'appartiennent pas à [Q1 – 1,5I ; Q3 + 1,5I] où I = Q3 – Q1 = 4

Q1 – 1,5I = 62,2 – 6 = 56,2 Q3 + 1,5I = 66,2 + 6 = 72,2

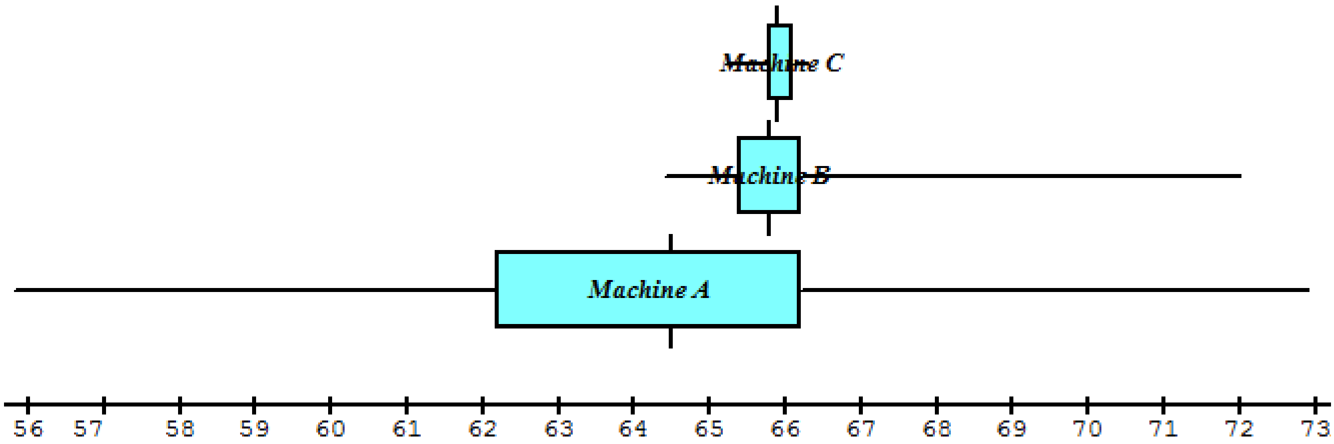
(C'est J.W. Tukey qui a qualifié de valeurs aberrantes les valeurs à l'extérieur de l'intervalle [Q1 – 1,5I ; Q3 + 1,5I]).

Les seules valeurs aberrantes sont 55,8 et 72,9 : le pourcentage de valeur aberrante est égal à 5 % : $\frac{2}{40} = \frac{5}{100}$

2) Machine B : Min = 64,4 Q1 = 65,4 Me = 65,8 Q3 = 66,2 Max= 72

Machine C : Min = 65,2 Q1 = 65,8 Me = 65,9 Q3 = 66,1 Max= 66,3

a)



b) la première série statistique ne porte que sur 40 valeurs, ce qui est peu.

les deux autres séries portent sur 1 200 valeurs ;

la machine C est plus régulière.

---

22 page 193 Contrôle de qualité

Organiser les données :

x_i	300	303	307	308	309	310	311	312	313	314	315	317	318	319	Total
n_i	1	1	1	3	2	2	3	1	1	1	1	1	1	1	20
$n_i x_i$	300	303	307	924	618	620	933	312	313	314	315	317	318	319	6213
$n_i x_i^2$	90000	91809	94249	284592	190962	192200	290163	97344	97969	98596	99225	100489	101124	101761	1930483

$m = \dfrac{6213}{20} = 310{,}65$ (en grammes)

$V = \dfrac{1930483}{20} - 310{,}65^2 = 20{,}727\,5$

$s \approx 4{,}55$ (en grammes)

L'intervalle $[m - s\ ;\ m + s]$ donne [306,1 ; 315,2], soit en arrondissant aux entiers des données : [307 ; 316] .

Cet intervalle contient : 15 barquettes.

On a donc : $\dfrac{15}{20} = \dfrac{75}{100}$ de barquettes dans cet intervalle.

Le lot est rejeté.

---

23 page 194 Contrôle de qualité

Organiser les données :

classes	[24,2 ;	[24,4 ;	[24,6 ;	[24,8 ;	[25,0 ;	[25,2 ;	[25,4 ;	[25,6 ;	[25,8 ;	Total

	24,4[	24,6[	24,8[	25,0[	25,2[	25,4[	25,6[	25,8[	26,0[	
centre : x_i	24,3	24,5	24,7	24,9	25,1	25,3	25,5	25,7	25,9	XXXX
n_i	5	13	24	19	14	10	8	5	2	100
$n_i x_i$	121,5	318,5	592,8	473,1	351,4	253	204	128,5	51,8	2494,6
$n_i x_i^2$	2952,45	7803,25	14642,16	11780,19	8820,14	6400,9	5202	3302,45	1341,62	62245,16

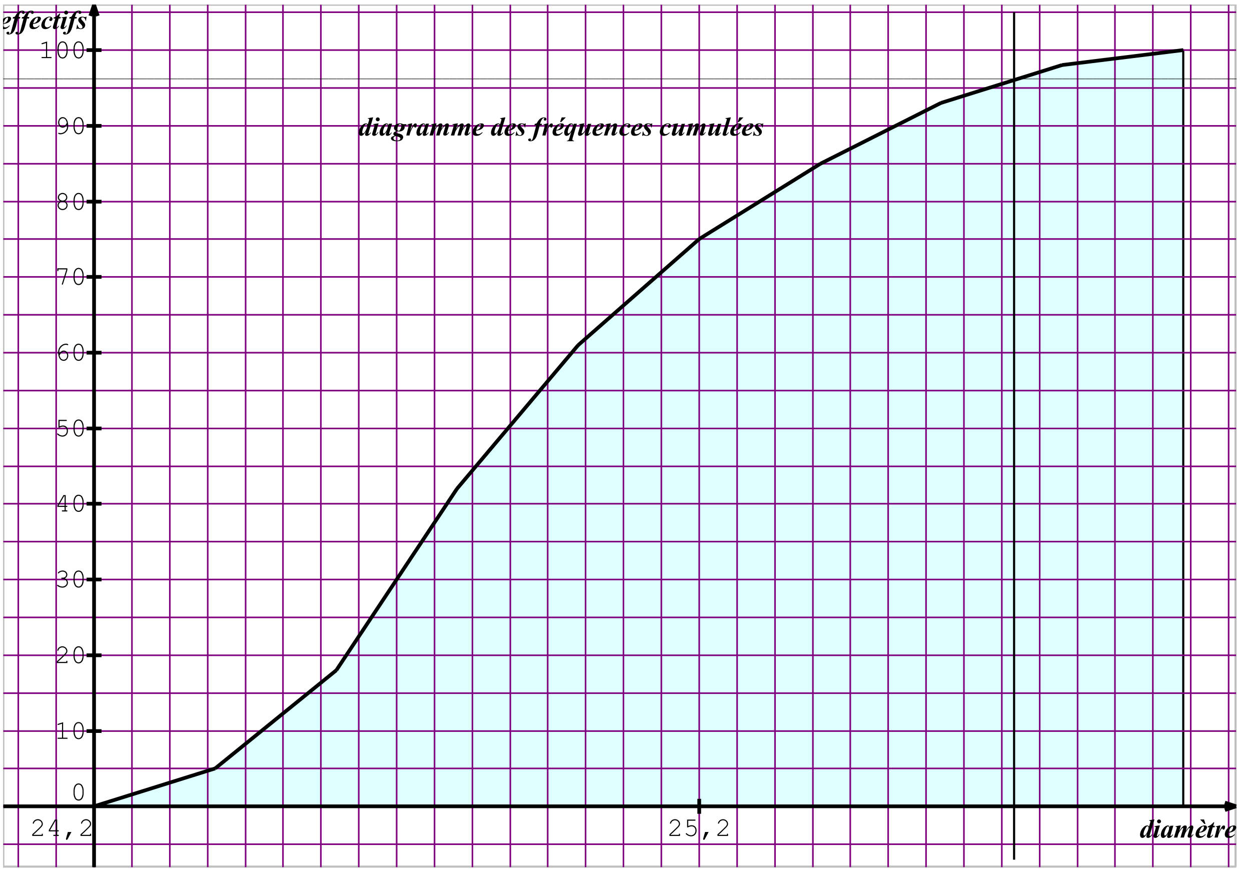
1) $\bar{x}$ = 24,946 mm

Variance = 622,4516 – 24,946² = 0,148684

s = 0,385596 mm

2) L'intervalle [$\bar{x} - 2s$; $\bar{x} + 2s$] donne [24,1748 ... ; 25,717..] plus de 90 % des effectifs.

(On suppose que les effectifs se répartissent régulièrement dans chaque classe, ce qui donne :



La production est bonne

---

24 page 194 évolution des âges

Année 1950						
Âges	[0 ; 20[	[20 ; 60[	[60 ; 65[	[65 ; 75[	[75 ; ...[	Total

Centre de la classe x_i	10	40	62,5	70	90	
fréquences en % f_i	30,1	53,7	4,8	7,6	3,8	100
$f_i x_i$	301	2148	300	532	342	3623
$f_i x_i^2$	3010	85920	18750	37240	30780	175700
âge moyen	36,23					
variance	444,39					
écart-type	21,08					

Année 1990						
Âges	[0 ; 20[	[20 ; 60[	[60 ; 65[	[65 ; 75[	[75 ; ...[	Total
Centre de la classe x_i	10	40	62,5	70	90	
fréquences en % f_i	27,8	53,2	5,1	7,1	6,8	100
$f_i x_i$	278	2128	318,75	497	612	3833,75
$f_i x_i^2$	2780	85120	19921,88	34790	55080	197691,88
âge moyen	38,34					
variance	507,15					
écart-type	22,52					

Année 2000						
Âges	[0 ; 20[	[20 ; 60[	[60 ; 65[	[65 ; 75[	[75 ; ...[	Total
Centre de la classe x_i	10	40	62,5	70	90	
fréquences en % f_i	25,6	53,8	4,6	8,8	7,2	100
$f_i x_i$	256	2152	287,5	616	648	3959,5
$f_i x_i^2$	2560	86080	17968,75	43120	58320	208048,75
âge moyen	39,6					
variance	512,72					
écart-type	22,64					

Année 2010						
Âges	[0 ; 20[	[20 ; 60[	[60 ; 65[	[65 ; 75[	[75 ; ...[	Total
Centre de la classe x_i	10	40	62,5	70	90	
fréquences en % f_i	24,3	53,0	6,0	7,9	8,8	100
$f_i x_i$	243	2120	375	553	792	4083
$f_i x_i^2$	2430	84800	23437,5	38710	71280	220657,5
âge moyen		40,83				
variance		539,49				
écart-type		23,23				

Année 2030						
Âges	[0 ; 20[	[20 ; 60[	[60 ; 65[	[65 ; 75[	[75 ; ...[	Total
Centre de la classe x_i	10	40	62,5	70	90	
fréquences en % f_i	22,6	48,1	6,1	11,2	12,0	100
$f_i x_i$	226	1924	381,25	784	1080	4395,25
$f_i x_i^2$	2260	76960	23828,13	54880	97200	255128,13
âge moyen		43,95				
variance		619,46				
écart-type		24,89				

Année 2050						
Âges	[0 ; 20[	[20 ; 60[	[60 ; 65[	[65 ; 75[	[75 ; ...[	Total
Centre de la classe x_i	10	40	62,5	70	90	
fréquences en % f_i	21,9	46,2	5,7	10,6	15,6	100
$f_i x_i$	219	1848	356,25	742	1404	4569,25
$f_i x_i^2$	2190	73920	22265,63	51940	126360	276675,63
âge moyen		45,69				

variance	678,95
écart-type	26,06

25 page 194

1)

a) $S = \sum_{k=0}^{k=5} 2k = 2\times0 + 2\times1 + 2\times2 + 2\times3 + 2\times4 + 2\times5$

(Le calcul donne $S = 30$)

Remarque : on peut factoriser 2, $S = 2 \sum_{k=0}^{k=5} k$

b) $S = \sum_{k=1}^{k=4} k^2 = 1^2 + 2^2 + 3^2 + 4^2$

(Le calcul donne $S = 30$)

c) $S = \sum_{k=4}^{k=8} k\times2^k = 4\times2^4 + 5\times2^5 + 6\times2^6 + 7\times2^7 + 8\times2^8$

(Le calcul donne $S = 3\,552$)

Remarque : On peut factoriser 2^4 : $S = 2^4 \times (4 + 5\times2 + 6\times2^2 + 7\times2^3 + 8\times2^4)$

2) a) $S = 1^3 + 2^3 + 3^3 + 4^3 + 5^3 + 6^3 = \sum_{k=1}^{k=6} k^3$

b) $S = 1\times2 + 2\times3 + 3\times4 + 4\times5 = \sum_{k=1}^{k=4} k(k+1)$

c) $S = (10-1)^2 + (10-2)^2 + (10-3)^2 + (10-4)^2 = \sum_{k=1}^{k=4} (10-k)^2$

Remarque : on peut développer S

$$\sum_{k=1}^{k=4} (10-k)^2 = \sum_{k=1}^{k=4} 10^2 - 2\times10\times \sum_{k=1}^{k=4} k + \sum_{k=1}^{k=4} k^2$$

26 page 194

Dans cet exercice, i est un indice (un numéro). Il n'entre pas dans les calculs. Les nombres qui entrent dans les calculs sont les nombres n_i et x_i .

Avec $r = 3$ i prend toutes les valeurs entières de 1 à 3

Indices (numéros) i	1	2	3
Valeurs de la série : x_i	x_1	x_2	x_3
Effectifs partiels : n_i	n_1	n_2	n_3

$d(x) = \sum_{i=1}^{r} n_i (x - x_i)^2$ pour tout réel x.

d est une fonction donnant la dispersion autour des valeurs x_i .

1 a) $r = 3$

$$d(x) = \sum_{i=1}^{3} n_i (x - x_i)^2$$

$$d(x) = n_1(x-x_1)^2 + n_2(x-x_2)^2 + n_3(x-x_3)^2$$

$$d(x) = (n_1+n_2+n_3)x^2 - 2((n_1 x_1+n_2 x_2+n_3 x_3)x + n_1 x_1^2 + n_2 x_2^2 + n_3 x_3^2$$

(avec le symbole $\sum$)

$$d(x) = \sum_{i=1}^{3} n_i (x - x_i)^2 = \sum_{i=1}^{3} n_i (x^2 - 2x_i x + x_i^2) = \sum_{i=1}^{3} n_i x^2 - \sum_{i=1}^{3} 2 n_i x_i x + \sum_{i=1}^{3} n_i x_i^2$$

$$= \left(\sum_{i=1}^{3} n_i\right) x^2 - 2\left(\sum_{i=1}^{3} n_i x_i\right) x + \sum_{i=1}^{3} n_i x_i^2$$

b) d est un polynôme du second degré.

$a = \left(\sum_{i=1}^{3} n_i\right) = \mathrm{N}$ $\quad b = -2\left(\sum_{i=1}^{3} n_i x_i\right)$ $\quad c = \sum_{i=1}^{3} n_i x_i^2$

Le coefficient de x^2 admet son minimum en $\alpha = -\dfrac{b}{2a} = \dfrac{n_1 x_1+n_2 x_2+n_3 x_3}{n_1+n_2+n_3} = \bar{x}$ *(moyenne de la série)*

La moyenne est la valeur qui minimise la fonction d.

Cette valeur minimale de d est la variance de la série multipliée par N (effectif total) : $\mathrm{N} \times \mathrm{V} = \sum_{i=1}^{3} n_i (\bar{x} - x_i)^2$

ou encore

$$d'(x) = 2 \times \left(\sum_{i=1}^{3} n_i\right) x - 2\left(\sum_{i=1}^{3} n_i x_i\right)$$

$d'(x)$ est une expression du premier degré qui s'annule en $\alpha = \dfrac{\left(\sum_{i=1}^{3} n_i x_i\right)}{\left(\sum_{i=1}^{3} n_i\right)} = \bar{x}$. ….

2) a) b) c)Les calculs précédents sont les mêmes en remplaçant 3 par r.

$$d(x) = \sum_{i=1}^{r} n_i (x - x_i)^2 = \sum_{i=1}^{r} n_i (x^2 - 2 x_i x + x_i^2) = \sum_{i=1}^{r} n_i x^2 - \sum_{i=1}^{r} 2 n_i x_i x + \sum_{i=1}^{r} n_i x_i^2$$

$$= \left(\sum_{i=1}^{r} n_i\right) x^2 - 2\left(\sum_{i=1}^{r} n_i x_i\right) x + \sum_{i=1}^{r} n_i x_i^2$$

le polynôme $d(x)$ du second degré a pour coefficients : $a = \left(\sum_{i=1}^{r} n_i\right)$, $b = -2\left(\sum_{i=1}^{r} n_i x_i\right)$ et $c = \sum_{i=1}^{r} n_i x_i^2$

Le minimum est atteint en $\alpha = -\dfrac{b}{2a} = \dfrac{\left(\sum_{i=1}^{r} n_i x_i\right)}{\left(\sum_{i=1}^{r} n_i\right)} = \bar{x}$ *(moyenne de la série)*

La moyenne est la valeur qui minimise la fonction d.

Cette valeur minimale de d est la variance de la série multipliée par N (effectif total) : $N \times V = \sum_{i=1}^{r} n_i (\bar{x} - x_i)^2$.

---

30 page 195

1) a)

Tireur A

Points : x_i	0	10	30	50	100	Total
Effectifs : n_i	1	6	2	11	5	25
$n_i x_i$	0	60	60	550	500	1170
$n_i x_i^2$	0	600	1800	27500	50000	79900

Étendue : 100

Moyenne par tir : $\dfrac{1170}{25} = 46{,}8$

Tireur B

Points : x_i	0	10	30	50	100	Total
Effectifs : n_i	2	8	3	4	8	25
$n_i x_i$	0	80	90	200	800	1170
$n_i x_i^2$	0	800	2700	10000	80000	93500

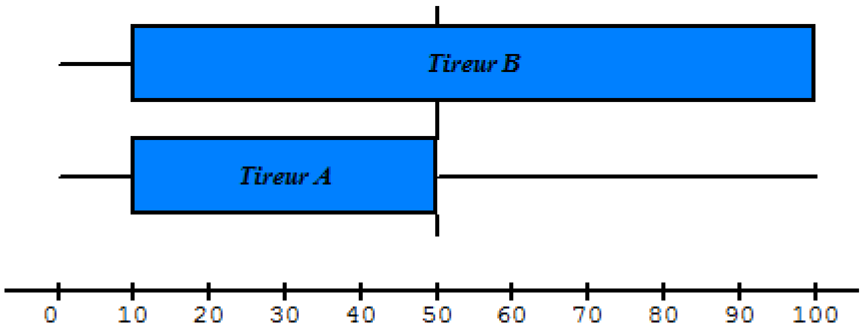
Étendue : 100

Moyenne par tir : $\dfrac{1170}{25} = 46{,}8$

b) Que doit-on observer ?
Le tireur B peut sembler plus régulier car les effectifs partiels paraissent plus équilibrés ….
Le tireur A peut sembler plus régulier car la majorité (11 tirs sur 25) des tirs est concentrée sur le « 50 ».
C'est cette dernière observation qui va apparaître ….

2 a)
Tireur A :
La médiane Me = 50 (13ème valeur)
Premier quartile Q1 = 10 (7ième valeur)
Troisième quartile Q3 = 50 (19ième valeur)
L'écart interquartile : 40

Tireur B :
La médiane Me = 30 (13ème valeur)
Premier quartile Q1 = 10 (7ième valeur)
Troisième quartile Q3 = 100 (19ième valeur)
L'écart interquartile : 90



(Tireur A : comme Me = Q3, on ne voit pas apparaître sur le diagramme les 25 % entre Me et Q3)
(Tireur B : comme Q3 = max, on ne voit pas apparaître les 25 % extérieurs à Q3)

b) Variance et écart-type :
Tireur A :
$V = \dfrac{79900}{25} - 46{,}8^2 = 1005{,}76$
écart-type : $\sigma \approx 31{,}713\,7$

Tireur B :
$V = \dfrac{93500}{25} - 46{,}8^2 = 1549{,}76$
écart-type : $\sigma \approx 39{,}367$

Les indicateurs de dispersion sont plus faibles pour le tireur A.

53 page 198 ***Travailler la démonstration***
1) ***Cas particulier***
La série statistique est partagée en deux sous-séries disjointes :
sous-série 1 : $x_1, x_2, \dots, x_k$ de moyenne $\bar{x}$ *(On a donc k valeurs)*
sous-série 2 : $y_1, y_2, \dots, y_r$ de moyenne $\bar{y}$ *(On a donc r valeurs)*
a) l'effectif total de la série est : $k + r$
b) La moyenne m de cette série est égale à : $m = \dfrac{(x_1+x_2+\dots+x_k)+(y_1+y_2+\dots+y_r)}{k+r}$

c) Par définition de la moyenne $\bar{x}$, on sait : $x_1 + x_2 + \ldots + x_k = k\,\bar{x}$.

d) De même, on sait que $y_1 + y_2 + \ldots + y_r = r\,\bar{y}$

En remplaçant ces sommes dans la définition de *m* au b), il vient : $m = \dfrac{k\,\bar{x}+r\,\bar{y}}{k+r}$

2) ***Cas général***.

Une série statistique X est partagée en *p* sous-séries X_i disjointes de moyennes et d'effectifs :
$(\bar{x}_1, n_1)$, $(\bar{x}_2, n_2)$, … , $(\bar{x}_p, n_p)$.

a) Effectif total de la série X est $N = n_1+n_2+\ldots+n_p$.

b) Pour chaque sous-série X_i , la somme des valeurs est $n_i.\bar{x}_i$.

Donc la somme des valeurs de la série X est : $S = \sum_{i=1}^{i=p} n_i.\bar{x}_i$

c) La moyenne $\bar{x}$ de la série *X* est donc : $\bar{x} = \dfrac{1}{N}\sum_{i=1}^{i=p} n_i.\bar{x}_i = \dfrac{n_1.\bar{x}_1+n_2.\bar{x}_2+\ldots+n_p.\bar{x}_p}{n_1+n_2+\ldots+n_p}$

---

54 page 198 Travailler la démonstration

Dans cet exercice, *i* est un indice (un numéro). Il n'entre pas dans les calculs. Les nombres qui entrent dans les calculs sont les nombres n_i et x_i .

Avec $r = 3$ *i* prend toutes les valeurs entières de 1 à 3

Indices (numéros) *i*	1	2	3
Valeurs de la série : x_i	x_1	x_2	x_3
Effectifs partiels : n_i	n_1	n_2	n_3

Variance de la série : $V = \dfrac{1}{N}\sum_{i=1}^{r} n_i(\bar{x}-x_i^2)$ (C'est la définition)

V est un nombre réel défini par les données. Ce n'est pas une valeur variable...

On est amené à faire un calcul algébrique (développement, factorisation, réduction, ….)

On a besoin de $(a - b)^2 = \ldots.$ *développement du carré de la différence de a* et *b.*

$a(b + c) = \ldots.$ *distributivité de la multiplication sur l'addition (pour développer)*

$ab + ac = \ldots..$ *idem (pour factoriser)*

Voir l'activité 4 de ce chapitre pour la " manipulation " du $\sum$.

a) $r = 3$

$V = \dfrac{1}{N}\sum_{i=1}^{3} n_i(\bar{x}-x_i^2) = \dfrac{1}{N}\left(\sum_{i=1}^{3} n_i(\overline{x^2}-2\,x_i\,\bar{x}+x_i^2)\right)$ *développement du carré de ......*

$$= \frac{1}{N} \sum_{i=1}^{3} n_i \overline{x^2} - \frac{1}{N} \sum_{i=1}^{3} 2 n_i x_i \overline{x} + \frac{1}{N} \sum_{i=1}^{3} n_i x_i^2$$ *distributivité de .... sur ....*

$$= \frac{1}{N} \left(\sum_{i=1}^{3} n_i \right) \overline{x}^2 - 2 \frac{1}{N} \left(\sum_{i=1}^{3} n_i x_i \right) \overline{x} + \frac{1}{N} \sum_{i=1}^{3} n_i x_i^2$$ *distributivité de .... sur .... (factorisation)*

Or, $N = \left(\sum_{i=1}^{3} n_i \right)$ et $\frac{1}{N} \left(\sum_{i=1}^{3} n_i x_i \right) = \overline{x}$, d'où

$$V = \overline{x}^2 - 2 \times \overline{x} \times \overline{x} + \frac{1}{N} \sum_{i=1}^{3} n_i x_i^2$$

$$= \frac{1}{N} \sum_{i=1}^{3} n_i x_i^2 - \overline{x}^2.$$

b) En remplaçant 3 par *r* dans la démarche :

on montre que $V = \frac{1}{N} \sum_{i=1}^{r} n_i x_i^2 - \overline{x}^2$. (Les propriétés utiles sont les mêmes)