

Séries statistiques à caractère quantitatif

1. Présentation d'une série statistique à caractère quantitatif	p2	3. Exemple	p4
2. Paramètres d'une série statistique à caractère quantitatif	p3	4. Répartition en classes	p9

1. Présentation d'une série statistique à caractère quantitatif

1.1. Rappels

Pour définir une série statistique à caractère quantitatif, on se donne une suite de p ($p \in \mathbb{N}^*$) nombres réels distincts classés dans l'ordre croissant correspondant aux p valeurs $x_1 \dots x_i \dots x_p$ du caractère étudié et une suite de p entiers naturels correspondant aux effectifs n_i associés aux valeurs x_i .

On donne souvent une telle série sous la forme d'un tableau.

Valeurs du caractère	x_1	x_2		x_i		x_p
Effectifs	n_1	n_2		n_i		n_p

$$x_1 < x_2 < \dots < x_i < \dots < x_p$$

1.2. Effectif total

$$\text{Effectif total: } n_1 + n_2 + \dots + n_i + \dots + n_p = \sum n_i$$

1.3. Effectifs cumulés croissants

Pour x_1 : n_1

Pour x_2 : $n_1 + n_2$

Pour x_3 : $n_1 + n_2 + n_3$

Pour x_i : $n_1 + n_2 + n_3 + \dots + n_i$

Pour x_p : $n_1 + n_2 + n_3 + \dots + n_i + \dots + n_p = N$

1.4. Fréquences

$$\text{On nomme fréquence de la valeur } x_i \text{ le nombre réel: } f_i = \frac{n_i}{N}$$

$$\text{On a: } f_1 + f_2 + \dots + f_p = 1$$

1.5. Fréquences cumulées croissantes:

Pour x_1 : f_1

Pour x_2 : $f_1 + f_2$

Pour x_3 : $f_1 + f_2 + f_3$

Pour x_i : $f_1 + f_2 + f_3 + \dots + f_i$

Pour x_p : $f_1 + f_2 + f_3 + \dots + f_i + \dots + f_p = 1$

1.6. Fréquences en pourcentage

On peut définir les fréquences en pourcentage en multipliant les fréquences par 100.

1.7. Représentations graphiques

Soit un nuage de points ou soit un diagramme (en bâtons, circulaire,)

2. Paramètres d'une série statistique à caractère quantitatif

2.1. Le mode

Le(s) mode(s) d'une série statistique est (sont) la (les) valeur(s) ayant le plus grand effectif.

2.2. La moyenne

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_i x_i + \dots + n_p x_p}{N}$$

Remarque: $\bar{x} = f_1 x_1 + f_2 x_2 + \dots + f_i x_i + \dots + f_p x_p$

2.3. La médiane

La médiane d'une série statistique partage les valeurs de cette série en deux parties de même effectif.

2.4. Détermination de la médiane

Pour déterminer la médiane, on peut écrire la liste des N valeurs du caractère dans l'ordre croissant c'est à dire si $n_1=2$ et $n_2=3$, on obtient pour le début de la liste: $x_1; x_1; x_2; x_2; x_2; \dots$

- Si N est un nombre impair: $N=2n+1$ (exemple: $17=2 \times 8+1$) alors la médiane est la $(n+1)^{\text{ième}}$ nombre de la liste (pour l'exemple: $9^{\text{ième}}$ nombre de la liste)
- Si N est un nombre pair: $N=2n$ (exemple: $22=2 \times 11$) alors la médiane est la demi-somme des $n^{\text{ième}}$ et $(n+1)^{\text{ième}}$ de la liste (pour l'exemple: la demi-somme de la $11^{\text{ième}}$ et de la $12^{\text{ième}}$ valeur).

2.5. Les quartiles

Le premier quartile est le plus petit nombre Q_1 de la liste (des N valeurs du caractère dans l'ordre croissant) tel que au moins un quart des nombres de la liste soit inférieurs ou égaux à Q_1 .

Le troisième quartile est le plus petit nombre Q_3 de la liste (des N valeurs du caractère dans l'ordre croissant) tel que au moins trois quarts des nombres de la liste soit inférieurs ou égaux à Q_3 .

2.6. Détermination des quartiles

- Si N est un multiple de 4: $N=4n$ (exemple: $N=24=4\times 6$) alors Q_1 est le $n^{\text{ième}}$ nombre de la liste et Q_3 est le $(3n)^{\text{ième}}$ nombre de la liste (pour l'exemple: Q_1 est le $6^{\text{ième}}$ nombre de la liste et Q_3 est le $18^{\text{ième}}$ nombre de la liste).
- Si N n'est pas un multiple de 4 (exemple: $N=29$). On considère le plus petit multiple de 4 strictement supérieur à N : $N'=4n'$ (pour l'exemple: $N'=32=4\times 8$, $n'=8$). Q_1 est le $n'^{\text{ième}}$ nombre de la liste (pour l'exemple Q_1 est le $8^{\text{ième}}$ nombre de la liste).

Puis on considère le plus petit multiple de 4 supérieur à $3N$ $N''=4n''$ (pour l'exemple $3\times 29=87$; $N''=88=4\times 22$; $n''=22$). Q_3 est le $n''^{\text{ième}}$ nombre de la liste. (pour l'exemple Q_3 est le $22^{\text{ième}}$ nombre de la liste)

2.7. Étendue

L'étendue de la série est la différence entre le plus grand nombre de la liste et le plus petit nombre de la liste: $x_p - x_1$

2.8. Écart inter-quartile

On nomme écart inter-quartile, la différence entre le troisième quartile et le premier quartile: $Q_3 - Q_1$

3. Exemple

3.1. Remarque

Dans le programme de seconde paru au BO du 23 juillet 2009, il est précisé que les élèves doivent: utiliser un logiciel (par exemple un tableur ou une calculatrice pour étudier une série statistique. Nous choisissons de vous proposer l'utilisation d'un tableur.

3.2. Énoncé

Dans une librairie, pour le rayon enfant, il existe pour les livres, huit tarifs différents: 10€; 13€; 15€; 17€; 20€; 23€; 25€ et 27€. Le nombre de livres vendus pour une semaine est respectivement 133; 115; 107; 215; 160; 10; 31; 29.

On définit ainsi une série statistique à caractère quantitatif.

On se propose d'utiliser un tableur pour déterminer tous les éléments et paramètres définis précédemment.

3.3. Présentation du tableur

a) On écrit dans un tableau (horizontalement) les deux suites de nombres, $p=8$

Les huit x_i sont écrits dans les cellules de B2 à I2.

Les huit n_i sont écrits dans les cellules de B3 à I3.

b) Pour calculer l'effectif total N , on doit calculer la somme des nombres des cellules de B3 à I3 et préciser la cellule où on place le résultat.

On veut écrire le résultat en K3(par exemple). Dans la cellule J3, on écrit: « effectif total » et dans la cellule K3, on écrit: « =somme(B3:I3) et on appuie sur entrée.

En K3 est alors écrit 800. $N=800$

Il y a donc 800 livres pour les enfants qui ont été vendus dans la semaine.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Prix en euros	10	13	16	17	20	23	25	27			
3	Nombre de Livres	133	115	107	215	160	10	31	29	Effectif total	800	

c) Pour obtenir les effectifs cumulés croissants dans la ligne 4. Dans la cellule B4, on écrit: « =B3 ». En C4, on veut calculer n_1+n_2 avec n_1 qui est dans la cellule B4 et n_2 qui est dans la cellule C3; Pour cela, dans la cellule C4, on écrit: « =B4+C3 » et on appuie sur entrée. On revient en C4 et on étire la formule jusque I4.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Prix en euros	10	13	16	17	20	23	25	27			
3	Nombre de Livres	133	115	107	215	160	10	31	29	Effectif total	800	
4	E.C.C.	133	248	355	570	730	740	771	800			

d) Dans la ligne 5, on veut écrire les fréquences c'est à dire dans la cellule B5, on veut calculer: $\frac{n_1}{N}$ avec n_1 qui est dans la cellule B3 et N qui est dans la cellule K3. Pour cela, on écrit dans la cellule B5: « =B3/\$K3 » et

appuie sur entrée. (\$ signifie que quand on va étirer la formule, on garde le nombre de la cellule K3 pour tous les calculs). On revient en B5 et on étire la formule jusque I5.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Prix en euros	10	13	16	17	20	23	25	27			
3	Nombre de Livres	133	115	107	215	160	10	31	29	Effectif total	800	
4	E.C.C.	133	248	355	570	730	740	771	800			
5	Fréquences	0.16625	0.14375	0.13375	0.26875	^{q3} 0.2	0.0125	0.03875	0.03625			

e) On veut maintenant calculer la somme des fréquences (qui est égale à 1). Pour cela, dans la cellule J5, on écrit: « somme des fréquences » et dans la cellule K5: « =somme(B5:I5) » et on appuie sur entrée.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Prix en euros	10	13	16	17	20	23	25	27			
3	Nombre de Livres	133	115	107	215	160	10	31	29	Effectif total	800	
4	E.C.C.	133	248	355	570	730	740	771	800			
5	Fréquences	0.16625	0.14375	0.13375	0.26875	^{q3} 0.2	0.0125	0.03875	0.03625	Somme des fréquences	1	

f) Dans la ligne 6, on veut calculer les fréquences cumulées croissantes. Dans la cellule B6, on écrit: « =B5 » puis on appuie sur entrée. Dans la cellule C6, on écrit: « =B6+C5 », on appuie sur entrée et on étire la formule jusque I6.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Prix en euros	10	13	16	17	20	23	25	27			
3	Nombre de Livres	133	115	107	215	160	10	31	29	Effectif total	800	
4	E.C.C.	133	248	355	570	730	740	771	800			
5	Fréquences	0.16625	0.14375	0.13375	0.26875	^{q3} 0.2	0.0125	0.03875	0.03625	somme des fréquences	1	
6	F.C.C.	0.16625	0.31	0.44375	0.7125	0.9125	0.925	0.96375	1			

g) Dans la ligne 7, on calcule les fréquences en pourcentage. Pour cela, en B7, on écrit: « =B5*100 » , on appuie sur entrée et on étire la formule jusque I7. En J7, on écrit: « somme des fréquences cumulées en % » et on K7, on tape: « =somme(B7:I7) » et on appuie sur entrée.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Prix en euros	10	13	16	17	20	23	25	27			
3	Nombre de Livres	133	115	107	215	160	10	31	29	Effectif total	800	
4	E.C.C.	133	248	355	570	730	740	771	800			
5	Fréquences	0.16625	0.14375	0.13375	0.26875	0.2	0.0125	0.03875	0.03625	Somme des fréquences	1	
6	F.C.C.	0.16625	0.31	0.44375	0.7125	0.9125	0.925	0.96375	1			
7	Fréquences %	16.625	14.375	13.375	26.875	20	1.25	3.875	3.625	Somme des fréquences%	100	

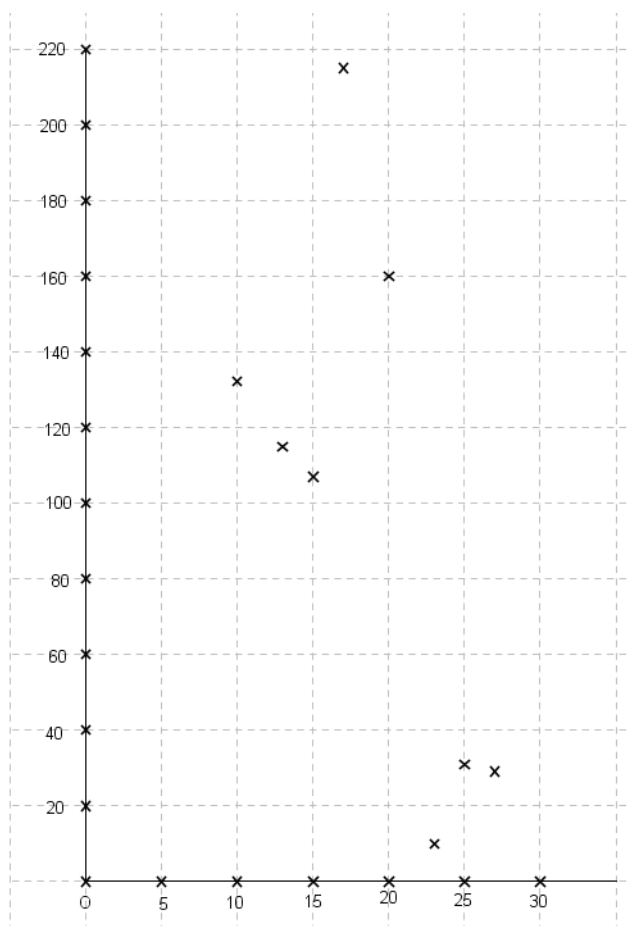
h) Dans la ligne 8, on écrit les fréquences cumulées en pourcentage croissantes. Pour cela, en B8, on écrit: « =B7 », on appuie sur entrée et en C8 on tape: « =B8+C7 », on appuie sur entrée et on étire la formule jusqu'en I8.

	A	B	C	D	E	F	G	H	I	J	K
1											
2	Prix en euros	10	13	16	17	20	23	25	27		
3	Nombre de Livres	133	115	107	215	160	10	31	29	Effectif total	800
4	E.C.C.	133	248	355	570	730	740	771	800		
5	Fréquences	0.16625	0.14375	0.13375	0.26875	0.2	0.0125	0.03875	0.03625	Somme des fréquences	1
6	F.C.C.	0.16625	0.31	0.44375	0.7125	0.9125	0.925	0.96375	1		
7	Fréquences %	16.625	14.375	13.375	26.875	20	1.25	3.875	3.625	Somme des fréquences%	100
8	F.C.C. en %	16.625	31	44.375	71.25	91.25	92.5	96.375	100		

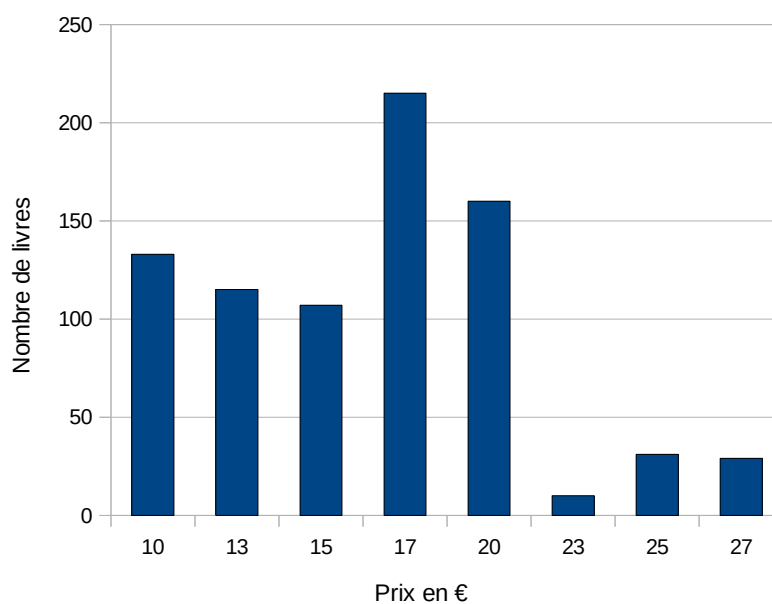
3.4. Représentations graphiques

a) Nuage de points

On considère les huit points ayant pour abscisses les valeurs du caractère et pour ordonnées leurs effectifs.



b) diagramme: par exemple le diagramme bâton (la hauteur est proportionnelle aux effectifs ou aux fréquences)



3.4. Détermination des paramètres de la série

a) Le mode de la série est **17**. (le plus grand effectif est 215 correspondant à la valeur 17)

Cela signifie que les livres les plus vendus sont ceux à 17€ durant la semaine considérée.

b) La médiane de la série est aussi **17**.

L'effectif total est 800 (nombre pair).

$$\frac{800}{2} = 400$$

La médiane est donc la demi-somme des nombre de la liste situés aux rangs 400 et 401. On regarde les effectifs cumulés croissants. Les nombres correspondants aux rangs 400 et 401 sont égaux à 17. Donc la médiane est 17.

Cela signifie que la moitié des livres vendues ont un prix inférieur ou égal à 17€.

c) Le premier quartile est **13**.

$$800 = 4 \times 200$$

Donc le premier quartile est le nombre au rang 200. On regarde les effectifs cumulés croissants. Le nombre correspondant au rang 200 est 13.

Cela signifie qu'au moins $\frac{1}{4}$ des livres vendus ont un prix inférieur ou égal à 13€.

d) Le troisième quartiles est **20**.

$$3 \times 200 = 600$$

Donc le troisième quartile est le nombre au rang 600. On regarde les effectifs cumulés croissants. Le nombre correspondant au rang 600 est 20.

Cela signifie qu'au moins $\frac{3}{4}$ des livres vendus ont un prix inférieur ou égal à 20€.

e) L'étendue est **17**.

$$27 - 17 = 10$$

La différence est le prix le plus élevé et le prix le moins élevé est 17.

f) L'écart inter-quartile est **7**.

$$20 - 13 = 7$$

La différence entre les deux quartiles est 7€.

g) Pour calculer la moyenne de la série, on peut utiliser la fonction sommeprod.

On peut continuer d'utiliser les lignes 10 et 11 des tableaux précédents. Mais pour diminuer la hauteur du tableau, on reprend les données dans les lignes 3,3 et 4.

Dans la cellule J5, on écrit: « somme des produits »

Dans la cellule K5, on écrit: « =sommeprod(B2:I2;B3:I3) et on appuie sur entrée

Dans la cellule J6, on écrit: « moyenne de la série »

Dans la cellule K6, on écrit: « =K10/K3 » et on tape sur entrée

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2	Prix en euros	10	13	16	17	20	23	25	27			
3	Nombre de Livres	133	115	107	215	160	10	31	29	Effectif total	800	
4										Somme des produits	13073	
5										Moyenne	16.34125	

La moyenne est 16,34.

Le prix moyen des livres est 16,34€.

4. Répartition en classes

4.1. Introduction

Lorsque les valeurs du caractère sont très nombreuses, on regroupe ces valeurs en intervalles que l'on nomme classes.

Ici, on suppose que toutes les classes ont la même amplitude.

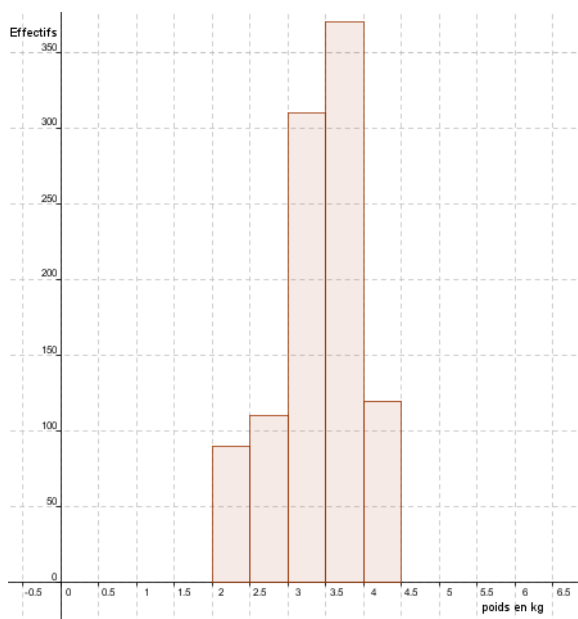
4.2. Exemple

Dans une maternité, on note le « poids » (masse en kg) de 1000 nouveaux nés. On regroupe les résultats sous forme de classe:

	A	B	C	D	E	F	G	H
1		1	2	3	4	5		
2	Classes	[2;2.5]	[2.5;3]	[3;3.5]	[3.5;4]	[4;4.5]		
3	Effectifs	90	110	310	370	120		

4.3. Histogramme

On représente cette série par un histogramme. On partage l'intervalle [2;4,5[en 5 parties de même amplitude et on construit 5 rectangles pour que les aires soient proportionnelles aux effectifs. (les bases des rectangles ayant la même amplitude, les hauteurs des rectangles sont proportionnelles aux effectifs).



4.4. Courbes des fréquences cumulées

On calcule les fréquences, puis les fréquences cumulées croissantes, puis les fréquences cumulées décroissantes.

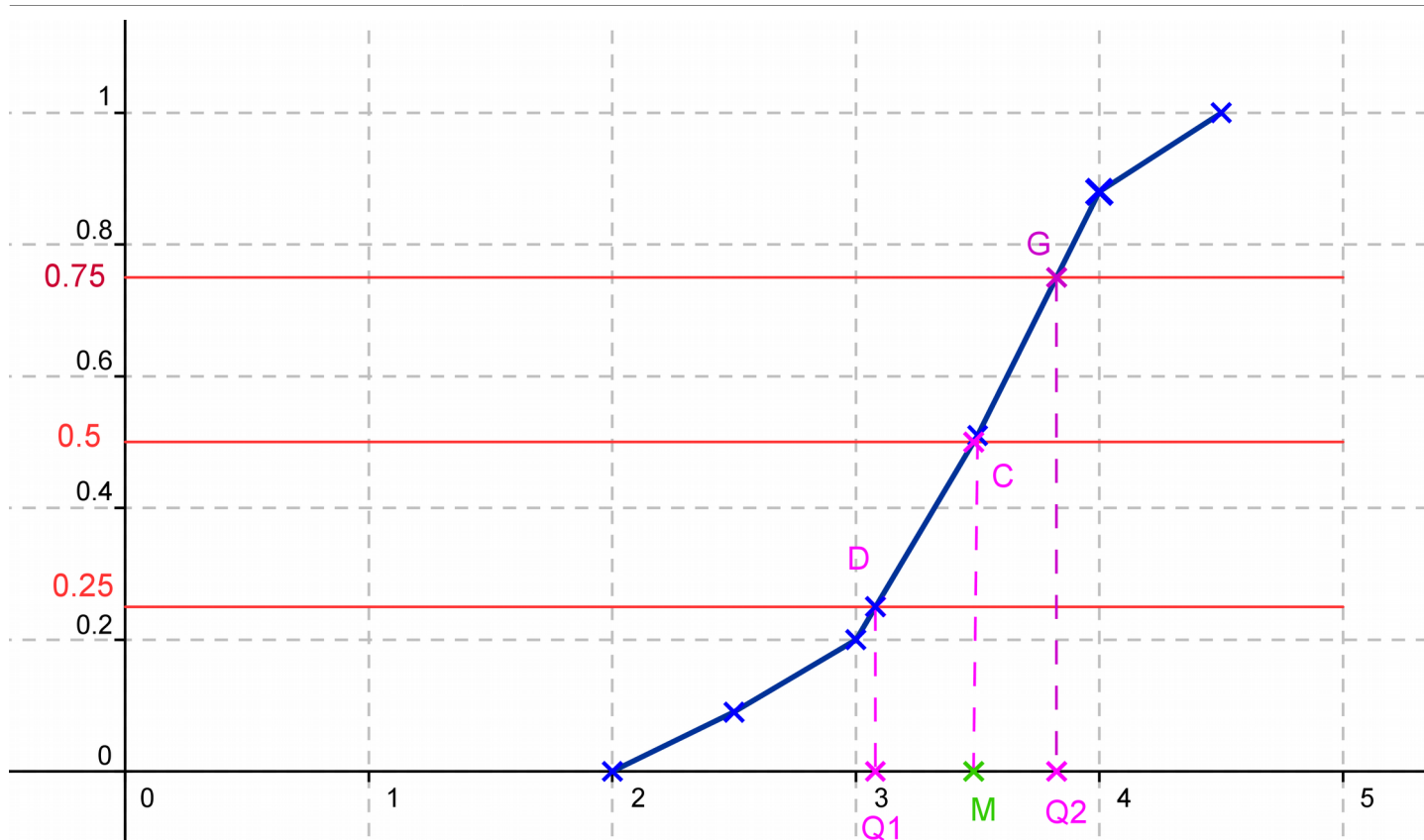
	A	B	C	D	E	F	G	H
1		1	2	3	4	5		
2	Classes	[2;2.5]	[2.5;3]	[3;3.5]	[3.5;4]	[4;4.5]		
3	Effectifs	90	110	310	370	120	Somme des effectifs	1000
4	Fréquences	0.09	0.11	0.31	0.37	0.12	Somme des fréquences	1
5	F.C.C.	0.09	0.2	0.51	0.88	1		
6	F.C.D.	1	0.91	0.8	0.49	0.12		

On suppose une répartition uniforme dans chaque classe. La courbe des fréquences cumulées croissantes est une courbe d'une fonction affine par morceaux.

On considère les points d'abscisses « les poids » et d'ordonnées « les fréquences cumulées »

(2;0) (2,5;0,09) (3;0,2) (3,5;0,51) (4;0,88) (4,5; 1)

On obtient:



La médiane est l'antécédent de 0,5. Le logiciel GéoGebra donne **3,48**. (abscisse du point M)

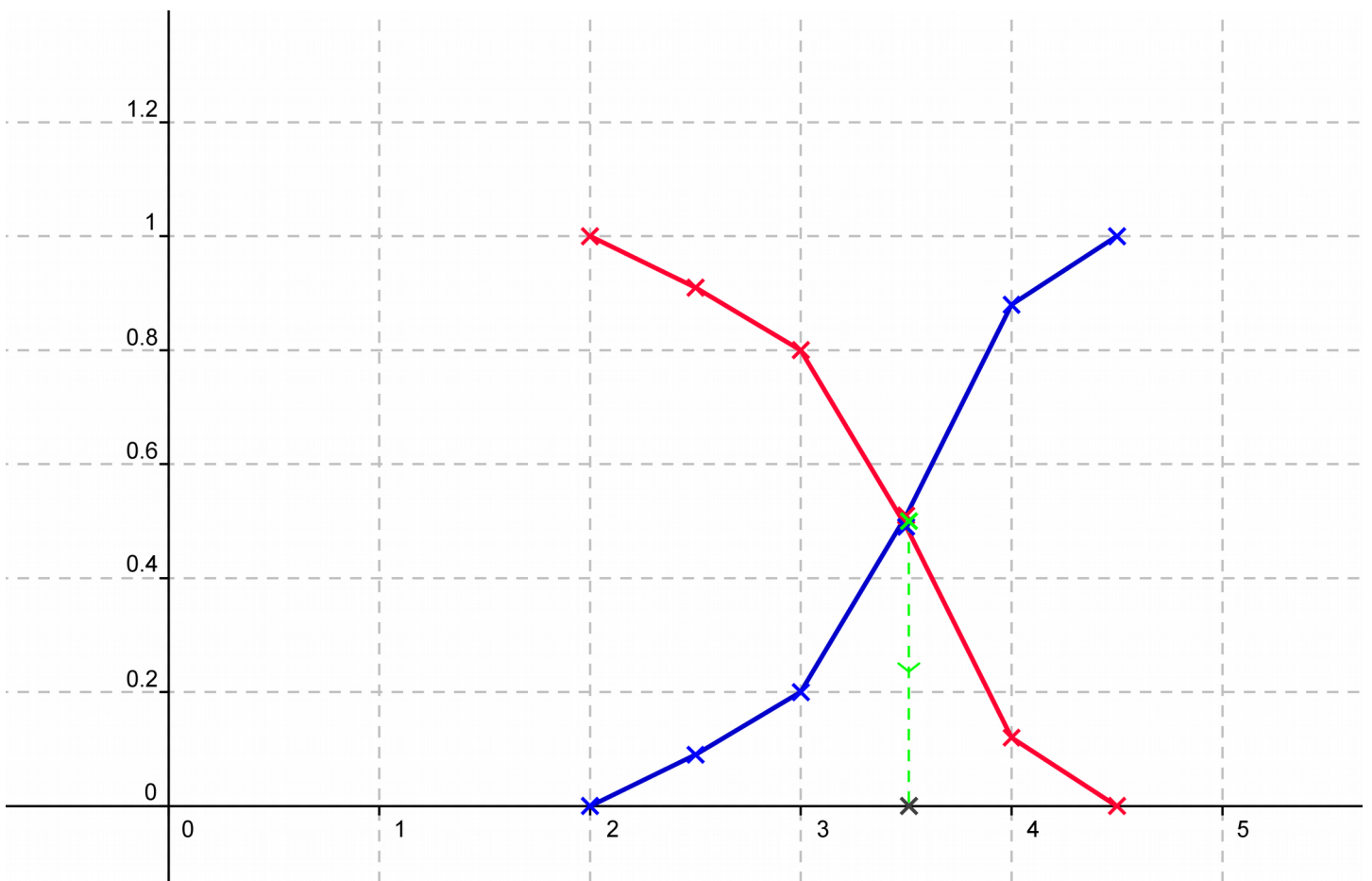
Le quartile Q_1 est l'antécédent de 0,25. Le logiciel GéoGebra donne **3,08**.

Le quartile Q_3 est l'antécédent de 0,75. Le logiciel GéoGebra donne **3,82**.

On définit de même la courbe des fréquences cumulées décroissantes.

On place les points:

(2;1) (2,5;0,91) (3;0,8) (3,5;0,49) (4;0,12) (4,5;0)



Les deux courbes sont symétriques par rapport à la droite d'équation: $y=0,5$

L'abscisse du point d'intersection des deux courbes est la médiane.

4.5. Calcul de la moyenne

Pour calculer la moyenne, on choisit le centre de la classe:

$$M = \frac{2,25 \times 90 + 2,75 \times 110 + 3,25 \times 310 + 3,75 \times 370 + 4,25 \times 120}{1000} \approx 3,41$$