Quartiles : programme officiel, calculatrice, tableur ou Wikipedia ?

J.Lejeune

I.R.E.M. de Basse-Normandie
lejeune@math.unicaen.fr

14 décembre 2007

Introduction :

Notion de quantile : passée dans le langage courant.

Introduction :

Notion de quantile : passée dans le langage courant.

- Exemple 1 : En 2003, le salaire médian, à temps complet était de 1849 euros c.-à.-d. la moitié des salariés à temps complet gagnait moins de 1849 euros et l'autre moitié plus de 1849 euros.
  $1849 = Me = q_{0.50} =$ quantile d'ordre $1/2 =$ médiane.

Introduction :

Notion de quantile : passée dans le langage courant.

- Exemple 1 : En 2003, le salaire médian, à temps complet était de 1849 euros c.-à.-d. la moitié des salariés à temps complet gagnait moins de 1849 euros et l'autre moitié plus de 1849 euros.
  $1849 = Me = q_{0.50} =$ quantile d'ordre $1/2$ = médiane.
- Exemple 2 : En 2003, les 10% des ménages les plus pauvres gagnaient moins de 609 euros par mois.
  $609 = D_1 = q_{0.10} =$ quantile d'ordre $1/10$ = premier décile.

Introduction :

Notion de quantile : passée dans le langage courant.

- Exemple 1 : En 2003, le salaire médian, à temps complet était de 1849 euros c.-à.-d. la moitié des salariés à temps complet gagnait moins de 1849 euros et l'autre moitié plus de 1849 euros.
  $1849 = Me = q_{0.50} =$ quantile d'ordre $1/2 =$ médiane.
- Exemple 2 : En 2003, les 10% des ménages les plus pauvres gagnaient moins de 609 euros par mois.
  $609 = D_1 = q_{0.10} =$ quantile d'ordre $1/10 =$ premier décile.
- Exemple 3 : Dans une étude sur 342 secteurs de gardes de la zone Vénissieux et Cours-la Ville, le quart de ces secteurs concernaient une population de moins de 34.41 h/km2 et le quart de ces secteurs une population de plus de 203.76 h/km2.
  $34.41 = Q_1 = q_{0.25} =$ quantile d'ordre $1/4 =$ premier quartile.
  $203.76 = Q_3 = q_{0.75} =$ quantile d'ordre $3/4 =$ troisième

Réflexions à partir de la brochure APMEP n°156

Commission Inter-IREM *Statistique et Probabilités*
Quartiles, déciles et tutti quantiles (*Jean-Claude Girard*)

Réflexions à partir de la brochure APMEP n°156

Commission Inter-IREM *Statistique et Probabilités*
Quartiles, déciles et tutti quantiles (*Jean-Claude Girard*)

Exemples de données de durée de vie d'ampoules de 100W

Jeu 1 : 6 ampoules : $46, 270, 293, 382, 630, 952$
Jeu 2 : 7 ampoules :$49, 90, 198, 302, 387, 547, 763$
Jeu 3 : 8 ampoules : $34, 47, 71, 263, 282, 622, 667, 968$
Jeu 4 : 9 ampoules : $39, 174, 196, 252, 331, 401, 456, 637, 944$

Calcul des quartiles

Trois méthodes :

1er quartile

	Programme	Calculatrice	Tableur
Jeu 1	270	270	275.75
Jeu 2	90	90	144
Jeu 3	47	59	65
Jeu 4	196	185	196

Calcul des quartiles

Trois méthodes :

1er quartile

	Programme	Calculatrice	Tableur
Jeu 1	270	270	275.75
Jeu 2	90	90	144
Jeu 3	47	59	65
Jeu 4	196	185	196

3ème quartile

	Programme	Calculatrice	Tableur
Jeu 1	630	630	568
Jeu 2	547	547	467
Jeu 3	622	644.5	633.25
Jeu 4	456	546.5	456

Définition mathématique pour une variable "continue" :

Soit X variable aléatoire réelle qui suit une loi de densité f non nulle sur I.
Soit F la fonction de répartition définie par :

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

Soit $p \in]0, 1|$.
Le quantile d'ordre p est le réel q_p appartenant à I tel que $F(q_p) = p$.
On écrit aussi $q_p = F^{-1}(p)$.

Exemple :

Une ampoule est vendue avec l'indication : durée moyenne = 1 an =365 jours.

X est la variable aléatoire qui mesure (en jours) la durée de vie de l'ampoule dont la loi est modélisée par la densité exponentielle $f(x) = \lambda e^{-\lambda x}$ si $x \geq 0$ et $f(x) = 0$ si $x < 0$.

Puisque $E(X) = \frac{1}{\lambda}$, $\lambda = \frac{1}{365}$ et

$F(x) = 1 - e^{-\lambda x} = 1 - e^{-\frac{x}{365}}$ si $x \geq 0$ et $F(x) = 0$ si $x < 0$.

q_p vérifie $1 - e^{-\lambda q_p} = p$ donc $q_p = -\dfrac{\ln(1-p)}{\lambda} = -365\ln(1-p)$

Exemple :

Une ampoule est vendue avec l'indication : durée moyenne = 1 an =365 jours.

X est la variable aléatoire qui mesure (en jours) la durée de vie de l'ampoule dont la loi est modélisée par la densité exponentielle $f(x) = \lambda e^{-\lambda x}$ si $x \geq 0$ et $f(x) = 0$ si $x < 0$.

Puisque $E(X) = \frac{1}{\lambda}$, $\lambda = \frac{1}{365}$ et

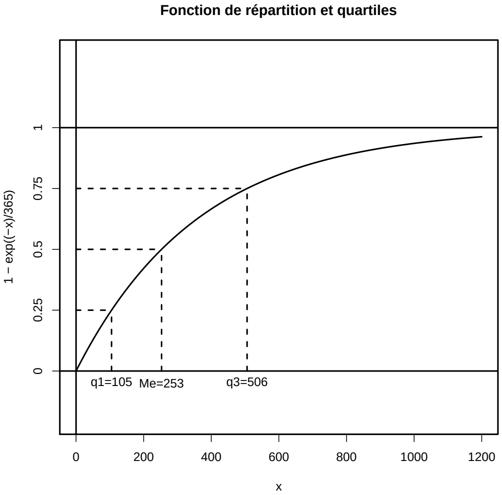
$F(x) = 1 - e^{-\lambda x} = 1 - e^{-\frac{x}{365}}$ si $x \geq 0$ et $F(x) = 0$ si $x < 0$.

q_p vérifie $1 - e^{-\lambda q_p} = p$ donc $q_p = -\dfrac{\ln(1-p)}{\lambda} = -365\ln(1-p)$

$Me = q_{0.50} = -365\ln(1 - 0.50) = 253$

$Q_1 = q_{0.25} = -365\ln(1 - 0.25) = 105$

$Q_3 = q_{0.75} = -365\ln(1 - 0.75) = 506$


Fonction de répartition et quartiles
1 − exp((−x)/365)
0
0.25
0.5
0.75
1
q1=105
Me=253
q3=506
0
200
400
600
800
1000
1200
x

Définition mathématique pour une variable discrète à partir d'un exemple

X =numéro tiré après tirage d'un dé équilibré

$P(X = i) = \frac{1}{6}$ pour $i = 1, 2, 3, 4, 5, 6$.

$F(x) = P(X \leq x) = 0$ si $x < 1$

$F(x) = P(X \leq x) = \frac{1}{6}$ si $1 \leq x < 2$

$F(x) = P(X \leq x) = \frac{2}{6}$ si $2 \leq x < 3$

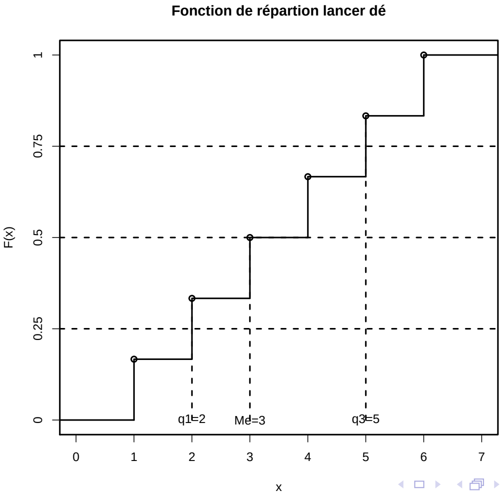
....

$F(x) = P(X \leq x) = \frac{5}{6}$ si $5 \leq x < 6$

$F(x) = P(X \leq x) = 1$ si $x \geq 6$

Définition :

$$q_p = \min\{x \ \ / F(x) \geq p\}$$

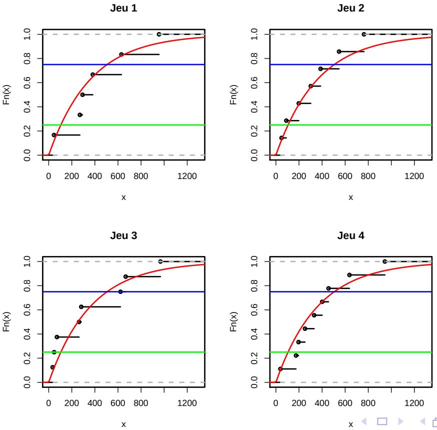


Remarque :
Pour une variable continue, on peut appliquer la même définition car
$\forall x \geq q_p \quad F(x) \geq F(q_p) = p$
et de plus F est continue au point q_p d'où la définition générale :

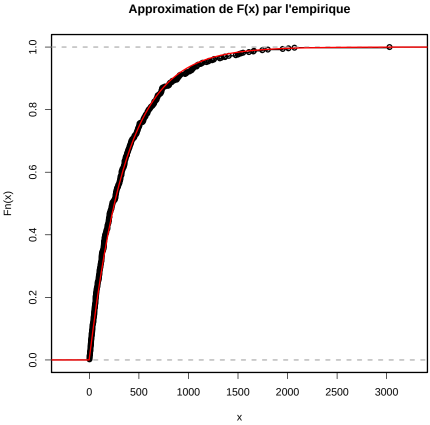
$$q_p = \min\{x \ \ /F(x) \geq p\}$$

Fonction de répartition empirique

$x_1, x_2, ..., x_n$ suite de données classées par ordre non-décroissant.
$\hat{F}(x) = \frac{1}{n} Card\{i \ \ / \ x_i \leq x\}$ = fréquence des données inférieures ou égales à x.



Fonction de répartition empirique = bonne approximation de $F(x)$
Simulation de 500 durées de vie pour des ampoules de durée moyenne 1 an



Nouveaux programmes de première :

Le premier (respectivement le troisième) quartile est le plus petit élément q_1 (respectivement q_3) des valeurs de la série ordonnée par ordre croissant, tel qu'au moins 25%*(respectivement* 75%*) de ces valeurs soient inférieures ou égales à q_1 (respectivement q_3).*

Nouveaux programmes de première :

Le premier (respectivement le troisième) quartile est le plus petit élément q_1 (respectivement q_3) des valeurs de la série ordonnée par ordre croissant, tel qu'au moins 25%(respectivement 75%) de ces valeurs soient inférieures ou égales à q_1 (respectivement q_3).

Jeu 1 : $46, 270, 293, 382, 630, 952$

$q_1 = 46$? Il y a 1 valeur inférieure ou égale à 46 soit $\frac{1}{6} = 16,66\%$

$q_1 = 270$? Il y a 2 valeurs inférieures ou égale à 270 soit $\frac{2}{6} = 33,33\%$

Donc $q_1 = 270$

Nouveaux programmes de première :

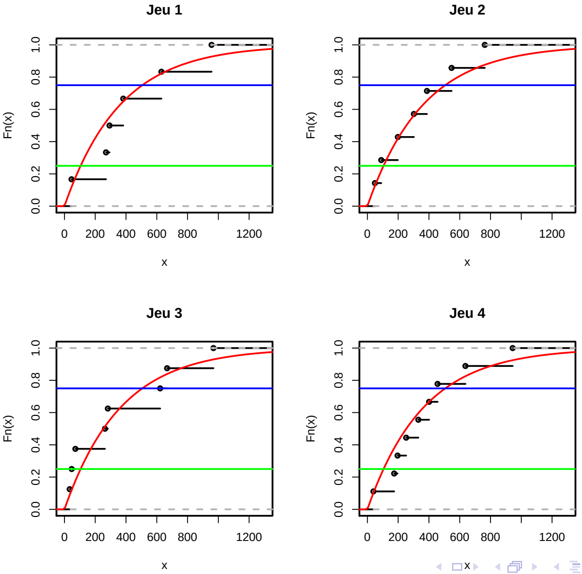
Le premier (respectivement le troisième) quartile est le plus petit élément q_1 (respectivement q_3) des valeurs de la série ordonnée par ordre croissant, tel qu'au moins 25%(respectivement 75%) de ces valeurs soient inférieures ou égales à q_1 (respectivement q_3).

Jeu 1 : $46, 270, 293, 382, 630, 952$

$q_1 = 46$? Il y a 1 valeur inférieure ou égale à 46 soit $\frac{1}{6} = 16,66\%$
$q_1 = 270$? Il y a 2 valeurs inférieures ou égale à 270 soit $\frac{2}{6} = 33,33\%$
Donc $q_1 = 270$

Jeu 3 : $34, 47, 71, 263, 282, 622, 667, 968$

$q_1 = 34$? Il y a 1 valeur inférieure ou égale à 34 soit $\frac{1}{8} = 12,5\%$
$q_1 = 47$? Il y a 2 valeurs inférieures ou égale à 47 soit $\frac{2}{8} = 25\%$
Donc $q_1 = 47$


Jeu 1
Jeu 2
Jeu 3
Jeu 4
Fn(x)
x
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1200

Un constat :

$\hat{q}_{0.25} =$ 1er quartile calculé suivant programme. Alors :

$$\hat{q}_{0.25} = \min\{x \ \ /\hat{F}(x) \geq 0.25\}$$

Un constat :

$\hat{q}_{0.25} =$ 1er quartile calculé suivant programme. Alors :

$$\hat{q}_{0.25} = \min\{x \ \ /\hat{F}(x) \geq 0.25\}$$

$\hat{q}_{0.75} =$ 3ème quartile calculé suivant programme. Alors :

$$\hat{q}_{0.75} = \min\{x \ \ /\hat{F}(x) \geq 0.75\}$$

Calcul par la calculatrice

Rappel : *"On ordonne" la série des observations par ordre croissant ; si la série est de taille $2n+1$, la médiane est la valeur du rang $n+1$ dans cette série ordonnée ; si la série est de taille $2n$ la médiane est la moyenne des valeurs des termes de rang n et $n+1$ dans cette série ordonnée "* (Document d'accompagnement des programmes de premières).
La médiane détermine deux sous-séries constituées des observations antérieures.
. sous-série 1 = observations $\leq$ médiane
. sous-série 2 = observations $\geq$ médiane
Si taille= $2n+1$, on exclut dans chaque sous-série une et une seule donnée : la donnée centrale de la série initiale (c.a.d. la médiane), sinon on n'exclut rien.
1er quartile = médiane de la nouvelle sous-série 1.
3ème quartile = médiane de la nouvelle sous-série 2.

Exemples :

Jeu 1 : $46, 270, 293, 382, 630, 952$

Taille $= 2n = 2 \times 3$. Médiane $= = \dfrac{x_3 + x_4}{2} = \dfrac{293 + 382}{2} = 337.5$

Sous-série 1 : $46, 270, 293$. 1er quartile $= 270$

Sous-série 2 : $82, 630, 952$. 3ème quartile $= 630$

Exemples :

Jeu 1 : $46, 270, 293, 382, 630, 952$

Taille $= 2n = 2 \times 3$. Médiane $= = \dfrac{x_3 + x_4}{2} = \dfrac{293 + 382}{2} = 337.5$

Sous-série 1 : $46, 270, 293$. 1er quartile $= 270$

Sous-série 2 : $82, 630, 952$. 3ème quartile $= 630$

Jeu 4 : : $39, 174, 196, 252, 331, 401, 456, 637, 944$

Taille $= 2n + 1 = 2 \times 4 + 1$. Médiane $= = x_5 = 331$

. sous-série 1 $= 39, 174, 196, 252, 331$

. sous-série 2 $= 331, 401, 456, 637, 944$

Taille impaire :

. nouvelle sous-série 1$= 39, 174, 196, 252$. 1er quartile $= \dfrac{174 + 196}{2} = 185$

. nouvelle sous-série 2 $= 401, 456, 637, 944$. 3ème quartile $= \dfrac{456 + 637}{2} = 546.5$

Calcul avec tableur

Calcul du quantile d'ordre p.

Ordonnée	x_1	x_2	$\cdots$	x_i	x_{i+1}	$\cdots$	x_n
Abscisse	0	$\frac{1}{n-1}$	$\cdots$	$\frac{i-1}{n-1}$	$\frac{i}{n-1}$	$\cdots$	1
Indice	1	2	$\cdots$	i	$i+1$	$\cdots$	n

Calcul avec tableur

Calcul du quantile d'ordre p.

Ordonnée	x_1	x_2	$\cdots$	x_i	x_{i+1}	$\cdots$	x_n
Abscisse	0	$\frac{1}{n-1}$	$\cdots$	$\frac{i-1}{n-1}$	$\frac{i}{n-1}$	$\cdots$	1
Indice	1	2	$\cdots$	i	$i+1$	$\cdots$	n

Pour $i = 1, 2, ..., n-1$, on relie par un segment de droite les points $\left(\frac{i-1}{n-1}, x_i\right)$ et $\left(\frac{i}{n-1}, x_{i+1}\right)$.

Calcul avec tableur

Calcul du quantile d'ordre p.

Ordonnée	x_1	x_2	$\cdots$	x_i	x_{i+1}	$\cdots$	x_n
Abscisse	0	$\frac{1}{n-1}$	$\cdots$	$\frac{i-1}{n-1}$	$\frac{i}{n-1}$	$\cdots$	1
Indice	1	2	$\cdots$	i	$i+1$	$\cdots$	n

Pour $i = 1, 2, ..., n-1$, on relie par un segment de droite les points $(\frac{i-1}{n-1}, x_i)$ et $(\frac{i}{n-1}, x_{i+1})$.

Ce polygone des fréquences cumulées inverse est donc le graphe de la fonction $H : [0,1] \longrightarrow \mathbb{R}$:

$H(p) = (1-\lambda_p)x_i + \lambda_p x_{i+1}$ si $\frac{i-1}{n-1} \leq p \leq \frac{i}{n-1}$ avec i défini ainsi :

$i = [r]$ où $r = (n-1)p + 1$

$\lambda_p = (n-1)p + 1 - i = r - [r]$

Le quantile d'ordre p est donné par :

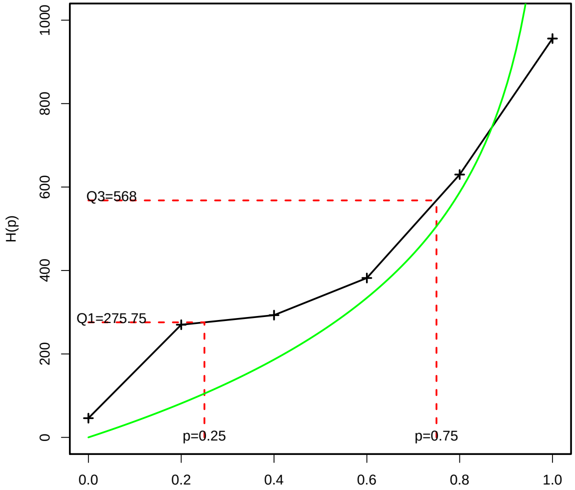
$\widetilde{q}_p = H(p)$

Exemple :

Jeu 1 : $46, 270, 293, 382, 630, 952$ $\quad n = 6$

Ordonnée	46	270	293	382	630	952
Abscisse	0	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$
Indice	1	2	3	4	5	6

On ajoute le graphe de F^{-1} en vert.



Recherche des quartiles

1er quartile : $Q_1 = q_{0.25} =?$

- on calcule $r = (n-1)p + 1 = (6-1) \times 0.25 + 1 = 2.25$
- alors $i = [r] = [2.25] = 2$ et $\lambda_{0.25} = r - [r] = 2.25 - 2 = 0.25$
- on cherche $x_i = x_2 = 270$ et $x_{i+1} = x_3 = 293$
- on trouve $q_{0.25} = H(0.25) = (1-\lambda_p)x_i + \lambda_p x_{i+1} = (1-0.25) \times 270 + 0.25 \times 293 =$ donc $Q_1 = 275.75$.

Recherche des quartiles

1er quartile : $Q_1 = q_{0.25} = ?$

- on calcule $r = (n-1)p + 1 = (6-1) \times 0.25 + 1 = 2.25$
- alors $i = [r] = [2.25] = 2$ et $\lambda_{0.25} = r - [r] = 2.25 - 2 = 0.25$
- on cherche $x_i = x_2 = 270$ et $x_{i+1} = x_3 = 293$
- on trouve $q_{0.25} = H(0.25) = (1-\lambda_p)x_i + \lambda_p x_{i+1} =$ $(1-0.25) \times 270 + 0.25 \times 293 =$ donc $Q_1 = 275.75$.

3ème quartile : $Q_3 = q_{0.75} = ?$

- on calcule $r = (n-1)p + 1 = (6-1) \times 0.75 + 1 = 4.75$
- alors $i = [r] = [4.75] = 4$ et $\lambda_{0.75} = r - [r] = 4.75 - 4 = 0.75$
- on cherche $x_i = x_4 = 382$ et $x_{i+1} = x_5 = 630$
- on trouve $q_{0.75} = H(0.75) = (1-\lambda_p)x_i + \lambda_p x_{i+1} =$ $(1-0.75) \times 382 + 0.75 \times 630 = 568$ donc $Q_3 = 568$.

... et Wikipedia ?

Début du document Wikipedia

En statistique descriptive, un quartile est chacune des 3 valeurs qui divisent les données triées en 4 parts égales, de sorte que chaque partie représente 1/4 de l'échantillon de population.

Calcul des quartiles

Voir à quantile pour les méthodes. Le quartile est calculé en tant que 4-quantile. Donc :

- le 1er quartile sépare les 25% inférieurs des données ;
- le 2e quartile est la médiane de la série ;
- le 3e quartile sépare les 75% inférieurs des données.

Méthode :

- Dans le cas continu on utilise la fonction représentative du polygone des fréquences cumulées. (voir à Statistiques élémentaires continues)
- Dans le cas discret on range les données par ordre croissant ensuite : Le quartile inférieur est la valeur du milieu du premier ensemble, dans lequel 25% des valeurs sont inférieures à Q1 et 75% lui sont supérieures. Le premier quartile prend la notation Q1. Le quartile supérieur est la valeur du milieu du deuxième ensemble, dans lequel 75% des valeurs sont inférieures à Q3 et 25% lui sont supérieurs. Le troisième quartile prend donc la notation Q3

Exemple :

Les valeurs dans l'ordre ascendant 1, 11, 15, 19, 20, 24, 28, 34, 37, 47, 50, 57.

Q1 est entre 15 et 19 donc : Q1 = 17

Q2 est entre 24 et 28 donc : Q2 = 26 (c'est la médiane)

Q3 est entre 37 et 47 donc : Q3 = 42

Fin du document Wikipedia

Quelques critiques...entre autres !

Quelques critiques...entre autres !

- Dans le cas discret, l'"ensemble" dont il est question est sans doute le sous ensemble $E =]15, 19[$ car 25% des valeurs de la série sont inférieures à n'importe $x \in E$ et 75% sont supérieures à $x \in E$. Pour le Jeu 3, on retrouve la définition des calculatrices. De toute façon, définition pas claire !

Quelques critiques...entre autres !

- Dans le cas discret, l'"ensemble" dont il est question est sans doute le sous ensemble $E =]15, 19[$ car 25% des valeurs de la série sont inférieures à n'importe $x \in E$ et 75% sont supérieures à $x \in E$. Pour le Jeu 3, on retrouve la définition des calculatrices. De toute façon, définition pas claire !
- Qu'en est-il pour le Jeu 2 : $49, 90, 198, 302, 387, 547, 763$?
  $E =]90, 198[$ ou bien $E =]49, 90[$?
  Dans le premier cas, plus de 25% des valeurs inférieures à $x \in E$ et moins de 75% sont supérieures à $x \in E$: Q1= 144 ;
  Dans le deuxième cas, moins de 25% des valeurs inférieures à $x \in E$ et plus de 75% sont supérieures à $x \in E$: Q1= 69.5.
  Définition inapplicable dans l'état si $n \neq 4k$.

Quelques critiques...entre autres !

- Dans le cas discret, l'"ensemble" dont il est question est sans doute le sous ensemble $E =]15, 19[$ car 25% des valeurs de la série sont inférieures à n'importe $x \in E$ et 75% sont supérieures à $x \in E$. Pour le Jeu 3, on retrouve la définition des calculatrices. De toute façon, définition pas claire !
- Qu'en est-il pour le Jeu 2 : $49, 90, 198, 302, 387, 547, 763$?
  $E =]90, 198[$ ou bien $E =]49, 90[$?
  Dans le premier cas, plus de 25% des valeurs inférieures à $x \in E$ et moins de 75% sont supérieures à $x \in E$: Q1= 144 ;
  Dans le deuxième cas, moins de 25% des valeurs inférieures à $x \in E$ et plus de 75% sont supérieures à $x \in E$: Q1= 69.5.
  Définition inapplicable dans l'état si $n \neq 4k$.

Pour poursuivre la polémique, voir le site MathemaTex :
http ://forum.mathematex.net/mathematiques-f5/mediane-quartiles-t2127.html