Note 10 de la page 23

La définition de la médiane et des quartiles donnée en page 23 de l'ouvrage est conforme à la définition recommandée par l'INSEE :

> Si on ordonne une distribution de salaires, de revenus, de chiffre d'affaires..., les quartiles sont les valeurs qui partagent cette distribution en quatre parties égales.
> Ainsi, pour une distribution de salaires :
> - le premier quartile (noté généralement Q1) est le salaire au-dessous duquel se situent 25 % des salaires ;
> - le deuxième quartile est le salaire au-dessous duquel se situent 50 % des salaires ; c'est la médiane ;
> - le troisième quartile (noté généralement Q3) est le salaire au-dessous duquel se situent 75 % des salaires.
>
> Le premier quartile est, de manière équivalente, le salaire au-dessus duquel se situent 75 % des salaires ; le deuxième quartile est le salaire au-dessus duquel se situent 50 % des salaires, et le troisième quartile le salaire au-dessus duquel se situent 25 % des salaires.

Toutefois cette définition ne définit pas la valeur des quartiles et de la médiane de manière unique !

Si l'on reprend l'exemple des 8 consommations de la série 1 :

4,2	4,9	5,1	5,9	6,4	6,4	7,6	8,3
	Q_1		médiane		Q_3		

Toute valeur comprise entre 4,9 et 5,1 peut être considérée comme le 1er quartile : les 2 consommations les plus petites (25 %) seront inférieures à cette valeur.
Toute valeur comprise entre 5,9 et 6,4 peut être considérée comme le 2ème quartile : les 4 consommations les plus petites (50 %) seront inférieures à cette valeur.
Toute valeur comprise entre 6,4 et 7,6 peut être considérée comme le 3ème quartile : les 6 consommations les plus petites (75 %) seront inférieures à cette valeur.

Pour le cas de la médiane, la convention retenue par tous les utilitaires de calcul statistique est la suivante :

- Si le nombre n des données, rangées par ordre croissant est impair, avec n = 2k + 1, la médiane est la donnée numéro : k + 1 ; en effet les k premières lui sont forcément inférieures ou égales, et les k dernières supérieures ou égales.
- Si le nombre n des données, rangées par ordre croissant est impair, avec n = 2k, la médiane est la moyenne des données numéro k et k+1.

C'est ce qui a été expliqué en page 22, avec 7 données, puis avec 8 données.

Pour le cas des quartiles 1 et 3, il n'existe pas à ce jour de convention internationale qui en fixe le mode de calcul de manière unique. On obtient effectivement des résultats différents avec les calculatrices ou le tableur excel, qui sont présentés dans le tableau ci-dessous :

	4,2	4,9	5,1	5,9	6,4	6,4	7,6	8,3
		Q_1				Q_3		
TI		5				7		
Casio		4,9				6,4		
Tableur (.INCLURE)		5,05				6,7		
Tableur (.EXCLURE)		4,95				7,3		

Les versions d'excel à partir de 2007 proposent d'ailleurs deux fonctions de calcul des quartiles :

- QUARTILE.INCLURE équivalent à QUARTILE (versions antérieures)
- QUARTILE.EXCLURE qui utilise une autre formule de calcul des quartiles.

Il faut aussi remarquer que la version la plus récente (début année 2016) de la calculatrice Casio GRAPH35+E propose deux modalités de calcul différentes, auxquelles on accède par le **SET UP**, après avoir activé le **Menu Stats**, sur la ligne **Q1Q3 Type** :

- OnD (OnData) activé par défaut donne les résultats du tableau ci-dessus
- Std donne les mêmes résultats que les TI

Bien évidemment, d'autres évolutions risquent d'avoir lieu sur les prochaines versions de calculatrices, de toutes marques, ou sur les tableurs !

Méthode de calcul des quartiles par les calculatrices TI

Dans le cas n = 2k (pair) ou n = 2k + 1 (impair)

- Q1 = médiane des k données les plus petites
- Q3 = médiane des k données les plus grandes

On peut observer que les calculatrices Casio donnent le même résultat lorsque n est impair (par exemple pour la série des 7 essais de consommation).

Méthode de calcul des quartiles par les calculatrices Casio

Cette méthode est conforme à la définition recommandée par l'éducation nationale pour les classes de lycée depuis les années 2000 :

> ***Premier quartile :*** *c'est le plus petit élément q des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 25% des données soient inférieures ou égales à q.*
> ***Troisième quartile :*** *c'est le plus petit élément q' des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 75% des données soient inférieures ou égales à q'.*

La précision importante apportée par cette définition est que **les 1^{er} et $3^{ème}$ quartiles sont des valeurs de la liste statistique**. On observe facilement que les résultats donnés par une calculatrice Casio sont conformes à cette définition :

4,9 est la plus petite valeur de la liste pour laquelle 2 données (25 %) sont inférieures ou égales à cette valeur.
6,4 est la plus petite valeur de la liste pour laquelle 6 données (75 %) sont inférieures ou égales à cette valeur.

La définition du deuxième quartile, dans les mêmes recommandations, en conformité avec celles des 1^{er} et $3^{ème}$ quartiles est la suivante :
c'est le plus petit élément q des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 50% des données soient inférieures ou égales à q

Mais avec cette définition, le $2^{ème}$ quartile n'est pas la médiane, contrairement aux recommandations de l'INSEE.
Par exemple le $2^{ème}$ quartile pour la série 1 des 8 consommations est 5,9.

Méthode de calcul des quartiles par les tableurs excel

Le calcul des quartiles sur tableur excel distingue 4 cas, selon le reste de la division de l'effectif n par 4.
Les valeurs obtenues sont dans la liste seulement lorsque n = 4k+1, et dans ce cas (seulement) les quartiles obtenus sont les mêmes que ceux donnés par une calculatrice Casio.
Dans les autres cas, les quartiles répondent à la définition suivante, avec la fonction QUARTILE ou QUARTILE.INCLURE.

Premier quartile : *c'est une valeur Q1, telle qu'au moins 25% des données soient inférieures ou égales à Q1.*
Troisième quartile : *c'est une valeur Q3, telle qu'au moins 25% des données soient supérieures ou égales à Q3.*

La valeur obtenue pour Q3 avec excel ne répond donc pas à la définition des classes de lycée lorsque n n'est pas de la forme 4k + 1.

La méthode de calcul est décrite ci-dessous ; elle est très technique et peut être sautée à la lecture ; dans ce cas il est conseillé d'aller directement à la conclusion de cette étude.

Chaque quartile est une moyenne pondérée de 2 données consécutives. Pour suivre au mieux la description de la méthode, il est conseillé d'ouvrir le fichier excel d'exemples de quartiles.

1^er^ cas : effectif total n = 4k (multiple de 4)

Les données rangées par ordre croissant sont réparties en 4 paquets de k données, et l'on note X_i la donnée de numéro i.
Le premier paquet comporte les données x_1 à x_k, le 2ème de X_{k+1} à X_{2k}, etc.

$Q1 = X_k + 3/4\,(X_{k+1} - X_k) \qquad = 0{,}25\,X_k + 0{,}75\,X_{k+1}$
$Q3 = X_{3k} + 1/4\,(X_{3k+1} - X_{3k}) \qquad = 0{,}75\,X_{3k} + 0{,}25\,X_{3k+1}$

2ème cas : effectif total n = 4k + 1

On a deux paquets de k données à gauche de X_{2k+1} (qui est la médiane), et deux paquets à droite de cette valeur.

$Q1 = X_{k+1}$ (plus petite valeur du 2ème paquet)
$Q3 = X_{3k+1}$ (plus grande valeur du 3ème paquet)

3ème cas : effectif total n = 4k + 2

Les 4 paquets de k données commencent à la 2ème donnée ; cela revient à isoler la première et la dernière.

$Q1 = X_{k+1} + 1/4\,(X_{k+2} - X_{k+1}) \qquad = 0{,}75\,X_{k+1} + 0{,}25\,X_{k+2}$
$Q3 = X_{3k+1} + 3/4\,(X_{3k+2} - X_{3k+1}) \qquad = 0{,}25\,X_{3k+1} + 0{,}75\,X_{3k+2}$

4ème cas : effectif total n = 4k + 3

Le premier paquet commence à la 2ème donnée, suivie du 3ème, puis de la médiane, puis des 2 derniers paquets avant la dernière valeur

$Q1 = (X_{k+1} + X_{k+2}) / 2 \qquad = 0{,}5\,X_{k+1} + 0{,}5\,X_{k+2}$
$Q3 = (X_{3k+2} + X_{3k+3}) / 2 \qquad = 0{,}5\,X_{3k+2} + 0{,}5\,X_{3k+3}$

Conclusions

Conclusion 1 : **l'objectif est une communication d'information claire.**

Les écarts entre les résultats affichés par différents outils pour les quartiles sont généralement minimes.

Ils n'ont aucune importance si l'on revient à l'objectif du calcul des paramètres statistiques : communiquer une information de synthèse.

Pour l'exemple des 1600 montants d'achat traité au chapitre 3, page 30, il est indifférent de dire que le premier quartile est 192,88 ou qu'il est de 192,99. Vouloir donner une valeur au centime près n'apporte pas d'information intéressante, et même vient brouiller le message.

Il vaut mieux retenir simplement que 25 % des montants d'achats sont inférieurs à 193 euros.

La définition large proposée par l'INSEE pour la médiane et les quartiles répond pleinement à cet objectif d'une communication la plus simple possible.

Conclusion 2 : garder le contrôle des outils

Les outils de calculs, tableur ou outil spécifique de gestion, ne sont que des outils.

Ils sont programmés ou paramétrés à partir de choix faits par les concepteurs.

Il appartient toujours à l'utilisateur de s'assurer et vérifier que les résultats obtenus sont conformes à l'objectif de calcul visé.