Deux ou trois choses que je sais des quartiles et des boites à moustaches(*)

Jacques Verdier(**)

Cet article fait suite à celui que j'ai publié dans le bulletin n° 430 de l'A.P.M.E.P. sous le titre « Deux ou trois choses que je sais de la médiane ». Ce dernier est disponible sur le Web, à l'adresse URL suivante :

http://www.ac- nancy-metz.fr/enseign/maths/APMEP/mediane.htm

Je vous invite à vous y reporter avant de lire celui-ci.

Les quartiles

Tout ce qui a été dit sur les problèmes de « définition rigoureuse » de la médiane dans l'article cité reste vrai pour les quartiles, et les déciles. Ce que l'élève doit retenir, c'est que 25% de la population se situe en dessous du premier quartile (Q1), 25% au-dessus du troisième quartile (Q3), et 50% entre les deux. De même pour les déciles : on utilise surtout Dl et D9, et les 80% « centraux » se trouvent entre Dl et D9 ; D9 – Dl s'appelle d'ailleurs l'écart interdécile.

Cependant, on peut vouloir chercher à définir de façon un peu plus rigoureuse ces quartiles (il en serait de même pour les déciles, problème que je n'aborderai pas ici : je laisserai le lecteur transposer).
Je rappelle que l'on travaille toujours sur les séries statistiques « sources » (c'est à dire qu'on ne les regroupe pas en classes : une statistique est une liste finie de valeurs)(1).
Dans toute la suite de mon exposé, j'utiliserai quatre statistiques : la série A compte 40 valeurs, la série B en compte 41, la série C 42 et la série D 43. Ces séries sont ordonnées (il n'est pas impossible que des valeurs consécutives soient égales). Pour simplifier les choses, je prendrai comme valeurs de la série A les entiers naturels de 21 à 60, pour la série B les entiers de 21 à 61, etc.
Je fais ce choix de travailler sur des exemples pour éviter de traduire les définitions choisies en « formules » mathématiques, la traduction de la langue courante vers la formule (dans la mesure où on l'estimerait nécessaire) étant plus aisée que l'inverse.

(*) Cet article est déjà paru dans le « PETIT VERT » (Bulletin de la Régionale Lorraine) n° 65 de mars 2001, sous le titre « *Honneur aux DAM ! (ou les avatars de la femme à barbe)* ». Pour ceux qui ne le savent pas, une DAM est une DEB qui a vieilli d'un an… (dans les instructions officielles, datant d'un an, on parle de Diagrammes en Boites ; depuis on a vu apparaître les Diagrammes à Moustaches).

(**) Lycée Varoquaux, 54-Tomblaine

(1) Le regroupement en classes est lui aussi un résumé de la série : ce n'est jamais lui qu'on utilise pour calculer moyenne, médiane, etc.

Définition « naturelle » des quartiles.

Une fois que l'on a admis la définition « naturelle » de la médiane (c'est la valeur centrale de la série classée si le nombre de termes est impair, sinon c'est la moyenne entre les deux valeurs « centrales »), on peut reprendre cette démarche pour définir les quartiles. Elle consiste à prendre pour premier quartile la médiane de la première moitié de la série, et comme troisième quartile la médiane de la seconde moitié.
J'ai appelé « naturelle » cette démarche parce que c'est celle qui a été donnée spontanément par mes élèves qui savaient déjà calculer les médianes (« *50% à gauche, 50% à droite* » !) et à qui je demandais de partager en quatre parts de 25%.

Voyons cependant ce que cela donne sur les quatre séries données en exemple.
Sur la série A (les 40 nombres entiers de 21 à 60), aucun problème : l'effectif étant divisible par 4, on coupe en quatre parts de 25% exactement. La première part va de 21 à 30, la seconde de 31 à 40, la troisième de 41 à 50 et la dernière de 51 à 60. La médiane vaut donc 40,5, le premier quartile 30,5 et le second quartile 50,5.
Sur la série C (nombres de 21 à 62), aucune difficulté non plus : $m = 41{,}5$ (moyenne des 21^{e} et 22^{e} valeurs), Q1 = 31 (11^{e} valeur) et Q3 = 52 (32^{e} valeur).
Sur les séries B et D, la médiane étant une valeur de la série, il faut décider ce qu'on appelle la « première moitié » et la « seconde moitié » ; je fais le choix suivant : la première moitié est constituée de toutes les valeurs dont le rang est strictement inférieur à celui de la médiane, et la seconde moitié des valeurs dont le rang est strictement supérieur à celui de la valeur médiane. La valeur qui a déterminé la médiane n'est donc ni dans la première, ni dans la seconde moitié.
Cette méthode est celle qui est utilisée par les calculatrices graphiques (TI82, TI83 et consœurs). Pour la série B, cela donne Q1 = 30, 5 (moyenne des 10^{e} et 11^{e} valeurs) et Q3 = 51,5 (moyenne des 31^{e} et 32^{e} valeurs).
Voir en annexe 2 une traduction plus formalisée de cette définition.

Les quartiles sur un tableur

Sur un tableur comme Excel, la syntaxe est =QUARTILE(*p1age_de_valeurs* ; k) où k = 1 pour le premier quartile Q1 et k = 3 pour Q3. Bien entendu, k = 2 redonne la médiane.
Tout se passe comme si le tableur créait un « rang fictif » pour faire ses calculs.
Prenons par exemple la série A : le rang du premier individu est 1 ; le rang du dernier est 40. Le « rang fictif » du milieu est donc la demi-somme de ces rangs, soit 20,5. La médiane est donc calculée en faisant la demi- somme de la 20^{e} et de la 21^{e} valeur : on trouve 40,5. Pour le premier quartile, le rang du premier est 1, et le « rang » de la médiane est 20,5 ; le milieu se trouve donc à 10,75. Ce rang fictif étant le barycentre de 10 et 11 affectés des coefficients 1 et 3, le quartile calculé sera le barycentre de la 10^{e} et de la 11^{e} valeurs affectées des coefficients 1 et 3, soit 30,75.
Voir en annexe 1 le tableau des résultats trouvés par le tableur sur les différents exemples.

Définition des quartiles par le GEPS[2]

La définition proposée par ces experts sur les différents documents d'accompagnement (programme actuel de Première L, futurs programmes de S ou de ES) s'appuie sur la définition de la fonction quantile dans le cours de statistique de l'I.N.R.I.A. de Grenoble.

> Le premier quartile 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.
>
> Le troisième quartile est le plus petit élément q' des valeur des termes de la série, ordonnées par ordre croissant, tel qu'au moins 75% des données soient inférieures ou égales à q'[3].

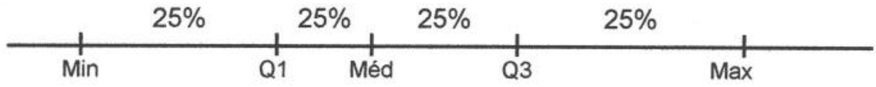
Autrement dit : on trie les valeurs de la série dans l'ordre croissant, et on parcourt cette liste de la gauche vers la droite ; dès que l'on atteint ou dépasse 25% de l'effectif, on est au premier quartile.
Avec cette définition, pour la série A, on trouve respectivement 30, 40 et 50 pour le premier, le second et le troisième quartile de la série A. Voir en annexe les résultats pour les trois autres séries.
Cette définition a deux inconvénients.
Le premier : avec cette définition, la médiane n'est pas le second quartile[4], ce qui est contraire au « bon sens commun ». Le GEPS s'en est bien rendu compte, qui nous propose ceci : « *Nous suggérons de ne pas définir le second quartile mais de manipuler {premier quartile, médiane, second quartile} ; il n'y a pas de raison de signaler qu'avec la définition adoptée, la médiane n'est pas le second quartile* ».
Le second : la définition de la médiane et des quartiles devrait être « essentiellement » symétrique. C'est-à-dire que l'ordre choisi pour trier (ordre croissant ou ordre décroissant) ne devrait pas avoir d'incidence sur le choix de la valeur où l'on place les « barres » de 25%, 50% et 75%.
Or la définition du GEPS n'est pas satisfaisante à ce point de vue: si l'on classait en ordre décroissant la série A, les quartiles vaudraient respectivement 31, 41 et 51 (au lieu de 30, 40 et 50).

Les boites à moustaches

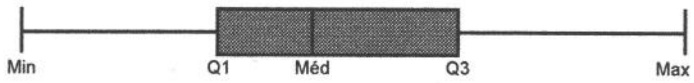
L'objectif est de donner un résumé pertinent d'une série statistique qui soit efficace, donc rapidement « visualisable ». Il s'appuie sur la médiane et les quartiles. La dénomination est très variable : boites ou diagrammes à moustaches, ou boite à pattes, ou diagramme en boite[5], ou encore diagramme de Tukey[6] ; en anglais box plot, ou box-and-whiskers plot.

(2) Groupe d'Experts pour les Programmes Scolaires (ex G.T.D.). C'est le groupe actuellement chargé de rédiger les programmes et les documents d'accompagnement.
(3) Citations du document du GEPS du 22/12/2000 intitulé « Quantiles et diagrammes en boite ».
(4) C'est, à notre connaissance, le seul « lieu d'enseignement » où la médiane n'est pas le second quartile.
(5) C'est la dénomination « officielle » du programme de Première L
(6) John Wilder TUKEY, né le 16/06/1915 à New Bedford (USA) et mort récemment, le 26/07/2000 à New Brunswick, fut après la seconde guerre mondiale simultanément professeur à Princeton et chercheur chez AT&T-Bell. Il travailla essentiellement dans le domaine de la

On passe de cette représentation (qui correspond aux valeurs trouvées pour la médiane et les quartiles) :



à celle-ci :



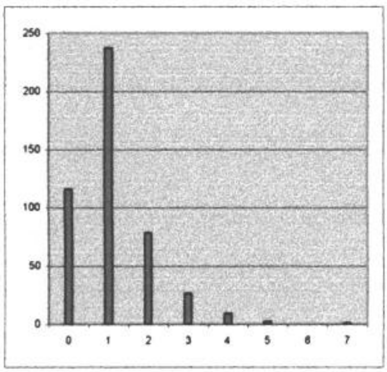
Il est bien entendu que cela n'a de sens que pour un caractère numérique, et que l'axe horizontal doit respecter la graduation. Le rectangle ici grisé correspond aux 50% « centraux » (ou mieux, « médians »). C'est ce « corps » de la boite qui est l'élément essentiel, les moustaches ne jouant qu'un rôle secondaire dans cette représentation. Des moustaches très courtes indiquent une très forte concentration d'individus sur un petit intervalle, au contraire de moustaches très longues (voir ci-après).
Les boites à moustaches permettent de comparer très facilement des échantillons correspondant au même caractère statistique, en les plaçant parallèlement les unes aux autres, relativement au même axe gradué. Selon les auteurs, ces boites sont placées horizontalement ou verticalement : les élèves doivent avoir rencontré les deux, pour ne pas être « surpris » le jour de l'examen.

Opportunité des boites à moustaches

Attention, il y a des exemples où médiane et boite à moustache ne sont pas du tout des indicateurs (résumés) pertinents

Premier exemple : calculer des déciles sur une série de 25 notes d'élèves !

Deuxième exemple : dans une commune (fictive) il y a 116 familles sans enfant, 237 familles de 1 enfant, 78 familles de 2 enfants, 26 familles de 3 enfants, 9 familles de 4 enfants, 2 familles de 5 enfants et 1 famille de 7 enfants. Les calculs donnent (quelle que soit la définition choisie) la médiane, le premier quartile et le troisième quartile tous trois égaux à 1.
Le meilleur résumé que l'on puisse faire est un petit tableau reprenant l'intégralité de l'information, ou de le remplacer par un diagramme en bâtons (voir figure). À la rigueur on pourrait donner le nombre moyen d'enfants par famille Mais surtout pas une boite à moustache !!!



statistique, en particulier dans le domaine de l'analyse de variance. On trouvera de plus amples renseignements sur http://wwwgroups.dcs.st-and.ac.uk/-history/Mathematicians/Tukey.html par exemple.

Faut-il élaguer les extrémités des moustaches ?

Prenons un exemple : dans une série statistique E de 100 valeurs (ordonnées), les 25 dernières valeurs sont 66, 66, 67, 67, …, 85, 86, 88, 91, 91, 92, 94, 95, 96 et 97 ; dans une autre série, F, les 25 dernières valeurs, sont 66, 66, 67, 67, …, 85, 86, 88, 91, 91, 92, 94, 95, 96 et 124. L'étendue du dernier quartile vaut 31 dans la série E, et 58 dans la série F ; de même l'étendue du dernier décile vaut 12 dans la série E, et 39 dans l'autre. On se rend compte combien un seul élément (ici le maximum) augmente la longueur de la moustache : autant la médiane et les quartiles (et donc l'intervalle inter-quartile) sont des indicateurs stables pour des variations des valeurs du caractère, autant les « moustaches » sont sensibles à une modification des valeurs extrêmes (ce qui importe, c'est donc bien le corps de la DAM et pas sa moustache !).

Encore un exemple, qui correspond à une activité faite en classe de première L. Les élèves avaient à couper des « bouts de ficelle » de 18 cm de long, sans aucun instrument de mesure à leur disposition (pas même une feuille de papier A4) Voici le relevé des longueurs obtenues

14	4											
15	0	2										
16	2	5	5									
17	0	0	0	1	5	4	4	6	9			
18	0	0	1	2	2	3	3	5	5	5	5	6	7
19~	0	0	0	0	0	1	1	3	5	5	5	8	
20	0	5	5	6	8							
21	0	2	5									
22	5	8	8	9								
23	2	2	2	5	9							
24	0	8										
25	5	8										
26	0	2	5									
27	5	5	7									
28	0											
29												
30	0											
31	0	5										
32	0	0										
33	5											
34												
35												
36	5											
37	5											
38												

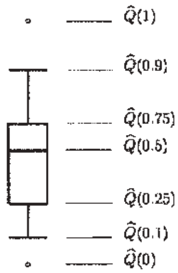
La méthode utilisée pour représenter les données est celle du diagramme « en tiges et feuilles ». Il se lit de la façon suivante: la plus petite valeur est 14,4 cm ; ensuite viennent 15,0 et 15,2 cm. Il n'y a aucune mesure dont la partie entière soit 29.

Et on a bien envie de dire que les valeurs 36,5 et 37,5 sont « aberrantes », et qu'elles ne devraient pas figurer sur le diagramme.

C'est pourquoi les statisticiens ont décidé de tailler les « pointes » des moustaches.

DAM élaguées

Les uns élaguent aux déciles : c'est la méthode préconisée par le GEPS. On coupe 10% de chaque côté (on arrête donc les moustaches à D1 et à D9, voir schéma ci-contre, issu de http://www.inrialpes.fr/sel/).

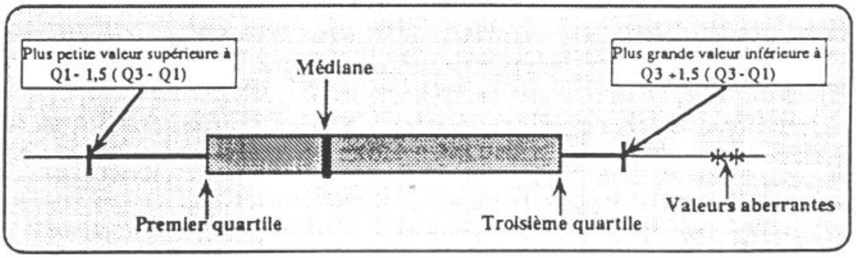


BAM élaguée aux déciles

Les autres (c'était le cas de Tukey, « inventeur » de ces diagrammes, c'est le cas de la nouvelle version de « Casimir », logiciel d'exploitation de l'évaluation à l'entrée en sixième, voir schéma ci-dessous, et des calculatrices graphiques que possèdent les élèves) ôtent tout ce qui dépasse **1,5 fois l'intervalle inter-quartile**.

D'autres – moins nombreux – retirent 10% en tout (soit 5% de chaque côté), ou encore 5% en tout (2,5% de chaque côté). Nous ne prendrons parti ni pour les uns, ni pour les autres, ni pour les « intégristes » qui ne veulent rien couper…

Avec l'exemple des bouts de ficelles ci-dessus, le premier quartile vaut 18,2 cm, la médiane vaut 19,5 cm, le troisième quartile vaut 24,8 cm. D'où un intervalle inter-quartile large de 6,6 cm. On « coupe » donc les moustaches à 18,2 - 1,5 * 6,6 = 8,3 cm, et à 24,8 + 1,5 * 6,6 = 34,7. Comme 8,3 est inférieur au minimum, la moustache ne sera pas élaguée à gauche. Par contre 34,7 est inférieur au maximum, et là, la moustache s'arrêtera à cette valeur. Cette méthode a « élagué » les deux valeurs 36,2 cm et 37,5 cm, considérées alors comme « aberrantes ».



BAM élaguée à 1,5 × (Q3 – Q1)

Bien sûr, pour que l'on se rende compte de ce qui a disparu de la série, on peut signaler sur l'axe les points où se trouvaient le maximum et le minimum, et – le plus souvent – les points correspondant aux valeurs ainsi « supprimées ».

En outre, nommer « valeurs aberrantes » les valeurs que l'on a fait disparaître peut être admissible quand il s'agit de relevés de mesures ou de contrôles de fabrication, mais pose des problèmes d'éthique quand il s'agit des poids ou des scores des élèves.

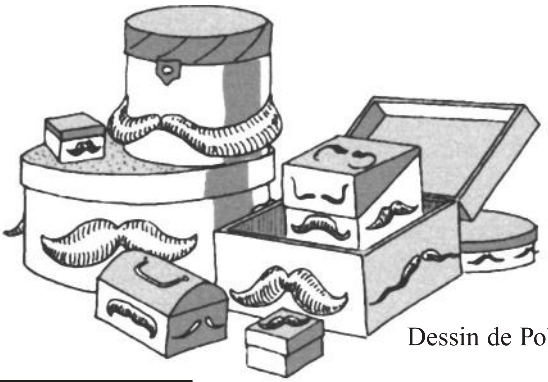
(7) Je ne décrirai pas plus ce que j'entends par là. Disons que c'est ici une estimation à la simple lecture du graphique représentant les valeurs, Mais le terme est utilisé dans le programme de Première L…

Et quand la statistique est (à peu près) « normale » ?

Dans le cas où on a une distribution sensiblement « gaussienne(7) », enlever 10% de chaque côté correspond (approximativement) à s'arrêter à la moyenne plus ou moins 1,3 fois l'écart-type. Mais arrêter la moustache aux valeurs déterminées par 1,5 fois l'intervalle inter-quartile correspond, sur une distribution « normale », à ne considérer comme « aberrantes » que 0,7% des valeurs (0,35% de chaque côté) ; ce qui me paraît plus raisonnable que d'en enlever 20% en tout…
Mais toutes les séries statistiques ne sont pas gaussiennes, loin de là.
La série des valeurs des ficelles du paragraphe précédent en est déjà un exemple.
Prenons comme autre exemple la population des 594 communes de la Meurthe-et-Moselle(8) : la plus petite commune (Leménil-Mitry) a 2 habitants ; la plus grosse (Nancy) en a 103 606. La médiane vaut 263,5 (donc la moitié des 594 communes ont 263 habitants ou moins) ; 50% des communes ont entre 130 et 659 habitants (intervalle inter-quartile) ; 10% des communes ont plus de 2404 habitants. Ces 10% des communes les plus peuplées représentent, à elles seules, près de 71% de la population du département. On voit à quoi conduirait, sur de telles séries, le « taillage » des pointes de moustaches !!!

Conclusion

Cet article avait pour objectif de montrer qu'il est parfois difficile de donner des définitions précises de ce que l'on veut (ou plutôt qu'il est difficile de choisir entre des définitions tout aussi valables les unes que les autres).
Ce que l'élève doit retenir, c'est qu'on dispose d'un outil permettant de représenter simplement entre quelle et quelle valeur sont les 50% « centraux » d'une population, et d'interpréter correctement cette information(9).
Sur une série statistique constituée de plusieurs centaines ou plusieurs milliers d'individus, peu importera la « définition » des quartiles choisie.
Cet article montre aussi, en filigrane, qu'il faut rester pertinent par rapport à ce que l'on étudie : il n'y a pas de « formule » universelle ; c'est aux mathématiques de s'adapter aux situations rencontrées, pour essayer d'en rendre compte le mieux possible.`



Dessin de Pol LE GALL

(8) Source: recensement de 1999.
(9) Voir en annexe 3 un exemple de contrôle où une telle interprétation est demandée dans un cas concret.

ANNEXE 1

Résultats donnés par les diverses définitions des quartiles

	Série A 40 valeurs (entiers de 21 à 60)	Série B 41 valeurs (entiers de 21 à 61)	Série C 42 valeurs (entiers de 21 à 62)	Série D 43 valeurs (entiers de 21 à 63)
Définition 'naturelle' et calculatrices TI :				
Q1	30,5	30,5	31	31
Médiane = Q2	40,5	41	41,5	42
Q3	50,5	51,5	52	53
Calculs faits par Excel :				
Q1	30,75	31	31,25	31,5
Médiane = Q2	40,5	41	41,5	42
Q3	50,25	51	51,75	52,5
Calculs faits avec 'Statistica' et 'StatGraphics' :				
Q1	30,5	31	31	31
Médiane = Q2	40,5	41	41,5	42
Q3	50,5	51	52	53
Définition du GEPS :				
Q1	30	31	31	31
Médiane	40,5	41	41,5	42
Q2	40	41	41	42
Q3	50	51	52	53
Remarque. Pour la définition du GEPS, si l'on ordonne les valeurs par ordre décroissant, on obtient pour la série A :Q1 = 31, Q2 = 41 (mais Médiane = 40,5) et Q3 = 51.				

ANNEXE 2

Calcul des quartiles selon leur définition « naturelle »

Ce qui suit est une formalisation de ce qui est énoncé dans le premier paragraphe. Cela correspond au cours qui est donné aux étudiants du département « Statistiques et Traitement Informatique de Données » de l'I.U.T. de Metz.
Il peut fournir l'occasion d'un petit calcul en classe de terminale S, où l'on travaille, en arithmétique, sur la division euclidienne.

Soit N l'effectif total de la statistique.

1º) Si N = $4k$, on partage la série (triée) en quatre « paquets » de k individus. Les valeurs des quartiles sont respectivement $(x_k + x_{k+1})/2$, $(x_{2k} + x_{2k+1})/2$, et $(x_{3k} + x_{3k+1})/2$.
En effet, chacun des quatre « paquets » représente exactement 25% de la population et aucun des trois quartiles n'est alors une valeur de la série.

2º) Si N = $4k + 1$, la médiane correspond à la valeur « centrale » : x_{2k+1}. Il reste alors, de part et d'autre, deux « paquets » de $2k$.
Les premier et troisième quartiles valent donc respectivement $(x_k + x_{k+1})/2$ et $(x_{3k+1}+ x_{3k+2})/2$.
Dans ce cas, seule la médiane est une valeur de la série.

3°) Si N = $4k + 2$, on partage la série en deux « paquets » de (2k + 1) individus, La médiane vaut alors $(x_{2k+1} + x_{2k+2})/2$, ce qui n'est pas une valeur de la série. Chacun des deux paquets est repartagé en deux , leur effectif étant impair, les quartiles sont donc des valeurs de la série : $Q1 = x_{k+1}$ et $Q3 = x_{3k+1}$.

4°) Si N = $4k + 3$, on partage alors la série en 4 paquets de k, séparés par les valeurs x_{k+1}, x_{2k+3} et x_{3k+3} qui sont respectivement les trois quartiles.

Avec cette méthode de calcul, la médiane est bien toujours égale au second quartile, et les valeurs calculées sont les mêmes selon que l'on range la suite dans l'ordre croissant ou dans l'ordre décroissant.
Cette méthode est très aisément programmable, puisqu'elle est basée sur le calcul du reste de la division de N par 4.

On pourrait l'étendre au calcul des déciles, en se basant sur le reste de la division de N par 10 : un excellent exercice pour les TS.

ANNEXE 3

Un exemple de contrôle sur les quartiles, boites à moustaches, etc.

Ce contrôle a été donné en classe de première littéraire (20 élèves présents).
Dans le premier exercice, on donnait aux élèves une série statistique (triée), et on leur demandait de calculer médiane, quartiles et déciles (le premier et le neuvième), puis de dessiner deux boites à moustaches, l'une non élaguée, l'autre élaguée aux déciles.

Voici l'énoncé du second exercice :

Une association de consommateurs relève les poids des baguettes dans trois boulangeries. En principe, les baguettes devraient peser 250 grammes.

Relevé de la boulangerie A :
Relevé portant sur 1248 baguettes. Poids minimum: 215 g. Poids maximum : 285 g. Poids médian: 250 g. Quartiles, 235 g et 265 g.

Relevé de la boulangerie B :
Relevé portant sur 908 baguettes. Poids minimum : 215 grammes. Poids maximum : 285 g. Poids médian : 255 g. Quartiles : respectivement 250 et 260 g.

Relevé de la boulangerie C :
Relevé portant sur 1035 baguettes. Poids minimum : 230 grammes. Poids maximum : 270 g. Poids médian : 250 g. Quartiles : respectivement 245 et 257 g.

1°) Réaliser les trois boites à moustache (non élaguées) sur le même graphique, l'une au-dessus de l'autre.

2°) Commenter les affirmations suivantes :

a) Comme le poids « normal » est 250 g, seul le boulanger B est dans la norme, c'est à dire qu'il est en règle vis-à-vis de la répression des fraudes.

b) Si vous achetez votre baguette chez le boulanger A, il y a une chance sur deux qu'elle ne fasse pas les 250 g attendus.
c) Si vous achetez votre baguette chez le boulanger C, il y a une chance sur deux qu'elle ne fasse pas les 250 g attendus.
d) Celui qui a le moins de variation dans ses poids est le boulanger C.
e) À la boulangerie B, les trois quarts des baguettes font au moins le poids réglementaire.
f) Il vaut mieux acheter son pain chez les boulangers A ou B, car le poids maximum relevé est 285 g (contre 270 g chez C).
g) Chez le boulanger C, les baguettes pèsent en moyenne 256 grammes.

QUELQUES COMMENTAIRES

En ce qui concerne les **boites à moustache,** elles sont bien réussies par 14 élèves sur 20. Deux élèves ont dessiné des boites en quatre parties de longueurs égales (puisqu'il y a quatre fois 25%) : leurs boites sont toujours superposables, quelle que soit la série donnée. Pour les autres, l'erreur la plus fréquente consiste en une échelle assez « élastique » des valeurs de la variable.

En ce qui concerne le **second exercice**, j'ai été surpris par la longueur des justifications et explications apportées, parfois jusqu'à 5 ou 6 lignes à chaque item. Mais ces explications sont souvent confuses, et sans aucun rapport avec ce qui est demandé.

<u>Question a</u> : Une se 5 réponses comme correctes. La majorité a répondu « Non, c'est 250 g » (erreur attendue). Je cite une des réponses : « *C'est faux : les baguettes ne sont pas proportionnelles les unes aux autres à chaque fois, à cause des quartiles* ».

<u>Question b et c</u> : Il n'y avait aucune ambiguïté, la moitié des baguettes étant au-dessus de 250 g et l'autre moitié au-dessous. Beaucoup de réponses font intervenir, dans une savante alchimie, les quartiles, les extrêmes, le nombre de baguettes testées ; en somme « l'âge du capitaine » (cf. Stella Baruk). Six élèves sur les 20 ont trouvé qu'une des deux affirmations était vraie et l'autre fausse.

<u>Question d</u> : Selon que l'on s'intéressait à l'étendue ou à l'intervalle inter-quartile, la réponse différait. J'ai bien sûr considéré comme correcte toute réponse clairement argumentée.

<u>Question e</u> : Il n'y a aucune ambiguïté possible ici (c'est la définition même du quartile), et je m'attendais à beaucoup de réponses exactes. Je n'en ai eu que 7 (je ne compte pas comme bonnes les réponses du type « *C'est vrai car l'intervalle inter-quartile est de 10 grammes* » ou autres).

<u>Question f</u> : Je m'attendais à une variété de réponses plus ou moins floues ou ambiguës, je n'ai pas été déçu. Certaines ont même fait intervenir le prix de vente dans leur argumentation…

<u>Question g</u> : Je considère seulement 5 réponses comme correctes. La majorité a répondu « Non, c'est 250 g » (erreur attendue). Je cite une des réponses : « *C'est faux : les baguettes ne sont pas proportionnelles les unes aux autres à chaque fois, à cause des quartiles* ».