Quartiles et dispersion

Cette note se rattache à cet extrait des programmes : « Le recours aux quartiles permet de préciser la dispersion d'une série par rapport à la seule notion d'étendue ». Elle annonce les notions d'intervalles interquartiles et de boîte à moustache qui figurent actuellement au programme de Première et concerne aussi les programmes de Troisième et de Seconde.

1 Résumé d'une série statistique

Soit $(\mathcal{S}) = (x_1, \cdots, x_n)$ une série statistique. On peut la résumer par le tableau suivant :

min	q_1	m	q_3	max	e	M

où min, q_1, m, q_3, max, e et M désignent respectivement le minimum, le premier quartile, la médiane, le troisième quartile, le maximum, l'étendue et la moyenne de $(\mathcal{S})$. Le résumé ne comprend donc que 7 termes alors que $(\mathcal{S})$ en comprend habituellement plusieurs milliers. On perd donc beaucoup d'information au bénéfice d'une grande simplicité [1].

On remarquera que min, q_1, m, q_3, max d'une part, M d'autre part sont des caractéristiques de $(\mathcal{S})$ de natures complètement différentes : en effet, M est une moyenne arithmétique (ou un barycentre) que l'on calcule à l'aide de l'addition et de la multiplication tandis que min, q_1, m, q_2, max ne dépendent que de la relation d'ordre $\leq$. L'étendue est une notion mixte en ce sens qu'elle se calcule à l'aide de $\leq$ et de la soustraction.

On peut dire que min, q_1, m, q_3, max et M sont des caractéristiques (ou paramètres) de position de la série statistique $(\mathcal{S})$ en ce sens que ces nombres nous renseignent sur la position qu'aurait $(\mathcal{S})$ si on la représentait sur un axe gradué. e est une caractéristique (ou paramètre) de dispersion en ce sens que plus l'étendue est grande, plus les valeurs qui composent $(\mathcal{S})$ sont dispersées.

Sans la moyenne, le résumé ci-dessus nous paraîtrait très insuffisant car la moyenne est la caractéristique de position la plus importante et la plus naturelle de $(\mathcal{S})$.

2 Quartiles et qualité de la dispersion d'une série statistique

Que peut-on dire des résumés suivants de deux séries statistiques ?

min	q_1	m	q_3	max	e	M
1,1	1,16	2,48	2,83	2,89	1,79	2,16

et

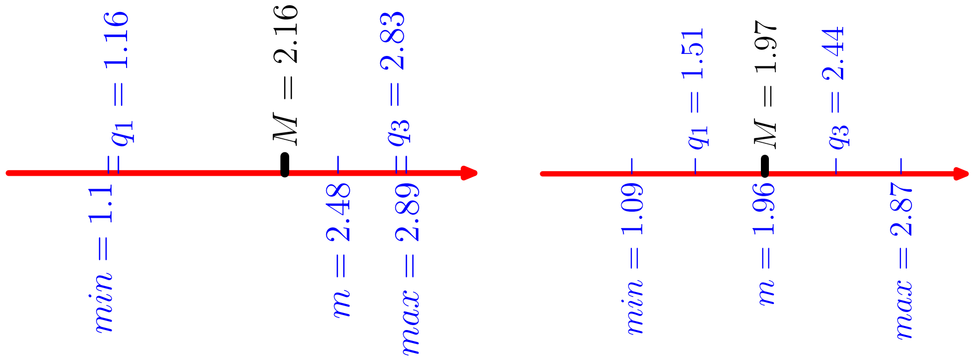
min	q_1	m	q_3	max	e	M
1,09	1,51	1,96	2,44	2,87	1,78	1,97

On peut dire que ces deux séries sont à peu près également dispersées puisque leurs étendues

1. Ce tableau est aussi riche qu'une boîte à moustache.

valent respectivement 1,79 et 1,78.

Mais elles sont dispersées de manières très différentes, ce que montrent les écarts $q_1 - min$, $m - q_1$, $q_3 - m$, $max - q_3$ [2]. Les choses apparaissent plus clairement si l'on place les points min, q_1, m, q_3, max et M sur un axe gradué. Cela donne :



<u>Deuxième série :</u> On voit que les points q_1, m et q_3 sont répartis régulièrement entre min et max. Comme entre deux valeurs consécutives se trouvent à peu près un quart des valeurs de cette série, on pourra se risquer à dire que la deuxième série est dispersée de manière assez homogène.

<u>Première série :</u> Au contraire, il y a à peu près un quart des valeurs de la première série qui sont très près de son minimum et à peu près un quart de ses valeurs qui sont très près de son maximum. La dispersion des valeurs de cette série est donc plus grande.

<u>On ne connaît pas la nature des deux séries résumées ci-dessus</u>. Comme il s'agit de séries réparties de manières très différentes, on peut faire penser qu'elles concernent deux grandeurs différentes. On doit noter à ce sujet que l'on pourrait, si l'on manquait de sens critique, comparer par exemple une série de poids de lombrics de nos jardins et une série de longueurs de sauts de kangourous en Australie. Si l'on découvrait par exemple une grande analogie ou au contraire une grande différence dans leur dispersion, on pourrait sans doute alimenter quelques articles de journaux bien sentis. Mais ce serait absurde.

Références

1 - Programme de l'enseignement des mathématiques, des SVT, de physique-chimie du collège, B.O. N°6 19 avril 2007, Hors-série, Annexe 2 Mathématiques

ftp://trf.education.gouv.fr/pub/edutel/bo/2007/hs6/MENE0750668A_annexe2.pdf

2 - Mathématiques Collège, Projet de document d'accompagnement – Probabilités – 17 mars 2008

http://eduscol.education.fr/D0015/doc_acc_clg_probabilites.pdf

3 - Programme de mathématiques de la classe de Seconde, Bulletin officiel n° 30 du 23 juillet 2009 :

http://media.education.gouv.fr/file/30/52/3/programme_mathematiques_seconde_65523.pdf.

2. Cela aurait paru plus convaincant si on avait interprété min par q_0 , max par q_4 et si la médiane m avait été définie comme q_2 !