Dialogue

Emmanuelle Boyer, professeur de mathématiques au lycée d'Aurillac

Claudine Schwartz, professeur des universités, statisticienne.

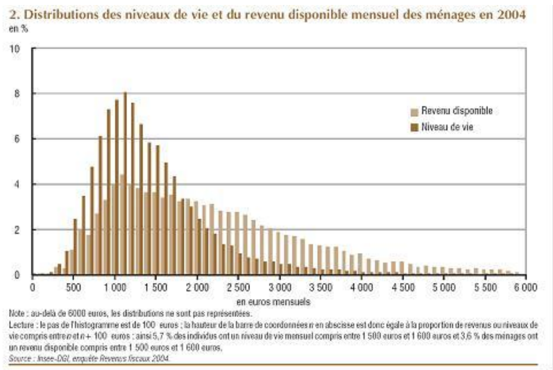
La lecture du document « éléments de corrigé » a suscité quelques questions et commentaires de Claudine Schwartz. Emmanuelle Boyer y a répondu, témoignant des enjeux et difficultés de la collaboration entre disciplines telle qu'elle est vécue sur le terrain des lycées.

Nous espérons que ce dialogue se poursuivra dans les lycées et sur le site Statistix.

Les numéros ci-dessous correspondent à ceux qui ont été placés dans le document « éléments de corrigés ».

(1) **C.S.** On fournit aux élèves le graphique reproduit ci-dessous. D'où vient-il ? Ce serait bien de le savoir : a-t-il été fait par les auteurs d'un ouvrage scolaire, ou par l'INSEE par exemple ? On aurait pu faire côte à côte, ou l'un sous l'autre les deux histogrammes correspondants. Je trouve ce graphique franchement difficile à interpréter parce-que dans un des diagrammes, l'unité est le ménage, dans l'autre, l'unité est l'individu.

E.B. C'est un graphique de l'INSEE, je n'ai pas trouvé de données permettant de le reprendre, mais il n'a pas posé de problèmes aux élèves.



Le graphique est dans l'étude :
http://www.insee.fr/fr/ffc/docs_ffc/ref/revpmen06b.pdf
Des compléments sur le sujet :
http://insee.fr/fr/ffc/docs_ffc/ref/revpmen06d.pdf
http://www.insee.fr/fr/ffc/docs_ffc/ref/revpmen06g.pdf

(2) **C.S.** J'ai du mal à avoir une lecture du décile aussi précise que « 11500 » sur le graphique

E.B : Chaque élève fournit une valeur plus ou moins proche de celle-ci ; dans le corrigé est donnée la valeur calculée par l'INSEE directement à partir des données brutes.

(3) **C. S.** Vous écrivez :

Attention : D1,Q1, Me, Q3 et D9 sont des valeurs du caractère de la série statistique.

Mais suivant les définitions, les quartiles et la médiane ne sont pas toujours des valeurs prises par la série

E.B. Pour les conventions sur les quartiles et les diagrammes en boîte, je reprends à l'oral les conventions en mathématiques au lycée et je me désole que l'on piège parfois les élèves au bac la dessus ce qui nous oblige à ne pas faire confiance aux valeurs trouvées à la calculatrice (qui de toute façon ne donnent pas toujours les mêmes résultats selon les types de calculatrices) …. Dans cette phrase, ce que je voulais dire, c'est que ces quantités ont même dimension (i.e. même unité, ici des euros) que les termes de la série.

C.S. Ce n'est pas évident que les conventions de choix sur la médiane et les quartiles soient les mêmes dans toutes les disciplines enseignées au lycée. On pourrait admettre au bac toutes les définitions usuelles. Si on dit une fois en mathématiques que la présence d'ex aequo complique la définition (on ne peut pas dire de la série 0 1 1 1 1 1 2 3 qu'il y a 50% des données inférieures ou égales à la médiane 1 et 50 % des données supérieures ou égales à 1) et qu'il y a des nuances dans les définitions, il faut installer l'image que la médiane coupe la série en deux , le premier quartile en ¼,3/4, etc.

(4) **C.S.** Vous écrivez :

En maths, on demandera souvent la valeur de l'écart interquartile : Q3-Q1 = 36000-16000 = 20000 €

Interprétation : l'écart des revenus annuels disponibles des ménages entre les 25 % les plus riches et les 25% les plus pauvres est supérieur à 20 000 €.

En SES , on demandera plutôt les rapports inter déciles : exemple $\frac{D9}{D1} = \frac{50000}{11500} \approx 4,3$

Interprétation : le revenu des 10% les plus riches est au moins 4,3 fois supérieur au revenu des 10 % les plus pauvres.

Ces formulations de l'interprétation, dans le cadre d'un enseignement (ce ne serait pas le cas si je voyais cela écrit dans la presse), me gênent un peu : que signifie « les revenus disponibles » puisqu'à l'intérieur de chaque décile, ces revenus varient et la variation est grande pour les hauts revenus ? Je dirais plutôt : le revenu d'un ménage parmi les 10% les plus riches est au moins 4,3 fois supérieur au revenu d'un ménage parmi 10 % les plus pauvres. C'est l'occasion d'activer en SES les réflexes sur les inégalités avec les fractions compte tenu que le neuvième décile est le plus petit salaire de la dernière tranche et le premier décile le plus grand salaire de la première tranche. Les interprétations des indices faites sur la base d'arguments de mathématiques élémentaires me paraissent propices à mieux comprendre lesdits indices.

Il me semble qu'on parle aussi dans les journaux du rapport entre la moyenne des salaires des 10% les plus riches et celle des 10% les plus pauvres, ce qui personnellement me parle plus.

E.B. Pour l'écart interquartile, c'est d'un commun accord avec le professeur de SES que nous avons choisi cette phrase. Mais on illustre à l'oral en prenant effectivement un individu dans un groupe et un autre dans l'autre ; en fait on passe un long moment à exploiter le diagramme en boîte et les commentaires viennent un peu en vrac après échange avec les élèves, c'est le moment clé. Pour D9/D1, les collègues de SES y tiennent et avec cette interprétation : ce sont leur tableaux « habituels », donc c'est peut-être seulement un intérêt pédagogique, mais c'est aussi ce que nous recherchons ! Et du coup nous (profs de maths) reprenons toujours avec votre phrase mais ce n'est pas vraiment le sens donné par les collègues de SES, je l'ai précisé à l'oral aux élèves toujours dans la discussion avec le diagramme en boîte.

(5) **C.S.** Vous écrivez :

Remarque : en SES, On emploie aussi le terme de *décile* pour désigner les ensembles (au nombre de 10) des données situées entre deux déciles (déciles pointés) par exemple pour D6 : ensemble des valeurs de la série obtenues entre 50 % et 60 % des effectifs. On obtiendrait pour D6 :] 25 000 ; 29 000]. Les 10 % des ménages ayant un revenu supérieur à celui des 50 % les moins riches et inférieur à celui des 40 % les plus riches, ont un revenu compris entre 25 000 € et 29 000 €.

Cela me parait assez naturel de nommer décile à la fois les valeurs et les intervalles (ce serait compliqué de toujours parler d'écart inter décile). Par contre « Les 10 % des ménages ayant un revenu supérieur à celui des 50% les moins riches et inférieur à celui des 40 % les plus riches » franchement, c'est dur comme formulation, il faut s'accrocher ! J'ai du mal à imaginer qu'on ne décroche pas une bonne partie des élèves avec ce genre de formulation, le décrochage portant sur l'idée de manipuler soi-même, activement, les pourcentages. Pourquoi ne pas dire :

50% des ménages ont un revenu inférieur à 25000 et 40% ont un revenu supérieur à 29000 et donc 10% ont un revenu entre 25 000 et 29 000 : cette tranche de salaires s'appelle aussi le 6ème décile.

E.B. Il faut trouver une formulation commune, c'est notre démarche, ou bien mettre en évidence face aux élèves ce que chacun des profs dans sa discipline met derrière comme « sens ». Là c'est clairement du domaine des SES. Pour l'interprétation, nous avons privilégié la vision par « tranche » des SES ; mais nous nous sommes souvent « affrontés » en salle des profs sur la vision des déciles en math et en SES !

Ce qui a été dur pour les élèves, alors que la collègue de SES ne s'y attendait pas, c'est sur les *au plus*, *au moins*, *pas plus*. Nous menons en maths un travail là dessus, avec la statistique et les probabilités, depuis le collège, mais ça reste dur pour certains élèves. En SES, pour interpréter les graphiques, on demande aux élèves de reproduire une phrase type en changeant les valeurs, sans avoir forcément besoin d'en comprendre le sens - j'exagère un peu ! - d'où l'intérêt de ce TP à deux voix.

(6) Vous écrivez

Ainsi, la moyenne des salaires de tous les ménages est égale au milieu de la moyenne des revenus des 50 % les plus riches et de la moyenne des 50% les moins riches.

Pourquoi milieu de moyennes plutôt que demi-somme ??

E.B. J'ai choisi « milieu » par rapport au diagramme en boîte, pour des élèves de ES, c'est pas facile ; la première phrase avec moyenne de moyenne et sous groupes est restée obscure, j'essaie toujours de revenir à du concret pour ne pas perdre ceux qui sont en difficulté ; mais dans un corrigé pour approfondissement c'est effectivement à discuter.

(7) **C.S.** Vous écrivez

Dans ce type de graphique où les abscisses sont les déciles (7) (cumul croissant de fréquences) et en ordonnées les parts d'une grandeur étudiée (cumul croissant de pourcentages) , les courbes obtenues s'appellent des courbes de Lorenz. Elles mettent en évidence la répartition de la grandeur étudiée pour permettre des comparaisons :

Les abscisses ne sont pas les déciles ; les déciles sont à la fois des tranches de revenus et les extrémités ce ces tranches, mais 10%, 20%, etc. sont des pourcentages et pas des déciles ! On peut rappeler que chaque point de la courbe est associé à un décile. On pourrait faire réfléchir les élèves sur l'intérêt qu'il y aurait à mettre les déciles en abscisses, et les mêmes ordonnées que dans la courbe de Lorenz. C'est toujours intéressant de se poser la question du pourquoi ce choix de 10%, 20 % et pas un autre qui a l'air sensé a priori. En particulier, si on mettait les valeurs des déciles, ce serait difficile de faire des comparaisons entre pays par exemple ou entre des quantités qui ne varient pas sur la même échelle.

E.B. Pour les collègues de SES, ce sont clairement ce qu'ils appellent les déciles en abscisses par abus de langage et cela me parait-être notre rôle de prof de math de travailler sur et avec cet abus de langage. Ce sont les profs de SES qui utilisent les courbes de Lorenz et ce sont eux qui ont besoin ; les tableaux viennent de l'Insee avec cette formulation. Nous nous mettons tous d'accord avec les élèves pour qu'il n'y ait pas d'ambiguïté de sens.

(8) **C.S.** Vous écrivez :

En utilisant un tableur, ajustons les données à l'aide d'un ajustement polynomial pour le revenu disponible 2004. (il existe bien d'autres ajustement pour une courbe de Lorenz) (9)

On obtient : $f(x) =$ ($3{,}3173x^5 - 6{,}6353x^4 + 4{,}663x^3 - 0{,}7235x^2 + 0{,}3781x - 0{,}0011$ trouvée au vidéo avec EXCEL en direct)

On ajuste une courbe passant par deux points fixes (l'origine et le point (1,1)) et dont on connait 9 points, par un polynôme de degré 6 ; comme la constante doit être nulle, on a 6 paramètres à estimer : ce n'est pas très risqué et assez inhabituel : on essaye d'être parcimonieux en paramètres. De plus, il y a beaucoup de décimales à ces paramètres ! Mais enfin, c'est de l'ajustement graphique avec un outil surpuissant pour la question.

(9) **C.S.** Vous écrivez aussi :

De même la courbe représentant du patrimoine peut être modélisée par la courbe représentant la fonction :

$$g(x) = \frac{1}{e^{0,6}} x^3 e^{(0,6\,x^4)} \quad \text{pour} \quad x \in [\,0\,,1\,]$$

Alors là, rien ne va plus : pourquoi une fonction pareille ? Je comprends la puissance 3 de *x* devant l'exponentielle pour si vous aviez derrière la tête d'avoir une fonction à primitive connue, mais pourquoi une puissance quatrième de la variable sous l'exponentielle ? **Et surtout, c'est de l'ajustement et pas un modèle.** Pour faire un modèle, on commence par réfléchir à ce qu'on va modéliser et c'est une étape difficile. Ici, on modélise la fonction de répartition des revenus, (ou sa dérivée la densité des revenus) puisque c'est elle qui est à la base de la courbe de Lorenz. Les fonctions candidates doivent avoir certaines propriétés, théoriques (comme celle que vous énoncez pour la courbe de Lorenz), mais on cherche aussi à avoir peu de paramètres et que ceux-ci, ou des fonctions de ceux-ci, soient interprétables dans le domaine traité. Ensuite, à partir de cette répartition des revenus, dont on a estimé les paramètres avec les données (ici la répartition empirique des revenus) on calcule l'indice de Gini (c'est simple à faire) et des tas d'autres choses. Par contre, revenir à la distribution des revenus à partir d'une équation de la courbe de Gini n'est pas naturel, voire impossible, donc modéliser la courbe de Lorenz serait « pauvre ».

E.B. Nous avons lu les ajustements au vidéoprojecteur ; c'est la méthode qui me paraissait importante et habituer les élèves à voir des paramètres non entiers ; j'ai galéré pour trouver une fonction pour le patrimoine, convexe, qui ne soit pas négative au début et qu'on puisse exploiter en terminale ; mais vous avez peut-être mieux. Je suis d'accord avec vous, j'ai fait de l'artificiel, et même en connaissance de cause. Mais c'est ce qu'on me demande de faire en terminale et comme je dis aux élèves : le jour du bac vous ne réfléchissez pas au sens ! Mes exercices types bac de référence sur les courbes de Lorenz sont : bac ES, la réunion 1998, France métropolitaine septembre 2003, France métropolitaine septembre 2002 et autres ; ils sont faciles à trouver sur le site de l'APMEP. Mes élèves les cherchent en autonomie avec le corrigé, il y en a aussi 5 autres exercices sur le prix d'équilibre et enfin, je choisi l'un des 8 pour le contrôle (il y a parfois un retirage si je trouve que l'exercice tiré au sort est trop dur). Ils sont ce qu'ils sont, artificiels certes, mais ils nous ont conforté dans notre travail (et sa justification)!

Calculer des indices de Gini avec des intégrales me parait artificiel ; j'ai beaucoup vu sur des cours du supérieur sur internet le calcul fait avec les sommes et les trapèzes, mais là c'est déjà trop long et les élèves sont lâchés par les maths ; alors j'ai abandonné volontairement les trapèzes et j'ai fait sans conviction le calcul d'intégrales, je crois que pour eux c'est plus simple d'être directement placé dans l'exercice de maths et de calculer l'intégrale d'un polynôme.

C.S. En l'absence de modèle (et un ajustement n'en est pas un), il convient de calculer l'aire sous la courbe de Lorenz en sommant les aires des trapèzes. Cela peut être amusant de faire établir aux élèves la formule qui donne cette aire A ; ce n'est qu'une fonction affine de la somme S des ordonnées de points, à savoir pour les déciles $A=h(0,5+S)$, avec $h=0,1$ pour une courbe de Lorenz tracée avec les points-déciles (si la courbe de Lorenz était tracée avec les points-centiles, on aurait $h=0,01$) ; il me semble un peu désolant de laisser les élèves lâchés pour ce type de calcul (c'est aussi l'occasion de leur dire qu'ils peuvent retrouver eux-mêmes l'aire d'un trapèze s'ils l'ont oubliée). On peut alors vérifier que les calculs par les trapèzes et en intégrant les fonctions ayant servi à l'ajustement des courbes donnent des résultats très voisins. Ce genre de calcul, en rendant les élèves actifs est bon et pour les maths et pour la compréhension de l'indice de Gini.

(10) **C.S**. Vous écrivez :

Remarque : d'autres graphiques peuvent être utilisés pour comparer des inégalités: type strobiloïdes (page 143 du livre « Sciences économiques et sociales terminale », ES BORDAS édition 2007) , chacun avec des avantages et des inconvénients et donc se complétant…

Dans ces graphiques on revient à des courbes représentant des répartitions de revenus (en inversant abscisses et ordonnées par rapport à ce qui est fait dans vos documents), avec une échelle particulière (la valeur de la médiane est par définition de cette échelle l'ordonnée 100) ; c'est astucieux, car cela permet de comparer des quantités diverses ou des courbes de pays différents.

E.B. Pour les strobiloïdes, cela m'a laissée perplexe quand je les ai vus dans leur manuel ; cette histoire de référence à la médiane qui n'est pas la même, mais je ne me suis pas vraiment plongée dedans ; les manuels de SES me laissent parfois perplexes : les manipulations des graphiques sont de plus en plus compliquées pour faire apparaître des conclusions adéquates. D'où l'importance des maths, pour essayer de rester critique.

Bilan d'E.BOYER

Le document proposé est fait pour les élèves et donc avec des choix pédagogiques discutables (voire très discutables). A chacun de modifier ou de compléter en vue de ce qu'il souhaite faire, ici, nous on présente ce qu'on a fait -et qui est déjà difficile pour des élèves de ES. La mise au point de la notion de déciles en maths et en SES a nécessité un réel travail de concertation, débroussaillage et mise au point pour chaque professeur impliqué, afin d'aider les élèves à créer des « images mentales » de référence pour la lecture des graphiques et tableaux, « images mentales » que nous avons expérimentées d'abord sur nous pour nous comprendre et dont la mise en évidence nous a pris du temps. Nous avons du faire face à des questions liées à des tournures des phrases différentes suivant les disciplines et faire des choix.

C'est un travail long qui nous a demandé du temps face aux élèves : 2 heures à deux voix et 2 fois 1heure pour la reprise de chaque professeur dans sa discipline et pour une conclusion qui nous parait indispensable dans notre démarche.
Nous aimerions faire une fiche d'autoévaluation, mais nous n'en avons toujours pas trouvé le temps !