ESD 2014 –08 : Prise de décision

1. Le sujet

A. L'exercice proposé au candidat

En France, la proportion d'individus de groupe sanguin O est de 43 %.
Une enquête statistique est menée pour déterminer si la proportion d'individus de groupe sanguin O dans une certaine région est identique à celle de la France.
On dispose d'un échantillon aléatoire de 200 résultats d'analyses de sang, réalisées dans des laboratoires de la région étudiée, sur lequel on a observé 120 individus de groupe O. Ce résultat nous amène-t-il à remettre en cause l'idée selon laquelle la proportion d'individus de groupe O dans cette région est identique à celle de la France ?

B. Les réponses de deux élèves

Élève 1
120 individus de groupe sanguin O sur 200 dans cette région, cela fait une fréquence de 60% ce qui est vraiment plus grand que 43%. Il est impossible que ce soit lié au seul hasard. Il semble certain que la proportion d'individus de groupe sanguin O est plus grande dans cette région qu'en France.

Élève 2
À l'aide d'un tableur, j'ai simulé le groupe sanguin de 200 individus avec la probabilité p = 0,43 que le groupe sanguin soit O. J'ai compté le nombre d'individus de groupe sanguin O et j'en ai trouvé 80 soit une fréquence de 0,40. J'ai appuyé un grand nombre de fois sur F9 et j'ai observé très rarement des fréquences supérieures à 0,50 et une seule fois une fréquence supérieure à 0,60. Cela paraît donc possible que les 60% de groupe O observés dans cette région soient du hasard mais cela me paraît peu probable. Je pense que la proportion d'individus de groupe O est supérieure dans cette région et qu'il faudrait étudier l'origine d'un tel écart par d'autres investigations.

C. Le travail à exposer devant le jury

1. Analysez la production de chaque élève en mettant en évidence ses compétences en probabilités.

2. Exposez une correction de l'exercice comme vous le feriez devant une classe de première S, en vous appuyant sur les productions d'élèves.

3. Proposez deux ou trois exercices sur le thème *prise de décision* à des niveaux de classe différents. On explicitera pour chacun d'eux l'objectif pédagogique.

2. Eléments de correction

1. Cet exercice présente une situation où l'on constate une « différence significative » par rapport à la valeur attendue d'une proportion. Il a précisément pour objectif de donner un sens à cette notion (dans quelle mesure peut-on ou non considérer l'écart constaté comme « significatif » ?
Il s'agit donc d'« Exploiter et faire une analyse critique d'un résultat d'échantillonnage », pour reprendre un intitulé des programmes de seconde.

Cet exercice peut être posé en seconde ou en première Les productions d'élèves ne nous éclairent pas sur le niveau de classe auquel il s'est réellement adressé, l'un des élèves paraît être un élève de seconde et l'autre de première. La correction de cet exercice sera assez différente suivant le niveau considéré.

L'élève 1 a une idée intuitive de la notion de « différence significative ». Il considère d'emblée l'écart « trop grand pour être dû au hasard », ce qui le dispense de tout traitement mathématique de la situation.
Aucune compétence dans le domaine des probabilités n'est apparente dans sa production (on note uniquement que cet élève sait calculer une fréquence …).

L'élève 2 propose un traitement mathématique pertinent à l'aide d'une simulation. Sa conclusion est argumentée, sans cependant que l'argument fourni soit un argument objectif (il s'agit d'une conjecture et non d'une preuve).
Cet élève met en œuvre des compétences ne relevant pas du seul domaine des probabilités :

- Elaborer et utiliser une simulation numérique prenant appui sur la modélisation et utilisant un logiciel (il modélise le choix d'un échantillon de 200 individus dont on relève le groupe sanguin).
- Effectuer des inférences pour obtenir de nouveaux résultats (il examine ce qu'il se passe lorsqu'on considère de nombreux échantillons).
- Emettre une conjecture (il raisonne à partir de la simulation qu'il a mise en place en considérant « possible mais peu probable » que l'écart soit du au hasard).

2. La production de l'élève 1 peut être exploitée pour lancer le débat, mais sans plus : « à partir de quelle valeur considères-tu qu'une fréquence est *vraiment* plus grande qu'une autre ? ». Mettre en évidence qu'un jugement subjectif ne permet pas de conclure, d'où la nécessité d'une norme qui soit acceptée par tous.

La production de l'élève 2, quant à elle, se prête à une exploitation plus intéressante :

- Cet élève a (parait-il, voir commentaire) obtenu une fois une fréquence plus grande que 0,6, ce qui prouve qu'il est « possible » que le résultat observé soit dû au hasard.
- L'élève a cependant rejeté cette hypothèse : pour quelle raison ?

Il s'agit de montrer aux élèves l'intérêt de quantifier le « nombre de fois qu'on appuie sur F9 » puis de se livrer à une étude statistique sur la répartition des fréquences observées. Mettre en place la notion de seuil (repérer l'intervalle dans lequel se trouvent 95 % des fréquences observées) et situer la fréquence de l'échantillon par rapport à ce seuil.

Une fois ce travail effectué, le professeur a le choix entre trois options.

- S'en tenir là, en faisant une synthèse du travail accompli : on a fait une simulation de l'expérience, que l'on répète un assez grand nombre de fois puis on a retenu l'intervalle excluant de part et d'autre les 2 ,5 % des valeurs les plus extrêmes. Faire remarquer l'inconvénient de cette méthode : l'intervalle obtenu dépend de l'expérience ; si on recommence dans les mêmes conditions, on ne trouvera pas exactement le même intervalle.
- Introduire l'intervalle de confiance au seuil 95 % tel qu'il est défini en classe de seconde. (on peut comparer avec la méthode précédente, en principe les résultats devraient être très similaires, et on gagne en temps et en simplicité d'application).
- Introduire l'intervalle de confiance au seuil 95 % issu de la loi binomiale *gilbertjulia* 2014 $B(200\ ;\ 0{,}43)$ conformément aux programmes de la classe de première.

3. Peut-être convient-il, en termes de prise de décision, de varier les options suivant le niveau de classe auquel on s'adresse. Etudier un cas où les méthodes de seconde ne peuvent pas s'appliquer.

Voir aussi l'exploitation d'une espérance mathématique comme aide à la décision (type joue-t-on ou ne soue-t-on pas).

Cependant, la notion de *prise de décision* est une notion difficile à maîtriser. Ne pas perdre de vue que tout test décisionnel, quel qu'il soit, donne une information neutre et objective. La décision finale nous appartient, et peut dépendre subjectivement des enjeux et des conséquences de la décision prise.

3. Commentaires

On note que l'auteur de l'exercice n'y est pas allé avec le dos de la cuiller dans la constitution de son échantillon. Un intervalle de confiance au seuil de 95 % que l'on peut retenir (type « programme de seconde ») est $[0{,}35\ ;\ 0{,}51]$. La valeur 0,6 en est très excentrée.

L'élève 1 risque de se dire « tout ça pour ça » et de prendre son professeur pour un enfonceur de portes ouvertes. Il n'aurait pas tout à fait tort …

Quant à l'élève 2 on peut s'interroger sur la chance inouïe qu'il a eue d'observer « *une fréquence* (égale ou) *supérieure à 0,60* » ne serait-ce que « *une seule fois* » même après un grand nombre d'appuis sur la touche F9. Un calcul montre en effet que : $\sum_{k=120}^{k=200} \binom{200}{k} 0{,}43^k . 0{,}57^{200-k} \underset{gilbertjulia\,2014}{=} 9{,}9\times10^{-7}$ à $0{,}1\times10^{-7}$ près. La probabilité d'obtenir au moins 0,60 est inférieure à 10^{-6} ...

On peut conjecturer qu'il s'agit d'un élève fictif et que le jury souhaitait voir comment les candidats exploitaient le fait que la fréquence 0,60 a été « réellement » (!) simulée.

Il me semble qu'il aurait été plus intéressant de proposer un échantillon constitué d'à peu près une centaine d'individus du groupe O, de façon que la prise de décision soit plus nuancée et que le traitement mathématique prenne tout son sens.