

## **Discrimination et statistique : analyse d'une argumentation statistique en terminale scientifique**

*Groupe « statistique et citoyenneté », IREM de Paris Nord.*

### **Objectifs**

#### **Méthodes**

Exploiter une situation réelle.  
Étude d'un texte judiciaire à contenu mathématique.

#### **Contenus**

Loi binomiale et fluctuations d'échantillonnage.

### **Durée**

Exercice « sur feuille » de 20 minutes (les copies sont corrigées mais non évaluées).  
Correction et discussion d'environ 20 minutes.

### **Organisation de l'activité**

Les élèves connaissent la loi binomiale.

- Exercice à traiter, à la fin d'un cours, sur une photocopie distribuée. Les élèves sont avertis que « cela ne sera pas noté ».
- Rendu des copies et correction, suivi d'une discussion, à la séance suivante.

## Séance 1 : recherche de l'exercice « sur feuille »

### Énoncé distribué aux élèves

Des arguments de type probabiliste peuvent être avancés et pris en compte dans les Cours de justice. Un accusé d'origine mexicaine, condamné pour vol et tentative de viol dans un comté du sud du Texas attaqua le jugement sous le motif que la désignation des jurés dans l'Etat du Texas était discriminatoire pour les Américains d'origine mexicaine. Son argument était que ceux-ci n'étaient pas suffisamment représentés dans les jurys populaires.

#### Attendu de la Cour Suprême des Etats-Unis (affaire Castaneda contre Partida)<sup>1</sup> :

« Si les jurés étaient tirés au hasard dans l'ensemble de la population, le nombre d'américains mexicains dans l'échantillon pourrait alors être modélisé par une **distribution binomiale**... Etant donné que **79,1 %** de la population est mexico-américaine, le nombre attendu d'américains mexicains parmi les **870** personnes convoquées en tant que grands jurés pendant la période de 11 ans est approximativement **688**. Le nombre observé est **339**. Bien sûr, dans n'importe quel tirage considéré, une certaine fluctuation par rapport au nombre attendu est prévisible. Le point essentiel cependant, est que le modèle statistique montre que les résultats d'un tirage au sort tombent vraisemblablement dans le voisinage de la valeur attendue... La mesure des fluctuations prévues par rapport à la valeur attendue est l'**écart type**, défini pour la distribution binomiale comme la racine carrée de la taille de l'échantillon (ici 870) multiplié par la probabilité de sélectionner un américain mexicain (ici 0,791) et par la probabilité de sélectionner un non américain mexicain (ici 0,209)... Ainsi, dans ce cas, l'écart type est approximativement de **12**. En règle générale pour de si grands échantillons, si la différence entre la valeur attendue et le nombre observé est plus grand que deux ou trois écarts types, alors l'hypothèse que le tirage du jury était au hasard serait suspect à un spécialiste des sciences humaines. Les données sur 11 années reflètent ici une différence d'environ **29** écarts types. Un calcul détaillé révèle qu'un éloignement aussi important de la valeur attendue se produirait avec moins d'**une chance sur  $10^{140}$** . »

#### Questions

Définir la variable aléatoire qui, dans cette situation, suit une loi binomiale.

.....  
 .....

Quels sont les paramètres de la loi binomiale ? .....

A quel calcul correspond la valeur 688 ? .....

Effectuer le calcul de l'écart type. ....

A quoi correspond la « différence de 29 écarts types » ? .....

.....  
 .....

A quel événement correspond la probabilité  $10^{-140}$  ? .....

.....  
 .....

<sup>1</sup> In : “Prove It with Figures (Statistics for Social Science and Behavioral Sciences)” - Hans Zeisel, D. H. et D. Kaye - Springer 2006.

.....  
 La constitution des jurys est-elle au hasard ? .....  
 .....

## Séance 2 : correction, compte-rendu des copies et discussion

### « Corrigé » distribué aux élèves

Définir la variable aléatoire qui, dans cette situation, suit une loi binomiale.

Il s'agit de la variable aléatoire  $X$  qui à 870 personnes tirées au sort dans la population (celle-ci étant très importante, on peut assimiler ces tirages à des tirages avec remise) associe le nombre de personnes d'origine mexicaine.

Quels sont les paramètres de la loi binomiale ?

$n = 870$  et  $p = 0,791$ .

A quel calcul correspond la valeur 688 ?

688 correspond au calcul de l'espérance :  $np \approx 688,17$ .

Effectuer le calcul de l'écart type.

On a  $\sigma = \sqrt{(870 \times 0,791 \times 0,209)} \approx 11,99 \approx 12$ .

A quoi correspond la « différence de 29 écarts types » ?

29 écarts types correspond à la différence entre la valeur observée de la variable aléatoire, c'est-à-dire 339, et l'espérance.

A quel évènement correspond la probabilité  $10^{-140}$  ?

Cela correspond à la probabilité d'observer une valeur inférieure d'au moins 29 écarts types de l'espérance, c'est-à-dire :  $P(X \leq 351)^2$

La constitution des jurys est-elle au hasard ?

La probabilité précédente est beaucoup trop faible pour considérer que la variable aléatoire binomiale introduite ici modélise correctement la situation.

### Analyse de quelques copies d'élèves

La copie suivante peut-être considérée comme une « bonne » copie.

<sup>2</sup> Plus précisément, pour une variable aléatoire  $X$  de loi binomiale de paramètres 870 et 0,791, on a :

$$P(X \leq 339) \approx 4,2 \cdot 10^{-145}.$$

Définir la variable aléatoire qui, dans cette situation, suit une loi binomiale.  
 La variable est le nombre de mexicain choisi parmi les ducs

Quels sont les paramètres de la loi binomiale ?  $B(770, 9,1\%)$

A quel calcul correspond la valeur 688 ?  $688 = E(X) = n \times p$

Effectuer le calcul de l'écart type.  $\sigma(X) = \sqrt{770 \times 0,09} = 11,99 \approx 12$

A quoi correspond la « différence de 29 écarts types » ?  $E(X) - 333 = 29 \sigma(X)$

A quel événement correspond la probabilité  $10^{-140}$  ? La probabilité que la valeur soit à un écart type de 29 écarts type.

La réponse suivante, bien que correcte, n'est pas argumentée, peut-être par manque de temps.

Définir la variable aléatoire qui, dans cette situation, suit une loi binomiale.  
 X suit la loi binomiale  $B(770, 0,091)$  car il s'agit d'expériences indépendantes qui peuvent être considérées comme indépendantes par rapport au nombre d'habitants de l'île avec 2 résultats possibles : résister (0,091) et mourir (0,909)

Quels sont les paramètres de la loi binomiale ?  $B(770, 0,091)$

A quel calcul correspond la valeur 688 ? L'espérance (moyenne)  $E(X) = 770 \times 0,091 = 688$

Effectuer le calcul de l'écart type.  $\sqrt{770 \times 0,091 \times (1 - 0,091)} = 11,99$

A quoi correspond la « différence de 29 écarts types » ?  $\frac{688 - 333}{12} = 29$

A quel événement correspond la probabilité  $10^{-140}$  ?  
 cela est une chance sur  $10^{140}$  qu'un tel événement ait lieu

Il est à noter que lors de la discussion qui a suivi la correction, est apparu le fait qu'une probabilité de  $10^{-140}$  ne semble pas immédiatement « humainement impossible » à un certain nombre d'élève. On peut songer à l'exemple des singes dactylographes de Borel. Il est en tout cas essentiel de signaler que face à une telle probabilité, il faut absolument chercher une autre explication possible.

### Bilan de l'activité

S'agissant d'une classe de terminale S à option « sciences de l'ingénieur », a priori peu « littéraire », le texte a été plutôt bien compris et l'aspect littéraire de l'exercice n'a pas vraiment posé de problème. En revanche c'est l'argumentation qui était souvent défaillante. Les élèves ont trop tendance à se contenter d'affirmations non justifiées, voire péremptoires : un tel exercice n'est donc pas superflu !

Les avis des élèves sur l'activité sont partagés. Plusieurs élèves ont trouvé ce travail « plus intéressant que ce que l'on fait d'habitude », parce qu'il s'agit de la « vraie vie ». Une élève a même affirmé que « s'il y avait eu plus souvent ce type d'exercice en

mathématiques, elle se serait peut-être davantage intéressée à la matière » (il s'agit d'une élève ayant des résultats généralement en dessous de la moyenne et travaillant peu). D'autres élèves cependant, ont trouvé ce travail « plus difficile » que les exercices classiques, et à la question « Que penseriez-vous de ce type d'exercice en maths au bac S (analyse d'un texte scientifique comme cela se fait en physique) ? », ils ont répondu qu'ils ne voulaient pas de ce type d'exercice à l'examen.

## Des prolongements

Les calculs précédents montrent qu'on ne peut pas considérer que les jurys résultent d'un tirage au sort où chaque élément de la population a les mêmes chances d'être choisi. Mais c'est tout ce que l'on peut dire et, en particulier, on n'a aucune information pour se prononcer sur les causes : on ne peut donc pas porter immédiatement des accusations de discrimination raciale.

L'étude statistique précédente doit inciter à enquêter plus sur les conditions de constitution des jurys.

On constatera :

- que pour être juré on doit maîtriser la langue anglaise (écrite et parlée) : ce qui n'est pas le cas d'une partie de la population d'origine hispanique :
- que pendant les 11 années correspondant à l'étude, la proportion des hispaniques dans la population a évolué,
- que la proportion d'hispaniques dans les jurys a également évolué au cours de ces 11 années.

La preuve statistique, telle qu'elle est apportée ici, est donc contestable au niveau du choix de la population de référence : celle-ci ne devrait pas être toute la population mais seulement la sous population de ceux qui ont plus de 18 ans, parlent lisent et écrivent l'anglais.

A titre de prolongement de cette activité, il pourrait être intéressant de faire rédiger aux élèves une contre argumentation statistique à partir de ces éléments.