Brochure n° 135
de la Commission inter-IREM
Lycées technologiques

Statistique & Citoyenneté

Le citoyen face au chiffre

UNIVERSITÉ PARIS 13

I.R.E.M. PARIS NORD
99 avenue J.B. Clément
93430 VILLETANEUSE

Brochure n° 135 de la Commission inter-IREM-Lycées technologiques

UNIVERSITÉ PARIS 13.- IREM.-
STATISTIQUE ET CITOYENNETÉ :
le citoyen face au chiffre.–
Jean-Paul Cardinal.- Aude Coupry.-
Françoise Delzongle.- Sébastien
Dherissard.- Philippe Dutarte.- Hatem
Maati.-
133 pages.- Villetaneuse 2007

Dépôt légal : 4ème trimestre 2007

Cette brochure a été réalisée dans le cadre des travaux du groupe « Statistique et citoyenneté » de l'I.R.E.M. de Paris-Nord :

Jean-Paul Cardinal – Directeur de l'IREM Paris-Nord,

Aude Coupry – lycée Gustave Eiffel à Cachan (94),

Françoise Delzongle – lycée Gustave Eiffel à Cachan (94) et IUFM de Créteil,

Sébastien Dherissard – lycée Gaston Bachelard à Chelles (77) ,

Philippe Dutarte – lycée Edouard Branly à Créteil (94),

Hatem Maati – lycée Gutenberg à Créteil (94).

Sommaire

Avant-propos

« Si la notion de *vérité statistique* devenait familière à tous ceux qui parlent ou écrivent au sujet de questions où la vérité statistique est la seule vérité, bien des sophismes et bien des paradoxes seraient évités. »

Emile Borel – *Le Hasard* – Alcan, 3e édition, 1914.

À un moment où se posent de façon aiguë les questions du rôle de l'enseignement des mathématiques, et pas seulement de la part de jeunes demandant « à quoi ça sert ? », de la définition d'une nécessaire « culture mathématique » et de la motivation des élèves, il nous semble pertinent de présenter des activités statistiques qui peuvent contribuer à une réponse. Comme on le constatera, on fait de « vraies maths » dans ces activités et on y acquiert des connaissances en informatique. Au delà du traitement statistique, des compétences de calcul, d'interprétation graphique, de raisonnement scientifique, inscrites dans les programmes officiels, sont sollicitées.

Le choix d'exemples réalistes, en prise avec le monde et les questions de société, motive indéniablement les élèves, en particulier les « non matheux », permettant de mettre en évidence les implications de notre enseignement. Ce n'est pas que les exemples ludiques, comme les jets de pièces et de dés, soient sans intérêt ; bien au contraire, ils constituent un outil pédagogique essentiel à l'étude de l'aléatoire, mais ils ne peuvent, à eux seuls, en justifier les enjeux. Les études statistiques développées ici suscitent des interrogations, en décelant des « tendances » qui, sans trancher sur la causalité des phénomènes, suggèrent des lieux où aller regarder « ce qui a pu se passer ». Elles débouchent donc sur un dialogue qui, dans la classe, entre élèves et professeurs de mathématiques ou d'autres disciplines, et avec l'appui de sources documentaires telles qu'Internet, peut être analogue à celui que doivent mener de vrais utilisateurs de la statistique.

Enfin les activités présentées s'appuient sur une pratique des mathématiques « par l'expérience », facilitée par les moyens informatiques. Les mathématiques ont un aspect pratique et expérimental, en particulier pour leur enseignement, pour conjecturer bien sûr, mais aussi pour comprendre et parfois pour emporter la conviction. Voir un théorème à l'œuvre peut être plus convainquant et plus éclairant que de lire sa démonstration, au moins dans un premier temps.

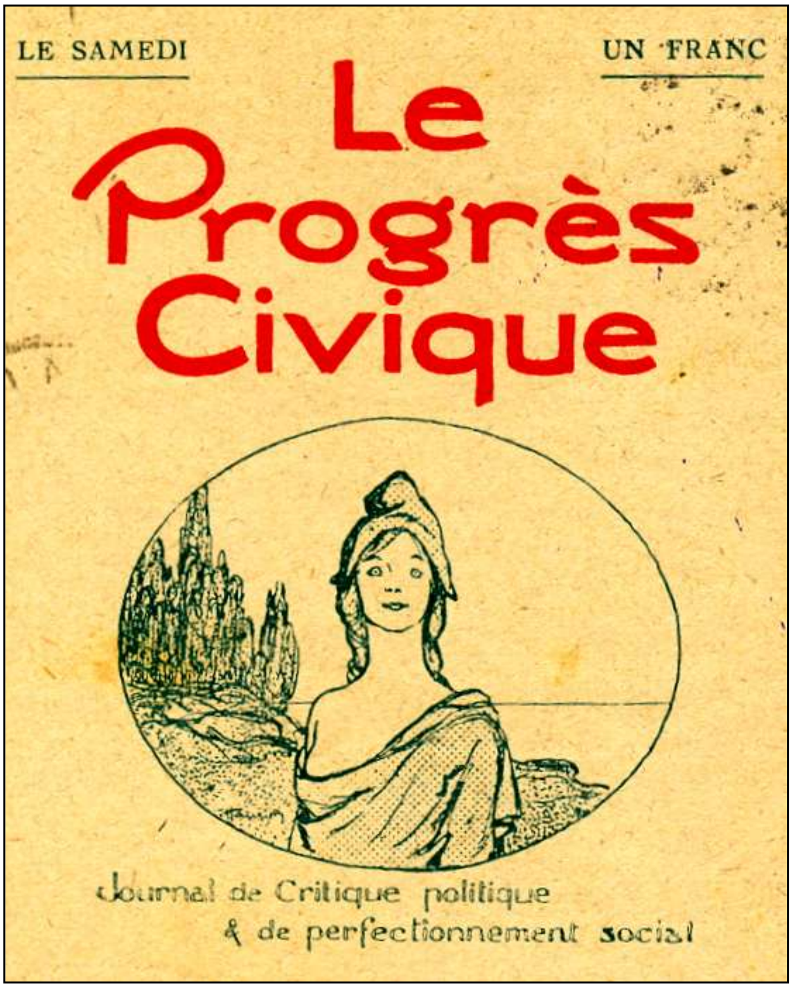
À partir de situations concrètes, ayant une forte résonance sociologique, les objectifs des activités développées dans cette brochure sont donc multiples :

– montrer l'utilité d'une formation mathématique pour « décrypter » le monde moderne ;
– développer l'intérêt pour les mathématiques par des activités ayant une signification forte et favorisant l'interdisciplinarité ;
– privilégier l'autonomie des élèves à mettre en place une démarche « scientifique » dans l'analyse d'une situation : formulation d'hypothèse, « construction » d'un « modèle », expérimentation (simulations…), conclusions ;
– initier à l'aléatoire et aux notions de risques.

Groupe « statistique et citoyenneté » de l'IREM Paris-Nord.

1

Initiation à l'aléatoire et simulation



Suggestion d'une séquence sur la simulation

La séquence proposée est constituée de 3 séances d'une heure minimum. Les deux premières séances nécessitent l'usage d'une calculatrice et peuvent être réalisées dans une banale salle de classe. La séance 3 est une séance d'initiation au tableur qui permet d'obtenir de grands échantillons nécessaire à la mise en œuvre des séances 4. Il est préférable d'utiliser les heures de modules mais elles peuvent être menées en classe entière (avec des groupes par exemple). La quatrième séance, qui utilise les capacités du tableur peut avoir des contextes très variés et donner lieu à un travail interdisciplinaire intéressant. On peut donc y revenir à différents moments dans l'année en demandant aux élèves de ré-activer leurs connaissances à partir des compte –rendus produits la première fois.

Ces Travaux Pratiques sont très riches. Ils peuvent être menés dés le début de l'année car ils permettent d'aborder des statistiques motivantes, qui donnent du sens à l'enseignement des mathématiques. Les statistiques descriptives font partie du programme de collège même si les élèves arrivent souvent sans avoir abordé ces notions et qu'il faudra y revenir plus tard.
Ce travail à partir de simulations obligent les élèves à entrer en réflexion approfondie très vite, ce qui est un enjeu majeur du lycée. Le travail personnel demandé (rédaction de compte rendus, récits de vécu en classe, petits exercice d'investissements seul devant sa calculatrice) a été imaginé après de longues années de pratique qui ont montré que la consigne usuelle de « relire le cours » est insuffisante pour se construire de réels savoirs avec de telles activités. L'élève doit rentrer une deuxième fois dans l'activité en produisant **un écrit** à partir de **sa prise de notes en classe** qui montrera qu'il a cherché à s'approprier ce que le professeur lui a fait découvrir lors de la séance de cours. Il va de soi que les productions attendues peuvent être très diverses et c'est surtout la qualité de l'investissement des élèves et le côté personnel du travail qui devra être évalué.
Les extraits de copies montrent la diversité des acquis des élèves à l'issue de la séquence, ce qui est tout à fait normal compte tenu des représentations initiales des élèves sur les sujets abordés.
Ils participent à la construction de nombreuses compétences mathématiques chez les élèves :

- Lire, donner du sens aux mots
- Elaborer et organiser une démarche en mobilisant des connaissances tant en statistiques que dans le domaine des nombres (nature d'un nombre, appartenance à un intervalle) et dans l'usage raisonné des outils de calcul dans l'esprit du programme de seconde.
- Exécuter , conjecturer, critiquer, justifier, conclure

Les élèves ont à produire des écrits de types différents :

- D'abord une prise de note en classe au cours de l'échange avec le groupe et le professeur, échanges oraux et écrits au tableau.
- Dans un deuxième temps **seul** , il produit un écrit de synthèse guidé, réalisé à partir du précédent qui permet de développer les compétences méthodologiques en mathématiques.

Le fait de le donner d'un cours sur l'autre permet de préciser la nécessité de faire un lien entre les cours qui n'est pas acquis au lycée. Il semble plus probable que cet écrit soit personnel puisqu'il ne ressemble pas au devoir maison qu'ils ont l'habitude d'avoir en mathématiques.
Un devoir maison de recherche avec des simulations à imaginer soi-même peut-être donné et il donne des résultats étonnants par leur originalité.

Objectifs de la séquence

- permettre aux élèves de s'approprier les méthodes de simulation d'un échantillon de taille n en utilisant l'outil calculatrice disponible en classe afin de préparer une séance au contenu plus riche avec mise en œuvre du tableur en salle informatique.
- observer la fluctuation des fréquences obtenues et conclure (le lien entre fréquence observée et probabilité qui est au programme de première et terminale est exploré dès la seconde dans l'esprit des programmes).
- Préparer l'étude de cas construits à partir de données réelles qui sera étudiée en salle informatique et permettre la découverte de quelques fonctionnalités très importantes du tableur.
- Développer des compétences méthodologiques sur le travail personnel.

Points particuliers concernant la mise en œuvre dans la classe :

Séance 1

Cette activité permet à l'élève d'utiliser une calculatrice scientifique telle qu'elle est préconisée au lycée en abordant d'emblée quelque chose de nouveau.
Elle permet aussi un travail sur la nature des nombres , les intervalles et la manipulations d'inégalités puisque RAND génère des décimaux dans [0,1[et INT permet d'obtenir l'entier immédiatement inférieur. Elle peut donc être menée en parallèle avec le chapitre de seconde sur les nombres.
Pour accompagner le travail des élèves en classe, un vidéo projecteur avec un émulateur de calculatrice sera très utile pour faire des synthèses et permettre à tous de progresser. L'élève aura des images mentales en tête qui devraient faciliter le travail maison entre deux séances.
La séance fait une place importante à l'oral et aux réactions du groupe en interaction avec l'enseignant. Les raisonnements sont notés sur le tableau et il faut attacher une place importante à la gestion de la prise de notes de élèves.
Dans le travail personnel demandé écrit explicitement sur le texte proposé aux élèves, la question 1 qui demande un **récit du déroulement de la séance** à pour objectif d'obliger les élèves à « évoquer » le travail de classe (cf travaux sur la gestion mentale de A De La Garanderie) à faire le lien entre activité de classe et activité seul. Des parents peuvent accompagner cette production en questionnant l'élève sur ce qui a eu lieu en classe, sans avoir de connaissances en mathématiques. La question 2 permet de voir ce que l'élève ressent comme **essentiel.** On donne souvent comme consigne de repérer l'essentiel pour apprendre sa leçon ce qui est très subjectif et doit être accompagné par le professeur. La lecture des copies d'élèves permet d'accéder immédiatement après le cours aux « souvenirs » des élèves et donner matière à des aides individualisées de début d'année qui pointent la nécessité d'être attentif , de participer etc..., attitudes fondamentales pour progresser en mathématiques. La question 3 a pour but de faire pointer à l'élève que le travail personnel est un moment où **on questionne** ce qu'on a vu en classe. Poser des questions nécessitent déjà une distance que les élèves n'auront peut-être pas en début de lycée, en effet ils pensent souvent qu'on attend d'eux des *réponses* et non des *questions*. La question 4 revient à une demande plus classique de faire une exercice d'application et elle permet de remettre l'élève en terrain connu. Sur ce point on ne pourra probablement pas éviter qu'ils copient sur le copain s'ils ne sont pas parvenus à faire seuls l'application mais on pourra travailler alors sur la manière de *« s'inspirer »* du travail de l'autre : copié-collé ou travail en groupe ?

Séance 2

Elle permet d'assurer la transition vers les activités plus complexes à données réelles. Dans l'exercice 1 l'introduction des lettes n_1, n_2 et N permet de reprendre les lettres en usages dans les formules de fréquences , de moyenne. Cela répond aussi au souci d'accompagner la prise de notes des élèves.
Le choix de l'étude de la proportion des daltoniens permet d'amener les élèves à distinguer des sous- populations (ici selon le sexe). Le caractère médical choisi (le daltonisme) ne présenta pas de risques de perturbations d'élèves qui se sentiraient touchés alors que lors de l'étude de la proportion de leucémies(WOBURN) , il faut être vigilant aux réactions de tous les élèves : lorsque quelqu'un de proche est atteint par une telle maladie on devient très vulnérable.

Points particuliers pour un bon déroulement de la séance

Exercice 1
Pour la question 2.c
Le professeur a préparé un transparent avec un tableau avec N lignes et n_1 colonnes qui sera complétée et commentée. Les élèves peuvent coder comme ils le souhaitent : D $\overline{D}$ ou 1 et 0 .
Il insiste sur l'observation de la fluctuation d'échantillonnage)
Pour la question 3
Le professeur pousse cette fois –ci à l'utilisation de INT(RAND +0,0045) pour tous les élèves.
Pour la question 4
Moyenne peu pertinente aux yeux du corps médical. Le professeur veillera à faire prendre des notes aux élèves sur ce point

Exercice 2
On rappelle que la probabilité qu'un échantillon avec remise de taille 20 contienne un nombre de femmes est inférieur où égale à 3 peut-être calculée avec la loi binomiale.
Les élèves doivent faire preuve de plus d'autonomie. Le parti pris de ne pas détailler plus les questions est volontaire. Les élèves doivent **transférer** la démarche proposée dans le premier exercice pour résoudre ce nouveau problème . Pour que les élèves abordent sans trop de difficultés la séance sur tableur il est indispensable de leur faire prendre en notes les indications qu'on leur donne à l'oral pour avancer et de les obliger à travailler cela à la maison. L'usage de l'outil informatique nécessite un travail de réflexion préalable que les élèves doivent appréhender.
Dans ce but , la fin de l'activité organise clairement le travail à la maison. Il est à noter que les collègues de physique savent depuis longtemps qu'un compte-rendu de TP individuel est indispensable pour que la séance de TP ne soit pas seulement un moment où on fait des manipulations.

Etude à la maison :
Le but de ces travaux écrits inspirés de ce qui est demandé par les collègues de sciences expérimentales après un TP a pour but d'obliger l'élève à faire des liens entre deux séances en classe. Ils peuvent être évalués de manière positive (sans barème au quart de point bien sûr !) lorsqu'on a le sentiment que le travail fait a été sérieux.
Il s'agit de faire comprendre aux élèves qu'en cours de sciences on est un acteur attentif et on essaie de retenir ce qui a été étudié. Il permet de valoriser l'effort d'étude personnelle.

Il permet de plus d'accéder un peu à ce qu'a construit l'élève lors de la séance. Le document rédigé peut prendre place dans le cahier de statistiques à la place des habituelles feuilles volantes que les élèves sont vite incapables de situer. On peut ainsi envisager de revenir sur ces simulations plusieurs fois dans l'année et mesurer mieux les progrès des élèves.
On a précisé aux élèves le travail demandé et les critères qui interviendront dans son appréciation afin de les aider à cerner un travail qui peut sembler très nouveau à un certain nombre.

séance 3

Elle permet d'aborder le travail sur les grands échantillons dont on aura besoin par la suite pour les autres activités de la brochure. C'est une séance d'initiation au tableur. On réinvestit les savoir et savoir –faire des 2 séances précédentes en substituant la fonction ALEA() à la fonction RAND et la fonction ENT à la fonction INT. On peut utiliser l'assistant graphique du tableur pour visualiser la fluctuation d'échantillonnage.

séance 4

On réinvestit les savoirs et savoir-faire développés lors des séances précédentes
On peut utiliser les activités :
– Inquiétudes à Woburn.
– L'affaire Castaneda.
– Présidentielles 2002.
etc...
On pourra en fin de séquence donner comme devoir maison :
– Pollution et sex ratio.

La encore la qualité du travail de réflexion , de rédaction de l'élève doit être prépondérante si on évalue de manière chiffrée afin que les élèves prennent en compte les remarques sur la copies comme autant de manière de progresser. On peut « remettre sur le métier » plusieurs fois dans l'année ces notions difficiles en faisant les activités 3 à différents moments de l'année. Ils ont alors à réviser seuls le contenu des deux premières séances. C'est là qu'ils peuvent mesurer l'intérêt d'un cahier de statistiques bien organisé. Lorsque les élèves ont revu seuls leurs activités, leur proposer de les refaire en autonomie en classe est un bon moyen d'évaluer le degré d'assimilation des notions.

Simuler des expériences aléatoires avec une calculatrice

Niveau

Seconde.

Situation étudiée

Différentes selon les séances :
Séance 1 : Jeu de pile ou face , tirages de boule dans une urne avec des proportions p données de boules blanches et noires
Séance 2 : Etude des proportions de daltoniens dans une population, parité homme femme dans les mairies d' arrondissements de PARIS
Séance 3 : Etude de la simulation d'un grand nombre de tirages dans une urne.

Type d'activité

Séquence comportant des exercices en modules , avec calculatrice et tableur, assortis d'un travail à la maison préalable à la première séance et entre les séances.

Durée

Séquence de 3 Séances sur semaines consécutives de 55 mn.

Objectifs

Initiation a l'aléatoire . Familiarisation avec les simulations.

Contenus mathématiques au programme

- Travail sur la nature des nombres.
- Ordre des nombres.
- Calcul de fréquences.
- Simulations avec calculatrice et tableur.
- Notion de fluctuation d'échantillonnage.
- Représentation graphique avec un tableur.

Compétences et attitudes

- Concevoir et mettre en œuvre des simulations simples.
- Interpréter un résultat donné par une calculatrice.
- Caractériser les éléments d'un intervalle.

Enjeux citoyens

Comparer les proportions d'individus présentant une caractéristique donnée dans un échantillon et dans une population pour éviter les inférences abusives sur le thème de la santé et sur le thème de la parité et pour s'interroger.

Organisation

Cette séquence peut être mise en œuvre des le début de l'année même lorsque le chapitre statistique descriptive de seconde n'a pas été traité car elle ne nécessite que des notions de collège.

Description des activités

Séance 1

Travail préalable à la maison

1) Prendre une pièce de monnaie .Effectuer une série de 10 lancers et noter les résultats obtenus dans le tableau ci-dessous :

N°lancé	1	2	3	4	5	6	7	8	9	10
Pile ou face										

2) Rechercher dans le mode d'emploi de votre calculatrice ou dans votre livre comment obtenir la fonction RANDOM (notée rand ou rn# *).*

Activité en classe

Le but de cette activité est de simuler certaines expériences aléatoires simples à l'aide de la calculatrice :

1) Fonction random de la calculatrice :

Utiliser à plusieurs reprises la fonction random de la calculatrice.
Qu'observe-t-on ?
La fonction random (rand ou rn#) de la calculatrice génère un nombre au hasard dans l'intervalle [0 ;1[.

2) Exemple de simulation d'un jeu de pile ou face :

On lance une pièce de monnaie équilibrée. Les résultats possibles pour cette expérience sont pile ou face. Coller le tableau donnant les résultats obtenus lors d'une série de 10 lancers.

a) Modélisation : comment la calculatrice remplace la pièce de monnaie.

On simule un lancé de la manière suivante :
en utilisant la fonction random si la calculatrice affiche un nombre de l'intervalle [0 ;0,5[on considère que l'on a obtenu face, si au contraire la calculatrice affiche un nombre de l'intervalle [0,5 ;1[on considère que l'on a obtenu pile.

b) Simulation de 10 lancers :

En utilisant la méthode ci-dessus simuler une série de 10 lancers et inscrire les résultats obtenus dans le tableau suivant :

N°lancé	1	2	3	4	5	6	7	8	9	10
Pile ou face										

c) Amélioration du modèle, fonction Int de la calculatrice :

Pour une lecture plus rapide on désire que la calculatrice affiche 0 si l'on obtient face et 1 si l'on obtient pile.
Pour cela on utilise la fonction partie entière (notée Int) de la calculatrice. (Rappeler comment l'obtenir sur la calculatrice)

- Effectuer à la calculatrice Int (6,8), Int (0,25), Int (π). Quelle est la nature du résultat obtenu ?

<u>définition</u> : la partie entière d'un nombre est le plus grand entier inférieur ou égal à ce nombre.

- Justifier que si l'on effectue $2\times rand$ à la calculatrice on obtient un nombre de l'intervalle[0 ;2[.
  Quels résultats peut-on obtenir si l'on effectue Int ($2\times rand$) ?
- Lorsque la calculatrice affiche 0, à quel intervalle appartient le nombre obtenu avec la fonction random ?
  Dans quelle situation du jeu sommes-nous ?
- Lorsque la calculatrice affiche 1, à quel intervalle appartient le nombre obtenu avec la fonction random ?
  Dans quelle situation du jeu sommes-nous ?

3) <u>Exemple de simulation d'un tirage dans une urne :</u>

   a) <u>Description du jeu :</u>

Une urne contient 3 boules noires et une boule blanche.
On choisit au hasard une boule dans l'urne.

   b) <u>Modélisation :</u>

On simule un tirage de la manière suivante :
si en utilisant la fonction random la calculatrice affiche un nombre de l'intervalle
[0 ; [on considère que l'on a tiré une boule noire, si au contraire la calculatrice
affiche un nombre de l'intervalle [;1[on considère que l'on a obtenu la blanche.

   c) <u>Amélioration du modèle :</u>

- Quels résultats peut-on obtenir en effectuant Int (rand + 0,25) ?
- Lorsque la calculatrice affiche 0, à quel intervalle appartient le nombre obtenu avec la fonction random ?
  Dans quel situation du jeu sommes-nous ?
- Lorsque la calculatrice affiche 1, à quel intervalle appartient le nombre obtenu avec la fonction random ?
  Dans quel situation du jeu sommes-nous ?

   d) <u>Simulation de dix tirages :</u>

   e) <u>Un autre cas :</u>

L'urne contient maintenant 7 boules noires et trois boules blanches.
Proposer une modélisation puis simuler une série de 10 tirages.

Travail personnel à la maison :

1) Raconter, en 10 lignes maximum, la séance 1.
2) Quels points vous semblent essentiels à retenir en mathématiques ?
3) Y a-t-il des questions que vous n'avez pas posées lors de la séance en classe ?
4) On cherche à simuler avec une calculatrice le tirage d'une boule dans une urne contenant 6 boules noires et quatre boules blanches.
   Noter les résultats obtenus dans le tableau ci-dessous :

N°tirage										
résultat										

On décrira avec soin la démarche pour le premier tirage.

Séance 2

Exercice 1

En France, le daltonisme touche environ 8% des hommes et 0,45 % des femmes.

1) Soit n_1 le nombre de garçons dans la classe , n_2 le nombre de filles dans la classe , N l'effectif de la classe.
n_1=
n_2=
N=
En supposant que cette proportion soit vérifiée combien y aurait-il de daltoniens dans la classe ? Est-ce le cas ?

2) On considère l'expérience aléatoire qui consiste à choisir au hasard n_1 hommes dans la population française et à comptabiliser le nombre de daltoniens.
a) Comment peut-on simuler cette expérience à l'aide de la calculatrice ? (on décrira par écrit la démarche)
b) Noter les résultats ainsi obtenus dans un tableau
c) Mettre en commun les résultats obtenus par tous les élèves ainsi que les différentes manières d'y parvenir
3) On considère l'expérience aléatoire qui consiste à choisir au hasard n_2 femmes dans la population française et à comptabiliser le nombre de daltoniennes.
Reprendre les questions du 2).
4) Si l'on ne distingue plus hommes et femmes on obtiendrait un taux d'environ 4, 2 % de daltoniens en France.
Ce chiffre vous semble t-il pertinent ?

Exercice 2 :

Paris compte 20 mairies d'arrondissement dont 17 sont dirigées par des hommes et 3 par des femmes.
Cette répartition ne paraît pas équitable, mais peut cependant être due au phénomène de fluctuation d'échantillonnage.
Le but de cette activité est d'étudier, grâce à des simulations, la fréquence d'apparition d'une

telle répartition dans le cas où l'équité serait respectée.

1) On considère l'expérience aléatoire qui consiste à choisir au hasard 20 personnes dans la population française et à comptabiliser le nombre de femmes obtenu.
   a) Comment peut-on simuler cette expérience à l'aide de la calculatrice ? (on détaillera la démarche avec soin)
   b) Noter les résultats ainsi obtenus.
   c) Mettre en commun les résultats obtenus par tous les élèves.
   Combien d'élèves ont obtenu un nombre de femmes inférieur ou égal à 3 ?
2) Quelle conjecture peut-on émettre suite à cette étude
3) Peut-on conclure avec certitude ?

Etude à la maison :

a) Compte rendu de l'activité en classe :

Ce travail d'écriture vise à vous aider à vous approprier le contenu du cours de mathématiques étudié en classe. Il vous sera très utile pour réussir le travail sur un ordinateur. Vous pouvez vous aider des indications ci-dessous pour réaliser le travail.

Critères d'appréciation de la qualité du travail :
Le sujet: « Simulations d'un échantillon sur une calculatrice »

- Qualité de la présentation et de l'écriture
- Le contexte de travail en classe est rapidement décrit
- Le sujet, les questions sont reformulées sur la feuille, afin que le compte rendu puisse être lu sans avoir le texte polycopié à portée de la main
- L'essentiel des démarches mathématiques explicitées à l'oral en classe ou partiellement écrites sur le tableau sont consignées dans le document
- Les points importants du cours de maths à retenir, pointés en classe, sont explicitement sur la feuille

b) Exercice d'application en autonomie

Une urne contient 3 boules bleues, 5 boules vertes et 2 jaunes.
On tire une boule au hasard. On veut simuler ce tirage sur une calculatrice .

a) Si on tape au clavier $10\times$RAND , à quel intervalle appartient le nombre obtenu ? Justifier
b) Si on tape au clavier INT($10\times$RAND), quelle est la nature du nombre obtenu ? Quels sont les résultats possibles
c) Faire une série de simulations de 10 tirages de 1 boule dans l'urne .

On précisera sur la feuille le **codage** utilisé. Inscrire les résultats trouvés dans le tableau suivant :

N°tirage										
résultat										

d) Calculer la fréquence d'apparition des boules bleues, des boules vertes , des boules jaunes lors de 10 tirages successifs.
e) Refaire ainsi 4 autres simulations de 10 tirages.

Simulation d'expériences aléatoires sur tableur

On s'appuie sur les savoir-faire développés lors des séances précédentes.
Le but de cette activité est de simuler certaines **expériences aléatoires simples** à l'aide du tableur.
On se propose de simuler **un grand nombre de tirages**.

Fonction Alea du tableur

Activité 1

Réflexion préalable

Situation de jeu étudiée :

Une urne contient 4 boules noires et 1 boule blanche.
On choisit au hasard une boule dans l'urne. On remet la boule dans l'urne et on procède à 100 tirages successifs.
Comment, intuitivement, évalueriez vous les chances d'obtenir une boule noire ?une boule blanche ?
....

Simulation du jeu

Le professeur montrera les manipulations sur le logiciel avec l'ordinateur et le vidéo projecteur. Soyez attentifs !

On cherche à simuler 100 échantillons de 100 tirages
Pour faciliter le repérage des colonnes on crée une ligne contenant les 100 premiers entiers. Afin de ne pas les taper tous , suivre la procédure suivante, après avoir observé comment procède le professeur au vidéo projecteur.

1) Entrer 1 en A1 , 2 en B1 et « faire glisser » jusqu'à la colonne CV.
2) Sachant que sur Excel la fonction RAND correspond à ALEA() et la fonction INT à ENT proposer la formule à entrer en A1 pour simuler le tirage d'une boule dans cette urne.
....
....
....
....
Taper cette formule dans (attention à la syntaxe !).
3) « Faire glisser » pour copier la formule de manière à obtenir un tableau avec 100 lignes et 100 colonnes(observer le professeur et refaire).
4) Taper dans la cellule A104 :
=SOMME(A2:A101)/100
Qu'a-t-on ainsi calculé ?
....
....
....
....
5) « Faire glisser » la formule jusqu'à la colonne CV.

On veut construire les points de coordonnées (numéro de l'échantillon ; fréquence de l'échantillon) .
Sélectionner les cellules de la ligne 104, de A104 à CV104, puis cliquer sur l'icône graphique et pour le choix du graphique suivre la démarche mise en œuvre par le professeur au vidéo-projecteur.
6) Combien avez-vous d'échantillons dont la fréquence n'appartient pas à l'intervalle [0,2 ;0,3] ? Tous les élèves ont-ils le même résultat ? Pourquoi ?
....
....
....
....
....
....
....
7) Enregistrez votre travail en suivant les consignes du professeur.

Activité 2 en autonomie

Reprendre les questions de l'activité 1 dans le cas d'une qui contiendrait 3 boules blanches et 5 noires. Rendre par écrit cette activité en explicitant le plus possible les démarches utilisées.

Q.C.M.

Une seule proposition est vraie sur chaque ligne.
Entourer la lettre de la proposition vraie.

Une urne contient 10 boules de 4 couleurs différentes : 2 boules rouges , 3 boules vertes, 3 boules jaunes, 2 boules bleues .

1) On effectue 10 tirages successifs d'une boule. Les fréquences d'apparition des boules bleues et rouges sur cet échantillon :

A. sont toujours toutes les deux égales à $\frac{2}{10}$;

B. sont toujours différentes
C. sont parfois égales

On se propose de simuler un tirage d'une boule dans l'urne à l'aide d'une calculatrice.
On utilise la fonction « Random » qui affiche un nombre décimal de l'intervalle [0 ;1[.

2) On lit le résultat affiché en appuyant une fois directement sur la touche Random

A. Si le résultat appartient à [0 ;0,25[on considère que la boule tirée est rouge, s'il appartient à [0.25 ;0,5[on considère que la boule tirée est verte, s'il appartient à [0.5 ;0,75[on considère que la boule tirée est jaune, s'il appartient à [0.75 ;1[on considère que la boule tirée est bleue.

B. Si le résultat appartient à [0 ;0,2[on considère que la boule tirée est rouge, s'il appartient à [0.2 ;0,5[on considère que la boule tirée est verte, s'il appartient à [0 ,5. ;0,8[on considère que la boule tirée est jaune, s'il appartient à [0.8 ;1[on considère que la boule tirée est bleue.

C. Si le résultat appartient à [1;1,2[on considère que la boule tirée est rouge, s'il appartient à [1.2 ;1,5[on considère que la boule tirée est verte, s'il appartient à [1.5 ;1,8[on considère que la boule tirée est jaune, s'il appartient à [1.8 ;2[on considère que la boule tirée est bleue.

3) Voici l'écran d'une calculatrice d'un élève ayant appuyé 3 fois sur la touche Random :
0,0887208491
0,5042087903
0.4996094888
A. dans cette situation on a tiré une boule bleue, puis une jaune, puis une verte ;
B. dans cette situation on a tiré une boule un rouge, puis une jaune, puis une jaune ;
C. dans cette situation on a tiré une boule rouge, puis une jaune, puis une verte.

On joue avec la règle suivante : si la boule est bleue on gagne sinon on perd.

4) On rappelle que la touche INT permet d'obtenir le plus grand entier inférieur ou égal à un nombre. On veut que la calculatrice affiche 1 si on gagne et 0 si on perd, pour cela on peut utiliser la formule :
A. INT (RAND+0,2)
B. INT (RAND+0,8)
C. INT (RAND+0,5) .

On utilise un ordinateur et la fonction « ALEA » qui permet de simuler 300000 tirages.

5) On calcule avec le tableur la fréquence d'apparition des boules bleues :
A. La fréquence d'apparition des boules bleues est forcément égale à 0,2 .
B. La fréquence d'apparition des boules bleues devrait être proche de 0,2 .
C. Je suis sur que la fréquence d'apparition de la boule bleue est supérieure ou égale à 0,2.

2

Santé et environnement

Tabac et risques d'infarctus

Niveau

Énoncé n° 1 : Troisième, seconde ou 1ère ST2S (information chiffrée).
Énoncé n° 2 : Terminales STG, ST2S, ES ou S.

Situation étudiée

Une campagne de la Fédération française de cardiologie donne l'information suivante : « *80% des victimes d'infarctus avant 45 ans sont fumeurs* ».
En quoi cette « donnée » indique-telle que fumer augmente le « risque » d'infarctus ? Comment peut-on calculer cette augmentation du risque ?

Type d'activité

Exercice.

Durée

55 minutes.

Objectifs

Contenus mathématiques au programme

Énoncé n° 1 : Proportions, calcul littéral.
Énoncé n° 2 : Probabilités conditionnelles.

Enjeux citoyens

Notion de « risque » sanitaire.
Montrer que des connaissances mathématiques (proportions ou probabilités conditionnelles) permettent de mieux comprendre des statistiques médicales.

Capacités et attitudes

Interprétation des calculs et des résultats.
Réflexion sur un sujet de société.

Organisation

Exercice en classe à faire rédiger soigneusement ensuite, ou en devoir à la maison, en groupes par exemple.

Description des activités

Exercice en termes de proportions (3ème – 2nde – 1ère ST2S)

Énoncé élève (niveau 3ème, seconde ou 1ère ST2S)

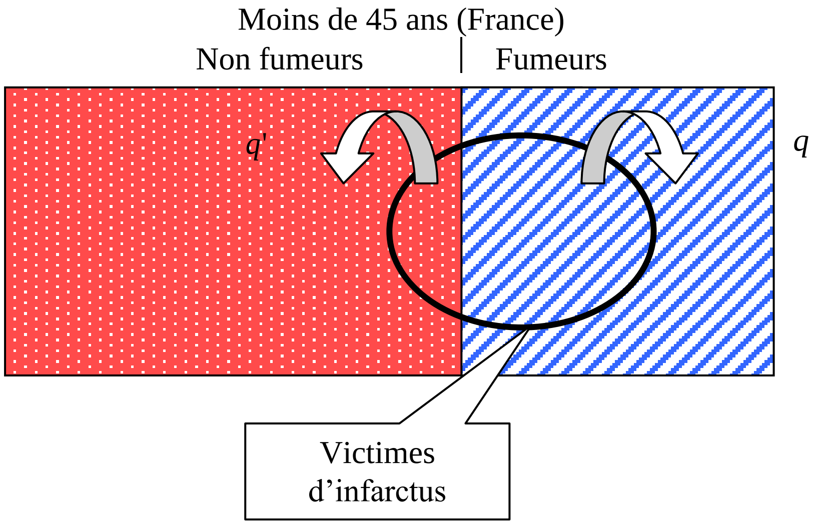
La Fédération française de cardiologie affiche l'information ci-dessous dans une campagne de presse.

1. Parmi les victimes d'infarctus ayant moins de 45 ans, quelle est la proportion de non-fumeurs ?
2. Pourquoi l'information donnée permet-elle de penser que fumer augmente le risque d'infarctus ?
3. On peut estimer qu'en France, parmi les moins de 45 ans, il y a environ 40% de fumeurs (ou d'anciens fumeurs).
a) On note i le nombre de cas d'infarctus observés chez les moins de 45 ans.
Exprimer en fonction de i le nombre d'infarctus parmi les fumeurs.
b) On note n le nombre de personnes de moins de 45 ans.
Montrer que proportion q d'infarctus parmi les fumeurs est $q = \dfrac{0{,}80 \times i}{0{,}40 \times n}$.
4. Donner, de même, l'expression de la proportion q' d'infarctus parmi les non-fumeurs.
5. Montrer que $\dfrac{q}{q'} = 6$.

On peut interpréter ce résultat en disant que pour les moins de 45 ans, un fumeur a 6 fois plus de risques d'avoir un infarctus qu'un non-fumeur.

Sources : Fédération française de cardiologie – www.tabac.gouv.fr – INSEE.
Pour la proportion des fumeurs en France, les estimations sont variables : aux alentours de 34% de la population. Le site tabac.gouv.fr affirme, à partir des données de l'INSEE, que la proportion de fumeurs dans la classe d'âge 18-25 ans était de 48% en 2000, 40% en 2003 et est remontée en 2005-2006.

Éléments de réponse (niveau troisième, seconde ou 1ère ST2S)



1. 20%.

2. On sait qu'il y a moins de 80% de fumeurs. On en déduit que les fumeurs sont sur-représentés parmi les cas d'infarctus.

3. $q = \dfrac{0{,}8 \times i}{0{,}4 \times n}$.

4. $q' = \dfrac{0{,}2 \times i}{0{,}6 \times n}$.

5. $\dfrac{q}{q'} = \dfrac{0{,}8 \times 0{,}6}{0{,}4 \times 0{,}2} = 6$.

Commentaires

Même si les outils mathématiques sont relativement élémentaires (proportions et calcul littéral – on peut s'aider d'un dessin), le résultat obtenu n'est pas évident. Sans ces outils, on ne peut comprendre que 80% et 40% conduisent à un risque multiplié par 6.

Les calculs de proportions menés ici ressemblent, dans leur esprit, à ceux réellement menés en épidémiologie (notion d'odds ratio). Même s'il s'agit d'une simplification des modèles utilisés, cet exercice n'en est pas une trahison.

Remarquons que le « risque » d'infarctus lui-même, qui correspondrait à q (pour un fumeur) et à q' (pour un non fumeur), n'a pas été calculé et est sans doute assez faible (on peut penser que la population n est grande par rapport au nombre d'infarctus i). Pour un risque relativement faible, le message « risque multiplié par 6 » est sans doute plus frappant. Par ailleurs si le risque individuel est relativement faible (et cela se discute car il s'agit d'un « risque évitable » que l'on peut ajouter à d'autres comme la trop grande vitesse sur les routes), il n'en est pas de même globalement pour la société (au niveau santé publique).

Exercice en termes de probabilités conditionnelles (terminales STG-ST2S-ES-S)

Énoncé élève (niveau terminale)

La Fédération française de cardiologie affiche l'information ci-dessous dans une campagne de presse.

1. Pourquoi l'information donnée permet-elle de penser que fumer augmente le risque d'infarctus ?

2. On prélève au hasard une personne dans la population des moins de 45 ans.
On note I l'événement « la personne prélevée a été victime d'un infarctus » et $\bar{I}$ l'événement contraire.
On note F l'événement « la personne prélevée est fumeur » et $\bar{F}$ l'événement contraire.
a) Quelle est des deux égalités suivantes, celle qui traduit le fait que « 80% des victimes d'infarctus avant 45 ans sont des fumeurs » $P_F(I) = 0{,}8$ ou $P_I(F) = 0{,}8$?
b) Calculer $P_I(\bar{F})$.

3. On peut estimer qu'en France, parmi les moins de 45 ans, il y a environ 40% de fumeurs. On a donc $P(F) = 0{,}4$.
On s'intéresse à $P_F(I)$ que l'on peut interpréter comme le risque d'infarctus pour un fumeur.
a) Montrer que $P_F(I) = \dfrac{P_I(F) \times P(I)}{P(F)}$.
b) En déduire que $P_F(I) = 2 \times P(I)$.

4. Montrer que $P_{\bar{F}}(I) = \dfrac{1}{3} P(I)$.

5. En déduire que $P_F(I) = 6 \times P_{\bar{F}}(I)$.
Interpréter le résultat obtenu.

Éléments de réponse (niveau terminale)

1. On sait qu'il y a moins de 80% de fumeurs. On en déduit que les fumeurs sont sur-représentés parmi les cas d'infarctus.
2. a) On a $P_I(F) = 0{,}8$.
b) On a $P_I(\overline{F}) = 1 - P_I(F) = 0{,}2$.
3. a) On a $P_F(I) = \dfrac{P(F \cap I)}{P(F)} = \dfrac{P_I(F) \times P(I)}{P(F)}$.
b) On en déduit que $P_F(I) = \dfrac{0{,}8}{0{,}4} \times P(I) = 2 \times P(I)$.
4. De même $P_{\overline{F}}(I) = \dfrac{P_I(\overline{F}) \times P(I)}{P(\overline{F})} = \dfrac{0{,}2}{0{,}6} \times P(I) = \dfrac{1}{3}\, P(I)$.
5. On peut interpréter ce résultat en disant que pour les moins de 45 ans, un fumeur a 6 fois plus de risques d'avoir un infarctus qu'un non-fumeur.

Commentaires

Si l'on compare à la version « proportions », on constate l'efficacité du calcul des probabilités conditionnelles (même si le formalisme est compliqué) évitant l'introduction d'effectifs inconnus.
Par ailleurs, les notions de risques étudiées ici se comprennent sans doute mieux en termes de probabilités, ce à quoi elles correspondent.

Inquiétudes à Woburn

Niveau

Seconde.

Situation étudiée

> Dans la petite ville américaine de Woburn la population s'interroge : elle a connu 3 cas de leucémies chez des jeunes filles et 9 chez de jeunes garçons en 10 ans. Doit-on en accuser le hasard ?

Type d'activité

Séance de « module » (demi groupes) utilisant le tableur.

Durée

1 heure.

Objectifs

Contenus mathématiques au programme

- Simulation avec un tableur.
- Notion de fluctuation d'échantillonnage (influence de la taille de l'échantillon).
- Travail sur les intervalles.

Enjeux citoyens

Etudier la démarche citoyenne de la population de Woburn, qui a permis de détecter qu'un problème de pollution était à l'origine de cas de leucémies infantiles.

Compétences et attitudes

Utiliser le tableur pour modéliser une situation concrète.
Réinvestir dans un cadre différent le travail des séances précédentes pour obtenir la formule =ENT(ALEA()+ p).
Interpréter ses résultats.
Mobiliser ses connaissances pour dégager la conclusion citoyenne de l'activité.

Organisation

Le chapitre de statistique descriptive a déjà été traité. Cette activité s'insère dans la séquence sur la simulation. Il s'agit d'une séance qui nécessite un travail préalable sur l'utilisation des fonctions random /alea et partie entière de la calculatrice et du tableur.

Description des activités

Séance de TP sur tableur

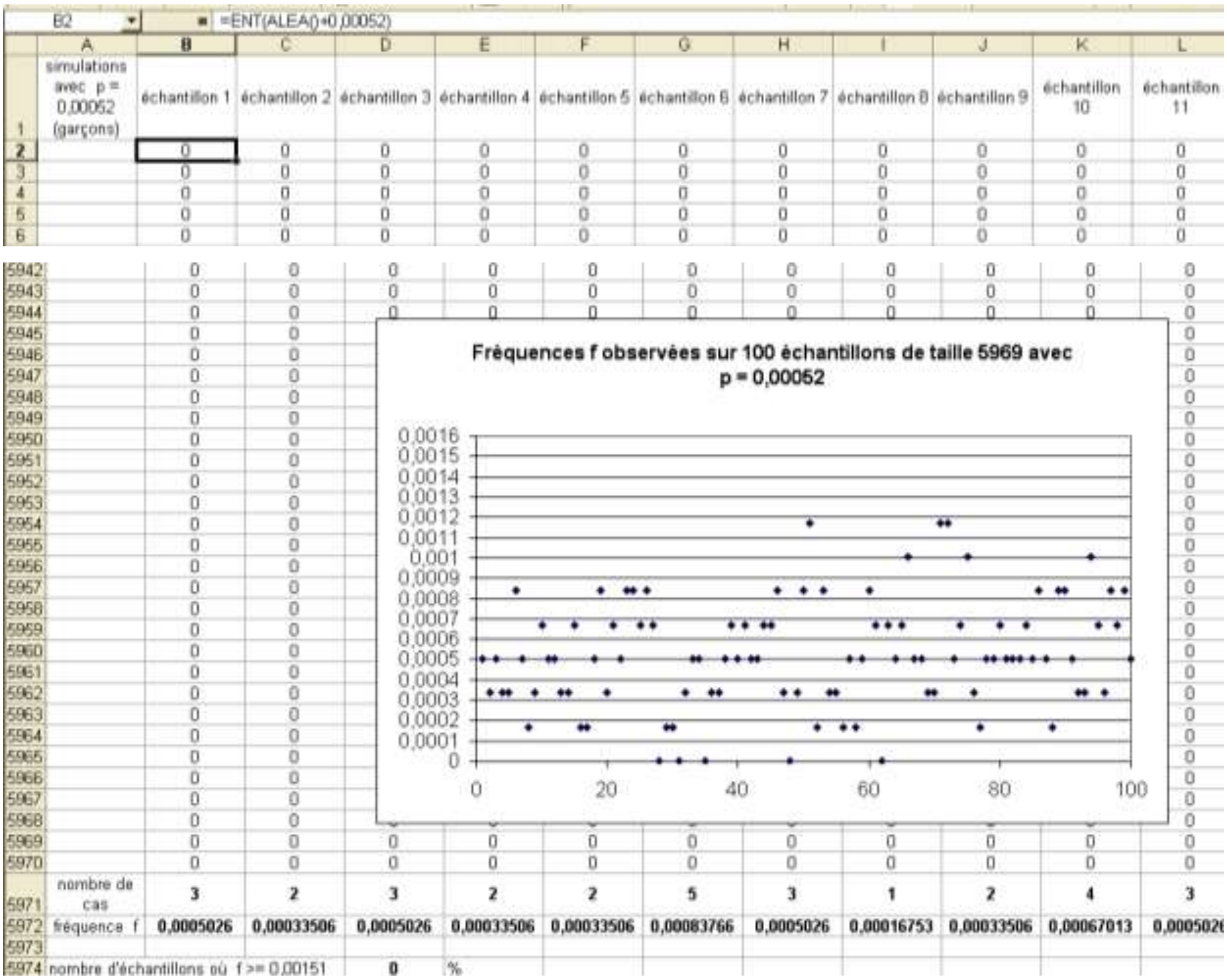
Introduction pour le professeur

Woburn est une petite ville industrielle du Massachusetts, au Nord-Est des Etats-Unis. Du milieu à la fin des années 1970, la communauté locale s'émeut d'un grand nombre de leucémies infantiles survenant dans certains quartiers de la ville. Les familles se lancent alors dans l'exploration des causes et constatent la présence de décharges et de friches industrielles ainsi que l'existence de polluants. Dans un premier temps, les experts gouvernementaux concluent qu'il n'y a rien d'étrange. Mais les familles s'obstinent et saisissent leurs propres experts. Une étude statistique montre qu'il se passe sans doute quelque chose « d'étrange ».
Le tableau suivant résume les données statistiques concernant les enfants de Woburn de moins de 15 ans, pour la période 1969-1979 (Sources : *Massachusetts Department of Public Health* et *Harvard University*).

Enfants entre 0 et 14 ans	Population de Woburn selon le recencement de 1970 n	Nombre de cas de leucémie infantile observés à Woburn entre 1969 et 1979	Fréquence des leucémies à Woburn f	Fréquence des leucémies aux Etats-Unis p
Garçons	5969	9	0,00151	0,00052
Filles	5779	3	0,00052	0,00038
Total	11748	12	0,00102	0,00045

La question statistique qui se pose est de savoir si le hasard seul peut raisonnablement expliquer les fréquences observées à Woburn, considérées comme résultant d'un échantillon prélevé dans la population américaine.
La population des Etats-Unis étant très grande par rapport à celle de Woburn, on peut considérer que l'échantillon résulte d'un tirage avec remise et simuler des tirages de taille n avec le tableur.

Dans le cas des garçons, simulons sur le tableur 100 échantillons de taille $n = 5969$ prélevés dans une population où $p = 0{,}00052$. On représente sur un graphique les 100 fréquences f des cas de leucémie observés sur les échantillons simulés. La taille de cette simulation peut demander quelques secondes de calcul, selon la puissance de l'ordinateur. Voici l'affichage que nous avons obtenu.



B2 =ENT(ALEA()+0,00052)

	A	B	C	D	E	F	G	H	I	J	K	L
1	simulations avec p = 0,00052 (garçons)	échantillon 1	échantillon 2	échantillon 3	échantillon 4	échantillon 5	échantillon 6	échantillon 7	échantillon 8	échantillon 9	échantillon 10	échantillon 11
2		0	0	0	0	0	0	0	0	0	0	0
3		0	0	0	0	0	0	0	0	0	0	0
4		0	0	0	0	0	0	0	0	0	0	0
5		0	0	0	0	0	0	0	0	0	0	0
6		0	0	0	0	0	0	0	0	0	0	0
5942		0	0	0	0	0	0	0	0	0	0	0
5943		0	0	0	0	0	0	0	0	0	0	0
5944		0	0	0	0	0	0	0	0	0	0	0
5945		0	0									0
5946		0	0									0
5947		0	0									0
5948		0	0									0
5949		0	0									0
5950		0	0									0
5951		0	0									0
5952		0	0									0
5953		0	0									0
5954		0	0									0
5955		0	0									0
5956		0	0									0
5957		0	0									0
5958		0	0									0
5959		0	0									0
5960		0	0									0
5961		0	0									0
5962		0	0									0
5963		0	0									0
5964		0	0									0
5965		0	0									0
5966		0	0									0
5967		0	0									0
5968		0	0									0
5969		0	0	0	0	0	0	0	0	0	0	0
5970		0	0	0	0	0	0	0	0	0	0	0
5971	nombre de cas	3	2	3	2	2	5	3	1	2	4	3
5972	fréquence f	0,0005026	0,00033506	0,0005026	0,00033506	0,00033506	0,00083766	0,0005026	0,00016753	0,00033506	0,00067013	0,000502(
5973												
5974	nombre d'échantillons où f >= 0,00151			0	%							

Cette simulation montre que plus de 95 % des fluctuations aléatoires des valeurs de *f* s'effectuent dans l'intervalle [0 ; 0,001].

L'instruction =NB.SI(B5972:CW5972;">=0,00151") entrée en cellule D5974 confirme qu'aucun échantillon simulé n'a montré une fréquence de leucémie infantile chez les garçons atteignant le niveau de 0,00151 observé à Woburn. On ne peut donc pas raisonnablement attribuer au seul hasard le niveau très « significativement » élevé des leucémies infantiles observées chez les garçons à Woburn.

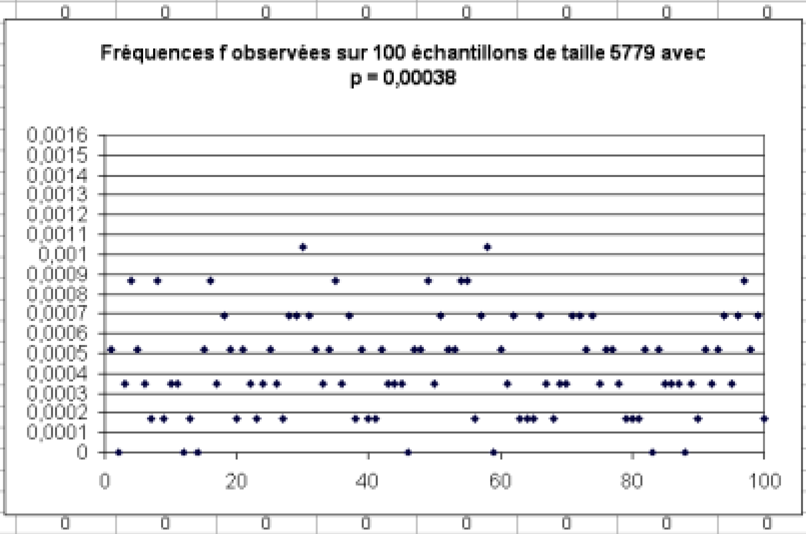
Pour ce qui est des filles, nous simulons de manière analogue sur le tableur 100 échantillons de taille n = 5779 prélevés dans une population où p = 0,00038. L'ordinateur affiche les résultats suivants :

B2 = =ENT(ALEA()+0,00038)

	A	B	C	D	E	F	G	H	I	J	K	L
1	simulations avec p = 0,00038 (filles)	échantillon 1	échantillon 2	échantillon 3	échantillon 4	échantillon 5	échantillon 6	échantillon 7	échantillon 8	échantillon 9	échantillon 10	échantillon 11
2		0	0	0	0	0	0	0	0	0	0	0
3		0	0	0	0	0	0	0	0	0	0	0
4		0	0	0	0	0	0	0	0	0	0	0
5		0	0	0	0	0	0	0	0	0	0	0

D5784 = =NB.SI(B5782:CW5782;">=0,00052")

	A	B	C	D	E	F	G	H	I	J	K	L
5752		0	0	0	0	0	0	0	0	0	0	0
5753		0	0	0	0	0	0	0	0	0	0	0
5754		0	0	0	0	0	0	0	0	0	0	0
5755		0	0	0	0	0	0	0	0	0	0	0
5756		0	0	0	0	0	0	0	0	0	0	0
5757		0	0	0								
5758		0	0	0								
5759		0	0	0								
5760		0	0	0								
5761		0	0	0								
5762		0	0	0								
5763		0	0	0								
5764		0	0	0								
5765		0	0	0								
5766		0	0	0								
5767		0	0	0								
5768		0	0	0								
5769		0	0	0								
5770		0	0	0								
5771		0	0	0								
5772		0	0	0								
5773		0	0	0								
5774		0	0	0								
5775		0	0	0								
5776		0	0	0								
5777		0	0	0								
5778		0	0	0								
5779		0	0	0								
5780		0	0	0	0	0	0	0	0	0	0	0
5781	nombre de cas	3	0	2	5	3	2	1	5	1	2	2
5782	fréquence f	0,00051912	0	0,00034608	0,0008652	0,00051912	0,00034608	0,00017304	0,0008652	0,00017304	0,00034608	0,00034608
5783												
5784	nombre d'échantillons où f >= 0,00052			25	%							



Cette fois, l'instruction =NB.SI(B5782:CW5782;">=0,00052") entrée en D5784 montre que 25 % des échantillons simulés avec p = 0,00038 font apparaître une fréquence f supérieure ou égale à celle observée avec les données de Woburn. On peut donc penser que le taux de leucémies infantiles observé chez les filles à Woburn n'est pas « significativement » élevé. Le hasard pourrait l'expliquer. La taille de l'échantillon est en tout cas trop faible pour mettre en évidence ici un phénomène « anormal ».

Le taux anormalement élevé de leucémies infantiles chez les garçons à Woburn est officiellement confirmé par le Département de Santé Publique du Massachusetts en avril 1980. Les soupçons se portent alors sur la qualité de l'eau de la nappe phréatique qui, par des forages, alimente la ville. On découvre alors le syndrome du trichloréthylène. Les industriels responsables de cette pollution sont traduits en justice, les familles obtiendront des « réparations » financières et la dépollution des sites sera engagée. Suite à cette affaire, le discours du nouveau maire montre bien le changement d'attitude des autorités : « notre première priorité, dira-t-il, est de nous assurer d'avoir un approvisionnement en eau propre et saine ».

Quelques remarques, pour approfondir un peu le traitement probabiliste de cet exemple.

Les faibles valeurs de p empêchent, dans cet exemple, d'utiliser la formule de fluctuation de plus de 95 % des fréquences au programme de seconde : [$p - \frac{1}{\sqrt{n}}$, $p + \frac{1}{\sqrt{n}}$]. Les intervalles obtenus sont beaucoup trop grands.
En revanche, avec les données concernant les enfants des deux sexes, on peut appliquer l'approximation par la loi normale, puisque $n \geq 30$ et $np = 5{,}3 \geq 5$. On sait alors qu'environ 95 % des fréquences obtenues sur les échantillons de taille $n = 11748$ dans une population où $p = 0{,}00045$ fluctuent à l'intérieur de l'intervalle :
[$p - 1{,}96 \times \sqrt{\frac{p(1-p)}{n}}$, $p + \sqrt{\frac{p(1-p)}{n}}$] c'est à dire [0,00032 ; 0,00057]. La fréquence $f = 0{,}00102$ observée avec les enfants de Woburn est loin d'appartenir à cet intervalle.
Lorsque p est trop petit, on peut considérer la variable aléatoire X correspondant au nombre d'observations et qui suit la loi binomiale de paramètres n et p. Vu la taille de n, il n'est pas très pratique d'utiliser ici cette loi, que l'on peut en revanche approcher par la loi de Poisson de paramètre $n \times p$ qui est bien adaptée aux cas rares.
Dans le cas des garçons, on peut ainsi calculer la probabilité d'observer un nombre de leucémies infantiles supérieur où égal à celui de Woburn en utilisant la loi de Poisson de paramètre $n \times p = 5969 \times 0{,}00052 \approx 3{,}1$. On obtient $P(X \geq 9) \approx 0{,}0047$. Ce que l'on peut interpréter en disant que l'explication « il ne se passe rien d'étrange » a moins de 0,5 % de chances d'être exacte.
Qui peut encore, après un tel exemple, affirmer que « la statistique est la forme la plus élaborée du mensonge » ?

Le TP sur tableur suivant a été pratiqué en seconde.
La classe est répartie en 14 groupes de 2 élèves par ordinateur – durée du TP : 1 heure.

Le TP est en deux parties. La première partie permet d'apporter aux élèves les outils nécessaires à un traitement statistique des données de Woburn, c'est-à-dire être capable de simuler sur un tableur les fluctuations des fréquences observées sur des échantillons de taille n.

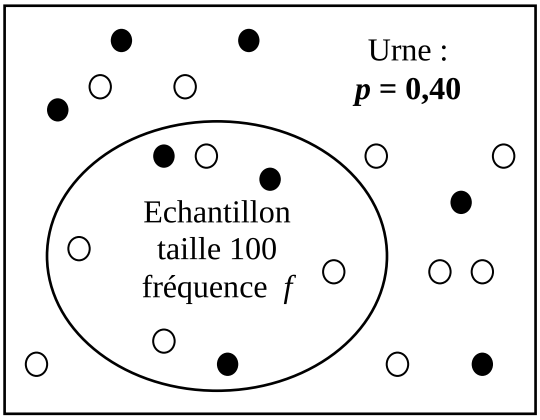
La seconde partie est plus « ouverte ». On demande aux élèves d'imaginer une procédure de simulation permettant de juger si les observations du nombre de cas de leucémies chez les garçons et les filles de Woburn sont à considérer comme « anormales » (le nombre de cas chez les garçons étant « significativement » élevé, mais pas le nombre de cas chez les filles).

Énoncé élève

Ouvrir un fichier Excel.

1 – Etude des fluctuations des échantillons

a. Tirage dans une urne

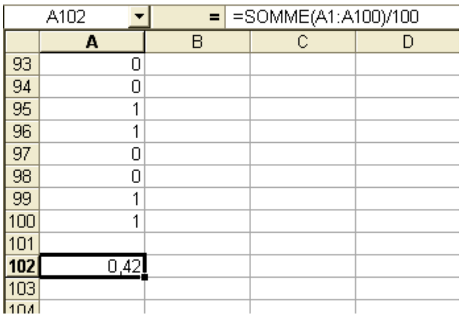


Une urne contient 40 % de boules noires et 60 % de boules blanches. On prélève au hasard une boule (chaque boule a les mêmes chances d'être prélevée).
Pour simuler le tirage d'une boule dans cette urne, il suffit avec Excel d'entrer dans la cellule A1 la ***formule*** : =ENT(ALEA() + 0,4) .

☞ Justifier cette formule sur la feuille réponse.

b. Echantillons de taille 100

On prélève au hasard avec remise 100 boules dans l'urne. Cela constitue un échantillon de taille 100.



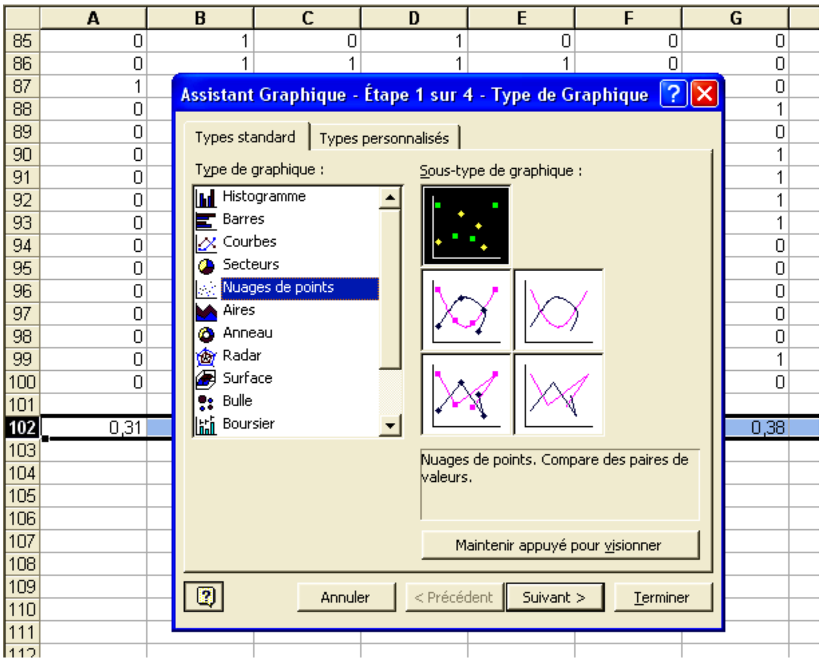
A102 = =SOMME(A1:A100)/100

	A	B	C	D
93	0			
94	0			
95	1			
96	1			
97	0			
98	0			
99	1			
100	1			
101				
102	0,42			
103				
104				

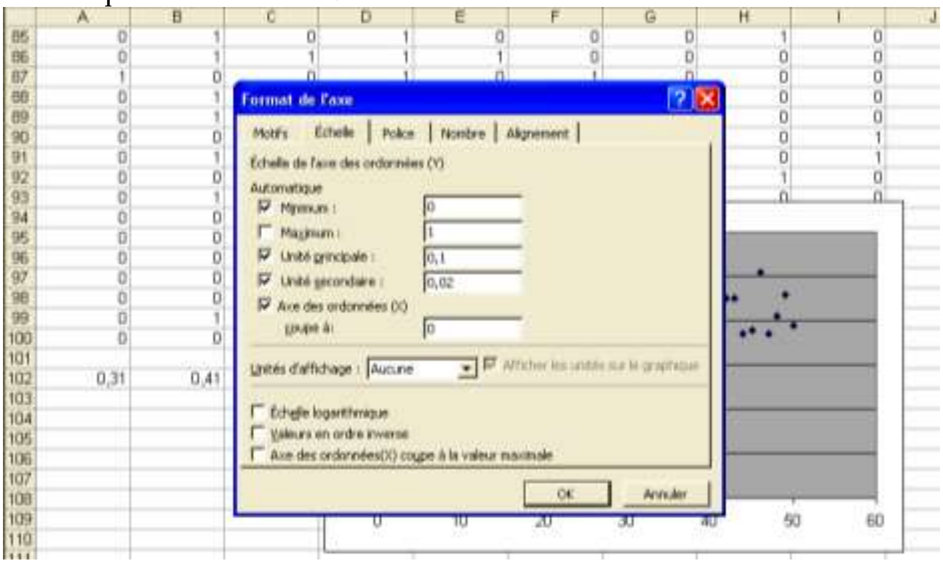
Pour simuler cet échantillon sur le tableur, il suffit de ***recopier vers le bas*** (pointeur de la souris en forme de croix noire) le contenu de la cellule A1 jusqu'en A100.
Calculez la fréquence *f* des boules noires observées sur cet échantillon en entrant en A102 la ***formule*** =SOMME(A1:A100)/100 .

☞ Faire plusieurs fois F9 et compléter la feuille réponse.

Pour visualiser les résultats de 50 échantillons à la fois, ***sélectionner*** (pointeur en forme de croix blanche) les cellules de A1 à A102 puis ***recopier vers la droite*** (pointeur en forme de croix noire) jusqu'à la colonne AX.



Sélectionner la ligne 102 puis cliquer sur l'icône de l'***Assistant graphique***, choisir ***Nuages de points*** et cliquer sur ***Terminer***.



En cliquant avec le bouton droit sur le graphique, ***Effacer*** la légende, puis par un clic droit sur l'axe des ordonnées, régler dans ***Format de l'axe*** le maximum à 1.

☞ Faire plusieurs fois F9 et compléter la feuille réponse.

c. Echantillons de taille 1000

Cliquer sur l'onglet **Feuil2** et procéder comme dans la partie b. pour étudier cette fois le comportement d'échantillons de taille 1000 extraits de l'urne précédente : simuler 50 échantillons de taille 1000 et représenter, avec la même échelle que précédemment, les 50 fréquences obtenues.

☞ Faire plusieurs fois F9 et compléter la feuille réponse.

2 – Inquiétudes à Woburn

Maisons à Woburn, où habitaient certains des enfants malades (Source : Internet.).

Woburn est une petite ville industrielle du Massachusetts, au Nord-Est des Etats-Unis. Du milieu à la fin des années 1970, la communauté locale s'émeut d'un grand nombre de leucémies infantiles survenant dans certains quartiers de la ville. Les familles se lancent alors dans l'exploration des causes et constatent la présence de décharges et de friches industrielles ainsi que l'existence de polluants.

Dans un premier temps, les experts gouvernementaux concluent qu'il n'y a rien d'étrange. Mais les familles s'obstinent et saisissent leurs propres experts.

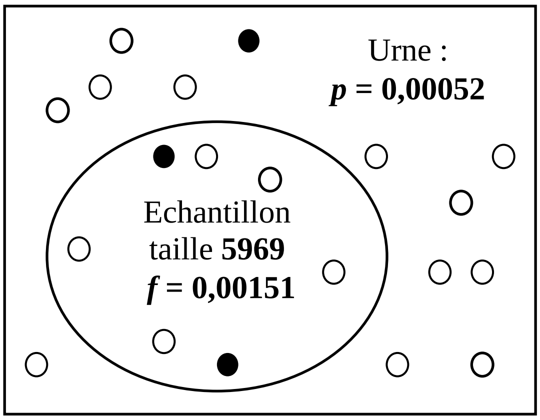
L'objectif de ce paragraphe est que vous apportiez votre expertise statistique !

a. Le cas des garçons

Les statistiques pour les garçons sont les suivantes :

Enfants entre 0 et 14 ans	Population de Woburn selon le recencement de 1970 n	Nombre de cas de leucémie infantile observés à Woburn entre 1969 et 1979	Fréquence des leucémies à Woburn f	Fréquence des leucémies aux Etats-Unis p
Garçons	**5969**	9	**0,00151**	**0,00052**

On peut considérer que les garçons de Woburn constituent un échantillon de taille n = 5969 extrait de la population des Etats-Unis.



Cliquer sur l'onglet **Feuil3**. Elaborer votre feuille de calcul pour étudier si les observations de Woburn, dans le cas des garçons, sont ou non « statistiquement anormales ».

☞ Compléter la feuille réponse.

b. Le cas des filles

Les statistiques pour les filles sont les suivantes :

Enfants entre 0 et 14 ans	Population de Woburn selon le recencement de 1970 *n*	Nombre de cas de leucémie infantile observés à Woburn entre 1969 et 1979	Fréquence des leucémies à Woburn *f*	Fréquence des leucémies aux Etats-Unis *p*
Filles	**5779**	3	**0,00052**	**0,00038**

Cliquer sur l'onglet **Feuil4**. Elaborer votre feuille de calcul pour étudier si les observations de Woburn, dans le cas des filles, sont ou non « statistiquement anormales ».

☞ Compléter la feuille réponse.

Feuille réponse

NOMS : ………………………………………………………………………………

1 – Etude des fluctuations des échantillons

a. Tirage dans une urne

Quels sont les résultats possibles de la formule =ENT(ALEA() + 0,4) ?

……………………………………………………………………………………………………

Pourquoi a-t-on 40 % de chances de voir afficher la valeur 1 par la formule =ENT(ALEA() + 0,4) ? (La réponse « parce qu'il y a écrit 0,4 » n'est pas correcte).

……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

b. Echantillons de taille 100

Pourquoi les échantillons de taille 100 ne donnent-ils pas la même fréquence ?

……………………………………………………………………………………………………
……………………………………………………………………………………………………

Sur 50 échantillons de taille 100 apparaissant sur le graphique, combien, en moyenne (faire plusieurs fois F9), donnent une fréquence f n'appartenant pas à l'intervalle [0,3 ; 0,5] ?

……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

Sur un échantillon de taille 100, est-il rare d'observer 45 % de boules noires ?

……………………………………………………………………………………………………

Sur un échantillon de taille 100, est-il rare d'observer 50 % de boules noires ? Est-ce possible ?

……………………………………………………………………………………………………
……………………………………………………………………………………………………

Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?

……………………………………………………………………………………………………
……………………………………………………………………………………………………

c. Echantillons de taille 1000

Quelle différence observez-vous, sur le graphique, entre le comportement des fréquences des échantillons de taille 100 et celui des fréquences des échantillons de taille 1000 ?

……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.

……………………………………………………………………………………………………
……………………………………………………………………………………………………

2 – Inquiétudes à Woburn

a. Le cas des garçons

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

....
....
....
....
....
....
....
....
....
....

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

....
....
....
....
....
....
....
....

b. Le cas des filles

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

....
....
....
....
....
....
....
....
....
....

Quelle est votre conclusion concernant les cas de leucémies observés chez les filles à Woburn (justifier) ?

....
....
....
....
....
....
....
....

Déroulement, éléments de réponse et commentaires

Première partie (traitée par tous les groupes)

Une urne contient 40% de boules noires. Simulation de tirages (avec remise) d'échantillons de taille 100 ou 1000.

Objectifs de cette partie :

– Comprendre le principe de la simulation (modèle de l'urne) pour être capable de l'adapter et de le réinvestir dans le cas des statistiques de leucémies à Woburn.

Cet objectif est atteint : **11** groupes sur 14 ont su adapter cette simulation au cas de Woburn. **3** groupes sur 14 n'ont pas traité la deuxième partie par manque de rapidité et surtout par manque de sérieux.

– Comprendre que d'un certain point de vue, le « hasard » ne se comporte pas n'importe comment : certaines observations sont plus ou moins rares selon la taille de l'échantillon.

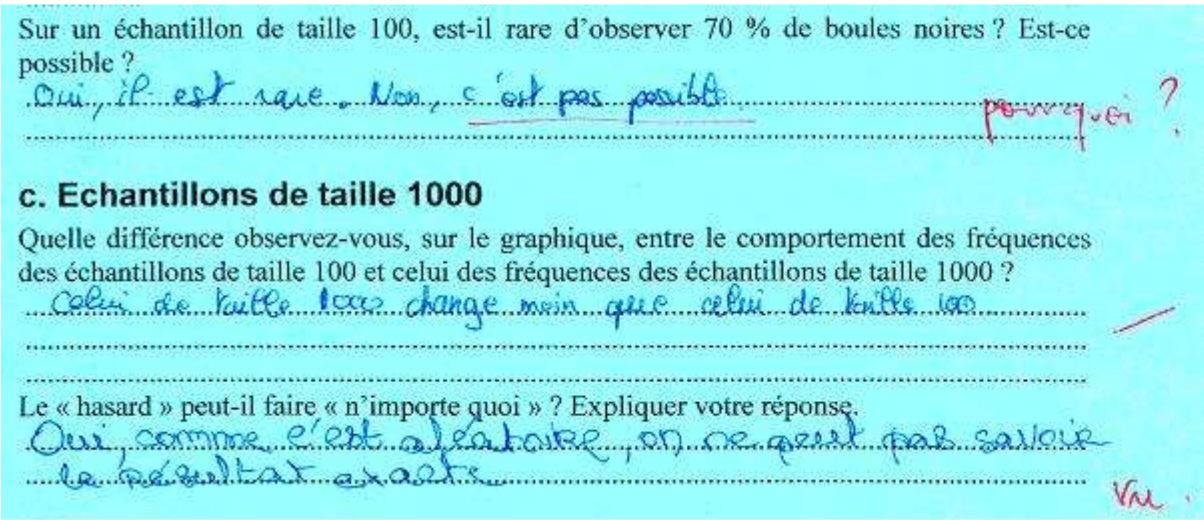
Il s'agissait, par l'observation des simulations, de lutter contre l'idée a priori répandue « hasard » = « n'importe quoi ». L'objectif est « atteint » au-delà du « raisonnable » puisque **11** groupes sur 14 qualifient « d'impossible » un événement (70% de boules noires sur un échantillon de taille 100) qui n'a pas été observé.

Exemples de réponses d'élèves

On s'intéresse aux questions :

- Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
- Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.

REPONSES DU TYPE « n'est possible que ce que j'observe » (11/14) :


Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
Oui, il est rare. Non, c'est pas possible. pourquoi ?

c. Echantillons de taille 1000

Quelle différence observez-vous, sur le graphique, entre le comportement des fréquences des échantillons de taille 100 et celui des fréquences des échantillons de taille 1000 ?
Celui de taille 1000 change moin que celui de taille 100

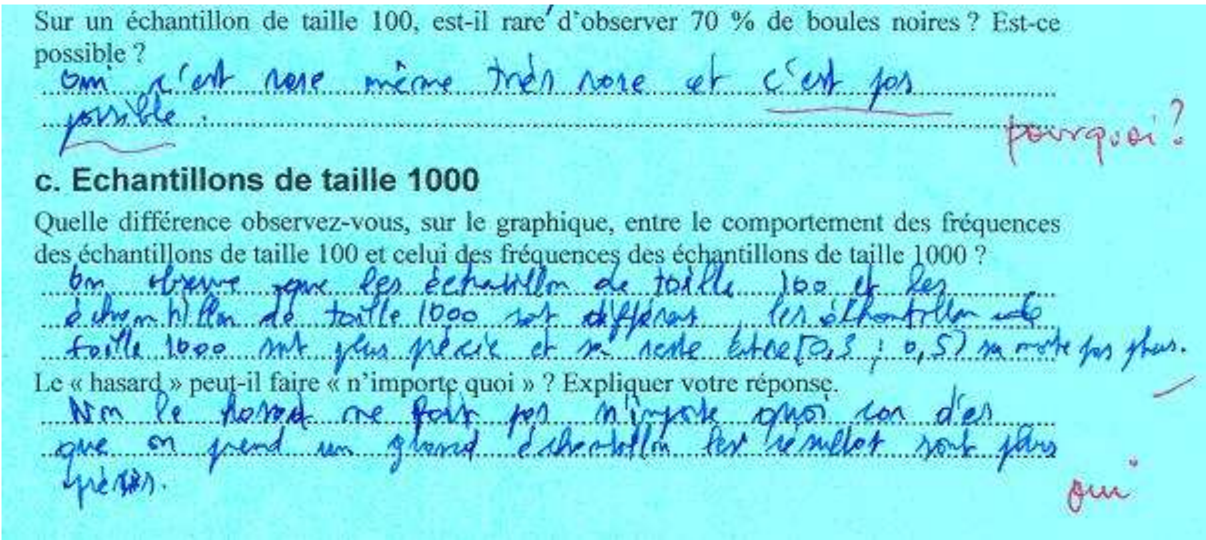
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
Oui, comme c'est aléatoire, on ne peut pas savoir le résultat exacte.
Vu.

Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?

Oui, il est rare. Non, c'est pas possible.

Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.

Oui, comme c'est aléatoire, on ne peut pas savoir le résultat exact.


Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
Oui c'est rare même très rare et c'est pas possible. pourquoi ?

c. Echantillons de taille 1000

Quelle différence observez-vous, sur le graphique, entre le comportement des fréquences des échantillons de taille 100 et celui des fréquences des échantillons de taille 1000 ?
On observe que les échantillon de taille 100 et les échantillon de taille 1000 sont différent. les échantillon de taille 1000 sont plus précis et ne reste entre [0,3 ; 0,5] ne sorte pas plus.

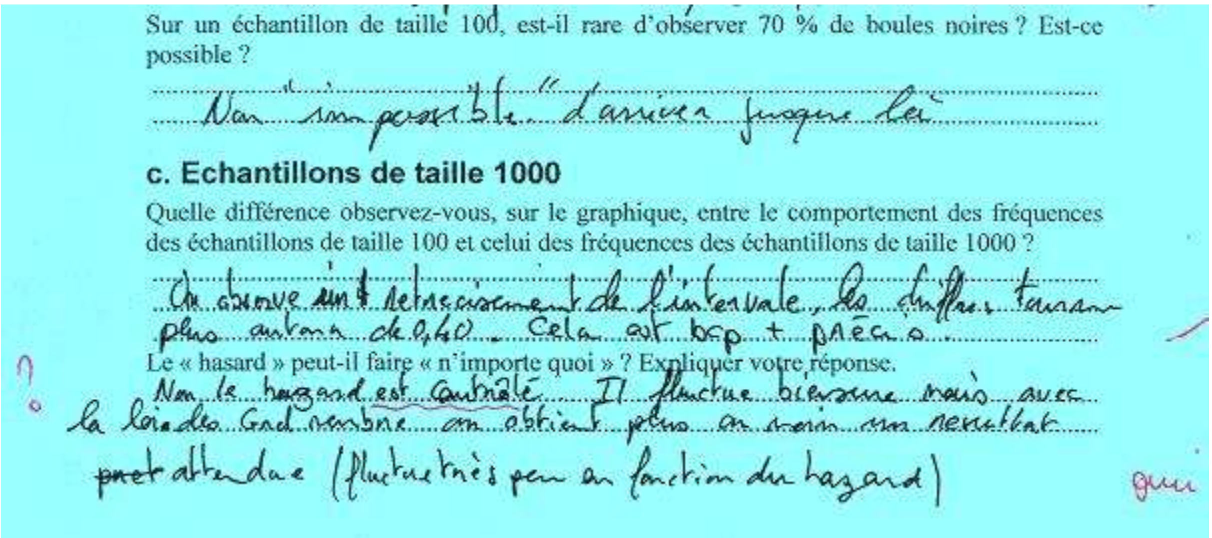
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
Non le hasard ne fait pas n'importe quoi car dès que on prend un grand échantillon les resultat sont plus précis. oui

Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?

Oui c'est rare, même très rare et c'est pas possible.

Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.

Non le hasard ne fait pas n'importe quoi car dès qu'on prend un grand échantillon les résultats sont plus précis.


Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
Non impossible d'arriver jusque là

c. Echantillons de taille 1000

Quelle différence observez-vous, sur le graphique, entre le comportement des fréquences des échantillons de taille 100 et celui des fréquences des échantillons de taille 1000 ?
On observe un rétrécissement de l'intervalle, les chiffres tournant plus autour de 0,40. Cela est bcp + précis.

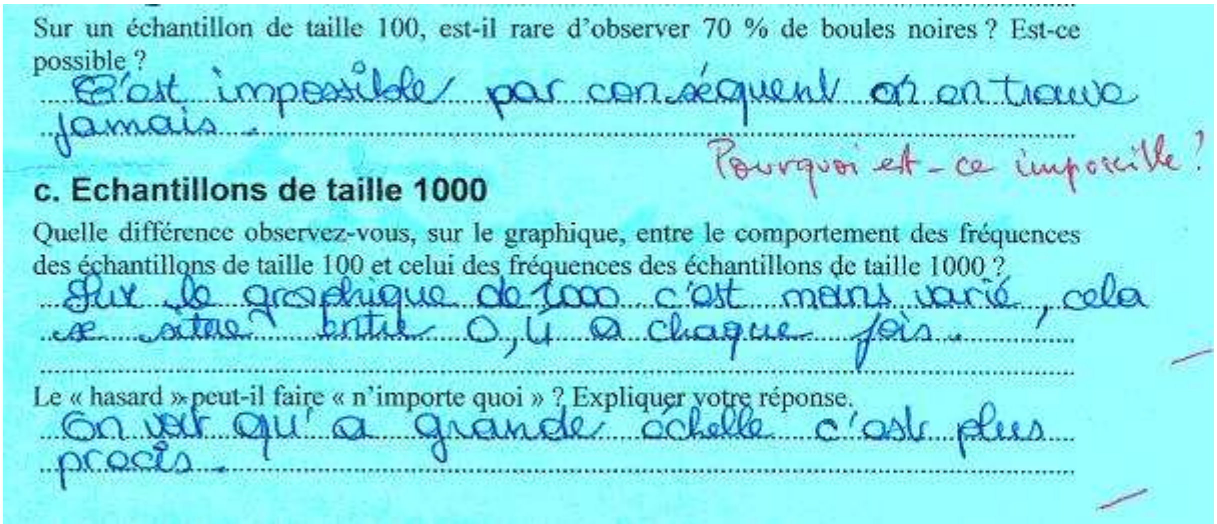
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
Non le hasard est contrôlé. Il fluctue bien sûr mais avec la loi des Grands nombres on obtient plus ou moins un résultat attendu (fluctue très peu en fonction du hasard) oui

Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?

Non « impossible » d'arriver jusque là.

Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.

Non le hasard est contrôlé. Il fluctue bien sûr, mais avec la loi des grands nombres, on obtient plus ou moins un résultat attendu (fluctue très peu en fonction du hasard).



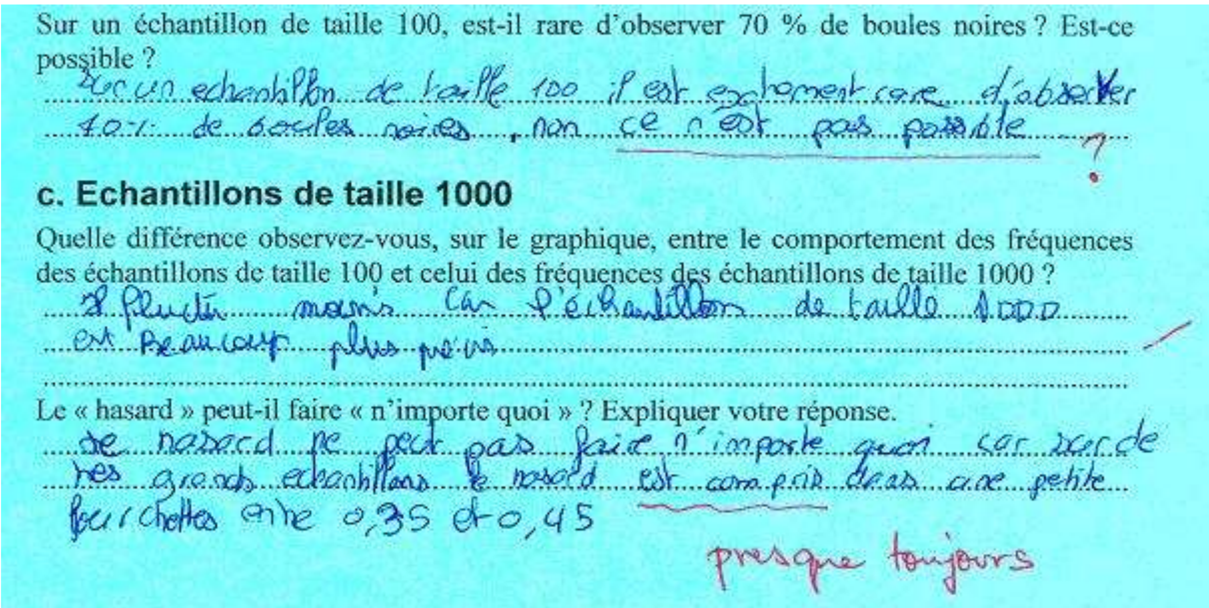
Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
C'est impossible par conséquent on en trouve jamais.
Pourquoi est-ce impossible ?

c. Echantillons de taille 1000

Quelle différence observez-vous, sur le graphique, entre le comportement des fréquences des échantillons de taille 100 et celui des fréquences des échantillons de taille 1000 ?
Sur le graphique de 1000 c'est moins varié, cela se situe entre 0,4 à chaque fois.

Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
On voit qu'à grande échelle c'est plus précis.

Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
C'est impossible par conséquent on en trouve jamais.
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
On voit qu'à grande échelle c'est plus précis.



Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
Sur un echantillon de taille 100 il est extrêmement rare d'observer 70 % de boules noires, non ce n'est pas possible
?

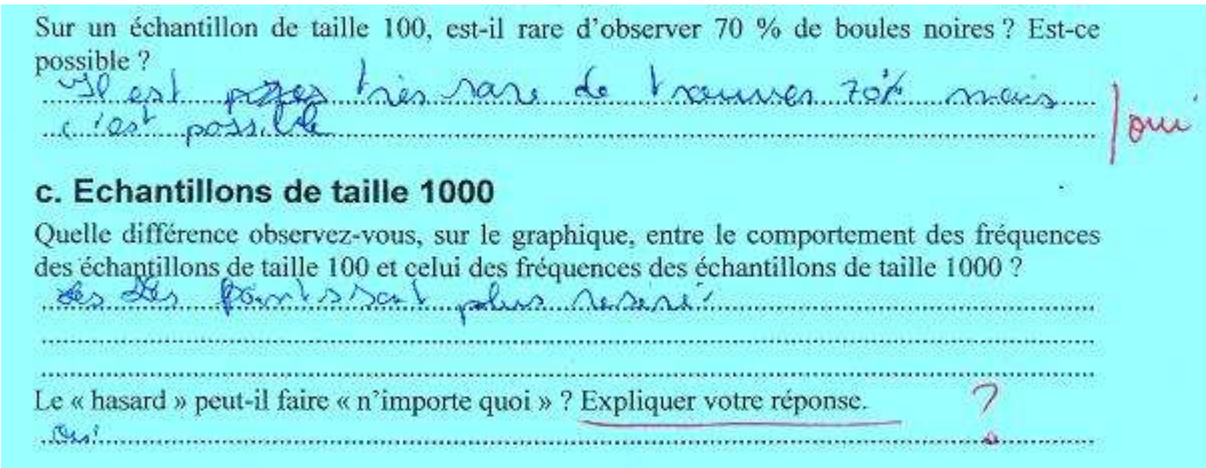
c. Echantillons de taille 1000

Quelle différence observez-vous, sur le graphique, entre le comportement des fréquences des échantillons de taille 100 et celui des fréquences des échantillons de taille 1000 ?
Il fluctue moins car l'échantillon de taille 1000 est beaucoup plus précis

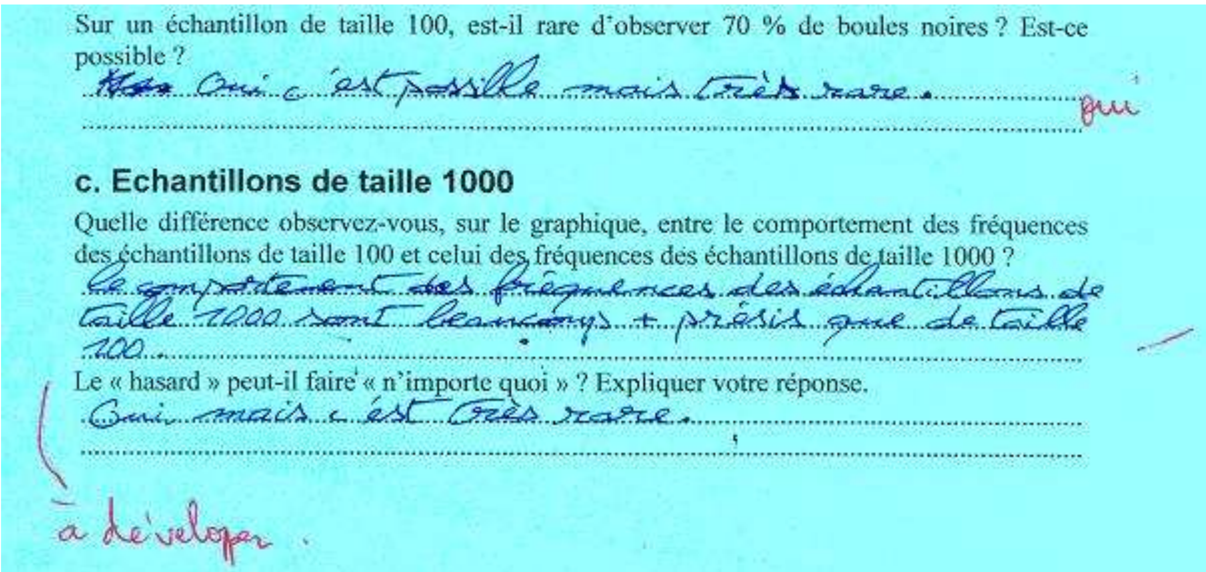
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
Le hasard ne peut pas faire n'importe quoi car sur de tres grands echantillons le hasard est compris dans une petite fourchette entre 0,35 et 0,45
presque toujours

Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
Sur un échantillon de taille 100 il est extrêmement rare d'observer 70 % de boules noires. non ce n'est pas possible.
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
Le hasard ne peut pas faire n'importe quoi car sur de très grands échantillons le hasard est compris dans une petite fourchette entre 0,35 et 0,45 [35% et 45% de boules noires].

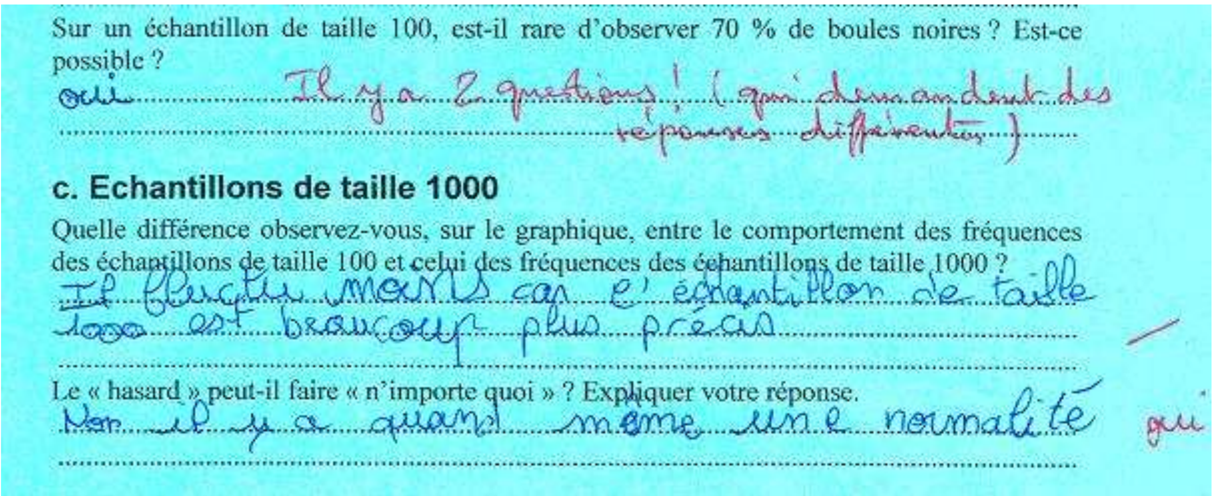
REPONSES DU TYPE « tout est possible » ou « tout est possible mais c'est rare » (3/14) :


Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
Il est ~~p~~ très rare de trouver 70% mais c'est possible
oui
c. Echantillons de taille 1000
Quelle différence observez-vous, sur le graphique, entre le comportement des fréquences des échantillons de taille 100 et celui des fréquences des échantillons de taille 1000 ?
les points sont plus resserrés
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse. ?
Oui !

Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
Il est très rare de trouver 70 % mais c'est possible.
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
Oui.


Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
Oui c'est possible mais très rare.
oui
c. Echantillons de taille 1000
Quelle différence observez-vous, sur le graphique, entre le comportement des fréquences des échantillons de taille 100 et celui des fréquences des échantillons de taille 1000 ?
Le comportement des fréquences des échantillons de taille 1000 sont beaucoups + présis que de taille 100.
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
Oui mais c'est très rare.
à développer.

Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
Oui c'est possible mais très rare.
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
Oui mais c'est très rare.



Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
oui Il y a 2 questions ! (qui demandent des réponses différentes)

c. Echantillons de taille 1000

Quelle différence observez-vous, sur le graphique, entre le comportement des fréquences des échantillons de taille 100 et celui des fréquences des échantillons de taille 1000 ?
Il fluctue moins car l'échantillon de taille 1000 est beaucoup plus précis

Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
Non il y a quand même une normalité qui

Sur un échantillon de taille 100, est-il rare d'observer 70 % de boules noires ? Est-ce possible ?
Oui.
Le « hasard » peut-il faire « n'importe quoi » ? Expliquer votre réponse.
Non, il y a quand même une normalité.

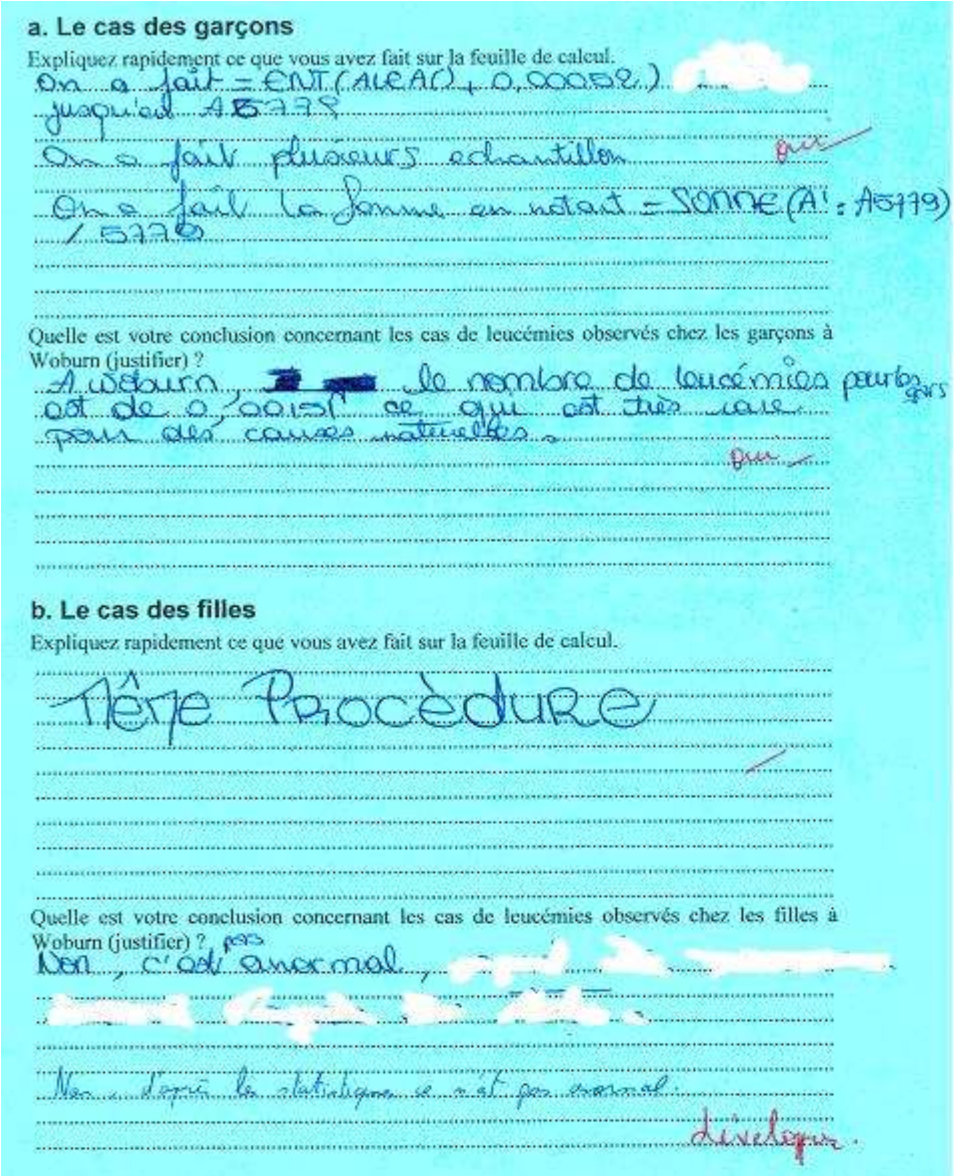
Seconde partie (traitée par 11 groupes sur 14)

Fréquences des leucémies chez les garçons et chez les filles à Woburn, en comparaison aux fréquences correspondantes aux Etats-Unis.

Objectifs :
– Penser à simuler des échantillons dont la taille correspond à la population de Woburn sur la base des fréquences observées aux Etats-Unis.
Cela a été fait par les 11 groupes qui se sont penchés sur cette parties. Ils étaient cependant aidés par un schéma.
– Constater que la fréquence chez les garçons de Woburn est tout à fait « anormale » alors que celle des filles ne peut pas être considérée comme « anormale ».
Seuls **3** groupes sur 11 fournissent une bonne réponse complète, plus ou moins argumentée. **1** groupe fournit une bonne réponse partielle (pas le temps d'étudier le cas des filles). **7** groupes sur 11 fournissent de mauvaises réponses, généralement non argumentées : il s'agit d'évidences du type « les leucémies sont rares » ou « il y a plus de leucémies à Woburn que la moyenne aux Etats-Unis ».

Exemples de réponses d'élèves

« BONNES » REPONSES (4/11) :


a. Le cas des garçons

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

On a fait =ENT(ALEA()+0,00052)
jusqu'à A5779

On a fait plusieurs échantillon oui

On a fait la somme en notant =SOMME(A1:A5779)
/5779

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

A Woburn, le nombre de leucémies pour les gars est de 0,00151 ce qui est très rare pour des causes naturelles.
oui

b. Le cas des filles

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

Même Procédure

Quelle est votre conclusion concernant les cas de leucémies observés chez les filles à Woburn (justifier) ?

Non, c'est pas anormal,

Non d'après la statistique ce n'est pas anormal.
développer.

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

A Woburn, le nombre [la fréquence] de leucémies pour les gars est de 0,00151 ce qui est très rare pour des causes naturelles.

Quelle est votre conclusion concernant les cas de leucémies observés chez les filles à Woburn (justifier) ?

Non, c'est pas anormal.

a. Le cas des garçons

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

= Ent(ALEA()+0,00052) puis tiré jusque A 5969

On a fait la somme. Somme(A1..A5969)/5969

On la tiré au 50 jusqu'a AX. On a fait le tableau des resultats des 50 et on obtient le resultat

oui

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

A l'aide du tableau on peut voir qu'obtenire 0,00151 est très rare

oui

Le fait d'obtenire 9 cas dans un même endroit ne sera pas dû au hazard donc elle est dû à une cause bien précise.

sans doute

oui

b. Le cas des filles

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

Même procédure

Quelle est votre conclusion concernant les cas de leucémies observés chez les filles à Woburn (justifier) ?

La fréquence à Woburn est de 0,00052

Avec le resultat on peut s'apercevoir que que le "hasard" normale se situe entre entre 0 et 0,001.

Donc pour les fille, ça peut très bien être le ~~Bas~~ hazard

Bien

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

A l'aide du tableau on peut voir qu'obtenir 0,00151 est très rare.

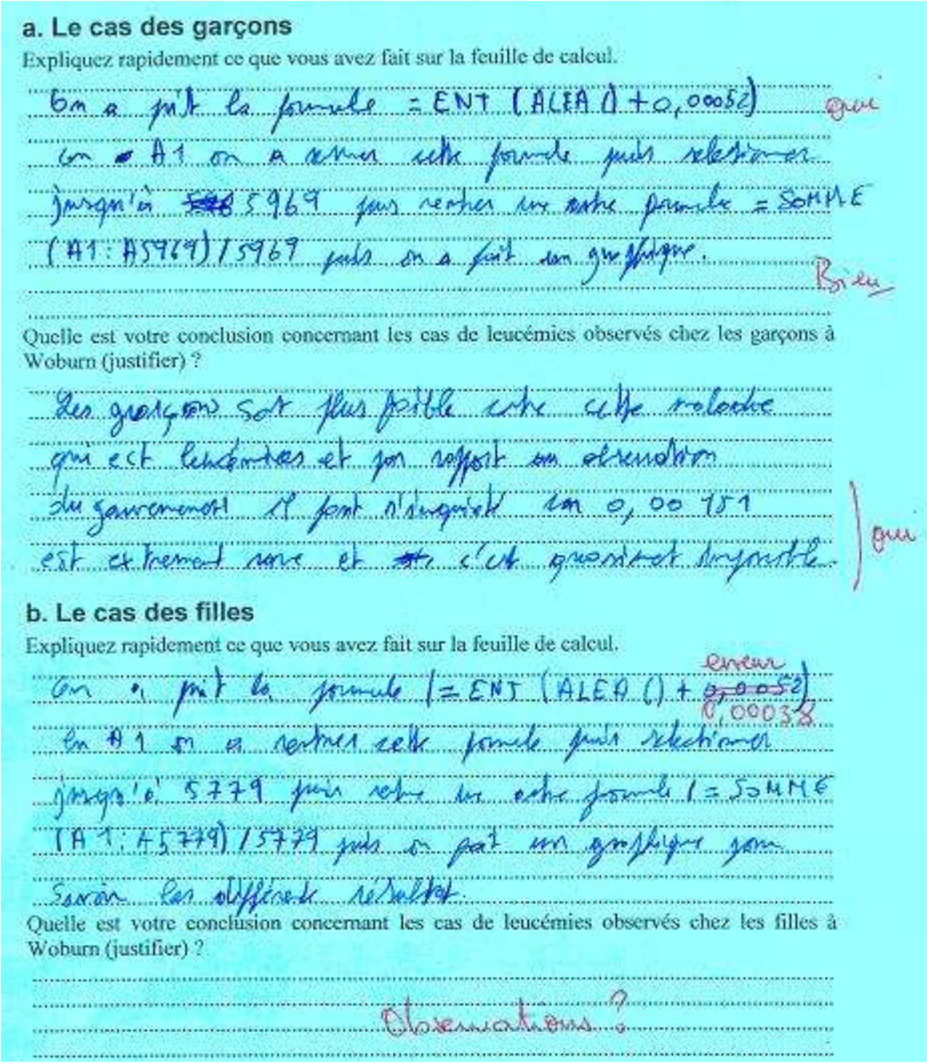
Le fait d'obtenir 9 cas dans un même endroit ne sera pas dû au hasard, donc c'est dû à une cause bien précise.

Quelle est votre conclusion concernant les cas de leucémies observés chez les filles à Woburn (justifier) ?

La fréquence à Woburn est de 0,00052.

Avec le résultat on peut s'apercevoir que le « hasard » normal se situe entre 0 et 0,001.

Donc pour les filles, ça peut très bien être le hasard.


a. Le cas des garçons

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

On a prit la formule =ENT (ALEA()+0,00052) que
en A1 on a entrer cette formule puis sélectionner
jusqu'à ~~5969~~ 5969 puis rentrer une autre formule =SOMME
(A1:A5969)/5969 puis on a fait un graphique. Bien

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

Les garçons sont plus faible contre cette maladie
qui est leucémies et par rapport aux obervations
du gouvernement il faut s'inquiéter car 0,00151
est extrement rare et ~~ça~~ c'est quasiment impossible. qui

b. Le cas des filles

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

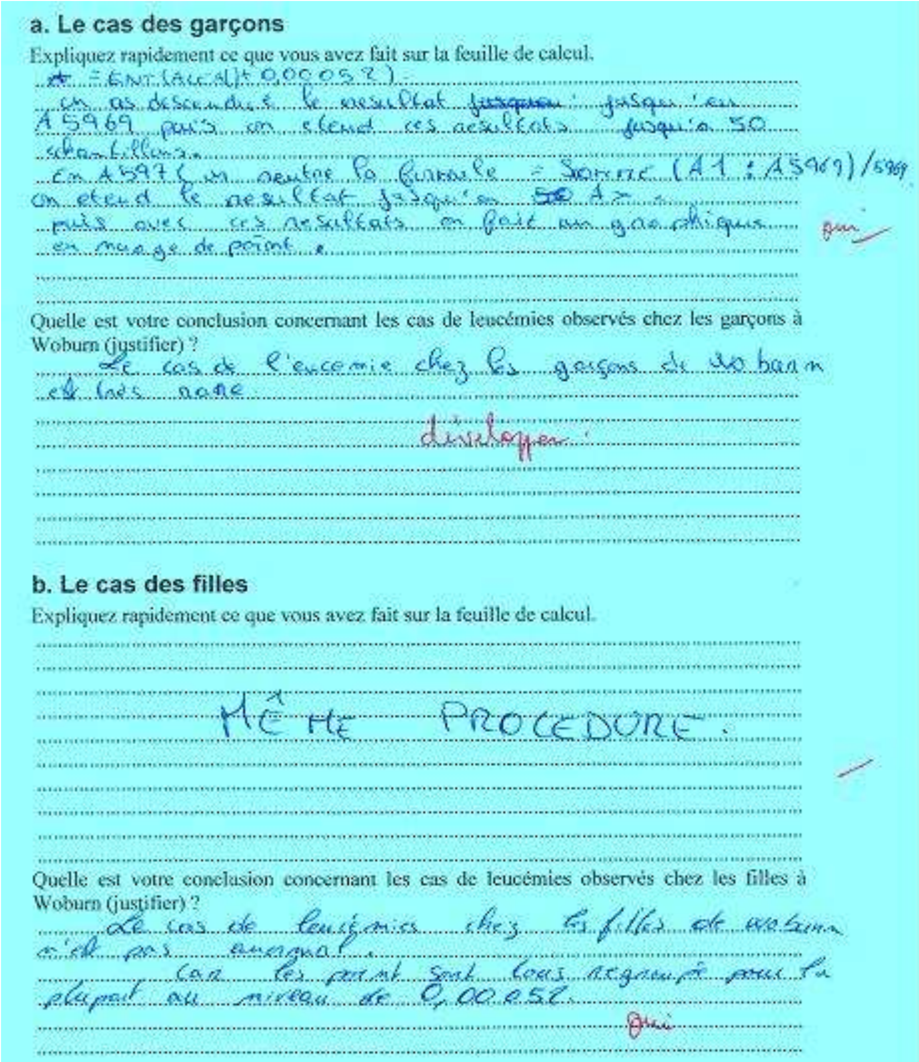
On a prit la formule (=ENT (ALEA()+ ~~0,00052~~ 0,00038) erreur
En A1 on a rentrer cette formule puis sélectionner
jusqu'à 5779 puis rentrer une autre formule (=SOMME
(A1:A5779)/5779 puis on fait un graphique pour
savoir les différents résultat.

Quelle est votre conclusion concernant les cas de leucémies observés chez les filles à Woburn (justifier) ?

Observations ?

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

Les garçons sont faibles contre cette maladie qui est la leucémie et par rapport aux observations du gouvernement il faut s'inquiéter car 0,00151 est extrêmement rare et c'est quasiment impossible.



a. Le cas des garçons

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

=ENT(ALEA()+0,00052)
on as descendus le resultat jusqu'en A5969 puis on etend ces resultats jusqu'a 50 echantillons.
En A5971 on rentre la formule =Somme(A1:A5969)/5969
on etend le resultat jusqu'en 50 A.
puis avec ces resultats on fait un graphique en nuage de point.

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

Le cas de l'eucemie chez les garçons de woburn est tres rare

développer !

b. Le cas des filles

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

MÊME PROCEDURE.

Quelle est votre conclusion concernant les cas de leucémies observés chez les filles à Woburn (justifier) ?

Le cas de leucémies chez les filles de woburn n'est pas anormal.
car les point sont tous regroupé pour la plupart au niveau de 0,00052.

oui

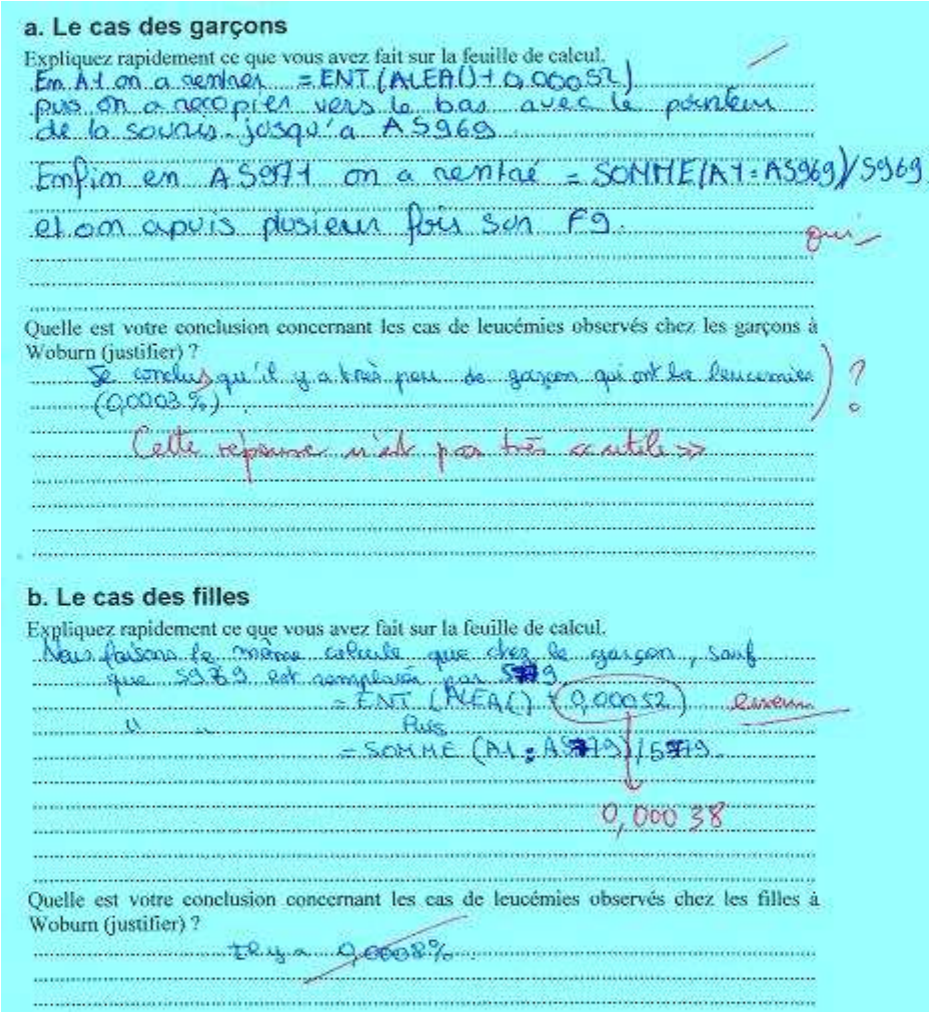
Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

Le cas de leucémie chez les garçons de Woburn est très rare.

Quelle est votre conclusion concernant les cas de leucémies observés chez les filles à Woburn (justifier) ?

Le cas de leucémie chez les filles de Woburn n’est pas anormal car les points sont tous regroupés pour la plupart au niveau de 0,00052 [valeur à Woburn].

« MAUVAISES » REPONSES OU ABSENCE DE REPONSE (7/11) :



a. Le cas des garçons

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

En A1 on a rentrer =ENT(ALEA()+0,00052)
puis on a recopier vers le bas avec le pointeur de la souris jusqu'a A5969
Enfin en A5971 on a rentrer =SOMME(A1:A5969)/5969
et on a puis plusieur fois sur F9. qui

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

Je conclus qu'il y a très peu de garçon qui ont la leucemie (0,0003 %) ?

Cette réponse n'est pas très utile

b. Le cas des filles

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

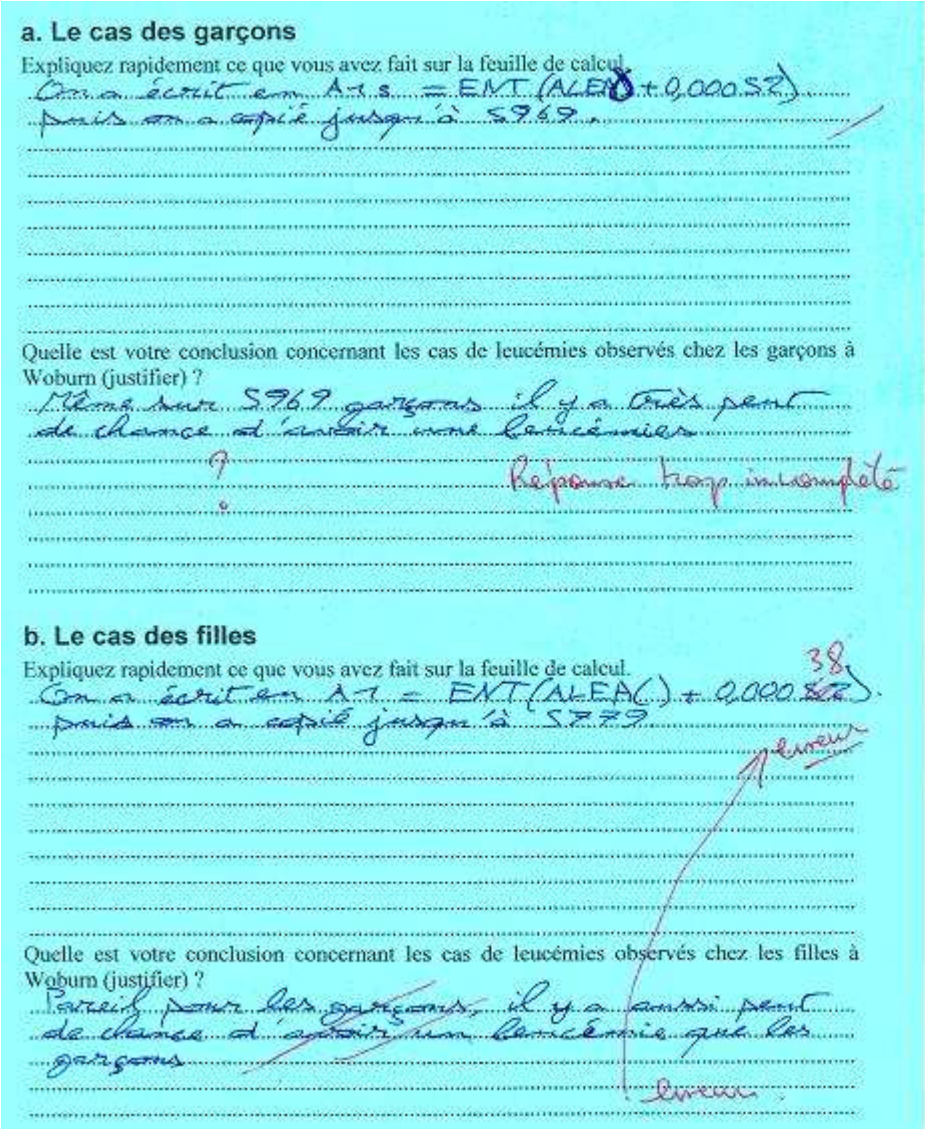
Nous faisons le même calcule que chez le garçon, sauf que 5969 est remplacée par 5[illegible]9
=ENT(ALEA()+(0,00052)) erreur
Puis
=SOMME(A1:A5[illegible]9)/5[illegible]9
0,000 38

Quelle est votre conclusion concernant les cas de leucémies observés chez les filles à Woburn (justifier) ?

~~Il y a 0,0008%~~

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

Je conclus qu'il y a très peu de garçons qui ont la leucémie.



a. Le cas des garçons

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

On a écrit en A1 = ENT (ALEA() + 0,00052)
puis on a copié jusqu'à 5969.

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

Même sur 5969 garçons il y a très peu de chance d'avoir une leucémie
?
Réponse trop incomplète

b. Le cas des filles

Expliquez rapidement ce que vous avez fait sur la feuille de calcul.

On a écrit en A1 = ENT(ALEA() + 0,000 ~~52~~ 38)
puis on a copié jusqu'à 5779
erreur

Quelle est votre conclusion concernant les cas de leucémies observés chez les filles à Woburn (justifier) ?

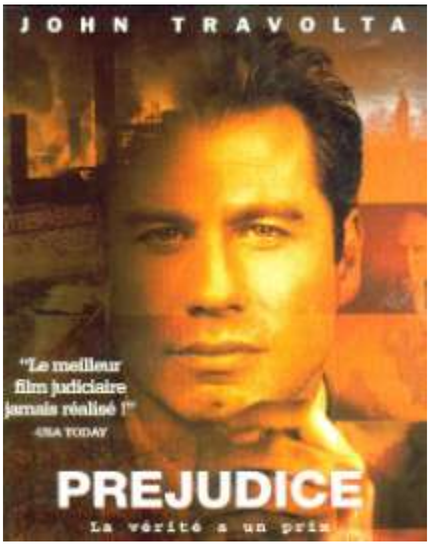
Pareil pour les garçons, il y a aussi peu de chance d'avoir une leucémie que les garçons
erreur

Quelle est votre conclusion concernant les cas de leucémies observés chez les garçons à Woburn (justifier) ?

Même sur 5969 garçons il y a très peu de chances d'avoir une leucémie.

Prolongement

Visionnage du film « Préjudice » en ECJS (Education civique juridique et sociale)



Le film « Préjudice » (*A Civil Action*) sorti en 1999, avec John Travolta dans le rôle de l'avocat, décrit le procès attenté par les familles des victimes de Woburn aux industriels responsables des pollutions. Il a été diffusé en cours d'ECJS, après le TP informatique réalisé en maths.

Un questionnaire a été rempli par les élèves après la projection. La première question revient sur l'aspect mathématique, à propos d'une réflexion du début du film, faisant état de « statistiques significatives ».

Texte du questionnaire d'ECJS

Le film *Préjudice* que vous allez voir raconte l'histoire du procès qui a opposé les parents des victimes de la pollution industrielle dans les années 1970 à Woburn dans le Massachussets à l'entreprise qui en était responsable.

Lisez les questions suivantes afin de pouvoir y répondre pendant le visionnage du film.

1) Qu'entend-t-on par « statistiques significatives » à propos de la fréquence des leucémies infantiles chez les garçons de Woburn ?
2) Identifiez le rôle de chacun des personnages importants de l'affaire.
3) En quoi l'argent joue-t-il un rôle essentiel ?
4) L'action des parents fait-elle preuve de civisme selon vous ? Pourquoi ?
5) Quel fut le dénouement de cette affaire ? Grâce à quelle intervention ?

Des réponses d'élèves à la première question de ce questionnaire

« *On entend que les morts des garçons de Woburn sont trop nombreuses par rapport à une autre ville.* »
« *On entend par statistiques significatives un nombre très élevé de cas qui ne s'expliquent pas statistiquement, ils sont trop nombreux pour être dus au hasard.* »

Pollution et sex ratio

Niveau

Seconde.

Situation étudiée

Dans un village du Canada situé dans une réserve indienne, on constate qu'entre 1999 et 2003, sur 132 naissances il n'y a eu que 46 garçons. Ce phénomène peut-il raisonnablement être du au seul hasard ou bien faut-il chercher d'autres causes (la proximité d'usines chimiques par exemple) ?

Type d'activité

Exercice en classe ou évaluation.

Durée

20 minutes.

Objectifs

Contenus mathématiques au programme

- Calcul de fréquences.
- Notion d'intervalle.
- Utilisation d'une formule qui permet de préciser la notion de fluctuation d'échantillonnage.

Enjeux citoyens

Comprendre comment une étude statistique peut permettre de détecter une anomalie touchant une certaine population (ici le sex-ratio chez une population vivant dans un réserve située au cœur d'industries chimiques) et par la suite encourager des recherches pour cerner les causes de cette anomalie.
Aborder la notion de « différence significative ».

Compétences et attitudes

Comprendre et savoir utiliser dans le contexte de l'activité les définitions données : sex ratio et « différence significative à 95% ».
Interpréter les calculs effectués : s'interroger…

Organisation

L'activité (évaluation ou exercice) se situe en fin de séquence. La formule donnant l'intervalle de fluctuation doit déjà avoir été travaillée dans le cadre d'une expérimentation.

Description de l’activité

L’énoncé suivant a été donné en devoir surveillé dans une classe de seconde en fin de chapitre de statistique.
Les données proviennent d’une étude effectuée au Canada et montrant une différence (très) significative sur le sex-ratio à la naissance (déficit de garçons) sur une population exposée à une pollution chimique. Dans ce cas particulier, l’inquiétude provient du fait que bien que ces industries canadiennes respectent les normes, une exposition prolongée à de faibles doses de polluants puisse avoir un impact sanitaire mesurable.

Énoncé élève

Le « sex-ratio » est le rapport du nombre de garçons à celui des filles à la naissance. Il est habituellement de 105 garçons pour 100 filles.

1) La fréquence habituelle des garçons est $p = \frac{105}{105+100} = 0{,}512$. Si cette proportion est exactement respectée, combien de garçons et de filles a-t-on sur 1000 naissances ?

2) Si l’on prélève des échantillons aléatoires de taille n dans une population où la fréquence étudiée est $p = 0{,}512$, dans plus de 95 % des cas, la fréquence f observée sera comprise dans l’intervalle $[0{,}512-\frac{1}{\sqrt{n}} ; 0{,}512+\frac{1}{\sqrt{n}}]$.

Si f n’appartient pas à cet intervalle, on dira que f présente une « différence significative » au niveau 0,95 avec $p = 0{,}512$.

a) Dans la réserve indienne d’Aamjiwnaang, située au Canada, il est né entre 1999 et 2003, n =132 enfants dont 46 garçons.
Que vaut la fréquence f des garçons pour cette période à Aamjiwnaang (arrondir à 10^{-3}) ?

b) La fréquence des garçons observée à Aamjiwnaang pour la période 1999-2003 présente-t-elle une « différence significative » avec $p = 0{,}512$ (justifier par un calcul) ?

Sources : Science et Vie février 2006 – Environmenthal Health Perspectives octobre 2005 (article en ligne).

Éléments de réponse

1) Si la proportion $p = 0{,}512$ est exactement respectée, on doit avoir, sur 1000 naissances, 512 garçons et 488 filles.

2) a) La fréquence de garçons f observée à Aamjiwnaang est $f = 46 / 132 = 0{,}348$.
b) La différence est significative lorsque la fréquence observée n’appartient pas à l’intervalle [0,424 ; 0,599].
0,348 n’appartient pas à cet intervalle donc la valeur observée présente une différence significative avec $p = 0{,}512$.

Déroulement et commentaires

Voici quelques exemples de copies d'élèves réalisées en devoir surveillé.

Les deux copies suivantes sont des exemples de réponses « satisfaisantes ».

1 SEX-RATIO

1) Si cette proportion est respecté, il y aura :
- 512 garçons
- 488 filles

... sur 1000 naissance. Oui

2)

a/ Ici le sex-ratio f vaut : $\frac{46}{132} = \boxed{0,348}$

par cette période

Oui

b) $\left[0,512 - \frac{1}{\sqrt{132}} \; ; \; 0,512 + \frac{1}{\sqrt{132}}\right]$ donc

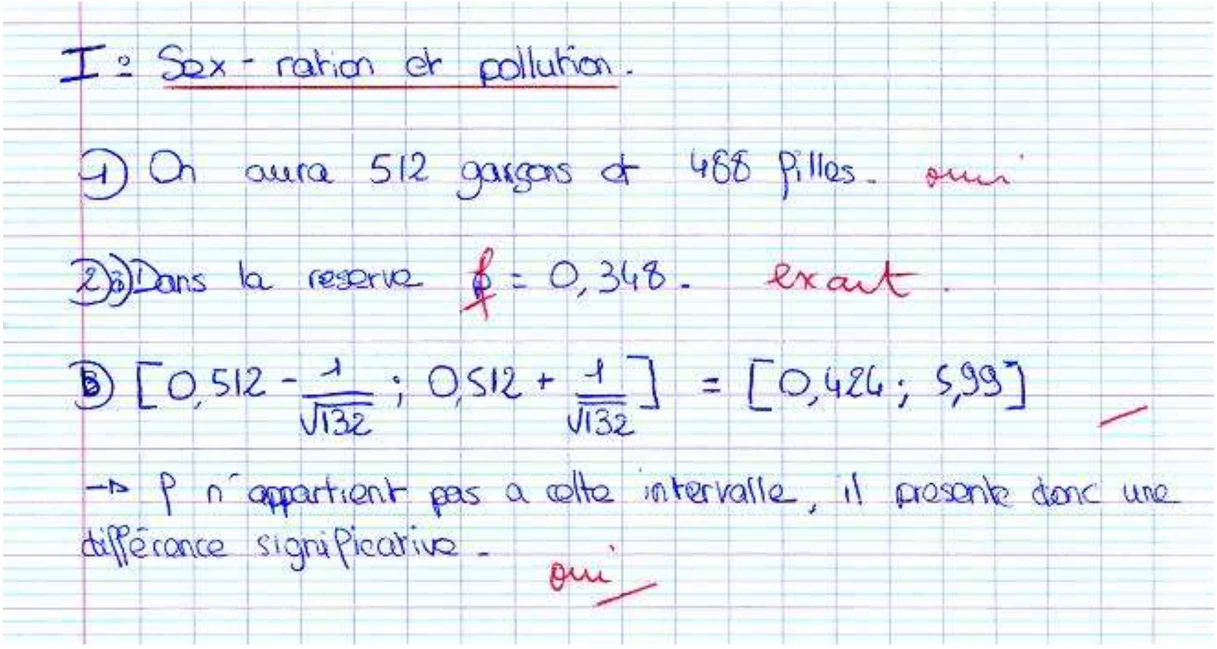
$I\,[0,424 \; ; \; 0,599]$

- Oui il y a donc une différence significative car le resultat obtenu (f normale) est très éloigner du resultat concret de la ville.

(normale [42,4 % et 59,9 %] alors que le resultat à 95 % de chance obtenu est de $\boxed{34,8\ \%}$

Bien

Autre exemple correct :



I : Sex-ration et pollution.

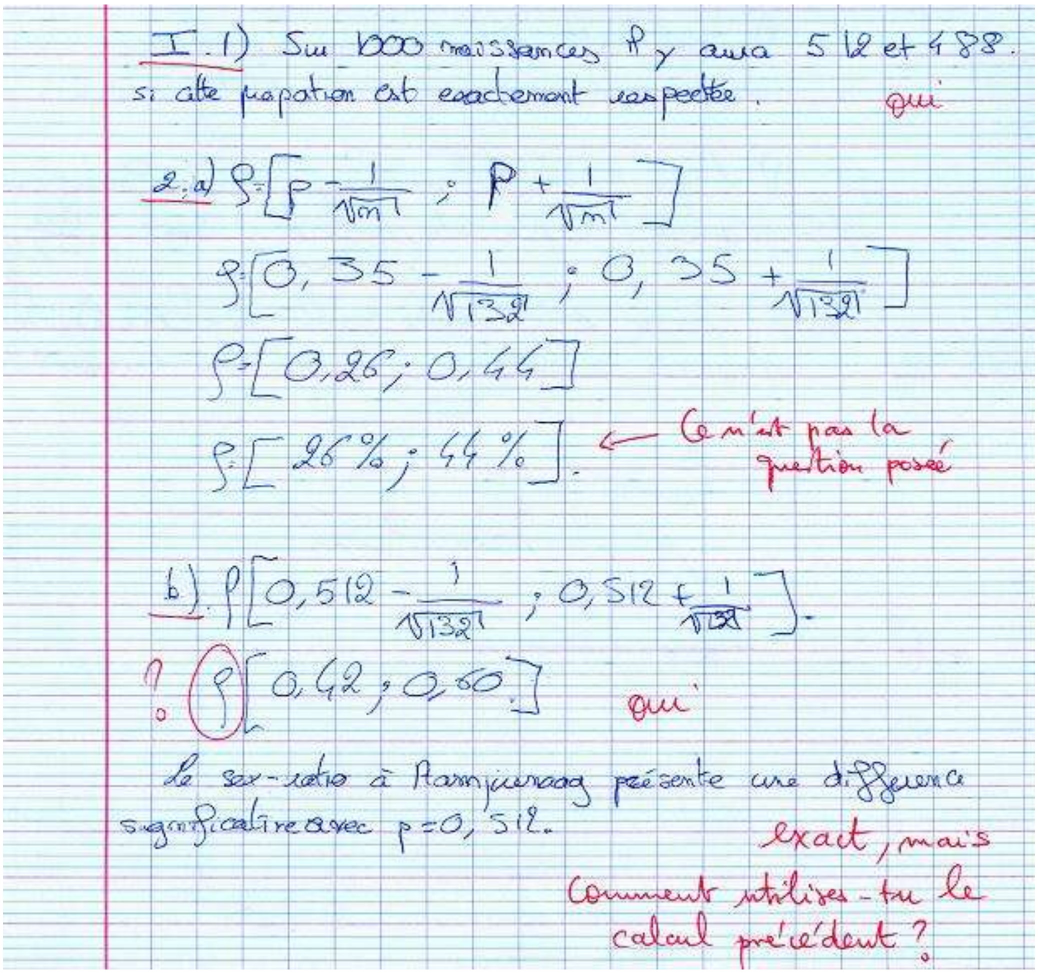
1) On aura 512 garçons et 488 filles. oui

2)a) Dans la reserve $f = 0,348$. exact.

b) $\left[0,512 - \frac{1}{\sqrt{132}} ; 0,512 + \frac{1}{\sqrt{132}}\right] = [0,426 ; 5,99]$

→ f n'appartient pas a cette intervalle, il presente donc une différance significative. oui

Dans la copie suivante, l'élève calcule à la fois un intervalle de confiance autour de la fréquence *f* observée (ce qui n'est pas la démarche demandée même si elle est ici équivalente) et un intervalle de fluctuation autour de la valeur *p* théorique. Les deux intervalles ont une partie commune et la conclusion de l'élève est confuse.



I.1) Sur 1000 naissances il y aura 512 et 488.
si cette proportion est exactement respectée. oui

2.a) $f \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$

$f \in \left[0,35 - \frac{1}{\sqrt{132}} ; 0,35 + \frac{1}{\sqrt{132}}\right]$

$f \in [0,26 ; 0,44]$

$f \in [26\,\% ; 44\,\%]$. ← Ce n'est pas la question posée

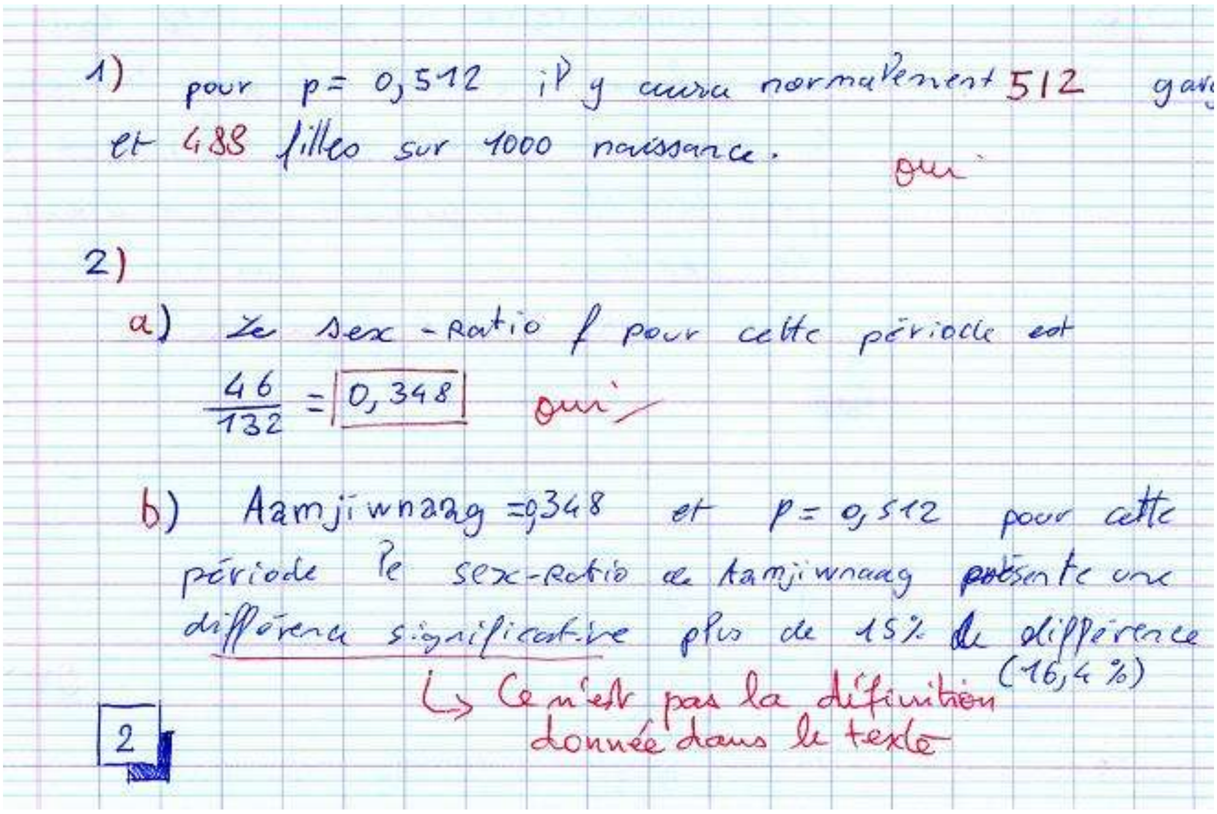
b) $f \left[0,512 - \frac{1}{\sqrt{132}} ; 0,512 + \frac{1}{\sqrt{132}}\right]$.

? $f [0,42 ; 0,60]$ oui

Le sex-ratio à Aamjiwnaang présente une différence significative avec $p = 0,512$.

exact, mais comment utilises-tu le calcul précédent ?

Dans la copie suivante, la « définition » donnée dans l'énoncé de « différence significative » n'est pas respectée. L'élève n'utilise pas, pour conclure, les fluctuations des échantillons de taille 132.



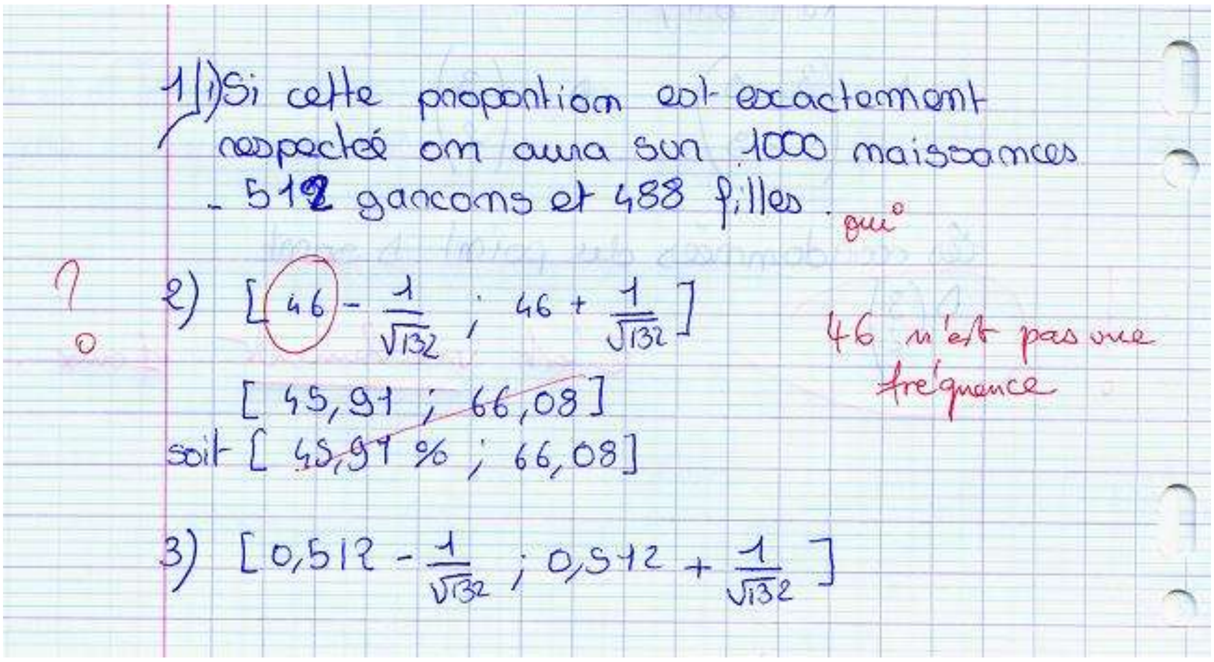
1) pour p = 0,512 il y aura normalement 512 garç[ons]
et 488 filles sur 1000 naissance.
oui

2)
a) Le sex-ratio f pour cette période est
$\frac{46}{132} = \boxed{0,348}$ oui

b) Aamjiwnaag = 0,348 et p = 0,512 pour cette période le sex-ratio de Aamjiwnaag présente une différence significative plus de 15% de différence (16,4 %)
↳ Ce n'est pas la définition donnée dans le texte

2

Dans cette copie, c'est la définition de « fréquence » qui n'est pas maîtrisée.



1) Si cette proportion est exactement respectée on aura sur 1000 naissances
- 512 garçons et 488 filles. oui

?
2) $\left[46 - \frac{1}{\sqrt{132}} ; 46 + \frac{1}{\sqrt{132}} \right]$ 46 n'est pas une fréquence
[45,91 ; 66,08]
soit [45,91 % ; 66,08]

3) $\left[0,512 - \frac{1}{\sqrt{132}} ; 0,512 + \frac{1}{\sqrt{132}} \right]$

Les erreurs les plus souvent rencontrées proviennent d'erreurs de calcul sur les fréquences ou de confusions entre estimation (intervalle centré sur la valeur observée) et test (intervalle d'acceptation de l'hypothèse centré sur la valeur correspondante de la fréquence). Dans l'ensemble, le caractère « anormal » des statistiques d'Aamjiwnaag a été vu et plusieurs élèves ont posé, à la fin du contrôle, la question de savoir s'il s'agissait d'une histoire « vraie ».

Bilan et perspectives

Tel que l'énoncé est présenté ici, il soulève immanquablement la question « Que peut-on tirer comme conclusion de ces statistiques ? ».

Une première version de l'énoncé ajoutait comme précision que la réserve d'Aamjiwnaang est située « à proximité de nombreuses industries chimiques ». Cela laissait entendre qu'il y avait une relation de causalité entre la pollution chimique et le « défaut » observé du sex-ratio, causalité qui n'est pas établie par l'étude statistique.

A la question « Que peut-on tirer comme conclusion ? », on peut seulement ici répondre que « cette étude pose question ». La statistique donne l'alerte, ce qui est déjà beaucoup. Le fait que la réserve soit située au cœur d'industries chimiques devient un élément troublant sur lequel on doit enquêter.

De façon générale, c'est la notion de « preuve statistique » qui est ici en jeu. Il ne s'agit pas d'une « preuve » au sens habituel mais d'un élément probant. Plutôt que de parler de « preuve statistique », les anglo-saxons disent plus justement *piece of evidence*.

D'autres explications possibles du déséquilibre du sex-ratio pourraient être liées au modes de vie de ces indiens ou à leur patrimoine génétique. Une étude statistique comparative a été menée sur des indiens de la même tribu vivant dans un autre environnement et a (dé)montré que ce n'était (sans doute) pas le cas. En revanche l'influence de certains produits chimiques sur le sex-ratio a été établie « statistiquement » par d'autres études.

Une recherche sur Internet permettra d'avoir d'autres éléments sur ce dossier (qui a fait polémique au Canada).

Maladies nosocomiales et facteur de risque

Niveau

Seconde (ou supérieur, en particulier $1^{\text{ère}}$ ST2S).

Situation étudiée

Une enquête nationale de 2001 indique un taux d'infections nosocomiales d'environ 7% des personnes hospitalisées.
On trouve alors ce commentaire sur un site Internet médical : « *Le risque de contracter une infection à l'hôpital est de 7%, c'est à dire que sur 100 personnes hospitalisées, sept d'entre elles auront une infection nosocomiale.* »
L'expérimentation montre que même en supposant que ces 100 personnes sont prises au hasard, on sera sans doute éloigné d'une observation de 7 infections, fluctuations obligent.
Au-delà, il s'agit de réfléchir à la notion de facteur de risque, de différence « significative », au sens que peut avoir, pour l'individu, une statistique générale.

Type d'activité

Exercice avec simulation sur tableur.

Durée

45 minutes.

Objectifs

Contenus mathématiques au programme

Fluctuation d'échantillonnage d'une fréquence.

Enjeux citoyens

Relativiser une statistique à la taille de l'échantillon dont elle provient.
Risques évitables et facteurs de risque.
Notion de différence significative.
Interprétation individuelle d'une statistique générale.

Capacités et attitudes

Expérimentation par simulation.
Esprit critique.

Organisation

Exercice nécessitant une simulation sur tableur (ou calculatrice).
Les élèves doivent savoir simuler 100 tirages avec remise dans une urne bicolore dont les proportions sont connues à l'aide d'une instruction du type =ENT(ALEA()+*p*) sur tableur ou int(rand+*p*) sur calculatrice.
Les élèves ont déjà vu la notion de fluctuation d'échantillonnage dans des contextes plus simples que celui-ci.

Description des activités

Énoncé élève

1. L'enquête nationale sur les maladies nosocomiales de 2001 a montré que le jour de l'étude, sur **305 656** patients étudiés, **21 010** souffraient d'une infection contractée à l'hôpital.
Calculer la fréquence p des infections nosocomiales sur cette enquête (arrondir à 10^{-4}).
2. En commentaire de cette enquête, on a relevé sur un site médical d'Internet l'affirmation suivante :
« *Le risque de contracter une infection à l'hôpital est de 7%, c'est à dire que sur 100 personnes hospitalisées, sept d'entre elles auront une infection nosocomiale.* »
a) Simuler des échantillons aléatoires de taille **100** prélevés dans une population avec un taux d'infection nosocomiale **7%**. Observer le nombre de cas d'infection.
b) Pourquoi faudrait-il dire : « *sur 100 personnes hospitalisées* ***prises au hasard****, sept d'entre elles ,* ***en moyenne****, auront une infection nosocomiale* » ?
3. Par définition, un **facteur de risque** agit en augmentant la fréquence de la maladie chez des sujets qui y sont exposés.
On a les observations suivantes (enquête 2001) :

Catégorie	Effectif	fréquence observée d'infections nosocomiales
Moins de 65 ans	n = **136 804**	f = **4,4** %
Femme	n = **172 096**	f = **6,7** %
Immunodépression	n = **28 106**	f = **13,5** %

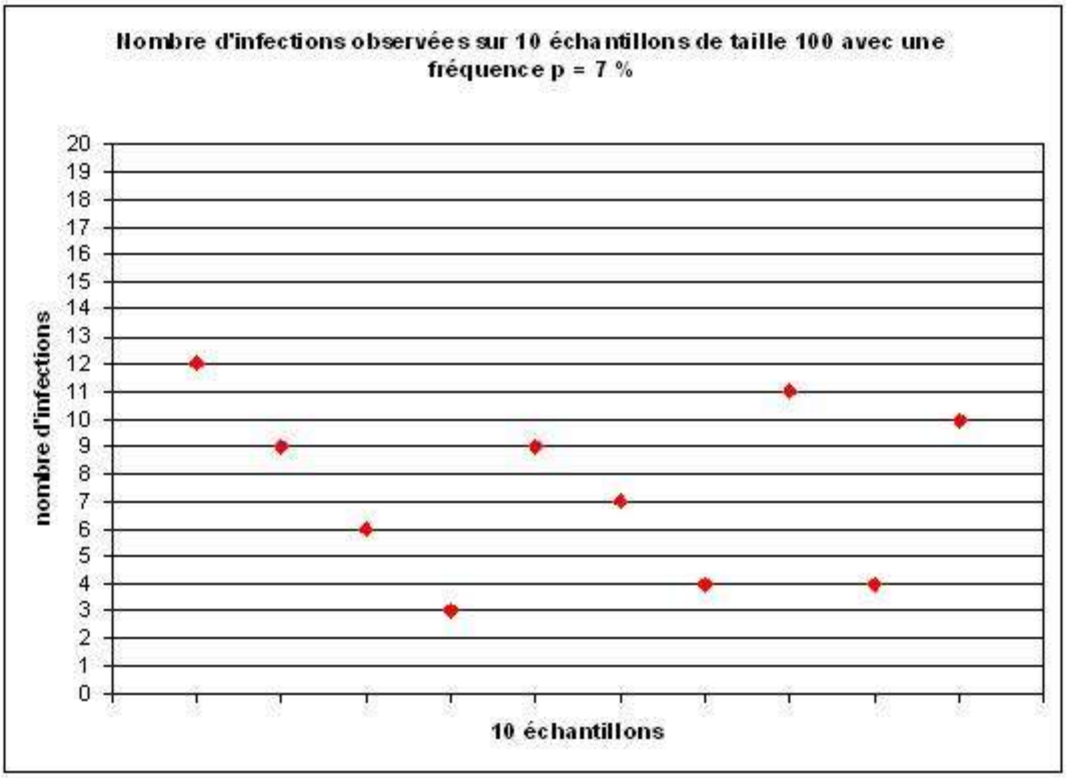
a) Pour chaque taille d'échantillon, calculer l'intervalle de fluctuation aléatoire des fréquences observées sur des échantillons de taille n (au niveau de 95%)

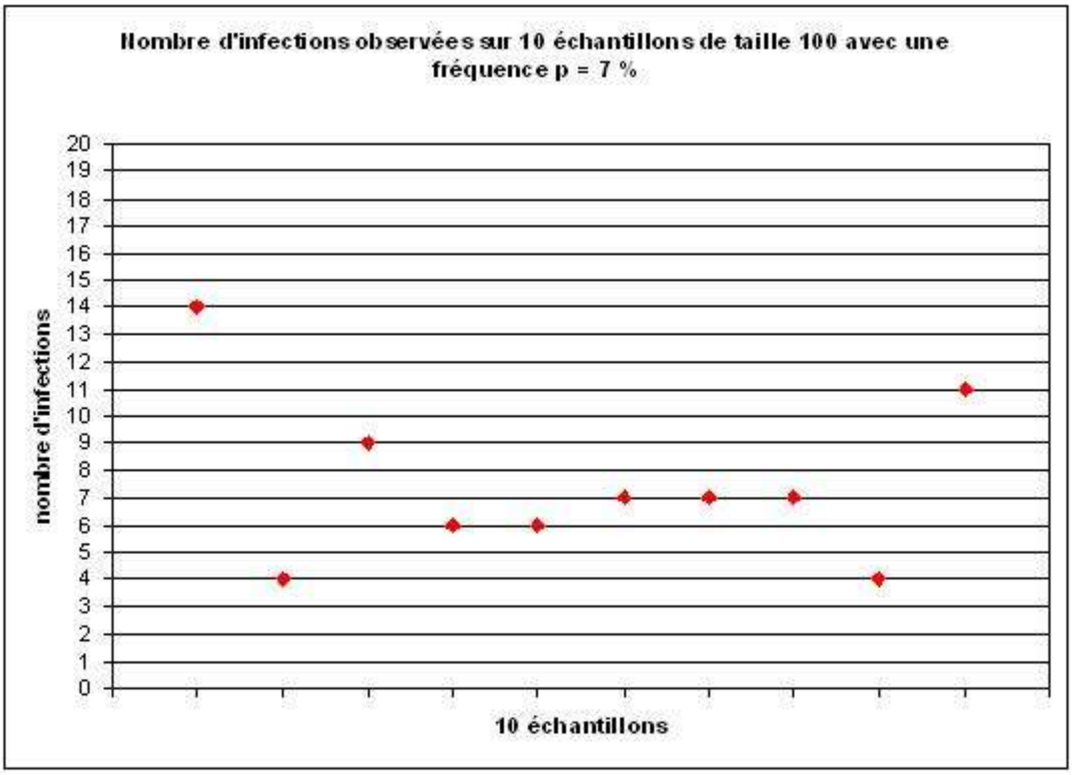
$$\left[\, p - \frac{1}{\sqrt{n}}\,,\ p + \frac{1}{\sqrt{n}}\,\right]$$

avec p = 6,87 % (arrondir les résultats à 10^{-3}).
b) On dira que la différence observée entre f et p est « significative » (au niveau de 95%) lorsque f est en-dehors de l'intervalle calculé précédemment.
Pour chaque catégorie, la différence observée est-elle « significative » ?
4. L'enquête nationale sur les maladies nosocomiales de 2006 a montré que le jour de l'étude, sur **358 467** patients étudiés, **17 820** souffraient d'une infection contractée à l'hôpital.
Calculer la fréquence f des infections nosocomiales sur cette enquête (arrondir à 10^{-4}).
A-t-on une diminution « significative » par rapport à la valeur p calculée à la question 1. avec les résultats de l'enquête de 2001 ?
5. Si vous entrez à l'hôpital, avez-vous 5% de « risque » de contracter une infection nosocomiale ?

Éléments de réponse

1. On a $p \approx 0{,}0687 = 6{,}87$ %.
2. a) Pour effectuer la simulation d'un échantillon aléatoire de taille 100 sous l'hypothèse p = 7%, il suffit, sur un tableur de recopier vers le bas 100 fois l'instruction =ENT(ALEA()+0,07) puis de faire la somme des 0 et des 1.
On peut, sous la forme d'un « nuage de points » (c'est le plus simple) afficher par exemple 10 échantillons de taille 100. On constate que le nombre de cas fluctue énormément.
Voici deux images d'écran.





Entre nous...

La valeur de n est ici un peut petite pour utiliser la formule $p \pm \frac{1}{\sqrt{n}}$ du programme de seconde, majorant l'intervalle des fluctuations dans 95% des échantillons. Les élèves seraient surpris d'obtenir un nombre négatif : [–0,3 ; 0,17] (entre – 3 et 17 cas). La formule, plus précise, fondée sur une approximation par la loi normale,

$p \pm 1{,}96\sqrt{\dfrac{p \times (1-p)}{n}}$, donne entre 2 et 12 cas dans 95% des échantillons, ce qui est plus conforme aux simulations.

b) L'affirmation « sur 100 personnes 7 seront infectées » est donc particulièrement fâcheuse puisque ce n'est pas le cas le plus souvent rencontré dans les simulations.
En revanche, il vrai qu'en moyenne (sur un nombre assez important d'échantillons), les simulations donnent 7 cas d'infections.
Il faudrait ajouter « prises au hasard dans la population française », c'est ce que l'on a simulé. Il est clair (du moins c'est ce que l'on voit dans la suite de l'exercice) que dans un service particulier, ou dans un hôpital particulier, on peut ne pas être « conforme » à ces statistiques nationales.

3. a) Les intervalles de fluctuation des fréquences (dans plus de 95% des cas) sous l'hypothèse $p = 0{,}0687$ sont :

- Moins de 65 ans : $[0{,}066 \,;\, 0{,}071]$.
- Femme : $[0{,}066 \,;\, 0{,}071]$.
- Immunodépression : $[0{,}063 \,;\, 0{,}075]$.

b) On constate que :

- Moins de 65 ans :
  $f = 0{,}044$ n'appartient pas à l'intervalle $[0{,}066 \,;\, 0{,}071]$, l'écart est significatif.
- Femme :
  $f = 0{,}067$ n'appartient pas à l'intervalle $[0{,}066 \,;\, 0{,}071]$ mais de peu, l'écart est peu significatif.
- Immunodépression :
  $f = 0{,}135$ % n'appartient pas à l'intervalle $[0{,}063 \,;\, 0{,}075]$, l'écart est significatif.

4. On procède, par exemple, comme à la question précédente.

On a $\left[p - \frac{1}{\sqrt{n}},\ p + \frac{1}{\sqrt{n}}\right] = \left[0{,}0687 - \frac{1}{\sqrt{358467}} \,;\, 0{,}0687 + \frac{1}{\sqrt{358467}}\right]$.

C'est-à-dire environ $[0{,}0670 \,;\, 0{,}0704]$.

En 2006 on a $f = \dfrac{17820}{358467} \approx 0{,}0497$.

La baisse est (bien sûr) significative.

5. C'est une question « piège » dont on se demande si elle a un sens. Mais c'est une question importante, à laquelle on « risque » d'être confronté !
L'essentiel est de comprendre que la réponse dépend des informations dont on dispose. On n'est pas dans le cas d'une personne prise au hasard dans la population des malades. Les réponses apportées dépendent dans quel état on entre à l'hôpital, dans quel service, éventuellement dans quel hôpital...

Déroulement et commentaires

Sources :
Le site médical sur lequel la phrase « incorrecte » a été relevée est www.caducee.net .
Les résultats des enquêtes nationales de prévalence des infections nosocomiales (2001 et 2006) sont accessibles sur le site de l'Institut de veille sanitaire (utiliser le moteur de recherche) [http ://www.invs.sante.fr/](http://www.invs.sante.fr/) .

Lorsque les élèves sont un peu entraînés à l'interprétation des simulations, la première partie de l'activité ne pose pas de problèmes. La suite est un peu plus délicate si l'on veut dépasser le simple niveau des calculs, mais l'enjeu citoyen est important.

Bilan

La surveillance statistique des taux d'infection permet de mieux prendre conscience des risques et de tout faire pour les réduire. On constate en particulier l'effet positif de l'enquête de 2001 en termes de sensibilisation aux résultats significativement en progrès de 2006.
En revanche, certains risques resteront inévitables et, surtout, le risque dépend fortement de la population à laquelle appartient le malade. Lorsqu'un citoyen s'interroge sur les risques d'infections nosocomiales qu'il encourre, il doit, autant que possible, relativiser ce risque selon son cas (type de service, type d'hôpital, type d'intervention, âge, état de santé...), la statistique globale de 5% (en 2006) ayant peu de sens vu l'hétérogénéité des cas.

Crues et notion de probabilité

Niveau

Énoncé n° 1 : troisième (programme 2008), seconde, première.
Énoncé n° 2 : terminale S.

Situation étudiée

Le magazine *Le Point* du 13/09/2002 écrit : « Une crue de la Seine comparable à celle de 1910 se produit en moyenne tous les cent ans et la probabilité d'une telle catastrophe augmente d'année en année ».
Une crue de la Seine analogue à celle de 1910, qualifiée de « centennale », a chaque année, par définition, une chance sur 100 de se produire.
Après expérimentation des fluctuations d'échantillons de taille 100 (un siècle), on reconsidèrera l'affirmation du journaliste.

Type d'activité

Travaux dirigés et expérimentation sur tableur.

Durée

50 minutes.

Objectifs

Contenus mathématiques au programme

Énoncé n° 1 : moyenne, notion de probabilité, fluctuations d'échantillonnage.
Énoncé n° 2 : loi binomiale.

Enjeux citoyens

Notion de probabilité et d'indépendance pour des phénomènes rares à l'échelle humaine.
Notion de risque.
Notion de modèle.
Comprendre les débats à propos des risques naturels.

Capacités et attitudes

Esprit critique. Remise en cause d'une opinion première.
Expérimentation par simulation : savoir mettre en place une simulation et en interpréter les résultats.

Organisation

Les élèves doivent savoir simuler la réalisation d'un événement ayant une probabilité p de se réaliser : tirage aléatoire dans une urne bicolore ayant une composition connue.
Pour l'énoncé n°2, en terminale S, les élèves doivent connaître la loi binomiale.

Description des activités

Énoncé n° 1 (seconde ; troisième en adaptant – fournir la simulation)

Énoncé élève

Le magazine *Le Point* du 13/09/2002 écrit : « Une crue de la Seine comparable à celle de 1910 se produit en moyenne tous les cent ans et la probabilité d'une telle catastrophe augmente d'année en année ».

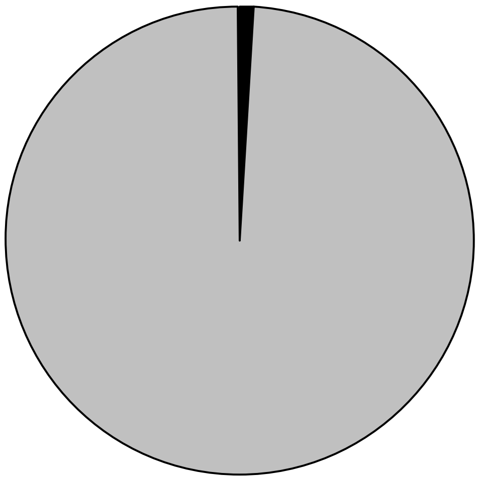
1. Une crue de la Seine analogue à celle de 1910 est qualifiée de « centennale ». La dernière crue centennale de la Seine s'est produite en 1910.
Pensez-vous que cette année, Paris :

- ☐ a **plus** de risque de subir la crue centennale qu'il y en avait en 1911 ;
- ☐ a **moins** de risque de subir la crue centennale qu'il y en avait en 1911 ;
- ☐ n'a **ni plus ni moins** de risque de subir la crue centennale qu'il y en avait en 1911 ;
- ☐ qu'on ne peut pas dire.

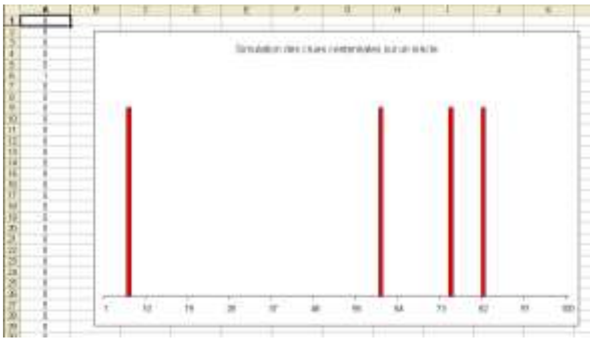
2. Par définition, une crue « centennale » a, chaque année, une chance sur 100 de se produire.
a) Si l'on voulait simuler l'arrivée d'une crue centennale à l'aide d'une roue de loterie, comment devrait-on partager cette roue ?
Montrer que l'instruction =ENT(ALEA()+0,01) permet de simuler sur un tableur l'arrivée d'une crue centennale.
b) Sur une feuille de calcul, simuler en colonne A les crues centennales survenant durant un siècle. Réaliser une représentation graphique en utilisant le « type histogramme ».
Appuyer sur F9 pour effectuer de nombreuses simulations.
– A-t-on dans la grande majorité des simulations, une et une seule crue centennale par siècle ?
– Dans un siècle, avez-vous observé plusieurs crues centennales à quelques années d'intervalle ?
c) Souhaitez-vous modifier votre réponse à la question 1 ? Si oui, comment ?
3. a) Faire la somme des 0 et des 1 de la colonne A. À quoi correspond cette somme ?
b) Après sélection de la colonne A, la recopier vers la droite jusqu'à la colonne IP pour simuler 250 siècles.
Calculer la moyenne du nombre de crues centennales par siècle et effectuer plusieurs fois F9. Que constatez-vous ?
4. Qu'est-ce qui est faux dans la phrase du journaliste citée au début ?
Comment expliquez-vous son erreur ?

Éléments de réponse

1. Opinion « a priori ». On répond « comme on le sent ».
2. a) Il faut partager la roue de loterie en deux secteurs dont l'un fait 3,6°.



L’instruction =ENT(ALEA()+0,01) donne 0 avec 99 chances sur 100 et 1 avec 1 chance sur 100. Il suffit de convenir que la valeur 1 correspond à la crue centennale.
b) On constate sur les simulations (voir les images d’écran suivantes) que :
– le cas d’une et une seule crue centennale par siècle est loin d’apparaître dans la grande majorité des simulation ;
– il se produit assez souvent des regroupement de 2 ou 3 crues centennales à quelques années consécutives.



c) La seule « bonne réponse » est la troisième : « ni plus ni moins de risque », par définition de « crue centennale ».
Remarque pour le professeur :
La réponse « on ne peut pas dire » n’est pas exacte. À partir du moment où l’on parle de crue « centennale » on est dans un modèle qui suppose en fait l’indépendance. On peut utiliser comme image qu’une roue de loterie n’a pas de mémoire.
En revanche il est vrai qu’on ne peut pas dire que ce modèle est exact (en fait tous les modèles sont faux...).
3.a) On calcule le nombre de crues centennales sur un siècle.
b) On constate que la moyenne du nombre de crues par siècle est très proche de 1. C’est une moyenne sur 250 siècles.
c) On peut considérer que le début de la phrase est vrai (en moyenne tous les cent ans), si l’on admet le modèle de la crue centennale (la roue de loterie n’a pas de mémoire).

En revanche la fin de la phrase est faux : la probabilité n'augmente pas du fait que cela fait longtemps que l'événement ne s'est pas produit. Il y a, par définition, indépendance du passé.
4. La notion de moyenne, associée ici à la crue centennale, n'est pas perceptible à l'échelle humaine et ne prend de sens que sur des centaines de siècles. Le journaliste n'a pas compris l'indépendance supposée du phénomène par rapport au passé.

Commentaire

Ce qu'il faut retenir à propos de la notion de « crue centennale » c'est « une chance sur 100 », sans mémoire, comme si l'on tournait une roue de loterie.

Énoncé n° 2 (terminale S)

Le magazine *Le Point* du 13/09/2002 écrit : « Une crue de la Seine comparable à celle de 1910 se produit en moyenne tous les cent ans et la probabilité d'une telle catastrophe augmente d'année en année ».

1. Une crue de la Seine analogue à celle de 1910 est qualifiée de « centennale ». La dernière crue centennale de la Seine s'est produite en 1910.
Pensez-vous que cette année, Paris :

☐ a **plus** de risque de subir la crue centennale qu'il y en avait en 1911 ;

☐ a **moins** de risque de subir la crue centennale qu'il y en avait en 1911 ;

☐ n'a **ni plus ni moins** de risque de subir la crue centennale qu'il y en avait en 1911 ;

☐ qu'on ne peut pas dire.

2. Par définition, une crue « centennale » a, chaque année, une chance sur 100 de se produire.
a) Montrer que l'instruction =ENT(ALEA()+0,01) permet de simuler sur un tableur l'arrivée d'une crue centennale.
b) Sur une feuille de calcul, simuler en colonne A les crues centennales survenant durant un siècle. Réaliser une représentation graphique en utilisant le « type histogramme ».
Appuyer sur F9 pour effectuer de nombreuses simulations.
– A-t-on dans la grande majorité des simulations, une et une seule crue centennale par siècle ?
– Dans un siècle, avez-vous observé plusieurs crues centennales à quelques années d'intervalle ?
c) Souhaitez-vous modifier votre réponse à la question 1 ? Si oui, comment ?
3. a) Faire la somme des 0 et des 1 de la colonne A. À quoi correspond cette somme ?
b) Après sélection de la colonne A, la recopier vers la droite jusqu'à la colonne IP pour simuler 250 siècles.
Calculer la moyenne du nombre de crues centennales par siècle et effectuer plusieurs fois F9. Que constatez-vous ?
4. On désigne par X la variable aléatoire qui, pour un siècle au hasard, donne le nombre de crues centennales.
a) Montrer que X suit une loi binomiale dont on donnera les paramètres.
b) Quel calcul peut justifier l'expression « en moyenne tous les cent ans » ?
5. a) Calculer à l'aide du tableur $P(X = 0)$; $P(X = 1)$; $P(X = 2)$; $P(X = 3)$.

b) Justifier que dans la majorité des cas, on n'a pas une crue centennale par siècle.
6. Qu'est-ce qui est faux dans la phrase du journaliste citée au début ?
Comment expliquez-vous son erreur ?

Éléments de réponse

4. a) On a la répétition de 100 épreuves aléatoires indépendantes (car chaque année la probabilité d'une crue est la même) avec deux issues possibles, crue centennale de probabilité 0,01 ou non, où X correspond au nombre total de crues.
Donc X suit la loi binomiale de paramètres $n = 100$ et $p = 0{,}01$.
b) On a $E(X) = 100 \times 0{,}01 = 1$.
5. a) L'instruction =LOI.BINOMIALE(A2;100;0,01;FAUX) fournit les résultats suivants.

k	P(X=k)
0	0,36603234
1	0,36972964
2	0,18486482
3	0,06099917

b) On a $P(X \neq 1) \approx 0{,}63$.

Bilan

Dans une revue technique de juin 2003[1], des « spécialistes », ingénieurs et hydrologues, insistent sur le fait « *qu'informer les citoyens sur les risques d'inondation par des messages clairs et compréhensibles est un enjeu social et économique fort mais complexe* ». Seule l'expérimentation par simulation permet de rapidement se rendre compte que sur l'échelle d'un siècle, les réalisations d'un événement ayant chaque année 1 chance sur 100 de se produire conduit à des observations très variables, et que la « moyenne », une fois tous les 100 ans, n'a de sens que pour un grand nombre de siècles.
Maintenant, il est possible que le modèle de la crue centennale (qui suppose l'indépendance) ne soit pas adapté à la situation... Il s'agit cependant d'un modèle simple réellement utilisé dans ce genre de situation.

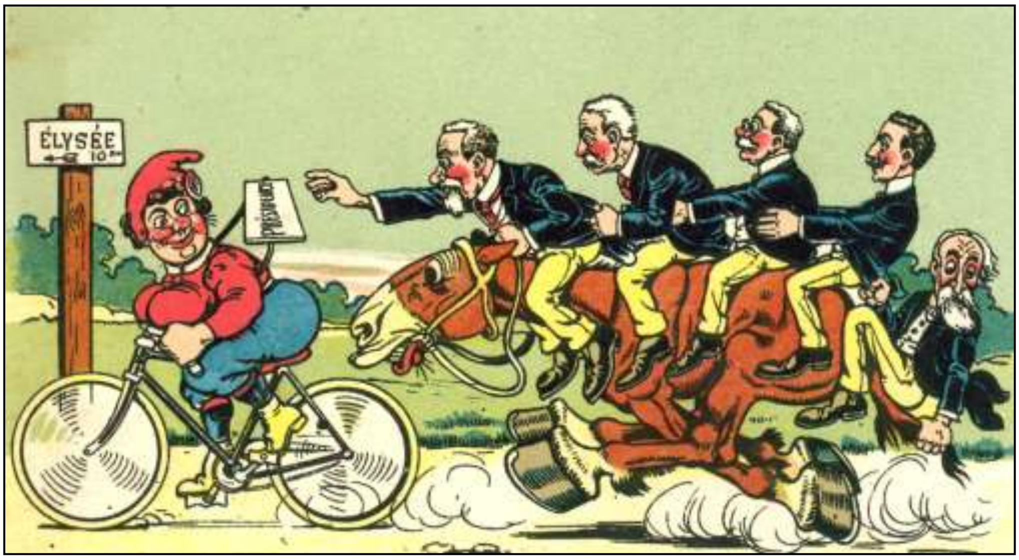
[1] Revue *Ingénieries – eau, agriculture, territoires* n°34 juin 2003, article intitulé *Risque d'inondation : une notion probabiliste complexe pour le citoyen* de N. Gendreau, F. Grelot, R. Garçon et D. Duband.

« Document à conserver » (jusqu'en 2010 ?) distribué par la mairie de Paris
pour se préparer à la prochaine crue centennale.

À supposer que le risque est le même qu'il l'était au début du XX^{e} siècle, on peut cependant affirmer, que les dégâts, le coût et la désorganisation seraient bien supérieurs.

3

Sondages



Petit lexique des sondages politiques

Niveau

Document d'information et de réflexion pour le professeur (mais lisible par un élève de terminale).

Situation étudiée

Peut-on croire un sondage politique ?

Objectifs

Contenus mathématiques

Échantillonnage aléatoire.
Intervalle de confiance.
Loi des grands nombres, loi binomiale, loi normale.

Enjeux citoyens

Les sondages sont un élément omniprésent de notre vie politique.
Quelle est la place des mathématiques dans les sondages ? Comprendre les notions mathématiques intervenant.
En quoi notre enseignement peut-il permettre d'éduquer le citoyen dans ce domaine ?
Que se cache-t-il derrière les mots « quotas », « stratifié », « représentatif », « 95% de confiance » ... ?

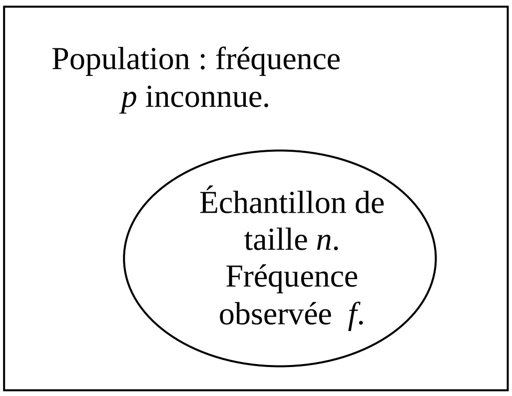
Lexique des sondages... et méthodologie

On aurait pu intituler ce paragraphe, selon un titre paru dans la presse :

« Dans les cuisines des sondeurs ».

Sondage aléatoire simple

C'est le type de « sondage » actuellement enseigné au lycée, dans le cadre d'un « thème d'étude » en classe de seconde générale, ou dans le chapitre « intervalles de confiance » de certaines sections de BTS (on effectue des sondages aléatoires simples dans l'industrie, pour le contrôle de qualité par exemple).



Il s'agit de tirer au hasard n éléments dans une population où la fréquence p d'un caractère est inconnue (par exemple le pourcentage p d'électeurs en faveur d'un candidat).

L’expression « au hasard » signifie que chaque échantillon de taille n a la même probabilité d’être tiré.
On peut facilement appliquer la théorie des probabilités à ce type de sondage.
En supposant que le tirage est effectué avec remise (on peut faire cette hypothèse si la taille n du sondage est faible devant celle de la population), on a la situation du schéma de Bernoulli.
Si n est « assez grand », l’approximation de la loi binomiale par la loi normale conduit à un « intervalle de confiance » : si on observe la fréquence f sur l’échantillon de taille n, on démontre que la fréquence correspondante inconnue p dans la population est située dans l’intervalle $\left[f - 1{,}96\sqrt{\dfrac{p(1-p)}{n}} \,,\; f + 1{,}96\sqrt{\dfrac{p(1-p)}{n}} \right]$ avec 95 % de « confiance ».
Cette expression signifie que sur un grand nombre d’échantillons de taille n, dans environ 95 % des cas, p est effectivement dans l’intervalle ci-dessus.
Comme p est inconnu, en remarquant que pour tout p de l’intervalle [0, 1], on a $1{,}96\sqrt{\dfrac{p(1-p)}{n}} \leq \dfrac{1{,}96}{2\sqrt{n}} \leq \dfrac{1}{\sqrt{n}}$, on peut majorer l’intervalle de confiance à 95 % en donnant une « fourchette » à plus de 95 % de confiance sous la forme :
$$\left[f - \frac{1}{\sqrt{n}} \,,\; f + \frac{1}{\sqrt{n}} \right].$$
C’est cette « fourchette » qui peut être expérimentée en seconde, par simulation.

Sondage par quotas

C’est la méthode pratiquée par les instituts de sondages français (voir les encarts méthodologiques publiés dans la presse, on a un exemple ci-dessous).
Cette méthode ne contient rien d’aléatoire (du moins maîtrisé) et par conséquent sa fiabilité ne peut être mathématiquement calculée, puisqu’on ne peut pas utiliser le calcul des probabilités. La fiabilité de la méthode des quotas n’est qu’empirique, fondée sur « l’expérience des sondages précédents ».


Méthodologie

- Etude réalisée auprès d’un échantillon de 1 006 personnes, représentatif de la population française âgée de 18 ans et plus.
- L’échantillon a été constitué selon la méthode des quotas, au regard des critères de sexe, d’âge, de catégorie socioprofession-nelle, après stratification par région et taille de la commune.
- Les interviews ont été conduites du 1er au 2 février 2007, par téléphone au domicile des personnes interrogées.

En étant assez optimiste, on peut considérer que la méthode des quotas conduit à une marge d’erreur, à 95 %, de l’ordre de celle d’un sondage aléatoire simple, c’est-à-dire environ $\dfrac{1}{\sqrt{n}}$, soit, pour n = 1000 personnes interrogées, $\dfrac{1}{\sqrt{1000}} \approx 3$ %.
C’est la raison pour laquelle, même si elle n’est pas pratiquée par les instituts de sondages français, l’enseignement des mathématiques de la méthode aléatoire est instructif (et cette méthode est utilisée dans les contrôles de qualité).

☺ Voir l’**exercice 1 « Fourchettes pour la cuisine des sondages »** des activités suivantes.

La méthode des quotas consiste à choisir un certain nombre de critères jugés importants pour le sujet du sondage : sexe, âge, catégorie socioprofessionnelle, région, taille de la commune..., puis à calculer le pourcentage de personnes appartenant à chaque catégorie selon les données du recensement.
Il s'agit alors d'obtenir autant de réponses que chaque quota ainsi calculé pour un échantillon de taille *n*. Pour atteindre les quotas de chaque catégorie, il ne s'agit pas de tirages au sort organisés de façon à maîtriser les probabilités (c'est beaucoup plus économique que d'interroger des personnes réellement au hasard). Évidemment des biais existent (que l'on peut chercher à corriger de façon plus ou moins empirique...), en particulier parce que répondent les personnes joignables qui veulent bien répondre. C'est un peu comme si un biologiste voulant tester un nouveau produit sur une souris le faisait sur la première souris qu'il peut attraper dans la cage. Il y a toutes les chances pour que cette souris soit la plus faible de toutes, la moins vive.

Sondage aléatoire stratifié (proportionnel / optimal)

Ces méthodes permettent d'améliorer la fiabilité du sondage aléatoire simple. Elles demeurent cependant des méthodes aléatoires, avec toutes les exigences d'un tirage « au hasard » et ne sont donc pas comparables à la méthode des quotas, malgré certaines confusions parfois entretenues (en particulier par les termes de « stratification » ou de « représentatif »). On suppose que pour toutes les personnes de la population, on peut avoir accès aux informations correspondant aux critères sélectionnés (sexe, âge, catégorie socioprofessionnelle, région, taille de la commune...). On effectue alors une partition de la population selon les critères retenus.
Un échantillon aléatoire stratifié proportionnel sera obtenu par tirage au sort dans chaque sous-ensemble (« strate ») de la population, en quantités proportionnelles aux effectifs de chaque sous-ensemble. On montre, qu'à taille d'échantillon égale, la précision peut-être considérablement améliorée par rapport à un sondage aléatoire simple, en particulier si les proportions, pour le caractère étudié, sont très différentes selon les strates.
Un échantillon aléatoire stratifié optimal tient compte de la dispersion selon les strates de la variable faisant l'objet de l'enquête. On gagne en précision en abandonnant la simple proportionnalité et en interrogeant davantage dans les strates à forte dispersion que dans celles très homogènes.

Marge d'erreur

Pour un sondage portant sur 1000 personnes, on parle parfois de « marge d'erreur de plus ou moins 3 % ». L'expression « marge d'erreur » pourrait laisser croire qu'au delà de cette marge, on a la certitude de ne pas trouver le pourcentage réel que l'on cherche à estimer. Ce qui est faux. Il vaudrait mieux parler de marge « d'incertitude à 95 % de confiance ».
Rappelons que pour la méthode des quotas, il est impossible d'évaluer sérieusement, c'est-à-dire mathématiquement, la marge d'incertitude. Pour un sondage aléatoire simple de 1000 personnes, la marge d'incertitude à 95 % de confiance, à partir d'une fréquence f calculée sur le sondage, est de plus ou moins $1{,}96\sqrt{\frac{p(1-p)}{1000}}$, que l'on peut approcher par $1{,}96\sqrt{\frac{f(1-f)}{1000}}$. Le tableau suivant donne quelques calculs :

Fréquence f calculée sur un sondage	10% ou 90%	20% ou 80%	30% ou 70%	40% ou 60%	50%
Marge d'incertitude à 95% de confiance pour un sondage aléatoire de taille 1000.	1,86%	2,48%	2,84%	3,04%	3,10%

On constate qu'évaluer la marge d'incertitude à 3% (à 95% de confiance) est bien adapté pour des fréquences observées entre 30% et 70%. Pour de petites ou de fortes fréquences, cette marge de 3% est une majoration parfois importante de l'incertitude.
A noter que si f est trop petite ou trop grande, la formule précédente cesse d'être valable, il faut avoir $n \times f$ et $n \times (1 - f)$ au moins supérieurs à 5 pour appliquer raisonnablement l'approximation d'une loi binomiale par une loi normale.

Échantillonnage

L'échantillonnage est la manière dont est constitué l'échantillon. Il faut distinguer l'échantillonnage aléatoire de celui qui ne l'est pas (comme dans la méthode des quotas). Avec un échantillonnage aléatoire, on peut utiliser le calcul des probabilités et donc estimer l'incertitude (encore faut-il ne pas avoir de non réponses, de fausses déclarations, d'enquêteurs peu sérieux...). Avec un échantillonnage non aléatoire, on ne peut pas estimer le risque d'erreur.
Il ne faut pas confondre « aléatoire » et « aveugle ». Quand on interroge les gens au téléphone par la méthode des quotas, on procède de façon « aveugle », donc avec une certaine dose d'aléatoire mais sans savoir laquelle ni dans quel sens. Quand on parle de méthode « aléatoire », il s'agit d'un « hasard complet », maîtrisé de sorte à s'assurer que chaque individu de la population a les mêmes chances d'être interrogé. Il faut pour cela une procédure très contrôlée : par exemple numéroter tous les individus et tirer les numéros au sort selon un procédé dont on sait qu'il respecte l'équiprobabilité, par exemple un générateur de nombres aléatoires.

Échantillon représentatif

Voilà une expression qui, si elle n'est pas précisée, peut signifier à peu près n'importe quoi !
Un échantillon constitué selon la méthode des quotas est évidemment « représentatif » des critères correspondants aux quotas (sexe, âge, catégorie socioprofessionnelle, région, taille de la commune...) selon lesquels il a été fabriqué. Mais on n'a aucun moyen de savoir jusqu'à quel point il est « représentatif » de ce pour quoi il a été prélevé, c'est-à-dire le sujet du sondage, l'opinion, le pourcentage que l'on cherche à évaluer. L'expression « représentatif de la population française », que l'on lit souvent dans la presse, prête évidemment à confusion. On a l'impression que l'échantillon est « représentatif » de tout ce que l'on veut.
En statistique, on désigne plutôt par « échantillon représentatif », un échantillon où le hasard permet d'éviter les biais inconnus et d'appliquer le calcul des probabilités. La méthode optimale pour obtenir un échantillon « représentatif » est celle du sondage aléatoire stratifié optimal.

Taux de réponse

La plupart des sondages sont effectués par téléphone. Dans ce cadre, Michel Lejeune évoque dans son article un taux de réponse de l'ordre de 10% à 20%. Avec un tel taux de non-réponses, le biais est sans doute non négligeable. Qui répond ? Qui refuse de répondre ? Le taux de non réponse n'est sans doute pas le même dans les différentes catégories d'opinion.

☺ Voir l'**exercice 2 « Non réponses »** des activités suivantes.

Défaut de couverture

C'est un autre biais important. La population sondée est-elle la population visée ? Si le sondage est effectué par Internet, s'il l'est par téléphone pendant les heures de travail, ... ce n’est certainement pas le cas. De toutes manières, des pans entiers de la populations sont hors d’atteinte.

Fausses déclarations

C’est une source importante de biais pour des questions sensibles et souvent difficile à évaluer. Il existe des méthodes d’interrogation aléatoire, ou de recoupement avec d’autres questions.

☺ Voir l’**exercice 3 « Éviter les fausses déclarations grâce au hasard »** des activités suivantes.

Redressement de l’échantillon

Il s’agit souvent de méthodes empiriques consistant à « corriger » certains biais constatés lors des études analogues précédentes. Par exemple, en 2002, certains sondeurs ont expliqué « redresser » les intentions de votes pour J.-M. Le Pen en les multipliant par 2. Comme on le constate, c’est une technique mathématique assez élémentaire...

☺ Voir l’**exercice 4 « Où l’on suspecte la méthode des quotas »** des activités suivantes.

Fourchettes, non réponses, fausses réponses et redressements... : la cuisine mathématique des sondages

Niveau

Exercice 1 : $3^{ème} – 2^{nde}$.
Exercice 2 : $3^{ème} – 2^{nde}$.
Exercice 3 : Seconde ou première.
Exercice 4 : Terminale S.

Situations étudiées

• Exercice 1 : Calcul des fourchettes du dernier sondage avant le premier tour de l'élection présidentielle de 2002. Confrontation avec les résultats du premier tour.

• Exercice 2 : Un sondage d'intentions de vote comporte un nombre important de « non-réponses ». On suppose que des enquêtes ont montré que 50% des électeurs d'extrême droite refusent de répondre. Comment utiliser cette « information » pour affiner les estimations du sondage ?

• Exercice 3 : De façon à éviter les fausses réponses, on souhaite préserver l'anonymat des sondés. Grâce au pile ou face, et à la loi des grands nombres permettant de supposer que l'on a à peu près autant de pile que de face, il sera possible d'exploiter globalement les réponses alors qu'on ignore le sens individuel de chacune.

• Exercice 4 : Pourquoi les résultats des sondages à propos de Lionel Jospin au premier tour de la présidentielle de 2002, « prouvent » qu'ils ne sont pas effectués de manière « indépendante » ?

Type d'activité

Exercices (en classe ou à la maison).

Durée

30 minutes par exercice.

Objectifs

Contenus mathématiques au programme

Exercice 1 : Pourcentages, ordre des nombres, notion d'intervalle et d'intersection.

Exercice 2 : Pourcentages, tableau à double entrée (éventuellement).

Exercice 3 : Notion de probabilité, pourcentage (fréquence), équation du premier degré, arbre de probabilité (éventuellement).

Exercice 4 : Loi binomiale, indépendance.

Enjeux citoyens

Montrer en quoi des notions mathématiques de base sont fondamentales pour analyser l'information.
Il s'agit non seulement de montrer en quoi les mathématiques permettent d'éviter les pièges ou les manipulations, mais aussi d'illustrer leur rôle positif pour mieux comprendre les situations.

Capacités et attitudes

Faire preuve d'esprit critique.
Lire un texte et rechercher l'information.
Illustrer ou organiser des données.

Organisation

Exercices utilisables en travaux dirigés ou en devoir.

Description des activités

Exercice 1 : Fourchettes pour la « cuisine » des sondages

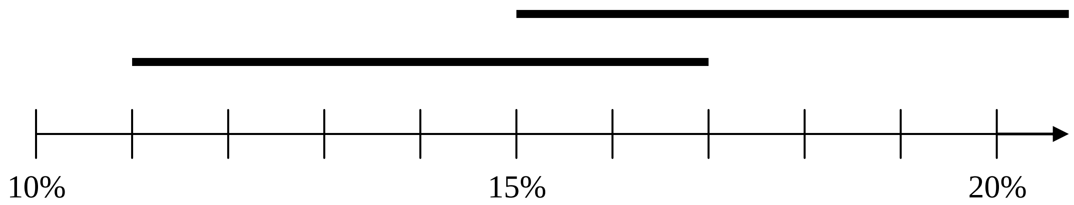
Énoncé élève

Voici un extrait d'article, publié dans le journal « Le Monde » par le statisticien Michel Lejeune, après le premier tour de l'élection présidentielle de 2002.
« *Pour les rares scientifiques qui savent comment sont produites les estimations, il était clair que l'écart des intentions de vote entre les candidats Le Pen et Jospin rendait tout à fait plausible le scénario qui s'est réalisé. En effet, certains des derniers sondages indiquaient 18 % pour Jospin et 14 % pour Le Pen. Si l'on se réfère à un sondage qui serait effectué dans des conditions idéales* [...], *on obtient sur de tels pourcentages une incertitude de plus ou moins 3 % étant donné la taille de l'échantillon* [...]. »

1. Si l'on tient compte de l'incertitude liée au sondage, entre quels pourcentages pourraient se situer réellement (à 95% de confiance) les deux candidats lorsque le sondage donne 18% pour l'un et 14% pour l'autre ?
2. Représenter sur un même graphique les deux « fourchettes » calculées à la question précédente. Peut-on prévoir l'ordre des candidats ?
3. Au premier tour de l'élection présidentielle de 2002, L. Jospin a obtenu 16,18% des voix et J.-M. Le Pen 16,86%.
Expliquer la phrase « *l'écart des intentions de vote entre les candidats Le Pen et Jospin rendait tout à fait plausible le scénario qui s'est réalisé* ».

Éléments de réponse

1. Pour L. Jospin, entre 15% et 21%. Pour J.-M. Le Pen, entre 11% et 17%.
2. Un dessin possible.



Si on utilise ces fourchettes, on ne peut pas prévoir l'ordre des candidats car elles ont une partie commune.
3. La phrase correspond au fait que les pourcentages obtenus à l'élection sont situés dans les fourchettes du sondage.

Commentaires

Voici un exemple où quelques notions mathématiques de base (pourcentage, représentation des nombres, intervalle, intersection) sont nécessaires à la bonne compréhension d'un article de presse de la rubrique société ou politique.

La lecture du texte est une des difficultés de l'exercice mais montre aussi l'intérêt de la situation.

L'aspect fluctuation d'échantillonnage (plus ou moins 3% sur un échantillon de taille mille, c'est-à-dire $\frac{1}{\sqrt{1000}} \approx 0{,}03$) n'est pas pris en compte dans cet exercice, qui se veut élémentaire. On fait « confiance » au statisticien. On pourra, en classe de seconde, expérimenter la pertinence de ce 3% (« à 95% de confiance grosso-modo) par simulation (voir l'activité présentée à ce propos au paragraphe suivant).

Exercice 2 : Non réponses

Énoncé élève

On a interrogé 1000 personnes auxquelles on demandait de répondre par « oui » ou par « non » à la question « à la prochaine élection, pensez-vous voter pour le candidat de l'extrême droite ? »

Les résultats sont les suivants :

Oui : **98** ; Non : **717** ; Non réponses : **185** ; Total : **1000**.

1. Quel est le pourcentage de « oui », par rapport aux personnes interrogées et par rapport aux répondants ?
2. On suppose que des études précédentes ont montré que parmi les électeurs d'extrême droite, **50%** refusent de répondre.

En admettant que ce pourcentage est ici respecté et qu'il n'y a aucune fausse déclaration, calculer le pourcentage d'électeurs d'extrême droite parmi les 1000 personnes interrogées.

3. Quel est, en utilisant le résultat de la question précédente, le pourcentage de non-réponses parmi les électeurs ne votant pas pour l'extrême droite ?

Éléments de réponse (c'est le cas de le dire...)

1. Il y a **9,8%** de « oui » par rapport aux 1000 personnes interrogées et $\frac{98}{98+717} \approx$ **12%** par rapport aux répondants.

2. Si l'on admet qu'il n'y a pas de fausses réponses, il y a 98 électeurs d'extrême droite qui répondent. En supposant que parmi les électeurs d'extrême droite 50% répondent et 50% ne répondent pas, il y a donc 98 + 98 = 196 électeurs d'extrême droite. Ce qui représente **19,6%** des 1000 personnes.
3. D'après la question précédente il y a 1000 – 196 = 804 électeurs ne votant pas pour l'extrême droite dont 717 ont répondus. Il y a donc 804 – 717 = 87 électeurs ne votant pas pour l'extrême droite qui n'ont pas répondu. (On peut faire aussi 185 – 98 = 87).
Le taux de non réponse chez ces électeurs est donc de $\frac{87}{804} \approx 10{,}8\%$.

Remarque : il peut être utile de visualiser la situation à l'aide d'un tableau à double entrée :

	électeurs d'extrême droite	non électeurs d'extrême droite	total
réponses	**98**	**717**	815
non réponses	98	87	**185**
total	196	804	**1000**

Commentaires

L'exercice montre bien le caractère « relatif » de l'information « pourcentage » : pourcentage par rapport à quoi ? Sur une certaine question, des pourcentages différents peuvent tous être justes s'ils n'ont pas la même base de calcul.
Du point de vue citoyen, il montre l'impact éventuellement important du phénomène des non-réponses sur l'image (déformée) que le sondage peut donner de la réalité.

Exercice 3 : Éviter les fausses réponses grâce au hasard

Énoncé élève

On souhaite interroger 1000 personnes sur leur intention de voter pour l'extrême droite.
Pour éviter les fausses déclarations, l'enquêteur préservera l'anonymat des réponses en proposant de réaliser en son absence la procédure suivante :

Lancer une pièce.

- *Si elle tombe sur pile, répondre par « oui » ou par « non » à la question : « avez-vous l'intention de voter pour l'extrême droite ? ».*
- *Si elle tombe sur face, relancer la pièce, si elle tombe sur pile, répondre « oui », si elle tombe sur face, répondre « non ».*

De cette façon, l'enquêteur ne peut pas savoir à quoi correspond la réponse « oui » ou « non » qu'on lui transmet.
L'enquête donne 338 réponses « oui » et 662 réponses « non ».

1. On suppose qu'il y a autant de pile que de face aux premiers lancers, comme aux seconds lancers. Combien y en a-t-il dans chaque cas ?
2. On désigne par x la fréquence des électeurs d'extrême droite, montrer que x vérifie l'équation $500\,x + 250 = 338$.
3. Calculer la fréquence des électeurs d'extrême droite.

Éléments de réponse

1. Il y a 1000 premiers lancers avec 500 faces et 500 piles.
Il y a 500 seconds lancers avec 250 piles et 250 faces.
2. Les 338 réponses « oui » correspondent aux électeurs d'extrême droite parmi les 500 qui ont fait pile au premier lancer, soit $500 \times x$, auxquels s'ajoutent les 250 personnes qui ont fait pile au second lancer. D'où $500\,x + 250 = 338$.

3. On en déduit que $x = \dfrac{338-250}{500} = 17{,}6\%$.

Remarque 1 : bien sûr, il n'y aura pas exactement autant de pile que de face, mais l'exercice montre qu'on peut « estimer » la proportion x sans rien connaître des réponses individuelles.

Remarque 2 : en classe de première, on peut proposer cet exercice dans le cadre des probabilités, à résoudre par exemple à l'aide d'un arbre.

Commentaires

Cet exercice montre bien comment le « hasard » (quand on en connaît une loi adaptée à la situation – ici, l'équiprobabilité du pile ou face) peut être un « allié » du statisticien. Situation paradoxale mais fréquente, qui permet d'éviter certains biais.

La méthode exposée dans cet exercice n'est pas utilisée (à notre connaissance) pour les sondages politiques. En revanche, ce type de méthode est utilisé aux États-Unis pour des enquêtes sur des sujets touchant à la vie privée ou à des questions sensibles (avortement, sexualité...).

Exercice 4 : Où l'on suspecte la méthode des quotas

Énoncé élève

A la suite du premier tour de l'élection présidentielle de 2002, le statisticien Michel Lejeune a mis en cause la méthode des quotas, utilisée en France pour effectuer les sondages, en affirmant :

« *Les sept derniers sondages publiés donnaient tous Jospin à 18 % : pour tout observateur avisé, cette constance est statistiquement invraisemblable avec des échantillons de taille 1000*. »

1. On suppose que le jour des sondages, il y avait effectivement 18% des électeurs en faveur de L. Jospin. On note X la variable aléatoire qui à tout tirage au hasard de 1000 électeurs associe le nombre de ceux en faveur de L. Jospin (la population étant très importante, on supposera que les tirages sont indépendants).

Quelle est la loi suivie par X ?

2. Les résultats des sondages sont donnés à 0,5% près. Pour quelles valeurs de X arrondira-t-on le résultat du sondage à 18% ?

3. A l'aide d'un tableur, on a calculé ci-dessous les probabilités $P(X \leq k)$ pour $k = 184$ et $k = 174$.

Quelle est la probabilité de l'événement : « le sondage affiche un résultat à 18% pour L. Jospin » ?

B1 = =LOI.BINOMIALE(184;1000;0,18;VRAI)

	A	B	C	D	E
1	P(X <= 184)	0,647			
2	P(X <= 174)	0,328			
3	Différence	0,319			
4					

4. En supposant que les 7 sondages ont été réalisés indépendamment et dans les mêmes conditions, quelle est la probabilité qu'ils donnent le même résultat de 18% ?

Éléments de réponse

1. La variable aléatoire X suit la loi binomiale de paramètres $n = 1000$ et $p = 0{,}18$.
2. Le résultat du sondage est arrondi à 18% pour X compris entre 175 et 184.
3. On a $P(175 \leq X \leq 184) = \mathrm{P}(X \leq 184) - P(X \leq 174) \approx 0{,}319$.

4. La probabilité que les 7 sondages donnent le même résultat est $(0{,}319)^7 \approx 0{,}0003$.
On en déduit que les sondages n'étaient sans doute pas effectués de façon indépendante.

Commentaires

Il s'agit d'une mise en situation « réelle » des connaissances de probabilité, assez élaborées, de terminale.

L'exemple montre que les sondages effectués selon la méthode des quotas (et redressements...) ne peuvent prétendre aux mêmes « bonnes » propriétés que les sondages aléatoires. En particulier, il n'y a aucune raison pour que l'indépendance des sondages soit respectée : les instituts utilisent les mêmes historiques et plus ou moins les mêmes « recettes » pour effectuer leurs « redressements ». Or le citoyen a le sentiment qu'il s'agit de sondages indépendants, c'est-à-dire qu'une accumulation de sondages allant dans le même sens confortent notre opinion. Ce serait le cas pour des sondages aléatoires où 10 sondages de taille 1000 valent pour un sondage de taille 10 000. Ici, ce n'est pas le cas.

Présidentielles 2002

Niveau

Seconde.

Situation étudiée

En 2002 , Jean-Marie Le Pen accède au second tour de l'élection présidentielle alors que les sondages précédant l'élection annonçaient un duel entre Lionel Jospin et Jacques Chirac.

Type d'activité

Activité n°1 : exercice en classe.
Activité n°2 : travaux pratiques sur tableur.

Durée

Activité n°1 : 30 minutes.
Activité n°2 : 1 heure.

Objectifs

Contenus mathématiques au programme

- Simulation avec un tableur.
- Notion de fluctuation d'échantillonnage.
- Notion de fourchette de confiance (thème d'étude).
- Travail sur les intervalles.
- Représentations graphiques.

Enjeux citoyens

Comprendre les sondages et s'interroger sur leur qualité.

Compétences et attitudes

Utiliser le tableur pour modéliser une situation concrète.
Comprendre et savoir utiliser une formule.
Représenter graphiquement des résultats.
Interpréter ses résultats.
Mobiliser ses connaissances pour dégager la conclusion citoyenne de l'activité.

Organisation

Le chapitre de statistique descriptive a déjà été traité. Cette activité s'insère dans la séquence sur la simulation. Le second exercice nécessite un travail préalable sur tableur indispensable à la compréhension des formules utilisées =(ENT(100*ALEA()+1)).

Description des activités

L'objectif de ces deux activités est de montrer la qualité toute relative de l'information fournie par un sondage. L'attitude de l'opinion vis à vis des sondages est souvent sans nuance : on leur prête des pouvoirs de prédiction qu'ils n'ont pas (en omettant souvent de fournir les « fourchettes ») et (ou) on déclare qu'ils se trompent 9 fois sur 10.

Activité 1 : exercice

Énoncé élève

Lors du premier tour des élections présidentielles, le dernier sondage publié par l'institut B.V.A. , effectué sur 1000 électeurs le vendredi 19/04/02, prévoyait :

Jacques Chirac	**19 %**
Lionel Jospin	**18 %**
Jean-Marie Le Pen	**14 %**

La surprise a été grande le dimanche 21/04/02 au vu des résultats, puisque Jean-Marie Le Pen figurait au second tour :

Jacques Chirac	**19,88 %**
Lionel Jospin	**16,18 %**
Jean-Marie Le Pen	**16,86 %**

1) On peut rappelle que la formule des fourchettes de sondage à plus de 95 % de confiance, calculée à partir d'une fréquence f obtenue sur un échantillon aléatoire de taille 1000 est :
$[\ f - \frac{1}{\sqrt{1000}}, f + \frac{1}{\sqrt{1000}}\ ]$.
Calculer les trois fourchettes à partir du sondage B.V.A. et les représenter sur un graphique.
2) En se basant sur ces fourchettes, peut-on « prévoir » l'ordre des candidats au premier tour de l'élection ?
3) Placer sur le même graphique les résultats du premier tour.
Doit-on considérer que le dernier sondage B.V.A. était « faux » ?

Éléments de réponse

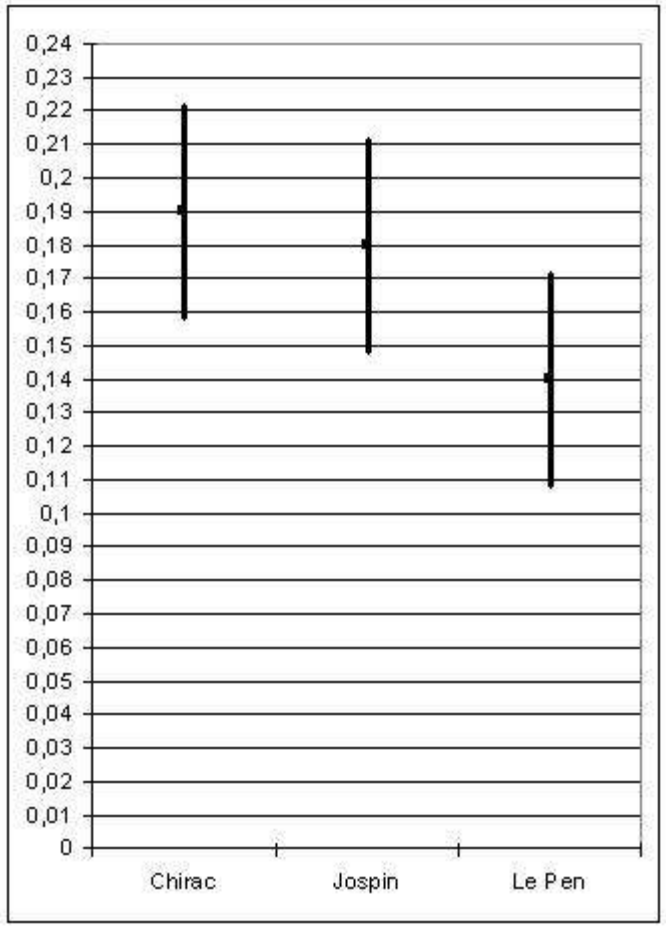
1) A partir du dernier sondage B.V.A., le calcul des fourchettes donne les estimations suivantes :

Jacques Chirac : $[\ 0{,}19 - \frac{1}{\sqrt{1000}}\ ;\ 0{,}19 + \frac{1}{\sqrt{1000}}\ ]$ soit environ [15, 8 % ; 22,2 %].

Lionel Jospin : $[\ 0{,}18 - \frac{1}{\sqrt{1000}}\ ;\ 0{,}18 + \frac{1}{\sqrt{1000}}\ ]$ soit environ [14,8 % ; 21,2 %].

Jean-Marie Le Pen : $[\ 0{,}14 - \frac{1}{\sqrt{1000}}\ ;\ 0{,}14 + \frac{1}{\sqrt{1000}}\ ]$ soit environ [10,8 % ; 17,2 %].

La représentation graphique de ces trois fourchettes est la suivante :

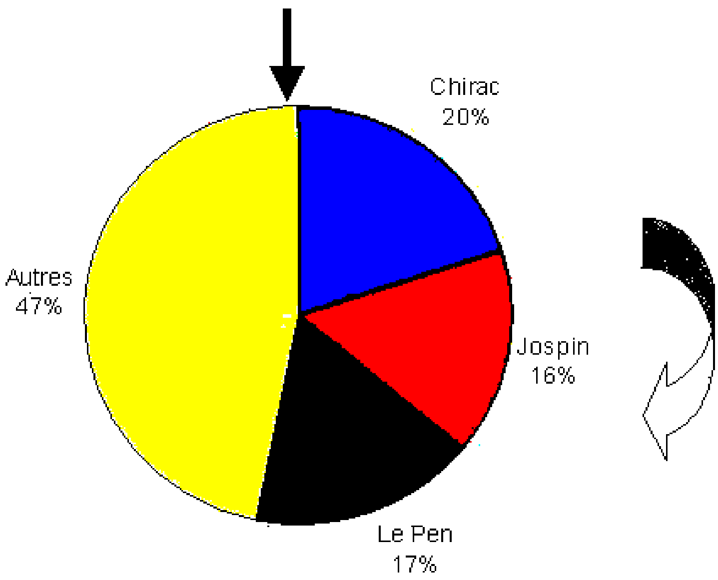


2) Ces fourchettes se chevauchent et ne permettent pas de prévoir l'ordre des candidats.
3) Le sondage est-il « faux », comme on a pu le lire dans la presse par la suite ?
Première réponse de statisticien, qui semble botter en touche, aucun sondage n'est jamais ni « vrai » ni « faux ». Cette logique n'est pas dans l'esprit de la statistique inductive (énervant). Les fourchettes sont à « 95 % de confiance », c'est à dire que la procédure est grosso modo fiable à 95 %.
Si l'on prend les chiffres du premier tour, on constate effectivement que les fourchettes sont « vraies » en ce sens qu'elles contiennent la valeur qu'elles sont censées estimer :
19,88 % $\in$ [15, 8 % ; 22,2 %] ;
16,18 % $\in$ [14,8 % ; 21,2 %] ;
16,86 % $\in$ [10,8 % ; 17,2 %].

Activité 2 : TP sur tableur

Énoncé élève

Même un sondage de taille 1000 le jour du premier tour de l'élection présidentielle de 2002, « sorti des urnes », aurait été assez indécis, compte tenu des résultats particulièrement serrés des candidats en deuxième et troisième position. On peut utilement expérimenter de tels sondages, par simulation, sur le tableur.
Le jour de l'élection, la structure de l'électorat correspond au « camembert » ci-dessous. La simulation d'un sondage aléatoire sur 1000 personnes consiste à faire tourner 1000 fois cette roue de loterie.



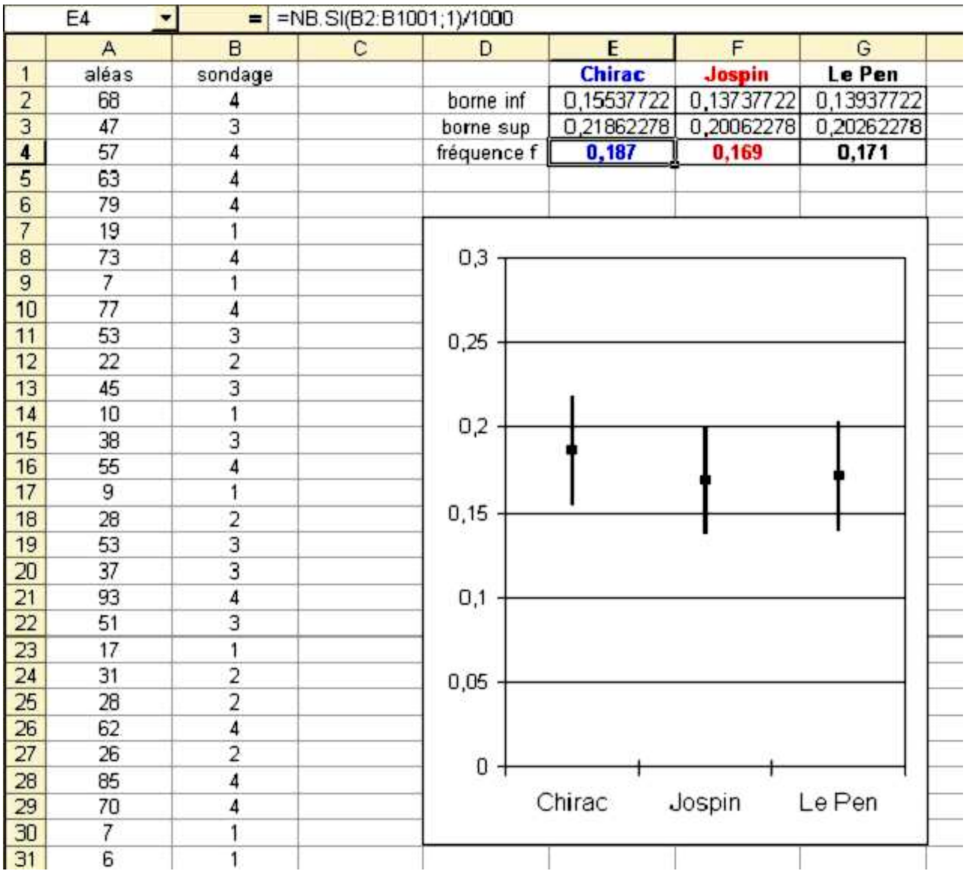
L'instruction ENT(1+100*ALEA()) simule le tirage au hasard d'un nombre entre 1 et 100.
Il suffit de convenir que si le nombre tiré est entre 1 et 20, il s'agit d'un électeur de Jacques Chirac, que s'il est entre 21 et 36, il s'agit d'un électeur de Lionel Jospin et que s'il est entre 37 et 53, il s'agit d'un électeur de Jean-Marie Le Pen.
Ainsi, les proportions 20 %, 16 % et 17 % sont respectées. Pour effectuer ce tri, on utilisera l'instruction :
SI(condition ; action si la condition est remplie ; action si la condition n'est pas remplie) .
Sur une feuille de calcul, entrer, comme ci-dessous, en A2 la formule
=ENT(1+100*ALEA()) et en B2 la formule =SI(A2<21;1;SI(A2<37;2;SI(A2<54;3;4))) .

B2 = =SI(A2<21;1;SI(A2<37;2;SI(A2<54;3;4)))

	A	B	C	D	E
1	aléas	sondage			
2	33	2			
3					

Sélectionne les deux cellules A2 et B2 puis les recopier vers le bas jusqu'à la ligne 1001 pour simuler un sondage de taille 1000.
Le calcul des trois fourchettes du sondage sera effectué dans les cellules de E2 à G4.
Entrer en E4 la formule =NB.SI(B2:B1001;1)/1000 qui calcule la fréquence *f* pour Jacques Chirac (on a convenu qu'il correspond aux valeurs 1 de la colonne B) puis en E2 et E3 les formules =E4-1/RACINE(1000) et =E4+1/RACINE(1000) fournissant les bornes de la fourchette de sondage à 95 % pour Jacques Chirac.

Procéder de même pour Lionel Jospin (valeurs 2 de la colonne B) et Jean-Marie Le Pen (valeurs 3 de la colonne B).
Après avoir sélectionné les cellules de E2 à G4 cliquer sur l'icône de l'assistant graphique pour demander un graphique « boursier » qui permet de visualiser les trois fourchettes, comme sur l'image ci-dessous.


E4 = =NB.SI(B2:B1001;1)/1000

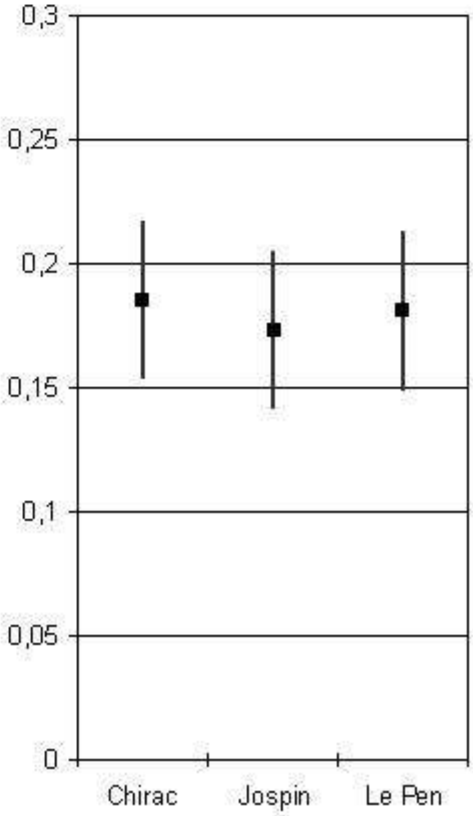
	A	B	C	D	E	F	G
1	aléas	sondage			Chirac	Jospin	Le Pen
2	68	4		borne inf	0,15537722	0,13737722	0,13937722
3	47	3		borne sup	0,21862278	0,20062278	0,20262278
4	57	4		fréquence f	0,187	0,169	0,171
5	63	4					
6	79	4					
7	19	1					
8	73	4					
9	7	1					
10	77	4					
11	53	3					
12	22	2					
13	45	3					
14	10	1					
15	38	3					
16	55	4					
17	9	1					
18	28	2					
19	53	3					
20	37	3					
21	93	4					
22	51	3					
23	17	1					
24	31	2					
25	28	2					
26	62	4					
27	26	2					
28	85	4					
29	70	4					
30	7	1					
31	6	1					

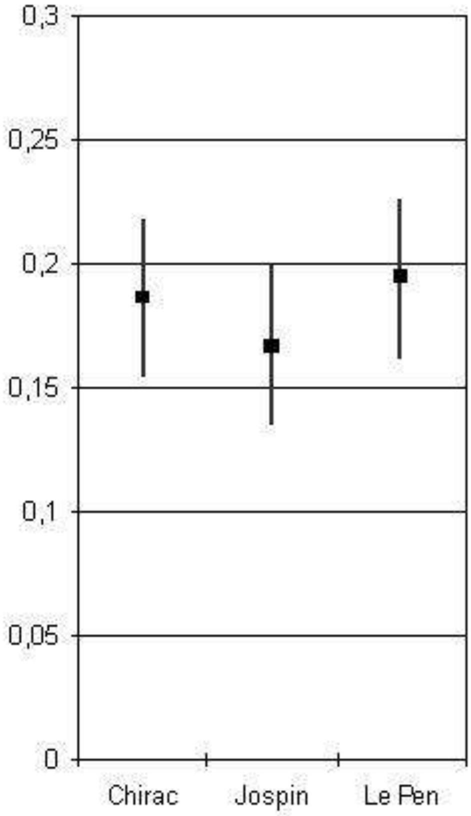
Il suffit ensuite de faire plusieurs fois F9 pour observer de nombreux sondages de taille 1000.

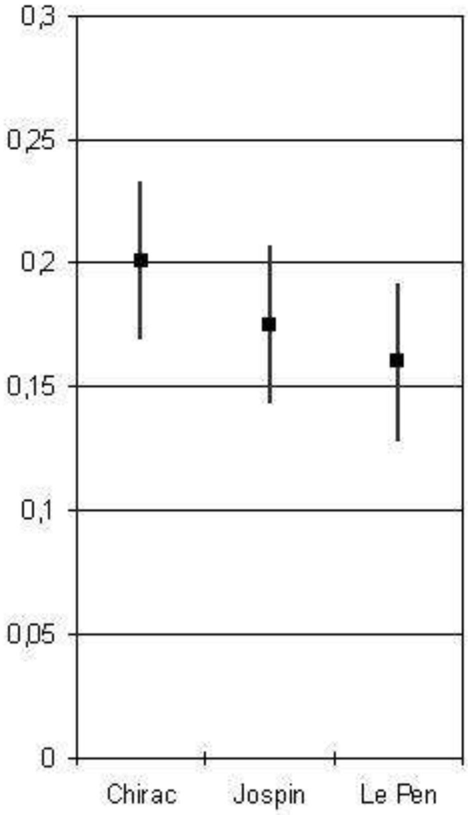
Les sondages simulés prévoient-il qui sera au premier rang le jour de l'élection ?
Observez-vous des sondages analogues au dernier sondage BVA qui donnait l'ordre Chirac-Jospin-Le Pen ?

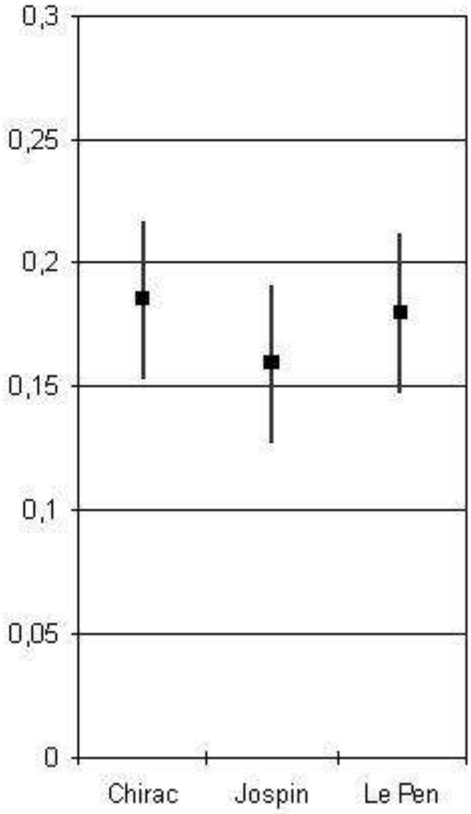
Éléments de réponse

On s'aperçoit que ces sondages ont beaucoup de difficultés pour distinguer les candidats en deuxième et troisième position, que les fluctuations des résultats sont importantes et que parmi eux figurent des sondages analogues à celui de B.V.A. le vendredi précédent le premier tour de 2002. Voici les résultats de quatre sondages de taille 1000 simulés le jour de l'élection (tels que les montre le tableur). On mesure le caractère instable de la situation...


0,3
0,25
0,2
0,15
0,1
0,05
0
Chirac
Jospin
Le Pen


0,3
0,25
0,2
0,15
0,1
0,05
0
Chirac
Jospin
Le Pen


0,3
0,25
0,2
0,15
0,1
0,05
0
Chirac
Jospin
Le Pen


0,3
0,25
0,2
0,15
0,1
0,05
0
Chirac
Jospin
Le Pen

Bilan

L’exemple de cette élection, qui ne laisse pas les élèves indifférents, suscite bien des questions. La formule de fourchette du programme de seconde est-elle valable pour les sondages électoraux ? Grosso modo, oui, même si les méthodes de sondage sont différentes de celles d’un pur sondage aléatoire, les incertitudes sont à peu près les mêmes. L’opinion n’a-t-elle pas évoluée entre le vendredi, jour de parution du sondage, et le dimanche, jour de l’élection ? Sans doute un peu mais pas au point de dire que le sondage était complètement erroné. Reste la question principale. Pourquoi les journalistes ne donnent-ils que rarement les fourchettes ? Par manque de prudence, pour aller plus vite, pour faire plus simple... Mais on est alors proche de la désinformation.

Élections et participation : la simulation pour encourager la participation

Niveau

Seconde.

Situation étudiée

Lors d'une élection, deux candidats s'affrontent. Le plus populaire a-t-il gagné d'avance ?
On tente d'anticiper les résultats d'une élection à partir de la donnée de la popularité des deux candidats et de la participation de leurs supporters.

Type d'activité

Séance de « module » (demi groupes) utilisant le tableur.

Durée

1 heure.

Objectifs

Contenus mathématiques au programme

- Simulation avec un tableur.
- Notion de fluctuation d'échantillonnage.
- Travail sur les intervalles et la notion de partie entière.
- Notion de fourchette de confiance.
- Calcul de la moyenne d'un groupe à partir de la moyenne de deux sous groupes.

Enjeux citoyens

Montrer l'importance décisive de se rendre aux urnes malgré la prolifération des sondages.

Capacités et attitudes

Utiliser le tableur pour modéliser une situation concrète. Interpréter ses résultats.
Mobiliser ses connaissances théoriques lors de l'analyse quantitative pour dégager la conclusion citoyenne de l'activité.

Organisation

Le chapitre de statistiques descriptives a déjà été traité. Cette activité s'insère dans la séquence sur la simulation. Il s'agit de la séance n°3 après un travail préalable sur l'utilisation des fonctions random et partie entière de la calculatrice pour simuler une expérience aléatoire.

Description des activités

Énoncé élève (niveau seconde)

Impact de la participation sur le résultat d'une élection municipale

But de l'activité : Utiliser la simulation d'évènements aléatoires pour montrer l'importance de se rendre aux urnes et analyser quantitativement le phénomène étudié.

I. Simulation sur tableur

Au deuxième tour des élections municipales de la ville de Nouillorc, les mille électeurs ont à choisir entre Ella Gagnet et Ilan Perdhu. On admet que sur les mille électeurs, 520 préfèrent Ella Gagnet contre 480 pour Ilan Perdhu.

1. A priori, qui devrait devenir le prochain maire de Nouillorc ? Justifier très brièvement.

2. On suppose que chaque supporter d'Ella Gagnet a 60% de chance de se rendre aux urnes. Comment utiliser les fonctions ALEA et ENT pour simuler qu'un supporter d'Ella Gagnet aille voter ou non ?
Comment simuler alors le nombre de suffrages que pourrait obtenir Ella Gagnet ?

3. De même, on suppose que chaque supporter d'Ilan Perdhu a 65% de chance de se rendre aux urnes. Comment utiliser les fonctions ALEA et ENT pour simuler qu'un supporter d'Ilan Perdhu aille voter ou non ?
Comment simuler alors le nombre de suffrages que pourrait obtenir Ilan Perdhu ?

4. En réalisant plusieurs simulations -Touche F9- peut-on voir qui va gagner ?

II. Analyse des proportions

On considère que lors d'une élection, qui vient d'avoir lieu, et où s'affrontaient deux candidats, le candidat 1 avait n_1 supporters avec un taux de participation parmi ses supporters égal à p_1 , et que le candidat 2 avait n_2 supporters avec un taux de participation parmi ses supporters égal à p_2.

1. Donner en fonction des nombres n_1, n_2, p_1 et p_2 le taux de participation lors de l'élection. (C'est une question de cours.)

2. Quels nombres faut il comparer pour savoir qui est le plus populaire ?

3. Quels nombres faut il comparer pour savoir qui aura le plus de voix ?

4. Exprimer en fonction de n_1, n_2 et p_1 la participation minimale pour que le candidat 2 emporte l'élection.

5. En vous appuyant sur les résultats précédents et en supposant que 60% des supporters d'Ella Gagnet et 65% des supporters d'Ilan Perdhu se rendent réellement aux urnes, pouvez vous dire qui va gagner les élections de Nouillorc ? Justifier.

Proposition de correction commentée :

I. Simulation sur tableur

1. Ella Gagnet devrait l'emporter puisqu'elle est la plus populaire.
Remarque : On pourra demander à l'oral aux élèves comment pourrait on savoir qui est le candidat le plus populaire, et dans quelle proportion, pour faire apparaître le terme de « sondages » et rendre ainsi le contexte plus concret.

2. L'instruction =ALEA(ENT+0.60) affiche 1 avec 60 chances sur 100 et 0 sinon.
Si 1 apparaît alors le supporter ira voter pour Ella Gagnet .
En « tirant la poignée » sur 520 lignes on simule l'attitude des 520 citoyens préférant Ella Gagnet, on utilise alors la fonction =SOMME (:).

3. De même avec =ALEA(ENT+0.65) pour un supporter d'Ilan Perdhu et en « tirant la poignée » sur 480 lignes.

4. En appuyant sur F9, on observe que les deux candidats peuvent gagner. Ce qui contredit l'affirmation initiale basée uniquement sur la popularité, mais sans la connaissance de la participation qui n'est fixée que le jour du scrutin.

Remarques :
Il est important alors de faire un bilan avec les élèves de cette simulation et d'expliquer que même si un candidat est annoncé gagnant par les sondages, il ne peut pas gagner tant que les électeurs n'ont pas voté ! La participation des supporters de chaque camp –qui n'est généralement pas mesurée par le sondage- importe autant que la popularité.
Pour annoncer la deuxième partie, plus difficile, il faut leur expliquer son but : montrer dans quelle mesure avoir des supporters plus citoyens, c'est-à-dire des supporters qui vont effectivement se rendre aux urnes peut compenser une popularité moindre.

II. Analyse des proportions

1. On a $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$.
Remarque : il s'agit ici de profiter la situation pour rappeler comment calculer la moyenne d'un groupe à partir de la moyenne de deux sous groupes.

2. et 3. Il s'agit ici surtout de manipuler les grandeurs littérales pour permettre aux élèves de mieux se les approprier : on compare n_1 et n_2 pour connaître le plus populaire et $n_1\ p_1$ avec $n_2\ p_2$ pour savoir qui a le plus de voix.

4. Pour que le candidat 2 l'emporte il faut que $n_2 > \dfrac{n_1 p_1}{p_2}$.

5. Pour le cas de Nouillorc, en supposant que 60% des supporters d'Ella Gagnet et 65% des supporters d'Ilan Perdhu se rendent réellement aux urnes, on a $n_2 = \dfrac{n_1 p_1}{p_2}$!
Autrement dit, l'élection pourrait se jouer à une voix près !

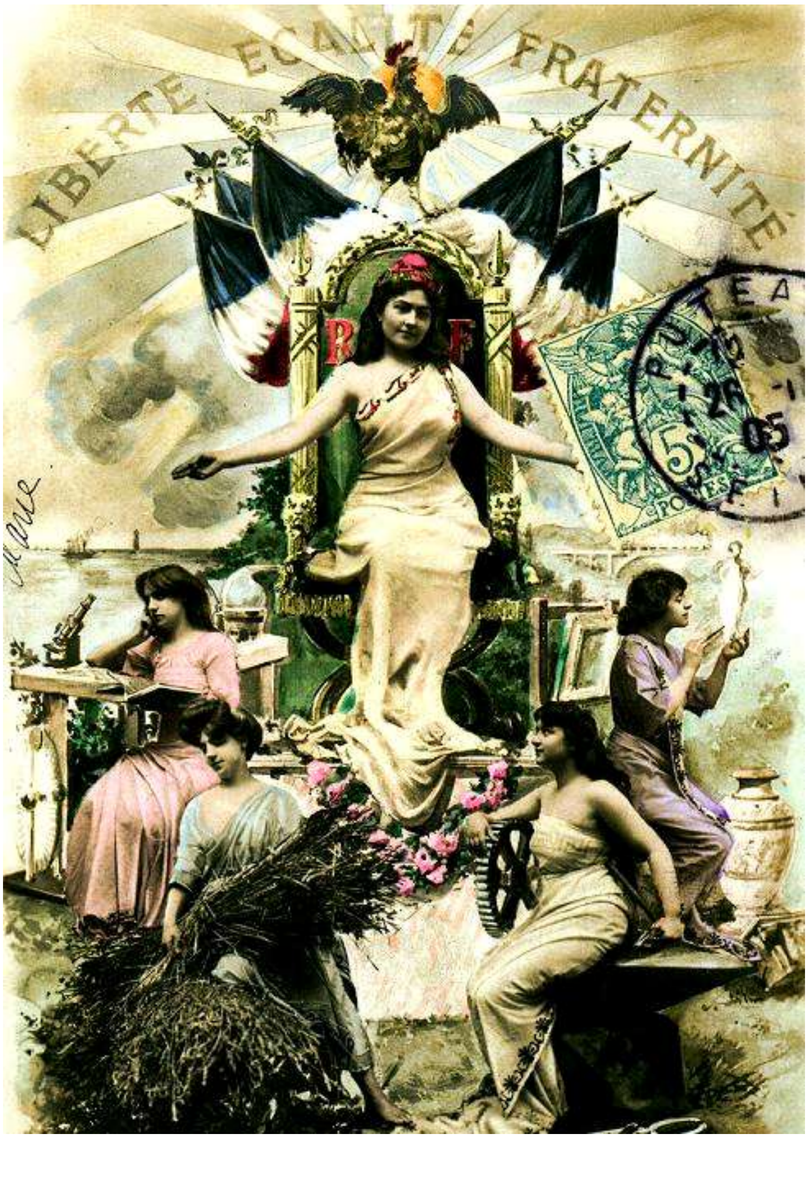
En conclusion

Le but est évidemment de faire réagir les élèves à ce qui ressemble à une provocation du professeur (et c'en est une !). Cette activité a pour enjeu citoyen de montrer que l'abstentionniste qui pense que voter est inutile à tort. Mais il faut éviter l'écueil inverse qui consisterait à penser que le sort de tout scrutin sera lié uniquement à leur vote et à leur participation individuelle. Autrement dit, cette activité est l'occasion d'expliquer pourquoi lors des campagnes électorales les partis s'attachent à impulser non seulement un mouvement de popularité en leur faveur : « Je suis le ou la meilleur(e) candidat(e) » mais aussi et surtout un **mouvement de participation** : « Allez aux urnes ! Allez voter pour moi ! ».

Lors d'une élection, sans la participation, la popularité n'est rien.

4

Économie et consommation



Le calcul de l'impôt

Niveau

Première ES enseignement de spécialité.

Situation étudiée

Calcul du quotient fiscal et du montant de l'impôt sur le revenu d'un foyer fiscal français classique. Progressivité de l'impôt. Inégalités des revenus.

Type d'activité

Séance de T.P. utilisant le tableur.

Durée

1 heure 30.

Objectifs

Contenus mathématiques au programme

- Calcul et construction de courbe avec un tableur.
- Fonctions affines par morceaux.
- Effectifs cumulés croissants.
- Tableaux à double entrée.
- Déciles.

Enjeux citoyens

Comprendre comment déterminer le montant de l'impôt sur le revenu d'un foyer fiscal classique. Comprendre pourquoi sa progressivité est facteur de cohésion sociale.

Capacités et attitudes

Utiliser le tableur pour déterminer le montant de l'impôt sur le revenu.
Interpréter ses résultats.
Analyser des documents à caractère économique pour en tirer une conclusion sociale.

Organisation

Le chapitre de traitement des données a déjà été abordé. Il est préférable qu'il ne s'agisse pas du premier cas de fonctions affines par morceaux pour ne pas multiplier les difficultés, la fin du sujet étant relativement difficile. Enfin, il est souhaitable également que cette activité se fasse en concertation avec l'enseignant de SES afin de faire écho avec le chapitre sur l'action des pouvoirs publics et les prélèvements obligatoires du programme de 1ère ES.

Description des activités

Énoncé élève

Calcul de l'impôt et distribution des revenus en France

Partie I : Calcul de l'impôt

Notations :

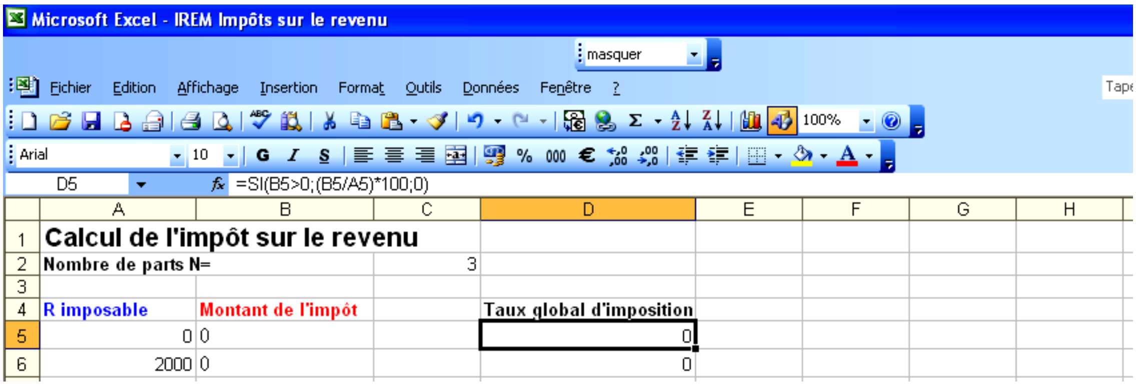
- R : Revenu imposable.
- N : Nombre de parts. Un couple marié ou pacsé représente deux parts. A la naissance du premier enfant, une demi part supplémentaire est ajoutée au quotient familial du couple. Le quotient est donc porté à 2,5 parts. Le calcul est le même pour le deuxième enfant. Mais à partir du troisième enfant, la part supplémentaire est portée non plus à une demi part mais à une part entière pour toute personne supplémentaire à charge.

1) Déterminer le nombre de parts d'un célibataire sans enfant, d'un couple marié sans enfant, d'un couple marié avec deux enfants à charge, d'un couple marié avec trois enfants à charge.
2) On appelle quotient familial et d'un foyer fiscal on notera Q le quotient du Revenu imposable R par le nombre de parts N ; ainsi $\mathrm{Q}=\dfrac{R}{N}$.
Déterminer le quotient familial :
a) de Roxane célibataire sans enfant ayant un revenu imposable de 10 000 €.
b) de Marine et Romain couple marié trois enfants et un revenu imposable de 20 000 €.
c) de leurs voisins Karim et Amina couple marié ayant deux enfants et un revenu imposable de 300 000 €.

Barème de l'impôt sur les revenus 2006

Tranches du quotient familial R/N en euros	Taux marginal d'imposition	Montant de l'impôt brut en euros
[0, 5614]	0 %	0
]5614, 11198]	5,5 %	$R \times 0{,}055 - 308{,}77 \times N$
]11198, 24872]	14 %	$R \times 0{,}14 - 1260{,}60 \times N$
]24872, 66679]	30 %	$R \times 0{,}3 - 5240{,}12 \times N$
]66679, + ∞[	40 %	$R \times 0{,}4 - 11908{,}02 \times N$

La suite du TP s’effectue sur une feuille de calcul à partir d’un fichier qui est fourni aux élèves.



On introduit en C2 le nombre de parts.
Les formules permettant le calcul de l’impôt sur tableur utilisent la fonction :
=SI(condition ; alors ; sinon). Elles sont pénibles à saisir et peuvent être fournies. On peut éventuellement en demander une explication. Dans la cellule B5 est entrée la formule suivante :
=SI(A5/C$2<5614;0;SI(A5/C$2<11198;A5/C$2*0,055-308,77;SI(A5/C$2<24872;A5/C$2*0,14-1260,6;SI(A5/C$2<66679;A5/C$2*0,3-5240,12;A5/C$2*0,4-11908,02)))).

1) Par recopie vers le bas, compléter la colonne A pour considérer des revenus annuels de 0 à 300 000 € avec un pas de 2 000.

2) Sélectionner la cellule B5, puis recopier vers le bas pour effectuer les calculs.

3) À l’aide de la feuille de calcul, et en modifiant le contenu de la cellule C2, donner le montant de l’impôt à payer dans chacun des trois cas ci-dessus.

4) Représenter l’impôt à payer en fonction du revenu entré en colonne A (utiliser « Nuage de points reliés par une courbe sans marquage des données ») dans le cas où N=3.
Justifier que le graphique est composé de segments de droites.

5) On appelle taux global d’imposition la part, en pourcentage, du revenu imposable consacrée au paiement de l’impôt. Calculer en colonne D le taux global d’imposition.

6) Etes-vous d’accord avec la phrase suivante : « A nombre de parts fixé, le montant de l’impôt augmente proportionnellement au revenu imposable » ?

7) Trouvez vous juste que le taux global d’imposition ne soit pas le même pour tous ?

Partie II : Etude de la distribution des salaires en France en 2004

Document 1 : Les revenus disponibles reçus pour chaque décile

Intervalle	**Part du revenu disponible (%)**	**Part cumulée de la population**	**Part cumulée du revenu disponible**
<= D1	3	10	3
>D1 et <=D2	4,5	20	7,5
>D2 et <=D3	5,5		
>D3 et <=D4	6,7		
>D4 et <=D5	7,9		
>D5 et <=D6	9,2		
>D6 et <=D7	10,7		
>D7 et <=D8	12,5		
>D8 et <=D9	15,2		
>D9	24,8		
ensemble des ménages	100		

Source : enquêtes revenus fiscaux en 2004, Insee-DGI.

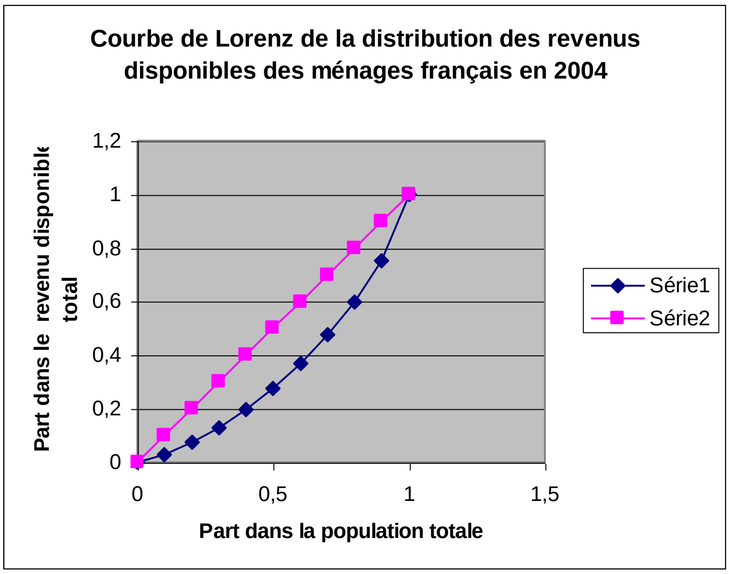
1) Lecture de la première ligne du tableau : Les 10% les plus pauvres ne gagnent que 3% des revenus.
Que signifie 24,8 à l'avant dernière ligne ?

2) Si les revenus disponibles étaient parfaitement égaux, quel serait le revenu disponible des 10% les plus riches ?

3) Compléter les deux dernières colonnes du tableau.

Document 2 : Courbe de Lorenz de la distribution des revenus disponibles des ménages français en 2004

Sur le document suivant, la série 1 correspond aux colonnes 3 et 4 du tableau précédent.
La courbe obtenue s'appelle **courbe de Lorenz**.



4) Que représente la série 2 ?

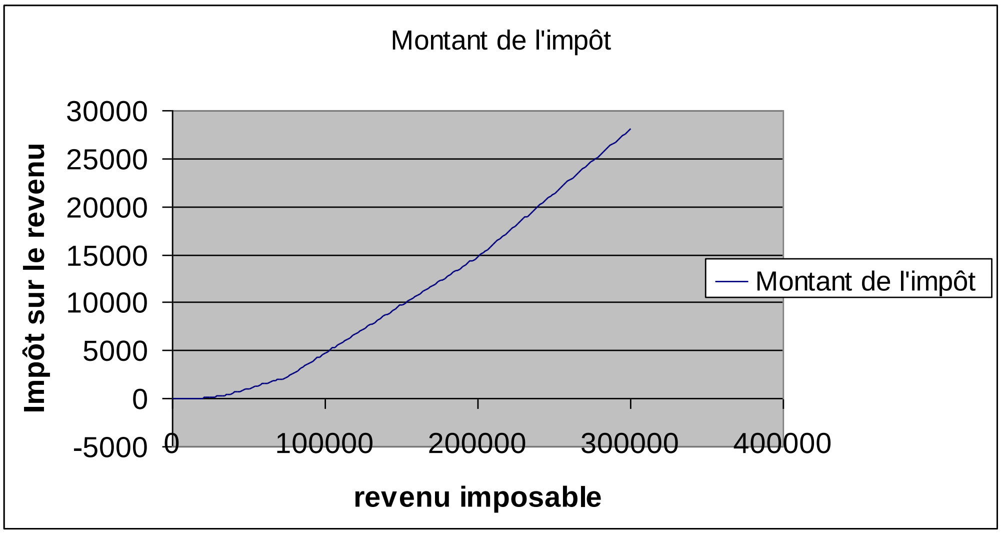
5) On appelle coefficient de Gini le double de la surface de la « lentille » située entre la courbe de Lorenz et la celle de la série 1.
Compléter la phrase suivante :
Plus le coefficient de Gini est grand, plus la répartition des revenus est………..

6) Pour vous, le caractère progressif de l'impôt sur le revenu est-il un facteur de distension ou de cohésion sociale ?

Éléments de réponse

Partie I : Calcul de l'impôt

3) Pour Roxane 241,23 € , pour Marine et Romain 0 € et pour Amina et Karim 28091,98 €.
4) Si on note I l'impôt, la fonction $R \mapsto I$ est affine par morceaux.
Graphique pour $N = 3$ parts.



6) Pour un nombre de parts fixé, plus les revenus sont importants, plus le montant de l’impôt est important et plus le taux global d’imposition est important.
Il s’agit surtout lors de cette question de mettre en relief que le lien entre l’impôt est et le revenu n’est pas une relation de proportionnalité. Autrement dit, le mode de prélèvement de l’impôt sur le revenu n’est pas égalitaire.
7) Il est intéressant de permettre aux élèves de s’exprimer sans support avant d’alimenter leur réflexion par la courbe de Lorenz.

Partie II : Etude de la distribution des salaires en France en 2004

1) Les revenus des 10% les plus riches représentent 24,8% des revenus totaux.
2) Si la distribution des salaires était égalitaire, ils auraient 10% des revenus et l’adjectif pauvre serait inutile : il n’y aurait ni riches ni pauvres !
4) La série 2 représente la distribution des revenus dans un système parfaitement égalitaire.
5) Le coefficient de Gini permet de mesurer l’inégalité des salaires.
6) Le caractère progressif de l’impôt sur le revenu est un facteur de cohésion sociale.

Bilan et perspectives

Cette activité permet aux élèves de comprendre comment on détermine le montant de l’impôt sur le revenu (ce qui est déjà un premier objectif en soi). Elle permet aussi de mettre en parallèle deux inégalités : tout le monde ne calcule pas son impôt avec la même formule, mais les salaires ne sont pas égaux non plus. Il est vraiment utile au professeur de mathématiques d’avoir présent à l’esprit certains points du programme de SES pour ne pas s’égarer dans un débat politique, le but étant seulement d’alimenter leur réflexion citoyenne et en aucun cas de dire si le montant du calcul est le plus juste qui soit, ce qui n’a d’ailleurs aucun sens mathématique.

En économie, les élèves de 1ère ES étudient le libéralisme qui est un système où les inégalités jouent aussi un rôle positif, dans la mesure où, selon cette théorie, elles stimulent le travail et l’innovation et donc la croissance. Cependant, on enseigne aussi que ce système ne peut pas fonctionner sans l’interventionnisme de l’état pour le réguler, notamment par ses modalités de prélèvement obligatoire afin de ne pas laisser les inégalités se creuser (cela serait source de forts conflits sociaux).

Enfin, l’indice de Gini peut aussi être abordé en Terminale dans le cadre du calcul intégral. En terminale ES, les élèves ont au programme de SES un chapitre sur les inégalités que cette activité pourrait illustrer.

Esprit critique et information chiffrée en économie

Niveau

Collège : $4^{ème}$ ou $3^{ème}$.
Lycée : seconde.

Situation étudiée

Publicités, graphiques, informations statistiques qui pourraient être extraites d'un journal (télévisé, papier ou radio…). Exemples de la vie quotidienne ou d'économie.

Type d'activité

Travaux dirigés, exercices en groupe, support pour un débat en classe, devoir maison pour une rédaction plus individuelle.

Durée

2 heures en classe (faire un choix d'exercices).
Travail de rédaction argumentée à la maison.

Objectifs

Contenus mathématiques au programme

Statistique descriptive, calculs de pourcentages, histogrammes.

Enjeux citoyens

Développer l'esprit critique et la lecture critique de l'élève face aux données chiffrées.
Détecter les pièges du marketing.
Favoriser l'échange, le débat, l'argumentation. Infirmer ou affirmer.

Compétences et attitudes

Faire preuve d'esprit critique face à une information chiffrée tirée de la vie quotidienne ou de l'économie du pays.
Se forger une opinion (à l'aide de calculs, graphique…) et l'argumenter.
Formuler oralement une réponse argumentée.
Participer à un débat et échanger des arguments à propos de la validité d'une solution.
Contrôler la pertinence ou la vraisemblance d'une solution.
Traiter les informations d'un document écrit incluant des représentations (histogramme, publicité)
Rédiger un texte pour communiquer le résultat d'une recherche individuelle ou collective.

Organisation

L'activité peut se situer en début de chapitre sur la statistique.
On peut répartir les élèves en groupes, laisser quelques minutes de réflexion et les laisser rédiger leurs réponses. Corriger en classe entière, après chaque exercice terminé, pour que les élèves acquièrent petit à petit les « bonnes » attitudes face aux exercices. Choisir quelques exercices qui seront à chercher et rédiger à la maison, ou déjà traités oralement en cours et à rédiger à la maison.

L'échange entre les élèves doit être favorisé pour les aider à développer leur esprit critique.

Description des activités

Énoncés élève

I – Publicités

Exercice 1 : « Petites annonces »

Voici la publicité d'un journal gratuit de petites annonces, « le 76 » (Seine-Maritime) paru en juin 1996. Que pensez-vous qu'un lecteur puisse conclure à sa lecture ?

1) Ce journal est très lu en Normandie.
☐ Vrai ☐ Faux
☐On ne peut pas savoir. Pourquoi ?
....
....

2) Les autres journaux de presse gratuite sont peu lus en Normandie.
☐ Vrai ☐ Faux
☐On ne peut pas savoir. Pourquoi ?
....
....

3) Les lecteurs des journaux de presse gratuite de la zone de diffusion lisent massivement ce journal.
☐ Vrai ☐ Faux
☐On ne peut pas savoir. Pourquoi ?
....
....

Exercice 2 : « du jambon ! »

Un exemple d'utilisation de « nombres inducteurs » est donné par une publicité dans le métro en 2006 :

« Deux produits achetés, un remboursé en bon d'achat » affirme l'annonce. Placé devant cette phrase, la plupart des personnes, parce que les 2 nombres « 2 » et « 1 » sont clairement mis en relief, pensent que cela équivaut à « 2 pour le pris d'un », ce qui est formellement vrai… sauf que pour avoir 2 articles, je dois tout de même débourser le prix de deux articles.

Quel est le pourcentage d'économie pour l'achat de 2 produits, de 3 produits ?

II – Pourcentages d'évolution et coefficients multiplicateurs

Exercice 3 : « c'est pas si évident que ça en a l'air…»

1) Le chômage en France source INSEE
a) Le chômage a augmenté de 11,9% entre 2002 et 2003, puis de 1,9% entre 2003 et 2004. Quelle est en pourcentage l'augmentation du chômage entre 2002 et 2004 ?
b) Il y avait en 2002, 2 396 000 chômeurs. Quel était le nombre de chômeurs en 2004 ?
c) Le chômage a augmenté de 11,9% entre 2002 et 2003, a aussi augmenté de 1,9% entre 2003 et 2004, puis a baissé de 0,6% entre 2004 et 2005. Sachant de plus qu'il y avait en 2002, 2 396 000 chômeurs, retrouver le nombre de chômeurs en 2005.

2) Entre 1991 et 1995, le coût moyen du pétrole brut importé en France a baissé de 24,3%, mais il a repris 30,7% de 1995 à 1997. Les automobilistes se sont inquiétaient de l'augmentation du coût moyen du pétrole brut importé en France entre 1991 et 1997. D'accord ou pas d'accord ?

3) Chaque année depuis 20 ans au pays de QuickDo, le poids moyen des adolescents augmente de 2%, soit une augmentation de 40% (il s'agit ici d'un exemple imaginaire). D'accord ou pas d'accord ?

4) La dette de la France, qui avait augmenté de 10,45% entre 2002 et 2003, n'a augmenté que de 7,17% entre 2003 et 2004. Le gouvernement se félicite de sa gestion exemplaire. Qu'en penses-tu ?
La dette de la France en 2002 était de 901,4 milliards d'euros.
Parmi les 3 affirmations suivantes, laquelle ou lesquelles sont justes ? Justifier.

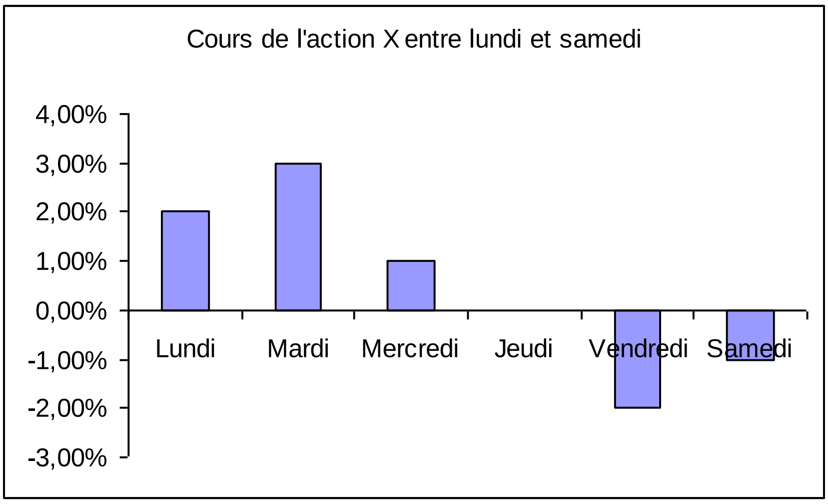
a) Diminution du déficit : l'augmentation de la dette est réduite entre 2003 et 2004 par rapport à la dette entre 2002 et 2003.
b) Diminution du déficit : de 94,2 milliards entre 2002 et 2003, il serait seulement de 71,4 milliards entre 2003 et 2004.
c) La dette a augmentée de plus de 165 milliards en 2 ans.

5) Fin décembre 2006, le maire d'une ville de Syldavie se réjouit :
« Depuis un an, affirme-t-il, le nombre de voitures brûlées ne cesse de diminuer de mois en mois. Et même, il commençait déjà à baisser en octobre 2005, peu de temps après mon élection. »
L'opposition ne l'entend pas de cette oreille, et fait publier par le journal local des statistiques montrant que les résultats de 2006 sont déplorables par rapport à 2005.
Là aussi, les 2 parties peuvent très bien avoir raison. Imagine une situation qui donne raison aux 2 parties.

IV – Avec des graphiques

Exercice 4

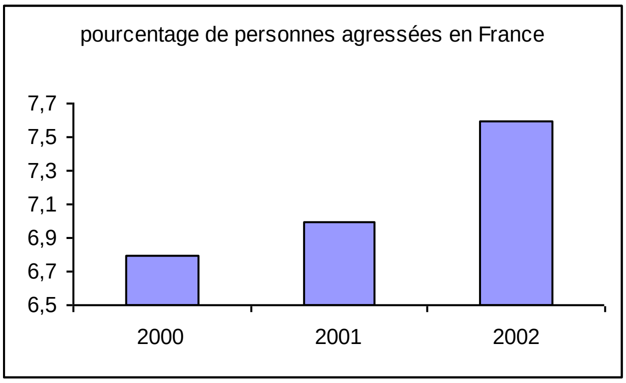
Un graphique d'une revue boursière donne les variations du cours d'une action, au jour le jour depuis quelques jours.



Aucune explication n'est donnée dans la revue, et le lecteur est face à ce mystère d'un cours négatif, en plus en pourcentages… Dans l'article qui suit ce graphique, on affirme cependant que le prix de l'action est plus élevé le samedi que le lundi à l'ouverture. Vrai ou faux ?
Aide toi d'un histogramme : l'action coûte à l'ouverture le lundi 100€.

Exercice 5

L’explosion de l’insécurité en France continue entre 2000 et 2002. Depuis 2000, le pourcentage de personnes agressées augmente de manière dramatique. Le graphique ci-dessous, qui donne le pourcentage de personnes agressées chaque année depuis 2000, montre bien l’explosion de violence dont nous sommes aujourd’hui victime.



Source : Insee, enquêtes permanentes sur les conditions de vie des ménages.

Refaire un histogramme dont l’axe des ordonnées est gradué à partir de 0 (1 carreau pour 1% des personnes agressées).
Critiquer.

V – Encore ! J’adore !

Exercice 6 : Méfiance !

Quels commentaires peut-on faire sur chacune des phrases suivantes ?
• 100% des gagnants au loto ont tenté leur chance.
• 100% lavable en machine.
• Le lit est l’endroit le plus dangereux du monde : 90% des personnes y meurent.
• Professeur Atoufair – médium – voyant – astrologue
Spécialiste de tous travaux de médium et du retour de l’affection ! Résultats à 100%.
• Publicité d’une école de conduite : remise de 10% sur présentation de la carte ; sauf inscription, code, conduite et conduite accompagnée.
• En moyenne un français a 1,99 jambes… Est-ce possible ?
• Pendant les soldes, un vêtement soldé affiche 2 démarques, la première -30% et la seconde -20%. Ce vêtement coûtait au départ 100€, il en coûte maintenant 54€. Il n’est pas vraiment à moitié prix, pourquoi ?
Trouver une phrase du même genre.

Exercice 7 : Diagnostic de performance énergétique, bien connu mais assez mal perçu

Depuis le 1er juillet 2007, toute signature d'un contrat de location ou d’un achat d'un logement doit s'accompagner d'un diagnostic de performance énergétique.
À cette occasion, les résultats d’une enquête réalisée par l’IFOP à la demande de l’ADEME ont été présentés. Elle porte sur la perception de ce dispositif par les particuliers et sur l’évaluation de son impact et de son influence dans l’acte d’achat et de location. Publié par le Secrétariat d'Etat à l'Ecologie, le sondage affirme que le DPE bénéficie d’une opinion favorable des ménages. En effet, selon l’enquête réalisée auprès d’un échantillon

de 873 ménages, interrogés par téléphone du 08 au 10 août, 67% en ont entendu parlé dont 63% déclarent connaître son contenu. 70% le savent obligatoire pour la vente.

Comment doit-on comprendre ces pourcentages ?

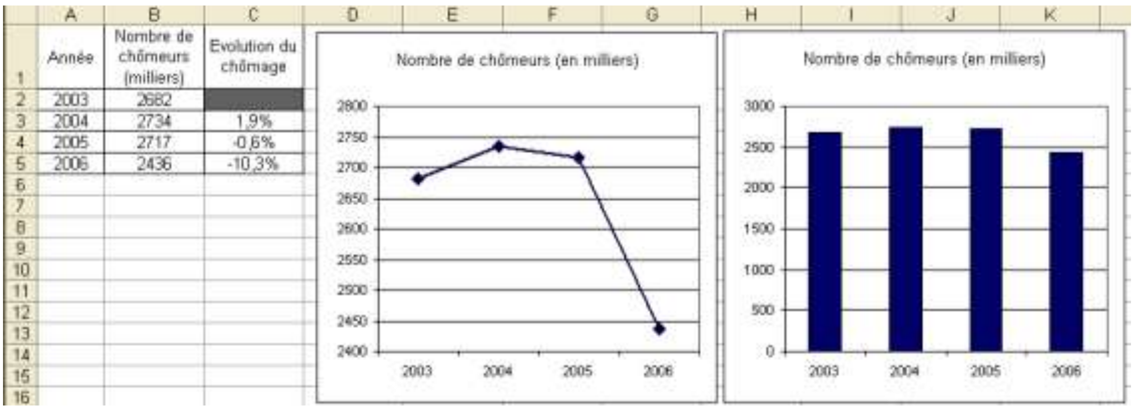
Exercice 8 : Egalité des chances

Une étude du ministère de l'éducation nationale indique qu'en 2004, 8,5% des élèves de classes préparatoires aux grandes écoles sont issus d'une famille d'employés (salariés non manuels) et 6,7% d'une famille d'artisans ou commerçants.

Peut-on en déduire l'affirmation suivante : « Un enfant d'employé a à peu autant de chances d'être en classe préparatoire aux grandes écoles, qu'un enfant d'artisan ou de commerçant » ? Justifier la réponse.

Exercice 9 : Le bon graphique avec les chiffres du chômage en France

La feuille de calcul suivante étudie l'évolution du nombre de chômeurs en France entre 2003 et 2006.



a) De quel pourcentage a baissé le nombre de chômeurs entre 2005 et 2006.
b) Les graphiques de la feuille de calcul sont-ils exacts ?
Sur quel graphique perçoit-on le mieux la réalité du pourcentage de baisse entre 2005 et 2006 ?

Exercice 10 : Les choix de graphiques

Le tableau ci-dessous représente l'évolution en pourcentage des chômeurs par rapport à la population active de la Syldavie (pays imaginaire chez Tintin), entre les mois de janvier et d'avril 2007.

mois	janvier	février	mars	avril	mai
% chômeurs	2,59%	2,66%	2,76%	2,92%	2,97%

1) Représenter graphiquement cette évolution en choisissant successivement les échelles suivantes :
Graphique 1 : en abscisses, les mois : 2 carreaux pour un mois.
En ordonnées, les pourcentages : 2,5 carreaux pour 1%, de 0% à 3%.
Graphique 2 : en abscisses, les mois : 1 carreau pour un mois.
En ordonnées, les pourcentages : 2,5 carreaux pour 0,1%, de 2,5% à 3%.

2) Le 15 juin 2007, des élections se déroulent en Syldavie opposant les Bleus, parti au gouvernement, et les Verts, parti d'opposition.
a) Si tu étais un candidat des Bleus, quel graphique choisirais-tu de présenter à la télévision ? Avec quel commentaire ?
b) Même question si tu étais un candidat des Verts.

Éléments de réponse

Exercice 1

Voici la publicité d'un journal gratuit de petites annonces, « le 76 » (Seine-Maritime) paru en juin 1996. Que pensez-vous qu'un lecteur puisse conclure à sa lecture ?

1) Ce journal est très lu en Normandie.

☐ Vrai ☐ Faux

■ On ne peut pas savoir. Pourquoi ?

<u>On ne connait pas la proportion des lecteurs de presse gratuite en Normandie par rapport aux lecteurs en Normandie.</u>

2) Les autres journaux de presse gratuite sont peu lus en Normandie.

☐ Vrai ☐ Faux

■ On ne peut pas savoir. Pourquoi ?

<u>Rien n'empêche les lecteurs de ce journal de lire d'autres journaux de presse gratuite.</u>

3) Les lecteurs des journaux de presse gratuite de la zone de diffusion lisent massivement ce journal.

■ Vrai ☐ Faux

☐ On ne peut pas savoir. Pourquoi ?

<u>En considérant que les lecteurs de moins de 15 ans sont minoritaires, cette affirmation est vraie : 87,5% des lecteurs de presse gratuite lisent ce journal</u>

Exercice 2

« Deux produits achetés, un remboursé en bon d'achat » affirme l'annonce. Placé devant cette phrase, la plupart des personnes, parce que les 2 nombres « 2 » et « 1 » sont clairement mis en relief, pensent que cela équivaut à « 2 pour le pris d'un », ce qui est formellement vrai <u>*(le 2ème et le 3ème ne coûtent que le prix d'un produit)*</u>
Sauf que pour avoir 2 articles (les 2 premiers), je dois tout de même débourser le prix de deux articles.

Quel est le pourcentage d'économie pour l'achat de 2 produits, de 3 produits ?
Pour 2 produits achetés, l'économie est de 0%.
Pour 3 produits achetés, l'économie est environ 33% (3 achetés au tarif « normal » dont un remboursé ; ce qui équivaut à 3 pour le prix de 2)

Exercice 3

1- Le chômage :

Nombre de chômeurs et taux de chômage (au sens du BIT), en moyenne annuelle

	2005	2004	2003	2002
Nombre de chômeurs (en milliers)	**2 717**	**2 734**	**2 682**	**2 396**
dont : hommes	*1 328*	*1 330*	*1 300*	*1 149*
femmes	*1 389*	*1 404*	*1 383*	*1 247*
Taux de chômage (en %)	**9,8**	**9,9**	**9,8**	**8,8**
Hommes	9,0	9,0	8,8	7,8
Femmes	10,8	11,0	11,0	10,1
15-29 ans	17,3	17,4	16,7	14,7
30-49 ans	8,3	8,3	8,2	7,5
50 ans ou plus	6,7	7,1	7,2	6,5
Cadres et professions intellectuelles supérieures	4,9	4,8	4,1	3,6
Professions intermédiaires	5,5	5,9	5,0	4,3
Employés	10,3	10,2	9,1	8,8
Ouvriers	12,5	12,3	10,9	9,9

Champ : France métropolitaine, individus de 15 ans et plus.
Source : Insee, enquêtes sur l'emploi.

a. Le chômage a augmenté de 11,9% entre 2002 et 2003, puis de 1,9% entre 2003 et 2004.
L'augmentation du chômage entre 2002 et 2004 est de : 1,119x1,019~1,14 soit 14%

b. Il y avait en 2002, 2 396 000 chômeurs.
Le nombre de chômeurs en 2004 est : 2 396 000x(1+14%)=2 732 065 chômeurs.

c. Le chômage a augmenté de 11,9% entre 2002 et 2003, a aussi augmenté de 1,9% entre 2003 et 2004, puis a baissé de 0,6% entre 2004 et 2005.
Sachant de plus qu'il y avait en 2002, 2 396 000 chômeurs, le nombre de chômeurs en 2005 est donc :
2 396 000.(1+11,9/100).(1+1,9/100).(1-0,6/100) = 2 715 673 chômeurs.

2- Entre 1991 et 1995, le coût moyen du pétrole brut importé en France a baissé de 24,3%, mais il a repris 30,7% de 1995 à 1997. Les automobilistes se sont inquiétaient de l'augmentation du coût moyen du pétrole brut importé en France entre 1991 et 1997. D'accord ou pas d'accord ?
Pas d'accord : de 1991 à 1997 le coût moyen du pétrole brut importé en France a baissé de 1% :

(1-24,3/100).(1+30,7/100)=0,99.

Consommation primaire de produits pétroliers et coût moyen du pétrole brut importé en France

indice 1973 = 100

	Coût moyen du pétrole brut importé en euros/tonnes	Consommation primaire en tonnes
1991	180,0	75,2
1995	136,3	75,3
1997	178,2	77,2

Sources : Minefi, Observatoire de l'énergie, Douanes.

3- Chaque année depuis 20 ans au pays de QuickDo, le poids moyen des adolescents augmente de 2%, soit une augmentation de 40% (il s'agit ici d'un exemple imaginaire). D'accord ou pas d'accord ?

Pas d'accord : $(1+2/100)^{20} = 1{,}49$ soit 49% d'augmentation.

4- La dette de la France :

❻ Dette publique*

	au 31/12/2002		au 31/12/2003		au 31/12/2004	
	Milliards d'€	% du PIB	Milliards d'€	% du PIB	Milliards d'€	% du PIB
Etat	735,0	47,5	798,8	50,4	839,9	51,0
Organismes divers d'administration centrale	43,9	2,8	55,3	3,5	90,2	5,5
Administrations locales	105,8	6,8	109,8	6,9	112,4	6,8
Administrations de sécurité sociale	16,8	1,1	31,7	2,0	24,5	1,5
Total administrations publiques	**901,4**	**58,2**	**995,6**	**62,8**	**1067,0**	**64,7**

* au sens du traité de Maastricht (cf. Définitions).

Source : Comptes nationaux, base 2000, Insee

Les 3 affirmations sont exactes.

a. Diminution du déficit : l'augmentation de la dette est réduite entre 2003 et 2004 par rapport à la dette entre 2002 et 2003. En effet l'augmentation de la dette entre 2002 et 2003 est égale à 71,4 milliards ce qui inférieur à l'augmentation de la dette entre 2003 et 2004 qui est de 94,2 milliards.
b. Diminution du déficit : de 94,2 milliards entre 2002 et 2003, il serait seulement de 71,4 milliards entre 2003 et 2004. Le déficit est égal à la (dette année n+1) – (dette année n).
c. La dette a augmentée de plus de 165 milliards en 2 ans.

La compréhension du vocabulaire est importante dans cet exercice.

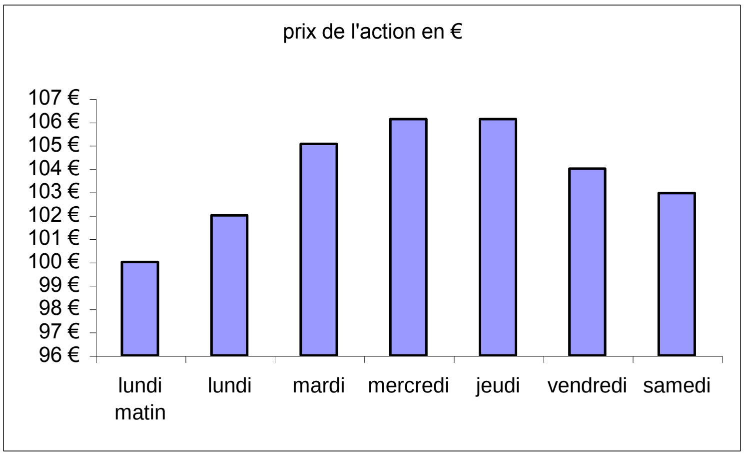
5- Fin décembre 2006, le maire d'une ville de Syldavie se réjouit :
« Depuis un an, affirme-t-il, le nombre de voitures brûlées ne cesse de diminuer de mois en mois. Et même, il commençait déjà à baisser en octobre 2005, peu de temps après mon élection. »
L'opposition ne l'entend pas de cette oreille, et fait publier par le journal local des statistiques montrant que les résultats de 2006 sont déplorables par rapport à 2005.
Là aussi, les 2 parties peuvent très bien avoir raison. Imagine une situation qui donne raison aux 2 parties.

Voici un exemple qui peut convenir :

Nombre de voitures brulées chaque mois de septembre 2005 à décembre 2006		
Mois	Voitures brulées chaque mois	effectifs cumulés croissant annuels
janv-05	30	30
févr-05	30	60
mars-05	30	90
avr-05	30	120
mai-05	30	150
juin-05	30	180
juil-05	30	210
août-05	30	240
sept-05	30	270
oct-05	49	319
nov-05	48	367
déc-05	47	414
janv-06	46	46
févr-06	45	91
mars-06	44	135
avr-06	43	178
mai-06	42	220
juin-06	41	261
juil-06	40	301
août-06	39	340
sept-06	38	378
oct-06	37	415
nov-06	36	451
déc-06	35	486

Exercice 4

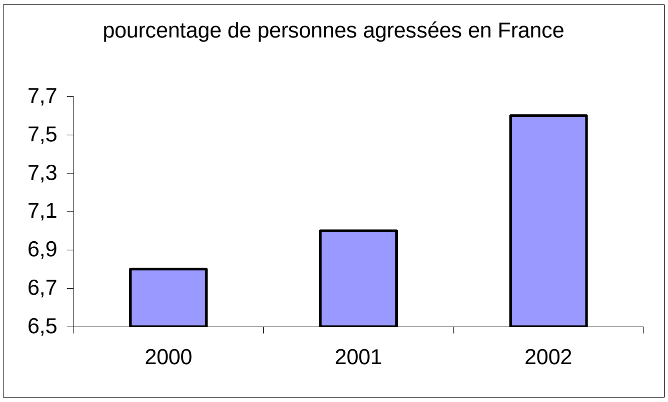
Aucune explication n'est donnée dans la revue, et le lecteur est face à ce mystère d'un cours négatif, en plus en pourcentages… Dans l'article qui suit ce graphique, on affirme cependant que le prix de l'action est plus élevé le samedi que le lundi à l'ouverture. C'est Vrai, comme le montre l'histogramme : l'action coûte à l'ouverture le lundi 100€ et 103€ à la fermeture le samedi.



Lundi matin à l’ouverture	100 €
lundi	102 €
mardi	105 €
mercredi	106 €
jeudi	106 €
vendredi	104 €
samedi	103 €

Exercice 5

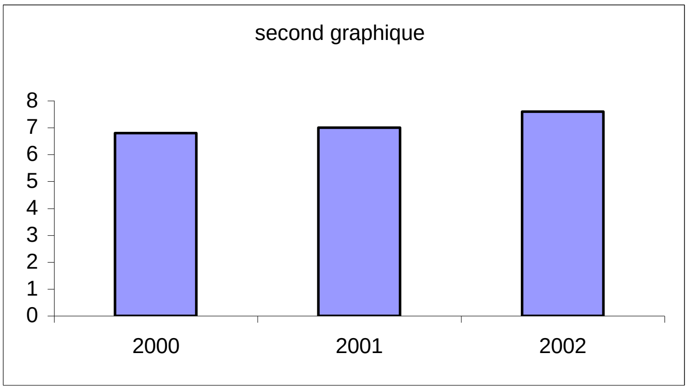
L’explosion de l’insécurité en France continue entre 2000 et 2002. Depuis 2000, le pourcentage de personnes agressées augmente de manière dramatique. Le graphique ci-dessous, qui donne le pourcentage de personnes agressées chaque année depuis 2000, montre bien l’explosion de violence dont nous sommes aujourd’hui victime.



Source : Insee, enquêtes permanentes sur les conditions de vie des ménages.

Refaire un histogramme dont l’axe des ordonnées est gradué à partir de 0 (1 carreau pour 1% des personnes agressées). Critiquer.

Voici l’histogramme attendu :



L’explosion est ici beaucoup moins nette.

Voici les sources :

Victimes de vols et d'agressions au cours des deux dernières années

En %

	1997	1998	1999	2000	2001	2002
Ménages concernés par des...						
cambriolages	3,6	3,2	2,7	2,7	2,5	2,5
vols de voiture (1)	12,7	11,1	9,7	10,7	9,4	10,2
Personnes concernées par des...						
vols (2)	5,4	4,8	4,3	4,7	4,7	5,2
agressions (physiques ou verbales)	5,4	6,3	4,4	6,8	7,0	7,6

(1) Y compris vols à la roulotte (vols dans une voiture en stationnement).
(2) En dehors des cambriolages et des vols de voiture (par exemple vol de sac à main, de portefeuille...).

Lecture : en 2002, 2,5 % des ménages ont déclaré avoir été victimes d'au moins un cambriolage au cours des deux dernières années, c'est-à-dire en 2000 ou 2001.
Champs : ensemble des ménages et des personnes de 15 ans ou plus.
Source : Insee, enquêtes permanentes sur les conditions de vie des ménages.

Il y avait 23,8 millions de ménages en France lors du recensement de 1999.

Exercice 6

Trouver une phrase du même genre.

- En moyenne un français a 1,99 jambes (il faut prendre en compte les unijambistes)
- Pendant les soldes, un vêtement soldé affiche 2 démarques, la première -30% et la seconde -20%. Ce vêtement coûtait au départ 100€, il en coûte maintenant 54€. Il n'est pas vraiment à moitié prix…

Exercice 7

Comment doit-on comprendre ces pourcentages ?
67% des ménages ont entendu parlé du DPE (Diagnostic de Performance Energétique) dont 63% (le « dont » signifie que 63% des 67% des ménages au courant) déclarent connaître son contenu : ça fait donc 0,67x0,63=0,42 soit 42% des ménages connaissent le contenu.
70% le savent obligatoire pour la vente. Ici il est impossible qu'il y ait seulement 63% des ménages au courant de l'existence du DPE et 70% des ménages qui savent que le DPE est obligatoire pour la vente. Donc il faut le comprendre : 0,63x0,70=0,44 soit 44% des ménages savent que le DPE est obligatoire pour la vente.

Exercice 8

Une étude de la MEN DEP indique qu'en 2004, 8,5% des élèves de classes préparatoires aux grandes écoles sont issus d'une famille d'employés (salariés non manuels) et 6,7% d'une famille d'artisans ou commerçants.

Peut-on en déduire l'affirmation suivante : « Un enfant d'employé a à peu autant de chances d'être en classe préparatoire aux grandes écoles, qu'un enfant d'artisan ou de commerçant » ? Justifier la réponse.

Répartition de la population de 15 ans et plus selon la catégorie socioprofessionnelle

en %

	2003	2004	2005
Catégorie socioprofessionnelle (PCS)	**Total**	**Total**	**Total**
Artisans, commerçants, chefs d'entreprise	**3,1**	**3,0**	**3,2**
Artisans	1,5	1,4	1,5
Commerçants et assimilés	1,3	1,3	1,4
Employés	**15,8**	**16,0**	**16,1**

* : résultats en moyenne annuelle en France métropolitaine.

Source : Insee, enquêtes Emploi.

Non, Un enfant d'employé n'a pas autant de chances d'être en classe préparatoire aux grandes écoles, qu'un enfant d'artisan ou de commerçant : soit n le nombre d'étudiants en classe préparatoire et soit p le nombre de personnes âgées de 18 ans en France

La proportion d'étudiants en classe préparatoire d'origine sociale « employés » parmi les jeunes âgés de 18 ans :

q_e =(0,085xn)/(0,16xp)

La proportion d'étudiants en classe préparatoire d'origine sociale « artisans, commerçants » parmi les jeunes âgés de 18 ans :

q_a =(0,067xn)/((0,014+0,013)xp) =(0,067xn)/(0,027xp)

Donc q_a / q_e = (0,16x0,067) / (0,085x0,027)=4,7 donc q_a =4,7 q_e

Donc un jeune d'origine sociale « employé » a 4,7 fois moins de chance d'être en classe préparatoire qu'un jeune d'origine sociale « ouvrier ».

Exercice 9

a) Entre 2005 et 2006 le nombre de chômeurs a baissé de 10,3%.

b) Les deux graphiques sont exacts (l'échelle en ordonnées est différente).

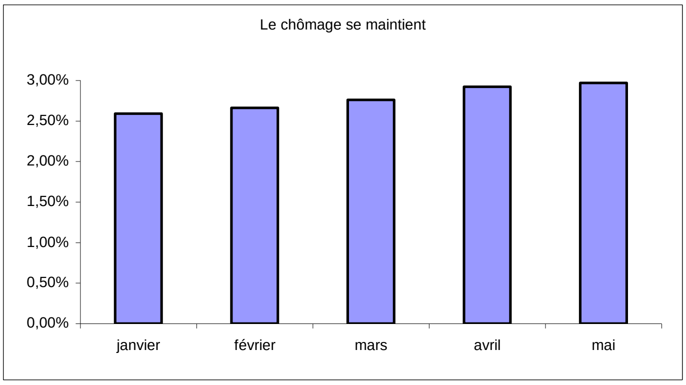
C'est sur le graphique de droite que l'on voit la baisse de 10% entre 2005 et 2006.

Exercice 10

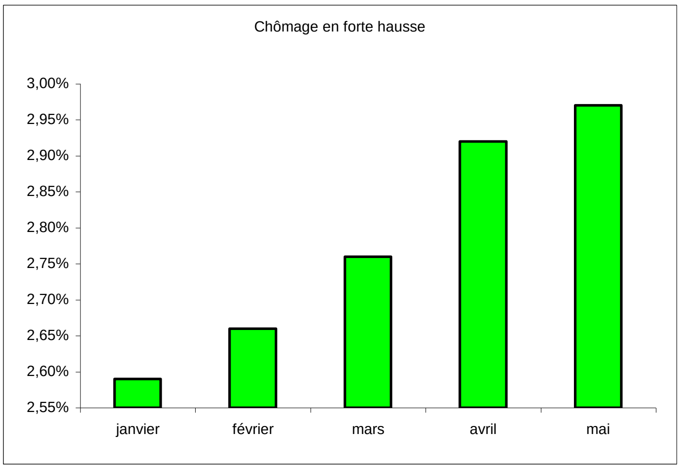
Le tableau ci-dessous représente l'évolution en pourcentage des chômeurs par rapport à la population active de la Syldavie (pays imaginaire chez Tintin), entre les mois de janvier et d'avril 2007.

mois	janvier	février	mars	avril	mai
% chômeurs	2,59%	2,66%	2,76%	2,92%	2,97%

Graphique 1 :



Graphique 2 :



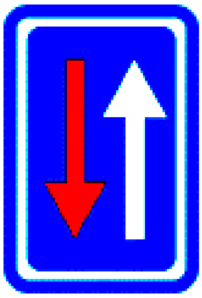
Sources :

- *Faire des mathématiques en seconde* – Elisabeth Hébert – CNDP Haute Normandie.
- *Statistiques, méfiez-vous* ! – Nicolas Gauvrit – Ellipses 2007.

5

Justice, discrimination et égalité des chances

Ascenseur social



Niveau

Énoncé n° 1 : Troisième, seconde.
Énoncé n° 2 : Terminales STG, ST2S, ES ou S.

Situation étudiée

On dispose des données suivantes :

Proportions d'élèves reçus à l'École polytechnique	1950	1990
d'origine défavorisée	21 %	7,8 %
d'origine favorisée	79 %	92,2 %

À la vue de ces chiffres, il semble que « l'ascenseur social » soit plus qu'en panne.
Il faut cependant, pour analyser ces chiffres, tenir compte de l'évolution de structure de la société comme le montre le tableau suivant :

Proportions de la population des 20-24 ans	1950	1990
d'origine défavorisée	90,8 %	68,2 %
d'origine favorisée	9,2 %	31,8 %

Comment tenir compte des deux tableaux pour voir réellement si la discrimination sociale est plus (ou moins) importante en 1990 qu'en 1950 ?

Type d'activité

Exercice.

Durée

Énoncé n°1 : 45 minutes.
Énoncé n°2 : 55 minutes.

Objectifs

Contenus mathématiques au programme

Énoncé n° 1 : Proportions, calcul littéral.
Énoncé n° 2 : Probabilités conditionnelles.

Enjeux citoyens

Favoriser l'analyse « critique » des données chiffrées en sensibilisant aux effets de « structure ».
Mettre en oeuvre des outils mathématiques (adaptés au niveau des élèves) permettant de quantifier la notion d'égalité (ou d'inégalité) des chances et son évolution entre deux périodes.

Pour l'énoncé n° 2 (terminale) :
Comprendre que tous les paramètres statistiques ne se valent pas, qu'ils peuvent conduire à des conclusions contradictoires (avec des calculs exacts et a priori « raisonnables ») et que la connaissance de leurs propriétés mathématiques permet d'en privilégier certains. La définition et les propriétés mathématiques d'un indicateur sont essentielles à connaître pour en comprendre le sens. Citons Claudine Schwartz (dans *Pratiques de la statistique* – Vuibert 2006) : « Notre société utilise de plus en plus d'indicateurs. [...] Ils n'ont pas de caractère absolu, leur interprétation est délicate, mais ils sont aujourd'hui nécessaires à l'appréhension des aspects variés du monde qui nous entoure. »

Capacités et attitudes

Esprit critique.
Interprétation des calculs et des résultats.
Réflexion sur un sujet de société.

Organisation

Exercice en classe ou en devoir à la maison.
Le second énoncé, plus long, peut être terminé à la maison.

Description des activités

Exercice en termes de proportions (3ème – 2nde)

Énoncé élève (niveau troisième, seconde)

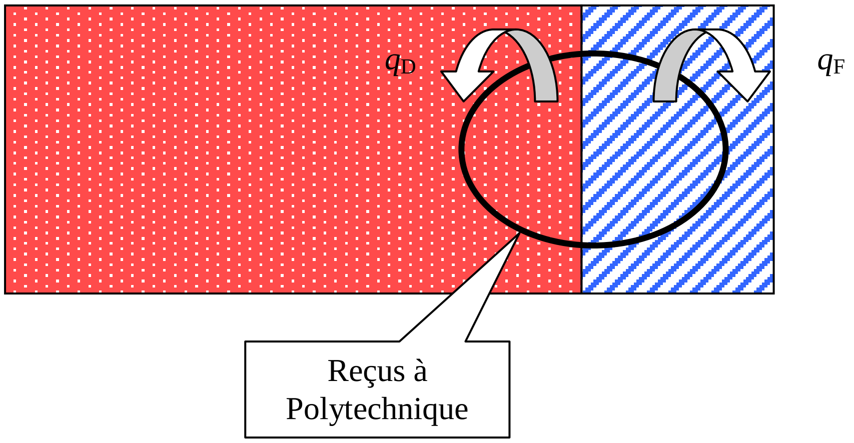
On dispose du tableau suivant, donnant, parmi les élèves reçus à l'École polytechnique, la proportion de ceux issus de classes « défavorisées » et ceux issus de classes « favorisées », pour les périodes 1950 et 1990.

Proportions d'élèves reçus à l'École polytechnique	1950	1990
d'origine défavorisée	21 %	7,8 %
d'origine favorisée	79 %	92,2 %

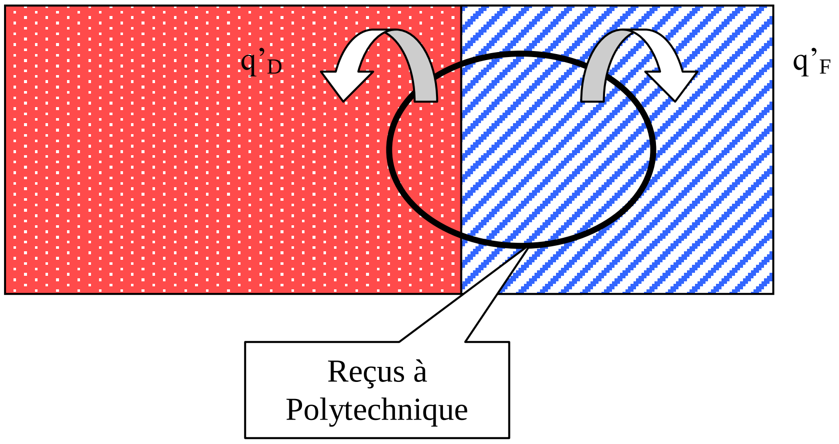
Pour étudier si la discrimination sociale est plus forte dans les années 1990 que dans les années 1950, il faut tenir compte de l'évolution de la composition de la société française entre ces deux périodes.

Proportions de la population des 20-24 ans	1950	1990
d'origine défavorisée	90,8 %	68,2 %
d'origine favorisée	9,2 %	31,8 %

1950 : population défavorisée | favorisée



1990 : population défavorisée | favorisée



1. Que signifie 21% dans le premier tableau ?
Que signifie 90,8% dans le second tableau ?

2. a) On note r le nombre de reçus à Polytechnique en 1950, exprimer en fonction de r le nombre de reçus à Polytechnique d'origine défavorisée en 1950.
b) On note n le nombre de jeunes de 20-24 ans en 1950.
Montrer que proportion q_D de reçus à Polytechnique parmi la population défavorisée en 1950 est $q_D = \dfrac{0{,}21 \times r}{0{,}908 \times n}$.

3. Donner, de même, l'expression de la proportion q_F de reçus à Polytechnique parmi la population favorisée en 1950.

4. On note $t = \dfrac{q_F}{q_D}$. Montrer que $t \approx 37$.

On peut interpréter ce résultat en disant qu'en 1950 un jeune « favorisé » a 37 fois plus de chances d'entrer à Polytechnique qu'un jeune « défavorisé ».

5. Pour comparer avec la situation en 1990, on utilise les mêmes notations que ci-dessus, avec un ' . On considère donc $t' = \dfrac{q_F'}{q_D'}$.

Calculer t'.

Peut-on considérer que « l'ascenseur social » s'est amélioré entre 1950 et 1990 ?
6. On souhaite utiliser le tableur (selon l'image d'écran suivante) pour calculer les quotients t et t' dans les quatre autres cas de ce tableau :

Proportions d'élèves d'origine défavorisée	1950	1990
ENA (École nationale d'administration)	18,3 %	6,1 %
ENS (Écoles normales supérieures)	23,9 %	6,1 %
HEC (Hautes études commerciales)	38,2 %	11,8 %
Grandes écoles (avec entrée sur concours)	29 %	8,6 %

On pourra organiser les calculs (ici pour l'École polytechnique) selon l'image d'écran, de sorte qu'il suffise de modifier le contenu des cellules B2 et C2.

	A	B	C	D
1	Proportion d'élèves reçus	1950	1990	
2	D	**0,21**	**0,078**	
3	F	0,79	0,922	
4				
5	Proportion des 20-24 ans	1950	1990	
6	D	0,908	0,682	
7	F	0,092	0,318	
8				
9		t	t'	
10		37,1283644	25,3509111	
11				

Interpréter les résultats obtenus.

Éléments de réponse (niveau troisième, seconde)

2. Le nombre de reçus d'origine défavorisée est $0,21 \times r$ et le nombre de 20-24 ans d'origine défavorisée est $0,908 \times n$.

3. On a $q_F = \dfrac{0,79 \times r}{0,092 \times n}$.

4. On a $q_D = \dfrac{0,21 \times r}{0,908 \times n}$. D'où $t = \dfrac{0,79}{0,092} \times \dfrac{0,908}{0,21} \approx \mathbf{37}$.

5. On a $q_F' = \dfrac{0,922 \times r'}{0,318 \times n'}$ et $q_D' = \dfrac{0,078 \times r'}{0,682 \times n'}$. D'où $t' = \dfrac{0,922}{0,318} \times \dfrac{0,682}{0,078} \approx \mathbf{25}$.

Pour ce qui concerne Polytechnique, on peut considérer que l'ascenseur social s'est (un peu) amélioré.

6. On peut entrer en B3 la formule =1-B2 puis la recopier vers la droite en C3.
On peut entrer en B10 la formule =(B3/B7)*(B6/B2) puis la recopier vers la droite en C10.
On obtient les résultats suivants (arrondis à l'unité).

	t	t'
ENA (École nationale d'administration)	44	33
ENS (Écoles normales supérieures)	31	33
HEC (Hautes études commerciales)	16	16
Grandes écoles (avec entrée sur concours)	24	23

On peut considérer globalement que les inégalités sociales de recrutement de ces écoles sont importantes et ont peu évolué entre 1950 et 1990.

Commentaires

• Il faut d'abord préciser qu'il ne s'agit ici que du cas des grandes écoles, qui ne constituent pas l'unique filière de « réussite » ou d'ascension sociale. Le cas du supérieur en général est différent, en particulier avec la création des BTS et des IUT.

• Source des données : Euriat et Thélot – *Le recrutement des élites scolaires depuis 40 ans* – Revue *Éducation et formation* – juin 1995.

L'exemple proposé a été cité par Jean-Louis Piednoir[2], en termes de proportions, pour illustrer le fait qu'il faut comparer ce qui est comparable et en particulier tenir compte de l'évolution de la constitution de la population sous-jacente. Le même exemple est beaucoup plus développé dans un ouvrage récent par Claudine Schwartz[3], en termes de probabilités conditionnelles, en particulier pour souligner le fait que la définition et les propriétés mathématiques d'un indicateur sont essentielles à connaître pour en comprendre le sens. Pour la petite histoire, signalons que Jacques Attali, par ailleurs polytechnicien, chargé par le ministre Claude Allègre d'un rapport sur l'organisation des études supérieures, citant les chiffres donnés ici en tête d'exercice, concluait à une « dé-démocratisation » du recrutement des élites. On a un peu moins de complexes à être nous mêmes parfois victimes d'erreurs dues à un « effet de structure » ou une variable cachée.

• La notion de classe défavorisée ou de classe populaire n'est bien sûr pas intrinsèque. Il est clair que la définition choisie influe sur les résultats, mais cela ne semble pas être l'élément essentiel ici.

Exercice en termes de probabilités conditionnelles (terminales STG-ST2S-ES-S)

Énoncé élève (niveau terminale)

On dispose du tableau suivant, donnant, parmi les élèves reçus à l'École polytechnique, la proportion de ceux issus de classes « défavorisées » et ceux issus de classes « favorisées », pour les périodes 1950 et 1990.

Proportions d'élèves reçus à l'École polytechnique	1950	1990
d'origine défavorisée	21 %	7,8 %
d'origine favorisée	79 %	92,2 %

Pour étudier si la discrimination sociale est plus forte dans les années 1990 que dans les années 1950, il faut tenir compte de l'évolution de la composition de la société française entre ces deux périodes.

Proportions de la population des 20-24 ans	1950	1990
d'origine défavorisée	90,8 %	68,2 %
d'origine favorisée	9,2 %	31,8 %

On prélève un jeune au hasard dans la population des 20-24 ans de 1950.

[2] Piednoir (Jean-louis) – Dutarte (Philippe) – *Enseigner la statistique au lycée : des enjeux aux méthodes* – IREM de Paris-Nord – 2001.

[3] Schwartz (Claudine) – *Pratiques de la statistique* – Vuibert 2006 – Chapitre 3 : *Des mots et des chiffres : probabilités conditionnelles, risques relatifs*. On consultera cet ouvrage pour mettre en perspective les activités proposées ici.

On note D l'événement « le jeune est issu de la classe défavorisée ».
On note F l'événement « le jeune est issu de la classe favorisée ».
On note R l'événement « le jeune a été reçu à l'École polytechnique ».
1. Donner la valeur des probabilités suivantes $P(D)$, $P(F)$, $P_R(D)$ et $P_R(F)$.
2. Justifier que $P_D(R) = \dfrac{P_R(D) \times P(R)}{P(D)}$ et établir la formule analogue pour $P_F(R)$.
3. En déduire que $P_F(R) \approx 37 \times P_D(R)$.
Traduire ce résultat par une phrase en français.
4. On prélève un jeune au hasard dans la population des 20-24 ans de 1990.
On note D' l'événement « le jeune est issu de la classe défavorisée ».
On note F' l'événement « le jeune est issu de la classe favorisée ».
On note R' l'événement « le jeune a été reçu à l'École polytechnique ».
Montrer qu'en 1990 la probabilité d'être reçu à Polytechnique était environ 25 fois plus importante pour un jeune favorisé que pour un jeune défavorisé.
Que peut-on en déduire comparativement à 1950 ?
5. Pour compléter cette étude on a considéré les quatre autres cas suivants :

Proportions d'élèves d'origine défavorisée (D)	1950	1990
ENA (École nationale d'administration)	18,3 %	6,1 %
ENS (Écoles normales supérieures)	23,9 %	6,1 %
HEC (Hautes études commerciales)	38,2 %	11,8 %
Grandes écoles (avec entrée sur concours)	29 %	8,6 %

On obtient les résultats suivants (arrondis à l'unité) :

Rapport entre les probabilités d'être reçu d'un jeune issu de la classe favorisée et d'un jeune issu de la classe défavorisée	1950	1990
ENA (École nationale d'administration)	44	33
ENS (Écoles normales supérieures)	31	33
HEC (Hautes études commerciales)	16	16
Grandes écoles (avec entrée sur concours)	24	23

Quels commentaires peut-on faire ?
6. On pourrait considérer qu'au lieu de faire le rapport des probabilités d'être reçu entre un jeune issu de la classe défavorisée et un jeune issu de la classe favorisée, il convient de faire le rapport des probabilités d'être reçu entre un jeune issu de la classe défavorisée et un jeune quelconque, c'est-à-dire $\dfrac{P_D(R)}{P(R)}$, pour reprendre le cas de l'École polytechnique en 1950.
Montrer que $\dfrac{P_D(R)}{P(R)} = \dfrac{P_R(D)}{P(D)}$ et en déduire la valeur de ce rapport (pour l'École polytechnique en 1950).
7. Calculer $\dfrac{P_{D'}(R')}{P(R')}$ pour l'École polytechnique en 1990. Comparer au résultat obtenu précédemment pour 1950. Quelle interprétation en fait-on du point de vue de l'ascenseur social ? Qu'aviez-vous conclu à propos de Polytechnique à la question 4 ?
8. Dans le cas des autres écoles, la situation est la suivante :

Rapport entre les probabilités d'être reçu d'un jeune issu de la classe défavorisée et d'un jeune quelconque	1950	1990
ENA (École nationale d'administration)	0,20	0,09
ENS (Écoles normales supérieures)	0,26	0,09
HEC (Hautes études commerciales)	0,42	0,17
Grandes écoles (avec entrée sur concours)	0,32	0,13

Lecture : en 1950, la probabilité d'un jeune défavorisé d'être reçu à l'ENA représente 20 % de celle d'un jeune quelconque.

Quel commentaire peut-on faire de ce tableau ?
9. Pour y voir un peu plus clair, on peut établir la relation liant les deux indicateurs utilisés.
Notons $t = \dfrac{P_F(R)}{P_D(R)}$ l'indicateur utilisé à la question 4 et $k = \dfrac{P_D(R)}{P(R)}$ l'indicateur utilisé à la question 6.
En utilisant que $P(R) = P(R \cap D) + P(R \cap F)$, montrer que l'on a :

$$\frac{1}{k} = P(D) + t \times (1 - P(D)).$$

10. Intéressons nous au cas de HEC où le premier indicateur est (pratiquement) constant pour les deux époques : $t = 16$. En notant x la proportion de jeunes de 20-24 ans d'origine défavorisée, on a $k = \dfrac{1}{x(1-t)+t} = \dfrac{1}{-15x+16}$.
Montrer que la fonction $x \mapsto \dfrac{1}{-15x+16}$ est croissante sur [0, 1].
En déduire que dans le cas de HEC, le déclin de l'ascenseur social, selon les utilisateurs de l'indicateur k, peut s'expliquer par la seule diminution du pourcentage des classes défavorisées entre 1950 et 1990.
Quel indicateur préconisez-vous pour rendre compte de l'ascenseur social dans la situation de cet exercice ?

Éléments de réponse (niveau terminale)

1. D'après les données, on a :
$P(D) = 0{,}908$; $P(F) = 0{,}092$; $P_R(D) = 0{,}21$ et $P_R(F) = 0{,}79$.
2. On a $P(D \cap R) = P_D(R) \times P(D) = P_R(D) \times P(R)$ d'où $P_D(R) = \dfrac{P_R(D) \times P(R)}{P(D)}$.
De même, $P_F(R) = \dfrac{P_R(F) \times P(R)}{P(F)}$.
3. On en déduit que $\dfrac{P_F(R)}{P_D(R)} = \dfrac{P_R(F) \times P(D)}{P_R(D) \times P(F)} = \dfrac{0{,}79 \times 0{,}908}{0{,}21 \times 0{,}092} \approx 37$.
Donc $P_F(R) \approx 37 \times P_D(R)$.
Ce résultat signifie qu'en 1950 la probabilité d'être reçu à Polytechnique était 37 fois plus importante pour un jeune favorisé que pour un jeune défavorisé.
4. Par un raisonnement analogue (qu'il est inutile de reproduire), on a :

$$P_{F'}(R') = \frac{0{,}922 \times 0{,}682}{0{,}078 \times 0{,}318} \approx 25.$$

On peut en déduire qu'en ce qui concerne le recrutement à Polytechnique, l'ascenseur social s'est un peu amélioré.

5. On peut considérer globalement que les inégalités sociales de recrutement de ces écoles sont importantes et ont peu évolué entre 1950 et 1990.
Il faut cependant préciser qu'il ne s'agit ici que des grandes écoles, le cas du supérieur en général est différent, en particulier suite à la création des BTS et des IUT.

6. On a $P(D \cap R) = P_D(R) \times P(D) = P_R(D) \times P(R)$ – déjà dit – d'où $\dfrac{P_D(R)}{P(R)} = \dfrac{P_R(D)}{P(D)}$.

On en déduit que $\dfrac{P_D(R)}{P(R)} = \dfrac{0{,}21}{0{,}908} \approx 0{,}23$ (en 1950, les chances d'être reçu à Polytechnique d'un jeune défavorisé représentent 23 % de celles d'un jeune quelconque).

7. On a $\dfrac{P_{D'}(R')}{P(R')} = \dfrac{0{,}078}{0{,}682} \approx 0{,}11$ (en 1990, les chances d'être reçu à Polytechnique d'un jeune défavorisé représentent 11 % de celles d'un jeune quelconque).
On en conclut que pour Polytechnique, l'ascenseur social fonctionne moins bien en 1990 qu'en 1950. Ce qui est le contraire de la conclusion obtenue à la question 4 !!

8. Dans tous les cas (ce qui peut nous mettre la puce à l'oreille) la situation s'est aggravée de façon importante.

9. On a $P(R) = P(R \cap D) + P(R \cap F) = P_D(R) \times P(D) + P_F(R) \times P(F)$.
D'où $\dfrac{1}{k} = \dfrac{P(R)}{P_D(R)} = P(D) + t \times P(F) = P(D) + t \times (1 - P(D))$.

10. Entre 1950 et 1990, la valeur de x décroît de 0,908 à 0,682. Comme l'indicateur k est une fonction croissante de x il décroît également.
Il convient donc de choisir le premier indicateur $t = \dfrac{P_F(R)}{P_D(R)}$.

Commentaire

Cet indicateur est « classique » et est par exemple utilisé en médecine où il se nomme le « risque relatif » de F. Si par exemple $t = 16$ pour une maladie R, la probabilité d'avoir R est 16 fois plus forte chez les F que chez les D.
Voir : Schwartz (Claudine) – *Pratiques de la statistique* – Vuibert 2006.

L'affaire Castaneda :
utiliser la simulation pour déceler une discrimination

Niveau

Seconde.

Situation étudiée

En 1976 au Texas, un accusé d'origine mexicaine conteste le jugement du tribunal au motif que la désignation des jurés est discriminatoire envers les Américains d'origine mexicaine. Les élèves analysent les arguments statistiques et probabilistes qui apparaissent dans l'attendu de la Cour Suprême des États-Unis.

Type d'activité

Séance de « module » (demi groupes) utilisant le tableur.

Durée

1 heure.

Objectifs

Contenus mathématiques au programme

Simulations, fluctuations d'échantillonnage et stabilisation des fréquences.
Utilisation du tableur.

Enjeux citoyens

Utiliser ses connaissances mathématiques pour analyser une situation réelle par simulation et diagnostiquer une discrimination.

Capacités et attitudes

Exploiter une situation réelle.
Mettre en oeuvre les connaissances et les techniques de cours.
Exprimer son opinion et l'argumenter.

Organisation

Séance n°3 après un travail préalable sur l'utilisation des touches random et partie entière de la calculatrice pour simuler une expérience aléatoire
Cette activité se situe donc en **fin de séquence** sur la simulation et les différentes notions vont pouvoir être réinvesties.

Description des activités

Document élève (énoncé original qui sera un peu modifié par la suite)

L'affaire Castaneda contre Partida

En Novembre 1976 dans un comté du sud du Texas, Rodrigo Partida était condamné à huit ans de prison.
Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que 79,1% de la population de comté était d'origine mexicaine, sur les 870 personnes convoqués pour être jurés lors d'une certaine période de référence, il n'y eût que 339 personnes d'origine mexicaine.

But du T.P. : On se propose de simuler 100 fois le tirage de 870 jurés pour voir s'il est possible que le hasard ne désigne que 339 Américains d'origine mexicaine et donc savoir s'il y a eu discrimination.

Partie A

1. Calculer 79,1% de 870. On arrondira à l'entier le plus proche. A quoi correspond ce nombre ?

2. On rappelle que la fonction Random de la calculatrice génère un nombre aléatoire entre 0 et 1. Sur le tableur, ALEA () génère un nombre aléatoire entre 0 et 1.
Expliquer alors pourquoi on peut simuler la désignation d'un juré de ce comté sur la cellule A1 à l'aide de la formule =ENT(ALEA() + 0,791).
On pourra s'aider de schémas pour représenter des intervalles.

Partie B

1. Après avoir rentré la formule et « tiré la poignée » verticalement, simuler le tirage au sort des 870 jurés.

2. Entrer la formule =SOMME(A1:A870) en cellule A871.
Qu'obtient-on ? En appuyant plusieurs fois sur la touche F9 (le tableur recalcule alors tout), indiquer autour de quel nombre le total de jurés d'origine mexicaine semble osciller.

3. Après avoir sélectionné toute la colonne A1:A871, « tirer la poignée » horizontalement pour simuler 100 fois le tirage de 870 jurés.

4. Représenter graphiquement les données de la dernière ligne.
(Commencer par sélectionner la ligne 871 puis utiliser l'assistant graphique.)

5. Obtient-on obligatoirement 688 jurés d'origine mexicaine ? Comment appelle-t-on ce phénomène ?

6. Quel phénomène explique que le nombre total de jurés d'origine mexicaine semble rester « assez proche » de 688 ?

Selon vous, y a-t-il eu discrimination à l'égard des Américains d'origine mexicaine ? Expliquer.

Déroulement de la séance

Avec le premier groupe de module, les élèves ont rapidement montré leurs difficultés à utiliser l'outil informatique (« J'ai perdu mon mot de passe », « C'est où Excel ? », « Moi j'ai que des 0, c'est normal ? »). On a donc du recourir au vidéo projecteur pour montrer aux élèves ce qu'ils devaient obtenir avec l'assistant graphique, contournant ainsi toutes les manipulations des questions 2 à 5 pour accéder directement aux résultats de la simulation.
Avec le second groupe, bien plus autonome, les difficultés techniques se sont réglées très rapidement ; les élèves pouvaient se faire assister au besoin par leur voisin. En revanche, la rédaction était individuelle.

Quelques extraits de copies commentés

Au cours de nos travaux au sein du groupe « statistique et citoyenneté » de l'IREM Paris-Nord, nous nous sommes attachés à comprendre les raisonnements des élèves et nous n'avons pas toujours été d'accord entre nous pour dire si tel élève qui avait écrit telle chose avait oui ou non compris (on n'est plus du tout dans le cadre habituel des raisonnements hypothético-déductifs). Nous restons persuadés que l'analyse des copies des élèves, même si c'est peut-être parfois plus difficile dans ce chapitre, constitue le meilleur moyen de comprendre la perception par nos élèves du hasard et donc de les faire progresser.
Les extraits ci-dessous proviennent des copies d'élèves remplies pendant le TP.

Pour la première question, on attendait non seulement que le calcul de pourcentage soit effectué correctement, mais aussi que les élèves indiquent que 688 correspond à une proportion de 79,1 % d'une population de 870 personnes.

1) $\frac{79,1 \times 870}{100} = 688$. Le nombre obtenu correspond au nombre de mexicains qu'il y a dans les 870 jurés.

Le calcul de pourcentage a été correctement effectué mais l'interprétation pose problème, comme dans la copie ci-dessous où l'élève affirme que 688 est le nombre qu'on « aurait du avoir », ce qui a conduit à **modifier l'énoncé** (voir à la fin).

1) 79,1 % de 870 c'est $\frac{79,1}{100} \times 870 = 688$

A quoi correspond 688 ?

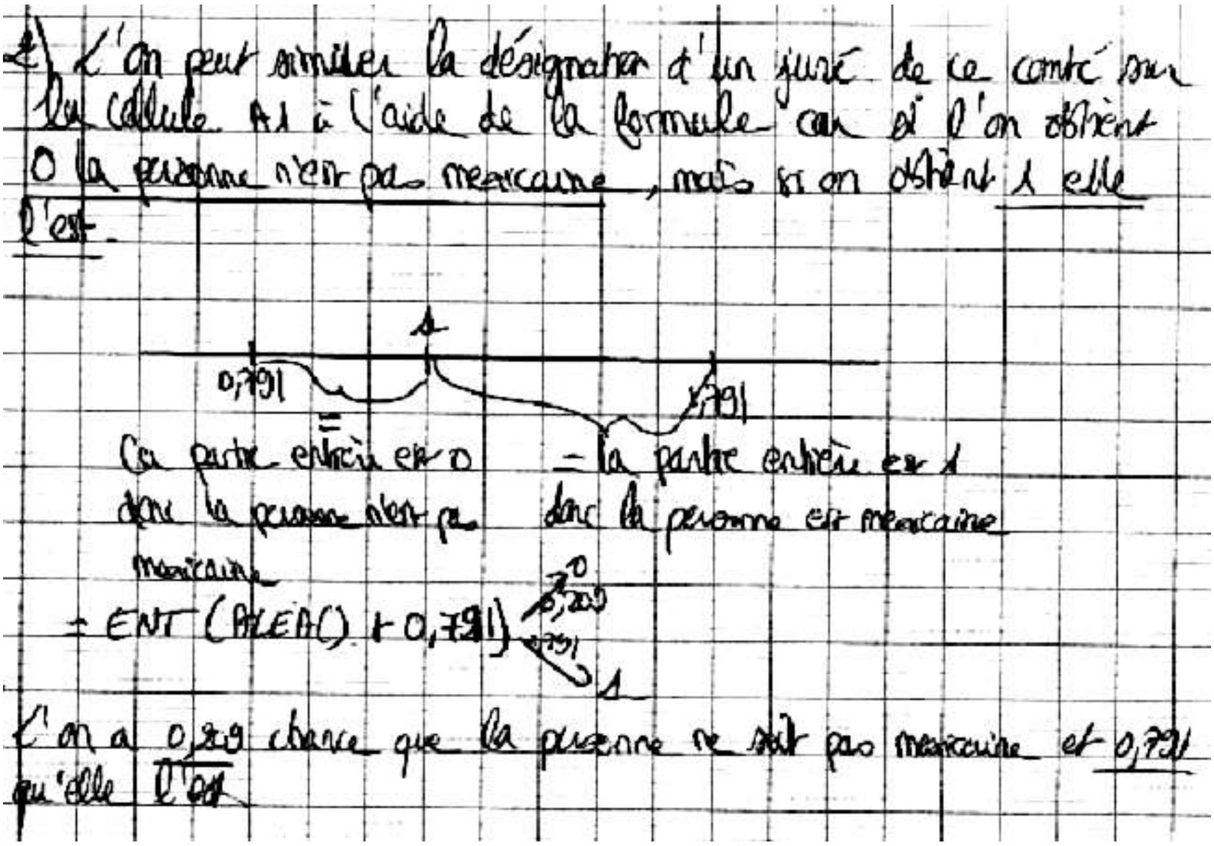
688 correspond au nombre de juré ~~non-mexicains~~. américain d'origine mexicaine qui aurait du avoir

Mais plus tard, l'élève semble considérer dans la même copie qu'il n'est pas anormal d'obtenir un nombre différent de 688 :

> 5) Non, on obtient pas toujours 688 jurés d'origine mexicaine, ce phénomène s'appelle la fluctuation d'échantillage

Ici, il est difficile de dire, à la lecture de la copie, si l'élève a eu besoin de lire la question pour se rappeler ce qu'est la fluctuation d'échantillonnage (la faute d'orthographe de l'élève plaide plutôt dans ce sens) ou si, plus simplement encore, l'élève n'a pas rédigé la réponse à la première question avec précision en essayant d'aller au plus court.

Pour la deuxième question, certains élèves ont su réinvestir les schémas réalisés lors des séances précédentes, en voici une illustration :



Cette copie montre que l'élève a su réinvestir le travail des séances précédentes.
Cependant, de nombreux élèves ont éludé la question dans le premier groupe.

Pour les questions 5 et 6, la formulation de la question doit être reprise, car les élèves se contentent, logiquement, d'énoncer les phénomènes de fluctuation d'échantillonnage et de stabilisation de fréquences sans aucun développement, comme dans la copie suivante.

> 5) L'on obtient pas obligatoirement 688 jurés d'origine mexicaine
> on appelle ce phénomène la fluctuation d'échantillonage.
> nomènes
> 6) Le phénomène qui explique le nombre total de jurés d'origine mexicaine qui semble rester "assez proche" de 688 est la stabilisation des fréquences.

Dans quelques copies cependant l'explication est plus détaillée :

> 5) Non on obtient pas obligatoirement 688 juries d'origine mexicaine tout depend de la simulation, la frequence varie donc c'est ce que l'on appelle la Fluctuation d'echantillonnage.
> 6) le phenomene qui explique que le nombre total de juré d'origine mexicaine sembla rester « assez proche » de 688 est la stabilisation des frequences. Plus le nombre de mexicain est grand plus la valeur est proche de 688.

Le point le plus intéressant est certainement la réponse des élèves à la dernière question : « Selon vos observations y a-t-il eu discrimination à l'égard des Américains d'origine mexicaine ? Expliquer ».

Tous les élèves concluent à la discrimination mais les justifications sont très diverses.
Dans cette copie, l'observation des résultats des simulations n'est pas du tout utilisée ; seule la première question reposant sur un « vrai calcul » a été exploitée, le travail sur les fluctuations ne sert pas d'argument :

> 7) Selon nos observations, il y a eu discriminations à l'égard des Americains d'origine mexicaine car selon la théorie il y a 79,1% de personnes d'origine mexicaine, or il n'y eu que 339 personnes d'origine mexicaine qui ont été convoquées.

En revanche, dans la copie suivante, l'élève synthétise tous les résultats établis ou observés :

Oui il y a eu discrimination à l'égard des américains d'origine mexicaine car au hasard on devrait avoir autour de 688 Américain d'origine mexicaine or il n'y a eût que 339 personnes d'origine mexicaine

Enfin, la copie qui suit est particulièrement satisfaisante : c'est l'une des rares copies où toute la séquence sur la simulation est réinvestie.

Selon mes observations il y a eu discrimination a l'egard des Americains d'origine mexicaine car cela est très peu probable d'obtenir ce resultat d'après les très grand nombre de simulation faite le nombre de mexicain reste dans les environs de 688

Prolongements

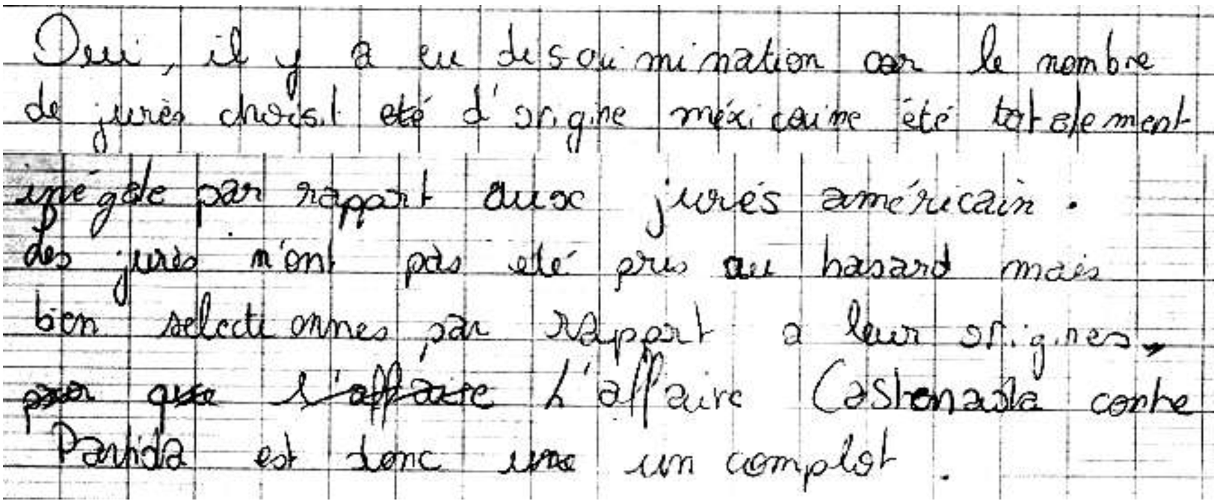
Lors de la séance suivante la décision de la cour d'appel a été donnée, ce qui aux yeux des élèves confirme la validité de leur travail.
Cette activité aurait du donner lieu à un débat en **ECJS** (éducation civique juridique et sociale) où l'attendu de la cour d'appel ci-dessous serait donné. Ce qui est aussi important pour tous, élève ou enseignant, est de comprendre complètement les motivations de la cour d'appel qui a finalement donné raison à la défense de Partida (voir ci-dessous).

Extrait de l'attendu de la cour suprême dans l'affaire Castaneda contre Partida

« 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) fois la probabilité de sélectionner un américain mexicain (ici

0,791) fois 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}. »

Cependant, la cour exclut une démonstration mathématique d'une discrimination raciale. Cela doit être souligné dans la mesure où les élèves semblent ressortir de l'activité avec la conviction d'avoir étudié un cas manifeste de racisme ou de complot comme dans cette copie :


Oui, il y a eu discrimination car le nombre de jurés choisit été d'origine méxicaine été totalement inégale par rapport aux jurés américain. Les jurés n'ont pas été pris au hasard mais bien selectionnes par rapport a leur origines, ~~par que~~ ~~l'affaire~~ L'affaire Castenada contre Partida est donc ~~une~~ un complot.

Pour contrer cette thèse, lors d'un **débat** par exemple, plusieurs pistes sont possibles. Certaines relèvent des mathématiques et d'autres entrent davantage dans le champ de l'éducation à la citoyenneté, ce qui ne doit pas nous interdire de les étudier en cours de mathématiques.

Concernant l'aspect statistique, l'un des principaux objectifs de cet enseignement n'est-il pas de développer l'**esprit critique** à l'égard des chiffres, en les faisant réagir à des phrases du type : « 100% des gagnants du loto ont tenté leur chance », « Momo machine lave 30% plus blanc », « plus de 99 % des français ont plus de jambes que la moyenne » etc.

Il est alors aussi de notre responsabilité d'être critique à l'égard des chiffres sur lesquels toute la simulation repose. Reprenons donc l'énoncé élève :

> « Alors que 79,1% de la population de comté était d'origine mexicaine, sur les 870 personnes convoqués pour être jurés lors d'une certaine période de référence, il n'y eût que 339 personnes d'origine mexicaine. »

Que signifie : « lors d'une certaine période de référence » ? En réalité, cette période de référence correspond à l'étude des listes des jurés lors des 11 années qui ont précédé le jugement de Partida. Sur ces listes, 339 citoyens d'origine mexicaine ont été comptabilisés sur 870. Or, non seulement il n'est pas raisonnable de penser que durant cette longue période la proportion de citoyens d'origine mexicaine est resté constante et égale à 79,1 %, mais en plus, sur la période de deux ans et demi précédent le jugement de l'affaire opposant le shérif Castaneda au prévenu Partida, la proportion de citoyens d'origine mexicaine ayant été membre d'un jury dans ce même comté était d'environ 56%. Cet argument a été soutenu par

l’accusation sans succès car la sous représentation des citoyens d’origine mexicaine restait trop importante.
Une deuxième piste pour contredire la thèse du complot repose sur l’étude des **modalités de sélection des jurés**. D’abord des listes de citoyens sont éditées de manière aléatoire mais ensuite, conformément à la loi, le juge du comté vérifie que le citoyen est alphabétisé et de « bonnes mœurs ». Ainsi, la loi elle-même impose une sous représentation d’une partie de la population qui est moins alphabétisée.
Enfin, il faut remarquer que la cour suprême ne conclut pas à la démonstration formelle de discrimination raciale. Elle précise, en effet: « Étant donné les nombreuses facettes de la motivation humaine, il serait peu approprié de prendre comme loi établie que des humains appartenant à un groupe ne pratiqueront pas de discrimination à l’égard des membres d’un autre groupe ».

Bilan

La prise de conscience par les élèves de l’**utilité de la statistique** pour répondre à ce problème de justice est indéniable. Il faut noter que cette activité qui a été réalisée en fin d’année, c’est-à-dire à une période où de nombreux élèves très en difficulté ne se sentent plus très impliqués, a néanmoins **vivement intéressé** tous les élèves à cause du sujet abordé. Ainsi, même quand ils n’ont pas su se sortir des aspects techniques du tableur, ils ont toujours essayé de voir ce qu’il fallait observer, avec pour obsession de savoir s’il y avait eu discrimination et donc répondre à la dernière question.
La statistique permet souvent de « récupérer » des élèves dont la motivation s’est étiolée au fil des semaines. Mais les élèves sont toujours très **sensibles aux discriminations et aux injustices**. Ceci explique, sans doute, l’intérêt que les élèves ont porté à cette activité.
Un seul regret cependant, celui de n’avoir pas pu exploiter davantage cette activité. Cette affaire peut faire l’objet d’un travail **interdisciplinaire** très fort entre les mathématiques, l’**ECJS**, voire l’**anglais** pour la traduction de certains extraits d’attendus (le programme de seconde d’anglais comporte un chapitre sur le lien social
ftp://trf.education.gouv.fr/pub/edutel/bo/2002/hs7/anglais.pdf
pour lequel les collègues d’anglais qui ~~m’~~ nous ont aidé dans la traduction de certains textes ont été très intéressés par cette activité), et le **français** où les élèves pourraient par exemple s’essayer à l’écriture de plaidoiries et montrer leurs progrès dans l’utilisation des figures de style. Toutes ces pistes doivent être étudiées suffisamment tôt dans l’année pour permettre une action coordonnée des enseignants intéressés.
Ainsi, il reste encore de nombreuses pistes à exploiter et surtout une réponse à trouver : Rodrigo Partida était-il coupable ou innocent ?

Après la séance et l’étude des copies des élèves, un nouveau document élève a été édité afin de tenir compte des imperfections décelées.

Nouveau « document élève » (énoncé modifié après expérimentation en classe)

L'affaire Castaneda contre Partida
L'ensemble des faits évoqués ci-dessous est réel.

En Novembre 1976 dans un comté du sud du Texas, Rodrigo Partida était condamné à huit ans de prison. Il attaqua ce jugement au motif que la désignation des jurés de ce comté était discriminante à l'égard des Américains d'origine mexicaine. Alors que 79,1% de la population de comté était d'origine mexicaine, sur les 870 personnes convoqués pour être jurés lors d'une certaine période de référence, il n'y eût que 339 personnes d'origine mexicaine.(*)

But du T.P.
On se propose de simuler 100 fois le tirage de 870 jurés pour voir s'il est vraisemblable que le hasard ne désigne que 339 Américains d'origine mexicaine dans une population où 79,1% est d'origine mexicaine.

Partie A : Simulation de la désignation d'un juré

1. A combien de jurés d'origine mexicaine pourrait-on s'attendre en choisissant au hasard 870 personnes dans la population ?
2. On rappelle que la fonction « Random » de la calculatrice génère un nombre aléatoire entre 0 et 1. Sur le tableur, on obtient la même fonction avec ALEA (). On rappelle que la fonction « Int » de la calculatrice renvoie la partie entière d'un nombre positif, elle est notée ENT sur le tableur.
Expliquer alors pourquoi on peut simuler la désignation d'un juré de ce comté sur la cellule A1 à l'aide de la formule =ENT(ALEA() + 0,791). *On pourra s'aider de schémas pour représenter des intervalles.*

Partie B : Simulation de 100 séries de 870 désignations de jurés

1. Après avoir rentré la formule et « tiré la poignée » verticalement, simuler le tirage des 870 jurés.
2. Dans la cellule A871 entrer la formule =SOMME(A1:A870). Qu'obtient-on ? En appuyant plusieurs fois sur la touche F9 (le tableur recalcule alors tout) indiquer autour de quel nombre le total de jurés d'origine mexicaine semble osciller.
3. Après avoir sélectionné toute la colonne A1:A871, « tirer la poignée » horizontalement pour simuler 100 fois le tirage de 870 jurés.
4. Représenter avec un nuage de points les données de la dernière ligne. (Commencer par sélectionner la ligne 871 puis utiliser l'assistant graphique).
5. Entre quelles bornes pourrait-on envisager le nombre de jurés mexicains sans qu'on puisse évoquer un problème de discrimination ?
6. Recopier et compléter la phrase : « d'après notre simulation de 100 élections de 870 jurés on n'a jamais obtenu moins de …. jurés d'origine mexicaine ».
7. Obtient-on obligatoirement 688 jurés d'origine mexicaine ? Qu'observez-vous ?
8. a) Est-il arrivé au hasard de distribuer un nombre de jurés d'origine mexicaine comparable à celui obtenu dans ce comté du Texas ?
b) Comment expliquez-vous cette situation ?

* source : *Prove it with figures* – H.Zeisel, D Kaye.

Statistique et discrimination : étude d'un document de justice

Niveau

Terminale S.

Situation étudiée

En 1976 au Texas, un accusé d'origine mexicaine conteste le jugement du tribunal au motif que la désignation des jurés est discriminatoire envers les Américains d'origine mexicaine. Les élèves analysent les arguments statistiques et probabilistes qui apparaissent dans l'attendu de la Cour Suprême des États-Unis.

Type d'activité

Exercice.

Durée

Exercice « sur feuille » de 20 minutes.
Correction et discussion d'environ 20 minutes.

Objectifs

Contenus mathématiques au programme

Loi binomiale, probabilités et fluctuations d'échantillonnage.

Enjeux citoyens

Étudier un texte judiciaire à contenu mathématique.
Voir en quoi des arguments mathématiques peuvent s'appliquer à des cas de société, en l'occurrence une question de discrimination.
Réfléchir à la notion de « preuve statistique » et au sens des très faibles probabilités.

Capacités et attitudes

Exploiter une situation réelle.
Lire un texte judiciaire à caractère scientifique.
Mettre en oeuvre les connaissances de cours.

Organisation

Les élèves (terminale S) connaissent la loi binomiale.
- Exercice à traiter, à la fin d'un cours, sur une photocopie distribuée.
- Rendu des copies et correction, suivi d'une discussion, à la séance suivante.

Description des activités

Séance 1 : recherche d'un exercice (20 minutes)

L'énoncé suivant est distribué aux élèves qui y réfléchissent de façon individuelle pendant une vingtaine de minutes. Les copies sont ramassées (mais ne seront pas « notées »).

Énoncé élève

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)[4] :

« *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} ? ....
....
....
....
La constitution des jurys est-elle faite au hasard ? ....
....

4 In : "Prove It with Figures (Statistics for Social Science and Behavioural Sciences)" - Hans Zeisel, D. H. et D. Kaye - Springer 2006.

Déroulement et commentaires

Travailler sur un texte d'aspect littéraire était complètement inhabituel mais le contexte a beaucoup motivé les élèves, y compris les « non matheux ». Tout le monde s'est rapidement mis au travail en étant tout de même rassuré par le fait que ce ne serait pas « noté ».

Séance 2 : correction et discussion (20 minutes)

Énoncé élève

Le corrigé suivant a été 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 : $P(X \leq 351)$ [5]
• *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.

Déroulement et commentaires

Les calculs purement techniques sur la loi binomiale ne posent pas trop de problèmes, c'est la réponse à la dernière question (et le sens que l'on peut donner à une probabilité de 10^{-140}) qui donne l'occasion d'une discussion assez riche.
Voici quelques exemples de copies.

La copie suivante peut-être considérée comme une « bonne » copie. L'élève a bien réagi au caractère invraisemblable de la probabilité obtenue sous l'hypothèse d'un tirage au sort « équiprobable » mais porte une accusation de « racisme » qui conduira à une discussion sur la manière dont les jurés sont « choisis ». Il y a bien tirage au sort mais on demande en particulier que les jurés comprennent et lisent correctement l'anglais, ce qui exclut une bonne partie de la population d'origine mexicaine.

[5] Plus précisément, pour une variable aléatoire X de loi binomiale de paramètres 870 et 0,791 , on a : $P(X \leq 339) \approx 4{,}2.10^{-145}$.

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

Quels sont les paramètres de la loi binomiale ? $B(870, 79,1\%)$

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

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

A quoi correspond la « différence de 29 écarts types » ? $E(X) - 339 = 29\ \sigma(X)$

A quel évènement correspond la probabilité 10^{-140} ? La probabilité que la valeur réel a un décalage de 29 écarts type.

La constitution des jurys est-elle aléatoire ? La constitution des jurys est donc très sélectifs. Il n'y a pas d'aléatoire.
RACISME

Pas "totalement" aléatoire : voir discussion.

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 $B(870 ; 0,791)$ car : il s'agit d'expériences aléatoires qui ~~...~~ peuvent être considérées comme indépendantes par rapport au nombre d'habitants du Texas avec 2 résultats possibles : réussite : mexicain (0,791) et ~~...~~ échec : américain (0,209)

Quels sont les paramètres de la loi binomiale ? $B(870 ; 0,791)$

A quel calcul correspond la valeur 688 ? L'espérance (moyenne) $E(X) = 870 \times 0,791 \approx 688$

Effectuer le calcul de l'écart type. $\sqrt{870 \times 0,791 \times (1 - 0,791)} = 11,99$

A quoi correspond la « différence de 29 écarts types » ? $\frac{688 - 339}{12} = 29$
(nb attendu − nb observé) / écart-type

A quel évènement correspond la probabilité 10^{-140} ? $10^{-140} = \frac{1}{10^{140}}$ donc c'est une chance sur 10^{140} qu'un tel éloignement ait lieu.

La constitution des jurys est-elle aléatoire ?
Non

Pourquoi ?

Dans la copie suivante (aux réponses partielles), l'élève ne fait pas le lien avec la probabilité valant 10^{-140}. Il semble qu'il n'y ait rien de choquant dans la constitution des jurys. A la question « La constitution des jurys est-elle aléatoire ? », on trouve comme réponse : « oui, car elle est soumise à un tirage au sort ». Cette réponse est évidemment exacte. C'est la raison pour laquelle on a modifié la dernière question de l'énoncé, qui de « La constitution des jurys est-elle aléatoire ? » est devenue « La constitution des jurys est-elle faite au hasard ? ».

Définir la variable aléatoire qui, dans cette situation, suit une loi binomiale. variable X : nombre de juré mexicain tirés au hasard dans l'échantillon d'américains mexicains

Quels sont les paramètres de la loi binomiale ? B(870 ; 79,1 %)

A quel calcul correspond la valeur 688 ? Espérance de E(X) = 870 × 79,1 % = 688

Effectuer le calcul de l'écart type. σ(X) = √(n.p.q) = √(870 × 79,1 % × 20,9 %) ≈ 12

A quoi correspond la « différence de 29 écarts types » ? Il y a 29 fois l'écart type entre les valeurs de la réalité et celle théorique

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

La constitution des jurys est-elle aléatoire ? Oui, car elle est soumis à un tirage au sort

Ce tirage est-il totalement « au sort » ?

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 (probabilité pour qu'une armée de singes tapant sur des machines à écrire produise l'œuvre de Victor Hugo).

Il est en tout cas essentiel de signaler que face à une telle probabilité, il faut absolument chercher une autre explication possible que le simple hasard. On a là une « preuve » statistique proche d'une quasi-certitude : il n'y a pas équiprobabilité dans le tirage au sort des jurés.

Bilan et perspectives

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 répondus 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.

Éléments de bibliographie

Internet

www.statistix.fr
http://www.inegalites.fr/

Livres et brochures

• Brochures APMEP :
Brochure n° 166 – ***Maths à Crédit*** – APMEP – IREM de Caen (niveau collège).
Brochure n° 156 – ***Statistique au lycée*** – APMEP 2005.
• Brochures IREM Paris-Nord :
Brochure n° 102 – ***Simulation et statistique en seconde*** – 2000.
Brochure n° 112 – ***Enseigner la statistique au lycée : des enjeux aux méthodes*** - 2001.
• Gilles Dowek – ***Peut-on croire les sondages ?*** – Les petites pommes du savoir – Le Pommier 2002.
• Philippe Dutarte – ***L'induction statistique au lycée illustrée par le tableur*** – Didier 2005.
• É. Janvresse et T. de la Rue – ***La loi des séries, hasard ou fatalité ?*** – Les petites pommes du savoir – Le Pommier 2007.
• Claudine Schwartz – ***Pratiques de la statistique*** – Vuibert 2006.
• Télérama hors série – ***Quelle égalité voulons-nous ?***