

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 ?

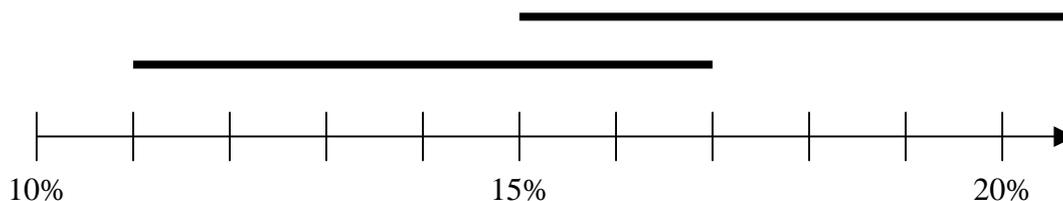
3. Au premier tour de l'élection présidentielle de 2002, L. Jospin a obtenu 16,18% des voix et J.-M. Le Pen 16,86%.

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 $\frac{98}{98 + 717} \approx \mathbf{12\%}$ 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 répondu. 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 $500x + 250 = 338$.
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 $500x + 250 = 338$.

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

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	$P(X \leq 184)$	0,647			
2	$P(X \leq 174)$	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 $P(175 \leq X \leq 184) = P(X \leq 184) - P(X \leq 174) \approx 0,319$.

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.