Probabilités conditionnelles
Problème de recherche : l'enquête de santé confidentielle

L'agence régionale de santé (ARS) veut estimer la proportion p de jeunes de 18 à 25 ans qui ont déjà pris des antibiotiques sans ordonnance (restes d'un ancien traitement, médicaments d'un proche…). Personne n'osera répondre « oui » franchement à un enquêteur. On utilise donc un protocole de **réponse aléatoire** :

— chaque jeune lance une pièce équilibrée, en cachette ;
— s'il obtient **pile**, il répond honnêtement à la question « Avez-vous déjà pris des antibiotiques sans ordonnance ? » ;
— s'il obtient **face**, il répond « oui », quoi qu'il en soit.

L'enquêteur ne voit que la réponse, jamais le résultat de la pièce. Sur 1 000 jeunes interrogés, on obtient 600 réponses « oui ».

La mission

1. Exprimer, en fonction de p, la probabilité qu'un jeune choisi au hasard réponde « oui ».
2. En déduire une estimation de la proportion p.
3. Un journaliste veut publier la liste des jeunes qui ont répondu « oui » en titrant « Ils prennent des antibiotiques sans ordonnance ». Quelle est la probabilité qu'un jeune ayant répondu « oui » l'ait réellement fait ? Le protocole protège-t-il les personnes interrogées ?

Pour aller plus loin

Une autre ARS utilise un dé : sur 1, 2, 3 ou 4, on répond honnêtement ; sur 5, on répond « oui » ; sur 6, on répond « non ». Elle obtient 30 % de « oui ». Estimer la proportion de jeunes concernés dans cette région.

Coups de pouce

Bloqué ? Des indices progressifs sont disponibles en ligne. N'en ouvrir qu'un à la fois, et seulement après avoir cherché !

Éléments de correction

On note P : « la pièce tombe sur pile », A : « le jeune a déjà pris des antibiotiques sans ordonnance », O : « le jeune répond oui ».

1. Probabilité de répondre oui

Arbre : P (0,5) puis A (p) → oui, ou $\overline{A}$ ($1-p$) → non ; $\overline{P}$ (0,5) → oui dans tous les cas.
$P(O) = P(P \cap O) + P(\overline{P} \cap O) = 0{,}5 \times p + 0{,}5 \times 1 = 0{,}5p + 0{,}5$.

2. Estimation de p

La fréquence observée des « oui » est $\frac{600}{1\,000} = 0{,}6$. On résout $0{,}5p + 0{,}5 = 0{,}6$: $p = 0{,}2$.
Environ **20 %** des jeunes de 18 à 25 ans auraient déjà pris des antibiotiques sans ordonnance.

3. Que signifie un « oui » ?

On cherche $P_O(A)$. Un jeune concerné répond toujours « oui » (avec pile comme avec face), donc $P(A \cap O) = P(A) = 0{,}2$. Un jeune non concerné répond « oui » s'il obtient face : $P(\overline{A} \cap O) = 0{,}8 \times 0{,}5 = 0{,}4$. On retrouve $P(O) = 0{,}6$, et

$$P_O(A) = \frac{P(A \cap O)}{P(O)} = \frac{0{,}2}{0{,}6} = \frac{1}{3} \approx 0{,}33.$$

Deux jeunes sur trois qui ont répondu « oui » ne sont pas concernés : **la liste du journaliste accuserait surtout des innocents**. Le protocole protège chaque personne interrogée, tout en donnant à l'ARS une estimation fiable pour l'ensemble de la population.

Pour aller plus loin

Avec le dé : $P(O) = \frac{4}{6}p + \frac{1}{6}$. On résout $\frac{2}{3}p + \frac{1}{6} = 0{,}3$: $\frac{2}{3}p = \frac{2}{15}$, donc $p = 0{,}2$. Ici aussi, environ 20 % des jeunes. Avantage de ce protocole : répondre « non » ne prouve pas non plus qu'on n'est pas concerné.

Ce qu'on a utilisé du chapitre

- Arbre pondéré ; formule des probabilités totales.
- Probabilité conditionnelle $P_O(A) = \dfrac{P(A \cap O)}{P(O)}$.