Pourcentages et fréquences
Problème de recherche : le test « fiable à 99 % »

Un laboratoire met au point un test de dépistage d'une maladie qui touche 1 % de la population. Il annonce fièrement :

— « Notre test est positif pour 99 % des personnes malades. »

— « Notre test est négatif pour 98 % des personnes qui ne sont pas malades. »

Un journal titre : « Test fiable à 99 % : si votre test est positif, vous êtes presque sûr d'être malade. »

La mission

1. Le titre du journal est-il correct ? Parmi les personnes dont le test est positif, quelle est la proportion de personnes réellement malades ?
2. On utilise ce test dans un service hospitalier où la maladie est beaucoup plus fréquente. À partir de quelle proportion de malades le titre du journal devient-il acceptable, c'est-à-dire qu'au moins 90 % des personnes testées positives sont malades ?

Pour aller plus loin

Pour la population générale (1 % de malades), le laboratoire peut améliorer l'un des deux pourcentages annoncés. Lequel faut-il améliorer en priorité ?

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

1. Sur 10 000 personnes

	Test positif	Test négatif	Total
Malades	99	1	100
Non malades	198	9 702	9 900
Total	297	9 703	10 000

Parmi les 297 personnes testées positives, 99 sont malades : la fréquence conditionnelle est $\frac{99}{297} = \frac{1}{3} \approx 33$ %. Le titre est **faux** : deux personnes testées positives sur trois ne sont pas malades. Les non-malades sont si nombreux que 2 % d'entre eux (198) dépassent largement le nombre de vrais malades détectés.

2. En faisant varier la proportion de malades

Pour une proportion p de malades et une population de 10 000 personnes : malades positifs $0{,}99 \times 10\,000p$, non-malades positifs $0{,}02 \times 10\,000(1-p)$. La fréquence cherchée est

$$F = \frac{0{,}99p}{0{,}99p + 0{,}02(1-p)}.$$

Le tableur donne par exemple : $p = 5$ % : $F \approx 72$ %; $p = 10$ % : $F \approx 85$ %; $p = 15$ % : $F \approx 89{,}7$ %; $p = 16$ % : $F \approx 90{,}4$ %.

Le titre devient acceptable à partir d'environ **15,4 %** de malades. (Calcul exact : $F \geqslant 0{,}9 \iff 0{,}99p \geqslant 0{,}9\,(0{,}97p + 0{,}02) \iff 0{,}117p \geqslant 0{,}018 \iff p \geqslant \frac{0{,}018}{0{,}117} \approx 0{,}154$.)

Pour aller plus loin

Avec 1 % de malades, le problème vient des faux positifs. Passer de 99 % à 100 % de détection des malades ne change presque rien ($F = \frac{100}{298} \approx 33$ %). En revanche, passer de 98 % à 99,9 % de tests négatifs chez les non-malades donne 9,9 faux positifs au lieu de 198 et $F = \frac{99}{108{,}9} \approx 91$ %. Il faut améliorer **le second pourcentage**.

Ce qu'on a utilisé du chapitre

— Tableau croisé d'effectifs; fréquences marginales et fréquences conditionnelles.

— Proportion d'une proportion; utilisation d'un tableur pour faire varier un paramètre.