

Énoncé

Le sujet

L'inférence bayésienne, qui fait le lien entre statistiques et intelligence artificielle, est souvent utilisée dans le cadre de diagnostics médicaux. En effet, elle permet d'établir la probabilité d'être porteur d'une maladie sachant qu'un certain nombre d'anomalies ont été détectées.

Parmi les femmes de 40 ans ayant effectué une mammographie, 1 % ont un cancer du sein. À la suite de mammographies sur échantillon, on a établi que :

- pour 82 % des femmes ayant un cancer du sein, la mammographie détecte une anomalie ;
- pour 9 % des femmes n'ayant pas de cancer du sein, la mammographie détecte une anomalie.

Lors des tests, plusieurs cas peuvent se présenter :

- les vrais positifs sont les sujets qui sont malades et à qui on détecte une anomalie ;
- les faux positifs sont les sujets qui ne sont pas malades et à qui on détecte une anomalie ;
- les vrais négatifs sont les sujets qui ne sont pas malades et à qui on ne détecte pas d'anomalie ;
- les faux négatifs sont les sujets qui sont malades et à qui on ne détecte pas d'anomalie.

La sensibilité d'un test est la probabilité que l'on détecte une anomalie à une personne malade.

La spécificité du test est la probabilité que l'on ne détecte aucune anomalie à une personne non malade.

On suppose que 10 000 femmes de 40 ans ont effectué une mammographie.

1. Préciser les caractéristiques (sensibilité, spécificité) d'une mammographie.
2. **a.** Calculer le nombre de femmes atteintes d'un cancer du sein dans cet échantillon.
b. Calculer le nombre de femmes à qui l'on détecte une anomalie, sachant qu'elles sont atteintes d'un cancer du sein.
c. Calculer le nombre de femmes à qui l'on détecte une anomalie, sachant qu'elles n'ont pas de cancer du sein.
d. Compléter le tableau ci-dessous :

| | Anomalie détectée | Pas d'anomalie détectée | Total |
|-------------|-------------------|-------------------------|--------|
| Malades | | | |
| Non malades | | | |
| Total | | | 10 000 |

3. Une femme de 40 ans a subi une mammographie qui a détecté une anomalie. Quelle est la probabilité qu'elle soit atteinte d'un cancer du sein ?

4. La valeur prédictive positive du test dans une population donnée, notée VPP, est la probabilité qu'un individu, à qui l'on détecte une anomalie, soit effectivement malade. De même, la valeur prédictive négative du test dans une population donnée, notée VPN, est la probabilité qu'un individu, à qui l'on ne détecte pas d'anomalie, soit sain. Calculer les valeurs prédictives positive et négative d'une mammographie chez les femmes de 40 ans.

5. La CNIL, dans son document *Comment permettre à l'Homme de garder la main ?* évoque l'utilisation de l'intelligence artificielle dans le diagnostic médical :

« On annonce [...] que l'intelligence artificielle serait supérieure à l'homme pour le diagnostic de certains cancers ou pour l'analyse de radiographies. Dans le cas où ces annonces s'avéreraient exactes, il pourrait donc devenir hasardeux pour un médecin d'établir un diagnostic ou de faire un choix thérapeutique autre que celui recommandé par la machine, laquelle deviendrait dès lors le décideur effectif. »

Quelles pourraient être les dérives de cette utilisation dans le domaine médical ?

La bonne méthode

1. L'énoncé donne le pourcentage de femmes à qui l'on détecte une anomalie, sachant qu'elles ne sont pas malades.
2. **a.** L'échantillon de l'étude est de 10 000 personnes. L'énoncé précise le pourcentage de malades.
b. Parmi les femmes atteintes d'un cancer du sein, il faut déterminer celles à qui l'on détecte une anomalie.
c. Déterminer le nombre de femmes qui n'ont pas de cancer du sein. Puis, parmi ces femmes-là, il faut déterminer celles à qui l'on a

défecté une anomalie.

d. Synthétiser les valeurs trouvées précédemment dans le tableau, puis déduire par addition ou soustraction les valeurs manquantes.

3. Il faut calculer la probabilité que la femme soit atteinte d'un cancer sachant qu'on lui a détecté une anomalie.

4. La VPP (ou VPN respectivement) correspond à la probabilité que la personne soit malade (ou saine) sachant qu'on lui a détecté (ou non) une anomalie. On peut également déduire la VPP de la question précédente.

5. Penser à ce qui peut se passer concrètement si le médecin ne traite pas une personne qui a été classée parmi les « non malades » par l'intelligence artificielle, alors qu'elle est effectivement malade.