

Récurrence des cancers thyroïdiens

Contexte des données

Les cancers thyroïdiens voient leurs incidences augmenter à l'échelle du monde, cela en particulier dû à une amélioration des techniques diagnostiques. Malgré le faible taux de mortalité des cancers thyroïdiens, les récurrences sont courantes et augmentent les risques de décès. C'est pourquoi il peut être intéressant d'identifier les caractéristiques qui augmentent le risque de récurrence des cancers de la thyroïde. L'association américaine pour la thyroïde (ATA) a établi le système de référence pour la classification de ces cancers. Ces classes sont au nombre au trois et permettent une catégorisation des risques de cancers : faible, intermédiaire, fort. Cette catégorisation simple permet un triage des patient es mais ne permet pas d'identification du risque de récurrence du cancer.

Ici on s'intéresse à une étude prospective sur 383 patient es qui ont été suivi es pendant au moins 10 ans au sein d'une fenêtre temporelle de 15 ans. On mesure un grand nombre de caractéristiques liées aux types de cancers thyroïdiens.

Descriptif des données

17 colonnes et 383 lignes:

Nom de la colonne	Type de variable	Description
Age	Nombre entier	Âge de la patiente/du patient
Gender	Chaîne de caractères	Genre du patient : "F" = femme ; "H" = homme
Smoking	Chaîne de caractères	Habitude tabagique présente du patient : "No" = non-fumeur ; "Yes" = fumeur
Hx Smoking	Chaîne de caractères	Habitude tabagique passée du patient : "No" = jamais fumé ; "Yes" = déjà fumé
Hx Radiothreapy	Chaîne de caractères	Historique de radio des régions du cours et de la tête : "No" = aucune radio ; "Yes" = au moins une radio de cette région
Thyroid Function	Chaîne de caractères	État de la thyroïde : "Clinical Hyperthyroidism" = hyperthyroïdie clinique ; "Clinical Hypothyroidism" = hypothyroïdie clinique ; "Euthyroid" = thyroïde normale ; "Subclinical Hyperthyroidism" = hyperthyroïdie subclinique ; "Subclinical Hypothyroidism" = hypothyroïdie subclinique
Physical Examination	Chaîne de caractères	Résultat de la palpation physique : "Diffuse goiter" = goître diffus ; "Multinodular goiter" = goître à plusieurs nodules ; "Normal" = thyroïde normale ; "Single nodular goiter-left" = goître nodulaire à gauche ; "Single nodular goiter-right" = goître nodulaire à droite (article Wikipédia)
Adenopathy	Chaîne de caractères	Présence d'adénopathie (état anormal d'un nœud lymphatique ; article Wikipédia) à l'examen physique : "No" = aucune ; "Bilateral" = bilatérale ; "Left" = gauche ; "Posterior" = postérieure ; "Right" = droite ; "Extensive" = à toutes les positions mentionnées précédemment

Nom de la colonne	Type de variable	Description
Pathology	Chaîne de caractères	Sous-type pathologique de cancer : "Follicular" = folliculaire ; "Hurthel cell" = cellule de Hürthle ; "Micropapillary" = micro-papillaire ; "Papillary" = papillaire
Focality	Chaîne de caractères	Focalité du cancer (présence en un seul ou plusieurs endroits de la thyroïde) : "Multi-Focal" = tumeurs multifocales ; "Uni-Focal" = tumeur unifocale
Risk	Chaîne de caractères	Classification de risque d'après l'ATA : "Low" = faible ; "Intermediate" = intermédiaire ; "High" = haut
T	Chaîne de caractères	Critère T dans la classification TNM (voir explications ci-dessous) : "T1a" ; "T1b" ; "T2" ; "T3a" ; "T3b" ; "T4a" ; "T4b"
N	Chaîne de caractères	Critère N dans la classification TNM (voir explications ci-dessous) : "N0" ; "N1a" ; "N1b"
M	Chaîne de caractères	Critère M dans la classification TNM (voir explications ci-dessous) : "M0" = aucune métastase ; "M1" = métastase à distance
Stage	Chaîne de caractères	Stade du cancer (voir explications ci-dessous) : 'I' ; 'II' ; 'III' ; 'IVA' ; 'IVB'
Response	Chaîne de caractères	Réponse initiale au traitement : 'Excellent' = excellente ; 'Indeterminate' = indéterminée ; 'Biochemical Incomplete' = réponse biochimique incomplète (indicateurs biochimiques du cancer qui ne déclinent pas ; ou pas assez vite) ; 'Structural Incomplete' = réponse structurelle incomplète (= persistance de la pathologie ou de tissus tumoraux)
Recurred	Chaîne de caractères	Récidive : "No" = non ; "Yes" = oui (localement ou du fait de métastases)

Classification TNM du cancer de la thyroïde

La classification TNM est une échelle internationale pour classer les cancers en fonction de leurs caractéristiques anatomiques. Elle est basée sur trois critères : le critère **T**, qui mesure l'étendue de la tumeur primaire ; le critère **N**, qui identifie dans quelle mesure la tumeur atteint les nœuds (ou ganglions) lymphatiques voisins ; le critère **M**, qui indique s'il y a ou non présence de métastase.

En fonction du cancer étudié, les critères d'évaluation de chaque classe varient. En fonction des niveaux observés sur l'échelle TNM on peut définir un "stade" synthétique du cancer, qui est repris dans la colonne **Stage** (référence pour la classification du cancer de la thyroïde : <https://www.pathologyoutlines.com/topic/thyroidstaging.html>).

Travail demandé

Objectifs

- On pourra chercher les relations entre biomarqueurs en fonction du stade de la maladie
- On pourra prédire le stade de maladie en fonction de l'ensemble des variables biologiques
- On pourra chercher à identifier le rôle que peut jouer le fait de fumer dans la nature des cancer thyroïdiens

Les techniques à mobiliser ici sont surtout les ACM (ou AFDM) étant donné la présence de beaucoup de variables qualitatives, et les arbres de classification pour prédire les catégories d'intérêt. Il est possible d'utiliser une régression logistique en simplifiant une des variables en variable binaire.

Exercice à rendre

Vous présenterez vos résultats lors de la dernière séance le 4 avril : 20 minutes de présentation et 10 minutes de questions. Votre présentation devra comporter les parties suivantes : - Présentation du contexte et de la question que vous souhaitez poser et répondre - Présentation de vos données - Présentation de la méthode choisie (pourquoi cette méthode, expliquez succinctement son fonctionnement) - Présentation des résultats - Conclusion

Références

Article original : Borzooei, S., Briganti, G., Golparian, M. et al. Machine learning for risk stratification of thyroid cancer patients: a 15-year cohort study. Eur Arch Otorhinolaryngol 281, 2095–2104 (2024). <https://doi.org/10.1007/s00405-023-08299-w>

Citation du jeu de données : Borzooei, S. & Tarokhian, A. (2023). Differentiated Thyroid Cancer Recurrence [Dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5632J>.