

Analyse de données épidémiologique

Objectifs :

Ce projet visait à nettoyer une base de données volumineuse et à réaliser des analyses sur des données de santé de patients admis en unité de soins intensifs aux États-Unis. Ce projet, réalisé en groupe de 4 personnes entre février et juin 2023, avait pour contrainte de se réaliser uniquement sur le logiciel de programmation R. La restitution des résultats était sous la forme d'un rapport d'une dizaine de pages, et d'une présentation orale avec support visuel.

Démarche méthodologique :

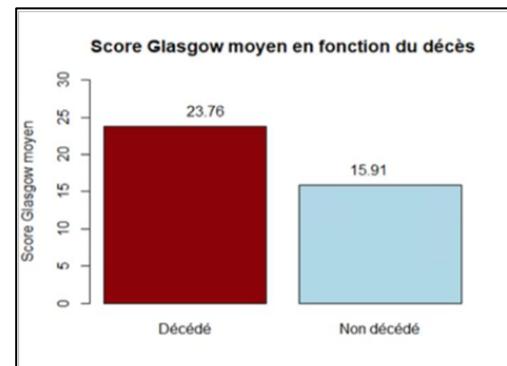
- Nettoyage de la base de données :

La première partie de notre travail consistait à comprendre l'organisation des données. Une fois familiarisé avec notre jeu de donnée, nous avons procédé au nettoyage. On avait beaucoup de valeurs manquantes, nous avons dû établir une règle : si une variable présentait plus de 5 % de valeurs manquantes, on la jugeait inexploitable et nous la supprimions. Nous avons également mis en forme nos données pour nous faciliter la suite du travail, en renommant par exemple les variables avec un libellé plus clair ou en supprimant les variables qui ne nous étaient d'aucune utilité pour la suite de nos analyses.

- Analyse des données :

En ce qui concerne l'analyse des données, notre objectif était de déterminer les variables qui avaient un impact sur la mort du patient. Pour cela, nous avons donc réalisé des tests statistiques. Par exemple, nous disposions d'une variable, le score Glasgow (compris entre 0 et 100), qui est un outil d'évaluation clinique utilisé pour évaluer rapidement et objectivement le niveau de conscience et l'état neurologique d'un patient. Afin de vérifier si ce score influe sur le décès du patient, nous avons d'abord calculé la moyenne pour les personnes décédées, puis pour les personnes non décédées, afin de voir s'il y avait une différence entre les deux groupes.

Nous avons ensuite réalisé le test de Student qui nous a permis de conclure que le score Glasgow avait un impact significatif sur la survie du patient.



Compétences développées

Ce projet fut le premier réalisé sur R, nous avons donc grandement augmenté nos compétences en programmation. Nous avons également augmenté nos compétences en analyse statistique de données via l'analyse descriptive, bivariée, ainsi que par les tests statistiques. Ce projet nous a permis de renforcer notre travail d'équipe, nos compétences en rédaction ainsi que la capacité à présenter des résultats statistiques.