

Test Technique Data Scientist 2025

1 Introduction

Ce test est à réaliser chez soi dans le temps imparti. Il n'y a pas de contraintes sur l'utilisation de langage ou de méthodes. Les données factices sont conçues selon le modèle de données de la BNDMR. Il n'existe pas de lien réel entre l'épidermolyse bulleuse et le plomb.

2 Scénario

La cellule opérationnelle de la BNDMR a été contactée par la filière maladie rare FIMARAD pour mener une étude approfondie de l'épidermolyse bulleuse (EB), une maladie dermatologique rare. D'après la nomenclature Orphanet des maladies rares, il y a plusieurs types d'EB dont l'EB simple et l'EB jonctionnelle (voir Annexe).

Nous souhaitons extraire les données démographiques liées à cette maladie à partir des données disponibles dans notre entrepôt de données de santé, et en tirer le maximum d'informations possible. En outre, une base complémentaire collectant le niveau de plomb dans le sang en $\mu\text{g/dL}$ nous a également été fournie par le professeur en charge de l'étude. Il nous est demandé de tester si, pour deux populations de patients traités avec le même agent, les différents types d'épidermolyses bulleuses affectent les effets du médicament.

3 Données

Pour cet exercice, trois fichiers de données sont fournis au format CSV :

- `patients.csv` : données démographiques des patients
- `diagnostics.csv` : données diagnostiques des patients
- `obs_plomb.csv` : données du niveau de plomb de patients sur 6 semaines

4 Attendus

- Explorer et décrire les données des patients atteints par EB.
- Évaluer l'impact du type d'EB sur l'évolution de la concentration de plomb chez les patients traités avec le même médicament.
- Présenter les résultats obtenus aux questions précédentes sous la forme d'un rapport.
Merci de fournir en annexe le code et/ou les documents tiers ayant permis sa génération.

Annexe : Relation entre les codes diagnostics (extrait)

