

draft

Lancelot Ravier

9 mai 2025

1 Introduction

Le projet expedition5300 a pour but d'analyser les effets de la vie en très haute altitude et les effets sur les habitants. Des recherches ont déjà été effectuées sur les adultes mais pas sur les enfants. Ici, on aura donc des données sur des enfants selon deux tranches d'âge.

- Etablir le risque, les conséquences et les marqueurs de l'anémie et de la déficience en fer chez l'enfant d'altitude.
- Etudier le développement de la masse d'hémoglobine et des volumes sanguins, de la viscosité sanguine et du système cardiovasculaire chez l'enfant d'altitude.

Hémoglobine : protéine permettant à l'oxygène de s'attacher au globule rouge
Hématocrite : taux de globule rouge dans le sang

2 Données

Les données ont été recueillies suivant plusieurs expéditions (2), par plusieurs personnes différentes. En conséquent, il y'a plusieurs fichiers de données à rassembler. Un travail a été fait en amont pour traiter les outliers donc pas besoin de porter d'analyses dessus. En ce qui concerne les différentes bases de données, elles possèdent les mêmes clés primaires donc il est possible de retrouver un patient dans les différentes bases de données par son code anonymisé de la manière suivante :

$id = 2 \text{ premières lettres du prénom} + \text{clé unique} + 2 \text{ premières lettres du nom}$

2.1 Préparation des données

On peut tout d'abord voir que les bases de données ne sont pas de la même taille :

- Anémie, Cardio, nfs_visco : 635
- template : 638

Il y'a donc 3 patients qui ne sont présents que dans la base de données template (BEJ0674, 640, QUJI740)

- Le 640 est sûrement quelqu'un qui s'est désisté avant de venir ?
- On a bien toutes les informations pour BEJ0674 et QUJI740 (peut être des gens qui se sont désistés ou qui se sont présentés deux fois)

Afin de continuer, il faut supprimer la ligne 638. Aussi, on supprimera les lignes appartenant aux deux autres identifiants vus au dessus.