

# Modèle linéaire et généralisations – TP1

## Charger les données

Le jeu de données contient 4 variables :

- **y** : mesure le risque sur la santé.
- **age**
- **tri** (triglycérides)
- **chol** (cholestérol)

Une valeur élevée de **y** signifie que le patient a un risque sérieux de développer une maladie dans les prochaines années. Cette mesure de risque est obtenue à l'aide d'analyses coûteuses. Notre objectif est de prédire **y** à partir de variables plus aisées à obtenir : **age**, **tri** et **chol**.

## Question 0

Charger les données du fichier "health.txt".

## Question 1

Tracer un graphe de corrélation entre les co-variables, identifier des relations entre les variables qui vous semblent intéressantes.

## Question 2

Estimer les paramètres d'un modèle linéaire exprimant **y** en fonction des autres covariables. Faire l'étude des diagnostics.

## Question 3

Essayer de détecter des relations non linéaires et des individus ayant des valeurs aberrantes ou une influence importante sur l'estimation.

## Question 4

Essayer d'ajouter des termes quadratiques pour obtenir un meilleur modèle. Refaire les diagnostics.

## Question 5

Proposer un modèle final et commenter.