

# État de santé d'une population

## Contexte des données

L'étude *National Health and Nutrition Examination Survey* (NHANES) est une enquête nationale américaine menée par le *National Center for Health Statistics* (NCHS), qui fait partie des *Centers for Disease Control and Prevention* (CDC). Débutée en 1960, cette étude est unique par son approche comprehensive de la santé et de la nutrition de la population américaine. Contrairement aux enquêtes traditionnelles qui reposent uniquement sur des questionnaires, NHANES combine des entretiens détaillés avec des examens médicaux et des analyses biologiques approfondies. L'objectif principal est de collecter des données représentatives sur la santé et l'état nutritionnel de la population des États-Unis, en échantillonnant chaque année un nouveau groupe représentatif de la population. Ces données sont essentielles pour comprendre les tendances épidémiologiques, développer des politiques de santé publique et guider la recherche médicale.

Ici, on a sélectionné un sous-ensemble de variables en rapport avec les indicateurs biochimiques et d'habitude de vie.

## Descriptif des données

10 colonnes et 2278 lignes:

Nom de la colonne	Type de variable	Description
SEQN	Nombre entier	Identifiant de la personne
age_group	Chaîne de caractères	Catégorie d'âge : "Adult" = adulte, < 65 ans ; "Senior" = âgée, 65 ans
RIDAGEYR	Nombre entier	Âge en années
RIAGENDR	Nombre entier	Genre de la personne : 1 = homme ; 2 = femme
PAQ605	Nombre entier	Activité physique régulière de la personne : 1 = activité physique modérée à importante par semaine ; 2 = pas d'activité physique régulière ; 7 = ne veut pas répondre
BMXBMI	Nombre réel	Indice de masse corporelle (IMC)
LBXGLU	Nombre entier	Glycémie à jeun en $\text{mg} \cdot \text{L}^{-1}$
DIQO10	Nombre entier	Diagnostic médical de diabète : 1 = oui ; 2 = non ; 3 = presque
LBXGLT	Nombre entier	Glycémie dans un test de tolérance au glucose pendant deux heures en $\text{mg} \cdot \text{L}^{-1}$
LBXIN	Nombre réel	Taux d'insuline sanguin en $\mu\text{unité} \cdot \text{mL}^{-1}$

## Travail demandé

### Objectifs

Plusieurs études peuvent être envisagées :

- On pourra chercher à comprendre l'influence conjointes des paramètres biologiques et sociologiques sur le fait d'être porteur du diabète

- On pourra étudier les associations entre diagnostique de diabète et autres paramètres (en particulier pour les individus ayant “presque” un diagnostique)
- On pourra contraster ces résultats par catégorie d’âge

## 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

## Référence

Jeu de données : NA, N. (2019). National Health and Nutrition Health Survey 2013-2014 (NHANES) Age Prediction Subset [Dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5BS66>.

Site de référence de l’enquête NHANES (documentation des variables) : <https://wwwn.cdc.gov/nchs/nhanes/search/DataPage.aspx?Component=Questionnaire&CycleBeginYear=2013>