

Exercice 1 - Lecture d'article

L'article de Wang et al. (2025) [[Wang et al., 2025](#)] est disponible sur le repo github du cours.

1. A quelle question les auteurs cherchent-ils à répondre ?
2. Avec quelle cohorte travaillent-ils ?
3. Quelles analyses statistiques ont été menées ?
4. Analyser les résultats obtenus avec le modèle de régression linéaire multiple.

Exercice 2 - Régression linéaire multiple

On s'intéresse aux données de l'article [[Martin et al., 2025](#)]. Dans ce papier, les auteurs s'intéressent aux maladies métaboliques (en particulier hépatiques) liées à une alimentation trop riche. L'étude porte sur des souris.

1. Appliquer un modèle de régression linéaire multiple pour prédire la variable synthétique "steatose" à partir des données mesurées sur le sérum sanguin.
2. Identifier les éventuels points leviers.

Exercice 3 - Etude observationnelle

On considère ici les données de la cohorte NHANES (<https://www.cdc.gov/nchs/nhanes/index.html>). On propose d'étudier les facteurs de risque liés au diabète. Les facteurs de risque de diabète de type 2 associés à un risque de diabète augmenté avec un lien de causalité démontré sont (entre autres) :

- l'âge [[Ricordeau et al., 2000](#)] ;
- l'origine géographique [[Fagot-Campagna et al., 2010](#)] ;
- un surpoids ;
- la sédentarité.

D'autres facteurs ont été identifiés mais la causalité n'a pas été démontrée à ce jour. Il s'agit (entre autres) :

- l'hypertension artérielle ;
- le tabagisme chronique [[Willi et al., 2007](#)].

En outre, il a été montré que la prévalence du diabète est plus forte chez les personnes en situation de précarité. Voir aussi https://sante.gouv.fr/IMG/pdf/referenciel_pratiques_diabete.pdf.

Dans cet exercice, nous nous intéressons aux liens qui peuvent exister entre les conditions de vie/santé et le risque de diabète, mesuré par la glycémie à jeun. Nous utilisons les données de la cohorte NHANES. Ces données sont recueillies entre 2021 et 2023 sur un échantillon d'environ 13000 personnes par des questionnaires et un bilan médical. Nous avons extrait des données de la base NHANES et constitué une table contenant les variables suivantes.

- LBDGLUSI : glycémie à jeun (variable à prédire)
- BMXWAIST : tour de taille
- smoker : fumeur
- RIDAGEYR : age
- genre
- race_eth : race pour l'origine géographique
- income : niveau de revenu

Répondre aux questions suivantes. Ne pas oublier de travailler avec un document Quarto ou R markdown pour éditer un rapport complet à la fin du TP.

1. Pourquoi cherche t'on à prédire la glycémie à jeun ?
2. Utiliser les outils de la régression linéaire multiple pour étudier les facteurs de risque de diabète.
3. Si vous avez tracé un histogramme ou un quantile-quantile plot de la variable LBDGLUSI, vous avez pu observer que sa distribution assez éloignée d'une distribution gaussienne. Quelle est la conséquence sur l'analyse menée à la question précédente ?
4. Utiliser une transformation de Box-Cos pour "gaussianiser" la variable à prédire et mettre à jour votre analyse pour la variable transformée.
5. Conclure

Références

- [Fagot-Campagna et al., 2010] Fagot-Campagna, A., Romon, I., Fosse, S., Roudier, C., et al. (2010). Prévalence et incidence du diabète, et mortalité liée au diabète en france. *Synthèse épidémiologique. Institut de veille sanitaire*.
- [Martin et al., 2025] Martin, D., Monbet, V., Leroyer, P., Oliviero, N., Turlin, B., Salim, Z., Fautrel, A., Ropert, M., Sire, O., and Loréal, O. (2025). Sequential integration of multi-modal data from serum improves the predictive performance of hepatic lipid accumulation in mice.
- [Ricordeau et al., 2000] Ricordeau, P., Weill, A., Vallier, N., Bourrel, R., Fender, P., and Allemand, H. (2000). L'épidémiologie du diabète en france métropolitaine. *Diabetes & metabolism*, 26 :11–24.
- [Wang et al., 2025] Wang, L., Wang, Y., Yang, C., Jiang, J., Wang, H., and Wu, M. (2025). Multiple linear regression model was constructed based on the influencing factors of vancomycin trough concentration. *Dose-Response*, 23(2) :15593258251313646.
- [Willi et al., 2007] Willi, C., Bodenmann, P., Ghali, W. A., Faris, P. D., and Cornuz, J. (2007). Active smoking and the risk of type 2 diabetes : a systematic review and meta-analysis. *Jama*, 298(22) :2654–2664.