

MODÈLES STATISTIQUES EN SANTÉ  
TESTS D'HYPOTHÈSES, ANOVA  
- M2 SNS -

### Exercice 1 - Lecture d'article

1. Dans l'article [Martin et al., 2021], à quel problème s'intéressent les auteurs ? Quel type de données est utilisé ? Est-ce une étude clinique ou une étude observationnelle ?
2. Relever les passages dans lesquels les auteurs utilisent l'ANOVA. Quelles sont les hypothèses de tests ? Quelles sont les conclusions ?
3. Qu'est ce qu'un test posthoc ?

### Préambule aux exercices suivants

Voir <https://www.scribbr.com/statistics/one-way-anova/> pour un exemple d'application de l'ANOVA à 1 facteur sous R.

Voir <https://www.scribbr.com/statistics/two-way-anova/> pour un exemple d'application de l'ANOVA à 2 facteurs sous R.

### Exercice 2 - Stéatose hépatique

On reprend (encore) l'étude sur la stéatose hépatique [Martin et al., 2025].

1. Observe t'on une différence significative entre les niveaux de glycémie moyen des souris des différents groupes ? Proposer puis mettre en oeuvre un test pour vérifier l'hypothèse "le niveau de glycémie des souris diffère selon le groupe".
2. Proposer puis mettre en oeuvre un (ou des) test(s) pour mettre en évidence quels groupes différent.
3. Reprendre le rapport (R markdown) du TD précédent et le mettre à jour avec les nouveaux résultats.

### Exercice 3 - Coût d'une assurance de santé

On considère le jeu de données `insurance.csv` (Source : <https://www.kaggle.com/mirichoi0218/insurance>) sur le coût des assurances de santé.

Le fichier reporte les données de 7 variables pour 1338 individus.

- `charges` : Individual medical costs billed by health insurance.
- `age` : age of primary beneficiary.
- `sex` : insurance contractor gender, female, male.
- `bmi` : Body mass index, providing an understanding of body, weights that are relatively high or low relative to height, objective index of body weight ( $kg/m^2$ ) using the ratio of height to weight, ideally 18.5 to 24.9.
- `children` : Number of children covered by health insurance / Number of dependents.
- `smoker` : Smoking.

— region : the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.

On n'a malheureusement pas d'information sur la façon dont les données ont été acquises. Elles ont donc uniquement un but pédagogique ici.

1. Proposer des graphiques pour montrer comment varie le coût de l'assurance (variable charges) respectivement en fonction du genre, du statut fumeur/non fumeur et de la région.
2. Proposer une analyse de la variance pour expliquer la variation du coût de l'assurance (variable charges) en fonction de la région. Poser les hypothèses de test et analyser les résultats.
3. Proposer une analyse de la variance pour expliquer la variation du coût de l'assurance (variable charges) en fonction du genre, du statut fumeur/non fumeur et de la région. Analyser les résultats.
4. Prendre en compte les interactions. Poser les hypothèses de test et analyser les résultats.
5. Créer 5 classes d'âge (moins de 25 ans, 25-35 ans, 35-45 and, 45-55 ans, plus de 55 ans) et faire une analyse de la variance pour expliquer la variation du coût de l'assurance en fonction de l'âge. Poser les hypothèses de test et analyser les résultats et mener l'analyse post hoc.
6. Conclure.

## Références

- [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.
- [Martin et al., 2021] Martin, L., Muscaritoli, M., Bourdel-Marchasson, I., Kubrak, C., Laird, B., Gagnon, B., Chasen, M., Gioulbasanis, I., Wallengren, O., Voss, A. C., et al. (2021). Diagnostic criteria for cancer cachexia : reduced food intake and inflammation predict weight loss and survival in an international, multi-cohort analysis. *Journal of cachexia, sarcopenia and muscle*, 12(5) :1189–1202.