



## EC 551 : Statistique descriptive et décisionnelle

### Projet

Guillaume Franchi

Année universitaire 2025-2026

## ■ Présentation du projet

---

Le jeu de données `olive_oil.csv`, disponible sur connect, contient des données relatives aux compositions en acides gras et au prix au litre de 572 huiles d'olives provenant de différentes régions d'Italie.

*Remarque : les données ont été simulées à des fins pédagogiques, et ne reflètent pas la réalité du marché.*

On s'intéresse aux variables pouvant expliquer le montant payé par un assuré.

Par binôme, vous devrez rendre sur connect pour le **mercredi 22 octobre 12h00** :

- un compte-rendu (*format : CR\_nombinome1\_nombinome2.pdf*);
- un script R (*format : script\_nombinome1\_nombinome2.R*);

permettant de répondre aux questions posées ci-dessous. On s'appuiera sur des représentations graphiques et des indicateurs statistiques présentés en cours.

*Remarque : Avant d'analyser les possibles liens entre deux variables, on effectuera une rapide analyse univariée des variables considérées.*

## ■ Questions

---

- 1) Quelles compositions en acides gras semblent expliquer le prix de l'huile d'olive ?
- 2) La région de provenance de l'huile semble-t-elle expliquer sa composition en acide linoléique ?
- 3) Créer une nouvelle variable qualitative correspondant à une gamme de prix :
  - $< 6$  : entrée de gamme ;
  - $[6, 8[$  : bon marché ;
  - $[8, 10[$  : moyenne gamme ;
  - $[10, 12[$  : gamme supérieure
  - $\geq 12$  : haut de gamme.

Existe-t-il un lien entre ces différentes gammes et la région de provenance de l'huile d'olive ?