



EC 56/51 : Statistique descriptive et décisionnelle

Projet

Guillaume Franchi

Année universitaire 2025-2026

■ Présentation du projet

Le jeu de données `coffee.csv`, disponible sur connect, contient les notes attribuées par un panel d’experts à 1 302 produits de café différents.

Ces notes sur l’arôme, le goût, l’acidité, le corps, l’équilibre et l’arrière-goût sont accompagnées de l’altitude à laquelle le café a poussé.

Enfin, une dernière variable qualitative « Label » indique si le café est labellisé « Biologique » (*Organic*) ou non.

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

Par binôme, vous devrez rendre sur connect pour le **vendredi 30 janvier 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.

■ Questions

- 1) Réaliser un rapide résumé statistique de chacune des variables du jeu de données. Certaines valeurs vous semblent-elles anormales ? Si oui, les retirer du jeu de données.
- 2) Quelle est la variable qui semble expliquer le plus la note de goût ? L’altitude semble-t-elle expliquer l’acidité du café ?
- 3) Le pays d’origine du café semble-t-il expliquer la note d’acidité de ce dernier ?
- 4) Existe-t-il un lien entre le pays d’origine du café et le fait que celui-ci soit « Biologique » ou non ?