

## Exercice d'auto-évaluation - Session 3

Un groupe industriel commercialisant du café souhaite comparer les cafés provenant de différents sites de production à partir de leur profil de composition physico-chimique : teneurs en acide chlorogénique (CGA), caféine (Cafeine), matières grasses (Fat), trigonelline (Trigonelline) et matières sèches (DM).

Les données sont fournies dans le fichier `coffee.txt`. Un script R important ces données est également à disposition. Dans un premier temps, on se concentre sur la question suivante : les taux de matières sèches sont-ils différents dans les sites de production 1 et 6 ?

1. Dans cette problématique, quelle est la variable réponse ? Quelle est la variable explicative ? Vous donnerez aussi la nature (quantitative ou catégorielle) de ces variables.
2. Proposez un modèle statistique adapté à cette question. Vous donnerez le nom courant de ce modèle.
3. Exprimez la question sous forme de test d'hypothèses dont vous donnerez les hypothèses  $H_0$  et  $H_1$ .
4. Donnez la valeur de la statistique de Student pour le test d'hypothèses de la question précédente.
5. Quelle est la distribution de cette statistique de test sous l'hypothèse  $H_0$  ?
6. Les teneurs moyennes en matières sèches sont-elles significativement différentes au seuil  $\alpha = 0.05$  ?
7. Quelle taille d'échantillon serait nécessaire pour détecter avec une probabilité de 0.90 une différence de teneurs moyennes en matières sèches à l'échelle de la population de 0.05 ?