

## TP n° 7 : ANOVA à 1 facteur

**Exercice 1.** On étudie le rendement en jus de 3 variétés de pommes. Pour chaque variété, on considère un échantillon de 4 arbres. Les résultats sont :

Golden	Delicious	Jonagold
48	52	53
46	50	51
52	49	55
50	49	57

La problématique est la suivante : existe-t-il une différence significative entre ces 3 variétés quant à la moyenne des rendements ? Pour ce faire, on utilise le test ANOVA à 1 facteur, le facteur considéré étant la variété d'une pomme. On propose les commandes suivantes :

```
rend = c(48, 46, 52, 50, 52, 50, 49, 49, 53, 51, 55, 57)
var = factor(rep(c("Golden", "Delicious", "Jonagold"), rep(4, 3)))
w = data.frame(rend, var)
attach(w)
plot(rend ~ var)
shapiro.test(rend[var == "Golden"])
shapiro.test(rend[var == "Delicious"])
shapiro.test(rend[var == "Jonagold"])
bartlett.test(rend ~ var)
reg = aov(rend ~ var)
summary(reg)
pairwise.t.test(rend, var, p.adj = "bonf")
```

Reproduire et comprendre l'enjeu de ces commandes.

**Exercice 2.** On s'intéresse aux heures par an passées à lire des ouvrages autres que ceux de fiction, pour un échantillon de 18 personnes, classées selon trois niveaux d'éducation :  $E_1$ ,  $E_2$  et  $E_3$ . Les résultats sont :

E1	E2	E3
8	12	20
10	14	24
12	16	16
4	10	28
6	11	30
8	12	32

Peut-on rejeter, au risque 5%, l'hypothèse selon laquelle le niveau d'éducation n'a pas d'effet sur le nombre d'heures de lecture ?

**Exercice 3.** Un expérimentateur veut calculer les paramètres de fidélité pour un paramètre qui est la teneur du pesticide Gamma-HCH en mugrammes par gramme dans la poudre de lait.

Cette teneur peut varier d'un producteur à l'autre en fonction en particulier de l'environnement dans lequel il a recueilli le lait. L'expérimentateur utilise des poudres de lait provenant de 3 producteurs :  $P_1$ ,  $P_2$  et  $P_3$ . Pour chacun des ces producteurs, il mesure la teneur en pesticide de 5 boîtes de poudre de lait choisies au hasard. Les résultats sont :

P1	P2	P3
44.7	42.3	47.0
44.6	41.6	47.4
44.8	42.9	47.4
45.8	43.5	47.2
46.4	43.7	46.2

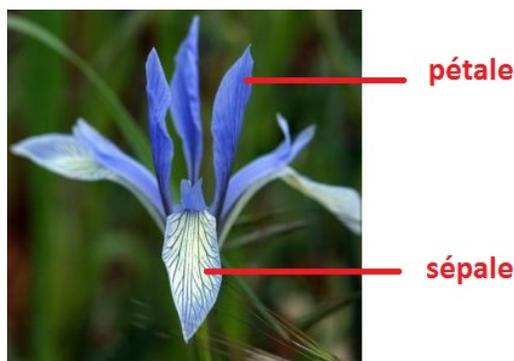
1. Peut-on supposer que, pour tout  $i \in \{1, 2, 3\}$ , la teneur du pesticide Gamma-HCH du producteur  $P_i$  est une  $var X_i$  suivant une loi normale ?
2. Peut-on affirmer, au risque 5%, que la dispersion de la teneur en pesticide diffère selon les producteurs ?
3. Y a-t-il un effet "producteur" significatif sur la teneur en pesticide ?

**Exercice 4.** On considère le jeu de données `iris` disponible dans R.

1. Reproduire les commandes suivantes :

```
data(iris)
attach(iris)
names(iris)
help(iris)
```

Comprendre le contexte de l'étude. Notons que ce jeu de données possède un facteur "espèce d'iris" avec 3 modalités : "setosa", "versicolor" et "virginica". Aussi, on précise :



2. Créer une data frame `w` ne contenant que les variables `Sepal.Length` et `Species`.
3. Utiliser `w` pour répondre à la problématique : peut-on affirmer qu'il y a une différence significative entre ces 3 espèces d'iris quant à la longueur moyenne du sépal ? Est-ce que les hypothèses standards du test ANOVA sont vérifiées ?

**Exercice 5.** On considère le jeu de données `PlantGrowth` disponible dans R. Peut-on affirmer que le facteur `group` influe de manière significative sur la variable `weight` ?