

TP ANOVA(1)

Dans ce TP, nous allons évaluer si la résistance à un pathogène donné varie pour des plantes de 4 lignées différentes de l'espèce modèle *Medicago truncatula*. Les données sont enregistrées dans le fichier excel « résistance ». Avant de comparer les scores moyens de résistance, nous effectuons une ANOVA à 1 facteur pour évaluer s'il y a un effet "lignée" sur le score de résistance. Si c'est le cas, nous comparerons ensuite les lignées 2 à 2 afin de déterminer la(les)quelle(s) se différencie(nt) des autres.

1) Exploration du jeu de données

Lecture du fichier

```
resistance <- read.csv2("resistance.csv")
attach(resistance)
table(lignee) # Affichage des effectifs des différents groupes (niveaux du facteur)
tapply(score,lignee,mean) # Affichage du score moyen par groupe avec la fonction tapply()
tapply(score,lignee,summary) # un résumé des données par groupe
boxplot(score~lignee,col="green",ylab="score") # Affichage de la distribution des scores pour
#chaque lignée avec une boîte à moustache
```

2) Tests préalables à l'ANOVA

Avant de procéder à l'ANOVA, il faut vérifier la normalité des données de chaque groupe, et l'homogénéité des variances des groupes:

Test d'adéquation à la loi normale ($H_0 = \text{"les données suivent la loi Normale"}$)

Test de Shapiro

```
shapiro.test(score[lignee=="HM008"])
shapiro.test(score[lignee=="A17"])
shapiro.test(score[lignee=="DZA45"])
shapiro.test(score[lignee=="HM013"])
```

Test d'homogénéité des variances intra-groupes ($H_0 = \text{"les variances sont égales"}$)

Test de Bartlett de comparaison de plus de 2 variances :

```
bartlett.test(score~lignee)
```

3) ANOVA à 1 facteur (le facteur est la variable "lignee") => $H_0 : \text{"les moyennes des différents groupes sont égales"}$

On utilise la commande aov() :

```
res = aov(score~lignee)
```

IMPORTANT : le tableau de l'ANOVA s'obtient avec la fonction summary() :

```
summary(res) # ou summary(aov(score~lignee))
```

On peut visualiser la différence des moyennes de chaque groupe avec la moyenne générale avec :

```
model.tables(res)
```

Si l'ANOVA détecte un effet significatif du facteur, on peut chercher les inégalités de moyennes : Tests de Student pour comparer les groupes 2 à 2

```
t.test(score[lignee=="DZA45"],score[lignee=="HM013"],var.equal=T)
```

```
t.test(score[lignee=="DZA45"],score[lignee=="HM008"],var.equal=T)
```

```
t.test(score[lignee=="DZA45"],score[lignee=="A17"],var.equal=T)
```

```
t.test(score[lignee=="A17"],score[lignee=="HM013"],var.equal=T)
```

```
t.test(score[lignee=="HM008"],score[lignee=="HM013"],var.equal=T)
```

```
t.test(score[lignee=="A17"],score[lignee=="HM008"],var.equal=T)
```

IMPORTANT : pour limiter le taux de faux positifs lors de tests multiples, et pour faire toutes ces comparaisons avec une seule commande, on fait un test de comparaisons multiples de Student avec correction de la p-valeur :

```
pairwise.t.test(score,lignee,p.adjust.method="bonferroni")
```

4) Conclusion de ces analyses ? Faites votre conclusion.....

REMARQUE IMPORTANTE, si les échantillons ne suivent pas la normalité

On peut faire une ANOVA non paramétrique avec le test de Kruskal-Wallis :

```
kruskal.test(score~lignee)
```

Puis des tests non paramétrique de Wilcoxon/Mann-Whitney, pour comparer les groupes 2 à 2 :

```
wilcox.test(score[lignee=="HM008"],score[lignee=="A17"])
```

```
wilcox.test(score[lignee=="HM008"],score[lignee=="DZA45"])
```

... et, pour les mêmes raisons que pour le test multiple de Student, faire un test multiple de Wilcoxon/Mann-Whitney :

```
pairwise.wilcox.test(score,lignee)
```

```
pairwise.wilcox.test(jitter(score),lignee,p.adjust.method="bonferroni")
```