Pratique d' ANOVA(1)

L'ANOVA(1) s'applique dès qu'on veut vérifier l'effet d'une variables qualitative (facteur, indépendante, exogène, explicative, dehors le modèle) sur une variable quantitative(d'intérêt,réponse, dépendante, endogène, à expliquer, dans le modèle)

Avant d'utiliser l'anova vous devez vérifier :

- les échantillons sont prélevés aléatoirement et indépendamment. Noter qu'il n'y a pas de test statistique simple pour étudier l'indépendance. Les conditions de l'expérience choisie nous déterminent si nous sommes dans le cas de l'indépendance.
- La normalité pour la variable qualitative dans tout les échantillons (**shapiro.test**)
- L'homoscedasticité :Les variances intra-groupe sont toutes égales(test de Bartlett de comparaison de plus de 2 variances : **bartlett.test**)

La commande qui permet d'établir une anova est :

```
aov(y ~ G, data=mydataframe) # y est la variable quantitative et G indique le facteur
summary(aov(y ~ G, data=mydataframe))
```

Dans la sortie R on accepte l'hypothèse nulle si la **p-value** est supérieure au risque **alpha** fixé à priori . Dans ce cas le facteur n'a aucune influence sur la variable dépendante

Exemple :

```
fr <- data.frame(var = c(10, 4, 5, 3, 3, 7, 2, 6, 2, 8, 5), group = factor(c("c","a","b", "a", "b", "b", "a", "b", "a", "c", "c")))
fit <- aov(var ~ group, fr)# analyse de variance
summary(fit)
```

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**

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(var,group,p.adjust.method="bonferroni")
```

Exercice d'application

On veut connaître l'effet de trois types de fertilisants sur la croissance des arbres d'une plantation. Pour cela, on extrait 3 échantillons (groupes) d'arbres et appliquer chaque fertilisant pour chaque échantillon. Les données sont regroupées dans les vecteurs suivants :

```
fert1<-c(23.3,24.4,24.6,24.9,25.0,26.2)
fert2<-c(18.9,21.1,21.1,22.1,22.5,23.5)
fert3<-c(22.5,22.9,23.7,24.0,24.0,24.5)
```

On se demande si pour la plantation, tous les fertilisants sont équivalents, ou bien y a-t-il un qui soit meilleur (ou moins bon) que les autres?

Réponse :

Population : les arbres

La variable d'intérêt : la croissance en cm/an par.

Le facteur : type de fertilisant.

1) Vérification des conditions :

   - L'indépendance des échantillon : il est possible quand en donnant un fertilisant à un arbre qu' il y ait des problèmes de diffusion aux arbres voisins (pluie, vent…). Une solution possible serait d'effectuer un maillage de la plantation (on définit ainsi des plateaux), on applique un fertilisant aux arbres qui sont dans le même plateau.

   - La normalité de la hauteur des arbre :

     ```
     shapiro.test(fert1)
     shapiro.test(fert2)
     shapiro.test(fert3)
     ```

     Les trois p-value sont supérieure à 0.05, donc on accepte l'hypothèse nulle pour les trois groupes i.e. pour les trois fertilisants les données sont issues d'une distribution normale.

   - L'homosédascticité :

     ```
     bartlett.test(var ~ group, data = fr)
     ```

     La p-value est supérieure à 0.05, donc on accepte l'hypothèse nulle qui suppose l'égalité des trois variances .

2) ANOVA(1) :

```
fr <- data.frame(var = c(fert1,fert2,fert3),  group = factor(c(rep("fert1",6), rep("fert2",6), rep("fert3",6))))
fit <- aov(var ~ group, fr)# analyse de variance
summary(fit)
```

3) Conclusion : p-value < 0.05 donc la croissance moyenne est significativement différente pour chaque fertilisant.

4) Test multiple

```
t.test(fert1,fert2,var.equal=T)
t.test(fert1,fert3,var.equal=T)
t.test(fert2,fert3,var.equal=T)
ou
pairwise.t.test(var,group,p.adjust.method="bonferroni")
```

REMARQUE IMPORTANTE :Si les échantillons ne suivent pas la normalité, on peut faire une ANOVA non paramétrique avec le test de Kruskall-Wallis. Puis des tests non paramétriques de Wilcoxon/Mann-Whitney, pour comparer les groupes 2 à 2 :

Exemple :

On reprend l'exemple vu dans le cours de pratique des test de moyennes

Dans une bibliothèque, il y a des livres de voyage, des ouvrages généraux et des livres de statistiques . On a trois échantillons :1. de voyage : 93, 98, 216, 249, 301, 319, 731 et 910 pages. 2.livres généraux 29, 39, 60, 78, 82, 112, 125, 170, 192, 224, 263, 275, 276, 286, 369 et 756 pages. 3. de statistiques : 126, 142, 156, 228, 245, 246, 370, 419, 433, 454, 478 et 503 pages. La question est " Ces échantillons proviennent-ils d'une même population ou au contraire un au moins des trois échantillons présentent une originalité en ce qui concerne la taille moyenne ? ".

```
livre <-c(93,98,216,249,301,319,731,910,29,39,60,78,82,112,125,170,192,224,263,275,276,286,369,756,126,142,156,228,245,246,370,419,433,454,478,503)
 group <- factor(rep(1:3,c(8,16,12)))

kruskal.test(livre~group)

voyage<-c(93, 98, 216, 249, 301, 319, 731 ,910)
generaux<-c(29, 39, 60, 78, 82, 112, 125, 170, 192, 224, 263, 275, 276, 286, 369 , 756)
stat<-c( 126, 142, 156, 228, 245, 246, 370, 419, 433, 454, 478 , 503)
wilcox.test(voyage,generaux)
wilcox.test(voyage,stat)
wilcox.test(generaux,stat)
```

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

```
pairwise.wilcox.test(livre,group)
```