Éléments de correction du TD numéro 13

Exercice 3

Question 1
Établir le tableau d'analyse de variance. Écrire le modèle associé à cette analyse.
Code R

```
-------------------------------------------------------------------
> cafe<- c(1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3,3,3)
> note<-c(0,2,3,3,5,6,3,3,5,6,6,8,4,6,7,7,8,10)
> cafe<- factor(cafe)
> test1<- data.frame(cafe,note)
> model1LM<-lm(note~cafe,test1)
> model1AOV<-anova(model1LM)
> model1AOV

Analysis of Variance Table
Response: note
          Df Sum Sq Mean Sq F value  Pr(>F)
cafe       2 44.111  22.056  5.3649 0.01748 *
Residuals 15 61.667   4.111
---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
----------------------------------------------------------------------------
```

Le modèle associé à cette analyse s'écrit :

Y_ij=mu+alpha_i+epsilon_ij   avec i=1 à I=3 et j =1 à J=6

où
mu = l'acidité moyenne
alpha_i = l'effet du café i
epsilon_ij = le résidu du modèle.
Nous ajoutons la contrainte sur les alpha_i à savoir : somme sur i des alpha_i=0.

Question 2
Tester l'hypothèse selon laquelle les trois cafés présentent une acidité identique. Donner vos conclusions, au risque de 5% (respectivement 1%).
Nous décidons d'intégrer au modèle précédent l'influence du juge sur la note d'acidité.

Les hypothèses testées sont
H_0: alpha_1=alpha_2=alpha_3=0
Contre
H_1: il existe au moins un i variant de 1 à 3, tel que alpha_i≠0

Pour réaliser ce test (ici il s'agit d'un test de Fisher), nous devons vérifier au préalable les trois hypothèses suivantes:

1 Indépendance des résidus
L'expérience est bien réalisée de façon indépendante car il y a 6 juges différents.

2 Normalité des résidus
Code R

```
-------------------------------------
residus<-residuals(modele1LM)
> shapiro.test(residus)
      Shapiro-Wilk normality test
data:  residus
W = 0.9502, p-value = 0.4284
-------------------------------------------
```

p-value=0.4284>0.05
Donc il y a normalité des résidus.

3 Égalité des variances des résidus ou homoscédasticité
Code R

```
------------------------------------------------------
> bartlett.test(residus~cafe,data=test1)
      Bartlett test of homogeneity of variances
data:  residus by cafe
Bartlett's K-squared = 0.0451, df = 2, p-value = 0.9777
---------------------------------------------------------------------------
```

p-value=0.9777 >0.05
Donc il y a égalité des variances des résidus.

Maintenant, nous pouvons tester l'hypothèse nulle H_0 contre l'hypothèse alternative H_1. D'après le tableau de l'ANOVA précédent, nous observons une p-value égale à 0.01748 qui est bien inférieure à 0.05. Nous décidons de rejeter l'hypothèse nulle H_0 au risque alpha=5%. Il y a donc des différences d'acidité significatives entre les trois cafés, càd certains cafés sont plus acides, en moyenne, que d'autres.
D'après le tableau de l'ANOVA précédent, nous observons une p-value égale à 0.01748 qui est bien supérieure à 0.01. Nous décidons de ne pas rejeter l'hypothèse nulle H_0 au risque beta qu'il faudrait calculer. Il n'y a donc pas de différence d'acidité entre les trois cafés, en moyenne.

Question 3
Faire une régression multiple sur les données. Donner ensuite le tableau d'analyse de la variance à deux facteurs en justifiant brièvement vos réponses.
Code R

```
--------------------------------------------------------------------------
>cafe<-c(1,1,1,1,1,1,2,2,2,2,2,2,3,3,3,3,3,3)
>juge<-c(1,2,3,4,5,6, 1,2,3,4,5,6,1,2,3,4,5,6)
>note<-c(0,2,3,3,5,6,3,3,5,6,6,8,4,6,7,7,8,10)
>cafe< factor(cafe)
>juge<-factor(juge)
>test2<-data.frame(cafe,juge,note)
>str(test2)
'data.frame':   18 obs. of  3 variables:
 $ cafe: Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 2 2 2 2 ...
 $ juge: Factor w/ 6 levels "1","2","3","4",..: 1 2 3 4 5 6 1 2 3 4 ...
 $ note: num  0 2 3 3 5 6 3 3 5 6 ...
>options(contrasts=c("contr.sum","contr.poly"))
>model2<-lm(note~cafe+juge,data=test2)
> residus<-residuals(model2)
```

```
> shapiro.test(residus)
        Shapiro-Wilk normality test
data:  residus
W = 0.9375, p-value = 0.2624
>bartlett.test(residus~cafe+juge,data=test2)
        Bartlett test of homogeneity of variances
data:  residus by cafe by juge
Bartlett's K-squared = 1.9874, df = 2, p-value = 0.3702
>summary(model2)
Call:
lm(formula = note ~ cafe + juge, data = test2)
Residuals:
     Min       1Q   Median       3Q      Max
-0.72222 -0.22222 -0.05556  0.23611  0.61111

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.11111    0.11915  42.895 1.14e-12 ***
cafe1       -1.94444    0.16851 -11.539 4.22e-07 ***
cafe2        0.05556    0.16851   0.330 0.748436
juge1       -2.77778    0.26644 -10.426 1.08e-06 ***
juge2       -1.44444    0.26644  -5.421 0.000292 ***
juge3       -0.11111    0.26644  -0.417 0.685468
juge4        0.22222    0.26644   0.834 0.423724
juge5        1.22222    0.26644   4.587 0.000999 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 0.5055 on 10 degrees of freedom
Multiple R-squared: 0.9758,     Adjusted R-squared: 0.9589
F-statistic:  57.7 on 7 and 10 DF,  p-value: 2.753e-07

> anova(model2)
Analysis of Variance Table
Response: note
          Df Sum Sq Mean Sq F value    Pr(>F)
cafe       2 44.111  22.056  86.304 4.925e-07 ***
juge       5 59.111  11.822  46.261 1.360e-06 ***
Residuals 10  2.556   0.256
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Le tableau de la variance est donné ci-dessus. Il manque uniquement la dernière ligne, la ligne « Totale » à rajouter.

Question 4
Quelle interpr□tation concr□te donner □ l□effet juge ? Est-il int□ressant, lorsque nous nous int□ressons uniquement □ l□effet caf□, de prendre en compte l□effet juge dans le mod□le d□analyse de la variance ?

Il y a un effet juge lorsque certains juges donnent des notes systématiquement élevées quel que soit le café et d’autres des notes systématiquement basses. En fait, les juges ne

se ressemblent pas. Ils ont des sensibilités différentes à l'acidité ou bien des références différentes ou encore des façons différentes d'utiliser l'échelle de notation.
Oui, il est intéressant de prendre en compte l'effet juge dans le modèle d'analyse de la variance puisqu'ils ne sont pas tous identiques, comme nous venons de le constater. Cela nous permet de distinguer la variabilité due aux juges de la variabilité résiduelle. Si jamais nous ne prenions pas en compte la variabilité juge, alors nous augmenterions la variabilité résiduelle.

Question 5
Écrire le modèle d'analyse de la variance. Conclure sur les effets des facteurs café et juge.

Le modèle d'analyse de la variance s'écrit :

$$Y_{ij}=\mu+\alpha_i+\beta_j+\epsilon_{ij}$$

où
mu = l'acidité moyenne
alpha_i = l'effet café i
beta_j = l'effet juge j
epsilon_ij = le résidu du modèle.
Nous ajoutons les deux contraintes suivantes : somme sur i des aplha_i=0 et somme sur j des beta_j =0.

Donc, à partir du tableau de la variance de la question 3, nous décidons de rejeter l'hypothèse nulle H_0 (p=4.925e-07<0.05) au risque alpha=5% pour le facteur café, c'est-à-dire il existe une différence, en moyenne, entre les trois types de cafés. Nous décidons également de rejeter l'hypothèse nulle H_0 (1.360e-06<0.05) au risque alpha=5% pour le facteur juge, c'est-à-dire il existe une différence, en moyenne, entre les six juges.

Question 6
Commenter les façons de noter des juges 1 et 3. Quel café achèteriez-vous si vous préférez les cafés peu acides ? Justifier.

Le juge 1 a tendance à mettre en moyenne des notes plus faibles que les autres juges.
Le juge 3 note dans la moyenne de l'ensemble des juges.
Code R

```
------------------------------------------------------------------------
>moy<-tapply(acide2$note,acide2$cafe,mean)
> moy
1        2        3
3.166667 5.166667 7.000000
------------------------------------------------------------------------
```

Si nous préférons les cafés peu acides, nous achèterons le café 1, jugé comme le moins acide parmi les trois testés.