MAT7381 Solution # 6.30

6.30 On compare 5 variétés de maïs, chacune plantée dans 4 terrains (des "blocs"). Les blocs sont divisés en 5 parcelles chacun, une parcelle étant consacrée à chaque variété. Une mesure de la quantité récoltée est prise dans chaque parcelle. Voici les données :

	Bloc				
Variété	1	2	3	4	Moyenne
A	21	25	28	35	27,25
B	23	29	33	34	29,75
C	15	21	22	28	21,50
D	16	19	25	25	21,25
E	12	18	21	25	19,00
Moyenne	17,4	22,4	25,8	29,4	23,75

Voici les données (on s'assure que « bloc » soit traité comme facteur)

```
   récolte variété bloc
1       21       A    1
2       25       A    2
3       28       A    3
4       35       A    4
5       23       B    1
6       29       B    2
7       33       B    3
8       34       B    4
9       15       C    1
10      21       C    2
11      22       C    3
12      28       C    4
13      16       D    1
14      19       D    2
15      25       D    3
16      25       D    4
17      12       E    1
18      18       E    2
19      21       E    3
20      25       E    4
```

On construit les colonne « variété » et « bloc » comme ceci :

```
bloc<-rep(1:4,5)
variété<-gl(5,4,labels=c("A","B","C","D","E"))
```

On développe un modèle additif :

```
> a<-lm(récolte~variété+bloc)
> summary(a)

            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.9000     0.8926  23.416 2.20e-11 ***
variétéB      2.5000     0.9979   2.505 0.027649 *
variétéC     -5.7500     0.9979  -5.762 8.99e-05 ***
variétéD     -6.0000     0.9979  -6.013 6.10e-05 ***
variétéE     -8.2500     0.9979  -8.267 2.68e-06 ***
bloc2         5.0000     0.8926   5.602 0.000116 ***
bloc3         8.4000     0.8926   9.411 6.87e-07 ***
bloc4        12.0000     0.8926  13.444 1.35e-08 ***

Residual standard error: 1.411 on 12 degrees of freedom
Multiple R-squared: 0.9679,     Adjusted R-squared: 0.9491
F-statistic: 51.63 on 7 and 12 DF,  p-value: 5.119e-08
```

En termes des paramètres μ, v (variété) et b (bloc), les moyennes sont :

	bloc1	bloc2	bloc3	bloc4	
Variété A	μ	μ +b2	μ +b3	μ +b4	$\mu +\sum b/4$
Variété B	μ +vb	μ +vb+b2	μ +vb+b3	μ +vb+b4	μ+vb $+\sum b/4$
Variété C	μ +vc	μ +vc+b2	μ +vc+b3	μ +vc+b4	μ +vc$+\sum b/4$
Variété D	μ +vd	μ +vd+b2	μ +vd+b3	μ +vd+b4	μ +vd$+\sum b/4$
Variété E	μ +ve	μ +ve+b2	μ +ve+b3	μ +ve+b4	μ +ve$+\sum b/4$
	$\mu +\Sigma v/4$	$\mu+\Sigma v/4$ +b2	$\mu+\Sigma v/4$ +b3	$\mu +\Sigma v/4$+b4	$\mu +\Sigma v/4+\sum b/4$

Elles sont estimées par

```
> M
      [,1]  [,2]  [,3]  [,4]
[1,] 20.90 25.90 29.30 32.90
[2,] 23.40 28.40 31.80 35.40
[3,] 15.15 20.15 23.55 27.15
[4,] 14.90 19.90 23.30 26.90
[5,] 12.65 17.65 21.05 24.65
```

Une façon de les obtenir est celle-ci :

```
> beta<-summary(a)$coefficients[,1]
> beta
(Intercept)   variétéB   variétéC   variétéD   variétéE      bloc2      bloc3      bloc4
      20.90       2.50      -5.75      -6.00      -8.25       5.00       8.40      12.00
```

On construit la matrice de design :

```
> X<-model.matrix(a)
> X
   (Intercept) variétéB variétéC variétéD variétéE bloc2 bloc3 bloc4
1            1        0        0        0        0     0     0     0
2            1        0        0        0        0     1     0     0
3            1        0        0        0        0     0     1     0
4            1        0        0        0        0     0     0     1
5            1        1        0        0        0     0     0     0
6            1        1        0        0        0     1     0     0
7            1        1        0        0        0     0     1     0
8            1        1        0        0        0     0     0     1
9            1        0        1        0        0     0     0     0
10           1        0        1        0        0     1     0     0
11           1        0        1        0        0     0     1     0
12           1        0        1        0        0     0     0     1
13           1        0        0        1        0     0     0     0
14           1        0        0        1        0     1     0     0
15           1        0        0        1        0     0     1     0
16           1        0        0        1        0     0     0     1
17           1        0        0        0        1     0     0     0
18           1        0        0        0        1     1     0     0
19           1        0        0        0        1     0     1     0
20           1        0        0        0        1     0     0     1
```

```
> muhat<-X%*%beta
> muhat
       [,1]
 [1,] 20.90
 [2,] 25.90
 [3,] 29.30
 [4,] 32.90
 [5,] 23.40
 [6,] 28.40
 [7,] 31.80
 [8,] 35.40
 [9,] 15.15
[10,] 20.15
[11,] 23.55
[12,] 27.15
[13,] 14.90
[14,] 19.90
[15,] 23.30
[16,] 26.90
[17,] 12.65
[18,] 17.65
[19,] 21.05
[20,] 24.65
```

a) Testez à 5% l'hypothèse qu'il n'y a pas de différence entre les variétés, en supposant qu'il n'y a pas d'interaction entre les variétés et les blocs.

La réponse découle immédiatement de l'ANOVA :

```
> anova(a)
Analysis of Variance Table

Response: récolte
          Df Sum Sq Mean Sq F value    Pr(>F)
variété    4 328.50  82.125  41.234 6.416e-07 ***
bloc       3 391.35 130.450  65.498 1.039e-07 ***
Residuals 12  23.90   1.992
```

On voit tout de suite que les différences entre les variétés sont hautement significatives.

b) Testez l'hypothèse qu'en moyenne les variétés A et B produisent autant que les trois autres.

On obtient les moyennes des carrés de façon à estimer la variance

```
> ms<-anova(a)$"Mean Sq"
> ms
[1]  82.125000 130.450000   1.991667
```

La variance σ^2 est estimée par MCR, la 3e composante de ce vecteur et son écart-type est estimé par

```
> sigma<-sqrt(ms[3])
> sigma
[1] 1.411264
```

La matrice de covariance de $\hat{\boldsymbol{\beta}}$ est $\sigma^2(\mathbf{X'X})^{-1}$. La matrice $(\mathbf{X'X})^{-1}$ est donnée par `summary(a)$cov.unscaled`. Voici donc la variance (estimée) $\mathbf{V}$ de $\hat{\boldsymbol{\beta}}$:

```
> V<-sigma^2*summary(a)$cov.unscaled
> round(V,4)
            (Intercept) variétéB variétéC variétéD variétéE   bloc2   bloc3   bloc4
(Intercept)      0.7967  -0.4979  -0.4979  -0.4979  -0.4979 -0.3983 -0.3983 -0.3983
variétéB        -0.4979   0.9958   0.4979   0.4979   0.4979  0.0000  0.0000  0.0000
variétéC        -0.4979   0.4979   0.9958   0.4979   0.4979  0.0000  0.0000  0.0000
variétéD        -0.4979   0.4979   0.4979   0.9958   0.4979  0.0000  0.0000  0.0000
variétéE        -0.4979   0.4979   0.4979   0.4979   0.9958  0.0000  0.0000  0.0000
bloc2           -0.3983   0.0000   0.0000   0.0000   0.0000  0.7967  0.3983  0.3983
bloc3           -0.3983   0.0000   0.0000   0.0000   0.0000  0.3983  0.7967  0.3983
bloc4           -0.3983   0.0000   0.0000   0.0000   0.0000  0.3983  0.3983  0.7967
```

L’hypothèse à tester est

[(μ+∑b/4)+(μ+vb +∑b/4)]/2=[(μ+vc +∑b/4)+(μ+vd +∑b/4)+(μ+ve +∑b/4)]/3

⇔ vb/2-(vc+vd+ve)/3 = 0. soit L’β = 0 où L est

```
> L
[1]  0.0000000  0.5000000 -0.3333333 -0.3333333 -0.3333333  0.0000000  0.0000000  0.0000000
```

$\mathbf{L'}\hat{\boldsymbol{\beta}} =$

```
> t(L)%*%beta
[1,] 7.916667
```

Var($\mathbf{L'}\hat{\boldsymbol{\beta}}$) est estimée par

```
> t(L)%*%V%*%L*sigma^2
[1,] 0.8264034
```

L’écart-type est estimé par

```
> sqrt(t(L)%*%V%*%L*sigma^2)
[1,] 0.9090673
```

Sous H_o, le rapport $\frac{\mathbf{L'}\hat{\boldsymbol{\beta}}}{\hat{\sigma}_{\mathbf{L'}\hat{\boldsymbol{\beta}}}}$ est de loi de Student à 12 degrés de liberté. Sa valeur est

```
> t(L)%*%beta/sqrt(t(L)%*%V%*%L*sigma^2)
[1,] 8.70856
```

La valeur *p* est

```
> 2*(1-pt(abs(t(L)%*%beta/sqrt(t(L)%*%V%*%L*sigma^2)),12))
[1,] 1.560907e-06
```

On rejette donc l’hypothèse énoncée en b).