

TD4. Exemples

Sophie Donnet

3 novembre 2017

R Markdown

Nous travaillons sur des données récoltées sur 200 personnes équirépartis en 4 groupes : A, B, C, et T de 50 personnes chacun. Au sein de chaque groupe, les individus ont respecté une organisation particulière de leurs repas sauf pour le groupe témoin où les individus n'ont rien changé à leurs habitudes. Le groupe A a eu 3 repas par jour supplémentés par des en-cas réguliers, le groupe B a eu 3 repas par jour (2 repas consistants et 1 léger), le groupe C a eu 2 repas par jour.

Le but est de comparer ces différentes organisations : ont-elles une influence sur le poids des individus ? Nous disposons des variables cal et prot donnant l'apport moyen quotidien en calories et en protéines des repas pour chaque individu. La variable P0 est le poids initial, la variable P3 le poids 3 mois après le début de l'expérience et la variable G3 la variation de poids au cours des 3 premiers mois.

```
don=read.table("nutri3.txt",header=TRUE,sep=";")
don=don[,2:7]
```

Un extrait des données est fourni ci-dessous.

```
head(don)
```

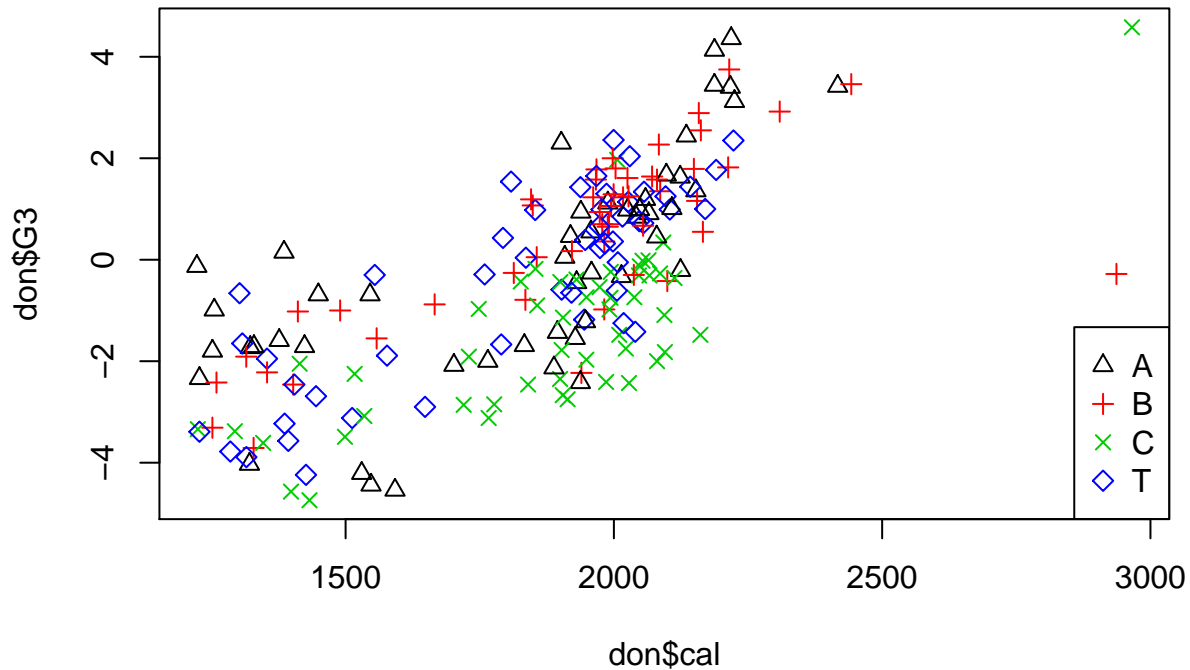
##	Group	P0	prot	cal	G3	P3
## 1	A	66.59	51.78	1930.62	-0.45	66.14
## 2	A	78.76	55.70	1958.28	-0.26	78.51
## 3	A	61.70	34.92	1423.51	-1.71	59.99
## 4	A	83.93	57.91	1938.44	0.94	84.87
## 5	A	73.54	56.05	1957.32	0.55	74.08
## 6	A	65.58	53.13	2134.98	2.44	68.01

```
tail(don)
```

##	Group	P0	prot	cal	G3	P3
## 195	T	51.83	32.98	1808.41	1.54	53.37
## 196	T	66.39	45.29	1946.59	0.39	66.77
## 197	T	52.18	31.84	1759.17	-0.29	51.90
## 198	T	60.74	49.53	2222.94	2.35	63.09
## 199	T	46.06	23.07	1577.23	-1.89	44.16
## 200	T	63.88	50.02	2039.85	-1.42	62.46

On s'intéresse à présent à la variation de poids en fonction du régime tout en tenant compte de la quantité moyenne de calories.

Figure 3 : Lien entre le groupe et la variation de poids en 3 mois



```
mod2=lm(G3~Group*cal,data=don)
summary(mod2)
```

```
##
## Call:
## lm(formula = G3 ~ Group * cal, data = don)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.7424 -0.8016  0.1228  0.7768  2.8689
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.7689253  0.9810945  -8.938 3.28e-16 ***
## GroupB       1.6034604  1.4386787   1.115  0.266
## GroupC      -1.0047663  1.5043960  -0.668  0.505
## GroupT      -1.1559747  1.4668354  -0.788  0.432
## cal          0.0047215  0.0005267   8.963 2.79e-16 ***
## GroupB:cal   -0.0007619  0.0007534  -1.011  0.313
## GroupC:cal   -0.0002977  0.0007980  -0.373  0.710
## GroupT:cal    0.0005381  0.0007936   0.678  0.499
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.208 on 192 degrees of freedom
## Multiple R-squared:  0.6349, Adjusted R-squared:  0.6216
```

```
## F-statistic: 47.69 on 7 and 192 DF, p-value: < 2.2e-16
mod0=lm(G3~1,data=don)
anova(mod0,mod2)

## Analysis of Variance Table
##
## Model 1: G3 ~ 1
## Model 2: G3 ~ Group * cal
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      199 767.02
## 2      192 280.05  7    486.97 47.694 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(mod2)

## Analysis of Variance Table
##
## Response: G3
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Group      3  97.03   32.34  22.1743 2.259e-12 ***
## cal        1 385.89  385.89 264.5599 < 2.2e-16 ***
## Group:cal   3   4.05    1.35   0.9252   0.4296
## Residuals 192 280.05    1.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

mod2b=lm(G3~Group+cal,data=don)
summary(mod2b)

##
## Call:
## lm(formula = G3 ~ Group + cal, data = don)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.3578 -0.7617  0.1743  0.7119  2.8506
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.4900837  0.5425196 -15.649  < 2e-16 ***
## GroupB       0.1491949  0.2428238   0.614    0.54
## GroupC      -1.5575048  0.2417631  -6.442 8.99e-10 ***
## GroupT      -0.1826719  0.2414697  -0.757    0.45
## cal          0.0045695  0.0002808  16.275  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.207 on 195 degrees of freedom
## Multiple R-squared:  0.6296, Adjusted R-squared:  0.622
## F-statistic: 82.87 on 4 and 195 DF, p-value: < 2.2e-16

anova(mod2b)

## Analysis of Variance Table
##
```

```
## Response: G3
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Group      3  97.03   32.34   22.20 2.073e-12 ***
## cal        1 385.89  385.89  264.86 < 2.2e-16 ***
## Residuals 195 284.10    1.46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(car)
```

```
## Loading required package: carData
```

```
Anova(mod2b)
```

```
## Anova Table (Type II tests)
##
## Response: G3
##           Sum Sq Df F value    Pr(>F)
## Group      92.42  3  21.146 6.665e-12 ***
## cal       385.89  1 264.865 < 2.2e-16 ***
## Residuals 284.10 195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(lsmeans)
```

```
## Warning: package 'lsmeans' was built under R version 3.5.3
```

```
## Loading required package: emmeans
```

```
## Warning: package 'emmeans' was built under R version 3.5.3
```

```
## The 'lsmeans' package is now basically a front end for 'emmeans'.
## Users are encouraged to switch the rest of the way.
## See help('transition') for more information, including how to
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.
```

```
lsmeans(mod2b, pairwise ~ Group, adjust = 'bonferroni')
```

```
## $lsmeans
##   Group lsmean    SE df lower.CL upper.CL
## A      0.0283 0.171 195  -0.403    0.459
## B      0.1775 0.172 195  -0.255    0.610
## C     -1.5292 0.171 195  -1.960   -1.099
## T     -0.1544 0.171 195  -0.586    0.277
##
## Confidence level used: 0.95
## Conf-level adjustment: bonferroni method for 4 estimates
##
## $contrasts
##   contrast estimate    SE df t.ratio p.value
## A - B      -0.149 0.243 195  -0.614 1.0000
## A - C       1.558 0.242 195   6.442 <.0001
## A - T       0.183 0.241 195   0.756 1.0000
## B - C       1.707 0.242 195   7.059 <.0001
## B - T       0.332 0.243 195   1.363 1.0000
## C - T      -1.375 0.242 195  -5.678 <.0001
##
```

P value adjustment: bonferroni method for 6 tests

1. Écrire le modèle `mod2` correspondant (sans oublier les hypothèses et les gammes de variation des indices).
2. Quelles sont les contraintes utilisées par le logiciel R ?
3. Sur la ligne `GroupeB:cal`, interpréter la valeur -0.0007619 dans la sortie `summary`.
4. Justifier le passage au modèle `mod2b`.
5. Le groupe a-t-il une influence sur la variation de poids (à justifier soigneusement) ? Détailler les hypothèses comparées par le test que vous utilisez.
6. Donner une estimation de la variation de poids moyenne pour un individu ayant suivi le régime A ayant eu un apport moyen de 2000 kcal ? Même question pour un individu ayant suivi le régime T avec le même apport de 2000 kcal.
7. Donner l'expression des moyennes ajustées dans ce modèle et leurs valeurs estimées pour les 4 régimes.
8. Expliquer en quoi consiste la méthode de Bonferroni pour les tests multiples. Indiquer quel(s) groupe(s) est (sont) significativement différent(s) des autres. Interpréter.