

Ancova. MtCars

Sophie Donnet

3 novembre 2017

Nous nous intéressons à la consommation de voitures en fonction de leur puissance et du type de transmission (automatique ou manuelle).

```
data(mtcars)
attach(mtcars)
don <- mtcars[,c("am", "mpg", "hp")]
don$am <- as.factor(don$am)
```

Un extrait des données est fourni ci-dessous.

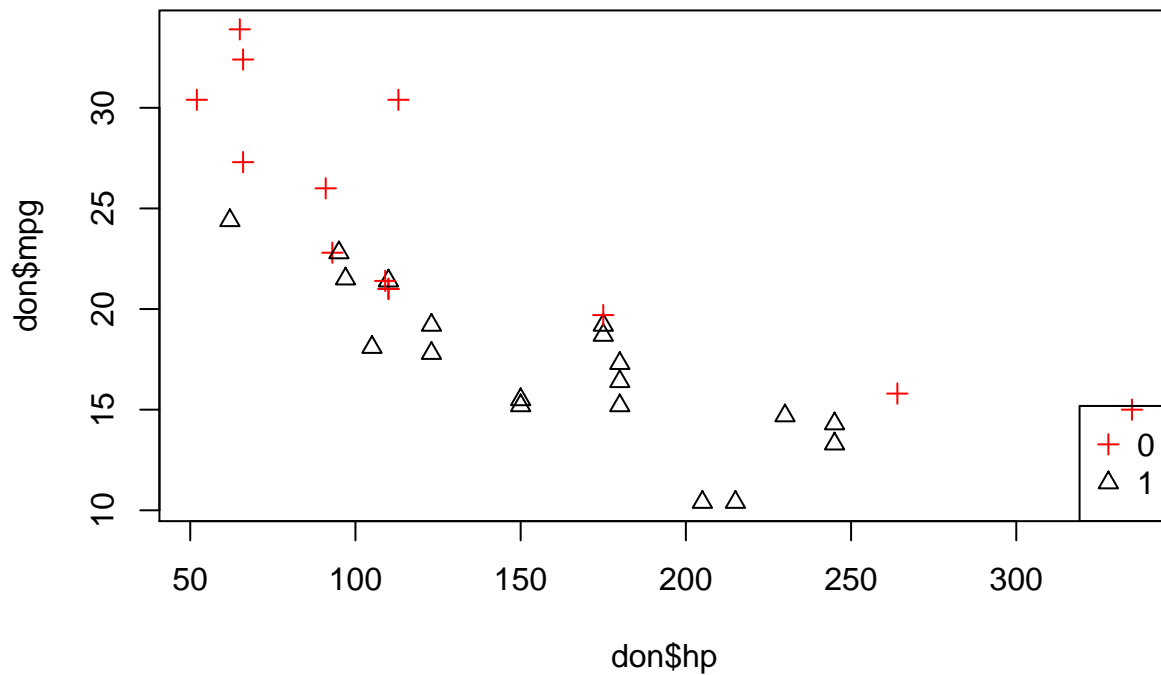
```
head(don)
```

	am	mpg	hp
## Mazda RX4	1	21.0	110
## Mazda RX4 Wag	1	21.0	110
## Datsun 710	1	22.8	93
## Hornet 4 Drive	0	21.4	110
## Hornet Sportabout	0	18.7	175
## Valiant	0	18.1	105

```
tail(don)
```

	am	mpg	hp
## Porsche 914-2	1	26.0	91
## Lotus Europa	1	30.4	113
## Ford Pantera L	1	15.8	264
## Ferrari Dino	1	19.7	175
## Maserati Bora	1	15.0	335
## Volvo 142E	1	21.4	109

On s'intéresse à présent à la variation de la consommation en fonction de la puissance et du type de transmission.



```
mod2 = lm(mpg~am*hp,data=don)
summary(mod2)
```

```
##
## Call:
## lm(formula = mpg ~ am * hp, data = don)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.3818 -2.2696  0.1344  1.7058  5.8752
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  26.6248479   2.1829432   12.197 1.01e-12 ***
## am1          5.2176534   2.6650931    1.958  0.0603 .
## hp          -0.0591370   0.0129449   -4.568 9.02e-05 ***
## am1:hp        0.0004029   0.0164602    0.024  0.9806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.961 on 28 degrees of freedom
## Multiple R-squared:  0.782, Adjusted R-squared:  0.7587
## F-statistic: 33.49 on 3 and 28 DF, p-value: 2.112e-09

mod0 = lm(mpg~1,data=don)
anova(mod0,mod2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ 1
## Model 2: mpg ~ am * hp
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      31 1126.05
## 2      28  245.43   3    880.61 33.488 2.112e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(mod2)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## am          1 405.15  405.15 46.2210 2.197e-07 ***
## hp          1 475.46  475.46 54.2419 5.088e-08 ***
## am:hp        1   0.01    0.01  0.0006   0.9806
## Residuals   28 245.43    8.77
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mod2b=lm(mpg~am + hp,data = don)
summary(mod2b)
```

```
##
## Call:
## lm(formula = mpg ~ am + hp, data = don)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.3843 -2.2642  0.1366  1.6968  5.8657
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26.584914   1.425094  18.655 < 2e-16 ***
## am1         5.277085   1.079541   4.888 3.46e-05 ***
## hp        -0.058888   0.007857  -7.495 2.92e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.909 on 29 degrees of freedom
## Multiple R-squared:  0.782, Adjusted R-squared:  0.767
## F-statistic: 52.02 on 2 and 29 DF,  p-value: 2.55e-10
```

```
anova(mod2b)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## am          1 405.15  405.15 47.871 1.327e-07 ***
## hp          1 475.46  475.46 56.178 2.920e-08 ***
## Residuals   29 245.44    8.46
## ---
```

```
## 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 ~ am, adjust = 'bonferroni')
```

```
## $lsmeans
```

```
##   am lsmean      SE df lower.CL upper.CL
```

```
## 0    17.9 0.676 29      16.3      19.5
```

```
## 1    23.2 0.822 29      21.3      25.2
```

```
##
```

```
## Confidence level used: 0.95
```

```
## Conf-level adjustment: bonferroni method for 2 estimates
```

```
##
```

```
## $contrasts
```

```
##   contrast estimate      SE df t.ratio p.value
```

```
## 0 - 1          -5.28 1.08 29 -4.888  <.0001
```

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 consommation moyenne (miles / gallon) pour un véhicule manuel de puissance 150? Même question pour un véhicule automatique?
7. 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.