

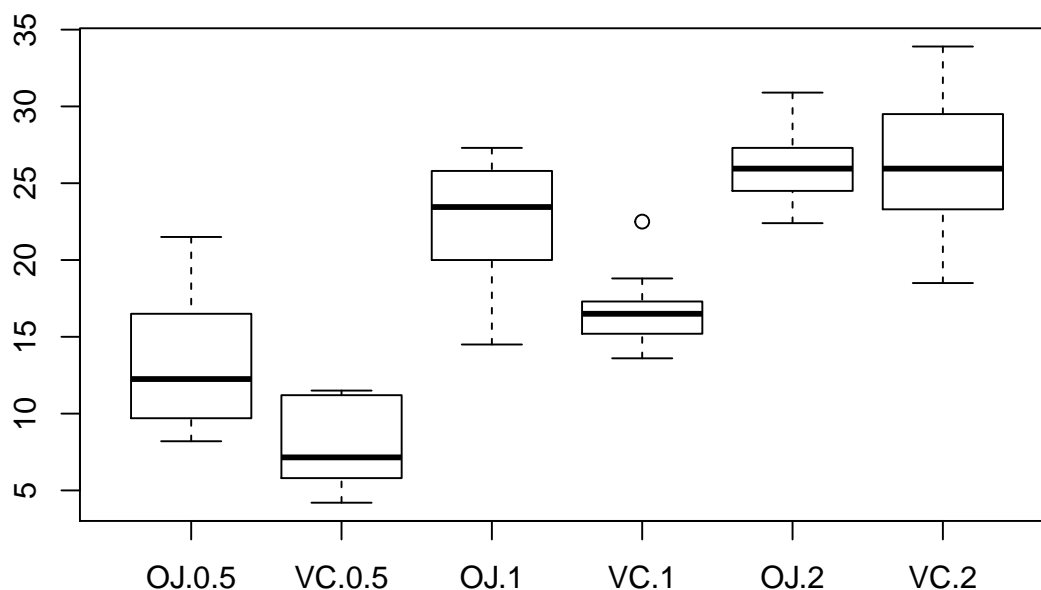
Tooth Growth (Anova 2 facteurs). Corrigé

Nous considérons le jeu de données `ToothGrowth`. La réponse Y est la longueur des odontoblastes (cellules intervenant dans la croissance des dents) chez $n = 60$ cochons de Guinée. Chaque animal a reçu une des trois doses possibles de vitamine C (0.5, 1, and 2 mg/day) par le biais d'une des deux méthodes d'administration (jus d'orange ou acide ascorbique). On s'intéresse à l'influence de ces facteurs sur la croissance dentaire. Le jeu de données est représenté par le box-plot suivant. (Pour l'exercice, nous avons transformé la variable quantitative `dose` en un facteur `doselevel`).

```
ToothGrowth$doselevel = as.factor(ToothGrowth$dose)
summary(ToothGrowth)
```

```
##      len      supp      dose      doselevel
##  Min.   : 4.20   OJ:30   Min.   :0.500   0.5:20
## 1st Qu.:13.07   VC:30   1st Qu.:0.500   1  :20
## Median :19.25           Median :1.000   2  :20
## Mean   :18.81           Mean   :1.167
## 3rd Qu.:25.27           3rd Qu.:2.000
## Max.   :33.90           Max.   :2.000
```

```
names(ToothGrowth)=c('len', 'suppfactor', 'dose', 'doselevel')
boxplot(len~suppfactor*doselevel,data=ToothGrowth)
```



```
table(ToothGrowth$doselevel, ToothGrowth$suppfactor)
```

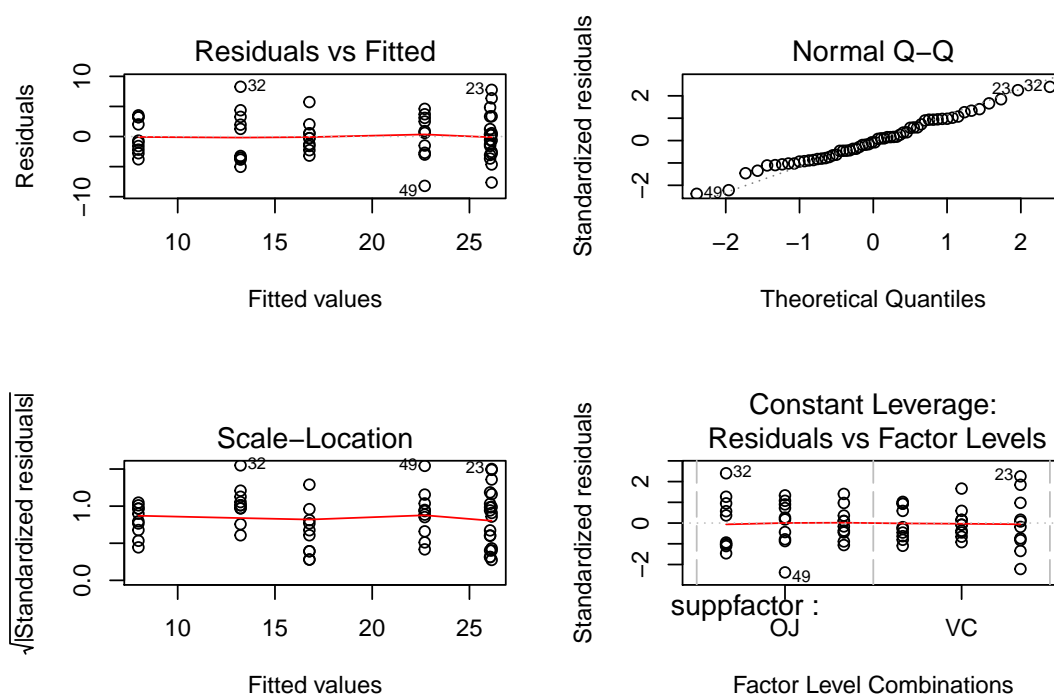
```
##
##      OJ VC
## 0.5 10 10
## 1   10 10
## 2   10 10
```

- Instruction 1

```
mod2=lm(len~suppfactor*doselevel,data=ToothGrowth)
```

```
summary(mod2)
```

```
##
## Call:
## lm(formula = len ~ suppfactor * doselevel, data = ToothGrowth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.20  -2.72  -0.27   2.65   8.27
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      13.230      1.148  11.521 3.60e-16 ***
## suppfactorVC       -5.250      1.624   -3.233 0.00209 **
## doselevel1         9.470      1.624    5.831 3.18e-07 ***
## doselevel2        12.830      1.624    7.900 1.43e-10 ***
## suppfactorVC:doselevel1 -0.680      2.297   -0.296 0.76831
## suppfactorVC:doselevel2  5.330      2.297    2.321 0.02411 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.631 on 54 degrees of freedom
## Multiple R-squared:  0.7937, Adjusted R-squared:  0.7746
## F-statistic: 41.56 on 5 and 54 DF,  p-value: < 2.2e-16
par(mfrow=c(2,2))
plot(mod2)
```



- **Instruction 2**
`anova(mod2)`

```
## Analysis of Variance Table
##
## Response: len
##              Df Sum Sq Mean Sq F value    Pr(>F)
## suppfactor      1  205.35   205.35   15.572 0.0002312 ***
## doselevel       2 2426.43  1213.22   92.000 < 2.2e-16 ***
## suppfactor:doselevel 2  108.32    54.16    4.107 0.0218603 *
## Residuals      54  712.11    13.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- **Instruction 3**
`Anova(mod2)`

```
## Anova Table (Type II tests)
##
## Response: len
##              Sum Sq Df F value    Pr(>F)
## suppfactor      205.35  1   15.572 0.0002312 ***
## doselevel      2426.43  2   92.000 < 2.2e-16 ***
## suppfactor:doselevel 108.32  2    4.107 0.0218603 *
## Residuals      712.11 54
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1. Écrire le modèle `mod2` correspondant (sans oublier les hypothèses et les gammes de variation des indices).

Anova à deux facteurs :

$$Y_{ijk} = \mu + \alpha_j + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$$i = 1, 2, j = 1, 2, k = 1 \dots, 10$$

2. Quelles sont les contraintes utilisées par le logiciel R ?

$$\alpha_1 = \beta_1 = \gamma_{1j} = \gamma_{i1} = 0$$

3. Les hypothèses du modèle linéaire sont elles vérifiées ?

Oui on regarde les plots des résidus comme d'hab

4. A-t-on un effet significatif des facteurs sur la croissance ? Quid de l'interaction ?

Voir tests des intructions 1 et 2 : tout est significatif