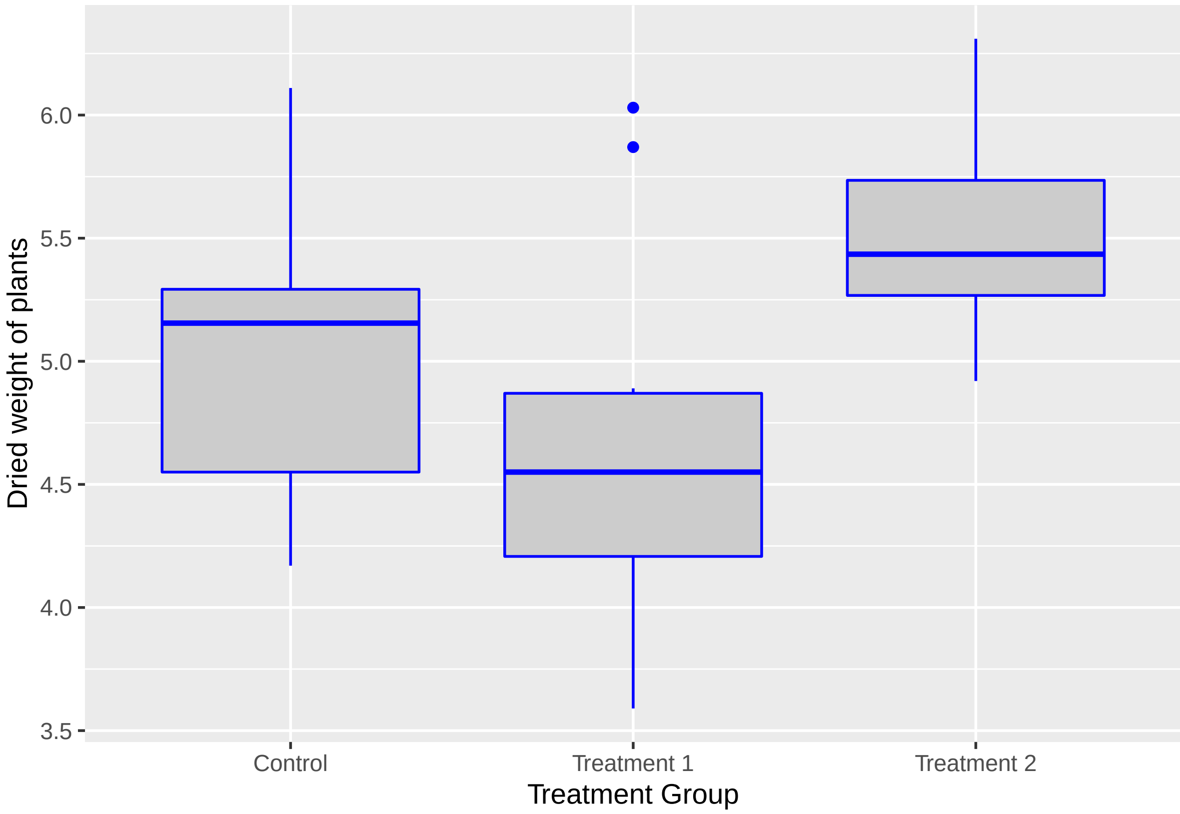
Plant Growth. Corrigé

On s'intéresse aux résultats d'une expérience cherchant à expliquer un rendement agricole (mesuré en poids sec de plantes) en fonction du traitement (2 traitements différents et un groupe de contrôle). On dispose de 30 observations réparties de façon équilibrées dans les trois modalités. Les données sont représentées sous la forme de boxplot sur la figure suivante :



Les instructions sont données ci-dessous.

- Instruction 1

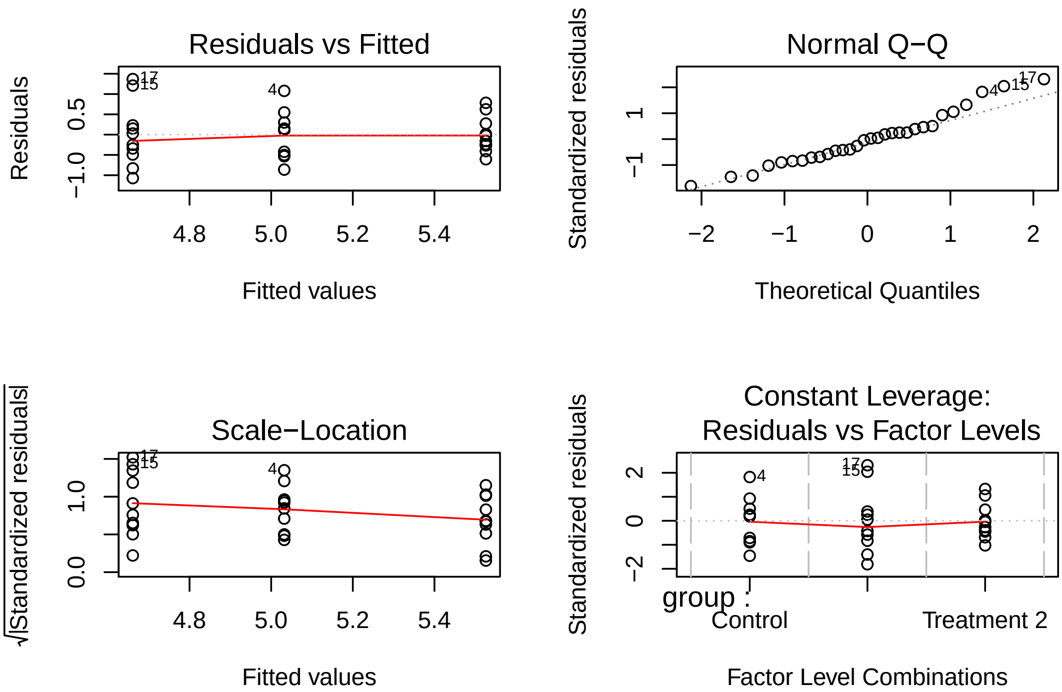
```
plant.mod1 = lm(weight ~ group, data = plant.df)
summary(plant.mod1)
```

```
##
## Call:
## lm(formula = weight ~ group, data = plant.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0710 -0.4180 -0.0060  0.2627  1.3690
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.0320     0.1971  25.527   <2e-16 ***
```

```
## groupTreatment 1  -0.3710     0.2788  -1.331   0.1944
## groupTreatment 2   0.4940     0.2788   1.772   0.0877 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6234 on 27 degrees of freedom
## Multiple R-squared:  0.2641, Adjusted R-squared:  0.2096
## F-statistic: 4.846 on 2 and 27 DF,  p-value: 0.01591
```

- Instruction 2

```r
par(mfrow=c(2,2))
plot(plant.mod1)
```



- Instruction 3

```r
contrasts(plant.df$group) <- contr.sum
plant.mod2 = lm(weight ~ group, data = plant.df)
summary(plant.mod2)
```

```
##
## Call:
## lm(formula = weight ~ group, data = plant.df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0710 -0.4180 -0.0060  0.2627  1.3690
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.0730     0.1138  44.573   <2e-16 ***
```

```
## group1       -0.0410     0.1610  -0.255   0.8009
## group2       -0.4120     0.1610  -2.560   0.0164 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6234 on 27 degrees of freedom
## Multiple R-squared:  0.2641, Adjusted R-squared:  0.2096
## F-statistic: 4.846 on 2 and 27 DF,  p-value: 0.01591
```

- Instruction 4

```
anova(plant.mod1)
```

```
## Analysis of Variance Table
##
## Response: weight
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## group      2  3.7663  1.8832  4.8461 0.01591 *
## Residuals 27 10.4921  0.3886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Instruction 5

```
test.treatment = pairwise.t.test(plant.df$weight,plant.df$group,

                p.adjust.method="bonferroni")
test.treatment
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  plant.df$weight and plant.df$group
##
##             Control Treatment 1
## Treatment 1 0.583   -
## Treatment 2 0.263   0.013
##
## P value adjustment method: bonferroni
```

1. Quelle contrainte est utilisée dans Instruction 1 ?

   Contrainte standard dans R : $\alpha_1 = 0$

2. Les hypothèses du modèle linéaire sont-elles vérifiées ?

   Analyse classique des 4 graphes de daignostique. HG : pas de tendance donc on n' a pas oublié de grande tendance. BG : pas de preuve d'hétéroscédasticité. HD : résidus gaussiens. BD : pas de points abérrants.

3. Que fait-on dans Instruction 3 ?

   On change la contrainte. On impose que la somme des α_i vaut 0

4. Rappelez l'expression de la statistique du test du modèle. Rejette-t-on $\mathcal{H}_0$ ici ?

$$\frac{(SCT - SCR)/(I-1)}{SCR/(n-I)} \sim \mathcal{F}_{I-1,n-1}$$

$I = 3$, $n = 30$ p-value vaut 0.015091. On rejette $\mathcal{B}_0$ au niveau 5%.

5. Que pensez-vous du R^2.

R^2 faible car modèle est pauvre. Mais oui il y a un effet du traitement du la taille des graines.

6. Interpréter les sorties Coefficients.

Il faut les interpréter comme des écarts à la référence. Donc soit le groupe 1 (ici controle) soit la moyenne des moyennes entre traitement.

7. Que fait-on dans l'$\boxed{\text{instruction 5}}$. Rappelez la statistique de test. A quoi correspond l'instruction Bonferroni ? Interprétez les résultats.

Si la correction de Bonferroni a été mise, alors c'est que pour chaque test on a multiplié la p-value par le nombre de tests faits : $I(I-1)/2$. On continue à comparer ce chiffre à 5%.