# Completely Randomized Designs - Lab

## Blake Pappas

## December 17, 2023

The dataset `PlantGrowth` contains results from an experiment to compare yields (as measured by dried weight of plants) obtained under a control and two different treatment conditions.

*Data Source:* Dobson, A. J. (1983) *An Introduction to Statistical Modelling.* London: Chapman and Hall.

Let's load the data first:

**Code:**

```
data(PlantGrowth)
str(PlantGrowth)
```

```
## 'data.frame':    30 obs. of  2 variables:
##  $ weight: num  4.17 5.58 5.18 6.11 4.5 4.61 5.17 4.53 5.33 5.14 ...
##  $ group : Factor w/ 3 levels "ctrl","trt1",..: 1 1 1 1 1 1 1 1 1 1 ...
```

1. Compute the treatment means and standard deviations.

**Code:**

```
(means <- tapply(PlantGrowth$weight, PlantGrowth$group, mean))
```

```
##  ctrl  trt1  trt2
## 5.032 4.661 5.526
```
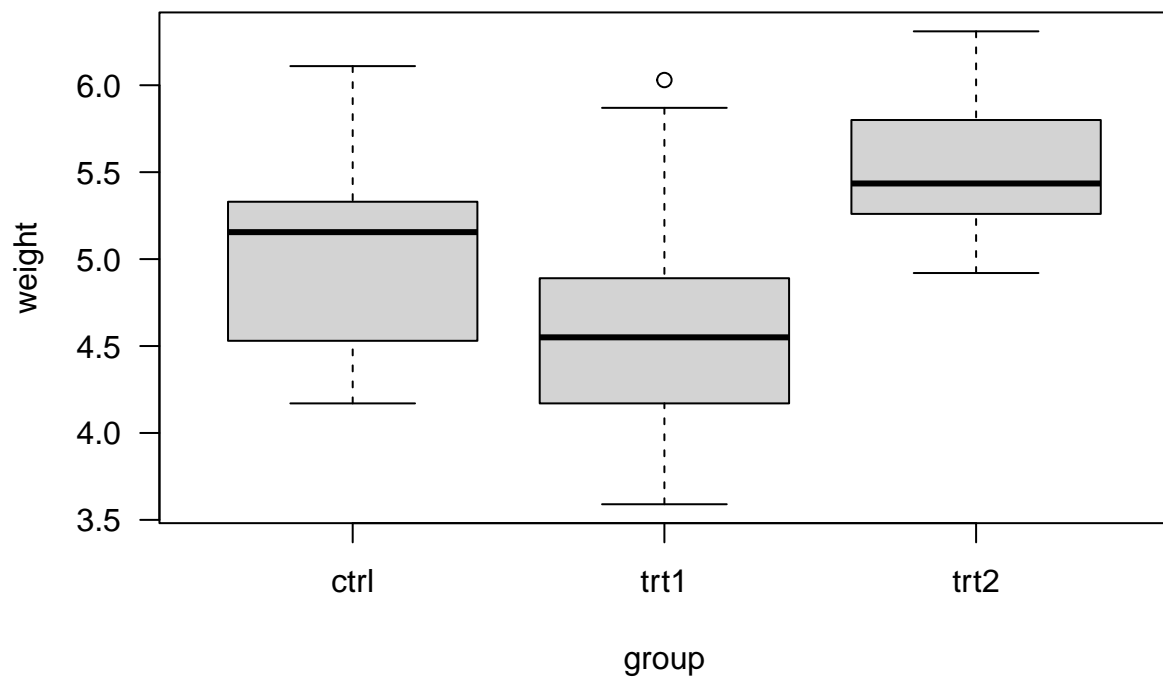
```
(vars <- tapply(PlantGrowth$weight, PlantGrowth$group, sd))
```

```
##      ctrl      trt1      trt2
## 0.5830914 0.7936757 0.4425733
```
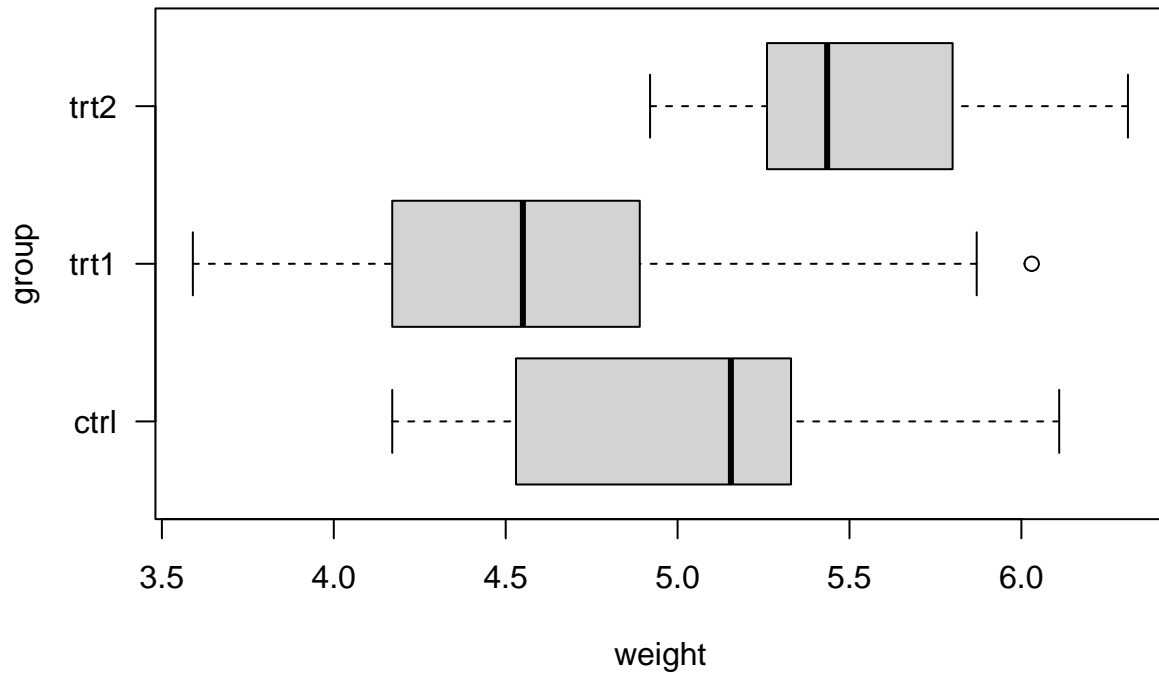
2. Make side-by-side boxplots by treatment.

**Code:**

```
boxplot(weight ~ group, data = PlantGrowth, las = 1)
```

```
boxplot(weight ~ group, data = PlantGrowth, las = 1, horizontal = T)
```

3. Write down the effects model and explain each term in the model (including the model assumptions regarding the random error).

**Answer:** $Y_{ij} : \mu + \alpha_i + \epsilon_{ij}$, $i = 1, \ldots, g$, $j = 1, \ldots, n_i$, $\epsilon_{ij} \sim N(0, \sigma\text{ ^2})$. $Y_{ij}$ **is the random variable that represents the response for the jth experimental unit to treatment i.** $\mu_i$ **is the mean response time for the ith treatment.** $\epsilon_{ij}$ **is the random error, which is assumed to be normal.**

4. Perform an overall F-test using ANOVA. State the hypotheses, p-value, decision, and conclusion.

**Code:**

```
AOV <- aov(weight ~ group, data = PlantGrowth)
summary(AOV)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## group         2  3.766  1.8832   4.846 0.0159 *
## Residuals    27 10.492  0.3886
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**Answer: The null hypothesis is** $H_0 : \alpha_i = 0$, **the alternative hypothesis is** $H_a : \alpha_i \neq 0$, **the** $\alpha$ **is 0.05, the F-statistic is 4.846, the p-value is 0.0159, and the decision is to reject the null hypothesis. The data shows evidence of differences the three treatment groups.**

5. Conduct pairwise comparisons using Tukey's HSD procedure.

**Code:**

```r
# HSD
HSD <- TukeyHSD(AOV, conf.level = 0.95)
HSD$group
```

```
##               diff        lwr       upr       p adj
## trt1-ctrl -0.371 -1.0622161 0.3202161 0.39087114
## trt2-ctrl  0.494 -0.1972161 1.1852161 0.19799599
## trt2-trt1  0.865  0.1737839 1.5562161 0.01200642
```
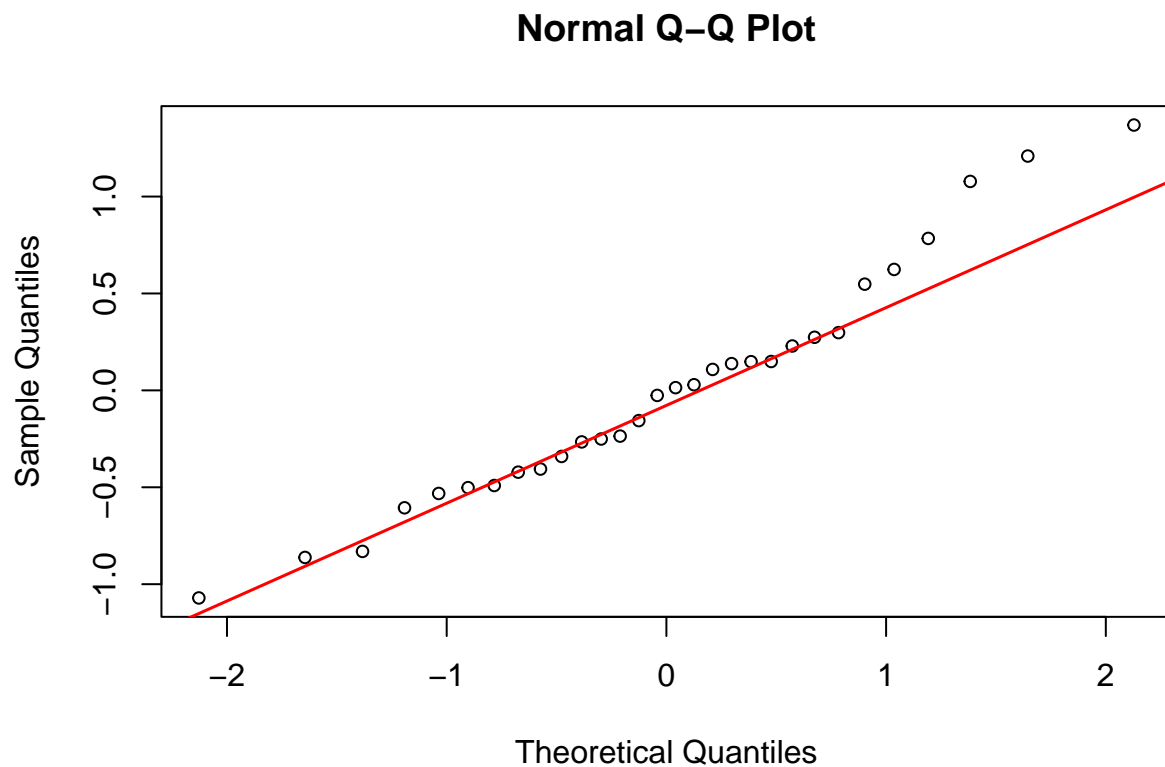
**Answer: The trt2-trt1 groups have statistically significant differences.**

6. Use qqplot to examine the normality assumption on error.

**Code:**

```r
mod1 <- lm(weight ~ group, data = PlantGrowth)

qqnorm(mod1$resid, cex = 0.8)
qqline(mod1$resid, col = "red", lwd = 1.5)
```
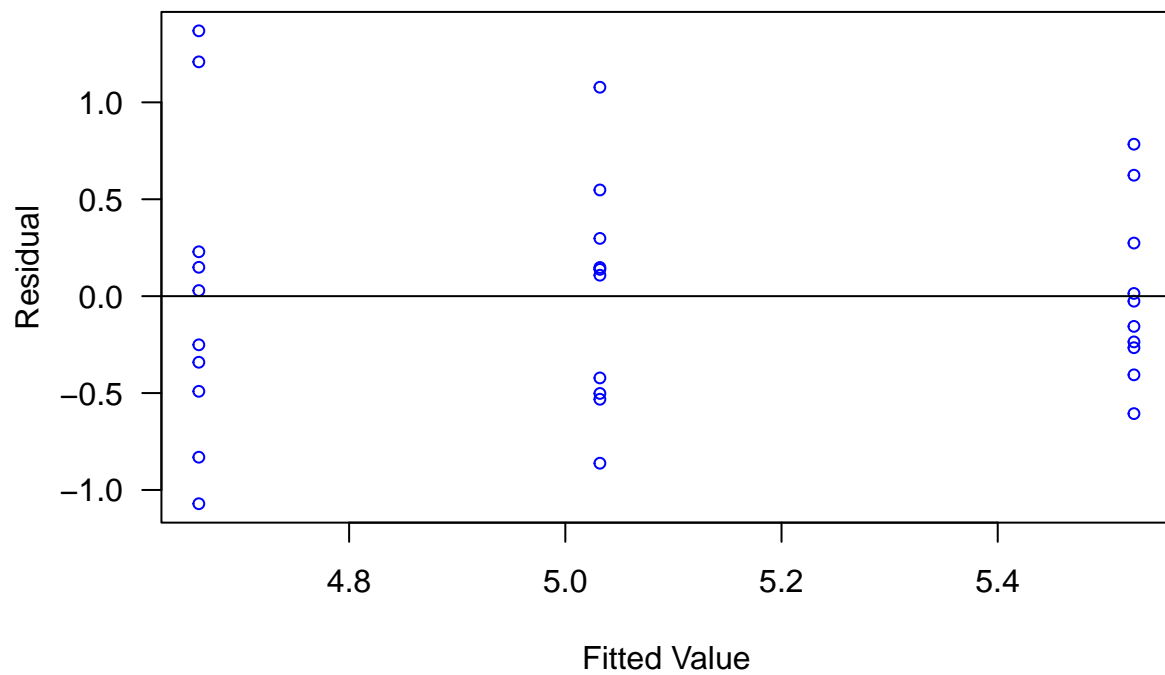
## Normal Q–Q Plot



**Answer: Based on the plot, the modeling assumption of normality appears to be reasonable. The pattern of the plot runs very close to the trend line.**
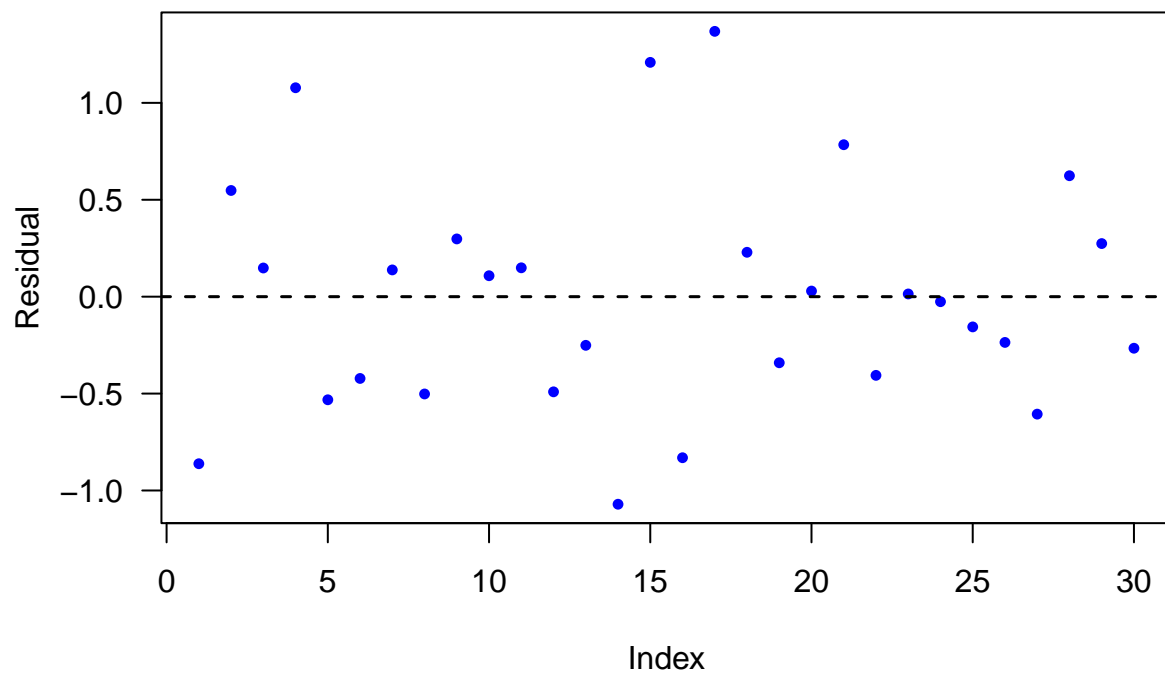
7. Make a residual plot to assess equal variance assumption.

**Code:**

```r
plot(mod1$fitted, mod1$resid, las = 1, xlab = "Fitted Value", ylab = "Residual", cex = 0.75, col = "blu
abline(h = 0)
```
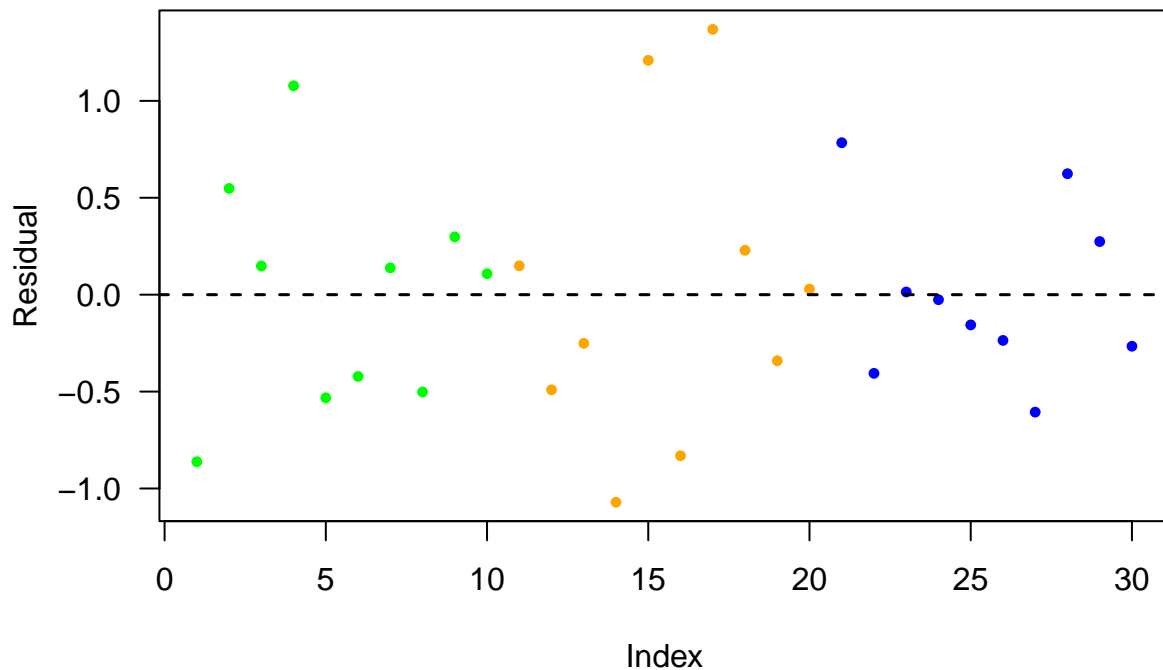


```r
plot(mod1$resid, col = "blue", las = 1, cex = 0.75, pch = 16, ylab = "Residual")
abline(h = 0, lty = 2, lwd = 1.5)
```

```r
# By Treatment
attach(PlantGrowth)
colorf <- as.factor(PlantGrowth$group)
colors <- c("green",
"orange", "blue")

plot(mod1$resid, col = colors[colorf], las = 1, cex = 0.75,
     pch = 16, ylab = "Residual")
abline(h = 0, lty = 2, lwd = 1.5)
```

**Answer: Based on the residual plots, the assumption of equal variances seems to be reasonable.**

8. Perform a Levene's test for equal variance.

**Code:**

```
# Levene's Test for Equal Variance
# install.packages("lawstat")
library("lawstat")
levene.test(PlantGrowth$weight, PlantGrowth$group, location = "mean")
```

```
##
##  Classical Levene's test based on the absolute deviations from the mean
##  ( none not applied because the location is not set to median )
##
## data:  PlantGrowth$weight
## Test Statistic = 1.237, p-value = 0.3062
```

```
# Brown-Forsythe Test
levene.test(PlantGrowth$weight, PlantGrowth$group, location = "median")
```

```
##
##  Modified robust Brown-Forsythe Levene-type test based on the absolute
##  deviations from the median
##
```

```
## data:  PlantGrowth$weight
## Test Statistic = 1.1192, p-value = 0.3412
```