DSA 8020 R Lab 7: Completely Randomized Designs

Whitney

March 24, 2021

Contents

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
```
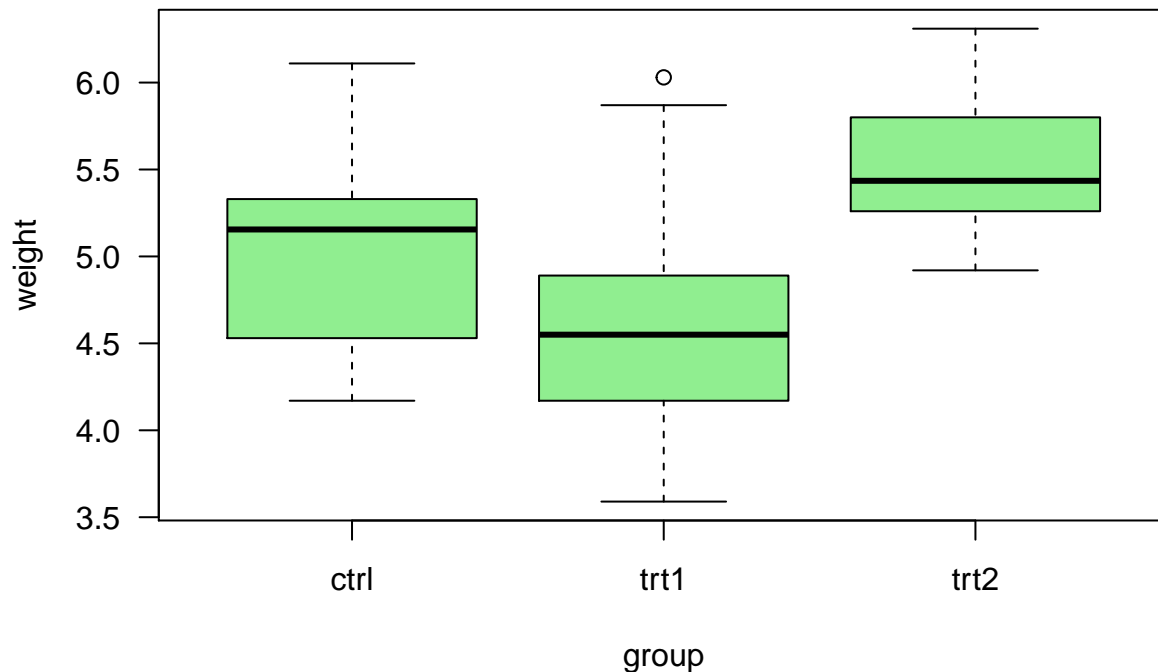
```
(sd <- tapply(PlantGrowth$weight, PlantGrowth$group, sd))
```

```
##      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, col = "lightgreen")
```

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

Answer:

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, 3, \ j = 1, \cdots, 10, \quad \epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$$

- μ: Overall mean
- α_i: ith treatment effect
- ϵ: independent random error term

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:

$H_0 : \alpha_1 = \alpha_2 = alpha_3 = 0$ $H_a :$ at least one of α_1, α_2, and $\alpha_3 \neq 0$

P-value = c(0.0159099583256229, NA) < 0.05 $\Rightarrow$ rejection H_0 and conclude that we have sufficient evidence that yield is affected by the treatment conditions at 0.05 level.

5. Conduct pairwise comparisons using Tukey's HSD procedure

Code:

```
(HSD <-TukeyHSD(AOV, conf.level = 0.95))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = weight ~ group, data = PlantGrowth)
##
## $group
##              diff        lwr       upr      p adj
## trt1-ctrl -0.371 -1.0622161 0.3202161 0.3908711
## trt2-ctrl  0.494 -0.1972161 1.1852161 0.1979960
## trt2-trt1  0.865  0.1737839 1.5562161 0.0120064
```
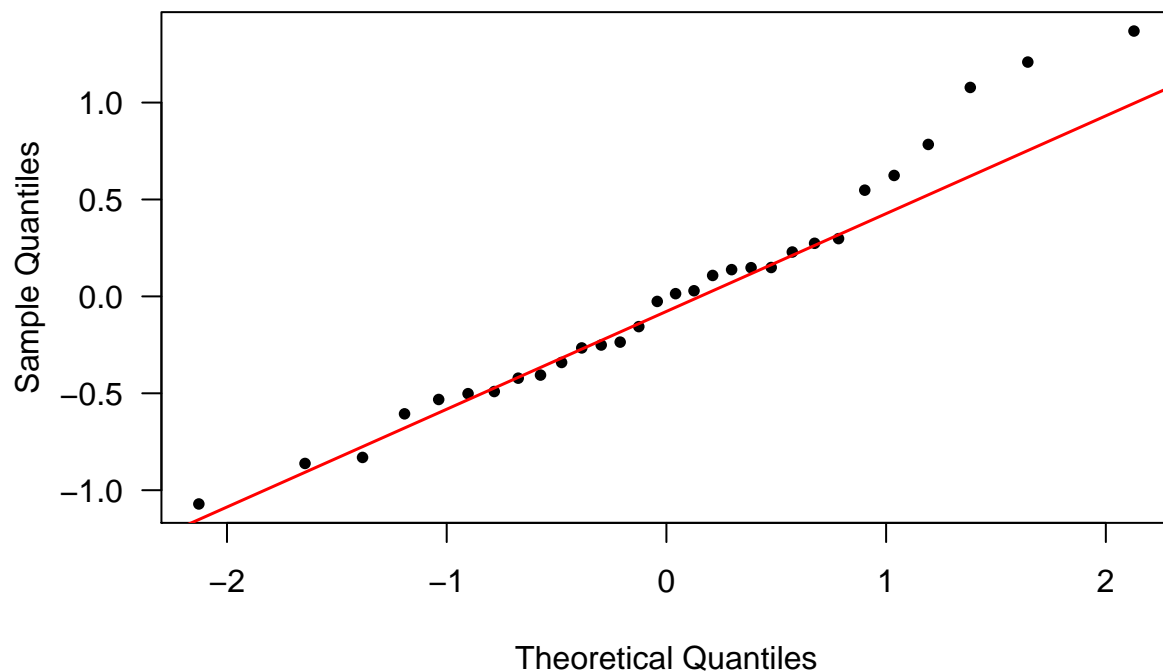
Answer:

Only the difference between treatment 1 and treatment 2 is significant.

6. Use qqplot to examine the normality assumption on error

Code:

```
qqnorm(AOV$resid, cex = 0.8, las = 1, pch = 16)
qqline(AOV$resid, col = "red", lwd = 1.5)
```
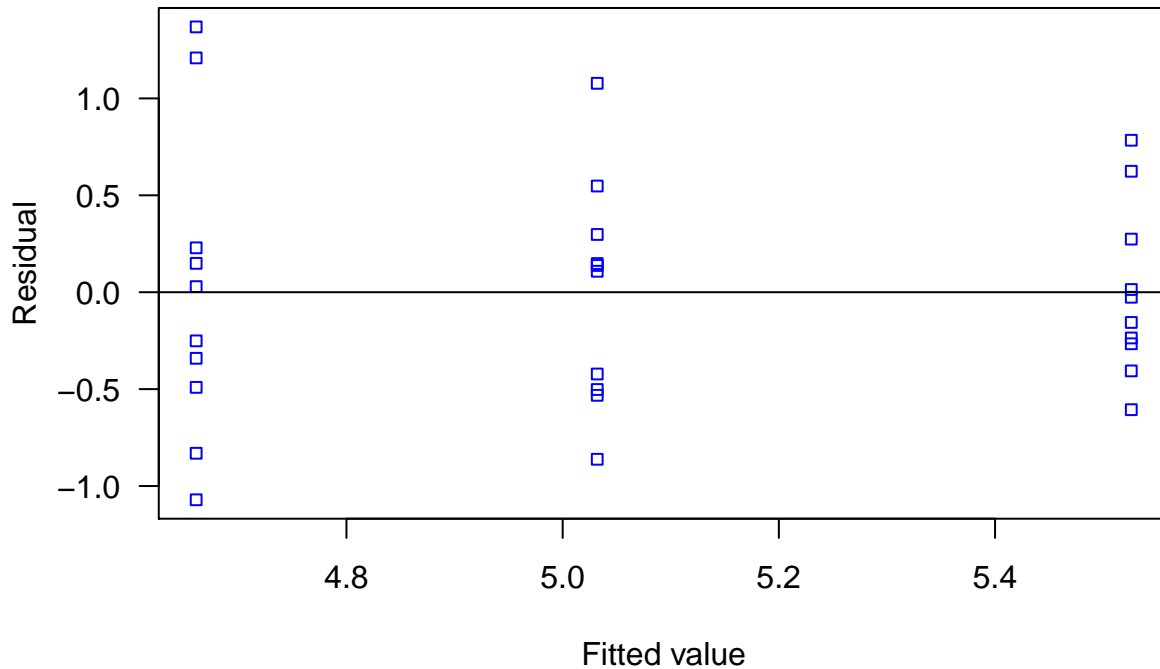
Normal Q–Q Plot

Answer:

It is reasonable to assume the random error term follows a normal distribution as the data points fall on an approximately straight line.

3

7. Make a residual plot to assess equal variance assmuption

Code:

```
plot(AOV$fitted, AOV$resid, las = 1, xlab = "Fitted value", ylab = "Residual",
     cex = 0.75, pch = 0, col = "blue")
abline(h = 0)
```


Answer:

The residuals have greater variance for smaller fitted values.

8. Perform a Levene's test for equal variance

Code:

```
library(car)
```

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

```
leveneTest(PlantGrowth$weight, PlantGrowth$group)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  1.1192 0.3412
##       27
```

```
med <- with(PlantGrowth, tapply(weight, group, median))
wt <- with(PlantGrowth, abs(weight - med[group]))
anova(lm(wt ~ group, PlantGrowth))
```

```
## Analysis of Variance Table
## 
## Response: wt
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      2 0.3495 0.17476  1.1192 0.3412
## Residuals 27 4.2161 0.15615
```

Answer:

Since the p-value = 0.3412 > 0.05, we conclude that we do not have sufficient evidence of a non-constant variance.