Homework №3

David Oniani

February 27, 2021

Contents

Setup

```r
library(MASS)
library(Stat2Data)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:MASS':
##
##     select

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(emmeans)
library(ggplot2)

# I like the minimal theme
theme_set(theme_minimal())

# Disable warnings (they clutter the document)
options(warn = -1)
```

Exercise 6.10

(a) $3 - 1 = 2$. Hence, we have 2 degrees of freedom for factor A.

(b) $3 - 1 = 2$. Hence, we have 2 degrees of freedom for factor B.

(c) $(3 - 1) \times (3 - 1) = 2 \times 2 = 4$. Hence, we have 4 degrees of freedom for Error.

Exercise 6.26

(a) This is an observational study as the study did not affect or control the subjects in any way.

(b)   – Factors of interest: word lists (note that the study suggested that within each word list, words are of equal hearing difficulty, but we do not know if this is the case across the word lists/groups).

  – Nuisance factors: volume and background noise

(c) Subjects with normal hearing.

(d) We know that the variability within blocks is less than the variability between blocks (since within blocks/word lists, words are of equal hearing difficulty). Hence, this is a randomized block design with 4 levels. Since there are 24 subjects in the study, the block size is 24.

```
data(HearingTest)

# Data processing
df1 = HearingTest

l1 = df1[df1$List == "L1",]$Percent
l2 = df1[df1$List == "L2",]$Percent
l3 = df1[df1$List == "L3",]$Percent
l4 = df1[df1$List == "L4",]$Percent

df2 = data.frame(c(1:24), l1, l2, l3, l4)
df2$Mean = rowMeans(df2[,-1])
colnames(df2) = c("Sub", "L1", "L2", "L3", "L4", "Mean")

mean_l1 = round(mean(df2$L1))
mean_l2 = round(mean(df2$L2))
mean_l3 = round(mean(df2$L3))
mean_l4 = round(mean(df2$L4))
mean_mean = round(mean(df2$Mean), 1)

tmp = data.frame(nrow(df2) + 1, mean_l1, mean_l2, mean_l3, mean_l4, mean_mean)
colnames(tmp) = c("Sub", "L1", "L2", "L3", "L4", "Mean")
df3 = rbind(df2, tmp)
rownames(df3) = c(c(1:24), "Mean")
```

Exercise 6.36

```r
data("Alfalfa")

# (a)
df4 = Alfalfa$Ht4
mean_df4 = mean(df4)
sd_df4 = sd(df4)

print(mean_df4)
```

```
## [1] 1.74
```

```r
print(sd_df4)
```

```
## [1] 1.105396
```

```r
lr = lm(Ht4 ~ Acid + Row, Alfalfa)
summary(lr)
```

```
##
## Call:
## lm(formula = Ht4 ~ Acid + Row, data = Alfalfa)
##
## Residuals:
##    Min    1Q Median    3Q   Max
## -0.860 -0.618  0.046  0.350  1.460
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.8860     0.5952   1.489   0.1749
## Acid3.0HCl   -0.3820     0.5510  -0.693   0.5078
## Acidwater     1.2040     0.5510   2.185   0.0604 .
## Rowb          0.4100     0.7114   0.576   0.5802
## Rowc          0.0900     0.7114   0.127   0.9024
## Rowd          1.1000     0.7114   1.546   0.1606
## Rowe          1.3000     0.7114   1.827   0.1050
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8712 on 8 degrees of freedom
## Multiple R-squared:  0.645,  Adjusted R-squared:  0.3788
## F-statistic: 2.423 on 6 and 8 DF,  p-value: 0.1229
```

```r
# (b)
# ANOVA
s = summary(aov(Ht4 ~ Acid + Row, data=Alfalfa))
print(s)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Acid         2  6.852   3.426   4.513 0.0487 *
## Row          4  4.183   1.046   1.378 0.3235
## Residuals    8  6.072   0.759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
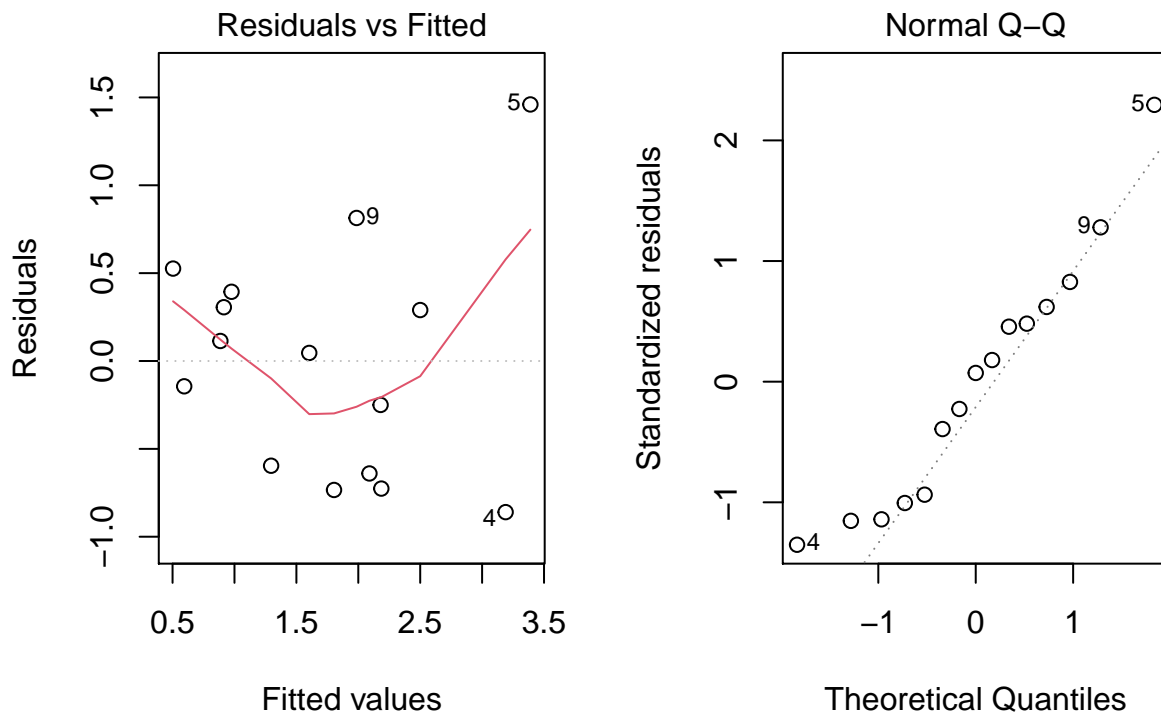
```
# (c)
# Residual analysis
par(mfrow = c(1,2))
plot(lr, which = 1:2)
```



```
# Reset
par(mfrow = c(1,1))
```

(a) The average for the growth in all 15 cups is 1.74 and the standard deviation for the growth in all 15 cups is approximately 1.11.

(b) Performed the ANOVA (see the code above).

(c) We can see that the acids are all independent of each other and thus, the condition for independence holds. From Residuals vs Fitted plot, we see that the red line is does not necessarily align with the black dotted line and is not flat. This means that there is some non-linear trend to the residuals. In the Normal QQ plot, we see that there are some data points that deviates form the dotted line. Hence, not all ANOVA conditions are met.

(d) Based on the ANOVA table, we can conclude that there is a significant difference in average growth due to the treatments. This is the case since the p-value is significant.

(e) Based on the ANOVA table, we cannot conclude that there is a significant difference in average growth due to the distance from the window. This is the case since the p-value is not significant.