Homework 7 - Stats 230 - 2way ANOVA : activity and alfalfa

Dhyey Mavani

date

PROBLEMS TO TURN IN: #6.4abc, #6.36bcde, #6.37, #6.38

The first four problems are conceptual.

Exercise 6.4 abc

6.4 part a:

SOLUTION: Here we can say that the response was the second-graders' score on the set of math problems

6.4 part b:

SOLUTION: The two factors here are noise and ADHD

6.4 part c:

SOLUTION: Here, whether or not the second-grader had ADHD is observational, while Noise is experimental. Also, we can see that both factors have 2 levels.

```
data(Alfalfa)
```

Exercise 6.36 bcde

6.36 part b: Two-way main effects ANOVA table

SOLUTION:

```
aovmodel <- aov(Ht4 ~ Acid + Row, data = Alfalfa)
summary(aovmodel)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Acid          2   6.85    3.43    4.51  0.049 *
## Row           4   4.18    1.05    1.38  0.324
## Residuals     8   6.07    0.76
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6.36 part c: Conditions for ANOVA

SOLUTION:

We have equal variance of errors because we can see the mostly equal spread of points around the x-axis in the residuals vs fitted plot.

Normality of errors is also satisfied because we can see the qq plot with most of the points falling along the line.

We can see that there is a slight heteroskedasticity with the increase in the fitted values, but in my opinion it is not too worrisome.

Finally, by the construction of the experiment, the samples are independent, and we can say that the treatment groups all have the same sample size so there is nothing to worry about there.

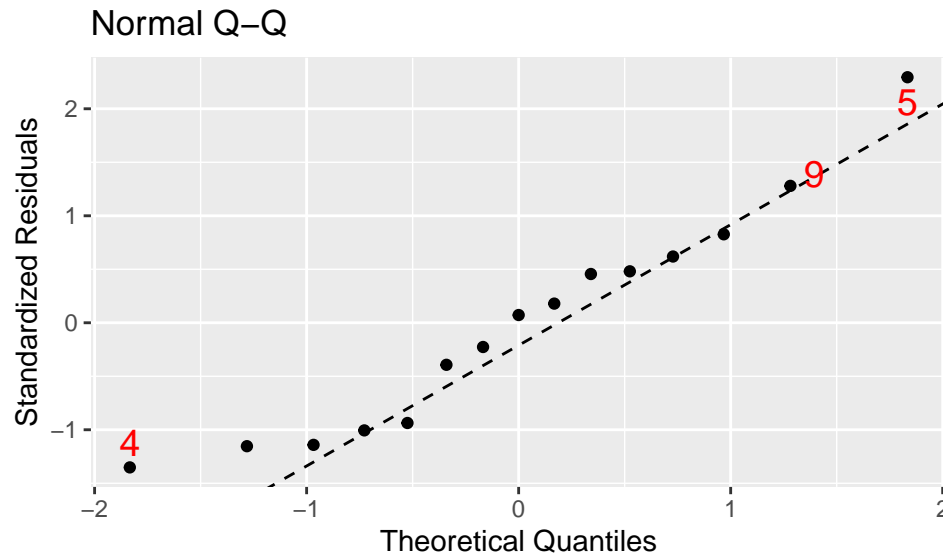Hence, all conditions for using the ANOVA model are satisfied.

```
model1 <- lm(Ht4 ~ Acid + Row, data = Alfalfa)
mplot(model1, which = 1:2)
```

[[1]]

'geom_smooth()' using formula 'y ~ x'

Residuals vs Fitted

##
[[2]]

Normal Q–Q
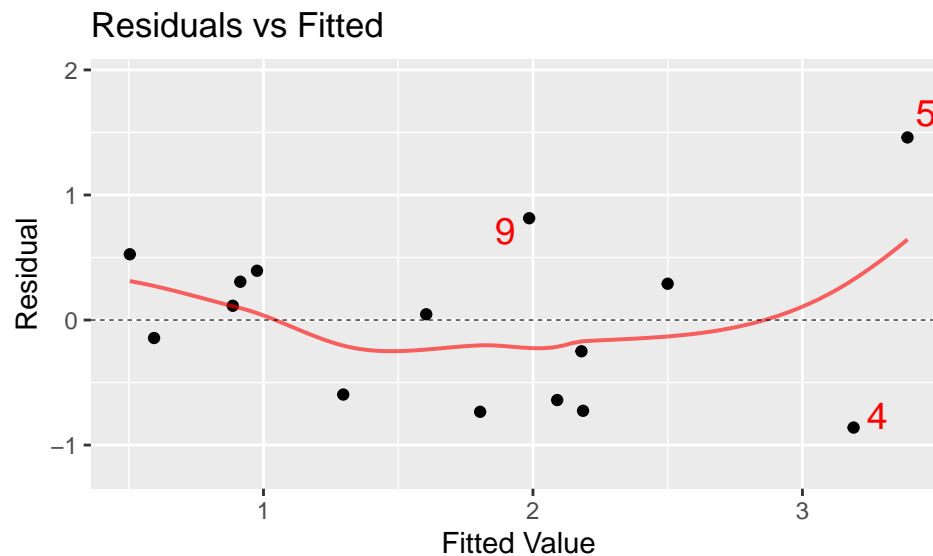
6.36 part d:

SOLUTION: We can say that there is a significant difference in average growth due to the acid treatments, because the p-value of the ANOVA F-test is $0.049 < 0.05$.

This tells me that there is a difference between at least 2 of the 3 treatment types.

6.36 part e:

SOLUTION: We can say that there is no significant difference in average growth due to the distance from the window.

We can conclude this because the p-value of the ANOVA F-test is 0.324, which is significantly greater than the significance level of 0.05.

Exercise 6.37 - continues 6.36 SOLUTION:

Fisher's LSD at6 the 0.05 significance level shows that there is significant difference in mean alfalfa growth after four days between the water and 3.0 HCl treatment levels. We can say this because the 95% confidence interval for that particular comparison does not include 0 inside the range.

Also, I wanted to note that, we observed previously that the factor of Acid is significant

```
PostHocTest(aovmodel, method = "lsd")
```

```
##
##   Posthoc multiple comparisons of means : Fisher LSD
##     95% family-wise confidence level
##
## $Acid
##                 diff     lwr.ci   upr.ci   pval
## 3.0HCl-1.5HCl -0.382 -1.6526522 0.888652 0.5078
## water-1.5HCl   1.204 -0.0666522 2.474652 0.0604 .
## water-3.0HCl   1.586  0.3153478 2.856652 0.0206 *
##
```

```
## $Row
##       diff     lwr.ci upr.ci   pval
## b-a  0.41 -1.230405 2.0504 0.5802
## c-a  0.09 -1.550405 1.7304 0.9024
## d-a  1.10 -0.540405 2.7404 0.1606
## e-a  1.30 -0.340405 2.9404 0.1050
## c-b -0.32 -1.960405 1.3204 0.6648
## d-b  0.69 -0.950405 2.3304 0.3605
## e-b  0.89 -0.750405 2.5304 0.2462
## d-c  1.01 -0.630405 2.6504 0.1934
## e-c  1.21 -0.430405 2.8504 0.1274
## e-d  0.20 -1.440405 1.8404 0.7857
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
data(Popcorn)
glimpse(Popcorn)
```

Exercise 6.38

```
## Rows: 12
## Columns: 3
## $ Unpopped <int> 26, 35, 18, 14, 8, 6, 47, 47, 14, 34, 21, 37
## $ Brand    <fct> Orville, Orville, Orville, Orville, Orville, Orville, Seaway,~
## $ Trial    <int> 1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6
```

```
head(Popcorn)
```

```
##   Unpopped   Brand Trial
## 1       26 Orville     1
## 2       35 Orville     2
## 3       18 Orville     3
## 4       14 Orville     4
## 5        8 Orville     5
## 6        6 Orville     6
```

```
#Strongly recommend you LOOK at the data and make sure R is treating variables the way you expect. #Thi
Popcorn <- mutate(Popcorn, Trial = factor(Trial))
```

6.38 part a:

SOLUTION: The mean number of unpopped kernels for the entire sample is 25.6 kernels. The effect for Orville is -7.75 kernels, and the effect for Seville is 7.75.

```
summary(Popcorn)
```

```
##     Unpopped          Brand    Trial
##  Min.   : 6.0   Orville:6   1:2
##  1st Qu.:14.0   Seaway :6   2:2
##  Median :23.5               3:2
##  Mean   :25.6               4:2
##  3rd Qu.:35.5               5:2
##  Max.   :47.0               6:2
```

```
aovmod2 <- aov(Unpopped ~ Brand + Trial, data = Popcorn)
model.tables(aovmod2)
```

```
## Tables of effects
##
##  Brand
## Brand
## Orville  Seaway
##   -7.75    7.75
##
##  Trial
## Trial
##      1       2       3       4       5       6
##  10.917  15.417  -9.583  -1.583 -11.083  -4.083
```

6.38 part b: Two-way ANOVA model and conditions

SOLUTION: There are rather few data points, and the qq plot shows weak fitting to the line. But it is worthy to note that the points mostly fall in place, so we have normality of errors.

The residuals vs fitted plot demonstrates slight heteroskedasticity as we can see in the plot, but the mostly even spread of points around the x-axis tells us that equal variance of errors is met.

Finally, we know that the samples are independently taken, and each brand has the same number of trials, so all conditions are satisfied and we are good!
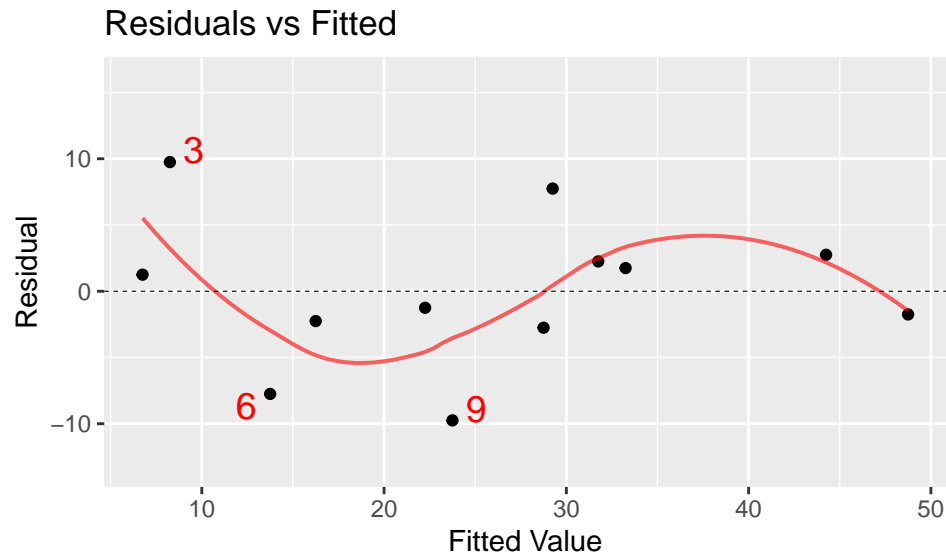
```
summary(aovmod2)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Brand        1    721     721   10.45  0.023 *
## Trial        5   1181     236    3.43  0.101
## Residuals    5    345      69
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
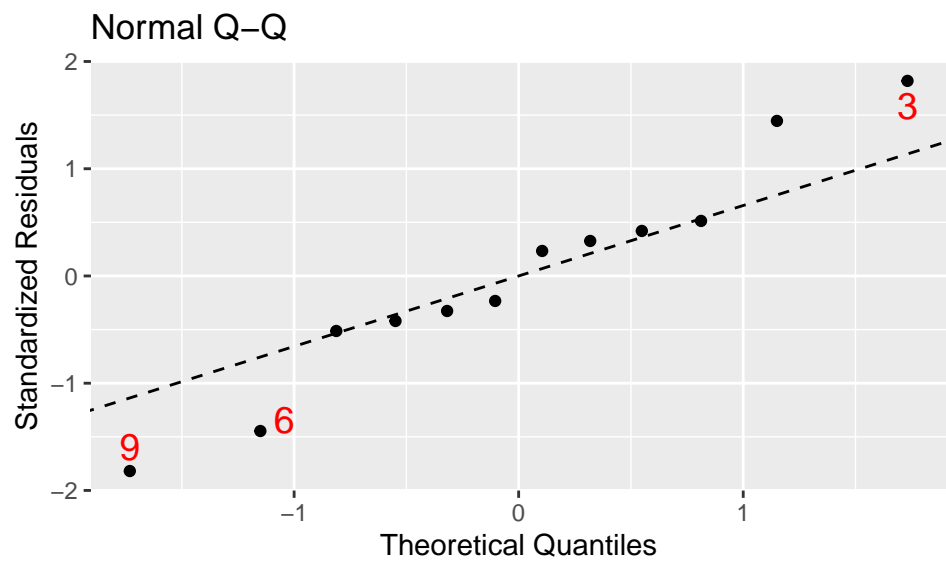
```
mod2 <- lm(Unpopped ~ Brand + Trial, data = Popcorn)
mplot(mod2, which = 1:2)
```

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

```
## `geom_smooth()` using formula 'y ~ x'
```

Residuals vs Fitted


```
##
## [[2]]
```

Normal Q–Q

6.38 part c:

SOLUTION: The brand of popcorn does appear to make a significant difference in the mean number of unpopped kernels.

We can say this because the p-value of the ANOVA F-test is $0.023 < 0.05$.

The trial does not make a significant difference in the mean number of unpopped kernels though as we can see that the p-value of 0.101 is greater than 0.05!