Math 328 Chapter 6 HW Part B

David Oniani

March 07, 2021

## Setup

library(Stat2Data)  
library(dplyr)

##   
## Attaching package: 'dplyr'

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

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

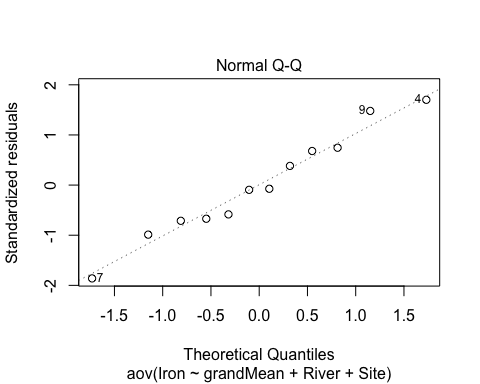
library (tidyr)  
library(emmeans)  
library(ggplot2)  
  
# I like the minimal theme  
theme\_set(theme\_minimal())  
  
# Disable warnings (they clutter the document)  
options(warn = -1)

## Exercise 6.30

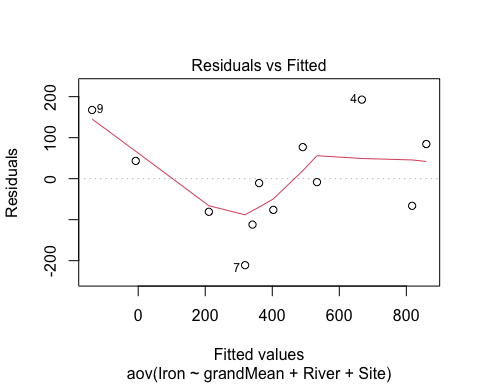
data("RiverIron")  
  
# (a) Fit the two-way additive model  
RiverIron$grandMean = mean(RiverIron$Iron)  
twoWayIron = aov(Iron ~ grandMean + River + Site, RiverIron)  
summary(twoWayIron)

## Df Sum Sq Mean Sq F value Pr(>F)   
## River 3 542458 180819 7.041 0.0216 \*  
## Site 2 442395 221198 8.613 0.0172 \*  
## Residuals 6 154083 25680   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# (b) Obtain a normal probability plot of residuals  
plot(twoWayIron, which = 2)



# (c) Obtain a plot of residuals versus fitted values  
plot(twoWayIron, which = 1)



# (d) Estimate the fitted value from the graph  
xval = twoWayIron$fitted.values[9]  
yval = twoWayIron$residuals[9]  
print(c(xval, yval))

## 9 9   
## -137.5833 167.5833

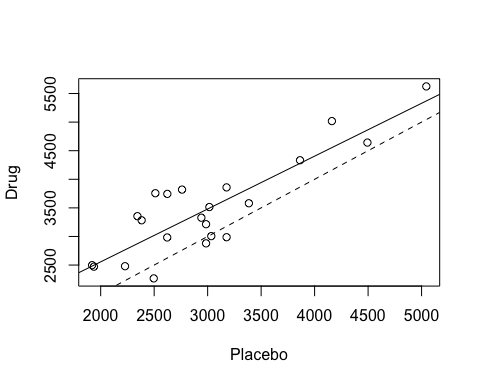
# Sort the values for analyzing differences in fitted values  
print(sort(twoWayIron$fitted.values))

## 9 8 6 7 5 12   
## -137.583333 -7.333333 210.750000 318.916667 341.000000 360.750000   
## 3 11 2 4 10 1   
## 403.083333 491.000000 533.333333 667.250000 817.250000 859.583333

1. After fitting the two-way additive model, we get that the F-score and p-value for River are 7.041 and 0.0172 respectively. F-score and p-value for Site are 8.613 and 0.0172 respectively.
2. Since the points are, for the most part, aligned with the dotted line, it seems like the normality assumption for ANOVA is not violated. That being said, it is true that some points are visibly outside the dotted line, but this is more of a judgment call and for me, it seems to be approximately normal.
3. It seems like the variation is not constant. Ideally, the plot should show a random scattering of points above and below the reference line at a horizontal 0. However, in this case, it does seem that there is some pattern in the data. The curvature is unmistakably to be present. The shape seems to suggest non-linearity (could also be seen as heteroscedastic, but, to me, it seems to be a non-linear trend), potentially quadratic relationship.
4. Eyeballing the leftmost fitted value, I could see the point (-135, 165). After estimating the point using R, we get that the point is (-137.5833, 167.5833). The reason it suggests that this point strongly suggests that the model is inappropriate is that the fitted value is a negative number with a large absolute value (significantly less than any other value). Also, surprisingly, both of these numbers (x and y coordinate of the leftmost point) are very similar. Their decimal expansions do not seem random as all 4 digits match. Interestingly, the similar trend can be observed win the rest of the points.

## Exercise 6.32

data("Contraceptives")  
  
# (a) Create the Anscombe plot for this dataset  
OralContraceptives = pivot\_wider (Contraceptives,  
 id\_cols = "ID",  
 names\_from = "Treatment",  
 values\_from = "EE")  
  
anscombe.plot = function (yvar, xvar) {  
 yname = deparse (substitute (yvar))  
 xname = deparse (substitute (xvar))  
 plot (yvar ~ xvar, ylab=yname, xlab=xname)  
 apfit = lm (yvar ~ xvar)  
 apfit  
 abline (apfit)  
 abline (0, 1, lty=2)  
 return (apfit)  
}  
  
with (OralContraceptives, anscombe.plot (Drug, Placebo))



##   
## Call:  
## lm(formula = yvar ~ xvar)  
##   
## Coefficients:  
## (Intercept) xvar   
## 707.1479 0.9243

1. You can see the generated plot above in the document. The slope is 0.9243.
2. As we got that the slope is “close” to 1 (with the actual value of 0.9243), we can apply the randomized complete block additive model (RCB additive model) and thus, there is no need for a transformation.

## Exercise 6.36

data("Alfalfa")  
  
# (a)  
df = Alfalfa$Ht4  
mean\_df = mean(df)  
sd\_df = sd(df)  
  
print(mean\_df)

## [1] 1.74

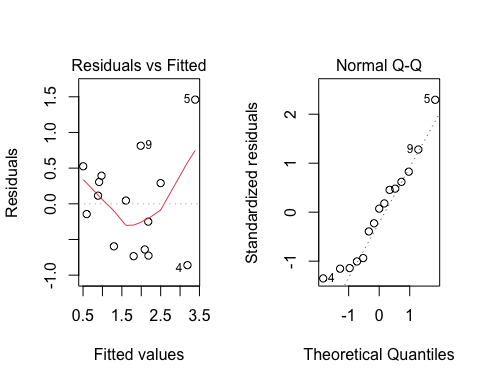
print(sd\_df)

## [1] 1.105396

twoWayAlfalfa = aov(Ht4 ~ Acid + Row, Alfalfa)  
  
# (b)  
# ANOVA  
summary(twoWayAlfalfa)

## 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(twoWayAlfalfa, which = 1:2)



par(mfrow = c(1,1))

1. 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.
2. Performed the ANOVA (see the code above).
3. 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.
4. 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 (0.0487 < 0.05).
5. 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 (0.3235 > 0.05).