Math 328 Chapter 8 HW

David Oniani

March 29, 2021

## Setup

library(emmeans)  
library(multcomp)

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##   
## Attaching package: 'TH.data'

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

library(pwr)

## Exercise 7.1

# k = g = 6  
# f = sqrt(zeta/g\*n) = sqrt(4\*n/g\*n) = sqrt(4/g) = sqrt(4/6) = sqrt(2/3)  
# sig.level = 0.05  
# power = 0.7  
res1 <- pwr.anova.test(k = 6, f = sqrt(2/3), sig.level = 0.05, power = 0.7)  
res1$n

## [1] 3.646913

Since we got that n = 3.646913, we should round it up (as the sample size cannot be a floating point number) and we get n = 4. Hence, the answer is that the smallest sample size is 4.

## Exercise 7.2

# k = g = 3  
# f = sqrt(((32/3 - 10)^2 + 2 \* (32/3 - 11)^2) / 12) = sqrt(2/3 / 12) = sqrt(1/18)  
# sig.level = 0.05  
# power = 0.9  
res2 <- pwr.anova.test(k = 3, f = sqrt(1/18), sig.level = 0.05, power = 0.9)  
res2$n

## [1] 76.93183

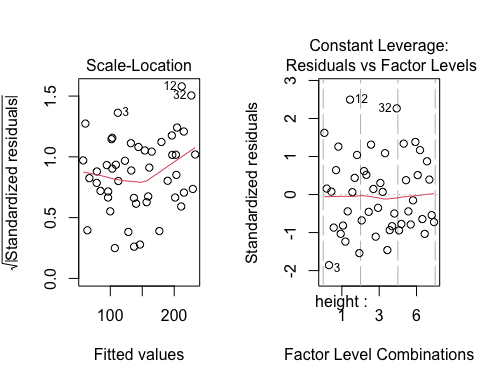
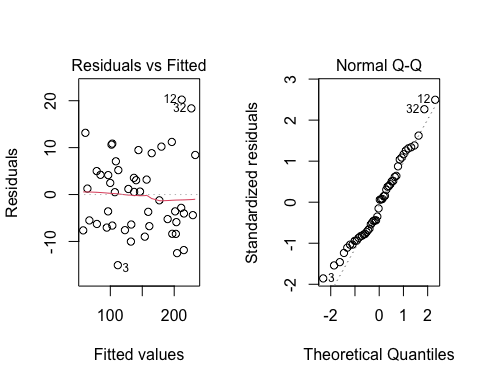
Since we got that n = 76.93, we should round it up (as the sample size cannot be a floating point number) and we get n = 77. Hence, the answer is that the sample size is 77.

## Exercise 8.6

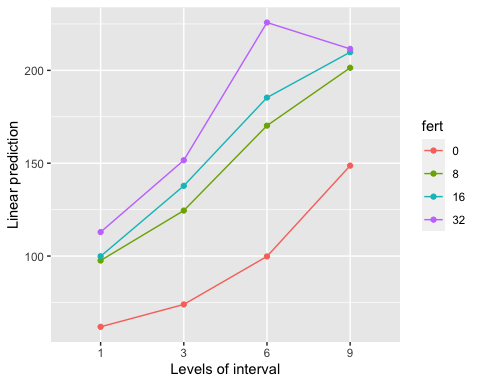
# Load the data  
load("oehlert.rdata")  
  
# Get the right dataset  
data <- pr08.6  
  
# Make height, fert, and interval variables factors  
data[, "height"] <- factor(data[, "height"], labels = c("1", "3", "6"))  
data[, "fert"] <- factor(data[, "fert"], labels = c("0", "8", "16", "32"))  
data[, "interval"] <- factor(data[, "interval"], labels = c("1", "3", "6", "9"))  
  
# Report the structure  
str(data)

## 'data.frame': 48 obs. of 4 variables:  
## $ height : Factor w/ 3 levels "1","3","6": 1 1 1 1 1 1 1 1 1 1 ...  
## $ fert : Factor w/ 4 levels "0","8","16","32": 1 1 1 1 2 2 2 2 3 3 ...  
## $ interval: Factor w/ 4 levels "1","3","6","9": 1 2 3 4 1 2 3 4 1 2 ...  
## $ cwt : num 74.1 65.4 96.7 147.1 87.4 ...

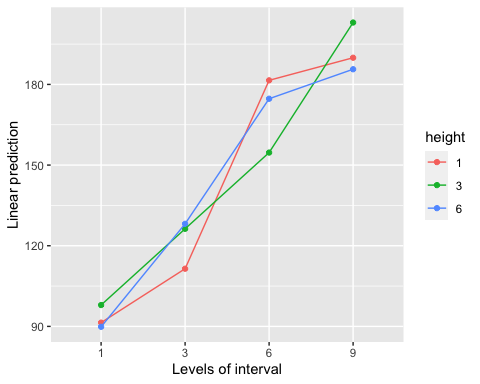
# Fit the initial model  
fit1 <- aov(cwt ~ .^2, data)  
  
# Residual analysis  
par(mfrow = c(1, 2))  
plot(fit1)



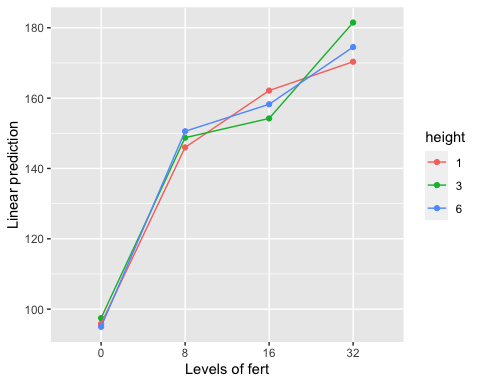
par(mfrow = c(1, 1))  
  
# Interaction plots  
emmip(fit1, fert ~ interval)



emmip(fit1, height ~ interval)



emmip(fit1, height ~ fert)



The Residuals vs Fitted plot shows a constant variance. However, the red line has some problems toward the right tail - it shifts down and does not follow the black dotted line. That being said, overall, the Residual vs Fitted plot looks good and the constant variance assumption does not seem to violated.

The Normal Q-Q plot shows that all, except for a few points in both tails, follow the dotted line and thus, the normality assumption is not violated. That being said, since at left tail, there are some number of points not aligning across the dotted line, a response transformation could potentially help.

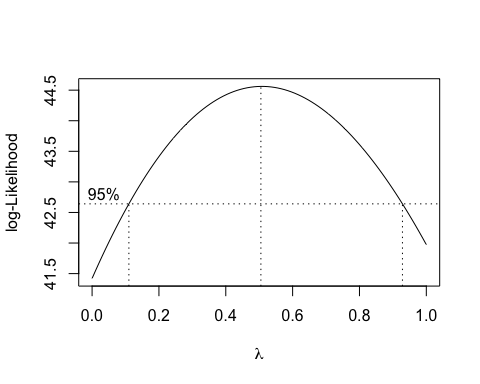
Scale-Location plot does not show a decreasing trend and shows that the variance is constant.

Residuals vs Factor Levels does not show any problems either.

In terms of interactions plots, all three interaction plots show some degree of interaction between the variables. The least degree of interaction seems to be between variables height and fert. However, further analysis will show the relationship between the variables in more detail.

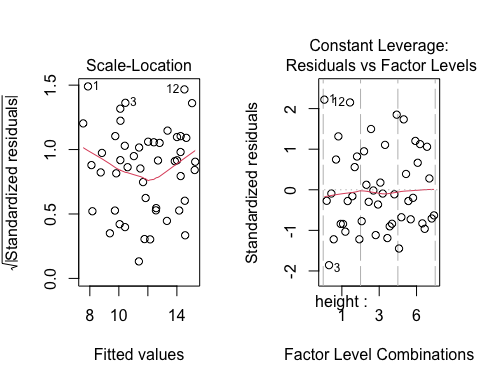
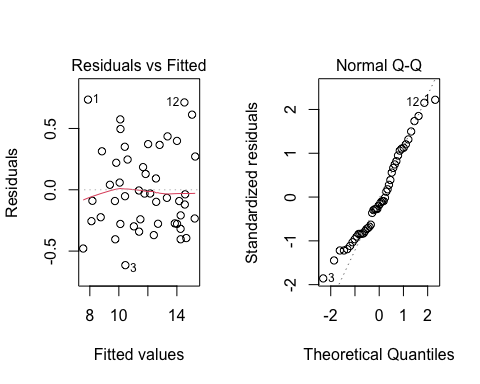
Given our analysis, let us now perform a Box-Cox analysis in order to find an appropriate power transformation for the response variable.

# Box-Cox analysis for the power transformation  
MASS::boxcox(fit1, lambda = seq(0, 1, 0.2))



From the plot, we can see that the lambda value is approximately 0.5 meaning that the response should be raised to 0.5 power and the model should then be refit. This is a square root transformation of the response.

# Refit the model, applying the right power transformation  
fit2 <- aov(sqrt(cwt) ~ .^2, data)  
  
# Residual analysis  
par(mfrow = c(1, 2))  
plot(fit2)



par(mfrow = c(1, 1))

Even after applying the transformation, the Residuals vs Fitted plot looks a bit better and the downward trend at the right tail is not as significant.

The Normal Q-Q plot looks a bit better and the normality assumption is satisfied.

Scale-Location plot has a small improvement and still shows constant variance.

Residuals vs Factor Levels plot still looks good.

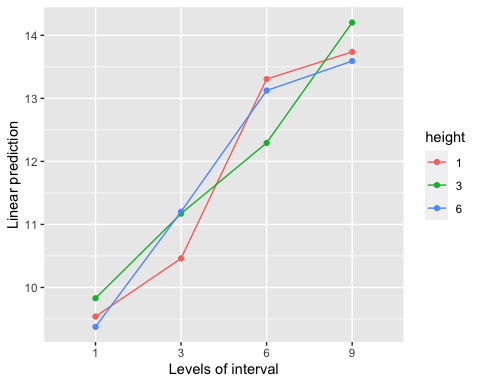
Even though we have performed a response transformation, it did not improve our model significantly as the model itself was fairly good from the very get-go.

Let us see the summary and if any, report significant interactions.

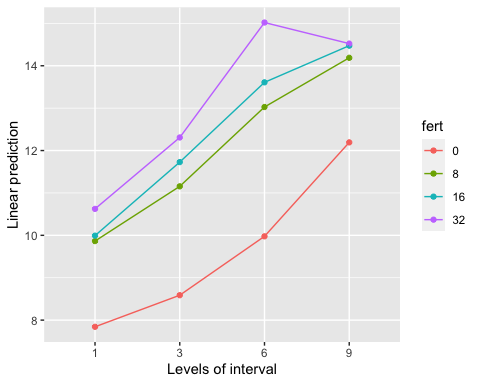
# Report the summary  
summary(fit2)

## Df Sum Sq Mean Sq F value Pr(>F)   
## height 2 0.10 0.05 0.176 0.8398   
## fert 3 82.22 27.41 93.820 3.51e-11 \*\*\*  
## interval 3 132.74 44.25 151.462 5.86e-13 \*\*\*  
## height:fert 6 0.54 0.09 0.306 0.9255   
## height:interval 6 4.87 0.81 2.780 0.0430 \*   
## fert:interval 9 6.87 0.76 2.612 0.0396 \*   
## Residuals 18 5.26 0.29   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Plot ONLY the significant interaction effects  
emmip(fit2, height ~ interval)



emmip(fit2, fert ~ interval)



It seems like variable fert is significant with the p-value of 3.51e-11.

Variable interval is also significant with the p-value of 5.86e-13.

Two interactions: height:interval and fert:interval are significant with the p-values of 0.043 and 0.040 respectively.

The plots of the significant interactions are also shown.

cld(emmeans(fit2, ~ height | interval), letters=LETTERS)

## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates

## interval = 1:  
## height emmean SE df lower.CL upper.CL .group  
## 6 9.38 0.27 18 8.81 9.94 1   
## 1 9.54 0.27 18 8.97 10.11 1   
## 3 9.83 0.27 18 9.26 10.40 1   
##   
## interval = 3:  
## height emmean SE df lower.CL upper.CL .group  
## 1 10.46 0.27 18 9.89 11.03 1   
## 3 11.17 0.27 18 10.60 11.74 1   
## 6 11.20 0.27 18 10.63 11.77 1   
##   
## interval = 6:  
## height emmean SE df lower.CL upper.CL .group  
## 3 12.29 0.27 18 11.73 12.86 1   
## 6 13.12 0.27 18 12.56 13.69 12   
## 1 13.31 0.27 18 12.74 13.87 2   
##   
## interval = 9:  
## height emmean SE df lower.CL upper.CL .group  
## 6 13.59 0.27 18 13.02 14.16 1   
## 1 13.74 0.27 18 13.17 14.31 1   
## 3 14.20 0.27 18 13.64 14.77 1   
##   
## Results are averaged over the levels of: fert   
## Results are given on the sqrt (not the response) scale.   
## Confidence level used: 0.95   
## Note: contrasts are still on the sqrt scale   
## P value adjustment: tukey method for comparing a family of 3 estimates   
## significance level used: alpha = 0.05

cld(emmeans(fit2, ~ fert | interval), letters=LETTERS)

## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates

## interval = 1:  
## fert emmean SE df lower.CL upper.CL .group  
## 0 7.84 0.312 18 7.19 8.50 1   
## 8 9.86 0.312 18 9.21 10.52 2   
## 16 9.99 0.312 18 9.34 10.65 2   
## 32 10.62 0.312 18 9.97 11.28 2   
##   
## interval = 3:  
## fert emmean SE df lower.CL upper.CL .group  
## 0 8.59 0.312 18 7.93 9.24 1   
## 8 11.16 0.312 18 10.50 11.81 2   
## 16 11.73 0.312 18 11.07 12.38 2   
## 32 12.31 0.312 18 11.65 12.96 2   
##   
## interval = 6:  
## fert emmean SE df lower.CL upper.CL .group  
## 0 9.98 0.312 18 9.32 10.63 1   
## 8 13.03 0.312 18 12.37 13.68 2   
## 16 13.61 0.312 18 12.95 14.26 2   
## 32 15.02 0.312 18 14.36 15.68 3   
##   
## interval = 9:  
## fert emmean SE df lower.CL upper.CL .group  
## 0 12.19 0.312 18 11.54 12.85 1   
## 8 14.19 0.312 18 13.53 14.84 2   
## 16 14.47 0.312 18 13.82 15.13 2   
## 32 14.52 0.312 18 13.87 15.18 2   
##   
## Results are averaged over the levels of: height   
## Results are given on the sqrt (not the response) scale.   
## Confidence level used: 0.95   
## Note: contrasts are still on the sqrt scale   
## P value adjustment: tukey method for comparing a family of 4 estimates   
## significance level used: alpha = 0.05

Above find the appropriate pairwise comparisons for the significant effects.

Let us now analyze emmeans with height and interval interaction.

In the first interval, means are approximately equal to each other.

In the third interval, height = 3 and height = 6 have the highest means that are also approximately equal to each other. height = 1 has the mean of 10.46, which is noticeably smaller than 11.17 or 11.20.

In the sixth interval, height = 1 (13.31) and height = 6 (13.12) have higher mean values than height = 3 (12.29).

In the ninth interval, height = 3 has the highest mean value of 14.20. Means at height = 1 and height = 6 and close and equal to 13.74 and 13.59 respectively.

Let us now analyze emmeans with fert and interval interaction.

Interestingly, in all intervals, we see clearly that the higher fert values have higher corresponding emmean values.