

Math 328 Chapter 7 HW

David Oniani

March 23, 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 7.14

It is clear that $MSAB < MSB < MSA$ (A has 10s squared, B has 5s squared, and AB has 4s squared) and each of the has 1 degree of freedom. Hence, $MSAB < MSB < MSA$ and we now need to find where MSE fits. SSE here equals to sum of square residuals over the degrees of freedom which is $60 / 4 = 15$. Now, since $MSAB = 8 * 4^2 = 128$, we have that $MSE < MSAB$. Finally, we have: $MSE < MSAB < MSB < MSA$.

Exercise 7.16

It is clear that $MSAB < MSA < MSB$ (B has 3s squared, A has 2s squared, and AB has 1s squared) and each of the has 1 degree of freedom. Hence, $MSAB < MSA < MSB$ and we now need to find where MSE fits. MSE here equals to sum of square residuals over the degrees of freedom which is $60 / 4 = 15$. Now, since $MSAB = 8 * 1^2 = 8$ and $MSA = 8 * 2^2 = 32$, we have that $MSAB < MSE < MSA$. Finally, we have: $MSAB < MSE < MSA < MSB$.

Exercise 7.34

```
# Dr. Phil's suggestion:
#   "For 7.34(c), I suggest plotting the data. Fitting a model using lm and
#   examining the parameter estimates could also be informative."

# Load the data
data("Swahili")

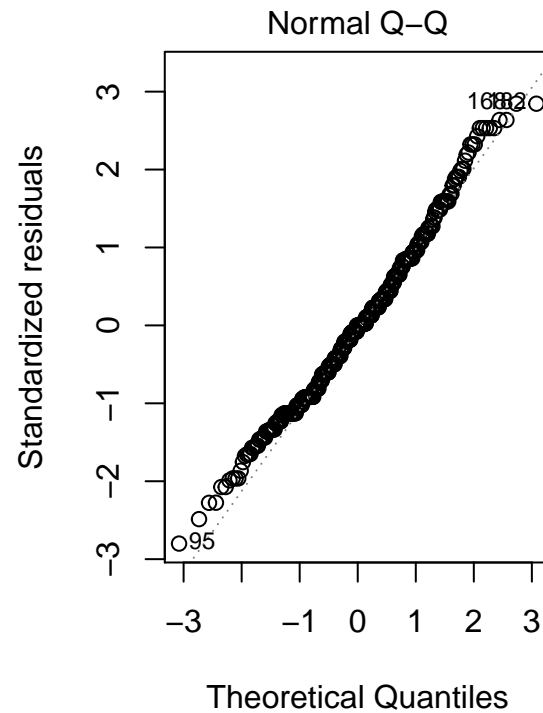
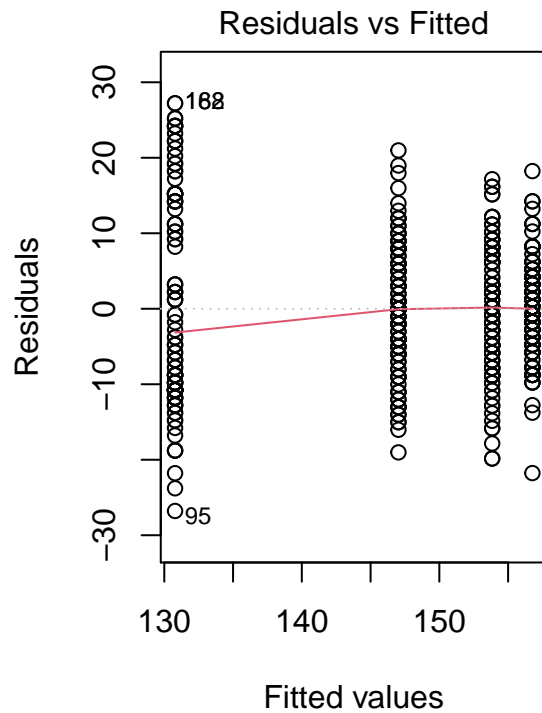
# (a)
modela = lm(Attitude.Score ~ factor(Province) * factor(Sex), data = Swahili)
summary(modela)

##
## Call:
## lm(formula = Attitude.Score ~ factor(Province) * factor(Sex),
##     data = Swahili)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -26.783  -7.210  -0.025   6.158  27.217
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   147.0250     0.8768  167.676 < 2e-16 ***
## factor(Province)PWANI           9.7417     1.2400   7.856 2.66e-14 ***
## factor(Sex)male                -16.2417     1.2400 -13.098 < 2e-16 ***
## factor(Province)PWANI:factor(Sex)male  13.3167     1.7537   7.594 1.66e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.605 on 476 degrees of freedom
## Multiple R-squared:  0.5254, Adjusted R-squared:  0.5224
## F-statistic: 175.6 on 3 and 476 DF, p-value: < 2.2e-16

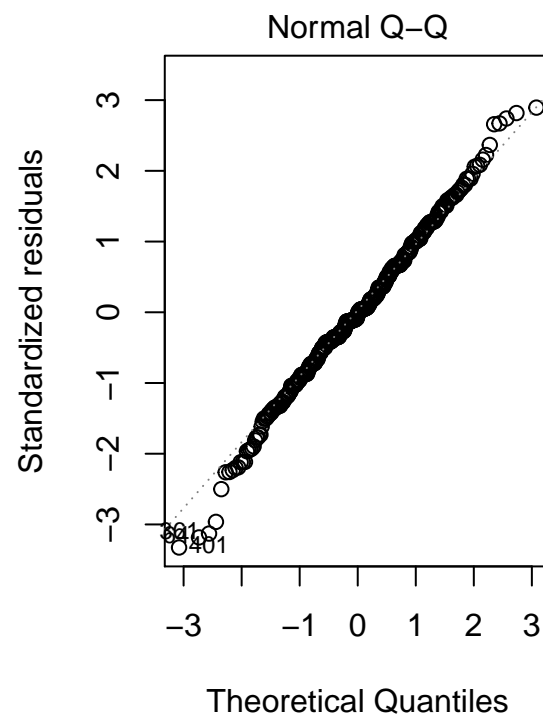
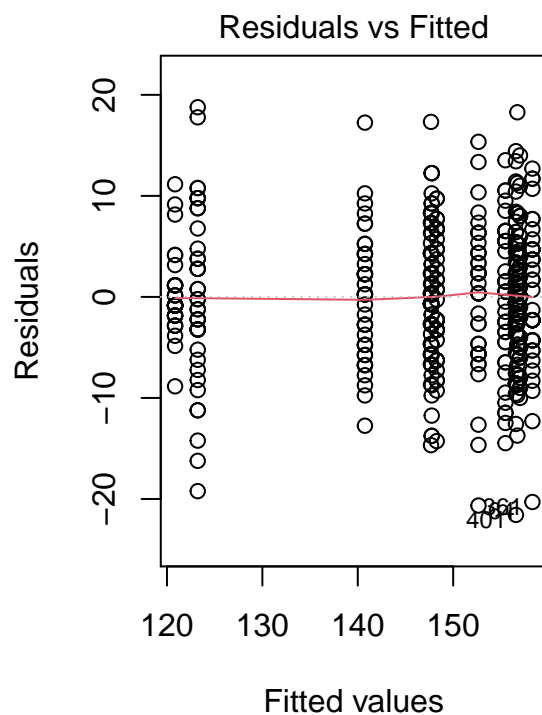
modela$coefficients

##                                (Intercept)                factor(Province)PWANI
##                                147.025000                      9.741667
##                                factor(Sex)male factor(Province)PWANI:factor(Sex)male
##                                -16.241667                      13.316667

# (b)
par(mfrow = c(1 : 2))
plot(modela, which = c(1 : 2))
```



```
# (c)
modelc = lm(Attitude.Score ~ factor(School) * factor(Province), data = Swahili)
plot(modelc, which = c(1 : 2))
```



```
summary(modelc)

##
## Call:
## lm(formula = Attitude.Score ~ factor(School) * factor(Province),
```

```
##      data = Swahili)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.575  -3.888  -0.225   4.275  18.775
##
## Coefficients: (12 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    120.850      1.039  116.341 <2e-16 ***
## factor(School)B    31.800      1.469   21.647 <2e-16 ***
## factor(School)C     2.375      1.469    1.617  0.107
## factor(School)D    19.900      1.469   13.546 <2e-16 ***
## factor(School)E    27.425      1.469   18.669 <2e-16 ***
## factor(School)F    26.825      1.469   18.260 <2e-16 ***
## factor(School)G    35.875      1.469   24.421 <2e-16 ***
## factor(School)H    34.625      1.469   23.570 <2e-16 ***
## factor(School)I    36.150      1.469   24.608 <2e-16 ***
## factor(School)J    37.450      1.469   25.493 <2e-16 ***
## factor(School)K    35.725      1.469   24.319 <2e-16 ***
## factor(School)L    26.900      1.469   18.312 <2e-16 ***
## factor(Province)PWANI      NA         NA      NA      NA
## factor(School)B:factor(Province)PWANI      NA         NA      NA      NA
## factor(School)C:factor(Province)PWANI      NA         NA      NA      NA
## factor(School)D:factor(Province)PWANI      NA         NA      NA      NA
## factor(School)E:factor(Province)PWANI      NA         NA      NA      NA
## factor(School)F:factor(Province)PWANI      NA         NA      NA      NA
## factor(School)G:factor(Province)PWANI      NA         NA      NA      NA
## factor(School)H:factor(Province)PWANI      NA         NA      NA      NA
## factor(School)I:factor(Province)PWANI      NA         NA      NA      NA
## factor(School)J:factor(Province)PWANI      NA         NA      NA      NA
## factor(School)K:factor(Province)PWANI      NA         NA      NA      NA
## factor(School)L:factor(Province)PWANI      NA         NA      NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.57 on 468 degrees of freedom
## Multiple R-squared:  0.7817, Adjusted R-squared:  0.7766
## F-statistic: 152.4 on 11 and 468 DF,  p-value: < 2.2e-16
```

- (a) Factors Province and Sex and both significant with the p-value being less than $2e-16$. This means that attitudes toward Swahili vary significantly based on Province and Sex. The coefficient for the factor Province is 9.742 which is a positive direction and has the standard error of 1.240. The coefficient for the factor Sex is -16.242, which is a negative direction and has the standard error value of 1.240. The interaction between Province and Sex is also significant with the p-value of approximately $1.66e-13$. This means that attitudes toward Swahili vary based on the interaction/relationship between Province and Sex (e.g., certain provinces have more people of a specific Sex). The interaction estimate is 13.317 and has a positive direction (Province:Sex). The standard error is 1.754. This means that further analysis is needed for males in PWANI Province in order to understand the interaction effect. In terms of explaining the main effects, they do not have a huge significance since we have an interaction. We need to first deal with the interaction, but the book does not suggest doing anything. But here is some interpretation of the results.
- (b) The normality condition seems to be met. There are some points at the tail of the plot that deviate from the dotted line, but the Normal QQ does look normal overall. 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. This is not the case in Residuals vs Fitted plot where points are stacked across vertical lines. Hence, we conclude that the normality condition is met, while equal variance condition is not met.

- (c) By just looking at Residuals vs Fitted plot, we can see that the equal variance condition is not met. Therefore, applying a two-way ANOVA is not appropriate in the first place. It is not a complete factorial design since equal sample size from each school is not enough. For the study to be a balanced complete factorial design, equal sample sizes must be taken for each of the possible factor combinations.

Exercise 7.38

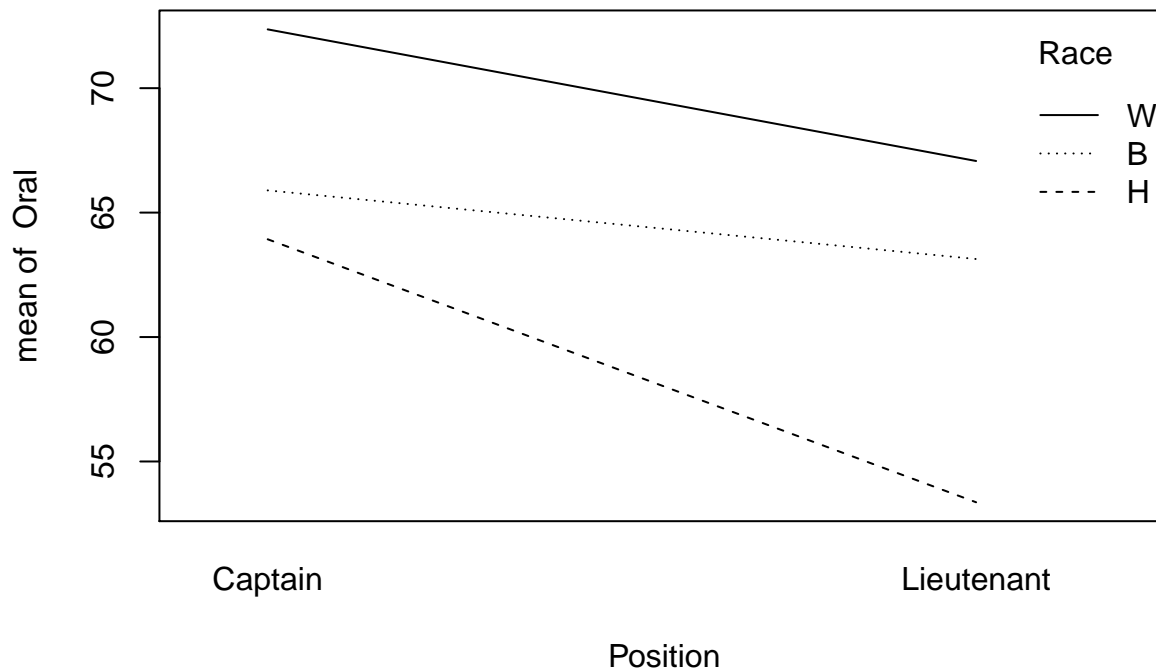
```
# Dr. Phil's Comment:
#   "For 7.38, use the Oral exam score as the response variable
#   (not Written or Combine)."
```

```
# Load the data
data(Ricci)
```

```
# (a)
lra = lm(Oral ~ Position, data = Ricci)
sa = summary(emmeans(lra, pairwise ~ Position), infer = c(T, T))
sa$contrasts
```

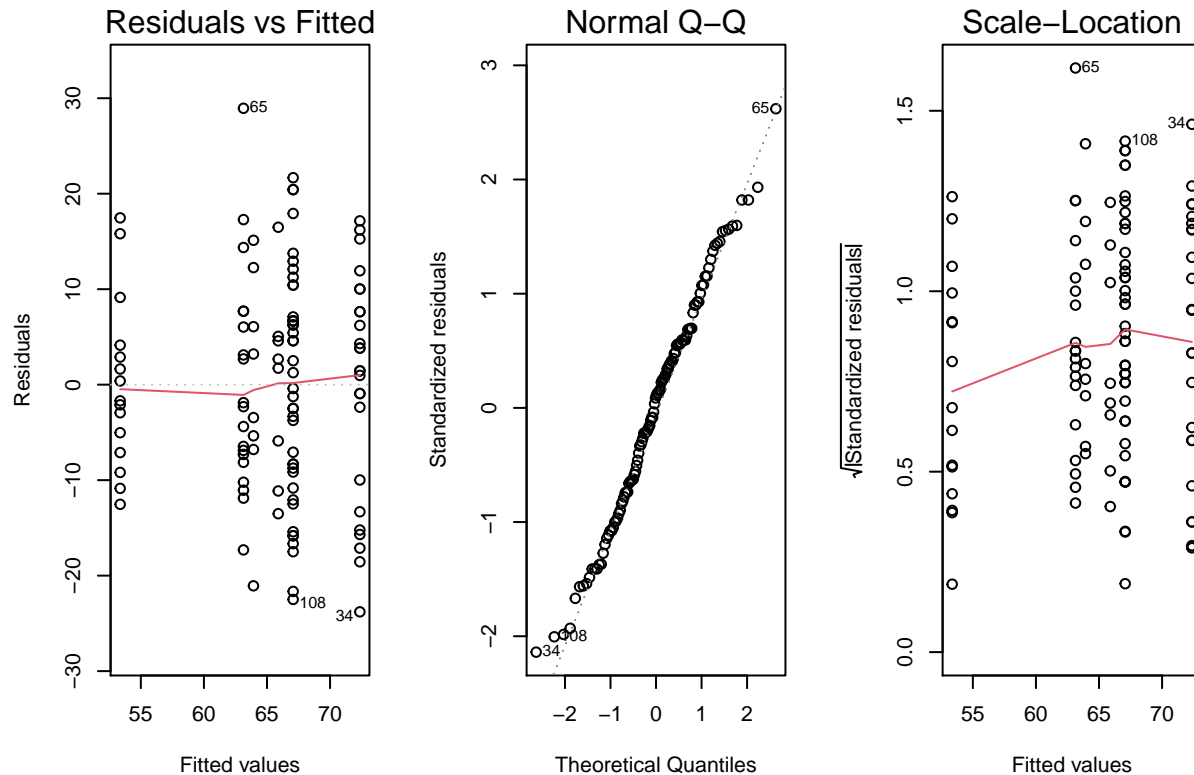
```
## contrast          estimate    SE  df lower.CL upper.CL t.ratio p.value
## Captain - Lieutenant      6.02 2.35 116     1.38    10.7 2.568  0.0115
##
## Confidence level used: 0.95
```

```
# (b)
par(mfrow = c(1, 1))
with(Ricci, interaction.plot(Position, Race, Oral))
```



```
# (c)
lrc = lm(Oral ~ factor(Position) * factor(Race), data = Ricci)
```

```
par(mfrow = c(1, 3))
plot(lrc, which = c(1 : 3))
```



```
summary(aov(Oral ~ factor(Position) * factor(Race), data = Ricci))
```

```
##               Df Sum Sq Mean Sq F value    Pr(>F)
## factor(Position)      1    971   970.8    7.534 0.007056 **
## factor(Race)           2   2475  1237.7    9.605 0.000141 ***
## factor(Position):factor(Race) 2    175    87.4    0.678 0.509519
## Residuals           112  14433   128.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The difference between the means of Captain and Lieutenant is 6.02 with the p-value of approximately 0.012. Hence, the difference is not significant.
- Analyzing the interaction plot, the more nonparallel the lines are, the greater the strength of the interaction. In our case, there seems to be some interaction between Position and Race as the lines are not parallel. If we extend these lines in both sides (left and right), every pair of lines will cross each other. Hence, there could be a potential interaction between Position and Race variables.
- By just looking Residuals vs Fitted, we can conclude that the equal variance condition is not met. The red line has a noticeable skew in the left tail and deviates from the dotted line. The normality condition also does not seem to be met since the left part of the Normal QQ seems to show a significant deviation from the dotted line. The Scale-Location also seems to show a decreasing trend. Hence, we conclude that fitting a two-way ANOVA model is not appropriate (it might be appropriate after a response transformation however, but the book does not ask for this).
- We concluded that the two-way ANOVA model is not appropriate, but we will still comment on the results. Position and Race are both significant with p-values of approximately 0.0071 and 0.0001 respectively. Interaction between Position and Race, however, is not significant with the p-value of

approximately 0.5095.