

STOR 590 HW8 Solution

Taebin Kim

4/6/2020

Page 234 Exercise 5

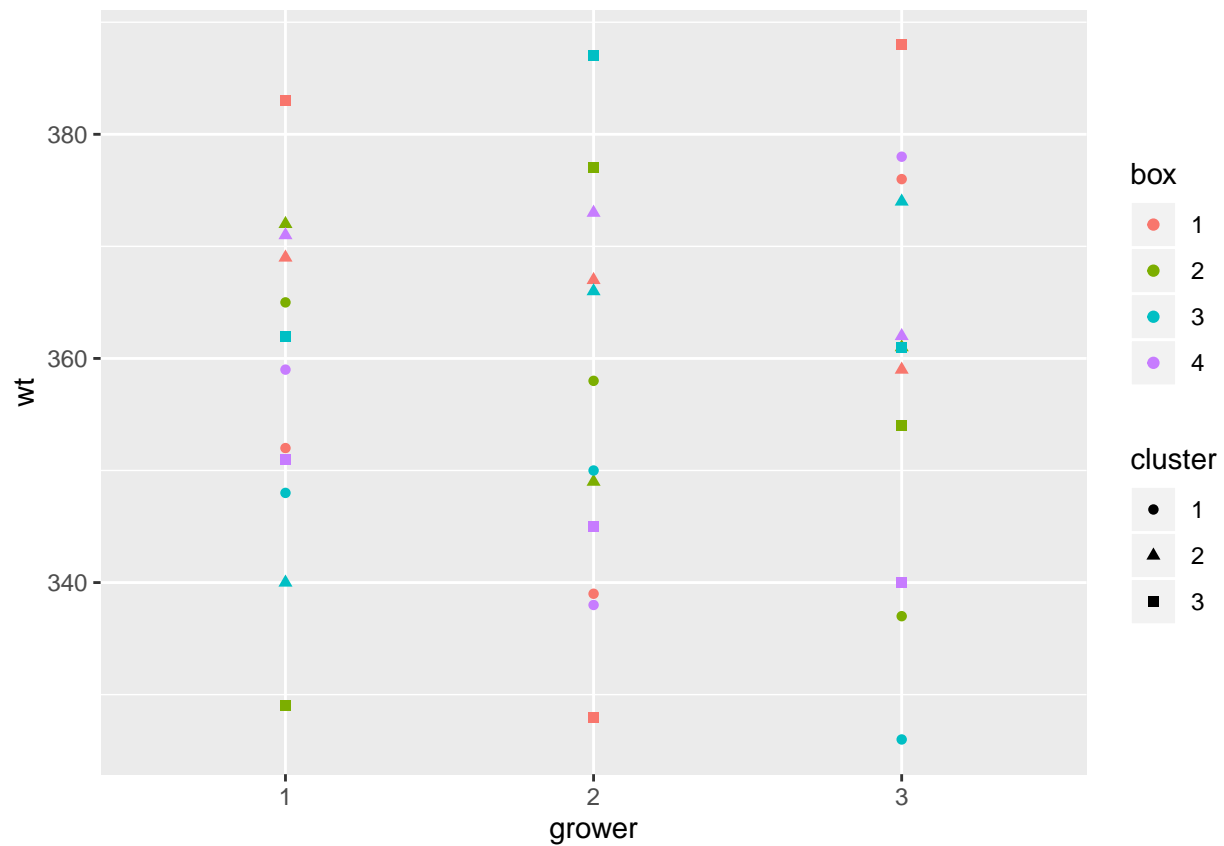
Part (a)

We plot the broccoli data.

```
library(faraway)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
data(broccoli)
ggplot(broccoli, aes(grower, wt, color = box, shape = cluster)) +
  geom_point()
```



The plot shows that the variation is mainly due to random effects rather than fixed effects.

Part (b)

We compute the mean weights within growers and within boxes.

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.2

##
## 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

broccoli %>%
  group_by(grower) %>%
  summarize(mean_wt = mean(wt))

## # A tibble: 3 x 2
##   grower mean_wt
##   <fct>   <dbl>
## 1 1      358.
## 2 2      356.
## 3 3      360.
```

```
broccoli %>%
  group_by(box) %>%
  summarize(mean_wt = mean(wt))
```

```
## # A tibble: 4 x 2
##   box mean_wt
##   <fct> <dbl>
## 1 1      362.
## 2 2      356.
## 3 3      357.
## 4 4      357.
```

Part (c)

We fit an appropriate mixed effects model.

```
library(lme4)

## Loading required package: Matrix
```

```
cmod <- lmer(wt ~ 1 + (1|grower) + (1|box), broccoli)
```

```
## boundary (singular) fit: see ?isSingular
```

```
summary(cmod)
```

```
## Fixed Effects:
## coef.est  coef.se
##   358.17    2.79
##
## Random Effects:
## Groups   Name      Std.Dev.
## box      (Intercept)  0.00
## grower   (Intercept)  0.00
## Residual                16.73
## ---
## number of obs: 36, groups: box, 4; grower, 3
## AIC = 308.1, DIC = 307.9
## deviance = 304.0
```

We treat all effects except the grand mean as random effects. The variation can be assigned to the variance between growers, the variance between boxes, and the variance between clusters.

Part (d)

We test if there is no variation attributable to growers.

```
library(RLRsim)
```

```
## Warning: package 'RLRsim' was built under R version 3.6.3
```

```
cmodr <- lmer(wt ~ 1 + (1|box), broccoli)
```

```
## boundary (singular) fit: see ?isSingular
```

```
cmods <- lmer(wt ~ 1 + (1|grower), broccoli)
```

```
## boundary (singular) fit: see ?isSingular
```

```
exactRLRT(cmods, cmod, cmodr)
```

```
## Observed RLRT statistic is 0, no simulation performed.
```

```
##
## simulated finite sample distribution of RLRT.
##
## (p-value based on 0 simulated values)
##
## data:
## RLRT = 0, p-value = 1
```

Since the p-value is large, we conclude that there may be no variation attributable to growers.

Part (e)

We test if there is no variation attributable to boxes.

```
exactRLRT(cmodr, cmod, cmods)
```

```
## Observed RLRT statistic is 0, no simulation performed.
```

```
##  
## simulated finite sample distribution of RLRT.  
##  
## (p-value based on 0 simulated values)  
##  
## data:  
## RLRT = 0, p-value = 1
```

Similarly, we conclude that there may be no variation attributable to boxes.

Part (f)

We compute confidence intervals for the SD components in the full model.

```
confint(cmod, method = "boot")
```

```
## Computing bootstrap confidence intervals ...
```

```
##  
## 387 message(s): boundary (singular) fit: see ?isSingular  
## 11 warning(s): Model failed to converge with max|grad| = 0.00219475 (tol = 0.002, component 1) (and ...)
```

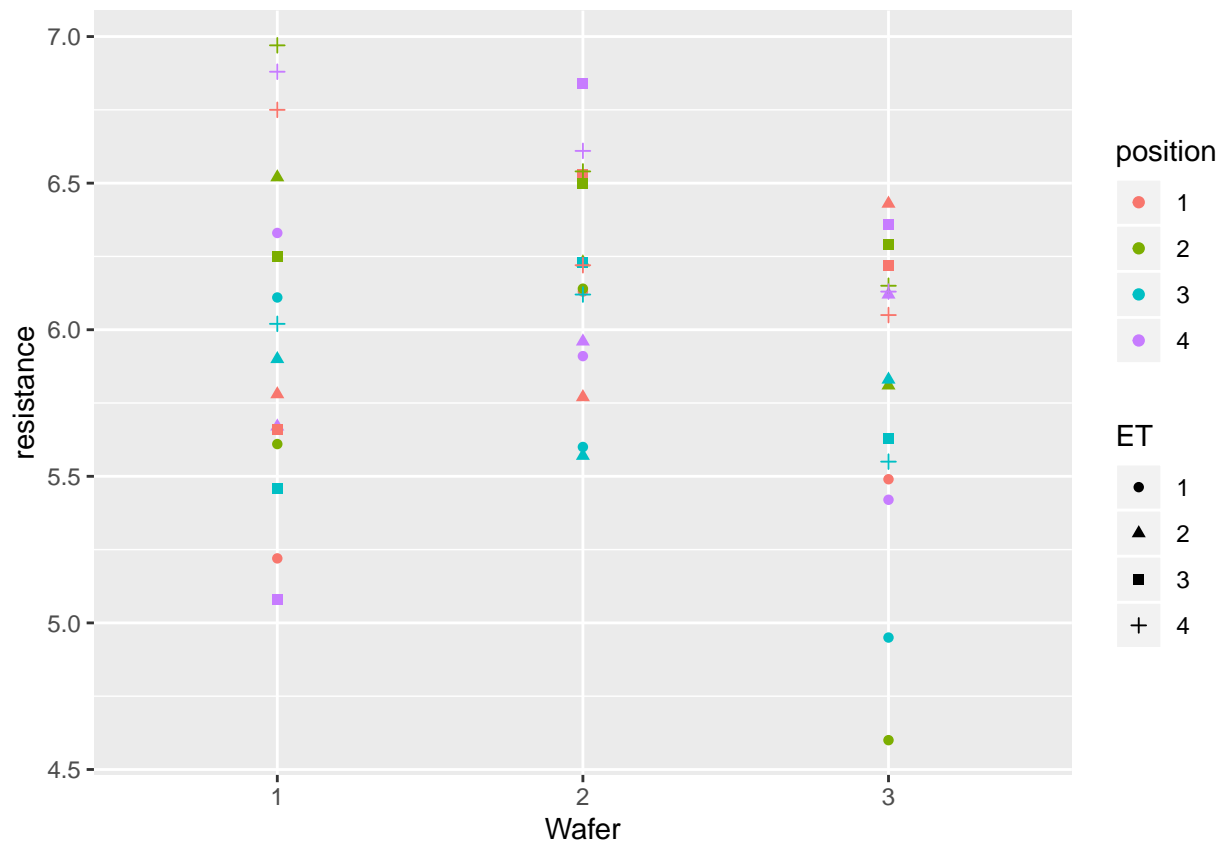
```
##           2.5 %      97.5 %  
## .sig01      0.00000  8.353004  
## .sig02      0.00000  8.383738  
## .sigma     12.61298 20.316450  
## (Intercept) 351.91770 364.456145
```

Page 235 Exercise 7

Part (a)

We plot the semicond data.

```
data(semicond)  
ggplot(semicond, aes(Wafer, resistance, shape = ET, color = position)) + geom_point()
```



We can observe that there is a difference in resistance for different ET and position.

Part (b)

We fit a fixed effects model with an interaction between ET and position.

```
lm1 <- aov(resistance ~ ET * position, semicond)
summary(lm1)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## ET           3   3.112   1.0374    4.782 0.00728 **
## position     3   1.129   0.3763    1.735 0.17965
## ET:position   9   0.809   0.0899    0.415 0.91775
## Residuals    32   6.942   0.2169
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-values show that only ET is significant. The problem in this model is that it does not take account of the random effects from other predictors.

Part (c)

We fit a model appropriate to the split plot design used here.

```
lmer1 <- lmer(resistance ~ ET * position + (1|Wafer) + (1|Grp), semicond)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =  
## control$checkConv, : Model failed to converge with max|grad| = 0.00340168  
## (tol = 0.002, component 1)
```

```
lmer2 <- lmer(resistance ~ ET * position + (1|Wafer), semicond)  
lmer3 <- lmer(resistance ~ ET * position + (1|Grp), semicond)  
exactRLRT(lmer2, lmer1, lmer3)
```

```
##  
## simulated finite sample distribution of RLRT.  
##  
## (p-value based on 10000 simulated values)  
##  
## data:  
## RLRT = 0.0001164, p-value = 0.3797
```

```
exactRLRT(lmer3, lmer1, lmer2)
```

```
##  
## simulated finite sample distribution of RLRT.  
##  
## (p-value based on 10000 simulated values)  
##  
## data:  
## RLRT = 7.5286, p-value = 0.0028
```

We can observe that the variation between the groups is not significant whereas the variation within the groups is significant.

Part (d)

We test for the effect of position.

```
library(pbkrtest)
```

```
## Warning: package 'pbkrtest' was built under R version 3.6.3
```

```
lmer4 <- lmer(resistance ~ ET + (1|Grp), semicond)  
KRmodcomp(lmer3, lmer4)
```

```
## F-test with Kenward-Roger approximation; time: 0.36 sec  
## large : resistance ~ ET * position + (1 | Grp)  
## small : resistance ~ ET + (1 | Grp)  
##      stat      ndf      ddf F.scaling p.value  
## Ftest  1.4533 12.0000 24.0000      1    0.21
```

The p-value indicates that the effect of position is not significant.

Part (e)

We find the level of ET that results in the highest resistance.

```
summary(lmer4)
```

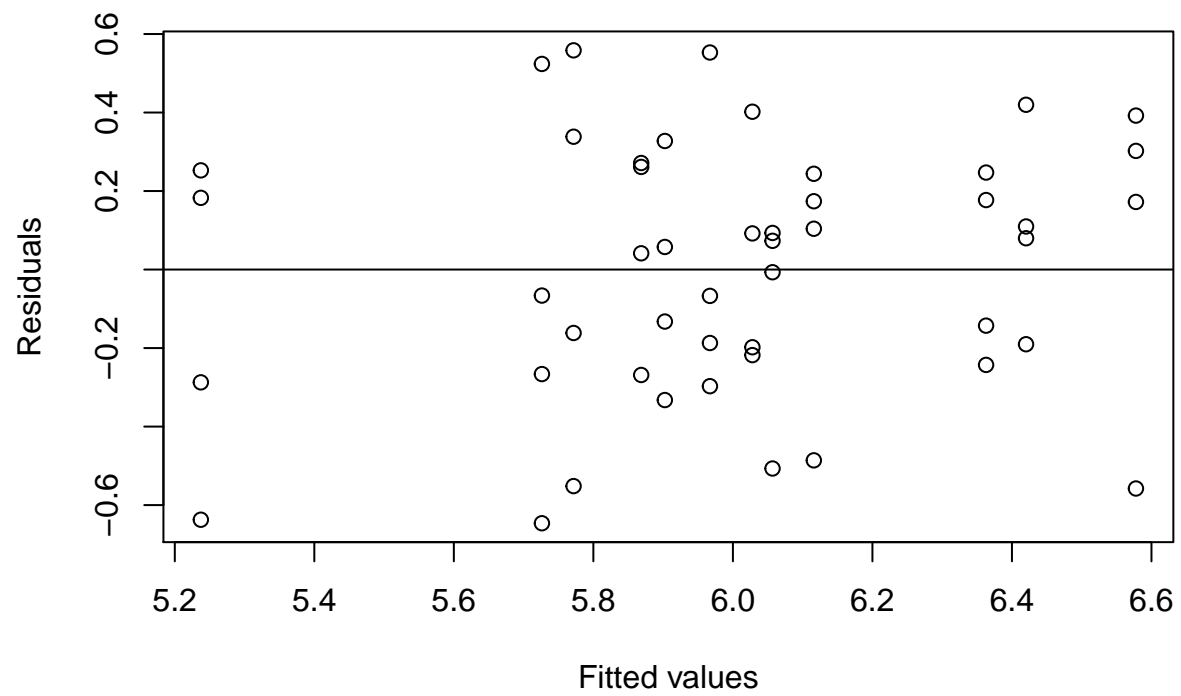
```
## Fixed Effects:
##               coef.est coef.se
## (Intercept)  5.63      0.21
## ET2          0.34      0.30
## ET3          0.46      0.30
## ET4          0.71      0.30
##
## Random Effects:
## Groups   Name      Std.Dev.
## Grp      (Intercept) 0.32
## Residual                0.36
## ---
## number of obs: 48, groups: Grp, 12
## AIC = 67.8, DIC = 43.8
## deviance = 49.8
```

`coef.est` shows that ET4 results in the highest resistance. However, we cannot be sure about this result since the difference of the coefficients between ET3 and ET4 is smaller than the standard error.

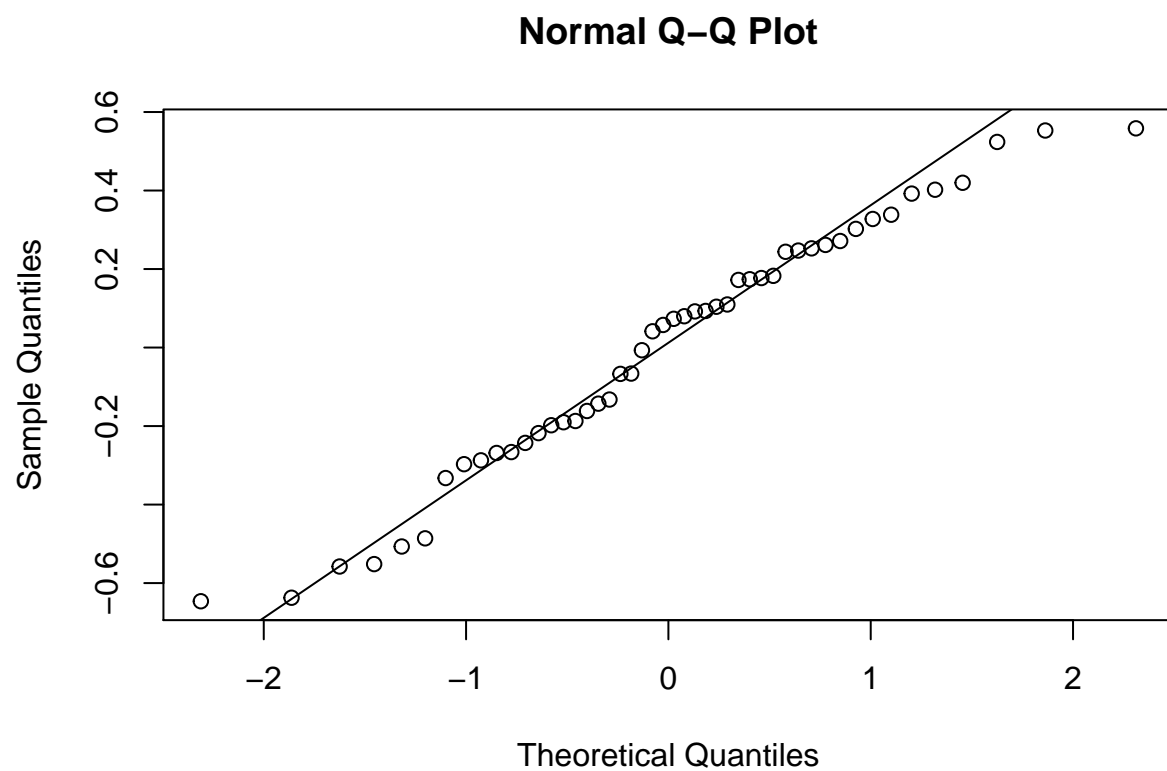
Part (f)

We make a plot of the residuals and fitted values. We also make a QQ plot.

```
plot(residuals(lmer4) ~ fitted(lmer4), xlab = "Fitted values", ylab = "Residuals")
abline(h=0)
```



```
qqnorm(residuals(lmer4))  
qqline(residuals(lmer4))
```

The plots show that our model is a good fit for the data.