HW11 Questions 1. ELMR Chapter 10, question 6 2. ELMR Chapter 10, question 8 3. ELMR Chapter 11, question 1 4. ELMR Chapter 11, question 5

6. An experiment was conducted to select the supplier of raw materials for production of a component. The breaking strength of the component was the objective of interest. Four suppliers were considered. The four operators can only produce one component each per day. A latin square design is used and the data is presented in breaking.

```
library(faraway)
br <- data.frame(breaking, package="faraway")
head(br)
```

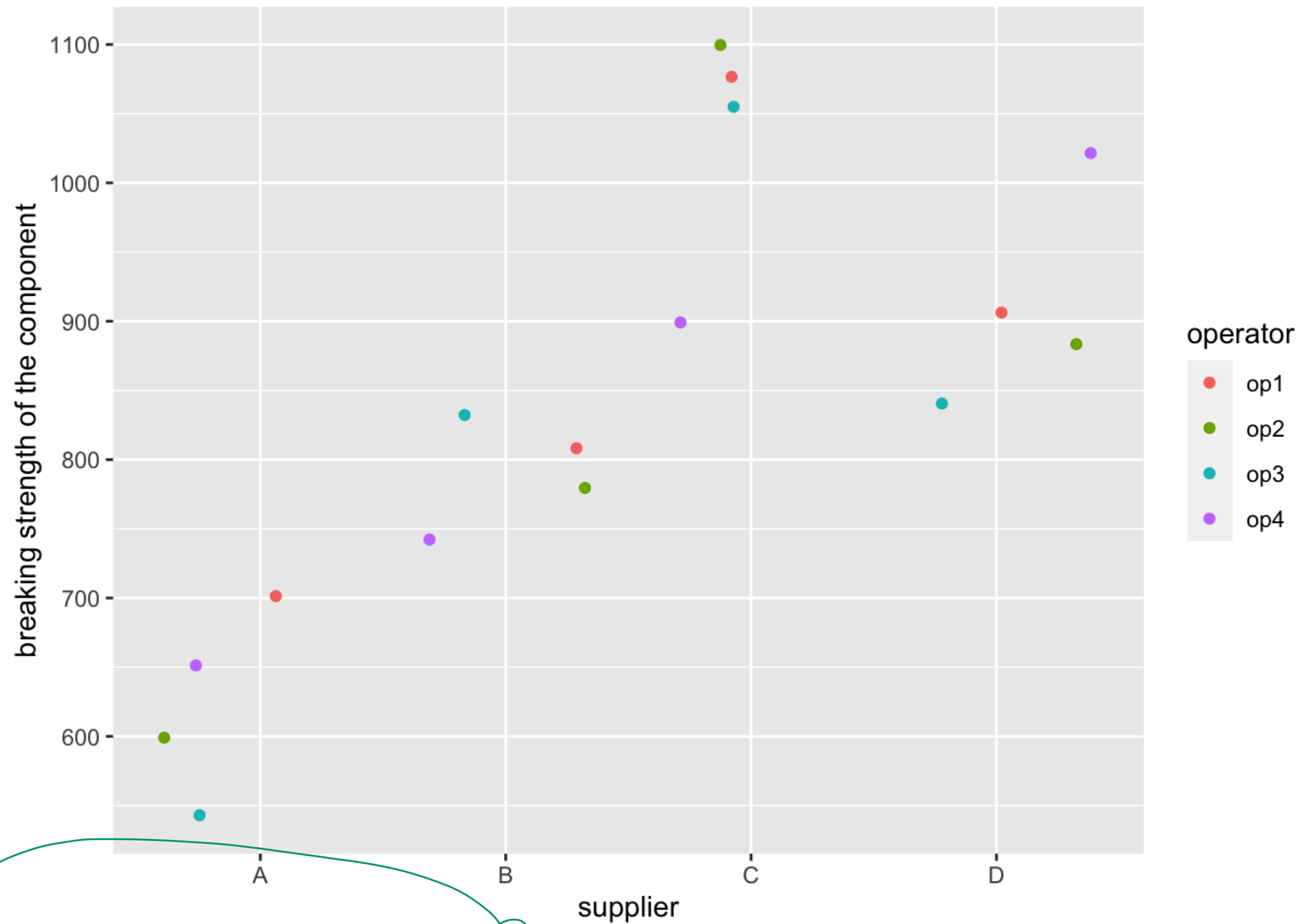
```
##      y operator  day supplier package
## 1  810      op1 day1        B faraway
## 2 1080      op1 day2        C faraway
## 3  700      op1 day3        A faraway
## 4  910      op1 day4        D faraway
## 5 1100      op2 day1        C faraway
## 6  880      op2 day2        D faraway
```

```
summary(br)
```

```
##       y          operator   day    supplier   package
##  Min.   : 540.0   op1:4    day1:4   A:4      Length:16
##  1st Qu.: 730.0   op2:4    day2:4   B:4      Class :character
##  Median : 835.0   op3:4    day3:4   C:4      Mode  :character
##  Mean   : 840.0   op4:4    day4:4   D:4
##  3rd Qu.: 938.8
##  Max.   :1100.0
```

a. Plot the data and interpret.

```
library(ggplot2)
ggplot(br, aes(x=supplier, y=y, color = operator)) +
xlab("supplier") +
ylab("breaking strength of the component") +
geom_point(position = position_jitter(),alpha=1)
```



We see apparent differences in supplier. (b) Fit a fixed effects model for the main effects. Determine which factors are significant. ← Answer: From page 4, supplier C (Estimated as 411.25) is expected to be significant.

p-value?

Also from page 5, anova of lmod tests the p-value as small to show that supplier factor is significant.

```
lmod <- lm(y ~ supplier, br)
anova(lmod)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## supplier   3 371137  123712  23.748 2.464e-05 ***
## Residuals 12  62513    5209
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see that supplier is the significant effect. (c) Fit a mixed effects model with operators and days as random effects but the suppliers as fixed effects. Why is this a natural choice of fixed and random effects?

```
library(tidyverse)
```

For each day, the fixed effects supplier has same condition to predict y.

```
## ── Attaching packages ─────────────────────────────────────────────────────── ti
dyverse 1.3.2 ──
## ✔ tibble  3.1.8      ✔ dplyr   1.0.10
## ✔ tidyr   1.2.0      ✔ stringr 1.4.1
## ✔ readr   2.1.2      ✔ forcats 0.5.2
## ✔ purrr   0.3.4
## ── Conflicts ────────────────────────────────────────────────────────── tidyver
se_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
library(lme4)
```

```
## Loading required package: Matrix
##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
mmod <- lmer(y ~ supplier + (1|operator:day), data = br,    control=lmerControl(chec
k.nobs.vs.nlev = "ignore",
     check.nobs.vs.rankZ = "ignore",
     check.nobs.vs.nRE="ignore"))
summary(mmod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: y ~ supplier + (1 | operator:day)
##    Data: br
## Control:
## lmerControl(check.nobs.vs.nlev = "ignore", check.nobs.vs.rankZ = "ignore",
##     check.nobs.vs.nRE = "ignore")
##
## REML criterion at convergence: 142.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.91940 -0.25992  0.05585  0.28570  0.76473
##
## Random effects:
##  Groups       Name        Variance Std.Dev.
##  operator:day (Intercept) 3927     62.67
##  Residual                 1282     35.81
## Number of obs: 16, groups:  operator:day, 16
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   622.50      36.09  17.250
## supplierB     167.50      51.04   3.282
## supplierC     411.25      51.04   8.058
## supplierD     291.25      51.04   5.707
##
## Correlation of Fixed Effects:
##           (Intr) spplrB spplrC
## supplierB -0.707
## supplierC -0.707  0.500
## supplierD -0.707  0.500  0.500
```

supplier C 35.81

Which supplier results in the highest breaking point? What is the nature of the variation between operators and days? (d) Test the operator and days effects.

```
library(RLRsim)
mmod2 <- lmer(y ~ 1 + (1|operator:day), data = br,    control=lmerControl(check.nobs.
vs.nlev = "ignore",
     check.nobs.vs.rankZ = "ignore",
     check.nobs.vs.nRE="ignore"))
exactRLRT(mmod2)
```

```
## Warning in exactRLRT(mmod2): Null distribution has mass 1 at zero.
```

```
##
##  simulated finite sample distribution of RLRT.
##
##  (p-value based on 10000 simulated values)
##
## data:
## RLRT = 2.8422e-14, p-value < 2.2e-16
```

e. Test the significance of the supplier effect.

```
anova(lmod)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## supplier   3 371137  123712  23.748 2.464e-05 ***
## Residuals 12  62513    5209
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

f. For the best choice of supplier, predict the proportion of components produced in the future that will have a breaking strength less than 1000.

```
ci_func <- function(model, n, level){
sd <- as.numeric(attr(VarCorr(model)$supplier, "stddev"))
sd_boot <- numeric(n)
for(i in 1:n){
y <- unlist(simulate(model))
bmod <- refit(model, y)
sd_boot[i] <- as.data.frame(VarCorr(bmod))$sdcor[1]
}
alpha = 1 - level
ci <- quantile(sd_boot, c(alpha/2, 1 - alpha/2))
return(ci)
}
ci_func(model = mmod, n = 1000, level = 0.98)
```

```
## Warning in optwrap(optimizer, ff, x0, lower = lower, control =
## control$optCtrl, : convergence code -4 from nloptwrap: NLOPT_ROUNDOFF_LIMITED:
## Roundoff errors led to a breakdown of the optimization algorithm. In this case,
## the returned minimum may still be useful. (e.g. this error occurs in NEWUOA if
## one tries to achieve a tolerance too close to machine precision.)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge: degenerate Hessian with 1 negative eigenvalues
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.613103 (tol = 0.002, component 1)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Mod
el is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Mod
el is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0810017 (tol = 0.002, component 1)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Mod
el is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0433428 (tol = 0.002, component 1)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.285082 (tol = 0.002, component 1)
```

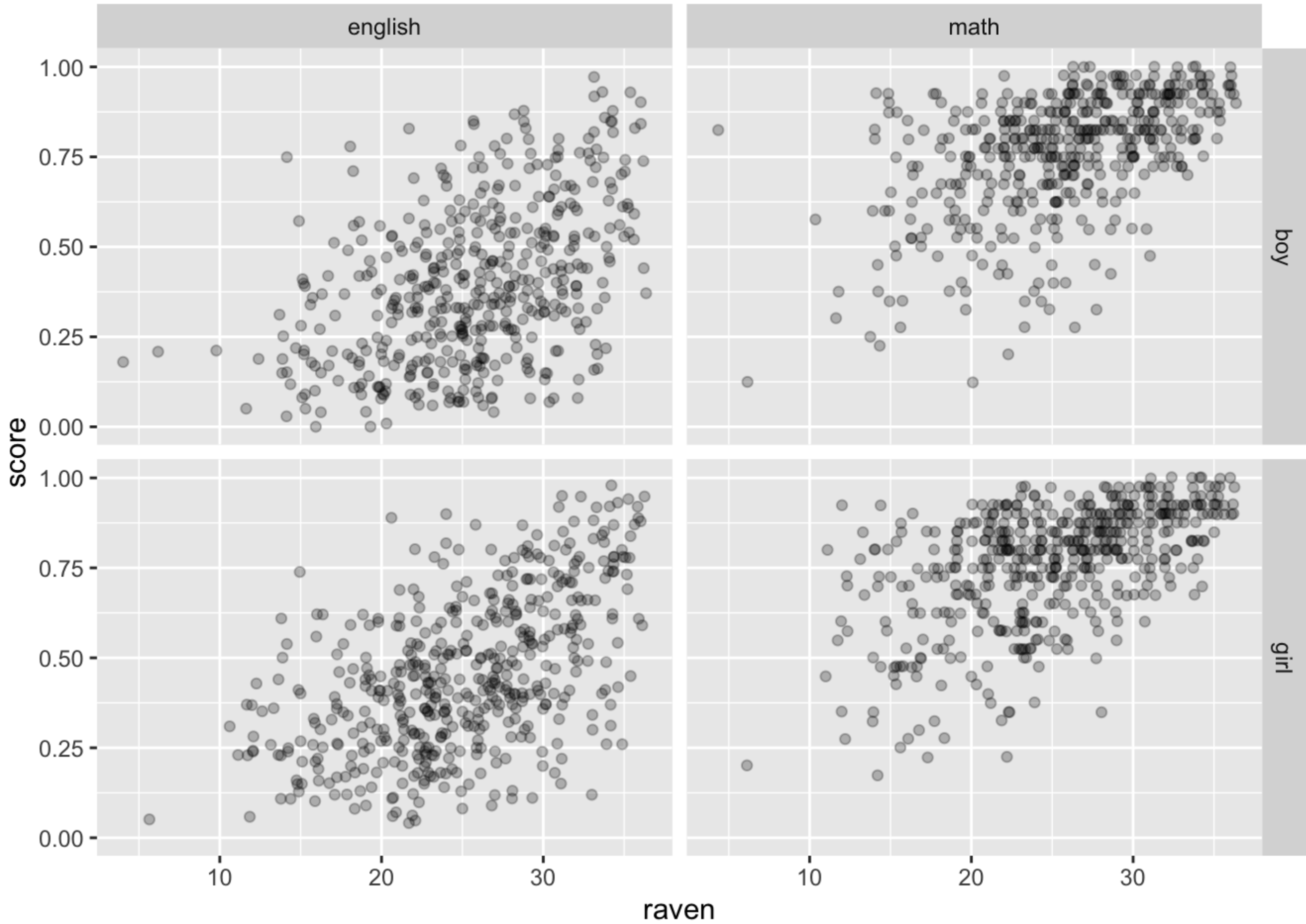
```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Mod
el is nearly unidentifiable: very large eigenvalue
##  - Rescale variables?
```

```
##         1%       99%
##   38.34379 107.39252
```

for the ratio in summary in p.4.

Since the interval does not include 0, the result supports the conclusion we got in mmod model, it is plausible that the SD for the random effect is nonzero, so the random effect may present in our model. \\ ###Q2 8. Redo the Junior Schools Project data analysis in the text with the final year English score as the response. Highlight any differences from the analysis of the final year Math scores.

```
data(jsp, package="faraway")
#final year is 2nd.
jspr <- jsp[jsp$year==2,]


#We set up the data in a format with one test score per line with an
#indicatorsubject identifying which type of test was taken. Scale the
#English and mathtest scores by their maximum possible values, 40 and 100,
#respectively, to aid comparison:
mjspr <- data.frame(rbind(jspr[,1:6],jspr[,1:6]), subject=factor(rep( c("english","ma
th"),c(953,953))), score=c(jspr$english/100, jspr$math/40))

#the final year English score as the response
library(ggplot2)
ggplot(mjspr, aes(x=raven, y=score)) + geom_jitter(alpha=0.25)+facet_grid(gender ~ su
bject)
```




```
#We now fit a model for the data that includes all the variables of interest #that in
corporates some of the interactions that we suspect might be present:
mjspr$craven <- mjspr$raven-mean(mjspr$raven)

library(lme4)
mmod <- lmer(score ~ subject*gender + craven*subject + social + (1|school) + (1|schoo
l:class) + (1|school:class:id),mjspr)
```

The model being fit for school i, class j, student k in subject l is: scorei_jkl = subject_l +gender_k +raven_k +social_k +(sub ject × gender)lk + (raven×sub ject)lk +school_i+class_j +student_k +ϵ_ijkl \ where the Raven score has been mean centered and school, class and student are random effects with the other terms, apart from ϵ, being fixed effects. The summary output:

```
summary(mmod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: score ~ subject * gender + craven * subject + social + (1 | school) +
##     (1 | school:class) + (1 | school:class:id)
##    Data: mjspr
##
## REML criterion at convergence: -1741.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.66538 -0.56918  0.00719  0.56409  2.58699
##
## Random effects:
##  Groups          Name        Variance  Std.Dev.
##  school:class:id (Intercept) 0.0102521 0.10125
##  school:class    (Intercept) 0.0005819 0.02412
##  school          (Intercept) 0.0022306 0.04723
##  Residual                    0.0135916 0.11658
## Number of obs: 1906, groups:
## school:class:id, 953; school:class, 90; school, 48
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)             0.4415780  0.0264593  16.689
## subjectmath             0.3665647  0.0077104  47.542
## gendergirl              0.0633509  0.0102541   6.178
## craven                  0.0173905  0.0009247  18.807
## social2                 0.0137536  0.0272303   0.505
## social3                -0.0207677  0.0289717  -0.717
## social4                -0.0707077  0.0258682  -2.733
## social5                -0.0504741  0.0288184  -1.751
## social6                -0.0878521  0.0306725  -2.864
## social7                -0.0994078  0.0316068  -3.145
## social8                -0.0816234  0.0423516  -1.927
## social9                -0.0473366  0.0274452  -1.725
## subjectmath:gendergirl -0.0591943  0.0107065  -5.529
## subjectmath:craven     -0.0037203  0.0009305  -3.998
```

```
##
## Correlation matrix not shown by default, as p = 14 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

we see that the math subject scores were about 37% higher than the English scores. Since gender has a significant interaction with subject, we must interpret these terms together. We see that on the English test, which is the reference level, girls score 6.3% higher than boys. We see that the scores are strongly related to the entering Raven score although the relation is slightly less strong for math than English.We also see the declining performance as we move down the social class scale as we found in the previous analysis. \ We can test the fixed effects using an F-test incorporating the Kenward-Roger F-test degrees of freedom adjustment:

```r
library(pbkrtest)
mmod <- lmer(score ~ subject*gender+craven*subject+social+ (1|school)
 +(1|school:class)+(1|school:class:id),mjspr, REML=FALSE)
mmodr <- lmer(score ~ subject*gender+craven+subject+social+(1|school)
 +(1|school:class)+(1|school:class:id),mjspr, REML=FALSE)
KRmodcomp(mmod, mmodr)
```

```
## large : score ~ subject + gender + craven + social + (1 | school) + (1 |
##     school:class) + (1 | school:class:id) + subject:gender +
##     subject:craven
## small : score ~ subject * gender + craven + subject + social + (1 | school) +
##     (1 | school:class) + (1 | school:class:id)
##          stat     ndf     ddf F.scaling   p.value
## Ftest  15.987   1.000 950.000         1 6.874e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

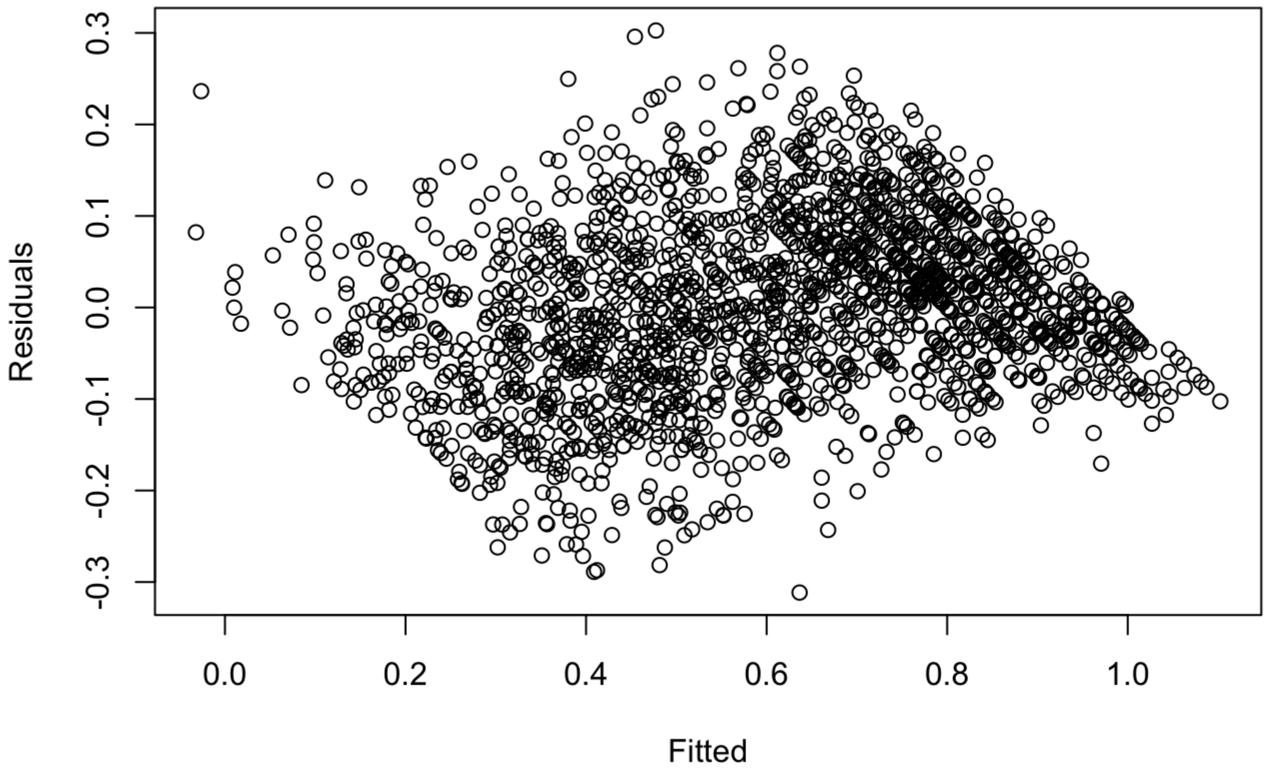
Here we test for a subject by gender interaction. We can see that this effect is strongly statistically significant. Moving to the random effects, we can see from Figure 11.9 that the standard deviation of the residual error in the math scores is smaller than that seen in the English scores. Perhaps this can be ascribed to the greater ease of consistent grading of math assignments or perhaps just greater variation is to be expected in English performance. The correlation between the English and math scores after adjusting for the other effects is also of interest. The last two terms in the model, student of component jkl , represent a 2$$2 covariance matrix for the residual scores for the two tests. We can compute the correlation as:

```r
0.101^2/(0.101^2+0.117^2)
```

```
## [1] 0.4269987
```

giving a moderate positive correlation between the scores. Various diagnostic plots can be made. An interesting one is:

```r
plot(resid(mmod) ~ fitted(mmod),xlab="Fitted",ylab="Residuals")
```




```
#diagd <- augment(mmod)
#ggplot(mmod, aes(x=.fitted,y=.resid)) + geom_point(alpha=0.3) + geom_hline(yintercep
t=0) + facet_grid(~ subject) + xlab("Fitted") +
#ylab("Residuals")
```

\ ###Q3 1. The ratdrink data consist of five weekly measurements of body weight for 27 rats. The first 10 rats are on a control treatment while 7 rats have thyroxine added to their drinking water. Ten rats have thiouracil added to their water.

```
library(lme4)
library(ggplot2)

# Fitting Models using lmer() ---------------------------------------------

ra <- data.frame(ratdrink, package="faraway")
head(ra)
```

```
##    wt weeks subject   treat package
## 1  57     0       1 control faraway
## 2  86     1       1 control faraway
## 3 114     2       1 control faraway
## 4 139     3       1 control faraway
## 5 172     4       1 control faraway
## 6  60     0       2 control faraway
```

```
summary(ra)
```

```
##        wt             weeks      subject          treat      package
##  Min.   : 46.0   Min.   :0   1       :  5   control   :50   Length:135
##  1st Qu.: 71.0   1st Qu.:1   2       :  5   thiouracil:50   Class :character
##  Median :100.0   Median :2   3       :  5   thyroxine :35   Mode  :character
##  Mean   :100.8   Mean   :2   4       :  5
##  3rd Qu.:122.5   3rd Qu.:3   5       :  5
##  Max.   :189.0   Max.   :4   6       :  5
##                              (Other):105
```

a. Plot the data showing how weight increases with age on a single panel, taking care to distinguish the three treatment groups. Now create a three-panel plot, one for each group. Discuss what can be seen.

```
library(tidyverse)
library(pbkrtest)
library(faraway)
library(RLRsim)
library(broom.mixed)
library(car)
```

```
## Loading required package: carData
```

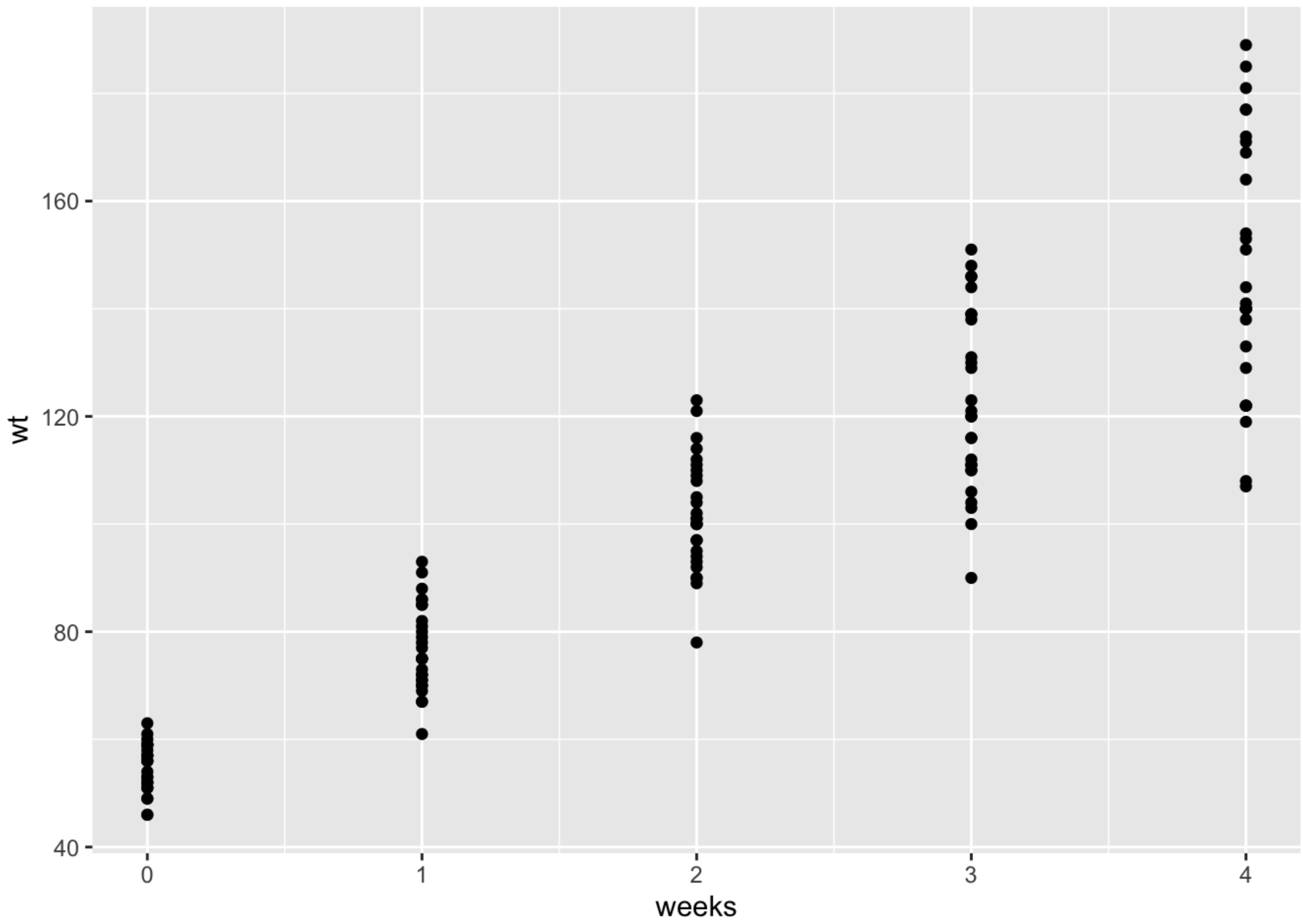
```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

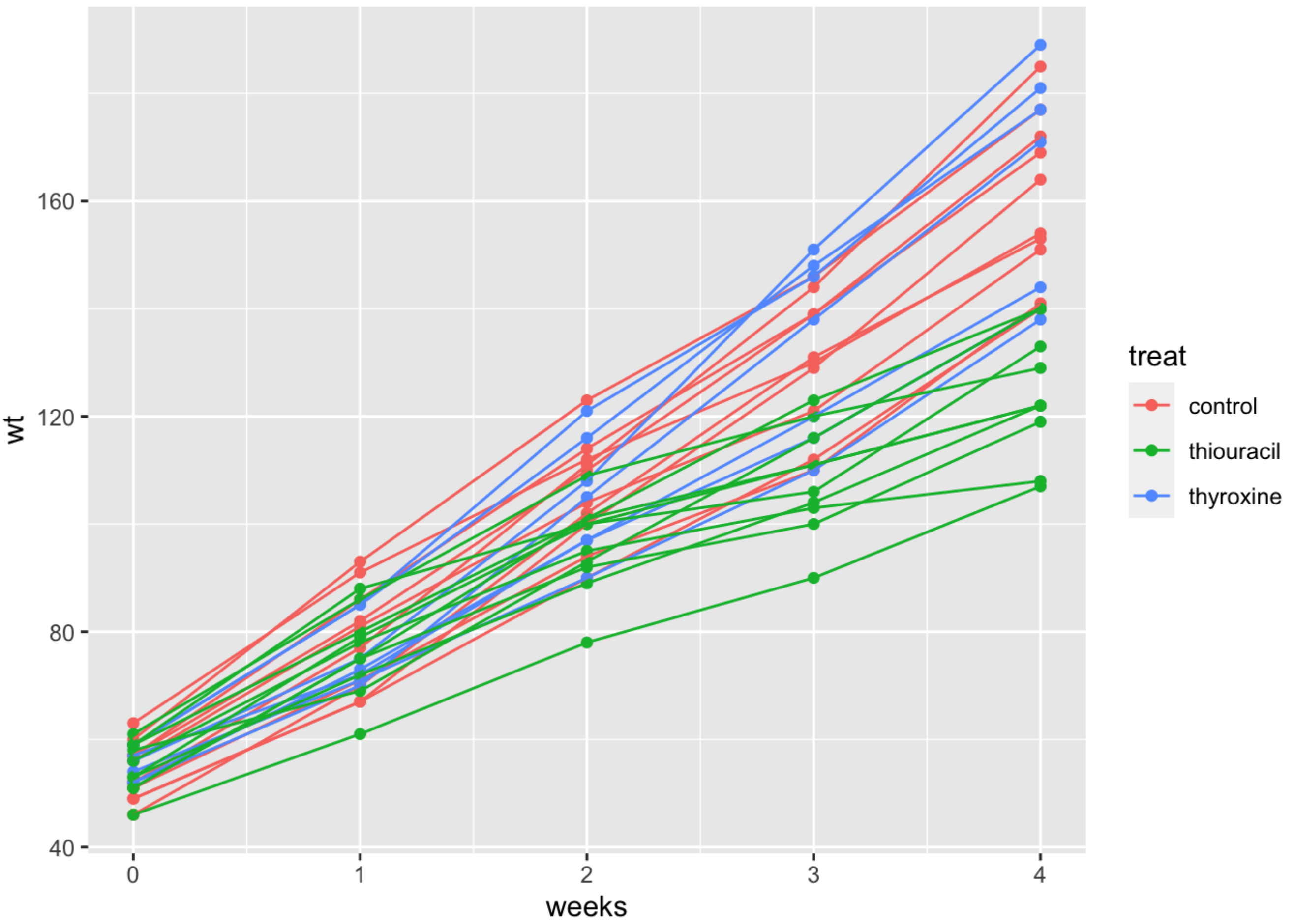
```
## The following objects are masked from 'package:faraway':
##
##     logit, vif
```

```
# Let's draw exploratory plots
# scatterplot with no grouping
ggplot(ra, aes(x=weeks, y=wt)) + geom_point()
```

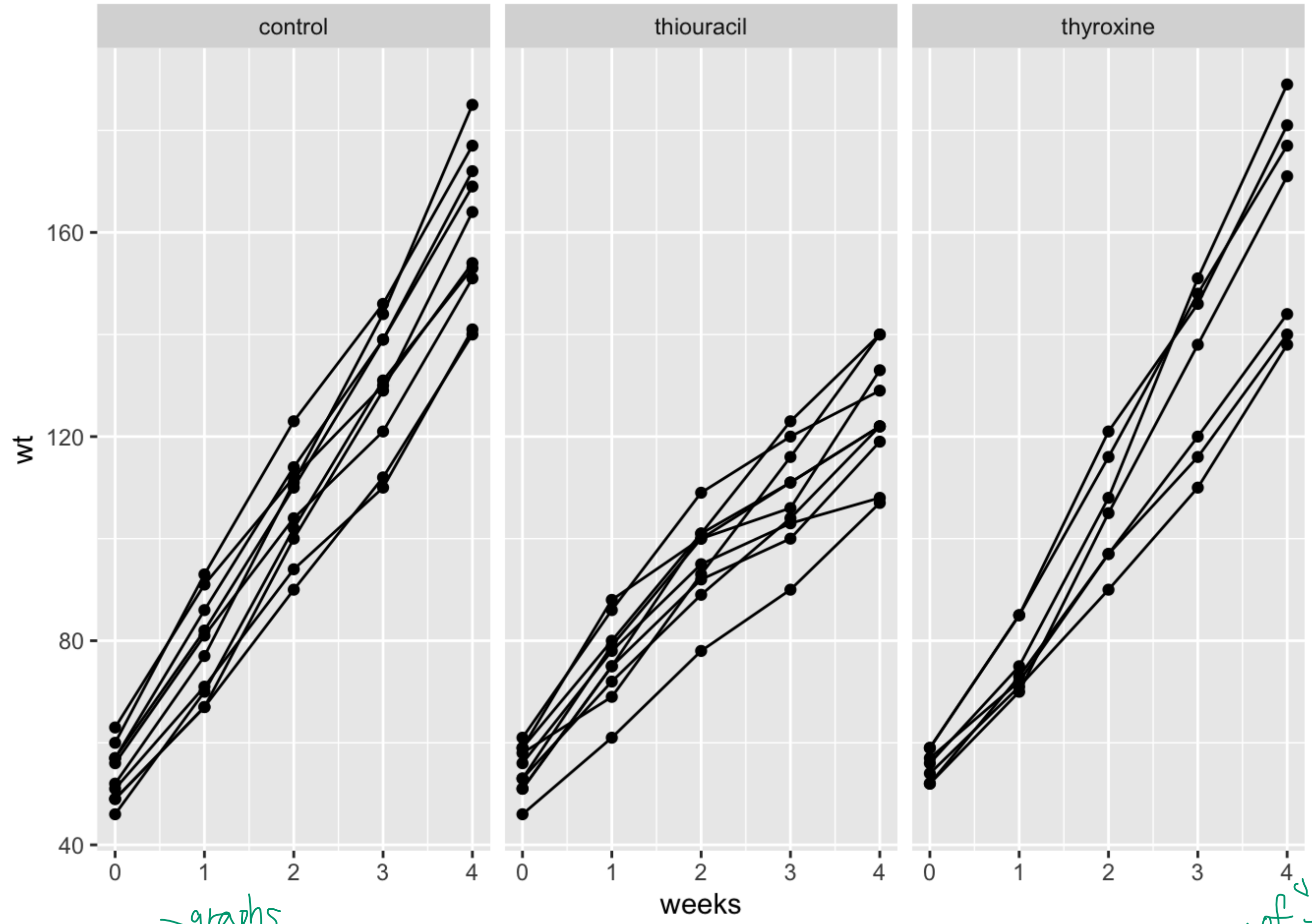



```
# wt-weeks colored by treat and with grouping
ggplot(ra, aes(x=weeks, y=wt, color=treat, group=subject)) +
  geom_point() + geom_line()
```

As weeks go, the wt show difference more.




```
ggplot(ra, aes(x=weeks, y=wt, group=subject)) +
  geom_point() + geom_line()+facet_wrap(~ treat)
```



From the above ~~graph~~ graphs, it seems that a linear model fits well. The slopes are different between subjects. of treat"

b. Fit a linear longitudinal model that allows for a random slope and intercept for each rat. Each group should have a different mean line. Give interpretation for the following estimates:

\ i. The fixed effect intercept term.

```
#####################
# Longitudinal data
#####################
#ra$cwt <- ra$wt-100 # Using the median wt
mmod <- lmer(wt ~ subject + (1|subject),ra)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## unable to evaluate scaled gradient
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Hessian is numerically singular: parameters are not uniquely determined
```

```
summary(mmod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: wt ~ subject + (1 | subject)
##    Data: ra
##
## REML criterion at convergence: 1135.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.6243 -0.7859  0.0368  0.6860  1.9239
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept)  149.2   12.21
##  Residual             1447.6   38.05
## Number of obs: 135, groups:  subject, 27
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   113.60      20.94   5.424
## subject2        6.20      29.62   0.209
## subject3        0.20      29.62   0.007
## subject4      -11.80      29.62  -0.398
## subject5      -11.00      29.62  -0.371
## subject6      -13.20      29.62  -0.446
## subject7      -20.20      29.62  -0.682
## subject8       -3.60      29.62  -0.122
## subject9      -22.00      29.62  -0.743
## subject10      -2.20      29.62  -0.074
## subject11       4.80      29.62   0.162
## subject12     -21.00      29.62  -0.709
## subject13       2.20      29.62   0.074
## subject14       3.40      29.62   0.115
## subject15     -15.60      29.62  -0.527
## subject16     -18.00      29.62  -0.608
## subject17      -6.40      29.62  -0.216
## subject18     -12.60      29.62  -0.425
## subject19     -19.00      29.62  -0.641
## subject20     -19.40      29.62  -0.655
## subject21     -17.60      29.62  -0.594
## subject22     -15.60      29.62  -0.527
## subject23     -26.20      29.62  -0.885
## subject24     -25.60      29.62  -0.864
## subject25     -18.40      29.62  -0.621
## subject26     -37.20      29.62  -1.256
## subject27     -25.60      29.62  -0.864
```

```
##
## Correlation matrix not shown by default, as p = 27 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```
## optimizer (nloptwrap) convergence code: 0 (OK)
## unable to evaluate scaled gradient
##  Hessian is numerically singular: parameters are not uniquely determined
```

\ ii. The interaction between thiouracil and week.

```
# fit model with interaction;
# implies different treatments lead to different slopes
# include random slopes and intercept
mmod2 <- lmer(wt ~ treat + weeks + treat:weeks + (weeks | subject),
             data=ratdrink)
# or mmod2 <- lmer(wt ~ treat * weeks + (weeks | subject), data=ratdrink)
summary(mmod2, corr=F)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: wt ~ treat + weeks + treat:weeks + (weeks | subject)
##    Data: ratdrink
##
## REML criterion at convergence: 878.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.83136 -0.54991  0.04003  0.58230  2.03660
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  subject  (Intercept) 32.49    5.700
##           weeks       14.14    3.760    -0.13
##  Residual             18.90    4.348
## Number of obs: 135, groups:  subject, 27
##
## Fixed effects:
##                        Estimate Std. Error t value
## (Intercept)             52.8800     2.0937  25.256
## treatthiouracil          4.7800     2.9610   1.614
## treatthyroxine          -0.7943     3.2628  -0.243
## weeks                   26.4800     1.2661  20.915
## treatthiouracil:weeks   -9.3700     1.7905  -5.233
## treatthyroxine:weeks     0.6629     1.9730   0.336
```

← "weeks" is the most effective term.

(*)

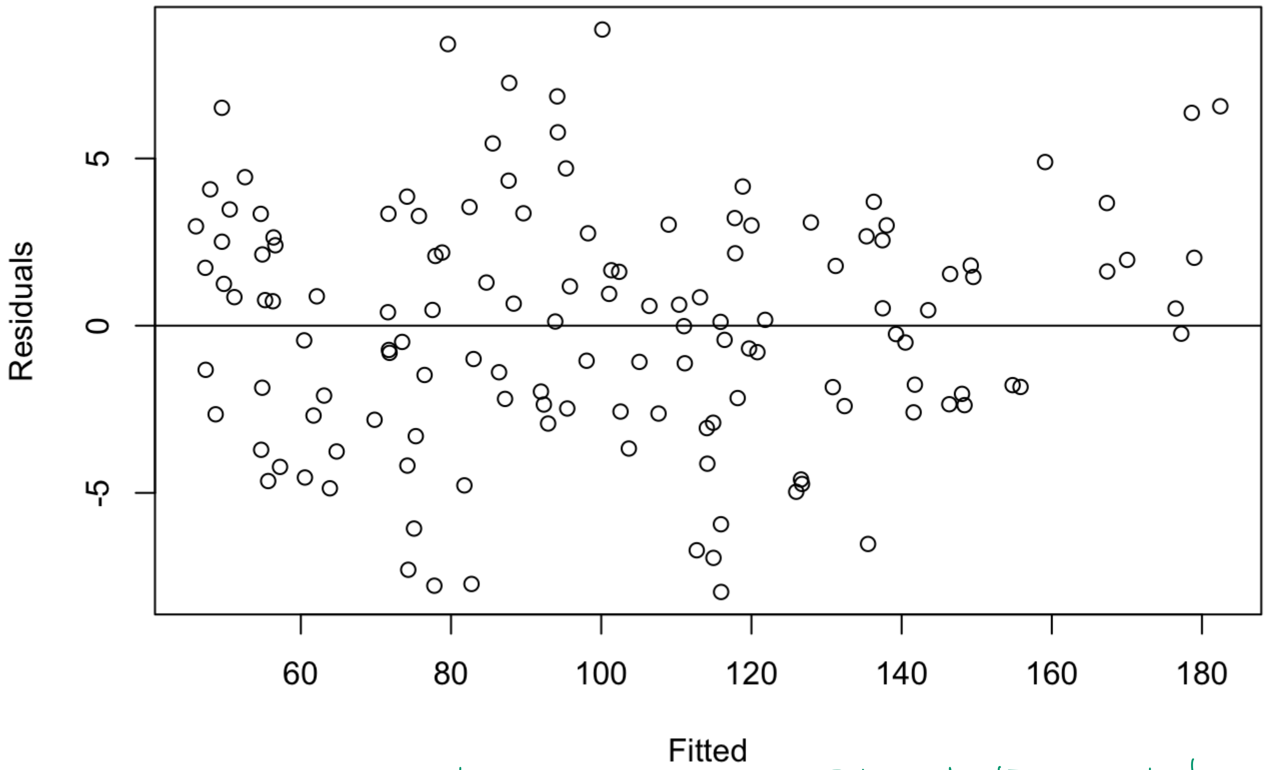
```
# Interpreting interaction:
# The intercept (52.88) and weeks (26.48) coefficients are the fitted line for
# the control group.

# Intercept + treatthiouracil is the intercept for the thiouracil group.
# 52.88 + 4.78 = 57.66.
# weeks + treatthiouracil:weeks is the slope for the thiouracil group.
# 26.48 - 9.37 = 17.11

# treatthiouracil:weeks = -9.37 means the trajectory for thiouracil is lower
# than the control group.
```

\ iii. The intercept random effect SD. ← That is 5.7 from above summary.

* \ (c) Check whether there is a significant treatment effect. p-value? The (*) in the p.16 is significant.

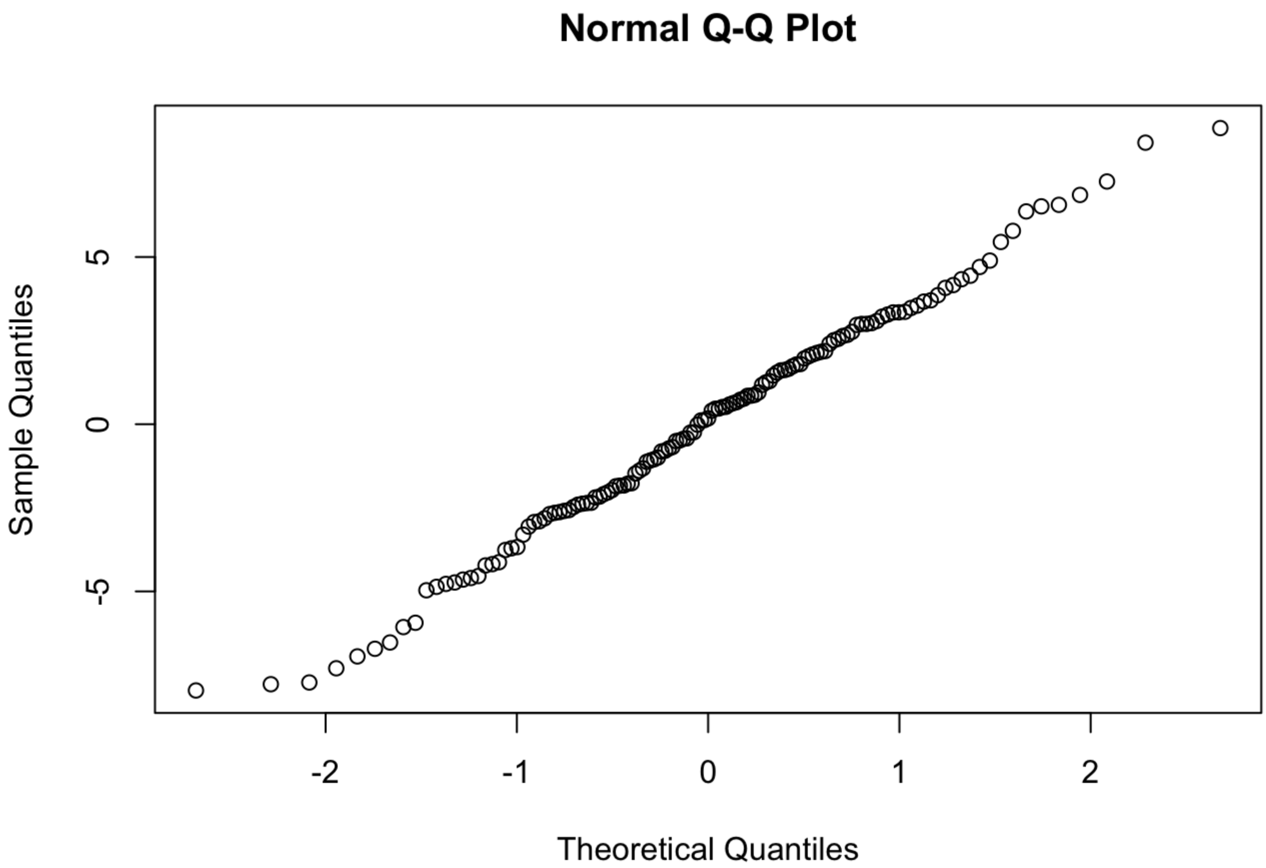
\ (d) Construct diagnostic plots showing the residuals against the fitted values and a QQ plot of the residuals. Interpret.

```
plot(resid(mmod2) ~ fitted(mmod2),xlab="Fitted",ylab="Residuals")
abline(h=0)
```



From above plot, it seems that there are some points with big residuals.

```
qqnorm(residuals(mmod2))
```



From above, q-q plot does not have problem.

```
#diagd1 <- augment(mmod2)
#ggplot(diagd1,aes(sample=.resid))+geom_qq()+facet_grid(~subject)+geom_qq_line()
```

\ (e) Construct confidence intervals for the parameters of the model. Which random effect terms may not be significant? Is the thyroxine group significantly different from the control group? Ans.

```
confint(mmod2, method="boot")
```

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

```
##
## 3 warning(s): Model failed to converge with max|grad| = 0.00308525 (tol = 0.002, c
omponent 1) (and others)
```

```
##                              2.5 %      97.5 %
## .sig01                   3.4862172   8.3078751
## .sig02                  -0.5994577   0.3669238
## .sig03                   2.6015599   5.0505488
## .sigma                   3.7044752   4.9476990
## (Intercept)             48.8874833  57.1930591
## treatthiouracil         -0.8642830  10.8276274
## treatthyroxine          -7.6393919   5.4635742
## weeks                   24.0453141  29.0094577
## treatthiouracil:weeks  -12.8908909  -5.8073502
## treatthyroxine:weeks    -3.2850948   4.5708160
```

Ans.
They don't contain 0.
So s.d. of random effect of this model is nonzero.
So yes, significantly different.

\ 5. The sleepstudy data found in the lme4 package describes the reaction times of subjects who are progressively sleep deprived.

```
library(lme4)
sb <- data.frame(sleepstudy, package="lme4")
head(sb)
```

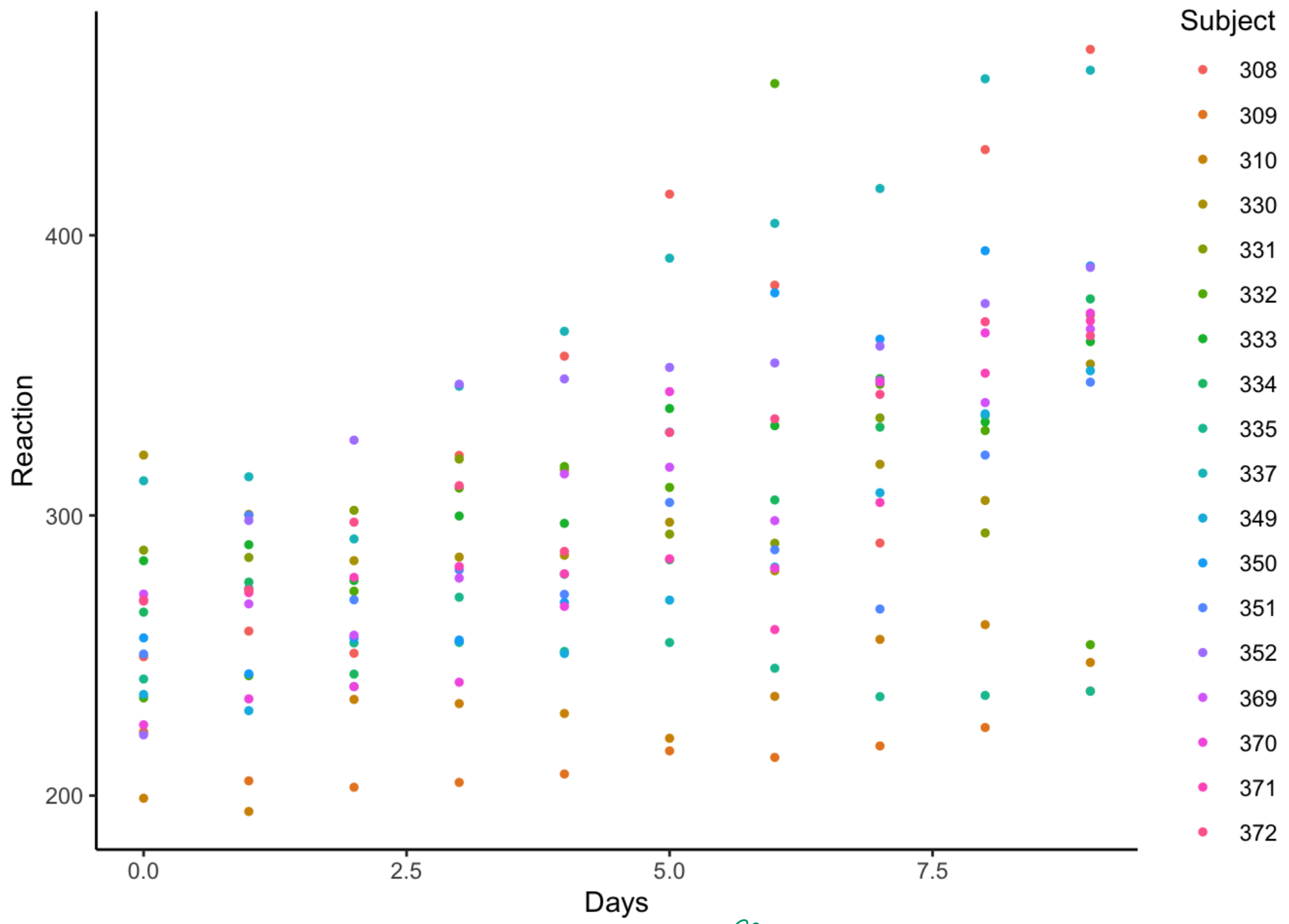
```
##   Reaction Days Subject package
## 1 249.5600    0     308    lme4
## 2 258.7047    1     308    lme4
## 3 250.8006    2     308    lme4
## 4 321.4398    3     308    lme4
## 5 356.8519    4     308    lme4
## 6 414.6901    5     308    lme4
```

```
summary(sb)
```

```
##     Reaction          Days        Subject      package         
##  Min.   :194.3   Min.   :0.0   308    : 10   Length:180        
##  1st Qu.:255.4   1st Qu.:2.0   309    : 10   Class :character  
##  Median :288.7   Median :4.5   310    : 10   Mode  :character  
##  Mean   :298.5   Mean   :4.5   330    : 10                     
##  3rd Qu.:336.8   3rd Qu.:7.0   331    : 10                     
##  Max.   :466.4   Max.   :9.0   332    : 10                     
##                                (Other):120                     
```

\ (a) Plot the data taking care to distinguish the trajectories of the different subjects. Comment on the pattern of variation.

```
ggplot(sb,aes(x=Days, y=Reaction, group=Subject, color=Subject)) +
    geom_point(size=1) + theme_classic()
```



\ For each subjects, reaction values are different.

b. Fit a mixed effects model that describes how the reaction time varies linearly with days and allows for random variation in both the slope and intercepts of the subject lines. Under this model, would it be unusual for an individual to have a reaction time that does not increase over time? → Ans.

```
#y:reaction, mixed:Days, random slope:Subject
model1 <- lmer(Reaction ~ Days + (Days|Subject), sb,REML=F)
summary(model1)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Reaction ~ Days + (Days | Subject)
##    Data: sb
##
##      AIC      BIC   logLik deviance df.resid
##   1763.9   1783.1   -876.0   1751.9      174
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9416 -0.4656  0.0289  0.4636  5.1793
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 565.48   23.780
##           Days         32.68    5.717   0.08
##  Residual             654.95   25.592
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  251.405      6.632  37.907
## Days          10.467      1.502   6.968
##
## Correlation of Fixed Effects:
##      (Intr)
## Days -0.138
```

Ans.

$\Rightarrow$ Yes, it is unnsual frome the estimate.

\ (c) Allow for quadratic effects in the previous model. Does the data support the inclusion of quadratic effects?

```
model2 <- lmer(Reaction ~ Days+ I(Days^2) + (Days+ I(Days^2)|Subject), sb,REML=F)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0171207 (tol = 0.002, component 1)
```

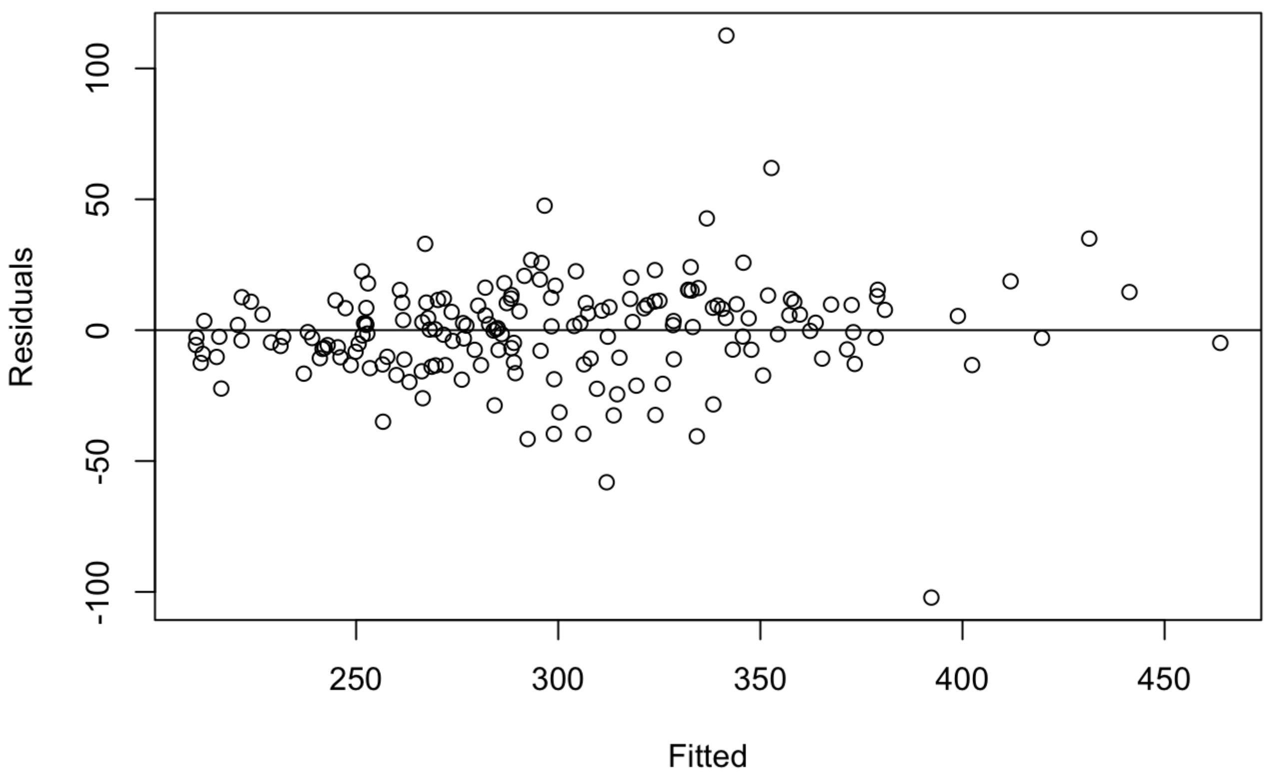
```
summary(model2)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: Reaction ~ Days + I(Days^2) + (Days + I(Days^2) | Subject)
##    Data: sb
##
##      AIC      BIC   logLik deviance df.resid
##   1757.6   1789.5   -868.8   1737.6      170
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4876 -0.4508  0.0434  0.4578  4.9453
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 742.232  27.244
##           Days        196.636  14.023   -0.40
##           I(Days^2)     1.958   1.399    0.45 -0.91
##  Residual             518.184  22.764
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 255.4494     7.6832  33.248
## Days          7.4341     3.9610   1.877
## I(Days^2)     0.3370     0.4041   0.834
##
## Correlation of Fixed Effects:
##           (Intr) Days
## Days      -0.525
## I(Days^2)  0.518 -0.925
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.0171207 (tol = 0.002, component 1)
```

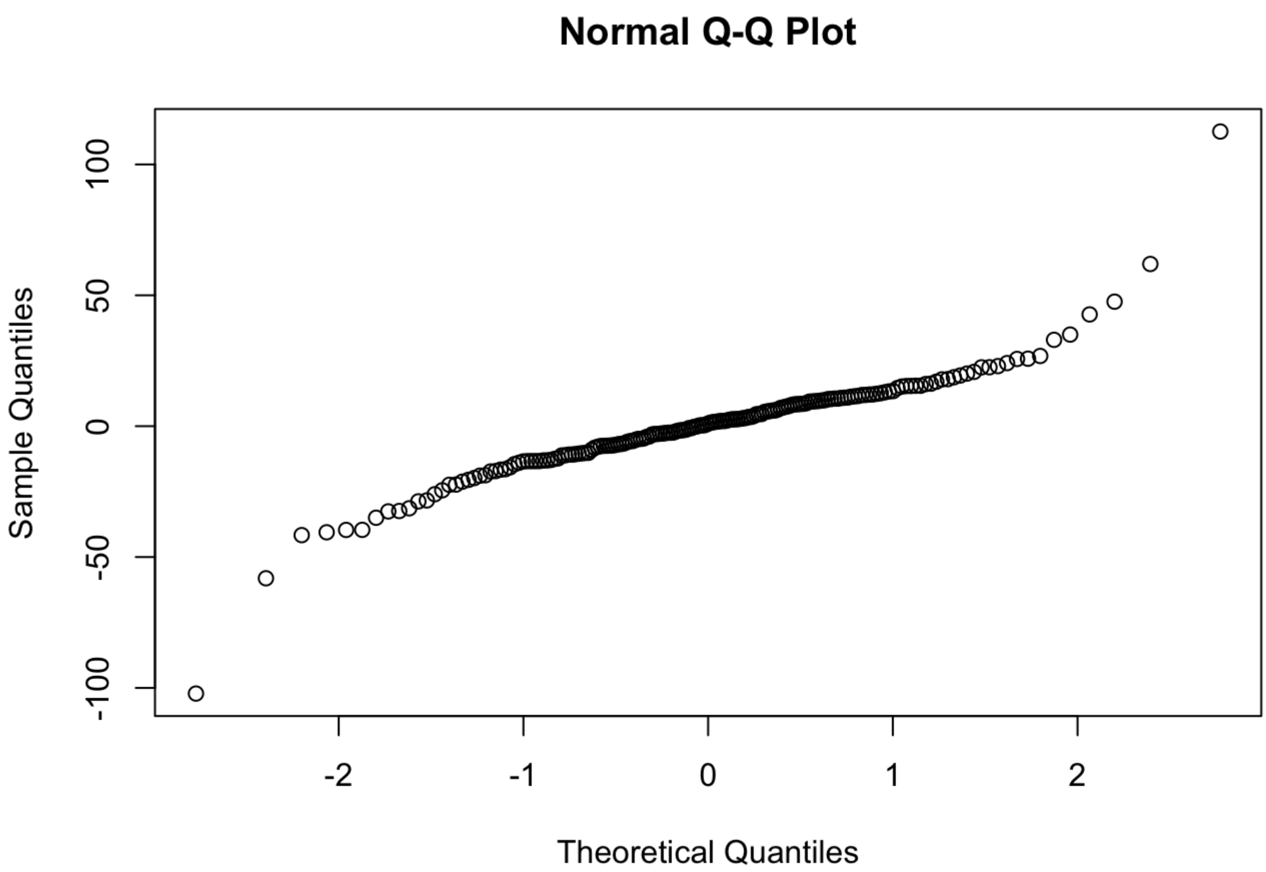
The quadratic term is not effective.

\ (d) Make the following diagnostic plots and interpret: (i) Residuals vs. Fitted plot, (ii) QQ plot of the residuals, (iii) QQ plot of both random effects, (iv) a scatterplot of the random effects.

```
plot(resid(model2) ~ fitted(model2),xlab="Fitted",ylab="Residuals")
abline(h=0)
```




```
qqnorm(residuals(model2))
```



```r
library(lme4)
ran<-data.frame(ranef(model2))
ranef(model2)
```

```
## $Subject
##     (Intercept)        Days   I(Days^2)
## 308   -3.866839  13.1142884 -0.40066580
## 309  -34.704532 -13.5801333  0.55945565
## 310  -43.892449  -2.3624930 -0.33100848
## 330   40.421240 -16.6024999  1.27741535
## 331   26.520255  -5.3154185  0.21561770
## 332  -31.456353  31.8124067 -3.61154530
## 333   16.252259   0.8447989 -0.13809302
## 334    6.183773  -9.8052440  1.23330587
## 335   -5.588652  -5.5174322 -0.63521731
## 337   36.174201   7.9272107  0.08163681
## 349  -16.344436  -6.7731006  0.92285521
## 350  -10.561230   3.5011380  0.39227624
## 351   10.785113  -7.4059731  0.47466911
## 352    1.207749  19.2959439 -1.77435069
## 369    5.937502  -1.1922872  0.23227059
## 370  -24.152156   2.2837934  0.33259800
## 371   16.156939 -12.8885585  1.33044131
## 372   10.927617   2.6635603 -0.16166125
##
## with conditional variances for "Subject"
```

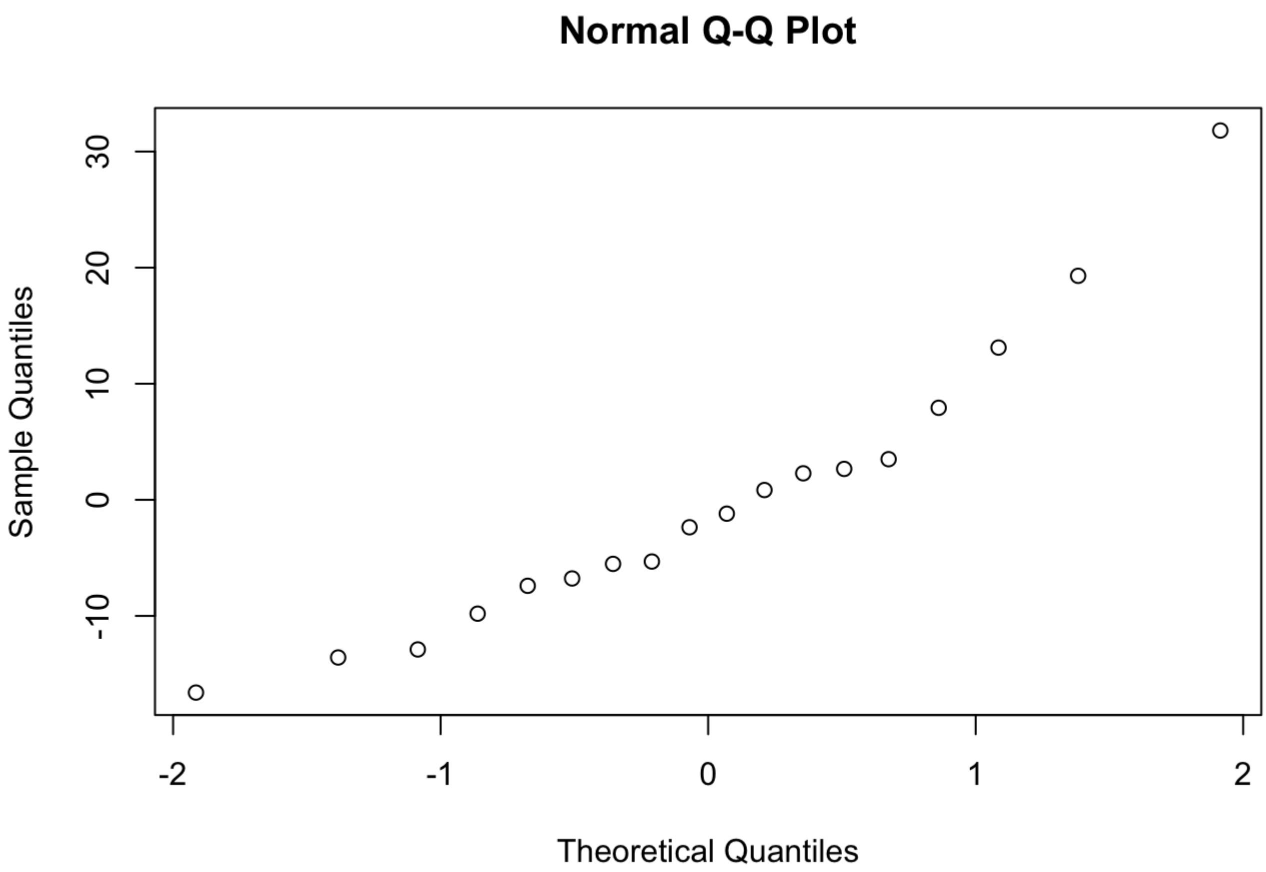
```r
ran
```

```
##     grpvar        term grp      condval     condsd
## 1  Subject (Intercept) 308  -3.86683875 14.4395875
## 2  Subject (Intercept) 309 -34.70453242 14.4395875
## 3  Subject (Intercept) 310 -43.89244917 14.4395875
## 4  Subject (Intercept) 330  40.42124012 14.4395875
## 5  Subject (Intercept) 331  26.52025539 14.4395875
## 6  Subject (Intercept) 332 -31.45635310 14.4395875
## 7  Subject (Intercept) 333  16.25225921 14.4395875
## 8  Subject (Intercept) 334   6.18377347 14.4395875
## 9  Subject (Intercept) 335  -5.58865239 14.4395875
## 10 Subject (Intercept) 337  36.17420068 14.4395875
## 11 Subject (Intercept) 349 -16.34443597 14.4395875
## 12 Subject (Intercept) 350 -10.56122973 14.4395875
## 13 Subject (Intercept) 351  10.78511268 14.4395875
## 14 Subject (Intercept) 352   1.20774872 14.4395875
## 15 Subject (Intercept) 369   5.93750181 14.4395875
## 16 Subject (Intercept) 370 -24.15215577 14.4395875
## 17 Subject (Intercept) 371  16.15693863 14.4395875
## 18 Subject (Intercept) 372  10.92761658 14.4395875
## 19 Subject        Days 308  13.11428841  7.4781772
## 20 Subject        Days 309 -13.58013327  7.4781772
## 21 Subject        Days 310  -2.36249297  7.4781772
## 22 Subject        Days 330 -16.60249989  7.4781772
## 23 Subject        Days 331  -5.31541850  7.4781772
## 24 Subject        Days 332  31.81240670  7.4781772
## 25 Subject        Days 333   0.84479891  7.4781772
## 26 Subject        Days 334  -9.80524404  7.4781772
## 27 Subject        Days 335  -5.51743223  7.4781772
## 28 Subject        Days 337   7.92721075  7.4781772
## 29 Subject        Days 349  -6.77310060  7.4781772
## 30 Subject        Days 350   3.50113799  7.4781772
## 31 Subject        Days 351  -7.40597307  7.4781772
## 32 Subject        Days 352  19.29594389  7.4781772
## 33 Subject        Days 369  -1.19228724  7.4781772
## 34 Subject        Days 370   2.28379336  7.4781772
## 35 Subject        Days 371 -12.88855852  7.4781772
## 36 Subject        Days 372   2.66356031  7.4781772
## 37 Subject   I(Days^2) 308  -0.40066580  0.8046666
## 38 Subject   I(Days^2) 309   0.55945565  0.8046666
## 39 Subject   I(Days^2) 310  -0.33100848  0.8046666
## 40 Subject   I(Days^2) 330   1.27741535  0.8046666
## 41 Subject   I(Days^2) 331   0.21561770  0.8046666
## 42 Subject   I(Days^2) 332  -3.61154530  0.8046666
## 43 Subject   I(Days^2) 333  -0.13809302  0.8046666
## 44 Subject   I(Days^2) 334   1.23330587  0.8046666
## 45 Subject   I(Days^2) 335  -0.63521731  0.8046666
## 46 Subject   I(Days^2) 337   0.08163681  0.8046666
## 47 Subject   I(Days^2) 349   0.92285521  0.8046666
## 48 Subject   I(Days^2) 350   0.39227624  0.8046666
## 49 Subject   I(Days^2) 351   0.47466911  0.8046666
## 50 Subject   I(Days^2) 352  -1.77435069  0.8046666
## 51 Subject   I(Days^2) 369   0.23227059  0.8046666
## 52 Subject   I(Days^2) 370   0.33259800  0.8046666
## 53 Subject   I(Days^2) 371   1.33044131  0.8046666
## 54 Subject   I(Days^2) 372  -0.16166125  0.8046666
```

```
#typeof(ran)
#ran2<-ran[,2]
#qqnorm(ran2)
```

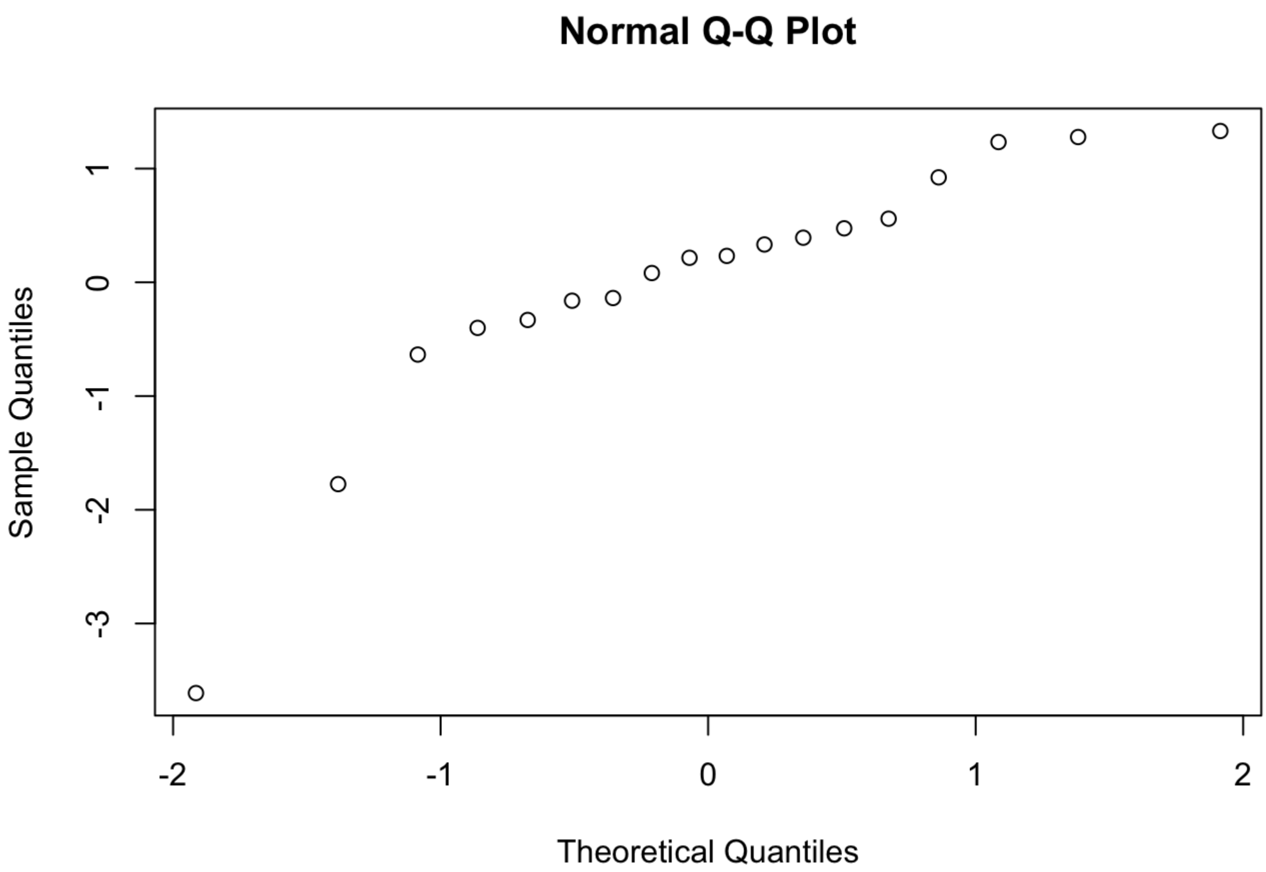
```
ranefdf<-data.frame( c(13.1142884, -0.40066580),
 c(-13.5801333,  0.55945565),
c (-2.3624930, -0.33100848),
c (-16.6024999,  1.27741535),
c (-5.3154185,  0.21561770),
c (31.8124067, -3.61154530),
 c (0.8447989, -0.13809302),
c (-9.8052440,  1.23330587),
c ( -5.5174322, -0.63521731),
c ( 7.9272107,  0.08163681),
c (-6.7731006,  0.92285521),
c  (3.5011380 , 0.39227624),
c  (-7.4059731 , 0.47466911),
c (19.2959439, -1.77435069),
c (-1.1922872,  0.23227059),
c  (2.2837934,  0.33259800),
c (-12.8885585,  1.33044131),
c  (2.6635603, -0.16166125))
qqnorm(ranefdf[1,])
```

```
## Warning in xtfrm.data.frame(x): cannot xtfrm data frames
```



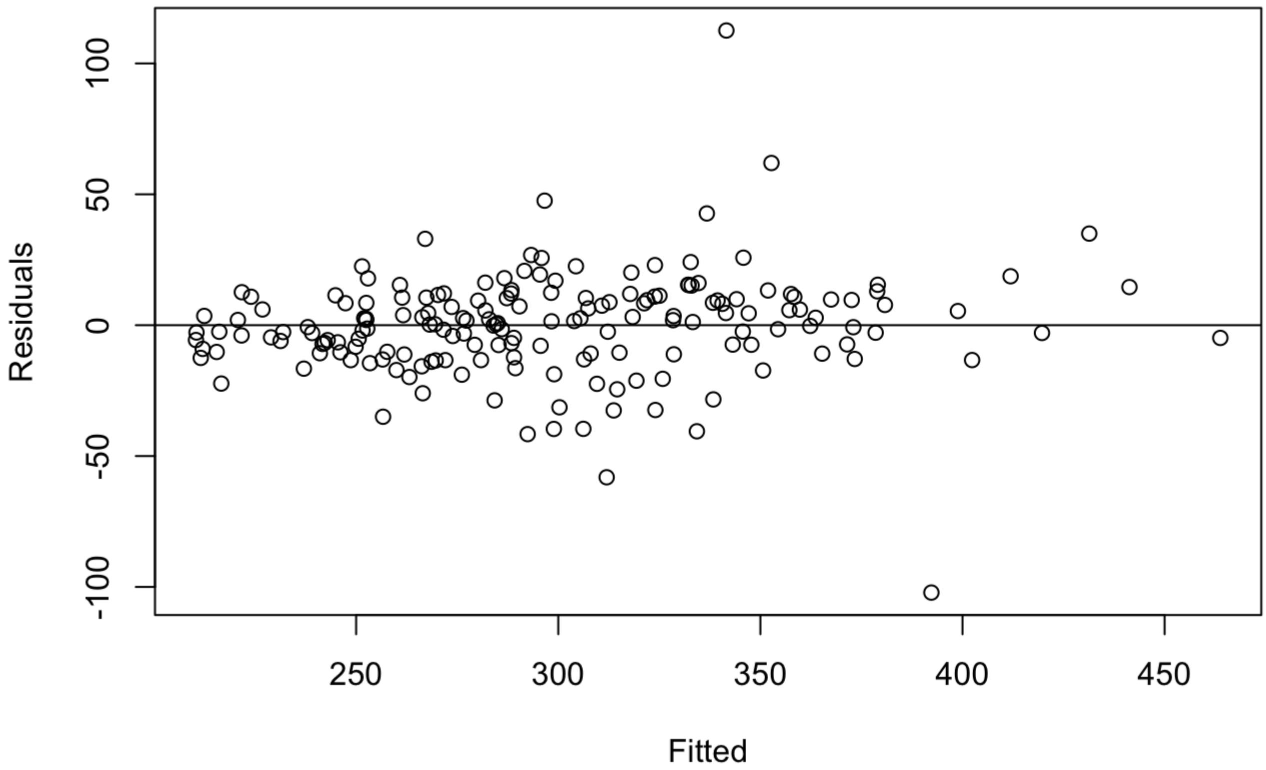
```
qqnorm(ranefdf[2,])
```

```
## Warning in xtfrm.data.frame(x): cannot xtfrm data frames
```



\ (e) Identify any outlying cases and mark these on top of your initial plot. Try refitting the model without these cases and identify the largest change in the model fit.

```
plot(resid(model2) ~ fitted(model2),xlab="Fitted",ylab="Residuals")
abline(h=0)
```



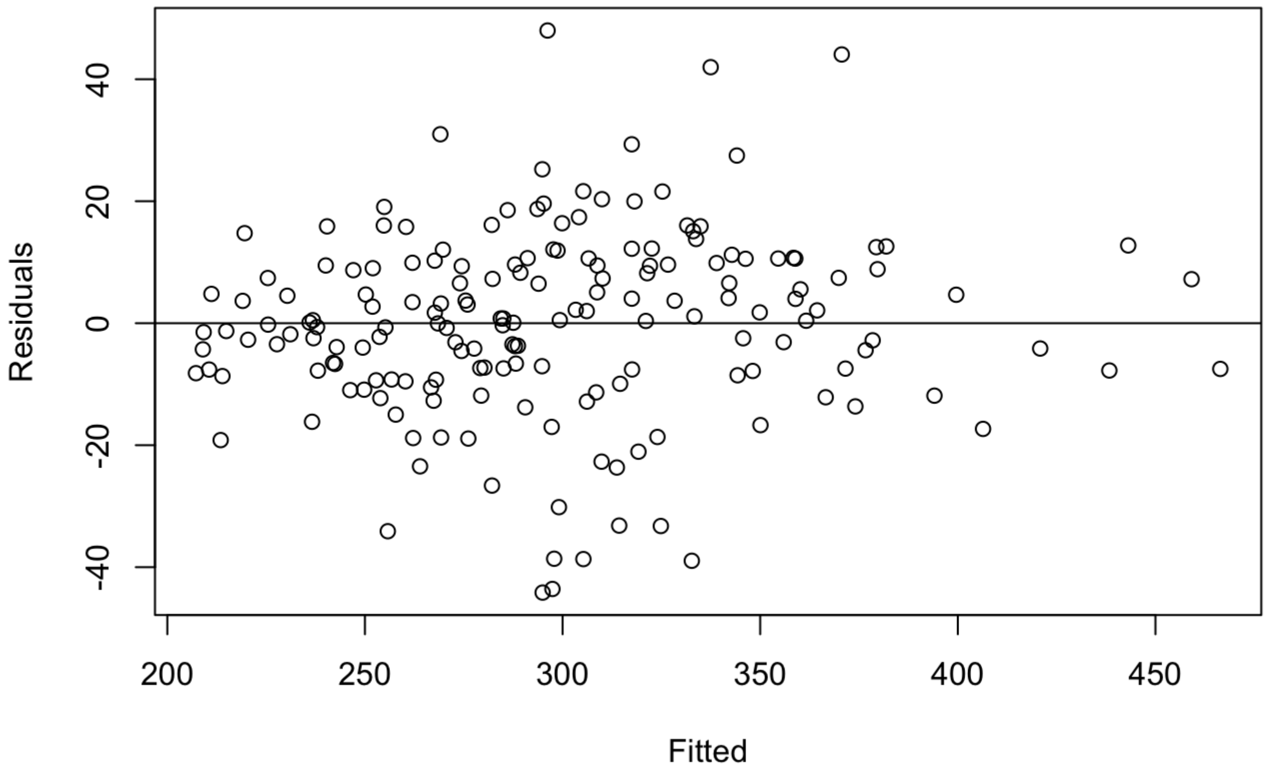

```
library(car)
outlierTest(model2)
```

```
##      rstudent unadjusted p-value Bonferroni p
## 57  5.429133         1.9416e-07   3.4948e-05
## 8  -4.924702         1.9983e-06   3.5969e-04
```

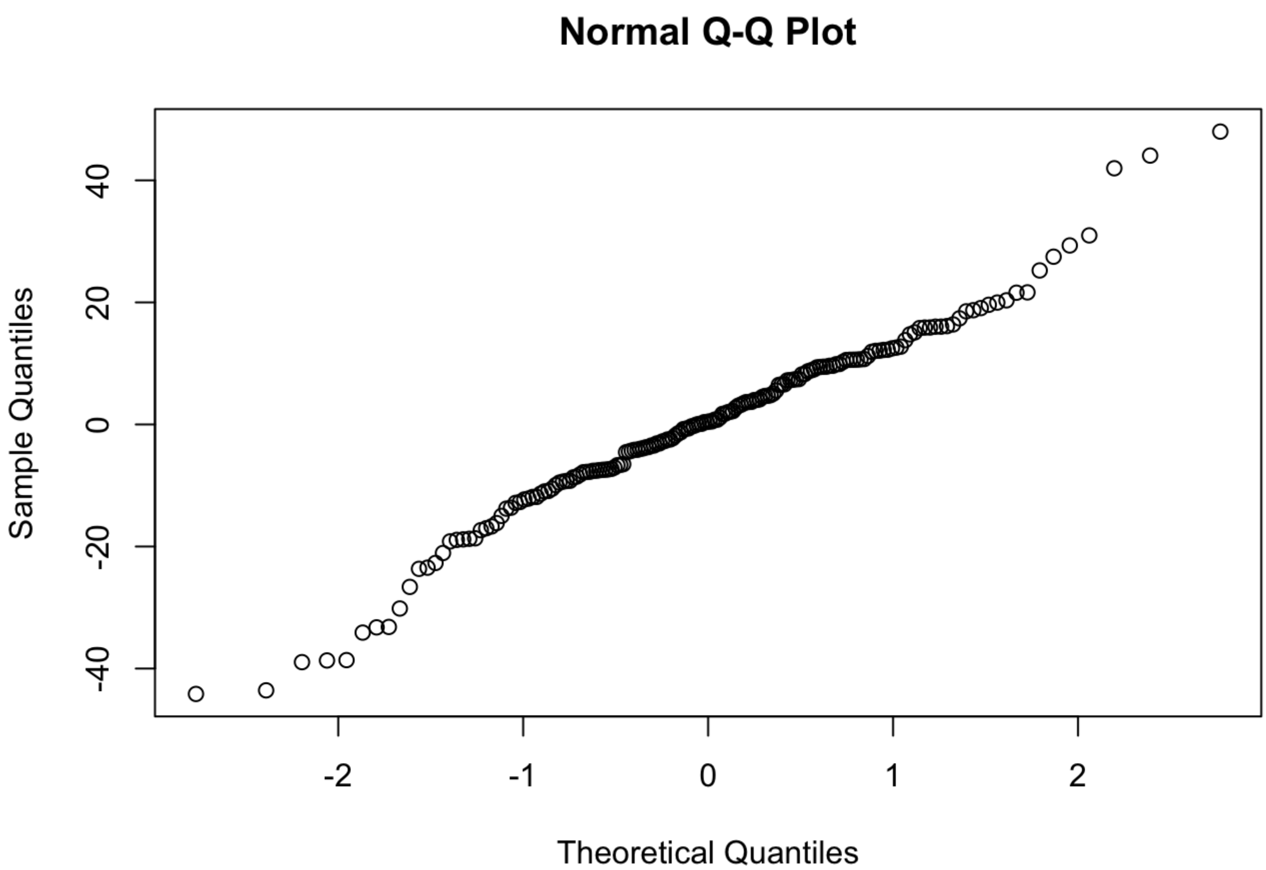
```
# we know that observations 8 and 57 are outliers from the previous assignment
sb_adj <- sb[-c(8, 57),]

model2.1 <- lmer(Reaction ~ Days+ I(Days^2) + (Days+ I(Days^2)|Subject), sb_adj,REML=
F)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0254985 (tol = 0.002, component 1)
```

```
plot(resid(model2.1) ~ fitted(model2.1),xlab="Fitted",ylab="Residuals")
abline(h=0)
```

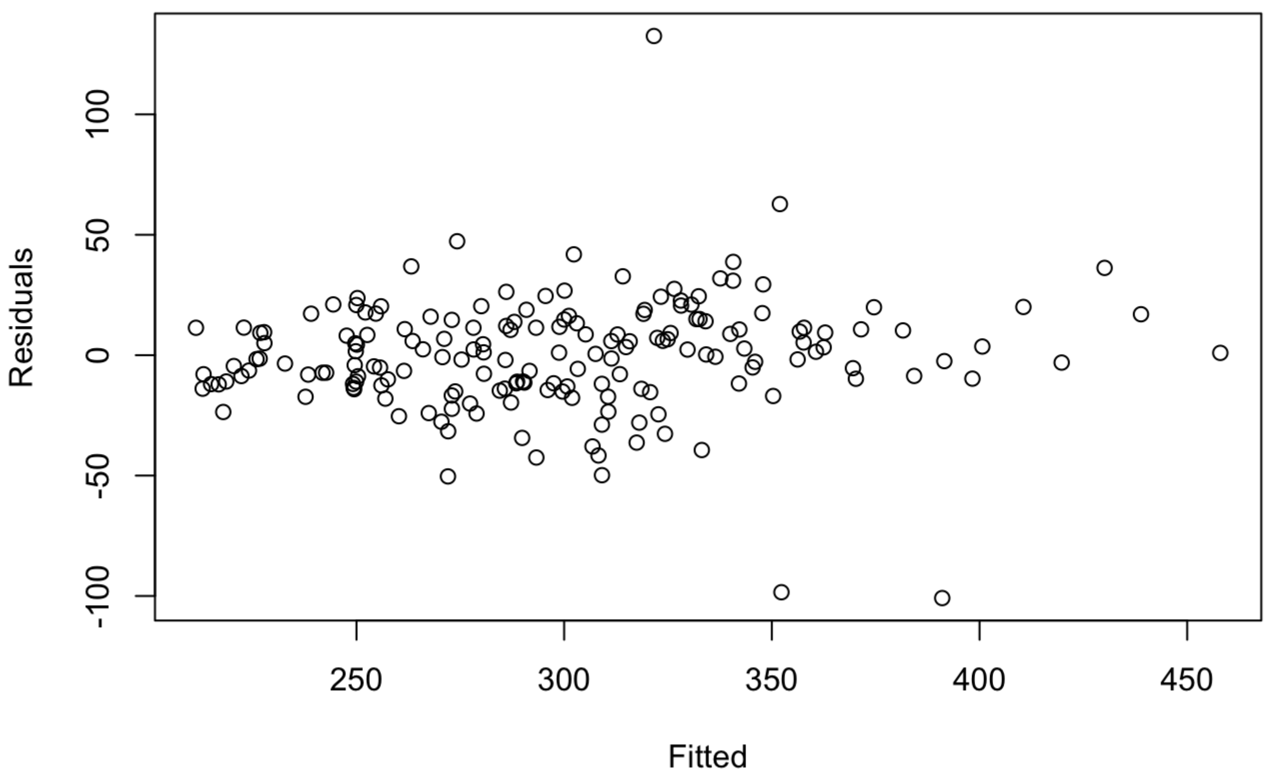


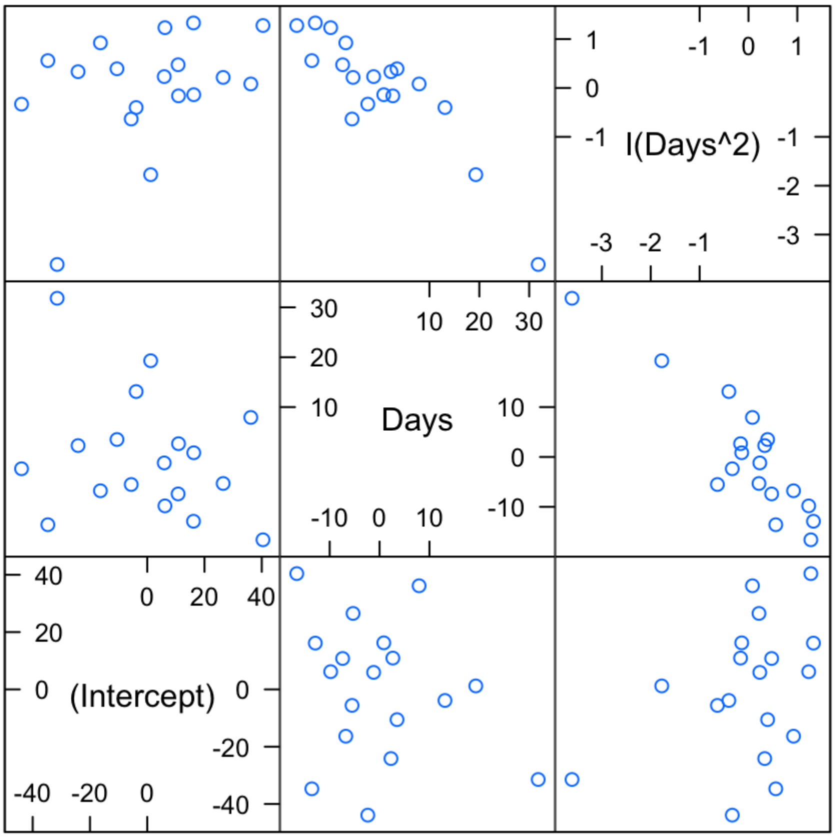

```
qqnorm(residuals(model2.1))
```



\ (f) Simulate the response under your first model and plot it. Does the simulated data look like the actual data? It seems actual data without some outlier as below graph.

```
plot(resid(model1) ~ fitted(model1),xlab="Fitted",ylab="Residuals")
```





Scatter Plot Matrix