HW 9

Ted Henson

4/13/2020

Chapter 11, questions 3 and 5, pages 252-254.

real page 268

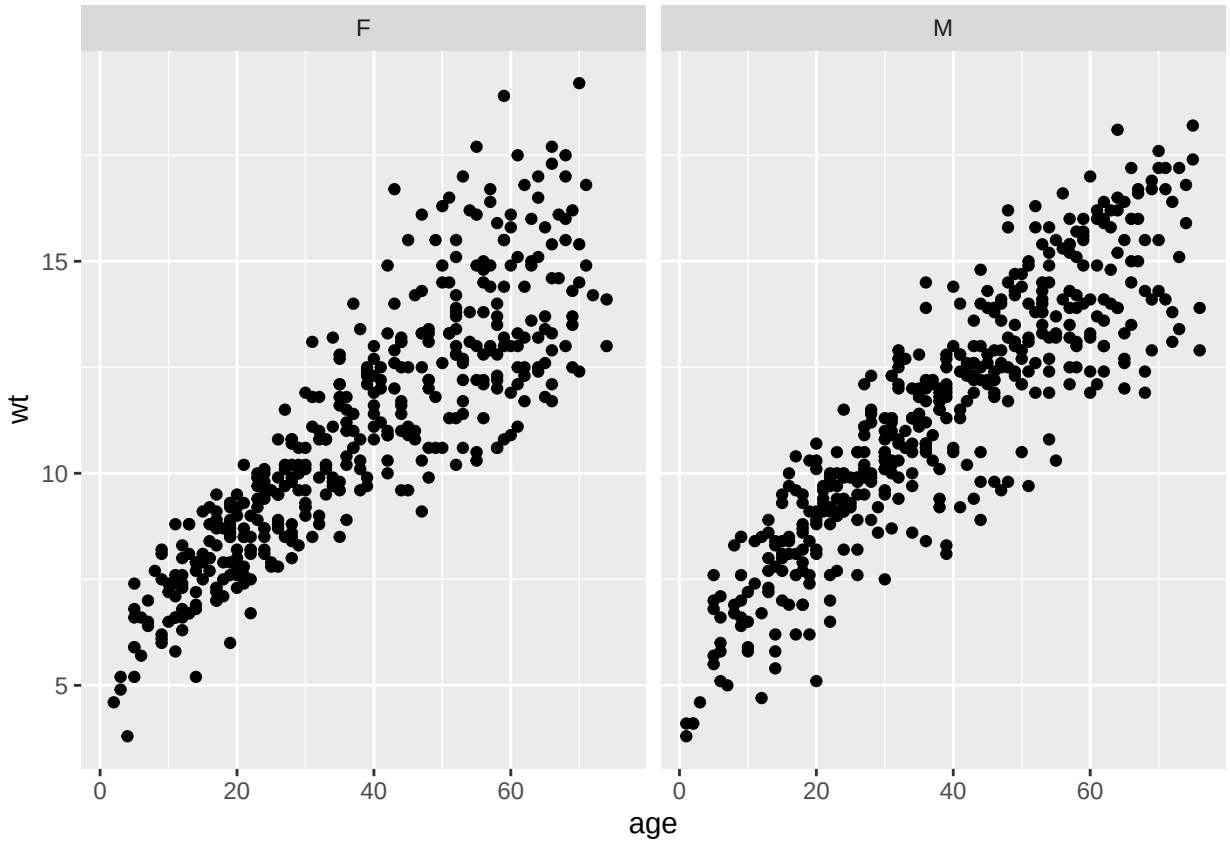
Question 3

a)

```
library(faraway)
library(lme4)
library(tidyverse)
data(nepali)
nepali = nepali %>% dplyr::select(-ht)
nepali = na.omit(nepali)
nepali$sex = ifelse(nepali$sex == 1, 'M', 'F')
head(nepali)
```

```
##       id sex   wt mage lit died alive age
## 1 120011   M 12.8   35   0    2     5  41
## 2 120011   M 12.8   35   0    2     5  45
## 3 120011   M 13.1   35   0    2     5  49
## 4 120011   M 13.8   35   0    2     5  53
## 6 120012   F 14.9   35   0    2     5  57
## 7 120012   F 15.1   35   0    2     5  61
```

```
library(ggplot2)
ggplot(nepali, aes(x=age, y = wt)) + geom_point() + facet_wrap(. ~ sex)
```



Weight generally icreases with age for both males and females. Female weight may be more variable at higher ages.

b)

```
mod = lm(wt~ age+sex+mage+lit+died, data = nepali)
summary(mod)
```

```
##
## Call:
## lm(formula = wt ~ age + sex + mage + lit + died, data = nepali)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6949 -0.8440  0.0890  0.7969  4.9531
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.435972   0.237558  18.673  < 2e-16 ***
## age          0.136296   0.002580  52.836  < 2e-16 ***
## sexM         0.353890   0.093455   3.787 0.000163 ***
## mage         0.048187   0.008194   5.881  5.8e-09 ***
## lit          1.022090   0.264399   3.866 0.000119 ***
## died        -0.054271   0.050250  -1.080 0.280435
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.37 on 871 degrees of freedom
## Multiple R-squared:  0.7843, Adjusted R-squared:  0.7831
## F-statistic: 633.5 on 5 and 871 DF,  p-value: < 2.2e-16
```

All terms except for the died variable appear to be significant in this model.

c)

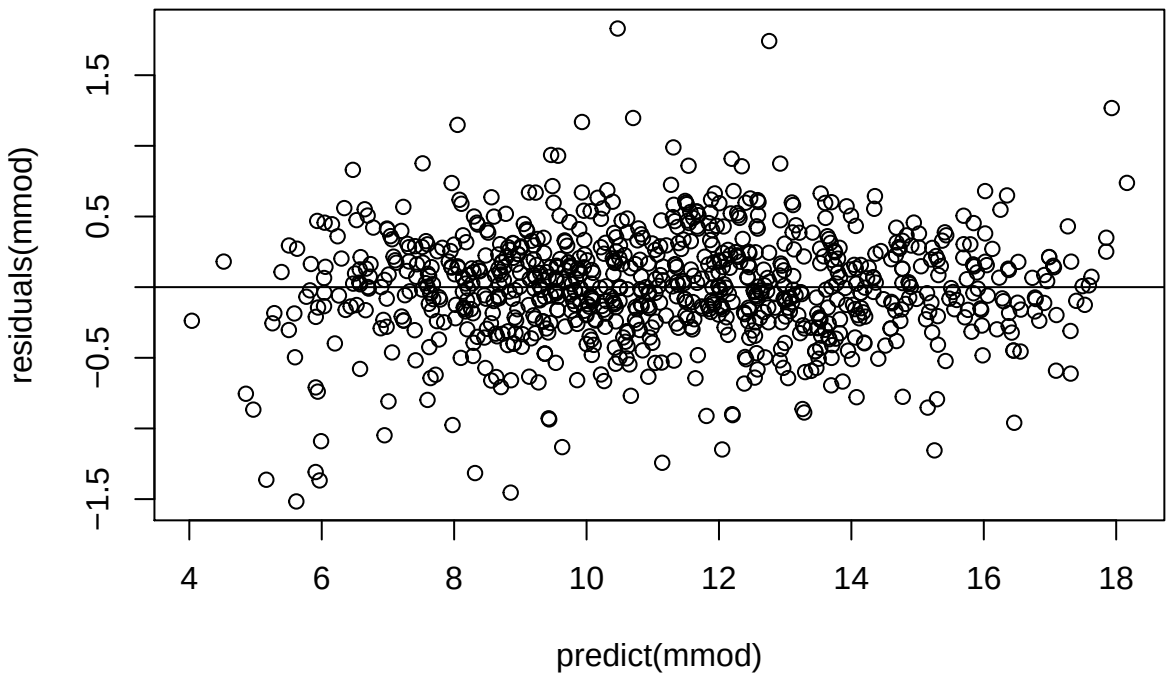
```
mmod=lmer(wt~ age*sex + lit + mage + (1|id),nepali)
summary(mmod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: wt ~ age * sex + lit + mage + (1 | id)
##    Data: nepali
##
## REML criterion at convergence: 1789.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4700 -0.4886  0.0007  0.5198  4.1880
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept) 1.766    1.329
##  Residual             0.191    0.437
## Number of obs: 877, groups:  id, 197
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 4.3812771  0.4859578   9.016
## age         0.1338857  0.0034868  38.398
## sexM        0.3698066  0.2640962   1.400
## lit         0.7560264  0.4716539   1.603
## mage        0.0508327  0.0157632   3.225
## age:sexM    0.0007549  0.0048474   0.156
##
## Correlation of Fixed Effects:
##          (Intr) age    sexM   lit    mage
## age      -0.182
## sexM     -0.304  0.478
## lit      -0.195 -0.010  0.063
## mage     -0.919 -0.086  0.024  0.145
## age:sexM  0.185 -0.714 -0.682  0.005  0.003
```

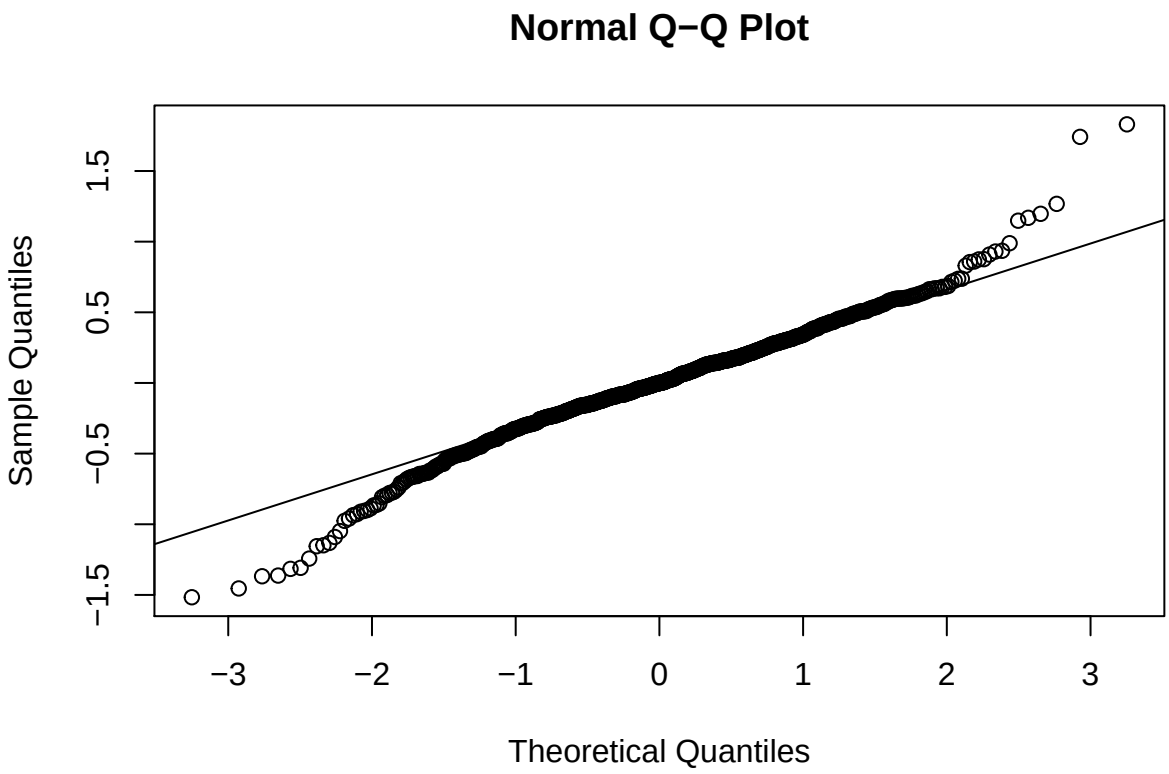
The predicted difference in child weight for a mother at age 25 compared to 15 would be 0.508327 (10*.0508327). The expected difference between identical twins would be .437, the standard deviation within a given id. This seems reasonable as identical twins often have almost identical physical features, including weight, due to their identical genetic makeup, but minor environmental differences cause minor differences in physical features, such as weight.

d)

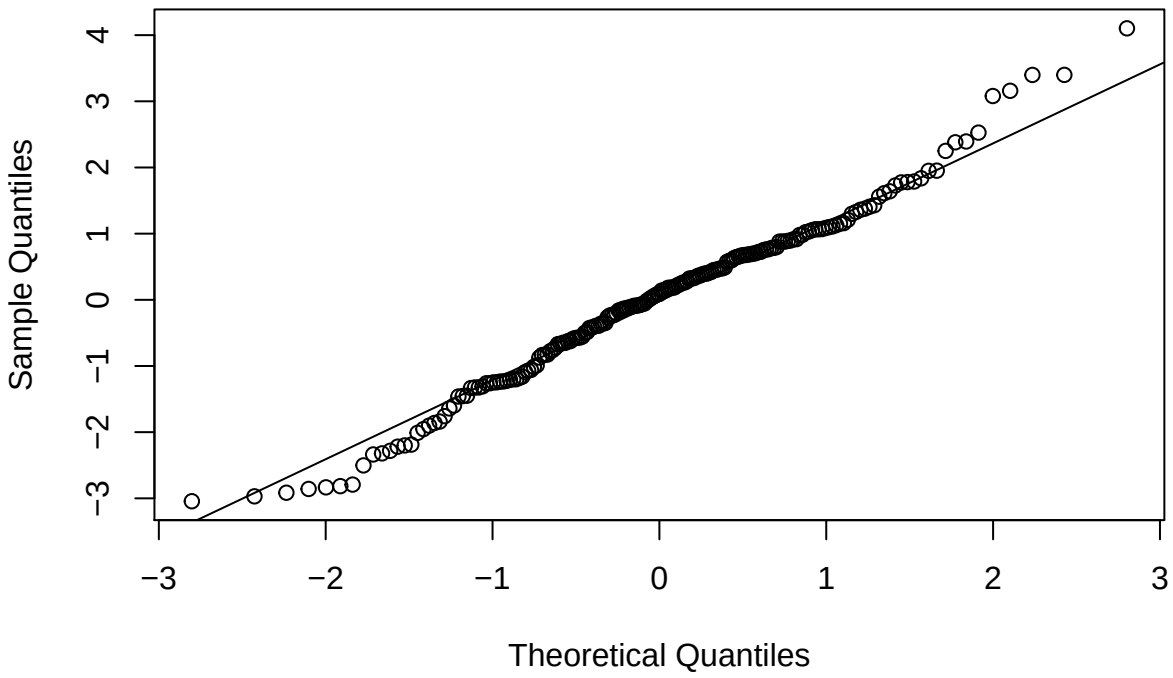
```
plot(predict(mmod), residuals(mmod))
abline(a = 0, b=0)
```




```
qqnorm(residuals(mmod))
qqline(residuals(mmod))
```



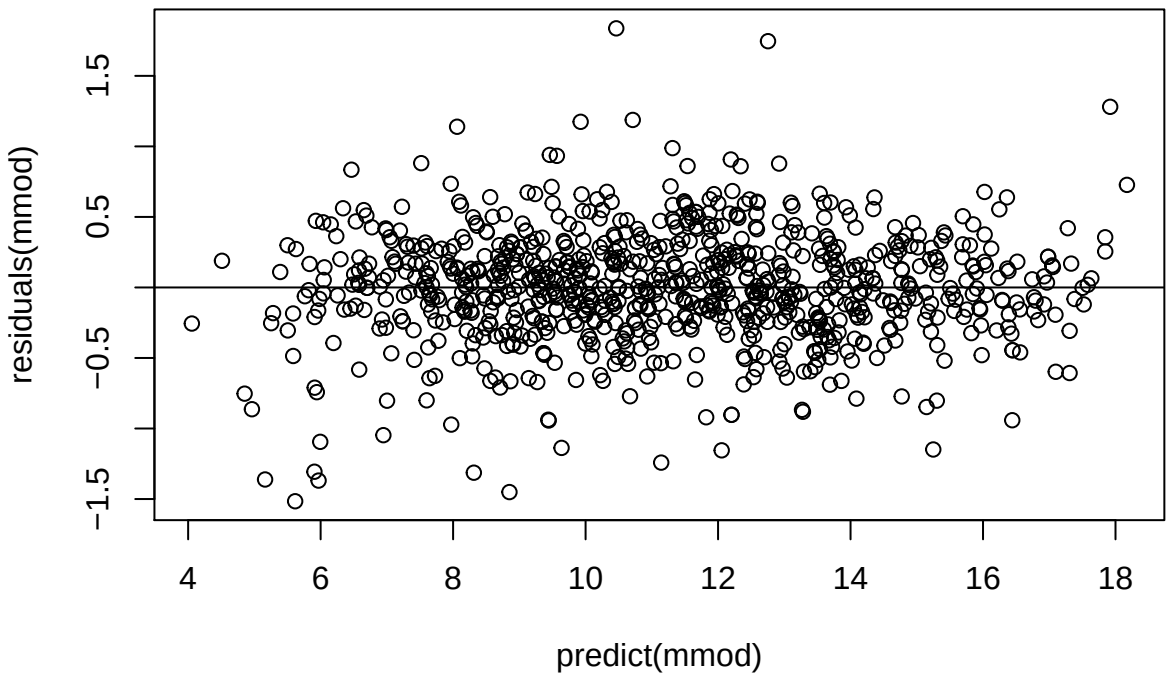

```
qqnorm(ranef(mmod)$"id"[[1]],main='')
qqline(ranef(mmod)$"id"[[1]],main='')
```



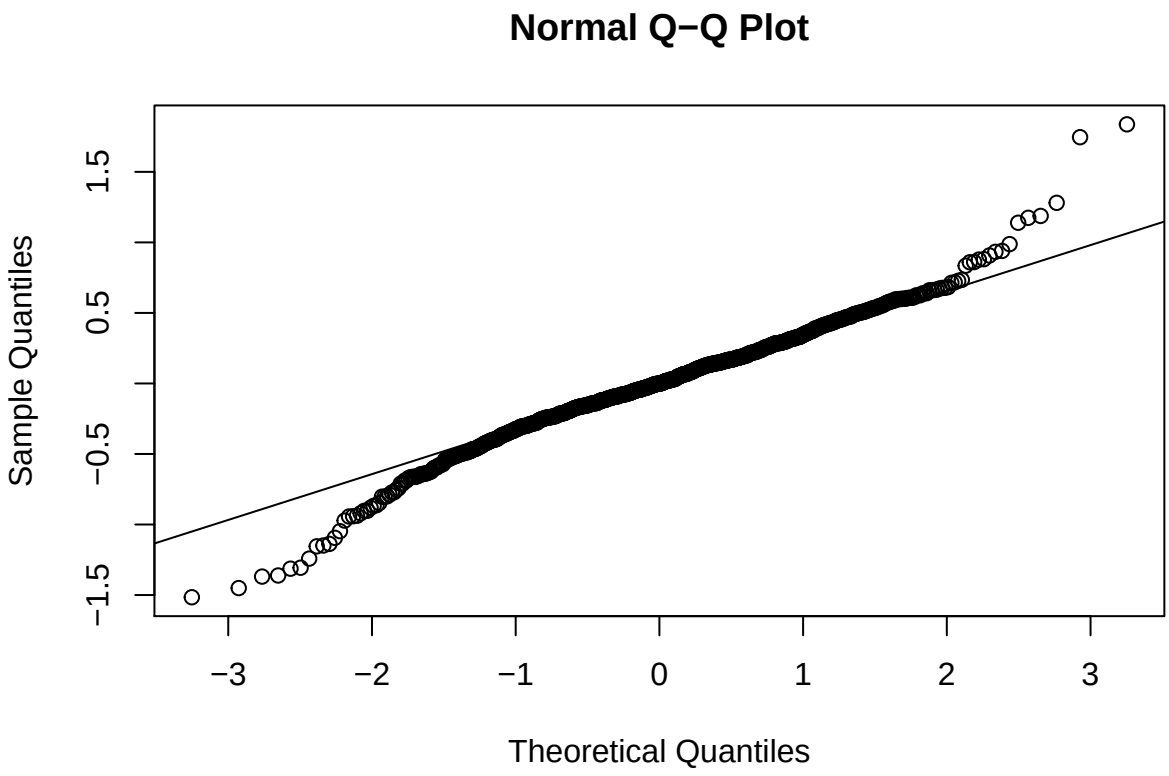
The residuals appear to be fairly normal, but slightly more negative towards the lower quantiles and more positive in the upper quantiles. The qqnorm plot of the random effects of the id has a similar trend.

e)

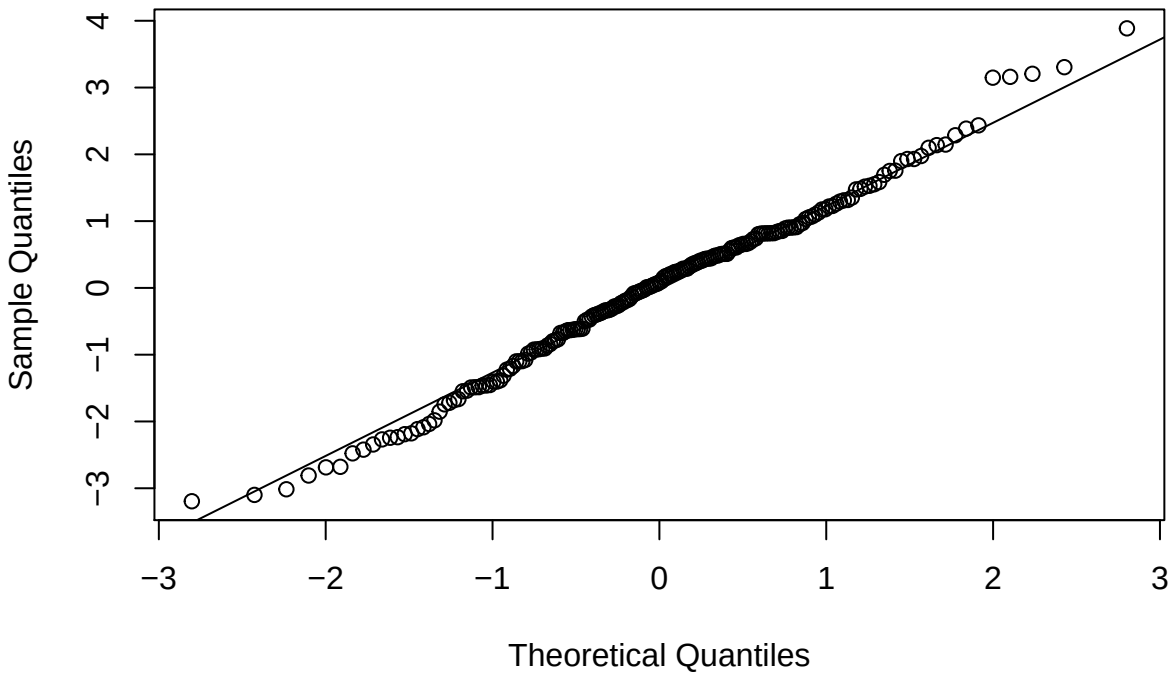
```
mmod=lmer(wt~ age + mage + (1|id),nepali)
plot(predict(mmod), residuals(mmod))
abline(a = 0, b=0)
```




```
qqnorm(residuals(mmod))
qqline(residuals(mmod))
```




```
qqnorm(ranef(mmod)$"id"[[1]],main='')
qqline(ranef(mmod)$"id"[[1]],main='')
```



The diagnostic plots show almost no improvement by including more terms as was done previously so the simplified model is most likely better.

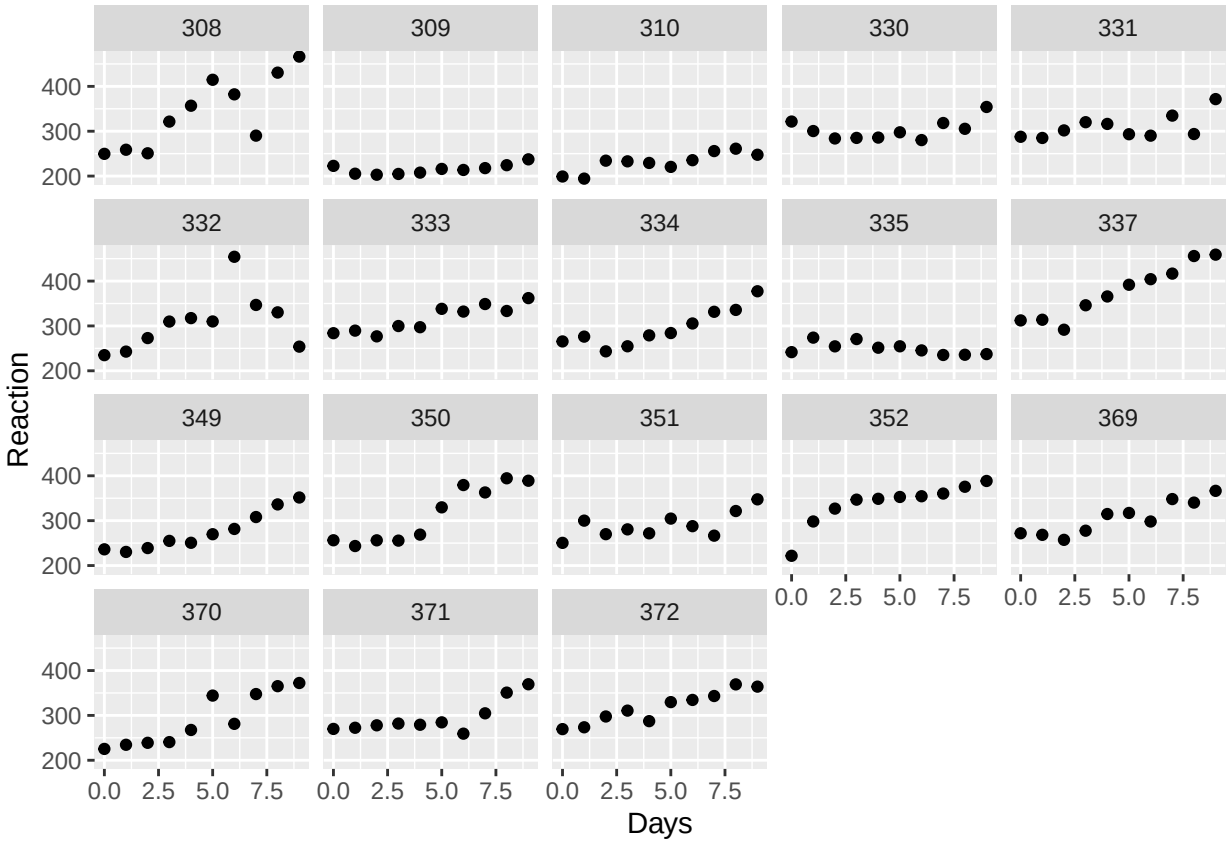
f)

g)

Question 5

a)

```
data(sleepstudy)
ggplot(sleepstudy, aes(x=Days, y = Reaction)) + geom_point() + facet_wrap(. ~ Subject)
```



The reaction generally increases with the number of days, but the strength and variability of the increase vary widely between subjects.

b)

```
mmod = lmer(Reaction ~ Days +  (1+Days|Subject),
            data = sleepstudy)
summary(mmod)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ Days + (1 + Days | Subject)
##    Data: sleepstudy
##
## REML criterion at convergence: 1743.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9536 -0.4634  0.0231  0.4633  5.1793
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 611.90   24.737
##           Days         35.08    5.923   0.07
##  Residual             654.94   25.592
## Number of obs: 180, groups:  Subject, 18
##
```

```
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  251.405      6.824  36.843
## Days          10.467      1.546   6.771
##
## Correlation of Fixed Effects:
##      (Intr)
## Days -0.138
```

Under this model it would be unusual for a subject's reaction time to decrease over time.

c)

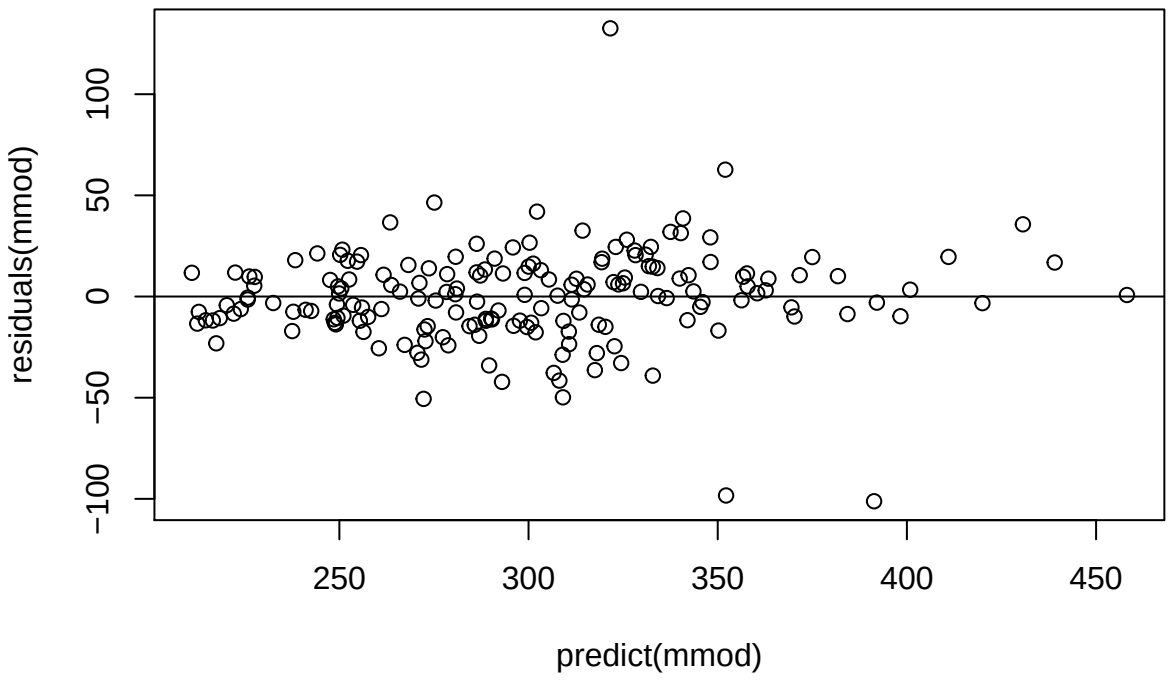
```
mmod2 = lmer(Reaction ~ Days + I(Days^2)+ (1+Days|Subject),
            data = sleepstudy)
summary(mmod2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ Days + I(Days^2) + (1 + Days | Subject)
##    Data: sleepstudy
##
## REML criterion at convergence: 1742.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0093 -0.4489  0.0422  0.5036  5.2702
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 613.12   24.761
##           Days         35.11    5.925   0.06
##  Residual             651.97   25.534
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 255.4494     7.5135  33.999
## Days          7.4341     2.8189   2.637
## I(Days^2)     0.3370     0.2619   1.287
##
## Correlation of Fixed Effects:
##           (Intr) Days
## Days      -0.418
## I(Days^2)  0.418 -0.836
```

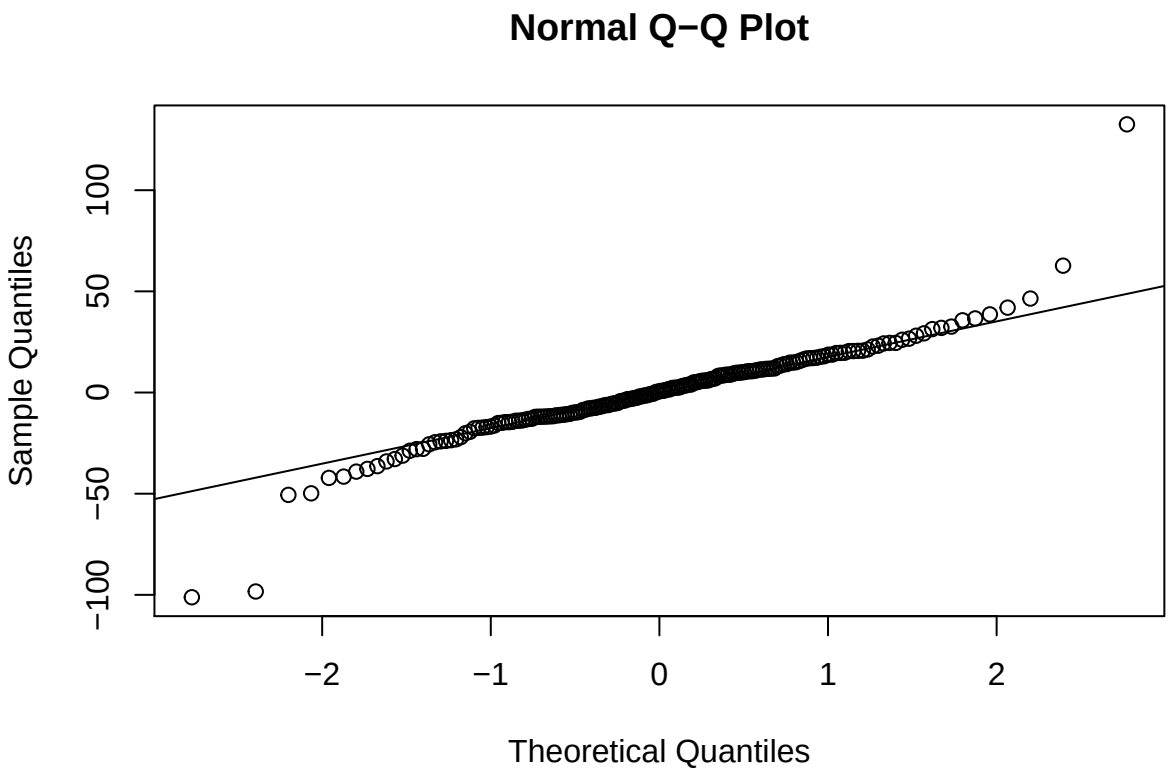
The model had a high standard error and low t value for the quadratic term.

d)

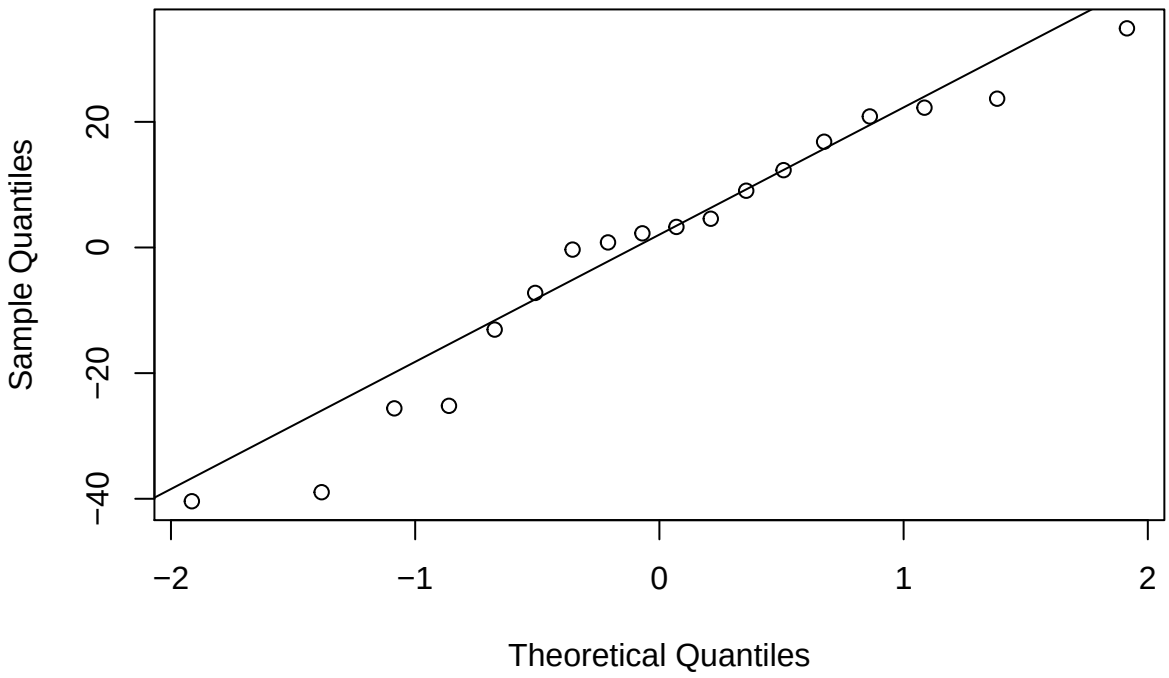
```
plot(predict(mmod), residuals(mmod))
abline(a = 0, b=0)
```



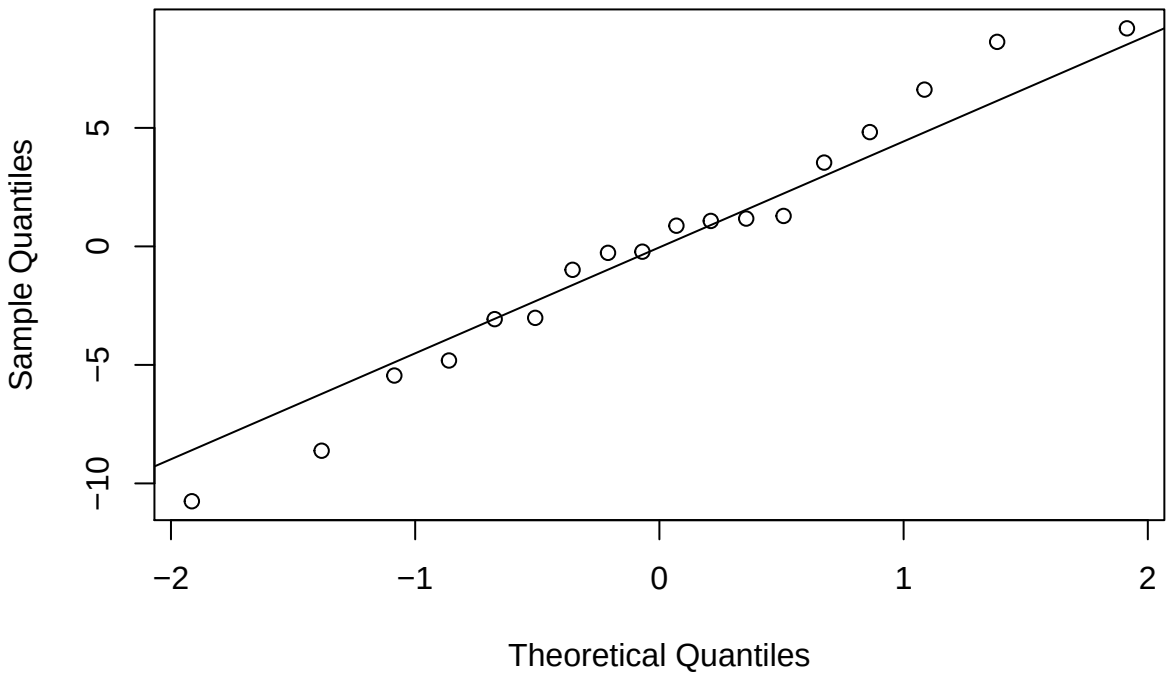

```
qqnorm(residuals(mmod))
qqline(residuals(mmod))
```



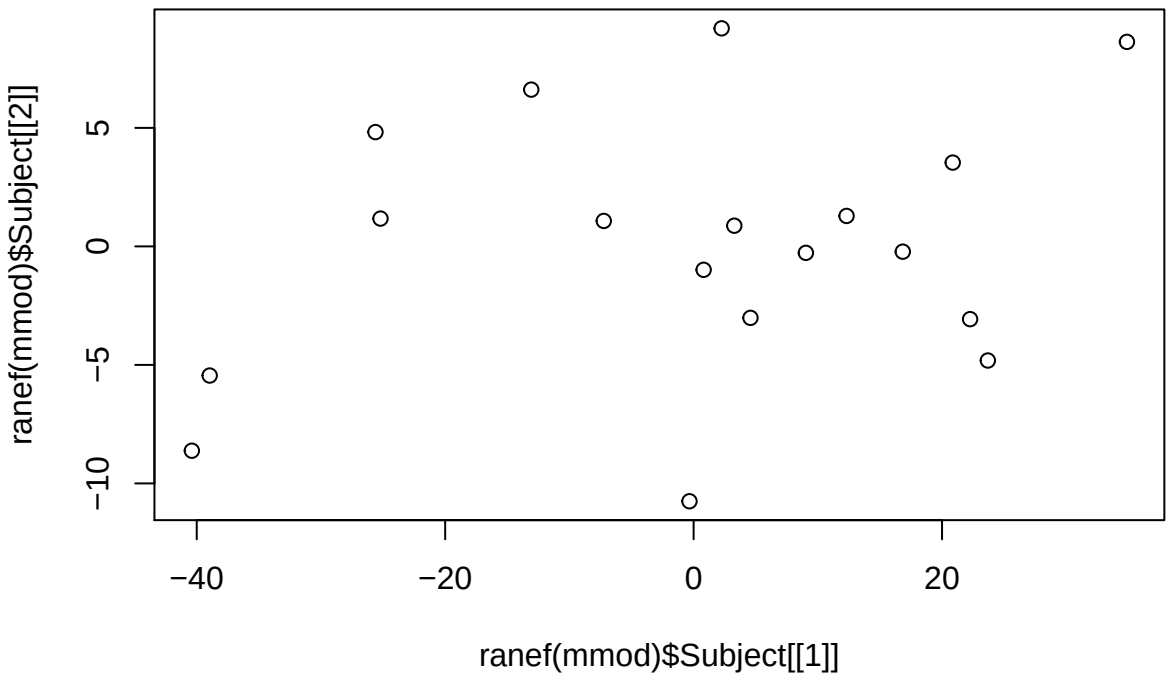

```
qqnorm(ranef(mmod)$"Subject"[[1]],main='')
qqline(ranef(mmod)$"Subject"[[1]],main='')
```




```
qqnorm(ranef(mmod)$"Subject"[[2]],main='')
qqline(ranef(mmod)$"Subject"[[2]],main='')
```




```
plot(ranef(mmod)$"Subject"[[1]], ranef(mmod)$"Subject"[[2]])
```



The residuals appear to be fairly normal and centered around zero, although there are a few potential outliers. The random effects of the intercept and slope of the days appear to be somewhat correlated.

e)

```
dat = sleepstudy[which(abs(residuals(mmod)) < 98),]
mmod3 = lmer(Reaction ~ Days +  (1+Days|Subject),
            data = dat)
summary(mmod3)
```

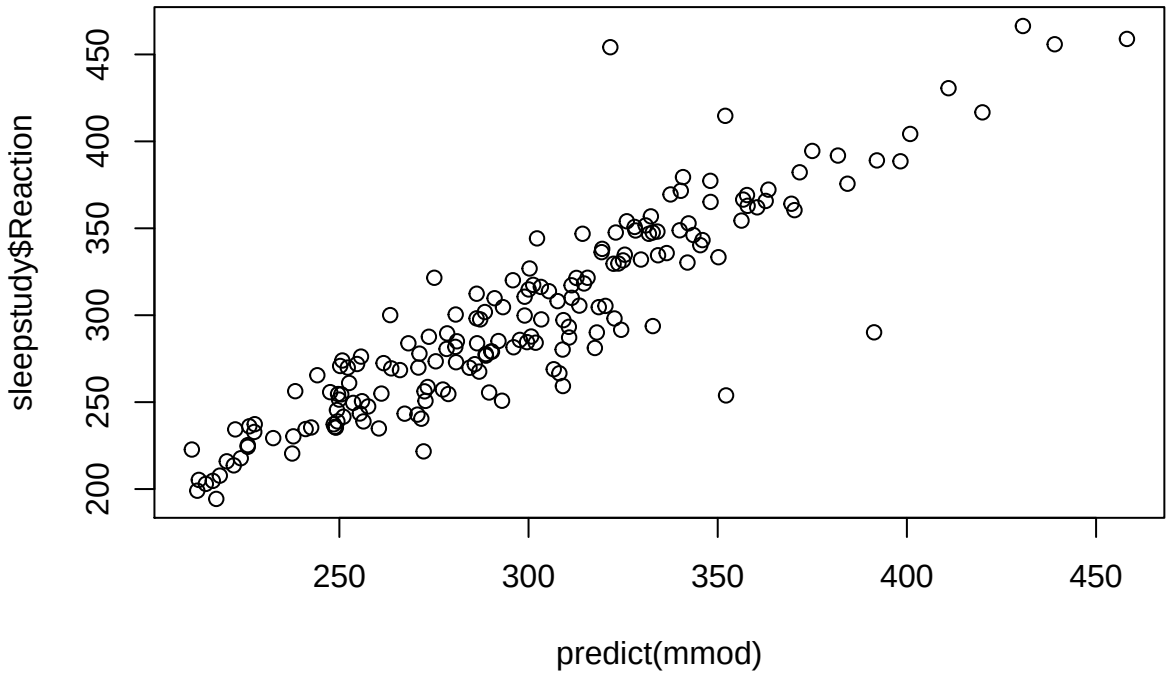
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ Days + (1 + Days | Subject)
##    Data: dat
##
## REML criterion at convergence: 1638.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.67893 -0.55470 -0.00955  0.56348  2.47740
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 705.27   26.557
##           Days         44.71    6.687   -0.06
##  Residual             374.16   19.343
## Number of obs: 177, groups:  Subject, 18
```

```
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  250.135      6.812  36.721
## Days          10.880      1.656   6.568
##
## Correlation of Fixed Effects:
##      (Intr)
## Days -0.157
```

The random intercept term for the subject had the largest increase. The random slope of the days had a small increase, and the fixed effects were almost the same as the initial model.

f)

```
plot(predict(mmod), sleepstudy$Reaction)
```



The predicted values are fairly close to the actual reaction values.