STA207 homework 6

Juanjuan Hu, Zhen Zhang

March 2, 2016

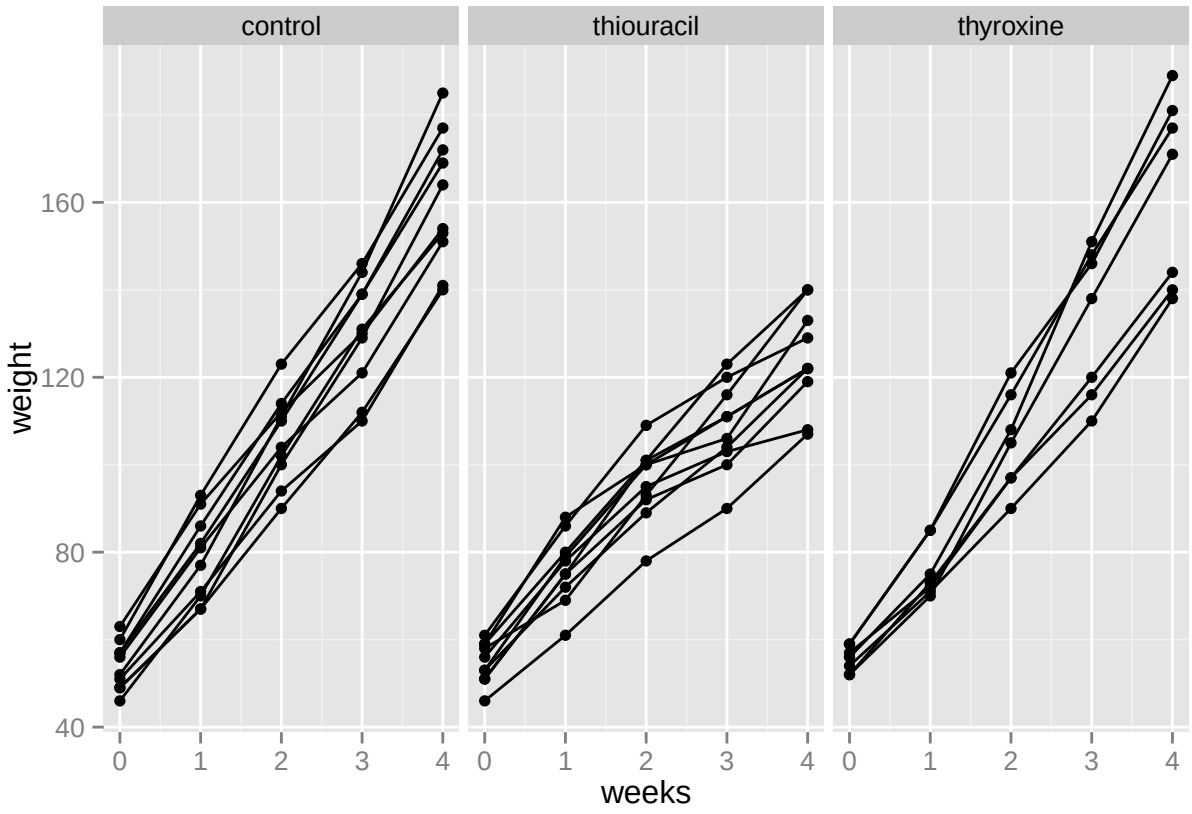
2

a

```
library(ggplot2)
library(lme4)
```

```
## Loading required package: Matrix
```

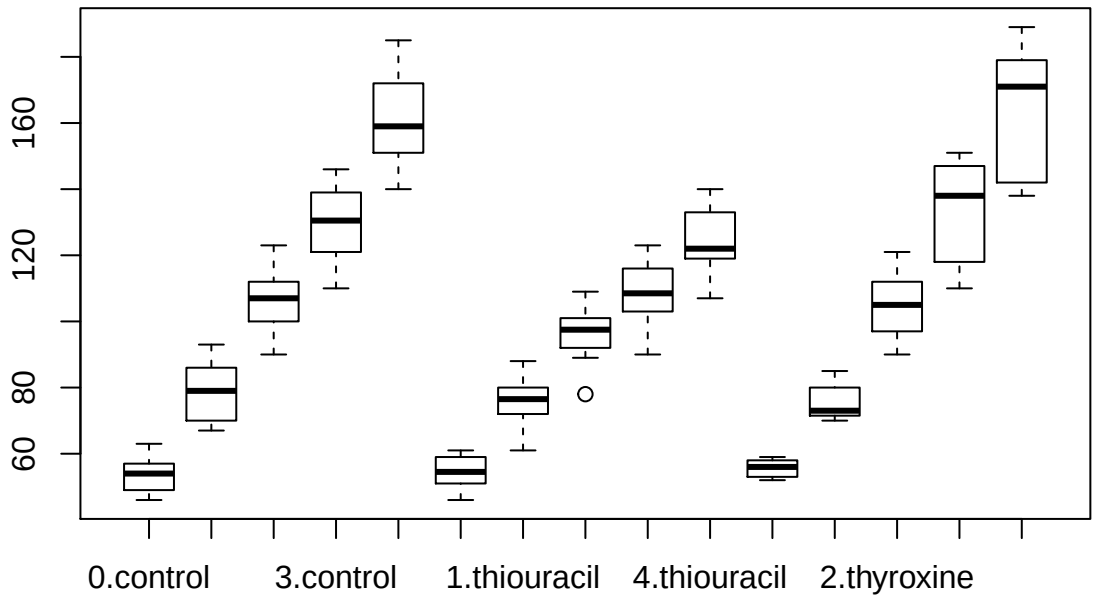
```
rat <- read.csv("ratdrink.csv")
ggplot(rat, aes(y = weight, x = weeks)) + facet_wrap(~ treat) + geom_point() + geom_line(aes(group = su
```



From the plot, it is easily seen that for the group thiouracil, the level of weight is smaller than the other two groups, and for the group thyroxine, some of the subjects' weight are more than that of the subjects in the control group, while some are less than that.

b

```
rat$subject <- as.factor(rat$subject)
rat$weeks <- as.factor(rat$weeks)
with(rat, boxplot(weight ~ weeks * treat))
```



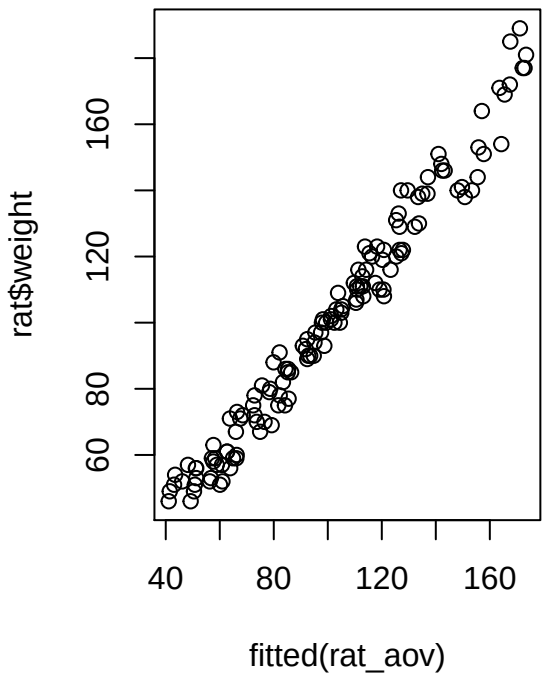
In terms of the mean, thyroxine is the biggest in the level of weight, followed by control group, with thiouracil the smallest level. This result is consistent with part a. For the variance, the big box of thyroxine indicates a larger variance, also the same with part a.

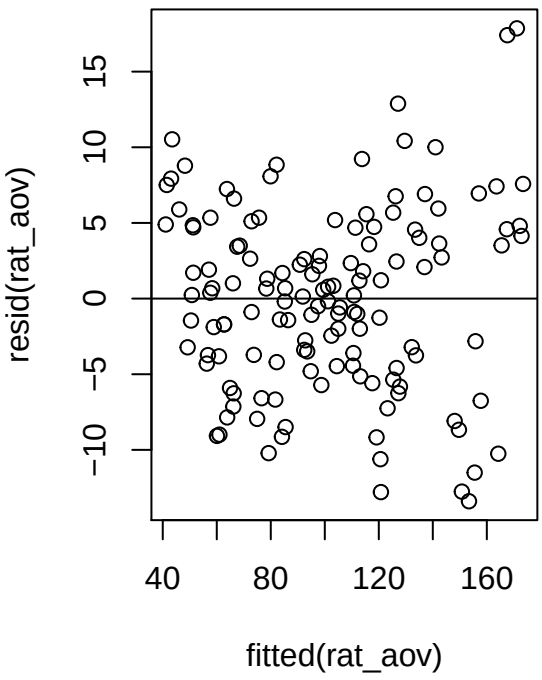
c

```
rat_aov <- lmer(weight ~ weeks * treat + (1|subject), data = rat)
```

The plot:

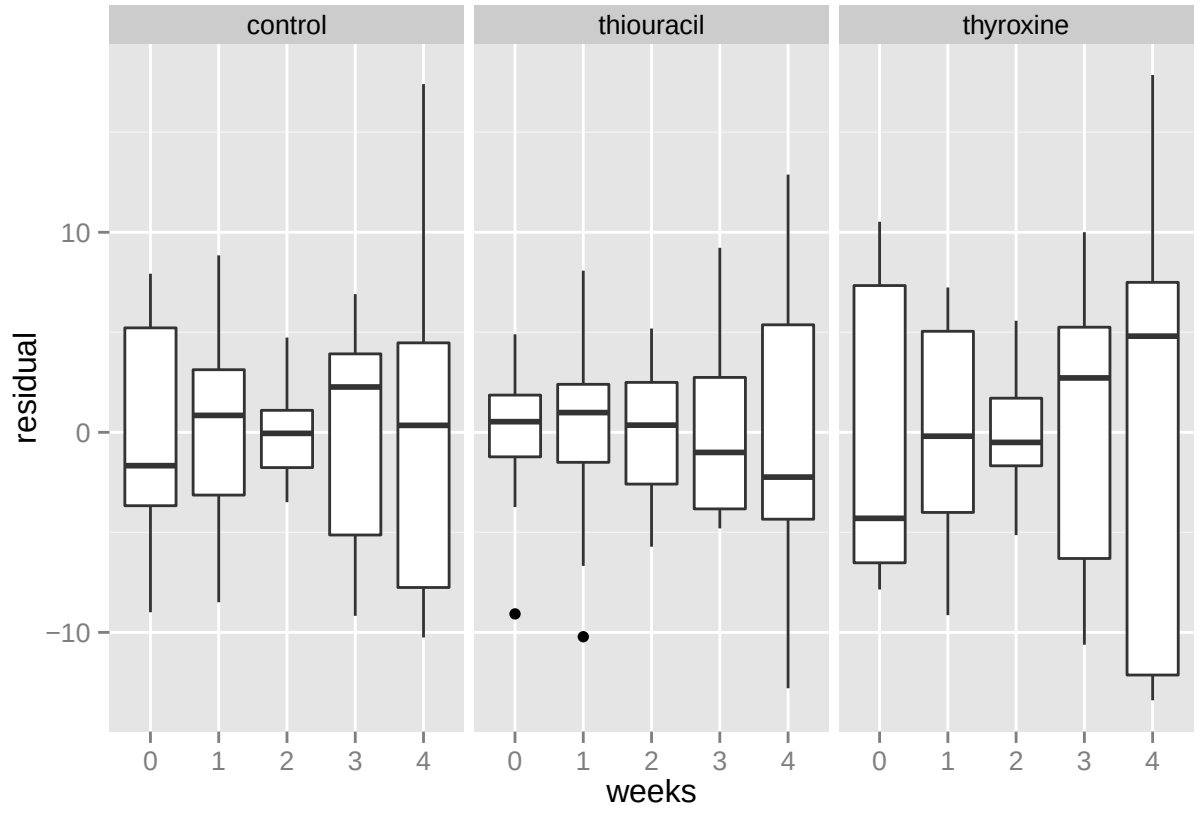
```
par(mfrow = c(1, 2))
plot(rat$weight ~ fitted(rat_aov))
plot(resid(rat_aov) ~ fitted(rat_aov))
abline(h = 0)
```





The boxplot:

```
rat_resid_dataframe <- cbind(rat, residual = resid(rat_aov))
ggplot(rat_resid_dataframe, aes(x = weeks, y = residual)) + facet_wrap(~treat) + geom_boxplot()
```



For the model plot, we can see that there is a straight line, which indicates a good fit. But in case of residual plot, we can see a trend of increasing variance. The boxplot here told us the almost the same pattern of residuals among different groups, fluctuate below and up 0. so this model is suitable.

d

Anova table:

```
anova(rat_aov)
```

```
## Analysis of Variance Table
##             Df Sum Sq Mean Sq  F value
## weeks        4 145188   36297 733.0981
## treat        2    770     385   7.7774
## weeks:treat  8   6403     800  16.1641
```

Parameter estimates and standard errors:

```
summary(rat_aov)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ weeks * treat + (1 | subject)
##    Data: rat
##
```

```
## REML criterion at convergence: 892.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.90369 -0.60408  0.03345  0.64957  2.53828
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 71.55    8.459
##  Residual             49.51    7.036
## Number of obs: 135, groups:  subject, 27
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)               54.0000     3.4794   15.52
## weeks1                    24.5000     3.1468    7.79
## weeks2                    52.0000     3.1468   16.52
## weeks3                    76.1000     3.1468   24.18
## weeks4                   106.6000     3.1468   33.88
## treatthiouracil            0.7000     4.9206    0.14
## treatthyroxine             1.5714     5.4222    0.29
## weeks1:treatthiouracil    -2.9000     4.4503   -0.65
## weeks2:treatthiouracil   -10.9000     4.4503   -2.45
## weeks3:treatthiouracil   -22.4000     4.4503   -5.03
## weeks4:treatthiouracil   -37.1000     4.4503   -8.34
## weeks1:treatthyroxine     -4.2143     4.9039   -0.86
## weeks2:treatthyroxine     -2.7143     4.9039   -0.55
## weeks3:treatthyroxine      1.0429     4.9039    0.21
## weeks4:treatthyroxine      0.6857     4.9039    0.14
##
## Correlation of Fixed Effects:
##             (Intr) weeks1 weeks2 weeks3 weeks4 trtthr trtthy wks1:trtthr
## weeks1      -0.452
## weeks2      -0.452  0.500
## weeks3      -0.452  0.500  0.500
## weeks4      -0.452  0.500  0.500  0.500
## treatthircl -0.707  0.320  0.320  0.320  0.320
## treatthyrxn -0.642  0.290  0.290  0.290  0.290  0.454
## wks1:trtthr  0.320 -0.707 -0.354 -0.354 -0.354 -0.452 -0.205
## wks2:trtthr  0.320 -0.354 -0.707 -0.354 -0.354 -0.452 -0.205  0.500
## wks3:trtthr  0.320 -0.354 -0.354 -0.707 -0.354 -0.452 -0.205  0.500
## wks4:trtthr  0.320 -0.354 -0.354 -0.354 -0.707 -0.452 -0.205  0.500
## wks1:trtthy  0.290 -0.642 -0.321 -0.321 -0.321 -0.205 -0.452  0.454
## wks2:trtthy  0.290 -0.321 -0.642 -0.321 -0.321 -0.205 -0.452  0.227
## wks3:trtthy  0.290 -0.321 -0.321 -0.642 -0.321 -0.205 -0.452  0.227
## wks4:trtthy  0.290 -0.321 -0.321 -0.321 -0.642 -0.205 -0.452  0.227
##             wks2:trtthr wks3:trtthr wks4:trtthr wks1:trtthy wks2:trtthy
## weeks1
## weeks2
## weeks3
## weeks4
## treatthircl
## treatthyrxn
## wks1:trtthr
```

```
## wks2:trtthr
## wks3:trtthr  0.500
## wks4:trtthr  0.500       0.500
## wks1:trtthy  0.227       0.227       0.227
## wks2:trtthy  0.454       0.227       0.227       0.500
## wks3:trtthy  0.227       0.454       0.227       0.500       0.500
## wks4:trtthy  0.227       0.227       0.454       0.500       0.500
##             wks3:trtthy
## weeks1
## weeks2
## weeks3
## weeks4
## treatthircl
## treatthyrxn
## wks1:trtthr
## wks2:trtthr
## wks3:trtthr
## wks4:trtthr
## wks1:trtthy
## wks2:trtthy
## wks3:trtthy
## wks4:trtthy  0.500
```

Stepwise:

```
first_set <- c("weeks", "treat", "weeks:treat")
sapply(first_set, function(x) {
  AIC(lmer(as.formula(sprintf("weight~%s + (1|subject)", x)), data = rat))
})
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
##       weeks       treat weeks:treat
##   1052.7907   1338.1493    926.1398
```

```
AIC(aov(weight ~ weeks * treat, data = rat))
```

```
## [1] 1046.712
```

So we should select "weeks:treat" first, with "1|subject". Now the second step:

```
second_set <- c("weeks", "treat", "weeks:treat")
sapply(second_set, function(x) {
  AIC(lmer(as.formula(sprintf("weight~weeks:treat + %s + (1|subject)", x)), data = rat))
})
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

```
##       weeks       treat weeks:treat
##    926.1398    926.1398    926.1398
```

```
AIC(aov(weight ~ weeks:treat + weeks, data = rat))
```

```
## [1] 1046.712
```

```
AIC(aov(weight ~ weeks:treat + treat, data = rat))
```

```
## [1] 1046.712
```

No improvement. So our final model becomes:

```
final1 <- lmer(weight ~ weeks:treat + (1|subject), data = rat)
```

```
## fixed-effect model matrix is rank deficient so dropping 1 column / coefficient
```

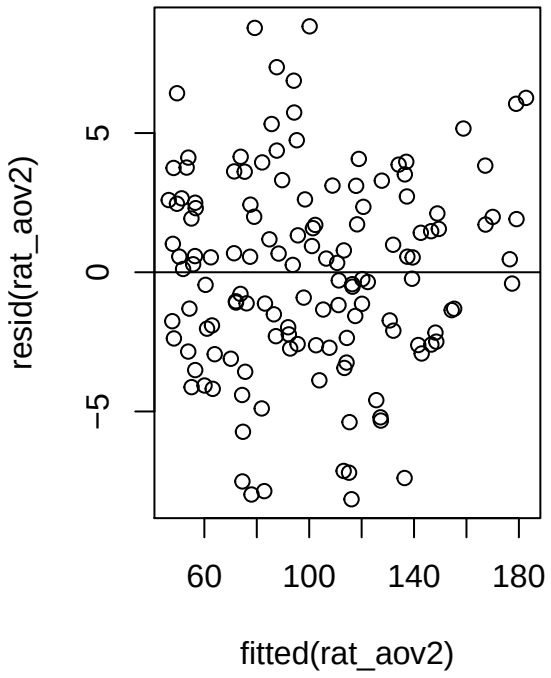
3

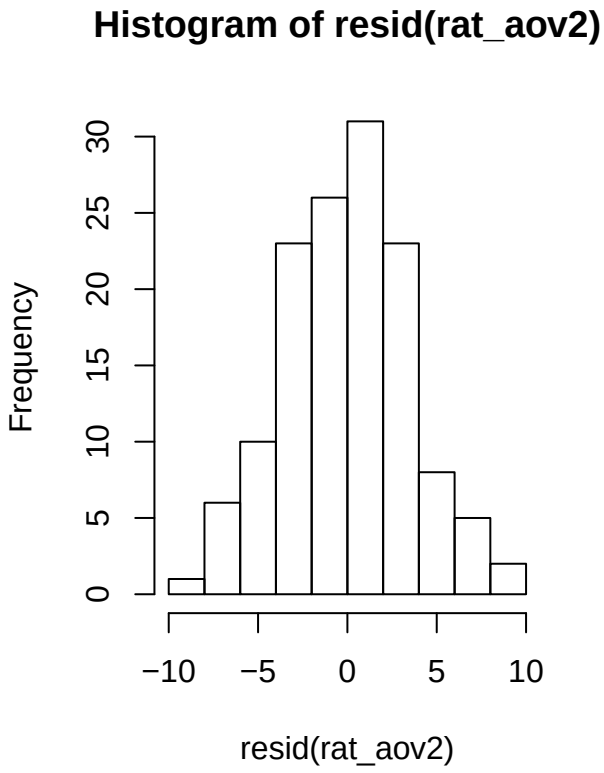
a

```
rat <- read.csv("ratdrink.csv")
rat$T <- rat$weeks
rat_aov2 = lmer(weight ~ (1|subject) + T + (0 + T|subject) + treat, data = rat)
```

plot:

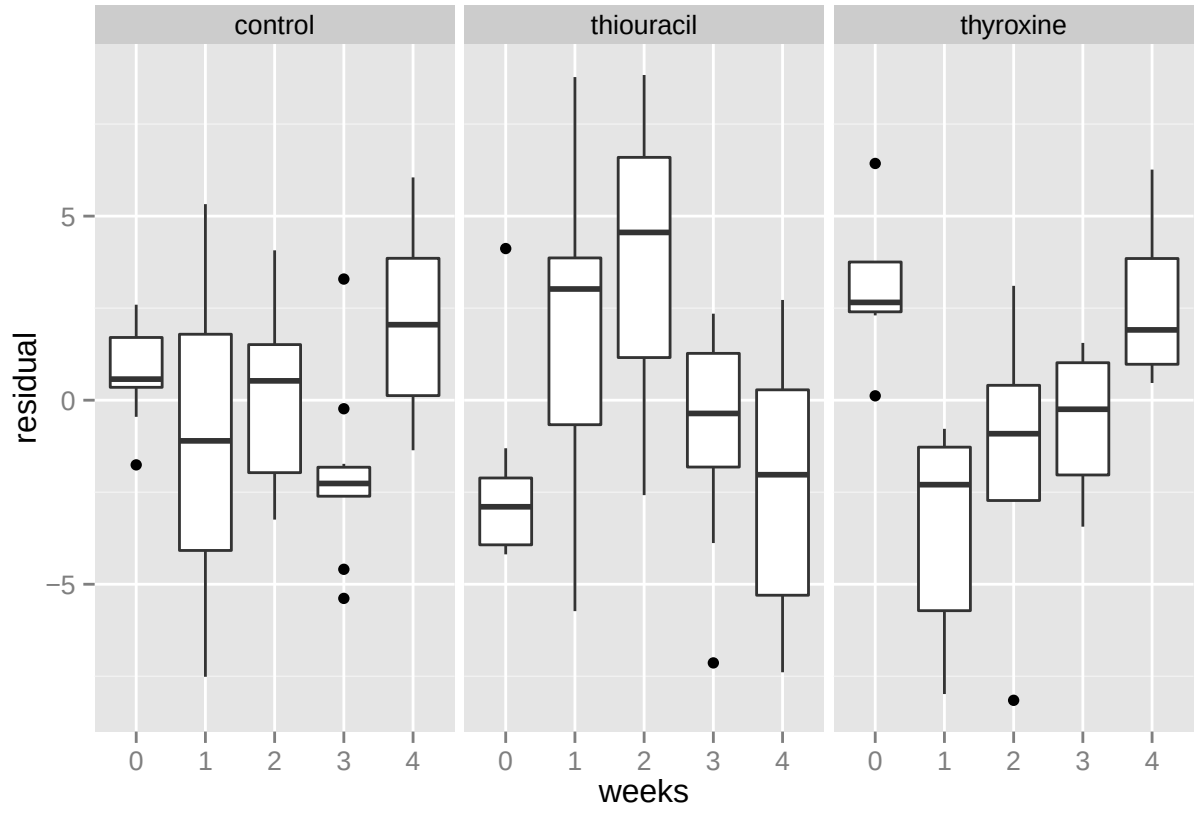
```
rat$subject <- as.factor(rat$subject)
rat$weeks <- as.factor(rat$weeks)
par(mfrow = c(1, 2))
plot(resid(rat_aov2) ~ fitted(rat_aov2))
abline(h = 0)
hist(resid(rat_aov2))
```





boxplot:

```
rat_resid_dataframe <- cbind(rat, residual = resid(rat_aov2))
ggplot(rat_resid_dataframe, aes(x = weeks, y = residual)) + facet_wrap(~treat) + geom_boxplot()
```



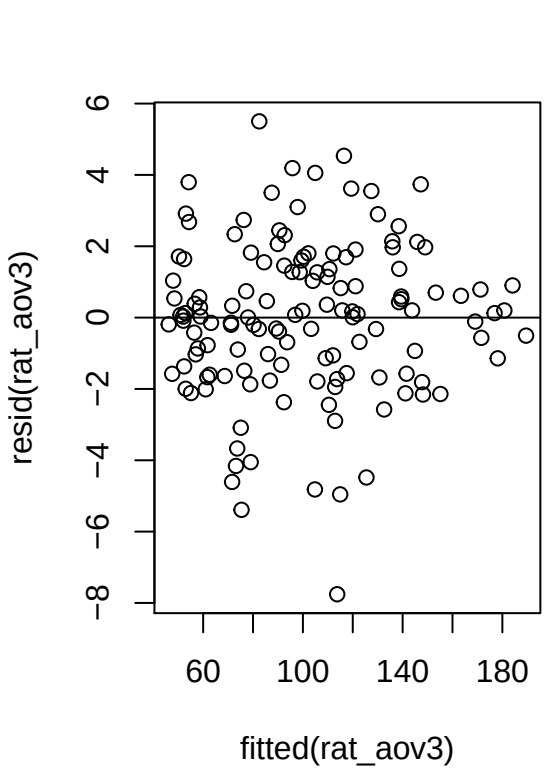
We can see the equal variance from fitted value versus residual plot, and is normal distributed residuals, and different residual pattern different control groups.

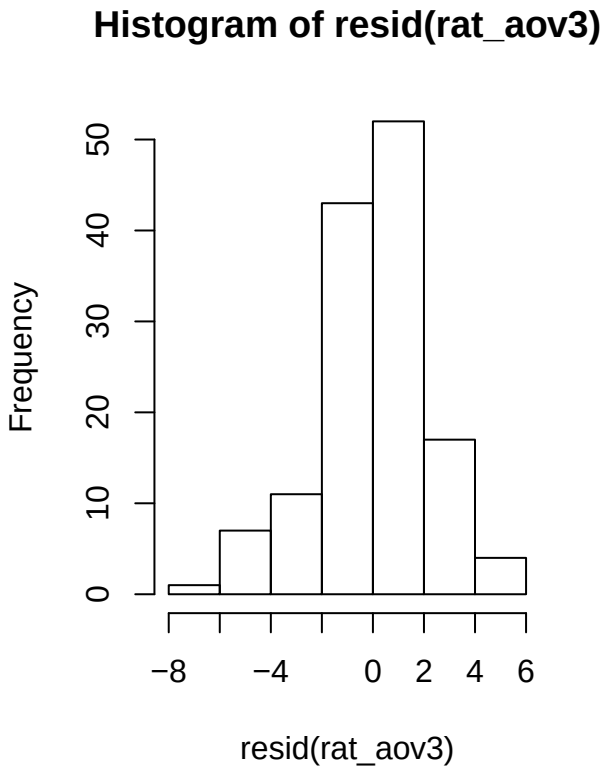
b

```
rat$T2 <- (rat$T)^2
rat_aov3 = lmer(weight ~ (1|subject) + T + T2 + (0 + T|subject) + (0 + T2|subject) + treat, data = rat)
```

plot:

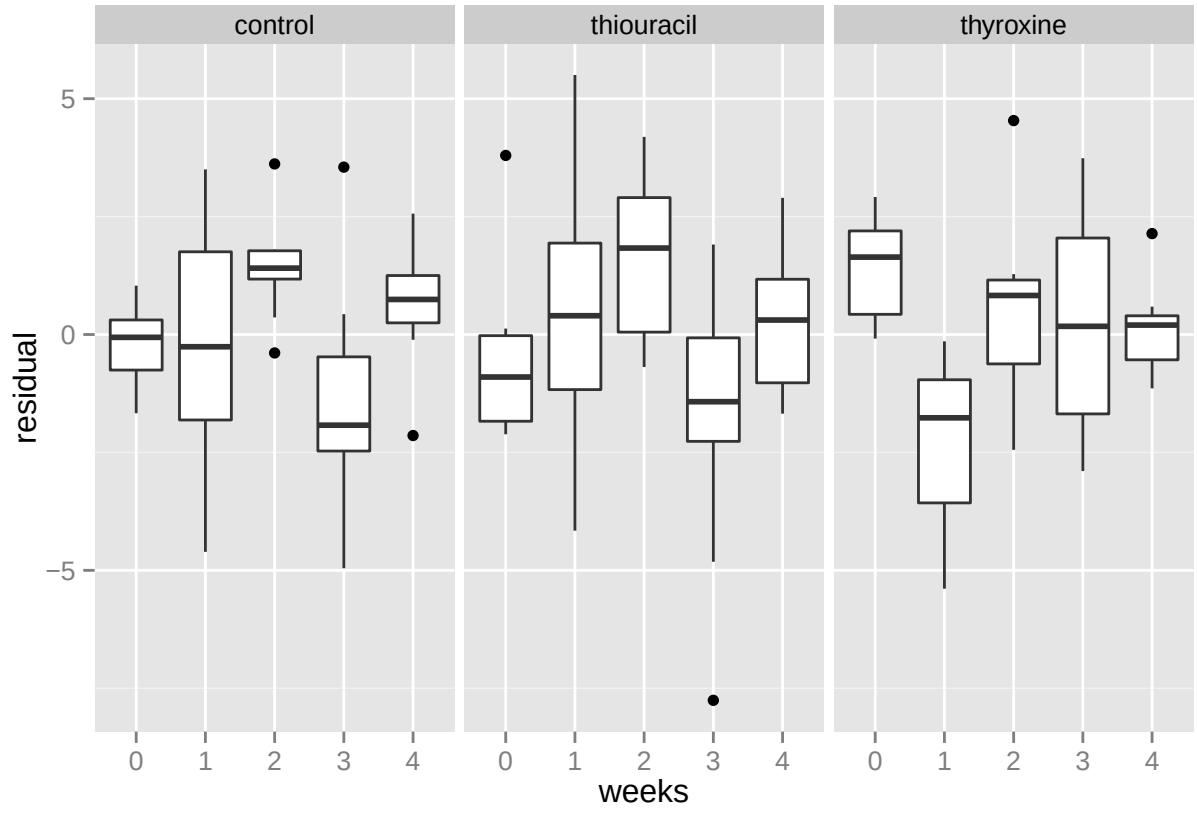
```
par(mfrow = c(1, 2))
plot(resid(rat_aov3) ~ fitted(rat_aov3))
abline(h = 0)
hist(resid(rat_aov3))
```





boxplot:

```
rat_resid_dataframe <- cbind(rat, residual = resid(rat_aov3))
ggplot(rat_resid_dataframe, aes(x = weeks, y = residual)) + facet_wrap(~treat) + geom_boxplot()
```



We can see the unequal variance from fitted value versus residual plot, and is not normal distributed residuals (left skewed), and different residual pattern different control groups.

c

```
AIC(rat_aov2)
```

```
## [1] 923.5493
```

```
AIC(rat_aov3)
```

```
## [1] 885.4248
```

Based on aic criterion, the second model is better.

```
anova(rat_aov3)
```

```
## Analysis of Variance Table
##       Df Sum Sq Mean Sq  F value
## T      1 5462.4  5462.4 618.0879
## T2     1    0.6     0.6   0.0694
## treat  2    1.7     0.9   0.0986
```

```
summary(rat_aov3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ (1 | subject) + T + T2 + (0 + T | subject) + (0 + T2 |
##     subject) + treat
##    Data: rat
##
## REML criterion at convergence: 867.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6089 -0.4807  0.0259  0.4747  1.8508
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  subject   (Intercept) 27.709   5.264
##  subject.1 T           14.427   3.798
##  subject.2 T2           1.984   1.409
##  Residual               8.838   2.973
## Number of obs: 135, groups:  subject, 27
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)     54.26187    1.85185  29.301
## T               22.85344    0.97012  23.557
## T2               0.08201    0.31124   0.263
## treatthiouracil  0.99096    2.59084   0.382
## treatthyroxine  -0.07879    2.85497  -0.028
##
## Correlation of Fixed Effects:
##             (Intr) T      T2     trtthr
## T           -0.141
## T2           0.081 -0.310
## treatthircl -0.700  0.000  0.000
## treatthyrxn -0.635  0.000  0.000  0.454
```

d

It is possible. The model is:

```
rat_aov4 = lmer(weight ~ (1|subject) + T + T2 + treat + treat:T + treat:T2, data = rat)
```

e

```
rat_aov4 = lmer(weight ~ (1|subject) + T + T2 + treat + treat:T + treat:T2, data = rat)
anova(rat_aov4)
```

```
## Analysis of Variance Table
##          Df Sum Sq Mean Sq   F value
```

```
## T         1 145093  145093 3015.0219
## T2        1      3       3    0.0528
## treat     2    749     374    7.7774
## T:treat   2   5873    2936   61.0181
## T2:treat  2    467     234    4.8542
```

```
summary(rat_aov4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ (1 | subject) + T + T2 + treat + treat:T + treat:T2
##    Data: rat
##
## REML criterion at convergence: 936.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9910 -0.5755  0.0700  0.6061  2.5878
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 71.83    8.475
##  Residual             48.12    6.937
## Number of obs: 135, groups:  subject, 27
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)          54.1086     3.3831  15.994
## T                    24.0229     2.4456   9.823
## T2                    0.6143     0.5863   1.048
## treatthiouracil       0.9086     4.7844   0.190
## treatthyroxine        0.6302     5.2721   0.120
## T:treatthiouracil    -1.6271     3.4586  -0.470
## T:treatthyroxine     -2.1861     3.8112  -0.574
## T2:treatthiouracil   -1.9357     0.8291  -2.335
## T2:treatthyroxine     0.7122     0.9137   0.780
##
## Correlation of Fixed Effects:
##              (Intr) T      T2     trtthr trtthy T:trtthr T:trtthy T2:trtthr
## T            -0.449
## T2            0.347 -0.959
## treatthircl -0.707  0.317 -0.245
## treatthyrxn -0.642  0.288 -0.222  0.454
## T:tretthrcl  0.317 -0.707  0.678 -0.449 -0.204
## T:trtthyrxn  0.288 -0.642  0.615 -0.204 -0.449  0.454
## T2:trtthrcl -0.245  0.678 -0.707  0.347  0.157 -0.959   -0.435
## T2:trtthyrx -0.222  0.615 -0.642  0.157  0.347 -0.435   -0.959    0.454
```

Next is the stepwise:

```
first_set <- c("T", "treat", "T:treat", "T2", "T2:treat")
sapply(first_set, function(x) {
  AIC(lmer(as.formula(sprintf("weight~%s + (1|subject)", x)), data = rat))
})
```

```
##        T    treat  T:treat       T2 T2:treat
## 1061.223 1338.149  971.753 1138.685 1095.132
```

```
sapply(first_set, function(x) {
  AIC(aov(as.formula(sprintf("weight~%s", x)), data = rat))
})
```

```
##        T    treat  T:treat       T2 T2:treat
## 1106.175 1351.663 1032.855 1152.999 1103.568
```

So "T:treat" is selected.

```
second_set <- c("T", "treat", "T2", "T2:treat")
sapply(second_set, function(x) {
  AIC(lmer(as.formula(sprintf("weight~%s + (1|subject) + T:treat", x)), data = rat))
})
```

```
##        T    treat       T2 T2:treat
## 971.7530 964.4490 973.8606 964.1273
```

```
sapply(second_set, function(x) {
  AIC(aov(as.formula(sprintf("weight~%s + T:treat", x)), data = rat))
})
```

```
##        T    treat       T2 T2:treat
## 1032.855 1034.401 1034.833 1032.194
```

So "treat" is selected.

```
third_set <- c("T", "T2", "T2:treat")
sapply(third_set, function(x) {
  AIC(lmer(as.formula(sprintf("weight~%s + (1|subject) + T:treat + treat", x)), data = rat))
})
```

```
##        T       T2 T2:treat
## 964.4490 966.5559 958.2299
```

```
sapply(third_set, function(x) {
  AIC(aov(as.formula(sprintf("weight~%s + T:treat + treat", x)), data = rat))
})
```

```
##        T       T2 T2:treat
## 1034.401 1036.378 1036.149
```

So "T2:treat" is selected.

```
third_set <- c("T", "T2")
sapply(third_set, function(x) {
  AIC(lmer(as.formula(sprintf("weight~%s + (1|subject) + T:treat + treat + T2:treat", x)), data = rat))
})
```

```
##        T       T2
## 958.2299 958.2299
```

```
sapply(third_set, function(x) {
  AIC(aov(as.formula(sprintf("weight~%s + T:treat + treat + T2:treat", x)), data = rat))
})
```

```
##        T       T2
## 1036.149 1036.149
```

No improvement.

So the final model:

```
final2 <- lmer(weight ~ (1|subject) + T:treat + treat + T2:treat, data = rat)
summary(final2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: weight ~ (1 | subject) + T:treat + treat + T2:treat
##    Data: rat
##
## REML criterion at convergence: 936.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9910 -0.5755  0.0700  0.6061  2.5878
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subject  (Intercept) 71.83    8.475
##  Residual             48.12    6.937
## Number of obs: 135, groups:  subject, 27
##
## Fixed effects:
##                     Estimate Std. Error t value
## (Intercept)          54.1086     3.3831  15.994
## treatthiouracil       0.9086     4.7844   0.190
## treatthyroxine        0.6302     5.2721   0.120
## T:treatcontrol       24.0229     2.4456   9.823
## T:treatthiouracil    22.3957     2.4456   9.157
## T:treatthyroxine     21.8367     2.9231   7.470
## treatcontrol:T2       0.6143     0.5863   1.048
## treatthiouracil:T2   -1.3214     0.5863  -2.254
## treatthyroxine:T2     1.3265     0.7008   1.893
##
## Correlation of Fixed Effects:
##             (Intr) trtthr trtthy T:trtc T:trtthr T:trtthy trtc:T2
## treatthircl -0.707
## treatthyrxn -0.642  0.454
## T:tretcntrl -0.449  0.317  0.288
## T:tretthrcl  0.000 -0.317  0.000  0.000
## T:trtthyrxn  0.000  0.000 -0.344  0.000  0.000
## trtcntrl:T2  0.347 -0.245 -0.222 -0.959  0.000    0.000
```

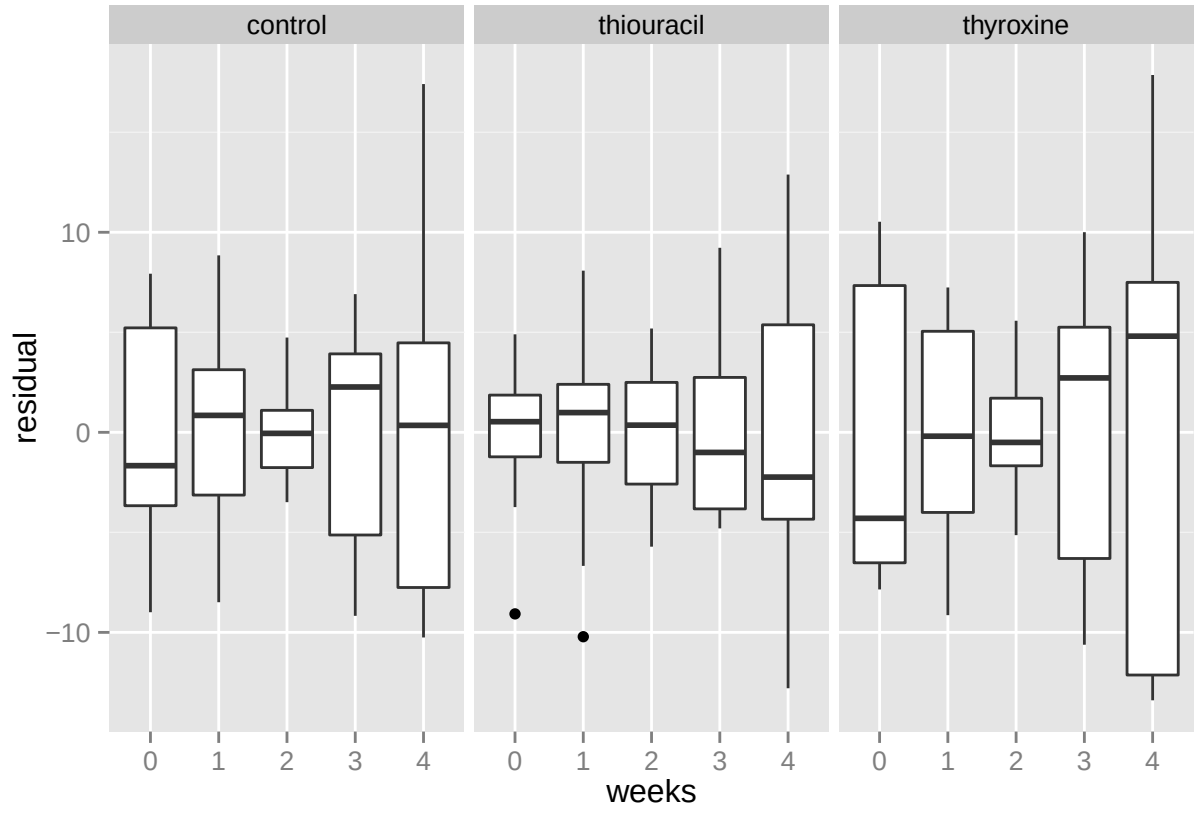
```
## trtthrcl:T2  0.000  0.245  0.000  0.000 -0.959    0.000    0.000
## trtthyrx:T2  0.000  0.000  0.266  0.000  0.000   -0.959    0.000
##             trtthr:T2
## treatthircl
## treatthyrxn
## T:tretcntrl
## T:tretthrcl
## T:trtthyrxn
## trtcntrl:T2
## trtthrcl:T2
## trtthyrx:T2  0.000
```

```
anova(final2)
```

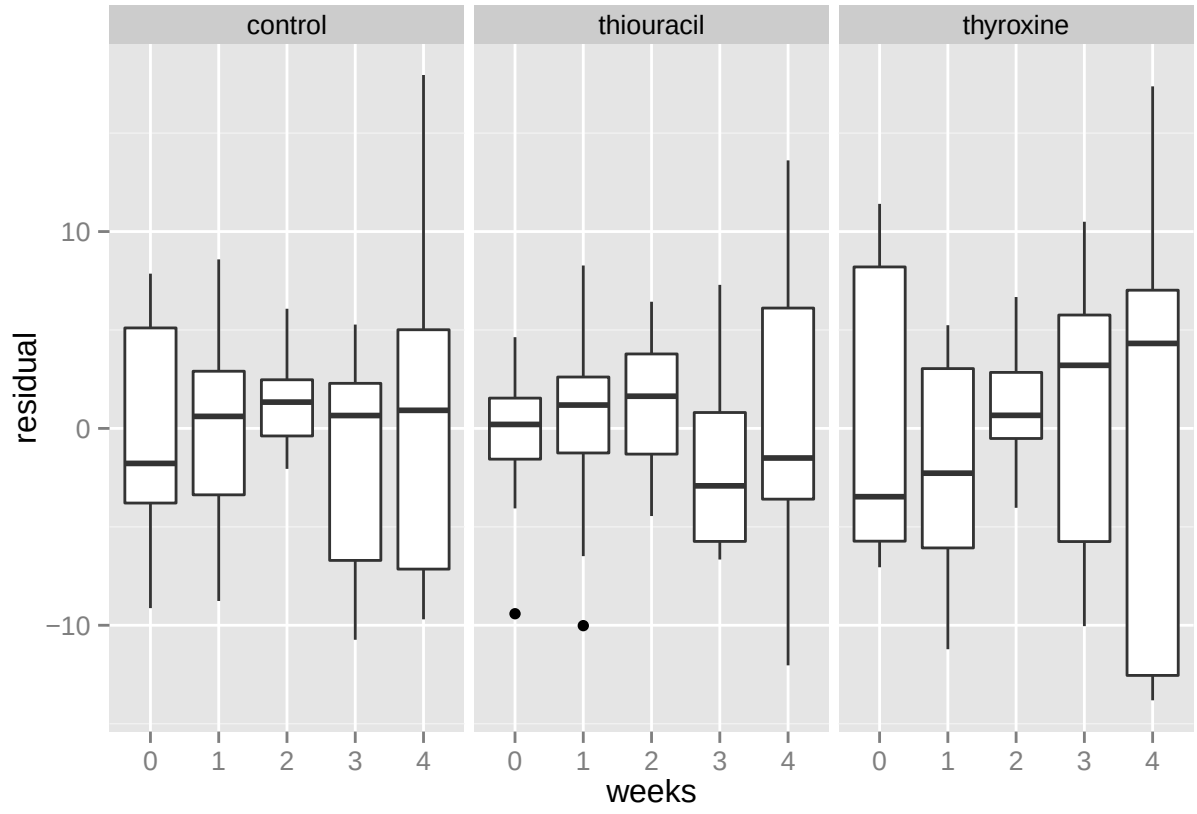
```
## Analysis of Variance Table
##          Df Sum Sq Mean Sq   F value
## treat     2    749     374    7.7774
## T:treat   3 150966   50322 1045.6860
## treat:T2  3    470     157    3.2537
```

f

```
# final1
rat_resid_dataframe <- cbind(rat, residual = resid(final1))
ggplot(rat_resid_dataframe, aes(x = weeks, y = residual)) + facet_wrap(~treat) + geom_boxplot()
```




```
# final2
rat_resid_dataframe <- cbind(rat, residual = resid(final2))
ggplot(rat_resid_dataframe, aes(x = weeks, y = residual)) + facet_wrap(~treat) + geom_boxplot()
```



They are almost the same pattern. The means are a little different, and the variances are a little different either.

g

```
rat_aov5 = lmer(weight ~ (1|subject) + T + T2 + treat + treat:T + treat:T2 + (0 + (treat:T)|subject) +
```

```
## 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
```