Biostat 200C Homework 4

Due 11:59PM May 23rd

Yenlin Lai

Q1. Balanced one-way ANOVA random effects model

Consider the balanced one-way ANOVA random effects model with a levels and n observations in each level

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \ldots, a, \quad j = 1, \ldots, n.$$

where α_i are iid from $N(0, \sigma_\alpha^2)$, ϵ_{ij} are iid from $N(0, \sigma_\epsilon^2)$.

1. Derive the ANOVA estimate for μ, σ_α^2, and σ_ϵ^2. Specifically show that

$$\begin{aligned}
\mathbb{E}(\bar{y}_{\cdot\cdot}) &= \mathbb{E}\left(\frac{\sum_{ij} y_{ij}}{na}\right) = \mu \\
\mathbb{E}(\text{SSE}) &= \mathbb{E}\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\cdot})^2\right] = a(n-1)\sigma_\epsilon^2 \\
\mathbb{E}(\text{SSA}) &= \mathbb{E}\left[\sum_{i=1}^{a}\sum_{j=1}^{n}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2\right] = (a-1)(n\sigma_\alpha^2 + \sigma_\epsilon^2),
\end{aligned}$$

   which can be solved to obtain ANOVA estimate

$$\begin{aligned}
\widehat{\mu} &= \frac{\sum_{ij} y_{ij}}{na}, \\
\widehat{\sigma}_\epsilon^2 &= \frac{\text{SSE}}{a(n-1)}, \\
\widehat{\sigma}_\alpha^2 &= \frac{\text{SSA}/(a-1) - \widehat{\sigma}_\epsilon^2}{n}.
\end{aligned}$$

Solution: See the pdf file attached.

2. Calculate the three estimates for the `pulp` example in class, check if your results match with the R output.

Solution: See the pdf file attached.

Q2. ELMR Exercise 11.1 (p251)

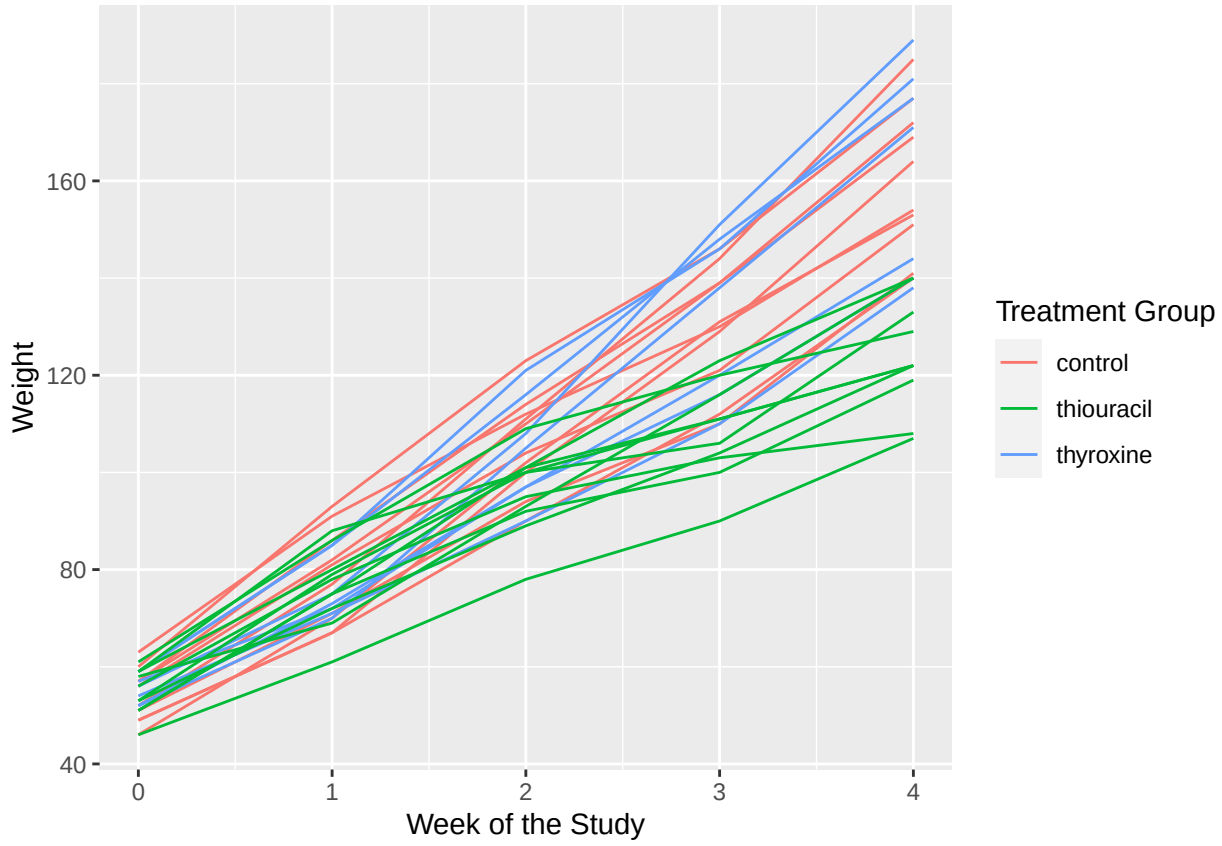
The `ratdrink` data consist of 5 weekly measurements of body weight for 27 rats. The first 10 rats are on a control treatment while 7 rats have thyroxine added to their drinking water and 10 rats have thiouracil added to their water.

```
help("ratdrink")
```

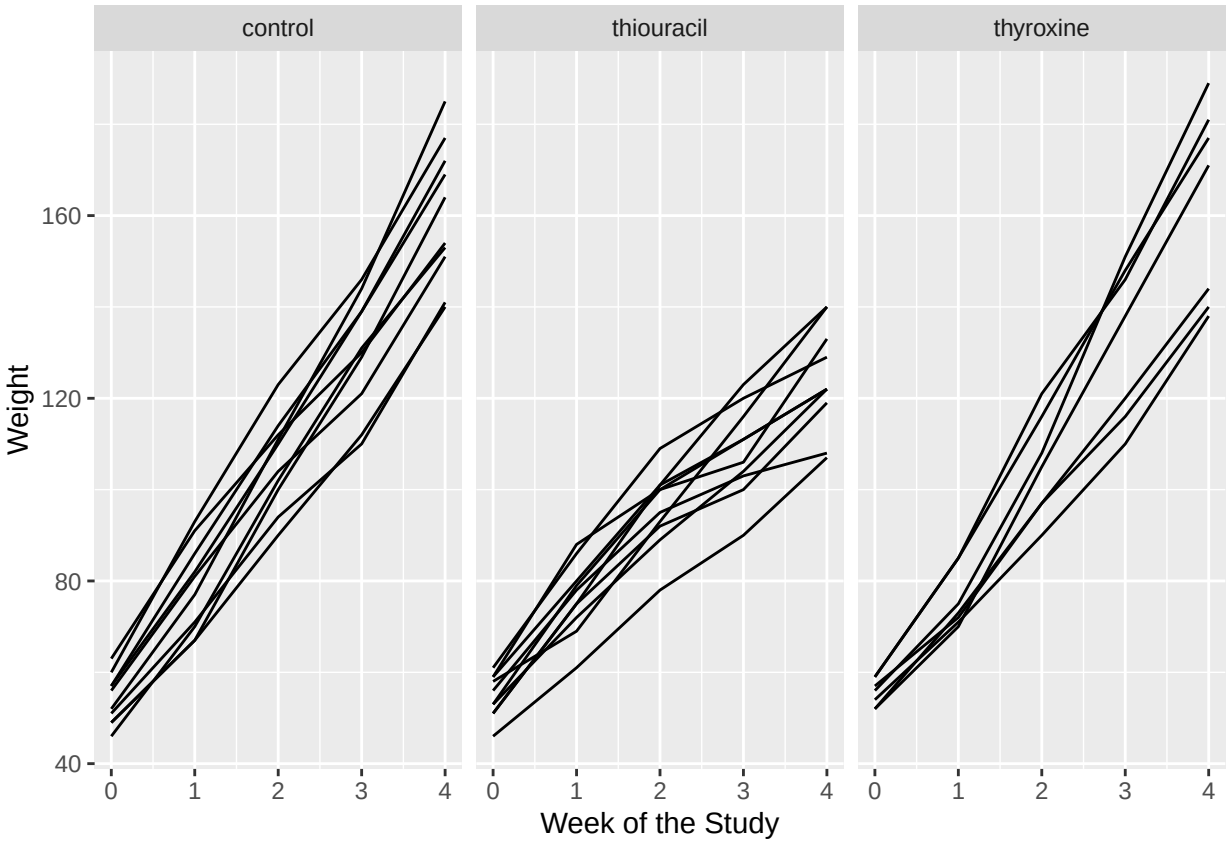
```
## starting httpd help server ... done
```

1. 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.

```
data("ratdrink")
library(ggplot2)
ggplot(data = ratdrink) +
  geom_line(mapping = aes(x = weeks, y = wt, group = subject,
                          colour = treat)) +
  labs(x = "Week of the Study", y = "Weight", colour = "Treatment Group")
```




```
ggplot(data = ratdrink) +
  geom_line(mapping = aes(x = weeks, y = wt, group = subject)) +
  facet_wrap(~ treat) +
  labs(x = "Week of the Study", y = "Weight")
```



From the plots shown above, we can observe that for control group and thyroxine group, they seem to have the similar pattern, while thyroxine group may have a higher variation within group. For thiouracil group, the slope is lower than the other two groups. That is to say, when the weeks of the study increase, the weight of the rat in thiouracil group increase slower than the rat in control and thyroxine group.

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

- The fixed effect intercept term.
- The interaction between thiouracil and week.
- The intercept random effect SD (standard deviation).

```
library(lme4)
```

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

```
## Warning: package 'Matrix' was built under R version 4.1.1
```

```
##
## Attaching package: 'Matrix'
```

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

```
longm <- lmer(wt ~ treat * weeks + (weeks | subject), data = ratdrink)
summary(longm)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: wt ~ 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
##
## Correlation of Fixed Effects:
##             (Intr) trtthr trtthy weeks  trtthr:
## treatthircl -0.707
## treatthyrxn -0.642  0.454
## weeks       -0.250  0.177  0.160
## trtthrcl:wk  0.177 -0.250 -0.113 -0.707
## trtthyrxn:w  0.160 -0.113 -0.250 -0.642  0.454
```

The parameter of fixed effect intercept term is 52.88, which means when all the explanatory variables are 0, and there are no interaction and random effect, the mean weight of the rat will be 52.88.

The parameter of interaction between thiouracil and week is -9.37. The coefficient of `treatthiouracil:weeks` is the change in slope of `treat` associated with a one-unit increase in `weeks`, and also the change in slope of `weeks` associated with a one-unit increase in `treatthiouracil` That is to say, when the treatment group is thiouracil, the slope of variable `weeks` will decrease by 9.37.

The parameter of intercept random effect SD (standard deviation) is 5.7. The standard deviation between subjects is 5.7.

3. Check whether there is a significant treatment effect.

```
library(pbkrtest)
mmod <- lmer(wt ~ treat * weeks + (weeks | subject), data = ratdrink
             , REML = TRUE)
```

```
mmodr <- lmer(wt ~  weeks + (weeks | subject), data = ratdrink
              , REML = TRUE)
KRmodcomp(mmod, mmodr)
```

```
## large : wt ~ treat * weeks + (weeks | subject)
## small : wt ~ weeks + (weeks | subject)
##          stat     ndf     ddf F.scaling   p.value
## Ftest  8.7125  4.0000 26.8141   0.94552 0.0001215 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a significant treatment effect because the p-value comparing the two models above is smaller than the significant level 0.05, which means that we cannot ignore the effect of the variable `treat`.

```
mmodt <- lmer(wt ~ treat + weeks + (weeks | subject), data = ratdrink
             , REML = TRUE)
KRmodcomp(mmodt, mmodr)
```

```
## large : wt ~ treat + weeks + (weeks | subject)
## small : wt ~ weeks + (weeks | subject)
##          stat     ndf     ddf F.scaling p.value
## Ftest  0.1014  2.0000 24.0000         1   0.904
```

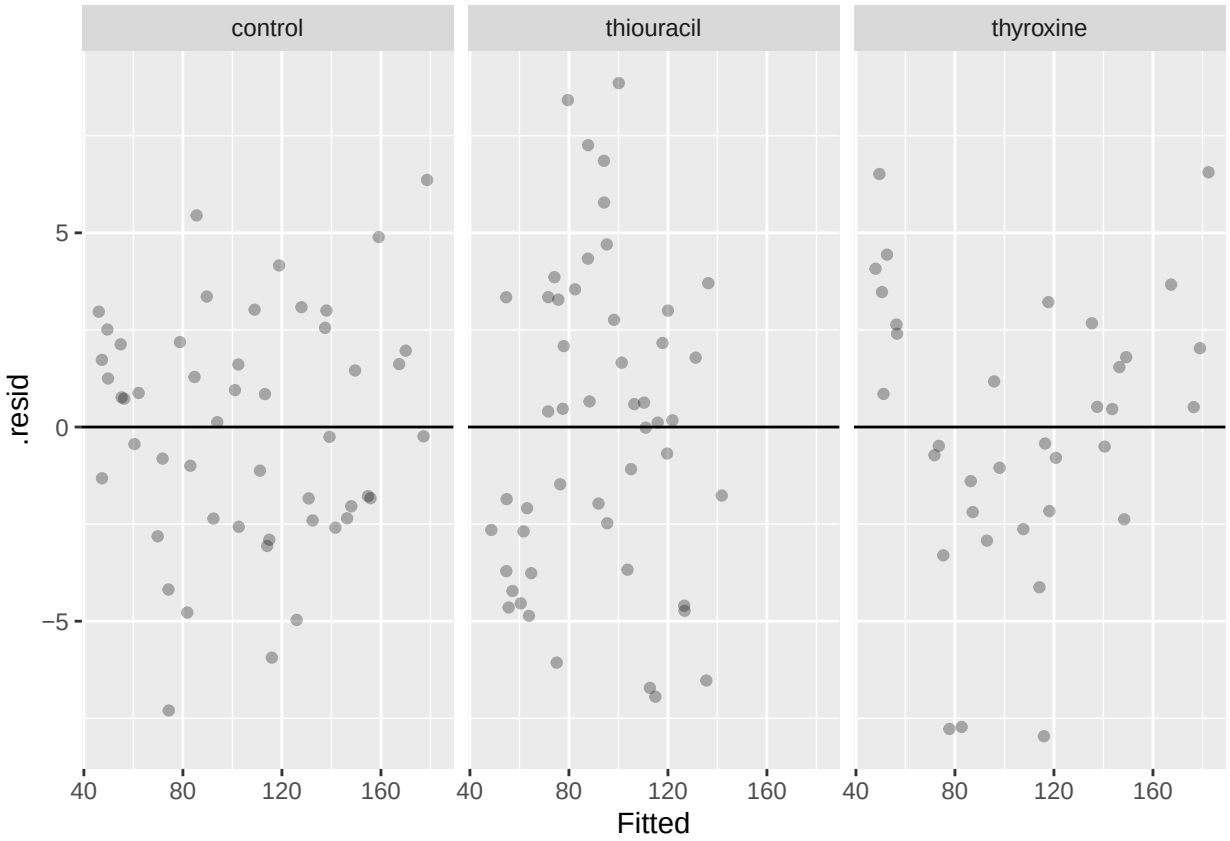
We get the conclusion that `treat` is not a significant effect when we do not add the interaction term. From the previous test we shown, we can know that the interaction term is significant, which also means `treat` play an important role in this model.

4. Construct diagnostic plots showing the residuals against the fitted values and a QQ plot of the residuals. Comment on the plots.

```
library(broom.mixed)
```

```
## Warning: package 'broom.mixed' was built under R version 4.1.3
```

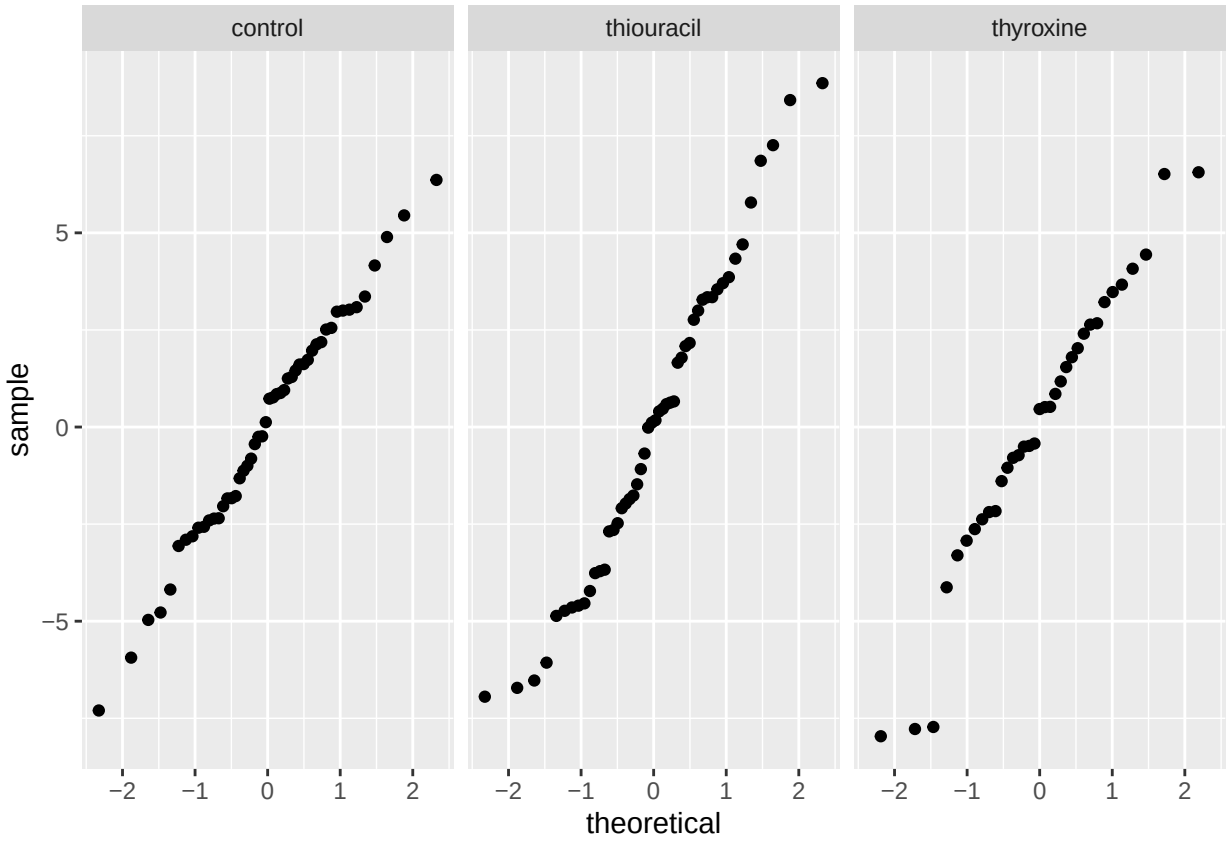
```
diagd <- augment(mmod)

diagd %>%
  ggplot(mapping = aes(x = .fitted, y = .resid)) +
  geom_point(alpha = 0.3) +
  geom_hline(yintercept = 0) +
  labs(x = "Fitted", ylab = "Residuals") +
  facet_grid(~ treat)
```



The residual plots are plotted by the different treatment group. For control group and thiouracil group, the residuals seem to randomly spread out here. It seems that they are homoscedastic, and there is no big problem with the outliers.

However, for thyroxine group, there is a pattern in the residual plot, which may suggest that we use the quadratic term here to better fit the model.

```
diagd %>%
  ggplot(mapping = aes(sample = .resid)) +
  stat_qq() +
  facet_grid(~ treat)
```



For control group and thiouracil group, QQ plots show that we don't see any outliers, and they don't violate the normality. However, for thyroxine group, QQ plot shows the violation of normal assumption. It suggests other transformations of the response.

5. 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?

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

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

##
## 2 warning(s): Model failed to converge with max|grad| = 0.00416213 (tol = 0.002, component 1) (and o
```

```
##                             2.5 %     97.5 %
## .sig01                  3.2702806  7.8160459
## .sig02                 -0.5596748  0.4411518
## .sig03                  2.5597076  5.1593973
## .sigma                  3.6633464  5.0189441
## (Intercept)            48.9072211 56.4702388
## treatthiouracil        -0.9590213 10.9208528
## treatthyroxine         -6.9761872  5.4760701
## weeks                  24.0797602 28.7524958
## treatthiouracil:weeks -12.5661067 -6.0029984
## treatthyroxine:weeks   -3.0336872  4.6471142
```

From the confidence intervals shown above, we can see that for the random effect terms `.sig02`, the 95% confidence interval contains 0. Therefore, the correlation between random intercept and slope is not significant.

The thyroxine group is not significantly different from the control group, because for the parameter `treatthyroxine`, the 95% confidence interval contains 0. It means that the effect of thyroxine itself could be the same as the control group.