

Q3

Hun Kang

2024-10-03

Q3.

```
library(lme4)
library(tidyverse)
library(ggplot2)
library(ggpubr)
library(cowplot)
```

```
dat = read.table(url("https://www2.stat.duke.edu/~pdh10/Teaching/610/Homework/cd4.dat"),header=TRUE)
dat$pid = as.factor(dat$pid)
head(dat)
```

```
##   pid  cd4 trt time
## 1   1  4.24  1  0.00
## 2   1  6.08  1  0.56
## 3   1  3.61  1  0.79
## 4   1  3.61  1  1.42
## 5   1  3.46  1  1.94
## 6   2  1.00  0  0.00
```

```
p0 = dat %>%
  add_count(pid, name = "count") %>%
  filter(count > 6) %>%
  filter(trt == 0) %>%
  ggplot(aes(x = time, y = cd4, color = pid)) +
  geom_line(show.legend = F, color="darkblue") +
  facet_wrap(~ pid, strip.position = "bottom") +
  theme_cowplot() +
  theme(
    strip.background = element_blank(),
    strip.text.x = element_blank(),
    axis.text.x = element_blank()
  ) +
  ggtitle("trt 0")
```

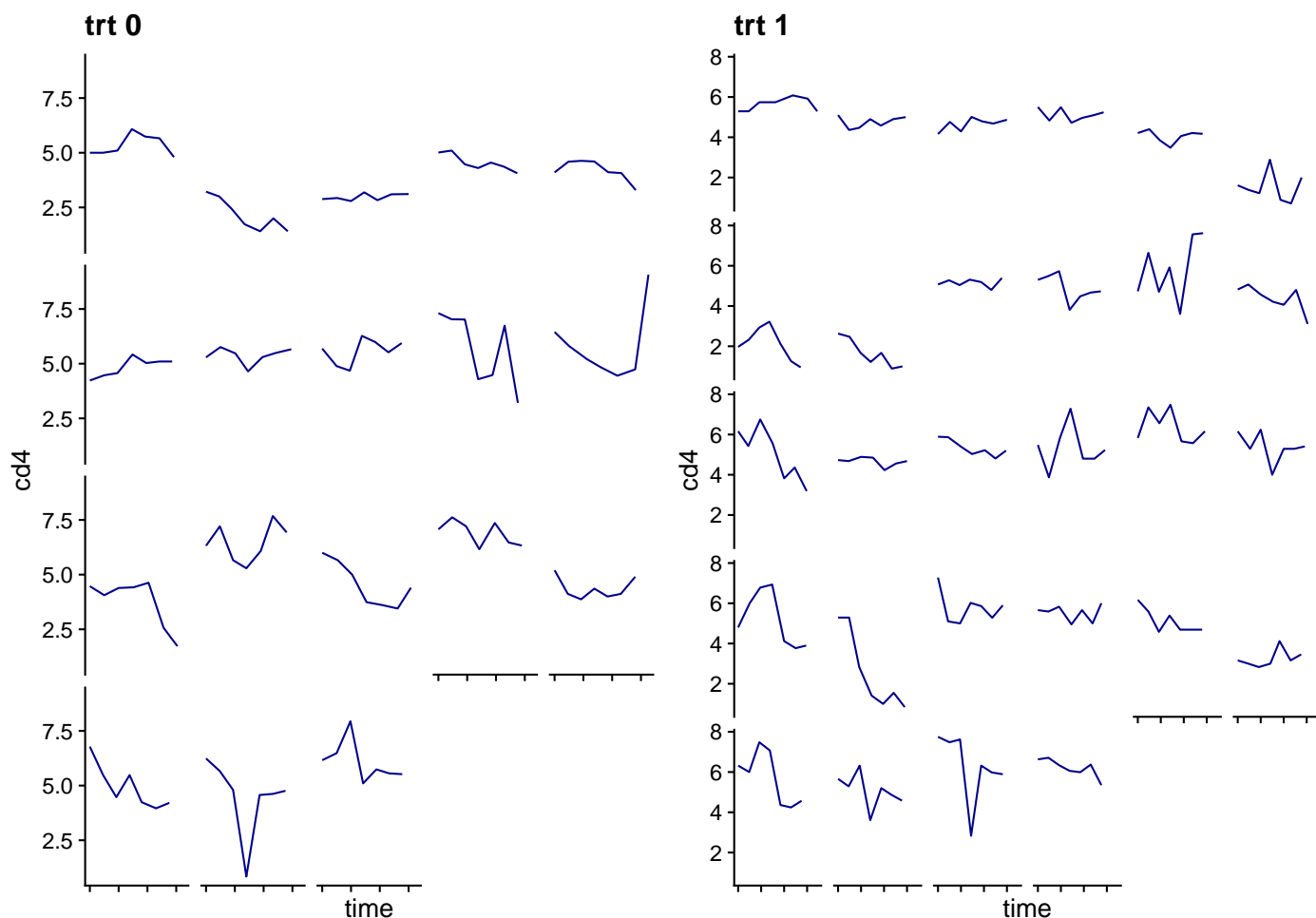
```
p1 = dat %>%
  add_count(pid, name = "count") %>%
  filter(count > 6) %>%
  filter(trt == 1) %>%
  ggplot(aes(x = time, y = cd4, color = pid)) +
  geom_line(show.legend = F, color="darkblue") +
  facet_wrap(~ pid, strip.position = "bottom") +
  theme_cowplot() +
  theme(
```

```

strip.background = element_blank(),
strip.text.x = element_blank(),
axis.text.x = element_blank()
)+
ggtitle("trt 1")

ggarrange(p0, p1)

```



a.

$$\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 1_{trt_{ij}=1} + \beta_2 time_{ij} + \beta_3 1_{trt_{ij}=1} time_{ij} \\
&\quad + u_j + v_j time_{ij} + \epsilon_{ij} \\
\epsilon_{ij} &\stackrel{iid}{\sim} N(0, \sigma^2) \\
\begin{bmatrix} u_j \\ v_j \end{bmatrix} &\stackrel{iid}{\sim} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_u^2 & \tau_{uv} \\ \tau_{uv} & \tau_v^2 \end{bmatrix} \right)
\end{aligned}$$

```

mod = lmer(cd4 ~ trt*time + (1+time | pid), data = dat)
summary(mod)

```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: cd4 ~ trt * time + (1 + time | pid)
## Data: dat
##
## REML criterion at convergence: 3113.6

```

```
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -5.0923 -0.3953  0.0056  0.4091  5.0152
##
## Random effects:
##   Groups      Name            Variance Std.Dev. Corr
##   pid        (Intercept)  1.9263     1.3879
##             time          0.3400     0.5831  -0.06
##   Residual                0.5147     0.7174
## Number of obs: 1072, groups:  pid, 250
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  4.82613    0.13726  35.160
## trt          -0.12783    0.19010  -0.672
## time         -0.30611    0.09928  -3.083
## trt:time     -0.08920    0.13574  -0.657
##
## Correlation of Fixed Effects:
##              (Intr) trt      time
## trt          -0.722
## time         -0.237  0.171
## trt:time     0.174 -0.243 -0.731
```

```
confint(mod, method = "profile")
```

```
## Computing profile confidence intervals ...
```

```
##              2.5 %      97.5 %
## .sig01        1.2474891  1.5336905
## .sig02       -0.3015859  0.2245137
## .sig03        0.4227834  0.7290044
## .sigma        0.6800690  0.7581767
## (Intercept)  4.5570316  5.0950726
## trt          -0.5002298  0.2448521
## time         -0.5022744 -0.1086288
## trt:time     -0.3565952  0.1774465
```

95% confidence intervals of `trt` and `trt:time` include 0, which means no effect. In the last HW, in the first approach the treatment effect was significant, but the model did not account for across group variation. In the second model the treatment effect could not be estimated because by treating group effects as fixed, the two were confounded. In the model above, by means of linear mixed effect model we pooled the group means (random slope and intercept) under the same distribution and came up with a prediction of groupwise effects as a function of the whole data. Therefore, we accounted for the group effects and estimated the treatment effects at the same time. It turned out that after considering group heterogeneity, the variation in the data leave less to support treatment effects.

b.

More shrinkage for groups of smaller sample sizes. Especially the slope is heavily restricted as our model yields weak treatment effects.

$$y_{ij} = \underbrace{(\beta_0 + \beta_1 1_{trt_{ij}=1} + u_j)}_{\text{Intercept}} + \underbrace{(\beta_2 + \beta_3 1_{trt_{ij}=1} + v_j)}_{\text{Slope}} time_{ij} + \epsilon_{ij}$$

```

# OLS estimates
fits = dat %>%
  group_by(pid) %>%
  do(model = lm(cd4 ~ time, data=..))
J = length(fits$pid)
alphas = betas = trt = nj = numeric(length=J)
for(j in 1:J){
  alphas[j] = coef(fits$model[[j]])[1]
  betas[j] = coef(fits$model[[j]])[2]
  tmp = filter(dat, pid == fits$pid[[j]])
  trt[j] = tmp[1, "trt"]
  nj[j] = nrow(tmp)
}
OLS = data.frame(intercept = alphas, slope = betas)

```

```

# LMM predictions
u = ranef(mod)$pid[, "(Intercept)"]
v = ranef(mod)$pid[, "time"]
b0 = fixef(mod)[ "(Intercept)" ]
b1 = fixef(mod)[ "trt" ]
b2 = fixef(mod)[ "time" ]
b3 = fixef(mod)[ "trt:time" ]
LMM = data.frame(
  intercept = b0 + b1*trt + u,
  slope = b2 + b3*trt + v
)

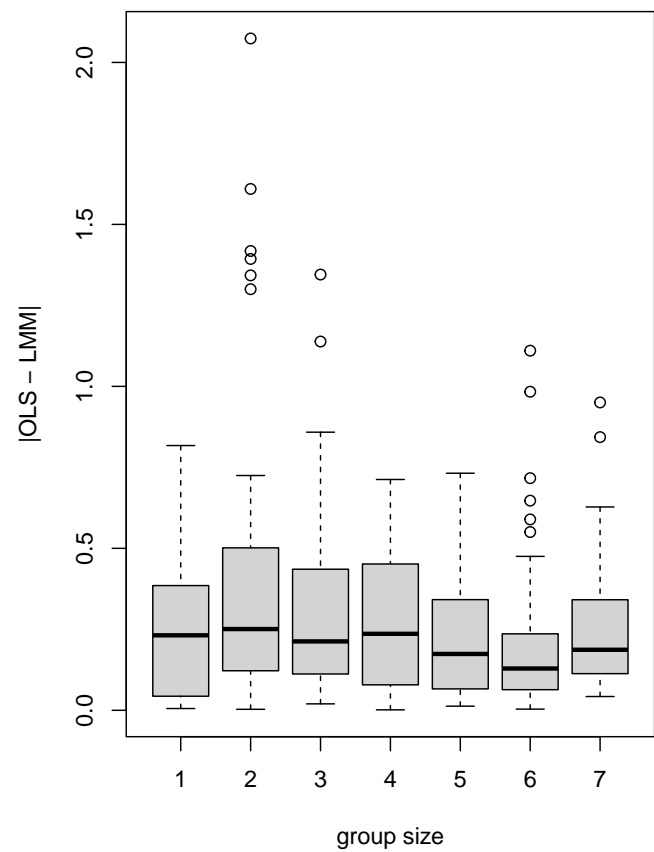
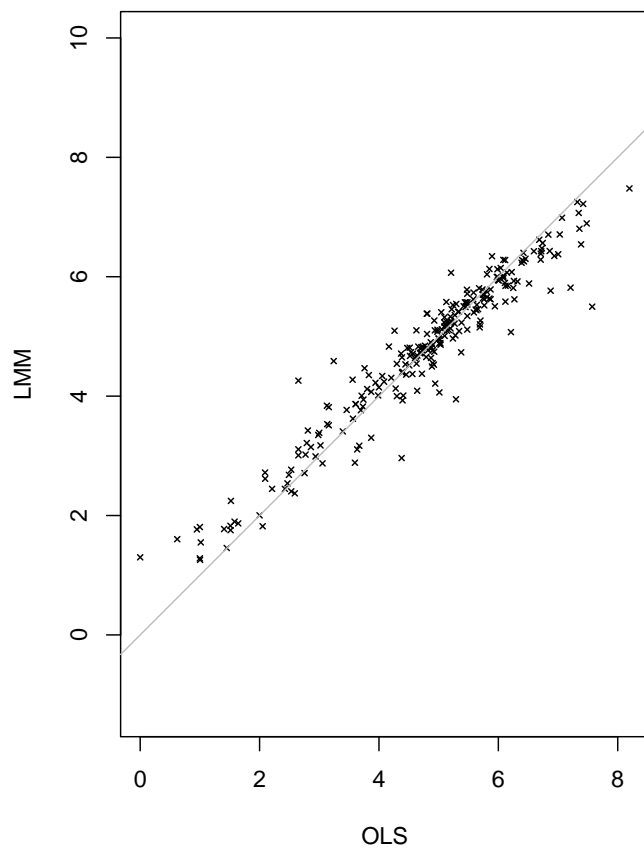
```

```

par(mfrow=c(1,2))
ols = OLS$intercept
lmm = LMM$intercept
plot(ols, lmm, pch=4, cex=0.5, asp=1,
     xlab = "OLS", ylab="LMM")
abline(a=0,b=1, col="grey")
boxplot(abs(ols-lmm) ~ nj,
        xlab="group size", ylab= "|OLS - LMM|")
mtext("Intercepts", outer=T, side=3, line=-2)

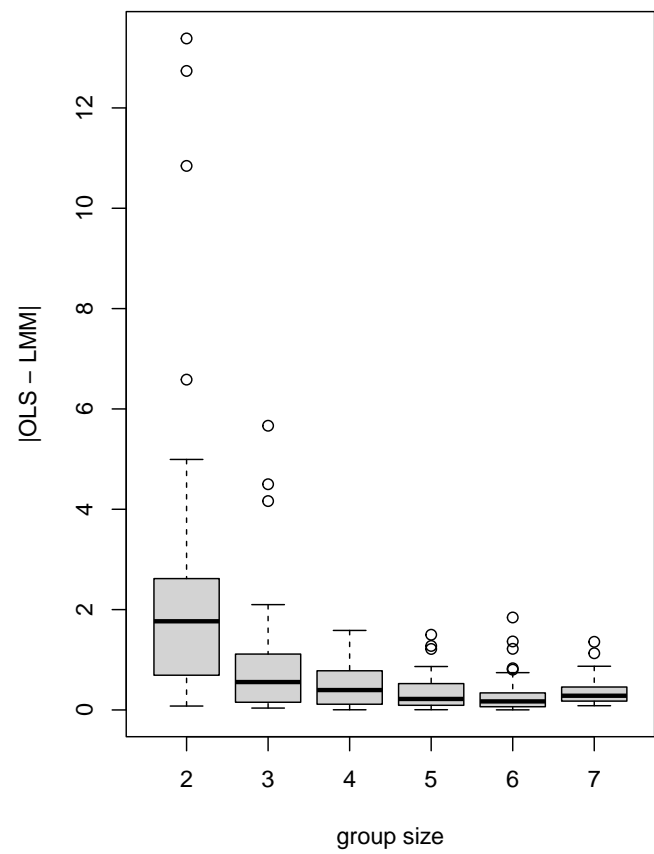
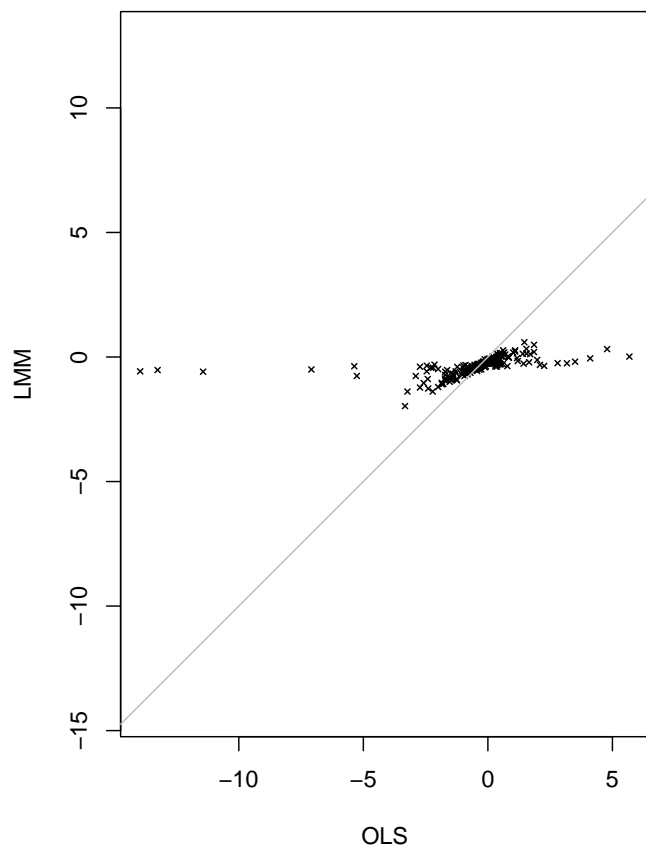
```

Intercepts



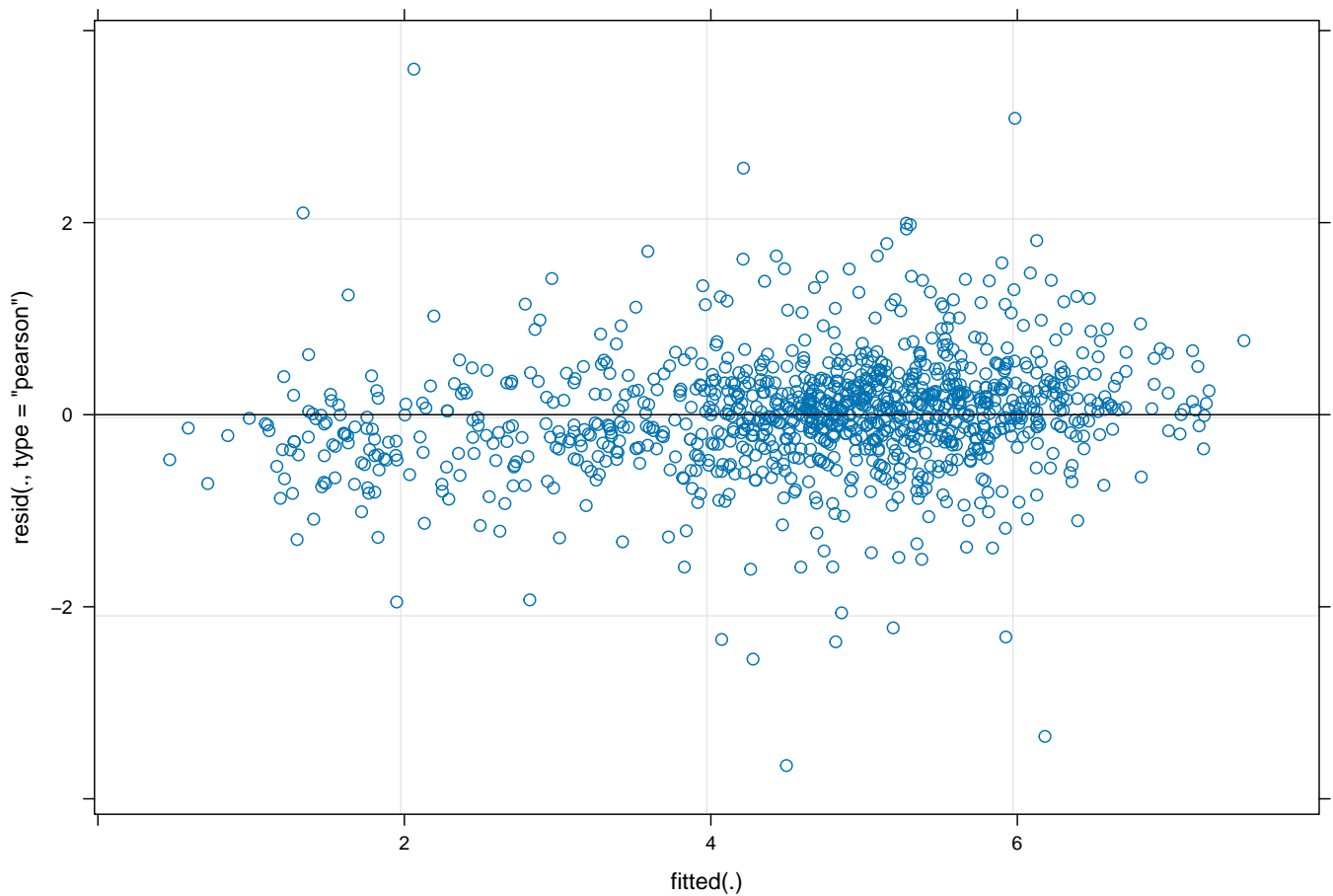
```
par(mfrow=c(1,2))
ols = OLS$slope
lmm = LMM$slope
plot(ols, lmm, pch=4, cex=0.5, asp=1,
     xlab = "OLS", ylab="LMM")
abline(a=0,b=1, col="grey")
boxplot(abs(ols-lmm) ~ nj,
       xlab="group size", ylab= "|OLS - LMM|")
mtext("Slopes", outer=T, side=3, line=-2)
```

Slopes



c.

```
plot(mod)
```



Overall looks fine around 0, but for some groups we see few points with large deviations.

```
dat$pred = fitted(mod)
dat$resid = resid(mod)
dat %>%
  add_count(pid, name = "count") %>%
  filter(count > 6) %>%
  ggplot(aes(x = pred, y = resid, color = pid)) +
  geom_line(show.legend = F, color="darkblue") +
  geom_point(show.legend = F, color="darkblue", size =0.8) +
  geom_hline(yintercept=0,linetype="dashed", color = "grey") +
  facet_wrap(~ pid, scales = "free_x", strip.position = "bottom") +
  theme_bw() +
  theme(
    strip.background = element_blank(),
    strip.text.x = element_blank(),
    axis.text.x = element_blank()
  ) +
  ggtitle("group size > 6")
```

group size > 6

