HW4

Andrew Pagtakhan

5/7/2021

```
library(lme4)
library(dplyr)
library(foreign)
library(ggplot2)

knitr::opts_chunk$set(echo = TRUE, tidy.opts = list(width.cutoff = 60),
                      tidy = TRUE, warning = FALSE)
```

```
# load data
cl <- read.dta("classroom.dta")
# convert to factor
cols_factor <- c("sex", "minority", "classid", "schoolid", "childid")
cl[cols_factor] <- lapply(cl[cols_factor], factor)
```

1.

```
summary(lme1 <- lmer(mathgain ~ mathkind + ses + minority + yearstea +
    (1 | schoolid), data = cl))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: mathgain ~ mathkind + ses + minority + yearstea + (1 | schoolid)
##    Data: cl
##
## REML criterion at convergence: 11403.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.0017 -0.6493 -0.0277  0.5669  4.1346
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schoolid (Intercept)  99.82    9.991
##  Residual             791.40   28.132
## Number of obs: 1190, groups:  schoolid, 107
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 281.27510   11.00687  25.555
## mathkind     -0.47046    0.02242 -20.986
## ses           5.49819    1.25197   4.392
## minority1    -7.71090    2.36056  -3.267
## yearstea      0.07857    0.09844   0.798
```

```
##
## Correlation of Fixed Effects:
##           (Intr) mthknd ses    mnrty1
## mathkind  -0.975
## ses        0.145 -0.171
## minority1 -0.313  0.165  0.158
## yearstea  -0.123  0.006 -0.025  0.047
```

```
round(logLik(lme1), 1)
```

```
## 'log Lik.' -5701.7 (df=7)
```

1a.

$MATHGAIN_{ijk} = \beta_0 + \beta_1 SES_{ijk} + \beta_2 SEX_{ijk} + \beta_3 MINORITY_{ijk} + \beta_4 YEARSTEA_{jk} + \zeta_{4k} YEARSTEA_{jk} + \zeta_{0k} + \varepsilon_{ijk}$

Assumptions (independent of one another):

$\zeta_{0k} \sim N(0, \sigma^2_\zeta)$

$\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$

$\rho(\zeta_{4k}, \zeta_{0k}) = 0$

```
summary(lme1a <- lmer(mathgain ~ mathkind + ses + minority +
    yearstea + (0 + yearstea | schoolid) + (1 | schoolid), data = cl))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: mathgain ~ mathkind + ses + minority + yearstea + (0 + yearstea |
##     schoolid) + (1 | schoolid)
##    Data: cl
##
## REML criterion at convergence: 11401.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.0463 -0.6525 -0.0299  0.5638  4.1634
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  schoolid   yearstea      0.07809  0.2794
##  schoolid.1 (Intercept)  92.48635  9.6170
##  Residual               783.78819 27.9962
## Number of obs: 1190, groups:  schoolid, 107
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 281.09447   11.00746  25.537
## mathkind     -0.47047    0.02242 -20.985
## ses           5.43383    1.25127   4.343
## minority1    -7.42511    2.36495  -3.140
## yearstea      0.07515    0.10571   0.711
##
## Correlation of Fixed Effects:
##           (Intr) mthknd ses    mnrty1
## mathkind  -0.975
```

```
## ses         0.143 -0.168
## minority1 -0.313  0.165  0.155
## yearstea  -0.120  0.008 -0.016  0.039
```

```
round(logLik(lme1a), 1)
```

```
## 'log Lik.' -5700.8 (df=8)
```

1c - 1e

We do not need school-specific varying coefficients for YEARSTEA. The p-value of the fixed effect for YEARSTEA increases, and does not change the decision to reject the null hypothesis.

```
anova(lme1, lme1a, refit = FALSE)
```

```
## Data: cl
## Models:
## lme1: mathgain ~ mathkind + ses + minority + yearstea + (1 | schoolid)
## lme1a: mathgain ~ mathkind + ses + minority + yearstea + (0 + yearstea |
## lme1a:     schoolid) + (1 | schoolid)
##       npar   AIC   BIC  logLik deviance Chisq Df Pr(>Chisq)
## lme1     7 11417 11453 -5701.7    11403
## lme1a    8 11418 11458 -5700.8    11402 1.631  1     0.2016
```

2.

```
summary(lme2 <- lmer(mathgain ~ mathkind + ses + minority + yearstea +
    (yearstea | schoolid), data = cl))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: mathgain ~ mathkind + ses + minority + yearstea + (yearstea |
##     schoolid)
##    Data: cl
##
## REML criterion at convergence: 11388.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.9714 -0.6142 -0.0406  0.5795  4.2447
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  schoolid (Intercept) 255.4654 15.9833
##           yearstea      0.5713  0.7558  -0.85
##  Residual             757.6597 27.5256
## Number of obs: 1190, groups:  schoolid, 107
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 282.59131   10.98180  25.733
## mathkind     -0.47273    0.02220 -21.298
## ses           5.32138    1.24233   4.283
## minority1    -8.13790    2.31306  -3.518
## yearstea      0.08622    0.12873   0.670
```

```
##
## Correlation of Fixed Effects:
##           (Intr) mthknd ses    mnrty1
## mathkind  -0.967
## ses        0.144 -0.173
## minority1 -0.310  0.165  0.166
## yearstea  -0.178  0.004 -0.012  0.054
```

```
round(logLik(lme2), 1)
```

```
## 'log Lik.' -5694.1 (df=9)
```

2a - 2c

We need school-specific varying coefficients for YEARSTEA The p-value of the fixed effect for YEARSTEA increases, and does not change the decision to reject the null hypothesis.

```
anova(lme1, lme2, refit = FALSE)
```

```
## Data: cl
## Models:
## lme1: mathgain ~ mathkind + ses + minority + yearstea + (1 | schoolid)
## lme2: mathgain ~ mathkind + ses + minority + yearstea + (yearstea |
## lme2:     schoolid)
##      npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lme1    7 11417 11453 -5701.7    11403
## lme2    9 11406 11452 -5694.1    11388 15.032  2  0.0005444 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

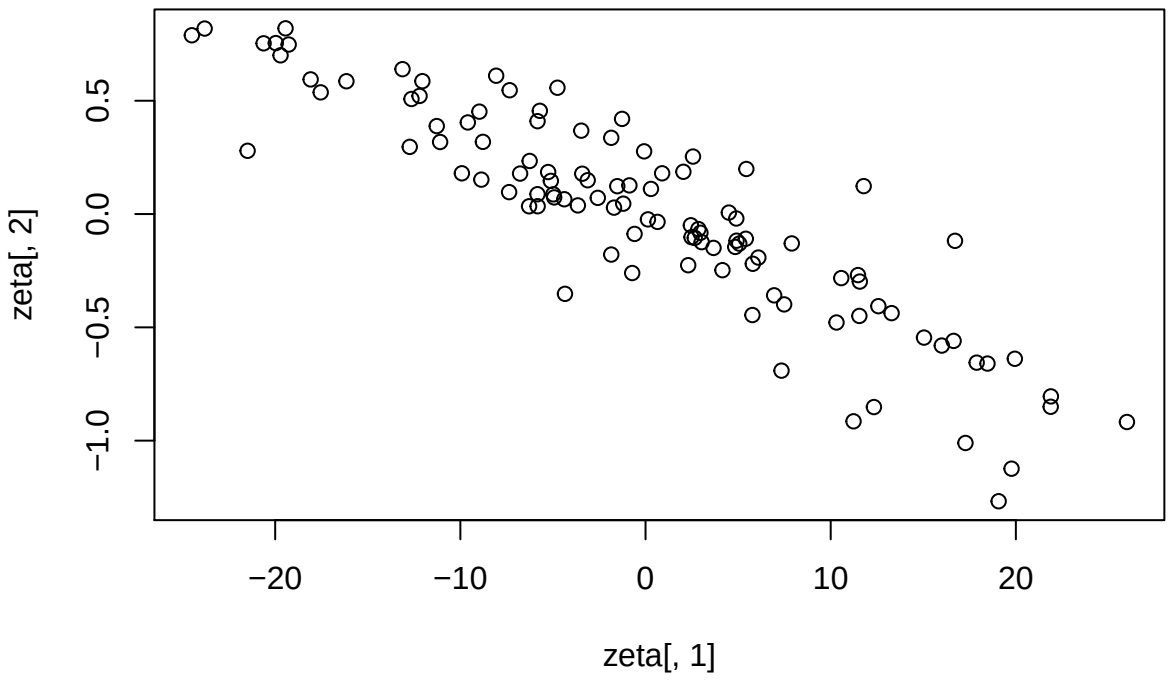
3a. - 3c.

$\rho(\zeta_{0k}, \zeta_{4k}) = -.89$. This matches what we see in the plot below

```
ranefs <- ranef(lme2)
zeta <- ranefs$schoolid
round(cor(zeta[, 1], zeta[, 2]), 2)
```

```
## [1] -0.9
```

```
plot(zeta[, 1], zeta[, 2])
```



4a.

```
head(zeta)
```

```
##   (Intercept)    yearstea
## 1   0.6481776 -0.03462942
## 2  10.5713769 -0.28262061
## 3  16.7133040 -0.11811983
## 4 -23.8103379  0.81862631
## 5  -1.2040642  0.04581467
## 6  19.9444312 -0.63870464
```

4b.

```
summary(lme4 <- lmer(mathgain ~ yearstea + (yearstea | schoolid),
    data = cl))
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: mathgain ~ yearstea + (yearstea | schoolid)
##    Data: cl
##
## REML criterion at convergence: 11771.1
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -4.8348 -0.6180 -0.0224  0.5419  5.5383
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  schoolid (Intercept)  218.2218 14.7723
##           yearstea       0.6075  0.7794  -0.79
##  Residual             1063.8480 32.6167
## Number of obs: 1190, groups:  schoolid, 107
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  56.3316     2.3247  24.232
## yearstea      0.1005     0.1438   0.699
##
## Correlation of Fixed Effects:
##          (Intr)
## yearstea -0.794
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00810792 (tol = 0.002, component 1)
```

4c.

```
zeta_ <- ranef(lme4)$schoolid
# add b0 and b4
zeta_$b0 <- summary(lme4)$coefficients["(Intercept)", "Estimate"]
zeta_$b4 <- summary(lme4)$coefficients["yearstea", "Estimate"]
# add schoolid
zeta_ <- zeta_ %>% mutate(schoolid = row_number())
```

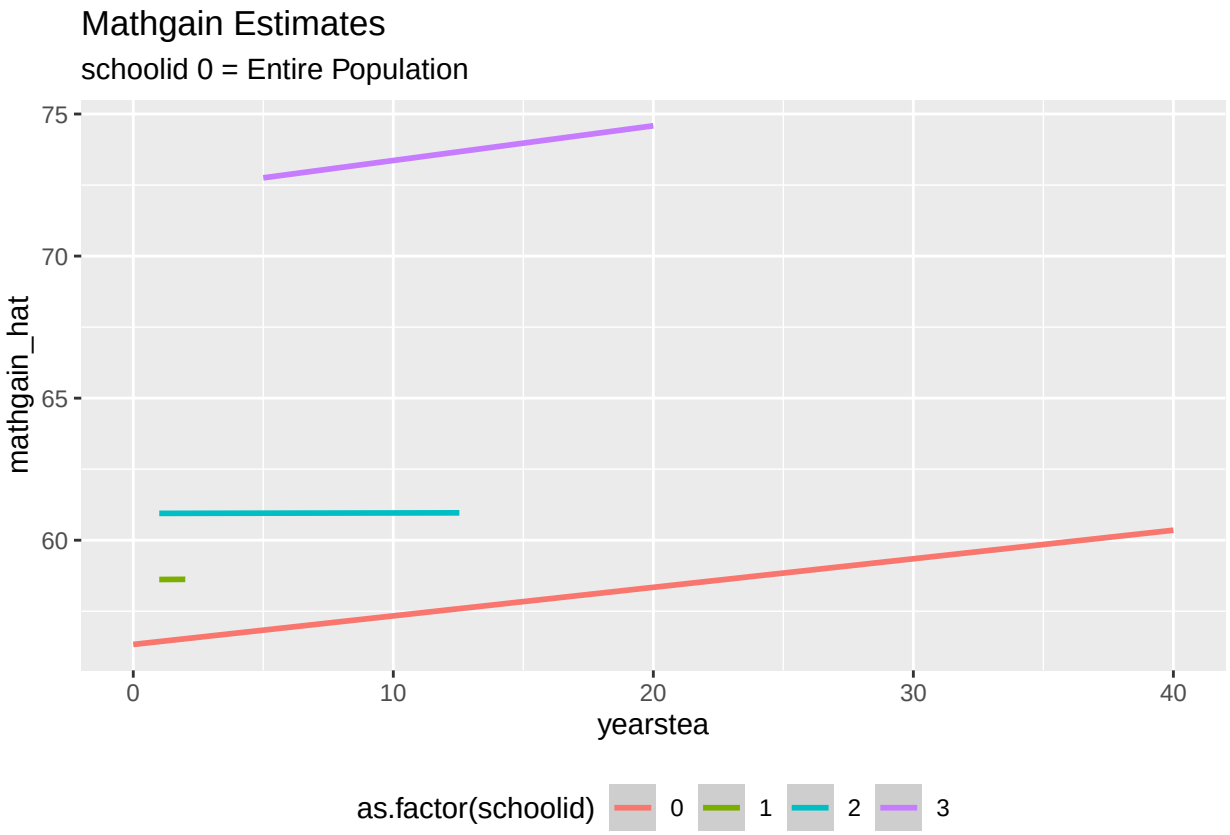
```
# join to class dataset
estimates <- zeta_ %>% left_join(cl %>% select(schoolid, yearstea,
    mathgain) %>% mutate(schoolid = as.integer(schoolid)), by = c(schoolid = "schoolid")) %>%
    mutate(zeta4 = yearstea.x, zeta0 = `(Intercept)`, yearstea = yearstea.y,
        mathgain_hat = b0 + (b4 + zeta4) * yearstea + zeta0) %>%
    select(-c(yearstea.x, `(Intercept)`, yearstea.y))

# entire population
estimates_all <- data.frame(b0 = estimates$b0, b4 = estimates$b4,
    schoolid = 0, mathgain = estimates$mathgain, zeta4 = 0, zeta0 = 0,
    yearstea = estimates$yearstea, mathgain_hat = estimates$b0 +
        estimates$b4 * estimates$yearstea)

# add to dataframe
estimates <- bind_rows(estimates, estimates_all)
```

```
estimates %>% filter(schoolid <= 3) %>% ggplot(aes(x = yearstea,
    y = mathgain_hat, color = as.factor(schoolid))) + geom_smooth(method = "lm") +
    xlim(0, 40) + labs(title = "Mathgain Estimates", subtitle = "schoolid 0 = Entire Population") +
    theme(legend.position = "bottom")
```

```
## `geom_smooth()` using formula 'y ~ x'
```


Mathgain Estimates
schoolid 0 = Entire Population
mathgain_hat
75
70
65
60
0
10
20
30
40
yearstea
as.factor(schoolid)
0
1
2
3