LabN

Kate Brandt

March 26, 2019

```
## KB Mar 25 2019
## Lab N - SOCI 709

# R code
# the data used here is in the Lecture N dropbox folder


rm(list=ls())
library(readstata13)
library(dplyr)
library(tidyverse)
library(lme4)
library(plm)

getwd()
```

[1] "C:/Users/kbran/OneDrive/Documents/R/soci709"

```
hsb2 <- read.dta13("./data/hsb2.dta")
save(hsb2, file="hsb2.RData")

head(hsb2)
```

minority female ses mathach size sector pracad disclim himinty 1 0 1 -1.09361696 5.876 842 0 0.35 1.597 0 2 0 1 -0.15361702 19.708 842 0 0.35 1.597 0 3 0 0 -0.09361702 20.349 842 0 0.35 1.597 0 4 0 0 -0.23361701 8.781 842 0 0.35 1.597 0 5 0 0 0.27638298 17.898 842 0 0.35 1.597 0 6 0 0 0.45638299 4.583 842 0 0.35 1.597 0 schoolid avg_ses 1 1224 -0.434383 2 1224 -0.434383 3 1224 -0.434383 4 1224 -0.434383 5 1224 -0.434383 6 1224 -0.434383

```
# m1,  a constant only linear model (to see if it works)
mod.m1 <- lm(mathach ~1  , data=hsb2 )
summary(mod.m1)
```

Call: lm(formula = mathach ~ 1, data = hsb2)

Residuals: Min 1Q Median 3Q Max -15.5799 -5.4729 0.3831 5.5691 12.2451

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.74785 0.08115 157.1 <2e-16 *** — Signif. codes: 0 '' ***0.001*** '' *0.01* '' 0.05 '.' 0.1 '' 1

Residual standard error: 6.878 on 7184 degrees of freedom

```
# m2
mod.m2 <- lmer(mathach ~ (1 | schoolid),  data=hsb2, REML=FALSE)
summary(mod.m2)
```

Linear mixed model fit by maximum likelihood ['lmerMod'] Formula: mathach ~ (1 | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

47121.8 47142.4 -23557.9 47115.8 7182

Scaled residuals: Min 1Q Median 3Q Max -3.06262 -0.75365 0.02676 0.76070 2.74184

Random effects: Groups Name Variance Std.Dev. schoolid (Intercept) 8.553 2.925
Residual 39.148 6.257
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 12.6371 0.2436 51.87

```
# m3
mod.m3 <- lmer(mathach ~ ses+ (1 | schoolid),  data=hsb2, REML=FALSE)
summary(mod.m3)
```

Linear mixed model fit by maximum likelihood ['lmerMod'] Formula: mathach ~ ses + (1 | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46728.4 46755.9 -23360.2 46720.4 7181

Scaled residuals: Min 1Q Median 3Q Max -3.09692 -0.73195 0.01945 0.75738 2.91422

Random effects: Groups Name Variance Std.Dev. schoolid (Intercept) 8.612 2.935
Residual 37.005 6.083
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 12.6362 0.2437 51.85 ses 2.1912 0.1086 20.17

Correlation of Fixed Effects: (Intr) ses 0.000

```
# m4
mod.m4 <- lmer(mathach ~ ses + (1 + ses | schoolid),  data=hsb2, REML=FALSE)
summary(mod.m4)
```

Linear mixed model fit by maximum likelihood ['lmerMod'] Formula: mathach ~ ses + (1 + ses | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46723.0 46764.3 -23355.5 46711.0 7179

Scaled residuals: Min 1Q Median 3Q Max -3.09688 -0.73198 0.01794 0.75445 2.89902

Random effects: Groups Name Variance Std.Dev. Corr schoolid (Intercept) 8.6206 2.9361
ses 0.6782 0.8235 0.02 Residual 36.7000 6.0581
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 12.6363 0.2437 51.85 ses 2.1932 0.1278 17.16

Correlation of Fixed Effects: (Intr) ses 0.009

```
# m4b: note that this modedl differs from model m4 becasuse
# in m4b to covariance between the random intercept and the random coefficient for
# ses is set to 0, whil in m4 they are allowed to be correlated (see the corresponding models in Stata)

mod.m4b <- lmer(mathach ~ ses + (1 | schoolid) + (0+ ses | schoolid),  data=hsb2, REML=FALSE)
summary(mod.m4b)
```

Linear mixed model fit by maximum likelihood ['lmerMod'] Formula: mathach ~ ses + (1 | schoolid) + (0 + ses | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46721.0 46755.4 -23355.5 46711.0 7180

Scaled residuals: Min 1Q Median 3Q Max -3.09615 -0.73103 0.01747 0.75419 2.90146

Random effects: Groups Name Variance Std.Dev. schoolid (Intercept) 8.6202 2.9360
schoolid.1 ses 0.6783 0.8236
Residual 36.7000 6.0581
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 12.6361 0.2437 51.85 ses 2.1926 0.1278 17.15

Correlation of Fixed Effects: (Intr) ses 0.000

```
# m5
mod.m5 <- lmer(mathach ~ ses + female+ (1 + ses | schoolid),  data=hsb2, REML=FALSE)
summary(mod.m5)
```

Linear mixed model fit by maximum likelihood [‘lmerMod’] Formula: mathach ~ ses + female + (1 + ses | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46675.7 46723.9 -23330.9 46661.7 7178

Scaled residuals: Min 1Q Median 3Q Max -3.13508 -0.73249 0.02731 0.75825 2.81968

Random effects: Groups Name Variance Std.Dev. Corr schoolid (Intercept) 8.220 2.8670
ses 0.627 0.7918 0.04 Residual 36.496 6.0412
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 13.2489 0.2538 52.20 ses 2.1544 0.1264 17.05 female -1.1757 0.1670 -7.04

Correlation of Fixed Effects: (Intr) ses
ses 0.003
female -0.342 0.044 convergence code: 0 Model failed to converge with max|grad| = 0.00273473 (tol = 0.002, component 1)

```
# m6
mod.m6 <- lmer(mathach ~ ses + female + avg_ses + (1 + ses | schoolid),  data=hsb2, REML=FALSE)
summary(mod.m6)
```

Linear mixed model fit by maximum likelihood [‘lmerMod’] Formula: mathach ~ ses + female + avg_ses + (1 + ses | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46518 46573 -23251 46502 7177

Scaled residuals: Min 1Q Median 3Q Max -3.1624 -0.7338 0.0247 0.7615 2.8896

Random effects: Groups Name Variance Std.Dev. Corr schoolid (Intercept) 2.4531 1.5662
ses 0.6203 0.7876 -0.19 Residual 36.5090 6.0423
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 13.2915 0.1671 79.547 ses 2.1530 0.1262 17.066 female -1.1675 0.1625 -7.186 avg_ses 5.7923 0.3478 16.653

Correlation of Fixed Effects: (Intr) ses female ses -0.091
female -0.507 0.042
avg_ses -0.011 0.005 0.039

```
hsb2$c_ses <- hsb2$ses*hsb2$avg_ses

hsb2 %>%
  filter(!is.na(c_ses)) %>%
  summarize(mean(c_ses), min(c_ses), max(c_ses), n=n())
```

```
mean(c_ses) min(c_ses) max(c_ses)    n
```

1 -1.346332e-10 -2.019714 1.520184 7185

```
hsb2 %>%
  slice(which.max(c_ses))
```

minority female ses mathach size sector pracad disclim himinty 1 1 0 -1.448542 1.927 643 0 0.5 0.715 1 schoolid avg_ses c_ses 1 4458 -1.049458 1.520184

```
# m7
mod.m7 <- lmer(mathach ~ ses + female + avg_ses + c_ses + (1 + ses | schoolid),  data=hsb2, REML=FALSE)
summary(mod.m7)
```

Linear mixed model fit by maximum likelihood [‘lmerMod’] Formula: mathach ~ ses + female + avg_ses + c_ses + (1 + ses | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46519.0 46581.0 -23250.5 46501.0 7176

Scaled residuals: Min 1Q Median 3Q Max -3.1771 -0.7336 0.0234 0.7600 2.8783

Random effects: Groups Name Variance Std.Dev. Corr schoolid (Intercept) 2.4527 1.5661
ses 0.5974 0.7729 -0.19 Residual 36.5117 6.0425
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 13.2922 0.1671 79.554 ses 2.1601 0.1258 17.176 female -1.1691 0.1625 -7.196 avg_ses 5.7655 0.3489 16.525 c_ses 0.3062 0.3130 0.979

Correlation of Fixed Effects: (Intr) ses female avg_ss ses -0.090
female -0.507 0.042
avg_ses -0.011 0.001 0.040
c_ses 0.003 0.058 -0.008 -0.079

```
# m8
hsb2$c_female <- hsb2$avg_ses*hsb2$female
# note m8a tests adding the random coefficient for female first.  The random coefficient s.e. are diffe
mod.m8a <- lmer(mathach ~ ses + female + (1 + female | schoolid),  data=hsb2, REML=FALSE)
summary(mod.m8a)
```

Linear mixed model fit by maximum likelihood [‘lmerMod’] Formula: mathach ~ ses + female + (1 + female | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46680.4 46728.5 -23333.2 46666.4 7178

Scaled residuals: Min 1Q Median 3Q Max -3.14924 -0.73331 0.03361 0.75395 2.82044

Random effects: Groups Name Variance Std.Dev. Corr schoolid (Intercept) 8.8084 2.9679
female 0.8956 0.9464 -0.33 Residual 36.6249 6.0518
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 13.2570 0.2627 50.461 ses 2.1571 0.1084 19.899 female -1.2200 0.1873 -6.515

Correlation of Fixed Effects: (Intr) ses
ses -0.016
female -0.433 0.044

```
mod.m8 <- lmer(mathach ~ ses + female + avg_ses + c_female + (1 + ses + female | schoolid),  data=hsb2,
summary(mod.m8)
```

Linear mixed model fit by maximum likelihood [‘lmerMod’] Formula: mathach ~ ses + female + avg_ses + c_female + (1 + ses + female |
schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46527.4 46610.0 -23251.7 46503.4 7173

Scaled residuals: Min 1Q Median 3Q Max -3.1728 -0.7315 0.0270 0.7590 2.8848

Random effects: Groups Name Variance Std.Dev. Corr
schoolid (Intercept) 3.19037 1.7862
ses 0.02311 0.1520 -1.00
female 0.90108 0.9492 -0.61 0.61 Residual 36.62759 6.0521
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 13.29523 0.18313 72.599 ses 2.15606 0.10908 19.765 female -1.20927 0.18148 -6.664 avg_ses 5.78621 0.44947 12.873 c_female -0.02468 0.45335 -0.054

Correlation of Fixed Effects: (Intr) ses female avg_ss ses -0.108
female -0.635 0.070
avg_ses -0.018 0.012 0.017
c_female 0.019 -0.012 0.019 -0.651 convergence code: 0 boundary (singular) fit: see ?isSingular

```
# m9 ...didn't converge with REML=FALSE
hsb2$ses_female <- hsb2$ses*hsb2$female
mod.m9 <- lmer(mathach ~ ses + female + avg_ses + c_female + ses_female + (1 + ses + female | schoolid)
summary(mod.m9)
```

Linear mixed model fit by maximum likelihood [‘lmerMod’] Formula: mathach ~ ses + female + avg_ses + c_female + ses_female + (1 +
ses + female | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46527.2 46616.7 -23250.6 46501.2 7172

Scaled residuals: Min 1Q Median 3Q Max -3.1885 -0.7339 0.0275 0.7599 2.8949

Random effects: Groups Name Variance Std.Dev. Corr
schoolid (Intercept) 3.18561 1.7848
ses 0.02044 0.1430 -1.00
female 0.89822 0.9477 -0.61 0.61 Residual 36.61700 6.0512
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 13.30296 0.18311 72.651 ses 1.98899 0.15696 12.672 female -1.21393 0.18144 -6.691 avg_ses 5.78364 0.44947 12.868 c_female -0.02481 0.45326 -0.055 ses_female 0.32251 0.21803 1.479

Correlation of Fixed Effects: (Intr) ses female avg_ss c_feml ses -0.092
female -0.635 0.060
avg_ses -0.018 0.008 0.017
c_female 0.019 -0.008 0.019 -0.650
ses_female 0.028 -0.720 -0.018 0.000 -0.001 convergence code: 0 boundary (singular) fit: see ?isSingular

```
# try it by removing the REML=FALSE
mod.m9 <- lmer(mathach ~ ses + female + avg_ses + c_female + ses_female + (1 + ses + female | schoolid)
summary(mod.m9)
```

Linear mixed model fit by REML [‘lmerMod’] Formula: mathach ~ ses + female + avg_ses + c_female + ses_female + (1 +
ses + female | schoolid) Data: hsb2

REML criterion at convergence: 46499.1

Scaled residuals: Min 1Q Median 3Q Max -3.1896 -0.7286 0.0282 0.7524 2.9706

Random effects: Groups Name Variance Std.Dev. Corr
schoolid (Intercept) 3.2453 1.8015
ses 0.6164 0.7851 -0.04
female 0.9782 0.9890 -0.60 -0.35 Residual 36.3487 6.0290
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 13.28427 0.18415 72.139 ses 1.97081 0.17156 11.488
female -1.18094 0.18310 -6.450 avg_ses 5.75387 0.45434 12.664 c_female 0.01926 0.45852 0.042 ses_female
0.35423 0.22570 1.569

Correlation of Fixed Effects: (Intr) ses female avg_ss c_feml ses -0.034
female -0.634 -0.023
avg_ses -0.018 0.007 0.017
c_female 0.020 -0.005 0.018 -0.654
ses_female 0.014 -0.679 -0.011 -0.003 0.001 convergence code: 0 Model failed to converge with max|grad| =
0.0032637 (tol = 0.002, component 1)

```
# note: this didn't converge in R...a "singular fit"...need to alter the model restricting the RE covar
mod.m9 <- lmer(mathach ~ ses + female +   + avg_ses + c_female + ses_female + (1  | schoolid) + (0 + se
                 (0 + female | schoolid),  data=hsb2, REML=FALSE)
summary(mod.m9)
```

Linear mixed model fit by maximum likelihood [‘lmerMod’] Formula: mathach ~ ses + female + +avg_ses +
c_female + ses_female + (1 |
schoolid) + (0 + ses | schoolid) + (0 + female | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46520.4 46589.2 -23250.2 46500.4 7175

Scaled residuals: Min 1Q Median 3Q Max -3.1688 -0.7362 0.0252 0.7550 2.9090

Random effects: Groups Name Variance Std.Dev. schoolid (Intercept) 2.4373 1.5612
schoolid.1 ses 0.6272 0.7920
schoolid.2 female 0.1130 0.3362
Residual 36.4678 6.0389
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 13.29974 0.16703 79.623 ses 1.97089 0.17192 11.464
female -1.17165 0.16516 -7.094 avg_ses 5.74441 0.41197 13.944 c_female 0.03404 0.41504 0.082 ses_female
0.35596 0.22644 1.572

Correlation of Fixed Effects: (Intr) ses female avg_ss c_feml ses -0.031
female -0.500 0.033
avg_ses -0.023 0.008 0.023
c_female 0.025 -0.007 0.022 -0.526
ses_female 0.021 -0.678 -0.004 -0.005 0.000

```
# m10
table(hsb2$sector)
```

0 1 3642 3543

```
hsb2$cath_female <- hsb2$sector*hsb2$female
mod.m10 <- lmer(mathach ~ ses + female + avg_ses + sector + cath_female + (1 + ses + female | schoolid)
summary(mod.m10)
```

Linear mixed model fit by maximum likelihood [‘lmerMod’] Formula: mathach ~ ses + female + avg_ses + sector + cath_female + (1 +
ses + female | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46510.1 46599.5 -23242.0 46484.1 7172

Scaled residuals: Min 1Q Median 3Q Max -3.1687 -0.7358 0.0256 0.7585 2.9172

Random effects: Groups Name Variance Std.Dev. Corr
schoolid (Intercept) 2.90371 1.7040
ses 0.09585 0.3096 1.00
female 1.03374 1.0167 -0.64 -0.64 Residual 36.58149 6.0483
Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 12.59542 0.23714 53.115 ses 2.15818 0.11120 19.407 female -1.21315 0.23093 -5.253 avg_ses 4.98045 0.34214 14.557 sector 1.57806 0.37399 4.220 cath_female -0.01591 0.37802 -0.042

Correlation of Fixed Effects: (Intr) ses female avg_ss sector ses 0.105
female -0.626 -0.009
avg_ses 0.193 -0.009 0.017
sector -0.658 0.008 0.393 -0.286
cath_female 0.387 -0.012 -0.610 0.026 -0.652 convergence code: 0 boundary (singular) fit: see ?isSingular

```
mod.m10 <- lmer(mathach ~ ses + female + avg_ses + sector + cath_female + (1 | schoolid) + (0 + ses | s
                  (0 + female | schoolid), REML = FALSE, data=hsb2)
summary(mod.m10)
```

Linear mixed model fit by maximum likelihood [‘lmerMod’] Formula: mathach ~ ses + female + avg_ses + sector + cath_female + (1 |
schoolid) + (0 + ses | schoolid) + (0 + female | schoolid) Data: hsb2

```
 AIC      BIC   logLik deviance df.resid
```

46505.5 46574.3 -23242.7 46485.5 7175

Scaled residuals: Min 1Q Median 3Q Max -3.1735 -0.7341 0.0271 0.7594 2.9038

Random effects: Groups Name Variance Std.Dev. schoolid (Intercept) 2.110556 1.45278 schoolid.1 ses 0.625995 0.79120 schoolid.2 female 0.004078 0.06386 Residual 36.505524 6.04198 Number of obs: 7185, groups: schoolid, 160

Fixed effects: Estimate Std. Error t value (Intercept) 12.7719 0.2183 58.495 ses 2.1521 0.1264 17.033 female -1.2531 0.2033 -6.164 avg_ses 5.2267 0.3537 14.775 sector 1.1523 0.3428 3.361 cath_female 0.1857 0.3353 0.554

Correlation of Fixed Effects: (Intr) ses female avg_ss sector ses -0.020
female -0.485 0.042
avg_ses 0.217 0.001 0.015
sector -0.676 0.007 0.306 -0.321
cath_female 0.303 -0.014 -0.605 0.032 -0.519

```
# m11
mod.m11 <- lm(mathach ~ ses + female + avg_ses + sector + cath_female ,  data=hsb2)
summary(mod.m11)
```

Call: lm(formula = mathach ~ ses + female + avg_ses + sector + cath_female, data = hsb2)

Residuals: Min 1Q Median 3Q Max -19.9999 -4.6149 0.1507 4.7697 17.8531

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.7730 0.1520 84.016 < 2e-16 ***ses 2.1485 0.1115 19.276 < 2e-16*** female -1.2839 0.2071 -6.199 6.01e-10 ***avg_ses 5.0486 0.1909 26.447 < 2e-16*** sector 1.3788 0.2228 6.189 6.39e-10 *** cath_female -0.1053 0.2949 -0.357 0.721
— Signif. codes: 0 '' ***0.001*** '' *0.01* '' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.233 on 7179 degrees of freedom Multiple R-squared: 0.1793, Adjusted R-squared: 0.1788 F-statistic: 313.8 on 5 and 7179 DF, p-value: < 2.2e-16

```
library(stargazer)
stargazer(mod.m9)
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Tue, Mar 26, 2019 - 9:32:13 AM

Table 1:

	Dependent variable:
	mathach
ses	1.971***
	(0.172)
female	−1.172***
	(0.165)
avg_ses	5.744***
	(0.412)
c_female	0.034
	(0.415)
ses_female	0.356
	(0.226)
Constant	13.300***
	(0.167)
Observations	7,185
Log Likelihood	−23,250.210
Akaike Inf. Crit.	46,520.410
Bayesian Inf. Crit.	46,589.210
Note:	*p<0.1; **p<0.05; ***p<0.01