STAT 196
week 1

Exam Data: Mixed effects models (gender gap)

prclim SFYS (ImList)

normexam $y = \alpha_0 + \alpha_1 \text{sexM} + \varepsilon$ by school

$\hat{\alpha}_0 = \bar{F}$ $\hat{\alpha}_1 = \bar{M} - \bar{F}$

Unconditional Model ggaplmer

Level 1 $y = \alpha_0 + \alpha_1 \text{sexM} + \varepsilon$

Level 2 $\alpha_0 = \gamma_{00} + u_0$ no Level 2 predictors (fixef / ranef)

$\alpha_1 = \gamma_{10} + u_1$

combined model $y \sim \text{sex}$ (fixedeff) [sex|school] random

---

Conditional model ggaplmer2 (correlate schavg)

Level 1 $y = \alpha_0 + \alpha_1 \text{sexM} + \varepsilon$ (random)

Level 2 $\alpha_0 = \gamma_{00} + \gamma_{01} \text{schavg} + u_0$ (fixed)

$\alpha_1 = \gamma_{10} + \gamma_{11} \text{schavg} + u_1$

combined model $y \sim \text{sex} * \text{schavg}$ (fixedeff)

Within-school ancova

ancovalmer

covariate standLRT

schavg = standLRT

Level 1 $y = \alpha_0 + \alpha_1 \text{sexM} + \alpha_2 \text{standLRT} + \varepsilon$ ancova

Level 2 $\alpha_0 = \gamma_{00} + u_0$ $\alpha_1 = \gamma_{10} + u_1$ $\alpha_2 = \gamma_{20} + u_2$

(could have additional predictors)

combined $y \sim \text{sexM} + \text{standLRT}$ (fixed)

---

effects ranef fixef

random - varies over schools (Level 2 units)

fixed - does not vary over schools

see plots

Week 4 p1

Exam Data Stat209 Week 4 multilevel data example

Exam obtained from the mlmRev package, a collection of datasets that Doug Bates uses for illustrating lme4 versus other computing (non-R) options.

The docs for the Exam data

Exam {mlmRev} R Documentation Exam scores from inner London

Description Exam scores of 4,059 students from 65 schools in Inner London.

Format A data frame with 4059 observations on the following 9 variables.

school School ID - a factor.

normexam Normalized exam score.

(note: publication explains: "General Certificate of Secondary Examination (GCSE) grades in mathematics and English")

schgend School gender - a factor. Levels are mixed, boys, and girls.

schavg School average of intake score.

vr Student level Verbal Reasoning (VR) score band at intake - a factor. Levels are bottom 25%, mid 50%, and top 25%.

intake Band of student's intake score - a factor. Levels are bottom 25%, mid 50% and top 25%.

standLRT [DRR note: this is the individual input or 'intake' score] Standardised LR test score.

(note: publication explains these are: "scores on a common reading test taken when theywere 11 years old-the London Reading Test (LRT)"

sex Sex of the student - levels are F and M.

type School type - levels are Mxd and Sngl.

student Student id (within school) - a factor

data list

---

pick off the mixed (i.e. coed) schools (i.e. schools that have both males and female students). The type ("Mxd") indicator gives 35 schools, but two of those appear to be errors (only one student of opposite gender) so the dataframe "mixed" in the output below has 33 schools, 2026 students (1028 males). Data are posted on webpage.

```
> mixed = read.table(file = "[]mExam", header = T)
> str(mixed)
'data.frame':   2169 obs. of  6 variables:
 $ school  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ normexam: num  0.261 0.134 -1.724 0.968 0.544 ...
 $ schavg  : num  0.166 0.166 0.166 0.166 0.166 ...
 $ standLRT: num  0.619 0.206 -1.365 0.206 0.371 ...
 $ sex     : Factor w/ 2 levels "F","M": 1 1 2 1 1 2 2 2 1 2 ...
 $ student : int  143 145 142 141 138 155 158 115 117 113 ...
> attach(mixed)
> table(sex, school)
    school
sex   1   3   4   5   9  10  12  13  14  15  17  19  20  22  23  26  28  32  33
  F  28  23  34  19  13  19  24  38 106  44  95  22  18  42  18  31  26  15  33
  M  45  29  45  16  21  31  23  26  92  47  31  33  21  48  10  44  31  27  44
    school
sex  34  38  42  43  45  46  47  50  51  54  55  56  59  61  62  63
  F   8  23  23  60  48  36   1  38  32   4  25  22  17  29  28  17
  M  18  31  35   1   5  47  81  35  26   4  26  16  30  35  43  13
```

M-F

```
> mixed = read.table(file = "[]mExam4347", header = T)
> attach(mixed)
> table(sex, school)
    school
sex   1   3   4   5   9  10  12  13  14  15  17  19  20  22  23  26  28  32  33  34
  F  28  23  34  19  13  19  24  38 106  44  95  22  18  42  18  31  26  15  33   8
  M  45  29  45  16  21  31  23  26  92  47  31  33  21  48  10  44  31  27  44  18

     38  42  45  46  50  51  54  55  56  59  61  62  63
     23  23  48  36  38  32   4  25  22  17  29  28  17
     31  35   5  47  35  26   4  26  16  30  35  43  13
                              small
> library(lme4)
> ggaplist = lmList(normexam ~ sex|school, data = mixed)
> gapCoef = coef(ggaplist)
> boxplot(gapCoef[,2])
```

$\bar{M} - \bar{F}$ for each school

grab coeff

see

norm ~ sex gives fem mean, male-fem for each school

Exam p2

```
> fivenum(gapCoef[,2])
[1] -0.6222566 -0.4010070 -0.2792353 -0.1167029  0.3610921
> sum(gapCoef[,2] > 0)    > sum(gapCoef[,2] < 0)
[1] 4                     [1] 29
> gapCoef
```

from lmList

	(Intercept)	sexM
1	0.74840566	-0.40100698
3	0.91842969	-0.11293860
4	0.40687566	-0.58503387
5	0.28464371	0.26023584
9	-0.26780577	-0.27179984
10	-0.06706601	-0.32633009
12	0.19844985	-0.55363628
13	-0.05019206	-0.48137258
14	0.15614029	-0.30253287
15	0.01967218	-0.11670288
17	-0.17425506	-0.28927331
19	0.57963902	-0.62225662
20	0.52556657	-0.04916964
22	-0.37649653	-0.22819645
23	-0.82430562	0.24268458
26	-0.22517959	-0.28438400
28	-0.72084548	-0.27923534
32	-0.07872164	-0.45517095
33	0.10602805	-0.05060315
34	-0.62088143	0.36109208
38	-0.17153152	-0.20208823
42	0.25940135	-0.39227368
45	-0.21855833	-0.07363659
46	-0.63300842	0.31011109
50	-0.25973997	-0.12939801
51	-0.02976871	-0.55390325
54	-0.55549500	-0.15626770
55	0.96429515	-0.48483622
56	0.18051960	-0.51014920
59	-0.83275722	-0.33891689
61	0.07845038	-0.23914873
62	0.22257551	-0.30478026
63	0.84624148	-0.25514608

SFYS

```
> tapply(normexam, list(school,sex), mean)
```

	F	M
1	0.74840566	0.34739868
3	0.91842969	0.80549108
4	0.40687566	-0.17815821
5	0.28464371	0.54487955
9	-0.26780577	-0.53960561
10	-0.06706601	-0.39339610
12	0.19844985	-0.35518643
13	-0.05019206	-0.53156464
14	0.15614029	-0.14639258
15	0.01967218	-0.09703070
17	-0.17425506	-0.46352837
19	0.57963902	-0.04261761
20	0.52556657	0.47639692
22	-0.37649653	-0.60469298
23	-0.82430562	-0.58162104
26	-0.22517959	-0.50956360
28	-0.72084548	-1.00008081
32	-0.07872164	-0.53389259
33	0.10602805	0.05542490
34	-0.62088142	-0.25978934
38	-0.17153152	-0.37361975
42	0.25940135	-0.13287234
45	-0.21855833	-0.29219492
46	-0.63300842	-0.32289733
50	-0.25973997	-0.38913798
51	-0.02976871	-0.58367196
54	-0.55549500	-0.71176270
55	0.96429515	0.47945893
56	0.18051960	-0.32962960
59	-0.83275722	-1.17167411
61	0.07845038	-0.16069835
62	0.22257551	-0.08220474
63	0.84624148	0.59109540

school means

w/in school

```
> ggaplmer = lmer(normexam ~ sex + (sex|school), data = mixed)
> summary(ggaplmer)
Linear mixed model fit by REML ['lmerMod']
Formula: normexam ~ sex + (sex | school)   Data: mixed
REML criterion at convergence: 5423.9
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 school   (Intercept) 0.196754 0.44357
          sexM        0.001262 0.03553  -1.00
 Residual             0.814182 0.90232
Number of obs: 2026, groups:  school, 33
Fixed effects:
            Estimate Std. Error t value
(Intercept)  0.02467    0.08334   0.296   γ̂00
sexM        -0.25886    0.04165  -6.214   γ̂10
```

mixed effects models

$$Y = \alpha_0 + \alpha_1 \text{isMale} + \varepsilon$$

$$\alpha_0 = \gamma_{00} + u_0$$

$$\alpha_1 = \gamma_{10} + u_1$$

no level-2 predictors

random - varies over units (schools)

fixed does not vary

compare

```
> mean(gapCoef[,1])
[1] 0.01177957
> mean(gapCoef[,2])
[1] -0.2386686
```

```
> confint(ggaplmer)
Computing profile confidence intervals ...
                 2.5 %     97.5 %
.sig01       0.3295245  0.5891989
.sig02      -1.0000000  1.0000000
.sig03       0.0000000  0.1794380
.sigma       0.8747990  0.9308519
(Intercept) -0.1416656  0.1903270
sexM        -0.3405181 -0.1766122
```

Inference (no p-values, afex package etc)

```
> str(mixed)
'data.frame':   2026 obs. of  6 variables:
 $ school  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ normexam: num  0.261 0.134 -1.724 0.968 0.544 ...
 $ schavg  : num  0.166 0.166 0.166 0.166 0.166 ...
 $ standLRT: num  0.619 0.206 -1.365 0.206 0.371 ...
 $ sex     : Factor w/ 2 levels "F","M": 1 1 2 1 1 2 2 2 1 2 ...
 $ student : int  143 145 142 141 138 155 158 115 117 113 ...
```

Is "ggap" associated w/ input vars?

```
> ggaplmer2 = lmer(normexam ~ sex*schavg + (sex|school), data = mixed)
```

```
> summary(ggaplmer2)
Linear mixed model fit by REML ['lmerMod']
Formula: normexam ~ sex * schavg + (sex | school)
   Data: mixed
REML criterion at convergence: 5406.3

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 school   (Intercept) 0.098371 0.31364
          sexM        0.003117 0.05583  -1.00
 Residual             0.815375 0.90298
Number of obs: 2026, groups:  school, 33

Fixed effects:
            Estimate Std. Error t value
(Intercept)  0.06209    0.06325   0.982
sexM        -0.26167    0.04233  -6.181
schavg       0.95884    0.20480   4.682
sexM:schavg -0.00288    0.14197  -0.020

Correlation of Fixed Effects:
            (Intr) sexM   schavg
sexM        -0.536
schavg       0.099 -0.027
sexM:schavg -0.029  0.023 -0.563
```

$$Y = \alpha_0 + \alpha_1 \text{isMale} + \varepsilon$$

$$\alpha_0 = \gamma_{00} + \gamma_{01}\,\text{schavg} + u_0$$

$$\alpha_1 = \gamma_{10} + \gamma_{11}\,\text{schavg} + u_1$$

"conditional model"

exercise: write out combined model

$\hat{\gamma}_{00}$ $\hat{\gamma}_{10}$ $\hat{\gamma}_{01}$ $\hat{\gamma}_{11}$

```
> confint(ggaplmer2, method = "boot", nsim = 1000, boot.type = "perc")
Computing bootstrap confidence intervals ...
                                    2.5 %     97.5 %
sd_(Intercept)|school         0.208978228  0.4213861
cor_sexM.(Intercept)|school -1.000000000  1.0000000
sd_sexM|school                0.006550685  0.1881426
sigma                         0.872457820  0.9310742
(Intercept)                  -0.073254826  0.1873577
sexM                         -0.339567709 -0.1741492
schavg                        0.566154943  1.3559651
sexM:schavg                  -0.290209374  0.2666750
```

profile bombed here

fixed random

Inference

```
> anova(ggaplmer,ggaplmer2)
refitting model(s) with ML (instead of REML)
Data: mixed
Models:
ggaplmer: normexam ~ sex + (sex | school)
ggaplmer2: normexam ~ sex * schavg + (sex | school)
          Df    AIC    BIC logLik deviance  Chisq Chi Df Pr(>Chisq)
ggaplmer   6 5428.1 5461.7  -2708   5416.1
ggaplmer2  8 5409.9 5454.9  -2697   5393.9 22.109      2  1.582e-05 ***
```

compare nested model fits