April 19th
linear mixed effects models

```
> library(lme4)
Loading required package: Matrix
Loading required package: lattice
> library(nlme)
> MathAchieve[1:10,]
Grouped Data: MathAch ~ SES | School
    School Minority    Sex    SES MathAch MEANSES
1     1224       No Female -1.528   5.876  -0.428
2     1224       No Female -0.588  19.708  -0.428
3     1224       No   Male -0.528  20.349  -0.428
4     1224       No   Male -0.668   8.781  -0.428
5     1224       No   Male -0.158  17.898  -0.428
6     1224       No   Male  0.022   4.583  -0.428
7     1224       No Female -0.618  -2.832  -0.428
8     1224       No   Male -0.998   0.523  -0.428
9     1224       No Female -0.888   1.527  -0.428
10    1224       No   Male -0.458  21.521  -0.428
```

Complete pooling

```
> lm0 <- lm(MathAch ~ SES, data=MathAchieve)
> summary(lm0)

Call:
lm(formula = MathAch ~ SES, data = MathAchieve)

Residuals:
     Min       1Q    Median       3Q      Max
-19.4382  -4.7580    0.2334   5.0649  15.9007

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.74740    0.07569  168.42   <2e-16 ***
SES          3.18387    0.09712   32.78   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.416 on 7183 degrees of freedom
Multiple R-squared: 0.1301, Adjusted R-squared:  0.13
F-statistic:  1075 on 1 and 7183 DF,  p-value: < 2.2e-16
```
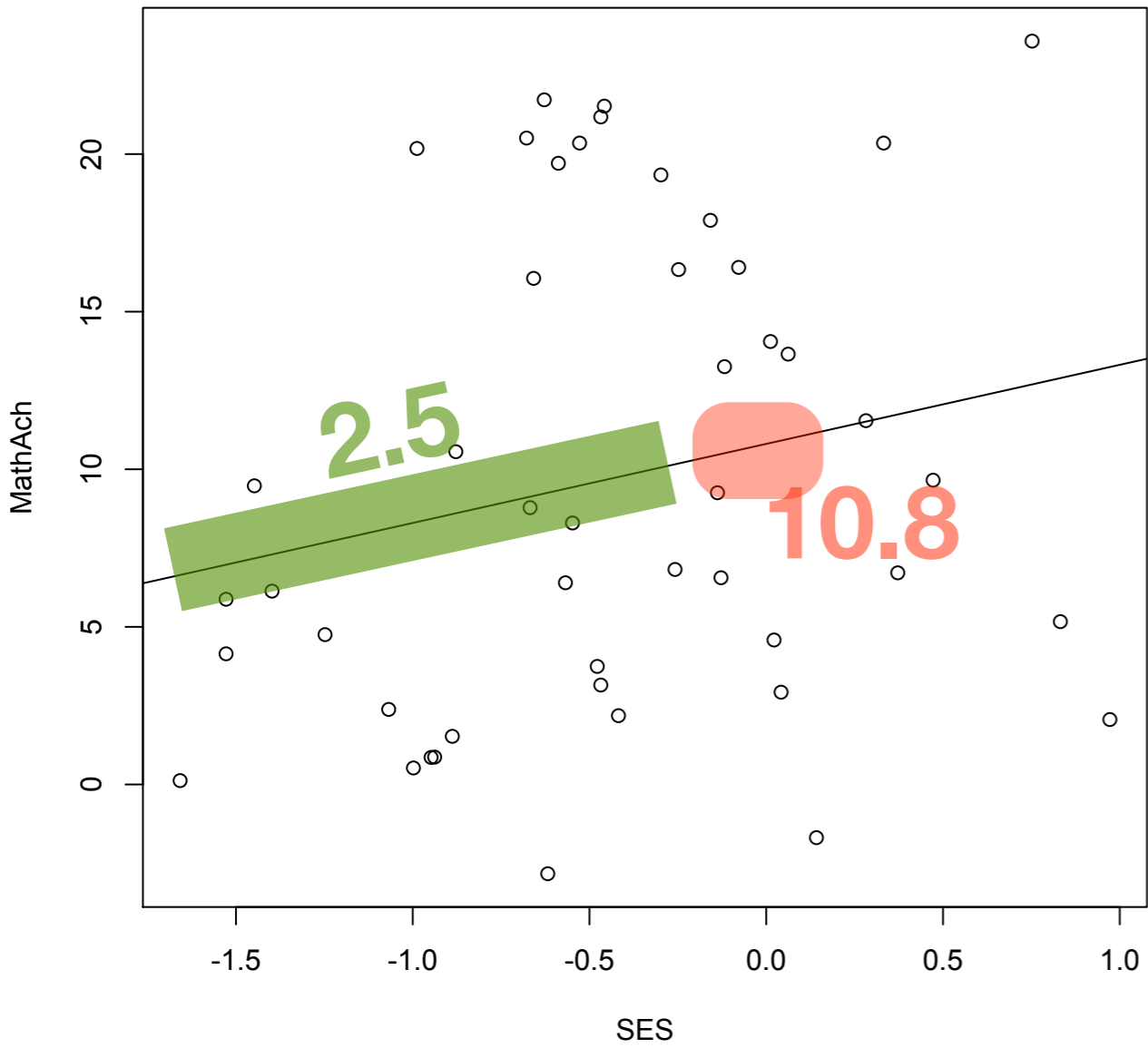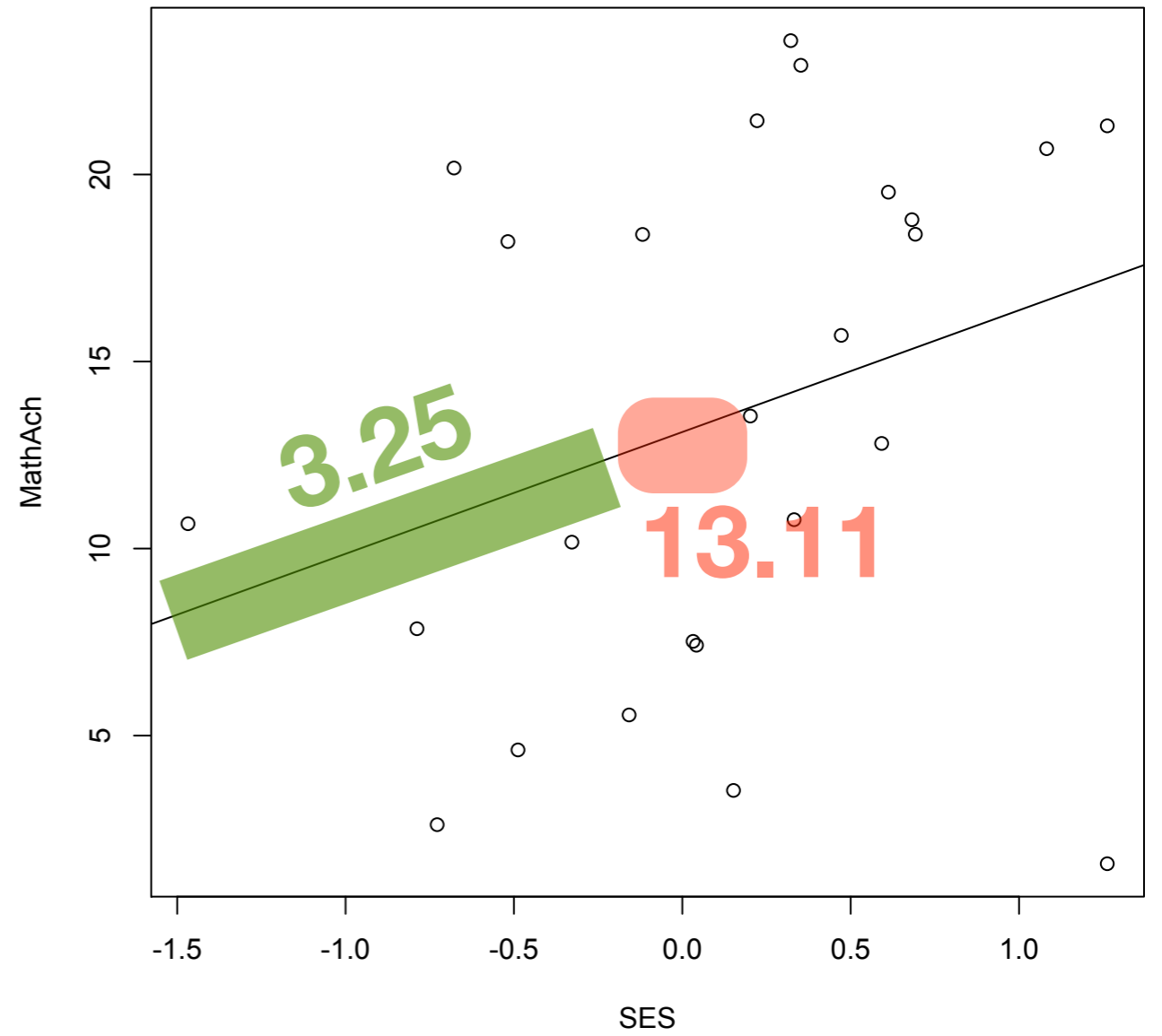
No pooling

School 1224

MathAch / SES

2.5   10.8

School 1228

MathAch / SES

3.25   13.11

160 separate models

$$Y_{ij} = \beta_{oj} + \beta_{1j}(X_{ij} - \bar{X}_{.j}) + \epsilon_{ij}$$

MathAchievement

0-SES achievement level for school j

per-SES improvement at school j

error term

Random effects

$$Y_{ij} = \beta_{oj} + \beta_{1j}(X_{ij} - \bar{X}_{.j}) + \epsilon_{ij}$$

dependent on something else

```
> attach(MathAchSchool)
> MathScores <- merge(MathAchieve,MathAchSchool,  by="School")
> head(MathScores)
  School Minority    Sex     SES MathAch MEANSES.x Size Sector PRACAD DISCLIM HIMINTY MEANSES.y
1   1224       No Female -1.528   5.876    -0.428  842 Public   0.35   1.597       0    -0.428
2   1224       No Female -0.588  19.708    -0.428  842 Public   0.35   1.597       0    -0.428
3   1224       No   Male -0.528  20.349    -0.428  842 Public   0.35   1.597       0    -0.428
4   1224       No   Male -0.668   8.781    -0.428  842 Public   0.35   1.597       0    -0.428
5   1224       No   Male -0.158  17.898    -0.428  842 Public   0.35   1.597       0    -0.428
6   1224       No   Male  0.022   4.583    -0.428  842 Public   0.35   1.597       0    -0.428
```

```
> attach(MathAchSchool)
> MathScores <- merge(MathAchieve,MathAchSchool,  by="School")
> head(MathScores)
  School Minority    Sex    SES MathAch MEANSES.x Size Sector PRACAD DISCLIM HIMINTY MEANSES.y
1   1224       No Female -1.528   5.876    -0.428  842 Public   0.35   1.597       0    -0.428
2   1224       No Female -0.588  19.708    -0.428  842 Public   0.35   1.597       0    -0.428
3   1224       No   Male -0.528  20.349    -0.428  842 Public   0.35   1.597       0    -0.428
4   1224       No   Male -0.668   8.781    -0.428  842 Public   0.35   1.597       0    -0.428
5   1224       No   Male -0.158  17.898    -0.428  842 Public   0.35   1.597       0    -0.428
6   1224       No   Male  0.022   4.583    -0.428  842 Public   0.35   1.597       0    -0.428
```

vs Catholic

```
> attach(MathAchSchool)
> MathScores <- merge(MathAchieve,MathAchSchool,  by="School")
> head(MathScores)
  School Minority    Sex     SES MathAch MEANSES.x Size Sector PRACAD DISCLIM HIMINTY MEANSES.y
1   1224       No Female  -1.528   5.876    -0.428  842 Public   0.35   1.597       0    -0.428
2   1224       No Female  -0.588  19.708    -0.428  842 Public   0.35   1.597       0    -0.428
3   1224       No   Male  -0.528  20.349    -0.428  842 Public   0.35   1.597       0    -0.428
4   1224       No   Male  -0.668   8.781    -0.428  842 Public   0.35   1.597       0    -0.428
5   1224       No   Male  -0.158  17.898    -0.428  842 Public   0.35   1.597       0    -0.428
6   1224       No   Male   0.022   4.583    -0.428  842 Public   0.35   1.597       0    -0.428
```

Catholicism indicator, W_j **vs** `Catholic`

```
> attach(MathAchSchool)
> MathScores <- merge(MathAchieve,MathAchSchool,  by="School")
> head(MathScores)
  School Minority     Sex     SES MathAch MEANSES.x Size Sector PRACAD DISCLIM HIMINTY MEANSES.y
1   1224       No  Female -1.528   5.876    -0.428  842 Public   0.35   1.597       0    -0.428
2   1224       No  Female -0.588  19.708    -0.428  842 Public   0.35   1.597       0    -0.428
3   1224       No    Male -0.528  20.349    -0.428  842 Public   0.35   1.597       0    -0.428
4   1224       No    Male -0.668   8.781    -0.428  842 Public   0.35   1.597       0    -0.428
5   1224       No    Male -0.158  17.898    -0.428  842 Public   0.35   1.597       0    -0.428
6   1224       No    Male  0.022   4.583    -0.428  842 Public   0.35   1.597       0    -0.428
```

Catholicism indicator, W_j **vs** `Catholic`

$$\beta_{oj} = \gamma_{00} + \gamma_{01}W_j + \epsilon_{oj}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + \epsilon_{1j}$$

Tell lme4 what to do

separate random effects
for intercept β_{oj}
and slope β_{1j}

```
> lme1.fm <- lmer(MathAch ~ SES + Sector + (1+ SES | School),MathScores)
```

```
> install.packages("arm")
> library(arm)
> display(lme1.fm)
lmer(formula = MathAch ~ SES + Sector + (1 + SES | School), data = MathScores)
            coef.est coef.se
(Intercept)  14.01     0.26
SES           2.39     0.12
SectorPublic -2.54     0.34

Error terms:
 Groups    Name         Std.Dev. Corr
 School    (Intercept)  1.99
           SES          0.66          0.55
 Residual               6.07
---
number of obs: 7185, groups: School, 160
AIC = 46615.9, DIC = 46593
deviance = 46597.4
```
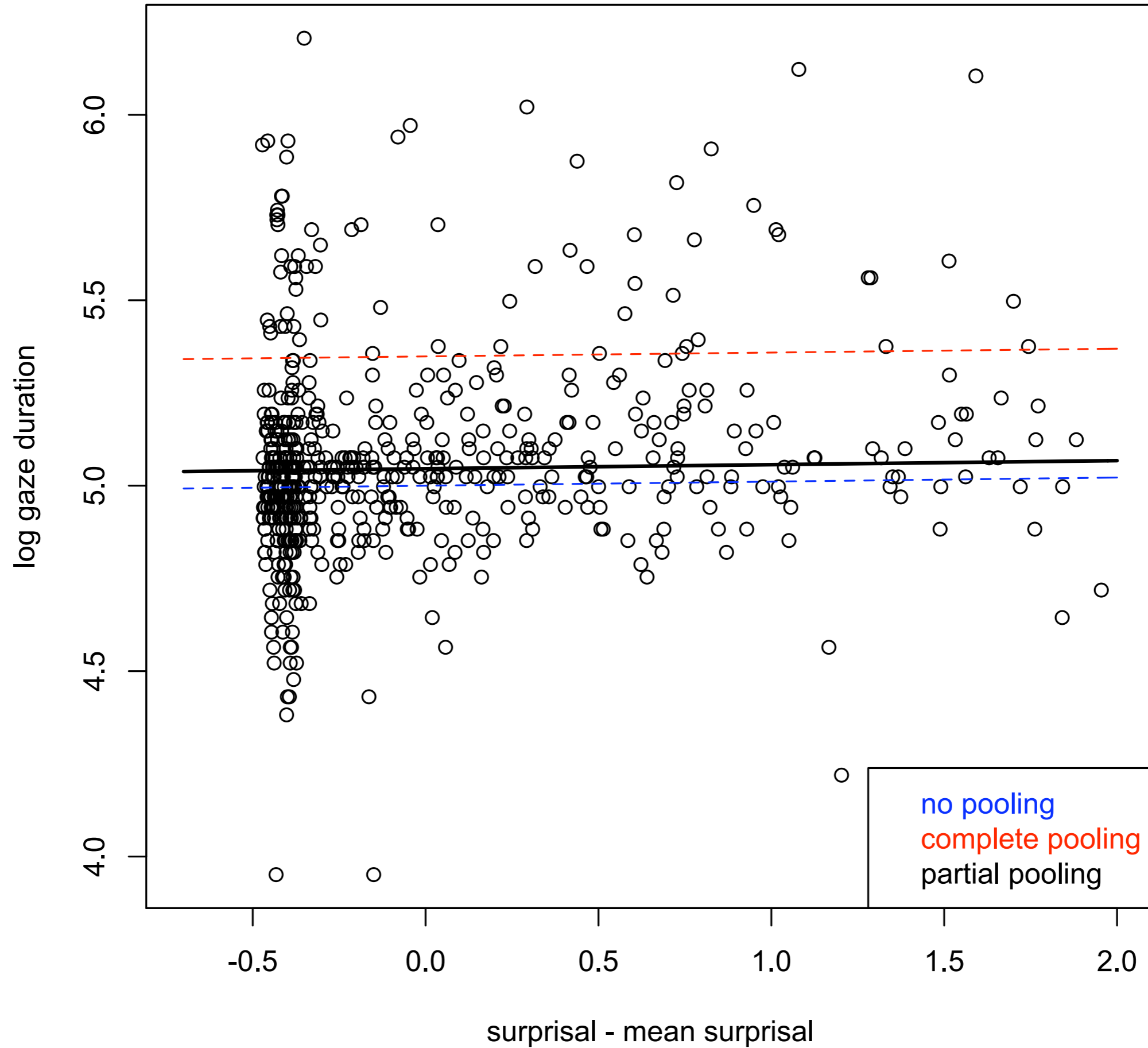
ϵ_{oj}

ϵ_{1j}
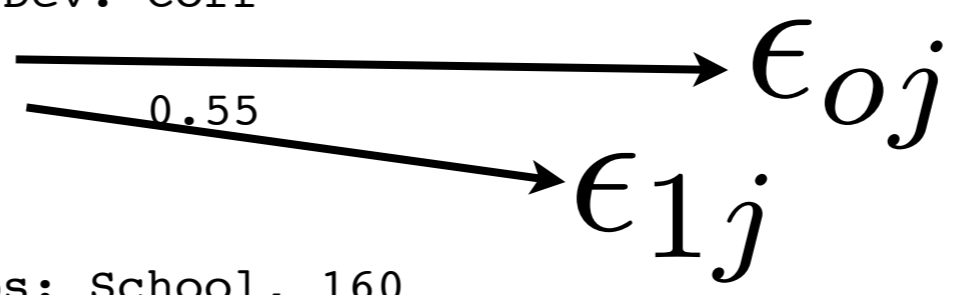
Kim, Kobele, Runner & Hale, to appear sometime in 2012

		Match	Mismatch
Voice	No ellipsis	(1-a) (1-b)	(2-a) (2-b)
	Ellipsis	(1-a) (1-b)	(2-a) (2-b)
Category	No ellipsis	(3)	(4-a) (4-b)
	Ellipsis	(3)	(4-a) (4-b)

(1) Voice match

    a. Active-Active: Jill betrayed Abby, and Matt did ~~betray Abby~~, too.

    b. Passive-Passive: Abby was betrayed by Jill, and Matt was ~~betrayed by Jill~~, too.

(2) Voice mismatch

    a. Active-Passive: Jill betrayed Abby, and Matt was ~~betrayed by Jill~~, too.

    b. Passive-Active: Abby was betrayed by Jill, and Matt did ~~betray Abby~~, too.

(3) Category match

    a. VP-VP: Everyone criticized Roy, but Kate didn't ~~criticize Roy~~

    b. VP-VP: The report criticized Roy, but Kate didn't ~~criticize Roy~~

(4) Category mismatch

    a. Noun-VP: The criticism of Roy was harsh, but Kate didn't ~~criticize Roy~~

    b. Adjective-VP: The report was critical of Roy, but Kate didn't ~~criticize Roy~~

Three random effects

```
NV.glmm <- lmer(LogeNorm ~ Ellipsis + Canonical + NPVP + Ellipsis:NPVP + (1|Subject) +
(1|Item) + (1|Trial), data=NV, family="gaussian")
```

Boston, Hale, Vasishth and Kliegl (still in reviewer-limbo)

Subject 44

log gaze duration

surprisal - mean surprisal

no pooling
complete pooling
partial pooling

Monday, April 19, 2010

CoNLL '05 data

```
vnum,verb,sense,args,A0loc,A1loc,A0size,A1size,A2size,A3size,N0size,N1size
1,take,01,(N0)(A0)(V)(A1)(AM)(N1),-1,1,2,2,0,0,17,25
0,say,01,(A0)(V)(A1),-1,1,4,14,0,0,0,0
1,expect,01,(N0)(A0)(V)(A1),-1,1,1,12,0,0,5,0
0,sell,01,(A0)(AM)(V)(A1)(AM),-2,1,4,2,0,0,0,0
0,say,01,(A0)(V)(A1),-1,1,9,18,0,0,0,0
0,increase,01,(A0)(V)(A1)(A4)(A3),-1,1,3,2,0,5,0,0
1,make,01,(A0)(V)(A1),-1,1,7,8,0,0,0,0
0,make,01,(A0)(V)(A1)(AM),-1,1,1,2,0,0,0,0
3,report,01,(N0)(A0)(V)(A1),-1,1,3,7,0,0,24,0
2,report,01,(N0)(A0)(V)(A1)(AM),-1,1,1,3,0,0,6,0
4,get,03,(N0)(A1)(V)(A2)(AM),0,-1,0,1,3,0,40,0
0,pay,01,(N0)(A0)(AM)(V)(A1)(A2)(N1),-2,1,2,1,8,0,1,30
1,report,01,(N0)(A0)(V)(A1)(AM),-1,1,7,8,0,0,5,0
3,call,11,(N0)(A0)(A0)(AM)(V)(A1)(N1),-2,1,2,2,0,0,15,23
4,say,01,(A1)(V)(A0),1,-1,18,24,0,0,0,0
2,say,01,(A1)(A0)(V),-1,-2,3,30,0,0,0,0
1,say,01,(A1)(A0)(V),-1,-2,2,11,0,0,0,0
3,say,01,(N0)(A0)(V)(A1),-1,1,1,33,0,0,14,0
4,have,03,(N0)(A1)(A0)(AM)(V)(A1),-2,1,4,29,0,0,12,0
1,say,01,(A1)(A0)(V),-1,-2,3,19,0,0,0,0
2,say,01,(A1)(A0)(V),-1,-2,2,15,0,0,0,0
.
.
.
```

amount of stuff *before* the clause of interest

length of the agent

Fit & compare models

```
vbarg <- read.csv("vbarg.txt")
vbarg$a0 <- log(vbarg$A0size+1)
vbarg$n0 <- log(vbarg$N0size+1)


a0n0.lme <- lmer(a0 ~ n0 + (1|verb),data=vbarg)
a0n0.lme2 <- lmer(a0 ~ n0 + (1+n0|verb),data=vbarg)
anova(a0n0.lme,a0n0.lme2)


Data: vbarg
Models:
a0n0.lme: a0 ~ n0 + (1 | verb)
a0n0.lme2: a0 ~ n0 + (1 + n0 | verb)
          Df    AIC    BIC logLik  Chisq Chi Df Pr(>Chisq)
a0n0.lme   4 58398 58431 -29195
a0n0.lme2  6 58083 58133 -29036 318.37      2  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Next class....

Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. Journal of Memory and Language, 59 (4): 434–446.

§7.4