Summary of Lectures 5+6 on Centering in Multi-level Models (MLMs)

$$Y_{ij} = \underbrace{\beta_0 + b_{0i}}_{\beta_{0i}} + \beta_1 X_{ij} + \varepsilon_{ij}$$

$$\varepsilon_{ij} \overset{iid}{\sim} G(0,\sigma^2)$$
$$b_{0i} \overset{iid}{\sim} G(0,\tau^2)$$
$$\varepsilon_{ij} \perp b_{0i}$$

Decompose $\quad X_{ij} = X_{ij} - \bar{X}_{..} + \bar{X}_{..}$

GRAND MEAN

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 \bar{X}_{..} + \beta_1 \underline{(X_{ij} - \bar{X}_{..})}$$

$$= (\beta_0 - \beta_1 \bar{X}_{..}) + b_{0i} + \beta_1 (X_{ij} - \bar{X}_{..})$$

$$= \underline{\beta_0^*} + \underline{b_{0i}} + \underline{\beta_1(X_{ij} - \bar{X}_{..})}$$

Centering with grand mean changes the intercept only; same plot except for center of X-axis

Centering with $\overline{X}_{i\bullet}$

$$X_{ij} = X_{ij} - \overline{X}_{i\bullet} + \overline{X}_{i\bullet}$$

cluster mean

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 (X_{ij} - \overline{X}_{i\bullet} + \overline{X}_{i\bullet}) + \varepsilon_{ij}$$

$$= \beta_0 + b_{0i} + \underbrace{\beta_{1W}}_{\substack{\text{within cluster}\\ \text{effect of}\\ X}} (X_{ij} - \overline{X}_{i\bullet}) + \underbrace{\beta_{1B}}_{\substack{\text{between cluster}\\ \text{effect of}\\ X}} \overline{X}_{i\bullet} + \varepsilon_{ij}$$

Contextual effect : effect of context ($\bar{x}_{i\cdot}$) for individuals with the same value of $\boxed{x_{ij}}$!!

(A) $$Y_{ij} = \beta_0 + b_{0i} + \beta_{1W}(x_{ij} - \bar{x}_{i\cdot}) + \underbrace{\beta_{1B}}_{?}\bar{x}_{i\cdot} + \varepsilon_{ij}$$

same value of $x_{ij} - \bar{x}_{i\cdot}$, <u>not</u> x_{ij}

Instead use:

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 x_{ij} + \beta_2 \bar{x}_{i\cdot} + \varepsilon_{ij}$$

contextual effect

(A): $Y_{ij} = \beta_0 + b_{0i} + \beta_{1W} X_{ij} - \beta_{1W} \overline{X}_{i\cdot} + \beta_{1B} \overline{X}_{i\cdot} + \varepsilon_{ij}$

$$= \beta_0 + b_{0i} + \beta_{1W} X_{ij} + (\beta_{1B} - \beta_{1W}) \overline{X}_{i\cdot} + \varepsilon_{ij}$$

$$\beta_2 = (\beta_{1B} - \beta_{1W}) \overline{X}_{i\cdot}$$

contextual effect

Compositional Effect

"when inter-group (inter-context) differences in the outcome (Y_{ij}) are attributable to differences in group (i) composition (i.e. in the characteristics of the individuals (Z_{ij}) that comprise the group (i)"

"inter-context differences" ⟹ estimate of contextual effect

are attributable to ⟹ contextual effects are reduced when Z_{ij} included in model

Objectives in Homework

II. does the composition of the school (ethnicity, gender, SES) explain the observed variation in outcome across schools?

I. to understand the variation in average math achievement across schools

III. do school-level variables account for variation in $\bar{Y}_j$ across schools?

$$Y_{ij} = \underbrace{\beta_0 + b_{0i}}_{\beta_{0i}} + \varepsilon_{ij} \qquad \begin{array}{l} \varepsilon_{ij} \sim G(0, \sigma^2) \\ b_{0i} \sim G(0, \tau^2) \end{array}$$

I. What measures variation across schools in average Y_{ij}? ______

Estimate $\underline{\tau^2}$ from data to answer I.

II. Is size (degree) of variation across schools explained by ethnicity, gender, SES differences in children across schools?

$$Y_{ij} = \beta_0 + b_{0i} + \beta_E(E_{ij} - \bar{E}_{\cdot\cdot}) + \beta_G(G_{ij} - \bar{G}_{\cdot\cdot}) + \beta_{SES}(SES_{ij} - SES_{\cdot\cdot}) + \epsilon_{ij}$$

Compare $\hat{\tau}^2_{I}$ with $\hat{\tau}^2_{II}$

Center by Grand or School mean?

III. Do school-level variables explain size of τ^2_{II} ?

Model II + $\beta_{PAT}(PAT_i - \overline{PAT}_{\cdot}) +$

$\beta_{\%Min}(\%Min_i - \overline{\%Min}_{\cdot}) +$

$\Rightarrow \hat{r}^2_{\mathrm{III}}$ - compare to $\hat{r}^2_{\mathrm{II}}$ and $\hat{r}^2_{\mathrm{I}}$

- BUT -

"Contextual effect" - vs compositional effects

- brief google scholar view

$$Y_{ij} = \beta_0 + b_{0i} + \beta_{1W}(X_{ij} - \bar{X}_{..}) + \beta_{1C}(\bar{X}_{i.} - \bar{X}_{..}) + \beta_Z(Z_{ij} - \bar{Z}_{..}) + \varepsilon_{ij}$$

$$\boxed{\text{Compare } \hat{\beta}_{1C} \text{ to } \hat{\beta}^*_{1C} \mid \beta_Z = 0}$$

Lecture 7

Linear Mixed Effects Models

-Review or Short overview

- Extension of traditional linear model
- * Some subset of the regression parameters vary randomly across subjects (schools, clusters, "independent units")
- Mean response is modeled as
  - Fixed effects: shared characteristics of the entire population
  - Subject ("Cluster")-specific effects: unique to ~~individuals~~ cluster
- The variation among clusters in regression parameters is how we induce correlation structure into the model
  - All observations from the same ~~subject~~ (cluster) share the same ~~subject~~ cluster-specific regression parameter(s) creating a link/correlation

Linear Mixed Effects Models

- In addition, these randomly varying or cluster subject-specific parameters
  - Allow us to distinguish between different sources of variation in the data
    - Variation in responses at baseline across clusters individuals
    - Variation in rate of change of responses across clusters individuals
    - Random variation in the measurement process within an individual cluster over time

Linear Mixed Effects Models

- Benefits of these models:
  - Partitioning variance into between vs. within cluster subject variation or more generally assigning variation to different levels or combinations of levels
  - Flexible in terms of handling imbalance: in number of observations per person and variation in measurement times
  - Modeling non-constant variance
  - Are valid given common missing data models
  - Prediction:
    - Can predict/describe population mean trajectories
    - Can predict individual trajectories

Linear Mixed Effects Models

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i$$

$$= \begin{bmatrix} 1 & t_{ij} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + 1 \cdot b_{0i} + \varepsilon_{ij}$$

X_{ij}

X_i is the design matrix for the fixed population-level effects

fixed effects

β is the vector of population-level association parameters

cluster

Z_i is the design matrix for the subject-specific or random effects

b_i is the vector of subject-specific parameters

random effects

$b_i \sim N(0, G)$, G is some covariance matrix

$\varepsilon_i \sim N(0, \sigma_\varepsilon^2 R)$, R is some correlation matrix

Assume b_i and ε_i are independent.

Random Intercept Model

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_i + \varepsilon_{ij}$$

$$= \beta_0 + b_{0i} + \beta_1 t_{ij} + \varepsilon_{ij}$$

$$= \beta_{0i} + \beta_1 t_{ij} + \varepsilon_{ij}$$

ε_{ij} iid $G(0, \sigma^2)$

β_{0i} iid $G(\beta_0, \tau^2)$

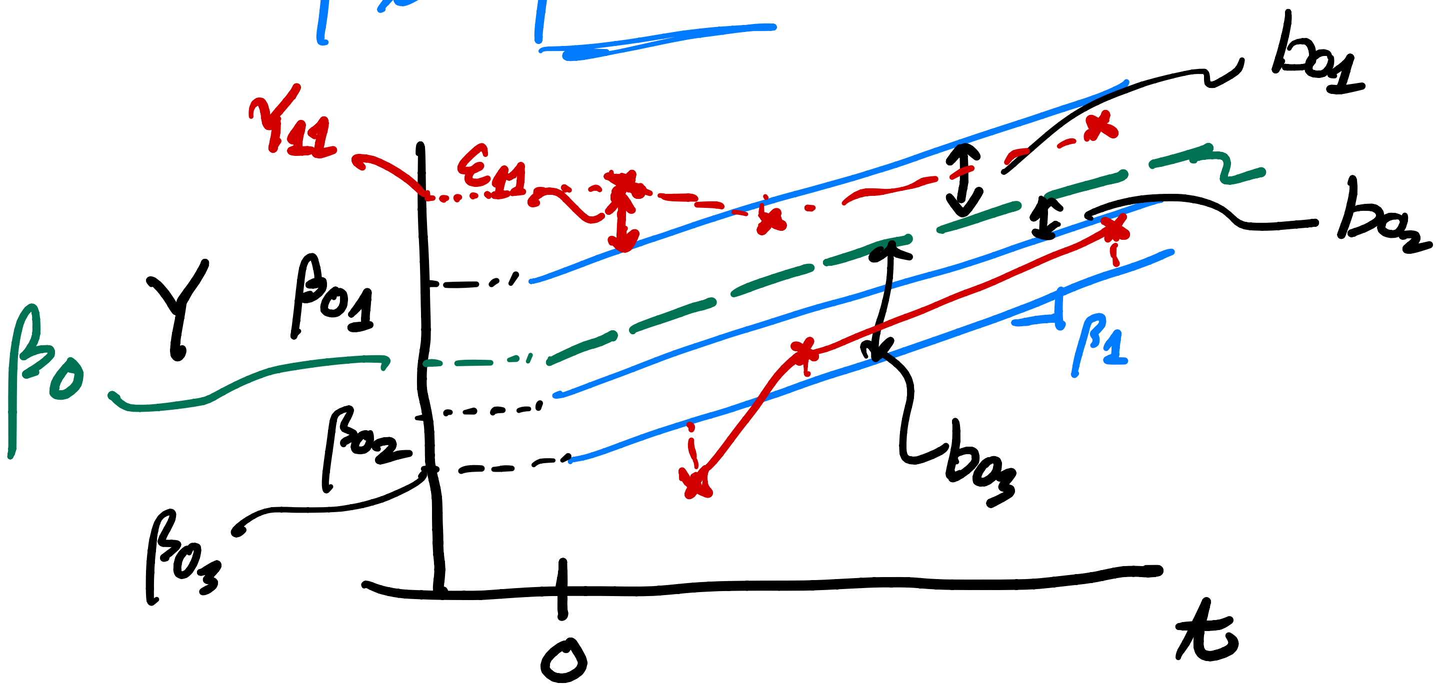
NOTE: reminder b_i are independent of ε_{ij}, simplest case is to assume ε_{ij} indepedent

Random Intercept Model

- Two-stage random effects formulation

$$Y_{ij} = \underline{\beta_{0i}} + \varepsilon_{ij} + \beta_1 t_{ij}$$

$$\beta_{0i} = \underline{\beta_0 + b_{0i}} \quad , \quad b_{0i} \sim G(0, \tau^2)$$



$$\underset{\sim}{\beta_i} = \beta_i + b_i$$

$$Y_{ij} = \underline{\beta_{0i}} + \underline{\beta_{1i}}\, t_{ij} + \varepsilon_{ij}$$

$$\begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix} \sim G$$

$$\underset{2\times 2}{D} = \mathrm{Var}\underset{2\times 1}{\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}}$$

$$G\left(\underset{2\times 1}{\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}}, \underset{2\times 2}{D}\right)$$

$$= \begin{pmatrix} \mathrm{Var}(b_{0i}) & \mathrm{Cov}(b_{0i}, b_{1i}) \\ & \mathrm{Var}(b_{1i}) \end{pmatrix}$$

$$\underset{n_i \times 1}{Y_i} = \underline{X_i}\beta + Z_i b_i + \varepsilon_i \qquad b_i \perp \varepsilon_i$$

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in_i} \end{pmatrix}$$

$$E(Y_i) = X_i \beta_j$$

$$\operatorname{Var} Y_i = \operatorname{Var}(Z_i b_i + \varepsilon_i) = \operatorname{Var}(Z_i b_i) + \operatorname{Var} \varepsilon_i$$

$$= \underline{ZDZ'} + \underline{\sigma^2 I}$$

Mean Responses

Cluster

- Subject-specific mean:

$$E(Y_{ij}|b_i) = \beta_0 + \beta_1 t_{ij} + b_i$$

- Population-level mean:

$$\mathrm{E}(Y_{ij}) = E[E(Y_{ij}|b_i)] = E[\beta_0 + \beta_1 t_{ij} + b_i] = \beta_0 + \beta_1 t_{ij}$$

population average =

average of

expected value for Y_{ij} given cluster intercept $\beta_0 + b_i$

$Eb_i = 0$

Variance Estimates

- Variance

$= E\{(\beta_0+\beta_1 t_{ij}+b_i+\varepsilon_{ij})^2\}$

$$Var(Y_{ij}) = Var(\beta_0 + \beta_1 t_{ij} + b_i + \varepsilon_{ij})$$

$$= Var(b_i + \varepsilon_{ij}) = \sigma^2{}_b + \sigma^2$$

$= E(b_i^2) + E(\varepsilon_{ij}^2) + 2\,cov(b_i, \varepsilon_{ij}) =$

- Covariance

$$Cov(Y_{ij}, Y_{ik}) = Cov(b_i + \varepsilon_{ij}, b_i + \varepsilon_{ik})$$

$= E\{(b_i+\varepsilon_{ij})(b_i+\varepsilon_{ik})\}$

$$= Cov(b_i, b_i) = \sigma^2{}_b$$

$= E(b_i^2) +$
$E(b_i\varepsilon_{ij}) +$
$E(b_i\varepsilon_{ik})$

- Correlation

$$Corr(Y_{ij}, Y_{ik}) = \frac{\sigma^2{}_b}{\sigma^2{}_b + \sigma^2}$$

Variance Estimates

- Linear random intercept model
  - Assumes constant variance, independent within subject residuals
  - Exchangeable correlation structure

$$Corr\left(Y_{ij}, Y_{ik}\right) = \frac{{\sigma^2}_b}{{\sigma^2}_b + \sigma^2}$$

- Within subject correlation can also be interpreted as:
  - The percentage of the total variation in Y that is attributable to differences across subjects relative to natural variation within subject.

Growth-curve modeling

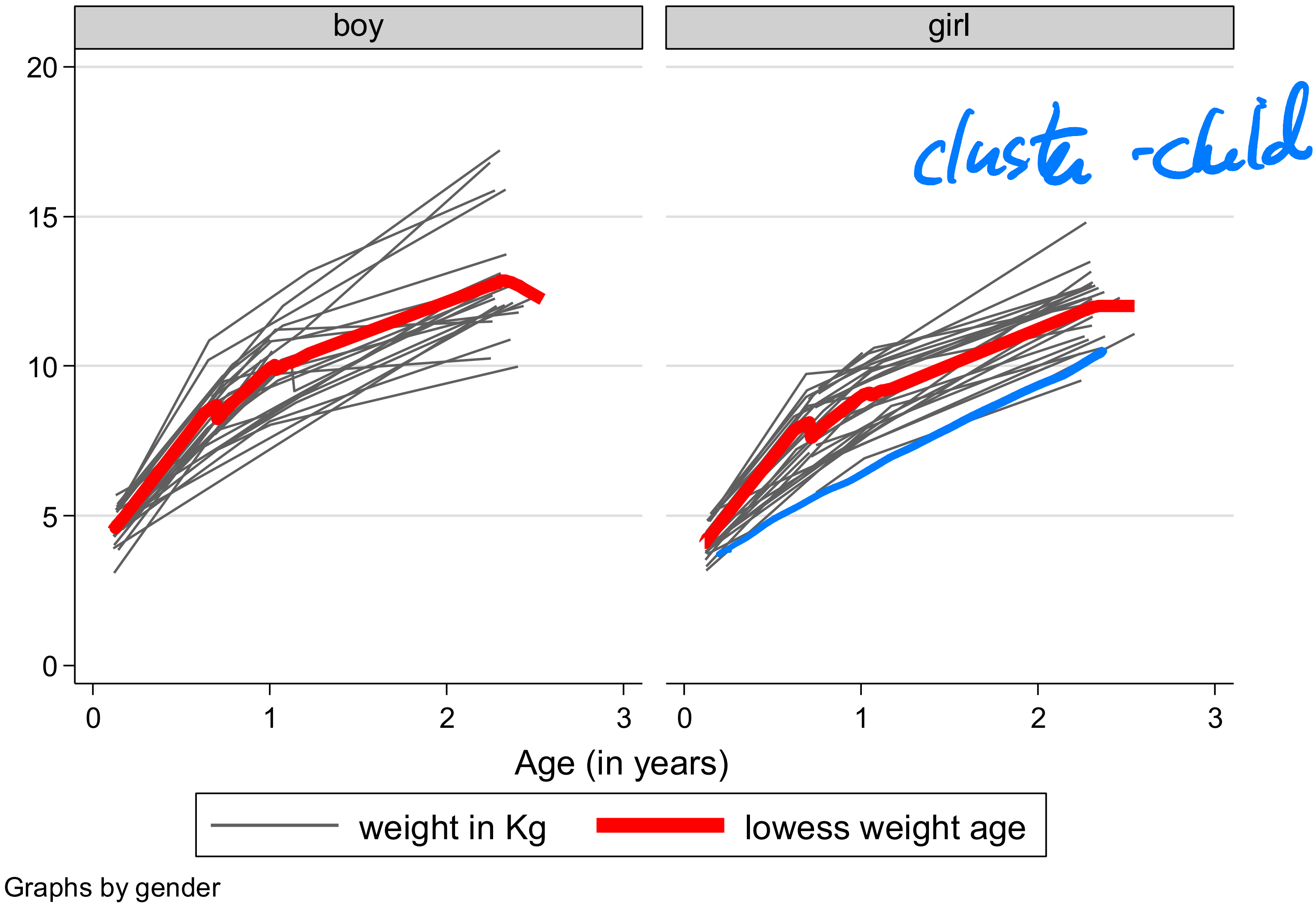
- "asian children weights.dta"

- Measurements of weight were recorded for children
up to 4 occasions at roughly 6 weeks, and then at 8,12, and 27 months

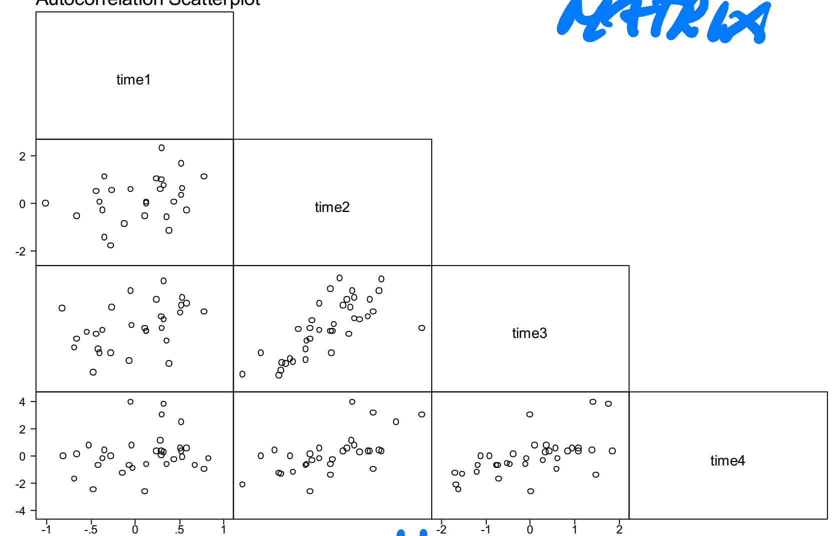
- Goal: We want to investigate the growth trajectories of children's weights as they get older

- Both the shape of the trajectories and the degree of variability are of interest

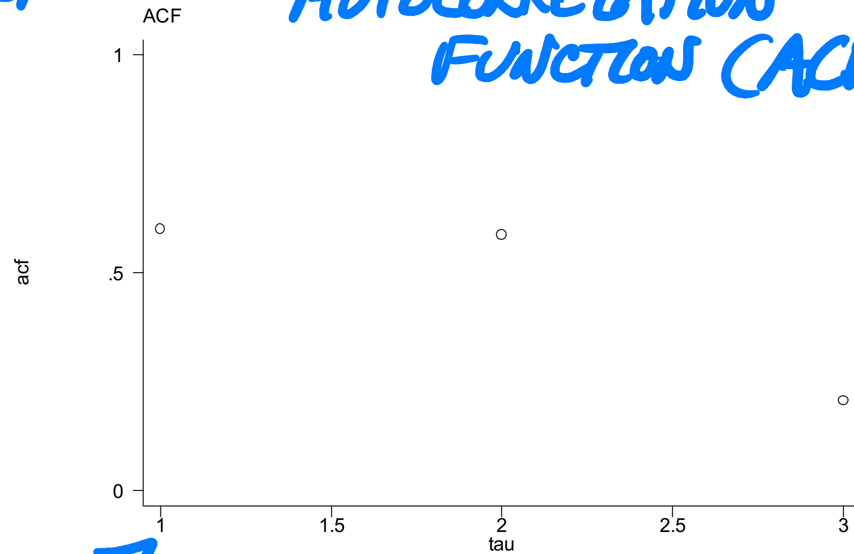
Observed Data



AUTOCORRELATION SCATTERPLOT MATRIX



AUTOCORRELATION FUNCTION (ACF)



AUTOCORRELATION MATRIX

if "STATIONARY"

```
. autocor wtres visit id

             |     time1      time2      time3      time4
-------------+--------------------------------------------
       time1 |    1.0000
       time2 |    0.3344     1.0000
       time3 |    0.4348     0.7388     1.0000
       time4 |    0.2060     0.6277     0.5938     1.0000
```

ACF

```
     +----------+
     |      acf |
     |----------|
  1. |  .600677 |
  2. | .5878199 |
  3. | .2059693 |
     +----------+
```

Random Intercept Model

- A reasonable model would be age, age-squared, gender and the interactions of the age terms and gender.

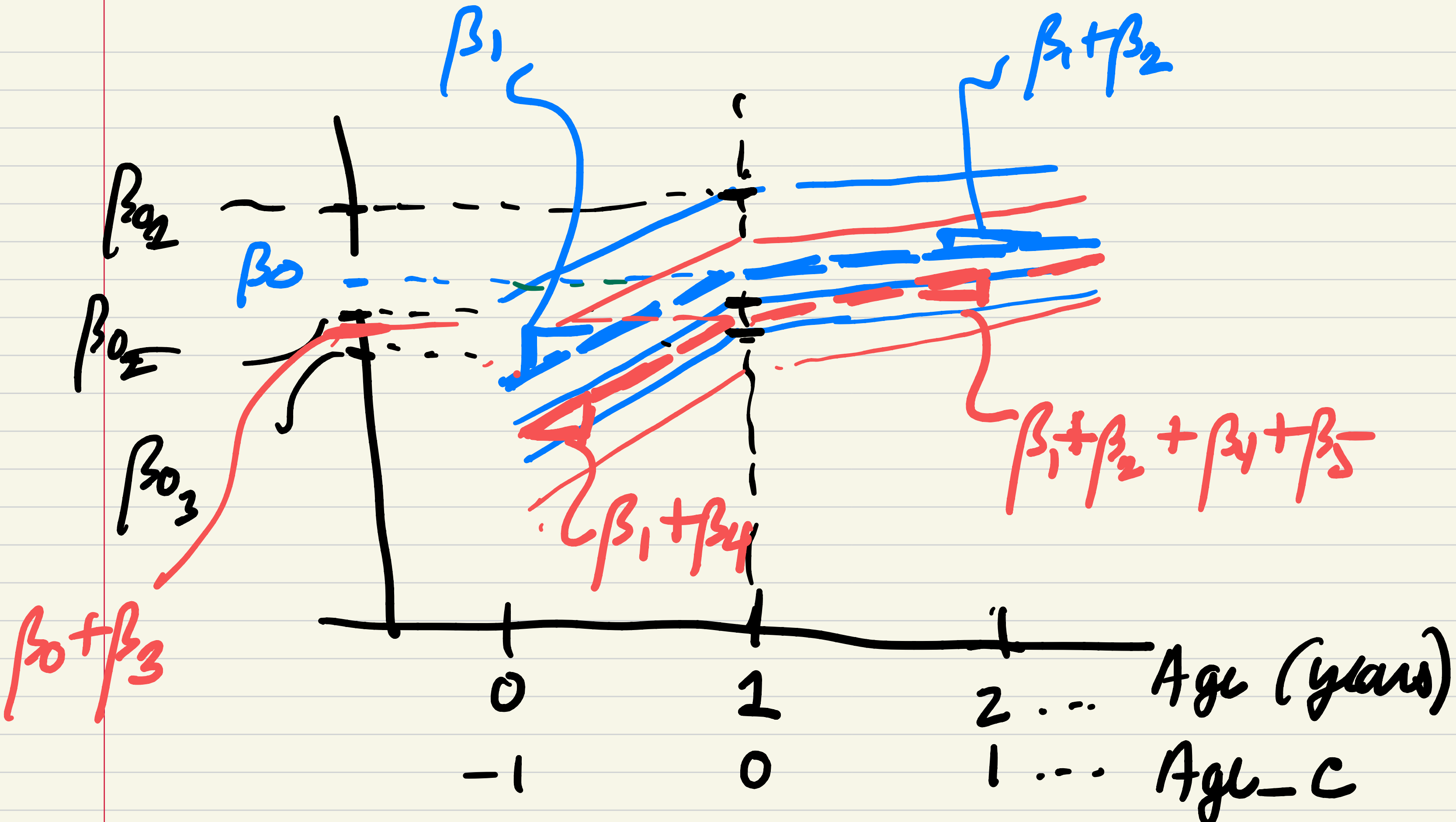
- Just to keep our model a bit more simple to interpret, we will use a linear spline model with knot at 1 year

$$u^+ = \begin{cases} u & \text{if } u \geq 0 \\ 0 & \text{if not} \end{cases}$$

```
gen age_c = age - 1
gen age_sp = (age_c>0)*age_c
gen age_c_girl = age_c*girl
gen age_sp_girl = age_sp*girl
mixed weight age_c age_sp girl age_c_girl age_sp_girl || id: , var
```

$$Wt_{ij} = \overbrace{\beta_{0i}}^{\beta_0 + b_{0i}} + \underline{\beta_1}(Age\_c_{ij}) + \beta_2(Age\_c_{ij})^+ + \beta_3 \text{♀}_i +$$

$$\beta_4\, Age\_c_{ij} * \text{♀}_i + \beta_5 (Age\_c_{ij})^+ * \text{♀}_i + \epsilon_{ij}$$



β_1
$\beta_1+\beta_2$
β_{0_2}
β_0
β_{0_E}
β_{0_3}
$\beta_0+\beta_3$
$\beta_1+\beta_4$
$\beta_1+\beta_2+\beta_4+\beta_5$
0
1
2 ...
Age (years)
-1
0
1 ...
Age_c

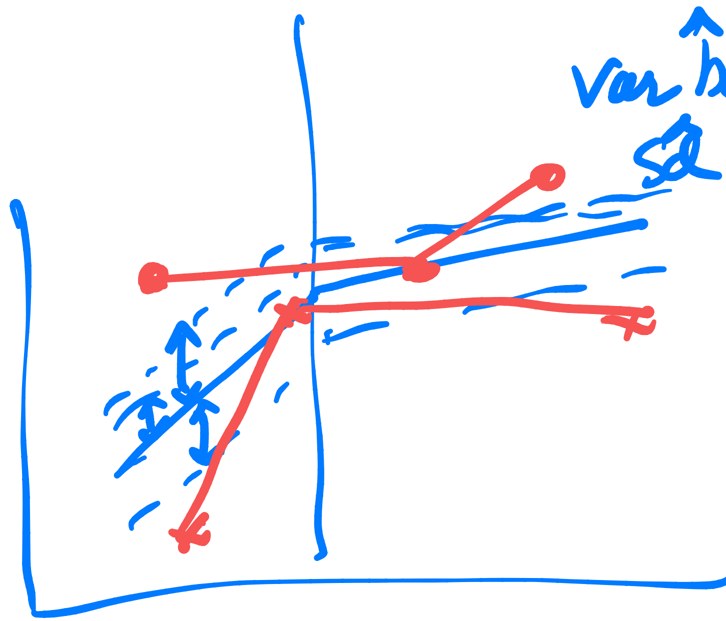
Random Intercept Model

```
Mixed-effects ML regression                     Number of obs      =       189
Group variable: id                              Number of groups   =        68

                                                Obs per group: min =         1
                                                               avg =       2.8
                                                               max =         4


                                                Wald chi2(5)       =   2471.22
Log likelihood = -265.42862                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       age_c |   6.466099   .2687824    24.06   0.000     5.939295    6.992903
      age_sp |    -4.5202   .3805891   -11.88   0.000    -5.266141   -3.774259
        girl |  -.9660133   .2890176    -3.34   0.001    -1.532477   -.3995493
  age_c_girl |  -.7398512   .3889635    -1.90   0.057    -1.502206    .0225032
 age_sp_girl |   .8015376   .5516641     1.45   0.146    -.2797043    1.882779
       _cons |   10.27181   .2032267    50.54   0.000     9.873495    10.67013
------------------------------------------------------------------------------
```

Random Intercept Model

weight	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age_c	6.466099	.2687824	24.06	0.000	5.939295	6.992903
age_sp	-4.5202	.3805891	-11.88	0.000	-5.266141	-3.774259
girl	-.9660133	.2890176	-3.34	0.001	-1.532477	-.3995493
age_c_girl	-.7398512	.3889635	-1.90	0.057	-1.502206	.0225032
age_sp_girl	.8015376	.5516641	1.45	0.146	-.2797043	1.882779
_cons	10.27181	.2032267	50.54	0.000	9.873495	10.67013

Interpretation:



$\text{var } \hat{b}_{0i} = .7$

$\text{sd } b_{0i} = \sqrt{.7} = \underline{.85}$

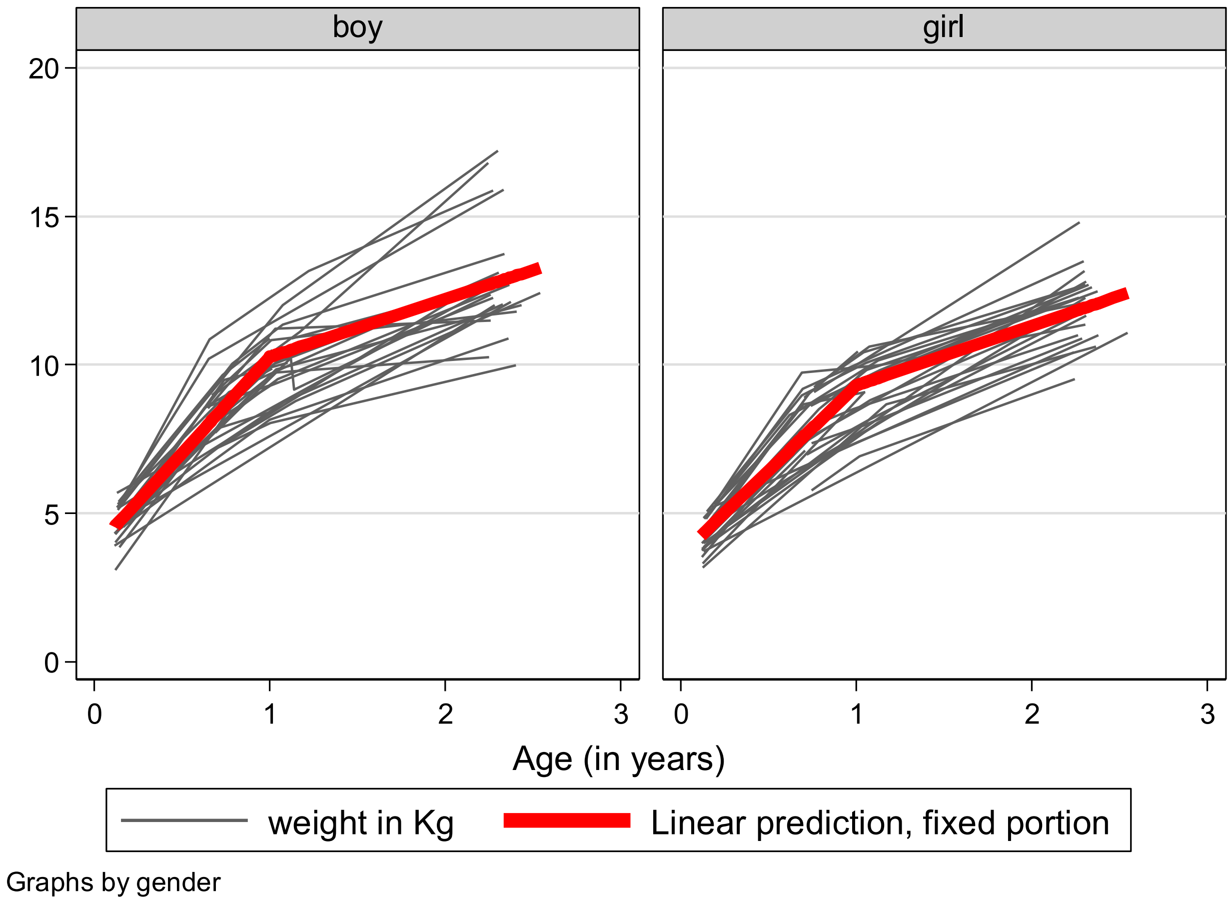
$\pm 2\text{sd} = \underline{1.7 \text{ kg}}$

Random Intercept Model

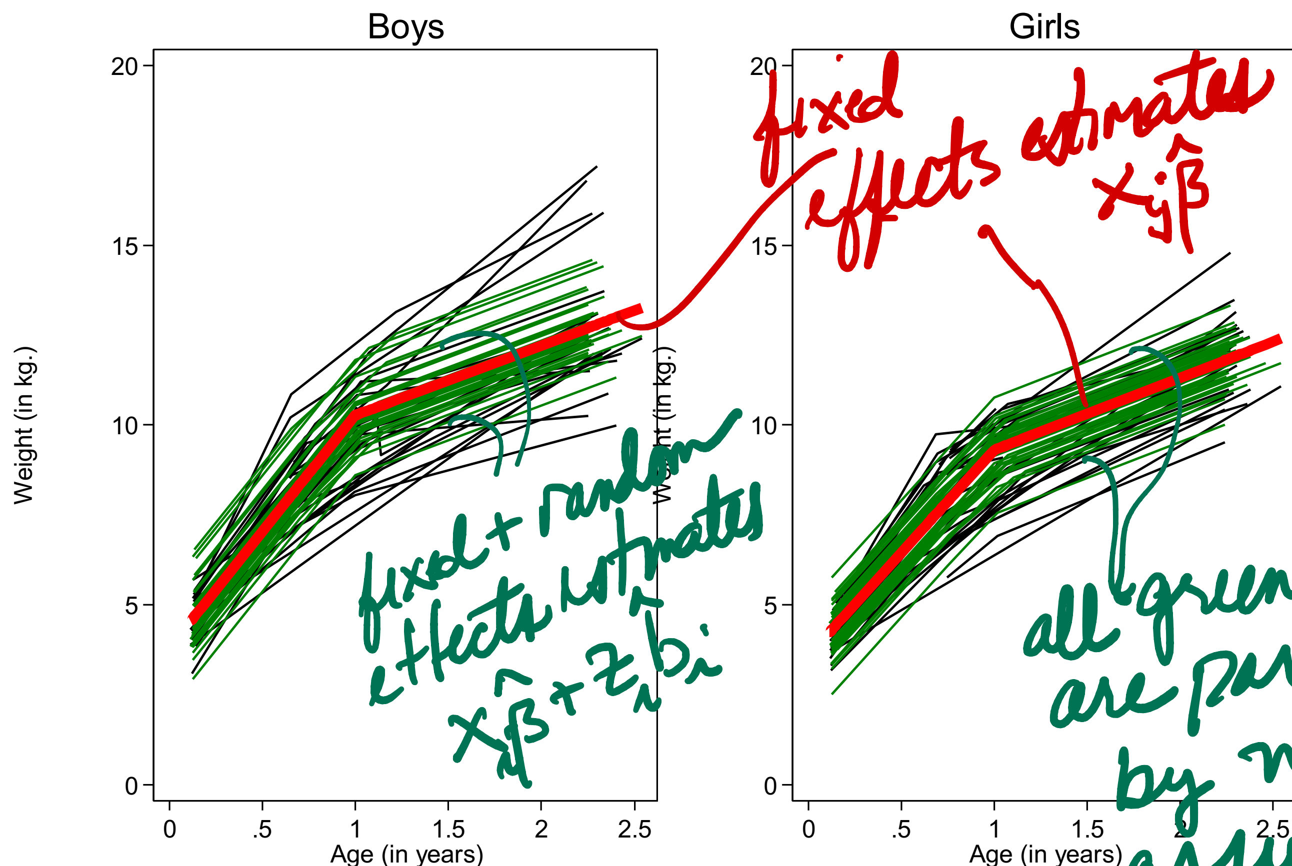
```
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                  var(_cons) |   .7161207   .1608949      .4610435    1.112322
-----------------------------+------------------------------------------------
               var(Residual) |   .5727228   .0730425      .4460526    .7353648
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =    56.33 Prob >= chibar2 = 0.0000
```

$$Corr(Y_{ij}, Y_{ik}) = \frac{0.72}{0.72 + 0.57} = 0.56$$

Predicted Mean Growth



Predicted Mean vs. Individual Growth



fixed effects estimates $X_i\hat{\beta}$

fixed + random effects estimates $X_i\hat{\beta} + Z_i\hat{b}_i$

all green curves are parallel by model assumption

$$\hat{b}_i = E(b_i \mid \underset{\sim}{Y}_i)$$

Summary of Lecture 7

- Considered an example of how to fit a marginal model where measurements are unequally spaced
  - This is a common data feature
- Random Effects models
  - In these models, the within subject correlation is generated by subject specific effects (intercept, slopes)
  - This translates to linear regression models where the intercept and slopes are defined separately for each subject

Summary of Lecture 7, Part I

- Random intercept model
  - Assume a single predictor: time, where time = 0 is baseline
  - The random intercept or subject specific intercept is the expected (mean) value of the response for each subject at baseline

    parallel
  - The model assumes that the mean response will change over time in the "same linear fashion" for each subject
  - If we assume the within subject residuals are independent, then this model is equivalent to a marginal exchangeable correlation model $\mathrm{cov}(y_{ij}, y_{ik}) = \tau^2/(\tau^2+\sigma^2)$
  - Allows us to estimate the proportion of variation in the response that is attributable to differences across subjects relative to variation over time within a subject.

proportion of total variance among clusters

Inner-London School Data

- At age 16, students take Graduate Certificate of Secondary Education (GCSE) exams
  - Scores derived from the GCSE are used for schools comparisons
  - However, schools should be compared based upon their “value added”; the difference in GCSE score between schools after controlling for achievements before entering the school
- One measure of prior achievement is the London Reading Test (LRT)
  - taken by these students at age 11 (G) (L)
- Goal: to investigate the relationship between GCSE and LRT and how this relationship varies across schools. Also identify which schools are most effective, taking into account intake achievement

Inner-London School data:

- Outcome: score exam at age 16 (gcse)
- Covariate: reading test score at age 11 prior to enrollment in the school (lrt)
- Data are clustered within schools:
  - 4059 students clustered within 65 schools
  - Level 1: Student
  - Level 2: School

Inner-London School Data

- Analysis Goals:
  - Estimate the school-specific relationship between the exam score at age 16 and the score at age 11
  - Investigate how this association varies across schools
  - Rank the schools in terms of "performance"
  - Is there evidence that gender modifies the association between score exam at age 16 and score at age 11?
    - Modification at level 1: two level 1 predictors interacting
  - Does the type of school (mixed/boys/girls) explain part of the observed variation across schools?
    - A level 2 predictor interacting with a level 1 predictor

Exploratory Data Analysis

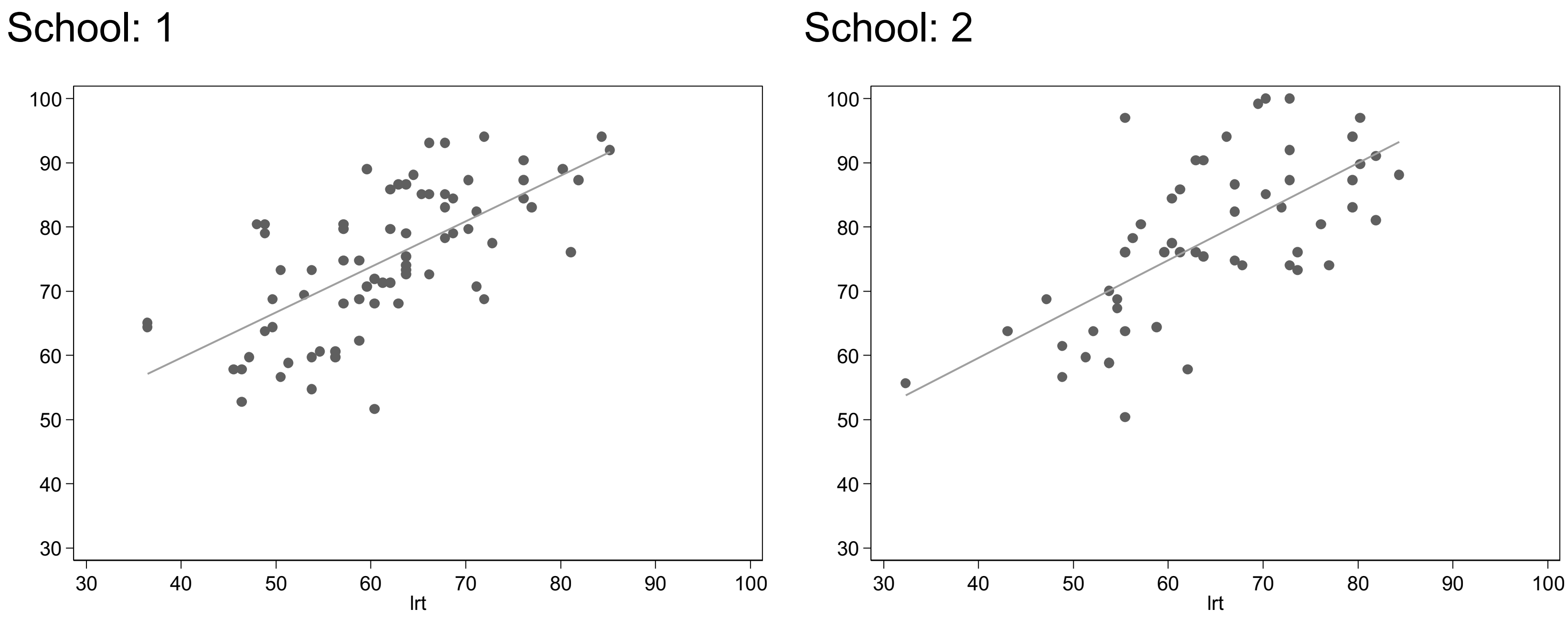
```
* Create some new variables

sort school student

* Generate the number of students within each school
by school: egen totalstudents = count(student)

* Generate a counter for the number of students within each school
by school: gen withinschoolcount = _n


* EDA

* What is the distribution of number of students in each school
summ totalstudents if withinschoolcount==1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
totalstude~s |        65    62.44615    29.74844          2        198
```

65 schools in the dataset:

Number of students ranges from 2 to 198, average 62

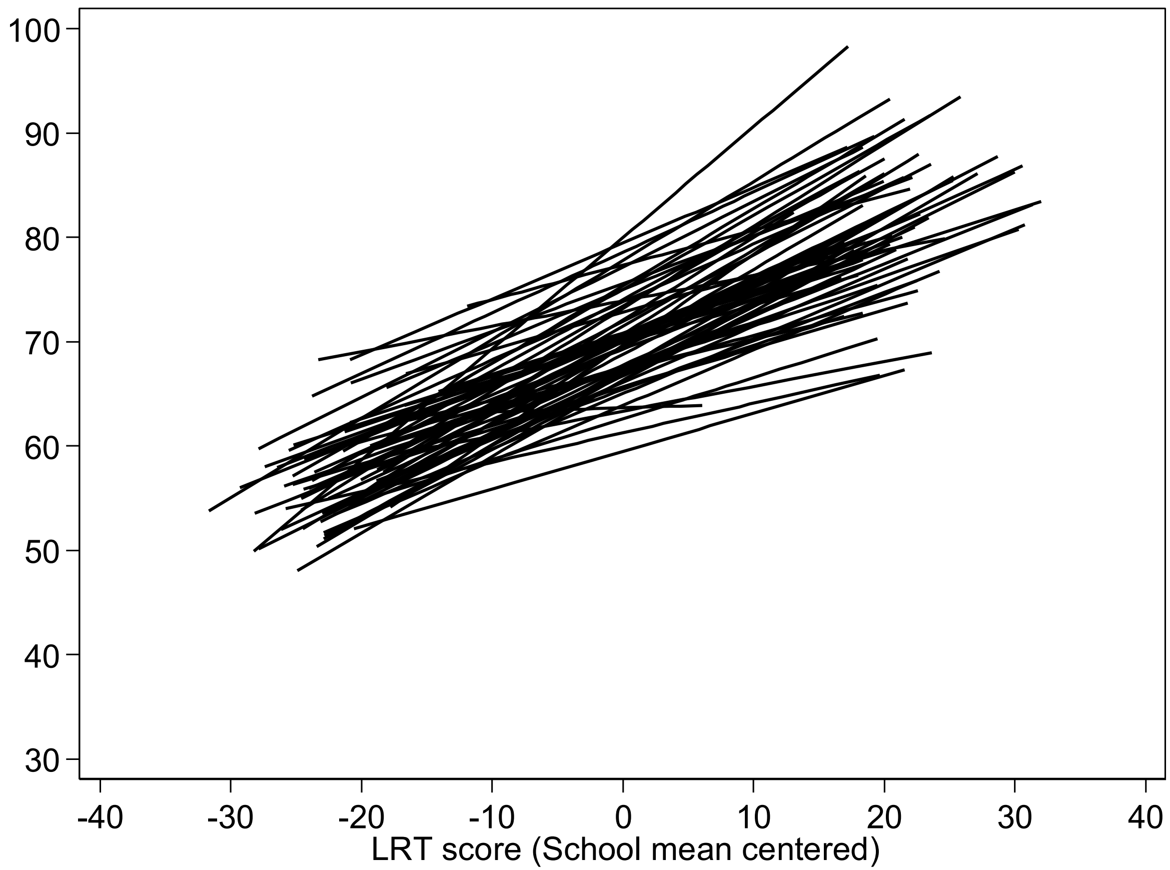
Exploratory Data Analysis

- Relationship between gcse and lrt among two schools (1 and 2)



School-specific relationships among schools with at least 5 students

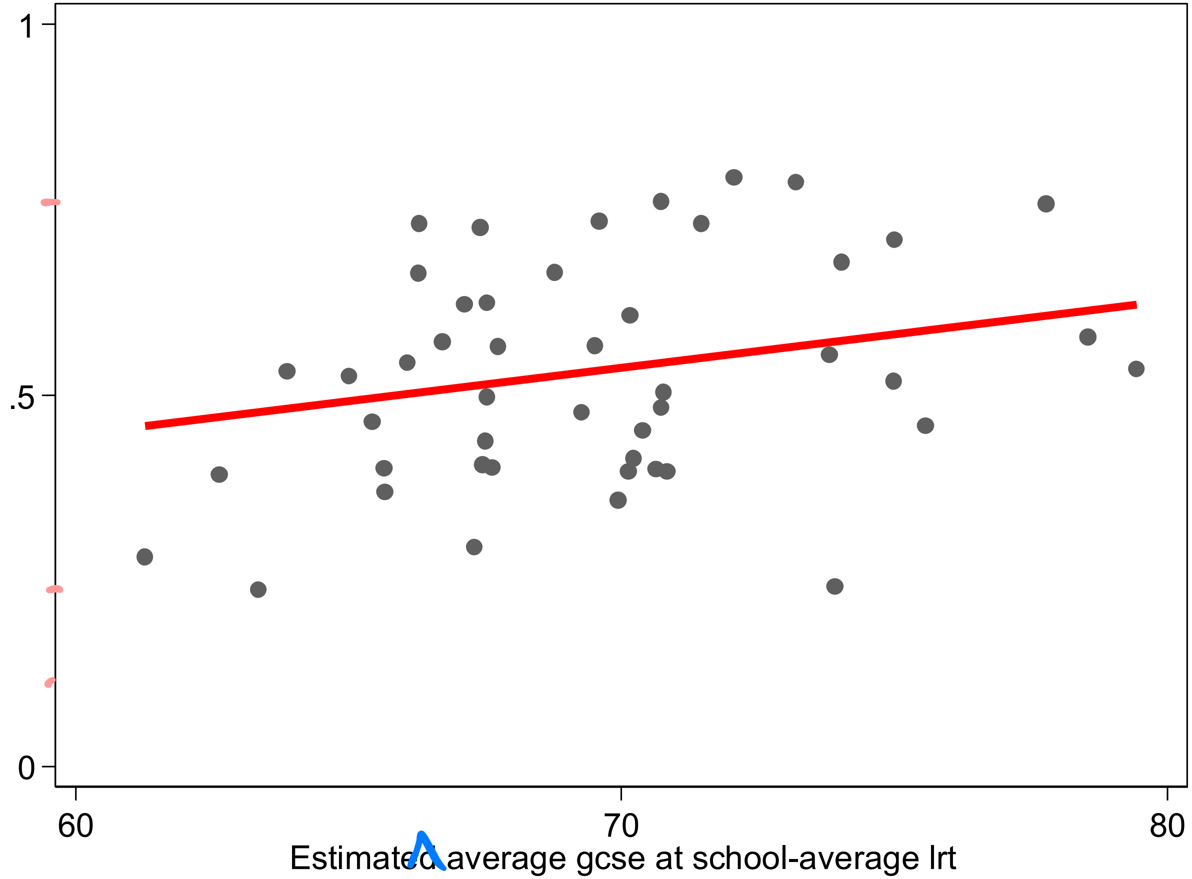
G_{ij}



$(L_{ij} - L_{\cdot j})$

Association between school-specific slope and intercept

$\hat{\beta}_1$



$\hat{\beta}_{0i}$

Linear regression model with random intercept and random slope

(G)

Y_{ij} gcse for student j in school i

(L)

x_{ij} lrt for student j in school i

$\bar{x}_{i.}$ average lrt for school i

$Y_{ij} = \beta_{0i} + \beta_{1i}(x_{ij} - \bar{x}_{i.}) + \varepsilon_{ij}$

$= [1, (x_{ij} - \bar{x}_{i\cdot})] \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + [1, (x_{ij} - \bar{x}_{i\cdot})] \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} + \varepsilon_{ij}$

$= \underset{1\times 2}{\underset{\sim}{X_{ij}}} \underset{2\times 1}{\underset{\sim}{\beta}} + \underset{1\times 2}{Z_{ij}} \underset{2\times 1}{b_i} + \varepsilon_{ij}$

$\beta_{0i} \sim N(\beta_0, \tau_1^2)$

$\beta_{1i} \sim N(\beta_1, \tau_2^2)$

$\text{cov}(\beta_{0i}, \beta_{1i}) = \tau_{12}$

$\varepsilon_{ij} \sim N(0, \sigma^2)$

here $Z_{ij} := X_{ij}$

usually $Z_{ij} \subseteq X_{ij}$

"is contained in"

Alternative representations of the same model

$$Y_{ij} = b_{0i} + \beta_0 + (b_{1i} + \beta_1)(x_{ij} - \bar{x}_{i.}) + \varepsilon_{ij}$$

$$b_{0i} \sim N(0, \tau_1^2)$$

$$b_{1i} \sim N(0, \tau_2^2)$$

$$\text{cov}(b_{0i}, b_{1i}) = \tau_{12}, \varepsilon_{ij} \sim N(0, \sigma^2)$$

$$Y_{ij} = \beta_{0i} + \beta_{1i}(x_{ij} - \bar{x}_{i.}) + \varepsilon_{ij}$$

$$\beta_{0i} = \beta_0 + b_{0i}, b_{0i} \sim N(0, \tau_1^2)$$

$$\beta_{1i} = \beta_1 + b_{1i}, b_{1i} \sim N(0, \tau_2^2)$$

$$\text{cov}(\beta_{0i}, \beta_{1i}) = \tau_{12}, \varepsilon_{ij} \sim N(0, \sigma^2)$$

Recall:

$E(Y_{ij})$ means the population average Y at a particular X_{ij}

or

$E(Y_{ij} | X_{ij})$

$E(Y_{ij} | b_i)$ means the expected value or "average" Y_{ij} at $X = X_{ij}$ given we are in cluster i whose intercept is $\beta_0 + b_i = \beta_{0i}$

If $Y_{ij} = X_{ij}\beta + Z_{ij}b_i + \varepsilon_{ij}$

$b_i \sim G(0, D)$

$\varepsilon_{ij} \sim G(0, \sigma^2)$

$\underline{b}_i \perp \underline{\varepsilon}_i$

then

$$E(Y_{ij}) = E(X_{ij}\beta + Z_{ij}b_i + \varepsilon_{ij})$$
$$= x_{ij} + Z_{ij}Eb_i + E\varepsilon_{ij} = x_{ij}\beta$$

$$E(Y_{ij} \mid b_i) = E(X_{ij}\beta + Z_{ij}b_i + \varepsilon_{ij}) = X_{ij}\beta + Z_{ij}b_i$$

$$= \underset{2\times 2}{Var}(b_i) = Var\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} = \begin{pmatrix} Var(b_{0i}) & Cov(b_{0i}, b_{1i}) \\ Cov(b_{0i}, b_{1i}) & Var(b_{1i}) \end{pmatrix}$$

$$Var(b_{0i}) \equiv \underbrace{E}_{\text{pop. average}}\left\{\left(b_{0i} - \underbrace{E(b_{0i})}_{0}\right)^2\right\} = E\{b_{0i}^2\}$$

$$= \tau_1^2$$

$$Cov(b_{0i}, b_{1i}) = \underbrace{E}_{\text{pop. average}}\left\{\left(b_{0i} - \underbrace{E(b_{0i})}_{0}\right)\left(b_{1i} - \underbrace{E(b_{1i})}_{0}\right)\right\}$$

$$= E(b_{0i} b_{1i}) = \tau_{12} \equiv \tau_{21}$$

$$\mathrm{Var}\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} = \begin{pmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{pmatrix}$$

"MIXED"

Random coefficient models induce heteroskedastic error structure!

○ - random

□ - fixed

$$Y_{ij} = (b_{0i} + \boxed{\beta_0}) + (b_{1i} + \boxed{\beta_1})(x_{ij} - \bar{x}_{i.}) + \varepsilon_{ij}$$

$$Y_{ij} = \boxed{(\beta_0 + \beta_1 x_{ij})} + (b_{0i} + b_{1i}(x_{ij} - \bar{x}_{i.})) + \varepsilon_{ij}$$

ξ_{ij}

$$\xi_{ij} = \underbrace{(b_{0i} + b_{1i}(x_{ij} - \bar{x}_{i.}))}_{} + \underbrace{\varepsilon_{ij}}_{}$$

$$\mathrm{var}(\xi_{ij}) = \tau_1^2 + 2\tau_{12}(x_{ij} - \bar{x}_{i.}) + \tau_2^2 (x_{ij} - \bar{x}_{i.})^2 + \sigma^2$$

$$\underbrace{\mathrm{Var}\, b_{0i} + 2 c_2 \mathrm{Cov}(b_{0i}, b_{1i}) + c_2^2 \mathrm{Var}\, b_{1i}}_{} + \underbrace{\mathrm{Var}\, \varepsilon_{ij}}_{}$$

The total residual variance is said to be heteroskedastic because it depends on x

$$\tau_2^2 = \tau_{12} = 0$$ Model with random intercept only

$$\mathrm{var}(\xi_{ij}) = \tau_1^2 + \sigma^2$$

$Var(c_1 Y_1 + c_2 Y_2)$ when $EY_1 = EY_2 = 0$, c_1, c_2 constants

$$= E\{[c_1 Y_1 + c_2 Y_2 - \overbrace{E(c_1 Y_1 + c_2 Y_2)}^{0}]^2\}$$

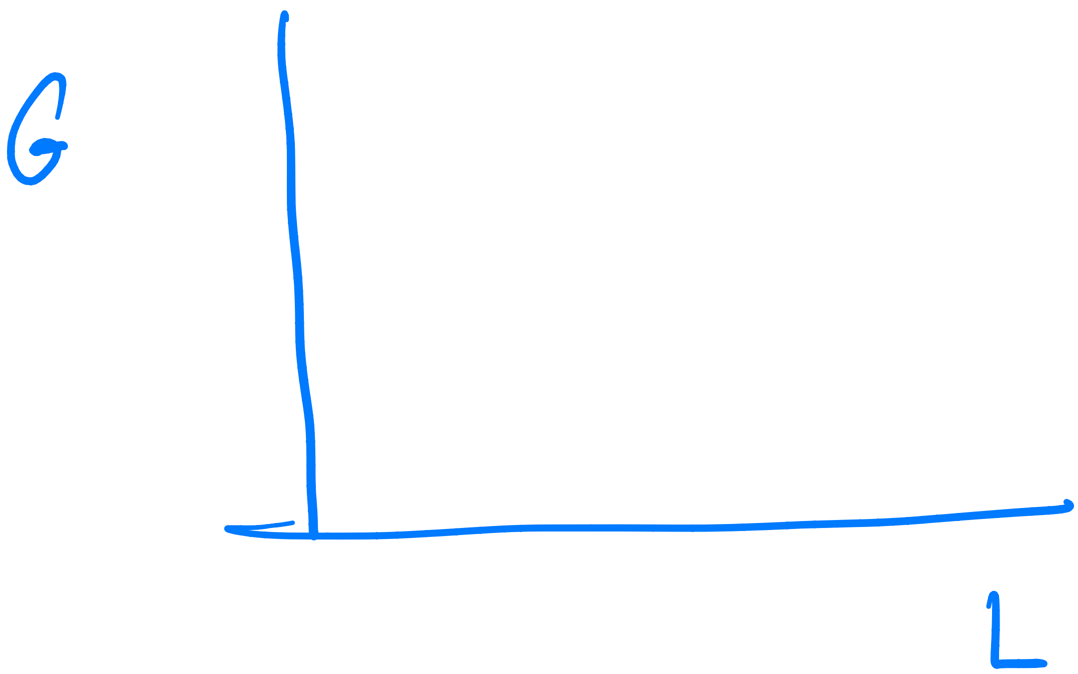
$$= E\{c_1^2 Y_1^2 + c_2^2 Y_2^2 + 2c_1 c_2 Y_1 Y_2\}$$

$$= c_1^2 Var\, Y_1 + c_2^2 Var\, Y_2 + 2c_1 c_2 Cov(Y_1, Y_2)$$

```
mixed gcse lrt_groupc || school: lrt_groupc, cov(uns) mle stddev


Mixed-effects REML regression                   Number of obs      =      4059
Group variable: school                          Number of groups   =        65

                                                Obs per group: min =         2
                                                               avg =      62.4
                                                               max =       198


                                                Wald chi2(1)       =    768.50
Log restricted-likelihood = -14020.718          Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        gcse |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  lrt_groupc |   .5529766   .0199474    27.72   0.000     .5138805    .5920727
       _cons |   69.83566   .5415451   128.96   0.000     68.77426    70.89707
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
               sd(lrt_gr~c)  |   .1202955   .0190615      .0881804    .1641069
                  sd(_cons)  |   4.220941   .3997349      3.505889    5.081833
        corr(lrt_gr~c,_cons) |   .5481782     .13837      .2241939    .7630512
-----------------------------+------------------------------------------------
              sd(Residual)   |   7.432694   .0838333      7.270187    7.598834
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =   835.32   Prob > chi2 = 0.0000
```

Interpretation of fixed effects

gcse	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
lrt_groupc	.5529766	.0199474	27.72	0.000	.5138805	.5920727
_cons	69.83566	.5415451	128.96	0.000	68.77426	70.89707



Interpretation of random effects

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
school: Unstructured				
sd(lrt_gr~c)	.1202955	.0190615	.0881804	.1641069
sd(_cons)	4.220941	.3997349	3.505889	5.081833
corr(lrt_gr~c,_cons)	.5481782	.13837	.2241939	.7630512
sd(Residual)	7.432694	.0838333	7.270187	7.598834

sd(lrt_gr~c) :

(a) $\approx \sum_{i=1}^{60} \hat{b}_{0i}^{2} / 60$

(b) $\approx \sum_{i=1}^{60} b_{0i}^{2} / 60$

Interpretation of random effects

```
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
school: Unstructured         |
               sd(lrt_gr~c) |   .1202955   .0190615      .0881804    .1641069
                  sd(_cons) |   4.220941   .3997349      3.505889    5.081833
        corr(lrt_gr~c,_cons) |   .5481782     .13837      .2241939    .7630512
-----------------------------+------------------------------------------------
               sd(Residual) |   7.432694   .0838333      7.270187    7.598834
------------------------------------------------------------------------------
```

$$Y_{ij} = X_{ij}\beta + Z_{ij} b_i + \varepsilon_{ij}$$

$$\underset{n_i\times 1}{\underset{\sim}{Y_i}} = \underset{n_i\times p}{X_i}\,\underset{p\times 1}{\beta} + \underset{n_i\times q}{Z_i}\,\underset{q\times 1}{\underset{\sim}{b_i}} + \underset{n_i\times 1}{\underset{\sim}{\varepsilon_i}}$$

$$\underset{\sim}{Y_i} \sim G(X_i\beta,\ Z_i D Z_i' + \sigma^2 I_{n_i})$$

$$E(\underset{2\times 1}{\underset{\sim}{b_i}} \mid \underset{n_i\times 1}{\underset{\sim}{Y_i}}) = E(\underset{\sim}{b_i}) + Cov(\underset{\sim}{b_i}, \underset{\sim}{Y_i})\, Var(\underset{\sim}{Y_i})^{-1}(Y_i - E\underset{\sim}{Y_i})$$

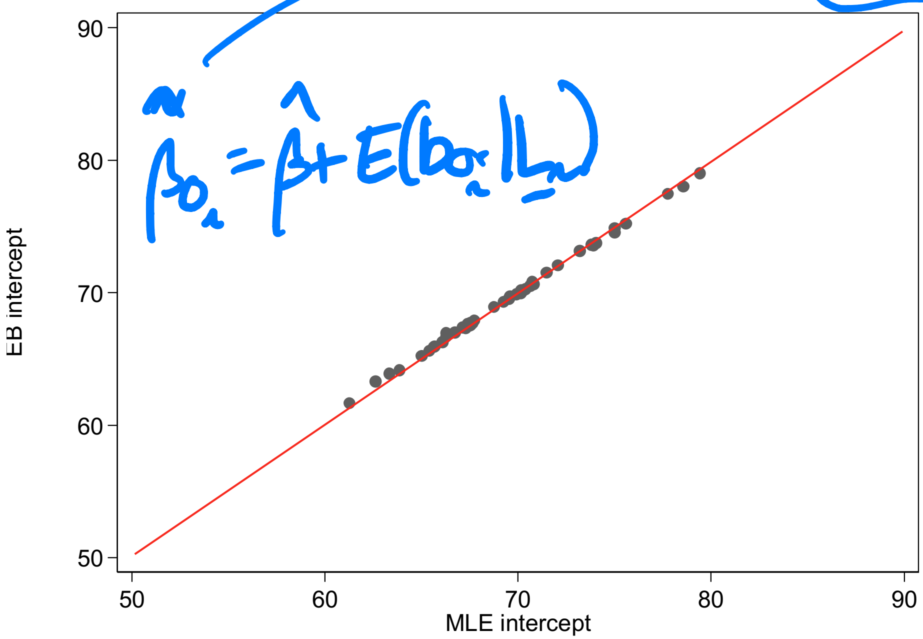
$$= 0 + DZ_i'(Z_i D Z_i' + \sigma^2 I)^{-1}(Y - X_i\beta)$$

EB – replace D, σ^2, β by their estimates

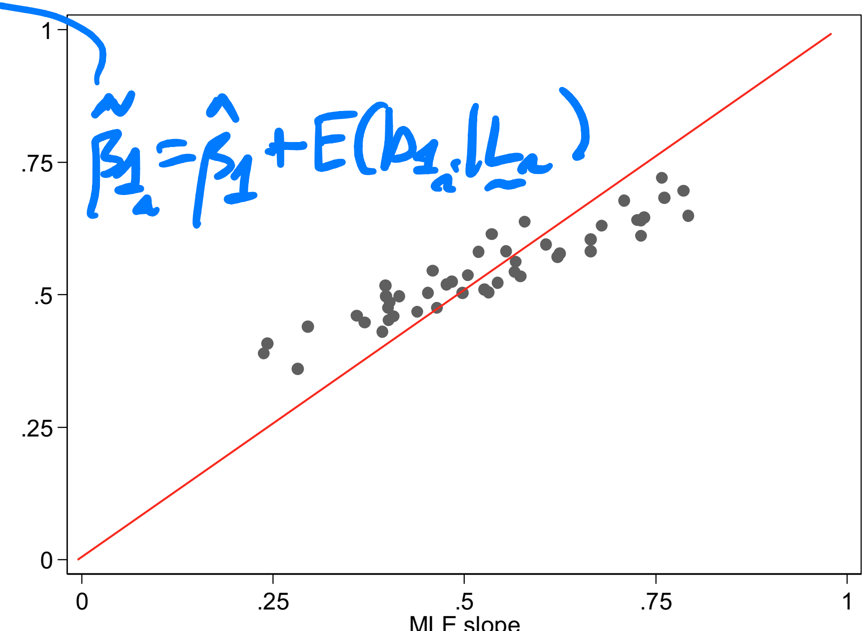
Impact of shrinkage on intercepts and slopes

Empirical Bayes

$\tilde{\beta}_{0_i} = \hat{\beta}_0 + E(b_{0i} \| L_i)$



$\tilde{\beta}_{1_i} = \hat{\beta}_1 + E(b_{1i} \| L_i)$

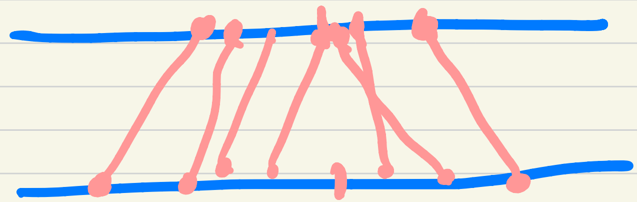


$$\hat{\beta}_{0_i} = \bar{G}_{i\cdot} - \hat{\beta}_{1_i} \bar{L}_{i\cdot}$$

$$\hat{\beta}_{1_i} = \frac{\sum_{j=1}^{n_i} (G_{ij} - \bar{G}_{i\cdot})(L_{ij} - \bar{L}_{i\cdot})}{\sum_j (L_{ij} - \bar{L}_{i\cdot})^2}$$

Which is corrects ?

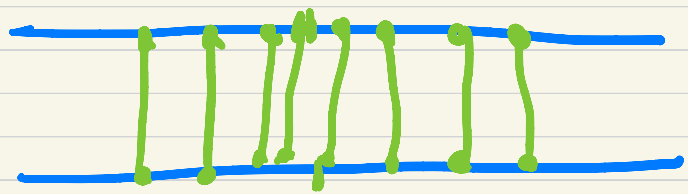
(a)



$\tilde{\beta}_{1i}$ (EB)

$\hat{\beta}_{1i}$ (ML)

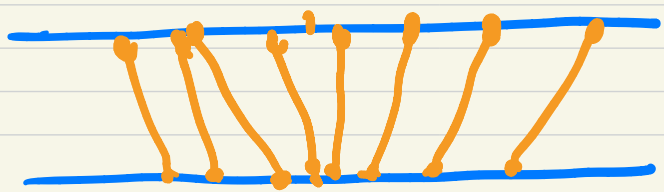
(b)



$\tilde{\beta}_{1i}$

$\hat{\beta}_{1i}$

(c)

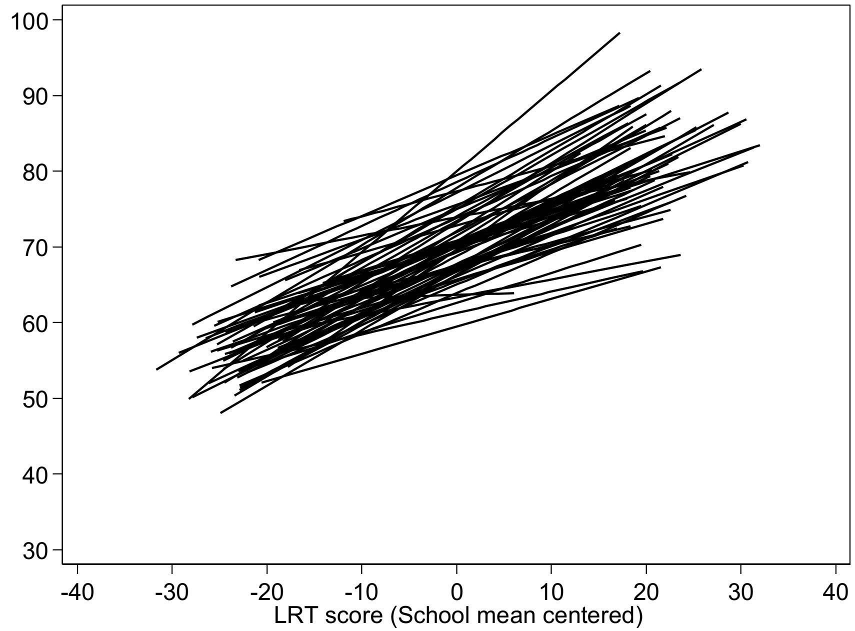


$\hat{\beta}_{1i}$

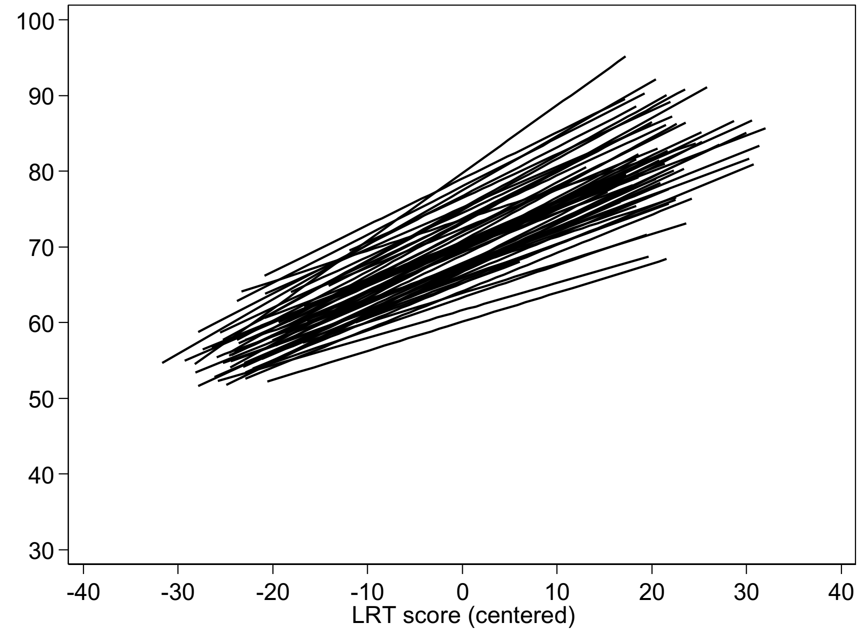
$\hat{\beta}_{0i}$

Impact of shrinkage on predicted association between gcse and lrt

individual curves



EB curves



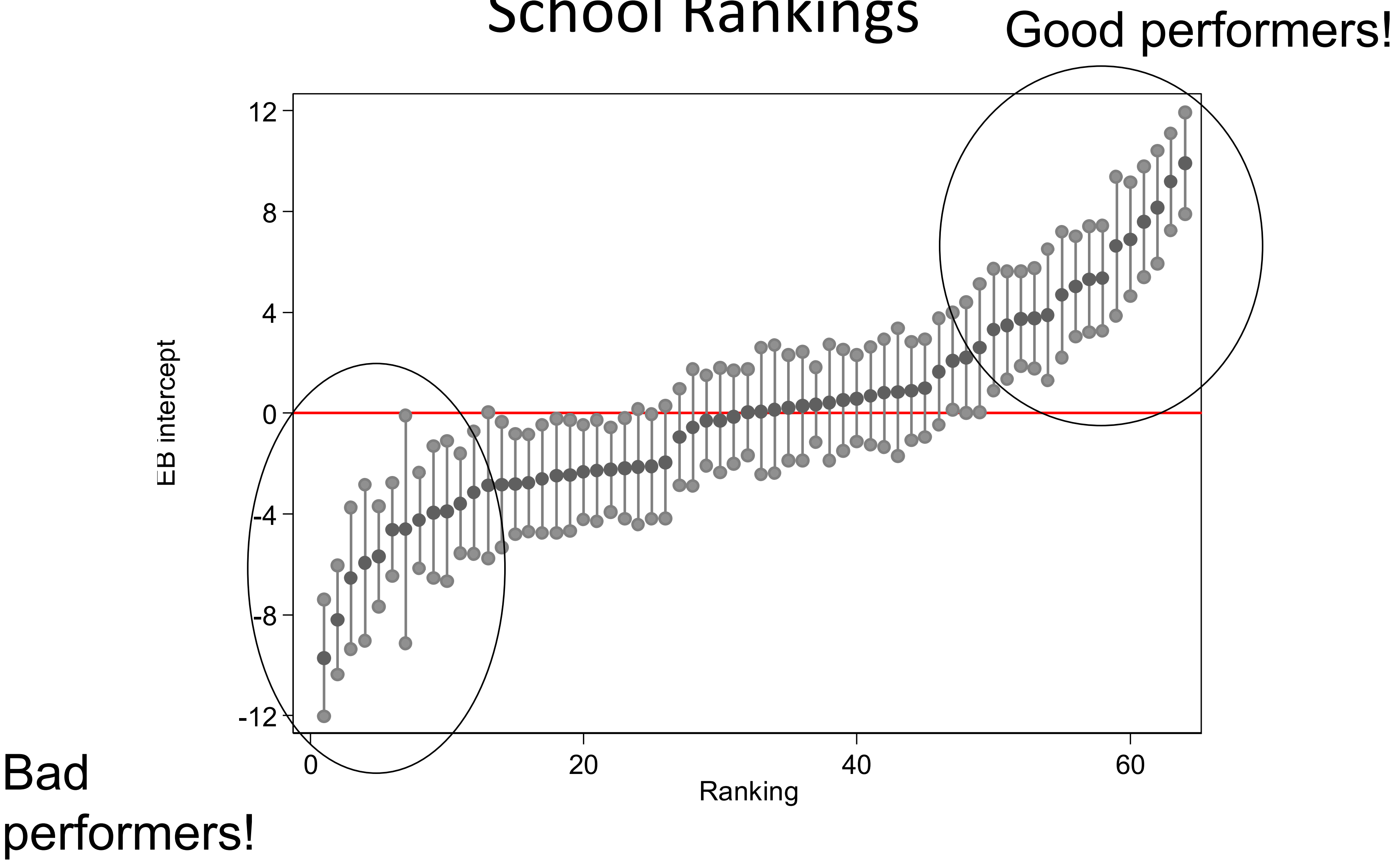
Goal: Rank the schools in terms of performance

What is the appropriate measure to use from the analysis?

- random intercept?
- random slope?

What is the interpretation of the EB estimates of the intercepts?

School Rankings



Lecture 7 Summary

- Reviewed linear random coefficient models for multilevel data
  - within cluster factor not time
- Since we are working in linear case, coefficients have both a marginal and conditional interpretation
- Random slope models allow for us to understand heterogeneity across clusters in within cluster associations
- Cluster-level summaries allow for ranking of clusters