Lecture 7

Linear Mixed Effects Models

- 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
- The variation among clusters in regression parameters is how we induce correlation structure into the model
  - All observations from the same subject share the same subject-specific regression parameter(s) creating a link/correlation

Linear Mixed Effects Models

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

Linear Mixed Effects Models

- Benefits of these models:
  - Partitioning variance into between vs. within 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

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

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

β is the vector of population - level association parameters

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

b_i is the vector of subject - specific parameters

$b_i \sim N(0, G), G$ 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

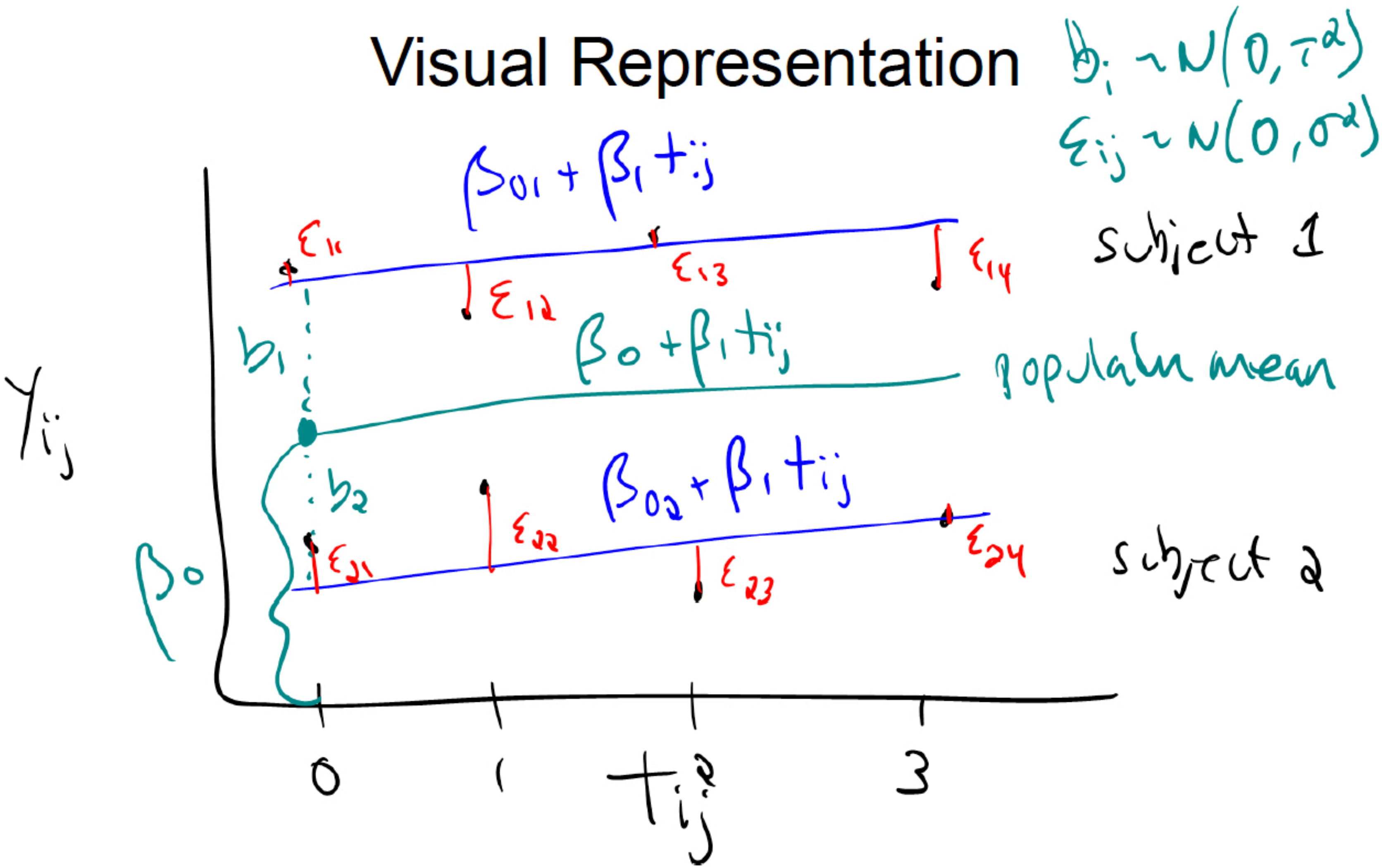
$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_i + \varepsilon_{ij}$$

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

Visual Representation



Mean Responses

- Subject-specific mean:
  $E(Y_{ij}|b_i) = \beta_0 + \beta_1 t_{ij} + b_i$

- Population-level mean:
  $\mathrm{E}(Y_{ij}) = E[E(Y_{ij}|b_i)] = E[\beta_0 + \beta_1 t_{ij} + b_i] = \beta_0 + \beta_1 t_{ij}$

Variance Estimates

- Variance

$$Var(Y_{ij}) = Var(\beta_0 + \beta_1 t_{ij} + b_i + \varepsilon_{ij})$$
$$= Var(b_i + \varepsilon_{ij}) = \sigma^2{}_b + \sigma^2$$

- Covariance

$$\mathrm{Cov}(Y_{ij}, Y_{ik}) = Cov(b_i + \varepsilon_{ij}, b_i + \varepsilon_{ik})$$
$$= Cov(b_i, b_i) = \sigma^2{}_b$$

- 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

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

- 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

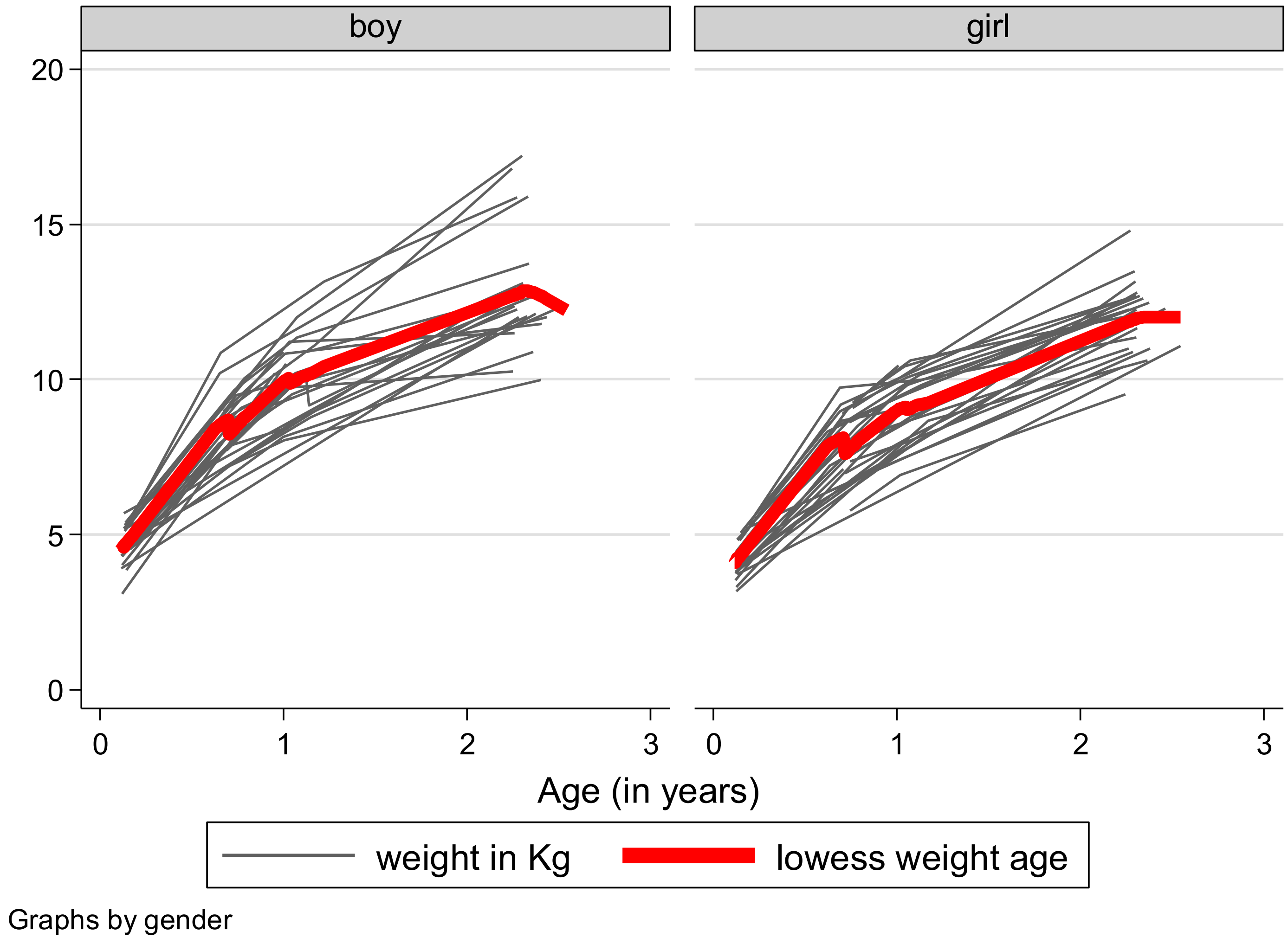
- “asian children weights.dta”

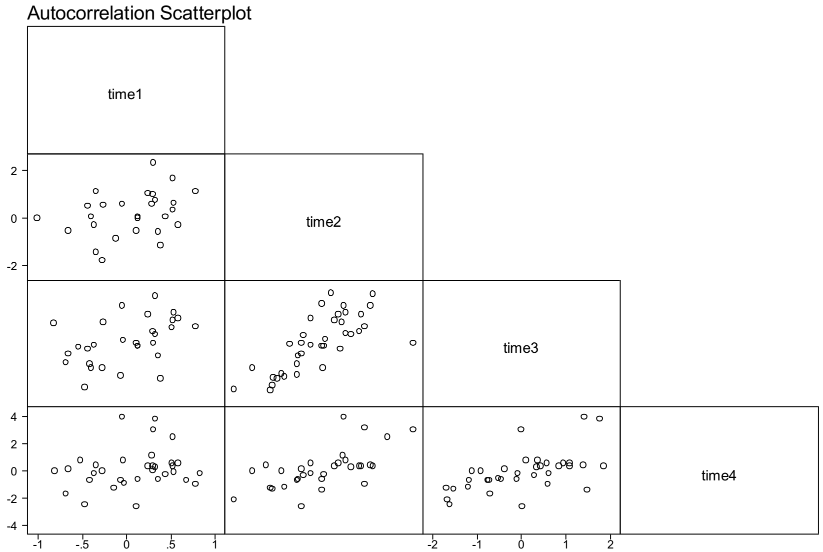
- Measurements of weight were recorded for children up to 4 occasions at roughly 6 weeks, and then at 8,12, and 27 months

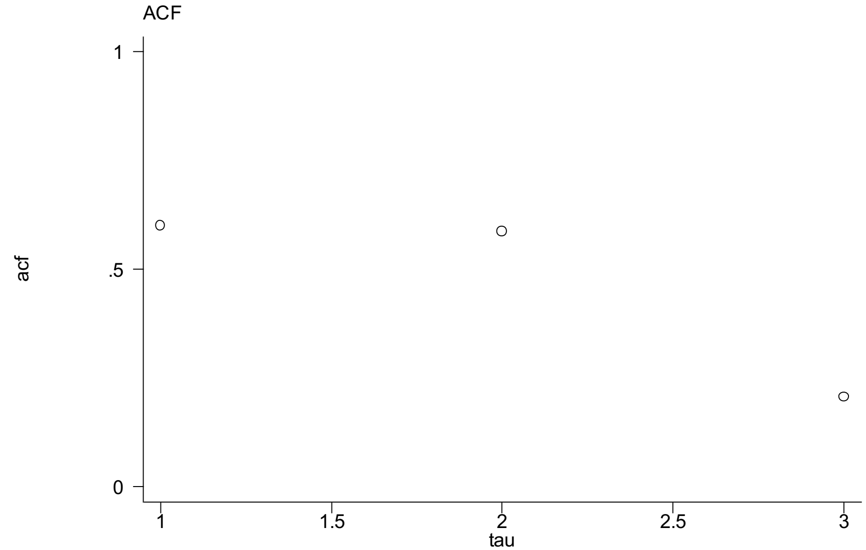
- Goal: We want to investigate the growth trajectories of children’s weights as they get older

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

Observed Data








```
. 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 |
     |-----------|
  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

```
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
```

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:

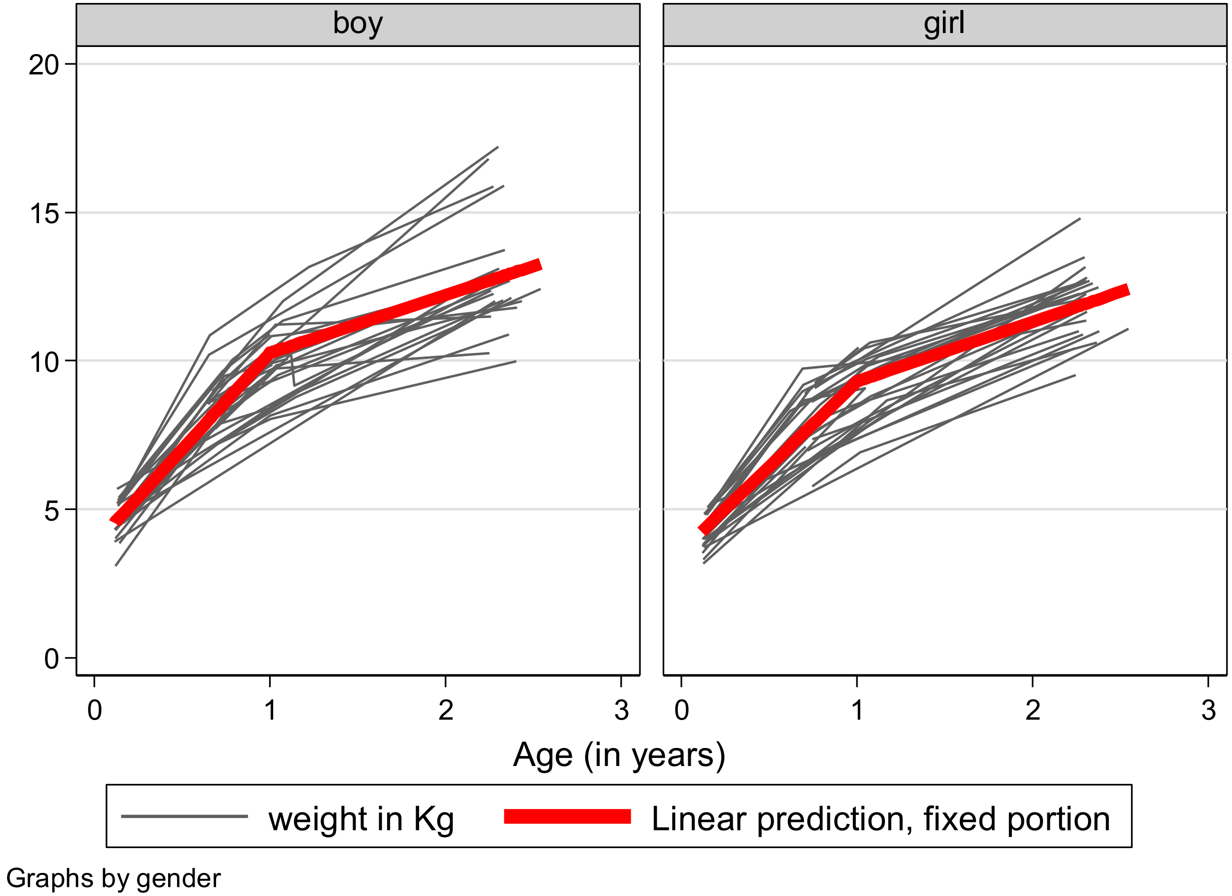
Random Intercept Model

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
id: Identity				
var(_cons)	.7161207	.1608949	.4610435	1.112322
var(Residual)	.5727228	.0730425	.4460526	.7353648

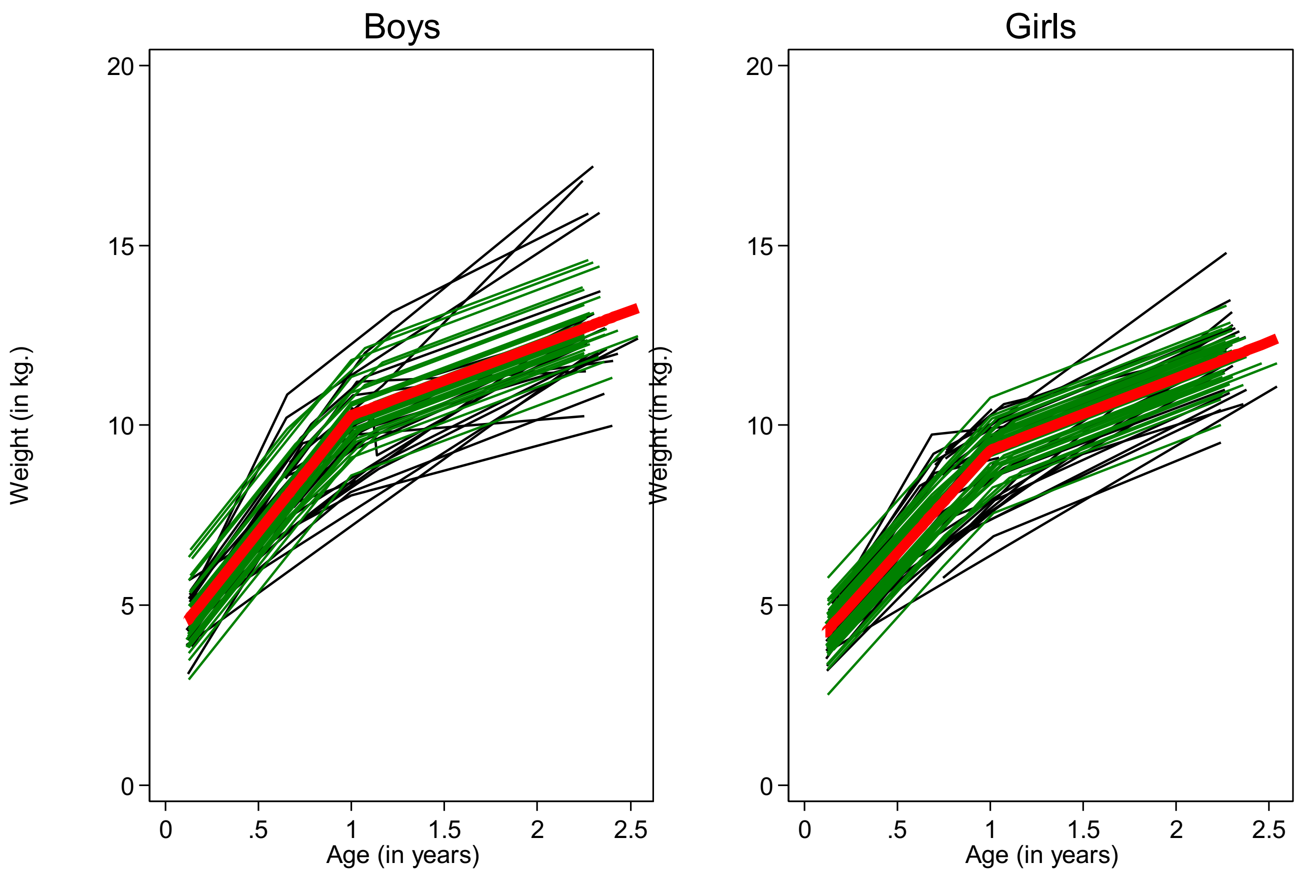
LR test vs. linear regression: chibar2(01) = 56.33 Prob >= chibar2 = 0.0000

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

Predicted Mean Growth



Predicted Mean vs. Individual Growth



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
  - 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
  - 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.

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
- 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
```

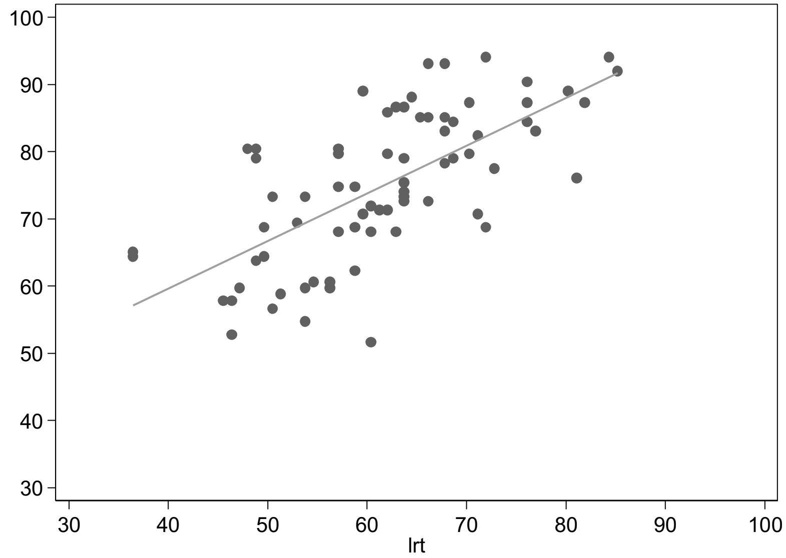
65 schools in the dataset:

Number of students ranges from 2 to 198, average 62

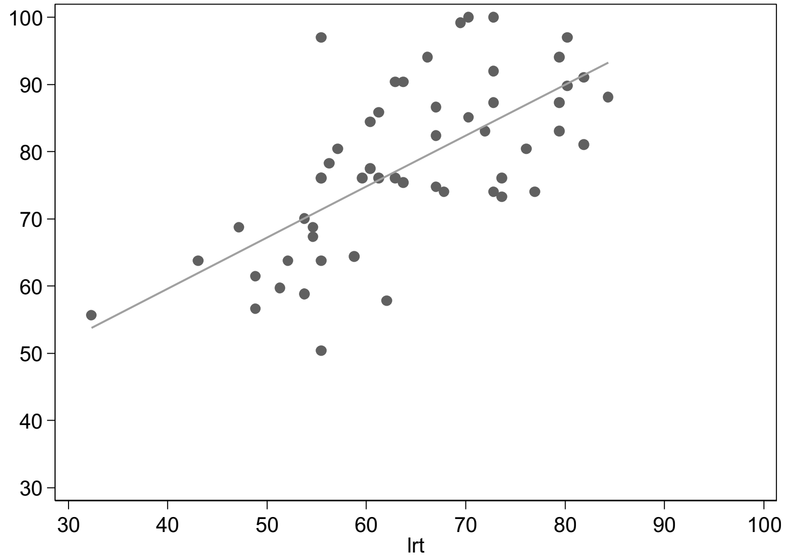
Exploratory Data Analysis

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

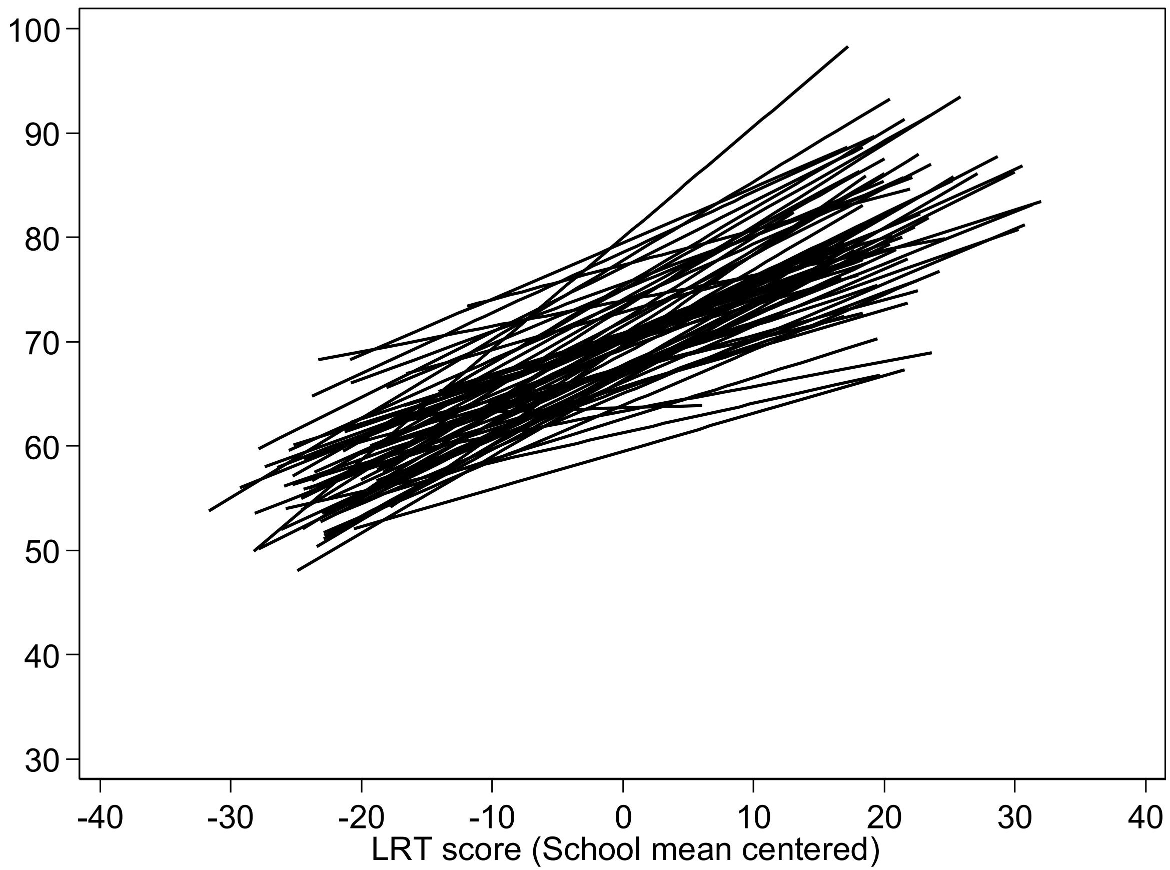
School: 1



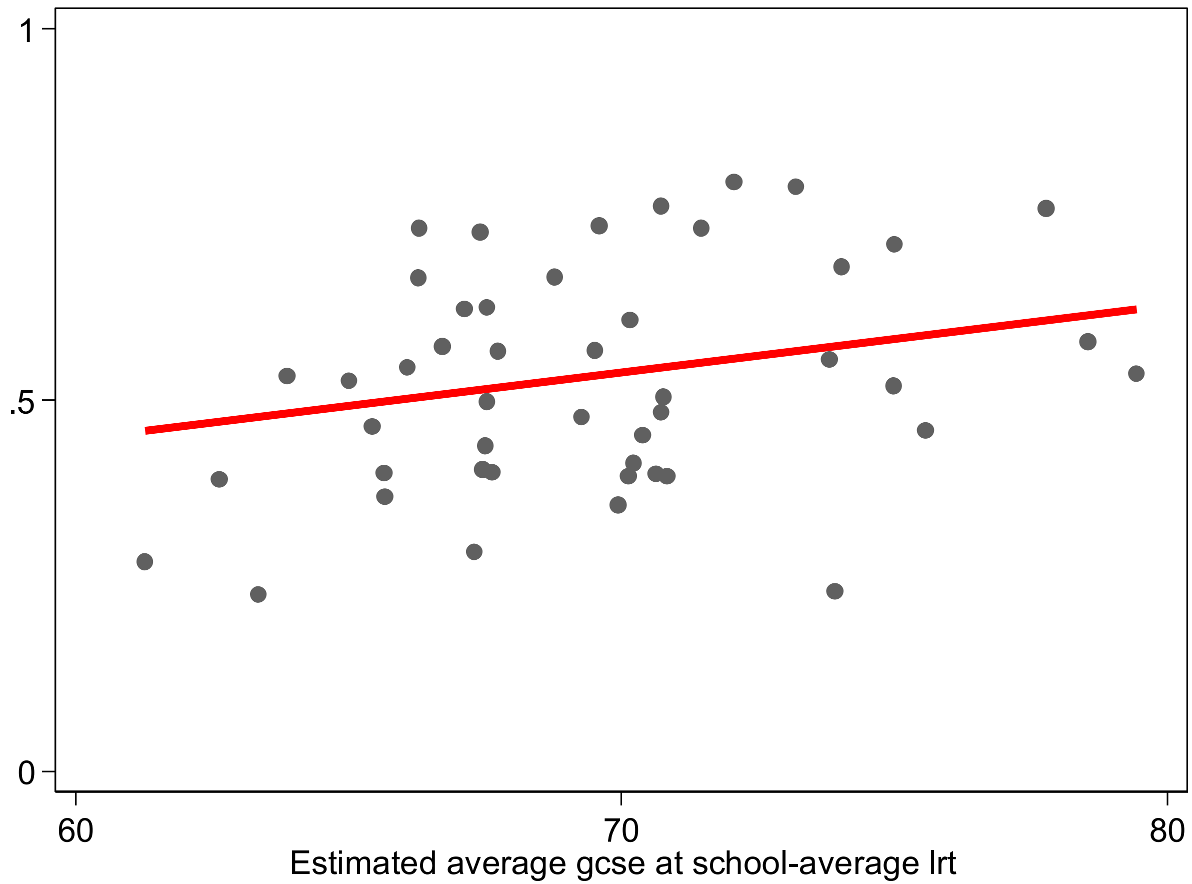
School: 2



School-specific relationships among schools with at least 5 students



Association between school-specific slope and intercept



Linear regression model with random intercept and random slope

Y_{ij} gcse for student j in school i

x_{ij} lrt for student j in school i

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

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

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

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

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

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

Alternative representations of the same model

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

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

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

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

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

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

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

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

Random coefficient models induce heteroskedastic error structure!

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

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

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

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

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

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

```
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

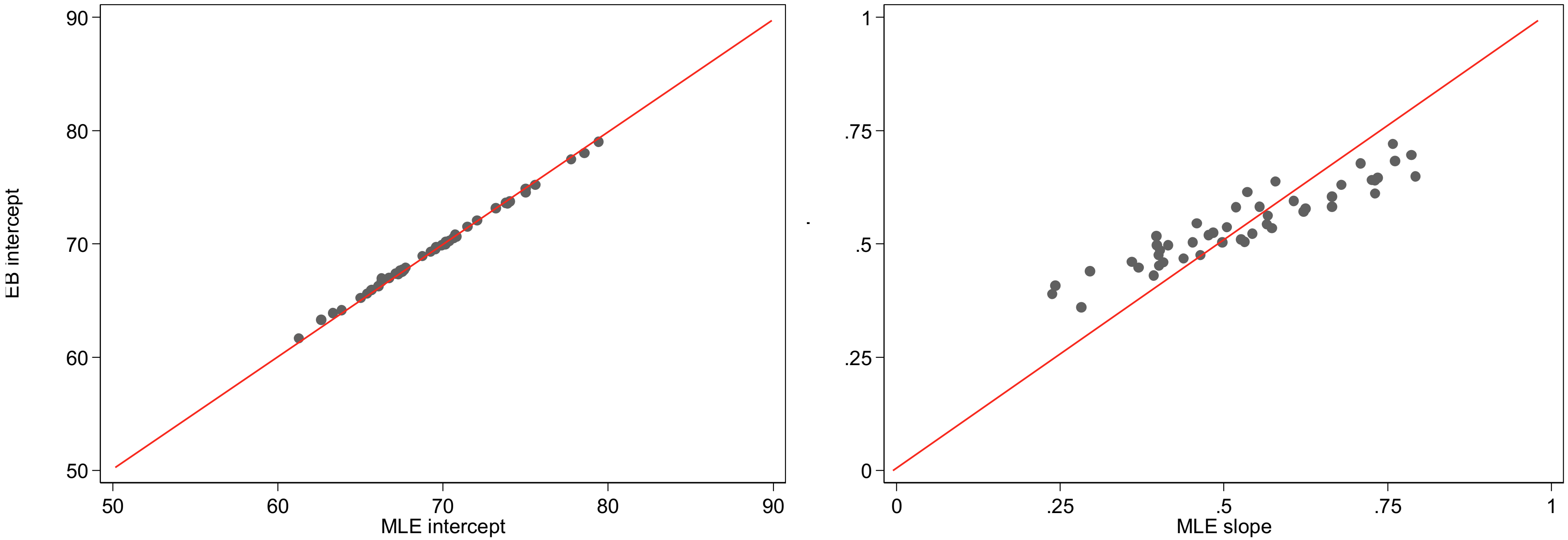
Interpretation of random effects

```
------------------------------------------------------------------------------
  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
------------------------------------------------------------------------------
```

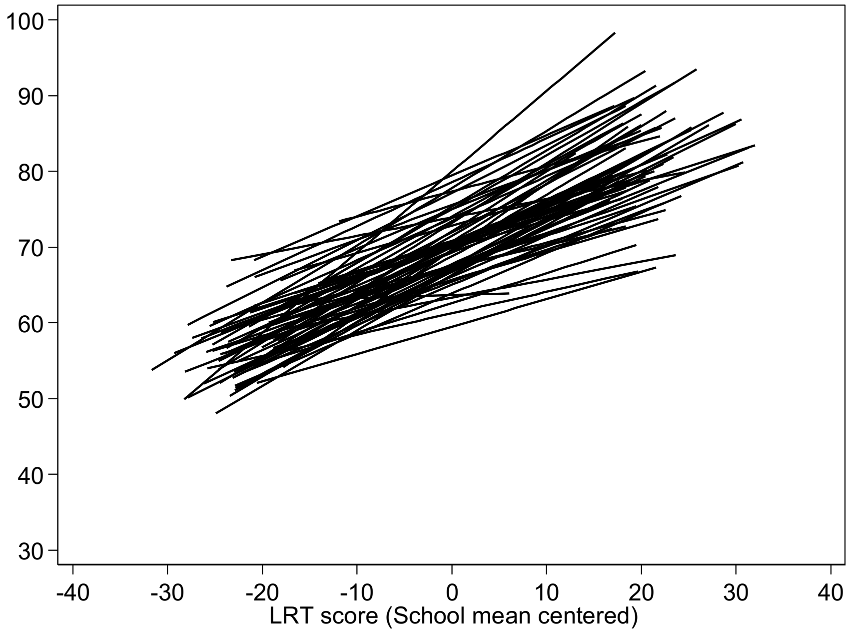
Interpretation of random effects

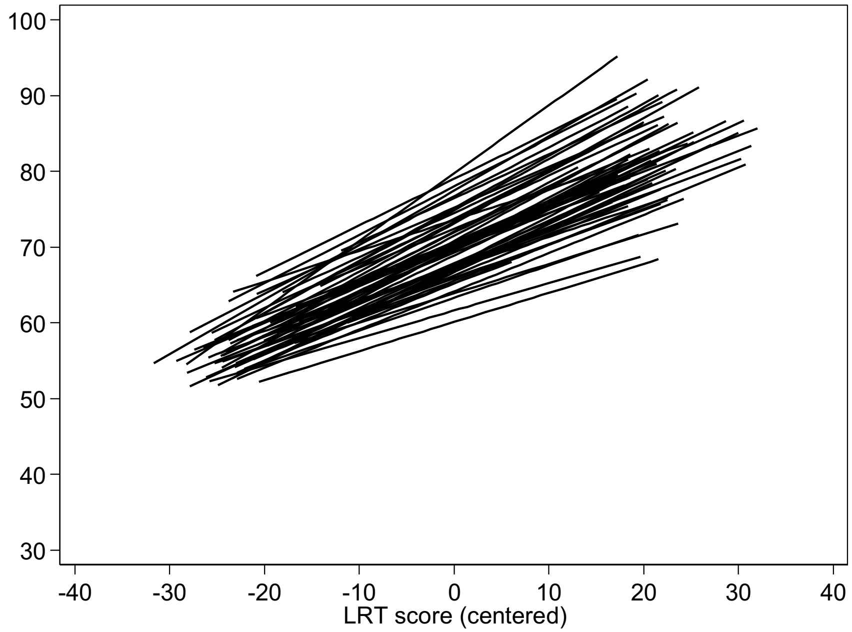
Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
school: Unstructured				
sd(lrt_gr~c)	.1202955	.0190615	.0881804	.1641069
sd(_cons)	4.220941	.3997349	3.505889	5.081833
corr(lrt_gr~c,_cons)	.5481782	.13837	.2241939	.7630512
sd(Residual)	7.432694	.0838333	7.270187	7.598834

Impact of shrinkage on intercepts and slopes



Impact of shrinkage on predicted association between gcse and lrt





Goal: Rank the schools in terms of performance

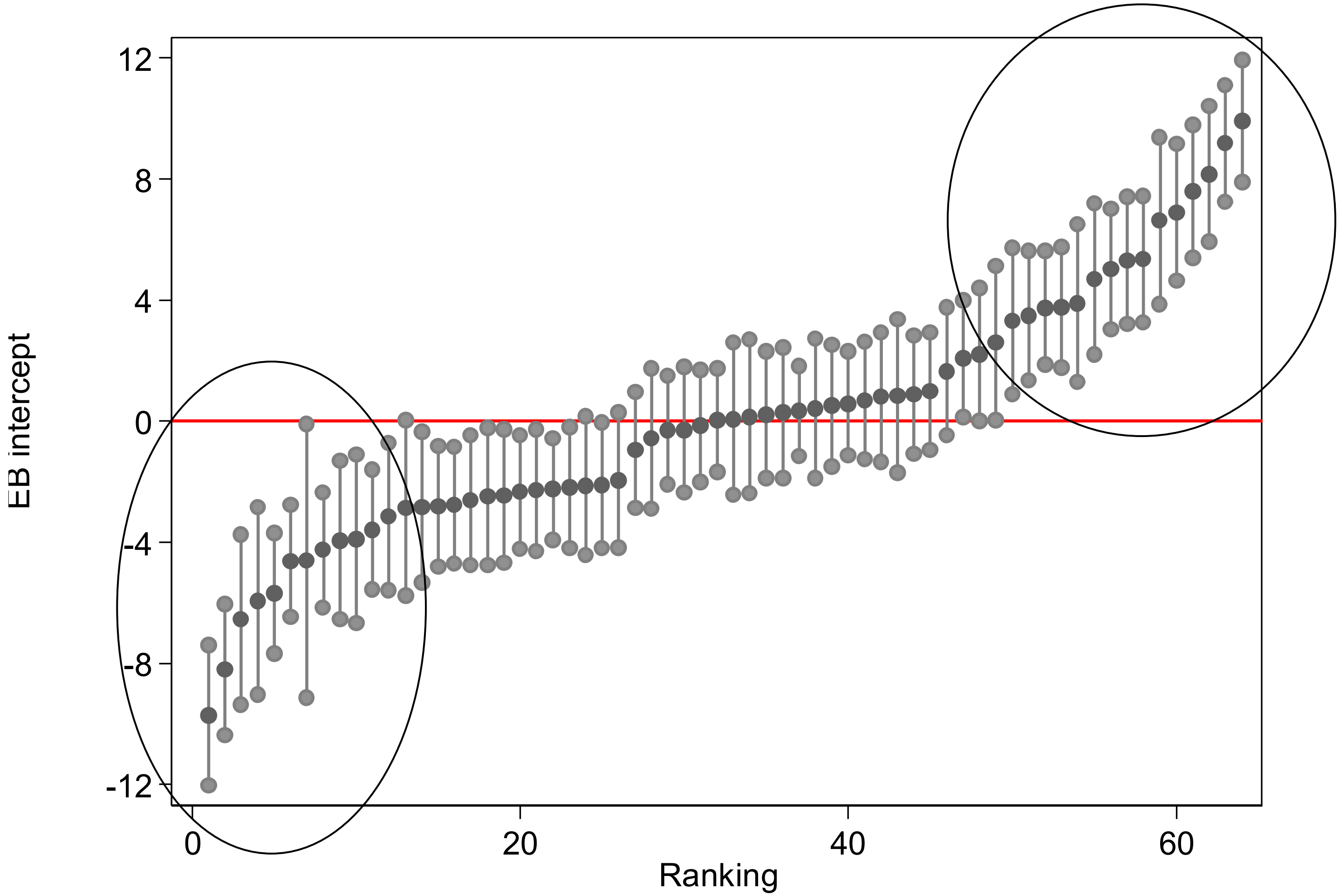
What is the appropriate measure to use from the analysis?

- random intercept?
- random slope?

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

School Rankings

Good performers!



Bad performers!

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