UMC Utrecht

Epidemiology and Big Data

Mixed Models 1: Introduction to Multilevel Models

Rebecca Stellato

Objectives for this week

- At the end of this week, the student will:
    - recognize multi-level and longitudinal study designs
    - be able to explain the difference between fixed and random effects, and know when to use random effects
    - know when to apply a linear mixed model
    - be able to perform linear mixed models using R

Overview Lecture 1: Multilevel Modelling

- Introduction to multilevel data
- Example: multilevel data (children within schools)
- The problem, and some possible solutions
- The mixed model solution
- Adding random effects (random intercept, random slope)
- Adding fixed effects (school- and child-level) to the model
- Interpretation of mixed models
- Summary

Examples of multilevel data

- Effect of school environment on exam results
  - Design: hierarchical, where the examination results of a random sample of students within a random sample of schools are compared
- Influence of race and sex on fetal heartbeat during pregnancy
  - Design: repeated measurements on different gestational ages during pregnancy, where the gestational ages were not the same for all women
- Multi-center hypertension trial
  - Design: hierarchical, with 193 patients in 27 centers, DBP measured 5 times per patient over the course of several weeks

Characteristics of multilevel data

- Hierarchical structure of data
  - children within (classrooms within) schools
  - patients within centers
  - measurements within patients
- Variation at all levels
- "Units" within a level expected to be correlated
- Variables can be measured at different levels
  - Level 2:
    - type of school (mixed vs. single-gender)
    - university vs. community hospital
  - Level 1:
    - reading ability of child at intake
    - gender of patient

Example: London Schools

- Data collected by Goldstein, Rasbash, et al (1993) on 4059 children in 65 schools in Inner London.

- Question: is examination achievement related to intake achievement level, pupil gender, school type and exam achievement of school (averaged over all pupils)?

- Subquestion: do girls do better at a mixed or all-girls' school?

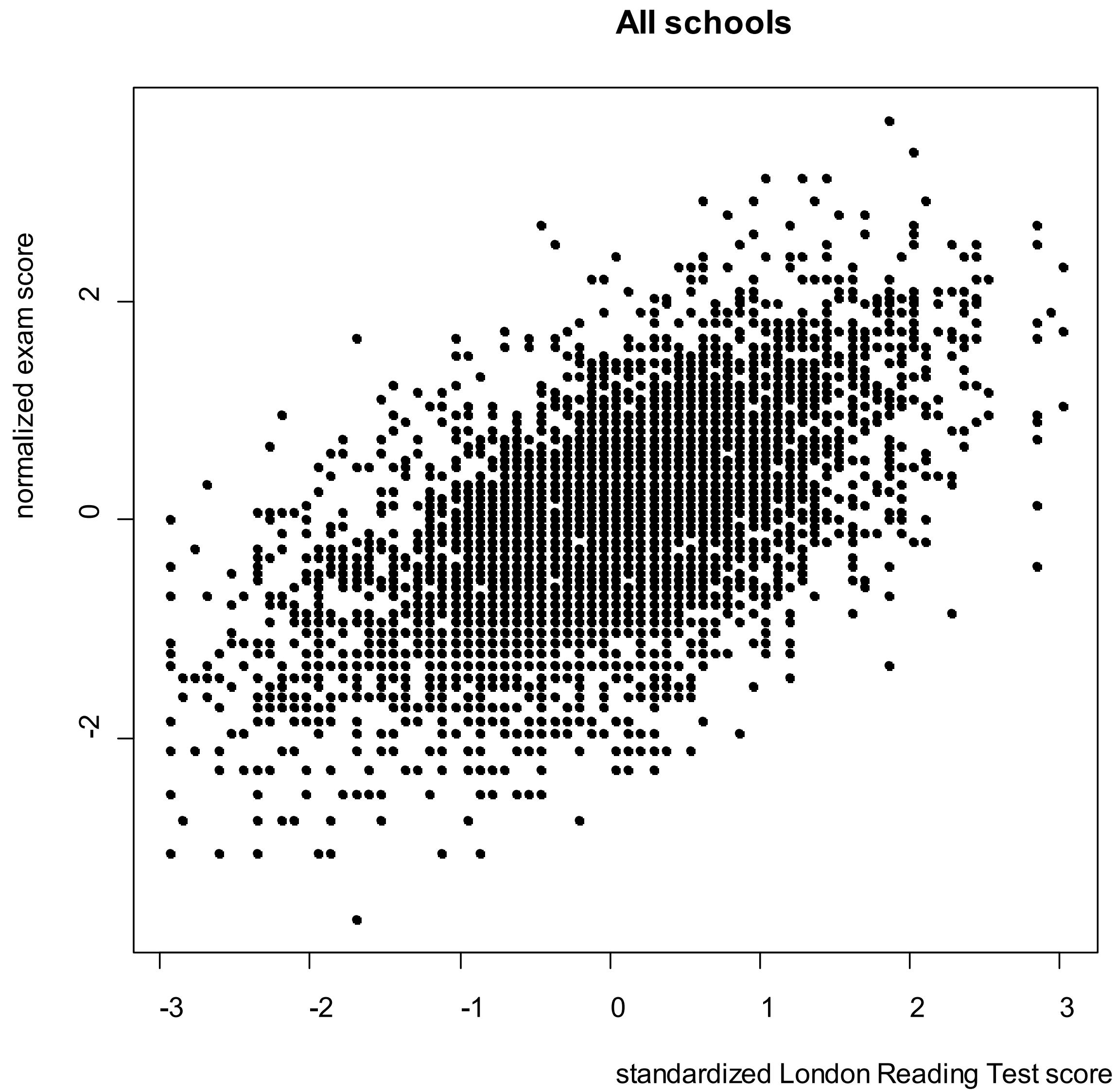
Example: London Schools

- Variables in dataset:
    - School ID
    - Student ID
    - Normalised exam score (outcome variable)
    - Standardised LR test score
    - Student gender
    - School gender
    - School average of intake score
    - Student level Verbal Reasoning (VR) score category at intake
    - Category of students' intake score (averaged)

London Schools

school	# boys	# children	school	# boys	# children
1	45	73	13	26	64
2	0	55	14	92	198
3	29	52	15	47	91
4	45	79	16	0	88
5	16	35	17	31	126
6	0	80	18	0	120
7	0	88	19	33	55
8	0	102	20	21	39
9	21	34	21	0	73
10	31	50	22	48	90
11	62	62	.	.	.
12	23	47	.	.	.
			.	.	.

London Schools

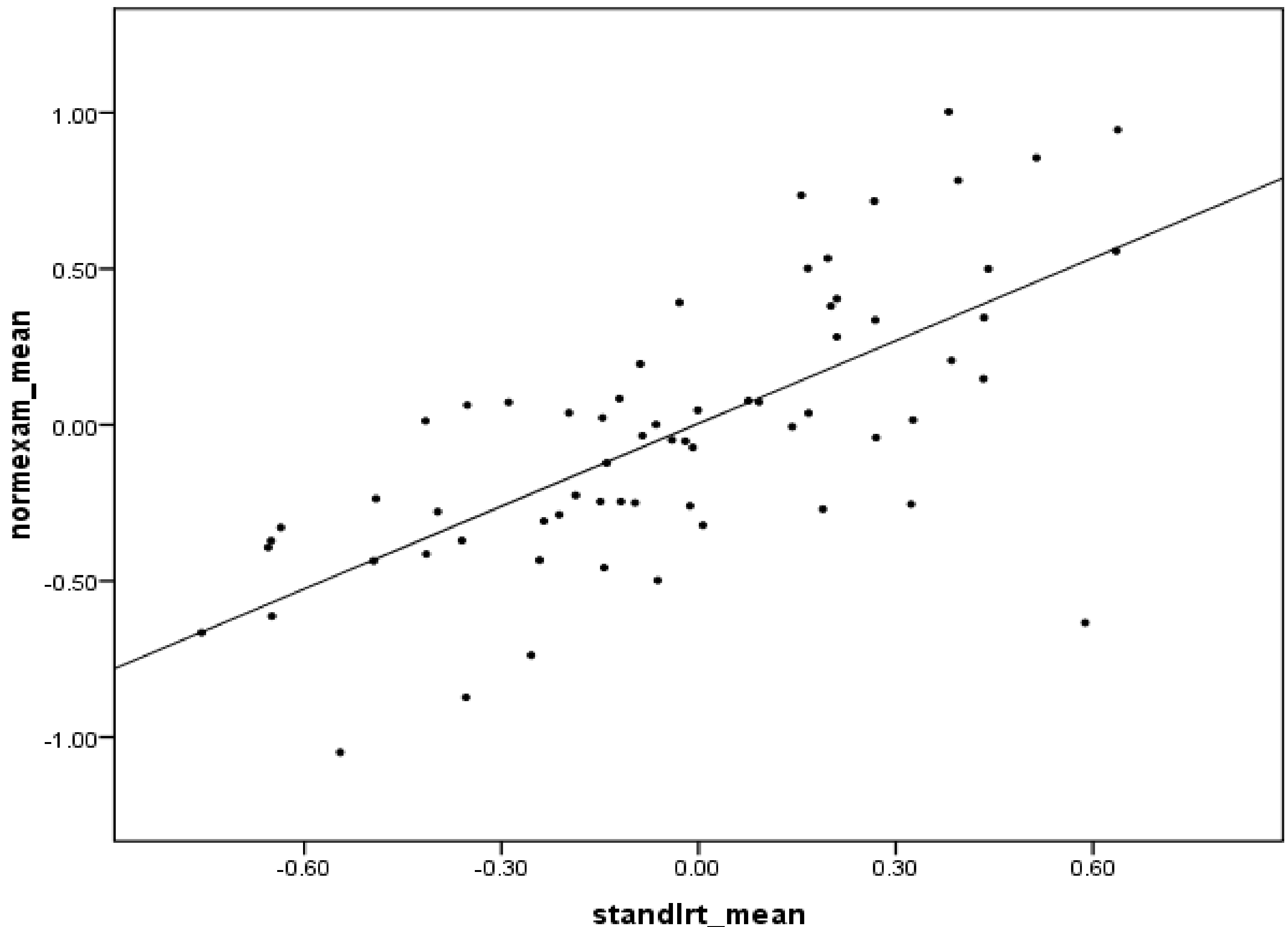


London Schools

- How to analyze relation between exam score and LRT score?
    1. linear regression, mean exam per school vs mean LRT (“aggregated data”)
    2. linear regression, all schools together (“disaggregated data”)
    3. linear regression per school (stratified analysis)
    4. linear regression, all schools together, regression with main effect and interactions to allow for different intercepts and slopes (fully stratified model)
    5. Linear mixed model

London Schools:

1. linear regression, aggregated mean exam vs mean LRT



London Schools:

1. linear regression, aggregated mean exam vs mean LRT

```
> agglondon= aggregate(london, by= list(london$school), FUN=mean)
> head(agglondon)
  Group.1 school student   normexam    standlrt    gender schgend   avslrt schav   vrband mixed
1       1      1    37.0 0.50120348 0.16617305 0.3835616       1 0.166170     2 1.712329     1
2       2      2    28.0 0.78309603 0.39514738 1.0000000       3 0.395150     3 1.636364     0
3       3      3    26.5 0.85543873 0.51415485 0.4423077       1 0.514160     3 1.519231     1
4       4      4    40.0 0.07362567 0.09176214 0.4303797       1 0.091764     2 1.746835     1
5       5      5    18.0 0.40360263 0.21052226 0.5428571       1 0.210520     3 1.657143     1
6       6      6    40.5 0.94456957 0.63765269 1.0000000       3 0.637660     3 1.462500     0
> aggmeanmodel= lm(normexam ~ standlrt, agglondon)
> summary(aggmeanmodel)

Call:
lm(formula = normexam ~ standlrt, data = agglondon)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.15787 -0.13819 -0.00342  0.19873  0.66268 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.004563   0.039737   0.115    0.909    
standlrt    0.883721   0.116016   7.617 1.67e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3191 on 63 degrees of freedom
Multiple R-squared:  0.4794,    Adjusted R-squared:  0.4712 
F-statistic: 58.02 on 1 and 63 DF,  p-value: 1.668e-10
```

estimate for intercept: 0.005 (se 0.040)

estimate for slope: 0.884 (se 0.116)

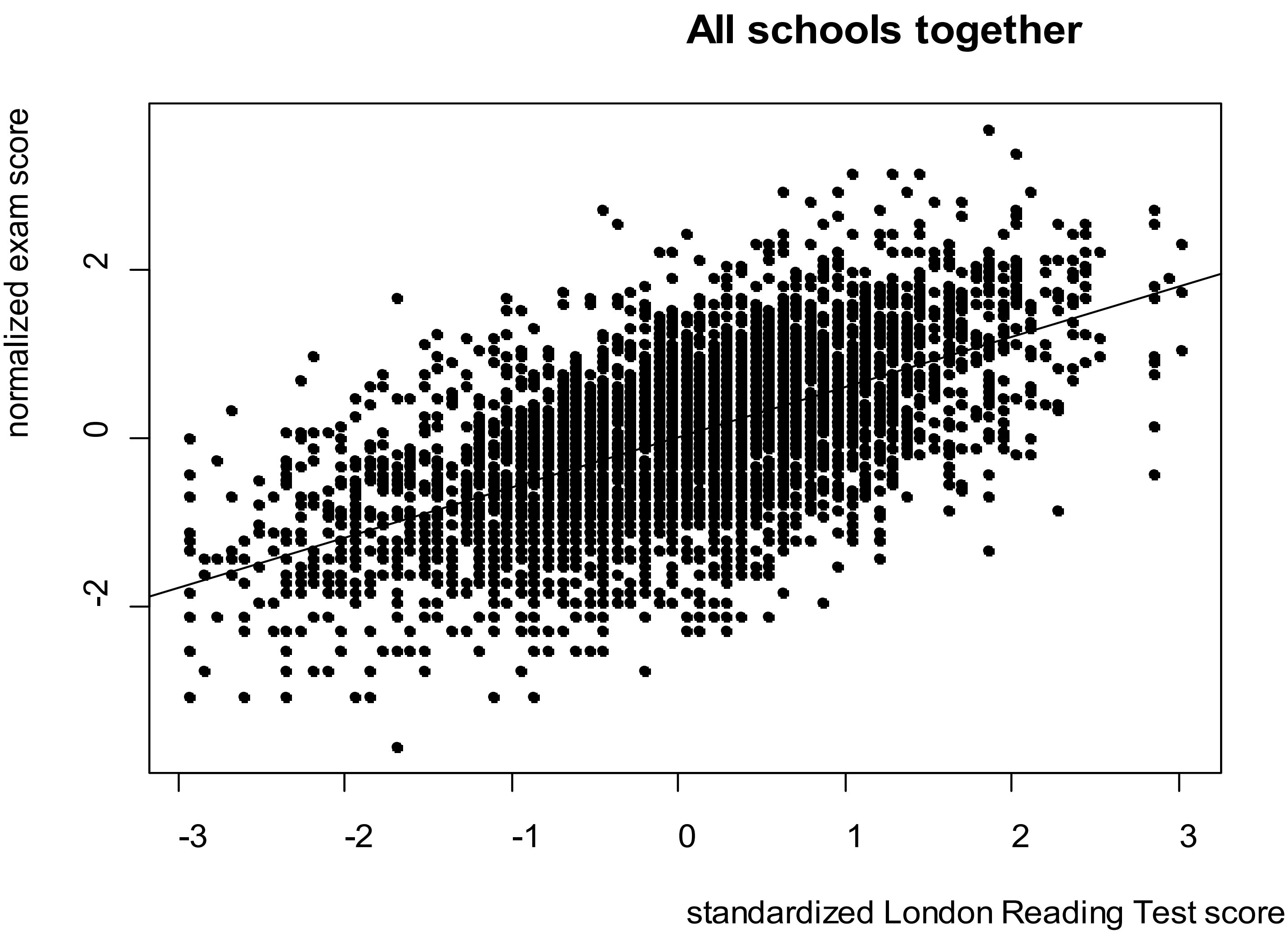
London Schools:

1. linear regression, aggregated mean exam vs mean LRT

- Disadvantages:
    - every school (regardless of sample size) given equal weight
    - N = 65
    - school-level variables possible, but not child-level variables
    - we can only make inference at school level, not child-level
    - possibility of “ecological fallacy”

London Schools

2. linear regression, all schools together



London Schools

2. linear regression, all schools together

```
> disagmod= lm(normexam ~ standlrt, data=london)
> summary(disagmod)

Call:
lm(formula = normexam ~ standlrt, data = london)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.65617 -0.51847  0.01265  0.54397  2.97399 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.001195   0.012642  -0.095    0.925    
standlrt     0.595055   0.012730  46.744   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8054 on 4057 degrees of freedom
Multiple R-squared:   0.35,     Adjusted R-squared:  0.3499 
F-statistic:  2185 on 1 and 4057 DF,  p-value: < 2.2e-16
```

estimate for intercept: - 0.001 (se 0.013)

estimate for slope: 0.595 (se 0.013)

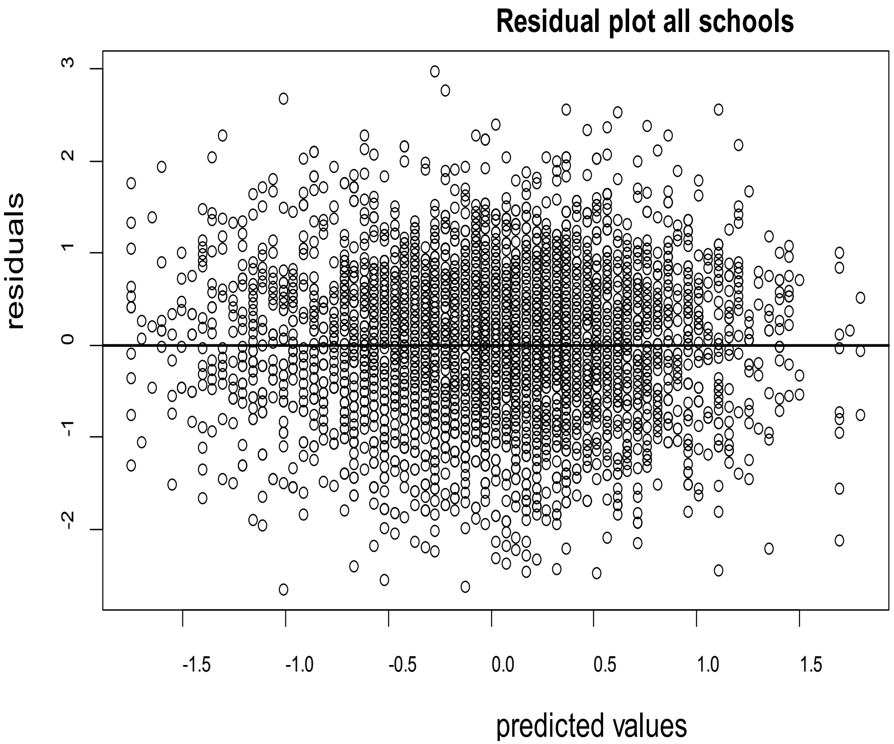
London Schools:

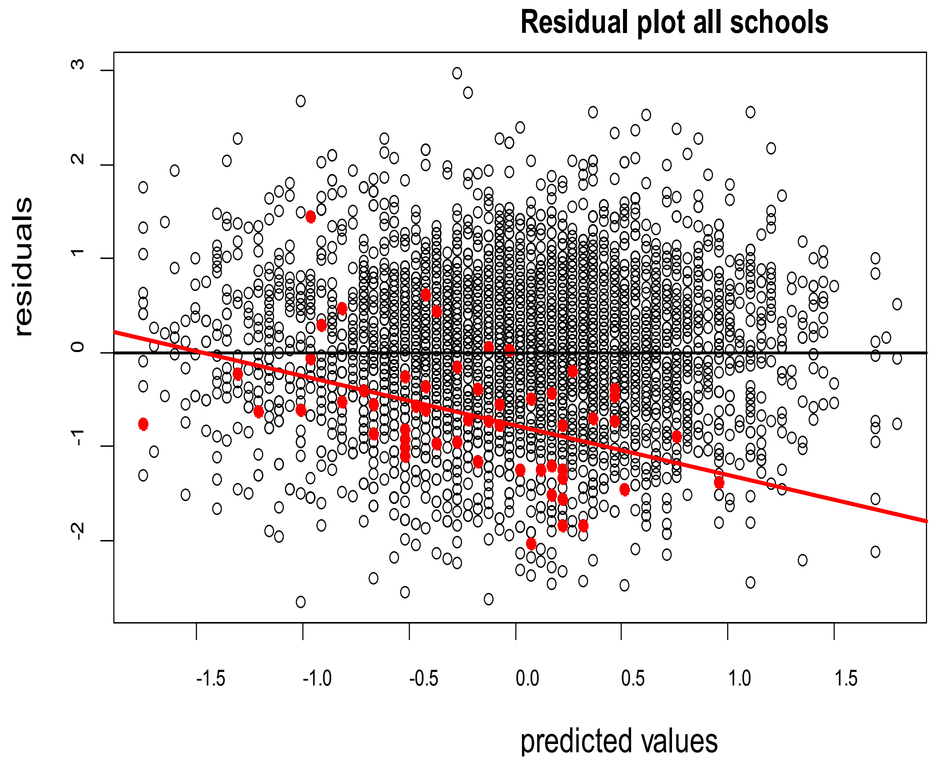
2. linear regression, all schools together

- Disadvantages:
  - inflates sample size, especially for level-2 variables
    - SE’s of level-2 variables tend to be underestimated → p-values too small, CI’s too narrow (type I error inflated)
    - SE’s of level-1 variables may be over- or underestimated
  - ignore correlated residuals (correlation of children within schools)

London Schools:

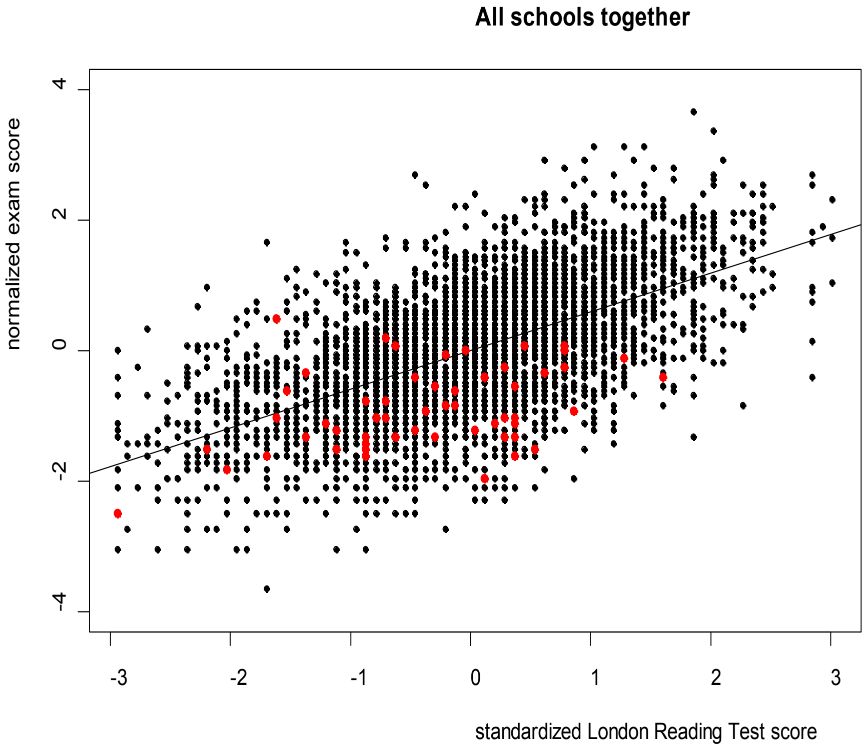
2. linear regression, all schools together

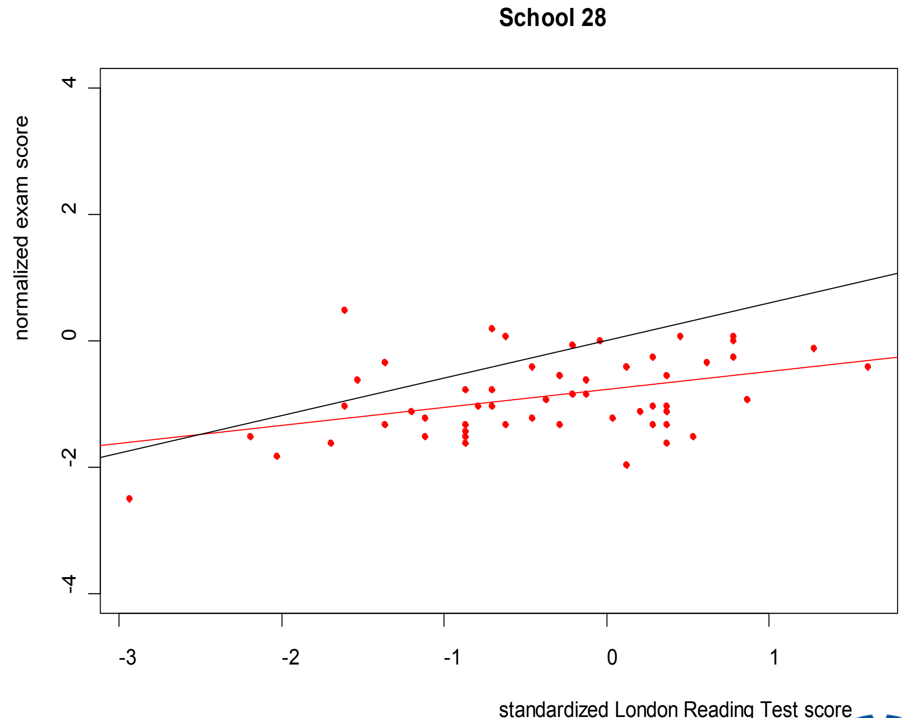




London Schools:

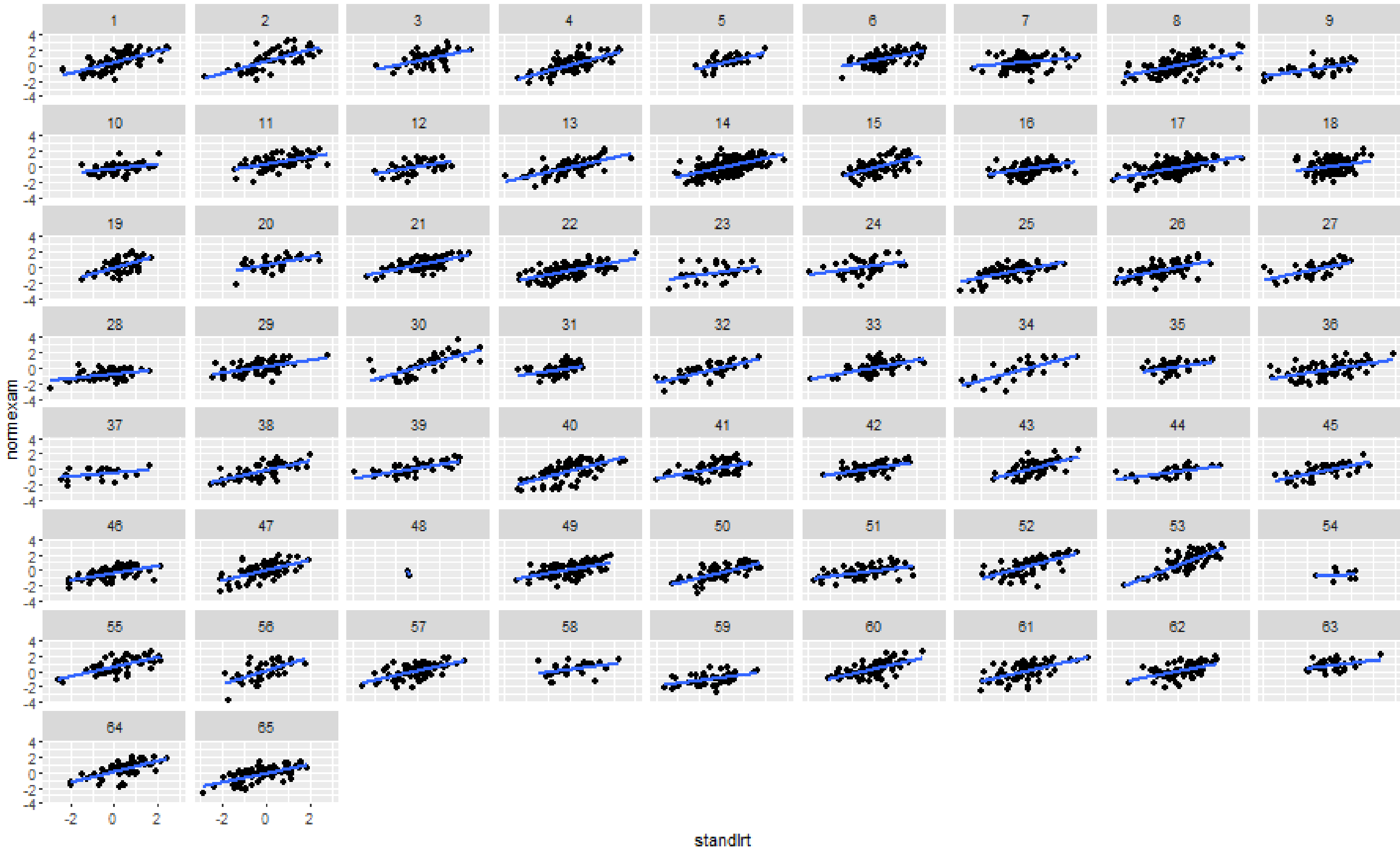
2. linear regression, all schools together





London Schools

3. linear regression per school



London Schools

3. linear regression per school

School	Intercept	slope
1	0.383330189	0.70934058
2	0.482275275	0.76128749
3	0.557750538	0.57898548
4	0.003753722	0.76144638
5	0.260443999	0.68001660
6	0.603206568	0.53534316
....		

- summary intercepts:
  - mean = -0.068; sd = 0.519; sem = 0.064
- summary slopes:
  - mean = 0.425; sd = 0.939; sem = 0.116

London Schools

3. linear regression per school

- Disadvantages:
  - 65 different regressions, how to combine the results?
    - mean slope: every school has equal weight
    - standard error of parameter estimate correct?
  - child-level variables possible, but not school-level variables

London Schools

4. all schools together, main effects and interactions

- Advantage over previous analysis:
  - now we can include both child- and school-level variables
  - residuals probably normally distributed (with constant variance?) around individual lines
- Disadvantages:
  - We wanted 1 intercept and 1 slope for LRT, but:
  - 65 schools, so 1 reference category and 64 estimates for intercepts (main effects per school) + 64 estimates for interactions (slopes per school)!
    - Which school is the reference?
  - We can't generalize beyond these 65 schools
  - This model uses 128 extra df for all those intercepts & slopes

London Schools: models so far

Model	overall/fixed slope LRT	s.e.
1. aggregated data	0.884	0.116
2. disaggregated data	0.595	0.013
3. regr. per school	0.425	0.116
4. school*LRT interactions	??	??

London Schools

5. Mixed Models

- Advantages:
    - sample size correct, account for correlation of children within schools
        - so: correct SE's/p-values/CI's
    - no need for 64 main effects and interactions
        - differences between schools captured one or more 'variance components'
    - both child-level and school-level variables simultaneously
        - so: inference for both children and schools
        - interactions between child- and school-level variables possible
    - examine variation at different levels
    - models work well in presence of missing outcomes (longitudinal)

Mixed Models

- Mixed models made up of
  - fixed effects
  - random effects
- Sometimes (inaccurately) called “random effects models”
- Also sometimes called “random coefficient” models
- Some variables (or: their coefficients) can be included as both “fixed” (of interest) and “random” (random variation across the level-2 units)

Mixed Models: what is a "fixed effect"?

- Fixed effect: variable of interest
  - overall intercept (not really of interest)
  - overall slope for LRT (to help make predictions of exam performance)
  - other fixed effects of interest:
    - gender (difference between boys and girls?)
    - type of school (boys', girls', mixed)
    - "achievement level" of school
    - …

Mixed Models: what is a “random effect”?

- A random intercept per school allows schools to have different intercepts
- A random effect for LRT per school allows the effect of LRT on exam score to differ per school (“random slope for LRT” = different slope for exam-LRT relation for each school)
- Random effect (“slope”) can also be for a categorical variable
  - difference between boys and girls on exam score could differ per school
  - treatment effect on an outcome can be thought to vary per center in a multi-center study
- All variables of interest are added as fixed
- Depending on theory, none/one/some fixed variables may also be modelled as random

Mixed Models: what is a "random effect"?

- Why "random effect"?
- Schools are *random* sample of all Inner London schools
  - intercepts (and LRT slopes) from these schools are a random sample from all possible intercepts and slopes
  - intercepts (and LRT slopes?) differ from one another, but
  - interest not in estimating the intercept and slope per school, thus
  - sufficient to estimate the variances of the intercepts and slopes
  - intercepts (and slopes) thought to come from normal distributions with mean 0 and variances $\sigma^2_{\upsilon 0}$ and $\sigma^2_{\upsilon 1}$, and covariance $\sigma_{\upsilon 01}$
  - in this way we only have to estimate 3 extra parameters, not 128

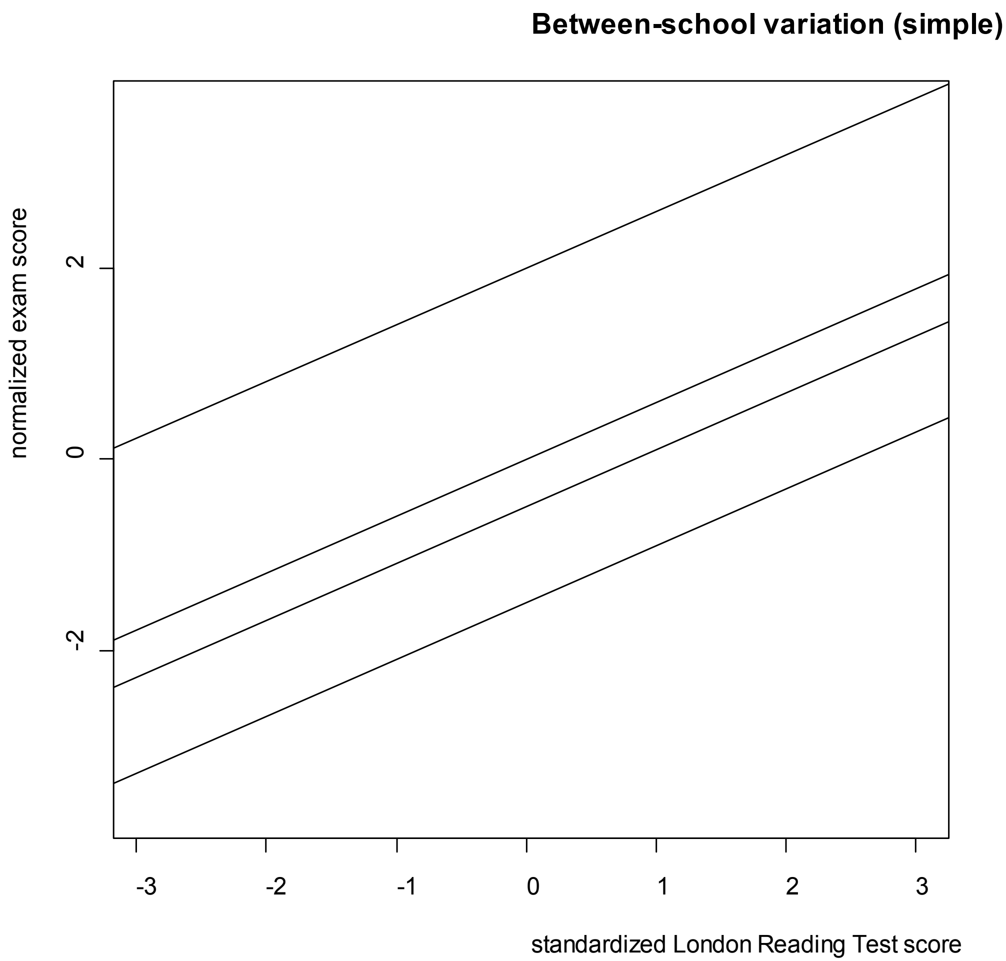
Interlude: some notation

- level-1 (child) model: $y_{ij} = b_{0i} + b_{1i} \cdot x_{1ij} + \varepsilon_{ij}$
- level-2 (school) model: $b_{0i} = \beta_0 + \upsilon_{0i}$; $b_{1i} = \beta_1 + \upsilon_{1i}$
- combine the two: $y_{ij} = \beta_0 + \upsilon_{0i} + \beta_1 \cdot x_{1ij} + \upsilon_{1i} \cdot x_{1ij} + \varepsilon_{ij}$
  - rewrite: $y_{ij} = (\beta_0 + \upsilon_{0i}) + (\beta_1 + \upsilon_{1i}) \cdot x_{1ij} + \varepsilon_{ij}$

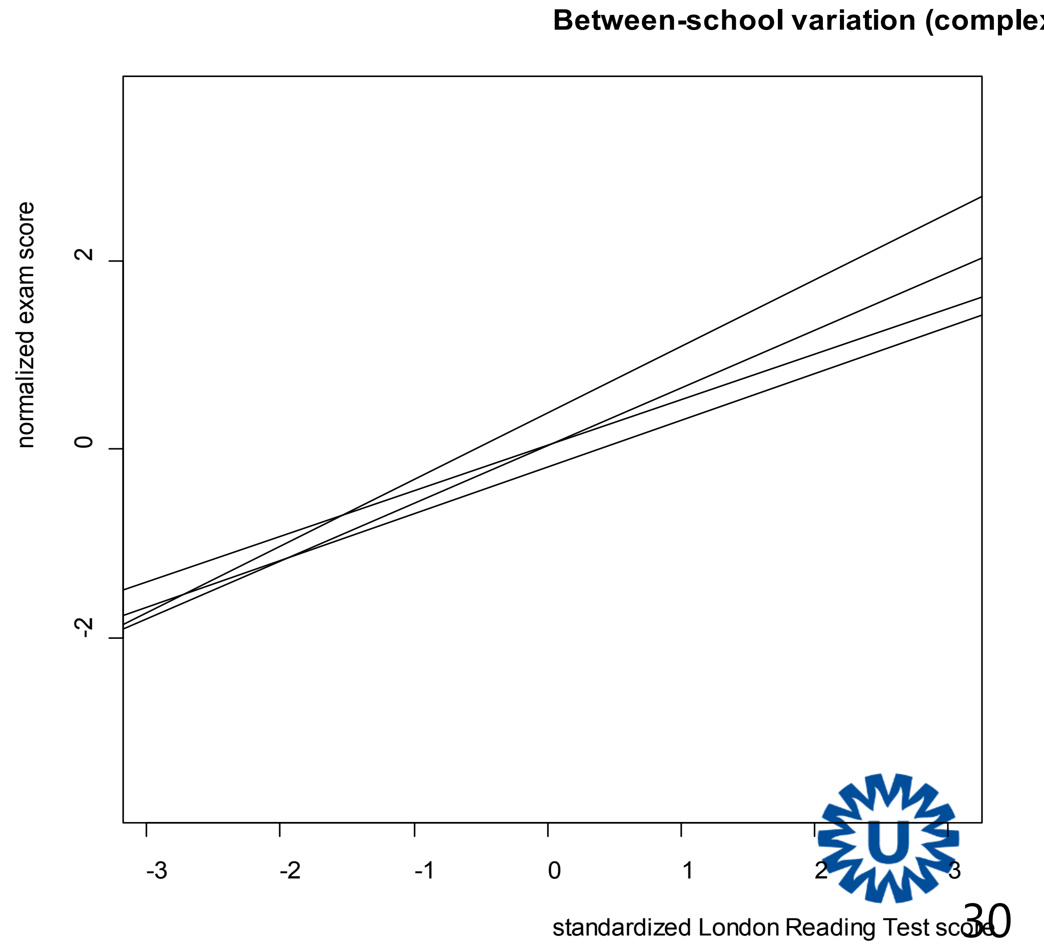
- y_{ij} : outcome (exam score) for jth child in ith school
- x_{1ij}: 1st explanatory var (LRT score) at level 1 (jth child in ith school)
- β_0, β_1, … : regression coefficients for overall effects of explanatory vars ("fixed effects")
- υ_{0i} : individual effect of ith school on intercept ("random effect")
- υ_{1i} : individual effect of ith school on slope (for LRT) ("random effect")
- ε_{ij} : level-1 residual (jth child in ith school)

Mixed Models: what is a “random effect”?

Random intercept only:

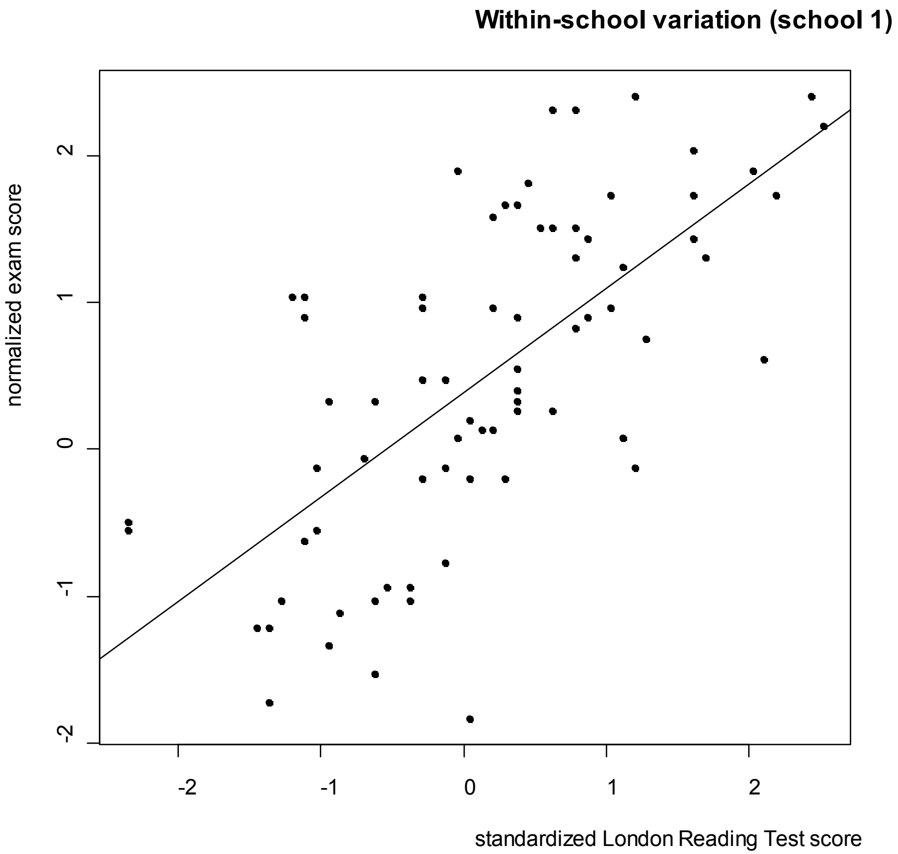


Random intercept + random slope:

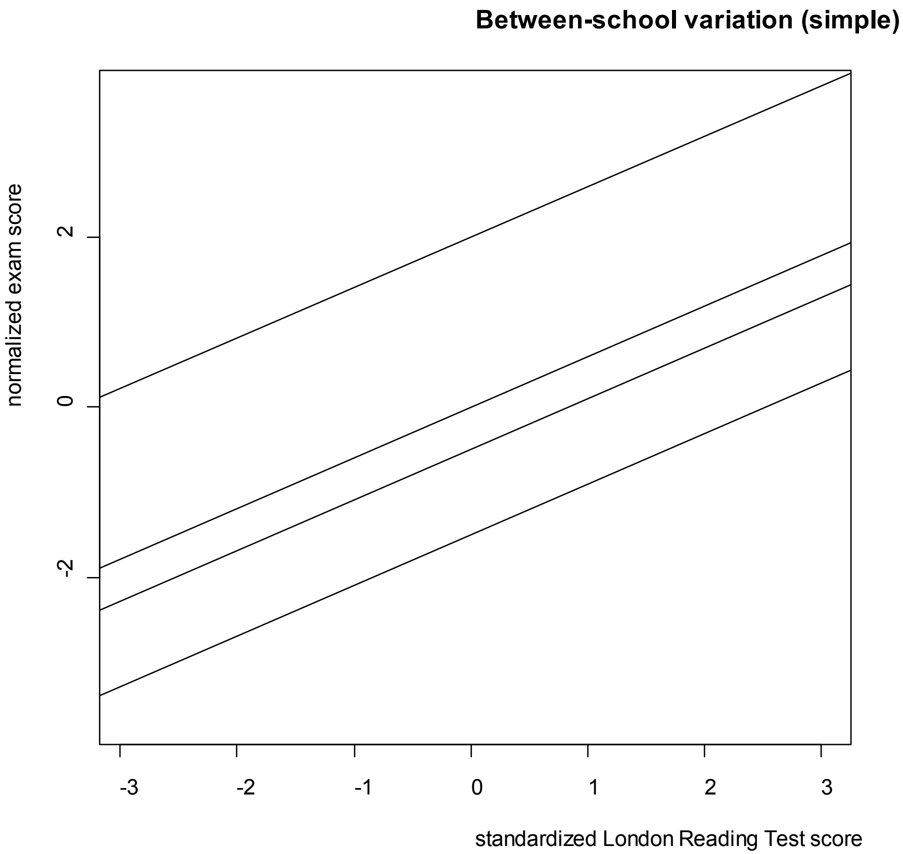


London Schools

$$y_{1j} = \beta_0 + \beta_1 X_{11j} + \varepsilon_{1j}$$



London Schools

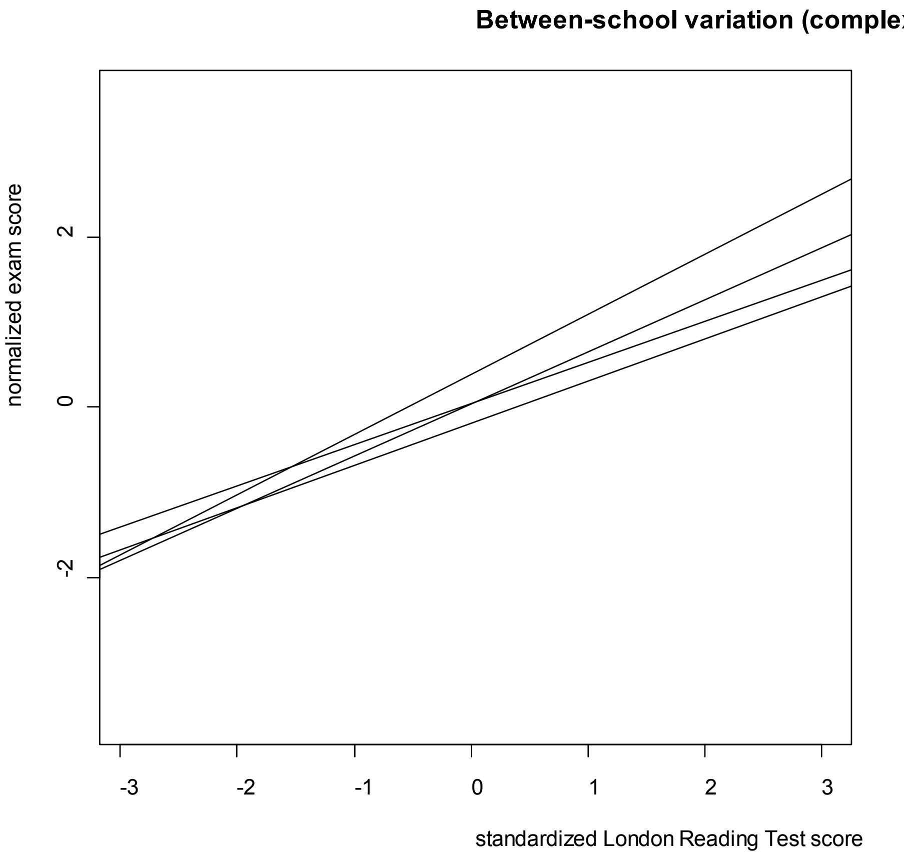


$$y_{1j} = \beta_0 + \upsilon_{01} + \beta_1 X_{11j} + \varepsilon_{1j}$$

$$y_{2j} = \beta_0 + \upsilon_{02} + \beta_1 X_{12j} + \varepsilon_{2j}$$

$$y_{\mathrm{i}j} = \beta_0 + \upsilon_{0i} + \beta_1 X_{1ij} + \varepsilon_{ij}$$

London Schools



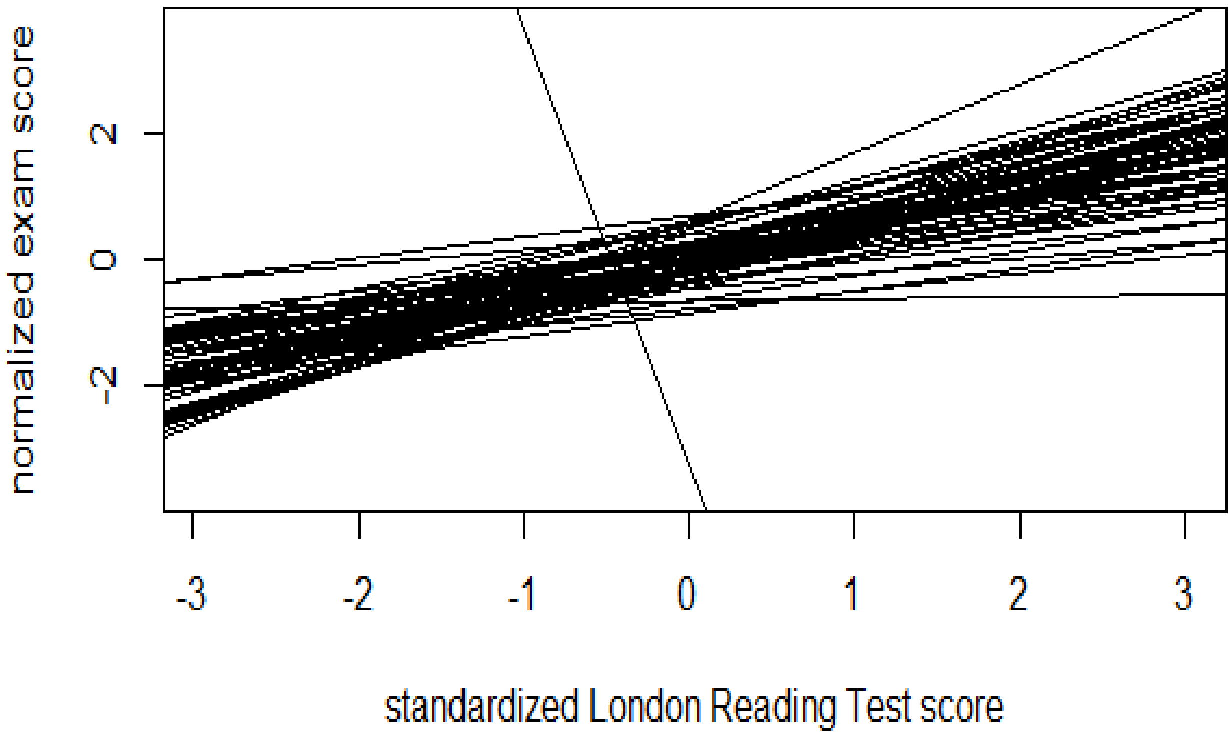
$$y_{1j} = \beta_0 + \upsilon_{01} + \beta_1 X_{11j} + \upsilon_{11} X_{11j} + \varepsilon_{1j}$$

$$y_{2j} = \beta_0 + \upsilon_{02} + \beta_1 X_{12j} + \upsilon_{12} X_{12j} + \varepsilon_{2j}$$

$$y_{ij} = \beta_0 + \upsilon_{0i} + \beta_1 X_{1ij} + \upsilon_{1i} X_{1ij} + \varepsilon_{ij}$$

London Schools

Graph per school ("spaghetti plot"):



Mixed Models: the model

- $y_{ij} = (\beta_0 + \upsilon_{0i}) + (\beta_1 + \upsilon_{1i}) \cdot x_{1ij} + \cdots + \varepsilon_{ij}$
- Where:
  - y_{ij}: outcome (exam score) for jth child in ith school
  - x_{1ij}: first explanatory variable (LRT score) at level 1 (jth child in ith school)
  - β_0, β_1, ... : regression coefficients for explanatory variables ("fixed effects")
  - υ_{0i}: random effect for the intercept in ith school
  - υ_{1i} : random effect for the slope (for LRT) in ith school
  - ε_{ij}: level-1 residual (jth child in ith school)
- Model assumptions:
  - $\varepsilon_{ij} \sim N(0, {\sigma_e}^2)$; $\upsilon_{0i} \sim N(0, {\sigma_{\upsilon 0}}^2)$; $\upsilon_{1i} \sim N(0, {\sigma_{\upsilon 1}}^2)$
  - ε_{ij} independent
  - $cov(\upsilon_{0i}, \upsilon_{1i}) = \sigma_{\upsilon 01}$
  - $cov(\varepsilon_{ij}, \upsilon_{0i}) = cov(\varepsilon_{ij}, \upsilon_{1i}) = 0$

Mixed models in R

Two packages used most frequently

- Package nlme
  - lme() for Gaussian models
  - gls() function for models with correlated errors
  - approximate (Wald) CI's via intervals() function in same package

- Package lme4
  - lmer() for Gaussian models
  - glmer() for generalized linear mixed models (day 2)
  - "profile likelihood" CI's via confint()

- See information on Blackboard

London Schools: mixed model

random intercept only

```
> sch.lme.1 <- lme(fixed=normexam~standlrt, random=~1 | school,
data=london, method="ML")
> summary(sch.lme.1)
Linear mixed-effects model fit by maximum likelihood
 Data: london
       AIC      BIC    logLik
  9365.213 9390.447 -4678.606

Random effects:
 Formula: ~1 | school
        (Intercept)  Residual
StdDev:   0.3035269 0.7521481
```

"`fixed=`" is optional; you could also just use:
`lme(normexam~standlrt, random=~1|school, data=london, method="ML")`

Watch out! R gives the standard deviation of the random effects, not the variance. Var(rand int) = $0.3035^2 = 0.092$; res var=$0.7521^2 = 0.565$

London Schools: mixed model

random intercept only

```
Fixed effects: normexam ~ standlrt
                Value  Std.Error   DF  t-value p-value
(Intercept) 0.0023871 0.04003241 3993  0.05963  0.9525
standlrt    0.5633697 0.01246844 3993 45.18366  0.0000
 Correlation:
         (Intr)
standlrt 0.008

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-3.7161719 -0.6304245  0.0286690  0.6844298  3.2680306

Number of Observations: 4059
Number of Groups: 65
```

London Schools: mixed model

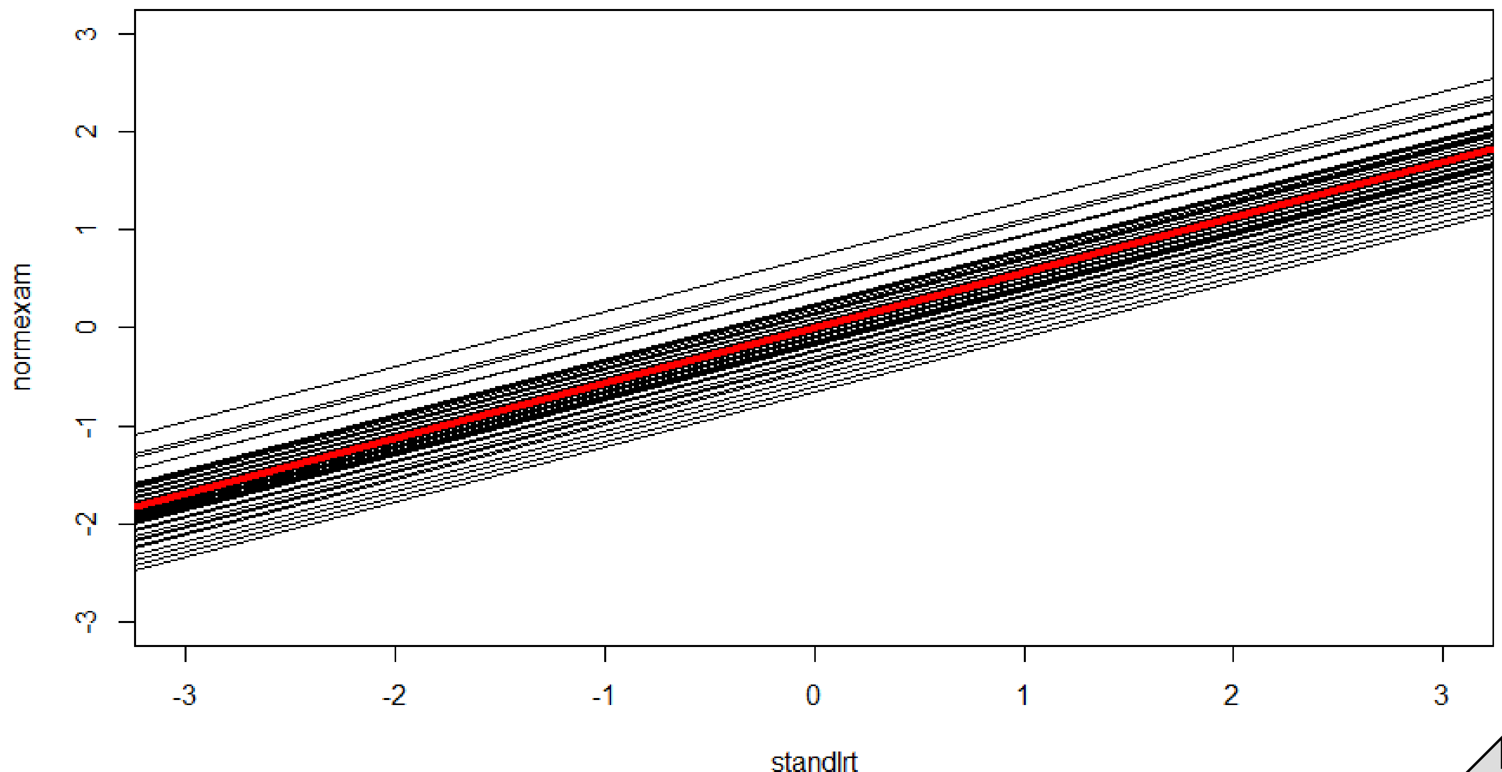
simplest model: only random intercept

- Estimate for fixed intercept is 0.0024
  - (est.) mean exam score for a child with standardized LRT = 0 (mean)
- Estimate for fixed slope is 0.563
  - for every unit (1 sd) increase in LRT score, the exam score increases on average by 0.563 sd (= units of exam score, because normalized)
- Estimate for random intercept (between-school) variance is 0.092
- Estimate for within-school (residual) variance is 0.566
  - In this model, more unexplained variance within than between schools

London Schools: mixed model

simplest model: only random intercept

Fitted model



London Schools: mixed model

random intercept + random slope

```
> sch.lme.2 <- lme(fixed=normexam~standlrt, random=~standlrt | school,
data=london, method="ML")
> summary(sch.lme.2)

Linear mixed-effects model fit by maximum likelihood
 Data: london
      AIC      BIC    logLik
  9328.84 9366.693 -4658.42

Random effects:
 Formula: ~standlrt | school
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 0.3007313 (Intr)
standlrt    0.1205753 0.497
Residual    0.7440777
```

This is equivalent to:
`random=~1+standlrt`

London Schools: mixed model

random intercept + random slope

```
Fixed effects: normexam ~ standlrt
                 Value  Std.Error   DF   t-value p-value
(Intercept) -0.0115074 0.03979173 3993 -0.289192  0.7724
standlrt     0.5567279 0.01994287 3993 27.916142  0.0000
 Correlation:
         (Intr)
standlrt 0.365

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.83123233 -0.63247485  0.03404163  0.68320636  3.45617450

Number of Observations: 4059
Number of Groups: 65
```

London Schools: mixed model

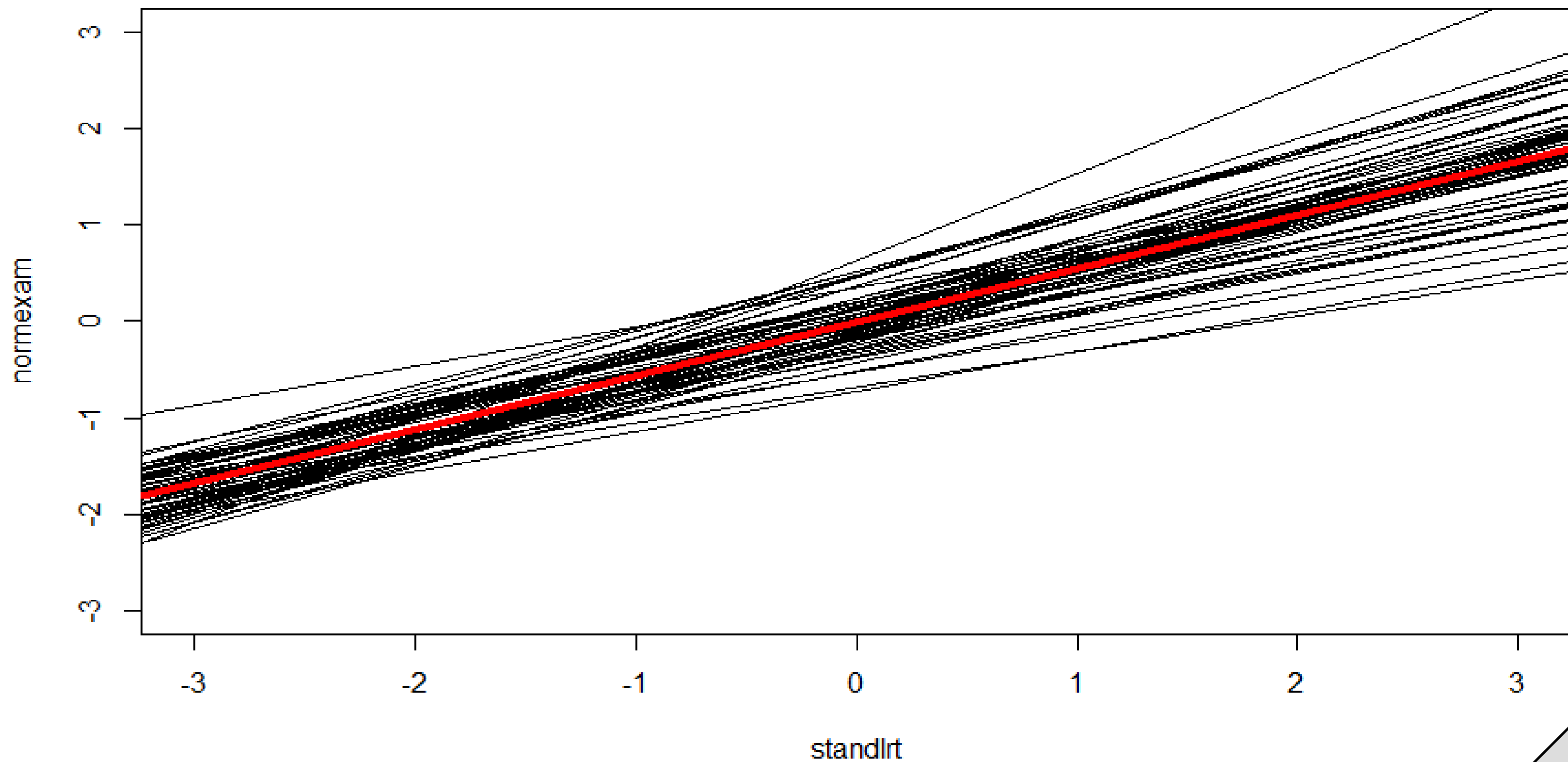
random intercept + random slope

- Interpreting the model:
  - Fixed intercept = -0.01: average exam score when stdLRT = 0 (so for a child with an average LRT score)
  - Fixed effect LRT = 0.56: for two children who differ by 1 SD in LRT score, the exam score will be (on average) 0.56 SD higher for the child with the higher LRT score
  - SD of random intercepts (0.30) and slopes (0.12) is much smaller than residual variance (0.74) - more variance within than between schools
  - Correlation intercept-slope (0.497) usually not interesting, but:
    - schools with higher mean exam score when stdLRT=0 (mean LRT) tend to have higher slope

London Schools: mixed model

random intercept + random slope

Fitted model



London Schools: comparing right & wrong models

Model	overall/fixed slope LRT	s.e.
1. aggregated data	0.884	0.116
2. disaggregated data	0.595	0.013
3. regr. per school	0.425	0.116
4. school*LRT interactions	??	??
5a. mixed model (random intercept) 5b. mixed model (random int + random slope LRT)	0.563 0.557	0.012 0.020

London Schools data

Aside: coding of categorical variables

- Gender: 0=boy, 1=girl
- Schavg (school average of intake score): 1=low, 2=mid, 3=high
- Schgend: 1= mixed school, 2=boys’ school, 3=girls’ school

London Schools:

adding a (fixed) child-level covariate

```
> sch.lme.3 <- lme(fixed=normexam~standlrt + factor(gender), random=~standlrt |
school, data=london, method="ML")
> summary(sch.lme.3)
Linear mixed-effects model fit by maximum likelihood
 Data: london
      AIC      BIC    logLik
  9301.358 9345.518 -4643.679

Random effects:
 Formula: ~standlrt | school
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 0.2936242 (Intr)
standlrt    0.1212575 0.533
Residual    0.7416710

Fixed effects: normexam ~ standlrt + factor(gender)
                     Value  Std.Error   DF   t-value p-value
(Intercept)     -0.1117670 0.04305229 3992 -2.596075  0.0095
standlrt         0.5529634 0.01998634 3992 27.667060  0.0000
factor(gender)1  0.1757988 0.03225659 3992  5.450011  0.0000
```

London Schools:

adding a child-level covariate

- On average, girls score 0.176 SD higher on exam than boys (holding stdLRT constant)

London Schools

adding (fixed) school-level covariates

```
> sch.lme.4 <- lme(normexam~standlrt + factor(gender) + factor(schgend) + factor(schav),
      random=~standlrt | school, data=london, method="ML")
> summary(sch.lme.4)

Random effects:
 Formula: ~standlrt | school
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 0.2660309 (Intr)
standlrt    0.1212542 0.499
Residual    0.7417279

Fixed effects: normexam ~ standlrt + factor(gender) + factor(schgend) +
factor(schav)
                      Value  Std.Error   DF   t-value p-value
(Intercept)      -0.2647657 0.08159434 3992 -3.244902  0.0012
standlrt          0.5515520 0.02006950 3992 27.482097  0.0000
factor(gender)1   0.1671313 0.03385088 3992  4.937282  0.0000
factor(schgend)2  0.1869684 0.09777600   60  1.912211  0.0606
factor(schgend)3  0.1570156 0.07780641   60  2.018029  0.0481
factor(schav)2    0.0668879 0.08534936   60  0.783696  0.4363
factor(schav)3    0.1742650 0.09876108   60  1.764511  0.0827
```

London Schools:

Adding child- and school-level covariates

Effect	estimate	se	p
Fixed Effects			
Intercept	-0.265	0.082	0.0012
norm. LRT	0.552	0.020	< 0.0005
girls (vs. boys)	0.167	0.034	< 0.0005
school avg: low	(ref)	0.100	
school avg: mid	0.067	0.085	0.436
school avg: high	0.174	0.099	0.083
school gender: mixed	(ref)		
school gender: boys	0.187	0.098	0.061
school gender: girls	0.157	0.078	0.048
(Co)variance Parameters:			
school intercept	0.266^2		
school slope	0.121^2		
corr int-slope	0.499		
residual variance	0.742^2		

London Schools: conclusions (so far)

- The reading score is a significant predictor of exam score
    - for every 1 SD higher on reading score, average increase of 0.552 SD on exam score
- Boys do significantly worse than girls on exam
    - boys score, on average, 0.167 SD lower on exam than girls
- School “level” (average exam score) does not appear to be predictive of exam score
- School gender may be predictive
    - average exam score at girls’ schools is 0.157 SD higher than at mixed schools
    - average exam score at boys’ schools is 0.174 SD higher than at mixed schools
- Note: these conclusions are based on the “Wald” p-values and are not necessarily to be trusted!

London Schools: conclusions (so far)

- Because the LRT score has been centered, the estimate for the intercept (-0.265) is the estimated average (normalized) exam score for:
  - a boy (ref) with
  - avg LRT score from
  - a school with low average score (ref) and
  - mixed school (ref)
- The residual variance is 0.550, much larger than the variances for the random intercept (0.071) and random slope (0.015), indicating more variation within schools than between.
- Adding child- and school-level covariates explains some of the variance between schools (variance intercepts 0.09 → 0.07)

London Schools: still to do

- We've made model assumptions, need to check them!
  - distribution of residuals
  - distribution of random effects (?)
- How to choose among models?
- How to answer subquestion (does gender of school have influence on effect of gender of pupil?)

Multilevel modelling, summary

- Account for correlation of measurements at different levels
  - children within schools, measurements within patients
- Allow us to include variables measured at different levels
  - child's gender, school's achievement or SES level
- We can model variation at different levels
  - more variation within than between schools
- Longitudinal data is a specific example of multi-level data
  - lecture 2: mixed models for longitudinal data
- How to build models, check assumptions?
  - lecture 3: technical issues in multilevel/longitudinal modelling
- Outcomes don't have to be continuous
  - lecture 4: models for binomial and Poisson data