**Notes on lab 1**

**Multilevel data**

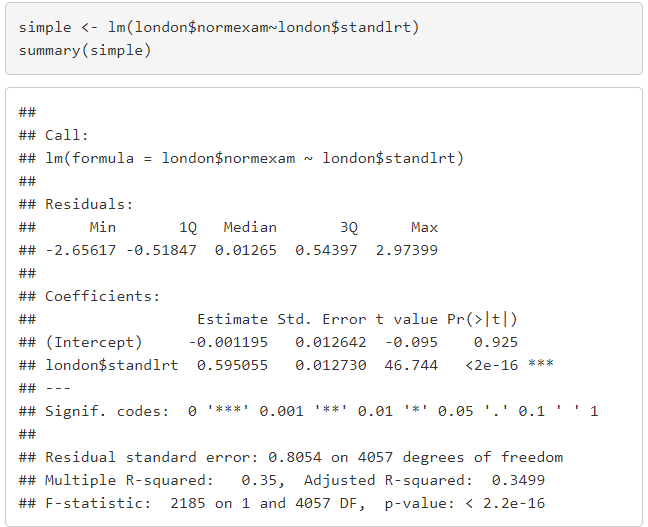
* Hierarchical structure so:
  + Children within schools
  + Patients within centers

With variation at all levels, correlation within levels

Variables can be measures at each level e.g.

Type of school = Level 2

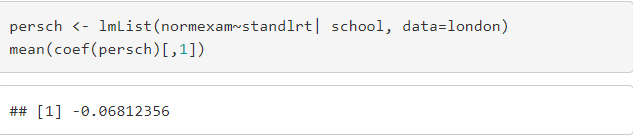
Score on LRT per child = Level 1

**Simple linear regression model with all schools together**

**Estimate for slope**: The normalized exam score is going up 0.595055 standard deviation for every one unit more of the standardized reading test with se 0.013. **Estimated intercept:** -0.001

**Disadvantages:**

* Inflates sample size
* SE of level 2 variables are underestimated -> p-value too small & CO too narrow
* SE of level 1 variables are over-/underestimated
* Always **assumes uncorrelated residuals** which is oftentimes unrealistic e.g. in this case one would assume no correlation between kids within the schools

**Linear regression per school**

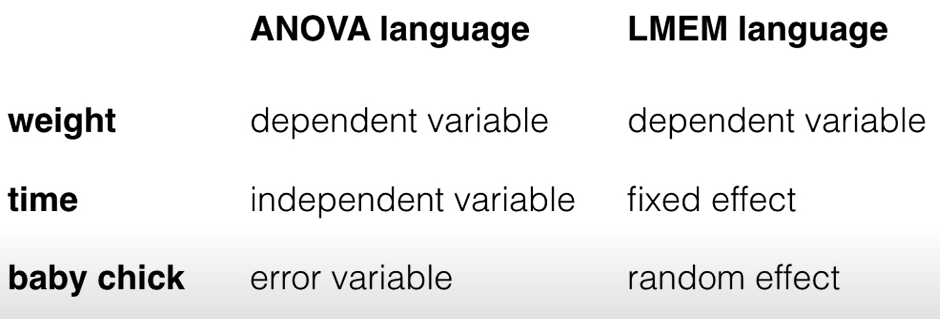
Taking the mean of the intercepts = -0.06812356 with

a standard deviation of the intercepts of 0.519.

Same for the slopes being averaged over all schools with a mean of 0.0435 and a sd of 0.939

**Disadvantages**

* As we do a linear regression for every school and simply took the average slope we are assuming that every school has an equal weight but the sized differed significantly
* We cannot analyze school-level variables as we averaged it (only child level)
  + No variation at the school level for each lm as it’s only analyzing one school

**Mixed Models**

**Fixed effect**= variable of interest

* Overall slope of London Reading Test to predict exam performance
* Gender
* Type of school,…

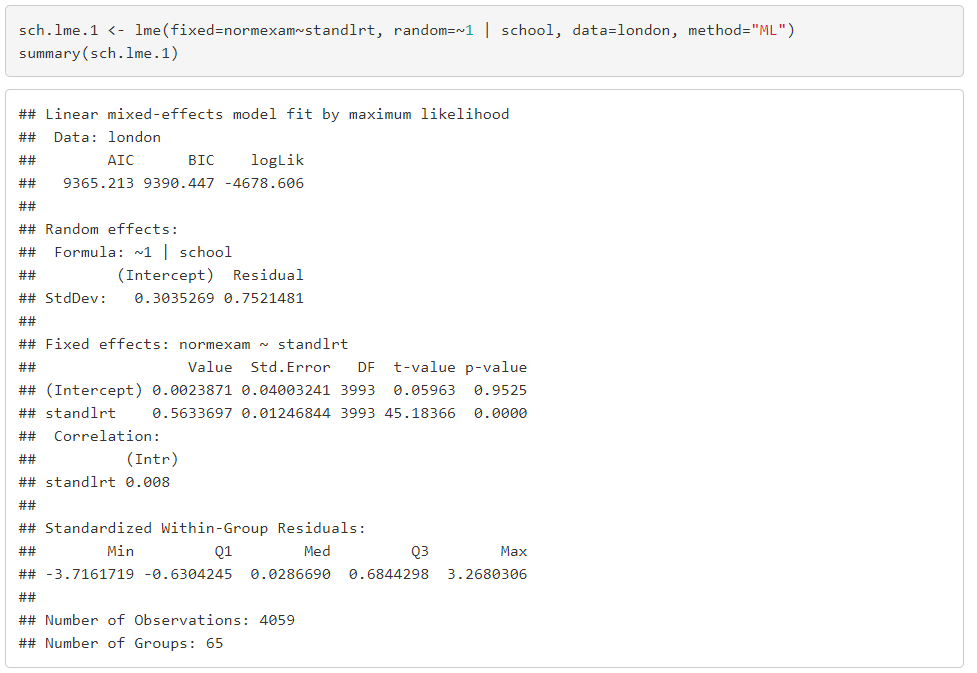
**Random effect**

* Allow for more than one source of variability in the data
* E.g. in a linear model the random variability could come from the individual students in a school where we use the LRT to predict their scores
* in mixed effects models we allow more sources of random variability in the data
* e.g. not just students but also variability from different schools

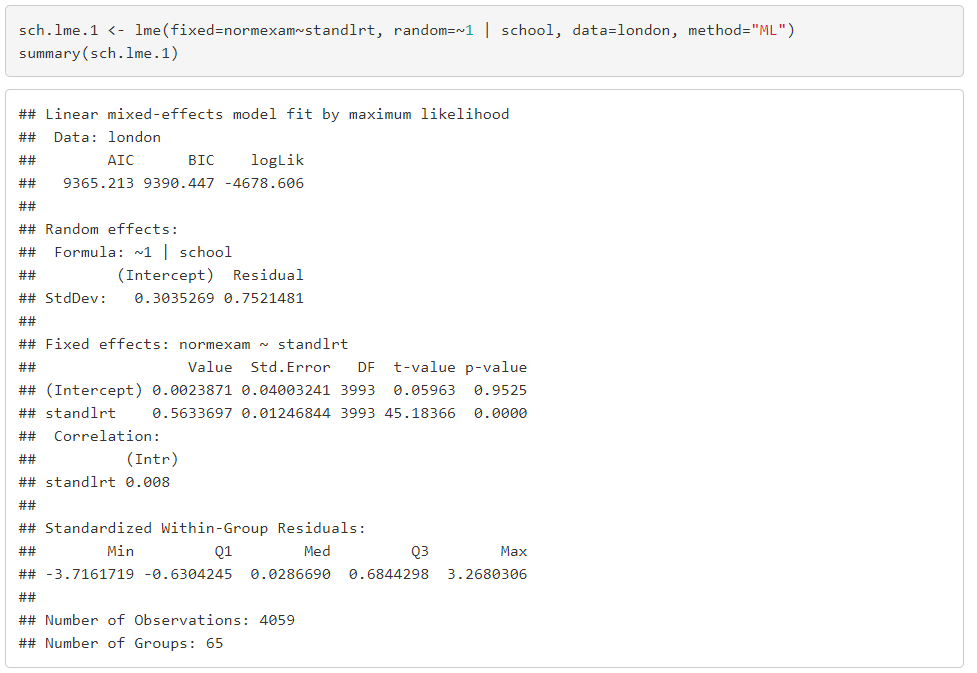
**random effects intercepts and slopes**:

* random intercepts allow the outcome to be higher or lower for each school
* random slopes allow fixed effects to vary for each school

By adding random effects we can incorporate the differences between schools and therefore make better assumptions/predictions.

**Linear mixed model with random intercept to predict exam scores using the LRT scores**

Standard deviation of the random effects (random intercept for every school and random=~1 means fixed forslope?) for the intercept and the residuals. So square to get variance= 0.3035^2=**0.092**; res var 0.75212^2=**0.565**

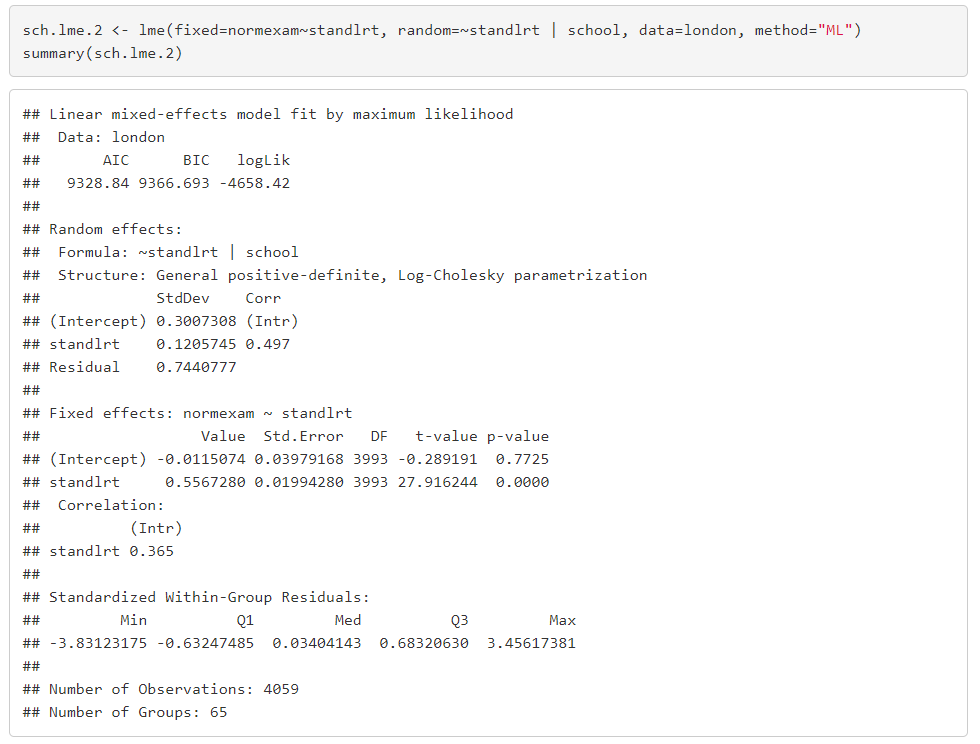


Estimated fixed effect of the standardized london reading test on exam scores while taking the number of kids into account.

**Estimate Fixed intercept** = 0.0024

**Estimate fixed slope** = 0.563

For every unit increase in the London reading test the exam score increases by 0.563 sd (sd because the reading test is normalized)

**Linear mixed model with random intercept & random slope**

Random = ~standlrt – random effect for slope

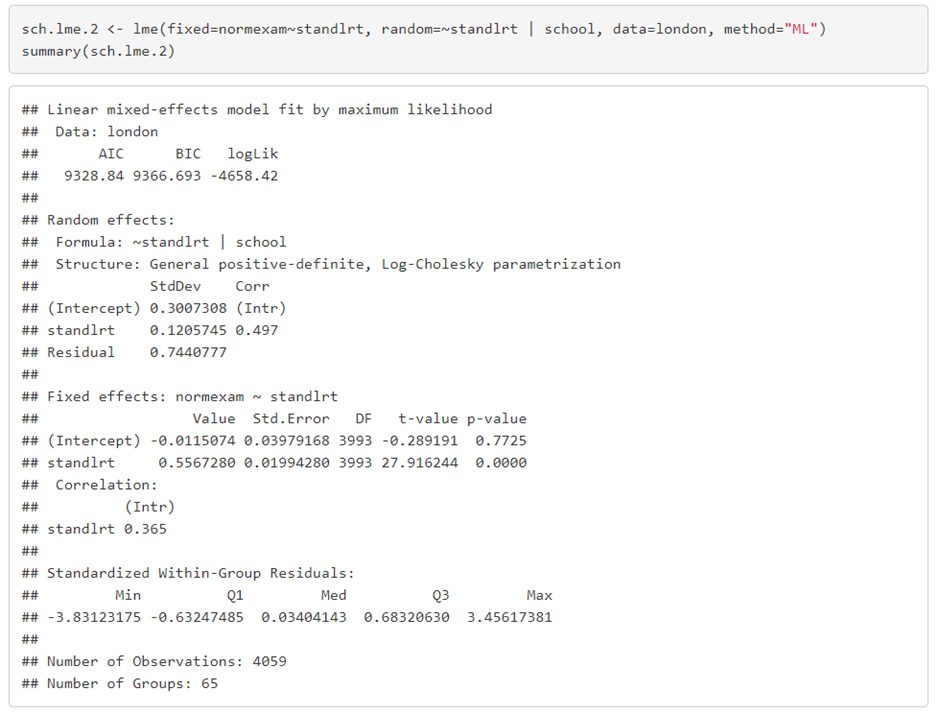
| school – random effect for intecept

**Estimated sd of intercept = 0.30**

**Estimated sd of slope = 0.12**

Both much smaller than **residual sd = 0.744** means more variance within than between schools

**Residual variance > variance of random intercept = more variation within than between**

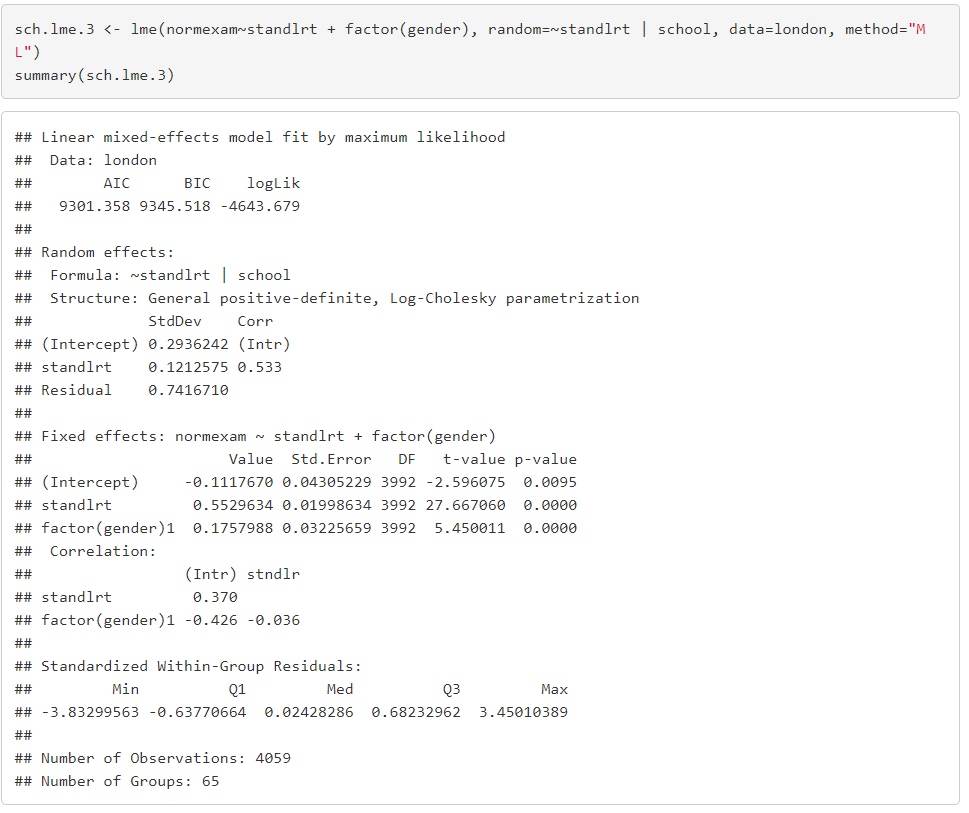
**estimate for fixed effect of average London reading test on overall test and its standard error**

**fixed effect LRT = 0.56**

the exam score will be 0.56 higher for each 1 SD in LRT score

**fixed intercept = -0.01 (=average exam score when stand. LRT =0)**

**Linear mixed model with random intercept & random slope + a fixed child-level covariate**

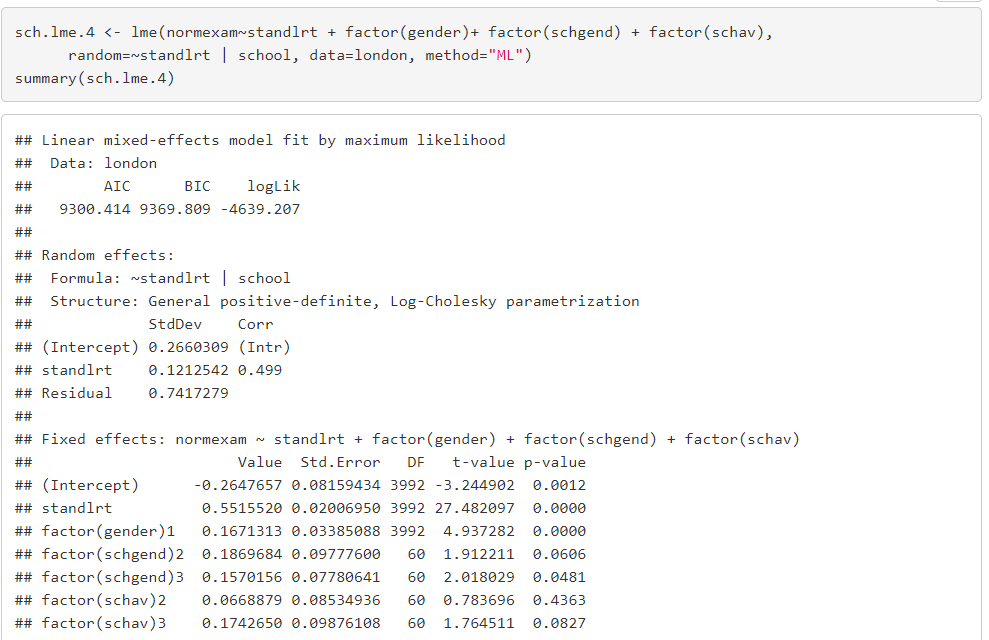
gender 0= boy (reference group) and 1= girl

**girls are 0.17 better than boys**

**Linear mixed model with random intercept & random slope + fixed child level covariate + fixed school-level covariates**

1= mixed schools

2= boys schools

3= girls schools

Comparison of all girls or all boys schools to mixed schools. The only significant p-value was for school type 3 = all girls schools, showing that **girls are scoring 0.157 better on average than mixed schools**

LRT has been normalized

* **the intercept (-0.265) is the estimated average exam score**
* **residual variance is 0.550**

Adding child-level covariates and school-level covariates explains some of the variance between schools (variance intercepts were lowered from 0.09 to 0.07)