

Assignment 1 Multilevel Modeling

Jovan Kalyango, Lela Roos, & Lina Kramer

Assignment 1 Multilevel Modeling

1

Multilevel analysis (modeling) is used for the statistical analysis of data with a hierarchical or clustered structure. For instance, educational data on students can be nested and include variables on a singular (student) level and a higher level (e.g., different schools). Usually data points within the same cluster (e.g., school) are more similar than data across different clusters. As a result, the assumption of independence may be violated, which standard statistical tests are not robust against. The degree of similarity within clusters is indicated by the intraclass correlation ρ .

Specifically, the data set we have at hand contains students that are nested in schools. Two individual-level variables are included for each student: Reading ability (as measured per the London Reading Test, LRT) and exam scores. As a cluster level variable, the data set includes the average reading score of students for each school. Because of the hierarchical structure of the data with two measurement levels, we can use multilevel regression analysis instead of traditional multiple regression analysis.

If we were to ignore the structure of the data and disaggregate the cluster level variable, we could perform a traditional regression analysis. However, there would likely be violations of the independence assumption. This is due to students from the same school likely having more similar scores to other students within that school than to students from different schools. This means exam scores and reading ability may not be independent of school. Violating the independence assumption may involve underestimation of standard errors and inflated type I error rates.

If we were to be interested in differences between schools, we could aggregate the individual level scores to the second- level (school-level). However, this would substantially decrease the effective sample size, hence lowering the statistical power (Hox et al., 2017).

Performing multilevel analysis instead allows a proper assessment of between school effects without disregarding the individual level information. We can assess the effect of individual LRT-scores on the outcome variable (exam scores) within schools, while the average LRT-scores can be used to asses their effects on exam scores between schools.

In summary, based on the available information from the data set, we think it is appropriate to use multilevel analysis in this cases.

2

Descriptive Statistics

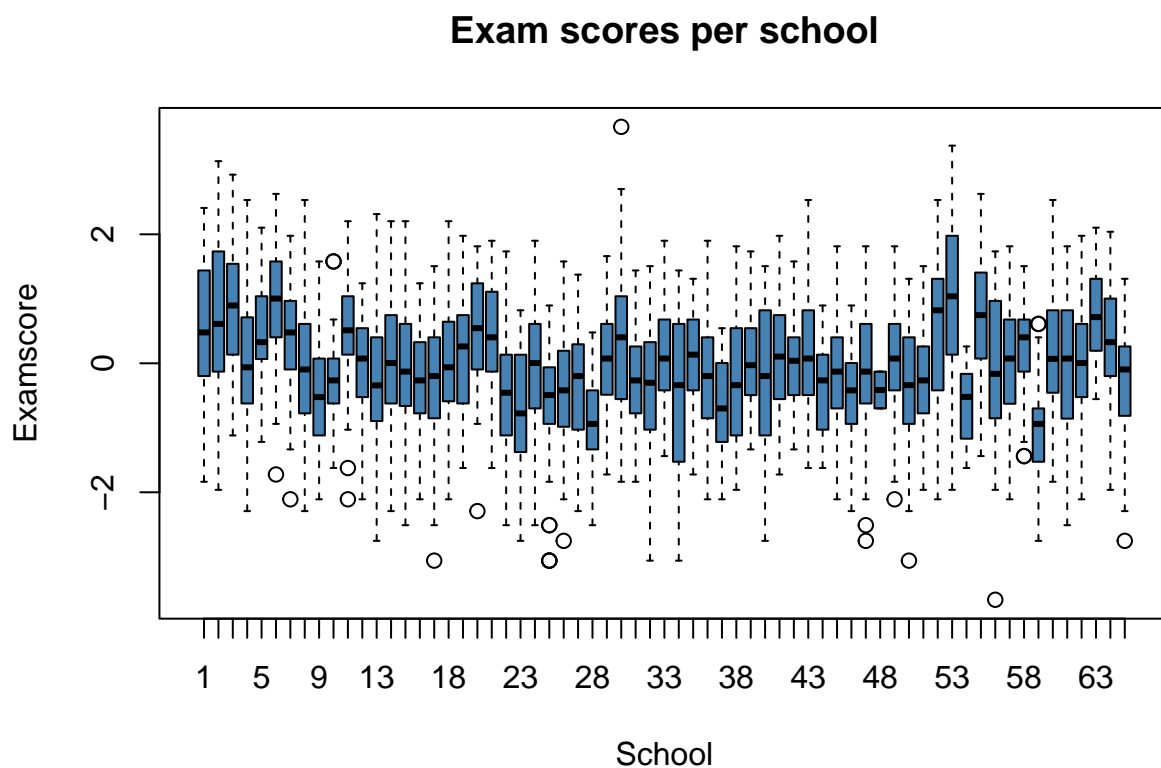
```
##      School      Student      Examscore      LRTscore
##  Min.    : 1.00   Min.      : 1.0   Min.      :-3.666100   Min.      :-2.93500
##  1st Qu.:14.00   1st Qu.: 16.0   1st Qu.: -0.699510   1st Qu.: -0.62071
##  Median :29.00   Median : 33.0   Median : 0.004322   Median : 0.04050
##  Mean   :31.01   Mean     : 38.7   Mean     :-0.000118   Mean     : 0.00181
##  3rd Qu.:47.00   3rd Qu.: 54.0   3rd Qu.: 0.678760   3rd Qu.: 0.61906
##  Max.    :65.00   Max.      :198.0   Max.      : 3.666100   Max.      : 3.01600
##
##      AvsLRT
##  Min.      :-0.755960
##  1st Qu.: -0.149340
##  Median : -0.020198
##  Mean     : 0.001811
##  3rd Qu.: 0.210520
##  Max.      : 0.637660
```

Our data set includes 4059 students from 65 different schools. We can view some descriptive statistics with the *summary()* function. Here, we see that the lowest amount of

students coming from one school is 1, and the highest amount is 198. We can also see that the variables 'Examscore,' 'LRTscore,' and 'AvsLRT' seem to indicate deviations from the average. Since the mean and medians are approximately zero for all variables, and it is possible to have a negative score. We also see that 'Examscore' ranges from -3.66 to 3.66, 'LRTscore' ranges from -2.94 to 3.02, and 'AvsLRT' ranges from -0.76 to 0.64.

Outlier checking

Observations considered as potential outliers by the *IQR* criterion are displayed as points in the boxplot. Based on this criterion, there are 17 potential outliers (2 above and 15 points below) the vertical line.



3

a)

Yes, from the the boxplot above we have an indication that scores are different per school. Therefore, we use multilevel regression analysis to accommodate these nested data and analyze the effect of school on a student's exam score. In the following steps we investigate how much variation in exam scores is explained on the school (cluster) level.

b)

H0: $u_{0j} = 0$ vs

HA: $u_{0j} > 0$

Or by means of deviance

H0: Deviance of the null/baseline model is equal to the deviance of intercept model;

$$Deviance_{m0} = Deviance_{m1}$$

Vs

H1: Deviance of the null/baseline model is larger than the deviance of intercept model;

$$Deviance_{m0} > Deviance_{m1}$$

c)

Level-1 equation.

$$examscore_{ij} = b_{0j} + e_{ij}$$

Level-2 equation

$$b_{0j} = \gamma_{00} + u_{0j}$$

Mixed Level equation

$$examscore_{ij} = \gamma_{00} + u_{0j} + e_{ij}$$

where i represents individual(student) and j represents school.

d)

The first model is the fixed-intercept model (model_0). We ignore the multilevel structure and only include an intercept.

The second model is the random intercept model (model_1). We add a variance component for the intercept at school level. This means exam scores are predicted an intercept and corresponding error terms at the cluster specific (school) and individual level.

An ANOVA comparing model_0 and model_1 revealed improved goodness-of-fit statistics for model_1 (*Deviance* = 11011, *AIC* = 11017) compared to model_0 (*Deviance* = 11509, *AIC* = 11513). A Chi-square difference test indicated significantly better model fit of model_1 for the data (

$$\chi^2_{(1)} = 498.72, p < 0.0001$$

). It follows that we reject the null hypothesis - the variance of the intercept between schools is not zero.

e)

The intraclass correlation (ICC) is 0.17, meaning that 17% of the variance is explained on the second level (school level). This also implies that if we randomly sampled two students from the same school, the expected correlation between their exam scores is 0.17.

f)

We have determined that there are differences in the observations between schools, with an ICC of 0.17. We think this value is high enough to continue with performing multilevel analysis, this will allow us to obtain both the explained variance separately at the both the individual level as well as the cluster level.

4

The two potential predictors of exam score are LRT-score on the student level and average LRT-score on the school-level. We first fit a model with only the student level predictor. The fixed effect model_2 includes the students level predictor LRT-score (for now without random slopes).

An ANOVA comparing model_2 to model_1 revealed improved goodness-of-fit for model_2 ($Deviance(4) = 9357.2, p < 0.001$; $AIC = 9365.2$). A Chi-square difference test

$$\chi^2_1 = 1653.4, p < 0.001$$

and Wald test ($LRT - score_{coef} = 0.56, p < 0.001$) confirmed that there is a significant fixed effect of LRT-score of exam score. Therefore, on average, a unit increase in the LRT-test leads to a 0.56 point increase in exam score. The first level $Pseudo - R^2 = 0.33$, this means that for students within the same school, 32% of variance in exam scores is explained by the model. For the second level $Pseudo - R^2 = 0.45$, meaning 45% of the variance in exam scores across schools is explained by the model.

In the next step, we add the second predictor variable, that is the average LRT scores per school, to our model. Our model_3 now includes a fixed effect of average LRT-score on the second level.

An ANOVA comparing model_3 to model_2 revealed improved goodness-of-fit indicators for model_3 ($Deviance(5) = 9347.6, p < 0.01$; $AIC = 9357.6$). A Chi-square

difference test ($\chi_1^2 = 9.623, p < 0.01$) and Wald test ($LRT - score_{coef} = 0.56, p < 0.001$; $AvsLRT_{coef} = 0.36, p < 0.01$) confirm that there is a significant fixed effect of average LRT-score on exam score. On average, a unit increase in the average LRT-test leads to a 0.56 point increase in exam score. 33% of the variance in exam scores is explained on the first (student) level ($Pseudo - R^2 = 0.33$) and 55% of variance is explained on the second (school) level ($Pseudo - R^2 = 0.55$). The value of the $Pseudo - R^2$ on the first level does not change in comparison to model_2 because we only added a level-2 predictor to the model.

So we conclude that both LRT-score on the student level and average LRT-score on the school-level are predictors of a students' exam score.

5

a)

$$H_0: u_{1j} = 0$$

vs

$$H_A: u_{1j} > 0$$

b)

To evaluate if the relation between LRT scores and exam scores is the same in all schools, we add random slopes to the model (model_4).

Allowing the slope of LRT to vary between schools results in improved goodness-of-fit indicators ($Deviance(7) = 9308.56, p < 0.001$; $AIC = 9324.40$). The Chi-square difference test confirms that the variance of the slope of LRT score between schools is significantly different from zero ($\chi_2^2 = 37.191, p < 0.001$).

c)

We conclude that the effect of LRT score on exam scores varies between schools. Therefore, we keep the random slope in the model.

6

a)

LRT score confidence interval at 0.95

$[0.51, 0.59]$

b)

LRT-score prediction interval at 0.95

$[0.31, 0.79]$

c)

By repeating the study infinitely many times, 95% of the constructed confidence intervals would contain the true slope of LRT-score. The predictive interval indicates that 95% of the regression coefficients of future LRT-score are predicted to lie in this interval. The difference between the two intervals is thus that the predictive interval refers to the distribution of the predicted slope coefficients within this particular model, whereas the confidence interval makes an assertion about a hypothetical sampling distribution of the slopes in the population. It also worth noting that the predictive interval is expected to be wider than the confidence interval.

7

To evaluate if average LRT scores explain a part of the different relations between LRT scores and exam scores in different schools, we fit a full multilevel regression model (model_5). In this model, we include the cross-level interaction term between the predictors LRT-score and average LRT-score. This model resulted in $Deviance(8) = 9302.9, p < 0.01$ and $AIC = 9318.9$. We did not center predictors, because their scales include zeros. This means the main effects remain meaningful and interpretable after adding the cross-level interaction.

The Chi-square difference test between model_4 and model_5 ($\chi^2_1 = 7.52, p < 0.01$) confirms a significant interaction between LRT scores and average LRT scores ($b = 0.16$). In conclusion, the average reading score explains part of the differences in the relationship between reading scores and exam scores in different schools. The higher the average reading, the stronger the positive relationship between the reading score and exam score. The Pseudo $R^2 = 0.23$ of the cross-level interaction indicates that 23% of the variance in this interaction is explained by the model.

8

We chose model_5 as the final model. This is because it only includes significant parameters and has the best model fit ($Deviance(8) = 9302.9$). Furthermore, model_5 has the lowest AIC (9318.9). Hence, it is the most appropriate model to fit the data set.

Level-1 equation.

$$examscore_{ij} = b_{0j} + b_{1j} * LRTscore_j + e_{ij}$$

Level-2 equation.

$$b_{0j} = \gamma_{00} + \gamma_{01} * AvgLRT_j + u_{0j}$$

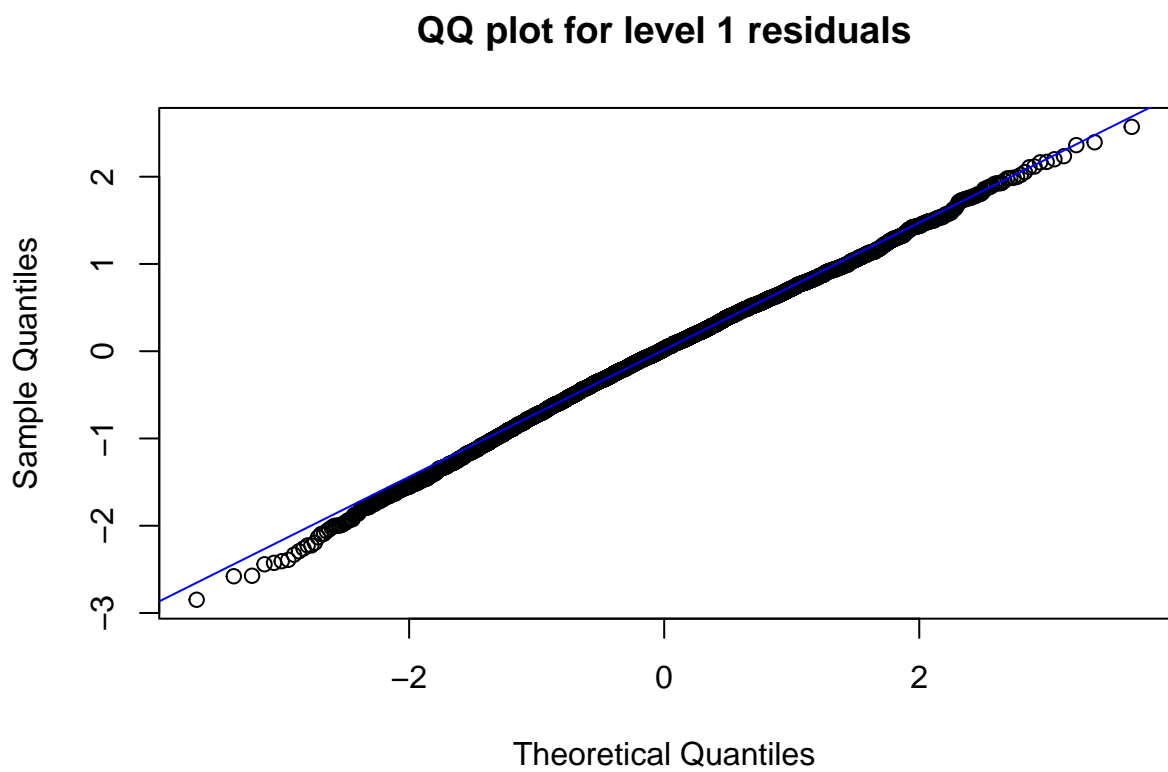
$$b_{1j} = \gamma_{10} + \gamma_{11} * AvgLRT_j + u_{1j}$$

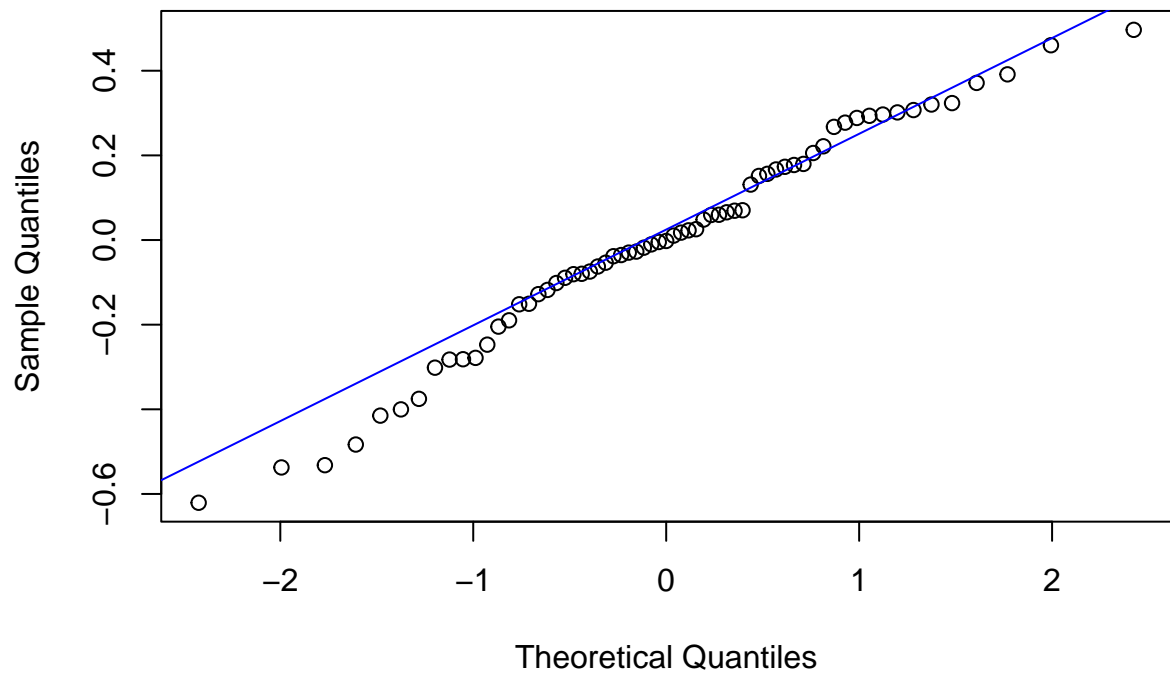
Mixed Level equation

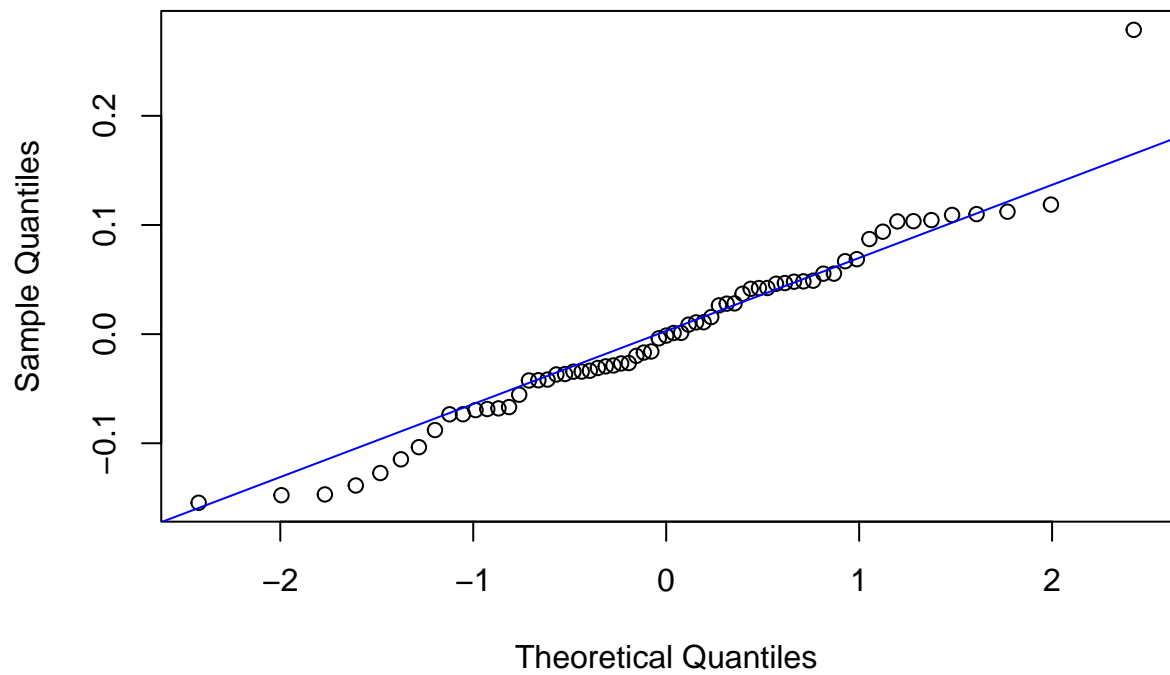
$$examscore_{ij} = \gamma_{00} + \gamma_{01} * AvgLRT_j + \gamma_{10} * LRTscore_j + \gamma_{11} * LRTscore_j * AvgLRT_j + u_{1j} * LRTscore_j + u_{0j} + e_{ij}$$

where i represents individual(student) and j represents school

9



QQ plot for random effect of intercept

QQ plot for random effect of LRTscore

The above plots shows no strong indications of a deviation from the normality assumption. So we can trust our previously made conclusions.

References

Hox, J. (1998). Multilevel Modeling: When and Why. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, Data Analysis, and Data Highways* (pp. 147–154). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-72087-1_17

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications* (3rd ed.). Routledge. <https://doi.org/10.4324/9781315650982>