TMA4315: Compulsory exercise 3: (Generalized) Linear Mixed Models

Group XX: Henrik Syversveen Lie, Mikal Stapnes, Oliver Byhring

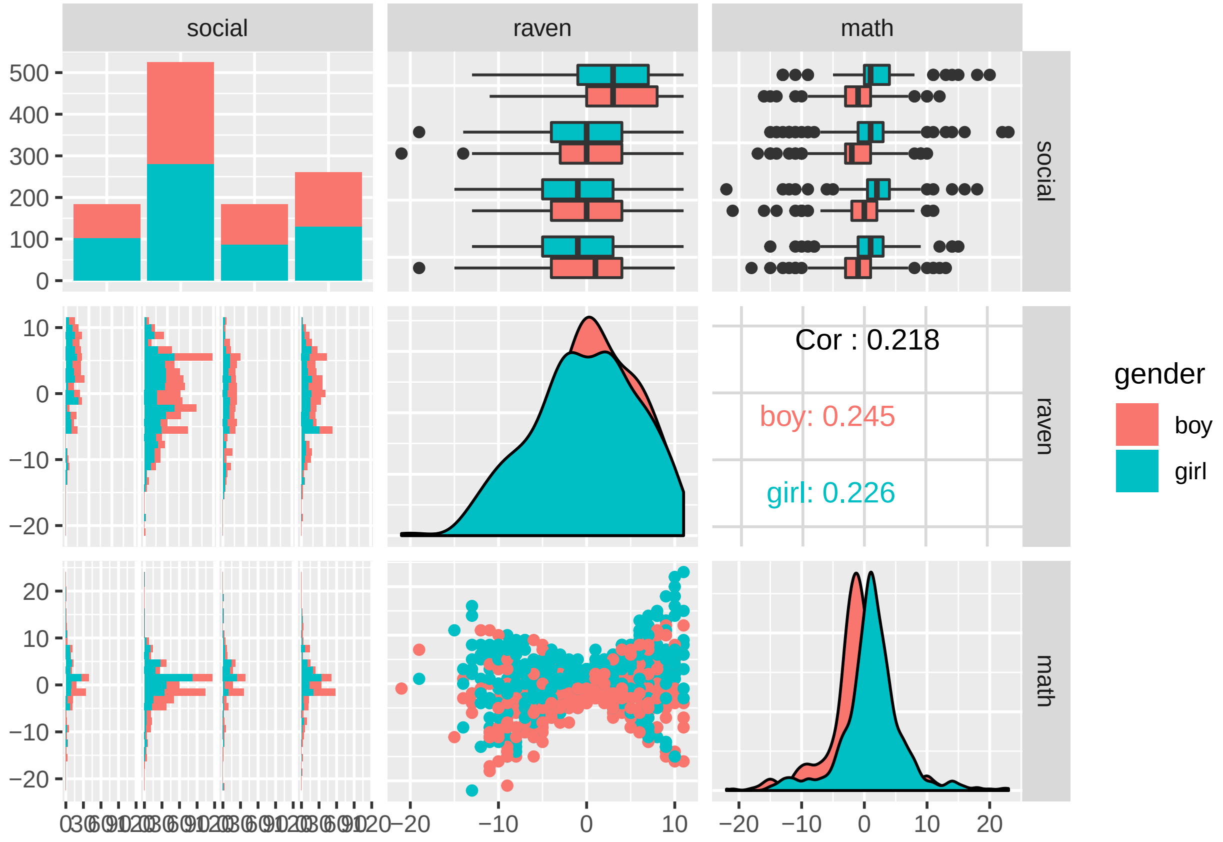
22.11.2018

Contents

a)

```
## [1] -1.256318
```

```
## [1] 1.188333
```



- Comment briefly on the plot you have created

First, we see that there is a positive correlation between the `raven` (test score) and `math` variable. This is as expected. Furthermore, we see that girls perform somewhat better in the math test than boys. Also, there is no evident correlation between social class and test scores, which is somewhat surprising.

```
##
## Call:
## lm(formula = math ~ raven + gender, data = dataset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.6704  -1.8791   0.1166   2.1166  19.6134
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.3131     0.2024  -6.488 1.29e-10 ***
## raven         0.1965     0.0240   8.188 6.98e-16 ***
## gendergirl    2.5381     0.2807   9.041  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.76 on 1151 degrees of freedom
## Multiple R-squared:  0.1105, Adjusted R-squared:  0.109
## F-statistic:  71.5 on 2 and 1151 DF,  p-value: < 2.2e-16
```

We fit a linear model with `math` as response, and `raven` and `gender` as covariates. Model for the kth student:

$$Y_k = \mathbf{x}_k\boldsymbol{\beta} + \epsilon_k$$

- Explain what the different parts of this model are called.

Y_k is called the response or random component, and will be the `math` score of student k. We assume the response to be normal distributed. The term $\mathbf{x}_k\boldsymbol{\beta}$ is the systematic component, or linear predictor. The final term ϵ_k is called the random error component, and is normal distributed with mean 0 and variance σ^2.

- Comment briefly on the parameter estimates you have found.

All parameter estimates are significant on a 0.001 level. We observe that the coefficient $\beta_{raven} = 0.1965$ and $\beta_{girl} = 2.5381$, which means that if $x_{raven} \to x_{raven} + 1$ our model would predict an increase in the math score of 0.1965. Similarly for `girl`, our model will predict a `math` score that is 2.5381 higher for a girl than for a boy assuming the remaining covariates are equal.

- What are we investigating with this model?

With this model we assume a linear relationship between the respone $Y =$ `math` and the covariates `raven` and `gender` and a normal distribution of the residuals,

$$Y_k = x_k^T\beta + \epsilon_k, \quad \epsilon_k \sim N(0, \sigma^2)$$

Under this assumption we investigate the significance and strength of our parameters β_{raven} and β_{girl}. We found both that the parameters are significant and that they are relatively strong.

b)

However, this model assumes that the distribution of `raven` and `gender` is independent of other covariates, which is not necessarily true. If there has been some gender distribution among the good and bad schools, this `school` effect will affect the parameter estimation β_{gender} and we will not be able to distinguish what should be attributed to the gender and what to the schools.

We therefore want to include in our model a random intercept $\gamma_{0,school}$ that seeks to remove the `school` factor as a contributing effect of the other parameters estimates.

we fit a new model

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{1}\gamma_{0i} + \epsilon_i.$$

- Explain what the different parts of this model are called and what dimensions the model components have.

The vector $\mathbf{Y}_i$ is still the response or random component and has dimension $n_i \times 1$. The term systematic component or linear predictor is now $\mathbf{X}_i\beta + \mathbf{1}\gamma_{0i}$. $\mathbf{X}_i$ is a $n_i \times p$ design matrix, β is a $p \times 1$ vector of fixed coefficients and $\mathbf{1}$ is a $n_i \times 1$ vector (design matrix for random effects) with every element equal 1. γ_{0i} is the random intercept, which is a normal distributed scalar with mean 0 and variance τ_0^2. Finally, the term ϵ_i is the random error component, which is a $n_i \times 1$ normal distributed vector with mean 0 and variance $I\sigma^2$.

- Write down distributional assumptions for γ_{0i} and ϵ_i.

We assume that each ϵ_i is independent, and that each of its elements are independent. This gives $\epsilon_i \sim N_{n_i}(0, I\sigma^2)$. Also, we assume γ_{0i} i.i.d. with $\gamma_{0i} \sim N_1(0, \tau_0^2)$.

- What do we assume about the dependency between the responses at school i, $\mathbf{Y}_i$, and school k, $\mathbf{Y}_k$?

We assume that responses between different schools i and k are conditionally independent. This means that if we construct a new global model for all clusters, with

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix},$$

then each $\mathbf{Y}_{ij}$ will be conditionally independent, or $\mathbf{Y}|\boldsymbol{\gamma} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma}, \boldsymbol{I}\sigma^2)$.

Now we fit the model in `R`:

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: math ~ raven + gender + (1 | school)
##    Data: dataset
##
## REML criterion at convergence: 6772.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4607 -0.4305 -0.0127  0.4083  4.2761
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  school   (Intercept)  3.879   1.969
##  Residual             19.220   4.384
## Number of obs: 1154, groups:  school, 49
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) -1.26915    0.34375  -3.692
## raven        0.21442    0.02331   9.197
## gendergirl   2.51119    0.26684   9.411
##
## Correlation of Fixed Effects:
##            (Intr) raven
## raven      -0.017
## gendergirl -0.404  0.034
```

- Compare the parameter estimates for `raven` and `gender` with the estimates from the linear model in a), and discuss.

We see that the parameters `gender` and `raven` (and the intercept) are larger for the random intercept model, and that they have smaller standard deviations. WHAT IS THERE TO DISCUSS??

- How do `gender` and `raven` score affect the math scores?

Again, we can say that for one female and one male student from the same school with equal `raven` score, the female student is expected to get 2.51119 better score on the math test. Also, if a student increases his/her `raven` score by one, we would expect an increase of 0.21442 to the math score.

- In the print-out from `summary(fitRI1)` there are no p-values. Why is this?

The distribution of the βs is assymptotically normal, but for finite sample sizes, we are not sure if the usual t-statistic is t-distributed. Also, if it is t-distributed, we do not know the degrees of freedom.

- Test the null-hypothesis $H_0: \ \beta_{\text{raven}} = 0$ against $H_1: \ \beta_{\text{raven}} \neq 0$ and provide a p-value for the test. (Yes, we have many observations and believe that we can calculate an asymptotic p-value even though the lmer package not by default want to report such a number.)

We assume we have enough observations to calculate an asymptotic p-value. We compute

```r
2 * (1 - pnorm(9.197))
```

```
## [1] 0
```

We observe that we get an asymptotic p-value of 0 for testing the hypothesis $H_0: \ \beta_{\text{raven}} = 0$ against $H_1: \ \beta_{\text{raven}} \neq 0$. We reject H_0 and conclude that $\beta_{\text{raven}} \neq 0$.

- Also provide a 95% confidence interval for the effect of the female gender on the math score.

We assume that we have enough observations for the β's to be normally distributed. We can then compute a 95% confidence interval for the effect of female gender to be

$$\beta_{\text{female}} \in [\hat{\beta}_{\text{female}} - z_{0.025}\text{sd}(\beta_{\text{female}}), \hat{\beta}_{\text{female}} + z_{0.025}\text{sd}(\beta_{\text{female}})].$$

And this gives

$$\beta_{\text{female}} \in [1.988, 3.034].$$

c)

We now continue with a random intercept model (school) with only raven as fixed effect (remove gender from our model).

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: math ~ raven + (1 | school)
##    Data: dataset
##
## REML criterion at convergence: 6856.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.2705 -0.4725 -0.0045  0.4603  4.4890
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  school   (Intercept)  4.002   2.001
##  Residual             20.711   4.551
## Number of obs: 1154, groups:  school, 49
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  0.03840    0.32071   0.120
## raven        0.20682    0.02418   8.554
##
## Correlation of Fixed Effects:
##       (Intr)
## raven -0.004
```

We rewrite our model as

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{1}\gamma_{0i} + \epsilon_i = \mathbf{X}_i\boldsymbol{\beta} + \epsilon_i^*.$$

We then write $\mathbf{V}_i = \text{Cov}(\epsilon_i^*) = \text{Cov}(\mathbf{1}\gamma_{0i}) + \text{Cov}(\epsilon_i) = \mathbf{1}\tau_0^2\mathbf{1}^T + \mathbf{I}\sigma^2$. This means that the matrix $\mathbf{V}_i$ will be on the form

$$\mathbf{V}_i = \begin{pmatrix} \tau_0^2+\sigma^2 & \tau_0^2 & \dots & \tau_0^2 \\ \tau_0^2 & \tau_0^2+\sigma^2 & \dots & \tau_0^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau_0^2 & \tau_0^2 & \dots & \tau_0^2+\sigma^2 \end{pmatrix},$$

an $n_i \times n_i$ matrix with tau_0^2 on the off-diagonal and $\tau_0^2+\sigma^2$ on the diagonal. This gives

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{V}_i).$$

In conclusion, we can say that the covariance between the responses Y_{ij} and Y_{il} from school i is τ_0^2, and the correlation is $\tau_0^2/(\tau_0^2+\sigma^2)$.

- What is this correlation for our fitted model `fitRI2`? Comment.

The `R` printout gives $\sigma^2 = 20.711$ and $\tau_0^2 = 4.002$, which gives a correlation of $4.002/(20.711+4.002) = 0.1619$. We observe that the correlation always has to be positive, because we only have square terms in the formula for the correlation. WHAT MORE IS THERE TO COMMENT??

- Write down the mathematical formula for $\hat{\gamma}_{0i}$ for your random intercept model and explain what the different elements in the formula means.

According to the module pages, it can be shown that after some calculations we get

$$\hat{\gamma}_{0i} = \hat{\mathbf{Q}}\mathbf{U}_i^T\hat{\mathbf{V}}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\beta}) = \dots = \frac{n_i\hat{\tau}_0^2}{\hat{\sigma}^2 + n_i\hat{\tau}_0^2}e_i.$$

Here, n_i is the number of observations in cluster i. $\hat{\tau}_0^2$ is the estimated variance of each γ_{0i}. $\hat{\sigma}^2$ is the estimated variance of ϵ_i. Finally, e_i is the average (raw, level 0) residual given by

$$e_i = \frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij} - \mathbf{x}_{ij}^T\hat{\boldsymbol{\beta}}).$$

- Explain what each of the six plots produced and displayed below can be used for (that is, why are we asking you to make these plots).

QQ-plot of random intercept:

Plots random effect quantiles on the y-axis vs. standard normal quantiles on the x-axis. Can be used to check normality assumption of the random effect.

Random intercept(RI)

Essentially the same plot as the last one, with random effect quantiles on the x-axis and schools on the y-axis. Displays all the random intercepts quantiles, and can be used to check normality assumption of random intercept.

Density of RI

Plots the density function of the random intercept vs. a theoretical normal distribution to check if the distribution of the random intercept replicates a normal distribution.

Residuals vs Fitted values

Plots the residuals vs. fitted values of the regression to check if the residuals are i.i.d. normal.

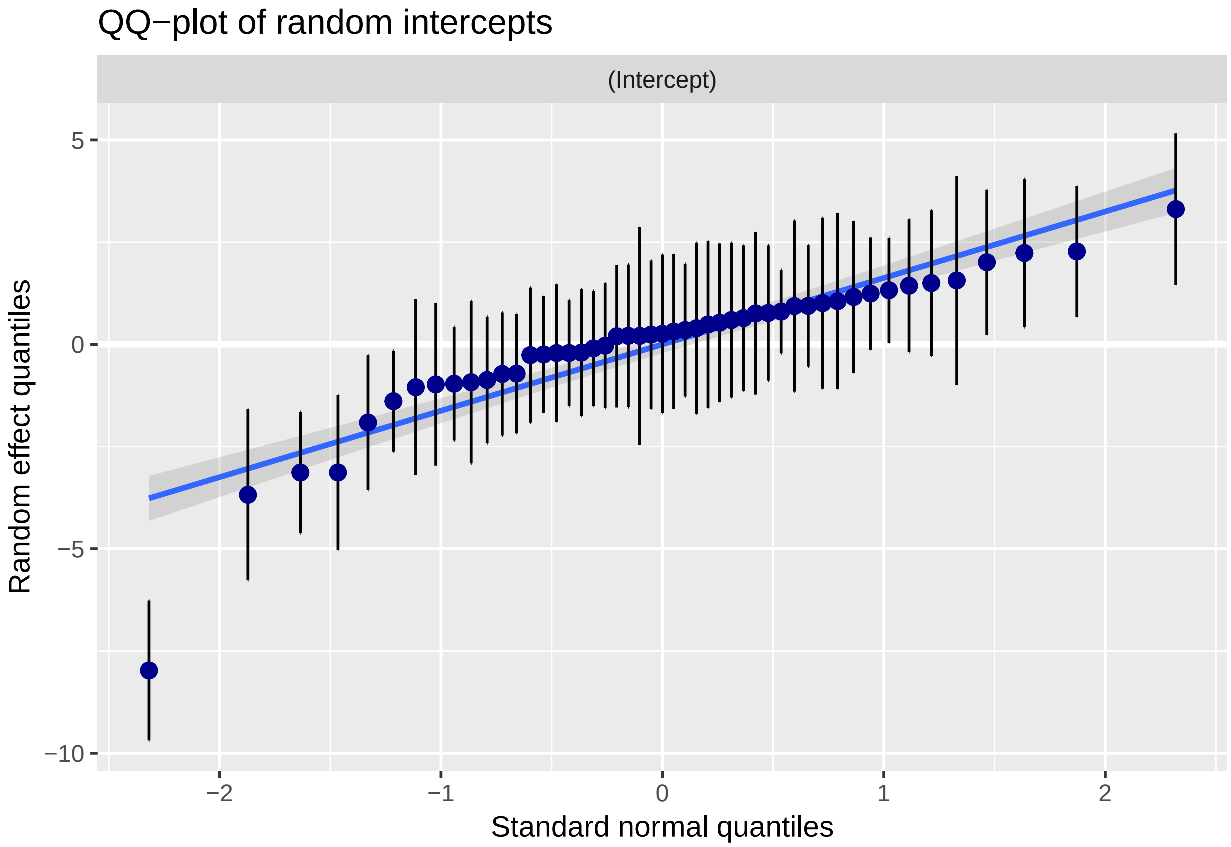
Nomral Q-Q

Plots standardized residuals of the regression vs. theoretical quantiles. Can be used to check normality assumption of the response Y.

Last plot

Plots the math score as a function of raven score for all the different schools. Can be used for prediction of a math score for a given school and given raven score.

- Comment on your findings.



From the first plot it seems like the random intercept is indeed normal, with only the quantile of one school really deviating from the normal line. From the density plot, we see that the distribution function is somewhat "thinner" than a normal distribution, with smaller tails and higher density for small absolute values of the random intercept. Moreover, the residuals seem to be homoscedastic and uncorrelated. The normal Q-Q plot also supports the assumption of normality of the residuals. From the last plot we see that one school has significant lower math scores than the rest, but apart from that there are no evident violations of our model assumptions.

d)

- Compare the model with and without the social status of the father using hypothesis test from the `anova` below (which is a likelihood ratio test - no, you need not look at the column called deviance since we have not talked about that). Which of the two models do you prefer?

```
fitRI3 <- lmer(math ~ raven + social + (1 | school), data = dataset,
    reml = T)
anova(fitRI2, fitRI3)
```

```
## Data: dataset
## Models:
## fitRI2: math ~ raven + (1 | school)
## fitRI3: math ~ raven + social + (1 | school)
##        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## fitRI2  4 6858.9 6879.1 -3425.4   6850.9
## fitRI3  7 6856.8 6892.1 -3421.4   6842.8 8.1175      3    0.04364 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cat("P(chi^2 > 8.1175) = ", 1 - pchisq(8.1175, 3))
```

```
## P(chi^2 > 8.1175) =  0.04364474
```

From the `anova` printout we get a reduction in `deviance` of 8.1175 when increasing the 3 additional parameters. Under model assumptions, the reduction in deviance will be χ^2-distributed with degrees of freedom $7 - 4 = 3$, which gives a p-value of 0.0436. As this is significant on a 0.05-level, we would prefer the larger model that includes the social status as a covariate.

- Why does the print-out say ???refitting model(s) with ML (instead of REML)??? (i.e. why do we not want REML when comparing models with the same random terms but with different fixed terms)?

While we do not observe this message when running ANOVA, we know that the REML estimation for the parameters (σ^2, τ_0^2) in $V = \sigma^2 I + UGU^T$ are found by maximizing the likelihood for $A^T Y$, where A is any $N \times (N - p)$ full-rank matrix with columns orthogonal to the columns of the design matrix X. However, as p differs between the two models, the transformations A_2 and A_3 will differ and the transformed observations $A_2^T Y$ and $A_3^T T$ will not be the same. As the observations differ, the likelihood of the respective saturated models will also differ and we cannot use the Likelihood Ratio Test. Instead we must use the Maximum-Likelihood (ML) estimation to conduct the Likelihood Ratio Test and use Restriced Maximum Likelihood (REML) for parameter estimation.

- Also comment on the AIC and BIC of the two models (automatically added in the print-out from `anova`).

Both AIC and BIC add a small penalty to the `deviance` for the use of additional parameters in the model. We would prefer the model with the least AIC/BIC, but we see here that the two measures do not agree on which model is better. However, as we here have nested models and can use the Likelihood Ratio Test, we prefer this over either measure and can disregard both the AIC and the BIC.

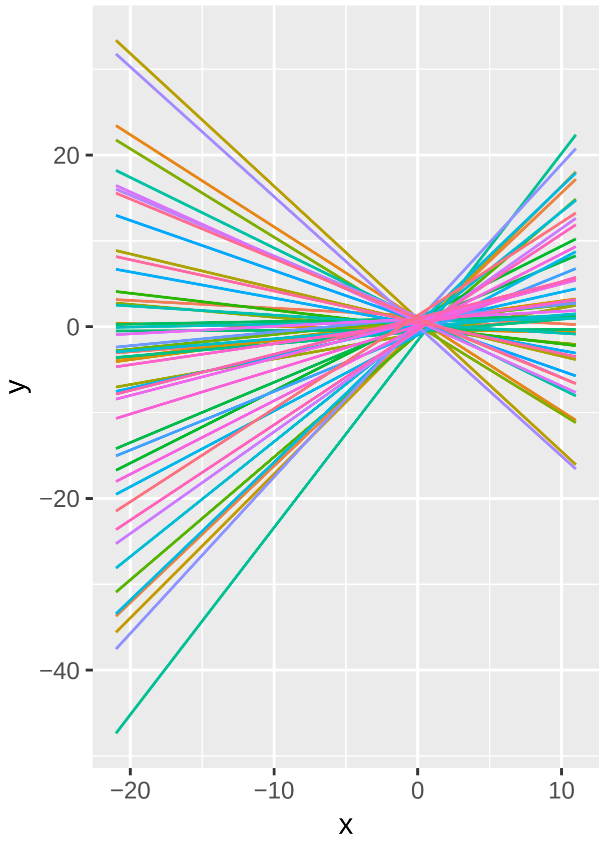
The last model we want to consider is a model with a random intercept an a random slopE for the `raven` score at each school.

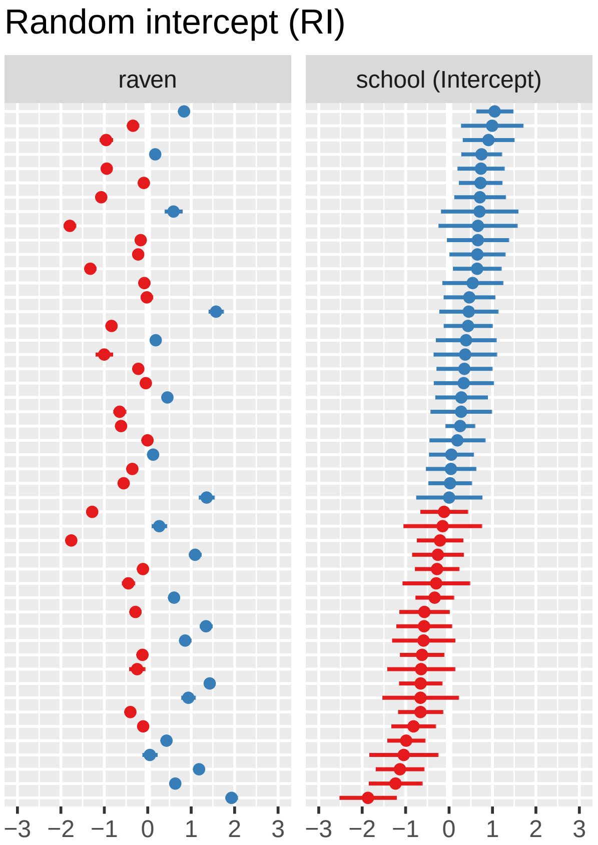
```
fitRIS <- lmer(math ~ raven + (1 + raven | school), data = dataset)
summary(fitRIS)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: math ~ raven + (1 + raven | school)
##    Data: dataset
##
```

```
## REML criterion at convergence: 4537.6
## 
## Scaled residuals: 
##      Min       1Q   Median       3Q      Max 
## -2.87462 -0.66206 -0.03913  0.65818  3.09716 
## 
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr 
##  school   (Intercept) 0.5519   0.7429        
##           raven       0.7293   0.8540   -0.40
##  Residual             2.2094   1.4864        
## Number of obs: 1154, groups:  school, 49
## 
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   0.2603     0.1183   2.200
## raven         0.2498     0.1223   2.042
## 
## Correlation of Fixed Effects:
##       (Intr)
## raven -0.356
```

```
df <- data.frame(x = rep(range(dataset$raven), each = 49), y = coef(fitRIS)$school[,
    1] + coef(fitRIS)$school[, 2] * rep(range(dataset$raven), each = 49),
    School = factor(rep(c(1:42, 44:50), times = 2)))
gg1 <- ggplot(df, aes(x = x, y = y, col = School)) + geom_line()
gg2 <- plot_model(fitRIS, type = "re", sort.est = "(Intercept)", y.offset = 0.4,
    dot.size = 1.5) + theme(axis.text.y = element_blank(), axis.ticks.y = element_blank()) +
    labs(title = "Random intercept (RI)")
ggarrange(gg1, gg2, ncol = 2, legend = FALSE)
```





- Write the mathematical formula for the random intercept and slope model and comment on what you see from fitting the model.

Using the model that includes both a random intercept and a random slope, we assume the model

$$Y_i = \beta_0 + \beta_1 raven + \gamma_{i0} + \gamma_{i1} raven + \epsilon_i, \epsilon_i \sim N_{n_i}(\mathbf{0}, \sigma^2 I_{n_i}), \gamma_i = (\gamma_{i0}, \gamma_{i1}) \sim N_2(\mathbf{0}, Q)$$

We now permit both a random effect for `raven`, γ_{i1} and a random intercept γ_{i0} for each school i. Observing the spread in the fitted lines $\hat{Y}_i(raven)$ we could support our claim that the inclusion of a random effect γ_{i1} is reasonable.

However, from the right plot we observe that the random effects tend to cancel each other out; where the random intercept is positive the random effects are mostly negative and vice versa. We observe from the summary of the model that the effects are negatively correlated, $Corr(\gamma_{i0}, \gamma_{i1}) = -0.40$. It is possible that the random effects “blow each other up”, and the inclusion of the random effect γ_{i1} may result in an overestimate of the fixed effect β_{raven} (0.2498 vs 0.2068).

SP??RRE OM DETTE HER ER RIMELIG

e)

- Why is it not suitable to use a linear mixed effects model?

We now wish our model to predict a probability in the range $[0, 1]$. It is no longer suitable to use a linear mixed effects model because assuming normality of our random effects is no longer reasonable. Firstly, the residuals ϵ_{ij} will then be in $[-1, 1]$ and thus bounded. Secondly, as good students largely pass but poor students might also pass it is reasonable to expect that the variance of the probability will be dependent on the fitted_value.

- What type of model would be more suitable? (hint: IL module 8)

To create a reasonable model we must instead consider som transformation$t(\cdot) : [0, 1] \to (-\infty, \infty)$. In this case, it will be suitable to consider a generalized linear mixed model with a binomial response. We then assume

$$\eta_{ij} \sim x_{ij}^T \beta + u_{ij}^T \gamma_i + \epsilon_{ij} \gamma_i \sim N(\mathbf{0}, Q) \epsilon_{ij} \sim N(0, \sigma^2)$$

and use the transformation

$$Y_{ij} = \frac{\exp \eta_{ij}}{1 + \exp \eta_{ij}},$$

as this is the canonical link of the binomial distribution.

- How would we add a random school intercept into this model (in which part of the model)?

To add a random school intercept into this model we simply let $\gamma_i = \gamma_{i0}$ denote each random school intercept and assume

$$\eta_{ij} \sim \beta_0 + \beta_1 raven + \gamma_i + \epsilon_{ij}, \gamma_i \sim N(0, \tau_0^2), \epsilon_{ij} \sim N(0, \sigma^2)$$

- What is the main challenge with this type of models? (hint: marginal model)

To do parameter estimation, we the consider the likelihood

$$L(\theta) = \prod_{i=1}^{m} f(y_{ij}|\theta) = \prod_{i=1}^{m} \int_{\gamma_i} f(y_{ij}|\gamma_i, \theta) f(\gamma_i) d\gamma_i,$$

and observe that, unfortunately, the expression $f(y_{ij}|\theta) = \int_{\gamma_i} f(y_{ij}|\gamma_i, \theta) f(\gamma_i) d\gamma_i$ only has a known distribution in special cases. Thus, to calculate the likelihood (and log-likelihood, deviance, BIC, AIC and other measures dependent on the likelihood) of a GLMM we must approximate this integral numerically.