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

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

18.11.2018

Contents

a) **1**

b) **3**

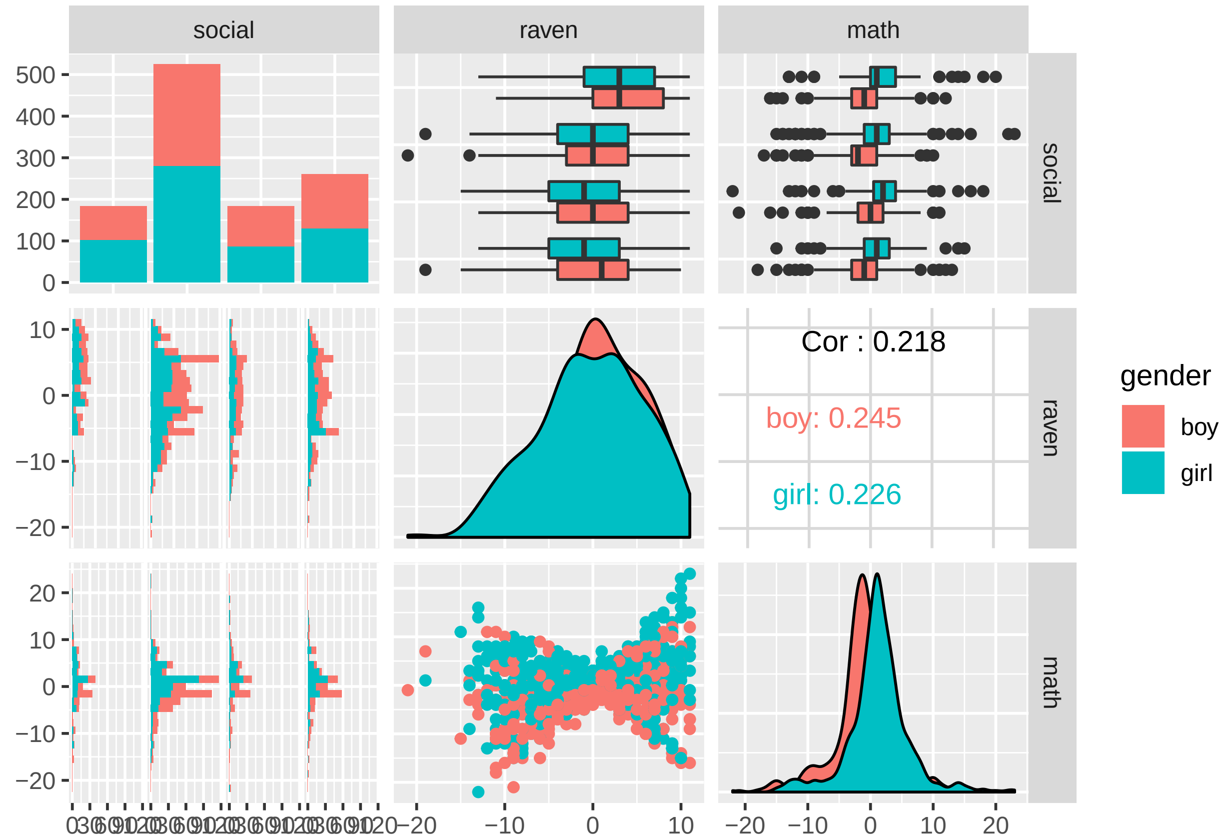
c) **3**

d) **3**

e) **3**

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
```

- Explain what the different parts of this model are called. ??? What is this supposed to mean
- 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

- For all

c)

- Write down the mathematical formula for the covariance and correlation between response Y_{ij} and Y_{il} from school i.
- Write down the mathematical formula for $\hat{\gamma}_{0i}$ for your random intercept model and explain what the different elements in the formula means.
- 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).
- Comment on your findings.

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?

e)