Bayesian Statistics Project

Edoardo Cortolezzis

2025-09-18

Problem C: Scores attained by students in Scotland

The R package mlmRev contains the ScotsSec dataset on scores attained by Scottish secondary school students on a standardized test taken at age 16. The data include 3435 observations on 6 variables. The help file description is as follows:

- verbal: The verbal reasoning score on a test taken by the students on entry to secondary school
- attain: The score attained on the standardized test taken at age 16
- primary: A factor indicating the primary school that the student attended
- sex: A factor with levels M and F
- social: The student's social class on a numeric scale from low to high social class
- second: A factor indicating the secondary school that the student attended

After performing some explorative analyses:

- Consider the binary variable attain01 which takes values 1 if attain is greater than 5 and 0 otherwise. Build a model for studying the effects of covariates on attain01 with rstan or rstanarm, taking into account the hierarchical structure of the data.
- Check the model fit and comment the results.
- Draw inference on school random effects. Does the primary school matter?
- [optional] Propose an alternative model for the variable attain (stan fit is not required).

Data loading and EDA

Let's start by loading the data, creating the **attain01** variable and transforming categorical variables into factors.

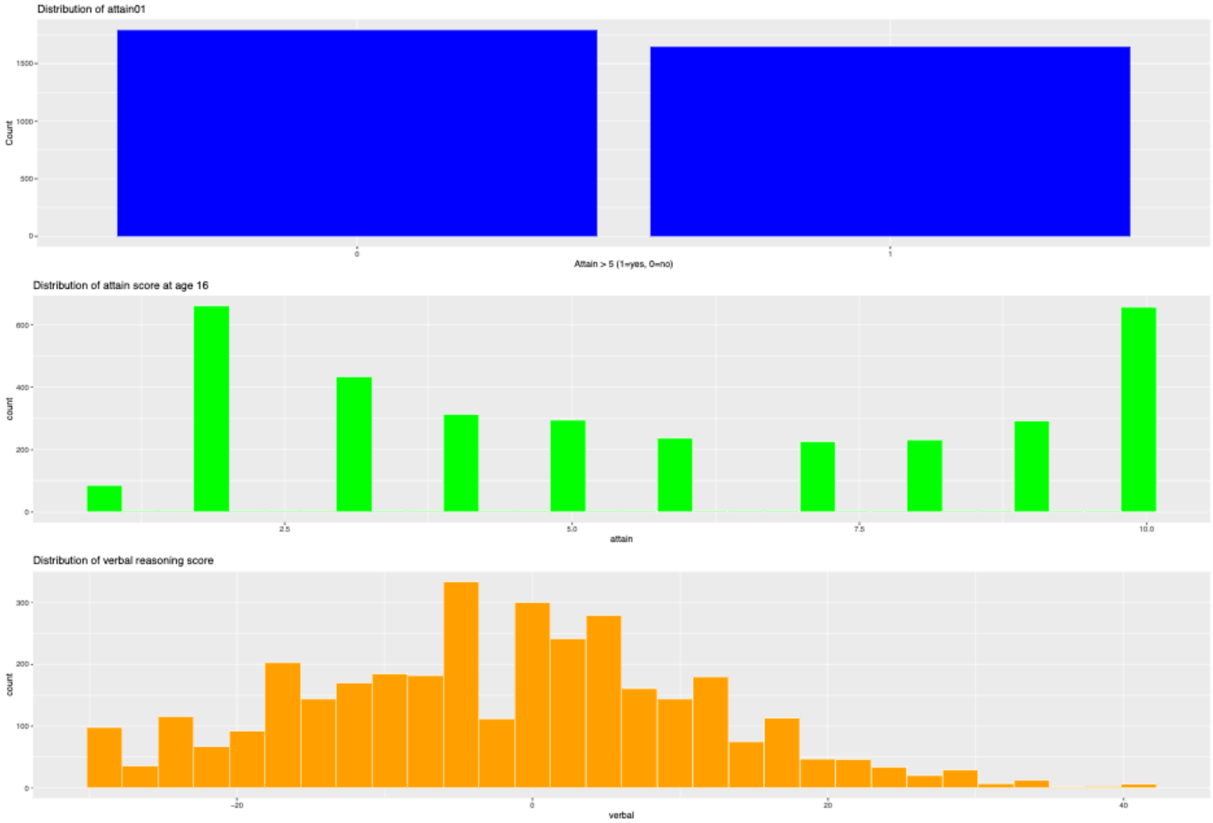
```r
data("ScotsSec", package = "mlmRev")
df <- ScotsSec %>%
  mutate(
    attain01 = ifelse(attain > 5, 1, 0),
    attain01 = factor(attain01),
    sex = factor(sex),
    primary = factor(primary),
    second = factor(second)
  )

# dataset structure summary
str(df)
```

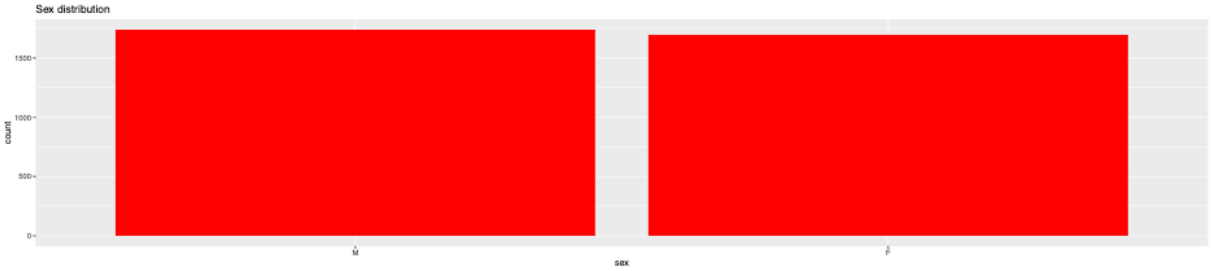
```
## 'data.frame':    3435 obs. of  7 variables:
##  $ verbal  : num  11 0 -14 -6 -30 -17 -17 -11 -9 -19 ...
##  $ attain  : num  10 3 2 3 2 2 4 6 4 2 ...
##  $ primary : Factor w/ 148 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ sex     : Factor w/ 2 levels "M","F": 1 2 1 1 2 2 2 1 1 1 ...
##  $ social  : num  0 0 0 20 0 0 0 0 0 0 ...
##  $ second  : Factor w/ 19 levels "1","2","3","4",..: 9 9 9 9 9 9 1 1 9 9 ...
##  $ attain01: Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 2 1 1 ...
```

Now that the data has been correctly loaded, we can proceed with some exploratory data analysis.

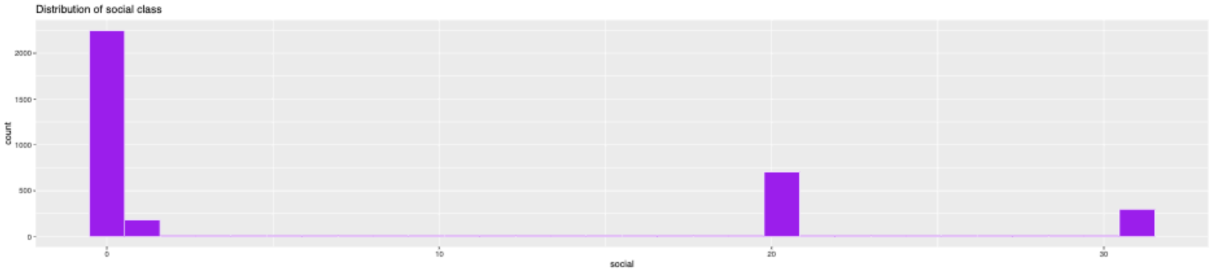
EDA: variables distributions, missing values and outliers

Distributions plots:

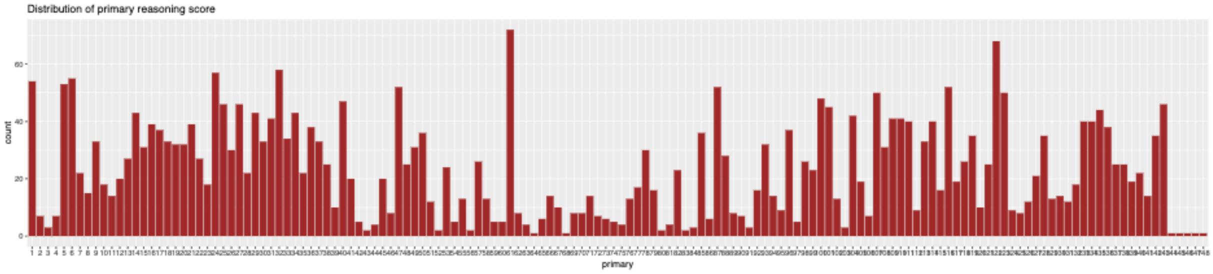




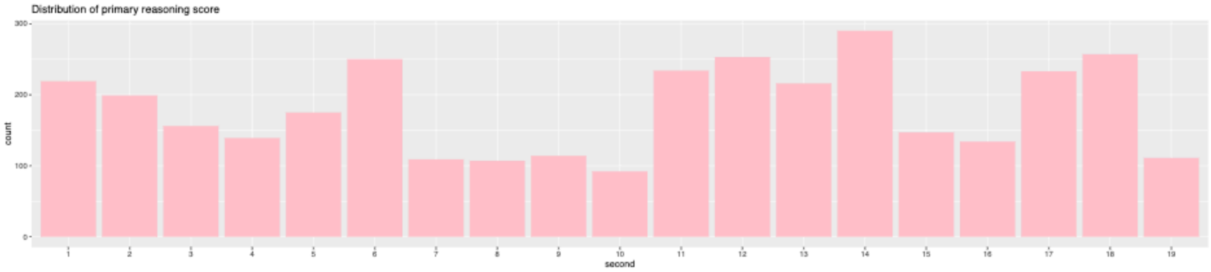
Sex distribution



Distribution of social class



Distribution of primary reasoning score



Distribution of primary reasoning score

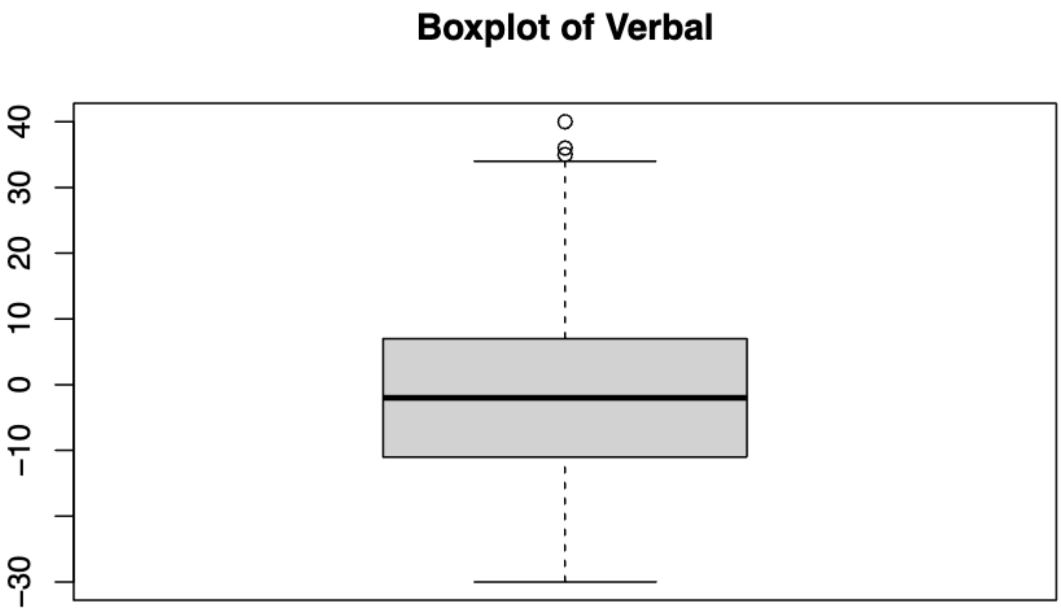
As we can see:

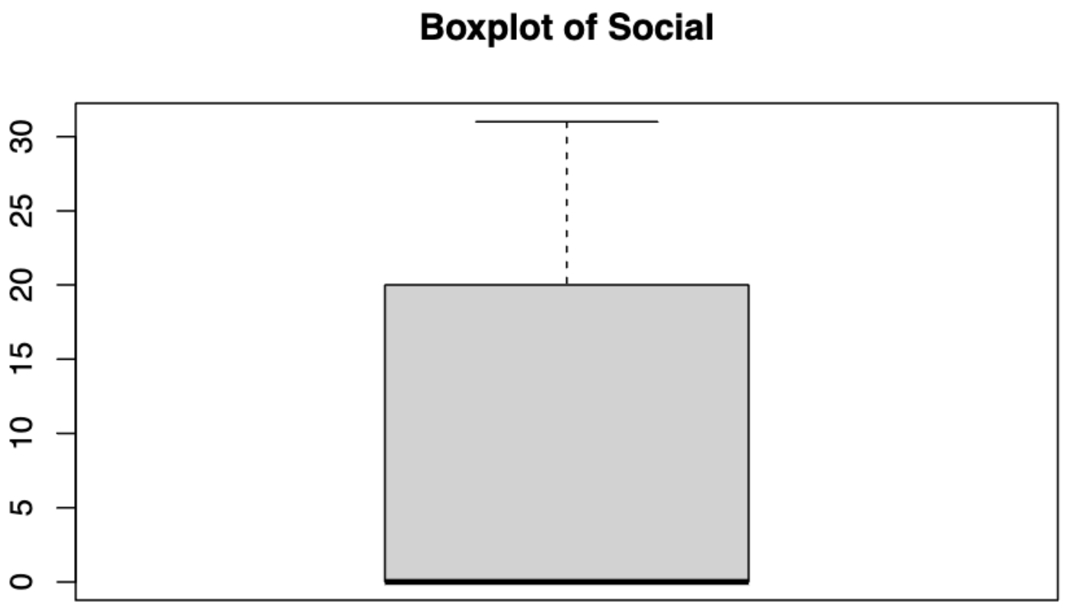
- `attain01` and `sex` seem quite balanced;
- `verbal` is roughly bell-shaped with some extreme values;
- `attain` has peaks at very low and very high scores;
- `social` is highly skewed, with most students in the lowest social class group;
- Primary schools are numerous and uneven in size, while secondary schools are fewer and more evenly distributed.

Missing Values:

```
## Total number of missing values in df: 0
```

Boxplots:


Boxplot of Verbal
40
30
20
10
0
−10
−30



There are some Outliers in `verbal` that must be dealt with.

```r
# Calculate quartiles and IQR
Q1 <- quantile(df$verbal, 0.25, na.rm = TRUE)
Q3 <- quantile(df$verbal, 0.75, na.rm = TRUE)
IQR_value <- IQR(df$verbal, na.rm = TRUE)

# Define bounds
lower_bound <- Q1 - 1.5 * IQR_value
upper_bound <- Q3 + 1.5 * IQR_value

# Identify outliers
outliers_verbal <- df[df$verbal < lower_bound | df$verbal > upper_bound, ]
outliers_verbal
```

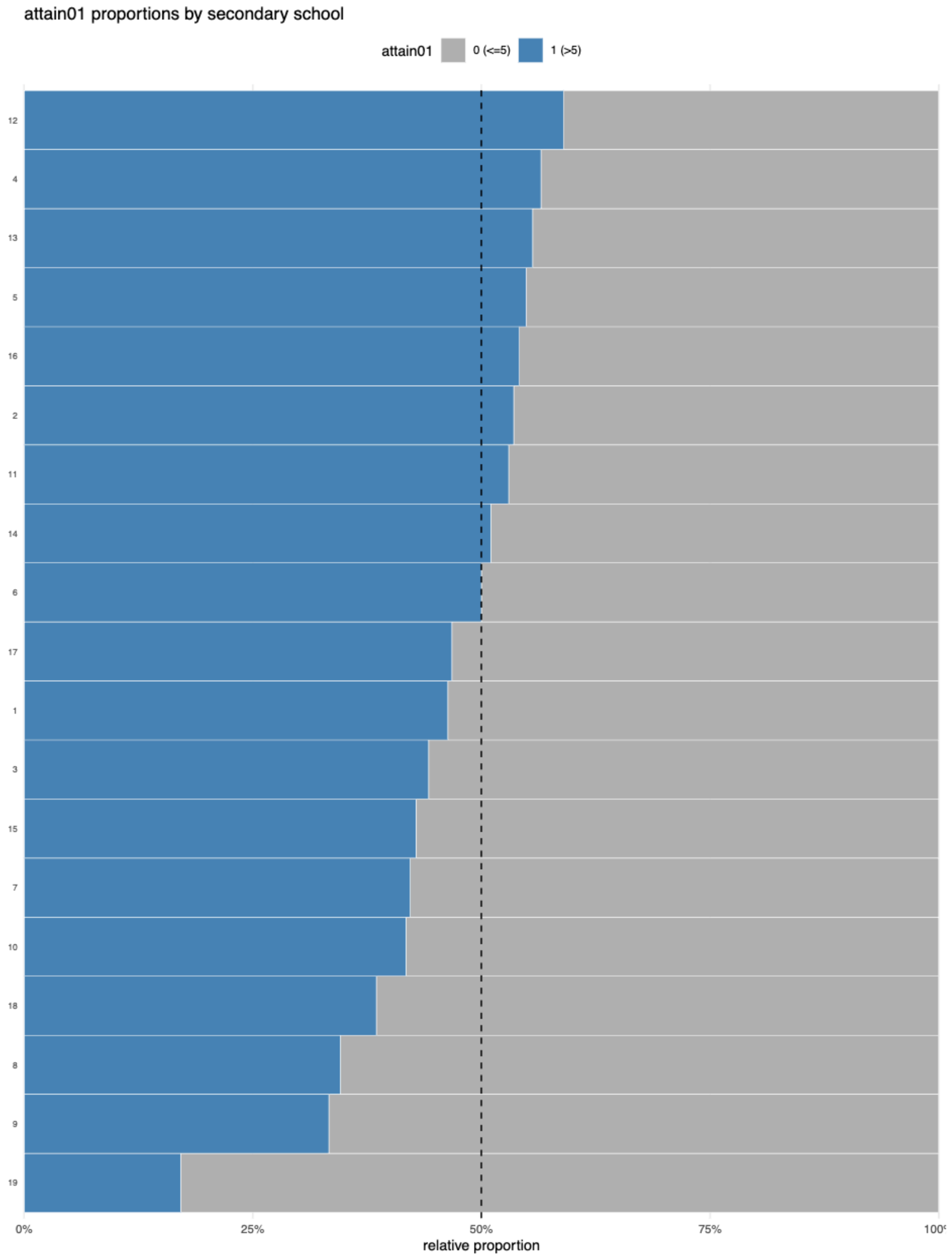
```
##      verbal attain primary sex social second attain01
## 451      35      9      17   M     31      4        1
## 621      40     10      22   F      0      5        1
## 1174     40     10      37   M      0      1        1
## 1355     36     10      47   M     31      2        1
## 1553     40     10      57   M      1     10        1
## 1896     40     10      85   F      1      5        1
## 1900     40     10      85   M      1     12        1
## 2052     36     10      93   F      1     12        1
## 2942     40     10     124   M     20     16        1
## 3356     35     10     142   F     31     13        1
## 3381      35       10       142    F        0      13          1
```

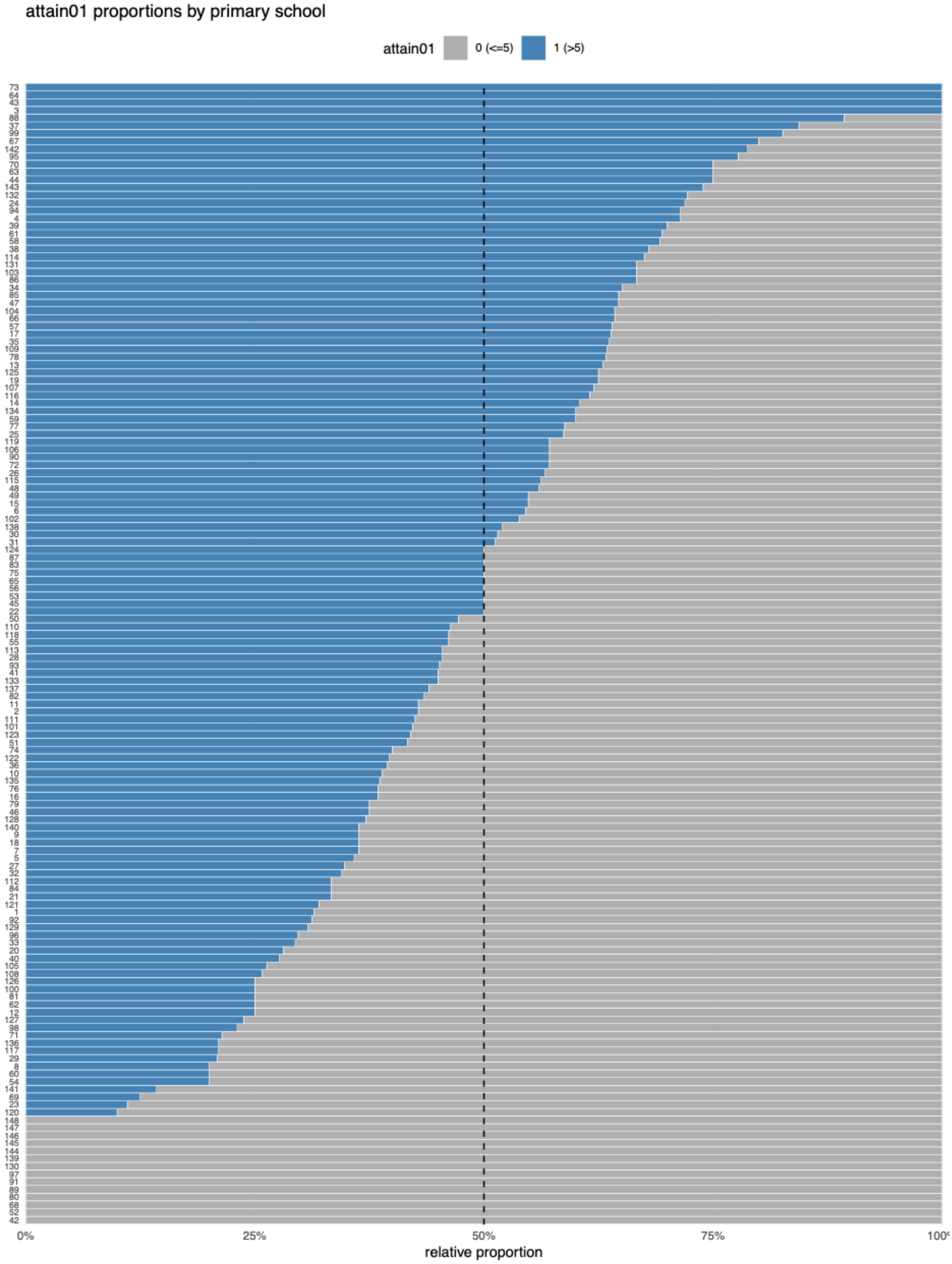
Since they are only 11 (out of 3k observations), we can simply remove them:

```
df <- df[df$verbal >= lower_bound & df$verbal <= upper_bound, ]
```

`attain01` distribution among different school groups

Before exploring the models, let's take a look at `attain01`'s distribution for the differnt groups of `primary` and `second`:


attain01 proportions by secondary school
attain01
0 (<=5)
1 (>5)
12
4
13
5
16
2
11
14
6
17
1
3
15
7
10
18
8
9
19
0%
25%
50%
75%
100
relative proportion



It looks like there are some major differences in `attain01`'s distribution among both primary and secondary school groups.

Therefore, it may be useful to explore some **hierarchical logistic regression modelss**.

Modeling

First thing first, let's investigate the fit of different models with increased complexity and various multilevel modeling approaches.

Let `attain01` be distributed as $y_i \sim \text{Bernoulli}(p_i)$, with $\text{logit}(p_i) = \eta_i$.

Individual level covariates: **verbal**, **sex** (baseline = M), **social**.

Schools: secondary $j(i)$, primary $k(i)$.

M0 (baseline): intercept-only

$$\eta_i = \alpha.$$

M1: random intercepts for **primary** and **secondary** (no covariates)

$$\eta_i = u_{0,\,j(i)}^{(\text{second})} + v_{0,\,k(i)}^{(\text{primary})},$$

with $u_j \sim \mathcal{N}(0, \sigma^2_{\text{second}})$, $v_k \sim \mathcal{N}(0, \sigma^2_{\text{primary}})$.

M2: fixed covariates only

$$\eta_i = \alpha + \beta_1 \text{verbal}_i + \beta_2 \cdot \text{sexF}_i \ + \beta_3 \text{social}_i.$$

M3: M2 + random intercepts for **primary**

$$\eta_i = \boldsymbol{\beta} \cdot \mathbf{x}_i + v_{0,\,k(i)}.$$

M4: M2 + random intercepts for **secondary**

$$\eta_i = \boldsymbol{\beta} \cdot \mathbf{x}_i + u_{0,\,j(i)}.$$

M5: M2 + (random intercepts **and** random slopes for *verbal* and *social*) by **primary**

$$\eta_i = +\boldsymbol{\beta} \cdot \mathbf{x}_i + \big(v_{0,\,k(i)} + v_{1,\,k(i)}\,\text{verbal}_i + v_{2,\,k(i)}\,\text{social}_i\big),$$

$\begin{pmatrix} v_{0k} \\ v_{1k} \\ v_{2k} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\text{prim}}).$

M6: M2 + (random intercepts **and** random slopes for *verbal* and *social*) by **secondary**

$$\eta_i = +\boldsymbol{\beta} \cdot \mathbf{x}_i + \big(u_{0,\,j(i)} + u_{1,\,j(i)}\,\text{verbal}_i + u_{2,\,j(i)}\,\text{social}_i\big),$$

$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\text{sec}}).$

M7: M2 + (random intercepts **and** random slopes for *verbal* and *social*) by **both** primary and secondary (cross-classified).

Where:

- α is the fixed intercept; $\boldsymbol{\beta}$ are fixed covariates;
- u and v are school-specific random effects;
- σ^2 and Σ are variance/covariance components.

Fitting the models

Let's start by fitting all the models:

```r
df <- df %>% mutate(attain01_num = as.integer(as.character(attain01)))

# formulas (use the numeric outcome!)
m0 <- attain01_num ~ 1
m1 <- attain01_num ~ 1 + (1 | primary) + (1 | second)
m2 <- attain01_num ~ verbal + sex + social
m3 <- attain01_num ~ verbal + sex + social + (1 | primary)
m4 <- attain01_num ~ verbal + sex + social + (1 | second)
m5 <- attain01_num ~ verbal + sex + social + (1 + verbal + social | primary)
m6 <- attain01_num ~ verbal + sex + social + (1 + verbal + social | second)
m7 <- attain01_num ~ verbal + sex + social +
                    (1 + verbal + social | primary) +
                    (1 + verbal + social | second)

# fits
fit_m0 <- stan_glm(m0, data = df, family = binomial("logit"))

fit_m1 <- stan_glmer(m1, data = df, family = binomial("logit"))

fit_m2 <- stan_glm(m2, data = df, family = binomial("logit"))

fit_m3 <- stan_glmer(m3, data = df, family = binomial("logit"))

fit_m4 <- stan_glmer(m4, data = df, family = binomial("logit"))

fit_m5 <- stan_glmer(m5, data = df, family = binomial("logit"))

fit_m6 <- stan_glmer(m6, data = df, family = binomial("logit"))

fit_m7 <- stan_glmer(m7, data = df, family = binomial("logit"))

fits <- list(M0=fit_m0, M1=fit_m1, M2=fit_m2, M3=fit_m3, M4=fit_m4, M5=fit_m5, M6=fit_m6, M7=fit_m7)
```

Model fit comparison (LOOIC)

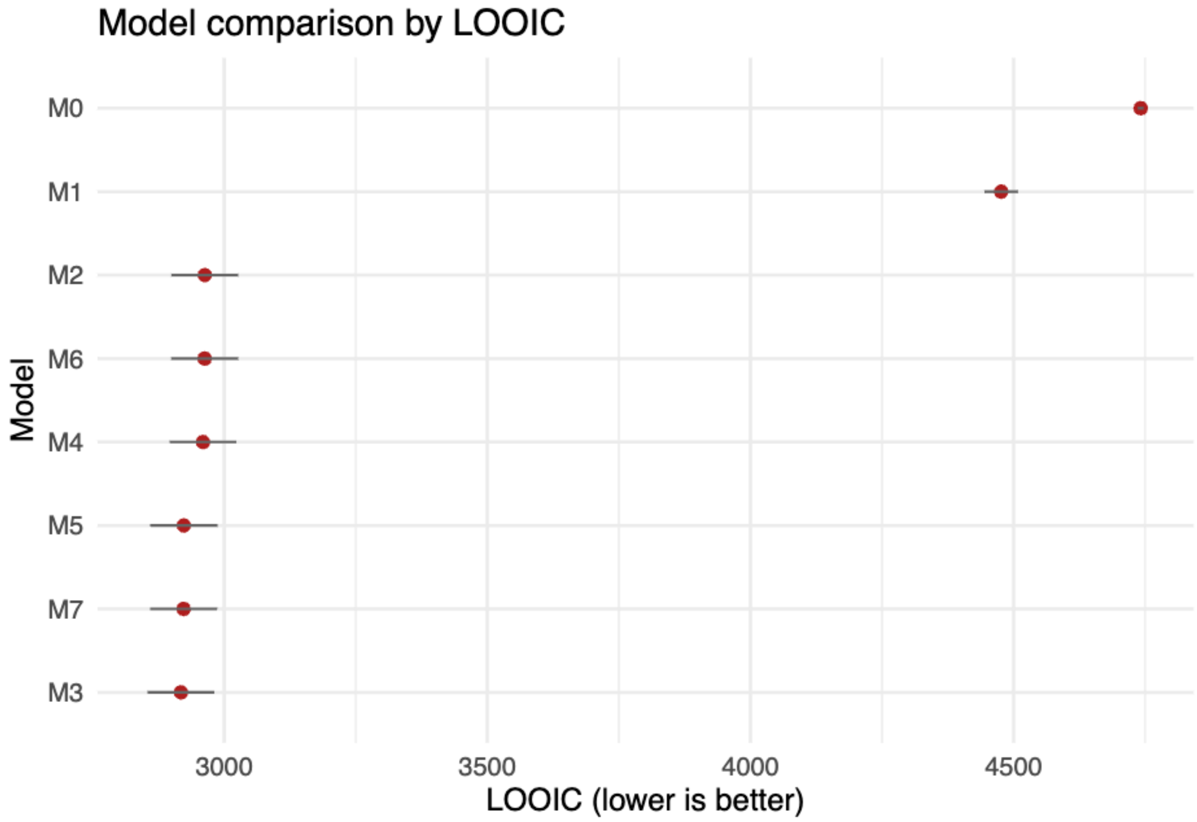
Now that all the models are fit, we can compare them in terms of Leave-One-Out Information Criterion (LOOIC):

```
##    model  elpd_loo  se_elpd     p_loo    looic k_gt_0.7 rank_looic
## 1    M3 -1458.905 31.86151 57.526227 2917.809        0          1
## 2    M7 -1461.444 31.96216 85.398279 2922.888        0          2
## 3    M5 -1461.616 31.93058 82.266357 2923.232        0          3
## 4    M4 -1479.831 31.76483 13.065254 2959.662        0          4
## 5    M6 -1481.478 31.77001 12.955570 2962.956        0          5
## 6    M2 -1481.544 31.75215  4.051887 2963.089        0          6
## 7    M1 -2238.252 15.88539 96.744007 4476.503        0          7
## 8    M0 -2370.770  2.66685  0.989453 4741.541        0          8
```

The best model (with lowest LOOIC) is:

```
## [1] "M3"
```

We can also have a visual comparison of LOOIC values across all the models:



From this overview, we can state that:

- M0 (intercept-only) and M1 (random intercepts only) have much higher LOOIC (≈ 4500 and ≈ 3900), because they fit the data very poorly compared to the other models.
- M2–M7 (all models including the fixed covariates verbal, sex, social) have dramatically lower LOOIC (≈ 2900–3000). This shows that those predictors explain a lot of variation in `attain01`. These models have very similar LOOIC values, hence it seems like adding random intercepts (M3, M4) or random slopes (M5–M7) doesn't clearly outperform the simple fixed-effect model M2, at least in terms of out-of-sample predictive accuracy.

Nonetheless, to investigate the random effect of **`primary`** and given the fact that it scored the lowest LOOIC, we'll opt for **M3**:

$$\eta_i = \boldsymbol{\beta} \cdot \mathbf{x}_i + v_{0,\,k(i)}.$$

Commenting the fit and results of the selected model

We can now explore the model fit, fixef() and ranef():

```
print(best_fit)
```

```
## stan_glmer
##  family:       binomial [logit]
##  formula:      attain01_num ~ verbal + sex + social + (1 | primary)
##  observations: 3424
## ------
##             Median MAD_SD
## (Intercept) -0.1    0.1  
## verbal       0.2    0.0  
## sexF         0.2    0.1  
## social       0.0    0.0  
## 
## Error terms:
##  Groups  Name        Std.Dev.
##  primary (Intercept) 0.49    
## Num. levels: primary 148 
## 
## ------
## * For help interpreting the printed output see ?print.stanreg
## * For info on the priors used see ?prior_summary.stanreg
```

```
fixef(best_fit)
```

```
## (Intercept)      verbal        sexF      social 
## -0.10504633  0.16535383  0.18846404  0.02551166
```

```
ranef(best_fit)
```

```
## $primary
##       (Intercept)
## 1    1.140058e-01
## 2   -2.894742e-03
## 3    2.151796e-01
## 4    2.150737e-01
## 5    4.276223e-02
## 6    5.816678e-01
## 7   -3.091292e-01
## 8   -2.378869e-01
## 9   -1.577945e-02
## 10   2.663168e-01
## 11   -2.614945e-02
## 12   -2.668648e-01
## 13   -1.787103e-01
## 14    1.774165e-01
## 15    2.933143e-01
## 16   -5.509673e-01
## 17    3.521842e-01
## 18    6.823688e-02
## 19    1.967658e-02
## 20   -2.887262e-01
## 21    1.371711e-01
## 22    1.640508e-01
## 23   -3.085712e-01
## 24    8.419485e-03
## 25    1.787774e-01
## 26    1.366879e-01
## 27   -1.706073e-01
## 28    2.296099e-01
## 29   -2.529229e-01
## 30    1.457296e-01
## 31    1.304587e-01
## 32   -2.201771e-01
## 33   -1.391302e-01
## 34    3.432663e-01
## 35    3.521971e-01
## 36   -8.273104e-02
## 37    5.567875e-01
## 38    1.344366e-01
## 39    1.111867e-01
## 40   -7.715728e-02
## 41   -4.587769e-02
## 42   -5.028842e-01
## 43    3.205453e-01
## 44    1.151053e-01
## 45   -3.039675e-01
## 46    2.096576e-01
## 47    5.411546e-01
## 48    1.501362e-01
## 49    4.789019e-01
## 50    9.216162e-02
## 51   -1.448094e-02
## 52   -5.702582e-02
## 53   -2.425414e-01
## 54   -6.382201e-02
## 55   -1.065280e-01
## 56   -2.741403e-02
## 57    3.476422e-01
## 58    5.415610e-02
## 59   -1.455726e-01
## 60   -4.731522e-02
## 61    5.614170e-01
## 62   -1.802673e-02
## 63    3.621268e-02
## 64    7.895084e-02
## 65   1.369575e-01
## 66  -1.635139e-01
## 67   1.162985e-01
## 68  -7.906728e-02
## 69  -5.833291e-01
## 70   2.848014e-01
## 71  -1.983716e-01
## 72   1.575049e-01
## 73   1.384160e-01
## 74  -1.064076e-01
## 75  -4.350305e-02
## 76  -1.474094e-01
## 77   7.528650e-03
## 78   2.410816e-01
## 79  -4.590403e-01
## 80  -8.322536e-02
## 81  -2.941253e-01
## 82  -7.595986e-02
## 83   2.473912e-02
## 84   1.153892e-01
## 85   2.414518e-01
## 86   1.041053e-01
## 87   4.195418e-03
## 88   5.361860e-01
## 89  -5.276672e-01
## 90  -6.845107e-02
## 91  -2.419788e-01
## 92  -4.869787e-01
## 93  -2.131198e-01
## 94   2.919344e-01
## 95   1.959847e-01
## 96  -4.634349e-01
## 97  -4.398972e-01
## 98  -1.764468e-01
## 99   4.538639e-01
## 100 -2.806660e-01
## 101 -2.703292e-02
## 102 -1.187932e-01
## 103  1.396385e-01
## 104  1.673623e-01
## 105 -3.282482e-01
## 106  6.447280e-02
## 107  2.333615e-01
## 108 -4.196315e-01
## 109  2.743591e-01
## 110 -1.032218e-01
## 111 -5.429524e-03
## 112 -3.370415e-01
## 113 -2.457362e-02
## 114  2.981550e-01
## 115  3.378881e-01
## 116  5.730434e-01
## 117 -7.436320e-02
## 118 -2.949592e-01
## 119  2.600116e-01
## 120 -2.939483e-01
## 121 -2.136354e-01
## 122 -5.088323e-01
## 123 -1.976468e-01
## 124 -9.790574e-02
## 125  9.952638e-02
## 126 -2.229622e-01
## 127 -2.975166e-01
## 128  2.245096e-02
## 129 -2.232171e-01
## 130 -2.854087e-01
## 131  3.559805e-01
## 132  5.191223e-01
## 133  2.197141e-02
## 134 -1.095017e-01
## 135 -2.017957e-01
## 136 -4.305810e-01
## 137 -2.204605e-01
## 138  2.397285e-01
## 139 -6.671380e-01
## 140 -2.693094e-01
## 141 -2.254451e-01
## 142  2.663948e-01
## 143  1.249300e+00
## 144 -1.770539e-02
## 145 -3.131736e-02
## 146  2.472468e-03
## 147 -1.782814e-02
## 148  3.514908e-05
##
## with conditional variances for "primary"
```

Hence, we can state that:

- Between-primary variation is substantial, with a random-intercept SD of 0.49. This means that depending on the primary school attended, the odds of achieving attain01 = 1 can shift markedly up or down.
- Averaging over the primary schools:

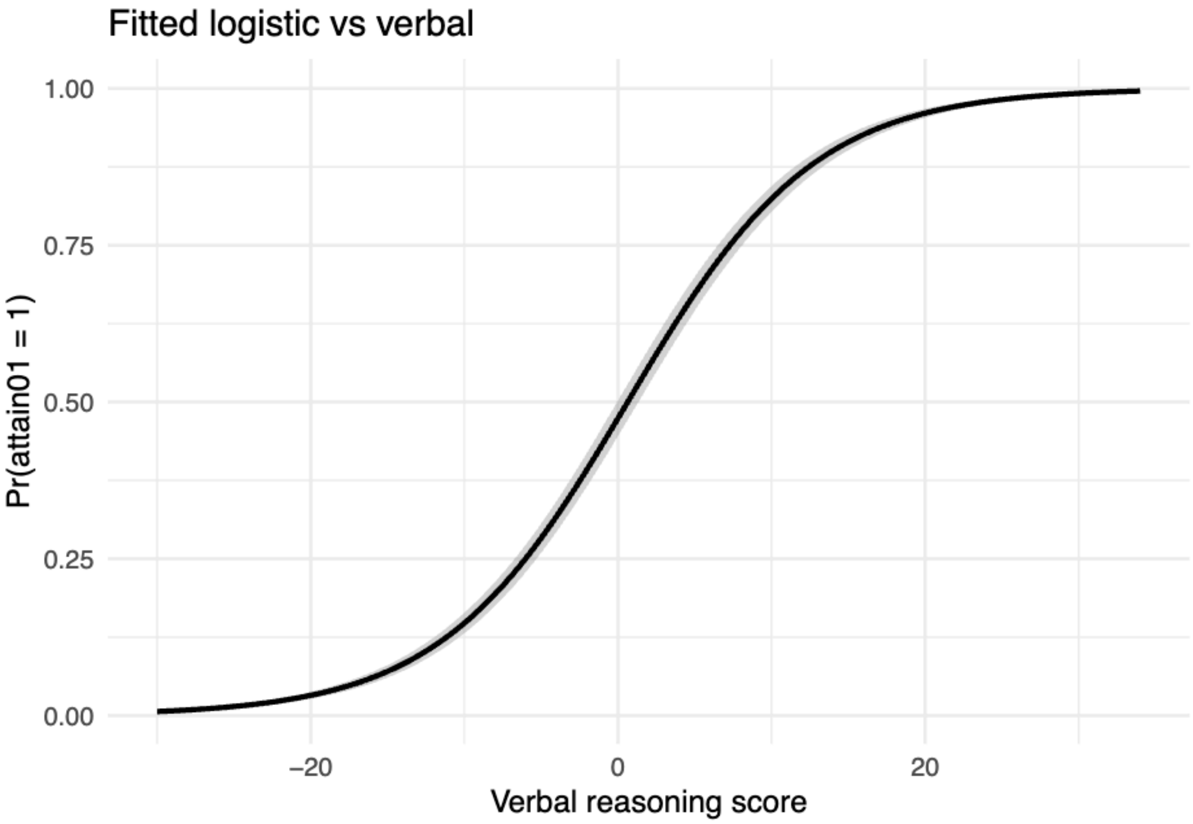
  – Sex: being female increases the odds by a factor of exp(0.187) ≈ 1.21, i.e. about +21% higher odds compared to males.

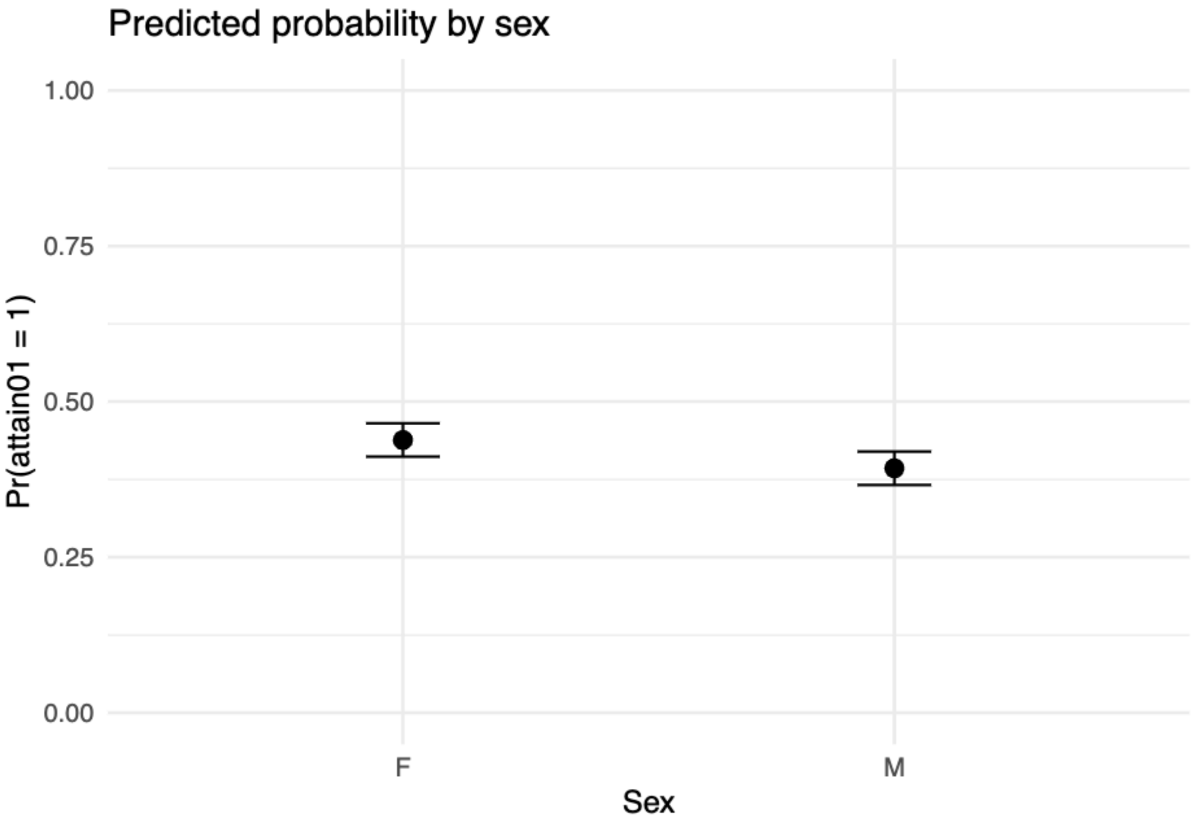
  – Verbal ability: each 1-point increase raises the odds by exp(0.165) ≈ 1.18, i.e. roughly an +18% increase. This is the strongest fixed predictor.

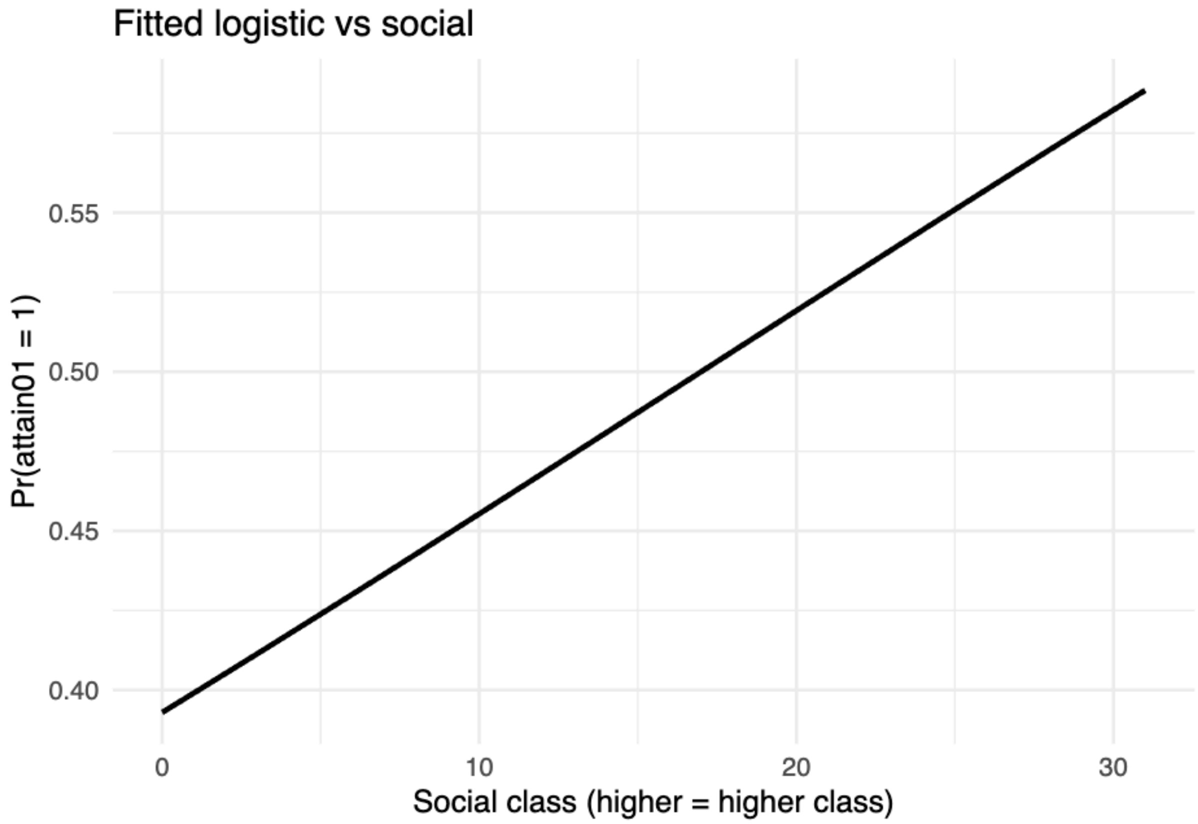
  – Social class: each 1-unit increase raises the odds by exp(0.026) ≈ 1.03, i.e. about a +3% increase. While positive, this effect is modest compared to verbal ability.

Overall, verbal reasoning is the most influential fixed effect coefficient, then sex and social class also contribute positively. However there is still a meaningful between-school heterogeneity contribuiting to the the determination of `attain01`.

We can also have a visual interpretation of the results by plotting the individual level variables againts the Pr(attain01=1):

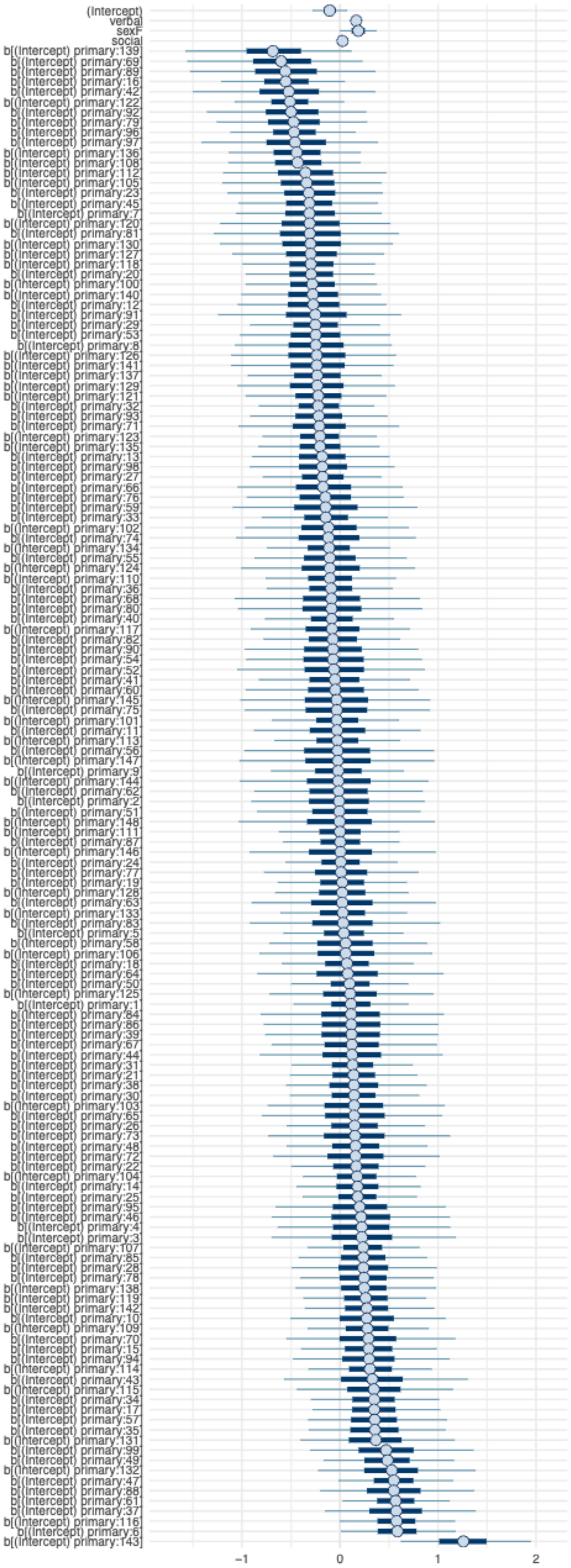

Fitted logistic vs verbal
1.00
0.75
0.50
0.25
0.00
Pr(attain01 = 1)
−20
0
20
Verbal reasoning score


Predicted probability by sex
1.00
0.75
0.50
0.25
0.00
Pr(attain01 = 1)
F
M
Sex



Inference on random effects and relevance of `primary` school groups

Finally, we can plot the fixed and random effects **50%–95%** credibility intervals:

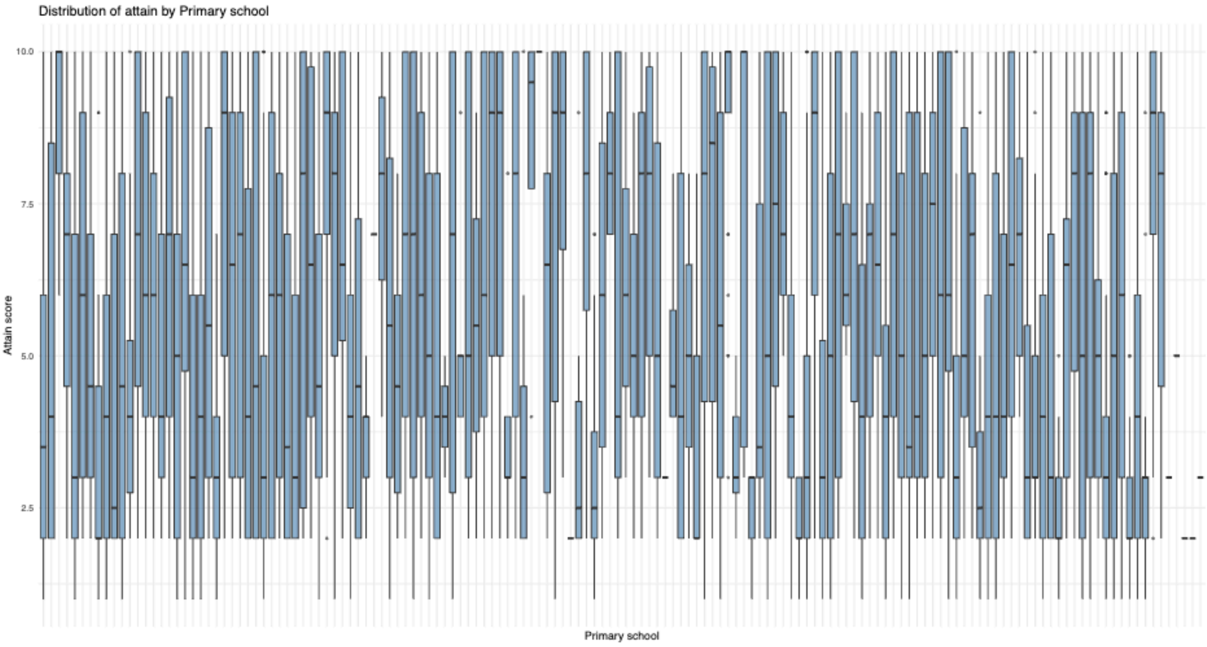

(Intercept)
verbal
sexF
social
b[(Intercept) primary:139]
b[(Intercept) primary:69]
b[(Intercept) primary:89]
b[(Intercept) primary:16]
b[(Intercept) primary:42]
b[(Intercept) primary:122]
b[(Intercept) primary:92]
b[(Intercept) primary:79]
b[(Intercept) primary:96]
b[(Intercept) primary:97]
b[(Intercept) primary:136]
b[(Intercept) primary:108]
b[(Intercept) primary:112]
b[(Intercept) primary:105]
b[(Intercept) primary:23]
b[(Intercept) primary:45]
b[(Intercept) primary:7]
b[(Intercept) primary:120]
b[(Intercept) primary:81]
b[(Intercept) primary:130]
b[(Intercept) primary:127]
b[(Intercept) primary:118]
b[(Intercept) primary:20]
b[(Intercept) primary:100]
b[(Intercept) primary:140]
b[(Intercept) primary:12]
b[(Intercept) primary:91]
b[(Intercept) primary:29]
b[(Intercept) primary:53]
b[(Intercept) primary:8]
b[(Intercept) primary:126]
b[(Intercept) primary:141]
b[(Intercept) primary:137]
b[(Intercept) primary:129]
b[(Intercept) primary:121]
b[(Intercept) primary:32]
b[(Intercept) primary:93]
b[(Intercept) primary:71]
b[(Intercept) primary:123]
b[(Intercept) primary:135]
b[(Intercept) primary:13]
b[(Intercept) primary:98]
b[(Intercept) primary:27]
b[(Intercept) primary:66]
b[(Intercept) primary:76]
b[(Intercept) primary:59]
b[(Intercept) primary:33]
b[(Intercept) primary:102]
b[(Intercept) primary:74]
b[(Intercept) primary:134]
b[(Intercept) primary:55]
b[(Intercept) primary:124]
b[(Intercept) primary:110]
b[(Intercept) primary:36]
b[(Intercept) primary:68]
b[(Intercept) primary:80]
b[(Intercept) primary:40]
b[(Intercept) primary:117]
b[(Intercept) primary:82]
b[(Intercept) primary:90]
b[(Intercept) primary:54]
b[(Intercept) primary:52]
b[(Intercept) primary:41]
b[(Intercept) primary:60]
b[(Intercept) primary:145]
b[(Intercept) primary:75]
b[(Intercept) primary:101]
b[(Intercept) primary:11]
b[(Intercept) primary:113]
b[(Intercept) primary:56]
b[(Intercept) primary:147]
b[(Intercept) primary:9]
b[(Intercept) primary:144]
b[(Intercept) primary:62]
b[(Intercept) primary:2]
b[(Intercept) primary:51]
b[(Intercept) primary:148]
b[(Intercept) primary:111]
b[(Intercept) primary:87]
b[(Intercept) primary:146]
b[(Intercept) primary:24]
b[(Intercept) primary:77]
b[(Intercept) primary:19]
b[(Intercept) primary:128]
b[(Intercept) primary:63]
b[(Intercept) primary:133]
b[(Intercept) primary:83]
b[(Intercept) primary:5]
b[(Intercept) primary:58]
b[(Intercept) primary:106]
b[(Intercept) primary:18]
b[(Intercept) primary:64]
b[(Intercept) primary:50]
b[(Intercept) primary:125]
b[(Intercept) primary:1]
b[(Intercept) primary:84]
b[(Intercept) primary:86]
b[(Intercept) primary:39]
b[(Intercept) primary:67]
b[(Intercept) primary:44]
b[(Intercept) primary:31]
b[(Intercept) primary:21]
b[(Intercept) primary:38]
b[(Intercept) primary:30]
b[(Intercept) primary:103]
b[(Intercept) primary:65]
b[(Intercept) primary:26]
b[(Intercept) primary:73]
b[(Intercept) primary:48]
b[(Intercept) primary:72]
b[(Intercept) primary:22]
b[(Intercept) primary:104]
b[(Intercept) primary:14]
b[(Intercept) primary:25]
b[(Intercept) primary:95]
b[(Intercept) primary:46]
b[(Intercept) primary:4]
b[(Intercept) primary:3]
b[(Intercept) primary:107]
b[(Intercept) primary:85]
b[(Intercept) primary:28]
b[(Intercept) primary:78]
b[(Intercept) primary:138]
b[(Intercept) primary:119]
b[(Intercept) primary:142]
b[(Intercept) primary:10]
b[(Intercept) primary:109]
b[(Intercept) primary:70]
b[(Intercept) primary:15]
b[(Intercept) primary:94]
b[(Intercept) primary:114]
b[(Intercept) primary:43]
b[(Intercept) primary:115]
b[(Intercept) primary:34]
b[(Intercept) primary:17]
b[(Intercept) primary:57]
b[(Intercept) primary:35]
b[(Intercept) primary:131]
b[(Intercept) primary:99]
b[(Intercept) primary:49]
b[(Intercept) primary:132]
b[(Intercept) primary:47]
b[(Intercept) primary:88]
b[(Intercept) primary:61]
b[(Intercept) primary:37]
b[(Intercept) primary:116]
b[(Intercept) primary:6]
b[(Intercept) primary:143]
−1
0
1
2

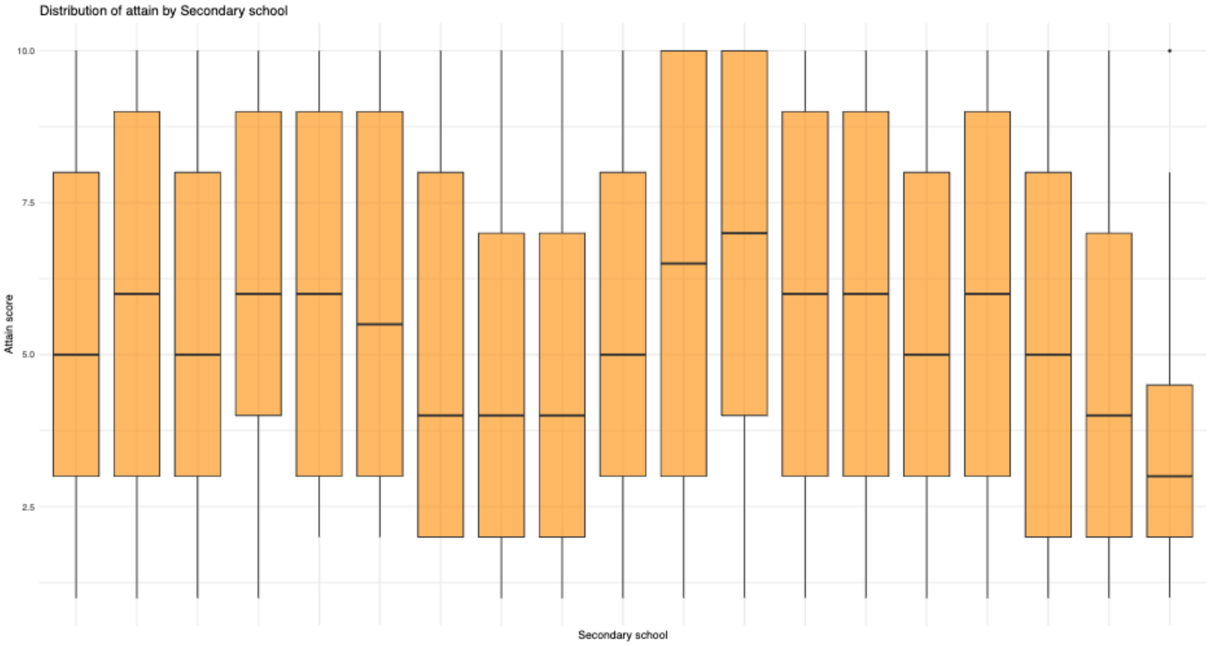
The primary-school random intercept has an SD of 0.49 on the log-odds scale, sign of heterogeneity among primary-school groups. The above plot shows that many schools have 50% intervals separated from 0, and some (e.g. Primary 143) have 95% CIs entirely after 0, indicating a strong positive effect. With a mean of ≈ +1.2, Primary 143 alone multiplies the odds by 3.3: an effect larger than any other fixed effect. Conversely, other schools show clearly negative shifts, cutting the odds roughly in half.

Does the primary school matter? Yes. Between-school differences are often greater than the effects of sex (+21%), verbal ability (+18% per point), or social class (+3% per unit). Moreover, including a primary random effect improved predictive performance (lowest LOOIC), showing that accounting for school-level groups is essential.

Optional: alternative model proposal for `attain`

Before proposing any model, let's look at the distributions of `attain` for each primary and secondary school.





From the unaligned boxplots in these plots it's clear how a **hierarchical Poisson/NB model** could be an interesting proposal.

For example, a **hierarchical NB model** with random effects at the intercepts level for both primary and secondary school could be the following:

Let `attain` be distributed as counts $y_i \sim \text{NegBin}(\lambda_i, \phi)$ with log link $\log \lambda_i = \eta_i$. We use the NB2 parameterization: $\mathbb{E}[y_i] = \lambda_i$, $\text{Var}(y_i) = \lambda_i + \lambda_i^2/\phi$ (over-dispersion $\phi > 0$).

Covariates: **verbal**, **sex** (baseline = M), **social**. Schools: secondary $j(i)$, primary $k(i)$ (cross-classified).

$$\eta_i = u_{0,\,j(i)}^{(\text{second})} + v_{0,\,k(i)}^{(\text{primary})} + \beta_1\,\text{verbal}_i + \beta_2\,\text{sexF}_i + \beta_3\,\text{social}_i,$$

With:

$$u_{0j}^{(\text{second})} \sim \mathcal{N}\big(\mu_{\text{j}}, \sigma^2_{\text{second}}\big), \qquad v_{0k}^{(\text{primary})} \sim \mathcal{N}\big(\mu_{\text{k}}, \sigma^2_{\text{primary}}\big).$$

$$\phi \sim \mathcal{N}^+(0, \gamma_\phi^2)$$

Where: α is the global intercept, β's are fixed effects at the coefficients level; u_{0j} and v_{0k} are school-specific random intercepts capturing residual heterogeneity at the secondary and primary levels; ϕ measures the amount of overdispersion.