

Assessing the Variability in Math Scores: The Role of Schools and Individual Factors Using a Bayesian Mixed Effects Model

Sara Pieruz 1940115

Bayesian Modelling a.y. 2024-2025

Introduction

Educational achievement is influenced by a complex interplay of individual characteristics, school environments, and broader socio-economic factors. Understanding these dynamics is critical for identifying inequities in education systems and informing policies to improve student outcomes. Among these, math performance is often considered a key indicator of educational success, as it is strongly linked to future academic and career opportunities.

The Programme for International Student Assessment (PISA) provides a unique opportunity to study these factors, offering rich data on student performance across various countries. This project focuses on students' math scores, examining how much of the variability can be attributed to differences between schools versus individual factors such as socio-economic status, math anxiety, gender, immigration status and others.

Using a Bayesian mixed-effects model, this analysis seeks to answer the following question: “How much of the variability in math scores among students can be attributed to differences between schools compared to individual factors?”

The Bayesian framework offers a robust method for handling uncertainty and includes prior knowledge in the analysis. By accounting for the hierarchical structure of the data, students grouped within schools, this study seeks to separate the effects of school-level and student-level factors on math performance.

The Dataset

The dataset used in this analysis is derived from the Programme for International Student Assessment (PISA) 2022, a large-scale international survey conducted by the OECD. PISA assesses the performance of 15-year-old students in reading, mathematics, and science across participating countries. For this study, the focus is on a subset of the data that includes students from Italy and contains variables capturing both individual and school-level characteristics.

The dataset includes the following variables:

- Country: Represents the country where the student resides. In this analysis, only students from Italy are included;
- SchoolId: A unique identifier for the school attended by each student. This variable allows for grouping students within schools;
- SCORE: Math scores derived from plausible values (PV1MATH to PV10MATH), which provide multiple estimates of student ability to account for measurement uncertainty;
- Male: A binary variable indicating the student's gender (1 = Male, 0 = Female);
- ANXMAT: A continuous index measuring math anxiety, with higher values indicating greater anxiety;
- ESCS: The Economic, Social, and Cultural Status index, a composite measure that captures a student's socio-economic background;
- TEACHSUP: An index measuring perceived teacher support, with higher values reflecting more positive perceptions of teacher assistance;
- REPEAT: A binary variable indicating whether the student has repeated at least one grade (1 = Yes, 0 = No);
- IMMIG: A binary variable indicating the immigration status of the student (1 = Immigrant, 0 = Native);
- PRIVATESCH: A binary variable indicating the type of school attended (1 = Private, 0 = Public). This information was obtained by merging the student dataset with the school questionnaire dataset.

The dataset includes 10552 students from 344 schools. The extraction and the manipulation done on the dataset is reported at the end of this document.

Descriptive Statistics

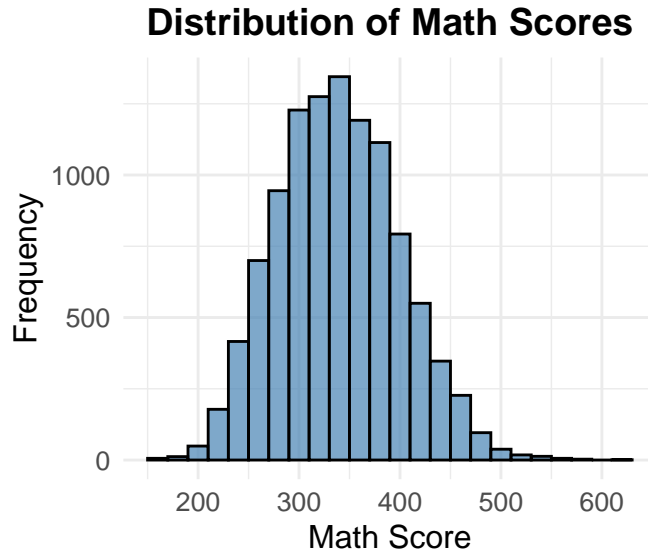
This section provides an overview of the dataset used in this analysis, highlighting key patterns and relationships between variables. By examining the distribution of math scores and their association with factors such as gender, school type, immigration status, and grade repetition, this section lays the foundation for understanding the variability in student performance. These insights will guide the subsequent modeling efforts to explore the contributions of individual and school-level factors to math achievement.

This first histogram displays the distribution of math scores for students in the dataset.

The distribution of math scores appears roughly symmetrical with a peak around 300-360. This suggests that most students score within this range, representing the majority of the population. The mean score is 338.9404 (Table 1) and the range is from approximately 150 to 600, with fewer students achieving very low or very high scores.

Table 1: Mean Math Score

Metric	Value
Mean Score	338.9404



If we consider the students' performance by gender, we can observe that:

- The proportion between males and females is well-balanced (Table 2);
- The distribution of Males' math scores (blue) appear to have slightly higher scores on average compared to females (pink);
- The mean math score for male students is approximately 346.79, while for female students it is 331.34 (Table 3);
- The shape of the distributions for both genders is fairly similar, but the center for males is slightly shifted to the right (better performance on average).

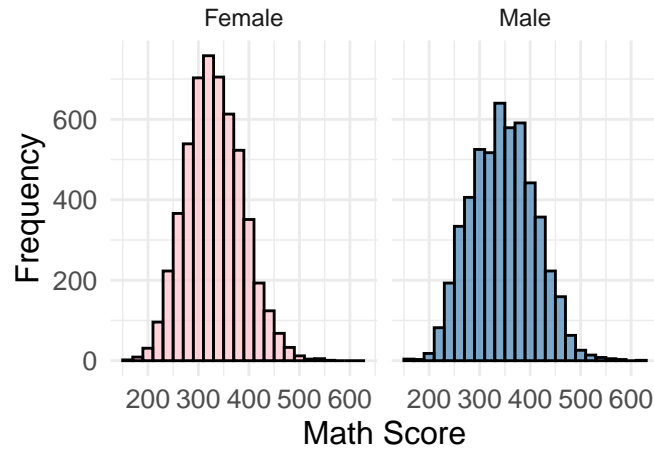
Table 2: Proportions of Gender

Gender (0 = Female, 1 = Male)	Proportion
0	0.5078658
1	0.4921342

Table 3: Mean Math Scores by Gender

Gender (0 = Female, 1 = Male)	Mean Score
0	331.3378
1	346.7860

Distribution of Math Scores by Gender



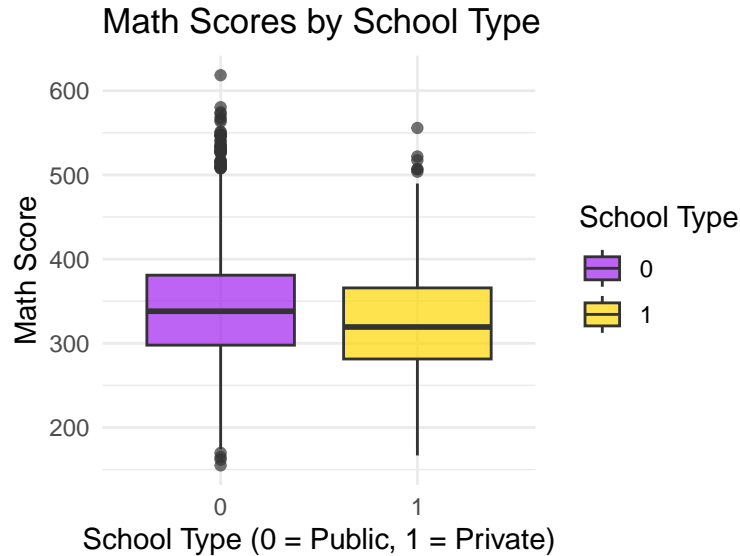
The following boxplot compares math scores between public and private school students.

We notice that students in private schools (yellow) tend to have lower scores compared to those in public schools (purple). However, we need to consider the fact that students in private schools are about the 9.6% of the whole sample (Table 4). The results might reflect characteristics of public school students more strongly than those of private school students. This could bias the overall descriptive statistics, such as mean scores and distributions.

In later stages, such as Bayesian mixed-effects modeling, the underrepresentation could lead to larger uncertainty in parameter estimates related to private schools, as fewer data points contribute to those estimates.

Table 4: Proportions of PRIVATESCH

PRIVATESCH	Proportion
0	0.9039045
1	0.0960955



The variables ESCS, ANXMAT, TEACHSUP, REPEAT, and IMMIG contain missing data, which necessitates addressing this issue to ensure the accuracy and validity of the analysis. Missing data can bias results and reduce statistical power if not handled properly. To address this, I employed multiple imputation using the mice package in R. This approach imputes missing values by generating plausible replacements based on the relationships among variables. For continuous variables like ESCS, ANXMAT, and TEACHSUP, a Bayesian normal imputation method was used (Gibbs sampling), while binary variables such as REPEAT and IMMIG were imputed using logistic regression (maximum likelihood estimation). Five imputed datasets were generated, and the first completed dataset was selected for analysis, ensuring a robust approach to handling missing data.

In this case, as we can see from the R output, all the variables are used as predictors to impute missing data for other variables, ensuring that the imputations are informed by the relationships in the dataset (each variable is not used to predict itself).

```
##      ...1   Country   SchoolId   SCORE   Male   ANXMAT   ESCS
##        0         0         0         0         0       803    101
##  TEACHSUP   REPEAT   IMMIG PRIVATE SCH
##      447       94      271         0
```

```
## Class: mids
## Number of multiple imputations: 5
## Imputation methods:
##   ESCS  ANXMAT TEACHSUP  REPEAT  IMMIG
## "norm" "norm"  "norm" "logreg" "logreg"
## PredictorMatrix:
##   ESCS  ANXMAT TEACHSUP  REPEAT  IMMIG
## ESCS    0      1      1      1      1
```

```
## ANXMAT      1      0      1      1      1
## TEACHSUP     1      1      0      1      1
## REPEAT       1      1      1      0      1
## IMMIG        1      1      1      1      0
```

```
##      ...1    Country    SchoolId      SCORE      Male      ANXMAT      ESCS
##      0      0      0      0      0      0      0
## TEACHSUP    REPEAT      IMMIG PRIVATESCH
##      0      0      0      0
```

As we can observe from Table 5 and Table 6, for the variables IMMIG and REPEAT we have the same problem as for PRIVATESCH. The percentage of non-native students is around 12% and the percentage of students that have repeated at least one year of school is around 9.6%. Therefore, the descriptive analysis can be biased by this sample proportion.

From the IMMIG boxplot we observe that natives (light blue) generally perform better in math compared to immigrant students (orange). Furthermore, immigrant students show slightly lower median scores and less variability.

From the REPEAT boxplot we observe that students who have repeated at least a grade (dark blue) have significantly lower scores compared to those who have not (dark red).

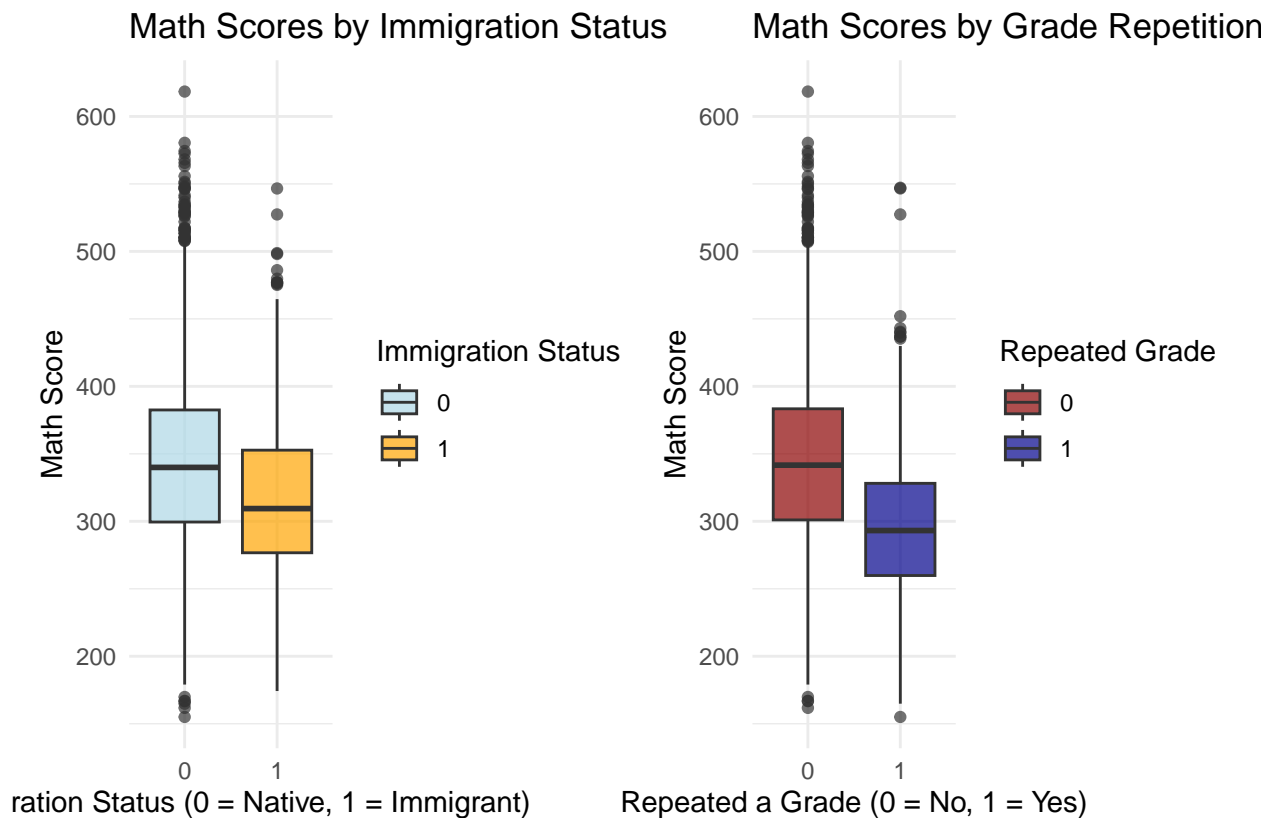
Both findings align with expectations. Grade repetition is often indicative of academic struggles, and in many educational contexts, being an immigrant can negatively impact student grades. However, the latter is not an inherent characteristic but rather the outcome of a combination of factors that influence academic performance.

Table 5: Proportions of IMMIG

IMMIG	Proportion
0	0.8791698
1	0.1208302

Table 6: Proportions of REPEAT

REPEAT	Proportion
0	0.9036202
1	0.0963798



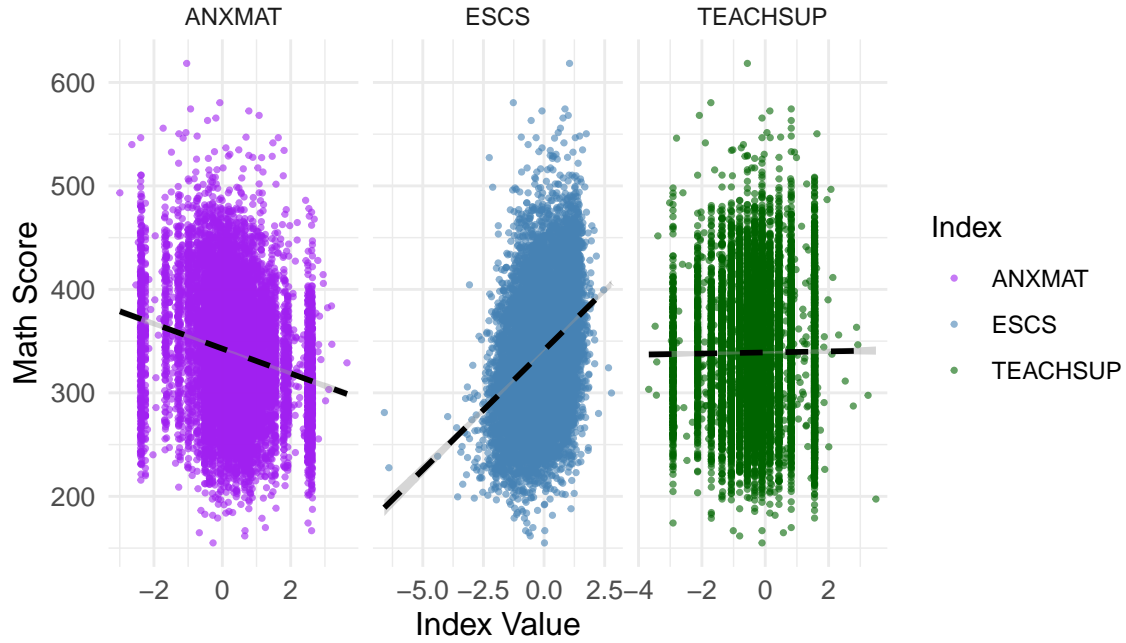
The following graph illustrates the relationship between Math Scores (SCORE) and three continuous indices: ANXMAT (Math Anxiety), ESCS (Economic, Social, and Cultural Status), and TEACHSUP (Teacher Support). Each index is represented in a separate facet, with points indicating individual students and a trend line highlighting the overall relationship for each variable.

Starting with ANXMAT, a negative relationship is evident: the trend line slopes slightly downward, indicating that higher levels of math anxiety are associated with lower math scores. However, the data points are widely dispersed, suggesting that while math anxiety negatively influences performance, other factors also contribute significantly to the variability in scores.

For the ESCS variable, we observe a clear positive relationship: the trend line slopes upward, showing that students with higher ESCS values tend to achieve higher math scores. This relationship appears stronger than that of ANXMAT, as the data points are more closely aligned with the trend line. This indicates that socioeconomic status has a consistent and significant positive effect on student performance in math.

Finally, TEACHSUP shows a weak relationship with math scores: the trend line is nearly flat, suggesting little to no association between teacher support and scores. Additionally, the data points are evenly distributed across the range of scores, implying that perceived teacher support does not strongly predict math performance within this sample.

Math Scores in Relation to ESCS, ANXMAT, and TEACHSUP



Model Specification

The objective of this section is to develop and specify Bayesian mixed-effects models to analyze the variability in math scores and evaluate the effects of individual and school-level factors. By employing a hierarchical structure, the models account for both individual-level predictors (e.g., socioeconomic status, math anxiety) and school-level clustering, allowing us to partition the variability in scores across these two levels. Four models are presented to systematically investigate the effects of predictors, assess the impact of random effects, and evaluate model fit.

Theoretical Framework

A Bayesian mixed-effects model extends the traditional linear model by incorporating both fixed effects and random effects. Fixed effects represent population-level parameters, while random effects capture group-level variability—in this case, differences between schools. The inclusion of random effects allows the model to account for the hierarchical structure of the data, where students are nested within schools. This approach provides greater flexibility and improves inference by borrowing strength across groups. The Bayesian framework further enhances this analysis by integrating prior knowledge and quantifying uncertainty through posterior distributions.

The general structure of a Bayesian mixed-effects model for nested data can be expressed as:

$$y_{ij} = \beta_0 + \sum_{k=1}^K \beta_k x_{ijk} + u_j + \epsilon_{ij},$$

Here, the components of the model follow these distributions:

- y_{ij} : Observed outcome (math score) for student i in school j , assumed to follow a normal distribution with mean μ_{ij} and variance σ^2 :

$$y_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma^2).$$

where the mean structure is given by:

$$\mu_{ij} = \beta_0 + \sum_{k=1}^K \beta_k x_{ijk} + u_j.$$

- β_k : Fixed-effect coefficients for predictors x_{ijk} , assigned weakly informative priors:

$$\beta_k \sim \mathcal{N}(0, \tau_\beta),$$

where τ_β is the precision of the prior, in this case it is set as a particularly small value to reflect a weakly informative prior.

- u_j : Random intercepts for schools, capturing school-specific deviations from the population-level mean, assumed to follow a normal distribution:

$$u_j \sim \mathcal{N}(0, \sigma_u^2),$$

where σ_u^2 represents the variance of the random effects across schools.

- ϵ_{ij} : Residual errors, capturing the variability within schools not explained by the predictors or random effects, assumed to follow a normal distribution:

$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2).$$

- σ^2 and σ_u^2 : Variance components, modeled using inverse-gamma priors (in the code as weakly informative gamma priors):

$$\sigma^2 \sim \text{InverseGamma}(a, b), \quad \sigma_u^2 \sim \text{InverseGamma}(a, b),$$

where a and b are hyperparameters reflecting prior beliefs about the variances.

By defining these distributions, the Bayesian framework captures the uncertainty in parameter estimates and allows for probabilistic interpretations. The hierarchical structure of the model facilitates the partitioning of variability in math scores into within-school and between-school components.

Model Formulations

The following four models are considered in this study:

Model 1: Null Model (Random Intercept Only) The null model includes only a random intercept for schools, with no predictors. This model serves as a baseline to decompose the total variability in math scores into within-school and between-school components:

$$y_{ij} = \beta_0 + u_j + \epsilon_{ij}.$$

Model 2: Random Effects with Key Individual Predictors The second model incorporates three key individual-level predictors: gender (**Male**), socioeconomic status (**ESCS**), and math anxiety (**ANXMAT**). This model evaluates how much of the within-school variability can be explained by these predictors:

$$y_{ij} = \beta_0 + \beta_1 \text{Male}_{ij} + \beta_2 \text{ESCS}_{ij} + \beta_3 \text{ANXMAT}_{ij} + u_j + \epsilon_{ij}.$$

Model 3: Full Random Effects Model The third model extends the second by including all available predictors: teacher support (**TEACHSUP**), grade repetition (**REPEAT**), immigration status (**IMMIG**), and school type (**PRIVATESCH**):

$$y_{ij} = \beta_0 + \beta_1 \text{Male}_{ij} + \beta_2 \text{ESCS}_{ij} + \beta_3 \text{ANXMAT}_{ij} + \beta_4 \text{TEACHSUP}_{ij} \\ + \beta_5 \text{REPEAT}_{ij} + \beta_6 \text{IMMIG}_{ij} + \beta_7 \text{PRIVATESCH}_{ij} + u_j + \epsilon_{ij}.$$

Model 4: Fixed Effects Only (No Random Effects) To assess the importance of random effects, a fourth model is specified without random intercepts. This model assumes no between-school variability:

$$y_{ij} = \beta_0 + \beta_1 \text{Male}_{ij} + \beta_2 \text{ESCS}_{ij} + \beta_3 \text{ANXMAT}_{ij} + \beta_4 \text{TEACHSUP}_{ij} \\ + \beta_5 \text{REPEAT}_{ij} + \beta_6 \text{IMMIG}_{ij} + \beta_7 \text{PRIVATESCH}_{ij} + \epsilon_{ij}.$$

Model Fit and Comparison

To compare these models, the Deviance Information Criterion (DIC) is used, with lower DIC values indicating better model fit. Additionally, the inclusion of Model 4 enables a direct comparison of the effects of incorporating random intercepts on the results. The systematic examination of these models allows us to identify key factors associated with math scores and the extent to which school-level variability influences student performance.

The results in the following table show that Model 3 provides the best fit to the data, as evidenced by its lowest DIC (110240.8). This model effectively captures both individual and school level factors, incorporating random effects to account for school clustering.

By contrast, Model 4, which excludes random effects, has a substantially higher DIC (114011.2), highlighting the importance of accounting for school-level variability.

The DIC values indicate that incorporating both random effects and a full set of predictors enhances model performance. Therefore, Model 3 will be selected for further interpretation and diagnostic analysis. Only the results from Model 3 will be presented and discussed; however, the outputs of the other models can be found at the end of the report.

Table 7: DIC Values for Different Models

Model	DIC
model1	111507.1
model2	110567.8
model3	110240.8
model4	114011.2

```
## Inference for Bugs model at "6", fit using jags,
## 1 chains, each with 10000 iterations (first 1000 discarded), n.thin = 10
## n.sims = 900 iterations saved. Running time = 322.3 secs
##      mu.vect sd.vect      2.5%      25%      50%      75%
## beta0      332.929  2.270   328.519   331.467   332.922   334.373
## beta1       12.554  0.952    10.668    11.943    12.538    13.211
## beta2        6.559  0.579     5.477     6.166     6.562     6.947
## beta3       -9.212  0.422   -10.016    -9.489    -9.229    -8.943
## beta4        0.616  0.402    -0.178     0.351     0.612     0.868
## beta5      -25.447  1.591   -28.514   -26.517   -25.480   -24.367
## beta6      -10.210  1.441   -13.089   -11.265   -10.213    -9.207
## beta7       -8.783  5.524   -19.553   -12.512    -8.845    -5.233
## sigma       44.108  0.309    43.501    43.911    44.102    44.326
## sigma_u     34.101  1.431    31.394    33.093    34.061    35.059
## deviance 109865.331 27.402 109810.933 109847.552 109865.950 109883.081
##      97.5%
## beta0      337.406
## beta1       14.354
## beta2        7.693
## beta3       -8.377
## beta4        1.415
## beta5      -22.204
## beta6       -7.438
## beta7        2.454
## sigma       44.709
## sigma_u     36.966
## deviance 109921.404
##
## DIC info (using the rule: pV = var(deviance)/2)
## pV = 375.4 and DIC = 110240.8
## DIC is an estimate of expected predictive error (lower deviance is better).
```

Interpretation of Results (Model 3)

The results of the Bayesian mixed-effects model provide insights into the factors influencing math scores. The intercept ($\beta_0 = 332.93$, 95% CI: [328.52, 337.41]) represents the baseline math score for a student with all predictors at their reference or mean levels. Key individual-level predictors include:

- Gender ($\beta_1 = 12.55$, 95% CI: [10.67, 14.35]): Male students tend to score higher than female students, reflecting a significant positive effect of gender on math performance. A more detailed interpretation is as follows: there is a 95% probability that the true effect of being male on math scores lies between 10.67 and 14.35. This positive range indicates that males consistently outperform females, and since the interval does not include 0, the effect is statistically significant.
- Socioeconomic status ($\beta_2 = 6.56$, 95% CI: [5.48, 7.69]): Higher ESCS is strongly associated with better math scores, highlighting the importance of socioeconomic background.
- Math anxiety ($\beta_3 = -9.21$, 95% CI: [-10.02, -8.38]): Increased math anxiety negatively impacts performance, with a substantial decrease in scores as anxiety rises.
- Teacher support ($\beta_4 = 0.62$, 95% CI: [-0.18, 1.42], non significant): The effect of teacher support is small and not statistically significant, as the credible interval includes zero.
- Grade repetition ($\beta_5 = -25.45$, 95% CI: [-28.51, -22.20]): Repeating a grade has a significant negative association with math scores, indicating a substantial disadvantage for these students.
- Immigration status ($\beta_6 = -10.21$, 95% CI: [-13.09, -7.44]): Immigrant students score lower on average compared to their native peers.
- School type ($\beta_7 = -8.78$, 95% CI: [-19.55, 2.45], non significant): Attending a private school has a negative effect on scores, though the effect is not statistically significant.

The estimated variance components indicate that school-level differences ($\sigma_u = 34.10$, 95% CI: [31.39, 36.97]) explain a substantial portion of the variability in scores, in addition to within-school variability ($\sigma = 44.11$, 95% CI: [43.50, 44.71]).

The model fit, as said before, suggests that this model captures both individual and school level influences effectively. Overall, the findings emphasize the importance of accounting for hierarchical structures in educational data and underscore the significant roles of socioeconomic background, math anxiety, and grade repetition in shaping student performance.

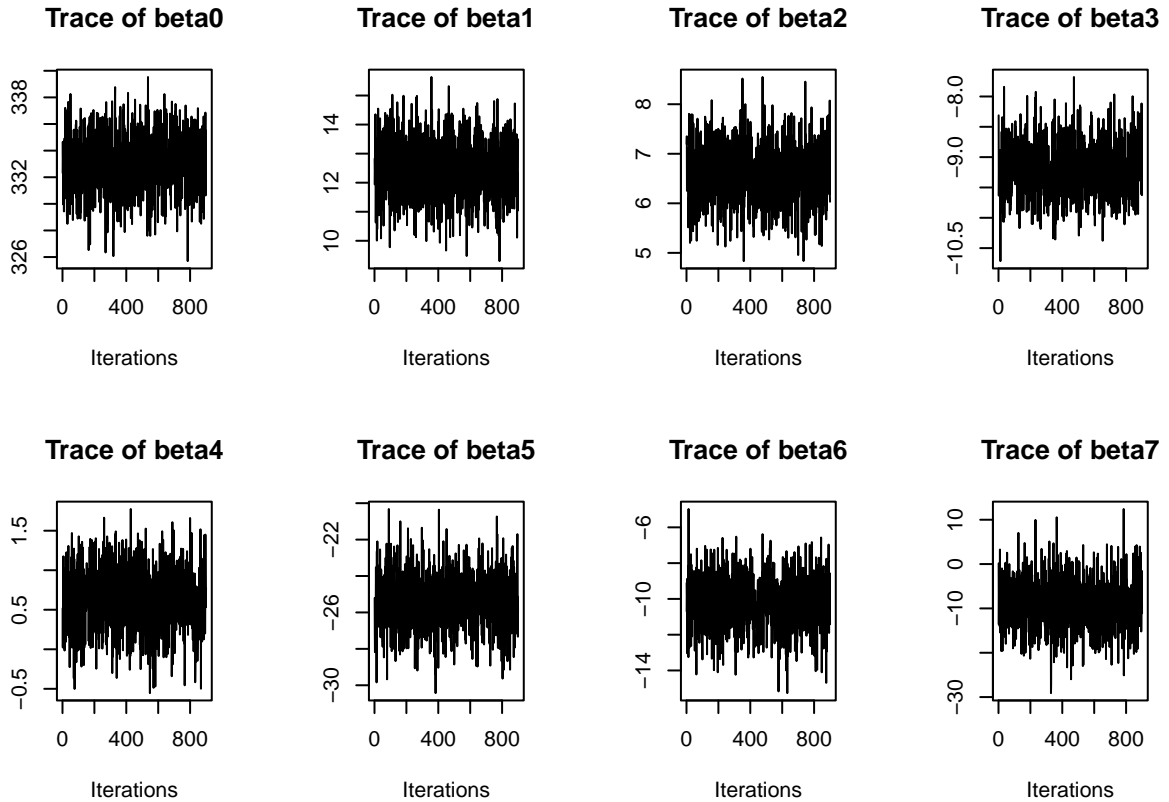
Diagnostics

To ensure the validity and reliability of the Bayesian mixed-effects models, we conducted several diagnostic checks to evaluate convergence, goodness-of-fit, and model assumptions. These diagnostics are essential for interpreting the results and identifying any potential issues with the model. Since the Model 3 exhibited the lowest Deviance Information Criterion (DIC), all diagnostics were performed on this model.

1. Convergence Diagnostics

We assessed the convergence of the Markov Chain Monte Carlo (MCMC) chains by analyzing the trace plots for the model parameters. The trace plots display the MCMC sampling process for the parameters β_i , where $i = 0, \dots, 7$. We find that:

- **Stability:** For all parameters, the trace plots demonstrate stability, with the samples fluctuating around a consistent mean value.
- **Good Mixing:** The chains exhibit good mixing, with no discernible trends or drift over iterations, indicating that the sampler has likely converged to the target posterior distribution.
- **Burnin Phase:** There is no evident “burnin” phase, where the chains transition to their stationary distribution, suggesting that the burnin period was sufficient.

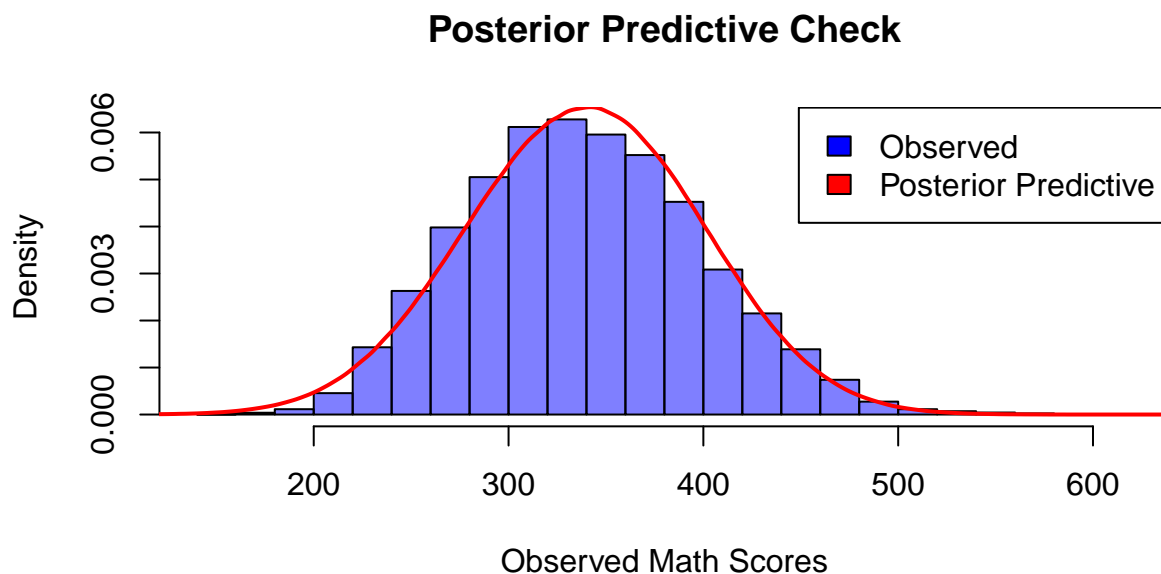


2. Posterior Predictive Checks

Posterior predictive checks were conducted to evaluate how well the model captures the observed data.

From the following plot, we observe that the posterior predictive distribution (red curve) aligns well with the observed distribution (blue bars).

So, the posterior predictive check indicates that the model effectively captures the overall distribution of math scores, including the mean, variance, and general shape of the observed data. This provides confidence in the model's adequacy for inference and prediction. However, it is important to note that this plot reflects the overall fit of the model and does not indicate whether the model fits all subgroups equally well (e.g., males vs. females, public vs. private schools).

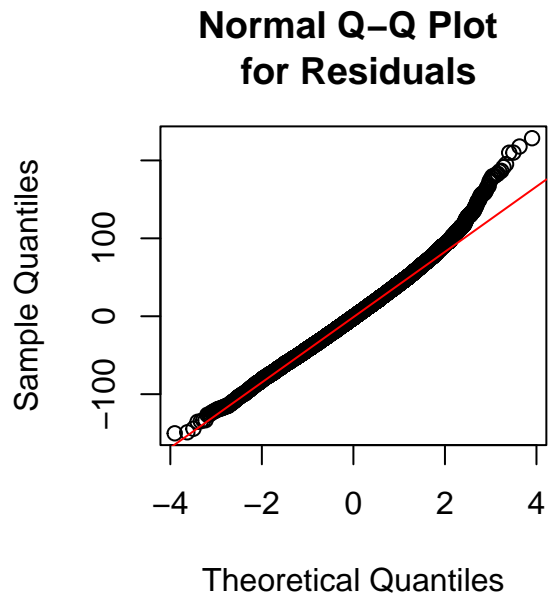
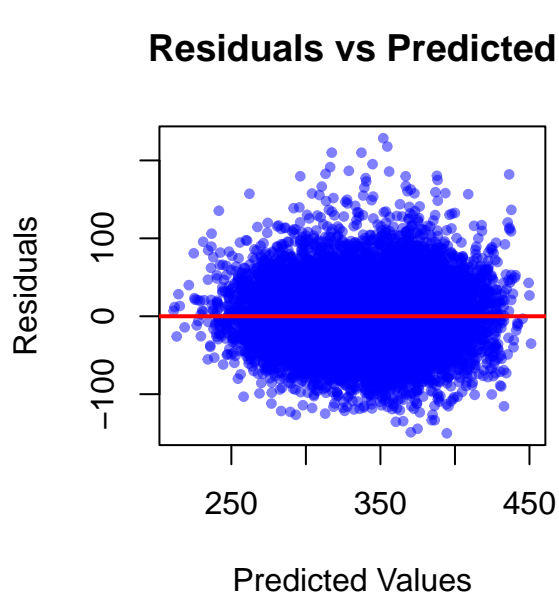


3. Residual Diagnostics

Residual diagnostics were performed to assess normality and homoscedasticity. The residuals are centered around zero and exhibit no clear patterns, suggesting that the model does not systematically underpredict or overpredict, and it satisfies the assumption of constant variance. The spread of residuals is consistent across the predicted range, indicating that the model effectively captures data variability.

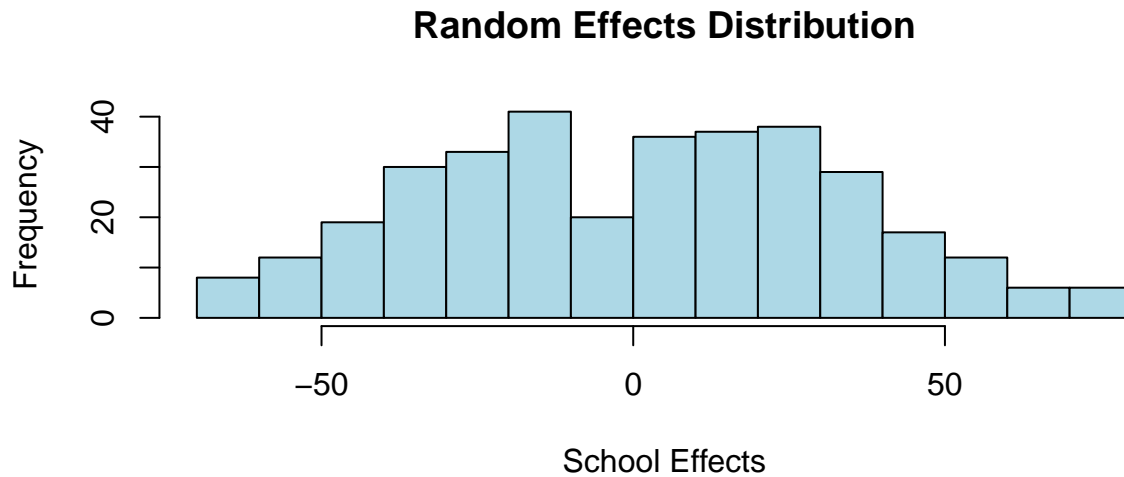
However, a few extreme residuals may signal outliers or cases where the model performs less accurately. While the overall fit of the model is confirmed by these diagnostics, further investigation into subgroup-level residual patterns is warranted.

Additionally, a Q-Q plot was used to assess the normality of the residuals. Most points align closely with the red diagonal line, indicating that the residuals generally adhere to the normality assumption. However, deviations are observed in the tails, where residuals are more extreme than expected, suggesting the presence of outliers or heavier tails, as seen in the residuals plot. These tail deviations may indicate the need for a more robust model, even if only the inference for extreme cases may be affected.



4. Interpretation of Random Effects

The histogram of school-level random effects demonstrates a symmetric distribution centered around zero, indicating that schools, on average, do not systematically increase or decrease student math scores relative to the overall mean. The effects range from approximately -50 to +50, reflecting substantial variability across schools. This supports the model's assumption of normally distributed random effects.

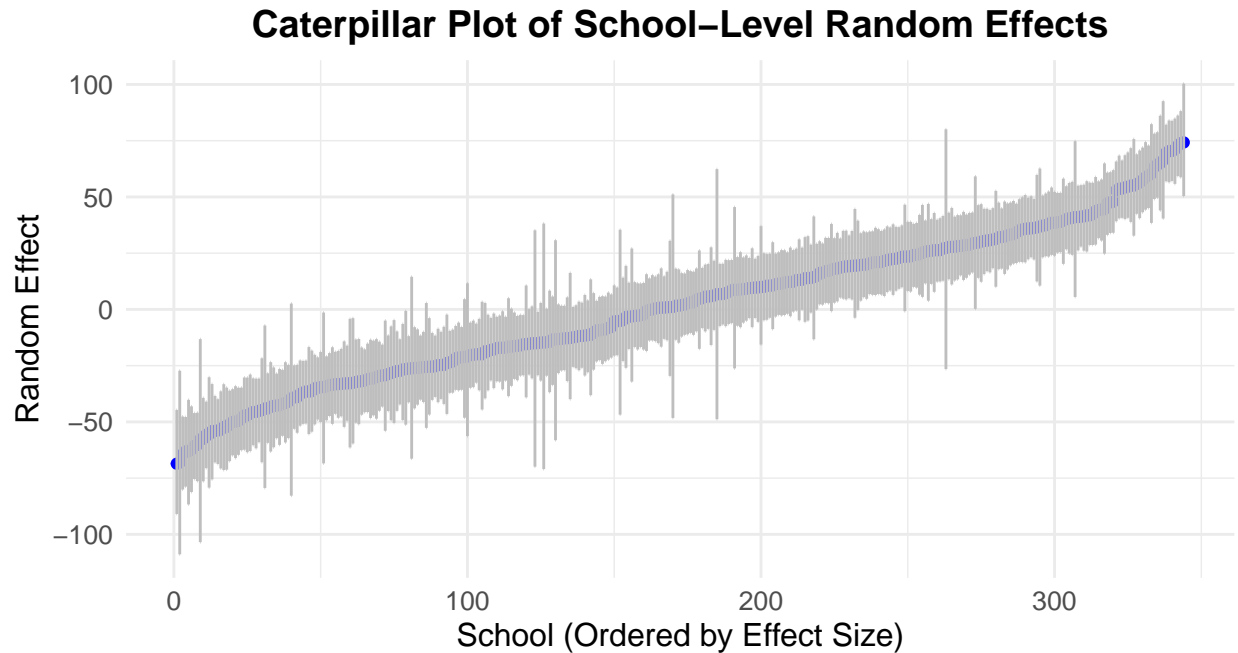


The caterpillar plot provides a more detailed visualization of the estimated random effects for each school, ordered by effect size. Each blue point represents the mean random effect for a specific school, while the vertical gray lines indicate the 95% credible intervals for the estimates. This plot is particularly useful for identifying variability among schools and assessing the precision of the random effect estimates.

From the plot, it is evident that the random effects are approximately symmetrically distributed around zero, consistent with the assumption of normally distributed school-level effects. Schools with positive random effects have above-average contributions to student math scores, while those with negative random effects contribute below the average. Notably, the range of effects spans from approximately -100 to $+100$, highlighting substantial variability in school contributions.

The length of the credible intervals varies across schools, reflecting differences in the precision of the estimates. Schools with fewer students tend to have wider intervals, indicating greater uncertainty in their random effect estimates, while schools with larger sample sizes have narrower intervals, demonstrating more precise estimates.

This plot effectively illustrates the hierarchical structure of the data and emphasizes the importance of accounting for school-level clustering in the model. Additionally, it highlights outliers and extreme cases, where certain schools have unusually high or low effects, which could be explored further to understand specific factors driving these differences.



5. Effective Sample Size (ESS)

The effective sample size (ESS) measures the number of independent samples obtained from the MCMC chains, accounting for autocorrelation. High ESS values ensure reliable posterior estimates. All parameters exhibit ESS values above 900, indicating excellent mixing and precision.

These values confirm the model’s robustness, with sufficient precision for all parameters.

```
##      beta0      beta1      beta2      beta3      beta4      beta5      beta6      beta7
## 1073.750  900.000 1007.752  900.000 1016.565  900.000  900.000  900.000
## deviance    mu[1]
##  900.000  900.000
```

Frequentist vs. Bayesian Approach

In this section, we compare the results from three models applied to the dataset: a simple linear regression model (LM), a frequentist linear mixed-effects model (LMM), and a Bayesian mixed-effects model (BMM). Each model incorporates the same predictors, but they differ in their underlying assumptions and methods for handling hierarchical data.

Simple Linear Regression Model

The LM assumes independence of observations and does not account for the nested structure of the data (students within schools). The model explains approximately 20.3% of the variability in math scores, as indicated by the adjusted R^2 . Key predictors, such as **Male** ($\beta_1 = 12.91$, $p < 0.001$), **ESCS** ($\beta_2 = 19.71$, $p < 0.001$), and **ANXMAT** ($\beta_3 = 10.05$, $p < 0.001$), show significant associations with math scores. However, the residual standard error ($RSE = 53.68$) is relatively high, reflecting unexplained variability due to ignoring the school-level clustering.

Frequentist Linear MixedEffects Model

The LMM introduces random intercepts to account for school-level variability. The model reveals that the variance in math scores can be attributed to both residual ($\sigma^2 = 1947$) and school-level ($\sigma_u^2 = 1157$) components. The inclusion of random effects reduces the residual variance compared to the LM ($RSE = 44.12$), indicating improved fit. Fixed-effect estimates are similar to those in the LM but adjusted for the hierarchical structure. For example, **ESCS** ($\beta_2 = 6.57$) remain significant but with smaller effect sizes, highlighting the impact of accounting for between-school variability. The REML criterion (110822.2) suggests better overall model fit compared to the LM.

Bayesian MixedEffects Model

The BMM, estimated using JAGS, provides probabilistic interpretations of parameter estimates. The posterior means for key predictors, such as **Male** ($\beta_1 = 12.49$, 95% CI: [10.61, 14.36]) and **ESCS** ($\beta_2 = 6.56$, 95% CI: [5.53, 7.66]), align closely with those from the LMM. Credible intervals offer a more intuitive understanding of uncertainty compared to frequentist confidence intervals. The inclusion of random effects is evident in the posterior mean estimate for the school-level standard deviation ($\sigma_u = 34.14$), consistent with the LMM. The DIC (110263.9) indicates a slightly better fit than the LMM, while providing richer inferential insights.

Overall, these findings highlight the importance of using mixed-effects models when analyzing hierarchical data, as they account for clustering and provide more reliable estimates. The Bayesian approach, while computationally intensive, delivers enhanced interpretability and precision, making it a valuable tool for understanding complex relationships in educational data.

```
## [1] "Linear model summary:"

##
## Call:
## lm(formula = SCORE ~ Male + ESCS + ANXMAT + TEACHSUP + REPEAT +
##     IMMIG + PRIVATESCH, data = data_ITA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -180.360  -36.424   -0.456   35.271  278.048
```

```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 337.80373    0.78539 430.112  <2e-16 ***
## Male        12.90891    1.06477  12.124  <2e-16 ***
## ESCS        19.70734    0.61147  32.230  <2e-16 ***
## ANXMAT      -10.05423    0.50130 -20.056  <2e-16 ***
## TEACHSUP     0.06619    0.47069   0.141    0.888
## REPEAT     -35.95963    1.82779 -19.674  <2e-16 ***
## IMMIG       -1.67292    1.69361  -0.988    0.323
## PRIVATESCH -16.11253    1.77961  -9.054  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 53.68 on 10544 degrees of freedom
## Multiple R-squared:  0.2031, Adjusted R-squared:  0.2026
## F-statistic: 384 on 7 and 10544 DF, p-value: < 2.2e-16

## [1] "Frequentist mixed-effects model summary:"

## Linear mixed model fit by REML ['lmerMod']
## Formula:
## SCORE ~ Male + ESCS + ANXMAT + TEACHSUP + REPEAT + IMMIG + PRIVATESCH +
##       (1 | SchoolId)
## Data: data_ITA
##
## REML criterion at convergence: 110822.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.3865 -0.6661 -0.0336  0.6150  5.1784
##
## Random effects:
## Groups Name Variance Std.Dev.
## SchoolId (Intercept) 1157 34.01
## Residual 1947 44.12
## Number of obs: 10552, groups: SchoolId, 344
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept) 334.4786    2.1227 157.569
## Male        12.3652    0.9687  12.765
## ESCS         6.5690    0.5695  11.534
## ANXMAT      -9.2020    0.4232 -21.745
## TEACHSUP     0.6389    0.4045   1.579
## REPEAT     -25.5711    1.5718 -16.269

```

```

## IMMIG      -10.3177      1.4610  -7.062
## PRIVATESCH -10.5015      5.4846  -1.915
##
## Correlation of Fixed Effects:
##           (Intr) Male   ESCS   ANXMAT TEACHS REPEAT IMMIG
## Male      -0.222
## ESCS       0.001 -0.059
## ANXMAT     -0.026  0.133  0.046
## TEACHSUP   -0.001  0.013  0.022  0.139
## REPEAT     -0.057 -0.067  0.081 -0.003 -0.023
## IMMIG      -0.081  0.025  0.270 -0.050 -0.015 -0.124
## PRIVATESCH -0.361 -0.011 -0.011  0.005 -0.010 -0.005  0.000

## [1] "Bayesian mixed-effects model summary:"

## Inference for Bugs model at "6", fit using jags,
## 1 chains, each with 10000 iterations (first 1000 discarded), n.thin = 10
## n.sims = 900 iterations saved. Running time = 322.3 secs
##           mu.vect sd.vect      2.5%      25%      50%      75%
## beta0      332.929  2.270    328.519    331.467    332.922    334.373
## beta1       12.554  0.952     10.668     11.943     12.538     13.211
## beta2        6.559  0.579      5.477      6.166      6.562      6.947
## beta3       -9.212  0.422    -10.016     -9.489     -9.229     -8.943
## beta4        0.616  0.402     -0.178      0.351      0.612      0.868
## beta5      -25.447  1.591    -28.514    -26.517    -25.480    -24.367
## beta6      -10.210  1.441    -13.089    -11.265    -10.213     -9.207
## beta7       -8.783  5.524    -19.553    -12.512     -8.845     -5.233
## sigma      44.108  0.309     43.501     43.911     44.102     44.326
## sigma_u     34.101  1.431     31.394     33.093     34.061     35.059
## deviance 109865.331 27.402 109810.933 109847.552 109865.950 109883.081
##           97.5%
## beta0      337.406
## beta1       14.354
## beta2        7.693
## beta3       -8.377
## beta4        1.415
## beta5      -22.204
## beta6       -7.438
## beta7        2.454
## sigma      44.709
## sigma_u     36.966
## deviance 109921.404
##
## DIC info (using the rule: pV = var(deviance)/2)
## pV = 375.4 and DIC = 110240.8
## DIC is an estimate of expected predictive error (lower deviance is better).

```

Conclusion

The primary goal of this study was to investigate how much of the variability in math scores among Italian students can be attributed to differences between schools compared to individual factors such as socioeconomic status (ESCS), math anxiety (ANXMAT), and gender (Male). By employing a hierarchical Bayesian mixed-effects model and comparing it with frequentist approaches, we were able to rigorously address this question while gaining a deeper understanding of the factors influencing student performance.

This analysis revealed that a significant portion of the variability in math scores is attributable to differences between schools, as indicated by the school-level random effects. This highlights the importance of school-level factors in shaping educational outcomes, even after accounting for individual-level predictors.

However, individual factors also play a critical role in explaining within-school variability. Among these, ESCS emerged as the most influential predictor, affirming the positive relationship between socioeconomic background and academic performance. Conversely, math anxiety showed a strong negative association with math scores, indicating the detrimental impact of psychological factors on academic achievement. Gender differences were also evident, with male students outperforming their female peers.

In addition to these findings, the inclusion of other predictors such as teacher support (TEACHSUP) and school type (PRIVATE SCH) provided nuanced insights, though their effects were weaker and less consistent across models. The mixed-effects models, both frequentist and Bayesian, outperformed the simple linear regression by accounting for the nested structure of the data, which is critical in educational research. The Bayesian model, in particular, offered probabilistic interpretations and credible intervals, enabling more robust and intuitive conclusions.

In answering the research question, we conclude that while individual factors like socioeconomic status and math anxiety are key drivers of student performance, school-level differences account for a substantial proportion of variability in math scores. This underscores the importance of targeted policies that address both individual and institutional disparities to improve educational outcomes. Future research could explore specific school-level characteristics (e.g., resources, teaching quality) to better understand the mechanisms driving between-school differences.

This study demonstrates the power of hierarchical models in educational research and highlights the value of Bayesian methods in providing richer inferential insights, particularly in the presence of complex, nested data structures.

R Code and Models Outputs

```
## FIRST DATA EXTRACTION FROM THE PISA22 DATASET STU_Q_22
data <- data.frame(STU_Q_22$CNT, STU_Q_22$CNTSCHID , # Country and schoolId
  STU_Q_22$ST004D01T, # Gender
  STU_Q_22$PV1MATH, STU_Q_22$PV2MATH, # plausible values for the math
  STU_Q_22$PV3MATH, STU_Q_22$PV4MATH, STU_Q_22$PV5MATH,
  STU_Q_22$PV6MATH, STU_Q_22$PV7MATH, STU_Q_22$PV8MATH,
  STU_Q_22$PV9MATH, STU_Q_22$PV10MATH,
  STU_Q_22$ESCS, # ESCS index
  STU_Q_22$ANXMAT, # ANXMAT index
  STU_Q_22$FAMSUP, # FAMSUP index
  STU_Q_22$TEACHSUP, # TEACHSUP index
  STU_Q_22$IMMIG,
  STU_Q_22$STRATUM, # stratum
  STU_Q_22$REPEAT # if the student has repeated at least 1 year of school
)

# To add the PRIVATESCH to the dataset
sch_q_subset <- CY08MSP_SCH_QQ[, c("CNTSCHID", "PRIVATESCH")]
# Merging the data
data <- merge(data, sch_q_subset, by.x = "STU_Q_22.CNTSCHID", by.y = "CNTSCHID",
  all.x = TRUE)

# RENAME AND RCODIFY THE VARIABLES

# Country
colnames(data)[colnames(data) == "STU_Q_22.CNT"] <- "Country"

# Gender
colnames(data)[colnames(data) == "STU_Q_22.ST004D01T"] <- "Gender"

# Male binary variable
data$Male <- ifelse(data$Gender==1,0,ifelse(data$Gender==2,1,NA))
data$Male <- as.factor(data$Male)

# SchoolId
colnames(data)[colnames(data) == "STU_Q_22.CNTSCHID"] <- "SchoolId"

# Plausible Values
colnames(data)[colnames(data) == "STU_Q_22.PV1MATH"] <- "PV1"
colnames(data)[colnames(data) == "STU_Q_22.PV2MATH"] <- "PV2"
colnames(data)[colnames(data) == "STU_Q_22.PV3MATH"] <- "PV3"
colnames(data)[colnames(data) == "STU_Q_22.PV4MATH"] <- "PV4"
colnames(data)[colnames(data) == "STU_Q_22.PV5MATH"] <- "PV5"
```

```

colnames(data)[colnames(data) == "STU_Q_22.PV6MATH"] <- "PV6"
colnames(data)[colnames(data) == "STU_Q_22.PV7MATH"] <- "PV7"
colnames(data)[colnames(data) == "STU_Q_22.PV8MATH"] <- "PV8"
colnames(data)[colnames(data) == "STU_Q_22.PV9MATH"] <- "PV9"
colnames(data)[colnames(data) == "STU_Q_22.PV10MATH"] <- "PV10"

# SCORE
data$SCORE <- as.numeric(rowMeans(data[, c(7,8,9,10,11,12,13,14,15,16)], na.rm = T))

# ANXMAT
colnames(data)[colnames(data) == "STU_Q_22.ANXMAT"] <- "Anxmat1"
data$ANXMAT <- ifelse(data$Anxmat1 > 5, NA, data$Anxmat1)

# FAMSUP
colnames(data)[colnames(data) == "STU_Q_22.FAMSUP"] <- "Famsup1"
data$FAMSUP <- ifelse(data$Famsup1 > 5, NA, data$Famsup1)

# TEACHSUP
colnames(data)[colnames(data) == "STU_Q_22.TEACHSUP"] <- "Teachsup1"
data$TEACHSUP <- ifelse(data$Teachsup1 > 5, NA, data$Teachsup1)

# REPEAT
colnames(data)[colnames(data) == "STU_Q_22.REPEAT"] <- "Repeat1"
data$REPEAT <- ifelse(data$Repeat1 == 0, 0,
                      ifelse(data$Repeat1 == 1, 1, NA))
data$REPEAT <- as.factor(data$REPEAT)

# IMMIG
colnames(data)[colnames(data) == "STU_Q_22.IMMIG"] <- "Immig1"
data$IMMIG <- ifelse(data$Immig1 == 1, 0,
                    ifelse(data$Immig1 == 2, 1,
                          ifelse(data$Immig1 == 3, 1, NA)))
data$IMMIG <- as.factor(data$IMMIG)

# STRATUM
colnames(data)[colnames(data) == "STU_Q_22.STRATUM"] <- "STRATUM"

# ESCS
colnames(data)[colnames(data) == "STU_Q_22.ESCS"] <- "ESCS"

# PRIVATESCH
colnames(data)[colnames(data) == "PRIVATESCH"] <- "PRIVATESCH1"
data$PRIVATESCH1 <- tolower(as.character(data$PRIVATESCH1))
data$PRIVATESCH <- ifelse(data$PRIVATESCH1 %in% c("private"), 1,
                          ifelse(data$PRIVATESCH1 %in% c("public"), 0, NA))

```

```

data$PRIVATESCH <- as.factor(data$PRIVATESCH)

# To save the new dataset
write.csv(data, "PISA.csv")

# FINAL DATA EXTRACTION (from the reduced dataset PISA.csv):
data <- data.frame(PISA$Country[PISA$Country=="ITA"], PISA$SchoolId[PISA$Country=="ITA"],
  PISA$SCORE[PISA$Country=="ITA"], PISA$Male[PISA$Country=="ITA"],
  PISA$ANXMAT[PISA$Country=="ITA"], PISA$ESCS[PISA$Country=="ITA"],
  PISA$TEACHSUP[PISA$Country=="ITA"], PISA$REPEAT[PISA$Country=="ITA"],
  PISA$IMMIG[PISA$Country=="ITA"], PISA$PRIVATESCH[PISA$Country=="ITA"])

# RENAME THE COLUMNS
colnames(data)[colnames(data) == "PISA.Country.PISA.Country.....ITA.."] <- "Country"
colnames(data)[colnames(data) == "PISA.SchoolId.PISA.Country.....ITA.."] <- "SchoolId"
colnames(data)[colnames(data) == "PISA.SCORE.PISA.Country.....ITA.."] <- "SCORE"
colnames(data)[colnames(data) == "PISA.Male.PISA.Country.....ITA.."] <- "Male"
colnames(data)[colnames(data) == "PISA.ANXMAT.PISA.Country.....ITA.."] <- "ANXMAT"
colnames(data)[colnames(data) == "PISA.ESCS.PISA.Country.....ITA.."] <- "ESCS"
colnames(data)[colnames(data) == "PISA.TEACHSUP.PISA.Country.....ITA.."] <- "TEACHSUP"
colnames(data)[colnames(data) == "PISA.REPEAT.PISA.Country.....ITA.."] <- "REPEAT"
colnames(data)[colnames(data) == "PISA.IMMIG.PISA.Country.....ITA.."] <- "IMMIG"
colnames(data)[colnames(data) == "PISA.PRIVATESCH.PISA.Country.....ITA.."] <- "PRIVATESCH"

# To save the final dataset to be used
write.csv(data, "data_ITA.csv")

## DATASET AND PACKAGES
setwd("C:/Users/Utente/OneDrive/Desktop/Bayesian Modelling/Final_Project")
library(knitr)
library(ggplot2)
library(mice)
library(patchwork)
library(tidyr)
library(rjags)
library(R2jags)
library(coda)
library(lme4)
library(readr)
data_ITA <- read_csv("data_ITA.csv")

## DESCRIPTIVE STATISTICS
# 1
# Mean score

```



```

mean_score <- mean(data_ITA$SCORE, na.rm = TRUE)
mean_score_table <- data.frame(
  Metric = "Mean Score",
  Value = mean_score
)
kable(mean_score_table, caption = "Mean Math Score")

ggplot(data_ITA, aes(x = SCORE)) +
  geom_histogram(binwidth = 20, fill = "steelblue", color = "black",
    alpha = 0.7) +
  labs(
    title = "Distribution of Math Scores",
    x = "Math Score",
    y = "Frequency"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10)
  )

# 2
# Table for Male
Male_table <- prop.table(table(data_ITA$Male))
kable(data.frame(
  Male = names(Male_table),
  Proportion = as.numeric(Male_table)
), caption = "Proportions of Gender",
col.names = c("Gender (0 = Female, 1 = Male)", "Proportion"))

mean_score_by_gender <- aggregate(SCORE ~ Male, data = data_ITA,
  FUN = mean, na.rm = TRUE)
kable(mean_score_by_gender,
  col.names = c("Gender (0 = Female, 1 = Male)", "Mean Score"),
  caption = "Mean Math Scores by Gender")

ggplot(data_ITA, aes(x = SCORE)) +
  geom_histogram(binwidth = 20, aes(fill = as.factor(Male)), color = "black",
    alpha = 0.7) +
  labs(
    title = "Distribution of Math Scores \n by Gender",
    x = "Math Score",
    y = "Frequency"
  )

```

```

) +
scale_fill_manual(
  values = c("0" = "pink", "1" = "steelblue"),
  labels = c("0" = "Female", "1" = "Male")
) +
facet_wrap(~ Male, labeller = labeller(Male = c(`0` = "Female", `1` = "Male")))) +
theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
  axis.title = element_text(size = 12),
  axis.text = element_text(size = 10),
  legend.position = "none"
)

# 3
# Table for PRIVATESCH
PRIVATESCH_table <- prop.table(table(data_ITA$PRIVATESCH))
kable(data.frame(
  PRIVATESCH = names(PRIVATESCH_table),
  Proportion = as.numeric(PRIVATESCH_table)
), caption = "Proportions of PRIVATESCH")

ggplot(data_ITA, aes(x = as.factor(PRIVATESCH), y = SCORE,
  fill = as.factor(PRIVATESCH))) +
  geom_boxplot(alpha = 0.7) +
  labs(
    title = "Math Scores by School Type",
    x = "School Type (0 = Public, 1 = Private)",
    y = "Math Score",
    fill = "School Type"
  ) +
  scale_fill_manual(values = c("0" = "purple", "1" = "gold")) +
  theme_minimal()

# NA
colSums(is.na(data_ITA))

imputation_methods <- c(
  "norm",      # ESCS: Continuous
  "norm",      # ANXMAT: Continuous
  "norm",      # TEACHSUP: Continuous
  "logreg",    # REPEAT: Binary
  "logreg"     # IMMIG: Binary
)

```

```

imputed_data <- mice(
  data = data_ITA[, c("ESCS", "ANXMAT", "TEACHSUP", "REPEAT", "IMMIG")],
  method = imputation_methods, # Bayesian normal imputation
  m = 5,                        # Number of multiple imputations
  maxit = 50,                   # Maximum number of iterations
  seed = 123,                   # For reproducibility
  printFlag = FALSE             # Suppress iterative output
)

# Check imputed data summary
summary(imputed_data)

data_complete <- complete(imputed_data, 1) # Choose the first imputed dataset

# Replace the original variables with imputed values
data_ITA$ESCS <- data_complete$ESCS
data_ITA$ANXMAT <- data_complete$ANXMAT
data_ITA$TEACHSUP <- data_complete$TEACHSUP
data_ITA$REPEAT <- data_complete$REPEAT
data_ITA$IMMIG <- data_complete$IMMIG
# Output interpretation:
# Predictor Matrix
# Defines which variables are used as predictors when imputing missing
# values for each variable

# Interpretation of Predictor Matrix
# Rows: The variable being imputed
# Columns: The variables used as predictors for imputation
# 0: The variable is not used as a predictor
# 1: The variable is used as a predictor

colSums(is.na(data_ITA))

# 4
# Table for IMMIG
immig_table <- prop.table(table(data_ITA$IMMIG))
kable(data.frame(
  IMMIG = names(immig_table),
  Proportion = as.numeric(immig_table)
), caption = "Proportions of IMMIG")

# Table for REPEAT
repeat_table <- prop.table(table(data_ITA$REPEAT))
kable(data.frame(
  REPEAT = names(repeat_table),

```

```

    Proportion = as.numeric(repeat_table)
  ), caption = "Proportions of REPEAT")

# Boxplot representations
p1 <- ggplot(data_ITA, aes(x = as.factor(IMMIG), y = SCORE,
                           fill = as.factor(IMMIG))) +
  geom_boxplot(alpha = 0.7) +
  labs(
    title = "Math Scores by Immigration Status",
    x = "Immigration Status (0 = Native, 1 = Immigrant)",
    y = "Math Score",
    fill = "Immigration Status"
  ) +
  scale_fill_manual(values = c("0" = "lightblue", "1" = "orange")) +
  theme_minimal()

p2 <- ggplot(data_ITA, aes(x = as.factor(REPEAT), y = SCORE,
                           fill = as.factor(REPEAT))) +
  geom_boxplot(alpha = 0.7) +
  labs(
    title = "Math Scores by Grade Repetition",
    x = "Repeated a Grade (0 = No, 1 = Yes)",
    y = "Math Score",
    fill = "Repeated Grade"
  ) +
  scale_fill_manual(values = c("0" = "darkred", "1" = "darkblue")) +
  theme_minimal()

combined_plot <- p1 + p2
combined_plot

# 5
data_long <- pivot_longer(
  data_ITA,
  cols = c(ESCS, ANXMAT, TEACHSUP), # columns to pivot
  names_to = "Index",              # column for variable names
  values_to = "Value"              # column for values
)

# Scatter plot with faceting
ggplot(data_long, aes(x = Value, y = SCORE, color = Index)) +
  geom_point(alpha = 0.6, size = 0.6) +
  geom_smooth(method = "lm", se = TRUE, linetype = "dashed", color = "black") +
  facet_wrap(~ Index, scales = "free_x") +

```

```

labs(
  title = "Math Scores in Relation to \n ESCS, ANXMAT, and TEACHSUP",
  x = "Index Value",
  y = "Math Score"
) +
scale_color_manual(values = c("ESCS" = "steelblue", "ANXMAT" = "purple",
                              "TEACHSUP" = "darkgreen")) +

theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
  axis.title = element_text(size = 12),
  axis.text = element_text(size = 10)
)

## MODELS

# Center the continuous predictors
data_ITA$ESCS <- as.vector(scale(data_ITA$ESCS, center = TRUE, scale = FALSE))
data_ITA$ANXMAT <- as.vector(scale(data_ITA$ANXMAT, center = TRUE, scale = FALSE))
data_ITA$TEACHSUP <- as.vector(scale(data_ITA$TEACHSUP, center = TRUE, scale = FALSE))

# JAGS MODEL 0
jags_data0 <- list(
  y = data_ITA$SCORE,                # Outcome variable
  school = as.numeric(as.factor(data_ITA$SchoolId)), # School grouping variable
  N = nrow(data_ITA),                # Number of students
  J = length(unique(data_ITA$SchoolId)) # Number of schools
)

model_code0 <- "
model {
  for (i in 1:N) {
    # Likelihood
    y[i] ~ dnorm(mu[i], tau)
    mu[i] <- beta0 + u[school[i]]

    # Posterior predictive distribution
    yrep[i] ~ dnorm(mu[i], tau) # Posterior predictions for y
  }

  # Random effects
  for (j in 1:J) {
    u[j] ~ dnorm(0, tau_u)
  }
}

```

```

# Priors for fixed effects
beta0 ~ dnorm(0, 0.001)

# Priors for variances
tau ~ dgamma(0.001, 0.001)
sigma <- 1 / sqrt(tau)

tau_u ~ dgamma(0.001, 0.001)
sigma_u <- 1 / sqrt(tau_u)
}
"

# Define parameters to monitor
params <- c("beta0", "beta1", "beta2", "sigma", "sigma_u")

# Run the JAGS model
set.seed(123)
jags_model0 <- jags(
  data = jags_data0,
  parameters.to.save = params,
  model.file = textConnection(model_code0),
  n.chains = 1,      # Number of chains
  n.iter = 10000,    # Total iterations
  n.burnin = 1000,   # Burn-in iterations
  n.thin = 10,       # Thinning factor
  DIC = TRUE         # Calculate DIC for model comparison
)

# JAGS MODEL 1
jags_data1 <- list(
  y = data_ITA$SCORE,      # Outcome variable
  x1 = data_ITA$Male,      # Predictor 1
  x2 = data_ITA$ESCS,      # Predictor 2
  x3 = data_ITA$ANXMAT,    # Predictor 2
  school = as.numeric(as.factor(data_ITA$SchoolId)), # School grouping variable
  N = nrow(data_ITA),      # Number of students
  J = length(unique(data_ITA$SchoolId)) # Number of schools
)

jags_data1$school <- as.numeric(as.factor(jags_data1$school))
jags_data1$J <- length(unique(jags_data1$school))

model_code1 <- "
model {
  for (i in 1:N) {
    # Likelihood

```

```

y[i] ~ dnorm(mu[i], tau)
mu[i] <- beta0 + beta1 * x1[i] + beta2 * x2[i] + beta3 * x3[i] + u[school[i]]

# Posterior predictive distribution
yrep[i] ~ dnorm(mu[i], tau) # Posterior predictions for y
}

# Random effects for schools
for (j in 1:J) {
  u[j] ~ dnorm(0, tau_u)
}

# Priors for fixed effects
beta0 ~ dnorm(0, 0.001)
beta1 ~ dnorm(0, 0.001)
beta2 ~ dnorm(0, 0.001)
beta3 ~ dnorm(0, 0.001)

# Priors for variances
tau ~ dgamma(0.001, 0.001)
sigma <- 1 / sqrt(tau)

tau_u ~ dgamma(0.001, 0.001)
sigma_u <- 1 / sqrt(tau_u)
}
"

# Define parameters to monitor
params <- c("beta0", "beta1", "beta2", "beta3", "sigma", "sigma_u")

set.seed(123)
jags_model1 <- jags(
  data = jags_data1,
  parameters.to.save = params,
  model.file = textConnection(model_code1),
  n.chains = 1,      # Number of chains
  n.iter = 10000,    # Total iterations
  n.burnin = 1000,   # Burn-in iterations
  n.thin = 10,       # Thinning factor
  DIC = TRUE         # Calculate DIC for model comparison
)

# JAGS MODEL 2
jags_data2 <- list(
  y = data_ITA$SCORE,      # Outcome variable

```

```

x1 = data_ITA$Male,           # Predictor 1
x2 = data_ITA$ESCS,          # Predictor 2
x3 = data_ITA$ANXMAT,        # Predictor 3
x4 = data_ITA$TEACHSUP,       # Predictor 4
x5 = data_ITA$REPEAT,        # Predictor 5
x6 = data_ITA$IMMIG,         # Predictor 6
x7 = data_ITA$PRIVATESCH,     # Predictor 7
school = as.numeric(as.factor(data_ITA$SchoolId)), # School grouping variable
N = nrow(data_ITA),           # Number of students
J = length(unique(data_ITA$SchoolId)) # Number of schools
)

jags_data2$school <- as.numeric(as.factor(jags_data2$school))
jags_data2$J <- length(unique(jags_data2$school))

model_code2 <- "
model {
  # Random effects for schools
  for (j in 1:J) {
    u[j] ~ dnorm(0, tau_u)
  }

  # Priors for fixed effects
  beta0 ~ dnorm(0, 0.001)
  beta1 ~ dnorm(0, 0.001)
  beta2 ~ dnorm(0, 0.001)
  beta3 ~ dnorm(0, 0.001)
  beta4 ~ dnorm(0, 0.001)
  beta5 ~ dnorm(0, 0.001)
  beta6 ~ dnorm(0, 0.001)
  beta7 ~ dnorm(0, 0.001)

  # Priors for variances
  tau ~ dgamma(0.001, 0.001)
  sigma <- 1 / sqrt(tau)

  tau_u ~ dgamma(0.001, 0.001)
  sigma_u <- 1 / sqrt(tau_u)

  for (i in 1:N) {
    # Likelihood
    y[i] ~ dnorm(mu[i], tau)
    mu[i] <- beta0 + beta1 * x1[i] + beta2 * x2[i] + beta3 * x3[i]
      + beta4 * x4[i] + beta5 * x5[i] + beta6 * x6[i] + beta7 * x7[i] + u[school[i]]
  }
}

```



```

    # Posterior predictive distribution
    yrep[i] ~ dnorm(mu[i], tau) # Posterior predictions for y
  }
}
"

# Define parameters to monitor
params <- c("beta0", "beta1", "beta2", "beta3", "beta4", "beta5",
            "beta6", "beta7", "sigma", "sigma_u")
set.seed(123)
jags_model2 <- jags(
  data = jags_data2,
  parameters.to.save = params,
  model.file = textConnection(model_code2),
  n.chains = 1,      # Number of chains
  n.iter = 10000,    # Total iterations
  n.burnin = 1000,   # Burn-in iterations
  n.thin = 10,       # Thinning factor
  DIC = TRUE         # Calculate DIC for model comparison
)

# JAGS 2 WITH MU AND U FOR DIAGNOSTICS
# Define parameters to monitor
params <- c("beta0", "beta1", "beta2", "beta3", "beta4", "beta5",
            "beta6", "beta7", "sigma", "sigma_u", "yrep", "mu", "u")

set.seed(123)
jags_model2mu <- jags(
  data = jags_data2,
  parameters.to.save = params,
  model.file = textConnection(model_code2),
  n.chains = 1,      # Number of chains
  n.iter = 10000,    # Total iterations
  n.burnin = 1000,   # Burn-in iterations
  n.thin = 10,       # Thinning factor
  DIC = TRUE         # Calculate DIC for model comparison
)

# JAGS 2 MODEL Without Random Effects
jags_data_no_random <- list(
  y = data_ITA$SCORE,      # Outcome variable
  x1 = data_ITA$Male,      # Predictor 1
  x2 = data_ITA$ESCS,      # Predictor 2
  x3 = data_ITA$ANXMAT,    # Predictor 3

```

```

x4 = data_ITA$TEACHSUP,      # Predictor 4
x5 = data_ITA$REPEAT,       # Predictor 5
x6 = data_ITA$IMMIG,        # Predictor 6
x7 = data_ITA$PRIVATESCH,   # Predictor 7
N = nrow(data_ITA)          # Number of students
)

model_code_no_random <- "
model {
  # Priors for fixed effects
  beta0 ~ dnorm(0, 0.001)
  beta1 ~ dnorm(0, 0.001)
  beta2 ~ dnorm(0, 0.001)
  beta3 ~ dnorm(0, 0.001)
  beta4 ~ dnorm(0, 0.001)
  beta5 ~ dnorm(0, 0.001)
  beta6 ~ dnorm(0, 0.001)
  beta7 ~ dnorm(0, 0.001)

  # Prior for variance
  tau ~ dgamma(0.001, 0.001)
  sigma <- 1 / sqrt(tau)

  for (i in 1:N) {
    # Likelihood
    y[i] ~ dnorm(mu[i], tau)
    mu[i] <- beta0 + beta1 * x1[i] + beta2 * x2[i] + beta3 * x3[i] +
              beta4 * x4[i] + beta5 * x5[i] + beta6 * x6[i] + beta7 * x7[i]
  }
}
"

# Define parameters to monitor
params_no_random <- c("beta0", "beta1", "beta2", "beta3", "beta4", "beta5",
                      "beta6", "beta7", "sigma")

# Run the JAGS model
set.seed(123)
jags_model_no_random <- jags(
  data = jags_data_no_random,
  parameters.to.save = params_no_random,
  model.file = textConnection(model_code_no_random),
  n.chains = 1,      # Number of chains
  n.iter = 10000,    # Total iterations
  n.burnin = 1000,   # Burn-in iterations

```

```

n.thin = 10,      # Thinning factor
DIC = TRUE       # Calculate DIC for model comparison
)

# DIC
dic_values <- c(
  model1 = jags_model0$BUGSoutput$DIC,
  model2 = jags_model1$BUGSoutput$DIC,
  model3 = jags_model2$BUGSoutput$DIC,
  model4 = jags_model_no_random$BUGSoutput$DIC
)

dic_table <- data.frame(
  Model = names(dic_values),
  DIC = as.numeric(dic_values)
)

kable(dic_table, caption = "DIC Values for Different Models")

# PRINTING MODEL 3 --> CHOSEN ONE
print(jags_model2)

## DIAGNOSTIC

# TRACEPLOTS FOR CONVERGENCE
mcmc_samples <- as.mcmc(jags_model2mu$BUGSoutput$sims.matrix)

par(mfrow=c(2,4))
traceplot(mcmc_samples[, c("beta0", "beta1", "beta2", "beta3",
                           "beta4", "beta5", "beta6", "beta7")])

# POSTERIOR PREDICTIVE
posterior_pred <- jags_model2mu$BUGSoutput$sims.list$yrep
observed <- jags_data2$y
hist(observed, breaks = 30, col = rgb(0, 0, 1, 0.5),
     main = "Posterior Predictive Check",
     xlab = "Observed Math Scores", prob = TRUE)
lines(density(as.vector(posterior_pred)), col = "red", lwd = 2)
legend("topright", legend = c("Observed", "Posterior Predictive"),
     fill = c("blue", "red"))

# NORMALITY AND HOMOSCHEDASTICITY
# Residuals
predicted_values <- apply(jags_model2mu$BUGSoutput$sims.list$mu, 2, mean)
residuals <- jags_data1$y - predicted_values

```

```

# Residuals vs Predicted values
par(mfrow=c(1,2))
plot(predicted_values, residuals, main = "Residuals vs Predicted",
      xlab = "Predicted Values", ylab = "Residuals", pch = 19,
      col = rgb(0, 0, 1, 0.5), cex = 0.6)
abline(h = 0, col = "red", lwd = 2)

# QQ plot
qqnorm(residuals, main = "Normal Q-Q Plot \n for Residuals")
qqline(residuals, col = "red")

# RANDOM EFFECTS
# Hist
random_effects, echo=FALSE, fig.height= 3}
random_effects <- apply(jags_model2mu$BUGSoutput$sims.list$u, 2, mean)
hist(random_effects, breaks = 20, col = "lightblue",
      main = "Random Effects Distribution", xlab = "School Effects")

# Caterpillar
# Extract random effects means and credible intervals
random_effects <- apply(jags_model2mu$BUGSoutput$sims.list$u, 2, mean) # Mean of r. eff.
lower_bounds <- apply(jags_model2mu$BUGSoutput$sims.list$u, 2,
                      function(x) quantile(x, 0.025)) # 2.5% quantile
upper_bounds <- apply(jags_model2mu$BUGSoutput$sims.list$u, 2,
                      function(x) quantile(x, 0.975)) # 97.5% quantile

# Dataframe for ggplot
random_effects_df <- data.frame(
  School = 1:length(random_effects),
  Effect = random_effects,
  Lower = lower_bounds,
  Upper = upper_bounds
)

# Order the schools by effect size
random_effects_df <- random_effects_df[order(random_effects_df$Effect), ]
random_effects_df$School <- 1:nrow(random_effects_df) # Reorder School IDs

ggplot(random_effects_df, aes(x = School, y = Effect)) +
  geom_point(color = "blue") +
  geom_errorbar(aes(ymin = Lower, ymax = Upper), width = 0.2, color = "gray") +
  labs(
    title = "Caterpillar Plot of School-Level Random Effects",
    x = "School (Ordered by Effect Size)",
    y = "Random Effect"
  )

```

```

) +
theme_minimal() +
theme(
  plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
  axis.title = element_text(size = 12),
  axis.text = element_text(size = 10)
)

# EFFECTIVE SAMPLE SIZE (ESS)
head(effectiveSize(mcmc_samples), 10) # Show the first 10 rows

## FREQUENTIST VS BAYESIAN
# Simple linear regression
simple_lm <- lm(SCORE ~ Male + ESCS + ANXMAT + TEACHSUP + REPEAT + IMMIG
               + PRIVATESCH, data = data_ITA)
print("Linear model summary:")
summary(simple_lm)

# Frequentist mixed-effects
frequentist_model <- lmer(SCORE ~ Male + ESCS + ANXMAT + TEACHSUP + REPEAT
                          + IMMIG + PRIVATESCH + (1 | SchoolId), data = data_ITA)
print("Frequentist mixed-effects model summary:")
summary(frequentist_model)

# JAGS " (Model 3)
print("Bayesian mixed-effects model summary:")
print(jags_model2)

print(jags_model0)

## Inference for Bugs model at "4", fit using jags,
## 1 chains, each with 10000 iterations (first 1000 discarded), n.thin = 10
## n.sims = 900 iterations saved. Running time = 183.79 secs
##      mu.vect sd.vect      2.5%      25%      50%      75%
## beta0      332.693   2.233   328.316   331.214   332.711   334.206
## sigma       46.901   0.316   46.283   46.696   46.896   47.114
## sigma_u     39.305   1.700   36.182   38.113   39.197   40.369
## deviance 111153.414 26.595 111104.247 111134.495 111153.487 111171.805
##           97.5%
## beta0      337.293
## sigma       47.544
## sigma_u     42.852
## deviance 111206.897
##
## DIC info (using the rule: pV = var(deviance)/2)

```

```
## pV = 353.7 and DIC = 111507.1
## DIC is an estimate of expected predictive error (lower deviance is better).
```

```
print(jags_model1)
```

```
## Inference for Bugs model at "5", fit using jags,
## 1 chains, each with 10000 iterations (first 1000 discarded), n.thin = 10
## n.sims = 900 iterations saved. Running time = 301.78 secs
##      mu.vect sd.vect      2.5%      25%      50%      75%
## beta0      327.902   2.075    324.017    326.449    327.969    329.267
## beta1       11.610   0.971     9.688     10.992     11.576     12.248
## beta2        8.713   0.546     7.678     8.334     8.713     9.075
## beta3       -9.457   0.426    -10.268    -9.767    -9.452    -9.166
## sigma      44.801   0.327    44.197    44.566    44.799    45.025
## sigma_u     35.449   1.498    32.629    34.389    35.408    36.436
## deviance 110195.795 27.276 110140.598 110178.573 110194.883 110214.783
##      97.5%
## beta0      331.909
## beta1       13.531
## beta2        9.767
## beta3       -8.616
## sigma      45.467
## sigma_u     38.341
## deviance 110248.780
##
## DIC info (using the rule: pV = var(deviance)/2)
## pV = 372.0 and DIC = 110567.8
## DIC is an estimate of expected predictive error (lower deviance is better).
```

```
print(jags_model2)
```

```
## Inference for Bugs model at "6", fit using jags,
## 1 chains, each with 10000 iterations (first 1000 discarded), n.thin = 10
## n.sims = 900 iterations saved. Running time = 322.3 secs
##      mu.vect sd.vect      2.5%      25%      50%      75%
## beta0      332.929   2.270    328.519    331.467    332.922    334.373
## beta1       12.554   0.952    10.668    11.943    12.538    13.211
## beta2        6.559   0.579     5.477     6.166     6.562     6.947
## beta3       -9.212   0.422    -10.016    -9.489    -9.229    -8.943
## beta4        0.616   0.402     -0.178     0.351     0.612     0.868
## beta5      -25.447   1.591    -28.514    -26.517    -25.480    -24.367
## beta6      -10.210   1.441    -13.089    -11.265    -10.213    -9.207
## beta7       -8.783   5.524    -19.553    -12.512    -8.845    -5.233
## sigma      44.108   0.309    43.501    43.911    44.102    44.326
```

```
## sigma_u      34.101   1.431    31.394    33.093    34.061    35.059
## deviance 109865.331 27.402 109810.933 109847.552 109865.950 109883.081
##              97.5%
## beta0        337.406
## beta1         14.354
## beta2          7.693
## beta3         -8.377
## beta4          1.415
## beta5        -22.204
## beta6         -7.438
## beta7          2.454
## sigma        44.709
## sigma_u      36.966
## deviance 109921.404
##
## DIC info (using the rule: pV = var(deviance)/2)
## pV = 375.4 and DIC = 110240.8
## DIC is an estimate of expected predictive error (lower deviance is better).
```

```
print(jags_model_no_random)
```

```
## Inference for Bugs model at "8", fit using jags,
## 1 chains, each with 10000 iterations (first 1000 discarded), n.thin = 10
## n.sims = 900 iterations saved. Running time = 197.42 secs
##      mu.vect sd.vect      2.5%      25%      50%      75%
## beta0    337.582  0.784   336.015   337.031   337.615   338.150
## beta1     13.045  1.062    10.914    12.334    13.073    13.784
## beta2     19.771  0.620    18.582    19.344    19.768    20.175
## beta3    -10.050  0.512   -11.042   -10.398   -10.060   -9.707
## beta4      0.081  0.489    -0.879    -0.258     0.075     0.411
## beta5    -35.736  1.787   -39.109   -37.001   -35.718   -34.532
## beta6     -1.460  1.701    -4.664    -2.671    -1.464    -0.310
## beta7    -16.046  1.746   -19.446   -17.265   -16.046   -14.799
## sigma     53.685  0.367    52.972    53.440    53.692    53.941
## deviance 114002.616 4.149 113996.274 113999.549 114001.921 114005.192
##              97.5%
## beta0    339.045
## beta1     15.097
## beta2     20.948
## beta3     -9.043
## beta4      1.038
## beta5    -32.036
## beta6      1.901
## beta7    -12.594
## sigma     54.382
```

```
## deviance 114012.199
##
## DIC info (using the rule:  $pV = \text{var}(\text{deviance})/2$ )
##  $pV = 8.6$  and  $DIC = 114011.2$ 
## DIC is an estimate of expected predictive error (lower deviance is better).
```