Predictors Of Student Performance

— Project Report —
Applied Bayesian Data Analysis

Syed Hamza Afzal Ashraf

Zeynep Beyza Aktepe

March 13, 2025

TU Dortmund University

Contents

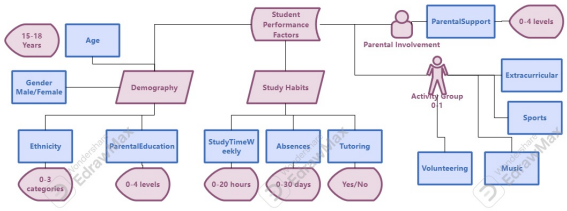

Figure 1: Factors influencing the student performance.

1 Introduction

In the dynamic and ever-evolving landscape of education, identifying the key factors that influence student performance is essential for developing effective learning strategies and fostering academic success. From study habits and classroom engagement to parental support and socio-economic background, a multitude of elements play a pivotal role in shaping student outcomes. Our project, *Predictors of Student Performance*, seeks to address the critical research question: *What are the predictors of student performance levels across subjects (Students) when performance (measured by Grade or GPA) is categorized as both ordinal and continuous?* By exploring the relationships between various predictors and academic performance, this study aims to uncover insights that can inform targeted interventions and enhance the classification of student achievement.

The motivation behind this study arises from the increasing use of data-driven decision-making in education. As schools and policymakers strive to improve outcomes of students performance, identifying key performance factors becomes of vital importance. Factors such as study time, parental education, extracurricular involvement, and attendance significantly impact a student's academic trajectory. By analyzing these relationships, educators and administrators can implement targeted interventions to support students effectively and foster a more inclusive learning environment.

To tackle this challenge, we employ analytical and Bayesian modeling techniques. Bayesian methods provide a structured approach to handling uncertainty while incorporating prior knowledge, making them particularly suitable for educational research. By utilizing Bayesian analysis, we can estimate the impact of different predictors on student performance while accounting for potential confounding variables. This approach not only quantifies the influence of key factors but also provides credible intervals to assess the reliability of our findings.

As illustrated in Figure 1, the factors influencing student performance can be broadly categorized into Demography, Study Habits, Activity Group, and Parental Involvement. In this project, we focus on identifying which of these factors are statistically significant predictors of academic performance. Through this analysis, we aim to determine which factors are most influential in predicting student performance and discuss their implications.

In Section 2, we provide details on data pre-processing, features and prior selection. Section 3 explores classification models, focusing on Bayesian approaches to predict student performance levels. Then, in Section 4, we assess model performance and convergence using key statistical measures. Additionally, Section 6 discusses the constraints of our analysis and opportunities for future improvements. Finally, Section 5 highlights the practical implications of our findings for educational strategies, while Section 7 presents our key takeaways from the project.

2 Data

The dataset used in this study was sourced from Kaggle and it is readily accessible. The dataset consists of 2392 observations and 15 variables. It contains several attributes related to students, capturing demographic details, study habits, parental involvement, and participation in extracurricular activities. These factors were considered as potential predictors of student performance levels. As independent variables, the dataset contains StudentID, Age, Gender, Ethnicity, ParentalEducation, StudyTimeWeekly, Absences, Tutoring, ParentalSupport, Extracurricular, Sports, Music and Volunteering. The potential target variables in the dataset are both GPA and GradeClass. As the dataset does not have any missing values, data clean-up was skipped.

This synthetic dataset which is licensed by Attribution 4.0 International (CC BY 4.0), was created approximately nine months ago, has not yet been utilized for Bayesian modeling on Kaggle, although a few machine learning and exploratory data analysis notebooks are available but no Bayesian model is applied. This presents a unique opportunity to apply Bayesian regression techniques to uncover new insights into the predictors of student performance.

2.1 Data Processing

For this analysis, the dataset was loaded into **R**, and `GradeClass` was converted into an ordered factor to reflect its ordinal nature. Additionally, various preprocessing and feature engineering steps were implemented to enhance the data for **clustering** and **Bayesian regression modeling**.

2.2 Target Variable

We consider two representations of student performance, as we predict two models for comparison (Continuous and Ordinal): The classification of `GradeClass` based on GPA thresholds is defined as follows:

$$\begin{aligned}
&\textbf{A} \quad (3.5 \leq \text{GPA} \leq 4.0) \\
&\textbf{B} \quad (3.0 \leq \text{GPA} < 3.5) \\
&\textbf{C} \quad (2.5 \leq \text{GPA} < 3.0) \\
&\textbf{D} \quad (2.0 \leq \text{GPA} < 2.5) \\
&\textbf{F} \quad (\text{GPA} < 2.0)
\end{aligned}$$

1. **Ordinal Variable**:
   - `GradeClass` (A, B, C, D, F) where A being the highest and F being the lowest.
2. **Continuous Variable**:
   - `GPA` (0–4 scale) 4 being the highest and 0 being the lowest.

The target variable `GradeClass` can be derived from the target variable `GPA` using predefined GPA ranges. However, the reverse is not entirely possible; only the range of `GPA` can be inferred from `GradeClass`, as `GradeClass` represents a categorical classification of the continuous `GPA` values.

This structured classification ensures that the analysis appropriately captures student performance levels for further modeling and interpretation.

2.3 Data Filtering

To standardize the variables `StudyTimeWeekly` and `Absences`, we apply z-score normalization:

$$\text{StudyTimeWeekly}_z = \frac{\text{StudyTimeWeekly} - \mu_{\text{StudyTimeWeekly}}}{\sigma_{\text{StudyTimeWeekly}}} \tag{1}$$

$$\text{Absences}_z = \frac{\text{Absences} - \mu_{\text{Absences}}}{\sigma_{\text{Absences}}} \tag{2}$$

where:

- $\mu_{\text{StudyTimeWeekly}}$ and $\sigma_{\text{StudyTimeWeekly}}$ are the mean and standard deviation of `StudyTimeWeekly`, respectively.
- μ_{Absences} and σ_{Absences} are the mean and standard deviation of `Absences`, respectively.

2.4 Clustering and Prior Selection

2.4.1 Hierarchical Clustering

We perform hierarchical clustering using Ward's method. First, we compute the Euclidean distance between each pair of students based on their study habits (StudyTimeWeekly, Absences) or Activity Group (Music, Volunteering, Sports and Extracurricular).

Next, we apply Ward's linkage criterion, which minimizes the increase in total within-cluster variance when merging clusters

To explore patterns in hierarchical clustering was performed using StudyTimeWeekly and Absences as key variables for study habits whereas for activity group, Volunteering, Music, Sports and Extracurricular were used as as key variables. Standardization was applied to these variables to ensure comparability.

2.4.2 Prior Selection

To incorporate prior knowledge, we define two sets of priors:

Narrow Priors (More Informative)

$$\beta_j \sim \mathcal{N}(0, 1), \quad u_{\text{Cluster\_HC}} \sim \text{Cauchy}(0, 1), \quad \text{Intercept} \sim t_3(0, 2) \tag{3}$$

Broad Priors (Weakly Informative)

$$\beta_j \sim \mathcal{N}(0, 10), \quad u_{\text{Cluster\_HC}} \sim \text{Cauchy}(0, 1), \quad \text{Intercept} \sim t_3(0, 2) \tag{4}$$

where:

- $\mathcal{N}(\mu, \sigma)$ is the normal distribution with mean μ and standard deviation σ.
- $\text{Cauchy}(0, 1)$ is a weakly informative prior for random effects.
- $t_3(0, 2)$ is a Student-t distribution for the intercept with 3 degrees of freedom.

These priors reflect different levels of belief strength about the regression coefficients before observing the data.

2.4.3 Model Formula

The equation below is for the Cluster Activity Group.

$$\begin{aligned} \text{GPA} = {} & \beta_0 + \beta_1 \text{Gender} + \beta_2 \text{ParentalEducation} + \beta_3 \text{ParentalSupport} \\ & + \beta_4 \text{StudyTimeWeekly} + \beta_5 \text{Absences} + \beta_6 \text{Tutoring} \\ & + \beta_7 \text{Ethnicity} + u_{\text{Cluster\_HC}} \end{aligned} \tag{5}$$

The equation below is for the Cluster StudyHabits:

$$\begin{aligned} \text{GPA} = {} & \beta_0 + \beta_1 \text{Gender} + \beta_2 \text{ParentalEducation} + \beta_3 \text{ParentalSupport} \\ & + \beta_4 \text{Extracurricular} + \beta_5 \text{Sports} + \beta_6 \text{Music} + \beta_7 \text{Volunteering} \\ & + \beta_8 \text{Ethnicity} + u_{\text{Cluster\_HC}} \end{aligned} \tag{6}$$

2.4.4 Implementation Details

In Section 2.2, we discuss the usage of the target variable. Here, we first convert it from Nominal to Ordinal. We then perform clustering process, as mentioned in Section 2.4.1.and then apply priors as shown in Equations (4) and (3). Equation (5) represents the first model that incorporates study habits, whereas Equation (6) extends the analysis by incorporating extracurricular activities. Furthermore, results of the implementation will be discussed later in the report in 3 section.

3 Models

We begin with applying clustering and then performing prior sensitivity analysis. After selecting the priors and clusters we proceed with comparisons of Ordinal and Continuous models. This analysis was conducted using Bayesian multilevel modeling in BRMS Bürkner, 2017, which interfaces with Stan Carpenter et al., 2017. In the project, we kept 2000 iterations for warmup and 2000 iterations for inference, having 4 chains and used 8 cores.

BRMS book Bürkner, 2024 provides an in-depth discussion on Bayesian regression models. The residual plots and posterior predictive checks presented in this report were also motivated by the methodologies outlined in this book.

3.1 Clustering and Prior Selection

In this section we discuss the formation of clusters and prior selection when target variable is set as GPA (Continuous). We first apply hierarchical clustering and then decide which K is suitable by the help of WSS along with silhouette score.

3.1.1 Study Habit Group

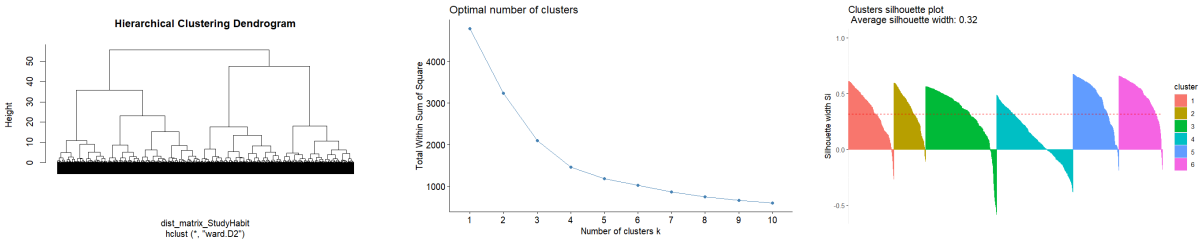

Figure 2: Dendogram,WSS and silhouette representations of study habits clustering.

In Figure 2, the leftmost image illustrates the dendorgram. By observing the plot, we estimate that approximately four clusters are formed when setting the cutoff height around 25. The significant height differences in the dendorgram indicate strong separations between certain student groups. To determine the optimal number of clusters, we apply

the Within-Cluster Sum of Squares (WSS) method. Using the Elbow method, as shown in the middle image of Figure 2, we observe that an optimal cutoff height is around 6.

The rightmost image in Figure 2 presents the silhouette score, which ranges between -1 and 1, where higher values indicate better clustering quality. While the silhouette scores are generally positive, some values approach -0.5, suggesting that certain data points might be misclassified.

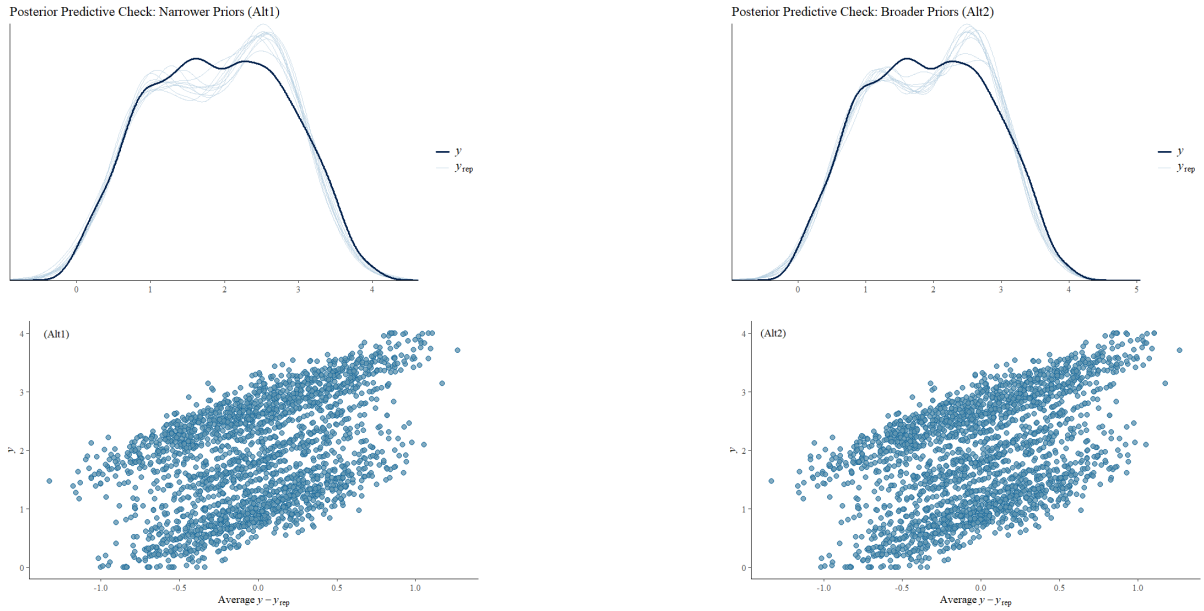

Figure 3: Posterior Predictive Checks (PPC) and Residuals for Study Habit Model. The first row shows PPC results, while the second row presents residual diagnostics.

In Figure 3, we assess the model performance after applying priors to Equation (5). The residual plots (second row in Figure 3) suggest poor model performance, indicating that the residuals do not follow the expected patterns. Additionally, the posterior predictive checks further confirm that the model is not a good fit, highlighting the need for potential refinements.

3.1.2 Activity Group

Just as section 3.1.1, In Figure 4, the first image shows a dendogram for which we try and identify which cutoff should be better and using WSS and silhouette score for k=5 and k=6, we find out that k=5 is a better cutoff height as it doesn't have any data points that might be a misclassification i.e no negative scores.

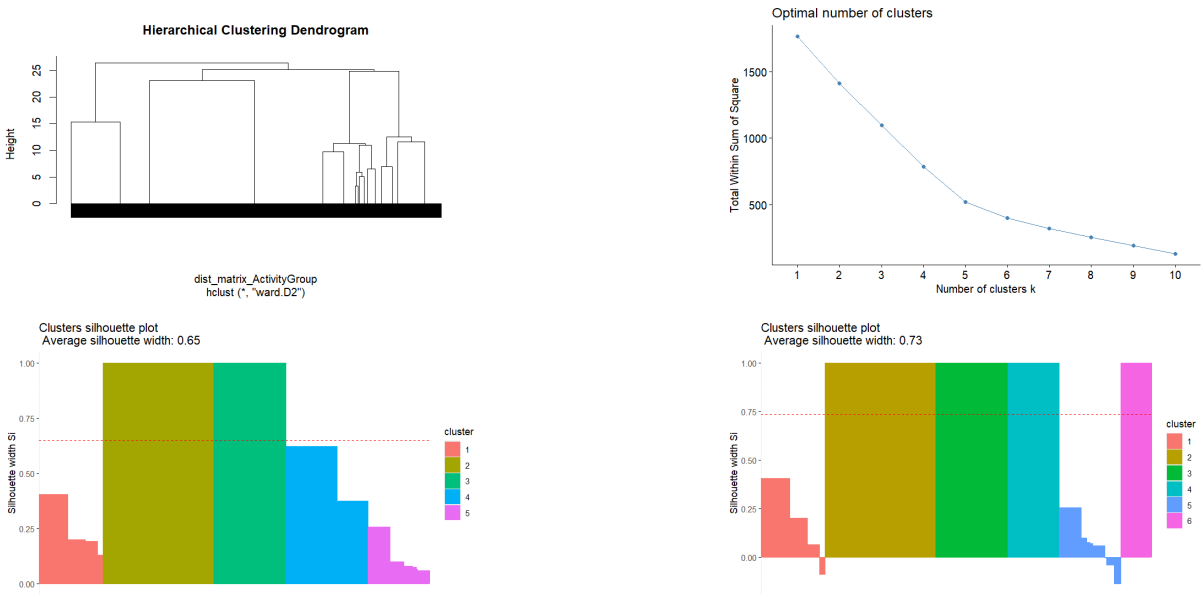

Figure 4: Dendogram,WSS and silhouette representations of Activity Group clustering.

In Figure 5, we assess the model performance after applying priors to Equation (6). The residual plots (second row in Figure 3) suggests desired model performance, indicating that the residuals follows the expected patterns. Additionally, the posterior predictive checks further confirm that the model is a good fit.

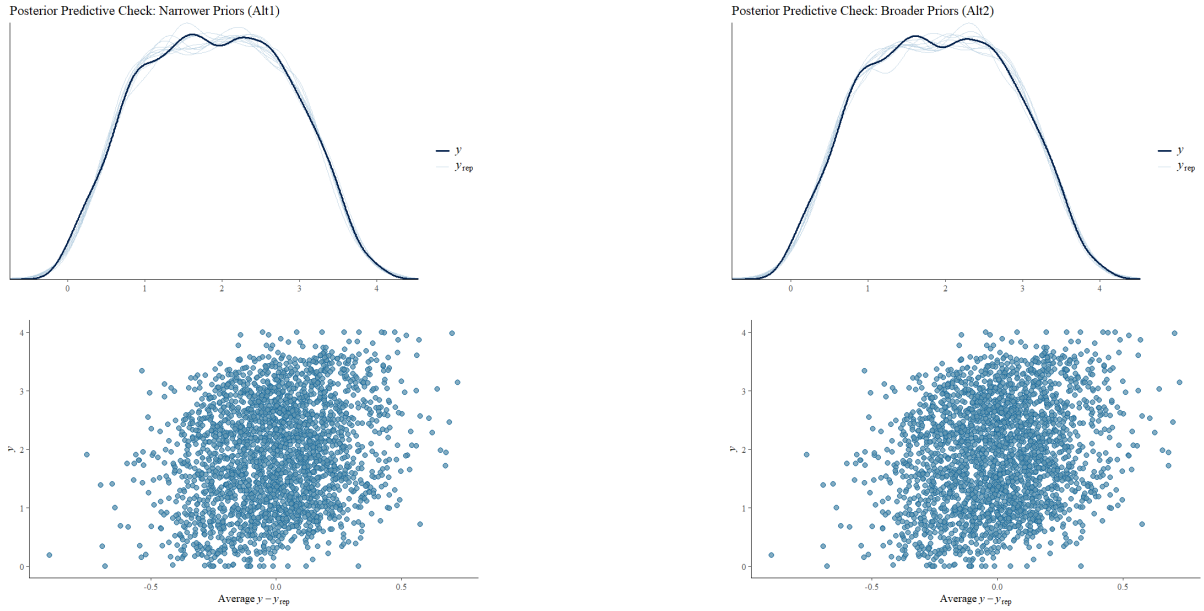

Figure 5: Posterior Predictive Checks (PPC) and Residuals for Activity Group Model. The first row shows PPC results, while the second row presents residual diagnostics.

The prior sensitivity analysis on Equation (5) and (6) demonstrates that the choice of

priors has minimal influence on the posterior estimates, indicating that the model is data-driven rather than prior-dependent. While Alt1 provides a more constrained approach, Alt2 allows for greater prior flexibility with only a minor increase in uncertainty.

3.2 Continuous and Ordinal Models

In this section, we discuss the comparison of predictors for both ordinal and continuous models. After obtaining satisfactory results for the priors and clustering, we proceed by using the Activity Group as the hierarchy and Narrow priors for the final model. In our pursuit of the optimal ordinal model, we have decided to exclude Gender and Parental Education from the final model comparisons. This decision is based on the findings from the Posterior Predictive Check, which suggested that the inclusion of these variables did not improve the model performance. Thus, the model formula in Equation (5) will be used for both ordinal and continuous models moving forward.

3.2.1 Continuous Model

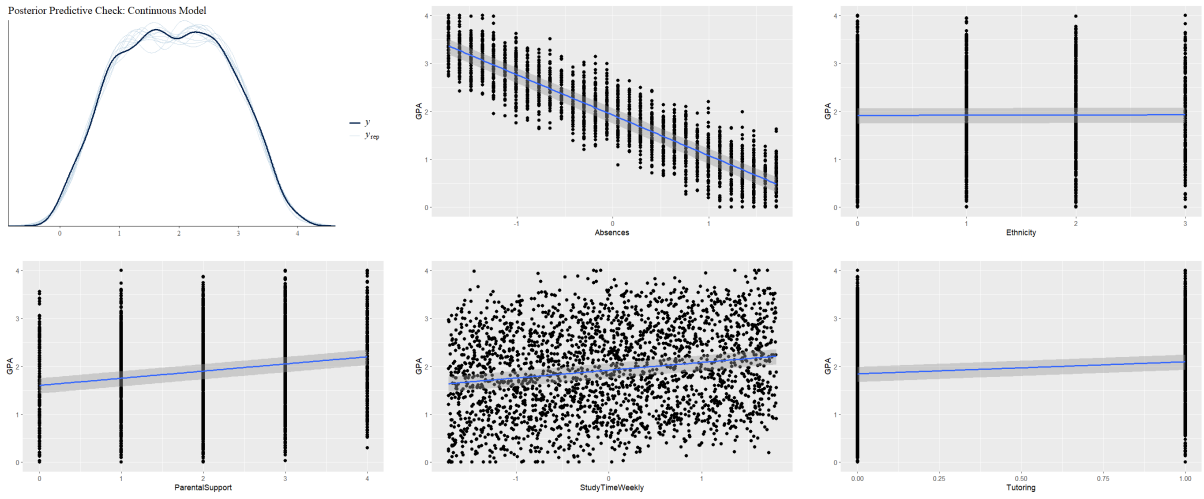

Figure 6: The first image is PPC for model and rest is conditional probabilities.

Analyzing the Figures 6 and 7 for both Continuous and Ordinal models, the conditional probabilities indicates that the most influential predictor is Absences, with a strong positive effect, indicating that missing school significantly increases the likelihood of lower academic performance. Tutoring, Parental Support, and Study Time Weekly all have negative coefficients, meaning they contribute positively to student performance. Ethnicity has a weak effect, meaning that once other factors are controlled for, its impact on GradeClass or GPA is minimal. Models suggests that reducing absences and increasing parental support, study time, and tutoring may help improve student performance.

3.2.2 Ordinal Model

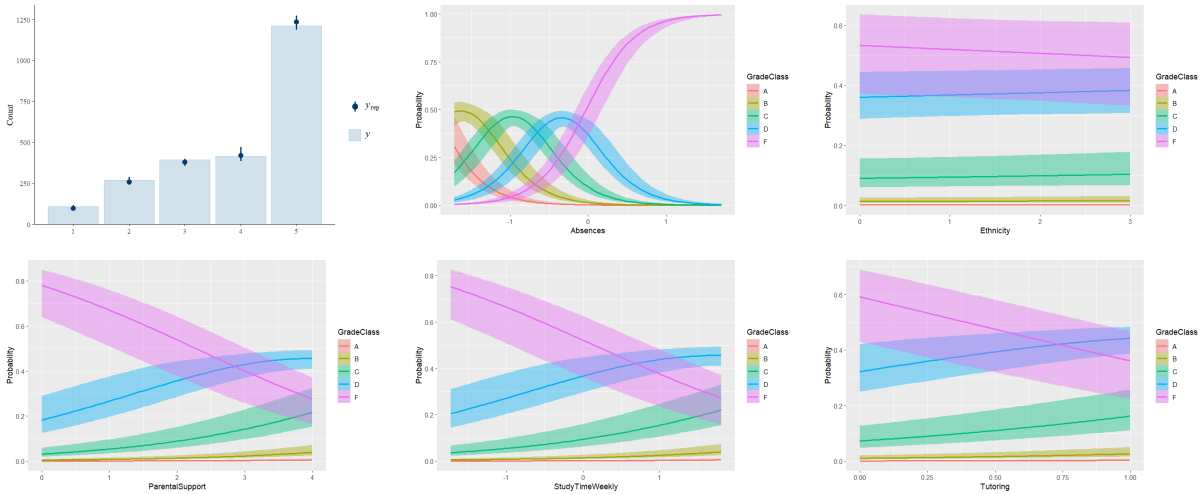

Figure 7: The first image is PPC for model and rest is conditional probabilities.

4 Convergence Diagnostics

4.1 Ordinal Model (Cumulative Logit)

The model equation of Ordinal model is

$$\begin{aligned}\text{logit}(P(Y \leq k)) = \alpha_k - (&\beta_1 \times \text{ParentalSupport} + \beta_2 \times \text{StudyTimeWeekly} \\ &+ \beta_3 \times \text{Absences} + \beta_4 \times \text{Tutoring} + \beta_5 \times \text{Ethnicity})\end{aligned} \tag{7}$$

α_k is the threshold (intercept) for the ordinal GradeClass model (A, B, C, D, E).

Table 1: Posterior Estimates for Ordinal Model

Predictor	**Estimate**	**SE**	**l-95% CI**	**u-95% CI**	**Rhat**	**Bulk ESS**	**Tail ESS**
Intercept[1]	-7.80	0.35	-8.43	-7.05	1.00	2963	3334
Intercept[2]	-5.61	0.33	-6.18	-4.91	1.00	2881	3144
Intercept[3]	-3.57	0.31	-4.11	-2.92	1.00	2874	3216
Intercept[4]	-1.57	0.30	-2.08	-0.90	1.00	2851	3255
Tutoring	-0.94	0.10	-1.13	-0.74	1.00	7300	6202
ParentalSupport	-0.56	0.04	-0.64	-0.47	1.00	7150	5358
StudyTimeWeekly	-0.59	0.05	-0.69	-0.50	1.00	7486	5913
Absences	3.20	0.09	3.02	3.38	1.00	5025	5757
Ethnicity	-0.05	0.04	-0.14	0.04	1.00	8067	5874
Cluster_HC	**Estimate**	**SE**	**l-95% CI**	**u-95% CI**	**Rhat**	**Bulk ESS**	**Tail ESS**
sd(Intercept)	0.49	0.27	0.21	1.21	1.00	2110	3056

4.2 Continuous Model (Gaussian)

Model equation of Continuous model is

$$\begin{aligned} \text{GPA} = \beta_0 + \beta_1 \times \text{ParentalSupport} + \beta_2 \times \text{StudyTimeWeekly} \\ + \beta_3 \times \text{Absences} + \beta_4 \times \text{Tutoring} + \beta_5 \times \text{Ethnicity} + \varepsilon \end{aligned} \tag{8}$$

Table 2: Posterior Estimates for Continuous Model

Predictor	**Estimate**	**SE**	**l-95% CI**	**u-95% CI**	**Rhat**	**Bulk ESS**	**Tail ESS**
Intercept	1.52	0.08	1.35	1.67	1.00	1987	2188
ParentalSupport	0.15	0.00	0.14	0.16	1.00	8942	4895
StudyTimeWeekly	0.16	0.00	0.15	0.17	1.00	10218	5189
Absences	-0.84	0.00	-0.85	-0.83	1.00	9353	5782
Tutoring	0.25	0.01	0.23	0.27	1.00	9104	5483
Ethnicity	0.00	0.00	-0.00	0.01	1.00	9032	5398
Cluster_HC	**Estimate**	**SE**	**l-95% CI**	**u-95% CI**	**Rhat**	**Bulk ESS**	**Tail ESS**
sd(Intercept)	0.16	0.09	0.07	0.41	1.00	2003	2696

4.3 Model Comparison

- The hierarchical variance for the ordinal model (0.49) is higher than for the continuous model (0.16), suggesting greater between-cluster variation in GradeClass than GPA.
- The effects of StudyTimeWeekly and ParentalSupport are significant in both models, but their interpretation differs: in the ordinal model, they influence the likelihood of being in a higher grade category, whereas in the continuous model, they directly affect GPA.
- The continuous model exhibits a divergent transition warning, indicating possible poor mixing, thus lower ESS values, particularly for the Intercept.
- Rhat values are all close to 1.00, suggesting good convergence, but Bulk ESS and Tail ESS are lower for the continuous model's intercept, which may indicate less efficient sampling.
- In the ordinal model, ESS values are generally high (e.g., 2963 for Intercept[1], 7300 for Tutoring), suggesting robust sampling for these parameters.
- In the continuous model, ESS values are lower for some parameters (e.g., 1987 for the Intercept), indicating less efficient sampling and potentially less reliable estimates.

5 Conclusion

Both models provide valuable insights, but the ordinal model may be more appropriate given the categorical nature of GradeClass. The continuous model requires diagnostic improvements to address divergence issues. During prior sensitivity analysis there were 0 divergent branches but a divergent branch emerged after we removed Gender and ParentalEducation from Equation (5). This might also be a hint that by removing these we are overfitting the data or the model requires further improvement. Based on Posterior predictive checks for both Ordinal and continuous models it is hard to decide which one is better as both the models perform almost identical. And since model types are different WAIC and LOO can't be used to differentiate between two different types of models as the target is of different type. The cumulative logit model assumes a smooth transition between grade classes, but real-world grading isn't always linear, therefore information loss is occurred. Continuous model should be preferred when high accuracy of GPA and low information loss is desired.

6 Limitations and Potential improvements

One improvement that can be made is by increasing adapting to a more optimized `adapt_delta` in BRMS in R. And since Absences has a very strong positive effect, its impact might be nonlinear. So, instead of treating it as a continuous variable, a threshold or polynomial transformation (e.g., I($\texttt{Absences}\hat{2}$)) can be made to capture nonlinearity.

7 Reflection on own Learnings

Throughout this project, We have gained a deeper understanding of Bayesian multilevel modeling, particularly in the context of student performance analysis. Working with hierarchical structures reinforced the importance of selecting appropriate priors and assessing their impact through prior sensitivity analysis. The process of diagnosing model convergence using Rhat and Effective Sample Size (ESS) provided valuable insights into the reliability of the posterior estimates. Additionally, performing posterior predictive checks and residual analysis emphasized the significance of model validation beyond point estimates. One of the key challenges encountered was dealing with divergent transitions in the Gaussian model, which required tuning the `adapt_delta` parameter to ensure stable sampling. This experience highlighted the computational complexities of Bayesian inference and the necessity of careful model specification. Overall, this project has enhanced my ability to apply Bayesian techniques to real-world problems, improving my skills in both statistical modeling and critical evaluation of results. Future improvements could involve incorporating more flexible priors, exploring alternative link functions, and refining model assumptions to better capture the nuances of student performance data.

References

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32.

Bürkner, P.-C. (2024). *Bayesian regression models using stan (brms)* [Accessed: March 13, 2025]. https://paulbuerkner.com/software/brms-book/brms-book.pdf