Analyzing Quiz Performance Trends: The Impact of Time, Gender, and Major

Sarah Jafari

Table of Contents

1. ObjectivesoftheStudy

The primary goal of this study is to analyze trends in quiz performance over time among students in a statistics class. In addition to this overarching aim, the study seeks to explore secondary objectives related to subgroup differences and potential outliers. Specifically, this study aims to:

- Determineifthereisanoveralltrendofimprovementinquizscoresacrossall students.
- Evaluatewhethergenderinfluencestrendsinquizperformance.
- Assesswhetherstudentsfromdifferentmajorsexhibitdifferingpatternsinquiz performance.
- Identifypotentialoutliersthatmayskewoverallpatterns.

Understanding these trends can provide valuable insights into how demographic factors, such as gender and academic major, affect learning outcomes over time. These findings could inform teaching strategies, support tailored academic interventions, and contribute to improving overall course design.

2. DataDescription

Dataset Overview:

The dataset (class.data) contains quiz scores from a statistics class, along with demographic information for each student. It includes scores from five quizzes administered throughout the semester and details about each student, their gender and academic major.

Variables:

The dataset contains the following variables:

- ID:Uniqueidentifierforeachstudent[Numeric]
- Gender:Studentgender,codedasMale(M)orFemale(F)[Categorical]
- Major:CodedasMathematics(Math),Statistics(Stat),orComputerScience(Comp). [Categorical]
- Quiz1-Quiz5:Scoresfromfivequizzesadministeredthroughoutthesemester. [Numeric]

Sample Size:

The dataset consists of 52 students:

- Genderdistribution:29malesand23females.
- Majordistribution:11Math,25Statistics,and16ComputerSciencestudents.

Preliminary Observations & Analysis

An initial exploration of the dataset reveals potential trends in quiz scores over time, as well as differences by gender and major. This variability provides an opportunity to examine whether scores improve over time and how demographic factors may influence performance.

Summary Statistics:

The following summary statistics provide an overview of quiz performance, including measures of central tendency, variability, and range:

Quiz	Mean	Median	Standard Deviation	Minimum Score	Maximum Score
1	65.60	64.5	15.79	31	97
2	69.04	70.5	14.34	42	96
3	73.88	77.0	18.29	11	100
4	77.42	80.0	12.14	55	92
5	77.67	80.0	13.30	38	99

The summary statistics reveal consistent improvement in quiz scores over time, with mean and median scores increasing across all five quizzes. Variability peaked in Quiz 3, as indicated by the highest standard deviation, but decreased in later quizzes, reflecting more consistent performance. The dataset contains no missing values, and initial variability in scores suggests potential subgroup differences, which are explored in subsequent analyses.

3. Exploratory Data Analysis (EDA)

After presenting summary statistics for the key variables in our dataset, we will now explore variations in quiz scores across gender and majors to identify meaningful patterns and trends.

Gender-based Performance Trends:

To explore variations in quiz scores across genders, we calculated the average scores for each quiz for male (M) and female (F) students. The analysis aimed to determine whether one gender generally performed better than the other and to assess overall trends in performance across the semester.

Findings

- InitialPerformance:Femalestudentsbeganthesemesterwithahigheraveragescore on Quiz 1 (68.15) compared to male students (63.04).

- Mid-SemesterTrends:MalestudentsoutperformedfemalesslightlyonQuiz3and Quiz 4, with marginal differences in average scores.
- End-of-SemesterPerformance:ByQuiz5,femalestudentsscoredanaverageof 78.50, while male students scored 76.85. The gap between genders narrowed significantly by the end of the semester.
- OverallImprovement:Bothgendersdemonstratedapositivelearningtrend,with consistent improvements in average quiz scores over time. This suggests that both groups adapted well to the course material as the semester progressed.

These results highlight notable patterns in gender-based performance, with female students maintaining a slight lead in most quizzes and male students showing significant progress over time. These trends warrant further examination in subsequent analyses to determine whether gender influences performance trends significantly.

Visualizing Gender-Based Performance Trends

To explore gender-based performance trends, a combined boxplot compared quiz score distributions between male and female students. Female students consistently had higher median scores in quizzes they led early on (e.g., Quiz 1 and Quiz 2), while male students showed significant improvement, narrowing the gap by Quiz 5. Outliers were primarily observed in Quiz 1 and Quiz 5 for males. Both genders demonstrated more consistent performance as the semester progressed, supporting earlier observations of overall improvement.

These visualizations support the earlier findings, showing that female students generally led in median scores, while male students demonstrated significant improvement, narrowing the performance gap by the end of the semester.

Major-based Performance Trends:

After analyzing gender-based trends, we examined quiz scores across academic majors to uncover patterns of performance among Math, Statistics, and Computer Science students. The average scores for each major were calculated and compared across all five quizzes.

Findings:

- ComputerScienceMajors:Consistentlyoutperformedothermajorsacrossallquizzes, peaking in Quiz 4 with an average score of 80.25.
- MathMajors:Showedthelowestaveragescores,particularlyinQuiz2(60.73). However, steady improvement was observed over time, culminating in an average score of 76.45 in Quiz 5.
- StatisticsMajors:Maintainedrelativelyhighandstableaveragescores,peakingin Quiz 5 with an average of 77.88.

The analysis revealed that Computer Science majors consistently outperformed their peers, followed by Statistics and then Math majors. These differences in performance may reflect variations in academic preparation, focus, or engagement among the three groups. Notably, Math majors demonstrated a steady improvement over time, suggesting a positive learning curve, while Statistics majors maintained relatively consistent performance throughout the semester, reflecting stability in their scores.

Visualizing Major-Based Performance Trends:

A grouped boxplot illustrates performance trends by major. Computer Science majors consistently performed well, with median scores between 75%-85%, indicating stable performance. Math majors exhibited wide score variability, with median scores ranging from 50%-70%, suggesting challenges with quiz difficulty or varying skill levels. Statistics majors started with lower scores but showed gradual improvement, particularly in Quizzes 4 and 5, where their median and upper quartile increased noticeably. Outliers were observed in all majors, especially in Math and Statistics, reflecting a few students who underperformed. Overall, Computer Science majors maintained stable performance, Statistics majors improved over time, and Math majors displayed the most variability.

Overall Trends:

The purpose of this analysis is to describe the trends in quiz scores over time, with a focus on how these trends differ by gender and major. The primary questions addressed are:

1. Isthereanoverallimprovementinquizscoresovertime?
2. Domaleandfemalestudentsshowsimilartrendsovertime?
3. Arethereanydifferencesintrendsbetweenstudentsfromdifferentmajors?

To visualize the answers, we have created a line plot that shows quiz score trends by gender and major. According to the line plot, an upward trend in quiz scores is evident across the semester, indicating overall improvement over time. Among Computer Science students, females consistently achieved high scores throughout the semester, while males also performed well, showing a significant increase between Quiz 2 and Quiz 4. Math students, both male and female, started with lower scores but demonstrated notable progress across the quizzes, with male Math students improving at a faster rate. Statistics students exhibited distinct patterns: females began with relatively low scores but showed steady and remarkable improvement, nearly closing the gap with other groups by Quiz 5, while males achieved the highest scores in Quiz 4.

Despite variations in performance trends by gender and major, the results highlight a consistent pattern of improvement for all students over time, suggesting adaptation to the course material and quizzes as the semester progressed.

Summary of EDA Insights:

The exploratory data analysis identified key trends in quiz performance across gender and major groups. Quiz scores improved over time, with female Computer Science students consistently performing well and male Statistics students showing the most improvement, peaking in Quiz 4. Math majors displayed the highest variability but showed steady progress. These findings provide a foundation for confirmatory data analysis, which will use regression modeling to quantify these relationships and test hypotheses regarding the influence of gender, major, and their interactions on quiz performance.

4. Confirmatory Data Analysis (EDA):

The Confirmatory Data Analysis (CDA) phase builds on the findings from the Exploratory Data Analysis (EDA) to rigorously test and quantify relationships observed in the data. Specifically, this section will use regression modeling to evaluate the influence of time (quiz number),gender,major,andtheirinteractionsonquizperformance.Byincorporating statistical methods, we aim to validate the observed trends, assess the significance of key predictors, and explore how these factors interact to explain variations in student performance. This analysis will provide deeper insights into group-specific dynamics and help answer critical questions raised during EDA.

Model Selection:

To investigate the factors influencing quiz performance, we consider multiple regression models with the following key components:

1. ResponseVariable:Quizscores(dependentvariable).
2. PredictorVariables:
    - QuizNumber:Anumericalvariablerepresentingtime(1,2,3,4,5)
    - Gender:Acategoricalvariablecodedas0=Male,1=Female
    - Major:Acategoricalvariablewiththreelevels:Math,Statistics,andComputer Science.
3. InteractionTerms:Tocapturegroup-specifictrends,interactionsbetweenpredictors will be included:
    - Gender-QuizNumber
    - Major-QuizNumber
    - Gender-Major

Candidate Models:

Model 1 - Base Model

The Base Model is designed to evaluate the main effects of key predictors on quiz scores. It provides initial insights into whether quiz performance improves over time and whether differences exist across gender or major groups. The model can be expressed as:

$$\beta0 + \beta1(\text{Quiz Number}) + \beta2(\text{Gender}) + \beta3(\text{Major}) + \epsilon$$

This model serves as a foundation for understanding the direct effects of time, gender, and major on quiz scores. By focusing on the main effects, this model allows us to:

- Testinitialhypotheses,suchaswhetherscoresimprovewitheachquizorvaryby demographic groups.
- Identifysignificantpredictorsofquizperformance.
- Provideabaselineframeworkforcomparisonwithmorecomplexmodels,suchas those incorporating interaction terms or non-linear effects.
  The insights from this model will guide refinements in future modeling, such as the inclusion of interactions or adjustments for individual variability.

Interpreting the Model Coefficients:

- QuizNumber($\beta1$):ThecoefficientforQuizNumberisapproximately3.25($p<0.001$), indicating a statistically significant positive relationship between quiz number and scores. This suggests that, on average, quiz scores increase by about 3.25 points per quiz, supporting the hypothesis of an improving trend over time for all students.
- Gender($\beta2$):ThecoefficientforGenderisapproximately0.87($p=0.648$),whichisnot statistically significant. This suggests that, after accounting for other variables, there is no strong evidence that gender influences overall quiz scores.
- Major($\beta3$):ThecoefficientforMathisapproximately-6.23($p=0.022$),indicatingthat Math students score significantly lower than Computer Science students, on average. The coefficient for Statistics is approximately -2.10 ($p=0.331$), which is not statistically significant. This suggests no meaningful difference between Statistics and Computer Science majors after accounting for other variables.

Model Performance:

- TheBaseModelexplains11.14%ofthevarianceinquizscores($R^2=0.1114$).
- Theresidualstandarderroris14.80,indicatingreasonablepredictionaccuracyforthis baseline model.

The Base Model confirms a significant positive trend in quiz scores over time, with students improving by an average of 3.25 points per quiz. While gender does not appear to have a meaningful impact, Math majors scored significantly lower than Computer Science students, highlighting a potential area for targeted support. These findings establish a solid foundation for further exploration in more complex models.

Model 2: Interaction Model

The Interaction Model extends the Base Model by introducing subgroup-specific terms through interactions between Quiz Number, Gender, and Major. This allows for an evaluation of whether performance trends differ across demographic groups.

$$\beta 0 + \beta 1(\text{Quiz Number}) + \beta 2(\text{Gender}) + \beta 3(\text{Major}) + \beta 4(\text{Gender-Quiz Number}) + \beta 5(\text{Major-Quiz Number}) + \beta 6(\text{Gender-Major}) + \epsilon$$

Interpreting the Model Coefficients:

- QuizNumber:Scorescontinuetoshowasignificantupwardtrend,withacoefficient of 2.51 (p=0.045), indicating an increase of approximately 2.51 points per quiz.
- Interactionterms:Noneoftheinteractiontermswerestatisticallysignificant(p>0.05), suggesting that the relationships between predictors and quiz scores do not vary meaningfully across subgroups.
- Major:MathstudentsscoredlowerthanComputerSciencestudents(-12.23),butthis was only marginally significant (p=0.091). Differences between Statistics and Computer Science students remained non-significant.
- Gender:Genderdifferencesremainedinsignificant(p=0.2),consistentwiththeBase Model findings.

Model Performance:

- Theinteractionmodelexplains12.24%ofthevarianceinquizscores(R^2=0.1224), with an adjusted R^2 of 0.0908.
- Theresidualstandarderroris14.83,comparabletotheBaseModel.

The Interaction Model adds complexity but does not improve explanatory power significantly. None of the interaction effects were statistically significant, and the model's goodness-of-fit metrics showed minimal improvement over the Base Model. As such, the Base Model remains preferable for its simplicity and clarity.

Base Model vs Interaction Model:

Both models capture the upward trend in quiz scores over time, with statistically significant coefficients for Quiz Number. However, the interaction model adds little explanatory power, as none of the interaction terms are statistically significant, and the R^2 improves only marginally (12.24% compared to 11.14% in the base model). The adjusted R^2 for the interaction model is slightly lower than that of the base model, indicating that the added complexity does not enhance the model's performance.

Residual errors are comparable between the two models, with similar ranges and standard errors (14.80 for the Base Model vs. 14.83 for the Interaction Model). Given the lack of significant interaction effects and the negligible improvement in model fit, the Base Model is preferable for its simplicity and clarity while effectively capturing the main trends in the data.

Explanatory Power and Limitations:

The low R^2 values in both models (Base Model and Interaction Model) indicate limited explanatory power, meaning that the predictors (quiz Number, gender, major) only explain a small proportion of the variability in quiz scores. This could be due to the following factors:

1. UnmeasuredFactors:
   - Variablessuchasstudyhabits,priorknowledge,quizdifficulty,motivation, and external support are likely influential but not included in the dataset. These unmeasured factors may contribute substantially to the variation in quiz scores.
2. HighVariabilityAmongStudents:
   - Quizscoresexhibitsignificantindividualdifferences,asnotedinthe Exploratory Data Analysis. While some students show consistent improvement, others may struggle or experience fluctuations due to personal or external circumstances. Linear models, which assume a uniform trend across all individuals, may fail to adequately capture such variability.

Checking for Multicollinearity:

To ensure the stability and interpretability of regression coefficients, we assessed multicollinearity using the Variance Inflation Factor (VIF) and a correlation matrix. The VIF values for all predictors were below 1.10, indicating no multicollinearity among predictors. This means the predictors in the model are not strongly correlated, and their individual effects can be interpreted without concerns about inflated standard errors. Additionally, pairwise correlations between numeric predictors were low, further confirming the independence of predictors in the model. These results validate the use of our selected predictors in the regression analysis.

Residual Diagnostics

To evaluate the validity of the assumptions underlying the Base Model, standard diagnostic plots were analyzed, including the Residuals vs. Fitted, Normal Q-Q, Scale-Location, and Residuals vs. Leverage plots. The key observations are summarized below:

Residuals vs. Fitted Plot

- Theresidualsarescatteredaroundzero,indicatingthattherelationshipbetweenthe predictors and the response variable is approximately linear.
- Minorpatternsandclusteringarepresent,suggestingareaswherethemodelcould potentially be refined.

Normal Q-Q Plot

- Mostresidualsaligncloselywiththediagonalline,supportingtheassumptionof normally distributed residuals.
- Smalldeviationsareobservedattheextremes(tails),suggestingslightnon-normality, but this is not severe enough to invalidate the model.

Scale-Location Plot

- Thespreadofresidualsisrelativelyconsistentacrosstherangeoffittedvalues,with only slight variations in spread. This indicates that the variability in residuals is generally stable.

Residuals vs. Leverage Plot

- Threeobservations(38,140and150)haveslightlyhigherleveragebutremainbelow the Cook’s distance threshold, meaning they do not overly influence the model.

Residual Assumption Validation:

The validity of the regression model was assessed by testing key assumptions. Initially, the Durbin-Watson test indicated autocorrelation among the residuals ($p<0.05$), violating the independence assumption. To address this, a lagged term for quiz scores was introduced into the Base Model, which resolved the issue ($p>0.05$). The Residuals vs. Fitted plot showed no clear patterns, confirming a linear relationship between predictors and the response variable. The Scale-Location plot displayed a uniform spread of residuals, and the Breusch-Pagan test ($p>0.05$) confirmed homoscedasticity. The Q-Q plot showed residuals aligning closely with the diagonal which supported normality. These results confirm that the updated model satisfies all key assumptions, ensuring the reliability of the analysis.

Investigation of Influential Points

During residual diagnostics, three influential observations (38, 148, and 150) corresponding to three students (scores of 11, 37, and 38 on Quizzes 3 and 5) were identified. These points showed moderate leverage but did not exceed the Cook’s distance threshold.

To evaluate their impact on the model, a new regression was fit after excluding these observations. The results of this updated model were compared with the original Base Model to determine whether these influential points affected the validity or robustness of the conclusions.

Comparison of Models

The table below summarizes the key differences between the original model (with influential points) and the updated model (without influential points):

Metric	Original Model	Updated Model
Intercept	62.96	64.95
Coefficientforquiz#	3.25(p<0.001)	3.42(p<0.001)
CoefficientforgenderM	N/A N/A N/A	0.035(p=0.985)
CoefficientformajorMath	14.88 0.0879	-5.54(p=0.032)
CoefficientformajorStat		-1.65(p=0.419)
ResidualStandardError		13.98
R^2		0.1248

Observations

1. TrendOverTime:
   - ThecoefficientforQuizNumberremainspositiveandhighlysignificantin both models (p < 0.001), supporting the hypothesis that quiz scores improve over time.
   - Theeffectsizeslightlyincreasesintheupdatedmodel(3.42vs.3.25), suggesting a stronger relationship between quiz number and scores when influential points are excluded.
2. ImpactofInfluentialPoints:
   - Excludingtheinfluentialpointsledtoamodestimprovementinmodelfit,as evidenced by an increase in R^2 (from 0.1114 to 0.1248) and a reduction in residual standard error (from 14.78 to 13.98).
   - Theseimprovementssuggestthattheinfluentialpointscontributedtolarger residuals and reduced the model’s explanatory power.
3. SubgroupEffects:
   - Genderremainsaninsignificantpredictorintheupdatedmodel(p=0.985), indicating no meaningful impact of gender on quiz scores.
   - MathmajorsexhibitsignificantlylowerscorescomparedtoComputerScience majors (−5.54, p=0.032), while Statistics majors do not differ significantly (−1.65, p=0.419).

Model Comparison Conclusion:

The updated model provides a clearer distinction in performance trends for Math majors while preserving the overall trend of improving quiz scores over time. However, excluding influential points has implications for the generalizability of the findings:

- IncludingInfluentialPoints:includingthesepointsreflectsrealstudentperformance and maintains the generalizability of the analysis.

- ExcludingInfluentialPoints:Removingthesepointsimprovesmodelfitandhighlights subgroup differences but may sacrifice generalizability.

The decision to include or exclude influential points should depend on the primary focus of the analysis. For broad insights into overall trends, it is preferable to include these points. However, for fine-tuned analysis emphasizing subgroup differences, excluding them is justifiable.

Model 3: Polynomial Model ($\beta0$ + $\beta1$(Quiz Number) + $\beta2$(Quiz Number^2) + $\beta3$(Gender) + $\beta4$(Major) $\beta5$(Gender-Quiz Number) + $\beta6$(Major-Quiz Number) + $\beta7$(Gender-Major) + $\beta8$(Quiz Number^2-Gender) + $\beta9$(Quiz Number^2-Major) + ϵ)

To explore potential non-linear trends in quiz scores, a polynomial model was fit, extending the interaction model with quadratic terms for time. This model aimed to capture curvature in performance trends across quizzes. Key results include:

- Thequadraticterms(poly(time,2))werenotstatisticallysignificant(p>0.15), suggesting that non-linear trends in quiz scores over time are not meaningful.
- Noneoftheinteractiontermsinvolvinggenderormajorshowedsignificantnon-linear effects (p>0.05).
- MathmajorsscoredsignificantlylowerthanComputerSciencemajors(−10.64, p=0.0151), while Statistics majors did not differ significantly (−2.34, p=0.4015).

Model Performance:

The model explained 13.92% of the variance in scores (R^2 = 0.1392), slightly higher than the interaction model (R^2 = 0.1248). However, the modest improvement in R^2 does not justify the added complexity of the polynomial terms.

The results indicate that quiz scores exhibit primarily linear trends, with no evidence to support significant non-linear effects over time. While the polynomial model offers a slight improvement in explanatory power, its added complexity reduces interpretability and efficiency. For these reasons, the simpler Interaction Model or Base Model remains preferable for analyzing quiz performance.

Model Comparison Using AIC

To identify the most appropriate model for explaining trends in quiz performance, we compared three models using the Akaike Information Criterion (AIC). AIC evaluates model fit while penalizing for complexity, allowing for the selection of a model that balances explanatory power with simplicity.

AIC Results:

Model	Degrees of Freedom	AIC

BasicModel	6	2145.18
InteractionModel	11	4
PolynomialModel	19	2151.93

2

- TheBaseModelhasthelowestAICvalue(2145.184),indicatingthatitprovidesthe best trade-off between model fit and complexity.

2162.91

- TheInteractionModelhasahigherAIC(2151.932),suggestingthatadding interaction terms increases complexity without a significant improvement in fit.

2

- ThePolynomialModelhasthehighestAIC(2162.912),demonstratingthatthe inclusion of quadratic terms further increases complexity but fails to enhance explanatory power.

Based on the AIC comparison, the Base Model is the most suitable choice for analyzing quiz performance. It effectively captures the main trend over time and subgroup differences while maintaining simplicity and interpretability. Although more complex models explored potential interactions and non-linear trends, they did not provide sufficient improvement in model fit to justify their added complexity. The Base Model's balance of explanatory power and simplicity makes it the optimal choice for this analysis.

Trends in Quiz Scores Over Time

To better understand the findings from the Base Model, trends in quiz scores over time were analyzed, focusing on variations by gender and major. Observations from the analysis are:

Plot 1: Overall Trend in Quiz Scores

A positive linear trend in quiz scores was observed across all students, with scores consistently improving as the semester progressed. This aligns with the Base Model findings, where Quiz Number was a significant predictor of scores ($p<0.001$). The consistent improvement suggests that students adapted to the course material and became more familiar with the quiz format over time.

Plot 2: Quiz Score Trends by Gender

Both male (M) and female (F) students exhibited similar positive trends in quiz scores over time, with no significant differences in performance. This is consistent with the regression analysis, where gender was not a statistically significant predictor ($p=0.648$). These findings indicate that gender does not play a significant role in influencing quiz scores within this dataset.

Quiz Score Trends by Major

Distinct trends were observed across the three academic majors:

- MathmajorsconsistentlyscoredlowerthanComputerScienceandStatisticsmajors, reflecting a significant difference identified in the regression analysis (p=0.032).
- ComputerScienceandStatisticsstudentsshowedcomparableperformance,with slightly higher average scores than Math students.

These findings support the Base Model, where Math was a significant predictor, indicating that Math majors performed worse on average compared to Computer Science majors.

Limitations and Recommendations

While the analyses provide valuable insights into quiz performance, several limitations must be acknowledged:

1. ModestExplanatoryPower:Themodelsexplainonlyasmallproportionofthevariance in quiz scores (values below 15%). This suggests the presence of additional unmeasured factors, such as prior academic performance, quiz-specific difficulty, study habits, or attendance, which could influence scores.
2. SimplisticModelingAssumptions:Theinearmodelsassumeuniformeffectsof predictors, potentially oversimplifying complex relationships. Interaction and polynomial terms provided limited improvement.
3. SmallSampleSize:Thedatasetincludesalimitednumberofstudentsandquizzes, restricting the generalizability of findings to broader contexts.
4. PotentialBiasinInfluentialObservations:Asmallnumberofinfluentialpointswere identified, and while their exclusion improved model fit slightly, they represent valid student performances and should not be dismissed without consideration.

Future work should collect additional predictors and consider more advanced modeling techniques, such as hierarchical models, to address these limitations.

Practical Implications

The findings from this analysis have several implications for course design and instruction:

1. PositiveTrendinScoresOverTime:Theupwardtrendinscoressuggeststhat students adapt well to course material and assessment formats, reflecting effective teaching strategies.
2. EquityinGenderPerformance:Thelackofsignificantgenderdifferenceshighlights equitable outcomes, underscoring the inclusivity of the course structure.
3. ChallengesforMathMajors:Mathmajorsconsistentlyunderperformedcomparedto their peers, indicating the need for tailored interventions, such as tutoring or additional resources, to bridge this gap.

Addressing these insights can enhance student learning outcomes and provide a framework for improving teaching practices in similar settings.

Conclusion

This study provided a detailed analysis of quiz performance trends, identifying key patterns and rigorously testing relationships. The findings confirm a positive trend in quiz scores over time, suggesting adaptation to course material and familiarity with quiz formats. Gender differences were insignificant, highlighting equitable performance, while Math majors scored significantly lower than their peers, underscoring the need for targeted interventions.

The Base Model was determined to be the most effective, balancing simplicity and explanatory power. While more complex models explored non-linear trends and subgroup-specific effects, they offered minimal improvement in fit. Residual diagnostics validated the robustness of the Base Model, with influential points having little impact on the conclusions.

Future work should expand the dataset and incorporate additional predictors to enhance explanatory power and generalizability. By addressing these gaps, future analyses can provide deeper insights into student performance, informing data-driven strategies to support learning and equity in education.
