

Statistical Assessment of the Relationship Between Subject Performance and CGPA

Zohaib Ali*

Institute of Mathematics & Computer Science, University of Sindh, Jamshoro
Email: zohaib.ali@usindh.edu.pk

Abdul Rafiu Alias Furkan

Subject Specialist GBHSS Mithani, Naushahro Feroze, Sindh Education and Literacy Department, Government of Sindh

Sumera Dero

Institute of Mathematics & Computer Science, University of Sindh, Jamshoro

Ubaidullah

SST Public School Rashidabad, District Tando Allahyar

ABSTRACT

This study investigates the relationship between course performance in core mathematics subjects and overall academic achievement measured through Cumulative Grade Point Average (CGPA). The dataset consists of 50 students' academic records collected from the official marksheets portal of the University of Sindh (7th and 8th semester results). The dependent variable (Y) represents students' combined CGPA of the 7th and 8th semesters, while the independent variables include the marks obtained in Functional Analysis (X_1), Numerical Analysis-II (X_2), and Rings and Fields (X_3). A multiple linear regression model is applied to examine how effectively these three subjects predict overall semester CGPA. The analysis aims to determine the strength of association, contribution of each subject to academic performance, and the predictive capability of the regression model. Statistical measures such as R, R^2 , Adjusted R^2 , standard errors, and t-statistics are used to evaluate model significance and goodness of fit. The findings of this study provide insight into the academic impact of advanced mathematics courses on overall student performance. The results may assist academic planners and educators in identifying key subjects that significantly influence CGPA and help in curriculum evaluation and academic performance forecasting.

Key Words: Multiple Linear Regression, Coefficient of Determination, Standard Error, t-Test, Confidence Intervals

INTRODUCTION

Academic performance is well known to be one of the main factors to indicate the intellectual growth of the students, their learning outcomes, and their success in the

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

future working life. In college, especially in science and mathematics subjects, the student achievement is assessed using Grade Point Average (GPA) or Cumulative Grade Point Average (CGPA). Universities, employers and policymakers use these indicators to measure academic competence and learning effectiveness. The variables that determine CGPA are hence a crucial topic in the education research, since it can be used to enhance teaching methods, curriculum, and student support mechanisms [1, 2].

In mathematics education, students are expected to gain good analytical, logical, and problem-solving skills. Functional Analysis, Numerical Analysis, and Rings and Fields are advanced mathematical topics that are vital in developing such capabilities. These topics are the basis of the modern mathematical theory and its use in science, engineering, economics, and technology. The level of performance in these courses is usually an indicator of conceptual knowledge, computational abilities and abstract thinking of the student. It is therefore necessary to analyze the correlation between marks achieved in these subjects and academic performance in general in order to compare their academic importance [3, 4].

In most universities, even among those that are in the developing world, these assessment systems are based on written exams and numerical grading. Although these systems give quantitative indicators of performance, in most cases they fail to demonstrate how certain students can do better than others [5, 6]. Issues like the quality of teaching, the structure of the curriculum, learning environment, the motivation of the students and the background of the student may all contribute to student results. Nevertheless, academic records are one of the trusted sources of data to be used in the empirical analysis. The analysis of such records using statistical techniques can be used to identify trends and relationships that cannot be identified using only qualitative techniques [7, 8].

The multiple linear regression analysis has emerged as one of the most popular statistical methods of analysis of the relationship amongst academic variables in educational research. It enables researchers to study the combined effect of multiple independent variables on one dependent variable [9-12]. Regression models could be applied to the context of academic performance, where the variables or subject areas that have the most impact on CGPA of students can be established. They can also be used in prediction and, in this way, educators can predict the results of students using early signs of performance results [13, 14].

Other past research works have examined the determinants of academic success in the form of regression and other statistical techniques. The researchers have studied the impact of attendance, socio-economic status, study habits, instructional approaches and previous academic success on student achievement. Most of such studies find that the subject specific performance is a very significant element in defining total academic outcomes, particularly in technical and scientific subjects. Nonetheless, empirical studies devoted to the specific topic of advanced mathematics classes and their influence on CGPA have not yet been conducted in relation to Pakistani universities [15-18].

Mathematics courses like Functional Analysis, Numerical Analysis II, and Rings and Fields are deemed as intellectually challenging in the mathematics field. Such topics

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

demand the students to learn abstract ideas, strict demonstrations, and complicated problem-solving methods. Functional Analysis focuses on physics and engineering, a field that stresses on the use of vector spaces, operators, and applications. Numerical Analysis II involves computational techniques to solve mathematical problems which are important in scientific computing. Rings and Fields present to students algebraic structures on which modern algebra and theoretical mathematics are founded. The achievement in these courses would represent theoretical and practical competence.

Students usually have problems mastering these subjects despite the importance of the subjects. In most institutions, there have been high failure rates and low average scores in the advanced mathematics courses. Such difficulties can be explained by the insufficient preparation, poor access to learning materials, inefficient teaching strategies, or deficient in academic orientation. Determining the relation of the performance in these subjects to the overall CGPA can enable the institutions to minimize these problems.

The last few years have provided new avenues to academic research based on data with the availability of online academic records and digital examination systems. Modern statistical tools have become a possibility to analyze large datasets of marks, grades, and academic histories of students. The analyses can allow researchers to make evidence-based suggestions on how the academic policies and practices can be enhanced [19]. This paper applies an empirical investigation of the academic performance of the students by using official examination data found on the University of Sindh marksheet portal [20].

The current research is based on the overall CGPA of 7th and 8th semesters in students as the dependent variable, which is their total performance in the last semesters at the undergraduate level in mathematics. The independent variables will include scores achieved in Functional Analysis, Numerical Analysis II and Rings and Fields. With the use of multiple linear regression analysis on this data, the research will attempt to establish the degree to which these core subjects affect the CGPA of students.

The research is especially important, as the end-of-year performance tends to be of a crucial factor that predetermines further opportunities of students, such as postgraduate school admission, job perspectives and academic honour. Knowledge about predictors of final CGPA will enable the educator to detect the students at-risk at an earlier stage and offer timely academic assistance. In addition, it can help curriculum developers to determine whether the existing course structures are helpful in achieving the desired learning outcomes among the students.

The other significant thing about this research is that it can be used in the local educational environment. Most global literature that investigates academic performance with statistical models has been done, but none of this work has concentrated on mathematics students of Pakistani universities. Different countries vary in terms of the educational systems, approaches to assessment and the learning environment, hence there is need to carry out localized research. The results of this study are thus hoped to be useful to the policymakers, administrators and teachers in the national higher education system.

In addition, quantitative research methods in this study enhance the incorporation of statistical analysis in the evaluation of education. This study motivates the use of

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

evidence-based methods in educational planning and decision-making by showing how the regression techniques could be applied to the analysis of academic data. These methods are needed to enhance both institutional effectiveness and student achievement in a more competitive learning environment.

The key questions to be addressed in the proposed study include investigation of the correlation between the performance in advanced mathematics courses and the overall CGPA, and determination of most significant some of the courses among the chosen ones, as well as determination of the predictive power of the regression model. The research also seeks to make recommendations on how teaching and learning practices can be improved based on findings that are obtained through the study.

Overall, the research is aimed at making a contribution to the concept of the academic performance in undergraduate mathematics programs and examining the influences of the most important theoretical and applied courses on the final CGPA of students. The study gives a systematic and objective evaluation of academic achievement through multiple linear regression analysis and official examination data. It is hoped that the results will help educators and administrators improve the design of the curriculum, instructional and student support services and eventually, achieve better outcomes in education.

Methodology

Research Design

This study adopts a quantitative research design to analyze the relationship between students' performance in selected mathematics subjects and their overall academic achievement. Multiple linear regression analysis is employed to model and quantify the effect of several independent variables on a single dependent variable.

Data and Variables

The dataset consists of academic records of undergraduate mathematics students.

Dependent Variable:

Y = Combined CGPA of the 7th and 8th semesters

Independent Variables:

- X_1 = Marks in Functional Analysis
- X_2 = Marks in Numerical Analysis II
- X_3 = Marks in Rings and Fields

Multiple Linear Regression Model

The general multiple linear regression model is defined as:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \varepsilon \quad (1)$$

Least Squares Estimation

The regression coefficients are estimated using the least squares method:

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

$$\min \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2)$$

The least squares method estimates regression parameters by minimizing the sum of squared differences between observed and predicted values.

Analysis of Variance (ANOVA)

The total variation in the dependent variable is decomposed as:

$$SST = SSR + SSE \quad (3)$$

Where:

$$SST = \sum (Y_i - \bar{Y})^2 \quad (4)$$

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2 \quad (5)$$

$$SSE = \sum (Y_i - \hat{Y}_i)^2 \quad (6)$$

SST measures the total variation in the dependent variable. SSR represents the variation explained by the regression model, while SSE measures the unexplained variation due to random error.

Coefficient of Determination

The coefficient of determination is defined as:

$$R^2 = \frac{SSR}{SST} \quad (7)$$

The multiple correlation coefficient is:

$$R = \sqrt{\frac{SSR}{SST}} \quad (8)$$

The adjusted coefficient of determination is given by:

$$\text{Adjusted } R^2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1} \quad (9)$$

Standard Error of Estimate

The standard error of estimate is defined as:

$$SEE = \sqrt{\frac{SSE}{n-p-1}} \quad (10)$$

The standard error of estimate measures the average deviation of observed values from predicted values and reflects the predictive accuracy of the regression model.

Standard Errors of Coefficients

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

The standard error of each regression coefficient is calculated as:

$$SE(b_i) = \sqrt{\text{Var}(b_i)} \quad (11)$$

Standard errors measure the precision of estimated coefficients. Smaller standard errors indicate more reliable estimates.

t-Test for Individual Significance

The statistical significance of each coefficient is tested using the t-statistic:

$$t(b_i) = \frac{b_i}{SE(b_i)} \quad (12)$$

The t-test evaluates whether an individual regression coefficient is significantly different from zero. A statistically significant t-value indicates that the corresponding independent variable has a meaningful impact on the dependent variable.

Confidence Intervals

The confidence interval for a regression coefficient is computed as:

$$b_i \pm t_{\alpha/2} \times SE(b_i) \quad (13)$$

F-Test for Overall Model Significance

The overall significance of the regression model is tested using the F-statistic:

$$F = \frac{SSR/p}{SSE/(n-p-1)} \quad (14)$$

The F-test determines whether the regression model provides a better fit to the data than a model with no predictors. A significant F-value indicates that the independent variables jointly influence the dependent variable.

Multicollinearity Diagnosis

Multicollinearity is assessed using the Variance Inflation Factor (VIF):

$$VIF_i = \frac{1}{1-R_i^2} \quad (15)$$

VIF quantifies the degree of correlation among independent variables. Higher VIF values indicate stronger multicollinearity, which can distort coefficient estimates.

Empirical Results

Regression Model Estimation

The estimated multiple linear regression model obtained from the analysis is:

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

$$Y = -1.1947 + 0.0106X_1 + 0.0225X_2 + 0.0231X_3 \quad (16)$$

The regression coefficients indicate that all three independent variables have a positive relationship with CGPA. Specifically:

A one-unit increase in Functional Analysis marks (X_1) increases CGPA by 0.0106 units, holding other variables constant.

A one-unit increase in Numerical Analysis II marks (X_2) increases CGPA by 0.0225 units.

A one-unit increase in Rings and Fields marks (X_3) increases CGPA by 0.0231 units. Among the three subjects, Rings and Fields has the highest regression coefficient, suggesting it has the strongest marginal impact on CGPA.

The intercept (-1.1947) represents the expected CGPA when all independent variables are zero. Although not practically meaningful in this academic context, it is statistically significant and necessary for model completeness.

Regression Model Summary

Table 1: Model Goodness-of-Fit Statistics

Statistic	Value
R	0.9614
R ²	0.9244
Adjusted R ²	0.9194
Standard Error of Estimate	0.2482

The table shows a very strong model fit, with $R^2 = 0.9244$ indicating that 92.44 % of the variation in CGPA is explained by the predictors. The adjusted R^2 (0.9194) being close to R^2 suggests the model is not overfitted. The standard error of estimate (0.2482) is small, indicating that predicted CGPA values are close to the observed values and the model has good predictive accuracy.

ANOVA Results

Table 2: Analysis of Variance for Regression Model

Source	Value
Total Sum of Squares (SST)	37.4736
Regression Sum of Squares (SSR)	34.6390
Error Sum of Squares (SSE)	2.8346
F-statistic	187.3753
Critical F-value	2.8068

Because $F > F_{\text{critical}}$, the null hypothesis that all regression coefficients are equal to zero is rejected. Therefore, the overall regression model is statistically significant at the 5% level.

Significance of Coefficients

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Table 3: t-test Results for Regression Coefficients

Coefficient	Estimate	Standard Error	t-value	Significant ($ t > 2.0129$)
b_0 (Intercept)	4.7993	1.1972	-4.8709	Yes
b_1 (Functional Analysis)	-0.3818	0.1308	4.7369	Yes
b_2 (Numerical Analysis II)	0.0046	0.0229	4.5296	Yes
b_3 (Rings and Fields)	0.0233	0.0076	3.7860	Yes

The critical t-value at the 5% significance level is 2.0129. Since the absolute value of each calculated t-statistic exceeds the critical value, all regression coefficients are statistically significant.

This means:

Functional Analysis significantly affects CGPA.

Numerical Analysis II significantly affects CGPA.

Rings and Fields significantly affects CGPA.

Thus, each subject independently contributes to predicting students' overall academic performance.

Confidence Intervals of Parameters

Table 4: 95% Confidence Intervals for Regression Coefficients

Parameter	Lower Bound	Upper Bound
b_0	-1.6885	-0.7010
b_1	0.0061	0.0151
b_2	0.0125	0.0325
b_3	0.0108	0.0354

Since none of the intervals include zero, this further confirms the statistical significance of all predictors.

Multicollinearity Diagnosis

The VIF matrix values range approximately between 3.55 and 4.12. Since these values are below the commonly accepted threshold of 5, there is no serious multicollinearity problem among the independent variables.

This indicates that Functional Analysis, Numerical Analysis II, and Rings and Fields contribute independently to the model without excessive overlap in explanatory power.

Discussion

The regression model estimated indicates that all the three subjects namely, Functional Analysis, Numerical Analysis II, and Rings and Fields are positively correlated with CGPA meaning that better marks in these courses increase overall academic performance. The coefficient of Rings and Fields is the highest one among them indicating that it will have the best marginal effect on CGPA. The model has been found to be very good with $R^2 = 0.9244$ and adjusted $R^2 = 0.9194$ meaning that these predictors account more than 92 per cent of the variance in CGPA. The small standard error of estimate (0.2482) implies that the predicted values of CGPA are similar to the values that are observed, which means that there is good predictive power. The results of ANOVA indicate that the total framework is statistically significant at a 5 per cent

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

level since the F-statistic (187.3753) is greater than the critical F-value (2.8068). The t-tests on the individual regression coefficients indicate that all the regression coefficients are significant, which supports the significance of each subject to predict CGPA. All coefficients have confidence limits that do not contain zero as well, which further supports the arguments of their significance. The values of VIF (3.55-4.12) indicate that there is no issue of multicollinearity, and, therefore, the predictors play an independent role in the model. Although the intercept is not practically significant, it is statistically required in order to be complete. Comprehensively, the outcomes show that performance which is subject specific is one of the determinants of academic success. Educators can employ the model to determine the courses which impact most on CGPA. These results offer empirical evidence to the use of specific academic interventions and curriculum planning. To sum up, the subject of advanced mathematics is important in determining overall academic performance of students.

Conclusion

This experiment was aimed at analyzing the correlation between performance of students in Functional Analysis, Rings and Fields, Numerical Analysis II, and performance in overall CGPA. The analysis of the multiple linear regression showed that these subjects are closely related to academic performance and such a relationship is strong and statistically significant. The large value of $R^2 = 0.9244$ shows that the predictors have more than 92 percent of the variations in CGPA. All the regression coefficients were positive and significant, which proves that each of the subjects leads to the creation of meaningful contributions to the overall performance. The the most significant marginal opportunity was enjoyed by rings and fields which underscored the significance of conceptual knowledge of algebra. Functional Analysis and Numerical Analysis II was also influential greatly, and proved the worth of theoretical and computational skills. The general model was of great importance and the diagnostics test showed that there is no severe multicollinearity and low standard error, which means that the predictions are reliable. The results can be used in the academic planning, curriculum development, and specific assistance to students with poor performance in central subjects. Early intervention of said-at risk students depending on performance in the subject can help enhance retention and graduation. Although the study is small (limited by the sample size and data in one institution), it offers a convenient framework to study academic performance. The future study can be extended to a series of universities and extend to academic and non-academic variables. On balance, performance on basic mathematics classes has a good predictive value on student performance, and regression models provide a powerful means of evidence-based decision-making in advanced education.

References

- [1] Kuncel, N. R., Hezlett, S. A., Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all?. *Journal of personality and social psychology*, 86(1), 148.
- [2] Richardson, M., Abraham, C., Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- analysis. *Psychological bulletin*, 138(2), 353.
- [3] National Research Council, Up, A. I. (2001). *Helping Children Learn Mathematics*. Mathematics Learning Study Committee. J. Kilpatrick, J. Swafford, and B. Findell, eds. Washington, DC: National Academy Press.
- [4] Schoenfeld, A. H. (2016). *Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics* (Reprint). *Journal of education*, 196(2), 1-38.
- [5] Biggs, J., Tang, C., Kennedy, G. (2022). *Teaching for quality learning at university* 5e. McGraw-hill education (UK).
- [6] Gibbs, G., Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and teaching in higher education*, (1), 3-31.
- [7] Astin, A. W. (1997). *What matters in college?: Four critical years revisited*. John Wiley Sons.
- [8] Tinto, V. (2012). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago press.
- [9] Cohen, J., Cohen, P., West, S. G., Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.
- [10] Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W. (2005). *Applied linear statistical models*. New York: McGraw, Irwin.
- [11] Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage publications limited.
- [12] Montgomery, D. C., Peck, E. A., Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley Sons.
- [13] James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R* (Vol. 103). New York: springer.
- [14] Peña-Ayala, A. (2014). *Educational data mining*. *Studies in Computational Intelligence*, 524.
- [15] Astin, A. W. (1997). *What matters in college?: Four critical years revisited*. John Wiley Sons.
- [16] Woessmann, L. (2008). *The economics of education. Human capital, family background and inequality*.
- [17] GMR, E. (2015). *Education for all global monitoring report*. Paris: UNESCO and Education for all Global monitoring.
- [18] Ginsburg, M., Archer, D., Barrera-Osorio, F., Lake, L., Vally, S., Wachter, N., Ulrick, J. (2018). CER moderated discussion on World Development Report 2018: Realizing the promise of education for development. *Comparative Education Review*, 62(2), 274-293.
- [19] Woessmann, L. (2008). *The economics of education. Human capital, family background and inequality*.
- [20] <https://exam.usindh.edu.pk/v2/course.php>