Modelling Student Enrolment Patterns in Region 2, The Gambia: A Computational Approach Using Gaussian Distribution, Regression, and Interpolation

Fatty, Fatou K[1], Prof. Liu Changhai[2], Dr. Aboubacar Jallow[3], Dr. Lamin Fatty[4], Mamadou Bah[5], Dawda Charreh[6], Demba Sm Yarbo[7], and Wali Collins[8], Djeneba Diakite[9]

[1]School of Education, Huazhong University of Science and Technology, Wuhan, Hubei Province, P.R. China & The University of Education, The Gambia, West Africa (ORCID: 0009-0001-5150-2807)

[2]School of Education, Huazhong University of Science and Technology, Wuhan, Hubei Province, P.R. China

[3]University of Education, The Gambia (ORCID: 0009-0000-1253-1165)

[4]The University of Law, Birmingham Campus

[5]University of The Gambia, School of Information Technology and Communication (ITC)

[6]University of The Gambia, School of Arts and Science, Department of Mathematics

[7]Registrar, University of Education

[8]University of Education, The Gambia

[9]Huazhong University of Science and Technology Wuhan, Hubei Province, P.R. China

DOI: https://doi.org/10.51584/IJRIAS.2026.11090165

Received: 04 October 2026; Accepted: 12 October 2026; Published: 26 October 2026

ABSTRACT

This study examines the enrolment patterns of senior secondary schools in Region 2 of The Gambia through a computational method that combines descriptive statistics, assessment of the Gaussian distribution, linear regression analysis, and correlation analysis alongside an evaluation of Newton's divided difference interpolation method. Enrolment data from eleven schools were analysed to examine cohort stability between Grade 11 and Grade 12, identify gender disparities, and evaluate the reliability of forecasting techniques. The results display strong gender disparities, with female students consistently outnumbering male students, and female progression being considerably more predictable than that of male students, who recorded high levels of attrition between Grade 11 and Grade 12. Distributional analysis showed that none of the enrolment datasets followed a normal (Gaussian) pattern, since the population of schools in the region is heterogeneous. Despite this non-normality, correlation and regression analyses indicated strong cohort stability, with Grade 11 enrolment explaining more than 75% of the variation in Grade 12 totals. Conversely, high-degree polynomial interpolation demonstrated high numerical instability in line with the Runge phenomenon, which made it inappropriate for enrolment prediction; lower-degree models showed improved stability but remained inferior to regression for forecasting purposes. Overall, the findings indicate the issues of male disengagement, the necessity of gender-based retention strategies, and the usefulness of simple and computationally effective educational planning models. The results of the study bring both new empirical and methodological evidence that will aid in making data-based decisions in the secondary education sector in The Gambia.

Keywords: Student Enrolment, Gaussian Distribution, Regression Analysis, Interpolation, Gender Disparity

INTRODUCTION

Education serves as an indispensable cornerstone of national development, empowering societies through knowledge, innovation, and human capital formation (Gajigo, 2012). However, within the educational landscape of The Gambia, persistent structural challenges most notably gender disparities in educational attainment continue to shape learning outcomes and institutional capacities. At the heart of effective educational administration lies enrolment data, which transcends mere administrative record-keeping to function as a vital instrument for strategic planning, resource allocation, and institutional foresight (Hussain et al., 2023). Enrolment figures offer a window into the operational realities of schooling systems, revealing how demographic distributions, urban-rural divides, and infrastructural variances impact student participation. In regions such as The Gambia's Region 2, where secondary schools span both densely populated urban centers and smaller, resource-constrained towns, enrolment naturally fluctuates. Nevertheless, when these variances measured by a high standard deviation (σ) become excessively wide among comparable institutions, they frequently signal systemic vulnerabilities, including unequal deployment of qualified educators, infrastructural deficits, or socioeconomic barriers that impede consistent attendance. Identifying these disparities equips policymakers with the empirical clarity needed to target interventions effectively.

Beyond institutional capacity, tracking educational progression particularly the cohort stability between Grades 11 and Grade 12 is crucial for monitoring student retention and mitigating dropout crises. Disaggregating this data by gender reveals critical systemic vulnerabilities, such as pronounced male attrition, which underscores the urgent need for targeted interventions aimed at alleviating non-financial socioeconomic pressures (Belanger, 2017). Despite the recognized importance of such data, historical investigations into Gambian educational trends have predominantly relied on basic descriptive counts or proportions, often lacking rigorous statistical validation. Crucially, foundational assumptions regarding data distribution, such as adherence to a Gaussian (normal) model, have frequently remained untested due to a scarcity of advanced computational analyses. Furthermore, previous predictive frameworks have rarely evaluated the mathematical reliability of their forecasting tools; as demonstrated in numerical analysis, high-degree polynomial interpolation at Equi spaced nodes can generate erratic oscillations and unrealistic projections of student populations (Belanger, 2017).

To bridge these analytical gaps, this study undertakes a comprehensive examination of Grade 11 and Grade 12 enrolment patterns in The Gambia, disaggregating data by gender through robust statistical measures, including means and cohort stability metrics. By moving beyond descriptive tallies, the empirical findings highlighted by a baseline regression model for total enrolment and corroborated by lower-degree interpolation tests offer a reliable framework for short-term educational budgeting and the estimation of missing administrative data. Crucially, the analysis uncovers a striking divergence in gender retention: male cohorts exhibit a high and erratic dropout rate ($\Delta\mu = 9.54$), whereas female cohorts demonstrate remarkable stability ($\Delta\mu = 1.09$). This stark contrast not only redefines our understanding of upper-secondary progression in The Gambia but also charts a clear path for evidence-based policymaking, emphasizing the imperative for gender-sensitive, community-specific interventions to safeguard male student retention.

LITERATURE REVIEW

Education enrolment data serves as a fundamental barometer for evaluating the vitality, inclusivity, and overall efficacy of a nation's schooling infrastructure (Altbach, Reisberg, & Rumbley, 2009; Carnoy, 1999; Hanushek & Woessmann, 2008; Schütz, Ursprung, & Wößmann, 2005; World Bank, 2018). Across the diverse and developing landscape of Sub-Saharan Africa, systemic challenges remain stark: fewer than half of all primary-school-age children successfully access formal education (Bentaouet Kattan & Burnett, 2004; Colclough, Al-Samarrai, Rose, & Tembon, 2003; UNESCO, 2015; World Bank, 2011), secondary school enrolment languishes below 20% in no fewer than 22 countries (Lewin, 2008; Oketch, 2016; UNESCO Institute for Statistics, 2018; World Bank, 2005), and fewer than 10% of the active workforce attain matriculation or upper-secondary completion (Filmer & Fox, 2014; Fox et al., 2013; Govender & Farlam, 2004; Johanson & Adams, 2004). Within The Gambia specifically, net enrolment at the primary level hovers around 65% having demonstrated a modest upward trajectory of approximately 1% per annum over the preceding two decades while secondary education historically lagged significantly behind at roughly 25% for high school students (Republic of The Gambia, 2016;

World Bank, 2012, 2020), a figure that contemporary estimates suggest has likely risen above 50% through sustained governmental and international interventions (AfDB, 2021; UNESCO, 2022). Far from functioning as inert administrative tallies, these enrolment metrics offer profound diagnostic insights into systemic efficiency, spatial and socio-economic access, and distributional equity (Bruns, Mingat, & Rakotomalala, 2003; Gajigo, 2012; Glewwe & Kremer, 2006; Hussain et al., 2023; Mingat, 2007). Nevertheless, translating these longitudinal metrics into reliable predictive insights is fraught with methodological hurdles; complex, high-degree numerical techniques such as high-order polynomial interpolation frequently exhibit extreme mathematical volatility and unpredictable behaviour, culminating in highly unrealistic student-count projections (Atkinson, 2008; Belanger, 2017; Burden & Faires, 2010; Cheney, Kincaid, & Clegg, 2012). To circumvent these spurious oscillations, robust regression analyses and lower-degree interpolation models provide substantially more dependable frameworks for charting historical trajectories, forecasting future enrolment demands, and ultimately furnishing policymakers with the empirical clarity required for evidence-based educational planning (Chatterjee & Hadi, 2015; Draper & Smith, 1998; Gujarati & Porter, 2009; Hyndman & Athanasopoulos, 2018; Montgomery, Peck, & Vining, 2021; Weisberg, 2005).

METHODOLOGY

3.1 Data Acquisition and Preprocessing Protocol

Enrolment data from 11 senior secondary schools in Region 2, The Gambia, were collected and recorded in an Excel sheet. We conducted all the analaysis in MATLAB using an Acer Aspire 3 laptop (16GB Ram, 12th Gen i5 and 477GB SSD), which provided a controlled environment that supported the reproducibility of results. We used standard software and functions to make sure that the variability which may arise due to different computational setups is minimised.

We imported the raw enrolment data into our workspace using MATLAB's readable function, making sure that the original structure and metadata of the source Excel sheet are preserved. We then went on to remove non-data rows, including the sub-titles of each column and totals, using their specific indexes.

enrolment data = enrolment data (2: end-1, :);

With this approach, we made sure that only individual school-level records were included in distribution modelling, which helped us avoid contamination from rows showing the total enrolment across the schools.

All the enrolment data we imported was in string format, so we used MATLAB's str2double to convert the columns for Grade 11 Male, Grade 12 Male, Grade 11 Female and Grade 12 Female to numeric vectors.

3.2 Gaussian Distribution and Descriptive Analysis

During this phase, we compute the mean of each of the 4 columns, Grade 11 Male, Grade 11 Female, Grade 12 Male, and Grade 12 Female, and their various totals using MATLAB's inbuilt mean () function. We also used std () to find the standard deviation of each of the groups mentioned above.

3.3 Procedural Uniformity

To perform all descriptive and normality analyses, we implemented a single function, which was named analyse distribution, the analyse distribution function takes the data of a particular group (Grade 11 Male, Grade 11 Female, Grade 12 Male, Grade 12 Female, and the total enrolment for both Grade 11 and Grade 12) and computes the basic statistics. We then used the analyse_distribution function on each of the groups. This was done to ensure that parameter estimation, theoretical modelling, and diagnostic plotting were performed consistently across all datasets, preventing methodological variation and supporting computational reproducibility.

3.4 Central Tendency and Dispersion

For every dataset, the sample mean (μ) and sample standard deviation (σ), the two most crucial parameters for the normal data model, were calculated. They provide an overview of each grade-gender group's average enrolment count (μ) and degree of variability (σ).

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

Sample Mean (μ)

For a dataset of n enrolment values

x1, x2, ..., xn

The sample mean is defined as:

μ=1/n ∑_(i=1)^n▒x_i

Where xi represents the enrolment value of the i-th school. ∑_(i=1)^n▒x_(i) represents total enrolment summed across all schools. n is the number of valid school records, and dividing by n gives the average enrolment per school, representing the central location of the distribution.

mu = mean(data);

Sample Standard Deviation (σ)

The spread of the data was quantified using the standard deviation:

σ=√(1/(n-1) ∑_(i=1)^n▒(x_i-μ)^2)

Each school's enrolment deviation from the mean is represented by the term $(xi - \mu)$, whereas the squared deviation $(xi - \mu)^2$ guarantees that every deviation adds positively to the measure of variability. The total variability in the dataset is captured by adding these squared deviations for each school. The unbiased estimator of the population variance is obtained by dividing this sum by $n - 1$ instead of n. The variance is then returned to the original scale of the data, expressed in the number of students, by taking the square root, which yields the standard deviation.

3.5 Theoretical Gaussian Fit

To visualise how closely each dataset followed a normal distribution, we constructed a theoretical Gaussian curve. MATLAB's norm generated the Normal Probability Density Function over a grid of 100 evenly spaced points spanning the observed data range using the estimated μ and σ.

x = Lin space(min(data), max(data), 100);

y = norm pdf (x, mu, sigma);

The above code produced a smooth curve representing the idealised distribution implied by the empirical parameters.

3.6 Normality Assessment (Q–Q Plot)

A quantile–quantile (Q–Q) plot was used as the primary tool for assessing normality. MATLAB's qqplot function compares the quantiles of the observed enrolment data with the quantiles of a perfect Gaussian distribution. If the plotted points lie along a straight diagonal, then the enrolment values follow a normal pattern. Meanwhile, systematic deviations, curvature, s-shaped patterns or heavy tails indicate departure from normality.

3.7 Predictive Modelling: Regression and Correlation Analysis

This section describes the statistical procedures that quantify the cohort stability between Grade 11 and Grade 12. The focus of this analysis is determining how strongly Grade 11 enrolment predicts Grade 12 enrolment for male, female and combined groups.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

Predictive modelling was used to assess whether Grade 11 enrolment serves as a trustworthy predictor of Grade 12 enrolment across schools. Enrolment in Grade 11 was the independent variable in this framework, and enrolment in Grade 12 was the dependent variable.

Pearson Correlation Coefficient (r)

$r=(\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}))/\sqrt{(\sum_{i=1}^{n}(x_i-\bar{x})^2 \sqrt{(\sum_{i=1}^{n}(y_i-\bar{y})^2))}}$

Where x_i is the grade 11 enrolment for school i, y_i is the grade 12 enrolment for school i. $\bar{x}$ represents the mean for Grade 11 enrolment, and $\bar{y}$ is the mean for Grade 12.

The numerator,

$\sum(x_i-\bar{x})(y_i-\bar{y})$

measures the covariance—whether increases in Grade 11 enrolment are associated with increases in Grade 12 enrolment.

The denominator rescales the covariance using the product of the standard deviations of both variables:

$\sqrt{(\sum(x_i-\bar{x})^2)}$ "and" $\sqrt{(\sum(y_i-\bar{y})^2)}$

The correlation coefficient r ranges from −1 to +1, indicating both the direction and strength of the relationship between two variables (Akoglu, 2018).

Coefficient of Determination (R^2)

$R^2=1-(SS_{"res"})/(SS_{"total"})$

Where the Residual Sum of Squares (model error)

$SS_{"res"} = \sum_{i=1}^{n}(y_i-(y_i))^2$

And the total sum of squares (total variation around the mean)

$SS_{"total"} = \sum_{i=1}^{n}(y_i-\bar{y})^2$

$\hat{(y_i)}=ax_i+b$ predicts the Grade 12 Value from the linear regression, where a is the slope and b is the intercept.

SS_total measures how much Grade 12 enrolment varies across schools, while SS_res measures the amount of variation not explained by the model.

The coefficient of determination, R^2, measures how well the predictor variable explains the variability of the response variable. An $R^2=0$ indicates that the predictor explains none of the variability, while an $R^2=1$ means it explains all the variability. Higher, stronger predictive power and more stability in the relationship between the variables are indicated by higher R^2 values.

Separate regression and correlation models were estimated for the following pairs:

Grade 11 Male → Grade 12 Male

Grade 11 Female → Grade 12 Female

Total Grade 11 → Total Grade 12

A more accurate analysis of gender-specific progression patterns were made possible by the disaggregation, which also made it possible to evaluate the cohort progression's consistency across demographic groups.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

To ensure statistical rigour and predictive performance transparency, each model included a scatterplot, a best-fit regression line, and the computation of r and R^2.

3.8 Numerical Analysis: Newton's Divided Difference Interpolation

When estimating hypothetical Grade 12 enrolment values from observed Grade 11 data, analysts are frequently confronted with a common hurdle: unevenly spaced data points. To overcome this, a method capable of handling irregular intervals is essential.

Newton's Divided Difference Interpolation was selected for this task precisely because it gracefully supports non-uniform x-values and constructs polynomials recursively, making it exceptionally well-suited for systematic stability testing. By leveraging this approach, the analysis establishes a controlled, highly adaptable environment to rigorously test whether higher-degree polynomials can produce stable and realistic enrolment predictions.

3.9 Newton's Divided Difference Polynomial Structure

General Polynomial Form

$$P(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \cdots + a_n(x - x_0)(x - x_1) \dots (x - x_{n-1})$$

The values $x_0, x_1, \dots, x_{n-1}$represents the Grade 11 enrollment values after removing duplicates, while $a_0, a_1, \dots, a_{n-1}$ are the divided-difference coefficients computed recursively from the Grade 12 values. In the Newton interpolation formula, each successive term introduces an additional factor $(x - xi)$,which increases the degree of the polynomial and builds the interpolating function step by step.

Recursive Divided Difference Formula

The coefficient *aj* is obtained from:

$$f[x_i, x_{i+1}, \dots, x_{i+j}] = \frac{f[x_{i+1}, \dots, x_{i+j}] - f[x_i, \dots, x_{i+j-1}]}{x_{i+j} - x_i}$$

This recursive structure is the basis for the MATLAB implementation used in the analysis, where the difference table *F* is constructed, and the diagonal elements form the coefficients.

3.11 Systematic Degree Evaluation

Polynomials of three different complexities were constructed for male, female and total groups, to diagnose which polynomial degree behaves reliably. A full degree polynomial which uses all the unique data points, which gave the highest complexity, and a third-order polynomial which uses the first four data points and a second-order polynomial which uses the first 3 data points were all used with the below code:

```
male_est_full = newton_interp (grade_11_male_unique, grade_12_male_unique, 64);

male_est_3    = newton_interp(grade_11_male_unique (1:4), grade_12_male_unique (1:4), 64);

male_est_2    = newton_interp (grade_11_male_unique (1:3), grade_12_male_unique (1:3), 64);
```

The goal here was not only to establish values but also to observe how polynomial complexity affects numerical stability when dealing with enrolment counts.

3.10 Diagnosis of Runge's Phenomenon

The full degree polynomial was purposefully used to test for Runge's phenomenon, a problem where a high-degree polynomial results in extreme over- or under-estimation, large oscillations close to the endpoints, unstable

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

or unrealistic function shapes, and extreme over- or under-estimation. Runge's instability arises because each added polynomial term amplifies sensitivity to small changes in the input data.

3.12 Practical Stability Assessment

The lower-degree polynomials (2nd and 3rd order) produced smooth, realistic curves, avoided negative or extreme values, and also maintained a shape which is consistent with actual enrolment patterns.

These outcomes show that lower-degree models provide the most stable and also practically meaningful estimates for administrative forecasting.

That is why the low-degree interpolation results were adopted as the preferred practical models, while the full-degree polynomial served as a diagnostic tool for identifying instability.

RESULTS

This chapter presents the results obtained from analysing the enrolment patterns of students in Region 2 using descriptive statistics, Gaussian distribution assessment, regression modelling, and interpolation techniques.

4.1 Descriptive Analysis: Measures of Central Tendency and Dispersion

This section presents the mean (μ) and standard deviation (σ) values for Grade 11 and Grade 12 enrolment for male, female, and total cohorts. Table **4.2 summarises all computed descriptive statistics.**

The results show that the mean enrolment for female cohorts is consistently higher across both grades compared to their male counterparts. In Grade 12, the female mean enrolment ($\mu \approx 63$) exceeded the male mean ($\mu \approx 40$) by approximately 22.91 students.

A comparison of Grade 11 and Grade 12 averages indicates a noticeable decline in male enrolment, with an average reduction of 9.54 students, while female enrolment saw a smaller decrease of 1.09 students.

The standard deviation measuring the variability levels shows that female cohorts had the highest dispersion ($\sigma \approx 25$), indicating substantial variability across schools. Grade 12 male enrolment had the lowest dispersion ($\sigma = 14.1338$), revealing a more clustered distribution.

Group	Mean (μ)	Std Dev (σ)
Grade 11 Male	49.3636	18.9962
Grade 11 Female	63.8182	26.4001
Grade 12 Male	39.8182	14.1338
Grade 12 Female	62.7273	24.9844
Grade 11 Total:	113.1818	42.1042
Grade 12 Total:	102.5455	37.8084

Table 4.1: Summary of Descriptive Statistics for Enrollment Distributions

4.3 Gaussian Distribution Assessment

The distributional characteristics of the enrollment data were assessed using separate histograms and Q-Q plots for each of the six cohorts: Grade 11 male, Grade 11 female, Grade 11 total, Grade 12 male, Grade 12 female, and Grade 12 total.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

The six histograms are shown in Figure 4.1a-4.1f, each overlaid with the theoretical Gaussian probability density function calculated with the help of the corresponding mean (μ) and standard deviation (σ).

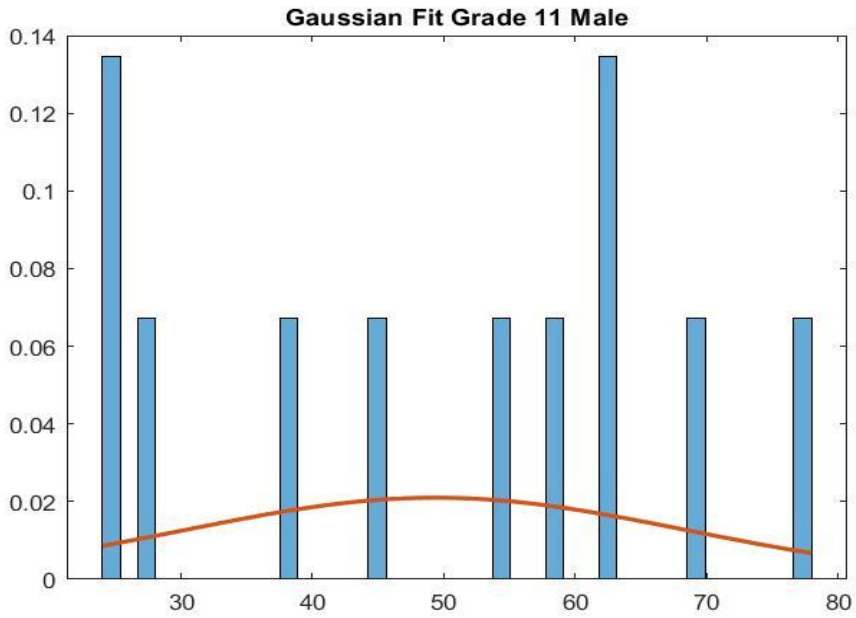


Fig 4.1(a)

The Grade 11 male enrolment histogram in Fig 4.1(a) had discrete spikes at enrolment at specific enrolment values. These peaks are not coincidental with the superimposed Gaussian curve, which denotes Grade 11 Male. The data of enrolment is not normallydistributed.

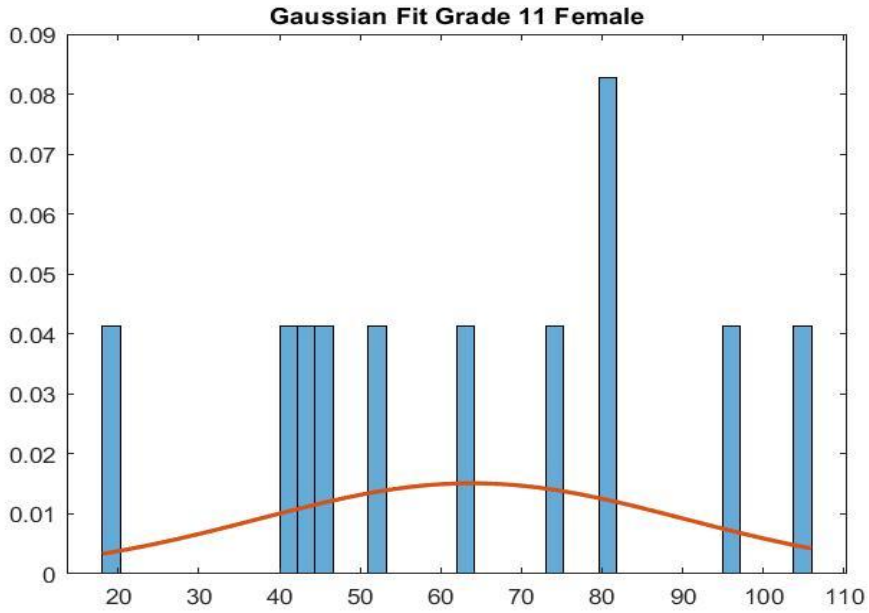


Fig 4.1(b)

Fig 4.1(b) represents the distribution of Grade 11 female enrolment and a fitted Gaussian curve. The histogram indicates the observed frequencies of the enrolment data, which are concentrated around several different peaks, notably in the 40-50 and 80 ranges. Theoretical normal distribution curve, which is superimposed in red, is a poor fit. is an approximation of the observed data, and this indicates that the values of the enrolment are not normally distributed.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

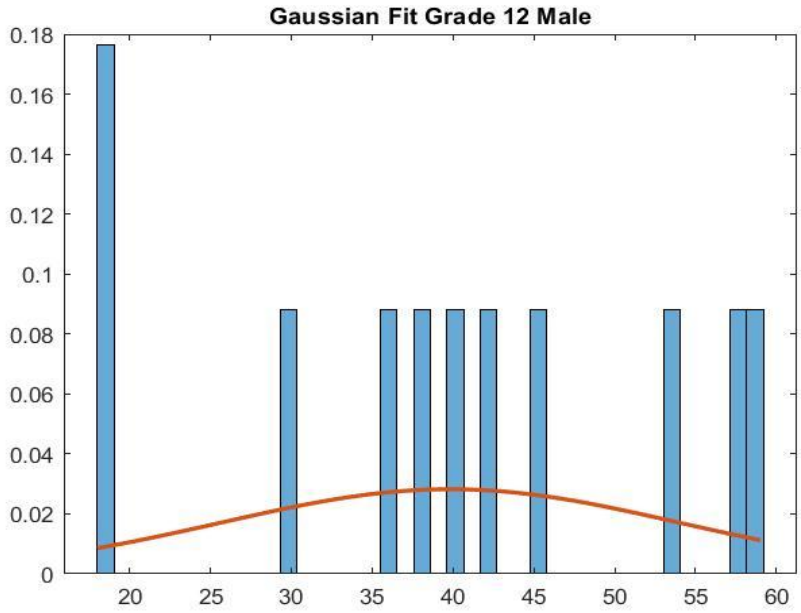


Fig 4.1(c)

The frequency distribution of the Grade 12 male in Fig 4.1(c) indicates that the frequency is mostly concentrated on the left side of the graph, in bin (18-20), with the highest frequency, which is about 0.18. Once the first spark is made, the rest of the data is. somewhat evenly distributed between the limits of 30 and 60. The superimposed theoretical normal distribution curve. The observed data is not well approximated by the red line. The curve grossly underestimates the peak at 18-20. and exaggerates the frequency in the 20-30 range, where no data were taken. This tendency indicates that the data is not normally distributed; it has highly skewed or bimodalcharacteristics.

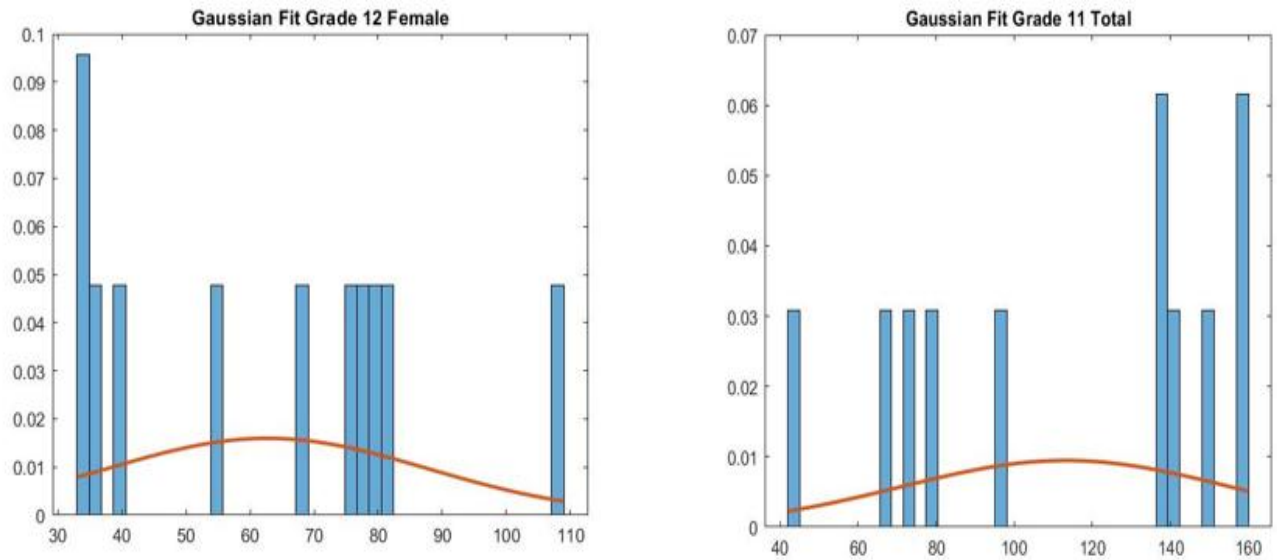


Fig 4.1(d)

Fig 4.1(d) and a represent the distribution of enrollment data of the Grade 12 Female students with a fitted Gaussian curve.

The histograms indicate the observed frequencies of enrolment grouped in a few discrete, separate bins. spanning the range of 30 to 110. The utmost proportionate frequency (about 0.095) is in the first. enrolment bin (around

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

33). There are also other prominent clusters between the range of 75 and 83 with four adjacent bins. have a frequency of the order of 0.05, and there is yet another peak at about 110. The data is characterised by a wide spread and multiple gaps where no enrolment figures were recorded. The theoretical normal distribution curve, overlaid in red, poorly approximates the observed data. The curve greatly underestimates the actual distribution of enrolment numbers. The frequency of enrolment numbers is low (30s) and middle-high (75-83), and the number of overestimates the data. frequency between the intermediate ranges that lack data. This noted trend indicates that the enrolment data for Grade 12 females are not normally shaped and take the form of a multimodal distribution.

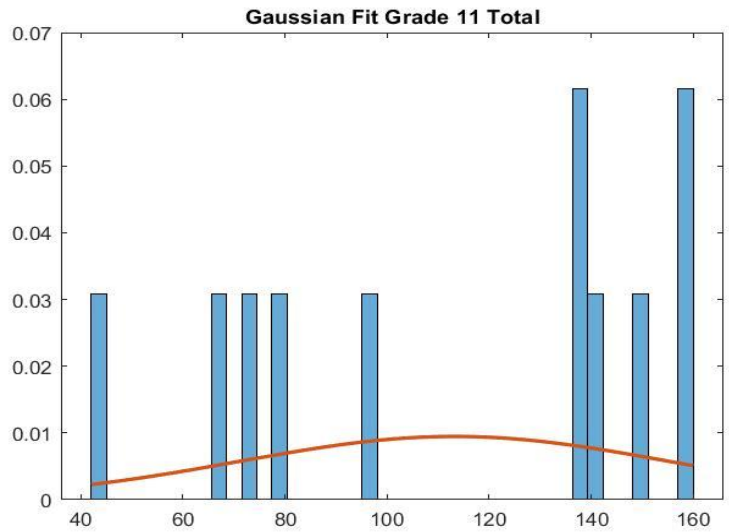


Fig 4.1(e)

The total enrolment information of Grade 11 participants (males and females combined) is plotted in. Fig. 4.1(e) and a fitted Gaussian curve.

The histogram indicates three major groups of enrolment values, with one low value around 40-45 and a second one around 65-100, with an approximate range of 65-100 and a large mass of the third around the 135-160 region. The upper cluster at bins around 140 and 160 shows the two highest relative frequencies (approximately 0.061). The data is characterised by a large range, large inner gaps, and a highly skewed distribution. The red theoretical normal distribution curve does not fit the observed data well. The red theoretical normal distribution curve approximates the observed data poorly. The fitted curve undervalues the frequencies of all three large data clusters and overvalues the frequency in the range of 110–130, where no enrolments were recorded. This observed pattern indicates that the total enrolment data are not normally distributed and suggests a multimodal or extremely irregular distribution.

Fig 4.1(f)

Fig. 4.1 (f) shows the distribution of total enrolment data for grade 12 and a fitted curve.

The figures of enrolment are grouped into a few discrete, discontinuous units within the scale of 50-170, as shown by the histogram. The data are skewed towards the lowest bin that illustrates a peak of approximately 55. maximum relative frequency (less than 0.062). After this first blossom of heaps, other small, thick groups. are clustered around 75-85 and four tightly clustered bins and around 115-125, and frequencies of approximately 0.03. The rest of the data points are on a higher side (140 and 165). The theoretical normal distribution curve, superimposed in red, poorly approximates the data. The fitted curve significantly underestimates the actual frequency of enrolment figures in the low (55) and middle-high (115–125) clusters while overestimating the frequency in the range of 90–110, where no data was recorded. This observed pattern indicates that the total enrolment data does not follow a normal distribution and exhibits characteristics of a highly skewed or multimodal distribution. **Q-Q Plots**

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

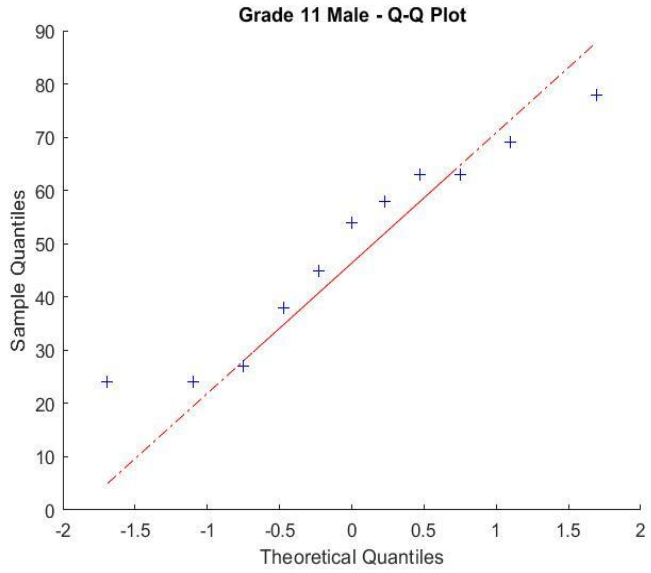


Fig 4.2(a)

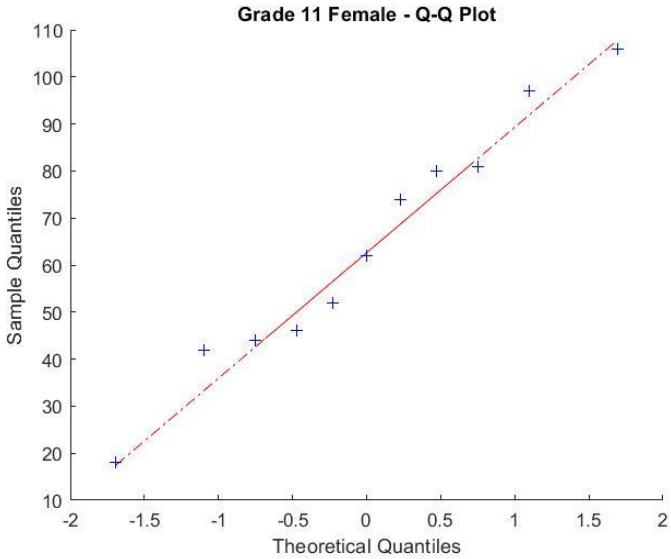


Fig 4.2(b)

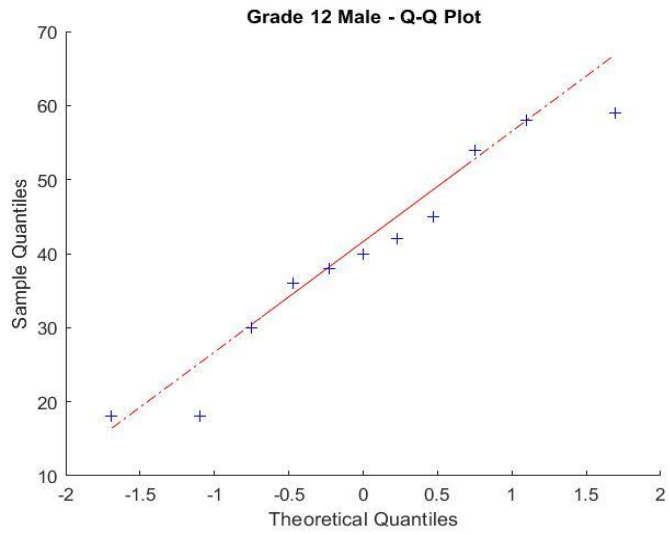


Fig 4.2(c)

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

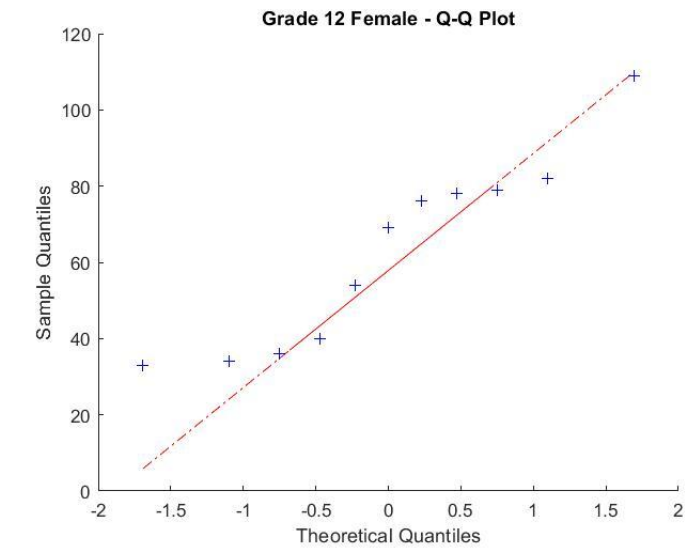


Fig 4.2(d)

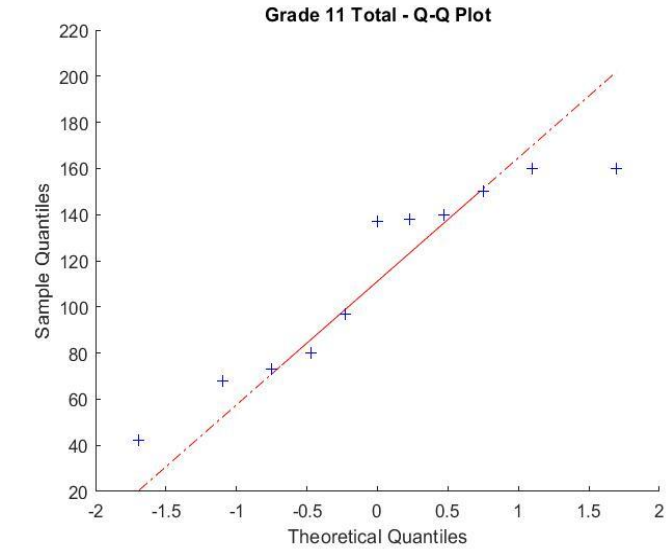


Fig 4.2(e)

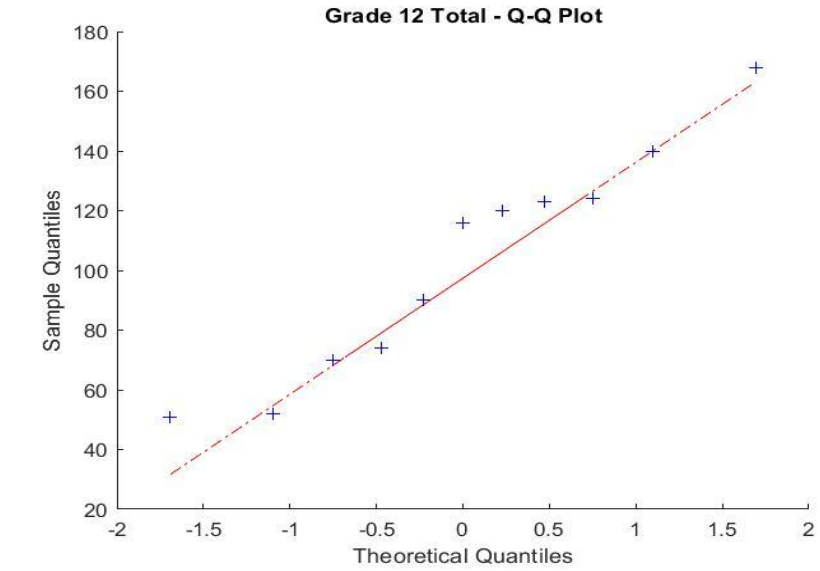


Fig 4.2(f)

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

The quantile-quantile (Q-Q) plots assessed the normality of the enrolment data of each subgroup, as presented in Figures 4.2a–4.2f. Examination of this plot shows that the enrolment data generally deviates from a theoretical normal distribution, where a close adherence of the sample quantiles to the straight line would indicate normality.

For the Grade 11 Male data shown in Fig. 4.2(a), the plot shows a pronounced S-shaped deviation. The data points on the lower (left) fall significantly below the theoretical line, which indicates the presence of heavier tails than a normal distribution and potential left-skewness due to the low point deviation. The Grade 11 female in Fig. 4.2(b) data provides the closest approximation to normality among the individual subgroups, as the majority of the central data points closely follow the theoretical line. However, minor deviations are still observed in some areas, suggesting slight non-normality in the tails. When these two subgroups are combined into the Grade 11 total data, the plot shows a clear overall deviation from the straight line, particularly at the extremities where the lower and upper points tend to fall below the line, suggesting the combined distribution still possesses heavier tails than the theoretical normal distribution. In the Grade 12 cohort, the Grade 12 Male data exhibits non-normality primarily due to a single data point on the far left (around Theoretical Quantile -1.5) that falls significantly below the line, suggesting a left-sided outlier or right-skewness impacting the distribution. The Grade 12 Female data display a distinct concave shape, where the middle points fall below the theoretical line while the lower (left) and upper (right) points are significantly above it. This pattern strongly indicates a heavy-tailed or right-skewed distribution. Finally, the Grade 12 Total data show the strongest linear pattern across all groups, with the majority of the data points lying very close to the theoretical straight line. While minor, negligible deviations are still observed at the extreme ends, the combined Grade 12 enrollment data approximates a normal distribution better than the Grade 11 combined data.

Overall, the Q-Q plots consistently show that all individual enrollment subgroups are not perfectly normally distributed, with varying patterns of skewness and heavy tails, though the Grade 11 Female and Grade 12 Total groups show the highest degree of linearity.

4.5 Correlation Analysis: Grade 11 vs. Grade 12 Enrollment

To assess the relationship between enrollment figures in Grade 11 and the corresponding enrollment figures in Grade 12, a Pearson product-moment correlation coefficient (r) was calculated for male, female, and total enrollment data. Scatter plots were generated to visually represent these relationships, with a linear best-fit line overlaid.

4.4 Male Enrolment Correlation

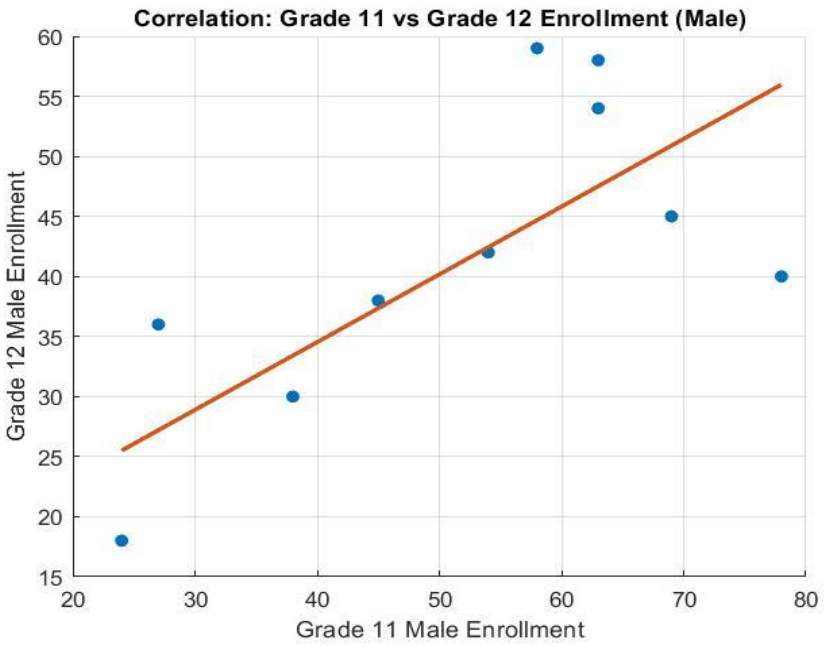


Fig 4.3(a)

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

When the correlation of male enrolment was analysed, a strong and positive linear relationship was found between grade. The 11 and Grade 12 registered numbers have a Pearson correlation coefficient of r = 0.7582. This result indicates that higher Grade 11 male enrolment is consistently associated with higher enrolment in Grade 12 male cohorts. The scattered plot shown in Fig. 4.3(a) confirms this, as most data points tend to follow an upward trend.

4.6 Female Enrolment Correlation

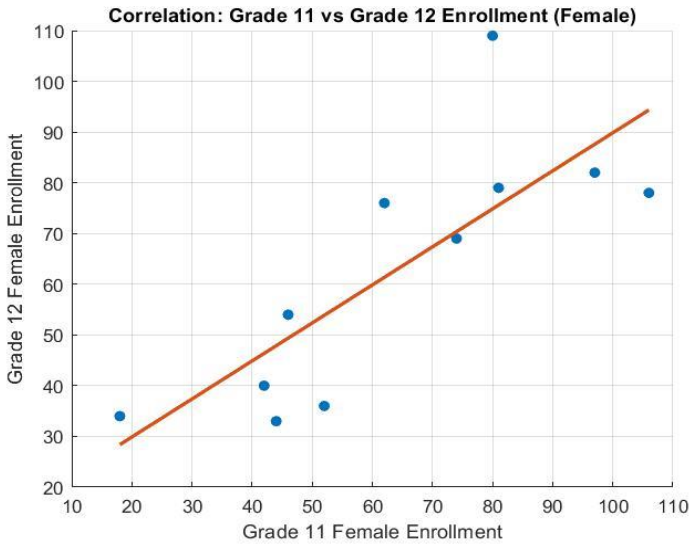


Fig 4.3(b)

A strong, positive linear relationship was found for Female Enrollment, yielding a correlation coefficient of r = 0.7921. The value is even higher than that of the male enrollment, suggesting a tighter association between Grade 11 Female enrollment and Grade 12 Female enrollment. *Fig 4.3(b)* confirms this as the data points follow an upward trend, close to the line of best fit.

4.7 Total Enrolment Correlation

The correlation for Total Enrolment (combined male and female) exhibited the strongest positive linear relationship of all groups, with r = 0.8673. This result indicates a very high degree of association, where the Overall enrolment in Grade 11 is highly predictive of the overall enrolment in Grade 12. As shown in the final scatter plot in Fig. 4.3(c), the data points are highly concentrated along the linear best-fit line, visually confirming a forceful positive correlation.

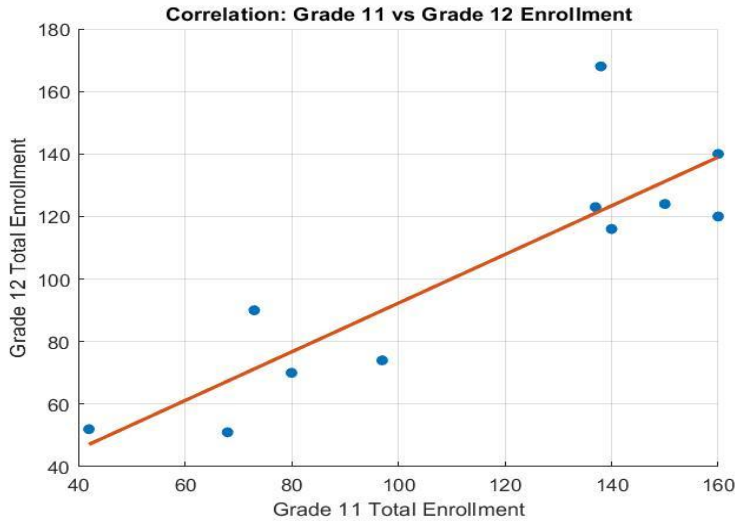


Fig 4.3(c)

4.8 Regression Analysis: Grade 11 Predicting Grade 12 Enrolment

A linear regression analysis was performed to determine to what extent Grade 11 enrolment figures predict Grade 12 enrolment figures for the male, female and total student populations. The coefficient of determination (R^2) was calculated for each model, showing the proportion of the variance in Grade 12 enrolment that is predictable from the Grade 11 enrolment data. The results for all three regressions are visualised in Figure 4.4.

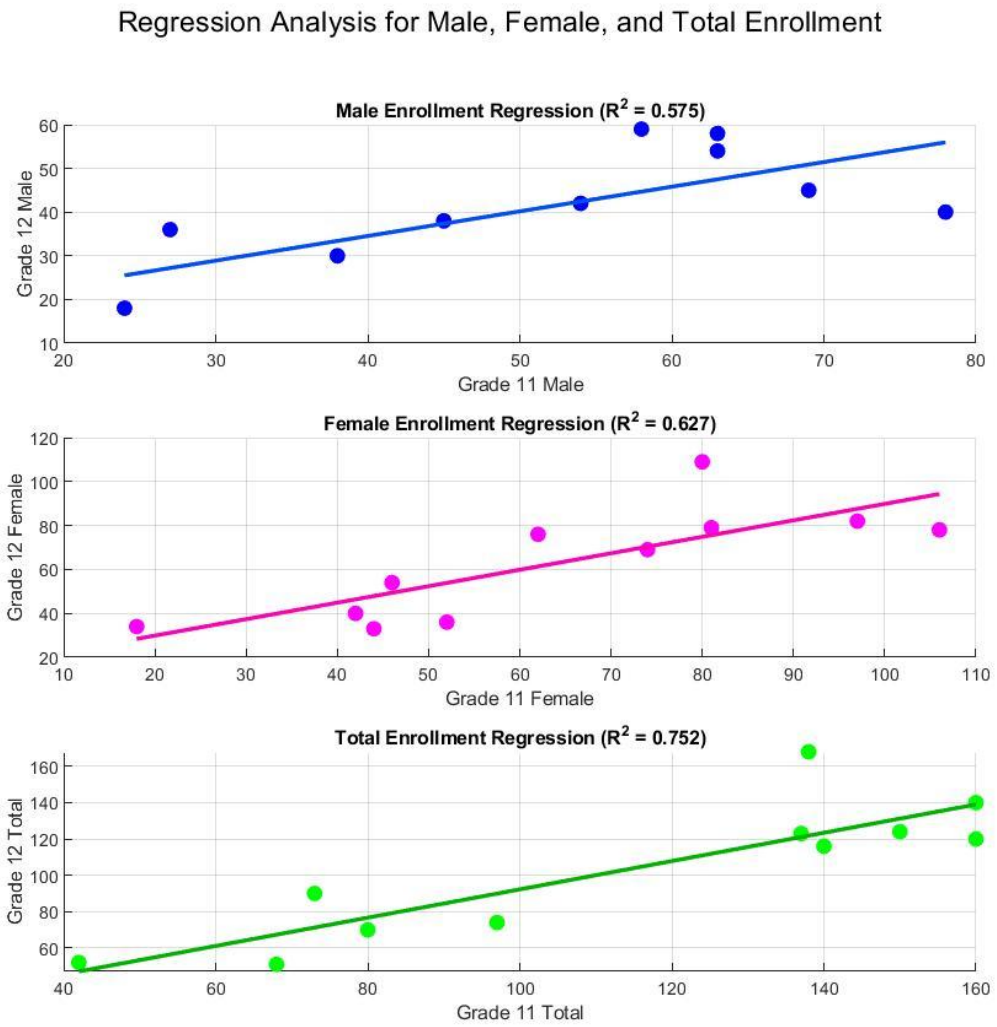


Fig 4.4

The regression analysis of male enrolment shows a moderately strong predictive relationship. The model yielded an R^2 score of 0.575 for male enrolment, indicating that approximately 57.5% of the variance in Grade 12 male enrolment is explained by the variance in Grade 11 male enrolment.

The model for female enrolment demonstrated a slightly stronger predictive power than the male model. The calculated R^2 value was 0.627, meaning that approximately 62.7% of the variance in Grade 12 female enrolment is accounted for by the Grade 11 female enrolment figures. The graphical illustration depicts a positive linear inclination, with the values tending to concentrate more along the regression line than in the male data.

The best predictive relationship among the three models was on the regression analysis of Total Enrolment (male and female combined). The model resulted in an R^2 value of 0.752, meaning that about 75.2 percent of the change in Grade 12 total enrolment can be accounted for based on the figures of Grade 11 total enrolment. The positive linear pattern that is tight in the scatter plot in Fig. 4.4 indicates that total Grade 11 enrolment is a strong predictor of total Grade 12 enrolment.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

4.9 Enrolment Data Interpolation Analysis

Newton's divided difference polynomial interpolation was performed to model the relationship between Grade 11 enrolment (x) and Grade 12 enrolment. Interpolation curves were generated using only the first three points, then the first four points and also using the full set of unique data points. The accuracy of this model was tested by estimating Grade 12 values at specific Grade 11 values.

4.10 Male Enrolment Interpolation

The male interpolation plot shown in Fig. 4.5(a) shows that the full-order polynomial (solid line) accurately fits the data points but exhibits large oscillations between points, particularly at the right extremities. This high-degree oscillation is characteristic of Runge's phenomenon and suggests the full polynomial is unreliable for predicting values between the data points. For the estimation point x=64, the full order estimate was 54.14 (full est.), while the 3rd-order and 2nd-order estimates (3rd est. and 2nd est.) were 56.75 and 54.97, respectively. For the estimation point x=45, the full-order estimate was 42, and the 3rd-order and 2nd-order models both estimated 38.

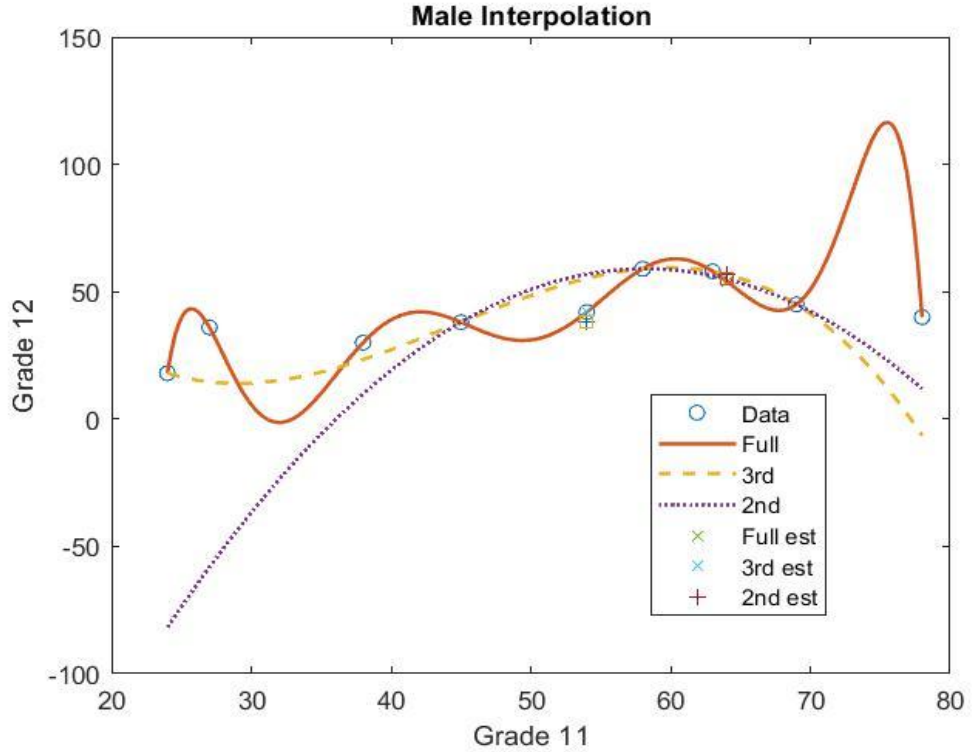


Fig 4.5(a)

4.11 Female Enrolment Interpolation

The female interpolation plot shown in *Fig 4.5(b)* demonstrate the most extreme case of Runge's phenomenon. The Full-order polynomial (solid line) produces massive oscillations with a peak Grade 12 estimate of approximately 3.6 x 10^4 (36,000) around a Grade 11 enrollment of 23. Given the actual Grade 12 values are small, the Full-order polynomial is entirely unsuitable for this data set. The 3rd-order and 2nd-order polynomials (dashed and dotted lines) provide much more stable and near-zero approximations across the central data range.

For the estimation point *x=50*, the Full-order estimate was 52.80, while the 3rd-order and 2nd-order estimates were much lower at 22.96 and -36.68, respectively. For the estimation point *x=52*, the Full-order estimate matched the actual data point at 36, while the 3rd-order and 2nd-order estimates were also 36, respectively, demonstrating that for this specific point, all three models yielded an accurate result.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

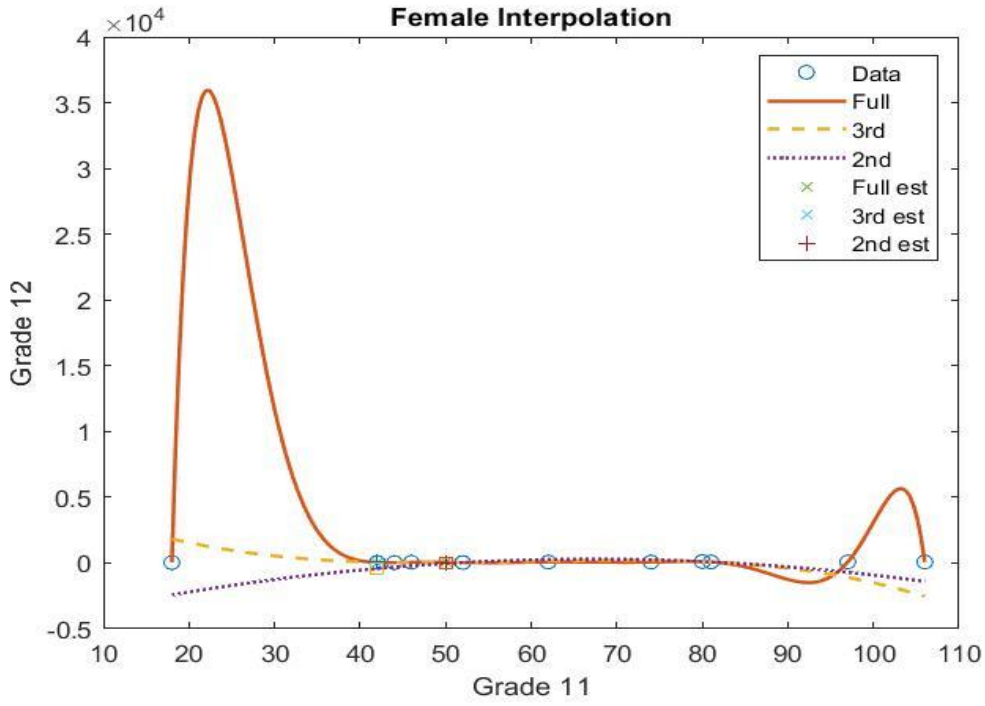


Fig 4.5 (b)

4.12 Total Enrolment Interpolation

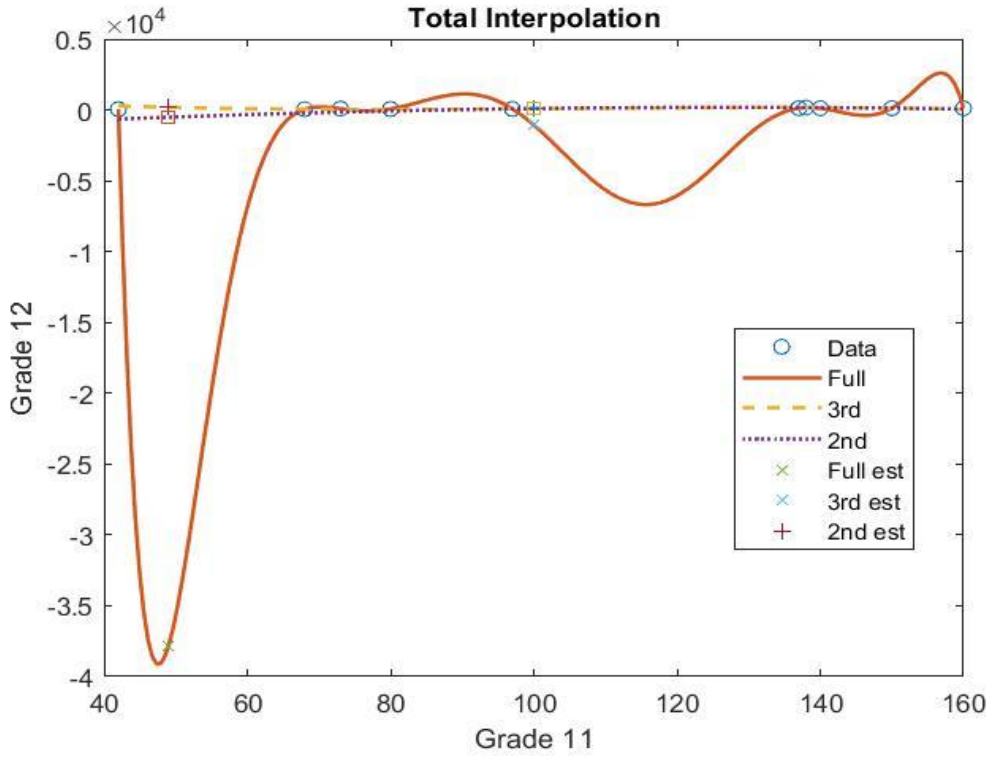


Fig 4.5 (c)

The Total Interpolation plot *Fig 4.5 (c)* also shows severe limitations. The Full-order polynomial (solid line) exhibits extremely large oscillations, with a trough dropping to approximately -3.8 x 10^4 at the lower end of the Grade 11 scale, making it unsuitable for interpolation.

For the estimation point *x=49*, the Full-order estimate was a highly inaccurate -37,853. The 3rd-order and 2nd-order models provided estimates of -1,068 and 188, respectively. At the estimation point, x=100, Full-order

estimated 84.39, 3rd-order and 2nd-order model estimated -516.39 and 93.70, respectively, once again, indicating a poor and unreliable predictive factor.

DISCUSSION

This paper assesses enrolment in 11 senior secondary schools within Region 2 of The Gambia using descriptive statistics, normality tests, correlation, regression analysis and polynomial interpolation. In this discussion, we explain the findings, place them in the context of literature, draw policy implications, discuss limitations and outline future research prospects.

5.1 Interpretation of Descriptive Patterns

The descriptive statistics show that there is an enduring gender disparity where male students outnumber the female students in Grade 11 and Grade 12, as well as in Grade 11 versus Grade 12 male attrition is significantly higher than female attrition. This aligns with recent patterns in The Gambia, where access and retention of girls have improved through programmes such as girls' scholarships and girl sensitisation (Gajigo, 2016). This is reinforced in the report of the United Nations in 2024: girls are now more numerous than boys in certain Gambian secondary schools, and it has been raised that there is a boy crisis in education (United Nations 2024).

The massive fall in enrolment of male students ($\Delta\mu \approx 9.54$) can be related to broader socioeconomic forces. UNESCO estimates that in most of the low- and middle-income countries, boys have a high risk of dropping out because of poverty, because of their employment requirements, or because of social behaviour that undervalues formal schooling (UNESCO, 2025). This is consistent with the observation made in this paper whereby male transition between Grade 11 and Grade 12 is less impressive, implying that male students might be being pulled out of the system by non-educational factors.

The standard deviations in female enrolment ($\sigma \approx 25–26$) are large, which means that there is a high amount of heterogeneity across schools. This would be the cause of such variability, probably differences in the size of the school, communities served, infrastructure and capacity of resources.

5.2 Normality Assessment and Distributional Behavior

The observation that the majority of the enrolment distributions deviate greatly from normality has methodological as well as practical consequences. Statistical models like parametric forecasting, models assuming Gaussian errors, rely on normality, yet, in education-administrative data, non-normal distributions are common, particularly when school sizes and contexts vary widely (Hussain et al., 2023). Models such as linear regression remain more appropriate given their robustness to non-Gaussian patterns, if normality is not assumed.

Also, the near-normal shape observed in the aggregated Grade 12 Total data in Fig. 4.2(f) may be explained by the Central Limit Theorem, where aggregation of heterogeneous school-level distributions tends to approximate normality. However, this superficial normality does not justify normality-based assumptions at the individual school level.

5.3 Cohort Stability Through Correlation Analysis

The correlation results (male: $r \approx 0.758$; female: $r \approx 0.792$; total: $r \approx 0.867$) show strong cohort stability between Grade 11 and Grade 12 enrolment. The high correlation observed in the total enrolment ($r \approx 0.867$) suggests that schools with large Grade 11 enrolment generally maintain large classes in Grade 12, despite dropouts. This stability is valuable for planners; Grade 11 enrolment can serve as a reliable predictor for Grade 12, facilitating resource allocation (e.g., teachers, classrooms) in the next academic year. It is also in line with all the literature on educational planning pointing to prior-year enrolment as an important forecasting variable (Hussain et al., 2023).

Additional clues can be made out of the gender-specific correlation difference, which is a little stronger in the female case; this difference may be because of the consistent retention mechanisms, which may be caused by

supportive policies. The less significant male correlation could be attributed to more non-school-related erratic dropout behaviour.

5.4 Predictive Capacity of Regression Models

The regression models yield strong predictive performance: R^2 values of ~0.752 for the total cohort, ~0.627 for females, and ~0.575 for males. These results demonstrate that Grade 11 enrolment accounts for a substantial portion of the variation in Grade 12 enrolment.

Given limited data infrastructure and resource constraints, education planners in The Gambia can rely on simple linear regression to project future Grade 12 numbers. Such models are computationally light, interpretable, and robust even when data do not meet normality assumptions.

Using straightforward, interpretable models is often more effective than over-engineered methods when working with real-world, noisy data (Hussain et al., 2023).

5.5 Interpretation of Interpolation and Numerical Stability

The outcome of the interpolation illustrates the major methodological risk: high-degree polynomials (i.e., full-order interpolation) resulted in extreme oscillations and implausible predictions (i.e., extrapolated values way beyond the possible enrolment). It is a classical instance of Runge's phenomenon, where interpolation by high-degree polynomials with equally spaced points is unstable, particularly around the extreme ends of the interval being interpolated (Runge, 1901).

The phenomenon described by Runge is due to the high flexibility of high-degree polynomials (free parameters), and the tendency of boundary oscillation nodes that are equidistant increases the oscillations of the boundary. Also, numerical theory demonstrates that the error bound is increasing rapidly with the degree because Lebesgue constants are increasing exponentially in degree.

This implies that full-degree interpolation is not accurate when it comes to the prediction of data in situations such as school enrolment, where the model fits the data points but has a rogue behaviour in other areas. Rather, in this study, less smooth, more realistic curves were obtained with lower-degree polynomial interpolation (e.g., 2nd or 3rd degree), but they were still inappropriate for long-term extrapolation compared with regression.

Applying some other methods, such as Chebyshev nodes (interpolation points that are not equally spaced) or piecewise polynomials (splines) to address the Runge phenomenon, are well-studied in numerical analysis. Chebyshev placement of nodes minimises oscillation since most nodes are clumped towards the interval boundaries and the maximum error is minimised. These strategies may be considered in future research or more sophisticated forecasting models.

5.6 Gender Differences and Policy-Relevant Interpretations

These gendered relations, which are the focus of this paper, have relevant policy implications. Even though the female enrolment is high and steady, the high rate at which boys drop out of Grade 11 to Grade 12 is a serious cause of concern since boys in Region 2 might be drifting out of secondary education because of social or economic factors.

This is in line with what is being experienced worldwide; according to UNESCO, boys are becoming more vulnerable to disengagement, especially in the less developed countries and the middle-income ones. In the case of The Gambia, the United Nations has warned of an imminent boy crisis and proposed some interventions, including male mentorship, second-chance programmes, and retention-specific initiatives.

The very high variability in female enrolment (high s) indicates an uneven improvement: there might be schools, which have an effective female-support policy, and there might be schools, which are left behind. The solution to this must be localised and context-specific policies, such as infrastructure, safe transportation, community participation, and incentives, and not national-wide ones.

5.6 Connection to Existing Literature

The research extends a number of lines of the current literature. According to Gajigo (2012), the scholarship and fee-removal policies increased the female secondary enrolment tremendously. This is reiterated in our findings but has also highlighted emerging issues of male retention. The trend of male dropout is correlated with the UNESCO results of gendered risks in education. The occurrence of the Runge phenomenon in our interpolation strongly supports the theory of numerical analysis. The problem with high-degree polynomials is that they are highly unstable with equispaced points (Runge, 1901).

This research is valuable to the planning of education as well as to the quality of best practice in forecasting by integrating educational data with the understanding of the numerical methods.

5.7 Limitations

Despite the value of this study, it has its weaknesses. The results of this study are only analysed in 11 secondary schools in the Region 2, and hence this restricts the extrapolation of the findings. The absence of longitudinal data would not enable us to track temporal trends and explain year to year variations. Simple linear regression was used; more complex and non-linear models (e.g., mixed-effects, time-series, or machine learning) might be able to capture more subtle dynamics. Besides this, we did not directly explore the underlying causes of dropout (e.g., socio-economic and family pressures) that would put quantitative patterns into perspective.

7 Recommendations for Future Research

To build on these results, future studies should utilise more schools (in different parts of The Gambia) and collect data over many years to do longitudinal studies to understand complex relationships and enhance predictive power; explore non-linear and hierarchical forecasting models (ARIMA, mixed-effects models, or machine learning methods) to understand complex dynamics and improve predictive power; investigate Chebyshev-node interpolation, spline interpolation, or constrained least-squares methods to avoid Runge oscillations and improve stability (e.g., mock-Chebyshev least-squares, De Marchi et al., 2015); and pilot and rigorously evaluate retention programmes tailored to male students (mentorship, second-chance, economic support), as well as initiatives to reduce variability in female enrolment across schools.

CONCLUSION

This paper studied the enrolment trends in the eleven senior secondary schools in Region 2 of The Gambia using descriptive statistics, measurement of Gaussian distribution, correlation analysis, regression analysis and interpretation of the results using a polynomial. The key objective was to test cohort stability, gender differences, and the predictability of computing methods in forecasting educational trends. These results provide a methodologically sound picture of the patterns of enrolment that define the Grade 11-Grade 12 transition in the area.

The study findings demonstrate a distinct gender disparity, with male students outnumbered by female students in both grade levels, and that the cohort progression of females is much more stable. The trend confirms the national and international findings that policy interventions in the form of scholarship, sensitisation programmes and reduction of fees have resulted in better female retention (Gajigo, 2012). Meanwhile, the drastic drop in male enrolment between Grade 11 and Grade 12 can be traced to the bigger issues of male non-enrolment in secondary school as indicated by UNESCO and the United Nations. The results show that The Gambia has a lot of work until it needs to maintain the gains it has achieved in the girl's education sector and at the same time deal with a new boy crisis that is being created by the socioeconomic pressures that have led to dropout among boys.

The lack of normality in nearly all the enrolment distributions supports the complexity of educational data in real life situations. The non-uniform distributions are associated with the variations in school size, catchment area, and demography and are a reminder that simple statistical assumptions are rarely true in the field of educational administration. Although the distribution was not normal, correlation analysis showed high cohort stability when pooling male and female data. This implies that prior-year enrolment will be a good predictor of

future cohorts, a fact which is vital in planning how teachers will work, how many will be in the classroom, and how much money will be spent.

This stability is supported by the regression model, in which the total enrolment model predicted the changes in Grade 12 enrolment more than 75 percent. The findings confirm that linear regression is a feasible and useful forecast method in situations such as The Gambia, where available data might not be sufficient to use more complicated modelling techniques. In the meantime, the interpolation analysis indicated that high-degree models of polynomials were seriously numerically unstable. These drastic swings and unrealistic forecasts of full-order Newton interpolation demonstrate the risks of using mathematically complex methods without considering their action on actual, discontinuous administrative data. Polynomials of lower degree gave more sensible results, but regression still gave a better prediction. Such results are consistent with the principles of numerical analysis and provide vital advice to policymakers and analysts interested in the computational approaches that are accurate and stable.

The study, in general, gives a fresh perspective on the education setting in Region 2, including the gendered nature of the process, statistical problems of the enrolment figures, and the advantages and drawbacks of the various computational forecasting methods. It also points to urgent policy needs: male students must receive specific retention strategies, schools with fluctuating female enrolment must have school-specific interventions, and there must be more focus on evidence-based planning. In addition, the article indicates a necessity for increased data gathering, such as longitudinal data and qualitative studies, which would enhance understanding of the dynamic nature of dropouts and improve predictive models.

REFERENCES

1. Akoglu, H. (2018). User's guide to correlation coefficients. Turkish Journal of Emergency Medicine, 18(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001
2. Baid, A., et al. (2011). Landscape analysis of evidence on learning outcomes for adolescents in fragile contexts. Inter-Agency Network for Education in Emergencies.
3. Belanger, C. (2017). External fake constraints interpolation: The end of Runge phenomenon with high-degree polynomials relying on equispaced nodes – Application to aerial robotics motion planning. Institute of Mathematics and its Applications. https://cdn.ima.org.uk/wp/wp-content/uploads/2017/03/Belanger-paper.pdf
4. Boyd, J. P., & Xu, F. (2009). Divergence (Runge phenomenon) for least-squares polynomial approximation on an equispaced grid and mock-Chebyshev subset interpolation. Applied Mathematics and Computation, 210(1), 158–168. https://doi.org/10.1016/j.amc.2008.12.002
5. Daly, A., et al. (2024). Bildungsherausforderungen in The Gambia: Eine Analyse von Zugang, Qualität und Abbruch [Educational challenges in The Gambia: An analysis of access, quality, and dropout]. Universität Marburg.
6. De Marchi, S., Dell'Accio, F., & Mazza, M. (2015). On the constrained mock-Chebyshev least-squares. Journal of Computational and Applied Mathematics, 280, 94–109. https://doi.org/10.1016/j.cam.2014.11.032
7. Dube, T., et al. (2017). Education financing and equity in Sub-Saharan Africa.
8. Gajigo, O. (2016). Closing the education gender gap: Estimating the impact of girls' scholarship program in The Gambia. Education Economics, 24(2), 167–188. https://doi.org/10.1080/09645292.2014.985287
9. Gambia Ministry of Basic and Secondary Education. (2025). The Gambia education finance brief 2025. https://www.edugambia.gm/data-area/publications/education-finance-brief/803-the-gambia-education-finance-brief-2025/download.html
10. Giordono, L., & Pugatch, T. (2017). Non-tuition costs, school access and student performance: Evidence from the Gambia. Journal of African Economies, 26(2), 140–168. https://doi.org/10.1093/jae/ejw033
11. Global Partnership for Education. (2020). Summative evaluation of GPE country-level support: The Gambia. https://www.globalpartnership.org/node/document/download?file=document/file/2020-04-summative-evaluation-gpe-country-level-support-education-the-gambia.pdf

12. Govender, P., & Farlam, P. (2004). Erasing the future: The tragedy of Africa's education. South African Institute of International Affairs. https://saiia.org.za/research/erasing-the-future-the-tragedy-of-africas-education-2/
13. Hua, et al. (2020).
14. Hussain, J., Rosangliana, D., & Vanlalruata. (2023). Student gross enrolment ratio forecasting: A comparative study using statistical method and machine learning. International Journal of Information and Education Technology, 13(3), 439–447. https://doi.org/10.18178/ijiet.2023.13.3.1824
15. Inter-Agency Network for Education in Emergencies. (2020). Landscape analysis of evidence on learning outcomes for adolescents in fragile contexts.
16. Karmaliani, R., et al. (2020). A Right to Play: Evaluation of school-based violence prevention in Hyderabad, Pakistan.
17. Kimmel, M., et al. (2004). Masculinities and gender norms: A global perspective.
18. Lifting Limits. (2019). Evaluation of the Lifting Limits programme in London primary schools.
19. Morishitwara, & Prakash. (2017).
20. Peremine, et al. (2017).
21. Rama, Jordan, & Yuan. (2021).
22. Riley, K. (2019).
23. Runge, C. (1901). Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten [On empirical functions and interpolation between equidistant ordinates]. Zeitschrift für Mathematik und Physik, 46, 224–243.
24. Stromquist, N. P. (2007). The gender socialization process in schools: A cross-national comparison. Background paper for the Education for All Global Monitoring Report 2008. UNESCO.
25. UNESCO. (2018). Gender-transformative education: A framework for action.
26. UNESCO. (2019). Global education monitoring report 2019: Gender report – Building bridges for gender equality.
27. UNESCO. (2022). Leave no child behind: Global report on boys' disengagement from education. https://www.unesco.org/en/articles/leave-no-child-behind-global-report-boys-disengagement-education
28. UNICEF. (2017). National study of out-of-school children in The Gambia. https://www.unicef.org/gambia/media/636/file/National-Study-of-Out-of-School-Children-in-The-Gambia-2017.pdf
29. UNICEF. (2019). Gender norms and education: A review of the evidence.
30. UNICEF. (2021). The Conditional Cash Transfer for Education (CCTE) Programme. https://www.unicef.org/turkey/en/conditionsi-cash-transfer-education-ccbe-programmeme
31. United Nations Development Programme. (2025). Gambia human development report 2025. https://www.undp.org/sites/g/files/zskgke326/files/2026-08/gambia_human_development_report_2025.pdf
32. United Nations Economic Commission for Africa. (n.d.). African gender and development index: Regional synthesis report.
33. United Nations in The Gambia. (2024). Advancing gender equality in education. https://gambia.un.org/en/275483-advancing-gender-equality-education
34. United Nations University. (2025). What are the root causes of global issues and why do they matter? UNU-EHS. https://unu.edu/ehs/article/what-are-root-causes-global-issues-and-why-do-they-matter
35. World Bank. (2017). Gambia education public expenditure review.
36. World Bank. (2024). Project performance assessment report: Early childhood development and lower basic education in The Gambia. https://documents1.worldbank.org/curated/en/099061524123016245/pdf/P162890119d06c08c1983e1ff5250d14522.pdf
37. World Food Programme. (2020). State of school feeding worldwide. Yuma, et al. (2019).

APPENDICES

Gaussian Code

```
%% ============================
%  LOAD & PREPARE THE DATA
%% ============================

% Read the Excel file into a table
enrollment_data = readtable("C:\Users\mbah1\Desktop\300LVL(Second Semester)↙
\Computational Mathematics\DATA FOR FATOU FATTY.xlsx");

% Remove the first row (header-like) and the last row (empty/total row)
enrollment_data = enrollment_data(2:end-1, :);

%% ============================
%  EXTRACT NUMERIC COLUMNS
%% ============================

% Convert text columns to numbers because MATLAB reads Excel columns as strings
grade_11_male    = str2double(enrollment_data.GRADE11);
grade_11_female  = str2double(enrollment_data.Var3);
grade_12_male    = str2double(enrollment_data.GRADE12);
grade_12_female  = str2double(enrollment_data.Var5);

%% ============================
%  COMPUTE TOTALS
%% ============================

grade_11_total = grade_11_male + grade_11_female;
grade_12_total = grade_12_male + grade_12_female;

%% ============================
%  RUN DISTRIBUTION ANALYSIS
%% ============================

analyze_distribution(grade_11_male,   "Grade 11 Male");
analyze_distribution(grade_11_female, "Grade 11 Female");
analyze_distribution(grade_12_male,   "Grade 12 Male");
analyze_distribution(grade_12_female, "Grade 12 Female");
analyze_distribution(grade_11_total,  "Grade 11 Total");
analyze_distribution(grade_12_total,  "Grade 12 Total");
```

```
%% ====================================================================
%   FUNCTION: analyze_distribution
%   Purpose: Computes statistics and plots histogram + normal curve + QQ
%% ====================================================================
function analyze_distribution(data, title_prefix)

    %% ---- Compute Basic Statistics ----
    mu = mean(data);      % Average of the dataset
    sigma = std(data);    % Standard deviation

    disp("===== " + title_prefix + " =====");
    disp("Mean (mu): " + mu);
    disp("Std Dev (sigma): " + sigma);
```

```
    disp(" ");

    %% ---- Histogram + Normal Distribution Curve ----
    figure;

    % Generate 100 points between the min and max of the data
    x = linspace(min(data), max(data), 100);

    % Normal distribution curve using computed mean and std
    y = normpdf(x, mu, sigma);

    % Plot histogram normalized to a probability density
    histogram(data, 'Normalization', 'pdf', 'NumBins', 40);
    hold on;

    % Plot Gaussian curve
    plot(x, y, 'LineWidth', 2);

    title('Gaussian Fit - ' + title_prefix);
    xlabel('Value');
    ylabel('Probability Density');
    hold off;

    %% ---- Q-Q Plot (Normality Test) ----
    figure;
    qqplot(data);
    title(title_prefix + " - Q-Q Plot");
    xlabel('Theoretical Quantiles');
    ylabel('Sample Quantiles');

end
```

Correlation Code

```
%% ============================================================
%   ENROLLMENT CORRELATION ANALYSIS (ALL IN ONE FILE)
%   Loads data, computes correlations, and plots relationships.
%% ============================================================

%% 1. LOAD DATA
load('enrollmentdata.mat');
% This file must contain:
% grade_11_male, grade_12_male,
% grade_11_female, grade_12_female,
% grade_11_total, grade_12_total


%% 2. COMPUTE CORRELATIONS

% --- Male ---
correlation_male = corrcoef(grade_11_male, grade_12_male);
correlation_male = correlation_male(1,2);    % extract actual correlation value

% --- Female ---
correlation_female = corrcoef(grade_11_female, grade_12_female);
correlation_female = correlation_female(1,2);

% --- Total ---
correlation_total = corrcoef(grade_11_total, grade_12_total);
correlation_total = correlation_total(1,2);

%% Display correlation results
disp("Correlation (Male): " + correlation_male);
disp("Correlation (Female): " + correlation_female);
disp("Correlation (Total): " + correlation_total);
```

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

```
%% 3. PLOTS + REGRESSION LINES
% ------------------------------------------------------------
%   Scatter Plot + Regression Line (Male)
% ------------------------------------------------------------
figure;
scatter(grade_11_male, grade_12_male, 'filled');
xlabel("Grade 11 Male Enrollment");
ylabel("Grade 12 Male Enrollment");
title("Correlation: Grade 11 vs Grade 12 Enrollment (Male)");
grid on;
hold on;

% Fit linear regression: y = p(1)*x + p(2)
p = polyfit(grade_11_male, grade_12_male, 1);

% Generate line for plotting
xline = linspace(min(grade_11_male), max(grade_11_male));
yline = polyval(p, xline);




plot(xline, yline, 'LineWidth', 2);
hold off;



% ------------------------------------------------------------
%   Scatter Plot + Regression Line (Female)
% ------------------------------------------------------------
figure;
scatter(grade_11_female, grade_12_female, 'filled');
xlabel("Grade 11 Female Enrollment");
ylabel("Grade 12 Female Enrollment");
title("Correlation: Grade 11 vs Grade 12 Enrollment (Female)");
grid on;
hold on;

p_f = polyfit(grade_11_female, grade_12_female, 1);
xline_f = linspace(min(grade_11_female), max(grade_11_female));
yline_f = polyval(p_f, xline_f);

plot(xline_f, yline_f, 'LineWidth', 2);
hold off;



% ------------------------------------------------------------
%   Scatter Plot + Regression Line (Total)
% ------------------------------------------------------------
figure;
scatter(grade_11_total, grade_12_total, 'filled');
xlabel("Grade 11 Total Enrollment");
ylabel("Grade 12 Total Enrollment");
title("Correlation: Grade 11 vs Grade 12 Enrollment (Total)");
grid on;
hold on;

p_t = polyfit(grade_11_total, grade_12_total, 1);
xline_t = linspace(min(grade_11_total), max(grade_11_total));
yline_t = polyval(p_t, xline_t);

plot(xline_t, yline_t, 'LineWidth', 2);
hold off;
```

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

Regression Code

```
load('enrollmentdata.mat');

p_male=polyfit(grade_11_male,grade_12_male,1);
male_slope=p_male(1);
male_intercept=p_male(2);
predicted_male=(male_slope*grade_11_male)+male_intercept;
SS_res_male=sum((grade_12_male-predicted_male).^2);
SS_total_male=sum((grade_12_male-mean(grade_12_male)).^2);
R2_male=1-(SS_res_male/SS_total_male);

p_female=polyfit(grade_11_female,grade_12_female,1);
slope_female=p_female(1);
intercept_female=p_female(2);
predicted_female=(slope_female*grade_11_female)+intercept_female;
SS_res_female=sum((grade_12_female-predicted_female).^2);
SS_total_female=sum((grade_12_female-mean(grade_12_female)).^2);
R2_female=1-(SS_res_female/SS_total_female);

p_total=polyfit(grade_11_total,grade_12_total,1);
total_slope=p_total(1);
total_intercept=p_total(2);
total_predicted=(total_slope*grade_11_total)+total_intercept;
SS_res_total=sum((grade_12_total-total_predicted).^2);
SS_total_total=sum((grade_12_total-mean(grade_12_total)).^2);
R2_total=1-(SS_res_total/SS_total_total);

disp(R2_male)
disp(R2_female)
disp(R2_total)
figure;

subplot(3,1,1);
scatter(grade_11_male,grade_12_male,60,'b','filled'); hold on;
plot(grade_11_male,predicted_male,'LineWidth',2,'Color',[0 0.3 1]);
title(['Male Enrollment Regression (R^2 = ' num2str(R2_male,'%.3f') ')']);
xlabel('Grade 11 Male');
ylabel('Grade 12 Male');
grid on;


subplot(3,1,2);
scatter(grade_11_female,grade_12_female,60,'m','filled'); hold on;
plot(grade_11_female,predicted_female,'LineWidth',2,'Color',[1 0 0.7]);
title(['Female Enrollment Regression (R^2 = ' num2str(R2_female,'%.3f') ')']);
xlabel('Grade 11 Female');
ylabel('Grade 12 Female');
grid on;

subplot(3,1,3);
scatter(grade_11_total,grade_12_total,60,'g','filled'); hold on;
plot(grade_11_total,total_predicted,'LineWidth',2,'Color',[0 0.7 0]);
title(['Total Enrollment Regression (R^2 = ' num2str(R2_total,'%.3f') ')']);
xlabel('Grade 11 Total');
ylabel('Grade 12 Total');
```

```
11/25/25 8:17 AM  C:\Users\mbah1\Matlab...\regression.m  2 of 2

grid on;

suptitle('Regression Analysis for Male, Female, and Total Enrollment');
```

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IX September 2026

Interpolation Code

```
load('enrollmentdata.mat');

[grade_11_male_unique, idx] = unique(grade_11_male, 'stable');
grade_12_male_unique = grade_12_male(idx);

male_est_full = newton_interp(grade_11_male_unique, grade_12_male_unique, 64);
male_est_full_actual = newton_interp(grade_11_male_unique, grade_12_male_unique, 45);

male_est_3 = newton_interp(grade_11_male_unique(1:4), grade_12_male_unique(1:4), 64);
male_est_3_actual = newton_interp(grade_11_male_unique(1:4), grade_12_male_unique(1:4), 45);

male_est_2 = newton_interp(grade_11_male_unique(1:3), grade_12_male_unique(1:3), 64);
male_est_2_actual = newton_interp(grade_11_male_unique(1:3), grade_12_male_unique(1:3), 45);

disp(male_est_full);
disp(male_est_full_actual);
disp(male_est_3);
disp(male_est_3_actual);
disp(male_est_2);
disp(male_est_2_actual);

figure;
xq = linspace(min(grade_11_male_unique), max(grade_11_male_unique), 200);
yq_full = arrayfun(@(t) newton_interp(grade_11_male_unique, grade_12_male_unique, t), xq);
yq_3 = arrayfun(@(t) newton_interp(grade_11_male_unique(1:4), grade_12_male_unique(1:4), t), xq);
yq_2 = arrayfun(@(t) newton_interp(grade_11_male_unique(1:3), grade_12_male_unique(1:3), t), xq);


plot(grade_11_male_unique, grade_12_male_unique, 'o');
hold on;
plot(xq, yq_full, '-', 'LineWidth', 1.5);
plot(xq, yq_3, '--', 'LineWidth', 1.5);
plot(xq, yq_2, ':', 'LineWidth', 1.5);
plot(64, male_est_full, 'x', 54, male_est_full_actual, 'x');
plot(64, male_est_3, '+', 54, male_est_3_actual, '+');
plot(64, male_est_2, 's', 54, male_est_2_actual, 's');

title('Male Interpolation');
xlabel('Grade 11');
ylabel('Grade 12');
legend('Data','Full','3rd','2nd','Full est','3rd est','2nd est');

[grade_11_female_unique, idx] = unique(grade_11_female, 'stable');
grade_12_female_unique = grade_12_female(idx);

female_est_full = newton_interp(grade_11_female_unique, grade_12_female_unique,
```

```
50);
female_est_full_actual = newton_interp(grade_11_female_unique,↙
grade_12_female_unique, 52);

female_est_3 = newton_interp(grade_11_female_unique(1:4), grade_12_female_unique(1:↙
4), 50);
female_est_3_actual = newton_interp(grade_11_female_unique(1:4),↙
grade_12_female_unique(1:4), 52);

female_est_2 = newton_interp(grade_11_female_unique(1:3), grade_12_female_unique(1:↙
3), 50);
female_est_2_actual = newton_interp(grade_11_female_unique(1:3),↙
grade_12_female_unique(1:3), 52);

disp(female_est_full);
disp(female_est_full_actual);
disp(female_est_3);
disp(female_est_3_actual);
disp(female_est_2);
disp(female_est_2_actual);

figure;
xq = linspace(min(grade_11_female_unique), max(grade_11_female_unique), 200);
yq_full = arrayfun(@(t) newton_interp(grade_11_female_unique,↙
grade_12_female_unique, t), xq);
yq_3 = arrayfun(@(t) newton_interp(grade_11_female_unique(1:4),↙
grade_12_female_unique(1:4), t), xq);
yq_2 = arrayfun(@(t) newton_interp(grade_11_female_unique(1:3),↙
grade_12_female_unique(1:3), t), xq);



plot(grade_11_female_unique, grade_12_female_unique, 'o');
hold on;
plot(xq, yq_full, '-', 'LineWidth', 1.5);
plot(xq, yq_3, '--', 'LineWidth', 1.5);
plot(xq, yq_2, ':', 'LineWidth', 1.5);
plot(50, female_est_full, 'x', 42, female_est_full_actual, 'x');
plot(50, female_est_3, '+', 42, female_est_3_actual, '+');
plot(50, female_est_2, 's', 42, female_est_2_actual, 's');

title('Female Interpolation');
xlabel('Grade 11');
ylabel('Grade 12');
legend('Data','Full','3rd','2nd','Full est','3rd est','2nd est');

[grade_11_total_unique, idx] = unique(grade_11_total, 'stable');
grade_12_total_unique = grade_12_total(idx);

total_est_full = newton_interp(grade_11_total_unique, grade_12_total_unique, 49);
total_est_full_actual = newton_interp(grade_11_total_unique, grade_12_total_unique,↙
100);

total_est_3 = newton_interp(grade_11_total_unique(1:4), grade_12_total_unique(1:4),↙
49);
```

```
total_est_2 = newton_interp(grade_11_total_unique(1:3), grade_12_total_unique(1:3),↙
49);
total_est_2_actual = newton_interp(grade_11_total_unique(1:3),↙
grade_12_total_unique(1:3), 100);

disp(total_est_full);
disp(total_est_full_actual);
disp(total_est_3);
disp(total_est_3_actual);
disp(total_est_2);
disp(total_est_2_actual);

figure;
xq = linspace(min(grade_11_total_unique), max(grade_11_total_unique), 200);
yq_full = arrayfun(@(t) newton_interp(grade_11_total_unique, grade_12_total_unique,↙
t), xq);
yq_3 = arrayfun(@(t) newton_interp(grade_11_total_unique(1:4),↙
grade_12_total_unique(1:4), t), xq);
yq_2 = arrayfun(@(t) newton_interp(grade_11_total_unique(1:3),↙
grade_12_total_unique(1:3), t), xq);

plot(grade_11_total_unique, grade_12_total_unique, 'o');
hold on;
plot(xq, yq_full, '-', 'LineWidth', 1.5);
plot(xq, yq_3, '--', 'LineWidth', 1.5);
plot(xq, yq_2, ':', 'LineWidth', 1.5);
plot(49, total_est_full, 'x', 100, total_est_full_actual, 'x');
plot(49, total_est_3, '+', 100, total_est_3_actual, '+');
plot(49, total_est_2, 's', 100, total_est_2_actual, 's');

title('Total Interpolation');
xlabel('Grade 11');
ylabel('Grade 12');
legend('Data','Full','3rd','2nd','Full est','3rd est','2nd est');

    function y_interp = newton_interp(x, y, x_query)
    x = x(:);
    y = y(:);
    n = length(x);
    F = zeros(n,n);
    F(:,1) = y;
    for j = 2:n
        for i = j:n
            F(i,j) = (F(i,j-1) - F(i-1,j-1)) / (x(i) - x(i-j+1));
        end
    end
    coeff = diag(F);
    y_interp = coeff(1);
    prod_term = 1;
    for j = 2:n
```

11/25/25 8:19 AM C:\Users\mbah1\Ma...\interpolation.m 4 of 4

```
        prod_term = prod_term * (x_query - x(j-1));
        y_interp = y_interp + coeff(j) * prod_term;
    end
    end
```