International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Senior High School Student Enrollment Forecasting Model: An Application of Time Series Analysis

Reñel M. Sabanal

Saint Mary's University
DOI: https://doi.org/10.55248/gengpi.4.923.52729

ABSTRACT

Prediction plays a vital role used for strategic and tactical decision-making undertaking that pave the way for efficient and effective management. The current study is to predict the number of senior high school enrollees basing the 8 semesters of SY 2018-2022 of Saint John Berchmans High School Incorporated by using the best fit model generated by Microsoft Excel. The data were gathered from the Learning Information System (LIS) of the said school. Only the ratio by strand has a significant linear relationship to the number of SHS enrollees but when the three independent variables (ratio by year level, strand, gender) are combined, these ratios are found to be predictors of the number of SHS enrollees. Sextic polynomial trend model was the best fit model for the number of SHS enrollees.

INTRODUCTION

Philippines has undertaken major educational reforms that transition and shift its 10-year basic education into the K-12 curriculum. Senior high school implementation aimed to equip the students with essential knowledge and skills that will help them prepare better for the chosen path in the higher education, employment, or entrepreneurship. The addition of two more years will better equip the students with the necessary skills, knowledge, and values needed for a successful future in their fields or course. Upon the implementation of this curriculum there were 5,965 public schools and 4,830 private high schools were approved to offer the Senior High School. Amidst the reservations for the Senior High School Program Implementation, the Department of Education stands firm for a positive response for the opening of classes. It was indeed a good start for the SHS as it reached over 1 million enrollees both public and private nationwide.

Saint John Berchmans High School Incorporated (SJBHSI) is a Private Catholic School in the municipality of Cordon, Isabela that implements Senior High School (SHS) level. Accountancy Business and Management (ABM) and Humanities and Social Sciences (HUMSS) were offered by SJBHSI to the students from 2018 up until now. The enrollment of SJBHSI in the SHS is growing. Enrollment of SHS at the SJBHSI were categorized into three: the ratio by year level (Grade 11 over Grade 12), ratio by strand (ABM over HUMSS) and ratio by gender (Male over Female).

This study was conducted at Saint John Berchmans High School Incorporated (SJBHSI) in Cordon, Isabela. The historical data of enrollment of SJBHSI from 2018 to 2022 per semester were used in this study. These data were gathered from the Learning Information System (LIS) of the said school with permission from the school principal. The aim of this study is to examine the linear relationship of the three ratios to the enrollment. It also aims the best fit model of the number of SHS enrollees at SJBHSI and predict the number of SHS enrollees of first semester of school year (SY) 2022-2023.

STATEMENT OF THE PROBLEM

This study aimed to explore the mathematical modeling of SHS enrollment at Saint John Berchmans High School Incorporated.

Especially, this study aimed to:

1. Determine the trend of the SHS enrollment of SJBHSI from SY 2018-2022.
2. Find if the enrollment has a significant linear relationship with the following variables;

a. Ratio by year level (Grade 11 over Grade 12)

b. Ratio by strand (ABM over HUMSS)

c. Ratio by gender (Male over Female)

d. Ratio by year level, ratio by strand, and ratio by gender.

3. Construct a time series model of the enrollment using the following models to predict the enrollment for the first semester of the school year 2022-2023.

a. Linear

b. Quadratic

c. Exponential

d. Polynomial (cubic, quartic, quintic, sextic)

e. Power

f. Moving Average

g. Exponential Smoothing

h. Autoregression

4. Determine the best fit model and predict the number of SHS enrollees in the first semester of the SY 2022-2023.

METHODS

Participants

SHS Enrollees of SJBHSI was used primarily as the participant in the study. The data was gathered from the first semester of 2018 to the second semester of 2022. The data were categorized by the ratios by year level, strand, and gender.

Data Sources

The data for the number of SHS enrollees were collected and tallied from the Learner Information System (LIS) of SJBHSI.

Procedure

The author explored the LIS of the chosen school to be used as variables in the regression analysis as it was a partial requirement in the Mathematical Modelling course. After recording the data to be used, it was organized first and run in MS Excel for analysis and interpretation.

Data Analysis

Observation of data per semester from 2018 to 2022 was done. A Scatter diagram was used to determine the trend of the number of the SHS enrollment at SJBHSI from SY 2018 to 2022. Simple linear regression was utilized to illustrate if there is a significant linear relationship between dependent and independent variables. Multiple linear regression was utilized to predict the number of SHS enrollees through the ratios by year level, strand, and gender. Constructing a time series model of the number of SHS enrollees. Determine the best fit model for predicting the number of SHS enrollees by observing the difference between r-square and the standard error.

RESULTS AND DISCUSSIONS

Section 1: Trend of the Number of SHS Enrollees at SJBHSI.

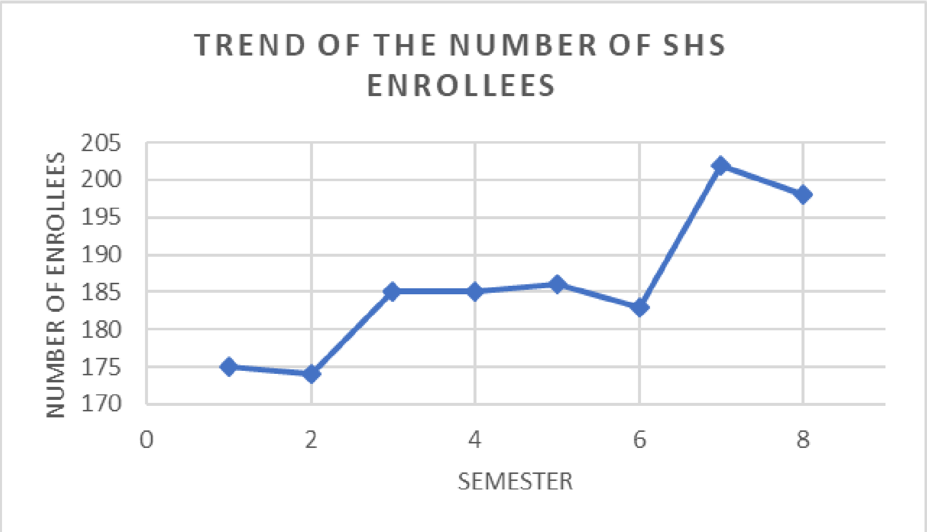

Figure 1: The Trend of the Number of SHS Enrollees from SY 2018 to 2022 per Semester

Figure 1 shows that there was an increase in the enrollment over 8 semesters. During the 7th Semester (the first semester of SY 2021-2022), there is a sudden increase of the number of the SHS enrollment. During the even number of semesters or the second semester of each school year, there is only a decrease of the number of SHS enrollees.

Section 2: Significant Linear Relationship of Number of SHS Enrollees

2.1 Ratio of Grade 11 Enrollees to Grade 12 Enrollees (Year Level)

Table 1 shows the simple linear regression of the number of SHS enrollees and the ratio by year level. The ratio by year level might still have a relationship, yet moderate, to the number of SHS Enrollees since the p-value is close at 0.05. The number of SHS Enrollees can be predicted using the equation $y = 62.9280x + 114.8896$, where x is the ratio by year level and y is the number of SHS Enrollees. There is an increase of the ratio (year level) by 62.928 per semester. The model can explain 43.74% of the variances.

Table 1: Linear Regression Result of the Number of SHS Enrollees and Ratio by Year Level

	Coefficients	Standard Error	T Stat	P-value
Intercept	114.8896	33.0434	3.4769	0.01
Ratio by Year Level	62.9280	29.1349	2.1599	0.07
$R^2 = 0.4374$	**significant at 0.01	*significant at 0.05		

2.2 Ratio of ABM Enrollees to HUMSS Enrollees (Strands)

Table 2 shows the linear regression result for the number of SHS Enrollees and the ratio of ABM enrollees to HUMSS enrollees. Table 2 presents that the ratio by strands is a significant predictor of the number of SHS Enrollees. Table 2 also shows that the ratio by strands significantly influences the number of SHS Enrollees, $p < 0.05$ as the result indicates that there exists a significant linear relationship between the ratio by strand and the number of SHS Enrollees. The number of SHS Enrollees can be predicted using the equation $y = -45.8519x + 223.6332$, where x is the ratio by strands and y is the number of SHS Enrollees. The model can explain 61.12% of the variances.

Table 2: Linear Regression Result of the Number of SHS Enrollees and Ratio by Strand

	Coefficients	Standard Error	T Stat	P-value
Intercept	223.6332	12.4738	17.9283	1.93633E-06
Ratio by Strand	-45.8519	14.9281	-3.0715	0.02*
$R^2 = 0.6112$	**significant at 0.01	*significant at 0.05		

2.3 Ratio of the Male Enrollee to the Female Enrollee (Gender)

Table 3 shows the linear regression result for the number of SHS Enrollees and the ratio of the male enrollees to female enrollees. There is no linear relationship between the number of SHS Enrollees and the ratio by gender. One reason can be the ratio by gender is decreasing over time but the number of enrollees is increasing over time. The number of SHS Enrollees can be predicted using the equation $y = 6.2497x + 181.4016$, where x is the ratio by gender and y is the SHS enrollment. The model can explain 00.12% of the variances.

Table 3: Linear Regression Result of the Number of SHS Enrollees and Ratio by Gender

	Coefficients	Standard Error	T Stat	P-value
Intercept	181.4016	52.5895	3.4494	0.01
Ratio by Gender	6.2497	71.2931	0.0877	0.93
$R^2 = 0.0012$	**significant at 0.01	*significant at 0.05		

2.4 Ratios by Year Level, Strand and Gender

Table 4 demonstrates the linear regression result of the three independent variables. When combined, the linear regression shows the ratios of year level, strands and genders have a significant linear relationship to the number of SHS Enrollees. The predictors indicate a significant effect on the number of SHS Enrollees, p<0.05. However, among the three variables, the ratio by genders influences the number of SHS Enrollees. The model that can be used in the prediction of the number of SHS Enrollees is $y = 160.9522x_1 + 12.9792x_2 - 190.4592x_3 + 132.7809$ where x_1 is the ratio by year level, x_2 is the ratio by strand, x_3 is the ratio by gender and y is the SHS enrollment. The model can explain 89.34% of the variances.

Table 4: Linear Regression Result of SHS Enrollment and the Three Variables

	Coefficients	Standard Error	T Stat	P-value
Intercept	132.7809	98.2356	1.3517	0.25
Ratio by Year Level x_1	160.9522	88.2132	1.8246	0.14
Ratio by Strand x_2	13.9792	43.0036	0.3251	0.76
Ratio by Gender x_3	-190.4592	64.3094	-2.9616	0.04*
P-value=0.02* $R^2 = 0.8934$		**significant at 0.01	*significant at 0.05	

Section 3: Time Series Model

a. Linear

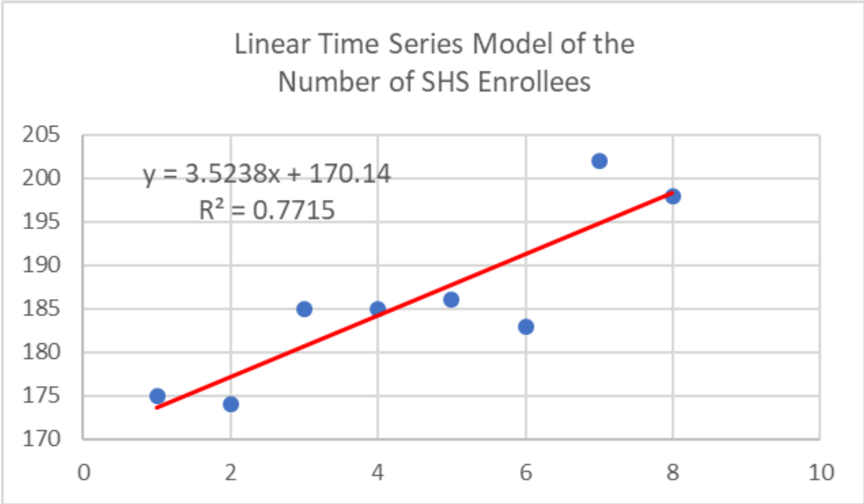

Figure 2: Linear Time Series Model of Number of SHS Enrollees

Figure 2 shows the linear times series model of the number of SHS enrollees from the 1st semester of 2018 to the 2nd semester of 2022. Figure 2 indicated that the model for linear rime series predicts the number of SHS enrollees is $y = 3.5238x + 170.14$. The coefficient of determination is 0.7715 indicating a unit change in time leads to a 77.15% change in the number of SHS enrollees.

b. Quadratic

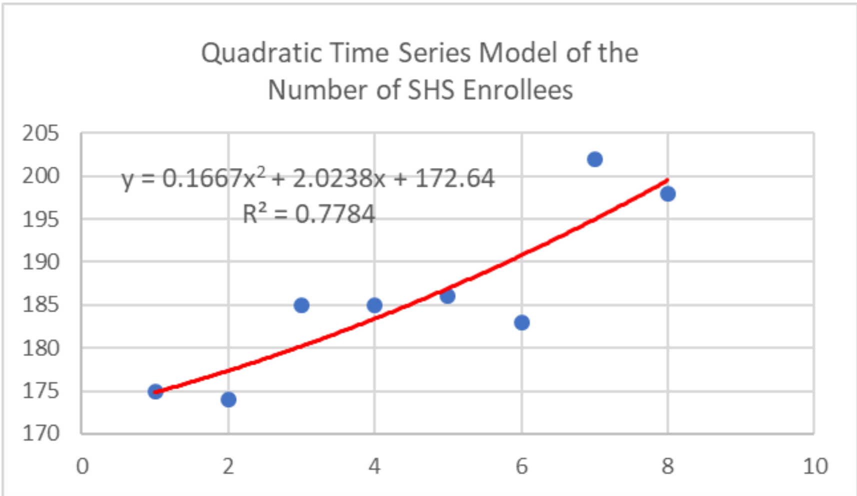

Figure 3: Quadratic Time Series Model of Number of SHS Enrollees

Figure 3 shows the quadratic times series model of the number of SHS enrollees from the 1st semester of 2018 to the 2nd semester of 2022. Figure 3 shows that the quadratic model for the time series of the number of SHS enrollees is $y = 0.1667x^2 + 2.0238x + 172.64$. The model has a coefficient of determination equal to 0.7784 which means that 77.84% of the variation can be explained by the quadratic model.

c. Exponential

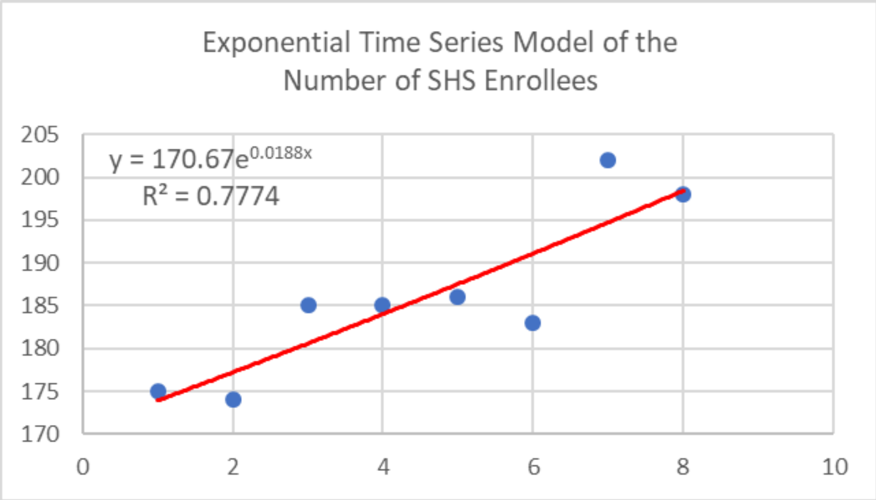

Figure 4: Exponential Time Series Model of Number of SHS Enrollees

Figure 4 shows the exponential times series model of the number of SHS enrollees from the 1st semester of 2018 to the 2nd semester of 2022. Figure 4 has $y = 170.67e^{0.0188x}$ as the model for exponential time series with the coefficient of determination equals 0.7774. This means that 77.74% is being explained by the model.

d. Polynomial (Cubic, Quartic, Quintic and Sextic)

d.1. Cubic

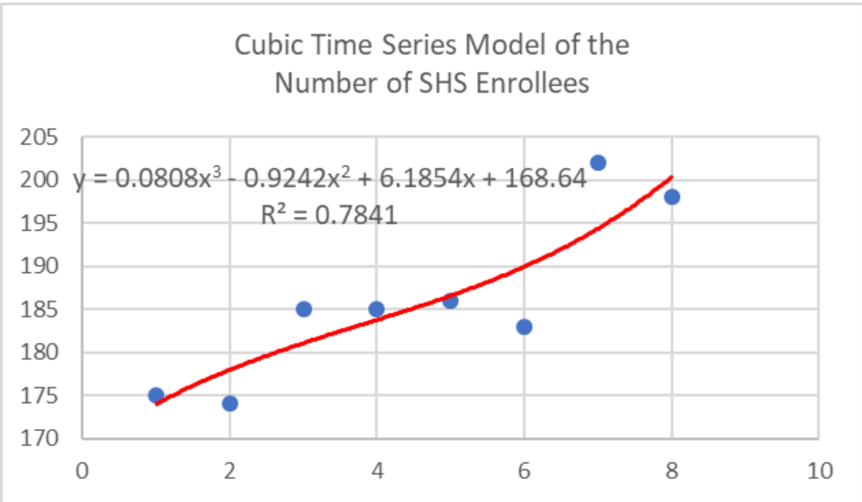

Figure 5: Cubic Time Series Model of Number of SHS Enrollees

Figure 5 shows the number of SHS enrollees from the 1st semester of 2018 to the 2nd semester of 2022 in a cubic polynomial-time series model. The model in Figure 5, $y = 0.0808x^3 - 0.9242x^2 + 6.1854x + 168.64$, describes the cubic polynomial-time series of the number of SHS enrollees. Given the coefficient of determination equals 0.7841, the model can explain 78.41% of the variation.

d.2. Quartic

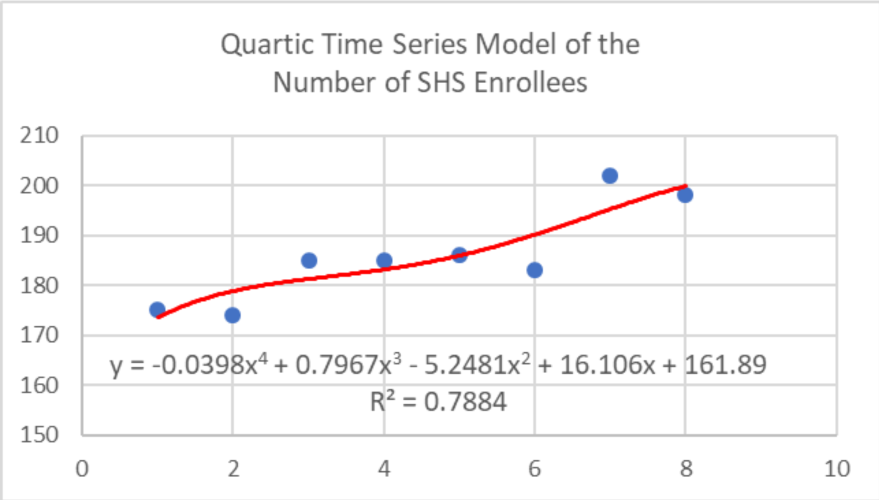

Figure 6: Quartic Time Series Model of Number of SHS Enrollees

Shown in figure 6 is the quartic polynomial-time series model of SHS enrollees from the 1st semester of 2018 to the 2nd semester of 2022. Figure 6 tells that the equation for the quartic polynomial-time series model of the number of SHS enrollees is $y = -0.0398x^4 + 0.7967x^3 - 5.2481x^2 + 16.106x + 161.89$ with coefficient of determination equivalent to 0.7884 indicating that 78.84% of the variation is being explained by the model.

d.3. Quintic

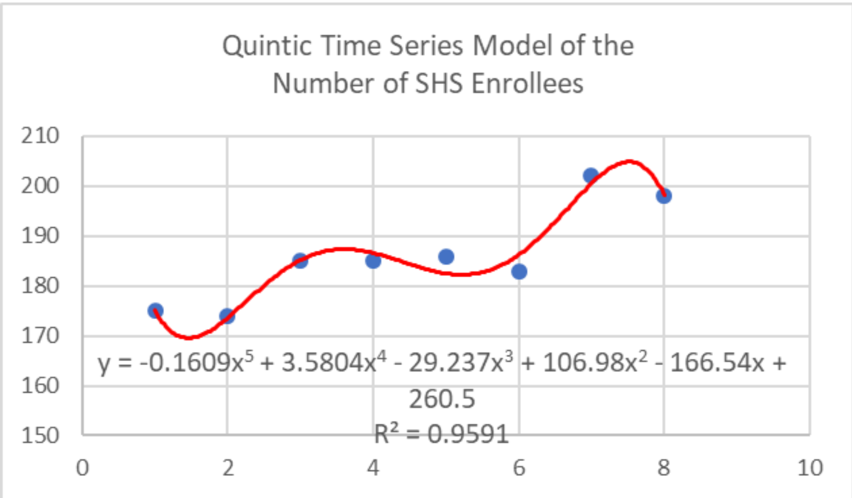

Figure 7: Quintic Time Series Model of Number of SHS Enrollees

Shown in Figure 7 is the quintic polynomial-time series model of SHS enrollees from the 1st semester of 2018 to the 2nd semester of 2022. The Figure 7 shows that the equation for quintic polynomial-time series model of the number of SHS enrollees is $y = -0.1609x^5 + 3.5804x^4 - 29.237x^3 + 106.98x^2 - 166.54x + 260.5$. The coefficient of determination is equivalent to 0.9591 indicating that 95.91% of the variation is being explained by the model.

d.4. Sextic

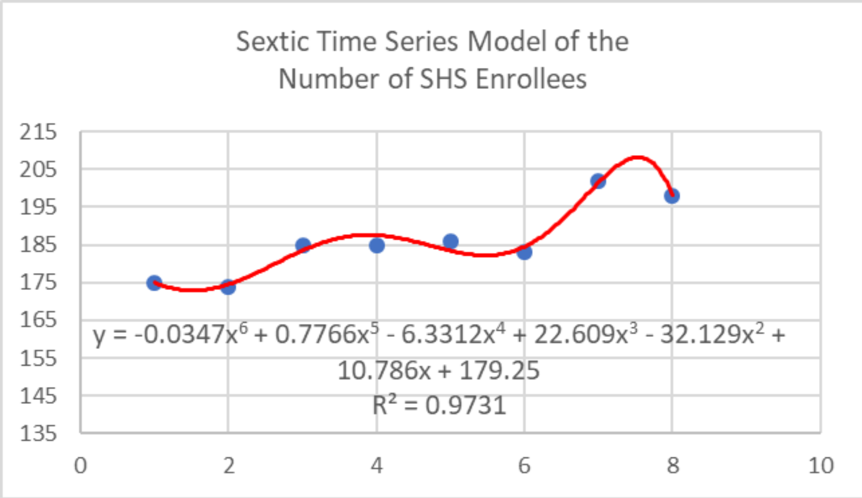

Figure 8: Sextic Time Series Model of Number of SHS Enrollees

Figure 8 shows the number of SHS enrollees from the 1st semester of 2018 to the 2nd semester of 2022 in a sextic polynomial-time series model. The Figure 8 shows that the equation for sextic polynomial-time series model of the number of SHS enrollees is $y = -0.0347x^6 + 0.7766x^5 - 6.3312x^4 + 22.609x^3 - 32.129x^2 + 10.786x + 179.25$. The coefficient of determination is equivalent to 0.9731 indicating that 97.31% of the variation is being explained by the model.

e. Logarithmic

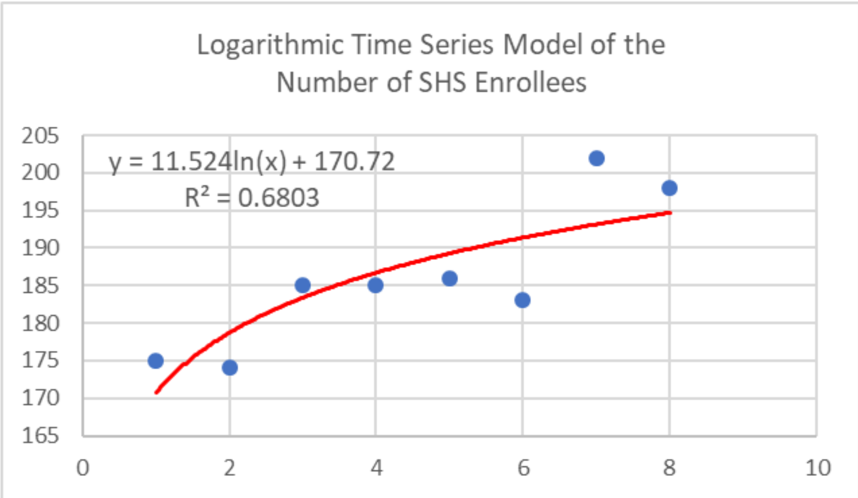

Figure 9: Logarithmic Time Series Model of Number of SHS Enrollees

Figure 9 shows the logarithmic times series model of the number of SHS enrollees from the 1st semester of 2018 to the 2nd semester of 2022. Figure 9 has $y = 11.524ln(x) + 170.72$ as the model for logarithmic time series with the coefficient of determination equals 0.6803. This means that 68.03% is being explained by the model.

f. Power

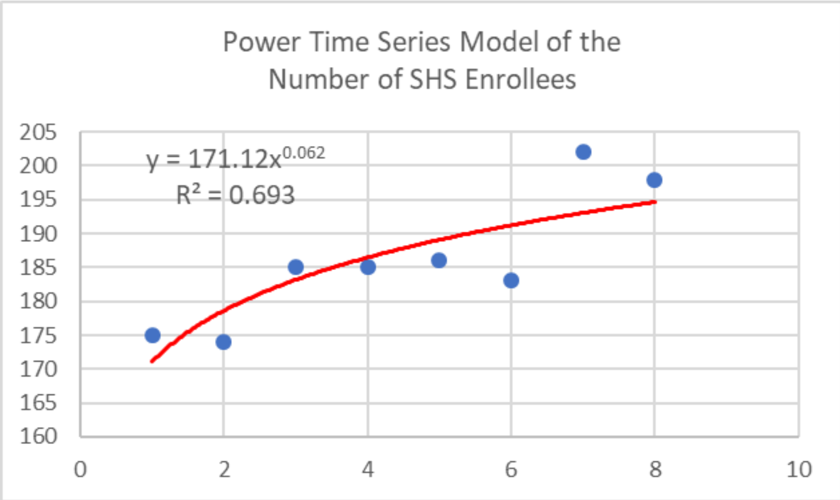

Figure 10: Power Time Series Model of Number of SHS Enrollees

Figure 10 shows the power times series model of the number of SHS enrollees from the 1st semester of 2018 to the 2nd semester of 2022. As indicated in Figure 10, the model for power time series to predict number of SHS enrollees is $y = 171.12x^{0.062}$. The coefficient if determination is 0.693 indicating that a unit change in time leads to 69.30% change in the number of SHS enrollees.

g. Moving Average

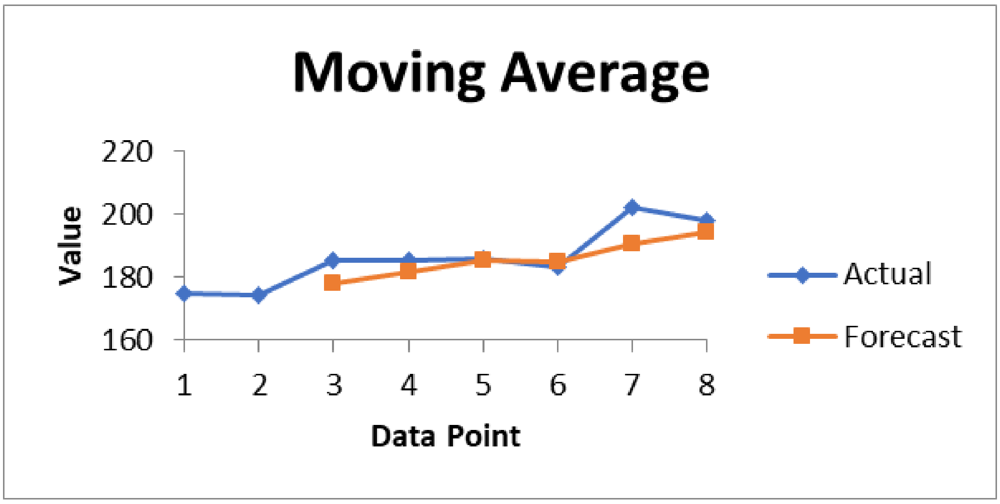

Figure 11: The Moving Average Model of Number of SHS Enrollees

Figure 11 shows the actual data gathered and its forecast. The model for moving average is $y=(Y_t + Y_{t+1} + Y_{t+2})/3$

h. Exponential Smoothing

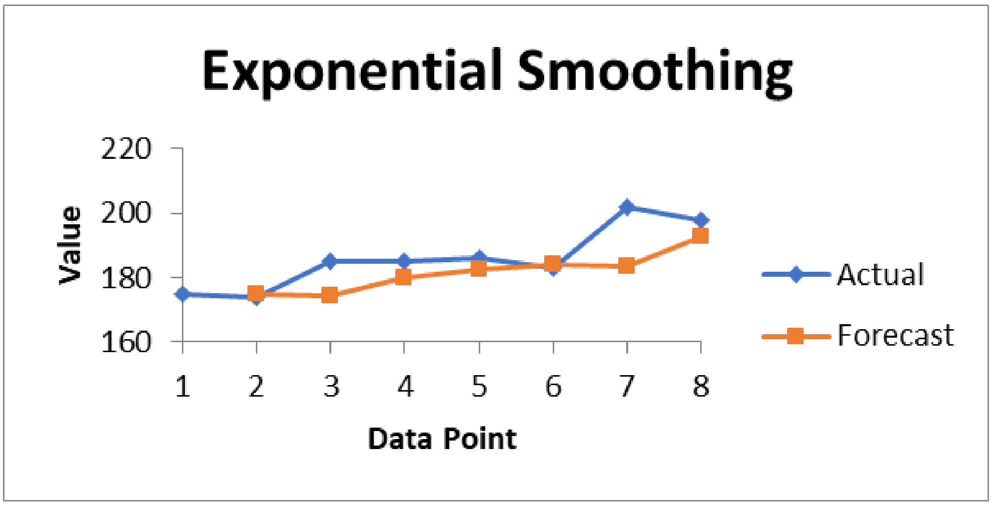

Figure 12: The Exponential Smoothing Tome Series Model of Number of SHS Enrollees

Figure 12 shows the exponential smoothing of the number of SHS enrollees. The model for exponential smoothing is $y=0.5y+0.5y_i$

i. Autoregression

i.1. First Autoregression

Table 5 shows the first autoregression result of the number of SHS enrollees from 1st semester of 2018 to 2nd semester of 2022. The first autoregression model can be used in predicting the number of SHS enrollees. The first autoregression model is $y = 76.640 + 0.601y_{n-1}$. This can explain 34.46% of the variances.

Table 5: First Autoregression Result of the Number of SHS Enrollees

	Coefficients	Standard Error	T Stat	P-value
Intercept	76.6397	68.4853	1.1191	0.31
X Variable 1	0.6020	0.3712	1.6215	0.17

$R^2 = 0.3446$ **significant at 0.01 *significant at 0.05

i.2. Second Autoregression

Table 6 demonstrate the second autoregression result of the number of SHS enrollees from 1st semester of 2018 to 2nd semester of 2022. The second autoregression model can be used in predicting the number of SHS enrollees. The second autoregression model is $y = 48.5363 + 0.2811y_{n-1} + 0.4911y_{n-2}$. This can explain 28.81% of the variances.

Table 6: Second Autoregression Result of the Number of SHS Enrollees

	Coefficients	Standard Error	T Stat	P-value
Intercept	48.5363	135.7577	0.3575	0.74
X Variable 1	0.2811	0.4664	0.6027	0.59
X Variable 2	0.4911	0.7842	0.6263	0.58

P-value=0.60 $R^2 = 0.2881$ **significant at 0.01 *significant at 0.05

Summary Table

Table 7 shows the summary of the time series models of the number of SHS enrollees from the 1st semester of 2018 to 2nd semester of 2022. Table 2 illustrates that sextic polynomial-time series has the highest value of R^2 and Second Autoregression has the least value of R^2.

Table 7: Equation of Time Series Models and R-square

Model	**Equation**	R^2
a. Linear	y=3.5238x+170.14	77.15%
b. Quadratic	$y=0.1667x^2+2.0238x+172.64$	77.84%
c. Exponential	$y=170.67e^{0.0188x}$	77.74%
d. Polynomial		
d.1. Cubic	$y=0.0808x^3-0.9242x^2+6.1854x+168.64$	78.41%
d.2. Quartic	$y=-0.0398x^4+0.7967x^3-5.2481x^2+16.106x+161.89$	78.84%
d.3. Quintic	$y=-0.1609x^5+3.5804x^4-29.237x^3+106.98x^2-166.54x+260.5$	95.91%
d.4. Sextic	$y=-0.0347x^6+0.7766x^5-6.3312x^4+22.609x^3-32.129x^2+10.786x+179.25$	97.31%
e. Power	$y=171.12x^{0.062}$	69.3%
f. Logarithmic	y=11.524ln(x)+170.72	68.03%
g. Moving Average	$y=(Y_t+Y_{t+1}+Y_{t+2})/3$	N/A
h. Exponential Smoothing	$y=0.5y+0.5y_i$	N/A
i. Autoregression		
i.1. First Auto-regression	$y = 76.640 + 0.601y_{n-1}$	34.46%
i.1. Second Auto-regression	$y = 48.5363 + 0.2811y_{n-1} + 0.4911y_{n-2}$	28.81%

Section 4: Best fit Model and Prediction

Table 8 demonstrate the accuracy of each model by subtracting R^2 and standard error. Table 8 also shows the prediction of SHS enrollment for the first semester of SY 2022-2023.

Table 8: Best Fit Model of the Number of SHS Enrollees

Model	R^2	***SE***	***Difference***	**Prediction for the 1st sem of SY 2022-2023**
a. Linear	77.15%	5.07	72.08%	202
b. Quadratic	77.84%	5.00	72.84%	204
c. Exponential	77.74%	5.05	72.69%	202
d. Polynomial				
d.1. Cubic	78.41%	4.93	73.48%	208
d.2. Quartic	78.84%	4.88	73.96%	201
d.3. Quintic	95.91%	2.15	93.76%	103
d.4. Sextic	97.31%	3.10	97.31%	33
e. Power	69.3%	5.92	63.38%	196
f. Logarithmic	68.03%	6.00	63.03%	196
g. Moving Average	N/A	2.61	N/A	N/A
h. Exponential Smoothing	N/A	4.75	N/A	N/A
i. Autoregression				
i.1. First Auto-regression	34.46%	8.39	26.07%	195
i.1. Second Auto-regression	28.81%	8.75	20.06%	203

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Summary

This study was conducted in order to determine the linear relationship of the number of SHS enrollees and the ratios by year level, strand and gender, to examine the best fit model for number of SHS enrollees from 2018-2019 per semester and use the best fit model for predicting the number of SHS enrollees of first semester of SY 2022-2023.

The trendline models were considered in the study created by Microsoft Excel. The difference of the r-square and the standard error was used to determine the best fit model for the number of SHS enrollees.

Summary of Findings

1. The trend of the SHS enrollment from 2018-2022 per semester is increasing over time.
2. Among the three independent variables, only the ratio by strand has a significant linear relationship on the number of SHS enrollees. When the three independent variables were combined together, there exist a significant linear relationship. However, only the ratio by gender suggested a significant influence to the number of SHS enrollees.
3. The time series models using the data in 2018-2022 are:

- $y=3.5238x+170.14$ for linear,
- $y=0.1667x^2+2.0238x+172.64$ for quadratic,
- $y=170.67e^{0.0188x}$ for exponential,
- $y=0.0808x^3-0.9242x^2+6.1854x+168.64$ for cubic polynomial,
- $y=-0.0398x^4+0.7967x^3-5.2481x^2+16.106x+161.89$ for quartic polynomial,
- $y=-0.1609x^5+3.5804x^4-29.237x^3+106.98x^2-166.54x+260.5$ for quintic polynomial,
- $y=-0.0347x^6+0.7766x^5-6.3312x^4+22.609x^3-32.129x^2+10.786x+179.25$ for sextic polynomial,
- $y=171.12x^{0.062}$ for power,
- $y=11.524\ln(x)+170.72$ for logarithmic,
- $y=(Y_t + Y_{t+1} + Y_{t+2})/3$ for moving average,
- $y=0.5y+0.5y_i$ for exponential smoothing,
- $y = 76.640 + 0.601Y_{n-1}$ for 1st autoregression, and
- $y = 48.5363 + 0.2811Y_{n-1} + 0.4911Y_{n-2}$ for 2nd autoregression.

4. The best fit model is the sextic polynomial model. Using this model, the number of SHS enrollees for the first semester of SY 2022-2023 is 33.

Conclusion

The findings of the study established that the ratios by year level, strand and gender, when combined, has an effect to the SHS enrollment. For the SHS enrollment, the best fit model was $y=-0.0347x^6+0.7766x^5-6.3312x^4+22.609x^3-32.129x^2+10.786x+179.25$ which was a sextic polynomial trend model. This study was conducted in order to help the chosen school to take future action of their enrollment.

Recommendations

Based on the findings of the study, the following recommendations are offered:

1. For the SJBHSI, the results in this research should not take as conclusive since the data from 2018-2022 are too limited, hence, suggesting a further study. This can help the school to plan financially and strategically in order to accommodate the growth of senior high school enrollment. This study can also help the school to have an idea about how many classrooms are still needed to be constructed for the next semester.
2. For the future researchers, the criteria for choosing the best fit trend model is insufficient since it is only the difference between the value of r-squared and the standard error. T-value, Mean Absolute Deviation (MAD), Mean Forecast Error (MFE), and Average Forecast Accuracy

(AFA) can be added to the criteria for selecting the best fit model from the given data. Seasonal Indices can also be added after selecting the best fit model.

REFERENCES

M. Koopmans (2011). Time Series in Education: The Analysis of Daily Attendance in Two High Schools. *Annual Convention of the American Educational Research Association*. https://files.eric.ed.gov/fulltext/ED546476.pdf

Y. Chen, R. Li & L. Hagedorn (2019). Undergraduate International Student Enrollment Forecasting Model: An Application of Time Series Analysis. *Journal of International Students Volume 9, Issue 1*. https://www.ojed.org/index.php/jis/article/view/266/544

J. Libo-on &A. Gadon (2021). Forecasting Enrollment: A Substitute to Career Guidance Campaign. *Romblon State University Research Journal*. https://ojs.rsu.edu.ph/index.php/rsurj/article/view/58/40

A. Dela Cruz, M. Basallo, B. Bere, J. Aguilar *et. al.* (2020). Higher Education Institution (HEI) Enrollment Forecasting using Data Mining Technique. *International Journal of Advanced Trends in Computer Science and Engineering Volume 9 No. 2*. https://www.academia.edu/download/63271852/ijatcse17992202020200511-73520-1ukh4h7.pdf

J. Libo-on. Predictors of Junior High School Enrollment for Proposed Appropriate Forecasting Model with Confirmatory Factor Analysis. *Journal of Business, Governance, & Information Technology pp 24-35*. http://web1.aup.edu.ph/urc/wp-content/uploads/2021/09/FIDERE-Journal-for-Business-Dec-20201.pdf#page=24