



## Forecasting Analysis of New Student Candidate Admissions Using the Simple Linear Regression Method

Sardo Pardingotan Sipayung<sup>1</sup>, Zakarias Situmorang<sup>2</sup>, Zekson Matondang<sup>3</sup> Masdiana Sagala<sup>4</sup>  
<sup>1,2,3,4</sup>Universitas Katolik Santo Thomas Medan Jl. Setiabudi No. 479 F Tanjungsari Medan

### Article Info

#### Article history:

Received, Nov 28, 2023

Revised, Des 20, 2023

Accepted, Des 27, 2023

#### Keywords:

Forecasting,  
Student Admission,  
Regression Method,  
Mean Absolute Percentage  
Error

### ABSTRACT

The development of science and technology facilitates various aspects of life, including forecasting. Forecasting student enrollment in private universities can maximize the use of resources for services, facilities, infrastructure, and improving human resources. The regression method is used to measure the effect of promotional costs on increasing student enrollment in the future. This forecasting will be valid if an accurate model is used. The results showed the level of accuracy using the MAPE (Mean Absolute Percentage Error) model of 2.229%. However, the level of accuracy can vary each due to differences in data.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Sardo Pardingotan Sipayung  
Universitas Katolik Santo Thomas Medan  
Jl. Setiabudi No. 479 F Tanjungsari Medan  
Email: [pinarsiphom@gmail.com](mailto:pinarsiphom@gmail.com)

## 1. INTRODUCTION

Admissions is an important aspect of higher education management. This process allows educational institutions to efficiently plan human resource and infrastructure requirements, as well as ensure the smooth running of the educational process. Therefore, forecasting admissions is crucial in the strategic planning of higher education institutions. In this context, the simple linear regression method is useful in performing forecasting. This method allows analyzing the linear relationship between independent variables (e.g., historical enrollment data) and dependent variables (future enrollment numbers). By understanding this relationship, colleges can make more accurate forecasts of future enrollment.

According to Muhammad et al. (2020), prediction is an action to estimate future conditions based on past data. One method that can be used for prediction is the linear regression method. This method consists of two types, namely simple linear regression and multiple linear regression. This study uses a simple linear regression method with one independent variable.

Several previous studies have tested the effectiveness of this method. Research by A. A. Azahra (2022) discusses the prediction analysis of the number of new student admissions using the simple linear regression method. The results of this analysis are reinforced by the MAPE method which produces a value of 6.04% for predicting the number of new students in the next 5 years. Similar research was conducted by Bhakti et al. (2020), which also discusses forecasting student admissions using the regression method, with the forecasting results strengthened by the MAPE method which produces a value of 2.7%. Another study by Amiruddin and Ishak (2020) showed an error rate of 4.24% for the informatics engineering study program and 7.69% for the law study

program. In this study, the application built has an accuracy rate of more than 90%, allowing it to predict the number of students precisely.

From previous research, there is a realization that the  $x$  variable used with a certain interim period still produces a significant gap from the forecasting results. Therefore, this study uses promotional costs as variable  $x$ . Promotion costs can be added to this study to expand the scope of analysis and improve the accuracy of predicting new student admissions. The use of promotional variables can provide valuable insights into the effectiveness of promotional efforts in attracting new prospective students to a particular college. This variable can include various promotional strategies such as advertisements in the mass media, educational exhibitions, school visits, and other promotional activities. By incorporating promotional variables into a simple linear regression model, the study can evaluate the extent to which the costs and types of promotions undertaken contribute to the number of new student admissions. This will assist universities in planning and optimizing their promotional budgets to achieve their desired student enrollment goals.

The purpose of this research is to apply the simple linear regression method in forecasting analysis of new student enrollment in higher education. Using historical enrollment data and other relevant variables, this research aims to develop a forecasting model that can assist universities in planning the capacity and resources needed to handle new students. This research also aims to improve the efficiency and effectiveness of admission management in higher education. By having an accurate forecasting model, universities can make better decisions regarding resource allocation, student recruitment, and promotion strategy development. The ultimate goal is to help universities achieve their admission targets more precisely and efficiently.

## 2. METHODS

This study uses linear regression to forecast student enrollment for the next year. Linear regression is an algorithm that models cause-and-effect relationships, with a cause variable ( $X$ ) and an effect variable ( $Y$ ). The stages that the author performs in this research include:

### 1. Data Collection

This research uses secondary data obtained from interviews with the Promotion and Admission Department of Santo Thomas Catholic University. Secondary data is data that has existed before and is usually obtained through a census or survey. This study uses data on promotion costs and the number of student admissions from 2019 to 2023.

### 2. Simple Linear Regression Method

Simple linear regression helps understand the relationship between variables. Variables that influence are called independent, while those that are influenced are called dependent.

### 3. Research Variables

#### a. Independent variable

Independent variables are variables that do not depend on other variables that cause changes in the dependent or dependent variable. or the emergence of the dependent variable. In this study, the cost of promoting student admissions is the independent variable that affects the dependent variable.

#### b. Bound Variable (Dependent)

The dependent variable (Dependent) is the variable that is influenced by the independent or independent variable. In this study, The number of student admissions from the previous year is the dependent variable which is influenced by promotional costs.

### 4. Data Analysis

#### a. Data Used

The data needed for this study are the number of student admissions from the previous year and the promotional costs each year. The available data includes student admissions from 2014 to 2019 as well as promotional cost data as listed in Table 1.

Tabel 1. Data Table 1. Revenue and Promotion Cost Data for the Last 5 Years

No	Year	Promotion Cost (Millions in Rupiah)	Number of Student Admission
1	2019	500	1200
2	2020	500	1100
3	2021	500	1238
4	2022	500	1375
5	2023	600	1536

b. Creating a regression table

In the first stage to perform simple linear regression calculations in forecasting the number of student admissions in 2024/2025 is to create a helper table to find X<sup>2</sup>, and XY.

Table 2. Regression Table

No	Year	Promotion Cost (Millions in Rupiah)	Number of Student Admission	X <sup>2</sup>	Y <sup>2</sup>	XY
1	2019	500	1200			
2	2020	500	1100			
3	2021	500	1238			
4	2022	500	1375			
5	2023	600	1536			

c. Simple Linear Regression Calculation Model

1. The regression equation model is as follows:

$$Y = a + bX$$

Where:

Y = dependent variable

a = coefficient a

b = coefficient b

x = forecasting value

5. Model for finding the values of a and b through:

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n(\sum X^2) - (\sum X)^2}$$

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

Correlation Coefficient Simple Linear Regression Correlation analysis is closely related to the regression method, but conceptually different from regression analysis. Correlation analysis will measure the level or strength of the relationship between two variables. Here is the formula for correlation:

$$r = \frac{(n \sum XY) - (\sum X \sum Y)}{\sqrt{(n \sum X^2 - (\sum X)^2)(n \sum Y^2 - \sum Y^2)}}$$

Where:

- n is the number of data pairs,
- X is the value of the independent variable,
- Y is the value of the dependent variable, and
- $\sum$  indicates the sum of each value.

MAPE (Mean Absolute Percentage Error) is used in this study to measure the accuracy of the regression model by comparing actual data and forecast data, so as to show how much error or gap between the two. The MAPE equation is as follows:

$$\text{MAPE} = \frac{n}{1} \sum_{t=1}^n \left[ \frac{|A_t - F_t|}{A_t} \right] \times 100\%$$

Where:

- $n$  is the number of periods or data observed
- $A_t$  is the actual value in the period  $t$ .
- $F_t$  is the forecast or predicted value in the period  $t$ .

MAPE measures the prediction error as a percentage of the actual value, and the lower the MAPE value, the more accurate the regression model.

### 3. RESULTS AND DISCUSSION

Based on the data in table 1, the first step is to calculate the forecast of student enrollment in the coming year using the linear regression method in 2020/2021 by first finding  $x_2$ ,  $y_2$ , and  $xy$ . The results of these calculations are in table 3.

Table 3. Regression Table Calculation Results

No	Tahun	Promotion Cost (Millions in Rupiah) X	Number of Student Admission Y	$X^2$	$Y^2$	XY
1	2019	500	1200	250.000	1.440.000	600.000
2	2020	500	1100	250.000	1.210.000	550.000
3	2021	500	1238	250.000	1.532.644	619.000
4	2022	500	1375	250.000	1.890.625	687.500
5	2023	600	1536	360.000	2.359.296	921.600

The next step is to find the values of  $a$  and  $b$  based on the model of finding the values of  $a$  and  $b$  in equations (2) and (3) with the results of calculating the values of  $a$  and  $b$  as follows:

$$a = \frac{(6449)(1,360,000) - (2600)(3,378,100)}{5(1,360,000) - (2600)^2}$$

$$a = -310.5$$

$$b = \frac{(5 \times 3,378,100) - (2600)(6449)}{5(1,360,000) - (6760000)}$$

$$b = 3.078$$

After obtaining the results from finding the values of  $a$  and  $b$ , equation (1) is formed using the simple linear regression method:

$$Y = -310.5 + 3.078x$$

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{(n \sum X^2 - (\sum X)^2)(n \sum Y^2 - (\sum Y)^2)}}$$

$$r = \frac{(5 \times 3,378,100) - (2600 \times 6449)}{\sqrt{((5 \times 1,360,000) - ((2600)^2))(5 \times 8,432,565 - (6449)^2)}}$$

$$r = 0.81$$

This means that there is a positive and strong relationship or correlation between promotional costs and student admissions. The greater the cost of promotion, the more student

admissions will increase in the following year. After knowing the correlation between the two variables X and Y, assumption data will be tried for variable X, namely promotional costs as a reason for forecasting student admissions in the following year. The assumed x value is 630.

$$Y = -310.5 + 3.078 x$$

$$Y = -310.5 + 3.078 \times 630$$

$$Y = 1628$$

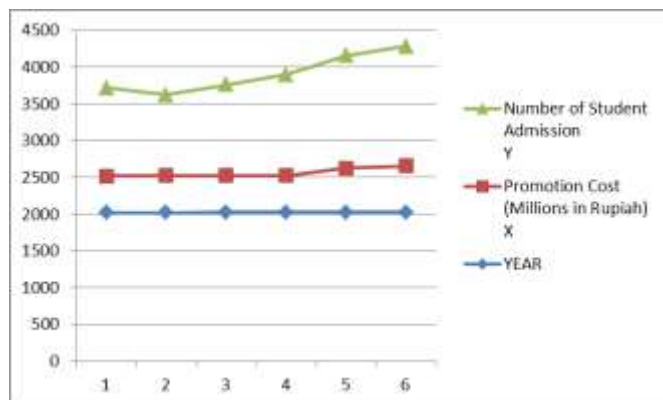


Figure 1. Data plot of Promotion Cost and Number of Students

The results of the Regression Graph obtained are as follows:

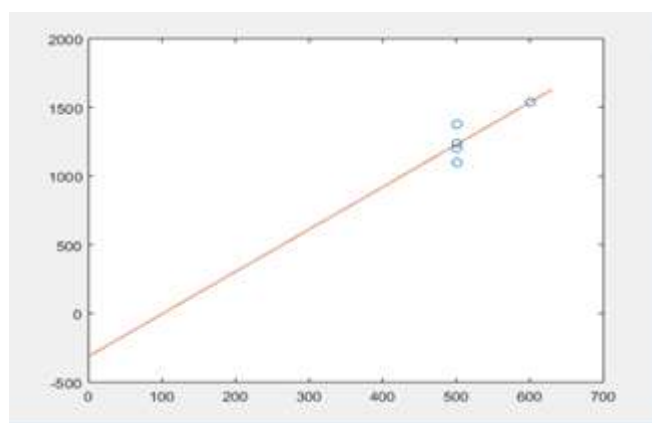


Figure 2. Regression Results

To determine the level of accuracy in calculations with the simple linear regression method, based on equation (5), the gap/error that will be generated with the last 3 years of data from the forecasting that has been done is as follows:

Year	Actual Data	Forecasting Data	Difference	Maape (%)
2021	1238	1228.25	33.75	0.0015
2022	1375	1228.25	-144.25	0.0213
2023	1536	1536	0	0
				0.0229

From the calculation, the gap value for forecasting the number of student admissions in the following year is 2.229%. For the results obtained between the forecasting value and the error value obtained, the forecasting value is  $Y = 1591$ . This proves that forecasting the number of students using the simple linear regression method is considered appropriate and improves the results obtained in previous research by [10] who said the use of the regression method was considered inappropriate, because the results obtained in this study got a very small gap/error value.

#### 4. CONCLUSIONS

The conclusion in the research conducted, that the author can draw conclusions based on the results obtained as follows: This research was made to assist in forecasting the number of student admissions in the coming year. The results obtained are only forecasts based on previous period data. Based on the results of calculations that have been carried out using data in the last 5 years and get the results of the equation of the regression method. Testing the correlation coefficient or the relationship between variables X and Y has a positive and strong correlation value, namely with the results of 0.81 of the correlation coefficient of promotional costs on increasing the number of student admissions can be done. The accuracy of the calculations carried out in the research on forecasting the number of student admissions using the regression method produces a gap / error value with a value of 2.229% using the MAPE (Mean Absolute Percentage Error) model improving the results of previous research conducted by Mulyono and RIFA, A., & Arifin, Z.. The conclusion provides answers to the problems identified in the Introduction section based on the results of discussion and testing. It can also present prospects for future development according to the results and discussion.

#### REFERENCES

- [1] Purba IS, Hartama D, and Kirana IO. Implementasi Algoritma Backpropagation dalam Memprediksi Jumlah Mahasiswa Baru pada AMIK-STIKOM Tunas Bangsa Pematangsiantar, Pros. Semin. Nas. Ris. Inf. Sci., Vol. 1, September 2019: 795, doi: 10.30645/senaris.v1i0.86.
- [2] Yordan, A., Putri, T. N., & Lamkaruna, D. H. PERAMALAN PENERIMAAN MAHASISWA BARU UNIVERSITAS SAMUDRA MENGGUNAKAN METODE REGRESI LINEAR SEDERHANA. J-TIFA, 2019; 2(1): 21-27.
- [3] Ritonga, A. S., & Atmojo, S. (2018). Pengembangan Model Jaringan Syaraf Tiruan untuk Memprediksi Jumlah Mahasiswa Baru di PTS Surabaya (Studi Kasus Universitas Wijaya Putra). Jurnal Ilmiah Teknologi Informasi Asia, 2018; 12(1): 15-24.
- [4] Kementerian Riset, "Statistik Pendidikan Tinggi 2017," Kementer. Riset, Teknol. dan Pendidik. Tinggi, pp. 1–7, 2018, doi: 10.1002/chem.200802548.
- [5] Siregar, Y. H. Prediksi Perilaku Pola Jumlah Mahasiswa Menggunakan Jaringan Syaraf Tiruan dengan Metode Backpropagation. JurTI (Jurnal Teknologi Informasi), 2018; 1(2): 145-152.
- [6] Bengnga, A., & Ishak, R. (2018). Prediksi Jumlah Mahasiswa Registrasi Per Semester Menggunakan Linier Regresi Pada Universitas Ichsan Gorontalo. ILKOM Jurnal Ilmiah, 2018; 10(2): 136-143.
- [7] Yasinta, S. A., & Yudihartanti, Y. Prediksi Daya Tersambung Dengan Metode Double Exponential Smoothing. Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi, 2017; 5(2): 1151-1162
- [8] Hidayat A. Pengertian Analisis Regresi Korelasi Dan Cara Hitung, Statiskian, 2012. <https://www.statistikian.com/2012/08/analisis-regresi-korelasi.html> (accessed Jan. 22, 2020).
- [9] RIFA, A., & Arifin, Z. Metode Regresi Linier Untuk Memprediksi Jumlah Pendaftar Calon Mahasiswa Baru Universitas Islam Indragiri. JURNAL ANALISIS MANAJEMEN, 2018; 4(2): 86-92.
- [10] Mulyono, "Analisis Regresi Sederhana," 2019. [https://bbs.binus.ac.id/management/2019/12/analisis-regresi-sederhana/#:~:text=Analisis regresi linier sederhana adalah,dengan variabel dependen \(Y\).](https://bbs.binus.ac.id/management/2019/12/analisis-regresi-sederhana/#:~:text=Analisis regresi linier sederhana adalah,dengan variabel dependen (Y).) (accessed Mar. 28, 2020).