ENSEMBLE
Journal of Applied Mathematics, Statistics, and Data Science
eISSN: xxxx-xxxx
Vol. X, No. X, XXXXXX
Page: 35-49

Analysis of Factors Affecting the Education Level of Provincial Children in Indonesia in 2023 Using the Multiple Linear Regression Method

Muhammad Zulfadhli[1*], Muhammad Yahya Matdoan[2]

[1]Departement of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Department of Statistics, Faculty of Science and Technology, University of Pattimura
Jl. Ir. M. Putuhena, Kampus Unpatti-Poka, Ambon, 97233, Indonesia

**E-mail Correspondence Author: ***muhammadzulfadhli23@gmail.com*

doi: xxxxxx

ARTICLE INFO

Article history

Received : Feb 14, 2026
Revised : Mar 17, 2026
Accepted : Apr 22, 2026

Keywords
Children's Education Level
Family Income
IIDN Residual Assumptions
Multiple Linear Regression
Per Capita Expenditure.

Abstract

Education is a critical driver for individual development and nation-building. This study aims to analyze the impact of economic and infrastructural factors, specifically average monthly per capita expenditure (X_1), family income (X_2), and access to educational facilities measured by the percentage of severely damaged schools (X_3), on the level of children's education (Y) across provinces in Indonesia in 2023. Utilizing secondary data from Statistics Indonesia (BPS), a multiple linear regression analysis was executed using Minitab software. The simultaneous F-test results indicate that the independent variables collectively exert a significant effect on the level of children's education. Partially, per capita expenditure shows a significant positive relationship with educational attainment, whereas family income exhibits a significant negative impact. Conversely, access to educational facilities does not significantly influence the response variable. Although the model successfully captured distinct regional disparities and outlier phenomena, the IIDN residual analysis reveals that the regression model does not fully satisfy the assumptions of being identical, independent, and normally distributed, yielding a coefficient of determination (R^2) of 18.81%. These findings emphasize the vital role of household economic allocation in educational attainment and suggest that policymakers should integrate broader contextual variables to mitigate regional educational disparities effectively.

1. INTRODUCTION

Education is a crucial pillar in the development of a nation and the main instrument in improving the quality of human resources (HR). Through quality education, a country can break the chains of poverty, promote sustainable economic growth, and create social stability. In the current era of globalization, strengthening the education sector is increasingly urgent so that the nation's competitiveness at the international level can continue to increase. Therefore, ensuring that every child has access to and is able to complete a decent level of education is a top priority in Indonesia's national development agenda.

As an archipelagic country with a very large and heterogeneous area, Indonesia faces a major challenge in the form of educational attainment gaps between provinces. Although the government has allocated a budget for the education function of 20% of the State Revenue and Expenditure Budget (APBN), the realization of children's education levels in various regions still shows significant disparities. This gap is strongly influenced by multidimensional factors, which can be broadly categorized into household economic aspects and supporting infrastructure aspects.

From an economic perspective, average per capita expenditure per month and family income are the main determinants that determine a household's ability to finance a child's education. Education costs include not only direct costs such as tuition or books, but also indirect costs including transportation, pocket money, and lost economic opportunities (*opportunity costs*). On the other hand, from an infrastructure perspective, the availability and feasibility of school physical facilities plays an equally important role. The condition of heavily damaged school buildings in several regions in Indonesia is allegedly a serious obstacle that reduces the sense of security, comfort, and effectiveness of the teaching and learning process, which ultimately has a negative impact on the continuity of children's education.

Although much research has been done on determinants of education levels, most previous studies have tended to focus on one aspect partially or use limited area coverage. In addition, post-pandemic dynamics in the 2023 data require the latest analysis to see the extent to which economic factors and current infrastructure conditions interact with each other affect the portrait of education at the provincial level. This study uses the Multiple Linear Regression Analysis method to model the linear relationship between these independent variables on children's education level in Indonesia. The use of this statistical method allows the identification of the most dominant variables as well as testing of residual assumptions to ensure the validity of the resulting model.

Based on this background, this study aims to analyze and evaluate the effect of average per capita expenditure, family income, and the percentage of schools with severely damaged conditions simultaneously or partially on the level of interprovincial children's education in Indonesia in 2023. The results of this research are expected to make an empirical contribution to the development of applied statistics and become a strategic consideration for policymakers (central and regional governments) in formulating a more targeted education equity program.

2. METHOD

2.1. Data Sources

The data used in this practicum is secondary data obtained from data sources obtained from https://www.bps.go.id/id website. Average per capita expenditure per month, monthly income, and access to educational facilities by province in 2023 as predictor variables (X_i), while children's education level as response variable (Y).

2.2. Research Variables

In this research, using different variables, here are the variables used.

Table 1. Research Variables

Variabel	Remarks	Scale	Units
X_1	Average Per Capita Expenditure Per Month By Province In 2023	Ratio	Rupiah
X_2	Net Income for a Family Month by Province in 2023	Ratio	Rupiah
X_3	Access to Educational Facilities by Province in 2023	Ratio	Presses
Y	Children's Education Level by Province in 2023	Ratio	Presses

2.3. Analysis Steps

The analysis steps carried out in this practicum are as follows.

1. Collecting data on average per capita expenditure per month, family income, access to educational facilities, and children's education level by province in 2023 in Indonesia through the official website of the Central Statistics Agency (BPS).
2. Analyze the characteristics of the data on average per capita expenditure per month, family income, access to educational facilities, and children's education level by province in 2023 in Indonesia using descriptive statistics.
3. Perform correlation tests on the data.
4. Perform multiple linear regression analysis on the data.
5. Interpret the results of the analysis.
6. Draw conclusions and suggestions.

2.4. Descriptive statistics

Descriptive statistics are statistics that can only be used to describe and also analyze data groups without being accompanied by drawing conclusions about which of the groups are large. Descriptive statistics are also often referred to as data descriptions that calculate the variance of raw data. And it is managed and described using tables or graphs or diagrams so that the data is clearer and easier to understand. (Hanafiah, 2020)(Hanafiah, 2020)**.**

2.4.1 Track-Track (*Mean)*

Average (*red*) is usually written in statistics using the symbol ( ) read exbar. Average ($\bar{X}$*red*) is a group explanation technique based on the average of the group. To find the average result ((Icam Sutisna, 2020)*red*) from a single data set, it can be searched by summing all the existing data and then dividing it by the amount of existing data. The formula for finding the average is as in Equation 2.1.(Icam Sutisna, 2020)

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} f_i} \tag{2.1}$$

Description :
$\bar{X}$ = Average
xi = Value i-i
fi = Frequency of the ith class
n = Total Data

2.4.2 Median

The median is the middle value of a group of data whose value of each observation has been arranged from smallest to largest. Insensitive to extreme values. The median is used to measure the centering if the distribution is convoluted (*Skewed*) clearly . The formula for the median is as Equation 2.2.(Icam Sutisna, 2020)

$$Me = \left[\frac{X\left(\frac{n}{2}\right) + \left(X\frac{n}{2}+1\right)}{2}\right] \tag{2.2}$$

Description:
And = median or Middle value
n = amount of data

2.4.3 Boxplot

Boxplot is a graphical display to illustrate the distribution of data using the median as well as the lower and upper quartiles (Widodo, 2017). *Outlier* we can easily detect through *boxplot* where each point above or below *whiskers mewakili outlier.* However, they have untapped potential they can be considered a kind of symbolic variable since they can summarize the data that stores pigtail information .(Widodo, 2017)

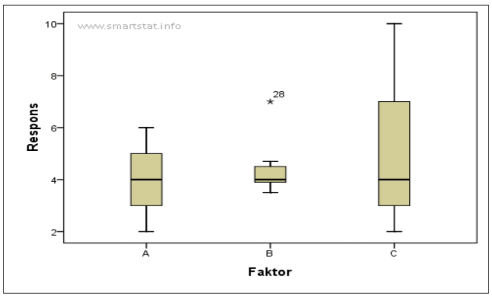


Figure 1. Example of a Boxplot
Source: https://accounting.binus.ac.id

2.4.4 Outlier

Outlier data can be defined as an observation that deviates so far from other observations. When researchers want to use parametric statistical methods, they are often constrained by the fulfillment of data normality assumptions. This is one of the reasons why data can be obtained (Sihombing P. R.-3., p. 2022)*Outlier* (isolated) checking *Outlier* The data is also carried out by looking at the standard value (z score) from the data. A data is considered a value *Outlier* if the value of the z score < -3.00 or > 3.00. The data is transformed using the formula in Equation 2.3.

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (2.3)$$

2.4.5 Scatterplot

Scatterplot is a graph used to visualize the relationship between two variables (Djemma & Sumangelipu, 2018). In this graph, each point represents an observation or data point, with the x-axis denoting one variable and the y-axis denoting the other. *Scatterplots* make it possible to identify patterns, relationships, or trends in the data, as well as help in determining the extent of the correlation between the two variables. Scatterplots are used to observe the relationships between variables (Djemma & Sumangelipu, 2018). The *Scatterplot* image is shown in Figure 2.2.

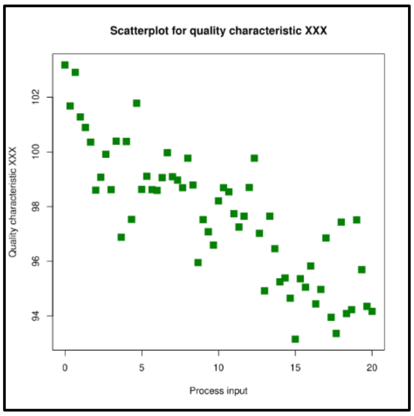


Figure 2. Scatterplot
Source : https://en.wikipedia.org/wiki/Scatter_plot

2.5 Correlation test

Correlation Test is a linkage method or research method that tries to connect one element/element with another element/element to create a new shape and form that is different from the previous one. Correlation tests are also used to measure and understand the relationship between two or more variables. This test provides information about the strength and direction of the relationship between these variables. Here is the hypothesis for the correlation test.(Rosyadi & Suyantiningsih, 2020)

$H0: \rho_{xy} = 0$ (There is no relationship between variable X and variable Y)

$H1: \rho_{xy} \neq 0$ (There is a relationship between variable X and variable Y)

A significant level of α is determined so that a critical area or a rejection area of H0 is obtained, namely if the th> calculation is tα/2,n-2 or *p-value* < α. The test statistics for the correlation test are as in Equation 2.4 and Equation 2.5.

$$t = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy^2}}} \tag{2.4}$$

$$r_{xy} = \frac{n\sum_{i=1}^{n} xiyi \sum_{i=1}^{n} xi \sum_{i=1}^{n} yi}{\sqrt{[n\,(\sum X^2) - (\sum X)^2][n\,(\sum Y^2) - (\sum X)^2]}} \tag{2.5}$$

Description :

T : Correlation Test Statistics

rxy : correlation coefficient

xi : observe ke-i, i = 1.2,..,n

yi : answer ke-i, i = 1.2,..,n

$\bar{x}$: average observation

$\bar{y}$: rat-rat responds

n : amount of data

2.6 Regresi Linear Berganda

Multiple linear regression is a regression technique that has many independent variables. One of the advantages of multiple linear regression analysis is that it is able to predict future conditions through the measurement of several independent variables (X) with non-independent variables (Y). The multiple regression parameter estimation is as Equation 2.5, Equation 2.6, and Equation 2.7.(Prasetyo, 2022)(Prasetyo, 2022)

$$Y = X\beta + \varepsilon \tag{2.5}$$

$$\beta = (XTX)^{-1}XTY \tag{2.6}$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots . + \beta kXk \tag{2.7}$$

Description :

Y : variable responds

X_1 : 1st predictor variable

XK : the variable predictor k-K

β : matriks parameter regresi

β_0 : constant (value Y when X = 0)

β_1 : Model Regressy K-1

BK : Model Regression K-K

ε : Error

2.6.1 Concurrent Tests

The concurrent test, also called the F test, is a statistical method used to test the influence of all intrinsic variables together on the dependent variables in a regression model. In other words, together look at how the free variable is influenced by the non-free variable. This test helps determine whether or not the regression model used is significant. The hypothesis for the simultaneous test is as follows.(Laili, 2023)

H_0 : $\beta_j = 0$, j = 0.1, . . , (variable X has no significant effect on variable Y)

H_1 : there is at least one $\beta_j \neq 0$ (variable X has a significant effect on variable Y)

A significant level of α is set so that a critical area or a rejection area of H0 is obtained if Fcalculate > F(1-α,n-k-1) or p-value < α. The Test statistics used for this analysis are shown in Table 2.

Table 2. Tabel ANOVA

Source Variasi	Degrees Free	Number of Squares	Average Amount Squares	F Calculate
Back	1	$SSR = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = b_1 SXY$	$\frac{SSR}{1}$	$\frac{MSR}{MSE}$
Galat	n-2	$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$	$\frac{SSE}{n-2}$	
Total	n-1	$SST = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = SYY$		

Description :

N : amount of data

$F_{hcalculation}$: test statistics

SSR : sum of the regression squares

SSE : sum squared error

SST : total number of squares

MSR : The Middle Square of Regression

MSE : The Error Is The Middle Side

2.6.2 Partial Test

A partial test, usually called a t-test, is a statistical method used to test the influence of each independent variable individually on the dependent variable in the regression method. This test helps determine whether each independent variable has a significant influence on the dependent variable. The partial test was carried out after a decision to reject H0 in the simultaneous test. Here is the hypothesis on the partial test.(Laili, 2023)

H_0 : $\beta_k = 0$, k = 0.1, . . n (Variable X has no significant effect on variable Y)

H_1 : $\beta_k \neq 0$, k= 0.1, . . n (Variable X has a significant effect on variable Y)

A significant level of α is set so that a critical area or a subverse area of H0 is obtained if the t-count is > tα or *p-value* < α. The Test statistics used (1−2)(n−k−1) are shown in Equation 2.8.

$$t\ hitung = \frac{\beta k}{\sqrt{var\ \beta k}} \quad (2.8)$$

Description :

βk : estimasi parameter *slope*

$var\ \beta k$: varians estimasi parameter *slope*

2.6.3 Benefits of the Regression Model

The goodness test of the model is also called "*goodness of fit*". This test is important to find out whether the model obtained is suitable or not. This test is used to measure how far the model is able to explain the variations of the dependent variables. The formula of the goodness of the model is as shown in Equation 2.9.

$$R^2 = 1 - \frac{SSerror}{SSTotal} = 1 - \frac{\sum(yi - yi\ hat)^2}{\sum(yi - yi\ bar)^2} \tag{2.9}$$

Description :
R2 : coefficient of determination
Yi : Observation of the first response
$\bar{Y}$: average varible Y
$\hat{y_i}$: Predictions of the first response

2.7 IIDN Residual Assumptions

The examination of IIDN residual assumptions (Identical, Independent, Normal Distribution) is a test that must be carried out whether the data used meets these three assumptions in conducting the test. The residual IIDN assumptions are described as follows.(Sudjana, 2006)

2.7.1 Identical Residual Assumption Checks

An examination of identical residual assumptions is performed to see if the residual meets the identical assumptions. A data is said to be identical if the residual plot is spread randomly and does not form a specific pattern. The average value of the variance is the same between one variance and another. The examination of identical residual assumptions is shown in figure 3.(Sudjana, 2006)

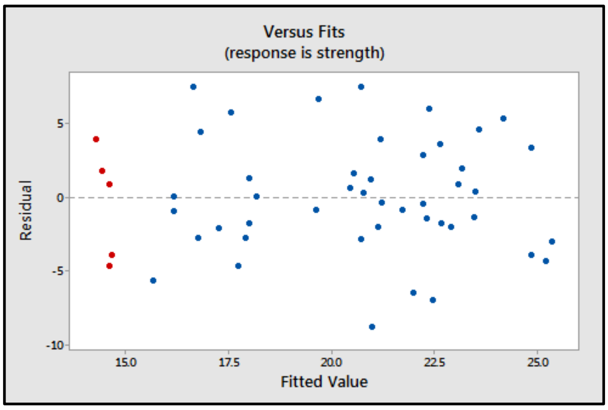


Figure 3. *Versus Fits*
(Source : *Minitab*)

2.7.2 Independent Residual Assumption Examination

An independent residual assumption check is performed to see if the residual meets the independent assumption. A data is said to be independent if its residual plot is spread randomly and does not form a specific pattern. The examination of independent residual assumptions is in figure 4.(Sudjana, 2006)

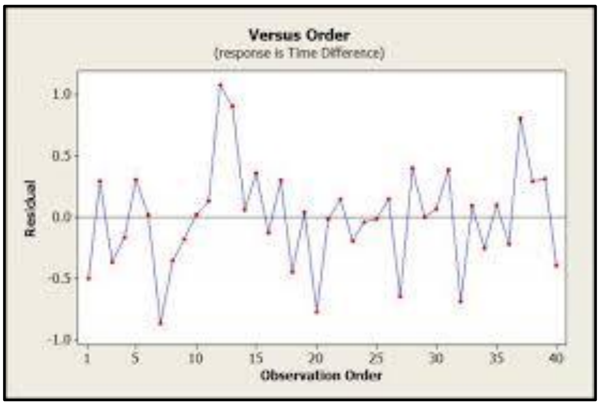
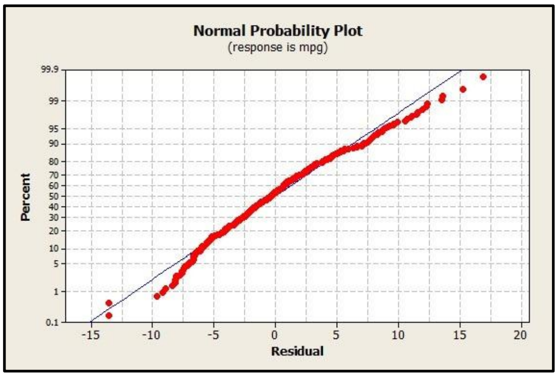


Figure 4. *Versus Order*
(Source : *Minitab*)

2.7.3 Normally Distributed Residual Assumption Testing

Testing of normally distributed residual assumptions is performed to see whether or not the residual meets the normally distributed assumptions. The normality of a data can be seen from the plot. If the plot is close to linear, it can be said that the data meets the assumption that it is normally distributed. The examination of the normal distributed residual assumptions is shown in Figure 5.(Sudjana, 2006)



Gambar 5. *Normal Probability Plot*
(Source : *Minitab*)

2.8 Average Per Capita Expenditure Per Month

Average per capita expenditure per month is the amount of money spent by each household member for consumption for one month. This expense includes expenses for food, clothing, and other necessities, as well as various types of services. Per capita expenditure is calculated by dividing the total household expenditure by the number of household members.(Nurul & Moh, 2010)

2.9 Family Income

Family income is the total amount of income received by all members of the household in a given period, usually one month or one year. This income includes various sources such as salary, wages, and income from investments. Family income is used to meet daily needs, such as food, clothing, education, and health. (Deti, 2015)

2.10 Educational Facilities

Educational facilities are the means and infrastructure that support the teaching and learning process in schools. The facility includes classrooms, laboratories, libraries, and other learning equipment needed to create an effective learning environment. The availability and completeness of educational facilities greatly affect the quality of learning and the effectiveness of the school.(Sholihatul, Siti, & Endang, 2022)

2.11 Children's Education Level

Andrew E. Sikula states that education level is a long-term process that uses systematic and organized procedures, in which the managerial workforce learns conceptual and theoretical knowledge for common purposes. Another opinion according to Azyumardi Azra states that the level of education is an activity of a person in developing his abilities, attitudes, and forms of behavior, both for the present life and at the same time preparation for future life through certain or unorganized organizations.(Garnika, E & Najwa, L, 2022)

3. RESULTS AND DISCUSSION

3.1. Characteristic Data

The characteristics of Average Monthly Per Capita Expenditure, Family Income, Access to Educational Facilities, and Children's Education Level by Province in 2023 in Indonesia are as follows.

3.1.1 Average CharacteristicsChildren 's Education Level

Analysis of the characteristics of the Average Children's Education Level by Province in 2023 can be seen in the *boxplot* image as follows.

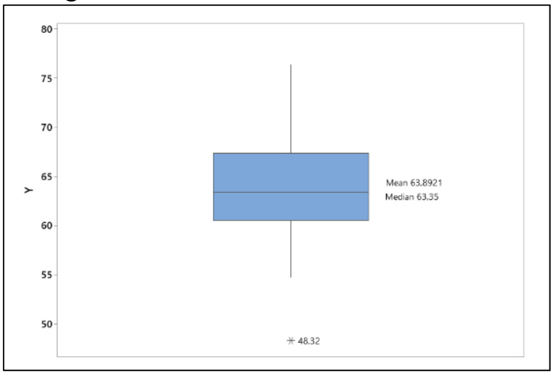


Figure 6 *Boxplot* of Average Children's Education Level

Figure 6 shows that the *mean* or average of the data on the Average Child Education Level by Province in 2023 is 63,982. The median or middle value of the Average Child Education Level data by Province in 2023, where 50% of the data is above 63.35 and 50% of the data is below 463.35. In Figure 4.1, it can be seen that the boxplot field is large, it shows the diversity of data on the Average Child Education Level by Province in 2023. There is also an improper median line in the middle of the *boxplot* field, which shows that *the boxplot* is not symmetrical. In the *boxplot* of the Average Child Education Level by Province in 2023, *outlier data was* found, namely 48.32% who have a percentage of children's education levels in Papua Province. *This outlier* is natural because Papua has serious problems in the region's education infrastructure, lack of education personnel, or limited access to education. This causes the level of education for children in the province to be much lower than other provinces in Indonesia.

3.1.2 Characteristics of Average Expenditure per Capita Per Month

An analysis of the characteristics of the Average Per Capita Expenditure per Month by Province in 2023 can be seen in the *boxplot* image as follows.

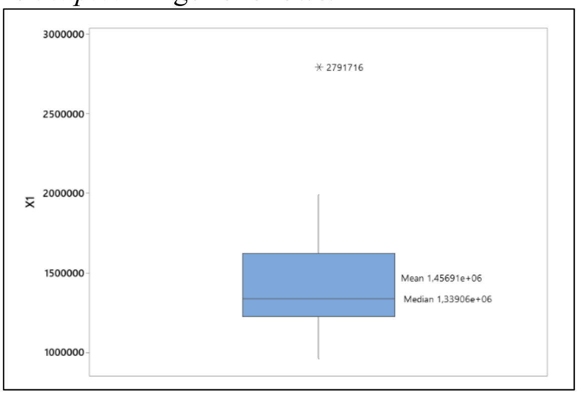


Figure 7. *Boxplot* Average Per Capita Expenditure Per Month

Figure 7 shows that the *mean* or average of the Per Capita Expenditure Per Month data by Province in 2023 is 1,457. The median or middle value of the Per Capita Expenditure Data per Month by Province in 2023, where 50% of the data is above 1,339 and 50% of the data is below 1,339. In Figure 4.2, it can be seen that the *boxplot* field is large, it shows the diversity of Per Capita Expenditure data per month by Province in 2023. There is also an improper median line in the middle of the *boxplot* field, which shows that *the boxplot* is not symmetrical. In the *Per* Capita Expenditure by Province boxplot in 2023, *outlier data was* found, namely 2791716 in DKI Jakarta Province. *This outlier* is natural because DKI Jakarta is the country's capital as well as a metropolitan city with a high cost of living and characteristics of people who tend to be consumptive or have a high lifestyle. This causes per capita expenditure in the province to be much higher than other provinces in Indonesia.

3.1.3 Family Income Characteristics

An analysis of the characteristics of Family Income by Province in 2023 can be seen in the *boxplot* image as follows.

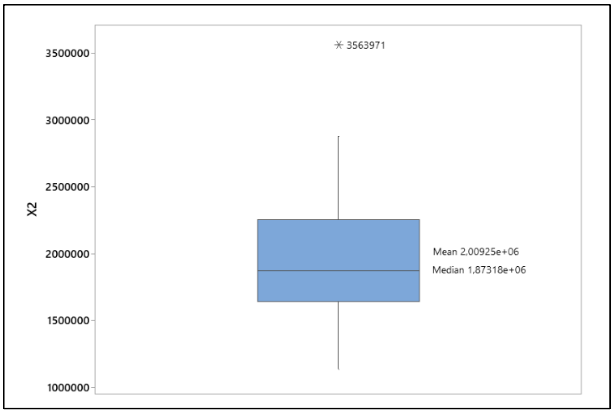


Figure 8. *Family Income* Boxplot

Figure 8 shows that the *mean* or average of the Family Income data by Province in 2023 is 2.009. The median or middle value of Family Income data by Province in 2023, where 50% of the data is above 1,873 and 50% of the data is below 1,873. In Figure 4.3, it can be seen that the *boxplot* field is large, it shows that the diversity of Family Income data by Province in 2023 is large. There is also an improper median line in the middle of the *boxplot* field, which shows that *the boxplot* is not symmetrical. In the *2023 Family Income by Province* boxplot, *outlier data was* found, namely 3563971 in DKI Jakarta Province. *This outlier* is natural because DKI Jakarta is the country's capital as well as a metropolitan city with a high cost of living and characteristics of people who tend to be consumptive or have a high lifestyle. This causes per capita expenditure in the province to be much higher than other provinces in Indonesia.

3.1.4 Characteristics of Heavily Damaged Educational Facility Access

An analysis of the characteristics of Access to Heavily Damaged Education Facilities by Province in 2023 can be seen in the *boxplot* image as follows.

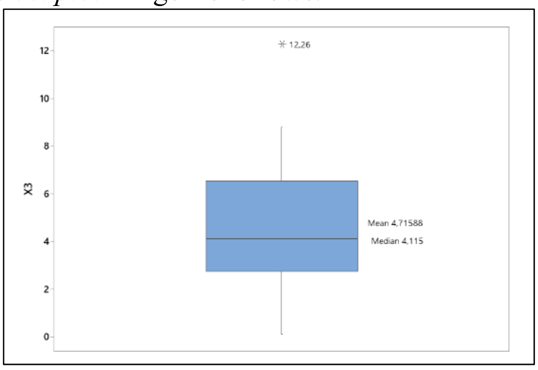


Figure 9 *Boxplot* Access to Heavily Damaged Education Facilities

Figure 9 shows that the *mean* or average of the data on Access to Heavily Damaged Education Facilities by Province in 2023 is 4,715. The median or middle value of the Data on Access to Severely Damaged Education Facilities by Province in 2023, where 50% of the data is above 4,115 and 50% of the data is below 4,115. In Figure 4.4, it can be seen that the boxplot field is large, it shows the diversity of Heavily Damaged Education Facility Access data by Province in 2023. There is also an improper median line in the middle of the *boxplot* field, which shows that *the boxplot* is not symmetrical. In the *boxplot* of Heavily Damaged Education Facility Access by Province in 2023, *outlier data was* found, namely 12.26%, which has a percentage of severely damaged educational facilities in Southeast Sulawesi Province. *This outlier* is natural because Southeast Sulawesi has serious problems in the education infrastructure in the region, such as lack of maintenance budget, geographical conditions that make it difficult to develop, or unequal attention to educational facilities in remote areas. This causes access to heavily damaged educational facilities in the province to be much higher than other provinces in Indonesia.

The existence *of outliers* in the Average Educational Level of Children, Expenditure Per Capita, Family Income, and Access to Heavily Damaged Education Facilities by province in 2023 shows that there is an inequality in distribution between regions in Indonesia. Therefore, in further analysis it is recommended to use a strong concentration measure such as the median, as well as consider information methods to reduce the influence of extreme values. In addition, paying attention to the presence of *outliers* can be better interpreted according to the purpose of the analysis.

3.2. Regresi Linear Berganda

Proper Multiple linear regression analysis of the effect of average per capita expenditure per month, family income, access to educational facilities on children's education level is as follows.

3.2.1 Parameter Estimation

Estimates of parameters from average per capita expenditure per month, family income, access to education facilities to children's education level are as follows.

$$Y = 59.63 + 0.000016\ X1 - 0.000010\ X2 + 0.174\ X3$$

The model shows that aspects of educational facilities have the most significant and positive influence. The child's education level will be worth 59.63% when the average per capita expenditure per month, family income, access to education facilities is zero. For every Rp 1 increase in monthly expenses, the child's education level will increase by 0.000016%, assuming that other variables are constant and because the value is small, it means that the influence of spending on children's education is not large directly, but can mean a large amount (for example, Rp 1 million => an increase of 0.016%). For every increase of Rp 1 in family income, the child's education level will actually decrease by 0.000010%, if other variables are constant (assuming the increase in income is not allocated for education). Every 1% increase in the quality of education facilities, will increase the child's education level by 0.174%.

3.3.2 Concurrent Test

The simultaneous test of the average per capita expenditure per month, family income, access to educational facilities to the child's education level is as follows.

H_0: b1 = b2 = b3 = 0 (average per capita expenditure per month, family income, access to educational facilities does not have a significant effect on the child's education level)

H_0: b1 = b2 = b3 = 0 (average per capita expenditure per month, family income, access to educational facilities have a significant effect on children's education level)

The significant level set is $\alpha = 0.05$ so that a critical area or a minus area of H_0 is obtained when $F_{\text{his calculated}} > F_{(0.05;\ 5;\ 31)}$ or P-value < 0.05. The table of values and simultaneous tests is as shown in Table 3.

Table 3 Concurrent Test

Fhcount	F(0.05; 5; 30)	P-value
2.32	0.116196	0.096

Table 3 shows that the results of the simultaneous test obtained a F_h value of 2.32 which is greater than the Ftable of 0.116196 and a P-value of 0.096 which is smaller than the α value of 0.05, so it was decided to reject H_0 which means that there is at least one factor that has a significant effect on the child's education level. With the decision to reject H_0, it will be continued with a partial test.

4.3.3. Partial Test

The partial test of the economic growth rate, the number of population, per capita expenditure, the Open Unemployment Rate (TPT), and the labor force on the GDP growth rate is as follows.

Hipotesis:

H_0: β1 = 0 (interception had no significant effect on children's education level)

H_1: β1 ≠ 0 (interception has a significant effect on children's education level)

The significant level set is α = 0.05 so that a critical area or a minus area of H_0 is obtained when $|t_h$ calculated$| > t_{(0.025; 30)}$ or P-value < 0.05. The partial table of interceptions is as shown in Table 4.

Table 4 Partial Test Interception

|Tcount|	t(0.025; 30)	P-*value*
9.17	2.360	0.000

Table 4 shows that the $t_{\text{hvalue is}}$ calculated by 9.17 which is greater than the table of 2.360 and the P-value of 0.000 which is smaller than the α value of 0.05, so it was decided to reject H_0 which means that interception has a significant effect on the child's education level.

Hipotesis :

H_0: β1 = 0 (average per capita expenditure has no significant effect on children's education level)

H_1: β1 ≠ 0 (average per capita expenditure has a significant effect on children's education level)

The significant level set is α = 0.05 so that a critical area or a minus area of H_0 is obtained when $|t_h$ calculated$| > t_{(0.025; 30)}$ or P-value < 0.05. The partial test table of average per capita expenditure is as shown in Table 4.4.

Table 5 Partial test of average expenditure per capita

|thcount|	t(0,025; 30)	P-*value*
24.82	2.360	0.000

Table 5 shows that the value of thet-value of 24.82 is greater than the ttable of 2.360 and the P-value of 0.000 which is smaller than the value of the α value of 0.05, so it was decided to reject H_0 which means that economic growth has a significant effect on the average expenditure per capita.

Hipotesis :

H_0: β2 = 0 (net income has no significant effect on the child's education level)

H_1: β2 ≠ 0 (net income has a significant effect on children's education level)

The significant level set is α = 0.05 so that a critical area or a minus area of H_0 is obtained when $|t_h$ calculated$| > t_{(0.025; 30)}$ or P-value < 0.05. The partial test table of net income is as shown in Table 6.

Table 6 Partial Test of Net Income

|thcount|	t(0.025; (34)	P-value
6.57	2.360	0.000

Table 6 shows that the value |thitung| of 6.57 which is larger than the ttable of 2.360 and the P-value of 0.006 which is smaller than the α value of 0.000, so it was decided to reject H_0 which means that the number of population has a significant effect on the GDP growth rate.

Hipotesis :

H_0: β3 = 0 (access to facilities has no significant effect on children's education level)

H_1: β3 ≠ 0 (access to facilities has a significant effect on children's education levels)

The significant level set is α = 0.05 so that a critical area or a minus area of H_0 is obtained when $|t_h$ calculated$| > t_{(0.025; 34)}$ or P-value < 0.05. The partial test table of per capita expenditure is as shown in Table 7.

Table 7 Partial Test of Access to Facilities

|Thitung|	$t(0.025; 34)$	P-*value*
-4.539	2.360	0.000

Table 7 shows that the value |thitung| of -4.539 which is smaller than the ttable of 2.360 and the P-value of 0.000. which is greater than the α value of 0.05, so it was decided to fail to reject H_0 which means that access to facilities does not have a significant effect on the child's education level.

4.3.4 Pros of the Model

The determination coefficient (R2) describes the magnitude of the response variable that can be explained by the predictor variable. It was obtained that the value of the determination coefficient or R *square* of the average data of the average per capita expenditure per month, family income, access to educational facilities to the child's education level was 18.81%, which means that the model can explain the variability of the data is only 18.81%, the variable of the average per capita expenditure per month, family income, access to educational facilities to the level of education of the child. Meanwhile, the remaining 81.19% cannot be explained by these three variables. This residual percentage reflects the influence of other factors outside the model.

4.4 IIDN Residual Assumption Examination

The following is an examination of identical residual assumptions, independent assumptions, and normal distribution at the child's education level, average per capita expenditure, net income, and access to provincial education in Indonesia in 2023. The following is an examination of identical, independent, and normally distributed residual assumptions presented in visual form.

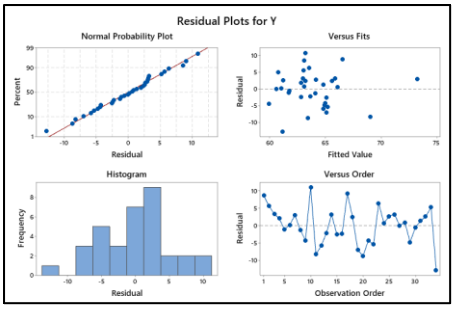


Figure 9. IIDN Residual Assumptions

Based on Figure 9 on the assumption of identical residual (*Versus Fits*) shows that the data does not spread randomly and forms a pattern. Independent residual assumptions also show that the data is spread randomly and does not form a pattern. Therefore, it can be concluded that the effect of average per capita expenditure per month, family income, and educational facilities on children's education level meets the residual assumption of IIDN.

4. CONCLUSION

Based on the results of the analysis of the effect of average per capita expenditure per month, family income, and educational facilities on children's education level, it is obtained as follows.

1. The characteristics of the *boxplot* data show that the four variables have considerable data diversity and asymmetrical distributions. *Outliers* were found in each variable, which reflects inequality between provinces. *Outliers* appear in provinces with extreme conditions, such as areas with low access to education, high cost of living, or poor educational infrastructure.
2. The results of *the scatterplot* and the average correlation test of per capita expenditure to children's education level showed that there was a positive relationship. The results of *scatterplots* and correlation tests of family income and access to educational facilities on children's education levels show that there is no linear relationship because the lines tend to form a horizontal trend.

3. The results of the multiple linear regression model show that if the average per capita expenditure increases, then the level of children's education tends to increase, while the increase in family income actually decreases the level of children's education. The variable of the percentage of schools that are severely damaged coefficient shows a positive, but not significant, effect on children's education levels. The results of the simultaneous test showed that average per capita expenditure, family income, and access to educational facilities together had a significant effect on children's education levels. The results of the partial test showed that average per capita expenditure and family income had a significant effect on children's education levels, while access to educational facilities (measured from severely damaged schools) had no significant effect. The goodness of the model is said to be not good, so there are many other factors that affect the level of children's education outside the model.

The results of the IIDN assumption examination on average per capita expenditure per month, family income, access to educational facilities, and children's education level meet identical, independent, and normally distributed assumptions.

Acknowledgments
The author would like to express gratitude to the Sepuluh Nopember Institute of Technology & University of Pattimura, as well as to the individuals and organizations that made personal contributions to this research.

Funding Information
There is no funding in this article.

REFERENCES

[1] A. M. Achmad, M. Zulfadhli, dan N. Nurfaudzan, "Binary logistic regression on the status of unemployment ratesa in West Java Province in 2023," *Likelihood: Journal of Statistics and Its Application*, vol. 1, no. 1, pp. 30–34, 2025.
[2] A. Aditia, "Covid-19: Epidemiologi, virologi, penularan, gejala klinis, diagnosa, tatalaksana, faktor risiko, dan pencegahan," *Jurnal Penelitian Perawat Profesional*, 2021. [Online]. Available: https://jurnal.globalhealthsciencegroup.com/index.php/JPPP/article/view/574
[3] A. Agresti, *Foundations of Linear and Generalized Linear Models*. Hoboken, NJ, USA: Wiley, 2015.
[4] S. M. Asnawati, *Analisis Inovasi dan Produk*. 2022.
[5] Badan Pusat Statistik (BPS), "Persentase penduduk miskin," 2020.
[6] Badan Pusat Statistik (BPS), "Tingkat Pengangguran Terbuka (TPT)," 2020. [Online]. Available: https://sirusa.bps.go.id/sirusa/index.php/indikator/44
[7] S. S. Bogusławski et al., "The outcomes of COVID-19 pneumonia in children: Clinical, radiographic, and pulmonary function assessment," *Pediatric Pulmonology*, 2023.
[8] R. Cahyandari, "Pengujian overdispersi pada model regresi Poisson (Studi kasus: kecelakaan lalu lintas mobil penumpang di Provinsi Jawa Barat)," *Jurnal Statistika*, 2014.
[9] Mayo Clinic, "Cough," 2022. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/cough
[10] I. I. Daqiqil, *Teori, Studi Kasus, dan Implementasi Menggunakan Python*. 2021.
[11] M. I. Dirwan, N. W. Surtini, dan M. Zulfadhli, "Learning by gaming: Extramural English gaming effect on Indonesian senior high students' pragmatic competence," *Journal of Pragmatics Research*, vol. 7, no. 1, pp. 1–20, 2025.
[12] A. N. Hidayati dan B. W., "Pelayanan puskesmas berbasis manajemen terpadu balita sakit dengan kejadian pneumonia balita," *Jurnal Kesehatan Masyarakat*, 2022.
[13] V. Silvia, *Statistika Deskriptif*. 2021.
[14] L. D. Martias, "Statistika deskriptif sebagai kumpulan informasi," *FIHRIS: Jurnal Ilmu Perpustakaan dan Informasi*, pp. 42–57, 2021.
[15] R. D. Mason, *Teknik Statistika untuk Bisnis dan Ekonomi*. Jakarta, Indonesia, 1996.
[16] J. F. Murray, *Textbook of Respiratory Medicine*. Amsterdam, Netherlands: Elsevier, 2016.
[17] Musbahuddin, *Analisis Data Penelitian dengan Statistik*. 2013.
[18] B. W. Otok, M. Zulfadhli, R. D. Pangesti, M. I. Kurniawan, A. E. P. Haryanto, dan I. Kurniawan, "Comparison of binary probit regression and Fourier series nonparametric logistic regression in modeling diabetes status at Hajj General Hospital Surabaya," *Barekeng*, vol. 20, no. 1, pp. 255–270, 2026.
[19] World Health Organization (WHO), "Pneumonia in children," 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia

[20] H. Purbaya, “Faktor-faktor yang berhubungan dengan kejadian pneumonia pada balita,” 2023.
[21] F. Rachmillah, “Prediksi jumlah kasus baru kusta dengan metode geographically weighted Poisson regression (GWPR),” *Jurnal Stikes Mandala*, 2015.
[22] R. T. Pamungkas, *Bahan Ajar Matematika Mememo (Mean, Median, Modus) untuk Siswa Kelas 6 SD*. CV Graf Literasi, 2020.
[23] S. Saida, M. Zulfadhli, dan M. Jurais, “Analisis faktor-faktor yang mempengaruhi *vaccine hesitancy* (keragu-raguan vaksin) pada mahasiswa di era pandemi COVID-19,” *Preventif: Jurnal Kesehatan Masyarakat*, vol. 13, no. 1, pp. 144–154, 2022.
[24] A. Saifuddin, *Buku Acuan Nasional Pelayanan Kesehatan Maternal dan Neonatal*. Jakarta, Indonesia: Yayasan Bina Pustaka Sarwono Prawirohardjo, 2020.
[25] S. Santoso, *Seri Solusi Bisnis Berbasis TI*. 2007.
[26] Y. R. Maharani dan S. Shinta, “Faktor risiko frekuensi kunjungan balita dengan kasus batuk,” *Jurnal Kesehatan Poltekkes Palembang*, 2020.
[27] D. Siagian, *Metode Statistik*. 2006.
[28] S. R. Wahyuningrum dan A. Muhlis, *Statistika Pendidikan Edisi Kedua (Dengan Statistika Al-Qur’an)*. 2019.
[29] M. H. Sriningsih, “Penanganan multikolinearitas dengan menggunakan analisis regresi komponen utama pada kasus impor beras di Provinsi Sulawesi Utara,” *Jurnal Ilmiah Sains*, vol. 18, no. 1, pp. 18–24, 2018.
[30] Dinas Kesehatan Provinsi Jawa Timur, *Profil Kesehatan Provinsi Jawa Timur*. Jawa Timur, Indonesia, 2022.
[31] K. A. Yulianingsih, “Penerapan regresi Poisson untuk mengetahui faktor-faktor yang memengaruhi jumlah siswa SMA/SMK yang tidak lulus UN di Bali,” *E-Jurnal Matematika*, vol. 1, no. 1, pp. 59–63, 2012.
[32] B. Yuniarto, *Analisis Regresi*. 2016.
[33] M. Zulfadhli, I. N. Budiantara, V. Ratnasari, A. S. Suriaslan, dan R. C. Dhewy, “A statistical inference framework for FSNBLR: Modeling underdeveloped regional status in Eastern Indonesia,” *MethodsX*, Art. no. 103746, 2025.
[34] M. Zulfadhli, I. N. Budiantara, V. Ratnasari, dan A. S. Suriaslan, “Semiparametric regression estimator of Fourier series for categorical data,” *IAENG International Journal of Applied Mathematics*, vol. 55, no. 7, pp. 2157–2164, 2025.
[35] A. Agresti, *An Introduction to Categorical Data Analysis*, 2nd ed. New York, NY, USA: John Wiley & Sons, Inc., 2007.
[36] M. Carapeto and W. Holt, “Testing for heteroscedasticity in regression models,” *J. Appl. Stat.*, vol. 30, no. 1, pp. 13–20, 2003.
[37] Z. Li and J. Yao, “Testing for heteroscedasticity in high-dimensional regressions,” *Econom. Stat.*, vol. 9, pp. 122–139, 2019.
[38] A. Račkauskas and D. Zuokas, “New tests of heteroskedasticity in linear regression model,” *Lith. Math. J.*, vol. 47, pp. 248–265, 2007.
[39] R. R. Wilcox and H. J. Keselman, “Detecting heteroscedasticity in a simple regression model via quantile regression slopes,” *J. Stat. Comput. Simul.*, vol. 76, no. 8, pp. 705–712, 2006.
[40] M. Yüce, “An asymptotic test for the detection of heteroskedasticity,” *Istanbul Univ. Econom. Stat. e-J.*, no. 8, pp. 33–44, 2008.