

Multivariate Regression Analysis of Factors Affecting the Human Development Index (HDI) and Gender Development Index (GDI) in East Java Province in 2024)

Zizia Aletha¹, Naila Widiana Ramdani², Baiq Tia Azhari³, Zakila Bilbina⁴, Alissa Chintyana^{5*}

^{1,2,3,4,5} *Department of Statistics, Hamzanwadi University, Indonesia*

*Corresponding author: alissachintyana@hamzanwadi.ac.id

*Submission date: 27 January 2026, Revision: 03 May 2026, Accepted: 20 October 2026

ABSTRACT

The Human Development Index (HDI) and the Gender Development Index (GDI) are important indicators to describe the development conditions in a region. This study aims to analyze the influence of socio-economic indicators on the HDI and GDI in regencies/cities in East Java Province. The data used comes from the official publication of the Central Statistics Agency (BPS) of East Java Province in 2024. The socio-economic variables studied include population density, Open Unemployment Rate (OUR), percentage of poor people (PPM), mean years of schooling (MYS), and Gross Regional Domestic Product (GRDP). The analysis was carried out using multivariate regression to estimate the influence of independent variables on two response variables, namely the HDI and GDI simultaneously. The results showed that in the HDI model, the MYS and GRDP variables had a significant positive effect on the HDI at a significance level of 5%, while population density, OUR, and PPM did not have a significant effect. In the GDI model, the OUR variable had a significant positive effect on the GDI at a significance level of 5%, while the other variables did not show a significant effect. In general, these results indicate that education and economic conditions play a role in increasing the HDI, while variations in the GPI are more influenced by the open unemployment rate. The multivariate regression model used provides a snapshot of the influence of socioeconomic variables on the HDI and GPI simultaneously.

KEYWORDS

Human Development Index, Gender Development Index, Multivariate Regression, East Java

1. INTRODUCTION

Differences in roles and division of responsibilities between women and men are generally acceptable as long as they do not lead to inequality or behavior that could harm either party. In terms of social development, the concept of gender equality is more widely used because it describes the balance of opportunities and benefits received by women and men in various development sectors. Gender inequality arises when individuals do not receive the same rights, opportunities, and development outcomes due to gender differences, whether in education, economics, or social matters. This condition not only affects individuals but also impacts the development process as a whole [1]

Development itself can be defined as a process of creating new conditions or improving existing conditions towards a better state. The main goal of development is to create an environment that supports people to live longer, healthier, and more productive lives. The Human Development Index (HDI) is an important indicator in describing the quality of development in a region. A high HDI value reflects an improving quality of life for the community, as seen in East Java Province, which is included in the high HDI category, indicating a relatively good population welfare condition [2].

According to data from the Central Statistics Agency (BPS), the Human Development Index (HDI) of East Java Province

has generally shown an increase in recent years. This indicates an improvement in the quality of human development in the region, with the HDI reaching relatively close to the national average. However, previous studies have shown that the HDI and GDI are influenced by socioeconomic factors such as education, labor market conditions, poverty, and regional economic activity [3]. Therefore, an analysis is needed to examine the influence of these factors on the level of human development and gender equality achieved in East Java Province.

The selection of East Java Province as the research area is based on its characteristics as a province with a large population and a high level of socioeconomic heterogeneity between districts/cities, thus providing a relevant context for examining the influence of socioeconomic factors on the HDI and GDI. Furthermore, although several previous studies have analyzed the factors influencing the HDI or GDI separately, studies that simultaneously analyze both indicators are still relatively limited, especially at the district/city level in East Java Province.

In previous research, the regression approach was generally used to analyze the influence of independent variables on a single response variable. However, in the context of human development and gender, the HDI and GDI are closely related, making analysis that considers both response variables simultaneously essential. Therefore, this study uses a multivariate regression approach, which allows testing the influence of independent variables on more than one response variable simultaneously [1]. This approach was chosen because it provides a more comprehensive picture of the influence of socioeconomic factors on the HDI and GDI compared to univariate regression analysis conducted separately.

2. LITERATURE REVIEW

2.1 Human Development Index (HDI)

The Human Development Index (HDI) is a composite indicator used to measure the success of human development through a multidimensional approach, which includes aspects of health, education, and a decent standard of living. The HDI was developed in response to the limitations of pure economic indicators such as Gross Domestic Product (GDP) which have not been able to describe the quality of human life as a whole. Methodologically, the HDI is calculated as the geometric mean of three dimensional indices, namely the health index, the education index, and the per capita expenditure index. The geometric approach is used to ensure that low achievement in one dimension cannot be fully compensated by high achievement in another dimension [4].

The HDI formula can be written as follows [5]:

$$IPM = \sqrt[3]{(I_{Kes} \times I_{Pend} \times I_{Peng})} \tag{1}$$

The health index is measured using life expectancy at birth, the education index is calculated from the expected length of schooling and the average length of schooling, while the expenditure index reflects the purchasing power of the community for the needs of a decent life [5].

2.2 Gender Development Index (GDI)

The Gender Development Index (GDI) is an indicator used to measure the gap in human development achievements between men and women [4].

The HDI is calculated based on the comparison of the HDI value of women to the HDI value of men, so this indicator directly reflects gender inequality in human development. The closer the number is to 100, the smaller the development gap between men and women [5].

Mathematically, IPG is formulated as follows [5]:

$$IPG = \left(\frac{IPM_{women}}{IPM_{men}} \right) \times 100 \tag{2}$$

The GDI is an important indicator in sustainable development planning because gender inequality has been proven to have a negative impact on the effectiveness of economic and social development in a region. The GDI is an important indicator in sustainable development planning because gender inequality has been proven to have a negative impact on the effectiveness of economic and social development in a region (gender inequality hampers economic development and inclusive growth) in many

recent studies that show a relationship between gender inequality and low economic growth, disparities in access to education, and limited participation of women in the workforce which are empirically related to slowing economic growth and social development [6].

2.3 Descriptive statistics

Descriptive statistics is the initial stage of data analysis that aims to provide a general overview of the characteristics of research data. This analysis is used to summarize data into representative numerical measures, thus facilitating the understanding of data patterns and distribution. Numerically, descriptive analysis is presented through summary statistics that include the minimum, maximum, average, and standard deviation values for each research variable. This statistical presentation aims to illustrate the measures of central tendency and dispersion of the data so that the characteristics and diversity of the data can be more clearly understood [7].

Data center measures (mean and median) are used to describe the typical value of each variable, while data distribution measures (minimum, maximum, quartiles, and standard deviation) are used to assess the level of variation and inequality of data between districts/cities. The presentation of these descriptive statistics serves as a basis for understanding the initial characteristics of the data, detecting possible outliers, and determining the feasibility of using further analysis in the form of multivariate regression. Descriptive statistics that include mean, median, minimum, maximum, and standard deviation are summaries of numerical data commonly used in quantitative research to describe the center and distribution of data before further analysis is carried out [8].

2.4 The relationship between HDI and IPG

The relationship between the HDI and the GDI reflects the link between the quality of human development and the level of gender equality. Regions with a high HDI generally have better access to education and health, which in turn can narrow the gender gap in human development. However, the relationship between the HDI and the GDI is not always linear because social, cultural, and public policy factors can influence women’s roles and access to development resources. Therefore, analyzing the relationship between the HDI and the GDI requires a statistical approach that is able to capture the structural linkages between development indicators [4].

2.5 Multicollinearity Detection (VIF)

Multicollinearity detection is a diagnostic procedure in regression analysis that aims to identify the presence of a strong linear relationship between independent variables. Multicollinearity is not tested through inferential statistics, but rather detected to assess the potential problem of high correlation that can cause large variance of regression coefficients and unstable parameter estimates. Empirical studies show that multicollinearity detection is an important initial stage before conducting regression modeling, especially in multiple linear regression and multivariate regression, to ensure the reliability of the interpretation of regression coefficients [9].

Multicollinearity is measured using the Variance Inflation Factor (VIF) [10]:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{3}$$

with R_j^2 is the coefficient of determination of the regression results of independent variable j against other independent variables. A VIF value smaller than 10 indicates no indication of serious multicollinearity between independent variables, so that the regression model can be considered worthy of interpretation [10].

2.6 Multivariate Regression

Multivariate regression is a statistical analysis method used to analyze the simultaneous influence of several independent variables on more than one dependent variable. This approach is particularly relevant when the dependent variables are correlated, such as the HDI and the GPI, so that separate analyses can produce less efficient estimates [11].

The multivariate regression model is generally formulated as [11]:

$$Y_1 = \beta_{10} + \beta_{11}X_1 + \beta_{21}X_2 + \dots + \beta_{p1}X_p + \varepsilon_1 \tag{4}$$

$$Y_2 = \beta_{20} + \beta_{12}X_1 + \beta_{22}X_2 + \dots + \beta_{p2}X_p + \varepsilon_2 \tag{5}$$

where:

Y_1 : Human Development Index (HDI)

Y_2 : Gender Development Index (GDI)

$X_1, X_2, \dots, X_p$: independent variables (Population Density, OUR, PPM, MYS, GRDP)

β_{kj} : regression coefficient that shows the influence of the kkkth independent variable on the jth dependent variable

$\varepsilon_1, \varepsilon_2$: Errors are correlated with each other, so the analysis is carried out multivariately

Multivariate regression allows researchers to understand the complex structure of relationships between variables and provides more comprehensive estimation results than univariate regression [11].

2.7 Significance Test (Wilk's Lambda)

Wilks' Lambda (Λ) is a multivariate test statistic used in MANOVA or multivariate regression to test whether independent variables simultaneously affect the dependent variable [12] .

This statistic measures the proportion of variation not explained by the model compared to the total variation. The value of Λ ranges from 0–1; Λ close to 0 $\rightarrow$ significant simultaneous effect, Λ close to 1 $\rightarrow$ weak effect [12].

Wilks' Lambda formula is stated as [12]:

$$\Lambda = \frac{|E|}{|E + H|} \quad (6)$$

where:

E : error matrix (residual)

H : matrix of variations explained by the model

$|\cdot|$: matrix determinant

Hypothesis on Wilks' Lambda [13]:

H_0 : All independent variables do not have a simultaneous effect on the dependent ($\beta = 0$)

H_1 : At least one independent variable has a simultaneous effect on the dependent variable ($\beta \neq 0$)

The Wilks' Lambda (Λ) test statistic can be transformed into an F-statistic [14]:

$$F = \frac{(1 - \Lambda)^{1/u} df_2}{\Lambda^{1/u} df_1} \quad (7)$$

where:

Λ : Wilks' Lambda, the ratio of the residual matrix determinant to the total variation

u : a correction factor calculated based on the number of dependent variables (p) and the number of independent variables (k), as well as the sample size (n).

df_1 : numerical degrees of freedom, usually related to the number of independent variables (k)

df_2 : degrees of freedom of the denominator, usually related to the number of samples (n) minus the number of independent and dependent variables.

If p-value < 0.05, H_0 is rejected. It is mean the independent variables are simultaneously significant.

Critical Areas:

Reject H_0 : Λ is small and p-value < α

Failed to reject H_0 : Λ approaches 1 or p-value $\geq \alpha$

2.8 Response Variable Dependency Test (Barlett's Test)

The Bartlett test is used to measure whether the dependent variables in a multivariate analysis are interdependent or significantly correlated. The assumption of independence between the response variables is necessary for a multivariate regression model or MANOVA to provide valid parameter estimates and for test statistics, such as Wilks' Lambda, to be interpreted correctly [14].

Mathematically, the Bartlett Test hypothesis is formulated as:

H_0 : The covariance matrix of the dependent (independent) variables is the same

H_1 : The covariance matrices of the dependent variables are not the same (dependent or correlated)

Bartlett's test statistic is expressed by the equation [14]:

$$X^2 = - \left(n - 1 - \frac{2p + 5}{6} \right) \ln |R| \tag{8}$$

where:
n: number of observations
p: number of dependent variables
 |*R*|: correlation matrix determinant
 The test statistic follows a chi-square distribution with degrees of freedom [15]:

$$df = \frac{p(p - 1)}{2} \tag{9}$$

If the p-value < α then *H*₀ is rejected, which indicates that there is a significant dependency between the dependent variables, and the multivariate model must consider the relationship between these variables. This test is important because the assumption of independence or homogeneity of covariance affects the reliability of multivariate statistics such as Wilks' Lambda, Lawley-Hotelling trace, and Pillai's Trace [15].

2.9 Multivariate Residual Normality Test

Multivariate normality of residuals is a basic assumption in multivariate regression which states that the residual vectors of all regression equations follow a multivariate normal distribution [15]. This assumption is important because it ensures that the regression parameter estimates are efficient and unbiased, guarantees the validity of multivariate significance tests such as Wilks' Lambda, and reduces the risk of inference errors that arise if the residual distribution deviates from normal [12].

If multivariate normality is not met, the distribution of the test statistic may deviate from the theoretical distribution, making the research conclusions less reliable. Therefore, multivariate residual normality testing is performed before multivariate hypothesis testing [13]. Testing was carried out using the Mahalanobis distance , which is a multivariate distance measure that takes into account the covariance between variables, so that it is able to evaluate the suitability of the residual distribution simultaneously [15].

If the residuals are multivariately normally distributed, then the squared Mahalanobis distance value will follow a Chi-square distribution with degrees of freedom equal to the number of response variables. Normality is assessed by comparing the empirical Mahalanobis distance value to the theoretical Chi-square quantile and looking at the proportion of observations that fall below that quantile [16].

Hypothesis:
*H*₀: Multivariate residuals follow a multivariate normal distribution
*H*₁: Multivariate residuals do not follow a multivariate normal distribution

Test Statistics (Mahalanobis distance) The test statistic used in multivariate residual normality testing is the squared Mahalanobis distance (*D*²) , which is formulated as follows [15]:

$$D^2 = (e_i - \bar{e})^T S^{-1} (e_i - \bar{e}) \tag{10}$$

where:
e_i: *i*-th residual vector
 $\bar{e}$: residual mean vector
S⁻¹: inverse residual covariance matrix

In conditions where the residuals are distributed normally multivariately, the values *D*² follow a Chi-square distribution with degrees of freedom equal to the number of response variables. The test decision is made by comparing the *D*² empirical value to the theoretical Chi-square quantile , and evaluating the proportion of observations that have a value *D*² smaller than that quantile [16].

2.10 Best Model Selection & Evaluation

Selecting the best model in multivariate regression aims to obtain the most efficient model in explaining the relationship between several dependent and independent variables, without including insignificant variables. The best model should be statistically significant, meet the assumptions of multivariate regression, and have a minimum number of parameters. Critical Area Evaluation[12]:

1. The results of the Mardia test calculated the p-value of skewness and kurtosis.
2. Reject H_0 if $p\text{-value} \leq \alpha \rightarrow$ residuals are not normal.
3. Fail to reject H_0 if $p\text{-value} > \alpha \rightarrow$ normal multivariate residuals, can be used for Wilks' Lambda testing and multivariate regression .

The model was performed using information criteria such as the Akaike Information Criterion (AIC) and simultaneous significance through Wilks' Lambda [17]. AIC is forionmulated as:

$$AIC = 2k - 2\ln(L) \tag{11}$$

with k = number of parameters, L = maximum likelihood of the model. Models with smaller AIC values are considered more efficient. The significance of the simultaneous model is assessed with Wilks' Lambda (Λ) in equation (6). A smaller Λ value indicates that the independent variables together explain the dependent variation better. The model selection process includes: constructing several candidate models, calculating AIC/BIC to compare efficiency, and testing simultaneous significance using Wilks' Lambda. The best model is one with a low AIC/BIC and a statistically significant Λ ($p\text{-value} < 0.05$), so it is efficient and can be interpreted reliably [17].

3. METHODOLOGY

3.1 Data Types and Sources

This study uses secondary data, that is, data obtained indirectly by the researcher through available official documents or publications. The secondary data in this study came from the Central Statistics Agency (BPS) of East Java Province, specifically the publication East Java in Figures 2025, which contains various social and economic indicators at the district/city level in East Java Province. The data used in this study is cross-sectional data, namely data collected over a certain period of time and representing the conditions of each district/city in East Java Province in 2024. The data is used to analyze variations in the Human Development Index (HDI) and the Gender Development Index (GDI) as well as the socio-economic factors that influence them.

3.2 Location and Time of Research

This research location covers all administrative districts/cities in East Java Province. This area was selected based on the availability of complete and consistent data across regions, as well as its relevance as one of the provinces with high levels of social and economic activity in Indonesia. The research period was carried out from the data collection stage to the preparation of the research report, namely in the period from December 2025 to January 2026. This time span includes data processing, statistical analysis, multivariate regression modeling, to interpretation of research results.

3.3 Research Variables

The variables used in this study consist of two dependent variables and five independent variables. Operational definitions, variable types, units, and descriptions of each variable are presented in **Table 1**.

3.4 Data Analysis Techniques

The data analysis stages in this study were conducted in accordance with the multivariate regression procedures discussed in the literature review. To clarify the sequence and interrelationships between the analysis stages, the entire research procedure is summarized in a flowchart presented in **Figure 1**.

Table 1. Operational Definition of Variables

Variables	Variable Types	Unit	Data Scale
Human Development Index (HDI)	Dependent (Y_1)	Index	Interval
Gender Development Index (GDI)	Dependent (Y_2)	Index	Interval
Population density	Independent (X_1)	People/km ²	Ratio
Open Unemployment Rate	Independent (X_2)	Percentage (%)	Ratio
Percentage of Poor Population	Independent (X_3)	Percentage (%)	Ratio
Average Length of Schooling	Independent (X_4)	Year	Ratio
Gross Regional Domestic Product (GRDP)	Independent (X_5)	Billion Rupiah	Ratio

4. RESULT & DISCUSSION

4.1 Data Description

The research data obtained from the East Java Province in Figures 2025 Publication is then described in general terms based on minimum, maximum, average and other characteristics as presented in the Table 2:

Table 2. Descriptive Analysis

	HDI	IPG	Population	Kep	OUR	PPM	MYS	GRDP
Min	66.72	82.67	411.0	1,560	3,060	5,080	8612	
1 st Qu	71.69	88.83	648.2	3,368	6,518	7,478	24653	
Median	74.56	91.64	862.5	4,075	9,215	8,060	44141	
Mean	75.31	91.57	1954.2	4,041	9,782	8,464	83516	
3 rd Qu	78.62	94.49	1214.0	4,735	12,043	9,832	101375	
Max	84.69	97.49	8698.0	6,490	20,830	12,110	772494	

Based on Table 2, the Human Development Index (HDI) of districts/cities in East Java Province in 2024 had an average value of 75.31 with a range of values from 66.72 to 84.69, which indicates that the level of human development in East Java is in the upper middle category but there are still disparities between regions. Also, the Gender Development Index (GDI) has an average value of 91.57 with a range of 82.67 to 97.49, indicating that gender equality is relatively good although not yet fully evenly distributed in every region.

Population density and Gross Regional Domestic Product (GRDP) show significant variation, reflecting differences in regional characteristics and economic activity between regions. These differences influence the level of development in each region. Meanwhile, the Open Unemployment Rate (OUR) and the Percentage of Poor Population (PPM) vary between districts/cities, indicating differences in employment conditions and community welfare, which can ultimately impact human

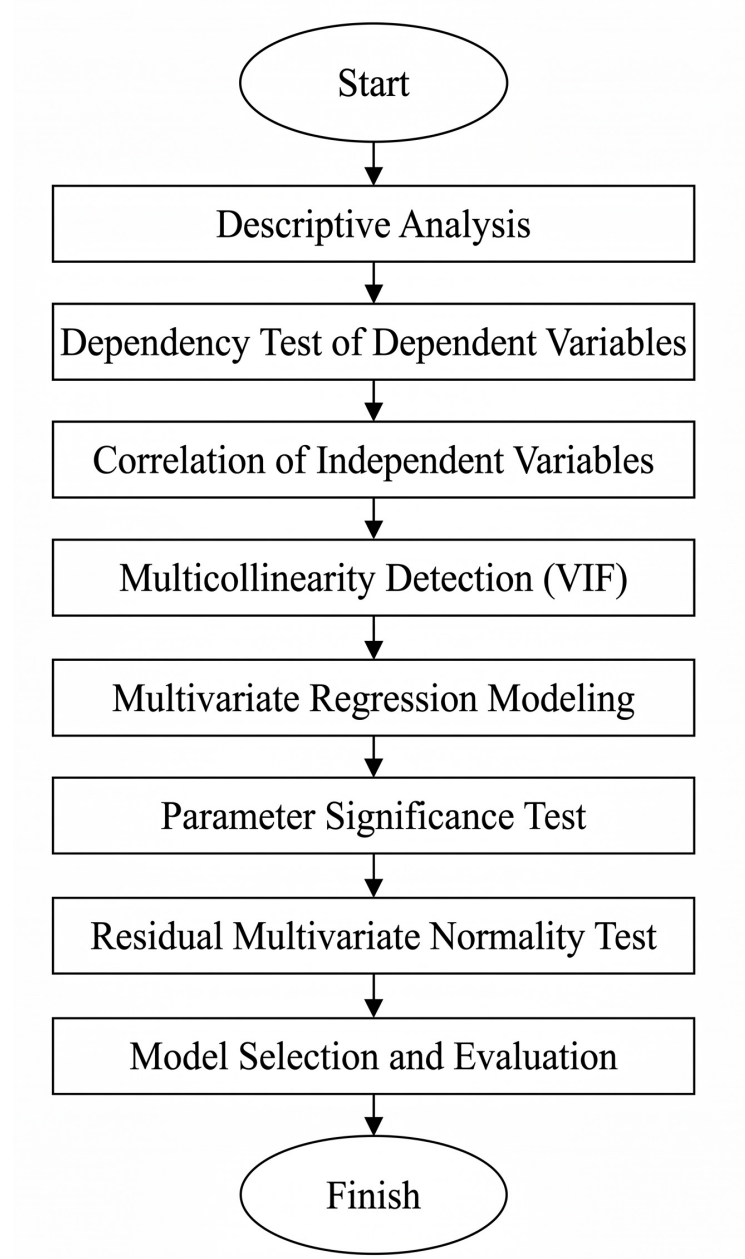


Figure 1. Flowchart of Data Analysis Stages

development.

Overall, these descriptive statistics indicate differences in social, economic, and human development conditions across districts/cities in East Java. Therefore, further analysis using multivariate regression is needed to examine the influence of each variable on the HDI.

4.2 The relationship between the Human Development Index (HDI) and the Gender Development Index (GDI)

As an initial stage of analysis, a Bartlett’s test of sphericity was conducted between the Human Development Index (HDI) and the Gender Development Index (GDI) to determine whether or not there was a significant relationship between the two response variables with the **Table 3**.

Based on Bartlett’s test, the Chi-square value was 27.43 with a degree of freedom (df) of 1 and a p-value of 1.63×10^{-7} ,

Table 3. Relationship between HDI and IPG

Statistics	Chi-square (χ^2)	Df	P-value
Mark	27.43	1	1.63×10^{-7}

which is less than 0.05. This indicates that the null hypothesis is rejected.

Thus, it can be concluded that there is a significant relationship between the Human Development Index (HDI) and the Gender Development Index (IPG) in East Java Province in 2024. This condition indicates that the two variables have an adequate relationship so that they meet the requirements for further analysis using the multivariate regression method.

4.3 Relationship Between Independent Variables

To determine the relationship between independent variables, a correlation analysis was performed, presented in the form of a corrrplot visualization. This makes it easier to identify the direction and strength of the relationship between the independent variables used in the study. The following visualization of the relationship between the independent variables is shown. **Figure**

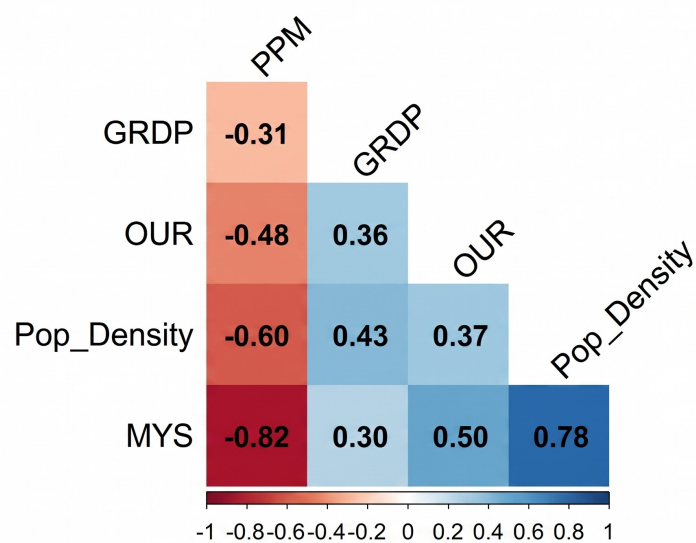


Figure 2. Visualization of Correlation Between Independent Variables Using Corrrplot

2, it can be seen that most of the independent variables have low to moderate correlations. The Mean Years of Schooling (MYS) variable shows a fairly strong positive correlation with population density, indicating that areas with higher population densities tend to have better mean years of schooling. Meanwhile, the Percentage of Poor Population (PPM) has a negative correlation with most other variables, especially with MYS, indicating that increasing education levels are associated with decreasing poverty levels.

4.4 Multicollinearity Detection (VIF)

Multicollinearity detection was performed using the Variance Inflation Factor (VIF) to ensure there was no strong linear relationship between independent variables before multivariate regression modeling was performed.

Based on **Table 4**, all independent variables have a VIF value of less than 10. The highest VIF value is found in the Mean Years of Schooling (MYS) variable of 5.486548, while other variables have VIF values below 5. This indicates that there is no multicollinearity problem in the model, so all independent variables are suitable for use in multivariate regression analysis.

Table 4. Variance Inflation Factor (VIF) Values of Independent Variables

Variables	VIF
Population Density	2.981636
OUR	1.467659
PPM	3.212921
MYS	5.486548
GRDP	1.368962

4.5 Multivariate Regression Model

4.5.1 Simultaneous Significance Test (Wilks' Lambda)

As an initial step in multivariate regression analysis, a simultaneous significance test using Wilks' Lambda was conducted to examine the effect of the independent variables simultaneously on the response variables, namely the Human Development Index (HDI) and the Gender Development Index (GDI). This test was used to determine whether the regression model had a significant overall effect.

Table 5. Results of Wilks' Lambda Significance Test

Wilks lambda	Approx F	df (num, den)	Pr (> F)
0.00011	145133	(2 , 31)	< 0.05

Based on the results of the Wilks' Lambda test, a p-value of < 0.05 was obtained, so it can be concluded that simultaneously the independent variables have a significant effect on the HDI and IPG.

4.5.2 Parameter Estimation and Model Equation

Based on the results of parameter estimation using a multivariate regression model, the model equation for each response variable is obtained as follows:

1. HDI Model $HDI = 54.75 + 0.00015PopulationDensity + 0.06636OUR - -0.04410PPM + 2.382MYS + 0.000003GRDP$
2. IPG Model $IPG = 83.41 + 0.0002899PopulationDensity + 0.8788OUR$

Based on these equations, the regression coefficient indicates the direction of influence of each independent variable on the response variable. A positive coefficient indicates a unidirectional relationship, while a negative coefficient indicates an inverse relationship. In the HDI model, the Percentage of Poor Population (PPM) variable has a negative coefficient. Meanwhile, in the GDI model, the PPM and Gross Regional Domestic Product (GRDP) variables also have negative coefficients.

4.5.3 Hypothesis Testing (Partial Test)

Based on **Table 6**, the variables Mean Years of Schooling (MYS) and Gross Regional Domestic Product (GRDP) have positive regression coefficients with relatively low significance levels. Meanwhile, the variables population density, open unemployment rate (OUR), and percentage of poor people (PPM) show significance values greater than the 5% level. These results will then be evaluated in the model selection process to obtain the best multivariate regression model.

Based on **Table 7**, the Open Unemployment Rate (OUR) variable has a positive regression coefficient with a relatively smaller significance level compared to other independent variables. Meanwhile, variables such as population density, percentage of poor people, mean years of schooling, and gross regional domestic product show a significance level greater than the 5% level. In addition, the gross regional domestic product variable shows a significance level of 10%. Therefore, its role in the model will be further considered at the model selection stage. It will then be used as a basis for the evaluation and selection of multivariate regression models to obtain the best model for the response variable of the Gender Development Index (IPG).

Table 6. HDI Regression Model

	<i>Estimate</i>	<i>Std. Error</i>	<i>t- value</i>	<i>Pr (> t)</i>
<i>Intercept</i>	54.75	2.15	25.471	$2e-16$ ***
X1	0.00015	0.00012	1,314	0.1981
X2	0.06636	0.1567	0.424	0.6745
X3	−0.04410	0.06470	−0.682	0.5004
X4	2.382	0.2092	11.386	$8.63e-13$ ***
X5	0.000003	0.000001	2.424	0.0212 *

Table 7. IPG Regression Models

	<i>Estimate</i>	<i>Std. Error</i>	<i>t- value</i>	<i>Pr (> t)</i>
<i>Intercept</i>	83.41	5.312	15,703	$< 2e-16$ ***
X1	0.0002899	0.0002846	1,019	0.3161
X2	0.8788	0.3869	2,271	0.0300 *
X3	−0.1832	0.1599	−1.146	0.2604
X4	0.7524	0.5169	1.456	0.1552
X5	−0.0000064	0.0000034	−1.921	0.0636

4.6 Best Model

In multivariate regression analysis, the selection of the best model is carried out to obtain the most optimal model in explaining the relationship between the independent variables and the response variable. Model selection was conducted using a stepwise method with the Akaike Information Criterion (AIC), where the model with the smallest AIC value was selected as the best model. The initial model used included all independent variables, namely Population Density, Open Unemployment Rate (OUR), Percentage of Poor Population (PPM), Mean Years of Schooling (MYS), and Gross Regional Domestic Product (GRDP).

4.6.1 Comparison of HDI Models

Table 8. Comparison of HDI Models

<i>Model</i>	<i>Variables</i>	<i>AIC</i>
Model awal	Population, OUR, PPM, MYS, GRDP	109.39
Stepwise model	MYS, GRDP	105.78

Based on **Table 8**, the resulting stepwise model has a lower AIC value than the initial model. Therefore, the model with the MYS and GRDP variables was selected as the best model for the HDI response variable.

4.6.2 Best Regression Model of HDI

Based on the parameter estimation results, the following regression model was obtained:

$HDI = 52.68 + 2.633MYS + 0.000004102GRDP$

Table 9. Best HDI Regression Models

<i>variables</i>	<i>Estimate</i>	<i>Std.Error</i>	<i>t-value</i>	<i>Pr (> t)</i>
Intercept	52.68	0.772	68,229	< 0.001
MYS	2.633	0.092	28,501	< 0.001
PDRB	0.000004102	0.000001200	3,418	0.00162

The variables Mean Years of Schooling (MYS) and Gross Regional Domestic Product (GRDP) have a positive and significant effect on the Human Development Index (HDI). This indicates that increasing levels of education and regional economic activity tend to increase the Human Development Index (HDI).

4.6.3 Comparison of IPG Models

Table 10. Comparison Table of IPG Models

<i>Model</i>	<i>Variables</i>	<i>AIC</i>
Initial Model	Population, OUR, PPM, MYS, GRDP	178.155
Stepwise model	MYS, OUR, GRDP	176.5024

Based on Table 10, the resulting stepwise model has a lower AIC value than the initial model. Therefore, the model with the MYS, OUR, and GRDP variables was selected as the best model for the IPG response variable.

4.6.4 Best Regression Model of IPG

Table 11. Best IPG Regression Models

<i>variables</i>	<i>Estimate</i>	<i>Std.Error</i>	<i>t-value</i>	<i>Pr (> t)</i>
Intercept	76.65	2.002	38,280	< 0.001
MYS	1.391	0.579	5,394	< 0.001
OUR	0.8779	0.3819	2,299	0.0278
GRDP	−0.00000492	0.00000311	−1,582	0.1228

Based on the parameter estimation results, the following regression model was obtained:

$GPI = 76.65 + 1.391MYS + 0.8779OUR - 0.00000492GRDP$

The variables Mean Years of Schooling (MYS) and the Open Unemployment Rate (OUR) have a positive and significant effect on the GDI. This indicates that increasing education levels and labor market dynamics contribute to improving gender development. Meanwhile, Gross Regional Domestic Product (GRDP) has a negative coefficient and does not significantly influence the GDI.

4.7 Multivariate Residual Normality Test

A multivariate residual normality test was conducted to assess the suitability of the residual distribution to the model assumptions. The test results are presented in the Table 12.

Based on the results in Table 12, the chi-square distribution quantile value is obtained 1.386294 and the proportion of Mahalanobis distances below this value is 0.4473684. This proportion value is close to 0.5, so it can be concluded that the residuals of the multivariate regression model tend to follow a multivariate normal distribution.

Table 12. Results of Multivariate Residual Normality Test

Criteria	Mark
Chi-square Quantile	1.386294
Mahalanobis Proportion	0.4473684

Based on the results of the multivariate residual normality test presented in the table above, a visual inspection was performed using the Multivariate Q–Q Plot to see the suitability of the residual Mahalanobis distance distribution to the chi-square distribution. The following is a visualization of the Multivariate Q–Q Plot.

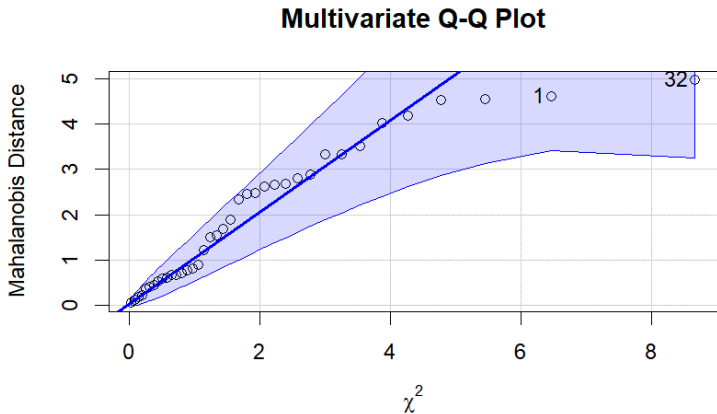


Figure 3. Q-Q Plot of Multivariate Residuals of Regression Model

Based on the Multivariate Q–Q Plot in **Figure 3**, it can be seen that most of the observation points are located around the diagonal line and are still within the confidence envelope. This indicates that the residual Mahalanobis distance has a pattern that is relatively consistent with the theoretical chi-square distribution. Although there are several points in the upper right that deviate slightly from the diagonal line, these deviations are still within reasonable limits and do not indicate the presence of extreme outliers (multivariate outliers). It can be concluded that the residuals of the multivariate regression model tend to be multivariately normally distributed, so that the assumption of multivariate normality of the residuals is met.

4.8 Comparison of Initial Model and Best Model

A comparison test between the initial model and the best model was conducted to evaluate whether the model simplification using the stepwise method maintained the model’s ability to explain variations in the HDI and GPI response variables. The test used was a multivariate ANOVA with the Pillai’s Trace approach.

Table 13. Comparison Table of Initial Model and Best Model

Model	Res.df	df	Gen.Var	Pillai	Approx F	Df (num, den)	P-value
Beginning	32	-	2.1116	-	-	-	-
Best	34	2	2.1029	0.10741	0.90806	(4, 64)	0.4647

The comparison results show that the Pillai value = 0.10741 with $F = 0.908$ and p-value = 0.465, so there is no significant difference between the two models. This indicates that simplifying the model to only the MYS, OUR, and GRDP variables does not reduce the model’s ability to explain variations in the two response variables, HDI and GDI. Thus, the best model can be used because it is more efficient and focuses on the predictors that have a dominant influence, while additional variables from the initial model do not provide a significant contribution to the model.

5. CONCLUSION

Based on the analysis, it can be concluded that human development and gender development in the districts/cities of East Java Province are influenced by different socioeconomic conditions. The results show that education, as reflected in the Mean Years of Schooling (MYS), is the most consistent factor influencing the Human Development Index (HDI) and the Gender Development Index (GDI). In addition, Gross Regional Domestic Product (GRDP) plays a role in increasing the HDI because it reflects the economic capacity of the region, while the Open Unemployment Rate (OUR) is related to the GDI, indicating that labor market conditions also influence the achievement of gender development in East Java.

The results of the inter-variable relationship test indicate that the Human Development Index (HDI) and the Gender Perception Index (GDI) have a directional and significant relationship. Regions with higher levels of human development generally also exhibit higher gender development outcomes. This finding confirms that efforts to improve people's quality of life not only impact general well-being but are also closely linked to increasing equality between men and women in development. Judging from the capabilities of the method used, the multivariate regression model proved quite effective in explaining the relationship between socioeconomic variables and the Human Development Index (HDI) and the Gender Perception Index (GDI). The variables used in the model were able to explain most of the variation in the HDI and GDI between regions and showed a significant simultaneous influence. Thus, the multivariate regression approach can be used as an appropriate analytical tool for simultaneously analyzing human development and gender development in East Java province.

REFERENCES

- [1] M. Tebiary and A. M. Balami, "Pemodelan Regresi Multivariat Pada Kasus Indeks Pembangunan Manusia dan Indeks Pembangunan Gender," *Parameter: J. Mat, Stat, & Ter.*, vol. 3, no. 01, pp. 61–70, Mar. 2024. [Online]. Available: <https://ojs3.unpatti.ac.id/index.php/parameter/article/view/12729>
- [2] D. I. Ramadhany, A. J. Pradani, and S. P. Wulandari, "Human development index using principal component analysis in east java in 2022," *Scientist: Journal of Natural Science*, vol. 2, no. 1, pp. 8–17, 2024.
- [3] N. Ramadhani and Y. P. Utomo, "Analisis faktor-faktor yang mempengaruhi indeks pembangunan manusia (ipm) di jawa timur tahun 2010-2020," *Jurnal Bisnis dan Manajemen*, vol. 3, no. 2, 2023.
- [4] United Nations Development Programme, "Human development report 2023/2024: Reimagining cooperation in a polarized world," United Nations Development Programme, Tech. Rep., 2024.
- [5] BPS Statistics, "Gender inequality index," Badan Pusat Statistik, 2024.
- [6] Q. Talafha, A. Bashayreh, S. Abdelhadi, and A. Y. Alsamman, "Consequences of gender inequality on economic development: an empirical approach for oecd countries," *Discover Sustainability*, 2025.
- [7] F. R. Faqih Rifaldy, M. N. A. S. M N Alfi Syahrin, M. I. P. M Irsan Prayoga, F. S. Fajar Syakbani, and M. F. Mhd. Furqan, "Visualisasi dan Statistik Deskriptif Beserta Analisis Dataset Status Gizi Balita Tahun 2022 Per Provinsi di Indonesia Menggunakan Pemrograman R," *JITET*, vol. 13, no. 3, Jul. 2025. [Online]. Available: <https://journal.eng.unila.ac.id/index.php/jitet/article/view/7025>
- [8] U. Sa'adah, B. C. Putra, and R. I. Permatasari, "Job satisfaction: The role of organizational culture as a regression variable on employees at pppptk boe malang," *NNOVATIVE: Journal Of Social Science Research*, vol. 5, pp. 3102–3116, 2025.
- [9] N. H. Chairunnisa and D. A. Nohe, "Overcoming multicollinearity in multiple linear regression using principal component analysis," *Exponential*, vol. 16, no. 1, pp. 1–9, 2025.
- [10] S. H. P. Ningrum, K. Hisan, T. P. Ramdhani, L. Luzianawati, M. D. R. Zindawi, and L. Harsyiah, "Regresi Komponen Utama dalam Mengatasi Multikolinieritas pada Faktor-Faktor yang Mempengaruhi Inflasi di Indonesia," *IJASDS*, vol. 2, no. 1, pp. 34–43, May 2025. [Online]. Available: <https://journal.unram.ac.id/index.php/ijasds/article/view/5827>

- [11] M. S. Pradana, N. Nensi, and N. E. Chandra, “Analisis Regresi Multivariate Berdasarkan Latar Belakang Santri,” *JMAT*, vol. 12, no. 1, p. 56, Mar. 2023. [Online]. Available: <https://ojs.unud.ac.id/index.php/jmat/article/view/76444>
- [12] D. Yulianto and E. A. Juniawan, “Integrating the scientific approach in stem-based learning: Advancing critical and creative mathematical thinking skills for future-oriented education,” *International Journal of Trends in Mathematics Education Research*, vol. 9, no. 2, pp. 1–13, 2025.
- [13] P.M. Parwatha, I.K. Sudarma, and I.W. Sukra Warpala, “Pengaruh Model Project Based Learning Berbantuan Liveworksheet Terhadap Kemampuan Berpikir Kreatif dan Hasil Belajar IPA,” *jurnal_tp*, vol. 15, no. 1, pp. 1–13, Apr. 2025. [Online]. Available: https://ejournal2.undiksha.ac.id/index.php/jurnal_tp/article/view/4170
- [14] B. G. Tabachnick and L. S. Fidell, “Multivariate Analysis of Variance (MANOVA),” in *International Encyclopedia of Statistical Science*, M. Lovric, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2025, pp. 1616–1619. [Online]. Available: https://doi.org/10.1007/978-3-662-69359-9_398
- [15] I. Sandamali, “Evaluation of Manova Test Statistics for Increasing Number of Group,” Ph.D. dissertation, Uppsala Universitet, 2025.
- [16] S. Wang, J. Liang, M. Zhou, and H. Ye, “Testing Multivariate Normality Based on F-Representative Points,” *Mathematics*, vol. 10, no. 22, p. 4300, Nov. 2022. [Online]. Available: <https://www.mdpi.com/2227-7390/10/22/4300>
- [17] W. Riansut, “A Study of the Effectiveness of Model Selection Criteria for Multiple Regression Model,” *Recent Science and Technology*, vol. 15, no. 1, pp. 198–212, Apr. 2023. [Online]. Available: <https://li01.tci-thaijo.org/index.php/rmutsvrj/article/view/249278>