East Asian Journal of Multidisciplinary Research (EAJMR)
Vol. 4, No. 9, 2025: 4603 - 4614

Implementation of Classical Assumption Test in Students' Decision Research in Choosing a College in Jambi City

Effiyaldi[1]*, Yossinomita[2], Eddy Suratno[3], Johni Paul Karolus Pasaribu[4] , Ronald Naibaho[5], Irfan Hassandi[6]
Universitas Dinamika Bangsa
Corresponding Author: Effiyaldi yldbf67@gmail.com

ARTICLEINFO

Keywords: Classical, Assumption, Test, Students, Decision

Received : 13, August
Revised : 14, September
Accepted: 24, September

ABSTRACT

In many research cases, a common problem is that the data obtained from the field, especially primary data obtained from respondents, is not normal. This study aims to measure and analyze data to ensure that the data used in a study produces normal data. Using the classical assumption test analysis method, the results indicate that the data used in this study is normal. The results of this study are expected to increase knowledge, especially for future researchers, regarding classical assumption tests, resulting in higher-quality research results and minimizing bias.

DOI prefix: https://doi.org/10.55927/eajmr.v4i9.399
ISSN-E: 2828 – 1519 (Online)
https://mtiformosapublisher.org/index.php/eajmr

INTRODUCTION

In research activities, especially quantitative research, the data and/or variables used must be truly sound and not related to unmet assumptions. Therefore, classical assumption tests are used to ensure that the data and/or variables used are not related to unmet assumptions.

Not all analyses can be successfully used in regression analysis. To make forecasts, we will use a regression model; a good model has the minimum possible forecast error. Therefore, the model must meet the assumptions, or assumptions, before being used.

This classical assumption test is conducted before conducting further analysis of the collected data. The purpose of this classical assumption test is to produce a regression model that meets the Blue criteria. These criteria indicate that the regression model is efficient, unbiased, consistent, and normally distributed. The best unbiased linear estimator can be used as a reliable and reliable estimator. Tests for normality, linearity, multicollinearity, autocorrelation, homoscedasticity, and autocorrelation are classical assumptions that must be met (Hamsar, 2023) (Maulid, 2022).

It is crucial for researchers to use the classical assumption test, especially those conducting quantitative research. Failure to do so will result in biased research results. Biased research results will not reflect the facts on the ground. Therefore, the research cannot address the underlying problem.

In the field, many researchers underestimate the results of this classical assumption test. Novice researchers are particularly concerned about selecting appropriate test methods for their data and/or statistical variables. Researchers often attempt to manipulate the data; the data is arranged in such a way that the statistical method used by the researcher can be applied (Maulid, 2022). Two important problems in regression analysis are multicollinearity and violation of the linearity assumption. These two problems can significantly alter the interpretation, predictive reliability, and accuracy of the model. When independent variables are highly correlated, multicollinearity occurs, leading to isolated variances and unstable coefficient estimates.(Imam Ghozali, 2005)

Similarly, inconsistencies between predictors and dependent variables lead to inaccurate estimates and reduce model validity. Both of these issues are typically caused by poor research design, data issues, or errors in model specification. Addressing these issues is crucial to ensure accuracy (Naufal et al., 2025) The result is that the resulting research results are biased, or in other words, the research results do not reflect the actual phenomenon and cannot answer the research problem. This renders the research conducted futile.

The study began in 2024. This research examines how classical assumption tests can generate relevant data and/or variables for research on student college choice in Jambi City.

THEORETICAL REVIEW

Normality Test

The normality test is used to predict errors (residuals), which are the differences between actual data and forecasted data. According to (Singgih Santoso, 2011), the normality test aims to determine whether the confounding variables or residuals have a normal distribution in a regression model. The current data should be normally distributed.

Two different methods can be used to determine whether the residuals have a normal distribution. The first is histogram analysis, also known as the Kolmogorov-Smirnov test (Agus Widarjono, 2010). This analysis compares observed data with a distribution comparable to a normal distribution. However, if the data used is small or limited, this method can mislead researchers. To get better results, use a normal probability plot. A normal distribution will form a straight diagonal line, and the residual data plot will be compared to a diagonal line. In cases where the residual data distribution is normal, the line representing the actual data will follow the diagonal line. 2. Statistical analysis, also known as the Jarque beta test, can be done by examining the kurtosis and skewness values of the residuals.

Heteroscedasticity Test

The heteroscedasticity test is typically used on cross-section data containing heteroscedasticity because this data collects various sizes, including small, medium, and large. The purpose of this test is to determine whether there is inequality in the regression model between residuals from one observation to another.

Situational changes not explained in the model specification cause heteroscedasticity (Arif Pratisto, 2004). Variation of disturbance variables that are not constant is called heteroscedasticity. Constant variation is expected from the remaining residuals. Heteroscedasticity occurs when the residual variance increases or decreases in a certain pattern.

There are several ways to determine the presence or absence of heteroscedasticity: 1. Look at the plot between the predicted value of the dependent variable, namely ZPRED, and its residual SRESID. Certain patterns in the scatterplot graph between SRESID and ZPRED are used to determine the presence or absence of heteroscedasticity. 3. The Glejser test proposes a method to regress the squared residual value against the independent variable (Gujarati, 2003). 4. The White test proposes a method to regress the squared residual with the squared variable, the independent variable, and the multiplication of the independent variable.

Multicollinearity Test

The multicollinearity test aims to identify the relationship or correlation between independent variables in multiple regression. It is possible that the independent variables have no correlation in a good regression model. Independent variables whose correlation value is equal to zero are called organic variables.

To determine whether or not there is multicollinearity in the regression model, it is as follows: first, the R square value produced by the empirical regression model estimation is very high; however, in particular, the influence of the independent variables on the dependent variable is very small. Second is to analyze the correlation matrix of each independent variable. There is a problem of multicollinearity if there is a fairly high correlation between the independent variables; this is usually more than 0.90. 3. Multicollinearity can be seen from two perspectives: a. Tolerance Value and b. Value against Tolerance. A good tolerance value here is <0.10. b. A good Variance Inflation (VIF) value here is >10, which means a low tolerance value is the same as a high VIF value (a tolerance value less than 0.10 is the same as a VIF value more than 10).

Linearity Test

The linearity test is used to evaluate the accuracy of the model specifications used . As states, the linearity test also seeks the appropriate model, whether it is linear, quadratic, or cubic. As explained, linearity refers to a positive or negative relationship.

There are several methods for conducting a linearity test. The first is the Durbin-Watson test, which is typically used to determine whether there is autocorrelation in a regression model. The second is the Ramsey test, which requires the assumption or belief that the correct function is linear. 3. Multiple Lagrange: This aims to obtain the arithmetic C-squared value or (n x R-squared).

METHODOLOGY

This study used data from 608 students from various universities in Jambi City. Data were collected randomly, with 608 respondents serving as a sample. Sampling was conducted using a snowball sampling method. Data were collected through a specially designed questionnaire using Google Forms, aligned with the research variables and indicators, and a Likert scale. The data were processed using SPSS.

Four variables, each independent and dependent, were used in this study. These variables are Tuition Fees (X1), Educational Facilities (X2), College Status (X3), and College Selection Decision (Y). Ten indicators were used to construct the questionnaire items.

RESULTS

Normality tests can be seen from histograms, plots, and Kolmogorov-Smirnov tests. The following are the test results by looking at the histogram.

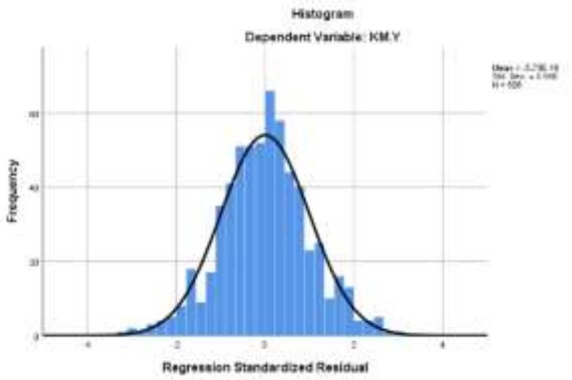

Figure 1. Histogram Dependent Variable: Keputusan Memilih (Y)

The figure shows a histogram of standardized residuals from a regression with the dependent variable KM.Y. This histogram is used to check the assumption of normality of the residuals, which is important for linear regression analysis. Several elements comprise the figure. The X-axis above shows the residuals (errors) from the normalized regression model, and the Y-axis below shows the frequency of those residuals. For the curved black line: The theoretical normal distribution, also known as the normal curve, serves as a comparison. An additional factor is the mean = -5.73E-16 ≈ 0, which indicates that the average of the residuals is very close to zero, as expected. While the standard deviation of the residuals is 0.998, which is the same as the normalized standard, and close to 1. And the number of observations in the regression model is N = 608.

The second way to see the normality of data is to look at the Plot. The following is a plot image of the day of data normality test results.

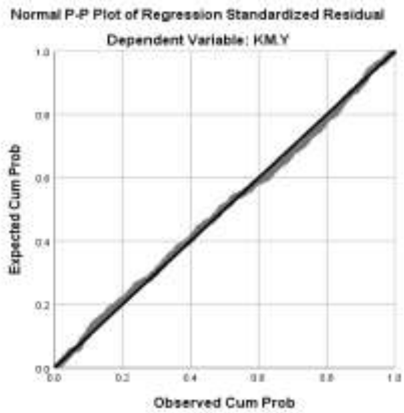

Figure 2. Normal Plot Dependent Variable: Keputusan Memilih (Y)

To test the assumption of normality of residuals in linear regression analysis, the dependent variable KM.Y is used as a normal P-P plot of standardized residual regression. In the figure, there are three axes. The X-axis (found along with probability) shows the actual cumulative probability of the standardized residuals, the Y-axis (expected along with probability) shows the expected cumulative probability if the residuals follow a normal distribution, the diagonal line is the ideal line where all points would be if the residuals were truly normally distributed, and the points are representations of the actual data (standardized residuals).

The next method for testing data is using the Kolmogorov-Smirnov test. The following image shows the results of the data normality test using the Kolmogorov-Smirnov test.

Table 1. One-Sample Kolmogorov-Smirnov Test

One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		608
Normal Parameters[a,b]	Mean	.0000000
	Std. Deviation	4.76794997
Most Extreme Differences	Absolute	.027
	Positive	.027
	Negative	-.023
Test Statistic		.027
Asymp. Sig. (2-tailed)		.200[c,d]

a. Test distribution is Normal.
b. Calculated from data.
c. Lilliefors Significance Correction.
d. This is a lower bound of the true significance.

According to the One-Sample Kolmogorov-Smirnov Test Output Table, there are several elements, as follows: N (number of samples) 608, residual mean 0.0000000, residual standard deviation 4.7679, most significant extreme difference in absolute terms 0.027, positive difference 0.027, negative difference -0.023, test statistic (K-S) 0.027, and sigma asymmetry (2-tailed) 0.200.
The explanation for the multicollinearity test is as follows:

Table 2. Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	10.555	1.614		6.539	.000		
	BP.X1	-5.537E-5	.000	-.026	-.738	.461	.990	1.010
	FP.X2	.326	.031	.371	10.373	.000	.974	1.027
	SPT.X3	.272	.034	.281	7.888	.000	.984	1.017

a. Dependent Variable: KM.Y

The results of the multicollinearity test shown in the coefficients table show that all variables have a tolerance value greater than 0.10 and a VIF value lower than 10, so it can be concluded that, although this regression model contains a moderate level of multicollinearity, in general this regression model can be accepted without multicollinearity problems. problems especially between FP.X2 and SPT.X3 because multicollinearity can cause unstable regression coefficients as well as inaccurate t-test significance values (significant variables may appear insignificant).

Thus, the regression model used is feasible and the estimation results obtained are reliable. Overall, this study proves that the variables FP.X2 and SPT.X3 have a positive and significant effect on KM.Y, while BP.X1 does not have a significant effect. that the value of "linearity statistics" shows a VIF value of less than 10 and a tolerance greater than 0.1. It can be concluded that there is no multicollinearity between the VIF values of the independent variables (Alita et al., 2021). It is different if the VIF value is high. By looking at the high VIF value, we can conclude that the variables in the regression model are strongly correlated with each other, which can interfere with stable and reliable estimation (Yhoga Hendrianto et al., 2023).

The third data test uses a scatterplot. The following is an image of a scatterplot.Untuk pengujian data yang ketiga adalah menggunakan Scaterplot. Berikut adalah gambar Scater Plot.

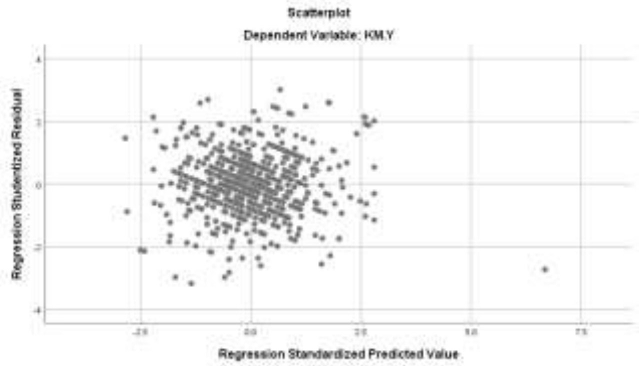

Figure 3. Scatterplot Regression Standardized Predicted Value dan Regression Studentized Residual

To evaluate two important assumptions in classical linear regression, a scatterplot between the standardized estimated values and the student's estimated values was used for the dependent variable KM.Y.
The linearity test can be explained as follows:

Table 3. ANOVA

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
KM.Y FP.X2	*Between Groups	(Combined)	4189.865	27	155.180	6.375	.000
		Linearity	3074.931	1	3074.931	126.330	.000
		Deviation from Linearity	1114.934	26	42.882	1.762	.012
	Within Groups		14117.503	580	24.341		
	Total		18307.368	607			

Source: Processed Data

The results of the linearity test using the ANOVA Table between the variables KM.Y and FP.X2 show a significance value of the linearity component of 0.000 (<0.05), with an F value of 126.330. This indicates that the two variables have a significant linear relationship, which indicates that the linearity assumption of the regression analysis is met. Therefore, to analyze the effect of the FP.X2 variable on KM.Y, a linear regression model can be used. By doing this step, the regression model will be able to provide a more accurate and representative estimate of the relationship between the variables studied. However, the significance value found in the deviation component from the straight line is 0.012 (<0.05), and the F value is 1.762.

DISCUSSION

As a type of continuous random variable distribution, the normal distribution has a bell-shaped curve or graph (Fiska R, 2025), as shown in Figure 1 above. This histogram is symmetrical and resembles a normal curve, also known as the black line. There is no significant skewness or outliers. This graph indicates that the assumption of normality is met, as the residuals are normally distributed. Regression results can be considered valid and reliable for inference (such as t-tests and F-tests) because the residuals follow a normal distribution.

Regression analysis was conducted in this study to determine the effect of the independent variables on the dependent variable KM.Y. One important assumption in classical linear regression is that the residuals—the difference between actual and predicted values—must be normally distributed. A standardized residual histogram is used to evaluate this assumption.

With a symmetrical bell-shaped pattern—indicating a normal curve—and most of the residual values concentrated around zero, the histogram shown shows that the residual distribution has a shape similar to a normal distribution. This indicates that there is no significant deviation from the standard. There is growing evidence that regression models can be used to predict independent variables and vice versa (Mardiatmoko, 2024). As evidence that there is no systematic bias in the regression model predictions, the descriptive statistics of the residuals show a mean value very close to zero. Furthermore, the standard deviation is very close to one, which is characteristic of well-normalized residuals. The large sample size indicates that the residuals follow a normal distribution.

Once the residual normality assumption is met, the regression model used in this study can be deemed suitable for use as a basis for statistical decision-making. The overall model testing process and the significance of the regression coefficients are required. Furthermore, this normality assumption ensures that the model parameter estimation results are efficient and unbiased. The residual histogram shows that the residual distribution of this regression model resembles a normal distribution, with the mean and standard deviation values consistent with expectations. Therefore, the residual normality assumption has been met. This indicates that the results of the regression analysis on the dependent variable KM.Y are valid.

Furthermore, in Figure 2, the plot points are located near the diagonal line. The line does not appear to curve significantly upward or downward. This pattern indicates that the residuals are normally distributed, or at least closely approximate a normal distribution, in the regression model. The normal P-P plot confirms the previous histogram results, indicating that the normality assumption has been met.

In classical linear regression analysis, testing the normality assumption of residuals is crucial. To test this assumption, a normal probability-probability plot (P-P plot) is used. This plot compares the observed cumulative probabilities with those expected from a normal distribution.

Because the residual points are closer to the diagonal line from the bottom left to the top right, the standardized residual distribution is closer to a normal distribution, as indicated by the normal P-P plot (Figure 2). There is no violation of the normality assumption, as there is no curvature or significant deviation from the diagonal line upward or downward. Therefore, the findings of this analysis support the previous residual histogram results, which showed a residual distribution pattern similar to a normal curve.

By meeting this residual normality assumption, the interpretation of regression coefficients, significance values, and hypothesis testing can be performed more accurately and reliably. This is why this assumption is important. The residual normality assumption has been met, and the normal P-P plot indicates that the residuals in the regression model for the dependent variable KM.Y are normally distributed. Therefore, the regression model used can be considered appropriate and valid for further analysis. The Kolmogorov-Smirnov test, or K-S test, is used to determine whether the residual distribution follows a normal distribution. The null hypothesis, or H0, results in normally distributed residual data, and the alternative hypothesis, or H1, results in non-normally distributed residual data.

Furthermore, a larger significance value indicates that H0 is accepted, meaning the residuals do not deviate significantly from the normal distribution. The test statistic value indicates how far the residual distribution deviates from the normal distribution. There is no significant deviation, as the value is small.

The significance value is obtained based on the results of the one-sample Kolmogorov-Smirnov test. Therefore, we can conclude that the residuals from the regression model are normally distributed. This result supports the validity of the regression model and reinforces previous findings from the normal P-P plot and residual histogram. This aligns with the findings of (Arnas & Harsono, n.d.); the normality test aims to determine whether the dependent and independent variables in a regression model have a normal or near-normal distribution. A good regression model is one with a normal or near-normal data distribution. The measuring instrument method is used to calculate the residuals of the dependent variable to confirm the assumption that the equation is normally distributed. In this study, data normality was tested using the Kolmogorov-Smirnov test by comparing the significance value with 0.050.

The scatterplot above shows that the points are randomly distributed around the zero horizontal line. There is no specific pattern, such as a curve or a V-shaped or butterfly-shaped pattern, which typically indicates a violation of the linearity or heteroscedasticity assumptions. Furthermore, the points are fairly evenly distributed from left to right, with consistent density.

The heteroscedasticity test is used to determine whether the variances in a regression model are proportional or unequal across the data. (Chaerul Rizky et al., n.d.) used a scatterplot to test for heteroscedasticity. The scatterplot between the standardized predicted values and the studentized residuals likely meets the heteroscedasticity and linearity assumptions. This indicates that the linear regression model used in this study is still valid. This is consistent with the results of (Ainiyah et al., 2016), who stated that the value of the test results can be seen from their significance values. Heteroscedasticity was not found in cases where the significance value was greater than 0.05.

Conversely, heteroscedasticity occurs if the significance value is <0.05. To test linearity, the Sig. linearity and Sig. linearity deviation values found in the ANOVA table can be used. The ANOVA results indicate a significant deviation from a linear relationship. In other words, due to the tendency for non-linear patterns in the data, the resulting relationship pattern is not always linear, even though the primary relationship between the variables is linear. According to the findings of (Naufal et al., 2025), when the relationship between the independent and dependent variables is non-linear, a violation of the linearity assumption occurs. Many studies use current theory to assume this relationship is linear. However, in practice, this is not always the case. The constructed regression model will produce inaccurate and misleading results if the relationship is non-linear. Bias in model specification is one type of violation that can occur because some important variables are not included. This situation occurs when the analysis ignores variables that should be present, which can result in incorrect or misleading conclusions and errors.

Overall, these results suggest that fixed linear regression can still be used as the basis for analysis; However, the results should be interpreted carefully. Linearity deviations can come from additional components that have not been included in the research model. As a result, researchers may consider adding control variables or using data transformations to refine the model. In this way, the regression model will be able to provide a more accurate and representative estimate of the relationship between the variables studied.

CONCLUSIONS AND RECOMMENDATIONS

Based on the discussion above, the following conclusions can be drawn:

1. The data presented in this study meets the assumption of normality and is normally distributed.
2. The distribution of points is fairly even from left to right, with consistent density, indicating no serious heteroscedasticity issues.
3. The assumptions of heteroscedasticity and linearity are met.
4. Although the primary relationship between variables is linear, the resulting relationship pattern is not perfectly linear, as there are still non-linear patterns in the data.

FURTHER STUDY

This study used a fairly large sample, but the results of the analysis showed that the relationship formed was not completely linear, because there was still a tendency for non-linear patterns in the data.

ACKNOWLEDGMENT

This research was completed thanks to the assistance of many parties. Therefore, I would like to express my gratitude to:

1. The Rectorate of Universitas Dinamika Bangsa for facilitating this research.
2. My wife, children, and extended family for their support and encouragement throughout this research.
3. My colleagues and colleagues who have encouraged me throughout this research.

REFERENCES

Agus Widarjono. (2010). *Analisis Statistika Multivariat Terapan*.

Ainiyah, N., Deliar, A., & Virtriana, R. (2016). The classical assumption test to driving factors of land cover change in the development region of northern part of west Java. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, *41*, 205–210. https://doi.org/10.5194/isprsarchives-XLI-B6-205-2016

Alita, D., Putra, A. D., & Darwis, D. (2021). Analysis of classic assumption test and multiple linear regression coefficient test for employee structural office recommendation. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(3), 295. https://doi.org/10.22146/ijccs.65586

Arif Pratisto. (2004). *Cara Mudah Mengatasi Statistik Dan Rancangan Percobaan Dengan SPSS 12.*

Arnas, Y., & Harsono, Y. (n.d.). The Effect Of Work Discipline And Work Environment On Employee Performance At Pt Axia Multi Sarana Kota Jakarta Selatan. In *International Journal of Economics,businessand Accounting Research (IJEBAR) Peer Reviewed-International Journal* (Vol. 7). https://jurnal.stie-aas.ac.id/index.php/IJEBAR

Chaerul Rizky, M., Ardian, N., & Sirait, E. (n.d.). *Analysis Of The Impact Of Training And Development On Increasing Community Capability Village In Human Resource Management In Kwala Serapuh Village, Langkat District*.

Fiska R. (2025). *Distribusi Normal: Pengertian, Parameter, Karakteristik dan Aturan Empiris.*

Gujarati, D. (2003). *Ekonometri Dasar Terjemahan: Sumarno Zain, Jakarta, Erlangga*.

Hamsar, I. A. (2023). *Asumsi Regresi Linear Klasik*.

Imam Ghozali. (2005). *Aplikasi Anaisis Multivariat Dengan SPSS*.

Mardiatmoko, G. (2024). The Application of the Classical Assumption Test in Multiple Linear Regression Analysis (a Case Study of the Preparation of the Allometric Equations of Young Makila). *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *8*(3), 724. https://doi.org/10.31764/jtam.v8i3.22179

Naufal, M. J., Perwira Ompusunggu, D., Sinaga, R. A., Dola, M., Sitohang, A., Gunawan, T. N., Simatupang, M., Salsabila, S., Simanullang, T., & Hutasoit, T. (2025). *A Theoretical Study of Multicollinearity and Linearity in Econometric Models for Economic Research*. https://journal.unismuh.ac.id/index.php/jeb

Maulid. (2022). *Kenali Uji Asumsi Klasik Pada Metode Statistik Regresi*. https://dqlab.id/kenali-uji-asumsi-klasik-pada-metode-statistik-regresi

Singgih Santoso. (2011). *Srtuctural Equation Medeling (SEM), Konsep dan APliaksi dengan AMOS 18*.

Yhoga Hendrianto, A., Haliza, A. N., & Firdausi, M. F. (2023). *Socius: Jurnal Penelitian Ilmu-Ilmu Sosial Statistical Analysis in Multicorrelation Test Conditions: Confronting the Incompatibility of Classical Assumptions*. https://doi.org/10.5281/zenodo.10420827