UNP JOURNAL MANAGEMENT CENTER

math.ppj.unp.ac.id
p-ISSN 2716-0726
e–ISSN 2716-0734

Indexed

Article History
Vol 3, No. 2, 2024

Subject Areas:
Statistics

Keywords:
Correlation analysis, Statistical regression, Climate variability

Author for correspondence:
Lita Lovia
e-mail: litalovia80@gmail.com

RANGKIANG MATHEMATICS JOURNAL
Journal of Mathematics and Mathematics Education

Linear Regression to Analyze Temperature and Air Humidity on Rainfall: A Case Study at Padang Panjang Geophysics Station (2020-2023)

Lita Lovia[1], Yessy Yusnita[2] and Alona Dwinata[3]

[1] Mathematics, Faculty Pharmacy, Sains and Technology, Dharma Andalas University, West Sumatera

[2] Civil Engineering, Faculty of Engineering, Institut Teknologi Padang, West Sumatera

[3] Mathematics Education, Faculty of Teacher Training and Education, Raja Ali Haji Maritime University, Tanjung Pinang Riau Islands

Abstract- This study aims to analyze the correlation between temperature, humidity, and Rainfall at the Geophysics Station of Padang Panjang from 2020 to 2023. Monthly data from the Meteorology, Climatology, and Geophysics Agency (BMKG) was used to evaluate the relationship between these meteorological variables. Statistical analysis, including correlation and multiple linear regression, was conducted using SPSS version 22. The results show a significant negative correlation between Rainfall and temperature, indicating that temperature tends to decrease as rainfall increases. In contrast, a positive correlation between Rainfall and humidity suggests that higher humidity levels are associated with increased Rainfall. However, the regression analysis reveals that temperature and moisture explain only 16.76% of the variation in Rainfall, indicating the potential influence of other factors not included in the model.

1. Introduction

Rainfall is an essential element in the global climate system that influences many aspects of human life, such as agriculture, clean water and energy availability. In Indonesia, which is located in a tropical region with high rainfall intensity, uncertainty in rainfall patterns can trigger various natural disasters, including floods, landslides and drought. Excessive Rainfall can damage agricultural land, disrupt water supplies, and trigger environmental disasters. On the other hand, a lack of Rainfall can cause drought, which

Disrupts food security and people's lives. Therefore, understanding the factors that influence Rainfall is crucial for predicting, managing and reducing the impact of increasingly extreme climate change. (Supriyanti, Damiri and Ramadhan, 2024)

Air temperature and air humidity are two main meteorological factors that are interrelated and play an essential role in the process of rainfall formation(Pipit Muliyah, Dyah Aminatun, Sukma Septian Nasution, Tommy Hastomo, Setiana Sri Wahyuni Sitepu, 2020). Air temperature affects the amount of water vapour that the atmosphere can store. When air temperatures increase, evaporation from the surface of seas, rivers and lakes also increase, producing more water vapour in the atmosphere, which has the potential to become rain. Extreme temperature increases are often associated with extreme weather, such as heavy rains or droughts, impacting ecosystems and communities. On the other hand, air humidity reflects the amount of water vapour in the atmosphere and plays an important role in cloud formation and precipitation. The higher the air humidity, the greater the potential for thick clouds to form, which can produce Rainfall. Previous research shows that high temperature and humidity often correlate with Rainfall. (Catursari, 2024)

Research on the relationship between temperature, air humidity and Rainfall is becoming increasingly crucial in facing the challenge of global climate change. Erratic changes in weather patterns due to global warming make the frequency and intensity of Rainfall challenging to predict, increasing the risk of natural disasters. In addition, temperature and humidity fluctuations caused by climate change can potentially exacerbate extreme weather conditions, such as tropical storms and heat waves, which significantly impact people's social and economic lives.

In Indonesia, which has a tropical climate with high and uneven Rainfall throughout the year, an in-depth understanding of the interactions between temperature, humidity and rainfall variables is necessary. Correlation analysis between these variables can provide essential insights in predicting rainfall patterns, which helps decision-making related to disaster mitigation, water resources management, and agricultural sector planning.

This research aims to analyze the relationship between temperature, air humidity and Rainfall at the Padang Panjang Geophysical Station from 2020 to 2023. Using regression and correlation statistical analysis methods, this research will evaluate how strong the relationship between these three variables is and how significant the influence is. Temperature and air humidity vary with Rainfall in this region. Through a better understanding of the factors that influence Rainfall, it is hoped that recommendations can be produced to improve preparedness for the impacts of extreme weather and more effectively manage climate risks in the future.

2. Methods

This research uses a quantitative approach with the Regression Analysis method. A quantitative approach was chosen because this research aims to measure the relationship between the independent and dependent variables and test the independent variable's significant influence on the dependent variable. Linear regression analysis is used to identify how much impact the independent variable has on the dependent variable.

The stages of this research are as follows:

Variable Identification

Variable identification is the stage of determining the dependent and independent variables based on data from https://sumbar.bps.go.id/id. This research involves three types of variables, namely the dependent variable Rainfall (Y), the independent variable temperature (X_1) and air humidity (X_2). The data used in this research was collected from the West Sumatra Central Statistics Agency's official website (BPS) from 2020 to 2023. The data taken came from the Padang Panjang Meteorological Station, which provides essential information regarding relevant meteorological variables to support the analysis in this research.

Data analysis

At this stage, a series of classical assumption tests are carried out to ensure the suitability of the data before further analysis. Next, multiple linear regression analysis was carried out to see the relationship between the independent and dependent variables. Hypothesis testing is carried out to assess the significance of the relationship between variables. All data analysis was carried out using SPSS version

22 software to obtain accurate and valid results. The following are the steps for data analysis:

a) Classical Assumption Testing

The classic test in linear regression analysis aims to ensure the model used is valid and meets the basic assumptions of regression. The steps that need to be taken are to test linearity, normality, multicollinearity, heteroscedasticity and autocorrelation.

b) Multiple Linear Regression Analysis

General form of multiple linear regression model with n independent variables

$$Y = B_0 + B_1X_1 + B_2X_2 + \cdots + B_nX_n + \varepsilon$$

Where :

Y	= Dependent variable (Y)
$X_1, X_2, \cdots, X_n$	= Independent variables (X)
B_0	= Constant
$B_1, B_2, \cdots B_n$	= Estimated Coefficient
ε	= Random Error

Least Squares Method

The least squares method is used to estimate the coefficient values. $B_0,\ B_1, B_2, \cdots B_n$ By minimizing the sum of the squares of the residuals (the difference between the predicted value and the actual value). The goal is to find the coefficient value that minimizes the sum of the squares of the differences between the value predicted by the model and the actual value of the dependent variable.

Hypothesis Testing.

In testing this hypothesis, a t-test for each regression coefficient is carried out to determine whether the independent variable significantly affects the dependent variable. In addition, the F test will be used to evaluate the overall significance of the regression model so that it can be seen whether the model built as a whole can significantly explain variations in the dependent variable.

Model Evaluation

The Coefficient of Determination, usually denoted R-squared (R^2), is a statistical measure used in regression analysis to describe how well an independent variable explains variation in the dependent variable.

Interpretation of Results:

Interpret the estimated coefficients, explaining the implications of the findings about the research question. Discuss the significance and direction of the relationships between variables.

3. Results and Discussion

Identify variables

The data used is 48 monthly data. The dependent variable (Y) in this research is Rainfall (mm3). The independent variables are temperature (X_1) and humidity (X_2). The data used is secondary data, with data sources from https://sumbar.bps.go.id/id for data from the Meteorology, Climatology & Geophysics Agency for Padang Panjang Geophysical Station, namely Rainfall, temperature and air humidity data from 2020 to 2023.

Data analysis

(a) Classical Assumption Testing

1. Linearity Test

The linearity test was conducted using SPSS software using the compare means method. The results of the linearity test can be seen in the following table:

Table 1. Anova Table for Linearity between Rainfall with Temperature

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Y * X1	Between Groups	(Combined)	403141.065	26	15505.426	.886	.619
		Linearity	111951.153	1	111951.153	6.401	.019
		Deviation from Linearity	291189.912	25	11647.596	.666	.835
	Within Groups		367303.325	21	17490.635		
	Total		770444.390	47			

The linearity test shows a significant linear relationship between variables temperature and Rainfall (Sig. value = 0.019 < 0.05). This indicates that temperature changes have a linear relationship with Rainfall, where increases or decreases in temperature tend to be followed by changes in Rainfall with predictable patterns. The deviation test from linearity shows no significant deviation from the linear pattern (Sig. value = 0.835 > 0.05). This means that the relationship between temperature and precipitation remains stable and consistent with a linear pattern, with no evidence to suggest that a non-linear model is a better fit. Therefore, a linear model is sufficient to describe the relationship between these two variables without needing a more complex model.

Table 2. Anova Table for Linearity between Rainfall with Air Humidity

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Y * X2	Between Groups	(Combined)	639869.815	40	15996.745	.858	.656
		Linearity	98021.535	1	98021.535	5.255	.056
		Deviation from Linearity	541848.280	39	13893.546	.745	.743
	Within Groups		130574.575	7	18653.511		
	Total		770444.390	47			

Sig. (0.056): p-value for linearity test. This is close to the 0.05 limit, meaning the linear relationship is almost significant but not strong enough to be considered statistically significant (p = 0.056 > 0.05). This shows a possible linear relationship between the Rainfall and air humidity but with the uncertainty of the normality test. Sig. (0.743): p-value for the deviation test from linearity. Since the p-value> 0.05, this indicates that there is no significant deviation from the linear relationship. So, if there is a relationship between Rainfall and air humidity, the relationship tends to be linear.

2. Normality Test

The normality test uses the Kolmogorov-Smirnov test with the help of SPSS 22. The results of the normality test can be seen in Figure 1.

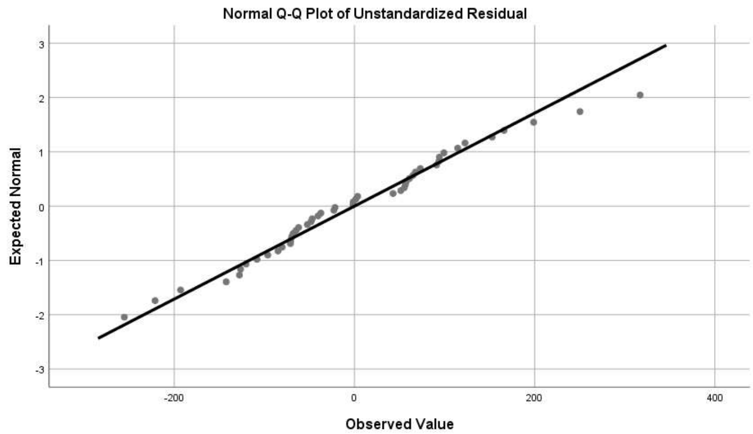

Figure 1. Normal Plot Q-Q of Residual

Based on the Normal Q-Q Plot of the unstandardized residuals in Figure 1, most of the residual points are around a straight line, which indicates that the residual distribution is close to a normal distribution. However, several points deviate in the tail of the distribution, both on the negative and positive sides, indicating a slight deviation from normality in some of the data. However, this deviation is relatively small and insignificant, so overall, the residual normality assumption is quite fulfilled. For further confirmation, checking the p-value of a normality test such as Shapiro-Wilk or Kolmogorov-Smirnov is necessary. If the p-value is more significant than 0.05, it can be concluded that the residual distribution is not significantly different from the normal distribution, so the normality assumption is acceptable.

Table 3. Tests of Kolmogorov-Smirnov

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.083	48	.200*	.984	48	.742

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Based on the normality test results using Kolmogorov-Smirnov in Table 3, a statistical value of 0.083 was obtained with degrees of freedom (df) of 48 and a significance value (p-value) of 0.200. Because the p-value is more significant than 0.05, it can be concluded that there is no significant difference between the residual distribution and the normal distribution. Thus, the assumption of residual normality in this regression model is fulfilled, which means that the residual distribution is considered to follow a normal distribution.

3. Multicollinearity Test

Table 4. Collinearity Statistics

Coefficients

Model		Collinearity Statistics: Tolerance	VIF
1	X1	.601	1.665
	X2	.601	1.665

a. Dependent Variable: Y

Based on the results of the multicollinearity test, a Tolerance value of 0.601 and a VIF of 1.665 were obtained for the two independent variables, namely X1 and X2. A tolerance value greater than 0.1 and a VIF value smaller than 10 indicate that there is no indication of a multicollinearity problem in this regression model. Thus, the independent variables are not strongly correlated, and the multicollinearity assumption is met. The regression model can be used further without worrying about multicollinearity affecting the analysis results.

4. Heteroscedasticity

The heteroscedasticity test uses the White test, and the following results are obtained:

Table 5. Anova Table for Heteroscedasticity

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	405393223.885	2	202696611.942	.496	.612[b]
	Residual	18397082907.508	45	408824064.611		
	Total	18802476131.393	47			

a. Dependent Variable: RES_1_SQ
b. Predictors: (Constant), X2, X1

Based on the ANOVA results obtained from the White test for heteroscedasticity, the F value of 0.496

indicates that the regression model used to explain the variation in the squared residuals is not significant. This value indicates that the independent variables (X1 and X2) cannot significantly explain the variations that occur in the residual square. Moreover, the p-value of 0.612 is much more significant than the commonly used significance level (0.05), so we cannot reject the null hypothesis stating the absence of heteroscedasticity in the model. Thus, there is no indication of heteroscedasticity in this regression model, which indicates that the assumption of homoscedasticity (constant residual variance) is met. This result is significant because heteroscedasticity can affect the validity of the regression coefficient estimates and the resulting statistical inferences.

math.ppj.unp.ac.id Rangkiang Mathematics Journal Vol 3 No 2, 2024

5. Autocorrelation test

Table 6. Durbin-Watson Test

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.409[a]	.168	.131	119.37788	1.994

a. Predictors: (Constant), X2, X1
b. Dependent Variable: Y

Durbin-Watson value = 1.994: Close to 2, indicating no autocorrelation in the residuals. This value shows that the regression model residuals tend to be independent. There was no indication of significant positive or negative autocorrelation. This means that the rainfall and temperature values at a particular time are not correlated with the Rainfall and temperature values at the previous time. Without a repeating pattern of prediction errors, this regression model is valid and meets the assumption of residual independence. Therefore, the prediction results from this model are reliable, and changes in Rainfall and temperature values at any time are considered random and not influenced by previous data.

(b) Linear Regression Model

The table presents the results of a linear regression with the dependent variable **Y** (Rainfall) and two independent variables, **X1** (temperature) and **X2** (air humidity)

Table 7. Table of Coefficients in linear regression output

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	965.835	1544.614		.625	.535
	X1	-62.481	42.279	-.259	-1.478	.146
	X2	9.185	8.362	.193	1.098	.278

a. Dependent Variable: Y

Based on the regression results presented in the coefficient table, two independent variables, temperature and air humidity, were analyzed to predict Rainfall. The constant (intercept) value obtained is 965,835, which shows that if the two independent variables are zero, the predicted Rainfall is 965,835 mm. Although this value provides a basic idea, it is practically less relevant because air temperature and humidity are unlikely to be at zero values in the context of weather. The Sig value shows the significance of this constant. = 0.535, more significant than 0.05, indicating that this constant is not statistically significant.

Furthermore, analysis of the temperature variable shows that every 1 degree Celsius increase in temperature will decrease Rainfall by 62,481 mm, which shows a negative relationship between temperature and Rainfall. However, the value of t = -1.478 and Sig. = 0.146 indicates that the effect of temperature on Rainfall is not statistically significant because the significance value is greater than 0.05. This means that although temperature tends to influence Rainfall, this relationship is not strong enough to be considered necessary in this model.

On the other hand, the air humidity variable shows a positive relationship with Rainfall, where every 1 unit increase in air humidity can increase Rainfall by 9,185 mm. The t value = 1.098 and Sig. = 0.278 also indicates that the effect of air humidity on Rainfall is not statistically significant because the significance value is higher than 0.05

The regression equation resulting from this analysis can be written as:

$$\boldsymbol{Y} = 965.835 - 62.481X_1 + 9.185X_2$$

(c) Hypothesis Testing

Table 8. ANOVA Table (Analysis of Variance Table), which shows the results of the regression significance test

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	129145.844	2	64572.922	4.531	.016[b]
	Residual	641298.546	45	14251.079		
	Total	770444.390	47			

a. Dependent Variable: Y

a. Predictors: (Constant), X2, X1

Based on the results of ANOVA analysis, the regression model that uses air humidity and temperature as predictor variables for Rainfall shows statistically significant results. The F value obtained is 4.531 with a significance level of 0.016, indicating that this regression model is essential at the 95% confidence level. This means that, overall, the two independent variables (air humidity and temperature) significantly influence the dependent variable (Rainfall). With a significance value below 0.05, we can conclude that this model is suitable for explaining variations in Rainfall caused by temperature and air humidity.

The total variance in Rainfall is 770,444,390 (Total Sum of Squares); this model can explain 129,145,844 (Sum of Squares Regression), equivalent to around 16.76% of the total variation in Rainfall. In other words, air humidity and temperature can explain about 16.76% of the variation in Rainfall. The remainder, amounting to 641,298,546 (Sum of Squares Residual) or around 83.24%, is a variation that this model cannot explain and may be caused by other variables not included in the model or by error (prediction error).

The Mean Square value for the regression is 64,572,922, which measures how much variation the two predictor variables explain on average. Meanwhile, the Mean Square value for the residual is 14,251,079, measuring how much average variation cannot be explained by the model.

Overall, although the model is significant, the model's ability to explain variations in Rainfall is limited. A large proportion of rainfall variation (more than 80%) cannot be explained by air humidity and temperature alone. This suggests that other factors may play an essential role in influencing Rainfall, which are not captured in this model. However, temperature and air humidity are significant predictors, although their contribution to rainfall prediction is tiny.

4. Conclusion

The conclusion of this research shows a relationship between air temperature, air humidity and Rainfall at the Padang Panjang Geophysical Station from 2020 to 2023. Multiple linear regression and correlation analysis reveal that Rainfall has a significant negative relationship with air temperature, where an increase in Rainfall tends to lower temperatures. In contrast, air humidity has a positive relationship with Rainfall, indicating that higher humidity leads to increased Rainfall. However, the regression results show that temperature and air humidity can only explain 16.76% of the variation in Rainfall, so other factors may play a role but are not included in this model.

Based on these findings, it is recommended that future research models include other potentially influential variables, such as wind speed, air pressure, or geographic factors, to improve the ability to predict Rainfall. In addition, the use of meteorological data from other weather stations in West Sumatra or more extended periods will help expand understanding of rainfall patterns in this region.

References

1. Catursari, S.A. (2024) ‘Analisis Pengaruh Tekanan Udara, Kelembaban Udara, Dan Suhu Udara Terhadap Curah Hujan Di Kota Surabaya’.
2. Harlan J (2018) Analisis Regesi Linear. Penerbit : Gunadarma. Pp : 13-20.
3. Mardani, A. et al. (2017) 'Recent Fuzzy Generalizations of Rough Sets Theory : A Systematic Review and Methodological Critique of the Literature', 2017.
4. Marni and Ishak Jumarang, M. (2016) ‘Analisis Hubungan Kelembaban Udara dan Suhu Udara Terhadap Parameter Tebal Hujan di Kota Pontianak’, Prisma Fisika, IV(03), pp. 80–83.
5. Muniarty Puji, salsabila, L.L. dkk (2024) Analisis Regresi, GetPress. Available at: https://www.semanticscholar.org/paper/ANALISIS-REGRESI-KUANTIL-Saidah-Yanuar/3acb55ead5001dcd6ed0b118c38dc6c9dc249668.
6. Pipit Muliyah, Dyah Aminatun, Sukma Septian Nasution, Tommy Hastomo, Setiana Sri Wahyuni Sitepu, T. (2020) ‘済無No Title No Title No Title’, Journal GEEJ, 7(2), pp. 398–406.
7. Regresi, P. et al. (2013) ‘Persamaan Regresi Prediksi Hujan Bulanan Di Pontianak Dengan Prediktor Suhu Dan Kelembapan Udara’, Forum Ilmiah, 10, p. 230.
8. Rusdi, W. et al. (2023) ‘Analisis Perbandingan Metode Fuzzy Tsukamoto dan Regresi Linier Berganda dalam Peramalan Jumlah Produksi Kopi’, Remik, 7(2), pp. 1016–1031. Available at: https://doi.org/10.33395/remik.v7i2.12267.
9. SEMBIRING, R. (no date) ANALISIS REGRESI. Penerbit ITB Bandung.
10. Supriyanti, K.R., Damiri, B.A. and Ramadhan, W.N. (2024) ‘Pengelompokan Faktor-Faktor yang Mempengaruhi Curah Hujan di Provinsi Sumatera Utara Menggunakan Metode Fuzzy C-Means’, Jurnal Statistika dan Komputasi, 3(1), pp. 1–10. Available at: https://doi.org/10.32665/statkom.v3i1.2623.
11. Suyono (2015) ANALISIS REGRESI UNTUK PENELITIAN. deepublish.
12. Trenberth, K.E., Fasullo, J.T. and Shepherd, T.G. (2015) 'Attribution of climate extreme events', Nature Climate Change, 5(8), pp. 725–730. Available at: https://doi.org/10.1038/nclimate2657.
13. Tukidi (2010) ‘Karakter Curah Hujan Di Indonesia’, Jurnal Geografi, 7(2), pp. 136–145. Available at: http://journal.unnes.ac.id/nju/index.php/JG/article/view/84.
14. Yessy Yusnita and Lita Lovia (2024) 'Mode Selection Model Based on Travel Time and Price of Goods with Regression Analysis Model', Rangkiang Mathematics Journal, pp. 41–47. Available at: https://doi.org/10.24036/rmj.v3i1.50.
15. Yusnita, Y, dkk (2020) ‘Model Pemilihan Moda Berdasarkan Variabel Kepemilikan Kendaraan dan Kategori Luas Lahan Parkir Dengan Teknik Analisis Regresi’, Phytagoras, 9(2), pp. 95–105.