JURNAL EMACS
[Engineering, MAthematics and Computer Science] Vol.6 No.2 May 2024: 151-156
e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v6i2.11466

Relationship Between Temperature and Humidity on Rainfall: A Multiple Linear Regression Analysis

Mohammad Dian Purnama[1*], Mutia Eva Mustafidah[2]

[1,2] Mathematics Department, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Surabaya,
Surabaya, Indonesia 60231
mohammaddian.20053@mhs.unesa.ac.id; mutia.20010@mhs.unesa.ac.id

*Correspondence: mohammaddian.20053@mhs.unesa.ac.id

Abstract – *One of the biggest challenges in tropical countries is flooding caused by heavy rainfall. Not only does it cause flooding, but rainfall also affects several sectors, especially agriculture. Areas that have a lot of rain-fed agricultural land, particularly rice fields, depend on rainfall because it determines crop yields. This study aimed to assess the relationship between rainfall and various factors such as temperature and humidity in Mojokerto district, where agricultural activities are a key pillar of the economy. The experiment using multiple linear regression with starts the data collection, classical assumption tests, modelling, and significance test. The experimental process involved collecting temperature and humidity data from weather stations across the 12 sub-districts and rainfall data from agricultural records. Based on the results, all predictor variables (temperature and humidity) showed a strong and positive relationship with rainfall, with a correlation coefficient of 0.760007. From simultaneous test indicated that both variables had a significant effect. At a significance level of 5% (0.05), the partial test revealed that only the humidity variable had a significant effect on the amount of rainfall. These factors together had a coefficient of determination of 0.57761, meaning the predictor variables accounted for 57.761% of the variation in rainfall, while the remaining 42.239% was due to other variables.*

Keywords: *Rainfall; Multiple Linear Regression; Relationship*

I. INTRODUCTION

In tropical regions especially Indonesia, rainfall exhibits a high diversity in climate elements. The characteristics of rainfall vary across different areas, influenced by factors such as geographical location, topography, the presence of mountains and valleys, and the structure and orientation of islands. Consequently, the distribution pattern of rainfall tends to be uneven across wide geographical scopes (Sari & Fitriyani, 2021).

Rainfall plays a crucial role in human life and is interconnected with other weather elements such as temperature and humidity. Air temperature reflects the level of molecular activity in the atmosphere, measured as the average kinetic energy of molecular movement (Putra, 2023). Humidity represents the total amount of water content in the air at a given time, expressed as a percentage of water vapor to the saturated vapor pressure at a specific temperature (Nurhayati et al., 2020).

Rainfall is measured as the height of rainwater on a flat surface that does not evaporate, infiltrate, or flow. For instance, a 1 mm rainfall means that one millimeter of water covers one square meter of a flat surface within a specific time frame (Umami et al., 2020). The Dictionary of Indonesian Language (KBBI) defines rainfall as the quantity of rain that falls in a particular area over a specific period.

High rainfall can lead to river flooding, increasing the risk of disasters in Mojokerto Regency. This region has numerous rain-fed agricultural areas, serving as the primary livelihood for the population. Agricultural conditions are significantly influenced by rainfall. Therefore, understanding the relationship between rainfall and influencing factors is crucial (Hidayatullah & Aulia, 2020).

Factors such as humidity, temperature, pressure, and wind speed can affect rainfall (Luthfiarta et al., 2020). In this context, statistical regression analysis can be employed to comprehend these relationships. This study utilizes a regression model with rainfall as the response variable, and temperature and humidity as predictor variables.

The primary objective of this research is to assess the impact of temperature and humidity on rainfall in Mojokerto district. By understanding these relationships, we can better predict and manage agricultural conditions and mitigate flood risks. The contributions of this study lie in providing actionable insights that align with the local climate conditions, thereby aiding in more directed and effective decision-making for agricultural and disaster management practices.

II. METHODS

The research methodology employed is multiple linear regression analysis. Regression analysis looks for a relationship between variables and expresses that relationship as a mathematical equation in order to evaluate research ideas. Multiple linear regression analysis seeks to determine how two or more predictor variables (X) affect a response variable (Y) (Wisudaningsi et. al, 2019).

Before conducting the analysis, the data undergoes classical assumption testing to determine if any classical assumption issues exist in an Ordinary Least Squares (OLS) linear regression model (Mardiatmoko, 2023). Classical assumption tests encompass checks for normality, autocorrelation, homoscedasticity, and multicollinearity within the data.

The experiment starts with data collection, normality test, autocorrelation test, Coefficient of Determination calculation, Multiple Correlation Coefficient calculation, modelling with multiple linear regression, simultaneous significance test with Analysis of variance (ANOVA), and partial significance test with t-test. The illustration of the experimental steps can be seen in Figure 1.

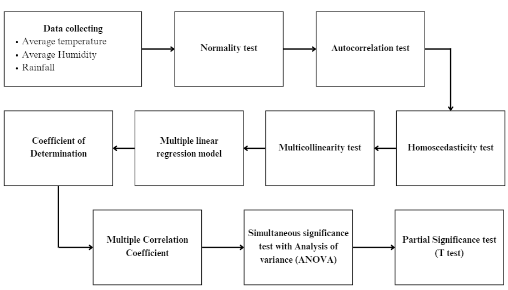

Figure 1. Steps of experiment

2.1 Data Acquisition

This experiment uses secondary data obtained from Badan Pusat Statistik (BPS) Kabupaten Mojokerto. The data includes average temperature and average humidity as predictor variables and rainfall as the response variable. All this data is sourced from 2022 in all sub-districts in Mojokerto district with a total of 12 sub-districts. The operational definitions of the variables used in this study can be seen in Table I.

Table I. Experiment variables

Variable	Definition
Y	Amount of Rainfall
X_1	Temperature
X_2	Humadity

2.2 Normality Test

For parametric statistical tests to be performed, the data used in the measurement of interval or ratio scale data must be regularly distributed. The purpose of the normality test is to determine if the residuals have a normal distribution (Quraisy, 2020). A well-fitted regression model assumes normally distributed residuals. Instead of testing each individual variable, this test is run on the residuals. For the normalcy test, the Shapiro-Wilk test is frequently employed, with the following hypothesis formulated:

H_0: The data has normally distributed

H_1: The data has not normally distributed

Based on the rules in this test decision is to accept , if the shapiro-wilk count value of the normality test is greater than the shapiro-wilk table value at the significance level (α) = 5% = 0.05.

2.3 Autocorrelation Test

The autocorrelation test evaluates whether residuals in period t and errors in the preceding period (*t-1*) inside a linear regression model are correlated when performing the subsequent classic assumption test. The author employs the Durbin-Watson test for autocorrelation testing, comparing the calculated Durbin-Watson statistic from the regression analysis with critical values from the Durbin-Watson table (Sam et al., 2021). The hypothesis for this test is as follows:

H_0 : There data has no autocorrelation

H_1 : There data has autocorrelation

The formula is:

$$d = \frac{\sum_{t=2}^{t=n}(\widehat{e_t} - \widehat{e_{t-1}})^2}{\sum_{t=1}^{n} \widehat{e_t}^2} \quad (1)$$

Where:

d = Durbin Watson value

e_t = Residual value of the period t regression equation

$e_{(t-1)}$ = Residual value of the regression equation period *t-1*

Reject if the value of d is calculated or the value of Durbin Watson is less than the value of Durbin Watson lower limit table (dL), which means there is a negative autocorrelation problem where d < dL, or the value of d count is more than the value of 4 − dU which means there is a positive autocorrelation problem where d > 4 − dU. Aside from the rejection area, it means accept H0.

2.4 Homoscedasticity Test

The homoscedasticity test examines whether residuals exhibit consistent variance across all observations in a regression model. It is a crucial classic assumption test that must be conducted in regression analysis. This test helps detect violations of the classical assumptions in linear regression, specifically the absence of heteroscedasticity (Maulina et al., 2022). The hypothesis formulation for the heteroskedasticity test is as follows:

H_0 : Data does not occur heterokedasticity in predictor variables

H_1 : Data heterokedasticity occurs in predictor variables

Based on the rule in this test decision is to accept, if the *p-value* of the heterokedasticity test is greater than the significance level value (α) = 5% = 0.05

2.5 Multicollinearity Test

In a regression analysis, the multicollinearity test looks for connections or correlations between the independent variables. The variance inflation factor (VIF) for each independent variable is examined in relation to the dependent variable in this test. In the event when the VIF value of each independent variable in the regression model is less than 10, multicollinearity issues are absent (Ghozali, 2018). The following is the hypothesis for the multicollinearity test:

H_0: No multicollinearity between variables

H_1: There is multicollinearity between variables

Multicollinearity detection involves assessing the Variance Inflation Factor (VIF). Multicollinearity is deemed absent in a model if the VIF value is below 10 and the tolerance value exceeds 0.01 (Yaldi et al., 2021). The formula as follows:

$$VIF = \frac{1}{1-r^2} \tag{2}$$

where r^2 is coefficient of determination.

2.6 Multiple Linear Regression Model

The multiple linear regression model can be expressed by the following (Rath et. al, 2020):

$$Y=a+b_1 X_1+b_2 X_2+...+b_n X_n \tag{3}$$

where:

Y = Predicted variable

a = Intercept coefficient

$b_1,b_2,...,b_n$ = Regression coefficient

$X_1, X_2,...,X_n$ = Predictor variable

2.7 Coefficient of Determination

The regression model's predictor variable's coefficient of determination indicates how well it can account for variation in the response variable. It quantifies the percentage of influence that predictor variables (X) have on the response variable (Y), calculated using the formula:

$$r^2 = \frac{(b_1 \sum X_1 Y)+(b_2 \sum X_2 Y)}{\sum Y^2} \tag{4}$$

Once known, to determine the value of the coefficient of determination with the formula $r^2 \times 100\%$.

2.8 Multiple Correlation Coefficient (

An index or value called the multiple correlation coefficient is used to assess how closely two variables are related to one response variable and one is related to two predictor variables when there are three or more variables. The magnitude of the value of the double correlation coefficient can be calculated by the formula:

$$r = \sqrt{r^2} = \sqrt{\frac{(b_1 \sum X_1 Y)+(b_2 \sum X_2 Y)}{\sum Y^2}} \tag{5}$$

The range of the correlation coefficient is -1 to +1. The correlation coefficient relationship's meaning or direction is shown by positive and negative signs. One variable rising will lead to an increase in other variables, and vice versa, according to the positive correlation value. According to the negative correlation value, a rise in one variable will cause a reduction in the other, and vice versam shown in Table II.

Table II. Interpretation of the Correlation Coefficient r value

Correlation Coefficient Interval	Relationship Level
0,00-0,19	Very Low
0,20-0,39	Low
0,40-0,59	Normal
0,60-0,79	Strong
0,80-1,00	Very Strong

2.9 F-Test (ANOVA)

Simultaneous significance testing is carried out with the aim of determining whether all parameter estimates in the model are significant overall. The test using Analysis of Variance (ANOVA) to evaluates the collective impact of all predictor variables on the response variable (Sofro et al., 2019). The formulation of the hypothesis for this test is as follows:

H_0: $\beta_1 = \beta_2 = 0$

H_1: $\beta_j \neq 0$ (At least nothing β is different)

With Significance level α = 5% = 0.05, in this case, Define test criteria. F_{value}, has compared to F_{table}. If F_{value} > F_{table}, have meaning rejected . The statistical tests is:

$$SS_T = \sum y^2 = \sum Y^2 - n\bar{Y}^2 \tag{6}$$

$$SS_R = b_1 \sum x_1 y + b_2 \sum x_2 y \tag{7}$$

$$SS_E = SS_T - SS_R \tag{8}$$

$$MS_R = SS_R / k \tag{9}$$

$$MS_E = SS_E / (n-k-1) \tag{10}$$

$$F_{value} = MS_R / MS_E \tag{11}$$

2.10 Partial Significance Test

Partial test is a testing method carried out on each parameter separately in a regression model. This test aims to assess the significance of each parameter in the model (Sofro et. al., 2019). This test using t-test with hypothesis testing criteria for partial tests as follows:

H_0: $\beta_j = 0$: (Predictor variable has no significant effect on the response variable).

H_1: $\beta_j \neq 0$: (Predictor variable has significant effect on the response variable).

With Significance level α = 5% = 0.05, in this case, t_{value}, has compared to t_{table}, with $t_{table} = t\ \alpha/2\ ;\ n - k - 1$. If $t_{value} > t_{table}$, have meaning rejected . The statistical tests is:

$$t_k = b_k / Sb_k \tag{12}$$

where:

t_k = t-value for predictor variable to k

b_k = Regression coefficient for predictor variable to k

Sb_k= Standard deviation of the regression coefficient of the predictor variable to k

Standard deviation (standard error) is a value that expresses how far the regression value deviates from the actual value.

$$S_e = \sqrt{\frac{\sum y^2 - (b_1(\sum x_1 y) + b_2(\sum x_2 y))}{n-m}} \tag{13}$$

$$Sb_k = \sqrt{\frac{S_e}{\sum x_i^2 - (1 - r_k^2)}} \tag{14}$$

where:

Sb_k = Standard deviation of the regression coefficient for the predictor variable to k

S_e = Standard error estimation

r_k^2 =Squared correlation between with other predictor variables.

III. RESULT AND DISCUSSION

3.1 Normality Test

The shapiro-wilk normality test yielded a value of 0.859 for the shapiro-wilk table and a probability value or P-value of 0.947969 when used to conduct a normality test. Because the shapiro-wilk calculated value of the normality test is greater than the shapiro-wilk table value (0.947969 > 0.859) at the significance level (α) = 5% = 0.05, so the decision is accept H_0, the sample comes from a normally distributed population.

3.2 Autocorrelation Test

By conducting an autocorrelation test using Durbin Watson, the result of the value dU = 1.5794, the value of 4 dU = 2.4206, and the calculated value of Durbin Watson (*dW*) is 2.232336. Since dU< dW < 4 dU (1.5794 < 2.232336 < 2.4206), so the decision is accept H_0 or there is no positive or negative autocorrelation.

3.3 Heteroskedasticity Test

The results of the heteroskedasticity test, which was conducted, were 0.729234, 0.595086, and 0.658569 for each variable's p-value. The resulting number indicates that the p-value exceeds the significance level value (α) of 5% = 0.05. Accepting means that there is no heteroskedasticity in the data for the predictor variables.

3.4 Multicollinearity Test

By conducting a multicollinearity test, the tolerant value is 0.689823 and the VIF value is 1.449647. Since the VIF is smaller than 10 (1.449647 < 10) and the tolerant value is greater than 0.01 (0.689823 > 0.01), so thedecision is accept H_0, The data does not occur multicollinearity.

3.5 Multiple Linear Regression Model

After the calculations carried out, the equation for multiple linear regression from equation 3 is obtained

$$Y = -6431{,}92 + 132{,}64\ X_1 + 36{,}31\ X_2$$

3.6 Coefficient of Determination

From the results of the calculations carried out, the value r^2 is 0.57761, this means that the contribution of influence from both variables is 57.761%.

3.7 Multiple Correlation Results

Based on the computed results, the correlation coefficient (*r*) of 0.760007 indicates a robust and positive relationship between temperature and humidity collectively affecting rainfall amounts.

3.8 Results of Simultaneous Significance Test

The simultaneous test used is that the variables average temperature and average humidity together have no significant effect on the variable amount of precipitation, while the alternative hypothesis is that the variables average temperature and average humidity together have an effect significant on the variable amount of rainfall.

Table III. Simultaneous Significance Test result

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Regression	104296	2	52148	6,15	0,021
Residuals	76268,8	9	8474,3		
Total	180564,8	11			

Based on the Table III, that have been carried out at the significance level value (α) = 5% = 0.05, a result of 6.15366 and a value of 4.26 were obtained. Based on the test criteria due $F_{value} > F_{table}$ (6.15366 > 4.26), then the initial hypothesis is rejected or is in the area of rejection. Therefore, with a P-value of 0.020688, which is less than α (the significance level), the results fall within the rejection region of the initial hypothesis. This indicates statistical significance, suggesting that the observed effect of the predictor variable (average humidity) on the response variable (amount of rainfall) is unlikely to have occurred by chance. From the two calculations carried out at the significance level value (α) = 5% = 0.05 shows that the variables of average temperature and average humidity together have an effect (significant) on the variable amount of rainfall. The combination of these two factors simultaneously can provide rainfall estimates that are more related to and influenced by air temperature and humidity. This information is crucial in the long term to anticipate the rainy season and help in mitigating the risk of crop failure due to rainfall in the area.

3.9 Results of Partial Significance Test

The partial test used is a testing method performed on each parameter separately in a regression model.

Table IV. Partial Significance Test

Variable	Standard deviation (Sb_k)	t_{value}
X_1 (average temperature)	79,80	1,66
X_2 (average humidity)	10,40	3,49
Total	90,2	5,15

At the significance level value (α) = 5% = 0.05 the t-table is 2.262. Based on the Table IV , that have been carried out, the results of the calculated t value on the variable of 1.66 because (1.66 < 2.262), then the initial hypothesis is confirmed or falls into the receiving region, indicating that there is no discernible relationship between the predictor variable (average temperature) and the response variable (amount of rainfall). While the results of the t value are calculated on the variable of 3.49 because (3.49 > 2.262). The partial test hypothesis is not supported, suggesting that the response variable (amount of rainfall) is significantly influenced by the predictor variable (average humidity). This rejection implies that the influence of average humidity on rainfall is supported statistically.

From the table 4 that have been done, the standard error for is 92.056 and the standard error for is 79.80 and the standard error for is 10.40. The calculation of the standard error carried out on each regression coefficient aims to measure the magnitude of deviation from each regression coefficient. The lower the default error, the more instrumental the variable will be in the model. So in this case the variable humidity plays a greater role in the multiple linear regression model performed.

From the partial test results, it was found that humidity partially has a significant influence on rainfall. So it is important to understand the contribution of moisture in isolation to rain formation. The implication is that agricultural management strategies based on accurate moisture forecasts can help farmers to optimally time planting and irrigation, reducing the risk of water shortage or excess water that can be detrimental to crop growth.

IV. CONCLUSION

There is a substantial correlation between the predictor variables (temperature and humidity) and the response variable (rainfall), as evidenced by the coefficient of determination, which shows that both variables had a 57.761% influence on the response variable. There was a significant degree of correlation between the predictor and response variables, as evidenced by the multiple correlation of 0.760007, the magnitude of which was obtained simultaneously between the predictor and responder variables.

Based on the simultaneous significance test at the significance level value (α) = 5% = 0.05, the result is a variable average temperature and average humidity together have an effect significant on the variable amount of rainfall. Based on the partial significance test, it is concluded that the predictor variable (average temperature) does not significantly affect the response variable (amount of rainfall). While for the predictor variable (average humidity) has a significant effect on the response variable (amount of rainfall).

From the partial test results, it was found that humidity has a significant influence on rainfall, emphasizing the importance of understanding the contribution of humidity in rainfall formation. Agricultural management strategies that use moisture forecasts can help farmers optimally organize planting and irrigation, reducing the risk of water shortage or excess that can be detrimental to crop growth. Meanwhile, the simultaneous combination of temperature and humidity provides more pertinent rainfall forecasts, useful for organizing the growing season over the long term and supporting risk mitigation due to extreme weather.

REFERENCES

Ghozali, I. (2018). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25*. Semarang: Badan Penerbit Universitas Diponegoro.

Hidayatullah, M. L., & Aulia, B. U. (2020). Identifikasi dampak perubahan iklim terhadap pertanian tanaman padi di Kabupaten Jember. *Jurnal Teknik ITS, 8*(2), 143-148.

Luthfiarta, A., Febriyanto, A., Lestiawan, H., & Wicaksono, W. (2020). Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda. *JOINS, 5*(1), 10–17.

Mardiatmoko, G. (2020). Pentingnya uji asumsi klasik pada analisis regresi linier berganda (studi kasus penyusunan persamaan allometrik kenari muda [canarium indicum l.]). *BAREKENG: Jurnal Ilmu Matematika Dan Terapan, 14*(3), 333-342.

Maulina, V., Harun, L., & Sutrisno, S. (2022). Pengaruh Minat Belajar dan Resiliensi Matematis Terhadap Kemampuan Pemecahan Masalah Matematis Siswa. *Imajiner: Jurnal Matematika Dan Pendidikan Matematika, 4*(4), pp. 347-354.

Nurhayati, E., Rizaldi, D. R., & Fatimah, Z. (2020). Pencegahan Penyebaran Covid-19 Melalui Inaktivasi Virus dalam Kajian Kinetika, Termodinamika dan Kesetimbangan. *Jurnal Ilmiah Profesi Pendidikan, 5*(2), 102-107.

Putra, R. M. (2023). *Prediksi Kelembaan Udara di Provinsi Jawa Timur Menggunakan Metode Regresi Linear Sederhana* (Masters's thesis), Available from E-Learning Universitas Islam Lamongan. (Unisla No. 2184)

Quraisy, A. (2020). Normalitas data menggunakan uji kolmogorov-smirnov dan saphiro-wilk: studi kasus penghasilan orang tua mahasiswa prodi pendidikan matematika unismuh makassar. *J-HEST Journal of Health Education Economics Science and Technology, 3*(1), 7-11.

Rath, S., Tripathy, A., & Tripathy, A. R. (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes & metabolic syndrome: clinical research & reviews, 14*(5), 1467-1474.

Sari, M., & Fitriyani, S. (2021). Analisis Spasial Temporal

Sosiodemografi Dan Variabilitas Iklim Terhadap Kejadian Tuberkulosis Paru BTA Positif di Provinsi Jawa Barat Tahun 2013-2017. *Jurnal Untuk Masyarakat Sehat (JUKMAS), 5*(2), 140-150.

Sam, M., Pamungkas, I., & Muzaini, M. (2021). Pengaruh Banyaknya Perusahaan Industri, Jumlah Tenaga Kerja, Jumlah Produksi Galian/Pertambangan Serta Luas Area Perkebunan dan Perikanan Terhadap PDRB Provinsi Sulawesi Selatan. *Infinity: Jurnal Matematika dan Aplikasinya, 2*(1), 72-83.

Sofro, A., Oktaviarina, A., & Maulana, D. A. (2019). Uji Hipotesis, Uji Z dan Uji T. *Metode Statistika* (pp. 35-54). Surabaya, Indonesia: Unesa Press.

Umami, A., & Sutikno, S. (2020). Perbandingan Estimasi Return Level Declustering dan Non Declustering pada Data Curah Hujan Ekstrem di Surabaya dan Mojokerto. *Jurnal Sains dan Seni ITS, 8*(2), 79-87.

Wisudaningsi, B. A., Arofah, I., & Belang, K. A. (2019). Pengaruh kualitas pelayanan dan kualitas produk terhadap kepuasan konsumen dengan menggunakan metode analisis regresi linear berganda. *Statmat: Jurnal Statistika Dan Matematika, 1*(1), 103-116.

Yaldi, E., Pasaribu, J. P. K., Suratno, E., Kadar, M., Gunardi, G., Naibaho, R., & Aryati, V. A. (2022). Penerapan uji multikolinieritas dalam penelitian manajemen sumber daya manusia. *Jurnal Ilmiah Manajemen dan Kewirausahaan (JUMANAGE), 1*(2), 94-102.