QUANTILE REGRESSION MODEL ON RAINFALL IN MAKASSAR 2019

Wahidah Sanusi[1], Sukarna[2*], Nur Harisahani[3]

[1,2,3]Department of Mathematics, FMIPA, Universitas Negeri Makassar
Jl. A.P. Pettarani, Makassar, 90222, Indonesia

Corresponding author's e-mail: * sukarna@unm.ac.id

ABSTRACT

Makassar is an area that has a monsoon rainfall pattern. This study aims to find a quantile regression model and determine the factors that significantly influence rainfall in the city of Makassar. This applied research applies a quantile regression model to rainfall data which is seasonal data. The advantage of this quantile regression model is that it can detect extreme rainfall conditions, such as heavy rain. The data used is daily data in 2019. The estimation results obtained 9 (nine) models from each quantile used. The best model is obtained based on the largest coefficient of determination (R^2), namely the 0.8th quantile (θ) of 0.28%. Furthermore, based on the model, it is found that the factor that humidity significantly influences rainfall in the city of Makassar. At the same time, the air temperature and wind speed have no significant effect on rainfall in the city of Makassar.

1. INTRODUCTION

Geographically, the city of Makassar, which is the capital of the province of South Sulawesi, is located between $119°24'17"$ east longitude and $5°8'6"$ south latitude [1]–[3]. Makassar city is located on the equator, which results in Makassar having a tropical climate with an average air temperature of $26.2℃ − 29.3℃$ with a humidity of 77 percent and an average wind speed of 5.2 knots [1]. Therefore, the rainfall in Makassar is quite varied throughout the year. Sometimes floods occur in the city of Makassar when the rainfall is quite high, meanwhile, when the rain is low some areas experience drought. [4], [5].

Rainfall is the amount of water that falls to the earth's surface. The degree of rainfall is stated by the amount of rainfall in a time unit. Grain of rain in the meteorology with diameter of more than 0.5 mm is called rain and diameter between 0.5-0.1 mm is called drizzle. The greater the size of the rain's grain is, the greater the rate of falling. Rainfall is influenced by several factors, including temperature, air humidity, wind speed, and air pressure. Rain classification view from the rate of fall [4], rainfall consists of: 1). Drizzle rain, if the rate of the falling range 0.5 m/sec., 2). Smooth rain, if the rate of the falling range 2.1 m/sec., 3). Normal rain, if the rate of the falling range 4-6.5 m/sec., 4). Heavy rain, if the rate of the falling range 8.1 m/sec.

The city of Makassar has a climate tropical with the average temperatures ranging between $26.2^0 C$ – $29.3^0 C$, air humidity range of 77 percent, and an average wind speed of 5.2 knots. In general, Makassar City has a rainy season from November to April and a dry season from May to October. Average annual rainfall around 256.08 mm/month [1], [4], [6]. Broadly speaking, Makassar has a tropical climate [7], [8] because this area is influenced by the western monsoon and equatorial rainfall patterns because of its location near the equator. The average monthly rainfall occurs in the rainy and dry seasons.

Based on weather conditions and rainfall, the city of Makassar has a monsoon rainfall pattern. The monsoon rainfall pattern has characteristics at the end and beginning of the year, having high rainfall with a peak that occurs around December, January, or February. Meanwhile, the lowest rainfall occurs around July, August, and September [9]. Makassar's high annual rainfall occurred in 1999 – 2001 in the range of 3947 mm to 4703 mm, while the lowest annual rainfall occurred in 1998, which was 781 mm. This low amount of rainfall is a drought phenomenon, as well as the impact of the El Nino event in the 1997/1998 period [8].

Regression analysis [10]–[12] is a statistical method that aims to model the relationship between two variables consisting of a dependent variable with one or more independent variables in a system. In linear regression, there are several methods of parameter estimation [10]; one of them is the OLS (Ordinary Least Square) method. Regression analysis with the OLS method is based on a function called the mean, showing the size of the concentration of a distribution. The approach with this method is only able to suspect the model of the conditional function of the mean and does not represent the overall data of distribution, so the mean approach is less precisely used as a nudity of the middle value, eventually developed quantile regression method [13].

Quantile regression is an approach in regression analysis that suspect of various quantile functions of a distribution Y as a function of X [14]. The main advantage of quantile regression compared to OLS regression is flexibility in data modeling with a heterogeneous conditional distribution. This method can be used to measure the expansion of the explanatory not only in the center of the data translation but also at the top or bottom of the tail of the distribution [15].

According to Gujarati [10], one famous method that can be used to get regression estimates is the Ordinary Least Square (OLS) or often referred to classic regression method. The OLS method defines parameter estimates as a value that minimizes the number of squares between observation and models. The other estimation method is the maximum likelihood. This method tries to maximize the likelihood to predict the exact estimation. Both methods are frequently used by frequenter's data analysis.

In regression analysis, normality violations often occur at the time of data containing a confirmation that causes the formation of data transportation to be no longer symmetrical. As a result, the smallest square method is less precise in performing uncetrical data analysis, it is developing median regression model. The median regression method was done with the Least Absolute Deviation (LAD) approach developed by replacing the mean on OLS with the median by considering when the data of the bell is not symmetrical. Along with the development of time, the median regression approach is also less precise used because it only looks at two data groups only. So that the quantile regression method can be used on more than two data groups. The approach of the quantile regression method is by separating or dividing data into two or more groups in sustained differences in estimated value in the quantiles [16].

Quantile is the value that divides the series of data sorted into the same parts. The quantile that divides the data into the foreign amount of two parts is called median, four parts called quartiles (Q_1, Q_2, Q_3), ten parts called desil $(D_1, D_2, \ldots D_9)$, and to be a hundred parts called percentiles $(P_1, P_2, \ldots P_{99})$ [17]. Suppose Y is a random variable with distribution function F_Y and θ is a constant where $0 < \theta < 1$. Quantile to-θ from F_Y, denoted as $q_Y(\theta)$ is the solution for $F_Y(q) = \theta$, that is $q_y(\theta) := F_y^{-1}(\theta) = inf\{y: F_y(y) \geq \theta\}$. So that $100\theta\%$ $(100(1-\theta)\%)$ from the period of opportunity Y is under (above) $q_Y(\theta)$.

The OLS method is used to determine classic regression coefficient [10] by inferring the number of error squares that minimize $\sum_{i=1}^{n} \varepsilon_i^2$. Estimator in the OLS method is obtained by minimizing $\sum \varepsilon_i^2$, with $\sum \varepsilon_i^2$ is the number of squares of error. As with the OLS method that minimize the number of square of error to find the alleged value for β, then in quantile regression, quantile to-θ from F_Y can be obtained by installing the following function against q.

The quantile regression analysis in the meteorological field can be applied to rainfall data, temperature, and climate change. Rainfall data is seasonal data so that at certain times there is a heavy rain. The heavy rainy incidence can be modeled using the top quantile regression analysis, especially the extreme value. The combination of any quantile values can explain the overall pattern of data, so it is beneficial to analyze certain parts of the conditional distribution [18].

According to Sinurat [19], rainfall is influenced by several factors such as temperature, air humidity, wind speed, and air pressure. This study aims to determine whether the temperature factors, air humidity, and wind speed affect rainfall in Makassar City using quantile regression Analysis.

This study will examine the form of quantile regression parameters of the quantile parameters to be used to find a quantile regulatory model of rainfall on 2019 in Makassar City based on temperature factors, air humidity, and wind speed. From the gained model will be known which factors are more influential to rainfall in Makassar City by 2019.

2. RESEARCH METHODS

This research is an applied research with a quantitative approach using quantile regression to find out the factors that influence rainfall in the city of Makassar, namely by collecting and then analyzing data using quantile regression analysis. This study was conducted for data study, namely rainfall data, temperature, humidity, and wind speed, in the daily season in 2019. The data used in this study is secondary data taken from the *Stasiun Meteorologi Maritim* Paotere, an online data center from the *Badan Meteorologi Klimatologi dan Geofisika*, Makassar City. Data analysis was carried out using the R studio software. The research method used is a combined literature study and applied. The literature is from collecting various references in the form of books, journals, and other sources related to the research conducted and data from secondary resources.

According to Sinurat [19], rainfall is influenced by several factors, including temperature, humidity, wind speed, and air pressure. The variables used in this study are divided into two, namely:

1. Response variable (Y) is the rainfall in the Makassar City.
2. Predictor variables (X) are:
   a. X_1 is the air temperature in Makassar City,
   b. X_2 is the air humidity in Makassar City,
   c. X_3 is the wind speed in Makassar City,

The steps were taken in this research as follows:

1. Looking for the parameter estimation form of the quantile regression model. As for the steps in finding the form of parameter estimation of the quantile regression model, that is

   Let's say the quantile regression equation $y_t = x_t'\beta + e_t$, with $Q_\theta(y_t|x_t) = x_t'\beta$ is a quantile to-θ $(0 < \theta < 1)$ of y with a certain value of x_t.

Minimize the absolute value of the error with weighting θ for positive errors and weighting $(1 - \theta)$ for negative errors

$$\widehat{\beta} = min_\beta\{\theta \sum_{t:y_t \geq x_t}|y_t - x_t'\beta| + (1-\theta) \sum_{t:y_t < x_t}|y_t - x_t'\beta|\}.$$

2. Collect rainfall data for the city of Makassar in 2019, along with data on the variables that affect it.
3. Modeling rainfall in the city of Makassar using quantile regression. The steps in modeling rainfall using quantile regression are:
   a. Determine the general model of quantile regression.
   b. Perform quantile regression parameter estimation.
   c. Get an estimator for each quantile.
4. Construct a confidence interval in quantile regression.
5. Test the assumptions of the quantile regression model.
6. Choose the best model.
7. Make results and discussion.
8. Make conclusions and suggestions.

3. RESULTS AND DISCUSSION

3.1. Quantile Regression Estimators

There are four variables explained in this research, those are rainfall, air humidity, air temperature, and wind speed. Relationships between those variables were analyzed using a regression model, especially quantile regression. At first glance, this paper provides a simple explanation of quantile regression

$$y_t = x_t'\beta + e_t$$

where $Q_\theta(y_t|x_t) = x_t'\beta$ is a quantile to-θ $(0 < \theta < 1)$ of y with a certain value of x_t. Estimator for β from quantile regression to-θ obtained by minimizing the absolute value of the error with a weighting θ for positive errors and weighting $(1 - \theta)$ for negative error, namely:

$$\hat{\beta} = argmin_\beta\{\theta \sum_{t;y_t \geq x_t}|y_t - x_t'\beta| + (1 - \theta)\sum_{t;y_t < x_t}|y_t - x_t'\beta|\}.$$

Or it can be written in the below equation:

$$\hat{\beta} = argmin_\beta \sum_{i=1}^T \rho_\theta u_t,$$

where $\rho_\theta(u_t) = \begin{cases} \theta_{u_t} & , jika\ u_t \geq 0 \\ (\theta - 1)_{u_t} & , jika\ u_t < 0 \end{cases}$

with $\hat{\beta}$: parameter estimator, θ: quantile index with $\theta \in (0,1)$, $\rho_\theta(u_t)$: loss function, and u_t: error of parameter estimator.

3.2. Quantile Regression Analysis Result

OLS estimation of a linear model with respect to y is obtained by minimizing the number of squares of error. Whereas the quantile regression estimation of a linear model with respect to y is obtained by minimizing the loss function value that is not symmetrical by minimizing the expected value $\rho_\theta(u)$. To find a solution, we can use the interior point method, the simplex method. **Table 1** gives the regression analysis results, and the following tables display the quantile regression analysis results.

Table 1. **Parameter Estimation Results and The Significance of the Regression Parameter**

Parameter	Coefficient	Std. Error	Significance
β_0	-6665.54	4074.71	0.103
β_1	183.06	116.39	0.117
β_2	21.98	16.35	0.180
β_3	60.28	173.85	0.729

Description: Significant on $\alpha = 0,05$, $R^2 = 0,8\%$

Table 2. Quantile Regression Parameter Estimation Results

Parameter	Quantile (θ) to-								
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
$\beta_0(\theta)$	0	0	0	1.001	2.976	5.988	2.465	-3.646	-19.870
$\beta_1(\theta)$	0	0	0	-0.263	-0.498	-0.883	-1.133	-1.736	-2.476
$\beta_2(\theta)$	0	0	0	0.097	0.169	0.297	0.468	0.868	1.487
$\beta_3(\theta)$	0	0	0	-0.061	-0.109	-0.321	-0.368	-1.389	0.727

Table 3. Parameter Significance Value

Parameter	Quantile (θ) to-	Significance Value	Description
β_0		1	Not significant
β_1		1	Not significant
β_2	0.1	1	Not significant
β_3		1	Not significant
β_0		1	Not significant
β_1		1	Not significant
β_2	0.2	1	Not significant
β_3		1	Not significant
β_0		1	Not significant
β_1		1	Not significant
β_2	0.3	1	Not significant
β_3		1	Not significant
β_0		0.67558	Not significant
β_1		0.00014	Significant
β_2	0.4	0.00000	Significant
β_3		0.54703	Not significant
β_0		0.47159	Not significant
β_1		0.00003	Significant
β_2	0.5	0.00000	Significant
β_3		0.53473	Not significant
β_0		0.39763	Not significant
β_1		0.00002	Significant
β_2	0.6	0.00000	Significant
β_3		0.28792	Not significant
β_0		0.88159	Not significant
β_1	0.7	0.01696	Significant

Parameter	Quantile (θ) to-	Significance Value	Description
β_2		0.00000	Significant
β_3		0.60196	Not significant
β_0		0.90655	Not significant
β_1		0.05098	Not significant
β_2	0.8	0.00000	Significant
β_3		0.29493	Not significant
β_0		0.85035	Not significant
β_1		0.41058	Not significant
β_2	0.9	0.00049	Significant
β_3		0.87150	Not significant

Table 4. Value of R^2 on Some Quantile

Quantile	R^2
0.1	0.0000
0.2	0.0000
0.3	0.0000
0.4	0.0001
0.5	0.0006
0.6	0.0013
0.7	0.0021
0.8	0.0028
0.9	0.0024

3.3. Discussion

Based on **Table 1**, the estimation results using the OLS method are obtained, the following model is obtained

$$Y = -6665.54 + 183.06X_1 + 21.98X_2 + 60.28X_3.$$

This model, which is explained in **Table 1**, we obtained the significance values of each parameter are 0.103, 0.117, 0.180, and 0.729, respectively. When the significant level of this model is 0.05, it means that no variables have a significant effect on rainfall in the city of Makassar. OLS method produces a coefficient of determination (R^2) of 0.0008169. This show that 0.08% percentage of rainfall in the city of Makassar can be explained by the model, while the rest is explained by the other variables outside the model.

The next step, **Table 2**, **Table 3**, and **Table 4** displayed results to model the rainfall using quantile regression. The quantile regression parameters are estimated in θ = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. From the estimation results shown in the **Table 2**, the quantile regression analysis models for each quantile are as follows:

1. $Y_{\theta(0.1)} = 0$
2. $Y_{\theta(0.2)} = 0$
3. $Y_{\theta(0.3)} = 0$
4. $Y_{\theta(0.4)} = 1.001 - 0.263\,X_1 + 0.097\,X_2 - 0.061\,X_3$
5. $Y_{\theta(0.5)} = 2.976 - 0.498\,X_1 + 0.169\,X_2 - 0.109\,X_3$
6. $Y_{\theta(0.6)} = 5.988 - 0.883\,X_1 + 0.297\,X_2 - 0.321\,X_3$
7. $Y_{\theta(0.7)} = 2.465 - 1.133\,X_1 + 0.468\,X_2 - 0.368\,X_3$
8. $Y_{\theta(0.8)} = -3.646 - 1.736\,X_1 + 0.868\,X_2 - 1.389\,X_3$
9. $Y_{\theta(0.9)} = -19.870 - 2.476\,X_1 + 1.487\,X_2 + 0.727\,X_3$

From the several quantiles shown, there are several independent variables that have a positive and negative effect on the dependent variable.

The next step is testing the quantile regression parameters, the results of which are shown in **Table 3**. Based on the results shown in the table, it can be seen that in quantiles 0.1, 0.2, and 0.3 there are no variables that have a significant effect on the level of rainfall. In quantiles 0.4, 0.5, 0.6, and 0.7 there are two predictor variables that affect the level of rainfall, namely the temperature variable and the humidity variable. In quantiles 0.8 and 0.9, only one predictor variable that affects the level of rainfall, namely the variable humidity. Because there is at least one significant parameter for each predictor variable, the variable has an effect on the response variable [20]. So that the model can be analyzed further and can be used as a quantile regression model.

Based on the result of the analysis obtained 9 (nine) models in each quantile used, so it is necessary to select the best model, which is selected based on the value of the coefficient of determination. The coefficient of determination for each model is shown in **Table 4**, where the largest coefficient of determination is the model in quantile 0.8, which is 0.28%. Larger than the other quantiles. So the model obtained is as follows:

$$Y_{\theta(0.8)} = -3.646 - 1.736X_1 + 0.868X_2 - 1.389X_3.$$

Then the MAE calculation is carried out to find out whether the model obtained is good for use in predictions.

$$
\begin{aligned}
\text{MAE} &= \frac{\sum_{i=1}^{n}|y_i - \hat{y_i}|}{n} \\
&= \frac{|8.5-21.3|+|2.3-9.5|+|0.4-8.7|+\cdots+|0-17.7|}{344} \\
&= \frac{109240.1}{344} \\
&= 317.56
\end{aligned}
$$

Based on the r-value and the value obtained from the MAE calculation, it is known that the model obtained is not very good for prediction. If the model is intended for estimation, it is necessary to overcome the existing heteroscedasticity problem beforehand if the data are not normally distributed or if the model has a heteroscedasticity problem from the model obtained.

4. CONCLUSIONS

Based on the result of the analysis and discussion, some conclusions can be drawn as follows:

1) Estimator for β from quantile regression to-θ obtained by minimizing the absolute value of the error with weighting θ for positive errors and weighting $(1 - \theta)$ for negative errors, namely:

$$\hat{\beta} = argmin_\beta\{\theta \sum_{t;y_t \geq x_t}|y_t - x_t'\beta| + (1-\theta)\sum_{t;y_t < x_t}|y_t - x_t'\beta|\}.$$

To find a solution to this equation, the interior point method and the simplex method can be used.

2) The rainfall quantile regression analysis model in the city of Makassar in 2019 is:

$$Y_{\theta(0.8)} = -3.646 - 1.736\,X_1 + 0.868\,X_2 - 1.389\,X_3.$$

With rainfall data for the city of Makassar as the response variable (Y) while air temperature (X_1), humidity (X_2), and wind speed (X_3) in the city of Makassar as the predictor variables (X).

3) From the quantile regression analysis model obtained, the predictor variable that has a significant effect on the level of rainfall in the city of Makassar is the air humidity variable, while the temperature and wind speed variables have no significant effect on the level of rainfall in the city of Makassar in 2019.

REFERENCES

[1]   Sukarna, E. Y. P. Ananda, and M. S. Wahyuni, "Rainfall Forecasting Model Using ARIMA and Kalman Filter in Makassar, Indonesia," *J. Phys. Conf. Ser.*, vol. 2123, no. 1, pp. 0–7, 2021, doi: 10.1088/1742-6596/2123/1/012044.

[2]   S. A. Thamrin, Aswi, Ansariadi, A. K. Jaya, and K. Mengersen, "Bayesian spatial survival modelling for dengue fever in Makassar, Indonesia," *Gac. Sanit.*, vol. 35, no. S1, pp. S59–S63, 2021, doi: 10.1016/j.gaceta.2020.12.017.

[3]   A. Aswi, S. Cramb, W. Hu, G. White, and K. Mengersen, "Spatio-Temporal Analysis of Dengue Fever in Makassar Indonesia: A Comparison of Models Based on CARBayes," in *Case Studies in Applied Bayesian Data Science, Lecture Notes in Mathematics*, vol. 2259, Switzerland: Springer Nature Switzerland AG, 2020, pp. 229–244.

[4]   Paida, "Deteksi Perubahan Pola Curah Hujan Kota Makassar," Universitas Hasanuddin, 2011.

[5]   C. R. Malino, M. Arsyad, and P. Palloan, "Analisis Parameter Curah Hujan dan Suhu Udara di Kota Makassar Terkait Fenomena Perubahan Iklim," *J. Sains dan Pendidik. Fis.*, vol. 17, no. 2, pp. 139–145, 2021, [Online]. Available: http://ojs.unm.ac.id/jsdpf.

[6]   F. Mazid, Wasilah, and Alfiah, "Pusat Pelatihan Bulutangkis dengan Pendekatan Arsitektur Bioklimatik di Kota Makassar," *TIMPALAJA Archit. Student Journals*, vol. 3, no. 2, pp. 143–152, 2021, [Online]. Available: https://journal.uin-alauddin.ac.id/index.php/timpalaja/article/view/18635/13110.

[7]   R. Maru *et al.*, "Analysis of The Heat Island Phenomenon in Makassar, South Sulawesi, Indonesia," *Am. J. Appl. Sci.*, vol. 12, no. 11, 2015, doi: 10.3844/ajassp.2015.

[8]   W. Sanusi, "Analisis Homogenitas Data Curah Hujan Tahunan Kota Makassar," *J. Sci. Pinisi*, vol. 2, no. 2, pp. 137–142, 2016.

[9]   S. Maulidani S, N. Ihsan, and Sulistiawaty, "Analisis Pola dan Intensitas Curah Hujan Berdasakan Data Observasi dan Satelit Tropical Rainfall Measuring Missions (TRMM) 3B42 V7 di Makassar," *J. Sains Dan Pendidik. Fis.*, vol. 11, no. 1, pp. 98–103, 2015.

[10]   D. N. Gujarati and D. C. Porter, *Basic Econometrics*, Fifth edit. New York: Douglas Reiner, 2009.

[11]   N. Fadhilah, E. Erfiani, and I. Indahwati, "Comparison of Functional Regression and Functional Principal Component Regression for Estimating Non-Invasive Blood Glucose Level," *Indones. J. Stat. Its Appl.*, vol. 5, no. 1, pp. 14–25, 2021, doi: 10.29244/ijsa.v5i1p14-25.

[12]   Saidah, F. Yanuar, and D. Devianto, "Analisis Regresi Kuantil," *J. Mat. UNAND*, vol. 5, no. 1, p. 103, 2016, doi: 10.25077/jmu.5.1.103-107.2016.

[13]   V. E. Wahyudi and I. Zain, *"Analisis IPM di Pulau Jawa Menggunakan Analisis Regresi Kuantil 1 1,2,"* vol. 2, no. 1, 2014.

[14]   R. Rahmawati, Widiarti, and P. Novianti, "Regresi Kuantil (Studi Kasus pada Data Suhu Harian)," pp. 1–8, 2011.

[15]   A. Djuraidah and A. H. Wigena, "Regresi Kuantil untuk Eksplorasi Pola Curah Hujan di Kabupaten Indramayu Regression for Exploring Rainfall Pattern in Indramayu Regency," *J. ILMU DASAR*, vol. 12, pp. 50–56, 2009.

[16]   Cintiani, *"Pemodelan Regresi Kuantil (Studi Kasus Pengeluaran Rumah Tangga untuk Konsumsi Rokok),"* Institut Teknologi Sepuluh November, 2017.

[17]   Harinaldi, *Prinsip-Prinsip Statistik untuk Teknik dan Sains*. Jakarta: Erlangga, 2005.

[18]   N. Astari, Amran, and A. K. Jaya, *"Pengujian Hipotesis pada Regresi Kuantil,"* Universitas Hasanuddin, 2017.

[19]   N. Sinurat, Sugianto, and W. Harjupa, *"Analisa Arah Angin menggunakan Equatorial Atmosphere Radar (EAR) dan Optical Rain Gauge (ORG) di Atas Kototabang Sumatera Barat,"* Universitas Riau, 2016.

[20]   R. S. Sari and I. N. Budiantara, *"Pemodelan Pengangguran Terbuka di Jawa Timur dengan Menggunakan Pendekatan Regresi Spline Multivariabel,"* *J. Sains Dan Seni Its*, vol. 1, no. 1, pp. 236–241, 2012.