*IOSR Journal of Mathematics (IOSR-JM)*
*e-ISSN: 2278-5728, p-ISSN:2319-765X. Volume 10, Issue 4 Ver. I (Jul-Aug. 2014), PP 49-54*
*www.iosrjournals.org*

# Circular Circular – Linear Regression Analysis of Order m in Circular Variable α and β against Linear Variable (Y)

M. Irpan Nurhab[1], Anang Kurnia[1], and I Made Sumertajaya[1]
[1]*Department of Statistics, Bogor Agricultural University, Indonesia*

***Abstract:*** *The development of data analysis is still predominanlty use linear statistics. Whereas the research world there are other types of data is data direction. One type of data direction is the data circular. Statistical analysis aimed at modeling the causal relationship between the independent variable and the dependent variable is regression analysis. So as to model the relationship between wind direction and cloud direction against rainfall is circular circular – linear multiple regression analysis. The purpose of this research it to build a model circular circular – linear regression analysis of order m in circular variable α and β against linear variable (Y). Data used in this research is the simulation data and secondary data obtained from the Meteorology, Climatology, and Geophysics in Bogor a cityof West Java, Indonesia. The data is the result of observations of wind direction, coulds direction, and rainfall in february 2014 and march 2014. From the analysis of the data showed that the best model to see the effect of wind direction and could direction against rainfall is circular circular – linear regression analysis to the order four better of the linear multiple regression analysis. It is seen from the sum square of error, p-value, and r-square.*
***Keywords:*** *data circular, circular circular – linear regression, rainfall.*

## I. Introduction

The development of data analysis is still predominanlty use linear statistics.whereas the research world there are other types of data is data direction. One type of data direction is the data circular, namely the measured data in the form of a angular or two-dimensional oriented unit time or degree direction. Data circular are met in nearly all branches of science, such as Biology, Geography, Geology-Geophysics, Medicine, Meteorology, Marine, and others. Some illustrations of data circular, namely the direction of come and go towards the birds, the wind direction and the direction of movement of the clouds, the time of arrival of patients (24 hours) in the emergency room in a house hospital, the number of occurrences in one year or in monthly, the time pattern crime within the daily and weekly, and others.

Circular data type of the data is direction, therefore less precise types of data circular were analyzed with statistics linear, so the data needs to be analyzed using circular statistics. One example of errors that can occur when the data circular were analyzed using statistical linear namely the calculation of the average direction.

Statistical analysis that aims to model the causal relationship between the independent variables with the dependent variable is regression analysis. This study lifting rainfall case. There are several factors that influence rainfall, such as wind direction and cloud direction. Wind direction and cloud direction is a kind of data circular, while rainfall is linear data, so as to model the relationship between wind direction and cloud direction with rainfall is circular - linear regression analysis. Since there are two independent variables circular and linear single dependent variable in this study used the term circular circular – linear regression analysis.
The purpose of this research it to build a model circular circular – linier regression analysis of order m in circular variable α and β against linear variable (Y).

## II. Literature Review

### 2.1 Data and Statistics Circular

Circular data is the measurement data values are repeated periodically. A return value will be found after seeing a period / full turn. The definition itself is circular variable characteristics of the data at the beginning and end of the measurement scale to meet each other [1]. The data type of circular divided into two types of data circular direction and the data type of the time [2].

A circular statistical distribution models and statistical techniques to analyze random variables that form cycles in nature. Circular statistical used in data measurement results in the form of direction and is usually expressed in angular size.

### 2.2 Circular Regression

Circular regression equation for the data is divided into three types, namely:
1. Circular-Linear Regression

Independentvariable is a variable circular α and dependent variable is α linear variable.
2. Circular-Linear Regression
Independentvariable is a linear variable and dependent variable is the variable circular α.
3. Circular-Circular Regression
Independentvariable is a circular variable and dependent variable α is α circular variable β [3].

**2.3Graphical representation of Circular Data**
An easily analyzed if the data can be depicted in a graph [4].

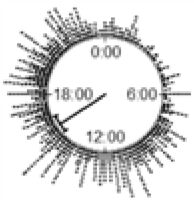


**Figure2 Transmit diagram**

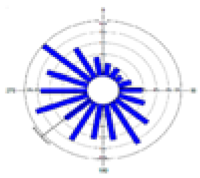

**Figure3 Cyclic histogram**

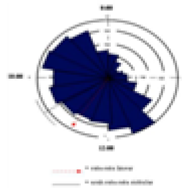

**Figure4 Rose diagram**

**2.4 Circular Descriptive Statistics**

To analyze the data circular there are two trigonometric functions are used as basis ie sines and cosines. Both basic trigonometric functions are used to help determine the position of the data. Both of these functions are used to align the two coordinate systems. Position that a direction may be determined by the polar coordinates or Cartesian coordinates. In Cartesian coordinates of point P expressed as (X, Y) or as the value of (r, θ) in polar coordinates where r is the distance of point P from the center point O [5]. polar coordinates can be converted into Cartesian coordinates using the following trigonometric equation:

$$x = r\cos\theta, y = r\sin\theta$$

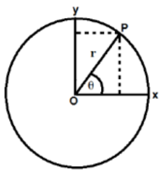


**Figure 5** The relationship between Cartesian coordinates and polar coordinates

In the analysis of circular note is the direction and not the magnitude of the vector, so it was taken to ease these vectors into unit vectors, ie vectors that have unit length with r = 1. Every direction associated with a point P on the circumference of a circle. In contrast, this point in the circumference of a circle can be expressed as an angle. If the point P lies in the circumference of a circle, and the change in polar coordinates and Cartesian coordinates

$$(1,\alpha) \Leftrightarrow (x = \cos\alpha, y = \sin\alpha) \qquad (2.1)$$

**2.5 Average direction**

The average direction of the circular sample data obtained by calculating the resultant vector of vectors of each sample unit. Direction of resultant vectors expressed towards the average of the sample data and the average length of the resultant concentration of each sample was expressed to the direction of the average. Suppose there samples $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \dots, \boldsymbol{\alpha_n}$ with n observations circular stated in the angular. Known transformation polar coordinates to Cartesian coordinates of each observation as follows,

$$(1,\alpha_i) \Leftrightarrow (x = \cos\alpha_i, y = \sin\alpha_i), i = 1,2,\dots,n$$

Retrieved resultant vector of the unit vectors by summing each component, ie

$$\mathbf{R} = \left( \sum_{i=1}^{n} \cos\alpha_i, \sum_{i=1}^{n} \sin\alpha_i \right) = (\mathbf{C}, \mathbf{S})$$

with,

$$R = \|R\| = \sqrt{C^2 + S^2} \ ,0 \le \mathrm{R} \le \mathrm{n} \text{ and } \overline{\mathrm{R}} = \frac{\mathrm{R}}{\mathrm{n}} ,0 \le \mathrm{R} \le 1$$

$\overline{\mathrm{R}}$ states the average length of the resultant vector and also shows the size of concentration of the data against the direction of the average.

Direction of the resultant vector **R** is the average direction of the circular denoted by $\bar{\alpha}$ and defined,

$$\cos\bar{\alpha}_0 = \frac{C}{R},\ \sin\bar{\alpha}_0 = \frac{S}{R},\ \text{so}\ \bar{\alpha}_0 = \arctan^* \frac{S}{C}$$

If all angles stated point in the same direction, it can be indicated that the data is concentrated andR closer to the value of n. Conversely, if the data can be spread throughout circles indicated that the data is not concentrated and R approaches the value 0.

**2.6 Variety Circular**

Defining circular sample variance is V = 1 -. $\bar{R}$ . The smaller the value the more concentrated circular variance data to a certain point. The value of V is in the interval [0,1] [6].

**2.7 Circular Circular-Linear Regression**

Circular circular - linear regression models between linear variable X with two independent variables circular α can be written [5]:

$$E(X) = A_0 + A_1\cos(\alpha_1 - \alpha_{01}) + A_2\cos(\alpha_2 - \alpha_{02}) \tag{2.2}$$

This equation can be described,

Suppose $B_{k1} = A_k\cos\alpha_{0k}$ dan $B_{k2} = A_k\cos\alpha_{0k}$ the form of circular circular linear regression model can be written:

$$Y = A_0 + B_{11}\cos\alpha_1 + B_{12}\sin\alpha_1 + B_{21}\cos\alpha_2 + B_{22}\sin\alpha_2$$

**2.8 Estimasasi Regression Coefficients**

regression coefficient $A_0, B_{11}, B_{12}, B_{21}, B_{22}$ will be estimated using the Least Squares Method. Least squares method choose parameter values such that the minimum value of SSE.. The solution of the system of equations is estimated least squares, ie $\hat{A}_0, \hat{B}_{11}, \hat{B}_{12}, \cdots, \hat{B}_{p1}, \hat{B}_{p2}$.

Model Circular Circular-Linear Regression when written in matrix form is

$$Y = Z\beta + \varepsilon \tag{2.3}$$

With

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix};\ \beta = \begin{bmatrix} A_0 \\ B_{11} \\ B_{12} \\ \vdots \\ B_{1m} \\ B_{21} \\ B_{22} \\ \vdots \\ B_{2m} \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} 1 & \cos\alpha_{11} & \sin\alpha_{11} & \cdots & \cos m\alpha_{11} & \sin m\alpha_{11} & \cos\alpha_{12} & \sin\alpha_{12} & \cdots & \cos m\alpha_{12} & \sin m\alpha_{12} \\ 1 & \cos\alpha_{21} & \sin\alpha_{21} & \cdots & \cos m\alpha_{21} & \sin m\alpha_{21} & \cos\alpha_{22} & \sin\alpha_{22} & \cdots & \cos m\alpha_{22} & \sin m\alpha_{22} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos\alpha_{n1} & \sin\alpha_{n1} & \cdots & \cos m\alpha_{n1} & \sin m\alpha_{n1} & \cos\alpha_{n2} & \sin\alpha_{n2} & \cdots & \cos m\alpha_{n2} & \sin m\alpha_{n2} \end{bmatrix}$$

Where,
**Y** = observation vector size (nx1)
**Z** = matrix size (nx(1+4m))
**β** = vector of regression coefficients measuring ((1+4m)x1)
**ε** = random error vector size (nx1)

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Then look for the least squares estimation vector $\hat{\beta}$ that minimizes the quadratic error function L.

$$\mathbf{L} = \sum \varepsilon_i^2 = \varepsilon'\varepsilon = (X - Z\beta)(X - Z\beta) = X'X - 2\beta'Z'X + \beta'Z'Z\beta \tag{2.4}$$

So that the vector alleged**β** is

$$\hat{\beta} = (Z'Z)^{-1}Z'X \tag{2.5}$$

**2.9Reduction in Sum of Squares Error (SSE)**

The most important thing in determining the order of the regression polynomial is reducing SSE when m increases. The decision was taken to trigonometric polynomial order-(m +1) by adding columns. With equations Sum of Squares Error,

$$SSE = \mathbf{X'X} - \hat{\beta}'\mathbf{Z'X} \tag{2.6}$$

In determining whether the taking rank (m +1), we first calculate the reduction SSE like (2.6), if the reduction was significantly greater then we decided to enter the order of (m +1).

## III. Methodology

### 3.1. Data

The data used in this study is a secondary data obtained from the Meteorology, Climatology, and Geophysics in Bogor, the data are observations of wind direction, cloud direction, and rainfall in the month of February 2014 and March 2014.

### 3.2.Method of Analysis

Step 1: **Creating a descriptive analysis of circular statistics for each variable wind direction (α) and cloud direction (β).**

a. Circular graphic representation of data for each variable wind direction (α) and cloud direction (β) with Circular Plot and Rose Diagram.

b. The average direction of circular and linear for each variable wind direction (α) and cloud direction (β)

$$\bar{\alpha} = \begin{cases} \arctan(S/C) & \text{if } C > 0, S \geq 0 \\ \pi/2 & \text{if } C = 0, S > 0 \\ \arctan(S/C) + \pi & \text{if } C < 0 \\ \arctan(S/C) + 2\pi & \text{if } C \geq 0, S < 0 \\ \text{undefined} & \text{if } C = 0, S = 0 \end{cases}$$

c. The length of the average vector for each circular variable wind direction (α) and cloud direction (β).

$$\bar{R} = \frac{R}{n}, \text{ where } R = \sqrt{C^2 + S^2}$$

d. Variety of circular and linear for each variable wind direction (α) and cloud direction (β).

V = 1 - $\bar{R}$ , with inteval V is at [0,1]

Step 2: **Circular circular - linear regression analysis to the wind direction (α) and cloud direction (β) as the independent variables on rainfall (Y) as the dependent variable.**

$$Y = A_0 + B_{11}\cos\alpha_1 + B_{12}\sin\alpha_1 + B_{21}\cos\alpha_2 + B_{22}\sin\alpha_2$$

Step 3: **Determination of the orderm of circular circular - linear regression polynomial.**

$$SSE = \mathbf{X'X} - \hat{\beta}'\mathbf{Z'X}$$

## IV. Results And Discussion

### 4.1. Descriptive Statistics Wind Direction (α) and Direction Cover (β) in the month February 2014 and March 2014 in Bogor.

**Table 1** Descriptive statistics Wind Direction (α) and Cloud Direction (β) in the month of February 2014 and March 2014 in Bogor

| **Variable** | Wind Direction (α) | Cloud Direction **(β)** |
|---|---|---|
| Number of observations | 58 | 58 |
| The average direction of the circular | $332{,}56^0$ | $313{,}70^0$ |
| The average direction of the linear | $284{,}19^0$ | $304{,}89^0$ |
| The length of the resultant | 51,76 | 48,47 |
| The average length of the resultant | 0,89 | 0,83 |
| Variety circular | 0,11 | 0,16 |
| Variety linear | 11539,53 | 2906,91 |

#### 4.2. Graphic representation of variables Wind Direction (α) and Variable Direction Cover (β)

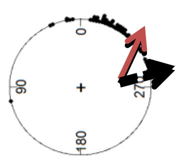


**Figure 6** Circular Plotof wind direction (α)

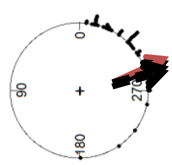


**Figure 7** Circular Plot of cloud diagram (β)

Figure 6Circular Plot visible red straight line is the direction of the circular average wind direction (α) of 332.560 mean wind direction (α) with circular statistics have a tendency toward the northwest and The dashed straight black is towards linear statistical average of the wind direction (α) of 284.190 means, wind direction (α) with linear statistics have a tendency toward the west. This suggests differences in the calculation of the average direction of data with circular statistics on the distribution of the data, while the linear statistics are slightly away from the distribution of the data. Figure 7Circular Plot visible red straight line is the direction of the circular average cloud direction (β) of 313.700 and The dashed straight black is towards linear statistical average cloud direction (β) of 304.890, mean cloud direction (β) with circular statistics and linear statistics have a tendency toward the northwest.

#### 4.3. Multiple Linear Regression and Regression Circular Circular - Linear To See Effects of Wind Direction (α) and Direction of Clouds (β) of the Rain (Y) in The MonthFebruary 2014 and March 2014 in Bogor

**Table 2** Multiple Linear Regression and Regression Circular Circular - Linear To See Effects of Wind Direction (α) and Direction of Clouds (β) of the Rain (Y) In February 2014 and March 2014 in Bogor

| TYPE | SSE | $R^2$ | P-Value |
|---|---|---|---|
| Multiple Linear Regression | 10244,8 | 1,7% | 0,630 |
| | | | |
| Circular Circular - LinearRegressionOrder 1 | 9480,5 | 9% | 0,278 |
| Circular Circular - LinearRegressionOrder 2 | 9141,6 | 12,3% | 0,560 |
| Circular Circular - LinearRegressionOrder 3 | 8094,1 | 22,3% | 0,401 |
| Circular Circular - LinearRegressionOrder 4 | 5423,9 | 47,9% | 0,014 |
| Circular Circular - LinearRegressionOrder 5 | 5290,9 | 49% | 0,061 |

In Table 2, the value of the coefficient of determination in multiple linear regression of 1.7%, ie by 1.7% variability of rainfall (Y) can be explained by wind direction (α) and cloud direction (β) in a linear relationship, and the rest influenced by other factors. While the value of the regression coefficient of determination circular circular - linear in Table 2 by 9% to 49% and a order 1 to order 5, meaning that 9% more variability in rainfall (Y) can be explained by wind direction (α) and cloud direction (β), and the rest is influenced by other factors. Seen that the circular circular – linear regression have much better results than linear regression to see the effect of wind direction (α) and cloud direction (β) againstrainfall (Y)

P-value in the multiple linear regression was 0.630, so the error rate α = 0.1, p-value 0.630> 0.1 (α). It can be interpreted, multiple linear regression model is not significant to see the effect of wind direction (α) and cloud direction (β) against the average rainfall (Y) with a confidence level of 90%. In the circular circular - linear regression, with an error rate α = 0.1, so that the circular circular – linear regression models, with a p-value 0.014 <0.1 (α), it can be interpreted, circular circular regression models - linear rank 4 very significant to see the effect of wind direction (α) and cloud direction (β) against the average rainfall (Y) with a confidence level of 90%.

To determine the best model, the first is a reduction in the value of SSE, SSEorder 4 - JKG order 5 = 5423.9 - 5290.9 = 133, indicating that the reduction of SSE very small, so that the circular circular - linear regression models order of four more both of the rank of five. So the best model to see the effect of wind direction (α) and cloud direction (β)against rainfall (Y) is $\hat{Y}_i$ = **- 16975 + 31092 cos $\alpha_i$ - 1516 sin $\alpha_i$ - 256 cos $\beta_i$ + 421 sin $\beta_i$ - 20433 cos 2$\alpha_i$ + 4283 sin 2$\alpha_i$ - 236 cos 2$\beta_i$ - 346 sin 2$\beta_i$ + 7738 cos 3$\alpha_i$ - 3940 sin 3$\alpha_i$ + 259 cos 3$\beta_i$ - 41 sin 3$\beta_i$ - 1149 cos 4$\alpha_i$ + 1192 sin 4$\alpha_i$ - 32,6 cos 4$\beta_i$ + 74,3 sin 4$\beta_i$**

## V. Conclusion

By looking at the value of Sum of Squares Error (SSE) between circular statistics and linear statistics are then SSE 9480.5 circular statistics are smaller than the linear statistical SSE 10244.8 indicates that circular models are better than linear models seen from SSE. $R^2$ for a much larger circular statistics is 47% compared to the linear statistics is only 1.7%, meaning that circular models better than the linear model of $R^2$. P-value for the circular statistics show a very significant level is 0.014 with a confidence level of 90%, whereas the linear statistical p-value 0.630 only means of linear statistical models is not very significant. Concluded the best model to see the effect of wind direction (α) and cloud direction (β) against rainfall (Y) in the month of February 2014 and March 2014 in Bogoris circular circular - linear regression with the order of four, where the model $\hat{Y}_i$ = **- 16975 + 31092 cos $\alpha_i$ - 1516 sin $\alpha_i$ - 256 cos $\beta_i$ + 421 sin $\beta_i$ - 20433 cos 2$\alpha_i$ + 4283 sin 2$\alpha_i$ - 236 cos 2$\beta_i$ - 346 sin 2$\beta_i$ + 7738 cos 3$\alpha_i$ - 3940 sin 3$\alpha_i$ + 259 cos 3$\beta_i$ - 41 sin 3$\beta_i$ - 1149 cos 4$\alpha_i$ + 1192 sin 4$\alpha_i$ - 32,6 cos 4$\beta_i$ + 74,3 sin 4$\beta_i$**

## References

[1]. Martin GK. 2008. Circular Statistics. Article Alley. http://www.articlealley.com/ circular-statistics-657388.html [17 February 2014].
[2]. Mardia, K.V. dan Jupp, P.E. 2000. Directional Statistics. New York: John Wiley & Sons.
[3]. Jammalamadaka S.R, Sarma Y.R. 1988. A Correlation Coefficient for Angular Variables. In Matusita, K. editor, Statistical Theory and Data Analysis II, pages 349—364. North Holland, Amsterdam.
[4]. Fisher N.I. 1993. Statistical Analysis of Circular Data. Cambridge: Cambridge University Press.
[5]. Jammalamadaka, S.R. dan A. SenGupta. 2001. Topics in circular Statistics. London: World Scientifics Publishing.
[6]. Mardia KV. 1976. Linier-Circular Correlation Coefficients and rythmometry. Biometrika, 63, 403-405.