REGRESSION MODELING IN RICE PRODUCTION AND DOMESTIC REGIONAL BRUTO PRODUCTS LAMPUNG PROVINCE

Reni Permata Sari

Nahdlatul Ulama Lampung University

renipermatasari71@gmail.com

ABSTACT

The agricultural sector is one sector that has an important role in the development of Lampung Province. Rice-specific food crops are a strategic commodity because they are the main foodstuff of the Indonesian people, and the Lampung community in particular. The need for an increase in the productivity of rice plants in Lampung is an important note for the Lampung Province government. One of the goals of agricultural development is to improve the welfare of farmers and improve food welfare. Considering rice commodity in supporting development in Lampung, it is necessary to study the relationship of rice production to the Gross Regional Domestic Product of the agricultural sector in Lampung Province. Results of research conducted on the Gross Regional Domestic Product of the agricultural sector in Lampung Province, namely: (1) Gross Regional Domestic Product of Lampung Province in the agriculture, agriculture, agriculture, and agriculture sectors related to a very strong positive linear relationship with the amount of production rice in Lampung Province, (2) The higher the amount of rice production in Lampung Province, the higher the Regional Gross Domestic Product of Lampung Province in the agriculture, animal husbandry, mining, and fishery sectors, and (3) Polynomial Model with a degree of Y = - 1521 + 185 x - 6.88 x2 + 0.0860 x3 with Y as the Gross Regional Domestic Product and X as the amount of rice production in Lampung Province is the best model that can explain the relationship between each.

Keywords: Rice Production, Gross Regional Domestic Product, Polynomial Model degree three

1. INTRODUCTION

1.1 Background of the Problem

Lampung Province is one of Indonesia's rice barns. The agricultural sector is one sector that has an important role in the economy of Lampung Province. Food crops, especially rice, are a strategic commodity because they are the staple food of the Indonesian people in general, and the Lampung community in particular. One of the goals of agricultural development is to improve the welfare of farmers and create food security. Not only agricultural science that must be socialized

to farmers, the government should also make efforts to analyze agriculture. Considering the importance of rice commodity in supporting development in Lampung, it is necessary to study the relationship of rice production to the Gross Regional Domestic Product (GRDP) of the agricultural sector in Lampung Province.

1.2 THEORY BASIS

1.2.1 Data Analysis Techniques

The techniques that are widely used for data analysis are:

1. Box plot

To make a boxplot only five statistics are needed, namely: minimum value, Q1 (first quartile), median, Q3 (third quartile), and maximum value (Lind, Douglas, 2009).

2. Bar Chart

A bar chart is a graph that displays classes on the horizontal axis and class frequencies on the vertical axis.

3. Scatter plot

A type of plot or diagram that uses Cartesian coordinates to display values, usually two variables for a data set.

4. Graph

Data graph is a visual presentation of data in the form of pictures, graphs or diagrams that usually come from tables. Histograms and frequency polygons are common ways of presenting a number of information derived from quantitative variables (Steel and Torrie, 1995).

1.2.2 Statistics Inference in Research

According to Yanto (2016), inference statistics relate to a series of techniques for studying, estimating, and drawing conclusions about population parameters through a portion of the data (sample data) selected with a particular technique from that population.

1.2.3 Descriptive Statistics

According to Walpole (1988), descriptive statistics are methods related to the collection and presentation of a group of data so as to provide useful information.

1.2.4 Regression Analysis

Regression analysis itself has an extended discussion that is, simple regression analysis, multiple regression analysis, and exponential regression analysis and principal component regression analysis. (Usman, 2009).

Simple linear regression is a regression analysis with only one predictor variable (X). While multiple linear regression is an analysis with more than one predictor variable (X) (Hendradi, 2006).

1.2.5 Correlation

The correlation coefficient (r_{xy}) is a statistic that is often used mainly in social science research that not only provides how two random variables are associated in a sample, but also has the nature of a close relationship with simple regression analysis (Cochran, 1997).

With this standardization, it is $-1 \leq r_{xy} \leq 1$ so that it is easy to interpret. If the linear relationship between X and Y is perfect then $r_{xy} = \pm 1$; +1 if the relationship is unidirectional and -1 is in the opposite direction. There is no linear relationship between X and Y marked rxy = 0 (Sembiring, 1995).

The hypothetical correlation coefficient is as follows:

i. $H_0 : r_{ij} = 0$ (there is no correlation between the dependent and independent variables)
ii. $H_1 : r_{ij} \neq 0$ (there is a correlation between the dependent and independent variables)
i. $H_0 : r_{ij} = 0$ (there is no correlation between the dependent and independent variables)
ii. $H_1 : r_{ij} > 0$ (there is a correlation between the dependent and independent variables)
i. $H_0 : r_{ij} = 0$ (there is no correlation between the dependent and independent variables)
ii. $H_1 : r_{ij} < 0$ (there is a correlation between the dependent and independent variables) (Sulaiman, 2004).

1.2.6 Coefficient of Determination

The coefficient of determination states the total amount of variation in the value of sample Y that has been explained by the predictor variables in the regression equation obtained through the least squares method. The number of regression squares, the variations explained, must increase at least if a new term is added to the model (Usman, 2009).

1.3 Research Analysis Methods

The steps of conducting research and data analysis are as follows:

1. Collecting data.

2. Creating a Scatterplot.

3. Conduct Linearity Tests.

4. Conduct Linear Regression Test.

5. Perform Polynomial Regression Test.

6. Selection of the Best Model.

1.4 Research Data

The data used in this study can be presented in the following table:

Table 1. Gross Regional Domestic Product and Rice Production Data in Lampung Province in 2000-2013

Tahun	PDRB (Rupiah)	Padi (Ton)
2000	10.388.765,00	1.946.406
2001	10.727.709,00	1.992.726
2002	10.871.433,00	1.951.109
2003	11.318.866,00	1.966.293
2004	11.951.916,00	2.091.996
2005	12.420.764,00	2.124.144
2006	13.184.537,31	2.129.914
2007	13.912.096,62	2.308.404
2008	14.327.563,43	2.341.075
2009	14.693.881,00	2.673.844
2010	14.851.400,00	2.807.676
2011	15.587.581,00	2.940.795
2012	16.242.780,00	3.101.455
2013	16.884.406,00	3.207.002

2.1 RESULTS AND DISCUSSION

Gross Regional Domestic Product data based on constant prices by business in the agriculture, animal husbandry, forestry and fisheries sectors as dependent variables (Y) in Rupiah and Rice Production in Lampung Province from 2000 to 2013 as free variables (X) in tons.

2.1.1 Scatterplot of Gross Regional Domestic Product against Rice Production

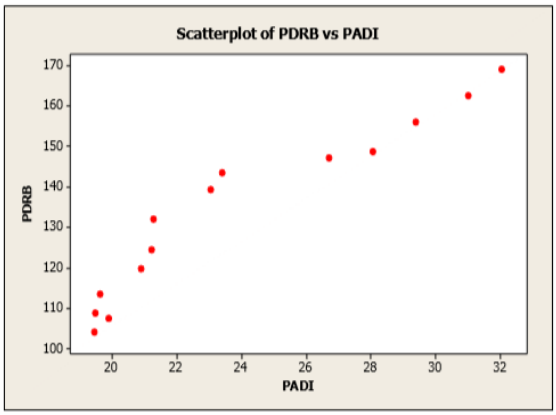


Figure 1. Gross Regional Domestic Product Scatterplot of Rice Production

From Figure 1, the scatter and points almost form a linear line. The linear relationship between the two variables above is a positive linear relationship. This means that if the value of variable X goes up, the value of variable Y will also go up, and vice versa if the value of variable X goes down, the value of variable Y will also go down.

2.1.2 Linearity Test

The linearity test result can be presented as follows:

```
Correlations: PDRB; PADI

Pearson correlation of PDRB and PADI = 0.952
P-Value = 0.000
```

Figure 2. Linearity Test Results

Pearson Correlation (r) value of 0.952. Because the value is positive, the linear relationship of the two variables is positive. This means that if the value of variable X goes up, the value of variable Y will also go up, and vice versa if the value of variable X goes down, the value of variable Y will also go down.

- Hypothesis Test:

• H_0: ($\rho = 0$)

There is no linear relationship between rice production in Lampung province in 2000 to 2013 and Lampung Province Gross Regional Domestic Product.

• H_1: ($\rho \neq 0$)

There is a linear relationship between rice production in Lampung Province in 2000 to 2013 with the Lampung Province Gross Regional Domestic Product.

- Significance level (α) = 5% = 0.05

Decision: Because the P-Value $<\alpha$ or 0,000 <0.005 then H_0 is rejected.

It can be concluded that there is a linear relationship between the results of rice production with Gross Regional Domestic Product in Lampung province. The linearity assumption is fulfilled.

2.1.3 Linear Regression

- Normal Probability Plot

You can see the Normal Probability Plot form as follows:

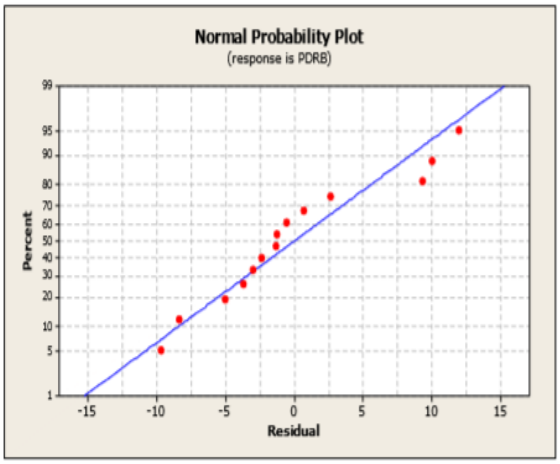


Figure 3. Normal Probability Plot in regression analysis forGross Regional Domestic Product data

Based on the normal probability graph, it appears that the points are around the estimation line. This means that Gross Regional Domestic Product data and rice production in Lampung Province between 2000 and 2013 were normally distributed so that the first assumptions were met.

- Resid The residual versus fits plot

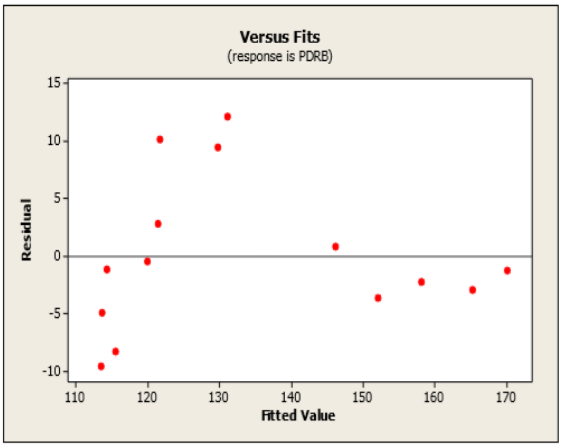


Figure 4. Plots of residual versus fits in the regression analysis for Gross Regional Domestic Product data

Based on the residual versus fits plot, it can be seen that the data is spread out and does not form a pattern like a funnel or something else. This means that the variety of data is homogeneous so that the second assumption is fulfilled.

- Resid Residual plot versus order

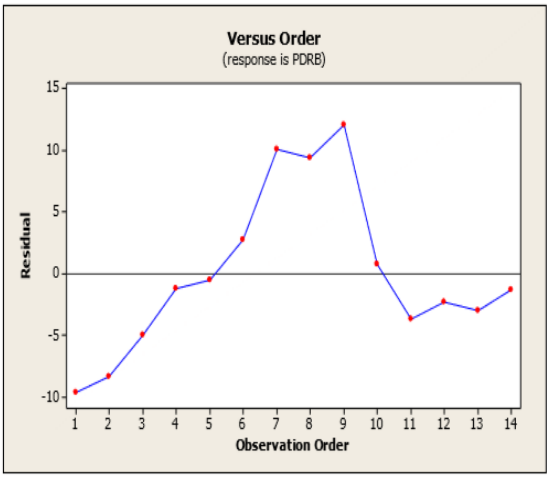


Figure 5. Plots of residual versus order in the regression analysis

for Gross Regional Domestic Product data

Based on the residual versus order plot, it appears that the data does not spread and form a pattern which means the data is not independent so the third assumption is not fulfilled.

The results of the Gross Regional Domestic Product regression analysis on rice production can be presented as follows:

Regression Analysis: pdrb versus padi

The regression equation is
pdrb = 2.61 + 4.49 padi

Predictor	Coef	SE Coef	T	P	VIF
Constant	2.609	1.015	2.57	0.025	
padi	4.4914	0.4162	10.79	0.000	1.000

S = 0.683114 R-Sq = 90.7% R-Sq(adj) = 89.9%

PRESS = 7.06893 R-Sq(pred) = 88.21%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	54.343	54.343	116.46	0.000
Residual Error	12	5.600	0.467		
Total	13	59.943			

Durbin-Watson statistic = 0.450693

Figure 6. Results of regression analysis and Analysis of Gross Regional Domestic Product Variance on rice production

➢ **Durbin-Watson Hypothesis Test:**
H0: $\rho = 0$; There is no autocorrelation
H1: $\rho \neq 0$; There is autocorrelation
D = 0.450693

From the Durbin-Watson table for n = 14 and k = 2, we obtain dU = 1.33 and dL = 0.97. Because 0 <D <dL then H_0 is rejected. This means that there is a positive autocorrelation in the model. There are other factors that influence the Y variable but are not included in the model. In addition, there are other models that are better than the model above. From the results of the regression analysis the regression model is obtained as follows:
pdrb (Gross Regional Domestic Product)
= 2.61 + 4.49 rice

➢ **Model Match Test:**
H_0: (Regression model that is formed is not significant)
H_1: (Significantly formed regression model)
Significance level $\alpha = 0.05$
Critical area: Reject H_0 if P-Value $<\alpha$

From the Analysis of Variance of Gross Regional Domestic Product data based on the results of rice production in Lampung Province, it can be seen that R-Sq of 89.9% means that 89.9% of the total variance of Gross Regional Domestic Product can be applied by the results of rice production in Lampung Province. The simple linear regression model has a large percentage to estimate or predict the value of GRDP. Besides that, it can be seen that the P-value is 0,000 but Durbin-Watson is 0.450693, so there is autocorrelation in the data.
Decision: P-Value (0,000) $<\alpha$ (0.05) then H_0 is rejected
Conclusion: Because P-value $<\alpha$, H0 is rejected

Because of the four assumptions above there is 1 assumption that is not fulfilled, the

simple linear regression model cannot be used to see the relationship between the Gross Regional Domestic Product and rice production data from 2000 to 2013.

2.1.4 Polynomial Regression

Because the Gross Regional Domestic Product data and rice production data cannot be analyzed with simple linear regression, a polynomial regression model approach is used. This research will be limited to using polynomial regression with two degrees and three degrees.

- Second Degree Polynomials

The polynomial equation of degree two can be written as follows:

$$Y = a + bx + cx^2$$

- Normal probability plot

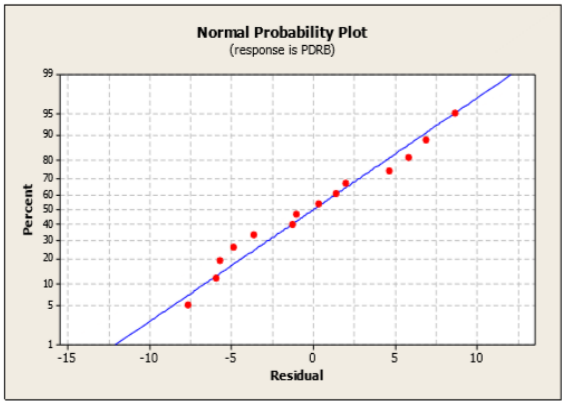


Figure 7. Normal Probability Plot of Gross Regional Domestic Product data on Second Degree Polynomials\

Normal probability can be seen that the points are around the estimation line. This shows that the Gross Regional Domestic Product data and rice production are normally distributed so that the first assumptions are met.

- Residual versus fit plot

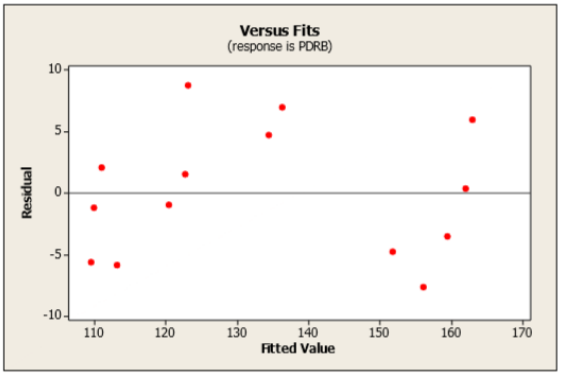


Figure 8. Plots of residual versus fit against Polynomial Gross Regional Domestic Product data

Based on the residual versus fits plot, it can be seen that the data is spread out and does not form a pattern like a funnel or something else. This means that the variety of data is homogeneous so that the second assumption is fulfilled.

- Residual plot versus order

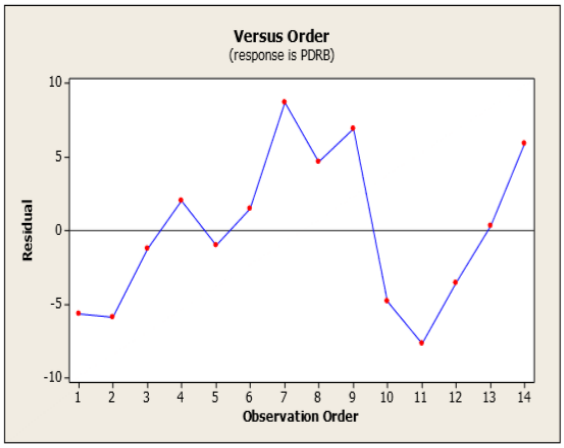


Figure 9. Plots of residuals versus orders against Gross Regional Domestic Product data on the Second Degree Polynomial

Based on the residual versus order plot, it appears that the data is spread out and forms a pattern which means the data is independent so that the third assumption is not fulfilled.

The results of the regression analysis and analysis of variance can be presented as follows:

Regression Analysis: PDRB versus PADI; padi^2

The regression equation is
PDRB = - 156 + 19.4 PADI - 0.294 padi^2

Predictor	Coef	SE Coef	T	P
Constant	-156.28	71.75	-2.18	0.052
PADI	19.379	5.827	3.33	0.007
padi^2	-0.2939	0.1148	-2.56	0.027

S = 5.64861 R-Sq = 94.1% R-Sq(adj) = 93.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	5643.3	2821.7	88.43	0.000
Residual Error	11	351.0	31.9		
Total	13	5994.3			

Source	DF	Seq SS
PADI	1	5434.3
padi^2	1	209.0

Durbin-Watson statistic = 0.943273

Figure 10. Results of regression analysis and analysis of variance in the Second Degree Polynomial

- Durbin-Watson Hypothesis Test:

H0: ρ = 0; There is no autocorrelation
H1: ρ ≠ 0; There is autocorrelation
D = 0.943273
From the Durbin-Watson table which corresponds to n = 14 and k = 3, we obtain dU = 1.7788 and dL = 0.7667. Because dL <D <dU, D is located in a decision-free region, meaning that there is no autocorrelation with the risk of a wrong decision. It could be that the model is wrong. Polynimial model with two degrees of Gross Regional Domestic Product on rice production, namely:
PRDB(Gross Regional Domestic Product) = - 156 + 19.4 RICE - 0.294 rice ^ 2

- Model Match Test:

H_0: (Regression model that is formed is not significant)
H_1: (Significantly formed regression model)
Significance level $\alpha = 0.05$
Areas of criticism: Reject H0 if P-Value $<\alpha$
From the Gross Regional Domestic Product data variance analysis table for polynomials of two degrees it can be seen that the R-Sq (Adj) of 93.1% means that 93.1% of the total variance of the total Gross Regional Domestic Product can be applied to the results of rice production in Lampung Province. This means that the polynomial regression model has a very large percentage to predict or predict the value of Gross Regional Domestic Product. In addition it can also be seen that the P-value of 0,000 but Durbin-Watson amounted to 0.943273, then the autocorrelation of the data occurs.
Decision: P-Value (0,000) $<\alpha$ (0.05) then H_0 is rejected.
Conclusion: Because of P-value $<\alpha$, H_0 is rejected.

Because of the four assumptions above there is one assumption that is not fulfilled, the polynomial model with a degree of two cannot be used to see the polynomial relationship between the Gross Regional

Domestic Product and rice production data in 2000 to 2013.

➢ Third Degree Polynomial

The polynomial equation of degree can be written as follows:

$$Y = a + bx + cx^2 + dx^3$$

- Normal probability plot

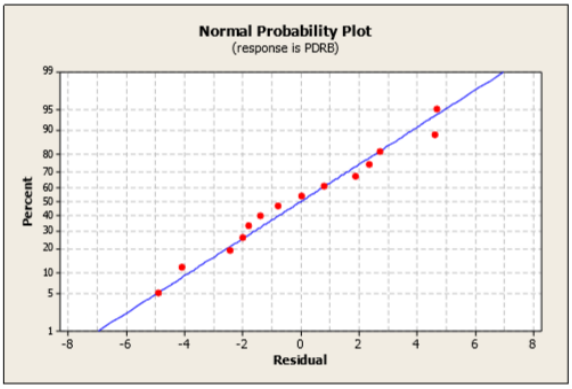


Figure 11. Normal Probability Plot of Gross Regional Domestic Product data on Third Degree Polynomials

Based on the normal probability graph, the points are located around the estimated line. This means that Gross Regional Domestic Product data and rice production in Lampung Province between 2000 and 2013 are normally distributed so that the first assumptions are met.

- Residual plot versus fits

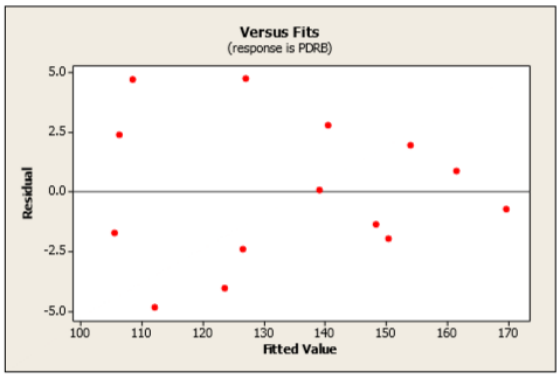


Figure 12. Residual plot versus fit of Gross Regional Domestic Product data on Third Degree Polynomials

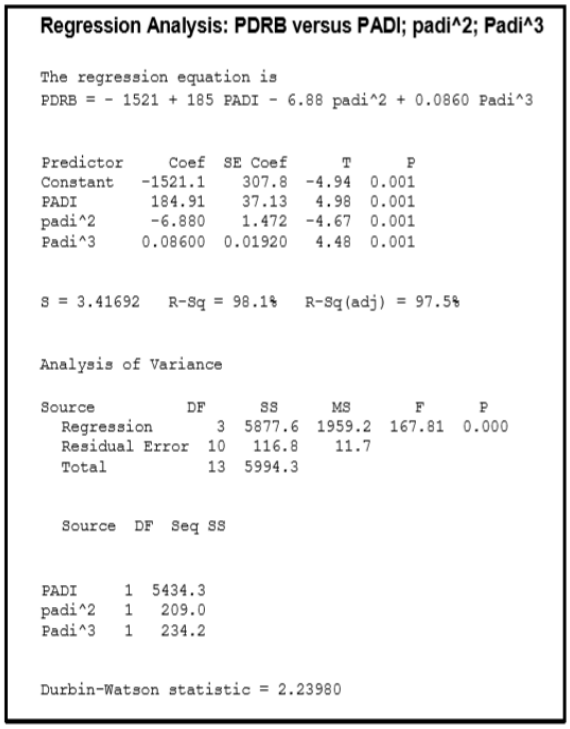
In the residual plot versus fit of the Gross Regional Domestic Product data on the Third Degree Polynomial, it is seen that the data is spread out and does not form a pattern like a funnel or something else. This means that the variety of data is homogeneous so that the second assumption is fulfilled.

- *Plot residual versus order*

Figure 13. Plots of residuals versus orders in Third Degree Polynomials

Based on the residual versus order plot of the Gross Regional Domestic Product data on the Third Degree Polynomial, it appears that the data is spread out and does not form a pattern which means that the data is independent so that the third assumption is fulfilled.

The results of the regression analysis and analysis of variance in the Third Degree Polynomial can be presented as follows:



Regression Analysis: PDRB versus PADI; padi^2; Padi^3

The regression equation is
PDRB = - 1521 + 185 PADI - 6.88 padi^2 + 0.0860 Padi^3

Predictor	Coef	SE Coef	T	P
Constant	-1521.1	307.8	-4.94	0.001
PADI	184.91	37.13	4.98	0.001
padi^2	-6.880	1.472	-4.67	0.001
Padi^3	0.08600	0.01920	4.48	0.001

S = 3.41692 R-Sq = 98.1% R-Sq(adj) = 97.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	5877.6	1959.2	167.81	0.000
Residual Error	10	116.8	11.7		
Total	13	5994.3			

Source	DF	Seq SS
PADI	1	5434.3
padi^2	1	209.0
Padi^3	1	234.2

Durbin-Watson statistic = 2.23980

Figure 14. Results of regression analysis and analysis of variance in Third Degree Polynomials

- Durbin-Watson Hypothesis Test:

H0: ρ = 0; There is no autocorrelation

H1: ρ ≠ 0; There is autocorrelation

D = 2.23980

From the Durbin-Watson table that corresponds to n = 14 and k = 4, we obtain dU = 2.0296 and dL = 0.6321. Because dU <D <4-dL, it does not reject H0, meaning that there is no autocorrelation in the model. The regression analysis model on the polynomial of degree three can be written as follows:

PRDB(Gross Regional Domestic Product) = - 1521 + 185 RICE - 6.88 rice ^ 2 + 0.0860 Rice ^ 3

- Model Match Test:

H_0: (Regression model that is formed is not significant)

H_1: (Significantly formed regression model)

Significance level $\alpha = 0.05$

Areas of criticism: Reject H0 if P-Value $<\alpha$

From the Gross Regional Domestic Product data variance analysis table on polymimials of three degrees it can be seen that the R-Sq (Adj) of 97.5% means that 97.5% of the total variance of the total Gross Regional Domestic Product can be applied to the results of rice production in Lampung Province. This means that the polynomial regression model with a degree of three percentages is very large to estimate or predict the value of Gross Regional Domestic Product. In addition it can also be seen that the P-value is 0,000

Decision: P-Value (0,000) <α (0.05) then H_0 is rejected.

Conclusion: Because the P-Value <α, then H_0 is rejected, so the regression model that is formed is significant or suitable for predicting Gross Regional Domestic Product.

Because the four assumptions above are fulfilled, a three-degree polynomial regression model can be used to see the shape of the polynomial relationship between Gross Regional Domestic Product and rice production in 2000 to 2013.

2.1.5 Selection of the Best Model

The selection of the best model is based on the correlation coefficient or the coefficient of determination symbolized by R2 or R-Sq (Adj). presented in the following table:

No	Model	Persamaan	R-Sq(Adj)	Keterangan
1	Regresi Linear Sederhana	$Y = 2.61 + 4.49\,x$	89,9 %	Terdapat Autokorelasi
2	Polinomial berderajat dua	$Y = -156 + 19.4\,x - 0.294\,x^2$	93,1%	Inkonklusiv Autokorelasi
3	**Polinomial berderajat tiga**	$Y = -1521 + 185\,x - 6.88\,x^2 + 0.0860\,x^3$	**97,5%**	**Semua asumsi Terpenuhi**

Table 2. Results of linear model analysis

The closer R-Sq (Adj) to 1, the better the match of data with the model, and vice versa the closer R-Sq (Adj) to 0, the better the match is. So as to determine the best model obtained by looking at the largest R-Sq (Adj) value. Based on the table above, the largest R-Sq (Adj) produced by a polynomial model with a degree of 97.5% with all assumptions fulfilled, so that the best model to describe the pattern of Gross Regional Domestic Product with rice production in Lampung Province is a three-degree polynomial model.

3.1 CONCLUSIONS AND SUGGESTIONS

3.1.1 Conclusions

Based on the results of the analysis conducted, it can be concluded that:

1. Gross Regional Domestic Product of Lampung Province agriculture, livestock, forestry, and fisheries sectors have a very strong positive linear relationship with the amount of rice production in Lampung Province.

2. The higher the amount of rice production in Lampung Province, the higher the Gross Regional Domestic Product of Lampung Province in the agriculture, livestock, forestry and fisheries sectors.

3. The three-degree Polynomial Model Y = -1521 + 185 x - 6.88 x2 + 0.0860 x3 with Y as the Gross Regional Domestic Product and X as the amount of rice production in

Lampung Province is the best model that can explain the relationship between the two.

3.1.2 Suggestions

The suggestions in this study are:

1. The need to provide socialization in agriculture to farmers to improve the ability of farmers.

2. The government should be able to stabilize the price of rice production to improve the welfare of farmers.

REFFERENCES

Cochran, W.G. 1997. *Sampling Tehnique.* New York: John Willeyl & Sons.

Hendradi, C.T. 2006. *Statistika Six Sigma dengan Minitab.* Yogyakarta: Andi.

Lind, Douglas. (2009). *Teknik-teknik Statistika dalam Bisnis dan Ekonomi*.

Jakarta:Salemba Empat.

Sembiring, R.K. 1995. *Analisis Regresi.* Bandung:ITB.

Steel,R.G.D. & Torrie,J.H. 1995. *Prinsip dan Prosedur Statistika Suatu Pendekatan*

Biometrik. Jakarta: PT. Gramedia Pustaka Utama.

Sulaiman, W. 2004. *Analisis Regresi menggunakan SPSS.* Yogyakarta: Andi.

Walpole, R. E. 1988. *Pengantar Statistika*. Edisi Ke-3. Jakarta: PT Gramedia Pustaka Utama

Usman, M. 2009. *Model Linear Terapan.*, Bandung: Sinar Baru Algensindo.

Yanto. 2016. *Statistika Inferensia untuk Penelitian dengan Minitab.*

Yogyakarta: ANDI.