

Prediction of Poor Inhabitant Number Using Least Square and Moving Average Method

Ningrum Ekawati

Informatics Engineering, STMIK Amikom Yogyakarta,
North Ring Road Street, Condong Catur Depok-Sleman-Indonesia, Telp : 0274 – 884201
email: ningrumekawati@gmail.com

Abstract

Poor population in South Kalimantan recently shows a decreased number for the last three years, compared to few previous years. The number of poor population differs from time to time. This dynamical scaled number has actually been a problem for South Kalimantan local government to take proper policies to solve this matter. It will then be necessary to predict potential number of poor population in the next year as the basis of subsequent policy making. This research will apply both Least Square and Moving Average methods as measurement to count prediction values. From the result, we can say that prediction analysis using those two methods is valid for predicting acquired number of potential people population based on its previous data due to its closest result to the actual condition. Reviewing the test result of last three years, the applied least square method shows validity of 92, 8%. Meanwhile, the applied moving average method shows validity of 98,8% both are considered valid.

Keywords: prediction, least square, moving average

Copyright © 2015 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Poverty is a main problem to which South Kalimantan local government concerns. According to Statistic center Agency (BPS), poor population is defined as those who averagely spend below the poverty line per capita per month. There are some indicators closely relating with the number of poor population in South Kalimantan province, such as service sector of Gross Regional Domestic Product (PDRB), service on constant price basis, the number of unemployment, the average spend per month per capita for durable goods and fruit category. The increasing development at South Kalimantan province decreases the number of poor population.

Based on South Kalimantan Province BPS data [14], the number of poor population on 2005 is 235.700 inhabitants; and 278.451 inhabitants on 2006, and 233.500 inhabitants on 2007, and 216.898 inhabitants on 2008, and 175.977 inhabitants on 2009, and 181.963 inhabitants on 2010, and 194.623 inhabitants on 2011, and 194.623 inhabitants on 2012, and 184.297 inhabitants on 2013, and 182.876 inhabitants on 2014. There are decreasing number of poor populations in South Kalimantan Province into 34, 32% from 2012 to 2014.

The problem lies in when predicting the decreasing number of South Kalimantan poor population for the last 3 years. However, there has been still no available prediction for the subsequent years, troubling the local government to put an action policy. This research is conducted to help BPS of South Kalimantan predict the number of poor population, making it easier to the local government to put an action and to increase people's life quality.

2. Research Method

Analysis time series is a statistical analysis method applied to predict a future condition. To make a correct and proper prediction, various data and information may be required to be observed on such a relatively long time period. As one of choices to describe a future trend, an analysis time series can be applied to reflect dynamic variable from one time to another [7]. From the previous researches using Least Square [1-6] and Moving Average [7-13] Methods, some results of analyzed data have been gained to help future prediction. Prediction is a management process to help decision making. It is described as an estimation process on the

unknown future situation. On the more general term, it is well known as a prediction referring to time series estimation or longitudinal type of data [8].

Least square method is often used to predict (Y), due to its thorough measurement. The trend line equation to be found is [1]:

$$Y = a_0 + bx \quad a = (\sum Y) / n$$

$$b = (\sum XY) / \sum X^2$$

With:

Y : Scaled data (time series) = Trend value prediction.

a₀ : Trend value on the basis year.

B : Averagely growing trend value on each year.

x : Time variable (day, week, month or year).

To conduct the measurement, a certain value on time variable (x) is required so it makes time variable value number zero or $\sum x = 0$. When analyzing data with Least Square Method, they are generally divided into two cases i.e. "even data" case and "odd data" case [1].

For odd "n", then:

The interval between two times gains a one-unit value

It is marked as negative when it is above 0

It is marked as positive when it is below 0

For even "n", then:

The interval between two times gains a two-unit value

It is marked as negative when it is above 0

It is marked as positive when it is below 0

When processing data tabulation for an "odd data" case, previous collected data from the last 9 years are required. Meanwhile, when processing data tabulation for an "even data" case, previous collected data from the last 10 years are required. Generally, linier line equation from time series analysis is:

$$Y = a + b X.$$

Explanation:

Y is a variable whose trend is to be searched.

X is a time variable (year).

Meanwhile, to find constant value (a) and parameter value (b) is:

$$a = \sum Y / N, \text{ dan } b = \sum XY / \sum X^2.$$

Moving Average method is a prediction approach by which we put some observed group of values, measure the average, and apply those average values as a prediction for subsequent period. The mathematic equation from this technique is [7]:

$$F_{t+1} = \frac{X_1 + X_2 + \dots + X_T}{T} \quad \text{or } \beta = \frac{\sum yt}{n}$$

Explanation:

F_{t+1} and β : Prediction for "t" period+ 1

X_T and $\sum yt$: Real value of "t" period

T and n : Moving average interval Duration moving average.

3. Results and Analysis

Data required on this research are those of poor population number on South Kalimantan province starting from 2005 to 2014.

Table 1. Population poor data, source from South Kalimantan BPS province.

No	Year	Number (inhabitant)
1	2005	235.700
2	2006	278.451
3	2007	233.500
4	2008	218.898
5	2009	175.977
6	2010	181.963
7	2011	194.623
8	2012	190.597
9	2013	184.297
10	2014	182.876

3.1. Least Square Method

3.1.1. "Odd Data" Case

Before measuring the prediction of acquired number of poor population on year 2015, we are going to test result measurement of those of 2012, 2013, and 2014 to consider the validity of Square least Method. To gain the result measurement of acquired number of poor population on 2012, that of 2013 to 2011 is required. Next, we continue to measure the variable value of X, XY, and X^2 .

Table 2. Data variable of poor population number, from year 2003 until 2011

No	Year	Number (Y)	X	XY	X^2
1	2003	258.960	-4	-1035840	16
2	2004	231.000	-3	-693000	9
3	2005	235.700	-2	-471400	4
4	2006	278.451	-1	-278451	1
5	2007	233.500	0	0	0
6	2008	218.898	1	218898	1
7	2009	175.977	2	351954	4
8	2010	181.963	3	545889	9
9	2011	194.623	4	778492	16
Total		2.009.072		-583458	60

Then, to measure the "a" value is:

$$a = \Sigma Y / N$$

$$a = 223230.22$$

And to measure "b" value is:

$$b = \Sigma XY / \Sigma X^2$$

$$b = -9724.3$$

After gaining the "a" and "b" value, the linear line equation then will be found. That is:

$$Y = a + b X (\text{for year 2012 } X \text{ value is } 5).$$

With that already measured linear line equation, we can predict the number of poor population on year 2012 by:

$$Y = 174608.72$$

It means the predicted number of poor population on 2012 is 174.608 inhabitants.

Then we are going to measure the acquired number of poor population on 2013. Data to be used are those of 2004 to those of 2012. With the same measurement approach, the prediction result of poor population on 2013 shows a number of 171.022 inhabitants. And for the result of acquired number of poor population on 2014, data to be used are those of 2005 to those of 2013. With the same measurement approach, the prediction result of poor population on 2014 shows a number of 161.789 inhabitants.

It is defined if the gap between Least Square Method measurement and real number is >40%, then such a measurement is considered invalid. Compared to the real result measured on 2012, there is a gap of 8, 39% (15.989 inhabitants), meaning that the result is valid. On 2013, there is a gap of 7, 2% (13.275 inhabitants), meaning that the result is also valid. And on 2014, we find a gap of 11, 53% (21,068 inhabitants), meaning that the result is also valid. We can see 3 valid results based on those three comparisons, meaning that the used Least Square Method is valid or close to the factual.

Then the measurement for prediction of acquired number of poor population on 2015 will be conducted. On data tabulation for "Odd Data" case, the required poor population data are those of five last years, starting from 2006 to 2014.

Table 3. Variable Data of poor population number, year 2006 to year 2014

No	Year	Number (Y)	X	XY	X ²
1	2006	278.451	-4	-1113804	16
2	2007	233.500	-3	-700500	9
3	2008	218.898	-2	-437796	4
4	2009	175.977	-1	-175977	1
5	2010	181.963	0	0	0
6	2011	194.623	1	194623	1
7	2012	190.597	2	381194	4
8	2013	184.297	3	552891	9
9	2014	182.876	4	731504	16
Total		1.841.182		-567865	60

Based on Table 3, the "a" and "b" values will be gained. To measure "a" and "b" values, below is the step:

To find "a" value is:

$$a = \Sigma Y / N$$

$$a = 204575.8$$

And to find "b" value is:

$$b = \Sigma XY / \Sigma X^2$$

$$b = -9464.42$$

After "a" and "b" values have been gained, the linear line equation will be gained, as following:

$$Y = a + b X(\text{for year 2015 the value of } X \text{ is } 5)$$

By the already measured linear line equation, we can predict the number of poor population on 2015, as following:

$$Y = 157253.7$$

It means that the number of poor population on year 2015 is 157.253 inhabitants.

Further, the trend analysis graphic with Least Square Method for the actual result comparison and the prediction result on last 3 years are:

Table 4. Variable Data of poor population number, year 2006 to year 2014.

Year	Actual result	Prediction
2012	190.597	174.608
2013	184.297	171.022
2014	182.876	161.789
2015		157.253

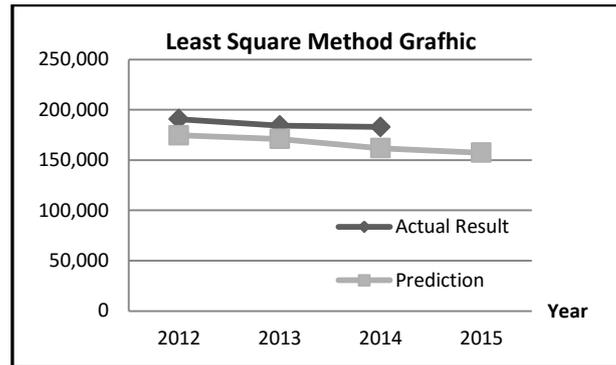


Figure 1. Prediction Graphic with Least Square Method

3.1.2. "Even Data" Case

Data required on the case of "even data" are those of poor population number on South Kalimantan Province starting from year 2005 to year 2014.

Table 5. Variable Data of poor population number, year 2005 to the year 2014

No	Year	Number (Y)	X	XY	X ²
1	2005	235.700	-9	-2121300	81
2	2006	278.451	-7	-1949157	49
3	2007	233.500	-5	-1167500	25
4	2008	218.898	-3	-656694	9
5	2009	175.977	-1	-175977	1
6	2010	181.963	1	181963	1
7	2011	194.623	3	583869	9
8	2012	190.597	5	952985	25
9	2013	184.297	7	1290079	49
10	2014	182.876	9	1645884	81
Total		2.076.882		-1415848	330

Based on Table 5, then we will gain the "a" and "b" values. To find those values, below are the steps:

To gain the "a" value is:

$$a = \frac{\sum Y}{N}$$

$$a = 207688.2$$

To gain the "b" value is:

$$b = \frac{\sum XY}{\sum X^2}$$

$$b = -4290.45$$

After those both values "a" and "B" have been gained, the linear line equation will also be gained, as following:

$$Y = Y = a + b X (\text{for year 2015 the value of X is 11})$$

With that equation, we can predict the number of poor population on 2015, such as following:

$$Y = 160493.3$$

It means the predicted number of poor population is 160.493 inhabitants.

From the measurement using Least Square method, the prediction of acquired number of poor people on year 2015 for "odd data" shows the number of 157.253 inhabitants. And that

of “even data” shows the number of 160.493 inhabitants. So the gap between those numbers is only 0, 98%. However, the result of predicted measurement of acquired number of poor population can be denied or aborted due to certain conditions such as natural disaster, disease epidemic, etc.

3.2. Moving Average Method

Before measuring the prediction of acquired number of poor population on 2015, a measurement result test of acquired number of poor population 2012, 2013, and 2014 will be conducted to figure out the validity of Single Moving Average method. To measure demand prediction of year 2015, first of all, we will measure the prediction of year 2014.

Table 6. Variable Data of poor population number, year 2003 to year 2014

No	Year	Number (inhabitant)
1	2003	258.960
2	2004	231.000
3	2005	235.700
4	2006	278.451
5	2007	233.500
6	2008	218.898
7	2009	175.977
8	2010	181.963
9	2011	194.623
10	2012	190.597
11	2013	184.297
12	2014	182.876

The next step is to define the variable value of poor population number using Single Moving Average Method for 2 periods, such as following:

Table 7. Variable Data of Poor Population number using Single Moving Average for 2 periods.

No	Year	Actual Result (inhabitants)	MA 2 Prediction Period
1	2003	258.960	-
2	2004	231.000	-
3	2005	235.700	244.980
4	2006	278.451	233.350
5	2007	233.500	257.076
6	2008	218.898	255.976
7	2009	175.977	226.199
8	2010	181.963	197.438
9	2011	194.623	178.970
10	2012	190.597	188.293
11	2013	184.297	192.610
12	2014	182.876	187.447
13	2015	-	183.587

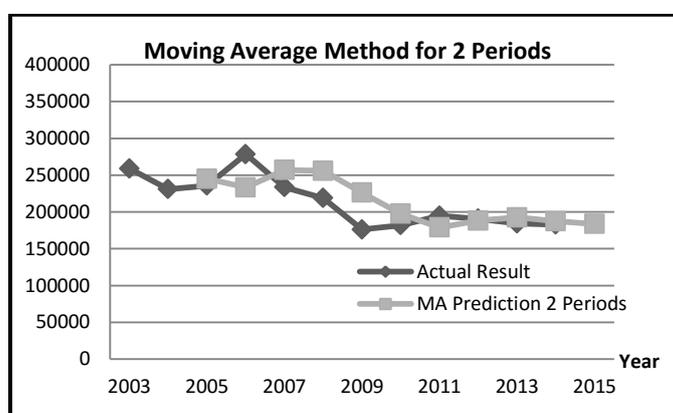


Figure 2. Prediction Graphic using Moving Average Method for 2 periods.

It is defined if the gapping value between prediction measurement and Single Moving Average for 2 periods with the actual results is >40%, then it is considered invalid. Compared to the real result gained on 2012, there is a gap of 1,21% (2.034 inhabitants), considered as valid. On 2013, there is a gap of 4,51% (8.313 inhabitants), meaning that the gained result is considered valid. On 2014, there is also a gap of 2, 50% (4.571 inhabitants), meaning that the gained result is also considered valid. From those three comparisons, we have gained 3 valid results. It brings to the statement that Single Moving Average Method is valid or closest the factual.

3.3. Result Comparison of Analysis Method of Least Square and Moving Average Trend Measurement

From those two used methods by analysis of Least Square and Moving average trends, it is considered invalid, when the acquired gap between prediction measurement and actual result is >40%. On the table 8, there is a comparison result from analysis method of of Least Square and Moving Average trends for last three years.

Table 8. Comparison result of Least Square and Single Moving Average method for 2 periods.

Year	Actual Result (inhabitants)	Prediction		Gap (%)		Conclusion
		Least Square	Moving Average	Least Square	Moving Average	
2012	190597	174608	188293	8,39	1,21	Valid
2013	184297	171022	192610	7,2	4,51	Valid
2014	182876	161789	187447	11,53	2,5	Valid
2015		157253	183587			

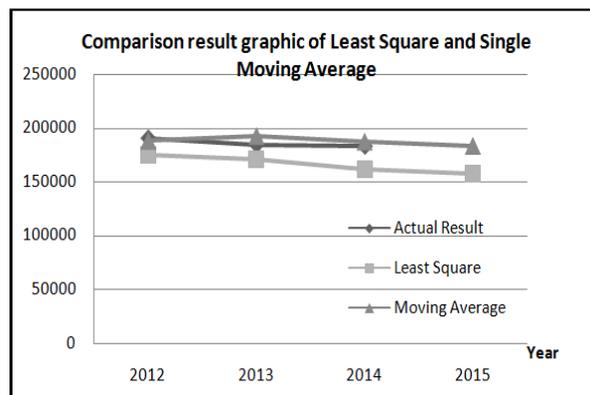


Figure 3. Comparison result graphic of Least Square and Single Moving Average method for 2 periods

4. Conclusion

From the research result, we may conclude some points as following:

1) Prediction analysis using analysis method of Least Square and Moving Average trends can be applied to predicting the number of poor population on South Kalimantan province for future period based on poor population data from South Kalimantan province BPS's one last year data due to its valid result or closest result to the factual.

2) From the measurement for the acquired number prediction of poor population on year 2012, the data show 174.608 inhabitants, and on 2013 the data show 171.022 inhabitants, and on 2014 the data show 161.789 inhabitants. It is considered invalid when the acquired gap between the measurement with Least Square Method and the actual result shows 40%. Compared to the actual result gained on 2012, there is a gap of 8, 39% (15.989 inhabitants), meaning that the result is considered valid. On year 2013, we find the gap of 7, 2% (13.275 inhabitants), meaning that the gained result is considered valid. On year 2014, there is a gap of

11, 53% (21,068 inhabitants), meaning that the gained result is also considered valid. From those three comparisons we have gained three valid results, meaning that the method of Least Square is valid or closes to the factual.

3) From the measurement result from prediction of acquired number of poor population on 2012, it shows 188.293 inhabitants, and on 2013 it shows 192.610 inhabitants, and on 2013 it shows 187.447 inhabitants. It is considered invalid when the acquired gap between measurement with Single Moving Average for 2 periods and the actual result is >40%. Compared to the factual result gained on 2012, there is a gap of 1,21 % (2.034 inhabitants), meaning that the result is valid. On 2013, we can see a gap of 4, 51% (8,313 inhabitants), meaning that the result is also valid. On 2014 the acquired gap is 2, 50% (4.571 inhabitants), meaning that the acquired result is also valid. From those three comparisons, we have gained three valid results, meaning that the Single Moving Average formula is valid or closest to the factual.

4) From the measured result, using Least Square and Single Moving Average methods, we may conclude that those two methods are proven to be valid or closest to the factual with the percentage result of Least Square method preciseness is 92, 8% and Moving Average method preciseness (single moving average for 2 periods) is 98,8%.

References

- [1] Muhammadl FR. Perancangan Aplikasi Peramalan Persediaan Obat-obatan Menggunakan Metode Least Square (Studi Kasus: Apotik Mutiara Hati). *Pelita Informatika Budi Darma*. 2014; 6(1): 49-51.
- [2] He M, Xiu Y. Estimation of the complex frequency of a harmonic signal based on a linear least squares method. *Geodesy and Geodynamics* 2015. 2015; 6(3): 220-221.
- [3] Sevcan AK, Mustafa P. *Least square support vector machine and minimum redundancy maximum relevance for diagnosis of breast cancer from breast microscopic images*. INTE 2014. Elazig. 2104: 4026-4027.
- [4] Shuai W, Wei S. *Forecasting Direction of China Security Index 300 Movement with Least Squares Support Vector Machine*. ITQM 2014. Beijing. 2014: 869-870.
- [5] Guohua T, Jianzhuo Y, Chen G, Suhua Y. *Prediction of water quality time series data based on least squares support vector machine*. International Conference on Advances in Computational Modeling and Simulation. Beijing. 2012: 1194-1195.
- [6] Kubra B, Hayri HD, Dursun ZS. *Ordinary Least Squares Regression Method Approach for Site Selection of Automated Teller Machines (ATMs)*. Spatial Statistics 2015: Emerging Patterns. Istanbul. 2015: 66-67.
- [7] Ingrid M, Tomas P. *Estimation of economic development in Papua New Guinea: linear trend analysis or moving average model*. Contemporary Issues in Business of Management and Education. Karvina. 2013: 450-453.
- [8] Lazim A. ARIMA Model for Gold Bullion Coin Selling Prices Forecasting. *International Journal of Advances in Applied Sciences (IJAAS)*. 2012; 1(4): 153-154..
- [9] Vicente R, Miguel AP, Aitziber O. *Dynamic portfolio management strategies based on the use of moving averages*. 2nd World Conference On Business of Economics And Management -WCBEM 2013. Bilbao. 2013: 1277-1278.
- [10] Kato M, Claude T. *Applying Moving Average Filtering for Non-interactive Differential Privacy Setting*. Complex Adaptive Systems of Science and Technology. Philadelphia. 2014: 409-410.
- [11] Hye KY, Na YM, Sung SK, Chaeshin C, Mee KK. *Forecasting the Number of Human Immunodeficiency Virus Infections in the Korean Population Using the Autoregressive Integrated Moving Average Model*. Korea Centers for Disease Control and Prevention. Osong. 2013: 358-359.
- [12] Arindrajit P, Jyoti PS, Paramartha D. *The Path Length Prediction of MANET using Moving Average Model*. International Conference on Computational Intelligence of Modeling Techniques and Applications (CIMTA) 2013. West Bengal. 2013: 882-883.
- [13] Rakesh K, Dalgobind M. Application of Proper Forecasting Technique in Juice Production. *Global Journal of Researches in Engineering Industrial Engineering*. 2013; 4(1): 42-43.
- [14] Data of poor population in South Kalimantan: <http://kalsel.bps.go.id/LinkTabelStatis/view/id/832>.