Application of Backward Method to Determine the Factor that Affect the Poverty Level of Sibolga City 2011-2020

Tika Afina Sari[1*], Suyanto[2]

[1]Student of Mathematics, Universitas Sumatera Utara, Indonesia
[2]Lecturer In Mathematics Education, Universitas Sumatera Utara, Indonesia
*Corresponding Author. E-mail : tikaafina@gmail.com,suyanto@usu.ac.id

Article Info	ABSTRACT
	The level of poverty is a condition that is financially impossible to meet an adequate standard of living of the local community, not only in terms of health and education, but also in terms of basic needs.The purpose of this study is to obtain information about the factors that influence the poverty rate of Sibolga City from 2011-2020 by applying the backward elimination method which sequentially selects variables that do not affect the equation model.The estimators obtained are $\hat{Y} = -11{,}783 + 1{,}264\,X_4 - 5{,}525X_5 - 0{,}273X_6$. The percentage of determination obtained is 83.4% with a significance level of 5 percentand the most influential factor of the independent variables that remain in the equation model is the average length of schooling.

To cite this article:

INTRODUCTION

Regression analysis is a statistical technique that is useful for examining and modeling the relationship between one Regression analysis is a mathematical model that can be used to determine the pattern of the relationship between two or more variables. The main purpose of regression analysis is to estimate the value of a variable (the dependent variable) if the value of another variable (the independent variable) has been determined (Algifari,2000).

Backward method is a method by including all variables into the model but then one by one the independent variables are removed from the model based on certain statistical significance criteria. The Backward method is a good regression method because in this method the behavior of the response variable is explained as well as possible by selecting the explanatory variable from the many explanatory variables available in the data. This method saves time and effort and is satisfying for those of us who want to see all the random variables in an equation not missing something (Delsen dkk,2019) . The Backward Elimination is a method that is used with a backward step, namely by eliminating insignificant traits in the equation model whose backward elimination algorithm is based on a linear regression model. The backward elimination method is a good regression method to use because one of the best methods to explain the relationship of the dependent variable very well is to choose an estimator from the many predictors available in the data (N, Draper,1992). According to the Sibolga City Government, it is time to accelerate the handling of poverty problems in Indonesia (N,Gilmore dkk, 2019). Especially in urban areas, people are said to be poor if there is an inability to fully access physical facilities and basic environmental infrastructure. In terms of the quality of housing and facilities, the standard quality is less than adequate, and the work is also not permanent. To find

out the causes of poverty in an area, it is necessary to apply a policy model that reduces the risk of poverty from year to year. One that can be applied is the backward elimination method. The subject of this research is the level of poverty in Sibolga City. In 2019 there were still 10.82 thousand people or 12.36% of the population who were below the poverty line, with the 2019 poverty line value of Rp. 425,236. This number has increased when compared to 2018 which was

10.81 thousand or 12.36% of the total population of Sibolga City, where Sibolga City was ranked 25 out of 34 cities in North Sumatra Province, namely in 2019 and 2020 (N,Gilmore dkk,2019). Metussin, H et al. (2019) has conducted a study using the backward elimination method, concluding that of all the analyzed variables, the remaining variables included in the equation are cognitive change, learning strategies, and self-skills, where these variables are several factors that can improve math achievement in Brunei (Haloui dkk, 2020).

RESEARCH METHOD

The research was carried out at the Sibolga Statistical Center Agency (BPS SIBOLGA) located on Jalan Tuanku Push Hutagalung No. 1. 2, Beringin City, Sibolga City, North Sumatra. The object of this study is secondary data such as minimum wage, unemployment rate, economic growth, human development index, average length of school, and the number of Sibolga City labor force for the period 2011-2020.

The data source applied to this study is secondary data. Secondary data is data obtained indirectly from existing data by obtaining quotations from researchers for research purposes either from written reports, books, documents, library research, or sources related to the research to be conducted. The data collection carried out in Sibolga BPS was done by recording from Sibolga BPS archives used to solve existing problems. The types of data applied are data obtained from companies in numerical form such as minimum wage data, unemployment rate, and other data related to this study.

RESULTS AND DISCUSSION

Sufficient Test of Data Sample

In research involving data collection, certain provisions must be observed to determine whether the sample size used is sufficient for the analysis criteria or not, while the assumptions made to conduct a sufficient sample adequacy test $\alpha = 0.05$ are as follows:

$$N' = \left[\frac{40\sqrt{N\sum_{t=1}^{n}Y_t{}^2 - \sum_{t=1}^{n}(\sum Y_t)^2}}{\sum_{t=1}^{n}Y_t}\right]^2$$

$$N' = \left[\frac{40\sqrt{N\sum Y_t{}^2 - (\sum Y_t)^2}}{\sum Y_t}\right]^2$$

$$N' = 2,962$$

Because the value of N' obtained is 2.96 smaller than the value of N = 10, then H0 is accepted because it meets the criteria for the hypothesis test, so it can be concluded that the data to be processed meets the standards for analysis

Multiple Regression Equation between Y and $X_1, X_2, X_3, X_4, X_5, X_6$

For multiple linear regression, the equation obtained by eliminating the equations of the least squares method, can be obtained with the help of the Statistical Product software and Social Science which is usually abbreviated as SPSS, which is shown in Table 2.

Table 2. Multiple Regression Equation between Y and $X_1, X_2, X_3, X_4, X_5, X_6$

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-46.446	76.250		-.609	.585
	X1	-1.810	3.341	-1.999	-.542	.626
	X2	-.045	.062	-.256	-.721	.523
	X3	-.198	.150	-.744	-1.319	.279
	X4	2.122	1.164	5.462	1.823	.166
	X5	-7.882	2.754	-3.582	-2.862	.064
	X6	-.232	.186	-1.254	-1.252	.299

a. Dependent Variable: Y

Then multiple regression equation formed is:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + b_5 X_5 + b_6 X_6$$

$$\hat{Y} = -46,446 - 1,810 X_1 - 0,045 X_2 - 0,198 X_3 + 2,122 X_4 - 7,882 X_5 - 0,232 X_6$$

B_1, β_2, β_3, β_5 and β_6 have negative values, which means that if the independent variable increases, the dependent variable decreases. β_4 have positive values, which means that if the independent variable increases, the dependent variable also increases.

Table 3. Partial Correlation Test and ANOVA between Y and $X_1, X_2, X_3, X_4, X_5, X_6$

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
X1	Between Groups	101.171	8	11.241	.486	.760
	Within Groups	13,865	1	5.971		
	Total	115.035	9			
X2	Between Groups	20.271	8	2.252	.926	.546
	Within Groups	7.121	1	1.113		
	Total	27.392	9			

X3	Between Groups	16.334	8	1.815	2.581	.358
	Within Groups	3.113	1	.419		
	Total	19.447	9			
X4	Between Groups	13.947	8	6.439	4.813	.012
	Within Groups	5.229	1	3.125		
	Total	19.176	9			
X5	Between Groups	55.928	8	.070	7.851	.008
	Within Groups	1.525	1	.003		
	Total	57.453	9			
X6	Between Groups	89.111	8	9.901	4.321	.003
	Within Groups	8.920	1	5.541		
	Total	98.031	9			

From table 2, it can be concluded that the F distribution obtained with a confidence level of 0.05 is the variable X1 with the smallest $F_{partial}$ = 0.486 while the $F_{table(6,9,0.05)} = 3,37$ value. Because the smallest partial $F_{partial} < F_{table}$ the variable X_1 will be eliminated from the equation model.

Multiple Regression Equation between Y and X_2,X_3,X_4,X_5,X_6

The results of the correlation coefficient of each variable when calculated using SPSS 20 are as follows:

Table 4. Multiple Regression Equation between Y and X_2,X_3,X_4,X_5,X_6

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6.306	16.367		-.385	.720
	X2	-.030	.051	-.175	-.599	.582
	X3	-.153	.114	-.576	-1.347	.249
	X4	1.573	.521	4.050	3.018	.039
	X5	-8.086	2.476	-3.675	-3.266	.031
	X6	-.303	.120	-1.634	-2.518	.065

a. Dependent Variable: Y

Then multiple regression equation formed is:

$$\hat{Y} = b_0 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6$$
$$\hat{Y} = -6,306 - 0,30X_2 - 0,153X_3 + 1,573X_4 - 8,086 X_5 - 0,303X_6$$

Table 5. Partial Correlation Test and ANOVA between Y and X_2,X_3,X_4,X_5,X_6

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
	Between Groups	20.271	8	2.252	.926	.546
X2	Within Groups	7.121	1	1.113		
	Total	27.392	9			
	Between Groups	16.334	8	1.815	2.581	.358
X3	Within Groups	3.113	1	.419		
	Total	19.447	9			
	Between Groups	13.947	8	6.439	4.813	.012
X4	Within Groups	5.229	1	3.125		
	Total	19.176	9			
	Between Groups	55.928	8	.070	7.851	.008
X5	Within Groups	1.525	1	.003		
	Total	57.453	9			
	Between Groups	89.111	8	9.901	4.321	.003
X6	Within Groups	8.920	1	5.541		
	Total	98.031	9			

Based on the table above, it is found that the distribution of F obtained with a significance level of 0.05 is the variable X2 with the smallest $F_{partial}$ = 0,926 and the $F_{table} = F_{(5,9;0,05)} = 3,48$ Because the smallest $F_{partial}$ t<F_{table} , the variable X_2will be eliminated from the equation model.

Multiple Regression Equation between Y and X_3,X_4,X_5,X_6
The results of the correlation coefficient of each variable when calculated using SPSS 20 are as follows:

Table 6. Multiple Regression Equation between Y and X_3,X_4,X_5,X_6
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	-11.513	12.946		-.889	.415
	X3	-.140	.104	-.528	-1.346	.236
1	X4	1.624	.480	4.181	3.381	.020
	X5	-7.908	2.295	-3.594	-3.446	.018
	X6	-.316	.110	-1.708	-2.872	.035

Then multiple regression equation formed is:

$$\hat{Y} = -11,513 - 0,140X_3 + 1,624X_4 - 7,908X_5 - 0,316X_6$$

have negative values, which means that if the independent variable increases, the dependent variable decreases.β4 have positive values, which means that if the independent variable increases, the dependent variable also increases.:

Table 7. Partial Correlation Test and ANOVA between Y and X_3,X_4,X_5,X_6
ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
X3	Between Groups	16.334	8	1.815	2.581	.358
	Within Groups	3.113	1	.419		
	Total	19.447	9			
X4	Between Groups	13.947	8	6.439	4.813	.012
	Within Groups	5.229	1	3.125		
	Total	19.176	9			
X5	Between Groups	55.928	8	.070	7.851	.008
	Within Groups	1.525	1	.003		
	Total	57.453	9			
X6	Between Groups	89.111	8	9.901	4.321	.003
	Within Groups	8.920	1	5.541		
	Total	98.031	9			

Based on the table above, it is found that the distribution of F obtained with a significance level of 0.05 is the variable X_3 with the smallest $F_{partial} = 2.581$ and the $F_{table\ (4,9,0.05)} = 3,63$. Because the smallest partial $F_{partial} < F_{table}$, the variable X3 will be eliminated from the equation model.

Multiple Regression Equation between Y and X_4, X_5, X_6
The results of the correlation coefficient of each variable when calculated using SPSS 20 are as follows:

Table 8. Multiple Regression Equation between Y and X4,X5,X6
Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-11.783	13.793		-.854	.426
	X4	1.264	.425	3.255	2.973	.025
	X5	-5.525	1.556	-2.511	-3.550	.012
	X6	-.273	.112	-1.472	-2.430	.051

a. Dependent Variable: Y

Then multiple regression equation formed is: ^
$$\hat{Y} = -11,783 + 1,264X_4 - 5,525X_5 - 0,273X_6$$
β_5 and β_6 have negative values, which means that if the independent variable increases, the dependent variable decreases. β_4 have positive values, which means that if the independent variable increases, the dependent variable also increases.

Table 9. Partial Correlation Test and ANOVA between Y X_3, X_4, X_5, X_6
ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
X4	Between Groups	13.947	8	6.439	4.813	.012

	Within Groups	5.229	1	3.125		
	Total	19.176	9			
X5	Between Groups	55.928	8	.070	7.851	.008
	Within Groups	1.525	1	.003		
	Total	57.453	9			
X6	Between Groups	89.111	8	9.901	4.321	.003
	Within Groups	8.920	1	5.541		
	Total	98.031	9			

Based on the table above, it is found that the F distribution obtained with a significance level of 0.05 is the variable X6 with the smallest $F_{partial}$ = 4,321 and the $F_{table(3,9,0.05)} = 3,63$. Because the smallest partial $F_{partial}$ >F_{table}, the variable X6 is not eliminated from the model and the iteration is stopped, so X4,X5,X6 automatically remain in the model. From the six independent variables that have been studied using the backward method above, it can be concluded that only 3 variables are in the regression model, namely X4,X5,X6. The estimation equation that can be formed between the dependent variable Y and the remaining independent variable Y is as follows:

$$\hat{Y} = -11,783 + 1,264X_4 - 5,525X_5 - 0,273X_6$$

Correlation Coefficient and Determination

The coefficient of determination obtained from the remaining variables X4,X5,X6 can use the formula in Equation (7). From each of the sigma values above, the value of the coefficient of determination will be sought as follows:

$$R^2 = \frac{\begin{matrix} (-1,810) \times (-1,531) + (-0,045) \times 2,940 + (-0,198) \times 6,082 + \\ 2,122 \times (-3,226) + (-7,882) \times (-0,811) + (-0,232) \times (-6,750) \end{matrix}}{3,058}$$

$$R^2 = \frac{2,55118}{3,058}$$

$$R^2 = 0,834$$

The calculation results based on SPSS 20 are as follows:

Table 10. Determination Coefficient
Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.913[a]	.834	.511	.40770

a. Predictors: (Constant), X6, X3, X2, X5, X4, X1

From the results of the calculation and display of SPSS 20 output based on the table above, the amount of R square (R^2) is 0.834, this value indicates that the contribution of the variables X_3,X_4,X_5,X_6 is 83.4% while the remaining 16.6% is determined by other factors. outside the model that was not detected in the research that has been done.

Residual analysis

The estimator formed by the backward elimination method can apply residual analysis. The following discussion is related to the estimator that has been obtained previously, it is proven that the assumptions (i), (ii), (iii).

Table 11. Spearman Rank Correlation Coefficient and Residue

No	Y	$\hat{Y}$	e_j	Rank $\hat{Y}$	rank e	D	d^2
1	13,18	12,8581	0,29301	6	3	3	9
2	13,00	13,0602	-0,08975	4	6	-2	4
3	12,90	13,3356	-0,46476	2	10	-8	64
4	12,26	12,4535	-0,22405	9	7	2	4
5	13,48	13,1393	0,30998	3	2	1	1
6	13,30	12,9194	0,34909	5	1	4	16
7	13,69	13,4583	0,2004	1	4	-3	9
8	12,38	12,6403	-0,29291	7	9	-2	4
9	12,36	12,5979	-0,2701	8	8	0	0
10	11,95	11,7287	0,18909	10	5	5	25
Jumlah	128,50	128,191	0	55	55	0	136,00

a. Assumption (i)

Based on table 11, the mean residual $e_j = 0$ is fulfilled

b. Assumption (ii)

Variance (e) = variance (ek)2. Proof of assumptions is carried out as follows:

$$r_s = 1 - 6 \left[\frac{\sum d^2}{n(n^2 - 1)} \right]$$

$$r_s = 1 - 6 \left[\frac{136}{10(10^2 - 1)} \right]$$

$$r_s = 0,178$$

After the value of the Spearman rank correlation has been obtained, it will then be tested using the t test as follows:

$$t_{count} = \frac{r_s \sqrt{n - 2}}{\sqrt{1 - r_s^2}}$$

$$t_{count} = \frac{0,178 \sqrt{10 - 2}}{\sqrt{1 - 0,178^2}}$$

$$t_{count} = 0,52$$

Because in this study using n = 10 and α= 0.05 then $t_{tabel} = t_{(n-2;1-\alpha)} = t_{(8,0.95)} = 1,86$ then compare tcount with ttable so that we get 0,52 < 1.86 or in other words it is obtained that $t_{count} < t_{tabel}$, then the assumption is fulfilled.

c.   Assumption (iii)

Covarian $(e_j, e_k) = 0 ; j \neq k$
This proof can be shown with a picture of the spread of several points that are below and above 0 on the Y axis, as follows
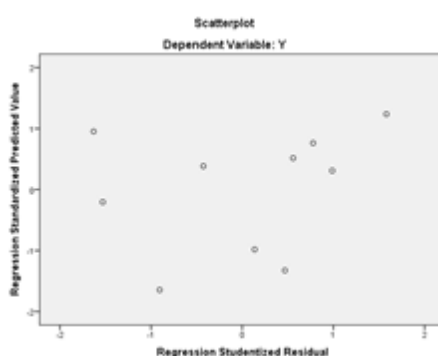
Figure 1. Residue Scatter Diagram

Based on Figure 1 above, it can be seen that the points are irregularly distributed and do not form a clear pattern where the points are above and below 0 on the Y axis, which proves that there is no heteroscedasticity in the model used, so that it is obtained conclusion that this assumption is met.

Conclusions

Based on what was discussed in the previous chapter, the conclusion is that of the six factors studied, the factors affecting the poverty level in Sibolga City are the human development index, the average length of school, and the number of labor force. As for the conjecture, $\hat{Y} = -11,783 + 1,264X_4 - 5,525X_5 - 0,273X_6$ means that if the human development index increases by one unit, then the poverty rate increases by 1.264 percent, if the average length of school falls by one unit, then the poverty rate drops by 5.525 percent.If the employment rate drops by one unit, the poverty rate will fall by 0.273 percent. The percentage of variation obtained using the backward method is 83.4% with a real degree of 0.05 meaning that the Y-value variation in the multiple linear regression equation is 83.4%, while the remaining 16.6% is affected by other variables outside the equation.

REFERENCES

Agresti, A. (1990). *Categorical Data Analysis.* New York: John Wiley and Sons.

Aji, C. M. (2014). Analisis Faktor-faktor yang Mempengaruhi Laju Pertumbuhan Penduduk Kota Semarang Tahun 2011 Menggunakan Geographically Weighted Logistic Regression. *Jurnal Gaussian*, Vol 3 No 2: 161-171.

Fatah, K. S., & Mahmood, R. F. (2016). Parameter Estiamtion for Binary Logistic

Regression Using Different Iterative Methods. *Journal of Zankoy Sulaimani,* Vol 19 No 2: 177-178.

Harlan, Johan. 2018. *Analisis Regresi Logistik*. Depok: Gunadarma.

.Hosmer, D. W., and Lemeshow, S. 2000.*Applied Logistic Regression*. New York: John Wiley and Sons, Inc.

Johnson, Richard A. dan Dean W. Wichern. 2007. *Applied Multivariate Statistical Analysis.* New Jersey: Pearson Education, Inc.

Kementerian Kesehatan RI. Profil Kesehatan Indonesia 2018. Jakarta: Pusat Data dan Informasi Kementrian Kesehatan RI; 2019.

Kim, J.S., and Dailey, R.J. 2008.*Biostatistic for Oral Healthcare*. USA: Blackwell Munksgaard.

Kuncoro, Mudrajad. 2004. *Metode Kuantitatif: Teori dan Aplikasi untuk Bisnis dan Ekonomi.* Yogyakarta: UPP AMP YKN.

Muaz, Fariz. Faktor-Faktor yang Mempengaruhi Kejadian Tuberkulosis Paru Basil Tahan Asam di Puskesmas Wilayah Kecamatan Serang Kota Serang tahun 2014. Jakarta: Universitas Syarif Hidayatullah.

Nirwana.S.R.A. 2015. Regresi Logistik Multinomial dan Penerapannya dalam Menentukan Faktor yang Berpengaruh pada Pemilihan Program Studi di Jurusan Matematika UNM.[Skripsi].Makassar:Universitas Negeri Makassar,Program Sarjana.

Peeters, B., Dewil, R., & Smets, I. Y. (2012). Improved Process Control of an Industrial Sludge Cemtrifuge-dryer Installation Through Binary Logistic Regression Modeling of The Fouling Issues. *Journal of Process Control*, Vol 22: 1390-1391.

Putri, E. I. S., Indriati, D. W., & Wahyunitisari, M. R. 2020. The Prevalence of Diabetes Mellitus among Hospitalized Tuberculosis Positive Cases in Hajj Hospital Surabaya. Malaysian Journal of Medicine and Health Sciences. 16(1): 235-239

Ranuh,I.G.N.2008. Pedoman Imunisasi di Indonesia. Edisi ketiga.Jakarta: Badan Penerbit Ikatan Dokter Anak Indonesia

RI, K. (2016). Peraturan Menteri Kesehatan Republik Indonesia Nomor 67 Tahun 2016 Tentang Penanggulangan Tuberkulosis. *Jakarta:Kementerian Kesehatan Republik Indonesia*.

Santoso, S. (2012).*Analisis SPSS pada Statistik Parametrik.*Jakarta: PT. ElexMedia Komputiondo.

Sugiarto,D.S.2006. *Metode statistika untuk bisnis dan ekonomi*.Jakarta: *Gramedia Pustaka Utama.*

Sumut, Dinkes. (2019). Profil Kesehatan Provinsi Sumatera Utara Tahun 2019. *Sumatera Utara.*

Suyo, J.2010, *Herbal Penyembuhan Gangguan Sistem Pernafasan*.,Yogyakarta: B First

Wijaya, I. M. K. (2013, December).Infeksi HIV (human immunodeficiency virus) pada penderita tuberkulosis.In *Prosiding Seminar Nasional MIPA*.

World Health Organization.Global Tuberculosis Report 2020. Geneva: World Health Organization; 2020.