M A T H L I N E
JURNAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA
ISSN 2502-5872 (Print)
ISSN 2622-3627 (Elektronik)

Volume 7 Nomor 2, Agustus 2022, halaman 185 – 196.

Application Of Truncated Spline Nonparametric Regression In Modeling Traffic Accident Rate In Palopo City

Penerapan Regresi Nonparametrik Spline Truncated Dalam Memodelkan Angka Kecelakaan Lalu Lintas Di Kota Palopo

Nurjaya Syamsualam[1], Rahmat Hidayat[2*]
[1]Universitas Cokroaminoto Palopo, Jl. Lamaranginang Kota Palopo, Sulawesi Selatan, nurjayasyamsualam.a@gmail.com
[2*]Universitas Cokroaminoto Palopo, Jl. Lamaranginang Kota Palopo, Sulawesi Selatan, dayatmath@gmail.com

ABSTRAK

Suatu kejadian di jalan raya yang dapat menimbulkan berbagai kerugian bagi pelaku maupun kerusakan terhadap benda disebut laka lantas (kecelakaan lalu lintas), yang umumnya timbul karena adanya factor berpengaruh. Penelitian ini bertujuan untuk memodelkan angka kecelakaan lalu lintas di Kota Palopo menggunakan regresi nonparametrik Spline truncated, memperoleh estimasi parameter model regresi, memperoleh metode untuk mencari titik knot ideal. Studi ini adalah regresi nonparametrik Spline truncated. Regresi nonparametrik Spline truncated merupakan salah satu pendekatan nonparametrik yang merupakan transformasi dari fungsi suku banyak berbagian-bagian. Hasil dari penelitian ini diperoleh estimasi parameter model regresi nonparametrik Spline truncated menggunakan metode kuadrat terkecil. Pemilihan titik knot pada model Spline dapat menjelaskan bentuk data (kurva regresi) sehingga model regresi terbaik dicapai berdasarkan pemilihan titik knot optimal dengan nilai GCV minimun. Model regresi Spline truncated yang memiliki nilai GCV terkecil yaitu 136,7351 sehigga disimpulkan bahwa angka kecelakaan lalu lintas di Kota Palopo dipengaruhi variabel x yaitu kepadatan penduduk.
Kata Kunci : Kecelakaan, Regresi Nonparametrik, GCV, Knot.

ABSTRACT

Traffic accident is an incident that occurs on the highway so that can cause various losses of the perpetrator as damage to objects. Traffic accident can caused by factors that have an effects. In this research aims to obtained an estimate of parameters using nonparametric Spline Truncated, obtain method to find the optimal knot point, and model the traffic accident rate in Palopo City using nonparametric Spline Truncated regression. The variable that have an effect is population density (x). Nonparametric Spline Truncated regression is the method we used in this research which one is modification of segmented polynomial function. The results of this research is the parameters estimation of the nonparametric Spline Truncated regression model was obtained using least squares method. The selection of knot point in the Spline model is able to interpret the data pattern (regression curve) so that the best regression model is obtain based on the selection of optimal knot point with a minimum GCV value is 136,7351. Based on that results the conclusion is the variable population density can affect traffict accident rate in Palopo City.
Keywords : Accident, Nonparamtric Regression, GCV, Knot.

How to Cite: Syamsualam, N., & Hidayat, R. (2022). Application Of Truncated Spline Nonparametric Regression In Modeling Traffic Accident Rate In Palopo City. Mathline: Jurnal Matematika dan Pendidikan Matematika , Vol. 7 No. 2, 185 – 196.

DOI: https://doi.org/10.31943/mathline.v7i2.275

PRELIMINARY

The increase in the quantity of use of private and public transportation occurs due to an increase in population and economic growth so that it indirectly causes traffic events such as traffic accidents. Accident cases tend to increase and are difficult to minimize with the number of vehicle movements and the increase or widening of the length of the road (Hobbs, 1995).

Palopo City is a city with a fairly large population with development and growth in various aspects, one of which is the economy. The increasing number of residents and the development of the economy in Palopo City have made the standard of living of people increase. Supported by the human need for vehicles, resulting in an increase in the number of private vehicle ownership in Palopo City. Increased ownership of private vehicles that are not in line with road facilities and infrastructure. According to Siahaan in Rahmayanti (2014) suggests that human negligence while on the road is the biggest cause of traffic accidents.

The bond between the independent variable (x) and the dependent variable (y) can be determined using one of the statistical methods, namely regression analysis. A total of three designs are commonly used in estimating the curve, namely if the data relationship pattern does not show the found regression curve/graph it will use non-pametric regression, if the data relationship pattern shows a known regression graph form it will use parametric regression, and if the data relationship pattern shows curves that are understood and not understood will use semiparametrics.

There is a method that will be used in this study, namely nonparametric regression. Nonparametric regression was used because the model of the relationship between the dependent and independent variables from the graph of the regression function curve was not found. In addition, nonparametric regression with a flexible spline approach in finding a regression curve estimation model on the data with no influence of subjectivity from the researcher. According to Hardle in Bidari and Budiantara (2021) Spline is a segmented polynomial where unequal polynomials are combined at knots and continue so that they are flexible. In the truncated spline, the knot points help to overcome the data that shows a significant up/down pattern so as to produce a relatively smooth curve.

Budiantara in Wijaya (2018) states that there is a transformation of the nature of the graphic pattern at different ranges at the fusion point called the knot point. Selection of the ideal knot point by reviewing the lowest GCV number can obtain the best Spline regression mode.

The objectives of this study are to obtain parameter estimates for nonparametric Spline truncated regression forms, obtain a method for finding the ideal knot point, and obtain a traffic accident rate model in Palopo City using nonparametric Spline truncated regression.

METHODS

Quantitative study is one of the research that will be conducted through the use of secondary data. The secondary data needed in this study were obtained from the Palopo City BPS office and the Palopo City Regional Police office. The dependent variable (y) and the independent variable (x) are the variables in this study. Where the independent variable (x) is population density and the dependent variable (y) is the number of accidents that occurred in Palopo City. The approach that will be used is Spline Truncated nonparametric regression because the data information on the number of traffic accidents in Palopo City and the variables that are thought to be influential do not show a special pattern, and the minimum Generalized Cross Validation value is used when choosing the smallest knot point.

The analysis steps carried out:

a. Describe the number of traffic accidents in Palopo City and the factors that are estimated to be influential.
b. Make descriptive statistics on the dependent variable and the independent variable.
c. Creating a scatterplot (correlation pattern) of traffic accident rates is defined as the dependent variable with population density defined as the independent variable. If there is no definite pattern, the method used is the nonparametric Spline regression approach.

Creating independent variables using Spline nonparametric regression.

a. Form a nonparametric regression form Spline dependent variable using one knot point.
b. The selection of the optimum knot point by taking into account the smallest GCV value.
c. Conducting significant parameter tests simultaneously.
d. Conducting identical assumption test and normality assumption test.
e. Interpreting the study outputs and drawing conclusions.

RESULTS AND DISCUSSION

RESULTS

1. Data Description

a. Characteristics of Traffic Accident Rates in Palopo City and Influential Factors The traffic accident rate in Palopo City and the estimated influential factors have the characteristics shown in table 1 in the form of data containing the average, variance, smallest value and largest value.

Table 1. Characteristics of traffic accidents in Palopo City and the factors that are estimated to be affected

Variable	Means	Varians	Minimum	Maximum
y	21,44	203,78	2,00	49,00
x	1333	1319215	187	3174

Based on table 1 the average traffic accident (y) in Palopo City is 21.44. The variance of the y variable is 203.78. The minimum value of the y variable is 2.00 and the maximum value of the y variable is 49.00. This shows that the range of the largest and smallest traffic accidents in Palopo City is quite far, namely 47.

The characteristic of variable x is population density which has an average of 1333 with a fairly high variance of 1319215, this shows the diversity of the data and the data interval for the variable x is quite large. The smallest population density is located in the Mungkajang sub-district, which is 187 and the largest is located in the Wara Timur sub-district, which is 3174.

b. Scatterplot of Traffic Accident Rates and Factors Estimated To Influence

The following is the form of the relationship between the number of traffic accidents and population density.

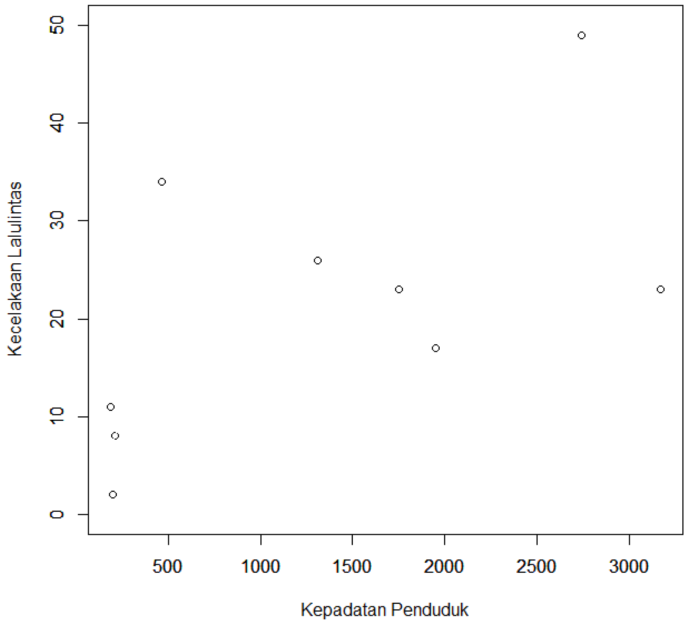

Figure 1. Scatterplot of traffic accident numbers with population density

Figure 1 shows the absence of a bond in the scatterplot of traffic accident numbers with population density. So that the pattern of the relationship between population density and traffic accidents will be modeled with nonparametric regression. Areas with a fairly high population density generally have a large number of accidents due to the large amount of mobility that occurs in the area.

2. Traffic Accident Modeling in Palopo City using SplineTruncated Nonparametric Regression

Based on the results of the initial analysis, it is found that between population density and traffic accidents will be modeled with nonparametric regression. The next step is to determine the optimum knot point with the smallest GCV value.

a. Choosing the Optimum Knot Point

Based on the smallest GCV value, the optimum knot point can be obtained on the traffic accident rate in Palopo City with the variable population density.

The following table is the result of the experiments carried out to find the ideal knot value that matches the smallest GCV value in the one knot point Spline model.

Table 2. Knot points and GCV values

Knot Points	GCV
⋮	⋮
436	136,7351
462	136,8224
461	136,8239
460	136,9102
459	136,9134
458	136,9982
457	137,0036
456	137,0867
455	137,0943
454	137,1755
458	137,1858
457	137,2778
470	137,3542
456	137,3706
471	137,4440
455	137,4640
454	137,5343
472	137,5580
473	137,6248
⋮	⋮

Table 2 is a number of iterations with the minimum optimization value at 463. The knot point with a minimum GCV for the population density variable is 463 with a GCV value of 136.7351. The following is a one knot point curve with minimum GCV that can be seen in figure 2.

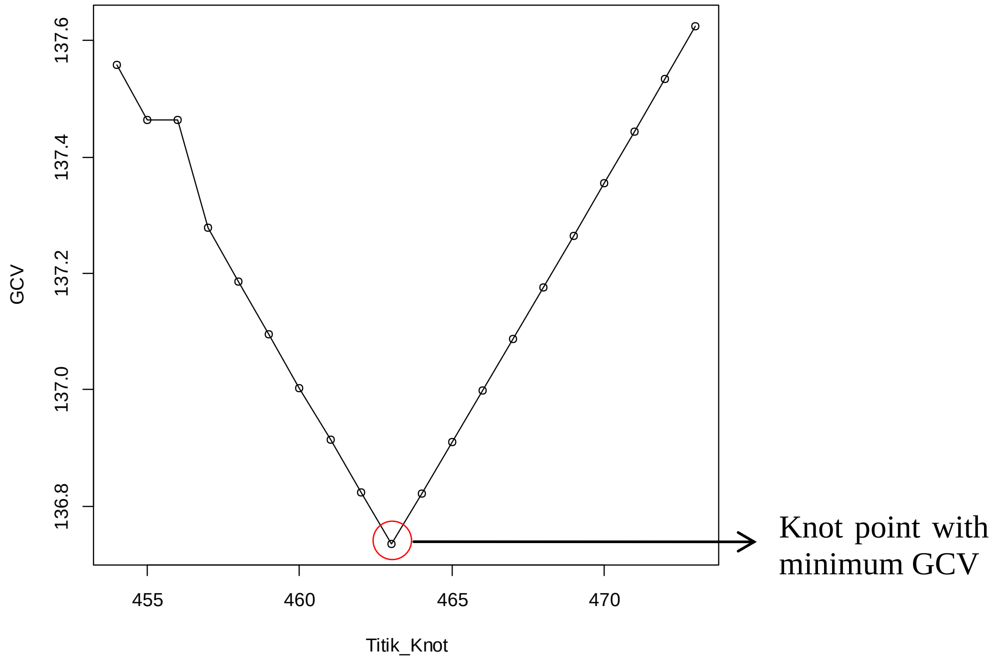

Figure 2. One knot point curve

b. Parameter Estimation

Parameter estimation using the least squares method consists of sorting out the ideal knot point by observing the smallest GCV value as presented in table 3 below.

Table 3. Estimation of spline model parameters

Parameter	**Estimation**
β_0	-8,21282
β_1	0,07669256
β_2	-0,07576433

From table 3 the spline regression equation model can be formed based on the estimated values of the Spline regression model parameters that have been loaded from the optimal knot point at the smallest GCV value. Here's the model that was formed.

$$\hat{y} = -8{,}21282 + 0{,}07669256x - 0{,}07576433(x - 463)_+$$

c. Simultaneous Significance Test

Processing the regression model simultaneously is called a simultaneous test. Following are the hypotheses for simultaneous significant tests:

H_0 : $\beta = 0$ (independent variable does not affect the dependent variable)

H_1 : $\beta \neq 0$ (independent variable affects the dependent variable)

The following is the ANOVA output for the nonparametric Spline truncated regression model.

Table 4. Spline model ANOVA

Model	**SoS**	**df**	**MS**	**F**	**Sig.**
Regression	3,668	1	3,668	7,131	0,032
Residual	3,600	7	0,514		
Total	7,269	8			

Table 4 shows a significant value of as much as 0.032. The value of (significance level) is 0.05 and the rejection area of H0 is carried out if the significant value is > . Because the significant value < α then the decision taken rejects H0. So it can be concluded that the independent variable affects the dependent variable significantly.

d. Residual Assumption Test

The feasibility of the regression model can be determined by testing the residual assumptions. When a regression model with a β significant value reaches the main standard but deviates from the residual assumption, it is not recommended to describe the pattern of binding between the independent variable and the dependent variable.

1) Identical Test

In the identical assumption test, the Glejser test is used to determine homoscedasticity in the residual variance and there is no heteroscedasticity. Glejser test output in the following ANOVA table.

Table 5. Glejser test results

Model	Number of Squares	df	Average Square	F	Sig.
Regresi	0,165	1	0.165	0,938	0,365
Residual	1,229	7	0.176		
Total	1,394	8			

Table 5 shows a significant value of 0.365. With a significant level of α = 0.05 and the rejection area is H_0 if the significant value is <0.05. From this result, the value is significant > α so the decision accepts H_0. So it can be concluded that there is no variance inequality from error for all observations of independent variables in the regression model so that identical assumptions are reached.

2) Normalitas Test

The normality assumption test was conducted to determine the data used were normally distributed using the Kolmogorov-Smirnov test. The hypothesis used is as follows:

H_0 : Residuals are normally distributed

H_1 : Residuals are not normally distributed

Here are the results of the normality test.

Table 6. Normality Test

One-Sample Kolmogorov-Smirnov Test		
		Population density
N		9
Normal Parameters	Std. Deviation	0,67086683
Test Statistic		0,142
Asymp. Sig. (2-tailed)		0,200

After analyzing in table 6, the Asimp value is obtained. Significant (2-tailed) is 0.200. By using the value of α = 0.05 and the rejection area H_0 if significant <α. Because it is significant > α then H_0 is accepted so that the residuals are normally distributed, so it is concluded that the Spline model has reached the normality assumption test. The residual normality plot is presented in the following figure.

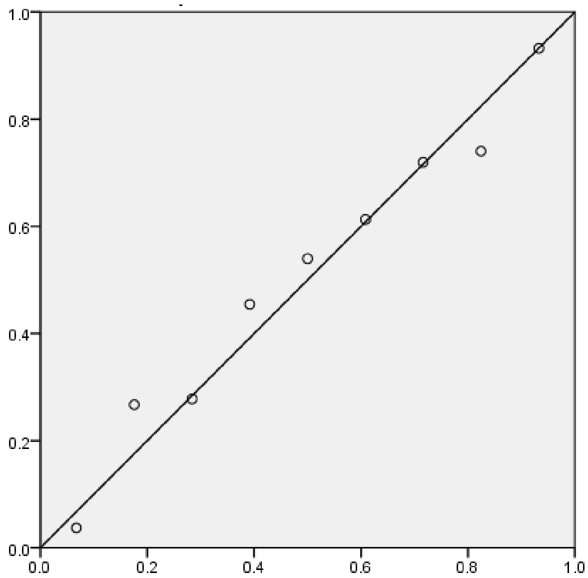

Figure 3. Residual Normality Plot

In Figure 3 it can be seen that the points in the image are data that are around the diagonal line spreading and following the position of the diagonal line, so it can be concluded that the normal assumptions of the Spline model are met.

DISCUSSION

The output of the analysis that has been carried out is that the best Spline nonparametric regression model in traffic accidents in Palopo City is formed at one knot with a GCV value of 136.7351. After entering the knot value and parameters, the best model is selected as follows:.

$$\hat{y} = -8{,}21282 + 0{,}07669256x - 0{,}07576433(x - 463)_+$$

From this model, if the number of traffic accidents in Palopo City is not influenced by population density, then traffic accidents decrease by -8.21282 units.

As for the partial function (truncated):

$$\hat{y} = \begin{cases} 0{,}07669256x, & x < 463 \\ 0{,}00092723x + 35{,}07888479, & x \geq 463 \end{cases}$$

Based on the model that has been obtained, in the first interval for $x < 463$ so $\hat{y} = 0{,}07669256x$ this shows that the population density is less than 463 people, this means that if there is an increase in population density of one person, the traffic accident rate in Palopo City will also increase by 0.07669256 units. There are 9 sub-districts included in this interval, namely East Wara sub-district, Telluwanua sub-district, North Wara sub-district, Wara sub-district, Bara sub-district, South Wara sub-district, West Wara sub-district, Sandana district, Mungkajang sub-district.

In the second interval, for $x \geq 463$ so $\hat{y} = 0{,}00092723x + 35{,}07888479$ This shows that for the population density there are 463 more people or equal to 463 people. So if there is an increase in population density by one person, the number of traffic accidents in Palopo City will also increase by 0.00092723 units. The following Spline regression model is presented in Figure 4.

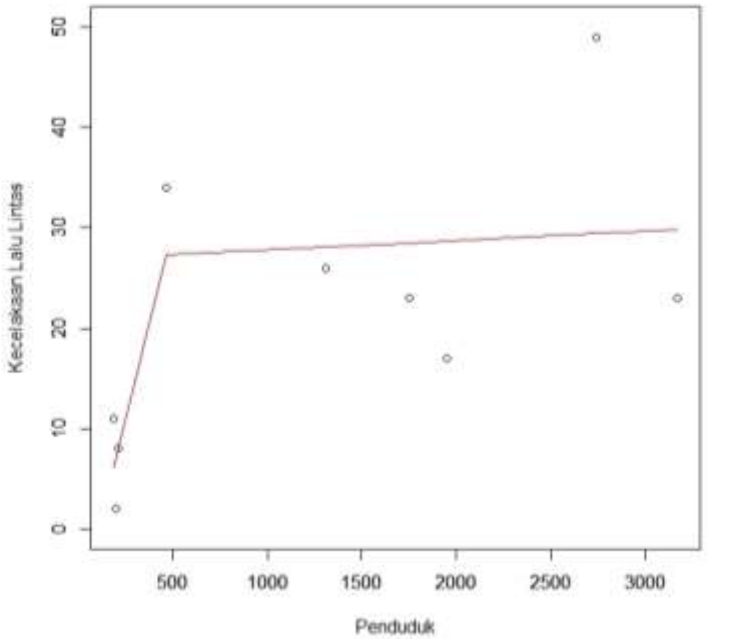

Figure 4. Spline regression model estimation curve.

CONCLUSION

Based on the results that have been obtained, it is concluded that:

1. Parameter estimation $\beta_0 = -8{,}21282$, $\beta_1 = 0{,}07669256$, and $\beta_2 = -0{,}07576433$ obtained based on the nonparametric Spline truncated regression approach using the least squares method.
2. The regression model was obtained based on the selection of the optimal knot point with a minimum Generalize Cross Validation (GCV) value.
3. The formed Spline truncated regression model is $\hat{y} = -8{,}21282 + 0{,}07669256x - 0{,}07576433(x - 463)_+$ by taking into account the smallest GCV value, which is 136.7351 which is in the one-point knot model, it

can be concluded that the number of traffic accidents in Palopo City is influenced by a variable x, namely population density.

REFERENCES

Adrianingsih, N. Y., Budiantara, I. N., & Purnomo, J. D. T. (2021). *Modeling with Mixed Kernel, Spline Truncated and Fourier Series on Human Development Index in East Java*. In IOP Conference Series: Materials Science and Engineering (Vol. 1115, No. 1, p. 012024). IOP Publishing.

Chand, A., Jayesh, S., & Bhasi, A. B. (2021). *Road Traffic Accidents: An Overview Of Data Source, Analysis Techniques And Contributing Factors*. Material Today: Proceedings, 47, 5135-5141.

Dani, A. T. R., Ratnasari, V., & Budiantara, I. N. (2021). *Optimal Knots Point and Bandwidth Selection in Modeling Mixed Estimator Nonparametric Regression*. In IOP Conference Series: Materials Science and Engineering (Vol. 1115, No. 1, p. 012020). IOP Publishing.

Dani, A. T. R., Adrianingsih, N. Y., Ainurrochmah, A., & Sriningsih, R. (2021). *Flexibility of Nonparametric Regression Spline Truncated on Data without a Specific Pattern*. Jurnal Litbang Edusaintech, 2(1), 37-43.

Gujarati, D. (2004). *Basic Economics fourth edition*. McGraw-Hill.Inc: New York

Hidayat, R., Yuliani, Y., & Sam, M. (2018). *Model regresi nonparametrik dengan pendekatan spline truncated*. Prosiding, 3(1).

Hidayat, R., & Ilyas, M. (2021). *Pemodelan angka kemiskinan dengan regresi linier spline berganda*. In Seminar Nasional Variansi (Venue Artikulasi-Riset, Inovasi, Resonansi-Teori, dan Aplikasi Statistika).

Hobbs. (1995). *Perencanaan dan tekni lalu lintas*. Universitas Gadjah Mada.

Igustin, E. D., & Budiantara, I. N. (2021). *Pemodelan Faktor-Faktor yang Mempengaruhi Total Fertility Rate di Indonesia Menggunakan Regresi Nonparametrik Spline Truncated*. Jurnal Sains dan Seni ITS, 9(2), D178-D185.

Kenneth, G. E. (2021). *Statistical Application Of Regression Techniques In Modeling Road Accident In Edo State, Nigeria*. Sch J Phys Math Stat, 1, 14-18.

Mayou, R. (2021). “*The psychiatry of road traffic accidents”*. The Psychiatry Of Road Traffic Accidents. 33-48.

Ningsih, S., & Dukalang, H. H. (2019). *Penerapan Metode Suksesif Interval pada Analsis Regresi Linier Berganda*. Jambura Journal of Mathematics, 1(1), 43-53.

Octavanny, M. A. D., Budiantara, I. N., Kuswanto, H., & Rahmawati, D. P. (2020). *Nonparametric regression model for longitudinal data with mixed truncated spline and fourier series*. In Abstract and Applied Analysis (Vol. 2020). Hindawi.

Poudel, S., Dhungana, S., & Dahal, R. (2021). *Pattern And Cause Of Road Traffic Accidents In Morang District.* Journal of Nepal Health Research Council, 19(1), 87-90.

Rahmayanti, W. (2014). *Pemodelan faktor penyebab kecelakaan lalu lintas dengan metode regresi logistik multinomial.* Surabaya: Institut Teknologi Sepuluh Nopember.

Ramli, M., Ratnasari, V., & Budiantara, I. N. (2020). *Estimation of matrix variance-covariance on nonparametric regression spline truncated for longitudinal data*. In Journal of Physics: Conference Series (Vol. 1562, No. 1, p. 012014). IOP Publishing.

Wijaya, A. F. H. (2018). *Analisis faktor-faktor yang mempengaruhi tingkat pengangguran terbuka (tpt) di provinsi aceh dengan regresi nonparametrik spline truncated*. Disertasi. ITS: *Tidak diterbitkan*

.