Cek Similarity/8. symomath published 22 september 2020).pdf

By Amin Tohari

83600901

Estimating model of the number of HIV and AIDS cases in East Java using bi-response negative binomial regression based on local linear estimator

Cite as: AIP Conference Proceedings 2264, 030003 (2020); https://doi.org/10.1063/5.0023451 Published Online: 22 September 2020

Amin Tohari, Nur Chamidah, and Fatmawati





ARTICLES YOU MAY BE INTERESTED IN

Kernel PCA and SVM-RFE based feature selection for classification of dengue microarray

AIP Conference Proceedings 2264, 030004 (2020); https://doi.org/10.1063/5.0023930

ne effects of hard water consumption on kidney-related diseases

AIP Conference Proceedings 2264, 020009 (2020); https://doi.org/10.1063/5.0023484

Hypertension risk modeling using penalized spline estimator approach based on consumption 🚮 salt, sugar, and fat factors

AIP Conference Proceedings 2264, 030005 (2020); https://doi.org/10.1063/5.0023456



Your Qubits. Measured.

Meet the next generation of quantum analyzers

- Readout for up to 64 qubits Operation at up to 8.5 GHz.
- mixer-calibration-free Signal optimization
 - with minimal latency





AIP Conference Proceedings 2264, 030003 (2020); https://doi.org/10.1063/5.0023451

2264. 030003

Estimating Model of The Number of HIV and AIDS Cases in East Java Using Bi-response Negative Binomial Regression Based on Local Linear Estimator

Amin Tohari^{1, b)}, Nur Chamidah^{2, a)} and Fatmawati^{2, c)}

¹Doctoral Student, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia.

² Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya, Indonesia.

^{a)}Corresponding author: nur-c@fst.unair.ac.id ^{b)}amin.tohari-20, 30 fst.unair.ac.id ^{c)}fatmawati@fst.unair.ac.id

Abstract. The number of HIV and AIDS cases is a count data correlated with each other. The common model to regress two response variables in the form of count data is bi-response Poisson regression. One infraction of assumptions in the Poisson regression is over dispersion. Negative binomial regression is to be solution to this over dispersion case. This study aims to estimate by 4 sponse negative binomial nonparametric regression model based on local linear estimators for modeling the number of HIV and AIDS cases in East Java. The results give the deviance value as the goodness of fit criterion of bi-response negative binomial nonparametric regression model by using local linear estimator of 2.263 which is smaller than the deviance value of the parametric regression model of 3.926. It means that the bi-response negative binomial nonparametric regression model by using local linear estimator is better than that by using parametric regression model approach to explain the relationship between HIV and AIDS cases affected by drug users in East Java.

INTRODUCTION

Regression analysis is one of procedurers in statistic that can be used to describe the functional relationship 22 ween response variables and predictor variables. While the form of regression function known and the relationship between the response variable 20 nd predictor variables are assumed to be certain curves, so parametric approach is appropriate to be used, while the relationship between the response variables and predictor variables is not in accordance to a particular pattern or there is no information about the form of regression, so nonparametric approach is appropriate to be used. The nonparametric regression approach has high usefulness because the regression function is not specified in certain forms, but it is assumed to be smooth. So it is estimated by using smoothing method based on data patterns [1].

Based on the type of response variable, there are various types of regression models. One of the regression models with discrete response variables is Poisson regression model. The response variable (Y) in the Poisson regression model as the meeting the assumption var(Y) = E(Y) that called equidispersion. To avoid the problem of over-dispersion in the Poisson regression model, in the parametric regression model has a rapidly developed regression model that does not assumed a Poisson distribution on the response variable. Some of the research include: [2] have used generalized Poisson to be applied to accident data; [3] and [4] have developed a negative binomial regression. In the case of bi-response, [5] has discussed the bivariate generalized Poisson model with moment estimates and maximum likelihood; [6] and [7] have discussed the bivariate negative binomial regression model.

Nonparametric approaches with continuous responses in bi-response and multi-response cases have also been developed by several researchers, including by [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], and [19]. Nonparametric regression models with discrete type response variables have not been developed by researchers yet.

[20] has used kernel estimators on Poisson regression; [21] has used a spline estimator in Poisson regression to model the relationship between socioeconomic status with mortality; [22] has discussed Poisson regression with a local linear estimator; and [23] has discussed Poisson regression with kernel estimator for machine order claims; [24] used kernel; [25] has used local linear estimator; [26] has used local polynomial estimators in generalized Poisson regression to model infant mortality based on age.

There is no development estimation theory used for negative binomial regression by using local linear estimator approach so far. The local linear approach is good for use at the parsimony principle model, so it is easily interpreted. In this study, bi-response negative binomial regression with local linear estimator approach will be implemented in the data the number of HIV and AIDS cases in East Java with predictors the number of drug users. In Indonesia, reports of new HIV cases have continued to increase every year since it was first reported (1987), in 2017 report data shows that the five provinces with the highest number of infections are East Java, Jakarta, West Java, Central Java and Papua [27]. The Number of HIV case and AIDS case in East Java is a count data that correlate with each other, so bi-response negative binomial regression is an appropriate model to use.

MATERIAL AND METHODS

In this study, the data used are secondary data obtained from the Health Department of East Java consists of 17 regencies. Steps used to estimate the model in modeling the number of HIV cases and AIDS cases in East Java with the number of drug users as predictor variable is as follows, The first step starts with conducting a correlation test between response variables, and then estimates bi-response negative binomial regression model with local linear estimator approach. Two response variables $(y_1 \text{ and } y_2)$ are known to be correlated. Suppose given (x_i, y_{ii}, y_{2i}) which is paired data, with i = 1, 2, ..., n and n is the amount of observation data. Regression models for the response variable y_{1i} and y_{2i} which are discrete random variable type and it is assumed that both response variables have a bivariate negative binomial distribution as follows [28]:

$$f(y_{v}, y_{2i} | x_{i}) = \frac{\Gamma\left(\frac{1}{\alpha} + y_{1i} + y_{2i}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_{1i} + 1)\Gamma(y_{2i} + 1)} \left[\mu_{i}(x_{i})\right]^{y_{i}} \left[\mu_{2}(x_{i})\right]^{y_{i}} \times \alpha^{-\frac{1}{\alpha}} \left(\frac{1}{\alpha} + \mu_{1}(x_{i}) + \mu_{2}(x_{i})\right)^{-\left(\frac{1}{\alpha} + y_{1i} + y_{2i}\right)}$$
(1)

where
$$\mu_i(x_i) = \exp(x_i^T \beta_1)$$
 (2)

$$\mu_2(x_i) = \exp(x_i^T \beta_2). \tag{3}$$

We estimate the parameters β_1 and β_2 in (2) and (3) by using locally weighted maximum likelihood the method. This method is done through the following steps:

1) Taking random samples size n:

$$(y_{1i}, y_{2i}, x_i)$$
 $i = 1, 2, \dots, n$

2) Determining the likelihood function as follows:

$$\ell\left(\underline{\beta}_{i},\underline{\beta}_{2},\alpha,x_{o}\right) = \prod_{i=1}^{n} \left\{ \frac{\Gamma\left(\frac{1}{\alpha} + y_{1i} + y_{2i}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(y_{1i} + 1\right)\Gamma\left(y_{2i} + 1\right)} \left[\mu_{i}\left(\underline{x}_{i}\right)\right]^{\gamma_{i}} \left[\mu_{2}\left(\underline{x}_{i}\right)\right]^{\gamma_{2i}} \times \alpha^{-\frac{1}{\alpha}} \left(\frac{1}{\alpha} + \mu_{1}\left(\underline{x}_{i}\right) + \mu_{2}\left(\underline{x}_{i}\right)\right)^{-\frac{1}{\alpha}+y_{1}+y_{2i}}\right)^{\kappa_{h}(\eta-s_{0})} \right\}$$

$$(4)$$

3) Determining the log-likelihood function as follows:

$$L(\underline{\beta}_{1}, \underline{\beta}_{2}, \alpha, x_{0}) = \sum_{i=1}^{n} K_{h}(x_{i} - x_{0}) \left\{ \ln \Gamma\left(\frac{1}{\alpha} + y_{1i} + y_{2i}\right) - \ln \Gamma\left(\frac{1}{\alpha}\right) - \ln \Gamma\left(y_{1i} + 1\right) - \ln \Gamma\left(y_{2i} + 1\right) + y_{1i} \ln\left(\mu_{1}(x_{i})\right) + y_{2i} \ln\left(\mu_{2}(x_{i})\right) - \frac{1}{\alpha} \ln \alpha - \left(\frac{1}{\alpha} + y_{1i} + y_{2i}\right) \ln\left(\frac{1}{\alpha} + \mu_{1}(x_{i}) + \mu_{2}(x_{i})\right) \right\}$$
(5)

4) Maximizing the log-likelihood function in (5): the first derivative $L(\beta_1, \beta_2, \alpha)$ to β_1 is:

$$\frac{dL\left(\underline{\beta}_{1},\underline{\beta}_{2},\alpha,x_{0}\right)}{d\underline{\beta}_{1}} = \sum_{i=1}^{n} K_{h}(x_{i}-x_{0}) \left\{ \left(\frac{y_{1i}}{\mu_{1}\left(\underline{x}_{i}\right)} - \frac{\left(\alpha^{-1} + y_{1i} + y_{2i}\right)}{\left(\alpha^{-1} + \mu_{1}\left(\underline{x}_{i}\right) + \mu_{2}\left(\underline{x}_{i}\right)\right)} \right\} \underline{x}_{i}\mu_{1}\left(\underline{x}_{i}\right) \right\}$$
(6)

the first derivative $L(\beta_1, \beta_2, \alpha)$ to β_2 is:

$$\frac{dL(\beta_{1},\beta_{2},\alpha,x_{0})}{d\beta_{2}} = \sum_{i=1}^{n} K_{h}(x_{i}-x_{0}) \left\{ \left(\frac{y_{2i}}{\mu_{2}(x_{i})} - \frac{(\alpha^{-1}+y_{1i}+y_{2i})}{(\alpha^{-1}+\mu_{1}(x_{i})+\mu_{2}(x_{i}))} \right) \mu_{2}(x_{i}) x_{i} \right\} \tag{7}$$

while the first derivative $L(\beta_1, \beta_2, \alpha)$ to α is:

$$\frac{dL\left(\underline{\beta}_{1},\underline{\beta}_{2},\alpha,x_{0}\right)}{d\alpha} = \sum_{i=1}^{n} K_{h}(x_{i}-x_{0}) \left\{ \psi\left(\frac{1}{\alpha}+y_{1i}+y_{2i}\right) - \psi\left(\frac{1}{\alpha}\right) + \frac{\ln\left(\frac{1}{\alpha}+\mu_{1}\left(\underline{x}_{i}\right)+\mu_{2}\left(\underline{x}_{i}\right)\right)}{\alpha^{2}} - \frac{\frac{1}{\alpha^{2}}\left[-\frac{1}{\alpha}\ln\alpha - \left(\frac{1}{\alpha}+y_{1i}+y_{2i}\right)\right]}{\left(\frac{1}{\alpha}+\mu_{1}\left(\underline{x}_{i}\right)+\mu_{2}\left(\underline{x}_{i}\right)\right)} \right\} \tag{8}$$

Because of equations (6), (7) and (8) do not provide explicit equations, we use the Newton-Raphson method to obtain the solutions to these equations as a numerical approach. Newton-Raphson iteration is an algorithm that utilizes first-order derivative vectors and second-order derivative matrices of maximized likelihood functions.

RESULTS AND DISCUSSION

Based on Pearson correlation test result, we get p-value of 0.001 that is less than $\alpha = 0.05$. It means that there is a significant correlation between the number of HIV and AIDS cases. Therefore, we can model the number of HIV cases and AIDS cases simultaneously by using bi-response negative binomial regression.

The following figure is plots of number of HIV (1st response) and AIDS (2nd response) cases versus drug users.

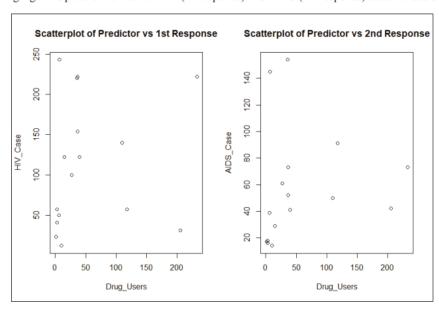


FIGURE. 1. Plots of number of HIV (1st response) and AIDS (2nd response) cases versus drug users.

Figure 1 shows that patterns of the number of HIV and AIDS cases over the drug users spread and they do not follow a specific pattern. So, bi-response negative binomial nonparametric regression is appropriate for modelling them.

In the estimation of nonparametric regression models based on local linear estimators, optimal bandwidth values are needed. The determination of the optimal bandwidth is based on the maximum value of the maximum likelihood cross-validation (MLCV) criteria. Table 1 gives the bandwidth and MLCV values.

TABLE 1. Values of bandwidth and MLCV

Bandwidth	MLCV
10.4	82.13287
10.5	82.26105
10.6	82.37577
10.7	81.35707
10.8	81. 44846

The plot of MLCV versus bandwidth is shown in Fig. 2.

BANDWIDTH PLOT

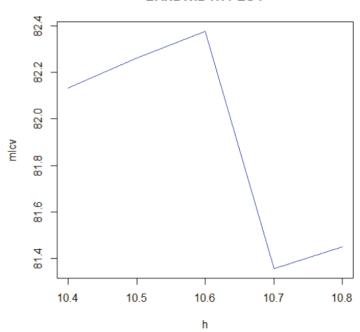


FIGURE. 2. Plot of MLCV versus Bandwidth

Based on Fig. 2, the optimal bandwidth is 10.6, with a maximum MLCV of 82.37577. After obtaining the optimal bandwidth value, we estimate the bi-response negative binomial nonparametric regression model by using local linear estimator. The estimated model has different coefficients depending on the location. One of the estimation results of the model obtained in Pasuruan Regency, which has the highest number of HIV sufferers as follows:

$$\hat{\mu}_1 = \exp(5.064 + 0.00046(x - 7)), 0 < x < 17.6$$
 (9)

$$\hat{\mu}_2 = \exp(4.377 + 0.00041(x - 7)), 0 < x < 17.6$$
(10)

Based on equation (9) we can interpret, for every addition of ten drug users will give a change in the number of HIV cases by 1.0046 times compared with the previous case. Based on equation (10) we can interpret, for every addition of ten drug users will provide a change in the number of AIDS cases by 1.0041 times compared with the previous case.

Goodness of fit test 4 bi-response negative binomial nonparametric regression model by using local linear estimator for the number of HIV and AIDS cases in East Java over the number of drug users gives deviance value of 2.623 which is smaller than $\chi^2_{(\alpha,2^{2n}-(p+1))} = 40.113$. It means that the model is appropriate. Next, we also estimate the number of HIV case and AIDS cases in East Java by using parametric regression model approach. Thus, we get the estimated model as follows:

$$\hat{\mu}_1 = \exp(4.316 + 0.00179x)$$
 and $\hat{\mu}_2 = \exp(3.614 + 0.0032x)$ (11)

based on the estimated model, we obtain deviance value of 3.926.

Based on equations (9) and (10), we can plot the estimated results for the both response variables and the observations shown in Fig. 3, where the red dot is the observation, the blue line is the parametric estimated results, and the green line is nonparametric estimated results.

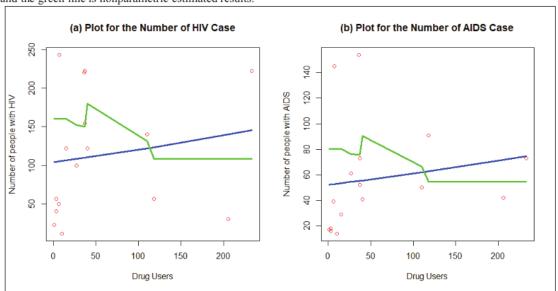


FIGURE. 3. Plots of observations and estimates of (a) HIV cases, and (b) AIDS cases

Based on Fig. 3, the estimated nonparametric regression model by using local linear estimator (green line) is closer to the observations data (red dot) than estimated parametric regression model (blue line). Likewise, based on goodness of fit criterion, i.e., deviance value of the estimated nonparametric regression by using local linear estimator is 2.623, which is less than the deviance value for the estimated parametric regression model of 3.926. It means that bi-response negative binomial nonparametric regression model by using local linear estimator is more appropriate than the estimation results by using the parametric regression model approach.

CONCLUSION

The estimated model we have obtained shows that estimation of the number of HIV and AIDS cases in East Java by using bi-response negative binomial nonparametric regression model based on local linear estimator has satisfied the deviance value as goodness of fit criterion which is less than chi-square value. Based on the estimated model by using local linear estimator we get local models for 17 regencies where they give different effects on the number of HIV and the number of AIDS cases affected by drug users.

ACKNOWLEDGMENT

The authors gratefully acknowledge to Director of the Directorate of Research and Public Service, the Directorate General of Reinforcing of Research and Development, the Ministry of Research, Technology, and 21 her Education of the Republic of Indonesia for funding this research via the Doctoral Dissertation Research Grant (Hibah Penelitian Disertasi Doktor) in the fiscal year 2019.

REFERENCES

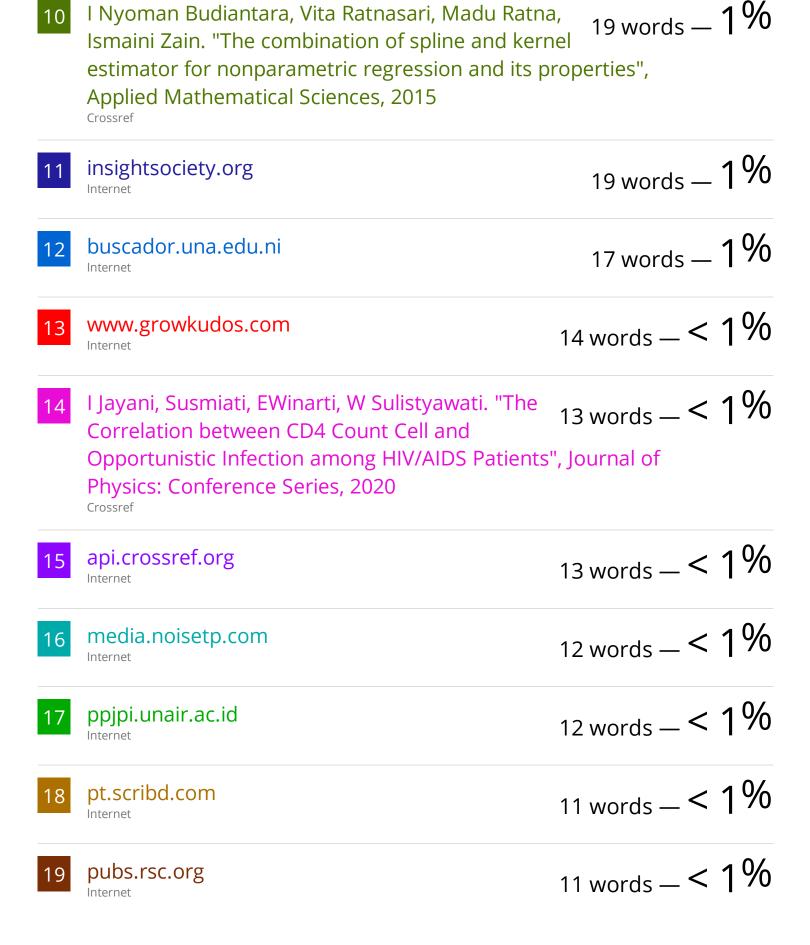
- 15
- [1] 23 L.Eubank, Nonparametric Regress 28 and Spline Smoothing. Second (Marcel Dekker, Inc., Texas, 1999)
- [2] F.Famoye, J.T. Wulu, and K.P.Singh, Journal of Data Science, 2, 287–95 (2004).
- [3] W.Greene, Economics Letters, 99, 585–590 (2008).
- [4] K.Månsson, Economic Modelling, Elsevier B.V. 29, 178–84 (2012).
- [5] F. Famoye, 25 istica Neerlandica, 64, 112–24 (2010).
- [6] F. Famoye, *Journal of Applied Statistics*, **37**, 969–981 (2010).
- [7] A.Tohari, N.Chamidah, and 16 mawati, "Modeling of HIV and AIDS in Indonesia Using Bivariate Negative Binomial Regression", in *IOP Conference Series: Materials Science and Engineering*, Vol. 546 (IOP Publishing, 2019), pp. 052079.
- [8] 7 Chamidah and T.Saifudin, Applied Mathematical Sciences, 7, 1839–1847 (2013).
- [9] 7 Chamidah and B.Lestari, Far East Journal of Mathematical Sciences (FJMS), 100, 1433–53 (2016).
- [10] N. Chamidah and M.Rifada, F. East Journal of Mathematical Sciences (FJMS), 99, 1233–1244 (2016).
- [11] N. Chamidah, and M.Rifada, "Estimation of Median Growth Curves for Children Up Two Years Old Based on Biresponse Local Linear Estimator", in AIP Conference Proceedings (AIP Publishing, 2016), p. 1718 010001.
- [12] B.Lestari, D.Anggraeni, and T.Saifudin, Far East Journal of Mathematical Sciences (FJMS), 108, 341–355 2918).
- [13] B.Lestari, Fatmawati and I.N.Budiantara, "Smoothing Spline Estimator in Multiresponse Nonparametric Regression for Predicting Blood Pressures and Pulse: A Theoretically Discussion", in *Proceeding of Global Conference on Engineering and Applied Science Sapporo, Hokkaido, Japan, (GCEAS, 2019)*, p. 71618, 81–
- [14] B. Lestari, Fatmawati and I.N.Budiantara, Spline Estimator and Its Asymptotic Properties in Multiresponse Nonparametric Regression Model. Songklanakarin Journal of Science and Technology (SJST), (2020) In Press.
- [15] N. Murbarani, Y. Swastika, A. Dwi, B. Aris and N. Chamidah. *Indonesian Journal of Statistics and Its Applications*, **3**, 139–47 (2019)
- [16] N. Chamidah, B.Zaman, L. Muniroh, L. and B. Lestari, "Estimation of Median Growth Charts for Height of Children in East Java Province of Indonesia Using Penalized Spline Estimator", in *Proceeding of Global Conference on Engineering and Applied Science Sapporo, Hokkaido, Japan, (GCEAS, 2019)* p. 71618: 68-78.
- [17] W. Ramadan, N. Chamidah, B. Zaman, B., L. Muniroh and B. Lestari, "Standard Growth Chart of Weight for Height to Determ 6: Wasting Nutritional Status in East Java Based on Semiparametric Least Square Spline Estimator", in *IOP Conference Series: Materials Science and Engineering (IOP Publishing, 2019)*, Vol. 546, p. 052)63.
- [18] A. Islamiyati, Fatmawati and N. Chamidah, Penalized Spline Estimator With Multi-Smoothing Parameters In Bi-Response Multi-Predictor Nonparametric Regression Model For Longitudinal Data. *Songklanakarin Journal of Science and* 17 *hnology*, (2020) **In-press**.
- [19] A. Puspitawati and N. Chamidah, "Coordinate of Property of Pr
- [20] D. Climov, J. Hart and L.Simar, Journal of Nonparametric Statistics, 14, 307–23 (2002)
- [21] J. Young,, P. Graham, and T.Blakely, American Journal of Epidemiology, 164, 282–291(2006)

- J.A. Santos, and M.M. Neves, Metrika, 68, 257-70 (2008) [22]
- J. Shim, an 26. Hwang Journal of the Korean Statistical Society, 40, 1–9. (2011) C.S. Chee, Computational Statistics and [19] Analysis, 109, 34–44. (2017) [23]
- [24]
- Darnah, M.I. Utoyo, and N. Chamidah, IOP Conference Series: Earth and Environmental Science (IOP [25] 10 lishing, 2019), Vol. 243. E.T. Astuti, I.N/ Budiantara, S. Sunaryo, and M. Dokhi, International Journal of Applied Mathematics and
- [26] Statistics, 3492–101(2013).
- Kemenkes. Laporan Perkembangan HIV-AIDS & Infeksi Menular Seksual (IMS) Triwulan IV Tahun 2017, [27]
- S. Cheon, S.H. Song, and B.C. Jung, Journal of the Korean Statistical Society, 38, 185–190. (2009) [28]

Cek Similarity/8. symomath published 22 september 2020).pdf

1	7	7		
		%		
SIMIL	ARITY	INDE		

SIMILARITY INDEX PRIMARY SOURCES				
2	journal.unhas.ac.id Internet	36 words — 1 %		
3	accmes.org Internet	35 words — 1 %		
4	ejournal.warmadewa.ac.id	26 words — 1%		
5	123dok.com Internet	24 words — 1 %		
6	arxiv.org Internet	24 words — 1 %		
7	www.pphmj.com Internet	24 words — 1 %		
8	www.coursehero.com Internet	21 words — 1 %		
9	www.lifescienceglobal.com	21 words — 1 %		



20	Benjamin Lasshof. "5. Results Summary", Walter de Gruyter GmbH, 2018 Crossref	10 words — <	1%
21	ijens.org Internet	10 words — <	1%
22	Zeidler, A "Refinements of empirical models to forecast the shear strength of persistent weak snow layers PART A: Layers of faceted crystals", Co Science and Technology, 200604 Crossref	9 words — <	1%
23	kkgpublications.com Internet	9 words — <	1%
24	www.iaeng.org Internet	9 words — <	1%
25	Xinling Xu, James W. Hardin. "Regression Models for Bivariate Count Outcomes", The Stata Journal: Promoting communications on statistics and State Crossref	8 words — < a, 2018	1%
26	dblp.uni-trier.de Internet	8 words — <	1%
27	hal.archives-ouvertes.fr	8 words — <	1%
28	Abass Mahama, Joseph A. Awuni, Franklin N. Mabe, Shaibu Baanni Azumah. "Modelling adoption intensity of improved soybean production technologies in Ghana - a Generalized Poisson approach to the company of the company		1%

Heliyon, 2020 Crossref

- Rahmat Hidayat, I Nyoman Budiantara, Bambang Widjanarko Otok, Vita Ratnasari. "Kernel-Spline Estimation of Additive Nonparametric Regression Model", IOP Conference Series: Materials Science and Engineering, 2019 Crossref
- Retna Apsari, Kristia Ningsih, M. Yasin Physics. "The utilization of the displacement sensor system to estimate cavities in dentures based on optical imaging", 2016 International Seminar on Sensors, Instrumentation, Measurement and Metrology (ISSIMM), 2016 Crossref
- Yang, Z.. "Testing overdispersion in the zeroinflated Poisson model", Journal of Statistical Planning and Inference, 20090901

 $_{6 \text{ words}}$ -<1%

EXCLUDE QUOTES OFF
EXCLUDE BIBLIOGRAPHY OFF

EXCLUDE MATCHES

OFF