ANALISIS SURVIVAL MODEL REGRESI SEMIPARAMETRIK PADA LAMA STUDI MAHASISWA

by Siti Alfiatur Rohmaniah

Submission date: 31-Aug-2022 12:03AM (UTC-0400)
Submission ID: 1889821782
File name: VIVAL_MODEL_REGRESI_SEMIPARAMETRIK_PADA_LAMA_STUDI_MAHASISWA.pdf (300.47K)
Word count: 2498
Character count: 15223

p-ISSN 2460-9986
e-ISSN 2476-9436

ANALISIS SURVIVAL MODEL REGRESI SEMIPARAMETRIK PADA LAMA STUDI MAHASISWA

Novita Eka Chandra[1)], Siti Alfiatur Rohmaniah[2)]

[1) 2)] Fakultas MIPA, Jurusan Matematika, Universitas Islam Darul Ulum Lamongan, Jl. Airlangga 03 Sukodadi Lamongan; Telp.024-8451279. Email: [1)]novitaeka@unisda.ac.id, [2)] nia0304@gmail.com

Abstrak

Pada analisis survival untuk menentukan hubungan antar variabel digunakan model regresi, salah satunya menggunakan model regresi semiparametrik. Model regresi semiparametrik merupakan model yang tidak memerlukan asumsi atau informasi distribusi data survival. Dengan begitu, model ini lebih fleksibel dalam penggunaannya. Dalam penelitian ini model regresi semiparametrik yang digunakan yaitu model regresi Cox *Proportional Hazard* (Cox PH). Estimasi parameter regresi Cox PH dapat dilakukan tanpa menentukan fungsi *baseline* hazard. Tujuan penelitian ini untuk menentukan faktor-faktor yang mempengaruhi lama studi mahasiswa. Apabila terdapat banyak mahasiswa yang lama studinya tidak tepat waktu, menunjukkan bahwa terjadi ketidakprofesionalan pada bidang akademis dari pendidiknya. Dengan demikian, masyarakat akan menilai rendahnya kualitas dari universitas, sehingga menimbulkan penurunan jumlah mahasiswa yang ingin kuliah di universitas tersebut. Sampel dalam penelitian ini yaitu mahasiswa angkatan 2014 Universitas Islam Darul Ulum Lamongan. Adapun variabel yang digunakan yaitu lama studi mahasiswa, gender, IPK, asal sekolah, organisasi dan bekerja. Berdasarkan hasil uji asumsi *Proportional Hazard* (PH) yang dilakukan, semua variabel independen telah memenuhi asumsi tersebut, sehingga variabel tersebut dapat digunakan ke dalam regresi Cox PH. Setelah dilakukan estimasi parameter dengan regresi Cox PH, faktor IPK dan organisasi mempengaruhi lama studi mahasiswa secara signifikan. Mahasiswa dengan IPK tinggi dan mengikuti organisasi lebih cepat menyelesaikan studinya.

Kata kunci: survival, regresi, semiparametrik, lama studi.

Abstract

In survival analysis to determine the relationship between variables is used a regression model, one of which uses the semiparametric regression model. The semiparametric regression model is a model that does not require assumptions or information on survival data distribution. That way, this model is more flexible in its use. In this study, the semiparametric regression model used the Cox Proportional Hazard (Cox PH) regression model. Estimation of Cox PH regression parameters can be done without determining the function baseline hazard. The purpose of study is to determine the factors that influence the duration of student studies. If there are many students whose studies have not been on time, it shows that there is a lack of professionalism in the academic field of the educator. Thus, community will assess the low quality of the university, resulting in a decrease in the number of students who want to study at the university. The samples in this study were students of class 2014 Universitas Islam Darul Ulum Lamongan. The variables have used the length of study for students, gender, GPA, school origin, organization, and work. Based on the results of the assumption Proportional Hazard (PH) conducted, all independent variables have fulfilled these assumptions, so that these variables can be used in Cox PH regression. After parameter estimation by Cox PH regression, the GPA and organizational factors significantly influence the duration of student study. Students with high GPA and participating in organizations more quickly complete their studies.

Keywords: *survival, regression, semiparametric, duration of the study.*

1. PENDAHULUAN

Analisis survival secara umum menurut (Kleinbaum dan Klein, 2012) merupakan prosedur statistika untuk menganalisis data yang variabelnya berupa waktu dari awal pengamatan hingga terjadinya suatu peristiwa. Analisis survival memiliki tujuan untuk menentukan hubungan antar variabel terhadap waktu, dengan waktu sebagai variabel dependen, dan memperkirakan fungsi *hazard* maupun fungsi survivalnya. Menurut (Chandra dkk, 2015) untuk menentukan hubungan antar variabel yaitu variabel independen dan dependen digunakan analisis regresi. Terdapat beberapa model regresi dalam analisis survival diantaranya model regresi parametrik dan model regresi semiparametrik. Model regresi parametrik merupakan model regresi yang didasarkan pada distribusi data, dimana fungsi *baseline hazard* mengikuti suatu distribusi dari data waktu survivalnya. Sebaliknya, model regresi semiparametrik merupakan model yang tidak berdasarkan pada bentuk distribusi data. Artinya, model regresi semiparametrik lebih fleksibel untuk digunakan. Sebelumnya, (Chandra dan Rohmaniah, 2019) meneliti mengenai analisis survival menggunakan model regresi parametrik pada data lama studi mahasiswa. Selanjutnya, pada penelitian ini dilakukan pengembangan dari model regresi parametrik yaitu dengan menggunakan model regresi semiparametrik, yaitu model regresi *Cox Proportional Hazard* (Cox PH). Regresi semiparametrik Cox PH tidak memerlukan fungsi *baseline hazard* yang terdapat pada model (Riyangkaryaman, 2017), dimana fungsi

p-ISSN 2460-9986
e-ISSN 2476-9436

baseline hazard adalah fungsi yang melibatkan waktu teteapi tidak melibatkan kovariat. Selain itu, model regresi semiparametrik Cox PH merupakan model yang aman untuk digunakan dan tidak ada ketakutan salah pilih model parametrik. Tujuan dari penelitian ini untuk menentukan faktor-faktor yang mempengaruhi lama studi mahasiswa dengan menggunakan regresi semiparametrik Cox PH. Lama studi mahasiswa sangatlah penting, karena apabila terdapat banyak mahasiswa yang lama studinya tidak tepat waktu, menunjukkan bahwa terjadi ketidakprofesionalan pada bidang akademis dari pendidiknya. Akibatnya, masyarakat akan menilai rendahnya kualitas dari universitas, bahkan menimbulkan penurunan jumlah mahasiswa yang ingin kuliah di universitas tersebut. Hasil dari penelitian ini diharapkan dapat dijadikan pertimbangan universitas untuk mengkontrol faktor penghambat lama studi mahasiswa, sehingga mahasiswa dapat lulus tepat waktu.

2. METODE

Data yang digunakan dalam analisis survival ini adalah data lama studi mahasiswa Universitas Islam Darul Ulum Lamongan (Chandra dan Rohmaniah, 2019). Adapun faktor-faktor yang dapat mempengaruhi lama studi mahasiswa diantaranya gender, IPK, asal sekolah, organisasi dan bekerja. Dalam penelitian ini, variabel dependennya adalah lama studi, sedangkan variabel independennya adalah gender, IPK, asal sekolah, organisasi dan bekerja. Metode yang digunakan dengan model regresi semiparametrik yaitu Cox PH. Penggolahan data menggunakan software R 3.5.2. Adapun langkah-langkahnya sebagai berikut :

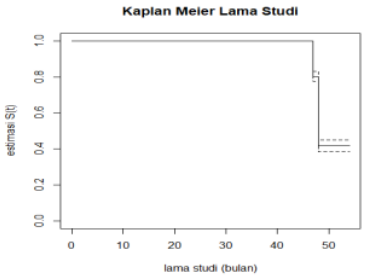
1. Pengujian Asumsi *Proportional Hazard* (PH)
   Pengujian asumsi PH menurut Collett (2015) ada dua cara yaitu menggunakan grafik *Log Minus Log* (LML) Survival dan menggunakan uji *Goodness of Fits*. Menurut (Kleinbaum dan Klein, 2010), jika grafik atau plot LML Survival antar kategori dalam satu variabel independen sejajar atau tidak saling berpotongan, maka asumsi PH terpenuhi dan variabel independen yang bersifat kategori dapat digunakan dalam model regresi semiparametrik Cox PH. Pengujian *Goodness of Fits* dilakukan untuk variabel independen yang berpengaruh terhadap model akhir regresi semiparametrik Cox PH yang didapatkan (Fitriani, 2018). Berikut hipotesis uji asumsi PH dengan menggunakan *Goodness of Fits* (Kleinbaum dan Klein, 2010):
   $H_0 : \rho = 0$ (asumsi PH terpenuhi)
   $H_1 : \rho \neq 0$ (asumsi PH terpenuhi)
   Asumsi PH dikatakan terpenuhi apabila *p-value* lebih besar dari 0,05.
2. Analisis Regresi Semiparametrik Cox PH
   Menurut (Collett, 2015), bentuk persamaan model regresi semiparametrik Cox PH yaitu
   $$h(t,X) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p) \quad (1)$$
   dengan $h_0(t)$ merupakan fungsi *baseline hazard*.
3. Pengujian Signifikansi Parameter
   Pengujian signifikansi parameter dilakukan dengan dua tahap, yaitu uji keseluruhan dan uji parsial. Berikut penjelasannya (Collett, 2015)
   a) Uji keseluruhan
      Uji keseluruhan digunakan untuk menentukan ada atau tidak pengaruh signifikan antara semua variabel independen terhadap model secara bersama-sama. Uji keseluruhan menggunakan uji *partial likelihood ratio* dan berikut bentuk hipotesisnya :
      $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$
      H_1 : minimal ada satu $\beta_j \neq 0, j = 1,2,\ldots,p$
      Apabila *p-value* lebih kecil dari 0,05, maka H_0 ditolak. Artinya, semua variabel independen bersama-sama mempengaruhi variabel dependen. Dengan kata lain, semua variabel independen mempengaruhi secara signifikan terhadap model.
   b) Uji parsial
      Uji parsial digunakan untuk menentukan ada atau tidak pengaruh signifikan antara setiap variabel independen terhadap variabel dependen. Uji parsial menggunakan uji *Wald* dan berikut bentuk hipotesisnya :
      $H_0 : \beta_j = 0$
      $H_1 : \beta_j \neq 0, j = 1,2,\ldots,p$
      Apabila *p-value* lebih kecil dari 0,05, maka H_0 ditolak. Artinya, variabel independen bersama-sama mempengaruhi variabel dependen. Dengan kata lain, variabel independen mempengaruhi secara signifikan terhadap model.
4. Interpretasi Model
   Pada Persamaan (1) model regresi semiparametrik Cox PH, fungsi *baseline hazard* tidak diketahui bentuknya. Untuk menginterpretasikan model dapat menggunakan nilai *Hazard Ratio* (HR). Hal ini dikarenakan nilai HR tidak bergantung pada fungsi *baseline hazard*. Nilai HR digunakan untuk menentukan resiko kegagalan suatu individu. Nilai HR dapat dinyatakan dalam bentuk
   $$HR = \frac{\hat{h}(t,X^*)}{h(t,X)} = \exp\left(\sum_{j=1}^{p} \beta_j (X_j^* - X_j)\right).$$
   Nilai HR berkisar dari nol sampai dengan tak berhingga. Interpretasi nilai HR sebagai berikut (Harlan, 2017):

p-ISSN 2460-9986
e-ISSN 2476-9436

a) HR < 1 mengindikasikan bahwa kovariat X merupakan faktor preventif terhadap terjadinya kegagalan.
b) HR = 1 mengindikasikan bahwa tidak ada asosiasi antara kovariat X dengan kejadian kegagalan.
c) HR > 1 mengindikasikan bahwa kovariat X merupakan faktor risiko terhadap terjadinya kegagalan.

3. HASIL DAN PEMBAHASAN

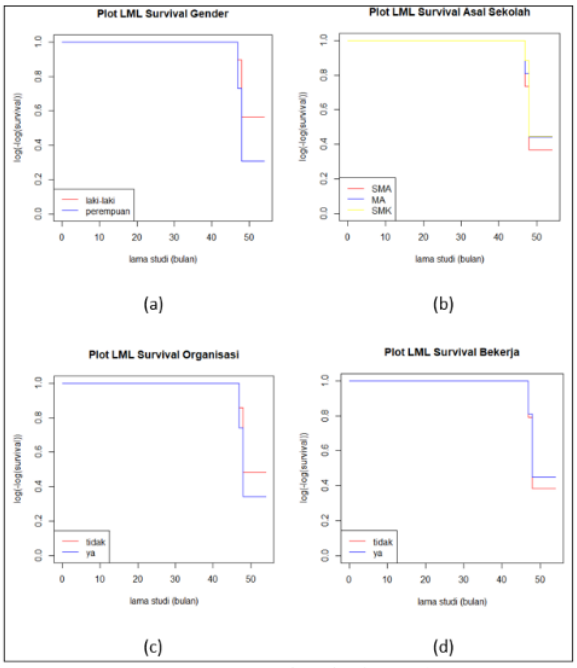
Uji pertama yang dilakukan adalah uji Kaplan Meier. Hal ini bertujuan untuk mencari estimasi fungsi survival $S(t)$ dan grafik fungsi survival lama studi mahasiswa. Berikut ini ditampilkan plot data antar kejadian lama studi mahasiswa dengan menggunakan software R.



Gambar 1. Kaplan Meier Lama Studi

Grafik pada Gambar 1 menunjukan grafik $S(t)$ dari lama studi mahasiswa. Dari grafik terlihat bahwa lama studi mahasiswa adalah 48 hari. Selain itu, berdasarkan hasil output tabel Kaplan Meier, diketahui bahwa mediannya sebesar 48 bulan, sehingga dapat dikatakan bahwa median lama studi mahasiswa 48 bulan.

Selanjutnya dilakukan uji asumsi PH untuk variabel independen yang bersifat kategorik (dalam penelitian ini yaitu variabel gender, asal sekolah, organisasi dan bekerja). Uji asumsi PH menggunakan grafik plot *Log Minus Log* (LML) Survival. Perhatikan Gambar 2 di bawah ini, terlihat bahwa plot LML Survival untuk semua variabel membentuk garis sejajar pada setiap kategorinya dan tidak ada yang saling berpotongan. Dengan demikian, asumsi PH sudah terpenuhi. Akibatnya, semua variabel independen yang bersifat kategorik dapat digunakan ke dalam model regresi semiparametrik Cox PH.



Gambar 2. Plot *Log Minus Log Survival* (a) Gender; (b) Asal Sekolah; (c) Organisasi; (d) Bekerja

p-ISSN 2460-9986
e-ISSN 2476-9436

Diketahui bahwa semua variabel independen telah memenuhi asumsi PH, selanjutnya dilakukan estimasi parameter dengan model regresi semiparametrik Cox PH. Hasil estimasi parameter dapat dilihat pada Tabel 1 di bawah ini.

Tabel 1. Estimasi Parameter Model Awal

Variabel	*Coef*	*Exp(coef)*	*P-value*	Keputusan
Gender_pr	0,050242	1,051526	0,63839	Ho diterima
IPK	3,879979	48,423182	<2e-16	Ho ditolak
Asal sekolah_MA	-0,439700	0,644229	4,06e-05	Ho diterima
Asal sekolah_SMK	-0,086132	0,917473	0,46202	Ho diterima
Organisasi_ya	0,246066	1,278984	0,00857	Ho ditolak
Bekerja_ya	-0,008619	0,991418	0,92729	Ho diterima
Likelihood ratio test = 587,2 on 6 df, p=<2e-16				

Kemudian dilakukan uji secara keseluruhan untuk menentukan variabel independen berpengaruh secara keseluruhan terhadap model. Berdasarkan hasil estimasi parameter pada Tabel 1, terlihat bahwa p-value <2e-16 lebih kecil dari taraf signifikansi 0,05, artinya seluruh variabel independen berpengaruh secara signifikan terhadap model, sehingga bentuk model awal regresi semiparametrik Cox PH sebagai berikut :

$$h(t,X) = h_0(t)\exp(0{,}050242X_{gender_pr} + 3{,}879979X_{IPK} - 0{,}439700X_{asal_sekolah_MA} - 0{,}086132X_{asal_sekolah_SMK} + 0{,}246066X_{organisasi_ya} - 0{,}008619X_{bekerja_ya}).$$

Selanjutnya dilakukan uji secara parsial untuk mengetahui ada atau tidak pengaruh yang signifikan variabel independen terhadap variabel dependen. Pada Tabel 1 terlihat bahwa *p-value* untuk variabel IPK dan organisasi_ya kurang dari 0,05, maka H_0 ditolak artinya variabel IPK dan organisasi_ya berpengaruh secara signifikan terhadap variabel lama studi. Sebaliknya, variabel gender_pr, asal sekolah_MA, asal sekolah_SMK dan bekerja_ya tidak berpengaruh secara signifikan terhadap variabel lama studi karena *p-value* lebih besar dari 0,05 maka H_0 diterima. Dengan begitu, disimpulkan bahwa tidak semua variabel independen berpengaruh terhadap model, sehingga dilakukan eliminasi *backward* untuk mendapatkan model terbaik.

Setelah dilakukan analisis ulang dan eliminasi *backward* dilakukan diperoleh model terbaik regresi semiparametrik Cox PH. Adapun hasil estimasi parameter untuk model terbaik sebagai berikut.

Tabel 2. Estimasi Parameter Model Terbaik

Variabel	*Coef*	*Exp(coef)*	*P-value*	Keputusan
IPK	3,77052	43,40271	<2e-16	Ho ditolak
Organisasi_ya	0,24529	1,27800	0,00718	Ho ditolak
Likelihood ratio test = 568,4 on 2 df, p=<2e-16				

Diperoleh bentuk model regresi semiparametrik Cox PH terbaik yaitu

$$h(t,X) = h_0(t)\exp(3{,}77052X_{IPK} + 0{,}24529X_{organisasi_ya}) \quad (2)$$

Terlihat pada Tabel 2, secara keseluruhan variabel independen berpengaruh signifikan terhadap model, hal ini dibuktikan dengan nilai *p-value* <2e-16 lebih kecil dari 0,05. Secara parsial, variabel IPK dan organisasi_ya memiliki nilai *p-value* kurang dari 0,05, maka H_0 ditolak artinya variabel IPK dan organisasi_ya berpengaruh signifikan terhadap variabel lama studi. Dengan demikian, variabel IPK dan organisasi_ya mempengaruhi secara signifikan terhadap model. Karena secara keseluruhan dan parsial variabel IPK dan organisasi_ya mempengaruhi model, maka dapat dikatakan model telah layak untuk digunakan, sehingga dapat dilakukan interpretasi dari model yang diperoleh.

Namun, sebelum dilakukan interpretasi model, terlebih dahulu dilakukan uji asumsi PH untuk variabel independen yang mempengaruhi model dengan menggunakan uji *Goodness of Fits*. Berikut hasil uji *Goodness of Fits*.

p-ISSN 2460-9986
e-ISSN 2476-9436

Tabel 3. Uji *Goodness of Fits*

Variabel	*P-value*	Keputusan
IPK	0,540	Ho diterima
Organisasi_ya	0,121	Ho diterima
Global p=0,223		

Terlihat pada Tabel 3, nilai p-value untuk variabel IPK dan organisasi_ya lebih besar dari 0,05, maka H_0 diterima artinya asumsi PH terpenuhi. Karena asumsi PH sudah terpenuhi, maka interpretasi model regresinya dengan menggunakan nilai HR dari masing-masing faktor yang mempengaruhi lama studi mahasiswa. Dalam hal ini nilai HR ditunjukan oleh nilai $\exp(\beta_j)$. Berdasarkan hasil pada Tabel 2, terlihat bahwa :

a. Untuk variabel IPK, nilai HRnya sebesar 43,40271 lebih besar dari 1 menunjukkan bahwa mahasiswa yang memiliki IPK tinggi lebih cepat menyelesaikan studinya dibandingkan mahasiswa yang memiliki IPK rendah.
b. Untuk variabel organisasi_ya, nilai HRnya sebesar 1,27800 lebih besar dari 1 menunjukkan bahwa mahasiswa yang mengikuti organisasi memiliki lama studi lebih singkat dibandingkan mahasiswa yang tidak mengikuti organisasi.

4. SIMPULAN

Berdasarkan analisis data dan pembahasan yang dilakukan diperoleh kesimpulan bahwa Persamaan (2) menunjukkan model regresi semiparametrik Cox PH dari faktor-faktor yang mempengaruhi lama studi mahasiswa. Adapun faktor-faktor yang mempengaruhi lama studi mahasiswa yaitu IPK dan organisasi. Mahasiswa yang memiliki IPK tinggi dan mengikuti organisasi lebih cepat menyelesaikan studinya, hal ini terlihat dari nilai HR yang diberikan lebih besar dari satu..

5. UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada DRPM Kemenristek Dikti yang telah membiayai penelitian ini melalui dana Hibah Penelitian Dosen Pemula tahun anggaran 2019.

6. DAFTAR PUSTAKA

Chandra, N. E. dan Rohmaniah, S. A. 2015. Regresi Nonparametrik Kernal Adjusted. Jurnal Matematika dan Pendidikan Matematika. 7(1) : 1-10.

Chandra, N. E. dan Rohmaiah, S. A. 2019. Analisis Survival Model Regresi Parametrik Lama Studi Mahasiswa. Jurnal Matematika. 9(1) : 1-10.

Collett, D. 2015. Modelling Survival Data in Medical Research. Chapman and Hall, USA.

Fitriani, I. D. 2018. Analisis Regresi Cox Proportional Hazard pada Identifikasi Faktor-Faktor yang Mempengaruhi Lama Studi Mahasiswa S1 FMIPA Universitas Islam Indonesia. Skripsi. Universitas Islam Indonesia. Yogyakarta

Harlan, J. 2017. Analisis Survival. Gunadarma. Jakarta.

Kleinbaum, D. G. dan Klein, M. 2010. Survival Analisis Third Edition. New York : Springer.

Riangkaryaman. 2017. Model Semiparametrik Cox PH dan Parametrik (Weibull dan Lognormal) Dalam Analisis Survival (Kasus : Rawat Inap Penderita DBD di RSUD Haji Kota Makassar). Skripsi. Universitas Negeri Makassar. Makassar.

ANALISIS SURVIVAL MODEL REGRESI SEMIPARAMETRIK PADA LAMA STUDI MAHASISWA

ORIGINALITY REPORT

15%	15%	6%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.lib.ntua.gr Internet Source	1%
2	repository.usd.ac.id Internet Source	1%
3	uim.ac.id Internet Source	1%
4	Submitted to State Islamic University of Alauddin Makassar Student Paper	1%
5	id.scribd.com Internet Source	1%
6	123dok.com Internet Source	<1%
7	Submitted to Universitas Pelita Harapan Student Paper	<1%
8	digilib.unsri.ac.id Internet Source	<1%

e-journal.ikip-veteran.ac.id

9	Internet Source	<1 %
10	journal.binadarma.ac.id Internet Source	<1 %
11	siteresources.worldbank.org Internet Source	<1 %
12	www.ejurnal.its.ac.id Internet Source	<1 %
13	docobook.com Internet Source	<1 %
14	ejournal.unibabwi.ac.id Internet Source	<1 %
15	iopscience.iop.org Internet Source	<1 %
16	eprints.perbanas.ac.id Internet Source	<1 %
17	koreascience.or.kr Internet Source	<1 %
18	repository.itk.ac.id Internet Source	<1 %
19	repository.pelitabangsa.ac.id Internet Source	<1 %
20	www.coursehero.com Internet Source	<1 %

21	digilib.uinsgd.ac.id Internet Source	<1 %
22	ejournal.unp.ac.id Internet Source	<1 %
23	repository.ar-raniry.ac.id Internet Source	<1 %
24	repository.trisakti.ac.id Internet Source	<1 %
25	repository.unri.ac.id Internet Source	<1 %
26	senatik.upgris.ac.id Internet Source	<1 %
27	www2.stat.unibo.it Internet Source	<1 %
28	Dspace.Uii.Ac.Id Internet Source	<1 %
29	Meykel Tulong, Charles Mongi, Mans Mananohas. "Regresi Logistik Multinomial Untuk Menentukan Faktor-Faktor Yang Mempengaruhi Pilihan Perguruan Tinggi Pada Siswa SMA dan SMK di Pulau Karakelang Kabupaten Kepulauan Talaud", d'CARTESIAN, 2018 Publication	<1 %
30	core.ac.uk	

Internet Source

<1%

31	eprints.iain-surakarta.ac.id Internet Source	<1%
32	journal.uny.ac.id Internet Source	<1%
33	ocs.unud.ac.id Internet Source	<1%
34	www.graduate.ubru.ac.th Internet Source	<1%
35	Novianti, Setyo Wira Rizki. "ANALISIS REGRESI LOGISTIK MULTINOMIAL PADA PEMILIHAN ALAT KONTRASEPSI WANITA (Studi Kasus di Puskesmas Sungai Kakap)", Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya, 2019 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		