

Cox Proportional Hazard Regression Analysis to Identify Factors Influencing Student Study Period

ATTILAH SUCI ANGGRAIRA¹, MUHAMMAD BAYU NIRWANA², SRI SUBANTI³

¹*Universitas Sebelas Maret, Jawa Tengah, Indonesia*
attilahsuci@student.uns.ac.id

²*Universitas Sebelas Maret, Jawa Tengah, Indonesia*
mbnirwana@staff.uns.ac.id

³*Universitas Sebelas Maret, Jawa Tengah, Indonesia*
sri_subanti@yahoo.co.id

Abstract

The length of study is defined as the amount of time students need to complete their educational program, from the time they enter college to the end of college. Usually, students are considered to graduate on time if they complete less than four years of education. By using survival analysis using the Cox Proportional Hazard regression model, this study aims to find the components that influence the length of student study. This study used data from 146 students who were randomly selected enrolled at Sebelas Maret University, Faculty of Mathematics and Natural Sciences, batch 2017 and 2018. The research variables consisted of nine predictor variables that were thought to influence the length of study of students. The nine factors are: student's gender, major, GPA, university entrance, scholarship, part-time job, region of origin, organization, and committee. The results show that majors and GPA significantly influence the duration of a student's study. The results of this study have important benefits for higher education administrators and policy makers to develop strategies that can help students complete study programs on time.

Keywords: survival analysis, cox regression, length of study, student, influencing factors



Copyright © 2023 The Author(s)

This is an open-access article under the CC BY-SA license.

Analisis Regresi Cox Proportional Hazard untuk Mengidentifikasi Faktor-Faktor yang Memengaruhi Lama Studi Mahasiswa

Abstrak

Lama studi didefinisikan sebagai jumlah waktu yang dibutuhkan mahasiswa untuk menyelesaikan program pendidikan mereka, mulai dari awal masuk kuliah hingga selesainya kuliah. Biasanya, mahasiswa dianggap lulus tepat waktu apabila mereka menyelesaikan pendidikan kurang dari empat tahun. Dengan menggunakan analisis survival yang menggunakan model regresi Cox Proportional Hazard, penelitian ini bertujuan untuk menemukan komponen yang memengaruhi lamanya studi siswa. Penelitian ini menggunakan data dari 146 mahasiswa yang dipilih secara acak yang terdaftar di Universitas Sebelas Maret Fakultas MIPA angkatan 2017 dan 2018. Variabel penelitian terdiri dari sembilan variabel prediktor yang diduga memengaruhi lamanya studi mahasiswa. Sembilan faktor tersebut: jenis kelamin mahasiswa, jurusan, IPK, jalur masuk universitas, beasiswa, kerja paruh waktu, asal daerah, organisasi, dan kepanitiaan. Hasil menunjukkan bahwa jurusan dan IPK memengaruhi secara signifikan durasi studi siswa. Hasil dari penelitian ini memiliki manfaat penting bagi administrator dan pembuat kebijakan perguruan tinggi untuk mengembangkan strategi yang dapat membantu mahasiswa menyelesaikan program studi dengan tepat waktu.

Kata kunci: analisis survival, regresi cox, lama studi, mahasiswa, faktor-faktor yang memengaruhi

PENDAHULUAN

Salah satu indikator keberhasilan perguruan tinggi dalam memenuhi tugasnya untuk memberikan pendidikan tinggi kepada masyarakat adalah lama studi mahasiswanya. Apabila lama studi berlangsung terlalu lama, hal ini dapat menyebabkan mahasiswa kehilangan semangat belajar, meningkatkan biaya pendidikan, dan memperbesar risiko putus sekolah. Fakultas perguruan tinggi memiliki banyak variabel yang memengaruhi kelulusan. Beberapa di antaranya adalah faktor internal, yang berasal dari mahasiswa sendiri, seperti tingkat kecerdasan, kemampuan belajar, dan motivasi mereka. Sementara itu, faktor eksternal berasal dari luar siswa, seperti kualitas pengajaran, kurikulum, kondisi lingkungan mereka, ketersediaan fasilitas sarana prasarana, dan lain sebagainya. Selain itu, diperkirakan bahwa elemen-elemen ini dapat memengaruhi jumlah waktu yang dihabiskan siswa untuk menyelesaikan studi mereka (Landong et al., 2014). Studi terdahulu menunjukkan bahwa lama studi mahasiswa dipengaruhi oleh

berbagai faktor seperti status organisasi, status pekerjaan, dan status nikah. Namun, penelitian yang dilakukan di Indonesia tentang faktor yang mempengaruhi durasi studi mahasiswa masih sangat terbatas, khususnya menggunakan model *survival analysis* (M. Z. Warli, 2017).

Analisis tahan hidup atau analisis survival merupakan metode statistik yang menganalisis data yang berhubungan dengan waktu, mulai dari pengamatan pertama penelitian yang sudah ditetapkan hingga pengamatan jika terjadi suatu kejadian atau pengamatan terakhir penelitian (Kleinbaum & Klein, 2012). Pengamatan pertama dalam penelitian yaitu pada saat terjadinya peristiwa pertama seperti awal memulai pendidikan atau didiagnosis penyakit, sedangkan pengamatan terakhir penelitian seperti keluar (*drop off*), meninggal, kelulusan, atau peristiwa lain yang relevan (Fitriani, 2018).

Tujuan utama analisis survival adalah untuk memahami dan mengestimasi hubungan antara variabel prediktor dan waktu yang dibutuhkan untuk terjadi peristiwa atau kejadian tertentu. (Klein & Moeschberger, 2003). *Cox Proportional Hazard* merupakan salah satu teknik analisis survival. Regresi Cox adalah teknik analisis statistik yang digunakan untuk menentukan faktor risiko yang memengaruhi waktu terjadinya suatu peristiwa. Model ini dapat digunakan untuk memodelkan waktu kejadian kematian, kelulusan, kejadian penyakit, atau kejadian lainnya.

Model regresi Cox didasarkan pada fungsi hazard, yaitu tingkat kejadian pada suatu waktu tertentu. Fungsi hazard dapat didefinisikan sebagai rasio probabilitas kejadian pada waktu tertentu dengan jumlah orang yang masih berisiko pada waktu itu. Dalam model regresi Cox, fungsi hazard diasumsikan sebagai kombinasi linear dari variabel prediktor (Kleinbaum & Klein, 2012). Misalnya, jika T merupakan variabel random tak negatif pada interval $[0, \infty]$ yang menunjukkan waktu seseorang sampai mengalami kejadian dalam suatu populasi, maka kemungkinan bahwa seseorang mengalami kejadian pada interval $(t+\Delta t)$ dijelaskan oleh fungsi hazard $h(t)$ (Iskandar, 2015).

Untuk dimensi fraktal, perkiraan yang tepat diterapkan. Ini adalah metode Maximum Likelihood Estimation (MLE), suatu model estimasi yang dapat mengatasi kelemahan metode demografi konvensional. Nilai MLE adalah nilai yang memaksimumkan fungsi likelihood n , dan nilai yang paling penting di MLE adalah nilai yang memaksimumkan probabilitas kejadian hasil sampel. Di sisi lain, metode estimasi MLE sangat populer untuk mengestimasi parameter distribusi (Dewi et al., 2019).

Dalam hal ini, model *Cox proportional Hazard* digunakan untuk menentukan apakah variabel prediktor memiliki dampak pada waktu survival, yang berdampak pada durasi studi siswa di Universitas Sebelas Maret (Setiani & Santoso, 2019). Penelitian ini memiliki tujuan untuk menghitung berapa lama mahasiswa menempuh pendidikan dari saat mulai kuliah di Universitas Sebelas Maret hingga mereka lulus. Model ini digunakan untuk membantu mengidentifikasi faktor-faktor yang memiliki pengaruh signifikan dalam menentukan durasi studi mahasiswa. Kasus ini menggunakan data lama studi mahasiswa S1 FMIPA UNS angkatan 2017 dan 2018. Data yang tertera mencakup mahasiswa yang lulus tepat waktu dalam waktu kurang dari sama dengan empat puluh delapan bulan, serta mahasiswa yang lulus dengan keterlambatan lebih dari empat puluh delapan bulan untuk menyelesaikan studi mereka.

METODE PENELITIAN

Penelitian ini menggunakan data primer yang diperoleh dari kuesioner yang didistribusikan ke mahasiswa FMIPA UNS angkatan 2017 dan 2018. Metode yang digunakan dalam penelitian ini adalah regresi *Cox Proportional Hazard*. Data *event* dalam penelitian ini yaitu mahasiswa lulus yang melakukan ujian skripsi kurang dari sama dengan 48 bulan. Data tersensor didapatkan jika mahasiswa yang lulus ujian skripsinya lebih dari 48 bulan, terhitung dari individu menjadi mahasiswa. Dalam penelitian ini, variabel respon yang digunakan terdiri dari waktu survival (waktu lama studi mahasiswa) dan variabel prediktornya adalah jenis kelamin, jurusan, jalur masuk universitas, IPK, asal daerah, beasiswa, kerja paruh waktu, organisasi, serta kepanitiaan.

Proses penelitian ini sebagai berikut:

- a. Analisis deskriptif pada data.
- b. Melakukan uji asumsi *proportional hazard*.
- c. Estimasi parameter dilakukan dengan metode (MLE) jika asumsi *proportional hazard* terpenuhi.
- d. Menentukan kombinasi variabel dalam model untuk menentukan model yang terbaik. Untuk memilih model terbaik, nilai AIC (*Akaike Information Criterion*) digunakan; nilai AIC terkecil adalah indikator model terbaik. (Harlyan et al., 2020).
- e. Pengujian signifikansi parameter dilakukan setelah memperoleh model regresi *Cox proportional hazard* terbaik, dengan *likelihood test* dan *wald test* serta menghitung nilai *hazard ratio*. Hazard rasio adalah perbandingan kecepatan

kejadian yang terjadi di satu kelompok dengan kecepatan yang terjadi di kelompok lain (Marisa et al., 2017).

HASIL DAN PEMBAHASAN

Sebelum melakukan analisis dengan lebih rinci, dilakukan statistika deskriptif untuk memberikan gambaran data yang digunakan. Tabel 1 tertampil statistika deskriptif:

Tabel 1. Statistika Deskriptif

		Frekuensi	Persentase
Informasi Kejadian	Terobservasi	124	84,9%
	Tersensor	22	15,1%
	Total	146	100%

Tabel 1 menunjukkan bahwa sekitar 124 mahasiswa (atau sekitar 84,9%) dari angkatan 2017 dan 2018 di FMIPA UNS dengan status terobservasi atau jumlah jumlah mahasiswa yang telah lulus kurang dari atau sama dengan 48 bulan. Sementara itu, terdapat 22 (atau sekitar 15,1%) mahasiswa FMIPA UNS angkatan 2017 dan 2018 yang membutuhkan waktu lebih dari 48 bulan (4 tahun) untuk menyelesaikan studinya, dan status mereka dikategorikan sebagai data tersensor.

Salah satu asumsi yang harus dipenuhi oleh model ini adalah asumsi *proportional hazard*. Asumsi ini dapat diuji melalui pendekatan *Goodness of Fit*, yang menggunakan *schoenfeld residual* (Sekar & Wuryandari, 2021). Metode *Goodness of Fit* digunakan untuk menguji asumsi *proportional hazard* ini. Metode ini memberikan pendekatan yang lebih objektif untuk mengevaluasi apakah variabel prediktor mengandung proporsional hazard atau tidak. Salah satu uji statistik yang digunakan dalam metode ini adalah nilai *residual Schoenfeld*. Untuk analisis ini, nilai ini dihitung dengan perangkat lunak Rstudio; hasilnya tercantum pada Tabel 2.

Tabel 2. Nilai Korelasi *Residual Schoenfeld* awal

Var	Korelasi	Pvalue	Putusan
Jk	0,431	0,511	Menerima H_0
Jrs	0,042	0,838	Menerima H_0
Jm	0,885	0,347	Menerima H_0
Ipk	2,941	0,086	Menerima H_0
Ad	0,436	0,509	Menerima H_0

Var	Korelasi	Pvalue	Putusan
Bea	0,815	0,367	Menerima H_0
Pt	1,305	0,253	Menerima H_0
Org	1,321	0,250	Menerima H_0
Kpn	0,967	0,326	Menerima H_0

Semua variabel prediktor yang memenuhi asumsi *proportional hazard* ditunjukkan pada Tabel 2. Dilihat dari nilai p-value untuk seluruh variabelnya lebih dari 0,05. Selanjutnya, menggunakan perangkat lunak Rstudio untuk melakukan estimasi parameter dengan metode regresi Cox *proportional hazard*. Hasilnya ditunjukkan pada Tabel 3:

Tabel 3. Hasil Estimasi Parameter Regresi Cox *Proportional Hazard*

Var	Coefisien	Exp (Coef)	SE (Coef)	z	p
jk	0.04405	1.04503	0.23477	0.188	0.851176
jrs	-0.70147	0.49586	0.27984	-2.508	0.012187
jm	0.00210	1.00210	0.12686	0.017	0.986793
ipk	2.11389	8.28040	0.57656	3.666	0.000246
ad	-0.15967	0.85242	0.19194	-0.833	0.404742
bea	-0.05095	0.95032	0.29270	-0.174	0.861800
pt	-0.19437	0.82335	0.29775	-0.653	0.513597
org	0.32555	1.38480	0.41234	0.790	0.429799
kpn	-0.27849	0.75692	0.39126	-0.712	0.476599

Model awal regresi Cox proportional hazard yang terbentuk berdasarkan hasil estimasi yang ditampilkan pada Tabel 3 di atas sebagai berikut:

$$\begin{aligned}
 h_t(t) = h_0(t) \exp(& 0,04405jk - 0,701478jrs + 0,00210jm \\
 & + 2,11389ipk - 0,15967ad - 0,05095bea \\
 & - 0,19437pt + 0,32555org - 0,27849kpn
 \end{aligned} \tag{1}$$

Selanjutnya, dilakukan pemilihan model terbaik menggunakan metode *backward elimination* dengan langkah-langkah manual menggunakan *software* Rstudio. Hasilnya tertampil di Tabel 4. Metode *backward elimination* mengeluarkan variabel yang tidak signifikan satu per satu dan melakukan prosedur ini secara terus menerus sampai tidak ada variabel prediktor yang tidak signifikan. (Dewi et al., 2018)

Tabel 4. Hasil Estimasi Parameter Terbaik pada Model Cox *Proportional Hazard*

Model	Uji Parsial	AIC
Model 1	ipk signifikan	1070.329
Model 2	ipk signifikan	1068.329
Model 3	ipk signifikan	1066.360
Model 4	ipk signifikan	1064.393
Model 5	jrs, ipk signifikan	1062.855
Model 6	jrs, ipk signifikan	1061.118
Model 7	jrs, ipk signifikan	1059.804
Model 8	jrs, ipk signifikan	1058.523

Berdasarkan tingkat signifikansi uji parsial yang tertampil pada Tabel 4, model ke-8 (model yang mencakup variabel jrs dan ipk) adalah model terbaik. Hal ini disebabkan seluruh variabel prediktor dalam model tersebut berdampak besar pada lamanya studi mahasiswa. Hal itu didukung dengan nilai AIC model ke-8 adalah nilai AIC terkecil dibandingkan dengan model lainnya. Oleh karena itu, diputuskan bahwa model terbaik yaitu model ke-7 atau model yang terdiri dari variabel jrs dan ipk karena memiliki tingkat kesalahan yang paling kecil. Model regresi Cox *proportional hazard* terbaik yang dikembangkan adalah sebagai berikut:

$$h_t(t) = h_0(t)\exp(-0,7414jrs + 1,8327ipk) \quad (2)$$

dengan $h_0(t)$ adalah fungsi *baseline hazard*, jrs adalah variabel jurusan dan ipk adalah variabel IPK.

Setelah itu, setiap parameter dalam model regresi Cox *proportional hazard* yang dibentuk diuji secara keseluruhan. *Likelihood ratio test* digunakan untuk menentukan variabel prediktor apa saja yang mempunyai pengaruh secara signifikan dilakukan secara serentak. Statistik ujinya mengikuti distribusi *chi-square*.

$$H_0: \beta_{jrs} = \beta_{ipk}$$

$$H_1: \text{minimal ada satu } \beta_k \neq 0 \text{ dengan } k = 1, 2, \dots, p$$

$$\alpha = 0,05$$

Tabel 5. Uji Serentak

Uji overall	<i>Likelihood Ratio Test</i>	P-value	Keputusan
	18,57	9,278e-05	Gagal Tolak H0

Berdasarkan Tabel 5, diperoleh statistik uji *likelihood ratio test* sebesar 14,21 dengan $p - value = 9,278e - 05$. Karena nilai $p - value < 0,05$ maka H_0 dapat ditolak. Dapat disimpulkan bahwa seluruh variabel prediktor berpengaruh secara serentak terhadap lama studi mahasiswa FMIPA UNS angkatan 2017 dan 2018.

Kemudian melakukan pengujian secara parsial dengan menggunakan uji *wald*. Pengujian ini dilakukan untuk menentukan apakah model dipengaruhi secara signifikan oleh variabel prediktor.

$$H_0: \beta_k = 0 \text{ dengan } k = jrs, ipk$$

$$H_1: \beta_k \neq 0 \text{ dengan } k = jrs, ipk$$

$$\alpha = 0,05$$

Berdasarkan hasil dari Rstudio, diperoleh Tabel 6 sebagai berikut. Keputusan diperoleh dengan membandingkan nilai $p - value$ dengan $\alpha = 0,05$.

Tabel 6. Uji Parsial

Variabel	P-value	Keputusan
Jrs	0,899	Tolak H0
Ipk	0,077	Tolak H0

Tabel 6 menunjukkan bahwa variabel yang paling signifikan terhadap lama studi mahasiswa FMIPA UNS angkatan 2017 dan 2018 adalah variabel jurusan dan IPK.

Hazard ratio digunakan untuk menggambarkan perbandingan hasil dari variabel-variabel ini terhadap lama studi mahasiswa FMIPA UNS angkatan 2017 dan 2018.

Tabel 7. Hazard Ratio

Hazard Ratio	
Variabel Jurusan	Variabel IPK
2,09887	13,0108

Tabel 7 menunjukkan bahwa lama studi mahasiswa diluar FMIPA mempunyai kecenderungan untuk menyelesaikan studi lebih cepat sebesar 2,09887 kali dibandingkan dengan mahasiswa yang berasal dari FMIPA. Selain itu, mahasiswa dengan nilai IPK > 3,50 memiliki potensi yang lebih tinggi untuk mengalami event (dinyatakan lulus kurang dari sama dengan 48 bulan) dibandingkan dengan mahasiswa dengan nilai IPK < 3,50. Hal ini dapat dilihat berdasarkan perbedaan

nilai *hazard ratio* antara dua mahasiswa yang memiliki nilai IPK yang berbeda, di mana nilai tersebut bernilai > 1 .

SIMPULAN

Hasil analisis lama studi mahasiswa FMIPA UNS angkatan 2017 dan 2018 yang telah dilakukan menunjukkan bahwa jurusan dan IPK merupakan variabel yang berpengaruh signifikan pada lama studi mahasiswa FMIPA UNS angkatan 2017 dan 2018. Mahasiswa yang berasal dari FMIPA cenderung membutuhkan waktu yang lebih panjang untuk lulus dibanding mahasiswa-mahasiswa yang di luar FMIPA UNS. Selain itu, mahasiswa dengan nilai IPK 3,91 berpotensi lebih tinggi untuk mengalami event (dinyatakan lulus kurang dari sama dengan 48 bulan) dibandingkan dengan mahasiswa dengan nilai IPK 2,51. Berikut adalah model regresi *Cox proportional hazard* yang terbentuk untuk variabel-variabel yang memengaruhi durasi studi mahasiswa FMIPA UNS angkatan 2017 dan 2018:

$$h_t(t) = h_0(t) \exp(0,04405jk - 0,701478jrs + 0,00210jm + 2,11389ipk - 0,15967ad - 0,05095bea - 0,19437pt + 0,32555org - 0,27849kpn)$$

Diharapkan penelitian ini bisa dikembangkan di penelitian selanjutnya dengan menggunakan metode regresi hazard aditif untuk membuat strategi yang dapat membantu mahasiswa menyelesaikan studi dengan tepat waktu.

DAFTAR PUSTAKA

- Dewi, A. Y., Dwidayati, N. K., & Agoestanto, Y. (2019). *Analisis Survival Model Regresi Cox Dengan Metode MLE Untuk Penderita Diabetes Mellitus*. *Unnes Journal of Mathematics*, 9(1), 31–40. <http://journal.unnes.ac.id/sju/index.php/ujm>
- Dewi, I. A. P. R., Suciptawati, N. L. P., & Tastrawati, N. K. T. (2018). *Aplikasi Regresi Cox Proportional Hazard pada Sintasan Pasien Diabetes Melitus*. *E-Jurnal Matematika*, 7(3), 278. <https://doi.org/10.24843/mtk.2018.v07.i03.p215>
- Harlyan, L. I., Yulianto, E. S., Fitriani, Y., & Sunardi. (2020). *Aplikasi Akaike Information Criterion (AIC) pada Perhitungan Efisiensi Teknis Perikanan Pukat Cincin di Tuban, Jawa Timur*. *Marine Fisheries*, Vol. 11, No.2, 181–188.
- I. D. Fitriani. (2018). *Analisis Regresi Cox Proportional Hazard Pada Identifikasi Faktor-Faktor yang Mempengaruhi Lama Studi Mahasiswa S1 FMIPA Universitas Islam Indonesia*.
- Iskandar, B. M. (2015). *Model Cox Proportional Hazard Pada Kejadian Bersama*.

-
- Klein, J. P., & Moeschberger, M. L. (2003). *Survival Analysis: Techniques For Censored And Truncated Data*. Springer.
- Kleinbaum, D. G., & Klein, M. (2012). *Statistics For Biology And Health Survival Analysis A Self-Learning Text Third Edition*.
<http://www.springer.com/series/2848>
- Landong, Hutahaean, P., Mukid, M. A., & Wuryandari, T. (2014). *Model Regresi Cox Proportional Hazards pada Data Lama Studi Mahasiswa (Studi Kasus di Fakultas Sains Dan Matematika*. 3(2), 173–181. <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- M. Z. Warli. (2017). *Pendekatan Regresi Cox Proporsional Hazard Dalam Penentuan Faktor-Faktor Yang Berpengaruh Terhadap Lama Studi Mahasiswa S-1 Matematika di Universitas Islam Negeri Alauddin Makassar*.
- Marisa, Yozza H, & Maiyastri. (2017). *Model Regresi Cox Proportional Hazard Pada Laju Tamat Mahasiswa Jurusan Matematika Universitas Andalas*.
- Sekar, K. A., & Wuryandari, T. (2021). *Model Regresi Cox Proportional Hazard pada Data Ketahanan Hidup Pasien Hemodialisa*. *Jurnal Gaussian*, 10(2), 303–314.
<https://ejournal3.undip.ac.id/index.php/gaussian/>
- Setiani, E., & Santoso, R. (2019). *Perbandingan Model Regresi Cox Proportional Hazard Menggunakan Metode Breslow dan Efron (Studi Kasus: Penderita Stroke di RSUD Tugurejo Kota Semarang)*. 8(1), 93–105.
<http://ejournal3.undip.ac.id/index.php/gaussian>