

PENERAPAN REGRESI COX UNTUK MENGANALISIS VARIABEL YANG BERPENGARUH TERHADAP DURASI STUDI MAHASISWA

Rhoudhotul Widyastuti¹, Dewi Wulandari^{2*}, Dina Prasetyowati³

^{1,2,3}Universitas PGRI Semarang

Email Korespondensi: dewiwulandari@upgris.ac.id

Email: rhou@gmail.com

Email: dinaprasetyowati@upgris.ac.id

ABSTRACT

This research aims to obtain the Cox regression equation in managing student study duration data and to determine variables or factors that significantly influence student study duration. We used data from 2017 FPMIPATI (Fakultas Pendidikan Matematika, Ilmu Pengetahuan Alam dan Teknologi Informasi) UPGRIS (Universitas PGI Semarang) Mathematics Education study program students which consisted of four predictor variables including gender, cumulative achievement index, parents' educational background, and organizational participation. From the results of the survival analysis, it was found that the factor that had a significant influence on the duration of student studies was the cumulative achievement index.

Keywords: *Survival Analysis, Study, Duration, Cox Regression*

ABSTRAK

Tujuan dari penelitian ini adalah mendapatkan persamaan regresi Cox dalam mengelolah data durasi studi mahasiswa dan dapat menentukan variabel atau faktor yang secara signifikan mempengaruhi durasi studi mahasiswa. Di dalam penelitian menggunakan data mahasiswa prodi Pendidikan Matematika FPMIPATI UPGRIS 2017 yang terdiri dari empat variabel prediktor antara lain jenis kelamin, indeks prestasi kumulatif, latar belakang pendidikan orang tua, dan keikutsertaan organisasi. Dari hasil analisis survival, diperoleh faktor yang memiliki pengaruh secara signifikan terhadap durasi studi mahasiswa adalah indeks prestasi kumulatif.

Kata kunci: *Analisis Survival, Pembelajaran, Durasi, Regresi Cox*

1. PENDAHULUAN

Lama (durasi) masa studi merupakan waktu yang dibutuhkan mahasiswa untuk menyelesaikan studinya, di mana durasi itu ditentukan sejak hari pertama aktif kuliah sampai mahasiswa mendapatkan surat keterangan lulus. Secara kualitas seluruh mahasiswanya yang lulus diharapkan untuk mempunyai nilai IPK yang bagus dan mampu lulus tepat waktu sesuai dengan standar kelulusan program sarjana yaitu empat tahun atau 48 bulan. Sedangkan secara kualitas, mahasiswa yang lulus sarjana diharapkan setara dengan mahasiswa yang masuk (Fitriana, 2016). Selama proses perkuliahan ternyata terdapat beberapa variabel yang dapat mempengaruhi lama studi mahasiswa.

Dalam analisis survival atau bisa disebut dengan waktu ketahanan hidup adalah suatu analisis untuk menghasilkan suatu data yang berupa waktu, di mana waktu tersebut dapat

diinterpretasikan dari awal kejadian hingga berakhirnya suatu event atau kejadian (Latan, 2014). Regresi Cox adalah suatu model bagian dari metode statistika yang dipergunakan untuk mendeskripsikan suatu keterkaitan antara variabel independen atau variabel bebas dan variabel dependen atau variabel terikat (Maiyanti, 2016).

Peneliti menggunakan regresi Cox proportional hazard untuk menentukan model regresi Cox dari data lama mahasiswa Pendidikan Matematika dan dapat menentukan variabel yang secara signifikan mempunyai pengaruh terhadap durasi masa studi S1 Pendidikan Matematika FPMIPATI UPGRIS angkatan tahun 2017. Lama studi mahasiswa dapat dihitung dari mahasiswa pertama kali masuk kuliah pada awal bulan September 2017 sampai dinyatakan lulus dengan gelar sarjana. Mahasiswa yang telah dinyatakan lulus sebelum waktu penelitian pada bulan Februari 2023 disebut terobservasi, sedangkan untuk mahasiswa yang belum dinyatakan lulus disebut tersensor.

Tahapan untuk menentukan model regresi Cox dimulai dengan pengecekan asumsi proportional hazard dengan menerapkan grafik plot log minus log survival dan pendekatan goodness of fit dilanjutkan analisis regresi Cox proportional hazard (Deo, dkk. 2021)

2. METODOLOGI

Analisis Survival adalah langkah-langkah yang dapat dipergunakan untuk mengidentifikasi data, di mana variabel-variabel respon berupa waktu awal terjadinya peristiwa atau *event* sampai berakhirnya peristiwa. Waktu *survival* bisa dihitung dalam bentuk hari, bulan, tahun dari awal terjadinya suatu peristiwa sampai berakhirnya peristiwa (Kristanti, 2019:53).

Analisis survival juga memiliki tujuan yaitu untuk mengidentifikasi suatu keterkaitan antara waktu awal terjadinya peristiwa dengan variabel yang diprediksi bisa memberikan pengaruh terhadap waktu ketahanan hidup (Prabawati, dkk, 2018).

Persamaan (1) berikut adalah persamaan fungsi distribusi kumulatif menurut Bain (1992), yang menggunakan variabel x dengan fkp $f(x)$ dan fungsi kepadatan peluang.

$$(x) = P(T \leq t) \int_{-\infty}^x f(x) dx \quad (1)$$

Variabel x dapat digantikan oleh variabel T di mana variabel T menjelaskan tentang waktu, maka fungsi distribusi kumulatifnya seperti persamaan (2).

$$F(t) = P(T \leq t) \int_0^t f(u) du \quad (2)$$

Menurut Lawless (2007), fungsi *survival* atau bisa dilambangkan dengan $S(t)$ adalah peluang dari suatu individu yang dapat *survive* terhadap waktu *survival* sampai berakhirnya waktu kejadian pada saat t ($t, 0$). Persamaan seperti persamaan (3).

$$S(t) = P(T \geq t) = \int_t^{\infty} f(x) dx \quad (3)$$

Fungsi *hazard* adalah ketidaklanjutan dari setiap individu yang mengalami kejadian dalam interval t hingga $t + \Delta t$. Menurut Lawless (1982:8), fungsi *hazard* biasanya dilambangkan dengan $h(t)$. Di mana setiap individu bisa mempertahankan hidup sampai dengan waktu t dapat dinyakatan dalam bentuk matematis seperti persamaan (4).

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \quad (4)$$

Menurut Collect yang dikutip Iskandar (2014:49), untuk menguji asumsi *proportional hazard* dapat digunakan tiga cara antara lain: pendekatan grafik *plot log minus log survival*, *goodness of fit* dan menambahkan variabel respon atau variabel dependen waktu. Untuk menguji asumsi *proportional hazard* dengan menggunakan grafik *plot log minus log survival*, apabila garis terlihat sejajar dan tidak saling memotong maka dapat dikatakan asumsi terpenuhi. Untuk mendukung hasil tersebut dapat melanjutkan dengan menguji asumsi menggunakan pendekatan *goodness of fit* di mana statistik uji nya menggunakan uji *residual schoenfeld* untuk mendapatkan keputusan yang lebih seimbang terhadap asumsi dapat terpenuhi atau tidak.

Menurut Rahayu (2015), estimasi parameter $\beta_1, \beta_2, \beta_3, \dots, \beta_p$ dapat menggunakan metode MPLE atau sering disebut dengan *maximum partial likelihood estimation*. Fungsi dari partial *likelihood* dapat dinyatakan pada persamaan (5).

$$L(\beta) = \prod_{i=1}^r \frac{\exp(\sum_{j=1}^p \beta_j x_{ij})}{\sum_{i \in R(t_j)} \exp(\sum_{j=1}^p \beta_j x_{ij})} \quad (5)$$

Berikut adalah fungsi *log partial likelihood* yang diperoleh dari persamaan (5) di atas.

$$\ln L(\beta) = \left[\left(\sum_{j=1}^p \beta_j x_{ij} \right) - \ln \left(\sum_{i \in R(t_j)} \exp \left(\sum_{j=1}^p \beta_j x_{ij} \right) \right) \right] \quad (6)$$

Di mana $R(t_j)$ adalah suatu himpunan dari semua individu yang mengalami kegagalan pada waktu t_j yang masih bertahan sampai waktu t_j sedangkan x_{ij} adalah variabel prediktor yang mengalami kegagalan pada waktu tertentu.

Untuk mendapatkan parameter β_j kita dapat memaksimumkan derivatif pertama dari fungsi *log partial likelihood* yang ditampilkan pada persamaan (6).

$$\frac{\partial \ln L(\beta)}{\partial (\beta_j)} = 0 \quad (7)$$

Dari persamaan (7), dengan menggunakan metode *Newton-Raphson*, dapat ditentukan solusinya.

Menurut Hosmer, dkk (2008), ada 2 uji parameter, yakni: uji simultan parameter dan uji parsial. Untuk Uji simultan parameter atau disebut dengan uji serentak digunakan statistik uji G di mana statistik uji tersebut berdistribusi chi-square.

$$G = -2 [\ln L(0) - \ln L(\hat{\beta}_j)] \quad (8)$$

Sedangkan uji parsial biasanya menggunakan statistik uji *wald*.

$$z^2 = \left(\frac{\hat{\beta}_j}{SE(\beta_j)} \right)^2 \quad (9)$$

Menurut Collect (2003), regresi *Cox* dapat dipergunakan untuk mengidentifikasi keterkaitan variabel prediktor dengan variabel respon di mana data yang digunakan adalah

data berupa waktu ketahanan hidup dari suatu individu tertentu. Persamaan (10) merupakan persamaan regresi *Cox*.

$$h(t, X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_p X_p) \quad (10)$$

Di mana X_i adalah sebuah nilai dari variabel prediktor dengan i bernilai 1, 2, 3, ..., p . β adalah sebuah koefisien dari model regresi. $h_0(t)$ adalah suatu fungsi *hazard* atau fungsi kegagalan suatu individu. Sedangkan $h(t, X)$ adalah suatu kelanjutan dari individu untuk mendapatkan suatu kejadian pada waktu t dengan karakteristik x .

Ragam penelitian yang digunakan dalam studi ini adalah penelitian kuantitatif. Data yang dihasilkan merupakan data sekunder, yaitu durasi masa studi mahasiswa S1 Pendidikan Matematika FPMIPATI UPGRIS 2017 yang diperoleh dari Tata Usaha (TU) FPMIPATI UPGRIS. Penelitian ini menggunakan *simple random sampling* untuk mengambil sampel penelitian. Terdapat beberapa variabel prediktor antara lain jenis kelamin, indeks prestasi kumulatif, latar belakang pendidikan orang tua, dan keikutsertaan organisasi. Untuk menganalisis data peneliti menggunakan metode regresi *Cox* dengan *software* bantu statistik (Dessai, 2018). Berikut adalah tahapan analisis data untuk mencapai tujuan penelitian: a) Analisis deskriptif; b) Pengecekan asumsi dengan menerapkan grafik *plot log minus log survival* dan pendekatan *goodness of fit*; c) Menentukan estimasi parameter; d) Pengujian parameter yang menggunakan uji simultan parameter dan uji parsial; e) Pemilihan model terbaik; dan f) Menarik kesimpulan.

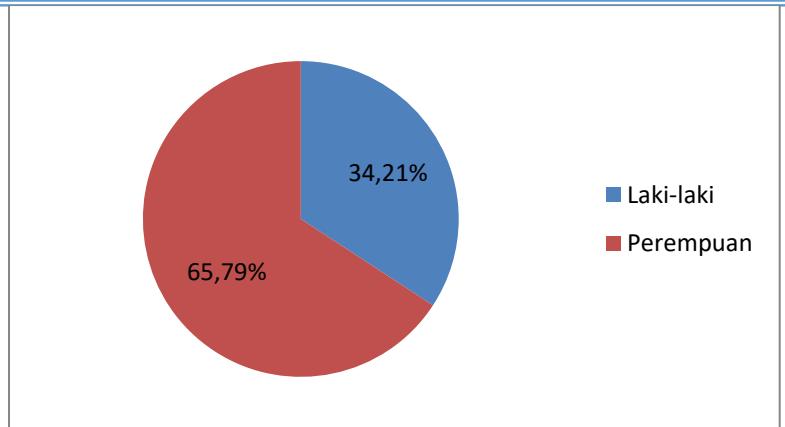
3. HASIL DAN PEMBAHASAN

Populasi yang diambil adalah seluruh mahasiswa Pendidikan Matematika FPMIPATI UPGRIS angkatan 2017 dengan jumlah 175 mahasiswa antara lain 24 mahasiswa laki-laki dan 151 mahasiswa perempuan. Sampel yang diambil sebanyak 38 mahasiswa.

Tabel 1. Frekuensi Data Mahasiswa

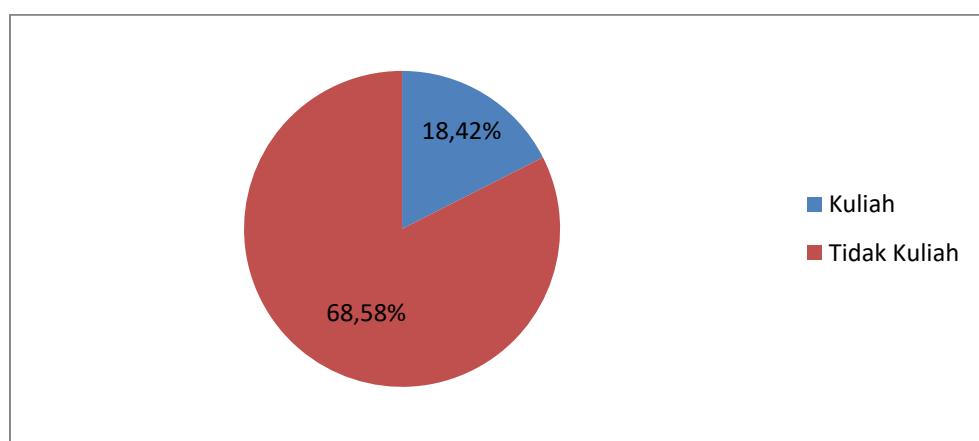
Informasi Kejadian	Frekuensi	Persentase
Terobservasi	28	73.68%
Tersensor	10	26.32%
Total	38	100%

Berdasarkan Tabel 1, terlihat bahwa mahasiswa yang terobservasi sebanyak 73.68% atau sebanyak 28 mahasiswa yang telah dinyatakan lulus sebelum bulan Februari 2023. Sedangkan untuk mahasiswa yang belum dinyatakan lulus atau tersensor sebanyak 10 mahasiswa atau 26.32%.



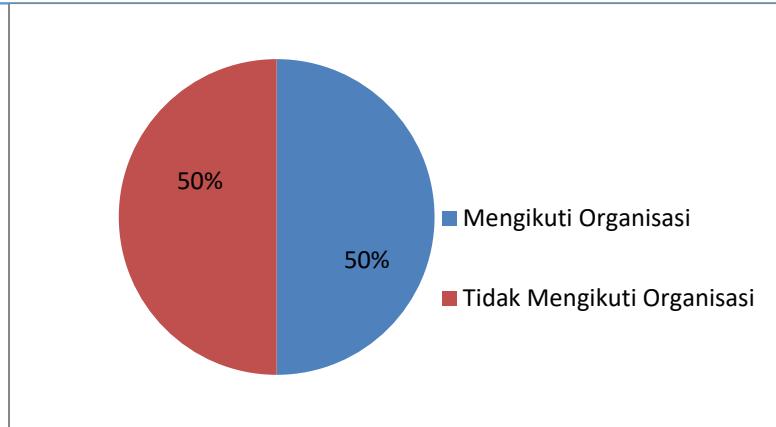
Gambar 1. Distribusi Frekuensi Variabel Jenis Kelamin

Variabel jenis kelamin dibagi menjadi dua kategori yaitu laki-laki dan perempuan (Gambar 1). Dari hasil pengambilan data mahasiswa Pendidikan Matematika FPMIPATI UPGRIS 2017, terdapat 13 mahasiswa laki-laki (34.21%) dan 25 mahasiswa perempuan (65.79%)



Gambar 2. Distribusi Frekuensi Variabel Indeks Prestasi Kumulatif

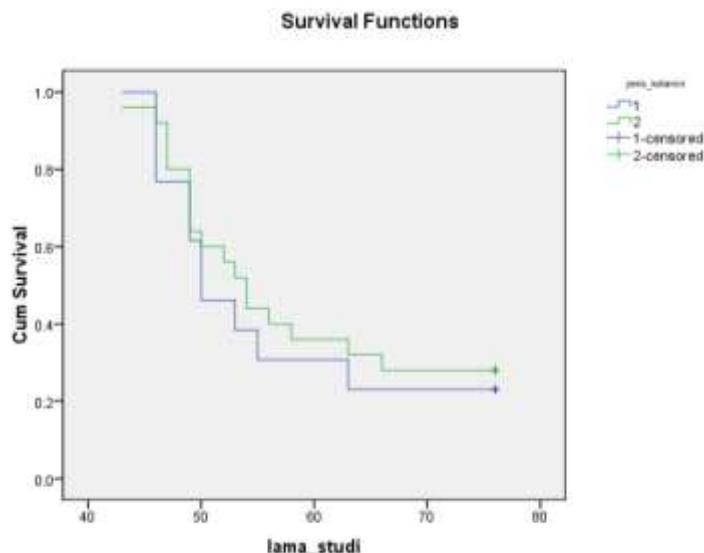
Dari gambar 2. Variabel latar belakang pendidikan orang tua dikategorikan menjadi dua yaitu kategori orang tua yang lulusan kuliah dan orang tua yang tidak lulusan kuliah. Dari hasil penelitian diperoleh sebanyak 7 mahasiswa (18.42%) yang latar belakang pendidikan orang tua nya lulusan kuliah dan sebanyak 31 mahasiswa (68.58%) yang latar belakang pendidikan orang tua nya tidak lulusan kuliah.



Gambar 3. Distribusi Frekuensi Variabel Keikutsertaan Organisasi

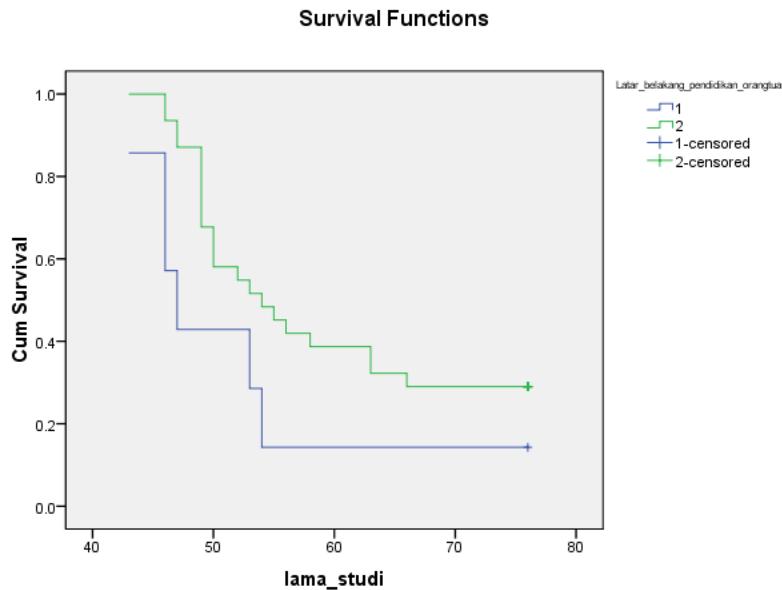
Pada Gambar 3. Variabel keikutsertaan organisasi dibagi menjadi dua kategori yaitu mahasiswa yang ikut serta dan tidak ikut serta dalam keorganisasian. Dari hasil penelitian di peroleh mahasiswa yang mengikuti dan tidak mengikuti organisasi berjumlah sama yaitu 19 mahasiswa.

Selanjutnya yaitu pengujian asumsi *proportional hazard* yang menggunakan dua pendekatan antara lain pendekatan grafik *plot log minus log survival* dan pendekatan *goodness of fit* yang menggunakan statistik uji *p-value*.



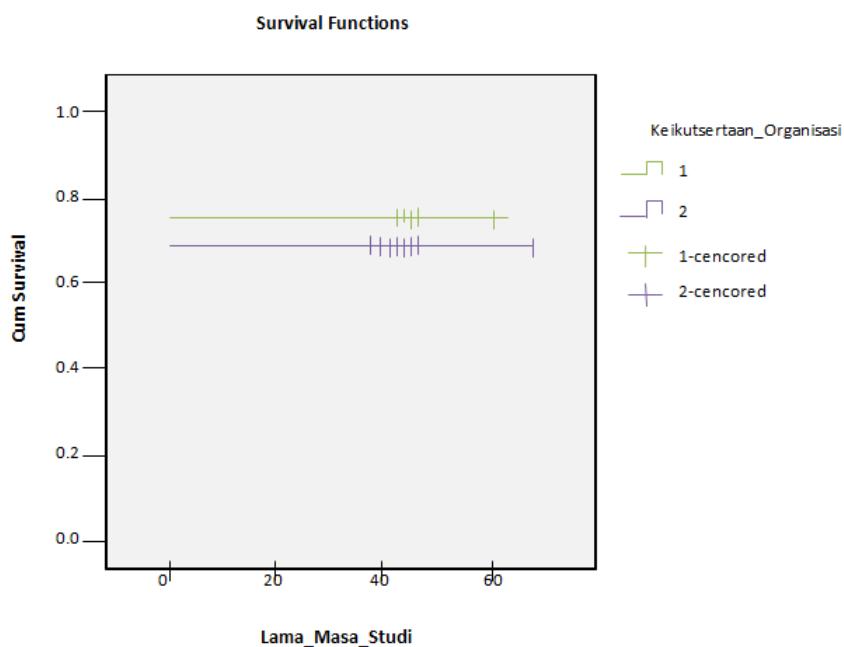
Gambar 4. Plot Log Minus Log Survival Variabel Jenis Kelamin (X1)

Dapat dilihat pada Gambar 4 bahwa grafik *survival* dari variabel jenis kelamin sejajar dan tidak saling memotong, maka dapat disimpulkan variabel jenis kelamin (X1) memenuhi asumsi *proportional hazard*.



Gambar 5. *Plot Log Minus Log Survival* Variabel Latar Belakang Pendidikan Orang Tua (X3)

Dapat dilihat dari Gambar 5, grafik *survival* untuk variabel latar belakang pendidikan orang tua sejajar dan tidak saling memotong, maka dapat disimpulkan variabel latar belakang pendidikan orang tua (X3) memenuhi asumsi.



Gambar 6. *Plot Log Minus Log Survival* Variabel Keikutsertaan Organisasi

Gambar 6 menunjukkan grafik *survival* untuk variabel keikutsertaan organisasi terlihat sejajar dan tidak berpotongan, maka dapat disimpulkan bahwa variabel keikutsertaan organisasi (X4) memenuhi asumsi *proportional hazard*.

Untuk mendukung jawaban dari pengujian asumsi *proportional hazard* yang menggunakan pendekatan grafik *plot log minus log survival* lebih baik dilanjutkan dengan pendekatan *goodness of fit* yang menggunakan statistik uji *residual schoenfeld* yang dilihat dari nilai *p-value*.

Tabel 2. Nilai *P-Value* dari Uji Asumsi *Proportional Hazard* menggunakan pendekatan *Goodness of Fit*

Variabel	Chi-Square	P-Value	Keputusan
Jenis kelamin	.238	.626	Gagal tolak H_0
Latar belakang pendidikan orang tua	2.794	.095	Gagal tolak H_0
Keikutsertaan Organisasi	.115	.735	Gagal tolak H_0

Tabel 2 memperlihatkan semua nilai *p-value* dari masing-masing variabel bernilai lebih besar dari taraf signifikan atau lebih dari 0,05. Jadi dapat diputuskan bahwa gagal tolak H_0 maka dapat diartikan semua variabel prediktor memenuhi asumsi *proportional hazard* dan dapat dilanjutkan ke dalam uji selanjutnya.

Selanjutnya adalah melakukan estimasi parameter. Tabel 3 berikut ini menunjukkan hasil estimasi parameter.

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

Variabel	β_j	Se (β_1)	$\chi^2 w$	P-Value	Keputusan
X1	.524	.418	1.575	.210	Terima H_0
X2	2.028	.772	6.896	.009	Tolak H_0
X3	-.014	.524	.001	.978	Terima H_0
X4	.725	.422	2.951	.086	Terima H_0

Selanjutnya, dapat dibentuk persamaan regresi Cox berdasarkan Tabel 3, yaitu sebagai berikut:

$$h(t, X) = \exp(0,524X1 + 2,028X2 - 0,014X3 + 0,725X4)h_0(t)$$

Dari model tersebut dilanjutkan dengan uji simultan parameter biasa disebut dengan uji sertentak yang digunakan untuk menguji apakah model fit. Di mana digunakan statistik uji *G*, dengan hipotesis seperti berikut:

H_0 : Model regresi *Cox proportional hazard* tidak layak digunakan

H_1 : Model regresi *Cox proportional hazard* layak digunakan

Diperoleh hasil perhitungan dari statistik uji *likelihood ratio test* atau uji *G* seperti berikut:

$$\chi^2 = l_0 - l_p$$

$$\chi^2 = 177,953 - 158,975$$

$$\chi^2 = 18,978$$

Tabel 4. Uji Simultan Model Awal Regresi *Cox Proportional Hazard*

Uji Overall (Score)	$\chi^2_{LR} \geq \chi^2_{4;0.05}$	<i>p</i> – value < α	Keputusan
	$18,978 \geq 9,488$	$.005 < .05$	Tolak H_0

Tabel 4 menunjukkan bahwa nilai $\chi^2_{LR} \geq \chi^2_{4;0.05}$ atau $p-value < \alpha$ maka diperoleh keputusan H_0 ditolak yang memiliki arti bahwa model layak digunakan. Selanjutnya dilakukan uji parsial yang digunakan untuk mengetahui apakah setiap variabel independen berpengaruh secara signifikan terhadap variabel dependen, di mana uji parsial menggunakan statistik uji *Wald*.

H_0 : Variabel independen tidak berpengaruh signifikan terhadap variabel dependen

H_1 : Variabel independen berpengaruh signifikan terhadap variabel dependen

Berdasarkan Tabel 3. Diperoleh bahwa keputusan H_0 diterima pada variabel jenis kelamin (X1), variabel latar belakang pendidikan orang tua (X2), dan variabel ikutsertaan organisasi (X4). Sedangkan untuk keputusan H_0 ditolak pada variabel indeks prestasi kumulatif (X2). Dari keputusan uji parsial diperoleh kesimpulan bahwa tidak semua variabel independen berpengaruh terhadap variabel dependen. Oleh karena itu, akan dilanjutkan dengan metode eliminasi *backward* yang digunakan untuk menghasilkan model terbaik dari regresi *Cox proportional hazard* dan mengetahui faktor yang berpengaruh secara signifikan terhadap model.

Tabel 5. Estimasi Terbaik

Variabel	β_j	Se (β_1)	$\chi^2 w$	P-Value	Keputusan
X2	1.629	.623	6.832	.009	Tolak H_0

Bersarkan Tabel 5. Diperoleh bahwa hasil estimasi parameter model terbaik dari regresi *Cox proportional hazard* sehingga menghasilkan model terbaik dari regresi *Cox proportional hazard* untuk lama studi mahasiswa Pendidikan Matematika FPMIPATI UPGRIS 2017 sebagai berikut:

$$h(t, X) = \exp(1,629X2)h_0(t)$$

Di mana X2 adalah variabel indeks prestasi kumulatif, jadi dari keempat faktor tersebut yang berpengaruh secara signifikan terhadap lama studi mahasiswa Pendidikan Matematika FPMIPATI adalah indeks prestasi kumulatif.

4. SIMPULAN

Kesimpulan yang dapat ditarik dari hasil pembahasan pada bagian sebelumnya adalah faktor yang dapat berpengaruh secara signifikan terhadap lama studi mahasiswa Pendidikan Matematika FPMIPATI Universitas PGRI Semarang angkatan 2017 adalah indeks prestasi kumulatif, dengan model regresi $h(t, X) = \exp(1,629X2)h_0(t)$.

5. DAFTAR PUSTAKA

- Bain, L. J., dan Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics* 2nd Edition. Boston. Duxburry Press.
- Collet, D. 2003. *Text in Statistical Science: Modelling Survival Data in Medical Research* Second Edition. USA. Chapman & Hall.
- Deo, S. V., Deo, V., dan Sundaram, V. 2021. Survival Analysis—Part 2: Cox Proportional Hazards Model. [Indian J Thorac Cardiovasc Surg.](https://doi.org/10.32493/sm.v7i1.41948) Vol. 37(2) Pp. 229–233.
- Dessai, S., Simha, V., dan Patil, V. 2018. Stepwise Cox regression analysis in SPSS. *Cancer*

Research, Statistics, and Treatment Vol 1(2) Pp. 167-170

- Fitriana, R. 2016. Analisis Survival Faktor-Faktor yang Mempengaruhi Lama Studi Mahasiswa Pendidikan Matematika Angkatan 2010 dengan Metode Regresi Cox Proportional Hazard. *Tugas Akhir Program Studi Statistika Terapan Dan Komputasi, Jurusan Matematika FMIPA UNNES.* <https://lib.unnes.ac.id/25050/>
- Hosmer, D. W., Lemeshow, S., dan May, S. 2008. *Applied Survival Analysis: Regression Modelling of Time to Event Data*. New Jersey. John Wiley.
- Kristanti, F. T. 2019. *Financial Distress 1st Edition*. Malang. Intelegensia Media.
- Latan, H. 2014. *Aplikasi Analisis Data Statistik Untuk Ilmu Sosial dan Sains dengan IBM SPSS*. Bandung. Alfabeta
- Lawless, J. F. 1982. *Statistical Models and Methods for Lifetime Data*. USA. John Wiley & Sons, Inc.
- Lawless, J. F. 2007. *The Statistical Analysis of Recurrent Event*. USA. Springer Science+Business Media Inc.
- Maiyanti, S. I., Cahyono, E. S., dan Eliyati, N. 2016. Aplikasi Regresi Cox Metode Backward untuk Menduga Faktor-faktor yang Berpengaruh terhadap Waktu Kelulusan Mahasiswa Bidik Misi Unsri. In *PRISMA, Prosiding Seminar Nasional Matematika* (pp.418-426).
- Prabawati, S., Nasution, Y. N., dan Wahyuningsih, S. 2018. Analisis Survival Data Kejadian Bersama dengan Pendekatan Efron Partial Likelihood (Studi Kasus: Lama Masa Studi Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman Angkatan 2011). *Jurnal Eksponensial* Vol. 09 Pp. 75–84.