JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI e-ISSN: 2614-8811
Published by Departement of Mathematics, Hasanuddin University, Indonesia
https://journal.unhas.ac.id/index.php/jmsk/index
Vol. 22, No. 3, May 2026, pp. 700-716
DOI: 10.20956/j.v22i3.50137

Cox Weibull Regression Modeling for Graduation Rate Analysis of Statistics Students at Mulawarman University

Pemodelan Regresi Cox Weibull untuk Analisis Laju Kelulusan Mahasiswa Program Studi Statistika Universitas Mulawarman

Azkadiaannzumi[1], Suyitno Suyitno[2*], Darnah[3]

[123]*Departemen of Statistics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Indonesia*
Email: [1]*azkadia07@gmail.com*, [2*]*suyitno.stat.unmul@gmail.com*, [3]*darnah.98@gmail.com*
**Corresponding Author*
Received: 3 March 2026, revised: 19 April 2026, accepted: 20 April 2026

Abstract

This study proposes a Cox regression model with a Weibull baseline hazard, representing a key statistical innovation, hereafter referred to as the Cox Weibull regression model. Parameter estimation was conducted separately, the regression parameters were estimated using the partial likelihood method, while the baseline hazard parameters were estimated using Maximum Likelihood Estimation (MLE). This study aims to identify factors influencing graduation rates and to examine variations in these rates among students in the Statistics Study Program at Universitas Mulawarman based on significant factors. The study used academic records of 77 students from the Statistics Study Program (cohorts 2016–2019), obtained from the Academic and Student Affairs Bureau, Faculty of Mathematics and Natural Sciences, Universitas Mulawarman. The best model selected using the backward elimination method, included three covariates and was considered adequate, as all covariates were statistically significant and the model yielded the lowest AIC value (462.34). The results indicate that cumulative grade point average (GPA), scholarship status, and university admission pathway significantly influence student's graduation rates. Higher GPA was associated with a shorter time to graduation, scholarship recipients graduated faster than non-recipients, and students admitted through the SNMPTN pathway graduated faster than those admitted through other pathways.

Keywords: Cox Weibull regression model, graduation rate, partial likelihood method

Abstrak

Penelitian ini mengusulkan model regresi Cox dengan hazard dasar Weibull, yang merupakan inovasi statistik utama, yang selanjutnya disebut sebagai model regresi Cox Weibull. Penaksiran parameter dilakukan secara terpisah, penaksiran parameter regresi menggunakan metode *partial likelihood*, sedangkan penaksiran parameter *baseline hazard* menggunakan metode *Maximum Likelihood Estimation* (MLE). Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap laju kelulusan dan mengkaji variasi laju kelulusan mahasiswa Program Studi

Statistika Universitas Mulawarman berdasarkan faktor-faktor yang signifikan. Penelitian ini menggunakan data rekam akademik dari 77 mahasiswa Program Studi Statistika (angkatan 2016–2019) yang diperoleh dari Biro Akademik dan Kemahasiswaan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman. Model terbaik dipilih menggunakan metode *backward elimination*, yaitu model yang melibatkan tiga kovariat yang menghasilkan model layak dengan seluruh kovariat signifikan secara statistik dan menghasilkan nilai AIC terendah (462,34). Hasil penelitian menunjukkan bahwa indeks prestasi kumulatif (IPK), status penerima beasiswa, dan jalur masuk perguruan tinggi berpengaruh signifikan terhadap laju kelulusan. IPK yang lebih tinggi berasosiasi dengan waktu kelulusan yang lebih singkat, mahasiswa penerima beasiswa lulus lebih cepat dibandingkan non-penerima, dan mahasiswa yang diterima melalui jalur SNMPTN lulus lebih cepat dibandingkan dengan jalur lainnya.

Kata kunci: *Cox Weibull regression model, graduation rate, partial likelihood method*

1. PENDAHULUAN

Penelitian ini mengkaji model regresi Cox, yaitu salah satu pemodelan dalam analisis *survival* yang berfokus memodelkan data waktu kejadian (*time-to-event*). Model regresi Cox adalah salah satu model statistika untuk memodelkan laju (*hazard rate*) *event* untuk mengalisis dan menjelaskan perubahan risiko kejadian (*event*) seiring berjalannya waktu [21]. Model regresi Cox pada penelitian ini mengusulkan fungsi *hazard* Weibull sebagai *baseline*, yang selanjunya dinamakan model regresi Cox Weibull. Model regresi Cox Weibull merupakan salah satu model *proportional hazard* (PH), yakni laju *event* suatu individu merupakan hasil perkalian dengan rasio tetap dari laju *event* individu lainnya [2], [6], [7]. Salah satu kekhasan data waktu adalah adanya data tersensor. Data waktu terdiri dari data lengkap dan data tersensor. Data waktu lengkap adalah data hasil pengamatan dari individu yang mengalami *event*. Sedangkan data waktu tersensor adalah data dimana individu yang diamati tidak mengalami *event* atau tidak dapat diamati hingga akhir penelitian [16].

Kekhasan model regresi Cox Weibull adalah penaksiran parameter *baseline* dan parameter regresi dilakukan secara terpisah. [1], [8]**.** Pemodelan data waktu yang lain yang berbasis distribusi Weibull dalam analisis *survival* yaitu model regresi Weibull. Berbeda dengan model regresi Cox Weibull, model *hazard rate* pada model regresi Weibull dikonstruksi dari fungsi *hazard* Weibull dengan parameter skala (*scale*) dinyatakan dalam fungsi parameter regresi dan kovariat [15], [25], [26]. Model regresi Weibull tidak memuat *baseline hazard*, sehingga penaksiran parameter yang terdiri dari parameter bentuk (*shape*) dan parameter regresi dilakukan secara simultan menggunakan metode *Maximum Likelihood Estimation* (MLE).

Penerapan model regresi Cox Weibull yang diusulkan pada penelitian ini adalah pada data waktu lama studi mahasiswa program studi Statistika Universitas Mulawarman. Laju kelulusan mahasiswa merupakan salah satu indikator yang mencerminkan kualitas penyelenggaraan pendidikan tinggi. Indikator ini dapat dilihat melalui lama studi mahasiswa, di mana lama studi yang lebih panjang mengindikasikan adanya penurunan laju kelulusan [20]. Prodi Statistika Universitas Mulawarman terus berupaya agar mahasiswanya dapat menyelesaikan studi dalam rentang 4 hingga 4,5 tahun. Namun, berdasarkan data yang diperoleh dari Biro Akademik dan Kemahasiswaan Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Mulawarman (Unmul), rata-rata lama studi mahasiswa Prodi Statistika angkatan 2016-2019 masih di atas 4,5 tahun yaitu 4,73 tahun. Hal ini menjadi tantangan bagi Prodi Statistika karena adanya indikasi laju kelulusan mahasiswa masih rendah yang akan berpengaruh terhadap akreditasi Prodi Statistika [17]. Oleh karena itu, penelitian ini mengusulkan untuk menganalisis faktor-faktor yang memengaruhi laju kelulusan dan mendeskripsikan variasi laju kelulusan mahasiswa melalui pemodelan regresi Cox Weibull. Hasil penelitian diharapkan dapat memberikan pertimbangan kepada Prodi Statistika FMIPA Unmul dalam menyusun program percepatan kelulusan mahasiswa.

Penelitian ini bertujuan mendapatkan model regresi Cox Weibull pada data waktu lama studi mahasiswa program studi Statistika Unmul. Melalui model regresi Cox Weibull (model *hazard rate*) dapat diidentifikasi faktor-faktor yang berpengaruh terhadap laju kelulusan dan dapat dikaji variasi laju kelulusan mahasiswa berdasarkan faktor-faktor yang signifikan. Melalui model regresi Cox Weibull laju kelulusan tidak hanya bergantung pada kovariat (faktor eksternal) tetapi juga bergantung pada waktu studi (masa studi) mahasiswa. Data waktu lama studi mahasiswa pada penelitian ini merupakan tipe data waktu tersensor kanan ke titik 5. Event pada penelitian ini adalah kelulusan mahasiswa dengan masa studi kurang atau tepat 5 tahun. Penaksiran parameter *baseline hazard* Weibull pada penelitian ini menggunakan metode MLE, sedangkan penaksiran parameter regresi atau koefisien kovariat (faktor eksternal) menggunakan metode *Maximum Partial Likelihood Estimation* (MPLE) dengan melibatkan status tersensor [1], [8]**.**

Kovariat pada penelitian ini terdiri dari kovariat numerik dan kovariat kategorik. Pemeriksaan multikolinieritas antar kovariat dilakukan secara terpisah dimana multikolinieritas antar kovariat numerik dideteksi berdasarkan nilai *Variance Inflation Factor* (VIF), sedangkan pendeteksian multikolinieritas antar kovariat kategorik yang diusulkan pada penelitian ini menggunakan uji independensi *chi square* [28]. Pemilihan model terbaik yang diusulkan dalam penelitian ini menggunakan metode *backward elimination*, yaitu proses seleksi model dengan menghapus kovariat satu per satu berdasarkan signifikansi pada pengujian parameter secara parsial [10], [14], [27], [29]**.** Model terbaik ditentukan berdasarkan kriteria *Akaike Information Criterion* (AIC) yang bergantung pada fungsi *log-likelihood* dari metode MPLE. Nilai AIC yang lebih kecil menunjukkan model yang lebih sesuai dengan pola data yang dianalisis [11], [12]. Kriteria pemilihan model terbaik adalah menghasilkan model yang layak, semua kovariat berpengaruh, model dengan kovariat siginifikan paling banyak, dan model mengasilkan AIC minimum. Pengujian signifikansi parameter regresi dilakukan secara simultan dan parsial. Uji simultan menggunakan *likelihood ratio test* dengan statistik uji mengikuti distribusi *chi square* sedangkan uji parsial menggunakan *Wald test* dengan statistik uji mengikuti distribusi normal baku [3].

Model regresi Cox Weibull umumnya digunakan untuk menganalisis data bidang medis dan kesehatan [1], [13], [18], [23], [27]. Penelitian terdahulu yang menerapkan model Cox Weibull pada data lama waktu rawat inap pasien yang diagnosis COVID-19 dilakukan oleh Azizy, Suyitno, dan Siringoringo [5]. Pada penelitian tersebut, pemilihan model terbaik dilakukan dengan mencoba untuk memodelkan seluruh kombinasi kovariat yang mungkin. Terdapat penelitian lain yang spesifik membahas laju kelulusan mahasiswa menggunakan model regresi Cox *proportional hazard*, tetapi *baseline hazard* dianggap konstan tidak sepesifik mengikuti pola distribusi tertentu, sehingga laju kelulusan hanya dipengaruhi oleh kovariat (faktor eksternal) tanpa mempertimbangkan adanya pengaruh waktu lama studi [4]. Pemodelan data waktu atau pemodelan data kontinu non-negatif dan berbasis distribusi Weibull pada kasus data yang berbeda dengan penelitian ini juga pernah dilakukan oleh peneliti terdahulu sebagai berikut: Liu dkk. [19] melalui Weibull *accelerated failure time (AFT) regression model*, Khairunnisa dkk. [15] melalui model regresi Weibull pada data waktu rawat inap pasien COVID-19, dan Suyitno dkk. [25], [26] melalui model regresi Weibull pada data waktu rawat inap pasien penderita *stroke*, dan model *geographically weighted Weibull regression* yang diterapkan pada pada spasial kontinu nonnegatif (bukan data waktu).

2. MATERIAL DAN METODE

2.1 Sumber dan Variabel Penelitian

Data penelitian ini merupakan data sekunder dari Biro Akademik dan Kemahasiswaan FMIPA Unmul. Ukuran sampel sebanyak 77 mahasiswa Prodi Statistika Angkatan 2016 – 2019. *Event* penelitian ini adalah kelulusan mahasiswa dengan lama studi kurang dari atau tepat 5 tahun.

Penelitian ini melibatkan variabel waktu (y), variabel status tersensor, dan kovariat yang disajikan pada Tabel 2.1. Berdasarkan data pada Tabel 2.1, bahwa nilai variabel status tersensor bukan merupakan hasil transformasi data waktu (data respon) menjadi data kategorik (nominal) biner. Kedua data waktu dan data status tersensor saling berhubungan dan dilibatkan pada penaksiran parameter model regresi Cox Weibull. Pemodelan regresi logistik pada data status tersensor sebagai data variabel respon kurang sesuai, karena status tersensor bisa berasal dari data tidak lengkap, ketika individu tidak dapat diamati atau hilang informasi hingga akhir penelitian [16].

Tabel 2.1 Variabel Penelitian

Variabel	Jenis Data	Satuan
Lama Studi (Y)	Numerik	Tahun
Status Tersensor (δ)	Kategorik	$\delta_i = \begin{cases} 1, y \leq 5 \\ 0, y > 5 \end{cases}$
Usia (X_1)	Numerik	Tahun
Jenis Kelamin (X_2)	Kategorik	$x_{i2} = \begin{cases} 1, \text{Laki-laki} \\ 0, \text{Perempuan} \end{cases}$
IPK (X_3)	Numerik	-
Beban Studi (X_4)	Numerik	SKS
Status Penerima Beasiswa (X_5)	Kategorik	$x_{i5} = \begin{cases} 1, \text{Penerima} \\ 0, \text{Bukan Penerima} \end{cases}$
Jalur masuk perguruan tinggi (PT) (X_6)	Kategorik	$x_{i6} = \begin{cases} 1, \text{SNMPTN} \\ 0, \text{non-SNMPTN} \end{cases}$

2.2 Penaksiran dan Pengujian Distribusi Weibull

Bentuk khusus distribusi Weibull adalah distribusi Weibull versi skala-bentuk dengan fungsi kepadatan peluang (FKP) diberikan oleh

$$f(y) = \frac{\gamma}{\lambda}\left(\frac{y}{\lambda}\right)^{\gamma-1} \exp\left[-\left(\frac{y}{\lambda}\right)^{\gamma}\right], \tag{2.1}$$

dengan $y \geq 0$ dan $0 < \gamma, \lambda < \infty$, dimana λ adalah parameter skala (*scale*) dan γ adalah parameter bentuk (*shape*) [3], [25]. Berdasarkan FKP pada persamaan (2.1), fungsi distribusi kumulatif diberikan oleh

$$F(y) = \int_0^y f(x)\, dx = 1 - \exp\left[-\left(\frac{y}{\lambda}\right)^{\gamma}\right]. \tag{2.2}$$

Fungsi *survival* dari distribusi Weibull versi skala-bentuk berdasarkan persamaan (2.2) adalah

$$S(y) = 1 - F(y) = \exp\left[-\left(\frac{y}{\lambda}\right)^{\gamma}\right], \tag{2.3}$$

dan fungsi *hazard* didefinisikan oleh

$$h(y) = \frac{f(y)}{S(y)} = \gamma\lambda^{-\gamma}y^{\gamma-1} \tag{2.4}$$

Parameter distribusi Weibull ditaksir menggunakan metode *Maximum Likelihood Estimation* (MLE). Misalkan, diberikan n data waktu yang saling bebas dan berdistribusi indentik, yaitu berdistribusi Weibull versi skala bentuk, dengan FKP diberikan oleh persamaan (2.1). Berdasarkan FKP (2.1), fungsi *likelihood* didefinisikan oleh

$$L(\boldsymbol{\theta},\mathbf{y}) = \prod_{i=1}^{n} f(\boldsymbol{\theta},\mathbf{y}) = \prod_{i=1}^{n} \left[\frac{\gamma}{\lambda}\left(\frac{y_i}{\lambda}\right)^{\gamma-1}\right] \exp\left[-\left(\frac{y_i}{\lambda}\right)^{\gamma}\right], \tag{2.5}$$

dengan $\boldsymbol{\theta} = [\lambda \quad \gamma]^T$ dan $\boldsymbol{y} = [y_1 \quad y_2 \quad \cdots \quad y_n]$. Penaksir *maximum likelihood* (ML) mudah diperoleh melalui fungsi *log-likelihood*, karena $\hat{\boldsymbol{\theta}}$ yang memaksimumkan fungsi *likelihood* juga memaksimumkan fungsi *log-likelihood*. Berdasarkan fungsi *likelihood* (2.5) maka fungsi *log-likelihood* didefinisikan oleh

$$\ell(\boldsymbol{\theta},\mathbf{y}) = \ln(L(\boldsymbol{\theta},\mathbf{y}))$$

$$\ell(\boldsymbol{\theta},\mathbf{y}) = \sum_{i=1}^{n} [\ln\gamma - \ln\lambda + (\gamma-1)(\ln y_i - \ln\lambda)] - \left(\frac{y_i}{\lambda}\right)^{\gamma}. \tag{2.6}$$

Maksimum fungsi *log-likelihood* (2.6) diperoleh dari turunan parsial orde pertama fungsi *log-likelihood* (2.6) terhadap seluruh parameter dan disamadengankan nol sehingga didapatkan persamaan *log likelihood*

$$\frac{\partial \ell(\boldsymbol{\theta},\mathbf{y})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \tag{2.7}$$

dengan ruas kiri pada persamaan (2.7) adalah vektor gradien berdimensi 2 dan $\mathbf{0}$ adalah vektor nol berdimensi 2. Berdasarkan fungsi *log-likelihood* (2.6) maka persamaan (2.7) adalah persamaan non-linear yang tidak bisa diselesaikan secara analitis. Penaksir ML dapat diperoleh secara numerik menggunakan metode Newton-Raphson. Algoritma iterasi Newton-Raphson adalah

$$\hat{\boldsymbol{\theta}}^{(q+1)} = \hat{\boldsymbol{\theta}}^{(q)} - \left[\mathbf{H}\left(\hat{\boldsymbol{\theta}}^{(q)}\right)\right]^{-1} \mathbf{g}\left(\hat{\boldsymbol{\theta}}^{(q)}\right), q = 0,1,2,\ldots \tag{2.8}$$

dengan $\mathbf{g}(\boldsymbol{\theta})$ adalah vektor gradien dan $\mathbf{H}(\boldsymbol{\theta})$ adalah matriks Hessian. Proses iterasi dimulai dari menentukan harga awal $\boldsymbol{\theta}^{(0)} = [\lambda^{(0)} \quad \gamma^{(0)}]^T$ dan iterasi dihentikan sampai dengan iterasi ke-$q+1$ apabila terpenuhi kondisi konvergen, yaitu $\|\hat{\theta}^{(q+1)} \quad \hat{\theta}^{(q)}\| < \varepsilon$, dengan ε menyatakan bilangan *real* positif berukuran *infinitesimal* (sangat kecil) misalnya 10^{-12} dan $\|.\|$ menyatakan norma vektor [15], [25].

Pengujian kecocokan data waktu distribusi Weibull dilakukan menggunakan uji Kolmogorov-Smirnov dengan rumusan hipotesis

$H_0 : F(y) = \hat{F}(y)$

(Data mengikuti distribusi Weibull dengan fungsi distribusi adalah $\hat{F}(y)$)

$H_1 : F(y) \neq \hat{F}(y)$

(Data tidak mengikuti distribusi Weibull dengan fungsi distribusi adalah $\hat{F}(y)$)

Statistik uji diberikan oleh

$$D = \max_{1 \le i \le n} \left|\hat{F}(y_i) - F_n(y_i)\right|, \tag{2.9}$$

dimana $F_n(y_i)$ adalah fungsi distribusi kumulatif empiris yang dibentuk dari sampel acak yang didefinisikan oleh

$$F_n(y_i) = \frac{\text{banyaknya data variabel } y \text{ yang } \le y_i}{n}, \tag{2.10}$$

dan $\hat{F}(y)$ adalah taksiran fungsi distribusi kumulatif berdasarkan persamaan (2.2). Daerah penolakan adalah menolak H_0 pada taraf signifikansi α jika dan hanya jika nilai D melebihi nilai kritis $D_{\alpha,n}$ [9], [15], [25].

2.3 Pemeriksaan Multikolinieritas

Pemeriksaan multikolinieritas pada kovariat numerik dilakukan dengan melihat nilai VIF. Nilai VIF lebih kecil dari 10 mengindikasikan tidak terjadi multikolinieritas antar kovariat pada model regresi. Pemeriksaan multikolinieritas antar kovariat (variabel bebas) berguna untuk menjamin penaksiran parameter regresi dapat dilakukan. Nilai VIF dihitung menggunakan rumus sebagai berikut

$$VIF = \frac{1}{1 - R_k^2}, \tag{2.11}$$

dengan

$$R_k^2 = 1 - \frac{\sum_{i=1}^{n}(x_{ki} - \hat{x}_{ki})^2}{\sum_{i=1}^{n}(x_{ki} - \bar{X}_k)^2}, k = 1,2, \dots, p \tag{2.12}$$

adalah koefisien determinasi dari hasil regresi kovariat ke-k terhadap kovariat sisanya [22], dimana x_{ki} adalah nilai pengamatan kovariat ke-k individu ke-i, $\hat{x}_{ki}$ adalah nilai taksiran kovariat ke-k individu ke-i berdasarkan model regresi linier, dan

$$\bar{X}_k = \frac{1}{n}\sum_{i=1}^{n} x_{ki}, \tag{2.13}$$

adalah rata-rata nilai pengamatan untuk kovariat ke-k [16], [26].

Pendeteksian multikolinieritas antar kovariat kategorik digunakan uji indepedensi dengan koreksi *Yates*. Uji ini digunakan karena masing-masing kovariat terdiri dari 2 kategori, sehingga membentuk tabel kontingensi 2×2. Uji *chi square* dengan koreksi *Yates* memiliki syarat yaitu tidak ada sel dengan frekuensi pengamatan bemilai nol dan tidak ada sel dengan frekuensi harapan kurang dari 5 [24]. Bentuk umum tabel kontingensi 2×2 disajikan pada Tabel 2.2.

Tabel 2.2 Tabel *Uji Chi Square* dengan Koreksi *Yates*

		Kovariat X_l	
		Kategori 1	**Kategori 2**
Kovariat X_k	**Kategori 1**	A	B
	Kategori 2	C	D

Hipotesis uji indepedensi adalah
H_0: Kovariat X_k dan X_l saling bebas
H_1: Kovariat X_k dan X_l tidak saling bebas
dimana $k \neq l$ dengan $k, l = 1,2, \dots, p$. Statistik uji adalah

$$Q = \frac{n\left(|AD - BC| - \frac{n}{2}\right)^2}{(A+B)(C+D)(A+C)(B+D)}, \tag{14}$$

dimana $Q \sim \chi_1^2$ dan

A	:	Frekuensi sampel yang termasuk dalam sel kovariat X_k kategori 1 dan kovariat X_l kategori 1
B	:	Frekuensi sampel yang termasuk dalam sel kovariat X_k kategori 1 dan kovariat X_l kategori 2
C	:	Frekuensi sampel yang termasuk dalam sel kovariat X_k kategori 2 dan kovariat X_l kategori 1
D	:	Frekuensi sampel yang termasuk dalam sel kovariat X_k kategori 2 dan kovariat X_l kategori 2
n	:	Frekuensi sampel penelitian

Daerah penolakan H_0 pada taraf signifikansi α adalah jika $Q > \chi^2_{(1-\alpha);1}$ [28].

2.4 Penaksiran Parameter Model Regresi Cox Weibull

Salah satu metode untuk mengetahui besar pengaruh dari kovariat terhadap waktu adalah model Cox *Proportional Hazard* (PH). Model umum Cox PH diberikan oleh

$$h(y) = h_0(y)\exp(\boldsymbol{\beta}^T\mathbf{x}_i) \tag{2.15}$$

dengan $\boldsymbol{\beta} = [\beta_1 \quad \beta_2 \quad \cdots \quad \beta_p]^T$ adalah vektor parameter regresi, $\mathbf{x}_i = [x_{1i} \quad x_{2i} \quad \cdots \quad x_{pi}]^T$ adalah vektor nilai kovariat untuk individu ke-i, dan $h_0(y)$ adalah fungsi *baseline hazard* [1], [27]. *Baseline hazard* $h_0(y)$ diinterpretasikan sebagai pengaruh waktu (y) terhadap *hazard* $h(y)$ tanpa dipengaruhi faktor eksternal (x). Berdasarkan asumsi bahwa variabel waktu mengikuti distribusi Weibull versi skala-bentuk maka model umum regresi Cox Weibull [5] diberikan oleh

$$h_i(y_i,\mathbf{x}_i) = \gamma\lambda^{-\gamma}y_i^{\gamma-1}\exp(\boldsymbol{\beta}^T\mathbf{x}_i), \tag{2.16}$$

Model umum Cox Weibull pada persamaan (2.16) terdiri dari dua bagian yaitu *baseline hazard* Weibull $h_0(y) = \gamma\lambda^{-\gamma}y^{\gamma-1}$ y dan komponen kovariat $\exp(\boldsymbol{\beta}^T\mathbf{x}_i)$ yang memuat parameter $\boldsymbol{\beta}$. Dua komponen tersebut dapat ditaksir secara terpisah. Penaksiran parameter pada *baseline hazard* dilakukan menggunakan metode MLE pada bagian 2.1, sedangkan parameter $\boldsymbol{\beta}$ dalam model Cox Weibull ditaksir dengan menggunakan metode *Maximum Partial Likelihood Estimation* (MPLE).

Misal diberikan n data sampel $y_1, y_2, \dots, y_n$ yang saling bebas, dimana terdapat n_1 diantaranya mencapai *event* dan n_2 adalah data tersensor $(n_1 + n_2 = n)$ dalam durasi waktu yang berbeda, dengan urutan waktu *event* $y_{(1)} < y_{(2)} < y_{(i)} \dots < y_{(n_1)}$ dan $\mathbf{x}_i$ adalah vektor data pengamatan kovariat dari individu ke-i dimana $i = 1,2, \dots, n$. Misal $R(y_i)$ adalah himpunan dari individu yang berisiko mengalami *event* pada waktu y_i, maka peluang individu mengalami *event* pada waktu y_i dengan syarat $y_i \in R(y_i)$ adalah

$$P(y_i) = \frac{h_i(y_i)}{\sum_{i\in R(y_i)} h_i(y_i)}. \tag{2.17}$$

Misalkan diberikan data waktu pengamatan $(y_i, \boldsymbol{x}_i)$, maka fungsi *partial likelihood* untuk model Cox Weibull pada persamaan (2.16) adalah

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\beta}^T\mathbf{x}_i)}{\sum_{i\in R(y_i)}\exp(\boldsymbol{\beta}^T\mathbf{x}_i)}, \tag{2.18}$$

dengan $\boldsymbol{\beta}^T\mathbf{x}_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}$. Penaksir *maximum partial likelihood* (MPL) model Cox Weibull mudah diperoleh melalui fungsi *log-partial likelihood*, karena $\hat{\boldsymbol{\beta}}$ yang memaksimumkan fungsi *likelihood* juga memaksimumkan fungsi *log-partiallikelihood* [1]. Fungsi *log-partial likelihood* berdasarkan persamaan (2.18) adalah

$$\begin{aligned}\ell(\boldsymbol{\beta}) &= \ln\big(L(\boldsymbol{\beta})\big)\\ &= \sum_{i=1}^{n}\left[\boldsymbol{\beta}^T\mathbf{x}_i - \ln\left(\sum_{i\in R(y_i)}\exp(\boldsymbol{\beta}^T\mathbf{x}_i)\right)\right].\end{aligned} \tag{2.19}$$

Maksimum fungsi *log-partial likelihood* diperoleh dari turunan parsial orde pertama fungsi *log-partial likelihood* (2.19) terhadap parameter $\boldsymbol{\beta}$ dan disamadengankan nol sehingga diperoleh persamaan

$$\frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \mathbf{0}, \tag{2.20}$$

dengan ruas kiri persamaan (2.20) adalah vektor gradien berdimensi p dan $\mathbf{0}$ adalah vektor nol berdimensi p.

Berdasarkan fungsi *log-partial likelihood* (2.19) maka persamaan (2.20) adalah persamaan non-linear sehingga penaksir eksak MPL tidak bisa diperoleh secara analitis dan dapat diperoleh secara numerik menggunakan metode Newton-Raphson. Algoritma iterasi Newton-Rapson adalah

$$\widehat{\boldsymbol{\beta}}^{(q+1)} = \widehat{\boldsymbol{\beta}}^{(q)} - \left[\mathbf{H}\left(\widehat{\boldsymbol{\beta}}^{(q+1)}\right)\right]^{-1} \mathbf{g}\left(\widehat{\boldsymbol{\beta}}^{(q)}\right), q = 0,1,2, \dots \qquad (2.21)$$

dengan $\mathbf{g}$ menyatakan vektor gradien pada persamaan (2.20) dan $\mathbf{H}(\boldsymbol{\beta})$ menyatakan matriks Hessian berukuran $p \times p$, yaitu matriks turunan parsial orde kedua dari fungsi *log-partiallikelihood* (2.19) terhadap seluruh kombinasi vektor $\boldsymbol{\beta}$**.** Proses iterasi dimulai dari $q = 0$ pada persamaan (2.21) sehingga dapat ditulis $\widehat{\boldsymbol{\beta}}^{(1)} = \widehat{\boldsymbol{\beta}}^{(0)} - \left[\mathbf{H}\left(\widehat{\boldsymbol{\beta}}^{(1)}\right)\right]^{-1} \mathbf{g}\left(\widehat{\boldsymbol{\beta}}^{(0)}\right)$, begitu seterusnya. Iterasi dihentikan sampai dengan iterasi $q + 1$ apabila terpenuhi kriteria konvergen, yaitu ketika $\left\|\widehat{\boldsymbol{\beta}}^{(q+1)} \quad \widehat{\boldsymbol{\beta}}^{(q)}\right\| < \varepsilon$, dengan ε adalah bilangan *real* positif yang berukuran *infinitesimal* (sangat kecil), misal 10^{-12} [3].

2.5 Pemeriksaan Asumsi *Proportional Hazard* (PH)

Pemeriksaan asumsi PH dilakukan terpisah antara kovariat kategorik dengan kovariat numerik. Pemeriksaan asumsi PH kovariat kategorik dilakukan dengan melihat pola plot $\ln\left(-\ln\left(S(y)\right)\right)$ terhadap waktu *survival*. Asumsi terpenuhi apabila garis antar kategori pada setiap kovariat sejajar atau tidak berpotongan [7], [11]. Sedangkan pemeriksaan asumsi PH kovariat numerik menggunakan metode *Goodness of Fit* (GoF) dengan menguji apakah terdapat korelasi atau hubungan yang signifikan antara residual *Schoenfeld* dengan peringkat (*rank*) waktu yang diurutkan mulai dari data terkecil. Hipotesis pengujian GoF adalah sebagai berikut

$H_0: \rho = 0$
(Tidak ada hubungan antara residual *Schoenfeld* dengan *rank* waktu)
$H_1: \rho \neq 0$
(Terdapat hubungan antara residual *Schoenfeld* dengan *rank* waktu)

Statistik uji GoF yaitu

$$Z_k = r_k\sqrt{n-1}, \qquad (2.22)$$

dengan

$$r_k = \frac{\sum_{i=1}^{n}(R_{ik} - \bar{R}_{ik})(RT_i - \overline{RT})}{\sqrt{\sum_{i=1}^{n}(R_{ik} - \bar{R}_{ik})^2}\sqrt{\sum_{i=1}^{n}(RT_i - \overline{RT})^2}}, \qquad (2.23)$$

dan

$$R_{ik} = \delta_i\left(x_{ik} - \frac{\sum_{i \in R(y_i)} x_{ik} \exp(\boldsymbol{\beta}^T \mathbf{x}_i)}{\sum_{i \in R(y_i)} \exp(\boldsymbol{\beta}^T \mathbf{x}_i)}\right), k = 1,2, \dots, p \qquad (2.24)$$

dimana R_{ik} adalah nilai residual *Schoenfeld* untuk kovariat ke-k dari individu dengan waktu *event* y_i, δ_i adalah status tersensor, x_{ik} adalah nilai pengamatan kovariat ke-k dari individu ke-i, $\bar{R}_k$ adalah nilai rata-rata residual *Schoenfeld* untuk kovariat ke-k, RT_i adalah *rank* waktu untuk waktu *event* ke-i yang telah diurutkan dari yang terkecil, dan $\overline{RT}$ adalah rata-rata *rank* waktu *event*. Daerah penolakan H_0 pada taraf signifikansi α adalah apabila nilai $|Z_k| > Z_{\alpha/2}$ atau $p_{value} < \alpha$ dengan $Z_{\alpha/2}$ adalah nilai kritis distribusi normal baku pada taraf signifikansi α [1], [3].

2.6 Pengujian Signifikansi Parameter Model Regresi Cox Weibull

Parameter regresi pada model Cox Weibull akan diuji secara simultan yang juga menilai kelayakan model dan secara parsial.

a. Uji Simultan

Hipotesis uji signifikansi parameter regresi secara simultan adalah

$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$ (Model Cox Weibull tidak layak)
$H_1: \exists \beta_k \neq 0, k = 1,2, \dots, p$ (Model Cox Weibull layak)

Statistik pengujian secara simultan menggunakan metode *likelihood ratio test* yaitu

$$G = -2\left[\ell\left(\widehat{\boldsymbol{\theta}}\right) - \ell\left(\widehat{\boldsymbol{\beta}}\right)\right], \qquad (2.25)$$

dimana $G \sim \chi_p^2$,

$$\ell(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \left[\ln \hat{\gamma} - \ln \hat{\lambda} + (\hat{\gamma} - 1)(\ln y_i - \ln \hat{\lambda}) - \left(\frac{y_i}{\hat{\lambda}}\right)^{\hat{\gamma}} \right], \qquad (2.26)$$

adalah nilai maksimum fungsi *log-likelihood* dengan asumsi H_0 benar, dan

$$\ell(\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \left[(\boldsymbol{\beta}^T \mathbf{x}_i) - \ln\left(\sum_{i \epsilon R(y_i)} (\widehat{\boldsymbol{\beta}}^{\mathbf{T}} \mathbf{x_i}) \right) \right], \qquad (2.27)$$

adalah nilai maksimum dari fungsi *log-partial likelihood* berdasarkan model regresi Cox Weibull yang diberikan oleh persamaan (2.19). Daerah penolakan H_0 pada taraf signifikansi α adalah apabila nilai $G > \chi^2_{(1-\alpha;p)}$ atau $p_{value} < \alpha$, dimana $p_{value} = P(G_V > G)$ dengan G_V adalah variabel acak berdistribusi χ_p^2 [25], [26].

b. Uji Parsial

Hipotesis uji signifikansi parameter regresi β_k secara parsial untuk k tertentu ($k = 1,2, \dots, p$) adalah sebagai berikut

$H_0: \beta_k = 0$
(Kovariat X_k tidak berpengaruh terhadap laju *event* (*hazard rate*))
$H_1: \beta_k \neq 0$
(Kovariat X_k berpengaruh terhadap laju *event* (*hazard rate*))

Statitik pengujian parameter secara parsial adalah statistik uji Wald yaitu

$$W = \frac{\hat{\beta}_k}{\sqrt{var(\hat{\beta}_k)}}, \qquad (2.28)$$

dimana $W \sim N(0,1)$ dengan $var(\hat{\beta}_k)$ adalah nilai variansi $\hat{\beta}_k$ yang diperoleh dari elemen ke-k diagonal utama matriks $-[\mathbf{H}(\widehat{\boldsymbol{\beta}})]^{-1}$. Daerah penolakan H_0 pada taraf signifikansi α pengujian ini adalah jika nilai $|W| > Z_{1-\alpha/2}$ atau jika $p_{value} < \alpha$, dimana $p_{value} = P(|Z| > W)$ [5], [15], [26].

2.7 Pemilihan Model Cox Weibull Terbaik

Tujuan pemilihan model Cox Weibull terbaik adalah untuk memperoleh model yang model sederhana tetapi optimal (parsimoni). Salah satu metode pemilihan model terbaik yang umum digunakan adalah metode *backward elimination* [10], [14], [27], [29]. Proses pemilihan model menggunakan *backward elimination* dimulai dari model lengkap yaitu model yang memuat seluruh kovariat dan kemudian kovariat tidak berpengaruh dieliminiasi satu persatu dimulai dari kovariat dengan nilai p_{value} terbesar hingga diperoleh model terbaik. Proses eliminasi didasarkan pada nilai $F_{parsial}$ dan berdasarkan nilai terkecil *Akaike Information Criterion* (AIC) [12], [29]. Perhitungan nilai AIC menggunakan persamaan berikut

$$AIC = -2\ell(\widehat{\boldsymbol{\beta}}) + 2k, \qquad (2.29)$$

dengan $\ell(\widehat{\boldsymbol{\beta}})$ adalah nilai maksimum dari fungsi *log-partiallikelihood* pada persamaan (2.15) dan k adalah banyaknya parameter regresi pada setiap model yang terbentuk [13].

2.8 *Hazard Ratio*

Pada model regresi Cox Weibull yang melibatkan kovariat kategorik, misalkan *hazard* individu pertama dinyatakan oleh $h(y, \mathbf{x}|x_k = 1)$ dan *hazard* individu kedua adalah $h(y, \mathbf{x}|x_k = 0)$ dengan asumsi nilai data waktu dan kovariat yang lain tetap (tidak mengalami perubahan), maka *hazard ratio* (HR) atau perbandingan *hazard rate* dua individu dihitung berdasarkan formula berikut [10]

$$\begin{aligned} HR(X_k) &= \frac{h(y, \mathbf{x}|x_k = 1)}{h(y, \mathbf{x}|x_k = 0)} \\ &= \frac{h_0(y)\exp\left(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k(1) + \cdots + \beta_p x_{pi}\right)}{h_0(y)\exp\left(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k(0) + \cdots + \beta_p x_{pi}\right)} \\ &= \exp(\beta_k - 0) \\ &= \exp(\beta_k) \end{aligned} \tag{2.30}$$

Pada model regresi Cox Weibull yang melibatkan kovariat numerik, misalkan nilai *hazard* individu pertama $adalah$ $h(y, \mathbf{x}|x_k + \mathrm{c})$ dimana c adalah besar perubahan kovariat dan nilai $hazard$ individu kedua adalah $h(y, \mathbf{x}|x_k)$, dengan asumsi nilai data waktu dan kovariat yang lain tetap (tidak mengalami perubahan), maka nilai HR dua individu berdasarkan kovariat numerik dihitung menggunakan formula berikut [14], [15], [25].

$$\begin{aligned} HR(X_k) &= \frac{h(y, \mathbf{x}|x_k + c)}{h(y, \mathbf{x}|x_k)} \\ &= \frac{h_0(y)\exp\left(\beta_1 x_{1i} + \beta_2 \beta_{2i} + \cdots + \beta_k(x_k + \mathrm{c}) + \cdots + \beta_p \beta_{pi}\right)}{h_0(y)\exp\left(\beta_1 x_{1i} + \beta_2 \beta_{2i} + \cdots + \beta_k(x_k) + \cdots + \beta_p \beta_{pi}\right)} \\ &= \exp(\beta_k(x_k + \mathrm{c}) - \beta_k x_k) \\ &= \exp(c\beta_k) \end{aligned} \tag{2.31}$$

3. HASIL DAN PEMBAHASAN

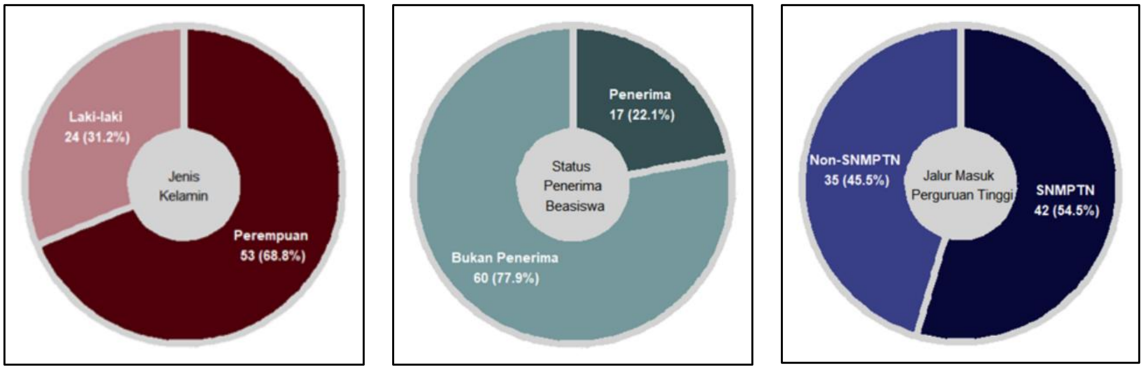
3.1 Deskripsi Data Penelitian

Deskripsi data penelitian untuk data waktu dan data kovariat numerik dinyatakan dengan nilai rata-rata, minimum, maksimum, dan simpangan baku pada Tabel 3.1, sedangkan deskripsi data kovariat kategorik dinyatakan dengan persentase setiap kategori melalui diagram lingkaran pada Gambar 3.1. Berdasarkan Tabel 3.1, rata-rata lama studi mahasiswa prodi statistika sebelum penyensoran data adalah 4,540 tahun. Lama studi tercepat adalah 3,750 tahun dan paling lama adalah 7 tahunj, dengan simpangan baku 0,718 tahun. Rata-rata lama studi setelah penyensoran data dengan maksimal studi 5 tahun menjadi 4,396 tahun, dengan simpangan baku 0,394 tahun.

Tabel 3.1 Deskripsi Data Penelitian Kovariat Numerik

Data	Rata-rata	Minimum	Maksimum	Simpangan Baku
Lama Studi (Y^*)	4,540	3,750	7,000	0,718
Lama Studi (Y)	4,396	3,750	5,000	0,394
Usia (X_1)	18,200	16,929	20,457	0,646
IPK (X_3)	3,479	2,800	3,930	0,216
Beban Studi (X_4)	146,900	144,000	151,000	1,757

Rata-rata usia mahasiswa saat masuk prodi statistika adalah 18,2 tahun, paling muda adalah 16,929 tahun, dan paling tua adalah 20,457 tahun dengan simpangan baku sebesar 0,646. Rata-rata IPK lulusan prodi statistika adalah sebesar 3,479, IPK terendah sebesar 2,8, dan tertinggi sebesar 3,930 dengan simpangan baku sebesar 0,216. Terakhir rata-rata beban studi yang diambil selama masa perkuliahan mahasiswa prodi statistika adalah 147 SKS, paling sedikit sebanyak 144 SKS, dan paling banyak 151 SKS dengan simpangan baku sebesar 1,757.



(a) (b) (c)

Gambar 3.1 Lulusan Prodi Statistika Berdasarkan Kovariat Kategorik

Gambar 3.1(a) menunjukkan proporsi lulusan prodi statistika berdasarkan jenis kelamin (X_2), gambar 3.1(b) menunjukkan proporsi lulusan prodi statistika berdasarkan status penerima beasiswa (X_5), dam gambar 3.1(c) menunjukkan proporsi lulusan prodi statistika berdasarkan jalur masuk PT (X_6). Persentase lulusan prodi statistika berjenis kelamin perempuan adalah sebesar 68,8% atau sebanyak 53 mahasiswa sedangkan 31,2% atau 24 mahasiswa lainnya berjenis kelamin laki-laki. Persentase lulusan prodi statistika yang berstatus sebagai penerima beasiswa adalah sebesar 77,9% atau sebanyak 60 mahasiswa sedangkan 22,1% atau 17 mahasiswa lainnya bukan penerima beasiswa. Terakhir persentase lulusan prodi statistika yang masuk perguruan tinggi melalui jalur SNMPTN adalah sebesar 54,5% atau sebanyak 42 mahasiswa sedangkan 45,5% atau 35 mahasiwa lainnya masuk melalui jalur non-SNMPTN yaitu SBMPTN dan SMMPTN.

3.2 Penaksiran Parameter dan Pengujian Distribusi Weibull

Parameter distribusi Weibull ditaksir menggunakan metode MLE yang diselesaikan secara numerik menggunakan metode Newton-Raphson pada data lama studi mahasiswa yang mengalami *event*. *Event* penelitian ini adalah kelulusan mahasiswa dengan lama studi kurang dari atau tepat 5 tahun. Hasil nilai taksiran parameter distribusi Weibull menggunakan *software* RStudio dapat dilihat pada Tabel 3.2.

Tabel 3.2 Nilai Taksiran Parameter Distribusi Weibull

Parameter	Taksiran
Bentuk (γ)	13,9714
Skala (λ)	4,4203

Berdasarkan Tabel 3.2, diperoleh fungsi distribusi kumulatif berdasarkan persamaan (2.2) adalah

$$\hat{F}(y) = 1 - \exp\left[-\left(\frac{y}{4,4203}\right)^{13,9714}\right]. \tag{3.1}$$

dan fungsi *hazard* Weibull berdasarkan persamaan (2.4) adalah

$$\hat{h}(y) = (1,3408 \times 10^{-8})y^{12,9714} \tag{3.2}$$

Pengujian distribusi Weibull dilakukan pada data lama studi menggunakan uji Kolmogorov-Smirnov dengan hipotesis adalah

$H_0 : F(y) = \hat{F}(y)$

(Data lama studi berdistribusi Weibull dengan $\hat{F}(y)$ diberikan oleh persamaan (3.1))

$H_1 : F(y) \neq \hat{F}(y)$

(Data lama studi berdistribusi Weibull dengan $\hat{F}(y)$ diberikan oleh persamaan (3.1))

Hasil perhitungan uji Kolmogorov-Smirnov memberikan nilai $D_{maks} = 0,105$ yang mana lebih kecil dari nilai $D_{(77;0,05)} = 0,155$ sehingga diputuskan H_0 diterima. Hal ini berarti bahwa lama studi mahasiswa Prodi Statistika Universitas Mulawarman berdistribusi Weibull dengan parameter bentuk (λ) sebesar 13,9714 dan parameter skala (γ) sebesar 4,4203.

3.3 Pemeriksaan Multikolinieritas

Pemeriksaan multikolinieritas menggunakan nilai VIF dilakukan antar kovariat numerik yaitu usia (X_1), IPK (X_3), dan beban studi (X_4). Hasil perhitungan nilai VIF disajikan pada Tabel 3.3

Tabel 3.3. Nilai VIF tiap Kovatiat Numerik

Kovariat	**Nilai VIF**
Usia (X_1)	1,027
IPK (X_3)	1,350
Beban Studi (X_4)	1,175

Berdasarkan Tabel 3.3, diketahui nilai VIF pada masing-masing kovariat numerik bernilai kurang dari 10 sehingga dapat disimpulkan bahwa antar kovariat numerik yaitu usia (X_1), IPK (X_3), dan beban studi (X_4) tidak terjadi multikolinieritas.

Pemeriksaan multikolinieritas antar kovariat kategorik yaitu jenis kelamin (X_1), status beasiswa (X_5), dan jalur masuk (X_6) menggunakan uji independensi dengan hipotesis sebagai berikut

H_0: Kovariat X_k dan X_l saling bebas
H_1: Kovariat X_k dan X_l tidak saling bebas

dimana $k \neq l$ dengan $k = 2{,}5$ dan $l = 5{,}6$. Hasil pengujian indepedensi menggunakan *software* Rstudio disajikan pada Tabel 3.4

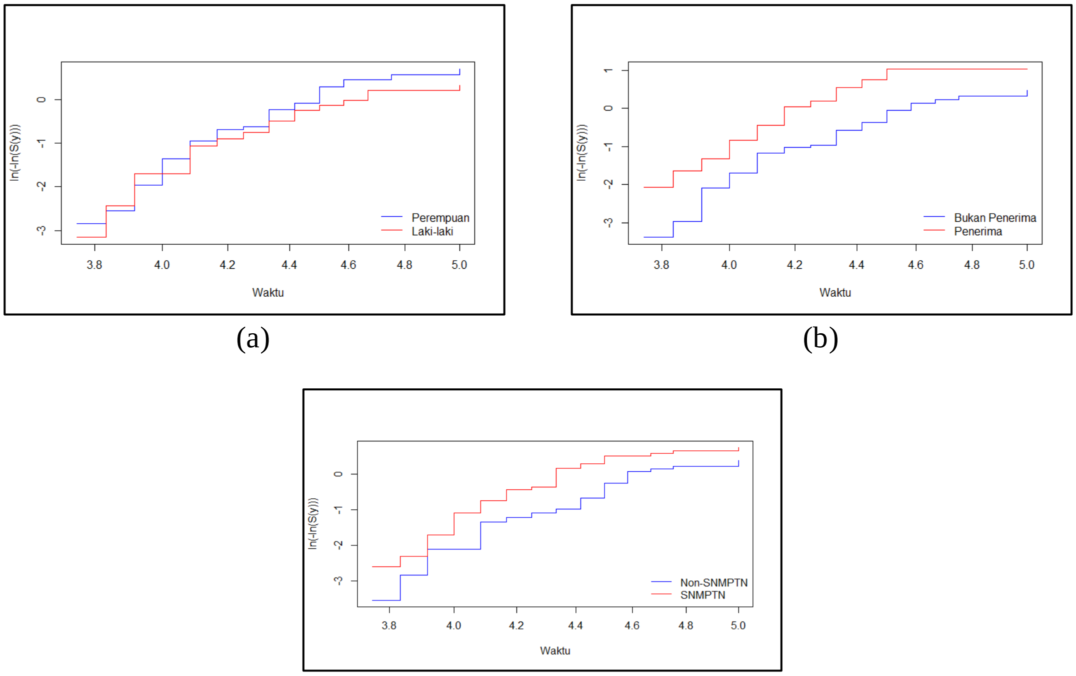
Tabel 3.4. Hasil Uji Indepedensi antar Kovariat Kategorik

Kovariat	**Kovariat**	$\boldsymbol{Q}$	$\boldsymbol{\chi^2_{0,95;1}}$	**Kesimpulan**
X_2	X_5	$3{,}093 \times 10^{-30}$	3,841	Saling bebas
	X_6	0,085		Saling bebas
X_5	X_6	3,171	3,841	Saling bebas

Berdasarkan Tabel 3.4, diketahui bahwa nilai statistik uji Q seluruh kovariat kategorik lebih kecil dari nilai kritis $\chi^2_{0,95;1}$ sebesar 3,841 sehingga disimpulkan bahwa kovariat jenis kelamin (X_1), status beasiswa (X_5), dan jalur masuk (X_6) saling bebas yang artinya tidak terjadi multikolinieritas antar kovariat kategorik.

3.4 Pemeriksaan Asumsi PH

Asumsi PH pada kovariat kategorik yaitu jenis kelamin (X_2), status penerima beasiswa (X_5), dan jalur masuk perguruan tinggi (X_6) diperiksa dengan melihat pola pada plot $\ln\left(-\ln\left(S(y)\right)\right)$ terhadap waktu *survival*. Asumsi terpenuhi apabila kurva tiap kategori pada masing-masing kovariat tampak sejajar dan tidak menyimpang tajam satu sama lain.



(a) (b)

(c)

Gambar 3.2 Plot $\ln\left(-\ln\left(S(y)\right)\right)$ terhadap waktu *survival*

Gambar 3.2(a) menunjukkan plot $\ln\left(-\ln\left(S(y)\right)\right)$ untuk kovariat jenis kelamin (X_2), gambar 3.2(b) menunjukkan plot $\ln\left(-\ln\left(S(y)\right)\right)$ untuk kovariat status penerima beasiswa (X_5), dan gambar 3.2(c) menunjukkan plot $\ln\left(-\ln\left(S(y)\right)\right)$ untuk kovariat jalur masuk perguruan tinggi (X_6). Berdasarkan gambar 3.2(a), 3.2 (b), maupun 3.2 (c), kurva merah dan biru pada masing-masing plot membentuk pola yang cenderung sejajar dan tidak menyimpang tajam satu sama lain. Hal ini berarti asumsi PH pada kovariat kategorik yaitu jenis kelamin (X_2), status penerima beasiswa (X_5), dan jalur masuk perguruan tinggi (X_6) terpenuhi sehingga seluruh kovariat kategorik dapat digunakan dalam pemodelan Cox Weibull.

Pengujian asumsi PH pada kovariat numerik yaitu usia (X_1), IPK (X_3), dan beban studi (X_4) menggunakan metode GoF dimana asumsi terpenuhi apabila antara residual *Schoenfeld* dengan peringkat waktu *survival* tidak saling berkorelasi pada kovariat yang diduga berpengaruh terhadap kelulusan mahasiswa. Hasil pengujian GoF dapat dilihat pada Tabel 3.5.

Tabel 3.5 Hasil Pengujian GoF

Kovariat	p_{value}
Usia (X_1)	0,910
IPK (X_3)	0,110
Beban Studi (X_4)	0,360

Hasil pengujian GoF seperti yang terlihat pada Tabel 3.5, didapatkan seluruh nilai p_{value} kovariat numerik lebih besar dari taraf signifikansi 0,05 sehingga disimpulkan bahwa kovariat usia (X_1), IPK (X_3), dan beban studi (X_4) memenuhi asumsi PH dan dapat digunakan dalam pemodelan Cox Weibull.

3.5 Penaksiran Parameter dan Pemilihan Model Terbaik

Model umum Cox Weibull dengan 6 kovariat berdasarkan persamaan (2.16) adalah sebagai berikut

$$\hat{h}(y,\mathbf{x}) = \gamma\lambda^{-\gamma}y^{\gamma-1}\exp\left(\hat{\beta}_1X_1 + \hat{\beta}_2X_2 + \hat{\beta}_3X_3 + \hat{\beta}_4X_4 + \hat{\beta}_5X_5 + \hat{\beta}_6X_6\right), \tag{3.3}$$

dimana parameter skala dan bentuk pada *baseline hazard* telah ditaksir sebelumnya menggunakan MLE pada bagian 3.1 dan diperoleh persamaan *baseline hazard* Weibull (3.2).

Pemilihan model terbaik menggunakan metode *backward elimination* dan dengan kriteria: semua kovariat dalam model berpengaruh signifikan, kovariat berpengaruh signifikan paling banyak, menghasilkan model Cox Weibull yang layak atau *fit*, dan memberikan nilai AIC yang paling kecil. Proses pemilihan model terbaik dapat dilihat pada Tabel 3.6.

Tabel 3.6 Pemilihan Model Terbaik dengan *Backward Elimination*

Kombinasi Kovariat	**Kovariat Berpengaruh**	**Uji Simultan**	**Nilai AIC**
X_1, X_2, X_3, X_4, X_5, dan X_6	X_3 dan X_6	Model Layak	467,05
X_1, X_3, X_4, X_5, dan X_6	X_3 dan X_6	Model Layak	465,09
X_3, X_4, X_5, dan X_6	X_3 dan X_6	Model Layak	463,27
X_3, X_5, dan X_6	X_3, X_5, dan X_6	Model Layak	462,34

Model terbaik berdasarkan proses *backward elimination* pada Tabel 3.6 adalah model dengan tiga kovariat yaitu IPK (X_3), status penerima beasiswa (X_5), dan jalur masuk perguruan tinggi (X_6) dimana ketiga kovariat ini menghasilkan model regresi Cox Weibull yang layak, seluruh kovariat berpengaruh (signifikan), kovariat berpengaruh paling banyak, dan memberikan nilai AIC terkecil, yaitu sebesar 462,34.

Nilai taksiran setiap parameter regresi Cox Weibull terbaik dengan metode MPLE disajikan pada Tabel 3.7, yang juga dilengkapi nilai $|W|$ dan p_{value} sebagai dasar penentuan daerah kritis pengujian parsial.

Tabel 3.7 Penaksiran Parameter Regresi Cox Weibull

Kovariat	$\hat{\boldsymbol{\beta}}_{\boldsymbol{k}}$	$\lvert \boldsymbol{W_k} \rvert$	$\boldsymbol{p_{value}}$
IPK (X_3)	2,0318	2,941	0,0033
Status Penerima Beasiswa (X_5)	0,6617	2,115	0,0312
Jalur masuk perguruan tinggi (X_6)	0,5653	2,104	0,0354

Model regresi Cox Weibull dengan parameter disajikam pada Tabel 3.7 merupakan model yang layak, dimana kelayakan model dibuktikan melalui uji statistik yang terdiri dari uji simultan dan parsial. Uji signifikansi parameter secara simultan dengan hipotesis

$H_0 : \beta_3 = \beta_5 = \beta_6 = 0$

$H_1 : \exists\beta_k \neq 0$, untuk $k = 3,5,6$

Statistik uji berdasarkan persamaan (2.21) adalah $G = 19,69$ dengan $p_{value} = 0,0002$. Nilai G sebesar 19,69 lebih besar dari nilai kritis $\chi^2_{0,05;3} = 7,815$ sehingga diputuskan H_0 ditolak dan disimpulkan bahwa model regresi Cox Weibull terbaik layak (*fit*).

Pengujian signifikansi parameter regresi secara parsial β_k dengan $k = 3,5,6$ dan hipotesis

$H_0: \beta_k = 0$
$H_1: \beta_k \neq 0$
Nilai $|W_k|$ seluruh kovariat pada Tabel 3.7 lebih besar dari nilai kritis $Z_{0,975} = 1{,}96$ atau p_{value} lebih kecil dari taraf signifikansi 0,05 sehingga disimpulkan IPK, status penerima beasiswa, dan jalur masuk perguruan tinggi masing-masing berpengaruh signifikan terhadap laju kelulusan mahasiswa Prodi Statistika Universitas Mulawarman.

Berdasarkan *baseline hazard* Weibull pada persamaan (3.2) dan hasil penaksiran parameter regresi pada Tabel 3.7, maka diperoleh model regresi Cox Weibull terbaik adalah

$$\hat{h}(y) = (1{,}3408 \times 10^{-8})y^{12{,}9714} \times \exp(2{,}0318X_3 + 0{,}6617X_5 + 0{,}5653\,X_6) \quad (3.4)$$

3.6 Interpretasi Model Regresi Cox Weibull

Hasil taksiran setiap parameter model regresi Cox Weibull terbaik (3.4) diinterpretasikan menggunakan *hazard ratio* (HR) berdasarkan persamaan (2.31) untuk kovariat IPK dan persamaan (2.30) untuk kovariat status penerima beasiswa dan jalur masuk perguruan tinggi.

Diketahui kovariat IPK (x_3) dalam penelitian ini adalah numerik (rasio), dengan mengacu persamaaan (2.31) perhitungan nilai HR berdasarkan kovariat IPK ketika terjadi kenaikan sebesar 0,1 adalah

$$HR(X_3) = \frac{h(y, \mathbf{x}|x_3 + 0{,}1)}{h(y, \mathbf{x})} = \exp(0{,}1 \times 2{,}0318) = 1{,}225$$

$HR(X_3) = 1{,}225$ menunjukkan bahwa peningkatan IPK sebesar 0,1 akan mempercepat laju kelulusan menjadi 1,225 kali lebih cepat atau laju kelulusan meningkat sebesar 22,5%. Hal ini berati mahasiswa dengan IPK lebih tinggi cenderung lulus lebih cepat dibandingkan mahasiswa dengan IPK lebih rendah.

Mengacu persamaan (2.30), perhitungan nilai HR berdasarkan kovariat status penerima beasiswa (x_5) yaitu

$$HR(X_5) = \frac{h(y, \mathbf{x}|x_5 = 1)}{h(y, \mathbf{x}|x_5 = 0)} = \exp(0{,}6617) = 1{,}938$$

Berdasarkan model PH [7], [25] nilai $HR(X_5) = 1{,}938$ menunjukkan bahwa $h(y, \mathbf{x}|x_5 = 1) = 1{,}938h(y, \mathbf{x}|x_5 = 0)$ yang berarti laju kelulusan mahasiswa dengan status penerima beasiswa adalah 1,938 kali laju kelulusan mahasiswa bukan penerima beasiswa. Nilai HR positip lebih dari 1 (sebesar 1,938) mengkonfirmasi bahwa kelulusan mahasiswa dengan status penerima beasiswa lebih cepat dibanding mahasiswa dengan status bukan penerima beasiswa.

Diketahui kovariat jalur masuk perguruan tinggi (X_6) dalam penelitian ini adalah kategorik (nominal) biner. Dengan mengacu pada persamaan (2.30, hasil perhitungan nilai *hazard ratio* (HR) kelulusan mahasiswa yang masuk perguruan tinggi melalui jalur SNMPTN ($x_6 = 1$) dan mahasiswa dengan jalur masuk perguruan tinggi non-SNMPTN ($x_6 = 0$), dengan mengasumsikan nilai data waktu dan kovariat lainnya tetap diperoleh

$$HR(X_6) = \frac{h(y, \mathbf{x}|x_6 = 1)}{h(y, \mathbf{x}|x_6 = 0)} = \exp(0{,}5653) = 1{,}760,$$

dengan $h(y, \mathbf{x})|x_6 = 1)$ menyatakan nilai *hazard rate* (laju kelulusan) mahasiswa yang masuk perguruan tinggi melalui jalur SNMPTN, dan $h(y, \mathbf{x})|x_6 = 0)$ menyatakan laju kelulusan mahasiswa yang masuk perguruan tinggi melalui jalur non-SNMPTN. Diketahui bahwa model regresi Cox Weibull merupakan model PH [7], [25] dan berdasarkan hasil perhitungan $HR(X_6)$ diperoleh informasi bahwa $h(y|x_6 = 1) = 1{,}760h(y|x_6 = 0)$, ini menunjukkan bahwa laju kelulusan mahasiswa yang masuk perguruan tinggi melalui jalur SNMPTN adalah 1,760 kali laju kelulusan mahasiswa yang masuk perguruan tinggi melalui jalur non-SNMPTN. Nilai $HR(X_6)$ sebesar 1,760 (lebih dari 1) menunjukkan bahwa mahasiswa yang masuk perguruan tinggi melalui jalur SNMPTN lulus lebih cepat dibanding mahasiswa yang masuk perguruan tinggi melalui jalur non-SNMPTN.

Hasil perhitungan nilai HR seluruh kovariat berpengaruh tersebut disajikan pada Tabel 3.8 berikut

Tabel 3.8 Nilai HR Setiap Kovariat Berpengaruh

Kovariat	$\hat{\beta}_k$	HR
IPK (X_3)	2,0318	1,225
Status Penerima Beasiswa (X_5)	0,6617	1,938
Jalur masuk perguruan tinggi (X_6)	0,5653	1,760

4. KESIMPULAN

Model regresi Cox Weibull yang menyatakan model laju kelulusan mahasiswa Prodi Statistika Universitas Mulawarman adalah

$$\hat{h}(y, \mathbf{x}) = (1{,}3408 \times 10^{-8})y^{12{,}9714} \exp(2{,}0318\, X_3 + 0{,}6617 X_5 + 0{,}5653\ X_6),$$

dengan y menyatakan waktu (lama) studi mahasiswa, X_3 adalah IPK, X_5 adalah status penerima beasiswa, X_6 adalah jalur masuk perguruan tinggi, dan $\hat{h}(y, \mathbf{x}) = (1{,}3408 \times 10^{-8})y^{12{,}9714}$ adalah fungsi *hazard* Weibull yang menyatakan *baseline*. Faktor-faktor yang berpengaruh signifikan terhadap laju kelulusan mahasiswa Prodi Statistika adalah IPK, status beasiswa, dan jalur masuk perguruan tinggi. Berdasarkan nilai *hazard ratio* dari setiap kovariat, diketahui bahwa peningkatan IPK akan mempercepat laju kelulusan mahasiswa. Laju kelulusan mahasiswa penerima beasiswa lebih cepat daripada laju kelulusan mahasiswa bukan penerima beasiswa. Laju kelulusan mahasiswa dengan jalur masuk SNMPTN lebih cepat daripada laju kelulusan mahasiswa dengan jalur masuk non-SNMPTN.

DAFTAR PUSTAKA

[1] Abdallh, G.Y.I. & Algamal, Z.Y., 2023. An investigational modeling approach for improving gene selection using regularized Cox regression model. *Математическая биология и биоинформатика*, Vol. 18, No. 2, 282-293. https://doi.org/10.17537/2023.18.282

[2] Abebe, K.B., 2025. Predictors associated with time to default for HIV/AIDS patients under HAART at Debre Tabor Referral Hospital: a Cox regression model. *Scientific Reports*, Vol. 15, No. 1, 2940. https://doi.org/10.1038/s41598-024-83335-1

[3] Akbar, I., Suyitno & Wahyuningsih, S., 2020. Model Regresi Cox Weibull dengan Metode Penaksiran Parameter Efron Partial Likelihood (Studi Kasus: Lama Perawatan Pasien Penderita Tuberkulosis di Puskesmas Loa Ipuh Tenggarong Tahun 2017). *Jurnal EKSPONENSIAL*, Vol. 11, No. 1, 1–8.

https://doi.org/https://doi.org/10.30872/eksponensial.v11i1.639

[4] Anggraira, A. S., Nirwana, M. B. & Subanti, S., 2023. Cox Proportional Hazard Regression Analysis to Identify Factors Influencing Student Study Period. *Radiant*, Vol. 4, No. 2, 85–94. https://doi.org/10.52187/rdt.v4i2.169

[5] Azizy, A., Suyitno & Siringoringo, M., 2023. Model Regresi Cox Proportional Hazard Weibull pada Data Waktu Rawat Inap Pasien Penderita Covid-19 di RSUD Abdul Wahab Sjahranie Samarinda. *Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya*, 143–160.

[6] Carroll, K. J., 2003. On The Use and Utility of The Weibull Model in The Analysis of Survival Data. *Controlled Clinical Trials*, Vol. 24, No. 6, 682–701. https://doi.org/10.1016/S0197-2456(03)00072-2

[7] Collet, D., 2015. *Modelling Survival Data in Medical Research*, Second Edition, Chapman and Hall/CRC, Bristol, UK. https://doi.org/https://doi.org/10.1201/b18041
[8] Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 34, No. 2, 187–220. https://doi.org/https://doi.org/10.1111/j.2517-6161.1972.tb00899.x
[9] Fajriati, N. A., Suyitno, S. & Wasono, W., 2022. Model Regresi Hazard Rate Weibull Kesembuhan Pasien Rawat Inap Demam Berdarah Dengue (DBD) di RSUD Panglima Sebaya Tanah Grogot. *Eksponensial*, Vol. 13, No. 1, 35-43. https://doi.org/10.30872/eksponensial.v13i1.878
[10] Farahdiba, S., Kartini, D., Nugroho, R.A., Herteno, R. & Saragih, T.H., 2023. Backward elimination for feature selection on breast cancer classification using logistic regression and support vector machine algorithms. *IJCCS (Indonesian J. Comput. Cybernetics Systems)*, Vol. 17, No. 4, 429-440. https://doi.org/10.22146/ijccs.88926
[11] Fathurahman, M., 2022. PEMILIHAN MODEL REGRESI BERNOULLI TERBAIK. *Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya*, 138–146.
[12] Gomes, A., Costa, B., Nunes, V. & Coelho, C., 2026. Cox Regression in Survival Analysis: Practical Insights for Clinicians. *Acta Médica Portuguesa*, 1-12 https://doi.org/10.20344/amp.23078
[13] Gómez, Y. M., Caimanque, W. E., Santibáñez, J. L., Zúñiga-Tapia, R., Gallardo, D. I. & Bourguignon, M., 2025. A new proportional hazard model with applications to breastfeeding data. *Scientific Reports*, Vol. 15, No. 1, 21869. https://doi.org/https://doi.org/10.1038/s41598-025-08219-4
[14] Irfandha, N. M., 2021. Survival Analysis in Patients with Dengue Hemorrhagic Fever (DHF) using Weibull Parametric Model in Indonesia: Case Study at the Regional General Hospital (RSUD) Makassar. *International Conference on Health and Well Being*, 72–83.
[15] Khairunnisa, S. F., Suyitno, S. & Mahmuda, S., 2023. Model Regresi Weibull pada Data Waktu Rawat Inap Pasien COVID-19 di RSUD Abdul Wahab Sjahranie Samarinda. *Jurnal Matematika, Statistika Dan Komputasi*, Vol. 19, No. 2, 286–303. https://doi.org/10.20956/j.v19i2.22266
[16] Klein, J. P. & Moeschberger, M. L., 2003. *Survival Analysis: Techniques for Censored and Truncated Data (Vol. 1230)*. Springer, New York.
[17] LAMSAMA, 2021. *Kriteria dan Prosedur IAPS-LAMSAMA 1.0*.
[18] Li, X., Marcus, D., Russell, J., Aboagye, E. O., Ellis, L. B., Sheeka, A., Park, W. H. E., Bharwani, N., Ghaem-Maghami, S. & Rockall, A. G., 2024. Weibull Parametric Model for Survival Analysis in Women with Endometrial Cancer using Clinical and T2-Weighted MRI Radiomic Features. *BMC Medical Research Methodology*, Vol. 24, No. 1, 2-9. https://doi.org/10.1186/s12874-024-02234-1
[19] Liu, E., Liu, R. Y. & Lim, K., 2023. Using the Weibull Accelerated Failure Time Regression Model to Predict Time to Health Events. *Applied Sciences*, Vol. 13, No. 24, 13041. https://doi.org/10.3390/app132413041
[20] Loder, A. K. F., 2024. Multiple Enrollment Policy: Survival Analyses and Odds of Graduating in at Least One University Degree Program. *Trends in Higher Education*, Vol. 3, No.3, 578–601. https://doi.org/10.3390/higheredu3030034
[21] Rachmawati, R., Afandi, N. & Alwansyah, M. A., 2025. Survival Analysis on Data of Students Not Graduating on Time Using Weibull Regression, Cox Proportional Hazards Regression, and Random Survival Forest Methods. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, Vol.19, No. 3, 2111–2126. https://doi.org/https://doi.org/10.30598/barekengvol19iss3pp2111-2126

[22] Rencher, A. C. & Schaalje, G. B., 2008. *Linear Models in Statistics*. John Wiley & Sons, Inc., New York.
[23] Sirohi, A., Alsaedi, B. S. O., Ahelali, M. H. & Jayaswal, M. K., 2023. Biased proportional hazard regression estimator in the existence of collinearity. *Heliyon*, Vol. 9, No. 11, 1-14. https://doi.org/10.1016/j.heliyon.2023.e21394
[24] Sulistyani, D. O. & Purhadi, 2013. Analisis Terhadap Faktor-Faktor yang Mempengaruhi Laju Perbaikan Kondisi Klinis Pasien Penderita Stroke dengan Regresi Cox Weibull. *Jurnal Sains Dan Seni Pomits*, Vol.2, No. 1, 72–77.
[25] Suyitno, Darnah, Dani, A. T. R. & Oktavia, N. T., 2025. Modeling the Hospitalization Time of Stroke Patients at Abdul Wahab Sjahranie Hospital Samarinda using the Weibull Regression Model. *MethodsX*, Vol. 14, 103082. https://doi.org/10.1016/j.mex.2024.103082
[26] Suyitno, S., Darnah, Hayati, M. N., Dani, A. T. R., Purnamasari, I., Goejantoro, R., Siringoringo, M., Nugraha, P. Y., Fauziyah, M., Fauziyah, Z. N., & Mislan, 2026. Geographically weighted Weibull regression modeling on dissolved oxygen data to analyze river water quality in East Kalimantan. *MethodsX*, Vol.16, 103745. https://doi.org/https://doi.org/10.1016/j.mex.2025.103745
[27] Wang, Y., Deng, Y., Tan, Y., Zhou, M., Jiang, Y. & Liu, B., 2023. A comparison of random survival forest and Cox regression for prediction of mortality in patients with hemorrhagic stroke. *BMC medical informatics and decision making*, Vol. 23, No. 1, 215:1-11 https://doi.org/10.1186/s12911-023-02293-2
[28] Wulansari, A. D., 2023. *Aplikasi Statistika Nonparametrik dalam Penelitian*. Thalibul Ilmi Publishing & Education.
[29] Yulianti, V. N. & Sembiring, P., 2023. Penerapan Metode Backward untuk Menentukan Persamaan Regresi Linier Berganda pada Dugaan Tindak Pidana di Kota Binjai. *FARABI: Jurnal Matematika Dan Pendidikan Matematika*, Vol. 6, No. 1, 1–9. https://doi.org/10.47662/farabi.v6i1.425