



Waiting Time to Get a Job: Comparison of Survival Analysis of Semi-Parametric & Parametric Models

Firdaus Harmajati , Muhammad Bayu Nirwana, Sugiyanto

Department of Statistics, Universitas Sebelas Maret, Surakarta 57126, Indonesia

 firdausharmajati@student.uns.ac.id

 <https://doi.org/10.53017/ujas.248>

Received: 11/03/2023

Revised: 27/03/2023

Accepted: 29/03/2023

Abstract

Education is very important along with the advancements in technology that required people to have in-depth and quality knowledge. A person who can get a job usually has a good educational background. How long it takes a person to get a job is affected by many factors. In statistical sciences, factors affecting how long a person gets a job after graduating from college can be studied using survival analysis. Survival analysis is a regression method used to examine the factors that influence the occurrence of one or more events. Several survival models can be used to estimate the impact of survival factors on how long a person gets a job, including the cox proportional hazard model and the accelerated failure time model. The cox proportional hazard model belongs to the semi-parametric approach where the hazard function essentially follows the non-parametric model while the independent variables follow the parametric model. The accelerated failure time model is a parametric model based on data distribution that can predict the time of the event in observed data. This study was conducted by comparing the cox proportional hazard model and the accelerated failure time model with the logistics distribution to determine if the conditions or data types are better identified with the cox proportional hazard model or the failure time model. The results obtained by considering the minimum AIC value, the accelerated failure time model has a better logistics distribution than the cox proportional hazard model.

Keywords: Survival analysis; Cox proportional hazard model; Accelerated failure time model

Perbandingan Analisis *Survival* Model Semiparametrik dan Model Parametrik Pada Waktu Tunggu Memperoleh Pekerjaan

Abstrak

Pendidikan sangatlah penting seiring dengan kemajuan teknologi yang menuntut manusia untuk memiliki pengetahuan luas dan berkualitas. Seorang yang mampu mendapatkan pekerjaan tentu biasanya dilatarbelakangi dengan pendidikan yang baik. Lama seseorang dalam memperoleh pekerjaan dipengaruhi oleh banyak faktor. Adapun dalam ilmu statistika faktor-faktor yang dapat mempengaruhi waktu tunggu seseorang memperoleh pekerjaan dapat dikaji menggunakan analisis *survival*. Analisis *survival* adalah metode regresi yang digunakan untuk melihat faktor-faktor yang berpengaruh pada terjadinya suatu peristiwa atau kejadian. Beberapa model *survival* dapat digunakan untuk menduga pengaruh dari faktor *survival* pada lama waktu tunggu memperoleh pekerjaan, salah satunya adalah model *cox proportional hazard* dan model *accelerated failure time*. Model *cox proportional hazard* termasuk ke dalam metode semiparametrik yang mana fungsi *hazard* dasarnya mengikuti model nonparametrik sedangkan, variabel-variabel independennya mengikuti model parametrik. Model *accelerated failure time* merupakan salah satu model parametrik yang didasarkan pada distribusi data yang dapat memprediksi waktu suatu kejadian yang terjadi pada data observasi. Penelitian kali ini dilakukan dengan perbandingan model *cox proportional hazard* dan model *accelerated failure time* berdistribusi logistik untuk mengetahui apakah kondisi atau tipe data lebih baik

digunakan pada model *cox proportional hazard* atau model *accelerated failure time*. Hasil yang diperoleh dengan mempertimbangkan nilai AIC terkecil, model *accelerated failure time* berdistribusi logistik lebih baik dibandingkan dengan model *cox proportional hazard*.

Kata kunci: Analisis survival; Model *cox proportional hazard*; Model *accelerated failure time*

1. Pendahuluan

Pendidikan sangatlah penting seiring dengan kemajuan teknologi yang menuntut manusia untuk memiliki pengetahuan luas dan berkualitas. Salah satu cara membentuk manusia berkualitas yaitu berdasarkan tingkat pendidikan yang semakin tinggi. Perguruan tinggi merupakan tingkat pendidikan yang mencetak dan mempersiapkan lulusan–lulusan yang mampu bersaing dan siap menghadapi dunia kerja. Seseorang setelah lulus pada jenjang perkuliahan salah satu tujuannya adalah mencari pekerjaan. Namun, dapat dilihat pada kenyataan saat ini bahwa seseorang memiliki gelar sarjana tidak lagi dapat menjamin seseorang untuk mudah mendapatkan pekerjaan. Menurut data Badan Pusat Statistik [1], menunjukkan per Februari 2022, tingkat pengangguran Indonesia tercatat sebesar 5,83% dari total penduduk usia kerja 208,54 juta orang. Dari 5,83% tersebut hampir 14% adalah penduduk dengan lulusan jenjang sarjana. Melihat dari hasil data yang tercatat, ternyata seseorang yang memiliki pendidikan yang tinggi belum tentu berhasil mendapatkan pekerjaan pertamanya secara cepat setelah lulus.

Menurut [2], analisis *survival* adalah suatu metode yang berhubungan dengan waktu, mulai dari *time origin* atau *start point* sampai dengan terjadinya suatu kejadian khusus atau *end point*. Salah satu tujuan analisis *survival* adalah mengetahui hubungan antara waktu *survival* dengan variabel-variabel yang diduga mempengaruhi waktu *survival* [3]. Dalam analisis *survival* dikenal tiga model regresi, yaitu model nonparametrik, semiparametrik, dan model parametrik. Regresi semiparametrik adalah regresi yang paling populer diantara metode regresi lainnya [4]. Regresi *cox proportional hazard (PH)* termasuk ke dalam metode semiparametrik yang mana fungsi *hazard* dasarnya mengikuti model nonparametrik sedangkan, variabel-variabel independennya mengikuti model parametrik. Regresi *cox PH* bertujuan untuk mengetahui efek dari beberapa variabel terhadap waktu *survival* secara bersama-sama. Secara umum, model regresi Cox dihadapkan pada situasi dimana kemungkinan kegagalan individu bertahan hidup pada suatu waktu yang dipengaruhi oleh satu atau lebih variabel bebas [5]. Model *accelerated failure time (AFT)* merupakan salah satu model parametrik yang didasarkan pada distribusi data yang dapat memprediksi waktu suatu kejadian yang terjadi pada data observasi [6]. Pada model *AFT* diharuskan untuk melakukan pengecekan asumsi distribusi, salah satunya yaitu distribusi logistik.

Peneliti sebelumnya yang berkaitan dengan model *cox PH* dilakukan oleh [7], untuk menganalisis identifikasi faktor-faktor yang mempengaruhi lama studi mahasiswa. Penelitian yang di dapat menyimpulkan bahwa faktor yang mempengaruhi lama studi mahasiswa adalah jurusan, IPK, dan nilai akreditasi SMA. Sedangkan penelitian sebelumnya dengan menggunakan model *AFT* yang dilakukan oleh [6], pada kasus kardiovaskular. Dari hasil penelitian, variabel yang paling dominan dalam menduga *survival time* yaitu variabel status diabetes melitus dengan waktu perawatan atau lama waktu pasien tersebut bertahan hidup hingga meninggal dunia adalah selama 11 hari, dengan presentase tertinggi yaitu 52,8%.

Penelitian kali ini dilakukan dengan perbandingan model *cox PH* dan model *AFT* berdistribusi logistik untuk mengetahui apakah kondisi atau tipe data lebih baik digunakan pada model *cox PH* atau model *AFT*. Untuk membandingkan kedua model tersebut akan

dipilih nilai *akaike's information criterion (AIC)* terbaik dari masing-masing model untuk menentukan hasil model yang terbaik.

2. Metode

Data data yang digunakan merupakan data kelulusan mahasiswa S1 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sebelas Maret tahun 2019 sampai dengan tahun 2020. Jumlah data yang digunakan dalam penelitian ini yaitu sebanyak 500. Variabel-variabel yang digunakan pada penelitian ini adalah variabel tidak bebas lamanya seseorang memperoleh pekerjaan: Variabel tidak bebas (Y) adalah lamanya seseorang memperoleh pekerjaan; Program studi (X_1); Jenis kelamin (X_2); Asal daerah (X_3); Waktu lulus (X_4); IPK (X_5); Keaktifan Organisasi (X_6).

2.1. Analisis *Survival*

Analisis *survival* merupakan metode statistika yang digunakan untuk mempelajari ketahanan hidup yang berhubungan dengan waktu. Waktu yang dimaksud yaitu mulai dari waktu awal (*time origin*) yang sudah ditentukan sampai pada suatu kejadian khusus (*failure evented point*), sehingga analisis *survival* memerlukan waktu *survival* pada suatu individu. Analisis *survival* bertujuan untuk mengetahui variabel yang mempengaruhi kejadian tertentu [8]. Dalam analisis *survival* terdiri dari 3 fungsi distribusi *survival time* yaitu sebagai berikut.

- a. Fungsi kepadatan peluang $f(t)$ merupakan peluang suatu individu mati atau gagal dalam interval waktu t sampai dengan Δt . Rumus fungsi kepadatan peluang ditunjukkan pada persamaan (1).

$$f(t) = \lim_{\Delta t \rightarrow 0} \left[\frac{F(t + \Delta t) - F(t)}{\Delta t} \right] \quad (1)$$

- b. Fungsi *survival* $S(t)$ merupakan probabilitas suatu individu yang masih dapat bertahan hidup sampai dengan waktu $t(t > 0)$. Rumus fungsi *survival* ditunjukkan pada persamaan (2).

$$S(t) = 1 - P(T < t) \quad (2)$$

- c. Fungsi *hazard* $h(t)$ merupakan peluang jika individu mengalami suatu peristiwa atau kejadian pada interval $(t + \Delta t)$. Rumus fungsi *hazard* ditunjukkan pada persamaan (3).

$$h(t) = \frac{f(t)}{S(t)} \quad (3)$$

2.2. Data Tersensor

Data tersensor merupakan data dimana nilai dari variabel terikat yang diteliti tidak memberikan informasi yang lengkap atau tidak dapat diamati secara penuh. Beberapa alasan terjadinya penyensoran pada data *survival* antara lain adalah subjek mengalami peristiwa di luar masa observasi, subjek menghilang pada masa observasi, subjek mengalami peristiwa lain sedemikian sehingga peristiwa yang menjadi perhatian tidak terjadi, dan subjek mengalami peristiwa yang menjadi perhatian namun disebabkan oleh alasan lain yang bukan menjadi perhatian peneliti [9]. Dalam analisis *survival* terdapat empat tipe penyensoran yaitu sebagai berikut.

- a. Sensor kanan adalah sensor yang terjadi dikarenakan objek pengamatan belum mengalami kejadian hingga akhir periode pengamatan, sedangkan waktu awal dari objek pengamatan dapat diamati secara penuh.

- b. Sensor kiri adalah sensor yang terjadi dikarenakan waktu awal dari subjek pengamatan tidak dapat teramati pada awal pengamatan, sementara kegagalan dapat diamati secara penuh sebelum penelitian berakhir.
- c. Sensor interval adalah sensor sensor yang waktu survival berada dalam suatu selang tertentu.
- d. Sensor acak adalah sensor yang terjadi jika individu yang diamati meninggal atau mengalami kejadian karena sebab yang lain, bukan disebabkan dari tujuan utama penelitian.

2.3. Fungsi Likelihood

Fungsi *likelihood* adalah besarnya peluang ketidakpastian terjadinya resiko, yang diperkirakan berdasarkan data historis frekuensi keseringan dari kejadian yang serupa. Rumus fungsi *likelihood* ditunjukkan pada persamaan (4).

$$L(\theta) = \prod_{i=1}^n \phi(t_i, x_i; \theta)^{y_i} S(t_i, y_i, x_i; \theta) \tag{4}$$

Estimasi dalam penelitian ini menggunakan estimasi parameter *maksimum partial likelihood* (MLE) yaitu dengan cara memaksimalkan fungsi likelihood. Rumus memaksimalkan fungsi *likelihood* ditunjukkan pada persamaan (5).

$$L(\beta) = \prod_{l=1}^e \frac{\exp(\beta^p x_{(l)})}{\sum_{f \in R(t_{(l)})} \exp(\beta^p x_f)} \tag{5}$$

2.4. Kejadian Bersama (Ties) dalam Partial Likelihood

Dalam analisis *survival* sering dijumpai adanya kejadian bersama atau yang sering disebut dengan *ties*. Keadaan ini terdapat dua atau lebih individu yang mengalami kejadian pada waktu yang bersamaan. Beberapa metode untuk menangani *ties*, yaitu metode *partial likelihood Breslow*, *partial likelihood Efron*, dan *partial likelihood Exact*. Dalam penelitian ini akan menggunakan pendekatan metode *partial likelihood Efron*. Penelitian yang dilakukan oleh [10] menguraikan metode dengan *pendekatan partial likelihood Efron* ini merupakan metode yang sedikit lebih intensif pada tingkat komputasinya dibanding dengan metode *Breslow*, akan tetapi metode ini memberikan hasil estimasi yang baik jika data kejadian bersama atau *ties* dalam ukuran besar. Pendekatan *Efron* secara umum memiliki bentuk persamaan seperti pada persamaan (6).

$$L(\beta_{Efron}) = \prod_{i \in D} \frac{\exp(\beta S_k)}{\prod_{i=1}^{d_k} [\sum_{i \in R_{t_i}} \exp(\beta X_i) - \frac{j-1}{d_k} \sum_{i \in D_k} \exp(\beta X_i)]} \tag{6}$$

2.5. Model Cox Proportional Hazard

Regresi *cox PH* model berdistribusi semiparametrik karena dalam model *Cox PH* tidak memerlukan informasi tentang distribusi yang mendasari waktu *survival*. Model regresi *cox PH* mengasumsikan bahwa *hazard ratio (HR)* bernilai konstan yaitu nilai *HR* tidak bergantung pada waktu. Sehingga menunjukkan perbandingan hazard kedua objek tetap sepanjang waktu [11]. Rumus untuk model *cox proportional hazard* ditunjukkan pada persamaan (7).

$$h_0(t, x) = h_0(t) \exp\left(\sum_i^p \beta_i X_i\right) \tag{7}$$

2.6. Pengujian Asumsi *Cox Proportional Hazard*

Asumsi *PH* digunakan untuk mengecek apakah suatu data cocok digunakan dalam model *cox PH*. Terdapat pendekatan yang digunakan untuk pengecekan asumsi *PH* yaitu sebagai berikut.

2.6.1. Pendekatan Grafik

Metode grafik yang digunakan dalam pengecekan asumsi *proportional hazard* yaitu menggunakan plot *log minus log survival* yaitu apabila masing-masing variabel untuk setiap respon menghasilkan grafik dengan bentuk garis sejajar pada setiap kategorinya maka pengujian asumsi *proportional hazard* terpenuhi dan dapat dilanjutkan untuk analisis selanjutnya [12].

2.6.2. Pendekatan *Goodnes of Fit*

Menurut [13] menyatakan bahwa dalam melakukan pengecekan asumsi *proportional hazard* dapat dilakukan dengan menggunakan penaksiran *goodness of fit (GOF)*. Metode ini akan memberikan hasil yang lebih objektif karena menggunakan pengujian secara statistik. Untuk pemeriksaan asumsi menggunakan uji statistik, metode residual yang digunakan yaitu *Schoenfeld residual* dimana residual tersebut terdefinisi untuk setiap individu yang mengalami *event*.

2.7. Model *Accelerated Failure Time*

Menurut [6], asumsi dasar untuk model *AFT* merupakan efek dari kovariat konstan dan mutliplikatif terhadap waktu *survival*, artinya kovariat mempengaruhi *survival* oleh suatu faktor konstan yang disebut *acceleration factor*. Berdasarkan model *AFT* tersebut, maka fungsi *hazard* untuk individu dengan variabel independen *Z* ditunjukkan pada persamaan (8).

$$h(x|Z) = \exp(\theta^t Z) h_0[\exp(\theta^t Z) x] \quad (8)$$

Pada penelitian ini akan menggunakan model *AFT* distribusi logistik dengan karena distribusi ini sering digunakan dan menghasilkan fungsi *hazard* tidak monoton. Fungsi distribusi kumulatif logistik memiliki bentuk tertutup sederhana, yang menjadi penting secara komputasi saat menyesuaikan data dengan penyensoran.

2.8. Pengujian Signifikansi Parameter

Pengujian signifikansi parameter bertujuan untuk memeriksa apakah variabel independen memiliki pengaruh dalam model. Pengujian parameter dalam pada uji serentak dilakukan menggunakan uji *partial likelihood ratio*. Persamaan untuk uji *partial likelihood ratio* ditunjukkan pada persamaan (9).

$$G = -2[\ln L_R - \ln L_f] \quad (9)$$

Uji parsial bertujuan untuk mengetahui variabel independen yang berpengaruh secara nyata terhadap variabel dependen. Uji parsial dilakukan menggunakan uji *wald*. Persamaan untuk uji *wald* ditunjukkan pada persamaan (10).

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (10)$$

2.9. Pemilihan Model Terbaik

Akaike's Information Criterion (AIC) adalah metode yang berguna untuk mendapatkan model terbaik yang ditemukan oleh Akaike. Menurut [14] besarnya AIC ditunjukkan pada persamaan (11).

$$AIC = -2L(\beta) + 2df \quad (11)$$

3. Hasil dan Pembahasan

Data data yang digunakan merupakan data kelulusan mahasiswa S1 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sebelas Maret (FMIPA UNS) tahun 2019 sampai dengan tahun 2020. Jumlah data yang digunakan dalam penelitian ini yaitu sebanyak 500. Variabel-variabel yang digunakan pada penelitian ini adalah variabel tidak bebas lamanya seseorang memperoleh pekerjaan: Variabel tidak bebas (Y) adalah lamanya seseorang memperoleh pekerjaan; Program studi (X_1); Jenis kelamin (X_2); Asal daerah (X_3); Waktu lulus (X_4); IPK (X_5); Keaktifan Organisasi (X_6)

3.1. Distribusi Frekuensi Data

Distribusi frekuensi adalah penyusunan suatu data mulai dari data terkecil hingga terbesar yang membagi banyaknya data kedalam beberapa kelas. Kegunaan data yang diubah dalam bentuk distribusi frekuensi yaitu untuk memudahkan penyajian, mudah dipahami dan mudah dibaca sebagai informasi yang hasilnya disajikan pada [Tabel 1](#).

Tabel 1. Distribusi frekuensi data

| | | Frekuensi | Presentase |
|--------------------|--------------|-----------|------------|
| Informasi Kejadian | Terobservasi | 305 | 61% |
| | Tersensor | 195 | 39% |
| Total | | 500 | 100% |

Dari [Tabel 1](#) dapat dilihat bahwa sebesar 61% atau sebanyak 305 terboservasi atau dengan kata lain seseorang yang telah lulus dapat memperoleh pekerjaannya dengan waktu tunggu kurang dari sama dengan 6 bulan, sedangkan yang belum mendapatkan pekerjaan ada sebesar 39% atau sebanyak 195 orang.

3.2. Eksplorasi Data Kuantitatif

Eksplorasi data kuantitatif bertujuan untuk memahami data sebelum dilakukan analisis. Pemahaman terhadap data yang akan dapat membantu dalam menentukan teknik-teknik analisis data terhadap data sebelum melakukan analisis yang dan hasilnya disajikan pada [Tabel 2](#).

Tabel 2. Eksplorasi data kuantitatif

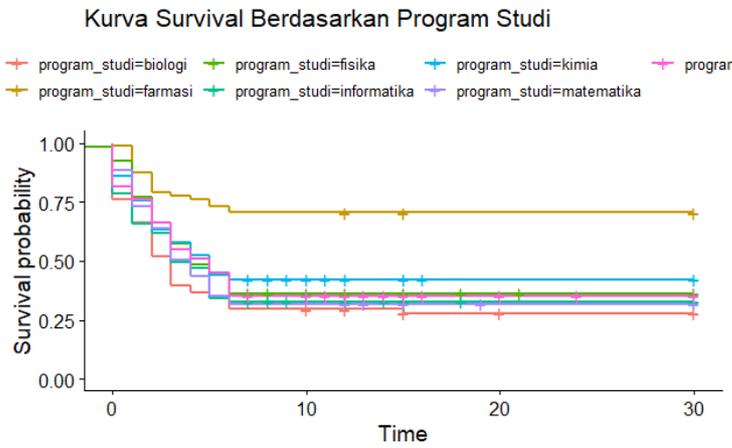
| Variabel | Mean | Median | Min | Max |
|---------------------------------|------|--------|------|------|
| Indeks Prestasi Kumulatif (IPK) | 3,40 | 3,41 | 2,53 | 3,90 |

Variabel independen yang bersifat kumulatif meliputi nilai IPK. Pada [Tabel 2](#) menunjukkan bahwa rata-rata nilai IPK lulusan FMIPA UNS tahun 2019 sampai tahun 2020 adalah sebesar 3,40 dengan IPK tertinggi sebesar 3,90 dan IPK terendah sebesar 2,53.

3.3. Pengujian Asumsi *Proportional Hazard* Pendekatan Grafik

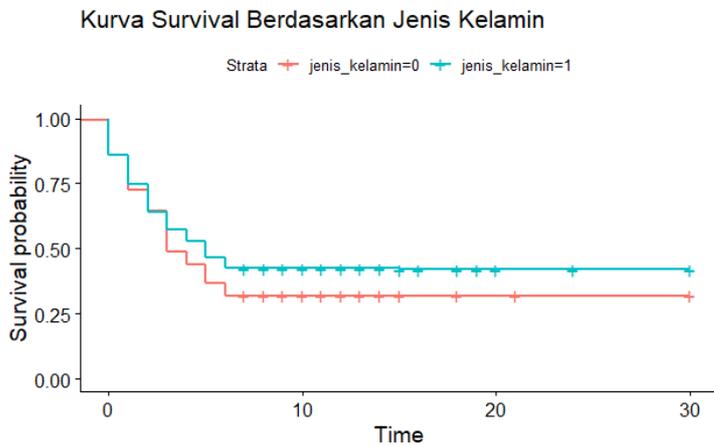
Dalam pengujian asumsi *cox PH* apabila plot antar kategori dalam satu variabel independen terlihat sejajar atau tidak bersilangan maka asumsi *proportional hazard* terpenuhi dan variabel independen yang bersifat kategorik dapat dimasukkan ke dalam model. Berikut merupakan uji asumsi *proportional hazard* untuk masing-masing variabel independen yang bersifat kategorik dalam penelitian ini.

[Gambar 1](#) menunjukkan bahwa variabel program studi (X_1) untuk kurva fungsi *survival* program studi statistika kebanyakan berada di bawah daripada kurva fungsi *survival* program studi lainnya.



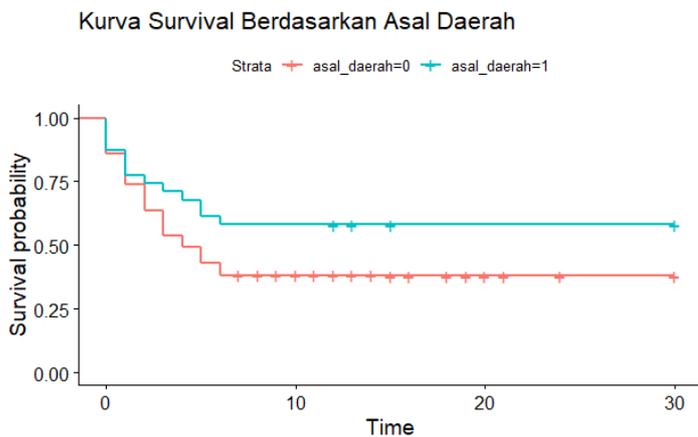
Gambar 1.
Kurva *survival* variabel program studi

Gambar 2 menunjukkan bahwa variabel jenis kelamin (X_2) untuk kurva fungsi *survival* jenis kelamin laki-laki selalu berada di bawah dari kurva jenis kelamin perempuan yang artinya jenis kelamin laki-laki peluang memperoleh pekerjaan lebih cepat dibandingkan dengan yang berjenis kelamin perempuan.



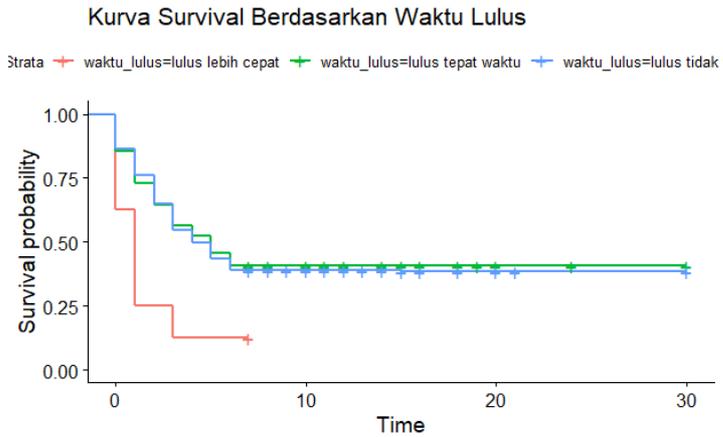
Gambar 2.
Kurva *survival* variabel jenis kelamin

Gambar 3 menunjukkan bahwa variabel asal daerah (X_3) untuk kurva fungsi *survival* lulusan berasal dari pulau Jawa selalu berada di bawah dari kurva dengan lulusan yang berasal dari luar Jawa yang artinya lulusan yang berasal dari pulau Jawa peluang memperoleh pekerjaan lebih cepat dibandingkan dengan lulusan yang berasal dari luar pulau Jawa.



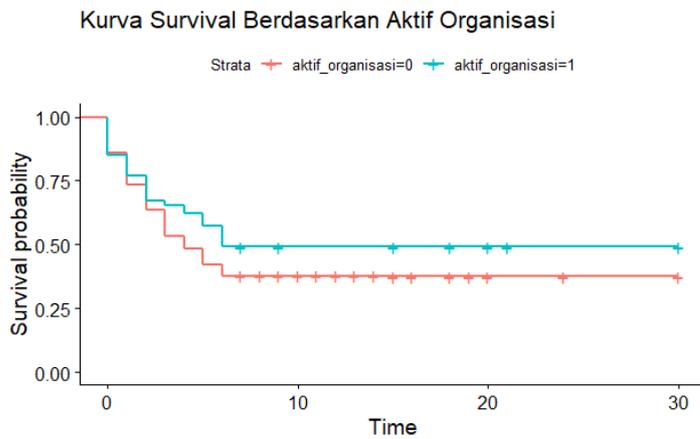
Gambar 3.
Kurva *survival* variabel asal daerah

Gambar 4 menunjukkan bahwa variabel waktu lulus (X_4) untuk kurva fungsi *survival* dengan waktu lulus kurang dari delapan semester selalu di bawah daripada kurva fungsi *survival* waktu lulus tepat waktu dan tidak tepat waktu.



Gambar 4.
Kurva *survival* variabel waktu lulus

Gambar 5 menunjukkan bahwa variabel aktif organisasi (X_6) untuk kurva fungsi *survival* dengan lulusan yang tidak aktif mengikuti organisasi kebanyakan berada di bawah daripada kurva fungsi *survival* lulusan dengan aktif berorganisasi.



Gambar 5.
Kurva *survival* variabel keaktifan organisasi

3.4. Estimasi Parameter *Cox Proportional Hazard*

Estimasi parameter dalam penelitian ini akan menggunakan pendekatan metode *partial likelihood Efron*. Hasil dari estimasi parameter disajikan pada **Tabel 3**.

Tabel 3. Estimasi parameter *cox proportional hazard*

| Variabel | Coef | P-Value | Keputusan |
|--------------------------------------|-------|-------------|-------------------|
| Program studi statistika $X_{1(1)}$ | -0,24 | 0,27 | Gagal Tolak H_0 |
| Program studi matematika $X_{1(2)}$ | 0,00 | 0,97 | Gagal Tolak H_0 |
| Program studi informatika $X_{1(3)}$ | -0,18 | 0,35 | Gagal Tolak H_0 |
| Program studi fisika $X_{1(4)}$ | -0,28 | 0,17 | Gagal Tolak H_0 |
| Program studi kimia $X_{1(5)}$ | -0,30 | 0,18 | Gagal Tolak H_0 |
| Program studi farmasi $X_{1(7)}$ | -1,37 | $2,1e^{-7}$ | Tolak H_0 |
| Jenis kelamin X_2 | -0,26 | 0,04 | Tolak H_0 |
| Asal daerah X_3 | -0,50 | 0,08 | Gagal Tolak H_0 |
| Lulus tepat waktu $X_{4(2)}$ | -0,84 | 0,03 | Tolak H_0 |
| Lulus tidak tepat waktu $X_{4(3)}$ | -0,77 | 0,06 | Gagal Tolak H_0 |
| IPK X_5 | 0,94 | 0,00 | Tolak H_0 |
| Keaktifan organisasi X_6 | -0,43 | 0,02 | Tolak H_0 |

Dengan demikian berdasarkan **Tabel 3** bahwa semua variabel independen kecuali variabel IPK berpengaruh secara signifikan terhadap model, sehingga didapatkan model awal regresi cox proportional hazard yang ditunjukkan pada persamaan (12).

$$h(t, X) = h_0(t) \exp(-0,24X_{1(1)} - 0,0X_{1(2)} - 0,18X_{1(3)} - 0,28X_{1(4)} - 0,30X_{1(5)} - 1,37X_{1(7)} - 0,26X_2 - 0,50X_3 - 0,84X_{4(2)} - 0,77X_{4(3)} + 0,43X_6) \tag{12}$$

3.5. Pengujian Model Terbaik *Cox Proportional Hazard*

Pada penelitian ini menggunakan eliminasi *backward* untuk menentukan pemilihan model terbaik. **Tabel 4** menunjukkan bahwa eliminasi *backward* ini dihentikan apabila semua variabel yang masuk ke dalam model sudah signifikan.

Tabel 4. Pengujian model terbaik *cox proportional hazard*

| Variabel | Coef | P-Value | Keputusan |
|----------------------------------|-------|-------------|-------------|
| Program studi farmasi $X_{1(7)}$ | -1,37 | $2,1e^{-7}$ | Tolak H_0 |
| Jenis kelamin X_2 | -0,26 | 0,04 | Tolak H_0 |
| Lulus tepat waktu $X_{4(2)}$ | -0,84 | 0,03 | Tolak H_0 |
| IPK X_5 | 0,94 | 0,00 | Tolak H_0 |
| Keaktifan organisasi X_6 | -0,43 | 0,02 | Tolak H_0 |

Berdasarkan hasil estimasi pada **Tabel 4** maka didapatkan model terbaik dari regresi *cox PH* yang ditunjukkan pada persamaan (13).

$$h(t, X) = h_0(t) \exp(-1,37X_{1(7)} - 0,26X_2 - 0,84X_{4(2)} + 0,90X_5 - 0,43X_6) \tag{13}$$

Pengujian serentak digunakan untuk mengetahui apakah variabel independen berpengaruh terhadap variabel dependen. Didapatkan hasil perhitungan uji G yang disajikan pada **Tabel 5**.

Tabel 5. Uji serentak *cox proportional hazard*

| Uji Overall | $G \geq X^2_{(0,05;5)}$ | $P - Value < \alpha$ | Keputusan |
|-------------|-------------------------|----------------------|-------------|
| | $59,90 \geq 11,07$ | $0,00 < 0,05$ | Tolak H_0 |

Berdasarkan hasil pengujian pada **Tabel 5** diperoleh nilai $G \geq X^2_{(0,05;5)}$ yaitu $59,90 \geq 11,07$ atau $P - Value < \alpha$ yaitu $0,00 < 0,05$ didapatkan keputusan bahwa H_0 ditolak yang artinya layak untuk digunakan.

3.6. Pengujian Asumsi *Cox Proportional Hazard* dengan *Goodness of Fit*

Pengujian asumsi *PH* juga dapat dilakukan dengan menggunakan metode *Goodness of Fit (GOF)*. **Tabel 6** menunjukkan statistik uji yang digunakan dalam metode ini adalah dengan melihat nilai *residual Schoenfeld*.

Tabel 6. Pengujian asumsi *cox proportional hazard* dengan *GOF*

| Variabel | chisq | P - Value | α | Keputusan |
|----------------------------------|-------|-----------|----------|---------------------------|
| Program studi farmasi $X_{1(7)}$ | 4,71 | 0,58 | > | Gagal Tolak H_0 |
| Jenis kelamin X_2 | 2,39 | 0,12 | > | Gagal Tolak H_0 |
| Lulus tepat waktu $X_{4(2)}$ | 1,03 | 0,59 | > | 0,05 Gagal Tolak H_0 |
| IPK X_5 | 6,81 | 0,00 | < | Tolak H_0 |
| Keaktifan organisasi X_6 | 1,05 | 0,30 | > | Gagal Tolak H_0 |

Berdasarkan **Tabel 6** dapat dilihat bahwa variabel independen yang signifikan terhadap model memiliki *P-Value* yang lebih besar daripada tingkat signifikansi ($\alpha = 0,05$) sehingga diperoleh keputusan H_0 gagal ditolak. Variabel independen yang signifikan terhadap model

yaitu variabel program studi farmasi ($X_{1(7)}$), jenis kelamin (X_2), waktu lulus tepat waktu ($X_{4(2)}$), dan aktif organisasi (X_6). Maka dapat disimpulkan bahwa asumsi *PH* untuk variabel independen yang signifikan dalam model terpenuhi.

3.7. Estimasi Parameter Model Accelerated Failure Time

Tabel 7 merupakan hasil perhitungan uji parsial untuk model *AFT*.

Tabel 7. Estimasi parameter accelerated failure time

| Variabel | Z | P-Value | Keputusan |
|--------------------------------------|-------|--------------|-------------------|
| Program studi statistika $X_{1(1)}$ | 0,33 | 0,73 | Gagal Tolak H_0 |
| Program studi matematika $X_{1(2)}$ | -0,29 | 0,77 | Gagal Tolak H_0 |
| Program studi informatika $X_{1(3)}$ | 0,82 | 0,41 | Gagal Tolak H_0 |
| Program studi fisika $X_{1(4)}$ | 1,02 | 0,30 | Gagal Tolak H_0 |
| Program studi kimia $X_{1(5)}$ | 0,85 | 0,77 | Gagal Tolak H_0 |
| Program studi farmasi $X_{1(7)}$ | 6,17 | $6,8e^{-10}$ | Tolak H_0 |
| Jenis kelamin X_2 | 1,82 | 0,69 | Gagal Tolak H_0 |
| Asal daerah X_3 | 1,78 | 0,76 | Gagal Tolak H_0 |
| Lulus tepat waktu $X_{4(2)}$ | 1,10 | 0,26 | Gagal Tolak H_0 |
| Lulus tidak tepat waktu $X_{4(3)}$ | 0,67 | 0,50 | Gagal Tolak H_0 |
| IPK X_5 | -2,32 | 0,02 | Tolak H_0 |
| Keaktifan organisasi X_6 | 2,13 | 0,03 | Tolak H_0 |

Dengan demikian berdasarkan Tabel 7 bahwa semua variabel independen kecuali variabel IPK berpengaruh secara signifikan terhadap model, sehingga didapatkan model accelerated failure time pada persamaan (14).

$$h(t, X) = h_0(t) \exp(0,81X_{1(1)} + 1,88X_{1(2)} - 0,75X_{1(3)} + 2,49X_{1(4)} + 2,22X_{1(5)} + 17,76X_{1(7)} + 2,74X_2 + 5,81X_3 + 5,18X_{4(2)} + 3,25X_{4(3)} - 9,57X_5 + 4,74X_6) \tag{14}$$

3.8. Pengujian Model Terbaik Accelerated Failure Time

Hasil estimasi parameter dari pemodelan terbaik model *AFT* dengan menggunakan metode eliminasi backward yang disajikan pada Tabel 8 di bawah ini.

Tabel 8. Pengujian model terbaik accelerated failure time

| Variabel | Coef | P-Value | Keputusan |
|----------------------------------|-------|--------------|-------------|
| Program studi farmasi $X_{1(7)}$ | 18,15 | $2,6e^{-10}$ | Tolak H_0 |
| Jenis kelamin X_2 | 2,94 | 0,04 | Tolak H_0 |
| IPK X_5 | -0,81 | 0,01 | Tolak H_0 |
| Keaktifan organisasi X_6 | 4,62 | 0,03 | Tolak H_0 |

Berdasarkan hasil estimasi pada Tabel 8 maka didapatkan model terbaik dari model *AFT* yang ditunjukkan pada persamaan (15).

$$h(t, X) = h_0(t) \exp(18,15X_{1(7)} + 2,94X_2 - 0,81X_5 + 4,62X_6) \tag{15}$$

Pengujian serentak digunakan untuk mengetahui apakah variabel independen berpengaruh terhadap variabel dependen. Didapatkan hasil perhitungan uji G yang disajikan pada Tabel 9.

Tabel 9. Uji serentak accelerated failure time

| Uji Overall | $G \geq X^2_{(0,05,4)}$ | $P - Value < \alpha$ | Keputusan |
|-------------|-------------------------|----------------------|-------------|
| | $68,68 \geq 9,48$ | $0,00 < 0,05$ | Tolak H_0 |

Berdasarkan hasil pengujian pada **Tabel 9** diperoleh nilai $G \geq X_{(0,05;5)}^2$ yaitu $68,68 \geq 11,07$ atau $P - Value < \alpha$ yaitu $0,00 < 0,05$ didapatkan keputusan bahwa H_0 ditolak yang artinya layak untuk digunakan.

3.9. Pemilihan Model Terbaik

Dalam menentukan pemilihan model terbaik adalah dengan melihat nilai *akaike's information criterion* (*AIC*) terkecil. **Tabel 10** merupakan hasil nilai *AIC* dari model *Cox PH* dan *AFT*.

Tabel 10. Pemilihan model terbaik

| Model | Nilai AIC |
|---------------------------------|-----------|
| <i>Cox Proportional Hazard</i> | 3512,71 |
| <i>Accelerated Failure Time</i> | 2823,14 |

Hasil menunjukkan bahwa model *AFT* lebih kecil dibandingkan model *Cox PH* yaitu sebesar 2823,14 sehingga model terbaik untuk waktu tunggu seseorang memperoleh pekerjaan setelah lulus adalah model *AFT*.

4. Kesimpulan

Model *AFT* dalam menduga *survival time* merupakan model yang lebih baik dari model *cox PH*, karena model *AFT* menghasilkan nilai *AIC* yang lebih kecil dibandingkan dari model *cox PH*. Faktor-faktor yang mempengaruhi waktu tunggu seseorang memperoleh pekerjaan setelah lulus dengan menggunakan metode *AFT* ada variabel program studi, jenis kelamin, IPK, dan Keaktifan organisasi.

Referensi

- [1] BPS, "Tingkat Pengangguran Terbuka (TPT) 2022," 9 Mei 2022. [Online]. Available: <https://www.bps.go.id/>.
- [2] H. Dukalang, Analisis Regresi COX Proportional Hazard pada Pemodelan Waktu Tunggu Mendapatkan Pekerjaan, Gorontalo: ejurnal.ung.ac.id, 2019.
- [3] T. Qomaria, Analisis Ketahanan Hidup Pasien Stroke Menggunakan Model COX Proportional Hazard, Jember, 2019.
- [4] Riangkaryaman, Model Semiparametrik Cox PH dan Parametrik (Weibull dan Lognormal) dalam Analisis Survival (Kasus: Rawat Inap Penderita DBD di RSUD Haji Kota Makassar), Makassar: Skripsi. Makassar: FMIPA Universitas Negeri Makassar, 2017.
- [5] I. A. Dewi, N. L. P. Suciptawati and N. K. T. Tastrawati, Aplikasi Regresi Cox Proportional Hazard Pada Sintasan Pasien Diabetes Melitus, E-Jurnal Matematika Vol. 7(3), 2018.
- [6] S. F. Fatmala, Model Accelerated Failure Time Untuk Menduga Survival Time Pada Kasus Kardiovaskular, Jember: Digital Repository Universitas Jember, 2019.
- [7] I. D. Fitriani, Analisis Cox Proportional Hazard Pada Identifikasi Faktor-Faktor Yang Mempengaruhi Lama Studi Mahasiswa S1 FMIPA Universitas Islam Indonesia, Yogyakarta: Universitas Islam Indonesia, 2018.
- [8] M. R. Pahlevi, Mustafid and T. Wuryandari, Model Regresi Cox Stratified Pada Data Ketahanan, Semarang: Jurnal Gaussian, Volume 3, Nomor 3, 2016.
- [9] A. Faruk, Aplikasi Regresi COx Pada Selang Kelahiran Anak Pertama di Provinsi Sumatera Selatan, Palembang: Jurnal Matematika Vol.7 No.1, Universitas Sriwijawa, 2017.
- [10] M. N. Susenati, Analisis Lama Waktu Mencari Kerja dengan Pendekatan Regresi Cox Proportional Hazard, Yogyakarta: Skripsi. Yogyakarta: FMIPA Universitas Islam Indonesia, 2015.

- [11] M. P. Paramitha, N. Satyahadewi and N. N. Debataraaja, MOdel Extended Cox Untuk Mengatasi Nonproportional Hazard, Buletin Ilmiah Mat. dan terapannya(Bimaster), 2022.
- [12] N. Nurmala, M. Fatekurrohman and D. Anggraeni, Comparison of Exact, Efron, and Breslow Parameter Approach Method on Hazard Ratio and Stratified Cox Regression Model, Journal of Physics Conference Series. 1008(7): 1-7, 2018.
- [13] Nurfain and S. W. Purnami, "Analisis Regresi Cox Extended Pada Pasien Kusta di Kecamatan Brondong Kabupaten Lamongan," *Jurnal sains dan Seni ITS*, Vols. 6(1): 2337-3520, 2017.
- [14] M. D. Stasinopoulos, R. A. Rigby and F. Bastiani, A distributional regression approach, London Metropolitan University, London, United Kingdom, 2018.



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)
