

Analisis Regresi Cox Proportional Hazards pada Kasus Pasien COVID-19

Ichwanul Kahfi Prasetya, Pradhikta Naufal Irfanda, dan Jerry Dwi Trijoyo Purnomo

Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

ichwankahfi@gmail.com, pradhikta.18062@mhs.its.ac.id, jerry@statistika.its.ac.id

Abstrak. Peningkatan jumlah kasus COVID-19 di berbagai negara-negara di dunia terus mengalami peningkatan. Tingginya jumlah penderita COVID-19 disebabkan karena ditemukan banyak virus yang bermutasi di dunia dengan berbagai varians, juga kurangnya kesadaran untuk menjaga protokol kesehatan, tidak menerapkan gaya hidup sehat, serta kurang berolahraga. Analisis statistika yang sering digunakan pada bidang kesehatan untuk mengetahui ketahanan hidup seseorang adalah analisis survival. Analisis survival merupakan analisis mengenai data yang diperoleh dari catatan waktu yang di-capai suatu obyek sampai terjadinya peristiwa khusus (failure event). Salah satu metode yang sering digunakan dalam analisis survival nonparametrik adalah analisis Kaplan Meier yang dilanjutkan dengan uji Log Rank. Penelitian ini menggunakan analisis Kaplan Meier dan Uji Log Rank untuk mengetahui perbedaan kurva survival pada setiap kategori variabel *age*, *sex*, *ECOG*, dan *Karnofsky*. Kemudian dilanjutkan dengan analisis Regresi Cox Proportional Hazards. Sehingga didapatkan kesimpulan bahwa hasil kurva KM dan Uji Log-Rank yaitu variabel Sex dan ECOG berpengaruh signifikan terhadap probabilitas survival pasien.

Keywords: Analisis Survival, COVID-19, Regresi Cox Proportional Hazards

1 Pendahuluan

1.1 Latar Belakang

Jumlah kasus COVID-19 masih terus mengalami peningkatan di berbagai negara-negara di dunia. Setiap negara juga memiliki kebijakan tersendiri untuk menahan penyebaran virus yang terjadi di wilayahnya. Dari kebijakan-kebijakan tersebut tentu ada yang tepat dan tidak. Ketepatan kebijakan tersebut menjadi salah satu faktor laju pertumbuhan kasus COVID-19 di negara tersebut. Menurut data yang dikumpulkan oleh WHO per tanggal 25 Juni 2021, total jumlah kasus terkonfirmasi positif COVID-19 di seluruh dunia telah mencapai 179.686.071 kasus dengan 3.899.172 kematian. Negara dengan jumlah kasus terbanyak antara lain Amerika (33.257.768 kasus), India (30.134.445 kasus), Brazil (18.169.881 kasus) dan seterusnya [1].

Tingginya jumlah penderita COVID-19 disebabkan karena ditemukan banyak virus yang bermutasi di dunia dengan berbagai varians. Selain itu karena kurangnya

kesadaran untuk menjaga protokol kesehatan, tidak menerapkan gaya hidup sehat, dan kurang berolahraga.

Analisis statistika yang sering digunakan pada bidang kesehatan untuk mengetahui ketahanan hidup seseorang adalah analisis survival. Analisis survival merupakan analisis mengenai data yang diperoleh dari catatan waktu yang dicapai suatu obyek sampai terjadinya peristiwa khusus (*failure event*). Salah satu metode yang sering digunakan dalam analisis survival nonparametrik adalah analisis Kaplan Meier yang dilanjutkan dengan uji Log Rank. Analisis Kaplan Meier digunakan untuk mengestimasi fungsi survival. Kemudian dari estimasi fungsi survival dapat dibentuk kurva survival Kaplan Meier. Sedangkan uji Log Rank digunakan untuk menguji apakah terdapat perbedaan atau tidak dalam kurva survival Kaplan Meier pada variabel-variabel yang mempunyai kategori dua atau lebih [5].

Berdasarkan pemaparan diatas, maka penelitian ini menggunakan analisis Kaplan Meier dan Uji Log Rank untuk mengetahui perbedaan kurva survival pada setiap kategori variabel *age*, *sex*, *ECOG*, dan *Karnofsky*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah dapat disusun rumusan masalah sebagai berikut.

1. Bagaimana statistika deskriptif yang dapat disajikan dari data variabel?
2. Bagaimana hasil yang diperoleh dari Kurva Kaplan Meier dan Uji Log-Rank?
3. Bagaimana hasil yang diperoleh dari analisis Regresi Cox PH?

1.3 Tujuan

1. Memperoleh hasil statistika deskriptif pada data variabel.
2. Memperoleh hasil dari *Kurva Kaplan Meier dan Uji Log-Rank*.
3. Memperoleh hasil dari analisis Regresi Cox PH.

1.4 Manfaat

Manfaat dari penelitian ini adalah memperoleh informasi untuk variabel yang berpengaruh pada probabilitas survival. Selain itu dapat memberikan informasi kepada pembaca terkait topik yang sedang diteliti.

1.5 Batasan masalah

Karena banyaknya data yang tersedia dalam sumber data utama yaitu kaggle.com dan terbatasnya waktu dalam melakukan penelitian, batasan masalah yang digunakan dalam penelitian ini adalah data yang digunakan merupakan data yang tidak mengandung *missing value*.

2 Landasan Teori

2.1 Statistika Deskriptif

Statistika deskriptif merupakan metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga dapat menghasilkan informasi yang berguna dengan menyusun tabel, grafik, diagram, dan besaran-besaran lain di majalah dan di koran. [2]

Rata-rata hitung (*mean*) adalah nilai rata-rata dari data-data yang tersedia. Rata-rata hitung dari populasi diberi simbol μ . Rata-rata hitung dari sampel diberi simbol \bar{x} [3]. Rata-rata hitung (*mean*) untuk data tunggal.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Di mana,

\bar{x} : rata-rata hitung (*mean*)

x_i : nilai data ke- i di mana $i=1,2,\dots,n$

n : jumlah sampel

Standar deviasi dan varians merupakan satu diantara teknik statistik yang digunakan untuk menjelaskan homogenitas kelompok. Varians merupakan jumlah kuadrat semua deviasi nilai-nilai individual terhadap rata-rata kelompok. Sedangkan akar dari varians disebut dengan standar deviasi atau simpangan baku[4].

Standar deviasi dan varians simpangan baku merupakan variasi sebaran data. Semakin kecil nilai sebarannya berarti variasi nilai data makin sama. Jika sebarannya bernilai 0, maka nilai semua datanya adalah sama. Semakin besar nilai sebarannya berarti data semakin bervariasi. Varians adalah salah satu ukuran dispersi atau ukuran variasi. Varians diberi simbol σ^2 untuk populasi dan untuk s^2 sampel, sedangkan standar deviasi diberi simbol σ untuk populasi dan s untuk sampel. Rumus untuk mencari varian adalah sebagai berikut.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (2)$$

Standar deviasi merupakan akar dari s yaitu sama seperti rumus di atas hanya saja dioperasikan pengakaran.

Keterangan :

\bar{x} : rata-rata hitung (*mean*)

x_i : nilai data ke- i di mana $i=1, 2, \dots, n$

n : jumlah sampel

Minimum adalah suatu nilai fungsi objek, yang menghasilkan nilai terendah pada daerah himpunan penyelesaian. Maksimum adalah suatu nilai fungsi objek, yang menghasilkan nilai tertinggi pada daerah himpunan penyelesaian[5].

2.2 Analisis Survival

Menurut Kleinbaum dan Klein, analisis survival merupakan kumpulan metode statistika yang digunakan untuk menganalisis data di mana *outcome* variabel yang

diteliti adalah waktu sampai suatu kejadian (*event*) muncul [6]. Tiga hal yang harus diperhatikan dalam menentukan waktu kegagalan menurut Cox dan Oakes dalam Inayati dan Purnami [7], adalah:

1. Waktu awal (*starting point*) pada penelitian tidak ambigu.
2. Skala pengukuran sebagai bagian dari waktu (*measurement scale for the passage of time*) yang konsisten.
3. Kejelasan definisi kejadian akhir (*ending event of interest*) pada penelitian.

2.3 Data Tersensor

Penyensoran adalah salah satu langkah yang harus dilakukan untuk mengatasi ketidaklengkapan suatu data pengamatan. Penyensoran terjadi apabila kita mempunyai informasi tentang waktu survival individu, tetapi tidak diketahui secara pasti waktu survivalnya, maka data tersebut termasuk data tersensor. Menurut Kleinbaum dan Klein terdapat tiga penyebab terjadinya data tersensor [6], yaitu:

1. Individu tidak mengalami kejadian yang diteliti sampai akhir pengamatan,
2. Individu yang hilang dari pengamatan selama masa penelitian,
3. Individu mengundurkan diri dari penelitian karena kematian (jika kematian bukan kejadian yang diteliti) atau alasan lainnya.

2.4 Fungsi Survival dan Fungsi Hazard

Menurut Kleinbaum dan Klein pada analisis survival terdapat dua macam fungsi utama yaitu fungsi survival ($S(t)$) dan fungsi hazard ($h(t)$) [7]. Fungsi survival $S(t)$ adalah probabilitas suatu objek bertahan setelah waktu ke- t , dinyatakan sebagai berikut:

$$S(t) = P(T > t) \quad t \geq 0 \quad (3)$$

Fungsi hazard $h(t)$ merupakan probabilitas suatu individu gagal pada interval waktu t . Dengan demikian fungsi hazard dapat diartikan sebagai kebalikan dari fungsi survival.

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

Sehingga hubungan antara fungsi survival dan fungsi hazard dapat dinyatakan sebagai berikut:

$$h(t) = \frac{f(t)}{S(t)} \quad (5)$$

2.5 Kurva Survival Kaplan Meier dan Uji Log Rank

Menurut Kleinbaum dan Klein analisis Kaplan Meier digunakan untuk menaksir fungsi survival. Berikut merupakan persamaan umum dari fungsi survival yang digunakan untuk membentuk kurva survival Kaplan Meier [7].

$$\hat{S}(t(j)) = \hat{S}(t(j-1)) \times P\hat{r}(T > t(j) | T \geq t(j)) \quad (6)$$

dimana,

$$\hat{S}(t(j-1)) = \prod_{i=1}^{j-1} P\hat{r}(T > t(j) | T \geq t(j)) \quad (7)$$

Hasil kurva survival Kaplan Meier yang terbentuk, kemudian dibandingkan apakah terapat perbedaan antar kurva survival menggunakan uji Log Rank. Uji Log Rank merupakan uji yang digunakan untuk membandingkan kurva survival dalam grup yang berbeda. Dengan hipotesis untuk uji Log Rank sebagai berikut:

H_0 : tidak ada perbedaan antar kurva survival

H_1 : paling sedikit ada satu perbedaan antar kurva survival

Statistik uji yang digunakan dalam uji Log Rank terbagi menjadi uji Log Rank dua grup dan uji Log Rank lebih dari dua grup. Statistik uji untuk uji Log Rank dua grup adalah sebagai berikut:

Log Rank statistics =

$$\frac{(O_i - E_i)^2}{Var(O_i - E_i)} \quad (8)$$

Statistik uji untuk uji Log Rank lebih dari dua grup adalah sebagai berikut.

Log Rank statistics =

$$\mathbf{d'V - 1d} \quad (9)$$

atau dengan rumus pendekatan Log Rank statistics

$$\frac{(O_i - E_i)^2}{Var(O_i - E_i)} \quad (10)$$

$$X^2 \approx \sum_i^G \frac{(O_i - E_i)^2}{E_i} \quad (11)$$

Hipotesis H_0 akan ditolak, jika nilai p-value kurang dari α atau Log Rank statistics $\approx \chi^2_{hitung}$ hitung lebih besar dari $\chi^2_{\alpha, df}$ dengan derajat bebas sama dengan $G-1$.

2.6 Regresi Cox Proportional Hazards

Menurut Collett, regresi Cox Proportional Hazards atau lebih dikenal sebagai model regresi Cox digunakan untuk mengetahui hubungan antara variabel dependen dengan variabel independen, dimana data yang digunakan pada regresi Cox Proportional Hazards berupa data waktu tahan hidup dari suatu individu [8]. Menurut Collet, model regresi Cox Proportional Hazards adalah sebagai berikut [8]

$$h_i(t|X) = h_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) = h_0(t) e^{\sum_{j=1}^p \beta_j x_{ji}} \quad (12)$$

dengan

$h_i(t|X)$: Fungsi kegagalan individu ke-i

$h_0(t)$: Fungsi kegagalan dasar

x_{ji} : Nilai variabel ke-j dari individu ke-i, dengan $j=1,2,\dots,p$ dan $i=1,2,\dots,n$

β_j : Koefisien regresi ke-j, dengan $j=1,2,\dots,p$

Fungsi kegagalan dalam model ini dapat mengambil bentuk apapun, tetapi fungsi kegagalan dari individu yang berbeda diasumsikan harus proporsional setiap waktu.

Menurut Kleinbaum dan Klein, apabila plot $\log\{-\log[S(t,x)]\}$ terhadap waktu ketahanan hidup antar kategori dalam satu variabel independen terlihat sejajar atau berhimpit maka asumsi proportional hazards terpenuhi dan variabel penjelas dapat dimasukkan ke dalam model. [6]

3 Metodologi Penelitian

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder. Data tersebut diperoleh dari website Kaggle. Data tersebut berjudul “*Lung-survival Dataset*” yang memiliki 7 variabel dan 228 observasi.

3.2 Variabel penelitian

Variabel penelitian yang digunakan pada analisis ini terdiri dari 7 variabel dan 338 observasi dengan deskripsi variabel penelitian sebagai berikut.

Tabel 1. Variabel Penelitian

Variabel	Jenis Data	Keterangan
Inst	Kategorik	Kode Fasilitas Kesehatan
Time	Numerik	Waktu Survival pasien
Status	Kategorik	Status Data, 1=Tersensor, 2=Tidak Tersensor
Age	Numerik	Usia Pasien dalam tahun
Sex	Kategorik	Jenis Kelamin, 1=Laki-laki, 2=Perempuan
ecog	Numerik	ECOG Score (0-5), 0=Sangat Baik, 5=Meninggal
karno	Numerik	Karnofsky Performance Score

3.3 Struktur Data

Struktur data yang digunakan pada penelitian ini adalah sebagai berikut.

Tabel 2. Struktur Data ‘*Phishing Detection Dataset*’

Obs	Variabel			
	X_1	X_2	...	X_m
1	$X_{1,1}$	$X_{1,2}$...	$X_{1,m}$
2	$X_{2,1}$	$X_{2,2}$...	$X_{2,m}$
3	$X_{3,1}$	$X_{3,2}$...	$X_{3,m}$
...
n	$X_{n,1}$	$X_{n,2}$...	$X_{n,m}$

3.4 Langkah Analisis

Langkah analisis yang dilakukan dalam praktikum ini adalah sebagai berikut.

1. Melakukan identifikasi terhadap permasalahan dan tujuan penelitian.
2. Mengambil data sekunder dilihat dari beberapa karakteristik yang sudah ditentukan.
3. Melakukan analisis statistika deskriptif

4. Membentuk Kurva Kaplan Meier
5. Melakukan Uji-Log-Rank
6. Melakukan Analisis Regresi Cox Proportional Hazard
7. Menarik kesimpulan dan saran.

4 Analisis dan Pembahasan

4.1 Karakteristik Data

Langkah awal sebelum melakukan analisis adalah perlu mengetahui karakteristik dari variabel yang digunakan dalam analisis. Berikut merupakan statistika deskriptif variabel numerik.

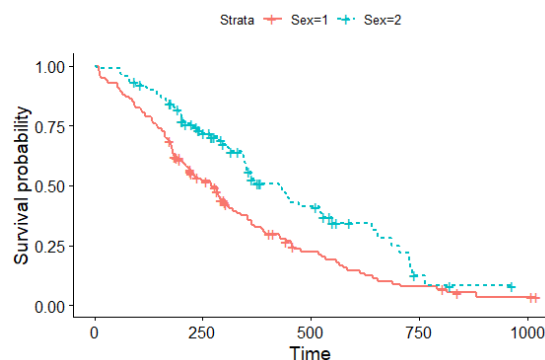
Tabel 3 Karakteristik Data Numerik

	Time	Age	Karno
Minimum	5.00	39.00	50.00
Median	255.50	63.00	80.00
Maximum	1022.00	82.00	100.00
Mean	305.20	62.45	81.97

Berdasarkan Tabel 2 didapatkan informasi bahwa pada variabel Time memiliki nilai [ada rentang 5 hingga 1022 dengan median 255,5 dan rata-rata 305,2. Kemudian variabel Age memiliki pada rentang 39 hingga 82 dengan median 63 dan rata-rata 62,45. Sedangkan variabel Karno memiliki pada rentang 50 hingga 100 dengan median 80 dan rata-rata 81,97

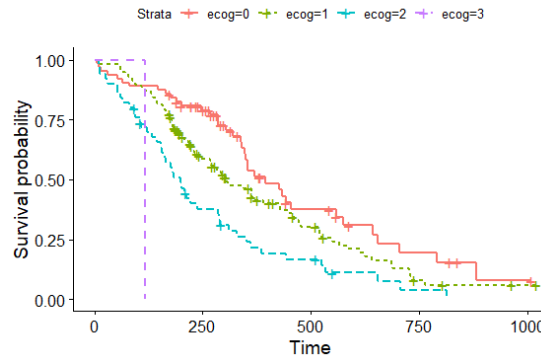
4.2 Kurva Kaplan Meier dan Uji Log-Rank

Kurva Kaplan Meier digunakan untuk melihat apakah terdapat perbedaan waktu survival dari variabel kategorik secara deskriptif. Kemudian dibuktikan dengan uji *Log-Rank* apakah terdapat perbedaan yang signifikan dari kategori tersebut. Berikut adalah hasil dari kurva Kaplan Meier dari variabel kategorik penelitian.



Gambar 1 Kurva Kaplan Meier Variabel Sex

Pada Gambar 1 secara visual dapat dilihat bahwa terdapat perbedaan pola kurva Kaplan Meier yang terbentuk. Pasien dengan kategori Sex=1 (laki-laki) cenderung memiliki probabilitas survival pasien dengan kategori Sex-2 (perempuan). Namun perlu dilakukan analisis lebih lanjut untuk mengetahui kebenarannya menggunakan uji log-rank.



Gambar 2 Kurva Kaplan Meier Variabel ECOG

Berdasarkan Gambar 2 menunjukkan bahwa probabilitas survival pasien cenderung berbeda terutama pada rentang waktu lebih dari 100 hari. Semakin kecil nilai ECOG maka probabilitas survival pasien cenderung lebih tinggi, sebaliknya semakin besar nilai ECOG maka probabilitas survival pasien cenderung lebih rendah. Namun pada rentang waktu 0 hingga 100 hari, probabilitas survival pasien sedikit berbeda dengan waktu setelahnya. Sehingga untuk mendapatkan kesimpulan yang lebih pasti, maka perlu dilakukan analisis lebih lanjut untuk mengetahui kebenarannya menggunakan uji log-rank.

Hipotesis

H0 : Kurva survival dari masing masing kategori sama.

H1 : Kurva survival dari masing masing kategori (minimal ada 1 yang) berbeda

Taraf Signifikansi = 0,05

Statistik Uji

Tabel 4 Statistik Uji dari Log Rank Test

Variabel	Chis-Square	df	P-value
Sex	10,3	1	0,001
ECOG	21,6	3	0,000

Berdasarkan Tabel 4 didapatkan masing-masing nilai p-value kurang dari taraf signifikansi (0,05) sehingga keputusan yang diambil adalah Tolak H0. Berdasarkan keputusan tersebut maka didapatkan kesimpulan bahwa kedua kurva survival berbeda, yang berarti masing-masing variabel kategori berpengaruh signifikan terhadap waktu survival pasien.

4.3 Regresi Cox PH

Untuk mengetahui pengaruh variabel penelitian terhadap waktu survival pasien, dapat dilakukan dengan menggunakan regresi Cox PH. Berikut adalah analisis regresi cox ph untuk data waktu survival pasien transpalansi.

1. Estimasi Parameter Regresi Cox PH

Dari hasil perhitungan pada software R, didapatkan parameter sebagai berikut.

Tabel 5 Estimasi Parameter Model Regresi Cox PH

Parameter	Coef	Exp(Coef)	SE(Coef)
ECOG	0,632	1,881	0,173
Sex	-0,585	0,557	0,169
Age	0,012	1,012	0,009
Karno	0,012	1,012	0,009

Didapatkan estimasi parameter yang ditunjukkan pada Tabel 5 dapat dilakukan uji serentak dan uji parsial terhadap hasil Regresi Stratified Cox tersebut

2. Uji Signifikansi Parameter

Terdapat 2 macam uji signifikansi parameter yang dilakukan pada model yaitu uji secara serentak kemudian uji secara parsial. Untuk pengujian secara serentak adalah sebagai berikut.

Hipotesis :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{minimal ada satu } \beta_p \neq 0$$

Taraf Signifikansi = 0,05

Statistik Uji :

Tabel 6. Uji Serentak Model Regresi Cox PH

Likelihood Ratio Test	Derajat Bebas	P-Value
32,45	4	0,000

Berdasarkan Tabel 6 didapatkan nilai p-value yang kurang dari 0,05 sehingga keputusannya yaitu menolak H_0 yang artinya terdapat minimal 1 parameter yang berpengaruh signifikan terhadap waktu survival. Sehingga terdapat treatment yang memberikan pengaruh atau perbedaan terhadap waktu survival pasien. Selanjutnya dilakukan pengujian secara parsial

Hipotesis

$$H_0 : \beta_i = 0 ; i = 1,2,3,4 \text{ (Parameter tidak berpengaruh terhadap } y \text{)}$$

$$H_1 : \beta_i \neq 0 ; i = 1,2,3,4 \text{ (Parameter berpengaruh signifikan terhadap } y \text{)}$$

Taraf Signifikansi = 0,05

Statistik Uji :

Tabel 7. Statistik Uji Signifikansi Parsial

Parameter	P-Value
ECOG	0,000
Sex	0,000
Age	0,185
Karno	0,199

Berdasarkan Tabel 7 terdapat dua variabel yaitu variabel ECOG dan Sex yang memiliki nilai kurang dari taraf signifikansi (0,05). Sedangkan dua variabel yang lain nilainya

lebih dari 0,05. Hal tersebut menandakan bahwa hanya dua variabel yang keputusannya menolak H_0 yaitu signifikans terhadap model. Sedangkan dua variabel yaitu Age dan Karno gagal menolak H_0 yang artinya variabel tidak berpengaruh signifikans.

3. Model Terbaik dan Odd Ratio

Dari hasil eliminasi model secara forward didapatkan model sebagai berikut.

$$\hat{h}(t, X) = \hat{h}_0(t) \exp(0,4858 \text{ ECOG} - 0,5601 \text{ Sex})$$

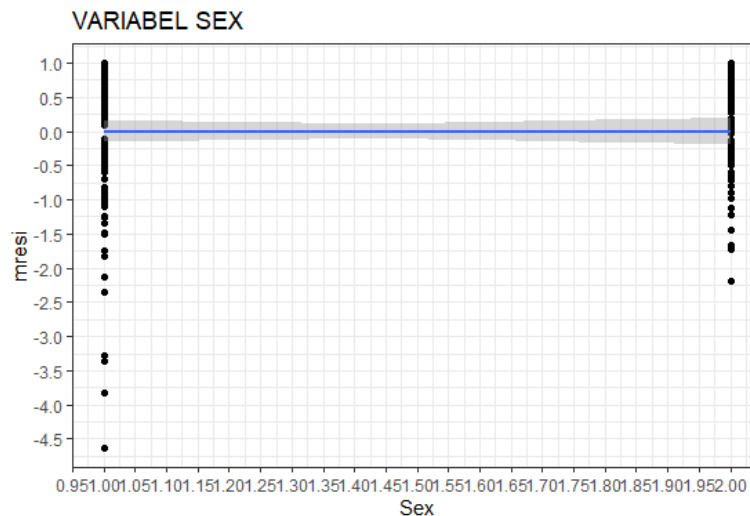
Tabel 8. Parameter Model Terbaik

	coef	exp(coef)	p-value
ECOG	0.4858	1,6254	0,000
Sex	-0,5601	0,5711	0,000

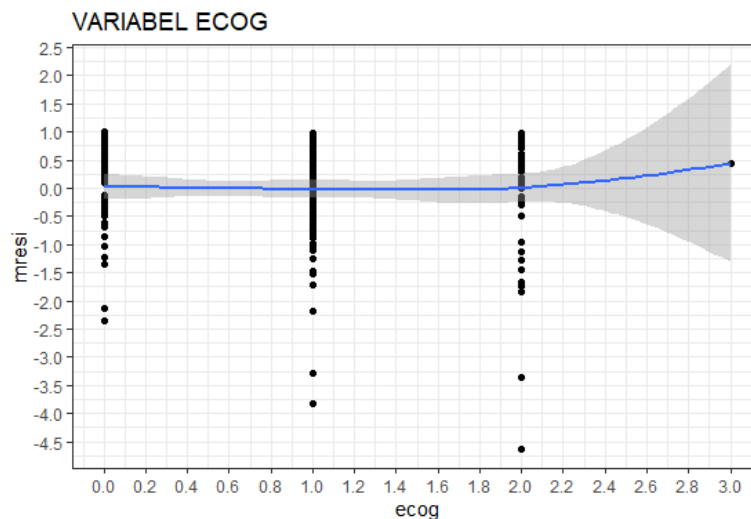
Berdasarkan Tabel 8 menunjukkan bahwa nilai p-value dari variabel yang parameter telah signifikans. Untuk mengetahui pengaruhnya dapat dilihat dari variabel odd ratio atau $\exp(\text{coef})$. Untuk variabel yang signifikans ECOG memiliki odd ratio 1,6254 yang artinya semakin besar nilai variabel ECOG maka resiko terjadi event semakin tinggi. Sedangkan untuk variabel Sex memiliki odd ratio 0.5711 yang artinya kategori dengan angka tinggi (2=laki-laki) memiliki resiko terjadi event lebih rendah dibandingkan kategori dengan angka rendah (1=perempuan)..

4. Uji Asumsi

Setelah didapatkan model terbaik maka akan diperiksa dari asumsinya apakah telah memenuhi atau belum yaitu terdapat uji asumsi linear, Uji Schoenfeld, dan Uji Grambsch-Therneau. Berikut dimulai dari hasil uji asumsi linear

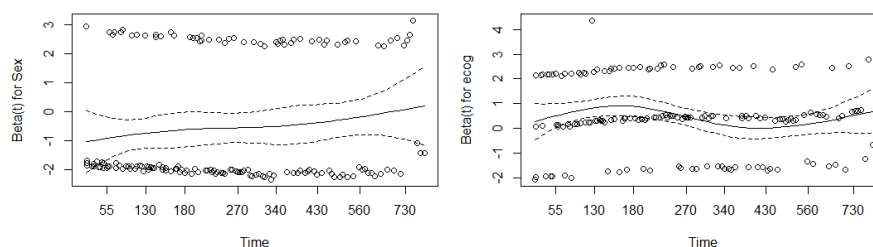


Gambar 3. Martine Residual untuk Variabel Sex



Gambar 4. Martinle Residual untuk Variabel ECOG

Berdasarkan Gambar 3 dan gambar 4, menunjukkan uji asumsi linieritas untuk variabel Sex dan ECOG menghasilkan garis lurus yang berada pada rentang -0,5 dan 0,5. Hal tersebut mengartikan asumsi kovariat yang bersifat linier telah terpenuhi. Selain itu, pengujian asumsi lainnya adalah residual Schoenfeld pada Gambar 5 dan Uji Grambsch-Therneau pada Tabel 7.



Gambar 5. Residual Schoenfeld Regresi Cox PH

Dari Gambar 5, menunjukkan bahwa residual-residual dari variabel Sex dan variabel ECOG cenderung banyak yang keluar dari garis *convidence interval*. Maka kesimpulan dari hasil *shoenfeld residuals plot* diperoleh hasil bahwa variabel Sex dan ECOG tidak memenuhi asumsi PH. Berikutnya ditampilkan hasil Uji Grambsch-Therneau untuk menguatkan uji asumsinya.

Tabel 9. Uji Grambsch-Therneau Regresi Cox PH

	chisq	df	p-value
ECOG	2,04	1	0,153
Sex	2,58	1	0,108
GLOBAL	4,64	2	0,098

Berdasarkan Tabel 9 didapatkan hasil uji Grambsch-Therneau yang menunjukkan parameter ECOG, Sex, dan GLOBE memiliki nilai p-value lebih dari 0,05 yang menandakan parameter telah konstan terhadap waktu dan memenuhi asumsi PH.

5 Penutup

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan maka didapatkan kesimpulan sebagai berikut.

1. Rata-rata variabel Time, Age, dan Karno berturut-turut adalah 305,2; 62,45; dan 81,97.
2. Hasil Kurva KM dan Uji Log-Rank yaitu variabel Sex dan ECOG berpengaruh signifikan terhadap probabilitas survival pasien
3. Model Regresi Cox Proportional Hazard yang terbentuk pada penelitian adalah $\hat{h}(t, X) = \hat{h}_0(t) \exp(0,4858 \text{ ECOG} - 0,5601 \text{ Sex})$

5.2 Saran

Saran Peneliti kepada penelitian berikutnya adalah untuk melakukan penelitian dengan menggunakan lebih banyak variabel dan metode yang lebih beragam supaya mendapatkan model yang lebih baik.

Daftar Pustaka

1. WHO. (2021). Retrieved from WHO: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
2. R. E. Walpole, Pengantar Statistika Edisi ke-3, Jakarta: Gramedia Pustaka Utama, 1995.
3. M. Muchson, Statistika Deskriptif, Bogor: Guepedia, 2012.
4. H. I, Pokok-Pokok Materi Statistik 1 (Statistika Deskriptif), Jakarta: Bumi Aksara, 1999.
5. L. X. Wei and Q. Y. Fang, "A Data Preprocessing Algorithm for Classification Model Base On Rough Sets," *International Conference on Solid State Devices and Materials Science*, pp. 2025-2029, 2012.
6. Kleinbaum, D. G., dan Klein, M. 2005. *Survival Analysis A Self-Learning Text* 2nd Edition. London: Springer.
7. Inayati, K. D., dan Purnami, S. W. 2015. Analisis Survival Nonparametrik Pada Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya Menggunakan Metode Kaplan Meier dan Uji Log Rank. *Jurnal Sains dan Seni ITS* Vol. 4, No.2: Hal. 199-204.
8. Collet, D. 2004. *Modelling Survival Data in Medical Research*. CRC Press.