

Final Project
Mata Kuliah Model Survival (A)
Kelompok C

No.	Nama Lengkap	NPM	Kontribusi	Tingkat Kontribusi
1.	Aghna Syahputra	2206051481	Aktif diskusi dalam kelompok, mengerjakan bagian konstruksi likelihood.	100%
2.	Dinia Feodora Kusumadewi	2206051411	Aktif diskusi dalam kelompok, mengerjakan bagian konstruksi likelihood.	100%
3.	Kayla Zahira Amadya	2206053890	Aktif diskusi dalam kelompok, mengerjakan bagian Cox Ph, melakukan EDA.	100%
4.	Khalila Izzatunnisa	2206051544	Aktif diskusi dalam kelompok, mengerjakan bagian konstruksi likelihood, melakukan EDA.	100%
5.	Najwa Putri Faradila	2206051355	Aktif diskusi dalam kelompok, mencari dataset, mengerjakan bagian Cox Ph.	100%

Pendahuluan

A. Latar Belakang

Hutan mempunyai peran yang penting dalam pengendalian perubahan iklim dan pemeliharaan biodiversitas. Pemahaman terkait faktor-faktor yang mempengaruhi kelangsungan hidup pohon dalam kondisi yang beragam menjadi acuan *project* kami ini. Untuk mengidentifikasi faktor-faktor tersebut, analisis survival digunakan untuk memprediksi waktu kelangsungan hidup pohon berdasarkan variabel-variabel lingkungan dan biologis yang berbeda.

Dataset “Tree Survival” yang kami gunakan mencakup informasi dari 2783 pohon dengan 24 variabel yang berbeda. Variabel-variabel yang dipilih, yakni “Species”, “Ligh_ISF”, dan “Sterile” dipilih berdasarkan hipotesis awal yang berkaitan dengan kelangsungan hidupnya.

Penggunaan model Survival Cox Proportional Hazards (Cox PH) dalam *project* ini memungkinkan penilaian simultan dari faktor-faktor secara statistik, tanpa memerlukan asumsi distribusi dari waktu kelangsungan hidup yang ditetapkan sebelumnya, sehingga model Cox PH sangat berguna untuk menganalisis data yang kompleks sehingga memudahkan kami untuk melihat pengaruh relatif yang signifikan secara statistik dari tiap variabel terhadap risiko kematian pohon.

B. Tujuan

Berikut ini adalah tujuan dari *project* ini.

- Mengidentifikasi dan menganalisis faktor-faktor lingkungan dan biologis yang paling berpengaruh terhadap kelangsungan hidup pohon, seperti jenis spesies, tingkat cahaya (Light_ISF), dan kondisi sterilisasi tanah (Sterile).
- Memahami signifikansi relatif dari setiap variabel dalam memprediksi kelangsungan hidup pohon.
- Memperkirakan fungsi kelangsungan hidup pohon dengan menggunakan model Cox PH, yang memungkinkan pengestimasian tanpa perlu asumsi tentang distribusi waktu hingga kejadian.

C. Metodologi

Berikut ini adalah metodologi yang kami lakukan untuk *project* ini.

1. Melakukan pencarian data, menyeleksi variabel, dan pengolahan data berupa menghilangkan outliers.
2. Melakukan analisis deskriptif dalam bentuk statistika deskriptif dan visualisasi data
3. Membangun model Survival Cox PH dan memvalidasi modelnya.
4. Melakukan analisis statistik lanjutan berupa uji signifikansi dan interpretasi koefisien.
5. Menyimpulkan hasil dari model dan penilaian variabel berdasarkan analisis data yang dilakukan.

D. Dataset

Pada *final project* kali ini, kami akan melakukan analisis terhadap kelangsungan hidup bibit di mana sifat fungsional bibit pohon yang merujuk pada karakteristik fisik, kimia, atau biologis dari bibit pohon mengatur respon kelangsungan hidup. Data yang kami gunakan bersumber dari Kaggle dengan judul [“Tree Survival Prediction”](#) dengan ukuran 2783 baris x 24 kolom,

	No	Plot	Subplot	Species	Light_ISF	Light_Cat	Core	Soil	Adult	Sterile	...	AMF	EMF	Phenolics	Lignin	NSC	Census	Time	Event	Harvest	Alive
0	126	1	C	Acer saccharum	0.106	Med	2017	Prunus serotina	I	Non-Sterile	...	22.00	NaN	-0.56	13.86	12.15	4	14.0	1.0	NaN	NaN
1	11	1	C	Quercus alba	0.106	Med	2017	Quercus rubra	970	Non-Sterile	...	15.82	31.07	5.19	20.52	19.29	33	115.5	0.0	NaN	X
2	12	1	C	Quercus rubra	0.106	Med	2017	Prunus serotina	J	Non-Sterile	...	24.45	28.19	3.36	24.74	15.01	18	63.0	1.0	NaN	NaN
3	2823	7	D	Acer saccharum	0.080	Med	2016	Prunus serotina	J	Non-Sterile	...	22.23	NaN	-0.71	14.29	12.36	4	14.0	1.0	NaN	NaN
4	5679	14	A	Acer saccharum	0.060	Low	2017	Prunus serotina	689	Non-Sterile	...	21.15	NaN	-0.58	10.85	11.20	4	14.0	1.0	NaN	NaN
...
2778	7165	17	B	Prunus serotina	0.111	Med	2017	Populus grandidentata	891	Non-Sterile	...	40.89	NaN	0.83	9.15	11.88	16	56.0	1.0	NaN	NaN
2779	7217	17	D	Quercus alba	0.118	Med	2017	Acer rubrum	1468	Non-Sterile	...	15.47	32.82	4.88	19.01	23.50	16	56.0	1.0	NaN	NaN
2780	7306	17	D	Quercus alba	0.118	Med	2017	Quercus rubra	1454	Non-Sterile	...	11.96	37.67	5.51	21.13	19.10	16	56.0	1.0	NaN	NaN
2781	7771	18	D	Quercus alba	0.161	High	2017	Sterile	1297	Sterile	...	16.99	22.51	4.28	19.38	21.36	33	115.5	NaN	NaN	NaN
2782	7401	18	A	Prunus serotina	0.141	High	2016	Populus grandidentata	118	Non-Sterile	...	60.46	NaN	1.00	9.04	11.82	16	56.0	1.0	NaN	NaN

2783 rows x 24 columns

dengan tipe data,

```

No          int64
Plot        int64
Subplot     object
Species     object
Light_ISF   float64
Light_Cat   object
Core        int64
Soil        object
Adult       object
Sterile     object
Conspecific object
Myco        object
SoilMyco    object
PlantDate   object
AMF         float64
EMF         float64
Phenolics   float64
Lignin      float64
NSC         float64
Census      int64
Time        float64
Event       float64
Harvest     object
Alive       object
dtype: object

```

Penjelasan mengenai kolom/atribut yang tertera di atas. Kolom/atribut yang akan dipilih untuk visualisasi dan analisis data akan ditandai dengan *highlight* hijau:

No	Kolom/Atribut	Penjelasan
1	No	Nomor identifikasi unik untuk setiap bibit.
2	Plot	Nomor plot lapangan tempat bibit ditanam (1-18).
3	Subplot	Subplot dalam plot utama tempat bibit ditanam (dibagi menjadi 5 subplot, A-E).
4	Species	Jenis spesies bibit pohon (Acer saccharum, Prunus serotina, Quercus alba, dan Quercus rubra).

5	Light_ISF	Tingkat cahaya yang diukur dengan perangkat lunak.
6	Light_Cat	Kategori tingkat cahaya yang dibuat dengan membagi rentang nilai 'Light_ISF' (low, med, and high).
7	Core	Tahun pengambilan inti tanah dari lapangan.
8	Soil	Spesies dari mana inti tanah diambil.
9	Adult	Pohon dari mana tanah diambil.
10	Sterile	Apakah tanah disterilkan atau tidak.
11	Conspecific	Apakah tanah berasal dari conspecific, heterospecific, atau sterilized specific.
12	Myco	Jenis mikoriza dari spesies bibit (AMF atau EMF).
13	SoilMyco	Jenis mikoriza dari spesies yang membudidayakan tanah (AMF atau EMF).
14	PlantDate	Tanggal bibit ditanam di pot lapangan.
15	AMF	Persentase kolonisasi fungsi mikoriza arbuskular pada akar halus bibit yang dipanen.
16	EMF	Persentase kolonisasi fungi mikoriza ektomikoriza pada ujung akar bibit yang dipanen.
17	Phenolics	Dihitung sebagai nmol ekuivalen asam galat per mg ekstrak kering.
18	NSC	Dihitung sebagai persentase massa kering karbohidrat non-struktural.
19	Lignin	Dihitung sebagai persentase massa kering lignin.
20	Census	Nomor sensus saat bibit mati atau dipanen.
21	Time	Jumlah hari saat bibit mati atau dipanen.
22	Event	Status kelangsungan hidup bibit pada waktu tertentu (0 = dipanen atau eksperimen berakhir atau 1 = mati).

23	Harvest	Menunjukkan apakah bibit dipanen untuk pengukuran sifat.
24	Alive	Menunjukkan apakah bibit masih hidup pada akhir musim tanam kedua.

Berikut ini adalah alasan dari memilih masing-masing variabel.

- Spesies (Jenis Spesies)
Pemilihan variabel ini karena memiliki potensi dalam perbedaan daya tahan terhadap penyakit, adaptasi terhadap kondisi lingkungan, dan kecepatan pertumbuhan.
- Light_ISF (Intensitas Cahaya)
Pemilihan variabel ini karena cahaya penting dalam mempengaruhi fotosintesis dan pertumbuhan pohon sehingga berpengaruh langsung pada kelangsungan hidup pohon.
- Sterile (Kondisi Sterilisasi Tanah)
Pemilihan variabel ini karena sterilisasi tanah dapat mempengaruhi ketersediaan nutrisi dan keberadaan patogen dalam tanah yang keduanya dapat mempengaruhi kelangsungan hidup pohon.

I. Pre-processing Data

A. Pre-processing Data

- Memilih variabel yang akan digunakan untuk visualisasi dan analisis data
Dari 24 kolom/atribut pada data asli, kami hanya akan memilih 5 kolom sebagai variabel yang akan digunakan untuk visualisasi dan analisis data yaitu 'Species', 'Light_ISF', 'Sterile', 'Time', dan 'Event'. Maka dari itu, kami akan membentuk *dataframe* baru yang hanya berisikan 5 variabel tersebut.

```
# Create dataframe of only selected variables
selected_columns = ['Species', 'Light_ISF', 'Sterile', 'Time', 'Event']
data = data[selected_columns]
data
```

	Species	Light_ISF	Sterile	Time	Event
0	Acer saccharum	0.106	Non-Sterile	14.0	1.0
1	Quercus alba	0.106	Non-Sterile	115.5	0.0
2	Quercus rubra	0.106	Non-Sterile	63.0	1.0
3	Acer saccharum	0.080	Non-Sterile	14.0	1.0
4	Acer saccharum	0.060	Non-Sterile	14.0	1.0
...
2778	Prunus serotina	0.111	Non-Sterile	56.0	1.0
2779	Quercus alba	0.118	Non-Sterile	56.0	1.0
2780	Quercus alba	0.118	Non-Sterile	56.0	1.0
2781	Quercus alba	0.161	Sterile	115.5	NaN
2782	Prunus serotina	0.141	Non-Sterile	56.0	1.0

2783 rows × 5 columns

2. Mengecek *missing values*

Akan dicek apakah terdapat *missing value* dari data,

```
# Check missing values
data.isnull().sum()

Species      0
Light_ISF    0
Sterile       0
Time          0
Event        1
dtype: int64
```

Terdapat satu *missing value* pada kolom 'Event'.

3. Menangani *missing value*

Kolom 'Event' merupakan kolom yang cukup krusial, sehingga nilainya harus ada untuk mengetahui apakah pada observasi tersebut terjadi *event* atau tidak. Maka dari itu, baris observasi dengan *missing value* pada kolom 'Event' akan dihapus.

```
# Remove row with missing value
data = data.dropna()
data
```

	Species	Light_ISF	Sterile	Time	Event
0	Acer saccharum	0.106	Non-Sterile	14.0	1.0
1	Quercus alba	0.106	Non-Sterile	115.5	0.0
2	Quercus rubra	0.106	Non-Sterile	63.0	1.0
3	Acer saccharum	0.080	Non-Sterile	14.0	1.0
4	Acer saccharum	0.060	Non-Sterile	14.0	1.0
...
2777	Quercus alba	0.122	Non-Sterile	56.0	1.0
2778	Prunus serotina	0.111	Non-Sterile	56.0	1.0
2779	Quercus alba	0.118	Non-Sterile	56.0	1.0
2780	Quercus alba	0.118	Non-Sterile	56.0	1.0
2782	Prunus serotina	0.141	Non-Sterile	56.0	1.0

2782 rows x 5 columns

4. Menangani *outliers*

Akan dilakukan penghapusan baris data untuk baris data pada kolom 'Light_ISF' yang memiliki *outlier*. Teknik penghapusan *outliers* yang digunakan adalah teknik *interquartile range*. Teknik ini akan dilakukan beberapa kali hingga baris data tidak berkurang setiap teknik dijalankan yang artinya data sudah bersih dari *outliers*.

```
# Remove outliers by Interquartile Range
Q1_Light_ISF = data['Light_ISF'].quantile(0.25)
Q3_Light_ISF = data['Light_ISF'].quantile(0.75)
IQR_Light_ISF = Q3_Light_ISF - Q1_Light_ISF

lower_bound_Light_ISF = Q1_Light_ISF - 1.5 * IQR_Light_ISF
upper_bound_Light_ISF = Q3_Light_ISF + 1.5 * IQR_Light_ISF

data = data[
    (data['Light_ISF'] >= lower_bound_Light_ISF) & (data['Light_ISF'] <= upper_bound_Light_ISF)
]

print(data)
```

	Species	Light_ISF	Sterile	Time	Event
0	Acer saccharum	0.106	Non-Sterile	14.0	1.0
1	Quercus alba	0.106	Non-Sterile	115.5	0.0
2	Quercus rubra	0.106	Non-Sterile	63.0	1.0
3	Acer saccharum	0.080	Non-Sterile	14.0	1.0
4	Acer saccharum	0.060	Non-Sterile	14.0	1.0
...
2777	Quercus alba	0.122	Non-Sterile	56.0	1.0
2778	Prunus serotina	0.111	Non-Sterile	56.0	1.0
2779	Quercus alba	0.118	Non-Sterile	56.0	1.0
2780	Quercus alba	0.118	Non-Sterile	56.0	1.0
2782	Prunus serotina	0.141	Non-Sterile	56.0	1.0

[2739 rows x 5 columns]

5. Menyimpan hasil *preprocessing*

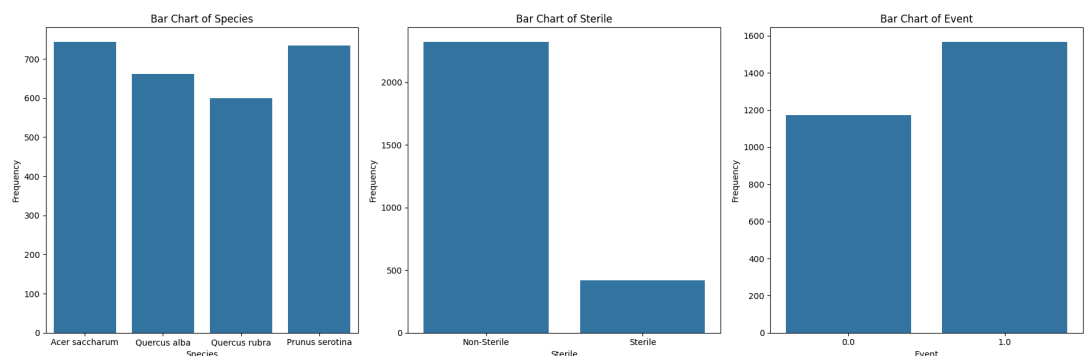
Data yang sudah melalui tahap *preprocessing* akan disimpan dalam file dengan format.csv,

```
# Save preprocessing result
data.to_csv('Tree_Data0.csv', index=False)
```

B. Statistika Deskriptif dan Visualisasi Data

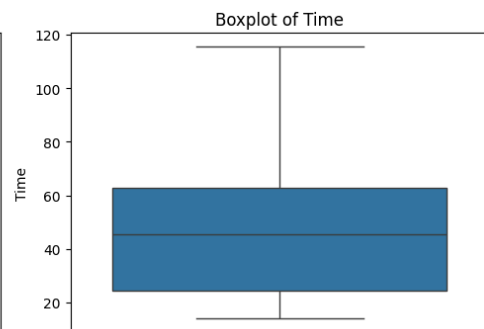
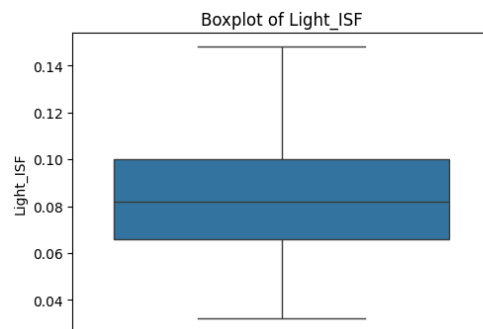
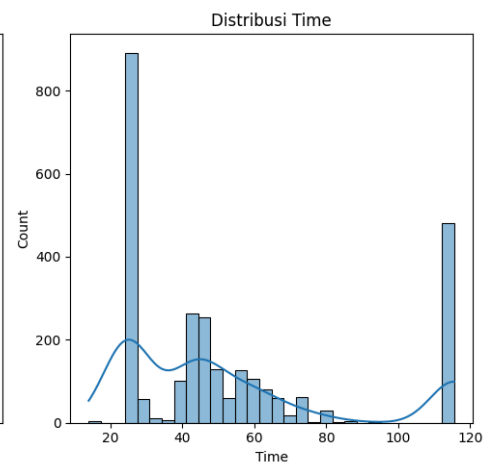
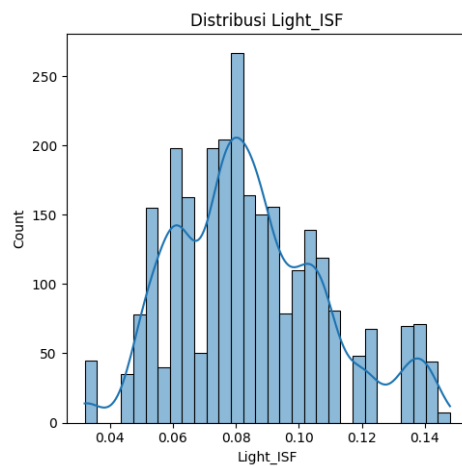
1. Variabel Kategorik

```
(Species
Acer saccharum    743
Prunus serotina   735
Quercus alba      661
Quercus rubra     600
Name: count, dtype: int64,
Sterile
Non-Sterile    2321
Sterile        418
Name: count, dtype: int64,
Event
1.0    1566
0.0    1173
Name: count, dtype: int64)
```



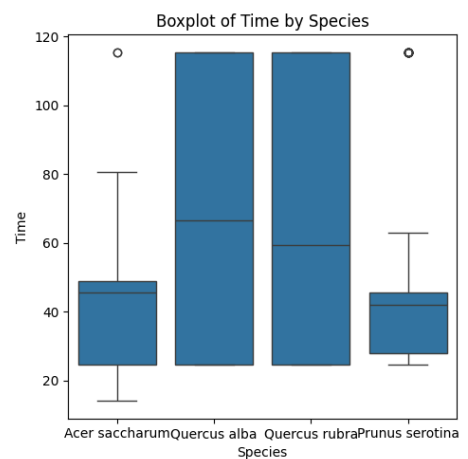
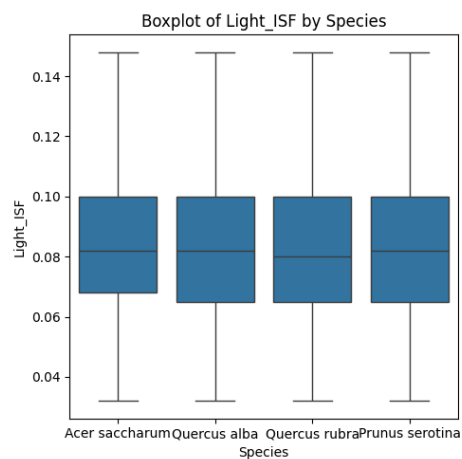
2. Variabel Numerik

	Light_ISF	Time
count	2739.000000	2739.000000
mean	0.084497	53.418766
std	0.023985	31.992995
min	0.032000	14.000000
25%	0.066000	24.500000
50%	0.082000	45.500000
75%	0.100000	63.000000
max	0.148000	115.500000

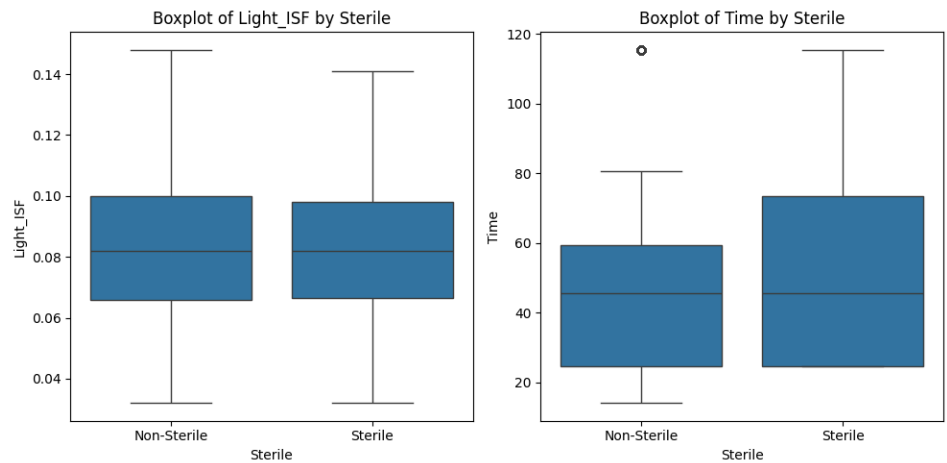


3. Variabel Numerik Berdasarkan Variabel Kategorik

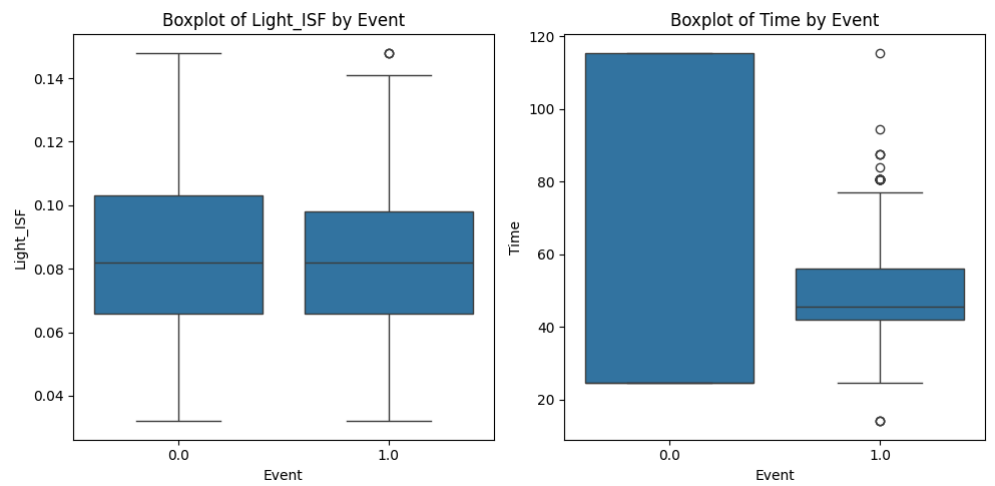
- Berdasarkan 'Species'



- Berdasarkan 'Sterile'



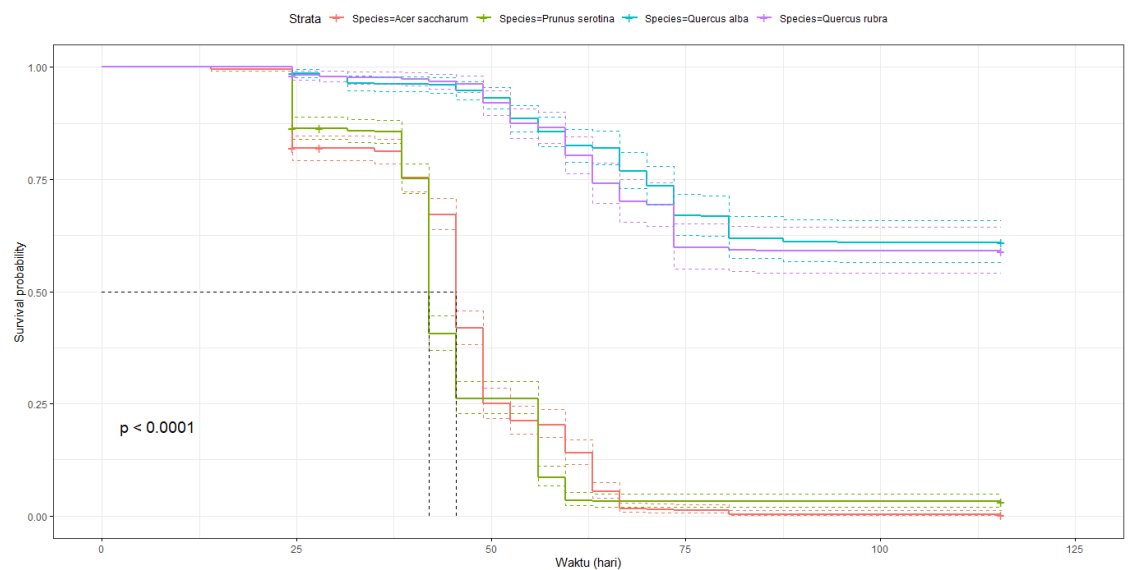
- Berdasarkan 'Event'



II. Analisis Data

A. Analisis Survival

- Species
- Plot Survival:



Terlihat bahwa, $p - value < 0.0001$ yang menunjukkan terdapat perbedaan dalam probabilitas ketahanan hidup antara keempat spesies secara signifikan.

Akan dilakukan pengujian menggunakan Uji Log Rank untuk menguji apakah secara statistik keempat spesies memiliki perbedaan ketahanan hidup.

H_0 : tidak terdapat perbedaan kondisi ketahanan hidup dari keempat spesies

H_1 : terdapat perbedaan kondisi ketahanan hidup dari keempat spesies

```
> uji_logrank=survdiff(Surv(Time,Event)~Species, data=data)
> uji_logrank
Call:
survdiff(formula = Surv(Time, Event) ~ Species, data = data)
```

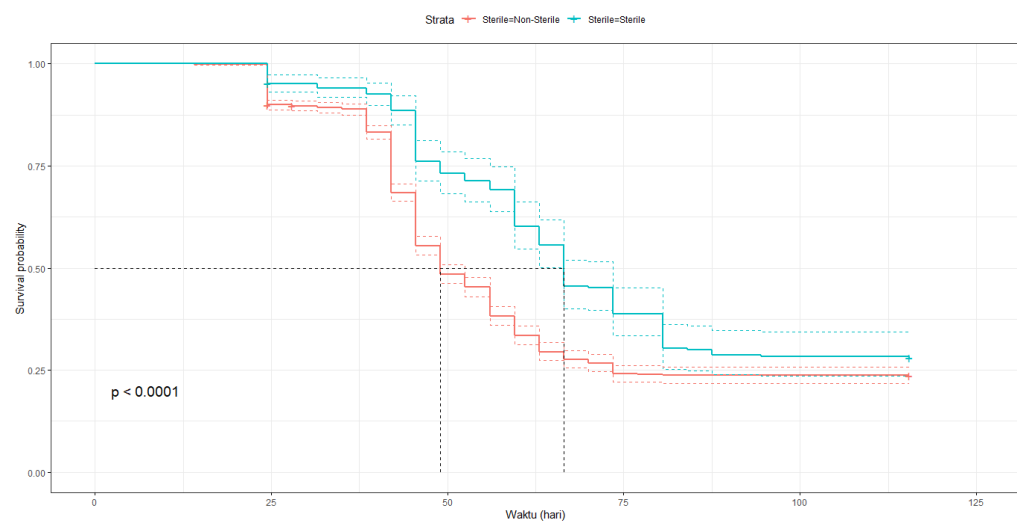
	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
Species=Acer saccharum	743	649	342	277	425
Species=Prunus serotina	735	603	282	366	538
Species=Quercus alba	661	165	510	233	411
Species=Quercus rubra	600	149	433	186	299

Chisq= 1387 on 3 degrees of freedom, p= <2e-16

Berdasarkan hasil Uji Log Rank, diperoleh nilai $p = <2e-16 < 0,05$, maka H_0 ditolak atau terdapat perbedaan kondisi ketahanan hidup antara keempat spesies.

- Sterile

Plot Survival:



Terlihat bahwa, $p - value < 0.0001$ yang menunjukkan terdapat perbedaan dalam probabilitas ketahanan hidup antara tanah yang disterilkan dan tidak disterilkan secara signifikan.

Akan dilakukan pengujian menggunakan Uji Log Rank untuk menguji apakah secara statistik tanah yang disterilkan dan tidak disterilkan memiliki perbedaan ketahanan hidup.

H_0 : tidak terdapat perbedaan kondisi ketahanan hidup dari tanah yang disterilkan dan tidak disterilkan

H_1 : terdapat perbedaan kondisi ketahanan hidup dari tanah yang disterilkan dan tidak disterilkan

```
> uji_logrank=survdiff(Surv(Time,Event)~Sterile, data=data)
> uji_logrank
Call:
survdiff(formula = Surv(Time, Event) ~ Sterile, data = data)

              N Observed Expected (O-E)^2/E (O-E)^2/V
Sterile=Non-Sterile 2321      1366      1291      4.34      28.2
Sterile=Sterile      418       200       275     20.39      28.2

      Chisq= 28.2 on 1 degrees of freedom, p= 1e-07
```

Berdasarkan hasil Uji Log Rank, diperoleh nilai $p = 1e-07 < 0,05$, maka H_0 ditolak atau terdapat perbedaan kondisi ketahanan hidup antara tanah yang disterilkan dan tidak disterilkan.

B. Regresi Cox-PH

- Model1:

```
> model1 <- coxph(Surv(Time, Event) ~ Light_ISF + as.factor(Species) + as.factor(Sterile),
data = data)
> summary(model1)
Call:
coxph(formula = Surv(Time, Event) ~ Light_ISF + as.factor(Species) +
as.factor(Sterile), data = data)

n= 2739, number of events= 1566

              coef exp(coef) se(coef)      z Pr(>|z|)
Light_ISF      -1.82634   0.16100  1.14673  -1.593   0.1112
as.factor(Species)Prunus serotina  0.13665   1.14643  0.05756   2.374   0.0176 *
as.factor(Species)Quercus alba    -2.37612   0.09291  0.09451 -25.142 < 2e-16 ***
as.factor(Species)Quercus rubra   -2.27875   0.10241  0.09725 -23.431 < 2e-16 ***
as.factor(Sterile)Sterile         -0.55997   0.57123  0.07757  -7.219 5.23e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
Light_ISF           0.16100      6.2111   0.01701   1.5238
as.factor(Species)Prunus serotina  1.14643   0.8723   1.02412   1.2834
as.factor(Species)Quercus alba     0.09291  10.7631   0.07720   0.1118
as.factor(Species)Quercus rubra     0.10241   9.7645   0.08464   0.1239
as.factor(Sterile)Sterile           0.57123   1.7506   0.49066   0.6650

Concordance= 0.758 (se = 0.006 )
Likelihood ratio test= 1524 on 5 df,  p=<2e-16
Wald test              = 1145 on 5 df,  p=<2e-16
Score (logrank) test = 1474 on 5 df,  p=<2e-16
```

Model ini mencakup variabel 'Light_ISF', 'Species', dan 'Sterile' sebagai faktor.

Interpretasi

- Nilai concordance adalah 0.758 dengan standard error 0.006. Concordance menunjukkan seberapa baik model memprediksi urutan kejadian (event) dengan benar. Nilai 0.758 menunjukkan bahwa model memprediksi urutan waktu kejadian dengan benar sekitar 75.8. Ini menunjukkan model memiliki performa prediksi yang cukup baik.
- Likelihood ratio test menghasilkan nilai chi-square 1524 pada 5 derajat kebebasan (df) dengan p-value $< 2e-16$. Tes ini mengukur apakah

model dengan prediktor yang diikutsertakan memberikan peningkatan yang signifikan dibandingkan model tanpa prediktor. *P-value* yang sangat kecil menunjukkan bahwa model dengan prediktor lebih baik secara signifikan dibandingkan model tanpa prediktor.

- c) Wald test menghasilkan nilai chi-square 1145 pada 5 df dengan *p-value* $< 2e-16$. Tes ini mengevaluasi signifikansi masing-masing koefisien dalam model. *P-value* yang sangat kecil menunjukkan bahwa setidaknya satu dari prediktor dalam model secara signifikan berkontribusi terhadap model.
- d) Score (logrank) test menghasilkan nilai chi-square 1474 pada 5 df dengan *p-value* $< 2e-16$. Tes ini juga mengevaluasi apakah ada perbedaan yang signifikan dalam model dibandingkan dengan model tanpa prediktor. *P-value* yang sangat kecil menunjukkan bahwa ada bukti kuat bahwa setidaknya satu dari prediktor dalam model berkontribusi secara signifikan terhadap risiko kejadian (event).

Pengecekan Asumsi Proporsionalitas Hazards (Cox-PH) untuk Model1

Selanjutnya, akan dilakukan pengujian asumsi proportional hazard dengan menggunakan Uji Global dengan hipotesis pengujian sebagai berikut,

H_0 : asumsi proporsionalitas hazard terpenuhi atau rasio hazard konstan sepanjang waktu

H_1 : asumsi proporsionalitas hazard tidak terpenuhi atau rasio hazard berubah seiring waktu

```
> uji_ph = cox.zph(model1)
> uji_ph
```

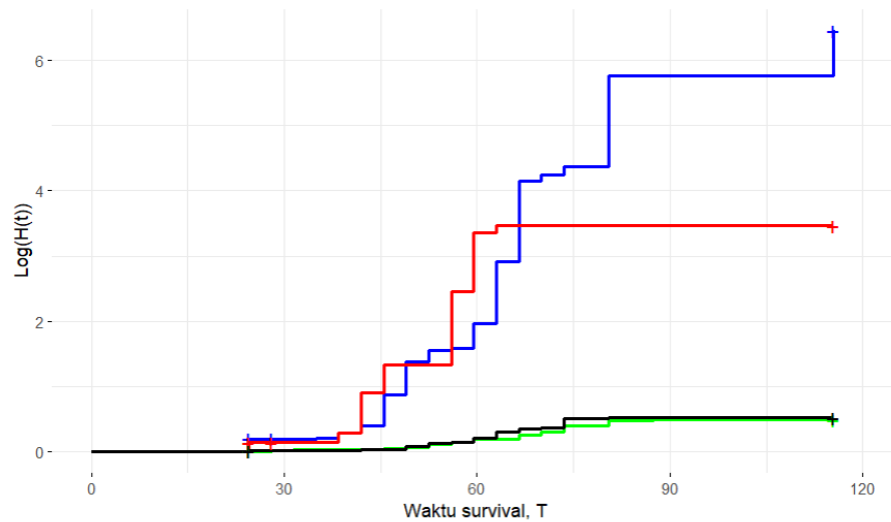
	chisq	df	p
Light_ISF	2.28	1	0.13
as.factor(Species)	39.51	3	1.4e-08
as.factor(Sterile)	15.87	1	6.8e-05
GLOBAL	55.37	5	1.1e-10

- a) Light_ISF menghasilkan *p-value* = 0.13 > 0.05 , menunjukkan bahwa terdapat bukti signifikan bahwa efek dari variabel 'Light_ISF' pada risiko tidak berubah seiring waktu, yang berarti memenuhi asumsi PH,
- b) as.factor(Species) menghasilkan *p-value* = 1.4e-08 < 0.05 , menunjukkan bahwa terdapat bukti signifikan bahwa efek dari variabel 'Species' pada risiko berubah seiring waktu, yang berarti pelanggaran terhadap asumsi PH,
- c) as.factor(Sterile) menghasilkan *p-value* = 6.8e-08 < 0.05 , menunjukkan bahwa terdapat bukti signifikan bahwa efek dari variabel 'Sterile' pada risiko berubah seiring waktu, yang berarti pelanggaran terhadap asumsi PH,
- d) GLOBAL menghasilkan *p-value* = 1.1e-10 < 0.05 , menunjukkan bahwa setidaknya satu variabel dalam model secara keseluruhan tidak memenuhi asumsi PH.

Pengecekan Asumsi dengan Plot Log(H(t))

Pemeriksaan asumsi PH untuk variabel Species

Strata Species=Acer saccharum Species=Prunus serotina Species=Quercus alba Species=Quercus rubra

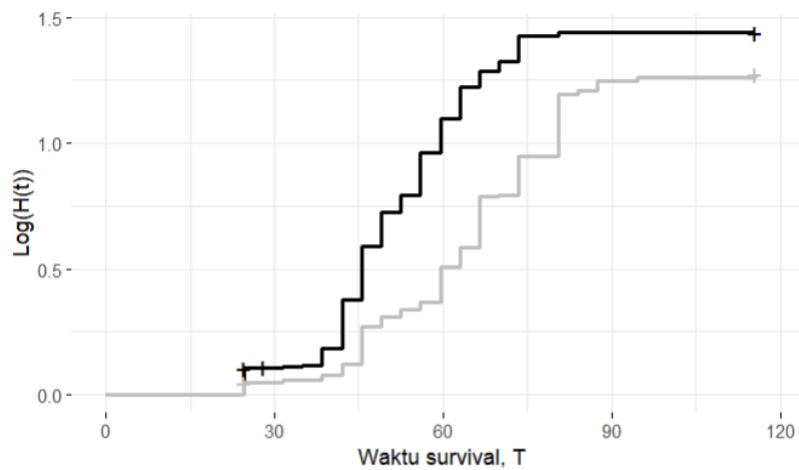


Number at risk

Species=Acer saccharum	743	516	88	2	0
Species=Prunus serotina	735	522	21	19	0
Species=Quercus alba	661	405	341	253	0
Species=Quercus rubra	600	345	283	208	0
	0	30	60	90	120

Pemeriksaan asumsi PH untuk variabel Sterile

Strata Sterile=Non-Sterile Sterile=Sterile



Number at risk

Sterile=Non-Sterile	2321	1533	572	405	0
Sterile=Sterile	418	255	161	77	0
	0	30	60	90	120

Ketika dilakukan pengecekan asumsi PH dengan plot, diperoleh bahwa asumsi PH tidak terpenuhi oleh 'Species' dan 'Sterile'. Hal ini terlihat dari ketidakparalelan dari fungsi $\log(H(t))$ untuk 4 spesies. Untuk 'Sterile', terlihat fungsi $\log(H(t))$ paralel di interval waktu 55-90 hari.

Maka dari itu, sekarang kita akan memfilter data hanya yang ada di interval 55-90 hari supaya memenuhi asumsi PH

```
# Filter data untuk waktu antara 55 dan 90 hari
data_filtered <- data[data$Time >= 55 & data$Time <= 90,]
```

Bentuk model dengan data yang sudah difilter, akan diterapkan metode backward untuk memilih model dengan variabel penjelas terbaik.

- Model2:

```
> model2 <- coxph(Surv(Time, Event) ~ Light_ISF + as.factor(Species) + as.factor(Sterile), data = data_filtered)
> summary(model2)
Call:
coxph(formula = Surv(Time, Event) ~ Light_ISF + as.factor(Species) +
      as.factor(Sterile), data = data_filtered)

n= 483, number of events= 483

              coef exp(coef) se(coef)      z
Light_ISF      0.5310    1.7007   2.4816  0.214
as.factor(Species)Prunus serotina  1.8815    6.5630   0.1497 12.570
as.factor(Species)Quercus alba    -1.3782    0.2520   0.1475  -9.345
as.factor(Species)Quercus rubra   -0.9519    0.3860   0.1477  -6.444
as.factor(Sterile)Sterile        -1.1130    0.3286   0.1262  -8.817
              Pr(>|z|)
Light_ISF      0.831
as.factor(Species)Prunus serotina < 2e-16 ***
as.factor(Species)Quercus alba    < 2e-16 ***
as.factor(Species)Quercus rubra   1.16e-10 ***
as.factor(Sterile)Sterile         < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95
Light_ISF      1.7007    0.5880   0.01313
as.factor(Species)Prunus serotina  6.5630    0.1524   4.89433
as.factor(Species)Quercus alba    0.2520    3.9677   0.18877
as.factor(Species)Quercus rubra    0.3860    2.5907   0.28896
as.factor(Sterile)Sterile         0.3286    3.0436   0.25654
              upper .95
Light_ISF      220.2706
as.factor(Species)Prunus serotina  8.8007
as.factor(Species)Quercus alba    0.3365
as.factor(Species)Quercus rubra    0.5156
as.factor(Sterile)Sterile         0.4208

Concordance= 0.816 (se = 0.014 )
Likelihood ratio test= 442.1 on 5 df,  p=<2e-16
Wald test              = 401.6 on 5 df,  p=<2e-16
Score (logrank) test = 555 on 5 df,  p=<2e-16
```

Terlihat variabel 'Light_ISF' memiliki nilai *p-value* > 0.05 yang berarti variabel ini tidak signifikan dalam menjelaskan model, sehingga menurut metode backward, variabel ini akan dihapus.

- Model3:

Pada model ini, variabel 'Light_ISF' akan dihapus karena pada model2 variabel ini terbukti tidak signifikan.

```

> model3 <- coxph(Surv(Time, Event) ~ as.factor(Species) + as.factor(Sterile), data
= data_filtered)
> summary(model3)
Call:
coxph(formula = Surv(Time, Event) ~ as.factor(Species) + as.factor(Sterile),
      data = data_filtered)

n= 483, number of events= 483

              coef exp(coef) se(coef)      z Pr(>|z|)
as.factor(Species)Prunus serotina  1.8796   6.5506  0.1494 12.581 < 2e-16 ***
as.factor(Species)Quercus alba    -1.3808   0.2514  0.1470 -9.394 < 2e-16 ***
as.factor(Species)Quercus rubra   -0.9544   0.3851  0.1473 -6.478 9.3e-11 ***
as.factor(Sterile)Sterile         -1.1134   0.3284  0.1262 -8.825 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
as.factor(Species)Prunus serotina  6.5506   0.1527   4.8879   8.7790
as.factor(Species)Quercus alba     0.2514   3.9782   0.1884   0.3353
as.factor(Species)Quercus rubra    0.3851   2.5970   0.2885   0.5140
as.factor(Sterile)Sterile          0.3284   3.0448   0.2565   0.4206

Concordance= 0.818 (se = 0.014 )
Likelihood ratio test= 442 on 4 df,  p=<2e-16
Wald test              = 401.5 on 4 df,  p=<2e-16
Score (logrank) test = 554.9 on 4 df,  p=<2e-16

```

Terlihat variabel pada model3 seluruh variabel memiliki nilai *p-value* < 0.05 yang berarti seluruh variabel signifikan dalam menjelaskan model.

- Model Cox Proportional Hazards nilai concordance sebesar 0.818 dengan Standard Error (SE) 0.014. Nilai concordance ini mengindikasikan bahwa sekitar 81.8% dari pasangan subjek yang dibandingkan memiliki prediksi risiko yang konsisten dengan hasil yang diamati.
- Likelihood Ratio Test menunjukkan nilai 442 dengan p-value kurang dari 2e-16, mengindikasikan bahwa model dengan variabel yang ditetapkan secara signifikan lebih baik dalam menjelaskan data dibandingkan dengan model kosong. Nilai p-value yang sangat kecil menegaskan signifikansi statistik yang kuat dari model.
- Wald Test memiliki nilai 401.5 pada 4 derajat kebebasan dengan p-value kurang dari 2e-16. Wald Test ini membuktikan signifikansi individu dari koefisien yang diestimasi dalam model, menyiratkan bahwa variabel-variabel yang termasuk dalam model memberikan kontribusi signifikan terhadap pemahaman risiko yang dihadapi subjek.
- Score (logrank) Test menunjukkan nilai 554.9 dengan p-value kurang dari 2e-16. Hasil ini menegaskan keseluruhan kecocokan model yang sangat baik terhadap data, mengkonfirmasi bahwa model secara signifikan mengidentifikasi perbedaan dalam risiko survival antara kelompok yang berbeda sesuai dengan faktor yang ditetapkan dalam model.

Pengecekan Asumsi Proporsionalitas Hazards (Cox-PH) untuk Model3

H_0 : asumsi proporsionalitas hazard terpenuhi atau rasio hazard konstan sepanjang waktu
 H_1 : asumsi proporsionalitas hazard tidak terpenuhi atau rasio hazard berubah seiring waktu

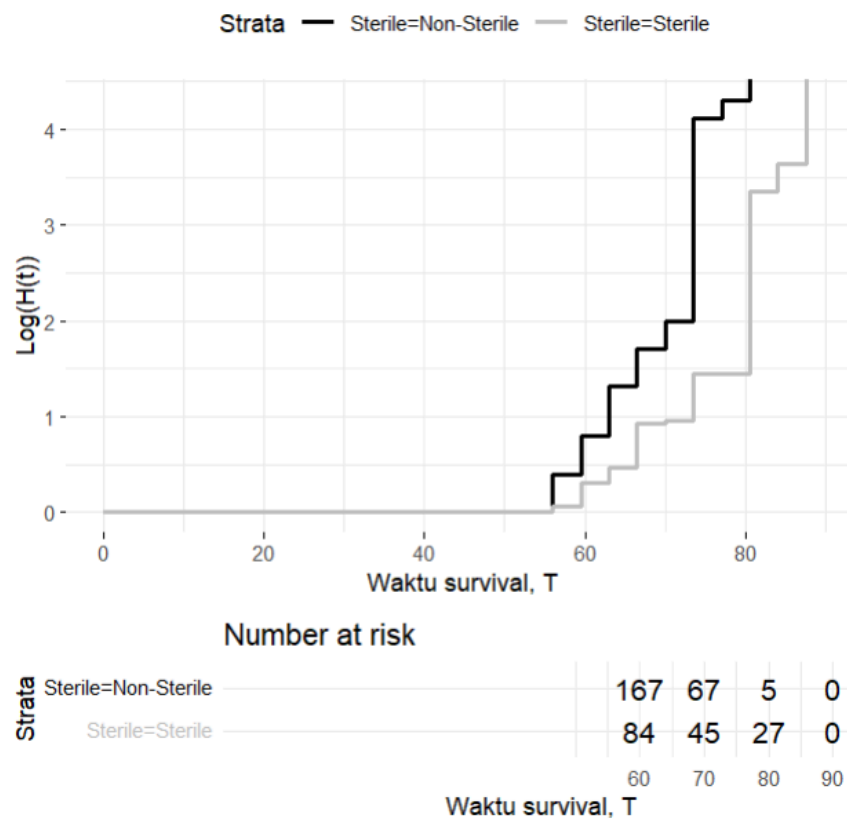
```
> uji_ph = cox.zph(model3)
> uji_ph
```

	chisq	df	p
as.factor(Species)	11.265	3	0.010
as.factor(Sterile)	0.366	1	0.545
GLOBAL	12.576	4	0.014

Perhatikan bahwa,

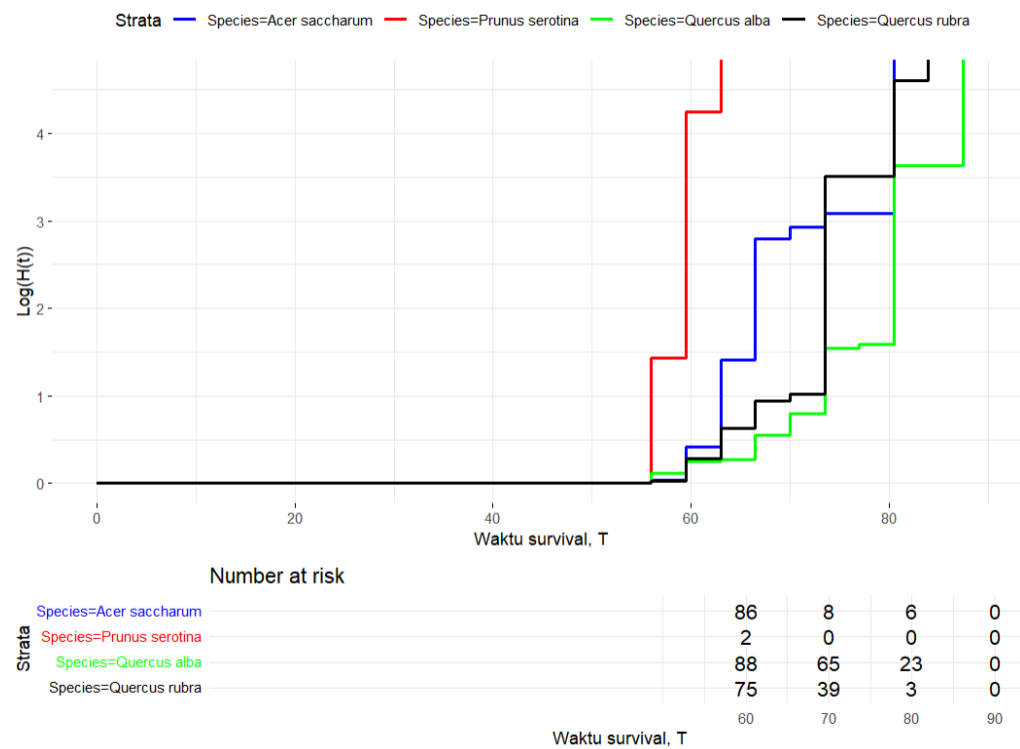
- as.factor(Species) menghasilkan $p\text{-value} = 0.010 < 0.05$, menunjukkan bahwa terdapat bukti signifikan bahwa efek dari variabel 'Species' pada risiko berubah seiring waktu, yang berarti pelanggaran terhadap asumsi PH,
- as.factor(Sterile) menghasilkan $p\text{-value} = 0.545 > 0.05$, menunjukkan tidak ada bukti signifikan bahwa efek dari 'Sterile' tidak berubah seiring waktu, yang berarti memenuhi asumsi PH,
- GLOBAL menghasilkan $p\text{-value} = 0.0414 < 0.05$, menunjukkan bahwa setidaknya satu variabel dalam model secara keseluruhan tidak memenuhi asumsi PH.

Pengecekan Asumsi dengan Plot Log(H(t))



Dari plot di atas, terlihat bahwa

- Kurva untuk 'Sterile' (hitam) dan 'Non-Sterile' (abu-abu) menunjukkan keparalelan yang lebih baik dibandingkan grafik 'Species'.
- Meskipun ada perbedaan jelas dalam level hazard, kurva kedua strata tersebut cenderung bergerak secara paralel sepanjang interval waktu yang diperlihatkan.
- Tidak ada lonjakan yang signifikan yang memisahkan kedua kurva pada interval waktu tertentu.



Dari plot di atas, terlihat bahwa

- Kurva untuk Acer saccharum, Prunus serotina, Quercus alba, dan Quercus rubra menunjukkan perbedaan yang signifikan dalam perilaku hazard sepanjang waktu, dengan setiap spesies menunjukkan lonjakan pada waktu yang berbeda.
- Acer saccharum (biru) stabil hingga sekitar 60 hari, kemudian mengalami peningkatan.
- Prunus serotina (merah) dan Quercus alba (hijau) menunjukkan peningkatan pada waktu yang berbeda; Prunus serotina meningkat setelah 60 hari dan Quercus alba setelah 70 hari.
- Quercus rubra (hitam) juga stabil hingga sekitar 70 hari sebelum mengalami peningkatan.
- Hal ini menunjukkan bahwa perilaku hazard dari masing-masing spesies berubah secara signifikan sepanjang waktu, yang mengindikasikan pelanggaran terhadap asumsi PH.

- Model4:

Pengajuan model dengan stratifikasi agar model Cox PH memenuhi asumsi

```
data$Species <- as.factor(data$Species)
data$Sterile <- as.factor(data$Sterile)
```

Call:

```
coxph(formula = Surv(Time, Event) ~ Sterile + strata(Species),
      data = data_filtered)
```

n= 483, number of events= 483

```
              coef exp(coef) se(coef)      z Pr(>|z|)
SterileSterile -1.1740    0.3091  0.1322 -8.882  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
              exp(coef) exp(-coef) lower .95 upper .95
SterileSterile    0.3091      3.235   0.2386   0.4005
```

Concordance= 0.637 (se = 0.019)

Likelihood ratio test= 93.26 on 1 df, p=<2e-16

Wald test = 78.89 on 1 df, p=<2e-16

Score (logrank) test = 83.81 on 1 df, p=<2e-16

Pengecekan Asumsi Proporsionalitas Hazards (Cox-PH) untuk Model4

H_0 : asumsi proporsionalitas hazard terpenuhi atau rasio hazard konstan sepanjang waktu

H_1 : asumsi proporsionalitas hazard tidak terpenuhi atau rasio hazard berubah seiring waktu

```
> uji.ph
              chisq df    p
Sterile 0.0615  1 0.8
GLOBAL  0.0615  1 0.8
```

Perhatikan bahwa

- as.factor(Species): p-value 0.8, menunjukkan tidak ada bukti signifikan bahwa efek dari 'Sterile' berubah sepanjang waktu, memenuhi asumsi PH,
- GLOBAL: p-value 0.8, menunjukkan bahwa variabel dalam model secara keseluruhan memenuhi asumsi PH

Model Matematika

$$h(t) = h_0(t|Species) \exp(-1.1740 \times Sterile), 55 \leq t \leq 90$$

Dengan mempertimbangkan stratifikasi berdasarkan Spesies, model ini mengontrol variasi yang disebabkan oleh perbedaan jenis Spesies. Maka dari itu, dapat disimpulkan jika suatu tanaman adalah sterile, risiko terjadinya event berkurang sebesar $e^{-1.1740} = 0.3091$ kali dibandingkan dengan tanaman yang tidak steril, yang berarti ada penurunan risiko sekitar 69% selama periode waktu dari hari ke-55 hingga hari ke-90. Nilai p (<2e-16) menunjukkan bahwa efek ini sangat signifikan secara statistik.

Model Cox Proportional Hazards yang telah dikembangkan dengan stratifikasi menunjukkan

- a) Nilai concordance sebesar 0.637 dengan Standard Error (SE) 0.019. Nilai concordance ini, walaupun lebih rendah dibandingkan model sebelumnya, mengindikasikan bahwa sekitar 63.7% dari pasangan subjek yang dibandingkan memiliki prediksi risiko yang konsisten dengan hasil yang diamati. Penurunan nilai concordance ini dapat diatribusikan pada penggunaan stratifikasi dalam model, yang membatasi keragaman data dalam setiap strata tetapi memberikan kecocokan yang lebih tepat terhadap asumsi dasar dari model Cox Proportional Hazards.
- b) Nilai Likelihood Ratio Test adalah 93.26 dengan p-value kurang dari $2e-16$. Hal ini menunjukkan bahwa model secara signifikan lebih baik dalam menjelaskan data dibandingkan model kosong (tanpa covariates). P-value yang sangat kecil ($p < 2e-16$) mengindikasikan bahwa hasil ini sangat signifikan secara statistik.
- c) Wald Test memiliki nilai 78.89 dengan p-value juga kurang dari $2e-16$. Wald test digunakan untuk menguji signifikansi koefisien regresi secara individual. Hasil ini mengindikasikan bahwa koefisien dalam model adalah signifikan secara statistik.
- d) Score (log rank) Test memiliki nilai 83.81 dengan p-value kurang dari $2e-16$. Uji ini digunakan untuk menguji keseluruhan signifikansi dari model regresi Cox. Nilai yang sangat tinggi dan p-value yang sangat kecil menunjukkan bahwa model secara signifikan lebih baik dalam menjelaskan data dibandingkan model kosong.

C. Konstruksi Likelihood Parsial

Likelihood parsial, dalam konteks model Cox Proportional Hazards adalah fungsi likelihood yang mengabaikan distribusi waktu dasar dari kejadian dan hanya fokus pada perbandingan risiko di antara individu pada setiap waktu kejadian. Hal ini memungkinkan model untuk mengestimasi efek kovariat pada hazard ratio tanpa memerlukan asumsi distribusi tertentu untuk waktu kejadian.

Untuk konstruksi likelihood parsial, akan digunakan subset data sebanyak 20 baris.

Time	Event	Light_ISF	Sterile_ Sterile
80.5	1	0.061	1
80.5	1	0.079	1
80.5	1	0.086	1
80.5	1	0.086	1
80.5	1	0.079	1
80.5	1	0.087	1
80.5	1	0.078	1

80.5	1	0.076	1
80.5	1	0.082	0
80.5	1	0.035	1
80.5	1	0.098	1
80.5	1	0.035	0
80.5	1	0.078	1
80.5	1	0.091	1
84	1	0.103	1
87.5	1	0.065	1
87.5	1	0.059	1
87.5	1	0.087	1
94.5	1	0.107	1
115.5	1	0.098	1

Dari 20 baris data di atas, terlihat terdapat 2 titik waktu dengan *ties* yaitu 80,5 dan 87,5, serta 2 titik waktu tanpa *ties* yaitu 84 dan 94,5. Akan dibentuk variabel dummy untuk variabel ‘Sterile’ dengan penjelasan

0 : jika berasal dari tanah yang tidak disterilkan

1 : jika berasal dari tanah yang disterilkan

Penjelasan variabel sebagai berikut:

x_1 : Light_ISF

x_2 : Sterile_Sterile

Bentuk umumnya adalah

$$h(t, \text{Light_ISF}, \text{Sterile_Sterile}) = h_0(t)e^{\beta_1 x_1 + \beta_2 x_2}$$

Titik waktu tanpa *ties*

Untuk data tanpa *ties*, rumus likelihood parsial yang digunakan adalah,

$$L(t_j) = \frac{h_0(t_j)e^{\beta x_j}}{\sum_{i \in R_j} h_0(t_j)e^{\beta x_j}} = \frac{e^{\beta x_j}}{\sum_{i \in R_j} e^{\beta x_j}}$$

Pada titik waktu $t = 84$, titik waktu yang masih dalam risiko adalah titik waktu ≥ 84 .

Maka, kontribusi likelihood parsial pada titik waktu $t = 84$ adalah

$$L(t_{15}) = \frac{e^{\beta_1 x_{1(15)} + \beta_2 x_{2(15)}}}{e^{\beta_1 x_{1(15)} + \beta_2 x_{2(15)}} + e^{\beta_1 x_{1(16)} + \beta_2 x_{2(16)}} + e^{\beta_1 x_{1(17)} + \beta_2 x_{2(17)}} + e^{\beta_1 x_{1(18)} + \beta_2 x_{2(18)}} + e^{\beta_1 x_{1(19)} + \beta_2 x_{2(19)}} + e^{\beta_1 x_{1(20)} + \beta_2 x_{2(20)}}}$$

$$L(t_{15}) = \frac{e^{-1.787467015}}{e^{0.103\beta_1 + \beta_2} + e^{0.065\beta_1 + \beta_2} + e^{0.059\beta_1 + \beta_2} + e^{0.087\beta_1 + \beta_2} + e^{0.107\beta_1 + \beta_2} + e^{0.098\beta_1 + \beta_2}}$$

Partial Likelihood Keseluruhan untuk Data tanpa *Ties*

Kontribusi parsial likelihood secara keseluruhan untuk data tanpa *ties* adalah,

$$L(t_j) = \frac{e^{-13.10168248X_{1(j)} - 0.43799372X_{2(j)}}}{\sum_{i \in R_j} e^{-13.10168248X_{1(i)} - 0.43799372X_{2(i)}}}$$

Titik waktu dengan ties

Akan dikonstruksi fungsi likelihood parsial untuk titik waktu dengan ties yaitu $t = 87,5$.

a. Metode Breslow

$$L(t_1) = \prod_{i=1}^D \frac{\exp(\beta'Z_i)}{[\sum_{l \in R_i} \exp(\beta'x_l)]^{d_i}}$$

Keterangan:

D : banyaknya event pada titik waktu tersebut

$\exp(\beta'Z_i)$: hazard rate dari bibit pohon ke- i pada titik waktu tersebut

$\sum_{l \in R_i} \exp(\beta'x_l)$: hazard rate dari himpunan risiko pada titik waktu tersebut

Sebelum dilakukan konstruksi likelihood dengan Metode Breslow, akan dikonstruksi hazard rate dari himpunan risiko pada waktu tersebut. Dimana pada titik waktu 87.5 merupakan titik waktu pertama dengan event yang tidak ada sensor, jumlah data himpunan risiko merupakan 20 dimana hal tersebut adalah semua bibit pohon dalam subset. Berikut perhitungan hazard rate dari himpunan risiko titik waktu tersebut.

$$\begin{aligned} \sum_{l \in R_i} \exp(\beta'x_l) &= e^{\beta x_1} + e^{\beta x_2} + e^{\beta x_3} + \dots + e^{\beta x_{20}} \\ &= e^{\beta_1 x_{1,1}} + e^{\beta_2 x_{2,1}} + e^{\beta_1 x_{1,2}} + e^{\beta_2 x_{2,2}} + \dots + e^{\beta_1 x_{1,20}} + e^{\beta_2 x_{2,20}} \\ &= e^{0.061\beta_1 + \beta_2} + e^{0.079\beta_1 + \beta_2} + e^{0.086\beta_1 + \beta_2} + e^{0.086\beta_1 + \beta_2} \\ &\quad + e^{0.079\beta_1 + \beta_2} + e^{0.087\beta_1 + \beta_2} + e^{0.078\beta_1 + \beta_2} + e^{0.076\beta_1 + \beta_2} \\ &\quad + e^{0.082\beta_1} + e^{0.035\beta_1 + \beta_2} + e^{0.098\beta_1 + \beta_2} + e^{0.035\beta_1} \\ &\quad + e^{0.078\beta_1 + \beta_2} + e^{0.091\beta_1 + \beta_2} + e^{0.103\beta_1 + \beta_2} + e^{0.065\beta_1 + \beta_2} \\ &\quad + e^{0.059\beta_1 + \beta_2} + e^{0.087\beta_1 + \beta_2} + e^{0.107\beta_1 + \beta_2} + e^{0.098\beta_1 + \beta_2} \end{aligned}$$

Lalu akan dicari hazard rate dari $t = 87,5$, berikut akan dilihat model pada $t = 87.5$.

Untuk pohon pertama : $\exp(\beta'Z_i) = e^{0.065\beta_1 + \beta_2}$

Untuk pohon kedua : $\exp(\beta'Z_i) = e^{0.059\beta_1 + \beta_2}$

Untuk pohon tiga : $\exp(\beta'Z_i) = e^{0.087\beta_1 + \beta_2}$

Maka, dengan menggunakan metode Breslow, kontribusi likelihood parsial pada titik waktu $t = 87,5$ adalah

$$\begin{aligned}
L(t_1) &= \prod_{i=1}^D \frac{\exp(\beta'Z_i)}{[\sum_{l \in R_i} \exp(\beta'x_l)]^{d_i}} \\
&= \left[\frac{\exp(\beta'Z_1)}{[\sum_{l \in R_1} \exp(\beta'x_l)]} \right] \left[\frac{\exp(\beta'Z_2)}{[\sum_{l \in R_1} \exp(\beta'x_l)]} \right] \left[\frac{\exp(\beta'Z_3)}{[\sum_{l \in R_1} \exp(\beta'x_l)]} \right] \\
&= \frac{\exp(\beta'Z_1)\exp(\beta'Z_2)\exp(\beta'Z_3)}{[\sum_{l \in R_1} \exp(\beta'x_l)]^3} \\
&= \frac{e^{0.065\beta_1+\beta_2} + e^{0.059\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2}}{[\sum_{l \in R_1} \exp(\beta'x_l)]^3} \\
&= \frac{e^{0.211\beta_1+\beta_2}}{[\sum_{l \in R_1} \exp(\beta'x_l)]^3} \\
&= \frac{e^{0.211\beta_1+\beta_2}}{[e^{0.061\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} + e^{0.076\beta_1+\beta_2} + e^{0.082\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.082\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} + e^{0.091\beta_1+\beta_2} + e^{0.103\beta_1+\beta_2} + e^{0.065\beta_1+\beta_2} + e^{0.059\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.107\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2}]^3}
\end{aligned}$$

b. Metode Efron

$$L(t_1) = \prod_{i=1}^D \frac{\exp(\beta'Z_i)}{\prod_{l \in R_i} [\sum_{k \in D_i} \exp(\beta'x_k)]^{d_i - \frac{j-1}{d_i} \sum_{k \in D_i} \exp(\beta'x_k)}}$$

Sebelum dilakukan konstruksi Metode Efron, akan dikonstruksi hazard rate dari himpunan risiko pada waktu tersebut. Dimana pada titik waktu 87.5 merupakan titik waktu pertama dengan event yang tidak ada sensor, jumlah data himpunan risiko merupakan 20 dimana hal tersebut adalah semua bibit pohon dalam subset. Berikut perhitungan hazard rate dari himpunan risiko titik waktu tersebut.

$$\begin{aligned}
\sum_{l \in R_i} \exp(\beta'x_l) &= e^{0.061\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} + e^{0.076\beta_1+\beta_2} + e^{0.082\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.082\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} + e^{0.091\beta_1+\beta_2} + e^{0.103\beta_1+\beta_2} + e^{0.065\beta_1+\beta_2} + e^{0.059\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.107\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2}
\end{aligned}$$

$$\begin{aligned}
&\sum_{l \in R_i} \exp(\beta'x_l) - \frac{1}{3} \sum_{k \in D_i} \exp(\beta'x_k) \\
&= e^{0.061\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} + e^{0.076\beta_1+\beta_2} + e^{0.082\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.082\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} + e^{0.091\beta_1+\beta_2} + e^{0.103\beta_1+\beta_2} + e^{0.065\beta_1+\beta_2} + e^{0.059\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.107\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} - \frac{1}{3} (e^{0.065\beta_1+\beta_2} + e^{0.059\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2})
\end{aligned}$$

$$\begin{aligned}
&\sum_{l \in R_i} \exp(\beta'x_l) - \frac{2}{3} \sum_{k \in D_i} \exp(\beta'x_k) \\
&= e^{0.061\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} + e^{0.076\beta_1+\beta_2} + e^{0.082\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.082\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} + e^{0.091\beta_1+\beta_2} + e^{0.103\beta_1+\beta_2} + e^{0.065\beta_1+\beta_2} + e^{0.059\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.107\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} - \frac{2}{3} (e^{0.065\beta_1+\beta_2} + e^{0.059\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2})
\end{aligned}$$

Maka, dengan menggunakan metode Efron, kontribusi likelihood parsial pada titik waktu $t = 87,5$ adalah

$$\begin{aligned}
 L(t_1) &= \prod_{i=1}^D \frac{\exp(\beta'Z_i)}{\prod_{l \in R_i} [\sum_{k \in D_i} \exp(\beta'x_k)]^{\frac{j-1}{d_i}}} \\
 &= \left[\frac{\exp(\beta'Z_1)}{\sum_{l \in R_1} \exp(\beta'x_l) - \frac{1-1}{3} \sum_{k \in D_1} \exp(\beta'x_k)} \right] \left[\frac{\exp(\beta'Z_2)}{\sum_{l \in R_1} \exp(\beta'x_l) - \frac{2-1}{3} \sum_{k \in D_1} \exp(\beta'x_k)} \right] \left[\frac{\exp(\beta'Z_3)}{\sum_{l \in R_1} \exp(\beta'x_l) - \frac{3-1}{3} \sum_{k \in D_1} \exp(\beta'x_k)} \right] \\
 &= \left[\frac{e^{0.065\beta_1 + \beta_2}}{\sum_{l \in R_1} \exp(\beta'x_l)} \right] \left[\frac{e^{0.059\beta_1 + \beta_2}}{\sum_{l \in R_1} \exp(\beta'x_l) - \frac{1}{3} \sum_{k \in D_1} \exp(\beta'x_k)} \right] \left[\frac{e^{0.087\beta_1 + \beta_2}}{\sum_{l \in R_1} \exp(\beta'x_l) - \frac{2}{3} \sum_{k \in D_1} \exp(\beta'x_k)} \right]
 \end{aligned}$$

Hasil hazard rate dari himpunan risiko untuk ketiga persamaan sesuai dengan konstruksi sebelumnya.

c. Metode Diskrit

$$L(t_1) = \prod_{i=1}^D \frac{\exp(\beta'Z_i)}{[\sum_{q \in Q_i} \exp(\beta'z_q^*)]}$$

$\sum_{q \in Q_i} \exp(\beta'z_q^*)$ = total hazard rate untuk seluruh kemungkinan kejadian pada himpunan risiko tersebut.

$$\begin{aligned}
 \sum_{q \in Q_i} \exp(\beta'z_q^*) &= \sum_{i>j=1}^{20} \sum_{j=1}^{20} \exp(\beta'Z_i) \exp(\beta'Z_j) \\
 &= \exp(\beta'Z_2) \exp(\beta'Z_1) + \exp(\beta'Z_3) \exp(\beta'Z_1) + \dots + \exp(\beta'Z_9) \exp(\beta'Z_1) + \dots + \\
 &\quad \exp(\beta'Z_{12}) \exp(\beta'Z_1) + \dots + \exp(\beta'Z_9) \exp(\beta'Z_2) + \dots + \exp(\beta'Z_{12}) \exp(\beta'Z_2) \\
 &\quad + \dots + \exp(\beta'Z_9) \exp(\beta'Z_3) + \dots + \exp(\beta'Z_{12}) \exp(\beta'Z_3) + \dots + \exp(\beta'Z_9) \exp(\beta'Z_4) \\
 &\quad + \dots + \exp(\beta'Z_{12}) \exp(\beta'Z_4) + \dots + \exp(\beta'Z_9) \exp(\beta'Z_5) + \dots + \exp(\beta'Z_{12}) \exp(\beta'Z_5) \\
 &\quad + \dots + \exp(\beta'Z_9) \exp(\beta'Z_6) + \dots + \exp(\beta'Z_{12}) \exp(\beta'Z_6) + \dots + \exp(\beta'Z_9) \exp(\beta'Z_7) \\
 &\quad + \dots + \exp(\beta'Z_{12}) \exp(\beta'Z_7) + \dots + \exp(\beta'Z_9) \exp(\beta'Z_8) + \dots + \exp(\beta'Z_{12}) \exp(\beta'Z_8) \\
 &\quad + \dots + \exp(\beta'Z_{12}) \exp(\beta'Z_9) + \dots + \exp(\beta'Z_{12}) \exp(\beta'Z_{10}) \\
 &\quad + \dots + \exp(\beta'Z_{12}) \exp(\beta'Z_{11}) \\
 &\quad + \exp(\beta'Z_{13}) \exp(\beta'Z_{12}) + \exp(\beta'Z_{14}) \exp(\beta'Z_{12}) + \exp(\beta'Z_{15}) \exp(\beta'Z_{12}) \\
 &\quad + \exp(\beta'Z_{16}) \exp(\beta'Z_{12}) + \exp(\beta'Z_{17}) \exp(\beta'Z_{12}) + \exp(\beta'Z_{18}) \exp(\beta'Z_{12}) \\
 &\quad + \exp(\beta'Z_{19}) \exp(\beta'Z_{12}) + \exp(\beta'Z_{20}) \exp(\beta'Z_{12}) + \exp(\beta'Z_{14}) \exp(\beta'Z_{13}) \\
 &\quad + \dots + \exp(\beta'Z_{20}) \exp(\beta'Z_{19}) \\
 &= e^{0.079\beta_1 + \beta_2} e^{0.061\beta_1 + \beta_2} + e^{0.086\beta_1 + \beta_2} e^{0.061\beta_1 + \beta_2} + \dots + e^{0.082\beta_1} e^{0.061\beta_1 + \beta_2} + \dots + \\
 &\quad e^{0.035\beta_1} e^{0.061\beta_1 + \beta_2} + \dots + e^{0.082\beta_1} e^{0.079\beta_1 + \beta_2} + \dots + e^{0.035\beta_1} e^{0.079\beta_1 + \beta_2} \\
 &\quad + \dots + e^{0.082\beta_1} e^{0.086\beta_1 + \beta_2} + \dots + e^{0.035\beta_1} e^{0.086\beta_1 + \beta_2} + \dots + e^{0.082\beta_1} e^{0.086\beta_1 + \beta_2}
 \end{aligned}$$

$$\begin{aligned}
& + \dots + e^{0.035\beta_1} e^{0.086\beta_1 + \beta_2} + \dots + e^{0.082\beta_1} e^{0.079\beta_1 + \beta_2} + \dots + e^{0.035\beta_1} e^{0.079\beta_1 + \beta_2} \\
& + \dots + e^{0.082\beta_1} e^{0.087\beta_1 + \beta_2} + \dots + e^{0.035\beta_1} e^{0.087\beta_1 + \beta_2} + \dots + e^{0.082\beta_1} e^{0.078\beta_1 + \beta_2} \\
& + \dots + e^{0.035\beta_1} e^{0.078\beta_1 + \beta_2} + \dots + e^{0.082\beta_1} e^{0.076\beta_1 + \beta_2} + \dots + e^{0.035\beta_1} e^{0.076\beta_1 + \beta_2} \\
& + \dots + e^{0.035\beta_1} e^{0.082\beta_1} + \dots + e^{0.035\beta_1} e^{0.035\beta_1 + \beta_2} + \dots + e^{0.035\beta_1} e^{0.098\beta_1 + \beta_2} \\
& + e^{0.078\beta_1 + \beta_2} e^{0.035\beta_1} + e^{0.091\beta_1 + \beta_2} e^{0.035\beta_1} + e^{0.103\beta_1 + \beta_2} e^{0.035\beta_1} \\
& + e^{0.065\beta_1 + \beta_2} e^{0.035\beta_1} + e^{0.059\beta_1 + \beta_2} e^{0.035\beta_1} + e^{0.087\beta_1 + \beta_2} e^{0.035\beta_1} \\
& + e^{0.107\beta_1 + \beta_2} e^{0.035\beta_1} + e^{0.098\beta_1 + \beta_2} e^{0.035\beta_1} + e^{0.091\beta_1 + \beta_2} e^{0.078\beta_1 + \beta_2} \\
& + \dots + e^{0.098\beta_1 + \beta_2} e^{0.107\beta_1 + \beta_2} \\
= & e^{0.14\beta_1 + 2\beta_2} + e^{0.147\beta_1 + 2\beta_2} + \dots + e^{0.143\beta_1 + \beta_2} + \dots + e^{0.096\beta_1 + \beta_2} + \dots + \\
& e^{0.161\beta_1 + \beta_2} + \dots + e^{0.114\beta_1 + \beta_2} + \dots + e^{0.168\beta_1 + \beta_2} + \dots + e^{0.121\beta_1 + \beta_2} + \dots + e^{0.168\beta_1 + \beta_2} \\
& + \dots + e^{0.121\beta_1 + \beta_2} + \dots + e^{0.161\beta_1 + \beta_2} + \dots + e^{0.114\beta_1 + \beta_2} + \dots + e^{0.169\beta_1 + \beta_2} + \dots + e^{0.122\beta_1 + \beta_2} \\
& + \dots + e^{0.16\beta_1 + \beta_2} + \dots + e^{0.113\beta_1 + \beta_2} + \dots + e^{0.158\beta_1 + \beta_2} + \dots + e^{0.111\beta_1 + \beta_2} + \dots + \\
& e^{0.035\beta_1} e^{0.082\beta_1} + \dots + e^{0.035\beta_1} e^{0.035\beta_1 + \beta_2} + \dots + e^{0.035\beta_1} e^{0.098\beta_1 + \beta_2} \\
& + e^{0.113\beta_1 + \beta_2} + e^{0.126\beta_1 + \beta_2} + e^{0.138\beta_1 + \beta_2} + e^{0.1\beta_1 + \beta_2} + e^{0.094\beta_1 + \beta_2} + e^{0.122\beta_1 + \beta_2} \\
& + e^{0.142\beta_1 + \beta_2} + e^{0.133\beta_1 + \beta_2} + e^{0.169\beta_1 + 2\beta_2} + \dots + e^{0.205\beta_1 + 2\beta_2}
\end{aligned}$$

Lalu akan dicari hazard rate dari $t = 87,5$, berikut akan dilihat model pada $t = 87.5$.

Untuk pohon pertama : $\exp(\beta'Z_i) = e^{0.065\beta_1 + \beta_2}$

Untuk pohon kedua : $\exp(\beta'Z_i) = e^{0.059\beta_1 + \beta_2}$

Untuk pohon ketiga : $\exp(\beta'Z_i) = e^{0.087\beta_1 + \beta_2}$

Maka, dengan menggunakan metode Diskrit, kontribusi likelihood parsial pada titik waktu $t = 87,5$ adalah

$$\begin{aligned}
L(t_1) &= \prod_{i=1}^D \frac{\exp(\beta'Z_i)}{[\sum_{q \in Q_i} \exp(\beta'Z_q^*)]} \\
&= \frac{\exp(\beta'Z_1)\exp(\beta'Z_2)}{\sum_{i>j=1}^{20} \sum_{j=1}^{20} \exp(\beta'Z_i)\exp(\beta'Z_j)} \\
&= \frac{\exp(\beta'Z_1)\exp(\beta'Z_2)\exp(\beta'Z_3)}{\exp(\beta'Z_2)\exp(\beta'Z_1) + \exp(\beta'Z_3)\exp(\beta'Z_1) + \dots + \exp(\beta'Z_9)\exp(\beta'Z_1) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_1) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_2) + \dots + \exp(\beta'Z_9)\exp(\beta'Z_2) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_3) + \dots + \exp(\beta'Z_9)\exp(\beta'Z_4) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_4) + \dots + \exp(\beta'Z_9)\exp(\beta'Z_5) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_5) + \dots + \exp(\beta'Z_9)\exp(\beta'Z_6) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_6) + \dots + \exp(\beta'Z_9)\exp(\beta'Z_7) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_7) + \dots + \exp(\beta'Z_9)\exp(\beta'Z_8) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_8) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_9) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_{10}) + \dots + \exp(\beta'Z_{12})\exp(\beta'Z_{11}) + \exp(\beta'Z_{13})\exp(\beta'Z_{12}) + \exp(\beta'Z_{14})\exp(\beta'Z_{12}) + \exp(\beta'Z_{15})\exp(\beta'Z_{12}) + \exp(\beta'Z_{16})\exp(\beta'Z_{12})}
\end{aligned}$$

$$\begin{aligned}
& + \exp(\beta'Z_{17})\exp(\beta'Z_{12}) + \exp(\beta'Z_{18})\exp(\beta'Z_{12}) + \exp(\beta'Z_{19})\exp(\beta'Z_{12}) \\
& + \exp(\beta'Z_{20})\exp(\beta'Z_{12}) + \exp(\beta'Z_{14})\exp(\beta'Z_{13}) + \dots + \exp(\beta'Z_{20})\exp(\beta'Z_{19}) \\
= & \frac{e^{0.065\beta_1+\beta_2} e^{0.059\beta_1+\beta_2} e^{0.087\beta_1+\beta_2}}{e^{0.079\beta_1+\beta_2} e^{0.061\beta_1+\beta_2} e^{0.086\beta_1+\beta_2} e^{0.061\beta_1+\beta_2} + \dots + e^{0.082\beta_1} e^{0.061\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.061\beta_1+\beta_2} + \dots +} \\
& e^{0.082\beta_1} e^{0.079\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.079\beta_1+\beta_2} + \dots + e^{0.082\beta_1} e^{0.086\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.086\beta_1+\beta_2} \\
& + \dots + e^{0.082\beta_1} e^{0.086\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.086\beta_1+\beta_2} + \dots + e^{0.082\beta_1} e^{0.079\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.079\beta_1+\beta_2} \\
& + \dots + e^{0.082\beta_1} e^{0.087\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.087\beta_1+\beta_2} + \dots + e^{0.082\beta_1} e^{0.078\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.078\beta_1+\beta_2} + \dots + \\
& e^{0.082\beta_1} e^{0.076\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.076\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.082\beta_1} + \dots + e^{0.035\beta_1} e^{0.035\beta_1+\beta_2} + \dots + \\
& e^{0.035\beta_1} e^{0.098\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} e^{0.035\beta_1} + e^{0.091\beta_1+\beta_2} e^{0.035\beta_1} + e^{0.103\beta_1+\beta_2} e^{0.035\beta_1} + e^{0.065\beta_1+\beta_2} e^{0.035\beta_1} \\
& + e^{0.059\beta_1+\beta_2} e^{0.035\beta_1} + e^{0.087\beta_1+\beta_2} e^{0.035\beta_1} + e^{0.107\beta_1+\beta_2} e^{0.035\beta_1} + e^{0.098\beta_1+\beta_2} e^{0.035\beta_1} \\
& + e^{0.091\beta_1+\beta_2} e^{0.078\beta_1+\beta_2} + \dots + e^{0.098\beta_1+\beta_2} e^{0.107\beta_1+\beta_2} \\
= & \frac{e^{0.221+3\beta_2}}{e^{0.14\beta_1+2\beta_2} + e^{0.147\beta_1+2\beta_2} + \dots + e^{0.143\beta_1+\beta_2} + \dots + e^{0.096\beta_1+\beta_2} + \dots + e^{0.14\beta_1+2\beta_2} + e^{0.147\beta_1+2\beta_2} + \dots + e^{0.143\beta_1+\beta_2} + \dots +} \\
& e^{0.143\beta_1+\beta_2} + \dots + e^{0.096\beta_1+\beta_2} + \dots + e^{0.161\beta_1+\beta_2} + \dots + e^{0.114\beta_1+\beta_2} + \dots + e^{0.168\beta_1+\beta_2} + \dots + \\
& e^{0.121\beta_1+\beta_2} + \dots + e^{0.168\beta_1+\beta_2} + \dots + e^{0.121\beta_1+\beta_2} + \dots + e^{0.161\beta_1+\beta_2} + \dots + e^{0.114\beta_1+\beta_2} + \dots + \\
& e^{0.169\beta_1+\beta_2} + \dots + e^{0.122\beta_1+\beta_2} + \dots + e^{0.16\beta_1+\beta_2} + \dots + e^{0.113\beta_1+\beta_2} + \dots + e^{0.158\beta_1+\beta_2} + \dots + \\
& e^{0.111\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.082\beta_1} + \dots + e^{0.035\beta_1} e^{0.035\beta_1+\beta_2} + \dots + e^{0.035\beta_1} e^{0.098\beta_1+\beta_2} + \\
& e^{0.113\beta_1+\beta_2} + e^{0.126\beta_1+\beta_2} + e^{0.138\beta_1+\beta_2} + e^{0.1\beta_1+\beta_2} + e^{0.094\beta_1+\beta_2} + e^{0.122\beta_1+\beta_2} + e^{0.142\beta_1+\beta_2} \\
& + e^{0.133\beta_1+\beta_2} + e^{0.169\beta_1+2\beta_2} + \dots + e^{0.205\beta_1+2\beta_2}
\end{aligned}$$

Partial Likelihood Keseluruhan untuk Data dengan Ties

Secara umum, dapat dibentuk partial likelihood keseluruhan dari 20 titik waktu. Akan dikonstruksi hasil dari partial likelihood dari 20 titik waktu yang dipilih dari metode Breslow, Efron, dan Diskrit. Akan dilihat hazard rate himpunan risiko selain waktu ties yang dipilih yaitu $t = 1$.

- $$\begin{aligned}
\bullet \sum_{i \in R_1} e^{\beta x_i} &= e^{0.061\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} \\
&+ e^{0.076\beta_1+\beta_2} + e^{0.082\beta_1} + e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1} + e^{0.078\beta_1+\beta_2} + e^{0.091\beta_1+\beta_2} + e^{0.103\beta_1+\beta_2} \\
&+ e^{0.107\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2}
\end{aligned}$$
- $$\begin{aligned}
\bullet \sum_{i \in R_2} e^{\beta x_i} &= e^{0.079\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} + e^{0.076\beta_1+\beta_2} \\
&+ e^{0.082\beta_1} + e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1} + e^{0.078\beta_1+\beta_2} + e^{0.091\beta_1+\beta_2} + e^{0.103\beta_1+\beta_2} + e^{0.107\beta_1+\beta_2} \\
&+ e^{0.098\beta_1+\beta_2}
\end{aligned}$$
- $$\begin{aligned}
\bullet \sum_{i \in R_3} e^{\beta x_i} &= e^{0.086\beta_1+\beta_2} + e^{0.086\beta_1+\beta_2} + e^{0.079\beta_1+\beta_2} + e^{0.087\beta_1+\beta_2} + e^{0.078\beta_1+\beta_2} + e^{0.076\beta_1+\beta_2} + e^{0.082\beta_1} \\
&+ e^{0.035\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2} + e^{0.035\beta_1} + e^{0.078\beta_1+\beta_2} + e^{0.091\beta_1+\beta_2} + e^{0.103\beta_1+\beta_2} + e^{0.107\beta_1+\beta_2} + e^{0.098\beta_1+\beta_2}
\end{aligned}$$

[illegible]

- $\sum_{i \in R_{19}} e^{\beta x_i} = e^{0.107\beta_1 + \beta_2} + e^{0.098\beta_1 + \beta_2}$
- $\sum_{i \in R_{20}} e^{\beta x_i} = e^{0.098\beta_1 + \beta_2}$

Maka, akan dilihat partial likelihood keseluruhan untuk setiap metode:

a. Metode Breslow

$$\begin{aligned}
 L(\beta) &= L_1(\beta)L_2(\beta)L_3(\beta).....L_{19}(\beta)L_{20}(\beta) \\
 &= \left[\frac{\exp(\beta'Z_1)}{\sum_{i \in R_1} \exp(\beta'x_i)} \right] \left[\frac{\exp(\beta'Z_2)}{\sum_{i \in R_2} \exp(\beta'x_i)} \right] \dots \left[\frac{\exp(\beta'Z_{16})\exp(\beta'Z_{17})\exp(\beta'Z_{18})}{[\sum_{i \in R_{16}} \exp(\beta'x_i)]^3} \right] \left[\frac{\exp(\beta'Z_{19})}{\sum_{i \in R_{19}} \exp(\beta'x_i)} \right] \left[\frac{\exp(\beta'Z_{20})}{\sum_{i \in R_{20}} \exp(\beta'x_i)} \right] \\
 &= \frac{\exp(\beta'Z_1)\exp(\beta'Z_2).....\exp(\beta'Z_{16})\exp(\beta'Z_{17})\exp(\beta'Z_{18})\exp(\beta'Z_{19})\exp(\beta'Z_{20})}{\frac{[\sum_{i \in R_1} \exp(\beta'x_i)][\sum_{i \in R_2} \exp(\beta'x_i)].....[\sum_{i \in R_{16}} \exp(\beta'x_i)]^3 [\sum_{i \in R_{19}} \exp(\beta'x_i)][\sum_{i \in R_{20}} \exp(\beta'x_i)]}{e^{0.061\beta_1 + \beta_2} e^{0.079\beta_1 + \beta_2} e^{0.065\beta_1 + \beta_2} e^{0.059\beta_1 + \beta_2} e^{0.087\beta_1 + \beta_2} e^{0.107\beta_1 + \beta_2} e^{0.098\beta_1 + \beta_2}}} \\
 &= \frac{[\sum_{i \in R_1} \exp(\beta'x_i)][\sum_{i \in R_2} \exp(\beta'x_i)].....[\sum_{i \in R_{16}} \exp(\beta'x_i)]^3 [\sum_{i \in R_{19}} \exp(\beta'x_i)][\sum_{i \in R_{20}} \exp(\beta'x_i)]}{e^{1.57\beta_1 + 18\beta_2}} \\
 &= \frac{[\sum_{i \in R_1} \exp(\beta'x_i)][\sum_{i \in R_2} \exp(\beta'x_i)].....[\sum_{i \in R_{16}} \exp(\beta'x_i)]^3 [\sum_{i \in R_{19}} \exp(\beta'x_i)][\sum_{i \in R_{20}} \exp(\beta'x_i)]}{e^{1.57\beta_1 + 18\beta_2}}
 \end{aligned}$$

Hazard rate pada persamaan sudah dikerjakan pada bagian sebelumnya. Untuk

$\sum_{i \in R_{16}} \exp(\beta'x_i)$ adalah.

$$\sum_{i \in R_{16}} \exp(\beta'x_i) = e^{0.065\beta_1 + \beta_2} + e^{0.059\beta_1 + \beta_2} + e^{0.087\beta_1 + \beta_2} + e^{0.107\beta_1 + \beta_2} + e^{0.098\beta_1 + \beta_2}$$

b. Metode Efron

$$\begin{aligned}
 L(\beta) &= L_1(\beta)L_2(\beta)L_3(\beta).....L_{19}(\beta)L_{20}(\beta) \\
 &= \left[\frac{\exp(\beta'Z_1)}{\sum_{i \in R_1} \exp(\beta'x_i)} \right] \left[\frac{\exp(\beta'Z_2)}{\sum_{i \in R_2} \exp(\beta'x_i)} \right] \dots \left[\frac{\exp(\beta'Z_{16})\exp(\beta'Z_{17})\exp(\beta'Z_{18})}{[\sum_{i \in R_{16}} \exp(\beta'x_i)][\sum_{i \in R_{16}} \exp(\beta'x_i) - \frac{1}{3} \sum_{k \in D_i} \exp(\beta'x_k)] [\sum_{i \in R_{16}} \exp(\beta'x_i) - \frac{2}{3} \sum_{k \in D_i} \exp(\beta'x_k)]} \right] \left[\frac{\exp(\beta'Z_{19})}{\sum_{i \in R_{19}} \exp(\beta'x_i)} \right] \left[\frac{\exp(\beta'Z_{20})}{\sum_{i \in R_{20}} \exp(\beta'x_i)} \right] \\
 &= \frac{\exp(\beta'Z_1)\exp(\beta'Z_2).....\exp(\beta'Z_{16})\exp(\beta'Z_{17})\exp(\beta'Z_{18})\exp(\beta'Z_{19})\exp(\beta'Z_{20})}{\frac{[\sum_{i \in R_1} \exp(\beta'x_i)][\sum_{i \in R_2} \exp(\beta'x_i)].....[\sum_{i \in R_{16}} \exp(\beta'x_i)][\sum_{i \in R_{16}} \exp(\beta'x_i) - \frac{1}{3} \sum_{k \in D_i} \exp(\beta'x_k)] [\sum_{i \in R_{16}} \exp(\beta'x_i) - \frac{2}{3} \sum_{k \in D_i} \exp(\beta'x_k)] [\sum_{i \in R_{19}} \exp(\beta'x_i)][\sum_{i \in R_{20}} \exp(\beta'x_i)]}{e^{0.061\beta_1 + \beta_2} e^{0.079\beta_1 + \beta_2} e^{0.065\beta_1 + \beta_2} e^{0.059\beta_1 + \beta_2} e^{0.087\beta_1 + \beta_2} e^{0.107\beta_1 + \beta_2} e^{0.098\beta_1 + \beta_2}}} \\
 &= \frac{[\sum_{i \in R_1} \exp(\beta'x_i)][\sum_{i \in R_2} \exp(\beta'x_i)].....[\sum_{i \in R_{16}} \exp(\beta'x_i)][\sum_{i \in R_{16}} \exp(\beta'x_i) - \frac{1}{3} \sum_{k \in D_i} \exp(\beta'x_k)] [\sum_{i \in R_{16}} \exp(\beta'x_i) - \frac{2}{3} \sum_{k \in D_i} \exp(\beta'x_k)] [\sum_{i \in R_{19}} \exp(\beta'x_i)][\sum_{i \in R_{20}} \exp(\beta'x_i)]}{e^{1.57\beta_1 + 18\beta_2}} \\
 &= \frac{[\sum_{i \in R_1} \exp(\beta'x_i)][\sum_{i \in R_2} \exp(\beta'x_i)].....[\sum_{i \in R_{16}} \exp(\beta'x_i)][\sum_{i \in R_{16}} \exp(\beta'x_i) - \frac{1}{3} \sum_{k \in D_i} \exp(\beta'x_k)] [\sum_{i \in R_{16}} \exp(\beta'x_i) - \frac{2}{3} \sum_{k \in D_i} \exp(\beta'x_k)] [\sum_{i \in R_{19}} \exp(\beta'x_i)][\sum_{i \in R_{20}} \exp(\beta'x_i)]}{e^{1.57\beta_1 + 18\beta_2}}
 \end{aligned}$$

Hazard rate pada persamaan sudah dikerjakan pada bagian sebelumnya.

c. Metode Diskrit

$$\begin{aligned}
L(\beta) &= L_1(\beta)L_2(\beta)L_3(\beta).....L_{19}(\beta)L_{20}(\beta) \\
&= \left[\frac{\exp(\beta'Z_1)}{\sum_{i \in R_1} \exp(\beta'x_i)} \right] \left[\frac{\exp(\beta'Z_2)}{\sum_{i \in R_2} \exp(\beta'x_i)} \right] \dots \left[\frac{\exp(\beta'Z_{16})\exp(\beta'Z_{17})\exp(\beta'Z_{18})}{\sum_{q \in Q_i} \exp(\beta'Z_q^*)} \right] \left[\frac{\exp(\beta'Z_{19})}{\sum_{i \in R_{19}} \exp(\beta'x_i)} \right] \left[\frac{\exp(\beta'Z_{20})}{\sum_{i \in R_{20}} \exp(\beta'x_i)} \right] \\
&= \frac{\exp(\beta'Z_1)\exp(\beta'Z_2).....\exp(\beta'Z_{16})\exp(\beta'Z_{17})\exp(\beta'Z_{18})\exp(\beta'Z_{19})\exp(\beta'Z_{20})}{\prod_{i \in R_1} \exp(\beta'x_i) \prod_{i \in R_2} \exp(\beta'x_i) \dots \prod_{i \in R_{19}} \exp(\beta'x_i) \prod_{i \in R_{20}} \exp(\beta'x_i)} \\
&= \frac{e^{0.061\beta_1 + \beta_2} e^{0.079\beta_1 + \beta_2} \dots e^{0.065\beta_1 + \beta_2} e^{0.059\beta_1 + \beta_2} e^{0.087\beta_1 + \beta_2} e^{0.107\beta_1 + \beta_2} e^{0.098\beta_1 + \beta_2}}{\prod_{i \in R_1} \exp(\beta'x_i) \prod_{i \in R_2} \exp(\beta'x_i) \dots \prod_{i \in R_{19}} \exp(\beta'x_i) \prod_{i \in R_{20}} \exp(\beta'x_i)} \\
&= \frac{e^{1.57\beta_1 + 18\beta_2}}{\prod_{i \in R_1} \exp(\beta'x_i) \prod_{i \in R_2} \exp(\beta'x_i) \dots \prod_{i \in R_{19}} \exp(\beta'x_i) \prod_{i \in R_{20}} \exp(\beta'x_i)}
\end{aligned}$$

Hazard rate pada persamaan sudah dikerjakan pada bagian sebelumnya.

III. Penutup

Pada *project* ini, kami telah mengidentifikasi dan menganalisis faktor-faktor yang dapat mempengaruhi kelangsungan hidup dari pohon. Melalui serangkaian analisis, kami ingin menunjukkan bahwa variabel seperti jenis spesies, tingkat cahaya, dan status steril tanah memiliki pengaruh terhadap probabilitas kelangsungan hidup pohon. Dengan menggunakan analisis Cox Proportional Hazards, kami membangun beberapa model hingga akhirnya mendapatkan suatu model terbaik yang menunjukkan bahwa sterilitas tanah secara signifikan mengurangi risiko kematian bibit dalam periode waktu tertentu.

Dengan menggunakan analisis Cox Proportional Hazards, kami membangun beberapa model hingga akhirnya mendapatkan suatu model terbaik yang menunjukkan bahwa sterilitas tanah secara signifikan mengurangi risiko kematian bibit dalam periode waktu tertentu.

Fungsi likelihood dalam analisis survival merupakan probabilitas kondisional dari data yang diberikan berdasarkan parameter model. Ini mengasumsikan bahwa data independen terhadap parameter. Dalam *project* ini, fungsi likelihood parsial memungkinkan kami untuk memperkirakan pengaruh variabel terhadap rasio hazard tanpa harus mengasumsikan distribusi tertentu untuk waktu hingga kejadian. Ini sangat penting dalam memberikan fleksibilitas dalam model Cox Proportional Hazards yang kami kembangkan, memperkuat analisis kami terhadap data yang mungkin termasuk data yang tersensor.

IV. Lampiran

Codes dan data yang kami gunakan dilampirkan dalam folder berikut ini:

Final Project_Model Survival (A)_Kelompok C

Sumber dataset: <https://www.kaggle.com/datasets/yekenot/tree-survival-prediction>