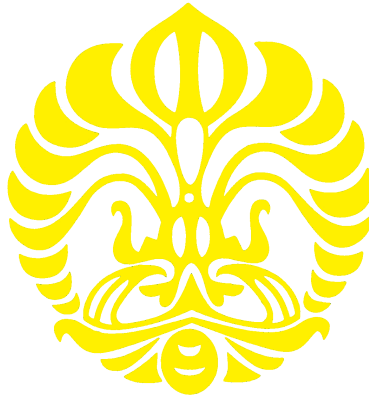


***PROJECT* MODEL SURVIVAL**

**Analisis Regresi Cox-Proportional Hazard pada Pemodelan
Kelangsungan Hidup Pasien Kanker Payudara**



Disusun oleh:

Kelompok 8

- 1. Azizah Annurohmah (2006572983)**
- 2. Diki Wahyudi (2106709131)**
- 3. Farah Ramadhani Putri (2106654832)**

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS INDONESIA
JUNI 2023**

DAFTAR ISI

ABSTRAK	ii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	2
1.4 Manfaat	2
BAB II METODE PENELITIAN	3
2.1 Variabel Penelitian	3
2.2 Data	3
2.3 Jenis Data	5
2.4 Metode Analisis Data	5
BAB III HASIL DAN PEMBAHASAN	8
3.1 Analisis Deskriptif	8
3.2 Analisis Regresi Cox-PH	14
3.3 Pengecekan Asumsi <i>Proportional Hazard</i> untuk Model Final	19
PENUTUP	21
4.1 Kesimpulan	21
DAFTAR PUSTAKA	22
<i>Link File</i>	22
LAMPIRAN	23
Lampiran I. Data Penelitian	23
PERSENTASE KONTRIBUSI KELOMPOK	27

ABSTRAK

Analisis Survival merupakan sebuah metode statistika yang memperhatikan waktu hingga terjadinya suatu kejadian (*event*). Tujuan dari analisis survival adalah menaksir probabilitas kelangsungan hidup, kematian, dan peristiwa lainnya sampai periode waktu tertentu. Salah satu Analisis Survival yang paling populer digunakan adalah regresi Cox-Proportional Hazard. Regresi Cox-Proportional Hazard merupakan salah satu regresi semiparametrik yang bertujuan untuk mengetahui kombinasi faktor-faktor yang berpengaruh terhadap variabel respon yang berupa waktu survival. Dalam penelitian ini, waktu survival yang digunakan adalah waktu lamanya hidup pasien kanker payudara setelah menjalani operasi. Data yang digunakan adalah data kelangsungan hidup pasien yang telah menjalani operasi kanker payudara antara tahun 1958 sampai 1970 di Rumah Sakit Billings Universitas Chicago. Berdasarkan hasil analisis dan pengujian diperoleh model regresi Cox-PH yang terbaik adalah $\hat{h}(t, x_1) = h_0(t) \exp(0.05462 \times x_1)$ dengan x_1 adalah banyaknya nodus aksila pada pasien. Berdasarkan kategori jumlah nodus aksila dan tahun dilakukannya operasi, pasien kanker payudara yang memiliki jumlah nodus aksila ≥ 10 setelah menjalani operasi memiliki kemungkinan 3.51750 kali untuk mengalami kematian dalam waktu 5 tahun dibandingkan pasien kanker payudara yang memiliki jumlah nodus aksila < 10 .

Kata kunci: Analisis Survival, Regresi Cox-Proportional Hazard, Kanker Payudara.

BAB I PENDAHULUAN

1.1 Latar Belakang

Kanker payudara adalah penyakit tumbuhnya sel abnormal yang tidak terkendali di jaringan payudara. Pertumbuhan sel abnormal ini dapat merusak sel normal di sekitar payudara dan dapat menjalar ke bagian tubuh yang lain. Pada tahun 2020, terdapat 2.3 juta perempuan yang didiagnosis dengan kanker payudara dan 685.000 meninggal. Tingkat mortalitas pada kanker payudara tidak banyak berubah pada tahun 1930 sampai 1970. Terdapat sedikit penurunan tingkat mortalitas pada tahun 1980 dengan berkembangnya teknologi untuk mendeteksi kanker payudara lebih dini.

Analisis regresi Cox-PH adalah metode statistik yang digunakan untuk menganalisis waktu kelangsungan hidup pasien atau waktu terjadinya suatu peristiwa, seperti kematian, dengan mempertimbangkan faktor-faktor risiko yang mempengaruhi. Metode ini relevan dalam konteks pemodelan kelangsungan hidup pasien kanker payudara, karena memungkinkan peneliti untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap prognosis pasien.

Project ini menggunakan Haberman's Survival Dataset. Dataset ini berdasarkan studi yang dilakukan pada tahun 1958 hingga 1970 di University of Chicago's Billings Hospital tentang tingkat survival pada pasien kanker payudara yang telah melakukan operasi. Dataset ini memiliki 4 variabel yaitu usia, tahun operasi, banyaknya nodus aksila positif yang terdeteksi, dan status bertahan hidup. Berdasarkan dataset Haberman's Survival Dataset akan dibuat model prediktif yang dapat memprediksi kelangsungan hidup pasien kanker payudara dan melihat hubungan antara umur, tahun operasi, dengan kelangsungan hidup pasien kanker payudara.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah dari penelitian yang dilakukan, maka dapat dibuat beberapa rumusan masalah sebagai berikut.

1. Bagaimana hubungan antara umur, tahun operasi, dan banyaknya nodus aksila positif yang terdeteksi, dengan kelangsungan hidup pasien kanker payudara?
2. Bagaimana model regresi Cox-PH untuk memprediksi kelangsungan hidup pasien kanker payudara?

1.3 Tujuan

Tujuan dari penelitian ini adalah

1. Mengidentifikasi faktor risiko yang memiliki hubungan signifikan dengan kelangsungan hidup pasien kanker payudara.
2. Mengukur kekuatan hubungan antara faktor-faktor risiko tersebut dengan risiko kelangsungan hidup pasien.
3. Mengembangkan model prediktif yang dapat memprediksi kelangsungan hidup pasien berdasarkan faktor-faktor risiko yang relevan.

1.4 Manfaat

Manfaat dari penelitian ini adalah

1. Bagi penulis
Menambah ilmu tentang Analisis Survival menggunakan salah satu metode semiparametrik, yaitu model regresi Cox-Proportional Hazard dan menerapkannya berdasarkan teori dan ilmu yang telah didapatkan selama dalam perkuliahan.
2. Bagi pembaca
Hasil penelitian diharapkan dapat berguna sebagai bahan evaluasi pengaplikasian model regresi Cox-Proportional dalam dunia nyata.

BAB II METODE PENELITIAN

2.1 Variabel Penelitian

Kami menggunakan data yang diambil dari UCI Machine Learning Repository melalui link berikut ini. <https://archive.ics.uci.edu/dataset/43/haberman+s+survival>. Dataset tersebut berisi kasus-kasus dari penelitian yang dilakukan antara tahun 1958 dan 1970 di Rumah Sakit Billings Universitas Chicago tentang kelangsungan hidup pasien yang telah menjalani operasi kanker payudara.

Keterangan variabel:

- **age** (*Age*): usia pasien pada saat menjalani operasi.
- **op_year** (*Operation Year*): tahun operasi pasien (tahun 1900-an); misalnya **op_year** = 60, berarti operasi terjadi pada tahun 1960.
- **axil_nodes**: jumlah nodus aksila positif yang terdeteksi. Nodus aksila merupakan kelenjar getah bening di daerah ketiak yang mengalirkan getah bening dari payudara dan daerah sekitarnya. Nodus aksila mengalirkan pembuluh getah bening dari kuadran lateral payudara, pembuluh getah bening superfisial dari dinding tipis dada dan perut di atas tingkat pusar, dan pembuluh dari ekstremitas atas. Nodus aksila terbagi dalam beberapa kelompok sesuai dengan lokasinya di ketiak. Kelenjar getah bening ini secara klinis signifikan pada kanker payudara, dan metastasis dari payudara ke kelenjar getah bening aksila dipertimbangkan dalam stadium penyakit.
- **surv_status** (*Survival Status*): status bertahan hidup. “1”; pasien bertahan 5 tahun atau lebih dan “2”; pasien meninggal dalam waktu 5 tahun.

2.2 Data

Import Library

```
library(readxl)
library(ggplot2)
library(survival)
library(survminer)
```

```
## Loading required package: ggpubr
```

```
##
## Attaching package: 'survminer'

## The following object is masked from 'package:survival':
##
##      myeloma
```

Import Data

```
haberman <- read_excel("D:/Materi Kuliah UI/Model Survival/Tugas Model Survival/haberman.xlsx")
colnames(haberman) <- c("age", "op_year", "axil_nodes", "surv_status")
```

Akan ditampilkan 6 baris pertama pada data tersebut.

```
head(haberman)
```

age	op_year	axil_nodes	surv_status
30	64	1	1
30	62	3	1
30	65	0	1
31	59	2	1
31	65	4	1
33	58	10	1

Untuk observasi pertama, pasien menjalani operasi kanker payudara saat usia 30 tahun pada 1964. Pasien tersebut terdeteksi memiliki 1 nodus aksila yang positif. Pasien tersebut bertahan 5 tahun atau lebih. Untuk observasi keempat, pasien menjalani operasi kanker payudara saat usia 31 tahun pada 1959. Pasien tersebut terdeteksi memiliki 2 nodus aksila yang positif. Pasien tersebut bertahan 5 tahun atau lebih.

Ukuran Data

```
dim(haberman)
```

```
## [1] 306 4
```

Terlihat bahwa data terdiri dari 306 observasi (baris) dan 4 variabel (kolom).

2.3 Jenis Data

- age: numerik, rasio
- op_year: numerik, rasio
- axil_nodes: numerik, rasio
- surv_status: kategorik, nominal

```
str(haberman)
```

```
## tibble [306 x 4] (S3: tbl_df/tbl/data.frame)
## $ age      : num [1:306] 30 30 30 31 31 33 33 34 34 34 ...
## $ op_year   : num [1:306] 64 62 65 59 65 58 60 59 66 58 ...
## $ axil_nodes : num [1:306] 1 3 0 2 4 10 0 0 9 30 ...
## $ surv_status: num [1:306] 1 1 1 1 1 1 1 2 2 1 ...
```

Karena `surv_status` masih bertipe numerik, variabel tersebut akan diubah menjadi variabel kategorik dalam bentuk factor di R.

```
haberman$surv_status <- as.factor(haberman$surv_status)
```

2.4 Metode Analisis Data

Fungsi Survival

Analisis survival didasarkan pada probabilitas bahwa peristiwa yang ingin diamati belum terjadi pada waktu t . Fungsi survival bergantung pada waktu, $S(t)$, umumnya digunakan untuk mewakili probabilitas tersebut.

$$S(t) = P(T > t)$$

Dengan kata lain, $S(t)$ sesuai dengan probabilitas bertahan hidup setelah waktu t . Dalam kasus ini, T merupakan waktu sampai pasien yang melakukan operasi kanker payudara bertahan hidup. Karena distribusi dari T tidak diketahui, akan digunakan estimasi nonparametrik yaitu estimasi Kaplan-Meier. Untuk semua t yang berada pada rentang data, taksiran fungsi survival dengan metode Kaplan-Meier (Product-Limit) adalah

$$\hat{S}(t) = \begin{cases} 1 & , t < t_1 \\ \prod_{t_i \leq t} [1 - \frac{d_i}{Y_i}] & , t_i \leq t \end{cases}$$

dengan $S(0) = 1$, dan $S(t)$ tidak terdefinisi untuk $t > t_{\max}$ dengan t_{\max} adalah waktu terbesar teramatinya suatu *event*.

Keterangan: misalkan *event* terjadi pada D titik waktu yang berbeda, yaitu pada waktu $t_1 < t_2 < \dots < t_D$. Pada waktu t_i terjadi sebanyak d_i *event*. R_i adalah himpunan yang memuat objek-objek yang berisiko mengalami event saat waktu t_i . Y_i adalah banyaknya anggota R_i . Jika penyensoran terjadi pada waktu t_i di mana terjadi *event*, maka penyensoran diasumsikan terjadi sesudah *event* terjadi, sehingga Y_i masih memperhitungkan objek yang tersensor. $\frac{d_i}{Y_i}$ adalah taksiran untuk probabilitas bersyarat bahwa subyek yang *survive* (belum mengalami *event* hingga awal waktu t_i) akan mengalami *event* saat t_i . $\frac{Y_i - d_i}{Y_i}$ adalah taksiran untuk probabilitas bersyarat bahwa subjek yang berhasil melewati waktu t_{i-1} akan *survive* melewati waktu t_i .

Fungsi Hazard

Fungsi hazard menyatakan tingkat bahaya sesaat, yakni, untuk suatu objek yang berhasil bertahan hingga waktu t , maka tingkat bahaya, atau risiko, dari objek tersebut untuk mengalami *event* secara tiba-tiba saat itu, dinyatakan dengan $h(t)$, yaitu

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{Pr(t < T \leq t + \Delta t | T > t)}{\Delta t}.$$

Fungsi hazard kumulatif, $H(t)$, dinyatakan sebagai $H(t) = \int_0^t h(u) du$, $t > 0$. Dengan menggunakan metode Nelson-Aalen, fungsi kumulatif hazard dapat ditaksir dengan perumusan berikut

$$\tilde{H}(t) = \begin{cases} 0 & , t < t_1 \\ \sum_{t_i \leq t} \frac{d_i}{Y_i} & , t_i \leq t \end{cases}$$

Regresi Cox-Proportional Hazard

Fungsi hazard pada waktu t dengan variabel penjelas x_1, \dots, x_p dinyatakan dengan model regresi Cox-PH sebagai berikut.

$$h(t, \mathbf{x}) = h_0(t)e^{\beta_1 x_1 + \dots + \beta_p x_p},$$

dengan $h_0(t)$ adalah baseline hazard yang bergantung pada waktu t , dan $e^{\beta_1 x_1 + \dots + \beta_p x_p}$ adalah komponen kontribusi dari variabel penjelas. $\beta_j, j = 1, \dots, p$ adalah koefisien regresi yang bersesuaian dengan variabel penjelas x_j .

BAB III HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif

Preprocessing

```
str(haberman)
```

```
## tibble [306 x 4] (S3: tbl_df/tbl/data.frame)
## $ age      : num [1:306] 30 30 30 31 31 33 33 34 34 34 ...
## $ op_year   : num [1:306] 64 62 65 59 65 58 60 59 66 58 ...
## $ axil_nodes : num [1:306] 1 3 0 2 4 10 0 0 9 30 ...
## $ surv_status: Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 2 2 1 ...
```

Jenis variabel pada R sudah sesuai.

Akan dicek apakah terdapat *missing value* atau tidak.

```
sapply(haberman, function(x) sum(is.na(x)))
```

```
##      age      op_year  axil_nodes surv_status
##      0          0          0          0
```

Terlihat bahwa data `haberman` tidak memiliki *missing value*.

Statistika Deskriptif

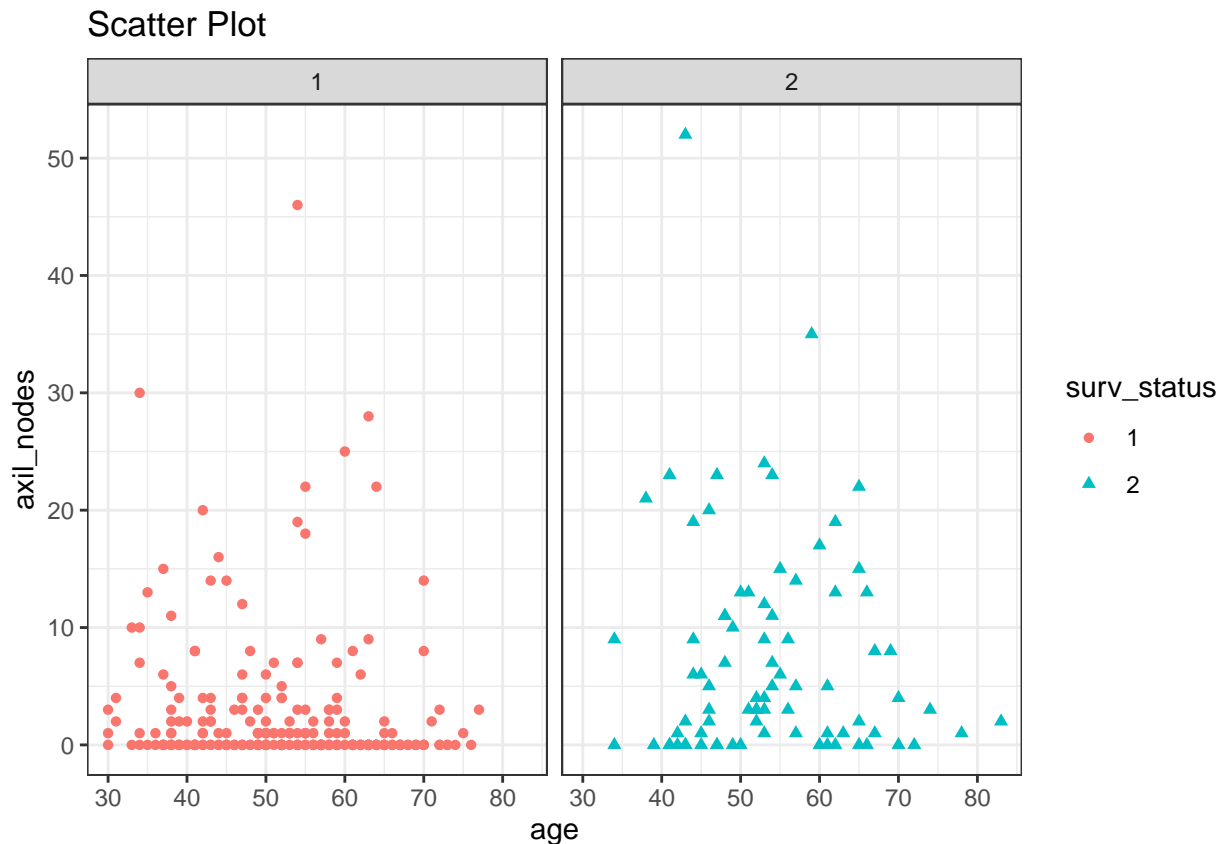
```
summary(haberman)
```

```
##      age      op_year      axil_nodes      surv_status
## Min.   :30.00  Min.   :58.00  Min.    : 0.000  1:225
## 1st Qu.:44.00  1st Qu.:60.00  1st Qu.: 0.000  2: 81
## Median :52.00  Median :63.00  Median : 1.000
## Mean   :52.46  Mean   :62.85  Mean    : 4.026
## 3rd Qu.:60.75  3rd Qu.:65.75  3rd Qu.: 4.000
## Max.   :83.00  Max.   :69.00  Max.    :52.000
```

Dalam data tersebut, banyaknya kelenjar getah bening yang terdeteksi mulai dari 0 sampai 52. Usia pasien berkisar antara 30-83 tahun, di mana pasien menjalani operasi pada waktu antara 1958–1969. Dari hasil `summary` di atas, terdapat 225 pasien dari 306 yang bertahan lebih dari 5 tahun dan hanya 81 pasien yang bertahan kurang dari 5 tahun.

Visualisasi

```
ggplot(data = haberman, aes(x = age, y = axil_nodes,
                             colour = surv_status, shape = surv_status)) +
  geom_point() + facet_wrap(~surv_status) + theme_bw() +
  ggtitle("Scatter Plot")
```



Dari visualisai tersebut, terlihat bahwa jumlah nodus aksila yang lebih tinggi cenderung memiliki probabilitas kelangsungan hidup yang lebih rendah (terdapat banyak pada `surv_status` kategori 2). Pasien dengan jumlah nodus aksila 0 memiliki peluang bertahan hidup yang lebih baik tanpa memandang usia. Pasien senior (usia lebih dari 50) dengan lebih dari 10 nodus memiliki peluang bertahan hidup yang rendah.

Fungsi Survival

Dengan menggunakan R, didapatkan hasil sebagai berikut.

```
fs.survival <- survfit(Surv(age, surv_status=="2") ~ 1, data = haberman)
summary(fs.survival)
```

```
## Call: survfit(formula = Surv(age, surv_status == "2") ~ 1, data = haberman)
```

```
##
```

```
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```
##    34    299      2   0.993 0.00471    0.984    1.000
```

```
##    38    282      1   0.990 0.00587    0.978    1.000
```

```
##    39    272      1   0.986 0.00688    0.973    1.000
```

```
##    41    263      3   0.975 0.00938    0.957    0.993
```

```
##    42    253      2   0.967 0.01077    0.946    0.989
```

```
##    43    244      4   0.951 0.01320    0.926    0.978
```

```
##    44    233      3   0.939 0.01480    0.911    0.969
```

```
##    45    226      3   0.927 0.01626    0.895    0.959
```

```
##    46    217      4   0.910 0.01806    0.875    0.946
```

```
##    47    210      3   0.897 0.01930    0.860    0.935
```

```
##    48    199      3   0.883 0.02053    0.844    0.924
```

```
##    49    192      2   0.874 0.02132    0.833    0.917
```

```
##    50    182      2   0.864 0.02214    0.822    0.909
```

```
##    51    170      2   0.854 0.02302    0.810    0.900
```

```
##    52    164      4   0.833 0.02470    0.786    0.883
```

```
##    53    150      6   0.800 0.02720    0.748    0.855
```

```
##    54    139      4   0.777 0.02875    0.723    0.835
```

```
##    55    126      2   0.765 0.02959    0.709    0.825
```

```
##    56    116      2   0.751 0.03051    0.694    0.814
```

```
##    57    109      3   0.731 0.03192    0.671    0.796
```

```
##    59     91      1   0.723 0.03257    0.662    0.789
```

```
##    60     83      2   0.705 0.03403    0.642    0.775
```

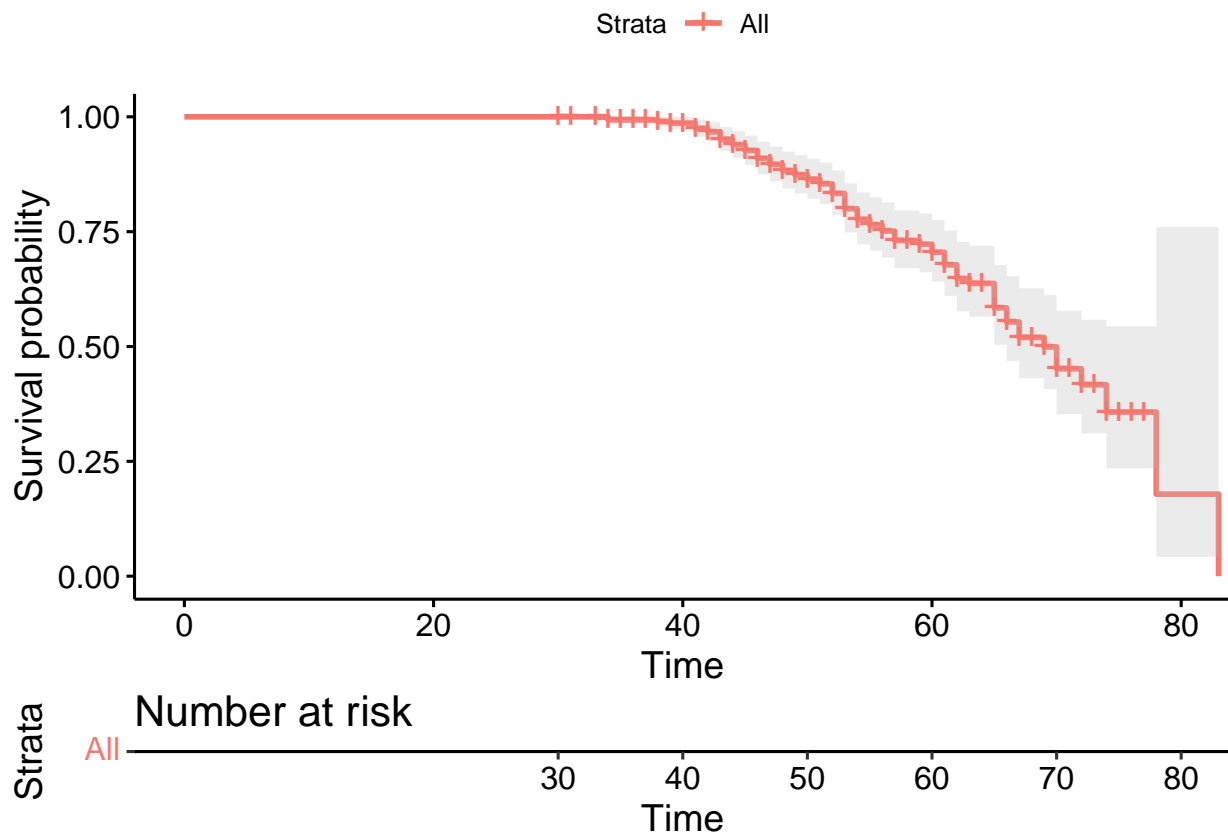
```
##    61     77      3   0.678 0.03621    0.610    0.753
```

```
##    62     68      3   0.648 0.03851    0.577    0.728
```

##	63	61	1	0.637	0.03932	0.565	0.719
##	65	48	4	0.584	0.04410	0.504	0.677
##	66	38	2	0.553	0.04684	0.469	0.653
##	67	33	2	0.520	0.04964	0.431	0.627
##	69	25	1	0.499	0.05183	0.407	0.612
##	70	21	2	0.452	0.05675	0.353	0.578
##	72	13	1	0.417	0.06211	0.311	0.558
##	74	7	1	0.357	0.07664	0.235	0.544
##	78	2	1	0.179	0.13199	0.042	0.760
##	83	1	1	0.000	NaN	NA	NA

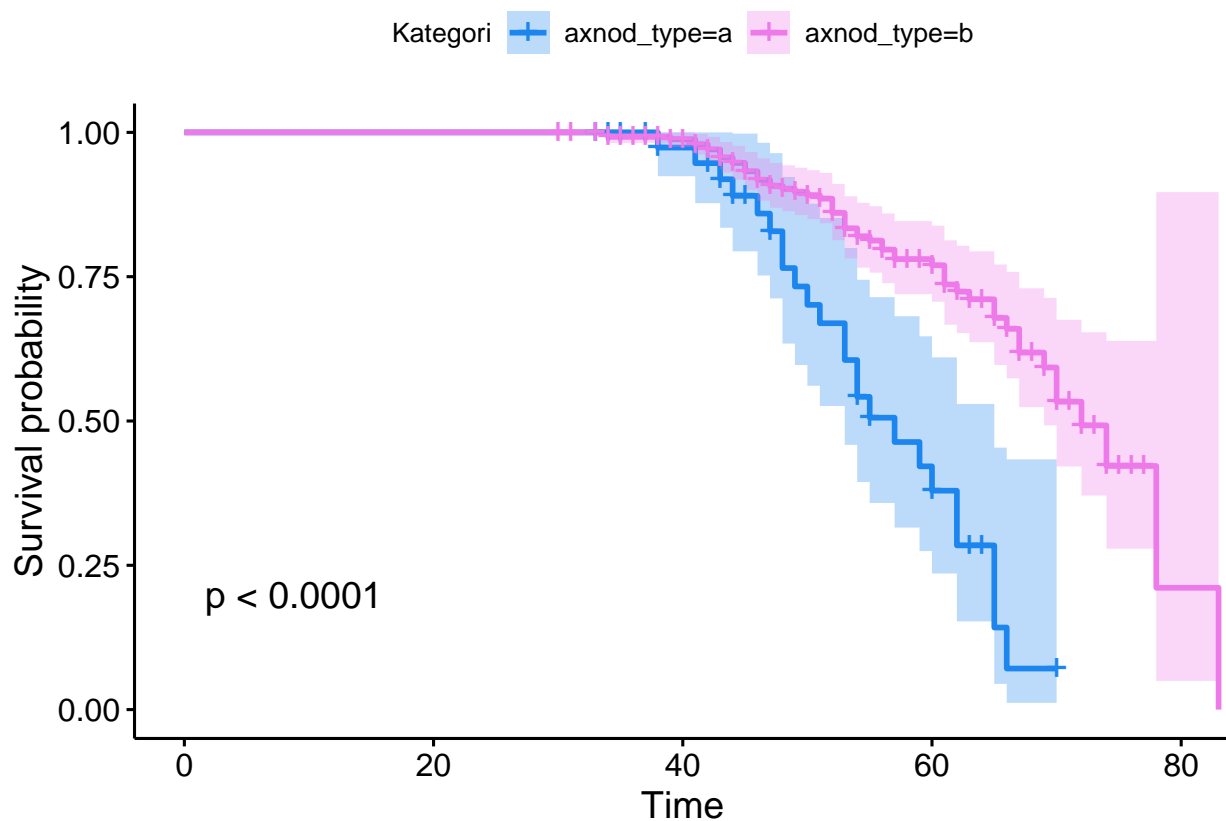
Hasil tersebut tidak efisien jika dibaca satu per satu. Oleh karena itu, hasil tersebut akan divisualisasikan.

```
ggsurvplot(fs.survival, conf.int = TRUE, pval = TRUE, risk.table = TRUE,
            main = "Kurva Fungsi Survival Kaplan-Meier untuk Pasien Kanker Payudara",
            risk.table.height = 0.15)
```



Dari hasil tersebut terlihat bahwa sekitar 50% pasien operasi kanker payudara *survive* dalam waktu lebih dari 5 tahun di usia lebih dari 65 tahun. Berdasarkan hasil visualiasi pada bagian sebelumnya, akan dibuat kurva fungsi survival berdasarkan `axil_nodes`, di mana akan dibuat menjadi dua kategori yaitu “a”: terdapat ≥ 10 nodus aksila dan “b”: terdapat < 10 nodus aksila. Akan dibuat variabel baru yaitu `axnod_type` berdasarkan kategori tersebut.

```
haberman$axnod_type <- ifelse(haberman$axil_nodes>=10, "a", "b")
haberman$axnod_type <- as.factor(haberman$axnod_type)
fs.survival2 <- survfit(Surv(age, surv_status=="2") ~ axnod_type, data = haberman)
ggsurvplot(fs.survival2, conf.int = TRUE, pval = TRUE, risk.table = FALSE,
  legend.title = "Kategori",
  palette = c("dodgerblue2", "orchid2"),
  main = "Kurva Fungsi Survival Kaplan-Meier",
  risk.table.height = 0.15)
```



Ada dua kurva fungsi survival, satu untuk setiap kategori. Dari kurva tersebut, terlihat bahwa pasien yang memiliki jumlah nodus aksila yang terdeteksi positif ≥ 10 lebih besar kemungkinannya untuk meninggal lebih cepat dibandingkan pasien yang memiliki nodus aksila yang terdeteksi positif < 10 .

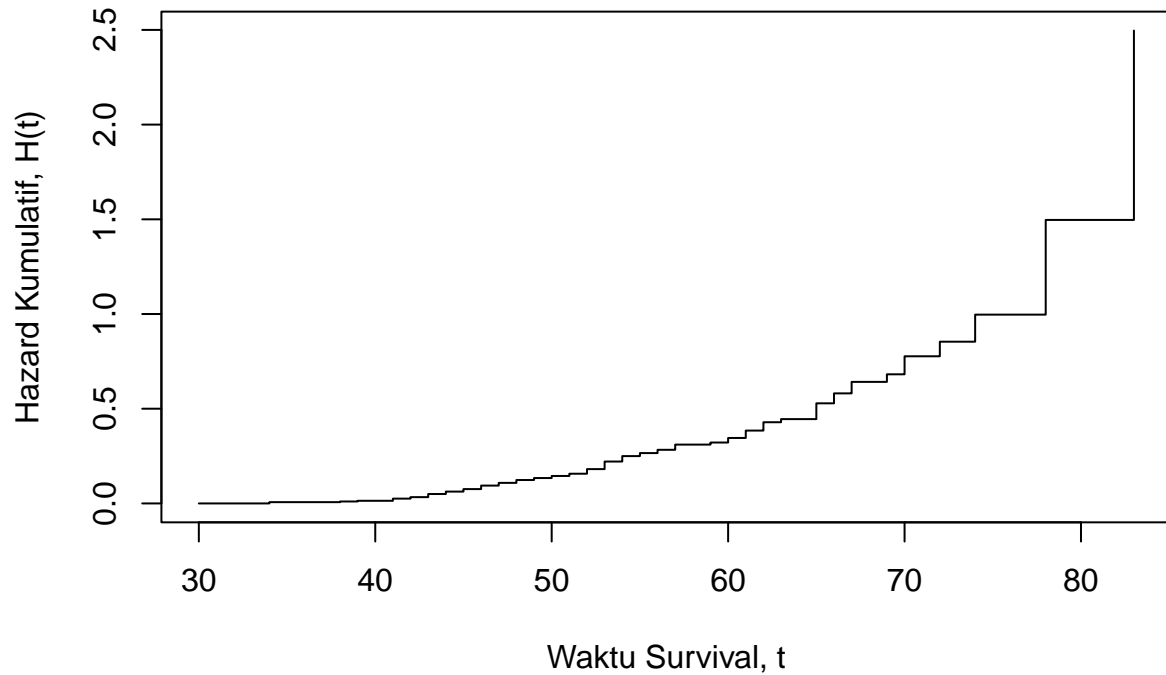
Sebelum umur 45 tahun, kedua kurva hampir sama. Pada titik di mana $t > 45$, dapat terlihat bahwa probabilitas survival berwarna biru lebih kecil daripada probabilitas survival berwarna ungu. Jadi, terdapat perbedaan antara fungsi survival berdasarkan jumlah nodus aksila yang terdeteksi positif pada pasien. Hal tersebut mengindikasikan bahwa `axnod_type` merupakan calon variabel penjelas yang cukup baik untuk waktu bertahan hidup pada pasien kanker payudara.

Fungsi Hazard

Akan diestimasi fungsi hazard dari data tersebut dengan menggunakan metode Nelson-Aalen.

```
fs.hazard <- fs.survival$n.event/fs.survival$n.risk
H.NA <- vector()
for (i in 1:length(fs.hazard)){
  H.NA[i] <- sum(fs.hazard[1:i])
}
plot(fs.survival$time, H.NA,
      xlab = "Waktu Survival, t", ylab = "Hazard Kumulatif, H(t)",
      main = "Fungsi H(t) berdasarkan Nelson-Aalen",
      type = "s", col = 1)
```


Fungsi H(t) berdasarkan Nelson–Aalen



Fungsi hazard tersebut meningkat seiring dengan bertambahnya usia yang menandakan semakin tua usia, maka risiko untuk mengalami kematian kurang dari 5 tahun setelah menjalani operasi makin tinggi.

3.2 Analisis Regresi Cox-PH

Untuk kasus ini, akan dibuat dua model regresi Cox-PH di mana variabel model pertama variabel penjelasnya yaitu `op_year` dan `axnod_type` (jenis data kategorik) dan model kedua variabel penjelasnya `op_year` dan `axil_nodes`, di mana jumlah nodus aksila tetap dalam jenis data numerik.

Model Pertama

Karena terdapat dua *level* pada variabel `axnod_type`, maka tidak diperlukan variabel *dummy*. Kategori “b” akan dijadikan *baseline* dalam model.

$$\text{axnod_type} = \begin{cases} 1 & , \text{kategori } a (\geq 10 \text{ nodus aksila}) \\ 0 & , \text{kategori } b (< 10 \text{ nodus aksila}) \end{cases}$$

```
haberman$axnod_type <- ifelse(haberman$axnod_type=="a", 1, 0)
```

Model regresi Cox-PH untuk data ini yaitu sebagai berikut.

$$h(t, \mathbf{x}) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_a)$$

dengan $x_1 := \text{op_year}$ dan $x_a := \text{axnod_type}$.

```
cox_ph1 <- coxph(Surv(age, surv_status=="2") ~ op_year + axnod_type, data = haberman)
summary(cox_ph1)
```

```
## Call:
## coxph(formula = Surv(age, surv_status == "2") ~ op_year + axnod_type,
##       data = haberman)
##
##      n= 306, number of events= 81
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## op_year      -0.02573   0.97459  0.03490 -0.737   0.461
## axnod_type    1.25775   3.51750  0.24903  5.051 4.41e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## op_year           0.9746      1.0261    0.9102    1.044
## axnod_type        3.5175      0.2843    2.1590    5.731
##
## Concordance= 0.569 (se = 0.041 )
## Likelihood ratio test= 21.57  on 2 df,  p=2e-05
## Wald test            = 25.79  on 2 df,  p=3e-06
## Score (logrank) test = 29.24  on 2 df,  p=4e-07
```

Terlihat bahwa pada pasien yang memiliki jumlah nodus aksila ≥ 10 ($x_a = 1$), hazard untuk mengalami kematian adalah 3.51750 kali dibandingkan dengan hazard untuk pasien yang memiliki jumlah nodus

aksila < 10 ($x_a = 0$). Hal ini mengimplikasikan bahwa pasien dengan jumlah nodus aksila < 10 memiliki kondisi yang lebih baik, dalam hal berisiko lebih rendah untuk mengalami kematian dalam waktu 5 tahun setelah operasi. Variabel `op_year` tidak signifikan karena $p\text{-value} = 0.461 > \alpha = 0.05$. Secara keseluruhan, model ini cukup baik dalam menjelaskan kondisi pasien yang menjalani operasi kanker payudara. Terlihat dari uji model, $p\text{-value}$ untuk uji rasio likelihood adalah $2e-05$, lebih kecil dari taraf signifikansi 0.05. Hal ini mengimplikasikan bahwa minimal salah satu dari variabel `year` atau `axnod_type` dapat menjelaskan lama waktu hingga pasien mengalami kematian. Namun, karena variabel `year` tidak signifikan, dugaan sementara hanya variabel `axnod_type` saja yang signifikan dalam menjelaskan lama waktu hingga pasien mengalami kematian.

Model Kedua

Model regresi Cox-PH untuk data ini yaitu sebagai berikut.

$$h(t, \mathbf{x}) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2)$$

dengan $x_1 := \text{op_year}$ dan $x_2 := \text{axil_nodes}$.

```
cox_ph2 <- coxph(Surv(age, surv_status=="2") ~ op_year + axil_nodes, data = haberman)
summary(cox_ph2)
```

```
## Call:
## coxph(formula = Surv(age, surv_status == "2") ~ op_year + axil_nodes,
##       data = haberman)
##
##      n= 306, number of events= 81
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## op_year      -0.02801   0.97238  0.03474 -0.806    0.42
## axil_nodes    0.05517   1.05672  0.01009  5.469 4.52e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
```

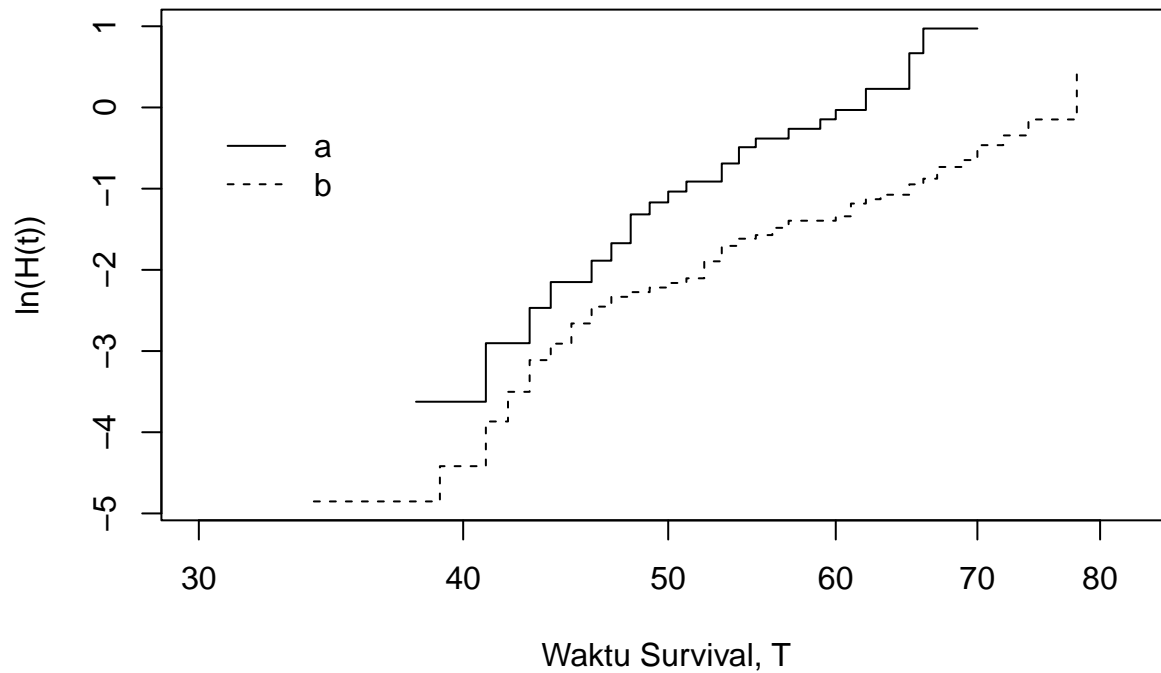
```
## op_year      0.9724      1.0284      0.9084      1.041
## axil_nodes    1.0567      0.9463      1.0360      1.078
##
## Concordance= 0.631 (se = 0.041 )
## Likelihood ratio test= 21.81 on 2 df, p=2e-05
## Wald test      = 30.08 on 2 df, p=3e-07
## Score (logrank) test = 32.88 on 2 df, p=7e-08
```

Terlihat bahwa variabel `axil_nodes` signifikan karena $p\text{-value} = 4.52e-08 < \alpha = 0.05$. Sama seperti model pertama, variabel `op_year` dalam model kedua juga tidak signifikan karena $p\text{-value} = 0.42 > \alpha = 0.05$. Secara keseluruhan, model ini cukup baik dalam menjelaskan kondisi pasien yang menjalani operasi kanker payudara karena $p\text{-value}$ untuk uji rasio likelihood adalah $2e-05$, lebih kecil dari taraf signifikansi 0.05.

Model Final

Karena dalam kedua model variabel `op_year` tidak signifikan, maka variabel `year` tidak akan dimasukkan dalam model final. Akan dilakukan pengecekan asumsi *proportional hazard* untuk variabel `axnod_type`.

```
haberman$axnod_type <- ifelse(haberman$axnod_type==1, "a", "b")
haberman$axnod_type <- as.factor(haberman$axnod_type)
plot(survfit(Surv(age, surv_status=="2") ~ axnod_type, data = haberman),
     fun = "cloglog", lty = 1:2, mark.time = FALSE,
     xlab = "Waktu Survival, T",
     ylab = "ln(H(t))")
legend(30, 0, legend = c("a", "b"), lty = 1:2, bty = "n")
```



Karena plot antara $\ln(H_a(t))$ dan $\ln(H_b(t))$ tidak paralel, asumsi PH untuk variabel `axnod_type` tidak terpenuhi. Selain itu, membuat variabel numerik menjadi variabel kategorik akan mengurangi kualitas informasi yang sesungguhnya. Oleh karena itu, untuk model final akan digunakan variabel `axil_nodes` dibandingkan dengan `axnod_type`. Model regresi Cox-PH final untuk kasus ini yaitu sebagai berikut.

$$h(t, x_1) = h_0(t) \exp(\beta_1 x_1) \quad (1)$$

dengan $x_1 := \text{axil_nodes}$.

```
cox_ph <- coxph(Surv(age, surv_status=="2") ~ axil_nodes, data = haberman)
summary(cox_ph)
```

```
## Call:
```

```
## coxph(formula = Surv(age, surv_status == "2") ~ axil_nodes, data = haberman)
```

```
##
```

```
## n= 306, number of events= 81
```

```
##
```

```
##               coef exp(coef) se(coef)      z Pr(>|z|)
## axil_nodes 0.05462   1.05614  0.01005 5.437 5.42e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## axil_nodes      1.056      0.9468      1.036      1.077
##
## Concordance= 0.657 (se = 0.038 )
## Likelihood ratio test= 21.16 on 1 df,  p=4e-06
## Wald test              = 29.56 on 1 df,  p=5e-08
## Score (logrank) test = 32.35 on 1 df,  p=1e-08
```

Didapatkan estimasi parameter β_1 yaitu $\hat{\beta}_1 = 0.05462$, sehingga taksiran untuk model regresi Cox-PH ini yaitu

$$\hat{h}(t, x_1) = h_0(t) \exp(0.05462 \times x_1). \quad (2)$$

Terlihat bahwa $\exp(0.05462) = 1.05614 > 1$ yang berarti semakin banyak jumlah nodus aksila yang terdeteksi positif, maka risiko untuk mengalami kematian pada pasien kanker payudara meningkat.

3.3 Pengecekan Asumsi *Proportional Hazard* untuk Model Final

Akan dilakukan pengujian asumsi *proportional hazard* untuk model pada persamaan (2).

H_0 : hazard proporsional

H_1 : hazard tidak proporsional

```
test.ph <- cox.zph(cox_ph)
test.ph
```

```
##               chisq df      p
## axil_nodes   3.41  1 0.065
## GLOBAL       3.41  1 0.065
```

Dari hasil tersebut, didapatkan $p - value = 0.065 > \alpha = 0.05$ yang berarti H_0 tidak ditolak. Oleh karena itu, **axil_nodes** memenuhi asumsi *proportional hazard* signifikan secara statistik.

PENUTUP

4.1 Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan menggunakan data kelangsungan hidup pasien kanker payudara yang telah menjalani operasi antara tahun 1958 sampai 1970 di Rumah Sakit Billings Universitas Chicago, didapatkan kesimpulan sebagai berikut.

1. Model Cox-Proportional Hazard terbaik untuk data kelangsungan hidup pasien kanker payudara yang telah menjalani operasi antara tahun 1958 sampai 1970 di Rumah Sakit Billings Universitas Chicago memiliki persamaan sebagai berikut.

$$\hat{h}(t, x_1) = h_0(t) \exp(0.05462 \times x_1)$$

di mana x_1 adalah jumlah banyaknya nodus aksila positif terdeteksi pada pasien kanker payudara setelah menjalani operasi. Dari taksiran fungsi hazard tersebut, terlihat bahwa memiliki banyak nodus aksila, menunjukkan peningkatan risiko kematian pada penderita kanker payudara.

2. Berdasarkan tipe nodus aksila sesuai jumlah nodus aksila yang terdeteksi (< 10 atau ≥ 10) dan tahun dilakukannya operasi, nilai hazard ratio nodus aksila sebesar $\exp(1.25775) = 3.51750$ yang artinya pasien kanker payudara yang memiliki jumlah nodus aksila positif terdeteksi ≥ 10 setelah menjalani operasi memiliki kemungkinan 3.51750 kali untuk mengalami kematian dalam waktu 5 tahun dibandingkan pasien kanker payudara yang memiliki jumlah nodus aksila positif terdeteksi < 10 .

DAFTAR PUSTAKA

- [1] Abdullah, S. (2022). *Analisis Survival: Konsep dan Aplikasi dengan R*. Bumi Aksara.
- [2] cancer.gov. *Definition of Cancer Terms*. Diakses pada 13 Juni 2023, dari <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/axillary-lymph-node>
- [3] Dukalang, Hendra. (2019). *Analisis Regresi COX Proportional Hazard pada Pemodelan Waktu Tunggu Mendapatkan Pekerjaan*. Jambura Jurnal of Mathematics: Volume 1 No. 1. Diakses dari <https://ejurnal.ung.ac.id/index.php/jjom/article/view/1744/1271>
- [4] en.wikipedia.org. *Axillary lymph nodes*. Diakses pada 13 Juni 2023, dari https://en.wikipedia.org/wiki/Axillary_lymph_nodes
- [5] Klein, J. P. and Moeschberger, M. L. (2005). *Survival Analysis: Techniques for Censored and Truncated Data*. Springer Science & Business Media.
- [6] who.int. (2021). *Breast cancer*. Diakses dari <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>

Link File

- archive.ics.uci.edu. Haberman's Survival. Data diambil dari *link* berikut. <https://archive.ics.uci.edu/dataset/43/haberman+s+survival>
- File *project*, R, dan lain-lain dapat diakses pada *link* berikut. https://drive.google.com/drive/folders/1l9mhIL5OWozdyxS8p_r32k7_7OfgMiZU?usp=drive_link

LAMPIRAN

Lampiran I. Data Penelitian

OperationAxillary Survival				OperationAxillary Survival				OperationAxillary Survival			
Age	Year	Nodes	Status	Age	Year	Nodes	Status	Age	Year	Nodes	Status
30	64	1	1	47	67	0	1	57	63	0	1
30	62	3	1	47	58	3	1	57	64	0	1
30	65	0	1	47	60	4	1	57	64	0	1
31	59	2	1	47	68	4	1	57	67	0	1
31	65	4	1	47	66	12	1	58	59	0	1
33	58	10	1	48	58	11	2	58	60	3	1
33	60	0	1	48	58	11	2	58	61	1	1
34	59	0	2	48	67	7	2	58	67	0	1
34	66	9	2	48	61	8	1	58	58	0	1
34	58	30	1	48	62	2	1	58	58	3	1
34	60	1	1	48	64	0	1	58	61	2	1
34	61	10	1	48	66	0	1	59	62	35	2
34	67	7	1	49	63	0	2	59	60	0	1
34	60	0	1	49	64	10	2	59	63	0	1
35	64	13	1	49	61	1	1	59	64	1	1
35	63	0	1	49	62	0	1	59	64	4	1
36	60	1	1	49	66	0	1	59	64	0	1
36	69	0	1	49	60	1	1	59	64	7	1
37	60	0	1	49	62	1	1	59	67	3	1
37	63	0	1	49	63	3	1	60	59	17	2
37	58	0	1	49	61	0	1	60	65	0	2
37	59	6	1	49	67	1	1	60	61	1	1
37	60	15	1	50	63	13	2	60	67	2	1
37	63	0	1	50	64	0	2	60	61	25	1
38	69	21	2	50	59	0	1	60	64	0	1
38	59	2	1	50	61	6	1	61	62	5	2
38	60	0	1	50	61	0	1	61	65	0	2
38	60	0	1	50	63	1	1	61	68	1	2

OperationAxillary Survival				OperationAxillary Survival				OperationAxillary Survival			
Age	Year	Nodes	Status	Age	Year	Nodes	Status	Age	Year	Nodes	Status
38	62	3	1	50	58	1	1	61	59	0	1
38	64	1	1	50	59	2	1	61	59	0	1
38	66	0	1	50	61	0	1	61	64	0	1
38	66	11	1	50	64	0	1	61	65	8	1
38	60	1	1	50	65	4	1	61	68	0	1
38	67	5	1	50	66	1	1	61	59	0	1
39	66	0	2	51	59	13	2	62	59	13	2
39	63	0	1	51	59	3	2	62	58	0	2
39	67	0	1	51	64	7	1	62	65	19	2
39	58	0	1	51	59	1	1	62	62	6	1
39	59	2	1	51	65	0	1	62	66	0	1
39	63	4	1	51	66	1	1	62	66	0	1
40	58	2	1	52	69	3	2	62	58	0	1
40	58	0	1	52	59	2	2	63	60	1	2
40	65	0	1	52	62	3	2	63	61	0	1
41	60	23	2	52	66	4	2	63	62	0	1
41	64	0	2	52	61	0	1	63	63	0	1
41	67	0	2	52	63	4	1	63	63	0	1
41	58	0	1	52	69	0	1	63	66	0	1
41	59	8	1	52	60	4	1	63	61	9	1
41	59	0	1	52	60	5	1	63	61	28	1
41	64	0	1	52	62	0	1	64	58	0	1
41	69	8	1	52	62	1	1	64	65	22	1
41	65	0	1	52	64	0	1	64	66	0	1
41	65	0	1	52	65	0	1	64	61	0	1
42	69	1	2	52	68	0	1	64	68	0	1
42	59	0	2	53	58	4	2	65	58	0	2
42	58	0	1	53	65	1	2	65	61	2	2
42	60	1	1	53	59	3	2	65	62	22	2
42	59	2	1	53	60	9	2	65	66	15	2
42	61	4	1	53	63	24	2	65	58	0	1

OperationAxillary Survival				OperationAxillary Survival				OperationAxillary Survival			
Age	Year	Nodes	Status	Age	Year	Nodes	Status	Age	Year	Nodes	Status
42	62	20	1	53	65	12	2	65	64	0	1
42	65	0	1	53	58	1	1	65	67	0	1
42	63	1	1	53	60	1	1	65	59	2	1
43	58	52	2	53	60	2	1	65	64	0	1
43	59	2	2	53	61	1	1	65	67	1	1
43	64	0	2	53	63	0	1	66	58	0	2
43	64	0	2	54	60	11	2	66	61	13	2
43	63	14	1	54	65	23	2	66	58	0	1
43	64	2	1	54	65	5	2	66	58	1	1
43	64	3	1	54	68	7	2	66	68	0	1
43	60	0	1	54	59	7	1	67	64	8	2
43	63	2	1	54	60	3	1	67	63	1	2
43	65	0	1	54	66	0	1	67	66	0	1
43	66	4	1	54	67	46	1	67	66	0	1
44	64	6	2	54	62	0	1	67	61	0	1
44	58	9	2	54	69	7	1	67	65	0	1
44	63	19	2	54	63	19	1	68	67	0	1
44	61	0	1	54	58	1	1	68	68	0	1
44	63	1	1	54	62	0	1	69	67	8	2
44	61	0	1	55	63	6	2	69	60	0	1
44	67	16	1	55	68	15	2	69	65	0	1
45	65	6	2	55	58	1	1	69	66	0	1
45	66	0	2	55	58	0	1	70	58	0	2
45	67	1	2	55	58	1	1	70	58	4	2
45	60	0	1	55	66	18	1	70	66	14	1
45	67	0	1	55	66	0	1	70	67	0	1
45	59	14	1	55	69	3	1	70	68	0	1
45	64	0	1	55	69	22	1	70	59	8	1
45	68	0	1	55	67	1	1	70	63	0	1
45	67	1	1	56	65	9	2	71	68	2	1
46	58	2	2	56	66	3	2	72	63	0	2

OperationAxillary Survival				OperationAxillary Survival				OperationAxillary Survival			
Age	Year	Nodes	Status	Age	Year	Nodes	Status	Age	Year	Nodes	Status
46	69	3	2	56	60	0	1	72	58	0	1
46	62	5	2	56	66	2	1	72	64	0	1
46	65	20	2	56	66	1	1	72	67	3	1
46	62	0	1	56	67	0	1	73	62	0	1
46	58	3	1	56	60	0	1	73	68	0	1
46	63	0	1	57	61	5	2	74	65	3	2
47	63	23	2	57	62	14	2	74	63	0	1
47	62	0	2	57	64	1	2	75	62	1	1
47	65	0	2	57	64	9	1	76	67	0	1
47	61	0	1	57	69	0	1	77	65	3	1
47	63	6	1	57	61	0	1	78	65	1	2
47	66	0	1	57	62	0	1	83	58	2	2

PERSENTASE KONTRIBUSI KELOMPOK

No.	Nama	NPM	Kontribusi	Tingkat Kontribusi
1.	Azizah Annurohmah	2006572983	Mengerjakan project	100%
2.	Diki Wahyudi	2106709131	Mengerjakan project	100%
3.	Farah Ramadhani Putri	2106654832	Mengerjakan project	100%