

TUGAS KELOMPOK
CASE STUDY COX REGRESSION



Anggota Kelompok 4 :

Divaya Syifa Susilobudi – 2106650790

Favian Sulthan Wafi – 2106706205

Wulan Akhsah – 2106637100

Model Survival B

DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS INDONESIA
TAHUN AJARAN 2022/2023

DAFTAR ISI

| | |
|--|----|
| DAFTAR ISI | i |
| Tabel Kontribusi | ii |
| BAB 1 Pendahuluan | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Tujuan | 2 |
| 1.3 Rumusan Masalah | 2 |
| 1.4 Informasi Data | 2 |
| BAB 2 Hasil dan Pembahasan | 4 |
| 2.1 <i>Preprocessing</i> dan Analisis Deskriptif | 4 |
| 2.1.1 <i>Preprocessing</i> | 4 |
| 2.1.2 Analisis Deskriptif | 5 |
| 2.2 Uji K-Sampel | 11 |
| 2.3 Analisis Model Regresi Cox-PH | 11 |
| 2.4 Pengecekan Asumsi Proportional Hazard(PH) | 13 |
| BAB 3 Penutup | 14 |
| 3.1 Kesimpulan | 14 |

Tabel Kontribusi

| Anggota | Jenis Kontribusi | Persentase Kontribusi |
|---|--|------------------------------|
| Divaya Syifa Susilobudi – 2106650790 | Aktif terlibat dalam diskusi. Membuat latar belakang, tujuan, rumusan masalah, dan informasi data, serta membuat format laporan | 100% |
| Favian Sulthan Wafi – 2106706205 | Aktif terlibat dalam diskusi. Mencari dataset. Membuat hasil dan pembahasan bagian uji hipotesis k-sampel dan analisis model regresi cox-PH, serta membuat kesimpulan dan daftar pustaka | 100% |
| Wulan Akhsah – 2106637100 | Aktif terlibat dalam diskusi. Membuat hasil dan pembahasan bagian preprocessing dan analisis deskriptif, serta merapikan format laporan | 100% |

BAB 1

Pendahuluan

1.1 Latar Belakang

Primary biliary cirrhosis (PBC) adalah penyakit hati autoimun kronis yang secara utama mempengaruhi saluran empedu kecil di dalam hati. Penyakit ini ditandai dengan kerusakan progresif pada saluran empedu tersebut, menyebabkan gangguan aliran empedu, peradangan, dan akhirnya sirosis hati. Gejalanya dapat meliputi kelelahan, gatal-gatal, kuning pada kulit dan mata (jaundice), dan nyeri perut. PBC lebih sering terjadi pada wanita di usia pertengahan. Gejalanya bisa berupa kelelahan, gatal-gatal, kulit dan mata kuning, serta nyeri perut. Prevalensi PBC di seluruh dunia berkisar antara 0,5 hingga 40 kasus per 100.000 orang, namun angka ini dapat bervariasi di berbagai wilayah dan kelompok populasi.

Analisis survival/analisis data ketahanan hidup adalah suatu metode yang berhubungan dengan waktu suatu individu / subjek mulai dari awal pengamatan sampai terjadinya kejadian. Kejadian yang diamati seperti kematian, timbul atau kambuhnya penyakit, rusak atau kembali bekerjanya sebuah mesin, dan kejadian pada individu yang berhenti atau mendapatkan pekerjaan (Kleinbaum & Klein, 2005: 4). Salah satu contoh analisis survival dalam bidang kesehatan adalah untuk menguji perbedaan resiko pada pasien kelompok tertentu dengan kelompok lainnya.

Terdapat beberapa cara untuk menganalisis data survival. Salah satunya adalah menggunakan metode parametrik, metode non-parametrik, dan metode semiparametrik. Metode parametrik mengasumsikan bahwa waktu survival mengikuti distribusi tertentu seperti Weibull, gamma, atau eksponensial. Metode non-parametrik digunakan jika data tidak mengikuti distribusi tertentu. Selain itu, analisis regresi juga dapat digunakan untuk memahami data survival. Regresi adalah metode statistik yang mempelajari hubungan antara beberapa variabel. Regresi Cox adalah salah satu metode regresi yang populer dalam analisis data survival. Metode ini merupakan metode semiparametrik di mana model hazard dasar mengikuti pola nonparametrik, sementara variabel independen mengikuti model parametrik. Tujuan dari regresi Cox adalah memahami hubungan antara waktu survival dan variabel-variabel yang mempengaruhinya. Regresi Cox juga dikenal sebagai

regresi Cox Proporsional Hazard karena mengasumsikan tingkat kematian tetap seiring berjalannya waktu, yang merupakan asumsi penting dalam metode ini.

Pada kasus ini, kami akan menggunakan metode Cox Proportional Hazard untuk mengetahui apakah ada perbedaan resiko kematian penderita PBC pada stadium yang berbeda.

1.2 Tujuan

Untuk mengetahui perbedaan risiko kematian penderita PBC dengan stadium yang berbeda menggunakan model regresi Cox-PH.

1.3 Rumusan Masalah

1. Apakah ada perbedaan risiko kematian penderita PBC pada stadium 1, 2, 3, dan 4?
2. Apakah risiko kematian penderita PBC meningkat seiring meningkatnya stadium?

1.4 Informasi Data

Data yang akan kami gunakan adalah dataset Primary Biliary Cirrhosis (PBC). Sumber dataset akan kami lampirkan di daftar pusaka. Berikut adalah detail mengenai variabel yang akan kami gunakan.

- **age:** usia dalam tahun
- **albumin:** serum albumin (g/dl)
- **alk.phos:** alkalin fosfatase (U/liter)
- **ascites:** keberadaan asites
- **ast:** aspartat aminotransferase (U/ml)
- **bili:** serum bilirunbin (mg/dl)
- **chol:** serum kolesterol (mg/dl)
- **copper:** tembaga dalam urine (ug/day)
- **edema:**
 - 0 = tidak ada edema
 - 0,5 = tidak diobati atau berhasil diobati
 - 1 = edema meskipun terapi diuretik
- **hepato:** keberadaan hepatomegali atau pembesaran hati
- **id:** nomor kasus

- **platelet:** jumlah platelet
- **protime:** waktu darah menggumpal yang terstandardisasi
- **sex:** m/f
- **spiders:** malformasi pembuluh darah di kulit
- **stage:** stadium histologis penyakit (butuh biopsy)
- **status:** status di titik akhir
 - 0 = tersensor
 - 1 = transplantasi
 - 2 = kematian
- **time** = jumlah hari antara pendaftaran dan awal kematian, transplantasi, atau analisis studi pada Juli 1986
- **trt:**
 - 1 = D-penicilmain
 - 2 = plasebo
 - NA = tidak dirandomisasi
- **trig:** trigliserida (mg/dl)

BAB 2

Hasil dan Pembahasan

2.1 *Preprocessing* dan Analisis Deskriptif

2.1.1 *Preprocessing*

Sebelum melakukan uji, akan dilakukan *preprocessing* pada data. Pertama dilakukan pemasukkan data setelah mengunduh dalam bentuk file .csv, kemudian dimuatkan ke dalam variabel “pbc”. Setelah itu, akan dilihat dimensi dari data untuk melihat jumlah kolom dan baris, serta akan ditampilkan tipe data sebagai berikut.

```
> pbc <- read.csv("C:/Users/ASUS/Downloads/pbc.csv") #mengimport dataset melalui device
> dim(pbc)
[1] 419 20
```

Ditampilkan enam contoh data:

```
> head(pbc)
  id time status trt      age sex ascites hepato spiders edema bili chol albumin copper alk.phos  ast trig
1  1  400      2   1 58.76523 f      1      1      1  1.0 14.5 261    2.60   156  1718.0 137.95 172
2  2 4500      0   1 56.44627 f      0      1      1  0.0  1.1 302    4.14    54   7394.8 113.52  88
3  3 1012      2   1 70.07255 m      0      0      0  0.5  1.4 176    3.48   210    516.0  96.10  55
4  4 1925      2   1 54.74059 f      0      1      1  0.5  1.8 244    2.54    64   6121.8  60.63  92
5  5 1504      1   2 38.10541 f      0      1      1  0.0  3.4 279    3.53   143    671.0 113.15  72
6  6 2503      2   2 66.25873 f      0      1      0  0.0  0.8 248    3.98    50    944.0  93.00  63
  platelet protime stage
1     190    12.2      4
2     221    10.6      3
3     151    12.0      4
4     183    10.3      4
5     136    10.9      3
6      NA    11.0      3
```

Akan dilihat apakah ada *missing values* pada kolom yang dianalisis.

```
> sum(is.na(pbc$stage)) ; sum(is.na(pbc$status)) ; sum(is.na(pbc$time))
[1] 6
[1] 0
[1] 0
```

Berdasarkan *output* tersebut terdapat *missing value*, maka akan dihapus terlebih dahulu dan dicek kembali apakah masih ada *missing value*.

```
> pbc <- pbc[!(is.na(pbc$stage)), ] ; pbc <- pbc[!(is.na(pbc$stage)), ] ; pbc <- pbc[!(is.na(pbc$stage)), ]
> sum(is.na(pbc$stage)) ; sum(is.na(pbc$status)) ; sum(is.na(pbc$time))
[1] 0
[1] 0
[1] 0
```

Karena sudah tidak ada, maka preprossessing dilanjutkan. Untuk memudahkan perhitungan, waktu akan diubah dari hari menjadi tahun. Kemudian, untuk data status *transplant* dijadikan tersensor karena pasien masih hidup setelah mendapatkan *transplant*.

```

> pbc$time <- pbc$time / 365
> pbc$status[pbc$status == 1] <- 0
> pbc$status[pbc$status == 2] <- 1
> table(pbc$status)

 0  1
256 157
> attach(pbc)
The following objects are masked from pbc (pos = 3):

  age, albumin, alk.phos, ascites, ast, bili, chol, copper, edema, hepato, id, platelet, protime,
  sex, spiders, stage, status, time, trig, trt

```

Fungsi attach digunakan untuk memanggil variabel-variabel pbc, tanpa perlu memanggil pbc.

2.1.2 Analisis Deskriptif

Setelah melakukan *preprocessing*, akan dibuat analisis deskriptifnya. Pada subbab ini akan dilihat perbedaan survival experience dari setiap stage pbc dan akan dilihat apakah ada tren dari fungsi hazard tiap stagenya. Pertama akan dilihat tipe data dan ringkasan dari *dataframe* pbc.

```

> str(pbc)
'data.frame':  413 obs. of  20 variables:
 $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ time    : num  1.1 12.33 2.77 5.27 4.12 ...
 $ status  : num  1 0 1 1 0 1 0 1 1 1 ...
 $ trt     : int  1 1 1 1 2 2 2 2 1 2 ...
 $ age     : num  58.8 56.4 70.1 54.7 38.1 ...
 $ sex     : chr  "f" "f" "m" "f" ...
 $ ascites : int  1 0 0 0 0 0 0 0 0 1 ...
 $ hepato  : int  1 1 0 1 1 1 1 0 0 0 ...
 $ spiders : int  1 1 0 1 1 0 0 0 1 1 ...
 $ edema   : num  1 0 0.5 0.5 0 0 0 0 0 1 ...
 $ bili    : num  14.5 1.1 1.4 1.8 3.4 0.8 1 0.3 3.2 12.6 ...
 $ chol    : int  261 302 176 244 279 248 322 280 562 200 ...
 $ albumin : num  2.6 4.14 3.48 2.54 3.53 3.98 4.09 4 3.08 2.74 ...
 $ copper  : int  156 54 210 64 143 50 52 52 79 140 ...
 $ alk.phos: num  1718 7395 516 6122 671 ...
 $ ast     : num  137.9 113.5 96.1 60.6 113.2 ...
 $ trig    : int  172 88 55 92 72 63 213 189 88 143 ...
 $ platelet: int  190 221 151 183 136 NA 204 373 251 302 ...
 $ protime : num  12.2 10.6 12 10.3 10.9 11 9.7 11 11 11.5 ...
 $ stage   : int  4 3 4 4 3 3 3 3 2 4 ...

```



```
> summary(pbc)
      id      time      status      trt      age      sex
Min.   : 1.0   Min.   : 0.1123   Min.   :0.0000   Min.   :1.000   Min.   :26.28   Length:413
1st Qu.:104.0   1st Qu.: 2.9918   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:42.74   Class :character
Median :207.0   Median : 4.6630   Median :0.0000   Median :1.000   Median :51.00   Mode  :character
Mean   :208.3   Mean   : 5.2454   Mean   :0.3801   Mean   :1.494   Mean   :50.65
3rd Qu.:310.0   3rd Qu.: 7.1479   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:58.00
Max.   :419.0   Max.   :13.1370   Max.   :1.0000   Max.   :2.000   Max.   :78.44
      NA's :101

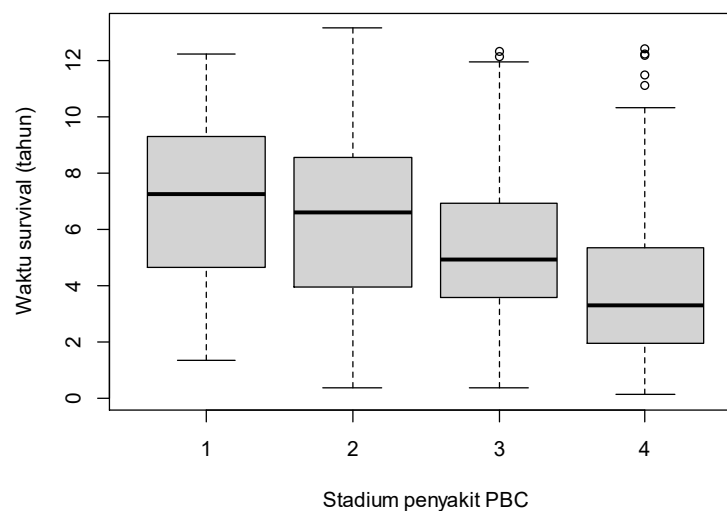
      ascites      hepato      spiders      edema      bili      chol
Min.   :0.000000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   : 0.300   Min.   :120.0
1st Qu.:0.000000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 0.800   1st Qu.:249.5
Median :0.000000   Median :1.0000   Median :0.0000   Median :0.0000   Median :1.400   Median :309.5
Mean   :0.07692   Mean   :0.5128   Mean   :0.2885   Mean   :0.1017   Mean   :3.221   Mean   :369.5
3rd Qu.:0.000000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.: 3.400   3rd Qu.:400.0
Max.   :1.000000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :28.000   Max.   :1775.0
      NA's :101      NA's :101      NA's :101      NA's :101      NA's :129      NA's :129

      albumin      copper      alk.phos      ast      trig      platelet
Min.   :1.96   Min.   : 4.00   Min.   :289.0   Min.   :26.35   Min.   :33.00   Min.   :62.0
1st Qu.:3.25   1st Qu.:41.25   1st Qu.:871.5   1st Qu.:80.60   1st Qu.:84.25   1st Qu.:188.2
Median :3.53   Median :73.00   Median :1259.0   Median :114.70   Median :108.00   Median :249.0
Mean   :3.50   Mean   :97.65   Mean   :1982.7   Mean   :122.56   Mean   :124.70   Mean   :255.9
3rd Qu.:3.78   3rd Qu.:123.00   3rd Qu.:1980.0   3rd Qu.:151.90   3rd Qu.:151.00   3rd Qu.:317.8
Max.   :4.64   Max.   :588.00   Max.   :13862.4   Max.   :457.25   Max.   :598.00   Max.   :563.0
      NA's :103      NA's :101      NA's :101      NA's :101      NA's :131      NA's :11

      protime      stage
Min.   : 9.00   Min.   :1.000
1st Qu.:10.00   1st Qu.:2.000
Median :10.60   Median :3.000
Mean   :10.74   Mean   :3.024
3rd Qu.:11.10   3rd Qu.:4.000
Max.   :18.00   Max.   :4.000
      NA's :2
```

Selanjutnya akan ditetapkan bahwa hanya akan digunakan stage sebagai variabel penjelas untuk waktu hingga terjadi kematian. Akan dicek boxplot dan perbandingan pasien yang hidup dan meninggal dari setiap stage.

```
> boxplot(time~stage,data=pbc,xlab="stadium penyakit PBC", ylab="waktu survival (tahun)")
```



Berdasarkan boxplot, terlihat bahwa terdapat kecenderungan pasien dengan tingkat stage (stadium) tinggi memiliki waktu survival yang lebih pendek daripada tingkat stage

yang lebih rendah. Seperti pada stage 4, memiliki waktu survival yang lebih pendek dibandingkan pasien pada tiga stage yang lainnya.

```
> table(pbc$status,pbc$stage)
```

```
      1  2  3  4
0  19 69 108 60
1   2 23  48 84
```

Berdasarkan tabel perbandingan, jumlah pasien yang masih bertahan hidup dan yang meninggal di masing-masing stage, diperoleh 2 dari 21 pasien stage 1 meninggal, 23 dari 92 pasien stage 2 meninggal, 48 dari 156 pasien stage 3 meninggal, dan untuk stage 4 terdapat 84 dari 144 pasien meninggal.

Kemudian akan dibuat estimasi fungsi survival dengan menggunakan Kaplan-Meier dan dibuat visualisasi sebagai berikut.

```
> st.stage <- survfit(Surv(time,status)~stage,data=pbc, conf.type="log-log")
> summary(st.stage)
call: survfit(formula = Surv(time, status) ~ stage, data = pbc, conf.type = "log-log")
```

```
      stage=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
1.34    21      1   0.952  0.0465   0.707   0.993
6.54    13      1   0.879  0.0824   0.585   0.970

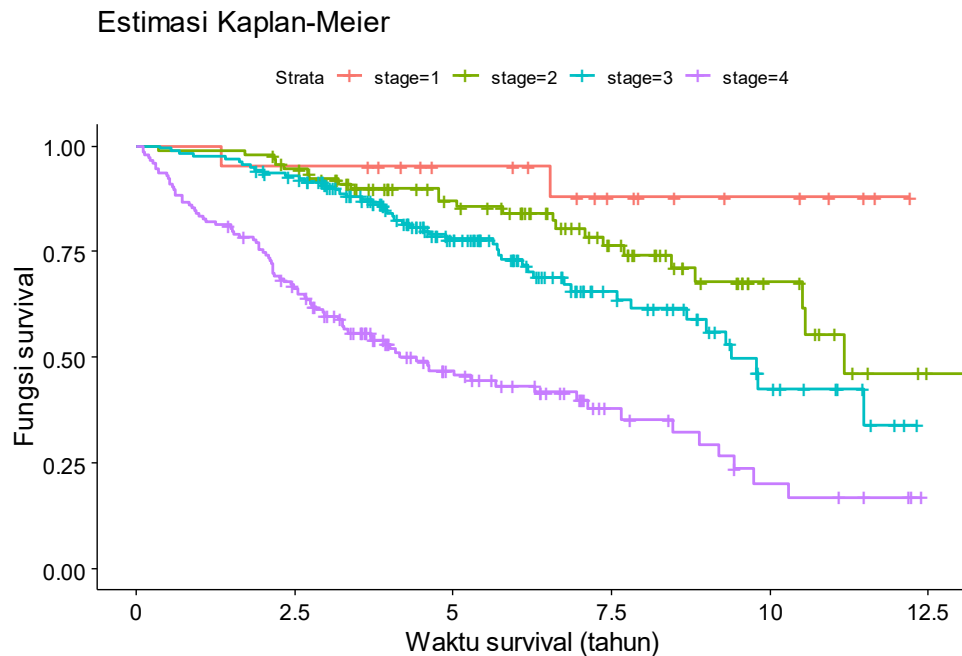
      stage=2
time n.risk n.event survival std.err lower 95% CI upper 95% CI
0.356    92      1   0.989  0.0108   0.925   0.998
1.712    91      1   0.978  0.0152   0.916   0.995
2.167    89      1   0.967  0.0186   0.902   0.989
2.189    88      1   0.956  0.0214   0.888   0.983
2.337    86      1   0.945  0.0238   0.873   0.977
2.712    84      1   0.934  0.0261   0.859   0.970
2.737    82      1   0.923  0.0281   0.844   0.962
3.200    78      1   0.911  0.0302   0.829   0.954
3.384    75      1   0.899  0.0321   0.814   0.946
4.770    64      1   0.885  0.0345   0.795   0.936
4.784    63      1   0.870  0.0367   0.777   0.926
5.060    61      1   0.856  0.0388   0.759   0.916
5.784    57      1   0.841  0.0409   0.741   0.905
6.575    47      1   0.823  0.0438   0.717   0.892
6.627    46      1   0.805  0.0464   0.695   0.879
7.077    41      1   0.786  0.0492   0.670   0.865
7.367    38      1   0.765  0.0521   0.644   0.850
7.704    34      1   0.743  0.0552   0.615   0.833
8.455    25      1   0.713  0.0605   0.575   0.813
8.827    21      1   0.679  0.0664   0.530   0.790
10.518   11      1   0.617  0.0843   0.431   0.758
10.556    10      1   0.555  0.0958   0.352   0.718
11.175     6      1   0.463  0.1163   0.232   0.666
```

| stage=3 | | | | | | | | |
|---------|--------|---------|----------|---------|-------|--------|-------|--------|
| time | n.risk | n.event | survival | std.err | lower | 95% CI | upper | 95% CI |
| 0.384 | 156 | 1 | 0.994 | 0.00639 | | 0.955 | | 0.999 |
| 0.542 | 155 | 1 | 0.987 | 0.00901 | | 0.950 | | 0.997 |
| 0.682 | 154 | 1 | 0.981 | 0.01100 | | 0.942 | | 0.994 |
| 0.893 | 153 | 1 | 0.974 | 0.01266 | | 0.933 | | 0.990 |
| 1.411 | 152 | 1 | 0.968 | 0.01410 | | 0.925 | | 0.987 |
| 1.636 | 151 | 1 | 0.962 | 0.01540 | | 0.916 | | 0.983 |
| 1.674 | 150 | 1 | 0.955 | 0.01658 | | 0.908 | | 0.978 |
| 1.814 | 149 | 1 | 0.949 | 0.01766 | | 0.900 | | 0.974 |
| 1.866 | 148 | 1 | 0.942 | 0.01867 | | 0.892 | | 0.970 |
| 2.008 | 145 | 1 | 0.936 | 0.01964 | | 0.884 | | 0.965 |
| 2.353 | 143 | 1 | 0.929 | 0.02056 | | 0.876 | | 0.960 |
| 2.584 | 141 | 1 | 0.923 | 0.02145 | | 0.868 | | 0.955 |
| 2.668 | 139 | 1 | 0.916 | 0.02230 | | 0.860 | | 0.950 |
| 2.959 | 133 | 1 | 0.909 | 0.02317 | | 0.851 | | 0.945 |
| 3.000 | 130 | 1 | 0.902 | 0.02402 | | 0.843 | | 0.940 |
| 3.156 | 125 | 1 | 0.895 | 0.02489 | | 0.834 | | 0.934 |
| 3.205 | 124 | 1 | 0.888 | 0.02571 | | 0.825 | | 0.929 |
| 3.321 | 123 | 1 | 0.881 | 0.02650 | | 0.817 | | 0.923 |
| 3.553 | 118 | 1 | 0.873 | 0.02730 | | 0.808 | | 0.917 |
| 3.726 | 114 | 1 | 0.865 | 0.02812 | | 0.799 | | 0.911 |
| 3.871 | 109 | 1 | 0.857 | 0.02896 | | 0.789 | | 0.905 |
| 3.910 | 106 | 1 | 0.849 | 0.02979 | | 0.780 | | 0.898 |
| 3.956 | 104 | 1 | 0.841 | 0.03061 | | 0.770 | | 0.892 |
| 4.049 | 103 | 1 | 0.833 | 0.03138 | | 0.761 | | 0.885 |
| 4.074 | 102 | 1 | 0.825 | 0.03212 | | 0.751 | | 0.878 |
| 4.208 | 100 | 1 | 0.817 | 0.03284 | | 0.742 | | 0.872 |
| 4.318 | 95 | 1 | 0.808 | 0.03360 | | 0.732 | | 0.865 |
| 4.540 | 88 | 1 | 0.799 | 0.03445 | | 0.721 | | 0.857 |
| 4.630 | 85 | 1 | 0.789 | 0.03530 | | 0.710 | | 0.849 |
| 4.893 | 80 | 1 | 0.780 | 0.03621 | | 0.698 | | 0.841 |
| 5.630 | 64 | 1 | 0.767 | 0.03764 | | 0.683 | | 0.832 |
| 5.701 | 63 | 1 | 0.755 | 0.03896 | | 0.669 | | 0.822 |
| 5.726 | 62 | 1 | 0.743 | 0.04020 | | 0.654 | | 0.812 |
| 5.767 | 61 | 1 | 0.731 | 0.04134 | | 0.640 | | 0.802 |
| 6.093 | 54 | 1 | 0.717 | 0.04273 | | 0.624 | | 0.792 |
| 6.181 | 52 | 1 | 0.704 | 0.04408 | | 0.607 | | 0.780 |
| 6.268 | 51 | 1 | 0.690 | 0.04532 | | 0.591 | | 0.769 |
| 6.756 | 43 | 1 | 0.674 | 0.04702 | | 0.572 | | 0.756 |
| 6.858 | 42 | 1 | 0.658 | 0.04856 | | 0.553 | | 0.743 |
| 7.586 | 32 | 1 | 0.637 | 0.05121 | | 0.528 | | 0.728 |
| 7.800 | 30 | 1 | 0.616 | 0.05372 | | 0.502 | | 0.711 |
| 8.685 | 24 | 1 | 0.590 | 0.05729 | | 0.469 | | 0.692 |
| 8.992 | 21 | 1 | 0.562 | 0.06107 | | 0.434 | | 0.672 |
| 9.301 | 18 | 1 | 0.531 | 0.06517 | | 0.396 | | 0.649 |
| 9.392 | 16 | 1 | 0.498 | 0.06903 | | 0.357 | | 0.623 |
| 9.792 | 15 | 1 | 0.465 | 0.07196 | | 0.320 | | 0.597 |
| 9.819 | 12 | 1 | 0.426 | 0.07566 | | 0.277 | | 0.567 |
| 11.482 | 5 | 1 | 0.341 | 0.09729 | | 0.163 | | 0.527 |

| stage=4 | | | | | | | | |
|---------|--------|---------|----------|---------|-------|--------|-------|--------|
| time | n.risk | n.event | survival | std.err | lower | 95% CI | upper | 95% CI |
| 0.112 | 144 | 1 | 0.993 | 0.00692 | | 0.9517 | | 0.999 |
| 0.118 | 143 | 1 | 0.986 | 0.00975 | | 0.9456 | | 0.997 |
| 0.140 | 142 | 1 | 0.979 | 0.01190 | | 0.9368 | | 0.993 |
| 0.195 | 141 | 1 | 0.972 | 0.01369 | | 0.9277 | | 0.989 |
| 0.211 | 140 | 1 | 0.965 | 0.01526 | | 0.9186 | | 0.985 |
| 0.258 | 139 | 1 | 0.958 | 0.01665 | | 0.9096 | | 0.981 |
| 0.301 | 138 | 1 | 0.951 | 0.01792 | | 0.9007 | | 0.977 |
| 0.304 | 137 | 1 | 0.944 | 0.01909 | | 0.8920 | | 0.972 |
| 0.359 | 136 | 1 | 0.938 | 0.02017 | | 0.8833 | | 0.967 |
| 0.490 | 135 | 1 | 0.931 | 0.02118 | | 0.8748 | | 0.962 |
| 0.510 | 134 | 1 | 0.924 | 0.02213 | | 0.8663 | | 0.957 |
| 0.523 | 133 | 1 | 0.917 | 0.02303 | | 0.8579 | | 0.952 |
| 0.529 | 132 | 1 | 0.910 | 0.02388 | | 0.8496 | | 0.947 |
| 0.567 | 131 | 1 | 0.903 | 0.02469 | | 0.8414 | | 0.941 |
| 0.592 | 130 | 1 | 0.896 | 0.02546 | | 0.8332 | | 0.936 |
| 0.605 | 129 | 1 | 0.889 | 0.02619 | | 0.8250 | | 0.930 |

| | | | | | | |
|--------|-----|---|-------|---------|--------|-------|
| 0.611 | 128 | 1 | 0.882 | 0.02689 | 0.8170 | 0.925 |
| 0.723 | 127 | 2 | 0.868 | 0.02820 | 0.8010 | 0.914 |
| 0.833 | 125 | 1 | 0.861 | 0.02882 | 0.7930 | 0.908 |
| 0.879 | 124 | 1 | 0.854 | 0.02941 | 0.7852 | 0.902 |
| 0.915 | 123 | 1 | 0.847 | 0.02998 | 0.7773 | 0.897 |
| 0.953 | 122 | 1 | 0.840 | 0.03053 | 0.7695 | 0.891 |
| 0.984 | 121 | 1 | 0.833 | 0.03106 | 0.7617 | 0.885 |
| 1.063 | 120 | 1 | 0.826 | 0.03156 | 0.7540 | 0.879 |
| 1.096 | 119 | 1 | 0.819 | 0.03205 | 0.7463 | 0.873 |
| 1.260 | 118 | 1 | 0.813 | 0.03253 | 0.7386 | 0.867 |
| 1.504 | 116 | 1 | 0.805 | 0.03299 | 0.7309 | 0.861 |
| 1.512 | 115 | 1 | 0.798 | 0.03344 | 0.7232 | 0.855 |
| 1.532 | 114 | 1 | 0.791 | 0.03387 | 0.7155 | 0.849 |
| 1.636 | 113 | 1 | 0.784 | 0.03429 | 0.7079 | 0.843 |
| 1.844 | 111 | 1 | 0.777 | 0.03470 | 0.7002 | 0.837 |
| 1.901 | 110 | 1 | 0.770 | 0.03510 | 0.6925 | 0.831 |
| 1.926 | 109 | 1 | 0.763 | 0.03548 | 0.6849 | 0.825 |
| 1.940 | 108 | 1 | 0.756 | 0.03585 | 0.6773 | 0.818 |
| 1.992 | 107 | 1 | 0.749 | 0.03620 | 0.6697 | 0.812 |
| 2.055 | 106 | 1 | 0.742 | 0.03654 | 0.6621 | 0.806 |
| 2.088 | 105 | 1 | 0.735 | 0.03687 | 0.6546 | 0.800 |
| 2.107 | 104 | 1 | 0.728 | 0.03719 | 0.6471 | 0.793 |
| 2.132 | 103 | 1 | 0.721 | 0.03749 | 0.6396 | 0.787 |
| 2.151 | 102 | 1 | 0.714 | 0.03779 | 0.6321 | 0.780 |
| 2.153 | 101 | 1 | 0.707 | 0.03807 | 0.6247 | 0.774 |
| 2.164 | 100 | 1 | 0.700 | 0.03834 | 0.6173 | 0.768 |
| 2.184 | 99 | 1 | 0.693 | 0.03860 | 0.6099 | 0.761 |
| 2.258 | 98 | 1 | 0.686 | 0.03884 | 0.6025 | 0.755 |
| 2.329 | 96 | 1 | 0.678 | 0.03909 | 0.5950 | 0.748 |
| 2.438 | 95 | 1 | 0.671 | 0.03933 | 0.5876 | 0.742 |
| 2.477 | 93 | 1 | 0.664 | 0.03956 | 0.5801 | 0.735 |
| 2.548 | 92 | 1 | 0.657 | 0.03978 | 0.5726 | 0.728 |
| 2.562 | 91 | 1 | 0.650 | 0.04000 | 0.5652 | 0.722 |
| 2.660 | 90 | 1 | 0.642 | 0.04020 | 0.5577 | 0.715 |
| 2.685 | 88 | 1 | 0.635 | 0.04040 | 0.5502 | 0.708 |
| 2.740 | 87 | 1 | 0.628 | 0.04059 | 0.5427 | 0.701 |
| 2.773 | 86 | 1 | 0.620 | 0.04077 | 0.5352 | 0.695 |
| 2.841 | 83 | 1 | 0.613 | 0.04096 | 0.5276 | 0.688 |
| 2.951 | 82 | 1 | 0.606 | 0.04113 | 0.5199 | 0.681 |
| 2.967 | 81 | 1 | 0.598 | 0.04130 | 0.5123 | 0.674 |
| 3.192 | 77 | 1 | 0.590 | 0.04149 | 0.5043 | 0.666 |
| 3.263 | 75 | 2 | 0.575 | 0.04185 | 0.4882 | 0.652 |
| 3.279 | 73 | 1 | 0.567 | 0.04201 | 0.4802 | 0.644 |
| 3.334 | 72 | 1 | 0.559 | 0.04215 | 0.4722 | 0.637 |
| 3.699 | 62 | 1 | 0.550 | 0.04243 | 0.4629 | 0.628 |
| 3.715 | 61 | 1 | 0.541 | 0.04268 | 0.4536 | 0.620 |
| 3.929 | 57 | 1 | 0.531 | 0.04297 | 0.4437 | 0.611 |
| 4.005 | 53 | 1 | 0.521 | 0.04331 | 0.4332 | 0.602 |
| 4.088 | 52 | 1 | 0.511 | 0.04363 | 0.4228 | 0.593 |
| 4.159 | 51 | 1 | 0.501 | 0.04391 | 0.4125 | 0.584 |
| 4.427 | 48 | 1 | 0.491 | 0.04422 | 0.4017 | 0.574 |
| 4.608 | 46 | 1 | 0.480 | 0.04452 | 0.3906 | 0.564 |
| 4.630 | 45 | 1 | 0.469 | 0.04479 | 0.3797 | 0.554 |
| 5.005 | 40 | 1 | 0.458 | 0.04519 | 0.3675 | 0.543 |
| 5.274 | 38 | 1 | 0.446 | 0.04557 | 0.3550 | 0.532 |
| 5.674 | 34 | 1 | 0.433 | 0.04608 | 0.3412 | 0.520 |
| 6.293 | 30 | 1 | 0.418 | 0.04675 | 0.3259 | 0.508 |
| 6.959 | 24 | 1 | 0.401 | 0.04793 | 0.3067 | 0.493 |
| 7.118 | 19 | 1 | 0.380 | 0.04984 | 0.2827 | 0.476 |
| 7.660 | 15 | 1 | 0.354 | 0.05255 | 0.2533 | 0.457 |
| 8.466 | 12 | 1 | 0.325 | 0.05585 | 0.2192 | 0.435 |
| 8.888 | 11 | 1 | 0.295 | 0.05806 | 0.1877 | 0.411 |
| 9.200 | 10 | 1 | 0.266 | 0.05929 | 0.1583 | 0.386 |
| 9.438 | 9 | 1 | 0.236 | 0.05960 | 0.1310 | 0.359 |
| 9.756 | 7 | 1 | 0.202 | 0.05988 | 0.1007 | 0.329 |
| 10.307 | 6 | 1 | 0.169 | 0.05864 | 0.0736 | 0.297 |

```
> ggsurvplot(st.stage,xlab="waktu survival (tahun)",ylab = "Fungsi survival",title="Estimasi Kaplan-Meier")
```



Berdasarkan hasil visualisasi fungsi survival berdasarkan stage, terlihat bahwa terdapat perbedaan antara fungsi survival untuk empat stage tersebut. Hal ini mengindikasikan bahwa stage merupakan calon variabel penjelas yang cukup baik untuk waktu terjadinya kematian (*event*). Pada plot tersebut juga terlihat bahwa pada mayoritas titik waktu pengamatan, fungsi survival untuk stage 1 lebih tinggi daripada stage 2. Namun, fungsi survival stage 2 lebih tinggi dari fungsi survival stage 3 dan fungsi survival stage 3 lebih tinggi dari fungsi survival stage 4. Hal ini mengimplikasikan bahwa pasien penderita pbc stage 4 lebih cepat atau memiliki resiko yang lebih besar untuk mengalami kematian (*event*) dibandingkan stage 3, stage 2, dan stage 1. Atau dapat disimpulkan bahwa semakin tinggi tingkatan stagenya, maka akan semakin beresiko untuk mengalami kematian (*event*).

2.2 Uji K-Sampel

Berdasarkan hasil analisis deskriptif, agar lebih yakin bahwa memang terdapat perbedaan fungsi survival yang signifikan dari setiap stadium, akan dilakukan uji k-sampel menggunakan bantuan *software* R untuk melihat apakah terdapat perbedaan risiko kematian penderita PBC dengan stadium yang berbeda (1,2,3,4). Hipotesis yang diajukan adalah sebagai berikut:

$H_0: h_1(t) = h_2(t) = h_3(t) = h_4(t), \text{ untuk } t \leq \tau$ (tidak ada perbedaan resiko kematian dari setiap stadium)

$H_1: h_j(t) \neq h_k(t), \text{ untuk beberapa } t \leq \tau, j \neq k, j, k = 1, 2, 3, 4$ (setidaknya terdapat satu stadium dengan risiko kematian yang berbeda)

Dengan menggunakan bantuan *software* R, dapat dilakukan uji log-rank test dengan hasil yang didapatkan sebagai berikut.

```
> survdiff(Surv(time,status)~stage, data=pbcc)
Call:
survdiff(formula = Surv(time, status) ~ stage, data = pbcc)
```

| | N | Observed | Expected | (O-E) ² /E | (O-E) ² /V |
|---------|-----|----------|----------|-----------------------|-----------------------|
| stage=1 | 21 | 2 | 11.4 | 7.77 | 8.45 |
| stage=2 | 92 | 23 | 44.1 | 10.08 | 14.20 |
| stage=3 | 156 | 48 | 61.4 | 2.92 | 4.81 |
| stage=4 | 144 | 84 | 40.1 | 48.00 | 65.51 |

```
Chisq= 70.3 on 3 degrees of freedom, p= 4e-15
```

Dari hasil di atas, terlihat bahwa stadium 1, 2, dan 3 memiliki nilai *expected* yang lebih tinggi dari *observed*. Sedangkan, untuk stadium 4, ternyata nilai *observed*-nya lebih tinggi dari nilai *expected*. Secara tidak langsung, hasil ini memberikan informasi bahwa terdapat perbedaan risiko kematian antarbeberapa stadium penderita PBC.

Hal ini dikonfirmasi dengan nilai p-valuenya ($p = 4 \times 10^{-15}$) yang lebih kecil dari taraf signifikansi ($\alpha = 0.05$) sehingga H_0 ditolak. Berarti, dapat disimpulkan bahwa terdapat perbedaan risiko kematian dari penderita PBC dengan stadium yang berbeda.

2.3 Analisis Model Regresi Cox-PH

Setelah mengetahui bahwa terdapat perbedaan risiko kematian dari pasien PBC dengan stadium yang berbeda, ingin diketahui lebih lanjut risiko kematian dari stadium

manakah yang lebih tinggi ataupun lebih rendah. Untuk menganalisis hal tersebut, akan dilakukan analisis dari model regresi Cox-PH.

Pertama-tama, akan dibentuk model regresi Cox-PH terlebih dahulu. Karena variabel stadium merupakan variabel kategorik, akan dibentuk variabel dummy Z_2, Z_3 , dan Z_4 dengan stadium=1 merupakan kategori acuan(*baseline*), sebagai berikut.

$$Z_2 = \begin{cases} 1, & \text{stadium} = 2 \\ 0, & \text{stadium} \neq 2 \end{cases}$$

$$Z_3 = \begin{cases} 1, & \text{stadium} = 3 \\ 0, & \text{stadium} \neq 3 \end{cases}$$

$$Z_4 = \begin{cases} 1, & \text{stadium} = 4 \\ 0, & \text{stadium} \neq 4 \end{cases}$$

Sehingga model Cox-PH yang dibentuk adalah

$$h(t, x) = h_o(t)e^{\beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4}$$

$i = 2, 3, 4 = \text{baseline hazard}$ yang bergantung waktu t

$\beta_i =$ koefisien regresi yang bersesuaian dengan variabel penjelas, $i = 2, 3, 4$

$Z_i =$ variabel penjelas (stadium), $i = 2, 3, 4$

Dengan menggunakan bantuan *software* R, didapatkan estimasi parameter β dari model Cox-PH yang telah dibentuk menggunakan metode Breslow, sebagai berikut.

```
> summary(coxph(Surv(time,status)~factor(stage), data=pbcc, method="breslow"))
Call:
coxph(formula = Surv(time, status) ~ factor(stage), data = pbcc,
      method = "breslow")

n = 413, number of events = 157

              coef exp(coef) se(coef)      z Pr(>|z|)
factor(stage)2  1.1028    3.0125  0.7374  1.496 0.134777
factor(stage)3  1.5291    4.6142  0.7224  2.117 0.034287
factor(stage)4  2.5313   12.5699  0.7168  3.532 0.000413

factor(stage)2
factor(stage)3 *
factor(stage)4 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
factor(stage)2      3.013    0.33195    0.710    12.78
factor(stage)3      4.614    0.21672    1.120    19.01
factor(stage)4     12.570    0.07956    3.085    51.22

Concordance= 0.694 (se = 0.02 )
Likelihood ratio test= 65.16 on 3 df,  p=5e-14
Wald test              = 59.6 on 3 df,  p=7e-13
Score (logrank) test = 70.25 on 3 df,  p=4e-15
```

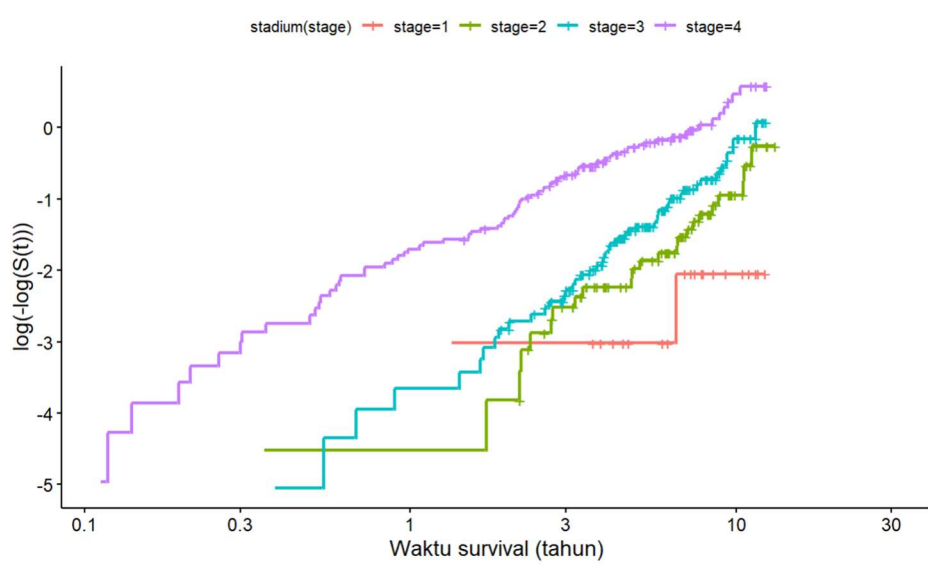
Dari hasil di atas, terlihat bahwa:

- Penderita PBC stadium 2 memiliki risiko kematian $\exp(\beta_2) = 3.0125$ kali dari penderita PBC stadium 1
- Penderita PBC stadium 3 memiliki risiko kematian $\exp(\beta_3) = 4.6142$ kali dari penderita PBC stadium 1
- Penderita PBC stadium 4 memiliki risiko kematian $\exp(\beta_4) = 12.5699$ kali dari penderita PBC stadium 1

Yang berarti terdapat keterurutan dari fungsi hazard penderita PBC, dengan stadium 1 memiliki risiko kematian yang paling rendah dan stadium 4 memiliki risiko kematian paling tinggi.

2.4 Pengecekan Asumsi Proportional Hazard(PH)

Untuk memastikan apakah benar stadium merupakan variabel penjelas yang signifikan terhadap risiko kematian, perlu dilakukan pengecekan asumsi PH menggunakan plot fungsi $\log(H(t))$ sebagai berikut.



Terlihat dari grafik bahwa stadium 2, 3, dan 4 memperlihatkan keparalelan yang cukup baik, terutama di tengah sampai akhir waktu. Akan tetapi, stadium 1 memperlihatkan ketidaksejajaran dengan setiap stadium lainnya. Hal ini mungkin disebabkan karena data pasien stadium yang terlalu sedikit. Oleh karena itu, jika ingin menggunakan hasil dari model Cox-PH ini, penggunaannya harus dibatasi pada waktu tengah dan akhir pengamatan saja, serta perlu ditambahkan data pasien stadium 1.

BAB 3

Penutup

3.1 Kesimpulan

Dari hasil analisis survival dataset Primary Biliary Cirrhosis (PBC), didapatkan kesimpulan sebagai berikut:

1. Terdapat perbedaan risiko kematian penderita PBC dengan stadium yang berbeda (1, 2, 3, 4).
2. Terdapat keterurutan dari fungsi hazard penderita PBC, yaitu stadium 1 memiliki risiko kematian yang paling rendah. dilanjutkan dengan stadium 2, kemudian 3, dan stadium 4 memiliki risiko kematian paling tinggi.
3. Penggunaan hasil dari model Cox-PH ini harus dibatasi di waktu tengah sampai akhir pengamatan saja, serta perlu ditambahkan data pasien stadium 1.

DAFTAR PUSAKA

Abdullah, Sarini. (2022). *Analisis Survival: Konsep dan Aplikasi dengan R*. Jakarta: Bumi Aksara

Abdullah, Sarini. “Studi kasus: pengolahan data survival dengan R dan interpretasinya” Youtube, diunggah oleh Prodi Statistika UI, 22 November 2020, <https://www.youtube.com/watch?v=HcvYAFougwM&t=584s>

Khadiv, Homayoon. (2021). *Primary Biliary Cirrhosis (pbc) Disease Dataset*. Diakses pada 10 Juni 2023 dari <https://www.kaggle.com/datasets/homayoonkhadivi/primary-biliary-cirrhosis-pbc-disease-dataset>

STHDA. *Cox Proportional-Hazards Model*. Diakses pada 10 Juni 2023 dari <http://www.sthda.com/english/wiki/cox-proportional-hazards-model>