

Homework 10

Xinyi Lin

5/5/2019

```
library(survival)
library(MASS)
library(survminer)
library(KMsurv)
library(tidyverse)
```

Problem 1

Input data

```
time = c(4, 12, 15, 21, 23, 2, 6, 8, 10, 19)
cens = c(1, 0, 1, 0, 1, 1, 0, 0, 1, 1)
group = c(rep(1, 5), rep(2, 5))
data1 = data.frame(time = time, cens = cens, group = group)
data1
```

##	time	cens	group
## 1	4	1	1
## 2	12	0	1
## 3	15	1	1
## 4	21	0	1
## 5	23	1	1
## 6	2	1	2
## 7	6	0	2
## 8	8	0	2
## 9	10	1	2
## 10	19	1	2

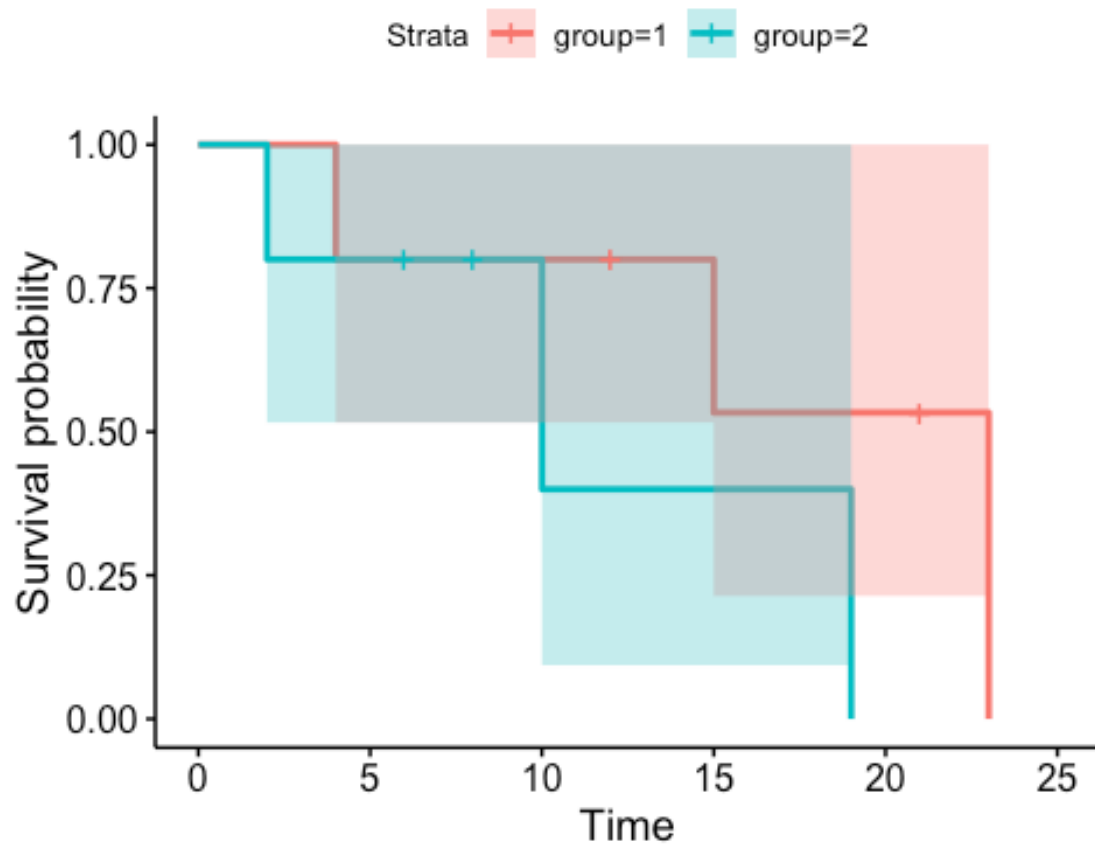
Using the log-rank test to test hypotheses.

$H_0: h_1(t) = h_2(t)$ for all t ; $H_1: h_1(t) \neq h_2(t)$ for some t .

```
survdif(Surv(time,cens)~group, data=data1) # Log rank test
```

```
## Call:
## survdiff(formula = Surv(time, cens) ~ group, data = data1)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## group=1  5          3     4.14     0.313     1.15
## group=2  5          3     1.86     0.697     1.15
##
##  Chisq= 1.1  on 1 degrees of freedom, p= 0.3
```

```
ggsurvplot( survfit(Surv(time, cens) ~ group, data = data1), conf.int=TRUE)
```



Calculating by using R, we can get that $Z^2 = 1.1$ and corresponding $p - value = 0.3$. As p -value is larger than 0.05, we fail to reject the null hypothesis and conclude that $h_1(t) = h_2(t)$ for all t . The plot of survival probability of two groups are shown above.

Problem 2

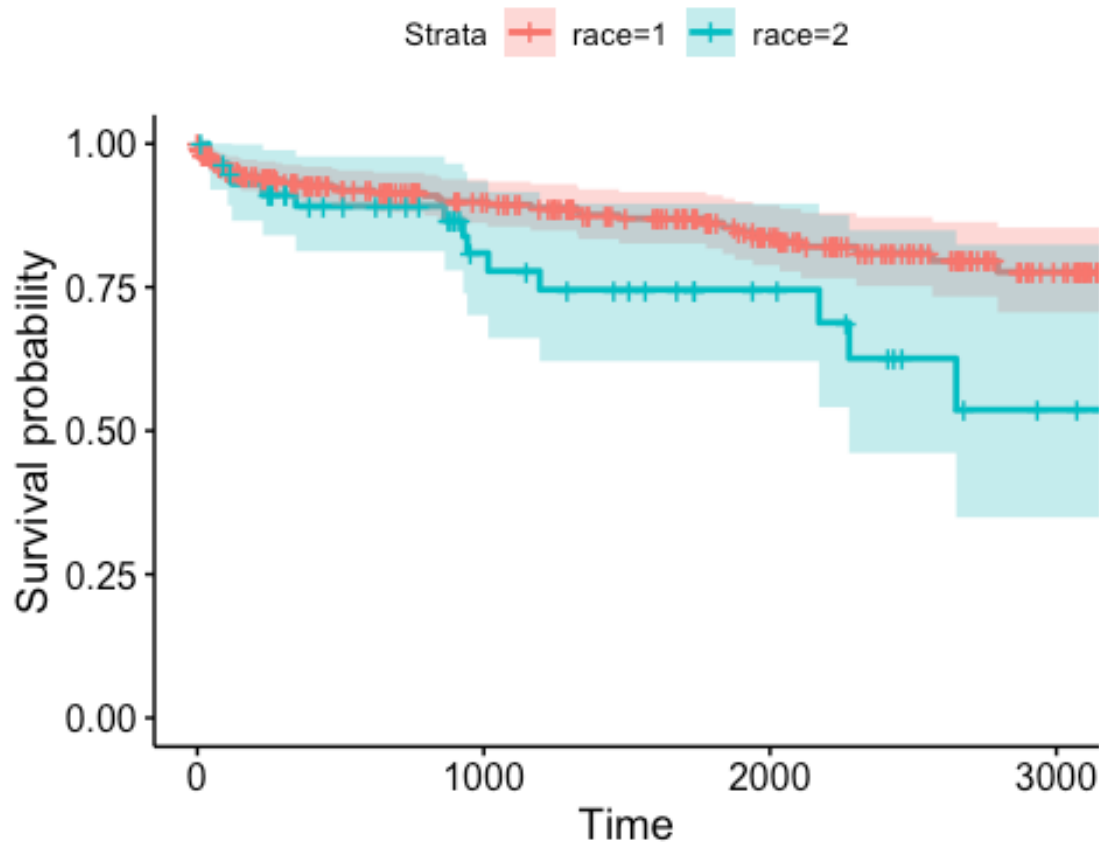
```
data("kidtran")
head(kidtran)
```

```
##   obs time delta gender race age
## 1   1   1     0      1    1  46
## 2   2   5     0      1    1  51
## 3   3   7     1      1    1  55
## 4   4   9     0      1    1  57
## 5   5  13     0      1    1  45
## 6   6  13     0      1    1  43
```

```
female_kid = kidtran %>%
  filter(gender == 2)
male_kid = kidtran %>%
  filter(gender == 1)
```

For female:

```
ggsurvplot( survfit(Surv(time, delta) ~ race, data = female_kid),  
conf.int=TRUE)
```



Using the log-rank test to test hypotheses.

$H_0: h_1(t) = h_2(t)$ for all t ; $H_1: h_1(t) \neq h_2(t)$ for some t .

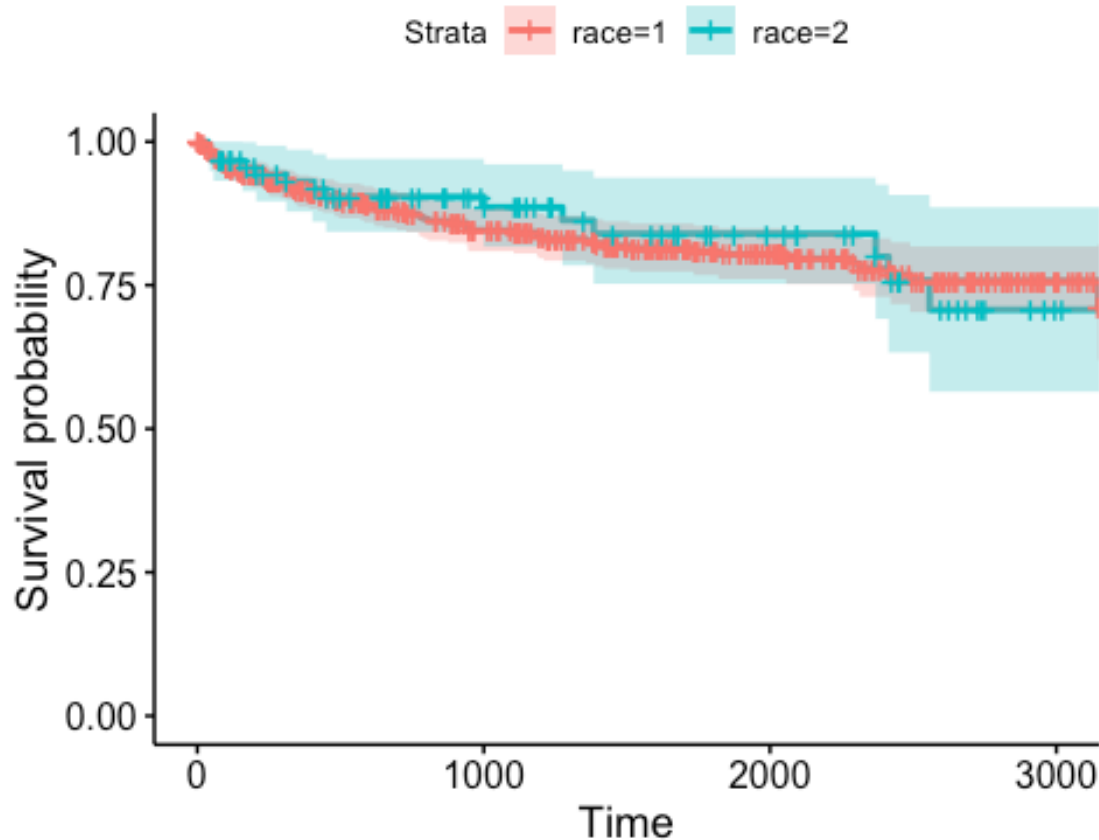
```
survdifff(Surv(time,delta) ~ race, data = female_kid)
```

```
## Call:  
## survdifff(formula = Surv(time, delta) ~ race, data = female_kid)  
##  
##      N Observed Expected (O-E)^2/E (O-E)^2/V  
## race=1 280      39   44.79    0.748    4.85  
## race=2  59      14    8.21    4.076    4.85  
##  
##  Chisq= 4.8  on 1 degrees of freedom, p= 0.03
```

Calculating by using R, we can get that $Z^2 = 4.8$ and corresponding p -value = 0.03. As p -value is smaller than 0.05, we reject the null hypothesis and conclude that $h_1(t) \neq h_2(t)$ for some t .

For male:

```
ggsurvplot( survfit(Surv(time, delta) ~ race, data = male_kid),  
conf.int=TRUE)
```



Using the log-rank test to test hypotheses.

$H_0: h_1(t) = h_2(t)$ for all t ; $H_1: h_1(t) \neq h_2(t)$ for some t .

```
survdifff(Surv(time,delta) ~ race, data = male_kid)
```

```
## Call:  
## survdifff(formula = Surv(time, delta) ~ race, data = male_kid)  
##  
##           N Observed Expected (O-E)^2/E (O-E)^2/V  
## race=1 432         73      71.9    0.0168    0.097  
## race=2  92         14      15.1    0.0801    0.097  
##  
##  Chisq= 0.1  on 1 degrees of freedom, p= 0.8
```

Calculating by using R, we can get that $Z^2 = 0.1$ and corresponding p -value = 0.8. As p -value is larger than 0.05, we fail to reject the null hypothesis and conclude that $h_1(t) = h_2(t)$ for all t .

Problem 3

Get data.

```
data("larynx")
larynx_data = larynx %>%
  mutate(z1 = ifelse(stage == 2, 1, 0),
         z2 = ifelse(stage == 3, 1, 0),
         z3 = ifelse(stage == 4, 1, 0))
head(larynx_data)

##   stage time age diagyr delta z1 z2 z3
## 1     1  0.6  77     76     1  0  0  0
## 2     1  1.3  53     71     1  0  0  0
## 3     1  2.4  45     71     1  0  0  0
## 4     1  2.5  57     78     0  0  0  0
## 5     1  3.2  58     74     1  0  0  0
## 6     1  3.2  51     77     0  0  0  0
```

We fit following model(using Breslow method for tie handling):

$$h(t) = h_0(t) \exp(\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3 + \beta_4 Z_4 + \beta_5 Z_1 \times Z_4)$$

```
fit=coxph(Surv(time,delta)~z1+z2+z3+age+z1*age,data=larynx_data,ties='breslow')
summary(fit)

## Call:
## coxph(formula = Surv(time, delta) ~ z1 + z2 + z3 + age + z1 *
##       age, data = larynx_data, ties = "breslow")
##
## n= 90, number of events= 50
##
##              coef exp(coef)    se(coef)      z Pr(>|z|)
## z1        -7.3820143  0.0006223  3.4027542 -2.169  0.0301 *
## z2         0.6218044  1.8622853  0.3558078  1.748  0.0805 .
## z3         1.7534270  5.7743576  0.4239595  4.136 3.54e-05 ***
## age         0.0059729  1.0059908  0.0148792  0.401  0.6881
## z1:age      0.1116674  1.1181409  0.0476728  2.342  0.0192 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## z1         0.0006223  1606.8231 7.900e-07  0.4903
## z2         1.8622853    0.5370 9.272e-01  3.7403
## z3         5.7743576    0.1732 2.516e+00 13.2550
## age         1.0059908    0.9940 9.771e-01  1.0358
## z1:age      1.1181409    0.8943 1.018e+00  1.2277
##
## Concordance= 0.682 (se = 0.045 )
## Rsquare= 0.235 (max possible= 0.988 )
```

```
## Likelihood ratio test= 24.11 on 5 df, p=2e-04
## Wald test = 23.77 on 5 df, p=2e-04
## Score (logrank) test = 27.98 on 5 df, p=4e-05
```

Explain of results:

$\beta_1 + \beta_5$: the log hazard ratio for subjects in stage II versus stage I is $(-7.382 + 0.112 \times a)$ given they have same age, where a is the age of subjects.

β_2 : the log hazard ratio for subjects in stage III versus stage I is 0.621 given they have same age.

β_3 : the log hazard ratio for subjects in stage IV versus stage I is 1.753 given they have same age.

β_4 : when patients are not in stage II, the log hazard ratio for subjects with one unit changes in age is 0.006 given they are in same stage.

$\beta_4 + \beta_5$: when patients are in stage II, the log hazard ratio for subjects with one unit changes in age is 0.118.

Relative risk:

For the hazard of dying for a stage II patient of age 50 is

$$h_2(t) = h_0(t) \exp(-7.382 \times 1 + 0.111 \times 1 \times 50)$$

For the hazard of dying for a stage I patient of age 50 is

$$h_1(t) = h_0(t) \exp(0)$$

So the hazard ratio is

$$HR(t) = \frac{h_2(t)}{h_1(t)} = 0.16$$