

p8131_hw10_xy2395

Jack Yan

4/30/2019

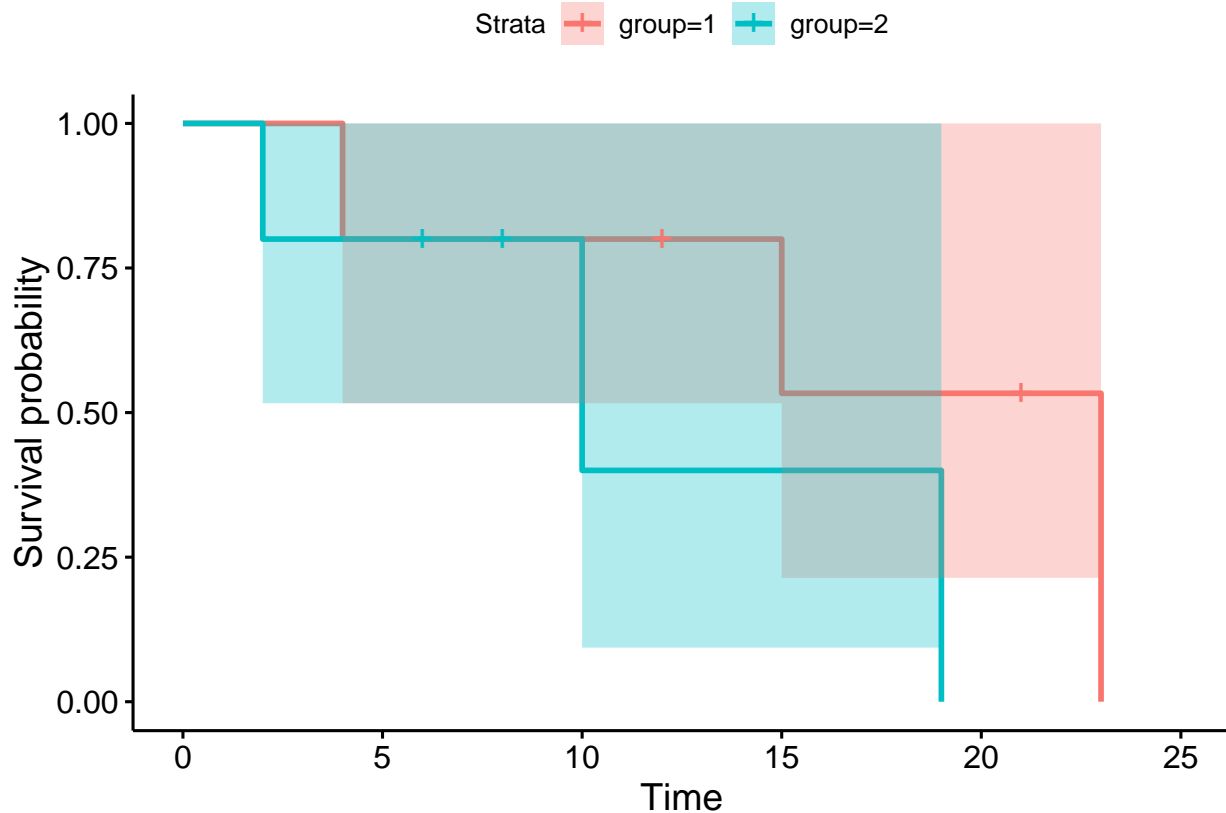
```
library(tidyverse)
library(survival)
library(survminer)
library(KMsurv)
```

Problem 1

```
data1 =
  tibble(
    time = c(4, 12, 15, 21, 23, 2, 6, 8, 10, 19),
    delta = c(1, 0, 1, 0, 1, 1, 0, 0, 1, 1),
    group = c(1, 1, 1, 1, 1, 2, 2, 2, 2, 2)
  )
Surv(data1$time, data1$delta, type = 'right')

## [1] 4 12+ 15 21+ 23 2 6+ 8+ 10 19

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



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

```
## Call:
## survdiff(formula = Surv(time, delta) ~ 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
```

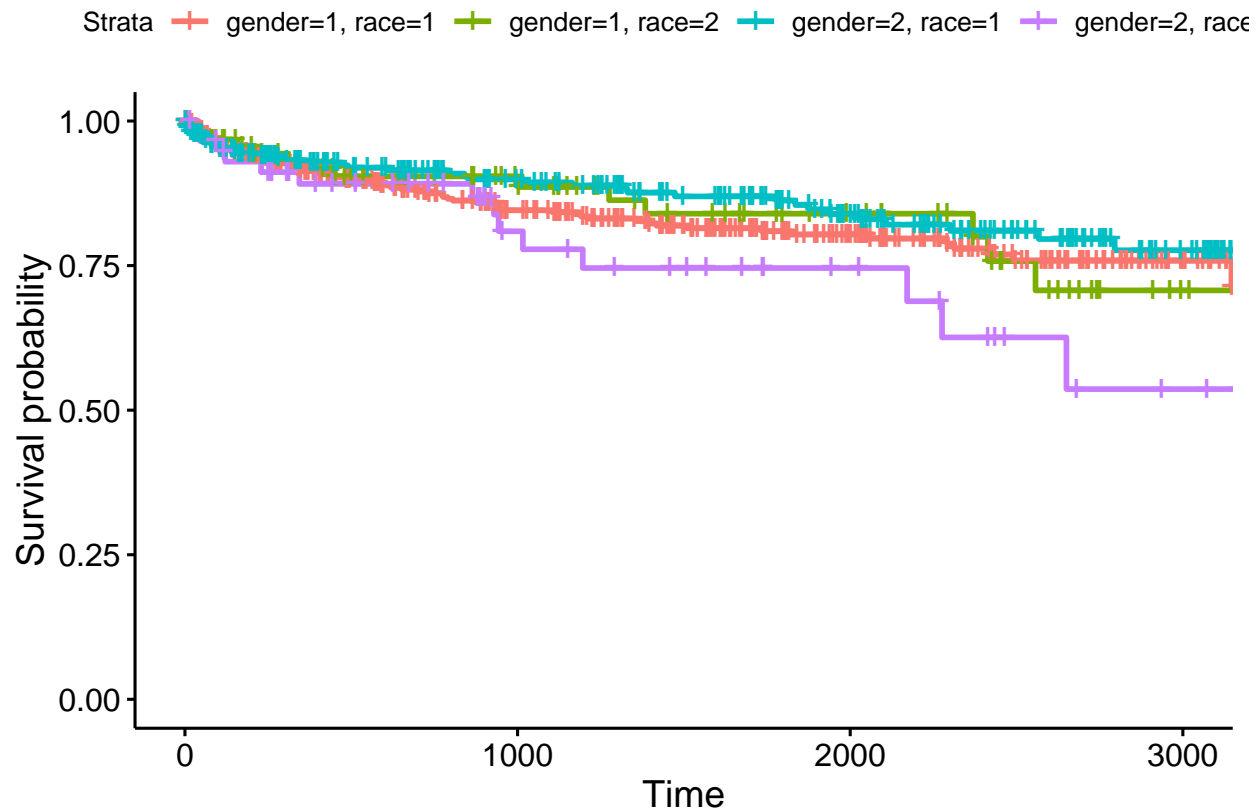
The chi-squared statistic is 1.1 and p-value is 0.3 > 0.05, so we fail to reject the null and conclude that the survival curves of the 2 groups are not different.

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
```

```
# plot
ggsurvplot(survfit(Surv(time, delta) ~ gender + race, data = kidtran), conf.int=F)
```



```
# Log-rank test
# male
survdif(Surv(time, delta)~ gender + race, data = kidtran, subset = (gender == 1))
```

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

```
# female
survdif(Surv(time, delta)~ gender + race, data = kidtran, subset = (gender == 2))
```

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

For males, the chi-squared statistic is 0.1 and p-value is $0.8 > 0.05$, so we fail to reject the null and conclude that the survival curve of different races is not different in males. On the other hand, for females, the chi-squared statistics is 4.8 and p-value is $0.03 < 0.05$, so we have enough evidence to reject the null and

conclude that the survival curves for different races is different in females.

Problem 3

```
data(larynx)
head(larynx)

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

data3 =
  larynx %>%
  mutate(z1 = if_else(stage == 2, 1, 0),
         z2 = if_else(stage == 3, 1, 0),
         z3 = if_else(stage == 4, 1, 0)) %>%
  rename(z4 = age)
fit = coxph(Surv(time, delta)~z1 + z2 + z3 + z4 + z1 * z4, data = data3)
summary(fit)
```

```
## Call:
## coxph(formula = Surv(time, delta) ~ z1 + z2 + z3 + z4 + z1 *
##       z4, data = data3)
##
##   n= 90, number of events= 50
##
##              coef exp(coef)    se(coef)      z Pr(>|z|)
## z1      -7.4898902  0.0005587  3.4169200 -2.192   0.0284 *
## z2       0.6251294  1.8684878  0.3558360  1.757   0.0790 .
## z3       1.7681449  5.8599725  0.4237680  4.172 3.01e-05 ***
## z4       0.0060299  1.0060481  0.0148750  0.405   0.6852
## z1:z4    0.1133273  1.1199985  0.0478711  2.367   0.0179 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## z1      0.0005587 1789.8556 6.898e-07  0.4526
## z2      1.8684878    0.5352 9.303e-01  3.7530
## z3      5.8599725    0.1706 2.554e+00 13.4464
## z4      1.0060481    0.9940 9.771e-01  1.0358
## z1:z4  1.1199985    0.8929 1.020e+00  1.2302
##
## Concordance= 0.682 (se = 0.04 )
## Likelihood ratio test= 24.49 on 5 df,  p=2e-04
## Wald test               = 24.14 on 5 df,  p=2e-04
## Score (logrank) test = 28.49 on 5 df,  p=3e-05
```

Explain results here. The relative risk of dying is 0.0005587 for a stage II patient of age 50 as compared to a stage I patient of the same age.