p8131_hw10_xy2395

Jack Yan

4/30/2019

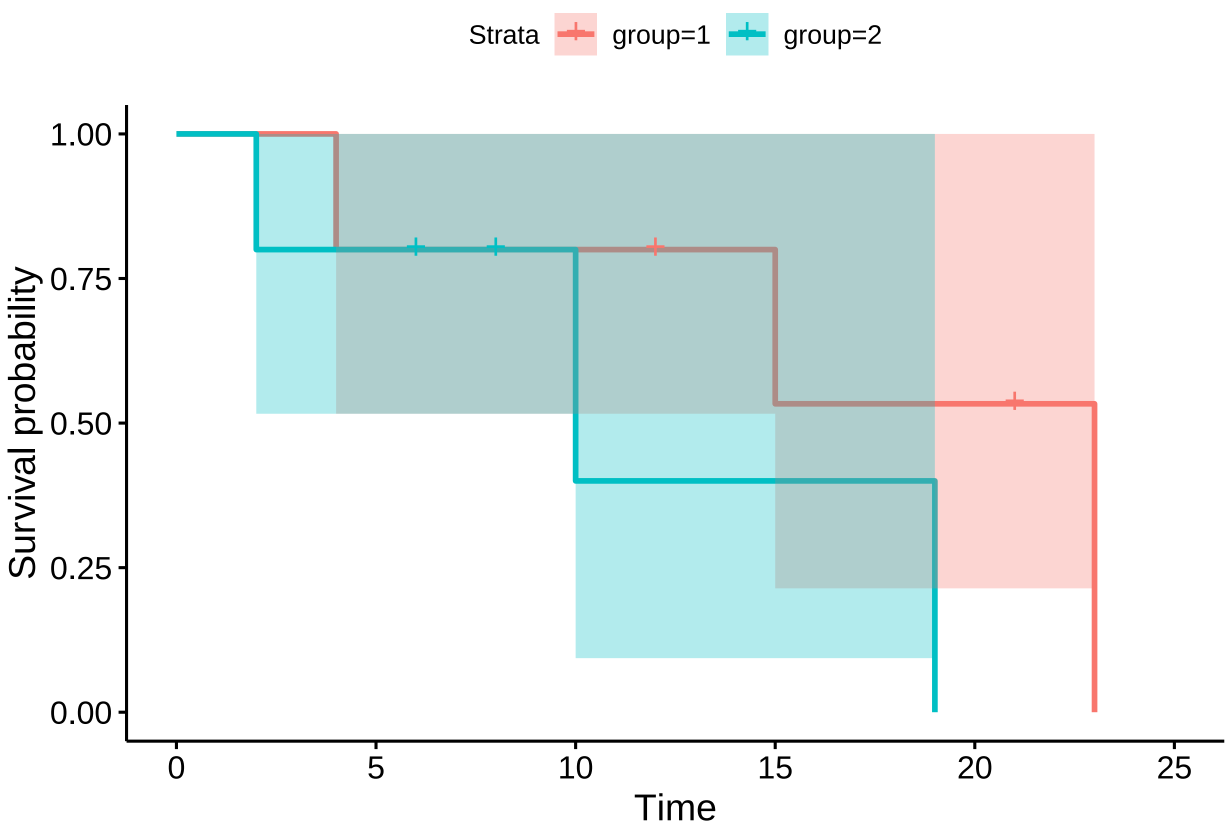
```
library(tidyverse)
library(survival)
library(survminer)
library(KMsurv)
```

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)
```



```r
# Log-rank test
survdiff(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
```

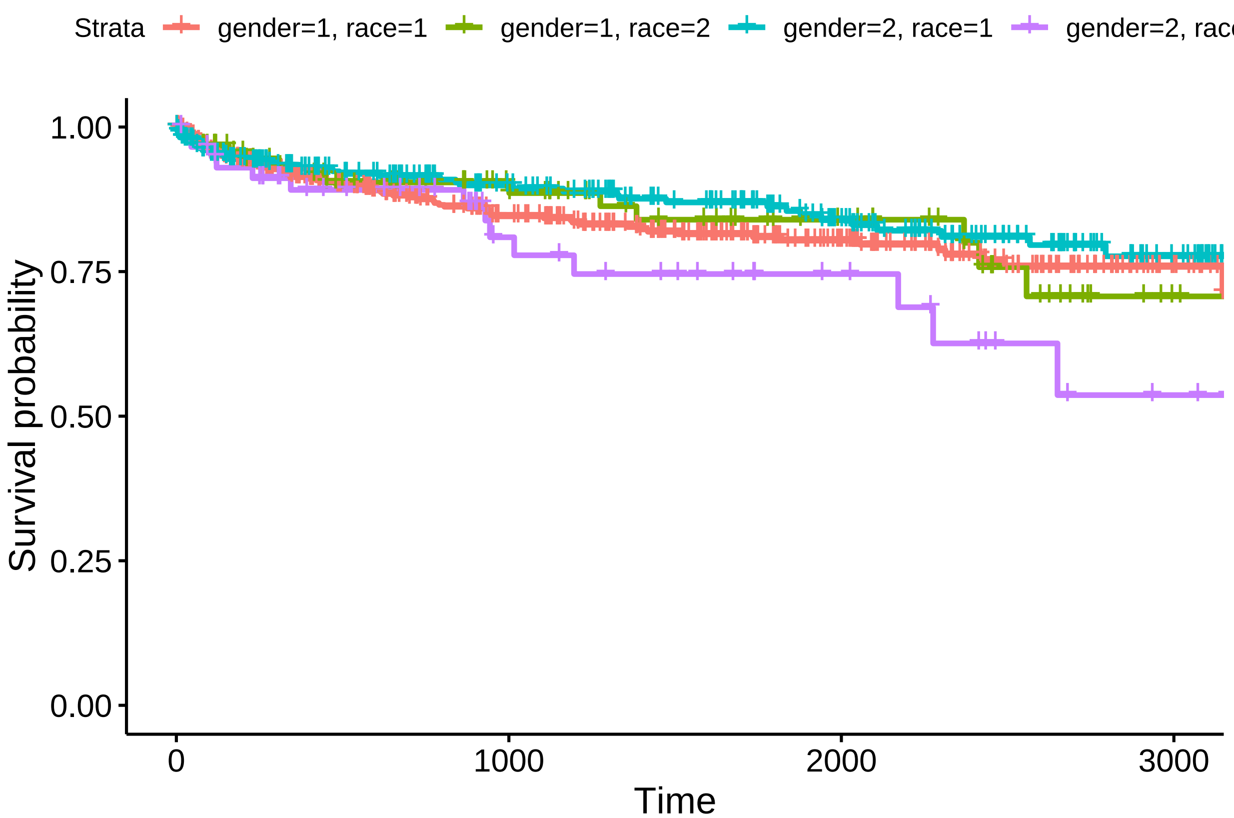
The chi-squared statistic is 1.1 and p-value is 0.3 >0.05, so we fail to reject the null and conclude that the survival curves of the 2 groups are not different.

Problem 2

```r
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
```

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




```
# Log-rank test
# male
survdiff(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
survdiff(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

```r
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
```

```r
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, ties = "breslow")
summary(fit)
```

```
## Call:
## coxph(formula = Surv(time, delta) ~ z1 + z2 + z3 + z4 + z1 *
##     z4, data = data3, 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 ***
## z4     0.0059729  1.0059908  0.0148792  0.401   0.6881
## z1:z4  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
## z4    1.0059908     0.9940 9.771e-01    1.0358
## z1:z4 1.1181409     0.8943 1.018e+00    1.2277
##
## Concordance= 0.682  (se = 0.04 )
## 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
```

The proportional hazards model is:

$$h_i(t) = h_0(t)e^{-7.38Z_1+0.62Z_2+1.75Z_3+0.006Z_4+0.11Z_1\times Z_4}$$

Interpretations:

The hazard ratio for subjects of age k in Stage II versus subjects in Stage I of the same age is $e^{-7.38+0.11\times k}$.

The hazard ratio for subjects in Stage III versus subjects in Stage I is 1.86, keeping age at diagnosis constant.

The hazard ratio for subjects in Stage IV versus subjects in Stage I is 5.77, keeping age at diagnosis constant.

If the patients are not in stage II, the hazard ratio for one-year increase in age at diagnosis is 1.0059908, for patients in the same stage. If the patients are in stage II, the hazard ratio for one-year increase in age at diagnosis is $e^{0.006+0.112} = 1.12$.

The relative risk of dying is $e^{-7.38+0.11\times 50} = 0.153$ for a stage II patient of age 50 as compared to a stage I patient of the same age.