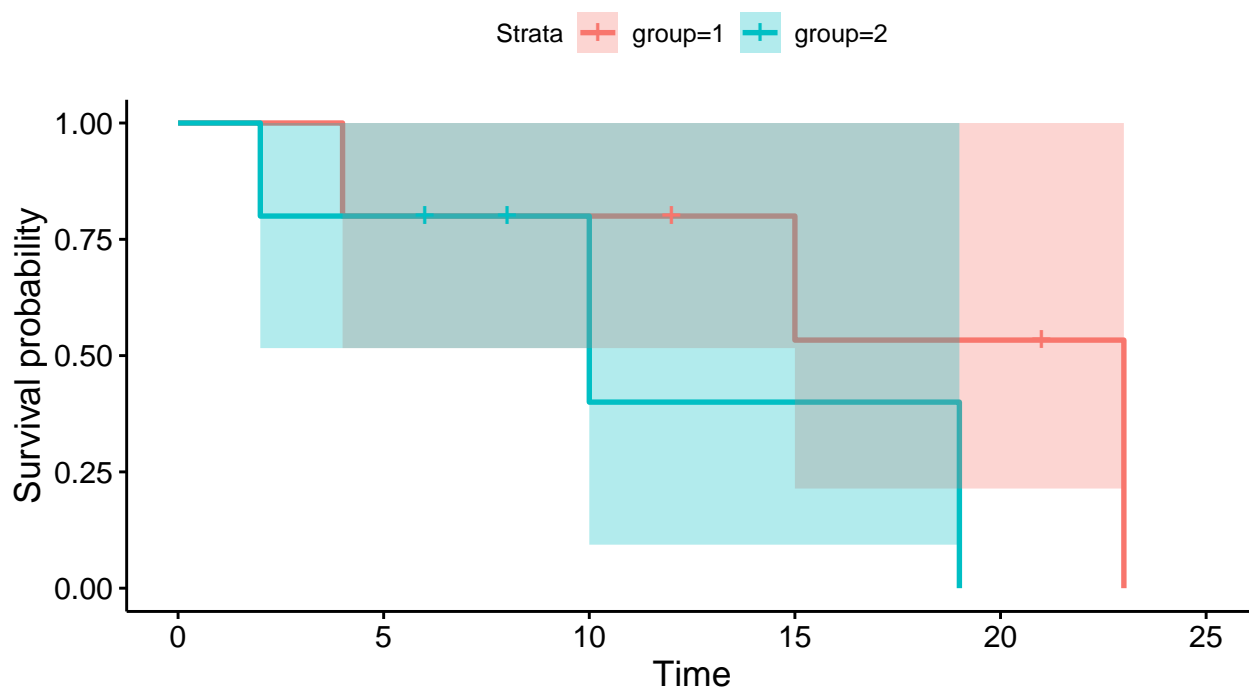# P8131 HOMEWORK 10

xc2474 Xinlei Chen

Problem 1

Suppose we have the following data:

Group 1: 4, 12+, 15, 21+, 23

Group 2: 2, 6+, 8+, 10, 19

Test H_0: $h_1(t) = h_2(t)$ for all t, $H_1 : h_1(t) \neq h_2(t)$ for some t using the log-rank test.

```
## 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
```
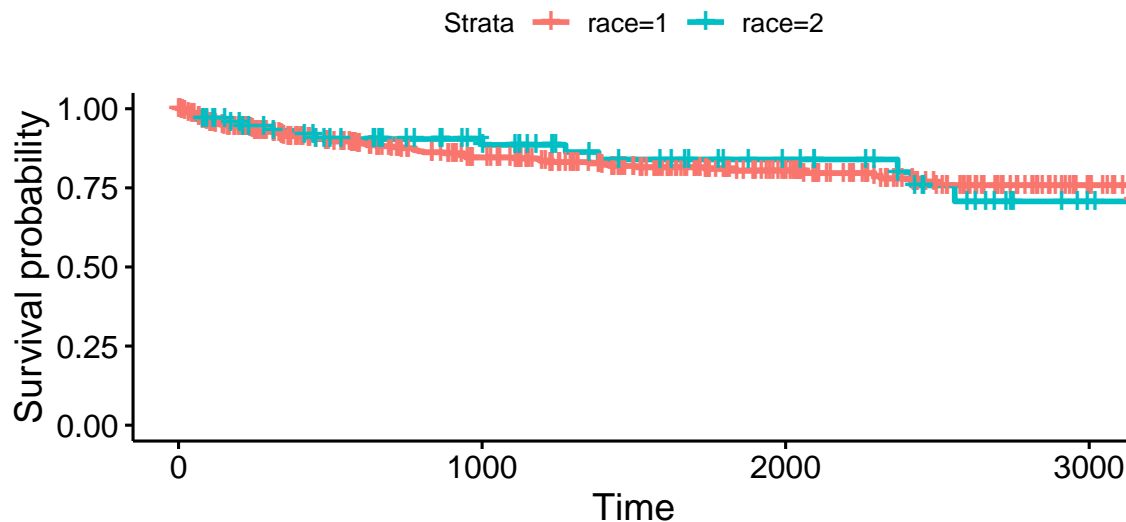

The test result shows that the p-value is 0.3 which is greater than 0.05, so we fail to reject the null hypothesis and state that the survival curves of 2 groups are not significantly different.

Problem 2

Consider the kidtran dataset in the KMsurv package where the death times of kidney transplant patients are reported. The patients can be classified by race (1=white and 2=black) and gender (1=male and 2=female) into one of four groups. Compare the survival curves for different races in each sex group.
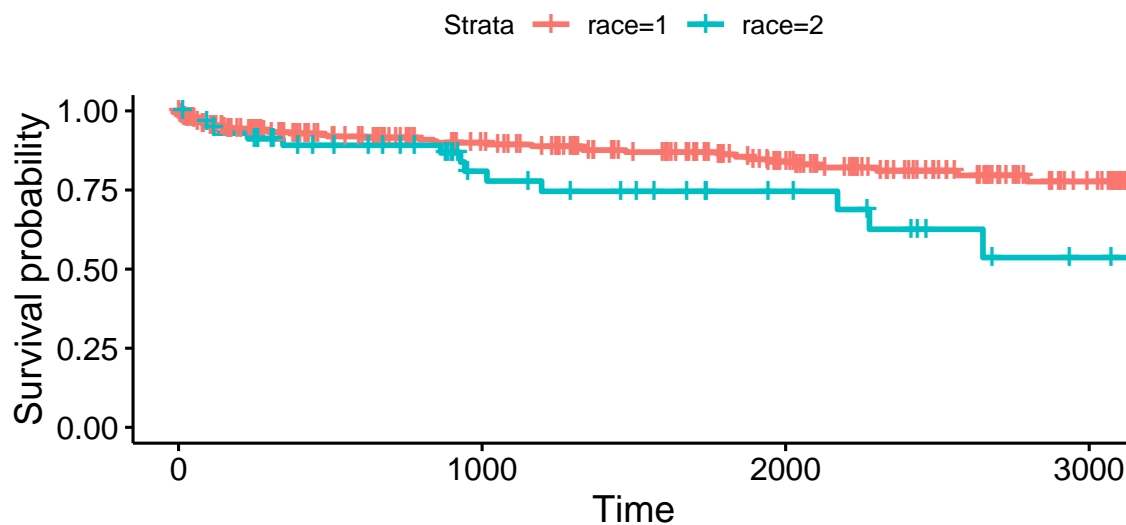
(1) For male group:


```
## 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
```

Comment:

For male group, the chi-squared statistic is 0.1 and the p-value is 0.8 which is greater than 0.05, so we fail to reject the null and state that the survival curves of different races is not significantly different in male group.

(2) For female group:

```
## 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
```

Comment:
For female group, the chi-squared statistics is 4.8 and p-value is 0.03 which is smaller than 0.05, so we reject
the null and state that the survival curves for different races is significantly different in female group.

Problem 3

Consider the larynx data in the KMsurv package. Let
$Z1 = 1$ if the patient is in stage II, 0 otherwise
$Z2 = 1$ if the patient is in stage III, 0 otherwise
$Z3 = 1$ if the patient is in stage IV, 0 otherwise
$Z4 =$ age of the patient (in years):
Fit a proportional hazards model using the covariates Z1, Z2, Z3, Z4 and the interaction term Z1 * Z4.
Explain your results. What is the relative risk of dying for a stage II parient of age 50 as compared to a
stage I patient of the same age?

```
## Call:
## coxph(formula = Surv(time, delta) ~ z1 + z2 + z3 + z4 + z1 *
##      z4, data = data2, 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:

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

Interpretations:

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

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

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

3 When 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 1.12.

Relative Risk:

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

Code

```r
knitr::opts_chunk$set(echo = FALSE)
library(tidyverse)
library(survival)
library(survminer)
library(KMsurv)
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)
  )

survdiff(Surv(time,delta)~group, data=data1)
ggsurvplot(survfit(Surv(time, delta) ~ group, data = data1), conf.int=TRUE)
data(kidtran)
# male
# plot
ggsurvplot(survfit(Surv(time, delta) ~ race,
                   data = subset(kidtran, gender == 1)),
           conf.int=FALSE)
# log-rank test
survdiff(Surv(time, delta)~ gender + race, data = kidtran, subset = (gender == 1))
# female
# plot
ggsurvplot(survfit(Surv(time, delta) ~ race,
                   data = subset(kidtran, gender == 2)),
           conf.int=FALSE)
# log-rank test
survdiff(Surv(time, delta)~ gender + race, data = kidtran, subset = (gender == 2))
data(larynx)
data2 = 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 = data2, ties = "breslow")
summary(fit)
```