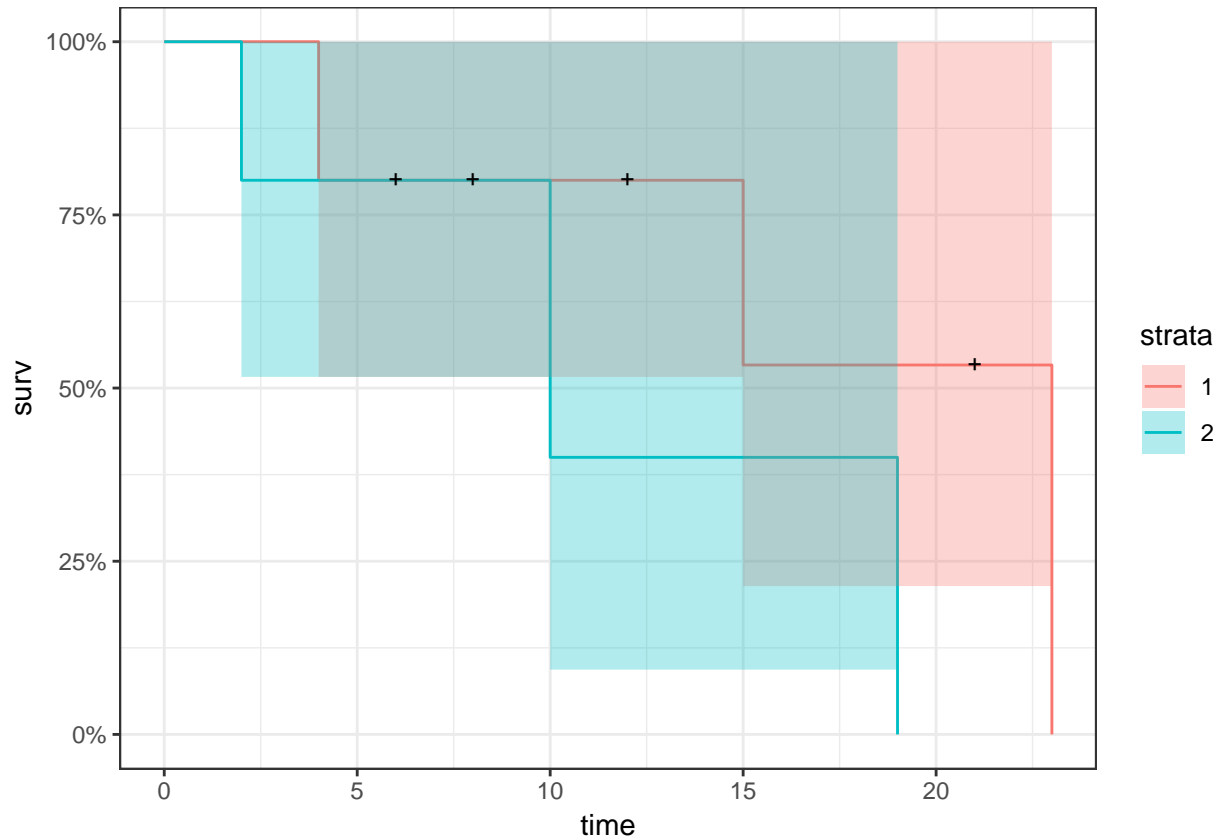# HW10_answer

Guojing Wu

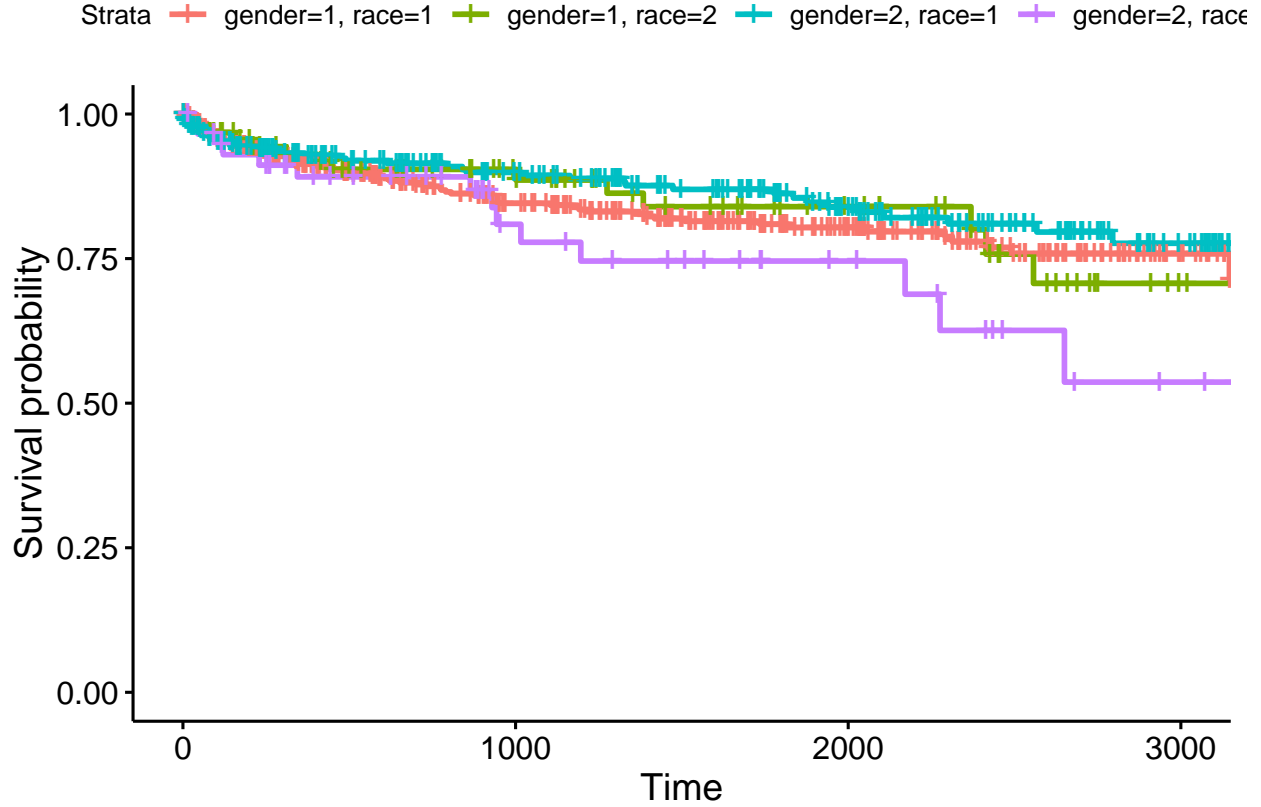*5/6/2019*

Problem 1. Log-rank test

Figure 1: KM curve for each group

We got the pvalue $= 0.284 > 0.05$, so we fail to reject the null and state that there is not much a difference between group1 and group2.

Problem 2.

Figure 2: KM curve for each strata

- For male patient, the pvalue $= 0.755 > 0.05$, so we fail to reject the null and state that there is not much a difference between black and white in male subgroup.

- For female patient, the pvalue $= 0.028 < 0.05$, so we reject the null and state that there is differences between black and white in female subgroup.

Problem 3

Table 1: coefficients for each predictor

Z1	Z2	Z3	Z4	Z1:Z4
-7.382014	0.6218044	1.753427	0.0059729	0.1116674

- At a given time point and fixed age k, the hazard ratio for subjects in Stage II versus subjects in Stage I is $e^{-7.38+0.11\times k}$.

- At a given time point, the log hazard ratio for subjects in Stage III versus subjects in Stage I is 0.622, keeping age at diagnosis constant.

- At a given time point, the log hazard ratio for subjects in Stage IV versus subjects in Stage I is 1.753, keeping age at diagnosis constant.

- At a given time point, for patients who are in stage II, the hazard ratio is $e^{0.006+0.112} = 1.12$ for 1 unit increase in age at diagnosis.

- At a given time point, for patients who are not in stage II, the log hazard ratio is 0.006, for 1 unit increase in age at diagnosis.

The relative risk of dying is $e^{-7.38+0.11\times50} = 0.166$ for a stage II patient of age 50 as compared to a stage I

patient of the same age.

Code

```r
knitr::opts_chunk$set(echo = F,
                      message = F,
                      warning = F,
                      comment = "")
library(tidyverse)
library(survival)
library(KMsurv) # contains many interesting data sets
library(ggfortify) # more beautiful plot
library(survminer)
theme_set(theme_bw())
dat.1 <- tibble(
  time = c(c(4, 12, 15, 21, 23), c(2, 6, 8, 10, 19)),
  cens = c(c(1, 0, 1, 0, 1), c(1, 0, 0, 1, 1)),
  group = c(rep(1, 5), rep(2, 5))
)
res1 = survdiff(Surv(time, cens) ~ group, data = dat.1)
autoplot(survfit(Surv(time, cens) ~ group, data = dat.1))
data("kidtran")
# autoplot(survfit(Surv(time, delta) ~ gender + race, data = kidtran))
ggsurvplot(survfit(Surv(time, delta) ~ gender + race, data = kidtran), conf.int=F)

res2.1 = survdiff(Surv(time, delta)~ gender + race, data = kidtran, subset = (gender == 1))
res2.2 = survdiff(Surv(time, delta)~ gender + race, data = kidtran, subset = (gender == 2))
data("larynx")
larynx = larynx %>%
  mutate(Z1 = ifelse(stage == 2, 1, 0),
         Z2 = ifelse(stage == 3, 1, 0),
         Z3 = ifelse(stage == 4, 1, 0),
         Z4 = age)

res3 = coxph(Surv(time, delta) ~ Z1 + Z2 + Z3 + Z4 + Z1 * Z4, data = larynx, ties = 'breslow')
t(res3$coefficients) %>% knitr::kable()
```