HW10

Zirui Zhang

2023-05-02

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

Question 1

```
# build data frame
time = c(4, 12, 15, 21, 23,
         2, 6, 8, 10, 19)
cens = c(0, 1, 0, 1, 0,
         0, 1, 1, 0, 0)
group = c(rep("group1", 5),
          rep("group2", 5))
df = data.frame(time, cens, group)
# log-rank test
survdiff(Surv(time, cens)~group, data = df)
```

```
## Call:
## survdiff(formula = Surv(time, cens) ~ group, data = df)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## group=group1 5        2     2.87     0.264      1.16
## group=group2 5        2     1.13     0.673      1.16
##
##  Chisq= 1.2  on 1 degrees of freedom, p= 0.3
```

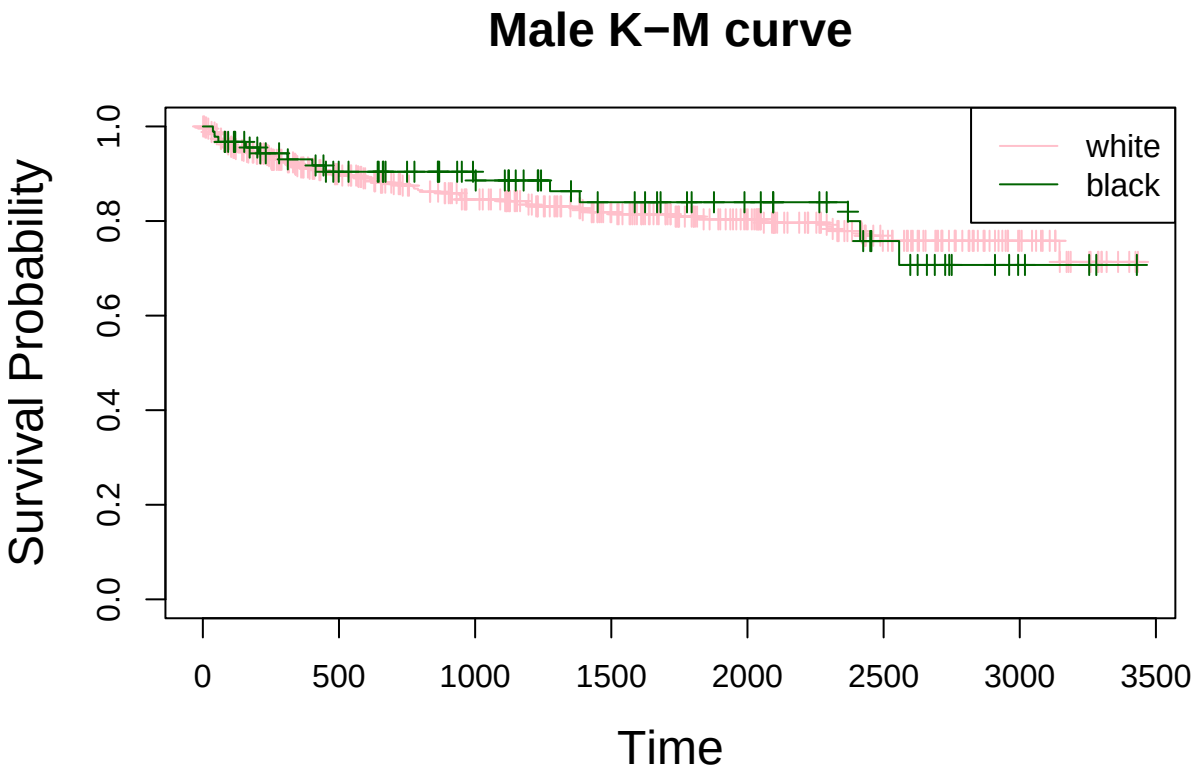
The p-value of the chi-sq test is 0.3>0.05, thus we fail to reject the null and conclude that there's no significant difference between the hazard function of the two groups.

Question 2

```
data(kidtran)
kidtran %>% head(10)
```

```
##    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
## 7    7   17     1      1    1  47
## 8    8   20     0      1    1  65
## 9    9   26     1      1    1  55
## 10  10   26     1      1    1  44
```

```
# for male group
KM1 = survfit(Surv(time, delta)~race, data = subset(kidtran, gender == 1), conf.type = "log")
plot(KM1, conf.int = F, mark.time = T,
     xlab = "Time", ylab = "Survival Probability", main = "Male K-M curve",
     col = c("pink", "dark green"),
     cex.lab = 1.5, cex.main = 1.5)
legend('topright', c("white", "black"), lty=1, col = c("pink", "dark green"))
```



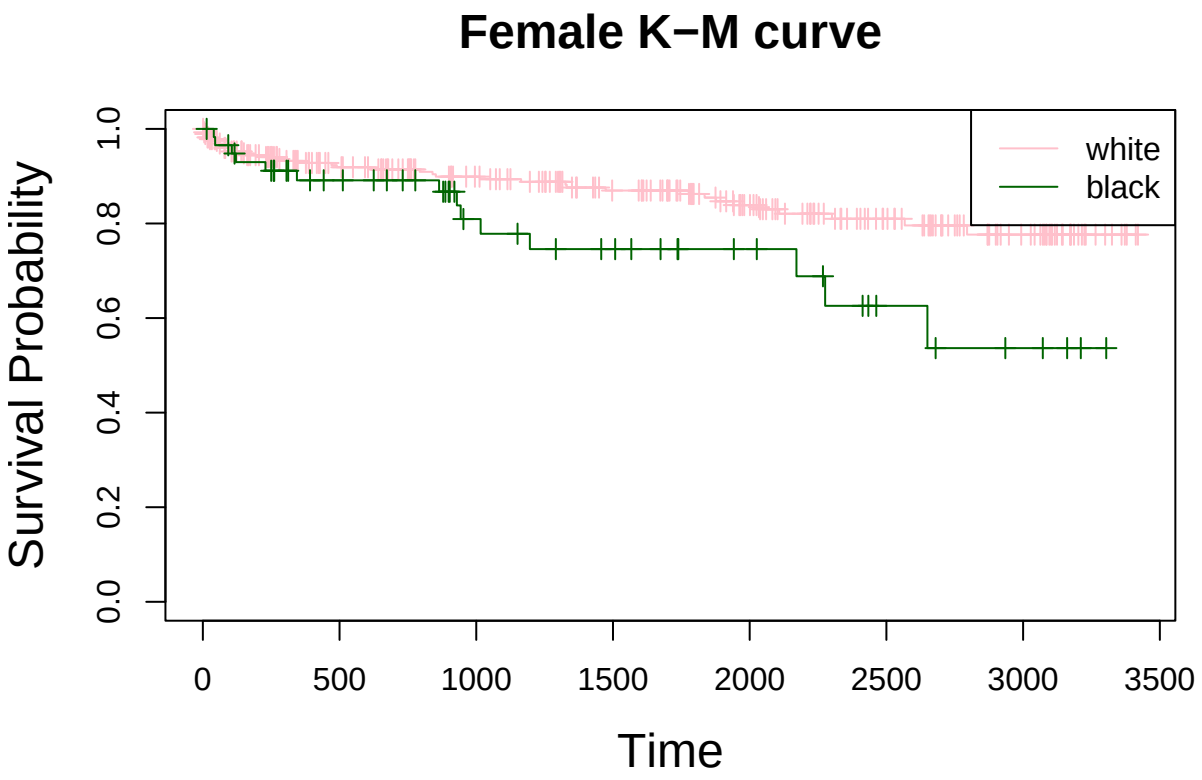

```
# log rank test
survdiff(Surv(time, delta)~race, data = subset(kidtran, gender == 1))
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ race, data = subset(kidtran,
##     gender == 1))
##
##          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
```

Among males, White people showed higher survival rates than Black people on the first half of time and lower after. The p-value from log-rank test is 0.8>0.05, thus we fail to reject the null and conclude that there's no significant difference between the survival rates of the two races among male patients.

```
# for female group
KM2 = survfit(Surv(time, delta)~race, data = subset(kidtran, gender == 2), conf.type = "log")
plot(KM2, conf.int = F, mark.time = T,
     xlab = "Time", ylab = "Survival Probability", main = "Female K-M curve",
     col = c("pink", "dark green"),
     cex.lab = 1.5, cex.main = 1.5)
legend('topright', c("white", "black"), lty=1, col = c("pink", "dark green"))
```




```
# log rank test
survdiff(Surv(time, delta)~race, data = subset(kidtran, gender == 2))
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ race, data = subset(kidtran,
##     gender == 2))
##
##          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
```

From the K-M plot we can see that White people have higher survival rates than Black people among females in most of the time.The p-value from log-rank test is 0.03<0.05, thus we reject the null and conclude that there exists significant difference between the survival rates of the two races among female patients.

Question 3

```
data(larynx)
larynx %>% head(10)
```

```
##    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
## 7      1  3.3  76     74     1
## 8      1  3.3  63     77     0
## 9      1  3.5  43     71     1
## 10     1  3.5  60     73     1
```

```
attach(larynx)
```

```
## The following object is masked _by_ .GlobalEnv:
##
##     time
```

```
Z1 = as.numeric(stage == 2)
Z2 = as.numeric(stage == 3)
Z3 = as.numeric(stage == 4)
Z4 = age
fit2 = coxph(Surv(time, delta)~Z1+Z2+Z3+Z4+Z1*Z4, data=larynx, method = 'breslow')
summary(fit2)
```

```
## Call:
## coxph(formula = Surv(time, delta) ~ Z1 + Z2 + Z3 + Z4 + Z1 *
##     Z4, data = larynx, method = "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
```

- Explanation:

Z1, Z3 and Z1*Z4 are significant variables. Holding the age, the HR between patients in stage II and patients in stage I is 0, HR between patients in stage III and patients in stage I is 1.86,HR between patients in stage IV and patients in stage I is 5.77. When a patient is in stage I, one unit increase in age would increase the probability of death by 1%.When a patient is in stage II, one unit increase in age would increse the probability of death by 12%.

- HR:

For a stage II patient of age 50, Z1=1, Z4=50; for a stage I patient of age 50, Z1=0, Z4=50. The probability of death at given time t of the former is 0.17 times the probability of death of the latter.