Problem 1

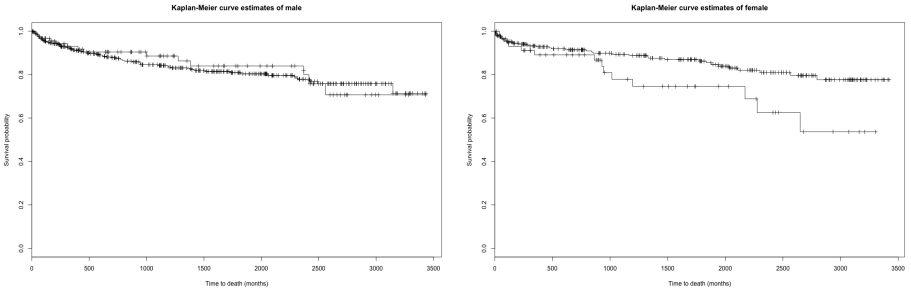
Suppose we have the following data:
Group 1: 4, 12+, 15, 21+, 23
Group 2: 2, 6+, 8+, 10, 19
Test $H_0 : h_1(t) = h_2(t)$ for all t, $H_1 : h_1(t) \neq h_2(t)$ for some t using the log-rank test.
(Optional, PhD required) Please also do the calculation by hand, and check your result.
Solution:
Based on the log-rank test result, Chisquare test statistics is 1.1 with 1 degree of freedom and P value is $0.3 > 0.05$. Therefore we fail to reject the null hypothesis at 5% level of significance and conclude the hazard rate functions are not significantly different between the two groups cross time.

event time	test statistics	variance
2	-0.5	0.25
4	0.44	0.25
10	-0.67	0.22
15	0.25	0.19
19	-0.67	0.22

Combine the information in the table, the final log rank test statistics is $Z = -1.07$, with P value 0.3.

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.
Solutions:



From the above figures, we can see that among males, the two survival curves of different races are crossed over time. The log rank test P value is 0.8, greater than 0.05. Thus there is no significant difference of survival time for white males and black males. Among females the two survival curves of different races are

different, white females has higher survival probability than black females. The log rank test P value is 0.03, meaning the difference is significant.

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 patient of age 50 as compared to a stage I patient of the same age?

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

           coef exp(coef)  se(coef)     z       p
Z1    -7.382014  0.000622  3.402754 -2.17   0.030
Z2     0.621804  1.862285  0.355808  1.75   0.081
Z3     1.753427  5.774358  0.423960  4.14 3.5e-05
Z4     0.005973  1.005991  0.014879  0.40   0.688
Z1:Z4  0.111667  1.118141  0.047673  2.34   0.019

Likelihood ratio test=24.11  on 5 df, p=2e-04
n= 90, number of events= 50
```

Stages are significant variables. Hazard ratio between stage II and stage I is almost 0 holding age constant; hazard ratio between stage III and stage I is 1.86 holding age constant; hazard ratio between stage IV and stage I is 5.77 holding age constant. Age effect is different in stage II and stage I and the difference is significant. In stage II for every unit increased in age, the hazard ratio is $exp(0.06 + 0.11) = 1.19$, in stage I for every unit increased in age, the hazard ratio is $exp(0.06) = 1.06$.
The relative risk of dying for a stage II patient of age 50 compared to a stage I patient of the same age is 0.15.

```
library(MASS)
library(survival)
library(KMsurv)
library(coxphw)

## question one
data <- data.frame( time = c(4,12,15,21,23,2,6,8,10,19),
                    event = c(1,0,1,0,1,1,0,0,1,1),
                    group = c(1,1,1,1,1,2,2,2,2,2))
fit1 <- survdiff(Surv(time, event) ~ group, data=data)

## question two
data('kidtran')
male <- subset(kidtran, gender==1)
female <- subset(kidtran, gender==2)
male$survObj <- with(male, Surv(time, delta==1))
female$survObj <- with(female , Surv(time, delta==1))

km_male <- survfit(survObj~race, conf.type="log", type="kaplan-meier", data=male)
summary(km_male)
plot(km_male , mark.time = T,
     main="Kaplan-Meier curve estimates of male",
     ylab = "Survival probability", xlab="Time to death (months)")
fit2 <- survdiff(Surv(time, delta) ~ race, data=male)

km_female <- survfit(survObj~race, conf.type="log", type="kaplan-meier", data=female)
summary(km_female)
plot(km_female , mark.time = T,
     main="Kaplan-Meier curve estimates of female",
     ylab = "Survival probability", xlab="Time to death (months)")
fit3 <- survdiff(Surv(time, delta) ~ race, data=female)

## question three
data('larynx')
Z1 <- (larynx$stage==2)*1
Z2 <- (larynx$stage==3)*1
Z3 <- (larynx$stage==4)*1
Z4 <- larynx$age
fit4 <- coxph(Surv(time,delta)~Z1+Z2+Z3+Z4 +Z1*Z4,method='breslow',data=larynx)
```