# Survival Third Assignment

## 6.2

**Plot the estimates:**

```r
data(alloauto)

haz.smooth.unif<-function(data,n.points=100, b=5)
{
  a<-survfit(Surv(data[,1],data[,2])~1, type="flem")
  H.na<--log(a$surv)
  m<-length(H.na)
  h.na<-(H.na-c(0,H.na[-m]))
  points<-(0:n.points)/n.points*max(data[,1]*(data[,2]==1))
  h.est<-NULL
  for(i in 1:n.points)
  {
    times<-a$time[abs(points[i]-a$time)<=b]
    haz<-h.na[abs(points[i]-a$time)<=b]
    h.est<-c(h.est,sum(0.5*haz)/b)
  }
  data.frame(time=points[-n.points], h.est=h.est)
}
haz.62<-haz.smooth.unif(alloauto[alloauto$type==1,c(1,3)],n.points=100,b=5)
ggplot(haz.62,aes(x=time,y=h.est))+geom_line()+labs(x="month",y="hazard")
```
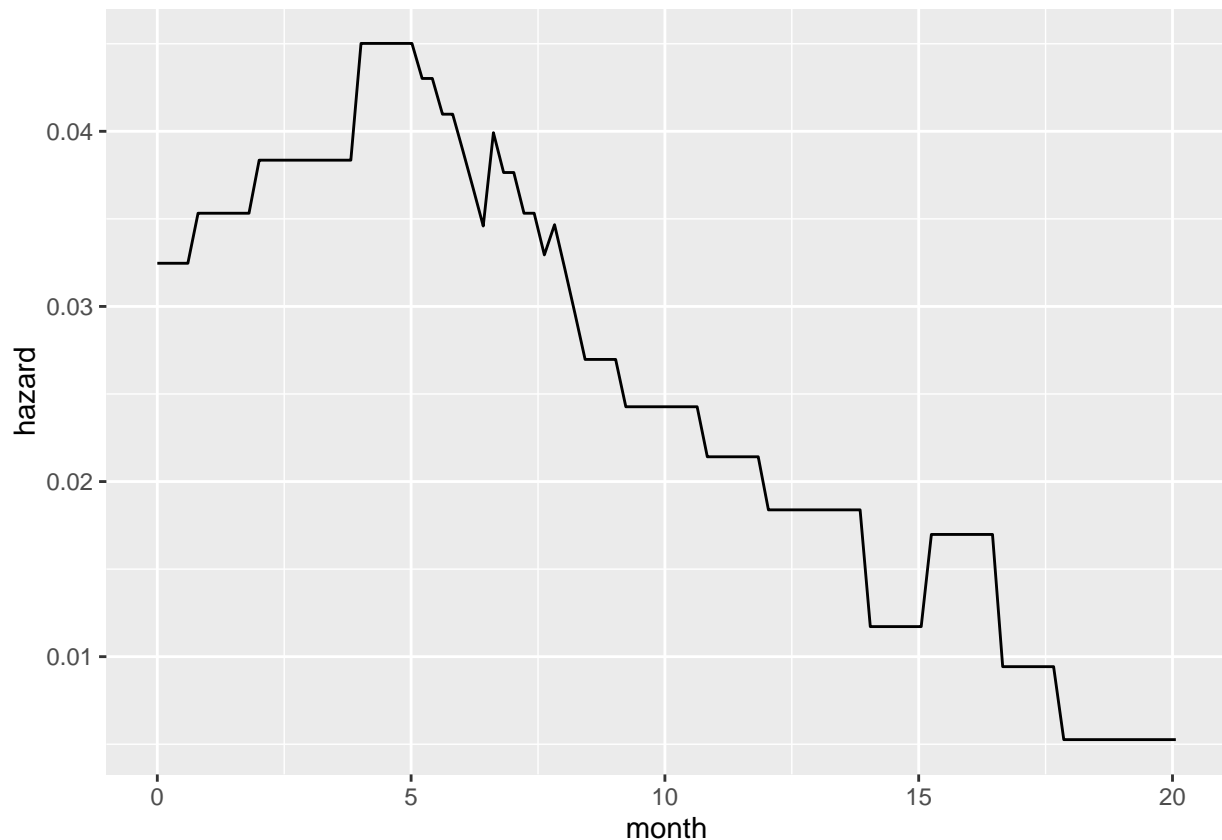
**hazard rate at given time**

h(1) = 0.0353218, h(3) = 0.0383521, h(5) = 0.0450187, h(7) = 0.0376479, h(9) = 0.0269733, h(11) = 0.0214135, h(13) = 0.0183832

**interpretation**

The hazard rate ascends from time 0 to around 5, and begins to decrease afterwards, which indicates tranplant using allo method is most risky at month 5. The situation gradually stablize after 5 months.

## 6.3

```
data("kidney")
haz63.biweight<-function(data,time.est = 3, b=5)
{
  a<-survfit(Surv(data[,1],data[,2])~1, type="flem")
  H.na<--log(a$surv)
  m<-length(H.na)
  h.na<-(H.na-c(0,H.na[-m]))

  times<-a$time[abs(time.est-a$time)<=b]
  haz<-h.na[abs(time.est-a$time)<=b]
  return(sum((1-((times-time.est)/b)^2)^2*(15/16)*haz)/b)
}

haz63.1<-haz63.biweight(kidney[kidney$type==1,c(1,2)])
haz63.2<-haz63.biweight(kidney[kidney$type==2,c(1,2)])

haz63.1.10<-haz63.biweight(kidney[kidney$type==1,c(1,2)],time.est = 10)
haz63.2.10<-haz63.biweight(kidney[kidney$type==2,c(1,2)],time.est = 10)
```

**Estimate at month 3**

$h_{type1}(3) = 0.0200272$, $h_{type2}(3) = 0.0200334$

**Estimate at month 10**

$h_{type1}(10) = 0.0387942$, $h_{type2}(10) = 0.0013934$

**note** here I do not apply boundary modification describe in §6.2, perhaps that's why estimates differ when time < 5 in ex6.2 and ex6.3.

## 7.1

```
sex<-c(0,0,0,0,1,1,1,1,1,rep(0,16),rep(1,15))###1 for male 0 for female
status<-c(rep(1,9),rep(0,31))###censoring status, 1 for death 0 for censored
age<-c(66,60,70,71,50,60,51,69,58,50,55,56,45,48,44,33,44,60,55,60,77,70,66,59,62,53,55,
                                 56,45,48,49,43,44,61,45,63,74,70,66,54)
futime<-c(74,76,77,81,59,66,69,71,71,68,72,60,55,51,55,51,50,70,60,72,80,75,70,63,63,68,62,
                                 63,51,61,55,51,54,70,60,72,80,76,72,70)
data71<-data.frame(sex,status,age,futime)
```

```
lifetable<-read.table("lifeus.txt");colnames(lifetable)<-c("age","male","female")
observe<-sum(status)
female.age<-data71[sex==0,]$age;male.age<-data71[sex==1,]$age
female.futime<-data71[sex==0,]$futime;male.futime<-data71[sex==1,]$futime
expect<-sum(-log(lifetable[female.futime-29,3]))+sum(-log(lifetable[male.futime-29,2]))-
                       sum(-log(lifetable[female.age-29,3]))-sum(-log(lifetable[male.age-29,2]))
chisq<-(observe-expect)^2/expect
chisq
```

## [1] 2.381146

```
1-pchisq(chisq,df=1)
```

## [1] 0.1228074

$H_0$: The group in interest have the same survival function as general American in 1989. $H_A$: Different survival function.

From the calculation, the statistic is 2.3811462, while under $\chi_1^2$ the p-value is 0.1228074. Conclude $H_0$.

### 7.3

```
data(burn)
survdiff(Surv(burn$T3,burn$D3)~burn$Z1,rho = 0)
```

```
## Call:
## survdiff(formula = Surv(burn$T3, burn$D3) ~ burn$Z1, rho = 0)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## burn$Z1=0 70       28     21.4      2.07      3.79
## burn$Z1=1 84       20     26.6      1.66      3.79
##
##  Chisq= 3.8  on 1 degrees of freedom, p= 0.0515
```

**a)**

$H_0 : h_0 = h_1$ for $\forall t$. $H_A : \exists t_0$ s.t $h_0(t_0) \neq h_1(t_0)$.

Chisq= 3.8 on 1 degrees of freedom, p= 0.0515. Based on the p-value calculated, reject $H_0$ on a 95% confidence level.

```
group0<-burn[burn$Z1==1,];group1<-burn[burn$Z1==0,]
fit0<-summary(survfit(Surv(group0$T3,group0$D3)~1))
fitnull<-summary(survfit(Surv(burn$T3,burn$D3)~1))
tD<-max(max(group0$T3*group0$D3),max(group1$T3*group1$D3))
D<-length(fitnull$time[fitnull$time<=tD])
di0<-rep(0,D);yi0<-rep(0,D)
for(i in 1:D){
  di0[i]<-nrow(group0[group0$T3*group0$D3==fitnull$time[i],])
  yi0[i]<-nrow(group0[group0$T3>=fitnull$time[i],])
}
numerator<-0;denominator<-0
for(i in 1:D){
  numerator<-numerator+fitnull$n.risk[i]*(di0[i]-yi0[i]*
                                          fitnull$n.event[i]/fitnull$n.risk[i])
```

```
    denominator<-denominator+fitnull$n.risk[i]^2*
      (yi0[i]/fitnull$n.risk[i])*(1-yi0[i]/fitnull$n.risk[i])*
        ((fitnull$n.risk[i]-fitnull$n.event[i])/(fitnull$n.risk[i]-1))*fitnull$n.event[i]
}
denominator<-sqrt(denominator)
Z<-numerator/denominator
Z
```

```
## [1] -1.692319
```

```
2*pnorm(Z)
```

```
## [1] 0.09058515
```

**b)**

$H_0 : h_0 = h_1$ for $\forall t$. $H_A : \exists t_0$ s.t $h_0(t_0) \neq h_1(t_0)$.

The test statistic Z = -1.6923191, p-value $= 2\Phi^{-1}(Z) = 0.0905851$, conclude $H_0$.

## 7.4

```
data(tongue);attach(tongue)
survdiff(Surv(time,delta)~type,rho = 0)
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ type, rho = 0)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## type=1 52       31     36.6     0.843      2.79
## type=2 28       22     16.4     1.873      2.79
##
##  Chisq= 2.8  on 1 degrees of freedom, p= 0.0949
```

**a)**

$H_0 : h_a = h_d$ for $\forall t$. $H_A : \exists t_0$ s.t $h_a(t_0) \neq h_d(t_0)$.

Logrank Chisq= 2.8 on 1 degrees of freedom, p= 0.0949. Based on the p-value calculated, conclude $H_0$ on a confidence level of 95%.

```
group1<-tongue[type==1,];group2<-tongue[type==2,]
fit1<-summary(survfit(Surv(group1$time,group1$delta)~1))
fitnull<-summary(survfit(Surv(tongue$time,tongue$delta)~1))
tD<-max(max(group1$time*group1$delta),max(group2$time*group2$delta))
D<-length(fitnull$time[fitnull$time<=tD])
di1<-rep(0,D);yi1<-rep(0,D)
for(i in 1:D){
  di1[i]<-nrow(group1[group1$time*group1$delta==fitnull$time[i],])
  yi1[i]<-nrow(group1[group1$time>=fitnull$time[i],])
}
numerator<-0;denominator<-0
for(i in 1:D){
  numerator<-numerator+fitnull$n.risk[i]*(di1[i]-yi1[i]*fitnull$n.event[i]/fitnull$n.risk[i])
  denominator<-denominator+fitnull$n.risk[i]^2*
```

```
     (yi1[i]/fitnull$n.risk[i])*(1-yi1[i]/fitnull$n.risk[i])*
       ((fitnull$n.risk[i]-fitnull$n.event[i])/(fitnull$n.risk[i]-1))*fitnull$n.event[i]
}
denominator<-sqrt(denominator)
Z<-numerator/denominator
Z*Z
```

```
## [1] 3.305493
```

```
1-pchisq(Z*Z,df=1)
```

```
## [1] 0.06904864
```

**b)**

$H_0 : h_a = h_d$ for $\forall t$. $H_A : h_a \neq h_d$

If the primary interest is in early time, using the Gehan test which assign heavier weight to early time is appropriate. $Z = 3.3054925$ the p-value is $0.0690486$. Conclude $H_0$ on a confidence level of 95%.

## 7.9

```
data("kidtran");attach(kidtran)
```

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

```
## The following objects are masked from tongue:
##
##      delta, time
```

```
survdiff(Surv(time,delta)~gender+race)
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ gender + race)
##
##                      N Observed Expected (O-E)^2/E (O-E)^2/V
## gender=1, race=1 432       73    69.25    0.2025    0.4013
## gender=1, race=2  92       14    14.52    0.0184    0.0205
## gender=2, race=1 280       39    47.39    1.4860    2.2531
## gender=2, race=2  59       14     8.84    3.0173    3.2245
##
##  Chisq= 4.7  on 3 degrees of freedom, p= 0.192
```

**a)**

$H_0 : h_{wm} = h_{wf} = h_{bm} = h_{bf}$. $H_A : h_i(t_0) \neq h_j(t_0)$ for at least one pair $i, j$ and time $t_0$

Logrank Chisq= 4.7 on 3 degrees of freedom, p= 0.192. Conclude $H_0$ on a confidence level of 95%.

```
male<-kidtran[gender==1,];female<-kidtran[gender==2,];detach(kidtran)
survdiff(Surv(male$time,male$delta)~male$race)
```

```
## Call:
## survdiff(formula = Surv(male$time, male$delta) ~ male$race)
```

```
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## male$race=1 432      73     71.9    0.0168     0.097
## male$race=2  92      14     15.1    0.0801     0.097
##
##  Chisq= 0.1  on 1 degrees of freedom, p= 0.755
```

```
survdiff(Surv(female$time,female$delta)~female$race)
```

```
## Call:
## survdiff(formula = Surv(female$time, female$delta) ~ female$race)
##
##                N Observed Expected (O-E)^2/E (O-E)^2/V
## female$race=1 280       39    44.79    0.748      4.85
## female$race=2  59       14     8.21    4.076      4.85
##
##  Chisq= 4.8  on 1 degrees of freedom, p= 0.0277
```

**b)**

**Test for male** $H_0 : h_{wm} = h_{bm}$. $H_A : h_{wm} \neq h_{bm}$

Logrank Chisq= 0.1 on 1 degrees of freedom, p= 0.755. Conclude $H_0$ on a confidence level of 95%.

**Test for female** $H_0 : h_{wf} = h_{bf}$. $H_A : h_{wf} \neq h_{bf}$

Logrank Chisq= 4.8 on 1 degrees of freedom, p= 0.0277. Reject $H_0$ on a confidence level of 95%.

```
attach(kidtran)
```

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

## The following objects are masked from tongue:
##
##     delta, time
```

```
survdiff(Surv(time,delta)~race+strata(gender))
```

```
## Call:
## survdiff(formula = Surv(time, delta) ~ race + strata(gender))
##
##            N Observed Expected (O-E)^2/E (O-E)^2/V
## race=1 712      112    116.7    0.188      1.13
## race=2 151       28     23.3    0.942      1.13
##
##  Chisq= 1.1  on 1 degrees of freedom, p= 0.287
```

```
detach(kidtran)
```

**Stratified test for gender** $H_0 : h_b = h_w$. $H_A : h_b > h_w$

$Z = 1.0488088$, while p-value $2(1 - \Phi(1.048)) = 0.2942661$. Conclude $H_0$ on a confidence level of 95%.

## 8.1

**a)**

$Z_1 = 1$ if the subject is in the category NHL-auto otherwise $Z_1 = 0$.

$Z_2 = 1$ if the subject is in the category HOD-allo otherwise $Z_2 = 0$.

$Z_3 = 1$ if the subject is in the category HOD-auto otherwise $Z_3 = 0$.

**b)**

$Z_1 = 1$ if the subject is in the category of HOD otherwise $Z_1 = 0$.

$Z_2 = 1$ if the subject is in the category of auto otherwise $Z_2 = 0$.

$Z_3 = Z_1 \times Z_2$, standing for interaction.

**c)**

$\beta_1 = 2$, $\beta_2 = 1.5$, $\beta_3 = -3$.

## 8.4

```r
data("bnct");attach(bnct)
```

```
## The following object is masked from tongue:
##
##      time
```

```r
Z1<-1*(trt==2);Z2<-1*(trt==3)
coxph4.1<-coxph(Surv(time,death)~Z1+Z2,method = "breslow")
coxph4.0<-coxph(Surv(time,death)~I(Z1+Z2),method = "breslow")
coxph4<-summary(coxph4.1)
detach(bnct)
```

**a)**

$b_1$ = -1.8119705, standard error = 0.559712.

$b_2$ = -3.5573706, standard error = 0.7582538.

**95% confidence interval**:

```r
coxph4$conf.int
```

```
##    exp(coef) exp(-coef)   lower .95 upper .95
## Z1 0.1633320    6.12250 0.054531004 0.4892140
## Z2 0.0285137   35.07086 0.006450996 0.1260319
```

Here we find thef 95% confidence interval for relative risk of radiated compared to untreated is [0.054531004 0.4892140]

**b)**

Likelihood ratio test= 27.37 on 2 df, p=1.139e-06

Wald test = 22.45 on 2 df, p=1.336e-05

Score (logrank) test = 31.74 on 2 df, p=1.282e-07

$H_0$: no effect, $H_A$: have effect. Reject $H_0$ on a confidence level of 95% on all tests.

```
###Z1=1 for radiated only, Z2=1 for radiated plus.
Z1<-1*(bnct[,1]==2);Z2<-1*(bnct[,1]==3)
advance<-Z1+Z2
bnct<-cbind(bnct,Z1,Z2,advance)
model0<-coxph(Surv(time,death)~advance,data=bnct,method = "breslow")
model1<-coxph(Surv(time,death)~Z1+Z2,data=bnct,method = "breslow")
LRtest<-2*(model1$loglik[2]-model0$loglik[2])
pvalue<-1-pchisq(LRtest,df=1)
LRtest
```

```
## [1] 8.439609
```

```
pvalue
```

```
## [1] 0.003671352
```

**c)**

I performed a likelihood ratio test comparing the full model to a reduced model. The statistic is 8.439609 while under $\chi^2_1$ the p-value is 0.0036714. Based on the p-value calculated, reject the $H_0$ on a confidence level of 95%.

```
Z1<-1*(bnct[,1]==1);Z2<-1*(bnct[,1]==3)
coxph4d<-coxph(Surv(time,death)~Z1+Z2,data=bnct,method = "breslow")
coxph4d<-summary(coxph4d)
coxph4d$conf.int[2,]
```

```
##     exp(coef)    exp(-coef)     lower .95     upper .95
##   0.028513702 35.070858764   0.006450996   0.126031883
```

**d)**

In order to find the relative confidence interval, I refit the model using only radiated as the baseline. The 95% confidence interval is [0.05051861,0.60327229].

```
summary(model0)
```

```
## Call:
## coxph(formula = Surv(time, death) ~ advance, data = bnct, method = "breslow")
##
##   n= 30, number of events= 27
##
##             coef exp(coef) se(coef)      z Pr(>|z|)
## advance -2.40737   0.09005  0.55440 -4.342 1.41e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##         exp(coef) exp(-coef) lower .95 upper .95
## advance   0.09005       11.1   0.03038    0.2669
##
## Concordance= 0.721  (se = 0.044 )
## Rsquare= 0.468   (max possible= 0.992 )
## Likelihood ratio test= 18.93  on 1 df,   p=1.355e-05
## Wald test            = 18.86  on 1 df,   p=1.41e-05
## Score (logrank) test = 26.17  on 1 df,   p=3.12e-07
```

**e)**

Likelihood ratio test= 18.93 on 1 df, p=1.355e-05

Reject the $H_0$ on a confidence level of 95%.

**f)**

Wald test = 18.86 on 1 df, p=1.41e-05

Reject the $H_0$ on a confidence level of 95%.

## 8.5

```
data("hodg")
HOD<-1*(hodg$dtype==2);allo<-1*(hodg$gtype==1)
Z1<-HOD*allo;Z2<-(1-HOD)*(1-allo);Z3<-HOD*(1-allo)
hodg<-cbind(hodg,Z1,Z2,Z3)
coxph5<-coxph(Surv(time,delta)~Z1+Z2+Z3,data = hodg,method = "breslow")
```

**a)**

Likelihood ratio test= 7.89 on 3 df, p=0.04825.

Reject the global hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ on a confidence level of 95%.

```
anova(coxph5)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(time, delta)
## Terms added sequentially (first to last)
##
##        loglik   Chisq Df Pr(>|Chi|)
## NULL -87.298
## Z1   -84.154  6.2879  1    0.01216 *
## Z2   -83.328  1.6514  1    0.19876
## Z3   -83.350 -0.0451  1    1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
data(hodg)
B1<-1*(hodg$gtype==2);B2<-1*(hodg$dtype==2);B3<-B1*B2
hodg<-cbind(hodg,B1,B2,B3)
coxph5b<-coxph(Surv(time,delta)~B1+B2+B3,data=hodg,method = "breslow")
```

**b)**

Likelihood ratio test= 7.89 on 3 df, p=0.04825.

Reject the global hypothesis $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ on a confidence level of 95%.

**Qusetion**

I notice that the Likelihood Ratio test for both model(8.5a,b) are identical, which confuses me, for numerical computation sedolm gets the exact same result. Below I list coeffecients, information matrix and Wald's test

result manually computed for both model. Despite difference in the former two, the test result is the same. Why this is happening? Is it because intrinsic algorithms used in R is to optimize test scores?

```
coxph5$coefficients
```

```
##       Z1       Z2       Z3
## 1.829744 0.663868 0.153713
```

```
coxph5$var
```

```
##           [,1]      [,2]      [,3]
## [1,] 0.4560590 0.2161858 0.2092039
## [2,] 0.2161858 0.3183967 0.2048986
## [3,] 0.2092039 0.2048986 0.3467307
```

```
t(coxph5$coefficients)%*%solve(coxph5$var)%*%coxph5$coefficients
```

```
##         [,1]
## [1,] 9.25845
```

```
coxph5b$coefficients
```

```
##       B1       B2       B3
##  0.663868  1.829744 -2.339899
```

```
coxph5b$var
```

```
##            [,1]      [,2]       [,3]
## [1,]  0.3183967  0.2161858 -0.3296839
## [2,]  0.2161858  0.4560590 -0.4630409
## [3,] -0.3296839 -0.4630409  0.7253530
```

```
t(coxph5b$coefficients)%*%solve(coxph5b$var)%*%coxph5b$coefficients
```

```
##         [,1]
## [1,] 9.25845
```

```
anova(coxph5b)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(time, delta)
## Terms added sequentially (first to last)
##
##        loglik  Chisq Df Pr(>|Chi|)
## NULL -87.298
## B1   -87.194 0.2070  1   0.649149
## B2   -86.995 0.3973  1   0.528481
## B3   -83.350 7.2899  1   0.006934 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**c)**

The point estimate is 1.9423, the 95% confidence interval is [0.6427, 5.870].

```
coxph5d<-coxph(Surv(time,delta)~0+Z2+Z3,data=hodg,method = "breslow")
matrix<-coxph5$var
diff<-coxph5$coefficients-c(0,coxph5d$coefficients)
```

```
Wald<-t(diff)%*%solve(matrix)%*%diff
Wald
```

```
##          [,1]
## [1,] 7.864931
```

```
1-pchisq(Wald,df=3)
```

```
##            [,1]
## [1,] 0.04888724
```

**d)**

To perform the first test comparing HOD Allo and NHL Allo. I fit the reduced model with dummy function for HOD Allo group $= 0$.

$H_0$:NHL Allo and HOD Allo have the same hazard rate. $H_A$: Different hazard rate.

Wald's test $= 7.8649306$, p-value $= 0.0488872$. Reject $H_0$.

```
####NHL auto as baseline, HOD auto D1,NHL allo D2, HOD allo D3
###Disease type (1=Non Hodgkin lymphoma, 2=Hodgkins disease)
###Graft type (1=allogenic, 2=autologous)
D1<-1*(hodg$dtype==2)*(hodg$gtype==2)
D2<-1*(hodg$dtype==1)*(hodg$gtype==1)
D3<-1*(hodg$dtype==2)*(hodg$gtype==1)
coxph5dA<-coxph(Surv(time,delta)~D1+D2+D3,data=hodg,method = "breslow")
coxph5d0<-coxph(Surv(time,delta)~0+D2+D3,data = hodg,method = "breslow")
diff<-coxph5dA$coefficients-c(0,coxph5d0$coefficients)
matrix<-coxph5dA$var
Wald<-t(diff)%*%solve(matrix)%*%diff
Wald
```

```
##          [,1]
## [1,] 1.025398
```

```
1-pchisq(Wald,df=3)
```

```
##           [,1]
## [1,] 0.7951066
```

To perform to second test comparing hazard rates between NHL Auto group and HOD Auto group. I refit the general model treating NHL Auto as baseline. In the reduced test, dummy function for HOD Auto group is set to 0.

$H_0$:NHL Auto and HOD Auto have the same hazard rate. $H_A$: Different hazard rate.

Wald's test $= 1.0253984$, p-value $= 0.7951066$. Conclude $H_0$ on a confidence level of 95%.

```
coxph5d0<-coxph(Surv(time,delta)~0+D2+D3,data = hodg,method = "breslow")
diff<-coxph5dA$coefficients-c(0,coxph5d0$coefficients)
matrix<-coxph5dA$var
Wald<-t(diff)%*%solve(matrix)%*%diff
```

**e)**

Wald $= 8.50$, while under $\chi^2_2$ the p-value is 0.014.