

Survival analysis hw03

統計碩一 105225020 wei ting sun 孫薇婷

Reproduce all results in Example 7.2 and Example 7.3 (do not use R packages .)

1.

The data showed below :

<i>Surgically Placed Catheter</i>
<i>Infection Times:</i> 1.5, 3.5, 4.5, 4.5, 5.5, 8.5, 8.5, 9.5, 10.5, 11.5, 15.5, 16.5, 18.5, 23.5, 26.5
<i>Censored Observations:</i> 2.5, 2.5, 3.5, 3.5, 3.5, 4.5, 5.5, 6.5, 6.5, 7.5, 7.5, 7.5, 7.5, 8.5, 9.5, 10.5, 11.5, 12.5, 12.5, 13.5, 14.5, 14.5, 21.5, 21.5, 22.5, 22.5, 25.5, 27.5
<i>Percutaneous Placed Catheter</i>
<i>Infection Times:</i> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 2.5, 2.5, 3.5, 6.5, 15.5
<i>Censored Observations:</i> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.5, 1.5, 1.5, 1.5, 2.5, 2.5, 2.5, 2.5, 3.5, 3.5, 3.5, 3.5, 3.5, 4.5, 4.5, 4.5, 5.5, 5.5, 5.5, 5.5, 5.5, 6.5, 7.5, 7.5, 7.5, 8.5, 8.5, 8.5, 9.5, 9.5, 10.5, 10.5, 10.5, 11.5, 11.5, 12.5, 12.5, 12.5, 12.5, 14.5, 14.5, 16.5, 16.5, 18.5, 19.5, 19.5, 19.5, 20.5, 22.5, 24.5, 25.5, 26.5, 26.5, 28.5

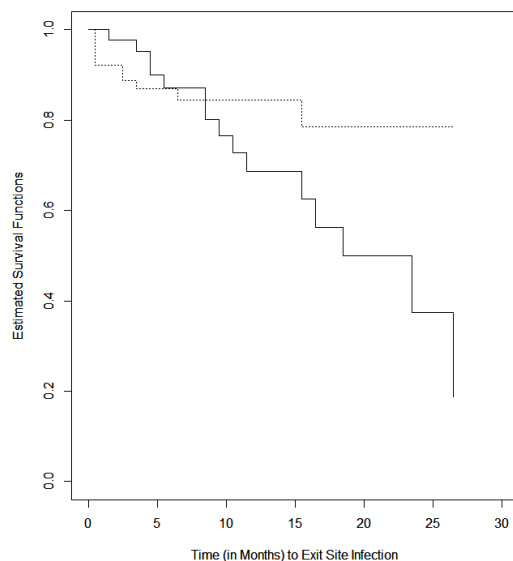
Group 1 : $(t_{i1}, \delta_{i1}), i = 1, \dots, n_1$

Group 2 : $(t_{i2}, \delta_{i2}), i = 1, \dots, n_2$

Times to infection (in months) of kidney dialysis patients with different catheterization procedures

We are interesting in testing if there is a difference in the time to cutaneous exit-site infection between patients whose catheter was placed surgically (group 1) as compared to patients who had their catheters placed percutaneously (group 2) .

(a) Figure 7.1



The picture shows the survival curves for the two samples which estimated (Infection-free) survival function for kidney dialysis patients with percutaneous (--) and surgical (—) placements of catheters .

Where

$$d_{ij} = \sum \delta_{kj} I(t_{kj} = t_i^*), j = 1 \text{ or } 2 \quad \text{number of death in group } j \text{ at } t_i^*$$

$$Y_{ij} = \sum_{k=1}^{n_i} I(t_{kj} \geq t_i^*) = \text{number of survivors in group } j \text{ at } t_i^*$$

$$d_j = \sum_{k=1}^{n_1+n_2} \delta_k I(t_k = t_j^*) = d_{i1} + d_{i2}$$

$$Y_i = \sum_{k=1}^{n_1+n_2} I(t_k \geq t_i^*) = Y_{i1} + Y_{i2}$$

(b)Table 7.2 construction of two-sample , log-rank test

t_i	Y_{i1}	d_{i1}	Y_{i2}	d_{i2}	Y_i	d_i	$Y_{i1}(\frac{d_i}{Y_i})$	$d_{i1} - Y_{i1}(\frac{d_i}{Y_i})$	$\frac{Y_{i1}}{Y_i}(1 - \frac{Y_{i1}}{Y_i})(\frac{Y_i - d_i}{Y_i - 1})d_i$
0.5	43	0	76	6	119	6	2.168	-2.168	1.326
1.5	43	1	60	0	103	1	0.417	0.583	0.243
2.5	42	0	56	2	98	2	0.857	-0.857	0.485
3.5	40	1	49	1	89	2	0.899	0.101	0.489
4.5	36	2	43	0	79	2	0.911	1.089	0.490
5.5	33	1	40	0	73	1	0.452	0.548	0.248
6.5	31	0	35	1	66	1	0.470	-0.470	0.249
8.5	25	2	30	0	55	2	0.909	1.091	0.487
9.5	22	1	27	0	49	1	0.449	0.551	0.247
10.5	20	1	25	0	45	1	0.444	0.556	0.247
11.5	18	1	22	0	40	1	0.450	0.550	0.248
15.5	11	1	14	1	25	2	0.880	0.120	0.472
16.5	10	1	13	0	23	1	0.435	0.565	0.246
18.5	9	1	11	0	20	1	0.450	0.550	0.248
23.5	4	1	5	0	9	1	0.444	0.556	0.247
26.5	2	1	3	0	5	1	0.400	0.600	0.240
SUM		15		11		26	11.036	3.964	6.211

The table above showed the calculations needed to construct the log-rank test .

Where we can see that

$H_0 =$ no difference between two groups

$H_1 =$ there is a difference between two groups

$$Z_1 = \sum \bar{W}(t_i) \left(d_{i1} - Y_{i1} \left(\frac{d_i}{Y_i} \right) \right) = 3.964$$

$$\text{Var}(Z_1) = \sum \bar{W}^2(t_i) \frac{Y_{i1}Y_{i2}}{Y_i^2} \frac{Y_i d_i}{Y_i - 1} d_i = 6.211$$

$$Z = \frac{Z_1}{\sqrt{\text{Var}(Z_1)}} = \frac{3.964}{\sqrt{6.211}} = 1.59$$

P – value = 0.1117 > $\alpha = 0.05$

Do not reject H_0 , so the log-rank test suggest that there is no difference between two procedures in the distribution of the time to exit-site infection .

```
> z
[1] 1.590442
> z1
[1] 3.963552
> var
[1] 6.210596
> z^2
[1] 2.529506
> 2*(1-pnorm(z))
[1] 0.1117352
```

(c)Table 7.3 comparison of two-sample tests

Test	$W(t_i)$	$Z_1(\tau)$	σ_{11}^2	χ^2	p-value
Log-Rank	1.0	3.964	6.211	2.53	0.112
Gehan	Y_i	-9.000	38861.81	0.002	0.964
Tarone-Ware	$Y_i^{\frac{1}{2}}$	13.203	432.831	0.403	0.526
Peto-Peto	$\bar{S}(t_i)$	2.469	4.358	1.399	0.237
Modified Peto-Peto	$\bar{S}(t_i)Y_i/(Y_i + 1)$	2.313	4.195	1.276	0.259
Fleming-Harrington p=0,q=1	$[1 - \hat{S}(t_{i-1})]$	1.413	0.207	9.668	0.002
Fleming-Harrington p=1,q=0	$\hat{S}(t_{i-1})$	2.550	4.690	1.387	0.239
Fleming-Harrington p=1,q=1	$\hat{S}(t_{i-1})[1 - \hat{S}(t_{i-1})]$	1.021	0.106	9.834	0.002
Fleming-Harrington p=0.5,q=0.5	$\hat{S}(t_{i-1})^{0.5}[1 - \hat{S}(t_{i-1})]^{0.5}$	2.470	0.657	9.285	0.002
Fleming-Harrington p=0.5,q=2	$\hat{S}(t_{i-1})^{0.5}[1 - \hat{S}(t_{i-1})]^2$	0.324	0.013	8.179	0.004

To further investigate these two treatments , we shall apply some of the other weight functions . The table summarizes the results of the tests .

We defined the common survival function as $\bar{S}(t_i) = \prod(1 - \frac{d_i}{Y_{i+1}})$.

$$\hat{S}(t) = \prod(1 - \frac{d_i}{Y_i})$$

The Gehan weights may have misleading results when the censoring patterns are different in the individual samples .

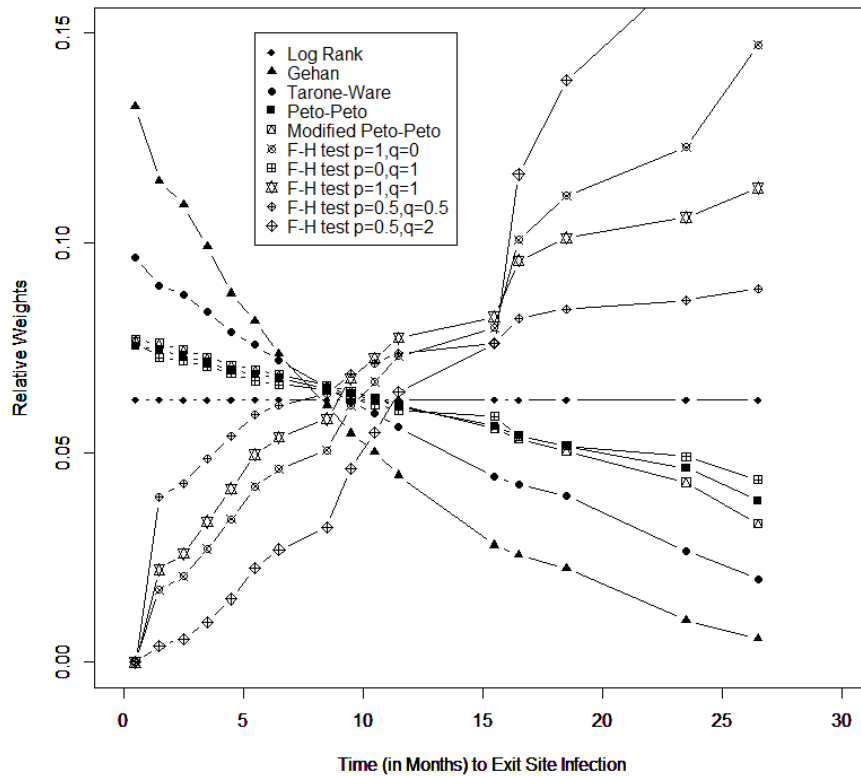
Fleming-Harrington : weighted function is

$$W(t_i) = \hat{S}(t_{i-1})^p [1 - \hat{S}(t_{i-1})]^q, p \geq 0, q \geq 0$$

When p=q=0 , we can lead to log-rank test ; when p=1, q=0, we can lead to Mann–Whitney–Wilcoxon test ; when q=0 and p>0 , these weights give the most weight to early departures between the hazard rates in the K populations, whereas, when p=0 and q>0, these tests give most weight to departures which occur late in time.

So we need to choose the appropriate p and q .

(d)Figure 7.2 Relative weights for comparison of observed and expected numbers of deaths for kidney dialysis patients .



The figure shows the relative weights these tests give to the comparisons at each time point , where we plot with (t_i^*, w_i) , $w_i = W(t_i) / \sum_{i=1}^D W(t_i)$. We can see that Gehan’s weight function gives very heavy weight to early comparisons at $t_i^* = 0.5$ and it leads to a negative test statistic .

The Fleming-Harrington tests put more weight on the late comparison and lead to significance tests because the two survival curves diverge for larger values of t .

The picture I draw is not similar to text book because the text book is wrong , it mistake to draw Fleming-Harrington p=0.5,q=2 as p=2,q=2 , so it does not look the same .

w1	w2	w3	w4	w5	w6	w7	w8	w9	w10
0.06250	0.13252	0.09662	0.07530	0.07689	0.00000	0.07645	0.00000	0.00000	0.00000
0.06250	0.11470	0.08989	0.07458	0.07606	0.01727	0.07260	0.02209	0.03937	0.00396
0.06250	0.10913	0.08768	0.07307	0.07448	0.02043	0.07189	0.02587	0.04261	0.00552
0.06250	0.09911	0.08356	0.07145	0.07275	0.02700	0.07042	0.03350	0.04849	0.00954
0.06250	0.08797	0.07873	0.06966	0.07084	0.03409	0.06884	0.04135	0.05387	0.01504
0.06250	0.08129	0.07568	0.06872	0.06981	0.04190	0.06710	0.04953	0.05896	0.02243
0.06250	0.07350	0.07196	0.06769	0.06867	0.04602	0.06618	0.05365	0.06136	0.02687
0.06250	0.06125	0.06569	0.06527	0.06602	0.05051	0.06518	0.05800	0.06380	0.03213
0.06250	0.05457	0.06200	0.06397	0.06456	0.06113	0.06281	0.06764	0.06890	0.04619
0.06250	0.05011	0.05942	0.06258	0.06304	0.06687	0.06152	0.07248	0.07132	0.05472
0.06250	0.04454	0.05602	0.06105	0.06134	0.07299	0.06016	0.07736	0.07368	0.06447
0.06250	0.02784	0.04429	0.05636	0.05580	0.07973	0.05865	0.08239	0.07604	0.07595
0.06250	0.02561	0.04248	0.05401	0.05330	0.10075	0.05396	0.09579	0.08199	0.11633
0.06250	0.02227	0.03961	0.05144	0.05044	0.11126	0.05162	0.10118	0.08426	0.13875
0.06250	0.01002	0.02657	0.04629	0.04290	0.12283	0.04903	0.10611	0.08629	0.16481
0.06250	0.00557	0.01981	0.03858	0.03310	0.14723	0.04359	0.11306	0.08907	0.22327

The relative weight I calculated is on above .

2.

I download the data from package-boot , first we separated the data into two groups –male and female .

These data are left-truncated by the individual's entry time into the retirement center . We now try to test the hypothesis that females live longer than males .

$$H_0: h_F(t) = h_M(t) , 777months \leq t \leq 1152$$

$$H_1: h_F(t) < h_M(t) , for all t \in [777,1152]$$

We can compute the Y_{iM} and Y_{iF} as the number of male and female , on the other hand , who is in the center at age t_i .

Here we use the log-rank test , where

$$Z_1 = \sum \bar{W}(t_i) \left(d_{i1} - Y_{i1} \left(\frac{d_i}{Y_i} \right) \right) = 9.692915$$

$$\text{Var}(Z_1) = \sum \bar{W}^2(t_i) \frac{Y_{i1}Y_{i2}}{Y_i^2} \frac{Y_i d_i}{Y_i - 1} d_i = 28.21195$$

$$Z = \frac{Z_1}{\sqrt{\text{Var}(Z_1)}} = \frac{9.692915}{\sqrt{28.21195}} = 1.824895$$

$$P - \text{value} = 0.03400844 < \alpha = 0.05$$

So the log-rank test suggest us to reject H_0 , which means we have enough evidence that males are dying at a faster rate than females .

```
> z1
[1] 9.692915
> var
[1] 28.21195
> z
[1] 1.824895
> 1-pnorm(z)
[1] 0.03400844
```

<code>

```
##figure 7.1
data1=c(1.5,3.5,4.5,4.5,5.5,8.5,8.5,9.5,10.5,11.5,15.5,16.5,18.5,23.5,26.5)
data1_c=c(2.5,2.5,3.5,3.5,3.5,4.5,5.5,6.5,6.5,7.5,7.5,7.5,7.5,8.5,9.5,
          10.5,11.5,12.5,12.5,13.5,14.5,14.5,21.5,21.5,22.5,22.5,25.5,27.5)
cens1=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,1)
time1=c(data1,data1_c)
dataa1=cbind(time1,cens1)

data2=c(0.5,0.5,0.5,0.5,0.5,0.5,2.5,2.5,3.5,6.5,15.5)
data2_c=c(0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,1.5,1.5,1.5,1.5,2.5,
          2.5,2.5,2.5,2.5,3.5,3.5,3.5,3.5,3.5,4.5,4.5,4.5,5.5,5.5,5.5,5.5,6.5,7.5,7.5,
          7.5,8.5,8.5,8.5,9.5,9.5,10.5,10.5,10.5,11.5,11.5,12.5,12.5,12.5,12.5,14.5,14.5,
          16.5,16.5,18.5,19.5,19.5,19.5,20.5,22.5,24.5,25.5,26.5,26.5,28.5)
cens2=c(0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
          ,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
          ,1,1,1,1,1,1,1,1)
time2=c(data2,data2_c)
dataa2=cbind(time2,cens2)

t=unique(sort(c(data1,data2)))
length(t)

y1=1:16 #number of unique death times
for(i in 1:16)
{
  y1[i]=length(time1[t[i]<=time1])
}
y1

d1=1:16
for(i in 1:16)
{
  d1[i]=sum(time1==t[i]&cens1==0)
```

```
}  
d1  
  
y2=1:16 #number of unique death times  
for(i in 1:16)  
{  
  y2[i]=length(time2[t[i]<=time2])  
}  
y2  
  
d2=1:16  
for(i in 1:16)  
{  
  d2[i]=sum(time2==t[i]&cens2==0)  
}  
d2  
  
st1=cumprod(1-d1/y1)  
st1=c(1,st1)  
t1=c(0,t)  
st2=cumprod(1-d2/y2)  
st2=c(1,st2)  
t2=c(0,t)  
plot(t1,st1,type='s',xlim=c(0,30),ylim=c(0,1),xlab='Time (in Months) to Exit Site  
Infection',ylab='Estimated Survival Functions')  
par(new=TRUE)  
plot(t2,st2,type='s',xlim=c(0,30),ylim=c(0,1),xlab='Time (in Months) to Exit Site  
Infection',ylab='Estimated Survival Functions',lty=9)  
  
##table 7.2  
data1=c(1.5,3.5,4.5,4.5,5.5,8.5,8.5,9.5,10.5,11.5,15.5,16.5,18.5,23.5,26.5)  
data1_c=c(2.5,2.5,3.5,3.5,3.5,4.5,5.5,6.5,6.5,7.5,7.5,7.5,7.5,8.5,9.5,  
          10.5,11.5,12.5,12.5,13.5,14.5,14.5,21.5,21.5,22.5,22.5,25.5,27.5)  
cens1=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,  
1,1)  
time1=c(data1,data1_c)  
dataa1=cbind(time1,cens1)
```



```

data2=c(0.5,0.5,0.5,0.5,0.5,0.5,2.5,2.5,3.5,6.5,15.5)
data2_c=c(0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,1.5,1.5,1.5,1.5,2.5,
2.5,2.5,2.5,2.5,3.5,3.5,3.5,3.5,4.5,4.5,4.5,5.5,5.5,5.5,5.5,6.5,7.5,7.5,
7.5,8.5,8.5,8.5,9.5,9.5,10.5,10.5,10.5,11.5,11.5,12.5,12.5,12.5,12.5,14.5,14.5,
16.5,16.5,18.5,19.5,19.5,19.5,20.5,22.5,24.5,25.5,26.5,26.5,28.5)
cens2=c(0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
,1,1,1,1,1,1,1,1)
time2=c(data2,data2_c)
dataa2=cbind(time2,cens2)

t=unique(sort(c(data1,data2)))

y1=1:16 #number of unique death times
for(i in 1:16)
{
  y1[i]=length(time1[t[i]<=time1])
}

d1=1:16
for(i in 1:16)
{
  d1[i]=sum(time1==t[i]&cens1==0)
}

y2=1:16 #number of unique death times
for(i in 1:16)
{
  y2[i]=length(time2[t[i]<=time2])
}

d2=1:16
for(i in 1:16)
{
  d2[i]=sum(time2==t[i]&cens2==0)
}

```

```

}

y=y1+y2

d=d1+d2

t
y1
d1
y2
d2
y
d
y1*(d/y)
d1-y1*(d/y)
y1/y*(1-y1/y)*(y-d)/(y-1)*d
sum(y1*(d/y))
z1=sum(d1-y1*(d/y))
var=sum(y1/y*(1-y1/y)*(y-d)/(y-1)*d)
z=z1/sqrt(var)
z
chisq=z^2
2*(1-pnorm(z))

##table 7.3
##log-rank
z1=sum(d1-y1*(d/y))
var=sum(y1/y*(1-y1/y)*(y-d)/(y-1)*d)
z=z1/sqrt(var)
z1
var
z^2
2*(1-pnorm(z))

## Gehan
z1=sum(y*(d1-y1*(d/y)))
var=sum(y^2*(y1/y*(1-y1/y)*(y-d)/(y-1)*d))
z=z1/sqrt(var)

```

```

z1
var
z^2
2*(pnorm(z))

##tarone-ware
z1=sum(y^(1/2)*(d1-y1*(d/y)))
var=sum(y*(y1/y*(1-y1/y)*(y-d)/(y-1)*d))
z=z1/sqrt(var)
z1
var
z^2
2*(1-pnorm(z))

##peto-peto
st=cumprod(1-d/(y+1))
z1=sum(st*(d1-y1*(d/y)))
var=sum(st^2*(y1/y*(1-y1/y)*(y-d)/(y-1)*d))
z=z1/sqrt(var)
z1
var
z^2
2*(1-pnorm(z))

##modified peto-peto
st=cumprod(1-d/(y+1))
z1=sum(st*y/(y+1)*(d1-y1*(d/y)))
var=sum((st*y/(y+1))^2*(y1/y*(1-y1/y)*(y-d)/(y-1)*d))
z=z1/sqrt(var)
z1
var
z^2
2*(1-pnorm(z))

##fleming-harrington (p=0,q=1)
p=0
q=1
st=cumprod(1-d/y)

```

```

st=c(1,st)
st=st[1:16]
wt=st^(p)*(1-st)^(q)
z1=sum(wt*(d1-y1*(d/y)))
var=sum(wt^2*(y1/y*(1-y1/y)*(y-d)/(y-1)*d))
z=z1/sqrt(var)
z1
var
z^2
2*(1-pnorm(z))

###fleming-harrington (p=1,q=0)
p=1
q=0
st=cumprod(1-d/y)
st=c(1,st)
st=st[1:16]
wt=st^(p)*(1-st)^(q)
z1=sum(wt*(d1-y1*(d/y)))
var=sum(wt^2*(y1/y*(1-y1/y)*(y-d)/(y-1)*d))
z=z1/sqrt(var)
z1
var
z^2
2*(1-pnorm(z))

###fleming-harrington (p=1,q=1)
p=1
q=1
st=cumprod(1-d/y)
st=c(1,st)
st=st[1:16]
wt=st^(p)*(1-st)^(q)
z1=sum(wt*(d1-y1*(d/y)))
var=sum(wt^2*(y1/y*(1-y1/y)*(y-d)/(y-1)*d))
z=z1/sqrt(var)
z1
var

```

```

z^2
2*(1-pnorm(z))

##fleming-harrington (p=0.5,q=0.5)
p=0.5
q=0.5
st=cumprod(1-d/y)
st=c(1,st)
st=st[1:16]
wt=st^(p)*(1-st)^(q)
z1=sum(wt*(d1-y1*(d/y)))
var=sum(wt^2*(y1/y*(1-y1/y)*(y-d)/(y-1)*d))
z=z1/sqrt(var)
z1
var
z^2
2*(1-pnorm(z))

##fleming-harrington (p=0.5,q=2)
p=0.5
q=2
st=cumprod(1-d/y)
st=c(1,st)
st=st[1:16]
wt=st^(p)*(1-st)^(q)
z1=sum(wt*(d1-y1*(d/y)))
var=sum(wt^2*(y1/y*(1-y1/y)*(y-d)/(y-1)*d))
z=z1/sqrt(var)
z1
var
z^2
2*(1-pnorm(z))
#####
### text book result (wrong!!!!!!!!!!!!)
w=c()
a1=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
w1=a1/sum(a1)
w2=y/sum(y)

```

```

w3=(y^0.5)/sum(y^0.5)

st1=cumprod(1-d/(y+1))
w4=st1/sum(st1)
w5=(st1*y/(y+1))/sum(st1*y/(y+1))

st2=cumprod(1-d/y)
st2=c(1,st2)
st2=st[1:16]
w6=(1-st2)/sum(1-st2)
w7=st2/sum(st2)
w8=(st2*(1-st2))/sum(st2*(1-st2))
w9=(st2^(0.5)*(1-st2)^(0.5))/sum(st2^(0.5)*(1-st2)^(0.5))
w10=(st2^(0.5)*(1-st2)^(2))/sum(st2^(0.5)*(1-st2)^(2))
w=cbind(w1,w2,w3,w4,w5,w6,w7,w8,w9,w10)

plot(t,w1,xlim=c(0,30),ylim=c(0,0.15),type='b',pch=18,xlab='Time (in Months) to Exit
Site Infection',ylab='Relative Weights')
par(new=TRUE)
plot(t,w2,xlim=c(0,30),ylim=c(0,0.15),type='b',pch=17,xlab='Time (in Months) to Exit
Site Infection',ylab='Relative Weights')
par(new=TRUE)
plot(t,w3,xlim=c(0,30),ylim=c(0,0.15),type='b',pch=16,xlab='Time (in Months) to Exit
Site Infection',ylab='Relative Weights')
par(new=TRUE)
plot(t,w4,xlim=c(0,30),ylim=c(0,0.15),type='b',pch=15,xlab='Time (in Months) to Exit
Site Infection',ylab='Relative Weights')
par(new=TRUE)
plot(t,w5,xlim=c(0,30),ylim=c(0,0.15),type='b',pch=14,xlab='Time (in Months) to Exit
Site Infection',ylab='Relative Weights')
par(new=TRUE)
plot(t,w6,xlim=c(0,30),ylim=c(0,0.15),type='b',pch=13,xlab='Time (in Months) to Exit
Site Infection',ylab='Relative Weights')
par(new=TRUE)
plot(t,w7,xlim=c(0,30),ylim=c(0,0.15),type='b',pch=12,xlab='Time (in Months) to Exit
Site Infection',ylab='Relative Weights')
par(new=TRUE)
plot(t,w8,xlim=c(0,30),ylim=c(0,0.15),type='b',pch=11,xlab='Time (in Months) to Exit

```

```

Site Infection',ylab='Relative Weights')
par(new=TRUE)
plot(t,w9,xlim=c(0,30),ylim=c(0,0.15),type='b',pch=10,xlab='Time (in Months) to Exit
Site Infection',ylab='Relative Weights')
par(new=TRUE)
plot(t,w10,xlim=c(0,30),ylim=c(0,0.15),type='b',pch=9,xlab='Time (in Months) to Exit
Site Infection',ylab='Relative Weights')
legend(5.5,0.15,legend=c("Log Rank","Gehan","Tarone-Ware","Peto-Peto","Modified
Peto-Peto","F-H test p=1,q=0","F-H test p=0,q=1","F-H test p=1,q=1","F-H test
p=0.5,q=0.5","F-H test p=0.5,q=2"),pch=c(18,17,16,15,14,13,12,11,10,9))

##2.
library(boot)
data=channing

data_m=data[data$sex=="Male",]
data_f=data[data$sex=="Female",]

t=unique(sort(c(data$exit,data$entry)))
length(t)

y1=1:315 #number of unique death times
for(i in 1:315)
{
  y1[i]=length(data_m$entry[data_m$entry<=t[i]&t[i]<=data_m$exit])
}
y1

d1=1:315
for(i in 1:315)
{
  d1[i]=sum(data_m$exit==t[i]&data_m$cens==1)
}
d1

y2=1:315 #number of unique death times
for(i in 1:315)

```

```

{
  y2=length(data_f$entry[data_f$entry<=t[i]&t[i]<=data_f$exit])
}
y2

d2=1:315
for(i in 1:315)
{
  d2[i]=sum(data_f$exit==t[i]&data_f$cens==1)
}
d2

y=y1+y2
d=d1+d2

y1*(d/y)
d1-y1*(d/y)
yyy=y1/y*(1-y1/y)*(y-d)/(y-1)*d
yyy=yyy[2:314]
sum(y1*(d/y))

z1=sum(d1-y1*(d/y))
var=sum(yyy)
z=z1/sqrt(var)
z
z1
var
z^2
1-pnorm(z)

```