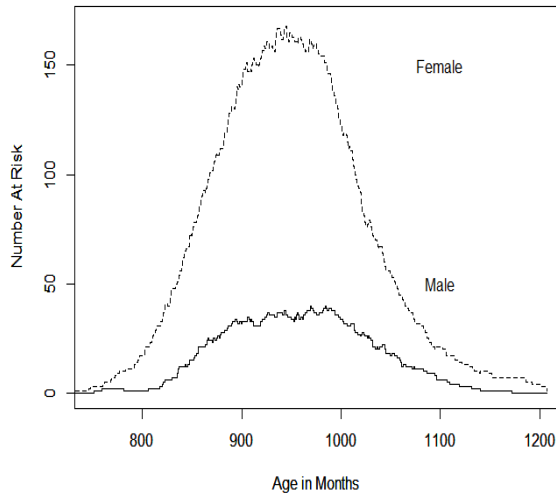# Survival analysis hw02
105225020 wei ting sun

1. read example 4.3 and reproduce Fig 4.10 and 4.11

(detailed calculations must be shown)

(a) Fig 4.10

I download the package (boot) for the <u>channing house data</u> .

The picture is where I separated the data into two group (male),(female) , and then draw the number of risk Y_i , but with the same time t_i^* .

The plot shows the number of individuals at risk as a function of the age at which individuals die for both males and females .

We can notice that the number at risk initially increases as more individuals enter into the study cohort and that this number decreases for later ages as individuals die or are censored .

```
> y1
  [1]   0   0   0   1   2   2   2   2   2   2   2   2   1   1   1   1   1   1   1   1   1   1   1   1   1   1   2   2   2
 [29]   2   2   2   2   2   2   2   3   3   3   4   5   5   6   6   6   6   6   6   7   7   7   7   9  11  12  12  12
 [57]  12  12  12  13  12  12  13  15  15  15  15  15  16  17  18  18  21  21  21  21  21  21  22  22  24  25  24  24
 [85]  24  23  25  25  24  24  25  26  25  27  28  28  28  29  30  31  31  31  31  32  33  33  34  34  33  32  32  33
[113]  32  33  33  33  33  33  35  34  33  34  33  33  31  31  31  31  31  31  31  33  33  34  34  35  35  36  37  37
[141]  36  36  36  36  36  35  35  35  37  37  37  37  36  36  38  36  36  35  35  35  34  34  33  35  34  36  37  36
[169]  34  35  35  34  36  36  37  37  38  38  38  40  39  39  38  37  36  36  36  37  37  38  39  39  40  39  38  38
[197]  39  39  38  38  38  38  36  36  36  35  35  34  34  34  33  32  32  32  31  31  31  31  32  32  32  29  28  28
[225]  28  27  27  26  27  28  28  27  26  26  25  26  25  25  24  24  21  21  20  20  21  20  20  21  21  22  22  22
[253]  20  18  18  18  17  17  18  18  17  17  16  16  16  15  14  12  12  13  13  12  12  12  11  11  12  11  11  10
[281]  10  10   9   9   9   9   9   8   6   6   6   6   6   5   4   4   4   4   3   3   3   2   2   2   2   1   1   1
[309]   1   1   0   0   0   0   0   0
```

```
> y2
  [1]    1    2    3    3    3    4    5    6    7    8    9   10   10   10   11   13   14   15   16   17   17
 [22]   17   21   21   21   21   22   23   24   26   28   28   29   30   31   31   33   33   35   37   38   41
 [43]   41   40   41   45   47   47   48   49   50   51   52   54   56   58   61   62   64   65   66   67   67
 [64]   72   73   72   74   76   78   78   80   81   82   85   85   89   90   93   92   94   96   98   99  101
 [85]  101  102  104  106  107  109  110  109  110  112  112  113  116  119  122  125  128  129  129  131  131
[106]  130  131  136  140  139  141  140  141  144  148  148  149  151  147  147  148  147  150  151  153  151
[127]  150  150  149  150  151  153  155  157  159  158  159  161  161  160  158  158  159  158  157  156  162
[148]  165  167  166  164  162  162  163  167  168  163  162  161  165  164  163  163  161  161  161  160  163
[169]  163  162  160  160  158  158  156  157  156  159  162  162  160  158  159  160  158  157  155  154  154
[190]  154  154  151  151  150  147  146  146  144  141  138  138  136  136  132  130  126  125  123  122  120
[211]  118  119  119  118  115  112  113  111  111  107  104  102  100   98   96   95   94   92   86   83   82
[232]   81   78   76   77   79   79   78   76   74   72   71   70   70   69   67   66   67   67   64   62   60
[253]   58   56   56   56   55   54   53   52   51   49   48   46   45   45   44   44   43   42   40   39   38
[274]   37   37   35   34   32   31   30   29   28   25   24   23   22   21   20   21   21   20   19   17   17
[295]   17   16   15   14   14   13   12   12   11   10    9   10   10    9    8    7    7    5    4    3    1
```

(b)Fig 4.11
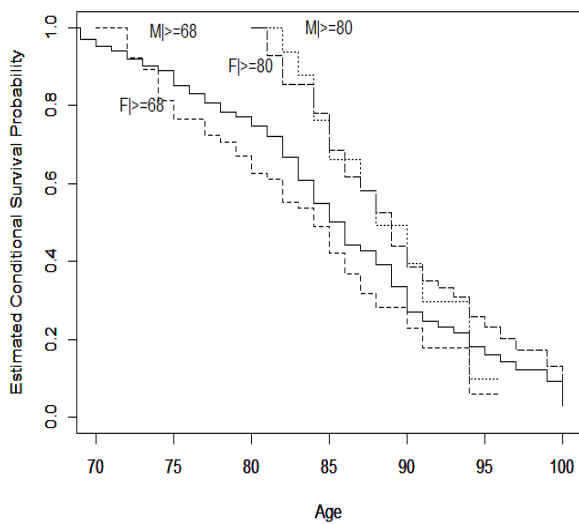
The picture in the left is I separated the data into four group (entry time>=68 & male) , (entry time>=68 & female) , (entry time>=80 & male),(entry time>=80 & female) , and then draw the survival function , that is what we called the conditional survival function.

We can also use the package 'survival' to draw the conditional survival function plot .

The plot shows the estimated probability of surviving beyond age t , given survival to 68 or 80 years for both males and females .

Where

$Y_i = the\ number\ of\ subjects\ at\ risk , i = 1, … ,4$

$d_i = the\ number\ of\ death , i = 1, … ,4$

$$\widehat{S(t_i^*)} = \prod_{j=1}^{i} (1 - \frac{d_i}{Y_i}) , i = 1, … ,4$$

```
> y1
 [1] 14 26 30 34 34 35 36 40 41 44 42 41 37 34 29 24 22 17 13
[20] 11  9  6  4  3  1
> d1
 [1] 0 2 1 3 2 0 2 1 2 3 1 4 1 3 4 3 3 2 0 2 2 0 0 0 2 0
> y2
 [1] 10 16 16 15 15 16 17 13 11 10  8  6  4  3  1
> d2
 [1] 0 1 1 2 2 0 2 2 0 2 2 0 0 2 0
> y3
 [1]  15  34  53  71  92 111 131 146 157 176 172 178 173 171
[15] 160 138 120  91  78  59  48  41  31  22  19  14  12  10
[29]   8   7   5   4   3
> d3
 [1]  0  1  1  1  2  2  2  6  4  5  5  3  5  6 12 12 12  8  9
[20]  2  4  6  6  2  1  1  2  1  1  1  0  1  2
> y4
 [1] 14 28 37 40 47 41 40 35 31 30 25 22 19 14 12 10  8  7  5
[20]  4  3
> d4
 [1] 0 2 3 0 4 5 4 2 3 5 3 2 1 1 2 1 1 1 0 1 2

> s_t1
 [1] 1.00000000 0.92307692 0.89230769 0.81357466 0.76571733
 [6] 0.76571733 0.72317748 0.70509804 0.67070301 0.62497326
[11] 0.61009295 0.55057168 0.53569137 0.48842448 0.42105559
[16] 0.36842364 0.31818405 0.28075063 0.28075063 0.22970506
[21] 0.17865949 0.17865949 0.17865949 0.05955316 0.05955316
> s_t2
 [1] 1.00000000 0.93750000 0.87890625 0.76171875 0.66015625
 [6] 0.66015625 0.58249081 0.49287684 0.49287684 0.39430147
[11] 0.29572610 0.29572610 0.29572610 0.09857537 0.09857537
> s_t3
 [1] 1.00000000 0.97058824 0.95227525 0.93886292 0.91845286
 [6] 0.90190416 0.88813463 0.85163595 0.82993821 0.80636042
[11] 0.78291971 0.76972444 0.74747807 0.72125077 0.66715696
[16] 0.60914331 0.54822898 0.50003303 0.44233691 0.42734244
[21] 0.39173057 0.33440414 0.26968076 0.24516433 0.23226094
[26] 0.21567087 0.17972573 0.16175316 0.14153401 0.12131487
[31] 0.12131487 0.09098615 0.03032872
> s_t4
 [1] 1.00000000 0.92857143 0.85328185 0.85328185 0.78066212
 [6] 0.68545942 0.61691348 0.58166128 0.52537148 0.43780957
[11] 0.38527242 0.35024765 0.33181357 0.30811260 0.25676050
[16] 0.23108445 0.20219889 0.17331334 0.17331334 0.12998500
[21] 0.04332833
```

2. Exercise 4.7

Consider a hypothetical study of the mortality experience of diabetics .

Thirty diabetics subjects are recruited at a clinic and followed until death or the end of the study . The subject's age at entry the study and their age at the end of study or death are given in the below . Of interest is estimating the survival curve for a 60- or for 70-year-old diabetic .

(a)(b)(c) <u>data is left truncated</u>

Left Truncated					
Time	Yi	Di	S(t|Alive at 60)	Di	S(t|Alive at 70)
58	2	0	1.0000000	0	1.0000000
59	3	0	1.0000000	0	1.0000000
60	5	1	0.8000000	0	1.0000000
61	6	0	0.8000000	0	1.0000000
62	9	1	0.7111111	0	1.0000000
63	10	1	0.6400000	0	1.0000000
64	10	0	0.6400000	0	1.0000000
65	10	2	0.5120000	0	1.0000000
66	10	1	0.4608000	0	1.0000000
67	12	0	0.4608000	0	1.0000000
68	13	2	0.3899077	0	1.0000000
69	14	2	0.3342066	0	1.0000000
70	13	2	0.2827902	2	0.8461538
71	12	2	0.2356585	2	0.7051282
72	12	2	0.1963821	2	0.5876068
73	11	1	0.1785292	1	0.5341880
74	9	1	0.1586926	1	0.4748338
76	7	1	0.1360222	1	0.4070004
77	5	1	0.1088178	1	0.3256003
78	4	0	0.1088178	0	0.3256003
79	3	0	0.1088178	0	0.3256003
80	1	0	0.1088178	0	0.3256003

I use R to separate the data into two group (exit age>=60) and (exit age >=70) , then calculate survival function at the left table .

Where

Y_i = *the number of subjects at risk*

d_i = *the number of death*

$$\widehat{S(t_i^*)} = \prod_{j=1}^{i}(1 - \frac{d_i}{Y_i})$$

(d) <u>data is right censored</u>

No Truncated					
Time	Yi	Di	S(t|Alive at 60)	Di	S(t|Alive at 70)
58	30	0	1.0000000	0	1.0000000
59	30	0	1.0000000	0	1.0000000
60	30	1	0.9666667	0	1.0000000
61	29	0	0.9666667	0	1.0000000
62	29	1	0.9333333	0	1.0000000
63	28	1	0.9000000	0	1.0000000
64	27	0	0.9000000	0	1.0000000
65	27	2	0.8333333	0	1.0000000
66	25	1	0.8000000	0	1.0000000
67	24	0	0.8000000	0	1.0000000
68	24	2	0.7333333	0	1.0000000
69	22	2	0.6666667	0	1.0000000
70	18	2	0.5925926	2	0.8888889
71	16	2	0.5185185	2	0.7777778
72	14	2	0.4444444	2	0.6666667
73	11	1	0.4040404	1	0.6060606
74	9	1	0.3591470	1	0.5387205
76	7	1	0.3078403	1	0.4617605
77	5	1	0.2462722	1	0.3694084
78	4	0	0.2462722	0	0.3694084
79	3	0	0.2462722	0	0.3694084
80	1	0	0.2462722	0	0.3694084

Y_i = *the number of subjects at risk*

d_i = *the number of death*

$$\widehat{S(t_i^*)} = \prod_{j=1}^{i}(1 - \frac{d_i}{Y_i})$$

But the data is not truncated anymore , we treated as censored data . The result showed at left table .

3. Exercise 4.8

The problem to compute the conditional survival function given that the patient entered the hospital at age at least 30.
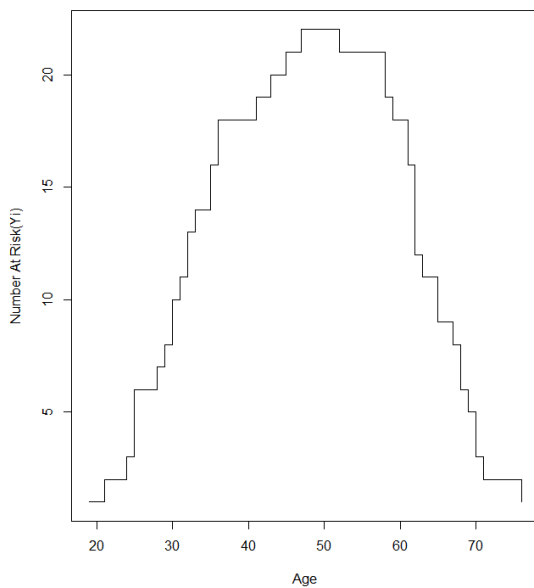
(a) Plot the number at risk , Y_i ,as a function of age .

The table is as below :

Where

$exit\ time = entry\ time +$

$time\ of\ follow-up(exit\_entry)$

Number At Risk(Yi) vs Age

```
> data
   gender entry exit cens exit_entry
1  Female    51   52    0          1
2  Female    58   59    0          1
3  Female    55   57    0          2
4  Female    28   50    0         22
5    Male    21   51    1         30
6    Male    19   47    0         28
7  Female    25   57    0         32
8  Female    48   59    0         11
9  Female    47   61    0         14
10 Female    25   61    1         36
11 Female    31   62    1         31
12   Male    24   57    1         33
13   Male    25   58    1         33
14 Female    30   67    1         37
15 Female    33   68    1         35
16   Male    36   61    0         25
17   Male    30   61    1         31
18   Male    41   63    0         22
19 Female    43   69    0         26
20 Female    45   69    0         24
21 Female    35   70    1         35
22   Male    29   63    1         34
23   Male    35   65    1         30
24   Male    32   67    0         35
25 Female    36   76    0         40
26   Male    32   71    1         39
```

| Time	Yi
19	1
21	2
24	3
25	6
28	7
29	8
30	10
31	11
32	13
33	14
35	16
36	18
41	19
43	20
45	21
47	22
48	22
50	22
51	22
52	21
55	21
57	21
58	19
59	18
61	16
62	12
63	11
65	9
67	8
68	6
69	5
70	3
71	2
76	1

Then I use R to calculate Y_i and t_i^* to draw the risk and age plot as left side .

We can notice that the number at risk initially increases as more individuals enter into the study cohort and that this number decreases for later ages as individuals die or are censored .

(b)Estimate the conditional survival function for a psychiatric patient who has survived to age 30 without entering a psychiatric hospital .

Time	Yi	Di	Left Truncated S(t|Alive at 30)
30	2	0	1.0000000
31	3	0	1.0000000
32	5	0	1.0000000
33	6	0	1.0000000
35	8	0	1.0000000
36	10	0	1.0000000
41	11	0	1.0000000
43	12	0	1.0000000
45	13	0	1.0000000
47	14	0	1.0000000
48	15	0	1.0000000
51	16	0	1.0000000
52	16	0	1.0000000
55	16	0	1.0000000
57	16	0	1.0000000
58	16	0	1.0000000
59	16	0	1.0000000
61	14	1	0.9285714
62	11	1	0.8441558
63	10	0	0.8441558
65	9	1	0.7503608
67	8	1	0.6565657
68	6	1	0.5471380
69	5	0	0.5471380
70	3	1	0.3647587
71	2	1	0.1823793
76	1	0	0.1823793
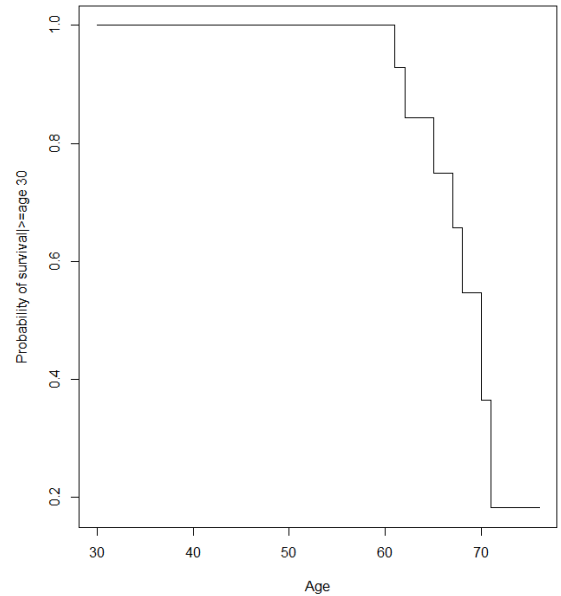
$Y_i = the\ number\ of\ subjects\ at\ risk$

$d_i = the\ number\ of\ death$

$$\widehat{S(t_i^*)} = \prod_{j=1}^{i}(1 - \frac{d_i}{Y_i})$$

Because we consider the time as both enter time and entry time , where the enter time need to >=30 , so the time is from 30-76 , but the exit time means the patient is dead or leave the experiment , so we consider enter>=30 , cens=1 , the first event occurred at time=61 . That is why survival function before time 59 is all 1 .

4. Estimate $\widehat{P}_j$ and plot $\widehat{S(t)}$

(process of convergence must be shown)

```
> pj
              p1         p2        p3         p4
 [1,]  1.666667e-01 0.3333333 0.3333333 1.666667e-01
 [2,]  1.111111e-01 0.3888889 0.3888889 1.111111e-01
 [3,]  7.407407e-02 0.4259259 0.4259259 7.407407e-02
 [4,]  4.938272e-02 0.4506173 0.4506173 4.938272e-02
 [5,]  3.292181e-02 0.4670782 0.4670782 3.292181e-02
 [6,]  2.194787e-02 0.4780521 0.4780521 2.194787e-02
 [7,]  1.463192e-02 0.4853681 0.4853681 1.463192e-02
 [8,]  9.754611e-03 0.4902454 0.4902454 9.754611e-03
 [9,]  6.503074e-03 0.4934969 0.4934969 6.503074e-03
[10,]  4.335382e-03 0.4956646 0.4956646 4.335382e-03
[11,]  2.890255e-03 0.4971097 0.4971097 2.890255e-03
[12,]  1.926837e-03 0.4980732 0.4980732 1.926837e-03
[13,]  1.284558e-03 0.4987154 0.4987154 1.284558e-03
[14,]  8.563718e-04 0.4991436 0.4991436 8.563718e-04
[15,]  5.709146e-04 0.4994291 0.4994291 5.709146e-04
[16,]  3.806097e-04 0.4996194 0.4996194 3.806097e-04
[17,]  2.537398e-04 0.4997463 0.4997463 2.537398e-04
[18,]  1.691599e-04 0.4998308 0.4998308 1.691599e-04
[19,]  1.127732e-04 0.4998872 0.4998872 1.127732e-04
[20,]  7.518216e-05 0.4999248 0.4999248 7.518216e-05
[21,]  5.012144e-05 0.4999499 0.4999499 5.012144e-05
[22,]  3.341430e-05 0.4999666 0.4999666 3.341430e-05
[23,]  2.227620e-05 0.4999777 0.4999777 2.227620e-05
[24,]  1.485080e-05 0.4999851 0.4999851 1.485080e-05
[25,]  9.900532e-06 0.4999901 0.4999901 9.900532e-06
[26,]  6.600355e-06 0.4999934 0.4999934 6.600355e-06
[27,]  4.400236e-06 0.4999956 0.4999956 4.400236e-06
[28,]  2.933491e-06 0.4999971 0.4999971 2.933491e-06
[29,]  1.955661e-06 0.4999980 0.4999980 1.955661e-06
[30,]  1.303774e-06 0.4999987 0.4999987 1.303774e-06
[31,]  8.691825e-07 0.4999991 0.4999991 8.691825e-07
[32,]  5.794550e-07 0.4999994 0.4999994 5.794550e-07
[33,]  3.863033e-07 0.4999996 0.4999996 3.863033e-07
[34,]  2.575356e-07 0.4999997 0.4999997 2.575356e-07
[35,]  1.716904e-07 0.4999998 0.4999998 1.716904e-07
[36,]  1.144602e-07 0.4999999 0.4999999 1.144602e-07
[37,]  7.630683e-08 0.4999999 0.4999999 7.630683e-08
[38,]  5.087122e-08 0.4999999 0.4999999 5.087122e-08
[39,]  3.391415e-08 0.5000000 0.5000000 3.391415e-08
[40,]  2.260943e-08 0.5000000 0.5000000 2.260943e-08
[41,]  1.507295e-08 0.5000000 0.5000000 1.507295e-08
[42,]  1.004864e-08 0.5000000 0.5000000 1.004864e-08
[43,]  6.699091e-09 0.5000000 0.5000000 6.699091e-09
[44,]  4.466061e-09 0.5000000 0.5000000 4.466061e-09
[45,]  2.977374e-09 0.5000000 0.5000000 2.977374e-09
[46,]  1.984916e-09 0.5000000 0.5000000 1.984916e-09
[47,]  1.323277e-09 0.5000000 0.5000000 1.323277e-09
[48,]  8.821849e-10 0.5000000 0.5000000 8.821848e-10
[49,]  5.881232e-10 0.5000000 0.5000000 5.881233e-10
[50,]  3.920821e-10 0.5000000 0.5000000 3.920822e-10
```
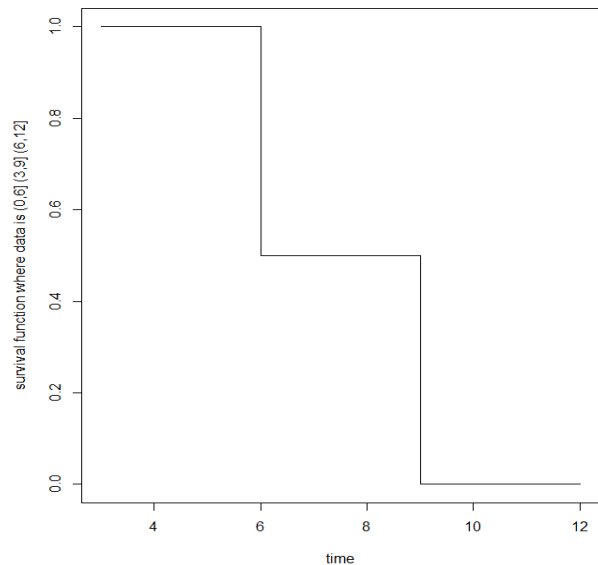

Step 1. I set $P_1 = 0.5 , P_2 = 0.5 , P_3 = 0.5 , P_4 = 0.5$

Step 2. $d_1 = \frac{P_1}{P_1+P_2} , d_2 = \frac{P_2}{P_1+P_2} + \frac{P_2}{P_2+P_3} , d_3 = \frac{P_3}{P_2+P_3} + \frac{P_3}{P_3+P_4} , d_4 = \frac{P_3}{P_3+P_4}$

Step 3. $Y_j = \sum_{k=j}^{m} d_k$

Step 4. $\widehat{S}(t) = \prod_{j=1}^{i} \left(1 - \frac{d_i}{Y_i}\right) , \widehat{P}_j = \widehat{S}(\tau_{j-1}) - \widehat{S}(\tau_j)$

$\rightarrow$ go to step 1 with $P_j = \widehat{P}_j$

The $\widehat{P}_j$ I estimate and survival plot showed above , we can see when J become bigger the value of $P_i, i = 1, ... ,4$ seemed would converge , and when j=18 ,

$\min_j \left| \widehat{P}_J - P_J \right| = 0.00008457994 < 0.0001$

```
> pj[17,]-pj[16,]
           p1            p2            p3            p4
 -0.0001268699  0.0001268699  0.0001268699 -0.0001268699
> pj[18,]-pj[17,]
           p1            p2            p3            p4
 -8.457994e-05  8.457994e-05  8.457994e-05 -8.457994e-05
```

<code>

```
##1.
##(a)
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

y2=1:315 #number of unique death times
for(i in 1:315)
{
   y2[i]=length(data_f$entry[data_f$entry<=t[i]&t[i]<=data_f$exit])
}
y2

plot(t,y1,type='s',xlim=c(750,1200),ylim=c(0,170),xlab='Age in
Months',ylab='Number At Risk')
par(new=T)
plot(t,y2,type='s',xlim=c(750,1200),ylim=c(0,170),lty=2,xlab='Age in
Months',ylab='Number At Risk')

##(b)
library(boot)
data=channing
data$entry=floor(data$entry/12)
data$exit=floor(data$exit/12)
data$time=data$exit-data$entry
```

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

data_m_68=data_m[data_m$entry>=68,]
data_m_80=data_m[data_m$entry>=80,]
data_f_68=data_f[data_f$entry>=68,]
data_f_80=data_f[data_f$entry>=80,]

t1=unique(sort(data_m_68$exit))
t2=unique(sort(data_m_80$exit))
t3=unique(sort(data_f_68$exit))
t4=unique(sort(data_f_80$exit))


y1=1:25 #number of unique death times
for(i in 1:25)
{
    y1[i]=length(data_m_68$entry[data_m_68$entry<=t1[i]&t1[i]<=data_m_68$exit])
}
y1

d1=1:25
for(i in 1:25)
{
d1[i]=sum(data_m_68$exit==t1[i]&data_m_68$cens==1)
}
d1

dy1=d1/y1
minus_dy1=1-d1/y1
s_t1=cumprod(minus_dy1)

y2=1:15 #number of unique death times
for(i in 1:15)
{
    y2[i]=length(data_m_80$entry[data_m_80$entry<=t2[i]&t2[i]<=data_m_80$exit])
}
y2
```

```r
d2=1:15
for(i in 1:15)
{
    d2[i]=sum(data_m_80$exit==t2[i]&data_m_80$cens==1)
}
d2

dy2=d2/y2
minus_dy2=1-d2/y2
s_t2=cumprod(minus_dy2)

y3=1:33 #number of unique death times
for(i in 1:33)
{
    y3[i]=length(data_f_68$entry[data_f_68$entry<=t3[i]&t3[i]<=data_f_68$exit])
}
y3

d3=1:33
for(i in 1:33)
{
    d3[i]=sum(data_f_68$exit==t3[i]&data_f_68$cens==1)
}
d3

dy3=d3/y3
minus_dy3=1-d3/y3
s_t3=cumprod(minus_dy3)

y4=1:21 #number of unique death times
for(i in 1:21)
{
    y4[i]=length(data_f_80$entry[data_f_80$entry<=t4[i]&t4[i]<=data_f_80$exit])
}
y4

d4=1:21
```

```
for(i in 1:21)
{
    d4[i]=sum(data_f_80$exit==t4[i]&data_f_80$cens==1)
}
d4

dy4=d4/y4
minus_dy4=1-d4/y4
s_t4=cumprod(minus_dy4)

y1
d1
y2
d2
y3
d3
y4
d4
s_t1
s_t2
s_t3
s_t4


plot(t1,s_t1,xlim=c(70,100),ylim=c(0,1),lty=2,type='s',xlab='Age',ylab='Estimated
Conditional Survival Probability')
par(new=T)
plot(t2,s_t2,xlim=c(70,100),ylim=c(0,1),lty=3,type='s',xlab='Age',ylab='Estimated
Conditional Survival Probability')
par(new=T)
plot(t3,s_t3,xlim=c(70,100),ylim=c(0,1),lty=1,type='s',xlab='Age',ylab='Estimated
Conditional Survival Probability')
par(new=T)
plot(t4,s_t4,xlim=c(70,100),ylim=c(0,1),lty=5,type='s',xlab='Age',ylab='Estimated
Conditional Survival Probability')

##using survival package
library(survival)
```

```
km2_1=survfit(Surv(data_m_68$entry,data_m_68$exit,data_m_68$cens)~1,type="k
aplan-meier")
km2_2=survfit(Surv(data_m_80$entry,data_m_80$exit,data_m_80$cens)~1,type="k
aplan-meier")
km2_3=survfit(Surv(data_f_68$entry,data_f_68$exit,data_f_68$cens)~1,type="kapla
n-meier")
km2_4=survfit(Surv(data_f_80$entry,data_f_80$exit,data_f_80$cens)~1,type="kapla
n-meier")

plot(km2_1,xlim=c(70,100),xlab="Age",ylab="Probability of survival")
par(new=T)
plot(km2_2,xlim=c(70,100),xlab="Age",ylab="Probability of survival")
par(new=T)
plot(km2_3,xlim=c(70,100),xlab="Age",ylab="Probability of survival")
par(new=T)
plot(km2_4,xlim=c(70,100),xlab="Age",ylab="Probability of survival")

##2.
entry=c(58,58,59,60,60,61,61,62,62,62,63,63,64,66,66,67,67,67,68,69,69,69,70,70,7
0,71,72,72,73,73)
exit=c(60,63,69,62,65,72,69,73,66,65,68,74,71,68,69,70,77,69,72,79,72,70,76,71,78,
79,76,73,80,74)
cens=c(1,1,0,1,1,0,0,0,1,1,1,0,1,1,1,1,1,1,1,0,1,1,0,1,0,0,1,1,0,1)
data=cbind(entry,exit,cens)
data=data.frame(data)

time=unique(sort(c(entry,exit)))
data_60=data[data$exit>=60,]
data_70=data[data$exit>=70,]

y=1:22 #number of unique death times
for(i in 1:22)
{
    y[i]=length(entry[entry<=time[i]&time[i]<=exit])
}
y

d1=1:22
```

```
for(i in 1:22)
{
    d1[i]=sum(data_60$exit==time[i]&data_60$cens==1)
}
d1

dy1=d1/y
minus_dy1=1-d1/y
s_t1=cumprod(minus_dy1)
s_t1

d2=1:22
for(i in 1:22)
{
    d2[i]=sum(data_70$exit==time[i]&data_70$cens==1)
}
d2

dy2=d2/y
minus_dy2=1-d2/y
s_t2=cumprod(minus_dy2)
s_t2

##
entry=c(58,58,59,60,60,61,61,62,62,62,63,63,64,66,66,67,67,67,68,69,69,69,70,70,7
0,71,72,72,73,73)
exit=c(60,63,69,62,65,72,69,73,66,65,68,74,71,68,69,70,77,69,72,79,72,70,76,71,78,
79,76,73,80,74)
cens=c(1,1,0,1,1,0,0,0,1,1,1,0,1,1,1,1,1,1,0,1,1,0,1,0,0,1,1,0,1)
data=cbind(entry,exit,cens)
data=data.frame(data)

time=unique(sort(c(entry,exit)))
data_60=data[data$exit>=60,]
data_70=data[data$exit>=70,]

y=1:22 #number of unique death times
for(i in 1:22)
```

```
{
   y[i]=length(entry[time[i]<=exit])
}
y


d1=1:22
for(i in 1:22)
{
   d1[i]=sum(data_60$exit==time[i]&data_60$cens==1)
}
d1

dy1=d1/y
minus_dy1=1-d1/y
s_t1=cumprod(minus_dy1)
s_t1

d2=1:22
for(i in 1:22)
{
   d2[i]=sum(data_70$exit==time[i]&data_70$cens==1)
}
d2

dy2=d2/y
minus_dy2=1-d2/y
s_t2=cumprod(minus_dy2)
s_t2

##3.
##(a)
gender=c('Female','Female','Female','Female','Male','Male',
            'Female','Female','Female','Female','Female',
            'Male','Male','Female','Female','Male','Male','Male',
            'Female','Female','Female','Male','Male','Male',
            'Female','Male')
entry=c(51,58,55,28,21,19,25,48,47,25,31,24,
```

```r
                25,30,33,36,30,41,43,45,35,29,35,32,36,32)
exit_entry=c(1,1,2,22,30,28,32,11,14,36,31,33,33,
                37,35,25,31,22,26,24,35,34,30,35,40,39)
cens=c(0,0,0,0,1,0,0,0,0,1,1,1,1,1,0,1,0,0,0,1,1,1,0,0,1)
exit=entry+exit_entry

data=cbind(gender,entry,exit,cens)
data=data.frame(data)

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

y=1:34 #number of unique death times
for(i in 1:34)
{
    y[i]=length(data$entry[entry<=t[i]&t[i]<=exit])
}
Y

plot(t,y,type='s',xlab='Age',ylab='Number At Risk(Yi)')

##(b)

gender=c('Female','Female','Female','Female','Male','Male',
            'Female','Female','Female','Female','Female',
            'Male','Male','Female','Female','Male','Male','Male',
            'Female','Female','Female','Male','Male','Male',
            'Female','Male')
entry=c(51,58,55,28,21,19,25,48,47,25,31,24,
            25,30,33,36,30,41,43,45,35,29,35,32,36,32)
exit_entry=c(1,1,2,22,30,28,32,11,14,36,31,33,33,
                37,35,25,31,22,26,24,35,34,30,35,40,39)
cens=c(0,0,0,0,1,0,0,0,0,1,1,1,1,1,0,1,0,0,0,1,1,1,0,0,1)
exit=entry+exit_entry

data2=data[entry>=30,]
entry=data2$entry
exit=data2$exit
```

```
cens=data2$cens
entry=as.vector(entry)
exit=as.vector(exit)
cens=as.vector(cens)

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

y=1:27 #number of unique death times
for(i in 1:27)
{
    y[i]=length(entry[entry<=t[i]&t[i]<=exit])
}
y

d=1:27
for(i in 1:27)
{
    d[i]=sum(exit==t[i]&cens==1)
}
d

dy=d/y
minus_dy=1-dy
s_t=cumprod(minus_dy)
s_t

plot(t,s_t,type='s',xlab="Age",ylab="Probability of survival|>=age 30")

##4.

p1=0.25
p2=0.25
p3=0.25
p4=0.25

data=matrix(c(1,0,0,1,1,0,0,1,1,0,0,1),ncol=4,nrow=3)
data2=matrix(c(p1,p1,p1,p2,p2,p2,p3,p3,p3,p4,p4,p4),ncol=4,nrow=3)
```

```r
colnames(data)=c('(0,t1)','(t1,t2)','(t2,t3)','(t3,t4)')
rownames(data)=c('aij','a2j','a3j')

data*data2

##
p=rep(0.25,4)
pj=matrix(,nrow=50,ncol=4)

y=c()


for(j in 1:50)
{

d[1]=p[1]/(p[1]+p[2])
d[2]=p[2]/(p[1]+p[2])+p[2]/(p[2]+p[3])
d[3]=p[3]/(p[2]+p[3])+p[3]/(p[3]+p[4])
d[4]=p[4]/(p[3]+p[4])

y[4]=d[4]
y[3]=d[3]+d[4]
y[2]=d[2]+d[3]+d[4]
y[1]=d[1]+d[2]+d[3]+d[4]

s_t=cumprod(1-d/y)


p[1]=1-s_t[1]
p[2]=s_t[1]-s_t[2]
p[3]=s_t[2]-s_t[3]
p[4]=s_t[3]-s_t[4]


pj[j,1]=p[1]
pj[j,2]=p[2]
pj[j,3]=p[3]
pj[j,4]=p[4]
```

```
}

colnames(pj)=c('p1','p2','p3','p4')
pj
time=c(3,6,9,12)
plot(time,s_t,type="s",ylab='survival function where data is (0,6] (3,9] (6,12]')
pj[17,]-pj[16,]
pj[18,]-pj[17,]
```