Homework 3

Xinyu Zhang

Due date: Tuesday, November 13

1. **Textbook problem 2.1** Listed below are values of survival time in years for 6 males and 6 females from the *WHAS100* study. Right-censored times are denoted by a "+" as a superscript.

   Males: 1.2, 3.4, 5.0^{+}, 5.1, 6.1, 7.1

   Females: 0.4, 1.2, 4.3, 4.9, 5.0, 5.1^{+}

   Using these data, compute the following **without a software package**:

   a. (2 pts) The Kaplan-Meier estimate of the survival function for each gender.
   b. (2 pts) Pointwise 95 % confidence intervals for the survival functions estimated in problem (1a).
   c. (2 pts) Pointwise 95 % confidence interval estimates of the 50th percentile of the survival time distribution for each gender.
   d. (2 pts) The estimated mean survival time for each gender using all available times, upto 7.1.
   e. (2 pts) A graph of the estimated survival functions for each gender computed in problem (1a) along with the point wise and overall 95 % limits computed in problem (1b).

```
> data=
+   data.frame(time=c(0.4,1.2,3.4,4.3,4.9,5,5.1,6.1,7.1),
+              n.risk=c(12,11,9,8,7,6,4,2,1),
+              d1.event=c(0,1,1,0,0,0,1,1,1),
+              c1.censor=c(0,0,0,0,0,1,0,0,0),
+              d0.event=c(1,1,0,1,1,1,0,0,0),
+              c0.censor=c(0,0,0,0,0,0,1,0,0))
>
> data
```

```
  time n.risk d1.event c1.censor d0.event c0.censor
1  0.4     12        0         0        1         0
2  1.2     11        1         0        1         0
3  3.4      9        1         0        0         0
4  4.3      8        0         0        1         0
5  4.9      7        0         0        1         0
6  5.0      6        0         1        1         0
7  5.1      4        1         0        0         1
8  6.1      2        1         0        0         0
9  7.1      1        1         0        0         0
```

We count the number of event and censors for both male(1) and female(0), then put them into a table.

```
> surv=data %>%
+   mutate(n1.risk=6-head(c(0,cumsum(d1.event+c1.censor)),-1),s1=cumprod(1-d1.event/n1.risk)) %>%
+   mutate(n0.risk=6-head(c(0,cumsum(d0.event+c0.censor)),-1),s0=cumprod(1-d0.event/n0.risk)) %>%
+   mutate(s1=replace_na(s1,tail(s1[!is.na(s1)],1)),s0=replace_na(s0,tail(s0[!is.na(s0)],1)))
>
> surv
```

```
  time n.risk d1.event c1.censor d0.event c0.censor n1.risk        s1
1  0.4     12        0         0        1         0       6 1.0000000
2  1.2     11        1         0        1         0       6 0.8333333
3  3.4      9        1         0        0         0       5 0.6666667
4  4.3      8        0         0        1         0       4 0.6666667
5  4.9      7        0         0        1         0       4 0.6666667
6  5.0      6        0         1        1         0       4 0.6666667
7  5.1      4        1         0        0         1       3 0.4444444
8  6.1      2        1         0        0         0       2 0.2222222
9  7.1      1        1         0        0         0       1 0.0000000
  n0.risk        s0
1       6 0.8333333
2       5 0.6666667
3       4 0.6666667
4       4 0.5000000
5       3 0.3333333
6       2 0.1666667
7       1 0.1666667
8       0 0.1666667
9       0 0.1666667
```

KM survival functin is s1 and s0 for male and female respectively.

```
> surv.ci=surv %>% mutate(var.log.s1=cumsum(d1.event/n1.risk/(n1.risk-d1.event)),
+                         var.log.s0=cumsum(d0.event/n0.risk/(n0.risk-d0.event))) %>%
+   mutate(lower.log.s1=log(s1)-1.96*sqrt(var.log.s1),upper.log.s1=log(s1)+1.96*sqrt(var.log.s1),
+          lower.log.s0=log(s0)-1.96*sqrt(var.log.s0),upper.log.s0=log(s0)+1.96*sqrt(var.log.s0)) %>%
+   mutate(lower.s1=pmax(0,exp(lower.log.s1)),upper.s1=pmin(1,exp(upper.log.s1)),
+          lower.s0=pmax(0,exp(lower.log.s0)),upper.s0=pmin(1,exp(upper.log.s0))) %>%
+   select(time,lower.s1,s1,upper.s1,lower.s0,s0,upper.s0)
>
> surv.ci
```

```
  time   lower.s1        s1 upper.s1   lower.s0        s0  upper.s0
1  0.4 1.00000000 1.0000000        1 0.58265096 0.8333333 1.0000000
2  1.2 0.58265096 0.8333333        1 0.37860252 0.6666667 1.0000000
3  3.4 0.37860252 0.6666667        1 0.37860252 0.6666667 1.0000000
4  4.3 0.37860252 0.6666667        1 0.22462704 0.5000000 1.0000000
5  4.9 0.37860252 0.6666667        1 0.10750490 0.3333333 1.0000000
6  5.0 0.37860252 0.6666667        1 0.02784821 0.1666667 0.9974708
7  5.1 0.16680493 0.4444444        1 0.02784821 0.1666667 0.9974708
8  6.1 0.04070163 0.2222222        1        NaN 0.1666667       NaN
9  7.1 0.00000000 0.0000000      NaN        NaN 0.1666667       NaN
```

Above is the 95% CI of KM survival function for s1 and s0 using Greenwood formula.

We draw a horizontal line across S(t)=0.5. If the line cross the KM survival function at t_median, then it is the median survival time. If the line crosses the lower and upper CI at t_lower and t_upper respectively, then the 95% CI for t_median is (t_lower, t_upper). The median survival time for male is 5.1. Its 95% CI is (3.4, NA). The median survial time for female is 4.3. Its 95% CI is (1.2,NA).

```
> e.s1=with(surv,sum(head(c(1,s1),-1)*(surv$time-head(c(0,surv$time),-1))))
> e.s0=with(surv,sum(head(c(1,s0),-1)*(surv$time-head(c(0,surv$time),-1))))
>
> e.s1
```

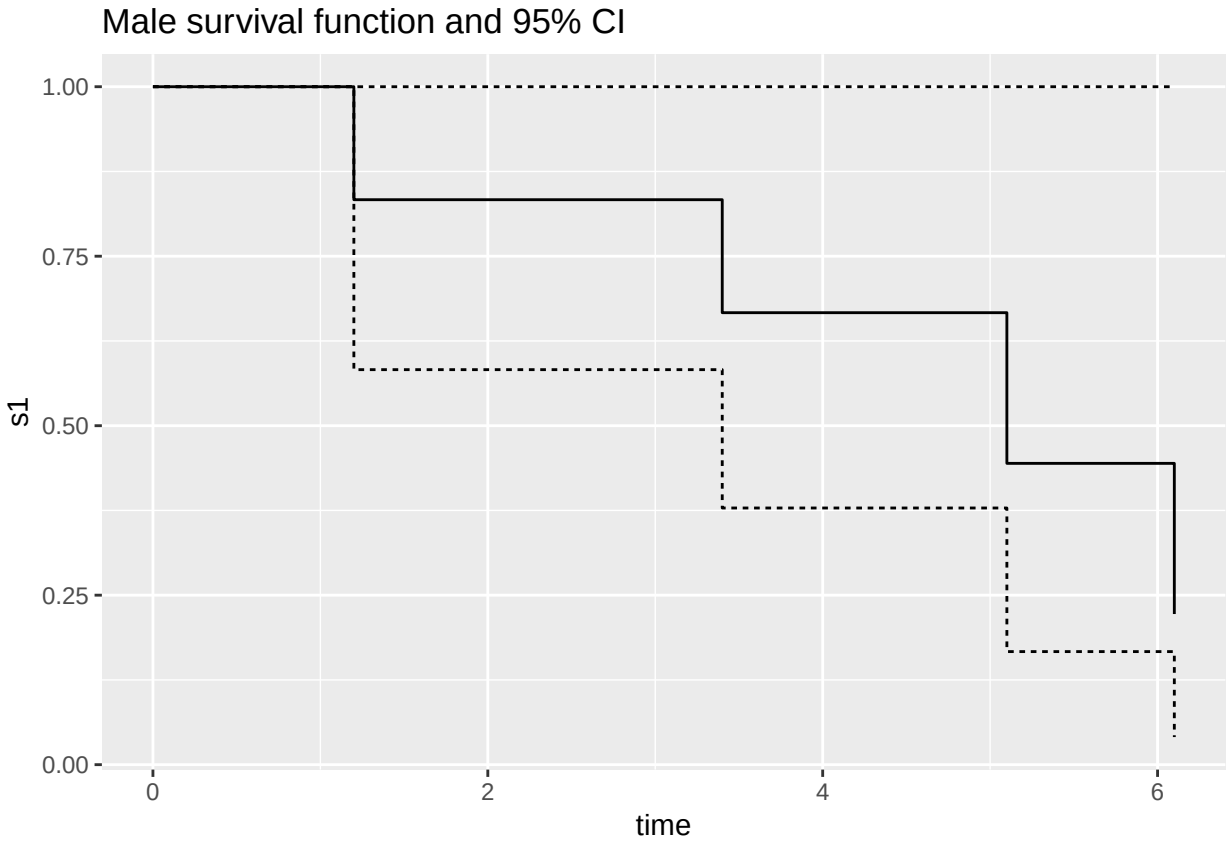
```
[1] 4.833333
```

```
> e.s0
```

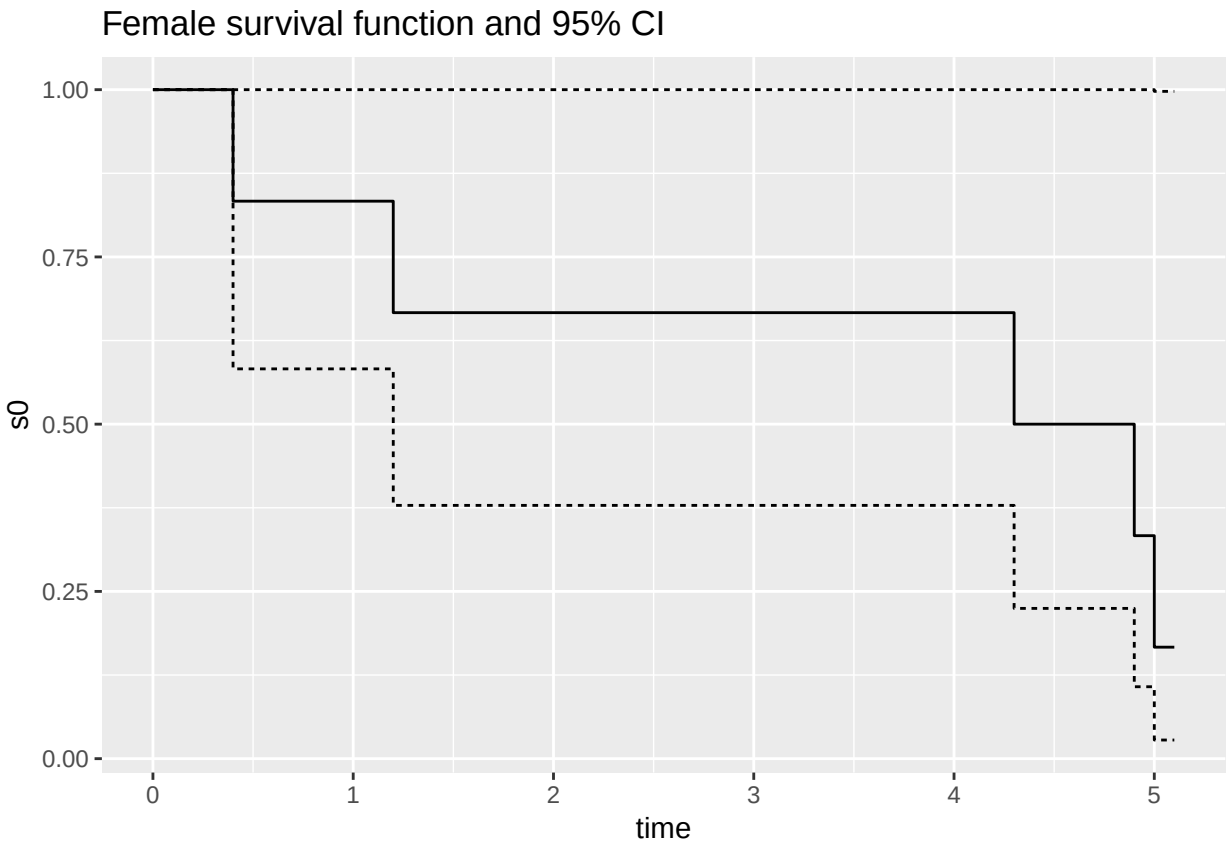
```
[1] 3.816667
```

The expected survival time up to 7.1 is 4.833333 and 3.816667 for male and female respectively.

```
> surv.ci %>% add_row(time=0,lower.s1=1,s1=1,upper.s1=1,lower.s0=1,s0=1,upper.s0=1,.before=1) %>%
+   filter(!is.na(upper.s1)) %>%
+   ggplot()+
+   geom_step(aes(time,s1,linetype="dashed"))+
+   geom_step(aes(time,lower.s1,linetype="solid"))+
+   geom_step(aes(time,upper.s1,linetype="solid"))+
+   theme(legend.position="none")+
+   ggtitle("Male survival function and 95% CI")
```




```
> surv.ci %>% add_row(time=0,lower.s1=1,s1=1,upper.s1=1,lower.s0=1,s0=1,upper.s0=1,.before=1) %>%
+   filter(!is.na(upper.s0)) %>%
+   ggplot()+
+   geom_step(aes(time,s0,linetype="dashed"))+
+   geom_step(aes(time,lower.s0,linetype="solid"))+
+   geom_step(aes(time,upper.s0,linetype="solid"))+
+   theme(legend.position="none")+
+   ggtitle("Female survival function and 95% CI")
```



Above is the KM survival curve for male and female.