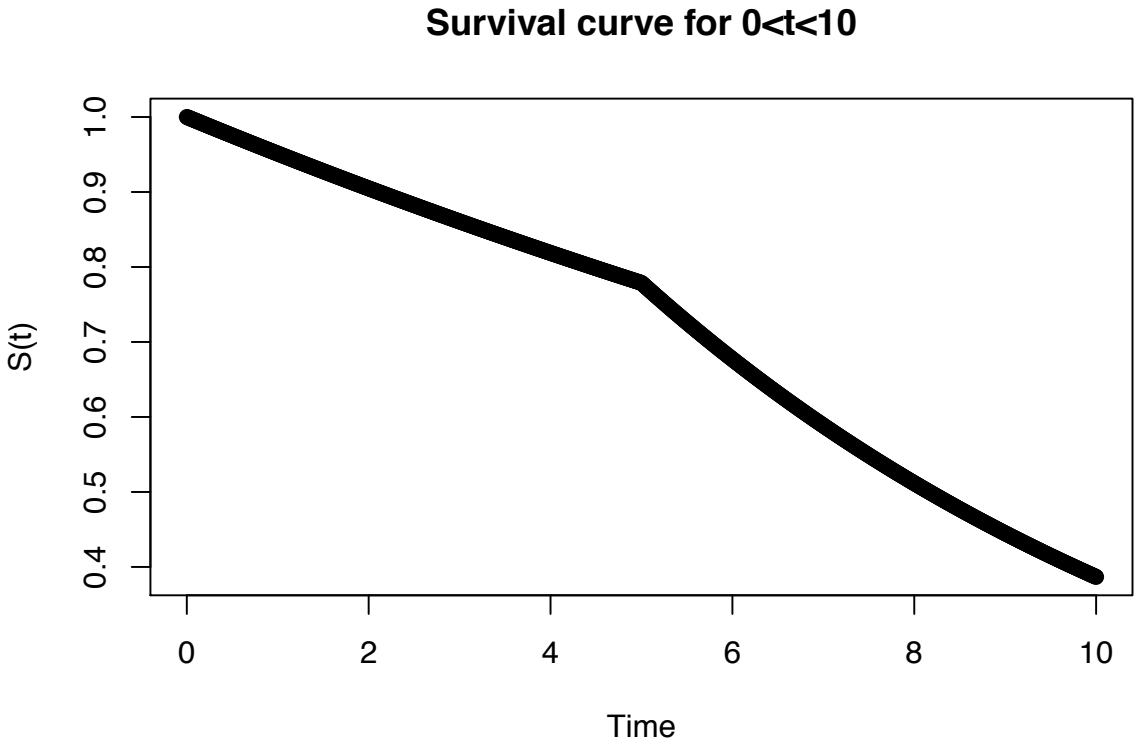
Survival Analysis HW 2

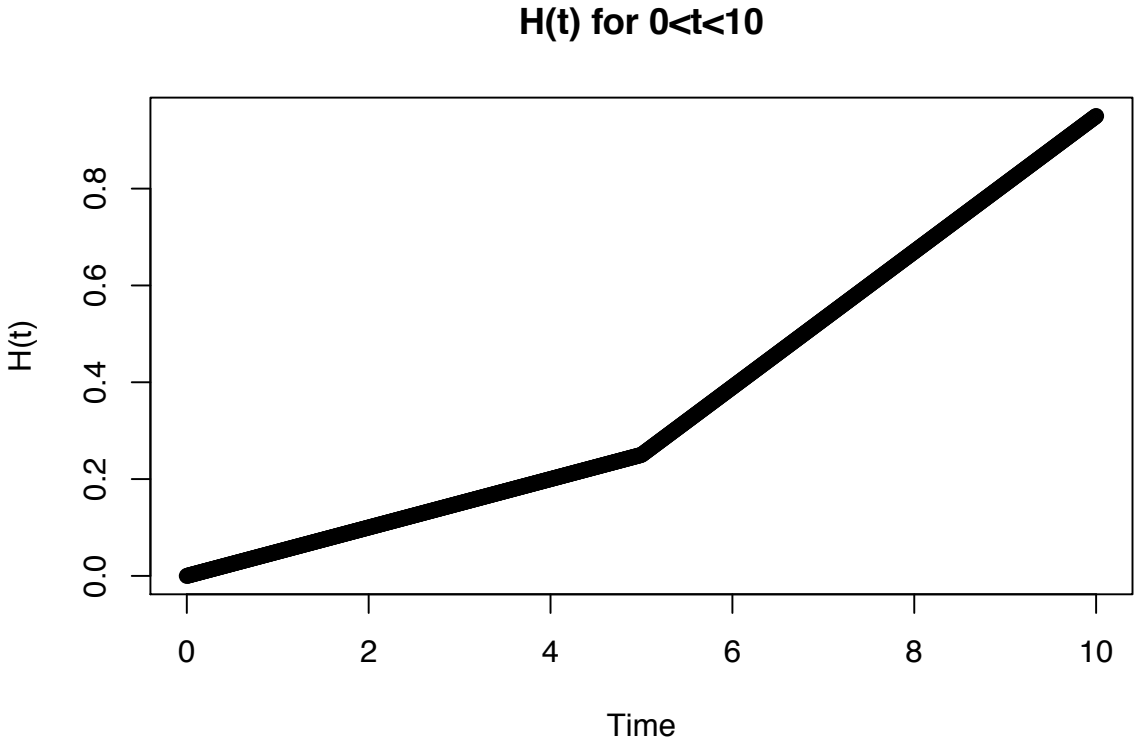
Joanna Chen

2/9/2020

Question 1



Although the homework didn't ask for, but I also ploted H(t) to get a better idea of the relationship between H(t) and S(t). Also it will be used in part (c).



Question 3

Data for this question come from an actual Phase II clinical trial on patients with advanced gastric cancer and para-aortic lymph node involvement. This single-arm study administered chemotherapeutic agent Xelox to patients prior to surgery. The primary outcome of interest is "progression-free survival," defined as the time from entry into the clinical trial to progression or death, whichever comes first. Survival times are reported in weeks. The data are provided in a CSV file (HW02_GastricCancer.csv).

a. Report the Kaplan-Meier estimate of the survival function using the data provided. [Note: You should report the survival probabilities at time 0 and at each unique event time in a table.]

time	n.risk	n.event	n.censor	Shat_KM
0	48	0	0	1.000
4	48	1	0	0.979
8	47	3	0	0.917
9	44	1	0	0.896
11	43	1	0	0.875
12	42	1	0	0.854
13	41	1	0	0.833
16	40	2	0	0.792
17	38	2	0	0.750
19	36	1	0	0.729
21	35	1	0	0.708
24	34	2	0	0.667
25	32	1	0	0.646
28	31	2	0	0.604
30	29	1	0	0.583

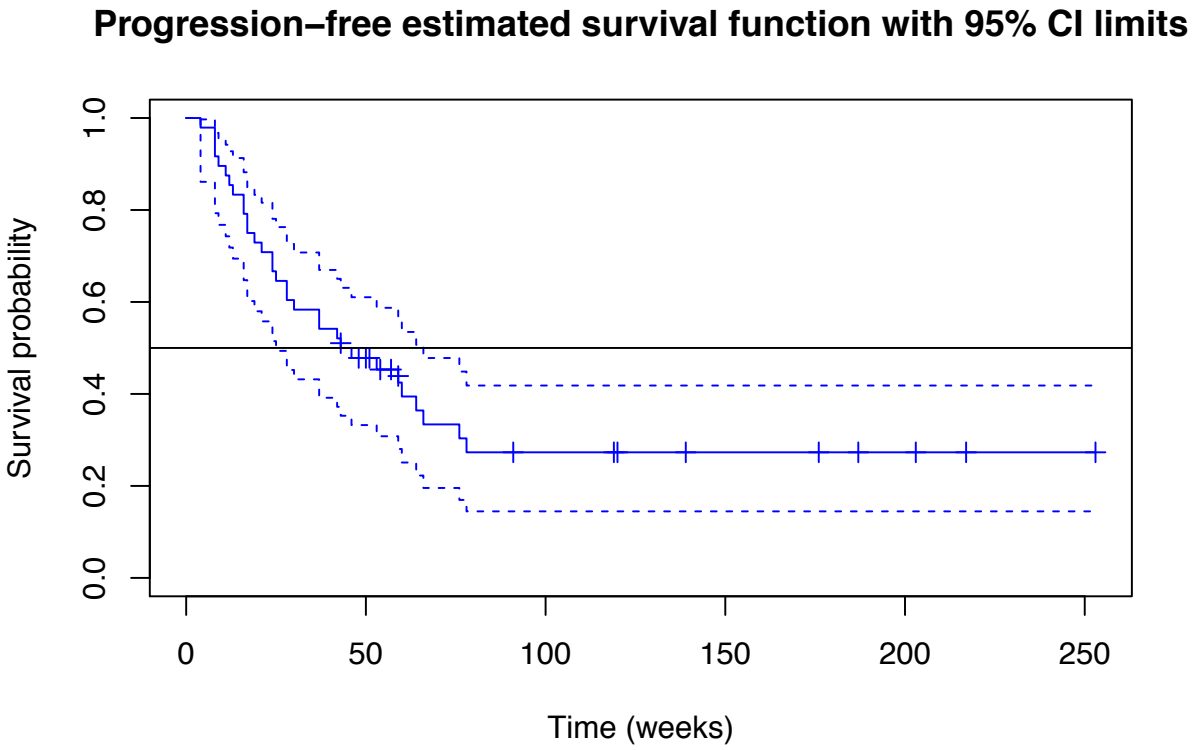
time	n.risk	n.event	n.censor	Shat_KM
37	28	2	0	0.542
42	26	1	0	0.521
43	25	1	1	0.500
46	23	1	0	0.478
53	19	1	3	0.453
59	16	1	3	0.425
60	14	1	0	0.394
64	13	1	0	0.364
66	12	1	0	0.334
76	11	1	0	0.303
78	10	1	0	0.273

b. Report the pointwise 95% confidence limits for the Kaplan-Meier survival function (your choice of CI). Be sure to indicate which type of CI you are reporting.

I am reporting Log-log CI.

time	n.risk	n.event	Shat_KM	std.error	lower 95% CI	upper 95% CI
0	48	0	1.000	0.000	1.000	1.000
4	48	1	0.979	0.021	0.861	0.997
8	47	3	0.917	0.040	0.793	0.968
9	44	1	0.896	0.044	0.768	0.955
11	43	1	0.875	0.048	0.743	0.942
12	42	1	0.854	0.051	0.718	0.928
13	41	1	0.833	0.054	0.694	0.913
16	40	2	0.792	0.059	0.647	0.882
17	38	2	0.750	0.062	0.602	0.850
19	36	1	0.729	0.064	0.580	0.833
21	35	1	0.708	0.066	0.558	0.816
24	34	2	0.667	0.068	0.515	0.781
25	32	1	0.646	0.069	0.494	0.763
28	31	2	0.604	0.071	0.452	0.726
30	29	1	0.583	0.071	0.432	0.708
37	28	2	0.542	0.072	0.392	0.670
42	26	1	0.521	0.072	0.372	0.650
43	25	1	0.500	0.072	0.353	0.631
46	23	1	0.478	0.072	0.332	0.610
53	19	1	0.453	0.073	0.308	0.587
59	16	1	0.425	0.073	0.280	0.562
60	14	1	0.394	0.074	0.251	0.535
64	13	1	0.364	0.074	0.223	0.507
66	12	1	0.334	0.074	0.196	0.478
76	11	1	0.303	0.073	0.170	0.449
78	10	1	0.273	0.072	0.145	0.418

c. Provide a graph of the estimated survival function computed in part (a) together with the pointwise 95% confidence limits computed in part (b).



d. If appropriate, report the median progression free survival time in these data.

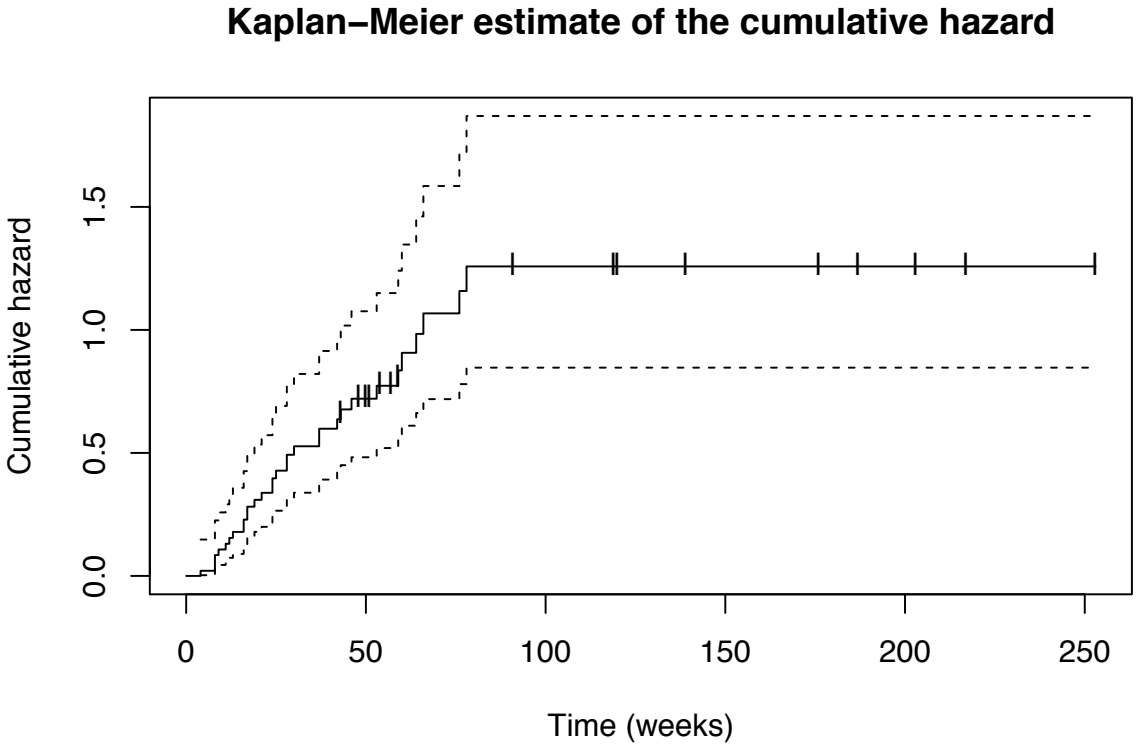
```
## Call: survfit(formula = Surv(timeWeeks, delta) ~ 1, data = gastric,
##     conf.type = "log-log")
##
##       n  events  median 0.95LCL 0.95UCL
##    48.0    32.0    44.5    25.0    66.0
```

The median progression free survival time is 44.5 weeks. We can also see through the plot by the time of the intersection point of survival probability = 0.5.

e. If appropriate, report the mean progression free survival time in these data.

Since the plot is positively skewed, the median is more appropriate.

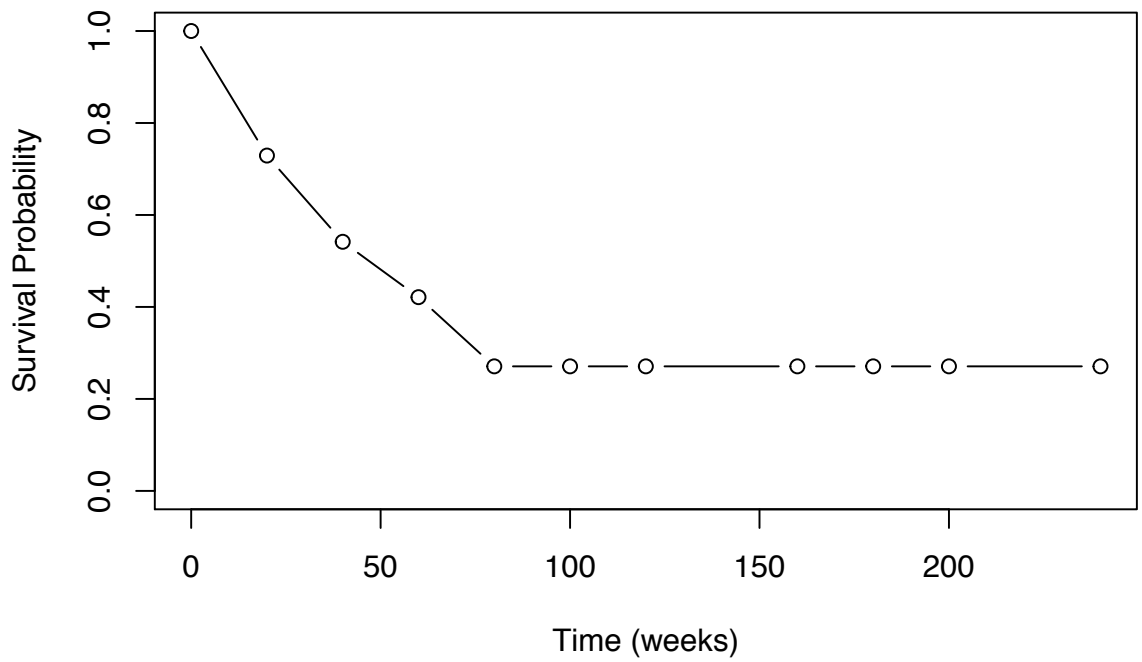
f. Provide a graph of the Kaplan-Meier estimate of the cumulative hazard. [Note: You may report the SAS auto-generated cumulative hazard plot or plot your own step function using a plotting procedure]. How do the graphs from part (c) and (e) visually compare?



Comparing our two plots, we can see that with cumulative hazard increasing, the survival probability decreasing. This fits the relationship that $H(T) = -log(S(t))$.

g. Provide the actuarial (life-table) estimate of the survival function using time intervals of 20 weeks and graph the estimated survival function. How does the life-table estimate visually compare to the Kaplan-Meier estimate of S(t) plotted in part (c)?

	nsubs	nlost	nrisk	nevent	surv
0-20	48	0	48.0	13	1.0000000
20-40	35	0	35.0	9	0.7291667
40-60	26	7	22.5	5	0.5416667
60-80	14	0	14.0	5	0.4212963
80-100	9	1	8.5	0	0.2708333
100-120	8	1	7.5	0	0.2708333
120-160	7	2	6.0	0	0.2708333
160-180	5	1	4.5	0	0.2708333
180-200	4	1	3.5	0	0.2708333
200-240	3	2	2.0	0	0.2708333
240-NA	1	1	0.5	0	0.2708333



Please see the "surv" column for the estimate.Compring with the Kaplan-Meier estimate of S(t) plotted in part (c), the plot survival times are grouped into intervals of fixed length (the reason why the Time = 140, 220 doesn't show on the graph is because there's no points in the interval). Unlike the Kaplan-Meier estimator, intervals are not created based on when events occurred. But overall their shape is similar.

Appendix

```
setwd("~/Desktop")
x = seq(0, 10, 0.01)
myfunction = function(x){(x>= 0 & x <=5)*(exp(-0.05*x)) +
                         (x> 5)*(exp(-0.25-0.14*(x-5)))}
plot(x, myfunction(x),main="Survival curve for 0<t<10", xlab="Time", ylab="S(t)")

x = seq(0, 10, 0.01)
myfunction = function(x){(x>= 0 & x <=5)*(0.05*x) +
                         (x> 5)*(0.25+0.14*(x-5))}

plot(x, myfunction(x),main="H(t) for 0<t<10", xlab="Time", ylab="H(t)")

gastric <- readr::read_csv("HW02_GastricCancer.csv")
library(survival)

# Surv(gastric$timeWeeks, gastric$delta)
kmsurv <- survfit(Surv(timeWeeks, delta) ~ 1, data=gastric, conf.type="none")
# summary(kmsurv)

output = cbind(summary(kmsurv)$time, summary(kmsurv)$n.risk,summary(kmsurv)$n.event, summary(kmsurv)$n.
output = rbind(c(0,48,0,0,1),output)
colnames(output)=c("time", "n.risk", "n.event", "n.censor", "Shat_KM")
knitr::kable(output)
```

```r
kmsurv_ci <- survfit(Surv(timeWeeks, delta) ~ 1, data=gastric, conf.type="log-log")

output1 = cbind(summary(kmsurv_ci)$time, summary(kmsurv_ci)$n.risk,summary(kmsurv_ci)$n.event, round(su
output1 = rbind(c(0,48,0,1,0,1,1),output1)
colnames(output1)=c("time", "n.risk", "n.event", "Shat_KM", "std.error", "lower 95% CI","upper 95% CI")
#output1
knitr::kable(output1)

plot(kmsurv_ci,  xlab="Time (weeks)",ylab="Survival probability", mark.time=T, conf.int=T, col="blue")
title("Progression-free estimated survival function with 95% CI limits")
abline(h=.5)

kmsurv_ci

plot(kmsurv_ci, conf.int=T, mark="|", xlab="Time (weeks)",
   fun="cumhaz", ylab="Cumulative hazard" )
title("Kaplan-Meier estimate of the cumulative hazard")

library(KMsurv)
library(nlme)

gastric = gastric[order(gastric$timeWeeks),] # Sort data by time

t20w = floor(gastric$timeWeeks/20) # Create time interval
tall = data.frame(t20w, gastric$delta)
die = gsummary(tall, sum, groups = t20w) # Count events
total = gsummary(tall, length, groups = t20w) # count subjects

ltab.data = data.frame(time = die[,1],die = die[,2],total = total[,2])

lt = length(t20w)
t20w[lt+1] = NA
nevent = ltab.data$die
nlost = ltab.data$total - nevent

mytable = lifetab(20*unique(t20w),nrow(gastric),nlost,nevent)
knitr::kable(mytable[,1:5])

plot(20*(unique(t20w)[-length(unique(t20w))]), mytable[,5], type="b",
   xlab="Time (weeks)", ylab="Survival Probability", ylim=c(0,1))
title("Life-table estimate of the survival function using time intervals of 20 weeks")
```