P8108 Homework 3

Ryan Wei, rw2844

2022-09-28

Problem 1

Table 1: Life table estimate

Interval	Time Period	d_i	c_i	n_i	n_i'	$S(t)$	$f(t)$	$h(t)$	$se(S(t))$
1	[0,4)	2	1	20	19.5	1.0000	0.0257	0.2703	0.0000
2	[4,8)	1	1	17	16.5	0.8974	0.0136	0.0156	0.0687
3	[8,12)	0	3	15	13.5	0.8430	0.0000	0.0000	0.0833
4	[12,16)	1	2	12	11.0	0.8430	0.0192	0.0238	0.0833
5	[16,20)	1	3	9	7.5	0.7664	0.0256	0.0357	0.1053
6	[20,24)	1	1	5	4.5	0.6642	0.0369	0.0625	0.1318
7	[24,∞)	0	3	3	1.5	0.5166	NA	NA	0.1657

Note: Survival functions estimated at the begining of each interval

Problem 2

Proof

Since

$$\hat{h}(t_{mi}) = \frac{d_i}{[(t_i - t_{i-1})(\frac{n'_i - d_i}{2})]}$$
$$\hat{f}(t_{mi}) = \frac{\hat{S}_L(t_{i-1}) - \hat{S}_L(t_i)}{t_i - t_{i-1}}$$
$$\hat{S}(t_{mi}) = \frac{\hat{S}_L(t_{i-1}) + \hat{S}_L(t_i)}{2}$$
$$\hat{S}_L(t_i) = \hat{S}_L(t_{i-1})(1 - \frac{d_i}{n'_i})$$

Therefore, based on the definition,

$$\begin{aligned}
\hat{h}(t_{mi}) &= \frac{\hat{f}(t_{mi})}{\hat{S}(t_{mi})} \\
&= \frac{2\hat{f}(t_{mi})}{\hat{S}_L(t_{i-1}) + \hat{S}_L(t_i)} \\
&= \frac{2(\hat{S}_L(t_{i-1}) - \hat{S}_L(t_i))}{(t_i - t_{i-1})(\hat{S}_L(t_{i-1}) + \hat{S}_L(t_i))} \\
&= \frac{2(\hat{S}_L(t_{i-1}) \times \frac{d_i}{n'_i})}{(t_i - t_{i-1})(\hat{S}_L(t_{i-1}) \times \frac{2n'_i - d_i}{n'_i})} \\
&= \frac{d_i}{(n'_i - \frac{d_i}{2})(t_i - t_{i-1})}
\end{aligned}$$

□

Problem 3

- Create life-table stratified by `rx`.

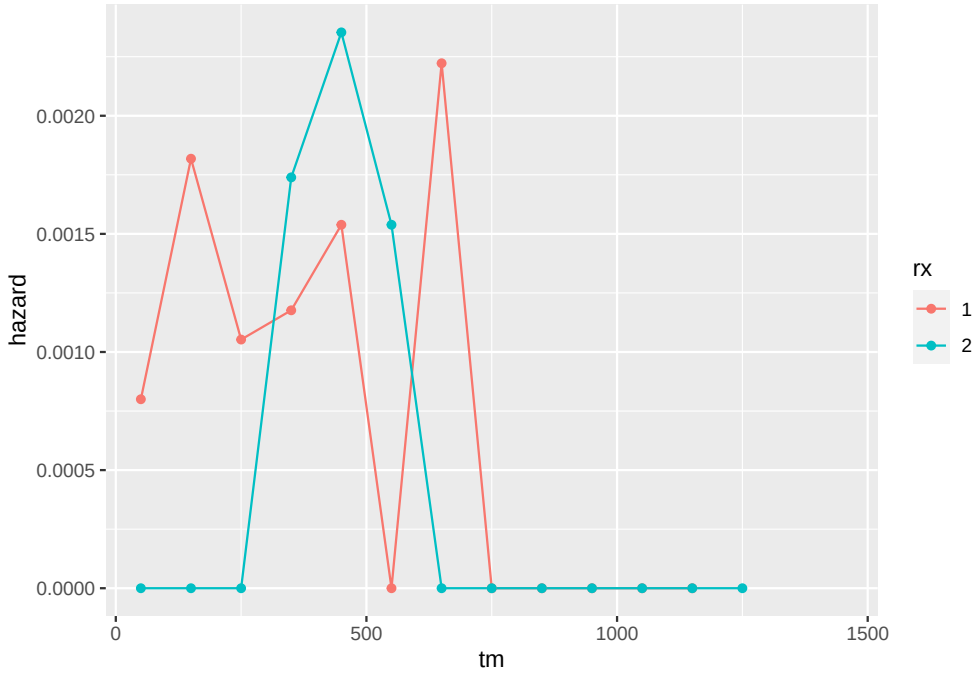
Table 2: Life table, rx=1

	tstart	tstop	nsubs	nlost	nrisk	nevent	surv	pdf	hazard	se.surv	se.pdf	se.hazard
0-100	0	100	13	0	13.0	1	1.0000000	0.0007692	0.0008000	0.0000000	0.0007391	0.0007994
100-200	100	200	12	0	12.0	2	0.9230769	0.0015385	0.0018182	0.0739053	0.0010007	0.0012803
200-300	200	300	10	0	10.0	1	0.7692308	0.0007692	0.0010526	0.1168545	0.0007391	0.0010512
300-400	300	400	9	0	9.0	1	0.6923077	0.0007692	0.0011765	0.1280077	0.0007391	0.0011744
400-500	400	500	8	2	7.0	1	0.6153846	0.0008791	0.0015385	0.1349320	0.0008364	0.0015339
500-600	500	600	5	0	5.0	0	0.5274725	0.0000000	0.0000000	0.1414241	NaN	NaN
600-700	600	700	5	0	5.0	1	0.5274725	0.0010549	0.0022222	0.1414241	0.0009851	0.0022085
700-800	700	800	4	0	4.0	0	0.4219780	0.0000000	0.0000000	0.1473220	NaN	NaN
800-900	800	900	4	2	3.0	0	0.4219780	0.0000000	0.0000000	0.1473220	NaN	NaN
900-1000	900	1000	2	0	2.0	0	0.4219780	0.0000000	0.0000000	0.1473220	NaN	NaN
1000-1100	1000	1100	2	1	1.5	0	0.4219780	0.0000000	0.0000000	0.1473220	NaN	NaN
1100-1200	1100	1200	1	1	0.5	0	0.4219780	0.0000000	0.0000000	0.1473220	NaN	NaN
1200-1300	1200	1300	0	0	0.0	0	0.4219780	NaN	NaN	0.1473220	NaN	NaN
1300-1400	1300	1400	0	0	0.0	0	NaN	NaN	NaN	NaN	NaN	NaN
1400-1500	1400	1500	0	0	0.0	0	NaN	NaN	NaN	NaN	NaN	NaN
1500-Inf	1500	Inf	0	0	0.0	0	NaN	NA	NA	NaN	NA	NA

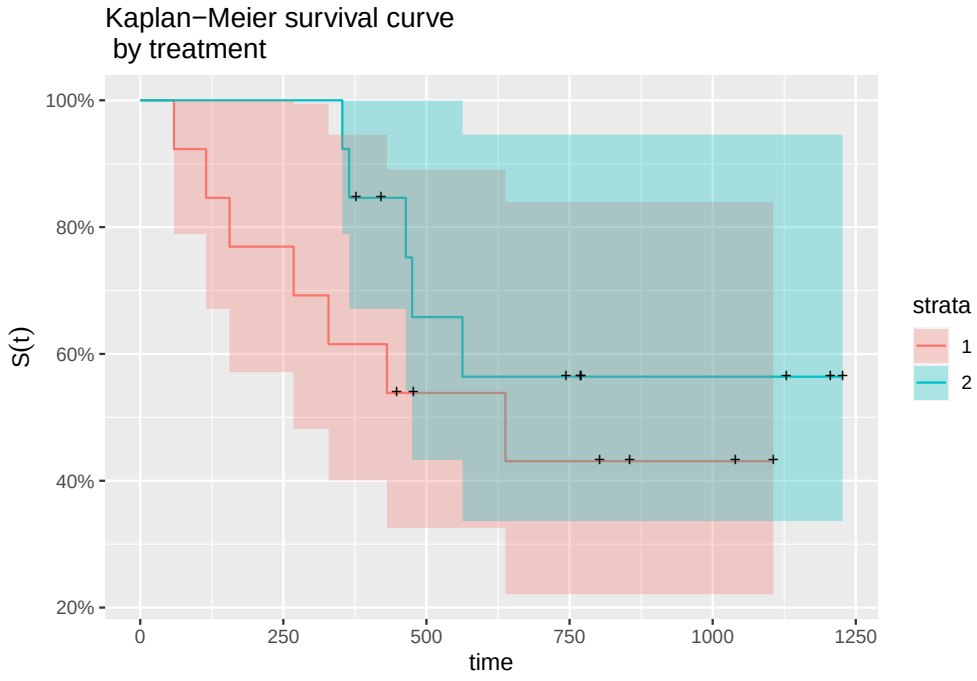
Table 3: Life table, rx=2

	tstart	tstop	nsubs	nlost	nrisk	nevent	surv	pdf	hazard	se.surv	se.pdf	se.hazard
0-100	0	100	13	0	13.0	0	1.0000000	0.0000000	0.0000000	0.0000000	NaN	NaN
100-200	100	200	13	0	13.0	0	1.0000000	0.0000000	0.0000000	0.0000000	NaN	NaN
200-300	200	300	13	0	13.0	0	1.0000000	0.0000000	0.0000000	0.0000000	NaN	NaN
300-400	300	400	13	1	12.5	2	1.0000000	0.0016000	0.0017391	0.0000000	0.0010369	0.0012251
400-500	400	500	10	1	9.5	2	0.8400000	0.0017684	0.0023529	0.1036919	0.0011323	0.0016522
500-600	500	600	7	0	7.0	1	0.6631579	0.0009474	0.0015385	0.1380074	0.0008990	0.0015339
600-700	600	700	6	0	6.0	0	0.5684211	0.0000000	0.0000000	0.1472614	NaN	NaN
700-800	700	800	6	3	4.5	0	0.5684211	0.0000000	0.0000000	0.1472614	NaN	NaN
800-900	800	900	3	0	3.0	0	0.5684211	0.0000000	0.0000000	0.1472614	NaN	NaN
900-1000	900	1000	3	0	3.0	0	0.5684211	0.0000000	0.0000000	0.1472614	NaN	NaN
1000-1100	1000	1100	3	0	3.0	0	0.5684211	0.0000000	0.0000000	0.1472614	NaN	NaN
1100-1200	1100	1200	3	1	2.5	0	0.5684211	0.0000000	0.0000000	0.1472614	NaN	NaN
1200-1300	1200	1300	2	2	1.0	0	0.5684211	0.0000000	0.0000000	0.1472614	NaN	NaN
1300-1400	1300	1400	0	0	0.0	0	0.5684211	NaN	NaN	0.1472614	NaN	NaN
1400-1500	1400	1500	0	0	0.0	0	NaN	NaN	NaN	NaN	NaN	NaN
1500-Inf	1500	Inf	0	0	0.0	0	NaN	NA	NA	NaN	NA	NA

- Plot hazard function by `rx` based on life-table estimate



- Plot K-M survival function by `rx`.



- What is the median survival time for each treatment group?

Table 4: Median survival time by K-M estimation

rx	Median Survival (95% CI)
1	638 (268,–)
2	– (475,–)

- Compare survival function estimations between K-M and F-H methods.

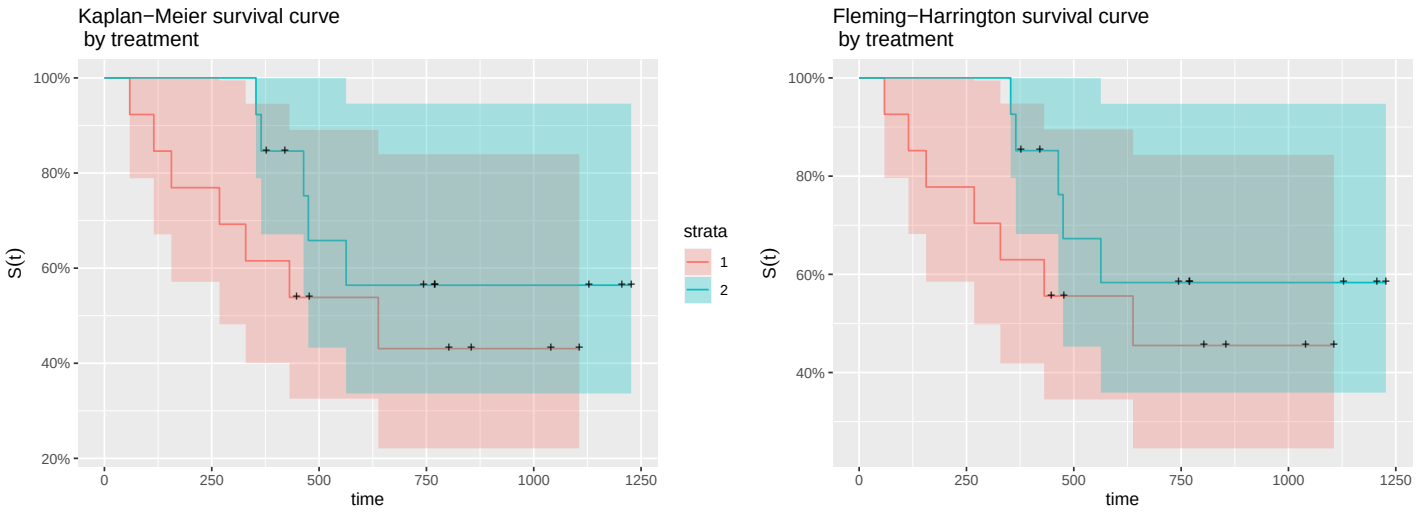


Table 5: Survival function estimations between K-M and F-H methods, rx = 1

	K-M estimation		F-H estimation	
Time	Survival	Survival Standard Error	Survival	Survival Standard Error
59	0.9231	0.0801	0.9260	0.0769
115	0.8462	0.1183	0.8519	0.1134
156	0.7692	0.1519	0.7779	0.1453
268	0.6923	0.1849	0.7039	0.1764
329	0.6154	0.2193	0.6298	0.2085
431	0.5385	0.2568	0.5558	0.2431
448	0.5385	0.2568	0.5558	0.2431
477	0.5385	0.2568	0.5558	0.2431
638	0.4308	0.3405	0.4551	0.3148
803	0.4308	0.3405	0.4551	0.3148
855	0.4308	0.3405	0.4551	0.3148
1040	0.4308	0.3405	0.4551	0.3148
1106	0.4308	0.3405	0.4551	0.3148

Table 6: Survival function estimations between K-M and F-H methods, rx = 2

	K-M estimation		F-H estimation	
Time	Survival	Survival Standard Error	Survival	Survival Standard Error
353	0.9231	0.0801	0.9260	0.0769
365	0.8462	0.1183	0.8519	0.1134
377	0.8462	0.1183	0.8519	0.1134
421	0.8462	0.1183	0.8519	0.1134
464	0.7521	0.1670	0.7623	0.1588
475	0.6581	0.2139	0.6728	0.2021
563	0.5641	0.2637	0.5832	0.2475
744	0.5641	0.2637	0.5832	0.2475
769	0.5641	0.2637	0.5832	0.2475
770	0.5641	0.2637	0.5832	0.2475
1129	0.5641	0.2637	0.5832	0.2475
1206	0.5641	0.2637	0.5832	0.2475
1227	0.5641	0.2637	0.5832	0.2475

- Describe your analyses and write conclusion based on you analyses.

1. In both stratums, there are 13 patients. However, there are more censored patients in group 2 (`rx = 2`) than group 1 (`rx = 1`). The hazard rate of group 1 is higher than group 2 for most of time.
2. According to different survival function estimations, both K-M curves and F-H curves show that the survival probability of group 2 is higher than group 1.
3. The median survival time of group 1 is 638. However, group 2 doesn't reach a median survival.
4. The estimation by K-M method and F-H method doesn't differ too much.

Reference

1. Edmunson, J.H., Fleming, T.R., Decker, D.G., Malkasian, G.D., Jefferies, J.A., Webb, M.J., and Kvols, L.K., Different Chemotherapeutic Sensitivities and Host Factors Affecting Prognosis in Advanced Ovarian Carcinoma vs. Minimal Residual Disease. Cancer Treatment Reports, 63:241-47, 1979.

Appendix: Code for this report

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
library(survival)
library(tidyverse)
library(biostat3)
library(ggfortify)
library(ggsurvfit)
library(gtsummary)
library(knitr)
library(kableExtra)
life_tab <- data.frame(Interval = c(1,2,3,4,5,6,7),
```

```
                  Period = c("[0,4)","[4,8)","[8,12)","[12,16)","[16,20)","[20,24)","[24,$\\infty$)"),
                  d_i = c(2,1,0,1,1,1,0),
                  c_i = c(1,1,3,2,3,1,3),
                  n_i = c(20,17,15,12,9,5,3),
                  n_i_prime = c(19.5,16.5,13.5,11,7.5,4.5,1.5),
                  st = c(1.000,0.8974,0.8430,0.8430,0.7664,0.6642,0.5166),
                  ft = c(0.0257,0.0136,0,0.0192,0.0256,0.0369,NA),
                  ht = c(0.2703,0.0156,0,0.0238,0.0357,0.0625,NA),
                  se_st = c(0,0.06870,0.0833,0.0833,0.1053,0.1318,0.1657))
kable(life_tab, col.names = c("Interval", "Time Period", "$d_i$", "$c_i$", "$n_i$", "$n_i^{\\prime}$",
ovarian.rx1 <- ovarian[ovarian$rx == 1, ]
ovarian.rx2 <- ovarian[ovarian$rx == 2, ]
ovarian.lt1 <- lifetab2(Surv(futime, fustat)~1, data = ovarian.rx1, breaks =seq(0, 1500, 100))
ovarian.lt2 <- lifetab2(Surv(futime, fustat)~1, data = ovarian.rx2, breaks =seq(0, 1500, 100))
ovarian.lt1 %>% kable(booktabs = T, caption = "Life table, rx=1") %>% kable_styling(latex_options = c("
ovarian.lt2 %>% kable(booktabs = T, caption = "Life table, rx=2") %>% kable_styling(latex_options = c("
# hazard: the estimated hazard rate at the midpoint of the intervals.
ovarian.lt1$rx = as.factor(1)
ovarian.lt2$rx = as.factor(2)
ovarian.lt <- rbind(ovarian.lt1, ovarian.lt2)
ovarian.lt$tm <- (ovarian.lt$tstart + ovarian.lt$tstop)/2
ovarian.lt %>%
  ggplot(data = ., aes(x = tm, y = hazard, group = rx, color = rx)) +
  geom_line() + geom_point()
ovarian.kmfit <-
  ovarian %>%
  survfit(Surv(futime,fustat)~rx, data = .)

ovarian.kmfit %>% autoplot() +
  ylab(bquote(S(t))) +
  ggtitle("Kaplan-Meier survival curve \n by treatment")


#ovarian.kmfit %>% ggsurvfit()
km_median =
  tibble(
    rx = as.factor(c(1,2)),
    mediansurvival = c("638 (268,--)", " -- (475,--)")
  )
km_median %>% kable(booktab = T, escape = F, col.names = c("rx", "Median Survival (95$\\%$ CI)"), capti
ovarian.fhfit <-
  ovarian %>%
  survfit(Surv(futime,fustat)~rx, data = ., type = "fh" )

par(mfrow=c(1,2))
ovarian.kmfit %>%
  autoplot() +
  ylab(bquote(S(t))) +
  ggtitle("Kaplan-Meier survival curve \n by treatment")
ovarian.fhfit %>%
  autoplot() +
  ylab(bquote(S(t))) +
  ggtitle("Fleming-Harrington survival curve \n by treatment")
```

```
survival_tab <- tibble(
  time = ovarian.fhfit$time,
  kmsurv = ovarian.kmfit$surv,
  kmstd = ovarian.kmfit$std.err,
  fhsurv = ovarian.fhfit$surv,
  fhstd = ovarian.fhfit$std.err,
  rx = c(rep(1,13), rep(2,13))
)
survival_tab %>% filter(rx == 1) %>% dplyr::select(-rx) %>% kable(longtable = T, booktabs = T, caption =

survival_tab %>% filter(rx == 2) %>% dplyr::select(-rx) %>% kable(longtable = T, booktabs = T, caption =
```