Survival Analysis HW 2

Victor Feagins

9/20/2021

## 4.1 Consider the survival time of the 30 melanoma patients in Table 3.1.

df.3.1 <- read\_excel("Table 3.1.xlsx")

## New names:  
## \* Uncensor -> Uncensor...10  
## \* Uncensor -> Uncensor...12

head(df.3.1, 10) %>%   
 flextable() %>%   
 set\_caption("First 10 Observations")

Table :First 10 Observations

| Patient | Age | Gender | Stage-II | Stage-III A | Stage-III B | Stage-IV A | Treatment | Remission Time | Uncensor...10 | Survival Time | Uncensor...12 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 59 | 0 | 0 | 0 | 1 | 0 | 1 | 33.7 | 0 | 33.7 | 0 |
| 2 | 50 | 0 | 0 | 0 | 1 | 0 | 1 | 3.8 | 1 | 3.9 | 1 |
| 3 | 76 | 1 | 0 | 0 | 1 | 0 | 1 | 6.3 | 1 | 10.5 | 1 |
| 4 | 66 | 0 | 0 | 0 | 1 | 0 | 1 | 2.3 | 1 | 5.4 | 1 |
| 5 | 33 | 1 | 0 | 0 | 1 | 0 | 1 | 6.4 | 1 | 19.5 | 1 |
| 6 | 23 | 0 | 0 | 0 | 1 | 0 | 1 | 23.8 | 0 | 23.8 | 0 |
| 7 | 40 | 0 | 0 | 0 | 1 | 0 | 1 | 1.8 | 1 | 7.9 | 1 |
| 8 | 34 | 1 | 0 | 0 | 1 | 0 | 1 | 5.5 | 1 | 16.9 | 0 |
| 9 | 34 | 1 | 0 | 0 | 1 | 0 | 1 | 16.6 | 0 | 16.6 | 0 |
| 10 | 38 | 0 | 1 | 0 | 0 | 0 | 1 | 33.7 | 0 | 33.7 | 0 |

### Compute and plot the PL estimates of the survivorship functions of the two treatment groups and check your results with Table 3.2 and Figure 3.1.

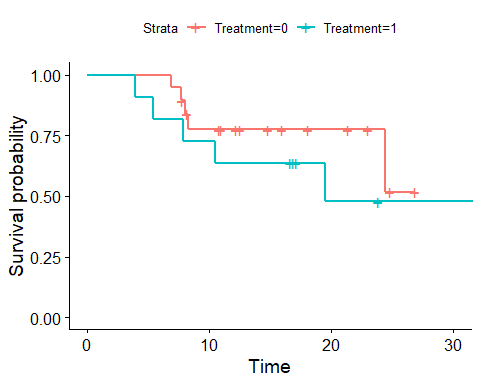
surv.model <- Surv(df.3.1$`Survival Time`, event = df.3.1$Uncensor...12)

km <- survfit(surv.model ~ Treatment, data = df.3.1)  
summary(km)

## Call: survfit(formula = surv.model ~ Treatment, data = df.3.1)  
##   
## Treatment=0   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 6.9 19 1 0.947 0.0512 0.852 1.000  
## 7.7 18 1 0.895 0.0704 0.767 1.000  
## 8.0 16 1 0.839 0.0854 0.687 1.000  
## 8.3 13 1 0.774 0.1003 0.601 0.998  
## 24.4 3 1 0.516 0.2211 0.223 1.000  
##   
## Treatment=1   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 3.9 11 1 0.909 0.0867 0.754 1.000  
## 5.4 10 1 0.818 0.1163 0.619 1.000  
## 7.9 9 1 0.727 0.1343 0.506 1.000  
## 10.5 8 1 0.636 0.1450 0.407 0.995  
## 19.5 4 1 0.477 0.1755 0.232 0.981

The survival times are the same with Table 3.2 in the book.

ggsurvplot(km)



The survival curves are the same as figure 3.1 in the book.

### Compute the variance of for every uncensored observation.

The following equation is used to calculate the variance of .

Where is number of observations and is number of deaths.

surv.data <- tidy(km)  
  
surv.var <- surv.data %>%  
 filter(n.event == 1) %>% #Looking only at uncensored observations  
 group\_by(strata) %>%   
 mutate(S.Var = estimate^2 \* cumsum(1/((max(n.risk)-cumsum(n.event))\*(max(n.risk)-cumsum(n.event) + 1))))  
  
surv.var %>%   
 select(time, strata, S.Var) %>%   
 flextable() %>%   
 set\_caption("Surival Function Variences")

Table :Surival Function Variences

| time | strata | S.Var |
| --- | --- | --- |
| 6.9 | Treatment=0 | 0.002624289 |
| 7.7 | Treatment=0 | 0.004956991 |
| 8.0 | Treatment=0 | 0.006943539 |
| 8.3 | Treatment=0 | 0.008414419 |
| 24.4 | Treatment=0 | 0.005008582 |
| 3.9 | Treatment=1 | 0.007513148 |
| 5.4 | Treatment=1 | 0.013523666 |
| 7.9 | Treatment=1 | 0.018031555 |
| 10.5 | Treatment=1 | 0.021036814 |
| 19.5 | Treatment=1 | 0.017256762 |

### Estimate the median survival times of the two groups

print(km)

## Call: survfit(formula = surv.model ~ Treatment, data = df.3.1)  
##   
## n events median 0.95LCL 0.95UCL  
## Treatment=0 19 5 NA 24.4 NA  
## Treatment=1 11 5 19.5 10.5 NA

For treatment 0 the median is undefined since the is still more then 50% of patients alive.

## 4.4 Consider the remission data of 42 patients with acute leukemia in Example 3.3.

## [1] "Placebo"

## [1] 1 1 2 2 3 4 4 5 5 8 8 8 8 11 11 12 12 15 17 22 23

## [1] "6-MP"

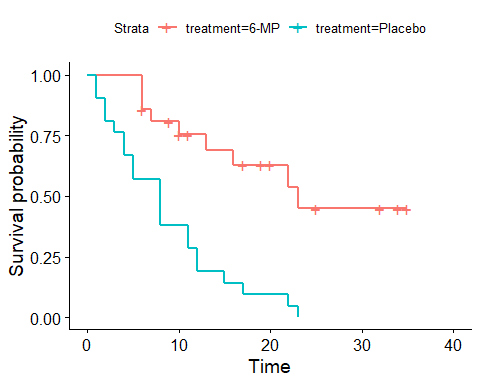
## [1] 6 6 6 7 10 13 16 22 23 6+ 9+ 10+ 11+ 17+ 19+ 20+ 25+ 32+ 32+  
## [20] 34+ 35+

### Compute and plot the PL estimates of at every time to relapse for the 6-MP and placebo groups.

km <- survfit(surv.model.3.3 ~ treatment, data = df.3.3)  
summary(km)

## Call: survfit(formula = surv.model.3.3 ~ treatment, data = df.3.3)  
##   
## treatment=6-MP   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 6 21 3 0.857 0.0764 0.720 1.000  
## 7 17 1 0.807 0.0869 0.653 0.996  
## 10 15 1 0.753 0.0963 0.586 0.968  
## 13 12 1 0.690 0.1068 0.510 0.935  
## 16 11 1 0.627 0.1141 0.439 0.896  
## 22 7 1 0.538 0.1282 0.337 0.858  
## 23 6 1 0.448 0.1346 0.249 0.807  
##   
## treatment=Placebo   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 1 21 2 0.9048 0.0641 0.78754 1.000  
## 2 19 2 0.8095 0.0857 0.65785 0.996  
## 3 17 1 0.7619 0.0929 0.59988 0.968  
## 4 16 2 0.6667 0.1029 0.49268 0.902  
## 5 14 2 0.5714 0.1080 0.39455 0.828  
## 8 12 4 0.3810 0.1060 0.22085 0.657  
## 11 8 2 0.2857 0.0986 0.14529 0.562  
## 12 6 2 0.1905 0.0857 0.07887 0.460  
## 15 4 1 0.1429 0.0764 0.05011 0.407  
## 17 3 1 0.0952 0.0641 0.02549 0.356  
## 22 2 1 0.0476 0.0465 0.00703 0.322  
## 23 1 1 0.0000 NaN NA NA

ggsurvplot(km)



### Compute the variances of in the 6-MP group and of in the placebo group.

surv.data <- tidy(km)  
  
surv.var <- surv.data %>%  
 filter(n.event == 1) %>% #Looking only at uncensored observations  
 group\_by(strata) %>%   
 mutate(S.Var = estimate^2 \* cumsum(1/((max(n.risk)-cumsum(n.event))\*(max(n.risk)-cumsum(n.event) + 1))))  
  
surv.var %>%   
 select(time, strata, S.Var) %>%   
 filter(time == 10 | time == 3) %>%   
 flextable() %>%   
 set\_caption("Surival Function Variences S(10) & S(3)")

Table :Surival Function Variences S(10) &amp; S(3)

| time | strata | S.Var |
| --- | --- | --- |
| 10 | treatment=6-MP | 0.004446435 |
| 3 | treatment=Placebo | 0.002134187 |

### Estimate the median remission times of the two treatment groups.

print(km)

## Call: survfit(formula = surv.model.3.3 ~ treatment, data = df.3.3)  
##   
## n events median 0.95LCL 0.95UCL  
## treatment=6-MP 21 9 23 16 NA  
## treatment=Placebo 21 21 8 4 12

## 4.9 Do a complete life-table analysis of the data given in Exercise Table 4.2, the Angina Pectoris data.

df.4.9 <- read\_excel("Female4.2.xlsx")  
head(df.4.9) %>%   
 flextable() %>%   
 set\_caption("First 5 Observations")

Table :First 5 Observations

| Interval | Number\_At\_Risk | Censored | Deaths |
| --- | --- | --- | --- |
| 0—1 | 555 | 0 | 82 |
| 1—2 | 473 | 8 | 30 |
| 2—3 | 435 | 8 | 27 |
| 3—4 | 400 | 7 | 22 |
| 4—5 | 371 | 7 | 26 |
| 5—6 | 338 | 28 | 25 |

This data is different format of survival data. R handles survival data where each row is an event. So we will need to transform the data by replicated the number of rows by the associated rows frequency

df.4.9.t <- rbind(data.d, data.c)  
df.4.9.t <- df.4.9.t %>%   
 tidyr::separate(Interval, c("Start", "End"), "—") %>%  
 mutate(Deaths = ifelse(is.na(End), 0, Deaths),# if we don't know when they died make it censored  
 End = ifelse(is.na(End), 15, End))

Once that transformation is done survival analysis is the same steps as the previous problems.

surv.model.4.9 <- Surv(#time = as.numeric(df.4.9.t$Start),  
 time = as.numeric(df.4.9.t$End),  
 event = df.4.9.t$Deaths)

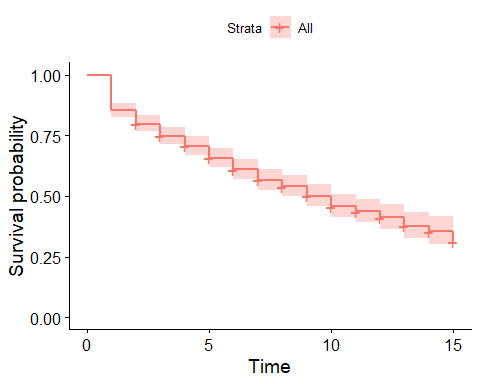
km <- survfit(surv.model.4.9 ~ 1, data = df.4.9.t)  
summary(km)

## Call: survfit(formula = surv.model.4.9 ~ 1, data = df.4.9.t)  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 1 555 82 0.852 0.0151 0.823 0.882  
## 2 473 30 0.798 0.0170 0.765 0.832  
## 3 435 27 0.749 0.0185 0.713 0.786  
## 4 400 22 0.707 0.0194 0.670 0.747  
## 5 371 26 0.658 0.0203 0.619 0.699  
## 6 338 25 0.609 0.0210 0.569 0.652  
## 7 285 20 0.566 0.0216 0.526 0.610  
## 8 234 11 0.540 0.0220 0.498 0.585  
## 9 191 14 0.500 0.0228 0.457 0.547  
## 10 153 13 0.458 0.0237 0.414 0.507  
## 11 113 5 0.438 0.0244 0.392 0.488  
## 12 86 5 0.412 0.0255 0.365 0.465  
## 13 58 5 0.377 0.0278 0.326 0.435  
## 14 35 2 0.355 0.0301 0.301 0.419  
## 15 24 3 0.311 0.0356 0.248 0.389

### Plot the three survival functions: survival, hazard and density functions.

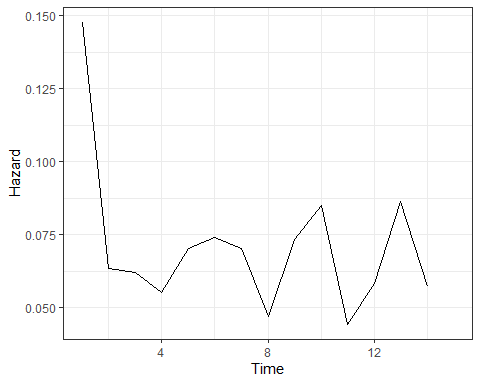
#### Survival Function

ggsurvplot(km)



#### Hazard Function

survival.table1 <- broom::tidy(km) %>% filter(n.event > 0)  
survival.table1 <- survival.table1 %>% mutate(hazard = n.event / (n.risk \* (lead(time) - time)))  
ggplot() +  
 geom\_line(data = survival.table1, aes(x = time, y = hazard)) +  
 labs(x = "Time", y = "Hazard")



### Probablity

survival.table1 <- survival.table1 %>% mutate(probability = estimate \* hazard)  
ggplot() +  
 geom\_line(data = survival.table1, aes(x = time, y = probability)) +  
 labs(x = "Time", y = "Probability Density")

