Lab4

Ziwen Zhang

2022-12-01

# Question 1

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

survtime <- c(59, 112, 152, 418, 434, 448, 461, 475, 477, 563, 638, 744, 769, 770, 803)  
status <- c(0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1)  
  
df <- data.frame(survtime, status)  
  
prob\_died <- c(1/15, 1/14, 0/13, 1/12, 1/11, 0/10, 0/9, 1/8, 0/7, 1/6, 1/5, 0/4, 0/3, 0/2, 0/1)  
  
prob\_surv <- 1 - prob\_died  
  
cumprod(prob\_surv) %>% round(5)

## [1] 0.93333 0.86667 0.86667 0.79444 0.72222 0.72222 0.72222 0.63194 0.63194  
## [10] 0.52662 0.42130 0.42130 0.42130 0.42130 0.42130

# Question 2

library(survival)

## Warning: package 'survival' was built under R version 4.2.2

library(flexsurv)

## Warning: package 'flexsurv' was built under R version 4.2.2

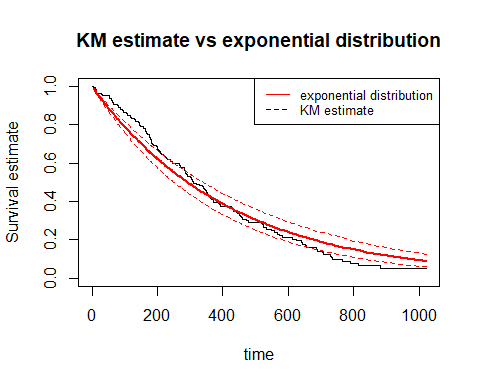
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.0 ✔ purrr 0.3.5  
## ✔ tibble 3.1.8 ✔ stringr 1.4.1  
## ✔ tidyr 1.2.1 ✔ forcats 0.5.2  
## ✔ readr 2.1.3

## Warning: package 'purrr' was built under R version 4.2.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

# estimate survival curve using exponential distribution  
ex\_fit <- flexsurvreg(Surv(time, status) ~ 1, data = cancer, dist = 'exp')  
  
plot(ex\_fit, main = "KM estimate vs exponential distribution", xlab = "time", ylab = "Survival estimate", conf.int = F)  
legend(x = "topright", legend=c("exponential distribution", "KM estimate"), col=c("red", "black"), lty=1:2, cex=0.8)

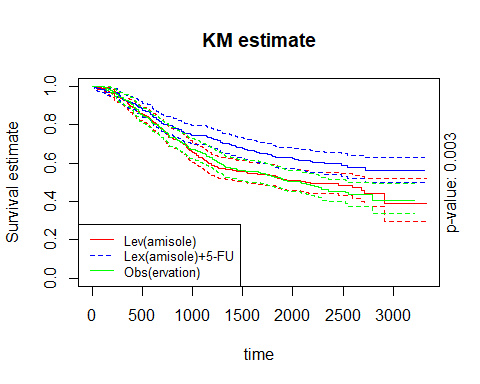
 When the time is short, the survival estimate from the KM method is slightly higher than that from the exponential distribution. When the time is longer than about 400 days, the survival estimate from the exponential distribution is slightly higher than that from the KM method.

# Question 3

head(colon)

## id study rx sex age obstruct perfor adhere nodes status differ extent  
## 1 1 1 Lev+5FU 1 43 0 0 0 5 1 2 3  
## 2 1 1 Lev+5FU 1 43 0 0 0 5 1 2 3  
## 3 2 1 Lev+5FU 1 63 0 0 0 1 0 2 3  
## 4 2 1 Lev+5FU 1 63 0 0 0 1 0 2 3  
## 5 3 1 Obs 0 71 0 0 1 7 1 2 2  
## 6 3 1 Obs 0 71 0 0 1 7 1 2 2  
## surg node4 time etype  
## 1 0 1 1521 2  
## 2 0 1 968 1  
## 3 0 0 3087 2  
## 4 0 0 3087 1  
## 5 0 1 963 2  
## 6 0 1 542 1

colon <- colon %>%  
 filter(etype == "2")  
  
colon\_Lev <- colon %>%  
 filter(rx == "Lev")  
  
colon\_Lev5FU <- colon %>%  
 filter(rx == "Lev+5FU")  
  
colon\_Obs <- colon %>%  
 filter(rx == "Obs")  
  
KM\_est\_Lev <- survfit(Surv(time, status) ~ 1, type = "kaplan-meier", conf.type ="log", data = colon\_Lev)  
  
KM\_est\_Lev5FU <- survfit(Surv(time, status) ~ 1, type = "kaplan-meier", conf.type ="log", data = colon\_Lev5FU)  
  
KM\_est\_Obs <- survfit(Surv(time, status) ~ 1, type = "kaplan-meier", conf.type ="log", data = colon\_Obs)  
  
  
plot(KM\_est\_Lev, main = "KM estimate", xlab = "time", ylab = "Survival estimate", col = "red")  
lines(KM\_est\_Lev5FU, col= "blue")  
lines(KM\_est\_Obs, col= "green")  
legend(x = "bottomleft", legend=c("Lev(amisole)", "Lex(amisole)+5-FU", "Obs(ervation)"), col=c("red", "blue", "green"), lty=1:2, cex=0.8)  
mtext("p-value: 0.003", side = 4)



logrank\_test <- survdiff(Surv(time, status) ~ rx, data = colon)  
logrank\_test

## Call:  
## survdiff(formula = Surv(time, status) ~ rx, data = colon)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## rx=Obs 315 168 148 2.58 3.85  
## rx=Lev 310 161 146 1.52 2.25  
## rx=Lev+5FU 304 123 157 7.55 11.62  
##   
## Chisq= 11.7 on 2 degrees of freedom, p= 0.003

library(survminer)

## Warning: package 'survminer' was built under R version 4.2.2

## Loading required package: ggpubr

##   
## Attaching package: 'survminer'

## The following object is masked from 'package:survival':  
##   
## myeloma

KM\_est\_colon <- survfit(Surv(time, status==0) ~ 1, type = "kaplan-meier", conf.type ="log", data = colon)  
   
surv\_median(KM\_est\_colon)

## Warning: `select\_()` was deprecated in dplyr 0.7.0.  
## ℹ Please use `select()` instead.  
## ℹ The deprecated feature was likely used in the dplyr package.  
## Please report the issue at <]8;;https://github.com/tidyverse/dplyr/issueshttps://github.com/tidyverse/dplyr/issues]8;;>.

## strata median lower upper  
## 1 All 2352 2313 2396

The estimated median follow-up time is 6.439 years.