

MAE0514 - Introdução a Análise de Sobrevida - Lista 2

Bruno de Castro Paul Schultze¹
Rubens Santos Andrade Filho²

Maio de 2021

Sumário

Questão 1	2
Questão 2	2
Questão 3	2
Questão 4	7
Questão 5	7
Questão 6	7
Questão 7	7
Questão 8	7

¹Número USP: 10736862

²Número USP: 10370336

Questão 1

a)

A variável do estudo é o tempo compreendido da exposição a um material cancerígeno até o desenvolvimento do tumor de um tamanho determinado nos ratos. Nesse caso, a origem é a exposição a um material cancerígeno e o evento de interesse é o desenvolvimento do tumor de um tamanho determinado.

b)

Para os rato A, B e C foram observados os tempos de falha, isto é, os tempos até os ratos desenvolverem o tumor de determinado tamanho.

Para o rato D foi observado uma censura aleatória à direita na vigésima semana, sua morte. Até a semana 20 o rato não tinha desenvolvido o tumor de um tamanho determinado.

Para os ratos E e F foram observados censuras à direita do tipo I na semana 30 por ser a do estudo. Entretanto no enunciado não está claro se todos os ratos foram expostos ao material cancerígeno ao mesmo tempo para dizermos se a censura é generalizada ou não.

Questão 2

Questão 3

Em um estudo clínico realizado com pacientes com câncer gástrico avançado (com metástase linfodonal), uma quimioterapia com Xeloda (capecitabina) e oxaliplatina foi administrada antes da cirurgia de 48 pacientes. Nesse tipo de ensaio clínico, é de interesse estudar e avaliar o tempo livre da doença, que é o tempo que o paciente fica bem, vivo e sem a doença. Assim, um dos objetivos é estudar o tempo decorrido entre o início do tratamento e óbito ou progressão da doença (o que ocorrer primeiro). Os dados do tempo livre da doença (em semanas) dos 48 pacientes estão disponíveis no arquivo `Lista2-Xelox.csv`, sendo que a variável delta é codificada como sendo 1 se o evento ocorreu e 0 se a observação é censurada.

(a) Calculamos o estimador da tabela de vida, considerando as seguintes faixas de tempo:

Faixa 1:	8 semanas (inclusive) ou menos
Faixa 2:	de 8 a 16 semanas (inclusive)
Faixa 3:	de 16 a 24 semanas (inclusive)
Faixa 4:	de 24 a 32 semanas (inclusive)
Faixa 5:	de 32 a 44 semanas (inclusive)
Faixa 6:	de 44 a 56 semanas (inclusive)
Faixa 7:	mais de 56 semanas

Dessa forma, consideramos os intervalos fechados à direita.

```
library(readr)
library(dplyr)
```

```

dados_raw <- readr::read_csv2('data/Lista2-Xelox.csv')

# limites dos intervalos
breaks <- c(0,8,16,24,32,44,56, Inf)

dados <- dados_raw %>%
  mutate(
    # define as faixas
    intervalo = cut(timeWeeks, breaks=breaks, right=TRUE),
    j = as.integer(intervalo)
  )

tabua <- dados %>%
  group_by(intervalo, j) %>%
  summarise(
    # numero de falhas no intervalo
    d = sum(delta),
    # numero de censuras no intervalo
    w = sum(delta==0)
  ) %>%
  ungroup() %>%
  mutate(
    # numero de obs em risco, que nao falharam até o fim do intervalo anterior
    n_estrela = sum(d+w) - cumsum(d+w) +w+d,
    # corrigindo o numero de ind. em risco
    n = n_estrela - w/2,
    # prop. de falhas no intervalo
    q_hat = d/n,
    p_hat = 1 - q_hat,
    # na tabua de vida, a estimativa de S do 1o intervalo = 1
    # depois o produtorio acumulado dos p_i
    s_hat = c(1, cumprod(p_hat)[-n()])
  )

tabua

```

```

## # A tibble: 7 x 9
##   intervalo     j     d     w n_estrela     n  q_hat p_hat s_hat
##   <fct>       <int> <dbl> <int>      <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1 (0,8]         1     4     0     48  48  0.0833 0.917 1
## 2 (8,16]        2     6     0     44  44  0.136  0.864 0.917
## 3 (16,24]       3     6     0     38  38  0.158  0.842 0.792
## 4 (24,32]       4     4     0     32  32  0.125  0.875 0.667
## 5 (32,44]       5     4     1     28  27.5 0.145  0.855 0.583
## 6 (44,56]       6     2     4     23  21  0.0952 0.905 0.498
## 7 (56,Inf]      7     6    11     17  11.5 0.522  0.478 0.451

```

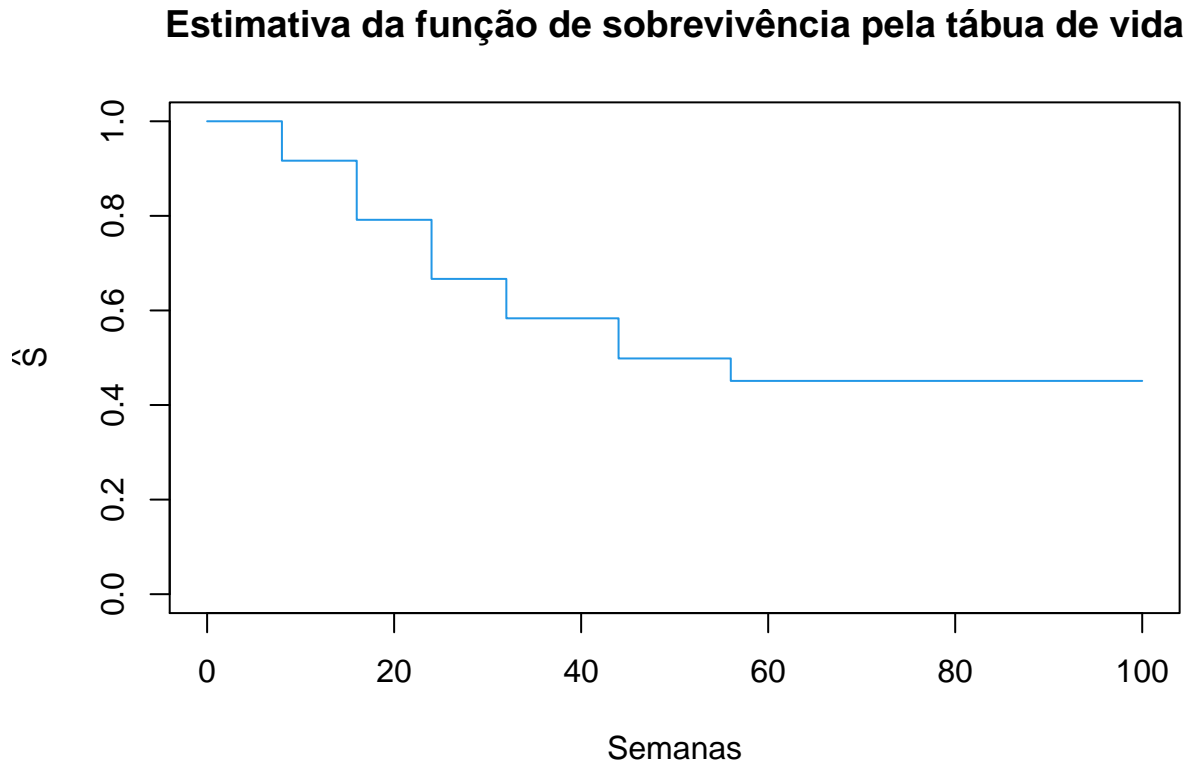
```

# library(KMsurv)
# fitlt = lifetab(breaks, 48, nlost = tab$w , nevent = tab$d)

x = rep(breaks, each=2)[2:15]
x[length(x)] <- 100 # substitui infinito
y = rep(tabua$s_hat, each=2)

```

```
plot(x, y, type="l", col=4, xlab="Semanas", ylab=expression(hat(S)),ylim = c(0,1),
     main = "Estimativa da função de sobrevivência pela tábua de vida", cex=.6)
```



Chama a atenção o fato da estimativa da função de sobrevivência não se aproximar de 0 à medida que aumentam o número de semanas. Isso acontece principalmente devido às 11 observações censuradas no último intervalo.

(b) Calcule o estimador Kaplan-Meier para os dados (você pode utilizar um software).

```
library(survival)

sfit <- survfit(Surv(dados$timeWeeks, dados$delta)~1, data=dados)

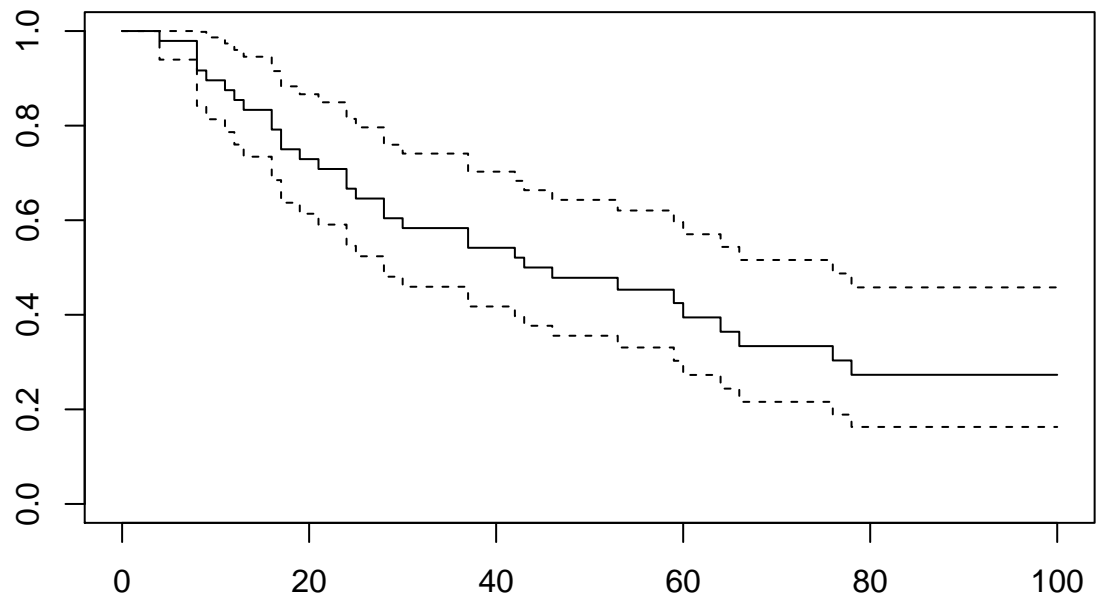
summary(sfit)
```

```
## Call: survfit(formula = Surv(dados$timeWeeks, dados$delta) ~ 1, data = dados)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    4      48      1   0.979  0.0206    0.940    1.000
##    8      47      3   0.917  0.0399    0.842    0.998
##    9      44      1   0.896  0.0441    0.813    0.987
##   11      43      1   0.875  0.0477    0.786    0.974
##   12      42      1   0.854  0.0509    0.760    0.960
```

##	13	41	1	0.833	0.0538	0.734	0.946
##	16	40	2	0.792	0.0586	0.685	0.915
##	17	38	2	0.750	0.0625	0.637	0.883
##	19	36	1	0.729	0.0641	0.614	0.866
##	21	35	1	0.708	0.0656	0.591	0.849
##	24	34	2	0.667	0.0680	0.546	0.814
##	25	32	1	0.646	0.0690	0.524	0.796
##	28	31	2	0.604	0.0706	0.481	0.760
##	30	29	1	0.583	0.0712	0.459	0.741
##	37	28	2	0.542	0.0719	0.418	0.703
##	42	26	1	0.521	0.0721	0.397	0.683
##	43	25	1	0.500	0.0722	0.377	0.663
##	46	23	1	0.478	0.0722	0.356	0.643
##	53	19	1	0.453	0.0727	0.331	0.620
##	59	16	1	0.425	0.0735	0.303	0.596
##	60	14	1	0.394	0.0742	0.273	0.570
##	64	13	1	0.364	0.0744	0.244	0.544
##	66	12	1	0.334	0.0742	0.216	0.516
##	76	11	1	0.303	0.0734	0.189	0.487
##	78	10	1	0.273	0.0720	0.163	0.458

```
plot(sfit,xmax=100, main="Estimativa da função de sobrevivência por Kaplan-Meier")
```

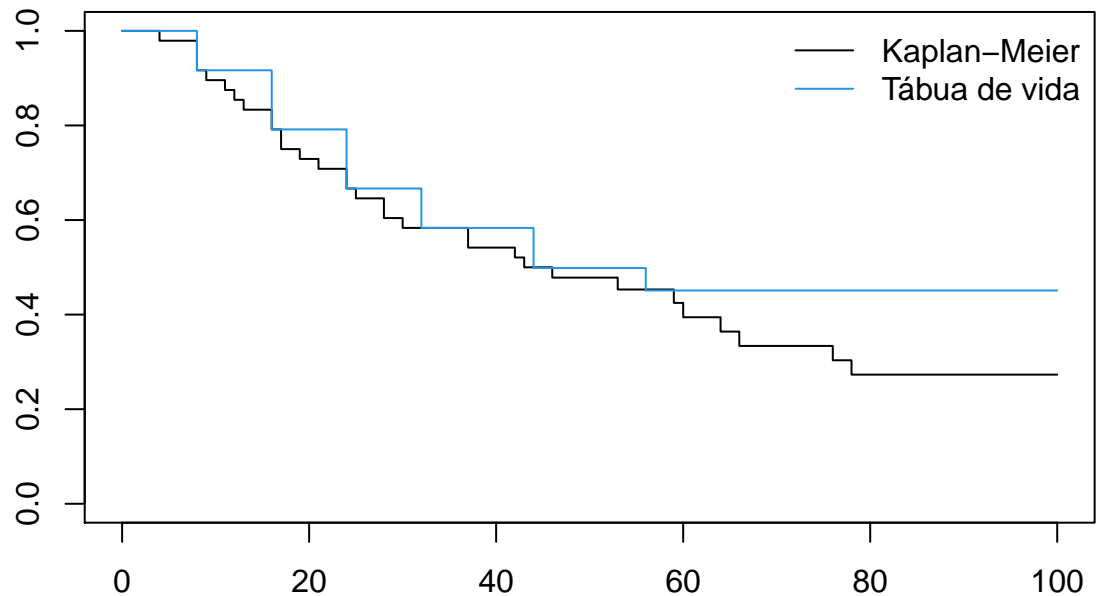
Estimativa da função de sobrevivência por Kaplan-Meier



(c) Coloque em um mesmo gráfico as duas curvas estimadas nos itens anteriores. Compare as curvas e comente.

```
plot(sfit,xmax=100, main="Estimativas da função de sobrevivência", conf.int = F)
lines(x, y, type="l", col=4)
legend("topright",legend=c("Kaplan-Meier","Tábua de vida"),lty = c(1, 1),
      col = c(1,4), bty="n")
```

Estimativas da função de sobrevivência

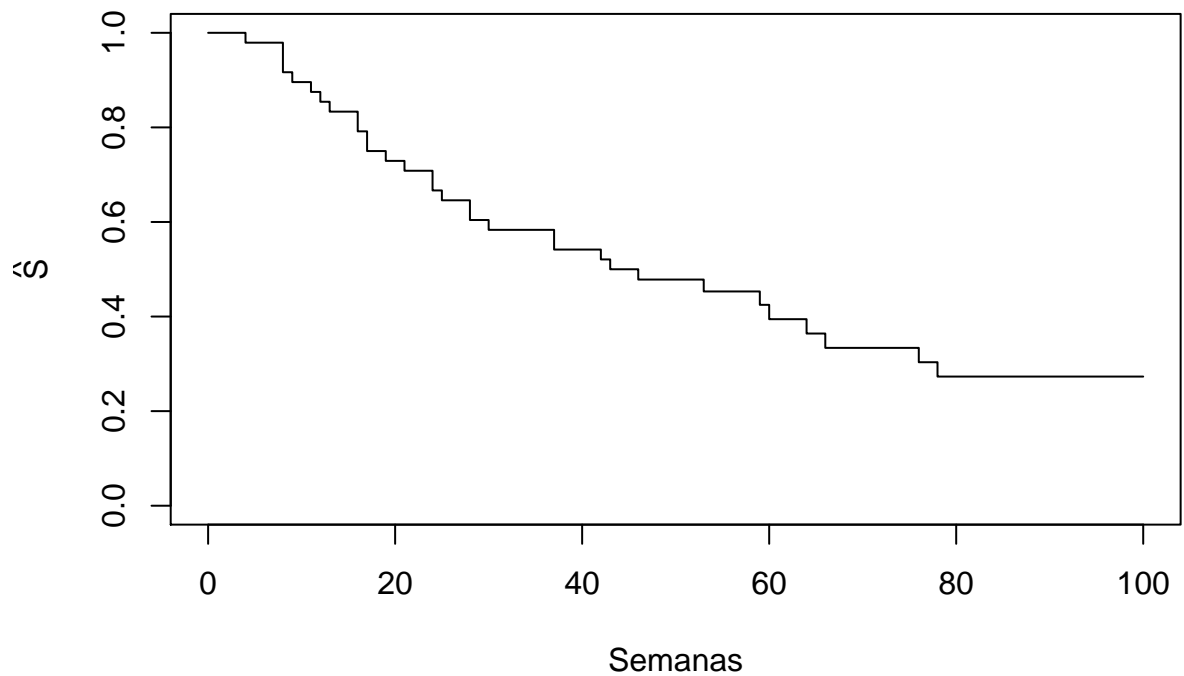


O método de Kaplan-Meier melhora visualmente a estimativa da curva da função de sobrevivência, principalmente nas semanas ≥ 56 , onde o número de censuras é maior. Além disso, enquanto que a tábua de vida

```
kmeier <- dados_raw %>%
  group_by(t=timeWeeks) %>%
  # numero de eventos e censuras em cada t
  summarise(d = sum(delta), w=sum(delta==0)) %>%
  ungroup() %>%
  mutate(
    # numero de individuos vivos até antes de cada instante t
    Y = sum(d+w) - (cumsum(d+w) - (d+w)),
    # estimate of the conditional probability that an individual who survives
    # to just prior to time ti experiences the event at time ti
    q = d/Y,
    # estimate of surv function
    s_hat = cumprod(1 - q)
  ) %>%
  filter(d!=0)
```

```
x <- c(0, rep(kmeier$t, each=2), 100)
y <- c(1, 1, rep(kmeier$s_hat, each=2))
plot(x, y, type="l", col=1, xlab="Semanas", ylab=expression(hat(S)), ylim = c(0,1),
     main = "Estimativa da função de sobrevivência por Kaplan-Meier", cex=.6)
```

Estimativa da função de sobrevivência por Kaplan-Meier



Questão 4

Questão 5

Questão 6

Questão 7

Questão 8