

ANÁLISE DE SOBREVIVÊNCIA

Teoria e aplicações em saúde

Caderno de Respostas

Capítulo 4

Estimação não-paramétrica

Exercício 4.1: Considere os dados de tempo de aleitamento de 15 crianças, descrito no exercício 2.1.

6 12 10 3 5 1 6 8 1 5 2 2 5 8 1

a. Complete a tabela Kaplan-Meier abaixo com estes dados

Resposta:

Vamos calcular a função de sobrevivência passo a passo (a função de risco será calculada no próximo exercício).

T	Intervalo	N(t)	R(t)	$\Delta(t)$	(a) $S_{km}(t)$	(b) $S_{km}(t)$	$S_{km}(t)$
0		0					1,0000
1	(0,1]	3	15	1	$1 \times (12/15)$	12/15	0,8000
2	(1,2]	2	12	1	$0,8 \times (10/12)$	10/15	0,6667
3	(2,3]	1	10	1	$0,67 \times (9/10)$	9/15	0,6000
5	(3,5]	3	9	2	$0,6 \times (6/9)$	6/15	0,4000
6	(5,6]	2	6	1	$0,4 \times (4/6)$	4/15	0,2667
8	(6,8]	2	4	2	$0,27 \times (2/4)$	2/15	0,1333
10	(8,10]	1	2	2	$0,13 \times (1/2)$	1/15	0,0667
12	(10,12]	1	1	2	$0,07 \times (0/1)$	0/15	0,0000

Como neste estudo não há relato de censura a estimativa de KM é equivalente a estimativa realizada no capítulo anterior e a função de sobrevivência de Kaplan-Meier (KM) pode ser obtida por:

$$S_{km}(t) = \frac{S_{km}(t-1) * R(t)}{R(t-1)} \quad \text{ou} \quad S_{km}(t) = \frac{R(t)}{N} = S(t)$$

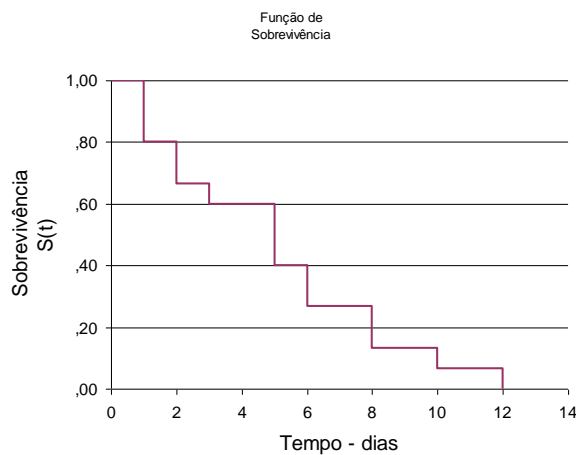
A diferença entre o método de K-M e o anterior, fica restrita ao nº de pessoas em risco em cada intervalo avaliado: quando há censura o indivíduo é retirado do grupo em risco.

No livro se considera que o paciente ainda está em risco no intervalo em que ocorre o evento, dessa forma apenas no intervalo seguinte o número de indivíduos é reduzido. Este formato difere do adotado pelo R, no qual a sobrevivência é 1 apenas no tempo zero. Observando o gráfico a seguir podemos observar que estas notações podem ser consideradas equivalentes.

b. Desenhe a curva Kaplan-Meier correspondente aos dados

Resposta: Para obter o efeito de escada no gráfico, elaboramos a tabela abaixo:

Tempo	S(t): sobrevivência
0	1,00000
1	1,00000
1	0,80000
2	0,80000
2	0,66670
3	0,66670
3	0,60000
5	0,60000
5	0,40000
6	0,40000
6	0,26670
8	0,26670
8	0,13330
10	0,13300
10	0,06670
12	0,06670
12	0,00000



- c. Qual a probabilidade de uma criança ser amamentada pelo menos até o sexto mês de vida?

Resposta: A probabilidade de uma criança ser amamentada pelo menos até o sexto mês de vida é $S(6) = 0,2667$

- d. Qual a probabilidade de ser amamentada por mais de 3 meses?

Resposta: A probabilidade de ser amamentada por mais de 3 meses é $S(3) = 0,6$

- e. Qual é a probabilidade de ser amamentada por mais de 10 meses?

Resposta: A probabilidade de ser amamentada por mais de 10 meses é $S(10) = 0,0667$

- f. Qual é o tempo mediano de aleitamento?

Resposta: Observando a tabela construída no item a, podemos dizer que o tempo mediano de aleitamento está entre 5 e 6 meses, mas observando o gráfico do item b verificamos que é igual a 5 meses.

Exercício 4.2: Ainda com os dados do aleitamento construa a tabela de risco acumulado usando os estimadores de Kaplan-Meier e Nelson-Aalen.

Resposta: Usando as formulas:

$$\Lambda_{km}(t) = -\ln(S_{KM}(t))$$

$$\Lambda_{NA}(t) = \Lambda(t-1) + [N(t)/R(t)]$$

$$\Lambda_{NA}(t) = \sum [N(t_j)/R(t_j)]$$

Completamos a tabela:

t	Intervalo	N(t)	R(t)	$\Delta(t)$	$\Lambda_{km}(t)$	$\Lambda_{NA}(t)$
0		0				
1	(0,1]	3	15	1	0,0000	0,0000
2	(1,2]	2	12	1	0,2231	0,2000
3	(2,3]	1	10	1	0,4055	0,3667
5	(3,5]	3	9	2	0,5108	0,4667
6	(5,6]	2	6	1	0,9163	0,8000
8	(6,8]	2	4	2	1,3218	1,1333
10	(8,10]	1	2	2	2,0149	1,6333
12	(10,12]	1	1	2	2,7081	2,1333

Exercício 4.3 Os dados completos de tempo de aleitamento (estudo fictício) estão presentes no arquivo leite.txt, Apêndice 12.4. Abra este banco de dados no R e:

- a. Construa a tabela Kaplan-Meier. Verifique qual é o tempo mediano de aleitamento na saída do R.

Resposta:

```
# Exercício 4.3
```

```
# Item a
```

```
leite <- read.table("leite.txt", header = T, sep = "")
```

```
# Carregando o pacote de Sobrevivência
```

```
library(survival)
```

```
# Criando o objeto que contém os dados de sobrevivência
```

```
y <- Surv(leite$tempo, leite$status)
```

```
# Obtendo as estimativas de sobrevivência por Kaplan-Meier (KM)
```

```
km <- survfit(y ~ 1, data = leite)
```

```
# Observando a Tabela de KM
```

```
summary(km)
```

```
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  1     15      3  0.8000  0.1033  0.6212  1.000
  2     12      2  0.6667  0.1217  0.4661  0.953
  3     10      1  0.6000  0.1265  0.3969  0.907
  5      9      3  0.4000  0.1265  0.2152  0.743
  6      6      2  0.2667  0.1142  0.1152  0.617
  8      4      2  0.1333  0.0878  0.0367  0.484
 10      2      1  0.0667  0.0644  0.0100  0.443
 12      1      1  0.0000      NA      NA      NA
```

Observando o tempo mediano estimado por KM

km

```
Call: survfit(formula = y ~ 1, data = leite)
records  n.max n.start  events  median 0.95LCL 0.95UCL
      15     15     15     15      5      2      8
```

Na saída do R observamos os mesmos resultados obtidos no exercício 4.1. O tempo mediano é igual a 5 meses, com intervalo de confiança de 95% de 2 a 8 meses. E a probabilidade de desmame em até 5 meses é 40%.

- b. Construa a curva de sobrevivência de aleitamento indicando o seu ponto mediano.

Resposta:

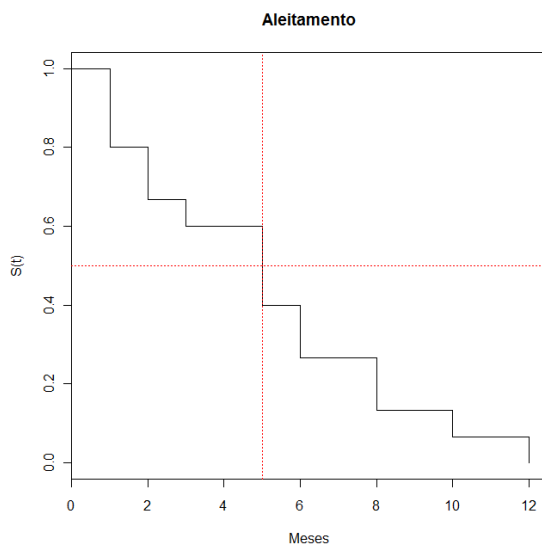
Item b

Fazendo o gráfico da sobrevivência com mediana

```
plot(km, conf.int = F, xlab = "Meses", ylab = "S(t)", main = "Aleitamento")
```

Adicionando a linha da mediana

```
abline(v=5,h=0.5,col=2,lty=3)
```



O tempo mediano para a ocorrência do desmame observado no gráfico do R é de 5 meses.

- c. Calcule as estimativas de Nelson-Aalen usando a função `coxph()`.

Resposta:

Item c

Obtendo as estimativas por Nelson-Aalen (NA)

Usando a função `coxph` para NA

```
y <- coxph (Surv(leite$tempo, leite$status))
```

```
na <- survfit(y ~ 1, data = leite)
```

Observando a Tabela de estimativas por NA

```
summary(na)
```

```
Call: survfit.coxph.null(formula = coxph(y ~ 1, data = leite))
```

time	n.risk	n.event	survival	std.err	lower	95% CI upper	95% CI
1	15	3	0.8065	0.1003	0.63208	1.000	
2	12	2	0.6776	0.1187	0.48070	0.955	
3	10	1	0.6131	0.1236	0.41291	0.910	
5	9	3	0.4197	0.1252	0.23384	0.753	

```

6      6      2  0.2909  0.1152      0.13384      0.632
8      4      2  0.1623  0.0933      0.05261      0.501
10     2      1  0.0984  0.0750      0.02212      0.438
12     1      1  0.0362  0.0455      0.00308      0.426

```

Observando o tempo mediano estimado por NA
na

```

Call: survfit.coxph.null(formula = coxph(y ~ 1, data = leite))
records  n.max n.start  events  median 0.95LCL 0.95UCL
      15     15     15     15      5      2      10

```

Observando a tabela de sobrevivência obtida pelo método de Nelson-Aalen encontra-se que o tempo mediano é também de 5 meses, mas a probabilidade de desmame em até 5 meses é 42%.

Exercício 4.4: Em que situação se deve dar preferência ao método de Nelson-Aalen?

Resposta: O método de Nelson-Aalen é mais indicado quando se analisa dados de amostra pequena. Quando a amostra é grande os resultados dos dois métodos são equivalentes.

Exercício 4.5: Suponha que os tempos de aleitamento de 60 bebês estejam agrupados em quatro comunidades diferentes:

Comunidade 1:	6	12	10	3	5	1	6	8	1	5	2	2	5	8	1
Comunidade 2:	5	12	10	4	4	3	6	9	2	6	4	1	7	10	1
Comunidade 3:	13	14	20	3	5	1	8	15	2	5	3	2	6	15	1
Comunidade 4:	1	16	20	1	1	1	1	2	2	13	3	1	1	14	2

O arquivo leite2.txt contém esses dados (descrito no Apêndice 12.4) contem esses dados. A variável grupo indica a comunidade à qual cada criança pertence. Carregue este arquivo no R e:

- Ajuste um modelo Kaplan-Meier estratificado por comunidade e compare o tempo mediano de aleitamento em cada comunidade.

Resposta:

```

# Exercício 4.5
# Ajustando um modelo KM estratificado por comunidade
# Lendo o novo banco de dados
leite2 <- read.table("leite2.txt", header = T, sep = "")
leite2
  crianca tempo status grupo
1         1      6      1      1
2         2     12      1      1
3         3     10      1      1
...
...
58        58      1      1      4
59        59     14      1      4
60        60      2      1      4

# Carregando o pacote de Sobrevivência
library(survival)
# item a
# Indicando a variável de sobrevivência

```

```

y2 <- Surv(leite2$tempo, leite2$status)
# Obtendo as estimativas por KM
km2 <- survfit(y2 ~ grupo, data = leite2)
# Tabela de estimativas de sobrevivência
summary(km2)
Call: survfit(formula = y2 ~ grupo, data = leite2)

```

```

      grupo=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  1     15      3  0.8000  0.1033  0.6212  1.000
  2     12      2  0.6667  0.1217  0.4661  0.953
  3     10      1  0.6000  0.1265  0.3969  0.907
  5      9      3  0.4000  0.1265  0.2152  0.743
  6      6      2  0.2667  0.1142  0.1152  0.617
  8      4      2  0.1333  0.0878  0.0367  0.484
 10      2      1  0.0667  0.0644  0.0100  0.443
 12      1      1  0.0000      NaN      NA      NA

```

```

      grupo=2
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  1     15      2  0.8667  0.0878  0.7106  1.000
  2     13      1  0.8000  0.1033  0.6212  1.000
  3     12      1  0.7333  0.1142  0.5405  0.995
  4     11      3  0.5333  0.1288  0.3322  0.856
  5      8      1  0.4667  0.1288  0.2717  0.802
  6      7      2  0.3333  0.1217  0.1630  0.682
  7      5      1  0.2667  0.1142  0.1152  0.617
  9      4      1  0.2000  0.1033  0.0727  0.550
 10      3      2  0.0667  0.0644  0.0100  0.443
 12      1      1  0.0000      NaN      NA      NA

```

```

      grupo=3
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  1     15      2  0.8667  0.0878  0.7106  1.000
  2     13      2  0.7333  0.1142  0.5405  0.995
  3     11      2  0.6000  0.1265  0.3969  0.907
  5      9      2  0.4667  0.1288  0.2717  0.802
  6      7      1  0.4000  0.1265  0.2152  0.743
  8      6      1  0.3333  0.1217  0.1630  0.682
 13      5      1  0.2667  0.1142  0.1152  0.617
 14      4      1  0.2000  0.1033  0.0727  0.550
 15      3      2  0.0667  0.0644  0.0100  0.443
 20      1      1  0.0000      NaN      NA      NA

```

```

      grupo=4
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  1     15      7  0.5333  0.1288  0.3322  0.856
  2      8      3  0.3333  0.1217  0.1630  0.682
  3      5      1  0.2667  0.1142  0.1152  0.617
 13      4      1  0.2000  0.1033  0.0727  0.550
 14      3      1  0.1333  0.0878  0.0367  0.484
 16      2      1  0.0667  0.0644  0.0100  0.443
 20      1      1  0.0000      NaN      NA      NA

```

```

# Tabela com eventos observados e mediana
km2

```

```

Call: survfit(formula = y2 ~ grupo, data = leite2)

```

```

      records n.max n.start events median 0.95LCL 0.95UCL
grupo=1     15    15     15     15      5      2      8
grupo=2     15    15     15     15      5      4     10
grupo=3     15    15     15     15      5      3     15

```

grupo=4 15 15 15 15 2 1 14

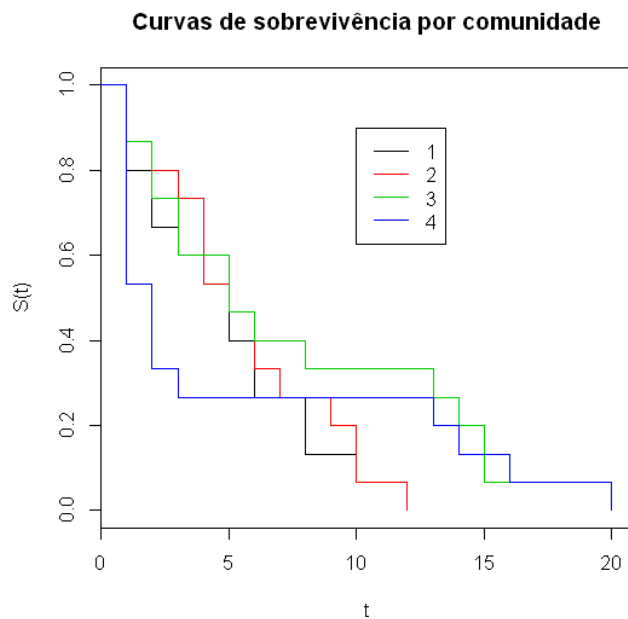
O tempo mediano de amamentação é menor para a comunidade 4 (igual a 2 meses) e é igual a 5 meses para as outras comunidades.

- b. Desenhe, no mesmo gráfico, as curvas de Kaplan-Meier, estratificadas por grupo. Como se comporta a curva de sobrevivência das outras comunidades quando comparadas a comunidade 1?

Resposta:

item b

```
# Fazendo o gráfico KM estratificado por comunidade
plot(km2, col = c(1:4), xlab = "t", ylab = "S(t)", conf.int = F)
legend(10, 0.9, c("1", "2", "3", "4"), lty = 1, col = c(1:4))
title("Curvas de sobrevivência por comunidade")
```



No gráfico pode-se observar que na comunidade 4 a probabilidade de amamentar é menor que em todas as outras (até o 6º mês, aproximadamente). Após 6 meses pouco mais 20% das crianças desta comunidade continuavam a ser amamentadas. O comportamento da comunidade 2 é semelhante ao da comunidade 1. A comunidade 3 apresenta maior probabilidade de amamentar, praticamente, ao longo de todo o tempo de estudo.

- c. Use os testes log-rank e de Peto para a variável grupo. Analise as funções de sobrevivência das quatro comunidades. A que conclusões você chega com esta análise? Existe diferença entre as comunidades quanto ao tempo de aleitamento?

Resposta:

item c

```
# Teste Log-rank (comando survdiff)
```

```
survdiff(y2 ~ grupo, data = leite2)
```

```
Call:
```

```
survdiff(formula = y2 ~ grupo, data = leite2)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
grupo=1	15	15	12.4	0.5489	0.8735
grupo=2	15	15	13.4	0.1862	0.3028


```

grupo=3 15      15      19.7      1.1220      2.1534
grupo=4 15      15      14.5      0.0182      0.0323

```

Chisq= 2.5 on 3 degrees of freedom, p= 0.478

Teste Peto (comando survdiff com rho=1)

survdiff(y2 ~ grupo, data = leite2, rho=1)

Call:

```
survdiff(formula = y2 ~ grupo, data = leite2, rho = 1)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
grupo=1	15	8.48	8.22	0.00865	0.0187
grupo=2	15	7.57	8.70	0.14764	0.3287
grupo=3	15	6.97	9.43	0.64499	1.5446
grupo=4	15	10.28	6.95	1.59872	3.3897

Chisq= 4.1 on 3 degrees of freedom, p= 0.254

Apesar da diferença visual, o teste do Log-rank e o Peto não rejeitam a hipótese nula de igual distribuição dos tempos de amamentação entre as comunidades. Desta forma não se pode afirmar que exista diferença significativa entre o tempo de amamentação entre as comunidades ($p > 0,10$).

Exercício 4.6: O banco de dados ipec.csv contém os dados de uma coorte de pacientes com Aids atendidos no IPEC e descritos no Apêndice 12.4. Para este estudo foi definido, como tempo de sobrevivência, o tempo entre o diagnóstico de Aids (critério CDC-1993) e o óbito.

a. Amplie a análise exploratória dos dados no R:

i. Qual é a idade média desses pacientes?

Resposta:

```
# Exercício 4.6
```

```
# Relendo o banco de dados
```

```
ipec <- read.table("ipec.csv", header = T, sep = ";")
```

```
# item a
```

```
# Ampliando a análise Exploratória
```

```
mean(ipec$idade)
```

```
[1] 36.55440
```

```
mean(ipec$idade [ipec$sexo == 'M'])
```

```
[1] 36.73611
```

```
mean(ipec$idade [ipec$sexo == 'F'])
```

```
[1] 36.02041
```

A idade média da coorte é 36,6 anos. Observa-se que as mulheres são um pouco mais jovens que os homens no momento do diagnóstico da Aids. A idade média das mulheres é 36,0 e dos homens é 36,7 anos. Observando gráficos de idade por sexo

```
# Gráfico da distribuição da idade por sexo:
```

```
boxplot(idade ~ sexo, main = "Idade por Sexo", ylab = "Idade", xlab = "Sexo", data=ipec)
```

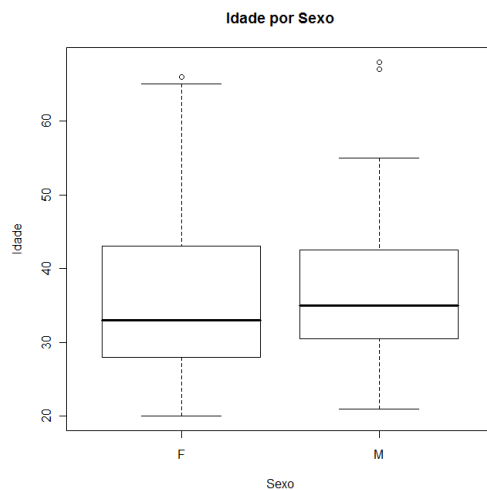
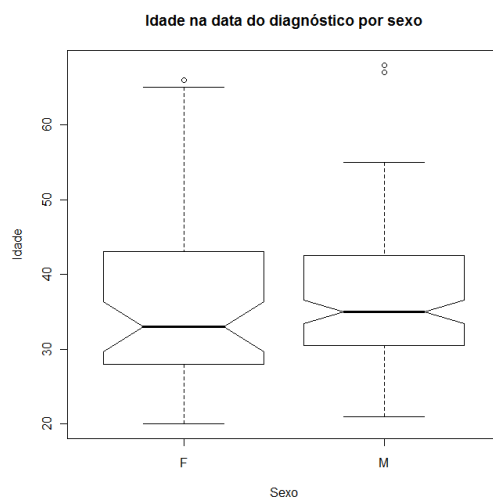


Gráfico da distribuição da idade por sexo:
`boxplot(idade ~ sexo,main="Idade na data do diagnóstico por sexo",notch = T,
ylab = "Idade", xlab = "Sexo",data=ipec)`



ii. Qual é a razão homens:mulheres nessa coorte?

Resposta:
`table(ipec$sexo)`

```
sexo
  F   M
49 144
```

Foram acompanhados 193 pacientes, sendo 49 mulheres e 144 homens. A relação homem:mulher nessa coorte de pacientes com Aids é igual a $144/49 = 2.94$, ou seja, quase 3 homens para cada mulher.

iii. Quantos receberam tratamento?

Resposta:
`table(ipec$tratam)`

```
tratam
  0   1   2   3
44 100 35 14
```

Dos 193 pacientes 100 receberam monoterapia, 35 receberam terapia combinada e 14 receberam terapia potente (HAART). Ou seja, 149 receberam terapia e 44 não morreram antes da terapia estar disponível.

iv. Quantos foram a óbito e quantos foram censurados?

Resposta:

```
table(ipec$status)
```

```
status
  0    1
103  90
```

```
table(ipec$obito)
```

```
obito
  I  N  S
21 80 92
```

```
table(ipec$status, ipec$obito)
```

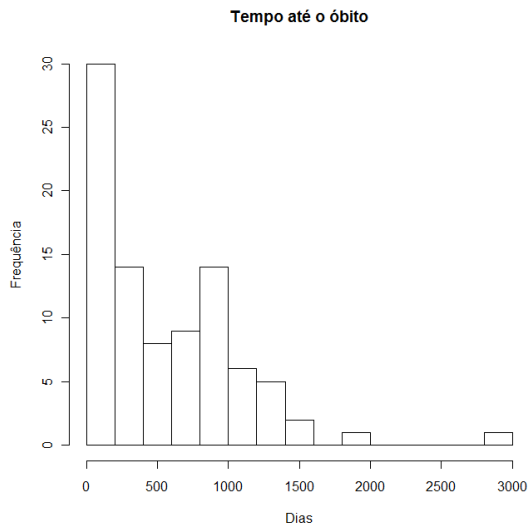
```
      obito
status I  N  S
  0    21 80  2
  1    0  0 90
```

Noventa pacientes foram a óbito e 103 foram censurados. Os motivos da censura foram: 2 óbitos por causa não relacionada a Aids, 80 estavam vivos no fim do estudo e 21 por perda de seguimento.

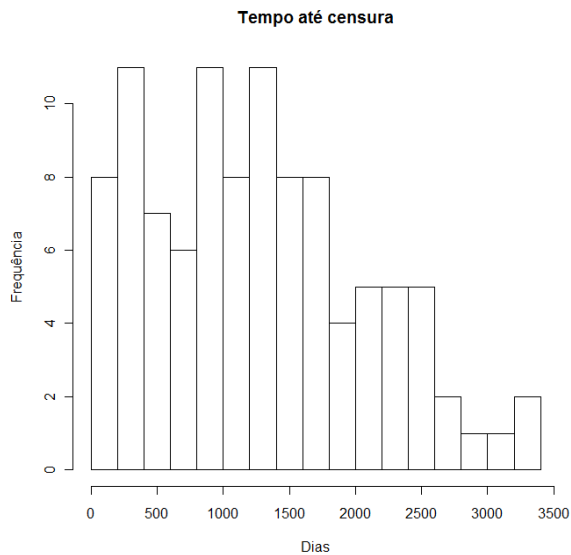
v. Os óbitos e as censuras se concentraram no fim do estudo ou ocorreram durante todo o período?

Resposta:

```
# Distribuição do tempo de sobrevivência dos pacientes que morreram
hist(ipec$tempo[ipec$status == 1], breaks = 12, main = "Tempo até o óbito",
ylab = "Frequência", xlab = "Dias")
```



```
# Distribuição do tempo de sobrevivência dos pacientes que sobreviveram
# até o fim do estudo
hist(ipec$tempo[ipec$status == 0], breaks = 12, ylab = "Frequência", xlab =
"Dias", main = "Tempo até censura")
```



Resposta:

Pelo histograma dos tempos dos pacientes que foram censurados observa-se que foram distribuídos ao longo do tempo. O mesmo não se pode dizer em relação os pacientes que foram a óbito. Os óbitos foram mais freqüentes até 1.000 dias de acompanhamento.

vi. Qual o tempo de sobrevivência mediano?

Resposta:

Análise de sobrevivência por Kaplan-Meier

Indicando as variáveis de sobrevivência

library(survival)

Fazendo KM para os dados do IPEC

KM <- survfit(Surv(tempo, status) ~ 1, data = ipec)

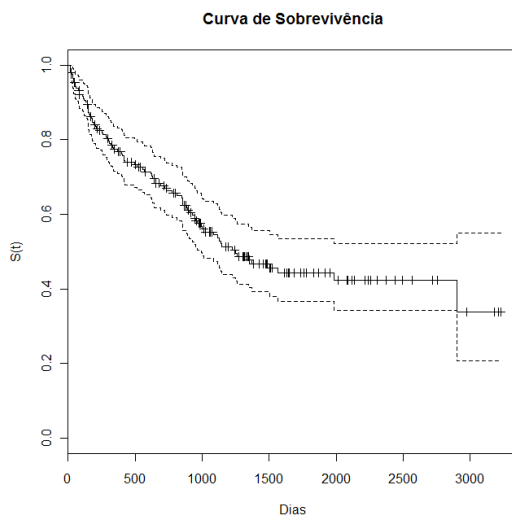
KM

Call: survfit(formula = Surv(tempo, status) ~ 1, data = ipec)

n	events	median	0.95LCL	0.95UCL
193	90	1247	992	Inf

Grafico KM Geral

plot(KM, ylab = "S(t)", xlab = "Dias", main = "Curva de Sobrevivência")



Resposta: O tempo mediano de sobrevivência na coorte foi igual a 1.247 [992, inf] dias (o limite superior não foi alcançado). Para os homens foi igual a 1.116 [887,1563] dias. Não foi possível calcular o tempo mediano para as mulheres porque mais de 50% delas permaneciam vivas ao final do estudo.

- b. Existe diferença entre homens e mulheres quanto ao tempo de sobrevivência pós-diagnóstico de Aids?

Resposta:

Análise de sobrevivência por Kaplan-Meier estratificada por sexo

```
KMsexo <- survfit(Surv(tempo, status) ~ sexo, data = ipec)
```

```
KMsexo
```

```
Call: survfit(formula = Surv(tempo, status) ~ sexo, data = ipec)
```

```

      n events median 0.95LCL 0.95UCL
sexo=F  49     16     Inf    1371     Inf
sexo=M 144     74    1116     887    1563

```

Gráfico da sobrevivência por sexo com intervalo de confiança (conf.int=T)

```
plot(KMsexo, lty = 1:2, col = 1:2, ylab = "S(t)", xlab = "Dias", conf.int=T)
```

```
legend(0, 0.4, c("Fem", "Masc"), lty = 1:2, col = 1:2)
```

```
title("Curvas de sobrevivência segundo sexo")
```

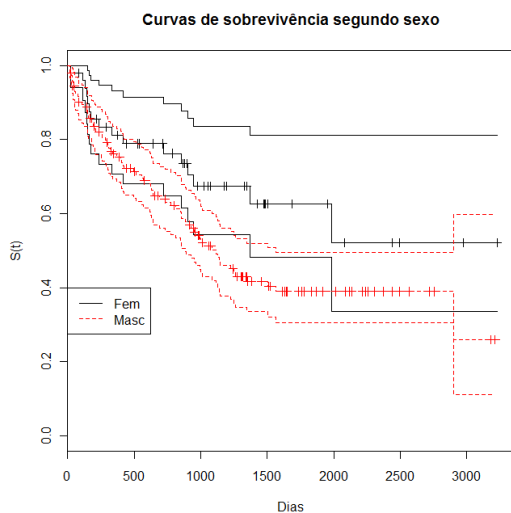
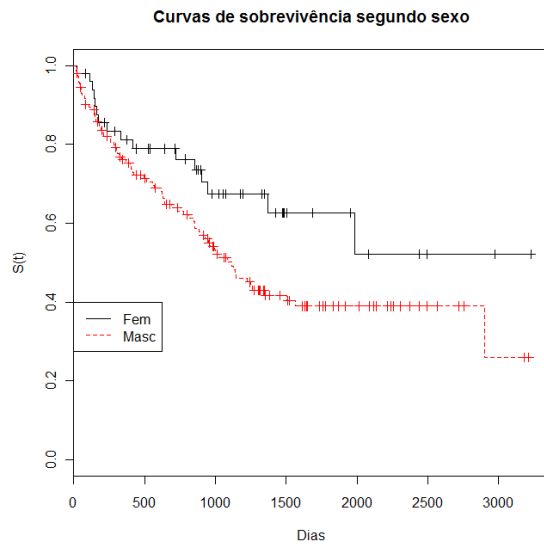


Gráfico da sobrevivência por sexo sem intervalo de confiança

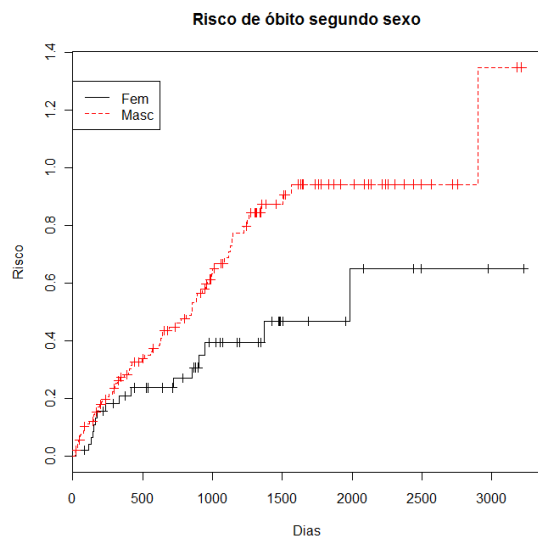
```
plot(KMsexo, lty = 1:2, col = 1:2, ylab = "S(t)", xlab = "Dias", conf.int = F)
```

```
legend(0, 0.4, c("Fem", "Masc"), lty = 1:2, col = 1:2)
```

```
title("Curvas de sobrevivência segundo sexo")
```



```
# Gráfico do risco de óbito por sexo sem intervalo de confiança
# Observe que agora é necessário indicar no comando plot a função de risco
# acumulado (cumhaz)
plot(KMsexo, lty = 1:2, fun = "cumhaz", col = 1:2, ylab = "Risco", xlab = "Dias",
conf.int = F)
legend(0, 1.3, c("Fem", "Masc"), lty = 1:2, col = 1:2)
title("Risco de óbito segundo sexo")
```



```
# Teste log-rank e peto para sexo
logrank <- survdiff(Surv(tempo, status) ~ sexo, data = ipec)
logrank
Call:
survdiff(formula = Surv(tempo, status) ~ sexo, data = ipec)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
sexo=F	49	16	24.5	2.93	4.03
sexo=M	144	74	65.5	1.09	4.03

Chisq= 4 on 1 degrees of freedom, p= 0.0447

```
peto <- survdiff(Surv(tempo, status) ~ sexo, data = ipec, rho = 1)
peto
Call:
```

```
survdif(formula = Surv(tempo, status) ~ sexo, data = ipec, rho = 1)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
sexo=F	49	12.1	18.2	2.011	3.54
sexo=M	144	55.1	49.0	0.746	3.54

```
Chisq= 3.5 on 1 degrees of freedom, p= 0.0598
```

No gráfico da sobrevivência por sexo observa-se que os homens sobreviveram menos. O mesmo se conclui se observarmos o gráfico da função de risco por sexo. Os resultados do teste log-rank, apesar de limítrofes ($p=0.047$), mostram uma diferença na sobrevivência após o diagnóstico de Aids entre homens e mulheres. Já o teste Peto, que dá maior peso as informações iniciais do estudo, não rejeita a hipótese nula de igualdade entre os gêneros quanto ao tempo de sobrevivência.

- c. Compare homens e mulheres segundo o tempo de sobrevivência mediano e percentis 75 e 90.

Resposta:

A saída padrão do R apresenta apenas a mediana com IC:

KM

```
Call: survfit(formula = Surv(tempo, status) ~ 1, data = ipec)
```

records	n.max	n.start	events	median	0.95LCL	0.95UCL
193	193	193	90	1247	992	NA

KMsexo

```
Call: survfit(formula = Surv(tempo, status) ~ sexo, data = ipec)
```

	n	events	median	0.95LCL	0.95UCL
sexo=F	49	16	Inf	1371	Inf
sexo=M	144	74	1116	887	1563

O comando `quantile` não trata as censuras se usado diretamente, podemos observar nos valores obtidos para a mediana, bastante diferentes (1247 x 852 dias):

Percentis sem censura

```
quantile(ipec$tempo)
```

0%	25%	50%	75%	100%
16	290	852	1348	3228

Percentis 50, 75 e 90 sem considerar a censura (não sofrem o evento)

```
quantile(ipec$tempo, c(0.5,0.75,0.9))
```

50%	75%	90%
852.0	1348.0	2088.8

Percentis 10, 25 e 50 sem considerar a censura (sofrem o evento)

```
quantile(ipec$tempo, c(0.1,0.25,0.5))
```

10%	25%	50%
84.2	290.0	852.0

Percentis 50, 75 e 90 para indivíduos que foram a óbito

```
quantile(ipec$tempo[status==1], c(0.5,0.75,0.9))
```

50%	75%	90%
413.00	890.75	1153.10

Percentis 50, 75 e 90 para indivíduos que foram censurados

```
quantile(ipec$tempo[status==0], c(0.5,0.75,0.9))
```

50%	75%	90%
-----	-----	-----

```
1242.0 1768.5 2423.8
```

Então, o que precisamos é aplicar *quantile* para os dados de sobrevida:

```
KM <- survfit(Surv(tempo, status) ~ 1, data=ipec)
```

```
# Percentis 90, 75 e 50 (tempo em dias nos quais 10%, 25% e 50%,  
respectivamente, dos pacientes sofreram o evento)  
quantile(KM, =c(0.1,.25,.5))
```

```
$quantile  
 10  25  50  
134 419 1247
```

```
$lower  
 10  25  50  
80 290 992
```

```
$upper  
 10  25  50  
183 688 NA
```

```
KMsexo <- survfit(Surv(tempo, status) ~ sexo, data=ipec)
```

```
# Percentis 90, 75 e 50 (tempo em dias nos quais 10%, 25% e 50%  
respectivamente, das mulheres – e dos homens - sofreram o evento)  
quantile(KMsexo, probs=c(0.1,.25,.5))
```

```
$quantile  
      10  25  50  
sexo=F 151 854 NA  
sexo=M 108 397 1116
```

```
$lower  
      10  25  50  
sexo=F 134 235 1371  
sexo=M  52 259  887
```

```
$upper  
      10  25  50  
sexo=F 721 NA  NA  
sexo=M 194 631 1563
```

Agora podemos observar que, tanto no tempo de sobrevida mediano quanto nos percentis 75 e 90, os homens dessa coorte apresentaram valores menores que as mulheres, o que significa que sobrevivem por menos tempo. As mulheres sequer alcançaram o tempo mediano e o limite inferior do percentil 75.

- d. Existe diferença na sobrevivência dos pacientes submetidos aos diferentes tipos de tratamento (variável *tratam*)?

Resposta:

```
# Análise de sobrevivência por Kaplan-Meier estratificada por tratamento:
```

```
KMtrat <- survfit(Surv(tempo, status) ~ tratam, data = ipec)
```

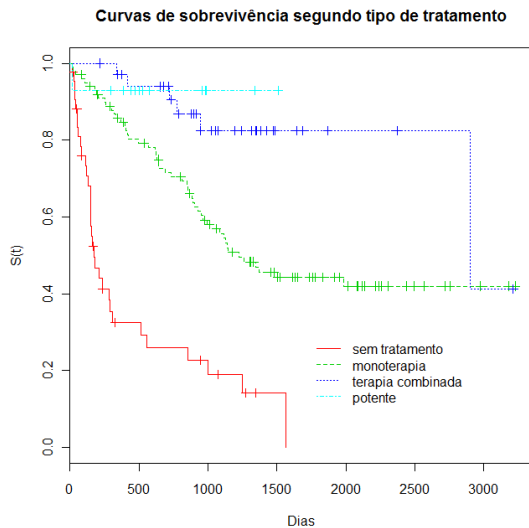
```
ipec$tratam <- factor(ipec$tratam, labels =c("Nenhum", "Mono", "Combinada", "Potente"))
```

```
KMtrat
```

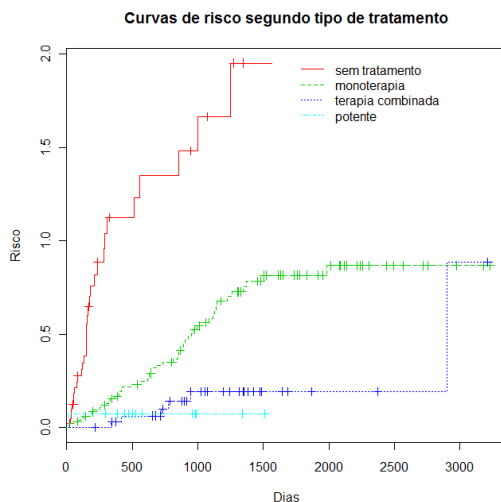
```
Call: survfit(formula = Surv(tempo, status) ~ tratam, data = ipec)
```


	records	n.max	n.start	events	median	0.95LCL	0.95UCL
tratam=Nenhum	44	44	44	32	173	151	516
tratam=Mono	100	100	100	51	1226	967	NA
tratam=Combinada	35	35	35	6	2898	2898	NA
tratam=Potente	14	14	14	1	NA	NA	NA

```
#Gráfico da sobrevivência por tratamento sem intervalo de confiança:
plot(KMtrat, lty = 1:4, col = 2:5, ylab = "S(t)", xlab = "Dias", conf.int = F)
legend(1700, 0.3, c("sem tratamento", "monoterapia", "terapia combinada",
"potente"), lty = 1:4, col = 2:5, bty = "n")
title("Curvas de sobrevivência segundo tipo de tratamento")
```



```
# Gráfico do risco de óbito por tratamento sem intervalo de confiança:
plot(KMtrat, lty = 1:4, fun = "cumhaz", col = 2:5, ylab = "Risco", xlab = "Dias", conf.int = F)
legend(1700, 2, c("sem tratamento", "monoterapia", "terapia combinada", "potente"),
lty = 1:4, col = 2:5, bty = "n")
title("Curvas de risco segundo tipo de tratamento")
```



```
# Testes log-rank e peto para tratamento:
Logrank.tto <- survdiff(Surv(tempo, status) ~ tratam, data = ipec)
Peto.tto <- survdiff(Surv(tempo, status) ~ tratam, data = ipec, rho=1)
Logrank.tto
```

Call:

```
survdif(formula = Surv(tempo, status) ~ tratam, data = ipec)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
tratam=0	44	32	9.51	53.229	61.560
tratam=1	100	51	54.73	0.254	0.659
tratam=2	35	6	20.27	10.050	13.049
tratam=3	14	1	5.49	3.674	3.950

```
Chisq= 69 on 3 degrees of freedom, p= 6.88e-15
```

Peto.tto

Call:

```
survdif(formula = Surv(tempo, status) ~ tratam, data = ipec,  
rho = 1)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
tratam=Nenhum	44	27.03	7.81	47.230	65.78
tratam=Mono	100	35.30	40.06	0.565	1.80
tratam=Combinada	35	3.89	14.98	8.206	13.70
tratam=Potente	14	1.00	4.37	2.597	3.43

```
Chisq= 72.2 on 3 degrees of freedom, p= 1.44e-15
```

No gráfico da função de sobrevivência por tratamento se observa que os pacientes que não alcançaram nenhum tratamento sobreviveram menos, os que receberam tratamento potente sobreviveram mais. O mesmo se observa no gráfico da função de risco por tratamento. Esta diferença é altamente significativa, tanto no teste log-rank ($p < 0,00001$), quanto no teste Peto ($p < 0,00001$). Portanto, ambos os testes rejeitaram a hipótese nula de igualdade no tempo de sobrevivência entre os diferentes tratamentos.

- e. Existe diferença na sobrevivência de pacientes com diferentes graus de escolaridade?

Resposta:

Análise de sobrevivência por Kaplan-Meier estratificada por escolaridade:

```
KMesc <- survfit(Surv(tempo, status) ~ escola, data = ipec)
```

```
KMesc
```

```
Call: survfit(formula = Surv(tempo, status) ~ escola, data = ipec)
```

```
11 observations deleted due to missingness
```

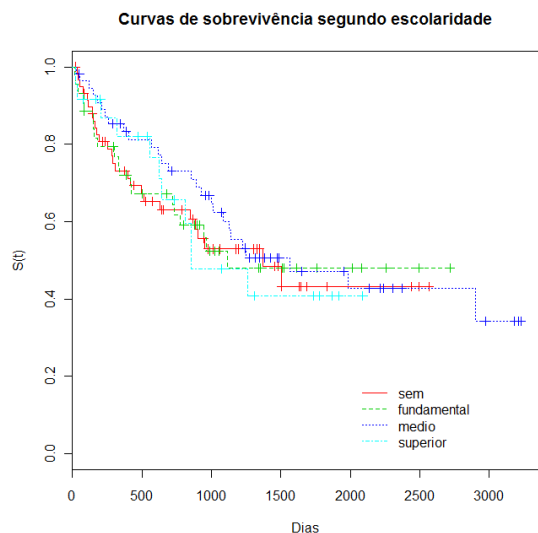
	records	n.max	n.start	events	median	0.95LCL	0.95UCL
escola=sem	59	59	59	26	1371	850	NA
escola=fundam	44	44	44	20	1116	733	NA
escola=medio	55	55	55	27	1563	1085	NA
escola=sup	24	24	24	11	855	644	NA

Gráfico da sobrevivência por escolaridade sem intervalo de confiança:

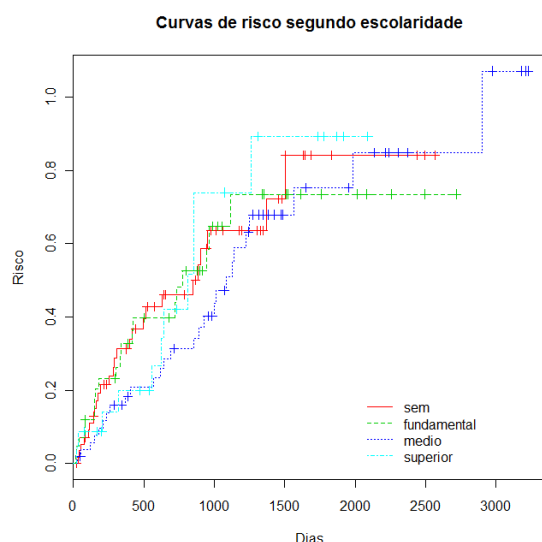
```
plot(KMesc, lty = 1:4, col = 2:5, ylab = "S(t)", xlab = "Dias", conf.int = F)
```

```
legend(2000, 0.2, c("sem", "fundamental", "medio", "superior"), lty = 1:4, col = 2:5,  
bty = "n")
```

```
title("Curvas de sobrevivência segundo escolaridade")
```



```
# Gráfico do risco de óbito por escolaridade sem intervalo de confiança:
plot(KMesc, lty = 1:4, fun = "cumhaz", col = 2:5, ylab = "Risco", xlab = "Dias", conf.int = F)
legend(2000, 0.2, c("sem", "fundamental", "medio", "superior"), lty = 1:4, col = 2:5, bty = "n")
title("Curvas de risco segundo escolaridade")
```



```
# Testes log-rank e peto para escolaridade:
Logrank.escol<- survdiff(Surv(tempo, status) ~ escola, data = ipec)
Peto.escol <- survdiff(Surv(tempo, status) ~ escola, data = ipec, rho=1)
Logrank.escol
```

```
Call:
survdiff(formula = Surv(tempo, status) ~ escola, data = ipec)
```

```
n=182, 11 observations deleted due to missingness.
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
escola=0	59	26	24.2	0.1348	0.1910
escola=1	44	20	19.2	0.0376	0.0490
escola=2	55	27	30.4	0.3785	0.6111
escola=3	24	11	10.3	0.0529	0.0605

```
Chisq= 0.6 on 3 degrees of freedom, p= 0.891
```

Peto.escol

Call:

```
survdif(formula = Surv(tempo, status) ~ escola, data = ipec,  
        rho = 1)
```

n=182, 11 observations deleted due to missingness.

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
escola=0	59	20.50	18.55	0.2048	0.3683
escola=1	44	15.96	14.56	0.1348	0.2228
escola=2	55	18.42	22.10	0.6112	1.2316
escola=3	24	8.16	7.84	0.0135	0.0195

Chisq= 1.3 on 3 degrees of freedom, p= 0.738

No gráfico da função de sobrevivência por grau de escolaridade observa-se que inicialmente os pacientes com escolaridade mais elevada (médio e superior) sobreviveram mais do que os de baixa escolaridade, mas que ao final tornam-se praticamente indistintas. Na curva de risco de óbito se observa que o risco dos mais escolarizados é menor inicialmente, mas ao final são todos similares. No entanto, esta diferença não foi significativa, nem no teste log-rank ($p>0,10$), nem no teste Peto ($p>0,10$). Portanto, não existe diferença na sobrevivência de Aids para os diferentes níveis de escolaridade.

- f. Faça os gráficos de sobrevivência e risco e aplique os testes de log-rank e Peto.

Resposta:

Dos gráficos e testes realizados nos itens anteriores podemos destacar os seguintes pontos:

Os homens apresentaram sobrevivência significativamente menor do que as mulheres ($p<0,05$). No entanto, se tivermos considerando mais relevantes os tempos de sobrevivência menores (início da curva), e usando o teste Peto, a diferença observada entre as curvas de sobrevivências dos dois sexos não pode ser considerada significativa.

A diferença na escolaridade dos pacientes não interferiu significativamente no tempo de sobrevivência dos pacientes desta coorte.

O tratamento ARV inicial submetido ao paciente é, sem dúvida, significativo no tempo de sobrevivência dos pacientes. Os pacientes que não tiveram acesso a nenhuma terapia tiveram uma sobrevida muito inferior aos que tiveram acesso aos tratamentos mais avançados. Os tempos medianos de sobrevivência calculados indicam claramente este fato, como pode ser também observado diretamente no gráfico. As medianas variaram de 173 a 2898 dias para pacientes sem tratamento e com terapia combinada, e sequer foi alcançada pelos pacientes que iniciaram o tratamento com terapia potente (isto é, mais da metade dos 14 pacientes que usaram esta terapia permaneciam vivos ao fim do estudo).