

Análise de Sobrevida
Teoria e Aplicações em Saúde

Caderno de Respostas

Marilia Sá Carvalho
Valeska Lima Andreozzi
Claudia Torres Codeço
Maria Tereza Serrano Barbosa
Silvia Emiko Shimakura

4

Estimação não-paramétrica

Exercícios

Exercício 4.1: Retorne aos dados de aleitamento do exercício 2.1. Construa, à mão, a tabela Kaplan-Meier e a curva Kaplan-Meier correspondente aos dados:

t_i	$R(t)$	$\Delta N(t)$	$\hat{S}_{KM}(t) = \prod_{t_i \leq t} \frac{R(t_i) - \Delta N(t_i)}{R(t_i)}$
1	15	3	$\frac{(15 - 3)}{15} = 0,8000$
2	12	2	$0,8000 \times \frac{(12 - 2)}{12} = 0,6667$
3	10	1	$0,6667 \times \frac{(10 - 1)}{10} = 0,6000$
5	9	3	$0,6000 \times \frac{(9 - 3)}{9} = 0,4000$
6	6	2	$0,4000 \times \frac{(6 - 2)}{6} = 0,2667$
8	4	2	$0,2667 \times \frac{(4 - 2)}{4} = 0,1333$
10	2	1	$0,1333 \times \frac{(2 - 1)}{2} = 0,0667$
12	1	1	$0,0667 \times \frac{(1 - 1)}{1} = 0,0000$

O gráfico do exercício a seguir corresponde a esta tabela.

1. Com base na tabela que você criou, responda: qual a probabilidade de uma criança ser amamentada pelo menos até o sexto mês de vida? Qual a probabilidade de ser amamentada por mais de 3 meses? Qual é a probabilidade de ser

amamentada por mais de 10 meses? Qual foi o tempo mediano de aleitamento?

Resposta:

- a probabilidade de uma criança ser amamentada pelo menos até o sexto mês de vida é $S(6) = 0,2667$
 - a probabilidade de ser amamentada por mais de 3 meses é $S(3) = 0,6$
 - a probabilidade de ser amamentada por mais de 10 meses é $S(10) = 0,0667$
 - o tempo mediano de aleitamento está entre 3 e 5 meses.
2. Construa a tabela Kaplan-Meier e a curva de sobrevida utilizando o pacote estatístico R. Os dados de aleitamento estão no arquivo *leite.txt*.

Lendo o banco de dados

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

Criando o objeto que contém os dados de sobrevida

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

Calculando a sobrevida por KM

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

Call: survfit(formula = y ~ 1, data = leite)

n	events	median	0.95LCL	0.95UCL
15	15	5	2	8

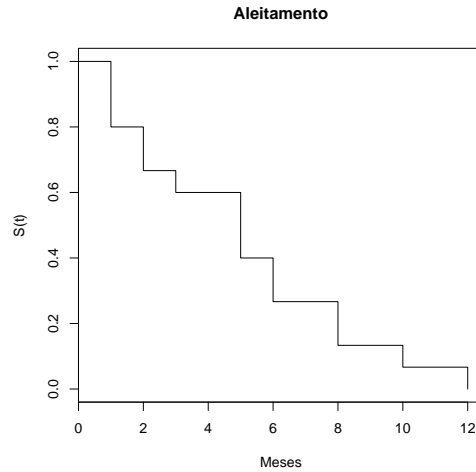
```
> summary(km)
```

Call: survfit(formula = y ~ 1, data = leite)

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	

Fazendo o gráfico da função de sobrevivência

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



Exercício 4.2: 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. A variável *grupo* indica a comunidade à qual cada criança pertence.

Lendo o bando 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
4	4	3	1	1

5	5	5	1	1
6	6	1	1	1
7	7	6	1	1
8	8	8	1	1
9	9	1	1	1
10	10	5	1	1
11	11	2	1	1
12	12	2	1	1
13	13	5	1	1
14	14	8	1	1
15	15	1	1	1
16	16	5	1	2
17	17	12	1	2
18	18	10	1	2
19	19	4	1	2
20	20	4	1	2
21	21	3	1	2
22	22	6	1	2
23	23	9	1	2
24	24	2	1	2
25	25	6	1	2
26	26	4	1	2
27	27	1	1	2
28	28	7	1	2
29	29	10	1	2
30	30	1	1	2
31	31	13	1	3
32	32	14	1	3
33	33	20	1	3
34	34	3	1	3
35	35	5	1	3
36	36	1	1	3
37	37	8	1	3
38	38	15	1	3
39	39	2	1	3
40	40	5	1	3
41	41	3	1	3
42	42	2	1	3
43	43	6	1	3
44	44	15	1	3
45	45	1	1	3
46	46	1	1	4
47	47	16	1	4
48	48	20	1	4
49	49	1	1	4
50	50	1	1	4
51	51	1	1	4

```

52    52    1    1    4
53    53    2    1    4
54    54    2    1    4
55    55   13    1    4
56    56    3    1    4
57    57    1    1    4
58    58    1    1    4
59    59   14    1    4
60    60    2    1    4

```

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

```

> y <- Surv(leite2$tempo, leite2$status)
> km <- survfit(y ~ grupo, data = leite2)
> km

```

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

	n	events	median	0.95LCL	0.95UCL
grupo=1	15	15	5	2	8
grupo=2	15	15	5	4	10
grupo=3	15	15	5	3	15
grupo=4	15	15	2	1	14

```
> summary(km)
```

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

grupo=1							
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	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	

grupo=2							
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	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	NA	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	NA	NA	NA

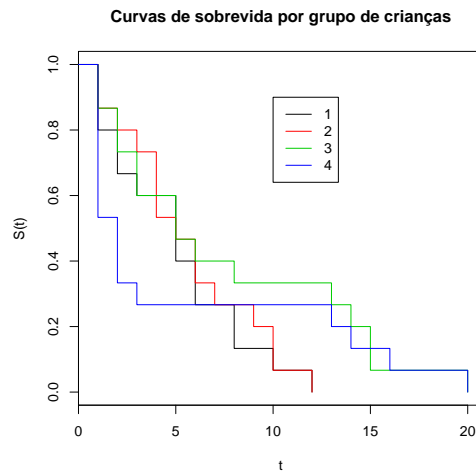
grupo=4

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	NA	NA	NA

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

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

```
> plot(km, 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 sobrevida por grupo de crianças")
```



Resposta: No gráfico de sobrevida pode-se ver que a comunidade 4 possui uma probabilidade de amamentar menor que a comunidade 1 até os 8 meses (aproximadamente). O comportamento da comunidade 2 é semelhante ao da comunidade 1 e a comunidade 3 possui um probabilidade maior de amamentar ao longo de todo o tempo de estudo.

3. Calcule o teste log-rank e de Peto para as quatro curvas.

```
> survdiff(y ~ grupo, data = leite2)
```

Call:

```
survdiff(formula = y ~ 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

```
> survdiff(y ~ grupo, data = leite2, rho = 1)
```

Call:

```
survdiff(formula = y ~ 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

```

A que conclusões você chega com esta análise? Existe diferença entre as comunidades quanto ao tempo de aleitamento?

Resposta: 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.

Exercício 4.3: O banco de dados *ipec.csv* contém os dados de uma coorte de 193 pacientes com Aids, da qual fazem parte os dados apresentados na Tabela 4.1. Para este estudo foi definido, como tempo de sobrevivência, o tempo entre o diagnóstico de Aids (critério CDC-1993) e o óbito.

1. Abra o banco de dados e observe as variáveis presentes. No Apêndice C1, você encontra um breve histórico da coorte e a descrição das variáveis.

```

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

```

```

[1] "id"      "ini"     "fim"     "tempo"   "status"  "sexo"   "escola"
[8] "idade"   "risco"   "acompan" "obito"   "anotrat" "tratam" "doenca"
[15] "propcp"

```

```

> attach(ipec)

```

```

The following object(s) are masked from ipec ( position 3 ) :

```

```

acompan anotrat doenca escola fim id idade ini obito propcp risco sexo status tempo

```

2. Faça uma análise exploratória dos dados no R. Como é esta coorte? Qual é a idade média? Qual é a razão de homens:mulheres? Quantos receberam tratamento? Quantos foram a óbito e quantos foram censurados? Faça um histograma dos tempos censurados: eles ocorreram no fim do estudo? Ou ocorreram durante todo o estudo? (Aproveite para praticar comandos no R para análise exploratória de dados).

Calculando a idade média:

```

> mean(idade)

```

```

[1] 36.55440

```

Número de homens e mulheres:

```
> table(sexo)
```

```
sexo
  F  M
49 144
```

A relação homem:mulher neste grupo de pacientes é igual a $144/49 = 2.93$, aproximadamente 3 homens para cada mulher.

Número de pacientes por grupo de tratamento:

```
> table(tratam)
```

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

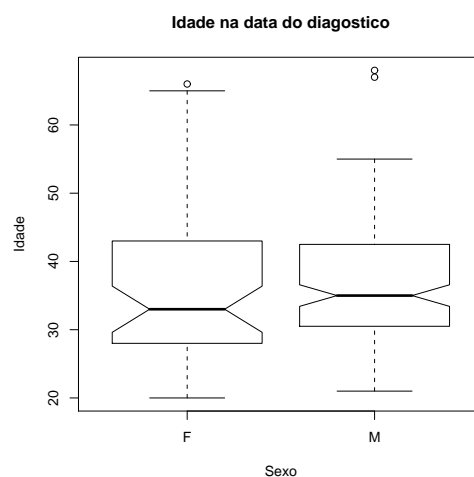
Número de óbitos e sobreviventes:

```
> table(status)
```

```
status
  0  1
103 90
```

Gráfico da distribuição da idade por sexo:

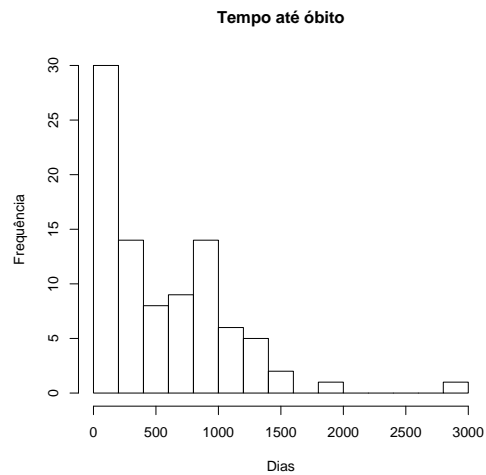
```
> boxplot(idade ~ sexo, main = "Idade na data do diagnóstico", notch = T,
+         ylab = "Idade", xlab = "Sexo")
```



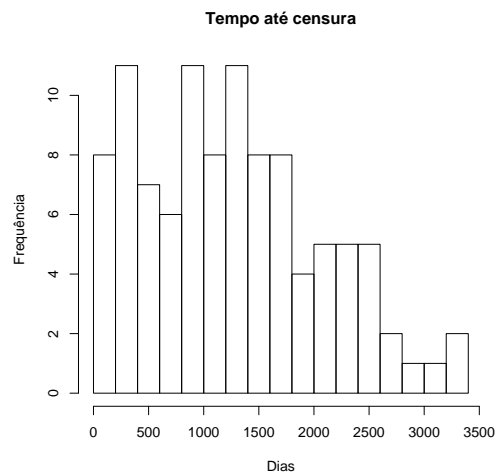
Não há diferença na idade por gênero.

Distribuição do tempo de sobrevivência dos pacientes que morreram e dos que sobreviveram até o fim do estudo

```
> hist(tempo[status == 1], breaks = 12, main = "Tempo até óbito",  
+       ylab = "Frequência", xlab = "Dias")
```



```
> hist(tempo[status == 0], breaks = 12, ylab = "Frequência", xlab = "Dias",  
+       main = "Tempo até censura")
```



3. Refaça no R a análise não paramétrica dos dados do Ipec apresentada no texto, estratificando o tempo de sobrevida pela variável sexo. Faça os gráficos das curvas de sobrevida e risco e calcule os testes log-rank e Peto (os comandos estão todos ao longo do texto). Existe diferença entre homens e mulheres quanto ao tempo de sobrevida pós-diagnóstico de Aids?

Lendo o banco de dados e calculando a sobrevida:

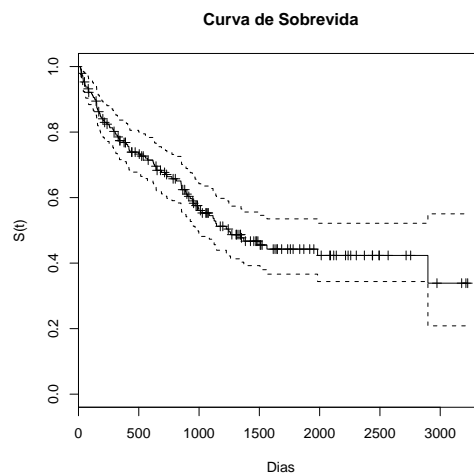
```
> require(survival)
> ipec <- read.table("ipec.csv", header = T, sep = ";")
> KM <- survfit(Surv(tempo, status) ~ 1, data = ipec)
> KM
```

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

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

Gráfico da sobrevida de Aids dos pacientes do IPEC com intervalo de confiança. Os tempos censurados estão assinalados no gráfico por um traço vertical sobre a curva.

```
> plot(KM, ylab = "S(t)", xlab = "Dias", main = "Curva de Sobrevida")
```



Calculando a sobrevida 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

Não foi possível calcular o tempo mediano para as mulheres porque menos de 50% delas haviam morridos ao fim do estudo.

Gráfico da sobrevida 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 sobrevida segundo sexo")
```

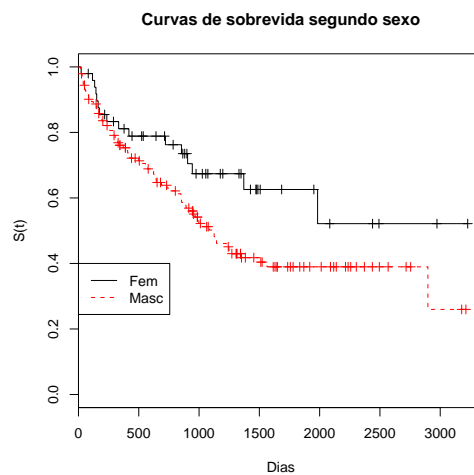
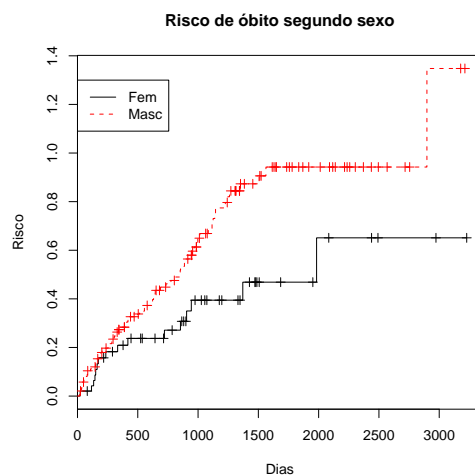


Gráfico do risco de óbito por sexo sem intervalo de confiança

```
> 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

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

Call:

```
survdiff(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

Resposta: Os resultados do teste log-rank, apesar de limítrofes ($p=0.047$), mostram uma diferença na sobrevivência pós diagnóstico de Aids entre homens e mulheres. Já o teste Peto, que dá maior peso às informações iniciais do estudo,

não rejeita a hipótese nula de igualdade entre os gêneros quanto ao tempo de sobrevida.

4. Refaça a análise acima, agora estratificando-a por tipo de tratamento. Existe diferença na sobrevivência dos pacientes submetidos aos diferentes tipos de tratamento (variável *tratam*)?

Calculando a sobrevida estratificada por tratamento:

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

Gráfico da sobrevida 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 sobrevida segundo tipo de tratamento")
```

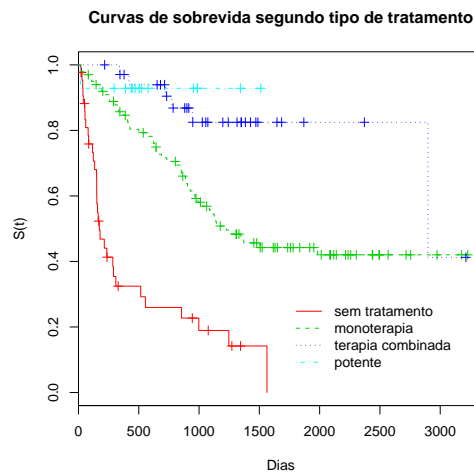
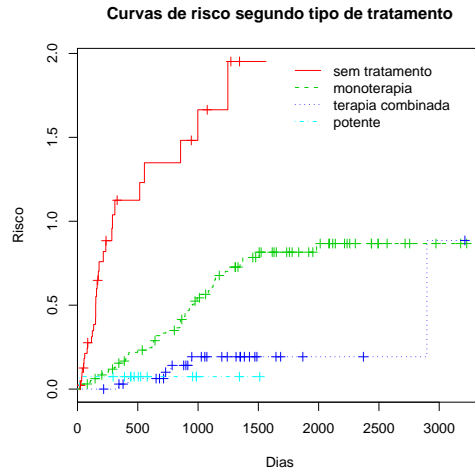


Gráfico do risco de óbito por tratamento sem intervalo de confiança:

```
> 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:

```
> survdiff(Surv(tempo, status) ~ tratam, data = ipec)
```

Call:

```
survdiff(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

```
> survdiff(Surv(tempo, status) ~ tratam, data = ipec, rho = 1)
```

Call:

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

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
tratam=0	44	27.03	7.81	47.230	65.78
tratam=1	100	35.30	40.06	0.565	1.80
tratam=2	35	3.89	14.98	8.206	13.70
tratam=3	14	1.00	4.37	2.597	3.43

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

Resposta: Ambos os testes rejeitaram a hipótese nula de igualdade no tempo de sobrevida entre os diferentes tratamentos.

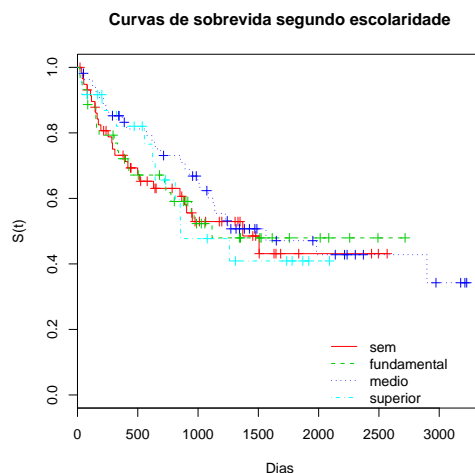
5. Estratifique os dados por nível de escolaridade. Existe diferença na sobrevivência de pacientes com diferentes graus de escolaridade?

Calculando a sobrevida estratificada por escolaridade:

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

Gráfico da sobrevida 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 sobrevida segundo escolaridade")
```



Testes log-rank e peto:

```
> survdiff(Surv(tempo, status) ~ escola, data = ipec)
```

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

```
> survdiff(Surv(tempo, status) ~ escola, data = ipec, rho = 1)
```

Call:

```
survdiff(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

Resposta: Não existe diferença na sobrevida de Aids para os diferentes níveis de escolaridade.