

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

# 7

## Modelos de regressão semiparamétricos

### Exercícios

**Exercício 7.1:** Os dados da coorte de transplante de medula óssea estão no arquivo *tmoclas.dat*. Abra o arquivo no R:

```
> tmo <- read.table("tmoclas.dat", header = T, sep = ",")
```

Nome das variáveis

```
> names(tmo)
```

```
[1] "id"      "sexo"    "idade"   "status"  "os"      "plaq"  
[7] "tempplaq" "deag"    "tempdeag" "decr"    "tempdecr" "fase"
```

Transformando as variáveis categóricas em factor

```
> tmo$sexo <- factor(tmo$sexo)  
> tmo$decr <- factor(tmo$decr)  
> tmo$deag <- factor(tmo$deag)  
> tmo$fase <- factor(tmo$fase)
```

1. Refaça os gráficos de Kaplan-Meier para as variáveis *sexo*, *deag*, *decr* e *fase* e verifique o pressuposto de proporcionalidade.

Este item tem o objetivo de familiarizar o leitor com os comandos do R. A interpretação dos resultados está discutida ao longo do texto. Ver seção 7.3.2.

Abrindo a biblioteca *survival* e criando o objeto *sobrevida*

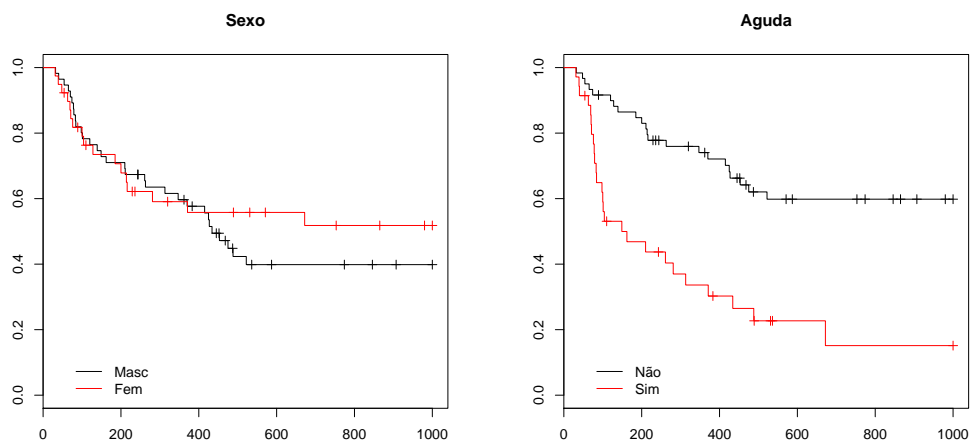
```
> require(survival)
> y <- Surv(tmo$os, tmo$status)
```

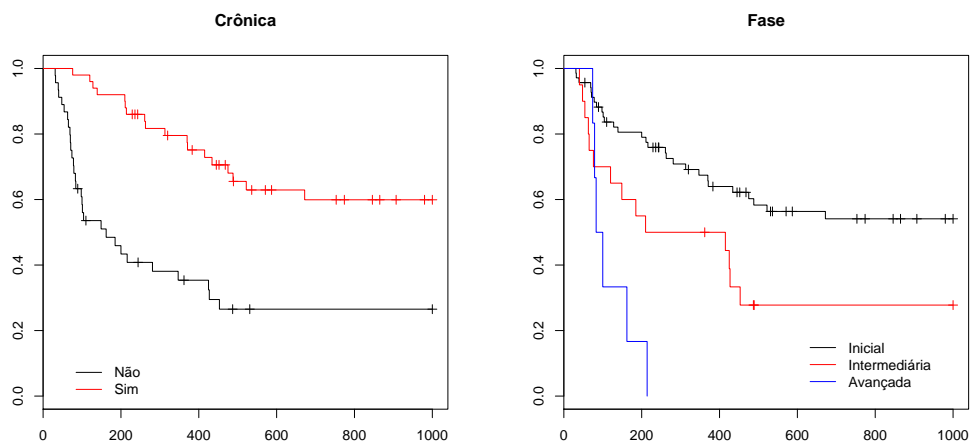
Estimando as curvas de sobrevivência pelo Kaplan-Meier por diversas variáveis

```
> KMsexo <- survfit(y ~ sexo, data = tmo)
> KMdeag <- survfit(y ~ deag, data = tmo)
> KMdecr <- survfit(y ~ decr, data = tmo)
> KMfase <- survfit(y ~ fase, data = tmo)
```

Gráfico das curvas de Kaplan-Meier

```
> par(mfrow = c(2, 2))
> plot(KMsexo, main = "Sexo", col = 1:2, legend.tex = c("Masc",
+ "Fem"))
> plot(KMdeag, main = "Aguda", col = 1:2, legend.tex = c("Não",
+ "Sim"))
> plot(KMdecr, main = "Crônica", col = 1:2, legend.tex = c("Não",
+ "Sim"))
> plot(KMfase, main = "Fase", col = c(1, 2, 4), legend.pos = c(600,
+ 0.2), legend.tex = c("Inicial", "Intermediária", "Avançada"))
```





2. Refaça os quatro modelos de Cox apresentados no texto e compare os modelos usando a análise de *deviance*. Veja os comentários da seção 7.4

Estimando os modelos de Cox

```
> mod1 <- coxph(y ~ idade + sexo, data = tmo)
> summary(mod1)
```

Call:

```
coxph(formula = y ~ idade + sexo, data = tmo)
```

n= 96

	coef	exp(coef)	se(coef)	z	p
idade	-0.0186	0.982	0.0141	-1.32	0.19
sexo2	-0.3299	0.719	0.3219	-1.02	0.31

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.982	1.02	0.955	1.01
sexo2	0.719	1.39	0.383	1.35

Rsquare= 0.022 (max possible= 0.984 )

Likelihood ratio test= 2.16 on 2 df, p=0.34

Wald test = 2.11 on 2 df, p=0.348

Score (logrank) test = 2.11 on 2 df, p=0.348

```
> mod2 <- coxph(y ~ idade + sexo + fase, data = tmo)
> summary(mod2)
```

Call:

```
coxph(formula = y ~ idade + sexo + fase, data = tmo)
```

```

n= 96
      coef exp(coef) se(coef)      z      p
idade -0.0255    0.975  0.0146 -1.752 8.0e-02
sexo2 -0.1641    0.849  0.3254 -0.504 6.1e-01
fase2  0.8932    2.443  0.3517  2.540 1.1e-02
fase3  1.9408    6.964  0.4884  3.974 7.1e-05

      exp(coef) exp(-coef) lower .95 upper .95
idade    0.975    1.026    0.947    1.00
sexo2    0.849    1.178    0.448    1.61
fase2    2.443    0.409    1.226    4.87
fase3    6.964    0.144    2.674   18.14

Rsquare= 0.165 (max possible= 0.984 )
Likelihood ratio test= 17.3 on 4 df, p=0.00169
Wald test          = 19.7 on 4 df, p=0.000577
Score (logrank) test = 23.5 on 4 df, p=0.000101

```

```

> mod3 <- coxph(y ~ idade + sexo + fase + deag, data = tmo)
> summary(mod3)

```

```

Call:
coxph(formula = y ~ idade + sexo + fase + deag, data = tmo)

```

```

n= 96
      coef exp(coef) se(coef)      z      p
idade -0.0163    0.984  0.0144 -1.134 0.26000
sexo2 -0.2053    0.814  0.3185 -0.645 0.52000
fase2  0.9815    2.668  0.3430  2.862 0.00420
fase3  1.5683    4.798  0.4978  3.150 0.00160
deag1  1.1848    3.270  0.3120  3.797 0.00015

      exp(coef) exp(-coef) lower .95 upper .95
idade    0.984    1.016    0.956    1.01
sexo2    0.814    1.228    0.436    1.52
fase2    2.668    0.375    1.362    5.23
fase3    4.798    0.208    1.809   12.73
deag1    3.270    0.306    1.774    6.03

Rsquare= 0.279 (max possible= 0.984 )
Likelihood ratio test= 31.4 on 5 df, p=7.77e-06
Wald test          = 32.6 on 5 df, p=4.43e-06
Score (logrank) test = 39.1 on 5 df, p=2.24e-07

```

```

> mod4 <- coxph(y ~ idade + sexo + fase + deag + decr, data = tmo)
> summary(mod4)

```

Call:  
 coxph(formula = y ~ idade + sexo + fase + deag + decr, data = tmo)

```

n= 96
      coef exp(coef) se(coef)      z      p
idade -0.00441    0.996  0.0149 -0.296 0.77000
sexo2 -0.22608    0.798  0.3329 -0.679 0.50000
fase2  0.64136    1.899  0.3765  1.703 0.08900
fase3  1.02796    2.795  0.5264  1.953 0.05100
deag1  1.25304    3.501  0.3307  3.789 0.00015
decr1 -0.97759    0.376  0.3404 -2.872 0.00410

```

```

      exp(coef) exp(-coef) lower .95 upper .95
idade    0.996      1.004    0.967    1.025
sexo2    0.798      1.254    0.415    1.532
fase2    1.899      0.527    0.908    3.972
fase3    2.795      0.358    0.996    7.844
deag1    3.501      0.286    1.831    6.693
decr1    0.376      2.658    0.193    0.733

```

```

Rsquare= 0.34 (max possible= 0.984 )
Likelihood ratio test= 39.9 on 6 df, p=4.8e-07
Wald test           = 37.8 on 6 df, p=1.21e-06
Score (logrank) test = 47.1 on 6 df, p=1.79e-08

```

### Análise de Deviance

```
> anova(mod1, mod2, mod3, mod4, test = "Chisq")
```

#### Analysis of Deviance Table

```

Model 1: y ~ idade + sexo
Model 2: y ~ idade + sexo + fase
Model 3: y ~ idade + sexo + fase + deag
Model 4: y ~ idade + sexo + fase + deag + decr
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1      94      395.93
2      92      380.78 2    15.14 0.0005146
3      91      366.67 1    14.11 0.0001726
4      90      358.20 1     8.47 0.0036015

```

3. Calcule o índice de prognóstico do modelo 4 para um indivíduo com 35 anos, do sexo feminino ( $sexo2 = 1$ ), na fase intermediária ( $fase2 = 1$ ,  $fase3 = 0$ ) da doença e que tenha tido doença do enxerto crônica, mas não aguda ( $decr = 1$ ,  $deag = 0$ ).

Colocando em um vetor as características do indivíduo na ordem em que aparecem no modelo (*idade*, *sexo*, *fase*, *deag*, *decr*)

```
> individuo <- c(35, 1, 1, 0, 0, 1)
> individuo
```

```
[1] 35 1 1 0 0 1
```

Colocando em outro vetor os coeficientes do modelo

```
> coeficientes <- mod4$coef
> coeficientes
```

```
      idade      sexo2      fase2      fase3      deag1      decr1
-0.004413461 -0.226083177  0.641355011  1.027956377  1.253039591 -0.977592832
```

Para calcular o índice de prognóstico basta multiplicar os valores das variáveis do paciente (vetor *individuo*) pelos coeficientes do modelo (vetor *coeficientes*) e somar

```
> ip <- sum(coeficientes * individuo)
> ip
```

```
[1] -0.7167921
```

Temos então que o índice de prognóstico do indivíduo com as características citadas anteriormente é igual a -0.717.

**Exercício 7.2:** Três modelos causais aninhados são propostos para explicar a sobrevida de pacientes em diálise. O primeiro considera apenas a variável idade, o segundo inclui as doenças de base e o terceiro inclui uma variável ambiental (tamanho da unidade de tratamento).

**Modelo I:** sobrevida = idade

**Modelo II:** sobrevida = idade + cdiab + congenita + crim

**Modelo III:** sobrevida = idade + cdiab + congenita + crim + grande

1. Utilizando o banco de dados *dialise.csv*, faça gráficos de Kaplan-Meier estratificados por cada covariável categórica e verifique (visualmente) se há fortes indícios de não proporcionalidade (o que inviabilizaria o ajuste do modelo de Cox tradicional).

```

> dialise <- read.table("dialise.csv", header = T, sep = ",")
> dialise$grande <- factor(dialise$grande)
> dialise$cdiab <- factor(dialise$cdiab)
> dialise$crim <- factor(dialise$crim)
> dialise$congenita <- factor(dialise$congenita)
> names(dialise)

[1] "unidade"  "idade"    "inicio"   "fim"      "status"   "tempo"
[7] "grande"   "causa"    "cdiab"    "crim"     "congenita"

> y <- Surv(dialise$tempo, dialise$status)

```

Estimando as curvas de Kaplan-Meier

```

> KMdiab <- survfit(y ~ cdiab, data = dialise)
> KMrim <- survfit(y ~ crim, data = dialise)
> KMcong <- survfit(y ~ congenita, data = dialise)
> KMgrande <- survfit(y ~ grande, data = dialise)

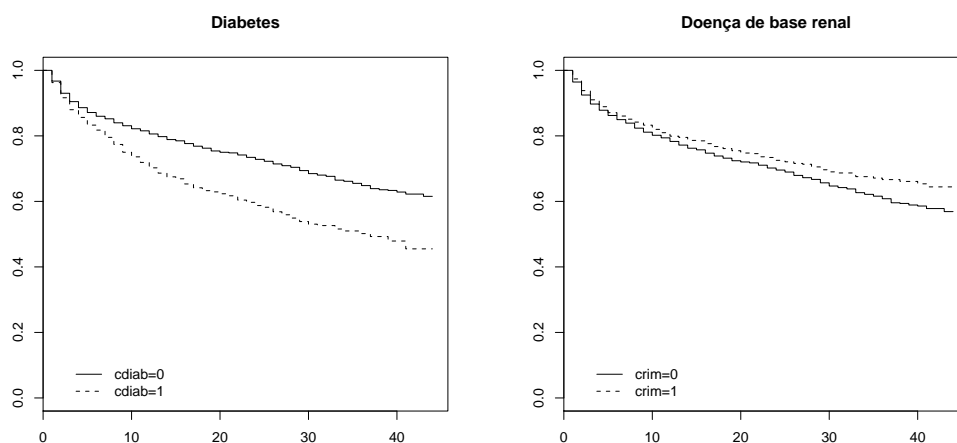
```

Gráficos das curvas KM

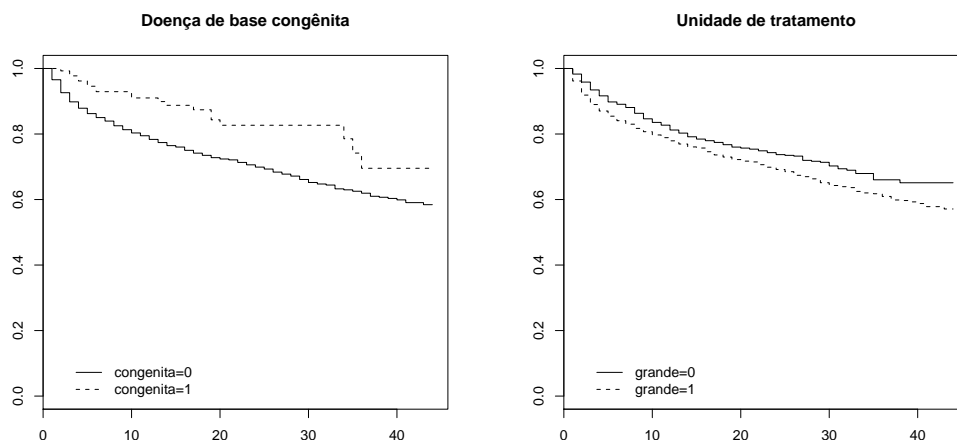
```

> par(mfrow = c(2, 2))
> plot(KMdiab, lty = c(1, 2), main = "Diabetes", mark.time = F,
+      legend.text = names(KMdiab$strata))
> plot(KMrim, lty = c(1, 2), main = "Doença de base renal", mark.time = F,
+      legend.text = names(KMrim$strata))
> plot(KMcong, lty = c(1, 2), main = "Doença de base congênita",
+      mark.time = F, legend.text = names(KMcong$strata))
> plot(KMgrande, lty = c(1, 2), main = "Unidade de tratamento",
+      mark.time = F, legend.text = names(KMgrande$strata))

```







**Resposta:** Podemos dizer que não há evidência de forte desvio do pressuposto de proporcionalidade. Apenas a variável congênita apresenta um padrão mais complexo, mas note que a tendência de queda é a mesma, em média, e os degraus observados na curva superior se deve, em grande parte, ao pequeno número de amostras. Em uma situação como esta, é interessante modelar com e sem esta variável, para verificar o efeito de sua inclusão/exclusão na estimativa das outras.

2. Ajuste cada modelo causal acima utilizando o modelo de riscos proporcionais de Cox, tomando o cuidado de interpretar os parâmetros a cada saída. (Lembre de acrescentar o argumento  $x = T$ , na especificação da função `coxph()`, para que depois se possam obter os índices de prognóstico.)

```
> modeloI <- coxph(y ~ idade, data = dialise, x = T)
> summary(modeloI)
```

Call:

```
coxph(formula = y ~ idade, data = dialise, x = T)
```

```
n= 6805
      coef exp(coef) se(coef)      z p
idade 0.0349      1.04 0.00173 20.2 0

      exp(coef) exp(-coef) lower .95 upper .95
idade      1.04      0.966      1.03      1.04
```

```
Rsquare= 0.062 (max possible= 0.98 )
Likelihood ratio test= 435 on 1 df, p=0
```

Wald test = 409 on 1 df, p=0  
Score (logrank) test = 415 on 1 df, p=0

```
> modeloII <- coxph(y ~ idade + cdiab + crim + congenita, data = dialise,  
+ x = T)  
> summary(modeloII)
```

Call:

```
coxph(formula = y ~ idade + cdiab + crim + congenita, data = dialise,  
x = T)
```

n= 6805

	coef	exp(coef)	se(coef)	z	p
idade	0.0344	1.035	0.00175	19.677	0.0e+00
cdiab1	0.2856	1.331	0.06031	4.736	2.2e-06
crim1	0.0303	1.031	0.06698	0.453	6.5e-01
congenita1	-0.7152	0.489	0.22618	-3.162	1.6e-03

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	1.035	0.966	1.032	1.039
cdiab1	1.331	0.752	1.182	1.498
crim1	1.031	0.970	0.904	1.175
congenita1	0.489	2.045	0.314	0.762

Rsquare= 0.067 (max possible= 0.98 )

Likelihood ratio test= 473 on 4 df, p=0  
Wald test = 438 on 4 df, p=0  
Score (logrank) test = 451 on 4 df, p=0

```
> modeloIII <- coxph(y ~ idade + cdiab + crim + congenita + grande,  
+ data = dialise, x = T)  
> summary(modeloIII)
```

Call:

```
coxph(formula = y ~ idade + cdiab + crim + congenita + grande,  
data = dialise, x = T)
```

n= 6805

	coef	exp(coef)	se(coef)	z	p
idade	0.0340	1.035	0.00175	19.386	0.0e+00
cdiab1	0.2955	1.344	0.06040	4.892	1.0e-06
crim1	0.0191	1.019	0.06708	0.285	7.8e-01
congenita1	-0.7274	0.483	0.22622	-3.216	1.3e-03
grande1	0.1770	1.194	0.06339	2.792	5.2e-03

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	1.035	0.967	1.031	1.038

cdiab1	1.344	0.744	1.194	1.513
crim1	1.019	0.981	0.894	1.163
congenita1	0.483	2.070	0.310	0.753
grande1	1.194	0.838	1.054	1.352

Rsquare= 0.068 (max possible= 0.98 )  
 Likelihood ratio test= 481 on 5 df, p=0  
 Wald test = 447 on 5 df, p=0  
 Score (logrank) test = 461 on 5 df, p=0

**Resposta:** Ajustando o Modelo I, encontramos que a idade apresenta-se como um fator de risco de 1.04. Isto é, cada ano de idade a mais na data de início da diálise implica em um risco 4% maior de óbito. No modelo II, a idade continua significativa, embora tenha perdido um pouco do efeito. A doença de base diabetes se mostrou um importante fator de risco, com um sobrerisco de 33%. A causa renal também se mostrou um fator de risco, porém seu efeito não foi significativo. A doença de base congênita se mostrou um fator protetor: pessoas sem doença congênita têm risco duas vezes maior de ir a óbito do que as com causa congênita. Este efeito protetor pode ser interpretado como um efeito indireto. Isto é, não é a causa congênita que protege, mas sim a ausência da diabetes ou da causa renal. Outra questão que deve ser levada em consideração é a prevalência baixa de pessoas com doená congêntas neste banco de dados que são somente 142 do total de 6805 pacientes. No modelo III, a inclusão da variável grande não alterou significativamente o efeito das outras variáveis. O tamanho da unidade se mostrou importante, com pacientes atendidos em unidades grandes tendo risco 19% maior de ir a óbito do que aqueles atendidos em unidades menores, com menos de 5 salas.

3. Compare os modelos usando a análise de *deviance* e escolha o modelo com melhor ajuste.

```
> anova(modeloI, modeloII, modeloIII, test = "Chisq")
```

Analysis of Deviance Table

```

Model 1: y ~ idade
Model 2: y ~ idade + cdiab + crim + congenita
Model 3: y ~ idade + cdiab + crim + congenita + grande
  Resid. Df Resid. Dev   Df Deviance P(>|Chi|)
1     6804    26073.1
2     6801    26034.9    3     38.2 2.507e-08
3     6800    26026.8    1      8.1 4.512e-03

```

**Resposta:** A inclusão das variáveis de doença de base (Modelo II) melhorou significativamente o ajuste do modelo, quando comparado com o modelo contendo apenas a idade (Deviance = 38.2,  $p < 0.0001$ ). A inclusão da variável tamanho da unidade (**grande**) melhorou ainda mais o ajuste (comparando os modelos II e III, temos Deviance = 8.2,  $p < 0.001$ ). Concluimos, então, que o Modelo III é o melhor modelo.

4. Qual o poder explicativo do modelo escolhido? Calcule a razão entre o  $R^2$  do modelo escolhido e o  $R^2$  máximo (ambos estão presentes na saída do comando `summary()`).

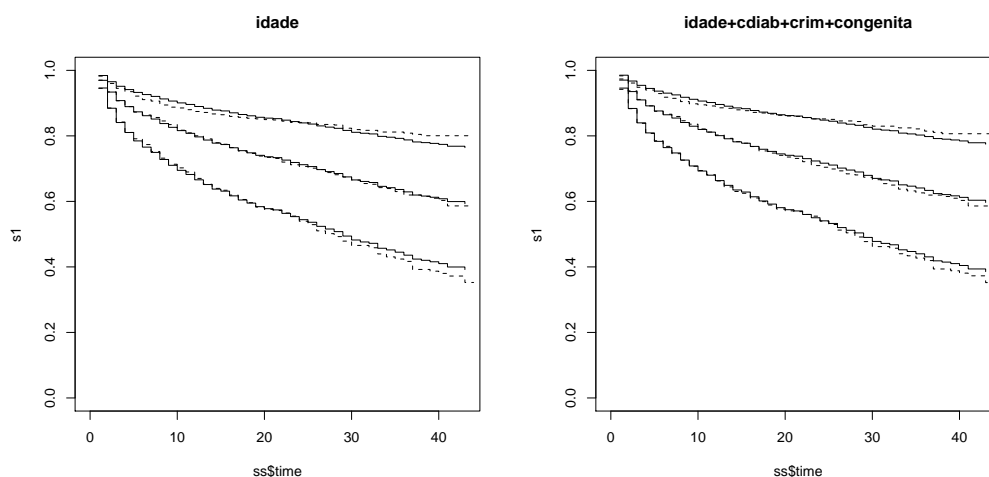
**Resposta:** O modelo III explicou  $0.068/0.98 \times 100 = 6.9\%$  da variância dos dados.

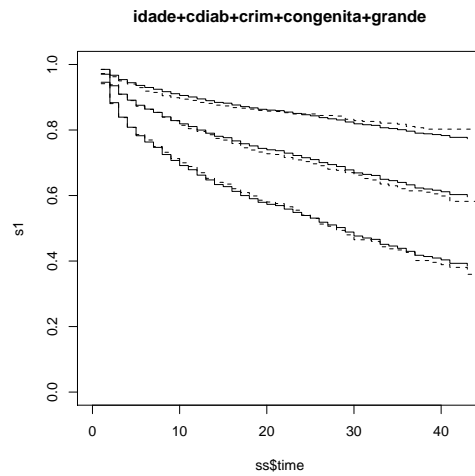
5. Faça o gráfico dos índices de prognóstico do modelo escolhido, utilizando a função `plot.pi(modelo escolhido)`. Avalie visualmente o ajuste do modelo à curva de Kaplan-Meier. A função `plot.pi()` não existe na biblioteca `survival` do R. Antes de utilizar a função `plot.pi()` precisamos executar o comando `source("Rfun.r")` para criá-la.

```
> source("Rfun.r")

> par(mfrow = c(2, 2))
> plot.pi(modeloI, main = "idade")
> plot.pi(modeloII, main = "idade+cdiab+crim+congenita")
> plot.pi(modeloIII, main = "idade+cdiab+crim+congenita+grande")
```

A linha sólida é o modelo ajustado e a linha pontilhada é o Kaplan-Meier

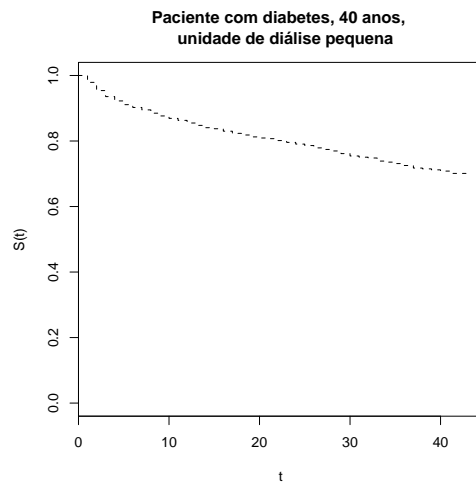




**Resposta:** O gráfico dos grupos com alto, médio e baixo índice de prognóstico indica que o modelo de Cox é capaz de distinguir entre os três grupos. As curvas estimadas acompanham de perto as curvas de Kaplan-Meier.

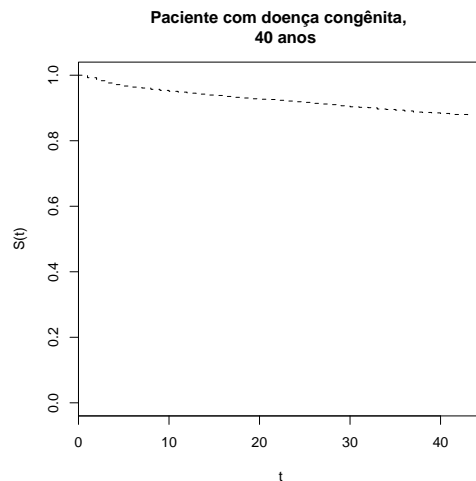
6. Podemos utilizar o modelo ajustado de Cox para obter estimativas de sobrevida para um paciente com determinado perfil. Por exemplo, o gráfico da curva de sobrevida (de acordo com o modelo III), para um paciente com causa de base diabetes, 40 anos de idade, e tratado em uma unidade pequena seria obtido com os comandos:

```
> paciente1 <- survfit(modeloIII, newdata = list(idade = 40, cdiab = factor(1,
+   levels = 0:1), crim = factor(0, levels = 0:1), congenita = factor(0,
+   levels = 0:1), grande = factor(0, levels = 0:1)))
> par(mfrow = c(1, 1))
> plot(paciente1, conf = F, lty = 2, ylab = "S(t)", xlab = "t",
+   main = "Paciente com diabetes, 40 anos, \n unidade de diálise pequena")
```



7. E um paciente com mesma idade e local de tratamento, mas com doença congênita?

```
> paciente2 <- survfit(modeloIII, newdata = list(idade = 40, cdiab = factor(0,
+   levels = 0:1), crim = factor(0, levels = 0:1), congenita = factor(1,
+   levels = 0:1), grande = factor(0, levels = 0:1)))
> plot(paciente2, conf = F, lty = 2, ylab = "S(t)", xlab = "t",
+   main = "Paciente com doença congênita, \n 40 anos")
```



**Exercício 7.3:** No exercício 6.3 ajustamos um modelo explicativo para sobrevida em Aids contendo as variáveis sexo, idade, tratamento e tipo de acompanhamento,

utilizando regressão paramétrica (banco de dados `ipec.csv`). Faça agora uma análise desses dados utilizando o modelo de Cox, considerando três modelos explicativos aninhados.

**Modelo I:**  $\text{sobrevida} = \text{idade} + \text{sexo}$

**Modelo II:**  $\text{sobrevida} = \text{idade} + \text{sexo} + \text{acompan}$

**Modelo III:**  $\text{sobrevida} = \text{idade} + \text{sexo} + \text{acompan} + \text{tratam}$

1. Faça gráficos de Kaplan-Meier estratificados por cada covariável categórica para verificar o pressuposto dos modelos de Cox.

Lendo os dados

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

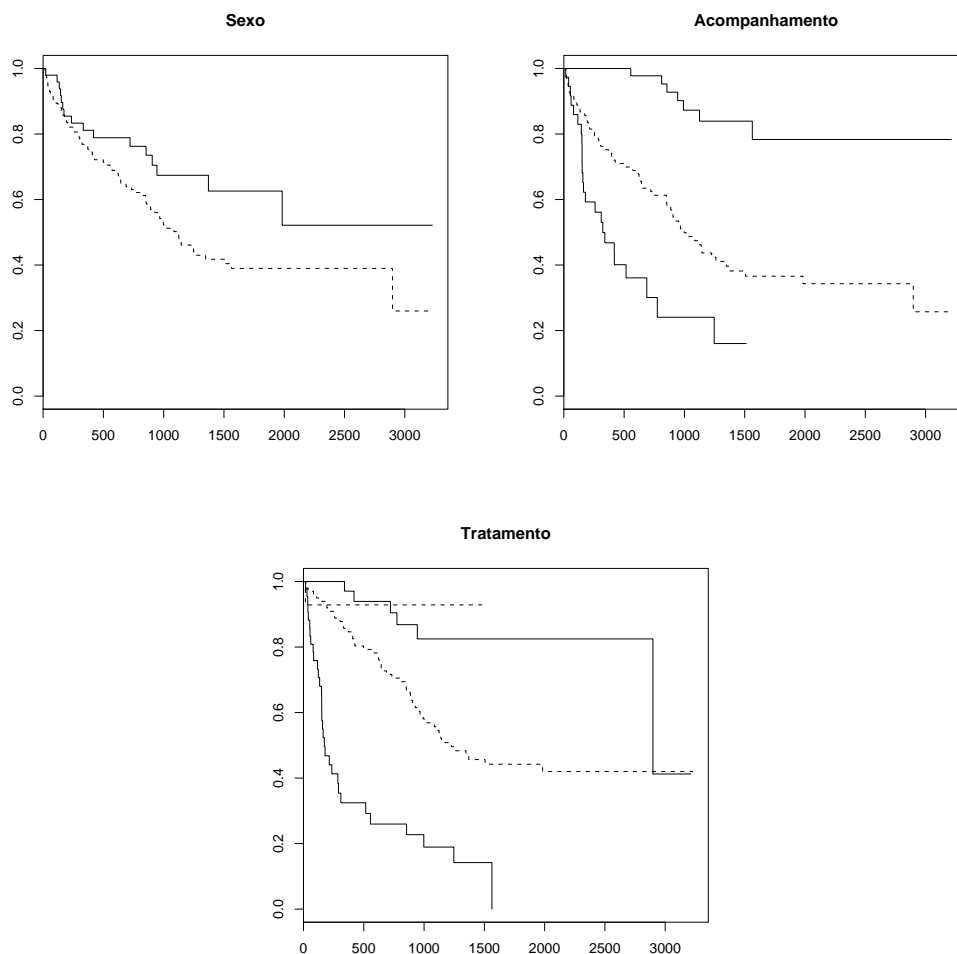
 [1] "id"      "ini"     "fim"     "tempo"   "status"  "sexo"   "escola"
 [8] "idade"   "risco"   "acompan" "obito"   "anotrat" "tratam" "doenca"
[15] "propcp"

> ipec$acompan <- factor(ipec$acompan)
> ipec$tratam <- factor(ipec$tratam)
```

Estimando os Kaplan-Meier

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

> par(mfrow = c(2, 2))
> plot(KMsexo, lty = c(1, 2), main = "Sexo", mark.time = F)
> plot(KMacompan, lty = c(1, 2), main = "Acompanhamento", mark.time = F)
> plot(KMtrat, lty = c(1, 2), main = "Tratamento", mark.time = F)
```



**Resposta:** Nas duas primeiras variáveis, não há mudança no tempo entre as categorias (não se cruzam e caem relativamente paralelas). A terceira variável tem muitas categorias e fica mais difícil de avaliar. Não há violação óbvia (linhas cruzando por exemplo), mas também não há um padrão claro de proporcionalidade. Vamos prosseguir com a análise, mas com cuidado.

2. Ajuste cada modelo causal acima utilizando o modelo de riscos proporcionais de Cox, tomando o cuidado de interpretar os parâmetros a cada saída. (Lembre de acrescentar o argumento  $x = T$ , na especificação do `coxph`, para que se possa utilizá-lo depois para obter os índices de prognóstico.)

```
> modeloI <- coxph(Surv(tempo, status) ~ idade + sexo, data = ipec,
```



```
+ x = T)
> summary(modeloI)
```

Call:

```
coxph(formula = Surv(tempo, status) ~ idade + sexo, data = ipec,
      x = T)
```

n= 193

	coef	exp(coef)	se(coef)	z	p
idade	-0.0127	0.987	0.0116	-1.10	0.270
sexoM	0.5562	1.744	0.2761	2.01	0.044

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.987	1.013	0.965	1.01
sexoM	1.744	0.573	1.015	3.00

Rsquare= 0.029 (max possible= 0.988 )

Likelihood ratio test= 5.64 on 2 df, p=0.0597

Wald test = 5.09 on 2 df, p=0.0783

Score (logrank) test = 5.18 on 2 df, p=0.0748

```
> modeloII <- coxph(Surv(tempo, status) ~ idade + sexo + acompan,
+ data = ipec, x = T)
> summary(modeloII)
```

Call:

```
coxph(formula = Surv(tempo, status) ~ idade + sexo + acompan,
      data = ipec, x = T)
```

n= 193

	coef	exp(coef)	se(coef)	z	p
idade	-0.00166	0.998	0.0120	-0.138	8.9e-01
sexoM	0.27218	1.313	0.2818	0.966	3.3e-01
acompan1	1.70732	5.514	0.4010	4.257	2.1e-05
acompan2	2.51763	12.399	0.4452	5.654	1.6e-08

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.998	1.0017	0.975	1.02
sexoM	1.313	0.7617	0.756	2.28
acompan1	5.514	0.1814	2.513	12.10
acompan2	12.399	0.0807	5.181	29.67

Rsquare= 0.231 (max possible= 0.988 )

Likelihood ratio test= 50.7 on 4 df, p=2.54e-10

Wald test = 37 on 4 df, p=1.79e-07

Score (logrank) test = 49 on 4 df, p=5.84e-10

```
> modeloIII <- coxph(Surv(tempo, status) ~ idade + sexo + acompan +
+   tratam, data = ipec, x = T)
> summary(modeloIII)
```

Call:

```
coxph(formula = Surv(tempo, status) ~ idade + sexo + acompan +
      tratam, data = ipec, x = T)
```

n= 193

	coef	exp(coef)	se(coef)	z	p
idade	0.00143	1.0014	0.0121	0.118	9.1e-01
sexoM	0.07424	1.0771	0.2858	0.260	8.0e-01
acompan1	1.67618	5.3451	0.4084	4.105	4.0e-05
acompan2	2.15300	8.6107	0.4672	4.608	4.1e-06
tratam1	-1.24192	0.2888	0.3011	-4.124	3.7e-05
tratam2	-2.09674	0.1229	0.4705	-4.456	8.3e-06
tratam3	-2.94502	0.0526	1.0188	-2.891	3.8e-03

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	1.0014	0.999	0.97792	1.026
sexoM	1.0771	0.928	0.61518	1.886
acompan1	5.3451	0.187	2.40080	11.900
acompan2	8.6107	0.116	3.44629	21.514
tratam1	0.2888	3.462	0.16007	0.521
tratam2	0.1229	8.140	0.04885	0.309
tratam3	0.0526	19.011	0.00714	0.387

Rsquare= 0.372 (max possible= 0.988 )

Likelihood ratio test= 89.8 on 7 df, p=1.11e-16

Wald test = 76.9 on 7 df, p=5.97e-14

Score (logrank) test = 100 on 7 df, p=0

**Resposta:** O modelo I indica que sexo masculino é um importante fator de risco, com razão de riscos de 1.744. Este modelo, porém, tem um poder explicativo muito baixo e não é significativamente melhor do que o modelo nulo (teste Wald = 5.09,  $p = 0.0783$ ). A inclusão da variável acompanhamento melhora muito o ajuste ( $R^2 = 0.231/0.988 \times 100 = 23.3\%$ ). Neste modelo, as variáveis demográficas se tornam não significativas. A variável *acompan* é um importante fator de risco, o que é razoável, pois pacientes internados têm maior risco de ir a óbito do que pacientes atendidos no ambulatório apenas (uma vez que o tipo de atendimento está associado com a gravidade do quadro clínico). O modelo III mostra que a variável tratamento tem forte efeito protetor, mesmo controlando por tipo de acompanhamento. Um paciente sem tratamento tem 3.5 vezes mais chance de ir a óbito, por unidade de tempo, do

que o paciente com monoterapia, 8.1 vezes mais chance do que um paciente com terapia combinada e 19 vezes mais chance do que o paciente recebendo terapia potente.

3. Compare os modelos usando a análise de deviance e o gráfico dos índices de prognóstico.

```
> anova(modeloI, modeloII, modeloIII, test = "Chisq")
```

Analysis of Deviance Table

Model 1: Surv(tempo, status) ~ idade + sexo

Model 2: Surv(tempo, status) ~ idade + sexo + acompan

Model 3: Surv(tempo, status) ~ idade + sexo + acompan + tratam

	Resid. Df	Resid. Dev	Df	Deviance	P(> Chi )
1	191	844.60			
2	189	799.50	2	45.10	1.612e-10
3	186	760.39	3	39.11	1.648e-08

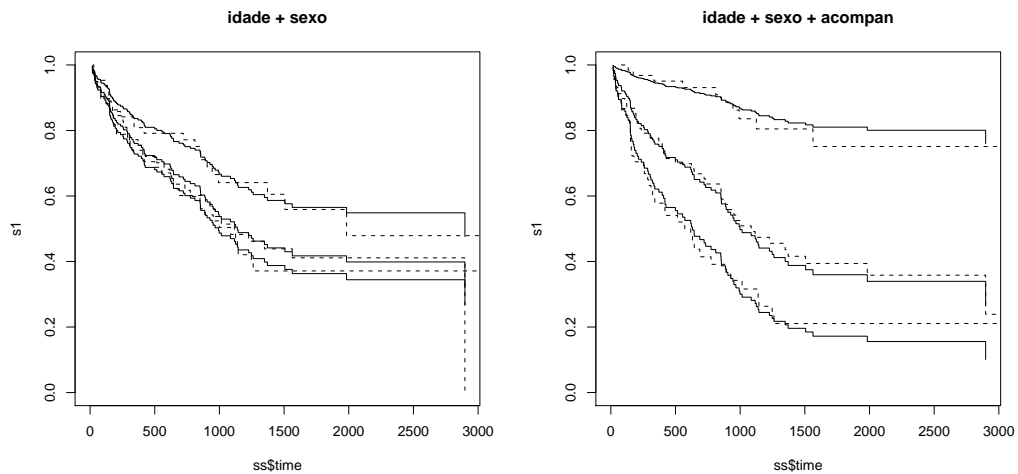
```
> par(mfrow = c(2, 2))
```

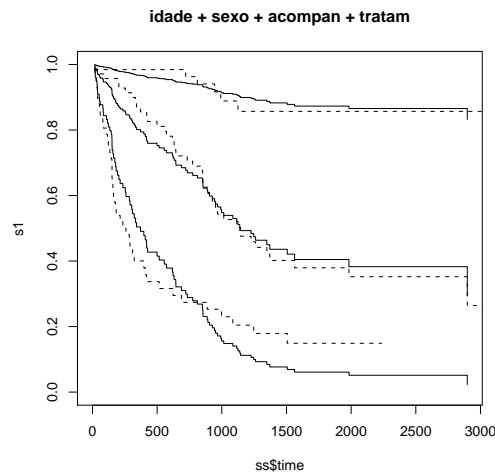
```
> plot.pi(modeloI, main = "idade + sexo")
```

```
> plot.pi(modeloII, main = "idade + sexo + acompan")
```

```
> plot.pi(modeloIII, main = "idade + sexo + acompan + tratam")
```

A linha sólida é o modelo ajustado e a linha pontilhada é o Kaplan-Meier





**Resposta:** A análise de deviance claramente apoia o modelo III, porém o gráfico de índices de prognóstico mostra que o modelo III está mais descolado da curva Kaplan-Meier do que o modelo II. Isto sugere que a variável tratamento, embora ajude na explicação da variância, provavelmente não teve seu efeito capturado pelo modelo de riscos proporcionais de Cox. Isto confirma a impressão de não proporcionalidade sugerida pela análise dos gráficos de Kaplan-Meier inicial. Concluindo, esta análise sugere que o tratamento tem efeito significativo na sobrevida, mas o valor quantitativo estimado do efeito não é confiável, devido ao desvio do pressuposto de proporcionalidade. Outros modelos, que relaxam o pressuposto de proporcionalidade, são necessários para uma melhor estimativa do efeito do tratamento na sobrevida em aids.

4. Qual o poder explicativo do modelo escolhido? Calcule a razão entre o  $R^2$  do modelo escolhido e o  $R^2$  máximo (ambos estão presentes na saída do comando `summary()`).

**Resposta:** Escolhendo o modelo III, temos que seu poder explicativo foi de  $0.372/0.988 \times 100 = 37.65\%$ .