

ANÁLISE DE SOBREVIVÊNCIA

Teoria e aplicações em saúde

Caderno de Respostas

Capítulo 6

Modelo semiparamétrico de sobrevivência

Exercício 6.1: Para os dados da coorte de transplante de medula óssea vistos nesse capítulo, refaça toda a análise utilizando o arquivo `tmoclas.dat` (Apêndice 12.4):

- a. Faça os gráficos de Kaplan-Meier para as variáveis `sexo`, `deag`, `decr` e `fase` e verifique o pressuposto de proporcionalidade.

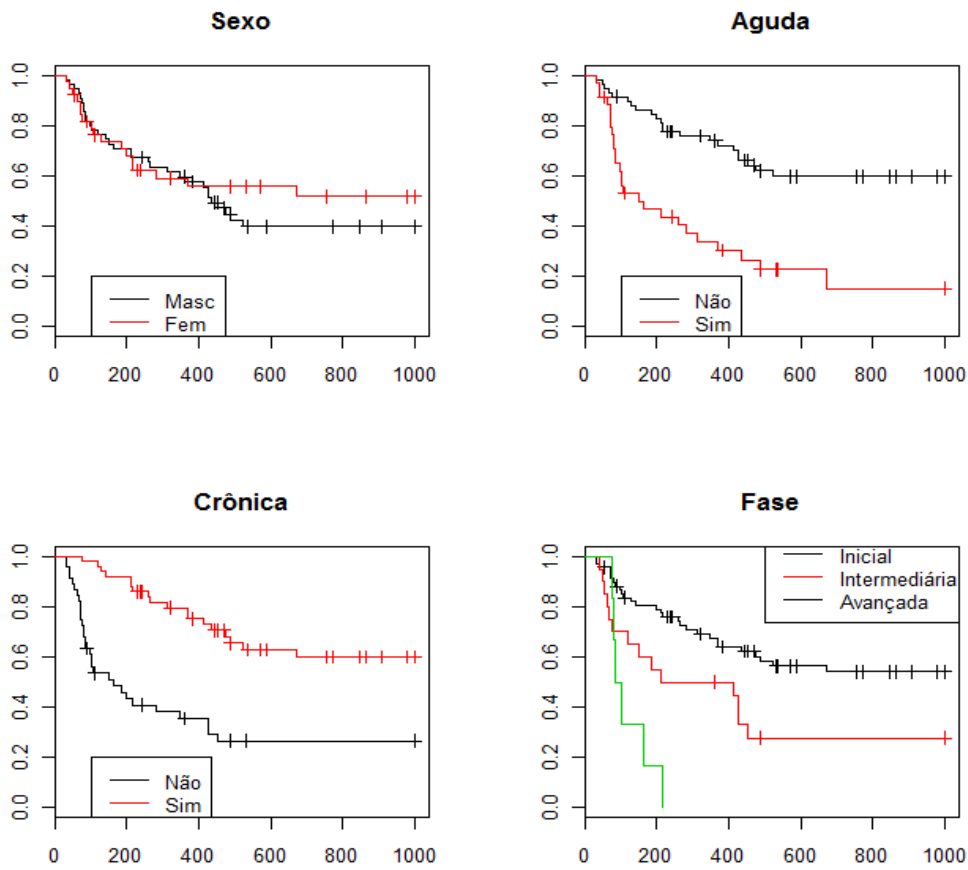
Resposta:

```
# Modelo de Cox
# Exercício 6.1
# Abrindo o arquivo com os dados do transplante de medula óssea
tmo <- read.table("tmoclas.dat", header = T)
# vendo o nome das variáveis
names(tmo)
# Transformando as variáveis categóricas em fator
tmo$sexo <- factor(tmo$sexo)
tmo$decr <- factor(tmo$decr)
tmo$deag <- factor(tmo$deag)
tmo$fase <- factor(tmo$fase)

# Abrindo a biblioteca survival e criando o objeto sobrevivência
library(survival)
y <- Surv(tmo$os, tmo$status)

# item a
# Estimando as curvas de sobrevivência pelo Kaplan-Meier para 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))
# SEXO
plot(KMsexo, main = "Sexo", col = 1:2)
legend(100, 0.2, c("Masc", "Fem"), col=1:2, lty=1)
# DOENÇA AGUDA
plot(KMdeag, main = "Aguda", col = 1:2)
legend(100, 0.2, c("Não", "Sim"), col=1:2, lty=1)
# DOENÇA CRONICA
plot(KMdecr, main = "Crônica", col = 1:2)
legend(100, 0.2, c("Não", "Sim"), col=1:2, lty=1)
# FASE DA DOENÇA
plot(KMfase, main = "Fase", col = 1:3)
legend(500, 1.1, c("Inicial", "Intermediária", "Avançada"), col=1:3, lty=1)
```



a. Ajuste os quatro modelos de Cox apresentados no texto e compare os modelos usando o teste da razão de verossimilhanças.

Resposta:

item b

Estimando os modelos de Cox

Modelo 1 – Idade e Sexo

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

```
summary(mod1)
```

Call:

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

```
n= 96, number of events= 49
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
idade	-0.02167	0.97857	0.01399	-1.548	0.122
sexo2	-0.37649	0.68626	0.32120	-1.172	0.241

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.9786	1.022	0.9521	1.006
sexo2	0.6863	1.457	0.3657	1.288

```

Concordance= 0.568 (se = 0.044 )
Rsquare= 0.03 (max possible= 0.986 )
Likelihood ratio test= 2.92 on 2 df, p=0.232
Wald test = 2.85 on 2 df, p=0.2408
Score (logrank) test = 2.85 on 2 df, p=0.2406

```

```

# Modelo 2 – Idade, Sexo e fase da doença
mod2 <- coxph(y ~ idade + sexo + fase, data = tmo)
summary(mod2)

```

```

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

```

```

n= 96, number of events= 49

```

	coef	exp(coef)	se(coef)	z	Pr(> z)
idade	-0.02786	0.97252	0.01441	-1.933	0.053230 .
sexo2	-0.20776	0.81240	0.32428	-0.641	0.521742
fase2	0.86502	2.37505	0.34873	2.481	0.013119 *
fase3	1.89091	6.62539	0.48690	3.884	0.000103 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.9725	1.0283	0.9454	1.000
sexo2	0.8124	1.2309	0.4303	1.534
fase2	2.3751	0.4210	1.1991	4.704
fase3	6.6254	0.1509	2.5513	17.205

```

Concordance= 0.649 (se = 0.044 )
Rsquare= 0.166 (max possible= 0.986 )
Likelihood ratio test= 17.41 on 4 df, p=0.00161
Wald test = 19.65 on 4 df, p=0.0005853
Score (logrank) test = 23.12 on 4 df, p=0.0001196

```

```

# Modelo 3 – Idade, Sexo, fase da doença e doença aguda
mod3 <- coxph(y ~ idade + sexo + fase + deag, data = tmo)
summary(mod3)

```

```

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

```

```

n= 96, number of events= 49

```

	coef	exp(coef)	se(coef)	z	Pr(> z)
idade	-0.01877	0.98140	0.01429	-1.314	0.188888
sexo2	-0.24258	0.78460	0.31841	-0.762	0.446149
fase2	0.94870	2.58234	0.34103	2.782	0.005405 **
fase3	1.52206	4.58164	0.49600	3.069	0.002150 **
deag1	1.13600	3.11429	0.30972	3.668	0.000245 ***

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.9814	1.0190	0.9543	1.009
sexo2	0.7846	1.2745	0.4204	1.464
fase2	2.5823	0.3872	1.3235	5.038

```
fase3      4.5816      0.2183      1.7331      12.112
deag1      3.1143      0.3211      1.6972      5.715
```

```
Concordance= 0.72 (se = 0.044 )
Rsquare= 0.272 (max possible= 0.986 )
Likelihood ratio test= 30.52 on 5 df, p=1.167e-05
Wald test = 31.81 on 5 df, p=6.491e-06
Score (logrank) test = 37.67 on 5 df, p=4.405e-07
```

```
# Modelo 4 – Idade, Sexo, fase da doença, doença aguda e doença crônica
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, number of events= 49
```

```
      coef exp(coef) se(coef)      z Pr(>|z|)
idade -0.005019  0.994993  0.014912 -0.337 0.736415
sexo2 -0.271984  0.761866  0.332115 -0.819 0.412816
fase2  0.593973  1.811171  0.371427  1.599 0.109784
fase3  0.938411  2.555918  0.522357  1.796 0.072416 .
deag1  1.190381  3.288332  0.327485  3.635 0.000278 ***
decr1 -1.061750  0.345850  0.338452 -3.137 0.001706 **
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
      exp(coef) exp(-coef) lower .95 upper .95
idade    0.9950    1.0050    0.9663    1.0245
sexo2    0.7619    1.3126    0.3974    1.4607
fase2    1.8112    0.5521    0.8746    3.7508
fase3    2.5559    0.3912    0.9182    7.1151
deag1    3.2883    0.3041    1.7307    6.2479
decr1    0.3459    2.8914    0.1782    0.6714
```

```
Concordance= 0.768 (se = 0.044 )
Rsquare= 0.345 (max possible= 0.986 )
Likelihood ratio test= 40.67 on 6 df, p=3.365e-07
Wald test = 38.46 on 6 df, p=9.113e-07
Score (logrank) test = 47.62 on 6 df, p=1.405e-08
```

```
# Selecionando o melhor modelo entre os modelos aninhados e de mesmo n (sem dados faltantes)
```

```
# Teste da Razão de verossimilhança
anova(mod1, mod2, mod3, mod4, test = "Chisq")
```

```
Analysis of Deviance Table
```

```
Cox model: response is y
```

```
Model 1: ~ idade + sexo
```

```
Model 2: ~ idade + sexo + fase
```

```
Model 3: ~ idade + sexo + fase + deag
```

```
Model 4: ~ idade + sexo + fase + deag + decr
```

```
loglik Chisq Df P(>|Chi|)
```

```
1 -201.94
```

```
2 -194.70 14.486 2 0.0007152 ***
```

```

3 -188.15 13.109 1 0.0002939 ***
4 -183.07 10.152 1 0.0014413 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

O modelo 4 é o mais adequado, a inclusão das variáveis foi significativa.

b. Qual o poder explicativo do modelo?

Resposta:

O modelo 4 apresentou um poder explicativo absoluto de 34%. Em relação ao máximo possível (98,4%) este modelo alcançou 34,9% de poder explicativo. A probabilidade de concordância estimada pelo modelo 4 teve alto valor discriminatório ou preditivo (76,8%).

c. Calcule o índice de prognóstico do modelo final para um indivíduo com 35 anos, do sexo feminino, na fase intermediária da doença e que tenha tido doença do enxerto crônica, mas não aguda.

Resposta:

```

# item c
# índice de prognóstico paciente com 35 anos, sexo feminino, fase intermediária e
# doença do enxerto crônica não aguda

# Colocando em um vetor as características do individuo na ordem
# em que aparecem no modelo (idade, sexo, fase, deag, decr)
# sexo feminino (sexo2 = 1)
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
-0.005019351 -0.271984459  0.593973494  0.938411452  1.190380544 -1.061749580

# 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] -0.9154378

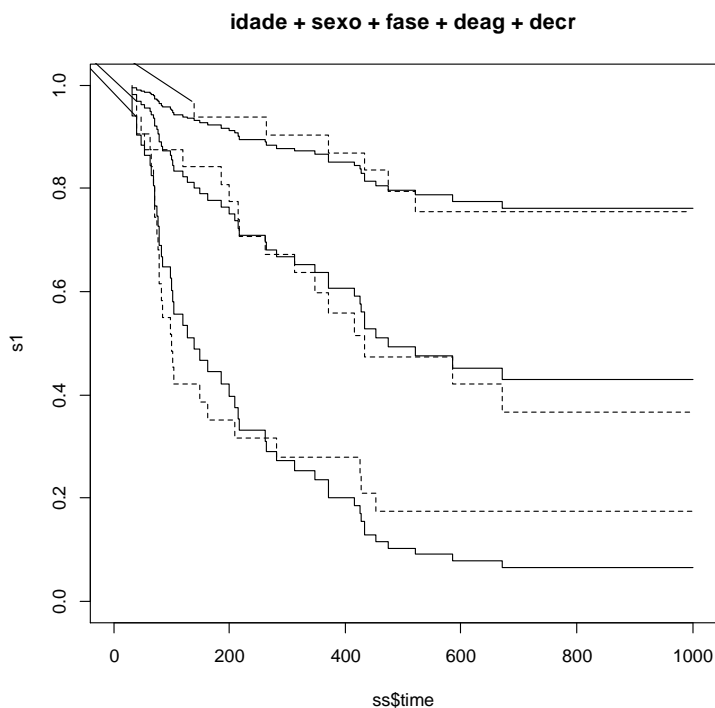
```

Temos então que o índice de prognóstico do indivíduo com as características citadas anteriormente é igual a -0.915.

d. 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.

Resposta:

```
# item e
# Executando a função Rfun
# para obter o gráfico Índice Prognóstico - IP
source("Rfun.r")
# utilizando a Função plot.pi (modelo escolhido)
# Acrescentando o argumento x = True, na especificação da função coxph()
# para ser possível rodar o Gráfico IP do modelo.
mod4 <- coxph(y ~ idade + sexo + fase + deag + decr, data = tmo, x=T)
par(mfrow = c(1, 1))
plot.pi(mod4, main = " idade + sexo + fase + deag + decr ")
# A linha sólida é o modelo ajustado e a linha pontilhada é o Kaplan-Meier
```



Exercício 6.2: Três modelos aninhados são propostos para explicar a sobrevivência de pacientes em diálise. O primeiro considera apenas a variável idade, o segundo inclui as doenças de base (cdiab, congênita, crim) e o terceiro inclui uma variável contextual que classifica o tamanho da unidade de tratamento (grande).

Modelo I: sobrevivência = idade

Modelo II: sobrevivência = idade + cdiab + congenita + crim

Modelo III: sobrevivência = idade + cdiab + congenita + crim + grande

- 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).

Resposta:

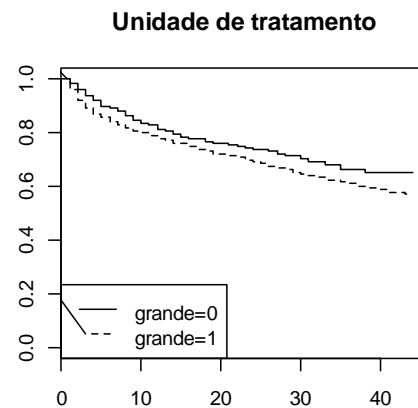
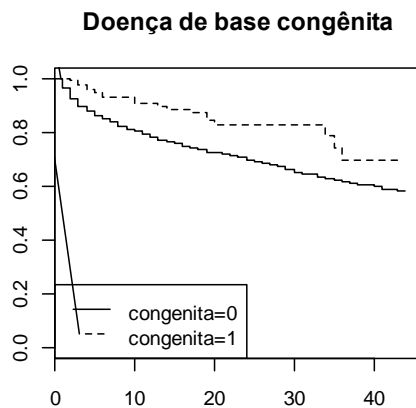
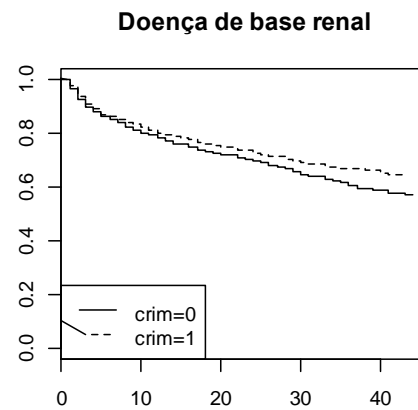
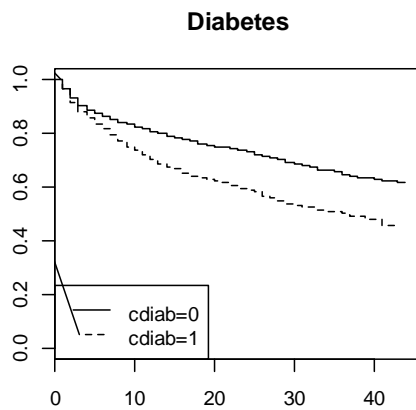
```
# Exercício 6.2
# item a
# lendo o banco
dialise <- read.table("dialise.csv", header = T, sep = ",")

# Acertando as variáveis
dialise$grande <- factor(dialise$grande)
dialise$cdiab <- factor(dialise$cdiab)
dialise$crim <- factor(dialise$crim)
dialise$congenita <- factor(dialise$congenita)
names(dialise)

# Sobrevivência
library(survival)
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(x="bottomleft", legend=names(KMdiab$strata), lty=c(1,2))
plot(KMrim, lty = c(1, 2), main = "Doença de base renal", mark.time = F)
legend(x="bottomleft", legend=names(KMrim$strata), lty=c(1,2))
plot(KMcong, lty = c(1, 2), main = "Doença de base congênita", mark.time = F)
legend(x="bottomleft", legend=names(KMcong$strata), lty=c(1,2))
plot(KMgrande, lty = c(1, 2), main = "Unidade de tratamento", mark.time = F)
legend(x="bottomleft", legend=names(KMgrande$strata), lty=c(1,2))
```

Observando o gráfico 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 tamanho da amostra. 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 variáveis.

- b. Ajuste cada modelo explicativo acima utilizando o modelo de riscos proporcionais de Cox, tomando o cuidado de interpretar os parâmetros a cada modelo.

Resposta:

item b

Rodando Cox com argumento $x = T$, para obter os IP dos modelos.

Modelo com idade

```
modelol <- coxph(y ~ idade, data=dialise, x=T)
```

```
summary(modelol)
```

Call:

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

```
n= 6805, number of events= 1603
```

```
      coef exp(coef) se(coef)      z Pr(>|z|)
idade 0.034903  1.035520 0.001727 20.22  <2e-16 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
      exp(coef) exp(-coef) lower .95 upper .95
idade      1.036      0.9657      1.032      1.039
```

```
Concordance= 0.644 (se = 0.008 )
```

```
Rsquare= 0.062 (max possible= 0.98 )
```

```
Likelihood ratio test= 435.1 on 1 df, p=0
```

```
Wald test = 408.7 on 1 df, p=0
```

```
Score (logrank) test = 415.2 on 1 df, p=0
```

```
# Modelo de sobrevivência com idade, cdiab, congenita e crim
```

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

```
Call:
```

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

```
n= 6805, number of events= 1603
```

```
      coef exp(coef) se(coef)      z Pr(>|z|)
idade      0.034448  1.035048 0.001751 19.677 < 2e-16 ***
cdiab1      0.285606  1.330568 0.060308  4.736 2.18e-06 ***
crim1       0.030331  1.030796 0.066975  0.453 0.65064
congenital1 -0.715243  0.489073 0.226181 -3.162 0.00157 **
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
      exp(coef) exp(-coef) lower .95 upper .95
idade      1.0350      0.9661      1.0315      1.0386
cdiab1      1.3306      0.7516      1.1822      1.4975
crim1       1.0308      0.9701      0.9040      1.1754
congenital1 0.4891      2.0447      0.3139      0.7619
```

```
Concordance= 0.649 (se = 0.008 )
```

```
Rsquare= 0.067 (max possible= 0.98 )
```

```
Likelihood ratio test= 473.3 on 4 df, p=0
```

```
Wald test = 438.5 on 4 df, p=0
```

```
Score (logrank) test = 451.3 on 4 df, p=0
```

```
# Modelo de sobrevivência acrescentando a variável grande
```

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

```
Call:
```

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

```

n= 6805, number of events= 1603

              coef exp(coef)  se(coef)      z Pr(>|z|)
idade          0.033995  1.034580  0.001754 19.386 < 2e-16 ***
cdiabl         0.295484  1.343777  0.060398  4.892 9.97e-07 ***
criml          0.019132  1.019316  0.067083  0.285  0.77550
congenital1   -0.727439  0.483145  0.226222 -3.216  0.00130 **
grandel        0.176994  1.193624  0.063390  2.792  0.00524 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
idade          1.0346    0.9666    1.0310    1.0381
cdiabl         1.3438    0.7442    1.1938    1.5126
criml          1.0193    0.9811    0.8937    1.1625
congenital1    0.4831    2.0698    0.3101    0.7527
grandel        1.1936    0.8378    1.0542    1.3515

Concordance= 0.652 (se = 0.008 )
Rsquare= 0.068 (max possible= 0.98 )
Likelihood ratio test= 481.4 on 5 df, p=0
Waldtest           = 447.4 on 5 df, p=0
Score (logrank) test = 460.8 on 5 df, p=0

```

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 doença congênitas 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 3 salas.

- c. Compare os modelos usando o teste da razão de verossimilhanças e escolha o modelo com melhor ajuste.

Resposta:

```
# item c
```

```
# Comparando os 3 modelos
```

```
anova(modelol, modeloll, modelolll, test = "Chisq")
```

```
Analysis of Deviance Table
```

```
Cox model: response is y
```

```
Model 1: ~ idade
```

```
Model 2: ~ idade + cdiab + crim + congenita
```

```

Model 3: ~ idade + cdiab + crim + congenita + grande
loglik   Chisq Df P(>|Chi|)
1 -13037
2 -13017 38.2457 3 2.507e-08 ***
3 -13013 8.0653 1 0.004512 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

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.

- d. Qual o poder explicativo do modelo escolhido? Calcule a razão entre o R2 do modelo escolhido e o R2 máximo.

Resposta:

Como o R2 do modelo ajustado (modelo III) foi 0,068, mas o R2 máximo foi 0,98, o modelo III explicou apenas $0.068/0.98 \times 100 = 6.9\%$ da variância dos dados. A concordância deste modelo foi de 65,2%, valor discriminatório bom, resultado comum observado em modelos de sobrevivência.

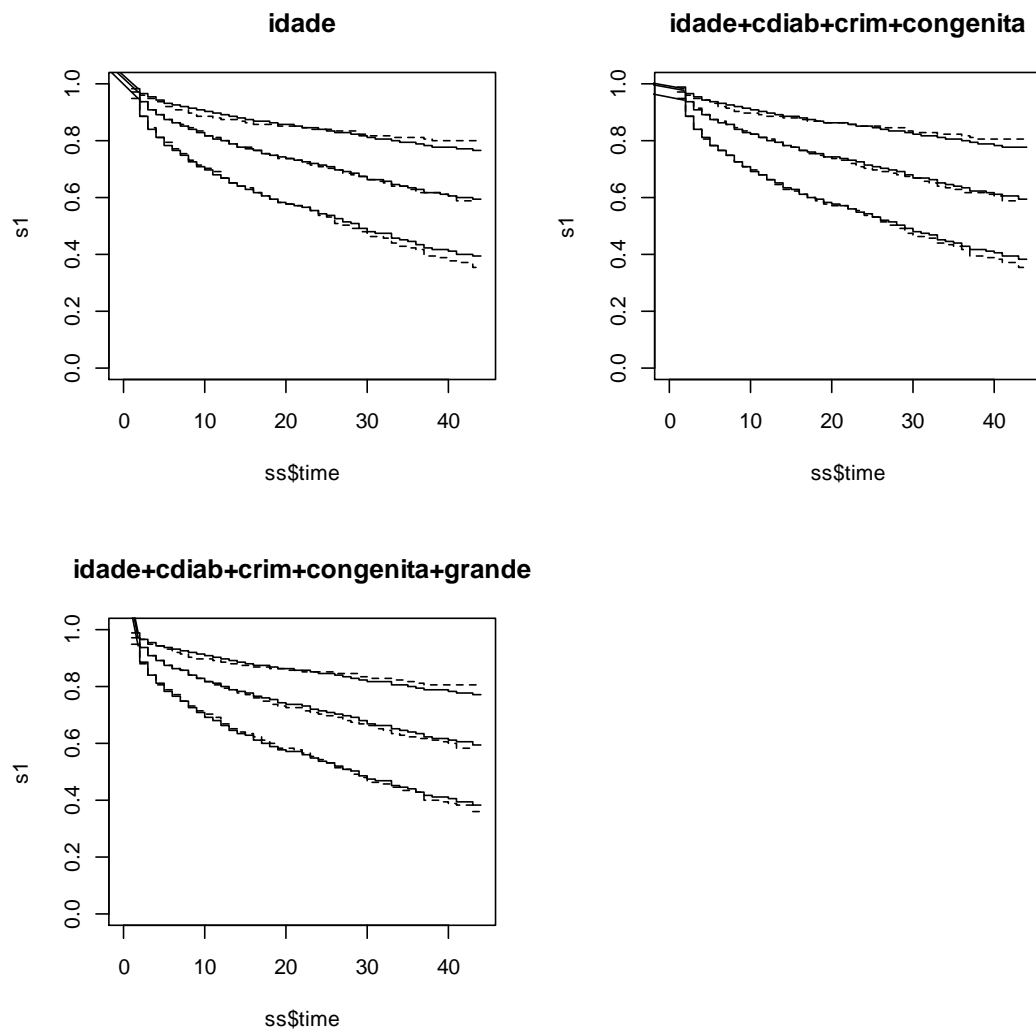
- e. Faça o gráfico dos índices de prognóstico.

Resposta:

```

# item e
# Executando a função Rfun
# para obter o gráfico Índice Prognóstico - IP
source("Rfun.r")
par(mfrow = c(2, 2))
# utilizando a Função plot.pi (modelo escolhido)
# Não se esqueça que para rodar o Gráfico PI é necessário que o modelo de
# Cox tenha sido rodado com a opção x= T
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

```



No gráfico IP a linha sólida é o modelo ajustado e a linha pontilhada é o Kaplan-Meier. 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.

OBS: A mensagem de erro abaixo, caso apareça na saída do R, não compromete os resultados e pode ser desconsiderada.

MENSAGEM DE ERRO NO PLOT.PI
 The solid line is the fitted model, the dashed one is the K-M

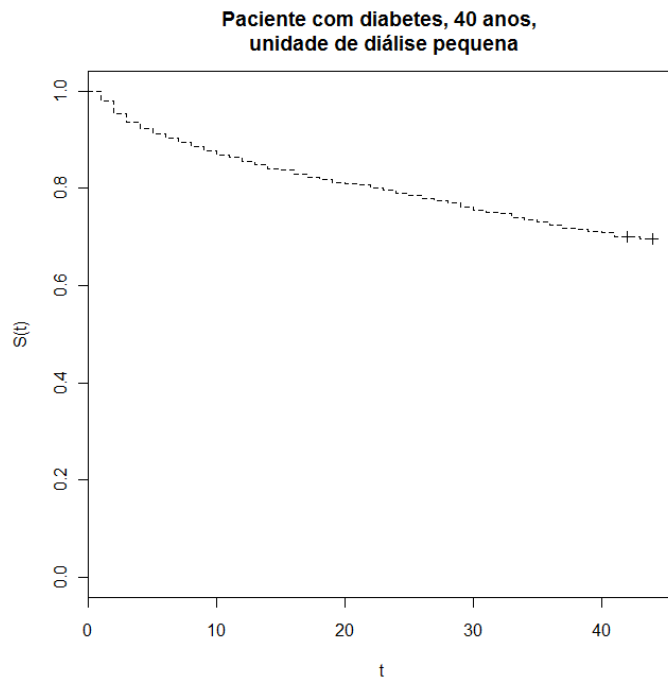
- f. Construa a curva de sobrevivência esperada pelo modelo III de um paciente com diabetes como doença de base, 40 anos de idade e tratado em uma unidade pequena.

Resposta:
 # item f

```

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")

```



- g.** Faça o mesmo para um paciente com mesma idade e local de tratamento, mas com doença de base congênita.

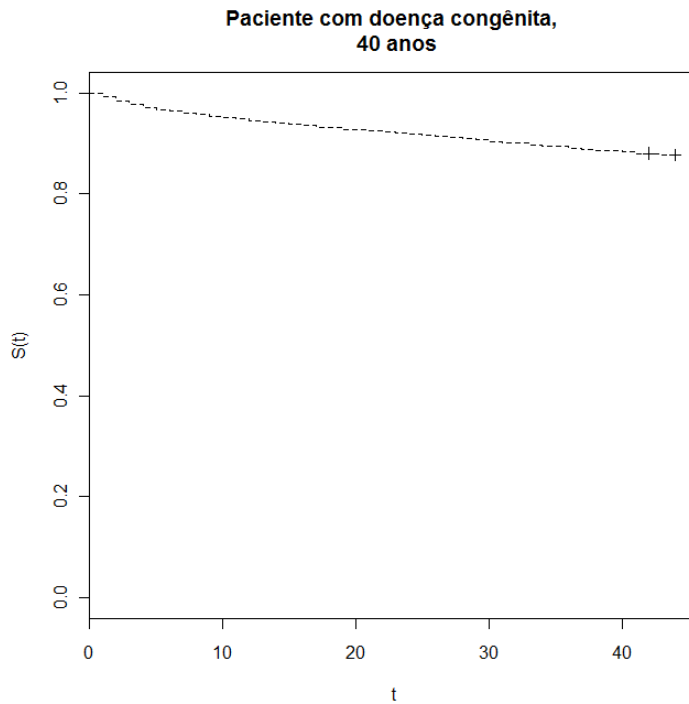
Resposta:

item g

```

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 6.3: No capítulo 5, ajustamos um modelo explicativo para sobrevivência em Aids contendo as variáveis sexo, idade, tratamento e tipo de acompanhamento, utilizando regressão paramétrica (banco de dados ipec.csv – Apêndice 12.4). Faça agora uma análise desses dados utilizando o modelo de Cox, considerando três modelos explicativos aninhados.

Modelo I: sobrevivência = idade + sexo

Modelo II: sobrevivência = idade + sexo + acompan

Modelo III: sobrevivência = idade + sexo + acompan + tratam

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

Resposta:

Exercício 6.3

item a

Kaplan-Meier

Lendo os dados

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

```
names(ipec)
```

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

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

```
# Estimando os ajustes de Kaplan-Meier
```

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

```
# Estratificando
```

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

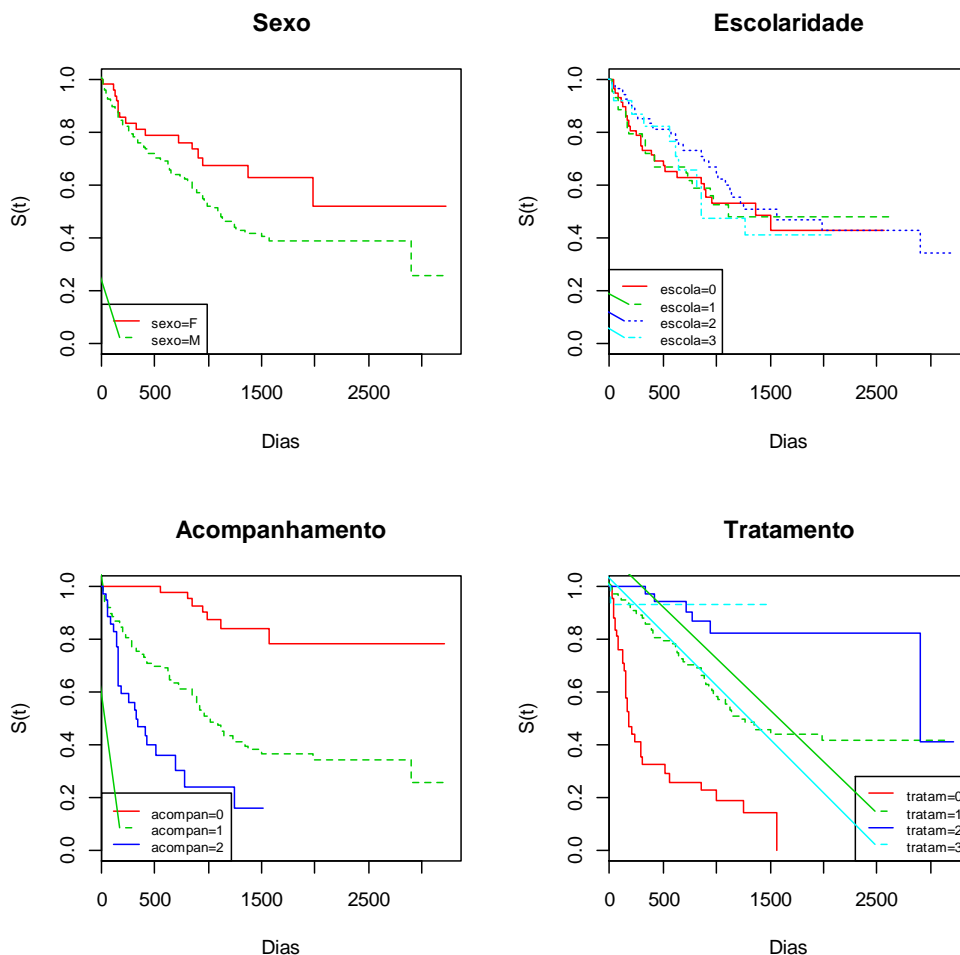
```
KMesc <- survfit(Surv(tempo, status) ~ escola, data = ipec)
```

```
KMacompan <- survfit(Surv(tempo, status) ~ acompan, data = ipec)
```

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

```
# Gráficos das variáveis
```

```
par(mfrow = c(2, 2))
plot(KMsexo, lty = 1:2, col = 2:3, main = "Sexo", ylab = "S(t)", xlab = "Dias", mark.time = F)
legend(x="bottomleft", legend=names(KMsexo$strata), lty=1:2, col=2:3, cex=0.7)
plot(KMesc, lty = 1:4, col = 2:5, main = "Escolaridade", ylab = "S(t)", xlab = "Dias",
conf.int = F, mark.time = F)
legend(x="bottomleft", legend=names(KMesc$strata), lty=1:4, col=2:5, cex=0.7)
plot(KMacompan, lty = c(1, 2), col = 2:4, main = "Acompanhamento", ylab = "S(t)", xlab = "Dias",
mark.time = F)
legend(x="bottomleft", legend=names(KMacompan$strata), lty=c(1,2), col=2:4, cex=0.7)
plot(KMtrat, lty = c(1, 2), col = 2:5, main = "Tratamento", ylab = "S(t)", xlab = "Dias",
mark.time = F)
legend(x="bottomright", legend=names(KMtrat$strata), lty=c(1,2), col=2:5, cex=0.7)
```



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.

- b. Ajuste cada modelo acima utilizando o modelo de riscos proporcionais de Cox, tomando o cuidado de interpretar os parâmetros a cada modelo.

Resposta:

item b

Modelos de Cox

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

```
summary(modelol)
```

Call:

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

n= 193, number of events= 90

	coef	exp(coef)	se(coef)	z	Pr(> z)
idade	-0.01274	0.98734	0.01162	-1.096	0.273
sexoM	0.55622	1.74407	0.27611	2.015	0.044 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.9873	1.0128	0.9651	1.010
sexoM	1.7441	0.5734	1.0152	2.996

Concordance= 0.556 (se = 0.033)

Rsquare= 0.029 (max possible= 0.988)

Likelihood ratio test= 5.64 on 2 df, p=0.05965

Wald test = 5.09 on 2 df, p=0.07834

Score (logrank) test = 5.18 on 2 df, p=0.07485

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

```
summary(modeloll)
```

Call:

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

n= 193, number of events= 90

	coef	exp(coef)	se(coef)	z	Pr(> z)
idade	-0.001659	0.998342	0.012013	-0.138	0.890
sexoM	0.272185	1.312830	0.281849	0.966	0.334
acompan1	1.707324	5.514184	0.401021	4.257	2.07e-05 ***
acompan2	2.517632	12.399195	0.445248	5.654	1.56e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.9983	1.00166	0.9751	1.022
sexoM	1.3128	0.76171	0.7556	2.281
acompan1	5.5142	0.18135	2.5126	12.101
acompan2	12.3992	0.08065	5.1808	29.675

Concordance= 0.711 (se = 0.033)

Rsquare= 0.231 (max possible= 0.988)
 Likelihood ratio test= 50.74 on 4 df, p=2.535e-10
 Wald test = 37.01 on 4 df, p=1.795e-07
 Score (logrank) test = 49 on 4 df, p=5.844e-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, number of events= 90

	coef	exp(coef)	se(coef)	z	Pr(> z)
idade	0.001428	1.001429	0.012120	0.118	0.90621
sexoM	0.074236	1.077061	0.285759	0.260	0.79503
acompan1	1.676179	5.345096	0.408363	4.105	4.05e-05 ***
acompan2	2.153003	8.610675	0.467205	4.608	4.06e-06 ***
tratam1	-1.241920	0.288829	0.301134	-4.124	3.72e-05 ***
tratam2	-2.096740	0.122856	0.470499	-4.456	8.33e-06 ***
tratam3	-2.945019	0.052601	1.018814	-2.891	0.00384 **

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	1.0014	0.9986	0.977921	1.0255
sexoM	1.0771	0.9285	0.615179	1.8857
acompan1	5.3451	0.1871	2.400800	11.9002
acompan2	8.6107	0.1161	3.446289	21.5141
tratam1	0.2888	3.4623	0.160072	0.5212
tratam2	0.1229	8.1396	0.048855	0.3089
tratam3	0.0526	19.0110	0.007141	0.3874

Concordance= 0.789 (se = 0.033)
Rsquare= 0.372 (max possible= 0.988)
 Likelihood ratio test= 89.84 on 7 df, p=1.11e-16
 Wald test = 76.87 on 7 df, p=5.973e-14
 Score (logrank) test = 100.1 on 7 df, p=0

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. A inclusão da variável tratamento melhora ainda mais o ajuste ($R^2 =$

$0.372/0.988 \times 100 = 37,7\%$) e tem resultado discriminatório muito bom (concordância = 0,789).

- c. Compare os modelos usando o teste da razão de verossimilhanças e escolha o modelo com melhor ajuste e o gráfico dos índices de prognóstico.

Resposta:

item c

Comparando os modelos

anova(modelol, modeloll, modelolll, test = "Chisq")

Analysis of Deviance Table

Cox model: response is Surv(tempo, status)

Model 1: ~ idade + sexo

Model 2: ~ idade + sexo + acompan

Model 3: ~ idade + sexo + acompan + tratam

	loglik	Chisq	Df	P(> Chi)
1	-422.30			
2	-399.75	45.097	2	1.612e-10 ***
3	-380.20	39.106	3	1.648e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Gráficos do Índice Prognóstico

par(mfrow = c(2, 2))

source("Rfun.r")

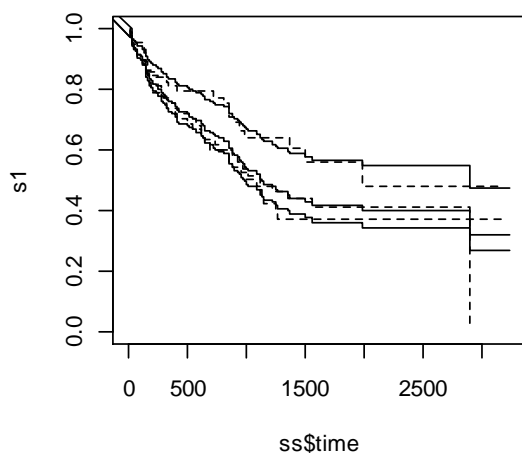
plot.pi(modelol, main = "Modelo I: idade + sexo")

plot.pi(modeloll, main = "Modelo II: idade + sexo + acompan")

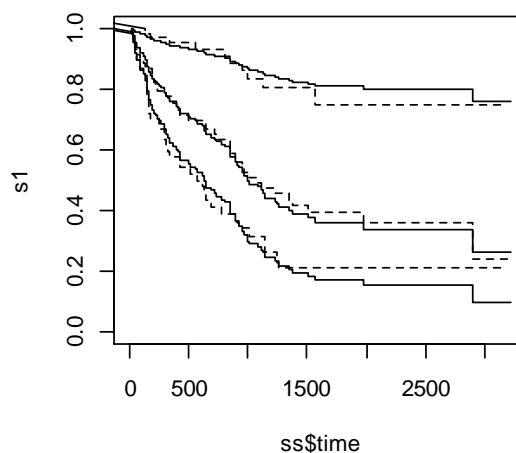
plot.pi(modelolll, main = "Modelo III: idade + sexo + acompan + tratam")

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

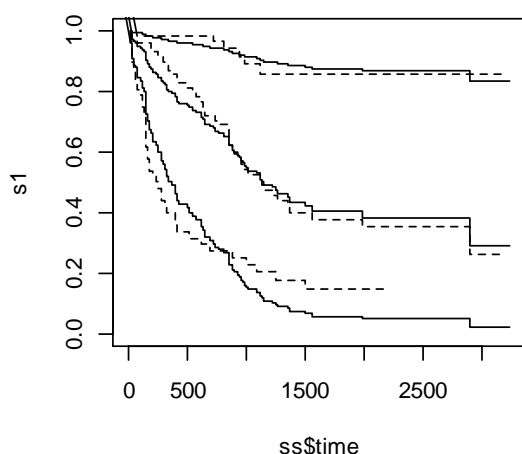
Modelo I: idade + sexo



Modelo II: idade + sexo + acompan



Modelo III: idade + sexo + acompan + trata



O teste da razão de verossimilhança 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 sobrevivência, 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 sobrevivência em Aids.

- d. Qual o poder explicativo do modelo escolhido? Calcule a razão entre o R2 do modelo escolhido e o R2 máximo.

Resposta: A inclusão da variável tratamento melhora ainda mais o ajuste, com poder explicativo de 37.7% ($0.372/0.988 \times 100$) e apresenta muito bom resultado discriminatório (concordância = 0,789).