

ANÁLISE DE SOBREVIVÊNCIA

Teoria e aplicações em saúde

Caderno de Respostas

Capítulo 9

Funções de suavização

9.1 Analise o comportamento da variável idade no risco de óbito de pacientes em hemodiálise (banco dialise.csv - Apêndice 12.4), analisado no Capítulo 6.

- a. Descreva o efeito aparente da idade no gráfico de resíduo martingale de um modelo com somente essa variável.

Resposta:

Exercício 9.1

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

item a

```
require (survival)
```

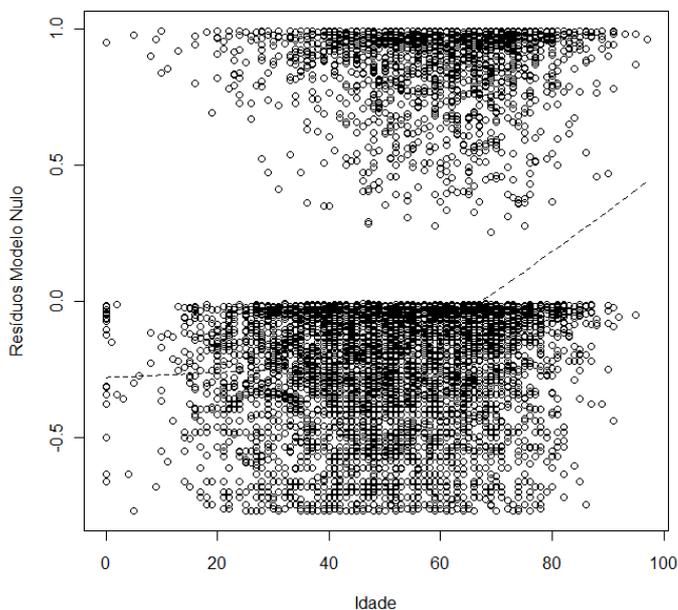
Residuo de Martingale do modelo Nulo

```
dialise.resnulo <- residuals (coxph(Surv(inicio, fim, status) ~ 1, data = dialise),"m")
```

Gráfico Residuo Modelo Nulo vs Idade

```
plot (dialise$idade,dialise.resnulo,xlab="Idade", ylab="Resíduos Modelo Nulo")
```

```
lines (lowess(dialise$idade, dialise.resnulo),lty=2)
```



Na análise da forma funcional da idade, avaliando o resíduo de Martingale do modelo nulo vs idade, observa-se uma forte inflexão entre os 40 e 60 anos, indicando que deve existir uma associação – não linear - entre o tempo de sobrevivência e a idade do indivíduo no início da diálise.

Modelo com Idade

```
dialise.idade <- coxph(Surv(inicio, fim, status) ~ idade, data = dialise)
```

```
summary(dialise.idade)
```

Call:

```
coxph(formula = Surv(inicio, fim, status) ~ idade, data = dialise)
```

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

| | coef | exp(coef) | se(coef) | z | Pr(> z) |
|-------|----------|-----------|----------|-------|------------|
| idade | 0.037252 | 1.037954 | 0.001738 | 21.43 | <2e-16 *** |

```
---
```

```

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

      exp(coef) exp(-coef) lower .95 upper .95
idade      1.038      0.9634      1.034      1.041

Concordance= 0.665 (se = 0.008 )
Rsquare= 0.069 (max possible= 0.974 )
Likelihood ratio test= 489.8 on 1 df, p=0
Wald test               = 459.2 on 1 df, p=0
Score (logrank) test = 467.6 on 1 df, p=0

```

A cada ano de vida o risco de óbito aumenta 3,8% (IC95% 3,4 – 4,1).

```

# Resíduos de Martingale - modelo com Idade – Verificação de pontos aberrantes
dialise.idade.res <- residuals ((dialise.idade),"m")
# Vendo alguns resíduos
dialise.idade.res [c(1,2,100,102)]

```

```

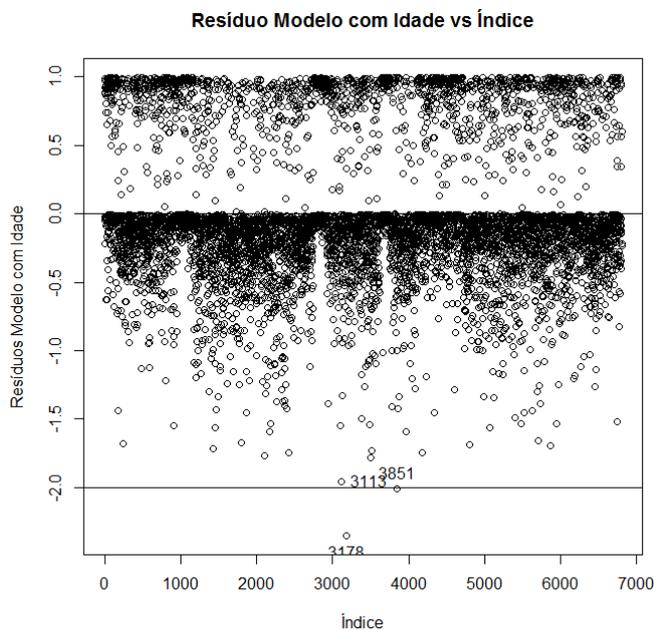
      1          2          100          102
-0.02036232 -0.03178973  0.98101907  0.97930329

```

```

# Gráfico Resíduo Modelo com Idade – Verificação de pontos aberrantes
plot (dialise.idade.res, main="Resíduo Modelo com Idade vs Índice",
      xlab="Índice", ylab="Resíduos Modelo com Idade")
abline(h=0)
abline(h=-2)

```



Os resíduos positivos de Martingale indicam que o tempo observado é maior que o estimado pelo modelo, os negativos indicam o inverso. Apenas os indivíduos 3178 e 3851 apresentam resíduos ≤ -2 , indicando que tiveram tempo de sobrevivência muito menor que o estimado pelo modelo.

b. Ajuste um modelo que permita que o efeito da variável idade varie suavemente.

Resposta:

item b

Modelo com função Spline

```
idade.spl <- coxph(Surv(inicio, fim, status) ~ pspline(idade), data = dialise)
summary(idade.spl)
```

Call:

```
coxph(formula = Surv(inicio, fim, status) ~ pspline(idade), data =
dialise)
```

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

```
                coef    se(coef) se2      Chisq DF   p
pspline(idade), linear 0.0369 0.00168 0.00168 482.1 1.00 0.000
pspline(idade), nonlin                10.8 3.08 0.014
```

```
                exp(coef) exp(-coef) lower .95 upper .95
ps(idade)3          0.997    1.0028    0.528    1.88
ps(idade)4          0.991    1.0094    0.342    2.87
ps(idade)5          0.988    1.0117    0.270    3.62
ps(idade)6          1.063    0.9410    0.269    4.20
ps(idade)7          1.356    0.7376    0.345    5.32
ps(idade)8          2.057    0.4862    0.533    7.94
ps(idade)9          3.097    0.3229    0.805   11.92
ps(idade)10         4.389    0.2279    1.139   16.91
ps(idade)11         6.161    0.1623    1.595   23.79
ps(idade)12        10.185    0.0982    2.607   39.80
ps(idade)13        18.866    0.0530    4.462   79.77
ps(idade)14        35.291    0.0283    6.604  188.59
```

Iterations: 6 outer, 19 Newton-Raphson

Theta= 0.96

Degrees of freedom for terms= 4.1

Concordance= 0.665 (se = 0.008)

Rsquare= 0.071 (max possible= 0.974)

Likelihood ratio test= 502 on 4.08 df, p=0

Wald test = 493 on 4.08 df, p=0

Observamos que a partir do segmento 10 (oitavo nó) a idade passa a ser fator de risco para o óbito em pacientes em diálise.

c. Interprete os resultados do ajuste da variável idade suavizada

Resposta:

O spline é significativo tanto na parte linear ($p < 0,001$) quanto na não linear ($p < 0,0014$). O risco passa a ser significativo do segmento 10 em diante.

No spline os segmentos são de mesmo tamanho e como foram indicados 11 nós (14-3) para um intervalo de idade de 0 a 97 anos, cada segmento tem tamanho 8,18 anos (97/11). Assim, a idade a partir da qual o risco passa a ser significativo corresponde a 70,5 anos (8,18 x 8).

Comparando o modelo com idade linear e idade suavizada.

Teste da Razão da verossimilhança.

anova (dialise.idade, idade.spl)

```

Analysis of Deviance Table
Cox model: response is Surv(inicio, fim, status)
Model 1: ~ idade
Model 2: ~ pspline(idade)
loglik  Chisq Df P(>|Chi|)
1 -12216
2 -12210 12.204 11 0.3485

```

Verificamos que a inclusão da idade suavizada não melhorou o ajuste.

d. Faça e interprete um gráfico do efeito da variável na sobrevivência

Resposta:

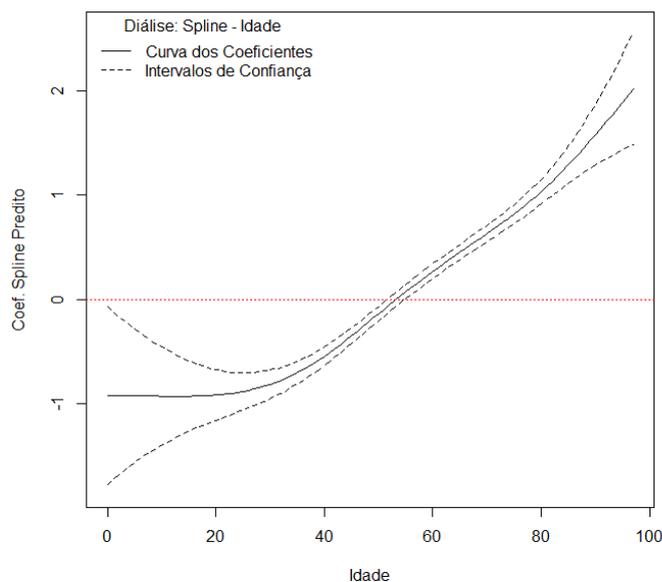
item c

Gráfico Spline Idade

```
termplot(idade.spl,term="pspline(idade)",se=T, ylab="Coef. Spline Predito", xlab =
"Idade", col.term=1, col.se=1)
```

```
abline(h=0,lty=3,col="red")
```

```
legend("topleft",title="Diálise: Spline - Idade", lty=1:2,bty="n", col=1, legend =
c("Curva dos Coeficientes","Intervalos de Confiança"))
```



Observando o gráfico podemos identificar a idade na qual o risco de morte dos pacientes em diálise se torna significativo é pouco antes dos **60 anos**. O modelo ajustado poderia ou deveria incluir, entre outras variáveis, a idade suavizada com spline ou em faixas etárias determinadas pelo spline, por exemplo, (0 – 30], (30 - 60] e (60 – 100].

CONTRADITÓRIO

9.2 Para estudar a sobrevivência de pacientes infartados foram analisadas as internações do SUS por essa causa. Foram considerados fatores de risco do paciente – sexo e idade – disponíveis no Sistema de Informações Hospitalares, registradas na AIH. Com os dados deste estudo (infarto.dat – Apêndice 12.4):

- a. Descreva a relação entre idade e tempo de sobrevivência, procurando indicar com que idade o efeito sofre mudança importante de comportamento.

Resposta:

Exercício 9.2

```
infarto <- read.table("infarto.dat", header = T)
```

```
names(infarto)
```

```
[1] "hospital" "id" "ini" "fim" "status" "sexo"
"idade" "natureza" "volume" "luti"
```

criando a variável tempo a partir das variáveis início e fim de cada observação

```
infarto$tempo <- infarto$fim - infarto$ini
```

```
dim (infarto)
```

```
[1] 3176 11
```

```
summary(infarto)
```

```

      hospital      id      ini      fim      status
k       : 368  Min.   :  1.0  Min.   :  0.0  Min.   : 60.0  Min.   :0.0000
i       : 337  1st Qu.: 835.8  1st Qu.:232.0  1st Qu.:283.0  1st Qu.:0.0000
j       : 246  Median :1652.5  Median :354.0  Median :404.0  Median :0.0000
m       : 240  Mean   :1658.3  Mean   :363.1  Mean   :413.6  Mean   :0.1814
x       : 213  3rd Qu.:2492.2  3rd Qu.:492.0  3rd Qu.:545.0  3rd Qu.:0.0000
u       : 169  Max.   :3322.0  Max.   :677.0  Max.   :735.0  Max.   :1.0000
(Other):1603
sexo    idade      natureza  volume      luti      tempo
F:1181  Min.   :15.00  C : 60  vg:3118  1a24:2088  Min.   : 0.0
M:1995  1st Qu.:51.00  PE : 927  vp: 58  25+ : 801  1st Qu.:60.0
      Median :61.00  PFU: 531  n   : 287  Median :60.0
      Mean   :60.97  PM :1658  Mean   :50.5
      3rd Qu.:70.00  3rd Qu.:60.0
      Max.   :99.00  Max.   :60.0

```

```
table(infarto$tempo)
```

```

 0   1   2   3   4   5   6   7   8   9  10  11  12  13  14
99  96  52  40  34  25  22  14  23  14  20  12  10   9   6
15  16  17  18  19  20  21  22  23  24  25  26  27  28  29
 8   7   5   3   8   6   5   3   3   2   1   2   3   1   5
30  31  32  33  34  35  36  37  38  39  41  44  45  46  47
 2   4   2   1   4   4   2   3   2   3   2   1   1   1   1
48  49  50  51  52  53  54  55  56  57  59  60
 2   1   1   3   4   1   2   3   1   2   1 2584

```

```
table(infarto$tempo,infarto$status)
```

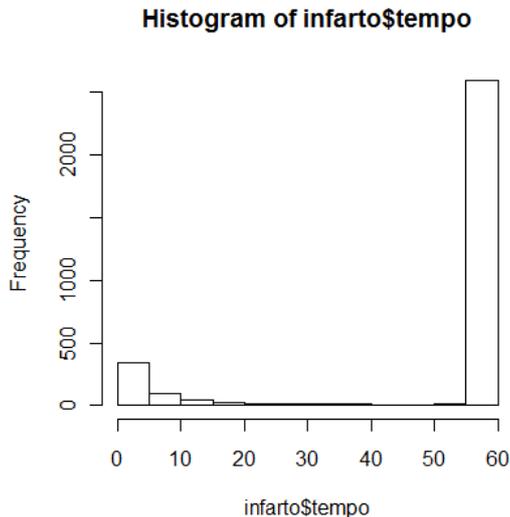
```

      0   1
0      2  97
1      0  96
2      0  52
3      0  40
.....
56     0   1
57     0   2
59     0   1
60 2584   0

```

Observamos que ocorreram 97 óbitos por infarto ocorridos no mesmo dia (tempo=0) e 96 com apenas um dia de internação. No outro extremo não encontramos nenhum óbito entre os 2.584 pacientes com 60 dias ou mais de internação.

hist(infarto\$tempo)



```
require (survival)
y <-Surv(infarto$ini, infarto$fim, infarto$status)
```

O R nos retorna a mensagem abaixo:

```
Mensagens de aviso perdidas:
In Surv(infarto$ini, infarto$fim, infarto$status) :
  Stop time must be > start time, NA created
```

Para entender melhor o que está acontecendo, vamos verificar os dados de sobrevivência (y):

```
y
[3108] (598,658+] [ NA,594 ] (603,663+] (603,605 ] (601,661+) ....
.....
[3173] (578,638+] (565,625+) [ NA,605 ] (577,581 ]....
```

Vamos observar os dados do paciente da linha 3175

tail(infarto)

| | hospital | id | ini | fim | status | sexo | idade | natureza | volume | luti | tempo |
|-------------|----------|-----------|------------|------------|----------|----------|-----------|-----------|-----------|-------------|----------|
| 3171 | y | 1841 | 558 | 563 | 1 | M | 87 | PE | vg | 1a24 | 5 |
| 3172 | y | 2877 | 574 | 634 | 0 | F | 75 | PE | vg | 1a24 | 60 |
| 3173 | y | 3028 | 578 | 638 | 0 | M | 61 | PE | vg | 1a24 | 60 |
| 3174 | y | 1081 | 565 | 625 | 0 | M | 55 | PE | vg | 1a24 | 60 |
| 3175 | y | 36 | 605 | 605 | 1 | M | 60 | PE | vg | 1a24 | 0 |
| 3176 | y | 86 | 577 | 581 | 1 | M | 64 | PE | vg | 1a24 | 4 |

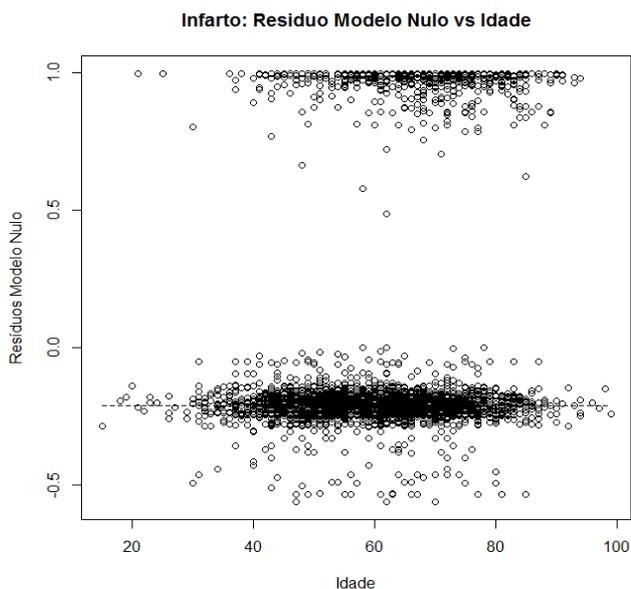
Pacientes com tempo zero de acompanhamento não podem ser considerados diretamente na análise. Para contornar este problema vamos considerar que todos os pacientes que morreram no mesmo dia da entrada na internação, viveram meio dia. Essa é uma decisão do pesquisador! O dado reflete a escala

de coleta. Nesse caso, a decisão poderia ser a exclusão dos pacientes que morreram antes de completar um dia, trabalhar com horas (se estiver disponível) ou adotar uma solução aproximada.

```
# Aumentado 0,5 dia o tempo dos pacientes com tempo=0
infarto$fim[infarto$fim == infarto$ini] <- infarto$fim[infarto$fim == infarto$ini] + 0.5
y <-Surv(infarto$ini, infarto$fim, infarto$status)
y
```

```
.....
[3166] (501,561.0+) (499,559.0+) (503,563.0+) (610,670.0+) (654,714.0+)
[3171] (558,563.0 ] (574,634.0+) (578,638.0+) (565,625.0+) (605,605.5 ]
[3176] (577,581.0 ]
```

```
# item a
# Resíduo de Martingale do modelo Nulo
infarto.resnulo <- residuals (coxph(Surv(ini, fim, status) ~ 1, data = infarto),"m")
# Gráfico Resíduo Modelo Nulo vs Idade
plot (infarto$idade,infarto.resnulo, main="Infarto: Residuo Modelo Nulo vs
Idade",xlab="Idade", ylab="Resíduos Modelo Nulo")
lines (lowess(infarto$idade, infarto.resnulo),lty=2)
```



Neste gráfico não é possível observar nenhuma relação não linear entre a idade e o tempo de sobrevivência.

```
# Modelo com Idade
infarto.idade <- coxph(Surv(ini, fim, status) ~ idade, data = infarto)
summary(infarto.idade)
```

```
Call:
coxph(formula = Surv(ini, fim, status) ~ idade, data = infarto)
```

```
n= 3176, number of events= 576
```

```
      coef exp(coef) se(coef)      z Pr(>|z|)
idade 0.047448  1.048592 0.003383 14.03  <2e-16 ***
---
```

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

```
      exp(coef) exp(-coef) lower .95 upper .95
idade      1.049      0.9537      1.042      1.056
```

```
Concordance= 0.671 (se = 0.012 )
Rsquare= 0.062 (max possible= 0.868 )
Likelihood ratio test= 204.1 on 1 df, p=0
Wald test = 196.8 on 1 df, p=0
Score (logrank) test = 201.2 on 1 df, p=0
```

A cada ano de vida o indivíduo tem o risco de morte por infarto significativamente aumentado em 4,9%.

```
# Resíduos do modelo com idade
infarto.idade.res <- residuals ((infarto.idade),"m")
# Vendo alguns resíduos
infarto.idade.res [c(1,2,3,4,5)]
```

```
      1      2      3      4      5
-0.1769143  0.9055075 -0.4624062 -0.1074866 -0.1237136
```

```
# Modelo com função Spline
idade.spl <- coxph(Surv(ini, fim, status) ~ pspline(idade), data = infarto)
summary(idade.spl)
```

```
Call:
coxph(formula = Surv(ini, fim, status) ~ pspline(idade), data = infarto)
```

```
n= 3176, number of events= 576
```

```
              coef  se(coef) se2      Chisq  DF  p
pspline(idade), linear 0.0475 0.00348 0.00348 186.62 1.00 0.000
pspline(idade), nonlin              6.77 3.07 0.084
```

```
      exp(coef) exp(-coef) lower .95 upper .95
ps(idade)3      0.958      1.044      0.2722      3.37
ps(idade)4      0.916      1.092      0.1044      8.03
ps(idade)5      0.895      1.117      0.0577     13.88
ps(idade)6      0.994      1.006      0.0503     19.65
ps(idade)7      1.234      0.810      0.0608     25.06
ps(idade)8      2.030      0.493      0.1029     40.06
ps(idade)9      3.505      0.285      0.1794     68.50
ps(idade)10     5.357      0.187      0.2736    104.88
ps(idade)11     7.261      0.138      0.3697    142.64
ps(idade)12     8.665      0.115      0.4349    172.66
ps(idade)13     8.422      0.119      0.3875    183.06
ps(idade)14     7.913      0.126      0.2707    231.28
```

```
Iterations: 5 outer, 15 Newton-Raphson
```

```
Theta= 0.881
```

```
Degrees of freedom for terms= 4.1
```

```
Concordance= 0.671 (se = 0.012 )
```

```
Rsquare= 0.065 (max possible= 0.868 )
```

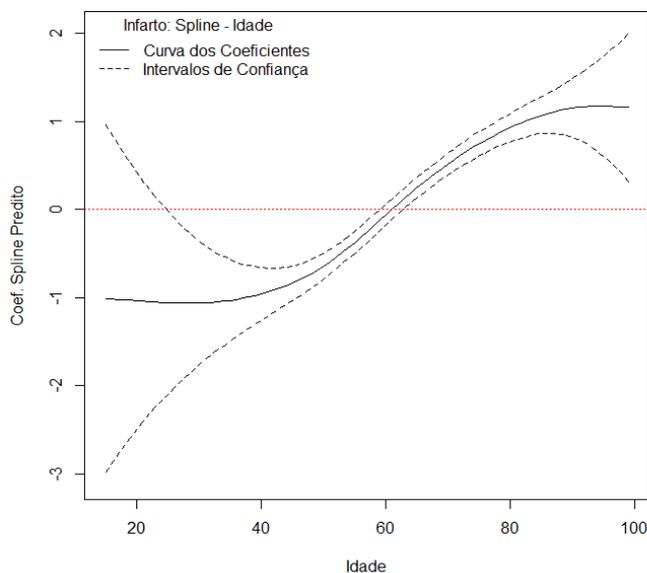
```
Likelihood ratio test= 212 on 4.07 df, p=0
```

```
Wald test = 193 on 4.07 df, p=0
```

O spline é significativo apenas na parte linear ($p < 0,001$), na não linear não há significância ($p > 0,08$). O risco não é significativo em nenhum dos segmentos. Neste caso não haveria necessidade de usar spline para incluir a idade no modelo.

Gráfico Spline Idade

```
termplot(idade.spl,term="pspline(idade)",se=T, ylab="Coef. Spline
Predito",xlab="Idade", col.term=1, col.se=1)
abline(h=0,lty=3,col="red")
legend("topleft",title="Infarto: Spline - Idade ",
lty=1:2,bty="n",col=1,legend=c("Curva dos Coeficientes",
"Intervalos de Confiança"))
```



Até próximo dos 40 anos o risco de morte por infarto é baixo (IC alto devido aos poucos casos) e se mantém estável, aumenta linearmente entre 40 e 80 anos. Após essa idade o intervalo de confiança volta a aumentar, por serem poucos casos.

b. Sugira uma hipótese biológica para explicar esse comportamento

Resposta:

Poucos infartos acontecem antes dos quarenta e por isso o intervalo de confiança até essa idade é imenso. Além disso, o infarto fulminante que ocorre entre os mais jovens não chega a ser internado, morre antes. A partir dos 50 anos, a idade tem relação linear com o desfecho: quanto mais velho mais morre. A mudança de inflexão da curva em idades avançadas indica que as pessoas de maior risco ou já faleceram, ou tiveram tempo de desenvolver circulação colateral. Idades mais avançadas têm como principal causa de óbito não mais o infarto, mas o AVC.

c. Indique outras formas de analisar sem o uso de spline

Resposta:

Outra forma de analisar a idade é usar os pontos de inflexão da curva para determinar pontos de corte para as faixas etárias. Ou adotar um modelo de regressão segmentado.

d. Escolha um ou mais pontos de corte para a variável idade, e ajuste uma regressão segmentada.

Resposta:

```
# item d
# Ajustando um modelo segmentado
# Criando a variável com a Idade em X categorias
# Idade segmentada (faixaet)
# faixaet <=40, 40< x <= 50, x>50

infarto$faixaet <- cut(infarto$idade,c(0,40,50,100))
table(infarto$faixaet)

(0,40] (40,50] (50,100]
      157      564      2455

table(infarto$faixaet, infarto$status)

      0      1
(0,40]  148     9
(40,50]  516    48
(50,100] 1936   519

# Modelo com idade categorizada
idade.cat <- coxph(Surv(ini, fim, status) ~ faixaet, data = infarto)
summary(idade.cat)

Call:
coxph(formula = Surv(ini, fim, status) ~ faixaet, data = infarto)

n= 3176, number of events= 576

      coef exp(coef) se(coef)      z Pr(>|z|)
faixaet(40,50]  0.4311    1.5389  0.3634  1.186  0.235
faixaet(50,100] 1.4627    4.3178  0.3363  4.349 1.37e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
faixaet(40,50]    1.539    0.6498    0.7549    3.137
faixaet(50,100]    4.318    0.2316    2.2335    8.347

Concordance= 0.576 (se = 0.009 )
Rsquare= 0.027 (max possible= 0.868 )
Likelihood ratio test= 86.63 on 2 df, p=0
Wald test = 63.41 on 2 df, p=1.699e-14
Score (logrank) test = 70.92 on 2 df, p=4.441e-16
```

Na faixa etária de 40 a 50 anos o risco de infartar é 54% maior do que até os 40 anos. Acima dos 50 o risco e 4 vezes maior.

```
# Modelo com idade segmentada (<50 e >=50)
infarto$idade50<- ifelse(infarto$idade>50, 0,infarto$idade)
infarto$idademais50<- ifelse(infarto$idade>50, infarto$idade-50,0)
# Verificando as novas variáveis
infarto$idade50
infarto$idademais50
# Modelo novo
idade.seg <- coxph(Surv(ini, fim, status) ~ idade50 + idademais50, data = infarto)
summary(idade.seg)
```

```
Call:
coxph(formula = Surv(ini, fim, status) ~ idade50 + idademais50,
      data = infarto)
```

```
n= 3176, number of events= 576
```

| | coef | exp(coef) | se(coef) | z | Pr(> z) |
|-------------|-----------|-----------|----------|--------|--------------|
| idade50 | -0.005608 | 0.994408 | 0.003711 | -1.511 | 0.131 |
| idademais50 | 0.048445 | 1.049637 | 0.004293 | 11.284 | <2e-16 *** |

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

| | exp(coef) | exp(-coef) | lower .95 | upper .95 |
|-------------|-----------|------------|---------------|--------------|
| idade50 | 0.9944 | 1.0056 | 0.9872 | 1.002 |
| idademais50 | 1.0496 | 0.9527 | 1.0408 | 1.059 |

```
Concordance= 0.672 (se = 0.012 )
```

```
Rsquare= 0.063 (max possible= 0.868 )
```

```
Likelihood ratio test= 205.2 on 2 df, p=0
```

```
Wald test = 202.9 on 2 df, p=0
```

```
Score (logrank) test = 224.2 on 2 df, p=0
```

Para cada ano a mais de idade, até os 50 anos, o efeito da idade não é significativo ($p = 0.131$). Acima dos 50 anos, cada ano a mais aumenta o risco em 5% (IC de 1.04 a 1.06)

e. Modele o efeito do sexo para cada classe de idade proposta. O que você conclui sobre ser o sexo masculino nas faixas etárias mais jovens?

Resposta:

```
# Modelo com idade segmentada e sexo (interação)
mod3<- coxph(formula = Surv(ini, fim, status) ~ idade50*sexo +
idademais50*sexo, data = infarto)
summary(mod3)
```

```
Call:
coxph(formula = Surv(ini, fim, status) ~ idade50 * sexo + idademais50 *
      sexo, data = infarto)
```

```
n= 3176, number of events= 576
```

| | coef | exp(coef) | se(coef) | z | Pr(> z) |
|-------------|-----------|-----------|----------|--------|--------------|
| idade50 | 0.003615 | 1.003622 | 0.005446 | 0.664 | 0.5068 |
| sexoM | -0.167634 | 0.845664 | 0.196177 | -0.855 | 0.3928 |
| idademais50 | 0.047706 | 1.048862 | 0.006048 | 7.888 | 3.11e-15 *** |

```

idade50:sexoM      -0.015954  0.984173  0.007467 -2.137   0.0326 *
sexoM:idademais50 -0.002254  0.997749  0.008688 -0.259   0.7953
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
idade50             1.0036     0.9964     0.9930     1.0144
sexoM                0.8457     1.1825     0.5757     1.2422
idademais50         1.0489     0.9534     1.0365     1.0614
idade50:sexoM       0.9842     1.0161     0.9699     0.9987
sexoM:idademais50   0.9977     1.0023     0.9809     1.0149

Concordance= 0.679 (se = 0.012 )
Rsquare= 0.067 (max possible= 0.868 )
Likelihood ratio test= 221.2 on 5 df, p=0
Wald test               = 210.9 on 5 df, p=0
Score (logrank) test = 242.1 on 5 df, p=0

```

A interação com sexo em menores de 50 anos indica efeito protetor para as mulheres.

f. Aponte vantagens e desvantagens de cada modelo ajustado

Resposta:

Modelando separadamente estima-se o efeito dentro de cada grupo. Pode ser desejável para estabelecer políticas e/ou protocolos diferenciados. Entretanto não permite comparar os estimadores entre os grupos, pois os intervalos de confiança, por exemplo, dizem respeito a amostras diferentes.

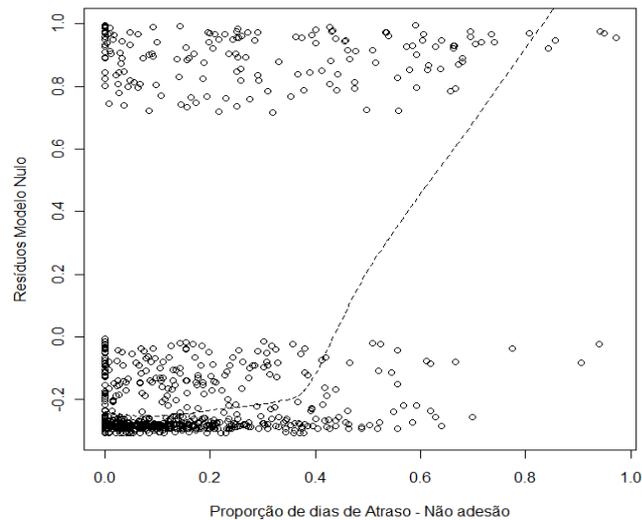
9.3 Refaça a análise apresentada no capítulo, utilizando o banco de dados adesão.dat descrito no Apêndice 12.4. Além disso, escolha um ou mais pontos de corte para a variável adesão e ajuste uma regressão segmentada.

Resposta:

```

# Exercício 9.3
adesao <- read.table("adesao.dat", header=T)
require (survival)
# Residuo de Martingale do modelo Nulo
adesao.resnulo <- residuals (coxph(Surv(ini, fim, status) ~ 1, data = adesao),"m")
# Gráfico Residuo Modelo Nulo vs Não adesão = proporção de dias de atraso
plot (adesao$propatraso,adesao.resnulo,xlab="Proporção de dias de Atraso - Não
adesão", ylab="Resíduos Modelo Nulo")
lines (lowess(adesao$propatraso, adesao.resnulo),lty=2, wd=2)

```



```
# Modelo com função Spline
adesao.spl <- coxph(Surv(ini, fim, status) ~ pspline(propatraso), data = adesao)
summary(adesao.spl)
```

Call:

```
coxph(formula = Surv(ini, fim, status) ~ pspline(propatraso),
      data = adesao)
n= 819
```

| | coef | se(coef) | se2 | Chisq | DF | p |
|---------------------------|------|----------|-------|-------|------|-------|
| pspline(propatraso), line | 3.8 | 0.309 | 0.306 | 151.4 | 1.00 | 0.000 |
| pspline(propatraso), nonl | | | | 10.9 | 3.06 | 0.013 |

| | exp(coef) | exp(-coef) | lower .95 | upper .95 |
|------------------|-----------|------------|-----------|-----------|
| ps(propatraso)2 | 0.581 | 1.7222 | 0.241 | 1.40 |
| ps(propatraso)3 | 0.416 | 2.4061 | 0.131 | 1.32 |
| ps(propatraso)4 | 0.611 | 1.6375 | 0.198 | 1.89 |
| ps(propatraso)5 | 1.017 | 0.9833 | 0.346 | 2.99 |
| ps(propatraso)6 | 1.448 | 0.6904 | 0.503 | 4.17 |
| ps(propatraso)7 | 2.242 | 0.4460 | 0.773 | 6.50 |
| ps(propatraso)8 | 3.596 | 0.2781 | 1.236 | 10.46 |
| ps(propatraso)9 | 6.141 | 0.1628 | 2.062 | 18.29 |
| ps(propatraso)10 | 9.763 | 0.1024 | 3.049 | 31.26 |
| ps(propatraso)11 | 13.735 | 0.0728 | 4.032 | 46.78 |
| ps(propatraso)12 | 19.015 | 0.0526 | 4.433 | 81.56 |
| ps(propatraso)13 | 26.538 | 0.0377 | 3.131 | 224.92 |

Iterations: 5 outer, 17 Newton-Raphson

Theta= 0.852

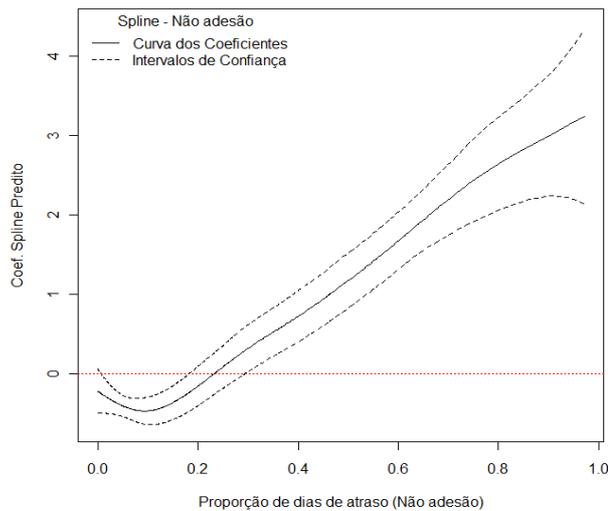
Degrees of freedom for terms= 4.1

Rsquare= 0.138 (max possible= 0.927)

Likelihood ratio test= 121 on 4.06 df, p=0

Wald test = 162 on 4.06 df, p=0

```
termplot(adesao.spl,term="pspline(propatraso)",se=T, ylab="Coef. Spline
Predito",xlab="Proporção de dias de atraso (Não adesão)", col.term=1, col.se=1)
abline(h=0,lty=3,col="red")
legend("topleft",title="Spline - Não adesão", lty=1:2,bty="n",col=1,legend=c("Curva
dos Coeficientes","Intervalos de Confiança"))
```



O ponto de inflexão da curva em torno de 10% pode determinar um ponto de corte, que representa uma adesão superior a 90%. A proporção de dias de atraso entre 10 e 30% (não adesão entre 70 e 90%) já apresenta um risco crescente de falha terapêutica, mas ainda não é significativo. A partir deste ponto, proporção de dias de atraso maior que 30% (Não adesão > 75%) o risco de falha passa a ser significativo.

Ajustando um modelo segmentado

```
adesao$pda <- cut(adesao$propatraso,c(0,.1,.30,1))
table(adesao$pda)
```

```
(0,0.1] (0.1,0.3] (0.3,1]
      291      258      164
```

Modelo com não adesão segmentada

```
adesao.seg <- coxph(Surv(ini, fim, status) ~ pda, data = adesao)
summary(adesao.seg)
```

Call:

```
coxph(formula = Surv(ini, fim, status) ~ pda, data = adesao)
```

```
n= 713, number of events= 148
(106 observations deleted due to missingness)
```

```
              coef exp(coef) se(coef)      z Pr(>|z|)
pda(0.1,0.3] 0.4676    1.5962  0.2265  2.064  0.039 *
pda(0.3,1]   1.6548    5.2320  0.2077  7.969 1.55e-15 ***
---
```

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

```
              exp(coef) exp(-coef) lower .95 upper .95
pda(0.1,0.3]    1.596    0.6265    1.024    2.488
pda(0.3,1]      5.232    0.1911    3.483    7.860
```

```
Rsquare= 0.095 (max possible= 0.925 )
```

```
Likelihood ratio test= 71.17 on 2 df, p=3.331e-16
```

```
Wald test = 75.72 on 2 df, p=0
```

```
Score (logrank) test = 90.4 on 2 df, p=0
```

No modelo apenas com a variável de adesão segmentada de acordo com o que foi observado no spline (3 categorias) encontramos que o risco de ocorrer falha terapêutica nos pacientes com média proporção de dias de atraso (entre 10 e 30% - adesão média) em relação aos pacientes com baixa proporção de dias de atraso (alta adesão) foi de 1,596 maior (IC95% 1,024 - 2,488), para os pacientes com alta proporção dias de atraso (>30% - baixa adesão) o risco foi de 5,232 (IC95% 3,483 – 7,860).