

# **ANÁLISE DE SOBREVIVÊNCIA**

## **Teoria e aplicações em saúde**

### **Caderno de Respostas**

---

#### **Capítulo 8**

#### **Covariáveis mudando no tempo**

**Exercício 8.1:** Em um estudo de sobrevivência de pacientes infectados pelo HIV, uma covariável importante é o momento em que a contagem de CD4 decresce abaixo de 200. Esse é um critério utilizado para classificar o indivíduo portador de HIV como caso de AIDS. A tabela abaixo mostra um trecho do banco de dados que resultaria de um estudo como esse, com desfecho óbito, onde id é o identificador do paciente e CD4 é um indicador do nível de CD4 (CD4=0, quando a contagem está acima de 200 e, CD4=1, quando a contagem está abaixo de 200).

id	inicio	fim	status	CD4
1	25	323	0	0
1	324	768	1	1
2	39	130	0	0
2	131	345	0	1
3	1	56	0	0
3	131	145	1	1

Com base nesses dados, descreva em palavras o que aconteceu com cada um dos três pacientes ao longo do período de acompanhamento.

**Resposta:**

- O paciente 1 entrou na coorte no 25º dia do estudo, no dia 324 o CD4 estava abaixo de 200 e no dia 768 ele morreu.
- O paciente 2 entrou na coorte no 39º dia do estudo, no dia 131 o CD4 estava abaixo de 200 e no dia 345 ele deixou de ser acompanhado, ou o estudo terminou (censura).
- O paciente 3 entrou na coorte no 1º dia do estudo, com CD4 acima de 200. Entre os dias 57 e 130 ele não foi acompanhado e no dia 131 o CD4 estava abaixo de 200. No dia 145 ele morreu.

**Exercício 8.2:** O tipo de tratamento empregado na terapia de um paciente com Aids pode ser considerado uma covariável tempo-dependente. Considerando 4 tipos possíveis de terapia (0 = sem anti-retrovirais; 1 = monoterapia, 2 = terapia combinada, 3 = terapia potente), pode-se observar o seguinte conjunto de dados:

id	inicio	fim	status	Terapia
1	2	124	0	0
1	124	213	0	1
1	213	230	1	3
2	21	210	1	3
3	17	56	0	0
3	56	99	0	1
3	99	154	0	2
3	154	255	1	3

- a. Com base nesses dados, descreva em palavras o que aconteceu com cada paciente ao longo do período de acompanhamento.

**Resposta:**

- O paciente 1 entrou na coorte no dia 2, sem nenhum tratamento, no dia 124 ele iniciou a monoterapia, no dia 213 iniciou a terapia potente e no dia 230 veio a falecer.

- O paciente 2 entrou na coorte no dia 21 já tomando a terapia potente e no dia 210 veio a falecer.
- O paciente 3 entrou na coorte no dia 17, sem nenhum tratamento, no dia 56 iniciou a monoterapia, no dia 99 iniciou a terapia combinada, no dia 154 iniciou a terapia potente e no dia 255 veio a falecer.

- b. Acrescente linhas a tabela acima para incorporar as informações do paciente 4, que entrou na coorte no dia 61, quando passou a ser tratado com a monoterapia. No dia 367 do estudo, ele muda para a terapia combinada, no dia 401 muda para a terapia potente, e vem a falecer no dia 460.

**Resposta:**

id	inicio	fim	status	terapia
4	61	367	0	1
4	367	401	0	2
4	401	460	1	3

- c. Faça o mesmo para o paciente 5, que entrou na coorte no dia 100 e veio a receber seu primeiro tratamento anti-retroviral (monoterapia) no dia 221. Dois meses depois, no dia 281, ele passa para a terapia potente e falece no dia 306.

**Resposta:**

id	inicio	fim	status	terapia
5	100	221	0	0
5	221	281	0	1
5	281	306	1	3

- d. Por fim, acrescente o paciente 6, que entrou na coorte no dia 47, recebendo monoterapia e passando para terapia combinada no dia 105. Esse paciente saiu da coorte (perda de seguimento) no dia 223.

**Resposta:**

id	inicio	fim	status	terapia
6	47	105	0	1
6	105	223	0	2

- e. Considere agora que estes 6 pacientes do banco devem ser analisados como sendo dados prevalentes. Que alteração o banco deveria sofrer?

**Resposta:**

Como o paciente mais antigo é o paciente 1 que iniciou o seu acompanhamento no segundo dia do estudo, a data de sua entrada deveria ser tomada como referência para os demais. Assim ele passaria a ter início no momento zero e todos os demais tempos seriam diminuídos de 2. A tabela passaria a ser:

id	inicio	fim	status	terapia
1	0	122	0	0
1	122	211	0	1
1	211	228	1	3
2	19	208	1	3
3	15	54	0	0
3	54	97	0	1
3	97	152	0	2
3	152	253	1	3
4	59	365	0	1
4	365	399	0	2
4	399	458	1	3
5	98	219	0	0
5	219	279	0	1
5	279	304	1	3
6	45	103	0	1
6	103	221	0	2

- f. Considere ainda que o paciente 2 teve um período afastado do estudo, por exemplo, não fez uso da terapia prescrita por conta própria por 2 meses e que isto ocorreu 2 após 100 dias de acompanhamento. Que alteração o paciente deveria sofrer e como o banco ficaria?

**Resposta:**

O paciente 2 deve ter o seu tempo dividido em três partes para ser possível representar a nova situação: a sua entrada na coorte no dia 19 (se considerarmos que os dados são prevalentes), quando recebe a terapia potente e é observado por 100 dias, ou seja, até o 119º dia. O segundo período que se inicia neste momento e termina quando o paciente retorna ao acompanhamento regular e volta a usar a terapia, após 179 dias do estudo. O terceiro e último período é iniciado aos 179 dias de estudo e termina com a ocorrência do evento (óbito) aos 208 dias do início do estudo e ao final de 29 dias de observação nesta fase.

id	inicio	fim	status	terapia	tempo
1	0	122	0	0	122
1	122	211	0	1	89
1	211	228	1	3	17
2	19	119	0	3	100
2	119	179	0	NA	60
2	179	208	1	3	29
3	15	54	0	0	39
3	54	97	0	1	43
3	97	152	0	2	55
3	152	253	1	3	101
4	59	365	0	1	306
4	365	399	0	2	34
4	399	458	1	3	59

5	98	219	0	0	121
5	219	279	0	1	60
5	279	304	1	3	25
6	45	103	0	1	58
6	103	221	0	2	118

**Exercício 8.3:** O banco tmopc.csv traz os dados do transplante de medula óssea com variáveis tempo dependentes, já estruturado no formato de processo de contagem (veja no Apêndice 12.4 o significado das variáveis).

a. Liste as primeiras 20 linhas.

**Resposta:**

# Exercício 8.3

# item a

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

```
tmopc[1:20, ]
```

```

  id sexo idade status inicio fim deag decr recplaq fasegr
1  1  1   2   31     0     0   9   0   0         0   CP1
2  1  1   2   31     0     9 3527  0   0         1   CP1
3  2  2   2   38     0     0  28   0   0         0   CP1
4  2  2   2   38     1    28  39   1   0         0   CP1
5  3  1   1   23     0     0  27   0   0         0   CP1
6  3  1   1   23     0    27  36   0   0         1   CP1
7  3  1   1   23     0    36 268   1   0         1   CP1
8  3  1   1   23     1   268 434   1   1         1   CP1
9  4  2   2    5     0     0  24   0   0         0   CP1
10 4  2   2    5     1    24  69   1   0         0   CP1
11 5  2   2   15     0     0  22   0   0         0   CP1
12 5  2   2   15     0    22  83   1   0         0   CP1
13 5  2   2   15     0    83 446   1   0         1   CP1
14 5  2   2   15     1   446 672   1   1         1   CP1
15 6  1   1   23     0     0  22   0   0         0   CP1
16 6  1   1   23     1    22  98   1   0         0   CP1
17 7  2   2   11     0     0  32   0   0         0   CP1
18 7  2   2   11     0    32 4025  0   0         1   CP1
19 8  1   1   30     0     0  45   0   0         0   Other
20 8  1   1   30     0    45  278  0   0         1   Other

```

b. Descreva, em palavras, o que aconteceu com os pacientes 5, 6 e 7.

**Resposta:**

- O paciente 5 do sexo feminino com 15 anos de idade, entrou no estudo ao ser transplantado (início=0), desenvolveu a doença enxerto aguda no dia 22 (deag=1); no dia 83 teve recuperação de plaquetas (recplaq=1); no dia 446 desenvolveu doença enxerto crônica e no dia 672 veio a falecer (status=1).
- O paciente 6 do sexo masculino com 23 anos de idade, entrou no estudo ao ser transplantado, no dia 22 desenvolveu a doença enxerto aguda e no dia 98 veio a falecer.
- O paciente 7, entrou no estudo ao ser transplantado, no dia 32 teve recuperação de plaquetas e no dia 4025 quando o estudo terminou ainda estava vivo.

- c. Acrescente linhas nesse banco de dados que correspondem a um paciente que não desenvolveu nenhuma doença enxerto, recuperou as plaquetas 10 dias após o transplante e saiu da coorte no dia 3.789 (perda de seguimento).

**Resposta:**

id	inicio	fim	status	deag	decr	recplaq
c	0	10	0	0	0	0
c	10	3789	0	0	0	1

- d. Faça o mesmo para um paciente que desenvolveu doença enxerto aguda no dia 20 e faleceu 10 dias depois, sem ter recuperado as plaquetas.

**Resposta:**

id	inicio	fim	status	deag	decr	recplaq
d	0	20	0	0	0	0
d	20	30	1	1	0	0

- e. Idem para um paciente que recuperou as plaquetas no dia 8, mas desenvolveu doença enxerto aguda no dia 31, depois desenvolveu a crônica no dia 210 e faleceu no dia 520.

**Resposta:**

id	inicio	fim	status	deag	decr	recplaq
e	0	8	0	0	0	0
e	8	31	0	0	0	1
e	31	210	0	1	0	1
e	210	520	1	1	1	1

**Exercício 8.4:** Continuando a análise iniciada no exercício anterior:

- a. Ajuste quatro modelos explicativos para a sobrevivência após transplante de medula óssea, tratando adequadamente as covariáveis tempo-dependentes (recplaq, deag e decr):

Modelo 1: sobrevivência = idade + sexo

Modelo 2: sobrevivência = idade + sexo + recplaq

Modelo 3: sobrevivência = idade + sexo + recplaq + deag

Modelo 4: sobrevivência = idade + sexo + recplaq + deag + decr

**Resposta:**

```
# EXERCICIO 8.4
```

```
# Modelos de Cox estendido – variáveis mudando no tempo
```

```
# item a
```

```
# Antes transformando as variáveis categóricas numéricas em fator
```

```
tmopc$sex <- factor(tmopc$sexo)
```

```
tmopc$recplaq <- factor(tmopc$recplaq)
```

```
tmopc$deag <- factor(tmopc$deag)
```

```
tmopc$decr <- factor(tmopc$decr)
```

```
library(survival)
```

```
# Ajustando os modelos de Cox estendido
```

```
# Apenas usando o banco modificado – Processo de Contagem
tmo.cox1 <- coxph(Surv(inicio,fim,status)~idade+sexo,data=tmpoc)
tmo.cox1
```

```
Call:
coxph(formula = Surv(inicio, fim, status) ~ idade + sexo, data = tmpoc)
```

	coef	exp(coef)	se(coef)	z	p
idade	-0.023	0.977	0.0132	-1.74	0.082
sexo	-0.378	0.685	0.3041	-1.24	0.210

```
Likelihood ratio test=3.58 on 2 df, p=0.167 n= 259, number of events=
54
```

```
summary(tmo.cox1)
```

```
Call:
coxph(formula = Surv(inicio, fim, status) ~ idade + sexo, data = tmpoc)
```

```
n= 259, number of events= 54
```

	coef	exp(coef)	se(coef)	z	Pr(> z )
idade	-0.02296	0.97730	0.01319	-1.740	0.0818 .
sexo	-0.37841	0.68495	0.30414	-1.244	0.2134

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

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.9773	1.023	0.9524	1.003
sexo	0.6849	1.460	0.3774	1.243

```
Concordance= 0.569 (se = 0.042 )
Rsquare= 0.014 (max possible= 0.823 )
Likelihood ratio test= 3.58 on 2 df, p=0.1672
Wald test = 3.5 on 2 df, p=0.1735
Score (logrank) test = 3.51 on 2 df, p=0.1733
```

```
tmo.cox2 <- coxph(Surv(inicio,fim,status)~idade+sexo+recplaq,data=tmpoc)
summary(tmo.cox2)
```

```
Call:
coxph(formula = Surv(inicio, fim, status) ~ idade + sexo + recplaq,
      data = tmpoc)
```

```
n= 259, number of events= 54
```

	coef	exp(coef)	se(coef)	z	Pr(> z )
idade	-0.01709	0.98306	0.01354	-1.261	0.207
sexo	-0.25477	0.77509	0.30813	-0.827	0.408
recplaq1	-2.22619	0.10794	0.45067	-4.940	7.82e-07 ***

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

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.9831	1.017	0.95731	1.0095
sexo	0.7751	1.290	0.42372	1.4179
recplaq1	0.1079	9.264	0.04462	0.2611

```
Concordance= 0.673 (se = 0.042 )
Rsquare= 0.109 (max possible= 0.823 )
```

Likelihood ratio test= 29.8 on 3 df, p=1.521e-06  
Wald test = 27.56 on 3 df, p=4.501e-06  
Score (logrank) test = 35.81 on 3 df, p=8.212e-08

```
tmo.cox3 <- coxph(Surv(inicio,fim,status)~idade+sexo+recplaq+deag,data=tmopc)
summary(tmo.cox3)
```

Call:

```
coxph(formula = Surv(inicio, fim, status) ~ idade + sexo + recplaq +
      deag, data = tmopc)
```

n= 259, number of events= 54

	coef	exp(coef)	se(coef)	z	Pr(> z )	
idade	-0.01086	0.98920	0.01325	-0.819	0.412560	
sexo	-0.32167	0.72494	0.30824	-1.044	0.296686	
recplaq1	-2.09390	0.12321	0.46919	-4.463	8.09e-06	***
deag1	0.98252	2.67117	0.28685	3.425	0.000614	***

---

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

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.9892	1.0109	0.96385	1.015
sexo	0.7249	1.3794	0.39621	1.326
recplaq1	0.1232	8.1165	0.04912	0.309
deag1	2.6712	0.3744	1.52243	4.687

Concordance= 0.738 (se = 0.042 )

Rsquare= 0.147 (max possible= 0.823 )

Likelihood ratio test= 41.13 on 4 df, p=2.527e-08

Wald test = 37.14 on 4 df, p=1.684e-07

Score (logrank) test = 48.44 on 4 df, p=7.652e-10

```
tmo.cox4 <- coxph(Surv(inicio,fim,status)~idade+sexo+recplaq+deag+decr,data=tmopc)
summary(tmo.cox4)
```

Call:

```
coxph(formula = Surv(inicio, fim, status) ~ idade + sexo + recplaq +
      deag + decr, data = tmopc)
```

n= 259, number of events= 54

	coef	exp(coef)	se(coef)	z	Pr(> z )	
idade	-0.01222	0.98785	0.01342	-0.911	0.362416	
sexo	-0.31706	0.72828	0.30749	-1.031	0.302473	
recplaq1	-2.08912	0.12380	0.46772	-4.467	7.95e-06	***
deag1	0.97622	2.65442	0.28634	3.409	0.000651	***
decr1	0.22134	1.24775	0.39031	0.567	0.570654	

---

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

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.9879	1.0123	0.9622	1.0142
sexo	0.7283	1.3731	0.3986	1.3306
recplaq1	0.1238	8.0778	0.0495	0.3096
deag1	2.6544	0.3767	1.5144	4.6526
decr1	1.2477	0.8014	0.5806	2.6814

Concordance= 0.732 (se = 0.042 )

Rsquare= 0.148 (max possible= 0.823 )

Likelihood ratio test= 41.45 on 5 df, p=7.605e-08



Wald test = 37.61 on 5 df, p=4.512e-07  
Score (logrank) test = 48.8 on 5 df, p=2.44e-09

b. Interprete os parâmetros dos modelos.

**Resposta:**

No primeiro modelo a estimativa do efeito da idade em que realizou o transplante, tem  $p < 0,10$  e pode ser considerado significativamente diferente de zero a este nível. Verifica-se que pelo exponencial da estimativa ter valor menor que 1, a sua interpretação é de que a idade tem um efeito protetor e que, por exemplo, uma pessoa que realizou o transplante 10 anos mais velha que a outra tem a cada dia depois do transplante 30% menos chance de morrer por esta causa.

No segundo modelo ao incluir a informação a respeito da recuperação de plaquetas, o efeito da idade desaparece. Assim, quando controlado por idade e sexo, a estimativa do efeito da recuperação das plaquetas indica um fortíssimo efeito protetor, diminuindo em 90% o risco do paciente que recuperou as plaquetas de morrer a cada unidade de tempo quando comparado a um paciente que não teve esta recuperação.

No terceiro modelo que inclui a informação do desenvolvimento da doença enxerto aguda, a estimativa do parâmetro referente ao efeito da recuperação da plaqueta continua indicando esta covariável como protetora, mas percebe-se que a estimativa do parâmetro referente ao desenvolvimento da doença enxerto aguda indica que, a cada unidade de tempo, este paciente tem quase três vezes mais (2,6) chances de morrer do que os que não desenvolveram esta doença.

No quarto modelo, a informação a respeito do desenvolvimento da doença enxerto crônica não altera as estimativas dos outros parâmetros de forma substancial, nem o efeito desta covariável é significativa (significativamente diferente de um).

c. Compare os quatro modelos utilizando o teste da razão de verossimilhança através do comando `anova()`. Qual modelo deve ser selecionado e porque?

**Resposta:**

```
# item c
# A seleção de covariáveis no modelo e a inferência são idênticas
# ao modelo clássico de Cox: usando a função desvio.
anova(tmo.cox1, tmo.cox2, tmo.cox3, tmo.cox4, test = "Chisq")
```

```
Analysis of Deviance Table
Cox model: response is Surv(inicio, fim, status)
Model 1: ~ idade + sexo
Model 2: ~ idade + sexo + recplaq
Model 3: ~ idade + sexo + recplaq + deag
Model 4: ~ idade + sexo + recplaq + deag + decr
  loglik   Chisq Df P(>|Chi|)
1 -222.18
```

```

2 -209.07 26.2230 1 3.042e-07 ***
3 -203.40 11.3297 1 0.0007628 ***
4 -203.24 0.3223 1 0.5702381
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Ao avaliar a qualidade dos ajustes, a partir da função desvio, verifica-se que a inclusão da variável recuperação de plaquetas (recplaq) fez decrescer a deviance em 26,22 e este decréscimo é significativo. A inclusão da covariável doença enxerto aguda (deag) também melhorou o ajuste de forma significativa, mas a inclusão da variável doença enxerto crônica (decr) não alterou a qualidade do ajuste. Assim, o modelo 3 (tmo.cox3) seria o escolhido entre os avaliados.

- d. Faça o diagnóstico do modelo selecionado, verificando os pressupostos e avaliando os resíduos de Schoenfeld e martingale. O pressuposto de proporcionalidade foi atendido por todas as covariáveis? Há observações mal ajustadas?

**Resposta:**

Os gráficos dos resíduos foram baseados no modelo escolhido no item anterior que é o modelo 3 (tmo.cox3).

```

# Análise de Resíduos do modelo escolhido (modelo 3)
# item d
# Resíduos de Schoenfeld – Análise global do Ajuste
res.sch <- cox.zph(tmo.cox3)
res.sch

```

	rho	chisq	p
idade	-0.0149	0.0106	0.9179
sexo2	-0.2658	4.0015	0.0455
recplaq1	-0.0626	0.2809	0.5961
deag1	-0.0418	0.0870	0.7681
GLOBAL	NA	4.9818	0.2892

Os valores do teste indicam bom ajuste global, e somente a variável sexo seria linearmente correlacionada ao tempo ( $p < 0,05$ ).

```

# Gráficos do Resíduos de Schoenfeld – Proporcionalidade das variáveis
# Conferindo a ordem das variáveis no modelo
tmo.cox3$call

```

```

coxph(formula = Surv(inicio, fim, status) ~ idade + sexo + recplaq +
      deag, data = tmopc)

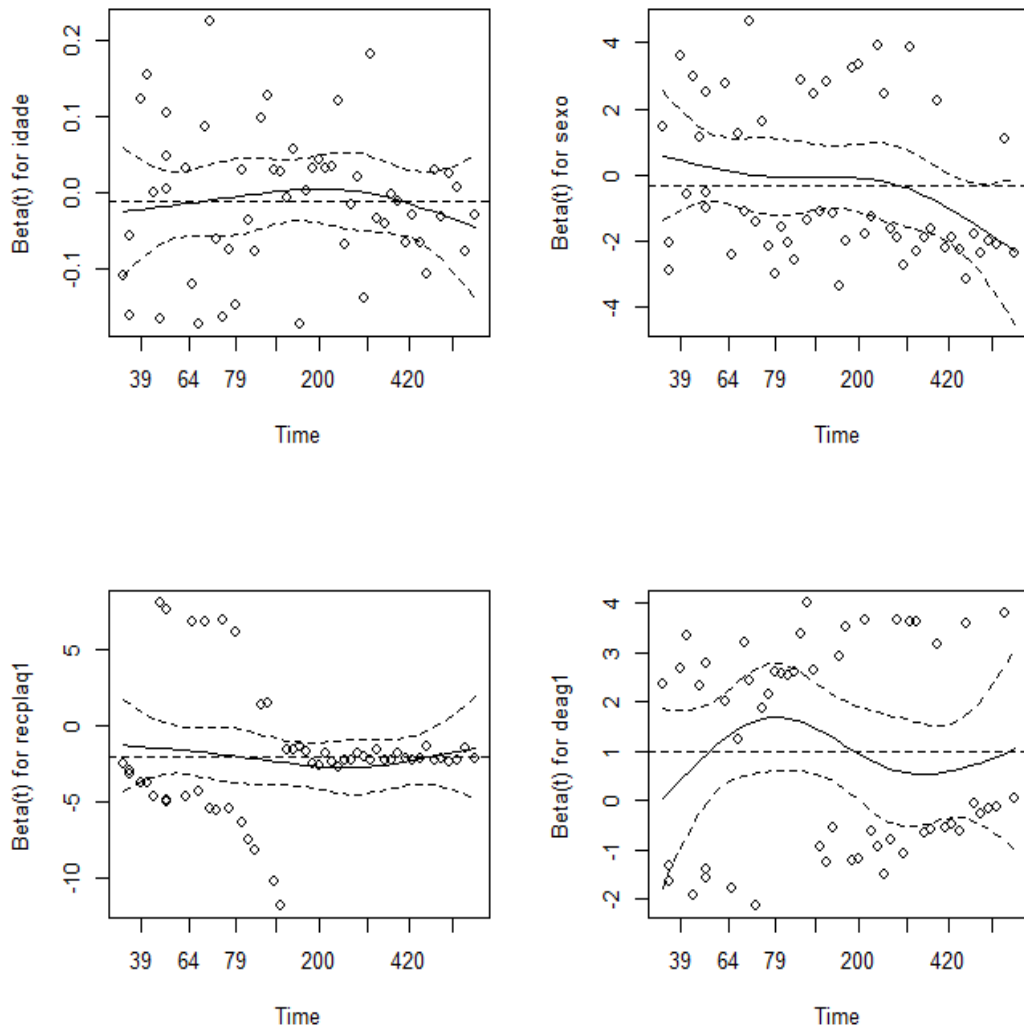
```

```

par (mfrow=c(2,2))
plot(res.sch[1])
abline(h = tmo.cox3$coef[1], lty = 2)
plot(res.sch[2])
abline(h = tmo.cox3$coef[2], lty = 2)
plot(res.sch[3])
abline(h = tmo.cox3$coef[3], lty = 2)
plot(res.sch[4])

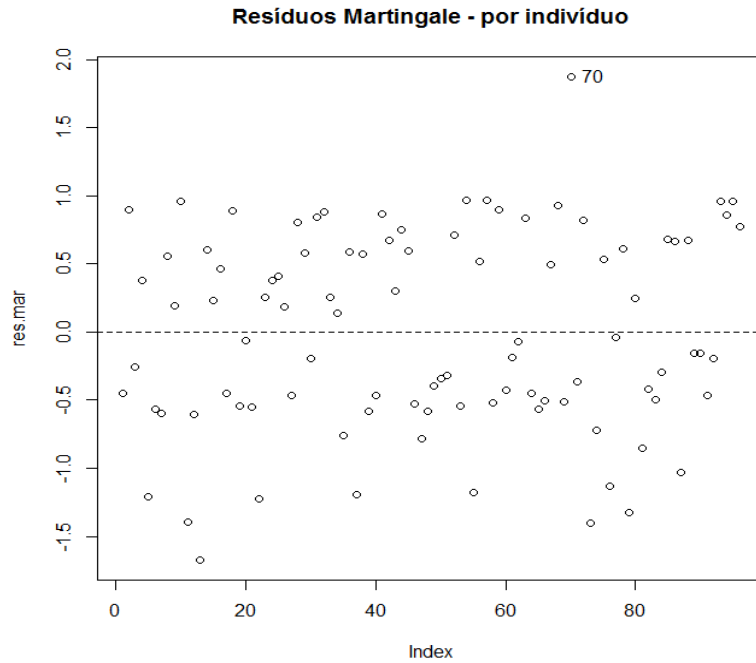
```

```
abline(h = tmo.cox3$coef[4], lty = 2)
```



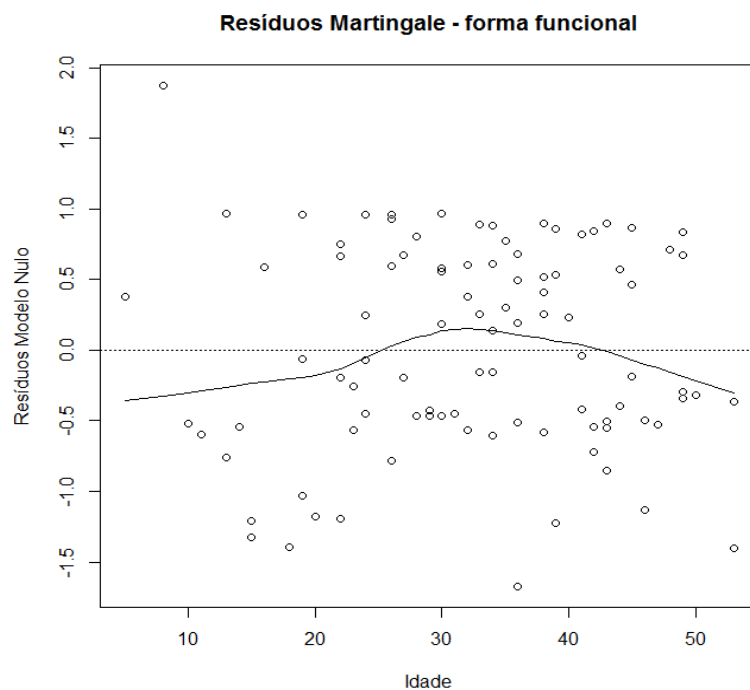
Graficamente, entretanto, nenhuma das covariáveis apresenta padrão de associação significativo com o tempo (os intervalos de confiança todos incluem o valor do parâmetro). Além disso, a curva visível no gráfico é causada por pouquíssimas observações, e não deve ser valorizada.

```
# Resíduos de Martingale - Identificando pontos aberrantes
res.mar <- resid(tmo.cox3, type = "martingale", collapse = tmo$pc$id)
plot(res.mar, main = "Resíduos Martingale - por indivíduo")
abline(h = 0, lty = 2)
identify(res.mar)
```



Os pontos estão homogeneamente distribuídos, se destaca apenas o indivíduo 70 ainda que com resíduo relativamente pequeno ( $< 2,0$ ).

```
# Resíduo Martingale – Avaliando a Forma Funcional
tmo.nulo <- coxph(Surv(inicio, fim, status) ~ 1, data = tmo)
res.nulo <- resid(tmo.cox3, type = "martingale", collapse = tmopc$id)
idade <- tmopc$idade[!duplicated(tmopc$id)]
plot(idade, res.nulo, main = "Resíduos Martingale - forma funcional", xlab =
"Idade", ylab = "Resíduos Modelo Nulo")
lines(lowess(idade, res.nulo))
abline(h=0, lty=3)
```



A única variável contínua para a qual se pode avaliar a forma funcional é a idade. A variação, ainda que presente, é pequena. Se a intensidade fosse maior, pelo formato poderia se investigar a adequação de uma função quadrática da idade. Pode-se considerar nesse caso que não existe associação entre a idade e o tempo.

**Exercício 8.5:** Refaça a análise de sobrevivência em Aids dos pacientes atendidos no HUGG/Unirio que está apresentado no texto (dados no arquivo gafcorr.dat – Apêndice 12.4):

- a. Leia, organize, explore o arquivo de dados, e descreva o que aconteceu com o paciente 35

**Resposta:**

```
# Exercício 8.5
```

```
# item a
```

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

```
names(muda)
```

```
[1] "reg"      "haart"    "ini"      "fim"      "sexo"     "escol"    "status"   [8]
"idade"
```

```
muda[1:20, ]
```

	reg	haart	ini	fim	sexo	escol	status	idade
1	1	S	1752	3297	M	Prim	0	27
2	2	S	1982	3297	F	Gin	0	45
3	3	S	2809	3297	F	Prim	0	41
4	4	N	942	1448	M	Gin	0	38
5	4	S	1448	1939	M	Gin	0	38
6	4	N	1939	1959	M	Gin	0	38
7	4	S	1959	3297	M	Gin	0	38
8	5	N	1031	3297	M	Prim	1	32
9	6	N	1337	1725	F	Univ	0	34
10	6	S	1725	3297	F	Univ	0	34
11	7	S	1808	2136	M	Sec	0	43
12	8	N	1057	1069	M	Prim	1	73
13	9	N	1557	2101	F	Sec	1	36
14	10	N	546	3297	M	Prim	1	31
15	11	N	2162	2988	F	Prim	0	38
16	11	S	2988	3297	F	Prim	0	38
17	12	S	1682	3297	F	Sec	0	32
18	13	N	1064	1666	M	Gin	0	33
19	13	S	1666	3297	M	Gin	0	33
20	14	S	3144	3297	F	Prim	0	45

```
muda[47:54, ]
```

```
> muda[47:54, ]
```

	reg	haart	ini	fim	sexo	escol	status	idade
47	32	N	665	804	F	Prim	1	36
48	33	S	1498	1820	M	Univ	0	76
49	33	S	2400	3297	M	Univ	0	76
50	34	N	686	3200	M	Sec	0	33
<b>51</b>	<b>35</b>	<b>N</b>	<b>769</b>	<b>1577</b>	<b>M</b>	<b>Sec</b>	<b>0</b>	<b>30</b>
<b>52</b>	<b>35</b>	<b>S</b>	<b>1577</b>	<b>1597</b>	<b>M</b>	<b>Sec</b>	<b>1</b>	<b>31</b>
53	36	S	3255	3297	F	Prim	0	52
54	37	N	1203	1341	M	Prim	1	31

**Errata:** o paciente 35 não deveria ter a idade alterada!. Esse erro vem do próprio banco de dados.

```
dim (muda)
[1] 1377      8
```

O paciente 35 é um homem de 30 anos de idade ao iniciar o acompanhamento no 769º dia do estudo, sem uso de HAART até o 1577º dia, sobreviveu por mais 20 dias, vindo a falecer no 1597º dia do estudo. O paciente foi acompanhado por 828 dias.

b. Ajuste um modelo de sobrevivência e analise os resultados encontrados

**Resposta:**

```
# item b
```

```
# Colocando a escolaridade universitária como referencia
```

```
muda$escol <- relevel(muda$escol, "Univ")
```

```
muda.cox <- coxph(Surv(ini, fim, status) ~ haart + idade + escol + sexo, data = muda)
```

```
muda.cox
```

```
summary(muda.cox)
```

```
Call:
```

```
coxph(formula = Surv(ini, fim, status) ~ haart + idade + escol + sexo, data = muda)
```

```
n= 1377, number of events= 184
```

	coef	exp(coef)	se(coef)	z	Pr(> z )	
haarts	-0.777857	0.459389	0.185076	-4.203	2.64e-05	***
idade	0.018461	1.018632	0.007542	2.448	0.0144	*
escolAnalf	-0.234203	0.791201	0.765471	-0.306	0.7596	
escolGin	0.536388	1.709820	0.326883	1.641	0.1008	
escolPrim	0.743847	2.104015	0.310752	2.394	0.0167	*
escolSec	0.326478	1.386078	0.339049	0.963	0.3356	
sexoM	0.225276	1.252668	0.169293	1.331	0.1833	

```
---
```

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

	exp(coef)	exp(-coef)	lower .95	upper .95
haarts	0.4594	2.1768	0.3196	0.6603
idade	1.0186	0.9817	1.0037	1.0338
escolAnalf	0.7912	1.2639	0.1765	3.5470
escolGin	1.7098	0.5849	0.9010	3.2448
escolPrim	2.1040	0.4753	1.1443	3.8687
escolSec	1.3861	0.7215	0.7132	2.6939
sexoM	1.2527	0.7983	0.8989	1.7456

```
Rsquare= 0.025 (max possible= 0.78 )
```

```
Likelihood ratio test= 35.07 on 7 df, p=1.083e-05
```

```
Wald test = 33.64 on 7 df, p=2.014e-05
```

```
Score (logrank) test = 34.43 on 7 df, p=1.433e-05
```

- A terapia HAART reduz significativamente o risco de óbito a menos da metade (54%);
- A cada ano a mais de idade no momento de diagnóstico de Aids o risco de morrer aumenta em 0,19%;
- Comparando com o grau de instrução superior (Univ), o analfabetismo não implica aumento significativo de risco ( $p > 0,07$ ), possivelmente por

serem apenas 20 pacientes entre os 963 (30 registros ao todo); ter apenas escolaridade até a quarta série do ensino fundamental dobra (2,104) o risco de morte por Aids significativamente quando comparado com quem tem escolaridade superior; mas não há diferença significativa entre nível universitário e nível secundário ( $p>0,30$ );

- Não há diferença de gênero no tempo de sobrevivência ( $p>0,10$ ).
- Observando a saída do summary podemos construir a tabela de resultados abaixo:

Variável	HR ajustado	p
Haart (ref sim)	0.459 [IC95% 0.320 - 0,660]	<0,0001
Idade (anos)	1,019 [IC95% 1,004; 1,034]	<0,01
Escolaridade (ref univer.)		
Analfabeto	0,791 [IC95% 0,176 – 3,547]	$p>0,75$
Primário	2,104 [IC95% 1,144 - 3,869]	$p<0,05$
Ginásio	1,710 [IC95% 0,901 - 3,245]	$p>0,10$
Secundário	1,386 [IC95% 0,713 - 2,694]	$p>0,33$
Sexo (ref masculino)	1,253 [IC95% 0,899; 1,746]	$p>0,18$

O modelo apresentou um Rsquare igual a 0.025, que em relação ao R máximo (max possible= 0.78) representa 0,064 (0,025/0,78), ou seja, o modelo ajustado explica apenas 6,4% da variação observada.

c. Faça o diagnóstico do modelo

**Resposta:**

# item c

```
muda.sch <- cox.zph(muda.cox)
```

```
muda.sch
```

```

          rho chisq    p
haartS   -0.0704 1.170 0.279
idade     0.0439 0.380 0.537
escolAnalf -0.0697 0.899 0.343
escolGin  -0.0978 1.796 0.180
escolPrim -0.0348 0.233 0.630
escolSec  -0.0791 1.172 0.279
sexoM     0.0435 0.388 0.534
GLOBAL    NA 5.877 0.554

```

```
plot(muda.sch[1], cex.lab = 1.5)
```

```
abline(h = muda.cox$coefficients[1], lty = 2)
```

```
plot(muda.sch[2], cex.lab = 1.5)
```

```
abline(h = muda.cox$coefficients[2], lty = 2)
```

```
plot(muda.sch[3], cex.lab = 1.5)
```

```
abline(h = muda.cox$coefficients[3], lty = 2)
```

```
plot(muda.sch[4], cex.lab = 1.5)
```

```
abline(h = muda.cox$coefficients[4], lty = 2)
```

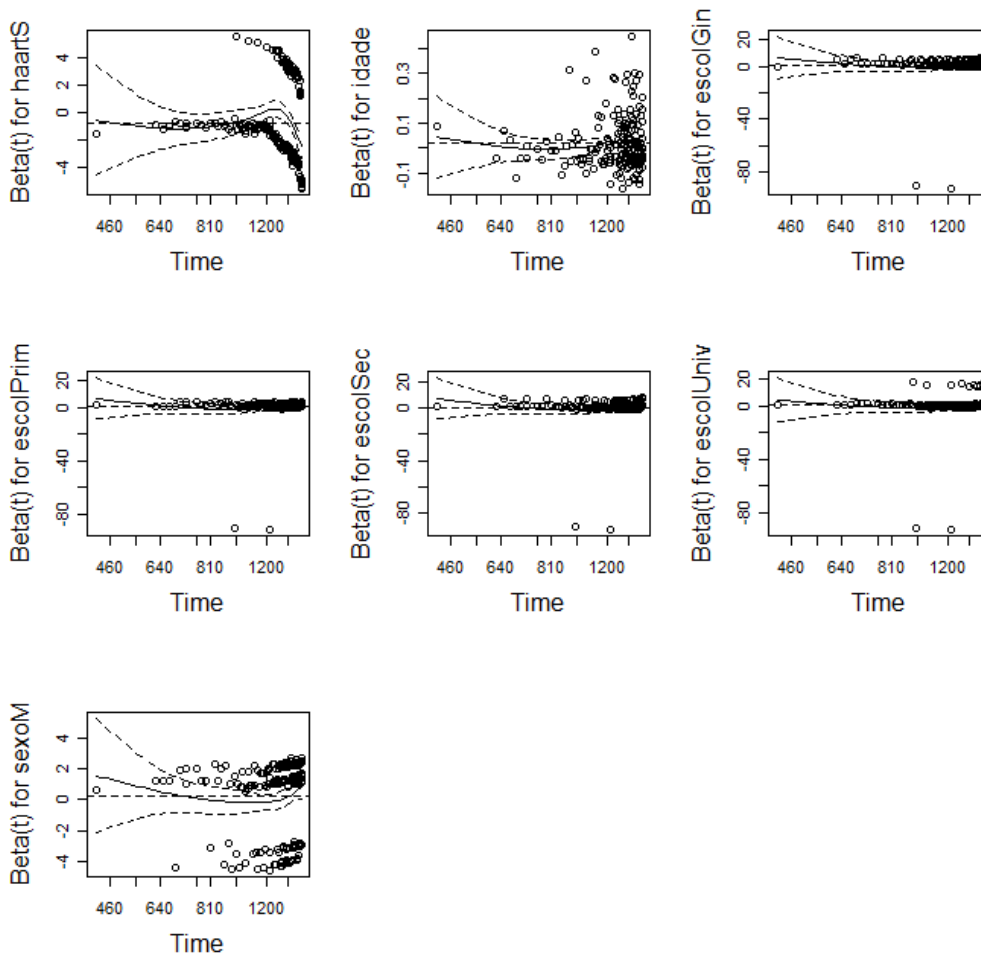
```
plot(muda.sch[5], cex.lab = 1.5)
```

```
abline(h = muda.cox$coefficients[5], lty = 2)
```

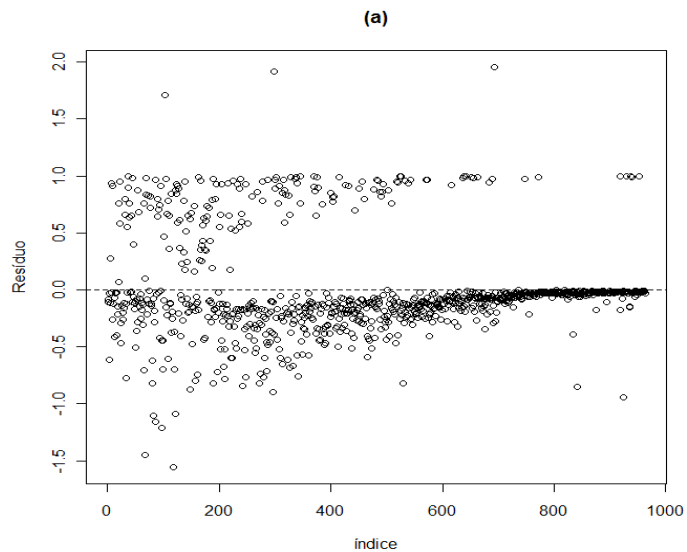
```
plot(muda.sch[6], cex.lab = 1.5)
```

```
abline(h = muda.cox$coefficients[6], lty = 2)
```

```
plot(muda.sch[7], cex.lab = 1.5)
abline(h = muda.cox$coefficients[7], lty = 2)
```



```
muda.mar <- resid(muda.cox, type = "martingale", collapse = muda$reg)
plot(muda.mar, xlab = "índice", ylab = "Resíduo", main = "(a)")
abline(h = 0, lty = 2)
```

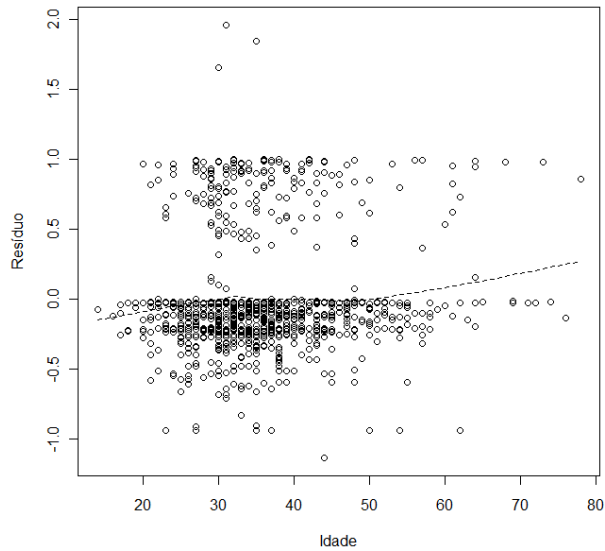




```

muda.nulo <- coxph(Surv(ini, fim, status) ~ 1, data = muda)
res.mar <- resid(muda.nulo, type = "martingale", collapse = muda$reg)
idade <- muda$idade[!duplicated(muda$reg)]
plot(idade, res.mar, xlab = "Idade", ylab = "Resíduo")
lines(lowess(idade, res.mar, iter = 0), lty = 2)

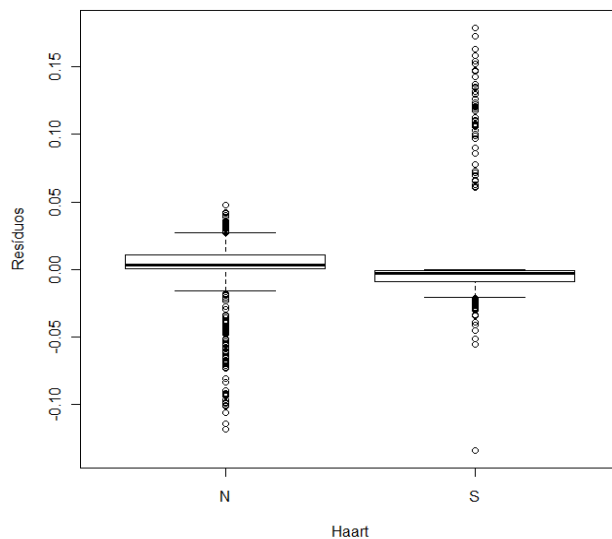
```



```

muda.score <- resid(muda.cox, type = "dfbetas", collapse = muda$reg)
mudaobs.score <- resid(muda.cox, type = "dfbetas")
indice <- !duplicated(muda$reg)
par(mfrow=c(1,1))
plot(muda$haart, mudaobs.score[, 1], xlab = "Haart", ylab = "Resíduos")

```



```

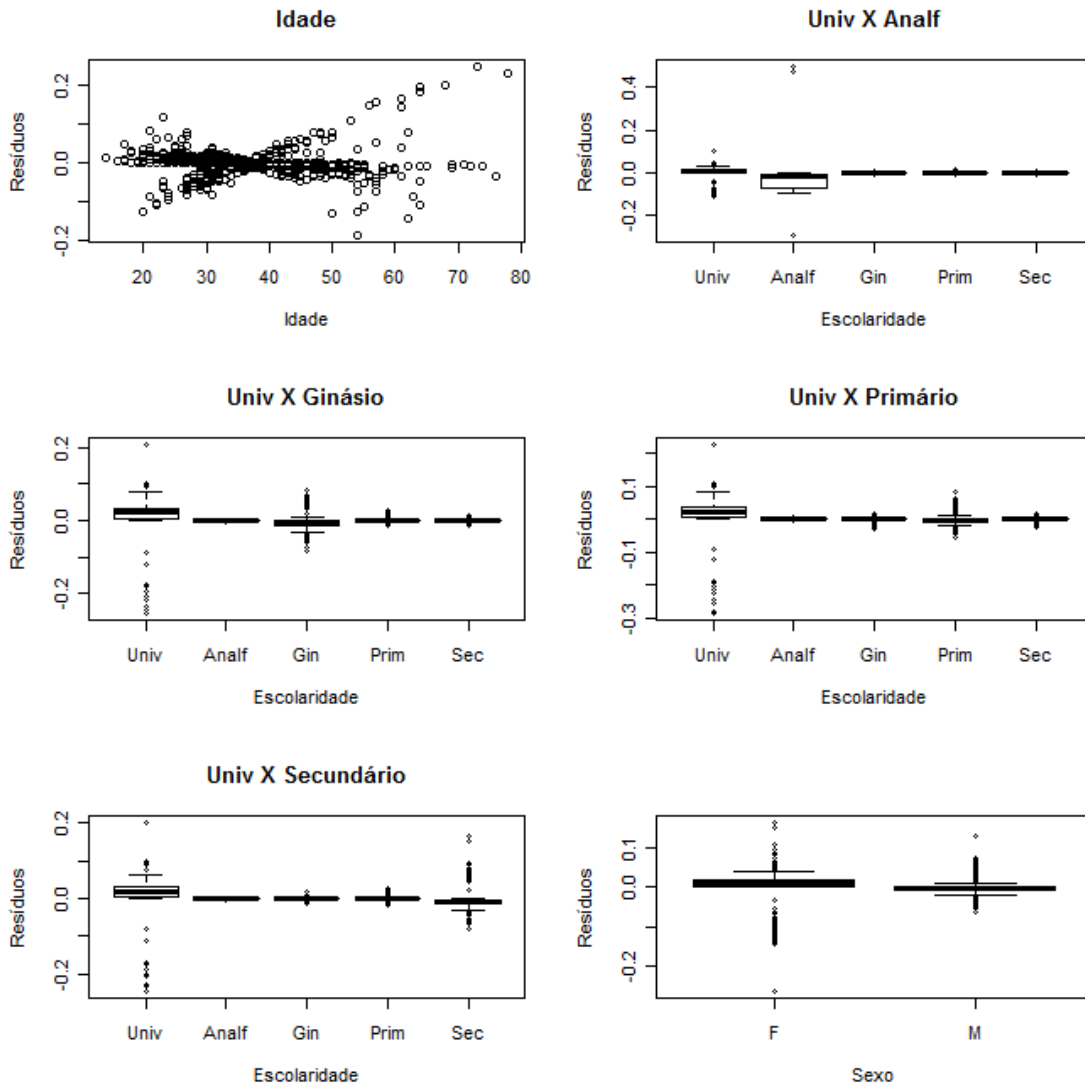
plot(muda$idade[indice], muda.score[, 2], xlab = "Idade", ylab = "Resíduos",
main = "Idade")
plot(muda$escol[indice], muda.score[, 3], xlab = "Escolaridade", ylab =
"Resíduos", main = "Univ X Analf")

```

```

plot(muda$escol[indice], muda.score[, 4], xlab = "Escolaridade", ylab =
"Resíduos", main = "Univ X Ginásio")
plot(muda$escol[indice], muda.score[, 5], xlab = "Escolaridade", ylab =
"Resíduos", main = "Univ X Primário")
plot(muda$escol[indice], muda.score[, 6], xlab = "Escolaridade", ylab =
"Resíduos", main = "Univ X Secundário")
plot(muda$sexo[indice], muda.score[, 7], xlab = "Sexo", ylab = "Resíduos")

```



Observando os resíduos de Schoenfeld do modelo ajustado, nota-se que apenas o tratamento, única variável a mudar no tempo, não atende a premissa de proporcionalidade. Os resíduos de Schoenfeld das demais variáveis sugerem proporcionalidade dos riscos. Essa observação é confirmada testando-se a hipótese nula de correlação entre as variáveis e o tempo.