

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

9

Covariáveis tempo-dependentes

Exercícios

Recomenda-se que o leitor revise o conceito de processo de contagem e sua representação gráfica, apresentados no Capítulo 2.

Exercício 9.1: 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 paciente portador de Aids. A tabela abaixo mostra um trecho do banco de dados que resultaria de um estudo como este, 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 paciente ao longo do período de acompanhamento.

Resposta:

- O paciente 1 entrou na coorte no dia 25, no dia 324 seu **CD4** estava abaixo de 200 e no dia 768 ele morreu.
- O paciente 2 entrou na coorte no dia 39, no dia 131 seu **CD4** estava abaixo de 200 e no dia 345 ele deixou de ser acompanhado, ou o estudo terminou.
- O paciente 3 entrou na coorte no dia 1, com **CD4** acima de 200. Entre os dias 56 e 130 ele não foi acompanhado e no dia 131 seu **CD4** estava abaixo de 200. No dia 145 ele morreu.

Exercício 9.2: 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), poder-se-ia 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

1. 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.

2. Acrescente linhas à 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

3. 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

4. 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

Exercício 9.3: Abra o arquivo *tmopc.csv* e liste as primeiras 20 linhas. Veja no Apêndice C.5 o significado das variáveis.

```
> tmo <- read.table("tmopc.csv", header = T, sep = ";")
> tmo[1:20, ]
```

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

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

Resposta:

- O paciente 5, entrou no estudo ao ser transplantado, no dia 22 desenvolveu a doença enxerto aguda; no dia 83 teve recuperação de plaquetas; no dia 446 desenvolveu doença enxerto crônica e no dia 672 veio a falecer.
- O paciente 6, 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.

2. Construa as linhas deste 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:

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

3. Construa as linhas para um paciente que desenvolveu doença enxerto aguda no dia 20 e faleceu 10 dias depois, sem ter recuperado as plaquetas.

Resposta:

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

4. Construa também as linhas para o 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:

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

Exercício 9.4: Ajuste quatro modelos causais para a sobrevida após transplante de medula óssea, usando covariáveis tempo-dependentes (*recplaq*, *deag* e *decr*):

Modelo 1: sobrevida = idade + sexo

Modelo 2: sobrevida = idade + sexo + recplaq

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

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

```
> tmo <- read.table("tmopc.csv", header = T, sep = ";")
> tmo$sex <- factor(tmo$sexo)
> tmo$recplaq <- factor(tmo$recplaq)
> tmo$deag <- factor(tmo$deag)
> tmo$decr <- factor(tmo$decr)
> tmo.cox1 <- coxph(Surv(inicio, fim, status) ~ idade + sexo, data = tmo)
> summary(tmo.cox1)
```

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

```
n= 259
      coef exp(coef) se(coef)      z      p
idade -0.0230    0.977   0.0132 -1.74 0.082
sexo  -0.3784    0.685   0.3041 -1.24 0.210

      exp(coef) exp(-coef) lower .95 upper .95
idade    0.977      1.02    0.952    1.00
sexo     0.685      1.46    0.377    1.24
```

Rsquare= 0.014 (max possible= 0.823)
Likelihood ratio test= 3.58 on 2 df, p=0.167
Wald test = 3.5 on 2 df, p=0.174
Score (logrank) test = 3.51 on 2 df, p=0.173

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

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

```
n= 259
      coef exp(coef) se(coef)      z      p
idade  -0.0171    0.983   0.0135 -1.261 2.1e-01
sexo   -0.2548    0.775   0.3081 -0.827 4.1e-01
recplaq1 -2.2262    0.108   0.4507 -4.940 7.8e-07

      exp(coef) exp(-coef) lower .95 upper .95
idade    0.983      1.02    0.9573    1.010
sexo     0.775      1.29    0.4237    1.418
recplaq1 0.108      9.26    0.0446    0.261
```

Rsquare= 0.109 (max possible= 0.823)
Likelihood ratio test= 29.8 on 3 df, p=1.52e-06
Wald test = 27.6 on 3 df, p=4.5e-06
Score (logrank) test = 35.8 on 3 df, p=8.21e-08

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

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

```
deag, data = tmo)
```

```
n= 259
```

	coef	exp(coef)	se(coef)	z	p
idade	-0.0109	0.989	0.0132	-0.82	4.1e-01
sexo	-0.3217	0.725	0.3082	-1.04	3.0e-01
recplaq1	-2.0939	0.123	0.4692	-4.46	8.1e-06
deag1	0.9825	2.671	0.2868	3.43	6.1e-04

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.989	1.011	0.9638	1.015
sexo	0.725	1.379	0.3962	1.326
recplaq1	0.123	8.116	0.0491	0.309
deag1	2.671	0.374	1.5224	4.687

```
Rsquare= 0.147 (max possible= 0.823 )
```

```
Likelihood ratio test= 41.1 on 4 df, p=2.53e-08
```

```
Wald test = 37.1 on 4 df, p=1.68e-07
```

```
Score (logrank) test = 48.4 on 4 df, p=7.65e-10
```

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

```
Call:
```

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

```
n= 259
```

	coef	exp(coef)	se(coef)	z	p
idade	-0.0122	0.988	0.0134	-0.911	3.6e-01
sexo	-0.3171	0.728	0.3075	-1.031	3.0e-01
recplaq1	-2.0891	0.124	0.4677	-4.467	7.9e-06
deag1	0.9762	2.654	0.2863	3.409	6.5e-04
decr1	0.2213	1.248	0.3903	0.567	5.7e-01

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	0.988	1.012	0.9622	1.014
sexo	0.728	1.373	0.3986	1.331
recplaq1	0.124	8.078	0.0495	0.310
deag1	2.654	0.377	1.5144	4.653
decr1	1.248	0.801	0.5806	2.681

```
Rsquare= 0.148 (max possible= 0.823 )
```

```
Likelihood ratio test= 41.5 on 5 df, p=7.6e-08
```

```
Wald test = 37.6 on 5 df, p=4.51e-07
```

```
Score (logrank) test = 48.8 on 5 df, p=2.44e-09
```


1. 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.

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 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 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 significativamente diferente de um.

2. Avalie a qualidade do ajuste global dos modelos, utilizando a análise de *deviance*

```
> anova(tmo.cox1, tmo.cox2, tmo.cox3, tmo.cox4, test = "Chisq")
```

Analysis of Deviance Table

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

Model 2: Surv(inicio, fim, status) ~ idade + sexo + recplaq

Model 3: Surv(inicio, fim, status) ~ idade + sexo + recplaq + deag

Model 4: Surv(inicio, fim, status) ~ idade + sexo + recplaq + deag + decr

	Resid. Df	Resid. Dev	Df	Deviance	P(> Chi)
1	257	444.36			
2	256	418.13	1	26.22	3.042e-07
3	255	406.81	1	11.33	7.628e-04
4	254	406.48	1	0.32	0.57

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

3. Faça a análise dos resíduos, avaliando os resíduos de Schoenfeld e martingale. O pressuposto de proporcionalidade foi atendido por todas as covariáveis? Há observações mal ajustadas? Nos comandos abaixo, substitua a palavra **mod** pelo modelo que você escolheu no item anterior.

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

Resíduos de Schoenfeld

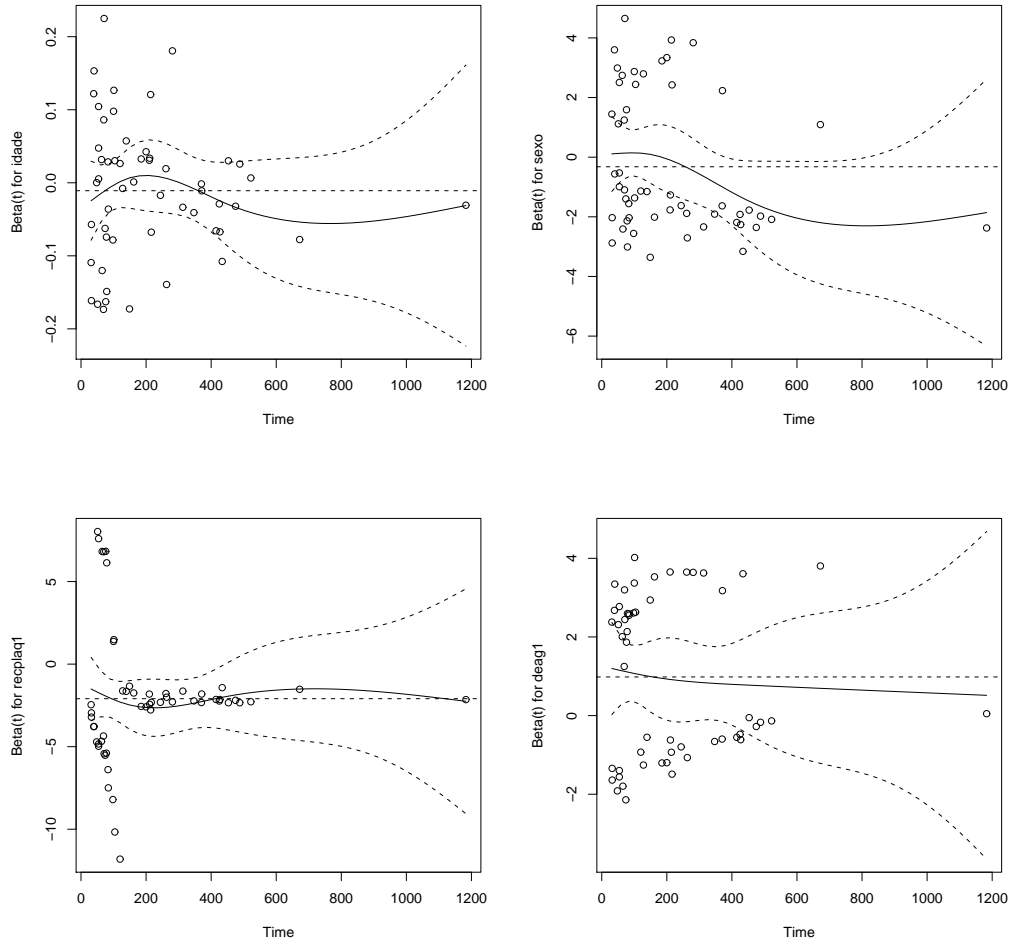
```
> res.sch <- cox.zph(tmo.cox3)
> res.sch
```

	rho	chisq	p
idade	-0.0655	0.2060	0.650
sexo	-0.2628	3.9107	0.048
recplaq1	-0.0157	0.0178	0.894
deag1	-0.0750	0.2797	0.597
GLOBAL	NA	4.5039	0.342

Os valores do teste indicam bom ajuste global, e somente a variável **sexo** seria linearmente correlacionada ao tempo.

```
> 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.

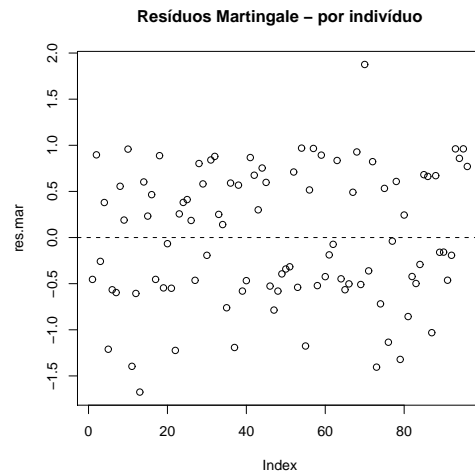


Resíduos de Martingale

```

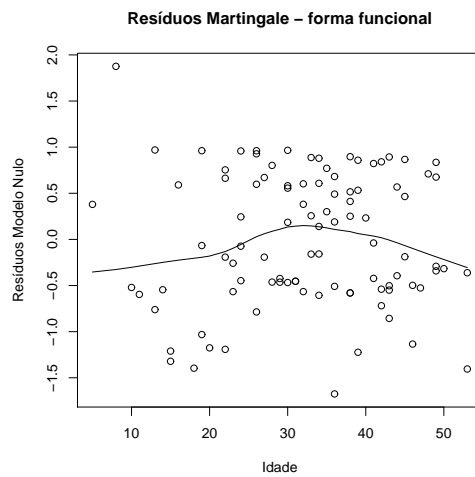
> res.mar <- resid(tmo.cox3, type = "martingale", collapse = tmo$id)
> plot(res.mar, main = "Resíduos Martingale - por indivíduo")
> abline(h = 0, lty = 2)

```



Pontos homogeneamente distribuidos.

```
> tmo.nulo <- coxph(Surv(inicio, fim, status) ~ 1, data = tmo)
> res.nulo <- resid(tmo.cox3, type = "martingale", collapse = tmo$id)
> idade <- tmo$idade[!duplicated(tmo$id)]
> plot(idade, res.nulo, main = "Resíduos Martingale - forma funcional",
+       xlab = "Idade", ylab = "Resíduos Modelo Nulo")
> lines(lowess(idade, res.nulo))
```



A única variável contínua para a qual pode-se avaliar a forma funcional é a idade. A variação, ainda que presente, é de pouca monta. Se a intensidade fosse maior, pelo formato poderia se investigar uma função quadrática da idade.

Exercício 9.5: Refaça a análise de sobrevida em Aids apresentada no texto (dados no arquivo *gafcorr.dat*).

Como este exemplo está bem discutido no texto, nos limitaremos aqui a listar comandos e saídas para que o leitor possa comparar com suas saídas.

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

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

> muda$escol <- relevel(muda$escol, "Univ")
> muda.cox <- coxph(Surv(ini, fim, censura) ~ haart + idade + escol +
+   sexo, data = muda)
> muda.cox
```

Call:

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

	coef	exp(coef)	se(coef)	z	p
haartS	-0.7779	0.459	0.18508	-4.203	2.6e-05
idade	0.0185	1.019	0.00754	2.448	1.4e-02
escolAnalf	-0.2342	0.791	0.76547	-0.306	7.6e-01
escolGin	0.5364	1.710	0.32688	1.641	1.0e-01
escolPrim	0.7438	2.104	0.31075	2.394	1.7e-02
escolSec	0.3265	1.386	0.33905	0.963	3.4e-01
sexoM	0.2253	1.253	0.16929	1.331	1.8e-01

Likelihood ratio test=35.1 on 7 df, p=1.08e-05 n= 1377

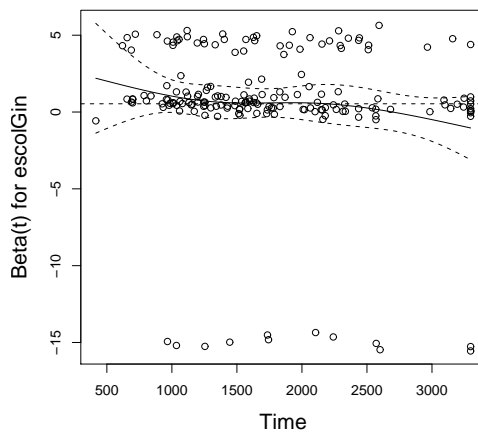
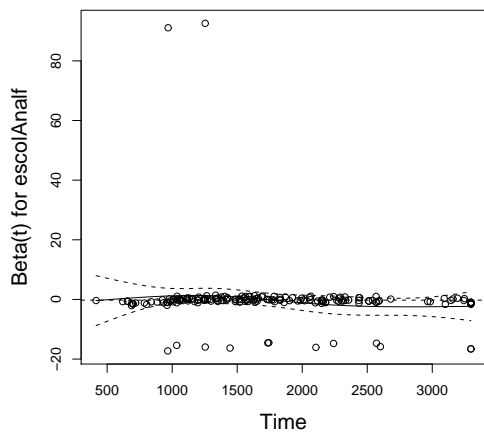
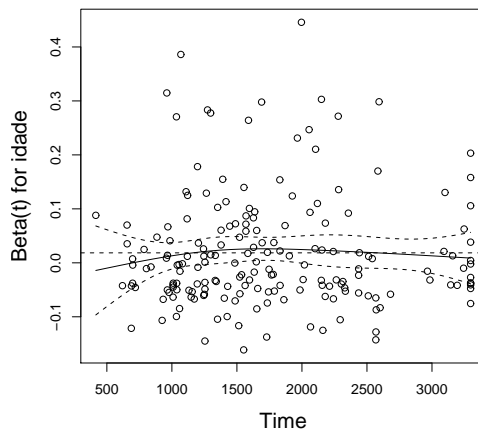
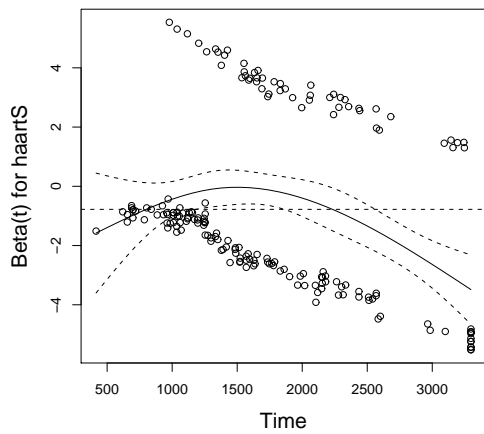
```
> muda.sch <- cox.zph(muda.cox)
> muda.sch
```

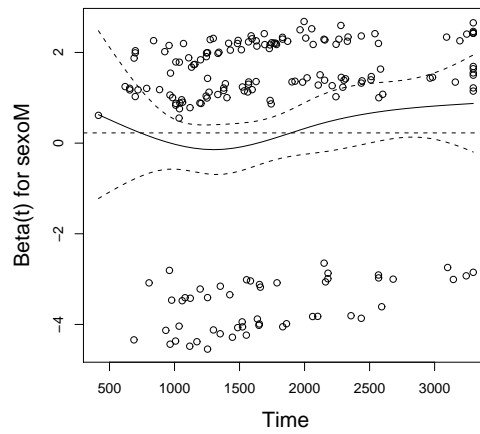
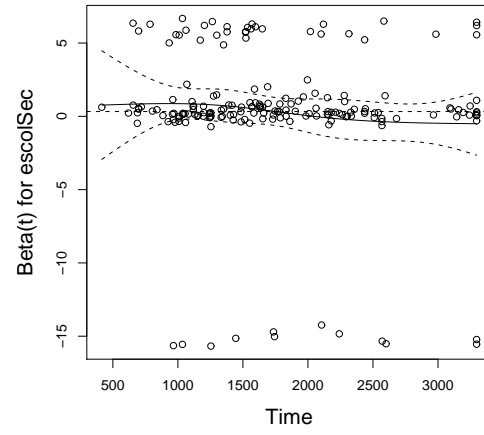
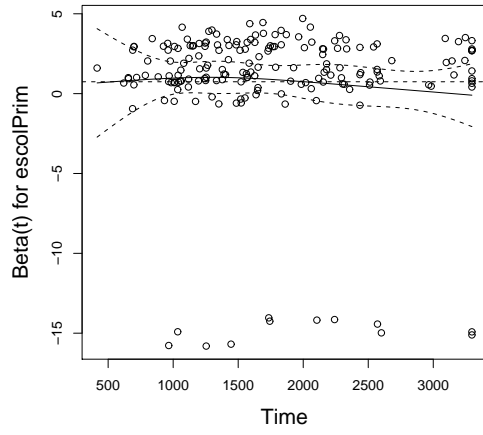
	rho	chisq	p
haartS	-0.26583	16.70605	4.36e-05
idade	0.00627	0.00775	9.30e-01
escolAnalf	-0.12455	2.86745	9.04e-02
escolGin	-0.12721	3.03844	8.13e-02
escolPrim	-0.07071	0.96321	3.26e-01
escolSec	-0.10421	2.03111	1.54e-01
sexoM	0.12002	2.94786	8.60e-02
GLOBAL	NA	24.41845	9.62e-04

```

> 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)

```

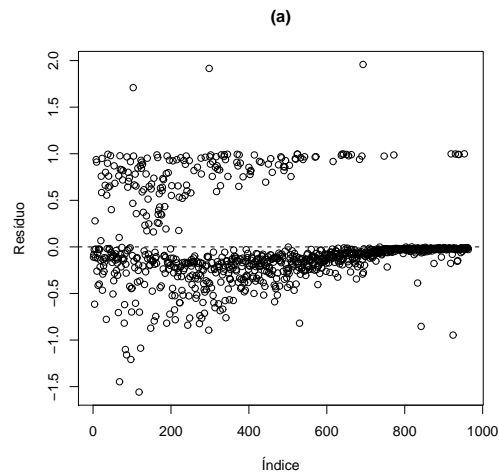




```

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

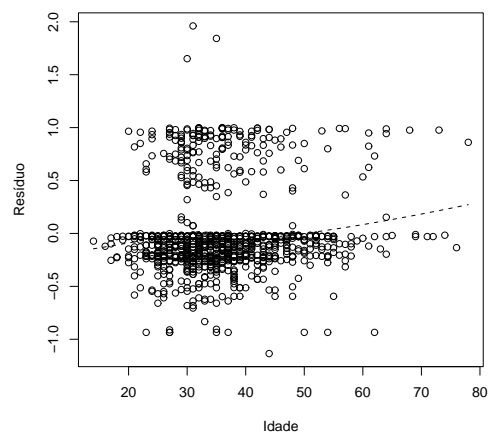
```



```

> muda.nulo <- coxph(Surv(ini, fim, censura) ~ 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 = "Residuo")
> 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)
> plot(muda$haart, mudaobs.score[, 1], xlab = "Haart", ylab = "Resíduos")
> plot(muda$idade[indice], muda.score[, 2], xlab = "Idade", ylab = "Resíduos")
> plot(muda$escol[indice], muda.score[, 3], xlab = "Escolaridade",

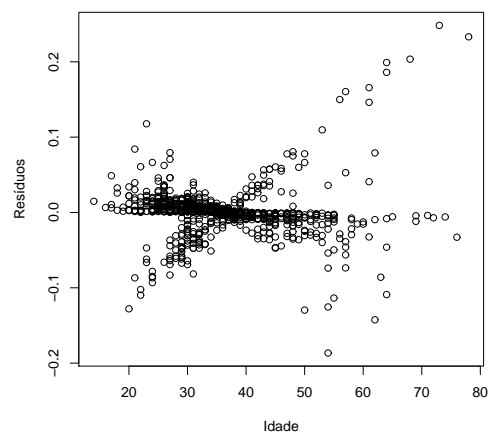
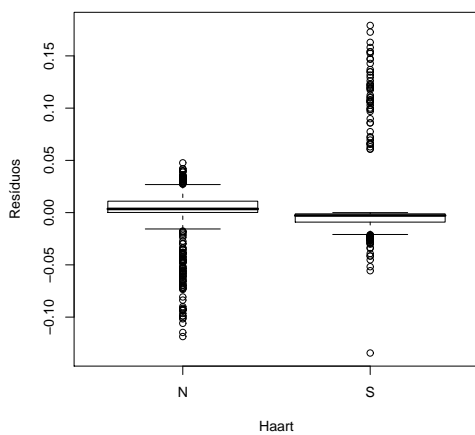
```



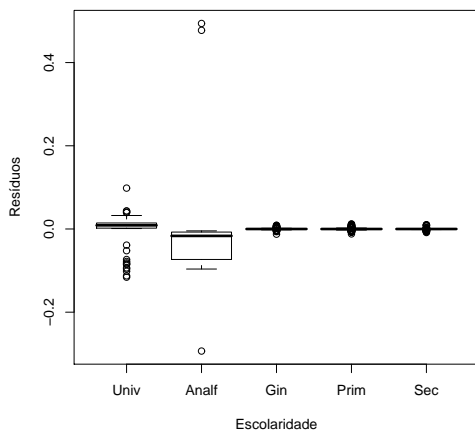
```

+   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")

```



Univ X Analf



Univ X Ginásio

