

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

8

Análise de resíduos para modelos de Cox

Exercícios

Neste capítulo, serão feitas as análises dos resíduos dos modelos ajustados no Capítulo 7. Por isso, é necessário primeiro realizar os exercícios do capítulo anterior.

Exercício 8.1: Encontramos, ao ajustar o modelo de Cox aos dados de transplante de medula óssea (banco de dados *tmoclas.dat*), que o melhor modelo explicativo da sobrevivência incluía as covariáveis *idade*, *sexo*, *fase*, *deag* e *decr*. Esse é o modelo que serve de exemplo neste capítulo. Refaça, no R, a análise de resíduos apresentada ao longo do texto.

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

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

> tmo$sexo <- factor(tmo$sexo)
> tmo$decr <- factor(tmo$decr)
> tmo$deag <- factor(tmo$deag)
> tmo$fase <- factor(tmo$fase)
> mod4 <- coxph(Surv(os, status) ~ idade + sexo + fase + deag +
+   decr, data = tmo)
> mod4
```

Call:

```
coxph(formula = Surv(os, status) ~ idade + sexo + fase + deag +  
      decr, data = tmo)
```

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

Likelihood ratio test=39.9 on 6 df, p=4.8e-07 n= 96

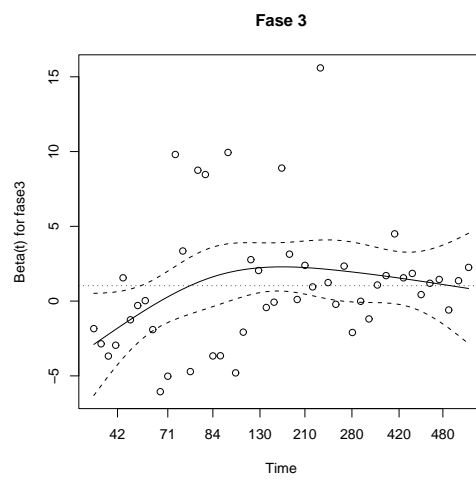
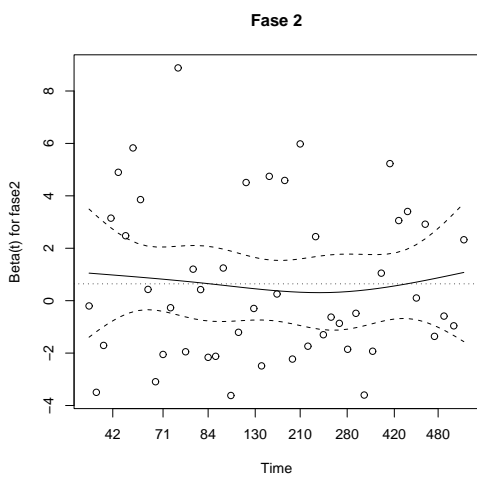
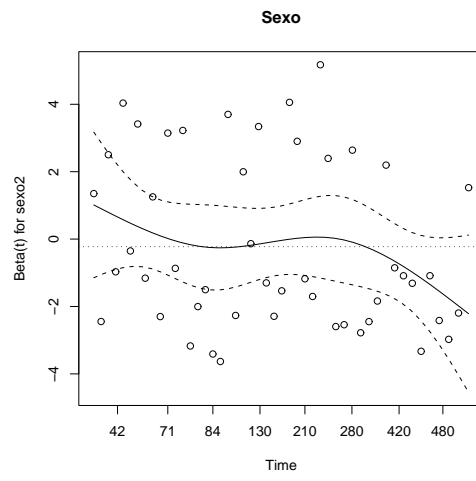
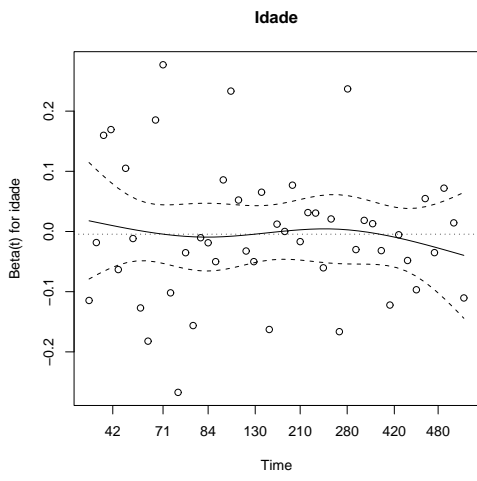
Calculando os resíduos de Schoenfeld

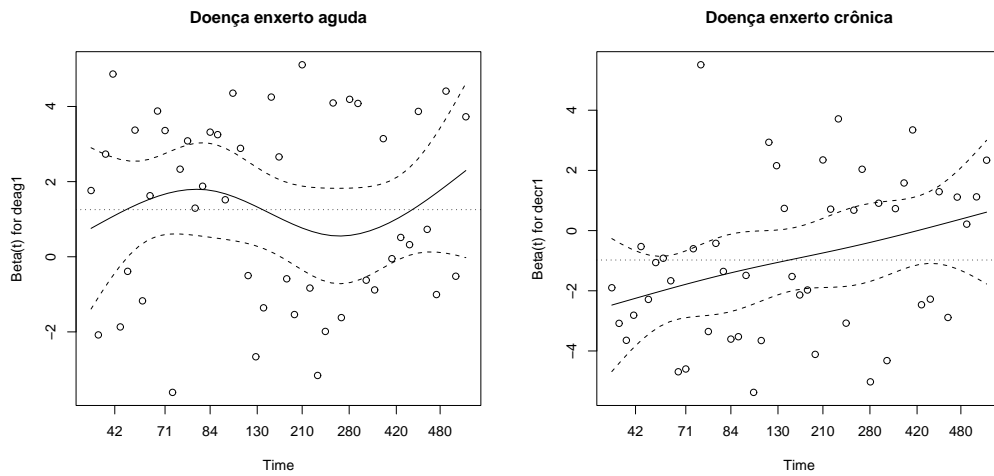
```
> zph <- cox.zph(mod4)  
> zph
```

	rho	chisq	p
idade	-0.0674	0.2547	0.6138
sexo2	-0.2260	2.8393	0.0920
fase2	-0.0317	0.0617	0.8039
fase3	0.2063	2.8416	0.0919
deag1	-0.0147	0.0117	0.9137
decr1	0.3341	6.4078	0.0114
GLOBAL	NA	13.1118	0.0413

Gráfico dos resíduos Schoenfeld

```
> par(mfrow = c(2, 3))  
> plot(zph[1], main = "Idade")  
> abline(h = mod4$coef[1], lty = 3)  
> plot(zph[2], main = "Sexo")  
> abline(h = mod4$coef[2], lty = 3)  
> plot(zph[3], main = "fase 2")  
> abline(h = mod4$coef[3], lty = 3)  
> plot(zph[4], , main = "Fase 3")  
> abline(h = mod4$coef[4], lty = 3)  
> plot(zph[5], main = "Doença enxerto aguda")  
> abline(h = mod4$coef[5], lty = 3)  
> plot(zph[6], main = "Doença enxerto crônica")  
> abline(h = mod4$coef[6], lty = 3)
```



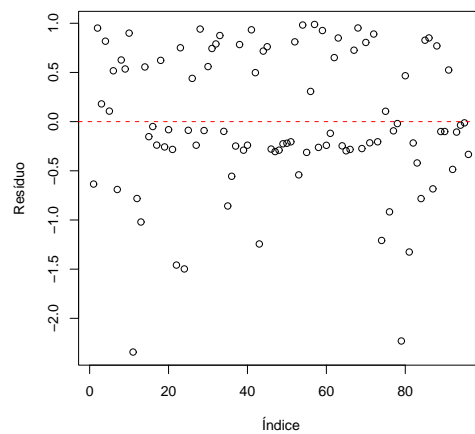


Resíduos de Martingale

```
> mod4.mar <- resid(mod4, "martingale")
```

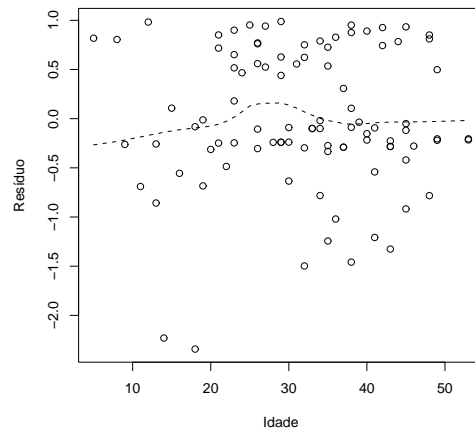
Identificando os pontos mal ajustados através do gráfico dos resíduos de Martingale x indivíduos

```
> plot(mod4.mar, xlab = "Índice", ylab = "Resíduo")
> abline(h = 0, col = 2, lty = 2)
```



Analisando a forma funcional através do gráfico dos resíduos de Martingale x idade

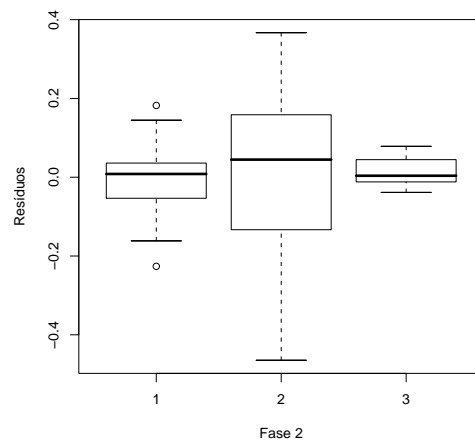
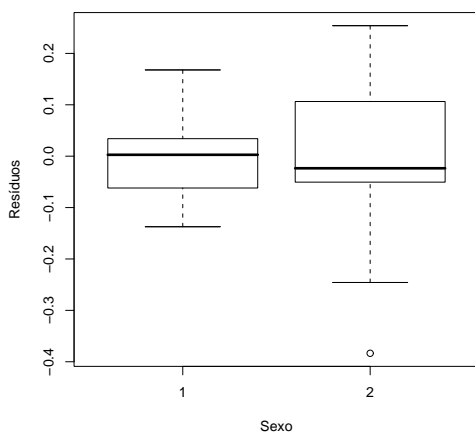
```
> plot(tmo$idade, mod4.mar, xlab = "Idade", ylab = "Resíduo")
> lines(lowess(tmo$idade, mod4.mar, iter = 0), lty = 2)
```

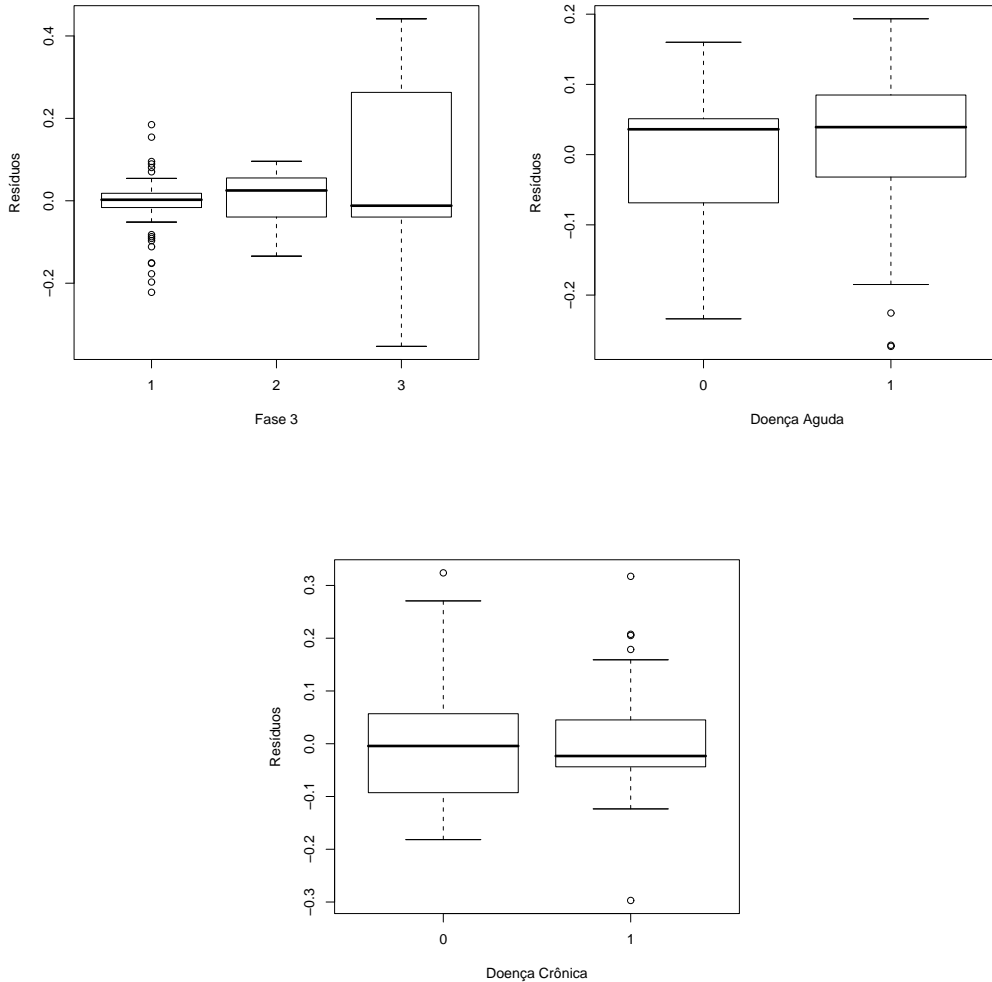


Gráficos dos Resíduos scores

```
> mod4.sco2 <- resid(mod4, type = "dfbetas")

> par(mfrow = c(3, 2))
> plot(tmo$sexo, mod4.sco2[, 2], xlab = "Sexo", ylab = "Resíduos")
> plot(tmo$fase, mod4.sco2[, 3], xlab = "Fase 2", ylab = "Resíduos")
> plot(tmo$fase, mod4.sco2[, 4], xlab = "Fase 3", ylab = "Resíduos")
> plot(tmo$deag, mod4.sco2[, 5], xlab = "Doença Aguda", ylab = "Resíduos")
> plot(tmo$decr, mod4.sco2[, 6], xlab = "Doença Crônica", ylab = "Resíduos")
```





Exercício 8.2: No estudo de sobrevida de pacientes em diálise (exercício 7.2), encontramos que o modelo contendo variáveis demográficas, clínicas e ambientais (*idade*, *cdiab*, *crim*, *congenita* e *grande*) foi o que melhor se ajustou aos dados. Pela análise visual do gráfico de Kaplan-Meier realizada no exercício, todas as variáveis pareciam atender ao pressuposto de Cox. A única que parecia levantar dúvidas era a variável *congenita*. Vamos reavaliar essas variáveis, agora utilizando os resíduos de Schoenfeld, fazendo uma análise visual dos resíduos e calculando o teste da correlação.

```
> dialise <- read.table("dialise.csv", header = T, sep = ",")
> y <- Surv(dialise$tempo, dialise$status)
```

```

> modeloIII <- coxph(y ~ idade + cdiab + crim + congenita + grande,
+   data = dialise)
> zph <- cox.zph(modeloIII)
> zph

```

	rho	chisq	p
idade	0.05367	5.0743	0.02428
cdiab	0.04419	3.1462	0.07610
crim	0.00773	0.0962	0.75641
congenita	0.05295	4.4936	0.03402
grande	-0.06718	7.2685	0.00702
GLOBAL	NA	19.7751	0.00138

Resposta: O teste da correlação linear sugere que idade, doença de base congênita e tamanho da unidade de tratamento não atendem o pressuposto de proporcionalidade. Os gráficos, por outro lado, mostram que a não proporcionalidade ocorre principalmente para os tempos muito longos (> 30 meses). É possível que, censurando estes valores, obtenha-se um modelo com proporcionalidade. A variável idade tem um padrão mais definido, no entanto, com maior variabilidade nos tempos menores.

Faça também o gráfico dos resíduos martingale *versus* o índice dos indivíduos. Há indicação de indivíduos mal ajustados pelo modelo?

```

> mod.mar <- resid(modeloIII, "martingale")
> plot(mod.mar, xlab = "Índice", ylab = "Resíduo")
> abline(h = 0, col = 2, lty = 2)

```

Resposta: O gráfico de resíduo martingale versus indivíduo sugere a ausência de indivíduos mal-ajustados pelo modelo.

Exercício 8.3: No ajuste do modelo de Cox aos dados de sobrevivência em Aids (exercício 7.3), vimos que a variável tratamento não parece atender ao pressuposto de Cox. Faça a análise de resíduos do modelo III proposto naquele exercício, e procure confirmar esse achado, calculando os resíduos de Schoenfeld.

Lendo os dados

```

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

```

```

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

```



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

Ajustando o modelo de Cox

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

Call:

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

```
n= 193
```

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

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

```
Rsquare= 0.372 (max possible= 0.988 )
Likelihood ratio test= 89.8 on 7 df, p=1.11e-16
Wald test = 76.9 on 7 df, p=5.97e-14
Score (logrank) test = 100 on 7 df, p=0
```

Calculando o resíduo de Schoenfeld e o teste da correlação linear

```
> zph <- cox.zph(modeloIII)
> zph
```

	rho	chisq	p
idade	-0.123	1.24	0.2649
sexoM	0.115	1.29	0.2564

```

acompan1 -0.214  3.97 0.0464
acompan2 -0.153  2.31 0.1288
tratam1   0.153  2.84 0.0920
tratam2   0.177  3.08 0.0794
tratam3  -0.141  1.83 0.1756
GLOBAL           NA 12.17 0.0950

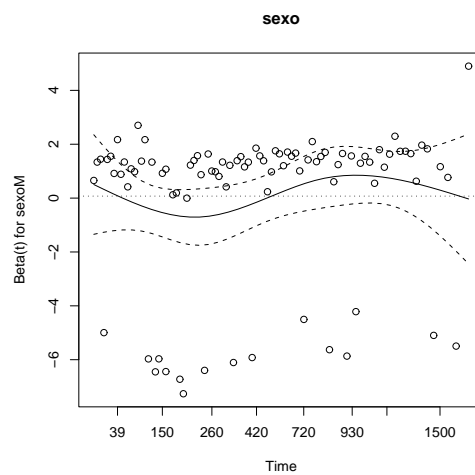
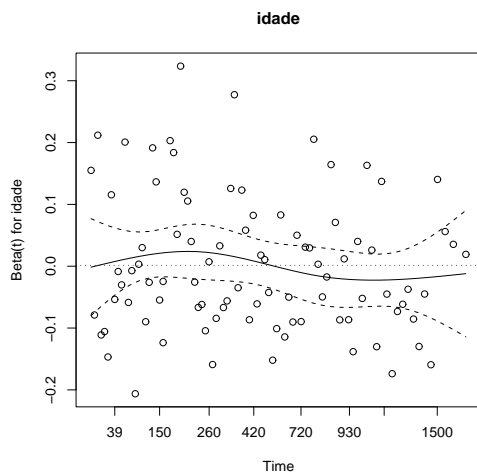
```

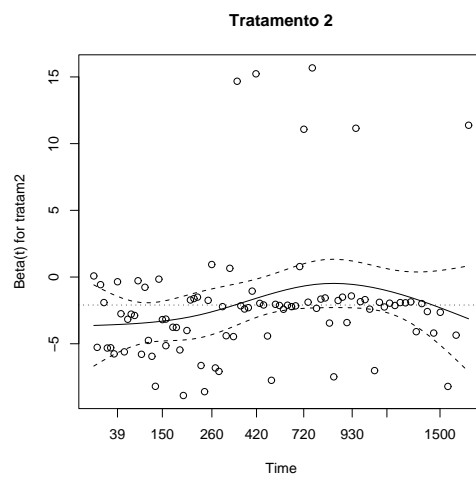
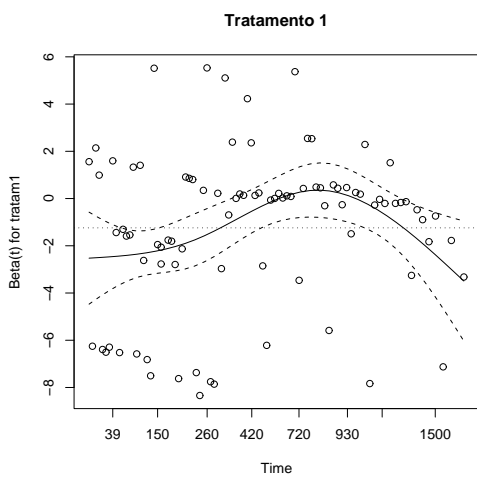
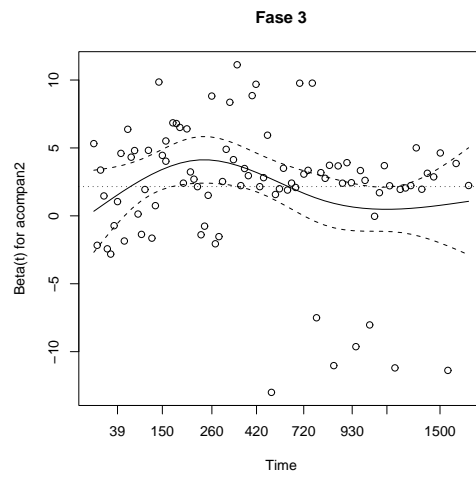
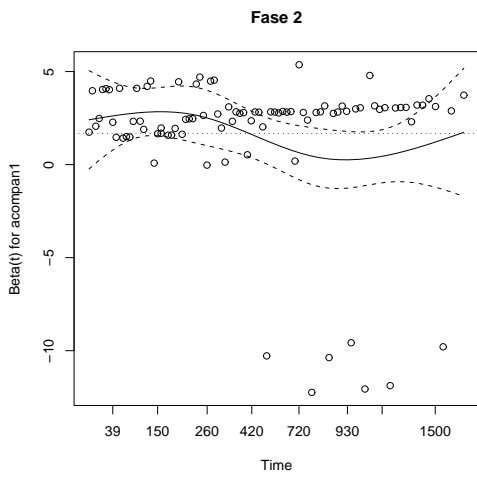
Gráficos dos resíduos de Schoenfeld

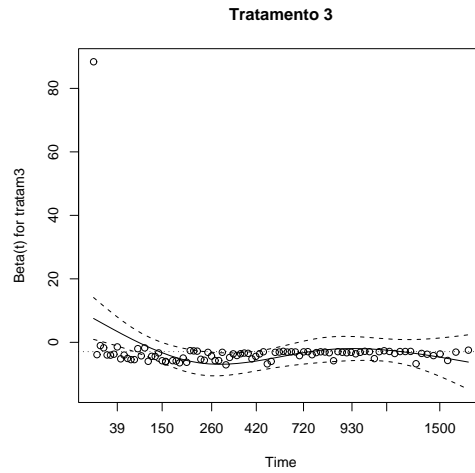
```

> par(mfrow = c(3, 3))
> plot(zph[1], main = "idade")
> abline(h = modeloIII$coef[1], lty = 3)
> plot(zph[2], main = "sexo")
> abline(h = modeloIII$coef[2], lty = 3)
> plot(zph[3], main = "Fase 2")
> abline(h = modeloIII$coef[3], lty = 3)
> plot(zph[4], main = "Fase 3")
> abline(h = modeloIII$coef[4], lty = 3)
> plot(zph[5], main = "Tratamento 1")
> abline(h = modeloIII$coef[5], lty = 3)
> plot(zph[6], main = "Tratamento 2")
> abline(h = modeloIII$coef[6], lty = 3)
> plot(zph[7], main = "Tratamento 3")
> abline(h = modeloIII$coef[7], lty = 3)

```

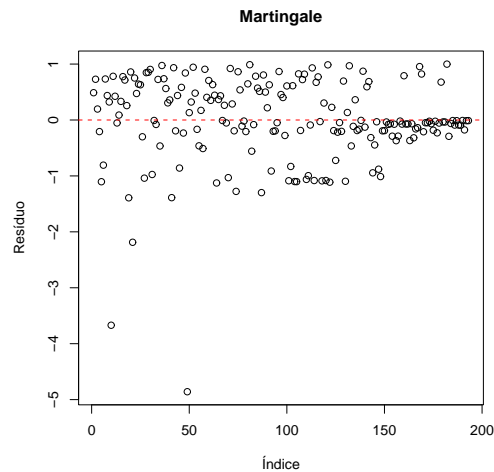






Resíduos de martingale

```
> mod.mar <- resid(modeloIII, "martingale")
> plot(mod.mar, xlab = "Índice", ylab = "Resíduo", main = "Martingale")
> abline(h = 0, col = 2, lty = 2)
```



Resposta: Observando o teste zph, vê-se que todas as variáveis apresentam correlação baixa isto é, menor do que 0.25. Algumas apresentam $p - valor < 0.05$, mas deve-se tomar cuidado com este indicador devido ao tamanho grande da amostra (n grande pode levar à indicação de valores pequenos da estatística a serem significativamente diferentes de zero. Se este efeito é "biologicamente" diferente de zero, é o que

importa). Quanto ao gráfico dos resíduos de Schoenfeld, nota-se que o tratamento 1 é a variável que mais foge do comportamento linear. As outras, embora contenham variação, estão contidas no intervalo de confiança (linha pontilhada).