

# Análise de Sobrevida - Extensões do Modelo de Cox

José Luiz Padilha

Julho de 2024

## Exemplo 1: Sinusite em Pacientes HIV

O primeiro exemplo envolve covariáveis que mudam de valor ao longo do período de acompanhamento. Para análise, faremos uso do Modelo de Cox com covariáveis dependentes do tempo.

### Descrição dos Dados

Neste estudo foram utilizadas informações provenientes de 91 pacientes HIV positivos e 21 HIV negativos, acompanhados no período entre março de 1993 a fevereiro de 1995.

Todos os pacientes incluídos no estudo foram encaminhados ao Centro de Treinamento e Referência em Doenças Infecto-parasitárias (CTR-DIP) da cidade de Belo Horizonte-MG, por pertencerem a grupos de comportamento de risco para adquirir o HIV ou por terem um exame elisa HIV positivo. Após a primeira consulta clínica, os pacientes foram encaminhados ao Serviço de Otorrinolaringologia da UFMG.

As doenças otorrinolaringológicas (ORL) avaliadas foram definidas com base nos estudos de prevalência destas manifestações na literatura em pacientes infectados pelo HIV. Quanto à infecção pelo HIV, os pacientes foram classificados de acordo ao seu Grupo de Risco, isto é,

1. HIV soronegativo: não possuem o HIV.
2. HIV soropositivo assintomático: possuem o vírus mas não desenvolveram o quadro clínico de AIDS e apresentam um perfil imunológico estável.
3. Com ARC (*AIDS Related Complex*): apresentam baixa imunidade e outros indicadores clínicos que antecedem o quadro clínico de AIDS.
4. Com AIDS: já desenvolveram infecções oportunistas que definem esta doença, segundo os critérios do CDC de 1987.

A cada consulta, a classificação do paciente foi reavaliada e, sendo assim, esta covariável depende do tempo pois os pacientes podem mudar de classificação longo do estudo. Cada paciente foi acompanhado por meio de consultas trimestrais. A frequência mediana foi de 4 consultas. A resposta de interesse foi o tempo, em dias, contado a partir da primeira consulta, até a ocorrência da sinusite. O objetivo foi identificar fatores de risco para esta manifestação.

As contagens de CD4 e CD8 também foram medidas no início do estudo mas não foram incluídas nas análises devido a falta de registro de ambas para em torno de 37% dos pacientes. Há ainda 23 valores perdidos para as variáveis Atividade Sexual e Uso de Cocaína.

Os fatores considerados na análise foram:

Idade do Paciente	medida em anos
Sexo do Paciente	0 - Masculino 1 - Feminino
Grupos de Risco	1 - Paciente HIV Soronegativo 2 - Paciente HIV Soropositivo Assintomático 3 - Paciente com ARC 4 - Paciente com AIDS
Atividade Sexual	1 - Homossexual 2 - Bissexual 3 - Heterossexual
Uso de Droga Injetável	1 - Sim 2 - Não
Uso de Cocaína por Aspiração	1 - Sim 2 - Não

A seguir, mostramos como os dados devem ser organizados para análise.

```
aids <- read.table("https://docs.ufpr.br/~giolo/Livro/ApendiceA/aids.txt", h = T)
dim(aids)
```

```
## [1] 133 12
```

```
tail(aids, 7)
```

```
##      pac id sex grp      ti      tf cens cd4  cd8 ats ud ac
## 127 109 30   1   2   0.0 130.5    0 277 1530   3  2  2
## 128 109 30   1   3 130.5 247.0    0 277 1530   3  2  2
## 129 109 30   1   4 247.0 296.0    0 277 1530   3  2  2
## 130 110 42   0   2   0.0 247.0    0 257  510   2  2  2
## 131 111 24   0   2   0.0  86.5    0  57  170  NA NA NA
## 132 111 24   0   3  86.5 192.5    1  57  170  NA NA NA
## 133 112 24   0   2   0.0 226.0    0  NA   NA   NA NA NA
```

## Ajuste do Modelo

Os resultados do ajuste do modelo de Cox incluindo a covariável Grupos de Risco, que depende do tempo foram:

```
library(survival)
fit <- coxph(Surv(ti[ti<tf], tf[ti<tf], cens[ti<tf]) ~ id[ti<tf] + sex[ti<tf] +
             factor(grp)[ti<tf] + factor(ats)[ti<tf] + factor(ud)[ti<tf] +
             factor(ac)[ti<tf], method = "breslow", data = aids)
fit
```

```
## Call:
```

```
## coxph(formula = Surv(ti[ti < tf], tf[ti < tf], cens[ti < tf]) ~
##      id[ti < tf] + sex[ti < tf] + factor(grp)[ti < tf] + factor(ats)[ti <
##      tf] + factor(ud)[ti < tf] + factor(ac)[ti < tf], data = aids,
##      method = "breslow")
```

```
##
##               coef exp(coef) se(coef)      z      p
## id[ti < tf]      -0.10087   0.90405  0.04379 -2.304 0.02125
## sex[ti < tf]       1.03602   2.81798  0.75928  1.364 0.17242
## factor(grp)[ti < tf]2 -0.30822   0.73475  1.42520 -0.216 0.82878
## factor(grp)[ti < tf]3  3.07416  21.63166  1.18410  2.596 0.00943
## factor(grp)[ti < tf]4  3.84194  46.61572  1.21095  3.173 0.00151
## factor(ats)[ti < tf]2  0.34412   1.41074  0.76758  0.448 0.65393
## factor(ats)[ti < tf]3 -0.85335   0.42598  0.80928 -1.054 0.29167
## factor(ud)[ti < tf]2 -0.15168   0.85926  1.20073 -0.126 0.89947
## factor(ac)[ti < tf]2   1.45439   4.28187  1.47348  0.987 0.32362
##
## Likelihood ratio test=32.8 on 9 df, p=0.0001447
## n= 97, number of events= 18
## (27 observations deleted due to missingness)
```

Com exceção das covariáveis idade e grupos de risco (esta última dependente do tempo), as demais parecem ser não significativas. Removendo-se estas covariáveis gradativamente chegou-se no modelo final para a ocorrência de sinusite.

Para o modelo final ajustado tem-se as estimativas:

```
fit1 <- coxph(Surv(ti[ti<tf], tf[ti<tf], cens[ti<tf]) ~ id[ti<tf] + factor(grp)[ti<tf],
              method = "breslow", data = aids)
summary(fit1)
```

```
## Call:
## coxph(formula = Surv(ti[ti < tf], tf[ti < tf], cens[ti < tf]) ~
##       id[ti < tf] + factor(grp)[ti < tf], data = aids, method = "breslow")
##
## n= 122, number of events= 26
## (2 observations deleted due to missingness)
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## id[ti < tf]      -0.07695   0.92593  0.03128 -2.460 0.013903 *
## factor(grp)[ti < tf]2 -0.73023   0.48180  1.00059 -0.730 0.465512
## factor(grp)[ti < tf]3  2.27261   9.70473  0.83711  2.715 0.006631 **
## factor(grp)[ti < tf]4  2.64906  14.14070  0.78967  3.355 0.000795 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## id[ti < tf]           0.9259   1.07999   0.87086   0.9845
## factor(grp)[ti < tf]2  0.4818   2.07556   0.06779   3.4243
## factor(grp)[ti < tf]3  9.7047   0.10304   1.88121  50.0644
## factor(grp)[ti < tf]4 14.1407   0.07072   3.00815  66.4725
##
## Concordance= 0.829 (se = 0.045 )
## Likelihood ratio test= 38.61 on 4 df, p=8e-08
## Wald test              = 25.9 on 4 df, p=3e-05
## Score (logrank) test = 39.18 on 4 df, p=6e-08
```

Idade e grupos de risco foram identificados como fatores de risco para a ocorrência de sinusite. Podemos fazer as seguintes interpretações para os resultados do ajuste:

- A cada aumento de 10 anos na idade do paciente, o risco de desenvolver sinusite diminui em 54% ( $1 - \exp\{-0.0769 \times 10\} = 0.5319$ ). Ou seja, pacientes mais jovens estão mais sujeitos a esta infecção.

- A taxa de ocorrência de sinusite para os pacientes HIV soropositivo assintomáticos não difere significativamente do grupo HIV soronegativo.
- A taxa de ocorrência do grupo com ARC é  $\exp \{2.273\} = 9.70$  vezes aquela do grupo HIV soronegativo.
- Para o grupo com AIDS, o risco de desenvolver sinusite é 14.14 vezes o risco do grupo HIV soronegativo. A precisão das estimativas associadas a estas duas últimas razões de taxas de falha é bastante reduzida, o que pode ser observado pela grande amplitude de seus respectivos intervalos e confiança.

## Exemplo 2: Hormônio de Crescimento

Nesta segunda aplicação, trataremos do caso em que a suposição de taxas proporcionais não é atendida. O modelo de Cox estratificado será usado para análise.

### Descrição dos Dados

O Hormônio do Crescimento (GH) é um importante agente do desenvolvimento humano e, quando sua deficiência é diagnosticada, são ministradas doses periódicas de acordo com o acompanhamento médico.

A deficiência de GH pode se manifestar em graus variados e ter muitas causas diferentes. Em crianças com defasagem no crescimento, é feita uma avaliação para verificar se a baixa estatura é devida à má secreção/ação do hormônio do crescimento. Em caso positivo, são ministradas doses do hormônio sintetizado e o desenvolvimento de cada indivíduo é acompanhado em intervalos regulares de tempo.

Um estudo foi realizado com 80 crianças participantes do Programa Hormonal do Crescimento da Secretaria de Saúde de Minas Gerais, diagnosticadas com deficiência do hormônio de crescimento. As mesmas tiveram um período de acompanhamento de, no mínimo, 19 meses em 31 de dezembro de 2002.

O objetivo do estudo consistiu em identificar fatores determinantes do crescimento de crianças com deficiência do GH. Foram coletadas 16 covariáveis que foram consideradas potencialmente importantes para descrever o crescimento de crianças. A seguir, uma descrição dessas variáveis.

Código	Variável	Descrição
V1	Sexo	1 se Masculino e 2 se Feminino
V2	Raça	1 se Branca e 2 se Negra
V3	Naturalidade	1 se Grande BH e 2 se Interior
V4	Tipo de parto	1 se Normal, 2 se Cesário e 3 se Fórceps
V5	Parto traumático	1 se Sim e 2 se Não
V6	Recém nascido	1 se AIG e 2 se PIG
V7	Apresentação	1 se Cefálica e 2 se Pélvica
V8	Renda	1: $\leq 2SM$ , 2: 2 a 5SM, 3: 5 a 10SM e 4: $>10SM$
V9	Diagnóstico/origem	1 se Idiopático e 2 se Orgânico
V10	Grau de deficiência	1 se Isolado e 2 se DMHH
V11	Peso ao nascimento	Entre 1250 e 4240 g
V12	Velocidade	Entre 0,5 e 5 cm/ano
V13	Idade óssea	Entre 0,3 e 13 anos
V14	Idade cronológica/inicial	Entre 2 e 21 anos
V15	Altura inicial	Entre 71 e 154,8 cm
V16	Dose inicial	Entre 0,20 e 0,71 mml/kg

AIG = adequado para a idade gestacional e PIG = pequeno para a idade gestacional,

SM = salário mínimo, BH = Belo Horizonte, DMHH = Deficiência múltipla de hormônios hipofisários.

Uma variável que norteia a decisão de alta clínica é a altura alvo. Ela é definida como a altura dos pais subtraída de 7 cm para meninas e somada de 7 cm para meninos. A variável resposta considerada é o tempo (em meses) até a altura alvo ser atingida.

```
hg2 <- read.table("https://docs.ufpr.br/~giolo/Livro/ApendiceA/hg2.txt", h=T)
dim(hg2)
```

```
## [1] 80 7
```

```
head(hg2)
```

```
##   tempos cens  raca  ialtura trauma recémnas renda
## 1     27    1    1   100.0     2        1     3
## 2      7    0    1   100.0     2        1     2
## 3     47    1    2   113.0     1        2     1
## 4     19    0    1   119.5    NA        1     4
## 5     12    0    1   125.0     1        1     2
## 6     10    0    1   120.5     2        1     2
```

## Ajuste do Modelo de Cox

Para seleção das covariáveis foi utilizado o modelo de Cox e a estratégia de construção de modelos proposta por Collett (para detalhes, veja a seção 4.5.3 do livro). Após a utilização dessa estratégia e discussões com os pesquisadores, as covariáveis selecionadas foram: raça, ocorrência de parto traumático, recém-nascido, renda e altura inicial, com a presença das possíveis interações entre essas covariáveis.

Uma análise inicial indicou que as três categorias de renda com valores iguais ou inferiores a 10 SM não diferiam entre si. A covariável renda foi então dicotomizada de acordo com os níveis  $\leq 10$  SM e  $>10$  SM.

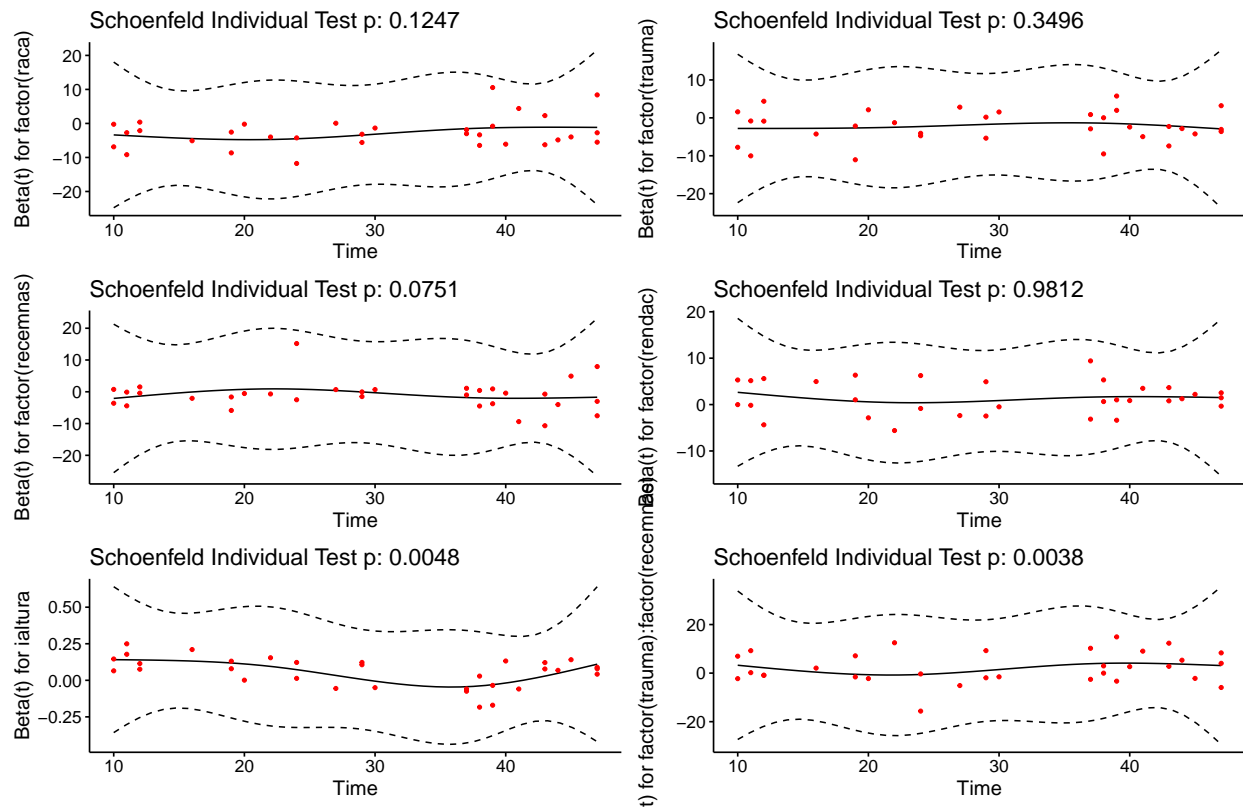
```
hg2$rendac <- ifelse(hg2$renda < 4, 1, 2)
fit1 <- coxph(Surv(tempos, cens) ~ factor(raca) + factor(trauma) + factor(recémnas) +
              factor(rendac) + ialtura + factor(trauma)*factor(recémnas), data = hg2,
              method = "breslow")
fit1
```

```
## Call:
## coxph(formula = Surv(tempos, cens) ~ factor(raca) + factor(trauma) +
##       factor(recémnas) + factor(rendac) + ialtura + factor(trauma) *
##       factor(recémnas), data = hg2, method = "breslow")
##
##               coef exp(coef) se(coef)      z      p
## factor(raca)2    -2.70671   0.06676  0.74966 -3.611 0.000306
## factor(trauma)2  -2.22402   0.10817  0.68573 -3.243 0.001182
## factor(recémnas)2 -1.07768   0.34038  0.81825 -1.317 0.187816
## factor(rendac)2    1.44405   4.23783  0.55807  2.588 0.009665
## ialtura           0.05809   1.05981  0.01747  3.326 0.000882
## factor(trauma)2:factor(recémnas)2  2.29680   9.94231  1.07065  2.145 0.031933
##
## Likelihood ratio test=31.22  on 6 df, p=2.3e-05
## n= 70, number of events= 32
## (10 observations deleted due to missingness)
```

Para confiarmos nos resultados do modelo ajustado acima, precisamos que a suposição de taxas de falhas proporcionais seja atendida. Podemos verificar tal pressuposto por meio da análise gráfica e de testes envolvendo os resíduos de Schoenfeld.

```
library(survminer)
zph <- cox.zph(fit1, transform = "identity")
ggcoxzph(zph)
```

Global Schoenfeld Test p: 0.003381



A análise dos resíduos e os resultados dos testes indicam que a suposição de taxas de falhas proporcionais não é atendida. A variável altura inicial está causando tal violação, e, assim, o modelo de Cox deve ser estratificado por esta variável.

## Ajuste do Modelo de Cox Estratificado

Considerando o modelo de Cox estratificado em que as crianças são consideradas de acordo com os estratos formados pelo valor mediano da altura inicial ( $< 120$  cm e  $\geq 120$  cm) e assumindo que o vetor  $\beta$  é comum para esses dois estratos, tem-se o modelo:

$$\lambda(t|x_{ij}) = \lambda_{0j}(t) \exp \{x'_{ij}\beta\},$$

para  $j = 1, 2$  e  $i = 1, \dots, n_j$ , em que  $n_j$  é o número de crianças no  $j$ -ésimo estrato. Os resultados são mostrados a seguir.

```
hg2$alt <- ifelse(hg2$ialtura < 120, 1, 2)
fit2 <- coxph(Surv(tempo, cens) ~ factor(raca) + factor(trauma) + factor(recemnas) +
             factor(rendac) + factor(trauma)*factor(recemnas) + strata(alt),
             data = hg2, method = "breslow")
fit2
```

## Call:

```
## coxph(formula = Surv(tempo, cens) ~ factor(raca) + factor(trauma) +
##       factor(recemnas) + factor(rendac) + factor(trauma) * factor(recemnas) +
##       strata(alt), data = hg2, method = "breslow")
##
```

	coef	exp(coef)	se(coef)	z	p
## factor(raca)2	-1.9846	0.1374	0.7251	-2.737	0.0062

```
## factor(trauma)2          -1.3535    0.2583    0.5886 -2.299 0.0215
## factor(recemnas)2        -0.4701    0.6250    0.8495 -0.553 0.5800
## factor(rendac)2          1.1896    3.2857    0.5253  2.264 0.0236
## factor(trauma)2:factor(recemnas)2  1.2618    3.5316    1.0550  1.196 0.2317
##
## Likelihood ratio test=16.88 on 5 df, p=0.004731
## n= 70, number of events= 32
## (10 observations deleted due to missingness)
```

Observamos que o efeito de interação entre trauma e recém-nascido, bem como o efeito de recém-nascido não são significativos. Removendo-se inicialmente a interação, e em seguida, a covariável recém-nascido, chegamos ao seguinte modelo.

```
fit3 <- coxph(Surv(tempo, cens) ~ factor(raca) + factor(trauma) + factor(rendac) +
              strata(alt), data = hg2, method = "breslow")
fit3
```

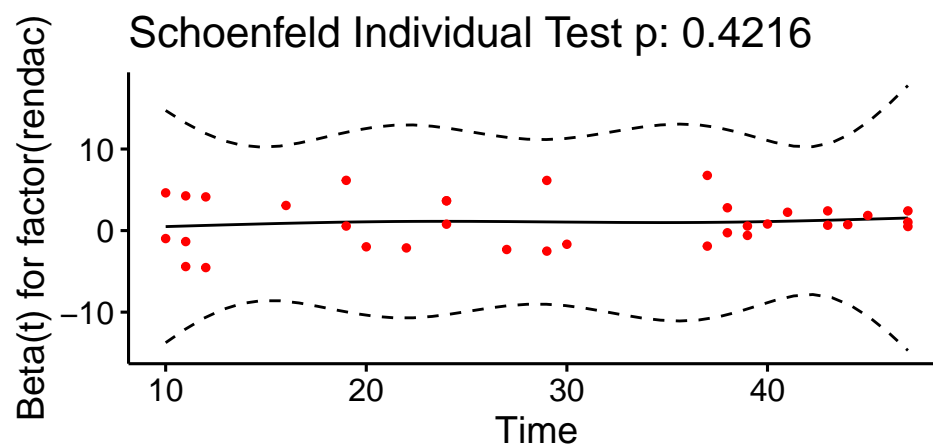
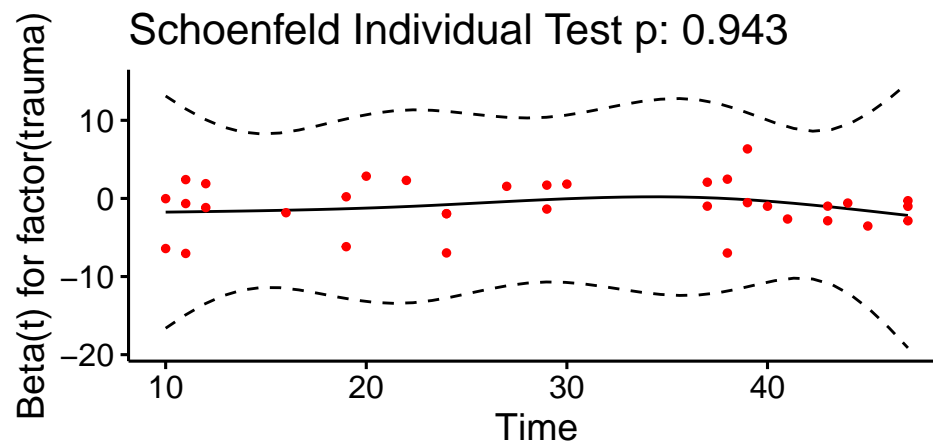
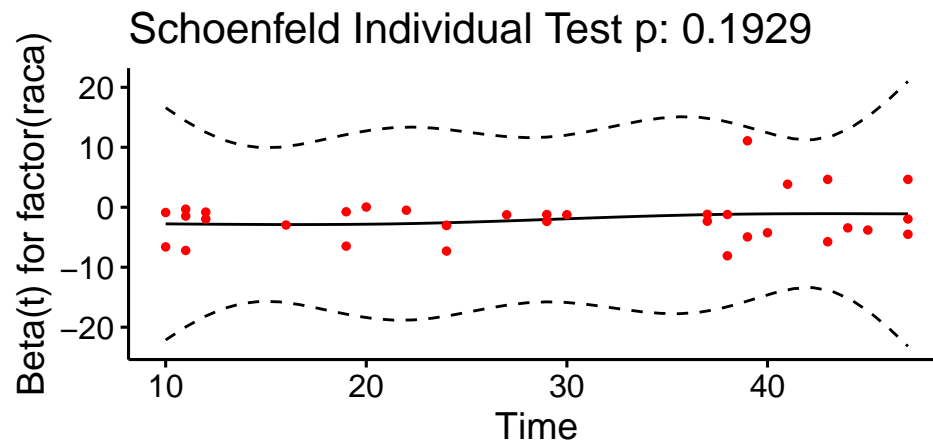
```
## Call:
## coxph(formula = Surv(tempo, cens) ~ factor(raca) + factor(trauma) +
##       factor(rendac) + strata(alt), data = hg2, method = "breslow")
##
##               coef exp(coef) se(coef)      z      p
## factor(raca)2  -1.9599    0.1409   0.6866 -2.855 0.00431
## factor(trauma)2 -1.0074    0.3652   0.5272 -1.911 0.05604
## factor(rendac)2  1.0338    2.8118   0.5050  2.047 0.04062
##
## Likelihood ratio test=14.6 on 3 df, p=0.00219
## n= 76, number of events= 34
## (4 observations deleted due to missingness)
```

Os resíduos padronizados de Schoenfeld e respectivos testes associados, apresentados a seguir, não mostraram evidências contra a hipótese nula de proporcionalidade das taxas de falha dentro dos estratos formados pelas duas categorias da covariável altura inicial.

```
zph <- cox.zph(fit3, transform = "identity")
ggcoxzph(zph)
```



Global Schoenfeld Test p: 0.4556



As seguintes interpretações podem ser obtidas a partir das estimativas fornecidas pelo ajuste do modelo de Cox estratificado.

```
summary(fit3)
```

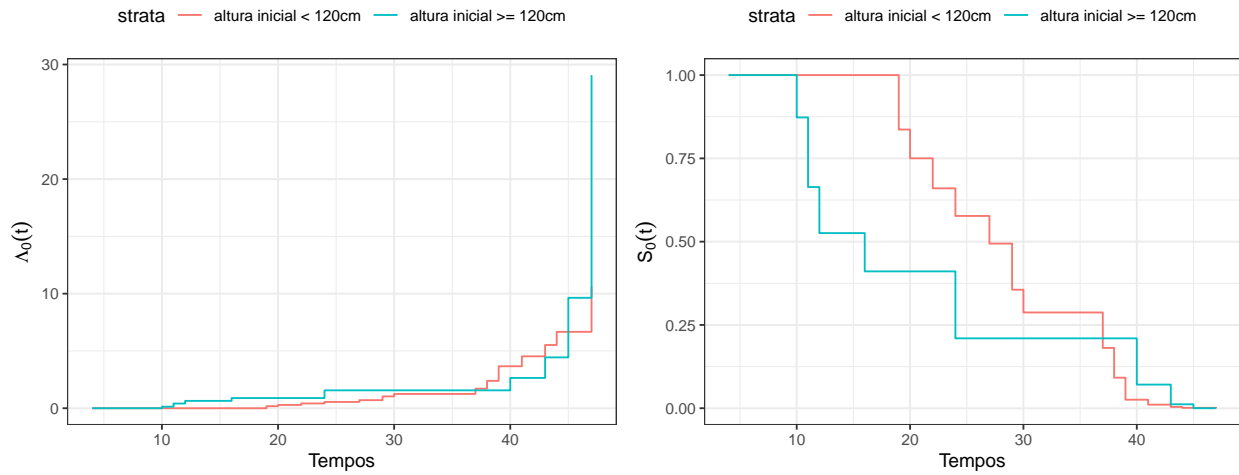
```
## Call:  
## coxph(formula = Surv(tempo, cens) ~ factor(raca) + factor(trauma) +
```

```
##      factor(rendac) + strata(alt), data = hg2, method = "breslow")
##
##      n= 76, number of events= 34
##      (4 observations deleted due to missingness)
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## factor(raca)2   -1.9599    0.1409   0.6866 -2.855  0.00431 **
## factor(trauma)2 -1.0074    0.3652   0.5272 -1.911  0.05604 .
## factor(rendac)2  1.0338    2.8118   0.5050  2.047  0.04062 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## factor(raca)2      0.1409      7.0987   0.03668   0.5411
## factor(trauma)2     0.3652      2.7384   0.12994   1.0263
## factor(rendac)2     2.8118      0.3556   1.04513   7.5649
##
## Concordance= 0.701 (se = 0.045 )
## Likelihood ratio test= 14.6 on 3 df,  p=0.002
## Wald test              = 11.37 on 3 df,  p=0.01
## Score (logrank) test = 12.53 on 3 df,  p=0.006
```

- A taxa de ocorrência da altura alvo em crianças brancas é estimada em  $\exp(1.96) = 7.10$  vezes a das crianças negras (IC 95%: 1.85 a 27.27).
- Para crianças que tiveram parto traumático, a taxa de atingir a altura alvo é cerca de  $\exp(1.01) = 2.74$  a das crianças que não tiveram parto traumático (IC 95%: 0.97 a 7.70).
- Finalmente, a taxa de uma criança de família com renda superior a 10 SM atingir a altura alvo é 2.81 vezes a das crianças com renda até 10 SM (IC 95%: 1.04 a 7.56).

As funções de taxa de falha acumulada de base e de sobrevivência de base, para os estratos da covariável altura inicial, são mostradas a seguir.

```
H0 <- basehaz(fit3, centered = F)
H0$S0 <- exp(-H0$hazard)
levels(H0$strata) <- c("altura inicial < 120cm", "altura inicial >= 120cm")
p1 <- ggplot(H0, aes(x = time, y = hazard, color = strata)) + geom_step() + theme_bw() +
  theme(legend.position = "top") + labs(x = "Tempos", y = expression(Lambda[0]*(t)))
p2 <- ggplot(H0, aes(x = time, y = S0, color = strata)) + geom_step() + theme_bw() +
  theme(legend.position = "top") + labs(x = "Tempos", y = expression(S[0]*(t)))
ggarrange(p1, p2, ncol = 2)
```



Tais funções se cruzam, reforçando as evidências de violação da suposição de taxas de falhas proporcionais para esta covariável. O ajuste do modelo estratificado mostrou que os fatores raça, ocorrência de parto traumático e renda influenciam o tempo até a criança atingir a altura alvo.