

MAE514 - Introdução à Análise de Sobrevida e Aplicações

Fernando Henrique Ferraz Pereira da Rosa
Vagner Aparecido Pedro Júnior

5 de dezembro de 2005

Lista 5

Os dados do arquivo `viarapida.xls` (obtenha o arquivo com Elaine, no CEA - leve um disquete em branco) referem-se a um estudo realizado no Instituto do Coração, do Hospital das Clínicas - FMUSP, para comparar dois tipos de protocolo de recuperação de pacientes submetidos a certo tipo de cirurgia: via-rápida e convencional. O protocolo via-rápida tem como filosofia uma maior integração entre as várias equipes que atendem os pacientes, com a finalidade de tentar diminuir o tempo de internação, melhorando a recuperação e diminuindo custos. Considere, para as questões que se seguem, apenas as seguintes variáveis:

- `delta` - indicador de censura (0 - censura, 1 - saída do centro cirúrgico)
- `idade` - idade do paciente na internação
- `Protocolo` - sendo 0 convencional e 1 via rápida
- `Raça` - sendo 1 Branco, 2 negro e 3 amarelo
- `Sexo` - sendo 0 feminino e 1 masculino
- `Tipac` - tipo de anomalia cardíaca, sendo 0 congênito e 1 coronariano
- `TCC` - tempo entre a entrada a saída do centro cirúrgico (horas)

1. Utilizando o programa de sua preferência, construa curvas de Kaplan-Meier para o TCC comparando sexo, raça, tipo de paciente (`Tipac`) e protocolo. Construa testes estatísticos adequados e interprete os resultados.

Na Figura 1 temos o gráfico de Kaplan-Meier de acordo com o sexo. Não se observa diferença aparente. Pelo teste de log-rank, obtemos:

```
> survdiff(via.Surv ~ SEXO, data = viarapida)
```

Call:

```
survdiff(formula = via.Surv ~ SEXO, data = viarapida)
```

```
n=143, 1 observations deleted due to missing.
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
SEXO=feminino	53	51	45.1	0.779	1.23
SEXO=masculino	90	86	91.9	0.382	1.23

Chisq= 1.2 on 1 degrees of freedom, p= 0.268

o que nos indica que não há diferença significativa a 5%.

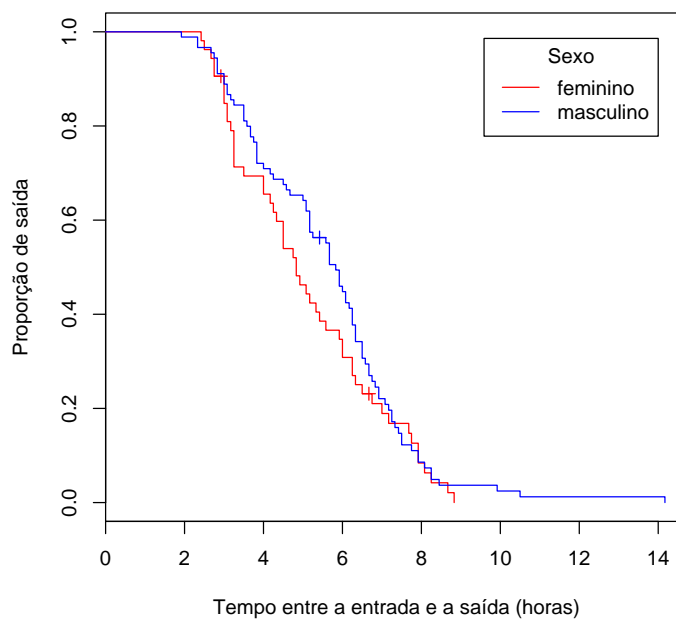


Figura 1: Gráfico do estimador de Kaplan-Meier da função de sobrevivência dividido por sexo.

Na Figura 2 temos o gráfico de Kaplan-Meier de acordo com a raça. Parece haver uma diferença grande entre a curva de sobrevivência de pacientes brancos e negros. Pelo teste de log-rank, obtemos:

```
> survdiff(via.Surv ~ RACA, data = viarapida)
```

Call:

```
survdiff(formula = via.Surv ~ RACA, data = viarapida)
```

n=143, 1 observations deleted due to missing.

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
RACA=branco	136	130	122.28	0.487	5.329
RACA=negro	4	4	10.84	4.314	5.907
RACA=amarelo	3	3	3.88	0.200	0.215

Chisq= 6.2 on 2 degrees of freedom, p= 0.045

que aponta para uma diferença significativa a 5%. Deve-se levar essa conclusão com cautela entretanto, pois o número de pacientes de raça negra ou amarela é muito baixo:

```
> table(viarapida$RACA)
```

branco	negro	amarelo
136	4	3

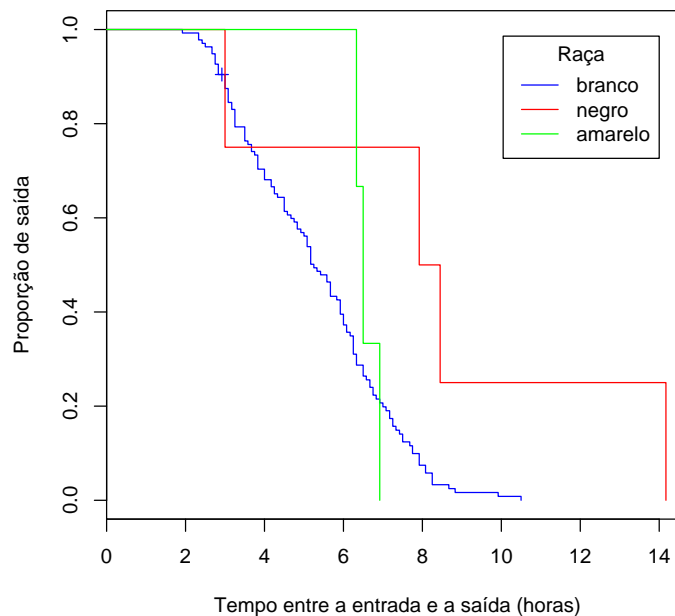


Figura 2: Gráfico do estimador de Kaplan-Meier da função de sobrevivência dividido por raça

Na Figura 3 temos o gráfico de Kaplan-Meier de acordo com o tipo de paciente. Aparenta-se uma clara diferença entre a sobrevivência de pacientes com anomalia congênita e coronariana. Em particular, pacientes com anomalia congênita tendem a ter proporção de saída menor. No teste de log-rank obtemos:

```
> survdiff(via.Surv ~ TIPAC, data = viarapida)
```

Call:

```
survdiff(formula = via.Surv ~ TIPAC, data = viarapida)
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
TIPAC=congénito	69	65	24.0	70.3	105
TIPAC=coronariano	75	73	114.0	14.8	105

Chisq= 105 on 1 degrees of freedom, p= 0

que aponta para uma diferença significativa a 5%. Concluímos assim que há diferença no tempo de saída do centro cirúrgico entre os dois tipos de pacientes.

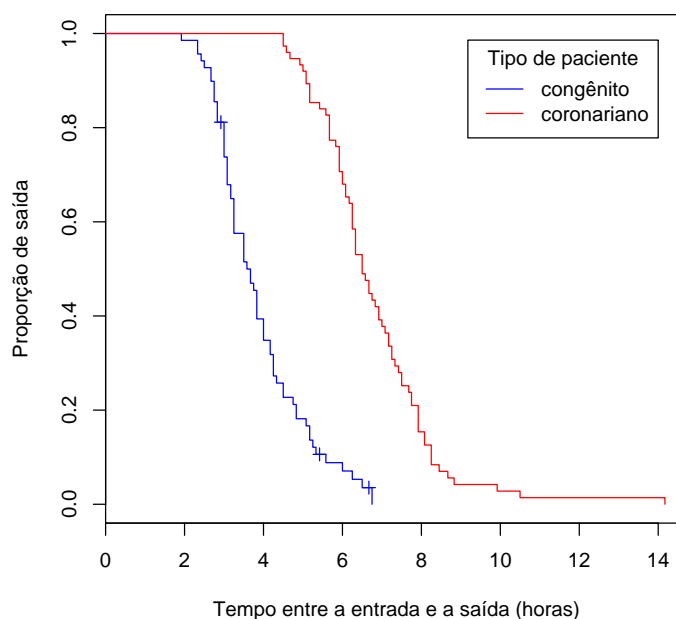


Figura 3: Gráfico do estimador de Kaplan-Meier da função de sobrevivência dividido por tipo de paciente

Na Figura 4 temos o gráfico de Kaplan-Meier de acordo com o protocolo de recuperação. Entre os instantes de tempo 2 e 8 horas, aparenta-se haver diferença entre via-rápida e condicional, implicando em uma menor proporção de saída. Dali em diante as duas curvas parecem mais próximas. No teste de log-rank obtemos:

```
> survdiff(via.Surv ~ Protocolo, data = viarapida)
```

Call:

```
survdiff(formula = via.Surv ~ Protocolo, data = viarapida)
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
Protocolo=convencional	58	56	73.1	4.01	9.15

Protocolo=via rápida 86 82 64.9 4.51 9.15

Chisq= 9.1 on 1 degrees of freedom, p= 0.00249

que aponta para uma diferença significativa a 5%. Concluimos assim que há diferença no tempo de saída do centro cirúrgico entre os dois protocolos.

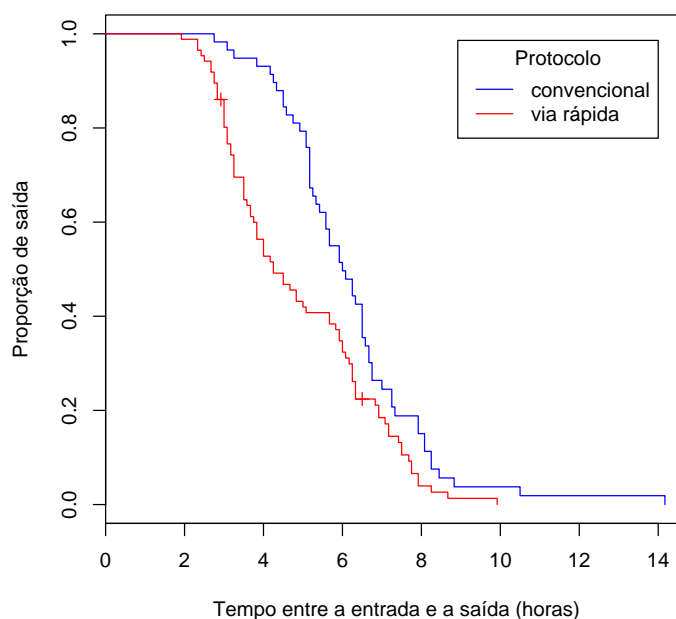


Figura 4: Gráfico do estimador de Kaplan-Meier da função de sobrevivência dividido por tipo de paciente

2. Ajuste um modelo Weibull com todas as variáveis descritas. Descreva e utilize algum método para avaliar a adequabilidade do modelo.

Começamos ajustando esse modelo no R, através do comando:

```
> viarap.weibull <- survreg(via.Surv ~ IDADE + Protocolo + RACA +
+ TIPAC + SEXO, data = viarapida)
> summary(viarap.weibull)
```

Call:

```
survreg(formula = via.Surv ~ IDADE + Protocolo + RACA + TIPAC +
SEXO, data = viarapida)
```

	Value	Std. Error	z	p
(Intercept)	1.601497	0.04161	38.492	0.00e+00
IDADE	-0.000422	0.00172	-0.245	8.06e-01
Protocolovia rápida	-0.205336	0.03972	-5.170	2.35e-07
RACAnegro	0.312484	0.11310	2.763	5.73e-03

RACAamarelo	-0.091738	0.12528	-0.732	4.64e-01
TIPACcoronariano	0.558970	0.09379	5.960	2.52e-09
SEXOmasculino	-0.057306	0.04038	-1.419	1.56e-01
Log(scale)	-1.555706	0.06421	-24.227	1.17e-129

Scale= 0.211

Weibull distribution

Loglik(model)= -224.2 Loglik(intercept only)= -291.2

Chisq= 133.97 on 6 degrees of freedom, p= 0

Number of Newton-Raphson Iterations: 6

n=143 (1 observations deleted due to missing)

Podemos verificar a bondade de ajuste de um modelo de Weibull através dos resíduos de Cox-Snell. Sob esse modelo, os resíduos devem ter uma distribuição exponencial com média igual a 1. Uma forma de verificar isso é através do gráfico dos resíduos de Cox-Snell pela função de sobrevivência empírica transformada, disponível na Figura 5. Quanto mais esse gráfico se aproxima duma reta passando pela origem, melhor o ajuste do modelo. Vemos que os valores estão razoavelmente próximos da reta passando pela origem, indicando que o modelo parece bem ajustado.

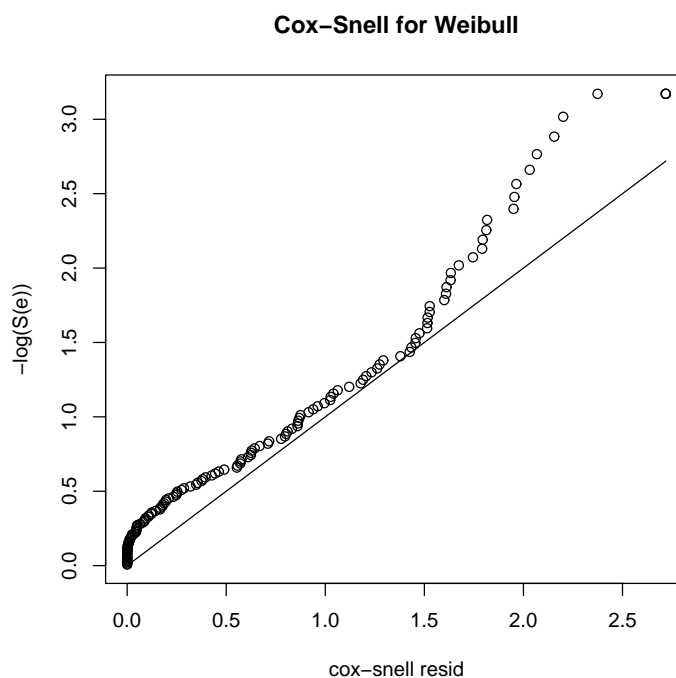


Figura 5: Gráfico dos resíduos de cox-snell para o modelo de Weibull

3. Verifique se o modelo Weibull pode ser reduzido para um modelo exponencial.

Começamos ajustando o modelo exponencial, apesar do parâmetro de escala na Weibull (0,21, bem diferente de 1) não oferecer suporte para isso.

```
> viarap.exp <- update(viarap.weibull, dist = "exponential")
> summary(viarap.exp)
```

Call:

```
survreg(formula = via.Surv ~ IDADE + Protocolo + RACA + TIPAC +
        SEX0, data = viarapida, dist = "exponential")
```

	Value	Std. Error	z	p
(Intercept)	1.50358	0.20901	7.1940	6.29e-13
IDADE	0.00141	0.00776	0.1813	8.56e-01
Protocolovia rápida	-0.17245	0.17953	-0.9605	3.37e-01
RACAnegro	0.19992	0.51830	0.3857	7.00e-01
RACAamarelo	-0.05375	0.59939	-0.0897	9.29e-01
TIPACcoronariano	0.45046	0.42501	1.0599	2.89e-01
SEX0masculino	-0.02647	0.18680	-0.1417	8.87e-01

Scale fixed at 1

Exponential distribution

Loglik(model)= -367.2 Loglik(intercept only)= -372.7

Chisq= 10.94 on 6 degrees of freedom, p= 0.09

Number of Newton-Raphson Iterations: 3

n=143 (1 observations deleted due to missing)

Obtemos agora o gráfico de $S(t)$ versus $\log(t)$, para os resíduos desse modelo, para verificar o ajuste. Temos na Figura 6 esse gráfico. Como a uma forte fuga de linearidade, concluímos que o ajuste exponencial não é adequado, de acordo com o que é indicado em Venables e Ripley (2002)¹.

¹VENABLES, W. B. and RIPLEY, B. R. (2002). *Modern Applied Statistics with S*. 4ed. Springer: New York.

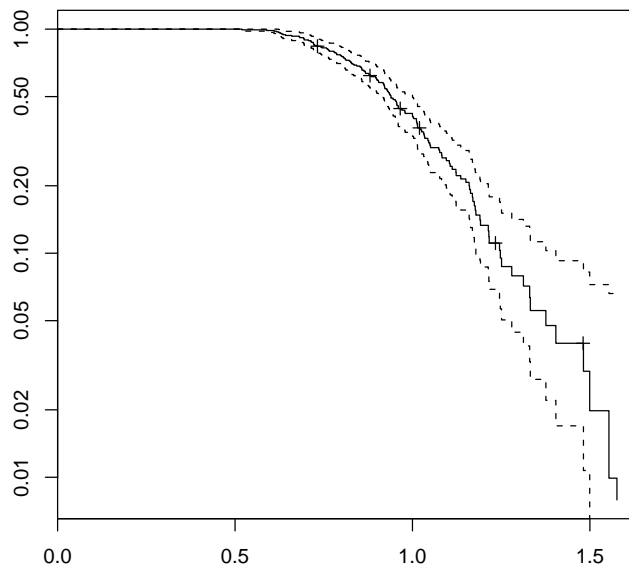


Figura 6: Gráfico na escala log de $S(t)$

4. Ajuste um modelo de riscos proporcionais de Cox. Verifique quais os fatores que influenciam no tempo de centro cirúrgico.

Ajustamos um modelo de riscos proporcionais de Cox através do comando no R:

```
> viarap.cox <- coxph(via.Surv ~ IDADE + Protocolo + RACA + TIPAC +
+   SEXO, data = viarapida)
> summary(viarap.cox)
```

Call:

```
coxph(formula = via.Surv ~ IDADE + Protocolo + RACA + TIPAC +
      SEXO, data = viarapida)
```

n=143 (1 observations deleted due to missing)

	coef	exp(coef)	se(coef)	z	p
IDADE	0.000684	1.001	0.00832	0.0823	9.3e-01
Protocolovia rápida	0.882482	2.417	0.20602	4.2834	1.8e-05
RACAnegro	-0.751004	0.472	0.61993	-1.2114	2.3e-01
RACAamarelo	0.433967	1.543	0.60133	0.7217	4.7e-01
TIPACcoronariano	-2.525291	0.080	0.50985	-4.9530	7.3e-07
SEXOmasculino	0.280007	1.323	0.19265	1.4535	1.5e-01

exp(coef) exp(-coef) lower .95 upper .95

IDADE	1.001	0.999	0.9845	1.017
Protocolovia rápida	2.417	0.414	1.6139	3.619
RACAnegro	0.472	2.119	0.1400	1.590
RACAamarelo	1.543	0.648	0.4749	5.016
TIPACcoronariano	0.080	12.495	0.0295	0.217
SEXOmasculino	1.323	0.756	0.9070	1.930

Rsquare= 0.555 (max possible= 1)
Likelihood ratio test= 116 on 6 df, p=0
Wald test = 97.3 on 6 df, p=0
Score (logrank) test = 124 on 6 df, p=0

Detecta-se a 5% o efeito de Protocolo e de Tipo de paciente, conforme havia sido observado na análise descritiva pelas curvas de Kaplan-Meier. Não se detecta o efeito de Raça, o que já era esperado, já que o número de observações na raças amarela e negra era tão baixo.

- Interprete e compare os parâmetros dos modelos finais obtidos para a abordagem paramétrica e semiparamétrica.

Na abordagem semiparamétrica, obtivemos um coeficiente para Protocolo de 0,88, implicando que $\exp(0,88) = 2,42$. Isso quer dizer que, fixadas as outras covariáveis, passar do protocolo convencional para a via rápida, aumenta os riscos de saída do centro cirúrgico em 2,42 vezes (em relação a manter-se o protocolo convencional). Isso indica que o protocolo via rápida é melhor que o protocolo convencional, no sentido de submeter os pacientes a um tempo menor de permanência.

Interpretando o coeficiente do tipo de paciente, temos que um paciente com tipo de anomalia cardíaca congênita, tem risco de saída do centro cirúrgico 12,49 vezes maior do que o mesmo paciente com anomalia cardíaca coronariana, fixadas as outras covariáveis e o instante de tempo. Isso quer dizer que pacientes com anomalia coronariana congênita saem da operação mais rápido.

No modelo de Weibull o risco relativo é dado por $\exp(\beta_k \times 1/\text{par. escala})$. Para a variável tipo de paciente, isso nos fornece que o risco de alta para pacientes com anomalia cardíaca congênita é 14,14 vezes maior do que para pacientes com anomalia cardíca coronariana. Para o Protocolo de recuperação, obtemos que o risco de alta para via rápida é 2,64 o risco de alta para via convencional.

Ambas as abordagens forneceram riscos relativos parecidos, mostrando que os modelos estão concordando nas inferências.

- Avalie a suposição de riscos proporcionais do Modelo de Cox.

Com o comando abaixo, realizamos o teste de riscos proporcionais no R:

```
> cox.zph(viarap.cox)

              rho  chisq      p
IDADE                0.1305  2.6693 0.102304
Protocolovia rápida -0.2965 14.3947 0.000148
```

RACAnegro	-0.1347	2.7360	0.098113
RACAamarelo	0.0648	0.5659	0.451904
TIPACcoronariano	0.0134	0.0285	0.865879
SEXOmasculino	0.0449	0.2986	0.584764
GLOBAL	NA	22.1772	0.001125

Onde rejeitamos, a 5%, que os riscos são proporcionais para Protocolo.
Para as outras variáveis a suposição de riscos proporcionais está adequada.

Sobre

A versão eletrônica desse arquivo pode ser obtida em <http://www.feferraz.net>

Copyright (c) 1999-2005 Fernando Henrique Ferraz Pereira da Rosa.
É dada permissão para copiar, distribuir e/ou modificar este documento sob os termos da Licença de Documentação Livre GNU (GFDL), versão 1.2, publicada pela Free Software Foundation;
Uma cópia da licença em está inclusa na seção intitulada "Sobre / Licença de Uso".