

Análise de Sobrevida

Fragilidade

Valeska Andreozzi¹

valeska.andreozzi@fc.ul.pt

&

Marilia Sá Carvalho²

cavalho@fiocruz.br

¹Centro de Estatística e Aplicações da Universidade de Lisboa, Portugal

²Escola Nacional de Saúde Pública e Programa de Computação Científica da Fundação
Oswaldo Cruz, Brasil

Julho, 2008

Programa

1 Fragilidade

Objetivos

- Conceituar fragilidade ou efeito aleatório.
- Indicar adequadamente a inclusão de efeitos aleatórios na modelagem da sobrevida.
- Identificar as distribuições mais utilizadas e as formas de estimação do efeito aleatório.
- Ajustar os modelos utilizando o R.
- Interpretar as estimativas obtidas no ajuste de modelo com efeitos aleatórios.

Algo não medido....

- Características importantes não medidas
- Características não mensuráveis
- Diferenças nos serviços de saúde prestados (utilizados)

heterogeneidade, correlação

Exemplo – Hemodiálise

Estudar a sobrevida de pacientes em hemodiálise tratados no RJ

Indivíduo

- Sexo, idade
- Causa da doença renal
- Características sócio-econômicas (reflete acesso aos serviços de saúde)
- Características nutricionais

Centros de diálise

- Oferta de serviço
- Público x Privado

Exemplo – Hemodiálise

DADOS: Sistema de Informação Ambulatorial de Alto Custo (APAC), 1998 a 2001 (mensal), 44 meses no Estado do Rio de Janeiro

Indivíduo Idade e sexo

Causa da Doença Renal: Hipertensão (base), diabetes, congênitas, renal, outras

Centro No de máquinas: 11 a 19 ou 20 e mais

Número de pacientes: até 50 e maior

Proporção de pacientes idosos, oferta de diálise cíclica, ...

Efeitos Aleatórios

Como levar em consideração na análise: **heterogeneidade** devido a informação não observada e **correlação intra-grupo**?

No exemplo da hemodiálise:

- Características sócio-econômicas e nutricionais não foram observadas
- Heterogeneidade dos pacientes \Rightarrow Modelos com superdispersão

O que fazer?

Efeitos Aleatórios

O que fazer?

- Obter mais informação (nem sempre possível)
- Considerar (**efeitos aleatórios individuais**) para tratar a superdispersão
- Considerar **efeito aleatório de grupo** para tratar da correlação intra-grupo
 - não melhoram a compreensão dos mecanismos que influenciam a sobrevida
 - mas aumentam a confiança nas estimativas sobre o que foi possível medir

Outros Exemplos

- Estudos multicêntricos
- Dados agrupados (mesma família, gêmeos, hospitais)
- Eventos recorrentes
- Estudo pareado

Efeitos Aleatórios

- Atuam multiplicativamente sobre o risco de base, da mesma forma que as covariáveis.
- É possível incluir dois níveis de efeito aleatórios, por exemplo, o do indivíduo e o da unidade, mas a estimação fica complicada.

Modelo de fragilidade

- Z é uma variável aleatória desconhecida que reflete a fragilidade (peculiaridade) do indivíduo ou do grupo
- média igual a 1 e variância ξ
- ξ grande \Rightarrow variabilidade não atribuível a covariável observada
- ξ pequeno \Rightarrow pouca heterogeneidade, não é necessário incluir fragilidade

Modelo de fragilidade

$$\lambda(t|\mathbf{x}) = z\lambda_0(t) \exp(\mathbf{x}\boldsymbol{\beta})$$

sendo

$\mathbf{x} \Rightarrow$ covariáveis

$\boldsymbol{\beta} \Rightarrow$ efeitos fixos

$Z = z \Rightarrow$ efeitos aleatórios

- Fragilidade $z_i > 1 \Rightarrow$ indivíduo (ou grupo) i tendem a experimentar o evento com uma taxa mais rápida que no modelo de Cox básico.
- Variância ξ se a proxima de 0 \Rightarrow modelo de Cox.

Quem é Z – Variável aleatória com distribuição "sensata"

Distribuição da fragilidade precisa ter plausibilidade biológica:

- *gama* – a distribuição da fragilidade entre os que sofreram o evento em um dado tempo e entre os sobreviventes também é *gama* (resultado teórico demonstrável)
- *lognormal* – escolha quando fragilidades relacionadas a covariáveis perdidas x_m : dado $Z = \exp(x_m\beta_m)$, ao aplicarmos o teorema do limite central a $x_m\beta_m$, teremos uma distribuição lognormal para Z .
- Ambas distribuições muito flexíveis, caracterizadas por dois parâmetros, variam desde forma exponencial até forma de sino similar à curva normal

Quem é Z – gama

- γ é parâmetro de forma
- α é parâmetro de escala

$$E(Z) = \gamma\alpha$$
$$Var(Z) = \gamma\alpha^2.$$

Sendo

$$\gamma = 1/\xi$$
$$\alpha = \xi$$

Então

$$E(Z) = 1$$
$$Var(Z) = \xi$$

Quem é Z - gama

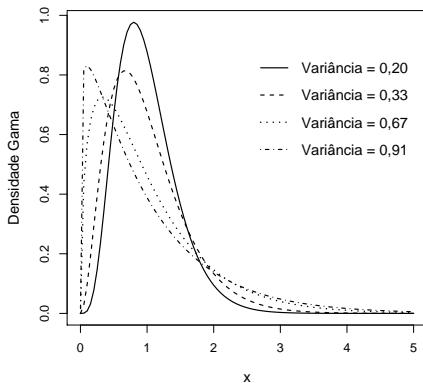


Figura: Distribuição gama com valor esperado 1 e diferentes variâncias ξ

Quem é Z – lognormal

- μ é parâmetro de locação
- σ^2 é parâmetro de escala

$$E(Z) = \exp(\mu + \sigma^2/2)$$

$$Var(Z) = \exp(\sigma^2) - 1$$

Se

$$\mu = -\sigma^2/2e$$

$$\sigma^2 = \log(1 + \xi)$$

Então

$$E(Z) = 1$$

$$Var(Z) = \xi$$

Quem é Z - lognormal

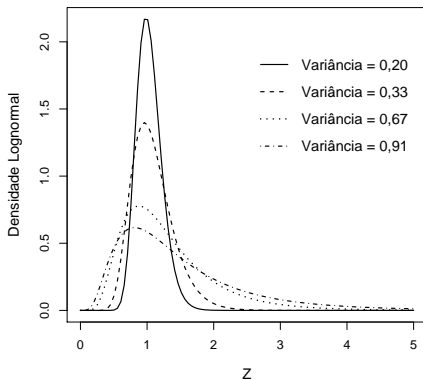


Figura: Distribuição lognormal com valor esperado 1 e diferentes variâncias ξ

Quem é Z

- *lognormal* tem a cauda direita mais pesada do que a *gama* \Rightarrow proporção maior de indivíduos com riscos muito altos
- cauda esquerda é mais leve que *gama* \Rightarrow menor a proporção de indivíduos com riscos muito baixos

Algoritmo EM

- Utiliza-se algoritmo Esperança-Maximização (EM) ou inferência bayesiana.
- Fragilidade tratada como dados não observáveis \Rightarrow estimadas no passo **E** do algoritmo
- $E(Z|t, \delta)$ como estimador.
- No passo **M** são obtidos os valores dos coeficientes de regressão que maximizam a verossimilhança parcial:

$$L_i(\boldsymbol{\beta}) = \left(\frac{z_i \exp(\mathbf{x}_i \boldsymbol{\beta})}{\sum_{j \in R(t_i)} z_j \exp(\mathbf{x}_j \boldsymbol{\beta})} \right)^{\delta_i}$$

Algoritmo EM

- Não implementado no R (nem em qq outro)
- Problemas na estimação da variância ξ
- Bom para estimar fragilidade individual

Verossimilhança Parcial Penalizada

- Implementado no R
- Somente para fragilidade compartilhada (ou multinível)

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(\mathbf{x}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\omega})$$

- \mathbf{x} e \mathbf{U} são matrizes de covariáveis
- $\boldsymbol{\beta}$ correspondem aos p efeitos fixos
- $\boldsymbol{\omega}$ é um vetor que contém q efeitos aleatórios
- o elemento u_{ij} da matriz \mathbf{U} é igual a 1 se o indivíduo i é um membro do grupo j , e é igual a zero caso contrário

Estimação

- Maximização da função de logverossimilhança parcial penalizada em função de β e ω .

$$LPP = l(\beta, \omega) - g(\omega, \theta)$$

- g é a função que impõe penalidades a valores de ω menos desejáveis.
- θ é um parâmetro de ajustamento de g
- Quando:
 - $g(\omega, \theta) = (1/\theta) \sum_i [\omega_i - \exp(\omega_i)] \Rightarrow$ modelo gama de efeitos aleatórios
 - $g(\omega, \theta) = (1/2\theta) \sum_i (\omega_i^2) \Rightarrow$ modelo lognormal, sendo θ a variância dos efeitos ω_i .

Ajuste no R

- θ pode ser fixado diretamente
- pode-se especificar graus de liberdade para o termo dos efeitos aleatórios de 0 (máxima penalidade) a q (nenhuma penalidade, sendo q o número máximo de grupos)
- critério de informação de Akaike (AIC)
- a variância da fragilidade gama pode ser estimada usando-se verossimilhança perfilada, com resultado idêntico ao do algoritmo EM
- fragilidades lognormais pode ser estimada com base numa equação de verossimilhança restrita aproximada

Testes de Hipóteses

- Teste de Wald
- Teste da Razão de Verossimilhança

ambas estatísticas seguem uma distribuição qui-quadrado

Fragilidade no R

- Tudo igual, incluindo o termo *frailty*

```
modelo <- coxph(Surv(inicio,fim,status) ~ covariáveis +  
frailty(grupo), data=dados)
```

- *grupo* \Rightarrow variável que indica o segundo nível, para o qual a fragilidade será estimada

Clínicas de hemodiálise

Objetivo Avaliar qualidade dos centros de hemodiálise

Dados APAC, 1998 a 2001 (mensal), 44 meses no Rio de Janeiro

Variáveis **Indivíduo** Idade e sexo

Causa da Doença Renal: Hipertensão (base), diabetes, congênitas, renal, outras

Centro No de máquinas: 11 a 19 x 20 ou mais

Número de pacientes: até 50 X 50 ou mais

Proporção de pacientes idosos, oferta de diálise cíclica, ...

Problemas Dado prevalente

Estrutura de dependência por centro

Mudança de centro

<http://dengue.procc.fiocruz.br/~sobrevida/dados/dialise.html>

Clínicas de hemodiálise

Objetivo Avaliar qualidade dos centros de hemodiálise

Dados APAC, 1998 a 2001 (mensal), 44 meses no Rio de Janeiro

Variáveis **Indivíduo** Idade e sexo

Causa da Doença Renal: Hipertensão (base), diabetes, congênitas, renal, outras

Centro No de máquinas: 11 a 19 x 20 ou mais

Número de pacientes: até 50 X 50 ou mais

Proporção de pacientes idosos, oferta de diálise cíclica, ...

Problemas Dado prevalente

Estrutura de dependência por centro

Mudança de centro

<http://dengue.procc.fiocruz.br/~sobrevida/dados/dialise.html>

Clínicas de hemodiálise

Objetivo Avaliar qualidade dos centros de hemodiálise

Dados APAC, 1998 a 2001 (mensal), 44 meses no Rio de Janeiro

Variáveis **Indivíduo** Idade e sexo

Causa da Doença Renal: Hipertensão (base), diabetes, congênitas, renal, outras

Centro No de máquinas: 11 a 19 x 20 ou mais

Número de pacientes: até 50 X 50 ou mais

Proporção de pacientes idosos, oferta de diálise cíclica, ...

Problemas Dado prevalente

Estrutura de dependência por centro

Mudança de centro

<http://dengue.procc.fiocruz.br/~sobrevida/dados/dialise.html>

Clínicas de hemodiálise

Objetivo Avaliar qualidade dos centros de hemodiálise

Dados APAC, 1998 a 2001 (mensal), 44 meses no Rio de Janeiro

Variáveis **Indivíduo** Idade e sexo

Causa da Doença Renal: Hipertensão (base), diabetes, congênitas, renal, outras

Centro No de máquinas: 11 a 19 x 20 ou mais
Número de pacientes: até 50 X 50 ou mais
Proporção de pacientes idosos, oferta de diálise cíclica, ...

Problemas Dado prevalente

Estrutura de dependência por centro

Mudança de centro

<http://dengue.procc.fiocruz.br/~sobrevida/dados/dialise.html>

Clínicas de hemodiálise

Objetivo Avaliar qualidade dos centros de hemodiálise

Dados APAC, 1998 a 2001 (mensal), 44 meses no Rio de Janeiro

Variáveis **Indivíduo** Idade e sexo

Causa da Doença Renal: Hipertensão (base), diabetes, congênitas, renal, outras

Centro No de máquinas: 11 a 19 x 20 ou mais

Número de pacientes: até 50 X 50 ou mais

Proporção de pacientes idosos, oferta de diálise cíclica, ...

Problemas Dado prevalente

Estrutura de dependência por centro

Mudança de centro

<http://dengue.procc.fiocruz.br/~sobrevida/dados/dialise.html>

Hemodiálise no R

```
> dial <- read.table("dialmenor.dat",header=T)
> head(dial)
```

	unidade	idade	sexo	inicio	fim	status	tempo	grande	causa
1	128	52	1	26	45	0	19	1	out
2	128	76	0	32	33	0	1	1	out
3	128	61	1	22	24	0	2	1	out
4	128	35	0	7	13	0	6	1	out
5	128	42	0	2	13	0	11	1	out
6	128	44	1	6	30	0	24	1	hip
7	128	41	1	1	6	1	5	1	out
8	128	39	1	10	13	0	3	1	out
9	128	57	0	7	45	0	38	1	out
10	128	71	1	16	33	0	17	1	out

```
> dial$causa <- relevel(dial$causa,ref="hip")
```

Hemodiálise no R

```
> uni.cox <- coxph(Surv(inicio,fim,status)~idade+causa,
  data=dial)
> summary(uni.cox)
```

```
Call:
coxph(formula = Surv(inicio, fim, status) ~idade + causa, data = dial)
n= 861
```

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	1.036	0.965	1.025	1.05
causacon	0.366	2.733	0.147	0.91
causadia	1.887	0.530	1.335	2.67
causaout	1.088	0.919	0.684	1.73
causaren	1.422	0.703	0.956	2.12

```
Rsquare= 0.085 (max possible= 0.929 )
Likelihood ratio test= 76.5 on 5 df, p=4.55e-15
Wald test = 71.6 on 5 df, p=4.82e-14
Score (logrank) test = 76.4 on 5 df, p=4.66e-15
```


Hemodiálise no R

Pacientes podem ter um risco comum associado à unidade de diálise em que são assistidos, a unidade foi incluída como variável *dummy*.

```
> summary(coxph(Surv(inicio,fim, status)~
  idade+causa+factor(unidade),data=dial))
```

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	1.03e+00	9.68e-01	1.0220	1.04
causacon	4.41e-01	2.27e+00	0.1354	1.44
causadia	1.78e+00	5.62e-01	1.2371	2.56
causaout	1.20e+00	8.34e-01	0.6694	2.15
causaren	1.22e+00	8.19e-01	0.8039	1.86
factor(unidade)217	8.84e-01	1.13e+00	0.0995	7.85
[...]				
factor(unidade)641	1.93e+00	5.17e-01	0.6901	5.42
factor(unidade)741	5.19e+00	1.93e-01	1.4148	19.05
[...]				
factor(unidade)5681	8.70e-07	1.15e+06	0.0000	Inf
factor(unidade)5692	3.34e+00	3.00e-01	0.2820	39.49

Rsquare= 0.22 (max possible= 0.929)

Hemodiálise no R

- Observe unidade com limite superior igual a infinito (sem ocorrência de óbito)

Warning message:

```
Loglik converged before variable 23 ; beta may be infinite.  
in: fitter(X, Y, strats, offset, init, control,  
weights = weights,
```

- *Rsquare* maior que o modelo sem as unidades (de 0.085 para 0.22)
- Neste modelo não podemos testar a hipótese de que a unidade número 741 tem um risco 5 vezes maior que a primeira unidade (referência) devido a proporção elevada de idosos, pois não é possível incluir covariáveis no nível do centro. (pois os valores não mudariam em relação a variável unidade)

Hemodiálise no R

Incluindo fragilidade

```
> summary(uni.gama<- coxph(Surv(inicio,fim, status) ~
  idade + causa + frailty(unidade), data=dial))
```

Call:

```
coxph(formula = Surv(inicio, fim, status) ~ idade + causa +
  frailty(unidade), data = dial)
```

n= 861

	coef	se(coef)	se2	Chisq	DF	p
idade	0.0325	0.00555	0.00552	34.26	1.0	4.8e-09
causacon	-0.9333	0.51228	0.48190	3.32	1.0	6.8e-02
causadia	0.5980	0.18320	0.18191	10.66	1.0	1.1e-03
causaout	0.1674	0.29820	0.29033	0.32	1.0	5.7e-01
causaren	0.2462	0.21095	0.20909	1.36	1.0	2.4e-01
frailty(unidade)				123.10	15.6	0.0e+00

Hemodiálise no R

Incluindo fragilidade (continuação)

[...]

	exp(coef)	exp(-coef)	lower .95	upper .95
idade	1.033	0.968	1.022	1.04
causacon	0.393	2.543	0.144	1.07
causadia	1.819	0.550	1.270	2.60
causaout	1.182	0.846	0.659	2.12
causaren	1.279	0.782	0.846	1.93

Iterations: 10 outer, 27 Newton-Raphson

Variance of random effect= 0.734 I-likelihood = -1059.8
 Degrees of freedom for terms= 1.0 3.8 15.6
 Rsquare= 0.216 (max possible= 0.929)
 Likelihood ratio test= 210 on 20.4 df, p=0
 Wald test = 54.4 on 20.4 df, p=6.09e-05

Hemodiálise no R

Interpretação

uni.cox

estima o efeito **médio populacional** da doença de base no risco de morte por insuficiência renal.

compara-se duas amostras aleatórias, por exemplo, pacientes cuja doença de base é diabetes e os hipertensos.

uni.gama

estima o efeito da doença de base dentro de uma **unidade típica**.

estima o risco relativo da diabetes dado que os pacientes são assistidos na **mesma unidade típica**, isto é, que possui efeito aleatório igual a zero.

Estimação

- *Iterations: 10 outer, 27 Newton-Raphson*
27 iterações necessárias para a variância dos efeitos aleatórios em cada iteração do efeitos fixos (total 10)

- Modelo não é bem estimado quando...

Warning message:

```
Inner loop failed to converge for iterations 6 in:
coxpenal.fit(X, Y, strats, offset, init = init,
control, weights = weights,
```

- Porque utiliza-se como padrão o argumento *sparse=TRUE*, que utiliza somente a diagonal da matriz de informação para facilitar os cálculos matriciais. Neste caso, recomenda-se ajustar o modelo sem usar essa opção, com a seguinte sintaxe:

```
> uni.gama<- coxph(Surv(inicio,fim,status) ~
idade+causa+frailty(unidade, sparse= FALSE),
data=dial)
```

Comparando os modelos

H_0 : efeitos aleatórios = 0

- Estatística de Wald segue distribuição qui-quadrado (*frailty(unidade)* = 123.10 com 15.6 gl);
- Razão de verossimilhança segue distribuição qui-quadrado com um grau de liberdade

$$2 \times \text{loglik}(\text{modelo.frail}) - \text{loglik}(\text{modelo.nofrail})$$

$$2 \times (-1059,8 - (-1099,963)) = 80,33$$

$\text{loglik}(\text{modelo.frail})$ = logverossimilhança parcial integrada fora os termos de fragilidade (*I-likelihood*)

$\text{loglik}(\text{modelo.nofrail})$ = logverossimilhança do modelo sem fragilidade

- Medida global de ajuste:

Likelihood ratio test = 210 on 20.4 df, $p=0$

Não disponível para modelos Gauss

Estimando os modelos

- distribuição gama \Rightarrow verossimilhança perfilada para θ (*method=EM*)
- distribuição normal \Rightarrow máxima verossimilhança restrita aproximada (*method=REML*)
- minimização AIC (Critério de Akaike)
- Não há método para identificar se uma distribuição é melhor que a outra, ou um método de estimação é melhor que o outro
- Avaliar consistência:
 - mesmas unidades com risco diferenciado
 - estimativas dos efeitos fixos e intervalos de confiança semelhantes

Estimando os modelos

- distribuição gama \Rightarrow verossimilhança perfilada para θ (*method=EM*)
- distribuição normal \Rightarrow máxima verossimilhança restrita aproximada (*method=REML*)
- minimização AIC (Critério de Akaike)
- **Não** há método para identificar se uma distribuição é melhor que a outra, ou um método de estimação é melhor que o outro
- Avaliar consistência:
 - mesmas unidades com risco diferenciado
 - estimativas dos efeitos fixos e intervalos de confiança semelhantes

Estimando os modelos

- distribuição gama \Rightarrow verossimilhança perfilada para θ (*method=EM*)
- distribuição normal \Rightarrow máxima verossimilhança restrita aproximada (*method=REML*)
- minimização AIC (Critério de Akaike)
- **Não** há método para identificar se uma distribuição é melhor que a outra, ou um método de estimação é melhor que o outro
- Avaliar consistência:
 - mesmas unidades com risco diferenciado
 - estimativas dos efeitos fixos e intervalos de confiança semelhantes

Comparando os modelos

	Efeito Gama (EM)			Efeito Gama (AIC)		
	$exp(\beta)$.95Inf	.95Sup	$exp(\beta)$.95Inf	.95Sup
idade	1.033	1.022	1.04	1.033	1.022	1.04
causacon	0.393	0.144	1.07	0.388	0.144	1.04
causadia	1.819	1.270	2.60	1.827	1.277	2.61
causaout	1.182	0.659	2.12	1.174	0.656	2.10
causaren	1.279	0.846	1.93	1.293	0.856	1.95

	Efeito Gauss (REML)			Efeito Gauss (AIC)		
	$exp(\beta)$.95Inf	.95Sup	$exp(\beta)$.95Inf	.95Sup
idade	1.033	1.022	1.04	1.033	1.022	1.04
causacon	0.419	0.157	1.12	0.415	0.157	1.09
causadia	1.814	1.267	2.60	1.820	1.273	2.60
causaout	1.168	0.660	2.07	1.161	0.658	2.05
causaren	1.247	0.824	1.89	1.254	0.829	1.90

Comparando os ajustes

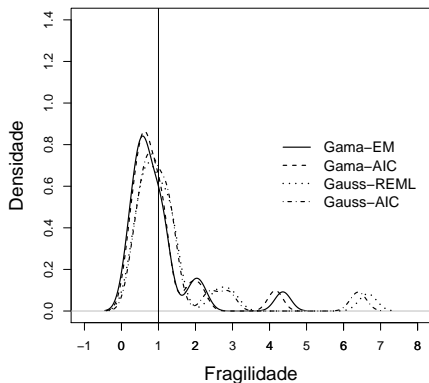
Medidas de ajuste	Cox Clássico	Gama EM	Gama AIC	Gauss EM	Gauss AIC
Variância do efeito aleatório	—	0,734	0,55	0,697	0,55
Wald Test ^{a,c}	—	123,10 (15,6)	118,81 (14,8)	128,26 (14,7)	123,82 (14,1)
Rsquare	0,085	0,216	0,215	0,216	0,215
Log-verossimil, <i>I-likelihood</i>	-1100	-1033	-1034	-1033	-1034
Qualidade do Ajuste ^{b,c}	76,5 (5)	210 (20,4)	208 (19,5)	210 (19,5)	209 (18,8)

^a Teste para a fragilidade (graus de liberdade entre parênteses)

^b Teste da razão de verossimilhança (graus de liberdade entre parênteses)

^c p-valor <0,001

Comparando os efeitos



Observar o escopo da fragilidade – gama chega perto do zero, gauss com valores muito altos e flutuação

Comparando os efeitos

