

Modelo paramétrico flexível para dados grupados e censurados: uma aplicação a sobrevida de idosos em Botucatu-SP

Liciane Vaz de Arruda Silveira ¹, José Eduardo Corrente ²

Introdução

O envelhecimento da população é um fenômeno mundial de importância quando se descreve padrões das doenças crônicas. Segundo a Organização das Nações Unidas (ONU), em seu último relatório técnico "Previsões sobre a população mundial", elaborado pelo Departamento de Assuntos Econômicos e Sociais, nos próximos 43 anos o número de pessoas com mais de 60 anos de idade será três vezes maior do que o atual. Os idosos representarão um quarto da população mundial projetada, ou seja, cerca de 2 bilhões de indivíduos (no total de 9,2 bilhões). No critério da Organização Mundial da Saúde (OMS), é considerado idoso o habitante de país em desenvolvimento com 60 anos ou mais e o habitante de país desenvolvido com mais de 65 anos (FELIX, 2012).

O envelhecimento populacional, tornou o estudo dos idosos da maior importância para o planejamento social da população que desponta neste século com este novo perfil.

Em relação aos riscos conhecidos de mortalidade entre idosos, uma vez retirada a influência da idade no tempo de sobrevivência, o sexo é o primeiro a ser lembrado, já que a sobrevivência maior do sexo feminino é verificada em quase todos os países americanos, asiáticos e europeus (CHOR et al, 1992). Antes de ser o sexo propriamente um fator de risco, o que acontece para essa atual geração, é que a mortalidade tem comportamento diferente nos dois sexos e por isso sempre é melhor sua análise separada. Das causas de mortalidade entre idosos, doenças cardiovascular e neoplasias malignas lideram como as mais prevalentes (RUIZ et al, 2003).

No município de Botucatu, São Paulo foram conduzidos inquéritos populacionais nos anos de 2003, 2006 e 2012, com o objetivo de avaliar a sobrevida de idosos (RUIZ et al, 2003).

Os dados obtidos com este estudo apresentaram muitos empates (mais que 25 por cento), sendo assim, resolveu-se tratá-los como grupados (CHALITA et al, 2002).

O objetivo deste trabalho foi desenvolver modelos flexíveis para dados grupados e censurados e aplicá-los aos dados da coorte de idosos, considerando como evento a morte por todas as causas.

O uso destes modelos torna-se uma alternativa interessante ao modelo de riscos proporcionais de Cox podendo resultar em melhores ajustes, principalmente, quando a suposição de riscos proporcionais não for satisfeita. Os modelos de Cox e os flexíveis foram aplicados aos dados e comparados os seus ajustes através do critério de informação de Akaike (AIC).

A análise foi feita separadamente para cada sexo, pois os segmentos feminino e masculino da população apresentam comportamentos bastante distintos com relação à sobrevida.

¹Depto. de Bioestatística, IBB,Unesp. e-mail: liciana.silveira@unesp.br

²Depto. de Bioestatística, IBB,Unesp. e-mail: jose.corrente@unesp.br

Métodos estatísticos

Modelos para dados agrupados e censurados

Considere que os tempos de vida, T_l , $l = 1, \dots, n$, onde n é o tamanho da amostra, são agrupados em k intervalos, $I_i = [a_{i-1}, a_i)$, $i = 1, \dots, k$, com $0 = a_0 < a_1 < \dots < a_k = \infty$ e, assuma que todas as censuras ocorrem no final do intervalo. Seja D_i o conjunto dos indivíduos que falharam no intervalo I_i ; R_i , o conjunto dos indivíduos sob risco no início de I_i , e Δ_{li} , a variável indicadora de falha do l -ésimo indivíduo em I_i , que assume valor zero quando a l -ésima observação é censurada no i -ésimo intervalo, e um em caso contrário, para $i = 1, \dots, k$, $l = 1, \dots, n$. Assim, pode-se construir a função de verossimilhança para tabela de vida com variáveis regressoras (\mathbf{x}), dada por:

$$\begin{aligned} L &= \prod_{i=1}^k \left[\prod_{l \in D_i} p_i(\mathbf{x}_l) \prod_{l \in R_i - D_i} (1 - p_i(\mathbf{x}_l)) \right] \\ &= \prod_{i=1}^k \prod_{l \in R_i} (p_i(\mathbf{x}_l))^{\Delta_{li}} (1 - p_i(\mathbf{x}_l))^{(1 - \Delta_{li})}, \end{aligned} \quad (1)$$

em que $p_i(\mathbf{x}_l)$ é a probabilidade do l -ésimo indivíduo falhar até a_i dado que sobreviveu até a_{i-1} na presença das variáveis regressoras \mathbf{x}_l .

A função (1) corresponde à função de verossimilhança para variáveis com distribuição de Bernoulli, já que a variável em questão, Δ_{li} , é binária (isto é, ocorreu a falha, ou não, no intervalo I_i).

A probabilidade $p_i(\mathbf{x}_l)$ pode ser modelada considerando o modelo de riscos proporcionais de Cox e os modelos paramétricos flexíveis.

Modelo de Cox para dados agrupados e censurados

Adotando o modelo de riscos proporcionais de Cox, para o tempo de vida T , tem-se:

$$p_i(\mathbf{x}_l) = 1 - \left[\frac{S_0(a_i)}{S_0(a_{i-1})} \right]^{\exp(\boldsymbol{\beta}' \mathbf{x}_l)}. \quad (2)$$

A estimação dos parâmetros é feita pelo método da máxima verossimilhança.

Modelo paramétrico flexível

Um método alternativo mais flexível comumente usado é o de ROYSTON P & PARMAR (2002). Este é modelado na escala log do risco cumulativo do modelo de riscos proporcionais de Cox ($H(t; \mathbf{x})$), PH

$$\log(H(t|\mathbf{x}_l)) = \log(-\log(S(t|\mathbf{x}_l))) = \log(H_0(t)) + \boldsymbol{\beta}' \mathbf{x}_l = s(u) + \boldsymbol{\beta}' \mathbf{x}_l \quad (3)$$

em que $H_0(t) = \int_0^t h_0(u)du$ é a função de risco básica acumulada.

A *spline* básica é especificada com base no log da função de risco acumulada de um modelo de riscos proporcionais ou com base log da função de chances proporcionais geral de um modelo de chances proporcionais. A relação linear com o log do tempo é relaxada através do uso de *splines* cúbicas restritas. A natureza restrita das *splines* força a função ser linear além dos nós de fronteira, k_{min} e k_{max} . A escolha padrão para os nós de fronteira

são o máximo e o mínimo dos tempos de sobrevivência não censurados. As funções *spline* são forçados a se juntar aos m nós internos, k_1, \dots, k_m , com primeira e segunda derivadas contínuas. As locações do nó interno são, geralmente, baseados em percentis dos tempos de sobrevivência não censurado. Pode-se, portanto, escrever uma função *spline* cúbica restrita de $\log(t)$, com vetor nó, k_0 e $u = \log(t)$.

$$s(u|\gamma, k_0) = \gamma_0 + \gamma_1 \nu_1 + \dots + \gamma_{m+1} \nu_{m+1} \quad (4)$$

com vetor de parâmetros γ e derivadas ν_j (conhecida como função base), em que $\nu_1 = u$

$$\nu_j = (u - k_j)_+^3 - \lambda_j (u - k_{min})_+^3 - (1 - \lambda_j)(u - k_{min})_+^3 \quad (5)$$

em que $j = 2, \dots, m+1$, $(u - k_j)_+^3$ é igual a $(u - k_j)^3$ para valores positivos e igual a

$$\lambda_j = (k_{max} - k_j)/(k_{max} - k_{min})$$

para valor 0.

Substituindo 4 em 3 tem-se (CROWTHER et al, 2014):

$$\log(H(t|\mathbf{x}_l)) = s(\log(t)|\gamma, k_0) + \beta' \mathbf{x}_l$$

Assim, pode-se obter as funções de risco e de sobrevivência, respectivamente:

$$h(t|\mathbf{x}_l) = \left[\frac{1}{t} \frac{ds(\log(t)|\gamma, k_0)}{d \log(t)} \right] \exp(\beta' \mathbf{x}_l)$$

$$S(t|\mathbf{x}_l) = \exp(-\exp(s(\log(t)|\gamma, k_0) + \beta' \mathbf{x}_l)) \quad (6)$$

Dada a natureza inteiramente paramétrica do modelo especificado, as derivadas da função spline, necessária na definição da função de risco, são facilmente calculadas (CROWTHER et al, 2014).

Modelo paramétrico flexível para dados grupados e censurados

Adotando o modelo paramétrico flexível (6), para o tempo de vida T, tem-se:

$$p_i(\mathbf{x}_l) = 1 - \left[\frac{S(a_i|\mathbf{x}_l)}{S(a_{i-1}|\mathbf{x}_l)} \right] \quad (7)$$

Assim,

$$p_i(\mathbf{x}_l) = 1 - \left[\frac{\exp(-\exp(s(\log(a_i)|\gamma, k_0) + \beta' \mathbf{x}_l))}{\exp(-\exp(s(\log(a_{i-1})|\gamma, k_0) + \beta' \mathbf{x}_l))} \right] \quad (8)$$

Os parâmetros são estimados pelo método da máxima verossimilhança utilizado o processo iterativo de Newton Raphison.

Resultados

O uso do modelo paramétrico flexível para dados grupados e censurados torna-se uma alternativa interessante ao modelo de riscos proporcionais de Cox, podendo resultar em melhores ajustes. Os modelos foram aplicados aos dados e comparados os seus ajustes através do critério de informação de Akaike (AIC). A análise foi feita separadamente para cada sexo, pois os segmentos feminino e masculino da população apresentam comportamentos bastante distintos com relação à sobrevida.(Tabela 1.)

Tabela 1. Valores de AIC para os modelos estudados.

Sexo	Cox	Paramétrico Flexível
feminino	929,39	509,6873
masculino	908,87	452,15

Comparando os valores dos critério de AIC, o modelo flexível, considerando cada sexo, se ajustou melhor aos dados. O modelo de Cox tem como pressuposição riscos proporcionais, e observou-se nestes dados que uma grande parte das covariáveis não apresentavam riscos proporcionais, sendo este um possível motivo de ter sido obtido um pior ajuste.

As covariáveis que foram significativas no ajuste do modelo Flexível foram: no caso feminino, idade superior a 69 anos ($p < 0,0001$), presença de diabetes ($p = 0,0029$) e ter doença do aparelho circulatório ($p < 0,0001$) e, no caso masculino, idade superior a 69 anos ($p = 0,0002$), presença de diabetes ($p = 0,0008$), ter doença do aparelho circulatório ($p < 0,0001$) e possuir outra fonte de renda ($p = 0,0259$). (Tabelas 2 e 3.)

Tabela 2. Ajuste do Modelo Flexível - Gênero feminino.

	Estimativa	Erro padrão	valor de p
idade > 69	0,5438	0,1363	<0,0001
diabetes	0,4870	0,1633	0,0029
outra renda	0,3735	0,1930	0,0530
ap. circ	1,2955	0,1499	<0,0001

Tabela 3. Ajuste do Modelo Flexível - Gênero masculino.

	Estimativa	Erro padrão	valor de p
idade > 69	0,5886	0,1588	0,0002
diabetes	1,0055	0,3008	0,0008
outra renda	0,4798	0,2154	0,0259
ap. circ	1,5823	0,1904	<0,0001

Nota-se que no caso do gênero feminino, ter outras fonte de renda não foi significativo para sobrevida dos idosos.

Agradecimentos

A Profa. Adjunta Tânia Ruiz por ceder os dados e a Fapesp pelo apoio financeiro.

Referencias Bibliográficas

BRESLOW, N. Covariance analysis of censored survival data. *Biometrics*, Washington, v.30, 89-90, 1974.

CHALITA, L.V.A.S., COLOSIMO, E. A. ; DEMÉTRIO, C.G. B. Likelihood Approximations And Discrete Models For Tied Survival Data. *Communications in Statistics: Theory and Methods*, v. 31, 1215-1229, 2002.

CHOR D, DUCHAIDE M.P., JOURDAN A.M.F. Diferencial de mortalidade em homens e mulheres em localidade da região sudeste, Brasil - 1960, 1970 e 1980. *Rev. Saúde Pública*, v. 26, 246-55, 1992 .

COLLETT, D. Modelling Survival Data in Medical Research. *Chapman and Hall*, London-2ed. 2003. 500p. .

COLOSIMO, E.A. & GIOLO, S.R. Análise de Sobrevida Aplicada. *Edgard Blucher*, São Paulo. 2006. 392p.

COX, D.R. Regression Models and Life-Tables. *Journal of the Royal statistical Society-B*, v.34, 187-220, 1972.

CROWTHER M.J., LOOK M.P., RILEY R.D. Multilevel mixed effects parametric survival models using adaptive Gauss-Hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Statistics in Medicine*, v.33, n.22, 3844-3858, 2014.

FELIX, J.S. "Economia da Longevidade: uma revisão da bibliografia brasileira sobre o envelhecimento populacional". 2012. 17p.

JÓIA, L.C. & RUIZ, T. Inquérito populacional sobre estilo e a qualidade de vida na Terceira Idade. *Revista de Atenção Básica Em Saúde*, v.2, 114-130, 2006.

KALBFLEISCH, J.D. & PRENTICE, R.L. The Statistical Analysis of Failure Time Data. *John Wiley & Sons, Inc.*, New York.2002. 447p.

KALBFLEISCH, J.D. & PRENTICE, R.L. Marginal Likelihood based on Cox's regression and life model. *Biometrika*, v.60, 267-278. 1973.

LAWLESS, J.F. Statistical Models and Methods for Lifetime Data. *Wiley*, New York. 2003. 579p.

PRENTICE, R.L. & GLOECKLER, L.A. Regression Analysis of Grouped Survival Data With Application to Breast Cancer Data. *Biometrics*, v.34, 57-67. 1978.

ROYSTON P. & PARMAR M.K.B. Flexible parametric proportional hazards and proportional odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*. v.21, n.15, 2175-2197, 2002.

RUIZ, T., CHALITA, LÍCIANA V.A.S., BARROS, M.B.A. Estudo de Sobrevida de um Coorte de Pessoas de 60 anos e mais no Município de Botucatu (SP). *Revista Brasileira de Epidemiologia*, v. 6, n.3, 227-236, 2003.