

ANÁLISE DE SOBREVIVÊNCIA COM A INCORPORAÇÃO DA INFORMAÇÃO ESPACIAL. Karina Lima Guimarães Firmino, Mário Hissamitsu Tarumoto. – Probabilidade e Estatística – Estatística - Departamento de Matemática, Estatística e Ciência da Computação – Faculdade de Ciências e Tecnologia – Campus de Presidente Prudente.

A análise de sobrevivência é uma área de pesquisa que cresceu muito nos últimos tempos, pela possibilidade de incorporar ao estudo observações parciais (censuras) a respeito do evento de interesse. Este fato é comum em muitas aplicações na área de ciências biológicas, ou ainda em outras áreas, cujo interesse é reduzir o tempo, e conseqüentemente o custo do experimento. Em outro contexto, a informação espacial também é de grande interesse dos pesquisadores: o tempo de reincidência do menor infrator pode estar relacionado ao local de moradia do mesmo e a expectativa de indivíduos pode estar relacionada às condições de vida, que por sua vez, depende de condições de infra-estrutura do bairro em que vive. Nesse sentido, a inclusão da informação espacial ao modelo é importante para verificar e modelar a influência do local no tempo de sobrevida.

Li e Lin (2005) propuseram uma classe de modelos de transformação normal semiparamétrica para dados de sobrevivência espacialmente correlacionados. Nesta classe de modelos, assume-se que os tempos de sobrevivência seguem marginalmente o modelo de riscos proporcionais de Cox, e a distribuição conjunta é definida pelo modelo normal multivariado com uma estrutura de correlação espacial, obtida por meio de transformações dos tempos em variáveis aleatórias normalmente distribuídas. Com isso, é possível ajustar os tempos de sobrevivência num modelo que comporte os parâmetros de dependência e cujos coeficientes regressores tenham interpretações populacionais.

O modelo de riscos proporcionais de Cox para a variável aleatória que define os tempos de sobrevivência de indivíduos observados, sujeitos à falha ou censura aleatória, é dado por:

$$I\{t|Z_i(.)\} = I_0(t) \exp\{\mathbf{b}' Z_i(t)\},$$

sendo \mathbf{b} o vetor de coeficientes de regressão, Z_i o vetor de covariáveis e $I_0(t)$ a função-base do risco.

Considere também $\Lambda_0(t) = \int_0^t I_0(s) ds$ como sendo a função-base do risco acumulada, e

$\Lambda_i(t) = \int_0^t I_i(s|Z_i) ds$ a função risco acumulada. Pode-se observar que $\Lambda_i(T_i)$ segue a

distribuição exponencial unitária, e fazendo-se uma transformação do tipo probito nessa função, é possível construir uma verossimilhança conjunta flexível para dados de sobrevivência espaciais, mantendo a distribuição marginal. Desta forma, assume-se então que T^* segue uma distribuição conjunta normal multivariada (LI e LIN apud WINKLER, 1995), ou seja:

$$T^* = \{T_i^*, i=1, 2, \dots, m\} \sim N_m(0, \Gamma),$$

sendo que T_i^* segue uma distribuição normal padrão e é dado pela expressão:

$$T_i^* = \Phi^{-1} \{1 - \exp[-\Lambda_i(T_i)]\},$$

em que $\Phi(.)$ representa a distribuição normal padrão acumulada.

A matriz de covariâncias Γ é uma matriz positiva definida com os elementos das diagonais (\mathbf{q}_{ii}) iguais a 1 e $\mathbf{q}_{ij} = \mathbf{q}_{ji}(a_i, a_j) = \mathbf{r}(T_i^*, T_j^*)$ os elementos da i -ésima linha e j -ésima coluna, que quantificam a correlação entre os pares de tempos T_i^* e T_j^* , $\mathbf{q}_{ij} \in I = [-1, 1]$. Diversos modelos podem ser utilizados para estudar a dependência espacial se um modelo paramétrico for especificado para estes elementos, como a função Matérn, a exponencial, e a esférica.

A estimação dos coeficientes de regressão e dos parâmetros de dependência espacial é feita através da solução de equações de estimação semiparamétrica, tendo em vista as dificuldades numéricas em trabalhar com a função de verossimilhança, que envolve uma dimensão de integrais do tamanho da amostra.

O método proposto foi utilizado no estudo de asma, numa região de Boston, na tentativa de associar a doença a fatores de risco, considerando também os endereços residenciais dos pacientes (latitude e longitude da residência). A coleta dos dados foi feita numa clínica de saúde, sendo possível identificar a idade na qual a doença teve seu início, e sendo assim, o modelo marginal de Cox para os tempos de início de asma é definido desta forma:

$$I(t) = I_0(t) \exp\{\mathbf{b}_L \times LRI + \mathbf{b}_M \times MEVAST + \mathbf{b}_C \times LOGMCOT\},$$

sendo LRI o índice de baixa respiração (Low Respiratory Index), que assume valores de 0 a 16 (índices altos indicam mau funcionamento respiratório), MEVAST uma covariável binária (1 se o paciente já teve asma, e 0 se nunca teve), e LOGMCOT a log-transformação de níveis maternos de *cotinine*.

A análise de dados foi feita para um total 606 pacientes, sendo que até o final do estudo foram observados 74 eventos. A estrutura de correlação espacial utilizada foi a função Matérn:

$$\mathbf{r}(d, \mathbf{a}) = \frac{\mathbf{a}_1}{2^{\mathbf{a}_3-1} \Gamma(\mathbf{a}_3)} (2\mathbf{a}_2 \sqrt{\mathbf{a}_3} d) K_{\mathbf{a}_3}(2\mathbf{a}_2 \sqrt{\mathbf{a}_3} d), 0 \leq \mathbf{a}_1 \leq 1, \mathbf{a}_2 \geq 0, \mathbf{a}_3 \geq 0,$$

em que $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, \mathbf{a}_1 é o parâmetro de escala local ($\mathbf{a}_1 = \lim_{d \rightarrow 0+} \mathbf{r}(d, \mathbf{a})$), \mathbf{a}_2 mede a diminuição global da correlação conforme a distância entre as observações aumenta, \mathbf{a}_3 é o parâmetro de suavização que caracteriza o comportamento da função quando próxima da origem, $K_{\mathbf{a}_3}(\cdot)$ é a função Bessel modificada de segundo tipo de ordem \mathbf{a}_3 (LI e LIN apud ABRAMOVITZ e STEGUN, 1965). A tabela 1 apresenta as estimativas para os parâmetros de interesse, de acordo com três valores fixados para o suavizador \mathbf{a}_3 .

As estimativas dos coeficientes de regressão indicaram que a única covariável significativa para o início de asma foi o índice de baixa respiração (LRI). Isso significa que crianças com má respiração estão mais propensas a desenvolver asma. A idade e níveis maternos de *cotinine* não são significativos no modelo. As estimativas para os parâmetros de dependência espacial ($\mathbf{a}_1, \mathbf{a}_2$) indicaram pequena correlação; o primeiro mede a correlação entre sujeitos numa proximidade geográfica, e de acordo com os autores, esta estimativa indicou baixa correlação. A interpretação do segundo parâmetro indicou que a correlação entre indivíduos diminui cerca de 90% a cada *km* aumentado na distância.

Tabela 1. Estimativas e erros padrões para os parâmetros de dependência espacial e coeficientes de regressão para determinados valores para o parâmetros de suavização para os tempos de ocorrência de asma

Parâmetro	$a_3=0,5$		$a_3=1$		$a_3=1,5$	
	Estimativa	Erro padrão	Estimativa	Erro padrão	Estimativa	Erro padrão
b_L	0,3121	0,044	0,3118	0,043	0,3124	0,0432
b_M	0,2662	0,3314	0,2644	0,3289	0,2676	0,3283
b_C	0,0294	0,1394	0,02531	0,127	0,0277	0,1288
a_1	1,68E(-3)	9,8E(-3)	0,74E(-3)	0,5E(-2)	0,72E(-3)	5,5E(-3)
a_2	2,2977	4,974	2,1917	4,7945	1,8886	6,5005

Fonte: Li e Lin (2005)

Referências Bibliográficas

ABRAMOVITZ, M., STEGUN I.A. (eds). **Handbook of Mathematical Functions**. New York: Dover Publications, 1965.

LI, Y.; LIN, X. Semiparametric normal transformation models for spatially correlated survival data. **Harvard University Biostatistics Working Paper Series**, 27, 2005.

WINKLER, G. **Image analysis, random fields, and dynamic Monte Carlo methods: a mathematical introduction**. New York: Springer, 1995.

FAPESP