



UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Emerson Rigoni

Henrique Aparecido Laureano

APLICAÇÃO DE MODELOS MULTIESTADOS PARA ANÁLISE DE DADOS DE SOBREVIVÊNCIA

Projeto de Pesquisa apresentado à disciplina
Laboratório A do Curso de Graduação em Estatística
da Universidade Federal do Paraná, como requisito
para elaboração do Trabalho de Conclusão de Curso.

Orientadora: Profa. Dra. Suely Ruiz Giolo

**CURITIBA
2015**

SUMÁRIO

1 INTRODUÇÃO	3
2 OBJETIVOS	4
2.1 Objetivo Geral	4
2.2 Objetivos Específicos	4
3 MATERIAIS E MÉTODOS	4
3.1 Materiais	4
3.1.1 Conjunto de Dados	4
3.1.2 Recursos Computacionais.....	7
3.2 Métodos	8
3.2.1 Modelos Multiestados - Funções Básicas.....	9
4 CRONOGRAMA DE ATIVIDADES	11
REFERÊNCIAS	12

1 INTRODUÇÃO

Análise de sobrevivência ou de confiabilidade é um conjunto de técnicas e métodos estatísticos utilizados para o estudo de dados relativos ao tempo até a ocorrência de um evento de interesse. O termo análise de sobrevivência é comumente utilizado na área médica e, neste caso, o evento de interesse é, entre outros: a morte, a cura ou a recorrência de um sintoma ou doença. Na área industrial, o termo mais utilizado é confiabilidade e, neste caso, o evento de interesse é: a falha, o defeito ou a recorrência de um defeito. A não ocorrência do evento caracteriza o que se denomina censura. Ou seja, censura é a observação parcial da resposta causada pela perda de acompanhamento por qualquer motivo que não o evento de interesse em estudo. Para estudos nas áreas citadas, a variável resposta é o tempo até a ocorrência do evento de interesse, denotada por T , a qual é não-negativa e usualmente contínua.

Os métodos estatísticos tradicionalmente utilizados para a análise de dados de sobrevivência, como por exemplo o estimador de Kaplan-Meier (KAPLAN; MEIER, 1958) e o modelo de Cox (COX, 1972), consideram a existência de apenas um tipo de desfecho, limitando sua aplicação quando se quer estudar as causas do evento de interesse. Sendo assim, para analisar as diferentes causas do evento de interesse, foram propostos os modelos multiestados ou uma subcategoria destes, que são os modelos de riscos competitivos.

Os modelos multiestados estendem a análise de sobrevivência clássica e são utilizados, segundo Putter *et al.* (2010), para duas finalidades:

- 1) Melhorar a compreensão do processo biológico da doença/recuperação de um paciente explicando como determinados prognósticos (covariáveis) influenciam nas diferentes fases do processo.
- 2) Predição: permitem aos médicos previsões mais precisas, do que os modelos clássicos de sobrevivência e, também, ajustar estas previsões ao longo do tempo, incorporando eventos intermediários.

Segundo Putter *et al.* (2010), apesar da aplicabilidade dos modelos multiestados em análise de sobrevivência, seu uso na área médica é limitado principalmente devido a: dificuldade de comunicação e entendimento dos modelos por parte dos médicos; falta de

softwares flexíveis e de uso amigável e; limitações dos pacotes existentes no R e em outros softwares.

2 OBJETIVOS

2.1 Objetivo Geral

Estudar e aplicar modelos multiestados para análise de dados de sobrevivência.

2.2 Objetivos Específicos

1. Revisar a literatura no que diz respeito à aplicação e modelagem multiestados no contexto de dados de sobrevivência;
2. Revisar os pacotes disponíveis no *software* R (R Core Team, 2015) para estimação e diagnóstico de modelos para dados de sobrevivência utilizando a abordagem multiestados;
3. Ajustar modelos multiestados a dois conjuntos de dados de sobrevivência, a princípio;
4. Comparar e discutir os resultados dos modelos ajustados.

3 MATERIAIS E MÉTODOS

3.1 Materiais

3.1.1 Conjuntos de Dados

O primeiro conjunto de dados que se pretende utilizar neste trabalho se refere a pacientes com doença arterial coronariana multiarterial, angina¹ estável e função ventricular² preservada. O tempo de acompanhamento foi de 10 anos e o objetivo foi

¹ Angina do peito: Dor constritiva intensa, no peito, frequentemente irradiada para o braço esquerdo, provocada por isquemia do miocárdio e resultante, quase sempre, de moléstia coronariana.

² Ventrículo: Cada uma das duas cavidades do coração, uma direita e outra esquerda, de grossas paredes musculares. Pela contração do ventrículo direito se injeta sangue na artéria pulmonar, e pela contração do esquerdo, na artéria aorta.

comparar três tratamentos: Intervenção Coronária Percutânea (ICP)³, Cirurgia de Revascularização Miocárdica (CRM)⁴, e Tratamento Médico (TM). Os eventos de interesse foram: mortalidade, infarto do miocárdio com onda Q⁵ ou angina refratária (resistente) que necessitasse de revascularização. Em uma única instituição, 611 pacientes foram randomizados para: CRM (203 pacientes), ICP (205 pacientes), ou TM (203 pacientes).

Segundo Hueb (2010) o tratamento de pacientes com Doença Arterial Coronariana (DAC) multiarterial estável e função ventricular preservada, permanece controverso. Os tratamentos incluem TM, ICP e CRM, todos os quais são usados em conjunto com a prevenção secundária agressiva. Em comparação com o CRM, vários estudos randomizados mostraram que o tratamento com ICP ou TM está associado a taxas mais elevadas de angina e de revascularização subsequente, mas não há diferenças significativas na mortalidade ou nas taxas de infarto do miocárdio (IM).

Foram incluídos no estudo pacientes com estenose⁶ proximal multiarterial coronariana de mais de 70%, por avaliação visual documentada angiograficamente, e isquemia documentada. A isquemia⁷ foi documentada por testes de estresse ou avaliação típica de angina estável da Sociedade Cardiovascular Canadense (classe II ou III).

Os doentes foram recrutados e randomizados se os cirurgiões, médicos assistentes e cardiologistas intervencionistas concordassem que a revascularização poderia ser alcançada por qualquer estratégia. Os pacientes assinaram um termo de consentimento informando que foram distribuídos aleatoriamente para um grupo de

³ Intervenção Coronária Percutânea (ICP) ou Angioplastia Coronária é o tratamento não cirúrgico das obstruções das artérias coronárias por meio de cateter balão, com o objetivo de aumentar o fluxo de sangue para o coração. Após a desobstrução da artéria coronária, por meio da angioplastia com balão, procede-se ao implante de uma prótese endovascular (para ser utilizada no interior dos vasos) conhecida como 'stent' (pequeno tubo de metal), usado para manter a artéria aberta.

⁴ Cirurgia de Revascularização Miocárdica (CRM) é um procedimento cardíaco, que cria um novo caminho para o sangue fluir ao redor (desvio) da obstrução de uma artéria, para que o sangue possa chegar ao coração. Existem dois procedimentos cirúrgicos de revascularização para a doença da artéria coronariana: cirurgia com o coração batendo e cirurgia com circulação extracorpórea.

⁵ Onda Q: Região específica de um Eletrocardiograma (ECG) que indica a despolarização septal. Septo: Designação genérica de divisória de órgãos ou de cavidades. Septo interventricular - Separa o ventrículo esquerdo do direito.

⁶ Estenose: estreitamento de qualquer canal ou orifício.

⁷ Isquemia: diminuição ou suspensão da irrigação sanguínea, numa parte do organismo, ocasionada por obstrução arterial ou por vasoconstrição.

tratamento. A metodologia foi aprovada por um comitê de ética e todos os procedimentos foram realizados de acordo com a Declaração de Helsinki⁸.

Os critérios clínicos de exclusão incluíram: angina refratária ou infarto agudo do miocárdio que necessitou de revascularização de emergência; aneurisma ventricular⁹, que exigiu reparo cirúrgico; fração de ejeção ventricular esquerda menor do que 40%; um histórico de ICP ou de revascularização miocárdica; lesão uniarterial e DAC normal ou mínimo. Os pacientes também foram excluídos do estudo se: tivessem uma história de doença cardíaca congênita, doença cardíaca valvular ou cardiomiopatia¹⁰; fossem incapazes de compreender ou de cooperar com os requisitos do protocolo ou de voltar para acompanhamento; tiveram estenose da artéria coronária de 50% ou mais; tinham gravidez suspeita ou conhecida; tinham alguma condição coexistente que contraindicava a revascularização do miocárdio ou ICP.

Efeitos adversos, clínicos ou outros, foram rastreados desde a randomização. Os pacientes foram avaliados com visitas de acompanhamento a cada 6 meses durante 10 anos. Os pacientes foram submetidos a teste ergométrico anual durante todo o estudo a menos que contraindicado. Exames de rotina incluíram testes de sangue e eletrocardiograma de rotina a cada 6 meses.

O segundo conjunto de dados foi extraído do livro de Pintilie (2006). Se trata de um estudo de células foliculares de linfoma, cujo objetivo é relatar os resultados a longo prazo do grupo de pacientes. Nesse estudo prospectivo 541 pacientes foram registrados para tratamento de linfoma entre 1967 e 1996 no *Princess Margaret Hospital* de Toronto, Canadá.

Além de uma variável identificadora dos pacientes, temos no conjunto de dados cinco variáveis que foram avaliadas na hora do diagnóstico: O estágio clínico (I ou II); Se o paciente seria tratado com radioterapia ou com radioterapia e quimioterapia (gerando assim dois diferentes tratamentos); a idade do paciente e seu nível de hemoglobina (g/l).

⁸ Declaração de Helsinque: conjunto de princípios éticos que regem a pesquisa com seres humanos. Foi redigida pela Associação Médica Mundial em 1964. Posteriormente foi revisada 6 vezes, sendo sua última revisão em outubro de 2008.

⁹ Aneurisma Ventricular: uma dilatação de uma parte do ventrículo resultante da cicatrização da perda da musculatura consequente à um infarto do miocárdio ou miocardite infecciosa.

¹⁰ Cardiomiopatia: condição em que o músculo cardíaco torna-se inflamado e ampliado. Por estar ampliado, o músculo cardíaco é esticado tornando-se fraco e não conseguindo bombear sangue tão rápido como deveria, havendo o risco de insuficiência cardíaca.

Ao longo do estudo mais seis variáveis foram mensuradas: se o paciente obteve uma resposta completa, ou não, após o tratamento; se houve uma recaída e em qual local a recaída ocorreu (local, distante ou ambos); o tempo do diagnóstico até a morte ou último acompanhamento, em anos; o tempo do diagnóstico até a primeira falha (resposta incompleta ao tratamento, recaída ou morte) ou último acompanhamento; se houve ou não censura; e se o paciente se encontrava vivo ou morto.

3.1.2 Recursos Computacionais

O *software* R (R CORE TEAM, 2015) será utilizado para ajustar os modelos aos dados descritos. O pacote a ser investigado com este propósito é o “*mstate*” cujas principais características são (Putter *et. al.*, 2010):

- Aplicável a modelos não paramétricos (com covariáveis) e semi-paramétricos (sem covariáveis). Um exemplo de modelo semi-paramétrico é o modelo de regressão de Cox onde, os coeficientes de regressão representam a parte paramétrica e o risco de base a parte não paramétrica.
- Contém funções para reorganizar os dados em formato adequado para análise multiestados e estimar os efeitos das covariáveis de forma flexível, no contexto dos modelos de regressão de Cox. Pode também ser combinado com outras funções do R (Ex.: para seleção de modelos e confecção de gráficos).
- Possui funções para estimar: as intensidades de transição de pacientes específicos e probabilidades de predição dinâmicas com os respectivos erros padrões associados.
- Modelos com riscos competitivos também podem ser analisados, pois estes são um tipo especial de modelos multiestados.
- Suas funções podem ser utilizadas de forma independente. Ou seja, suas entradas não necessariamente dependem de outras funções no pacote, conferindo flexibilidade ao pacote.

Um dos principais objetivos do “*mstate*” é permitir a predição dinâmica, sendo que uma predição inicial pode ser atualizada pela incorporação de informações sobre a chegada em outros estados em pontos de tempo posteriores. Isso ajuda a responder à pergunta clínica: como dar um prognóstico para um dado paciente conhecendo o valor das covariáveis e como atualizar esta predição se eventos intermediários ocorrerem?

3.2 Métodos

A estratégia de regressão proposta será fundamentada nos modelos multiestados os quais serão utilizados para modelar os dados relativos ao tempo até o evento de interesse. Nesta estratégia de modelagem, todos os indivíduos iniciam em um estado de partida (ex.: pós-diagnóstico, transplante, cirurgia, etc.) e, terminam em um estado absorvente ou final (Ex.: morte ou recidiva). O estado inicial e final pode ser diferente entre os indivíduos. Entre os estados inicial e final ou absorvente, estados intermediários podem ser visitados, possivelmente mais de uma vez. Alguns indivíduos são censurados antes que atinjam um estado final ou absorvente. A Figura 1 mostra um exemplo de representação gráfica das transições entre estados de um Modelo Multiestados para pacientes com Transplante de Medula Óssea.

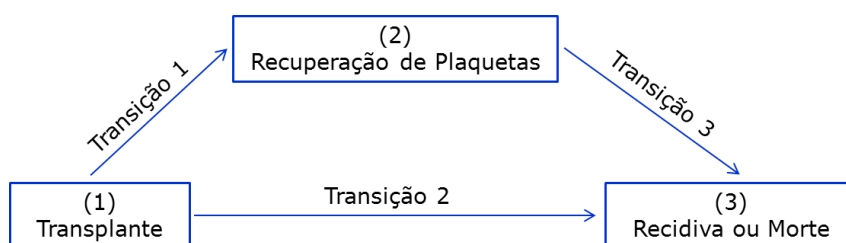


Figura 1 - Representação gráfica das transições entre estados de um Modelo Multiestados para pacientes com Transplante de Medula Óssea.

Os modelos de riscos competitivos são uma subcategoria dos modelos multiestados. Possuem um estado inicial e pelo menos 2 estados absorventes e não há estados intermediários. Suas transições são frequentemente apontadas como causas da falha. A Figura 2 mostra um exemplo de representação gráfica das transições entre estados de um modelo de riscos competitivos para pacientes com Transplante de Medula Óssea.

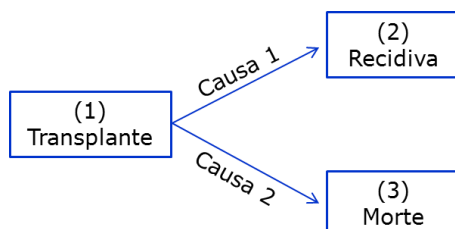


Figura 2 - Representação gráfica das transições entre estados de um Modelo de Riscos Competitivos para pacientes com Transplante de Medula Óssea.

Devido às características dos dados e as especificidades das análises de interesse, este trabalho aplicará somente a modelagem inerente aos modelos multiestados. A seguir são apresentadas as principais funções dos modelos multiestados.

3.2.1 Modelos Multiestados - Funções Básicas

Considere $S = \{1, 2, \dots, s\}$ um conjunto de estados e $X(t)$ um processo aleatório Markoviano¹¹ em S . No contexto de modelos multiestados, a intensidade de transição ou taxa de falha/risco, denotada por $\alpha_{gh}(t)$, é definida como a probabilidade instantânea de uma transição do estado “g” para o estado “h” no tempo t , isto é,

$$\alpha_{gh}(t) = \lim_{\Delta t \rightarrow 0} \frac{P[X(t + \Delta t) = h \mid X(t) = g]}{\Delta t}. \quad (1)$$

Em consequência, a intensidade de transição ou taxa de falha/risco acumulada, $A_{gh}(t)$, é dada pela Equação (2):

$$A_{gh}(t) = \int_0^t \alpha_{gh}(u) du. \quad (2)$$

Assume-se que o estado futuro depende apenas do estado atual (processo Markoviano), assim como que: (a) existem $K > 0$ transições possíveis e (b) se uma transição direta do estado “g” para o estado “h” é impossível, então $A_{gh}(t) = 0$.

As intensidades de transição ou taxas de falha/risco acumuladas, $A_{gh}(t)$, podem ser representadas por uma matriz $S \times S$, cuja diagonal principal é formada pelos elementos $A_{gg}(t)$, conforme Equação (3).

$$A_{gg}(t) = - \sum_{h \neq g} A_{gh}(t). \quad (3)$$

Para indivíduos com probabilidade de transição nula é assumido que estes permanecem no estado “g”. Os elementos da matriz de probabilidade de transição, denotada por $P(s, t)$, são definidos conforme a Equação (4), em que $P_{gh}(s, t)$ corresponde

¹¹ Processo Markoviano: processo no qual o estado futuro depende apenas do estado atual.

a probabilidade de transição de estado “g” para o estado “h” no intervalo de tempo (s, t]. Cada elemento $P_{gh}(s, t)$ combina transições diretas e indiretas do estado “g” para o estado “h”.

$$P_{gh}(s, t) = P[X(t) = j | X(s) = h]. \quad (4)$$

Em modelos Markovianos, dado $A(t)$, a matriz $P(s, t)$ das probabilidades de transição pode ser calculada por meio da Equação (5), em que I é a matriz identidade e $dA(u)$ a derivada da matriz de intensidade de transição ou taxa de falha/risco acumulada.

$$P(s, t) = \prod_{(s, t]} [I + dA(u)]. \quad (5)$$

No contexto de modelos com riscos Ccompetitivos, a função de incidência acumulada $P_{1h}(0, t)$, expressa a probabilidade de falhar de causa “h” antes do tempo “t”. Modelos de sobrevivência clássicos, que consideram apenas a transição de “vivo” para “morto”, podem ser considerados como modelos de 2 estados e, neste caso, a probabilidade de sobrevivência pode ser determinada por meio da Equação (6) em que: $P_{12}(0, t)$ é a probabilidade acumulada de transição do estado 1 para o estado 2 no tempo “t” e $P_{11}(0, t)$ é a probabilidade acumulada de permanecer no estado 1 no tempo “t”.

$$S(t) = 1 - P_{12}(0, t) = P_{11}(0, t). \quad (6)$$

REFERÊNCIAS

ANDERSEN, P.K.; BORGAN, Ø.; GILL, R.D.; KEIDING, N. **Statistical Models Based on Counting Processes**, 2nd ed., Springer Series in Statistics, Springer, 1993.

ALLIGNOL, A.; BEYERSMANN, J.; SCHUMACHER, M. An R package for the Nelson-Aalen estimator in multistate models, **R News**, 8 (2), 48–50, 2008.

COX, D.R. Regression models and life tables. **Journal Royal Statistical Society, Series B**, v. 34, N° 2, p. 187-220, 1972.

FIOCCO, M.; PUTTER, H.; Van HOUWELINGEN, H.C. Reduced rank proportional hazards regression and simulation-based prediction for multi-state models, **Statistics in Medicine**, N° 27, p. 4340–4358, 2008.

HUEB, W.; LOPES, N.; GERSH, B. J.; SOARES, P. R.; RIBEIRO, E. E.; PEREIRA, A. C.; FAVARATO, D.; ROCHA, A. S. C.; HUEB, A. C.; RAMIRES, J. A.F. Ten-Year Follow-Up Survival of the Medicine, Angioplasty, or Surgery Study (MASS II): A Randomized Controlled Clinical Trial of 3 Therapeutic Strategies for Multivessel Coronary Artery Disease. **Circulation**, 2010, 122, 949-957, ISSN: 0009-7322.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, Vol. 53, N° 282, p. 457-481, 1958.

PINTILIE, M. **Competing risks: a practical perspective**. Chichester: John Wiley & Sons, Ltd, 2006.

PUTTER, H.; de WREEDE, L. C.; FIOCCO, M. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. **Computer Methods and Programs in Biomedicine**, N° 99, p. 261–274, 2010.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria, 2015. ISBN 3-900051-07-0. Disponível em: <http://www.R-project.org>.