

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Camille Alves Caetano Rezende do Prado

Aplicação do Modelo de Regressão de Cox em dados de pacientes chagásicos

Curitiba

2024

Camille Alves Caetano Rezende do Prado

Aplicação do Modelo de Regressão de Cox em dados de pacientes chagásicos

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. José Luiz Padilha da Silva

Curitiba
2024

Aplicação do Modelo de Regressão de Cox em dados de pacientes chagásicos

Application of the Cox Regression Model to data from chagasic patients

Camille Alves Caetano Rezende do Prado¹, José Luiz Padilha da Silva²

¹Aluna do programa de Especialização em Data Science & Big Data, camilleacrp@gmail.com

²Professor do Departamento de Estatística - DEST/UFPR, jlpadilha@ufpr.br

Pacientes portadores de doença de Chagas estão sujeitos a desenvolver disfunções cardíacas progressivas ao longo do tempo, com risco de morte súbita. A análise de sobrevivência, com a utilização do modelo de regressão semiparamétrico de Cox, permite entender os fatores que aumentam esse risco. Neste trabalho, propôs-se estimar os efeitos que as covariáveis de um modelo ajustado têm sobre o risco de óbito, com base em um conjunto de dados de pacientes chagásicos. O modelo ajustado apresentou boa concordância e atendeu a suposição de razão de riscos proporcionais. As estimativas dos riscos para as covariáveis foram coerentes com o que se encontra na literatura.

Palavras-chave: Análise de Sobrevivência; Modelo de Cox; Doença de Chagas

Patients with Chagas disease are subject to developing progressive cardiac dysfunction over time, with a risk of sudden death. Survival analysis, using the semiparametric Cox regression model, allows us to understand the factors that increase this risk. In this work, we proposed to estimate the effects that the variables of an adjusted model have on the death risk based on a set of data from patients with Chagas disease. The adjusted model showed good agreement and met the proportional hazards ratio assumption. The risk estimates for the variables were consistent with what is found in the literature.

Keywords: Survival Analysis; Cox Model; Chagas Disease

1. Introdução

A doença de Chagas foi descrita por Carlos Chagas em 1909 e é causada pela infecção do *Trypanosoma cruzi*, um protozoário parasita [1][2]. Ela pode ser transmitida vetorialmente, por triatomíneos hematófagos infectados, ou de forma indireta, seja via oral (pela ingestão de alimentos contaminados), vertical (de mãe para bebê, durante gravidez ou parto), acidental (contato de pele ferida ou mucosas com material contaminado), por transfusão de sangue e transplante de órgãos. O principal vetor da doença está presente em toda a América Latina, mas atualmente, por conta da globalização, a doença também é registrada em outros continentes [2]. No Brasil a doença urbanizou-se próximo aos anos 80 com a urbanização e fluxos migratórios rurais para as grandes cidades, com grande predomínio de casos nas regiões Nordeste, Centro-Oeste e Sudeste [3][4]. Segundo dados do portal do Ministério da Saúde, o Brasil registrou, de 2010 a 2019, 45.409 óbitos causados pela doença. Além disso, é uma das

doenças tropicais negligenciadas que geram mais custos devido a sua elevada morbimortalidade e ao tratamento de longo prazo, no qual procedimentos cirúrgicos costumam ser requeridos [4][2]. A maior parte dos infectados na fase aguda da doença ou não apresentam sintomas ou apresentam sintomas não específicos, e assim, sendo mal diagnosticados e não recebendo tratamento adequado, transicionam para a fase crônica da doença e podem permanecer sem sintomas ao longo da vida. Ainda que assintomáticos, podem apresentar anormalidades cardiovasculares e gastrointestinais com o tempo. A cardiomiopatia é a manifestação mais severa do estágio crônico da doença e um reconhecimento precoce pode prevenir, dentre outras coisas, o risco de derrame e morte súbita [3][1]. Logo, identificar os indivíduos que estão sob maiores riscos é um objetivo terapêutico importante [2].

Na análise de sobrevivência, em que a variável dependente é o tempo até determinado evento, o modelo de regressão semiparamétrico de Cox permite a estimação dos efeitos das covariáveis sobre o risco de ocorrer

o evento, sem ser necessário que se assuma uma distribuição de probabilidade sobre o tempo [5][6]. Também é chamado de modelo de riscos proporcionais, pois a suposição básica para o uso deste modelo é de que a razão das funções de risco de dois indivíduos diferentes é constante no tempo. Portanto, assume-se no modelo que os efeitos das covariáveis atuam multiplicativamente na função de risco [7].

Sendo assim, a proposta deste trabalho é ajustar um modelo de regressão de Cox em dados de pacientes chagásicos e estimar os efeitos das covariáveis sobre o tempo até o óbito.

2. Análise exploratória

Os dados utilizados neste estudo são oriundos do Hospital das Clínicas da Universidade Federal de Minas Gerais, coletados de pacientes com sorologia positiva para doença de Chagas entre os anos de 1999 e 2019. As variáveis são referentes a sexo, idade e a informações clínicas provenientes de exames de ecocardiograma e de eletrocardiograma. Foram analisados utilizando a linguagem R (versão 4.4.0) por meio do software RStudio (versão 2024.04.0). De 469 observações, 177 são do sexo feminino e 267 do masculino, com idade média de 51,34 anos (com a idade mínima sendo 19 anos e a máxima de 84 anos), sendo 141 o número de eventos (óbitos).

As 28 variáveis do conjunto de dados, com exceção das variáveis referentes ao registro do tempo e das ocorrências de óbito, foram inicialmente avaliadas por meio da observação de suas curvas de sobrevivência obtidas pelo estimador de Kaplan-Meier. Para isso, as variáveis contínuas foram categorizadas em intervalos iguais. Foram encontradas diferenças estatisticamente significativas (p -valor<0,05) em 19 delas, das quais três foram desconsideradas por apresentarem dados faltantes em porcentagem superior ou igual a 30%. Não foi realizada imputação dos dados faltantes para a aplicação do modelo.

3. Ajuste do modelo

As 16 covariáveis (Figura 1), então, foram ajustadas conjuntamente e avaliadas individualmente pelos resultados do teste de razão de verossimilhanças. Após isto, foram ajustados modelos reduzidos por remoção de uma única variável por vez, excluindo-se do modelo as covariáveis com menor significância de efeito geral até restar apenas as que apresentaram p -valor significa-

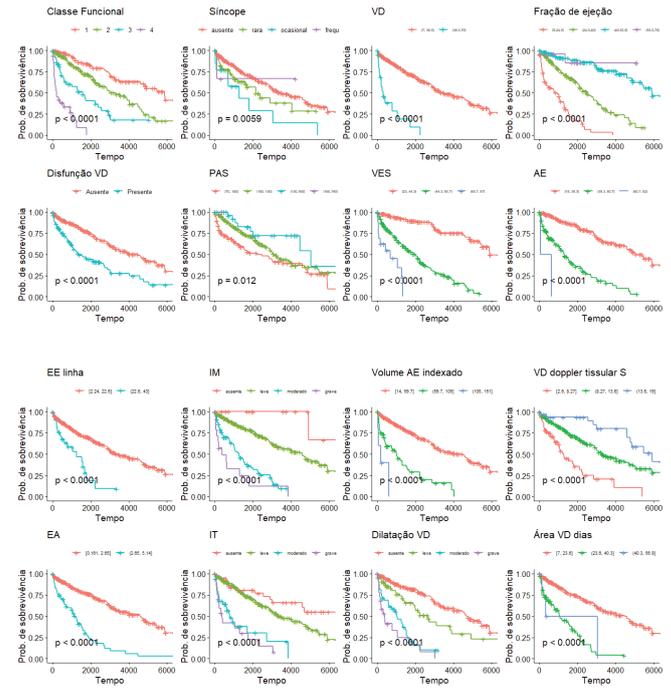


Figura 1: Curvas de sobrevivência obtidas pelo estimado de Kaplan-Meier das covariáveis elegidas para ajuste.

tivo. Por fim, verificou-se se covariáveis anteriormente removidas poderiam apresentar efeito significativo ao serem novamente ajustadas ao modelo.

O modelo final (Figura 2) apresentou concordância de 0.83, demonstrando bom poder discriminatório, com 54 observações deletadas devido a dados faltantes. Nos testes de razão de verossimilhanças, de Wald e log-rank pontuou, respectivamente, 229.9, 205.8 e 284.1.

4. Diagnóstico do modelo

A suposição de razão de riscos proporcionais foi avaliada por meio do teste de resíduos de Schoenfeld [5][7][8]. O teste não rejeita a hipótese nula de riscos proporcionais e a análise gráfica dos resíduos de Schoenfeld não demonstrou tendência ao longo do tempo (Figura 3). Assim, esse resultado aponta que o pressuposto básico do modelo é atendido.

5. Discussão

Fatores como o sexo e a idade podem estar associados com a progressão do dano cardíaco [2], porém, neste conjunto de dados, não apresentaram efeitos significativos nas curvas de sobrevivência de Kaplan-

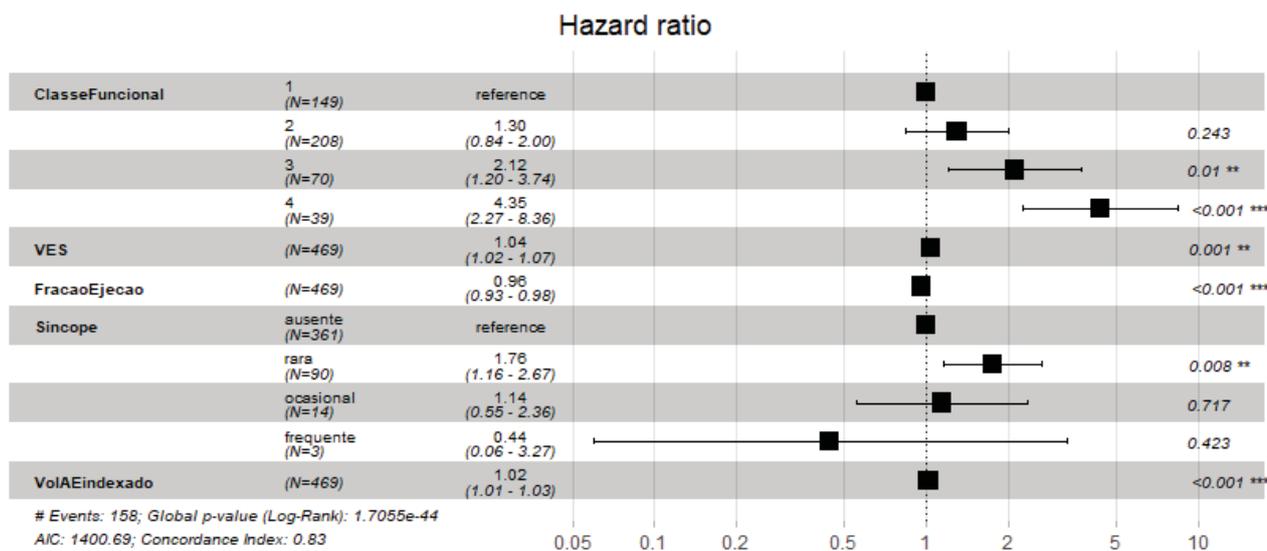


Figura 2: Resultados do modelo ajustado final, com a razão de riscos e intervalos de confiança para cada covariável.

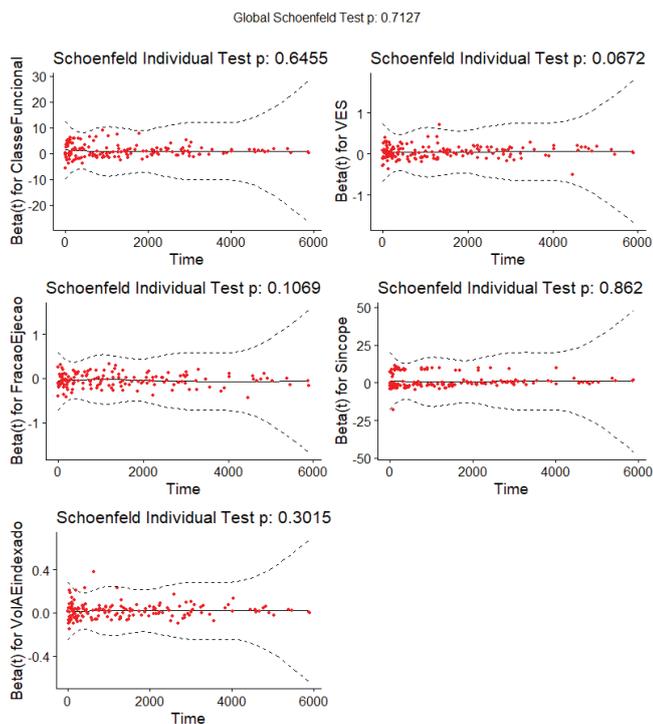


Figura 3: Gráfico do teste de resíduos de Schoenfeld.

Meier. Outros fatores como insuficiência cardíaca, disfunção ventricular, fração de ejeção e volume do átrio esquerdo, são preditores consistentes do risco de óbito [2][3][9], sendo o primeiro mensurado segundo a classificação NYHA (New York Heart Association). A variável para disfunção ventricular apresentou significância sobre as curvas de sobrevivência de Kaplan-Meier (Figura 1), mas não foi incluída no ajuste final. Das co-

variáveis ajustadas no modelo (Figura 2), interpreta-se por suas estimativas que o risco de óbito para pacientes da Classe Funcional 3 e 4 (classificação NYHA) é de, nesta ordem, 2.12 e 4.35 vezes maior que o de pacientes da Classe Funcional 1. Observa-se um efeito protetor da covariável referente a fração de ejeção, com redução do risco de 4% para cada acréscimo de unidade, e sobrerisco de 4% e de 2% para cada acréscimo de unidade nas covariáveis referentes ao volume de ejeção sistólico (“VES”) e ao volume indexado do átrio esquerdo (“VolAEindexado”), respectivamente. A estimativa para pacientes que têm síncope raramente é de um sobrerisco de 76%. As estimativas de risco atribuídas aos pacientes que apresentam síncope ocasionais e frequentes não são significativas, pois seus intervalos de confiança englobam a unidade [5]; o mesmo ocorrendo para pacientes da Classe Funcional 2.

6. Conclusão

Neste trabalho foi realizada uma aplicação do modelo de regressão semiparamétrico de Cox para estimar os efeitos que as variáveis, do conjunto de dados aqui estudado, apresentam sobre o risco de óbito, que é o evento de interesse. Conclui-se que os efeitos de risco e proteção das covariáveis ajustadas encontram suporte na literatura e nota-se que não analisar a interação entre as covariáveis é uma limitação deste trabalho, o que pode servir como investigação em estudos futuros.

Referências

- [1] Hochberg, N. S.; Montgomery, S. P. *Annals of Internal Medicine*. *Annals of internal medicine*, v. 176, n. 2, p. ITC17–ITC32, fev. 2023.
- [2] Nunes, M. C. P. et al. Chagas disease: an overview of clinical and epidemiological aspects. *Journal of the American College of Cardiology*, v. 62, n. 9, p. 767–776, 27 ago. 2013.
- [3] Almeida, D. R. Insuficiência cardíaca na doença de Chagas. 2004.
- [4] Alves, S. M. M. et al. *Advances in Clinical Practice, Diagnosis and Treatment of Chronic Chagas' Heart Disease*. *ABC Heart Fail Cardiomyop*, v. 3, n. 1, p. e20230029, 10 jul. 2023.
- [5] Carvalho, M. S. et al. Análise de sobrevivência: teoria e aplicações em saúde. Em: *Análise de sobrevivência: teoria e aplicações em saúde*. [s.l.: s.n.]. p. 434–434.
- [6] Miot, H. A. Análise de sobrevivência em estudos clínicos e experimentais. *Jornal Vascular Brasileiro*, v. 16, n. 4, p. 267–269, 2017.
- [7] Colosimo, E. A.; Giolo, S. R. *Análise de Sobrevivência Aplicada*. [s.l.] Editora Blucher, 2021.
- [8] Schoenfeld, D. Partial residuals for the proportional hazards regression model. *Biometrika*, v. 69, n. 1, p. 239–241, 1 abr. 1982.
- [9] Rocha, E. S. DA et al. Volume de átrio esquerdo como preditor de morte em miocardiopatia chagásica dilatada. *Rev. bras. ecocardiogr*, p. 29–36, 2004.