



Universidade de Brasília
IE - Departamento de Estatística
Estágio Supervisionado 1

Análise de Sobrevivência para Dados Grupados

Daniel Lima Viegas

Proposta de Projeto Final

Orientador: Prof.^a Juliana Betini Fachini Gomes

Brasília
Setembro de 2017

Sumário

1	Introdução	3
2	Objetivo	3
2.1	Objetivo Geral	4
2.2	Objetivos Especificos	4
3	Metodologia	5
3.1	Material	5
3.2	Métodos	5
4	Cronograma	5
	Referências	8

1 Introdução

A análise de sobrevivência é um tópico importante utilizado em diversas áreas, como biologia, engenharia, medicina, entre outros. O principal objetivo desta análise é explicar ou prever o tempo até a ocorrência do evento estudado, esse tempo é chamado de tempo de falha. A principal diferença desta técnica de modelagem para as demais é a capacidade de levar em consideração também os tempos onde não foi possível observar o evento de interesse, esse tipo de ocorrência é chamado de censura.

Dentro os tipos de censuras existentes, a mais genérica é a censura intervalar. Esse tipo de censura ocorre quando não é possível determinar o tempo de ocorrência, mas se tem o intervalo de tempo onde ele ocorreu. Por exemplo, no estudo sobre o tempo até uma lâmpada queimar, deixa-se a lâmpada ligada até que ela queime, em um dia, ela está funcionando, o pesquisador sai da área onde está acontecendo o experimento e quando retorna, a lâmpada está queimada. Neste caso, sabe-se que o intervalo de tempo onde a lâmpada queimou é entre o tempo em que o pesquisador saiu e o que ele voltou. O objetivo deste trabalho é estudar o comportamento de dados grupados, que é um caso particular da censura intervalar.

Em diversos estudos de sobrevivência, estuda-se o relacionamento de covariáveis e o tempo, tendo o objetivo de realizar as análises estatísticas e tentando encontrar o melhor uso dessas variáveis para a criação de um modelo de regressão para dados censurados.

Um dos objetivos deste trabalho é sugerir um modelo de regressão para dados grupados, ou seja, considerando-se a grande quantidade de empates. Para a análise, irão ser utilizadas funções relativas ao estudo de sobrevivência, sejam algumas delas para calcular a probabilidade de que o indivíduo permaneça no experimento além do tempo em que ele já está, ou a taxa de falha de um indivíduo, onde essas funções serão calculadas a partir de métodos não-paramétricos.

2 Objetivo

2.1 Objetivo Geral

O objetivo deste trabalho é propor um modelo de regressão para dados de sobrevivência com censura intervalar quando ocorrem muitos empates.

2.2 Objetivos Específicos

Como a Análise de Sobrevivência exige o estudo de algumas técnicas não-paramétricas para o cálculo de seus estimadores e devido a não normalidade dos dados, apresentam-se os seguintes objetivos específicos:

- Estudar a metodologia de análise de sobrevivência;
- Revisar a bibliografia a respeito de estudos de sobrevivência com dados grupados;
- Estudar as metodologias computacionais presentes nos pacotes de análise de sobrevivência no software estatístico R, com auxílio da IDE RStudio;
- Estudar o banco de dados para a aplicação de um possível modelo;
- Aplicar métodos para verificar a relação entre a variável tempo e as covariáveis presentes no banco;
- Propor um modelo de regressão para dados censurados para

3 Metodologia

3.1 Material

A fim de propor um modelo de regressão para dados grupados, irá ser utilizado um banco de dados cedido pelo Instituto de Saúde Coletiva da Universidade Federal da Bahia. Esses dados foram obtidos a partir de um estudo conduzido por Barreto et al.(1994). O banco de dados é formado por 1207 crianças com idade entre 6 e 48 meses no início do estudo, que receberam placebo ou vitamina A. Dentre as variáveis se encontra a variável tempo, que é o tempo entre a primeira dose de placebo ou vitamina A e a ocorrência de diarreia na criança, a variável idade, tipo de tratamento e a variável sexo.

3.2 Métodos

A função de sobrevivência e a Taxa de risco acumulado são funções essenciais na análise de sobrevivência, por isso, primeiramente será necessário estimá-las.

Para a estimação dessas funções irá ser usado o estimador de Kaplan-Meier, foi escolhido por ser um estimador de máxima verossimilhança tendo assim sua consistência assintótica. Além disso, esse é um dos poucos métodos que consideram a censura.

O método da máxima verossimilhança também será usado para a estimação dos parâmetros do modelo a ser proposto, assim como os parâmetros de uma possível distribuição observada da variável tempo.

4 Cronograma

O cronograma a seguir, foi organizado da seguinte forma:

1. Escolha do tema a ser abordado;
2. Estudo da metodologia de análise de sobrevivência;

3. Estudo de modelos de sobrevivência para dados grupados;
4. Desenvolvimento da proposta de projeto final;
5. Entrega da proposta final do projeto final;
6. Entrega do relatório parcial ao orientdo para correção;
7. Ajuste de modelo de dados de sobrevivência para dados grupados;
8. Descrever resultados para o relatório final;
9. Correção do relatório final;
10. Entrega do relatório final a banca examinadora; e
11. Apresentação do relatório final para a banca examinadora.

Tabela 1: Cronograma 2/2017

	2/2017					
Atividades	Julho	Agosto	Setembro	Outubro	Novembro	Dezezebro
1						
2						
3						
4						
5						
6						

Tabela 2: Cronograma 1/2018

	1/2018					
Atividades	Janeiro	Fevereiro	Março	Abril	Maiio	Junho
7						
8						
9						
10						
11						

Referências

- Barreto, M. L., Santos, L. M. P., Assis, A. M. O., Araújo, M. P. N., Franzena, G. G., Santos, P. A. B., & Fiaccone, R. L. (1994). Effect of vitamin a supplementation on diarrhoea acute lower-respiratory-tract infection in young children in brazil. *Lancet* 344, pages 228–231.
- Colosimo, E, A. & Giolo, S. R. (2006). *Análise de Sobrevivência Aplicada*. Editora Bucher.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*, (2nd edition ed.). John Wiley and sons.