



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.<sup>a</sup> Juliana Betini Fachini Gomes

Brasília  
Setembro de 2017

# Sumário

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Objetivo</b>	<b>3</b>
2.1	Objetivo Geral . . . . .	3
2.2	Objetivos Especificos . . . . .	4
<b>3</b>	<b>Metodologia</b>	<b>5</b>
3.1	Material . . . . .	5
3.2	Métodos . . . . .	5
<b>4</b>	<b>Cronograma</b>	<b>6</b>
	<b>Referências</b>	<b>8</b>

# 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 em que 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 para analisar o banco de dados de Barreto et al. (1994).

## 2 Objetivo

### 2.1 Objetivo Geral

O objetivo deste trabalho é propor um modelo de regressão para dados grupados para analisar os dados de Barreto et al. (1994).

## 2.2 Objetivos Especificos

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 grupados.

## 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

Para realizar uma análise preliminar no tempo até a ocorrência de diarreia na criança, é necessário estimar a função de sobrevivência para a análise descritiva da variável. Para essa estimação, será usado o estimador de Kaplan-Meier. Este estimador foi escolhido por ser um estimador de máxima verossimilhança, possuindo assim as propriedades de um estimador deste tipo. A partir da estimação da função de sobrevivência, pela propriedade da invariância, pode-se estimar a taxa de risco acumulado. Por meio desta função, encontra-se possíveis distribuições para a variável resposta.

Neste trabalho, tem-se como um objetivo verificar o efeito de covariáveis, como sexo e idade, e a resposta, tempo. Para tal, será proposto um modelo de regressão, que é uma extensão da distribuição de probabilidade assumida para a variável resposta.

Para a estimação dos parâmetros do modelo, não pode-se usar alguns métodos de estimação, tais como o método de mínimos quadrados e de momentos, pois eles não levam em consideração a censura presente nos dados de sobrevivência. Sendo assim, para incorporar a censura na análise dos dados, será utilizado uma adaptação do método de máxima verossimilhança.

Para a análise dos dados, será utilizado o software estatístico R por meio da IDE Rstudio.

## 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 agrupados;
4. Desenvolvimento da proposta de projeto final;
5. Entrega da proposta final do projeto final;
6. Entrega do relatório parcial ao orientador para correção;
7. Ajuste de modelo de dados de sobrevivência para dados agrupados;
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

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

Tabela 2: Cronograma 1/2018

	<b>1/2018</b>						
<b>Atividades</b>	<b>Janeiro</b>	<b>Fevereiro</b>	<b>Março</b>	<b>Abril</b>	<b>Maio</b>	<b>Junho</b>	<b>Julho</b>
<b>2</b>							
<b>7</b>							
<b>8</b>							
<b>9</b>							
<b>10</b>							
<b>11</b>							

## 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.
- Hashimoto, E. M. (2008). Modelo de regressão para dados com censura intervalar e dados de sobrevivência grupados. Master's thesis, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*, (2nd edition ed.). John Wiley and sons.
- Resende, V. S. (2017). Modelo de regressão log-beta burr iii para dados grupados. Master's thesis, Universidade de Brasília. Departamento de Estatística - Instituto de Ciências Exatas.