

# Análise de Dados Amostrais Complexos

*Djalma Pessoa e Pedro Nascimento Silva*

*2016-12-27*



# Contents

<b>Prefácio</b>	<b>5</b>
Agradecimentos . . . . .	5
<b>1 Introdução</b>	<b>7</b>
<b>2 Referencial para Inferência</b>	<b>9</b>
<b>3 Estimação Baseada no Plano Amostral</b>	<b>11</b>
<b>4 Efeitos do Plano Amostral</b>	<b>13</b>
<b>5 Ajuste de Modelos Paramétricos</b>	<b>15</b>
<b>6 Modelos de Regressão</b>	<b>17</b>
<b>7 Testes de Qualidade de Ajuste</b>	<b>19</b>
<b>8 Testes em Tabelas de Duas Entradas</b>	<b>21</b>
<b>9 Agregação vs. Desagregação</b>	<b>23</b>
<b>10 Pacotes para Analisar Dados Amostrais</b>	<b>25</b>
<b>11 Placeholder</b>	<b>27</b>

```
library(survey)
```

```
## Loading required package: grid
## Loading required package: Matrix
## Loading required package: survival
##
## Attaching package: 'survey'
## The following object is masked from 'package:graphics':
##
##      dotchart
```

```
source("~/\\GitHub\\adac\\data\\ppv1.R")
```



# Prefácio

Uma preocupação básica de toda instituição produtora de informações estatísticas é com a utilização “correta” de seus dados. Isso pode ser interpretado de várias formas, algumas delas com reflexos até na confiança do público e na própria sobrevivência do órgão. Do nosso ponto de vista, como técnicos da área de metodologia do IBGE, enfatizamos um aspecto técnico particular, mas nem por isso menos importante para os usuários dos dados.

A revolução da informática com a resultante facilidade de acesso ao computador, criou condições extremamente favoráveis à utilização de dados estatísticos, produzidos por órgãos como o IBGE. Algumas vezes esses dados são utilizados para fins puramente descritivos. Outras vezes, porém, sua utilização é feita para fins analíticos, envolvendo a construção de modelos, quando o objetivo é extrair conclusões aplicáveis também a populações distintas daquela da qual se extraiu a amostra. Neste caso, é comum empregar, sem grandes preocupações, pacotes computacionais padrões disponíveis para a seleção e ajuste de modelos. É neste ponto que entra a nossa preocupação com o uso adequado dos dados produzidos pelo IBGE.

O que torna tais dados especiais para quem pretende usá-los para fins analíticos? Esta é a questão básica que será amplamente discutida ao longo deste texto. A mensagem principal que pretendemos transmitir é que certos cuidados precisam ser tomados para utilização correta dos dados de pesquisas amostrais como as que o IBGE realiza.

O que torna especiais dados como os produzidos pelo IBGE é que estes são obtidos através de pesquisas amostrais complexas de populações finitas que envolvem: **probabilidades distintas de seleção, estratificação e conglomeração das unidades, ajustes para compensar não-resposta e outros ajustes**. Os pacotes tradicionais de análise ignoram estes aspectos, podendo produzir estimativas incorretas tanto dos parâmetros como para as variâncias destas estimativas. Quando utilizamos a amostra para estudos analíticos, as opções disponíveis nos pacotes estatísticos usuais para levar em conta os pesos distintos das observações são apropriadas somente para observações independentes e identicamente distribuídas (IID). Além disso, a variabilidade dos pesos produz impactos tanto na estimação pontual quanto na estimação das variâncias dessas estimativas, que sofre ainda influência da estratificação e conglomeração.

O objetivo deste livro é analisar o impacto das simplificações feitas ao utilizar procedimentos e pacotes usuais de análise de dados, e apresentar os ajustes necessários desses procedimentos de modo a incorporar na análise, de forma apropriada, os aspectos aqui ressaltados. Para isto serão apresentados exemplos de análises de dados obtidos em pesquisas amostrais complexas, usando pacotes clássicos e também pacotes estatísticos especializados. A comparação dos resultados das análises feitas das duas formas permitirá avaliar o impacto de ignorar o plano amostral na análise dos dados resultantes de pesquisas amostrais complexas.

## Agradecimentos

A elaboração de um texto como esse não se faz sem a colaboração de muitas pessoas. Em primeiro lugar, agradecemos à Comissão Organizadora do SINAPE por ter propiciado a oportunidade ao selecionar nossa proposta de minicurso. Agradecemos também ao IBGE por ter proporcionado as condições e os meios usados

para a produção da monografia, bem como o acesso aos dados detalhados e identificados que utilizamos em vários exemplos.

No plano pessoal, agradecemos a Zélia Bianchini pela revisão do manuscrito e sugestões que o aprimoraram. Agradecemos a Marcos Paulo de Freitas e Renata Duarte pela ajuda com a computação de vários exemplos. Agradecemos a Waldecir Bianchini, Luiz Pessoa e Marinho Persiano pela colaboração na utilização do processador de textos. Aos demais colegas do Departamento de Metodologia do IBGE, agradecemos o companheirismo e solidariedade nesses meses de trabalho na preparação do manuscrito.

Finalmente, agradecemos a nossas famílias pela aceitação resignada de nossas ausências e pelo incentivo à conclusão da empreitada.

## Chapter 1

# Introdução





## Chapter 2

# Referencial para Inferência



## Chapter 3

# Estimação Baseada no Plano Amostrai



## Chapter 4

# Efeitos do Plano Amostral



## Chapter 5

# Ajuste de Modelos Paramétricos





## Chapter 6

# Modelos de Regressão



## Chapter 7

# Testes de Qualidade de Ajuste



## Chapter 8

# Testes em Tabelas de Duas Entradas



## Chapter 9

# Agregação vs. Desagregação





## Chapter 10

# Pacotes para Analisar Dados Amostrais



## Chapter 11

# Placeholder



# Bibliography