

Sumário

Capa	5
Informações editoriais	7
Bem-vindo	9
Agradecimentos	11
Lista de Siglas	13
Lista de Tabelas	15
Lista de Figuras	17
Arquivos de dados usados no livro	19
Errata	21
1 Introdução	23
1.1 Motivação	23
1.2 Objetivos do livro	23
2 Conceitos e Cadastros	25
2.1 Fontes de dados e tipos de pesquisas	26
2.2 Alguns conceitos fundamentais	26
2.3 Abordagens alternativas para pesquisas por amostragem	26
2.4 Planejamento e execução de pesquisas por amostragem	26
2.5 Cadastros	26
2.6 Exercícios	26
3 Visão Geral da Amostragem e Estimação	27
3.1 Definições e notação para população de pesquisa e parâmetros selecionados	28
3.2 Amostra	28
3.3 Amostragem probabilística	28

3.4	Estatísticas, estimadores e estimativas	28
3.5	A distribuição de aleatorização	28
3.6	Estimadores não viciados para o total populacional	28
3.7	Teoria básica	28
3.8	Exercícios	28
4	Amostragem Aleatória Simples	29
4.1	Planos amostrais e algoritmos de seleção	30
4.2	Amostragem aleatória simples com reposição	30
4.3	Amostragem aleatória simples sem reposição	30
4.4	Determinação do tamanho da amostra	30
4.5	Exercícios	30
5	Estimação de Proporções	31
5.1	Parâmetros populacionais	32
5.2	Estimação sob amostragem aleatória simples com reposição . . .	32
5.3	Estimação sob amostragem aleatória simples sem reposição . . .	32
5.4	Distribuição amostral exata de estimadores de proporção sob AASC e AAS	32
5.5	Intervalos de confiança para proporções na amostragem aleatória simples	32
5.6	Intervalos de confiança utilizando a aproximação Normal	32
5.7	Cálculo do tamanho da amostra	32
5.8	Estimação de proporções para variáveis com mais de duas categorias	32
5.9	Exercícios	32
6	Estimação de Razões e Funções de Totais	33
6.1	Razões populacionais como parâmetros de interesse	33
6.2	Razão de médias <i>versus</i> média de razões	33
6.3	Outras funções de totais	33
6.4	Estimadores de razões de totais	33
6.5	Estimadores para outras funções de totais	33
6.6	Análise do vício do estimador da razão	33
6.7	Erro quadrático médio de \hat{R}	33
6.8	Exercícios	33
7	Estimação para Domínios de Estudo	35
7.1	Domínios de estudo	36
7.2	Parâmetros de interesse para domínios	36
7.3	Estimação de parâmetros para domínios sob AAS	36
7.4	Intervalos de confiança para os parâmetros para o domínio . . .	36
7.5	Propriedades condicionais (fixando n_c)	36
7.6	Estimação do total no domínio quando tamanho do domínio é conhecido	36
7.7	Comparação da eficiência dos estimadores de total no domínio . .	36
7.8	Estimação de proporções dentro do domínio	36

7.9	Exercícios	36
8	Amostragem Sistemática Simples	37
8.1	Ideia básica	38
8.2	Método de seleção da amostra	38
8.3	Composição das amostras sistemáticas simples	38
8.4	Probabilidades de inclusão na amostragem sistemática simples .	38
8.5	Variáveis aleatórias indicadoras - propriedades	38
8.6	Estimação de totais sob amostragem sistemática simples	38
8.7	Estimação de médias sob amostragem sistemática simples	38
8.8	Estimação de uma proporção na amostragem sistemática simples	38
8.9	Variância dos estimadores sob amostragem sistemática simples .	38
8.10	Estimação de variâncias dos estimadores na amostragem sistemática simples	38
8.11	Alternativas para seleção com amostragem sistemática simples .	38
8.12	Exercícios	38
9	Outros Métodos de Amostragem com Equiprobabilidade	39
9.1	Amostragem binomial ou de Bernoulli	40
9.2	Amostragem inversa simples	40
9.3	Exercícios	40
10	Amostragem com Probabilidades Proporcionais ao Tamanho	41
10.1	Motivação	42
10.2	Estimação do total populacional	42
10.3	Amostragem PPT com reposição	42
10.4	Amostragem PPT de Poisson	42
10.5	Amostragem sequencial de Poisson	42
10.6	Amostragem sistemática com PPT	42
10.7	Amostragem PPT de Pareto	42
10.8	Exercícios	42
11	Amostragem Estratificada	43
11.1	O método geral	44
11.2	Amostragem estratificada simples	44
11.3	Alocação da amostra nos estratos	44
11.4	Definição dos limites dos estratos	44
11.5	Número de estratos na amostragem estratificada simples	44
11.6	Exercícios	44
12	Amostragem Conglomerada	45
12.1	Conceituação básica	47
12.2	Amostragem de áreas	47
12.3	Definições básicas e notação	47
12.4	Amostragem conglomerada em um estágio	47
12.5	Amostragem conglomerada em vários estágios	47

12.6	Notação na amostragem conglomerada em um estágio	47
12.7	Amostragem conglomerada em um estágio simples	47
12.8	Efeito do plano amostral	47
12.9	Amostragem conglomerada com PPT em um estágio	47
12.10	Amostragem conglomerada em dois estágios	47
12.11	Ideias básicas na escolha de unidades compostas	47
12.12	Exercícios	47
13	Estimadores de Calibração	49
13.1	Motivação	51
13.2	O problema	51
13.3	Estimadores de regressão como estimadores de calibração	51
13.4	Propriedades dos estimadores de calibração	51
13.5	Estimadores de razão como casos particulares da calibração . . .	51
13.6	Alguns problemas práticos com calibração	51
13.7	Seleção de variáveis para calibração com não resposta	51
13.8	Indicadores de qualidade da calibração	51
13.9	Um guia para aplicação de calibração	51
13.10	Observações finais	51
13.11	Exercícios	51

Capa

Informações editoriais

Publicação 15 de outubro de 2020

Última atualização 11 de novembro de 2022

Permitida a reprodução total ou parcial com a devida citação de fonte.

Comunicação com os autores

Pedro Luis do Nascimento Silva - pedronsilva@gmail.com Antonio José Ribeiro Dias - vermelho2@gmail.com

Revisão

Sonia Albieri

Capa

Ubiratã Oliveira dos Santos

Ficha catalográfica

S586a Silva, Pedro Luis do Nascimento
Amostragem : teoria e prática usando R / Pedro Luis do Nascimento Silva,
Zélia Magalhães Bianchini, Antonio José Ribeiro Dias – Rio de Janeiro : [s.n.],
2021.
1 v.; il.
Inclui referências
Disponível em: <<https://amostragemcomr.github.io/livro/>>.
ISBN: 978-65-00-54993-5
1. Amostragem (Estatística). 2. Amostragem – Métodos estatísticos. I.
Bianchini, Zélia Magalhães. II. Dias, Antonio José Ribeiro. III. Título.
CDU: 311.213.2

Bem-vindo

Este é um livro escrito para apoiar a aprendizagem de **Amostragem**. Nosso objetivo principal é orientar um leitor no caminho para aprender os conceitos, as principais ideias e a usar as ferramentas de amostragem para resolver problemas.

Nossa escolha de temas a incluir no livro foi guiada, em grande parte, por nossa experiência com a coleção de pesquisas por amostragem do IBGE, onde trabalhamos por vários anos, e nossa atuação como professores na graduação e na pós graduação da ENCE. Também reflete nossa perspectiva quanto ao melhor caminho para aprender a trabalhar com **Amostragem**.

Nossa abordagem não busca ser exaustiva e, por esse motivo, são poucas as provas que incluímos dos resultados aqui discutidos. Fizemos a escolha deliberada de não apresentar demonstrações da maioria dos resultados relacionados a valor esperado e variância de estimadores. Essas demonstrações podem intimidar e afastar alguns e, além disso, cremos que estão bem cobertas em diversos outros livros sobre o tema. Aos leitores interessados em verificar os resultados, recomendamos a consulta ao excelente livro de Särndal et al. (1992) ou então às muitas referências incluídas ao longo do texto.

Em contraste, escolhemos enfatizar a apresentação de exemplos e de ferramentas computacionais, algo que não tem cobertura tão ampla na literatura sobre **Amostragem**. Nesse contexto, optamos também por enfatizar o uso de comandos ou recursos básicos do R, em lugar de pacotes mais avançados que estão disponíveis.

O livro está organizado em treze capítulos, nominados a seguir:

- 1) Introdução
- 2) Conceitos e Cadastros
- 3) Visão Geral da Amostragem e Estimação
- 4) Amostragem Aleatória Simples
- 5) Estimação de Proporções
- 6) Estimação de Razões e Funções de Totais
- 7) Estimação para Domínios de Estudo

- 8) Amostragem Sistemática Simples
- 9) Outros Métodos de Amostragem com Equiprobabilidade
- 10) Amostragem com Probabilidades Proporcionais ao Tamanho

- 11) Amostragem Estratificada
- 12) Amostragem Conglomerada
- 13) Estimadores de Calibração

Cada um dos capítulos é autocontido e vários deles podem ser omitidos num primeiro curso. Com exceção do Capítulo 13, o material do livro pode ser coberto num curso com cerca de 45 horas de duração, como ministrado várias vezes na pós-graduação da ENCE. Caso necessário, algum(ns) dos Capítulos 5, 6 ou 7 pode(m) ser suprimidos ou separados para estudo individual. Os Capítulos 8 e 9 podem ser omitidos sem prejuízo da sequência. O conteúdo central do livro é formado por todos os capítulos não citados neste parágrafo. Tal conteúdo formaria, a nosso ver, o mínimo para cobertura num primeiro curso, no nível de graduação, sobre **Amostragem**.

Nossa opção ao escolher essa forma de publicação (livro em formato de hipertexto, hospedado na internet) se deve a três fatores principais: primeiro, não pretendemos comercializar o livro e, sim, torná-lo de acesso livre e aberto, como é a filosofia do software que usamos para sua elaboração e produção (R + RStudio + R Markdown + Github); segundo, essa forma de publicação permitirá atualizações mais rápidas e frequentes do conteúdo, o que favorece a correção de erros, revisões do texto, inclusão de exemplos ou tópicos novos etc.; terceiro, podemos usar esse caminho disponibilizar os dados que usamos para exemplos e exercícios. Esperamos que essa escolha não afaste os leitores que ainda gostam de livros em papel, como nós. . .

O leitor de qualquer livro precisa reconhecer que não é possível começar do zero: é preciso contar com conhecimento prévio de algumas ideias e conceitos básicos essenciais à compreensão do material tratado. Nossa abordagem pressupõe que o leitor está familiarizado com um curso básico de introdução à probabilidade e à inferência estatística, no nível tratado, por exemplo, em Magalhães e Lima (2004) e Magalhães (2006).

Agradecimentos

Nossa gratidão é dirigida, em primeiro lugar, ao IBGE e à ENCE, instituições que nos propiciaram a maior parte das oportunidades de aprendizagem sobre o tema ao longo de nossas carreiras.

Agradecemos em especial a Sonia Albieri, que fez a revisão do livro, e que emprestou seu olhar competente e experimentado tanto nas questões técnicas como nos detalhes do idioma para reduzir os erros, imprecisões e falta de padronização que caracterizaram as primeiras versões do texto. É claro, todos os defeitos que porventura persistem são por nossa conta...

Na fase final de elaboração e revisão do livro, perdemos um querido mestre e amigo, o Prof. Djalma Galvão Carneiro Pessoa. Djalma foi quem nos induziu ao uso do R e nos ensinou o caminho das pedras com a elaboração de material em R Markdown. Muito antes disso, foi nosso professor, mentor e consultor, com uma generosidade sem par no compartilhamento do seu vasto saber e no seu apetite por enfrentar desafios de ajudar a resolver problemas que envolvessem Estatística e R. Deixamos aqui uma singela homenagem e um agradecimento por tudo que fez por nós ao longo de nossas vidas.

Por último, mas não menos importantes, agradecimentos aos cônjuges, de quem sempre roubamos tempo para dedicar a este projeto: Denise Britz, Waldecir Bianchini e Laura Motta, listados aqui na mesma ordem dos autores. Ao Waldecir Bianchini, uma gratidão especial pela contribuição no apoio computacional para o aprendizado da linguagem Latex e do R Markdown.

Como esperamos que este seja um livro vivo, agradecemos desde já aos leitores que se dispuserem a nos escrever com comentários e sugestões que nos permitam aprimorar o seu conteúdo sempre que necessário.

Boa leitura a todos!

Lista de Siglas

Placeholder

Lista de Tabelas

Placeholder

Lista de Figuras

Placeholder

Arquivos de dados usados no livro

Dê um *click* no nome do arquivo que deseja baixar

Nome do arquivo

Aves.csv

Empresas.csv

Rendimentos.csv

Fazendas_dat.rds

MunicBR_dat.rds

Obs.: No caso dos arquivos *CSV*, após visualisá-los, *click* no botão direito do *mouse* e utilize o comando *Salvar como*

Errata

Na tabela abaixo são apresentadas algumas correções feitas no texto do livro, tanto em erros localizados pelos autores como por leitores que tiveram a gentileza de nos comunicar. Os erros podem aparecer em versões que, eventualmente, possam ter sido baixadas por leitores antes das datas que aparecem na coluna “Data”.

Caso deseje comunicar algum problema encontrado ou sugestão, favor enviar e-mail para Pedro Luis do Nascimento Silva: pedronsilva@gmail.com ou Antonio José Ribeiro Dias: vermelho2@gmail.com.

Item	Data	Versão corrigida
Exemplo 4.4	10/11/2020	<code>med_amo <- sqrt(var_med_amo)</code>
Exemplo 4.4	10/11/2020	<code>med_amo <- 100 * dp_med_amo / med_amo</code>
Equação 10.18	18/01/2021	$\widehat{Y}_{AP} = \frac{n}{n-1} \sum_{i \in s} \left(\frac{y_i}{\lambda_i} - \frac{\sum_{k \in s} y_k (1 - \lambda_k) / \lambda_k}{\sum_{k \in s} (1 - \lambda_k)} \right)^2 (1 - \lambda_i)$
GSBPM (2013)	29/01/2021	Generic Statistical Business Process Model: GSBPM: version 5.0
Tabela 3.5	05/05/2020	100 6,52084692454763e+241 (última linha)
Tabela 12.3	22/06/2021	$Y_i / N_i = \frac{1}{N_i} \sum_{j \in C_i} y_{ij}$

Capítulo 1

Introdução

Placeholder

1.1 Motivação

1.2 Objetivos do livro

Capítulo 2

Conceitos e Cadastros

Placeholder

2.1 Fontes de dados e tipos de pesquisas

2.2 Alguns conceitos fundamentais

2.3 Abordagens alternativas para pesquisas por amostragem

2.4 Planejamento e execução de pesquisas por amostragem

2.4.1 Especificação das necessidades da pesquisa

2.4.2 Planejamento da pesquisa

2.4.3 Elaboração dos instrumentos de coleta e dos sistemas

2.4.4 Coleta dos dados

2.4.5 Processamento da pesquisa

2.4.6 Análise dos dados da pesquisa

2.4.7 Disseminação dos resultados da pesquisa

2.4.8 Arquivamento das informações da pesquisa

2.4.9 Avaliação da pesquisa

2.5 Cadastros

2.5.1 Cadastros importantes no IBGE

2.5.2 Defeitos de cadastros

2.5.3 Regras de associação das unidades da população à unidade cadastral

2.6 Exercícios

Capítulo 3

Visão Geral da Amostragem e Estimação

Placeholder

- 3.1** Definições e notação para população de pesquisa e parâmetros selecionados
- 3.2** Amostra
- 3.3** Amostragem probabilística
- 3.4** Estatísticas, estimadores e estimativas
- 3.5** A distribuição de aleatorização
- 3.6** Estimadores não viciados para o total populacional
- 3.7** Teoria básica
 - 3.7.1** Estimador linear de total
 - 3.7.2** Propriedades do estimador de Horvitz-Thompson
 - 3.7.3** Estimação da média populacional
- 3.8** Exercícios

Capítulo 4

Amostragem Aleatória Simples

Placeholder

- 4.1 Planos amostrais e algoritmos de seleção
- 4.2 Amostragem aleatória simples com reposição
 - 4.2.1 Estimação do total e média populacionais sob AASC
- 4.3 Amostragem aleatória simples sem reposição
 - 4.3.1 Algoritmo “convencional” para selecionar uma AAS
 - 4.3.2 Algoritmo de Hájek para selecionar AAS
 - 4.3.3 Algoritmo de Fan, Muller e Rezucha para selecionar AAS
 - 4.3.4 Probabilidades de inclusão sob AAS
 - 4.3.5 Estimador não viciado do total e média populacionais sob AAS
 - 4.3.6 Distribuição da média amostral
 - 4.3.7 Comparação dos planos de amostragem aleatória simples com e sem reposição
- 4.4 Determinação do tamanho da amostra
 - 4.4.1 Tamanho amostral para custo fixado
 - 4.4.2 Tamanho amostral para precisão fixada
- 4.5 Exercícios

Capítulo 5

Estimação de Proporções

Placeholder

- 5.1 Parâmetros populacionais
- 5.2 Estimação sob amostragem aleatória simples com reposição
- 5.3 Estimação sob amostragem aleatória simples sem reposição
- 5.4 Distribuição amostral exata de estimadores de proporção sob AASC e AAS
- 5.5 Intervalos de confiança para proporções na amostragem aleatória simples
- 5.6 Intervalos de confiança utilizando a aproximação Normal
- 5.7 Cálculo do tamanho da amostra
 - 5.7.1 Cálculo do n utilizando outras formas de representar o erro amostral
- 5.8 Estimação de proporções para variáveis com mais de duas categorias
- 5.9 Exercícios

Capítulo 6

Estimação de Razões e Funções de Totais

Placeholder

- 6.1 Razões populacionais como parâmetros de interesse
- 6.2 Razão de médias *versus* média de razões
- 6.3 Outras funções de totais
- 6.4 Estimadores de razões de totais
- 6.5 Estimadores para outras funções de totais
- 6.6 Análise do vício do estimador da razão
- 6.7 Erro quadrático médio de \widehat{R}
- 6.8 Exercícios

Capítulo 7

Estimação para Domínios de Estudo

Placeholder

- 7.1 Domínios de estudo
- 7.2 Parâmetros de interesse para domínios
- 7.3 Estimação de parâmetros para domínios sob AAS
- 7.4 Intervalos de confiança para os parâmetros para o domínio
- 7.5 Propriedades condicionais (fixando n_c)
- 7.6 Estimação do total no domínio quando tamanho do domínio é conhecido
- 7.7 Comparação da eficiência dos estimadores de total no domínio
- 7.8 Estimação de proporções dentro do domínio
- 7.9 Exercícios

Capítulo 8

Amostragem Sistemática Simples

Placeholder

- 8.1 Ideia básica
- 8.2 Método de seleção da amostra
- 8.3 Composição das amostras sistemáticas simples
- 8.4 Probabilidades de inclusão na amostragem sistemática simples
- 8.5 Variáveis aleatórias indicadoras - propriedades
- 8.6 Estimação de totais sob amostragem sistemática simples
- 8.7 Estimação de médias sob amostragem sistemática simples
- 8.8 Estimação de uma proporção na amostragem sistemática simples
- 8.9 Variância dos estimadores sob amostragem sistemática simples
- 8.10 Estimação de variâncias dos estimadores na amostragem sistemática simples
- 8.11 Alternativas para seleção com amostragem sistemática simples
- 8.12 Exercícios

Capítulo 9

Outros Métodos de Amostragem com Equiprobabilidade

Placeholder

9.1 Amostragem binomial ou de Bernoulli

9.1.1 Método de seleção da amostra

9.1.2 Probabilidades de inclusão na amostragem binomial

9.1.3 Estimação do total sob amostragem binomial

9.1.4 Estimação da média sob amostragem binomial

9.1.5 Exemplos de aplicação da amostragem binomial

9.2 Amostragem inversa simples

9.2.1 Método de seleção da amostra inversa simples

9.2.2 Amostragem inversa simples para varredura por amostragem

9.2.3 Estimação de Horvitz-Thompson na amostragem inversa simples

9.2.4 Estimação baseada no estimador de Murthy na amostragem inversa simples

9.2.5 Exemplos de aplicação da amostragem inversa simples

9.3 Exercícios

Capítulo 10

Amostragem com Probabilidades Proporcionais ao Tamanho

Placeholder

10.1 Motivação

10.2 Estimação do total populacional

10.3 Amostragem PPT com reposição

10.3.1 Estimação do total sob amostragem PPT com reposição

10.4 Amostragem PPT de Poisson

10.4.1 Estimação do total sob amostragem de Poisson

10.5 Amostragem sequencial de Poisson

10.5.1 Estimação com amostragem sequencial de Poisson

10.6 Amostragem sistemática com PPT

10.6.1 Amostragem PPT sistemática com ordenação

10.7 Amostragem PPT de Pareto

10.7.1 Estimação do total sob amostragem PPT de Pareto

10.8 Exercícios

Capítulo 11

Amostragem Estratificada

Placeholder

11.1 O método geral

11.2 Amostragem estratificada simples

11.2.1 Método de seleção

11.2.2 Parâmetros populacionais

11.2.3 Estimação do total e da média populacional

11.2.4 Intervalos de confiança

11.3 Alocação da amostra nos estratos

11.3.1 Alocação proporcional

11.3.2 Alocação ótima

11.3.3 Tamanho da amostra estratificada simples

11.3.4 Comparação de alternativas de alocação da amostra

11.3.5 Alocação potência

11.3.6 Alguns problemas com alocação ótima

11.4 Definição dos limites dos estratos

11.5 Número de estratos na amostragem estratificada simples

11.6 Exercícios

Capítulo 12

Amostragem Conglomerada

Placeholder

- 12.1 Conceituação básica
- 12.2 Amostragem de áreas
- 12.3 Definições básicas e notação
- 12.4 Amostragem conglomerada em um estágio
- 12.5 Amostragem conglomerada em vários estágios
- 12.6 Notação na amostragem conglomerada em um estágio
- 12.7 Amostragem conglomerada em um estágio simples
 - 12.7.1 Estimação do total sob AC1S
 - 12.7.2 Estimação da média sob AC1S
- 12.8 Efeito do plano amostral
- 12.9 Amostragem conglomerada com PPT em um estágio
- 12.10 Amostragem conglomerada em dois estágios
 - 12.10.1 AC2 com AAS nos 2 estágios
 - 12.10.2 AC2 com PPT com reposição no primeiro estágio e AAS no segundo estágio
 - 12.10.3 AC2 com PPT de Poisson sequencial no primeiro estágio e AAS no segundo estágio
 - 12.10.4 Tamanho de amostra na amostragem conglomerada em dois estágios
- 12.11 Ideias básicas na escolha de unidades compostas
- 12.12 Exercícios

Capítulo 13

Estimadores de Calibração

Placeholder

13.1 Motivação**13.2 O problema****13.3 Estimadores de regressão como estimadores de calibração****13.4 Propriedades dos estimadores de calibração****13.4.1 Propriedades dos estimadores para grandes amostras****13.5 Estimadores de razão como casos particulares da calibração****13.5.1 Estimador de razão combinada sob AES****13.5.2 Estimador de razão separada sob AES****13.5.3 Comparação dos estimadores de razão combinada e separada****13.6 Alguns problemas práticos com calibração****13.6.1 Tamanhos de amostra pequenos****13.6.2 Pesos negativos ou extremos****13.6.3 Grande número de variáveis de pesquisa ou de grupos de modelagem****13.6.4 Grande número de variáveis auxiliares****13.6.5 Não resposta e calibração****13.7 Seleção de variáveis para calibração com não resposta****13.8 Indicadores de qualidade da calibração****13.9 Um guia para aplicação de calibração****13.10 Observações finais****13.11 Exercícios**

Magalhães, M. N. (2006). *Probabilidade e Variáveis Aleatórias* (Segunda ed, p. 411). EDUSP.

- Magalhães, M. N. e Lima, A. C. P. (2004). *Noções de Probabilidade e Estatística* (6a. edição, p. 392). EDUSP.
- Särndal, C. E.; Swensson, B. e Wretman, J. H. (1992). *Model Assisted Survey Sampling*. Springer-Verlag.