

Capítulo 1 Introdução

1.1 Motivação

A citação abaixo é atribuída ao famoso estatístico norte-americano W. E. Deming:

“In God we trust. All others must bring data.”

Esta citação traduz uma tendência bastante atual. Vivemos numa era em que a disponibilidade e o acesso a dados não têm precedentes na história da humanidade. Dados são demandados e utilizados para orientar a tomada de decisões em todos os níveis. Por tais motivos, a Estatística nunca esteve tanto em evidência. A evolução tem sido tão rápida nesse tema que deu origem à criação de uma nova ciência (*data science*) - ver Donoho (2017) - e profissão (*cientista de dados*).

Tais tendências confirmam que o conhecimento do mundo ao nosso redor depende muito de dados e de estatísticas, aqui entendidas como resumos dos dados. Além disso, os métodos estatísticos fornecem:

- a. **orientações para a obtenção de dados relevantes, atuais, precisos e custo-efetivos;**
- b. **ferramentas para modelagem e extração de conhecimento relevante dos dados; e**
- c. **orientações de como usar os dados e estatísticas para apoiar a tomada de decisões.**

É nesse contexto que se coloca nossa motivação para escrever um livro sobre *Amostragem*, a parte da Estatística que lida com os métodos para extrair amostras de populações, e para obter estimativas de parâmetros relevantes dessas populações a partir das amostras observadas.

Nossa experiência no Instituto Brasileiro de Geografia e Estatística (IBGE) e com assessoria para o planejamento e realização de pesquisas por amostragem nos convenceu da importância da *Amostragem* para a obtenção, a custo justo, de dados atuais, relevantes e confiáveis. Nossa experiência como docentes de cursos de Amostragem na Escola Nacional de Ciências Estatísticas (ENCE) e também em outras instituições nos mostrou que há espaço para um novo texto nessa área, que adote uma abordagem mais moderna para ensinar. Nesse livro adotamos uma abordagem que faz uso da ampla disponibilidade de dados

públicos e de software aberto e livre para mostrar como fazer para implementar os vários métodos de amostragem e de estimação cobertos no texto. O uso extensivo de exemplos ancorados na prática de pesquisas feitas no Brasil também é um dos diferenciais que nosso livro tem.

Uma questão de fundo é a da relevância da *Amostragem* como área de ensino e pesquisa na Estatística. Por um lado, o advento dos chamados *big data* ou *dados orgânicos*, na definição de Groves (2011), parece representar séria ameaça à cultura e prática das pesquisas por amostragem. Isto se dá porque agora é possível encontrar situações em que dados *universais, frequentes e de baixo custo* poderiam ser usados para responder perguntas sobre as populações de interesse, sem o necessário recurso ao uso da *Amostragem* para sua obtenção. Por outro lado, há evidências de que, sozinhos, dados orgânicos nunca serão suficientes para responder a todas as perguntas de interesse. Os motivos podem ser um ou mais dos seguintes:

- a. o número de variáveis ou medidas disponíveis para cada unidade nalgumas dessas novas fontes pode ser pequeno, ou as variáveis medidas podem não ser as que permitem responder adequadamente às perguntas de interesse;**
- b. a qualidade das medidas e respostas disponíveis pode ser insatisfatória por motivos ligados à forma de captura das informações;**
- c. a cobertura da população de interesse pode ser precária ou difícil de avaliar, levando ao risco de inferências enviesadas por efeito de seletividade não controlada.**

Em todos esses casos, pesquisas amostrais podem ser necessárias para suplementar as informações das fontes *orgânicas* de modo a permitir obter respostas adequadas às perguntas de interesse sobre as populações especificadas.

Além das questões já apontadas, haverá também situações em que a *Amostragem* será necessária para permitir analisar os dados orgânicos em tempo adequado ou com custo computacional acessível. Algumas das fontes de dados desse tipo são tão vastas que não é factível pensar que podem ser exploradas de forma exaustiva, e técnicas de amostragem serão então necessárias para extrair amostras adequadas dos dados orgânicos para exploração e análise.

É nesse contexto formado por novos desafios e oportunidades que apresentamos neste livro o conteúdo de um curso de *Amostragem* para aplicações em pesquisas de populações finitas. O material evoluiu da nossa experiência de ensinar na ENCE, tanto em disciplinas do curso

de Graduação em Estatística, como também na Pós-graduação em População, Território e Estatísticas Públicas, além de diversas outras experiências na apresentação de cursos curtos de capacitação ministrados sobre o tema.

1.2 Objetivos do livro

Nosso livro pretende cobrir o conteúdo essencial para um curso básico de *Amostragem*, em nível similar ao de textos como Bolfarine e Bussab (2005), Cochran (1977) ou Lohr (2010). Trataremos dos métodos fundamentais de *Amostragem*, sempre que possível ilustrando com exemplos sua utilização em pesquisas realizadas no Brasil ou também em outras partes.

Pretendemos que um leitor atento e dedicado consiga, ao final da leitura e estudo prático e interativo do material, usar de forma confiante os vários métodos de amostragem e de estimação apresentados e discutidos no livro. Pretendemos também que possa usar ferramentas de software modernas para implementar os vários métodos apresentados. Para docentes de cursos de *Amostragem*, pretendemos que o uso do livro sirva de apoio como material didático de referência e por fornecer uma boa coleção de exercícios para fixação da aprendizagem.

Tomando em conta as facilidades disponíveis com base nas modernas ferramentas computacionais usadas, todos os códigos de R usados na implementação dos exemplos e na solução de exercícios são fornecidos junto do material do livro. Também são fornecidos todos os dados usados nos exemplos e exercícios que não podem ser encontrados facilmente nas fontes originais.

1.3 Estrutura do livro

Este livro está organizado em quatorze capítulos, incluindo esta introdução.

1. **Introdução**
2. **Conceitos e Cadastros**
3. **Visão Geral da Amostragem e Estimação**
4. **Amostragem Aleatória Simples (AAS)**
5. **Estimação de Proporções**
6. **Estimação de Razões e Funções de Totais**

- 7. Estimação para Domínios de Estudo**
- 8. Amostragem Sistemática Simples (AS)**
- 9. Amostragem Binomial ou de Bernoulli (AB)**
- 10. Amostragem com Probabilidades Desiguais**
- 11. Amostragem Inversa**
- 12. Amostragem Estratificada**
- 13. Amostragem de Conglomerados em Um ou Mais Estágios**
- 14. Estimadores de Calibração**

Cada um dos capítulos é auto-contido, e vários deles podem ser saltados. Por exemplo, os capítulos sobre Amostragem Sistemática, Amostragem Binomial e Amostragem Inversa podem ser omitidos num curso que tenha carga horária insuficiente para cobrir todo o material. O capítulo sobre Estimadores de Calibração tem grande importância prática, e pode ser apresentado antes do tratamento da Amostragem de Conglomerados, se necessário.

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).

Nossa abordagem não busca ser exaustiva, e por esse motivo, são poucas as provas que incluímos dos resultados aqui discutidos. Aos leitores interessados em verificar os resultados, recomendamos a consulta ao excelente livro de Särndal, Swensson, e Wretman (1992), ou então às muitas referências incluídas ao longo do texto.