

Mauricio Teixeira Leite de Vasconcellos

Escola Nacional de Ciências Estatísticas,
Fundação Instituto Brasileiro de Geografia e
Estatística. Rio de Janeiro, RJ, Brasil

Correspondência | Correspondence:
Maurício T. L. de Vasconcellos
R. André Cavalcanti, 106
20231-050 Rio de Janeiro, RJ, Brasil
E-mail: mtlv@ibge.gov.br

Comentário: Eunice Pinho de Castro Silva, o despontar da pesquisa por amostragem na saúde pública brasileira

Eunice Pinho de Castro Silva, the rise of public health research by sampling in Brazil

O artigo de Eunice Pinho de Castro Silva, objeto do presente comentário, representa um marco na história da pesquisa por amostragem na área de saúde pública no Brasil, ampliando o saber-fazer vigente à época, com elementos que surpreendem por sua atualidade e pelo cuidado e rigor com que foram descritos.

Meus comentários procuram, inicialmente, situar a contribuição trazida por esse artigo no contexto da pesquisa por amostragem no Brasil à época de sua publicação, buscando, preliminarmente, aspectos históricos do desenvolvimento desta ferramenta estatística. Estabelecido o contexto, os comentários são voltados ao artigo em si e divididos em dois pontos: os elementos básicos do desenho da amostra e o método de estimação.

Breve histórico da pesquisa por amostragem

A pesquisa por amostragem tem seu nascimento no ano de 1895, quando Kiar⁹ apresentou suas idéias sobre o método representativo na reunião do *International Statistical Institute (ISI)*. Kiar pressupunha ser possível, a partir de um conjunto de unidades relativamente pequeno, representar toda uma população e obter informações sobre essa população. Apesar da simplicidade de suas idéias, o método representativo – hoje conhecido como amostragem probabilística – não teve, inicialmente, grande aceitação. Na reunião do ISI⁸ em 1926, a aplicação desse método nas pesquisas realizadas pela estatística oficial dos países foi rejeitada pelo grupo de trabalho constituído por representantes dos órgãos nacionais de estatística. Eles tinham grande preocupação com o uso de um método novo e desconhecido da população, que poderia gerar críticas sobre a validade de suas pesquisas e diminuir a credibi-

lidades desses órgãos, ponto fundamental para a obtenção de respostas acuradas.

Em uma das diversas comemorações do centenário do artigo de Kiar, Leslie Kish¹⁰ sustenta que o artigo de Kiar pode ser considerado como a data de nascimento da pesquisa por amostragem, tornando a pesquisa estatística economicamente viável. Embora pesquisas estatísticas já eram realizadas desde 1820 quando Adolphe Quételet (1796-1874), matemático e estatístico belga, Secretário Permanente do Observatório Real de Bruxelas, fundou a estatística baseada no conceito de população. De fato, Theodore Porter¹¹ mostra que o nome e o campo da Estatística surgiram quando Quételet mostrou que as leis de probabilidade podiam ser aplicadas a populações reais de indivíduos.

Kish dividiu os 100 anos desde o trabalho de Kiar em duas partes iguais, “*quando, no fim da segunda guerra mundial, a pesquisa por amostragem decolou*” (Kish,¹⁰ 1996:16). E prossegue, afirmando que a pesquisa por amostragem decolou em consequência dos avanços acelerados da modelagem em amostragem e de sua consolidação em livros textos (Yates,¹³ 1949; Deming,⁴ 1950; Cochran,² 1953; Hansen et al.,⁶ 1953; Sukhatme,¹² 1954), que sistematizaram o conhecimento e facilitaram a criação e ampliação dos cursos de amostragem nas universidades.

A pesquisa por amostragem no Brasil

As pesquisas domiciliares de consumo de alimentos surgem no País na década de 1930 com os trabalhos pioneiros de Josué de Castro¹ e Horace Davis,³ e tiveram um grande impulso nos anos seguintes com a legislação do salário-mínimo, que determinava a necessidade de estudos sobre o custo de vida para fixa-

ção do valor do salário-mínimo local. No entanto, nenhuma das pesquisas realizadas nos anos 30 e 40 adotou técnicas de amostragem probabilística.

No âmbito do Sistema Estatístico Nacional, a pesquisa por amostragem nasceu no Censo Demográfico 1960 (IBGE,⁷ 2003) com a seleção, em cada setor censitário, de uma amostra sistemática de 25% dos domicílios do setor, para aplicação do Boletim da Amostra, um questionário mais abrangente do que o usado para os demais domicílios.

Somente em 1967, data da publicação do artigo de Eunice Pinho de Castro Silva, a Pesquisa Nacional por Amostra de Domicílios (PNAD) deu seus primeiros passos. Começou, inicialmente nos Estados da Guanabara e Rio de Janeiro (região I) e São Paulo (região II), e paulatinamente foi incluindo as demais regiões PNAD até que, em 1973, seu processo de implantação foi concluído, sem contudo considerar as áreas rurais das macrorregiões Norte e Centro-Oeste. A cobertura completa do território nacional foi completada na PNAD 2004, com a inclusão das áreas rurais dos estados da região Norte, exceto Tocantins cuja área rural já vinha sendo abrangida pela PNAD.

Elementos básicos do desenho da amostra

A autora iniciou o artigo apresentando as restrições operacionais que determinaram o desenho de amostra utilizado e justificou sua publicação pelo procedimento apresentado, que considerou útil para os que convivem com restrições análogas. O artigo estabeleceu claramente os elementos básicos da amostra selecionada, contrapondo os procedimentos ideais a aplicar com os que foram possíveis aplicar diante das limitações de tempo e custo a que estava subordinada a seleção da amostra.

A população-alvo foi descrita (mulheres com idade de 15 a menos de 50 anos), assim como as limitações que impediam sua estratificação por variáveis socioeconômicas. As três etapas do desenho de amostra ideal foram apresentadas: seleção de áreas, de domicílios e de mulheres da população-alvo. Em seguida, os passos que seriam necessários para seleção da amostra ideal foram descritos, deixando claro aos leitores a impossibilidade de sua aplicação diante das limitações operacionais, abrindo espaço para a descrição da engenhosa solução aplicada.

O tamanho da amostra foi estabelecido em função dos recursos financeiros disponíveis, como ocorre até hoje na maioria das pesquisas por amostragem. Em seguida, foram descritos os demais elementos básicos da amostra, tais como: a estratificação geográfica;

a seleção das áreas, ou pontos de partidas no mapa; a seleção, em campo, do domicílio inicial que definia o segmento do quarteirão que o entrevistador deveria percorrer até selecionar os três domicílios com moradores da população-alvo; e a seleção da mulher a ser entrevistada, caso houvesse mais de uma moradora no domicílio que pertencesse à população-alvo.

O artigo detalhou todo o processo de seleção dos pontos de partida, dando à comunidade científica um roteiro didático dos procedimentos de seleção do primeiro estágio da amostra, e estabeleceu as regras operacionais para o trabalho de campo.

Sobre este último ponto, é interessante observar que as regras definidas seguem os mesmos princípios adotados nos Censos Demográficos e demais pesquisas domiciliares do Instituto Brasileiro de Geografia e Estatística (IBGE). Por exemplo, o percurso de quarteirões no sentido horário, tendo a área de pesquisa (os domicílios) sempre ao lado direito do pesquisador; a entrada em vilas e retorno à face do quarteirão; o percurso em edifícios, do andar mais alto ao mais baixo. Ao observar estas regras propostas internacionalmente e endossadas pelas Nações Unidas, a autora produziu um texto que serve de referência até os dias de hoje.

Em relação às não-respostas, o artigo definiu claramente o protocolo a adotar em cada caso específico. De forma resumida, as regras estabelecidas buscavam fazer outras visitas para obter a entrevista ou, em caso de impossibilidade, substituir o domicílio por outro na seqüência do percurso estabelecido para o segmento do quarteirão.

É fato que qualquer solução de tratamento de não-respostas introduz viés na amostra. No entanto, a não-resposta é uma ocorrência inevitável em qualquer pesquisa domiciliar e as soluções atualmente adotadas são: aumento do tamanho da amostra; substituição de unidades; ou amostragem inversa (Haldane⁵). Em todos os casos, correções compensatórias devem ser feitas nos pesos amostrais. Mas o relevante é que a autora, novamente, forneceu um guia didático para a comunidade científica lidar com essas situações inerentes às pesquisas domiciliares.

Métodos de estimação

Dos elementos básicos da descrição de um desenho de amostra, o único ausente no artigo foi a descrição dos métodos de estimação ou de cálculo dos pesos amostrais.

No entanto, a descrição apresentada mostra uma pre-

ocupação de estabelecer um método de seleção que conduzisse a uma amostra autoponderada: a alocação do tamanho da amostra nos estratos geográficos (subdistritos) foi proporcional; a seleção dos pontos de partidas foi proporcional à área dos estratos desenhados nos mapas; e foi selecionado um número fixo de domicílios a entrevistar para cada ponto de partida. Esse esquema conduz a um peso amostral constante.

É curioso observar que o desenho da amostra da

PNAD, nascido à mesma época do artigo, também tem um esquema probabilístico que conduz a uma amostra autoponderada, apesar de o processo de estimação abandonar esta característica e calibrar os pesos amostrais para totais projetados da população brasileira.

Concluindo, gostaria de parabenizar a Revista de Saúde Pública pelo seu quadragésimo aniversário e pela publicação de artigos pioneiros na seção de Clássicos dos Primeiros Dez Anos.

REFERÊNCIAS

1. Castro J. O valor da alimentação: estudo econômico das condições de vida das classes operárias no Recife. *Bol Min Trab Ind.* 1935;5:177-31.
2. Cochran WG. *Sampling techniques.* New York: Wiley; 1953.
3. Davis HB. Padrão de vida dos operários da cidade de São Paulo. *Rev Arq Munic São Paulo.* 1935;13:113-66.
4. Deming WE. *Some theory of sampling.* New York: Wiley; 1950.
5. Haldane JBS. On a method of estimating frequencies. *Biometrika.* 1945;33:222-5.
6. Hansen MH, Hurwitz WN, Madow WG. *Sample survey methods and theory.* New York: Wiley; 1953. v. 2.
7. Instituto Brasileiro de Geografia e Estatística - IBGE. *Metodologia do Censo Demográfico 2000.* Rio de Janeiro; 2003.
8. International Statistical Institute - ISI. Report on the representative method in statistics. *Bull Int Stat Inst.* 1926;22:359-438.
9. Kiar AN. Observations et expériences concernant les dénombrements représentatifs. *Bull Int Stat Inst.* 1895;9:176-205.
10. Kish L. The hundred years' wars of survey sampling. In: *Proceedings of the Conference to Commemorate 100 Years of Sample Survey; 1995; Rome.* Roma: Centro d'Informazione e Stampa Universitaria; 1996. p. 15-27.
11. Porter TM. *The rise of statistical thinking, 1820-1900.* Princeton (NJ): Princeton University Press; 1986.
12. Sukhatme PV. *Sampling theory of surveys with applications.* Ames: Iowa State College Press; 1954.
13. Yates F. *Sampling methods for censuses and surveys.* London: Charles Griffin; 1949.