

Desenho da amostra e participação no Estudo Educatel

Sample design and participation in the
Educatel Study

Diseño de la muestra y participación en el
Estudio Educatel

Marcel de Toledo Vieira ¹

Rafael Moreira Claro ²

Ada Ávila Assunção ³

doi: 10.1590/0102-311X00167217

Resumo

O Educatel Brasil 2015/2016 foi um estudo transversal, realizado por entrevista telefônica, com o objetivo de oferecer informações a respeito da saúde e absenteísmo de professores da Educação Básica no país. O plano amostral adotado, com representatividade nacional, teve como base o método de amostragem estratificada simples, com estratificação definida visando a atender aos domínios de análise estabelecidos para o estudo (cinco regiões geográficas, duas áreas censitárias, quatro faixas etárias, sexo, três dependências administrativas da escola, cinco tipos de vínculo e seis etapas de ensino) e seleção por amostragem aleatória simples de professores dentro de cada estrato. O sorteio dos professores foi baseado no Censo Escolar de 2014, realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Dos 2.229.269 professores registrados no censo, foram selecionados 13.243. Foram realizadas 119.378 ligações telefônicas, as quais permitiram identificar 7.642 professores elegíveis (57,7% do total inicialmente sorteado). Ao final, foram completadas 6.510 entrevistas, 85,2% de taxa de resposta. Ao término da coleta, pesos amostrais foram associados a cada um dos professores entrevistados. Esses fatores de ponderação estão conectados não apenas ao desenho amostral do Educatel, mas também aos termos de ajuste para tratamento dos casos de não resposta ocorridos durante o processo de coleta dos dados.

Amostragem Aleatória e Sistemática; Inquéritos e Questionários; Professores Escolares; Ensino Fundamental e Médio

Correspondência

M. T. Vieira

Departamento de Estatística, Instituto de Ciências Exatas,
Universidade Federal de Juiz de Fora.

Rua José Lourenço Kelmer s/n, Juiz de Fora, MG
36036-900, Brasil.

marcel.vieira@ice.ufjf.br

¹ Instituto de Ciências Exatas, Universidade Federal de Juiz de
Fora, Juiz de Fora, Brasil.

² Escola de Enfermagem, Universidade Federal de Minas Gerais,
Belo Horizonte, Brasil.

³ Faculdade de Medicina, Universidade Federal de Minas
Gerais, Belo Horizonte, Brasil.



Introdução

O planejamento de amostras probabilísticas eficientes para abordar populações em pesquisas é tratado de forma exaustiva na literatura especializada ^{1,2,3}. Tal cenário atende tanto à demanda de redução dos custos dos inquéritos quanto àquela do aumento de sua agilidade (seja para a coleta dos dados, seja para a divulgação dos resultados). Por outro lado, sabe-se que o percentual de participação em estudos epidemiológicos vem reduzindo ao longo dos anos, com efeitos indesejáveis sobre a validade interna da pesquisa ⁴. Com isso, planos amostrais complexos que buscam a combinação de diferentes métodos probabilísticos para a seleção de uma amostra são considerados avançados em relação ao método da amostragem aleatória simples (AAS). Amostragens complexas têm sido adotadas com frequência cada vez maior, principalmente na área da saúde quando se tem grandes amostras. Exemplificando, a *Pesquisa Nacional de Saúde* (PNS), de base domiciliar, abrange o território nacional com o objetivo de estudar a situação de saúde e os estilos de vida da população brasileira; para tanto se vale de processo de amostragem por conglomerados em três estágios de seleção (são selecionados setores censitários no 1º estágio; domicílios nos setores em um 2º; e um adulto entre os moradores do domicílio no 3º estágio) ⁵. A *Pesquisa Nacional por Amostra de Domicílios* (PNAD), que também busca periodicamente informações sobre a saúde da população brasileira, utiliza plano amostral semelhante, por conglomerados que ora considera dois estágios e ora considera três estágios de seleção ⁶. Já o estudo *Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico* (VIGITEL), conduzido por entrevista telefônica nas 26 capitais de estados e Distrito Federal do país – com o objetivo de monitorar a magnitude e a variação na frequência dos principais fatores de risco e proteção contra as doenças e os agravos não transmissíveis –, vale-se de estratégia distinta, com o sorteio de linhas telefônicas fixas (realizado com base em cadastros fornecidos pelas principais operadoras do país) sucedido pela seleção de um adulto entre os residentes de cada domicílio contatado ⁷.

Em termos gerais, a adoção de um plano amostral complexo viabiliza obter estimativas com medidas de precisão pré-estabelecidas em inquéritos populacionais, adicionando a vantagem dos baixos custos e agilidade na coleta se comparada aos estudos que empregam planos amostrais menos sofisticados ². Por essa razão, uma parcela expressiva dos estudos epidemiológicos de grande porte no país se vale atualmente de planos amostrais dessa natureza. Ao reduzir o número do contingente dos participantes, adiciona a vantagem de interromper menos pessoas em seus cotidianos para participar da pesquisa. Sabe-se das dificuldades da pesquisa de campo e das dificuldades das pessoas que, apesar de convidadas a participar, se veem diante de restrições temporais ou de outro tipo qualquer. Dessa feita, a redução do tamanho amostral propicia mais vantagens para a logística da pesquisa.

O Educatel Brasil 2015/2016 foi um estudo transversal com representatividade nacional, realizado por meio de entrevista telefônica, valendo-se de amostra probabilística dos professores da Educação Básica do país, cuja coleta foi realizada entre o último trimestre de 2015 e o primeiro trimestre de 2016. O objetivo foi oferecer à comunidade acadêmica e aos gestores públicos uma fonte de dados sobre essa população ⁸, fonte esta de caráter inédito ao abordar os temas saúde, prevalência de morbidades e acidentes, absenteísmo e seus motivos. Definiu-se absenteísmo no Estudo Educatel como pelo menos uma falta ao trabalho por pelo menos um dia nos últimos 12 meses.

Neste artigo, são apresentados os procedimentos adotados para a seleção da amostra do Educatel, bem como as bases para o cálculo do seu tamanho amostral. Além disso, são discutidos os aspectos: (a) do cadastro de professores utilizado para a seleção da amostra; e (b) dos objetivos da pesquisa em termos da definição dos domínios (ou subpopulações) para os quais são desejadas estimativas com precisão controlada. Uma vez finalizada a coleta, procedeu-se a análise das características da amostra efetivada a fim de identificar as frequências de participação (de elegíveis) por macrorregiões, situação da escola (rural ou urbana), faixa de idade, sexo, dependência administrativa da escola e vínculo de trabalho. Pesos amostrais, ajustes para não resposta e procedimentos de estimação são também abordados.

O Educatel foi aprovado pelo Comitê de Ética em Pesquisa (COEP) da Faculdade de Medicina da Universidade Federal de Minas Gerais (CAAE: 48129115.0.0000.5149).

População-alvo e plano amostral

A população-alvo do EducateL é composta por professores que atuavam (em sala de aula) na Educação Básica no Brasil em 2015. Sendo assim, os procedimentos de amostragem empregados visaram a obter uma amostra probabilística do universo de professores em escala nacional.

Tendo em vista a complexidade logística e o alto custo envolvidos na realização de entrevistas presenciais quando se almeja a representatividade territorial com tal magnitude, optou-se pela utilização de entrevista telefônica assistida por computador (CATI) ⁸.

A estrutura do plano de amostragem adotado baseou-se no método de amostragem estratificada simples. Utilizou-se estratégia de estratificação definida, com vistas a atender aos domínios de análise (subgrupos populacionais) pré-estabelecidos para o estudo, conforme será detalhado adiante. O trabalho de planejamento da amostra levou em consideração os resultados de estudos que enfocaram grupos de professores e a experiência metodológica do VIGITEL ⁷.

O sucesso dos inquéritos por amostragem probabilística depende da identificação prévia da população de pesquisa e de seus domínios de análise, a fim de garantir tamanhos amostrais suficientes para a obtenção de estimativas relativas a cada um dos domínios pertinentes. Sendo assim, visando ao espalhamento da amostra e captação da heterogeneidade da população de pesquisa, a estratificação da população foi definida segundo um plano no qual foram combinadas as categorias das seguintes variáveis: (a) grandes regiões: Norte, Nordeste, Centro-oeste, Sudeste e Sul; (b) situação da escola: urbana e rural; (c) faixas etárias: ≤ 34 anos, entre 35 e 44 anos, entre 45 e 54 anos e ≥ 55 anos; (d) sexo: masculino e feminino; (e) dependência administrativa da escola: estadual, municipal, privada e outras; (f) tipo de vínculo: concursado/efetivo/estável, contrato temporário, rede privada, contrato CLT e outros; e (g) etapa de ensino: infantil, fundamental, médio, EJA (educação de jovens e adultos), profissional e outros. É importante notar que as variáveis de estratificação (c), (d), (f) e (g) refletem as características dos professores, e as variáveis de estratificação (a), (b) e (e) refletem características das escolas onde os professores trabalham.

Ressalta-se que o número de estratos possíveis com base na combinação de todas as categorias das variáveis listadas anteriormente é 9.600. Entretanto, apenas 7.650 desses estratos possíveis de fato existiam na população de pesquisa, ou seja, apresentavam um ou mais professores. Dessa forma, adotou-se a seguinte estratégia de alocação da amostra nos estratos existentes: (a) nos 1.549 estratos em que o tamanho populacional era 1, o tamanho da amostra foi fixado em 1 também; e (b) nos demais estratos existentes a amostra foi alocada de maneira proporcional ao tamanho populacional do estrato, arredondando sempre para o inteiro superior mais próximo, sendo que este arredondamento fez com que em outros 5.504 estratos o tamanho da amostra fosse também fixado em 1. Cabe ressaltar que, em princípio, a estimação de variâncias de estimadores depende de tamanhos amostrais de tamanho maior ou igual a 2. Sendo assim, a existência de estratos com tamanho amostral unitário exigirá a adoção de métodos de estimação que sejam capazes de lidar com tal característica. Esses métodos estão disponíveis em softwares tais como R (pacote *survey*), Stata (comandos *svy*) e SAS, por exemplo.

Faz-se necessário aqui também fazer a distinção entre estratos (para fins de seleção e espalhamento da amostra) e domínios de análise (para fins de posterior estimação) pré-estabelecidos para o estudo, com o objetivo de se garantir níveis de precisão aceitáveis pré-estabelecidos. No Estudo EducateL foram definidas como domínios de análise as categorias (marginais) das variáveis de estratificação (e não todas as combinações de todas as suas categorias). Para cada um desses domínios de análise, a alocação da amostra nos estratos existentes da forma como foi descrita antes resultou, após as perdas na amostra final realizada, na produção de estimativas de prevalência de absenteísmo com margem de erro de no máximo: (a) 3% para as categorias de grandes regiões; (b) 2,5% para as categorias de áreas censitárias; (c) 3,5% para as categorias de faixas etárias; (d) 2% para categorias de sexo; (e) 4% para as categorias de dependências administrativas da escola; (f) 4% para as categorias de tipo de vínculo; e (g) 4,5% para as categorias de etapas de ensino.

O cadastro básico para a seleção da amostra de professores para o Estudo EducateL tomou os dados do *Censo Escolar* de 2014, do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), que informa os números dos telefones das escolas. Esse cadastro trazia informações da população de pesquisa – 2.229.269 professores –, incluindo uma variável de identificação e outras que permitiram a construção das variáveis de estratificação citadas: sexo, idade, local de funcionamento,

etapa de ensino, rede administrativa e área censitária. Adicionalmente, variáveis de caracterização das escolas foram também obtidas, removendo a necessidade de que fossem novamente aferidas no Educatel, como: tamanho da escola segundo o número de professores, acesso ao abastecimento de água, disponibilidade de água filtrada, abastecimento de energia elétrica, esgoto sanitário, destinação do lixo, equipamentos e dependências da escola ⁸.

Vale mencionar que os dados do *Censo Escolar* de 2014 eram a base de dados mais atualizada sobre a população estudada disponível na época do sorteio amostra do Educatel, permitindo boa aproximação ao universo de professores atuantes em 2015 (ano de realização da coleta). Dessa forma, a população de pesquisa do Estudo Educatel é constituída professores que constavam do cadastro do *Censo Escolar* de 2014 e que atuavam (em sala de aula) na mesma escola em 2015. A população de pesquisa é, portanto, um subconjunto da população-alvo definida anteriormente. Essa distinção é necessária, uma vez que todas as inferências estatísticas realizadas com base nos dados coletados pelo Estudo Educatel são dirigidas à população de pesquisa definida segundo o cadastro considerado. Além disso, limitou-se também a população de pesquisa aos professores que responderam a pelo menos uma das primeiras 15 tentativas de contato (feitas em dias e horários variados, incluindo sábados e períodos noturnos), e também àqueles que trabalhavam em escolas com telefone operante.

Ainda assim, dado o período entre a realização do *Censo Escolar* de 2014 e a sua efetiva utilização no Educatel (pouco mais de um ano), algumas discrepâncias foram encontradas entre a listagem de escolas, os professores sorteados e o quadro de efetivos identificados no contato por telefone. Falecimentos, aposentadorias, mobilidade no mercado de trabalho foram mais frequentes do que o esperado. Vale observar que os microdados de cada censo são disponibilizados no semestre seguinte ao ano de referência da coleta, fato que inviabilizou a utilização do *Censo Escolar* de 2015 (divulgado apenas quando a coleta de dados do Educatel já se encontrava em andamento). Em síntese, a lista de participantes elegíveis para o estudo que foi produzida incluiu professores não mais inseridos nas escolas, que portanto não mais pertenciam à população-alvo do estudo. Doravante, essa situação será considerada o “problema (1) do cadastro”. Além disso, aqueles inseridos no mercado após 2014, ou seja, após a realização do *Censo Escolar* daquele ano, não estavam, obviamente, incluídos no cadastro utilizado. Dessa feita, a equipe de pesquisa deparou-se com o problema de cobertura do cadastro. Doravante, essa situação será considerada “problema (2) do cadastro”. Visando a dimensionar a magnitude dos problemas (1) e (2), com vistas a traçar estratégias para compensar possíveis vieses, foram considerados os dados do *Censo Escolar* de 2013 para os esclarecimentos cabíveis.

Comparando os cadastros do *Censo Escolar* de 2013 com os do censo de 2014, estimou-se em 16,62% a proporção de professores no cadastro considerado que poderiam ser classificados como problema (1) do cadastro, e em 24,28% a proporção de professores novos que não estariam contemplados no cadastro, ou seja, classificados como problema (2).

Optou-se pela utilização de amostragem estratificada pela possibilidade de classificação da população em estratos de acordo com características conhecidas da população de pesquisa. A seleção dos professores pertencentes a diferentes estratos foi realizada de forma independente (entre os estratos), considerando-se o método da AAS sem reposição, resultando então na adoção do método da amostragem estratificada simples. Essa estratégia permite maior homogeneidade dos subgrupos definidos de acordo com a estratificação do que seria encontrado se considerada toda a população. A adoção de planos amostrais que envolvam estratificação da amostra pode resultar no aumento da precisão das estimativas em relação a planos por AAS com o mesmo tamanho de amostra, além de permitir a estimação tanto para a população como um todo quanto para os subgrupos. Além disso, reduz-se o erro amostral, pois quanto mais homogêneos os subgrupos quanto às características dos seus componentes, maior a eficiência do dito procedimento amostral. Em síntese, o método para a seleção da amostra do Educatel foi elaborado e conduzido dessa forma.

A adoção de um desenho amostral alternativo que envolveria a seleção de conglomerados definidos como escolas foi também considerada. Entretanto, decidiu-se pela seleção direta de professores (mesmo com os desafios da adoção de um cadastro desatualizado) para evitar os efeitos de conglomeração que poderiam ter resultado no aumento do tamanho da amostra, de forma substancial. Efeitos de planos amostrais de desenhos conglomerados com frequência podem chegar ou ser maiores do que 4, o que poderia resultar no aumento do tamanho da amostra em 4 vezes, ou mais.

Cálculo do tamanho da amostra

O parâmetro de interesse central do Educatel, prevalência de absenteísmo por doença, norteou as definições fundamentais da amostragem, haja vista a necessidade de coerência entre o plano amostral, utilização de estimadores e natureza do que se buscou dimensionar. Levando em conta a problemática, o universo populacional, o estado da arte sobre a morbidade dos professores, as seguintes definições foram elaboradas para o cálculo do tamanho da amostra: (i) nível de 95% de confiança; (ii) 38% de prevalência (P) de, pelo menos, um episódio de ausência ao trabalho⁹; (iii) 0,99% de erro máximo (B) previsto (margem de erro) para a estimativa da prevalência de absenteísmo para toda a população de professores no Brasil, conforme definido em (ii); (iv) taxa de não entrevista ($tx1$) devido à recusa (ou outras formas de não resposta) de no máximo 20%; (v) taxa de não aplicação ($tx2$) de questionário devido ao problema (1) do cadastro de no máximo 20%; e (vi) correção para populações finitas. A margem de erro relatada anteriormente foi definida pela coordenação geral do estudo, levando-se em consideração aspectos tais como o orçamento, a logística da coleta e o cronograma.

O cálculo inicial do tamanho da amostra baseou-se em AAS, para que então fosse considerado o efeito do plano amostral (EPA). O *design effect* ou *deff* (EPA) de Kish é uma medida do efeito da amostragem sobre a variância dos estimadores¹⁰. O EPA é estimado baseando-se na razão entre a variância obtida considerando-se o plano amostral empregado e a variância obtida considerando AAS. Em função da ausência de pesquisas sobre professores da Educação Básica, com desenhos amostrais semelhantes, no plano hipotético, considerou-se o valor do EPA igual a 1 em procedimento respaldado pela literatura especializada¹¹. Enfim, sob tais premissas construiu-se o pressuposto segundo o qual não haveria perda de eficiência amostral se adotada a amostragem estratificada quando comparada à AAS. Dessa forma, foi considerada a seguinte expressão para o cálculo do tamanho da amostra do Estudo Educatel,

$$n = EPA \times tx1 \times tx2 \times \frac{NPQ}{(N-1)D + PQ}$$

com

$$D = \frac{B^2}{z_{\alpha/2}^2}.$$

Nessa expressão tem-se: EPA = efeito do plano amostral definido como igual a 1; $tx1 = 1,20$; $tx2 = 1,20$; $N =$ é o tamanho da população de pesquisa e é igual a 2.229.269 professores; $P = 0,38$; $Q = (1 - 0,38) = 0,62$; $B = 0,99$; e $z_{\alpha/2} = 1,96$, considerando o nível de 95% de confiança.

Dessa forma, o tamanho da amostra calculado foi de 13.243 professores. Ressalta-se que esse tamanho de amostra já leva em consideração ajustes visando a compensar possíveis perdas, por meio das $tx1$ e $tx2$. Sendo assim, foi necessário prever um número largamente superior quando do sorteio inicial, haja vista os resultados preliminares de um piloto que tentou observar a consistência da lista preparada à luz dos microdados do *Censo Escolar* de 2014 fornecidos pelo INEP. Por essa razão, 13.243 professores, distribuídos entre 11.042 escolas, foram selecionados de modo a aumentar as margens para a obtenção do número mínimo de entrevistas necessário para o sucesso da pesquisa.

Respeitando a definição apresentada anteriormente da população de pesquisa, foram considerados não elegíveis os professores que, na ocasião do contato para a participação no estudo, não mais trabalhavam na escola identificada no *Censo Escolar* de 2014 (durante o sorteio da amostra), aqueles que não responderam a 15 tentativas de contato (feitas em dias e horários variados, incluindo sábados e períodos noturnos), além daqueles que trabalhavam em escolas sem telefone ou nas quais o número do telefone do cadastro encontrava-se inoperante.

Participação no Estudo Educatel

Foram realizadas 119.378 ligações telefônicas, que permitiram identificar 7.642 professores elegíveis (57,7% do total inicialmente sorteado). Ao final, foram completadas 6.510 entrevistas, 85,2% de taxa de resposta. Estima-se em 1,18% a margem de erro permitida com base no tamanho da amostra realizado de 6.510 professores entrevistados. Foram necessárias em média 19 ligações por entrevista completa; o tempo médio das entrevistas foi de 12 minutos.

O desempenho do estudo, no que diz respeito ao percentual de respostas em cada uma das categorias das variáveis utilizadas na estratificação da amostra, pode ser avaliado por meio dos dados apresentados na Tabela 1.

Tabela 1

Número de professores sorteados, de professores elegíveis, de entrevistas realizadas e taxa de sucesso e de recusa segundo regiões geográficas, áreas censitárias, faixas de idade, sexo, dependências administrativas, vínculos de trabalho e etapas do ensino. Educatel Brasil, 2015/2016.

Variáveis	Linhas telefônicas		Total de entrevistas realizadas	Taxa (%) *	
	Sorteadas	Elegíveis (%)		Sucesso	Recusa **
Regiões geográficas					
Norte	2.604	1.093 (42,0)	1.000	91,5	3,8
Nordeste	2.640	1.338 (50,7)	1.150	85,9	7,0
Centro-oeste	2.499	1.562 (62,5)	1.297	83,0	5,6
Sudeste	2.880	1.940 (67,4)	1.671	86,1	7,6
Sul	2.620	1.709 (65,2)	1.392	81,5	9,3
Área censitária					
Urbana	9.358	5.842 (62,4)	4.979	85,2	7,1
Rural	3.885	1.800 (46,3)	1.531	85,1	6,2
Faixa de idade (anos)					
≤ 34	5.014	2.578 (51,4)	2.218	86,0	6,4
35-44	3.849	2.293 (59,6)	1.944	84,8	7,0
45-54	2.969	1.881 (63,4)	1.604	85,3	7,0
≥ 55	1.404	889 (63,3)	744	83,7	8,0
Sexo					
Masculino	4.916	2.793 (56,8)	2.394	85,7	6,6
Feminino	8.327	4.849 (58,2)	4.116	84,9	7,1
Dependência administrativa					
Estadual	3.406	2.062 (60,5)	1.780	86,3	6,2
Municipal	3.481	1.733 (49,8)	1.427	82,3	8,8
Privada	1.481	964 (65,1)	798	82,8	8,4
Outras	4.875	2.883 (59,1)	2.505	86,9	5,8
Vínculo de trabalho					
Concursado/Efetivo/Estável	4.370	2.805 (64,2)	2.381	84,9	7,3
Contrato temporário	2.931	1.379 (47,0)	1.180	85,6	6,8
Rede privada	1.481	964 (65,1)	798	82,8	8,4
Contrato CLT	418	224 (53,6)	187	83,5	4,9
Outros	4.043	2.270 (56,1)	1.964	86,5	6,1
Etapas do ensino					
Infantil	1.275	692 (54,3)	569	82,2	10,1
Fundamental	2.698	1.505 (55,8)	1.256	83,5	8,2
Médio	1.247	783 (62,8)	664	84,8	5,5
EJA	775	411 (53,0)	346	84,2	5,8
Profissional	592	376 (63,5)	322	85,6	6,6
Outros	6.656	3.875 (58,2)	3.353	86,5	6,3
Total	13.243	7.642 (57,7)	6.510	85,2	6,9

CLT: Consolidação das Leis do Trabalho; EJA: Educação de Jovens e Adultos.

Fonte: Núcleo de Estudos Saúde e Trabalho, Faculdade de Medicina, Universidade Federal de Minas Gerais.

* Taxas calculadas sobre o total de professores elegíveis;

** Taxa de recusa total, considerando-se a recusa tanto dos professores quanto das escolas.

Pesos amostrais e expansão da amostra

Uma vez finalizada a coleta de dados, as características da amostra efetiva foram cuidadosamente avaliadas. Os pesos amostrais conforme definidos estão conectados não apenas ao desenho amostral do Educatel, mas também aos termos de ajuste para tratamento dos casos de não resposta completa ocorridos durante o processo de coleta dos dados, sendo eles:

$$\pi_{hi} = n_h / N_h$$

a probabilidade de seleção do professor i pertencente ao estrato h ; e

$$d_{hi} = \frac{1}{\pi_{hi}}$$

o peso amostral básico do professor i pertencente ao estrato h .

De forma simples, o peso amostral básico reflete apenas os aspectos principais do desenho amostral, ou seja, o inverso das probabilidades de seleção dos professores. Entretanto, a ocorrência de não resposta, comum em pesquisas de grande porte, foi também observada no Educatel. Para o tratamento desses casos, consideramos o mecanismo gerador do tipo aleatório (MAR)¹², com o padrão de não resposta dependente das seguintes variáveis de estratificação: (a) grandes regiões, (b) localização (urbana, rural) e (d) sexo, da seguinte maneira:

$$\varphi_c = n'_c / n_c$$

o escore de propensão de respostas nos grupos c formados pela combinação das categorias das variáveis de estratificação (a), (b) e (d), em que n' é o tamanho da amostra realizada nos grupos c ; e

$$W_{hi} = \frac{d_{hi}}{\varphi_c}$$

o peso amostral ajustado para não resposta no estrato h .

Esses pesos amostrais devem ser utilizados na estimação de quaisquer medidas descritivas de interesse, calculadas com base nos dados da amostra do Educatel. Sua adoção assegura a abrangência dos diversos aspectos do esquema de seleção adotado, incluindo a estratificação, bem como permite ajustamentos para os efeitos da não resposta. A não consideração dos pesos amostrais na análise pode resultar em vício nas estimativas. Estimativas que consideram os pesos amostrais podem ser produzidas valendo-se de software como R (pacote *survey*), SUDAAN, SAS (PROC SURVEY), Stata (módulo *survey*) ou SPSS (pacote SPSS *Complex Samples*), por exemplo.

A estimação da precisão das estimativas de uma pesquisa, por sua vez, é um importante instrumento de avaliação da qualidade dos resultados produzidos. Medidas de precisão incluem o coeficiente de variação e intervalos de confiança, que têm como base os erros padrão estimados. Sendo assim, para o cálculo de estimativas de erros padrão, de proporções produzidas, recomenda-se para a análise da base de dados do Educatel a adoção de procedimentos já implantados nos pacotes citados anteriormente. Tais procedimentos produzem uma aproximação razoável das verdadeiras estimativas de precisão, que poderiam ser obtidas se o esquema de seleção adotado fosse considerado de maneira integral¹¹.

Considerações finais

As principais dificuldades de membros da comunidade acadêmica e gestores na compreensão e utilização de dados de um inquérito residem possivelmente na pouca exposição de características de seu delineamento. Este artigo busca atenuar essa dificuldade, apresentando uma exposição abrangente do desenho da amostra e da participação dos professores no Educatel, o mais abrangente estudo sobre questões de saúde e absenteísmo entre professores da Educação Básica já realizado no país.

Uma vez que a AAS não se mostrou viável em um inquérito de grande porte com as características do Educatel, fez-se necessário buscar um método de amostragem que permitisse representar os diversos domínios populacionais pretendidos e suas estimativas com níveis de precisão pré-estabelecidos. O desenho amostral probabilístico adotado foi o de amostragem estratificada simples. Essa estratégia permitiu maior homogeneidade dos subgrupos definidos na estratificação, favorecendo a precisão

das estimativas, a redução do erro amostral e a estimação destas tanto para a população como um todo quanto para os seus subgrupos. Por fim, pesos amostrais ajustados para correção da não resposta completa de professores foram calculados e discutidos.

A adoção de metodologia estatística apropriada para a análise de dados amostrais complexos, já disponível em diversos software, garante a produção de inferências estatísticas válidas e precisas para o absentismo de professores da Educação Básica no Brasil e para domínios específicos de análise pré-estabelecidos, tendo em vista os dados amostrais do Educatel.

Colaboradores

M. T. Vieira participou da concepção, delineamento, seleção da amostra, cálculo dos pesos amostrais, bem como da redação e revisão do artigo. R. M. Claro e A. A. Assunção participaram da concepção, delineamento, redação e revisão do artigo.

Agradecimentos

À Secretaria de Articulação de Sistemas de Ensino do Ministério da Educação, ao Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Conflitos de interesse

Os autores declaram não ter conflitos de interesse.

Referências

1. Bolfarine H, Bussab WDO. Elementos de amostragem. São Paulo: Editora Edgard Blücher; 2005.
2. Cochran WG. Sampling techniques. 3rd Ed. New York: John Wiley & Sons; 1977.
3. Särndal CE, Swensson B, Wretman J. Model assisted survey sampling. New York: Springer-Verlag; 1992.
4. Silva TLN, Klein CH, Souza AM, Barufaldi LA, Abreu GA, Kuschnir MCC, et al. Participação no Estudo de Riscos Cardiovasculares em Adolescentes – ERICA. Rev Saúde Pública 2016; 50 Suppl 1:3s.
5. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional de Saúde 2013: percepção do estado de saúde, estilos de vida e doenças crônicas. Brasil, grandes regiões e unidades da federação. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2014.
6. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional por Amostra de Domicílios Contínua: notas metodológicas. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2014.
7. Ministério da Saúde. Vigitel Brasil 2014. Saúde suplementar: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico. Brasília: Ministério da Saúde; 2015.
8. Assunção AA, Medeiros AM, Claro RM, Vieira MT, Maia EG, Andrade JM. Hipóteses, delineamento e instrumentos do Estudo Educatel Brasil, 2015/2016. Cad Saúde Pública 2019; 35 Suppl1:e00108618.
9. Santos MN, Marques AC. Condições de saúde, estilo de vida e características de trabalho de professores de uma cidade do sul do Brasil. Ciênc Saúde Coletiva 2013; 18:837-46.
10. Kish L. Survey sampling. New York: Wiley; 1965.
11. Pessoa DGC, Nascimento Silva PL. Análise de dados amostrais complexos. São Paulo: Associação Brasileira de Estatística; 1998.
12. Rubin DB. Inference and missing data. Biometrika 1976; 63:581-92.

Abstract

Educatel Brazil 2015/2016 was a cross-sectional study conducted by telephone interview with the aim of producing information on health and absenteeism in Brazilian schoolteachers. The nationally representative sampling plan was based on the simple stratified sampling method, with stratification defined to meet the analytical domains established for the study (five major geographic regions, two census areas, four age brackets, sex, three types of school administration, five types of teacher employment, and six grade levels) and selection by simple random sampling of teachers in each stratum. Teacher selection was based on the 2014 School Census conducted by the Brazilian National Institute for Educational Studies and Research "Anísio Teixeira". Of the 2,229,269 teachers recorded in the Census, 13,243 were selected. A total of 119,378 telephone calls were made, identifying 7,642 eligible teachers (57.7% of the total initially sampled). A total of 6,510 interviews were finally completed, for a response rate of 85.2%. At the end of data collection, sample weights were assigned to each of the teachers interviewed. These weighting factors are connected not only to the Educatel sample design, but also to the adjustment terms for treatment of non-response during the data collection process.

Random and Systematic Sampling; Surveys and Questionnaires; School Teachers; Education, Primary and Secondary

Resumen

Educatel Brasil 2015/2016 fue un estudio transversal, realizado mediante entrevista telefónica, con el objetivo de ofrecer información respecto a la salud y absentismo de profesores de Educación Básica en el país. El plan de muestras adoptado, representativo nacionalmente, tuvo como base el método de muestra estratificada simple, con una estratificación definida, con el fin de atender los dominios de análisis establecidos para el estudio (cinco regiones geográficas, dos sectores censitarios, cuatro franjas de edad, sexo, tres dependencias administrativas de la escuela, cinco tipos de vínculo y seis etapas de enseñanza), y selección por muestra aleatoria simple de profesores dentro de cada estrato. El sorteo de los profesores se basó en el Censo Escolar de 2014, realizado por el Instituto Nacional de Estudios e Investigaciones Educativas "Anísio Teixeira". De los 2.229.269 profesores registrados en el censo, fueron seleccionados 13.243. Se realizaron 119.378 llamadas telefónicas, que permitieron identificar a 7.642 profesores elegibles (57,7% del total inicialmente sorteado). Al final, se completaron 6.510 entrevistas, con un 85,2% de tasa de respuesta. Al término de la recogida, se asociaron los pesos de las muestras a cada uno de los profesores entrevistados. Estos factores de ponderación están conectados, no sólo por el diseño de muestras de Educatel, sino también por los términos de ajuste para el tratamiento de los casos de no respuesta producidos durante el proceso de recogida de datos.

Muestreo Aleatorio y Sistemático; Encuestas y Cuestionarios; Maestros; Educación Primaria y Secundaria

Recebido em 25/Set/2017
Versão final reapresentada em 03/Abr/2018
Aprovado em 20/Abr/2018