

Pesquisa Nacional de Saúde do Escolar - PeNSE

Nota técnica 01/2022

Controle estatístico de confidencialidade para ampliar a segurança da privacidade do informante no arquivo de microdados da PeNSE 2019

Esta Nota técnica tem por objetivo fornecer aos usuários em geral da Pesquisa Nacional de Saúde do Escolar - PeNSE 2019 informações sobre o tratamento de confidencialidade a que foram submetidos os microdados, que estão disponibilizados no portal do IBGE, estruturando-se em três eixos: (1) motivações para o tratamento de confidencialidade aplicado aos microdados; (2) procedimentos adotados; (3) implicações do tratamento para os resultados.

1. Motivações para o tratamento de confidencialidade aplicado aos microdados

As políticas de ampliação da segurança da privacidade do informante adotadas pelo IBGE têm seguido forte tendência internacional na produção de estatísticas oficiais. No Brasil, com a entrada em vigor da Lei Geral de Proteção de Dados Pessoais¹ (LGPD), em setembro de 2020, o IBGE vem gradualmente revendo e modernizando os processos referentes à garantia de privacidade dos informantes das pesquisas e demais operações estatísticas, o que reforça o seu compromisso com os Princípios Fundamentais das Estatísticas Oficiais² e o Código de Boas Práticas das Estatísticas do IBGE³.

A Pesquisa Nacional de Saúde do Escolar - PeNSE 2019 contou com um plano amostral desenhado para atender a uma nova demanda, referente ao aumento da precisão das estimativas para escolas privadas fora dos municípios das Capitais, o que exigiu maior cobertura dessas subpopulações.

É inerente ao processo de seleção amostral algum nível de proteção, uma vez que a investigação de apenas parte da população produz um conjunto de dados com quantidade de informação reduzida em relação ao cadastro de toda a população, permitindo, contudo, produzir

¹ BRASIL. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: jun. 2022.

² UNITED NATIONS. Statistical Commission. Report on the special session (11-15 April 1994). New York, 1994. 34 p. (E/1994/29 and E/CN.3/1994/18). Economic and Social Council, official records, 1994, suppl. n. 9. Disponível em: <http://unstats.un.org/unsd/statcom/94report.pdf>. Acesso em: jun. 2022.

³ IBGE. Código de Boas Práticas das Estatísticas do IBGE. 2 ed. Rio de Janeiro: IBGE, 2021. Disponível em <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101744.pdf>. Acesso em: jun. 2022.

estimativas suficientemente precisas para registrar o fenômeno de interesse. Ainda que a essência do Controle Estatístico de Confidencialidade (CEC – Statistical Disclosure Control) seja a redução da quantidade de informações disponíveis com a preservação da utilidade do conjunto de dados resultante, o plano amostral da pesquisa não é desenhado com esse objetivo. Com efeito, para algumas subpopulações específicas, a confidencialidade pode ser comprometida em favor de atingir a precisão desejada das estimativas.

Além dos aspectos de confidencialidade e precisão das estimativas, vale ressaltar que a PeNSE investiga alguns quesitos extremamente sensíveis em uma população de indivíduos menores de idade, o que motivou a realização de um estudo com o objetivo de identificar riscos graves à confidencialidade e buscar o nível de proteção adequado para o conjunto de microdados resultante. Mais especificamente, o conjunto de microdados da PeNSE 2019 foi submetido ao processo do CEC proposto por Hundepool et al. (2012)⁴. A metodologia adotada buscou mitigar os riscos mais severos à confidencialidade, sem comprometer, no entanto, a utilidade dos microdados e consequentemente produzir menor impacto possível aos usuários.

2. Procedimentos adotados

Como resultado desse processo, no conjunto de microdados disponível ao público foram adotados os seguintes procedimentos:

- ✓ A variável B01003 (Idade) foi agregada em 4 faixas (menores de 13 anos, de 13 a 15 anos, de 16 a 17 anos e de 18 anos ou mais).
- ✓ Vinte e uma (21) variáveis de escola foram totalmente suprimidas em função do seu potencial de reidentificação das unidades de análise.

E01P02A - Qual é o TOTAL de alunos(as) matriculados(as) na escola?

E01P04A - Qual é a quantidade TOTAL de salas de aula da escola?

E01P06A - A escola tem sala ou laboratório de informática para os(as) alunos(as) EM CONDIÇÕES DE USO?

E01P23 - A escola tem conselho escolar?

E01P24B - NOS ÚLTIMOS 12 MESES, com que frequência o conselho escolar se reuniu?

E01P15A - A escola tem quadra de esportes EM CONDIÇÕES DE USO?

E01P16A - Quantas quadras de esporte EM CONDIÇÕES DE USO a escola tem?

E01P17B - Quantas das quadras de esporte EM CONDIÇÕES DE USO da escola são cobertas?

E01P20 - A escola tem piscina EM CONDIÇÕES DE USO?

E01P36A - A escola tem material, acessórios ou equipamentos para prática de atividade física EM CONDIÇÕES DE USO? Exemplos: bola, corda, rede, raquete etc.

E01P21 - A escola tem vestiário para os(as) alunos(as) EM CONDIÇÕES DE USO?

.....
⁴ HUNDEPOOL, A. et al. Statistical Disclosure Control. Wiley, 2012. DOI 10.1002/9781118348239.

E01P37A - A escola tem vestiários separados por sexo EM CONDIÇÕES DE USO para os(as) alunos(as)?

E01P45A - A escola oferece comida/merenda para os(as) alunos(as)?

E01P46A1 - A escola oferece comida/merenda para alunos de quais etapas de ensino?

E01P46A2 - A escola oferece comida/merenda para alunos de quais turnos?

E01P47A - A escola tem cozinha (espaço onde são preparados os alimentos) EM CONDIÇÕES DE USO?

E01P48A - A escola possui refeitório (espaço estruturado destinado à alimentação, com assentos e mesas) EM CONDIÇÕES DE USO?

E01P50A - A escola tem água potável (adequada para beber) para consumo dos(as) alunos(as)?

E01P52A - Qual é a principal fonte de água potável (adequada para beber) da escola?

E01P53A. - A escola tem banheiros para os(as) alunos(as) EM CONDIÇÕES DE USO?

E01P54A - A escola tem banheiros separados por sexo EM CONDIÇÕES DE USO para os(as) alunos(as)?

- ✓ Nas variáveis B01001A (Sexo), B01003 (Idade) e/ou B08013A (Alguma vez na vida você engravidou?) houve supressão de registros em 0,3% dos casos.
- ✓ Na variável B01002 (Cor ou raça) houve supressão em 1,9% dos registros.

A categoria 9 (sem resposta) foi atribuída às respectivas variáveis dos registros suprimidos e não há meio de diferenciar esses casos dos marcados como sem resposta durante a aplicação do questionário. A proporção de registros afetados pela supressão parcial foi bastante reduzida e em um conjunto de variáveis limitado.

3. Implicações do tratamento para os resultados

Apesar da inviabilidade de explorar todos os cruzamentos e subpopulações possíveis, espera-se que, de maneira geral, os resultados encontrados a partir dos microdados modificados não apresentarão grandes diferenças em comparação com os microdados originais. Ainda assim, é importante ressaltar que há a possibilidade de que as análises a partir dos microdados modificados difiram dos resultados da pesquisa, em especial para a variável B01002 (Cor ou raça), mais impactada nesse tratamento e na qual a distribuição de supressões entre as categorias foi mais desigual.

Dessa forma, é esperado que, na ampla maioria dos casos, a expansão dos microdados públicos produza estimativas iguais ou muito próximas dos valores divulgados na publicação e tabelas de resultados disponibilizados no portal do IBGE. Entretanto, cruzamentos ou recortes mais restritos podem implicar discrepâncias de magnitude desconhecida. O usuário que, em

razão de necessidades específicas, não puder prescindir do uso do conjunto de microdados completo, deverá fazê-lo através da Sala de Acesso a Dados Restritos⁵ do IBGE.

13 de julho de 2022

Diretoria de Pesquisas

.....
⁵ IBGE. Guia do Usuário da Sala de Acesso a Dados Restritos. Rio de Janeiro: IBGE, 2021. Disponível em <https://www.ibge.gov.br/acesso-informacao/sala-de-acesso-a-dados-restritos.html>. Acesso em: jun. 2022.