



Censo Demográfico 2022

Notas metodológicas 01/2025

Educação: Resultados preliminares da amostra

Presidente da República

Luiz Inácio Lula da Silva

Ministra do Planejamento e Orçamento

Simone Nassar Tebet

**INSTITUTO BRASILEIRO
DE GEOGRAFIA E
ESTATÍSTICA - IBGE**

Presidente

Marcio Pochmann

Diretora-Executiva

Flávia Vinhaes Santos

ÓRGÃOS ESPECÍFICOS SINGULARES

Diretoria de Pesquisas

Gustavo Junger da Silva

Diretoria de Geociências

Maria do Carmo Dias Bueno

Diretoria de Tecnologia da Informação

Marcos Vinícius Ferreira Mazoni

Centro de Documentação e Disseminação de Informações

José Daniel Castro da Silva

Escola Nacional de Ciências Estatísticas

Paulo de Martino Jannuzzi

Ministério do Planejamento e Orçamento
Instituto Brasileiro de Geografia e Estatística - IBGE

Censo Demográfico 2022

Notas metodológicas 01/2025

Educação: Resultados preliminares da amostra



Rio de Janeiro
2025

Instituto Brasileiro de Geografia e Estatística - IBGE

Av. Franklin Roosevelt, 166 - Centro - 20021-120 - Rio de Janeiro, RJ - Brasil

© IBGE. 2025

Em virtude do prazo disponível para o cumprimento do cronograma editorial, os originais desta publicação não foram submetidos aos protocolos completos de normalização e editoração, sendo o seu conteúdo finalizado pela Unidade Responsável.

Sumário

Introdução.....	4
Identificação de erro de coleta no quesito de frequência à escola ou creche e procedimentos adotados	5
Metodologia de identificação de casos suspeitos	5
Imputação de dados.....	6
Aspectos relativos à data de referência, data de entrevista e impactos nos indicadores educacionais.....	8

Introdução

Essa breve nota, divulgada em conjunto com publicação “Censo Demográfico 2022: Educação: Resultados preliminares da amostra”, visa descrever de forma sucinta o procedimento especial de crítica e imputação de dados realizado para o bloco temático de Educação do questionário de Amostra do Censo Demográfico 2022, em função de um erro de aplicação do quesito frequência à escola ou creche, identificado nas primeiras semanas da coleta.

Adicionalmente, a última seção tece considerações sobre a data de referência e datas de entrevista do Censo Demográfico 2022, e sua relação com a produção de indicadores educacionais.

Identificação de erro de coleta no quesito de frequência à escola e procedimentos adotados

Logo nos primeiros dias de coleta do Censo Demográfico 2022, os sistemas de acompanhamento dos dados coletados indicavam que, em uma alta proporção de entrevistas, adultos maiores com 30 ou mais anos de idade haviam sido identificados como frequentando escola ou creche, resultado incompatível com o observado em registros administrativos ou em levantamentos amostrais do IBGE.

A partir de observações realizadas em campo, foi identificado que alguns recenseadores interpretavam incorretamente as opções de resposta do quesito “FREQUENTA ESCOLA OU CRECHE?”, assinalando a opção “1 – SIM” em situações nas quais deveria ser assinalada a opção “2 - NÃO, MAS JÁ FREQUENTOU”.

A partir dessa hipótese de que esse erro de interpretação era responsável pela alta proporção de adultos identificados como frequentando escola, ao longo das primeiras semanas de coleta foram realizados ajustes voltados à prevenção desse equívoco: o layout do quesito foi alterado, visando facilitar a visualização das opções de resposta; foi emitido um boletim informativo a toda a rede de coleta, reforçando a necessidade de atenção ao preenchimento correto dessa informação; por fim, o aplicativo de coleta foi atualizado para que houvesse exibição de uma mensagem de alerta toda vez que o recenseador assinalava que um morador de 30 anos ou mais frequentava escola.

Após a implementação dessas medidas, o acompanhamento das informações coletadas indicou que a proporção de adultos frequentando escola se reduziu sensivelmente nas novas entrevistas, atingindo um patamar compatível com as expectativas. No entanto, naquele momento já havia um estoque de entrevistas realizadas nas primeiras semanas de coleta e com possível erro na aplicação desse quesito – esse conjunto de entrevistas demandaram um tratamento especial nas fases de crítica e imputação de dados, descrito a seguir.

Metodologia de identificação de casos suspeitos

Partindo da hipótese de que o problema poderia ter origem na confusão de entendimento e preenchimento do questionário por parte de alguns recenseadores, a metodologia adotada para corrigir os dados focou na análise da concentração destes casos em recenseadores específicos.

A partir da premissa de que os dados estavam consistentes a partir da sexta semana de coleta, após as intervenções relatadas acima, definiu-se que seriam classificados como casos

suspeitos de preenchimento incorreto por parte dos recenseadores os que se encaixassem simultaneamente em todos os critérios abaixo:

1. Registro com entrevista realizada nas 5 primeiras semanas de coleta;
2. Registro de indivíduos acima de 18 anos com resposta “sim” à pergunta de frequência escolar, para o qual o curso frequentado não fosse de Educação de Jovens e Adultos (EJA);
3. Registro com entrevista realizada por recenseadores que coletaram ao menos 5 indivíduos com mais de 18 anos frequentando escola;
4. Taxa de frequência escolar para indivíduos maiores de 18 anos do recenseador específico maior que a referência do percentil 95 dos dados a partir da 6ª semana de coleta (P95=33% de frequência escolar para maiores de 18 anos)

A partir desses critérios, foram identificados 435 588 casos suspeitos que correspondiam a estes critérios, representando 2,7% dos registros das pessoas com mais de 18 anos. Simulando a retirada desses indivíduos do conjunto de pessoas que frequentavam escola, a frequência escolar por idade verificada nas 5 primeiras semanas de recenseamento ficou muito próxima ao observado a partir da sexta semana, reforçando a hipótese de erro nesses registros.

Imputação de dados

O tratamento destes dados foi realizado por imputação probabilística via CANCEIS (*CANadian Census Edit and Imputation System*), um sistema de edição e imputação de dados desenvolvido pelo Instituto Nacional de Estatística do Canada, StatCan, e usualmente empregado pelo IBGE para solução de problemas como esse.

Nesse processo de imputação probabilística, foram considerados como possíveis doadores da informação de frequência escolar os domicílios com características similares.

Como o fluxo do questionário era diferente conforme a resposta ao “FREQUENTA ESCOLA OU CRECHE?”, nos registros para os quais a resposta a esse quesito foi imputada para “NÃO, MAS JÁ FREQUENTOU”, foi necessário preencher os quesitos do fluxo determinado por essa resposta. Para isso, foi realizada uma compatibilização do fluxo de respostas, adaptando as respostas originalmente obtidas para os quesitos do fluxo “SIM” (“QUAL É O CURSO QUE FREQUENTA?”, “QUAL É O ANO QUE FREQUENTA?” e “QUAL É A SÉRIE QUE FREQUENTA?”) para obter as respostas dos quesitos do fluxo determinado pela resposta para “NÃO, MAS JÁ FREQUENTOU” (“QUAL FOI O CURSO MAIS ELEVADO QUE FREQUENTOU ANTERIORMENTE?”, “QUAL FOI O ÚLTIMO ANO QUE CONCLUIU COM APROVAÇÃO?” e “QUAL FOI A ÚLTIMA SÉRIE QUE CONCLUIU COM APROVAÇÃO?”).

Os quesitos do fluxo determinado pela resposta para “NÃO, MAS JÁ FREQUENTOU” que não puderam ser determinados pela forma descrita no parágrafo acima passaram por nova imputação probabilística.

Considerações relativas à data de referência e à data de entrevista

Anualmente, o IBGE divulga indicadores educacionais da população brasileira produzidos através da PNAD contínua, pesquisa amostral domiciliar. Embora a PNAD Contínua realize entrevista ao longo de todos os meses do ano, os indicadores educacionais são produzidos através das entrevistas referentes ao segundo trimestre do ano (meses de abril, maio e junho). Esse período é considerado preferível para aferição de indicadores educacionais, por localizar-se ainda na primeira metade do ano letivo e não coincidir com períodos de férias escolares.

No caso dos indicadores educacionais produzidos através do Censo Demográfico 2022, a data de referência é a mesma utilizada para coleta das demais informações do Censo, isto é, a meia noite entre os dias 31 de julho e 1º de agosto de 2022. Ainda que os quesitos do bloco de educação estivessem redigidos no tempo verbal presente e não fizessem referência explícita a data de referência, objetivava-se investigar as características educacionais da população recenseada tal qual elas se encontravam na data de referência.

Essas diferenças nos períodos de referências de diferentes levantamentos devem ser sempre levadas em consideração em eventuais comparações entre os resultados obtidos em diferentes pesquisas.

Cabe lembrar também que a coleta do Censo Demográfico 2022 prolongou-se até maio de 2023. Para parte da população recenseada, portanto, as entrevistas foram realizadas até 10 meses depois da data de referência, intervalo de tempo que abarcou férias escolares e mudança de ano letivo, período no qual ocorrem alterações das séries e níveis de ensino frequentados pelos alunos. O distanciamento entre a data de referência e a data de entrevista é um possível indutor de ruído nas informações coletadas.