



RELATÓRIO DE AMOSTRAGEM DO SAEB 2021

DIRETORIA DE AVALIAÇÃO
DA EDUCAÇÃO BÁSICA
DAEB

INEP MINISTÉRIO DA
EDUCAÇÃO

REPÚBLICA FEDERATIVA DO BRASIL

MINISTÉRIO DA EDUCAÇÃO | **MEC**

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS
EDUCACIONAIS ANÍSIO TEIXEIRA | **INEP**

DIRETORIA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA | **DAEB**





RELATÓRIO DE AMOSTRAGEM DO SAEB 2021

Brasília-DF
Inep/MEC
2023



Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep)
É permitida a reprodução total ou parcial desta publicação, desde que citada a fonte.

DIRETORIA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA (DAEB)

COORDENAÇÃO-GERAL DE MEDIDAS DA EDUCAÇÃO BÁSICA
(CGMEB)

EQUIPE TÉCNICA

Giordano Alan Barbosa Sereno

Joelson Severo dos Santos Azevedo

Laene Ascenso Lustosa

DIRETORIA DE ESTUDOS EDUCACIONAIS (DIREDE)

COORDENAÇÃO-GERAL DE EDITORAÇÃO E PUBLICAÇÕES (CGEP)
Priscila Pereira Santos

DIVISÃO DE PERIÓDICOS (DPE)

Roshni Mariana de Mateus

DIVISÃO DE PRODUÇÃO EDITORIAL (DPR)

Ricardo César Blezer

APOIO EDITORIAL

Janáina da Costa Santos

REVISÃO LINGUÍSTICA

Ana Claudia de Medeiros Rios

NORMALIZAÇÃO

Aline do Nascimento Pereira

REVISÃO GRÁFICA

Daniel Caixeta

PROJETO GRÁFICO CAPA/MIOLO

Raphael C. Freitas

DIAGRAMAÇÃO

José Miguel dos Santos

Dados Internacionais de Catalogação na Publicação (CIP)

Brasil. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep).
Relatório da amostragem do Saeb 2021 [recurso eletrônico]. – Brasília, DF : Inep,
2023.

33 p. il.

ISBN 978-65-5801-116-3

1. Avaliação da educação básica. 2. Levantamento de dados. 3. Informações
estatísticas. I. Título.

CDU 371.26

SUMÁRIO

ESTA PUBLICAÇÃO POSSUI SUMÁRIO INTERATIVO
PARA RETORNAR AO SUMÁRIO, CLIQUE NO NÚMERO DA PÁGINA EM CADA SEÇÃO

APRESENTAÇÃO.....	4
.....	
1	DESCRIÇÃO DO UNIVERSO DE REFERÊNCIA DO SAEB 20216
1.1	Objeto do Levantamento..... 7
1.2	Sistema de Referência 8
1.3	População de Referência da Amostra do Saeb 2021..... 8
2	PLANO AMOSTRAL 10
2.1	Domínios de interesse10
2.2	Plano Amostral.....12
2.3	Seleção das turmas nas escolas selecionadas15
3	EXPANSÃO DA AMOSTRA 16
3.1	Esquema geral de ponderação16
3.2	Descrição da ponderação da amostra20
3.3	Utilização dos pesos na estimação22
3.4	Estimação da precisão das estimativas23
3.5	Descrição da ponderação da parte censitária24
.....	
BIBLIOGRAFIA.....	27
APÊNDICE	28
ANEXO.....	33



APRESENTAÇÃO

O objetivo principal deste relatório é apresentar o processo de amostragem do Sistema de Avaliação da Educação Básica (Saeb) 2021. Além disso, divulgar informações relativas ao cálculo dos pesos para expansão dos resultados (pesos amostrais), tanto os obtidos a partir dos dados coletados por amostragem quanto os provenientes da aplicação censitária. O Saeb foi criado em 1990 como uma única avaliação e reestruturado, em 2005, quando houve a instituição da Aneb e da Avaliação Nacional do Rendimento Escolar (Anresc), conhecida como Prova Brasil. Enquanto a Aneb manteve as características, os objetivos e os procedimentos da avaliação da educação básica efetuada até 2005 pelo Saeb, a Prova Brasil foi criada com o objetivo de avaliar a qualidade do ensino ministrado nas escolas de ensino fundamental das redes públicas. Em 2013, com a inclusão da Avaliação Nacional da Alfabetização (ANA), o Saeb passou a ser composto por três avaliações externas em larga escala: Aneb, Anresc (Prova Brasil) e ANA.

Em 2019, às vésperas de completar três décadas de realização, passou por uma reestruturação para se adequar à Base Nacional Comum Curricular (BNCC). A BNCC tornou-se referência na formulação dos itens do 2º ano (língua portuguesa e matemática) e do 9º ano do ensino fundamental, no caso dos testes de ciências da natureza e ciências humanas, aplicados de forma amostral. As siglas ANA, Aneb e Anresc deixam de existir e todas as avaliações passam a ser identificadas pelo nome Saeb, acompanhado das etapas, das áreas de conhecimento e dos tipos de instrumentos envolvidos. A avaliação da alfabetização passa a ser realizada no 2º ano do ensino fundamental, primeiramente, de forma amostral. Em 2021 é implementada a avaliação da educação infantil, com aplicação de questionários eletrônicos exclusivamente para professores e diretores – secretários municipais e estaduais também passaram a responder. O Quadro 1 resume as principais mudanças do Saeb, desde sua criação.

QUADRO 1

EVOLUÇÃO DE CARACTERÍSTICAS CENTRAIS DO SAEB

EDIÇÃO	DISCIPLINAS AVALIADAS	POPULAÇÃO-ALVO	ANOS ESCOLARES AVALIADOS	TIPO DE APLICAÇÃO
1990	Língua Portuguesa, Matemática, Ciências Naturais e Redação	Escolas públicas	1ª, 3ª, 5ª e 7ª séries do EF	Amostrал
1993	Língua Portuguesa, Matemática, Ciências Naturais e Redação	Escolas públicas	1ª, 3ª, 5ª e 7ª séries do EF	Amostrал
1995	Língua Portuguesa (leitura) e Matemática	Escolas públicas e privadas	4ª e 8ª séries do EF, 3ª série do EM	Amostrал
1997	Língua Portuguesa, Matemática, Ciências Naturais, Física, Química e Biologia	Escolas públicas e privadas	4ª e 8ª séries do EF, 3ª série do EM	Amostrал
1999	Língua Portuguesa, Matemática, Ciências Naturais, Física, Química, Biologia, História e Geografia	Escolas públicas e privadas	4ª e 8ª séries do EF, 3ª série do EM	Amostrал
2001	Língua Portuguesa e Matemática	Escolas públicas e privadas	4ª e 8ª séries do EF, 3ª série do EM	Amostrал
2003	Língua Portuguesa e Matemática	Escolas públicas e privadas	4ª e 8ª séries do EF, 3ª série do EM	Amostrал
2005	Língua Portuguesa e Matemática	Escolas públicas e privadas	4ª série/5º ano, 8ª série/9º ano do EF, 3ª série do EM	Parte censitária e parte amostral
2007	Língua Portuguesa e Matemática	Escolas públicas e privadas	4ª série/5º ano, 8ª série/9º ano do EF, 3ª série do EM	Parte censitária e parte amostral
2009	Língua Portuguesa e Matemática	Escolas públicas e privadas	4ª série/5º ano, 8ª série/9º ano do EF, 3ª série do EM	Parte censitária e parte amostral
2011	Língua Portuguesa e Matemática	Escolas públicas e privadas	5º ano/4ª série, 9º ano/ 8ª série do EF, 3ª e 4ª séries do EM	Parte censitária e parte amostral
2013	Língua Portuguesa e Matemática	Escolas públicas e privadas	5º ano/4ª série, 9º ano/ 8ª série do EF, 3ª e 4ª séries do EM	Parte censitária e parte amostral
2015	Língua Portuguesa e Matemática	Escolas públicas e privadas	5º e 9º anos do EF, 3ª e 4ª séries do EM	Parte censitária e parte amostral
2017	Língua Portuguesa e Matemática	Escolas públicas e privadas	5º e 9º anos do EF, 3ª e 4ª séries do EM (tradicional e integrado)	Parte censitária e parte amostral
2019	Língua Portuguesa, Matemática, Ciências Humanas e Ciências da Natureza	Escolas públicas e privadas	2º, 5º e 9º anos do EF, 3ª e 4ª séries do EM (tradicional e integrado)	Parte censitária e parte amostral
2021	Língua Portuguesa, Matemática, Ciências Humanas e Ciências da Natureza	Escolas públicas e privadas	2º, 5º e 9º anos do EF, 3ª e 4ª séries do EM (tradicional e integrado) educação infantil	Parte censitária e parte amostral

Fonte: Adaptado de Lustosa (2020, p. 32-33).

O plano amostral do Saeb 2021 tem um desenho parecido com o adotado na sua última edição, limitando-se às escolas privadas para o 5º ano do Ensino Fundamental e para 3ª e 4ª séries do Ensino Médio, e abrangendo as escolas públicas e privadas quando se trata do 2º ano e do 9º ano do Ensino Fundamental. A amostra do Saeb para o ensino médio inclui tanto as escolas que oferecem ensino médio na modalidade tradicional quanto na de ensino médio integrado, a fim de permitir a obtenção de resultados para o ensino médio tradicional e para o ensino médio integrado, exclusivamente, mas, também, para as duas modalidades de ensino conjuntamente. Também inclui as amostras do 9º ano do ensino fundamental para avaliação de Ciências Humanas e Ciências da Natureza e do 2º ano do ensino fundamental para avaliação de Língua Portuguesa e Matemática.

O Relatório de Amostragem do Saeb 2021 está organizado da seguinte forma: o capítulo 1 trata do universo de referência da avaliação; o capítulo 2 explicita a sistemática do plano amostral e da seleção das turmas das escolas da amostra e o capítulo 3 dispõe sobre a metodologia da expansão dos resultados. Para a avaliação da educação infantil, foi desenvolvido documento específico juntamente com seus microdados.

1 DESCRIÇÃO DO UNIVERSO DE REFERÊNCIA DO SAEB 2021

Este capítulo descreve a população (ou universo) de referência do Saeb 2021, indicando de quais unidades dessa população foi selecionada a amostra e quais compuseram a parte censitária.

Primeiramente, é conveniente lembrar a distinção entre população de interesse (ou população alvo) e população de referência (ou população de pesquisa). A primeira é a que se gostaria de cobrir com uma pesquisa ou um estudo e é usualmente declarada de um modo solto. A segunda é aquela em que a pesquisa ou estudo vai efetivamente cobrir e que se baseia num sistema de referência (cadastro) bem documentado e que permite identificar claramente e localizar seus membros. A população-alvo do Saeb 2021 foi estabelecida pela Portaria nº 250, de 05 de julho de 2021, conforme detalhamento do Quadro 2.

QUADRO 2

DETALHAMENTO DA POPULAÇÃO-ALVO DO SAEB 2021

(continua)

Rede de ensino	Mínimo de matrículas	Disciplinas	Ano Escolar	Tipo de aplicação
Escolas públicas localizadas em zonas urbanas e rurais	10	Língua Portuguesa e Matemática	5º e 9º anos do ensino fundamental e 3ª e 4ª séries do ensino médio (tradicional e integrado)	Censitária
Escolas privadas localizadas em zonas urbanas e rurais	10	Língua Portuguesa e Matemática	5º e 9º anos do ensino fundamental e 3ª e 4ª séries do ensino médio (tradicional e integrado)	Amostral
Escolas públicas e privadas localizadas em zonas urbanas e rurais	10	Ciências Humanas e Ciências da Natureza	9º ano do ensino fundamental	Amostral
Escolas públicas e privadas localizadas em zonas urbanas e rurais	10	Língua Portuguesa e Matemática	2º ano do ensino fundamental	Amostral

QUADRO 2

DETALHAMENTO DA POPULAÇÃO-ALVO DO SAEB 2021

(conclusão)

Rede de ensino	Mínimo de matrículas	Disciplinas	Ano Escolar	Tipo de aplicação
Instituições privadas, públicas e conveniadas com o poder público, localizadas em zonas urbanas e rurais	-	-	Creche ou pré-escola da etapa da educação infantil ¹	Amostral

Fonte: Elaboração própria com base na Portaria nº 250/2021.

* O detalhamento sobre a avaliação da educação infantil não será tratado neste relatório.

Para delimitação da população de referência à Portaria nº 250/2021, foram consideradas as seguintes exclusões:

- I. escolas com menos de 10 estudantes matriculados nas etapas avaliadas;
- II. as turmas multisseriadas;
- III. as turmas de correção de fluxo;
- IV. as turmas de educação de jovens e adultos;
- V. as turmas de ensino médio Normal/Magistério;
- VI. as classes, as escolas ou os serviços especializados de educação especial não integrantes do ensino regular; e
- VII. as escolas indígenas que não ministrem a Língua Portuguesa como primeira língua.

1.1 Objeto do Levantamento

O Saeb 2021 é uma avaliação aplicada a alunos das escolas públicas e privadas, dos 2º, 5º ano e 9º anos do ensino fundamental e das 3ª e 4ª séries do ensino médio¹, em todas as Unidades da Federação. Envolve as áreas do conhecimento de Matemática (MT) e Língua Portuguesa (LP) e, para o 9º ano do ensino fundamental, abrangem Ciências Humanas (CH) e Ciências da Natureza (CN). Assim, foram considerados quatro universos ou populações de referência para avaliação amostral do Saeb 2021:

- alunos do 2º ano do ensino fundamental;
- alunos do 5º ano do ensino fundamental;
- alunos do 9º ano do ensino fundamental; e
- alunos da 3ª e 4ª séries do ensino médio.

¹ No caso do ensino médio, como algumas escolas certificam na 4ª série como etapa de conclusão, foram consideradas as 3ª e 4ª séries do ensino médio, indistintamente, abrangendo o ensino médio tradicional e o ensino médio integrado.

A amostra do Saeb 2021 buscou garantir, *a priori*, estimativas confiáveis das médias de proficiências para os seguintes domínios de interesse:

- por rede: estadual, municipal e privada para o 2º ano do ensino fundamental e, para o 9º ano do ensino fundamental, tratando-se das avaliações de CH e CN. Para os demais universos da avaliação, a amostra alcança apenas a rede privada. Para a dependência administrativa federal, a aplicação é censitária, em função do número reduzido de escolas;
- por unidade da federação;
- por localização: urbana, rural;
- por área: capital, interior;
- modalidade de ensino: tradicional, integrado para o ensino médio.

1.2 Sistema de Referência

O sistema de referência adotado é o Censo Escolar de 2021, em sua versão preliminar, realizado pelo MEC/Inep, para identificação das escolas que possuíam alunos matriculados nos anos escolares mencionados. Geralmente o sistema de referência para seleção da amostra para a aplicação do Saeb é o Censo Escolar do ano anterior. No entanto, o cenário de pandemia de covid-19 adiou o cronograma de aplicação do Saeb, permitindo que o Censo Escolar de 2021, em sua versão preliminar, fosse utilizado como referência única para seleção da amostra para esta edição.

1.3 População de Referência da Amostra do Saeb 2021

Para a edição de 2021, utilizou-se o mesmo padrão investigado desde o Saeb 2013. Desse modo, o universo da pesquisa/avaliação é o seguinte:

Todos os alunos de turmas regulares matriculados em 2021 nas escolas declaradas no Censo Escolar deste ano, nas localizações urbanas e rurais em um dos anos escolares de interesse, excluindo os alunos das escolas com menos de 10 alunos por ano escolar e outros critérios específicos.

Para as amostras limitadas às escolas privadas, a avaliação mantém os padrões investigados. Uma vez que o estudo para as escolas públicas foi ampliado e aplicado de forma censitária, é possível obter resultados comparáveis aos dos estudos de anos anteriores, bastando, para isso, limitar a parte da amostra usada para comparação à que fornece cobertura para a população de referência equivalente à utilizada naqueles estudos.

Partindo da população de interesse, são realizadas exclusões em função da diversidade de cenários educacionais observados no contexto das escolas brasileiras, até que possa chegar à realidade da população de referência, mostrada na Tabela 1.

TABELA 1**CONTAGEM DE ALUNOS, TURMAS E ESCOLAS DA POPULAÇÃO DE REFERÊNCIA PARA O DIMENSIONAMENTO DA AMOSTRA DO SAEB 2021 (VERSÃO PRELIMINAR DO CENSO ESCOLAR DE 2021)**

Ano Escolar	Alunos	Turmas	Escolas
2º ano EF	2.565.447	119.509	58.455
5º ano EF	2.816.441	120.065	60.985
9º ano EF	2.767.935	100.109	48.336
3ª e 4ª séries EM	2.370.941	77.717	26.824
Total	10.520.764	417.400	93.575

Fonte: elaboração própria a partir da versão preliminar do Censo Escolar 2021

As exclusões realizadas para delimitação da população de referência são detalhadas no apêndice, permitindo o entendimento do impacto de cada uma das exclusões efetuadas. Para o 2º ano, por exemplo, o universo de interesse em 2021 era composto por 2.770.047 alunos, distribuídos em 161.065 turmas e 96.654 escolas, conforme mostra a linha denominada Universo 1 da Tabela A.2º ano – apêndice.

Quando são excluídos os alunos das turmas multisseriadas ou de correção de fluxo, de escolas indígenas e escolas com menos de 10 alunos, conforme descreve a linha rotulada como Universo 2, no contexto do 2º ano, este contingente passa a ser formado por 2.565.447 alunos, 119.509 turmas regulares e 58.455 escolas. Este procedimento reduziu universo inicial em 7,39% do número de alunos, porcentagem bem menor que a redução do universo de escolas (39,52%).

Verificamos ainda que o número médio de alunos de 2º ano por turmas multisseriadas ou de correção de fluxo, ou indígena, é de 4,7 alunos por turma excluída por esses motivos. Este fato indica a grande dispersão dos alunos em turmas pequenas espalhadas por várias escolas, sendo um argumento favorável para a exclusão.

Por questões operacionais, na aplicação das provas do Saeb, foram excluídos do universo de pesquisa as escolas com menos de 10 alunos no ano escolar em questão, que totalizaram, no 2º ano, 47.201 alunos em 7.846 escolas – um impacto de 1,7% do total de alunos e 8,1% do total de escolas, em relação ao Universo 1. Portanto, a linha denominada Universo 2 indica o universo de referência para o Saeb 2021, considerando os critérios definidos acima.

Assim, no 2º ano do ensino fundamental, o universo original de aluno reduz-se para 92,61% do número inicial, enquanto o número de escolas reduz-se para 60,48%. Este procedimento vai ao encontro do objetivo da pesquisa em diminuir sensivelmente o número de escolas sem excluir muitos alunos.

Apresentamos ainda, naquela tabela, a decomposição dos números deste universo por dependência administrativa, destacando as escolas privadas, estaduais e municipais, que formaram o universo a ser amostrado no Saeb – 2º ano, e as escolas federais, avaliadas de forma censitária.

Explicações semelhantes podem ser construídas a partir da análise das tabelas correspondentes a cada um dos anos escolares avaliados. Observe, entretanto, que as reduções para o 9º ano do ensino fundamental e para as 3ª e 4ª séries do ensino médio não são tão acentuadas como as observadas para os 2º e 5º anos do ensino fundamental.

2 PLANO AMOSTRAL

Este capítulo descreve o plano amostral utilizado no Saeb 2021. Para esta edição não houve mudanças em relação à anterior. Foram mantidas a avaliação do 2º ano do ensino fundamental e as avaliações de Ciências Humanas e Ciências da Natureza para o 9º ano do ensino fundamental, com aplicação amostral, tanto para alunos da rede pública quanto da rede privada e a estratificação da amostra de alunos do ensino médio da rede privada, por modalidade de ensino, a fim de permitir que os resultados sejam calculados para o ensino médio tradicional e integrado, separadamente, bem como em conjunto. Além disso, o plano amostral preservou as avaliações com série histórica mais longa, abrangendo língua portuguesa e matemática, aplicadas para alunos dos 5º e 9º anos do ensino fundamental e 3ª e 4ª séries do ensino médio das escolas da rede privada.

2.1 Domínios de interesse

O principal objetivo do plano amostral considerado é permitir a produção de estimativas com precisão controlada para a média e os respectivos erros de medida em relação à proficiência dos alunos em duas áreas de conhecimento: Língua Portuguesa (LP) e Matemática (MT). Para o 9º ano, planeja-se que as estimativas sejam calculadas com precisão controlada para Ciências Humanas (CH) e Ciências da Natureza (CN). Além das estimativas globais para o conjunto dos alunos de cada uma das subpopulações definidas anteriormente (em cada um dos anos escolares), a avaliação visa a produzir estimativas para alguns domínios de interesse (estratos ou grupos de alunos para os quais se pretende publicar resultados).

O universo do Saeb é dividido por subpopulações, segundo o ano escolar em que o aluno está matriculado, a saber: 2º, 5º e 9º anos do ensino fundamental regular e 3ª e 4ª séries do ensino médio regular nas modalidades de ensino tradicional ou integrada. Sabendo-se que os alunos de escolas públicas com pelo menos 10 alunos matriculados nesses anos escolares participam do Saeb de forma censitária, não há necessidade de selecionar amostras para estas populações. Como exceção, temos as escolas públicas do 2º ano e para a avaliação de CH e CN do 9º ano que participam do Saeb de forma amostral.

Em seguida, o primeiro critério de estratificação utilizado para delimitação dos domínios de interesse é o da dependência administrativa da escola. Como para as avaliações de LP e MT do 5º ano, 9º ano do ensino fundamental e das 3ª e 4ª séries do ensino médio, as escolas públicas (federais, estaduais e municipais) são avaliadas de forma censitária e mantidas em estratos separados das escolas privadas, considerou-se apenas a categoria privada. Já para o 2º ano e para CH e CN do 9º ano foram consideradas as escolas estaduais, municipais e privadas. Sendo assim, foram trabalhadas as seguintes subpopulações:

- 2º ano do ensino fundamental regular pública ou privada (F2ParPub);
- 5º ano do ensino fundamental regular privada (F5Par);
- 9º ano do ensino fundamental regular pública ou privada (F9ParPub);
- 3ª e 4ª séries do ensino médio regular tradicional ou integrado privada (M34TIPar).

Para os alunos do 9º ano do ensino fundamental, a amostra abrange tanto escolas públicas quanto escolas privadas em função das avaliações de CH e CN. No entanto, para LP e MT, a aplicação amostral limitou-se às escolas privadas, enquanto as públicas participam de forma censitária.

Os critérios apresentados a seguir foram aplicados dentro de cada uma das quatro subpopulações. O segundo critério é o da localização geográfica da escola. Em todas as subpopulações definiu-se que seriam

consideradas as Unidades de Federação (UF) para a produção de estimativas com precisão controlada para os 26 estados e para o Distrito Federal, totalizando 27 estratos nesta categoria. O terceiro critério é o da localização urbana ou rural da escola, dentro da Unidade de Federação. O quarto critério é o da área da escola, ou seja, se as escolas estão localizadas na capital ou no interior dentro da UF.

Além desses critérios, para os 2º e 9º anos do ensino fundamental, foi aplicada a estratificação por dependência administrativa estadual, municipal ou privada. As escolas federais são avaliadas de forma censitária em função de sua quantidade reduzida. Para o ensino médio, foi acrescido o critério de estratificação, segundo a modalidade de ensino, podendo ser tradicional ou integrado.

O Quadro 3 apresenta um resumo dos critérios de estratificação que foram considerados inicialmente para a produção de estimativas com precisão controlada.

QUADRO 3

CRITÉRIOS PARA DEFINIR OS DOMÍNIOS DE INTERESSE PARA O SAEB 2021

Critério	Quantidade	Descrição
F2ParPub		Escolas com, pelo menos, 10 alunos no ano escolar
UF	27	26 estados + o DF
Localização	2	Urbana ou Rural
Área	2	Capital ou Interior
Dep. Adm	3	Estadual, Municipal ou Privada
F5Par		Escolas com, pelo menos, 10 alunos no ano escolar
UF	27	26 estados + o DF
Localização	2	Urbana ou Rural
Área	2	Capital ou Interior
F9ParPub		Escolas com, pelo menos, 10 alunos no ano escolar
UF	27	26 estados + o DF
Localização	2	Urbana ou Rural
Área	2	Capital ou Interior
Dep. Adm	3	Estadual, Municipal ou Privada
M34TIPar		Escolas com, pelo menos, 10 alunos no ano escolar
UF	27	26 estados + o DF
Localização	2	Urbana ou Rural
Área	2	Capital ou Interior

Fonte: elaboração própria

Combinando todos estes critérios, temos para este plano amostral um total de 972 possíveis domínios de interesse. Entretanto, como em alguns desses estratos não havia nenhuma escola na população, houve redução do número de estratos. Assim, o universo de referência apresenta 252 estratos no 2º ano do ensino fundamental, 78 no 5º ano do ensino fundamental, 270 no 9º ano do ensino fundamental e 104 nas 3ª e 4ª séries do ensino médio regular, totalizando 704 domínios de interesse considerados para o sorteio da amostra.

É importante destacar que, mesmo que um estrato apresente um número reduzido de escolas, ele foi mantido no plano amostral. Para a definição do plano amostral nos 704 domínios de interesse, não estão contempladas as escolas que fazem parte da aplicação censitária do Saeb.

2.2 Plano Amostral

O planejamento amostral do Saeb 2021 tem como base o plano adotado desde o Saeb 2003. Para manter em um único documento o plano detalhado, com adaptações para essa edição, reproduzimos partes do relatório original da edição de 2003 (Silva; Andrade; Bussab, 2003). A população de alunos está dividida em quatro subpopulações fundamentais, segundo o ano escolar em que estão matriculados. A seleção direta de alunos individualmente não é factível do ponto de vista prático. Sendo assim, foi necessária a adoção de alguma forma de amostragem conglomerada de alunos.

Para seleção dos alunos a serem testados como parte da amostra do Saeb 2021, o plano amostral adota a amostragem conglomerada em duas etapas, tendo as escolas como unidades primárias de amostragem, considerando as características do universo a ser pesquisado na avaliação e a disponibilidade de um cadastro, obtido a partir dos dados da versão preliminar do Censo Escolar de 2021. Na primeira etapa do plano amostral, foram selecionadas escolas que contenham alunos de cada um dos anos escolares considerados, a saber: 2º, 5º e 9º anos do ensino fundamental e 3ª e 4ª séries do ensino médio tradicional ou integrado.

Na segunda etapa, foram selecionadas turmas dentro das escolas selecionadas na primeira etapa, em cada um dos anos escolares. Uma vez selecionada uma turma para participar da avaliação, todos os alunos da turma, presentes no dia da avaliação, seriam submetidos às provas nas áreas de Língua Portuguesa e Matemática.

Vale ressaltar que as unidades primárias de amostragem são as escolas. Por esta razão, uma mesma escola pode participar dos universos (e, portanto, das amostras) de mais de um ano escolar, desde que tenha turmas e alunos de mais de um ano escolar considerado.

Para especificar completamente o plano amostral, foram executadas as seguintes operações:

- Estratificar o universo por ano escolar a ser avaliado em quatro subpopulações distintas:
 - escolas privadas ou públicas oferecendo 2º ano do ensino fundamental (F2ParPub);
 - escolas privadas oferecendo 5º ano do ensino fundamental (F5Par);
 - escolas privadas ou públicas oferecendo 9º ano do ensino fundamental (F9ParPub);
 - escolas privadas oferecendo 3ª e 4ª séries do ensino médio tradicional ou integrado (M34TIPar).
- Estratificar as escolas em cada subpopulação segundo os domínios de interesse. As escolas do cadastro obtido a partir do Censo Escolar 2021 preliminar foram estratificadas nos 704 domínios de interesse.
- Calcular o tamanho da amostra de escolas e turmas para cada um dos domínios de interesse, em cada um dos anos escolares.

O dimensionamento da amostra foi efetuado, refletindo a ideia básica de oferecer precisão igual para as estimativas de proficiência para cada área do conhecimento, em cada domínio de interesse, considerando, também, que seriam avaliadas duas áreas (LP e MT) em cada ano escolar. Decidiu-se seguir o mesmo padrão utilizado nas últimas edições do Saeb, ou seja, calcular o tamanho da amostra de escolas e turmas em cada ano escolar e domínio de interesse, usando a seguinte regra:

Garantir um erro máximo, por UF, de 4 pontos absolutos na escala (250,50) para o 5º ano do ensino fundamental, 9º ano do ensino fundamental e 3ª/4ª séries do ensino médio tradicional ou integrado, e de 4 pontos absolutos na escala (750,50) para o 2º ano do ensino fundamental.

O cálculo da amostra foi baseado no Censo Escolar 2021 preliminar e nos resultados do Saeb 2019. As informações da média e da variância, por UF, foram obtidas do Saeb 2019 e o número de turmas e de alunos/turma obtidos do Censo Escolar 2021 preliminar.

O cálculo do tamanho da amostra considerou a variância e a média das áreas de Língua Portuguesa e Matemática, e, para o 9º ano, Ciências Humanas e Ciências da Natureza, tomando como amostra final aquela que gerou a maior quantidade de turmas. Prevendo possíveis perdas durante a aplicação das provas e supondo que, em decorrência do cenário de pandemia, as perdas poderiam ser maiores que o usual, nos 2º, 5º e 9º anos foram acrescidos 50% no número de turmas, enquanto nas 3ª e 4ª séries de ensino médio, o tamanho da amostra em número de turmas foi dobrado. Além disso, considerando que, para a subpopulação do 9º ano, todos os alunos da amostra seriam avaliados em Língua Portuguesa e Matemática, mas metade dos alunos seriam submetidos às provas de Ciências Humanas e outra metade às de Ciências da Natureza, o tamanho da amostra foi multiplicado por dois, possibilitando a seleção de amostras suficientemente grandes e independentes para estas duas áreas de conhecimento.

Após o cálculo do tamanho da amostra em cada UF, criou-se a variável porte, dividindo as escolas em pequenas e grandes, a fim de garantir escolas com diferentes tamanhos na amostra, de acordo com o seguinte critério: a) Pequenas escolas com uma ou duas turmas no ano escolar avaliado; b) Grandes escolas com três ou mais turmas no ano escolar avaliado.

Em seguida, alocou-se o número de turmas de forma proporcional ao tamanho em cada combinação de dependência administrativa, localização, área, modalidade de ensino, conforme as particularidades da avaliação de cada subpopulação, criando categorias denominadas como domínios de interesse. Ressalta-se que em todo domínio de interesse, com somente uma turma, esta seria selecionada automaticamente na amostra.

Finalmente, calculou-se o número de escolas a selecionar por estrato de tamanho (domínio de interesse + porte da escola), dividindo-se o número de turmas a avaliar pelo número de turmas a selecionar por escola em cada estrato (uma turma por escola nas escolas pequenas, e duas turmas por escola nas grandes). O número de escolas selecionadas por estrato de tamanho foi sempre arredondado para múltiplos de dois. Tal arredondamento se deu para facilitar a etapa de seleção da amostra de escolas, cujo método dependia da seleção de pares de escolas. Devido ao arredondamento e ao fato de que todos os domínios de interesse devem ter, pelo menos, uma escola na amostra, procurou-se fazer alguns ajustes para obter o número de turmas calculado inicialmente. A Tabela 2 apresenta o dimensionamento da amostra para os quatro grandes domínios de interesse.

TABELA 2

ALOCAÇÃO ESPERADA DA AMOSTRA PARA O SAEB 2021

Subpopulação	F2ParPub	F5Par	F9ParPub	M34TIPar	Total
Alunos na população	2.564.282	477.370	2.763.729	286.149	6.091.530
Domínios de interesse	252	78	270	104	704
Alocação da amostra de escolas	902	1.028	8.144	2.129	12.203
Nº de turmas nas escolas sorteadas (censo)	2.298	1.964	18.868	3.267	26.397
Alocação da amostra de turmas	1.270	1.201	10.631	2.362	15.464
Média de alunos por turma das escolas sorteadas	21	21	28	30	27
Alocação da amostra de alunos	26.578	25.338	298.853	71.613	422.383
Fração amostral prevista (porcentagem de alunos)	1,0%	5,3%	10,8%	25,0%	6,9%

Fonte: elaboração própria (reflete plano amostral com base no Censo Escolar 2021 - preliminar)

Antes de efetuar a seleção, as escolas dentro de cada estrato foram ordenadas de acordo com os códigos de CEP correspondentes. Essa ordenação, se utilizada com sorteio sistemático, garantiria um efeito de estratificação implícita por localização geográfica, quando aplicável, e um maior espalhamento da amostra de escolas. Entretanto, para evitar os problemas técnicos decorrentes do uso de amostragem sistemática, decidiu-se por utilizar um método de seleção de escolas denominado de Amostragem Sequencial de Poisson (Ohlsson, 1998).

Este método de amostragem é bastante simples e pode ser descrito, para uma população genérica de tamanho N , de onde se deseja selecionar uma amostra de n unidades com probabilidades proporcionais a uma medida de tamanho t , usando o algoritmo descrito no Anexo B. Para o Saeb, são adotadas probabilidades proporcionais ao número de turmas nas escolas.

Para conseguir o efeito implícito da estratificação geográfica na amostragem sistemática, após a ordenação das escolas em cada estrato, as escolas foram divididas em “zonas de amostragem” dentro das quais foram selecionadas duas escolas por Amostragem Sequencial de Poisson, usando o algoritmo acima indicado. Em cada estrato de tamanho, o número de zonas de amostragem definidas foi sempre igual ao número de escolas desejadas na amostra, dividido por dois.

Este esquema de amostragem preserva o espalhamento da amostra que seria conseguido com amostragem sistemática sem, entretanto, resultar nas mesmas dificuldades técnicas daquele método (necessidade de uso de métodos indiretos para estimação de variância). Ao mesmo tempo, preserva a simplicidade do processo de seleção da amostra.

Outra característica importante desse processo de amostragem é a possibilidade de controlar a rotação da amostra de escolas em edições consecutivas do Saeb, denominada de esquema de rotação parcial controlada. Para isso foi necessária a informação dos números aleatórios permanentes do Saeb anterior, os quais foram gerados por meio de uma distribuição uniforme. Nesta amostragem, criou-se um número aleatório para cada escola que não fazia parte do processo de amostragem de edições anteriores do Saeb, usando o mesmo critério de geração de números aleatórios usado anteriormente.

Para sequência do plano amostral, selecionaram-se turmas dentro das escolas amostradas com duas ou mais turmas no ano escolar, utilizando-se amostragem aleatória simples. No caso das escolas

com apenas uma turma no ano escolar, não é necessária esta etapa de amostragem, pois a única turma da escola é escolhida automaticamente para a avaliação. A seleção aleatória deve garantir distribuição das turmas por turno na amostra aproximadamente proporcional e similar à distribuição na população.

O processo descrito acima se repete de maneira independente para cada subpopulação, de forma que uma escola selecionada para participar da avaliação em um ano escolar não necessariamente participará para todos os anos escolares oferecidos, mas apenas aquelas para as quais foi selecionada.

A Tabela 3 apresenta um resumo dos números de municípios e escolas incluídas nas amostras principais de cada uma das subpopulações, bem como o número de turmas esperado. Observe que a linha de total tem valores nas colunas escolas e municípios que não representam a soma das outras linhas, mas, sim, os números de escolas distintas nas quatro amostras, e o número de municípios com, ao menos, uma escola em uma das amostras dos anos escolares considerados.

TABELA 3

NÚMERO DE ESCOLAS, TURMAS E MUNICÍPIOS SELECIONADOS POR SUBPOPULAÇÃO A PARTIR DA VERSÃO PRELIMINAR DO CENSO ESCOLAR 2021

Subpopulação	Escolas	Turmas Esperadas	Municípios
F2ParPub	902	1.270	415
F5Par	1.028	1.201	418
F9ParPub	8.144	10.631	2.417
M34TIPar	2.123	2.362	664
Total	9.952	15.464	2.543

Fonte: elaboração própria (reflete o plano amostral com base no Censo Escolar 2021)

⁽¹⁾ O total da coluna de escolas e de municípios registram o número total de escolas/municípios distintas(os) efetivamente selecionadas(os) para a amostra do Saeb 2021.

2.3 Seleção das turmas nas escolas selecionadas

Ao final do processo de ajuste da amostra das escolas, foram escolhidas as turmas por meio do método de amostragem aleatória simples. Nas escolas de porte grande, foram selecionadas duas turmas e, nas de porte pequeno, selecionada uma turma, conforme descrito anteriormente. Salienta-se que, para a edição de 2021 do Saeb, tanto a seleção das escolas quanto a seleção das turmas foram realizadas utilizando a versão preliminar do censo escolar 2021 como cadastro.

Os resultados apresentados nas Tabelas 2 e 3, na seção anterior, tiveram algumas alterações devido aos ajustes na seleção das escolas, em função da negativa de algumas escolas em participar da avaliação e as eventuais substituições realizadas. Após a realização dos ajustes, a fração amostral representa 7,2% dos alunos da população. Os resultados atualizados estão na Tabela 4.

TABELA 4

ALOCAÇÃO DA AMOSTRA APÓS O AJUSTE DAS ESCOLAS PARA O SAEB 2021

Subpopulação	F2ParPub	F5Par	F9ParPub	M34TIPar	Total
Alunos na população	2.567.585	480.070	2.748.280	287.514	6.083.449
Domínios de interesse	253	78	270	104	705
Alocação da amostra de escolas	903	1.024	8.110	2.079	9.896
Nº de turmas nas escolas sorteadas (censo)	2.281	1.949	18.873	3.172	26.275
Alocação da amostra de turmas	1.268	1.198	10.706	2.310	15.482
Média de alunos por turma das escolas sorteadas	22,5	22,4	29,1	30,9	28,3
Alocação da amostra de alunos	28.542	26.776	311.447	71.266	438.031
Fração amostral prevista (porcentagem de alunos)	1,1%	5,6%	11,3%	24,8%	7,2%

Fonte: Elaboração própria (reflete o plano amostral atualizado após confirmação das escolas).

Ressalta-se que todas as escolas participantes do Saeb 2021 estão na versão preliminar do Censo Escolar 2021, entretanto, após a consolidação dos dados finais desse censo, a situação das escolas e dos alunos é verificada a fim de garantir que os resultados da avaliação estejam alinhados com a população escolar daquele ano. Essa condição é importante para o cálculo dos pesos dos alunos, ao final da aplicação do Saeb 2021.

3 EXPANSÃO DA AMOSTRA

Este capítulo descreve os procedimentos adotados para ponderação (cálculo dos pesos) da amostra realizada e, também, a expansão utilizada para parte censitária, bem como os procedimentos necessários para estimação das medidas de interesse com base na amostra. A descrição é muito similar em formato e conteúdo aos textos anteriores que detalham os procedimentos de ponderação adotados nas edições mais recentes do Saeb.

Assim, apresenta-se o processo de cálculo dos pesos da parte censitária e da amostra, pesos que serão posteriormente empregados para estimar algumas quantidades de interesse dos usuários do levantamento, bem como os métodos propostos para estimar essas quantidades e respectivas medidas de precisão, de modo a permitir sua correta implementação pelos usuários dos dados do Saeb 2021.

O capítulo está dividido em três seções: a seção 1 apresenta o esquema geral de ponderação; a 2 trata dos métodos adotados para a ponderação da amostra principal do Saeb 2021 e a 3 dos métodos adotados para a ponderação da parte censitária da avaliação.

3.1 Esquema geral de ponderação

Toda amostra extraída de uma população finita requer que seus dados sejam adequadamente ponderados para permitir considerar medidas de interesse referentes ao conjunto da população. Isto se torna mais visível quando se trata de estimar totais populacionais, mas não é menos verdadeiro, mesmo quando se pretende estimar outras quantidades, tais como médias, proporções etc. Quando amostragem probabilística

é empregada para selecionar as unidades pesquisadas, o método usual (natural) de ponderar as observações da amostra consiste em dar a cada unidade um peso que é igual ao inverso da respectiva probabilidade de inclusão na amostra. A ponderação do Saeb 2021 foi realizada de duas formas, pois a ponderação de dados com característica amostral é diferente daqueles que têm característica censitária.

Na amostra selecionada para o Saeb 2021, existem três principais unidades de análise: alunos, turmas e escolas. A principal unidade de análise é o aluno, mas também são coletados dados, por meio dos questionários, de professor e diretor. Por causa da diversidade de unidades, o esquema de ponderação da amostra torna-se mais complexo, exigindo para cada uma o seu próprio sistema de pesos para apuração dos resultados. Isto é consequência não só do número de unidades de análise distintas, mas, também, da decisão de tentar fazer com que as contagens amostrais ponderadas de cada um dos tipos de unidade de análise coincidam com contagens populacionais correspondentes obtidas do Censo Escolar 2021. Essa propriedade imposta ao esquema de ponderação é chamada de calibração da amostra.

Para calcular os pesos que serão associados aos alunos da amostra do Saeb 2021, foram adotados os mesmos princípios das edições anteriores do Saeb, em função do interesse de assegurar que as estimativas amostrais ponderadas provenientes do Saeb fossem comparáveis com contagens populacionais disponíveis, com base no censo escolar do mesmo ano de realização do Saeb.

Seguindo a referência das edições anteriores, decidiu-se considerar os dados já disponíveis do Censo Escolar 2021, após aplicação de exclusões equivalentes às efetuadas na montagem do cadastro de seleção da amostra, baseado na versão preliminar do Censo Escolar de 2021. O cadastro assim obtido é chamado doravante de “cadastro de expansão da amostra”. A escolha dessa alternativa para o cadastro de expansão tem a vantagem de fazer com que os dados ponderados no Saeb 2021 se ajustem aos últimos dados divulgados do Censo Escolar 2021, facilitando a interpretação dos resultados. Por outro lado, tem a desvantagem de refletir mudanças no universo de escolas, turmas e alunos que a amostra do Saeb 2021 não é capaz de captar com precisão. Por exemplo, nos totais desse cadastro estão incluídas escolas e turmas novas, criadas durante o período de retificação do censo escolar, que não poderiam participar da seleção da amostra nem tampouco do levantamento de dados do Saeb 2021. Dessa forma, alguns estratos ficam descobertos pela amostra e o total populacional estimado se aproxima dos totais do censo escolar, mas não coincide totalmente. Portanto, vale reforçar que a amostra prevista é baseada na versão preliminar do Censo Escolar 2021 e não captura a mudança de alunos entre escolas.

Foram utilizadas contagens em nível do estrato de expansão para efetuar a calibração dos pesos – por estrato de expansão se entende o estrato de seleção construído com as regras definidas na seleção da amostra do Saeb 2021, usando as informações obtidas do cadastro de expansão. Esta opção permite que todas as contagens para níveis que representem agregação de estratos de expansão sejam calibradas automaticamente na amostra. A desvantagem dessa opção, quando a comparamos com calibração em níveis mais agregados, é que as amostras dentro de cada estrato são menores, aumentando o risco de problemas localizados, ao efetuar a calibração. Esses mesmos riscos estavam embutidos nas calibrações anteriores.

Por último, na base de dados havia escolas que não declararam turma regular no Censo Escolar 2021 ou mencionaram menos de 10 alunos em turmas regulares no ano escolar. Por não fazer parte do universo de expansão, o qual condiciona que a escola tenha 10 ou mais alunos em turmas regulares no ano escolar de estudo, essas escolas foram excluídas do cálculo do peso. Os alunos que participaram da aplicação e pertencem a escolas com menos de 10 alunos foram mantidos na base e, a eles, foi atribuído peso nulo.

O processo de expansão do universo se inicia com o cálculo de um peso que é igual ao inverso da respectiva probabilidade de inclusão na amostra. Para os alunos incluídos na aplicação censitária do Saeb 2021, considera-se probabilidade de inclusão igual a 1. Em seguida, esse peso inicial passa pelo processo de

calibração da amostra, a fim de completar as contagens populacionais correspondentes obtidas do Censo Escolar 2021, cobrindo as perdas naturais do processo de amostragem e coleta dos dados.

A calibração da amostra se dá por meio de transformações lineares que ajustam o peso inicial de forma que sua soma totaliza o número de alunos declarados no Censo Escolar 2007 e ocorre de forma gradual, iniciando pelas turmas, passando para as escolas e os domínios de interesse, aumentando, aos poucos, sua abrangência até que se chegue ao total nacional.

Para os universos da amostra do Saeb 2021 (escolas privadas de 5º ano do ensino fundamental, 3ª e 4ª séries do ensino médio e escolas públicas e privadas de 2º ano e 9º ano do ensino fundamental), a amostragem tem como estrato a Unidade da Federação, a rede de ensino (estadual, municipal ou privada), a localização (urbana e rural) e a área (capital ou interior). Para o ensino médio, considera-se também a modalidade de ensino (tradicional ou integrada).

A expansão do universo submetido à aplicação censitária do Saeb 2021 foi realizada usando como estrato município, dependência administrativa (federal, municipal e estadual) e localização (urbana e rural). Nas situações em que estratos não tinham alunos com prova válida, a expansão foi realizada por município e dependência administrativa, e, posteriormente, foi expandida apenas para o município. Se mesmo assim, a expansão dos estratos não tivesse sido contemplada, fez-se a expansão para a UF.

Os universos de expansão do Censo Escolar 2021, que esse trabalho tomou como referência, estão apresentados nas Tabelas de 5 a 9.

TABELA 5

DADOS DO CENSO ESCOLAR 2021 DO 2º ANO PARA EXPANSÃO

2º Ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.568.789	92,59	119.936	74,18	58.675	60,45
Universo 2A: Escolas privadas	473.696	18,44	26.992	22,51	16.012	27,29
Universo 2B: Escolas públicas federais	1.204	0,05	58	0,05	22	0,04
Universo 2C: Escolas públicas estaduais	321.229	12,51	13.707	11,43	6.386	10,88
Universo 2D: Escolas públicas municipais	1.772.660	69,01	79.179	66,02	36.255	61,79

Fonte: Elaboração própria (contagem do Censo Escolar 2021)

TABELA 6

DADOS DO CENSO ESCOLAR 2021 DO 5º ANO PARA EXPANSÃO

5º ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.821.684	93,66	120.489	76,52	61.186	64,57
Universo 2A: Escolas privadas	480.070	17,01	23.929	19,86	15.056	24,61
Universo 2B: Escolas públicas federais	1.635	0,06	66	0,05	24	0,04
Universo 2C: Escolas públicas estaduais	388.635	13,77	16.056	13,33	7.372	12,05
Universo 2D: Escolas públicas municipais	1.951.344	69,16	80.438	66,76	38.734	63,31

Fonte: Elaboração própria (contagem do Censo Escolar 2021)

TABELA 7

DADOS DO CENSO ESCOLAR 2021 DO 9º ANO PARA EXPANSÃO

9º ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.752.550	97,74	100.398	90,86	48.463	84,88
Universo 2A: Escolas privadas	408.181	14,83	16.379	16,31	11.066	22,83
Universo 2B: Escolas públicas federais	4.270	0,16	155	0,15	37	0,08
Universo 2C: Escolas públicas estaduais	1.208.316	43,90	41.378	41,21	16.543	34,14
Universo 2D: Escolas públicas municipais	1.131.783	41,12	42.486	42,32	20.817	42,95

Fonte: Elaboração própria (contagem do Censo Escolar 2021)

TABELA 8

DADOS DO CENSO ESCOLAR 2021 DAS 3ª E 4ª SÉRIES DO ENSINO MÉDIO TRADICIONAL PARA EXPANSÃO

3ª e 4ª séries	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.141.071	99,62	70.833	98,26	24.915	95,33
Universo 2A: Escolas privadas	281.365	13,14	9.654	13,63	6.887	27,64
Universo 2B: Escolas públicas federais	5.667	0,26	199	0,28	48	0,19
Universo 2C: Escolas públicas estaduais	1.844.433	86,15	60.645	85,62	17.853	71,66
Universo 2D: Escolas públicas municipais	9.606	0,45	335	0,47	127	0,51

Fonte: Elaboração própria (contagem do Censo Escolar 2021)

TABELA 9

DADOS DO CENSO ESCOLAR 2021 DAS 3ª E 4ª SÉRIES DO ENSINO MÉDIO INTEGRADO PARA EXPANSÃO

3ª e 4ª séries	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	193.123	99,83	6.967	99,19	1.932	97,48
Universo 2A: Escolas privadas	6.149	3,18	286	4,11	147	7,61
Universo 2B: Escolas públicas federais	70.634	36,57	2.483	35,64	502	25,98
Universo 2C: Escolas públicas estaduais	113.356	58,70	4.095	58,78	1.264	65,42
Universo 2D: Escolas públicas municipais	2.984	1,55	103	1,48	19	0,98

Fonte: Elaboração própria (contagem do Censo Escolar 2021)

3.2 Descrição da ponderação da amostra

O cálculo da ponderação da amostra do Saeb 2021 foi feito, respeitando-se a estrutura do plano amostral usado, seguindo a mesma ordem da hierarquia utilizada para a seleção da amostra. Primeiramente, as escolas selecionadas foram classificadas em seus estratos de origem, sendo que tais estratos foram definidos em função da Unidade da Federação, rede de ensino (pública e privada), localização (urbana e rural), área (capital e interior) e modalidade de ensino (tradicional e integrado) – no caso do ensino médio – a que pertencem. Para maiores detalhes, observar a descrição da estratificação feita no Capítulo 2.

Toda a ponderação de escolas, turmas e alunos em cada estrato foi efetuada independentemente de que ocorreu nos demais estratos. As contagens populacionais para calibração dos pesos em cada estrato foram obtidas do cadastro de expansão derivado do Censo Escolar 2021. Assim, a descrição a seguir é do processo de cálculo dos pesos usados para um estrato de expansão qualquer. Omite-se a identificação do estrato para simplificar a notação.

Para um estrato de seleção qualquer da amostra do Saeb 2021, representa-se por:

- m , o número de escolas na amostra;
- M , o número de escolas no mesmo estrato segundo o cadastro de seleção; e
- M' , o número de escolas no mesmo estrato segundo o cadastro de expansão.

Seja d_i uma variável que indica se a escola i teve alunos testados (portanto, participou do levantamento ou não), isto é:

$$d_i = \begin{cases} 1 & \text{se a escola } i \text{ teve alunos testados} \\ 0 & \text{se a escola } i \text{ não teve alunos testados} \end{cases} \quad \text{para } i = 1, 2, \dots, m.$$

Peso da Escola

Representa-se por T_i o tamanho (número de turmas) da escola i . Conforme o processo de seleção adotado, a escola i tem probabilidade de seleção no estrato igual a $m \times p_i$,

Sendo:

m para o Saeb igual a dois, pois para cada estrato são selecionadas duas escolas; e

p_i é a probabilidade de seleção da escola i no estrato de k escolas:

$$p_i = \frac{T_i}{\sum_k^M T_k}, \text{ para todo } i = 1, 2, \dots, M$$

Os valores p_i satisfazem a relação: $\sum_{i=1}^M p_i = 1$. Consequentemente, o peso inicial (básico do plano amostral) da escola i da amostra pode ser obtido, calculando-se o inverso da respectiva probabilidade de seleção, que, nesse caso, é dado por

$$w_i = \frac{1}{m \times p_i}, \text{ para } i = 1, \dots, m$$

Observa-se que os pesos iniciais são calculados e atribuídos para todas as escolas da amostra selecionada, mesmo aquelas sem preenchimento do respectivo questionário ou que terminaram sem alunos testados. Tais pesos podem ser utilizados para estimar o número de escolas do cadastro (de seleção ou de

expansão). Ao estimar para o cadastro, devem ser excluídas as escolas que não tiveram alunos testados (escolas com $d_i = 0$) e que, posteriormente, terão seus respectivos pesos finais zerados, uma vez que não deverão contribuir na expansão ou análise dos dados.

Assim, estima-se o número de escolas no estrato usando

$$\hat{M} = \sum_{i=1}^m w_i \times d_i .$$

Essa estimativa não igualará, em geral, o total conhecido de escolas do cadastro de expansão, M' . Para que as estimativas amostrais do total de escolas possam reproduzir exatamente esse total, é suficiente ajustar ou calibrar os pesos das escolas “ativas” da amostra (aquelas com $d_i = 1$) mediante sua multiplicação por um fator de ajuste igual a $\frac{M'}{\hat{M}}$. Assim, o peso ajustado ou calibrado da escola i foi obtido como

$$c_i = d_i \times w_i \times \frac{M'}{\hat{M}}, \text{ para } i = 1, \dots, m.$$

Ao usar os pesos c_i para estimar o total de escolas do cadastro, será reproduzido exatamente o total conhecido M' – esta propriedade é denominada calibração da amostra.

Peso da Turma

Faz-se o cálculo dos pesos das turmas da amostra de maneira semelhante. Primeiramente, são definidos os pesos iniciais das turmas selecionadas. O peso inicial da turma j da escola i da amostra é obtido, multiplicando-se o peso calibrado da escola i pelo inverso da probabilidade de seleção/inclusão da turma dentro da escola, isto é,

$$w_{ij} = \frac{T_i^*}{n_i} \times c_i$$

Em que:

T_i^* é o número total de turmas da escola i no ano escolar de interesse, obtido na operação de verificação cadastral das escolas da amostra;

n_i é o número de turmas selecionadas na escola i ; e

w_{ij} é o peso inicial da turma j da escola i , para $j = 1, \dots, n_i$ e $i = 1, \dots, m$.

Note que, se $c_i = 0$, então todas as turmas da escola i terão seus pesos automaticamente iguais a zero, pois se não houve alunos testados na escola, não pode haver alunos testados nas turmas dessa escola. Note também que, conforme o plano amostral adotado, n_i pode tomar apenas dois valores: 1 se a escola tiver até duas turmas ($T_i^* \leq 2$) e 2 se a escola tiver três ou mais turmas ($T_i^* > 2$). Para a amostra de Ciências Humanas e Ciências da Natureza, no 9º ano do ensino fundamental, o valor de w_{ij} foi multiplicado por $\frac{1}{2}$, considerando que metade das turmas foi selecionada para a amostra de cada uma dessas áreas de conhecimento.

Como ocorrem com os pesos simples de escolas, os pesos simples de turmas podem ser usados para estimar o número de turmas do cadastro de expansão:

$$\hat{T} = \sum_i^m \sum_j^n w_{ij} \times d_{ij}$$

onde d_{ij} é a variável indicadora de haver alunos testados na turma j da escola i da amostra, para $j = 1, \dots, n_j$ e $i = 1, \dots, m$.

Novamente, o ajuste dos pesos de turmas é obtido simplesmente multiplicando os pesos iniciais das turmas “ativas” pela razão entre o número total de turmas obtido do cadastro de expansão, denotado T' , e o número estimado com os pesos iniciais. Assim, os pesos calibrados de turmas ficam iguais a

$$c_{ij} = d_{ij} \times \frac{T'}{\hat{T}} \times w_{ij}, \text{ para } j = 1, \dots, n_j \text{ e } i = 1, \dots, m.$$

Cálculo do Peso dos Alunos

Finalmente, chegamos à descrição dos pesos de alunos. Como todos os alunos de cada turma selecionada são avaliados nas duas disciplinas, não há amostragem dentro de turma e cada aluno participa da avaliação com peso inicial igual a 1 na sua turma, fazendo com que seu peso inicial seja igual ao peso c_{ij} de sua turma. Essa seria a regra caso não ocorressem perdas de alunos dentro das turmas, mas tais perdas são observadas. Portanto, é preciso que tais pesos incorporem uma correção inicial devida à não resposta ou não participação de alunos. Isto pode ser feito usando os números efetivos de alunos testados em cada turma.

O peso inicial do aluno k , da turma j da escola i da amostra é calculado como

$$w_{ijk} = c_{ij} \times d_{ijk} \times \frac{r_{ij}}{a_{ij}}$$

Em que:

d_{ijk} é a variável indicadora de que o aluno k , da turma j , da escola i da amostra foi testado;

r_{ij} é o número total de alunos na turma j da escola i ; e

a_{ij} é o número de alunos testados, para $k = 1, \dots, n_{ij}$, $j = 1, \dots, n_j$ e $i = 1, \dots, m$. Alunos não testados recebem imediatamente peso igual a zero.

Tais pesos não garantem a calibração para o número de alunos do cadastro de expansão, pois alguns estratos do Censo Escolar 2021 não estavam contemplados na versão preliminar do Censo Escolar 2021, conforme levantado anteriormente. Os pesos calibrados de alunos são dados por

$$C_{ijk} = \frac{A'}{\hat{A}} w_{ijk}, \text{ para } k = 1, \dots, n_{ij}, j = 1, \dots, n_j \text{ e } i = 1, \dots, m$$

Em que A' é o número total de alunos da unidade de federação, rede e área do cadastro de expansão, e $\hat{A} = \sum_i^m \sum_j^{n_j} \sum_k^{n_{ij}} w_{ijk}$ é a estimativa desse número obtida da amostra, considerando os pesos iniciais de alunos testados, segundo a unidade de federação, rede e área.

3.3 Utilização dos pesos na estimação

Os pesos, definidos e calculados, conforme a descrição apresentada na seção 2, devem ser utilizados para estimar todas as medidas descritivas de interesse, e ser considerados quando forem ajustados modelos com os dados da amostra pesquisada. Seu uso garante a compensação dos efeitos da amostragem com taxas

diferenciadas nos diversos estratos da pesquisa, correção parcial dos efeitos da não resposta e a calibração para os totais populacionais conhecidos obtidos do cadastro de expansão. Ignorar tais pesos nas análises pode provocar vício nas estimativas e deve ser evitado.

Para estimar totais, a ideia é adotar estimadores “lineares” do tipo

$$\hat{Y} = \sum_h \sum_i \sum_j \sum_k c_{hijk} \times y_{hijk}$$

Em que:

c_{hijk} é o peso calibrado associado ao estrato h , escola i , turma j e aluno k ,

y_{hijk} é o valor de uma variável de interesse y associado ao estrato h , escola i , turma j e aluno k .

Esta fórmula é válida tanto para estimativas de totais baseadas na amostra de escolas, turmas ou alunos, bastando adotar o conjunto de pesos e unidades apropriados para cada caso.

Para calcular médias, a ideia é adotar estimadores do tipo razão:

$$\bar{Y}_c = \frac{\sum_h \sum_i \sum_j \sum_k c_{hijk} \times y_{hijk}}{\sum_h \sum_i \sum_j \sum_k c_{hijk}}$$

Proporções podem ser acomodadas nesta última expressão, definindo-se variáveis indicadoras que tomam valor 1, quando a unidade possui o atributo ou característica de interesse (por exemplo, alunos do sexo masculino) e zero, em caso contrário, usando-se o estimador de médias acima. Contagens também podem ser acomodadas da mesma forma, utilizando o estimador de total. Estes estimadores de totais e médias podem ser facilmente calculados com o R (ex. pacote *Survey*), o Stata (comandos *svy*), o SAS (*Proc Survey*), o SPSS (*Complex Sample*), entre outros.

Estimação de outras medidas, tais como quantis de distribuições e coeficientes em modelos de regressão deve ser feita considerando os pesos, mas descrever estimadores para tais quantidades “complexas” está fora do escopo deste relatório. Para detalhes, sugere-se consultar Pessoa e Silva (1998), Skinner, Holt e Smith (1989), Shah et al. (1995), Chambers e Skinner (2003), Heeringa, West e Berglund (2017), Lehtonen e Pahkinen (2004) e Lumley (2004).

3.4 Estimação da precisão das estimativas

Uma das etapas fundamentais em qualquer pesquisa amostral é a estimação da precisão das estimativas. Esta atividade é fundamental para permitir conhecer e avaliar a qualidade das estimativas produzidas. Na amostra do Saeb 2021, recomenda-se adotar estimadores de variâncias e erros-padrões que possam ser obtidos usando pacotes computacionais. Tais estimadores oferecem apenas uma aproximação dos valores que seriam obtidos se fosse considerado exatamente o plano amostral adotado e o processo de calibração dos pesos. Tal aproximação é recomendada em virtude da sua simplicidade de implementação e de permitir evitar a programação sob medida, sendo as estimativas calculáveis através do R no pacote *survey*, do Stata nos comandos *svy*, do SAS no *Proc Survey*, do SPSS no módulo *Complex Sample*, entre outros.

Para estimar variâncias dos estimadores de total, recomenda-se adotar o estimador:

$$\hat{V}(\hat{Y}) = \sum_h \frac{m_h}{m_h - 1} \times \sum_i^{m_h} (\hat{Y}_{hi} - \bar{Y}_h)^2$$

Em que

$$\hat{Y}_{hi} = \sum_j \sum_k c_{hijk} \times y_{hijk}$$

\hat{Y}_{hi} é a parcela da estimativa do total da variável y correspondente às Unidades Primárias de Amostragem (UPA) i da amostra no estrato h , m_h é o número de unidades primárias de amostragem participantes no estrato h ; e

$$\bar{\hat{Y}}_h = \frac{\sum_i^{m_h} \hat{Y}_{hi}}{m_h}$$

é a média dos \hat{Y}_{hi} no estrato h .

Na maioria dos estratos, as UPAs foram as escolas; em uns poucos estratos, as escolas foram todas incluídas na amostra, e as Unidades Primárias de Amostragem nesses casos foram as turmas. Os erros-padrões podem ser obtidos, calculando-se a raiz quadrada das estimativas de variância obtidas por esse método. Esta forma simplificada e aproximada de estimação de variâncias consiste em imaginar que o sorteio de unidades primárias de amostragem teria sido feito por amostragem com probabilidades proporcionais ao tamanho com reposição, e é comumente denominada como “Método do Conglomerado Primário” – ver Pessoa e Silva (1998).

Para estimar variância de médias, recomenda-se empregar a expressão:

$$\hat{V}(\bar{y}_c) = \sum_h \frac{m_h}{m_h - 1} \times \sum_i^{m_h} (e_{hi} - \bar{e}_h)^2$$

Em que

$$e_{hi} = \frac{1}{\sum_h \sum_i \sum_j \sum_k c_{hijk}} \left[\sum_j \sum_k c_{hijk} (y_{hijk} - \bar{y}_c) \right]$$

e

$$\bar{y}_c = \frac{\sum_h \sum_i \sum_j \sum_k c_{hijk} \times y_{hijk}}{\sum_h \sum_i \sum_j \sum_k c_{hijk}}$$

3.5 Descrição da ponderação da parte censitária

A aplicação censitária poderia dispensar o uso da ponderação dos resultados, uma vez que todos os alunos teriam peso igual a 1. No entanto, em função das perdas observadas no processo de coleta dos dados, e assumindo-se que a perda é aleatória, o processo de ponderação é recomendado para compensação das taxas de não resposta.

O peso dos **alunos** avaliados de forma censitária foi expandido para o estrato município, dependência administrativa (federal, estadual e municipal) e localização (urbana e rural). No caso de ter havido perda total de algum estrato, ou seja, nenhuma escola de um determinado estrato participou da aplicação censitária do Saeb 2021, foi feita a expansão para o município e, posteriormente, para a Unidade da Federação. Dessa forma, a soma dos pesos totaliza a quantidade de alunos do Censo Escolar 2021. Anteriormente à expansão para o

estrato, foi feita a expansão dos alunos para as turmas e para as escolas. O esquema de expansão adotado segue abaixo:

$$Peso_{turma} = \frac{Alunos_{(turma|censo)}}{Alunos_{(turma|validos)}}$$

Em que:

$Alunos_{(turma|censo)}$ é o número total de alunos na turma declarado no Censo Escolar 2021;

$Alunos_{(turma|validos)}$ é o número total de alunos da turma que realizaram a prova; e

$Peso_{turma}$ é o peso do aluno expandido para a turma.

Após o cálculo do peso do aluno expandido para turma, calcula-se o peso do aluno expandido para escola, dado por:

$$Peso_{aux} = \frac{Alunos_{(escola|censo)}}{\sum Peso_{turma}(escola)}$$

$$Peso_{escola} = Peso_{turma} * Peso_{aux}$$

Em que:

$Alunos_{(escola|censo)}$ é o número total de alunos na escola declarado no Censo Escolar 2021;

$\sum Peso_{turma}(escola)$ é a soma dos pesos dos alunos de todas as turmas na escola, ou seja, representa o total de alunos da escola que tiveram nota calculada; e

$Peso_{escola}$ é o peso do aluno expandido para a escola.

Finalmente o peso dos alunos expandido para o estrato é dado por

$$Peso_{aux2} = \frac{Alunos_{(estrato|censo)}}{\sum Peso_{escola}(estrato)}$$

$$Peso_{estrato} = Peso_{escola} * Peso_{aux2}$$

Em que:

$Alunos_{(estrato|censo)}$ é o número total de alunos no estrato declarado no Censo Escolar 2021;

$\sum Peso_{escola}(estrato)$ é a soma dos pesos de todos os alunos das escolas no estrato, representando o total de alunos do estrato que tiveram nota calculada.

Como a soma dos pesos não soma a quantidade de alunos do Censo Escolar 2021, fez-se a expansão por município e dependência administrativa, por município e, por fim, expandiu-se para a Unidade da Federação, caso algum município tenha ficado de fora da aplicação. A seguir são apresentadas as fórmulas da continuação do processo de expansão.

Expansão para o município e dependência administrativa:

$$Peso_{aux3} = \frac{Alunos_{(Mun_Dep|censo)}}{\sum Peso_{estrato}(Mun_Dep)}$$

$$Peso_{Mun_Dep} = Peso_{estrato} * Peso_{aux3}$$

Expansão para o município:

$$Peso_{aux4} = \frac{Alunos_{(Mun|censo)}}{\sum Peso_{Mun_Dep}(Mun)}$$
$$Peso_{Mun} = Peso_{Mun_Dep} * Peso_{aux4}$$

Expansão para a Unidade da Federação (UF):

$$Peso_{aux5} = \frac{Alunos_{(UF|censo)}}{\sum Peso_{Mun}(UF)}$$
$$Peso_{Aluno} = Peso_{Mun} * Peso_{aux5}$$

Ao final dos passos descritos acima, o aluno receberá um peso dado por *Peso_Aluno*. Salienta-se que alguns alunos ficaram com peso menor do que 1. Isso ocorre principalmente entre alunos da aplicação censitária, pois a expansão é por turma. Por exemplo, uma escola declarou que tem três turmas, sendo que todas fizeram a prova, só que uma delas tem apenas um aluno no Censo Escolar 2021 e três alunos nessa turma fizeram a prova. Com isso, o peso dos alunos dessa turma será 1/3.

BIBLIOGRAFIA

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Portaria nº 250, de 5 de julho de 2021. Estabelece as diretrizes de realização do Sistema de Avaliação da Educação Básica (Saeb) no ano de 2021. *Diário Oficial da União*, Brasília, DF, 6 jul. 2021. Seção 1, p. 39.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Microdados do Censo Escolar da Educação Básica 2021*. Brasília, DF, 2023.

CHAMBERS, R. L. SKINNER, C. J. *Analysis of survey data*. West Sussex: John Wiley & Sons, 2003.

HEERINGA, S. G.; WEST, B. T.; BERGLUND, P. A. *Applied survey data analysis*. 2. ed. Boca Raton: CRC Press, 2017.

LEHTONEN, R.; PAHKINE, E. *Practical methods for design and analysis of complex sample*. 2. ed. West Sussex: John Wiley & Sons, 2004.

LUMLEY, T. Analysis of complex survey samples. *Journal of Statistical Software*, [S.l.], v. 9, n. 6, Apr. 2004. Available in: <<https://www.jstatsoft.org/article/view/v009i08>>. Access in: 11 Dec. 2019.

LUSTOSA, L. A. *Um estudo sobre o plano amostral do Saeb*. 164 f. 2020. Dissertação (Mestrado em Gestão e Avaliação da Educação Pública) – Universidade Federal de Juiz de Fora, Juiz de Fora, 2020.

OHLSSON, E. Sequential Poisson Sampling. *Journal of Official Statistics*, v. 14, n. 2, p. 149-162, 1998.

PESSOA, D. G. C.; SILVA, P. L. N. *Análise de dados amostrais complexos*. São Paulo: Associação Brasileira de Estatística, 1998.

SHAH, B. V. et al. *Statistical methods and mathematical algorithms used in SUDAAN*. Research Triangle Park, NC: Research Triangle Institute. 1995.

SKINNER, C.J., HOLT, D. E SMITH, T.M.F. *Analysis of Complex Surveys*. Chichester: John Wiley & Sons, 1989.

APÊNDICE

Memória da Construção dos Universos de Escolas do Saeb 2021

Universos do 2º ano do ensino fundamental

Os universos do 2º ano foram construídos por meio da seleção das escolas com matrículas de alunos no 2º ano do ensino fundamental, diurno e noturno, excluindo turmas de atendimento educacional especializado exclusivo ou de atividade complementar e turmas de educação de jovens e adultos.

Universos do 5º ano do ensino fundamental

Os universos do 5º ano foram construídos por meio da seleção das escolas com matrículas de alunos no 5º ano do ensino fundamental, diurno e noturno, excluindo turmas de atendimento educacional especializado exclusivo ou de atividade complementar e turmas de educação de jovens e adultos.

Universos do 9º ano do ensino fundamental

Os universos do 9º ano foram construídos por meio da seleção das escolas com matrículas de alunos no 9º ano do ensino fundamental, diurno e noturno, excluindo turmas de atendimento educacional especializado exclusivo ou de atividade complementar e turmas de educação de jovens e adultos.

Universos das 3ª e 4ª séries do ensino médio (tradicional e integrado)

Os universos das 3ª e 4ª séries do ensino médio foram construídos por meio da seleção das escolas com matrículas de alunos na 3ª série e 4ª série do ensino médio tradicional, bem como do ensino médio integrado, diurno e noturno, excluindo turmas de atendimento educacional especializado exclusivo ou de atividade complementar, turmas de educação de jovens e adultos e turmas de ensino médio Normal/Magistério.

TABELA A

2º ANO

2º ano do ensino fundamental	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 1	2.770.047	100,00	161.065	100,00	96.654	100,00
Escolas exclusivamente multisseriadas/Correção de fluxo	152.460	5,50	32.547	20,21	30.313	31,36
Universo 1A	2.617.587	94,50	128.518	79,79	66.341	68,64
Escolas indígenas	559	0,02	60	0,04	40	0,04
Universo 1B	2.617.028	94,48	128.458	79,76	66.301	68,60
Alunos em turma multisseriada/Correção de fluxo	4.380	0,16	1.023	0,64	748	0,77
Universo 1C	2.612.648	94,32	127.435	79,12	66.301	68,60
Escolas com menos de 10 alunos	47.201	1,70	7.926	4,92	7.846	8,12
Universo 2	2.565.447	92,61	119.509	74,20	58.455	60,48
Universo 2A: Escolas privadas (universo amostral)	471.091	18,36	26.796	22,42	15.898	27,20
Universo 2B: Escolas públicas federais	1.165	0,05	56	0,05	21	0,04
Universo 2C: Escolas públicas estaduais (universo amostral)	321.867	12,55	13.654	11,43	6.365	10,89
Universo 2D: Escolas públicas municipais (universo amostral)	1.771.324	69,05	79.003	66,11	36.171	61,88

Universo 1 - Alunos de 2º ano do ensino fundamental

Universo 1A - Universo 1 excluindo escolas que têm somente turmas multisseriadas/de correção de fluxo, sem declarar turmas regulares de 2º ano do EF

Universo 1B - Universo 1A excluindo alunos de escolas indígenas que não ministram a Língua Portuguesa como primeira língua

Universo 1C - Universo 1B excluindo alunos de turmas multisseriadas/de correção de fluxo em escolas que declararam turmas regulares do 2º ano do EF

Universo 2 - Alunos de 2º ano do EF de Ensino Regular em escolas com 10 ou mais alunos matriculados

Para o 2º ano do EF:

TP_ETAPA_ENSINO = 15 (ensino fundamental de 9 anos - 2º Ano)

Para as exclusões:

TP_UNIFICADA = 2, 3 e 4 identificam turmas de correção de fluxo ou multisseriadas.

TP_LINGUA_INDIGENA = 1 identifica escolas que ministrem o ensino somente em Língua Indígena

Fonte: Elaboração própria (contagem da versão preliminar do Censo escolar 2021)

TABELA A

5º ANO

5º ano do ensino fundamental	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 1	3.006.269	100,00	156.833	100,00	94.388	100,00
Escolas exclusivamente multisseriadas/Correção de fluxo	133.176	4,43	27.576	17,58	25.826	27,36
Universo 1A	2.873.093	95,57	129.257	82,42	68.562	72,64
Escolas indígenas	673	0,02	67	0,04	44	0,05
Universo 1B	2.872.420	95,55	129.190	82,37	68.518	72,59
Alunos em turma multisseriada/Correção de fluxo	10.484	0,35	1.536	0,98	1.139	1,21
Universo 1C	2.861.936	95,20	127.654	81,39	68.518	72,59
Escolas com menos de 10 alunos	45.495	1,51	7.589	4,84	7.533	7,98
Universo 2	2.816.441	93,69	120.065	76,56	60.985	64,61
Universo 2A: Escolas privadas (universo amostral)	477.370	16,95	23.759	19,79	14.948	24,51
Universo 2B: Escolas públicas federais	1.591	0,06	64	0,05	23	0,04
Universo 2C: Escolas públicas estaduais	389.130	13,82	16.002	13,33	7.359	12,07
Universo 2D: Escolas públicas municipais	1.948.350	69,18	80.240	66,83	38.655	63,38

Universo 1 - Alunos de 5º ano do ensino fundamental

Universo 1A - Universo 1 excluindo escolas que têm somente turmas multisseriadas/de correção de fluxo, sem declarar turmas regulares de 5º ano do EF

Universo 1B - Universo 1A excluindo alunos de escolas indígenas que não ministram a Língua Portuguesa como primeira língua

Universo 1C - Universo 1B excluindo alunos de turmas multisseriadas/de correção de fluxo em escolas que declararam turmas regulares do 5º ano do EF

Universo 2 - Alunos de 5º ano do EF de Ensino Regular em escolas com 10 ou mais alunos matriculados

Para o 5º ano do EF:

TP_ETAPA_ENSINO = 18 (ensino fundamental de 9 anos - 5º ano)

Para as exclusões:

TP_UNIFICADA = 2, 3 e 4 identificam turmas de correção de fluxo ou multisseriadas.

TP_LINGUA_INDIGENA = 1 identifica escolas que ministrem o ensino somente em Língua Indígena

Fonte: Elaboração própria (contagem da versão preliminar do Censo escolar 2021)

TABELA A

9º ANO

9º ano do ensino fundamental	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 1	2.831.299	100,00	110.135	100,00	56.915	100,00
Escolas exclusivamente multisseriadas/Correção de fluxo	28.255	1,00	5.348	4,86	4.767	8,38
Universo 1A	2.803.044	99,00	104.787	95,14	52.148	91,62
Escolas indígenas	928	0,03	64	0,06	46	0,08
Universo 1B	2.802.116	98,97	104.723	95,09	52.102	91,54
Alunos em turma multisseriada/Correção de fluxo	9.674	0,34	833	0,76	619	1,09
Universo 1C	2.792.442	98,63	103.890	94,33	52.102	91,54
Escolas com menos de 10 alunos	24.507	0,87	3.781	3,43	3.766	6,62
Universo 2	2.767.935	97,76	100.109	90,90	48.336	84,93
Universo 2A: Escolas privadas (universo amostral)	405.922	14,67	16.270	16,25	10.999	22,76
Universo 2B: Escolas públicas federais	4.206	0,15	153	0,15	37	0,08
Universo 2C: Escolas públicas estaduais (universo amostral)	1.227.693	44,35	41.329	41,28	16.526	34,19
Universo 2D: Escolas públicas municipais (universo amostral)	1.130.114	40,83	42.357	42,31	20.774	42,98

Universo 1 - Alunos de 9º ano do Ensino fundamental

Universo 1A - Universo 1 excluindo escolas que têm somente turmas multisseriadas/de correção de fluxo, sem declarar turmas regulares de 9º ano do EF

Universo 1B - Universo 1A excluindo alunos de escolas indígenas que não ministram a Língua Portuguesa como primeira língua

Universo 1C - Universo 1B excluindo alunos de turmas multisseriadas/de correção de fluxo em escolas que declararam turmas regulares do 9º ano do EF

Universo 2 - Alunos de 9º ano do EF de Ensino Regular em escolas com 10 ou mais alunos matriculados

Para o 9º ano do EF:

TP_ETAPA_ENSINO = 41 (ensino fundamental de 9 anos - 9º ano)

Para as exclusões:

TP_UNIFICADA = 2, 3 e 4 identificam turmas de correção de fluxo ou multisseriadas.

TP_LINGUA_INDIGENA = 1 identifica escolas que ministrem o ensino somente em Língua Indígena

Fonte: Elaboração própria (contagem da versão preliminar do Censo escolar 2021)

TABELA A

3ª E 4ª SÉRIES EMT

3ª e 4ª séries do ensino médio (tradicional)	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 1	2.187.856	100,00	72.115	100,00	26.100	100,00
Escolas indígenas	760	0,03	53	0,07	32	0,12
Universo 1A	2.187.096	99,97	72.062	99,93	26.068	99,88
Escolas com menos de 10 alunos	7.542	0,34	1.190	1,65	1.170	4,48
Universo 2	2.179.554	99,62	70.872	98,28	24.898	95,39
Universo 2A: Escolas privadas (universo amostral)	279.972	12,85	9.595	13,54	6.846	27,50
Universo 2B: Escolas públicas federais	5.413	0,25	192	0,27	47	0,19
Universo 2C: Escolas públicas estaduais	1.884.523	86,46	60.750	85,72	17.877	71,80
Universo 2D: Escolas públicas municipais	9.646	0,44	335	0,47	128	0,51

Universo 1 - Alunos de 3ª e 4ª séries do ensino médio (tradicional)

Universo 1A - Universo 1 excluindo alunos de escolas indígenas que não ministram a Língua Portuguesa como primeira língua

Universo 2 - Alunos de 3ª e 4ª séries do Ensino médio (tradicional) de Ensino Regular em escolas com 10 ou mais alunos matriculados

Turmas multisseriadas/de correção de fluxo não estão associadas às etapas do Ensino médio

Para a 3ª e 4ª Séries do ensino médio (tradicional e integrado):

TP_ETAPA_ENSINO = 27 (ensino médio - 3º ano/3ª série)

TP_ETAPA_ENSINO = 28 (ensino médio - 4º ano/4ª série)

Para as exclusões:

TP_LINGUA_INDIGENA = 1 identifica escolas que ministrem o ensino somente em Língua Indígena

Fonte: Elaboração própria (contagem da versão preliminar do Censo escolar 2021)

TABELA A

3ª/4ª SÉRIE EMI

3ª e 4ª Séries do ensino médio (integrado)	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 1	191.692	100,00	6.901	100,00	1.973	100,00
Escolas indígenas	17	0,01	2	0,03	1	0,05
Universo 1A	191.675	99,99	6.899	99,97	1.972	99,95
Escolas com menos de 10 alunos	288	0,15	54	0,78	46	2,33
Universo 2	191.387	99,84	6.845	99,19	1.926	97,62
Universo 2A: Escolas privadas (universo amostral)	6.177	3,23	288	4,21	147	7,63
Universo 2B: Escolas públicas federais	66.166	34,57	2.334	34,10	485	25,18
Universo 2C: Escolas públicas estaduais	116.059	60,64	4.120	60,19	1.275	66,20
Universo 2D: Escolas públicas municipais	2.985	1,56	103	1,50	19	0,99

Universo 1 - Alunos de 3ª e 4ª séries do ensino médio (integrado)

Universo 1A - Universo 1 excluindo alunos de escolas indígenas que não ministram a Língua Portuguesa como primeira língua

Universo 2 - Alunos de 3ª e 4ª séries do ensino médio (integrado) de ensino regular em escolas com 10 ou mais alunos matriculados

Turmas multisseriadas/de correção de fluxo não estão associadas às etapas do ensino médio

Para a 3ª e 4ª séries do ensino médio (integrado):

TP_ETAPA_ENSINO = 32 (Curso Técnico Integrado (ensino médio Integrado) 3ª série)

TP_ETAPA_ENSINO = 33 (Curso Técnico Integrado (ensino médio Integrado) 4ª série)

Para as exclusões:

TP_LINGUA_INDIGENA = 1 identifica escolas que ministrem o ensino somente em Língua Indígena

Fonte: Elaboração própria (contagem da versão preliminar do Censo Escolar 2021)

ANEXO

ALGORITMO PARA AMOSTRAGEM SEQUENCIAL DE POISSON

Primeiro, obtenha para cada unidade no cadastro um número pseudo-aleatório, isto é, um número gerado com base na distribuição uniforme no intervalo (0;1), de forma que a geração para diferentes unidades no cadastro seja independente. Seja X_i o número aleatório gerado para a unidade i do cadastro, $i = 1, 2, \dots, N$.

Em seguida, calcule os valores relativos da medida de tamanho t a ser empregada para selecionar a amostra. Denotando por T_p o tamanho t da unidade i do cadastro, calcule os valores de

$$p_i = T_i / \sum_{k=1}^N T_k$$

para todo $i = 1, 2, \dots, N$. Observe que os valores p_i satisfazem a relação:

$$\sum_{i=1}^N p_i = 1$$

Calcule os “números aleatórios modificados” dados por $Z_i = X_i / p_i$ para todas as unidades do cadastro, $i = 1, 2, \dots, N$.

Ordene as unidades do cadastro, segundo os números aleatórios modificados, do menor para o maior valor.

Inclua na amostra as n primeiras unidades na lista ordenada, isto é, aquelas com os n menores valores dos números aleatórios modificados Z_i .

