



RELATÓRIO DE AMOSTRAGEM DO SAEB 2019

**DIRETORIA DE AVALIAÇÃO
DA EDUCAÇÃO BÁSICA
DAEB**



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 2019

Brasília-DF
Inep/MEC
2021



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 INSTRUMENTOS E MEDIDAS (CGIM)

EQUIPE TÉCNICA

Bruno Alves Pinheiro de Sousa

Ester Pereira Neves de Macedo

Giordano Alan Barbosa Sereno

Joelson Severo dos Santos Azevêdo

Laene Ascenso Lustosa

ELABORAÇÃO

Laene Ascenso Lustosa

DIRETORIA DE ESTUDOS EDUCACIONAIS (DIRED)

COORDENAÇÃO DE EDITORAÇÃO E PUBLICAÇÕES (COEP)

REVISÃO

Thaiza de Carvalho dos Santos

NORMALIZAÇÃO

Aline do Nascimento Pereira

PROJETO GRÁFICO CAPA/MIOLO

Marcos Hartwich/Raphael Freitas

DIAGRAMAÇÃO E ARTE-FINAL

Daniel Caixeta

REVISÃO GRÁFICA

Lilian dos Santos Lopes

Valéria Borges

Dados Internacionais de Catalogação na Publicação (CIP)

Brasil. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep).
Relatório de amostragem do Saeb 2019 [recurso eletrônico]. – Brasília, DF : Inep,
2021.

47 p. il.

ISBN 978-65-5801-019-7

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

CDU 371.26

SUMÁRIO

APRESENTAÇÃO	5
1 DESCRIÇÃO DO UNIVERSO DE REFERÊNCIA DO SAEB 2019.....	7
1.1 Objeto do levantamento	8
1.2 Sistema de referência	9
1.3 População de referência da amostra do Saeb 2019	9
2 PLANO AMOSTRAL.....	13
2.1 Estratos de interesse.....	13
2.2 Plano amostral	15
2.3 Seleção das turmas nas escolas selecionadas.....	19
3 EXPANSÃO DA AMOSTRA.....	21
3.1 Esquema geral de ponderação	21
3.2 Descrição da ponderação da amostra	26
3.2.1 Peso da escola.....	26
3.2.2 Peso da turma.....	27
3.2.3 Cálculo do peso dos alunos	28
3.3 Utilização dos pesos na estimação.....	29

4	ESTIMAÇÃO DA PRECISÃO DAS ESTIMATIVAS.....	31
4.1	Descrição da ponderação da parte censitária.....	32
	REFERÊNCIAS	35
	APÊNDICE.....	37
	ANEXO.....	45



APRESENTAÇÃO

O objetivo principal deste relatório é apresentar o processo de amostragem do Sistema de Avaliação da Educação Básica (Saeb) 2019. Além disso, traz informações relativas ao cálculo dos pesos para expansão dos resultados, tanto os obtidos por amostragem quanto aqueles provenientes da aplicação censitária.

O Saeb foi criado em 1990 como uma única avaliação e reestruturado em 2005, quando houve a instituição da Avaliação Nacional da Educação Básica (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 da rede pública. Em 2013, com a inclusão da Avaliação Nacional da Alfabetização (ANA), o Saeb passou a ser composto por três avaliações externas em larga escala: a Aneb, a Anresc (Prova Brasil) e a ANA.

Em 2019, às vésperas de completar três décadas de realização, o Saeb passa por uma nova reestruturação para se adequar à Base Nacional Comum Curricular (BNCC). A BNCC torna-se referência na formulação dos itens do 2º ano – Língua Portuguesa (LP) e Matemática (MT) – e do 9º ano do ensino fundamental, no caso dos testes de Ciências da Natureza (CN) e Ciências Humanas (CH), 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. Começa a avaliação da educação infantil, em estudo-piloto, com aplicação de questionários eletrônicos exclusivamente para professores e diretores. Secretários municipais e estaduais também passam a responder questionários eletrônicos.

O plano amostral do Saeb 2019 tem um desenho parecido com o adotado nas suas últimas edições, limitando-se às escolas privadas; entretanto, foram incluídos ajustes para adequação ao novo formato da avaliação. A amostra do Saeb para o ensino médio passou a incluir as escolas que oferecem o ensino médio integrado, a fim de permitir a obtenção de resultados, de forma separada, para o ensino médio tradicional e para o ensino médio integrado, e também para as duas modalidades de ensino de forma conjunta. Também foram incluídas as amostras do 9º ano do ensino fundamental para avaliação de CH e CN e do 2º ano do ensino fundamental para avaliação de LP e MT.

O Relatório de Amostragem do Saeb 2019 está organizado da seguinte forma: o primeiro capítulo trata do universo de referência da avaliação; o segundo explicita a sistemática do plano amostral e da seleção das turmas das escolas da amostra; e o terceiro dispõe sobre a metodologia da expansão dos resultados. O detalhamento sobre o estudo-piloto para avaliação da educação infantil não será tratado neste relatório.

1 DESCRIÇÃO DO UNIVERSO DE REFERÊNCIA DO SAEB 2019

Este capítulo descreve a população (ou universo) de referência do Saeb 2019, indicando quais unidades dessa população foram retiradas para compor a amostra e quais unidades 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 população de interesse é a que se gostaria de cobrir com uma pesquisa ou com um estudo, e é usualmente declarada de um modo separado. A população de referência é a que a pesquisa ou estudo vai efetivamente cobrir, e se baseia num sistema de referência bem documentado que permite identificar e localizar claramente seus membros. A população alvo do Saeb 2019 foi estabelecida pela Portaria nº 366, de 29 de abril de 2019, conforme detalhamento do Quadro 1.

QUADRO 1
DETALHAMENTO DA POPULAÇÃO ALVO DO SAEB 2019

(continua)

Rede de ensino	Mínimo de matrículas	Disciplinas	Séries	Tipo de aplicação
Escolas públicas localizadas em zonas urbanas e rurais	10	Língua Portuguesa e Matemática	5º ano e 9º ano 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 ano e 9º ano do ensino fundamental e 3ª e 4ª séries do ensino médio (tradicional e integrado)	Amostrai

QUADRO 1

DETALHAMENTO DA POPULAÇÃO ALVO DO SAEB 2019

(conclusão)

Rede de ensino	Mínimo de matrículas	Disciplinas	Séries	Tipo de aplicação
Escolas públicas e privadas localizadas em zonas urbanas e rurais	10	Ciências Humanas e Ciências da Natureza	9º ano do ensino fundamental	Amostrai
Escolas públicas e privadas localizadas em zonas urbanas e rurais	10	Língua Portuguesa e Matemática	2º ano do ensino fundamental	Amostrai
Instituições públicas ou conveniadas com o setor público, localizadas em zonas urbanas e rurais	-	-	Creche ou pré-escola da etapa da educação infantil ¹	Amostrai

Fonte: Elaborado pela Daeb/Inep com base na Portaria nº 366/2019 (Brasil. Inep, 2019).

Nota: 1 - O detalhamento sobre o estudo-piloto para avaliação da educação infantil não será tratado neste relatório.

Para delimitação da população de referência, a Portaria nº 366/2019 considerou as seguintes exclusões:

- I as escolas com menos de dez 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 2019 é uma avaliação aplicada a alunos de escolas públicas e particulares, do 2º ano, 5º ano e 9º ano do ensino fundamental e das 3ª e 4ª séries do ensino médio¹, em todas

¹ No caso do ensino médio, algumas escolas certificam os alunos como concluintes na 4ª série, enquanto outras certificam ao fim da 3ª série. Portanto, a 3ª e a 4ª séries foram consideradas indistintamente como etapas de conclusão do ensino médio, abrangendo a modalidade tradicional e a integrada.

as unidades da Federação (UFs). A avaliação envolve as áreas do conhecimento de Matemática e Língua Portuguesa e, para o 9º ano do ensino fundamental, envolve também Ciências Humanas e Ciências da Natureza. Assim, foram considerados quatro universos ou populações de referência para avaliação amostral do Saeb 2019:

- I alunos do 2º ano do ensino fundamental;
- II alunos do 5º ano do ensino fundamental;
- III alunos do 9º ano do ensino fundamental; e
- IV alunos das 3ª e 4ª séries do ensino médio.

A amostra do Saeb 2019 buscou garantir, *a priori*, estimativas confiáveis para os seguintes estratos de interesse:

- I por rede: estadual, municipal e privada para o 2º ano do ensino fundamental, e para CH e CN no 9º ano do ensino fundamental. 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;
- II por unidade da Federação;
- III por localização: urbana, rural;
- IV por área: capital, interior; e
- V modalidade de ensino (no caso do ensino médio): tradicional, integrado.

1.2 SISTEMA DE REFERÊNCIA

O sistema de referência adotado é o Censo Escolar 2018, realizado pelo Ministério da Educação/Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (MEC/Inep) para identificação das escolas que possuíam, em 2018, alunos matriculados nas séries mencionadas. Assim, operacionalmente, a população investigada será formada pelos alunos declarados no Censo Escolar 2019, pertencentes a escolas ativas no Censo Escolar 2018. Vale notar que as escolas participam da definição da população de referência com as características existentes em 2018. Não são esperadas grandes modificações nas características das escolas entre os dois períodos, principalmente em relação à criação de novas escolas, mas registra-se que as escolas criadas e implantadas após o Censo Escolar 2018 e com alunos nas séries de interesse não farão parte da população de referência da amostra do Saeb 2019.

1.3 POPULAÇÃO DE REFERÊNCIA DA AMOSTA DO SAEB 2019

Para a edição de 2019, 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 2019 nas escolas do Censo Escolar 2018, nas localizações urbanas e rurais, em uma das séries de interesse, excluindo os alunos das escolas com menos de dez alunos por série e outros critérios específicos.

Para as amostras que estão limitadas às escolas particulares, a avaliação mantém os mesmos padrões investigados. Uma vez que o estudo para as escolas públicas foi ampliado e aplicado censitariamente, 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 a 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 se possa chegar à realidade da população de referência, apresentada na Tabela 1.

TABELA 1
CONTAGEM DE ALUNOS, TURMAS E ESCOLAS DA POPULAÇÃO DE REFERÊNCIA
PARA O DIMENSIONAMENTO DA AMOSTRA DO SAEB 2019

Série	Alunos	Turmas	Escolas
2º ano EF	2.645.107	122.079	60.816
5º ano EF	2.882.901	121.216	62.565
9º ano EF	2.624.438	96.791	48.278
3ª/4ª séries EM	2.183.422	74.040	26.307
Total	10.335.868	414.126	94.697

Fonte: Elaborada pela Daeb/Inep com base em dados do Censo Escolar 2018 (Brasil. Inep, 2018).

Essas exclusões realizadas para delimitação da população de referência são detalhadas no Apêndice A, permitindo o entendimento do impacto de cada uma das exclusões efetuadas. Para o 2º ano, por exemplo, o universo de interesse em 2018 era composto por 2.864.968 alunos, distribuídos em 166.845 turmas e 101.311 escolas, conforme mostra a linha denominada “Universo 1” da Tabela A – Apêndice A.

Quando são excluídos os alunos das turmas multisseriadas ou de correção de fluxo, de escolas indígenas e escolas com menos de dez alunos, conforme descreve a linha rotulada como “Universo 2”, no contexto do 2º ano, esse contingente passa a ser formado por 2.645.107 alunos, 122.079 turmas regulares e 60.816 escolas. Tal procedimento eliminou 7,67% de alunos do universo inicial, porcentagem bem menor que a de escolas eliminadas (39,97%).

Verificamos ainda que o número médio de alunos de 2º ano por turmas multisseriadas ou de correção de fluxo, ou indígena, é de apenas 4,70, entre turmas excluídas por esses motivos. Esse 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 dez alunos na série em questão, que totalizaram, no 2º ano, 39.079 alunos em 6.286 escolas, um impacto de apenas 1,36% do total de alunos e 6,20% do total de escolas em relação ao Universo 1. Portanto, a linha denominada “Universo 2” indica o universo que seria investigado em 2018, com os critérios definidos para o Saeb 2019.

Assim, no 2º ano do ensino fundamental, o universo de alunos se reduz para 92,33% do número inicial, enquanto que o número de escolas se reduz para um pouco mais da metade, isto é, para 60,03%. Esse procedimento vai ao encontro do objetivo da pesquisa de diminuir sensivelmente o número de escolas sem deixar de lado muitos alunos.

Apresentamos ainda na Tabela A (Apêndice A), a decomposição dos números desse 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 que são avaliadas de forma censitária.

Explicações semelhantes podem ser retiradas analisando as tabelas correspondentes a cada uma das séries avaliadas. Observa-se, 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 o 2º ano e o 5º ano do ensino fundamental.



2 PLANO AMOSTRAL

Este capítulo descreve o plano amostral utilizado no Saeb 2019, considerando as mudanças em relação à edição anterior. As principais novidades são a avaliação do 2º ano do ensino fundamental e as avaliações de CH e CN para o 9º ano do ensino fundamental, com aplicação amostral tanto para alunos da rede pública quanto da rede privada, bem como 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 o integrado, separadamente e em conjunto.

2.1 ESTRATOS DE INTERESSE

O principal objetivo do plano amostral é produzir estimativas confiáveis para a média e os respectivos erros de medida da habilidade dos alunos em duas áreas de conhecimento: Língua Portuguesa e Matemática. Para o 9º ano, espera-se que as estimativas também sejam confiáveis para Ciências Humanas e Ciências da Natureza. Além das estimativas globais para o conjunto dos alunos de cada um dos universos definidos anteriormente (em cada uma das séries/anos), a avaliação também visa produzir estimativas para alguns estratos de interesse (grupos de alunos para os quais se pretende publicar resultados separados).

O primeiro grande critério de estratificação se refere ao ano/série em que o aluno está matriculado: 2º ano, 5º ano e 9º ano do ensino fundamental regular e 3ª e 4ª séries do ensino médio regular nas modalidades de ensino tradicional ou integrado. Sabendo-se que os alunos de escolas públicas com pelo menos dez alunos matriculados nesses anos/séries participam do Saeb de forma censitária, não há necessidade de selecionar amostra para essas populações. Como exceção temos as escolas públicas do 2º ano e a avaliação de CH e CN do 9º ano, que participam do Saeb de forma amostral.

O segundo critério utilizado é o da dependência administrativa da escola. Como para as avaliações de LP e MT do 5º ano, do 9º ano do ensino fundamental e das 3ª/4ª séries do ensino médio as escolas públicas (federais, estaduais e municipais) são avaliadas de forma censitária e mantidas em estratos separados das particulares, utilizou-se apenas a categoria “particular”. Já para o 2º ano e para CH e CN do 9º ano, foram consideradas as escolas estaduais, municipais e particulares. Assim, trabalhou-se com os seguintes grandes estratos:

- I 2º ano do ensino fundamental regular público ou particular (F2ParPub);
- II 5º ano do ensino fundamental regular particular (F5Par);
- III 9º ano do ensino fundamental regular particular (F9ParPub); e
- IV 3ª/4ª séries do ensino médio regular tradicional ou integrado particular (M34TIPar).

Para os alunos do 9º ano do ensino fundamental, a amostra abrange tanto escolas públicas quanto escolas particulares em função da valiação de CH e CN, no entanto, para LP e MT, a aplicação amostral limitou-se às escolas particulares, enquanto as escolas públicas participam censitariamente.

Os critérios que virão são aplicados dentro de cada um dos quatro grandes estratos. O terceiro critério é o da localização geográfica da escola. Em todos os estratos definiu-se que seriam utilizadas as UFs, ou seja, obter-se-iam estimativas para os 26 estados e para o Distrito Federal, totalizando 27 estratos nessa categoria. O quarto critério é o da localização urbana ou rural da escola, dentro da UF. O quinto critério é o da área da escola, ou seja, escolas localizadas na capital ou no interior dentro da UF e localização urbana ou rural.

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

O Quadro 2 apresenta um resumo dos critérios de estratificação que foram controlados inicialmente para a produção de estimativas confiáveis.

QUADRO 2

CRITÉRIOS PARA DEFINIR OS ESTRATOS DE INTERESSE PARA O SAEB 2019

(continua)

Critério	Quantidade	Descrição
F2ParPub		Escolas com pelo menos dez alunos na série
UF	27	26 estados + DF
Localização	2	Urbana ou rural
Área	2	Capital ou interior
Dep. Adm	3	Estadual, municipal ou privada
F5Par		Escolas com pelo menos dez alunos na série
UF	27	26 estados + DF
Localização	2	Urbana ou rural
Área	2	Capital ou interior

QUADRO 2

CRITÉRIOS PARA DEFINIR OS ESTRATOS DE INTERESSE PARA O SAEB 2019

(conclusão)

Critério	Quantidade	Descrição
F9ParPub		Escolas com pelo menos dez alunos na série
UF	27	26 estados + DF
Localização	2	Urbana ou rural
Área	2	Capital ou interior
Dep. Adm	3	Estadual, municipal ou privada
M34TIPar		Escolas com pelo menos dez alunos na série
UF	27	26 estados + DF
Localização	2	Urbana ou rural
Área	2	Capital ou interior

Fonte: Elaborado pela Daeb/Inep.

Combinando todos esses critérios, temos para este plano amostral um total de 972 possíveis estratos de interesse. Entretanto, como em alguns deles não havia nenhuma escola, houve uma redução do número de estratos. Assim, o universo de referência apresenta 246 estratos no 2º ano do ensino fundamental, 77 estratos no 5º ano do ensino fundamental, 249 estratos no 9º ano do ensino fundamental e 97 estratos nas 3ª/4ª séries do ensino médio regular, totalizando 669 estratos de interesse considerados para o sorteio da amostra. É importante destacar que mesmo que um estrato apresente um número reduzido de escolas, este foi mantido no plano amostral. Para a definição do plano amostral nos 669 estratos de interesse, não estão contempladas as escolas que fazem parte da aplicação censitária do Saeb.

2.2 PLANO AMOSTRAL

O planejamento amostral do Saeb 2019 acompanha de perto o plano adotado desde o Saeb 2003. Para manter em um único documento o plano detalhado, com adaptações para esta edição, reproduzimos partes daquele relatório (Silva; Andrade; Bussab, 2003). A população de alunos está dividida em quatro subpopulações fundamentais (que poderiam ser vistas como quatro estratos), segundo a série em que estão matriculados. No plano amostral, não seria factível, do ponto de vista prático, selecionar alunos individualmente, sendo necessário adotar alguma forma de amostragem conglomerada de alunos.

Considerando as características do universo a ser pesquisado na avaliação e a disponibilidade de um cadastro de escolas, obtido a partir dos dados do Censo Escolar 2018, foi adotado um plano amostral em duas etapas, para seleção dos alunos a serem testados como parte da amostra do Saeb 2019. Na primeira etapa do plano amostral, foram selecionadas escolas que continham alunos de cada uma das séries consideradas (2º ano, 5º ano e 9º ano 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 uma das séries. Uma vez selecionada uma turma para participar da avaliação, todos os alunos, presentes no dia da avaliação, seriam submetidos às provas nas áreas de LP e MT.

Vale ressaltar que as unidades primárias de amostragem para cada subpopulação são as escolas. Por essa razão, uma mesma escola pode participar dos universos (e, portanto, das amostras) de mais de uma série, desde que tenha turmas e alunos de mais de uma das séries consideradas.

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

- I Estratificar o universo por série a ser avaliada em quatro subpopulações distintas:
 - i escolas particulares ou públicas oferecendo 2º ano do ensino fundamental (F2ParPub);
 - ii escolas particulares oferecendo 5º ano do ensino fundamental (F5Par);
 - iii escolas particulares ou públicas oferecendo 9º ano do ensino fundamental (F9ParPub); e
 - iv escolas particulares oferecendo 3ª/4ª séries do ensino médio tradicional ou integrado (M34TIPar).
- II Estratificar as escolas em cada subpopulação, segundo os estratos de interesse. As escolas do cadastro obtido a partir do Censo Escolar 2018 foram estratificadas nos 669 estratos de interesse.
- III Calcular o tamanho da amostra total de escolas e turmas para cada um dos estratos de interesse, em cada uma das séries.

O dimensionamento da amostra foi efetuado considerando a ideia básica de oferecer precisão igual para as estimativas de proficiência para cada área do conhecimento, em cada estrato de interesse, e considerando também que seriam avaliadas duas áreas (LP e MT) em cada ano/série. 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 série e estrato de interesse usando a seguinte regra:

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

O cálculo da amostra foi baseado no Censo Escolar 2018 e nos resultados do Saeb 2017. As informações da média e da variância, por UF, foram obtidas do Saeb 2017, e o número de turmas e de alunos/turma foi obtido do Censo Escolar 2018. Como restrição do cálculo do tamanho da amostra para o 2º ano do ensino fundamental, utilizou-se a média e a variância da ANA 2016, que avaliou a alfabetização no 3º ano do ensino fundamental, por ser a referência mais próxima disponível para o universo do 2º ano do ensino fundamental.

Tendo em vista que as aplicações do Saeb 2017 e da ANA 2016 foram realizadas nas áreas de Língua Portuguesa e Matemática, e não havendo referências de variabilidade para Ciências Humanas e Ciências da Natureza no contexto do 9º ano, o cálculo do tamanho da amostra considerou a variância e a média das duas áreas, tomando como amostra final aquela que gerou a maior quantidade de turmas. Prevendo, ainda, possíveis perdas durante a aplicação das provas, nos 2º, 5º e 9º anos foram acrescidos 25% no número de turmas, nas 3ª/4ª séries de ensino médio foram acrescidas 50% de turmas. 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 suficientes e independentes para essas duas áreas de conhecimento.

Vale salientar que devido ao caráter censitário da aplicação da avaliação nas escolas públicas de 5º ano do ensino fundamental e das 3ª/4ª séries do ensino médio, a amostra de turmas de escolas particulares para essas séries correspondeu a 40% do total de turmas na amostra.

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 os seguintes critérios: *i)* pequenas escolas com uma ou duas turmas; e *ii)* grandes escolas com três ou mais.

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

Finalmente, calculou-se o número de escolas a selecionar por estrato de tamanho (estrato 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 a selecionar 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 estratos 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 as quatro subpopulações.

TABELA 2
ALOCUÇÃO ESPERADA DA AMOSTRA PARA O SAEB 2019

(continua)

Universo	F2ParPub	F5Par	F9Par	M34TIPar	Total
Alunos na população	2.643.695	490.079	2.620.448	277.667	6.031.889
Estratos de interesse	246	77	249	97	669

TABELA 2

ALOCUÇÃO ESPERADA DA AMOSTRA PARA O SAEB 2019

(conclusão)

Universo	F2ParPub	F5Par	F9Par	M34TIPar	Total
Alocação da amostra de escolas	3.062	1.261	4.928	1.343	10.594
Nº de turmas nas escolas sorteadas (censo)	6.832	2.337	11.253	2.282	22.704
Alocação da amostra de turmas	3.869	1.461	6.389	1.532	13.251
Média de alunos por turma das escolas sorteadas	21	22	28	31	25
Alocação da amostra de alunos	83.007	31.823	176.389	47.101	338.320
Fração amostral prevista (alunos)	3,14	6,49	6,73	16,96	5,61

Fonte: Elaborada pela Daeb/Inep (reflete plano amostral com base no Censo Escolar 2018).

Antes de efetuar a seleção, as escolas dentro de cada estrato foram ordenadas de acordo com os códigos de endereçamento postal (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 utilizar um método de seleção de escolas denominado “amostragem sequencial de Poisson” (Ohlsson, 1998).

Esse 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 A.

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 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.

Esse esquema de amostragem preserva o espalhamento da amostra que seria conseguido com amostragem sistemática sem, entretanto, enfrentar as 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. Nessa amostragem, criou-se um número aleatório para cada escola que não fazia parte

do processo de amostragem no Saeb (Aneb) 2017, usando o mesmo critério de geração de números aleatórios utilizado naquela edição do Saeb.

Para sequência do plano amostral, selecionaram-se turmas dentro das escolas amostradas com duas ou mais turmas na série, utilizando-se amostragem aleatória simples. No caso das escolas com apenas uma turma na série, não é necessária essa etapa de amostragem, pois a única turma da escola é escolhida automaticamente para a avaliação. Essa seleção de turmas dentro de cada escola da amostra é efetuada após a atualização, pelas secretarias estaduais de educação, das informações sobre turmas existentes nas escolas incluídas na amostra na etapa anterior. A seleção aleatória deve garantir distribuição das turmas por turno na amostra aproximadamente proporcional e similar à distribuição na população.

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. Note-se que na linha “total” há 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 das séries consideradas.

TABELA 3
NÚMERO DE ESCOLAS, TURMAS E MUNICÍPIOS SELECIONADOS POR SUBPOPULAÇÃO
A PARTIR DO CENSO ESCOLAR 2018

Subpopulação	Escolas	Turmas esperadas	Municípios
F2ParPub	3.062	3.869	1.264
F5Par	1.261	1.461	509
F9ParPub	4.928	6.389	1.807
M34TIPar	1.339	1.532	519
Total	7.735	13.251	2.205

Fonte: Elaborada pela Daeb/Inep (reflete o plano amostral com base no Censo Escolar 2018).

Nota: O total da coluna de “Escolas” e de “Municípios” registra o número total de escolas/municípios distintas(os) efetivamente selecionadas(os) para a amostra do Saeb 2019.

2.3 SELEÇÃO DAS TURMAS NAS ESCOLAS SELECIONADAS

Ao final do processo de ajuste da amostra das escolas, foram selecionadas as turmas por meio do método de amostragem simples ao acaso. Nas escolas de porte grande, foram selecionadas duas turmas e nas escolas de porte pequeno foi selecionada uma turma, conforme descrito anteriormente. Salienta-se que a seleção das escolas foi realizada utilizando o Censo Escolar 2018, já a seleção das turmas foi realizada utilizando o Censo Escolar 2019, pois é efetuada em um período posterior.

Os resultados apresentados nas Tabelas 2 e 3 tiveram algumas alterações devido aos ajustes na seleção das escolas, em função da negativa de algumas escolas em participar da avaliação e do número de turmas e alunos declarados no Censo de 2019. Os resultados atualizados são apresentados na Tabela 4.

TABELA 4

ALOCÇÃO DA AMOSTRA APÓS O AJUSTE DAS ESCOLAS PARA O SAEB 2019

Universe	F2ParPub	F5Par	F9Par	M34TIPar	Total
Alunos na população	2.649.028	507.216	2.596.247	279.895	6.032.386
Estratos de interesse	244	76	247	97	664
Alocação da amostra de escolas	2.795	1.136	4.632	1.206	7.225
Nº de turmas nas escolas sorteadas (censo)	6.288	2.181	16.506	2.038	27.013
Alocação da amostra de turmas	3.585	1.340	6.187	1.393	12.505
Média de alunos por turma das escolas sorteadas	23,5	24,2	29,3	32,4	27,4
Alocação da amostra de alunos	84.273	32.386	181.036	45.106	342.801
Fração amostral prevista (alunos)	3,181	6,385	6,973	16,115	5,683

Fonte: Elaborada pela Daeb/Inep (reflete o plano amostral atualizado para o Censo Preliminar 2019).

Ressalta-se que todas as escolas participantes do Saeb 2019 estão no Censo Escolar 2018, entretanto, após a consolidação dos dados finais do Censo Escolar 2019, 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 2019.



3 EXPANSÃO DA AMOSTRA

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

Assim, apresenta-se neste capítulo o processo de cálculo dos pesos da parte censitária e da amostra, que serão posteriormente empregados para estimar algumas quantidades de interesse dos usuários do levantamento, bem como os métodos propostos para estimar essas quantidades e as respectivas medidas de precisão, de modo a permitir sua correta implementação pelos usuários dos dados do Saeb 2019.

O capítulo está dividido em três seções: a primeira apresenta o esquema geral de ponderação; a segunda trata dos métodos adotados para a ponderação da amostra principal do Saeb 2019; e a terceira trata 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 estimar medidas de interesse referentes ao conjunto da população. Isso 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 a 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 igual ao inverso da respectiva probabilidade de inclusão na amostra. A ponderação do Saeb 2019 foi realizada de duas formas, pois a ponderação de dados com característica amostral é diferente daquela que tem característica censitária.

Na amostra utilizada no Saeb 2019, 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, diretor e secretário municipal de educação. Por causa da diversidade de unidades, o esquema de ponderação da amostra se torna mais complexo, exigindo para cada uma o seu próprio sistema de pesos para apuração dos resultados. Isso 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 2019. Essa propriedade imposta ao esquema de ponderação se chama “calibração da amostra”.

Para calcular os pesos que serão associados aos alunos da amostra do Saeb 2019, adotou-se os mesmos princípios das edições anteriores, 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.

Ainda seguindo a referência das edições anteriores, decidiu-se considerar os dados já disponíveis no Censo Escolar 2019, após aplicação de exclusões equivalentes às efetuadas na montagem do cadastro de seleção da amostra, baseado no Censo Escolar 2018. 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 2019 se ajustem aos últimos dados divulgados no Censo Escolar 2019, 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 2019 não é capaz de captar com precisão. Por exemplo, nos totais desse cadastro estão incluídas escolas e turmas novas, criadas após o Censo Escolar 2018, que não poderiam participar da seleção da amostra, tampouco do levantamento de dados do Saeb 2019. 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 no Censo Escolar 2018, o qual não captura a mudança de alunos entre escolas.

Também, decidiu-se utilizar contagens em nível do estrato de expansão para efetuar a calibração dos pesos, em que, por estrato de expansão, entende-se o estrato de seleção construído com as regras definidas na seleção da amostra do Saeb 2019, mas usando as informações obtidas no cadastro de expansão. Essa 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 2019 ou declaram menos de dez alunos em turmas regulares na série/ano. Por não fazer parte do universo de expansão, o qual condiciona que a escola tenha dez ou mais alunos em turmas regulares na série/ano de estudo, essas escolas foram excluídas do cálculo do peso. Os alunos que participaram da aplicação e pertencem a escolas com menos de dez alunos foram mantidos na base e ficaram sem peso.

O processo de expansão do universo se inicia com o cálculo de um peso igual ao inverso da respectiva probabilidade de inclusão na amostra. Para os alunos incluídos na aplicação censitária do Saeb 2019, 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 2019, cobrindo as perdas naturais do processo de amostragem e da coleta dos dados. As contagens populacionais podem não se completarem perfeitamente em função da ausência de dados para alguns estratos, caso as perdas do processo de coleta de dados atinjam um estrato por inteiro fazendo com que não haja retorno de nenhum dado para o grupo.

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 2019 e ocorre de forma gradual, iniciando-se pelas turmas, passando para as escolas e os estratos de interesse, aumentando aos poucos sua abrangência até que se chegue ao total nacional.

Para os universos da amostra do Saeb 2019 (escolas particulares de 5º ano do ensino fundamental e 3ª/4ª séries do ensino médio e escolas públicas e particulares 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 particular), 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 2019 foi realizada usando como estrato “município”, “dependência administrativa” (federal, municipal e estadual) e “localização” (urbana e rural). Nas situações em que estratos não tinham nenhum aluno com prova válida, a expansão foi realizada por “município” e “dependência administrativa”, e posteriormente foi expandida apenas para “município”. Se, mesmo assim, a expansão dos estratos não tivesse sido contemplada, fez-se a expansão para a unidade da Federação.

Os universos de expansão do Censo Escolar 2019, que este trabalho tomou como referência, estão apresentados nas Tabelas 5 a 8.

TABELA 5
DADOS DO CENSO ESCOLAR 2019 DO 2º ANO PARA EXPANSÃO

(continua)

2º ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.650.359	92,52	122.264	73,99	60.680	60,93
Universo 2A: Escolas privadas	550.692	20,78	29.417	24,06	17.303	28,52
Universo 2B: Escolas públicas federais	1.331	0,05	59	0,05	23	0,04

TABELA 5

DADOS DO CENSO ESCOLAR 2019 DO 2º ANO PARA EXPANSÃO

(conclusão)

2º ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2C: Escolas públicas estaduais	354.516	13,38	14.893	12,18	6.985	11,51
Universo 2D: Escolas públicas municipais	1.743.820	65,80	77.895	63,71	36.369	59,94

Fonte: Elaborada pela Daeb/Inep (contagem do Censo Escolar 2019).

Nota: Os percentuais da linha Universo 2 têm como totalizador as quantidades totais da população alvo, antes de serem aplicadas as exclusões.

TABELA 6

DADOS DO CENSO ESCOLAR 2019 DO 5º ANO PARA EXPANSÃO

5º ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.877.003	93,44	121.332	75,87	62.413	64,53
Universo 2A: Escolas privadas	507.216	17,63	24.402	20,11	15.447	24,75
Universo 2B: Escolas públicas federais	1.514	0,05	61	0,05	24	0,04
Universo 2C: Escolas públicas estaduais	432.398	15,03	17.517	14,44	8.094	12,97
Universo 2D: Escolas públicas municipais	1.935.875	67,29	79.352	65,40	38.848	62,24

Fonte: Elaborada pela Daeb/Inep (contagem do Censo Escolar 2019).

Nota: Os percentuais da linha Universo 2 têm como totalizador as quantidades totais da população alvo, antes de serem aplicadas as exclusões.

TABELA 7

DADOS DO CENSO ESCOLAR 2019 DO 9º ANO PARA EXPANSÃO

9º ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.600.864	97,45	96.336	90,38	48.069	84,91
Universo 2A: Escolas privadas	415.187	15,96	16.493	17,12	10.952	22,78
Universo 2B: Escolas públicas federais	4.617	0,18	165	0,17	37	0,08
Universo 2C: Escolas públicas estaduais	1.195.517	45,97	41.299	42,87	16.811	34,97
Universo 2D: Escolas públicas municipais	985.543	37,89	38.379	39,84	20.269	42,17

Fonte: Elaborada pela Daeb/Inep (contagem do Censo Escolar 2019).

Nota: Os percentuais da linha Universo 2 têm como totalizador as quantidades totais da população alvo, antes de serem aplicadas as exclusões.

TABELA 8

DADOS DO CENSO ESCOLAR 2019 DAS 3ª/4ª SÉRIES PARA EXPANSÃO

3ª/4ª séries	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.123.745	99,56	72.803	98,08	26.390	95,03
Universo 2A: Escolas privadas	279.895	13,18	9.824	13,49	6.932	26,27
Universo 2B: Escolas públicas federais	66.842	3,15	2.496	3,43	516	1,96
Universo 2C: Escolas públicas estaduais	1.765.481	83,13	60.063	82,50	18.797	71,23
Universo 2D: Escolas públicas municipais	11.527	0,54	420	0,58	145	0,55

Fonte: Elaborada pela Daeb/Inep (contagem do Censo Escolar 2019).

Nota: Os percentuais da linha Universo 2 têm como totalizador as quantidades totais da população alvo, antes de serem aplicadas as exclusões.

3.2 DESCRIÇÃO DA PONDERAÇÃO DA AMOSTRA

O cálculo da ponderação da amostra do Saeb 2019 foi feito respeitando a estrutura do plano amostral utilizado, seguindo a mesma ordem da hierarquia aplicada 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, da rede de ensino (pública e particular), da localização (urbana e rural), da área (capital e interior) e, no caso do ensino médio, da modalidade de ensino (tradicional e integrado) a que pertencem. Para mais detalhes, consultar a descrição da estratificação feita no segundo capítulo.

Toda a ponderação de escolas, turmas e alunos em cada estrato foi efetuada independentemente do 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 2019. Assim, a descrição que se apresenta é do processo de cálculo dos pesos usado 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 2019, 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:

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

3.2.1 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:

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

Observa-se que 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 nenhum aluno testado (escolas com $d_i = 0$) e que posteriormente terão seus respectivos pesos finais zerados, visto 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' . Essa propriedade é denominada “calibração da amostra”.

3.2.2 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 na série 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 $i = 1, \dots, m$.

Nota-se que, se $c_i=0$, então todas as turmas da escola i terão seus pesos automaticamente igualados a zero, pois se não houve alunos testados na escola, não pode haver alunos testados nas turmas dessa escola. Nota-se 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 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, esses pesos simples de turmas podem ser usados para estimar o número de turmas do cadastro de expansão:

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

Em que: d_{ij} é a variável indicadora de haver alunos testados na turma j , da escola i da amostra, para $j=1, \dots, n_i$ 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 no cadastro de expansão, denotado T' , e o número estimado com os pesos iniciais. Assim, os pesos calibrados de turmas ficam iguais a:

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

3.2.3 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 temos 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 esses pesos incorporem uma correção inicial devido a não resposta ou não participação de alunos. Isso 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 $k=1, \dots, n_{ij}$, $j=1, \dots, n_i$ 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 2019 não estavam contemplados no Censo Escolar 2018, conforme levantado anteriormente. Os pesos calibrados de alunos são dados por:

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

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

3.3 UTILIZAÇÃO DOS PESOS NA ESTIMAÇÃO

Os pesos, definidos e calculados conforme a descrição apresentada na seção 3.2, devem ser utilizados para estimar todas as medidas descritivas de interesse, e também devem ser considerados quando forem ajustados os 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 também 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 ; e
- y_{hijk} é o valor de uma variável de interesse y associado ao estrato h , escola i , turma j e aluno k .

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

Para estimar 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 0 em caso contrário, utilizando-se o estimador de médias acima. Contagens também podem ser acomodadas da mesma forma, utilizando o estimador de total. Esses estimadores de totais e médias podem ser facilmente calculados com 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) e Shah *et al.* (1995).

4 ESTIMAÇÃO DA PRECISÃO DAS ESTIMATIVAS

Uma das etapas fundamentais em qualquer pesquisa amostral é a estimação da precisão das estimativas. Essa atividade é fundamental para permitir conhecer e avaliar a qualidade das estimativas produzidas. Na amostra do Saeb 2019, recomenda-se adotar estimadores de desvios-padrões que possam ser obtidos usando pacotes computacionais. Tais estimadores oferecem apenas uma aproximação dos valores que seriam obtidos se fossem considerados 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 SPSS, no módulo *Complex Sample*, entre outros.

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

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

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 poucos estratos, as escolas foram incluídas na amostra, e as UPAs nesses casos foram as turmas. Os desvios-padrões podem ser obtidos calculando-se a raiz quadrada das estimativas de variância obtidas por esse método.

Essa 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 “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:

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

e

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

4.1 DESCRIÇÃO DA PONDERAÇÃO DA PARTE CENSITÁRIA

O peso dos alunos avaliados de forma censitária, como descrito anteriormente, foi expandida para os estratos “município”, “dependência administrativa” (federal, estadual e municipal) e “localização” (urbana e rural). No caso de ter havido perda de algum estrato, ou seja, nenhuma escola de um determinado estrato participou da aplicação censitária do Saeb 2019, 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 2019. Anterior à expansão para o estrato, foi feita a expansão dos alunos para as turmas e para as escolas.

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

Em que:

- $Alunos_{(turma|censo)}$ é o número total de alunos na turma declarado no Censo Escolar 2019;
- $Alunos_{(turma|válidos)}$ 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 2019;

- $\Sigma Peso_{turma(escola)}$ 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:

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

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

Em que:

- $Alunos_{(estrato|censo)}$ é o número total de alunos no estrato declarado no Censo Escolar 2019; e
- $\Sigma Peso_{turma(escola)}$ 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 2019, fez-se a expansão por município e dependência administrativa, por município e, por fim, por unidade da Federação, caso algum município tenha ficado fora da aplicação. São apresentadas as fórmulas da continuação do processo de expansão.

Expansão para município e dependência administrativa:

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

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

Expansão para município:

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

$$Peso_{Mun} = Peso_{Mun_Dep} * Peso_{aux4}$$

Expansão para unidade da Federação:

$$Peso_{aux5} = \frac{Alunos_{(UF|censo)}}{\Sigma Peso_{Mun(UF)}}$$

$$Peso_Aluno = Peso_{Mun} * Peso_{aux5}$$

Ao final dos passos descritos, o aluno receberá um peso dado por " $Peso_Aluno$ ". Salienta-se que alguns alunos ficaram com peso menor que 1. Isso ocorre principalmente entre alunos da aplicação censitária, em que a expansão é por turma. Por exemplo, uma escola declarou que tem três turmas, sendo que as três fizeram a prova, só que uma das turmas tem apenas um aluno no Censo Escolar 2019 e três alunos nessa turma fizeram a prova, com isso, o peso dos alunos dessa turma será 1/3.



REFERÊNCIAS

ANDRADE, D. F. *Memorando técnico*: relatório de todas as atividades realizadas no plano amostral do Saeb 2011. Para: MEC/Inep. Projeto: MEC-INEP-SAEB 2011, [s.l.], jun. 2012.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Portaria nº 366, de 29 de abril de 2019. Estabelece diretrizes para o planejamento e a operacionalização do Sistema de Avaliação da Educação Básica (Saeb) no ano de 2019. *Diário Oficial da União*, Brasília, 2 maio 2019. Seção 1, p. 47.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Censo Escolar*: microdados do Censo Escolar da Educação Básica. Brasília, 2018.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Censo Escolar*: microdados do Censo Escolar da Educação Básica. Brasília, 2019.

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.

SILVA, P. L. N.; BUSSAB, W.O.; ANDRADE, D.F. *Memorando Técnico* – Fundação Cesgranrio nº 01/03: definição do universo a ser investigado. Para: MEC/Inep. Projeto: MEC-Inep-Saeb03, [s.l.], maio, 2003.

SKINNER, C. J.; HOLT, D.; SMITH, T. M. F. *Analysis of complex surveys*. Chichester: John Wiley & Sons, 1989.



APÊNDICE



MEMÓRIA DA CONSTRUÇÃO DOS UNIVERSOS DE ESCOLAS DO SAEB 2019

Universos do 2º ano do ensino fundamental

Os universos do 2º ano foram construídos através da seleção das escolas com matrículas de alunos no 2º ano do ensino fundamental, diurno e noturno. Foram excluídas 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 através da seleção das escolas com matrículas de alunos no 5º ano do ensino fundamental, diurno e noturno. Foram excluídas 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 através da seleção das escolas com matrículas de alunos no 9º ano do ensino fundamental, diurno e noturno. Foram excluídas turmas de atendimento educacional especializado exclusivo ou de atividade complementar e turmas de Educação de Jovens e Adultos.

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

Os universos da 3ª e 4ª séries do ensino médio foram construídos através 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. Foram excluídas 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 DO ENSINO FUNDAMENTAL

2º ano do ensino fundamental	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 1	2.864.968	100,00	166.845	100,00	101.311	100,00
Escolas exclusivamente multisseriada/correção de fluxo	174.049	6,08	36.796	22,05	34.165	33,72
Universo 1A	2.690.919	93,92	130.049	77,95	67.146	66,28
Escolas indígenas	647	0,02	74	0,04	44	0,04
Universo 1B	2.690.272	93,90	129.975	77,90	67.102	66,23
Alunos em turma multisseriada/correção de fluxo	6.086	0,21	1.543	0,92	1.137	1,12
Universo 1C	2.684.186	93,69	128.432	76,98	67.102	66,23
Escolas com menos de dez alunos	39.079	1,36	6.353	3,81	6.286	6,20
Universo 2	2.645.107	92,33	122.079	73,17	60.816	60,03
Universo 2A: Escolas privadas (universo amostral)	549.254	20,76	28.995	23,75	17.000	27,95
Universo 2B: Escolas públicas federais	1.412	0,05	61	0,05	22	0,04
Universo 2C: Escolas públicas estaduais (universo amostral)	353.235	13,35	15.066	12,34	7.210	11,86
Universo 2D: Escolas públicas municipais (universo amostral)	1.741.206	65,83	77.957	63,86	36.584	60,16

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 ensino fundamental.

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 ensino fundamental.

Universo 2 – Alunos de 2º ano do ensino fundamental de ensino regular em escolas com dez ou mais alunos matriculados.

Para o 2º ano do ensino fundamental:

TP_ETAPA_ENSINO = 15 (ensino fundamental de nove 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 do Censo escolar 2019).

TABELA B

5º ANO DO ENSINO FUNDAMENTAL

5º ano do ensino fundamental	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 1	3.095.569	100,00	161.598	100,00	98.364	100,00
Escolas exclusivamente multisseriada/correção de fluxo	156.186	5,05	31.240	19,33	29.256	29,74
Universo 1A	2.939.383	94,95	130.358	80,67	69.108	70,26
Escolas indígenas	656	0,02	67	0,04	44	0,04
Universo 1B	2.938.727	94,93	130.291	80,63	69.064	70,21
Alunos em turma multisseriada/correção de fluxo	15.836	0,51	2.547	1,58	1.782	1,81
Universo 1C	2.922.891	94,42	127.744	79,05	69.064	70,21
Escolas com menos de dez alunos	39.990	1,29	6.528	4,04	6.499	6,61
Universo 2	2.882.901	93,13	121.216	75,01	62.565	63,61
Universo 2A: Escolas privadas (universo amostral)	490.079	17,00	23.536	19,42	15.043	24,04
Universo 2B: Escolas públicas federais	1.543	0,05	62	0,05	24	0,04
Universo 2C: Escolas públicas estaduais	448.384	15,55	18.205	15,02	8.394	13,42
Universo 2D: Escolas públicas municipais	1.942.895	67,39	79.413	65,51	39.104	62,50

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 ensino fundamental.

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 ensino fundamental.

Universo 2 – Alunos de 5º ano do ensino fundamental de ensino regular em escolas com dez ou mais alunos matriculados.

Para o 5º ano do ensino fundamental:

TP_ETAPA_ENSINO = 18 (ensino fundamental de nove 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 do Censo escolar 2019).

TABELA C

9º ANO DO ENSINO FUNDAMENTAL

9º ano do ensino fundamental	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 1	2.702.651	100,00	107.678	100,00	56.493	100,00
Escolas exclusivamente multisseriada/correção de fluxo	25.637	0,95	4.745	4,41	4.192	7,42
Universo 1A	2.677.014	99,05	102.933	95,59	52.301	92,58
Escolas indígenas	604	0,02	52	0,05	40	0,07
Universo 1B	2.676.410	99,03	102.881	95,55	52.261	92,51
Alunos em turma multisseriada/correção de fluxo	25.618	0,95	2.084	1,94	1.497	2,65
Universo 1C	2.650.792	98,08	100.797	93,61	52.261	92,51
Escolas com menos de dez alunos	26.354	0,98	4.006	3,72	3.983	7,05
Universo 2	2.624.438	97,11	96.791	89,89	48.278	85,46
Universo 2A: Escolas privadas (universo amostral)	416.889	15,88	16.361	16,90	10.746	22,26
Universo 2B: Escolas públicas federais	3.990	0,15	154	0,16	36	0,07
Universo 2C: Escolas públicas estaduais (universo amostral)	1.210.339	46,12	41.729	43,11	17.235	35,70
Universo 2D: Escolas públicas municipais (universo amostral)	993.220	37,85	38.547	39,82	20.261	41,97

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 ensino fundamental.

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 ensino fundamental.

Universo 2 – Alunos de 9º ano do ensino fundamental de ensino regular em escolas com dez ou mais alunos matriculados.

Para o 9º ano do ensino fundamental:

TP_ETAPA_ENSINO = 41 (ensino fundamental de nove 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 do Censo Escolar 2019).

TABELA D

3ª/4ª SÉRIES DO ENSINO MÉDIO (TRADICIONAL E INTEGRADO)

3ª e 4ª séries do ensino médio (tradicional e integrado)	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 1	2.192.469	100,00	75.406	100,00	27.641	100,00
Escolas exclusivamente multisseriada/correção de fluxo	-	-	-	-	-	-
Universo 1A	2.192.469	100,00	75.406	100,00	27.641	100,00
Escolas indígenas	437	0,02	30	0,04	25	0,09
Universo 1B	2.192.032	99,98	75.376	99,96	27.616	99,91
Alunos em turma multisseriada/correção de fluxo	-	-	-	-	-	-
Universo 1C	2.192.032	99,98	75.376	99,96	27.616	99,91
Escolas com menos de dez alunos	8.610	0,39	1.336	1,77	1.309	4,74
Universo 2	2.183.422	99,59	74.040	98,19	26.307	95,17
Universo 2A: Escolas privadas (universo amostral)	277.667	12,72	9.717	13,12	6.840	26,00
Universo 2B: Escolas públicas federais	60.813	2,79	2.413	3,26	485	1,84
Universo 2C: Escolas públicas estaduais	1.833.043	83,95	61.479	83,03	18.826	71,56
Universo 2D: Escolas públicas municipais	11.899	0,54	431	0,58	156	0,59

Universo 1 – Alunos de 3ª e 4ª séries do ensino médio (tradicional e 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 (tradicional e integrado) de ensino regular em escolas com dez ou mais alunos matriculados.

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

Para as 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).

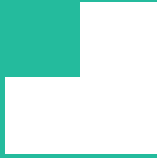
TP_ETAPA_ENSINO = 32 (curso técnico integrado – ensino médio integrado – 3ª série).

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 do Censo Escolar 2019).



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, $i=1,2,...,N$.

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

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

para todo $i=1,2,...,N$. 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, $i=1,2,...,N$.

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 .

Fonte: Andrade (2012).



CC BY-NC

VENDA PROIBIDA

