



**MINISTÉRIO DA EDUCAÇÃO**  
**INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO**  
**TEIXEIRA**

**SAEB 2017**

**RELATÓRIO DA AMOSTRAGEM**  
**DA ANEB 2017**

**Outubro de 2018**

---

## **DIRETORIA DE AVALIAÇÃO DA EDUCAÇÃO BÁSICA**

Coordenação Geral de Instrumentos e Medidas

Coordenação Geral do Sistema Nacional de Avaliação da Educação Básica

Coordenação Geral de Exames para Certificação

## SUMÁRIO

<b>APRESENTAÇÃO.....</b>	<b>5</b>
<b>CAPÍTULO 1 - DESCRIÇÃO DO UNIVERSO DE REFERÊNCIA DA ANEB 2017.....</b>	<b>6</b>
1. Objeto do Levantamento .....	6
2. Sistema de Referência .....	7
3. População de Referência da ANEB 2017 .....	7
<b>CAPÍTULO 2 - PLANO AMOSTRAL .....</b>	<b>9</b>
1. Estratos de Interesse .....	9
2. Plano Amostral .....	10
3. Seleção das turmas nas escolas selecionadas .....	14
<b>CAPÍTULO 3 – EXPANSÃO DA AMOSTRA .....</b>	<b>16</b>
1. Esquema geral de ponderação .....	16
<b>2. Descrição da ponderação da amostra.....</b>	<b>19</b>
3. Utilização dos pesos na estimação.....	23
4. Estimação da precisão das estimativas.....	24
5. Descrição da ponderação da parte censitária (Prova Brasil).....	26
<b>BIBLIOGRAFIA.....</b>	<b>28</b>
<b>ANEXO A.....</b>	<b>29</b>
<b>ANEXO B.....</b>	<b>33</b>



## APRESENTAÇÃO

O objetivo deste relatório é apresentar o processo de amostragem da Avaliação Nacional da Educação Básica - Aneb 2017. Esta avaliação faz parte do Sistema de Avaliação da Educação Básica - Saeb, 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: a Aneb, a Anresc (Prova Brasil) e a ANA.

O plano amostral da Aneb 2017 tem um desenho parecido com o adotado nas edições do Saeb de 1999, 2001, 2003, 2005, 2009, 2011, 2013 e 2015. Foram mantidas as modificações referentes aos novos critérios de estratificação introduzidos em 2013 e as semelhanças dos planos amostrais anteriores, o que permite aos usuários daquelas edições refazerem as suas análises sem muitas dificuldades. Entretanto, para a edição de 2017, a amostra foi limitada às escolas particulares, em função da ampliação da Anresc (Prova Brasil), que passou a alcançar também as instituições públicas de ensino médio e as instituições públicas que oferecem o 5º ou o 9º ano do ensino fundamental a um número de 10 a 19 alunos.

Este relatório 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 da amostra.

## CAPÍTULO 1 - DESCRIÇÃO DO UNIVERSO DE REFERÊNCIA DA ANEB 2017

Este capítulo descreve a população (ou universo) de referência da Aneb 2017, indicando quais unidades dessa população foram retiradas para compor a amostra da Aneb e quais unidades compuseram a parte censitária, a Anresc (Prova Brasil), que também é incorporada à Aneb.

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 estudo, e é usualmente declarada de um modo solto. A *população de referência* é a que a pesquisa ou estudo vai efetivamente cobrir, e se baseia num sistema de referência bem documentado e que permite identificar claramente e localizar os seus membros.

### 1. Objeto do Levantamento

A Aneb é uma avaliação aplicada a alunos das escolas públicas e particulares, do 5º ano e do 9º ano do ensino fundamental<sup>1</sup> e da 3ª série do ensino médio<sup>2</sup>, em todas as Unidades da Federação. A avaliação envolve as áreas do conhecimento de Matemática (MT) e Língua Portuguesa (LP). Assim, foram considerados três universos ou populações de referência para a Aneb 2017:

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

A Aneb 2017 buscou garantir, a priori, estimativas confiáveis para os seguintes estratos de interesse:

- por rede: pública (estadual e municipal) e privada. Para a dependência administrativa federal haverá um único estrato Brasil;
- por unidade da federação;
- por localização: urbana, rural;
- por área: capital, interior.

---

<sup>1</sup> Como ainda existem instituições que oferecem o Ensino Fundamental de 8 anos, sempre que foram selecionadas turmas de 5º ano do Ensino Fundamental de 9 anos, subentende-se do texto que também foram selecionadas turmas de 4ª série do Ensino Fundamental de 8 anos. E da mesma forma, sempre que foram selecionadas turmas de 9º ano do Ensino Fundamental de 9 anos estão incluídas as turmas de 8ª série do Ensino Fundamental de 8 anos.

<sup>2</sup> No caso do Ensino Médio, como algumas escolas certificam na 4ª série, quando mencionamos a 3ª série do Ensino Médio também estamos incluindo a 4ª série.

## 2. Sistema de Referência

O Sistema de Referência adotado é o Censo Escolar de 2016, realizado pelo MEC/INEP. Valeu-se principalmente de informações das escolas que possuíam, em 2016, alunos nas séries mencionadas. Assim, operacionalmente, a população investigada será formada pelos alunos de 2017, pertencentes a escolas ativas no Censo Escolar de 2016. Vale notar que as escolas participam da definição da população de referência com as características existentes no ano de 2016, não se espera 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 de 2016 e com alunos nas séries de interesse não farão parte da população de referência da Aneb 2017.

## 3. População de Referência da ANEB 2017

Para a edição de 2017, utilizou-se o mesmo padrão investigado no Saeb 2013. Desse modo, o universo da pesquisa/avaliação é o seguinte:

*Todos os alunos de turmas regulares matriculados em 2017 nas escolas do Censo Escolar de 2016 nas localizações urbanas e rurais em uma das três séries de interesse, excluindo os alunos das escolas com menos de 10 alunos por série e outros critérios específicos.*

Apesar de a amostra ter sido limitada às escolas particulares, ao manter os mesmos padrões investigados, uma vez que o estudo para as escolas públicas foi ampliado e aplicado censitariamente na Anresc, 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.

**Tabela 1.** Contagem de alunos, turmas e escolas da população de referência para o sorteio das escolas da amostra (dados do Censo Escolar de 2016)

Série	Alunos	Turmas	Escolas
5º Ano EF	2.867.993	120.765	63.229
9º Ano EF	2.756.508	101.134	48.815
3ª Série EM	2.145.274	71.088	24.828
<b>Total</b>	<b>7.769.775</b>	<b>292.987</b>	<b>136.872</b>

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 tabela 1. Estas exclusões são detalhadas no Anexo A, permitindo o entendimento do impacto de cada uma das exclusões efetuadas.

Para o 5º ano, o universo de interesse em 2016 era composto por 3.102.847 alunos, distribuídos em 166.016 turmas e 103.349 escolas, conforme mostra a linha denominada Universo 1 da Tabela A.5º Ano – Anexo A.

Quando são excluídos os alunos das turmas multisseriadas, os alunos de aceleração, de escolas indígenas e escolas com menos de 10 alunos, conforme descreve a linha rotulada como Universo 2, este contingente passa a ser formado por 2.867.993 alunos, 120.765 turmas regulares e 63.229 escolas. Este procedimento elimina 7,57% de alunos do universo inicial, porcentagem bem menor que a de escolas eliminadas (38,82%).

Verificamos ainda que o número médio de alunos de 5º ano por turmas multisseriadas, ou de aceleração, ou indígena é de apenas 5,03, entre as escolas excluídas por esse motivo. 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 da Aneb foram excluídos do universo de pesquisa as escolas com menos de 10 alunos na série em questão, que totalizaram 43.556 alunos em 7.132 escolas, um impacto de apenas 1,40% do total de alunos e 6,90% do total de escolas, em relação ao Universo 1. Portanto, a linha denominada Universo 2 indica o universo que seria investigado em 2016 com os critérios definidos acima para o Aneb 2017.

Assim, no 5º ano do ensino fundamental, o universo original de aluno reduz-se para 92,43% do número inicial, enquanto que o número de escolas reduz-se para um pouco menos da metade, isto é, para 61,18%. Este procedimento vai ao encontro do objetivo da pesquisa em diminuir sensivelmente o número de escolas sem deixar de lado muitos alunos.

Apresentamos ainda naquela tabela, a decomposição dos números deste universo para a dependência administrativa Federal, para as escolas que formarão o universo a ser amostrado na ANEB e para as escolas que formarão o universo Anresc (Prova Brasil).

Explicações semelhantes podem ser retiradas analisando as tabelas correspondentes ao 9º Ano, ver Tabela A.9º Ano e à terceira série do ensino médio, ver Tabela A.3ª Série EM. Observe, entretanto, que as reduções não são tão acentuadas como as observadas para o 5º ano.



## CAPÍTULO 2 - PLANO AMOSTRAL

Este capítulo descreve o plano amostral utilizado na Aneb 2017, em que não houve mudanças em relação à edição anterior, além da já mencionada limitação da amostra às escolas particulares.

### 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 (LP) e Matemática (MT). 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 refere-se ao ano/série em que o aluno está matriculado, a saber: 5º ano e 9º ano do Ensino Fundamental Regular e 3ª série do Ensino Médio Regular. Sabendo-se que os alunos de escolas públicas com pelo menos 10 alunos matriculados nesses anos/séries participam da Anresc (Prova Brasil), que tem caráter censitário, não há necessidade de selecionar amostra para estas populações.

O segundo critério utilizado é o da dependência administrativa da escola. Como 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. Assim, trabalhou-se com os seguintes grandes estratos:

- 1) 5º ano do ensino fundamental regular particular (F5/4Par);
- 2) 9º ano do ensino fundamental regular particular (F9/8Par);
- 3) 3ª do ensino médio regular particular (M3RPar).

Os critérios que virão a seguir são aplicados dentro de cada um dos três grandes estratos.

O terceiro critério é o da localização geográfica da escola. Em todos os estratos definiu-se que seriam utilizados as Unidades de Federação (UF), ou seja, obter-se-ia estimativas para os 26 estados e para o Distrito Federal, totalizando 27 estratos nesta categoria.

O quarto critério é o da localização urbana ou rural da escola, dentro da Unidade de Federação.

O quinto critério é o da área da escola, ou seja, escolas localizadas na capital ou no interior dentro da Unidade de Federação e localização urbana ou rural.

O Quadro 1, a seguir, apresenta um resumo dos critérios de estratificação que foram controlados inicialmente para a produção de estimativas confiáveis.

**Quadro 1.** Critérios para definir os estratos de interesse para o SAEB/ANE2015

<b>Critério</b>	<b>Quantidade</b>	<b>Descrição</b>
<b>F5/4Par</b>		<b>Escolas com pelo menos 10 alunos na série</b>
UF	27	26 estados+ DF
Localização	2	Urbana ou Rural
Área	2	Capital ou Interior
<b>F9/8Par</b>		<b>Escolas com pelo menos 10 alunos na série</b>
UF	27	26 estados+ DF
Localização	2	Urbana ou Rural
Área	2	Capital ou Interior
<b>M3RPar</b>		<b>Escolas com pelo menos 10 alunos na série</b>
UF	27	26 estados+ DF
Localização	2	Urbana ou Rural
Área	2	Capital ou Interior

Combinando todos estes critérios, temos para este plano amostral um total de 324 possíveis estratos de interesse. Entretanto, como em alguns desses estratos havia nenhuma escola, houve uma redução do número de estratos. Assim, o universo de referência apresenta 76 estratos no 5º ano do Ensino Fundamental, 77 estratos no 9º ano do Ensino Fundamental e 71 estratos na 3ª série do Ensino Médio Regular, totalizando 225 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 225 estratos de interesse, não estão contempladas as escolas que fazem parte da Anresc (Prova Brasil), pelo fato de sua aplicação ter caráter censitário.

## **2. Plano Amostral**

O planejamento amostral da Aneb 2017 acompanha de perto o plano adotado desde o Saeb 2003. Para manter em um único documento o plano detalhado, com adaptações para essa edição, reproduzimos abaixo partes daquele relatório, Silva *et al.* (2003).

A população de alunos está dividida em três subpopulações fundamentais (que poderiam ser vistas como três 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 de 2016, foi adotado um plano amostral em duas etapas, para seleção dos alunos a serem testados como parte da Aneb 2017. Na primeira etapa do plano amostral, foram selecionadas escolas que contenham alunos de cada

uma das séries consideradas na Aneb 2017, a saber, 5º ano e 9º ano do Ensino Fundamental e 3ª série do Ensino Médio.

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 da turma, presentes no dia da avaliação, seriam submetidos às provas nas áreas de Língua Portuguesa e Matemática.

Para as escolas de Ensino Médio, a segunda etapa não foi aplicada, em decorrência da necessidade de gerar resultados por escolas, conforme estabelecido na Portaria INEP Nº 447, de 24 de maio de 2017, que só seria possível com a aplicação a todos os alunos, uma vez que não seria viável calcular o tamanho de amostra adequado a cada uma das escolas. Assim, para o Ensino Médio, todos os alunos da escola, 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 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:

- *Estratificar o universo por série a ser avaliada em três sub-populações distintas:*
  - escolas particulares oferecendo 5º ano do ensino fundamental (F5/4Par);
  - escolas particulares oferecendo 9º ano do ensino fundamental (F9/8Par);
  - escolas particulares oferecendo 3ª série do ensino médio regular (M3RPar).
- *Estratificar as escolas em cada subpopulação segundo os estratos de interesse.* As escolas do cadastro obtido a partir do Censo Escolar 2016 foram estratificadas nos 225 estratos de interesse.
- *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 edições últimas edições na Aneb (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 5 pontos absolutos na escala (250,50) para o 5º ano do Ensino Fundamental, 9º ano do Ensino Fundamental e 3ª série do Ensino Médio Regular.*

O cálculo da amostra foi baseado no Censo Escolar 2016 e na Anresc (Prova Brasil) 2015. As informações da média e da variância, por UF, foram obtidas da Prova Brasil 2015 e o número de

turmas e de alunos/turma foram obtidos do Censo Escolar 2014. Como restrição do cálculo do tamanho da amostra para a 3ª série do Ensino Médio Regular, utilizou-se a média e variância do 9º ano, pois a Prova Brasil é aplicada somente no 5º ano e no 9º ano.

Tendo em vista que a aplicação da Aneb 2015 foi realizada nas áreas de Língua Portuguesa e Matemática, o cálculo do tamanho da amostra considerou a variância e a média das duas áreas, obtidas na Prova Brasil 2015, tomando como amostra final aquela que gerou a maior quantidade de turmas. Prevendo, ainda, possíveis perdas, durante a aplicação das provas, nos 5º ano e 9º ano foram acrescidos 25% de turmas, na 3ª série regular foram acrescidos 50% de turmas,.

Vale salientar que devido ao caráter censitário da aplicação da avaliação nas escolas públicas, a amostra de turmas de escolas particulares 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 o seguinte critério:

- a) Pequenas escolas com 1 ou 2 turmas;
- b) Grandes escolas com 3 ou mais.

Em seguida, alocou-se o número de turmas proporcional ao tamanho em cada combinação de Localização e Área (Urbana/Capital, Urbana/Interior, Rural/Capital, Rural/Interior) e Porte. Ressalta-se que todo estrato Localização/Área 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 com uma ou duas turmas, e duas turmas por escola nas demais). 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 os três grandes estratos de interesse.

**Tabela 2.** Alocação esperada da amostra para a Aneb 2017

<b>Universo</b>	<b>F5/4Par</b>	<b>F9/8Par</b>	<b>M3RPar</b>	<b>Total</b>
Alunos na população	471.666	411.779	308.487	1.191.932
Estratos de interesse	76	77	72	225
Alocação da amostra de escolas	1.081	903	995	2.979
Nº de turmas nas escolas sorteadas (censo)	1.986	1.662	1.887	5.535
Alocação da amostra de turmas	1.256	1.064	1.887	4.207
Média de alunos por turma das escolas sorteadas	21,95	26,95	32,75	27,13
Alocação da amostra de alunos	27.564	28.678	61.796	118.038
Fração amostral prevista (alunos)	5,84	6,96	20,03	9,90

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 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, 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. No entanto, 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 no Saeb (Aneb) 2015, usando o mesmo critério de geração de números aleatórios que foi usado 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 esta etapa de amostragem, pois a única turma da escola é escolhida automaticamente para a avaliação. Esta 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 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 três 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 2016

<b>Sub-população</b>	<b>Escolas</b>	<b>Turmas Esperadas</b>	<b>Municípios</b>
F5/4Par	1.081	1.256	449
F9/8Par	903	1.064	395
M3RPar	995	1.887	415
<b>Total (1)</b>	<b>1.882</b>	<b>4.207</b>	<b>649</b>

(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 da ANEB 2017.

### 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 2016, já a seleção das turmas foi realizada utilizando o censo escolar 2017, pois é efetuada em um período posterior.

Mais uma vez destaca-se que para as escolas do ensino médio deste procedimento não foi efetuado e todos os alunos presentes no dia da avaliação, seriam submetidos às provas nas áreas de Língua Portuguesa e Matemática.

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 do número de turmas e alunos declarados no censo preliminar de 2017. Os resultados atualizados são apresentados na Tabela 4.

**Tabela 4.** Alocação da amostra após o ajuste das escolas para o ANEB 2017

<b>Universo</b>	<b>F5/4Par</b>	<b>F9/8Par</b>	<b>M3RPar</b>	<b>Total</b>
Alunos na população	496.027	410.803	300.436	1.207.266
Estratos de interesse	80	82	74	236
Alocação da amostra de escolas	973	813	911	1.719
Nº de turmas nas escolas sorteadas (censo)	1.142	968	1.667	3.777
Alocação da amostra de turmas	1.561	1.100	1.204	3.865
Média de alunos por turma das escolas sorteadas	17,9	25,3	48,5	29,5
Alocação da amostra de alunos	27.909	27.802	58.443	114.154
Fração amostral prevista (alunos)	0,056	0,068	0,195	9,456

Ressalta-se que todas as escolas participantes da Aneb 2017 estão no censo escolar 2016. Essa condição é importante para o cálculo dos pesos dos alunos ao final da aplicação da Aneb 2017.

## **CAPÍTULO 3 – EXPANSÃO DA AMOSTRA**

Este capítulo descreve os procedimentos adotados para ponderação (cálculo dos pesos) da amostra realizada. Também descreve a expansão utilizada para parte censitária (Prova Brasil – Anresc 2017), bem como dos procedimentos necessários para estimação das medidas de interesse com base na amostra. Esta descrição é muito similar em formato e conteúdo aos textos anteriores que descrevem os procedimentos de ponderação adotados no SAEB 2009, no SAEB 2011, no SAEB 2013 e no SAEB 2015.

Assim, apresenta-se neste capítulo 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 da Aneb 2017.

O capítulo está dividido em três seções: seção 1 apresenta o esquema geral de ponderação; a seção 2 trata dos métodos adotados para a ponderação da amostra principal da Aneb 2017; a seção 3 trata dos métodos adotados para a ponderação da parte censitária da Anresc (Prova Brasil) 2017.

### **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. 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 da Aneb e da Anresc 2017 foi realizada de duas formas, pois a Aneb tem característica amostral enquanto que a Anresc (Prova Brasil) tem característica censitária.

Na amostra utilizada na Aneb 2017, 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 escola. 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 2017. Essa propriedade imposta ao esquema de ponderação se chama de calibração da amostra.

Para calcular os pesos que serão associados aos alunos da amostra da ANEB 2017 adotou-se os mesmos princípios das edições anteriores do SAEB (1997, 1999, 2001, 2003, 2005, 2009, 2011, 2013 e 2015), pois o INEP manifestou interesse em 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 dentro do espírito das edições anteriores, decidiu-se considerar os dados já disponíveis do Censo Escolar 2017, após aplicação de exclusões equivalentes às efetuadas na montagem do cadastro de seleção da amostra, baseado no Censo Escolar de 2016. 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 na ANEB 2017 se ajustem aos últimos dados divulgados do Censo Escolar 2017, facilitando a interpretação dos resultados. Por outro lado, tem a desvantagem de refletir mudanças no universo de escolas, turmas e alunos que a amostra da ANEB 2017 não é capaz de captar com precisão. Por exemplo, nos totais deste cadastro estão incluídas escolas e turmas novas, criadas após o Censo Escolar 2016, que não poderiam participar da seleção da amostra nem tampouco do levantamento de dados na ANEB 2017. Portanto, vale reforçar que a amostra prevista é baseada no Censo Escolar 2016, 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, onde por estrato de expansão se entende o estrato de seleção construído com as regras definidas na seleção da amostra da ANEB 2015, mas 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 2017 ou declaram menos de 10 alunos em turmas regulares na série/ano. Por não fazer parte do universo de expansão, o qual condiciona que a escola tenha 10 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 10 alunos foram colocados na base e ficaram sem peso.

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 ANRESC 2017, onde a aplicação é censitária, 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 2017, 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 2017 e ocorre de forma gradual iniciando 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 ANEB 2017 (escolas particulares de 5º ano e 9º ano do Ensino Fundamental e particulares de 3ª série do Ensino Médio), a amostragem tem como estrato a Unidade da Federação, a Rede de Ensino (Pública ou Particular), a Localização (Urbana e Rural) e a Área (Capital ou Interior).

A ANEB 2017 suprimiu a amostragem de escolas públicas de 5º Ano e 9º Ano com 10 a 19 alunos e de escolas públicas de 3ª série do Ensino Médio, passando a incluí-las no universo da ANRESC 2017, com aplicação censitária.

A expansão do universo da ANRESC 2017 foi realizada usando como estrato Município, dependência administrativa (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 o Município. Se mesmo assim, a expansão dos estratos não tivesse sido contemplada, fez-se a expansão para a unidade da federação.

Como os universos dos alunos das escolas Federais têm caráter censitário, o número de respondentes da prova é grande o suficiente para expandir para estratos menores, assim o cálculo do peso destes alunos adotou o mesmo esquema da ANRESC 2017.

Os universos de expansão do Censo Escolar 2017, que esse trabalho tomou como referência, estão apresentados nas tabelas abaixo.

**Tabela 5.** Dados do Censo Escolar 2017 do 5º Ano para expansão

5º Ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	<b>2.891.346</b>	<b>92,78</b>	<b>122.024</b>	<b>73,98</b>	<b>63.258</b>	<b>62,61</b>
Universo 2A: Escolas Privadas (amostral)	479.601	16,59	22.921	18,78	14.772	23,35
Universo 2B: Escolas Federais	1.526	0,05	61	0,05	24	0,04
Universo 2C: Escolas Estaduais e Municipais	2.410.219	83,36	99.042	81,17	48.462	76,61

**Tabela 6.** Dados do Censo Escolar 2017 do 9º Ano para expansão

9º Ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	<b>2.564.498</b>	<b>97,04</b>	<b>95.370</b>	<b>89,80</b>	<b>48.225</b>	<b>85,39</b>
Universo 2A: Escolas Privadas (amostral)	406.120	15,37	15.962	15,03	10.583	18,74
Universo 2B: Escolas Federais	4.052	0,15	159	0,15	35	0,06
Universo 2C: Escolas Estaduais e Municipais	2.154.326	81,52	79.249	74,62	37.607	66,59

**Tabela 7.** Dados do Censo Escolar 2017 do 3ª série para expansão

3ª Série	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	<b>2.072.633</b>	<b>99,63</b>	<b>68.771</b>	<b>98,33</b>	<b>24.781</b>	<b>95,58</b>
Universo 2A: Escolas Privadas (amostral)	298.710	14,41	10.075	14,65	6.894	27,82
Universo 2B: Escolas Federais	5.548	0,27	208	0,30	53	0,21
Universo 2C: Escolas Estaduais e Municipais	1.768.375	85,32	58.488	85,05	17.834	71,97

## 2. Descrição da ponderação da amostra

### (ANEB: 5º e 9º Ano Particular e 3ª Série Regular Particular)

O cálculo da ponderação da amostra principal da ANEB 2017 foi feito respeitando-se a estrutura do plano amostral utilizado, seguindo à 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 particular), localização (urbana e rural) e área (capital e interior) a que pertencem. Para maiores detalhes, veja a descrição da estratificação feita no Capítulo 2.

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 2017.

Assim, a descrição que se apresenta a seguir 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 ANEB 2017 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 2 escolas; e

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

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

Observe que os valores  $p_i$  satisfazem a relação:  $\sum_{i=1}^M p_i = 1$ . Conseqüentemente, 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

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

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, de 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 este 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, se vai reproduzir 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,

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

Onde:

$T_i^*$  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  $i = 1, \dots, m$ .

Note 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. 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 ( $T'_i > 2$ ).

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:

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

onde  $d_{ij}$  é a variável indicadora de haver alunos testados na turma  $j$  da escola  $i$  da amostra, para  $j=1, \dots, n_i$  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 do cadastro de expansão, denotado  $T'$ , e o número estimado com os pesos iniciais. Assim, os pesos calibrados de turmas ficam iguais a

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

### 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 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 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

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

Onde:

$d_{ijk}$  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  $k=1,...,n_{ij}$ ,  $j=1,...,n_i$  e  $i=1,...,m$ . 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 2017 não estavam contemplados no Censo Escolar 2016, conforme levantado anteriormente. Os pesos calibrados de alunos são dados por

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

onde  $A'$  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, rede e área.

### 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 também 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 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

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

Onde:

$c_{hijk}$  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 assim descrita é válida tanto para estimativas de totais baseadas na amostra de escolas, ou turmas e ou ainda de alunos, bastando adotar o conjunto de pesos e unidades apropriados para cada caso.

Para estimar médias, a idéia 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, e usando-se o estimador de médias acima. Contagens também podem ser acomodadas da mesma forma, usando o estimador de total. Estes 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. Esta atividade é fundamental para permitir conhecer e avaliar a qualidade das estimativas produzidas. Na ANEB 2017 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 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 SPSS no módulo *Complex Sample*, entre outros.

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

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

onde



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

$\hat{Y}_{hi}$  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 UPA's 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 desvios 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:

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

onde

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

e

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

## 5. Descrição da ponderação da parte censitária (Prova Brasil)

O peso dos **alunos** da Prova Brasil, como descrito anteriormente, foi expandida para o estrato Município, Dependência Administrativa (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 Prova Brasil, foi feita a expansão para o Município e posteriormente para a Unidade da Federação, dessa forma, a soma dos pesos totalizam a quantidade de alunos do Censo Escolar 2015. Anteriormente a expansão para o estrato, foi feita a expansão dos alunos para as turmas e para as escolas. O esquema de expansão adotado segue abaixo:

$$\text{Peso}_{\text{turma}} = \frac{\text{Alunos}_{(\text{turma}|\text{censo})}}{\text{Alunos}_{(\text{turma}|\text{validos})}}$$

Onde:

$\text{Alunos}_{\text{turma}/\text{censo}}$  é o número total de alunos na turma declarado no Censo Escolar 2015;

$\text{Alunos}_{\text{turma}|\text{validos}}$  é o número total de alunos da turma que realizaram a prova; e

$\text{Peso}_{\text{turma}}$  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:

$$\text{Peso}_{\text{aux}} = \frac{\text{Alunos}_{(\text{escola}|\text{censo})}}{\sum \text{Peso}_{\text{turma}(\text{escola})}}$$

$$\text{Peso}_{\text{escola}} = \text{Peso}_{\text{turma}} * \text{Peso}_{\text{aux}}$$

Onde:

$\text{Alunos}_{\text{escola}/\text{censo}}$  é o número total de alunos na escola declarado no Censo Escolar 2015;

$\sum \text{Peso}_{\text{turma}(\text{escola})}$  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

$\text{Peso}_{\text{escola}}$  é o peso do aluno expandido para a escola.

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

$$\text{Peso}_{\text{aux2}} = \frac{\text{Alunos}_{(\text{estrato}|\text{censo})}}{\sum \text{Peso}_{\text{escola}(\text{estrato})}}$$

$$\text{Peso}_{\text{estrato}} = \text{Peso}_{\text{escola}} * \text{Peso}_{\text{aux2}}$$

Onde:

$\text{Alunos}_{\text{estrato}/\text{censo}}$  é o número total de alunos no estrato declarado no Censo Escolar 2015;

$\sum \text{Peso}_{\text{escola}(\text{estrato})}$  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 2017, 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 é apresentada as fórmulas da continuação do processo de expansão.

Expansão para o Município e Dependência Administrativa:

$$\text{Peso}_{\text{aux3}} = \frac{\text{Alunos}_{(\text{Mun\_Dep}|\text{censo})}}{\sum \text{Peso}_{\text{estrato}(\text{Mun\_Dep})}}$$

$$\text{Peso}_{\text{Mun\_Dep}} = \text{Peso}_{\text{estrato}} * \text{Peso}_{\text{aux3}}$$

Expansão para o Município:

$$\text{Peso}_{\text{aux4}} = \frac{\text{Alunos}_{(\text{Mun}|\text{censo})}}{\sum \text{Peso}_{\text{Mun\_Dep}(\text{Mun})}}$$

$$\text{Peso}_{\text{Mun}} = \text{Peso}_{\text{Mun\_Dep}} * \text{Peso}_{\text{aux4}}$$

Expansão para a Unidade da Federação (UF):

$$\text{Peso}_{\text{aux5}} = \frac{\text{Alunos}_{(\text{UF}|\text{censo})}}{\sum \text{Peso}_{\text{Mun}(\text{UF})}}$$

$$\text{Peso\_Aluno} = \text{Peso}_{\text{Mun}} * \text{Peso}_{\text{aux5}}$$

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 na Prova Brasil, onde a expansão é por turma. Por exemplo, uma escola declarou que tem 3 turmas, sendo que as três turmas fizeram a prova, só que uma das turmas tem apenas 1 aluno no Censo Escolar 2017 e 3 alunos nessa turma fizeram a prova, com isso, o peso dos alunos dessa turma será 1/3.

## **BIBLIOGRAFIA**

Andrade, D.F. (2011). Plano Amostral para o SAEB 2011

Andrade, D.F. (2011). Plano Amostral do SAEB-2011 – Definição do Universo a ser Investigado. Brasília, DF. MEC/INEP.

Andrade, D.F. (2011). Plano Amostral do SAEB-2011 – Definição do Plano Amostral. Brasília, DF. MEC/INEP.

Andrade, D.F., Silva, P.L.N. e Bussab, W.O. (1999). Plano Amostral SAEB-99 – Definição do Universo a Ser Investigado, no. 01/99.

Andrade, D.F., Silva, P.L.N. e Bussab, W.O. (2001). Plano Amostral para o SAEB-01.

Bussab, W.O., Cardoso, R.H. e Rabello, M.C.R. (2005). Sistema de Ponderação SAEB2005

Bussab, W.O., Miazaki, E.S. e Rabello, M.C.R. (2009). Plano Amostral para o SAEB2009

Bussab, W.O., Andrade, D.F., Silva, P.L.N. e Freitas, M.P.S. (1999). Plano Amostral do SAEB99 – Definição do Plano Amostral – no. 02/99. INEP (1999). Projeto básico para a execução do SAEB99 e pré-testagem de itens para o Banco Nacional de Itens. Brasília, DF: MEC/INEP.

Bussab, W.O., Andrade, D.F., Nishimura, R. (2009). Plano Amostral do SAEB2009 – Definição do Universo da Pesquisa. Memorando Técnico no. 01/09. Brasília, DF. MEC/INEP.

Ohlsson, E. (1998). Sequential Poisson Sampling. *Journal of Official Statistics*, 14, p. 149-162.

Pessoa, D.G.C. e Silva, P.L.N. (1998). *Análise de Dados Amostrais Complexos*. São Paulo: Associação Brasileira de Estatística.

Shah, B.V et al. (1995). *Statistical Methods and Mathematical Algorithms Used in SUDAAN*. Research Triangle Park, NC: Research Triangle Institute.

Silva, P.L.N. (2003). Plano Amostral do SAEB-99: Procedimentos de Estimção com a Amostra Realizada.

Silva, P.L.N., Andrade, D.F. e Bussab, W.O. (2003A). Plano Amostral SAEB-2003: Memo: No.01/03- Definição do Universo a Ser Investigado.

Silva, P.L.N., Bussab, W.O., Andrade, D.F. e Freitas, M.P.S. (2000). Plano Amostral SAEB-99 – Procedimentos de Estimção com a Amostra Realizada, no. 04/99.

Silva, P.L.N., Freitas, M.P.S., Bussab, W.O. e Andrade, D.F. (1999). Plano Amostral SAEB-99 – Avaliação e Substituição de Escolas Perdidas, no. 03/99.

Skinner, C.J., Holt, D. e Smith, T.M.F. (1989). *Analysis of Complex Surveys*. Chichester: John Wiley & Sons.

## **ANEXO A**

### **Memória da Construção dos Universos de Escolas da ANEB 2017**

#### **Universos do 5º ano - EF**

Os universos do 5º ano foram construídos através da seleção das escolas com matrículas de alunos no 5º ano do EF, diurno e noturno. Excluindo escolas que tenham somente EJA.

#### **Universos do 9º ano - EF**

Os universos do 9º ano foram construídos através da seleção das escolas com matrículas de alunos no 9º ano do EF, diurno e noturno. Excluindo escolas que tenham somente EJA.

#### **Universos da 3ª série - EM**

Os universos da 3ª série do Ensino Médio foram construídos através da seleção das escolas com matrículas de alunos na 3ª série e 4ª série, diurno e noturno. Excluindo escolas que tenham somente EJA.

---

**Tabela A.5º Ano**

5º ano (4ª série) – EF	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
<b>Universo 1</b>	<b>3.102.847</b>	<b>100,00</b>	<b>166.016</b>	<b>100,00</b>	<b>103.349</b>	<b>100,00</b>
Escolas Multisseriada/Aceleração/Língua Indígena	175.087	5,64	35.269	21,24	32.988	31,92
Universo 1A	2.927.760	94,36	130.747	78,76	70.361	68,08
Alunos Multisseriada/Aceleração	16.211	0,52	2.787	1,68	1.943	1,88
Universo 1B	2.911.549	93,83	127.960	77,08	70.361	68,08
Escolas com menos de 10 alunos	43.556	1,40	7.195	4,33	7.132	6,90
<b>Universo 2</b>	<b>2.867.993</b>	<b>92,43</b>	<b>120.765</b>	<b>72,74</b>	<b>63.229</b>	<b>61,18</b>
Universo 2A: Escolas Particulares (Universo amostral)	471.666	15,20	22.540	13,58	14.630	14,16
Universo 2B: Escolas Federais	1.579	0,05	62	0,04	23	0,02
Universo 2C: Escolas Públicas (estaduais e municipais)	2.394.748	77,18	98.163	59,13	48.576	47,00

Universo 1 - Alunos de 5º ano (4ª série) do EF

Universo 1A - Universo 1 excluindo escolas que têm somente turmas multisseriadas/de aceleração, sem declarar turmas regulares de 5º ano (4ª série) do EF

Universo 1B - Universo 1A excluindo alunos de turmas multisseriadas/de aceleração

Universo 2 - Alunos de 5º ano (4ª série) do EF de Ensino Regular em escolas com 10 ou mais alunos matriculados

Para o 5º ano do EF:

FK\_COD\_ETAPA\_ENSINO = 7 (Ensino Fundamental de 8 anos – 4ª série)

FK\_COD\_ETAPA\_ENSINO = 18 (Ensino Fundamental de 9 anos – 5º ano)

FK\_COD\_ETAPA\_ENSINO=12

FK\_COD\_ETAPA\_ENSINO=13

FK\_COD\_ETAPA\_ENSINO=22

FK\_COD\_ETAPA\_ENSINO=23

Os códigos 12, 13, 22, 23 são alunos de correção de fluxo ou multisseriada em escolas de Ensino Fundamental de 8 ano ou de 9 anos.

**Tabela A.9º Ano**

9º ano (8ª série) – EF		Alunos		Turmas		Escolas	
		Número	%	Número	%	Número	%
<b>Universo 1</b>		<b>2.812.998</b>	<b>100,00</b>	<b>110.129</b>	<b>100,00</b>	<b>56.543</b>	<b>100,00</b>
	Escolas Multisseriada/Aceleração/Língua Indígena	25.155	0,89	4.230	3,84	3.771	6,67
Universo 1A		2.787.843	99,11	105.899	96,16	52.772	93,33
	Alunos Multisseriada/Aceleração	5.001	0,18	789	0,72	592	1,05
Universo 1B		2.782.842	98,93	105.110	95,44	52.772	93,33
	Escolas com menos de 10 alunos	26.334	0,94	3.976	3,61	3.957	7,00
<b>Universo 2</b>		<b>2.756.508</b>	<b>97,99</b>	<b>101.134</b>	<b>91,83</b>	<b>48.815</b>	<b>86,33</b>
Universo 2A: Escolas Particulares (Universo amostral)		411.779	14,64	16.115	14,63	10.531	18,62
Universo 2B: Escolas Federais		4.180	0,15	159	0,14	36	0,06
Universo 2C: Escolas Públicas (estaduais e municipais)		2.340.549	83,20	84.860	77,06	38.248	67,64

Universo 1 - Alunos de 9º ano (8ª série) do EF

Universo 1A - Universo 1 excluindo escolas que têm somente turmas multisseriadas/de aceleração, sem declarar turmas regulares de 9º ano (8ª série) EF

Universo 1B - Universo 1A excluindo alunos de turmas multisseriadas/de aceleração

Universo 2 - Alunos de 9º ano (8ª série) do EF de Ensino Regular em escolas com 10 ou mais alunos matriculados

Para o 9º ano do EF:

FK\_COD\_ETAPA\_ENSINO = 11 (Ensino Fundamental de 8 anos – 8ª série)

FK\_COD\_ETAPA\_ENSINO = 41 (Ensino Fundamental de 9 anos – 9º ano)

FK\_COD\_ETAPA\_ENSINO=12

FK\_COD\_ETAPA\_ENSINO=13

FK\_COD\_ETAPA\_ENSINO=22

FK\_COD\_ETAPA\_ENSINO=23

Os códigos 12, 13, 22, 23 são alunos de correção de fluxo ou multisseriada em escolas de Ensino Fundamental de 8 ano ou de 9 anos.

**Tabela A.3ª série EM**

3ª série (4ª série) – EM	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
<b>Universo 1</b>	<b>2.152.525</b>	<b>100,00</b>	<b>72.174</b>	<b>100,00</b>	<b>25.893</b>	<b>100,00</b>
Língua Indígena	386	0,02	30	0,04	24	0,09
Universo 1A	2.152.139	99,98	72.144	99,96	25.869	99,91
Escolas com menos de 10 alunos	6.865	0,32	1.056	1,46	1.041	4,02
<b>Universo 2</b>	<b>2.145.274</b>	<b>99,66</b>	<b>71.088</b>	<b>98,50</b>	<b>24.828</b>	<b>95,89</b>
Universo 2A: Escolas Particulares (Universo amostral)	308.487	14,33	10.314	14,29	6.910	26,69
Universo 2B: Escolas Federais	6.063	0,28	227	0,31	58	0,22
Universo 2C: Escolas Públicas (estaduais e municipais)	1.830.724	85,05	60.547	83,89	17.860	68,98

Universo 1 - Alunos de 3ª série (4ª série) do EM de Ensino Regular

Universo 2 - Alunos de 3ª série (4ª série) do EM de Ensino Regular em escolas com 10 ou mais alunos matriculados

Para a 3ª série do EM:

FK\_COD\_ETAPA\_ENSINO = 27 (Ensino Médio – 3ª série)

FK\_COD\_ETAPA\_ENSINO = 28 (Ensino Médio – 4ª série)



## ANEXO B

### 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  $T_i$  o tamanho  $t$  da unidade  $i$  do cadastro, calcule os valores de

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

para todo  $i=1,2,\dots,N$ . 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  $Z_i$ .