



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

**SAEB 2015**

**RELATÓRIO DA AMOSTRAGEM**  
**DA ANEB 2015**

**Dezembro de 2017**

---

## **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 2015 .....</b>	<b>6</b>
1. Objeto do Levantamento .....	6
2. Sistema de Referência .....	7
3. População de Referência da ANEB 2015 .....	7
<b>CAPÍTULO 2 - PLANO AMOSTRAL .....</b>	<b>9</b>
1. Estratos de Interesse .....	9
2. Plano Amostral .....	11
3. Seleção das turmas nas escolas selecionadas .....	15
<b>CAPÍTULO 3 – EXPANSÃO DA AMOSTRA .....</b>	<b>16</b>
1. Esquema geral de ponderação .....	16
2. Descrição da ponderação da amostra.....	19
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 2015. 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 2015 tem um desenho muito parecido com o adotado nas edições do Saeb de 1999, 2001, 2003, 2005, 2009, 2011 e 2013. 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, não foi repetido o estudo das escolas e turmas que possuem a 3ª série do ensino médio integrado.

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 2015

Este capítulo descreve a população (ou universo) de referência da Aneb 2015, 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 privadas, 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 2013:

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

A Aneb 2015 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 2014, realizado pelo MEC/INEP. Valeu-se principalmente de informações das escolas que possuíam, em 2014, alunos nas séries mencionadas. Assim, operacionalmente, a população investigada será formada pelos alunos de 2015, pertencentes a escolas ativas no Censo Escolar de 2014. Vale notar que as escolas participam da definição da população de referência com as características existentes no ano de 2014, 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 2014 e com alunos nas séries de interesse não farão parte da população de referência da Aneb 2015.

## 3. População de Referência da ANEB 2015

Para a edição de 2015, 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 2015 nas escolas do Censo Escolar de 2014 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 não ter sido repetido o estudo das escolas e turmas que possuem a 3ª série do ensino médio integrado, ao manter os mesmos padrões investigados, 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 2014)

Série	Alunos	Turmas	Escolas
5º Ano EF	2.976.905	124.035	64.332
9º Ano EF	2.898.214	104.093	48.393
3ª Série EM	2.117.852	69.873	24.369
<b>Total</b>	7.992.971	298.001	137.094

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 2014 era composto por 3.235.330 alunos, distribuídos em 173.791 turmas e 107.68 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 2.976.905 alunos, 124.035 turmas regulares e 64.332 escolas. Este procedimento elimina 7,99% de alunos do universo inicial, porcentagem bem menor que a de escolas eliminadas (40,26%).

Verificamos ainda que o número médio de alunos de 5º ano por turmas multisseriadas, ou de aceleração, ou indígena é de apenas 5,01, 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 43.625 alunos em 7.174 escolas, um impacto de apenas 1,35% do total de alunos e 6,66% do total de escolas, em relação ao Universo 1. Portanto, a linha denominada Universo 2 indica o universo que seria investigado em 2014 com os critérios definidos acima para o Aneb 2015.

Assim, no 5º ano do ensino fundamental, o universo original de aluno reduz-se para 92,01% do número inicial, enquanto que o número de escolas reduz-se para um pouco menos da metade, isto é, para 59,74%. 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 2015, em que não houve mudanças em relação à edição anterior, além das já mencionadas a respeito do Ensino Médio Integrado.

### **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 5º ano e 9º ano de escolas públicas com pelo menos 20 alunos matriculados nesses anos/séries participam da Anresc (Prova Brasil), que tem caráter censitário, para estas populações, somente necessitam ser amostrados alunos de escolas públicas com 10 a 19 alunos.

O segundo critério utilizado é o da dependência administrativa da escola: federal, pública (estadual e municipal) e particular. Como as escolas federais são avaliadas de forma censitária e mantidas em estratos separados das demais dependência administrativa de ensino, utilizaram-se, em uma primeira instância, apenas as categorias Pública (estadual e municipal) e Particular. Assim, trabalhou-se com os seguintes grandes estratos:

- 1) 5º ano do ensino fundamental regular pública (F5/4Pub);
- 2) 5º ano do ensino fundamental regular particular (F5/4Par);
- 3) 9º ano do ensino fundamental regular pública (F9/8Pub);
- 4) 9º ano do ensino fundamental regular particular (F9/8Par);
- 5) 3ª do ensino médio regular pública (M3RPub);
- 6) 3ª do ensino médio regular particular (M3RPar).

Os critérios que virão a seguir são aplicados dentro de cada um dos seis 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. Para os estratos F5/4Pub e F9/8Pub, o critério da área (capital ou interior) foi desconsiderado, por se entender que as estimativas deste estrato já estão atendidas na Anresc (Prova Brasil).

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/ANEB2015

<b>Critério</b>	<b>Quantidade</b>	<b>Descrição</b>
<b>F5/4Pub</b>		<b>Escolas com 10 a 19 alunos na série</b>
UF	27	26 estados+ DF
Localização	2	Urbana ou Rural
<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>F9/8Pub</b>		<b>Escolas com 10 a 19 alunos na série</b>
UF	27	26 estados+ DF
Localização	2	Urbana ou Rural
<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>M3RPub</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 540 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 131 estratos no 5º ano do Ensino Fundamental, 132 estratos no 9º ano do Ensino Fundamental e 161 estratos na 3ª série do Ensino Médio Regular, totalizando 424 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 424 estratos de interesse, não estão contempladas as escolas que fazem parte da Anresc (Prova Brasil) nem as escolas federais, pelo fato de sua aplicação ter caráter censitário.

## **2. Plano Amostral**

O planejamento amostral da Aneb 2015 acompanha bem 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 seis subpopulações fundamentais (que poderiam ser vistas como seis estratos), segundo a série em que estão matriculados e a dependência administrativa da escola. 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 2014, foi adotado um plano amostral em duas etapas, para seleção dos alunos a serem testados como parte da Aneb 2015. Na primeira etapa do plano amostral, foram selecionadas escolas que contenham alunos de cada uma das séries consideradas na Aneb 2015, 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.

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 seis sub-populações distintas:*
  - escolas públicas oferecendo 5º ano do ensino fundamental (F5/4Pub);
  - escolas privadas oferecendo 5º ano do ensino fundamental (F5/4Par);
  - escolas públicas oferecendo 9º ano do ensino fundamental (F9/8Pub);
  - escolas privadas oferecendo 9º ano do ensino fundamental (F9/8Par);
  - escolas públicas oferecendo 3ª série do ensino médio regular (M3RPub);

- escolas privadas 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 2014 foram estratificadas nos 424 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 na Aneb (Saeb) 2013, 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 2014 e na Anresc (Prova Brasil) 2013. As informações da média e da variância, por UF, foram obtidas da Prova Brasil 2013 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 2013, 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 de 5º ano e 9º ano do Ensino Fundamental com pelo menos 20 alunos, a amostra de turmas de escolas privadas 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.

Para conseguir melhor distribuição das escolas, foi alocado o número de turmas proporcional ao tamanho na dependência administrativa, garantindo uma diferença da amostra entre as dependências administrativas de, no máximo, 20%. Se a diferença fosse maior que 20%, seria

fixada a proporção 60% para a dependência com maior quantidade de turmas e 40% para a dependência com menor quantidade de turmas. Este procedimento foi aplicado nos estratos M3RPub e M3RPar, enquanto para os estratos F5/4Par e F9/8Par, já que as escolas públicas fazem parte da Anresc (Prova Brasil), foi necessário amostrar somente escolas privadas, com tamanho da amostra igual a 40% do total.

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) dentro de cada Dependência Administrativa 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.

Ressalta-se que essa metodologia de seleção das escolas não foi realizada no estrato F5/4Pub e nem no estrato F9/8Pub, pois trata-se de escolas com 10 a 19 alunos na série onde o interesse é completar a informação para as escolas desses estratos que já estão participando da Anresc (Prova Brasil). Com isso, o critério adotado para a seleção das escolas foi o de selecionar, em cada UF, quatro escolas na localização urbana e seis escolas na localização rural. Consequentemente, para o 5º ano e 9º ano a metodologia de seleção das escolas, descrita anteriormente, foi usada somente para escolas privadas.

A Tabela 2 apresenta o dimensionamento da amostra para os seis grandes estratos de interesse.

**Tabela 2.** Alocação esperada da amostra para a Aneb 2015

<b>Universo</b>	<b>F5/4Pub</b>	<b>F5/4Par</b>	<b>F9/8Pub</b>	<b>F9/8Par</b>	<b>M3RPub</b>	<b>M3RPar</b>	<b>Total</b>
Alunos na população	148.343	475.908	105.084	430.343	1.803.100	308.766	3.271.544
Estratos de interesse	54	77	54	78	88	73	424
Alocação da amostra de escolas	267	1.318	257	911	1.273	1.018	5.044
Nº de turmas nas escolas sorteadas (censo)	270	2.077	260	1.472	4.523	1.597	10.199
Alocação da amostra de turmas	270	1.562	260	1.122	2.093	1.259	6.566
Média de alunos por turma das escolas sorteadas	14,6	21,6	14,2	26,9	29,3	29,6	25,9
Alocação da amostra de alunos	3.929	33.701	3.684	30.210	61.236	37.281	170.041
Fração amostral prevista (alunos)	2,65	7,08	3,51	7,02	3,40	12,07	5,20

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) 2013, usando o mesmo critério de geração de números aleatórios que foi usado naquela edição do Saeb.

Para seqüência 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 2014

Sub-população	Escolas	Turmas Esperadas	Municípios
F5/4Pub	267	270	237
F5/4Par	1.318	1.562	531
F9/8Pub	257	260	237
F9/8Par	911	1.122	398
M3RPub	1.273	2.093	792
M3RPar	1018	1259	412
Total (1)	3.890	6.566	1.453

(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 2015.

### 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 2014, já a seleção das turmas foi realizada utilizando o censo escolar 2015, pois é efetuada em um período posterior.

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. Os resultados atualizados são apresentados na Tabela 4.

**Tabela 4.** Alocação da amostra após o ajuste das escolas para o ANEB 2015

Universo	F5/4Pub	F5/4Par	F9/8Pub	F9/8Par	M3RPub	M3RPar	Total
Alunos na população	147026	473.665	106312	400.593	1.733.602	297.451	3.158.649
Estratos de interesse	54	76	53	79	88	70	420
Alocação da amostra de escolas	148	1.319	195	920	1.322	1.020	3.788
Nº de turmas nas escolas sorteadas (censo)	149	2.477	196	1.741	5.935	1.866	12.364
Alocação da amostra de turmas	148	1.561	195	1.100	2.163	1.204	6.371
Média de alunos por turma das escolas sorteadas	16,7	24,5	16,6	29,1	30,5	34,1	28,7
Alocação da amostra de alunos	2.466	38.185	3.241	31.963	66.020	41.033	182.908
Fração amostral prevista (alunos)	0,02	0,08	0,03	0,08	0,04	0,14	5,79

Ressalta-se que todas as escolas participantes da Aneb 2015 estão no censo escolar 2014. Essa condição é importante para o cálculo dos pesos dos alunos ao final da aplicação da Aneb 2015.

## **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 2015), 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 e no SAEB 2013.

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 2015.

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 2015; a seção 3 trata dos métodos adotados para a ponderação da parte censitária da Anresc (Prova Brasil) 2015.

### **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 2015 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 2015, 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 2015. 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 2015 adotou-se os mesmos princípios das edições anteriores do SAEB (1997, 1999, 2001, 2003, 2005, 2009, 2011 e 2013), 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 2015, após aplicação de exclusões equivalentes às efetuadas na montagem do cadastro de seleção da amostra, baseado no Censo Escolar de 2014. 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 2015 se ajustem aos últimos dados divulgados do Censo Escolar 2015, 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 2015 não é capaz de captar com precisão. Por exemplo, nos totais deste cadastro estão incluídas escolas e turmas novas, criadas após o Censo Escolar 2014, que não poderiam participar da seleção da amostra nem tampouco do levantamento de dados na ANEB 2015. Portanto, vale reforçar que a amostra prevista é baseada no Censo Escolar 2014, 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 2015 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 2015 e das escolas federais, 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 2015, 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 2015 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.

A ANEB 2015 manteve a amostragem de escolas públicas de 5º Ano e 9º Ano com 10 a 19 alunos em turmas regulares tendo como estrato a Unidade da Federação e a Localização (Urbana e Rural) e com o peso dos alunos sendo expandido para este estrato. Para os outros universos da ANEB 2015 (escolas particulares de 5º ano e 9º ano do Ensino Fundamental e escolas públicas ou 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 expansão do universo da ANRESC 2015 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 2015.

Os universos de expansão do Censo Escolar 2015, que esse trabalho tomou como referência, estão apresentados nas tabelas abaixo.

**Tabela 5.** Dados do Censo Escolar 2015 do 5º Ano para expansão

5º Ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.917.391	100,00	122.433	100,00	63.720	100,00
Universo 2A: Escolas ANEB Privadas	477.628	16,37	22.481	18,36	14.532	22,80
Universo 2B: Escolas Federais	1.473	0,05	61	0,04	23	0,03
Universo 2C: Escolas ANRESC	2.290.802	78,52	89.789	77,33	39.137	61,42
Universo 2D: Escolas ANEB Públicas	147.488	5,05	10.102	8,25	10.028	15,73

**Tabela 6.** Dados do Censo Escolar 2015 do 9º Ano para expansão

9º Ano	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.753.800	100,00	100.905	100,00	48.241	100,00
Universo 2A: Escolas ANEB Privadas	402.642	14,62	15.587	15,45	10.079	20,89
Universo 2B: Escolas Federais	4.325	0,16	175	0,17	35	0,07
Universo 2C: Escolas ANRESC	2.239.922	81,34	77.699	77,00	30.733	63,71
Universo 2D: Escolas ANEB Públicas	106.911	3,88	7.444	7,38	7.394	15,33

**Tabela 7.** Dados do Censo Escolar 2013 do 3ª série para expansão

3ª série	Alunos		Turmas		Escolas	
	Número	%	Número	%	Número	%
Universo 2	2.057.989	100,00	69.645	100,00	24.532	100,00
Universo 2A: Escolas ANEB	2.051.309	99,68	69.389	99,63	24.470	99,75
Universo 2B: Escolas Federais	6.680	0,32	256	0,37	62	0,25

## 2. Descrição da ponderação da amostra

**(ANEB: 5º e 9º Ano Particular, 5º e 9º Ano Pública com 10 a 19 alunos na escola, 3ª Série Regular)**

O cálculo da ponderação da amostra principal da ANEB 2015 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 para o 5º ano e 9º ano particular e 3ª série, 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 o 5º ano e 9º ano públicas com 10 a 19 alunos na escola os estratos foram definidos para unidade da federação e localização (urbana e rural). 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 2015. 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 2015 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:

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

Observe 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

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

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 iguais a zero, pois se não houve alunos testados na escola, não pode haver alunos testados nas turmas dessa escola. Note também que, conforme o plano amostral adotado,  $n_i$  pode tomar apenas dois valores: 1 se a escola tiver até duas turmas ( $T'_i \leq 2$ ) e 2 se a escola tiver três ou mais turmas ( $T'_i > 2$ ).

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 2015 não estavam contemplados no Censo Escolar 2014, 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 2015 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:

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

Onde:

$Alunos_{turma/censo}$  é o número total de alunos na turma declarado no Censo Escolar 2015;

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

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

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

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

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

Onde:

$Alunos_{escola/censo}$  é o número total de alunos na escola declarado no Censo Escolar 2015;

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

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

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

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

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

Onde:

$Alunos_{estrato/censo}$  é o número total de alunos no estrato declarado no Censo Escolar 2015;

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

Como a soma dos pesos não soma a quantidade de alunos do Censo Escolar 2015, 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:

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

$$Peso_{Mun\_Dep} = Peso_{estrato} * Peso_{aux3}$$

Expansão para o Município:

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

$$Peso_{Mun} = Peso_{Mun\_Dep} * Peso_{aux4}$$

Expansão para a Unidade da Federação (UF):

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

$$Peso\_Aluno = Peso_{Mun} * Peso_{aux5}$$

Ao final dos passos descritos acima, o aluno receberá um peso dado por  $Peso\_Aluno$ .

Salienta-se que alguns alunos ficaram com peso menor do que 1. Isso ocorre principalmente 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 2015 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 2015**

#### **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	%
Universo 1	3.235.330	100	173.791	100	107.680	100
Escolas Multisseriada/Aceleração/Língua Indígena	195.276	6,04	39.005	22,44	36.174	33,59
Universo 1A	3.040.054	93,96	134.786	77,56	71.506	66,41
Alunos Multisseriada/Aceleração	19.524	0,60	3.495	2,01	0	0,00
Universo 1B	3.020.530	93,36	131.291	75,55	71.506	66,41
Escolas com menos de 10 alunos	43.625	1,35	7.256	4,18	7.174	6,66
Universo 2	2.976.905	92,01	124.035	71,37	64.332	59,74
Universo 2A: Escolas ANEB (Universo amostral)	475.908	15,99	22.385	18,05	14.567	22,64
Universo 2B: Escolas Federais	1.521	0,05	61	0,05	22	0,03
Universo 2C: Escolas ANRESC (Prova Brasil)	2.351.133	78,98	91.388	73,68	39.633	61,61
Universo 2D: Escolas Públicas com 10 a 19 alunos	148.343	4,98	10.201	8,22	10.110	15,72

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	%
Universo 1	2.951.312	100	112.723	100	55.787	100
Escolas Multisseriada/Aceleração/Língua Indígena	20.560	0,70	3.693	3,28	3.317	5,95
Universo 1A	2.930.752	99,30	109.030	96,72	52.470	94,05
Alunos Multisseriada/Aceleração	5.611	0,19	846	0,75	0	0,00
Universo 1B	2.925.141	99,11	108.184	95,97	52.470	94,07
Escolas com menos de 10 alunos	26.927	0,91	4.091	3,63	4.077	7,31
Universo 2	2.898.214	98,20	104.093	92,34	48.393	86,75
Universo 2A: Escolas ANEB (Universo amostral)	430.343	14,85	16.264	15,62	10.294	21,27
Universo 2B: Escolas Federais	4.506	0,16	164	0,16	37	0,08
Universo 2C: Escolas ANRESC (Prova Brasil)	2.358.281	81,37	80.326	77,17	30.793	63,63
Universo 2D: Escolas Públicas com 10 a 19 alunos	105.084	3,63	7.339	7,05	7.269	15,02

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	%
Universo 1	2.124.920	100	70.939	100	25.418	100
Escolas Língua Indígena	300	0,01	26	0,04	23	0,09
Universo 1A	2.124.620	99,99	70.913	99,96	25.395	99,91
Escolas com menos de 10 alunos	6.768	0,32	1.040	1,47	1.026	4,04
Universo 2	2.117.852	99,67	69.873	98,50	24.369	95,87
Universo 2A: Escolas ANEB (Ensino Regular)	2.111.866	99,72	69.662	99,70	24.316	99,78
Universo 2B: Escolas Federais	5.986	0,28	211	0,30	53	0,22

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  $Z_i$ .