

RELATÓRIO METODOLÓGICO TIC EDUCAÇÃO

INTRODUÇÃO

O Comitê Gestor da Internet no Brasil (CGI.br), por meio do Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br), do Núcleo de Informação e Coordenação do Ponto BR (NIC.br), apresenta a metodologia da Pesquisa sobre o uso das tecnologias de informação e comunicação nas escolas brasileiras – TIC Educação.

A pesquisa com escolas de áreas urbanas foi desenhada tendo como referencial metodológico o trabalho realizado pela International Association for the Evaluation of Educational Achievement – IEA (2009), divulgado em duas publicações: *Sites 2006 (Technical Report – Second Information Technology in Education Study)* e *Sites 2006 (User Guide for the International Database)*. Sua primeira edição ocorreu em 2010 e, ao longo de sua trajetória, aspectos pertinentes à metodologia e ao questionário foram adaptados a fim de atender às especificidades do universo escolar do Brasil e às necessidades dos diferentes setores da sociedade, como governo, academia, organizações da sociedade civil e setor privado.

A coleta de dados em escolas urbanas e rurais é realizada por meio de métodos diferentes, sendo as urbanas abordadas presencialmente e as rurais, por telefone. A seguir, são detalhados os aspectos metodológicos para ambos os âmbitos da pesquisa.

OBJETIVOS DA PESQUISA

O objetivo da pesquisa TIC Educação é identificar o acesso, o uso e a apropriação das tecnologias de informação e comunicação em escolas brasileiras, tanto no que diz respeito à prática pedagógica quanto à gestão escolar.

METODOLOGIA DA PESQUISA EM ESCOLAS URBANAS

CONCEITOS E DEFINIÇÕES

POPULAÇÃO-ALVO

A população-alvo do estudo é composta pelas escolas públicas (estaduais e municipais) e particulares em atividade, localizadas em áreas urbanas do Brasil e que oferecem ensino na modalidade regular em pelo menos um dos níveis de ensino e séries. Três séries são investigadas: 4ª série / 5º ano do Ensino Fundamental I, 8ª série / 9º ano do Ensino Fundamental II e 2º ano do Ensino Médio.

Também fazem parte da população-alvo todos os diretores das escolas, bem como os alunos matriculados e os profissionais que exercem as funções de coordenação pedagógica (coordenadores pedagógicos) e de docência (professores) relacionados com as turmas das séries e dos níveis de ensino considerados na pesquisa.

UNIDADES DE ANÁLISE E REFERÊNCIA

Para atingir o objetivo da pesquisa, são investigadas várias dimensões relacionadas às unidades de referência e análise. São elas:

- **Escolas localizadas em áreas urbanas:** perfil em termos de infraestrutura e práticas envolvendo TIC;
- **Diretores de escolas localizadas em áreas urbanas:** perfil de uso do computador e da Internet; uso das TIC nas atividades administrativas e de gestão; interação com a comunidade e percepção sobre as limitações para a integração das TIC à educação;
- **Coordenadores pedagógicos (pessoas em funções de coordenação pedagógica) de escolas localizadas em áreas urbanas:** perfil de uso do computador e da Internet; uso das TIC nas atividades administrativas, de coordenação pedagógica e no projeto pedagógico da escola; e percepção sobre as limitações para a integração das TIC à educação;
- **Professores (pessoas em funções docentes) de escolas localizadas em áreas urbanas:** perfil profissional e de uso do computador e da Internet; habilidades e capacitação para o uso das TIC; uso das TIC nas atividades gerais e de ensino-aprendizagem; uso de conteúdos educacionais digitais e percepção sobre as limitações para a integração das TIC à educação;
- **Alunos de escolas localizadas em áreas urbanas:** perfil de uso do computador e da Internet; habilidades para o uso das TIC; capacitação para o uso das TIC; atividades realizadas com o uso de computador e/ou Internet na escola.

DOMÍNIOS DE INTERESSE PARA ANÁLISE E DIVULGAÇÃO

Para as unidades de análise e referência, os resultados são divulgados para domínios definidos com base nas variáveis e nos níveis descritos a seguir.

Para todas as unidades de análise:

- **Região:** corresponde à divisão regional do Brasil, segundo critérios do Instituto Brasileiro de Geografia e Estatística (IBGE), nas macrorregiões Centro-Oeste, Nordeste, Norte, Sudeste e Sul;
- **Dependência administrativa:** corresponde ao tipo de subordinação administrativa da escola – pública municipal, pública estadual ou particular.

Com exceção da escola:

- **Sexo:** corresponde à divisão em feminino ou masculino.

Para alunos e professores:

- **Série:** corresponde ao nível de ensino do processo educacional em que o respondente leciona ou estuda.

Para diretores, coordenadores pedagógicos e professores:

- **Faixa etária:** corresponde às faixas de idade em anos completos do respondente no dia da entrevista, expressa em anos completos;
- **Renda familiar:** corresponde à soma da renda mensal de todos os membros do domicílio do respondente expressa em três faixas de múltiplos do SM, considerando o SM vigente definido pelo Ministério do Trabalho e do Emprego no mês anterior ao dia da entrevista;
- **Renda pessoal:** corresponde à soma da renda mensal total do respondente expressa em três faixas de múltiplos do SM, considerando o SM vigente definido pelo Ministério do Trabalho e do Emprego no mês anterior ao dia da entrevista.

INSTRUMENTOS DE COLETA

INFORMAÇÕES SOBRE OS INSTRUMENTOS DE COLETA

As entrevistas em escolas urbanas são realizadas com questionários estruturados específicos para os públicos abordados na pesquisa: alunos, professores, coordenadores pedagógicos e diretores. O questionário da unidade de análise Escolas é respondido pelos diretores. Mais informações sobre os instrumentos de coleta podem ser encontradas no Relatório de Coleta de Dados da pesquisa.

PLANO AMOSTRAL

A amostra de escolas urbanas é probabilística e selecionada em múltiplos estágios, que dependem do público-alvo da pesquisa. O primeiro estágio de seleção da amostra de escolas é realizado a partir da estratificação do universo da pesquisa segundo capitais e grandes regiões brasileiras. Em cada estrato, foram selecionadas localidades que consistem em aglomerados de municípios vizinhos segundo características de número de turmas (a serem descritas em detalhes a seguir).

Para as capitais e localidades selecionadas, as escolas são divididas segundo os anos de ensino que estão disponíveis: 5^o ou 9^o anos do Ensino Fundamental, ou 2^o ano do Ensino Médio, construindo três listas para a seleção de escolas – as escolas-série. Sendo assim, uma escola que tenha turmas nos 5^o e 9^o anos do Ensino Fundamental é incluída no cadastro para seleção da amostra do 5^o ano e do 9^o ano. Isso significa que as escolas que possuem turmas em mais de uma série de interesse participam mais de uma vez do processo de amostragem. Em cada uma das listas, a seleção da amostra de escolas-série é realizada de forma independente – 2^o estágio de seleção. Esse estágio traz para a pesquisa as escolas-série que fazem parte da amostra e, a partir dessa amostra, são atribuídos os respondentes para as unidades de análise Escolas e Diretores.

Nas escolas-série selecionadas, são realizadas as listagens de turmas, coordenadores pedagógicos e professores de Língua Portuguesa, Matemática ou de anos iniciais de Ensino Fundamental (para as turmas de 5^o ano). A partir dessas listas, são selecionados os coordenadores pedagógicos e professores para responder à pesquisa – 3^o estágio de seleção.

Para as turmas selecionadas no terceiro estágio, é realizada a listagem de alunos matriculados para a seleção de alunos que devem responder à pesquisa – 4^o estágio de seleção. Essa é a última etapa para a construção da amostra que atende a todos os universos de interesse da pesquisa.

CADASTRO E FONTES DE INFORMAÇÃO

O cadastro utilizado para seleção das escolas-série é o Censo Escolar da Educação Básica, coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Esse cadastro contém dados referentes a todas as escolas de Educação Básica no Brasil. A partir do cadastro mais recente de escolas do Inep (divulgado em março de cada ano), são separadas aquelas que satisfazem a todas as condições de elegibilidade para participar da população da pesquisa, ou seja, escolas em funcionamento, localizadas em área urbana, que não estejam em áreas diferenciadas e que possuem a população-alvo do estudo.

As escolas potencialmente elegíveis também têm seus cadastros de turmas analisados para permitir identificar quais poderiam fazer parte da amostra. Isso é necessário porque a cobertura da pesquisa se refere apenas às turmas de educação regular. Turmas de Educação Infantil, Educação Especial, Educação de Jovens e Adultos (EJA) e de Ensino Profissionalizante estão fora do escopo da pesquisa.

As escolas elegíveis que tenham sido criadas durante o ano em que se dá a coleta de dados não fazem parte da população da pesquisa. Para todas as demais unidades de referência (diretores,

coordenadores pedagógicos, professores e alunos), as condições de elegibilidade são aplicadas conforme a situação das escolas no ano da coleta, após atualização cadastral a ser realizada em cada estabelecimento escolar selecionado para a amostra.

CONSTRUÇÃO DAS LOCALIDADES

A criação das localidades como agregação de municípios vizinhos é realizada segundo a metodologia Skater¹ disponível no programa TerraView². Ao todo, o Brasil possui 5.570 municípios nas 27 unidades da federação (UF). Estes municípios são agregados (a seus vizinhos) segundo os seguintes critérios:

- Uma localidade deveria pertencer apenas a uma UF;
- O número médio de escolas em uma localidade não deveria diferir de forma relevante da média de escolas por município na UF;
- As capitais deveriam ficar isoladas e constituir um aglomerado próprio; e
- Os aglomerados resultantes deveriam ser heterogêneos internamente, isto é, contendo municípios com muitas escolas e municípios com poucas escolas.

O método é executado separadamente por UF, garantindo que não haja criação de aglomerados que ultrapasassem as fronteiras de duas ou mais regiões distintas.

De forma a obter heterogeneidade interna dos aglomerados, a variável de medida para a agregação de municípios é definida como:

$$A_{ij} = \left| E_{ij} - \sum_i \frac{E_{ij}}{M_j} \right|,$$

onde:

E_{ij} é o número de escolas do município i da UF j ; e

M_j é o número de municípios da UF j .

Para garantir que as capitais fiquem separadas em aglomerados únicos, utiliza-se na metodologia uma variável CAP definida como:

$$CAP = 1 \times (1 - I[capital]) + 100.000 \times I[capital].$$

¹ A metodologia Skater consiste na construção de aglomerados considerando a similaridade para medidas (estatísticas/dados) das unidades iniciais e a vizinhança geográfica dessas unidades (Assunção, Lage, & Reis, 2002).

² O software TerraView é um programa de informações e aplicações geográficas gratuito desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (Inpe).

Como resultado do método, a partir dos 5.570 municípios são construídas 2.253 localidades. Todas as capitais, exceto Macapá, constituem um aglomerado isolado. Os municípios de Macapá e Itaubal formam um aglomerado, uma vez que Itaubal está integralmente circunscrito ao município da capital amapaense. A Tabela 1 apresenta o número de localidades segundo as regiões brasileiras.

TABELA 1
NÚMERO DE LOCALIDADES, SEGUNDO REGIÃO

Região	Número de localidades
Norte	210
Nordeste	821
Sudeste	587
Sul	461
Centro-Oeste	174
Total	2 253

DIMENSIONAMENTO DA AMOSTRA

O tamanho da amostra da pesquisa TIC Educação em escolas urbanas tem por objetivo proporcionar uma melhor leitura dos resultados diante da heterogeneidade das unidades de referência e análise. As características de seleção da amostra, descritas em detalhes mais adiante, resultam em um tamanho de amostra variável.

ESTRATIFICAÇÃO DA AMOSTRA DE LOCALIDADES

A população-alvo da pesquisa é estratificada segundo municípios das capitais e grandes regiões brasileiras.

ALOCAÇÃO DA AMOSTRA POR GRANDES REGIÕES, CAPITAIS E UNIDADES DA FEDERAÇÃO

A estratégia geral de amostragem da pesquisa envolve a seleção de uma amostra de escolas em cada série de interesse. Para tanto, é considerada a seleção de localidades nos estratos e seleção de escolas-séries dentro das localidades selecionadas. A unidade primária de amostragem é a localidade nos estratos de regiões e as escolas-série nos estratos de capitais. A Tabela 2 apresenta a alocação da amostra por regiões:

TABELA 2
ALOCÇÃO DA AMOSTRA DE LOCALIDADES, SEGUNDO ESTRATOS

Estrato	Número de localidades
Norte	30
Nordeste	30
Sudeste	30
Sul	30
Centro-Oeste	30
Capitais	27
Total	177

A amostra de localidades nas regiões é estratificada por unidades federativas dentro das regiões, de forma a obter informação de escolas em todas as UF. A estratificação por grandes regiões, UF e capitais considerou:

- Capitais (um estrato em cada uma das 27 unidades da federação);
- Unidades da federação (27).

Ao todo, são 54 estratos finais. Todas as capitais são localidades presentes com certeza na amostra e 30 localidades são selecionadas aleatoriamente em cada grande região brasileira.

O tamanho da amostra de localidades por UF é determinado pela proporção de população da UF em relação à população da região, de tal forma que:

$$n_h = 30 \times \frac{P_{hl}}{\sum_l P_{hl}},$$

onde:

n_h é o tamanho da amostra de localidades no estrato h ; e

P_{hl} é a população da localidade l no estrato h .

A distribuição da amostra de localidades por UF é apresentada na Tabela 3.

TABELA 3
TAMANHO DA AMOSTRA DE LOCALIDADES, SEGUNDO UNIDADE DA FEDERAÇÃO

UF	Capital	Interior
Rondônia	1	2
Acre	1	2
Amazonas	1	6
Roraima	1	2
Pará	1	14
Amapá	1	2
Tocantins	1	2
Maranhão	1	3
Piauí	1	2
Ceará	1	4
Rio Grande do Norte	1	2
Paraíba	1	2
Pernambuco	1	5
Alagoas	1	2
Sergipe	1	2
Bahia	1	8
Minas Gerais	1	7
Espírito Santo	1	2
Rio de Janeiro	1	6
São Paulo	1	15
Paraná	1	11
Santa Catarina	1	7
Rio Grande do Sul	1	12
Mato Grosso do Sul	1	6
Mato Grosso	1	8
Goiás	1	16
Distrito Federal	1	0
Total	27	150

Após a seleção das localidades, busca-se no Censo Escolar todas as escolas nos municípios que compõem a amostra (inclusive as capitais). A partir dessa base, as escolas são divididas em três grupos distintos para seleção de escolas-série (segundo estágio): escolas com 5º ano do Ensino Fundamental, escolas com 9º ano de Ensino Fundamental e escolas com 2º ano do Ensino Médio. Esses conjuntos de escolas separadas por série dão origem às populações de escolas-série. Dessa forma, uma escola que tenha turmas em mais de um nível de ensino pesquisado tem maior chance de participar da amostra final de escolas, pois participa mais de uma vez do processo de amostragem.

Uma vez selecionadas as amostras de escolas-série em cada localidade, são selecionadas amostras das demais unidades de referência e análise de interesse (diretores, alunos,

coordenadores pedagógicos e professores), isto é, o plano amostral é implementado seguindo as etapas para a seleção das unidades de referência.

Enquanto as escolas-série são selecionadas na segunda etapa, a seleção das demais unidades de referência considera cada unidade escola-série como um conglomerado, dentro do qual é feito o cadastramento em campo e a seleção das demais unidades de referência, conforme detalhado adiante.

Os estratos para a seleção das unidades escola-série são definidos considerando a estratificação por UF/capital e a identificação da localidade selecionada. Dessa forma, todas as localidades dentro de um estrato têm pelo menos uma escola-série selecionada para a pesquisa. Essa alocação garante que os domínios de análise região e nível de ensino estejam contemplados pela estratificação. A dependência administrativa é considerada no processo de seleção da amostra de escolas-série, conforme explicitado no tópico seleção de escolas-série.

O tamanho total da amostra de escolas-série varia de acordo com as localidades que são selecionadas no primeiro estágio da amostra. Usualmente, os tamanhos das amostras por escolas-série se aproximam de 500 escolas em cada nível de ensino (Ensino Fundamental I, Ensino Fundamental II ou Ensino Médio), procurando alocar 100 escolas-série por grande região brasileira.

SELEÇÃO DA AMOSTRA

PRIMEIRO ESTÁGIO: SELEÇÃO DE LOCALIDADES

A seleção das localidades em cada estrato é realizada por meio de amostragem aleatória simples. Conforme já mencionado, todas as capitais são localidades presentes com certeza na amostra, e 30 localidades são selecionadas aleatoriamente em cada grande região brasileira. A probabilidade de seleção de cada localidade é dada por:

$$P_{hl} = \begin{cases} \frac{n_h}{L_h} & , \text{ se não é município de capital,} \\ 1 & , \text{ se é município de capital,} \end{cases}$$

onde:

P_{hl} é a probabilidade de seleção da localidade l no estrato h ;

n_h é o tamanho da amostra de localidades no estrato h ; e

L_h é o total de localidades no estrato h .

SEGUNDO ESTÁGIO: SELEÇÃO DA AMOSTRA DE ESCOLAS

A partir da seleção de localidades, são separadas no Censo Escolar as escolas da população-alvo da pesquisa. Todas as localidades da amostra têm escolas selecionadas, garantindo

espalhamento e presença de amostra em todas as unidades da federação. Para cada localidade é alocada uma amostra de escolas-série proporcional ao número de unidades escolas-séries presentes na localidade. Observa-se a distribuição de 100 escolas-série a selecionar em cada grande região brasileira. A alocação é apresentada no Relatório de Coleta de Dados da pesquisa.

As escolas pertencentes à população-alvo da pesquisa e situadas na amostra de localidades selecionadas são divididas em três bases de escolas-série: 5ª ano ou 9ª ano do Ensino Fundamental e 2ª ano do Ensino Médio. Em cada uma dessas bases é feita uma seleção independente de escolas para participação na amostra. A seleção é feita utilizando o método de Amostragem Sequencial de Poisson – do inglês, *Sequential Poisson Sampling* (Ohlsson, 1998) – dentro de cada localidade para cada base de escola-série. A medida de tamanho m_{hld}^{es} utilizada é calculada padronizando-se por dependência administrativa. Para efetuar a padronização, utiliza-se a média de turmas por escola por tipo de dependência. Essa média é dada por:

$$m_{hld}^{es} = \frac{\sum_d T_{hld}^{es}}{\sum_d E_{hld}^{es}},$$

onde:

m_{hld}^{es} é a média do número de turmas da dependência d na localidade l do estrato h ;

T_{hld}^{es} é o número de turmas da dependência d na localidade l do estrato h ; e

E_{hld}^{es} é o número de escolas da dependência d na localidade l do estrato h .

A medida padronizada para cada escola na lista de escolas-série é dada por:

$$m_{hldk}^{es} = \frac{T_{hldk}^{es}}{m_{hld}^{es}},$$

onde:

T_{hldk}^{es} é o número de turmas da escola k da dependência d na localidade l no estrato h .

A probabilidade de seleção da escola k , na lista de escolas-série, é dada por:

$$p_{hldk}^{es} = n_{hl}^{es} \times \frac{m_{hldk}^{es}}{\sum_d \sum_k m_{hldk}^{es}},$$

onde:

p_{hldk}^{es} é a probabilidade de seleção da escola k da dependência d na localidade l no estrato h ; e

n_{hl}^{es} é o tamanho da amostra de escolas-série na localidade l no estrato h .

TERCEIRO ESTÁGIO: SELEÇÃO DA AMOSTRA DE TURMAS

A partir da amostra de escolas-série, o número de turmas existentes em cada escola-série é levantado por telefone ou presencialmente, por meio do preenchimento de um formulário de arrolamento. No dia de realização das entrevistas, o entrevistador confere as informações obtidas anteriormente e, caso haja diferença, é considerada a informação mais atualizada. Essas informações são utilizadas para o cálculo do peso das turmas.

Para as escolas com uma, duas ou três turmas na série de interesse, é selecionada apenas uma turma e, no caso de escolas com quatro ou mais turmas, são selecionadas duas delas. A seleção das turmas é realizada de forma aleatória em cada escola-série, considerando o número de turmas a serem selecionadas.

QUARTO ESTÁGIO: SELEÇÃO DOS RESPONDENTES

SELEÇÃO DA AMOSTRA DE ALUNOS

Fixou-se em dez o número de alunos a serem entrevistados em cada série e em cada uma das escolas. Nos casos em que a escola apresenta até três turmas para a série selecionada, os dez alunos são selecionados por amostragem inversa simples na turma selecionada. No caso de a escola ter mais de três turmas na série selecionada, a amostra de alunos é obtida mediante seleção de cinco estudantes por amostragem inversa em cada uma das duas turmas selecionadas. A seleção de alunos é realizada mediante um procedimento simples, que depende apenas da obtenção da lista de frequência com os nomes dos matriculados em cada turma junto à administração da escola ou com um professor.

A partir da lista de frequência de uma turma selecionada, os alunos matriculados são numerados de um até o número total de alunos na turma. Para cada uma, foi previamente gerada uma lista com faixas de seleção com números aleatórios, variando de um ao número total de alunos matriculados na turma. De posse dessa informação, o entrevistador percorre a lista na ordem previamente definida para selecionar os alunos até que sejam entrevistados cinco ou dez, conforme o caso para a turma em questão.

SELEÇÃO DA AMOSTRA DE PROFESSORES

Em cada turma selecionada são entrevistados os professores que ministram as disciplinas de Português e Matemática. Nas turmas de 4ª série / 5º ano, são entrevistados os professores dos anos iniciais que ministram as disciplinas básicas. Nas escolas selecionadas para a realização das entrevistas em duas turmas na mesma série, é entrevistado um professor de cada disciplina por turma, totalizando dois professores por turma. E nas escolas selecionadas para a realização de entrevistas em uma turma na mesma série, são entrevistados dois professores de cada disciplina na turma, totalizando quatro professores. Já no caso de a escola ser selecionada para a realização de entrevistas em duas ou três séries, os mesmos procedimentos descritos acima são utilizados para cada série selecionada. Os professores são listados de acordo com a disciplina e a turma que lecionam, sendo selecionados aleatoriamente.

SELEÇÃO DA AMOSTRA DE COORDENADORES PEDAGÓGICOS

É entrevistado um coordenador pedagógico por série. Nos casos em que exista mais de um coordenador pedagógico para a série selecionada, é realizada a listagem de coordenadores pedagógicos, e um deles é selecionado aleatoriamente.

SELEÇÃO DA AMOSTRA DE DIRETORES

Ainda que alguma escola seja selecionada para a realização de entrevistas em mais de uma série, apenas um diretor por escola é entrevistado.

COLETA DE DADOS EM CAMPO

CRITÉRIOS PARA COLETA DE DADOS

A coleta de dados é feita por meio da visita presencial às escolas selecionadas e de entrevistas conduzidas com os diretores, coordenadores pedagógicos, professores e alunos selecionados para a amostra. Na maior parte dos casos, é realizado um agendamento prévio por telefone com o diretor ou responsável para que a visita dos entrevistadores não interfira no cotidiano da escola. Além disso, busca-se agendar uma data de realização das entrevistas em que o diretor, o coordenador pedagógico e os professores selecionados estejam presentes. Nos casos de difícil contato por telefone, os entrevistadores vão pessoalmente às escolas e fazem o agendamento e os arrolamentos no próprio local, e em alguns casos, as entrevistas são realizadas no mesmo dia.

Desse modo, na data agendada, os entrevistadores são enviados à escola e realizam as entrevistas seguindo os procedimentos e os questionários estruturados para cada público.

Cabe destacar que a pesquisa conta com o apoio institucional do Ministério da Educação (MEC), do Conselho Nacional de Secretários de Educação (Consed) e da União Nacional dos Dirigentes Municipais de Educação (Undime), que encaminham ofícios às escolas selecionadas antes e durante a realização da coleta de dados, a fim de informá-las sobre a pesquisa e de solicitar o apoio dos responsáveis no sentido de autorizar a realização das entrevistas.

PROCESSAMENTO DOS DADOS

PROCEDIMENTOS DE PONDERAÇÃO

A ponderação da pesquisa parte do cálculo de pesos básicos derivados das probabilidades de seleção em cada estágio, sobre os quais são aplicadas correções de não resposta. Os pesos das escolas são calibrados para os totais conhecidos da população-alvo da pesquisa.

PESO DAS ESCOLAS

O peso básico para cada escola é calculado a partir do inverso da probabilidade de seleção de escolas que consideram duas etapas de seleção: seleção de localidades e seleção de escolas no cadastro de escolas-série em localidades selecionadas.

De acordo com o que foi descrito anteriormente:

$$q_{hldk}^{es} = p_{hl} \times p_{hldk}^{es},$$

onde:

q_{hldk}^{es} é a probabilidade de selecionar a escola k da dependência d na localidade l do estrato h .

Algumas escolas possuem até três séries de interesse da pesquisa. Sendo assim, elas têm uma probabilidade maior de serem selecionadas, já que estavam presentes em cada um dos cadastros de escolas-série. Dessa forma, a probabilidade de uma escola ser selecionada para participar da pesquisa independentemente da escola-série é dada por:

$$\begin{aligned} q_{hldk} &= q_{hldk}^{5^a} + q_{hldk}^{9^a} + q_{hldk}^{2^a} \\ &\quad - 2 \times (q_{hldk}^{5^a} \times q_{hldk}^{9^a}) - 2 \times (q_{hldk}^{5^a} \times q_{hldk}^{2^a}) - 2 \times (q_{hldk}^{9^a} \times q_{hldk}^{2^a}) \\ &\quad + (q_{hldk}^{5^a} \times q_{hldk}^{9^a} \times q_{hldk}^{2^a}). \end{aligned}$$

Como o peso é o inverso da probabilidade de inclusão da escola na amostra, tem-se que o peso básico da escola w_{hldk} é dado por:

$$w_{hldk} = \frac{1}{q_{hldk}}.$$

CORREÇÃO DE NÃO RESPOSTA

Para efeito de correção de não resposta (cada unidade de análise pode possuir um contingente diferente de escolas respondentes), é considerado o ajuste dentro de cada estrato através da fórmula:

$$w_{hldk}^* = w_{hldk} \times \frac{\sum_{k \in s} w_{hldk}}{\sum_{k \in r} w_{hldk}},$$

onde:

w_{hldk}^* é o peso com correção de não resposta da escola k da dependência d na localidade l do estrato h ;

s é o conjunto de escolas selecionadas da dependência d na localidade l do estrato h ; e

r é o conjunto de escolas respondentes da dependência d na localidade l do estrato h .

CALIBRAÇÃO

Os pesos para as escolas respondentes, em todas as unidades de análise (escolas, diretores, coordenadores pedagógicos, professores e alunos), recebem o ajuste de não resposta (o número de respondentes é distinto em cada base) e a calibração para os totais de escolas por unidade da federação, dependência administrativa e série. Os totais das variáveis de calibração são obtidos no cadastro do Censo Escolar para a população-alvo da pesquisa, de onde foram selecionadas as amostras. O método utilizado é o ajuste iterativo sobre marginais, também conhecido por pós-estratificação multivariada incompleta ou *raking*. O peso final das escolas é: w_{hldk}^{*C} .

PESO DOS DIRETORES

O peso da unidade de análise Diretor é exatamente o mesmo peso calculado para a escola correspondente, uma vez que, quando há resposta da escola, sempre há resposta do diretor.

Logo, o peso final dos diretores é dado por: w_{hldk}^{*C} .

PESO DOS COORDENADORES PEDAGÓGICOS

O peso do coordenador pedagógico é calculado em duas etapas. A primeira é o cálculo do peso da escola respondente para essa unidade de análise (como descrito anteriormente). A segunda é o cálculo do inverso da probabilidade de seleção de cada coordenador pedagógico, por meio da razão entre o número de coordenadores listados e o de respondentes. O peso final do coordenador pedagógico é obtido pelo produto desses dois pesos.

$$w_{hldk}^C = w_{hldk}^{*C} \times \frac{C_{hldk}}{C_{hldk}^r},$$

onde:

w_{hldk}^C é o peso final para coordenadores na escola k da dependência d na localidade l do estrato h ;

C_{hldk} é o número de coordenadores listados na escola k da dependência d na localidade l do estrato h ; e

C_{hldk}^r é o número de coordenadores respondentes na escola k da dependência d na localidade l do estrato h .

PESO DOS PROFESSORES

O peso final dos professores é calculado em três etapas. A primeira utiliza o peso básico calibrado para as escolas em que é realizada pelo menos uma entrevista com professores. A segunda é o peso da turma que é obtido pela razão entre o número de turmas existentes e respondentes em cada série para cada escola, corrigido pela razão entre o número de turmas existentes e respondentes no estrato. O peso final da turma é dado por:

$$w_{hldk}^{**T^a} = w_{hldk}^{*C} \times \frac{T_{hldk}^a}{T_{hldk}^{ra}} \times \frac{T_{hl}^a}{T_{hl}^{ra}},$$

onde:

$w_{hldk}^{**T^a}$ é o peso da turma T da série a da escola k da dependência d na localidade l do estrato h ;

T_{hldk}^a é o número de turmas existentes na série a na escola k da dependência d na localidade l do estrato h ;

T_{hldk}^{ra} é o número de turmas respondentes na série a na escola k da dependência d na localidade l do estrato h ;

T_{hl}^a é o número de turmas existentes na série a na localidade l do estrato h ; e

T_{hl}^{ra} é o número de turmas respondentes na série a na localidade l do estrato h .

Outra fonte levada em consideração para o cálculo do peso do professor é a disciplina lecionada. Em cada turma, são levantadas as informações do número de professores existentes para cada uma das disciplinas de interesse. Dessa forma, para cada disciplina e para cada turma, obtém-se a razão entre o número de professores existentes e o número de entrevistas realizadas. O produto dos três fatores (escola, turma e disciplina) resulta no peso final de cada um dos professores entrevistados, dado por:

$$w_{hldk}^{P^a} = w_{hldk}^{**P} \times w_{hldk}^{**T^a} \times \frac{P_{hldk}^{xa}}{P_{hldk}^{xra}},$$

onde:

$w_{hldk}^{P^a}$ é o peso final do professor p da série a da escola k da dependência d na localidade l do estrato h ;

w_{hldk}^{**P} é o peso corrigido para escolas com professores respondentes da escola k da dependência d na localidade l do estrato h ;

$w_{hldk}^{**T^a}$ é o peso da turma T da série a da escola k da dependência d na localidade l do estrato h ;

P_{hldk}^{xa} é o número de funções docentes existentes da disciplina x na série a na escola k da dependência d na localidade l do estrato h ; e

P_{hldk}^{xra} é o número de funções docentes respondentes da disciplina x na série a na escola k da dependência d na localidade l do estrato h .

PESO DOS ALUNOS

Para o cálculo do peso final dos alunos, é estabelecido o produto entre os pesos das escolas que tiveram alunos entrevistados, o peso da turma e o peso básico dos alunos. O peso da turma é obtido do mesmo modo como descrito em Peso dos Professores. A seleção de alunos é realizada por amostragem inversa simples, desse modo, o peso básico dos alunos é igual ao inverso da probabilidade de inclusão de cada aluno na amostra. O produto entre as três componentes resulta no peso final dos alunos, dado por:

$$w_{hldk}^{A^a} = w_{hldk}^{**A} \times w_{hldk}^{**T} \times \frac{(A_{hldk}^{ra} - 1)}{(A_{hldk}^{sa} - 1)} \times \frac{A_{hldk}^a}{A_{hldk}^{ra}},$$

onde:

$w_{hldk}^{A^a}$ é o peso final do aluno A da série a respondentes da escola k da dependência d na localidade l do estrato h ;

w_{hldk}^{**A} é o peso corrigido para escolas com alunos respondentes da escola k da dependência d na localidade l do estrato h ;

w_{hldk}^{**T} é o peso da turma T da série a da escola k da dependência d na localidade l do estrato h ;

A_{hldk}^a é o número de alunos existentes na série a da escola k da dependência d na localidade l do estrato h ;

A_{hldk}^{sa} é o número de alunos selecionados na série a da escola k da dependência d na localidade l do estrato h ; e

A_{hldk}^{ra} é o número de alunos respondentes na série a da escola k da dependência d na localidade l do estrato h .

ERROS AMOSTRAIS

Os cálculos das medidas ou estimativas dos erros amostrais dos indicadores da pesquisa TIC Educação levam em consideração o plano amostral empregado na pesquisa. Foi utilizado o Método do Conglomerado Primário (do inglês, *Ultimate Cluster*) para estimação de variâncias para estimadores de totais em planos amostrais de múltiplos estágios. Proposto por Hansen, Hurwitz e Madow (1953), o método considera apenas a variação entre informações disponíveis no nível das unidades primárias de amostragem (UPA), tratando-as como se tivessem sido selecionadas do estrato com reposição da população.

Com base nesse conceito, é possível considerar a estratificação e a seleção com probabilidades desiguais, tanto das unidades primárias quanto das demais unidades de amostragem. A premissa para permitir a aplicação desse método é que estejam disponíveis estimadores não viciados dos totais da variável de interesse para cada um dos conglomerados primários selecionados. Esse método fornece a base para vários dos pacotes estatísticos especializados em cálculo de variâncias considerando o plano amostral.

A partir das variâncias estimadas, divulgam-se os erros amostrais expressos pela margem de erro. Para a divulgação, as margens de erros foram calculadas para um nível de confiança de 95%. Isso significa que, se a pesquisa for repetida várias vezes, em 95% delas o intervalo de confiança conterá o verdadeiro valor populacional. Outras medidas derivadas dessa estimativa de variabilidade são comumente apresentadas, tais como: erro padrão, coeficiente de variação ou intervalo de confiança.

O cálculo da margem de erro considera o produto do erro padrão (raiz quadrada da variância do estimador) pelo valor 1,96 (valor da distribuição normal que corresponde ao nível de significância escolhido de 95%). Esses cálculos são feitos para cada variável de cada uma das tabelas, o que assegura que todas as tabelas de indicadores possuem margens de erro relacionadas a cada estimativa apresentada em cada célula da tabela.

METODOLOGIA DA PESQUISA EM ESCOLAS RURAIS

CONCEITOS E DEFINIÇÕES

POPULAÇÃO-ALVO

A população-alvo do estudo é composta pelas escolas públicas (estaduais e municipais) e particulares em atividade, localizadas em áreas rurais do Brasil. Também fazem parte da população-alvo todos os diretores ou responsáveis pela administração do estabelecimento escolar rural.

UNIDADE DE ANÁLISE E REFERÊNCIA

Para atingir o objetivo da pesquisa, são investigadas várias dimensões relacionadas às unidades de referência e análise. São elas:

- **Escolas localizadas em áreas rurais:** perfil em termos de infraestrutura e práticas envolvendo TIC;
- **Responsáveis pela escola localizada em área rural:** perfil de uso do computador e da Internet; uso de TIC nas atividades administrativas e de gestão; interação com a comunidade e percepção sobre as limitações para a integração das TIC à educação.

DOMÍNIOS DE INTERESSE PARA ANÁLISE E DIVULGAÇÃO

Para as unidades de análise e referência, os resultados são divulgados para domínios definidos com base nas variáveis e níveis descritos a seguir.

Para todas as unidades de análise:

- **Região:** corresponde à divisão regional do Brasil, segundo critérios do Instituto Brasileiro de Geografia e Estatística (IBGE), nas macrorregiões Centro-Oeste, Nordeste, Norte, Sudeste e Sul;
- **Dependência administrativa:** corresponde ao tipo de subordinação administrativa da escola – pública municipal, pública estadual ou particular;

Para os responsáveis pela escola, acrescentam-se:

- **Sexo:** corresponde à divisão em feminino ou masculino;
- **Faixa etária:** corresponde às faixas de idade em que se encontrava o respondente no dia da entrevista, expressa em anos completos;
- **Renda familiar:** corresponde à soma da renda mensal de todos os membros do domicílio do respondente em três faixas de renda, iniciando-se pelo salário mínimo vigente definido pelo Ministério do Trabalho e do Emprego no ano da coleta;

- **Renda pessoal:** corresponde à soma da renda mensal total do respondente em três faixas de renda, iniciando-se pelo salário mínimo vigente definido pelo Ministério do Trabalho e do Emprego no ano da coleta.

INSTRUMENTOS DE COLETA

INFORMAÇÕES SOBRE OS INSTRUMENTOS DE COLETA

As entrevistas são realizadas com questionários estruturados específicos para os diretores ou responsáveis das escolas que foram selecionadas para responder à pesquisa. Mais informações sobre os instrumentos de coleta podem ser encontradas no Relatório de Coleta de Dados da pesquisa.

PLANO AMOSTRAL

A amostra de escolas rurais é probabilística estratificada simples em um estágio. Esse estágio de seleção da amostra de escolas é realizado a partir da estratificação do universo de pesquisa segundo grandes regiões brasileiras e localidades³. Em cada estrato, são selecionadas escolas localizadas em áreas rurais.

CADASTRO E FONTES DE INFORMAÇÃO

O cadastro utilizado para seleção das escolas é o Censo Escolar da Educação Básica, coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Esse cadastro contém dados referentes a todas as escolas de Educação Básica no Brasil. A partir do cadastro de escolas do Inep mais recente (divulgado em março de cada ano), são separadas as escolas que satisfazem a todas as condições de elegibilidade para participar da população da pesquisa, ou seja, escolas em funcionamento, localizadas em área rural de dependência municipal, estadual ou privada. As escolas elegíveis que tenham sido criadas durante o ano em que se dá a coleta de dados não fazem parte da população de pesquisa.

DIMENSIONAMENTO DA AMOSTRA

O tamanho da amostra da pesquisa TIC Educação, com a coleta de dados em escolas localizadas em áreas rurais, tem por objetivo proporcionar uma melhor leitura dos resultados diante da heterogeneidade das unidades de referência e análise. As características de seleção resultam em um tamanho variável da amostra, descrito detalhadamente no Relatório de Coleta de Dados da pesquisa.

³ As localidades são agrupamentos de municípios construídos para a pesquisa em escolas urbanas. Ver Construção das Localidades.

CRITÉRIOS PARA DESENHO DA AMOSTRA

A amostra da pesquisa é desenhada utilizando a técnica de amostragem estratificada, que visa melhorar a precisão das estimativas e garantir a inclusão de subpopulações de interesse. As variáveis de estratificação são região e uma variável indicadora de localidades presentes na amostra da pesquisa TIC Educação em escolas urbanas.

ALOCAÇÃO DA AMOSTRA

A estratégia geral de amostragem da pesquisa envolve a seleção de uma amostra de escolas segundo dois grupos: escolas rurais em localidades, conforme seleção realizada para a coleta de dados para a pesquisa TIC Educação em escolas urbanas, e escolas rurais que compõem o restante do universo de instituições da pesquisa, ou seja, aquelas que estão em localidades que não foram selecionadas para a pesquisa TIC Educação em escolas urbanas.

A amostra de escolas rurais é obtida por amostragem aleatória simples sem reposição em cada estrato. Dessa forma, dentro de cada estrato, as probabilidades de seleção são iguais. A tabela contendo a alocação da amostra de escolas rurais está disponível no Relatório de Coleta de Dados da pesquisa.

SELEÇÃO DA AMOSTRA

Dentro de cada estrato, as escolas rurais são selecionadas por amostragem aleatória simples. Dessa forma, a probabilidade de seleção de cada escola rural é dada por:

N é o tamanho total da população;

N_h é o tamanho da população do estrato h ;

n é o tamanho da amostra; e

n_h é o tamanho da amostra dentro de cada estrato h . De tal forma, que:

$$n_h = n \times \frac{N_h}{N}.$$

Logo, as probabilidades de inclusão das escolas i para cada estrato h são dadas por:

$$\pi_{ih} = \frac{n_h}{N_h}.$$

COLETA DE DADOS EM CAMPO

MÉTODO DE COLETA

As escolas são contatadas por meio da técnica de Entrevista Telefônica Assistida por Computador (do inglês, *Computer-Assisted Telephone Interviewing* – CATI).

Como parte da edição 2016 da pesquisa TIC Educação, foi conduzido um estudo piloto com escolas que, segundo o Censo Escolar, estavam localizadas em áreas rurais (CGI.br, 2017). Por meio daquele estudo, foi possível observar que algumas escolas de áreas rurais não possuíam telefone, entre outras dificuldades de contato. Desse modo, a fim de obter informações sobre escolas com tais características, foram selecionadas algumas instituições para a realização de uma abordagem presencial, ou melhor, pelo método CAPI (do inglês, *Computer-Assisted Personal Interviewing*).

Devido ao alto custo da aplicação desse método em áreas rurais, as escolas selecionadas deveriam pertencer às mesmas localidades já selecionadas para a coleta de dados da pesquisa em escolas localizadas em áreas urbanas, que também é realizada presencialmente. Além disso, esse grupo de escolas deveria ser composto apenas por instituições as quais não foi possível fazer contato por outro meio, como o telefone.

Em cada escola pesquisada, busca-se entrevistar o diretor ou o principal responsável, isto é, aquela pessoa que mais conheça a instituição como um todo.

PROCESSAMENTO DOS DADOS

PROCEDIMENTOS DE PONDERAÇÃO

A ponderação da pesquisa parte das probabilidades de seleção das escolas rurais, na qual são aplicadas correções de não resposta por estratos. Os pesos das escolas são calibrados para os totais conhecidos da população-alvo da pesquisa.

PESO DAS ESCOLAS

A cada escola da amostra é associado um peso amostral básico, obtido pela razão entre o tamanho da população e o tamanho da amostra no estrato final correspondente. O peso básico é calculado a partir do inverso da probabilidade de seleção de escolas em cada estrato, expresso pela equação:

$$w_{ih} = \frac{N_h}{n_h},$$

onde:

w_{ih} é peso básico da escola i no estrato h ;

N_h é o tamanho da população do estrato h ; e

n_h é o tamanho da amostra dentro de cada estrato h .

Para corrigir os casos nos quais não se obtém a resposta de todos os selecionados, é realizada uma correção de não resposta. Como cada estrato pode possuir um contingente diferente de escolas respondentes, é considerado o ajuste dentro de cada estrato através da fórmula:

$$w_{ih}^* = w_{ih} \times \frac{N_h}{n_h^r},$$

onde:

w_{ih}^* é o peso com correção de não resposta da escola i no estrato de h ;

w_{ih} é peso básico da escola i no estrato h ;

N_h é o tamanho da população do estrato h ; e

n_h^r é o total de escolas respondentes no estrato h .

CALIBRAÇÃO

Após a primeira correção básica de não resposta, é feita uma comparação entre as estimativas da proporção obtida pela amostra para dois indicadores presentes no Censo Escolar – escolas com computador e escolas com acesso à Internet – e as estimativas obtidas com o universo-alvo da pesquisa. Essa comparação tem por objetivo identificar possíveis vieses de resposta/não resposta associados à capacidade de se contatar as escolas (devido à metodologia CATI amplamente utilizada na coleta de dados).

Dessa forma, caso verificado algum viés associado a não resposta diferencial, os pesos corrigidos para não resposta são calibrados. Assim, é ajustado um modelo logístico para previsão da probabilidade de resposta a partir de variáveis conhecidas do universo de pesquisa. A partir das variáveis identificadas como significantes no modelo, é implementado o método *raking*. O peso final das escolas é denominado por: w_{ih}^{**} .

ERROS AMOSTRAIS

Os cálculos das medidas ou estimativas dos erros amostrais dos indicadores da coleta de dados em escolas rurais levaram em consideração o plano amostral empregado na pesquisa. Foi utilizado o Método do Conglomerado Primário (do inglês, *Ultimate Cluster*) para estimação de variâncias para estimadores de totais em planos amostrais de múltiplos estágios. Proposto por Hansen, Hurwitz e Madow (1953), o método considera apenas a variação entre informações disponíveis no nível das unidades primárias de amostragem (UPA), tratando-as como se tivessem sido selecionadas do estrato com reposição da população.

Com base nesse conceito, pode-se considerar a estratificação e a seleção com probabilidades desiguais para as UPA. As premissas para permitir a aplicação desse método são: que haja estimadores não viciados dos totais da variável de interesse para cada um dos conglomerados primários selecionados; e que pelo menos dois deles sejam selecionados em cada estrato (se a amostra for estratificada no primeiro estágio). Esse método fornece a base para vários pacotes estatísticos especializados em cálculo de variâncias considerando o plano amostral.

A partir das variâncias estimadas, optou-se por divulgar os erros amostrais expressos pela margem de erro. Para a divulgação, as margens de erros foram calculadas para um nível de confiança de 95%. Assim, se a pesquisa fosse repetida, em 19 de cada 20 vezes o intervalo conteria o verdadeiro valor populacional. Outras medidas derivadas dessa estimativa de variabilidade são comumente apresentadas, tais como: erro padrão, coeficiente de variação ou intervalo de confiança.

O cálculo da margem de erro considera o produto do erro padrão (raiz quadrada da variância) pelo valor 1,96 (valor da distribuição normal que corresponde ao nível de significância escolhido de 95%). Esses cálculos foram feitos para cada variável de cada uma das tabelas, o que assegura que todas as tabelas de indicadores possuem margens de erros relacionadas a cada estimativa apresentada em cada célula da tabela.

DISSEMINAÇÃO DOS DADOS

Os resultados da TIC Educação são apresentados de acordo com as variáveis descritas no item Domínios de interesse para análise e divulgação. Arredondamentos fazem com que, em alguns resultados, a soma das categorias parciais difira de 100% em questões de resposta única. O somatório de frequências em questões de respostas múltiplas usualmente é diferente de 100%. Vale ressaltar que, nas tabelas de resultados, o hífen (–) é utilizado para representar a não resposta ao item. Por outro lado, como os resultados são apresentados sem casa decimal, as células com valor zero significam que houve resposta ao item, mas ele é explicitamente maior do que zero e menor do que um.

Nesta publicação, estão apresentadas as tabelas de alguns indicadores selecionados para cada unidade de análise. As tabelas de totais e margens de erros calculadas para cada indicador estão apenas disponíveis para *download* no *website* do Cetic.br.⁴

⁴ Mais informações no *website* do Cetic.br. Recuperado em 20 julho, 2017, de <http://data.cetic.br/cetic/>

REFERÊNCIAS

- Assunção, R., Lage, J., & Reis, E. (2002). Análise de conglomerados espaciais via árvore geradora mínima. *Revista Brasileira de Estatística*, 62(220), 1-23.
- Bolfarine, H., & Bussab, W. O. (2005). *Elementos de amostragem*. São Paulo: Blucher.
- Cochran, W. G. (1977). *Sampling techniques* (3ª ed.). Nova Iorque: John Wiley & Sons.
- Comitê Gestor da Internet no Brasil – CGI.br (2017). *Pesquisa sobre o uso das tecnologias de informação e comunicação nas escolas brasileiras: TIC Educação 2016*. São Paulo: CGI.br.
- Hansen, M. H., Hurwitx, W. N., & Madow, W. G. (1953). *Sample survey methods and theory*. Nova Iorque: Wiley.
- International Association for the Evaluation of Educational Achievement – IEA (2009). *Sites 2006 Technical Report*. Recuperado em 20 março, 2013, de http://www.iea.nl/fileadmin/user_upload/Publications/Electronic_versions/SITES_2006_Technical_Report.pdf
- International Association for the Evaluation of Educational Achievement – IEA (2009). *Sites 2006 User Guide for the International Database*. Recuperado em 20 março, 2013, de http://pub.iea.nl/fileadmin/user_upload/Publications/Electronic_versions/SITES_2006_IDB_User_Guide.pdf
- Lumley, T. (2010). *Complex surveys: a guide to analysis using R*. Nova Jersey: John Wiley & Sons.
- Ministério da Educação – MEC & Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2011). *Plano de Desenvolvimento da Educação e Sistema Nacional de Avaliação da Educação Básica 2011*. Recuperado em 25 julho, 2012, de http://www.oei.es/quipu/brasil/RelatorioSaeb2003_3.pdf
- Ministério da Educação – MEC & Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep (2015). *Censo Escolar 2014*. Recuperado em 5 novembro, 2015, de <http://portal.inep.gov.br/microdados>
- Ohlsson, E. (1995). Coordination of samples using permanent random numbers. In: B. G. Cox. *Business survey methods* (pp. 153-170). Nova Iorque: John Wiley.
- Ohlsson, E. (1998). Sequential Poisson Sampling. *Journal of Official Statistics*, 14, pp. 149-162.
- Särndal, C., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling*. Nova Iorque: Springer Verlag.
- Thompson, S. K. (1999). *Sampling*. Nova Iorque: John Wiley & Sons.
- Vasconcellos, M. T., Silva, P. L., & Szwarcwald, C. L. (2005). Sampling design for the World Health Survey in Brazil. *Cadernos de Saúde Pública*, 21.