

GUIA PARA DIVULGAÇÃO
DE ERROS AMOSTRAIS
NAS PESQUISAS
POR AMOSTRAGEM
PROBABILÍSTICA
REALIZADAS PELO IBGE

Presidente da República
Jair Messias Bolsonaro

Ministro da Economia
Paulo Roberto Nunes Guedes

Secretário Especial do Tesouro e Orçamento
Esteves Pedro Colnago Junior

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

Presidente
Eduardo Luiz G. Rios Neto

Diretora-Executiva
Marise Maria Ferreira

ÓRGÃOS ESPECÍFICOS SINGULARES

Diretoria de Pesquisas
Cimar Azeredo Pereira

Diretoria de Geociências
Claudio Stenner

Diretoria de Informática
Carlos Renato Pereira Cotovio

Centro de Documentação e Disseminação de Informações
Carmen Danielle Lins Mendes Macedo

Escola Nacional de Ciências Estatísticas
Maysa Sacramento de Magalhães

UNIDADE RESPONSÁVEL

Diretoria de Pesquisas
Coordenação de Métodos e Qualidade
Andrea Borges Paim

Ministério da Economia
Instituto Brasileiro de Geografia e Estatística - IBGE
Diretoria de Pesquisas
Coordenação de Métodos e Qualidade

Guia para Divulgação de Erros Amostrais nas Pesquisas por Amostragem Probabilística Realizadas pelo IBGE



Rio de Janeiro
2021

Instituto Brasileiro de Geografia e Estatística - IBGE
Av. Franklin Roosevelt, 166 - Centro - 20021-120 - Rio de Janeiro, RJ - Brasil

© IBGE. 2021

Capa - Ubiratã O. Santos/Eduardo Sidney Araújo

Coordenação de Marketing/Centro de Documentação e Disseminação
de Informações - CDDI

Sumário

Apresentação	5
Introdução	7
Definições	9
Medidas de erros amostrais	13
Erro padrão estimado	14
Coeficiente de variação estimado	14
Intervalo de confiança estimado	15
Formas de divulgação da precisão das estimativas	17
Modelos de função de variância.....	20
Tabelas com valor da medida de precisão ao lado das estimativas	22
Tabelas separadas com medidas de precisão.....	24
Tabelas com indicador de intervalo de valor para coeficiente de variação ao lado da estimativa.....	26
Considerações finais	29
Referências	33

Apresentação

Com este guia, o Instituto Brasileiro de Geografia e Estatística - IBGE divulga as formas de apresentação dos erros amostrais associados às estimativas obtidas a partir de pesquisas por amostragem probabilística realizadas pelo Instituto.

Este guia está em consonância com o Código de Boas Práticas das Estatísticas do IBGE, em sua segunda edição, publicada em 2021, especialmente no que concerne ao Princípio 14 - Precisão e exatidão e ao Princípio 17 - Acessibilidade e clareza.

Com isso, o IBGE reafirma sua missão institucional, de “Retratar o Brasil com informações necessárias ao conhecimento de sua realidade e ao exercício da cidadania”.

Cimar Azeredo Pereira

Diretor de Pesquisas



Introdução

Em dezembro de 2013, foi publicado o Código de Boas Práticas das Estatísticas do IBGE, cuja revisão deu origem à segunda edição, divulgada em 2021. O Código é um instrumento orientador e regulador, constituído por um conjunto de diretrizes que têm por finalidade promover uma conduta profissional padronizada na aplicação das melhores práticas estatísticas, fundamentais para a credibilidade institucional e, portanto, para o reconhecimento e a confiança da sociedade nas informações que a Instituição produz.

Tal como previsto no Código, em seu Princípio 14 - Precisão e exatidão, indicador 14.2, “o IBGE deve analisar e documentar os erros amostrais e não amostrais de suas estatísticas oficiais” (IBGE, 2021, p. 55). Ademais, no Princípio 17 - Acessibilidade e clareza, o indicador 17.4 estabelece que “o IBGE deve disponibilizar informações sobre a metodologia de suas operações estatísticas, incluindo a qualidade de seus produtos” (IBGE, 2021, p. 62).

Complementarmente a essas disposições, em 2019, foi publicado o Guia para Elaboração de Relatórios Metodológicos, o qual relaciona o conteúdo que deve constar nos relatórios que descrevem a metodologia das operações estatísticas realizadas pelo Instituto (IBGE, 2019). Um dos itens constantes nesse Guia, em seu subitem 7.2, trata dos aspectos de qualidade e recomenda que sejam apresentadas “informações sobre o grau em que as estatísticas descritivas produzidas pelo IBGE refletem, de forma precisa e confiável, a realidade do fenômeno que se pretendeu mensurar”, entre elas o erro amostral (IBGE, 2019, p. 16). Tais in-

formações também são apresentadas no Sistema de Metadados do IBGE e no relatório de qualidade gerado a partir desse Sistema. Portanto, é fundamental que informações sobre a qualidade dos produtos estatísticos, em especial sobre a sua precisão, sejam disponibilizadas para permitir que os usuários avaliem e utilizem as estatísticas produzidas pelo IBGE de maneira adequada.

O Guia ora divulgado descreve formas sugeridas para a apresentação dos erros amostrais associados às estimativas obtidas a partir de pesquisas por amostragem probabilística, exemplificando, quando possível, as operações estatísticas do IBGE que as utilizam.



Definições

Qualidade

Qualidade pode ser definida como adequação para o uso ou adequação para os propósitos.

Acurácia

Acurácia é uma das dimensões de qualidade identificadas com o conceito de adequação para o uso ou adequação para os propósitos.

A acurácia de uma informação estatística é definida como a proximidade entre a estimação de um valor e o seu verdadeiro valor não conhecido e pode ser mensurada pelo erro total, incluindo:

- Erros de amostragem ou erros amostrais; e
- Erros de não amostragem, os quais podem ser classificados em:
 - a) Erro de cobertura;
 - b) Erro de medida;
 - c) Erro de elaboração (entrada de dados, crítica, imputação, expansão, tabulação);
 - d) Erro de não resposta (total ou parcial); ou
 - e) Erro de suposição do modelo.

Erros de amostragem ou erros amostrais

Erros de amostragem ou erros amostrais derivam do fato de que apenas uma amostra de unidades da população objetivo é investigada. Estimativas baseadas em uma pesquisa por amostragem variam de amostra para amostra e tipicamente são diferentes dos resultados que seriam obtidos se a investigação fosse censitária.

Medidas de erros de amostragem

Erro padrão, coeficiente de variação e intervalo de confiança são formas diferentes de se apresentar os erros de amostragem associados a uma informação estatística estimada ou, simplesmente, estimativa. A escolha da medida a ser usada depende do público a que se destina.

Parâmetro

Parâmetro é uma função dos valores de uma determinada variável da população.

Estatística

Estatística é uma função dos valores de uma determinada variável observados numa amostra.

Estimador

Estimador de um parâmetro de uma dada variável da população é toda estatística que, aplicada aos valores observados numa amostra, fornece uma estimativa para o parâmetro.

Estimativa

Estimativa é o valor que o estimador assume para uma dada amostra.

Erro padrão do estimador

Erro padrão do estimador é uma medida de dispersão dos valores desse estimador, obtidos com cada uma das amostras possíveis, selecionadas sob as mesmas condições.

Erro padrão de um estimador é o desvio padrão desse estimador, ou a raiz quadrada da variância do estimador.

Erro padrão estimado de uma estatística (ou estimativa)

Erro padrão estimado de uma estatística (ou estimativa) é o valor obtido do desvio padrão do estimador a partir de uma amostra específica.

É uma medida da incerteza associada à estimativa: quanto menor o valor do erro padrão estimado, maior a precisão da estimativa.

Coefficiente de variação

Coefficiente de variação é uma medida de dispersão relativa que representa o desvio padrão de um conjunto de dados expresso como porcentagem da média. É obtido pela razão entre o desvio padrão e a média do conjunto de dados.

Coefficiente de variação de um estimador

Coefficiente de variação de um estimador representa a dispersão dos valores desse estimador obtidos com cada uma das amostras possíveis, selecionadas sob as mesmas condições, expressa como porcentagem da esperança do estimador.

Coefficiente de variação estimado de uma estatística (ou estimativa)

Coefficiente de variação estimado de uma estatística (ou estimativa) é o valor obtido do coeficiente de variação do estimador a partir de uma amostra específica.

Quanto menor o valor do coeficiente de variação estimado, maior a precisão da estimativa.

Intervalo de confiança

Intervalo de confiança é um intervalo aleatório que contém o parâmetro (quantidade de interesse) com probabilidade fixada, chamada de nível de confiança.

Em sentido estrito, um intervalo de confiança para um parâmetro populacional é um intervalo com uma proporção p associada, de tal forma que, se a amostragem for repetida inúmeras vezes e o intervalo de confiança for recalculado para cada amostra de acordo com o mesmo método, uma proporção p dos intervalos de confiança conterá o parâmetro em questão.

O intervalo de confiança é a forma predominante de estimativa por intervalo.

Intervalo de confiança estimado de uma estatística (ou estimativa)

Intervalo de confiança estimado de uma estatística (ou estimativa) é construído a partir do erro padrão estimado e de uma medida do nível de confiança.

Dizer que um intervalo tem 95% de confiança significa que, se forem feitas repetidas seleções de amostras, seguindo um mesmo plano amostral, as estimativas obtidas dessas amostras levarão a intervalos que, em 95% das vezes, incluirão o valor verdadeiro do parâmetro estimado.

É uma forma de apresentar a precisão de uma estimativa. Em vez de estimar o parâmetro por um único valor, é dado um intervalo de estimativas prováveis. O quanto essas estimativas são prováveis será determinado pelo nível de confiança.

Quanto menor o tamanho do intervalo de confiança, maior a precisão da estimativa.



Medidas de erros amostrais

As conclusões de uma pesquisa por amostragem probabilística devem ser apoiadas nas estimativas produzidas, as quais, por sua vez, são calculadas considerando-se os pesos amostrais, associados às unidades da amostra, de acordo com o plano amostral definido e com os estimadores usados. Essas estimativas têm uma variabilidade que é inerente ao processo de amostragem. Assim, a avaliação dos chamados erros de amostragem, ou erros amostrais, é um ponto fundamental, pois dela decorre o grau de confiança nas conclusões analíticas que subsidiam a tomada de decisão.

Para cada estimativa derivada da pesquisa, é possível obter uma medida de precisão que auxilia na análise e na interpretação dos dados, por meio de uma estimativa do erro amostral. Há, porém, uma questão que deve ser considerada, qual seja: os diferentes tipos de indicadores que podem ser obtidos.

Independentemente do plano amostral adotado para a realização da pesquisa, estimativas de totais, médias e proporções, em geral, possuem fórmulas de variância associadas que estão implementadas em algum pacote estatístico, seja para cálculo direto, seja aproximado. O mesmo não ocorre com estimativas de diferença e de variação, em pesquisas repetidas, ou em estimativas de taxas. De forma geral, o cálculo da variância desses tipos de estimativas envolve o cálculo simultâneo de variâncias e covariâncias, o que nem sempre pode ser obtido pelos pacotes estatísticos diretamente (LILA; FREITAS, 2007). Nessas situações, pode-se divulgar, junto à estimativa de diferença ou de

variação, um indicador de significância estatística, ou divulgar o próprio erro amostral, dependendo das demandas e necessidades dos usuários.

No caso de estimativas de indicadores complexos, como, por exemplo, o índice de Gini, o cálculo de variâncias demanda programação específica. Isso faz com que nem sempre esses indicadores possam ser divulgados com medidas de erro amostral associadas.

A escolha da medida de precisão que será utilizada na divulgação dos resultados de uma operação estatística deve considerar as necessidades dos usuários, observando-se as vantagens e desvantagens de cada uma dessas medidas.

Os erros amostrais podem ser avaliados por meio do erro padrão estimado, do coeficiente de variação estimado, ou do intervalo de confiança estimado, e suas principais vantagens e desvantagens são elencadas a seguir.

Erro padrão estimado

Vantagens:

- Constitui medida básica de erro da estimativa, sempre positiva e na mesma escala, e unidade de medida da própria estimativa; e
- É amplamente conhecida e utilizada na área de Estatística e com propriedades bastante conhecidas.

Desvantagens:

- Não permite uma interpretação direta pelo usuário;
- Não é uma medida intuitiva para leigos ou o público em geral; e
- Geralmente, necessita que seja feito mais um cálculo para verificar se a estimativa correspondente é de boa qualidade.

Coeficiente de variação estimado

Vantagens:

- É um valor padronizado que permite verificar, facilmente, a qualidade da estimativa;
- Não depende da unidade de medida da estimativa correspondente; e
- Facilita a comparação da qualidade de diversas estimativas.

Desvantagens:

- Pode assumir valor negativo, caso a estimativa seja negativa, o que dificulta a interpretação pelo usuário;
- Caso a estimativa seja muito pequena, pode assumir um valor muito grande; e
- No caso de diferença entre estimativas próxima a zero, um valor do coeficiente de variação excepcionalmente grande pode trazer dúvidas infundadas sobre a qualidade das estimativas originais.

Intervalo de confiança estimado

Vantagens:

- Possibilita uma interpretação fácil, pois transmite, rapidamente, um intervalo onde se acredita que está o valor verdadeiro que se busca medir; e
- Permite verificar a qualidade da estimativa pela amplitude do intervalo.

Desvantagens:

- Normalmente, é interpretado de forma errônea, sendo considerado um intervalo de probabilidade quando não é;
- É necessário assumir a normalidade do estimador, uma hipótese muito forte quando se trabalha com amostras pequenas ou quantidades de interesse com taxas próximas a 0% ou a 100%; e

- A definição do nível de confiança pode não satisfazer o grau de precisão que todos os usuários necessitam em seus trabalhos analíticos, ocasionando esforço de cálculo extra para alguns deles.



Formas de divulgação da precisão das estimativas

A forma de divulgação dos erros amostrais em uma operação estatística deve ser adequada para atender aos seus usuários, observando-se o nível de detalhe requerido e permitindo que tenham acesso a toda informação de que necessitam. Desse modo, a forma como esses erros são apresentados combina uma medida de precisão com a quantidade de informação disponibilizada.

As seguir, são apresentadas as formas de divulgação da precisão das estimativas sugeridas para adoção nas operações estatísticas do IBGE, com exemplos de uso.

Tabelas resumo com medida de precisão associada a tamanhos de estimativa

Em pesquisas cujas estimativas são basicamente contagens de unidades com determinadas características, como é o caso das pesquisas por amostra de domicílios, que levantam informações sobre unidades domiciliares e seus moradores, uma das formas de apresentação dos erros amostrais é a utilização de tabelas-resumos contendo medida de precisão associada a diversos tamanhos de estimativa.

Para estimativas que são contagens, o erro amostral é diretamente relacionado com a própria estimativa; portanto, a divulgação da precisão para determinados tamanhos de estimativa que contemplem características mais raras (estimativas pequenas) até as mais comuns (estimativas grandes) fornece informação suficiente para os usuários avaliarem a qualidade dos resul-

tados da operação estatística. Para essa forma de divulgação, os erros amostrais podem ser apresentados tanto em termos de erro padrão quanto de coeficiente de variação.

É recomendado que as tabelas-resumos sejam construídas para os temas investigados e para os domínios de interesse, especialmente níveis geográficos para os quais são gerados resultados, como, por exemplo Brasil, Unidades da Federação e Regiões Metropolitanas, pois a precisão das estimativas pode ser diferenciada para esses domínios.

A divulgação da medida de precisão das estimativas por meio de tabelas-resumos é apropriada, especialmente quando a operação estatística produz muitas tabelas de resultados, para diversos domínios, tornando o cálculo do erro amostral para todas as estimativas muito custoso. Como o erro amostral não é divulgado para todas as estimativas produzidas, devem ser disponibilizadas instruções para cálculo e informações necessárias, no arquivo de microdados público, para que o usuário possa calcular a medida de precisão, caso a aproximação obtida pela tabela não seja suficiente. Se a operação estatística não disponibiliza seus microdados por questões de sigilo estatístico, recomenda-se que as tabelas-resumos sejam as mais detalhadas possível.

Exemplo de uso no Censo Demográfico

Na divulgação dos resultados provenientes da amostra do Censo Demográfico 2010, optou-se por divulgar o erro padrão e o coeficiente de variação (CV). Em cada publicação temática, foram apresentadas tabelas-resumos com valores de erro padrão e coeficientes de variação calculados para alguns tamanhos de estimativas de características de pessoas e de domicílios para o Brasil, as Grandes Regiões e as Unidades da Federação, como exemplificado na Figura 1, que reproduz tabela divulgada em uma das publicações temáticas de resultados do Censo Demográfico 2010.

Figura 1 - Exemplo da divulgação de erro amostral no Censo Demográfico 2010

Tabela 1 - Erro-padrão e estimativa do CV aproximados para alguns tamanhos de estimativas de características de pessoas e domicílios - Região Norte

Características de pessoas			Características de domicílios		
Tamanho de estimativa	Erro-padrão aproximado	CV (%)	Tamanho de estimativa	Erro-padrão aproximado	CV (%)
100	10	10,0	100	10	10,0
500	22	4,5	500	22	4,5
1 000	32	3,2	1 000	32	3,2
2 000	45	2,2	2 000	45	2,2
5 000	71	1,4	5 000	71	1,4
10 000	100	1,0	10 000	100	1,0
20 000	141	0,7	20 000	141	0,7
50 000	223	0,4	50 000	222	0,4
100 000	315	0,3	100 000	312	0,3
150 000	386	0,3	150 000	380	0,3
200 000	444	0,2	200 000	436	0,2
500 000	696	0,1	250 000	484	0,2
1 000 000	968	0,1	500 000	662	0,1
2 000 000	1 322	0,1	1 000 000	867	0,1
3 000 000	1 560	0,1	2 000 000	1 003	0,1
4 000 000	1 730	0,0	3 000 000	874	0,0
5 000 000	1 850	0,0	4 000 000	159	0,0
6 000 000	1 932	0,0	4 025 533	0	0,0
7 000 000	1 978	0,0			
8 000 000	1 991	0,0			
9 000 000	1 973	0,0			
10 000 000	1 923	0,0			
15 000 000	904	0,0			
15 864 454	0	0,0			

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de População e Indicadores Sociais e Coordenação de Métodos e Qualidade.

Nota: Fração amostral efetiva: 10,84%.

Para outros níveis geográficos, foram fornecidas as informações necessárias para que o próprio usuário obtivesse a estimativa da medida de precisão aproximada da estimativa de interesse.

Modelos de função de variância

Uma alternativa de apresentação dos erros amostrais em operações estatísticas cujas estimativas são contagens é a utilização de modelos de regressão do tipo Função de Variância Generalizada (WOLTER, 2007). Como nas tabelas-resumos, nessa forma de apresentação, são divulgadas aproximações dos erros amostrais, relacionando o tamanho da estimativa com o seu nível de precisão, expresso pelo coeficiente de variação (CV). Enquanto nas tabelas os erros são divulgados para um conjunto limitado de tamanhos de estimativa, nos modelos de regressão, o usuário tem a possibilidade de calcular o erro aproximado para qualquer tamanho de estimativa.

É recomendado que, para cada domínio de publicação dos resultados, especialmente para níveis geográficos diferentes, seja ajustado um modelo de regressão para explicar os coeficientes de variação das estimativas como função das próprias estimativas. O modelo ajustado é do tipo $CV=AY^B$, onde Y é o valor da estimativa, CV é o respectivo coeficiente de variação, e A e B são os parâmetros do modelo. Os coeficientes de regressão ajustados, a e b , devem ser publicados juntamente com os resultados da operação estatística.

Para o ajuste dos modelos de regressão em cada domínio, devem ser aleatoriamente selecionadas células das tabelas de resultados, dos diversos temas, que correspondam a estimativas de tamanhos variados. Dessa forma, o coeficiente de variação para cada uma das células selecionadas deve ser estimado de acordo com o plano amostral, e os valores obtidos utilizados na modelagem.

Essa forma de divulgação é interessante para operações estatísticas com uma quantidade muito grande de tabelas de resultados, permitindo que os usuários obtenham boas aproximações para o nível de precisão das estimativas apenas realizando cálculos simples. Disponibilizam-se, assim, mais informações que nas tabelas-resumos, mas são necessários esforços adicionais na estimação dos parâmetros do modelo.

Apesar de possibilitar o cálculo do coeficiente de variação aproximado para qualquer estimativa, alguns usuários especializados podem desejar calcular a medida de precisão sem o uso do modelo. Dessa forma, é necessário que sejam disponibilizadas todas as informações para permitir tais cálculos, nos casos em que a operação estatística divulga os seus microdados.

Exemplo de uso na Pesquisa Nacional por Amostra de Domicílios - PNAD

A precisão das estimativas da PNAD¹ era medida pelo coeficiente de variação, calculado para um conjunto selecionado de estimativas de totais de variáveis categóricas. As estimativas dos coeficientes de variação eram divulgadas, no volume Brasil, por meio do ajuste de um modelo de regressão ($CV=AY^B$) para cada domínio de publicação dos resultados (Brasil, Grandes Regiões, Unidades da Federação e Regiões Metropolitanas), como exemplificado na Figura 2, que reproduz tabela divulgada no volume Brasil de resultados da PNAD 2009.

Figura 2 - Exemplo da divulgação de erro amostral na PNAD 2009

Tabela 2 - Coeficientes de regressão, por tipo de estimativa, segundo a situação do domicílio - Brasil - 2009

Situação do domicílio	Coeficientes de regressão	
	A	B
Pessoas		
Total	1 913,7128	(-) 0,4682
Urbana	770,7994	(-) 0,4064
Rural	304,5823	(-) 0,3050
Famílias e domicílios		
Total	3 033,6896	(-) 0,4953
Urbana	1 524,2050	(-) 0,4444
Rural	873,5733	(-) 0,3733

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios 2009.

¹ A pesquisa foi encerrada em 2016, com a divulgação das informações referentes a 2015.

Tabelas com valor da medida de precisão ao lado das estimativas

Outra alternativa de divulgação da medida de precisão que permite uma avaliação rápida e direta da qualidade das estimativas é a disposição, ao lado da estimativa, do seu erro amostral, o qual pode ser expresso pelo erro padrão, pelo coeficiente de variação ou pelo intervalo de confiança. A escolha da medida utilizada deve ser feita avaliando-se as vantagens e desvantagens já descritas neste guia, as necessidades dos usuários, bem como as exigências dos eventuais acordos e convênios firmados.

O inconveniente dessa forma de apresentação é a necessidade de aumento do número de colunas da tabela de resultados para poder acomodar o erro amostral: no caso do erro padrão e do coeficiente de variação, dobra-se a quantidade de colunas, e, no caso do intervalo de confiança, triplica-se. Por esse motivo, apresentar o erro amostral ao lado das estimativas é apropriado para operações estatísticas cujas tabelas de resultados são enxutas, com poucas colunas, evitando-se a elaboração de tabelas complexas para leitura, o que acaba por eliminar a vantagem dessa forma de divulgação que é a avaliação direta da qualidade das estimativas.

A divulgação do erro amostral ao lado das estimativas pode ser feita para todo o plano amostral da operação estatística, ou para tabelas selecionadas, consideradas as mais importantes.

A escolha de quantas e quais tabelas são apresentadas com medida de precisão deve considerar a quantidade de tabelas produzidas, a complexidade de cálculo, a necessidade dos usuários e a possibilidade de esses calcularem o erro amostral para estimativas não contempladas na escolha. É recomendado que as operações estatísticas do IBGE que não disponibilizam microdados divulguem o máximo de estimativas com o erro amostral associado, permitindo a avaliação da qualidade de seus resultados pelos usuários.

Exemplo de uso na Pesquisa Nacional de Saúde do Escolar - PeNSE

Na área dos estudos sobre saúde, é comum analisar resultados utilizando-se intervalos de confiança associados às estimativas. Por esse motivo e em função do convênio firmado entre o IBGE e o Ministério da Saúde para a realização da pesquisa, nas publicações de resultados da PeNSE, as estimativas são apresentadas com os respectivos intervalos de confiança estimados, considerando-se um nível de confiança de 95%, como exemplificado na Figura 3, que reproduz tabela divulgada na publicação de resultados da PeNSE 2015.

Figura 3 - Exemplo da divulgação de erro amostral na PeNSE 2015

Tabela 1.1.4.1 (parte) - Estimativas total e percentual de escolares frequentando o 9º ano do ensino fundamental em escolas que informaram oferecer comida (merenda escolar/almoço) para os escolares, por dependência administrativa da escola, com indicação do intervalo de confiança de 95%, segundo as Grandes Regiões e as Unidades da Federação - 2015

Grandes Regiões e Unidades da Federação	Estimativa total de escolares frequentando o 9º ano do ensino fundamental em escolas que informaram oferecer comida (merenda escolar/almoço) para os escolares		
	Total		
	Total	Intervalo de confiança de 95%	
Limite inferior		Limite superior	
Brasil	2 299 536	2 248 735	2 350 336
Norte	233 213	225 175	241 251
Rondônia	19 978	18 822	21 134
Acre	13 681	12 468	14 895
Amazonas	55 580	51 781	59 380
Roraima	7 676	6 909	8 442
Pará	104 750	98 082	111 419
Amapá	10 934	10 192	11 677
Tocantins	20 612	19 286	21 938
Nordeste	621 186	600 195	642 178
Maranhão	91 026	83 377	98 674
Piauí	39 815	36 786	42 844
Ceará	109 433	101 691	117 174
Rio Grande do Norte	34 528	31 975	37 081
Paraíba	39 627	36 834	42 419
Pernambuco	94 509	84 616	104 402
Alagoas	33 093	30 675	35 511
Sergipe	21 036	19 136	22 936
Bahia	158 120	144 285	171 955
Sudeste	998 788	956 410	1 041 167
Minas Gerais	272 393	256 992	287 794
Espírito Santo	42 211	39 615	44 806
Rio de Janeiro	143 987	134 716	153 259
São Paulo	540 197	501 908	578 486
Sul	275 757	261 568	289 945
Paraná	118 188	108 485	127 891
Santa Catarina	67 762	61 844	73 681
Rio Grande do Sul	89 806	81 312	98 300
Centro-Oeste	170 591	161 751	179 431
Mato Grosso do Sul	27 115	24 729	29 501
Mato Grosso	40 336	35 982	44 690
Goiás	71 894	67 547	76 241
Distrito Federal	31 247	25 364	37 129

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de População e Indicadores Sociais, Pesquisa Nacional de Saúde do Escolar, Amostra 1, 2015.

Tabelas separadas com medidas de precisão

Para evitar o aumento do número de colunas para apresentação do erro amostral, o que pode dificultar a leitura das tabelas, uma outra alternativa de divulgação é a criação de tabelas separadas contendo as medidas de precisão, as quais podem ser qualquer uma das já mencionadas.

Nessa forma de apresentação, a avaliação da qualidade das estimativas não ocorre diretamente como na alternativa anterior, pois o usuário precisa observar as duas tabelas em conjunto: a tabela de resultados e a tabela de erros amostrais.

Como na alternativa anterior, devem ser consideradas as necessidades dos usuários e todas as questões operacionais de cálculo para definir quais tabelas têm a sua correspondente de erros amostrais produzida – podem ser apenas as principais, ou todo o plano tabular. É importante também levar em conta se os usuários têm acesso às informações necessárias para calcular a medida de precisão; caso não seja possível calculá-la, recomenda-se que mais estimativas tenham o seu erro amostral divulgado.

Como as duas formas de divulgação são equivalentes, a escolha por uma delas deve ser feita observando-se aspectos visuais da apresentação, a complexidade das tabelas, o número de tabelas cujas medidas de precisão serão apresentadas, bem como o próprio retorno dos usuários.

Exemplo de uso na Pesquisa de Orçamentos Familiares - POF

Essa pesquisa, a partir da edição correspondente a 2002-2003, passou a divulgar o erro amostral para estimativas selecionadas de seu plano tabular em tabelas separadas. Para cada tabela de resultados selecionada, foi produzida outra tabela contendo os coeficientes de variação correspondentes, como exemplificado na Figura 4, que reproduz tabela divulgada em uma das publicações temáticas de resultados da POF 2017-2018.

Figura 4 - Exemplo da divulgação de erro amostral na POF 2017-2018

Tabela 3 - Coeficientes de variação do rendimento total e variação patrimonial médio mensal familiar, por classes de rendimento total e variação patrimonial mensal familiar, segundo os tipos de origem do rendimento, com indicação do número e tamanho médio das famílias - Brasil - período 2017-2018

Origem do rendimento, número e tamanho médio das famílias	Coeficientes de variação do rendimento total e variação patrimonial médio mensal familiar (%)							
	Total	Classes de rendimento total e variação patrimonial mensal familiar (R\$) (
		Até 1 908 (2)	Mais de 1 908 a 2 862	Mais de 2 862 a 5 724	Mais de 5 724 a 9 540	Mais de 9 540 a 14 310	Mais de 14 310 a 23 850	Mais de 23 850
Rendimento total e variação patrimonial	1,7	0,4	0,2	0,2	0,3	0,3	0,6	3,4
Rendimento total	1,6	0,4	0,2	0,2	0,3	0,5	0,8	3,7
Rendimento do trabalho	2,1	1,3	1,1	0,7	1,0	1,6	2,3	5,0
Empregado	1,9	1,8	1,6	1,0	1,5	2,2	3,8	4,7
Empregador	10,4	19,8	11,3	7,3	6,5	8,5	9,4	16,8
Conta própria	2,4	2,3	2,6	2,1	3,4	6,5	7,7	11,0
Transferência	1,7	1,4	1,6	1,6	2,6	4,5	5,8	6,9
Aposentadoria e pensão do INSS	1,6	2,4	2,1	2,0	3,4	5,5	6,6	8,0
Aposentadoria, pensão da previdência pública	4,5	12,1	8,3	5,8	6,6	7,4	8,8	9,7
Aposentadoria, pensão da previdência privada	12,2	35,8	30,0	23,5	16,8	18,6	19,9	21,0
Programas sociais federais	2,2	2,6	3,7	4,4	9,6	20,9	30,9	54,4
Pensão alimentícia, mesada ou doação	4,7	4,6	6,5	6,0	11,4	21,2	24,3	29,0
Outras transferências	3,0	5,7	5,7	4,0	6,1	10,8	18,0	14,0
Rendimento de aluguel	10,8	13,1	11,2	7,7	7,2	12,2	13,2	20,8
Outras rendas	7,7	10,8	9,4	8,4	11,6	20,3	28,7	32,6
Rendimento não monetário	1,4	1,0	1,1	0,9	1,5	2,5	3,9	7,3
Variação patrimonial	5,9	4,8	4,5	3,3	4,2	7,7	9,7	10,4
Número de famílias	0,9	1,8	1,6	1,4	2,1	3,9	4,8	6,1
Tamanho médio da família (pessoas)	0,3	0,7	0,7	0,6	0,8	1,1	1,7	1,8

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa de Orçamentos Fam

Nota: O termo família está sendo utilizado para indicar a unidade de investigação da pesquisa, unidade de Consumo, conforme descrito na Introdução da publicação.

(1) Inclui os rendimentos monetário e não monetário e a variação patrimonial. (2) Inclusive sem rendimento.

Tabelas com indicador de intervalo de valor para coeficiente de variação ao lado da estimativa

Uma outra forma alternativa de apresentação dos erros amostrais que pretende minimizar o inconveniente do aumento de colunas nas tabelas de resultados para inclusão da medida de precisão e, ao mesmo tempo, permitir a avaliação da qualidade de maneira direta, sem a necessidade de observação de tabela separada, é a inserção de um indicador do intervalo de valor do coeficiente de variação ao lado de cada estimativa na tabela de resultados, por meio de uma letra.

Como o indicador proposto para o intervalo de valor é apenas uma letra, que deve ser posicionada ao lado da estimativa na tabela, apesar de ser necessário aumentar o número de colunas para a inserção dessa letra, as novas colunas têm largura que não interfere na visualização da tabela, não prejudicando, assim, a sua leitura. Além disso, a letra ao lado da estimativa mostra o seu nível de precisão de maneira rápida e objetiva.

Para essa forma de divulgação dos erros amostrais, cabem as observações já feitas sobre quantas e quais tabelas terão a medida de precisão apresentada.

Uma desvantagem de apresentar o coeficiente de variação dessa forma é não ser possível que o usuário construa diretamente intervalos de confiança ou obtenha o valor do erro padrão, necessário para uso em algumas análises, pois não tem disponível o valor estimado desse coeficiente, apenas o intervalo que compreende esse valor. Em todas as outras formas de divulgação, quando o valor da medida de precisão é disponibilizado, o usuário pode derivar as demais medidas que não foram apresentadas. Para contornar essa desvantagem, sugere-se utilizar o ponto médio do intervalo de valor.

A definição dos intervalos de valor do coeficiente de variação apresentados a seguir considerou as experiências de

institutos oficiais de estatística de outros países e a aplicação dessa forma de divulgação no IBGE. Foram propostas algumas versões, das mais detalhadas, que permitem uma análise mais fina da qualidade das estimativas, até versões simplificadas, que facilitam a análise por possuírem menos intervalos, permitindo maior familiaridade com os intervalos e suas respectivas letras indicadoras.

O quadro a seguir apresenta a versão dos intervalos de valor para os coeficientes de variação recomendada para ser utilizada nas operações estatísticas do IBGE.

Quadro 1 - Classificação das estimativas quanto à precisão

Indicador	Intervalo do coeficiente de variação (CV) (%)
Z	0
A	Até 5
B	Mais de 5 a 15
C	Mais de 15 a 30
D	Mais de 30 a 50
E	Mais de 50

Fonte: IBGE, Diretoria de Pesquisas.

Exemplo de uso na Pesquisa Industrial Anual - Empresa - PIA-Empresa

As pesquisas anuais por empresas, dentre elas a PIA-Empresa, divulgam os erros amostrais para as estimativas de algumas tabelas gerais, e a forma de apresentação de tais erros nessas pesquisas é feita por meio de um indicador do intervalo de valor do coeficiente de variação estimado para cada estimativa, tal como o quadro simplificado apresentado anteriormente. A Figura 5 reproduz tabela divulgada na publicação de resultados da PIA-Empresa 2018.

Figura 5 - Exemplo da divulgação de erro amostral na PIA 2018

Tabela 1.1 (parte) - Variáveis selecionadas das empresas industriais com 1 ou mais pessoas ocupadas, segundo as divisões de atividades - Brasil - 2017-2018

Códigos da CNAE 2.0	Divisões de atividades	Número de empresas ativas	Pessoal ocupado em 31.12
2017			
	Total	312 848 A	7 648 609 A
B	Indústrias extrativas	6 361 B	188 342 A
05	Extração de carvão mineral	16 A	3 980 A
06	Extração de petróleo e gás natural	41 A	2 910 A
07	Extração de minerais metálicos	239 B	76 827 A
08	Extração de minerais não-metálicos	5 867 B	78 061 A
09	Atividades de apoio à extração de minerais	198 A	26 564 A
C	Indústrias de transformação	306 487 A	7 460 267 A
10	Fabricação de produtos alimentícios	42 131 A	1 745 618 A
11	Fabricação de bebidas	2 971 A	169 727 A
12	Fabricação de produtos do fumo	160 A	18 153 A
13	Fabricação de produtos têxteis	9 636 B	258 613 A
14	Confeção de artigos do vestuário e acessórios	40 365 B	586 419 A
15	Preparação de couros e fabricação de artefatos de couro, artigos para viagem e calçados	9 545 B	343 795 A
16	Fabricação de produtos de madeira	12 534 B	178 774 A
17	Fabricação de celulose, papel e produtos de papel	3 832 B	189 516 A
18	Impressão e reprodução de gravações	16 010 B	117 845 A
19	Fabricação de coque, de produtos derivados do petróleo e de biocombustíveis	297 A	168 705 A
20	Fabricação de produtos químicos	7 929 A	311 562 A
21	Fabricação de produtos farmoquímicos e farmacêuticos	516 A	105 398 A
22	Fabricação de produtos de borracha e de material plástico	11 400 A	378 696 A
23	Fabricação de produtos de minerais não-metálicos	25 945 A	415 324 A
24	Metalurgia	2 597 A	205 803 A
25	Fabricação de produtos de metal, exceto máquinas e equipamentos	38 750 A	449 179 A
26	Fabricação de equipamentos de informática, produtos eletrônicos e ópticos	3 052 A	132 997 A
27	Fabricação de máquinas, aparelhos e materiais elétricos	4 325 A	207 589 A
28	Fabricação de máquinas e equipamentos	12 862 A	335 061 A
29	Fabricação de veículos automotores, reboques e carrocerias	5 115 A	430 450 A
30	Fabricação de outros equipamentos de transporte, exceto veículos automotores	1 038 A	83 696 A
31	Fabricação de móveis	19 102 B	260 287 A
32	Fabricação de produtos diversos	12 258 B	167 964 A
33	Manutenção, reparação e instalação de máquinas e equipamentos	24 117 B	199 096 A

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Serviços e Comércio, Pesquisa Industrial Anual - Empresa 2018.

Nota: Adotou-se como referencial de linha os códigos da CNAE 2.0.



Considerações finais

As conclusões de uma pesquisa por amostragem probabilística devem ser apoiadas nas estimativas produzidas, as quais têm uma variabilidade que é inerente ao processo de amostragem.

Para cada estimativa derivada da pesquisa, é possível obter uma estimativa do erro amostral, seja erro padrão, seja coeficiente de variação, ou intervalo de confiança, mesmo que o seu cálculo seja difícil de implementar. É importante ressaltar que, numa mesma pesquisa, geralmente de múltiplos propósitos, as estimativas geradas para parâmetros associados a variáveis distintas podem ter níveis de precisão diferentes, posto que a precisão também varia de acordo com o nível de desagregação da variável de interesse, seja de classificação, seja geográfico.

Assim, a avaliação e a divulgação dos chamados erros amostrais são fundamentais, pois do seu conhecimento decorre o grau de confiança nas conclusões analíticas que subsidiam a tomada de decisão. No caso específico de comparações entre estimativas ao longo do tempo ou mesmo entre localidades, é fundamental conhecer os erros amostrais ou a significância estatística para que as conclusões sobre as eventuais variações sejam corretamente interpretadas.

Vale destacar duas situações distintas, em relação ao cálculo e à divulgação de erros amostrais associados às estimativas, as quais fazem parte das práticas adotadas pelo IBGE:

- Dependendo do plano amostral usado para a realização da pesquisa e do estimador definido para a obtenção dos resultados, o IBGE escolhe e aplica um método para a es-

timação dos erros amostrais e disponibiliza informações necessárias para que os usuários repliquem os cálculos. Essas informações são obtidas a partir dos microdados das pesquisas e do sistema de metadados; e

- Em outros casos, o cálculo de variâncias é realizado utilizando-se um método que não pode ser replicado por usuários externos por questões relacionadas com o sigilo estatístico das informações individuais. Nessas situações, o usuário pode ser orientado a usar métodos aproximados para a obtenção de estimativas de erro amostral, quando possível.

Conforme mencionado, a escolha da medida de precisão a ser usada depende do público a que se destina e não carece de padronização, assim como a forma de divulgação. Para definição da medida e da forma, devem ser consideradas as características, bem como as vantagens e desvantagens de cada uma delas, apresentadas nesse guia, e em qual das duas situações descritas anteriormente a operação se enquadra.

Recomenda-se, apenas, que, em uma determinada operação estatística, seja utilizada a mesma medida de precisão, podendo ser adotada mais de uma forma de divulgação, dependendo da estimativa produzida. Assim, uma operação estatística que produz indicadores que são contagens de unidades com certa característica e que também produz médias ou totais de variáveis numéricas, por exemplo, pode divulgar a precisão dos primeiros por meio do modelo de função de variância, e, para os demais, pode apresentar os coeficientes de variação estimados correspondentes.

Apesar de não ser necessário padronizar a forma de divulgação dos erros amostrais, é recomendável que as áreas responsáveis por todas as operações estatísticas do IBGE façam um esforço para estimar o erro amostral de cada vez mais estimativas produzidas pelo Instituto, buscando como meta alcançar a sua

totalidade. Dessa maneira, permitirão uma avaliação mais detalhada da qualidade dos resultados dessas operações, tanto pela equipe, quanto pelos usuários de suas estatísticas.

Além disso, notas técnicas, relatórios metodológicos e informações do sistema de metadados, disponibilizados a cada divulgação de resultados das pesquisas por amostragem probabilística que o IBGE produz, devem conter a metodologia usada para a obtenção dos dados, das estimativas e de seus erros amostrais.

Referências

CONJUNTO mínimo de indicadores padrão de qualidade a ser aplicado no MERCOSUL. Rio de Janeiro: IBGE, 2014. 19 p. (Textos para discussão. Diretoria de Pesquisas, n. 52). Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv87506.pdf>. Acesso em: out. 2021.

IBGE. *Código de boas práticas das estatísticas do IBGE*. 2. ed. Rio de Janeiro, 2021. 67 p. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101744>. Acesso em: out. 2021.

IBGE. *Guia para a divulgação de erros amostrais associados às estimativas provenientes de pesquisas por amostragem probabilística, realizadas pelo IBGE*. Rio de Janeiro, 2016. Documento interno.

IBGE. *Guia para elaboração de relatórios metodológicos*. Rio de Janeiro, 2019. 20 p. Disponível em: <https://www.ibge.gov.br/institucional/codigos-e-principios.html>. Acesso em: out. 2021.

JURAN, J. M.; GRZYNA, F. M. *Quality planning and analysis: from product development through use*. 2nd. ed. New York: McGraw-Hill, 1980. 629 p.

LILA, M. F.; FREITAS, M. P. S. de. *Estimação de intervalos de confiança para estimadores de diferenças temporais na pesquisa mensal de emprego*. Rio de Janeiro: IBGE, 2007. 101 p. (Textos para discussão. Diretoria de Pesquisas, n. 22). Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/trabalho/9180-pesquisa-mensal-de-emprego.html?t=notas-tecnicas>. Acesso em: out. 2021.

STATISTICS CANADA. *Statistics Canada quality guidelines*. 5th ed. Ottawa, Oct. 2009. 89 p. Disponível em: <https://unstats.un.org/unsd/dnss/docs-nqaf/Canada-12-539-x2009001-eng.pdf>. Acesso em: out. 2021.

WOLTER, K. Generalized variance functions. *In: WOLTER, K. Introduction to variance estimation*. New York: Springer, 2007. p. 272-297.

Equipe técnica

Diretoria de Pesquisas

Coordenação de Métodos e Qualidade

Andréa Borges Paim

Gerência de Metodologia Estatística

Marcus Vinicius Morais Fernandes

Marcos Paulo Soares de Freitas

Centro de Documentação e Disseminação de Informações

Coordenação de Produção

Marisa Sigolo

Gerência de Editoração

Estruturação textual

Leonardo Ferreira Martins

Diagramação textual

Aline Carneiro Damacena

Programação visual

Thiago Goulart da Silva

Gerência de Documentação

Pesquisa e normalização documental

Aline Loureiro de Souza

Ana Raquel Gomes da Silva

Isabella Carolina do Nascimento Pinto

Lioara Mandoju

Nadia Bernuci dos Santos

Elaboração de quartas capas

Ana Raquel Gomes da Silva

Gerência de Gráfica

Ednalva Maia do Monte

Newton Malta de Souza Marques

Se o assunto é **Brasil**,
procure o **IBGE**.



/ibgecomunica



/ibgeoficial



/ibgeoficial



/ibgeoficial

www.ibge.gov.br 0800 721 8181

