



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DO SERTÃO PERNAMBUCANO**
Pró-Reitoria de Pesquisa, Inovação e Pós-Graduação (PROPIP)
Pós-Graduação Lato Sensu em Ensino de Ciências da Natureza e Matemática

**DIMENSIONAMENTO AMOSTRAL, MARGEM DE ERRO E
PRECISÃO EM PESQUISAS POLÍTICO-ELEITORAIS**

Edward Bernard Bastiaan de Rivera y Rivera

202227170001

Petrolina – PE

2024

Edward Bernard Bastiaan de Rivera y Rivera

DIMENSIONAMENTO AMOSTRAL, MARGEM DE ERRO E PRECISÃO EM PESQUISAS POLÍTICO-ELEITORAIS

Trabalho de Conclusão de Curso apresentado ao curso de Pós-Graduação Lato Sensu em Ensino de Ciência da Natureza e Matemática, ofertado pelo Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano, como parte dos requisitos para obtenção do título de Especialista em Ensino de Ciências da Natureza e Matemática.

Orientador: Prof. Me. Renan Fernandes de Moraes

Petrolina – PE

2024

Dados Internacionais de Catalogação na Publicação (CIP)

R622 Rivera y Rivera, Edward Bernard Bastiaan de.

Dimensionamento Amostral, Margem de Erro e Precisão em Pesquisas Político-Eleitorais / Edward Bernard Bastiaan de Rivera y Rivera. - Petrolina, 2024.
59 f.

Trabalho de Conclusão de Curso (Especialização em Ensino de Ciências da Natureza e Matemática) -Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano, Campus Pró-reitoria de Pesquisa, Inovação e Pós-graduação, 2024.
Orientação: Prof. Msc. Renan Fernandes de Moraes.

1. Ensino da Matemática. 2. Estatística. 3. Software R. 4. Objetos de Aprendizagem. 5. Planilhas Eletrônicas. I. Título.

CDD 372.7



ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

Na presente data realizou-se a sessão pública de defesa do Trabalho de Conclusão de Curso intitulada **DIMENSIONAMENTO AMOSTRAL, MARGEM DE ERRO E PRECISÃO EM PESQUISAS POLÍTICO-ELEITORAIS**, sob orientação de Renan Fernandes de Moraes, apresentada pelo aluno **Edward Bernard Bastiaan de Rivera y Rivera (202227170001)** do Curso **ESPECIALIZAÇÃO EM ENSINO DE CIÊNCIAS DA NATUREZA E MATEMÁTICA (Petrolina)**. Os trabalhos foram iniciados às 19h05 pelo Professor presidente da banca examinadora, constituída pelos seguintes membros:

- **Renan Fernandes de Moraes** (Presidente)
- **Luciana Nunes Cordeiro** (Examinadora Interna)
- **Junio Moreira de Alencar** (Examinador Externo)

A banca examinadora, tendo terminado a apresentação do conteúdo do Trabalho de Conclusão de Curso, passou à arguição do candidato. Em seguida, os examinadores reuniram-se para avaliação e deram o parecer final sobre o trabalho apresentado pelo aluno, tendo sido atribuído o seguinte resultado:

Aprovado

Reprovado

Nota (quando exigido): 100

Observação / Apreciações:

Proclamados os resultados pelo presidente da banca examinadora, foram encerrados os trabalhos e, para constar, eu **Renan Fernandes de Moraes** lavrei a presente ata que assino juntamente com os demais membros da banca examinadora.

Documento assinado digitalmente
gov.br JUNIO MOREIRA DE ALENCAR
Data: 03/07/2024 20:49:46-0300
Verifique em <https://validar.iti.gov.br>

Petrolina / PE, 26/06/2024

Documento assinado digitalmente
gov.br LUCIANA NUNES CORDEIRO
Data: 18/08/2024 10:03:58-0300
Verifique em <https://validar.iti.gov.br>

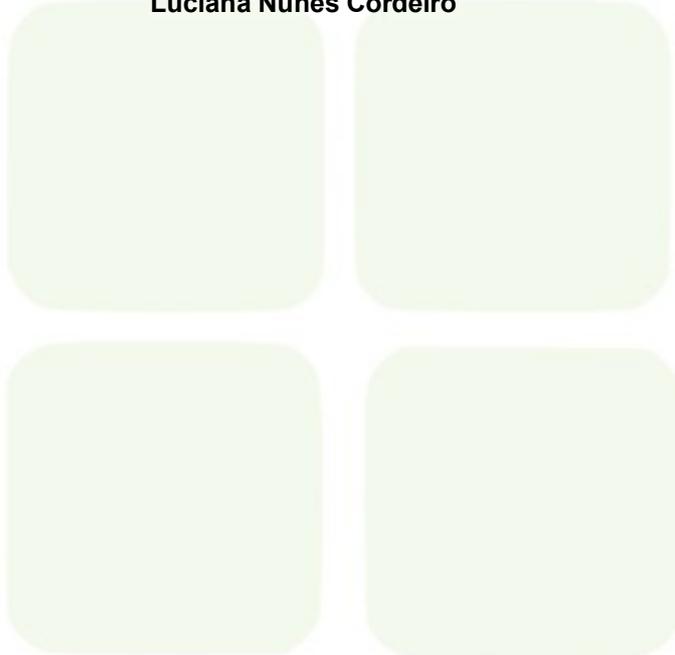
Junio Moreira de Alencar

Renan Fernandes de
Moraes:05397732397

Digitally signed by Renan Fernandes de Moraes 05397732397
DN: CN=Renan Fernandes de Moraes 05397732397, OU=IFSERTA/COPE - Instituto Federal do Sertão Pernambucano, O=IFPEdu, C=BR
Reason: I agree to the terms defined by the placement of my signature in this document
Location: Olinda - PE
Date: 2024.06.20 20:45:03-0300
Email: ifpe@educacao.pe.gov.br

Renan Fernandes de Moraes

Luciana Nunes Cordeiro



AGRADECIMENTOS

Quero expressar meu sincero agradecimento a Deus, o ser supremo que me concedeu a vida. Fui abençoado com talentos e oportunidades para seguir adiante em minha jornada acadêmica. Cada conquista e superação foram possíveis graças à Sua presença constante em minha vida. Espero poder continuar contando com a Sua proteção em todas as áreas da minha vida.

Maria Clotilde, é um privilégio ter uma mãe como você, que sempre compartilhou das minhas conquistas e me apoiou em momentos de desafios nessa área acadêmica. Desde minha tenra infância, você e minha vó Maria (*in memoriam*) foram mulheres fortes, lutando sozinhas, tanto financeiramente quanto emocionalmente, para me proporcionar uma base sólida. Foi graças a essa força e dedicação que pude chegar até aqui e minha caminhada continuará - com êxito.

Gostaria de expressar minha imensa gratidão pela oportunidade obtida no Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano - IFSertãoPE. Ser selecionado para fazer parte dessa instituição renomada e comprometida com a excelência educacional é uma honra e um privilégio. Agradeço de coração a todos os responsáveis pela minha admissão, bem como à equipe que trabalha incansavelmente para tornar o IFSertãoPE um ambiente acolhedor e propício ao aprendizado.

Gostaria de agradecer imensamente à Coordenadora Dr^a. Luciana Nunes Cordeiro por todo o seu apoio e dedicação. Sempre esteve presente, oferecendo suporte e solucionando dúvidas com paciência e clareza. Sua expertise e comprometimento foram essenciais para tornar a experiência educacional enriquecedora e proveitosa. Sou grato pela sua liderança inspiradora e pela forma como incentivou cada um de nós a alcançar nosso potencial máximo. Seu profissionalismo e carinho fizeram toda a diferença em minha formação como educador. Espero poder continuar contando com a sua orientação e sabedoria em minha jornada pessoal, profissional e acadêmica. Será sempre lembrada com carinho e admiração por tudo o que fez por mim e pelos demais colegas ao longo desse curso, com a honra de tê-la em minha banca com sugestões importantes que levarei para a minha vida.

Gostaria de expressar minha gratidão ao meu orientador, Prof. Me. Renan Fernandes de Moraes, que abraçou o tema que escolhi e por seu apoio e incentivo durante a pesquisa com motivação para que eu continuasse avançando. Sua paixão pela área reflete-se em seu trabalho, e sou grato por fazer parte do seu grupo de orientandos.

Gostaria de expressar minha sincera gratidão aos professores das disciplinas que tive o privilégio de cursar, Prof.^a Dr^a. Luciana Nunes Cordeiro (Introdução a EaD), Prof. Dr. Paulo Roberto Freire de Paula (Ensino por Investigação e Experimentação), Prof. Dr. Gabriel Kafure

da Rocha (História e Filosofia das Ciências), Prof. Dr. Ericleiton Rodrigues de Macedo (Teorias de Aprendizagem e Ensino de Ciências), Prof.^a Dr.^a Clecia Simone Gonçalves Rosa Pacheco (Interdisciplinaridade e Avaliação no Ensino de Ciências), Prof. Me. Mário Cezar Augusto de Almeida Bezerra (Metodologia Científica e Escrita Acadêmica), Prof. Dr. Francisco Kelsen de Oliveira (Tecnologias de Informação e Comunicação para o Ensino de Ciências da Natureza e Matemática), Prof. Me. Jean Lúcio Santos Evangelista (Metodologias Ativas), Prof.^a Dr.^a Danielle Juliana Silva Martins (Formação de Professores), Prof. Me. Deivid Andrade Porto (Tendências e Metodologias para o Ensino de Ciências Naturais e Matemática; Tópicos Especiais em Ensino de Ciências da Natureza e Matemática 1; Tópicos Especiais em Ensino de Ciências da Natureza e Matemática 2). A dedicação, conhecimento e paixão pelo ensino foram inspiradores e fundamentais para o meu aprendizado. Por meio do comprometimento de vocês, pude adquirir novas habilidades e conhecimentos que certamente farão diferença em minha vida pessoal e profissional. Muito obrigado por tornarem o processo de aprendizado significativo e enriquecedor. Sou extremamente grato por fazerem parte da minha jornada educacional. Agradeço ao Prof. Dr. Júnio Moreira de Alencar (IFCE) pela disponibilidade de participar de minha banca de defesa e seus comentários e sugestões.

RESUMO

O objetivo deste trabalho é analisar as pesquisas eleitorais mais amplamente divulgadas no 2º Turno das Eleições Presidenciais de 2022, considerando a construção de intervalos de confiança e suas premissas-chave, como o dimensionamento amostral, a margem de erro e o nível de confiança, investigando-se potenciais discrepâncias com avaliação estatística. Os dados foram coletados a partir do Sistema de Registro de Pesquisas Eleitorais (PesqEle) do Tribunal Superior Eleitoral (TSE) e a delimitação da amostra foi realizada por meio dos agregadores Poder360 e CNN. Assim, investiga-se em que medida as pesquisas eleitorais geram estimativas condizentes com as premissas divulgadas, adotando-se ferramentas tecnológicas como planilhas eletrônicas e software R, com procedimentos replicáveis no contexto empírico e educacional. Deste modo, duas lacunas são preenchidas. A primeira refere-se ao Ensino de Matemática e suas Tecnologias com foco em uma visão integrada aplicada aos diferentes aspectos da realidade, propondo-se a inserção da Estatística Inferencial logo no Ensino Médio, uma vez que proporciona habilidades analíticas que fortalecem o pensamento crítico, com comunicação não somente por símbolos e conectivos lógicos, mas por meio da linguagem corrente, bem como a situação-problema mais abordada pelas mídias sociais e profissionalmente, cujo entendimento é de suma importância à plena participação democrática. A segunda relaciona-se à pesquisa aplicada, acerca do uso da estatística na condução e divulgação de pesquisas eleitorais, investigando-se possíveis falhas que exercem influência sobre intervalos construídos, complementando-se a literatura nacional e internacional. Entre os resultados obtidos, os maiores investimentos por unidade amostral não garantem pesquisas que reflitam com precisão os resultados eleitorais; os institutos de pesquisa, em geral, adotam tamanhos amostrais que não correspondem à margem de erro divulgada e ao nível de confiança estabelecido, com discrepâncias percentuais de dois dígitos ou mais; no caso de amostras subdimensionadas, aproximadamente dois terços das estimativas intervalares não contemplam os verdadeiros parâmetros tanto sob margem de erro divulgada quanto efetiva; no caso de amostras superdimensionadas, cerca de dois terços das estimativas intervalares contêm a proporção real de votos válidos no respectivo candidato a partir da margem de erro divulgada, mas essa parcela cai para a metade com a margem de erro efetiva. Deste modo, ressalta-se tanto a importância do dimensionamento amostral adequado quando da representatividade amostral.

Palavras-chave: ensino da matemática; estatística; software R; objetos de aprendizagem; planilhas eletrônicas.

ABSTRACT

The objective of this study is to analyze the most widely reported electoral polls from the second round of the 2022 Brazilian Presidential Elections, focusing on the construction of confidence intervals and their key assumptions, such as sample size, margin of error, and confidence level, while investigating potential discrepancies through statistical assessment. Data were collected from the Electoral Poll Registration System (PesqEle) of the Superior Electoral Court (TSE), with the sample defined using aggregators Poder360 and CNN. The study examines the extent to which electoral polls generate estimates consistent with disclosed assumptions, employing technological tools such as spreadsheets and R software, with procedures replicable in both empirical and educational contexts. Consequently, two gaps are addressed. The first pertains to Mathematics Education and its Technologies, emphasizing an integrated approach to real-world applications and proposing the inclusion of Inferential Statistics into the Brazilian high school curriculum, as it fosters analytical skills that enhance critical thinking. This approach facilitates communication not only via symbols and logical connectors but also through current language, addressing the most frequently discussed problem situation in social media and professional settings, whose understanding is crucial for full democratic participation. The second relates to applied research on the use of statistics when conducting and reporting electoral polls, exploring potential distortions that impact the properties of confidence intervals, thereby contributing to both national and international literature. Among the results obtained, higher investments per sample unit do not ensure that polls accurately reflect electoral outcomes; research institutes typically adopt sample sizes that do not align with the disclosed margin of error and confidence level, leading to percentage discrepancies of two digits or more. Regarding undersized samples, approximately two-thirds of the interval estimates do not encompass the true parameter under both the disclosed and actual margins of error. As for oversized samples, about two-thirds of interval estimates include the true proportion of valid votes for the respective candidate based on the disclosed margin of error; however, this amount declines to half when considering actual margins of error. Thus, the importance of sample sizing and representativeness is underscored.

Keywords: mathematics education; statistics; R software; learning objects; spreadsheets.

LISTA DE TABELAS

Tabela 01 - Relação entre p^* , q^* e p^*q^* para n_p	25
Tabela 02 – Pesquisas Eleitorais: Início, Término, Registro e Divulgação.....	29
Tabela 03 – Tamanho Amostral, Nível de Confiança, Margem de Erro e Valores.....	30
Tabela 04 – Razão entre Valor (R\$) e Tamanho Amostral Efetivo (n_e).....	32
Tabela 05 – Ranking: Razão entre Valor (R\$) e Tamanho Amostral Efetivo (n_e).....	33
Tabela 06 – Variação Percentual entre Tamanho Amostral Previsto e Efetivo.....	36
Tabela 07 – Ranking: Variação Percentual entre Tamanho Amostral Previsto e Efetivo.....	37
Tabela 08 – Variação Percentual entre Margem de Erro Divulgada e Prevista.....	40
Tabela 09 – Ranking: Variação Percentual entre Margem de Erro Divulgada e Prevista.....	41
Tabela 10 – Estimativas Pontuais de Votos Totais e Válidos por Pesquisa.....	44
Tabela 11 – Precisão das Estimativas Intervalares em Relação ao Parâmetro Populacional na Perspectiva da Margem de Erro Divulgada (E_d): Lula.....	45
Tabela 12 – Precisão das Estimativas Intervalares em Relação ao Parâmetro Populacional na Perspectiva da Margem de Erro Divulgada (E_d): Bolsonaro.....	47
Tabela 13 – Precisão das Estimativas Intervalares em Relação ao Parâmetro Populacional na Perspectiva da Margem de Erro Efetiva (E_e): Lula.....	50
Tabela 14 – Precisão das Estimativas Intervalares em Relação ao Parâmetro Populacional na Perspectiva da Margem de Erro Efetiva (E_e): Bolsonaro.....	52

LISTA DE FIGURAS

Figura 01 – Ranking da Razão do Valor pelo Tamanho Amostral Efetivo ($R\$/n_e$)	34
Figura 02 – Ranking: Variações Percentuais Positivas (n_p e n_e).....	38
Figura 03 – Ranking: Variações Percentuais Negativas (n_p e n_e)	38
Figura 04 – Ranking: Variações Percentuais Negativas (E_d e E_p).....	42
Figura 05 – Ranking: Variações Percentuais Positivas (E_d e E_p).....	42

SUMÁRIO

1. INTRODUÇÃO	12
1.1 Tema e Justificativa	12
1.2 Problema de Pesquisa	14
1.3 Objetivos de Pesquisa	15
1.4 Organização do Trabalho	16
2. REFERENCIAL TEÓRICO	18
2.1 Trabalhos Seminais em Pesquisas Político-Eleitorais na Condução do Estudo	18
2.2 Contribuições em Pesquisas Político-Eleitorais: Contexto Brasileiro	19
2.3 Contribuições em Pesquisas Político-Eleitorais: Contexto Internacional	21
3. PROCEDIMENTOS METODOLÓGICOS	24
3.1 Técnica de Coleta de Dados.....	24
3.2 Procedimentos de Análise de Dados.....	24
4. RESULTADOS	28
5. CONSIDERAÇÕES FINAIS	53
REFERÊNCIAS	54
APÊNDICE A – Dimensionamento Amostral (n_p) em Planilha Eletrônica e Software R	56
APÊNDICE B – Margem de Erro Prevista (E_p) em Planilha Eletrônica e Software R	57
APÊNDICE C – Intervalo de Confiança em Planilha Eletrônica e Software R.....	58
APÊNDICE D – Teste de Hipótese em Planilha Eletrônica e Software R.....	59

1 INTRODUÇÃO

1.1 Tema e Justificativa

A Base Nacional Comum Curricular (BNCC) define os saberes essenciais a serem desenvolvidos ao longo da Educação Básica, caracterizando-se como um orientador curricular a partir do qual redes estaduais, municipais e do distrito federal se baseiam para elaboração e cumprimento dos Projetos Políticos Pedagógicos (PPPs), planos de aulas e materiais adotados nos estabelecimentos de ensino, com o objetivo de superação da fragmentação de políticas educacionais para um patamar comum de aprendizagem. Entretanto, uma consideração mais atenta ao documento evidencia que as competências e habilidades listadas não estabelecem um padrão de conteúdos mínimos para serem incorporados de forma estruturada, fornecendo margem para que sejam contemplados de forma superficial, insuficiente e circular. Assim, o documento tem limitações nos fundamentos e respectivo desenvolvimento ao longo dos anos.

Especificamente na área de Matemática e suas Tecnologias no Ensino Médio, para o qual este trabalho se direciona, considera-se o desenvolvimento de habilidades relativas à Estatística previstas na BNCC, compreendendo a interpretação de estatísticas divulgadas pela mídia, planejamento e execução de pesquisa amostral com o uso de tecnologias como planilhas eletrônicas como estimuladores do pensamento computacional. Assim, objetiva-se que o Ensino Médio tenha como foco a visão integrada da Matemática aplicada aos diferentes aspectos da realidade, considerando-se tanto as vivências cotidianas dos estudantes como os avanços tecnológicos e exigências do mercado de trabalho. Entretanto, o ensino de Estatística é fragmentado com conteúdos de probabilidade e estatística descritiva, sem interrelação para compreensão e interpretação de algo bem divulgado na mídia: as pesquisas político-eleitorais.

Em Estatística no Ensino Médio, conforme orientações da BNCC, as mesmas habilidades compõem todos os seus anos, envolvendo conteúdos de Probabilidade e Estatística Descritiva. Ao mesmo tempo, a Estatística Inferencial, essencial na interrelação entre ambas, não está presente na estrutura de habilidades explicitadas. Assim, conceitos importantes como margem de erro e intervalo de confiança, presentes na mídia pelas pesquisas de opinião, como notadamente nas pesquisas eleitorais, não são abordados ou considerados como aprendizagem essencial em nenhum momento. Deste modo, fundamentos de Estatística Inferencial, chaves na interrelação entre probabilidade e estatística descritiva, alinhados com a interpretação de estatísticas divulgadas pela mídia, pensamento computacional e mundo do trabalho, mantêm-se desconhecidos, o que repercute na efetividade da participação cidadã, pilar da democracia.

Este trabalho propõe a inserção da Estatística Inferencial como componente do Ensino de Matemática no Ensino Médio, adotando-se a situação-problema mais amplamente abordada pelas mídias sociais e profissionalmente, que são os intervalos de confiança para a proporção populacional, a partir de análises das principais pesquisas brasileiras para o 2º Turno das Eleições Gerais de 2022. Deste modo, busca-se tornar os procedimentos replicáveis por meio de planilha eletrônica para fins educacionais, estimulando-se o desenvolvimento do pensamento matemático e computacional, tendo a realidade e curiosidade como referências, no contexto das pesquisas eleitorais amplamente divulgadas, porém ainda pouco compreendidas pelos cidadãos, para o conhecimento deste ramo da Estatística e desenvolvimento de análises críticas sobre tamanho amostral, margem de erro e seus impactos na precisão das estimativas intervalares.

Portanto, este trabalho é desenvolvido prioritariamente para o Ensino de Matemática, em que se busca preencher a lacuna entre Probabilidade e Estatística Descritiva estimulando-se saberes pertinentes à Estatística Inferencial, inseridos nos contextos midiático e profissional, fortalecendo assim o pensamento analítico e crítico voltado à participação política-democrática. Favorecendo a Educação Aberta, com práticas que prezam a equidade e qualidade, priorizando softwares livres, foram desenvolvidas estruturas em planilhas eletrônicas e na linguagem aberta computacional R, situadas nos Apêndices com as respectivas notas explicativas, oportunizando a replicação dos procedimentos analíticos aqui conduzidos em qualquer outra eleição desejada. Tais recursos educacionais abertos (REAs) estão sob licença Creative Commons (CC BY-SA), permitindo distribuição e obras derivadas com atribuição do devido crédito ao autor original.

Outra lacuna que se busca preencher, relacionada à pesquisa aplicada, é sobre o uso da estatística na condução e divulgação de pesquisas eleitorais, investigando-se possíveis falhas que exercem influência sobre os intervalos de confiança construídos, permeando diferenças no dimensionamento amostral previsto a partir da margem de erro e nível de confiança propostos, diferenças na margem de erro divulgada em relação à margem efetiva a partir dos dados adotados, e respectivos impactos na precisão das estimativas intervalares, analisando-se, no contexto dessas diferenças estatísticas, a probabilidade de cada pesquisa eleitoral conter o verdadeiro parâmetro observado, ou seja, os resultados efetivos das eleições observados em relação aos votos válidos para cada candidato no 2º Turno das Eleições Gerais de 2022. Assim, as contribuições deste trabalho incluem o Ensino de Matemática e participação democrática.

1.2 Problema de Pesquisa

Uma das aplicações mais divulgadas da Estatística Inferencial refere-se às pesquisas de opinião apresentadas pela mídia regularmente em períodos eleitorais. Embora termos como “pontos percentuais para mais ou para menos” e “nível de confiança” sejam utilizados de forma corriqueira em diversos veículos de comunicação e mídias sociais para acompanhamento das estimativas de intenções de voto para cada candidato, como se fossem conceitos amplamente conhecidos pela sociedade, esses saberes não são tradicionalmente abordados na área de Matemática e suas Tecnologias no Ensino Médio. No entanto, ao se depararem com conceitos como desvio padrão, os discentes buscam estabelecer uma conexão com os termos aos quais foram expostos pela imprensa, tal como ocorre com a “margem de erro”. Mesmo em cursos subsequentes, não é raro esses conceitos não serem abordados nas disciplinas de Estatística.

Em um contexto em que não se estabelecem interrelações entre Estatística Descritiva e Probabilidade por meio da Estatística Inferencial, pouco ou superficialmente abordada nos diferentes níveis de ensino e cursos em que os conhecimentos estatísticos se fazem presentes, os saberes que fornecem conceitos e ferramentas-chave, úteis tanto no campo profissional quanto para atuação cidadã, são sistematicamente prescindidos no contexto educacional brasileiro. Assim, as possibilidades analíticas não são exploradas no ensino médio ou superior. Tradicionalmente e conforme a BNCC, em Estatística, enfatizam-se cálculos de medidas de tendência central (média, moda e mediana), de dispersão (amplitude, variância e desvio padrão), cálculos de probabilidade e gráficos como histograma, de caixa, ramos e folhas. Muitas vezes, de forma rotineira e distante de aplicações e análises estatísticas efetuadas em contexto real.

Consequentemente, esses conhecimentos, sem a articulação da Estatística Inferencial, não fornecem a compreensão do conteúdo estatístico mais divulgado pelos veículos de imprensa, que são as estimativas das pesquisas de opinião, com margem de erro e nível de confiança, em períodos eleitorais, uma vez que não se trabalha com este ramo da Estatística na área de Matemática e suas Tecnologias e, especificamente, com intervalos de confiança para a proporção que interrelaciona saberes de Estatística Descritiva e Probabilidade. Por outro lado, como esses conteúdos não são amplamente conhecidos pela sociedade, pesquisas eleitorais, embora questionadas quando distantes ou opostas dos resultados verificados, não passam pelo escrutínio cidadão sobre a lisura dos procedimentos e premissas divulgadas pelos institutos de pesquisa, diminuindo o potencial de interpretações corretas das estimativas e suas limitações.

Deste modo, visando à construção de uma visão integrada da Matemática aplicada à realidade, articulando-se seus diferentes campos, estabelece-se a seguinte questão-problema:

Em que medida as pesquisas eleitorais para o 2º Turno das Eleições Gerais de 2022 geram estimativas condizentes com o dimensionamento amostral e margem de erro divulgados, e como estimativas geradas influenciam a precisão e interpretação de intervalos de confiança?

Para responder esse problema de pesquisa, coletaram-se dados do Sistema de Registro de Pesquisas Eleitorais (PesqEle) do Tribunal Superior Eleitoral (TSE), delimitando-se a amostra de pesquisas eleitorais investigadas por meio dos agregadores Poder360 e CNN. Portanto, analisam-se diferenças entre dimensionamento amostral previsto e divulgado, na margem de erro, e os respectivos efeitos na construção e precisão dos intervalos de confiança resultantes.

1.3 Objetivos de Pesquisa

O objetivo deste trabalho é analisar as pesquisas eleitorais mais divulgadas para o 2º Turno das Eleições Gerais de 2022, considerando-se a construção de intervalos de confiança e suas premissas-chave, como dimensionamento amostral, margem de erro e nível de confiança, investigando-se potenciais discrepâncias entre informações divulgadas pelos institutos de pesquisa e dados implícitos evidenciados a partir da avaliação estatística. Deste modo, organizam-se as pesquisas por término e ranqueamento, investigando-se recursos financeiros aplicados por unidade amostral, diferenças percentuais entre tamanho amostral previsto pela estatística e efetivamente adotado, assim como diferenças entre margem de erro divulgada e implícita a partir dos dados utilizados. Em seguida, desenvolvem-se intervalos de confiança e testes de hipóteses, avaliando se estimativas contemplam ou não os parâmetros populacionais.

No âmbito educacional e de pesquisa aplicada, perseguem-se os objetivos específicos:

- i) Investigar a razão entre valores empregados nas pesquisas e tamanho amostral efetivo.
- ii) Analisar as variações percentuais entre tamanhos amostrais previstos calculados e efetivos.
- iii) Analisar as variações percentuais entre margens de erro divulgadas e efetivas calculadas.
- iv) Construir estimativas intervalares a partir das margens de erro divulgadas e efetivas.
- v) Avaliar probabilisticamente a hipótese de cada pesquisa conter o parâmetro populacional.

No desenvolvimento deste trabalho, preza-se pela possibilidade de replicação dos resultados a partir de planilhas eletrônicas, de forma a estimular os discentes no desenvolvimento do pensamento computacional como preconizado pela BNCC em Matemática e suas Tecnologias no Ensino Médio. Adicionalmente, este estudo, que interrelaciona aspectos teóricos e operacionais, também exerce contribuição para níveis subsequentes como o Ensino Superior, assim como àqueles com curiosidade sobre o desenvolvimento de pesquisas eleitorais e identificação de potenciais falhas na sua condução, cujos benefícios envolvem interpretações

mais adequadas das estimativas, distinção entre limitações estatísticas e falhas nos procedimentos, assim como utilização de ferramentas analíticas desenvolvidas neste trabalho para investigação de resultados *vis-à-vis* as premissas divulgadas em cada pesquisa eleitoral.

1.4 Organização do Trabalho

Este trabalho estrutura-se em cinco etapas. Na primeira, analisa-se a razão entre os valores (R\$) empregados pelo tamanho amostral efetivo de cada pesquisa eleitoral. Trata-se de um tratamento preliminar dos dados coletados no Sistema de Registro de Pesquisas Eleitorais (PesqEle) do Tribunal Superior Eleitoral (TSE), de forma a se observar os custos relativos de cada sondagem, buscando-se possíveis padrões como institutos e/ou contratantes que empregam valores por entrevista mais/menos consideráveis, assim como períodos ao longo da sequência de sondagens em que somas mais/menos elevadas foram despendidas na condução de pesquisas de opinião, caso identificado tal padrão temporal. Em seguida, para visualização mais clara dos dados calculados, constrói-se um ranking de valores empregados, com número de identificação, data de término, instituto e contratante das respectivas pesquisas analisadas.

Na segunda etapa de análise, calculam-se os dimensionamentos amostrais previstos por meio de ferramentas estatísticas a partir de informações como proporção esperada, margem de erro e nível de confiança. Comparando-se os tamanhos amostrais gerados analiticamente com aqueles efetivamente adotados em cada pesquisa eleitoral, analisam-se as diferenças percentuais envolvendo estas variáveis, a partir das quais desenvolve-se um ranking em ordem decrescente com a magnitude e sinal das variações encontradas entre o dimensionamento amostral previsto por meio de ferramentas analíticas e efetivamente empregado, buscando-se possíveis padrões entre institutos de pesquisa. Ressalta-se a importância desse procedimento, cujos impactos repercutirão nas análises posteriores sobre margem de erro, em que se esperam resultados diametralmente opostos em relação aos verificados no dimensionamento amostral.

Na terceira etapa, calculam-se as margens de erro efetivas por meio das ferramentas analíticas desenvolvidas a partir de informações como proporção esperada, tamanho amostral e nível de confiança. Deste modo, estrutura-se um ranking com os valores de margem de erro calculados, evidenciando-se diferenças em relação às margens divulgadas oficialmente. Assim, espera-se que, considerando-se a relação inversa entre margem de erro e tamanho amostral, as maiores diferenças percentuais positivas (negativas) entre margem de erro divulgada e efetiva sejam identificadas nas pesquisas eleitorais em que se verifiquem as maiores diferenças percentuais negativas (positivas) entre dimensionamento amostral previsto e efetivo. Portanto,

um ranking em ordem crescente das discrepâncias entre as margens de erro divulgadas e efetivas manterão pesquisas nas mesmas posições, porém com magnitudes e sinais diferentes.

Na quarta etapa, calculam-se as estimativas de votos válidos e constroem-se intervalos de confiança primeiramente a partir das margens de erro divulgadas oficialmente, buscando-se investigar possíveis padrões quando o verdadeiro parâmetro populacional, isto é, os resultados verificados no 2º Turno das Eleições Gerais de 2022, está incluso ou não no intervalo estimado, uma vez que dimensionamentos amostrais abaixo (acima) do previsto estatisticamente geram margens de erro efetivas acima (abaixo) daquelas divulgadas pelos institutos, impactando nos intervalos de confiança e seu potencial em contemplar ou não o verdadeiro parâmetro. Em seguida, efetuam-se os mesmos procedimentos a partir das margens de erro efetivas previamente calculadas estatisticamente. Espera-se que pesquisas com subdimensionamento amostral apresentem maior parcela de intervalos de confiança sem o parâmetro populacional.

Na quinta etapa, efetua-se uma análise probabilística com base em testes de hipóteses, cujos resultados possuem relação com os intervalos de confiança e análises sobre se o parâmetro populacional, isto é, os verdadeiros resultados verificados nas Eleições, estão “dentro” ou “fora” das estimativas intervalares por meio dos dados provenientes das pesquisas eleitorais. Entretanto, um diferencial dos testes de hipóteses, baseados em estatísticas padronizadas construídas a partir das informações sobre dimensionamento amostral efetivo e estimativas de votos válidos para cada pesquisa, é permitir o cálculo preciso sobre a probabilidade de cada pesquisa eleitoral, por meio das informações-chave mencionadas, em rejeitar um determinado valor hipotético estabelecido quando este é verdadeiro. Nesse contexto, os valores hipotéticos estabelecidos serão os percentuais de cada candidato no 2º Turno das Eleições Gerais 2022.

2 REFERENCIAL TEÓRICO

2.1 Trabalhos Seminais em Pesquisas Político-Eleitorais na Condução do Estudo

Dugas (1991), no artigo “A Problemática das Pesquisas Político-Eleitorais: o Currículo de Matemática para a Compreensão Social”, salienta o uso amplo de dados estatísticos ao longo de campanhas políticas, ao mesmo tempo em que verifica a incapacidade do público em compreender adequadamente as informações divulgadas. Dentre as razões apresentadas para esse fenômeno estão a descontextualização e impregnação positivista no Ensino de Matemática, que não demonstra aos discentes as maneiras nas quais estes saberes afetam suas vidas, direta ou indiretamente. Deste modo, a autora defende mudanças curriculares a partir da concepção de uma educação tecnológica que revele a natureza política da Matemática, ao contrário da sua abordagem como saberes “neutros”, de forma a se compreender origens e implicações dos cálculos na vida social e política, a exemplo de técnicas estatísticas nas campanhas eleitorais.

Nesse contexto, Dugas (1991) ressalta que a extensa divulgação das pesquisas eleitorais está imbricada em uma teia de poder, ignorando-se o fato de que não somente o público, mas muitos profissionais das mídias não compreenderem os jargões estatísticos. Embora o ambiente analítico seja o ensino norte-americano, há semelhanças no caminho predominante percorrido pelo discente na área de Matemática e suas Tecnologias, na qual a Estatística não é atribuída importância semelhante aos saberes de Álgebra, Geometria, Trigonometria e mesmo Cálculo. Entretanto, uma pessoa não é matematicamente instruída ao desconhecer a Estatística, cuja ausência dos saberes ou descontextualização é negação de sua compreensão social. Portanto, espera-se as tecnologias facilitem a transição da linguagem abstrata vigente da Matemática para a compreensão de conceitos, aplicações e interpretações, com seu respectivo caráter humano.

Outro trabalho-chave no desenvolvimento do presente estudo é a investigação de Ferreira (2012), “Como Contestar Pesquisa Eleitoral?”, que salienta a necessidade de escolas e universidades contemplarem de forma sistemática o ensino das ferramentas estatísticas utilizadas na construção das estimativas intervalares para a proporção, averiguando se o dimensionamento amostral preenche as premissas divulgadas como margem de erro e nível de confiança, indo além das abordagens isoladas de probabilidade e estatística descritiva, predominantes no Ensino Médio e subsequentemente. Assim, à semelhança das concepções de Dugas (1991), defende-se a disseminação desses conhecimentos estatísticos de forma que a sociedade, deparada com as pesquisas político-eleitorais, saiba interpretá-las adequadamente usando informações divulgadas e exerça avaliações embasadas de seus atributos e limitações.

Ferreira (2012) verifica que o tamanho das amostras adotadas pelas pesquisas eleitorais, no contexto observado para o Governo Municipal de Goiânia de 2012, não é suficiente para extração dos resultados divulgados, uma vez que o nível de confiança e margem de erro estabelecidos pelos institutos correspondem a uma amostra relativamente maior, entrevistando-se, portanto, uma quantidade menor de eleitores do que seria ideal a partir dos critérios supostamente utilizados. Deste modo, cálculos são efetuados para demonstrar qual seria o tamanho amostral ideal de acordo com os níveis de confiança e margens de erro adotadas e, de forma semelhante, a partir do dimensionamento amostral adotado, calculam-se as margens de erro que deveriam ter sido utilizadas. Assim, falhas nos procedimentos comprometem a margem de erro e a construção de intervalos de confiança das estimativas pontuais para as proporções.

2.2 Contribuições em Pesquisas Político-Eleitorais: Contexto Brasileiro

Trabalhos estatísticos em relação a pesquisas político-eleitorais brasileiras permanecem escassos na literatura, centrando-se em poucos pesquisadores. Como maneiras de explicar resultados observados frequentemente fora das margens de erro divulgadas, as temáticas compreendem, embora de forma não sistemática, o dimensionamento amostral incompatível com a margem de erro e nível de confiança previamente estabelecidos, possíveis instabilidades nas preferências eleitorais a partir de contextos políticos, diferenças entre eleições municipais e nacionais, em que as primeiras possuem tipicamente menos entrevistas no plano amostral, e análises de práticas irregulares de véspera, como boca de urna, não captadas pelas pesquisas, mas que influenciam resultados e erros observados, assim como discrepâncias nas eleições legislativas com menor cobertura e em regiões de arranjos de poder familiares e oligárquicos.

A partir da investigação de pesquisas para as Eleições Municipais de 2012, Ferreira (2013) reflete que uma das principais falhas de institutos não terem êxito em suas pesquisas políticas está centrada no tamanho da amostra, adotando-se dimensionamento amostral inadequado para os critérios estabelecidos. Ressalta que, como operam com margem de erro de $\pm 2\%$ a $\pm 5\%$, o que significa uma folga de intervalo de 4% a 10%, um parâmetro fora deste intervalo é ocorrência de erro. Deste modo, registram-se falhas na coleta de dados a partir dos principais indicadores de pesquisas de opinião: tamanho amostral inadequado à margem de erro, diretamente relacionado ao tamanho amostral e ao valor crítico da distribuição normal padrão, obtido a partir do nível de confiança originalmente estabelecido. Assim, erros acima do estatisticamente aceitável decorre por falhas no dimensionamento amostral e margem de erro.

Examinando-se pesquisas para presidente e governadores em 2010, Gramacho (2015a), usando a média do erro/desvio absoluto entre intenções e votos válidos, verifica que os institutos de pesquisa estimaram resultados fora das margens de erro informadas à Justiça Eleitoral, com erro médio situando-se um ponto percentual acima da média das margens informadas às autoridades eleitorais. Entretanto, não se observa um padrão generalizado de erros, ou concentração destes em poucos levantamentos, verificando-se que mais da metade das pesquisas incorreram em erro nas estimativas acima do previsto. De forma complementar, analisa-se que sondagens concluídas sete dias antes tenderam a erros verificados maiores do que outras, com as mesmas características, com término na véspera. Estimando-se o desvio, negativo/positivo, não se verificou desfavorecimento sistemático para nenhum grupo político.

Em análise das eleições brasileiras de 2014, Gramacho (2015b) reflete a problemática dos questionamentos de atores políticos sobre a qualidade das pesquisas eleitorais. Embora observe a intenção de votos como retrato propenso ao voto estratégico, influência de fatos novos como debates ou pesquisas mais recentes, e atitudes individuais, verifica que algo ignorado é a discrepância entre erro amostral, diferença entre estatística amostral e parâmetro populacional, e o erro total, acima desta margem de erro prevista. A partir de cada pesquisa, verifica-se a média dos valores absolutos da diferença da intenção de votos válidos em cada candidato e seu respectivo percentual de votos válidos. Na interpretação de discrepâncias, considera-se a instabilidade nas preferências eleitorais no contexto abordado, mas reitera a precisão amostral dada pela combinação associada entre margem de erro e intervalo de confiança estabelecidos.

Meireles e Russo (2022) ressaltam o tamanho amostral como fator de influência na precisão das estimativas. Analisam que, em anos de eleição municipal, a maioria dos planos amostrais previa cerca de 450 entrevistas, resultando em margem de erro de aproximadamente 4,6%, supondo um nível de confiança de 95%, exceto em pesquisas conduzidas em grandes capitais, por alguns institutos tradicionais ou em eleições nacionais e estaduais. Assim, dentre fatores relacionados com erros nas pesquisas, o primeiro centra-se no tamanho das amostras adotadas, com impacto direto na margem de erro das estimativas geradas, enquanto um fator secundário relaciona-se com atitudes de eleitores que se inclinam para um lado ou outro a partir de novas informações ou com base em quem está na frente das pesquisas. Deste modo, quanto maior o tamanho amostral e proximidade temporal, maior a tendência à minimização de erros.

Estendendo-se a discussão sobre erros nas pesquisas brasileiras, Lloyd e Turgeon (2021) salientam que práticas irregulares tardias, como a boca de urna, não é detectável nas sondagens em si, mas pelas diferenças entre as pesquisas de véspera e resultados eleitorais. Neste sentido, argumenta-se que se trata de um fenômeno ainda pouco estudado pela comunidade acadêmica.

A partir do erro das pesquisas por meio da média dos desvios absolutos entre votos previstos e reais para cada candidato, os autores verificam erros maiores nas pesquisas em eleições com muitos eleitores indecisos e desequilíbrios financeiros entre candidatos, e em regiões nas quais a prática é historicamente mais comum, a partir de arranjos familiares e oligárquicos de poder. Adicionalmente, erros de pesquisas são maiores em eleições legislativas do que nas executivas, no contexto de muitos candidatos para os quais há menos informações e cobertura pela mídia.

No intuito de se atingir grau mais elevado de precisão, Brito e Adeodato (2023) adotam inteligência artificial para a previsão de resultados. Assim, embora reconheçam as pesquisas eleitorais como mecanismos “de facto” para esse fim, salientam-se vantagens (e desafios) no uso de informações provenientes das mídias sociais, como volume massivo de informações em tempo real, alcance, velocidade e baixo custo em relação às sondagens pessoais ou por telefone. Porém, há carência e discussão remanescente de processos bem definidos e replicáveis, sujeitos à manipulação de informações. Deste modo, aplica-se um modelo de aprendizagem de máquina às eleições presidenciais do Brasil (2018), Argentina (2019), Colômbia (2018) e México (2018) a partir de dados de postagens nos perfis dos candidatos no Twitter, Facebook, e Instagram, assim como pesquisas eleitorais, constatando-se alto grau de precisão em se prever resultados.

2.3 Contribuições em Pesquisas Político-Eleitorais: Contexto Internacional

Embora não haja um modelo ou estrutura única considerada universal na análise de pesquisas político-eleitorais, um ponto de inflexão foram as sondagens das eleições de 1948 nos Estados Unidos, nas quais institutos como Gallup e Crossley apontavam vitória de Thomas Dewey (Republicano) sobre Harry Truman (Democrata) por 5 pontos percentuais em uma corrida em que Truman ganhou por 4 pontos, representando um erro de 9 pontos. Neste sentido, um dos métodos clássicos propostos por Mosteller et al. (1949) consistiu em calcular a média das diferenças absolutas entre a estimativa e voto real de cada candidato, utilizado até os dias de hoje. Vertentes mais recentes apontam ao uso de outras informações além das pesquisas eleitorais com o objetivo de se alcançar maior precisão nas previsões, como dados de redes sociais e mercados de previsão, ou a associação destes com as próprias pesquisas disponíveis.

Caballé (2013) avalia as estimativas dos resultados das eleições ao parlamento espanhol publicados nos principais jornais nacionais. O grau de precisão é medido pela distância entre as previsões e os resultados nas urnas, avaliando-se a relação entre tamanho amostral e qualidade da estimação, assim como a relação da direção do erro com a inclinação política do respectivo veículo. Ao avaliar o tamanho amostral informado, sugerindo-se maior esforço e

orçamento, não se observa relação entre qualidade preditiva com tamanho amostral ou margem de erro divulgadas pelos periódicos. No entanto, verifica-se relação negativa esperada entre maior dimensionamento amostral documentado e margem de erro, com aderência semelhante, porém não idêntica, à prevista pelas ferramentas estatísticas. Adicionalmente, não há evidência clara de inclinação de erro a determinado partido associado a afinidades políticas de veículos.

Graefe (2023) argumenta que a combinação de informações de diferentes naturezas tende a reduzir erros de previsão, como a associação de estimativas das próprias pesquisas eleitorais com as geradas por mercados preditivos de apostas, julgamento de especialistas e modelos econométricos. O autor mostra que o agregador dessa diversidade de informações, denominado Pollyvote, nas eleições presidenciais dos Estados Unidos de 2020, apresentou média de erros absolutos, calculadas para previsões dos últimos 100 dias até as eleições, de 0,2 pontos percentuais, caracterizando maior precisão do que cada método ou previsões específicas. Ao considerar as eleições presidenciais de 2004 a 2020, o agregador representou média absoluta de erros de 0,8, inferior a qualquer método e 47% abaixo do valor de 1,5 das pesquisas eleitorais. Salienta que a combinação de informações minimiza erro de sondagens/métodos individuais.

Zhou et al. (2021) refletem sobre falhas sistemáticas das pesquisas tradicionais em prever os resultados das eleições presidenciais no mundo. Focando-se nas eleições argentinas de 2019, investigam sobre os motivos pelos quais as sondagens falharam de forma significativa. Os autores relatam falhas na representação demográfica e inconsistências nas respostas dos entrevistados, compreendendo a tendência de não se expressar a verdade sobre preferências em determinado candidato. Deste modo, propõem um método de inteligência artificial na qual se extraem opiniões das redes sociais, como o Twitter, reestruturando-se os pesos dos respondentes de acordo com o censo nacional, de forma a se evitar viés de respondentes mais jovens. Assim, diferentemente das pesquisas tradicionais, capta-se o grau de lealdade para um candidato ao longo do tempo, demonstrando precisão nas estimativas do resultado observado.

Avaliando-se como contexto influencia a precisão das pesquisas eleitorais, Sohlberg e Branham (2020) avaliam como características específicas de eleições podem torná-las mais fáceis ou difíceis de prevê-las. Neste sentido, estimam como erros são impactados pelo grau de participação, mudança eleitoral e compra de votos em uma variedade de pesquisas de 44 países ao longo de 71 anos, iniciando-se em 1942, associando-se com dados sobre graus de compra de votos. Assim, a métrica de erros são os respectivos desvios absolutos entre os resultados nas urnas e a média das estimativas das sondagens para cada partido no final da campanha. Os resultados sugerem que o grau de participação não está sistematicamente associado ao erro,

porém volatilidade eleitoral e compra de votos exercem impacto significativo nesta variável. Portanto, quanto menor a compra de votos e instabilidades no cenário, melhor para pesquisas.

Investigando as eleições gerais dos Estados Unidos, Panagopoulos (2021) usa três medidas: a média de diferenças absolutas entre a estimativa e resultado para cada candidato, também conhecida como “Medida de Mosteller 3”; a diferença entre duas margens, em que a primeira é entre as estimativas de voto dos candidatos líderes na pesquisa, e a segunda é entre os resultados reais das urnas, conhecida como “Medida de Mosteller 5”; e método alternativo desenvolvido por Martin, Traugott e Kennedy (2005) baseado no logaritmo natural da razão de intenções de voto de um candidato e outro sobre a razão de seus respectivos votos nas urnas, conhecida como “Medida A”, com a vantagem de que um sinal positivo indica viés para o primeiro candidato e negativo, ao segundo. Os resultados indicam que as eleições de 2020 estão entre as de menor precisão dos últimos 25 anos, e viés sistemático detectado aos Democratas.

3. PROCEDIMENTOS METODOLÓGICOS

3.1 Técnica de Coleta de Dados

Utilizaram-se dados do Sistema de Registro de Pesquisas Eleitorais (PesqEle) do Tribunal Superior Eleitoral (TSE) referentes ao 2º Turno das Eleições Gerais Brasileiras de 2022, em que os institutos devem registrar suas respectivas pesquisas até 05 (cinco) dias antes da divulgação, conforme Lei nº 9.504/1997 e Resolução nº 23.600/2019 do TSE, delimitando-se então a amostra por meio da listagem nos agregadores Poder360 e CNN para análise das sondagens mais divulgadas, perfazendo 47 (quarenta e sete) registros. Assim, dados coletados de pesquisas político-eleitorais, no período de término de 05/10 a 29/10/2022, para estruturação e aplicação das ferramentas estatísticas são: números de identificação (ID); institutos executores e contratantes; datas de início, término, registro e divulgação; níveis de confiança (c); margens de erro divulgadas (E_d); tamanhos amostrais efetivos (n_e); e valores totais empregados (R\$).

3.2 Procedimentos de Análise de Dados

Seguindo-se a organização do trabalho proposta, analisa-se inicialmente a razão entre valores investidos nas pesquisas e tamanho amostral efetivo ($R\$/n_e$), ranqueando-se as pesquisas nas quais se empregam um maior montante por unidade amostral em busca de possíveis padrões em relação a institutos responsáveis, contratantes e/ou datas de término das sondagens. Caracteriza-se como uma análise inicial exploratória, uma vez que montantes mais significativos não são, necessariamente, lastreados em dimensionamento amostral adequado à margem de erro e nível de confiança propostos. Procede-se com uma representação gráfica desse ranqueamento, classificando-se de forma decrescente as pesquisas em relação aos valores empregados por unidade amostral, permitindo-se estabelecer interrelações subsequentes sobre presença potencial de discrepância entre dimensionamento efetivo e previsto estatisticamente.

Na segunda etapa, analisam-se as variações percentuais entre os tamanhos amostrais previstos calculados e efetivos. Nesse sentido, calcula-se, para cada pesquisa político-eleitoral, o dimensionamento amostral previsto (n_p) a partir da margem de erro (E_d) e nível de confiança (c) estabelecidos, investigando-se discrepâncias positivas ou negativas entre dimensionamento amostral previsto e efetivo, $\Delta\% = n_e/n_p - 1$, a partir da determinação do tamanho amostral:

$$n_p = p \cdot q \cdot (z_c/E_d)^2 \quad (1)$$

em que p^* é a estimativa preliminar da proporção amostral \hat{p} de certo atributo; $q^* = 1 - p^*$, que significa a parcela que não possui tal atributo em questão; e z_c , o valor crítico encontrado na Tabela Normal Padrão (ou software), sendo o valor de Z , em módulo, que contém, na cauda superior da distribuição, área de probabilidade equivalente à metade do nível de significância ($\alpha/2$) e, entre este e seu análogo negativo, área igual ao nível de confiança ($1 - \alpha$) escolhido.

No contexto deste trabalho, a característica é a proporção amostral com intenção de voto em determinado candidato. Na prática, o valor de \hat{p} é desconhecido antes análise amostral. Assim, adota-se uma estimativa preliminar que resulte no maior dimensionamento amostral possível e que, conseqüentemente, seja suficiente para atender as premissas da margem de erro e nível de confiança estabelecidos. A lógica de se escolher $p^* = 0,5$ tem o seguinte raciocínio:

Tabela 01 – Relação entre p^* , q^* e p^*q^* para n_p

p^*	q^*	p^*q^*
0,10	0,90	$(0,10)(0,90) = 0,09$
0,30	0,70	$(0,30)(0,70) = 0,21$
0,50	0,50	$(0,50)(0,50) = 0,25$
0,70	0,30	$(0,70)(0,30) = 0,21$
0,90	0,10	$(0,90)(0,10) = 0,09$

Nota: Valores hipotéticos p^* , q^* e p^*q^* resultantes na aplicação em $n_p = p^*q^*(z_c/E_d)^2$ para determinação do dimensionamento amostral previsto. p^* é a estimativa preliminar da proporção amostral \hat{p} ; $q^* = 1 - p^*$; z_c , o valor crítico de Z correspondente ao nível de confiança ($1 - \alpha$) estabelecido; e E_d , a margem de erro divulgada.
Fonte: Elaborado pelo autor.

Deste modo, a adoção de $p^* = 0,50$ como estimativa preliminar da proporção amostral \hat{p} garante que o dimensionamento amostral não seja subestimado, fornecendo o maior valor de n_p que atenda as premissas de margem de erro e nível de confiança estabelecidos. Usualmente, em um planejamento amostral, usa-se tal valor como estimativa preliminar de \hat{p} para um intervalo de confiança da proporção populacional p com margem de erro e precisão almejadas.

Na terceira etapa, investigam-se as variações percentuais entre margens de erro divulgadas e as estatisticamente previstas. Assim, para cada pesquisa político-eleitoral, calcula-se a margem de erro prevista (E_p) a partir do tamanho amostral efetivo (n_e) e nível de confiança (c) informados, além da estimativa preliminar $p^* = 0,5$, analisando-se discrepâncias negativas ou positivas entre margem de erro divulgada e prevista, $\Delta\% = E_p/E_d - 1$, usando-se o cálculo:

$$E_p = z_c \sqrt{p^*q^*/n_e} \quad (2)$$

em que o termo p^*q^* é o produto das estimativas preliminares da proporção populacional p com e sem a característica-alvo analisada, adotadas como estimativas planejadas de $\hat{p}\hat{q}$; z_c , o valor

crítico, em módulo, na Distribuição Normal Padrão, cuja área de probabilidade na cauda à direita equivale à metade do inverso do nível de confiança; e n_e , o tamanho amostral efetivo.

Na sequência, constroem-se estimativas intervalares com margens de erro divulgadas (E_d) e efetivas (E_e), conforme abaixo. Deste modo, os intervalos de confiança são construídos e analisados com dados divulgados nas pesquisas político-eleitorais de caráter nacional mais disseminadas na mídia, assim como margens de erro efetivas calculadas estatisticamente, investigando-se interrelações entre precisão e discrepâncias verificadas nos dimensionamentos.

$$IC_{ij} = \hat{p}_{ij} \pm E_d \quad (3)$$

$$IC_{ij} = \hat{p}_{ij} \pm E_e \quad (4)$$

em que IC_{ij} é o intervalo de confiança da pesquisa i para o candidato j ; \hat{p}_i , a proporção verificada na amostra de intenção de voto em determinado candidato; E_d , a margem de erro divulgada no registro no PesqEle; E_e , a margem de erro efetiva, $z_c \sqrt{\hat{p}\hat{q}/n_e}$, em que $\hat{p}\hat{q}$ é o produto do valor das proporções amostrais de p e seu inverso; e z_c , o valor crítico para dado nível de confiança, investigando se os verdadeiros parâmetros (resultados eleitorais) estão contidos nos intervalos.

Na quinta etapa, avalia-se probabilisticamente a hipótese de cada pesquisa conter o parâmetro populacional, isto é, a proporção de votos válidos verificada para cada candidato. Deste modo, testa-se a hipótese bilateral das estimativas das sondagens serem estatisticamente iguais ou diferentes à verdadeira proporção de votos em cada candidato. A hipótese nula testada é $H_0: p = p_0$ contra a hipótese alternativa $H_a: p \neq p_0$, a partir da estatística de teste Z a seguir:

$$Z_{ij} = \frac{\hat{p}_{ij} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n_e}}} \quad (5)$$

em que \hat{p}_{ij} é a proporção amostral de intenção de voto em determinado candidato; p_0 , a proporção-alvo no teste de hipótese que, neste contexto, é o próprio resultado da eleição em relação a votos válidos verificados para cada candidato à presidência. Assim, a estatística Z fornece a distância padronizada entre a proporção amostral (\hat{p}) e a proporção no resultado das eleições (p_0). Quanto maior Z , em módulo, mais distante \hat{p}_{ij} é de p_0 , tendendo à rejeição de H_0 .

Entretanto, para que se ocorra ou não a rejeição da hipótese nula, é necessário não apenas conhecer o valor da estatística Z , mas se ter uma métrica da probabilidade de se obter um valor mais extremo do que aquele obtido na amostra, tomando-se como referência a proporção de votos de cada candidato obteve efetivamente a partir dos resultados eleitorais. Assim, quanto menor tal probabilidade, mais distante a estatística amostral do valor hipotético que, neste caso, é a proporção de votos válidos para cada candidato na eleição. Tal medida é conhecida como p-valor, interpretada também como a probabilidade de se cometer o Erro do

Tipo I, que é a probabilidade de se rejeitar a hipótese nula quando esta é verdadeira. Deste modo, quanto menor o p-valor, menor a probabilidade de se cometer um erro ao se rejeitar a hipótese nula, podendo-se afirmar que, baseado na amostra, o resultado eleitoral não seria p_0 .

Portanto, o p-valor é a área de probabilidade na Distribuição Normal Padrão, no teste de hipótese bilateral do tipo “=” ou “≠”, que se encontra à direita da estatística Z positiva e à esquerda da respectiva estatística negativa. Alternativamente, é a área de probabilidade à cauda direita da estatística Z em módulo, multiplicada por dois, para se considerar ambas as caudas da distribuição. E a rejeição ou não da hipótese nula ocorre comparando-se o p-valor com o nível de significância (α) escolhido, que, além de ser o inverso do nível de confiança ($1 - \alpha$), é a tolerância máxima estabelecida pelo pesquisador em se cometer o Erro Tipo I mencionado. Assim, os níveis usuais de significância utilizados são de 10%, 5% e 1%. Portanto, um p-valor igual ou menor que α resulta na rejeição de $H_0: p = p_0$. Na medida em que a rejeição de H_0 atinge níveis menores de significância, mais contundente é a distância estatística entre \hat{p} e p_0 .

A relação entre intervalo de confiança e teste de hipótese bilateral é que, rejeições da hipótese nula a um nível de significância α implica em um intervalo de confiança que não contempla o verdadeiro parâmetro a um nível de confiança $1 - \alpha$. Deste modo, quando a hipótese nula é rejeitada ao nível de significância α , o intervalo de confiança gerado ao nível de confiança $1 - \alpha$ não compreenderá a proporção verificada nos resultados da eleição, situando-se o parâmetro fora da estimativa intervalar. Alternativamente, se a hipótese nula não é rejeitada ao nível de significância α , o intervalo de confiança construído ao nível de $1 - \alpha$ contemplará a proporção de votos válidos para determinado candidato na eleição. Portanto, há um padrão esperado entre intervalos de confiança e resultados eleitorais que se encontram fora (dentro) da margem de erro e a rejeição (não rejeição) da hipótese nula $H_0: p = p_0$ a ser testada.

Relacionados às análises estatísticas efetuadas, de forma a possibilitar a replicação para fins didáticos, Apêndices são apresentados com procedimento em planilha eletrônica e rotinas em software R. No Apêndice A, determina-se o tamanho amostral planejado (n_p) a partir da estimativa preliminar (p^*), nível de confiança (c) e margem de erro (E_d). No Apêndice B, determina-se a margem de erro prevista (E_p) a partir da estimativa planejada (p^*), nível de confiança (c) e tamanho amostral efetivo (n_e). No Apêndice C, desenvolve-se o intervalo de confiança para a proporção ($IC = \hat{p} \pm E_e$), com seus respectivos limites inferior e superior, a partir do tamanho amostral (n_e), proporção amostral (\hat{p}) e nível de confiança (c). No Apêndice D, desenvolve-se o teste de hipótese para a proporção ($H_0: p = p_0$), resultando na estatística de teste Z e p-valor correspondente para se comparar posteriormente com nível de significância.

4. RESULTADOS

Conforme organização inicial de dados a partir dos agregadores Poder360 e CNN, assim como a base de dados PesqEle do TSE, os registros das pesquisas analisadas referentes ao 2º Turno das Eleições Gerais de 2022 estão listados na Tabela 02, em que a primeira coluna corresponde ao número de identificação; a segunda, ao instituto responsável; a terceira, à data de início; a quarta, à data de término; a quinta, à data de registro; e a sexta, à data de divulgação. As informações estão no contexto da Lei nº 9.504/1997 e Resolução nº 23.600/2019, em que a antecedência mínima deve ser de cinco dias completos entre a data de registro e divulgação. Adicionalmente, as datas de início e término permeiam esse intervalo temporal, com término ocorrendo tipicamente um dia antes da data de divulgação. Legalmente, a pesquisa eleitoral é diferente de enquetes ou sondagens, pois entende-se que estas não seguem método científico.

Informações como nível de confiança (c), margem de erro (E_d), tamanho amostral (n_e) e unidades monetárias investidas (R\$), conforme requeridas por lei, estão dispostas na Tabela 03. Seguindo-se a questão-problema, a investigação direciona-se à interrelação entre tais dados, uma vez que o dimensionamento amostral pode estar acima ou abaixo do previsto na geração de estimativas condizentes com o nível de confiança e margem de erro divulgados. De outro modo, para um mesmo nível de confiança, um tamanho amostral subdimensionado resulta em margem de erro efetiva superior à divulgada por determinado instituto, comprometendo as características formalmente divulgadas ao público. Assim, embora argumente-se que a margem de erro seja conhecida após a obtenção da estimativa para a proporção, é possível controlar a margem de erro máxima previamente mediante ferramentas analíticas conhecidas na Estatística.

Valores monetários absolutos tanto representam a disposição a pagar do respectivo solicitante quanto a magnitude do trabalho conduzido, refletido pelo número de entrevistas. Analisando-se investimentos nas pesquisas com o dimensionamento amostral, valores mais expressivos não acompanham, necessariamente, maiores tamanhos amostrais. Deste modo, não se verifica uma associação entre as duas variáveis como esperada. Adicionalmente, é necessário analisar se a razão entre valores empregados (R\$) e tamanho amostral (n_e) é mantida na medida em que o tamanho amostral aumenta para dado instituto. Considerando um mesmo nível de confiança e margem de erro, existe variabilidade de entrevistas conduzidas, com indícios de propriedades estatísticas diferentes das divulgadas, além de registros de valores empregados mais elevados do que outros institutos que trabalharam com dimensionamento mais expressivo.

Tabela 02 – Pesquisas Eleitorais: Início, Término, Registro e Divulgação

ID	Instituto	Início	Término	Registro	Divulgação
BR-04217/2022	Veritá	28/10/2022	29/10/2022	23/10/2022	29/10/2022
BR-08297/2022	Datafolha	28/10/2022	29/10/2022	23/10/2022	29/10/2022
BR-05256/2022	Ipec	23/10/2022	29/10/2022	23/10/2022	29/10/2022
BR-05765/2022	Quaest	28/10/2022	29/10/2022	23/10/2022	29/10/2022
BR-01820/2022	MDA	26/10/2022	28/10/2022	23/10/2022	29/10/2022
BR-09573/2022	Paraná Pesquisas	26/10/2022	28/10/2022	23/10/2022	29/10/2022
BR-04920/2022	Veritá	25/10/2022	27/10/2022	22/10/2022	28/10/2022
BR-07309/2022	Brasmarket	23/10/2022	27/10/2022	22/10/2022	28/10/2022
BR-04208/2022	Datafolha	25/10/2022	27/10/2022	21/10/2022	27/10/2022
BR-07903/2022	Futura	24/10/2022	26/10/2022	20/10/2022	26/10/2022
BR-05418/2022	Gerp	21/10/2022	26/10/2022	21/10/2022	27/10/2022
BR-01560/2022	Nervera	21/10/2022	25/10/2022	20/10/2022	26/10/2022
BR-08584/2022	Brasmarket	21/10/2022	25/10/2022	20/10/2022	26/10/2022
BR-00470/2022	Quaest	23/10/2022	25/10/2022	20/10/2022	26/10/2022
BR-01159/2022	PoderData	23/10/2022	25/10/2022	20/10/2022	26/10/2022
BR-08044/2022	Ipespe	22/10/2022	24/10/2022	19/10/2022	25/10/2022
BR-00525/2022	Paraná Pesquisas	20/10/2022	24/10/2022	19/10/2022	25/10/2022
BR-06043/2022	Ipec	18/10/2022	24/10/2022	18/10/2022	24/10/2022
BR-08487/2022	Brasmarket	19/10/2022	23/10/2022	18/10/2022	24/10/2022
BR-06415/2022	Nervera	18/10/2022	22/10/2022	17/10/2022	23/10/2022
BR-04043/2022	Veritá	17/10/2022	20/10/2022	15/10/2022	21/10/2022
BR-08523/2022	Futura	17/10/2022	19/10/2022	14/10/2022	20/10/2022
BR-00053/2022	Ideia	14/10/2022	19/10/2022	14/10/2022	20/10/2022
BR-05389/2022	Brasmarket	15/10/2022	19/10/2022	14/10/2022	20/10/2022
BR-02276/2022	Paraná Pesquisas	15/10/2022	19/10/2022	14/10/2022	20/10/2022
BR-07340/2022	Datafolha	17/10/2022	19/10/2022	13/10/2022	19/10/2022
BR-08917/2022	PoderData	16/10/2022	18/10/2022	13/10/2022	19/10/2022
BR-04387/2022	Quaest	16/10/2022	18/10/2022	13/10/2022	19/10/2022
BR-06307/2022	Ipespe	17/10/2022	18/10/2022	12/10/2022	18/10/2022
BR-02707/2022	Ipec	11/10/2022	17/10/2022	11/10/2022	17/10/2022
BR-05514/2022	MDA	14/10/2022	16/10/2022	11/10/2022	17/10/2022
BR-04850/2022	Veritá	11/10/2022	15/10/2022	10/10/2022	16/10/2022
BR-01682/2022	Datafolha	13/10/2022	14/10/2022	08/10/2022	14/10/2022
BR-06012/2022	Nervera	08/10/2022	12/10/2022	07/10/2022	13/10/2022
BR-06280/2022	Futura	10/10/2022	12/10/2022	07/10/2022	13/10/2022
BR-07942/2022	Ipespe	10/10/2022	12/10/2022	08/10/2022	14/10/2022
BR-07106/2022	Quaest	10/10/2022	12/10/2022	07/10/2022	13/10/2022
BR-08438/2022	Paraná Pesquisas	08/10/2022	12/10/2022	07/10/2022	13/10/2022
BR-09241/2022	PoderData	09/10/2022	11/10/2022	06/10/2022	12/10/2022
BR-02322/2022	Gerp	06/10/2022	11/10/2022	05/10/2022	11/10/2022
BR-01120/2022	Ipespe	08/10/2022	10/10/2022	05/10/2022	11/10/2022
BR-02853/2022	Ipec	04/10/2022	10/10/2022	04/10/2022	10/10/2022
BR-02012/2022	Datafolha	05/10/2022	07/10/2022	01/10/2022	07/10/2022
BR-07940/2022	Quaest	03/10/2022	05/10/2022	30/09/2022	06/10/2022
BR-08253/2022	PoderData	03/10/2022	05/10/2022	30/09/2022	06/10/2022
BR-02736/2022	Ipec	03/10/2022	05/10/2022	29/09/2022	05/10/2022
BR-08263/2022	Futura	03/10/2022	05/10/2022	30/09/2022	06/10/2022

Nota: Dados coletados do Tribunal Superior Eleitoral (TSE) para o 2º Turno das Eleições Gerais de 2022. Os números de identificação (ID) são das pesquisas mais divulgadas a partir dos agregadores de pesquisas eleitorais Poder360 e CNN. As pesquisas devem ser registradas no Sistema de Registro de Pesquisas Eleitorais (PesqEle) até cinco dias antes de sua divulgação, seguindo a Lei nº. 9.504/1997 e Resolução nº 23.600/2019 do TSE.

Fonte: Elaborado pelo autor.

Tabela 03 – Tamanho Amostral, Nível de Confiança, Margem de Erro e Valores

ID	Instituto	c	E _d	n _e	R\$
BR-04217/2022	Veritá	95,00%	2,00%	17552	119.000,00
BR-08297/2022	Datafolha	95,00%	2,00%	8320	617.974,00
BR-05256/2022	Ipec	95,00%	2,00%	4272	514.446,81
BR-05765/2022	Quaest	95,00%	2,00%	2000	198.200,00
BR-01820/2022	MDA	95,00%	2,20%	2002	168.000,00
BR-09573/2022	Paraná Pesquisas	95,00%	2,00%	2400	50.000,00
BR-04920/2022	Veritá	95,00%	2,00%	30180	119.000,00
BR-07309/2022	Brasmarket	95,00%	2,00%	2400	32.000,00
BR-04208/2022	Datafolha	95,00%	2,00%	4592	615.874,00
BR-07903/2022	Futura	95,00%	2,20%	2000	93.435,01
BR-05418/2022	Gerp	95,55%	2,18%	2095	30.000,00
BR-01560/2022	Nervera	95,00%	1,00%	7500	65.000,00
BR-08584/2022	Brasmarket	95,00%	2,00%	2400	32.000,00
BR-00470/2022	Quaest	95,00%	2,00%	2000	112.865,23
BR-01159/2022	PoderData	95,00%	2,00%	5000	103.715,00
BR-08044/2022	Ipespe	95,45%	3,00%	1100	46.200,00
BR-00525/2022	Paraná Pesquisas	95,00%	2,20%	2020	50.000,00
BR-06043/2022	Ipec	95,00%	2,00%	3008	356.901,84
BR-08487/2022	Brasmarket	95,00%	2,00%	2400	32.000,00
BR-06415/2022	Nervera	95,00%	1,00%	4500	65.000,00
BR-04043/2022	Veritá	95,00%	2,00%	20220	119.000,00
BR-08523/2022	Futura	95,00%	2,20%	2000	93.435,01
BR-00053/2022	Ideia	95,00%	3,00%	1500	33.802,73
BR-05389/2022	Brasmarket	95,00%	2,00%	2400	32.000,00
BR-02276/2022	Paraná Pesquisas	95,00%	2,20%	2020	50.000,00
BR-07340/2022	Datafolha	95,00%	2,00%	2912	467.894,00
BR-08917/2022	PoderData	95,00%	2,00%	8431	103.715,00
BR-04387/2022	Quaest	95,00%	2,00%	2000	112.865,23
BR-06307/2022	Ipespe	95,45%	3,00%	1100	46.200,00
BR-02707/2022	Ipec	95,00%	2,00%	3008	356.901,84
BR-05514/2022	MDA	95,00%	2,20%	2002	168.000,00
BR-04850/2022	Veritá	95,00%	2,00%	5528	119.000,00
BR-01682/2022	Datafolha	95,00%	2,00%	2898	385.874,00
BR-06012/2022	Nervera	95,00%	1,00%	4500	65.000,00
BR-06280/2022	Futura	95,00%	2,20%	2000	93.435,01
BR-07942/2022	Ipespe	95,45%	2,20%	2000	84.000,00
BR-07106/2022	Quaest	95,00%	2,00%	2000	112.865,23
BR-08438/2022	Paraná Pesquisas	95,00%	2,20%	2020	190.000,00
BR-09241/2022	PoderData	95,00%	2,00%	5000	103.715,00
BR-02322/2022	Gerp	95,55%	2,18%	2095	30.000,00
BR-01120/2022	Ipespe	95,45%	3,00%	1100	46.200,00
BR-02853/2022	Ipec	95,00%	2,00%	2000	237.301,76
BR-02012/2022	Datafolha	95,00%	2,00%	2884	467.894,00
BR-07940/2022	Quaest	95,00%	2,00%	2000	112.865,23
BR-08253/2022	PoderData	95,00%	2,00%	3500	103.715,00
BR-02736/2022	Ipec	95,00%	2,00%	2000	237.301,76
BR-08263/2022	Futura	95,00%	2,20%	2000	93.435,01

Nota: Nível de confiança (c), margem de erro divulgada (E_d), tamanho amostral efetivo (n_e) e valor (R\$) coletados a partir do Sistema de Registro de Pesquisas Eleitorais (PesqEle) do Tribunal Superior Eleitoral (TSE), em que institutos devem registrar cada pesquisa, conforme Lei n.º 9.504/1997 e Resolução n.º 23.600/2019 do TSE.

Fonte: Elaborado pelo autor.

Na Tabela 04, tem-se o cálculo da razão entre valores despendidos em cada pesquisa (R\$) e seu respectivo tamanho amostral efetivo (n_e) e, na Tabela 05, apresenta-se o ranking construído, em que os quatro maiores valores empregados por entrevista pertencem aos contratantes Grupo Folha e Grupo Globo em conjunto, com pesquisa conduzida pelo Datafolha, seguidos pelo Grupo Globo individualmente com o desenvolvimento pelo Ipec, ocupando da quinta à nona posição. Nota-se que Datafolha, também situada na décima quarta posição, com os mesmos contratantes, possui valor por entrevista de menos de sessenta por cento do valor de suas próprias pesquisas situadas nas primeiras posições, e um maior número de entrevistas. Entretanto, a relação inversa entre tamanho amostral e valores não se mantém, pois a terceira pesquisa no ranking possui dimensão amostral maior em relação à seguinte do mesmo instituto.

Nas posições intermediárias encontram-se Quaest e Futura como realizadores, ocupando da décima quinta à vigésima segunda posição, com contratantes Banco Genial e Banco Modal, respectivamente, seguidos pelo Ipespe. Observa-se também pesquisa Quaest na décima posição, com valor 75,61% mais elevado do que as demais pesquisas do mesmo instituto e contratante, assim como, na décima primeira posição, situa-se o Paraná Pesquisas com partido como contratante, com valor empregado por entrevista 380% mais elevado em relação às segundas pesquisas mais dispendiosas do mesmo instituto a partir dessa métrica, de mesmo tamanho amostral, mas realizada a partir de recursos do próprio instituto. Na sequência, há registros dos institutos MDA, Datafolha, Quaest, Futura e Ipespe, com o mesmo tamanho amostral, exceto para Ipespe que possui margem de erro de 3 pontos percentuais e nível de confiança de 95,45%.

Nas últimas posições, situam-se PoderData, com os mesmos valores absolutos, nível de confiança e margem de erro, porém com amplitude amostral acima de 4.000; Paraná Pesquisas, com pesquisas de recursos próprios de valores abaixo daquela contratada por partido; Ideia, com registro único e margem de erro de três pontos percentuais; Veritá, com registros de valores absolutos e critérios idênticos, exceto no tamanho amostral com amplitude acima de 20.000; Nervera, de pesquisas com a menor margem de erro, de um ponto percentual, porém amplitude amostral de 3.000; Gerp, com margem de 2,18% e nível de confiança de 95,55%; e Brasmart, com regularidade nos critérios informados e valores empregados, mantendo-se a mesma razão $R\$/n_e$ nos quatro registros analisados. Assim, dada a variabilidade amostral e de diretrizes estabelecidas, é necessária a análise de discrepâncias estatísticas nestes critérios estabelecidos.

Tabela 04 – Razão entre Valor (R\$) e Tamanho Amostral Efetivo (n_e)

ID	Término	Instituto	Contratante	R\$/n _e
BR-04217/2022	29/10/2022	Veritá	Veritá	6,78
BR-08297/2022	29/10/2022	Datafolha	Grupo Folha e Grupo Globo	74,28
BR-05256/2022	29/10/2022	Ipec	Grupo Globo	120,42
BR-05765/2022	29/10/2022	Quaest	Banco Genial	99,10
BR-01820/2022	28/10/2022	MDA	CNT	83,92
BR-09573/2022	28/10/2022	Paraná Pesquisas	Paraná Pesquisas	20,83
BR-04920/2022	27/10/2022	Veritá	Veritá	3,94
BR-07309/2022	27/10/2022	Brasmarket	Brasmarket	13,33
BR-04208/2022	27/10/2022	Datafolha	Grupo Folha e Grupo Globo	134,12
BR-07903/2022	26/10/2022	Futura	Banco Modal	46,72
BR-05418/2022	26/10/2022	Gerp	Gerp	14,32
BR-01560/2022	25/10/2022	Nervera	Nervera	8,67
BR-08584/2022	25/10/2022	Brasmarket	Brasmarket	13,33
BR-00470/2022	25/10/2022	Quaest	Banco Genial	56,43
BR-01159/2022	25/10/2022	PoderData	PoderData	20,74
BR-08044/2022	24/10/2022	Ipespe	Ipespe	42,00
BR-00525/2022	24/10/2022	Paraná Pesquisas	Paraná Pesquisas	24,75
BR-06043/2022	24/10/2022	Ipec	Grupo Globo	118,65
BR-08487/2022	23/10/2022	Brasmarket	Brasmarket	13,33
BR-06415/2022	22/10/2022	Nervera	Nervera	14,44
BR-04043/2022	20/10/2022	Veritá	Veritá	5,89
BR-08523/2022	19/10/2022	Futura	Banco Modal	46,72
BR-00053/2022	19/10/2022	Ideia	Ideia	22,54
BR-05389/2022	19/10/2022	Brasmarket	Brasmarket	13,33
BR-02276/2022	19/10/2022	Paraná Pesquisas	Paraná Pesquisas	24,75
BR-07340/2022	19/10/2022	Datafolha	Grupo Folha e Grupo Globo	160,68
BR-08917/2022	18/10/2022	PoderData	PoderData	12,30
BR-04387/2022	18/10/2022	Quaest	Banco Genial	56,43
BR-06307/2022	18/10/2022	Ipespe	Ipespe	42,00
BR-02707/2022	17/10/2022	Ipec	Grupo Globo	118,65
BR-05514/2022	16/10/2022	MDA	CNT	83,92
BR-04850/2022	15/10/2022	Veritá	Veritá	21,53
BR-01682/2022	14/10/2022	Datafolha	Grupo Folha e Grupo Globo	133,15
BR-06012/2022	12/10/2022	Nervera	Nervera	14,44
BR-06280/2022	12/10/2022	Futura	Banco Modal	46,72
BR-07942/2022	12/10/2022	Ipespe	XP Investimentos	42,00
BR-07106/2022	12/10/2022	Quaest	Banco Genial	56,43
BR-08438/2022	12/10/2022	Paraná Pesquisas	PP	94,06
BR-09241/2022	11/10/2022	PoderData	PoderData	20,74
BR-02322/2022	11/10/2022	Gerp	Gerp	14,32
BR-01120/2022	10/10/2022	Ipespe	Ipespe	42,00
BR-02853/2022	10/10/2022	Ipec	Grupo Globo	118,65
BR-02012/2022	07/10/2022	Datafolha	Grupo Folha e Grupo Globo	162,24
BR-07940/2022	05/10/2022	Quaest	Banco Genial	56,43
BR-08253/2022	05/10/2022	PoderData	PoderData	29,63
BR-02736/2022	05/10/2022	Ipec	Grupo Globo	118,65
BR-08263/2022	05/10/2022	Futura	Banco Modal	46,72

Nota: Cálculos da razão entre valor (R\$) e tamanho amostral efetivo (n_e) realizados a partir das informações divulgadas no Sistema de Registro de Pesquisas Eleitorais (PesqEle) do Tribunal Superior Eleitoral (TSE). Números idênticos até a segunda casa decimal não implicam necessariamente em igualdade nas demais casas.

Fonte: Elaborado pelo autor.

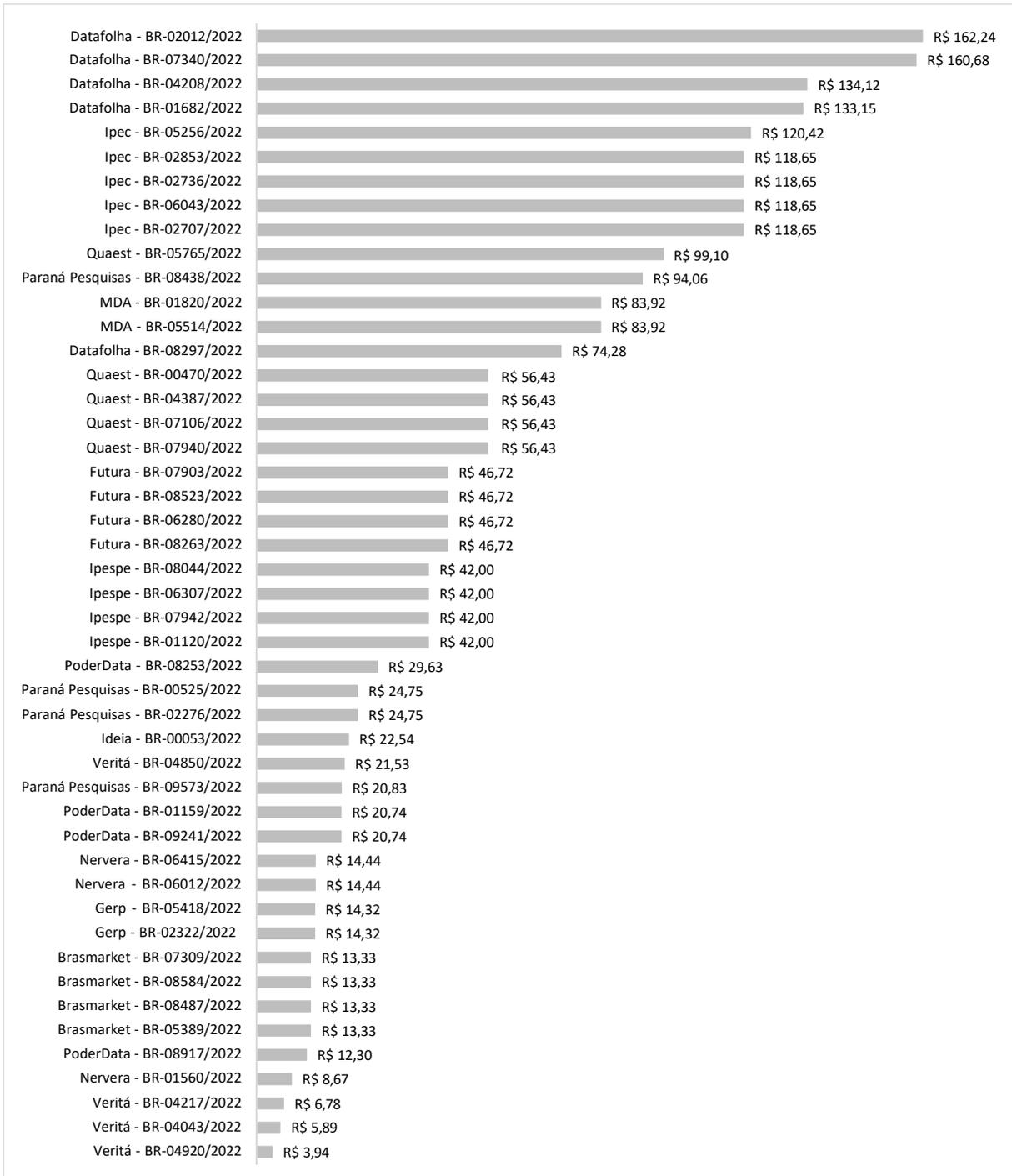
Tabela 05 – Ranking: Razão entre Valor (R\$) e Tamanho Amostral Efetivo (n_e)

Ranking	ID	Término	Instituto	Contratante	R\$/ n_e
1º	BR-02012/2022	07/10/2022	Datafolha	Grupo Folha e Grupo Globo	162,24
2º	BR-07340/2022	19/10/2022	Datafolha	Grupo Folha e Grupo Globo	160,68
3º	BR-04208/2022	27/10/2022	Datafolha	Grupo Folha e Grupo Globo	134,12
4º	BR-01682/2022	14/10/2022	Datafolha	Grupo Folha e Grupo Globo	133,15
5º	BR-05256/2022	29/10/2022	Ipec	Grupo Globo	120,42
6º	BR-02853/2022	10/10/2022	Ipec	Grupo Globo	118,65
7º	BR-02736/2022	05/10/2022	Ipec	Grupo Globo	118,65
8º	BR-06043/2022	24/10/2022	Ipec	Grupo Globo	118,65
9º	BR-02707/2022	17/10/2022	Ipec	Grupo Globo	118,65
10º	BR-05765/2022	29/10/2022	Quaest	Banco Genial	99,10
11º	BR-08438/2022	12/10/2022	Paraná Pesquisas	PP	94,06
12º	BR-01820/2022	28/10/2022	MDA	CNT	83,92
13º	BR-05514/2022	16/10/2022	MDA	CNT	83,92
14º	BR-08297/2022	29/10/2022	Datafolha	Grupo Folha e Grupo Globo	74,28
15º	BR-00470/2022	25/10/2022	Quaest	Banco Genial	56,43
16º	BR-04387/2022	18/10/2022	Quaest	Banco Genial	56,43
17º	BR-07106/2022	12/10/2022	Quaest	Banco Genial	56,43
18º	BR-07940/2022	05/10/2022	Quaest	Banco Genial	56,43
19º	BR-07903/2022	26/10/2022	Futura	Banco Modal	46,72
20º	BR-08523/2022	19/10/2022	Futura	Banco Modal	46,72
21º	BR-06280/2022	12/10/2022	Futura	Banco Modal	46,72
22º	BR-08263/2022	05/10/2022	Futura	Banco Modal	46,72
23º	BR-08044/2022	24/10/2022	Ipespe	Ipespe	42,00
24º	BR-06307/2022	18/10/2022	Ipespe	Ipespe	42,00
25º	BR-07942/2022	12/10/2022	Ipespe	XP Investimentos	42,00
26º	BR-01120/2022	10/10/2022	Ipespe	Ipespe	42,00
27º	BR-08253/2022	05/10/2022	PoderData	PoderData	29,63
28º	BR-00525/2022	24/10/2022	Paraná Pesquisas	Paraná Pesquisas	24,75
29º	BR-02276/2022	19/10/2022	Paraná Pesquisas	Paraná Pesquisas	24,75
30º	BR-00053/2022	19/10/2022	Ideia	Ideia	22,54
31º	BR-04850/2022	15/10/2022	Veritá	Veritá	21,53
32º	BR-09573/2022	28/10/2022	Paraná Pesquisas	Paraná Pesquisas	20,83
33º	BR-01159/2022	25/10/2022	PoderData	PoderData	20,74
34º	BR-09241/2022	11/10/2022	PoderData	PoderData	20,74
35º	BR-06415/2022	22/10/2022	Nervera	Nervera	14,44
36º	BR-06012/2022	12/10/2022	Nervera	Nervera	14,44
37º	BR-05418/2022	26/10/2022	Gerp	Gerp	14,32
38º	BR-02322/2022	11/10/2022	Gerp	Gerp	14,32
39º	BR-07309/2022	27/10/2022	Brasmarket	Brasmarket	13,33
40º	BR-08584/2022	25/10/2022	Brasmarket	Brasmarket	13,33
41º	BR-08487/2022	23/10/2022	Brasmarket	Brasmarket	13,33
42º	BR-05389/2022	19/10/2022	Brasmarket	Brasmarket	13,33
43º	BR-08917/2022	18/10/2022	PoderData	PoderData	12,30
44º	BR-01560/2022	25/10/2022	Nervera	Nervera	8,67
45º	BR-04217/2022	29/10/2022	Veritá	Veritá	6,78
46º	BR-04043/2022	20/10/2022	Veritá	Veritá	5,89
47º	BR-04920/2022	27/10/2022	Veritá	Veritá	3,94

Nota: Cálculos da razão entre valor (R\$) e tamanho amostral efetivo (n_e) realizados a partir das informações divulgadas no Sistema de Registro de Pesquisas Eleitorais (PesqEle) do Tribunal Superior Eleitoral (TSE). Números idênticos até a segunda casa decimal não implicam necessariamente em igualdade nas demais casas.

Fonte: Elaborado pelo autor.

Figura 01 – Ranking da Razão do Valor pelo Tamanho Amostral Efetivo (R\$/n_e)



Nota: Cálculos da razão entre valor (R\$) e tamanho amostral efetivo (n_e) realizados a partir das informações divulgadas no Sistema de Registro de Pesquisas Eleitorais (PesqEle) do Tribunal Superior Eleitoral (TSE). Números idênticos até a segunda casa decimal não implicam necessariamente em igualdade nas demais casas.

Fonte: Elaborado pelo autor.

Na sequência, apresentam-se as variações percentuais entre tamanho amostral previsto (n_p) e tamanho amostral efetivo (n_e), em que $\Delta\% = n_e/n_p - 1$. O termo n_p é obtido, a partir da Tabela 06, por meio de análise estatística a partir de $n_p = p \cdot q \cdot (z_c/E_d)^2$, baseada no nível de confiança e margem de erro estipulados, assim como estimativa planejada para a proporção amostral que, por padrão, é 0,5 para se gerar um dimensionamento amostral mínimo que resulte em estimativas intervalares de acordo com a margem de erro máxima previamente informada.

Portanto, variações percentuais positivas entre n_p e n_e resultam em margens de erro menores do que as divulgadas originalmente e intervalos de confiança mais precisos. Por outro lado, discrepâncias percentuais negativas por dimensionamento amostral menor do que o previsto pela análise estatística refletem em margens maiores e imprecisão das estimativas intervalares.

Na Tabela 07 e Figuras 02 e 03, apresenta-se o ranking de variações percentuais entre tamanho amostral previsto e efetivo. As três maiores variações positivas são do instituto Veritá, cuja pesquisa-líder possui variação de 1156,98%. Na quarta e quinta posição, respectivamente, encontram-se pesquisas dos institutos PoderData e Datafolha, com variações acima de 200%. Observa-se que as pesquisas com os menores custos por entrevista desses dois institutos resultaram nos registros com as maiores variações percentuais entre dimensionamento previsto e efetivo, evidenciando diluição de custos aliada à maior precisão nesses casos, embora, para Datafolha, amostras maiores não resultem necessariamente em valores menores empregados por entrevista. Até a décima posição, verificam-se registros dos institutos Veritá, PoderData, Datafolha e Ipec, cujas pesquisas somente dos dois últimos estão entre os cinco maiores R\$/ n_e .

Nas dez maiores variações negativas, com discrepâncias acima de 15% entre o tamanho mínimo previsto estatisticamente e o aplicado efetivamente na pesquisa, verificam-se registros de pesquisas realizadas pelos institutos Nervera (três observações), Ipec (duas observações) e Quaest (cinco observações). No caso do primeiro instituto, as pesquisas encontram-se entre as de menores R\$/ n_e , aliando-se baixos valores por entrevista ao dimensionamento amostral insuficiente. No caso do segundo instituto, ambos os registros de maior discrepância percentual entre n_p e n_e estão também entre os dez maiores valores R\$/ n_e . Assim, aplicaram-se valores relativamente expressivos por entrevista, porém com tamanho amostral abaixo do mínimo previsto estatisticamente para a margem de erro máxima e nível de confiança divulgados. No caso do terceiro instituto, as pesquisas também estão entre os 20 maiores valores por entrevista.

Tabela 06 – Variação Percentual entre Tamanho Amostral Previsto e Efetivo

ID	Instituto	c	E _d	n _p	n _e	Δ% (n _p , n _e)
BR-04217/2022	Veritá	95,00%	2,00%	2401	17552	631,03%
BR-08297/2022	Datafolha	95,00%	2,00%	2401	8320	246,52%
BR-05256/2022	Ipec	95,00%	2,00%	2401	4272	77,93%
BR-05765/2022	Quaest	95,00%	2,00%	2401	2000	-16,70%
BR-01820/2022	MDA	95,00%	2,20%	1985	2002	0,86%
BR-09573/2022	Paraná Pesquisas	95,00%	2,00%	2401	2400	-0,04%
BR-04920/2022	Veritá	95,00%	2,00%	2401	30180	1156,98%
BR-07309/2022	Brasmarket	95,00%	2,00%	2401	2400	-0,04%
BR-04208/2022	Datafolha	95,00%	2,00%	2401	4592	91,25%
BR-07903/2022	Futura	95,00%	2,20%	1985	2000	0,76%
BR-05418/2022	Gerp	95,55%	2,18%	2124	2095	-1,37%
BR-01560/2022	Nervera	95,00%	1,00%	9604	7500	-21,91%
BR-08584/2022	Brasmarket	95,00%	2,00%	2401	2400	-0,04%
BR-00470/2022	Quaest	95,00%	2,00%	2401	2000	-16,70%
BR-01159/2022	PoderData	95,00%	2,00%	2401	5000	108,25%
BR-08044/2022	Ipespe	95,45%	3,00%	1112	1100	-1,08%
BR-00525/2022	Paraná Pesquisas	95,00%	2,20%	1985	2020	1,76%
BR-06043/2022	Ipec	95,00%	2,00%	2401	3008	25,28%
BR-08487/2022	Brasmarket	95,00%	2,00%	2401	2400	-0,04%
BR-06415/2022	Nervera	95,00%	1,00%	9604	4500	-53,14%
BR-04043/2022	Veritá	95,00%	2,00%	2401	20220	742,15%
BR-08523/2022	Futura	95,00%	2,20%	1985	2000	0,76%
BR-00053/2022	Ideia	95,00%	3,00%	1068	1500	40,45%
BR-05389/2022	Brasmarket	95,00%	2,00%	2401	2400	-0,04%
BR-02276/2022	Paraná Pesquisas	95,00%	2,20%	1985	2020	1,76%
BR-07340/2022	Datafolha	95,00%	2,00%	2401	2912	21,28%
BR-08917/2022	PoderData	95,00%	2,00%	2401	8431	251,15%
BR-04387/2022	Quaest	95,00%	2,00%	2401	2000	-16,70%
BR-06307/2022	Ipespe	95,45%	3,00%	1112	1100	-1,08%
BR-02707/2022	Ipec	95,00%	2,00%	2401	3008	25,28%
BR-05514/2022	MDA	95,00%	2,20%	1985	2002	0,86%
BR-04850/2022	Veritá	95,00%	2,00%	2401	5528	130,24%
BR-01682/2022	Datafolha	95,00%	2,00%	2401	2898	20,70%
BR-06012/2022	Nervera	95,00%	1,00%	9604	4500	-53,14%
BR-06280/2022	Futura	95,00%	2,20%	1985	2000	0,76%
BR-07942/2022	Ipespe	95,45%	2,20%	2067	2000	-3,24%
BR-07106/2022	Quaest	95,00%	2,00%	2401	2000	-16,70%
BR-08438/2022	Paraná Pesquisas	95,00%	2,20%	1985	2020	1,76%
BR-09241/2022	PoderData	95,00%	2,00%	2401	5000	108,25%
BR-02322/2022	Gerp	95,55%	2,18%	2124	2095	-1,37%
BR-01120/2022	Ipespe	95,45%	3,00%	1112	1100	-1,08%
BR-02853/2022	Ipec	95,00%	2,00%	2401	2000	-16,70%
BR-02012/2022	Datafolha	95,00%	2,00%	2401	2884	20,12%
BR-07940/2022	Quaest	95,00%	2,00%	2401	2000	-16,70%
BR-08253/2022	PoderData	95,00%	2,00%	2401	3500	45,77%
BR-02736/2022	Ipec	95,00%	2,00%	2401	2000	-16,70%
BR-08263/2022	Futura	95,00%	2,20%	1985	2000	0,76%

Nota: Nível de confiança (c), margem de erro divulgada (E_d), tamanho amostral previsto (n_p) e tamanho amostral efetivo (n_e). Aplicação de $n_p = p^*q^*(z_c/E_d)^2$, em que $p^* = 0,5$, $q^* = 1 - p^*$ e z_c retrata o valor crítico para c.

Fonte: Elaborado pelo autor.

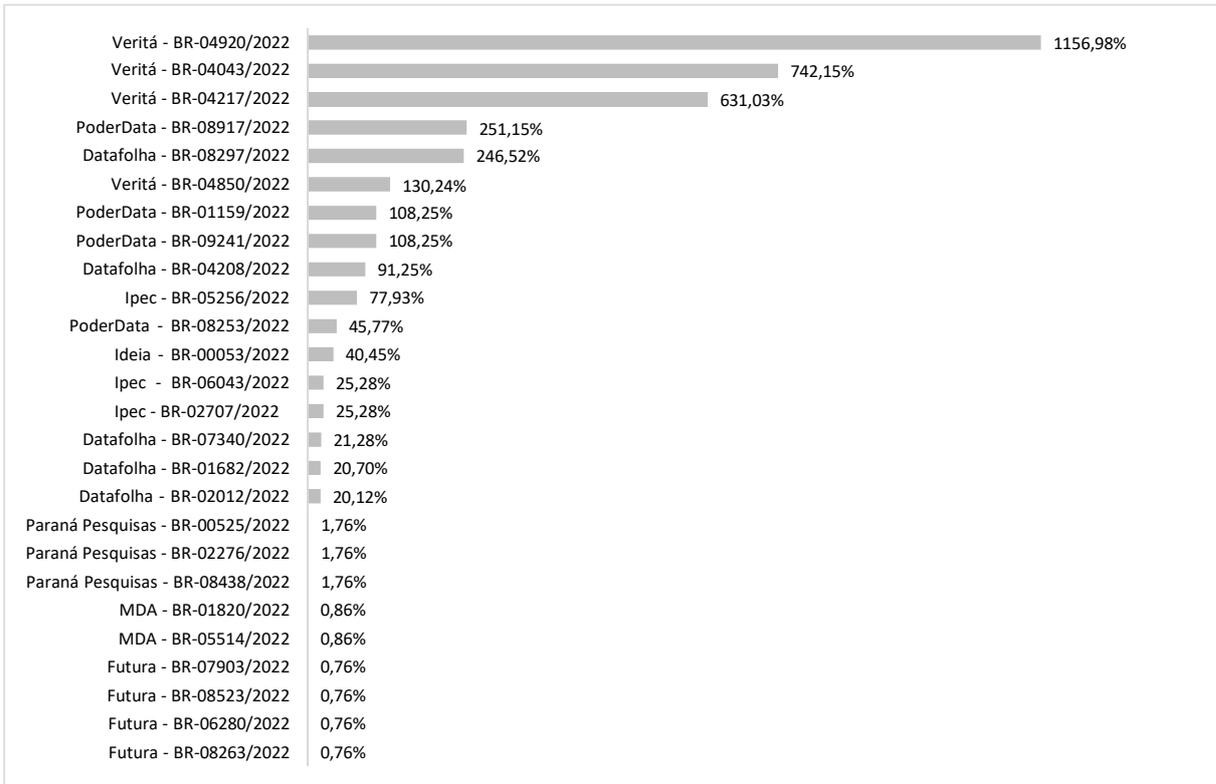
Tabela 07 – Ranking: Variação Percentual entre Tamanho Amostral Previsto e Efetivo

Ranking	ID	Instituto	c	E _d	n _p	n _e	Δ% (n _p , n _e)
1º	BR-04920/2022	Veritá	95,00%	2,00%	2401	30180	1156,98%
2º	BR-04043/2022	Veritá	95,00%	2,00%	2401	20220	742,15%
3º	BR-04217/2022	Veritá	95,00%	2,00%	2401	17552	631,03%
4º	BR-08917/2022	PoderData	95,00%	2,00%	2401	8431	251,15%
5º	BR-08297/2022	Datafolha	95,00%	2,00%	2401	8320	246,52%
6º	BR-04850/2022	Veritá	95,00%	2,00%	2401	5528	130,24%
7º	BR-01159/2022	PoderData	95,00%	2,00%	2401	5000	108,25%
8º	BR-09241/2022	PoderData	95,00%	2,00%	2401	5000	108,25%
9º	BR-04208/2022	Datafolha	95,00%	2,00%	2401	4592	91,25%
10º	BR-05256/2022	Ipec	95,00%	2,00%	2401	4272	77,93%
11º	BR-08253/2022	PoderData	95,00%	2,00%	2401	3500	45,77%
12º	BR-00053/2022	Ideia	95,00%	3,00%	1068	1500	40,45%
13º	BR-06043/2022	Ipec	95,00%	2,00%	2401	3008	25,28%
14º	BR-02707/2022	Ipec	95,00%	2,00%	2401	3008	25,28%
15º	BR-07340/2022	Datafolha	95,00%	2,00%	2401	2912	21,28%
16º	BR-01682/2022	Datafolha	95,00%	2,00%	2401	2898	20,70%
17º	BR-02012/2022	Datafolha	95,00%	2,00%	2401	2884	20,12%
18º	BR-00525/2022	Paraná Pesquisas	95,00%	2,20%	1985	2020	1,76%
19º	BR-02276/2022	Paraná Pesquisas	95,00%	2,20%	1985	2020	1,76%
20º	BR-08438/2022	Paraná Pesquisas	95,00%	2,20%	1985	2020	1,76%
21º	BR-01820/2022	MDA	95,00%	2,20%	1985	2002	0,86%
22º	BR-05514/2022	MDA	95,00%	2,20%	1985	2002	0,86%
23º	BR-07903/2022	Futura	95,00%	2,20%	1985	2000	0,76%
24º	BR-08523/2022	Futura	95,00%	2,20%	1985	2000	0,76%
25º	BR-06280/2022	Futura	95,00%	2,20%	1985	2000	0,76%
26º	BR-08263/2022	Futura	95,00%	2,20%	1985	2000	0,76%
27º	BR-09573/2022	Paraná Pesquisas	95,00%	2,00%	2401	2400	-0,04%
28º	BR-07309/2022	Brasmarket	95,00%	2,00%	2401	2400	-0,04%
29º	BR-08584/2022	Brasmarket	95,00%	2,00%	2401	2400	-0,04%
30º	BR-08487/2022	Brasmarket	95,00%	2,00%	2401	2400	-0,04%
31º	BR-05389/2022	Brasmarket	95,00%	2,00%	2401	2400	-0,04%
32º	BR-08044/2022	Ipespe	95,45%	3,00%	1112	1100	-1,08%
33º	BR-06307/2022	Ipespe	95,45%	3,00%	1112	1100	-1,08%
34º	BR-01120/2022	Ipespe	95,45%	3,00%	1112	1100	-1,08%
35º	BR-05418/2022	Gerp	95,55%	2,18%	2124	2095	-1,37%
36º	BR-02322/2022	Gerp	95,55%	2,18%	2124	2095	-1,37%
37º	BR-07942/2022	Ipespe	95,45%	2,20%	2067	2000	-3,24%
38º	BR-05765/2022	Quaest	95,00%	2,00%	2401	2000	-16,70%
39º	BR-00470/2022	Quaest	95,00%	2,00%	2401	2000	-16,70%
40º	BR-04387/2022	Quaest	95,00%	2,00%	2401	2000	-16,70%
41º	BR-07106/2022	Quaest	95,00%	2,00%	2401	2000	-16,70%
42º	BR-02853/2022	Ipec	95,00%	2,00%	2401	2000	-16,70%
43º	BR-07940/2022	Quaest	95,00%	2,00%	2401	2000	-16,70%
44º	BR-02736/2022	Ipec	95,00%	2,00%	2401	2000	-16,70%
45º	BR-01560/2022	Nervera	95,00%	1,00%	9604	7500	-21,91%
46º	BR-06415/2022	Nervera	95,00%	1,00%	9604	4500	-53,14%
47º	BR-06012/2022	Nervera	95,00%	1,00%	9604	4500	-53,14%

Nota: Nível de confiança (c), margem de erro divulgada (E_d), tamanho amostral previsto (n_p) e tamanho amostral efetivo (n_e). Aplicação de $n_p = p^*q^*(z_c/E_d)^2$, em que $p^* = 0,5$, $q^* = 1 - p^*$ e z_c retrata o valor crítico para c.

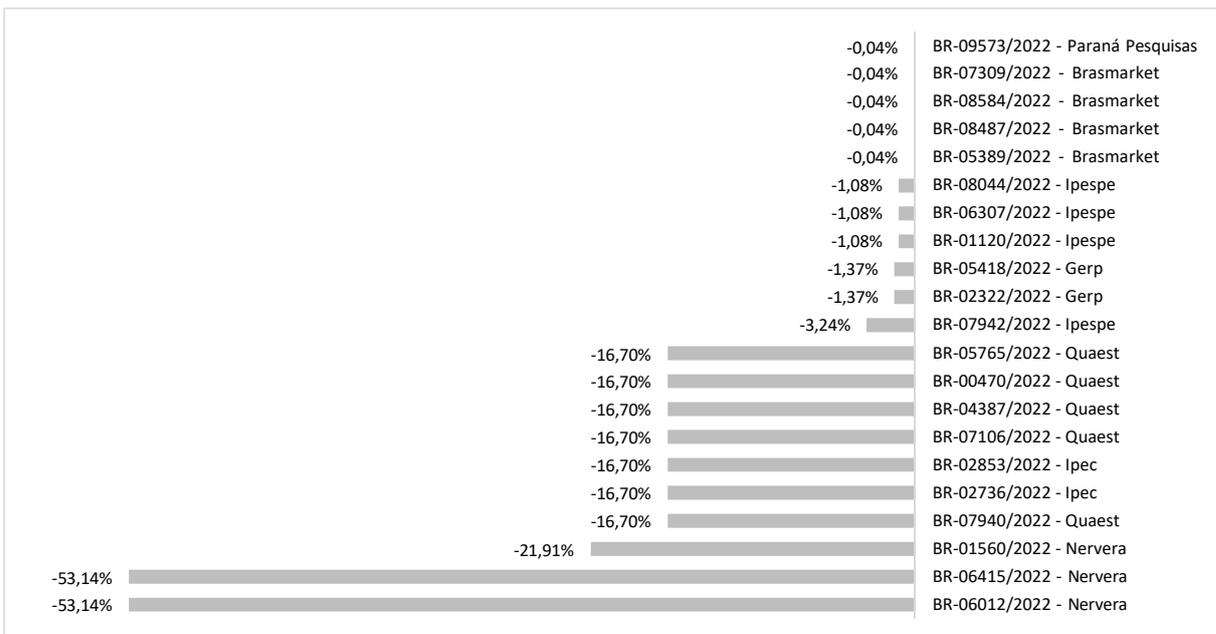
Fonte: Elaborado pelo autor.

Figura 02 – Ranking: Variações Percentuais Positivas (n_p e n_e)



Nota: Cálculos de diferenças percentuais entre tamanho amostral previsto (n_p) a partir de $n_p = p \cdot q \cdot (z_c/E_d)^2$ e tamanho amostral efetivo (n_e). Diferenças percentuais positivas maiores refletem em margens de erro menores.
Fonte: Elaborado pelo autor.

Figura 03 – Ranking: Variações Percentuais Negativas (n_p e n_e)



Nota: Cálculos de diferenças percentuais entre tamanho amostral previsto (n_p) a partir de $n_p = p \cdot q \cdot (z_c/E_d)^2$ e tamanho amostral efetivo (n_e). Diferenças percentuais negativas maiores refletem em margens de erro maiores.
Fonte: Elaborado pelo autor.

Na etapa seguinte, investigam-se as variações percentuais entre a margem de erro divulgada (E_d) no PesqEle e a margem de erro prevista (E_p) por meio de ferramenta estatística.

O termo E_p , apresentado a partir da Tabela 08, é obtido por $E_p = z_c \sqrt{p^*q^*/n_e}$, cujo resultado baseia-se no produto das estimativas preliminares que possuem valor de 0,5 por padrão, no valor crítico correspondente ao nível de confiança estabelecido e no tamanho amostral efetivo. Deste modo, mantendo-se os demais parâmetros constantes, quanto maior o nível de confiança, maior a margem prevista; quanto maior o tamanho amostral, menor será a E_p ; e, para qualquer valor diferente de 0,5 na estimativa planejada, menor a E_p resultante. Adicionalmente, variações percentuais positivas entre E_d e E_p decorrem de dimensionamento amostral efetivo abaixo do previsto estatisticamente, refletindo em margens de erro maiores que as divulgadas ao público.

Apresenta-se então o ranking de variações percentuais entre E_d e E_p na Tabela 09 e nas Figuras 04 e 05 contemplando, respectivamente, variações negativas e positivas identificadas. Se disposto de forma decrescente, o ranking seria diametralmente oposto ao das discrepâncias entre tamanho amostral previsto e efetivo, como previsto e observado. Deste modo, seguindo a estrutura de ambos os rankings, pesquisas com dimensionamento amostral efetivo acima (abaixo) do previsto refletem em margem de erro estatisticamente prevista abaixo (acima) daquela registrada no PesqEle. Assim, pesquisas com as maiores variações percentuais negativas nessa métrica tendem a gerar estimativas mais precisas do que a divulgação indica.

Por outro lado, pesquisas com as maiores variações percentuais positivas tendem a produzir intervalos de confiança menos precisos do que o registro e a comunicação ao público sugerem.

Deste modo, E_p representa a margem de erro máxima prevista, analisando-se, nas últimas posições, registros de margem de erro divulgada de um ponto percentual que não possuem tamanho amostral suficiente para mantê-la com o nível de confiança estabelecido e discrepância amostral atingindo nível acima de 45%, como no caso de três observações para Nervera, assim como pesquisas de valores mais elevados por entrevista em termos relativos com defasagem amostral acima de 9%, como duas pesquisas Ipec e cinco observações Quaest. Adicionalmente, analisa-se variação percentual de 0,69%, em dois registros Gerp; 0,50% em três pesquisas Ipespe, em que a própria empresa é a contratante, porém de 1,64% no caso de contratante externo; e 0,02% em registros da Brasmarket e Paraná Pesquisas. Os procedimentos para a viabilização de n_p e E_p em planilha eletrônica e software R estão nos Apêndices A e B.

Tabela 08 – Variação Percentual entre Margem de Erro Divulgada e Prevista

ID	Instituto	c	n _e	E _d	E _p	Δ% (E _d , E _p)
BR-04217/2022	Veritá	95,00%	17552	2,00%	0,7397%	-63,02%
BR-08297/2022	Datafolha	95,00%	8320	2,00%	1,0744%	-46,28%
BR-05256/2022	Ipec	95,00%	4272	2,00%	1,4993%	-25,03%
BR-05765/2022	Quaest	95,00%	2000	2,00%	2,1913%	9,57%
BR-01820/2022	MDA	95,00%	2002	2,20%	2,1902%	-0,44%
BR-09573/2022	Paraná Pesquisas	95,00%	2400	2,00%	2,0004%	0,02%
BR-04920/2022	Veritá	95,00%	30180	2,00%	0,5641%	-71,79%
BR-07309/2022	Brasmarket	95,00%	2400	2,00%	2,0004%	0,02%
BR-04208/2022	Datafolha	95,00%	4592	2,00%	1,4462%	-27,69%
BR-07903/2022	Futura	95,00%	2000	2,20%	2,1913%	-0,40%
BR-05418/2022	Gerp	95,55%	2095	2,18%	2,1950%	0,69%
BR-01560/2022	Nervera	95,00%	7500	1,00%	1,1316%	13,16%
BR-08584/2022	Brasmarket	95,00%	2400	2,00%	2,0004%	0,02%
BR-00470/2022	Quaest	95,00%	2000	2,00%	2,1913%	9,57%
BR-01159/2022	PoderData	95,00%	5000	2,00%	1,3859%	-30,70%
BR-08044/2022	Ipespe	95,45%	1100	3,00%	3,0151%	0,50%
BR-00525/2022	Paraná Pesquisas	95,00%	2020	2,20%	2,1804%	-0,89%
BR-06043/2022	Ipec	95,00%	3008	2,00%	1,7868%	-10,66%
BR-08487/2022	Brasmarket	95,00%	2400	2,00%	2,0004%	0,02%
BR-06415/2022	Nervera	95,00%	4500	1,00%	1,4609%	46,09%
BR-04043/2022	Veritá	95,00%	20220	2,00%	0,6892%	-65,54%
BR-08523/2022	Futura	95,00%	2000	2,20%	2,1913%	-0,40%
BR-00053/2022	Ideia	95,00%	1500	3,00%	2,5303%	-15,66%
BR-05389/2022	Brasmarket	95,00%	2400	2,00%	2,0004%	0,02%
BR-02276/2022	Paraná Pesquisas	95,00%	2020	2,20%	2,1804%	-0,89%
BR-07340/2022	Datafolha	95,00%	2912	2,00%	1,8160%	-9,20%
BR-08917/2022	PoderData	95,00%	8431	2,00%	1,0673%	-46,64%
BR-04387/2022	Quaest	95,00%	2000	2,00%	2,1913%	9,57%
BR-06307/2022	Ipespe	95,45%	1100	3,00%	3,0151%	0,50%
BR-02707/2022	Ipec	95,00%	3008	2,00%	1,7868%	-10,66%
BR-05514/2022	MDA	95,00%	2002	2,20%	2,1902%	-0,44%
BR-04850/2022	Veritá	95,00%	5528	2,00%	1,3181%	-34,10%
BR-01682/2022	Datafolha	95,00%	2898	2,00%	1,8204%	-8,98%
BR-06012/2022	Nervera	95,00%	4500	1,00%	1,4609%	46,09%
BR-06280/2022	Futura	95,00%	2000	2,20%	2,1913%	-0,40%
BR-07942/2022	Ipespe	95,45%	2000	2,20%	2,2361%	1,64%
BR-07106/2022	Quaest	95,00%	2000	2,00%	2,1913%	9,57%
BR-08438/2022	Paraná Pesquisas	95,00%	2020	2,20%	2,1804%	-0,89%
BR-09241/2022	PoderData	95,00%	5000	2,00%	1,3859%	-30,70%
BR-02322/2022	Gerp	95,55%	2095	2,18%	2,1950%	0,69%
BR-01120/2022	Ipespe	95,45%	1100	3,00%	3,0151%	0,50%
BR-02853/2022	Ipec	95,00%	2000	2,00%	2,1913%	9,57%
BR-02012/2022	Datafolha	95,00%	2884	2,00%	1,8248%	-8,76%
BR-07940/2022	Quaest	95,00%	2000	2,00%	2,1913%	9,57%
BR-08253/2022	PoderData	95,00%	3500	2,00%	1,6565%	-17,18%
BR-02736/2022	Ipec	95,00%	2000	2,00%	2,1913%	9,57%
BR-08263/2022	Futura	95,00%	2000	2,20%	2,1913%	-0,40%

Nota: Nível de confiança (c), tamanho amostral efetivo (n_e), margem de erro divulgada (E_d) e margem de erro prevista (E_p). Aplicação $E_p = z_c \sqrt{p^*q^*/n_e}$, em que $p^* = 0,5$, $q^* = 1 - p^*$ e z_c consiste no valor crítico para c.

Fonte: Elaborado pelo autor.

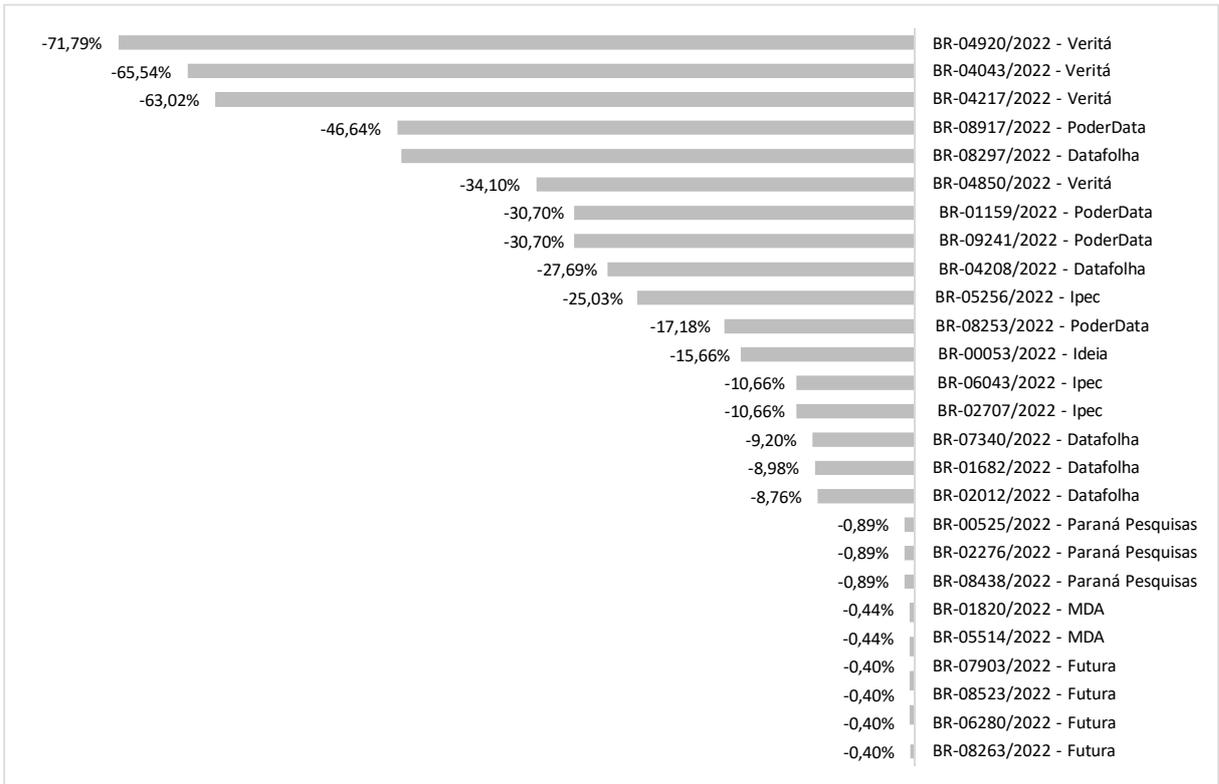
Tabela 09 – Ranking: Variação Percentual entre Margem de Erro Divulgada e Prevista

Ranking	ID	Instituto	c	n _e	E _d	E _p	Δ% (E _d , E _p)
1º	BR-04920/2022	Veritá	95,00%	30180	2,00%	0,5641%	-71,79%
2º	BR-04043/2022	Veritá	95,00%	20220	2,00%	0,6892%	-65,54%
3º	BR-04217/2022	Veritá	95,00%	17552	2,00%	0,7397%	-63,02%
4º	BR-08917/2022	PoderData	95,00%	8431	2,00%	1,0673%	-46,64%
5º	BR-08297/2022	Datafolha	95,00%	8320	2,00%	1,0744%	-46,28%
6º	BR-04850/2022	Veritá	95,00%	5528	2,00%	1,3181%	-34,10%
7º	BR-01159/2022	PoderData	95,00%	5000	2,00%	1,3859%	-30,70%
8º	BR-09241/2022	PoderData	95,00%	5000	2,00%	1,3859%	-30,70%
9º	BR-04208/2022	Datafolha	95,00%	4592	2,00%	1,4462%	-27,69%
10º	BR-05256/2022	Ipec	95,00%	4272	2,00%	1,4993%	-25,03%
11º	BR-08253/2022	PoderData	95,00%	3500	2,00%	1,6565%	-17,18%
12º	BR-00053/2022	Ideia	95,00%	1500	3,00%	2,5303%	-15,66%
13º	BR-06043/2022	Ipec	95,00%	3008	2,00%	1,7868%	-10,66%
14º	BR-02707/2022	Ipec	95,00%	3008	2,00%	1,7868%	-10,66%
15º	BR-07340/2022	Datafolha	95,00%	2912	2,00%	1,8160%	-9,20%
16º	BR-01682/2022	Datafolha	95,00%	2898	2,00%	1,8204%	-8,98%
17º	BR-02012/2022	Datafolha	95,00%	2884	2,00%	1,8248%	-8,76%
18º	BR-00525/2022	Paraná Pesquisas	95,00%	2020	2,20%	2,1804%	-0,89%
19º	BR-02276/2022	Paraná Pesquisas	95,00%	2020	2,20%	2,1804%	-0,89%
20º	BR-08438/2022	Paraná Pesquisas	95,00%	2020	2,20%	2,1804%	-0,89%
21º	BR-01820/2022	MDA	95,00%	2002	2,20%	2,1902%	-0,44%
22º	BR-05514/2022	MDA	95,00%	2002	2,20%	2,1902%	-0,44%
23º	BR-07903/2022	Futura	95,00%	2000	2,20%	2,1913%	-0,40%
24º	BR-08523/2022	Futura	95,00%	2000	2,20%	2,1913%	-0,40%
25º	BR-06280/2022	Futura	95,00%	2000	2,20%	2,1913%	-0,40%
26º	BR-08263/2022	Futura	95,00%	2000	2,20%	2,1913%	-0,40%
27º	BR-09573/2022	Paraná Pesquisas	95,00%	2400	2,00%	2,0004%	0,02%
28º	BR-07309/2022	Brasmarket	95,00%	2400	2,00%	2,0004%	0,02%
29º	BR-08584/2022	Brasmarket	95,00%	2400	2,00%	2,0004%	0,02%
30º	BR-08487/2022	Brasmarket	95,00%	2400	2,00%	2,0004%	0,02%
31º	BR-05389/2022	Brasmarket	95,00%	2400	2,00%	2,0004%	0,02%
32º	BR-08044/2022	Ipespe	95,45%	1100	3,00%	3,0151%	0,50%
33º	BR-06307/2022	Ipespe	95,45%	1100	3,00%	3,0151%	0,50%
34º	BR-01120/2022	Ipespe	95,45%	1100	3,00%	3,0151%	0,50%
35º	BR-05418/2022	Gerp	95,55%	2095	2,18%	2,1950%	0,69%
36º	BR-02322/2022	Gerp	95,55%	2095	2,18%	2,1950%	0,69%
37º	BR-07942/2022	Ipespe	95,45%	2000	2,20%	2,2361%	1,64%
38º	BR-05765/2022	Quaest	95,00%	2000	2,00%	2,1913%	9,57%
39º	BR-00470/2022	Quaest	95,00%	2000	2,00%	2,1913%	9,57%
40º	BR-04387/2022	Quaest	95,00%	2000	2,00%	2,1913%	9,57%
41º	BR-07106/2022	Quaest	95,00%	2000	2,00%	2,1913%	9,57%
42º	BR-02853/2022	Ipec	95,00%	2000	2,00%	2,1913%	9,57%
43º	BR-07940/2022	Quaest	95,00%	2000	2,00%	2,1913%	9,57%
44º	BR-02736/2022	Ipec	95,00%	2000	2,00%	2,1913%	9,57%
45º	BR-01560/2022	Nervera	95,00%	7500	1,00%	1,1316%	13,16%
46º	BR-06415/2022	Nervera	95,00%	4500	1,00%	1,4609%	46,09%
47º	BR-06012/2022	Nervera	95,00%	4500	1,00%	1,4609%	46,09%

Nota: Nível de confiança (c), tamanho amostral efetivo (n_e), margem de erro divulgada (E_d) e margem de erro prevista (E_p). Aplicação $E_p = z_c \sqrt{p^*q^*/n_e}$, em que $p^* = 0,5$, $q^* = 1 - p^*$ e z_c consiste no valor crítico para c.

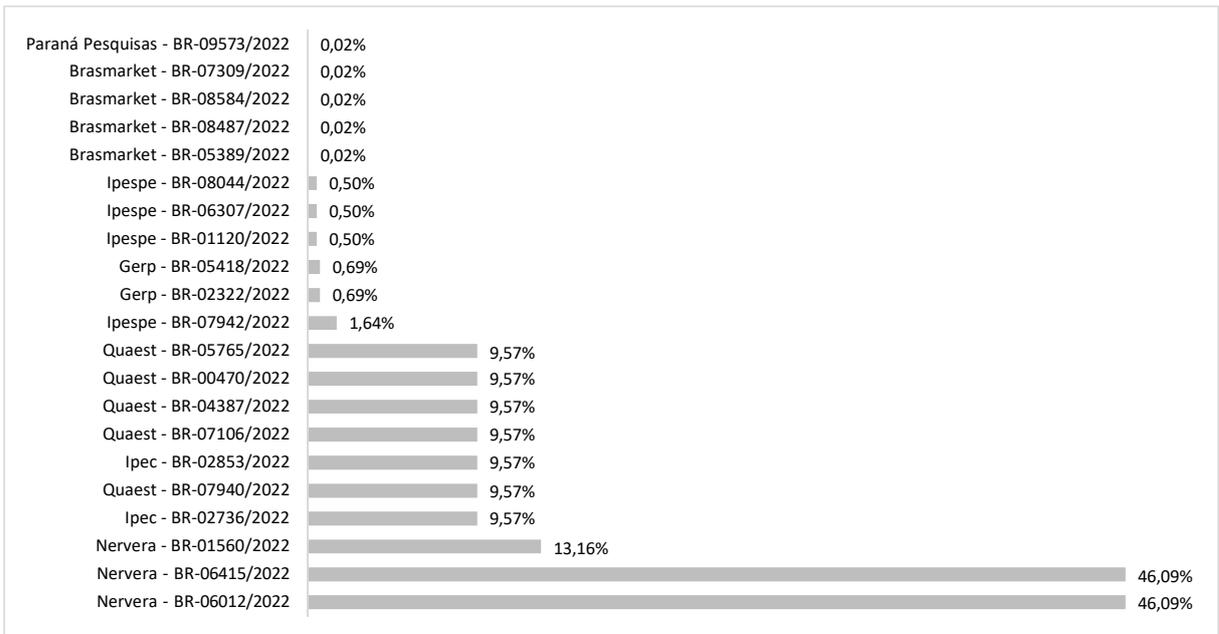
Fonte: Elaborado pelo autor.

Figura 04 – Ranking: Variações Percentuais Negativas (E_d e E_p)



Nota: Cálculos de diferenças percentuais entre margem de erro divulgada (E_d) e margem de erro prevista (E_p) a partir de $E_p = z_c \sqrt{p \cdot q} / n_e$. Diferenças negativas refletem em margens de erro divulgadas maiores que previstas. Fonte: Elaborado pelo autor.

Figura 05 – Ranking: Variações Percentuais Positivas (E_d e E_p)



Nota: Cálculos de diferenças percentuais entre margem de erro divulgada (E_d) e margem de erro prevista (E_p) a partir de $E_p = z_c \sqrt{p \cdot q} / n_e$. Diferenças positivas refletem em margens de erro divulgadas menores que previstas. Fonte: Elaborado pelo autor.

Na etapa seguinte, analisam-se as estimativas de votos válidos \hat{p} para cada candidato, desenvolvendo-se intervalos de confiança que inicialmente seguem margens de erro divulgadas. Assim, dimensionamento n_e acima de n_p tende a gerar intervalos que contenham o parâmetro populacional, em que a margem de erro efetiva é inferior à divulgada pelo dimensionamento amostral aplicado acima do previsto estatisticamente. Na Tabela 11 de precisão das estimativas intervalares para Lula, é observado se o parâmetro está contido no intervalo resultante, aliado a informações geradas analiticamente sobre dimensionamento amostral e margem de erro, investigando-se a frequência de parâmetros inseridos nos intervalos de confiança para registros com dimensionamento amostral acima e abaixo do previsto, o que significa investigar a precisão das estimativas intervalares com margem de erro abaixo e acima da prevista, respectivamente.

Em relação ao conjunto amostral, verifica-se que, das 25 pesquisas que possuem intervalos de confiança, $IC = \hat{p}_{IL} \pm E_d$, com o parâmetro populacional, 18 possuem tamanho amostral acima do previsto, representando 72% dos registros com esse resultado. Por outro lado, apenas 7 registros com n_e abaixo de n_p possuem estimativas contendo o parâmetro, correspondendo a 28% das pesquisas com tal resultado. Portanto, um dimensionamento amostral que segue, no mínimo, as características divulgadas de margem de erro e nível de confiança impacta na propriedade de intervalos de confiança contemplarem o parâmetro populacional. Considerando o total de 26 pesquisas com dimensionamento acima do previsto, 69,23% tiveram intervalos de confiança com o parâmetro, enquanto, do total de 21 registros com n_e abaixo de n_p , 33,33% geraram estimativas com o parâmetro populacional considerado.

Quanto aos demais registros analisados, observa-se que, das 22 pesquisas com intervalos de confiança sem o parâmetro, 14 possuem tamanho amostral abaixo do previsto, o que constitui 63,63% dos registros com esse resultado. Em contrapartida, 8 registros com n_e acima de n_p geraram estimativas sem o respectivo parâmetro, correspondendo a 36,36% dos registros com tal resultado. Deste modo, dimensionamentos amostrais abaixo do previsto, que não atendem requisitos de margem de confiança e nível de confiança divulgados, tendem a produzir estimativas intervalares que não contemplam o parâmetro populacional. Considerando o total de 21 pesquisas com dimensionamento amostral abaixo do previsto, 66,67% tiveram intervalos de confiança sem o parâmetro. Por outro lado, do total de 26 pesquisas com tamanho amostral acima do previsto, apenas 30,77% geraram estimativas sem o referido parâmetro populacional.

Tabela 10 – Estimativas Pontuais de Votos Totais e Válidos por Pesquisa

ID	Instituto	Término	Totais Lula	Totais Bolsonaro	Válidos Lula	Válidos Bolsonaro
BR-04217/2022	Veritá	29/10/2022	45,90%	47,80%	48,99%	51,01%
BR-08297/2022	Datafolha	29/10/2022	49,00%	45,00%	52,13%	47,87%
BR-05256/2022	Ipec	29/10/2022	50,00%	43,00%	53,76%	46,24%
BR-05765/2022	Quaest	29/10/2022	46,00%	43,00%	51,69%	48,31%
BR-01820/2022	MDA	28/10/2022	46,90%	44,90%	51,09%	48,91%
BR-09573/2022	Paraná Pesquisas	28/10/2022	47,10%	46,30%	50,43%	49,57%
BR-04920/2022	Veritá	27/10/2022	46,10%	49,00%	48,48%	51,52%
BR-07309/2022	Brasmarket	27/10/2022	42,70%	48,20%	46,97%	53,03%
BR-04208/2022	Datafolha	27/10/2022	49,00%	44,00%	52,69%	47,31%
BR-07903/2022	Futura	26/10/2022	46,60%	47,20%	49,68%	50,32%
BR-05418/2022	Gerp	26/10/2022	43,00%	47,00%	47,78%	52,22%
BR-01560/2022	Nervera	25/10/2022	52,40%	46,00%	53,25%	46,75%
BR-08584/2022	Brasmarket	25/10/2022	41,50%	48,00%	46,37%	53,63%
BR-00470/2022	Quaest	25/10/2022	48,00%	42,00%	53,33%	46,67%
BR-01159/2022	PoderData	25/10/2022	49,00%	44,00%	52,69%	47,31%
BR-08044/2022	Ipespe	24/10/2022	50,00%	44,00%	53,19%	46,81%
BR-00525/2022	Paraná Pesquisas	24/10/2022	46,30%	45,90%	50,22%	49,78%
BR-06043/2022	Ipec	24/10/2022	50,00%	43,00%	53,76%	46,24%
BR-08487/2022	Brasmarket	23/10/2022	41,80%	47,70%	46,70%	53,30%
BR-06415/2022	Nervera	22/10/2022	52,00%	46,20%	52,95%	47,05%
BR-04043/2022	Veritá	20/10/2022	46,70%	49,40%	48,60%	51,40%
BR-08523/2022	Futura	19/10/2022	45,90%	46,90%	49,46%	50,54%
BR-00053/2022	Ideia	19/10/2022	50,00%	46,00%	52,08%	47,92%
BR-05389/2022	Brasmarket	19/10/2022	40,80%	45,50%	47,28%	52,72%
BR-02276/2022	Paraná Pesquisas	19/10/2022	46,90%	44,50%	51,31%	48,69%
BR-07340/2022	Datafolha	19/10/2022	49,00%	45,00%	52,13%	47,87%
BR-08917/2022	PoderData	18/10/2022	48,00%	44,00%	52,17%	47,83%
BR-04387/2022	Quaest	18/10/2022	47,00%	42,00%	52,81%	47,19%
BR-06307/2022	Ipespe	18/10/2022	49,00%	43,00%	53,26%	46,74%
BR-02707/2022	Ipec	17/10/2022	50,00%	43,00%	53,76%	46,24%
BR-05514/2022	MDA	16/10/2022	48,10%	41,80%	53,50%	46,50%
BR-04850/2022	Veritá	15/10/2022	46,40%	48,70%	48,79%	51,21%
BR-01682/2022	Datafolha	14/10/2022	49,00%	44,00%	52,69%	47,31%
BR-06012/2022	Nervera	12/10/2022	51,10%	46,50%	52,36%	47,64%
BR-06280/2022	Futura	12/10/2022	46,90%	46,50%	50,21%	49,79%
BR-07942/2022	Ipespe	12/10/2022	49,00%	43,00%	53,26%	46,74%
BR-07106/2022	Quaest	12/10/2022	49,00%	41,00%	54,44%	45,56%
BR-08438/2022	Paraná Pesquisas	12/10/2022	47,60%	44,10%	51,91%	48,09%
BR-09241/2022	PoderData	11/10/2022	48,00%	44,00%	52,17%	47,83%
BR-02322/2022	Gerp	11/10/2022	48,00%	46,00%	51,06%	48,94%
BR-01120/2022	Ipespe	10/10/2022	50,00%	43,00%	53,76%	46,24%
BR-02853/2022	Ipec	10/10/2022	51,00%	42,00%	54,84%	45,16%
BR-02012/2022	Datafolha	07/10/2022	49,00%	44,00%	52,69%	47,31%
BR-07940/2022	Quaest	05/10/2022	48,00%	41,00%	53,93%	46,07%
BR-08253/2022	PoderData	05/10/2022	48,00%	44,00%	52,17%	47,83%
BR-02736/2022	Ipec	05/10/2022	51,00%	43,00%	54,26%	45,74%
BR-08263/2022	Futura	05/10/2022	54,00%	38,00%	58,70%	41,30%

Nota: Estimativas pontuais em relação ao total de entrevistados (considerando outras respostas como “branco”, “nulo”, “não votará”, “não sabe”, “não respondeu”) seguidas pelas estimativas de votos válidos em cada candidato. Valores padronizados em duas casas decimais para se manter a consistência na execução de resultados posteriores. Fonte: Elaborado pelo autor.

Tabela 11 – Precisão das Estimativas Intervalares em Relação ao Parâmetro Populacional na Perspectiva da Margem de Erro Divulgada (E_d): Lula

ID	$\Delta\%$ (n_p, n_e)	$\Delta\%$ (E_d, E_p)	E_d	\hat{p}	IC _{inferior}	IC _{superior}	Parâmetro
BR-04217/2022	631,03%	-63,02%	2,00% ^{AC}	48,99%	46,99%	50,99%	Dentro
BR-08297/2022	246,52%	-46,28%	2,00% ^{AC}	52,13%	50,13%	54,13%	Dentro
BR-05256/2022	77,93%	-25,03%	2,00% ^{AC}	53,76%	51,76%	55,76%	Fora
BR-05765/2022	-16,70%	9,57%	2,00% ^{AB}	51,69%	49,69%	53,69%	Dentro
BR-01820/2022	0,86%	-0,44%	2,20% ^{AC}	51,09%	48,89%	53,29%	Dentro
BR-09573/2022	-0,04%	0,02%	2,00% ^{AB}	50,43%	48,43%	52,43%	Dentro
BR-04920/2022	1156,98%	-71,79%	2,00% ^{AC}	48,48%	46,48%	50,48%	Fora
BR-07309/2022	-0,04%	0,02%	2,00% ^{AB}	46,97%	44,97%	48,97%	Fora
BR-04208/2022	91,25%	-27,69%	2,00% ^{AC}	52,69%	50,69%	54,69%	Dentro
BR-07903/2022	0,76%	-0,40%	2,20% ^{AC}	49,68%	47,48%	51,88%	Dentro
BR-05418/2022	-1,37%	0,69%	2,18% ^{AB}	47,78%	45,60%	49,96%	Fora
BR-01560/2022	-21,91%	13,16%	1,00% ^{AB}	53,25%	52,25%	54,25%	Fora
BR-08584/2022	-0,04%	0,02%	2,00% ^{AB}	46,37%	44,37%	48,37%	Fora
BR-00470/2022	-16,70%	9,57%	2,00% ^{AB}	53,33%	51,33%	55,33%	Fora
BR-01159/2022	108,25%	-30,70%	2,00% ^{AC}	52,69%	50,69%	54,69%	Dentro
BR-08044/2022	-1,08%	0,50%	3,00% ^{AB}	53,19%	50,19%	56,19%	Dentro
BR-00525/2022	1,76%	-0,89%	2,20% ^{AC}	50,22%	48,02%	52,42%	Dentro
BR-06043/2022	25,28%	-10,66%	2,00% ^{AC}	53,76%	51,76%	55,76%	Fora
BR-08487/2022	-0,04%	0,02%	2,00% ^{AB}	46,70%	44,70%	48,70%	Fora
BR-06415/2022	-53,14%	46,09%	1,00% ^{AB}	52,95%	51,95%	53,95%	Fora
BR-04043/2022	742,15%	-65,54%	2,00% ^{AC}	48,60%	46,60%	50,60%	Fora
BR-08523/2022	0,76%	-0,40%	2,20% ^{AC}	49,46%	47,26%	51,66%	Dentro
BR-00053/2022	40,45%	-15,66%	3,00% ^{AC}	52,08%	49,08%	55,08%	Dentro
BR-05389/2022	-0,04%	0,02%	2,00% ^{AB}	47,28%	45,28%	49,28%	Fora
BR-02276/2022	1,76%	-0,89%	2,20% ^{AC}	51,31%	49,11%	53,51%	Dentro
BR-07340/2022	21,28%	-9,20%	2,00% ^{AC}	52,13%	50,13%	54,13%	Dentro
BR-08917/2022	251,15%	-46,64%	2,00% ^{AC}	52,17%	50,17%	54,17%	Dentro
BR-04387/2022	-16,70%	9,57%	2,00% ^{AB}	52,81%	50,81%	54,81%	Dentro
BR-06307/2022	-1,08%	0,50%	3,00% ^{AB}	53,26%	50,26%	56,26%	Dentro
BR-02707/2022	25,28%	-10,66%	2,00% ^{AC}	53,76%	51,76%	55,76%	Fora
BR-05514/2022	0,86%	-0,44%	2,20% ^{AC}	53,50%	51,30%	55,70%	Fora
BR-04850/2022	130,24%	-34,10%	2,00% ^{AC}	48,79%	46,79%	50,79%	Fora
BR-01682/2022	20,70%	-8,98%	2,00% ^{AC}	52,69%	50,69%	54,69%	Dentro
BR-06012/2022	-53,14%	46,09%	1,00% ^{AB}	52,36%	51,36%	53,36%	Fora
BR-06280/2022	0,76%	-0,40%	2,20% ^{AC}	50,21%	48,01%	52,41%	Dentro
BR-07942/2022	-3,24%	1,64%	2,20% ^{AB}	53,26%	51,06%	55,46%	Fora
BR-07106/2022	-16,70%	9,57%	2,00% ^{AB}	54,44%	52,44%	56,44%	Fora
BR-08438/2022	1,76%	-0,89%	2,20% ^{AC}	51,91%	49,71%	54,11%	Dentro
BR-09241/2022	108,25%	-30,70%	2,00% ^{AC}	52,17%	50,17%	54,17%	Dentro
BR-02322/2022	-1,37%	0,69%	2,18% ^{AB}	51,06%	48,88%	53,24%	Dentro
BR-01120/2022	-1,08%	0,50%	3,00% ^{AB}	53,76%	50,76%	56,76%	Dentro
BR-02853/2022	-16,70%	9,57%	2,00% ^{AB}	54,84%	52,84%	56,84%	Fora
BR-02012/2022	20,12%	-8,76%	2,00% ^{AC}	52,69%	50,69%	54,69%	Dentro
BR-07940/2022	-16,70%	9,57%	2,00% ^{AB}	53,93%	51,93%	55,93%	Fora
BR-08253/2022	45,77%	-17,18%	2,00% ^{AC}	52,17%	50,17%	54,17%	Dentro
BR-02736/2022	-16,70%	9,57%	2,00% ^{AB}	54,26%	52,26%	56,26%	Fora
BR-08263/2022	0,76%	-0,40%	2,20% ^{AC}	58,70%	56,50%	60,90%	Fora

Nota: Intervalos de confiança (IC) com limite superior e inferior usando a margem de erro divulgada (E_d). Construção a partir de $IC = \hat{p} \pm E_d$, em que \hat{p} é a estimativa pontual da proporção considerando votos válidos. Parâmetro populacional para Lula é de 50,90%, conforme resultado do 2º turno das Eleições Gerais 2022. Sobrescritos AC/AB representam E_d acima/abaixo da margem prevista E_p pelo dimensionamento amostral n_e .

Fonte: Elaborado pelo autor.

Desenvolve-se análise semelhante a partir da Tabela 12, relacionando-se as estimativas de votos válidos \hat{p} às variações percentuais sobre dimensionamento amostral e margem de erro previstos e efetivos para Bolsonaro usando-se margens de erro divulgadas pelos institutos. Portanto, no contexto em que a margem de erro divulgada E_d é menor que a prevista E_p , intervalos de confiança gerados tendem a não incluir o parâmetro populacional, refletindo que o tamanho amostral n_e não corresponde à margem de erro e nível de confiança estabelecidos. Deste modo, investiga-se, dentre as estimativas intervalares com e sem o referido parâmetro, a proporção destas com dimensionamento amostral acima e baixo do previsto, assim como, em relação às pesquisas com n_e acima e abaixo de n_p , a proporção destas que geraram intervalos de confiança com o parâmetro populacional, vis-à-vis àquelas sem esse parâmetro considerado.

Considerando a amostra de registros, observa-se que, das 25 pesquisas com intervalos de confiança $IC = \hat{p}_{iB} \pm E_d$ que contemplam o parâmetro populacional, 18 possuem tamanho amostral acima do previsto, o que caracteriza 72% dos registros com esse resultado. Em contrapartida, somente 7 pesquisas com n_e abaixo de n_p geraram estimativas contendo o parâmetro, representando 28% dos registros com tal característica. Deste modo, se a pesquisa não possui tamanho amostral que corrobore a margem de erro e nível de confiança divulgados, as estimativas intervalares estarão comprometidas, pois os critérios se interrelacionam e não serão aqueles atributos comunicados. Em relação aos 26 registros com tamanho amostral acima do previsto, 69,23% produziram intervalos de confiança contendo o parâmetro, ao passo que, de 21 pesquisas com n_e abaixo de n_p , 33,33% contêm estimativas intervalares com o parâmetro.

Observando-se os demais registros, das 22 pesquisas cujos intervalos de confiança não contêm o parâmetro populacional, 14 foram conduzidas com dimensionamento amostral efetivo abaixo do previsto, compreendendo 63,63% dos registros com esse resultado. Por outro lado, 8 pesquisas com n_e acima de n_p resultaram em estimativas sem englobar o parâmetro em análise, representando 36,36% dos registros com tal resultado. Portanto, quando o dimensionamento amostral está abaixo do previsto para manter a margem de erro e nível de confiança divulgados, as pesquisas terão suas propriedades alteradas e intervalos de confiança serão comprometidos. Em relação às 21 pesquisas com n_e abaixo de n_p , 66,67% possuem estimativas sem o parâmetro enquanto, das 26 pesquisas com n_e acima de n_p , apenas 30,77% geraram estimativas sem este. Procedimentos em planilha eletrônica e software R para estas análises constam no Apêndice C.

Tabela 12 – Precisão das Estimativas Intervalares em Relação ao Parâmetro Populacional na Perspectiva da Margem de Erro Divulgada (E_d): Bolsonaro

ID	$\Delta\%$ (n_p, n_e)	$\Delta\%$ (E_d, E_p)	E_d	\hat{p}	IC _{inferior}	IC _{superior}	Parâmetro
BR-04217/2022	631,03%	-63,02%	2,00% ^{AC}	51,01%	49,01%	53,01%	Dentro
BR-08297/2022	246,52%	-46,28%	2,00% ^{AC}	47,87%	45,87%	49,87%	Dentro
BR-05256/2022	77,93%	-25,03%	2,00% ^{AC}	46,24%	44,24%	48,24%	Fora
BR-05765/2022	-16,70%	9,57%	2,00% ^{AB}	48,31%	46,31%	50,31%	Dentro
BR-01820/2022	0,86%	-0,44%	2,20% ^{AC}	48,91%	46,71%	51,11%	Dentro
BR-09573/2022	-0,04%	0,02%	2,00% ^{AB}	49,57%	47,57%	51,57%	Dentro
BR-04920/2022	1156,98%	-71,79%	2,00% ^{AC}	51,52%	49,52%	53,52%	Fora
BR-07309/2022	-0,04%	0,02%	2,00% ^{AB}	53,03%	51,03%	55,03%	Fora
BR-04208/2022	91,25%	-27,69%	2,00% ^{AC}	47,31%	45,31%	49,31%	Dentro
BR-07903/2022	0,76%	-0,40%	2,20% ^{AC}	50,32%	48,12%	52,52%	Dentro
BR-05418/2022	-1,37%	0,69%	2,18% ^{AB}	52,22%	50,04%	54,40%	Fora
BR-01560/2022	-21,91%	13,16%	1,00% ^{AB}	46,75%	45,75%	47,75%	Fora
BR-08584/2022	-0,04%	0,02%	2,00% ^{AB}	53,63%	51,63%	55,63%	Fora
BR-00470/2022	-16,70%	9,57%	2,00% ^{AB}	46,67%	44,67%	48,67%	Fora
BR-01159/2022	108,25%	-30,70%	2,00% ^{AC}	47,31%	45,31%	49,31%	Dentro
BR-08044/2022	-1,08%	0,50%	3,00% ^{AB}	46,81%	43,81%	49,81%	Dentro
BR-00525/2022	1,76%	-0,89%	2,20% ^{AC}	49,78%	47,58%	51,98%	Dentro
BR-06043/2022	25,28%	-10,66%	2,00% ^{AC}	46,24%	44,24%	48,24%	Fora
BR-08487/2022	-0,04%	0,02%	2,00% ^{AB}	53,30%	51,30%	55,30%	Fora
BR-06415/2022	-53,14%	46,09%	1,00% ^{AB}	47,05%	46,05%	48,05%	Fora
BR-04043/2022	742,15%	-65,54%	2,00% ^{AC}	51,40%	49,40%	53,40%	Fora
BR-08523/2022	0,76%	-0,40%	2,20% ^{AC}	50,54%	48,34%	52,74%	Dentro
BR-00053/2022	40,45%	-15,66%	3,00% ^{AC}	47,92%	44,92%	50,92%	Dentro
BR-05389/2022	-0,04%	0,02%	2,00% ^{AB}	52,72%	50,72%	54,72%	Fora
BR-02276/2022	1,76%	-0,89%	2,20% ^{AC}	48,69%	46,49%	50,89%	Dentro
BR-07340/2022	21,28%	-9,20%	2,00% ^{AC}	47,87%	45,87%	49,87%	Dentro
BR-08917/2022	251,15%	-46,64%	2,00% ^{AC}	47,83%	45,83%	49,83%	Dentro
BR-04387/2022	-16,70%	9,57%	2,00% ^{AB}	47,19%	45,19%	49,19%	Dentro
BR-06307/2022	-1,08%	0,50%	3,00% ^{AB}	46,74%	43,74%	49,74%	Dentro
BR-02707/2022	25,28%	-10,66%	2,00% ^{AC}	46,24%	44,24%	48,24%	Fora
BR-05514/2022	0,86%	-0,44%	2,20% ^{AC}	46,50%	44,30%	48,70%	Fora
BR-04850/2022	130,24%	-34,10%	2,00% ^{AC}	51,21%	49,21%	53,21%	Fora
BR-01682/2022	20,70%	-8,98%	2,00% ^{AC}	47,31%	45,31%	49,31%	Dentro
BR-06012/2022	-53,14%	46,09%	1,00% ^{AB}	47,64%	46,64%	48,64%	Fora
BR-06280/2022	0,76%	-0,40%	2,20% ^{AC}	49,79%	47,59%	51,99%	Dentro
BR-07942/2022	-3,24%	1,64%	2,20% ^{AB}	46,74%	44,54%	48,94%	Fora
BR-07106/2022	-16,70%	9,57%	2,00% ^{AB}	45,56%	43,56%	47,56%	Fora
BR-08438/2022	1,76%	-0,89%	2,20% ^{AC}	48,09%	45,89%	50,29%	Dentro
BR-09241/2022	108,25%	-30,70%	2,00% ^{AC}	47,83%	45,83%	49,83%	Dentro
BR-02322/2022	-1,37%	0,69%	2,18% ^{AB}	48,94%	46,76%	51,12%	Dentro
BR-01120/2022	-1,08%	0,50%	3,00% ^{AB}	46,24%	43,24%	49,24%	Dentro
BR-02853/2022	-16,70%	9,57%	2,00% ^{AB}	45,16%	43,16%	47,16%	Fora
BR-02012/2022	20,12%	-8,76%	2,00% ^{AC}	47,31%	45,31%	49,31%	Dentro
BR-07940/2022	-16,70%	9,57%	2,00% ^{AB}	46,07%	44,07%	48,07%	Fora
BR-08253/2022	45,77%	-17,18%	2,00% ^{AC}	47,83%	45,83%	49,83%	Dentro
BR-02736/2022	-16,70%	9,57%	2,00% ^{AB}	45,74%	43,74%	47,74%	Fora
BR-08263/2022	0,76%	-0,40%	2,20% ^{AC}	41,30%	39,10%	43,50%	Fora

Nota: Intervalos de confiança (IC) com limite superior e inferior usando a margem de erro divulgada (E_d). Construção a partir de $IC = \hat{p} \pm E_d$, em que \hat{p} é a estimativa pontual da proporção considerando votos válidos. Parâmetro populacional para Bolsonaro é de 49,10%, conforme resultado do 2º turno das Eleições Gerais 2022. Sobrescritos AC/AB representam E_d acima/abaixo da margem prevista E_p pelo dimensionamento amostral n_e .

Fonte: Elaborado pelo autor.

Nas etapas subsequentes, avalia-se a precisão das estimativas intervalares usando-se margens de erro efetivas (E_e) e testando-se a hipótese nula $H_0: p = p_0$ para avaliar a distância estatística padronizada entre proporção amostral \hat{p} e valor hipotetizado p_0 que, no contexto, é a proporção verificada de votos válidos no respectivo candidato. Se a estatística de teste Z padronizada for significativa, em módulo, equivale a afirmar que, se o parâmetro hipotetizado p_0 fosse verdadeiro (sabe-se que o é, de fato), a proporção amostral \hat{p} não seria proveniente da mesma população, sendo uma amostra não representativa. De outro modo, equivale a considerar que a pesquisa político-eleitoral sugere que o parâmetro é diferente da proporção já verificada, refletindo na rejeição da hipótese nula, pois o p-valor, sendo a probabilidade de se rejeitar H_0 quando esta é verdadeira ou de se obter uma estatística de teste tão ou mais extrema, é pequeno.

A diferença entre E_p e E_e decorre do fato de que a primeira é a margem de erro máxima sobre a qual a pesquisa político-eleitoral exerce controle por meio do tamanho amostral, nível e confiança e da estimativa preliminar $E_p = z_c \sqrt{p^*q^*/n_e}$ em que, por padrão, $p^* = 0,5$, enquanto a segunda provém das proporções amostrais \hat{p} encontradas $E_e = z_c \sqrt{\hat{p}\hat{q}/n_e}$. Portanto, E_e não excede E_p , sendo menor ou, no máximo, equivalente à margem prevista analiticamente, ao contrário da margem de erro divulgada E_d estabelecida pelo próprio instituto e não segue, necessariamente, as premissas de margem de erro e nível de confiança divulgados ao público, conforme verificado nas análises anteriores. Considera-se que, nos casos em que n_p exceda n_e , as estimativas intervalares tendem a não contemplar o parâmetro, assim como outros aspectos metodológicos em coleta de dados tendam a estar comprometidos, repercutindo nas estimativas.

As rejeições da hipótese nula $H_0: p = p_0$, quando ocorrem, são analisadas a partir dos níveis de significância α de 1%, 5% e 10%, comparando-os com o p-valor da estatística de teste. Assim, p-valor abaixo desses níveis reflete na rejeição de H_0 para dado nível, em que 1% indica que a probabilidade de se rejeitar H_0 quando esta é verdadeira é de apenas 1% ou menos. Portanto, quanto menor o p-valor, mais distante a proporção amostral para tal candidato se encontra do parâmetro conhecido pelo resultado eleitoral (valor central de uma distribuição normal padronizada), implicando, no caso de rejeição de H_0 , em intervalo de confiança sem o parâmetro para um nível de confiança de $1 - \alpha$. No contexto de tamanhos amostrais mais expressivos, o intervalo de confiança será mais restrito, analisando se as tendências acerca do dimensionamento amostral e estimativas intervalares permanecem sob avaliação mais rigorosa.

Na Tabela 13, considerando os 26 registros de tamanho amostral n_e acima do previsto n_p , em que $E_p < E_d$ por propriedade e $E_e < E_d$ nesse contexto, verifica-se que 13 (ou 50%) destes geraram estimativas $IC = \hat{p}_{iL} \pm E_e$ contendo o parâmetro. Assim, comparando-se com análises anteriores, 5 intervalos que continham o parâmetro a partir da margem divulgada E_d passaram a não o contemplar com E_e . De outro modo, 13 registros geraram intervalos sem o parâmetro, ao passo que, com E_d , eram apenas 8 registros com esse resultado no conjunto amostral de pesquisas com dimensionamento amostral efetivo n_e acima do previsto, enquanto 18 registros geraram estimativas intervalares com o parâmetro. Nota-se que as 5 referidas pesquisas com alteração de resultado são de Veritá, Datafolha e PoderData, com 2 registros nos últimos, situando-se, conforme análise prévia, entre as 10 de maior variabilidade entre n_p e n_e .

Considerando-se os resultados das 21 pesquisas de tamanho amostral n_e abaixo do previsto n_p , em que $E_p > E_d$ por propriedade e geralmente $E_e > E_d$, exceto em casos nos quais a diferença percentual entre n_p e n_e é pequena (menos de 1%) e $\hat{p} \neq 0,5$, como os registros da Brasmaket, observa-se que 7 (ou 33,33%) geraram estimativas contemplando o parâmetro. Deste modo, relacionando-se com análises prévias, 14 pesquisas que não possuíam o parâmetro com a margem divulgada E_d permaneceram sem o parâmetro sob E_e . Portanto, no caso de dimensionamento amostral abaixo do previsto analiticamente para atender à margem de erro e intervalo de confiança estabelecidos, as frequências de pesquisas que compreendem ou não o parâmetro são mantidas, mesmo com margem geralmente mais ampla. Assim, essa tendência difere do caso $n_e > n_p$, em que intervalos sem o parâmetro aumentam, porém não predominam.

Avaliando-se o p-valor, que é a probabilidade de se rejeitar a hipótese nula quando esta é verdadeira, dentre as pesquisas de estimativas intervalares que contém o parâmetro com E_e , não se rejeita $H_0: p = p_0$ para 15 registros, ao passo que se rejeita a hipótese nula para 5 deles, porém ao nível de significância limítrofe de 10%. Esta tendência é esperada, pois a não rejeição de H_0 significa que, se $p_0 = 50,90\%$ for o parâmetro populacional de votos válidos para Lula, não seria incomum a obtenção de \hat{p} como proporção amostral. Deste modo, a não rejeição da hipótese nula implica que a respectiva \hat{p} é estatisticamente próxima de p_0 , em que este é o valor central de uma distribuição normal padrão. Dentre os registros sem o parâmetro, rejeições de H_0 ocorre 1 vez ao nível de 10%; 7 vezes ao nível de 5%; e 19 vezes ao nível de 1%. Rejeições de H_0 refletem em probabilidade igual ou menor que α de se obter \hat{p} mais extremo caso $p = p_0$.

Tabela 13 – Precisão das Estimativas Intervalares em Relação ao Parâmetro Populacional na Perspectiva da Margem de Erro Efetiva (E_e): Lula

ID	E_e	\hat{p}	IC _{inferior}	IC _{superior}	Parâmetro	Z	p-valor	Sig.
BR-04217/2022	0,7395% ^{AC}	48,99%	48,25%	49,73%	Fora	-5,06	0,0000	***
BR-08297/2022	1,0734% ^{AC}	52,13%	51,06%	53,20%	Fora	2,24	0,0248	**
BR-05256/2022	1,4951% ^{AC}	53,76%	52,26%	55,26%	Fora	3,74	0,0002	***
BR-05765/2022	2,1901% ^{AB}	51,69%	49,50%	53,88%	Dentro	0,71	0,4797	
BR-01820/2022	2,1897% ^{AC}	51,09%	48,90%	53,28%	Dentro	0,17	0,8650	
BR-09573/2022	2,0003% ^{AB}	50,43%	48,43%	52,43%	Dentro	-0,46	0,6451	
BR-04920/2022	0,5638% ^{AC}	48,48%	47,92%	49,04%	Fora	-8,41	0,0000	***
BR-07309/2022	1,9967% ^{AB}	46,97%	44,97%	48,97%	Fora	-3,85	0,0001	***
BR-04208/2022	1,4441% ^{AC}	52,69%	51,25%	54,13%	Fora	2,43	0,0153	**
BR-07903/2022	2,1913% ^{AC}	49,68%	47,49%	51,87%	Dentro	-1,09	0,2751	
BR-05418/2022	2,1928% ^{AB}	47,78%	45,59%	49,97%	Fora	-2,86	0,0043	***
BR-01560/2022	1,1292% ^{AB}	53,25%	52,12%	54,38%	Fora	4,07	0,0000	***
BR-08584/2022	1,9951% ^{AB}	46,37%	44,37%	48,37%	Fora	-4,44	0,0000	***
BR-00470/2022	2,1864% ^{AB}	53,33%	51,14%	55,52%	Fora	2,17	0,0297	**
BR-01159/2022	1,3839% ^{AC}	52,69%	51,31%	54,07%	Fora	2,53	0,0113	**
BR-08044/2022	3,0090% ^{AB}	53,19%	50,18%	56,20%	Dentro	1,52	0,1287	
BR-00525/2022	2,1804% ^{AC}	50,22%	48,04%	52,40%	Dentro	-0,61	0,5410	
BR-06043/2022	1,7818% ^{AC}	53,76%	51,98%	55,54%	Fora	3,14	0,0017	***
BR-08487/2022	1,9960% ^{AB}	46,70%	44,70%	48,70%	Fora	-4,12	0,0000	***
BR-06415/2022	1,4583% ^{AB}	52,95%	51,49%	54,41%	Fora	2,75	0,0059	***
BR-04043/2022	0,6889% ^{AC}	48,60%	47,91%	49,29%	Fora	-6,54	0,0000	***
BR-08523/2022	2,1912% ^{AC}	49,46%	47,27%	51,65%	Dentro	-1,29	0,1977	
BR-00053/2022	2,5281% ^{AC}	52,08%	49,55%	54,61%	Dentro	0,91	0,3606	
BR-05389/2022	1,9974% ^{AB}	47,28%	45,28%	49,28%	Fora	-3,55	0,0004	***
BR-02276/2022	2,1797% ^{AC}	51,31%	49,13%	53,49%	Dentro	0,37	0,7124	
BR-07340/2022	1,8144% ^{AC}	52,13%	50,32%	53,94%	Dentro	1,33	0,1843	
BR-08917/2022	1,0663% ^{AC}	52,17%	51,10%	53,24%	Fora	2,33	0,0197	**
BR-04387/2022	2,1878% ^{AB}	52,81%	50,62%	55,00%	Dentro	1,71	0,0875	*
BR-06307/2022	3,0087% ^{AB}	53,26%	50,25%	56,27%	Dentro	1,57	0,1174	
BR-02707/2022	1,7818% ^{AC}	53,76%	51,98%	55,54%	Fora	3,14	0,0017	***
BR-05514/2022	2,1848% ^{AC}	53,50%	51,32%	55,68%	Fora	2,33	0,0200	**
BR-04850/2022	1,3177% ^{AC}	48,79%	47,47%	50,11%	Fora	-3,14	0,0017	***
BR-01682/2022	1,8178% ^{AC}	52,69%	50,87%	54,51%	Dentro	1,93	0,0539	*
BR-06012/2022	1,4592% ^{AB}	52,36%	50,90%	53,82%	Fora	1,96	0,0501	*
BR-06280/2022	2,1913% ^{AC}	50,21%	48,02%	52,40%	Dentro	-0,62	0,5371	
BR-07942/2022	2,2313% ^{AB}	53,26%	51,03%	55,49%	Fora	2,11	0,0348	**
BR-07106/2022	2,1826% ^{AB}	54,44%	52,26%	56,62%	Fora	3,17	0,0015	***
BR-08438/2022	2,1788% ^{AC}	51,91%	49,73%	54,09%	Dentro	0,91	0,3639	
BR-09241/2022	1,3846% ^{AC}	52,17%	50,79%	53,55%	Dentro	1,80	0,0724	*
BR-02322/2022	2,1945% ^{AB}	51,06%	48,87%	53,25%	Dentro	0,15	0,8835	
BR-01120/2022	3,0066% ^{AB}	53,76%	50,75%	56,77%	Dentro	1,90	0,0578	*
BR-02853/2022	2,1810% ^{AB}	54,84%	52,66%	57,02%	Fora	3,52	0,0004	***
BR-02012/2022	1,8222% ^{AC}	52,69%	50,87%	54,51%	Dentro	1,92	0,0545	*
BR-07940/2022	2,1845% ^{AB}	53,93%	51,75%	56,11%	Fora	2,71	0,0067	***
BR-08253/2022	1,6549% ^{AC}	52,17%	50,52%	53,82%	Dentro	1,50	0,1329	
BR-02736/2022	2,1833% ^{AB}	54,26%	52,08%	56,44%	Fora	3,01	0,0026	***
BR-08263/2022	2,1579% ^{AC}	58,70%	56,54%	60,86%	Fora	6,98	0,0000	***

Nota: Construção de IC = $\hat{p} \pm E_e$, em que \hat{p} é a estimativa pontual da proporção considerando votos válidos e E_e , a margem de erro efetiva. Parâmetro populacional para Lula é de 50,90%, conforme resultado do 2º turno das Eleições Gerais 2022. Significância (sig.) aos níveis de 1% (***), 5% (**) e 10% (*) a partir do cálculo da estatística Z e p-valor do teste bilateral com a hipótese nula (H_0) de que o parâmetro populacional é igual a 50,90%. Sobrescritos AC/AB representam dimensionamento efetivo n_e acima/abaixo do dimensionamento previsto n_p .

Fonte: Elaborado pelo autor.

Na Tabela 14, avaliam-se as 26 pesquisas com dimensionamento amostral n_e acima do previsto n_p , o que implica em $E_p < E_d$, pois a margem prevista resultante de n_e situa-se abaixo de n_p decorrente da margem de erro divulgada. Assim, $E_e < E_d$, dado que o valor máximo que margem de erro efetiva pode atingir é $E_e = E_p$, se a proporção amostral $\hat{p} = 0,50$. Deste modo, verifica-se que 13 (ou 50%) destas produziram intervalos $IC = \hat{p}_{iB} \pm E_e$ incluindo o parâmetro. Alternativamente, 13 registros com essa característica produziram intervalos de confiança sem o parâmetro adotando-se a perspectiva de E_e , enquanto 8 registros incorreram em tal resultado com E_d . Nota-se que as 5 pesquisas com alteração de resultado pertencem a Veritá, Datafolha e PoderData, com dois registros para os últimos, estando estas entre as 10 pesquisas com maior dimensionamento efetivo em relação ao previsto, e uma delas entre as 3 de maior razão $R\$/n_e$.

Avaliando-se os resultados dos 21 registros com tamanho amostral n_e abaixo do previsto n_p em que $E_p > E_d$ por propriedade e $E_e > E_d$, exceto quando $\Delta\%(n_p, n_e) < |1\%|$ como nos registros da Brasmarket, 7 (ou 33,33%) geraram estimativas sem o parâmetro. Assim, os 14 intervalos de confiança que não englobaram o resultado verificado nas eleições usando-se E_d mantiveram as conclusões também a partir de E_e . Deste modo, no contexto de dimensionamento amostral abaixo do previsto, a frequência de intervalos sem o parâmetro se mantém, mesmo considerando margens de erro efetivas tipicamente mais amplas. Nesse sentido, observa-se a importância de um dimensionamento amostral que mantenha a margem de erro máxima divulgada, sendo condição necessária frequentemente não seguida pelos institutos para se ter estimativas intervalares que englobem o parâmetro independentemente da adoção de E_d ou E_e .

Examinando-se o p-valor, caracterizado pela probabilidade de cometer o Erro Tipo I, dentre intervalos gerados sob E_e , não se rejeita $H_0: p = p_0$ para 15 registros, enquanto H_0 é rejeitado em 5 pesquisas ao nível de significância limítrofe de 10%. Assim, a não rejeição da hipótese nula para dado nível de significância implica que estimativas intervalares com nível de confiança de $1 - \alpha$ incluem o parâmetro hipotetizado. De outro modo, a não rejeição de H_0 implica que, se $p_0 = 49,10\%$ for o parâmetro populacional de votos válidos para Bolsonaro, determinado \hat{p} seria estatisticamente próximo de p_0 , não sendo um valor incomum ou extremo. Dentre os registros que não englobam o parâmetro em análise, rejeição ao nível de 10% ocorre 1 vez; ao nível de significância de 5%, ocorrem 7 vezes; e ao nível de 1% de significância, 19 vezes. Portanto, rejeições de H_0 ao nível de α refletem em intervalos sem p_0 quando $c = 1 - \alpha$.

Tabela 14 – Precisão das Estimativas Intervalares em Relação ao Parâmetro Populacional na Perspectiva da Margem de Erro Efetiva (E_e): Bolsonaro

ID	E_e	\hat{p}	IC _{inferior}	IC _{superior}	Parâmetro	Z	p-valor	Sig.
BR-04217/2022	0,7395% ^{AC}	51,01%	50,27%	51,75%	Fora	5,06	0,0000	***
BR-08297/2022	1,0734% ^{AC}	47,87%	46,80%	48,94%	Fora	-2,24	0,0248	**
BR-05256/2022	1,4951% ^{AC}	46,24%	44,74%	47,74%	Fora	-3,74	0,0002	***
BR-05765/2022	2,1901% ^{AB}	48,31%	46,12%	50,50%	Dentro	-0,71	0,4797	
BR-01820/2022	2,1897% ^{AC}	48,91%	46,72%	51,10%	Dentro	-0,17	0,8650	
BR-09573/2022	2,0003% ^{AB}	49,57%	47,57%	51,57%	Dentro	0,46	0,6451	
BR-04920/2022	0,5638% ^{AC}	51,52%	50,96%	52,08%	Fora	8,41	0,0000	***
BR-07309/2022	1,9967% ^{AB}	53,03%	51,03%	55,03%	Fora	3,85	0,0001	***
BR-04208/2022	1,4441% ^{AC}	47,31%	45,87%	48,75%	Fora	-2,43	0,0153	**
BR-07903/2022	2,1913% ^{AC}	50,32%	48,13%	52,51%	Dentro	1,09	0,2751	
BR-05418/2022	2,1928% ^{AB}	52,22%	50,03%	54,41%	Fora	2,86	0,0043	***
BR-01560/2022	1,1292% ^{AB}	46,75%	45,62%	47,88%	Fora	-4,07	0,0000	***
BR-08584/2022	1,9951% ^{AB}	53,63%	51,63%	55,63%	Fora	4,44	0,0000	***
BR-00470/2022	2,1864% ^{AB}	46,67%	44,48%	48,86%	Fora	-2,17	0,0297	**
BR-01159/2022	1,3839% ^{AC}	47,31%	45,93%	48,69%	Fora	-2,53	0,0113	**
BR-08044/2022	3,0090% ^{AB}	46,81%	43,80%	49,82%	Dentro	-1,52	0,1287	
BR-00525/2022	2,1804% ^{AC}	49,78%	47,60%	51,96%	Dentro	0,61	0,5410	
BR-06043/2022	1,7818% ^{AC}	46,24%	44,46%	48,02%	Fora	-3,14	0,0017	***
BR-08487/2022	1,9960% ^{AB}	53,30%	51,30%	55,30%	Fora	4,12	0,0000	***
BR-06415/2022	1,4583% ^{AB}	47,05%	45,59%	48,51%	Fora	-2,75	0,0059	***
BR-04043/2022	0,6889% ^{AC}	51,40%	50,71%	52,09%	Fora	6,54	0,0000	***
BR-08523/2022	2,1912% ^{AC}	50,54%	48,35%	52,73%	Dentro	1,29	0,1977	
BR-00053/2022	2,5281% ^{AC}	47,92%	45,39%	50,45%	Dentro	-0,91	0,3606	
BR-05389/2022	1,9974% ^{AB}	52,72%	50,72%	54,72%	Fora	3,55	0,0004	***
BR-02276/2022	2,1797% ^{AC}	48,69%	46,51%	50,87%	Dentro	-0,37	0,7124	
BR-07340/2022	1,8144% ^{AC}	47,87%	46,06%	49,68%	Dentro	-1,33	0,1843	
BR-08917/2022	1,0663% ^{AC}	47,83%	46,76%	48,90%	Fora	-2,33	0,0197	**
BR-04387/2022	2,1878% ^{AB}	47,19%	45,00%	49,38%	Dentro	-1,71	0,0875	*
BR-06307/2022	3,0087% ^{AB}	46,74%	43,73%	49,75%	Dentro	-1,57	0,1174	
BR-02707/2022	1,7818% ^{AC}	46,24%	44,46%	48,02%	Fora	-3,14	0,0017	***
BR-05514/2022	2,1848% ^{AC}	46,50%	44,32%	48,68%	Fora	-2,33	0,0200	**
BR-04850/2022	1,3177% ^{AC}	51,21%	49,89%	52,53%	Fora	3,14	0,0017	***
BR-01682/2022	1,8178% ^{AC}	47,31%	45,49%	49,13%	Dentro	-1,93	0,0539	*
BR-06012/2022	1,4592% ^{AB}	47,64%	46,18%	49,10%	Fora	-1,96	0,0501	*
BR-06280/2022	2,1913% ^{AC}	49,79%	47,60%	51,98%	Dentro	0,62	0,5371	
BR-07942/2022	2,2313% ^{AB}	46,74%	44,51%	48,97%	Fora	-2,11	0,0348	**
BR-07106/2022	2,1826% ^{AB}	45,56%	43,38%	47,74%	Fora	-3,17	0,0015	***
BR-08438/2022	2,1788% ^{AC}	48,09%	45,91%	50,27%	Dentro	-0,91	0,3639	
BR-09241/2022	1,3846% ^{AC}	47,83%	46,45%	49,21%	Dentro	-1,80	0,0724	*
BR-02322/2022	2,1945% ^{AB}	48,94%	46,75%	51,13%	Dentro	-0,15	0,8835	
BR-01120/2022	3,0066% ^{AB}	46,24%	43,23%	49,25%	Dentro	-1,90	0,0578	*
BR-02853/2022	2,1810% ^{AB}	45,16%	42,98%	47,34%	Fora	-3,52	0,0004	***
BR-02012/2022	1,8222% ^{AC}	47,31%	45,49%	49,13%	Dentro	-1,92	0,0545	*
BR-07940/2022	2,1845% ^{AB}	46,07%	43,89%	48,25%	Fora	-2,71	0,0067	***
BR-08253/2022	1,6549% ^{AC}	47,83%	46,18%	49,48%	Dentro	-1,50	0,1317	
BR-02736/2022	2,1833% ^{AB}	45,74%	43,56%	47,93%	Fora	-3,01	0,0026	***
BR-08263/2022	2,1579% ^{AC}	41,30%	39,14%	43,46%	Fora	-6,98	0,0000	***

Nota: Construção de IC = $\hat{p} \pm E_e$, em que \hat{p} é a estimativa pontual da proporção considerando votos válidos e E_e , a margem de erro efetiva. Parâmetro populacional para Bolsonaro é de 49,10%, conforme resultado do 2º turno das Eleições Gerais 2022. Significância (sig.) aos níveis de 1% (***), 5% (**) e 10% (*) a partir do cálculo da estatística Z e p-valor do teste bilateral com a hipótese nula (H_0) de que o parâmetro populacional é igual a 49,10%. Sobrescritos AC/AB representam dimensionamento efetivo n_e acima/abaixo do dimensionamento previsto n_p .

Fonte: Elaborado pelo autor.

5. CONSIDERAÇÕES FINAIS

Ao contrário do pensamento convencional, estudantes que não conhecem Estatística não podem ser considerados instruídos em Matemática. Conforme orientações da BNCC, propõe-se uma visão integrada à realidade, prezando pelo raciocínio, comunicação não somente por símbolos e conectivos lógicos, mas pela linguagem corrente. Assim, diferentes habilidades são organizadas por competências, acrescentando-se sugestões sobre organização curricular. Entretanto, as características implicam em conteúdos que não possuem estrutura sistemática de desenvolvimento ao longo dos anos, gerando-se lacuna para que sejam contemplados de forma superficial, insuficiente e circular. Assim, as mesmas habilidades de Probabilidade e Estatística Descritiva compõem todo o Ensino Médio, enquanto saberes como margem de erro e intervalo de confiança, essenciais para compreensão das pesquisas político-eleitorais, não são abordados.

Preenchendo-se essa lacuna no Ensino da Matemática, aplicável em nível médio e mesmo superior, investiga-se em que medida as pesquisas para o 2º Turno das Eleições Gerais de 2022 geram estimativas condizentes com o dimensionamento amostral e margem de erro divulgados pelos institutos responsáveis, e como as estimativas geradas influenciam a precisão e interpretação de intervalos de confiança. Deste modo, atribui-se ênfase à investigação, explicação e justificativa dos problemas resolvidos a partir de processos de argumentação, em contextos reais, conforme preconizado pela BNCC, e estimula-se o pensamento computacional por meio de planilhas eletrônicas e software R, aberto e utilizado de forma cada vez mais ampla. Busca-se também preencher lacuna em pesquisa aplicada sobre falhas estatísticas dos institutos, a partir de discrepâncias entre dimensionamento amostral, margem de erro e nível de confiança.

Dentre os resultados obtidos, investimentos monetários por pesquisa não acompanham necessariamente maiores tamanhos amostrais e constata-se influência de contratantes sobre valores despendidos. Adicionalmente, verifica-se amplitude relevante em tamanhos amostrais sem a esperada alteração das premissas estatísticas divulgadas, assim como discrepâncias entre dimensionamentos amostrais previstos analiticamente para sustentar a margem de erro e nível de confiança divulgados ao público e aqueles efetivamente adotados, repercutindo na precisão dos intervalos de confiança. No caso de pesquisas superdimensionadas, estas tendem a incluir o parâmetro sob margem divulgada e a serem menos penalizadas mesmo sob margem efetivas mais estreitas, com resultados corroborados por testes de hipóteses. Assim, avaliar criticamente a construção de intervalos de confiança é essencial como instrumento de exercício da cidadania.

REFERÊNCIAS

- BRASIL. **Lei nº 9.504, de 30 de setembro de 1997**. Estabelece normas para as eleições. Diário Oficial da União, Brasília, DF, 1º de outubro de 1997. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/19504.htm. Acesso em: 05 mai. 2023.
- BRASIL. **Ministério da Educação**. Base Nacional Comum Curricular. Brasília, 2018. Disponível em: <http://basenacionalcomum.mec.gov.br/a-base>. Acesso em: 05 mai. 2023.
- BRASIL. Tribunal Superior Eleitoral. **Resolução nº 23.600, de 12 de dezembro de 2019**. Dispõe sobre as pesquisas eleitorais. Brasília, DF: TSE, 2019. Disponível em: <https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-600-de-12-de-dezembro-de-2019>. Acesso em: 05 mai. 2023.
- BRITO, K.; ADEODATO, P. J. L. Machine learning for predicting elections in Latin America based on social media engagement and polls. **Government Information Quarterly**. Volume 40, Issue 1, p. 101-782, jan. 2023. DOI: <https://doi.org/10.1016/j.giq.2022.101782>
- CABALLÉ, A.; GRIMA, P.; MARCO-ALMAGRO, L. ¿Aciertan los sondeos electorales? Análisis sobre la bondad de predicción de los sondeos electorales publicados en la prensa / Are Election Polls Right? Analysis of the Accuracy of Election Polls Predictions Published in the Press. **Revista Española de Investigaciones Sociológicas**, 143, p. 25–46, jul./set. 2013. DOI: <https://doi.org/10.5477/cis/reis.143.25>
- DUGAS, L. S. A problemática das pesquisas político-eleitorais: o currículo de matemática para a compreensão social. **Caderno de Pesquisa**, São Paulo, n. 76, p. 18-23, fev. 1991. Disponível em: http://educa.fcc.org.br/scielo.php?script=sci_arttext&pid=S0100-15741991000100002&lng=pt&nrm=iso. Acesso em: 05 mai. 2023.
- FERREIRA, M. N. Como contestar pesquisa eleitoral? **Revista de Produção Científica da PUC-GO**, v. I, p. 2484-2501, 2012. ISSN 2176-0705.
- FERREIRA, M. N. Erros e falhas de pesquisas eleitorais e suas influências na decisão do voto do eleitor. **Revista de Produção Científica da PUC-GO**, v. I, p. 3166-3113, 2013. ISSN 2176-0705.
- GRAEFE, A. Embrace the differences: revisiting the PollyVote method of combining forecasts for U.S. presidential elections (2004 to 2020). **International Journal of Forecasting**, Volume 39, Issue 1, p. 170-177, jan./mar. 2023. DOI: <https://doi.org/10.1016/j.ijforecast.2021.09.010>
- GRAMACHO, W. G. A margem das margens? precisão das pesquisas pré-eleitorais brasileiras em 2010. **Opinião Pública**, 19(1), p. 65–80, out. 2015(a). Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/op/article/view/8641440>
- GRAMACHO, W. G. Surveys pré-eleitorais nas eleições brasileiras de 2014: erros, acertos e polêmicas. **Revista de Estudos Brasileiros**, v. 2, n. 2, p. 115-31, fev. 2015(b). Disponível em: <https://bibliotecadigital.tse.jus.br/xmlui/handle/bdtse/4376>
- LLOYD, R.; TURGEON, M. Polling in new democracies and electoral malpractice: the case of Brazil. **International Journal of Public Opinion Research**, Volume 33, Issue 4, mai. 2021. DOI: <https://doi.org/10.1093/ijpor/edab005>

MARTIN, E.; TRAUGOTT, M.; KENNEDY, C. A Review and Proposal for a New Measure of Poll Accuracy. **Public Opinion Quarterly**, Volume 69, Issue 3, p. 342-369, jan. 2005. DOI: <https://doi.org/10.1093/poq/nfi044>

MEIRELES, F.; RUSSO, G. Pesquisas eleitorais no Brasil: tendências e desempenho. **Estudos Avançados**, 36, 106, p. 117-131, set./out. 2022. DOI: <https://doi.org/10.1590/s0103-4014.2022.36106.007>

MOSTELLER, F. Measuring the error. In: MOSTELLER, F.; HYMAN, H.; MCCARTHY, P. et al. (Orgs.). **The pre-election polls of 1948**. Report to the committee on analysis and pre-election polls and forecast. New York: Social Science Research Council. p. 54-80, 1949.

PANAGOPOULOS, C. Polls and Elections Accuracy and Bias in the 2020 U.S. General Election Polls. **Presidential Studies Quarterly**, 51, p. 214-227, mar. 2021. DOI: <https://doi.org/10.1111/psq.12710>

SOHLBERG, J.; BRANHAM, J. A. Just a Difficult Election to Poll? How Context Affects Polling Accuracy. **Survey Methods: Insights from the Field**, fev. 2020. DOI: <https://doi.org/10.13094/SMIF-2020-00013>

ZHOU, Z.; SERAFINO, M.; COHAN, L. et al. Why polls fail to predict elections. **Journal of Big Data**, 8, 137, out. 2021. DOI: <https://doi.org/10.1186/s40537-021-00525-8>

APÊNDICE A – Dimensionamento Amostral (n_p) em Planilha Eletrônica e Software R

A.1) Planilha Eletrônica:

	A	B	C	D	E	F	G
1	Determinação do Dimensionamento Amostral (n_p)						
2							
3	Dados						
4	estimativa planejada: p^*	0,5					
5	nível de confiança: c	95,00%					
6	margem de erro: E_d	2,00%					
7							
8	Valores Preliminares						
9	p^*q^*	0,25	<-- =B4*(1-B4)				
10	Z_c	1,9600	<-- =ABS(INV.NORMP.N((1-B5)/2))				
11							
12	Resultado						
13	n_p	2401,00	<-- =ARREDONDAR.PARA.CIMA(B9*(B10/B6)^2;0)				

Nota: Estrutura desenvolvida em planilha eletrônica para cálculo de $n_p = p^*q^*(z_c/E_d)^2$. Os dados a serem inseridos pelo usuário são relativos à estimativa planejada p^* , nível de confiança c e margem de erro E_d . Os valores preliminares resultantes são o produto da proporção planejada e seu inverso, p^*q^* , e o valor crítico que corresponde ao nível de confiança estabelecido, Z_c , de forma a se atingir o resultado do dimensionamento amostral previsto n_p .
Fonte: Elaborado pelo autor sob licença CC BY-SA.

A.2) Rotina em Software R:

```
# Rotina em R para cálculo do dimensionamento amostral previsto (Np)
dimensionamento_amostal <- function(proporcao_preliminar, margem_erro,
nivel_confianca) {
  # Calcula o tamanho da amostra necessário
  dimensionamento_amostal <- (qnorm(1 - (1 - nivel_confianca) / 2) /
margem_erro)^2 * proporcao_preliminar * (1 - proporcao_preliminar)
  # Arredonda o tamanho da amostra para cima
  dimensionamento_amostal <- ceiling(dimensionamento_amostal)
}
# Aplicação da rotina. Digite as premissas em decimal.
proporcao_preliminar <- 0.5 # Proporção preliminar
margem_erro <- 0.02 # Margem de erro desejada
nivel_confianca <- 0.95 # Nível de confiança estabelecido
Np <- dimensionamento_amostal(proporcao_preliminar, margem_erro,
nivel_confianca)
print(paste("O tamanho amostral necessário é:", Np))
```

Nota: Procedimento em R desenvolvido para cálculo do dimensionamento amostral previsto $n_p = p^*q^*(z_c/E_d)^2$. As premissas são a estimativa planejada da proporção (p^*), margem de erro divulgada (E_d) e nível de confiança c .
Fonte: Elaborado pelo autor sob licença CC BY-SA.

APÊNDICE B – Margem de Erro Prevista (E_p) em Planilha Eletrônica e Software R

B.1) Planilha Eletrônica:

	A	B	C	D	E	F	G
1	Determinação da Margem de Erro Prevista (E_p)						
2							
3	Dados						
4	estimativa planejada: p^*	0,5					
5	nível de confiança: c	95,00%					
6	tamanho amostral: n_e	17552					
7							
8	Valores Preliminares						
9	p^*q^*	0,25	<-- =B4*(1-B4)				
10	Z_c	1,9600	<-- =ABS(INV.NORMP.N((1-B5)/2))				
11							
12	Resultado						
13	E_p	0,7397%	<-- =B10*RAIZ(B9/B6)				

Nota: Estrutura desenvolvida em planilha eletrônica para cálculo de $E_p = z_c \sqrt{p^*q^*/n_e}$. Os dados a serem inseridos pelo usuário são relativos à estimativa planejada p^* , nível de confiança c e tamanho amostral n_e . Os valores preliminares resultantes são o produto da proporção planejada e seu inverso, p^*q^* , e o valor crítico que corresponde ao nível de confiança estabelecido, Z_c , de forma a se atingir o resultado em relação à margem de erro prevista E_p .
Fonte: Elaborado pelo autor sob licença CC BY-SA.

B.2) Rotina em Software R:

Rotina em R para cálculo da margem de erro prevista (E_p)

```
margem_erro <- function(proporcao_preliminar, nivel_confianca,
tamanho_amostral) {
  # Calcula o valor crítico com base no nível de confiança
  valor_critico <- qnorm(1 - (1 - nivel_confianca) / 2)

  # Calcula a margem de erro prevista
  margem_erro <- (valor_critico * sqrt((proporcao_preliminar * (1 -
proporcao_preliminar)) / tamanho_amostral))
}
```

```
# Aplicação da rotina. Digite as premissas em decimal.
proporcao_preliminar <- 0.5 # Proporção preliminar
nivel_confianca <- 0.95 # Nível de confiança estabelecido
tamanho_amostral <- 17552 # Tamanho amostral
```

```
Ep <- margem_erro(proporcao_preliminar, nivel_confianca, tamanho_amostral)
print(paste("A margem de erro é:", Ep*100, "%"))
```

Nota: Procedimento em R desenvolvido para cálculo da margem de erro prevista $E_p = z_c \sqrt{p^*q^*/n_e}$. As premissas são a estimativa preliminar da proporção (p^*), tamanho amostral efetivo (n_e) e nível de confiança estabelecido c .
Fonte: Elaborado pelo autor sob licença CC BY-SA.

APÊNDICE C – Intervalo de Confiança em Planilha Eletrônica e Software R

C.1) Planilha Eletrônica:

	A	B	C	D	E	F	G
1	Intervalo de Confiança para Proporção (IC = $\hat{p} \pm E_e$)						
2							
3	Dados						
4	tamanho amostral: n_e	17552					
5	proporção amostral: \hat{p}	48,99%					
6	nível de confiança: c	95,00%					
7							
8	Valores Preliminares						
9	Z_c	1,9600	<-- =ABS(INV.NORMP.N((1-B6)/2))				
10	E_e	0,7395%	<-- =B9*RAIZ(B5*(1-B5)/B4)				
11							
12	Resultado						
13	Limite inferior	48,25%	<-- =B5-B10				
14	Limite superior	49,73%	<-- =B5+B10				

Nota: Estrutura em planilha eletrônica para cálculo de IC = $\hat{p} \pm E_e$. As premissas são o tamanho amostral n_e , a proporção amostral \hat{p} , e nível de confiança estabelecido c para obtenção dos limites intervalares inferior e superior. Fonte: Elaborado pelo autor sob licença CC BY-SA.

C.2) Rotina em Software R:

```
# Rotina em R para cálculo de intervalo de confiança para proporção
IC <- function(tamanho_amostral, proporcao_amostral, nivel_confianca) {
  # Calcula a margem de erro efetiva
  margem_erro <- qnorm(1 - (1 - nivel_confianca) / 2) *
sqrt((proporcao_amostral * (1 - proporcao_amostral)) / tamanho_amostral)

  # Calcula o limite inferior e superior do intervalo de confiança
  limite_inferior <- proporcao_amostral - margem_erro
  limite_superior <- proporcao_amostral + margem_erro

  # Formata a margem de erro e o intervalo de confiança
  margem_erro_formatada <- paste(round(margem_erro * 100, 4), "%")
  intervalo_confianca_formatado <- paste("[", round(limite_inferior * 100,
4), "%; ", round(limite_superior * 100, 4), "%]")

  resultado <- list(margem_erro = margem_erro_formatada,
intervalo_confianca = intervalo_confianca_formatado)
  return(resultado)
}

# Aplicação da rotina. Digite as premissas percentuais em decimal.
tamanho_amostral <- 17552 # Tamanho amostral
proporcao_amostral <- 0.4899 # Proporção amostral
nivel_confianca <- 0.95 # Nível de confiança estabelecido

resultado <- IC(tamanho_amostral, proporcao_amostral, nivel_confianca)

print(paste("A margem de erro é:", resultado$margem_erro))
print(paste("O intervalo de confiança para a proporção é:",
resultado$intervalo_confianca))
```

Nota: Procedimento em R para obtenção de IC = $\hat{p} \pm E_e$, em que o tamanho amostral n_e , proporção amostral \hat{p} e nível de confiança c são premissas para geração da margem de erro efetiva E_e e dos limites intervalares resultantes. Fonte: Elaborado pelo autor sob licença CC BY-SA.

APÊNDICE D – Teste de Hipótese em Planilha Eletrônica e Software R

D.1) Planilha Eletrônica:

	A	B	C	D	E	F	G
1	Teste de Hipótese para Proporção (Ho: p = p₀)						
2							
3	Dados						
4	tamanho amostral: n _e	17552					
5	proporção amostral: \hat{p}	48,99%					
6	valor hipotetizado: p ₀	50,90%					
7	nível de significância: α	5,00%					
8							
9	Valor Preliminar						
10	Estatística de teste Z	-5,062	<-- =(B5-B6)/RAIZ(B6*(1-B6)/B4)				
11							
12	Resultado						
13	p-valor	0,0000	<-- =2*(1-DIST.NORMP.N(ABS(B10);VERDADEIRO))				
14	Conclusão	Rejeitar Ho	<-- =SE(B13<=B7; "Rejeitar Ho"; "Não rejeitar Ho")				

Nota: Estrutura desenvolvida em planilha eletrônica para $H_0: p = p_0$ e decisão com base na estatística Z e p-valor. As premissas são o tamanho amostral n_e , proporção amostral \hat{p} , valor hipotetizado p_0 e nível de significância α .
Fonte: Elaborado pelo autor sob licença CC BY-SA.

D.2) Rotina em Software R:

```
# Rotina em R para cálculo de teste de hipótese para proporção
hipotese_proporcao <- function(tam_am, prop_am, valor_hip, nivel_sig) {
  # Calcula a estatística de teste (Z-score)
  Z <- (prop_am - valor_hip) / sqrt((valor_hip * (1 - valor_hip)) / tam_am)

  # Calcula o p-valor formatado
  valor_p <- 2 * (1 - pnorm(abs(Z)))
  valor_p_form <- sprintf("%.4f", valor_p)

  # Define conclusão com base no p-valor e nível de significância
  if (valor_p <= nivel_sig) {
    conclusao <- "Rejeita-se a hipótese nula"
  } else {
    conclusao <- "Falha-se em rejeitar a hipótese nula"
  }
  resultado <- list(Z = Z, valor_p = valor_p_form, conclusao = conclusao)
  return(resultado)
}

# Aplicação da rotina. Digite as premissas percentuais em decimal.
tam_am <- 17552 # Tamanho amostral
prop_am <- 0.4899 # Proporção amostral
valor_hip <- 0.5090 # Valor hipotetizado
nivel_sig <- 0.05 # Nível de significância

resultado <- hipotese_proporcao(tam_am, prop_am, valor_hip, nivel_sig)
print(paste("Estatística de teste (Z-score):", resultado$Z))
print(paste("Valor p:", resultado$valor_p))
print(paste("Conclusão:", resultado$conclusao))
```

Nota: Procedimento desenvolvido em R para $H_0: p = p_0$ e decisão baseado na estatística Z e p-valor respectivo. As premissas são o tamanho amostral n_e , proporção amostral \hat{p} , valor hipotetizado p_0 e nível de significância α .
Fonte: Elaborado pelo autor sob licença CC BY-SA.