PESQUISAS ELEITORAIS: UMA DISCUSSÃO DE CENÁRIOS AMOSTRAIS CONSTRUÍDOS A PARTIR DAS DISTRIBUIÇÕES DICOTÔMICAS E POLITÔMICAS

ELECTORAL RESEARCH: A DISCUSSION OF SAMPLE SCENARIOS BUILT FROM THE DICHOTOMIC AND POLITOMIC DISTRIBUTIONS

ENCUESTAS ELECTORALES: UNA DISCUSIÓN DE ESCENARIOS DE MUESTRAS CONSTRUIDOS DE DISTRIBUCIONES DICTÓMICAS Y POLITÓMICAS

Julio Cesar Guimarães de Paula[1]

Resumo: O presente trabalho tem por objetivo discutir o desenho amostral utilizado nas pesquisas eleitorais no Brasil. Do ponto de vista da teoria estatística, todo desenho amostral se constrói a partir de uma distribuição de probabilidade. Assim, a estimação de variáveis categóricas se relaciona a dois tipos de distribuição de probabilidade: a dicotômica e a politômica. A primeira ambienta-se em duas categorias *(distribuição binomial)*, já a segunda, em três ou mais. O trabalho mostra que o uso da distribuição binomial pelos institutos diminui a magnitude amostral com impactos nos erros e nos intervalos de confiança. Assim, se propõe o uso da distribuição multinomial indexada as correções de Bonferroni pela elevar as qualidades das estimativas eleitorais no Brasil.

Palavras-chave: Pesquisas eleitorais; Desenho amostral; Distribuição de probabilidade; Correção de Bonferroni.

Abstrat: This paper aims to discuss the sample design used in electoral research in Brazil. From the point of view of statistical theory, every sample design is constructed from a probability distribution. Thus, the estimation of categorical variables is related to two types of probability distribution: dichotomous and polytomous. The first is used in two categories (Binomial Distribution), while the second, in three or more. The work shows that the use of Binomial Distribution by the institutes reduces the sample magnitude with impacts on errors and confidence intervals. Thus, it is proposed to use the Multinomial Distribution indexed to Bonferroni's corrections for raising the quality of electoral estimates in Brazil.

Keywords: Electoral Research; Sample research design; Probability distribution; Bonferroni Correction.

Resumen: Este documento tiene como objetivo discutir el diseño de la muestra utilizada en las encuestas electorales en Brasil. Desde el punto de vista de la teoría estadística, cada diseño de muestra se construye a partir de una distribución de probabilidad. Por lo tanto, la estimación de variables categóricas se relaciona con dos tipos de distribución de probabilidad: dicotómica y politómica. El primero se usa en dos categorías (distribución binomial), mientras que el segundo, en tres o más. El trabajo muestra que el uso de la distribución binomial por parte de los institutos reduce la magnitud de la muestra con impactos en los errores y los intervalos de confianza. Por lo tanto, se propone utilizar la distribución multinomial indexada a las correcciones de Bonferroni para elevar la calidad de las estimaciones electorales en Brasil.

Palabras clave: Investigación electoral; Diseño de muestras de encustas electorales; Distribución de probabilidade; Corrección de Bonferroni.

[1] Doutorando em Ciência Política de Universidade Federal de Minas Gerais (UFMG) e pesquisador do Centro de Estudos Legislativos da mesma instituição (CEL-DCP-UFMG). ORCID: https://orcid.org/0000-0003-1243-225X. E-mail: juliogcpaula@gmail.com

Introdução

Pesquisas eleitorais podem ser definidas como, pesquisas quantitativas nas quais normalmente são aplicados questionários estruturados a uma amostra pretensamente representativa do eleitorado. Na atualidade se configuram como a melhor informação que eleitores e candidatos têm sobre os rumos do processo eleitoral. Assim, nas democracias contemporâneas as pesquisas eleitorais, além funcionarem como mecanismos de vocalização de preferências, garantem a presença das distintas vontades no espaço público e possibilitam a relação de congruência entre o estado e os cidadãos.

Nesse sentido, no Brasil atual percebemos o destaque midiático dado aos institutos de pesquisa. Esses protagonizam discussões públicas sobre a assertividade de seus levantamentos. Nesse cenário de dúvidas sobre a qualidade dos resultados publicizados pelos institutos, uma questão se coloca: as pesquisas devem antecipar os resultados eleitorais? Ou devem ser entendidas como um retrato de um momento específico de um processo mais amplo? Essas problematizações levam em conta aspectos importantes que se relacionam a consolidação de preferencias eleitorais em democracias de massa, como é o caso da brasileira. Preferencias pouco consolidadas, alta volatilidade eleitoral, pouca estruturação do sistema partidário são aspectos que tornam a projeção de resultados eleitorais tarefa extremamente difícil. Esse fato não exime os institutos de pesquisa de uma análise mais detida sobre os fundamentos metodológicos utilizados em seus levantamentos. Assim, é nesse horizonte que se enquadra o presente trabalho.

Sob esse prisma, o estudo em tela se propõe a uma análise dos erros amostrais a partir das distribuições dicotômicas e politômicas. Para a consecução de tal objetivo o trabalho além desta introdução está dividido da seguinte maneira. A primeira seção apresenta um levantamento histórico sobre as dissonâncias entre pesquisas e resultados eleitorais e discute a consolidação das técnicas de amostragem a partir da experiência norte-americana. Na segunda apresento alguns fundamentos das pesquisas eleitorais no Brasil.

A seção seguinte se atem aos fundamentos da avaliação das pesquisas eleitorais no Brasil. Na terceira seção se discute as distribuições dicotômica e politômica de probabilidade e os desenhos amostrais construídos a partir dessas distribuições. Na seção seguinte são calculadas as correções de Bonferroni para cenários com mais de duas categorias. Por fim, são apresentadas as conclusões.

1. Apontamentos sobre amostragem: as dissonâncias entre pesquisas e resultados

A revista *Literary Digest* foi fundada em 1890, desde a sua fundação conduziu a realização de enquetes eleitorais que visavam prever os resultados dos pleitos presidenciais norte-americanos. O periódico apontou corretamente a vitória de *Woodrow Wilson* nas eleições de 1916 e estimou corretamente os resultados nos quatro pleitos posteriores. Já nas eleições de 1936 a metodologia adotada pela revista não alcançou os resultados esperados. O levantamento apontou

equivocadamente a vitória do candidato republicano Alfred Landom, sobre o candidato a reeleição, o democrata Franklin D. Roosevelt (FERRAZ, 1996). O desenho proposto pelos pesquisadores se baseava no envio de cartões de resposta aos eleitores. Esses foram amostrados com base nos registros telefônicos e de proprietários de automóveis. Segundo dados da época foram enviados cerca de 10 milhões de cartões, a taxa de resposta chegou a aproximadamente 25% do total da amostra. A diferença entre os dados e o resultado eleitoral chegou a 19 pontos percentuais. O erro nas estimativas foi atribuído a inconsistência do método de amostragem utilizado. O universo dos donos de automóveis não era representativo do universo de eleitores. Muitos eleitores atingidos pela crise de 1929 não tinham automóveis e consequentemente não poderiam ser amostrados (ABEP, 2008).

Nesse sentido, a estimação equivocada realizada para o pleito de 1936 motivou um amplo debate na sociedade norte-americana sobre os fundamentos metodológicos das pesquisas eleitorais. Os estudiosos da época mostraram que o viés de uma amostra pouco representativa e a baixa participação proporcional da população amostrada, foram os motivos pelos quais, o resultado apontado pela revista se destoou do resultado aferido pelas urnas (FERRAZ, 1996; ABEP, 2008). Na mesma eleição George Gallup e seu instituto de pesquisa o *American Institute of Public Opinion* acertaram o resultado com uma amostra demograficamente representativa de 3.000 entrevistas. O survey conduzido por Gallup foi o marco inicial no uso de métodos científicos na elaboração e realização de pesquisas eleitorais (ABEP, 2008).

Nas eleições de 1948 a cientificidade dos procedimentos de amostragem também foi questionada, os mais importantes institutos de pesquisa norte-americanos não apontaram a vitória do democrata Harry Truman para a presidência da república. As explicações pelo erro se fundamentaram em duas dimensões. A primeira se relaciona a volatilidade das preferências dos eleitores. A segunda, a distância entre a realização das pesquisas e das eleições, já que os levantamentos foram realizados duas semanas antes da eleição, fato que não captou as alterações nas preferências no período final do processo eleitoral (FERRAZ, 1996; ABEP, 2008). A outra dimensão dos erros está ligada a uma discussão dos fundamentos da amostragem. O uso de amostras probabilísticas e o uso de amostragem por cotas era o cerne do debate. O relatório do *Social Science Research Council* fez uma defesa veemente da amostragem probabilísticas (FERRAZ, 1996). A partir daí a maneira como as amostras eram construídas se tornou o núcleo das discussões sobre a robustez inferencial nas pesquisas eleitorais. É importante destacar que a inferência é o processo pelo qual se generalizam informações obtidas de uma amostra da população de interesse. Assim, nesse processo de generalização a qualidade das inferências podem ser afetadas por dois fatores, erros não amostrais e erros amostrais (ABEP, 2008).

Os erros não amostrais estão ligados a logística da pesquisa, definição inadequada da população de interesse, questionários mal elaborados (perguntas que induzem a determinadas respostas, falta de objetividade, ordem inadequada, vocabulário inacessível etc.) e entrevistadores

mal treinados. Já os erros amostrais, são ligados a construção do plano amostral, tamanho, homogeneidade e estratificação. Nesse sentido, podemos dizer que amostras representativas são aquelas em que a proporção de ocorrência dos dois tipos de erros é minimizada ou quantificada. No caso dos erros amostrais pesquisas realizadas através de métodos probabilísticos devem sempre incluir em suas estimativas um erro, a chamada margem de erro nas pesquisas eleitorais (FERRAZ, 1996; ABEP, 2008). A partir da compreensão da importância das amostras probabilistas e da estimação dos erros. A próxima seção se dedicará ao exame desses pressupostos nas pesquisas eleitorais brasileiras.

2. Os fundamentos das pesquisas eleitorais no Brasil

No caso brasileiro o debate sobre os fundamentos das pesquisas eleitorais e sua capacidade em prever os resultados das urnas é recente. A criação do Instituto Brasileiro de Opinião Pública, o IBOPE, em 1942 é o marco fundador das pesquisas e consequentemente o início do debate sobre a qualidade das inferências para os contextos eleitorais (*Idem*). Os anos de interrupção democrática arrefeceram o debate, que só foi retomado após a estabilização do calendário eleitoral, após as eleições de 1982 para os governos estaduais. Assim, as discussões sobre o quão as pesquisas refletem os resultados eleitorais também passaram a nortear os debates sobre eleições no Brasil.

Discussões sobre possíveis erros e inconsistências amostrais também pautaram o debate brasileiro sobre a validade das inferências aferidas pelas pesquisas. Em geral nas pesquisas de intenção de voto realizadas no Brasil adota-se, o procedimento de múltiplos estágios[2], com estratificação[3] e conglomerados[4] nos primeiros estágios (regiões, municípios, setores censitários) e em seguida, no último estágio, incorporam-se cotas[5] por sexo, idade, grau de instrução etc., definidas segundo critérios do IBGE e do Tribunal Superior Eleitoral. As cotas são utilizadas para que o entrevistador aplique entrevistas a pessoas que têm uma probabilidade de resposta baixa, evitando-se, com isto, possíveis vieses na amostra. Por exemplo, são definidas cotas de PEA

[2] Um procedimento de amostragem pode ser realizado em várias etapas e neste caso tem-se uma amostragem em múltiplos estágios. O objetivo é combinar os diversos tipos de amostragens utilizando as vantagens de cada tipo. Numa amostragem em três estágios, por exemplo, pode-se nos dois primeiros estágios empregar técnicas de aleatorização e no último amostragem por cotas.

[3] A amostragem estratificada é o desenho adequado quando o pesquisador pretende estudar uma população a partir de características específicas. O objetivo é definir grupos representativos, combinando amostragem aleatória e estratificação. A amostra estratificada foi desenvolvida como uma forma de aumentar a precisão do processo amostral, reduzindo o grau de heterogeneidade presente na amostra aleatória simples. A estratificação amostral tem como característica a menor variação dos dados dentro de cada estrato do que entre os estratos.

[4] A falta de listas disponíveis para grandes populações é uma dificuldade para a utilização da amostragem aleatória simples. A amostragem por conglomerados é uma maneira de corrigir esse problema. O pesquisador constrói múltiplos estágios de seleção onde os estágios iniciais são os chamados conglomerados e o mesmo princípio de aleatoriedade se aplica a cada um deles. Um conglomerado é uma unidade que aglomera pessoas. Ao se sortear aleatoriamente um conglomerado, a lógica é a mesma do sorteio aleatório de indivíduos.

[5] O Anexo (1) mostra os desenhos amostrais utilizados pelos principais institutos de pesquisa nas últimas eleições presidenciais.

(População Economicamente Ativa) e Não-PEA para impor que pessoas que trabalham e que não trabalham pertençam à amostra (ABEP, 2008).

Segundo King et. al. (2001) em pesquisas eleitorais as taxas de não-respostas variam entre 50% até 90%. Essas altas taxas segundo os autores estão relacionadas a dois fatores: (1) os temas abordados no questionário, tais como: racismo, desigualdade, adesão a democracia, assuntos que dependendo da maneira como são apresentados, podem aumentar a taxa de não-respostas e (2) as localidades onde as pesquisas são realizadas. Sob esse prisma, Smith (1983) indica a comparação das taxas de não-respostas com dados oficiais, como os obtidos pelos censos demográficos, por exemplo. Na esteira dos estudos que tomam as variáveis socioeconômicas com variáveis independentes das taxas de não-resposta, Henkel (2012) encontra resultados importantes, quando avalia as taxas de não-resposta de estudantes das escolas estaduais do Pará, à uma pesquisa que procurava avaliar o posicionamento dos estudantes sobre aspectos do sistema político brasileiro. Segundo o autor, a estrutura sociodemográfica, o ciclo de vida dos respondentes e suas experiências com relação às políticas públicas, ajudam a explicar as taxas de não-resposta.

No limite, a amostragem por cotas (MOSER, et. al., 1953) mitiga a probabilidade de não-respostas pelo fato de trazer para o cenário indivíduos que teriam dificuldade de serem pesquisados. O método utilizado – de estabelecer cotas dentro dos setores selecionados probabilisticamente - pode ser considerado como aproximadamente alto-ponderado e sua função é evitar as possíveis distorções que seriam introduzidas pelos entrevistadores, caso não houvesse cotas. Embora seja amplamente conhecido o fato de que amostragem por cotas não possibilita o cálculo do erro amostral (ou margem de erro), pois ela não atende aos princípios da aleatoriedade estatística, os institutos de pesquisa adotam conscientemente o modelo de amostragem em múltiplos estágios, envolvendo cotas no último estágio, como um modelo aproximado em relação a um modelo ideal (do ponto de vista probabilístico).

Problemas como o tempo de realização e o custo de execução da pesquisa, bem como as imposições da Lei Eleitoral[6], que obriga a declaração da margem de erro no momento do registro da pesquisa junto à Justiça Eleitoral, são os principais argumentos utilizados pelos institutos pela escolha do método aplicado. Assim, a divulgação das pesquisas está legalmente condicionada ao registro das seguintes informações no Tribunal Eleitoral com um prazo mínimo de 5 dias antes do conhecimento dos resultados:

[6] Em 1988, com base em recursos apresentados pelos meios de comunicação e de produção de pesquisas, as restrições foram suspensas, e em 1990 foram retiradas da legislação (Resolução 16.402/1990). Foi nesse momento que a legislação eleitoral avançou para o campo da regulação das informações, no sentido de dar transparência tanto sobre os agentes envolvidos no processo político, quanto sobre os parâmetros metodológicos de produção dos dados. Nas mudanças mais recentes, a Reforma Política parcial realizada nos anos de 2005 e 2006 definiu novas regras para a realização das campanhas eleitorais e de divulgação de pesquisas, válidas a partir das eleições municipais de 2008. Naquela lei (lei 11.300/06) foi definida a restrição da divulgação para o período dos 15 dias anteriores ao pleito. No entanto, em 8 de novembro de2007, através da Resolução 22.623 do Tribunal Superior Eleitoral, foi estabelecida uma regulamentação aberta, definindo que pesquisas realizadas em data anterior ao dia das eleições, poderão ser divulgadas a qualquer momento, inclusive no dia das eleições.

1) O contratante da pesquisa;
2) O valor e origem dos recursos;
3) A metodologia e o período de realização da pesquisa;
4) O plano amostral e ponderação quanto a sexo, idade, grau de instrução e nível econômico do entrevistado; área física de realização do trabalho, intervalo de confiança e margem de erro;
5) O sistema interno de controle e verificação, conferência e fiscalização da coleta de dados e do trabalho de campo;
6) O questionário completo aplicado ou a ser aplicado;
7) O nome de quem pagou pela realização do trabalho;
8) Documentos de comprovação do registro da empresa;
9) O nome do estatístico responsável pela pesquisa e seu registro no Conselho Regional de Estatística;
10) Número do registro da empresa responsável pela pesquisa no Conselho Regional de Estatística.

A partir do entendimento de como os desenhos amostrais são formatados pelos institutos de pesquisa do Brasil e de como a Lei Eleitoral regula a sua realização e divulgação. A seção seguinte propõe uma discussão sobre os critérios que fundamentam a avaliação dos resultados das pesquisas eleitorais.

3. Como avaliar pesquisas eleitorais no Brasil?

É notório que as pesquisas eleitorais se popularizaram no Brasil nas últimas décadas (MENDES, 1991; SILVA, et. al, 2019), ao mesmo tempo vemos um aumento na descrença com relação aos seus resultados (ALMEIDA; 2008; BRAUN; 2009). Conceitos como amostragem e inferência entraram no inconsciente coletivo da população e direcionam as críticas às dissonâncias entre as pesquisas e os resultados eleitorais. Um trabalho canônico nessa área é o livro de Frederick Mosteller (1949), intitulado *The Pre-election Polls of 1948: Report to the Committee on Analysis of Pre-Election Polls and Forecasts*, a pesquisa é o resultado de vários estudos sobre os erros das pesquisas eleitorais nas eleições norte-americanas de 1948. Mosteller *et. al.* (1949) apresentou oito formas para se mensurar a precisão das pesquisas eleitorais. Os métodos propostos pelos autores podem ser divididos em dois grupos: *(i) os que se centram na diferença entre as porcentagens absolutas de votos obtidas pelos candidatos e as estimadas pelos institutos e (ü) os que se ocupam das distâncias relativas entre os candidatos.*

Nesse sentido, os trabalhos de Gramacho (2013; 2015) utilizam a metodologia proposta pelo estatístico norte-americano para entender o caso brasileiro. A partir das idiossincrasias do sistema político brasileiro, em especial o multipartidarismo, o autor utiliza o "Método 3" de

Mosteller *et. al.* (1949). Denominado de (MM3) O método consisti em mensurar o erro de cada pesquisa eleitoral a partir da diferença entre a média dos valores absolutos da intenção de votos estimada para cada candidato e percentual de votos válidos obtidos pelo candidato (GRAMACHO, 2013). O processo de cálculo do (MM3) é o seguinte:

1) Porcentagem sem decimais da estimativa de votos feita pelo instituto X para os candidatos ***- (Intenção de Voto, I.V.);***
2) Porcentagem sem decimais do resultado obtido pelos candidatos nas eleições ***(Votação Total, V.T.);***
3) Extrai-se o valor absoluto da diferença entre ***(I.V. - V.T. = Erro);***
4) Calcula-se a média aritmética dessas diferenças: ***Método Mosteller 3 (MM3).***

Para fins ilustrativos as tabela 1, 2, 3 e 4 apresentam os cálculos do (MM3) para os dois principais institutos de pesquisa do Brasil, IBOPE e DATAFOLHA para as eleições 2018[7]. O IBOPE apresenta (MM3) inferior ao DATAFOLHA. Os cálculos são apresentados abaixo.

As variáveis (*I.V 1, 2, 3)* apresentam as intenções de votos para os candidatos a presidência em 2018. A variável (*Média I.V.*) representa a média das intenções de votos, já V.T. são os votos obtidos pelos respectivos candidatos. O Erro é a diferença entre as variáveis *(Média I.V)* e *(V.T.).* O MM3 calculado para os levantamentos 2.34 do IBOPE, 0,34 acima dos 2 pontos informados pelo instituto. A metodologia do instituto é apresentada no Anexo (1).

[7] As eleições de 2018 apresentaram características especial por diversos fatores, o candidato do Partido dos Trabalhadores (PT) o ex-presidente Luís Inácio Lula da Silva não teve sua candidatura deferida pela justiça eleitoral por conta da Lei da Ficha Limpa. A não homologação da candidatura do ex-presidente colocou na disputa o seu candidato a vice. O ex-ministro da educação e o ex-prefeito de São Paulo Fernando Haddad. Essa conjuntura fez com que o nome do candidato do Partido dos Trabalhadores (PT) Fernando Haddad só aparecesse nas pesquisas eleitorais a partir do 24/09 de 2018.

Tabela 1 – Cálculo do MM3 para o IBOPE ELEIÇÕES DE 2018 (1º Turno)

Candidatos	Partido	I.V. 1	I.V. 2	I.V. 3	Média I.V.	V.T.	Erro
Jair Bolsonaro	PSL	38	41	45	41,3	46	-4,7
Fernando Haddad	PT	28	25	28	27,0	29	-2,0
Ciro Gomes	PDT	12	13	14	13,0	12	1,0
Geraldo Alckmin	PSDB	8	8	4	6,7	4	2,7
João Amoêdo	NOVO	4	3	3	3,3	2	1,3
Marina Silva	REDE	3	3	2	2,7	1	1,7
Henrique Meirelles	MDB	2	2	1	1,7	1	0,7
Guilherme Boulos	PSOL	2	2	1	1,7	0	1,7
Cabo Daciolo	PATRIOTA	2	2	1	1,7	1	0,7
Alvaro Dias	PODEMOS	1	1	1	1,0	0	1,0
João Goulart Filho	PPL	0	0	0	0,0	0	0,0
Vera Lúcia	PSTU	0	0	0	0,0	0	0,0
Eymael	DC	0	0	0	0,0	0	0,0
Método Mosteller 3 (MM3): 2,34							

Fonte: Elaboração própria a partir de dados do site "Poder 360".

Os dados da tabela 2 seguem a mesma lógica da tabela anterior, as variáveis (*I.V 1, 2, 3)* apresentam as intenções de votos para os candidatos a presidência em 2018. A variável (*Média I.V.*) representa a média das intenções de votos, já V.T. são os votos obtidos pelos respectivos candidatos. O Erro é a diferença entre as variáveis *(Média I.V)* e *(V.T.)*. O MM3 calculado para os levantamentos 4,2 do DATAFOLHA, 4,2 acima dos 2 pontos informados pelo instituto. A metodologia do instituto é apresentada no Anexo (1).

Tabela 2 – Cálculo do MM3 para o DATAFOLHA - ELEIÇÕES DE 2018 (1º Turno)

Candidatos	Partido	I.V. 1	I.V. 2	I.V. 3	Média I.V.	V.T.	Erro
Jair Bolsonaro	PSL	33	39	40	37	46	-9
Fernando Haddad	PT	21	26	25	24	29	-5
Ciro Gomes	PDT	11	13	15	13	12	1
Geraldo Alckmin	PSDB	9	9	8	9	4	5
João Amoêdo	Novo	8	4	3	5	2	3
Marina Silva	Rede	5	3	3	4	1	3
Henrique Meirelles	MDB	4	2	2	3	1	2
Guilherme Boulos	Podemos	3	2	2	2	0	2
Cabo Daciolo	Psol	2	1	1	1	1	0
Alvaro Dias	Patriota	2	1	1	1	0	1
João Goulart Filho	PSTU	2	0	0	1	0	1
Vera Lúcia	DC	0	0	0	0	0	0
Eymael	PPL	0	0	0	0	0	0
Método Mosteller 3 (MM3): 4.2							

Fonte: Elaboração própria a partir de dados do site "Poder 360º".

O 2º turno das eleições de 2018 se transcorreu entre os dias 07/10 e 28/10. Os levantamentos do IBOPE foram realizados respectivamente nos dias 15, 23 e 27 de outubro. O erro estimado pela metodologia apresenta pelo instituto foi de 2 pontos percentuais e o (MM3) ficou abaixo, 1,4. Isso demostra que as estimativas de 2º turno são mais assertivas do que as do 1º.

Tabela 3 – Cálculo do MM3 para o IBOPE ELEIÇÕES DE 2018 (2º Turno)

Candidatos	Partido	I.V. 1	I.V. 2	I.V. 3	I.V. 4	Média I.V.	V.T.	Erro
Jair Bolsonaro	PSL	59	57	56	54	56,5	55,13	1,4
Fernando Haddad	PT	41	43	44	46	43,5	44,87	-1,4
Método Mosteller 3 (MM3): 1,4								

Fonte: Elaboração própria a partir de dados do site "Poder 360º".

Os levantamentos do DATAFOLHA foram realizados respectivamente nos dias 10, 18, 25 e 27 de outubro. O erro estimado pela metodologia apresenta pelo instituto foi de 2 pontos percentuais e o (MM3) ficou no limiar 1,9. Os dados do DATAFOLHA também se apresentaram mais assertivos no 2º. Explicações estatísticas para tal assertividade serão apresentadas na próxima seção.

Tabela 4 – Cálculo do MM3 para o DATAFOLHA ELEIÇÕES DE 2018 (2º Turno)

Candidatos	Partido	I.V. 1	I.V. 2	I.V. 3	I.V. 4	Média I.V.	V.T.	Erro
Jair Bolsonaro	PSL	58	59	56	55	57	55,13	1,9
Fernando Haddad	PT	42	41	44	45	43	44,87	-1,9
Método Mosteller 3 (MM3): 1,9								

Fonte: Elaboração própria a partir de dados do site "Poder 360º".

Segundo Gramacho (2013) em contextos multipartidários como é o caso do Brasil o (MM3) apresenta uma limitação importante na medida em que apresenta o conjunto dos resultados de cada pesquisa eleitoral e não especificamente para os candidatos tomados individualmente. Para responder a essa limitação o autor desenvolveu o (MM3C) que é o *Método de Estimação do Erro para cada Candidato*, seus valores utilizam o mesmo processo de cálculo do (MM3) até o terceiro passo, dos quatro apontados acima.

A partir dos dois métodos o (MM3) e o (MMEC), Gramacho (2013; 2015) desenvolve dois trabalhos com o objetivo de discutir as incongruências entre as pesquisas eleitorais e os resultados das urnas, nas eleições majoritárias nos pleitos de 2010 e 2014. Os resultados revelam erros superiores as margens informadas a Justiça Eleitoral, as maiores discrepâncias foram encontradas em: *(i) pesquisas realizadas com maior antecedência, (ii) pesquisas realizadas no 1º turno, (iii) em disputas pouco competitivas, (iv) quando o número de candidatos é reduzido e (v) nas eleições*

para governadores. No entanto, uma avaliação da qualidade das pesquisas eleitorais pautada única e exclusivamente na predição correta do resultado eleitoral é uma avaliação no mínimo precipitada.

Dessa forma, a construção de modelos preditivos para resultados eleitorais não deve considerar apenas as intenções de voto dos eleitores. Assim, algumas dimensões podem ser utilizadas para a construção de modelos mais assertivos, tais com: *(i) Pesquisas eleitorais anteriores, (ii) Histórica de brancos e nulos; (iii) Histórico de voto consolidado em regiões específicas do país; (iv) Padrão de rejeição de partidos e candidatos; (v) Índices de partidarismo e consolidação do sistema partidário.* Assim, avaliar pesquisas eleitorais pelo seu grau assertividade seria negligenciar uma gama de fatores que interferem na decisão do voto dos eleitores.

Sob esse prisma, Gramacho (2013) estima um modelo de regressão linear no qual a variável dependente são os valores do (MM3) calculados para as 153 pesquisas eleitorais analisados pelo autor. Das variáveis independentes mobilizadas os maiores escores estimados nos quatro modelos apresentados, são relacionados a variável *(2º Turno)*, que se refere ao turno de realização da eleição. Vários fatores podem explicar esse resultado, e a adequabilidade da distribuição de probabilidade ao desenho amostral pode ser um deles.

Com relação ao desenho amostral esse ponto foi o que orientou as discussões sobre a qualidade das pesquisas não apenas no Brasil, como também nos Estados Unidos. Contudo, para além de uma discussão sobre as características amostrais o que se propõe aqui é um debate sobre o erro amostral, popularmente conhecido como *margem de erro*. Tome-se como exemplo as eleições majoritárias brasileiras que se dão em dois turnos. No primeiro turno invariavelmente há mais de dois candidatos, variável de interesse é por definição politômica, ou seja, possui três ou mais possibilidades de escolha. E o cálculo amostral utilizado pelos institutos de pesquisa leva em conta uma distribuição dicotômica. Essa escolha gera importantes impactos no tamanho das amostras e na estimação dos erros, fatos que comprometem as inferências, não necessariamente a predição dos resultados. Assim, a próxima seção será dedicada aos fundamentos formais das duas distribuições e seus impactos no tamanho da amostra.

4. Distribuições dicotômicas e politômicas

A estimação de proporções ambienta-se em questões de dois tipos: dicotômica e politômica. As questões dicotômicas são as que apresentam dois itens, as questões politômicas são as que apresentam mais de dois itens. A teoria estatística avançou nos aspectos relacionados as questões dicotômicas, porém, em cenários eleitorais as questões são politômicas, ou seja apresentam mais de duas possibilidades aos respondentes. Mesmo em um cenário de segundo turno onde existem apenas dois candidatos, as possibilidades de abstenção elevam o número de categorias (SILVA, 2012; ASSUMÇÃO, 2017).

4.1. Questões dicotômicas

A amostragem de proporções em questões dicotômicas fundamenta-se na distribuição binomial, que diz respeito a um experimento aleatório que consiste em repetidas tentativas que apresentam apenas dois resultados possíveis (tentativas de Bernoulli) e possui as seguintes características (AGRESTI *et. al*., 2012):

1) As tentativas são independentes, ou seja, o resultado de uma não altera o resultado da outra;
2) Cada repetição do experimento admite apenas dois resultados: sucesso ou fracasso;
3) A probabilidade de sucesso (p), em cada tentativa, é constante;
4) As probabilidades para as duas categorias são as mesmas para cada observação;
5) Representamos as probabilidades por π para a categoria 1 e (1 - π) para a categoria 2.

Na Distribuição Binomial a variável aleatória **X** possui parâmetros $\boldsymbol{n}$ **e** $\boldsymbol{\theta \epsilon [0,1]}$ **se X (ω)** $\boldsymbol{\epsilon}$ **{0,1,....,n}** com

$$\mathbb{P}(\mathbf{X} = \boldsymbol{K}) = \frac{\boldsymbol{n}!}{\boldsymbol{K}!(\boldsymbol{n}-\boldsymbol{k})!}\, \boldsymbol{\theta}^{\boldsymbol{k}}(\mathbf{1} - \boldsymbol{\theta})^{\boldsymbol{n}-\boldsymbol{k}}. \tag{4.1}$$

A esperança e a variância são dadas[8] por:

$$\boldsymbol{\mu} = \boldsymbol{E}(\boldsymbol{X}) = \boldsymbol{np} \tag{4.2}$$

$$\boldsymbol{\sigma^2} = \boldsymbol{V}(\boldsymbol{x}) = \boldsymbol{np}(\mathbf{1} - \boldsymbol{p}) \tag{4.3}$$

O cálculo do tamanho da amostra $\boldsymbol{n}$**,** para casos dicotômicos**,** seria modelado pela equação abaixo:

$$\boldsymbol{n} = \frac{\boldsymbol{N}\,.\,\boldsymbol{z}^2.\boldsymbol{p}.\boldsymbol{q}}{(\boldsymbol{N}-\mathbf{1})\,.\boldsymbol{e}^2 + \boldsymbol{z}^2.\boldsymbol{p}.\boldsymbol{q}} \tag{4.4}$$

4.2. Questões Politômicas

No caso de se terem questões politômicas, a distribuição que fundamenta o processo de amostragem é a multinomial, que é uma generalização da distribuição binomial para mais de duas proporções. Situações que podem ser modeladas pela probabilidade acima questões de múltipla

[8] A média de uma variável aleatória discreta é a média ponderada dos valores possíveis de X, onde os pesos são as probabilidades. De forma similar a variância usa f(x) com peso para multiplicar cada desvio quadrado $(x - \mu^2)$. (ASSUMÇÃO, 2017).

escolha, sejam de resposta única ou múltipla, escala de Likert, escala numérica, etc. (SILVA, 2012; ASSUMÇÃO, 2017). Sejam, então, as probabilidades $\boldsymbol{\theta 1,...,\theta 2}$, **satisfazendo a** $\mathbf{0 \leq \theta i \leq 1}$, **para** $\mathbf{i=1,...,n}$, e $\sum_{i=1}^{k} \theta_{1=1}$. Então, a probabilidade conjunta de se obterem as quantidades $(\boldsymbol{n_1,...,n_k})$, a partir de uma amostra de tamanho m, é dada por:

$$\mathbb{P}\left(N = \left(n_1, n_2, \ldots . n_k\right)\right) = \frac{m!}{x_1! x_2 \ldots . x_k!} \theta_1^{n1}\ \theta_2^{n2} \ldots . \theta_3^{n3} \tag{4.5}$$

Assim, os equacionamentos dos parâmetros de amostragem (esperança e variância) que utilizamos para amostragem de questões dicotômicas são válidos para a amostragem de questões politômicas. Em função da equivalência demonstrada acima, a esperança de $\boldsymbol{n_i}$ é $\mathbf{m \cdot \theta_1}$ e a sua variância é $\mathbf{m \cdot \theta_i \cdot (1-\theta_i)}$, que são equivalentes ao caso binomial.

Desde a redemocratização o eleitor foi submetido a cenários eleitorais onde há mais de dois candidatos, essa situação deve ser modelada por uma distribuição multinomial e teria tamanho de amostra igual a:

$$\boldsymbol{n = \max_{p.q} . \frac{N\ .\ z^2\ .p.q}{(N-1)\ .e^2 + z^2\ .p.q}} \tag{4.6}$$

Em que é **N** é o tamanho da população, q é igual a **(1-p)**, e é a margem de erro, e z é o fator da distribuição normal padronizada correspondente ao nível de significância α. Geralmente, o produto **p.q** é obtido do histórico de trabalhos anteriores ou, quando totalmente desconhecido, substituído por 0,25, valor máximo que proporcionará um cálculo conservador do tamanho da amostra.

5. As distribuições de probabilidade e as correções de Bonferroni

Silva (2012) em discussão análoga a que se pretende nesse trabalho, *"apontou que a estimação de intervalos de confiança para as* $\boldsymbol{k}$ *classes precisa considerar que as estimativas de precisão são simultaneamente dadas para as* $\boldsymbol{k}$ *classes. Significaria distribuir o nível de significância global* $\boldsymbol{\alpha}$ *pelos k intervalos de estimação"* (SILVA, 2012: 125). O autor utiliza o método de Bonferroni para a correção do nível de significância. Em linhas gerais o método distribui o nível de significância global de modo igualitário pelas **k** categorias de interesse. Por exemplo, nas últimas eleições presidenciais tivemos 13 candidatos e se levarmos em conta votos brancos, nulos indecisos e abstenções, teremos no mínimo 15 classes a serem testadas em conjunto com um nível de significância global de 0,95 **(α = 0,05)**, o nível de significância corrido por Bonferroni para cada classe será de $\boldsymbol{\alpha_{15} = \alpha/k}$ = 0,05/15 = 0,0033. A razão entre o valor **α**

(para um nível de significância global de 0,95) e o número de **k** torna o seu o valor de $\boldsymbol{\alpha_k}$ menor na medida em que categorias são introduzidas nos testes. Assim, pode-se concluir que as correções de Bonferroni causam três impactos imediatos: *(i) diminuição da área de rejeição* ***(z);*** *(ii) Aumento do intervalo de confiança e (iii) e o consequente aumento da amostra* ***(n).***

O exemplo abaixo presenta uma situação hipotética para o cálculo amostral, tendo como referência as correções de Bonferroni para dados categóricos com **k=15**. O universo é o número de eleitores da cidade de Belo Horizonte, o nível de confiança é o mais utilizado pelos institutos de pesquisa 95% e a margem de erro de 2%. Os cálculos tomam como referência as formulas (3.4) e (3.6).

Exemplo 1:

O valor **z** a ser inserido no exemplo 1 corresponde a razão entre o nível de significância global (0,05) e as 15 **k** categorias em questão **(0,05/15 = 0,003334),** esse valor representará um **z = 2,712986**. O valor **p.q = 0,25** corresponde ao cenário mais pessimista da construção amostral.

$$\boldsymbol{n = \frac{1.956.410\ .\ 2,712986^2\ .\ 0,25}{(1.956.410 - 1)\ .\ 0,02^2 + 2,712986^2\ .\ 0,25} = 4589,39\ \cong 4589}$$

Exemplo 2:

A equação abaixo apresenta o mesmo cálculo amostral para duas categorias. O valor **z** a ser inserido na equação 2 corresponde a razão entre o nível de significância global (0,05) e as 2 **k** categorias em questão **(0,05/2 = 0,025),** esse valor representará um **z = 1,959964**.

$$\boldsymbol{n = \frac{1.956.410\ .\ 1,959964^2\ .\ 0,25}{(1.956.410 - 1)\ .\ 0,02^2 + 1,959964^2\ .\ 0,25} = 2397,97\ \cong 2398}$$

O exemplo 1 apresentou a correção do nível de significância global (0,05) pelas 15 categoriais introdução na equação. As correções elevaram a magnitude amostral a um valor quase duas vezes superior ao cálculo sem correções exemplificado no exemplo 2. É importante notar com o valor (n) no segundo exemplo se aproxima muito dos cálculos amostrais realizados pelos institutos de pesquisa brasileiros que giram em torno de 2000 entrevistados.

A tabela 5 ilustra os cálculos acima e mostra os valores das correções de Bonferroni para **α** para as **k** categorias. Assim, os valores das correções pelas **k** categorias se apresentam mais conservadores do que as estimativas não corrigidas, fato que mitiga as possibilidades de erro do tipo I.

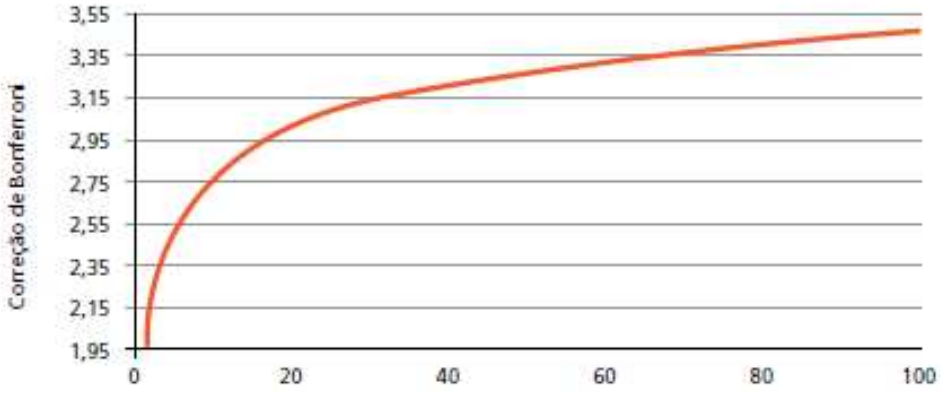
Tabela 5 – Correções de α Bonferroni

k	α = 5%	α = 10%
2	0,05000	0,10000
3	0,02500	0,05000
4	0,01667	0,03333
5	0,01250	0,02500
6	0,01000	0,02000
7	0,00833	0,01667
8	0,00714	0,01429
9	0,00625	0,01250
10	0,00556	0,01111

Fonte: Silva (2012).

O valor **z** da distribuição normal, não obstante aumente bastante, não explode com o quantitativo maior de itens **k**, não comprometendo assim o tamanho da amostra. No gráfico 1, para **α** = 5%, foram plotados os valores de **z** em função de **k**, pela formulação de Bonferroni, que a opção mais conservadora a resultar em z mais elevado. Tomando-se, como parâmetro de comparação o k=2, em que convencionalmente z = 1,9599, podemos verificar quão reduzido é o avanço de z com aumento de k, a ponto de z será penas ainda 3,4780, quando k atinge 100. Nesse caso, o aumento do tamanho da amostra, considerando a população suficientemente grande, seria de 216% (=(3,4780/1,9566)2-1), mesmo considerando um aumento de k=2 para k=100 (SILVA, 2012).

Gráfico 1 – Variação da correção de Bonferroni em função do número de itens k, para α = 0,05



Fonte: Silva (2012)

Conclusões

As pesquisas eleitorais já fazem parte da rotina democrática das sociedades modernas. Com a divulgação de informações sobre os rumos dos processos eleitorais, elas diminuem as assimetrias informacionais entre os indivíduos. Assim, os números apresentados são discutidos e problematizados não apenas por especialistas, mas também pela sociedade como um todo. O efeito colateral dessa popularização está no aumento dos questionamentos relacionados a assertividade dos *surveys* pré-eleitorais. A discussão se vincula ao papel das pesquisas em prever o resultado das urnas. Um pressuposto importante do presente trabalho está em tomar as pesquisas eleitorais como levantamentos estaque, na medida em que seus resultados refletem conjunturas especificas e seus resultados não devem ser medidos apenas pela assertividade ou não dos resultados eleitorais.

O que o presente trabalho propõe não é uma discussão das pesquisas eleitorais a partir dos resultados das urnas e sim uma discussão sobre a lógica amostral empregada pelos institutos. Os cálculos realizados nos *Exemplos 1 e 2* demostram uma clara sub-representação amostral no desenho realizado a partir de uma distribuição binomial. Em cenários onde a consolidação do voto e institucionalização do sistema partidário apresentam índices mais robustos, essa sub-representação pode não apresentar impactos efeitos. Porém, no caso brasileiro onde as preferencias são cada vez mais voláteis e o sistema partidário passa por elevada deterioração, amostras subestimadas podem nos propiciar fotos desfocadas dos cenários eleitorais.

Tem-se ciência aqui dos custos financeiros e logísticos de um aumento na magnitude amostral para os institutos. O fato é que a democracia brasileira nos coloca a cada dia novas questões e novos desafios, as experiências recentes comprovaram que o processo democrático se apresenta cada vez mais complexo e as pesquisas eleitorais devem acompanhar esse caminho, que parece cada vez mais inexorável.

Referências

ASSOCIAÇÃO BRASILEIRA DE EMPRESAS DE PESQUISA. **Publicação de Pesquisas Eleitorais**. São Paulo: ABEP, 2008.

AGRESTI, Alan; FINLAY, Barbara. **Métodos estatísticos para as ciências sociais**. 4. ed. Porto Alegre: Penso, 2012.

ALMEIDA, Alberto Carlos de. **A cabeça do eleitor**: estratégia de campanha, pesquisa e vitória eleitoral. São Paulo: Record, 2008.

ASSUMÇÃO, R. **Fundamentos Estatísticos da Ciência de Dados**: voltado para aplicações. BOOKWEBSITE.COM, 2017.

BRAUN, Cecilia; Maíra STRAW, C. (Org.). **Opinion Pública**: una mirada desde América Latina. Buenos Aires: Planeta, 2009.

FERRAZ, Cristiano. **Crítica Metodológica às Pesquisa Eleitorais no Brasil,** Dissertação de Mestrado, UNICAMP, 1996.

GRAMACHO, Wladimir. G. À margem das margens? A precisão das pesquisas pré-eleitorais

brasileiras em 2010. **Opinião Pública**, v. 19, n. 1, p. 65-80, 2013.

GRAMACHO, Wladimir G. A pesquisa governamental de opinião pública: razões, limites e a experiência recente no Brasil. **Revista do Serviço Público**, v. 65, n.1, p. 49-64, 2014.

HENKEL, Karl. Análise da não resposta em *surveys* políticos. **Opinião Pública**, Campinas, v. 18, n. 1, p. 216-238, jun., 2012.

KING, Gary, et. al. Analyzing incomplete political science data: an alternative algorithm for multiple imputation. **American Political Science Review**, v. 95, n. 1, p.49-69, 2001.

MENDES, Antonio Manuel Teixeira. O papel das pesquisas eleitorais. **Novos Estudos Cebrap**, São Paulo, v. 1, n. 29, p. 28-33, mar. 1991

MOSTELLER, Frederick. Measuring the error. In: MOSTELLER, Frederick, et. al. **The pre-election polls of 1948**. Report to the committee on analysis and pre-election polls and forecast. New York: Social Science Research Council, 1049, p. 54-80.

Moser, Claus, et. al. An experimental study of quota sampling. **Journal of the Royal Statistical Society**: Series A 116, p.349–405, 1953.

SILVA, Ângelo Henrique Lopes da. Estimativa de proporções em questões politômicas. **Revista do TCU**, n.125, p.18-27, 2012.

SILVA, Bruno Fernando da; GONCALVES, Ricardo Dantas. Pesquisas eleitorais afetam receitas de campanha: a correlação entre expectativa de vitória e financiamento de campanha em disputas ao Senado. **Revista de Sociologia e Política**, Curitiba, v. 27, n. 71, p.2-17, 2019.

SMITH, T. W. The hidden 25 percent: an analysis of nonresponse on the 1980 General Social Survey. **Public Opinion Quarterly**, v. 47, n. 3, p. 386-404, 1983.

Anexo (1)

IBOPE INTELIGENCIA PESQUISA E CONSULTORIA LTDA

Metodologia de pesquisa

Pesquisa quantitativa, que consiste na realização de entrevistas pessoais, com a aplicação de questionário estruturado junto a uma amostra representativa do eleitorado em estudo. Pesquisa realizada no estado do Rio de Janeiro.

Plano amostral

Plano amostral e ponderação quanto a sexo, idade, grau de instrução e nível econômico do entrevistado; intervalo de confiança e margem de erro:

Representativa dos votantes da área em estudo, elaborada em três estágios. No primeiro estágio fez-se um sorteio probabilístico dos municípios pesquisados, pelo método PPT (Probabilidade Proporcional ao Tamanho), com base na população de votantes (TSE 2018, excluindo abstenção 1º turno 2010 e 2014) de cada município. Em um segundo estágio, dentro dos municípios sorteados foram selecionados, também pelo método PPT, os locais de votação com base no número de votantes de cada local de votação. No terceiro estágio, dentro dos locais de votação sorteados, os respondentes foram selecionados através de quotas amostrais, proporcionais em função de variáveis significativas, a saber: Sexo, Idade, de acordo com o perfil dos votantes. Devido à metodologia amostral adotada, a pesquisa é auto-ponderada, ou seja, as proporções do universo pesquisado estão previstas na amostra, não sendo necessário qualquer tipo de ponderação quanto à sexo, idade, grau de instrução e nível econômico. O intervalo de confiança estimado é de 99% e a margem de erro máxima estimada considerando um modelo de amostragem aleatório simples, é de 03 (três) pontos percentuais para mais ou para menos sobre os resultados encontrados no total da amostra.

Sistema interno de controle e verificação

Para a realização da pesquisa, utiliza-se uma equipe de entrevistadores e supervisores contratados pelo IBOPE INTELIGÊNCIA PESQUISA E CONSULTORIA LTDA. devidamente treinados para o trabalho. Após os trabalhos de campo, os questionários são submetidos a uma fiscalização de cerca de 20% (vinte por cento) dos questionários aplicados pelos entrevistadores para verificação das respostas e da adequação dos entrevistados aos parâmetros amostrais.

DATA FOLHA: INSTITUTO DE PESQUISAS LTDA

Metodologia de pesquisa

Pesquisa do tipo quantitativo, por amostragem, com aplicação de questionário estruturado e abordagem pessoal em pontos de fluxo populacional. O conjunto do eleitorado brasileiro com 16 anos ou mais foi tomado como universo da pesquisa.

Plano amostral

Plano amostral e ponderação quanto a sexo, idade, grau de instrução e nível econômico do entrevistado; intervalo de confiança e margem de erro:

Universo: Eleitorado brasileiro, com 16 anos ou mais. Tamanho da amostra: A amostra prevista é de 18.060 entrevistas. Técnica de amostragem: A amostra é estratificada por região geográfica e natureza dos municípios (capital, região metropolitana ou interior). Em cada estrato, num primeiro estágio, são sorteados os municípios que farão parte do levantamento. Num segundo estágio, são sorteados os bairros e pontos de abordagem onde serão aplicadas as entrevistas. Por fim, os entrevistados são selecionados aleatoriamente para responder ao questionário, de acordo com cotas de sexo e faixa etária. Nesta amostra, os tamanhos dos estratos foram desproporcionalizados para permitir detalhamento das seguintes unidades da federação e suas capitais: SP, RJ, MG, além do DF. Nos resultados finais, as corretas proporções serão restabelecidas através de ponderação. Os dados utilizados para definição e seleção da amostra são baseados nos dados fornecidos pelo TSE ¿ Tribunal Superior Eleitoral (eleitorado de agosto de 2018) e IBGE (Estimativa 2018). Os dados relativos a sexo e faixa etária são: Sexo masculino: 47%, feminino: 53%, 16 a 24 anos 15%, 25 a 34 anos 21%, 35 a 44 anos 21%, 45 a 59 anos 24% e 60 anos ou mais 19%. Ponderação dos resultados: No processamento dos dados é realizada ponderação referente à proporção de cada cidade na amostra para correta representação das regiões geográficas. Está prevista a eventual ponderação para correção nos tamanhos dos segmentos considerando as variáveis sexo e faixa etária. Para as variáveis grau de instrução e nível econômico do entrevistado (renda familiar mensal), o fator previsto para ponderação é 1 (resultados obtidos em campo). Área física: Serão realizadas entrevistas em 341 municípios, localizados nas seguintes unidades da federação: Acre, Alagoas, Amazonas, Amapá, Bahia, Ceará, Distrito Federal, Espírito Santo, Goiás, Maranhão, Minas Gerais, Mato Grosso, Mato Grosso do Sul, Pará, Paraíba, Paraná, Pernambuco, Piauí, Rio de Janeiro, Rio Grande do Norte, Rondônia, Roraima, Rio Grande do Sul, Santa Catarina, São Paulo, Sergipe e Tocantins. A relação completa dos municípios e bairros pesquisados será encaminhada a esse tribunal posteriormente até o sétimo dia seguinte à data de registro da pesquisa, conforme a Resolução 23.549/2017 do TSE, no art.2º §6º. Margem de Erro: A margem de erro máxima prevista é de 2 pontos percentuais para mais ou para menos, considerando um nível de confiança de 95%. Os intervalos de confiança serão calculados considerando os resultados obtidos para um nível de confiança de 95%.

Sistema interno de controle e verificação

Os pesquisadores envolvidos na realização desta pesquisa são treinados pelo Instituto e recebem instruções específicas para cada projeto realizado. A coleta será feita com a utilização de tablet e questionário eletrônico. São checados, no mínimo, 30% dos questionários de cada pesquisador, seja in loco por supervisores de campo ou, posteriormente, por telefone. Internamente, todo o material é verificado e codificado. Antes do processamento final e emissão dos resultados, realiza-se processo de consistência dos dados.

Anexo (2)

- Amostragem a partir de uma Distribuição Binomial

$$\boldsymbol{n = \frac{Z^2_{1-\alpha/k}}{4\ x\ \varepsilon^2}}$$

- Amostragem a partir de uma Distribuição Multinomial

$$\boldsymbol{n = \frac{Z^2_{1-\alpha/2k}}{4\ x\ \varepsilon^2}}$$

Onde:

- $\boldsymbol{Z^2_{1-\alpha/2}}$ **=** Valor tabelado da curva normal padrão;

- n = Tamanho da amostra;

- ε = erro máximo admitido;

- 1 – α = nível de confiança;

- K = número de categorias.

Artigo recebido em: 2020-03-31
Artigo reapresentado em: 2020-05-13
Artigo aceito em: 2020-06-05