AMOSTRAGEM E ANÁLISE DE DADOS: UM ESTUDO DE CASO SOBRE PESQUISAS ELEITORAIS BRASILEIRAS DE 2022

Mariane Patrícia Sargi (UFU) - marianesargi@gmail.com

Jéssica Rayse de Melo Silva (UFU) - jessicar@ufu.br

Carlos Roberto Souza Carmo (FACIC-UFU) - carlosjj2004@hotmail.com

Resumo:

Por meio de um estudo de caso único, no qual foram analisados os dados de pesquisas eleitorais realizadas por seis institutos para o segundo turno das eleições presidenciais de 2022, esta investigação teve por objetivo avaliar como a definição da amostra de pesquisa pode influenciar os resultados do processo de análise de dados. Para tanto, comparativamente aos resultados reais observados nas eleições, foram calculados o erro absoluto de cada pesquisa, as quantidades totais e médias de entrevistas e seus respectivos desvios-padrão, além do Coeficientes de Variação de Pearson. Em seguida, utilizou-se o coeficiente de correlação de Spearman para avaliar quais das medidas de tendência central e de dispersão apresentaram relação com o comportamento dos erros absolutos observados nas pesquisas eleitorais que compuseram este estudo de caso. Os resultados desta investigação indicam que não há evidências de que os institutos de pesquisa eleitoral no Brasil apresentaram algum tipo de viés doméstico em seus resultados, e que a distribuição geográfica no processo de amostragem parece ter interferido nos resultados das pesquisas eleitorais, uma vez que a seleção de cidades da região sudeste do Brasil foi mais significativo para a precisão do resultado das pesquisas eleitorais do que a quantidade de entrevistas realizadas, comparativamente aos resultados reais observados nas respectivas eleições. Esses achados do estudo contribuem para a compreensão da metodologia de amostragem das pesquisas eleitorais no Brasil, fornecendo subsídios para a discussão sobre a importância de avaliação do processo de amostragem.

Palavras-chave: *amostragem; análise; métodos quantitativos aplicados.*

Área temática: *Contabilidade Aplicada ao Setor Público e ao Terceiro Setor*

AMOSTRAGEM E ANÁLISE DE DADOS: UM ESTUDO DE CASO SOBRE PESQUISAS ELEITORAIS BRASILEIRAS DE 2022

Resumo:
Por meio de um estudo de caso único, no qual foram analisados os dados de pesquisas eleitorais realizadas por seis institutos para o segundo turno das eleições presidenciais de 2022, esta investigação teve por objetivo avaliar como a definição da amostra de pesquisa pode influenciar os resultados do processo de análise de dados. Para tanto, comparativamente aos resultados reais observados nas eleições, foram calculados o erro absoluto de cada pesquisa, as quantidades totais e médias de entrevistas e seus respectivos desvios-padrão, além do Coeficientes de Variação de Pearson. Em seguida, utilizou-se o coeficiente de correlação de Spearman para avaliar quais das medidas de tendência central e de dispersão apresentaram relação com o comportamento dos erros absolutos observados nas pesquisas eleitorais que compuseram este estudo de caso. Os resultados desta investigação indicam que não há evidências de que os institutos de pesquisa eleitoral no Brasil apresentaram algum tipo de viés doméstico em seus resultados, e que a distribuição geográfica no processo de amostragem parece ter interferido nos resultados das pesquisas eleitorais, uma vez que a seleção de cidades da região sudeste do Brasil foi mais significativo para a precisão do resultado das pesquisas eleitorais do que a quantidade de entrevistas realizadas, comparativamente aos resultados reais observados nas respectivas eleições. Esses achados do estudo contribuem para a compreensão da metodologia de amostragem das pesquisas eleitorais no Brasil, fornecendo subsídios para a discussão sobre a importância de avaliação do processo de amostragem.

Palavras-chave: amostragem; análise; métodos quantitativos aplicados.

Área temática do evento: Mercados Financeiro, de Crédito e de Capitais.

1 INTRODUÇÃO

Tanto na tomada de decisões quanto na pesquisa científica, o processo de análise de dados para produção de informações frequentemente esbarra em limitações relacionadas à disponibilidade e à qualidade das observações realizadas e/ou coletadas acerca de determinado objeto de interesse e estudo.

As dificuldades e os custos de se realizar o levantamento dos dados de toda uma população, portanto, de se realizar um censo, levam tomadores de decisão, analistas e pesquisadores, entre outros interessados, a recorrerem ao processo de amostragem, que consiste basicamente na seleção de parte de uma população para levantamento de um conjunto de dados referentes a determinado objeto de interesse ou estudo.

Por outro lado, as desvantagens inerentes à opção por não realizar a coleta e a análise dos dados de toda uma população (censo) estão diretamente relacionadas à qualidade da amostra coletada, ou seja, se a amostra é boa, a análise é boa, sendo que, o contrário também é verdadeiro. De forma análoga, poderia se afirmar que a qualidade dos resultados de uma pesquisa baseada na análise de dados está para amostra assim como a matéria-prima utilizada na produção de determinado bem tem influência direta na qualidade do produto final.

Nesse contexto, a presente proposta de pesquisa será conduzida a partir do seguinte questionamento direcionador: como a definição da amostra de pesquisa pode influenciar os resultados do processo de análise de dados, especialmente, à luz de fenômenos contemporâneos relacionados ao dia a dia da sociedade brasileira?

Se por um lado a análise de dados baseada em um censo é mais precisa, por outro, a utilização de amostras pode conter certa margem de erro. Contudo, sob a ótica de fatores como conveniência, tempo e custo, entre outras dificuldades, os benefícios do processo de amostragem superam o censo.

Dessa forma, ao considerar "como é dispendioso, difícil e por vezes impraticável ter acesso a toda uma população, costuma-se escolher uma amostra e estudá-la" (DOWNING; CLARK, 2006, p.3). Sendo que, "para evitar predições imprecisas, é essencial que a amostra represente efetivamente a população da qual foi extraída" (DOWNING; CLARK, 2006, p.3). Percebe-se que a identificação de amostras significativamente representativas de uma determinada população pode ser considerada um dos fatores críticos do processo de análise de dados e produção de informações.

Assim, o estudo acerca dos métodos de seleção de amostras se justifica pela inexorável relação entre a amostragem e a validade dos resultados do processo de análise de dados. Pois, se a escolha adequada da amostra "diz respeito à identificação do comportamento acerca de um todo (universo maior de determinado objeto de estudo, denominado de população) a partir de uma parcela menor desse todo (denominado de amostra)", não existe qualquer garantia que o simples levantamento de um elevado número de observações seja capaz de permitir realizar inferências adequadas sobre comportamento de determinado grupo de objetos, indivíduos, fenômenos, etc. (CARMO; LIMA, 2018, p. 94).

2 REFERENCIAL TEÓRICO

A produção de pesquisas experimentais inicia-se com a observação da realidade e posteriormente à escolha do método para processamento e análise de dados, de modo que as informações obtidas e resultados alcançados assegurem ao pesquisador um valor científico (SILVA, 2004). Tão importante quanto a escolha do método, seja o quantitativo, o levantamento ou experimentação, será a adoção adequada dos procedimentos para a realização da pesquisa de forma que garanta resultados confiáveis (BOLFARINE; BUSSAB, 2005).

O objeto de estudo em análise no presente trabalho de pesquisa é o procedimento de amostragem, sendo que, a qualidade do procedimento em si caracteriza-se como o alvo do respectivo questionamento direcionador. Nesse sentido, uma vez que a amostra é a porção de um todo, é primordial distinguir alguns conceitos para que não haja confusão do que se pretende ao longo do processo de levantamento dos dados necessários ao processo de pesquisa e análise quantitativa. Dessa forma, se a amostra é uma porção de um todo, a esse todo denomina-se população.

Fávero e Belfiore (2017, p. 171) explicam que "população é conjunto com todos os indivíduos, objetos ou elementos a serem estudados, que apresentam uma ou mais características em comum". Para Silva (2004, p. 18), "a população de estudo é definida como a parte identificável e acessível do conjunto (população-objeto) para o qual idealmente se deseja a informação". Bolfarine e Bussab (2005, p. 6-7) afirmam que "[...] se entende por população a reunião de todas as unidades elementares, [...]" sendo que, "[...] a unidade elementar, ou simplesmente elemento de uma população, é o objeto ou entidade portadora das informações que pretende-se coletar". Assim, a amostra é a unidade elementar, ou seja, a fração da população.

Nota-se que o traço comum nos conceitos de população descritos é o fato dela se caracterizar como o conjunto ou reunião de elementos que atendam ao objetivo perseguido. Logo, população não é um mero quantitativo ou conjunto de escolhas fortuitas, são conjuntos com, no mínimo, um agrupamento de características comuns que trazem consigo elementos próprios do objeto de interesse pesquisado.

A população de interesse, isso é, aquela a ser pesquisada, pode ainda ter duas características próprias: finitude ou infinitude (WERKEMA, 2014). Segundo Werkema (2014, p. 10), "uma população finita possui um número limitado de elementos" e, de modo contrário, "uma população infinita possui um número não limitado de elementos".

A pesquisa realizada com base em toda a população é definida como um censo (FÁVERO; BELFIORE, 2017). De forma análoga a Fávero e Belfiore (2017), Akanime e Yamamoto (2013, p. 8) explicam que, "um levantamento estatístico (pesquisa) que abrange todos os elementos de uma população é denominado censo". Nesse sentido, pode-se concluir que censo é a análise de toda a população devidamente delimitada pelo objeto de interesse. Porém, nem sempre é possível realizar um censo, quer seja pelo elevado custo, pelo tempo demandado para a coleta de dados ou quadro de pessoal disponível para tanto. Dessa forma, torna-se necessária a realização de estudos baseados em amostras, no lugar do censo.

Percebe-se que a amostra, "[...] é uma parte da população, retirada segundo uma regra conveniente" (COSTA, 2015, p. 15). Sendo que, "o propósito da amostra é o de fornecer informações que permitam descrever os parâmetros do universo, da maneira mais adequada possível. A boa amostra permite a generalização de seus resultados dentro de limites aceitáveis de dúvidas" (BOLFARINE; BUSSAB, 2005, p.14). Dessa maneira, a amostra é uma porção ou fração da população pesquisada, a partir da qual pode-se realizar inferências acerca das características da população.

A qualidade da amostra pode determinar uma melhor representatividade da população. E, para tanto, devem ser levados em conta os fatores relacionados ao tamanho da amostra e à metodologia empregada na sua obtenção (FONSECA; MARTINS, 2012).

Com relação específica à metodologia utilizada para obtenção de uma amostra, denominado por amostragem, "o pressuposto básico para que se utilize determinada regra de amostragem é que ela gere amostras representativas, isto é, com todas as características básicas e importantes do universo: que seja uma verdadeira miniatura da população" (COSTA, 2015, p.15). Nesse sentido, o processo de amostragem deve propiciar a extração de informações e a realização de inferências acerca de valores populacionais, tomando por base apenas uma parcela de toda a população (SILVA, 2004).

De uma maneira geral, as técnicas de amostragem podem ser classificadas em amostragem probabilística e não probabilística. Acerca da amostragem probalística, Silva (2004) e Costa (2015) possuem entendimento semelhante e, segundo eles, a seleção dos elementos amostrais é aleatória e todo elemento tem possibilidades reais e iguais de pertencer à amostra escolhida. De forma análoga, Fonseca e Martins (2012) complementam afirmando que, devido àquela probabilidade (real) descrita por Silva (2004) e Costa (2015), somente mediante o processo de amostragem probabilística é que podem ser realizadas inferências generalizáveis para toda a população com base na respectiva amostra.

A amostragem não probalística diverge da probalística na forma pela qual os elementos que compõem a amostra podem não ser escolhidos (COSTA, 2015). Consequentemente, uma vez que não existe a probabilidade real e igual de se escolher os elementos integrantes da amostra, o processo de amostragem não-probabilística não permite que inferências sejam generalizadas para a respectiva população (FONSECA; MARTINS, 2012).

Dessa forma, no processo amostragem não-probalística, "as amostras são obtidas de forma não aleatória, ou seja, a probabilidade de alguns ou de todos os elementos da população pertencerem à amostra é desconhecida" (FÁVERO; BELFIORE, 2017, p. 178). Devido a isso, "[...] não é possível estimar o erro amostral e nem generalizar os resultados da amostra para a população [...]" (FÁVERO, BELFIORE, 2017, p. 178).

Amostragem probabilística tem como principais métodos: a amostragem aleatória simples sem reposição, amostragem aleatória simples com reposição, amostragem sistemática, amostragem estratificada e amostragem por conglomerado, enquanto principais técnicas.

A amostragem aleatória simples consiste no método mais básico e mais usualmente utilizado. Segundo Costa (2015, p. 16), "neste processo, todos os elementos da população têm igual probabilidade de serem escolhidos, não só antes de ser iniciado, como até completar-se o processo de coleta".

Além disso, na amostragem aleatória simples sem reposição, cada elemento sorteado só integrará a amostra uma única vez, posto que, após sorteado, ele deixa de integrar a respectiva população (SILVA, 2004).

A amostragem aleatória simples com reposição difere-se da amostragem aleatória simples sem reposição quando há reposição do elemento sorteado, isso é, o elemento sorteado volta a compor a população e pode ser sorteado novamente (BOLFARINE; BUSSAB, 2005).

Tem-se, portanto, que a amostragem aleatória simples com ou sem reposição diferenciam-se pela recolocação ou não do mesmo elemento para sorteio.

Na amostragem sistemática para a composição da amostra há uma regra predeterminada para a escolha dos elementos (VIRGILITO, 2017). Ou seja, os elementos da população apresentam algum de organização que permite sua retirada de forma sistematizada.

O processo de amostragem aleatória estratificada ocorre quando a população de interesse apresenta subdivisões devido à presença de características comuns aos seus elementos, de forma que é possível agrupá-los em estratos com características homogêneas, e, dessa maneira, a amostra é retirada desses estratos de maneira aleatória (FONSECA; MARTINS, 2012).

A amostragem por conglomerados acontece quando os elementos da população são associados em grupos e, por isso, ocorre o sorteio de alguns destes grupos para a composição da amostra (SILVA, 2004).

As técnicas de amostragem probabilística não se esgotam por aqui, entretanto os tipos descritos representam aqueles mais encontrados na literatura disponível sobre o tema.

Dentre as diversas técnicas de amostragem não probabilística recorrentemente abordadas na literatura sobre o tema destacam-se: amostragem por conveniência, amostragem por julgamento, amostragem por quotas e amostragem bola de neve.

A amostragem por conveniência ou acidental decorre da acessibilidade e/ou facilidade de encontrar os elementos para composição da amostra (FONSECA; MARTINS, 2012). Existe o julgamento do pesquisador quanto a escolha da amostra na amostragem por conveniência, portanto, esse tipo de amostra não traduz uma representação adequada da população.

Na amostragem por julgamento ou amostragem intencional o pesquisador escolhe de modo predeterminado os elementos ou grupos que irão compor a amostra (COSTA, 2015). O risco da amostragem intencional é a subjetividade, isso é, o pesquisador pode equivocar-se quanto aos elementos convenientes na amostra inabilitando-os e, elegendo os inconvenientes.

A amostragem por quotas se dá, inicialmente, com a identificação de características de interesse na população em estudo; em seguida, identificam-se cotas da população de acordo com essas características, e, finalmente, selecionam-se os componentes da amostra de forma proporcionalmente equivalente àquelas cotas estabelecidas inicialmente (FÁVERO; BELFIORE, 2017; VIRGILITO, 2017).

Na amostragem bola de neve identificam-se aqueles elementos da população com características próprias, de difícil acesso ou, até mesmo, desconhecidas, e, a partir deles, identificam-se outros indivíduos com tais características, repetindo-se esse processo até que se

esgotem os indivíduos com aquelas características iniciais, ou ainda, que não proporcionem novas informações e/ou evidências ao processo de pesquisa (FÁVERO; BELFIORE, 2017).

As técnicas de amostragem não probabilística ou não aleatória são muitas e bem variadas, sendo que, o traço comum a todas elas é justamente a utilização algum tipo de procedimento que inviabiliza a generalização das inferências realizadas para a respectiva população; cabendo ressaltar que, apesar de não esgotar esse conjunto de possibilidades, as técnicas abordadas nessa seção secundária são aquelas mais recorrentes na literatura disponível sobre o tema.

Assim, a pesquisa realizada com base em levantamentos amostrais parte da identificação da população e das suas características de interesse, para, em seguida, tomar-se a decisão acerca do respectivo planejamento amostral. Tudo isso sem perder de vista a relação custo-benefício, ou seja, deve-se avaliar se o custo do processo de amostragem é maior ou menor que o benefício esperado acerca das inferências que podem ser generalizadas ou não para a respectiva população.

Nesse contexto, as pesquisas eleitorais utilizam a estatística, que por meio da sua metodologia deve fornecer informações seguras, imparciais e isentas de interferência política. Assim, tais pesquisas devem coletar dados que permitam realizar a tomada de decisões apoiadas em determinado nível de credibilidade e considerando a respectiva margem de erro (IGNÁCIO, 2010). Por isso, as técnicas empregadas nas pesquisas eleitorais devem garantir a sua confiabilidade, de modo que, levando-se em conta o respectivo procedimento amostral, seja possível calcular o seu erro total e avaliar a respectiva qualidade (BERNDT, 1989).

Em 1935, George Gallup começou a desenvolver pesquisas eleitorais de maneira mais científica e, ao valer-se dos conceitos e técnicas de amostragem próprios da ciência estatística, introduziu a ideia de que uma amostra grande não é precisamente uma amostra de qualidade (CHALEGRE; ALMEIDA; PASSOS, 2006).

Ao realizarem comparativos entre as pesquisas de opinião e os resultados finais da votação em eventos como o referendo da independência escocesa em 2014, as eleições presidenciais dos EUA e o referendo do Brexit em 2016, Pekar, *et. al*. (2021) observam que possíveis discrepâncias entre pesquisas e os respectivos resultados eleitorais sejam decorrentes da falta representatividade amostral, a queda na taxa de resposta nas urnas, além da questão relacionada à preferências e rejeições decorrentes de fatores sociais como intolerância, imigração, liberalismo versus conservadorismo, entre outros. Nesse contexto, os métodos de previsão eleitoral, baseados em pesquisas de intenção de voto, desempenham um papel importante no processo de previsão e tomada de decisões voltadas para o resultado de uma eleição (LEVENE; FENNER, 2021), por isso, qualquer que seja a modelagem utilizada nesse tipo de previsão, a amostragem assume especial relevância, uma vez que ela constitui-se na "matéria-prima" básica do planejamento e da tomada de decisões baseada em inferências realizadas a partir de um conjunto de dados.

Ao analisar as fontes de erros nas pesquisas eleitorais, Berndt (1989) pondera que tais erros podem apresentar origens em aspectos metodológicos e aspectos inerentes às medidas de erro propriamente ditas. Especificamente em relação ao processo metodológico, Berndt (1989) destaca fatores relacionados à amostragem, mensuração e projeção de resultados. Nesse sentido, o distanciamento da amostragem probabilística pode ocasionar dúvidas quanto as classes populacionais que possui, podendo os elementos das classes estarem sub ou super-representados (BERNDT, 1989). Adicionalmente, Berndt (1989, p.83) ainda faz a seguinte ponderação relacionada aos erros das pesquisas eleitorais: "é incorreto reduzir o erro total da pesquisa apenas ao erro amostral"; e ainda completa: "pelo contrário, a concorrência de 'vie-

ses' conceituados na estatística como erros sistemáticos, sempre na mesma direção, com grande frequência sobrepujam em dimensões os erros amostrais".

Nesse sentido, Pickup e Johnston (2008) avaliaram o erro nas pesquisas da campanha presidencial dos Estados Unidos da América de 2004 comparando os métodos frequentista e bayesianos. Pickup e Johnston (2008) destacam, ao menos, duas possibilidades do distanciamento do resultado das urnas para as pesquisas: um, erro em relação a limitação do tamanho amostral, o outro, a diferença na seleção dos elementos da amostra de uma empresa para outra, o que denominado de erro de viés doméstico.

Em outras palavras, se por um lado as diferenças entre os resultados da pesquisas e os resultados eleitorais efetivos podem ocorrer devido a forma de cálculo do respectivo erro, por outro lado esse erro pode ser decorrente do mal dimensionamento das respectivas amostras e/ou devido ao mal planejamento do processo amostral em si, à forma como questionamentos são formulados, ponderações, etc.. Sendo que, esse tipo de erro (mal planejamento do processo amostral) pode acontecer tanto de uma empresa de pesquisa para outra, como no segmento de pesquisas como um todo (PICKUP; JOHNSTON, 2008).

Lauderdale *et al*. (2020) utilizaram métodos de regressão multinível e pós-estratificação (MRP) nas pesquisas pré-eleitorais do referendo do Reino Unido sobre a adesão à UE em 2016, às eleições presidenciais dos EUA em 2016 e às eleições gerais do Reino Unido em 2017, com o objetivo de formar estimativas probabilísticas consistentes. Para Lauderdale *et al*. (2020), a ocorrência de erros de pesquisa periódicos e de alto perfil não surpreende devido às evidências de erros amostrais decorrentes da metodologia tradicional de pesquisa eleitoral, que normalmente não utiliza amostras probabilísticas e com baixas taxas de resposta.

Turgeon e Rennó (2012) buscaram realizar previsões eleitorais diante de um número pequeno de eleições presidenciais realizada no Brasil, para tanto, analisaram as eleições de 1994, 1998, 2002, 2006 e 2010, cujo objetivo foi aumentar o tamanho da amostra utilizando dados subnacionais, ou seja, os dados nacionais foram divididos em unidades menores (unidades da federação). Obviamente, Turgeon e Rennó (2012) observaram que os dados estaduais foram melhores no processo de previsão dos resultados observados nas urnas. Até porque, conforme Bhattacharyya e Dey (2021) observam, numa eleição em grande escala, é necessário que a amostra coletada não seja muito pequena. Mas, em contraposição, deve-se levar em conta a ideia de que uma amostra grande não é precisamente uma amostra de qualidade (CHALEGRE; ALMEIDA; PASSOS, 2006). Por isso, e conforme já dito, a presente pesquisa leva em conta o fato de que a qualidade da amostra assume especial relevância, uma vez que ela se constitui na "matéria-prima" básica do planejamento voltada para a tomada de decisões baseadas em inferências realizadas de dados.

3 METODOLOGIA DE PESQUISA

Após o levantamento acerca da teoria adjacente para embasamento teórico sobre a temática relacionada às técnicas de amostragem, implementou-se um estudo de caso único, no qual foram abordados e analisados os dados das pesquisas eleitorais realizadas por sete institutos, cujos resultados referentes ao percentual de votos válidos para cada candidato foram divulgados na véspera do segundo turno das eleições presidenciais de 2022 (29/10/2022) pelo jornal Estado de Minas (2022), comparativamente ao resultado real observado ao final da referida eleição, isso é, 50,90% dos votos válidos para o candidato Luiz Inácio Lula da Silva e 49,10% dos votos válidos para o candidato Jair Messias Bolsonaro, conforme detalhamento descrito na Tabela 1.

Tabela 1 – Resultados das pesquisas eleitorais divulgados na véspera do segundo turno das eleições presidenciais de 2022 (29/10/2022)

Instituto	Registro da pesquisa eleitoral no Tribunal Superior Eleitoral	Resultado da pesquisa (% dos votos válidos)	
		Lula	Bolsonaro
Ipec	BR-05656/2022	54,00%	46,00%
Datafolha	BR-08297/2022	52,00%	48,00%
Quaest	BR-05765/2022	52,00%	48,00%
Atlas	BR-04838/2022	53,40%	46,60%
Brasmarket	BR-08584/2022	46,40%	53,60%
Paraná	BR-09573/2022	50,40%	49,60%
CNT/MDA	BR-01820/2022	51,10%	48,90%

Fonte: elaborado pelos autores a partir dos dados do Estado de Minas (2022) e dos resultados oficiais da eleição presidencial de 2022.

Contudo, cabe observar que a pesquisa BR-04838/2022, desenvolvida pelo Instituto Atlas, teve sua coleta de dados realizada via web. Logo, por ter seguido um processo amostral distinto das demais, ela foi excluída do estudo de caso ora proposto, analisando-se apenas aquelas seis pesquisas restantes, ou seja: BR-05656/2022; BR-08297/2022; BR-05765/2022; BR-08584/2022; BR-09573/2022; e, BR-01820/2022.

Do ponto de vista metodológico, a adoção do método do estudo de caso se justifica quando o pesquisador não tem controle sobre eventos caracterizados como seu objeto de estudo, e ainda, quando se trata de fenômenos contemporâneos oriundos da vida real, sobre os quais são formuladas questões de pesquisa do tipo “como” e/ou “por que” (YIN, 2001).

Nesse sentido, ao considerar toda a polêmica relacionada aos resultados das pesquisas realizadas nas eleições de 2022, para as quais cogitou-se até a hipótese de se instaurar uma Comissão Parlamentar de Inquérito (CPI) para investigar a atuação dos institutos de pesquisas (CLAVERY, 2022; JOHNSON, 2022; MELO, 2022), o estudo de caso ora proposto abordou um fenômeno social contemporâneo que gerou debate por parte de diversos segmentos da sociedade brasileira mesmo após o término das referidas eleições, além do fato do objeto deste estudo poder ser analisado sob uma perspectiva eminentemente empírica.

No processo de análise de dados, inicialmente, foi identificado o erro absoluto (em percentual de votos válidos) de cada uma das pesquisas cujos dados foram utilizados na composição do estudo de caso proposto. O erro absoluto desconsidera o sinal atribuído a variação ocorrida entre o resultado da pesquisa e o resultado observado nas eleições, desprezando-se o sinal negativo.

Após a identificação do erro absoluto de cada pesquisa, foram identificadas as respectivas quantidades totais de entrevistas, as quantidades médias de entrevistas e os seus desvios-padrão, e ainda, seus Coeficientes de Variação (CV) de Pearson, por município e por estado e distrito federal. Esse conjunto de indicadores teve por finalidade permitir traçar um perfil geral acerca de como foi composto o plano amostral, municipal e estadual, das pesquisas integrantes do estudo de caso analisado.

De forma complementar, estimou-se o grau de correlação entre o erro absoluto (em percentual de votos válidos) e aquelas quantidades médias de entrevistas, desvios-padrão, CV de Pearson e totais de entrevistas. Tal procedimento teve por objetivo avaliar quais daquelas medidas de tendência central e de dispersão (médias de entrevistas, desvios-padrão e CV de Pearson) e quantidades totais de entrevistas, por município e estado, poderiam apresentar relação com o comportamento dos erros absolutos observados nas pesquisas eleitorais que compuseram o estudo de caso desta investigação científica.

Adicionalmente, foram identificados cinco estados com maior quantidade relativa (percentual estadual) de entrevistas realizadas em cada um, e ainda, foi calculado o total relativo (percentual por pesquisa) que esses cinco estados representavam comparativamente a todas as

entrevistas realizadas em cada uma das pesquisas eleitorais integrantes do estudo de caso proposto para esta investigação. Esse procedimento permitiu avaliar como as entrevistas foram distribuídas ao longo do território nacional, de acordo com a plano amostral de cada uma das pesquisas eleitorais integrantes do estudo de caso analisado.

Enquanto medida relativa de dispersão (percentual), o Coeficientes de Variação de Pearson (CV de Pearson) é calculado a partir da razão entre o desvio-padrão amostral e média amostral, multiplicada por cem, e serve para avaliar o grau de dispersão da média (MARTINS, 2010). Sendo que, um CV até 15% indica baixa dispersão, um CV entre 15% e 30% indica média dispersão, e a partir de 30% o CV indica elevada dispersão (MARTINS, 2010).

Para realizar a análise de correlação foi utilizado o coeficiente de correlação de Spearman, que é um teste não-paramétrico destinado a avaliar a força da relação entre o comportamento de duas variáveis, cuja série de dados não apresenta distribuição normal (SIEGEL; CASTELLAN JR, 2006). O coeficiente de correlação de Spearman assume valores entre -1,00 e +1,00 (um negativo e um positivo), em que: quanto mais próximo de +1,00, maior a correlação em mesmo sentido entre as variáveis analisadas; por outro lado, quanto mais próximo de -1,00, maior a correlação entre as variáveis analisadas, porém, em sentido contrário; e, quanto mais próximo de 0,00 (zero), menor a intensidade do relacionamento entre o comportamento das variáveis analisadas (FIELD, 2009).

Para interpretação do coeficiente de correlação, Martins (2010) observa que um coeficiente a partir de 0,70 pode ser considerado alto. Nesse sentido, Bisquerra, Sarriera e Martinez (2004) afirmam que um coeficiente de correlação entre 0,60 e 0,80 pode ser considerado elevado, o que corrobora com a avaliação de Martins (2010).

Dessa maneira, a presente pesquisa pode ser classificada como uma investigação científica de natureza quantitativa baseada em um estudo de caso único, cujo metodologia analítica caracterizou-se pela adoção de métodos quantitativos aplicados às ciências sociais.

4 ANÁLISE DOS DADOS E APRESENTAÇÃO DOS RESULTADOS

Inicialmente, cabe considerar que esta investigação científica não pretendeu, em momento algum, estabelecer qualquer juízo de valor acerca do processo metodológico implementado pelas pesquisas eleitorais que compuseram o estudo de caso cujos dados foram analisados. Nesse sentido, assumiu-se como ponto de partida deste processo de análise de dados o distanciamento entre o resultado de cada uma daquelas pesquisas e o resultado efetivo das eleições propriamente ditas (coluna "erro percentual absoluto", da Tabela 2). Logo, é inegável que os resultados dessas pesquisas eleitorais refletiram as observações realizadas com base nas respectivas amostras.

Ao contrário das hipóteses levantadas por Berndt (1989) acerca das origens de erros ocorridos nas estimativas realizadas com base em pesquisas eleitorais, não se considerou que o distanciamento entre o percentual dos votos válidos detectado pelas pesquisas eleitorais e o resultado real aferido após o término das eleições possa ser atribuído a aspectos metodológicos. Ou seja, levou-se em conta somente a diferença entre o percentual real e o percentual apurado pelos institutos. E, nesse sentido, observou uma amplitude desse erro absoluto, distanciamento entre previsto e real), de 4,30% (amplitude = $erro_{max} - erro_{min}$ = 4,50% - 0,20% = 4,30%), conforme os dados resumidos na Tabela 2.

Tabela 2 – Amostra dos dados utilizados para composição do estudo de caso [a]

Pesquisas	Resultado da pesquisa (% dos votos válidos)		Variação observada [b] (% real - % pesquisa)		Erro percentual absoluto
	Lula	Bolsonaro	Lula	Bolsonaro	
BR-08584/2022	46,40%	53,60%	4,50%	-4,50%	4,50%

BR-05656/2022	54,00%	46,00%	-3,10%	3,10%	3,10%
BR-08297/2022	52,00%	48,00%	-1,10%	1,10%	1,10%
BR-05765/2022	52,00%	48,00%	-1,10%	1,10%	1,10%
BR-09573/2022	50,40%	49,60%	0,50%	-0,50%	0,50%
BR-01820/2022	51,10%	48,90%	-0,20%	0,20%	0,20%

(a) Pesquisas classificadas em ordem decrescente de erro percentual absoluto.

(b) Variação calculada com base no resultado real observado ao final da referida eleição, isso é, 50,90% dos votos válidos para o candidato Luiz Inácio Lula da Silva e 49,10% dos votos válidos para o candidato Jair Messias Bolsonaro.

Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Dentre as hipóteses levantadas por Pickup e Johnston (2008) acerca dos motivos do distanciamento entre o resultado das urnas e o resultado de pesquisas eleitorais, destacam-se possíveis limitações em relação ao tamanho amostral, e ainda, a ocorrência de problemas relacionados ao processo de seleção dos elementos da amostra de uma empresa de pesquisa para outra, denotando o chamado erro de viés doméstico. Contudo, a variação observada em relação aos percentuais de erro absoluto (Tabela 2), comparativamente à variação no total de municípios participantes de cada pesquisa (Tabela 3), permitem observar que as pesquisas eleitorais integrantes do estudo de caso proposto apresentaram planos amostrais e metodologias distintos entre si, descaracterizando-se a possibilidade de algum tipo de viés, ou ainda, alguma característica metodológica comum a todas às pesquisas, de tal forma que pudesse direcionar os respectivos resultados.

Outro ponto a ser observado é o fato de que todas as pesquisas integrantes do estudo de caso em análise apresentaram CV de Pearson acima de 30%, conforme informado na Tabela 3. Essa evidência demonstra uma elevada dispersão (MARTINS, 2010) em torno da quantidade média de entrevistas aplicadas nos respectivos munícipios, sendo que, as pesquisas que mais se distanciaram do resultado efetivo da eleição foram aquelas com maior CV de Person, e, consequentemente, o inverso também é verdadeiro, isso é, aquelas pesquisas cujo CV de Person foi menor apresentaram um resultado mais próximo do resultado das urnas.

Tabela 3 – Medidas de tendência central e de dispersão dos dados do estudo de caso, totalizações e sua correlação com o erro percentual absoluto

Pesquisas e análise de correlação com o erro percentual absoluto	Entrevistas/município				Entrevistas/estado e DF			
	Média (un.)	Desvio-Padrão (un.)	CV de Pearson	Total (un.)	Média (un.)	Desvio-Padrão (un.)	CV de Pearson	Total (un.)
BR-08584/2022	4,51	15,13	335%	532	19,70	12,70	64%	27
BR-05656/2022	18,18	24,37	134%	235	8,70	12,66	145%	27
BR-08297/2022	32,84	53,56	163%	226	9,37	15,94	170%	27
BR-05765/2022	16,67	9,07	54%	120	4,62	5,63	122%	26
BR-09573/2022	14,46	15,20	105%	166	6,15	5,52	90%	27
BR-01820/2022	14,59	8,14	56%	138	5,52	5,99	109%	25
Correlação	-0,03	0,38	0,67	0,75	0,67	0,55	-0,03	0,60

Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Ainda segundo as informações contidas nas Tabelas 2 e 3, pode-se observar que quanto maior a quantidade total de municípios integrantes de cada pesquisa, maior erro percentual absoluto. Todavia, quanto menor a quantidade de municípios, mais os resultados das pesquisas se aproximaram do resultado final das eleições. Essa evidência pode ajudar a corroborar com as ponderações de Chalegre, Almeida e Passos (2006) de que nem sempre o tamanho da amostra

pode influenciar a qualidade dos seus resultados, isso se o distanciamento entre o resultado da pesquisa e o resultado efetivo da eleição fosse utilizado como único parâmetro de qualidade.

Ao proceder à análise de correlação entre o erro percentual absoluto e aquelas informações sobre médias e quantidades de municípios e estados, ambas descritas também na Tabela 3, pode-se constatar que o total de municípios integrantes da amostra de pesquisa foi a variável que apresentou maior correlação com erro percentual absoluto. E, uma vez que, as pesquisas com menor quantidade de municípios foram aquelas com menor distanciamento entre seus resultados e o resultado efetivo das urnas, pode-se inferir que a variável-chave nesse processo pode não ter sido a quantidade total de entrevistas realizadas, mas, por outro lado, os municípios e consequentemente os estados nos quais foram realizadas as respectivas entrevistas.

Nesse sentido, ao analisar a participação relativa dos 5 estados com maior quantidade de entrevistas, por pesquisa eleitoral, pode-se observar que, as pesquisas que concentraram de 45% a 55% das suas entrevistas em estados como São Paulo (SP), Minas Gerais (MG), Bahia (BA) e Paraná (PR), tiveram seus resultados mais próximos dos resultados observados nas urnas, conforme demonstram a informações contidas na Tabela 4. Sendo que, ainda segundo as informações resumidas na Tabela 4, a maior concentração das entrevistas ocorreu nos estados da região sudeste.

Tabela 4 – Análise da participação relativa dos 5 estados com maior quantidade de entrevistas

Pesquisa	1º lugar	2º lugar	3º lugar	4º lugar	5º lugar	Soma
BR-08584/2022	PA (≈10,53%)	TO (≈8,27%)	RS (≈6,95%)	MG (≈6,20%)	PR (≈6,02%)	≈37,97%
BR-05656/2022	MG (≈24,26%)	SP (≈14,89%)	RJ (≈12,34%)	RS (≈5,53%)	BA (≈5,11%)	≈62,13%
BR-08297/2022	SP (≈31,22%)	MG (≈15,42%)	RJ (≈7,51%)	BA (≈5,14%)	PR (≈4,35%)	≈63,64%
BR-05765/2022	SP (≈20,83%)	MG (≈15,00%)	MA (≈8,33%)	BA (≈6,67%)	RJ (≈5,83%)	≈56,67%
BR-09573/2022	SP (≈14,46%)	MG (≈8,43%)	RS (≈8,43%)	BA (≈7,83%)	PR (≈7,23%)	≈46,39%
BR-01820/2022	SP (≈20,29%)	MG (≈11,59%)	BA (≈8,70%)	RJ (≈6,52%)	PR (≈5,80%)	≈52,59%

Fonte: elaborado pelos autores, com base nos dados da pesquisa.

Dessa maneira, mesmo considerando a pequena diferença na votação final obtida pelos dois candidatos que concorreram ao segundo das eleições brasileiras de 2022, a análise dos dados que integraram o estudo de caso realizado nesta investigação mostra que quanto mais uniforme a quantidade média de entrevistas por município, e quanto maior a concentração desses municípios em estados da região sudeste, especialmente São Paulo (SP) e Minas Gerais (MG), juntamente com Bahia (BA), na região nordeste, e Paraná (PR), na região Sul, menor a distância entre o resultado da pesquisa e o resultado efetivo das eleições.

5 CONSIDERAÇÕES FINAIS

Ao problematizar como a definição da amostra de pesquisa pode influenciar os resultados do processo de análise de dados, esta pesquisa utilizou como base um fenômeno social contemporâneo que gerou discussões em vários setores da sociedade brasileira: as eleições presidenciais de 2022. Para tanto, utilizaram-se os dados das pesquisas de intenção de voto divulgadas por seis institutos na véspera do segundo turno do referido pleito. Cabe ressaltar que não se realizou qualquer avaliação subjetiva em relação ao processo metodológico adotado pelos institutos, partindo-se do pressuposto de que os resultados obtidos nas referidas pesquisas eleitorais foram representativos das observações feitas com base nas amostras correspondentes.

Diante dos resultados obtidos, pode-se concluir que, ao menos no que tange ao processo de amostragem analisado no presente estudo de caso, não há evidências de que os institutos de pesquisa eleitoral no Brasil incorreram em qualquer tipo de viés doméstico em suas

determinações. Isso pode ser pontuado porque as características amostrais utilizadas nas pesquisas eleitorais variaram significativamente entre os institutos, sugerindo que não há uma preferência ou inclinação política inerente à escolha da amostra.

Embora exista uma tendência natural em conjecturar que o tamanho da amostra possa influenciar significativamente a qualidade dos resultados de uma pesquisa, isso não se constitui uma verdade absoluta. No estudo de caso analisado, os resultados demonstraram que, quanto menor a quantidade total de municípios integrantes de cada pesquisa, mais os seus resultados se aproximaram do resultado efetivo das eleições. Isso sugere que a investigação do processo de amostragem demanda a associação de outras métricas para a determinação da qualidade das amostras, o que significa não utilizar o tamanho amostral como base única de avaliação.

Os achados ainda sugerem que a distribuição geográfica no processo de amostragem também parece ter interferido nos resultados das pesquisas eleitorais presidenciais de 2022. Observou-se que a seleção de cidades da região Sudeste do Brasil foi mais determinante para a precisão das pesquisas eleitorais do que a quantidade de entrevistas realizadas, uma vez que as pesquisas que concentraram as entrevistas em cidades dessa região apresentaram menor distância entre o resultado de intenção de voto e o resultado real das eleições. Esse achado fornece indícios de que, para o fenômeno em questão, a quantidade total de entrevistas realizadas de fato não foi o fator determinante para os distanciamentos observados entre os levantamentos realizados pelos institutos e o resultado efetivo do pleito, mas sim os municípios e, consequentemente, os estados em que as entrevistas foram conduzidas.

Ressalta-se que a definição de amostras é um processo complexo que requer uma série de considerações metodológicas e estatísticas para garantir a representatividade da população estudada. Nesse contexto, acredita-se que a definição das amostras não foi uma escolha arbitrária dos institutos de pesquisa, mas envolveu um processo que considerou diversos fatores para a precisão e a confiabilidade dos resultados.

Embora esta pesquisa tenha observado, dentre outros indicadores, o erro percentual absoluto das pesquisas, destaca-se que a qualidade de uma amostra não pode ser determinada somente com base na diferença entre o resultado da pesquisa de intenção de voto e o resultado real da eleição. Fatores como a técnica de amostragem utilizada, a margem de erro, a metodologia de coleta de dados, entre outros, podem influenciar nos resultados obtidos em cada uma das pesquisas e não foram considerados neste estudo. Portanto, sugere-se, para pesquisas futuras, a adoção de uma abordagem mais ampla para a avaliação do processo de amostragem, a fim de aprofundar na avaliação da representatividade dos resultados obtidos em relação à população estudada.

Referências

AKANIME, Carlos Takeo; YAMAMOTO, Roberto Katsuhiro. **Estudo dirigido de estatística descritiva**. 3. ed. rev. São Paulo: Érica, 2013.

BERNDT, Alexander. Erros nas pesquisas eleitorais. **Revista de Administração**, São Paulo, v. 24, n. 3, p. 82-88, jul./set. 1989. Disponível em: https://www.periodicos.usp.br/rausp/article/view/180259/167100. Acesso em: 17 fev. 2023.

BHATTACHARYYA, Arnab; DEY, Palash. Predicting winner and estimating margin of victory in elections using sampling. **Artificial Intelligence**, [*s. l.*], v. 296, e-article 103476, July 2021. Disponível em: https://doi.org/10.1016/j.artint.2021.103476. Acesso em: 17 fev. 2023.

BISQUERRA, Rafael; SARRIERA, Jorge Castellá; MARTINEZ, Francesc. **Introdução à estatística**: enfoque informático com o pacote estatístico SPSS. Porto Alegre: Artmed, 2004.

BOLFARINE, Heleno; BUSSAB, Wilton de Oliveira. **Elementos de amostragem**. São Paulo: Blucher, 2005.

CARMO, Carlos Roberto Souza; LIMA, Adriano Davidson de. Métodos quantitativos e pesquisa contábil: um estudo de caso relacionado a pequenas amostras de dados. **CONTABILOMETRIA -** ***Brazilian Journal of Quantitative Methods Applied to Accounting***, Monte Carmelo, v. 5, n. 1, p. 92-10,009, jan.-jun./2018. Disponível em: https://revistas.fucamp.edu.br/index.php/contabilometria/article/view/10,0025/0. Acesso em: 07 nov. 2022.

CHALEGRE, Ágata Mascarenhas; ALMEIDA, Jocely Santos Caldas; PASSOS, Rubislan Oliveira. A importância das pesquisas de opinião para a elaboração de uma estratégia eleitoral. **Diálogos & Ciência: Revista Eletrônica da Faculdade de Tecnologia e Ciências**, [*s. l.*], ano 4, n. 7, p. 1-16, jun. 2006. Disponível em: https://cdn.administradores.com.br/app/uploads/2022/01/29173029/academico_734_190226_151223.pdf. Acesso em: 03 fev. 2023

CLAVERY, Elisa. Deputados da base do governo protocolam na Câmara pedido de CPI sobre institutos de pesquisa. **TV Globo (g1)**, Brasília, Política, publicação eletrônica em 21/10/2022-20h17. Disponível em: https://g1.globo.com/politica/noticia/2022/10/21/deputados-da-base-do-governo-protocolam-na-camara-pedido-de-cpi-sobre-institutos-de-pesquisa.ghtml. Acesso em: 08 nov. 2022.

COSTA, Giovani Glaucio de Oliveira. **Curso de estatística básica:** teoria e prática. 2 ed. São Paulo: Atlas, 2015.

DOWNING, Douglas; CLARK, Jeffrey. **Estatística aplicada**. Tradução de Alfredo Alves de Faria. 2. ed. São Paulo: Saraiva, 2006.

DUARTE, Cíntia Barbosa; GOMES, Marcella Furtado de Magalhães. As mudanças jurídicas das pesquisas eleitorais no Brasil. **Cadernos da Escola do Legislativo**, Belo Horizonte, v. 12, n. 18, p. 99-126, jan./jun. 2010. Disponível em: https://dspace.almg.gov.br/handle/11037/1301. Acesso em: 30 jan. 2023.

ESTADO DE MINAS. Lula x Bolsonaro: veja o que apontam as pesquisas na véspera do 2º turno. **Estado de Minas**, Belo Horizonte, seção: Política, publicação eletrônica em 29/10/2022-19:48, atualizado 30/10/2022-09:39. Disponível em: https://www.em.com.br/app/noticia/politica/2022/10/29/interna_politica,1414014/lula-x-bolsonaro-veja-o-que-apontam-as-pesquisas-na-vespera-do-2-turno.shtml. Acesso em: 29 out. 2022.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de análise de dados**. Rio de Janeiro: Elsevier, 2017.

FIELD, Andy. **Descobrindo a estatística usando SPSS**. 2. ed. Porto Alegre: Artmed, 2009.

FONSECA, Jairo Simon da; MARTINS, Gilberto de Andrade. **Curso de estatística**. 6 ed. São Paulo: Atlas, 2012.

IGNÁCIO, Sérgio Aparecido. Importância da estatística para o processo de conhecimento e tomada de decisão. **Revista Paranaense de Desenvolvimento**, Curitiba, n. 118, p. 175-192, jan./jun. 2010. Disponível em: https://ipardes.emnuvens.com.br/revistaparanaense/article/view/89. Acesso em: 30 jan. 2023.

JOHNSON, Jones. CPI das pesquisas: senador diz ter assinaturas e deputado apresenta projeto. **Jornal do Commercio**, Recife, Politíca, notícia publicada em 06/10/2022-14:19. Disponível em: https://jc.ne10.uol.com.br/colunas/jamildo/2022/10/15094633-cpi-das-pesquisas-senador-diz-ter-assinaturas-e-deputado-apresenta-projeto-saiba-mais.html. Acesso em: 07 nov. 2022.

LAUDERDALE, Benjamin E.; BAILEY, Delia; BLUMENAU, Jack; RIVERS, Douglas. Model-based pre-election polling for national and sub-national outcomes in the US and UK. **International Journal of Forecasting**, [*s. l.*], v. 36, n. 2, p. 399-413, Abr./Jun. 2020. Disponível em: https://www.sciencedirect.com/science/article/pii/S016920701930189X. Acesso em: 16 fev. 2023.

LEVENE, Mark; FENNER, Trevor. A stochastic differential equation approach to the analysis of the 2017 and 2019 UK general election polls. **International Journal of Forecasting**, [*s. l.*], v. 37, N. 3, p. 1227-1234, Jul./Set. 2021. Disponível em: https://www.sciencedirect.com/science/article/pii/S0169207021000303. Acesso em: 16 fev. 2023.

MARTINS, Gilberto de Andrade**. Estatística geral e aplicada**. 3. ed. São Paulo: Atlas, 2010.

MELO, Eduarda. Pesquisas eleitorais: CPI das pesquisas vai ser criada pela câmara? **Jornal do Commercio**, Recife, Politíca, publicação eletrônica em 04/10/2022-9:22. Disponível em: https://jc.ne10.uol.com.br/colunas/jamildo/2022/10/15092654-cpi-das-pesquisas-vai-ser-criada-pela-camara-veja-o-que-parlamentares-disseram.html. Acesso em: 07 nov. 2022.

PEKAR, Viktor; NAJAFI, Hossein; BINNER, Jane M.; SWANSON, Riley; RICKARD, Charles; FRY, John. Voting intentions on social media and political opinion polls. **Government Information Quarterly**, [*s. l.*], v. 39, n. 4, e-article. 101658, Out. 2022. Disponível em: https://www.sciencedirect.com/science/article/pii/S0740624X21000940. Acesso em: 16 fev. 2023.

PICKUP, Mark; JOHNSTON, Richard. Campaign trial heats as election forecasts: measurement error and bias in 2004 presidential campaign polls. **International Journal of Forecasting**, [*s. l.*], v. 24, n. 2, p. 272-284, Apr./Jun. 2008. Disponível em: https://www.sciencedirect.com/science/article/pii/S0169207008000307 . Acesso em: 30 jan. 2023.

PICKUP, Mark; JOHNSTON, Richard. Campaign trial heats as electoral information: evidence from the 2004 and 2006 canadian federal elections. **Electoral Studies**, [*s. l.*], v. 26, n. 2, p. 460-476, Jun. 2007. Disponível em: https://www.sciencedirect.com/science/article/pii/S0261379407000212. Acesso em: 30 jan. 2023.

SIEGEL, Sidney; CASTELLAN JR., John. **Estatística não-paramétrica para ciências do comportamento**. 2. ed. Porto Alegre: Artmed, 2006.

SILVA, Nilza Nunes. **Amostragem probalística:** um curso introdutório. 2. ed. São Paulo: Editora da Universidade de São Paulo, 2004.

TURGEON, Mathieu; RENNÓ, Lucio. Forecasting Brazilian presidential elections: solving the *N* problem. **International Journal of Forecasting**, [*s. l.*], v. 28, n. 4, p. 804-812, Oct./Dez. 2012. Disponível em: https://www.sciencedirect.com/science/article/pii/S0169207012000416. Acesso em: 16 fev. 2023.

VIRGILITO, Salvatore Benito. **Estatística aplicada**. São Paulo: Saraiva, 2017.

WERKEMA, Cristina. **Inferência estatística:** como estabelecer conclusões com confiança no giro do pdca e dmaic. Rio de Janeiro: Elsevier, 2014.

YIN, Robert K.. **Estudo de caso**: planejamento e métodos. Tradução de Daniel Grassi. 2. ed.. Porto Alegre: Bookman, 2001.