



UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE MATEMÁTICA
PROGRAMA PROFISSIONAL DE MESTRADO
EM REDE NACIONAL - PROFMAT



Bruna de Oliveira Bueno Molina

Um pouco da matemática por trás da margem de erro

Maringá-PR, Brasil

2020

Bruna de Oliveira Bueno Molina

Um pouco da matemática por trás da margem
de erro

Maringá-PR, Brasil

2020

Bruna de Oliveira Bueno Molina

Um pouco da matemática por trás da margem de erro

Dissertação de Mestrado apresentada ao Departamento de Matemática da Universidade Estadual de Maringá, como parte dos requisitos exigidos para obtenção do título de Mestre pelo Programa de Mestrado Profissional em Matemática em Rede Nacional (PROFMAT).

Universidade Estadual de Maringá - UEM

Orientador: Prof. Dr. Rodrigo Martins

Coorientador: Profa. Dra. Claudete Matilde Webler Martins

Maringá-PR, Brasil

2020

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá - PR, Brasil)

M722p

Molina, Bruna de Oliveira Bueno

Um pouco da matemática por trás da margem de erro / Bruna de Oliveira Bueno Molina.
-- Maringá, PR, 2020.
53 f.figs., tabs.

Orientador: Prof. Dr. Rodrigo Martins.

Coorientadora: Profa. Dra. Claudete Matilde Webler Martins.

Dissertação (Mestrado Profissional) - Universidade Estadual de Maringá, Centro de Ciências Exatas, Departamento de Matemática, Programa de Pós-Graduação em Matemática (PROFMAT) - Mestrado Profissional, 2020.

1. Pesquisa eleitoral - Margem de erro - Brasil. 2. Significância (Estatística). 3. Eleições - Brasil - 2018. I. Martins, Rodrigo, orient. II. Martins, Claudete Matilde Webler, coorient. III. Universidade Estadual de Maringá. Centro de Ciências Exatas. Departamento de Matemática. Programa de Pós-Graduação em Matemática (PROFMAT) - Mestrado Profissional. IV. Título.

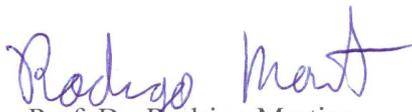
CDD 23.ed. 519.5

BRUNA DE OLIVEIRA BUENO MOLINA

UM POUCO DA MATEMÁTICA POR TRÁS DA MARGEM DE ERRO

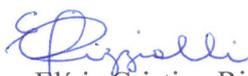
Dissertação apresentada ao Programa de Mestrado Profissional em Matemática em Rede Nacional do Departamento de Matemática, Centro de Ciências Exatas da Universidade Estadual de Maringá, como parte dos requisitos necessários para a obtenção do título de Mestre em Matemática tendo a Comissão Julgadora composta pelos membros:

COMISSÃO JULGADORA:



Prof. Dr. Rodrigo Martins

UEM - Universidade Estadual de Maringá (Orientador)



Profa. Dra. Eliris Cristina Rizzioli.

UNESP - Universidade Estadual Paulista Júlio de Mesquita Filho



Profa. Dra. Luciene Parron Gímenes Arantes

UEM - Universidade Estadual de Maringá

Aprovada em: 07 de fevereiro de 2020

Local de defesa: Auditório do DMA, Bloco F67, campus da Universidade Estadual de Maringá.

Agradecimentos

Agradeço primeiramente a Deus, que me capacitou até aqui, agradeço aos meus orientadores Claudete Matilde Webler Martins e Rodrigo Martins pela paciência e dedicação que tiveram comigo. Agradeço a meu marido Antonio, que sempre me incentivou a seguir em frente, agradeço a Minha filha Júlia, que chegou para me ensinar o que é o amor. Agradeço aos meus familiares que sempre me apoiaram. Agradeço as minhas amigas, Aline, Elaine, Nelydi, Sandra e Simone, pelos momentos que passamos juntas, agradeço a todos os professores do curso PROFMAT-UEM, que me ensinaram e inspiraram. Por fim, agradeço a todos que de alguma forma contribuíram para elaboração deste trabalho.

*“...Diga o fraco, eu sou forte!”
(Bíblia Sagrada, Joel 3; 10)*

Resumo

Neste trabalho apresentamos as informações básicas para se compreender o significado da margem de erro de uma pesquisa eleitoral. Além disso, apresentamos os resultados das pesquisas em contra-ponto ao resultado final da última eleição majoritária.

Palavras-chave: Margem de erro, significância, pesquisa eleitoral.

Abstract

In this paper we present the basic information to understand the meaning of the margin of error of an electoral survey. In addition, we present the results of the election polls and, on the other hand, the final result of the last majority election.

Keywords: margin of error, significance, election poll.

Lista de ilustrações

Figura 1 – Interpretação gráfica da função densidade de probabilidade	17
Figura 2 – Função densidade de probabilidade de $f(x) = 2x$, para $0 \leq x \leq 1$	18
Figura 3 – Gráfico da função de Distribuição F	19
Figura 4 – Distribuição Normal	22
Figura 5 – Áreas sob a Curva Normal Padrão de 0 a z	23
Figura 6 – Área sob a curva normal entre $z = 0$ e $z = 1, 2$	24
Figura 7 – Resultado das eleições presidenciais (1º turno)	32
Figura 8 – Pesquisa para presidente primeiro turno (boca de urna)	33
Figura 9 – Evolução de intenção de votos válidos no primeiro turno de 2018	34
Figura 10 – Bolsonaro eleito presidente no 2º turno	35
Figura 11 – Resultado das eleições presidenciais 2º turno	36
Figura 12 – Pesquisas presidente no 2º turno	37
Figura 13 – Resultado das eleições para governador no estado do Paraná	38
Figura 14 – Pesquisa de boca de urna para governador do estado do Paraná	39
Figura 15 – Pesquisa eleitoral para governador do estado do Paraná	40
Figura 16 – Resultado das eleições para senado estado de Minas Gerais no primeiro turno	41
Figura 17 – Pesquisas eleitorais para o senado de Minas Gerais no primeiro turno	42
Figura 18 – Pesquisas eleitorais para o senado de Minas Gerais no primeiro turno	43
Figura 19 – Descrição das pesquisas Data Folha para senado de Minas Gerais	44
Figura 20 – Pesquisa IBOPE para senado de Minas Gerais	46
Figura 21 – Pesquisa Boca de Urna para senado de Minas Gerais	47
Figura 22 – Pesquisa Boca de Urna para senado de Minas Gerais	48

Sumário

	INTRODUÇÃO	10
1	INTEGRAIS IMPRÓPRIAS	12
2	TEORIA DA MARGEM DE ERRO	15
3	UM EXEMPLO DE MARGEM DE ERRO	31
4	CONCLUSÃO	49
	REFERÊNCIAS	52

Introdução

Podemos dizer que os primeiros sinais da estatística deram início no século XVII, em estudos sobre a taxa de mortalidade, onde se fazia o uso de dados de nascimentos, casamentos, migrações, entre outros. A coleta dos dados era feita pelas paróquias, dando origem ao que chamamos hoje de censo demográfico.

Com a ideia de se representar em uma miniatura a população, Andres Nicolas Kiaer, com seu método de representatividade, na reunião do Instituto Internacional de Estatística em 1895, em Berna, Suíça, foi o primeiro a apresentar o método de amostragem. As contribuições de Ronald Aylmer Fisher à estatística começaram com o artigo (Fisher, 1912) no qual empregou o método da máxima verossimilhança no ajustamento de curvas de frequência ¹, apresentou também estudos sobre distribuições amostrais, desenvolveu os testes de significância, amostragem estratificada, entre outros.

O professor Arthur Lyon Bowley (1869-1957) considerou a amostra representativa de dois métodos: seleção intencional ou ao acaso (amostra aleatória).

Em 1925, os estatísticos Jerzy Neyman e Egon Sharpe Pearson deram início a teoria do Teste de Hipótese, onde temos duas hipóteses: a nula, é a que será testada, e a alternativa, é aquela que será aceita caso o teste indique que a hipótese nula deva ser rejeitada. Levando aos erros, do tipo I, também chamado de nível de significância, que é a probabilidade de se rejeitar a hipótese nula quando ela é verdadeira. E o erro do tipo II, que é a probabilidade de não rejeitar a hipótese nula quando esta for falsa.

Corrado Gini (1884 – 1965) e seu assistente Luigi Galvani, fizeram outra aplicação da amostragem representativa. Para obter uma amostra dos dados do censo de 1921, eles decidiram retirar os dados de 29 das 214 unidades administrativas em que a Itália estava dividida. Nas 29 unidades continham seus valores médios de sete características, próximas as médias do país inteiro. Mas quando consideradas outras características, Gini e Galvani encontraram grandes diferenças entre os dados da amostra e do país. Concluindo que os controles eram ineficientes. Sob este cenário foi publicado o famoso artigo de Neyman (1934), que contou com a presença de Bowley, Fisher e E. S. Pearson, onde Neyman mostra a superioridade da amostragem aleatória estratificada sobre a seleção intencional.

Após 1935, na apresentação do famoso artigo de Neyman, intitulado "On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection"(Neyman, 1934), juntamente com outros trabalhos posteriores, garantiram a Neyman o reconhecimento de sua preeminência na teoria dos

¹ Uma amostra verossímil é uma amostra que fornece a melhor informação possível sobre um parâmetro de interesse da população

levantamentos por amostragem, proporcionando, posteriormente, o seu desenvolvimento em diversos países.

Influenciada pelo uso de computadores, atualmente a estatística se apresenta muito matematizada, dificultando sua leitura, muitas vezes, pelos estatísticos. Mas as máquinas de contar já estavam presentes no desenvolvimento teórico da estatística com Fisher e Neyman nas décadas de 40 e 50.

Em uma pesquisa estatística é, na maioria das vezes, inviável que se tome toda a população, muitas vezes se tornando impossível, como no caso de se analisar doenças no sangue de uma pessoa, teríamos que retirar todo o sangue do paciente, o levando a óbito. Para estes casos fazemos uso da amostra aleatória que, no exemplo acima citado, se trata de apenas alguns mililitros de sangue do paciente para que se possa inferir um resultado, se possui ou não alguma doença. Para garantir uma melhor representação da população fazemos uso da margem de erro, que nada mais é do que um limite, para mais ou para menos, que os resultados podem variar. Afirmando que o resultado de uma pesquisa não é uma certeza absoluta, e que pode variar dentro da margem de erro. Assim como pode também não representar uma verdade.

A amostragem aleatória faz parte de nossas vidas, em diversos meios podemos nos deparar com o uso de amostras para se analisar a população em questão. Sem nenhuma intenção de sermos partidários, trataremos das pesquisas eleitorais de 2018 para presidente do Brasil, governador do Paraná e senador de Minas Gerais. Em pesquisas eleitorais apenas uma pequena parcela dos eleitores é entrevistada, e se a amostra for bem distribuída e sem nenhuma manipulação, os dados desta amostra podem prever o resultado de uma disputa eleitoral, apenas uma parte da população representa a intenção de todo o país. Deixando claro que não tivemos nenhum viés político na elaboração deste trabalho, iremos comparar as pesquisas estatísticas e o resultado das eleições de 2018, falando sobre os acertos e os erros das mesmas.

Esta dissertação está dividida em três capítulos. No primeiro capítulo apresentamos alguns resultados preliminares sobre integrais impróprias, no segundo capítulo incluímos as teorias da margem de erro, algumas definições, exemplos, teoremas, que nos ajudam a entender a teoria matemática e estatística existente por trás de uma pesquisa com margem de erro. No terceiro capítulo abordamos um exemplo de margem de erro que, neste caso, trouxemos alguns resultados de pesquisas eleitorais das eleições de 2018.

1 Integrais Impróprias

Sabemos que a Matemática e a Estatística andam juntas, por este motivo, vamos iniciar este trabalho com algumas definições matemáticas. Estamos considerando que o leitor já possui os conhecimentos prévios até Funções Integráveis e, vamos definir a seguir Integrais impróprias, baseados no texto [10].

Definição 1.1. *Integrais com limites infinitos de integração são chamadas integrais impróprias do tipo I.*

1. Se $f(x)$ é contínua em $[a, \infty)$, então

$$\int_a^{\infty} f(x)dx = \lim_{b \rightarrow \infty} \int_a^b f(x)dx.$$

2. Se $f(x)$ é contínua em $(-\infty, b]$, então

$$\int_{-\infty}^b f(x)dx = \lim_{a \rightarrow -\infty} \int_a^b f(x)dx.$$

3. Se $f(x)$ é contínua em $(-\infty, \infty)$, então

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^c f(x)dx + \int_c^{\infty} f(x)dx,$$

onde c é qualquer número real.

Em todos os casos, se o limite é finito, dizemos que o limite da integral imprópria converge e que o limite é o valor da integral imprópria. Se o limite não existe, dizemos que a integral imprópria diverge.

Exemplo 1.2. *Vamos calcular $\int_1^{+\infty} \frac{1}{x^2} dx$. Notemos que, pela Definição 1.1,*

$$\int_1^{+\infty} \frac{1}{x^2} dx = \lim_{b \rightarrow +\infty} \int_1^b \frac{1}{x^2} dx.$$

Como

$$\int_1^b \frac{1}{x^2} dx = \left[-\frac{1}{x} \right]_1^b = -\frac{1}{b} + 1,$$

resulta que

$$\int_1^{+\infty} \frac{1}{x^2} dx = \lim_{b \rightarrow +\infty} \left[-\frac{1}{b} + 1 \right] = 1.$$

Como

$$\int_1^{+\infty} \frac{1}{x^2} dx = 1,$$

a integral imprópria dada acima é convergente.

Exemplo 1.3. A integral imprópria $\int_1^{+\infty} \frac{1}{x} dx$ é divergente. Vejamos que,

$$\int_1^b \frac{1}{x} dx = [\ln x]_1^b = \ln b.$$

Assim,

$$\int_1^{+\infty} \frac{1}{x} dx = \lim_{b \rightarrow +\infty} \ln b = +\infty.$$

Definição 1.4. Integrais de funções que se tornam infinitas em um ponto dentro do intervalo de integração são integrais impróprias do tipo II.

1. Se $f(x)$ é contínua em $(a, b]$ e descontínua em a , então

$$\int_a^b f(x) dx = \lim_{c \rightarrow a^+} \int_c^b f(x) dx.$$

2. Se $f(x)$ é contínua em $[a, b)$ e descontínua em b então

$$\int_a^b f(x) dx = \lim_{c \rightarrow b^-} \int_a^c f(x) dx.$$

3. Se $f(x)$ é descontínua em c onde $a < c < b$ então

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx.$$

Em todos os casos, se o limite é finito, dizemos que a integral imprópria converge e que o limite é o valor da integral imprópria. Se o limite não existe, dizemos que a integral imprópria diverge.

Exemplo 1.5. Vamos calcular $\int_0^1 \frac{1}{\sqrt{x}} dx$. Temos $f(x) = \frac{1}{\sqrt{x}}$ é não limitada em $(0, 1]$ e integrável em $[a, 1]$, para $0 < a < 1$; de acordo com a definição anterior,

$$\int_0^1 \frac{1}{\sqrt{x}} dx = \lim_{a \rightarrow 0^+} \int_a^1 \frac{1}{\sqrt{x}} dx = \lim_{a \rightarrow 0^+} [2 - 2\sqrt{a}] = 2$$

ou seja, $\int_0^1 \frac{1}{\sqrt{x}} dx = 2$,

Quando não podemos resolver uma integral imprópria diretamente, tentamos determinar se ela é convergente ou divergente. Se ela converge, podemos então utilizar métodos numéricos para ter seu valor aproximado. Os principais testes para convergência ou divergência são o teste de comparação direta e o teste de comparação por limite.

Teorema 1.6. (Teste de comparação) Sejam f e g contínuas em $[a, \infty)$ com $0 \leq f(x) \leq g(x)$ para qualquer $x \geq a$. Então

1. $\int_a^\infty f(x) dx$ converge se $\int_a^\infty g(x) dx$ converge.

2. $\int_a^\infty g(x) dx$ diverge se $\int_a^\infty f(x) dx$ diverge.

Exemplo 1.7. $\int_1^\infty \frac{\operatorname{sen}^2 x}{x^2} dx$ converge pois $0 \leq \frac{\operatorname{sen}^2 x}{x^2} \leq \frac{1}{x^2}$ em $[1, \infty)$ e $\int_1^\infty \frac{1}{x^2} dx$ converge.

Exemplo 1.8. $\int_1^\infty \frac{1}{\sqrt{x^2 - 0,1}} dx$ diverge pois $\frac{1}{\sqrt{x^2 - 0,1}} \geq \frac{1}{\sqrt{x^2}} = \frac{1}{x}$ em $[1, \infty)$ e $\int_1^\infty \frac{1}{x} dx$ diverge.

Teorema 1.9. (Teste de comparação no limite) Se as funções positivas f e g são contínuas em $[a, \infty)$ e se

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = L, \quad 0 < L < \infty$$

, então

$$\int_a^\infty f(x) dx \quad \text{e} \quad \int_a^\infty g(x) dx$$

são ambas convergentes ou ambas divergentes.

Exemplo 1.10. Vamos mostrar que $\int_1^\infty \frac{3}{e^x - 2} dx$ converge. As funções $f(x) = \frac{1}{e^x}$ e $g(x) = \frac{3}{e^x - 2}$ são positivas e contínuas em $[1, \infty)$ e $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow \infty} \frac{\frac{1}{e^x}}{\frac{3}{e^x - 2}} = \lim_{x \rightarrow \infty} \frac{e^x - 2}{3e^x} = \lim_{x \rightarrow \infty} \frac{1}{3} - \frac{2}{3e^x} = \frac{1}{3}$. Portanto, como a integral $\int_1^\infty \frac{1}{e^x} dx = \int_1^\infty e^{-x} dx$ converge, pelo Teorema 1.9, $\int_1^\infty \frac{3}{e^x - 2} dx$ também converge.

2 Teoria da margem de erro

Neste capítulo apresentamos os conceitos básicos que envolvem a significância estatística. Fizemos um compilado do conteúdo teórico que envolve o tema, baseado nos textos [1], [3], [10], [13], [14], [15], [19] e [22].

Iniciamos apresentando alguns conceitos estatísticos que serão utilizados posteriormente.

Definição 2.1. *Chamamos de população o conjunto de todos os elementos, ou resultados, sob investigação. E de amostra, um subconjunto desta população. Em geral, usamos n para indicar o número de elementos da amostra, isto é, o seu tamanho.*

A seguir, apresentamos um exemplo.

Exemplo 2.2. *Numa eleição estamos interessados apenas nas pessoas que de fato votam, assim a população é formada pelos eleitores. Por outro, lado numa pesquisa eleitoral é inviável perguntar sua escolha a todos os eleitores, assim perguntamos à apenas um subgrupo dos eleitores, neste caso, uma amostra de tamanho n eleitores.*

Podemos obter os dados amostrais através de levantamentos amostrais (probabilísticos ou não probabilísticos), planejamento de experimentos ou levantamentos observacionais. Vamos descrever essas técnicas abaixo nos itens 1, 2 e 3, a seguir.

1-Levantamentos amostrais: a amostra é obtida de uma população bem definida, por meio de processos bem protocolados e controlados pelo pesquisador. Pode ser subdividido em dois grupos: levantamentos probabilísticos, onde cada elemento da população tem uma probabilidade conhecida e diferente de zero de ser selecionado para compor a amostra. Um sorteio aleatório para determinar a amostra, é um exemplo. E levantamentos não probabilísticos: são aqueles em que a seleção dos elementos da população para compor a amostra depende de julgamentos pré-definidos. Não existe um controle estatístico, portanto, não se pode garantir o cálculo da margem de erro. Uma pesquisa feita apenas com mulheres com mais de 30 anos e ensino superior completo, é um exemplo de levantamentos não probabilísticos, ou ainda, uma amostra de voluntários para o uso de determinado cosmético.

2- Planejamento de experimentos: o objetivo é analisar o efeito de uma variável sobre a outra, tem a interferência do pesquisador com o intuito de medir o efeito desejável.

3- Levantamentos observacionais: os dados são coletados sem interferência do pesquisador.

Exemplo 2.3. *Um exemplo de planejamento de experimentos são os testes de eficiência de medicamentos, usados por uma determinada amostra de ratos, em um determinado ambiente, com controle de fatores externos.*

Exemplo 2.4. *O número de acidentes em determinado feriado na cidade de Maringá, constitui um levantamento observacional.*

A seguir, vamos definir o que são espaço amostral S , variável aleatória e variável aleatória discreta.

Definição 2.5. *Consideremos um experimento qualquer, e seja S o espaço amostral associado a tal experimento, ou seja, S é o conjunto de todos os possíveis resultados de tal experimento. Suponhamos, agora, que a cada resultado possível de tal experimento seja associado um número X . Pois bem, a variável X obtida dessa forma denomina-se variável aleatória. Se o conjunto de todos os valores de X for finito ou enumerável, dizemos que X é uma variável aleatória discreta.*

Exemplo 2.6. *Para exemplificar a Definição 2.7, vamos considerar o experimento que consiste em lançar uma moeda. Neste caso, o espaço amostral é o conjunto $\{\text{cara; coroa}\}$. Se ao associarmos o número 0 ao resultado cara e ao coroa o 1, a variável aleatória X poderá assumir qualquer valor do conjunto finito $\{0, 1\}$, e X será então uma variável aleatória discreta. Supondo a moeda honesta, a probabilidade $p(x)$ de cada valor x de X é $\frac{1}{2}$, ou seja, $p(0) = \frac{1}{2}$ e $p(1) = \frac{1}{2}$; é usual a notação $P(X = x)$ para representar a probabilidade de a variável aleatória X ser igual a x : $P(X = x) = p(x)$. Observemos que $p(0) + p(1) = 1$.*

Vamos definir, agora, função densidade de probabilidade.

Definição 2.7. *Seja uma função definida para todo x real e integrável em todo o intervalo $[a, b]$, com a e b reais e $a < b$. Dizemos que f é uma função densidade de probabilidade se as seguintes condições estiverem satisfeitas:*

$$a) f(x) \geq 0; \quad (2.1)$$

$$b) \int_{-\infty}^{+\infty} f(x)dx = 1. \quad (2.2)$$

Vejamos um exemplo de função densidade de probabilidade.

Exemplo 2.8. *Sejam $a < b$ dois reais quaisquer e f a função dada por*

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{se } a \leq x \leq b \\ 0, & \text{se } x < a \quad \text{ou } x > b \end{cases}$$

De $b > a$, segue que $f(x) \geq 0$, para todo $x \in \mathbb{R}$. Por outro lado,

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{+\infty} \frac{1}{b-a}dx = 1.$$

Logo a função dada é uma função densidade de probabilidade.

A seguir, definimos probabilidade de uma variável aleatória que não é discreta mas que admite função densidade de probabilidade e, em seguida, damos um exemplo.

Definição 2.9. *Sejam X uma variável aleatória e f uma função densidade de probabilidade. Dizemos que a variável aleatória X tem densidade de probabilidade f se a probabilidade de X pertencer ao intervalo (a, b) , com $a < b$ quaisquer ($a = -\infty$ ou $b = +\infty$), for dada por*

$$P(a < X < b) = \int_a^b f(x)dx, \quad (2.3)$$

respectivamente,

$$P(-\infty < X < b) = P(X < b) = \int_{-\infty}^b f(x)dx,$$

ou

$$P(a < X < +\infty) = P(X > a) = \int_a^{+\infty} f(x)dx.$$

A seguir, apresentamos o gráfico de $y = f(x)$ que representa a função de densidade de probabilidade.

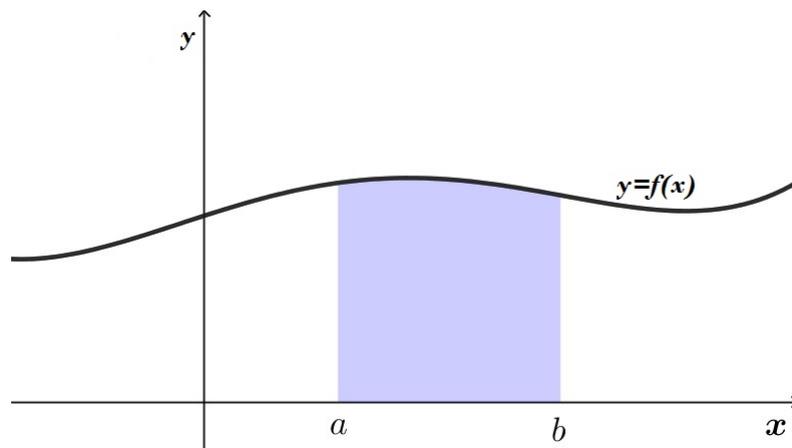


Figura 1 – Interpretação gráfica da função densidade de probabilidade

Fonte: O autor

Desse modo, a probabilidade de X estar entre a e b nada mais é do que a área da região limitada pelo gráfico de $y = f(x)$, pelas retas $x = a$, $x = b$ e pelo eixo x . De $\int_{-\infty}^{+\infty} f(x)dx = 1$ e $f(x) \geq 0$ para todo x , resulta que a probabilidade de a variável aleatória X pertencer ao intervalo (a, b) é tal que $0 \leq P(a \leq X \leq b) \leq 1$. Observemos que $\int_{-\infty}^{\infty} f(x)dx$ é um valor aproximado para a probabilidade de a variável aleatória X estar compreendida entre x e $x + dx$.

Exemplo 2.10. *Se $f(x) = 2x$, para $0 \leq x \leq 1$, e zero fora desse intervalo, vemos que $f(x) \geq 0$, para qualquer $x \in \mathbb{R}$, e a área sob o gráfico de $f(x)$ é unitária. Logo, a função f pode representar a função densidade de uma variável aleatória contínua X .*

E $P(0 \leq X \leq \frac{1}{2})$ é igual a área do triângulo de base $\frac{1}{2}$ e altura 1 (conforme Figura 2), ou seja,

$$P(0 \leq X \leq \frac{1}{2}) = \int_0^{\frac{1}{2}} 2x dx = x^2 \Big|_0^{\frac{1}{2}} = \left(\frac{1}{2}\right)^2 = \frac{1}{4} = \frac{1}{2} \left(\frac{1}{2} \cdot 1\right) = \frac{1}{4}.$$

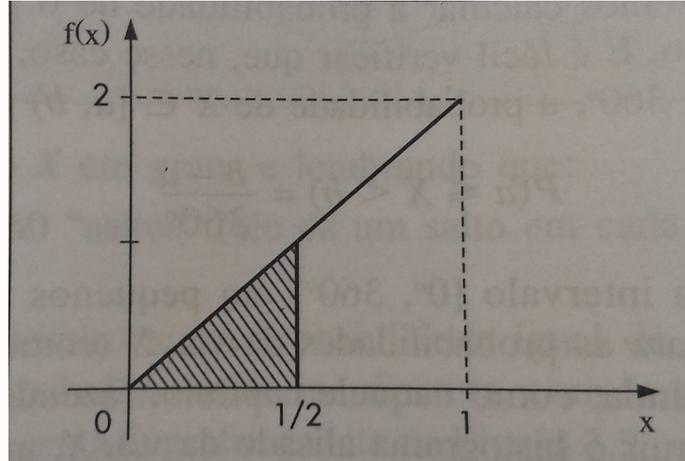


Figura 2 – Função densidade de probabilidade de $f(x) = 2x$, para $0 \leq x \leq 1$

Fonte: Morettin, Pedro Alberto (2010, p.166)

A seguir, definimos variável aleatória contínua.

Definição 2.11. Dizemos que uma variável aleatória X é contínua se, para todo a real, a probabilidade de $X = a$ for zero. Se X é uma variável aleatória que admite função densidade de probabilidade f , então X será uma variável aleatória contínua, pois para todo a real $P(X = a) = \int_a^a f(x)dx = 0$.

Definimos função de distribuição e, em seguida, apresentamos um exemplo.

Definição 2.12. Seja X uma variável aleatória. A função dada por $F(x) = P(X \leq x)$, com x real, é denominada função de distribuição da variável aleatória X . Se X for uma variável aleatória contínua, com densidade de probabilidade f , temos

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt, \quad (2.4)$$

para todo x real.

Se X for uma variável aleatória contínua com função densidade de probabilidade f , então a sua função de distribuição F é uma função contínua e $F'(x) = f(x)$ em todo x em que f for contínua. E ainda, a probabilidade da variável aleatória X pertencer ao intervalo $[a, b]$ é

$$P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x)dx. \quad (2.5)$$

Observemos que, se F for uma função de distribuição, temos, necessariamente,
 $\lim_{x \rightarrow +\infty} F(x) = 1$ e $\lim_{x \rightarrow -\infty} F(x) = 0$.

Exemplo 2.13. Consideremos a função densidade de probabilidade dada por $f(x) = \frac{1}{x^2}$, se $x \geq 1$ e $f(x) = 0$, se $x < 1$. Temos que $F(x) = \int_{-\infty}^x f(t)dt$, segue que $F(x) = 0$, se $x \leq 1$ e $F(x) = \int_1^x \frac{1}{t^2}dt$, se $x > 1$, ou seja,

$$F(x) = \begin{cases} 0 & \text{se } x \leq 1 \\ 1 - \frac{1}{x} & \text{se } x > 1. \end{cases}$$

Podemos observar, na Figura 3, que $\lim_{x \rightarrow +\infty} F(x) = 1$.

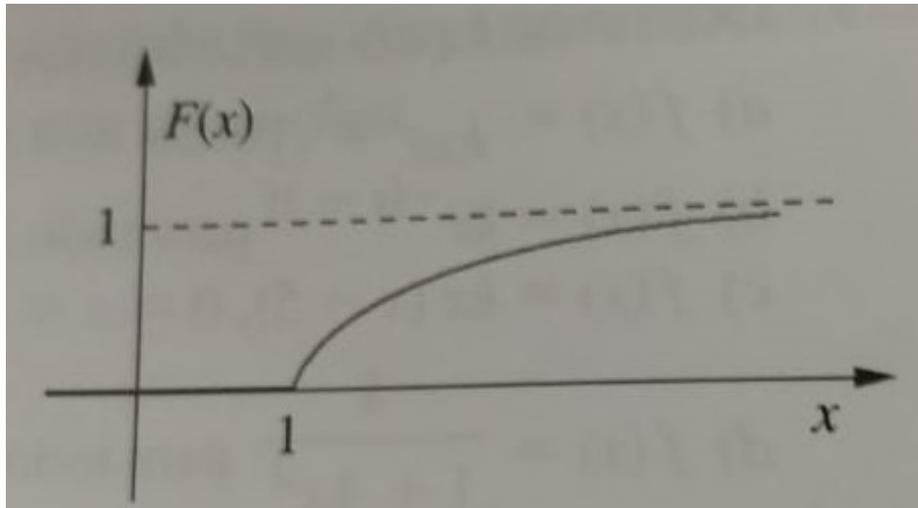


Figura 3 – Gráfico da função de Distribuição F

Fonte: Guidorizzi, Hamilton Luiz (2011, p50)

A seguir, definimos o que é uma amostra aleatória.

Definição 2.14. Uma amostra aleatória de tamanho n de uma variável aleatória X é um conjunto de n variáveis aleatórias independentes X_1, \dots, X_n , cada uma com a mesma distribuição de X .

Podemos ter amostragem aleatória simples com reposição, se for permitido que a unidade possa ser sorteada mais de uma vez. E sem reposição, se a unidade sorteada for removida da população.

Definição 2.15. Uma estatística é uma característica da amostra, ou seja, uma estatística T é uma função de X_1, \dots, X_n . Estatística é uma característica numérica da amostra.

Vamos, agora, definir as estatísticas mais comuns.

Considerando uma coleção de n números reais em que o número x_1 aparece repetido n_1 vezes, x_2 aparece n_2 vezes, ..., x_k aparece n_k vezes, de tal modo que $\sum_{i=1}^k n_i = n$; pois bem, a *média aritmética* \bar{x} desses números é dada por

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \sum_{i=1}^k n_i f_i, \text{ onde } f_i = \frac{n_i}{n}. \quad (2.6)$$

A distância entre os números x_i e \bar{x} é $|x_i - \bar{x}|$; assim, o quadrado da distância de x_i a \bar{x} é $(x_i - \bar{x})^2$. A média aritmética dos quadrados das distâncias de x_i a \bar{x} , i de 1 a k , é por definição, a *variância* de tais números:

$$\text{variância} = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 n_i = \sum_{i=1}^k (x_i - \bar{x})^2 f_i. \quad (2.7)$$

A raiz quadrada da variância denomina-se *desvio padrão* de tais números:

$$\text{desvio padrão} = \sqrt{\sum_{i=1}^k (x_i - \bar{x})^2 f_i}. \quad (2.8)$$

Observemos que quanto maior o desvio padrão, mais afastados estarão os números x_i da média \bar{x} , e, quanto menor o desvio padrão, mais concentrados em torno da média \bar{x} estarão os números x_i .

Consideremos, agora, uma variável aleatória discreta X com possíveis valores $x_1, x_2, x_3, \dots, x_k$ e probabilidades $p(x_1), p(x_2), \dots, p(x_k)$, respectivamente, o *valor esperado* ou *média de X* , que se indica por $E(X)$ ou, simplesmente, por μ é

$$E(X) = \sum_{i=1}^k x_i p(x_i). \quad (2.9)$$

Por outro lado, a *variância de X* , que se indica por $Var(X)$ ou, simplesmente, por σ^2 , $\sigma > 0$, é por definição, dada por

$$Var(X) = \sum_{i=1}^k (x_i - E(X))^2 p(x_i). \quad (2.10)$$

Observemos que se $p(x_i) = \frac{n_i}{n}$, para i de 1 a k , o valor esperado de $E(X)$ nada mais é do que a média \bar{x} , e $Var(X)$ nada mais é do que a variância dos números $x_1, x_2, x_3, \dots, x_k$, onde x_i aparece repetido n_i vezes e $\sum_{i=1}^k n_i = n$.

A raiz quadrada de $Var(X)$ é o *desvio padrão* σ da variável aleatória X

$$\sigma = \sqrt{Var(X)}. \quad (2.11)$$

O maior valor da amostra é definido por

$$X_{(k)} = \max(X_1, \dots, X_k). \quad (2.12)$$

O menor valor da amostra por

$$X_{(1)} = \min(X_1, \dots, X_k). \quad (2.13)$$

A amplitude amostral é dada por

$$W = X_{(k)} - X_{(1)}. \quad (2.14)$$

A i -ésima maior observação da amostra é dada por

$$X_{(i)} = a. \quad (2.15)$$

Exemplo 2.16. Dada uma amostra de tamanho 5, com os valores amostrais 7, 9, 1, 6, 2, então, ordenando a amostra, temos 1, 2, 6, 7, 9.

A média amostral é $\bar{X} = \frac{7+9+1+6+2}{5} = 5$.

A variância da amostra é $S^2 = \frac{(1-5)^2 + (2-5)^2 + (6-5)^2 + (7-5)^2 + (9-5)^2}{5-1} = \frac{46}{4} = 11,5$.

O maior valor da amostra é $X_{(5)} = 9$.

O menor valor da amostra é $X_{(1)} = 1$.

A amplitude total da amostra é $W = X_{(5)} - X_{(1)} = 9 - 1 = 8$.

Observando que, para dx suficientemente pequeno, $f(x)dx$ é praticamente a probabilidade de ocorrência de x , portanto, as seguintes definições de valor esperado e variância para uma variável aleatória contínua são dadas.

Definição 2.17. Seja X uma variável aleatória contínua X , com função densidade de probabilidade f . Definimos o valor esperado $E(X)$ de X por

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx \quad (2.16)$$

e a variância $Var(X)$ de X por

$$Var(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)dx. \quad (2.17)$$

Definição 2.18. Um parâmetro é uma medida usada para descrever uma característica da população.

Se retirarmos amostras de uma população, identificada pela variável aleatória X , seriam parâmetros a média μ e sua variância $\text{Var}(X)$, por exemplo.

A seguir, temos a definição da distribuição normal, que é a distribuição de probabilidades mais importante da estatística.

Definição 2.19. Dizemos que a variável aleatória contínua X tem distribuição normal, com média μ e variância σ^2 , $\sigma > 0$, se a sua função de densidade de probabilidade for dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}. \quad (2.18)$$

A notação $X \sim N(\mu, \sigma^2)$ é usada para indicar que a variável aleatória X tem distribuição normal, com média μ e variância σ^2 (ou desvio padrão σ). O gráfico de $f(x)$ tem a seguinte forma, veja Figura 4.

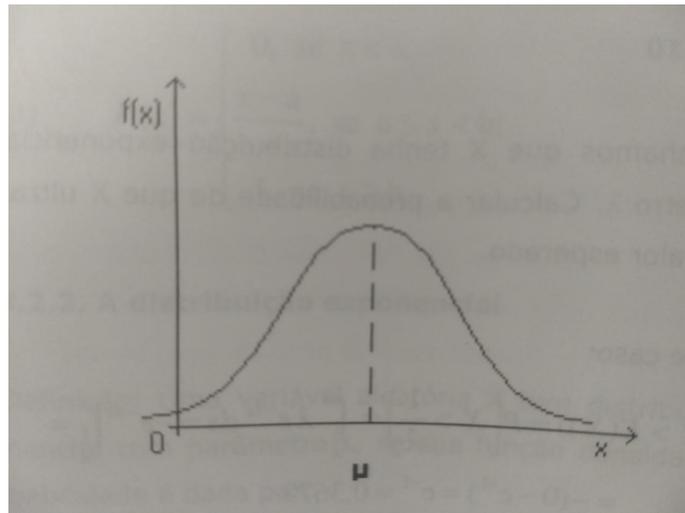


Figura 4 – Distribuição Normal

Fonte: Azevedo, Paulo Roberto Medeiros (2005.p88)

Temos ainda que $X = \mu$ é o ponto de máximo de $f(x)$; $f(x)$ tende para zero quando x tende para mais ou menos infinito; $f(x)$ é simétrica ao redor de $x = \mu$, isto é, $f(\mu + x) = f(\mu - x)$, entre os pontos $\mu - 3\sigma$ e $\mu + 3\sigma$ a área sob o gráfico de $f(x)$ é igual a 99,74%, ou seja, entre estes pontos está praticamente toda área sob o gráfico de $f(x)$; e se $X \sim N(\mu, \sigma^2)$, mostra-se que $E(X) = \mu$ e $V(X) = \sigma^2$.

Consideremos agora a tabela dada na Figura 5. Esta tabela nos dá a área sob a curva normal padrão de 0 a z . Para encontrar o valor da área tomamos a diferença entre o intervalo desejado ($z - 0$), no eixo x . Na tabela, descemos pela coluna denominada z , até a primeira casa decimal do valor obtido pela diferença do intervalo, ($z - 0$), em seguida

vamos até o valor da segunda casa decimal, obtida na diferença ($z - 0$), que será uma das dez colunas seguintes, enumeradas de 0 a 9. a interseção da linha obtida em z com a coluna, de 0 a 9, será o valor correspondente a área da região sob a curva $y = f(x)$.

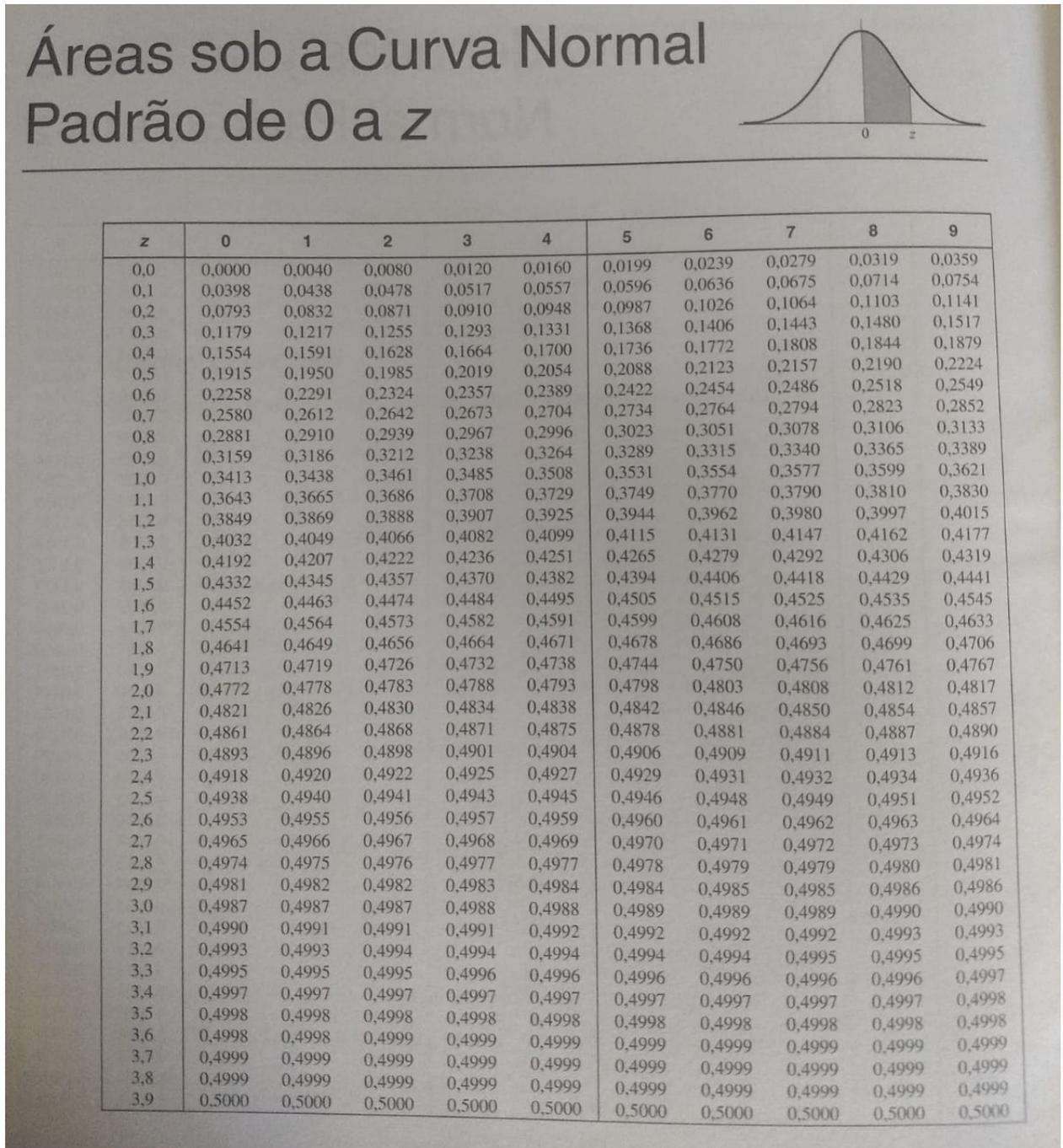


Figura 5 – Áreas sob a Curva Normal Padrão de 0 a z

Fonte: Spiegel, M.R., Schiller, J., Srinivasan, A (2013.p.414)

Vejamos um exemplo.

Exemplo 2.20. Para encontrar a área sob a curva normal padrão entre $z = 0$ e $z = 1,2$.

Usaremos a tabela de Áreas sob a Curva Normal Padrão de 0 a z (Figura 5). Para isso desça pela coluna assinalada z até alcançar o valor 1,2. Então prossiga à direita na coluna assinalada com um zero. O resultado, 0,3849, é a área solicitada e representa a probabilidade de que Z esteja entre 0 e 1,2. Portanto,

$$P(0 \leq Z \leq 1,2) = \int_0^{1,2} e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} = 0,3849$$

Vejamos, graficamente, o exemplo acima. A área sombreada correspondia a 0,3849.

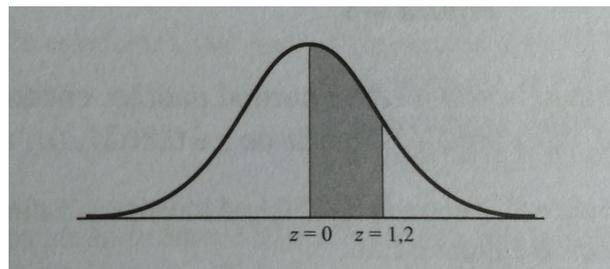


Figura 6 – Área sob a curva normal entre $z = 0$ e $z = 1,2$

Fonte: Spiegel, M.R., Schiller, J., Srinivasan, A (2013.p.123)

A inferência Estatística é um ramo da estatística cujo objetivo é fazer afirmações a partir de um conjunto de valores representativo (amostra) sobre um universo (população). Portanto, a inferência faz uma afirmação sobre a população através de uma amostra retirada da mesma.

Teorema 2.21. *Seja X uma variável aleatória contínua, com distribuição normal, média μ e variância σ^2 , temos que*

i) $E(X) = \mu$;

ii) $Var(X) = \sigma^2$.

iii) *A distribuição de \bar{x} aproxima-se de uma distribuição normal com valor esperado μ e variância $\frac{\sigma^2}{n}$, quando n tende ao infinito.*

Demonstração: Vejamos que

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (2.19)$$

Temos

$$\int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} (x-\mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Com a mudança de variável $s = x - \mu$, temos

$$\int_{-\infty}^{+\infty} (x - \mu) e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} s \cdot e^{-\frac{s^2}{2\sigma^2}} ds = 0,$$

pois o integrando da segunda integral é uma função ímpar. Segue que

$$\int_{-\infty}^{+\infty} x e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = \mu \int_{-\infty}^{+\infty} e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx.$$

Assim,

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = \mu,$$

pois

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = 1.$$

Portanto, $E(X) = \mu$.

E, ainda,

$$Var(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x - \mu)^2 e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx.$$

Tendo em vista a simetria do gráfico do integrando em relação à reta $x = \mu$, resulta

$$Var(X) = \frac{2}{\sigma\sqrt{2\pi}} \int_{\mu}^{+\infty} (x - \mu)^2 e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx.$$

Fazendo $f(x) = x - \mu$, $g'(x) = (x - \mu)e^{-\frac{(x - \mu)^2}{2\sigma^2}}$ e, integrando por partes, vem

$$\int_{\mu}^s f(x)g'(x)dx = -\sigma^2 \left[(x - \mu)e^{-\frac{(x - \mu)^2}{2\sigma^2}} \right]_{\mu}^{\sigma} + \sigma^2 \int_{\mu}^{\sigma} se^{-\frac{(x - \mu)^2}{2\sigma^2}} dx.$$

Como $\lim_{s \rightarrow +\infty} (s - \mu)e^{-\frac{(s - \mu)^2}{2\sigma^2}} = 0$, resulta

$$Var(X) = \frac{2\sigma^2}{\sigma\sqrt{2\pi}} \int_{\mu}^{+\infty} e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = \sigma^2.$$

Assim, $Var(X) = \sigma^2$. ■

Exemplo 2.22. Consideremos a população de cinco números: 2, 3, 6, 8, 11. E todas as amostras de tamanho dois que possam ser extraídas sem reposição desta população. Então a média da população é dada por

$$\mu = \frac{2 + 3 + 6 + 8 + 11}{5} = \frac{30}{5} = 6;$$

O desvio padrão da população é dado por

$$\begin{aligned} \text{Var}(\bar{X}) &= \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5} = \\ &= \frac{16 + 9 + 0 + 4 + 25}{5} = 10,8. \end{aligned}$$

Vamos encontrar a média da distribuição amostral das médias. Para isto, notemos que existem 10 amostras de tamanho dois, extraídas sem reposição:

$$(2, 3), (2, 6), (2, 8), (2, 11), (3, 6), (3, 8), (3, 11), (6, 8), (6, 11), (8, 11).$$

As médias amostrais correspondentes são 2,5; 4,5; 6,5; 4,5; 5,5; 5,5; 7,7; 8,5; 9,5.

A média da distribuição amostral das médias é

$$\mu_{\bar{x}} = \frac{2,5 + 4 + 5 + 6,5 + 4,5 + 5,5 + 5,5 + 7 + 7 + 8,5 + 9,5}{10} = 6.$$

Corolário 2.23. Se X_1, \dots, X_n for uma amostra aleatória simples da população X com média μ e variância σ^2 finita e

$$\bar{X} = \frac{(X_1 + \dots + X_n)}{n}, \quad (2.20)$$

então

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1). \quad (2.21)$$

Chamemos de e a variável aleatória que mede a diferença entre a média da população \bar{X} e a média amostral μ , isto é,

$$e = \bar{X} - \mu, \quad (2.22)$$

e é chamado erro amostral da média.

Corolário 2.24. A distribuição de e aproxima-se de uma distribuição normal com média 0 e variância $\frac{\sigma^2}{n}$, isto é,

$$\frac{\sqrt{ne}}{\sigma} \sim N(0, 1). \quad (2.23)$$

O Teorema 2.23 afirma que \bar{X} aproxima-se de uma normal quando n tende para o infinito.

Podemos determinar o tamanho da amostra a ser trabalhada de modo a obtermos um erro previamente estipulado, com determinado grau de confiança, usando

$$n = \left(\frac{Z_\gamma \sigma}{\varepsilon} \right)^2, \quad (2.24)$$

sendo, n o número de indivíduos na amostra, ε a margem de erro, Z_γ o valor crítico que corresponde ao grau de confiança e σ é o desvio padrão populacional.

Ou ainda, podemos usar a proporção populacional(p)

$$n = \frac{Z_\gamma^2 p(1-p)}{\varepsilon^2}, \quad (2.25)$$

p é a proporção de indivíduos que pertence a categoria que estamos interessados em estudar e $(1-p)$ é a proporção de indivíduos que não pertence a categoria que estamos interessados em estudar.

A seguir definiremos estimação por ponto

Definição 2.25. *Sejam X_1, \dots, X_n uma amostra de uma variável aleatória X e θ um parâmetro desconhecido da distribuição de X . Um estimador pontual de θ é definido como sendo qualquer função de X_1, \dots, X_n .*

Definição 2.26. *(Consistência) Considerando a amostra X_1, \dots, X_n de uma variável aleatória X e θ um parâmetro da sua distribuição, temos que um estimador T de θ , baseado em X_1, \dots, X_n , é dito consistente se esse satisfaz às duas seguintes condições:*

i) T é não viciado, ou seja, $E(T) = \theta$.

ii) A variância de T se aproxima de zero quando n tende ao infinito.

Definição 2.27. *(Estimação por intervalo) Um intervalo que contenha um parâmetro θ , com certa probabilidade $(1-\alpha)$, é chamado de intervalo de confiança para θ , com coeficiente de confiança $(1-\alpha)$.*

Definição 2.28. *(Intervalo de confiança para a média populacional) Consideremos X_1, \dots, X_n uma amostra aleatória de uma variável X com $E(X) = \mu$ e $V(X) = \sigma^2$. Sabemos que para n suficientemente grande, a distribuição de*

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0; 1). \quad (2.26)$$

Seja Z_α o valor da tabela da normal padrão tal que $P(-Z_\alpha < Z < Z_\alpha) = 1 - \alpha$, isto é, $P(-Z_\alpha < \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} < Z_\alpha) = 1 - \alpha \Rightarrow P(\bar{X} - Z_\alpha \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_\alpha \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$. Assim, se diz que $(\bar{X} - Z_\alpha \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_\alpha \frac{\sigma}{\sqrt{n}})$ é um intervalo de confiança para μ , com coeficiente de confiança $1 - \alpha$.

A tabela a seguir apresenta valores de Z_α correspondentes aos vários níveis de confiança usados na prática. Usaremos esta tabela no próximo exemplo.

Níveis de confiança	Z_α
99,73%	3,00
99%	2,58
98%	2,33
96%	2,05
95,45%	2,00
95%	1,96
90%	1,645
80%	1,28
68,27%	1,00
50%	0,6745

Valores de Z_α correspondentes a vários níveis de confiança usados na prática.

Exemplo 2.29. Na medida do tempo de reação, um psicólogo estima que o desvio padrão é de 0,05 segundos. Qual deve ser o tamanho coletado da amostra das medidas para que estejamos 95% confiantes de que o erro na sua estimativa da média do tempo de reação não irá exceder a 0,01 segundos?

Os limites de 95% são $\bar{X} \pm \frac{1,96\sigma}{\sqrt{n}}$ e, o erro da estimativa é $\frac{1,96\sigma}{\sqrt{n}}$. Assumindo $\sigma = s = 0,05$ segundos, vemos que este erro será igual a 0,01 segundos se $\frac{(1,96)(0,05)}{\sqrt{n}} = 0,01$, isto é, $\sqrt{n} = \frac{(1,96)(0,05)}{0,01} = 9,8$ ou $n = 96,04$. Portanto, podemos estar 95% confiantes de que o erro na estimativa será menor do que 0,01 se n for 97 ou maior.

A seguir vamos definir teste de hipótese e significância estatística.

Definição 2.30. A hipótese nula (H_0) é aquela que será testada. É uma afirmação que o parâmetro é tal como especificado (isto é, a afirmação é verdadeira). É uma suposição de que não haja diferença, ou seja, $H_0 : \theta = \theta_0$.

Definição 2.31. A hipótese alternativa (H_1) é aquela que será aceita caso o teste indique que H_0 deva ser rejeitada. A hipótese alternativa geralmente representa a suposição que o pesquisador quer provar, sendo H_0 formulada com expresso propósito de ser rejeitada. E poderá ter uma das seguintes formas: unilateral a esquerda: $H_1 : \theta < \theta_0$; unilateral a direita: $H_1 : \theta > \theta_0$ ou bilateral: $H_1 : \theta \neq \theta_0$.

Definição 2.32. (ERRO DO TIPO I) Rejeitar a hipótese nula quando essa é verdadeira. Chamamos de α a probabilidade de cometer esse erro, isto é,

$\alpha = P(\text{erro do tipo I}) = P(\text{rejeitar } H_0 | H_0 \text{ é verdadeira})$. Sendo α também chamado de nível de significância do teste.

Nos casos mais conservadores o nível de significância do teste usualmente é fixado em 0,01 ou 0,05, pois garantem uma maior amostra.

Definição 2.33. (ERRO DO TIPO II) Um erro é dito do tipo II, quando não se rejeita H_0 e H_0 é falsa. A probabilidade de cometer esse erro é denotada por β , ou seja,

$$\beta = P(\text{erro do tipo II}) = P(\text{não rejeitar } H_0 | H_0 \text{ é falsa}).$$

Definição 2.34. Dada uma amostra aleatória, o teste de hipótese diz, usando uma estatística T , rejeita-se ou não a hipótese nula. Essa hipótese é rejeitada se o valor de T pertence a uma certa região, denominada de região crítica (RC) ou região de rejeição do teste (RR). Podemos escrever a probabilidade do erro do tipo I da seguinte forma $P(T \in RC | H_0 \text{ é verdadeira}) = \alpha$.

A seguir, descrevemos os cinco passos para a construção de um teste de hipótese.

Passo 1: Fixe as hipóteses nula (H_0) e a alternativa (H_1);

Passo 2: Decida qual estatística será usada para julgar a hipótese nula;

Passo 3: Fixe o nível de significância do teste α , usando-o em seguida para definir a região de rejeição;

Passo 4: Use as informações da amostra para calcular o valor da estatística citada no passo 2;

Passo 5: Se o valor citado no passo anterior pertence à região de rejeição, rejeite H_0 , caso contrário, não rejeite.

Tais passos são aplicados no exemplo a seguir.

Exemplo 2.35. Um candidato a prefeito Y de uma certa cidade afirma que 60% dos eleitores são favoráveis à sua candidatura. Um outro candidato, no entanto, deseja contestar essa afirmação e, para isto contratou uma pesquisa de opinião. O instituto contratado usou uma amostra de 200 eleitores e constatou que dos eleitores entrevistados 110 eram favoráveis ao candidato Y . podemos acreditar, ao nível de 5%, que Y tem realmente 60% da preferência dos eleitores?

Passo 1: Colocamos à prova a afirmação do candidato Y , isto é, $H_0 : p = 0,60$, sendo p a proporção de eleitores favoráveis a Y . Sabemos que se esta hipótese não for verdadeira o outro candidato espera uma porcentagem menor, nunca maior. Portanto, a hipótese alternativa neste caso é dada por $H_1 : p < 0,60$, ou seja, o teste é unilateral à esquerda.

Passo 2: A estatística a ser usada aqui é: $Z = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{200}}}$,

onde \hat{P} é a proporção dos 200 eleitores que são favoráveis ao candidato Y e, \hat{P} tem distribuição de, aproximadamente, $N(p, \frac{p(1-p)}{200})$, o que significa que a distribuição de Z se aproxima de $N(0,1)$.

Passo 3: Fixando $\alpha = 5\%$ e sendo este um teste unilateral à esquerda, temos $0,05 = P(\text{Rejeitar } H_0 | H_0 \text{ é verdadeira}) = P(Z < z_0 | p = 0,60)$. Através da tabela da normal padrão (veja Figura 5) e fazendo a subtração de $0,5 - 0,05 = 0,45$, procuramos o valor mais próximo a este na tabela e obtemos que $z_0 = -1,65$, ou seja, neste caso a região de rejeição é definida pelo conjunto dos valores da normal padrão menores que $-1,65$.

Passo 4: Dado que 110 dos eleitores entrevistados eram favoráveis a Y , a proporção amostral fica: $\hat{P} = \frac{110}{200} = 0,55$. Assim, o valor da estatística do teste para os dados observados e considerando H_0 verdadeira, será $Z = \frac{0,55-0,60}{\sqrt{\frac{0,60(1-0,60)}{200}}} = -1,445$.

Passo 5: Do resultado acima, vemos que o valor observado de Z não pertence a região da rejeição. Portanto, não temos motivo para rejeitar a hipótese nula, isto é, há evidências de que o candidato Y tem realmente 60% da preferência do eleitorado.

3 Um exemplo de margem de erro

Uma pesquisa feita dentro dos níveis de significância de teste, provavelmente não apresenta grandes surpresas, seus resultados finais não fugirão daquilo que se obteve em suas pesquisas.

O Brasil possui 209 milhões de habitantes, 27 estados com o Distrito Federal e 5570 municípios. Como uma entrevista feita com 3 mil pessoas, aproximadamente, pode representar a opinião de toda essa população?

Segundo site do Data Folha Uol, em suas pesquisas eleitorais, é admitido o erro como um valor fixo, sua margem de erro é de dois pontos percentuais para mais ou para menos. Ao se definir o tamanho da amostra leva-se em conta o grau de precisão desejável e o detalhamento necessário para a leitura dos dados. Suas amostras variam em torno de 2 mil a 3 mil entrevistados, não possuindo um número máximo de entrevistados.

Exemplo 3.1. *(Eleições 2018 - Presidente da República - 1º turno) Vamos observar, a seguir, as pesquisas realizadas no ano de 2018 para os candidatos à presidência da República. Na Figura 7 temos o resultado do primeiro turno das eleições presidenciais e na Figura 8 o resultado de uma pesquisa de boca de urna. Comparando os dados apresentados nas duas tabelas, percebe-se que no primeiro turno tivemos o resultado das eleições que valida a pesquisa de boca de urna, ou seja, o resultado da eleição ficou dentro dos limites da margem de erro desta pesquisa. A última pesquisa (veja Figura 9) foi realizada nos dias 3 e 4 de outubro de 2018; teve como margem de erro 2 pontos percentuais para mais ou para menos; contou com 10930 eleitores entrevistados em 389 municípios, com nível de confiança de 95% e os contratantes foram TV Globo e Folha de S. Paulo. E ainda, as pesquisas anteriores mostravam o candidato Jair Bolsonaro em primeiro lugar, mas o resultado não se enquadrava dentro dos limites da margem de erro das mesmas, que apontavam este candidato com resultados variando entre 37% e 41%, na última pesquisa realizada. Se observarmos os dados apresentados pela pesquisa temos $Z_\gamma = 1,96$ $\epsilon = 0,02$ $p = 0,5$ Logo,*

$$n = \frac{Z_\gamma^2 p(1-p)}{\epsilon^2} = \frac{(1,96)^2 \cdot 0,5(1-0,5)}{0,02^2} = \frac{3,8416 \cdot 0,5 \cdot 0,5}{0,0004} = \frac{0,9604}{0,0004} = 2401 \text{ ou}$$

seja, teriam que entrevistar no mínimo 2401 eleitores na última pesquisa.

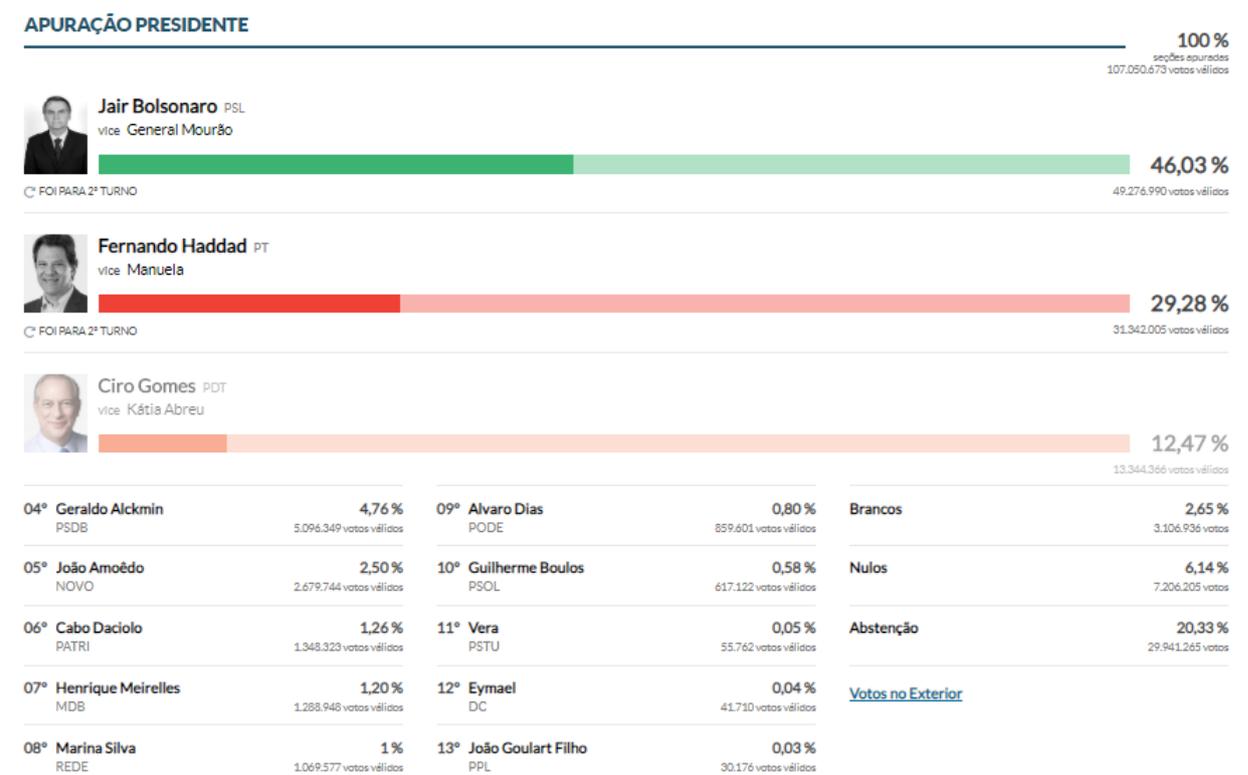


Figura 7 – Resultado das eleições presidenciais (1º turno)

Fonte: página Política no Estadão

Pesquisa para presidente - Primeiro turno

Estimulada, em %



Pesquisa realizada pelo Ibope/Boca de Urna em 7/out com 30.000 entrevistados (Brasil). Contratada por: IBOPE INTELIGENCIA PESQUISA E CONSULTORIA. Registro no TSE: BR-07052/2018 . Margem de erro: 2 pontos percentuais. Confiança: 99%. *Não sabe / Não respondeu

Figura 8 – Pesquisa para presidente primeiro turno (boca de urna)

Fonte: Página Pesquisas Eleitorais em Gazeta do Povo.

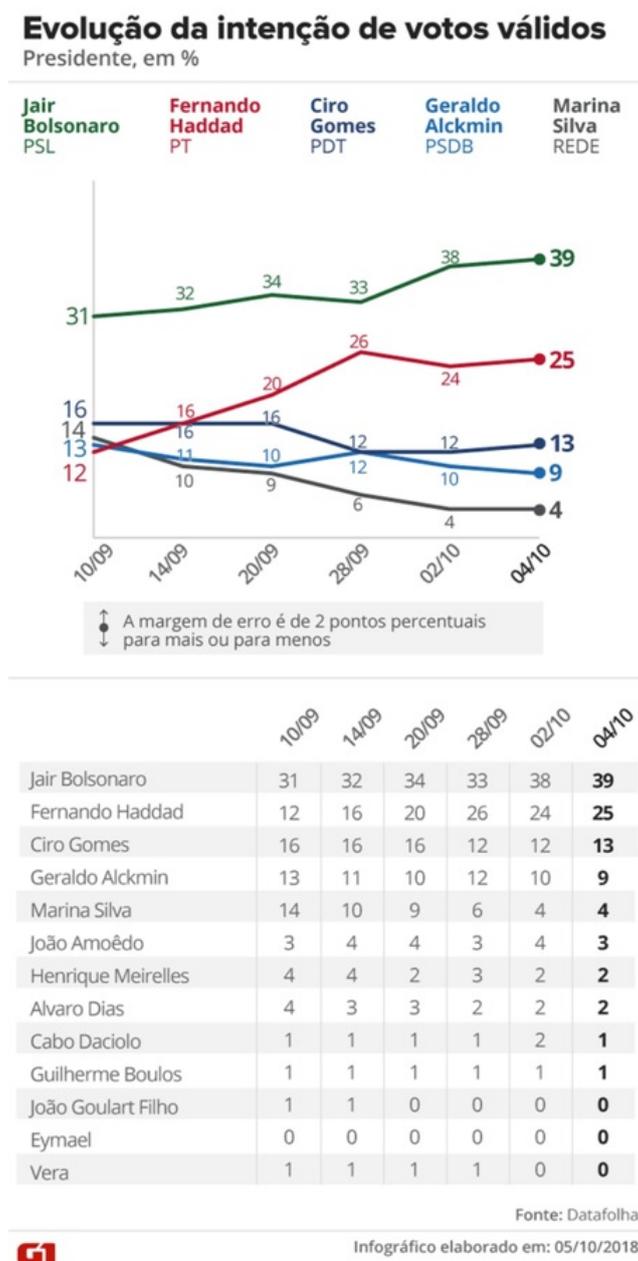


Figura 9 – Evolução de intenção de votos válidos no primeiro turno de 2018

Fonte: Página Eleições 2018 em Globo.com.

Exemplo 3.2. *(Eleições 2018 - Presidente da República - 2º turno) Tivemos um episódio de segundo turno disputado entre Jair Bolsonaro e Fernando Haddad; teve como resultado final o candidato Jair Bolsonaro eleito com 55% dos votos (veja Figuras 10 e 11). Se compararmos o resultado final com a pesquisa realizada no dia 27 de outubro (veja Figura 12), podemos dizer que o resultado final estava dentro dos limites da margem de erro da pesquisa que apontava Bolsonaro na frente, com 54% e Haddad com 46% dos votos válidos. Com probabilidade de os resultados retratarem a realidade de noventa e cinco por cento, e margem de erro de dois pontos - para mais ou para menos, foram entrevistados 3010 eleitores nos dias 26 a 27 de outubro. Se observarmos os dados apresentados pela pesquisa*

temos $Z_\gamma = 1,96$ $\epsilon = 0,02$ $p = 0,5$ Logo,

$$n = \frac{Z_\gamma^2 p(1-p)}{\epsilon^2} = \frac{(1,96)^2 \cdot 0,5(1-0,5)}{0,02^2} = \frac{3,8416 \cdot 0,5 \cdot 0,5}{0,0004} = \frac{0,9604}{0,0004} = 2401 \text{ ou}$$

seja, teriam que entrevistar no mínimo 2401 eleitores nesta pesquisa.

Jair Bolsonaro é eleito presidente com 57,8 milhões de votos

Candidato pelo PSL, capitão reformado do Exército derrotou o petista Fernando Haddad no 2º turno e vai governar o Brasil pelos próximos 4 anos.



Jair Messias Bolsonaro, do PSL, foi eleito o 38º presidente da República neste domingo (28) ao derrotar em segundo turno o petista Fernando Haddad, interrompendo um ciclo de vitórias do PT que vinha desde 2002. A vitória foi confirmada às 19h18, quando, com 94,44% das seções apuradas, Bolsonaro alcançou 55.205.640 votos (55,54% dos válidos) e não podia mais ser ultrapassado por Haddad, que naquele momento somava 44.193.523 (44,46%). Com 100% das seções apuradas, Bolsonaro recebeu 57.797.847 votos (55,13%) e Haddad, 47.040.906 (44,87%). (MAZUI, 2018).

Figura 10 – Bolsonaro eleito presidente no 2º turno

Fonte: Página Eleições 2018 em Globo.com.

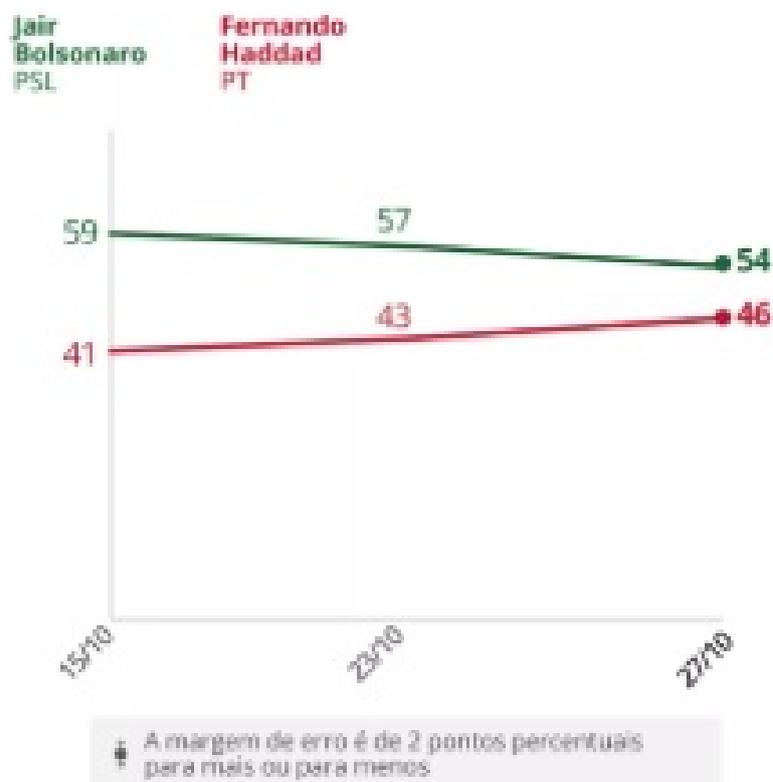


Figura 11 – Resultado das eleições presidenciais 2º turno

Fonte: Página Eleições 2018 em Globo.com.

Votos válidos - 2º turno

Intenção de voto para presidente, em %



Fonte: IBOPE



Infográfico elaborado em: 27/10/2018

Pesquisa Ibope - 27-10 - Votos válidos no segundo turno da eleição presidencial — Foto: Arlê/G1

Votos totais

Nos votos totais, os resultados foram os seguintes:

- Jair Bolsonaro (PSL): 47%
- Fernando Haddad (PT): 41%
- Em branco/nulo: 10%
- Não sabe: 2%

Figura 12 – Pesquisas presidente no 2º turno

Fonte: Página Eleições 2018 em Globo.com.

Exemplo 3.3. (Eleições 2018 - governador do Paraná) Ainda se tratando das eleições do

ano de 2018, podemos citar, as eleições para governador do estado do Paraná. O resultado se apresentou conforme as expectativas, já que a pesquisa de boca de urna se confirmou dentro dos limites da margem de erro (veja Figuras 13, 14 15). Podemos observar que as pesquisas realizadas mostravam o candidato Ratinho Junior em primeiro lugar, seguido pela candidata Cida Borghetti, sendo a última pesquisa entre os dias 4 e 6 de outubro com margem de erro de 2 pontos percentuais, ou seja, o resultado da eleição deveria apontar Ratinho Junior entre 42% e 46% dos votos válidos. A vitória do candidato se confirmou, mas não dentro dos limites da margem de erro. Se observarmos os dados apresentados pela pesquisa de boca de urna temos $Z_\gamma = 2,58$ $\epsilon = 0,03$ $p = 0,5$ Logo,

$$n = \frac{Z_\gamma^2 p(1-p)}{\epsilon^2} = \frac{(2,58)^2 \cdot 0,5(1-0,5)}{0,03^2} = \frac{6,65 \cdot 0,5 \cdot 0,5}{0,0009} = \frac{1,66}{0,0009} = 1844 \text{ ou seja,}$$

teriam que entrevistar no mínimo 1844 eleitores na pesquisa de boca de urna.

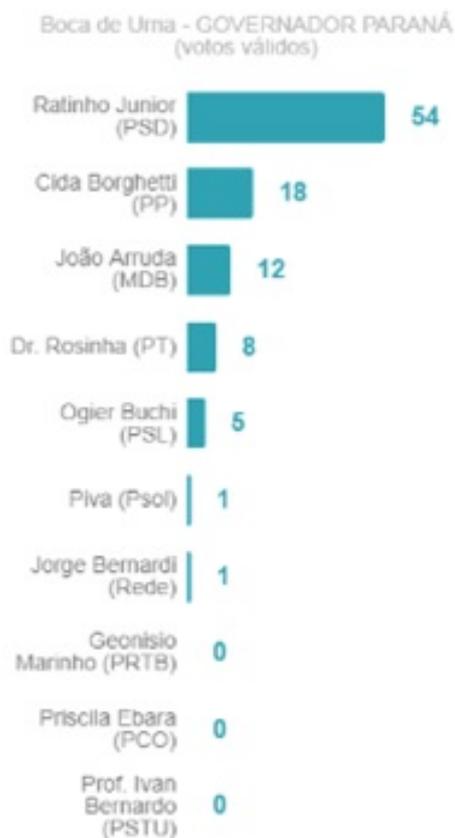


Figura 13 – Resultado das eleições para governador no estado do Paraná

Fonte: Página Governador do Paraná, Eleições 2018 em UOL.

Pesquisa para governador do Paraná

Estimulada, em %



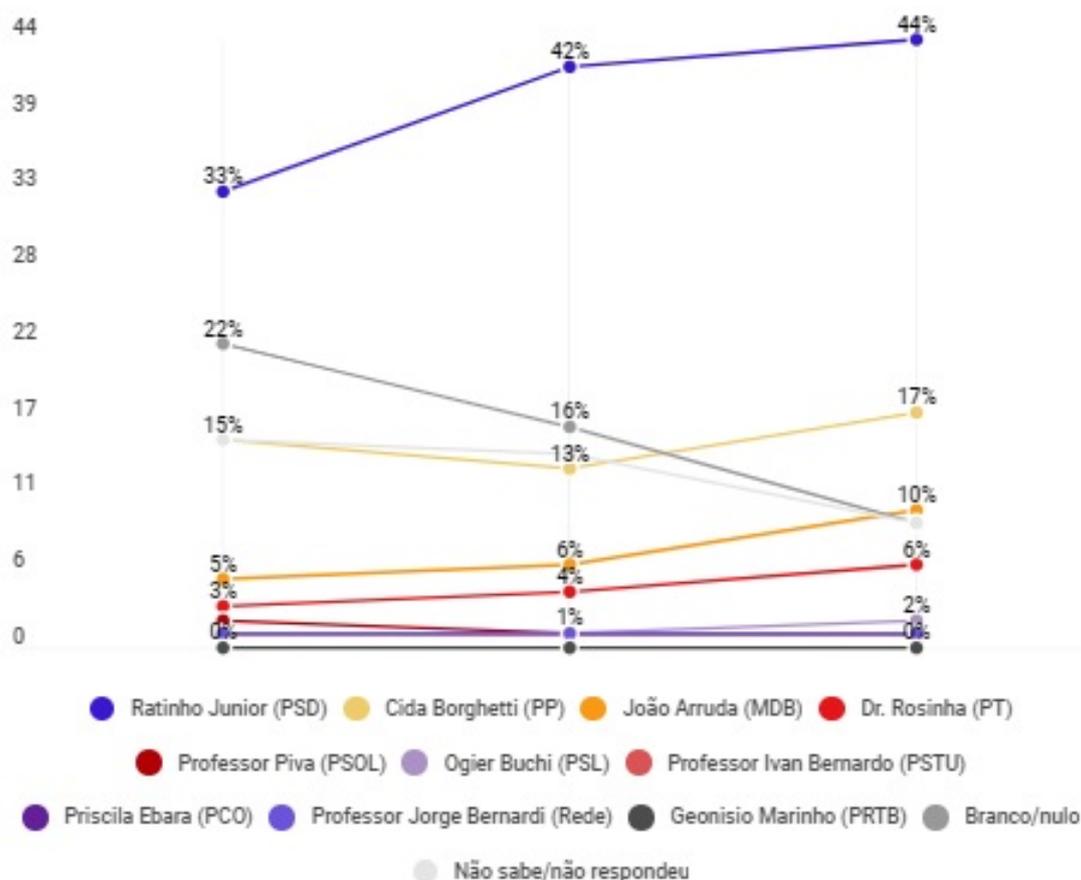
Pesquisa realizada pelo Ibope/Boca de Urna de 7/out a 7/out/2018 com 2.600 entrevistados (Paraná). Contratada por: IBOPE INTELIGENCIA PESQUISA E CONSULTORIA. Registro no TSE: PR-05072/2018. Margem de erro: 3 pontos percentuais. Confiança: 99%. *Não sabe / Não respondeu

Figura 14 – Pesquisa de boca de urna para governador do estado do Paraná

Fonte: Página Eleições 2018 em Gazeta do Povo.

Veja abaixo os resultados das pesquisas da corrida eleitoral para o governo do Paraná realizadas após os registros de candidaturas.

Ibope



Ibope (data 19-21/8, registro BR-06574/2018 / PR-04869/2018, amostra 1008, margem de erro 3 p.p.; data 1-4/9 registro BR-01988/2018 / PR-04985/2018; amostra 1204; margem de erro 3 p.p.; data 24-26/9, registro PR-07128/2018 / BR-03369/2018, amostra 1.204, margem de erro 3 p.p) 6/10 (data 4-6/10, registro BA-07427/2018 / BR-07032/2018, amostra 2.002, margem de erro 2 p.p)

Ibope (data 19-21/8, registro BR-06574/2018 / PR-04869/2018, amostra 1008, margem de erro 3 p.p.; data 1-4/9 registro BR-01988/2018 / PR-04985/2018; amostra 1204; margem de erro 3 p.p.; data 24-26/9, registro PR-07128/2018 / BR-03369/2018, amostra 1.204, margem de erro 3 p.p) 6/10 (data 4-6/10, registro BA-07427/2018 / BR-07032/2018, amostra 2.002, margem de erro 2 p.p)

Figura 15 – Pesquisa eleitoral para governador do estado do Paraná

Fonte: Página Eleições em Veja.

Olhando para os exemplos acima, vemos que os resultados nas pesquisas se formalizaram nas eleições. De frente com esses acertos, o que pode ter ocorrido no resultado da candidata ao senado de Minas Gerais, Dilma Rousseff?

Exemplo 3.4. (Eleições 2018 - Senador de Minas Gerais - 1º turno) O resultado da eleição, no primeiro turno, trouxe Rodrigo Pacheco e o Jornalista Carlos Viana, eleitos para a disputa do segundo turno, como podemos ver na Figura 16. Mas as pesquisas mostravam algo diferente. As pesquisas apontavam a candidata Dilma Rousseff em primeiro lugar com margem entre 24% e 30% dos votos válidos (veja Figuras 17 e 18). Ao final da eleição a mesma obteve apenas 13% dos votos válidos (veja Figura 16), ficando em quarto lugar na disputa. A pesquisa de boca de urna (veja Figura 20), mostrou exatamente esse cenário, a candidata, entre os limites de 12 e 18 por cento dos votos válidos.

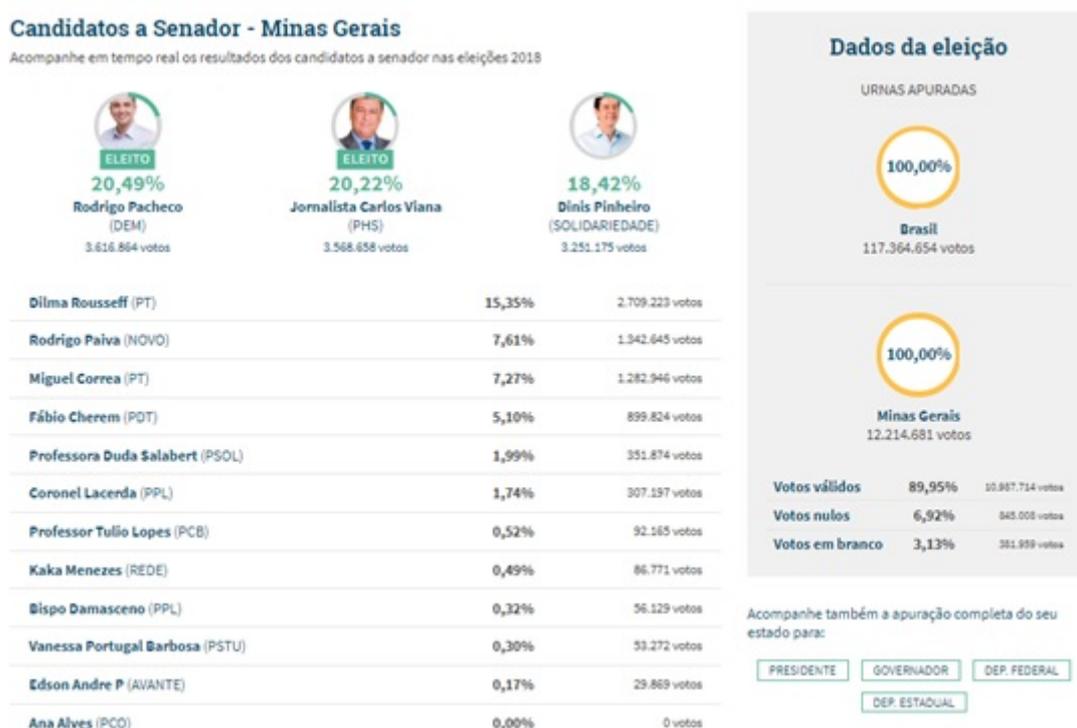


Figura 16 – Resultado das eleições para senado estado de Minas Gerais no primeiro turno

Fonte: Página Eleições 2018 em Gazeta do Povo.

Dilma é líder isolada com 27%, mas cai vantagem pelo Senado em MG, diz Datafolha

Segunda vaga tem empate técnico entre candidatos do DEM, do PHS e do SD



A ex-presidente Dilma Rousseff (PT), que é candidata ao Senado por Minas Gerais - Douglas Magno - 21.set.2018/AFP

Daniel Carvalho

BELO HORIZONTE A ex-presidente Dilma Rousseff (PT) chega à reta final da disputa pelo Senado em Minas Gerais como líder isolada porém com sua vantagem diminuindo, segundo pesquisa Datafolha divulgada nesta quinta-feira (4).

A petista tem 27% das intenções de voto. Tinha 28% [na semana passada](#) e, há duas semanas, tinha 29%.

Figura 17 – Pesquisas eleitorais para o senado de Minas Gerais no primeiro turno

Fonte: Página Eleições 2018 em Folha UOL.

No intervalo entre as duas pesquisas, o juiz Sergio Moro, responsável pela Lava Jato, [levantou o sigilo da delação](#) de Antonio Palocci. O ex-ministro afirma que Dilma e o ex-presidente Lula estiveram em uma reunião na qual foi acertado pagamento de R\$ 40 milhões em propina para a campanha dela em 2010. Dilma e Lula negam as acusações.

Pesquisa Datafolha

Senador - Minas Gerais - Intenção de voto - Estimulada

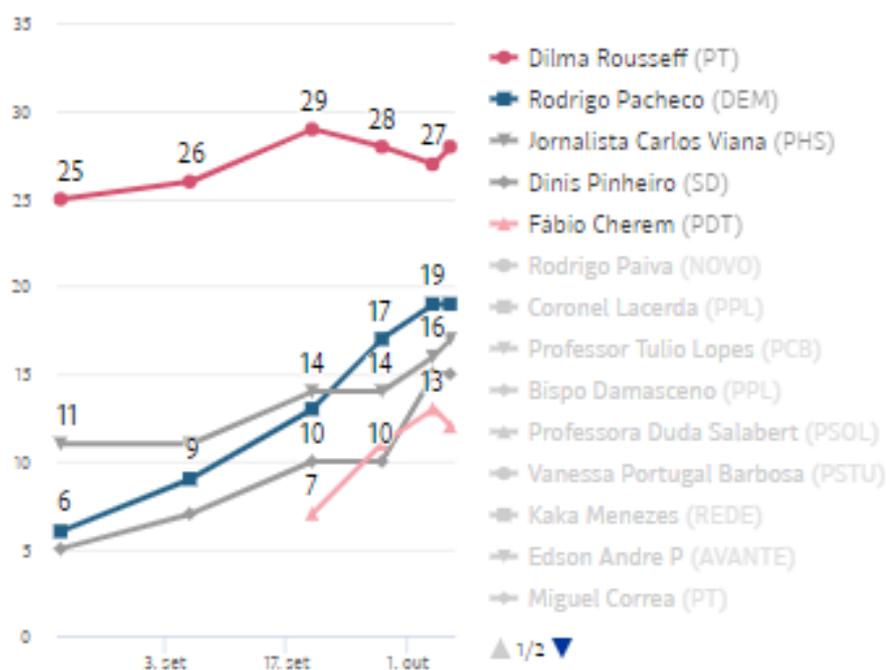


Figura 18 – Pesquisas eleitorais para o senado de Minas Gerais no primeiro turno

Fonte: Página Eleições 2018 em Folha UOL.

Na disputa pela segunda vaga, três candidatos estão tecnicamente empatados.

Rodrigo Pacheco (DEM) tem 19% das intenções de voto (tinha 17% na semana passada e 13% há duas semanas); o Jornalista Carlos Viana (PHS) 16% (tinha 14% na semana passada, o mesmo que há duas semanas); Dinis Pinheiro (SD) tinha 10% na semana passada, o mesmo que há duas semanas, e agora tem 15%.

Segundo o Datafolha, Fábio Cherem (PDT) tem 13% (tinha 11%) e, em relação a Pacheco, está no limite da margem de erro. Por isso, informa o instituto, é baixa a probabilidade de estar tecnicamente empatado com o candidato do DEM.

Rodrigo Paiva (Novo) tem 9% (tinha 8%); Coronel Lacerda (PPL) tem 5% (tinha 6%); Bispo Damasceno (PPL) tem 5% (mesmo índice anterior); Miguel Correa (PT) tem 4% (tinha 5%); Professor Tulio Lopes (PCB) tem 4% (mesmo índice anterior); Professora Duda Salabert (PSOL) tem 3% (tinha 4%); Vanessa Portugal (PSTU) tem 2%; Ana Alves (PCO) tem 2%; e com 1%, cada um, Kaka Menezes (Rede) e Edson Andre (Avante).

Para a primeira vaga ao Senado, uma parcela de 15% dos eleitores declarou que pretende votar em branco ou nulo (eram 17%) e 15% estão indecisos (eram 14%). Já, para a segunda vaga, 20% pretendem votar em branco ou nulo (eram 24%) e 25% estão indecisos (mesmo índice anterior).

A pesquisa foi encomendada pela Folha e pela TV Globo e foi registrada no TSE (Tribunal Superior Eleitoral) com o número MG-02076/2018. O Datafolha realizou, nos dias 03 e 04 de outubro de 2018, 1.585 entrevistas presenciais com eleitores em 72 municípios de Minas Gerais. A margem de erro máxima é de 3 pontos percentuais, para mais ou para menos, para o total da amostra.

O nível de confiança de 95%. Ou seja, considerando a margem de erro, a chance do resultado retratar a realidade é de 95%.

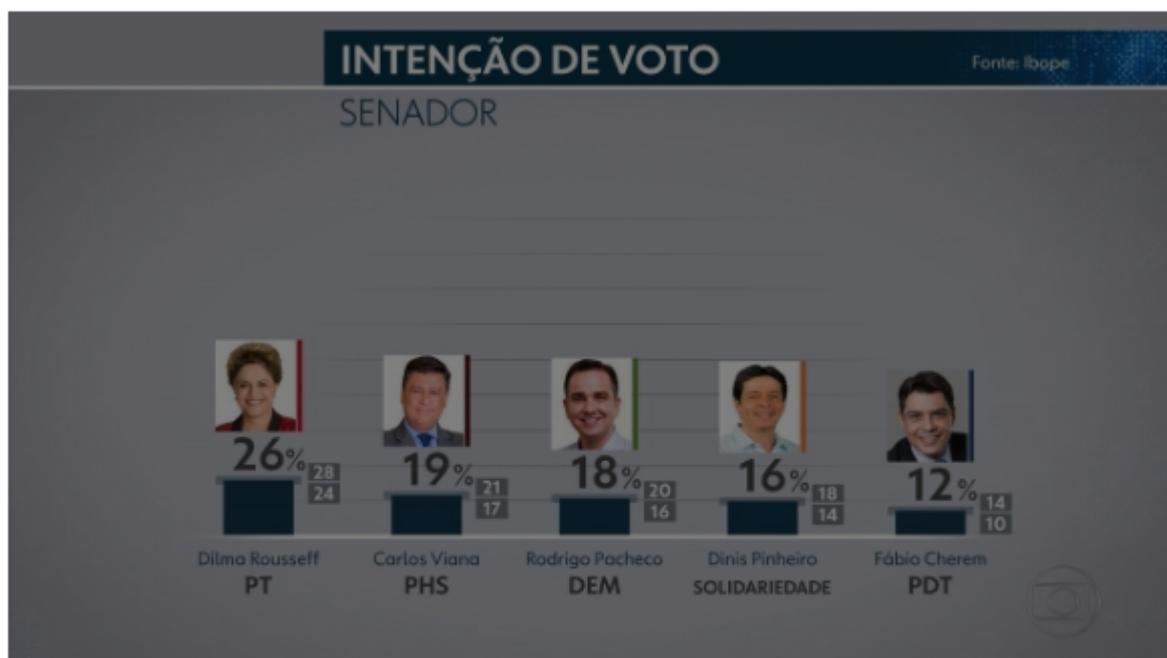
Figura 19 – Descrição das pesquisas Data Folha para senado de Minas Gerais

Fonte: Página Eleições 2018 em Folha UOL.

Se observarmos os dados apresentados pela pesquisa temos: $Z_\gamma = 1,96$ $\epsilon = 0,03$ $p = 0,5$
Logo,

$$n = \frac{Z_\gamma^2 p(1-p)}{\epsilon^2} = \frac{(1,96)^2 \cdot 0,5(1-0,5)}{0,03^2} = \frac{3,8416 \cdot 0,5 \cdot 0,5}{0,0009} = \frac{0,9604}{0,0009} = 1067.$$

ou seja, teriam que entrevistar no mínimo 1067 eleitores nesta última pesquisa.



Ibope Senado - Minas Gerais, votos válidos: Dilma 20%, Carlos Viana 14% e Pacheco 14%

Sobre a pesquisa

- **Margem de erro:** 2 pontos percentuais para mais ou para menos
- **Quem foi ouvido:** 2.002 eleitores com 16 anos ou mais
- **Quando a pesquisa foi feita:** entre os dias 4 e 6 de outubro.
- **Registro no TRE:** MG - 01559/2018
- **Registro no TSE:** BR - 08924/2018
- O nível de confiança utilizado é de 95%. Isso quer dizer que há uma probabilidade de 95% de os resultados retratarem o atual momento eleitoral, considerando a margem de erro.

Figura 20 – Pesquisa IBOPE para senado de Minas Gerais

Fonte: Página Eleições 2018 em Globo.com

Boca de urna mostra vagas para o Senado indefinidas em Minas

A ex-presidente Dilma Rousseff (PT), que estava em primeiro lugar nas pesquisas eleitorais, apareceu na quarta posição

EM Estado de Minas

Postado em 07/10/2018 17:19 / atualizado em 07/10/2018 17:34



A pesquisa boca de urna realizada pelo IBOPE durante a votação deste domingo apontou um cenário indefinido para a eleição ao Senado em Minas Gerais.

Diferentemente do que apontavam as pesquisas, o jornalista Carlos Viana (PHS) aparece em primeiro lugar em percentual de votos válidos, com 21% dos votos válidos. Na sequência estão Dinis Pinheiro (SD), com 18%, Rodrigo Pacheco (DEM), com 18% e Dilma Rousseff, com 15%.

A margem de erro é de três pontos percentuais para mais ou para menos.

Nas últimas pesquisas divulgadas, a ex-presidente Dilma Rousseff aparecia em primeiro lugar, com 20% das intenções de voto. Na sequência, embolados na segunda posição, estavam Carlos Viana e Dinis Pinheiro, com 14%, e Dinis Pinheiro, com 12%.

Figura 21 – Pesquisa Boca de Urna para senado de Minas Gerais

Fonte: Página Política em Estado de Minas.

Veja os números do Ibope Minas

Carlos Viana (PHS):	21%
Dinis Pinheiro (SOLIDARIEDADE):	18%
Rodrigo Pacheco (DEM):	18%
Dilma Rousseff (PT):	15%
Miguel Correa (PT):	7%
Rodrigo Paiva (NOVO):	7%
Fábio Cherem (PDT):	4%
Duda Salabert (PSOL):	3%
Coronel Lacerda (PPL):	2%
Bispo Damasceno (PPL):	1%
Edson Andre P (AVANTE):	1%
Kaka Menezes (REDE):	1%
Professor Tulio Lopes (PCB):	1%
Vanessa Portugal Barbosa (PSTU):	1%
Ana Alves (PCO):	0%

Figura 22 – Pesquisa Boca de Urna para senado de Minas Gerais

Fonte: Página Política em Estado de Minas.

Aqui pudemos ver que as pesquisas feitas anteriormente caíram na margem de erro, pois toda pesquisa tem a chance de errar, porque não há uma precisão por se tratar de uma amostragem. Podemos ter nos deparado também com falhas humanas, do tipo, um questionário mal elaborado e, ou tendencioso, a distribuição da retirada da amostra, sabendo que Minas Gerais tem 853 municípios e somente em 72 foram feitas as entrevistas. Com 95% de confiança a pesquisa estatística não é absoluta, livre de erros, mas nos trás uma tendência, quando bem realizada, podendo nos mostrar resultados muito próximos dos reais, e válidos para toda a população em questão.

4 Conclusão

Divulgado no dia sete de outubro de 2018, a pesquisa de boca de urna apontava Jair Messias Bolsonaro com 45% dos votos válidos e Fernando Haddad com 28% dos votos válidos, sendo confirmado horas depois que o segundo turno das eleições seria disputado entre Bolsonaro com 46,03% dos votos válidos e Haddad com 29,28% dos votos válidos. Estando o resultado dentro da margem de erro, que era de 2 pontos percentuais para mais ou para menos, da pesquisa de boca de urna, confirmou-se o que as pesquisas realizadas a partir do dia 14 de setembro de 2018 apontavam.

No segundo turno, as pesquisas apontavam Jair Bolsonaro como futuro presidente, com 54% dos votos válidos e Fernando Haddad com 46%, com margem de erro de dois pontos percentuais para mais ou para menos, e nível de significância de 95%. Na última pesquisa foram entrevistados 9.173 eleitores em 341 municípios, divulgada no dia 27 de outubro de 2018. No dia seguinte se confirmou o que as pesquisas nos mostravam, Bolsonaro foi eleito com 55,13% dos votos válidos e Haddad teve 44,87% dos votos válidos.

Ainda se tratando das eleições, para governador do estado do Paraná tivemos no dia sete de outubro de 2018, Ratinho Junior eleito no primeiro turno com 59,99% dos votos válidos seguido pela candidata Cida Borghetti com 15,53% dos votos válidos. Confirmando a pesquisa de boca de urna que apontava Ratinho Junior com 54% e Cida Borguetti com 18% dos votos válidos, com margem de erro de 5 pontos percentuais para mais ou para menos. Na pesquisa do dia 6 de outubro, com margem de erro de dois pontos percentuais para mais ou para menos, com uma amostra de 2002 entrevistados, Ratinho estava com 44% e Cida Borguetti com 17% dos votos válidos, mostrava a vitória de Ratinho Junior mas, não foi assertiva, dentro da margem de erro, quanto a porcentagem de votos dele.

Já as pesquisas para senador no estado de Minas Gerais apontavam a candidata Dilma Rousseff em primeiro lugar, disparada, em todas elas. Na última pesquisa, que foi dia 06 de outubro de 2018, Dilma estava com 27% dos votos, Rodrigo Pacheco com 19% e Carlos Viana com 17% dos votos válidos, sendo a margem de erro de três pontos percentuais para mais ou para menos, em uma pesquisa com 95% de significância, ou seja, havia 95% de chance dos resultados se tornarem reais.

A pesquisa de boca de urna, no dia sete de outubro, veio para contestar os resultados anteriores, apresentando a candidata Dilma em quarto lugar, com 15% dos votos, e Carlos Viana em primeiro lugar com 21% dos votos válidos. Seguida pelo resultados das eleições, que trazia Rodrigo Pacheco com 20,49% e Carlos Viana com 20,22% dos votos válidos, porcentagem de votos que já era esperada, seus resultados ficaram dentro da margem de erro da última pesquisa. O que teria acontecido com a candidata Dilma Rousseff?

A mesma ficou em quarto lugar com 15% dos votos válidos, não estando nem perto do esperado, já que a margem de erro das pesquisas eram de três pontos percentuais para mais ou para menos, e a mesma se encontrava com 27% dos votos. Na pesquisa foram entrevistados 1585 pessoas em 72 municípios de Minas Gerais. Na pesquisa IBOPE divulgada no dia 06 de outubro de 2018 podemos ver a candidata Dilma Rousseff com 26%, Carlos Viana com 19% e Rodrigo Pacheco com 18% dos votos válidos, com 2 pontos percentuais para mais ou para menos na margem de erro e nível de significância de 95%, onde foram entrevistados 2002 eleitores. Nos mostrando praticamente os mesmos resultados da pesquisa Datafolha. Os resultados destas pesquisas caíram dentro dos 5% de chance da pesquisa não retratar a realidade, pois nem mesmo dentro da margem de erro, a candidata Dilma se encontrava.

Uma pesquisa prevê um resultado, mas não é uma verdade absoluta, é apenas um retrato momentâneo da realidade em determinado local. Acontece que muitas vezes a amostra entrevistada não representa toda a população, uma pesquisa centralizada, em um único ponto da cidade, por exemplo, não vai representar toda a cidade, que se distribui com diferentes níveis sociais, graus de escolaridade e até mesmo a idade dos moradores.

Temos ainda que a margem de erro estabelecida pode não ser suficiente para retratar a amostra tomada. Quando escolhemos a amostra, por exemplo, de 50 mulheres com mais de 30 anos e ensino superior completo, não estamos tomando uma amostra aleatória, e nesse caso a margem de erro deve ser maior, pois essa amostra não representa toda a população. A margem de erro nada mais é, do que um limite, para mais ou para menos, do resultado obtido, e é pré estabelecido pelo pesquisador, assim como o nível de significância, que é a porcentagem da pesquisa retratar a realidade. Com estes números fixos, podemos calcular o tamanho da amostra que precisamos entrevistar, para realizar a pesquisa. A amostra deve representar toda a população, ou seja, deve ser bem distribuída entre as cidades, tomando o cuidado de se estabelecer pontos distintos da mesma, e no caso de uma pesquisa estadual, tomar diversas cidades, com perfis distintos, tomando o cuidado para não induzir o resultado da pesquisa.

Mesmo com uma amostra aleatória bem distribuída e bem representativa da população, podemos não ter o resultado esperado, simplesmente pelo fato de que a pesquisa não possui 100% de significância, ou seja, está sujeita ao erro. Isso porque, muitas vezes, não conseguimos tomar toda a população, por exemplo, para saber se uma laranjeira produz laranjas doces, não vamos provar todas as laranjas que esta árvore produz, pegamos apenas uma amostra desta laranjeira, e a mesma representará todas as laranjas que esta árvore produz nesta época.

Uma pesquisa estatística é extraordinária, com apenas uma amostra podemos representar uma população, a partir disso, podemos fazer exames com apenas uma amostra de sangue, podemos verificar a durabilidade e resistência de produtos, a saúde das pessoas,

animais e plantas, entre outras coisas como neste caso, o resultado das eleições. A estatística é fundamental e indispensável no nosso cotidiano, já não sabemos e não conseguimos viver sem ela, pois em quase tudo se faz necessário o uso de amostras para nos inferir resultados em relação a uma população em questão.

Referências

- [1] AZEVEDO, Paulo Roberto Medeiros, **Introdução a estatística**. Natal: EDUFRN - Editora da UFRN, 2005.
- [2] CARVALHO, Daniel. Dilma é líder isolada com 27%, mas cai vantagem pelo Senado em MG, diz Datafolha. **UOL**. 4.out.2018. Disponível em: <https://www1.folha.uol.com.br/poder/2018/10/dilma-e-lider-isolada-mas-cai-vantagem-em-disputa-pelo-senado-em-mg-diz-datafolha.shtml>. Acesso em: 13 de fev. de 2019.
- [3] CASTANHEIRA, Nelson Pereira. **Estatística aplicada a todos os níveis**. 2. ed. rev. e atual. Curitiba: IBPEX, 2003.
- [4] ELEIÇÕES 2018. Candidatos a Senador: Minas Gerais. **Gazeta do Povo**. Disponível em: <https://www.gazetadopovo.com.br/apuracao/resultado-eleicoes-2018/minas-gerais/senador/>. Acesso em: 13 de dez. de 2018.
- [5] ELEIÇÕES 2018. Pesquisas Eleitorais, Boca de urna. **Gazeta do Povo**. 07 de out. de 2018. Disponível em: <https://especiais.gazetadopovo.com.br/eleicoes/2018/pesquisas-eleitorais/ibope/boca-de-urna-governo-pr-ratinho-54-cida-18-joao-arruda-12/>. Acesso em: 04 de abr. de 2019.
- [6] G1. Eleições 2018, Ibope para Presidente. **Globo.com**. 27 de out. de 2018. Disponível em: <https://g1.globo.com/politica/eleicoes/2018/noticia/2018/10/27/ibope-para-presidente-votos-validos-bolsonaro-54-haddad-46.ghtml>. Acesso em: 10 de dez. de 2018.
- [7] G1. Eleições 2018, Pesquisa Data Folha Para Presidente. **Globo.com**. 04 de out. de 2018. Disponível em: <https://g1.globo.com/politica/eleicoes/2018/noticia/2018/10/04/pesquisa-datafolha-para-presidente-bolsonaro-35-haddad-22-ciro-11-alckmin-8-marina-4.ghtml>. Acesso em: 04 de abr. de 2019.
- [8] G1. Eleições 2018, Presidente. **Globo.com**. Disponível em: <https://g1.globo.com/politica/eleicoes/2018/apuracao/presidente.ghtml>. Acesso em: 10 de dez. de 2018.
- [9] G1 MINAS. Ibope Senado - Minas Gerais, votos válidos. **Globo.com**. Belo Horizonte. 06 de out. de 2018. Disponível em: <https://g1.globo.com/mg/minas-gerais/eleicoes/2018/noticia/2018/10/06/ibope-senado-minas-gerais-votos-validos->

- dilma-20-carlos-viana-14-pacheco-14-e-dinis-pinheiro-12.ghtml. Acesso em 09 de out. de 2019.
- [10] GUIDORIZZI, Hamilton Luiz. **Um curso de Cálculo**. vol 2. 5ed. 7 reimpr. Rio de Janeiro: LTC, 2008.
- [11] MAZUI, Guilherme. Eleições 2018, Presidente. **Globo.com**. Brasília. 28 de out. de 2018. Disponível em: <https://g1.globo.com/politica/eleicoes/2018/noticia/2018/10/28/jair-bolsonaro-e-eleito-presidente-e-interrompe-serie-de-vitorias-do-pt.ghtml>. Acesso em: 10 de dez. de 2018.
- [12] MEMÓRIA, José Maria Pompeu. **Breve História da Estatística**. Brasília. Embrapa, 2004.
- [13] MORETTIN, Luiz Gonzaga. **Estatística Básica: probabilidade e inferência**. Volume único. São Paulo: Pearson Prentice Hall, 2010.
- [14] MORETTIN, P.A., Bussab, W.O., **Estatística Básica**. 6. ed. São Paulo: Saraiva, 2010.
- [15] NUNES, E.M.A., Almeida, W.M., **Estatística aplicada usando excel**. Maringá: Eduem, 2016.
- [16] PESQUISAS Eleitorais, Eleições 2018. **Gazeta do Povo**. Disponível em: <https://especiais.gazetadopovo.com.br/eleicoes/2018/pesquisas-eleitorais/ibope/boca-de-urna-aponta-jair-bolsonaro-e-fernando-haddad-no-2o-turno/>. Acesso em: 04 de abr. de 2019.
- [17] POLÍTICA. Boca de urna mostra vagas para o Senado indefinidas em Minas. **Estado de Minas**. 07 de out. de 2018. Disponível em: https://www.em.com.br/app/noticia/politica/2018/10/07/interna_politica,995173/boca-de-urna-mostra-vagas-para-o-senado-indefinidas-em-minas.shtml. Acesso em: 20 de fev de 2019.
- [18] POLÍTICA. Eleições 2018, Apuração do primeiro turno para presidente. **Estadão**[S.D]. Disponível em: <https://politica.estadao.com.br/eleicoes/2018/cobertura-votacao-apuracao/primeiro-turno>. Acesso em: 26 de set. de 2019.
- [19] SPIGEL, M.R., Schiller, J., Srinivasan, A., **Probabilidade e estatística**. 3. ed., Porto Alegre: Bookman, 2013.
- [20] UOL. Eleições 2018, Governador do Paraná. **UOL**. Disponível em: <https://placar.eleicoes.uol.com.br/2018/1turno/pr/>. Acesso em: 23 de abr. de 2019.

-
- [21] VEJA. Eleições. Eleições 2018: todas as pesquisas para governador do Paraná. **Veja**. 10 de set de 2018. Disponível em: <https://veja.abril.com.br/politica/eleicoes-2018-todas-as-pesquisas-para-governador-do-parana/>. Acesso em: 04 de abr. de 2019.
- [22] VIEIRA, S., Wada,R., **Estatística: introdução ilustrada**. São Paulo: Atlas, 1985.