

CÁLCULO DO TAMANHO DA AMOSTRA PARA UMA PESQUISA ELEITORAL

Raquel Oliveira dos Santos, Luis Felipe Dias Lopes

Programa de Pós-Graduação em Estatística e Modelagem Quantitativa
CCNE – UFSM, Santa Maria – RS

e-mails: raquel.os@ibest.com.br, lflopes@ccne.ufsm.br

Home Page: http://planeta.terra.com.br/educacao/felipe_lopes

RESUMO

Numa pesquisa de intenção de voto o universo a ser entrevistado é todos os eleitores acima de 15 anos aptos a votar naquela eleição. Em função dos problemas de custo torna-se impraticável consultar todos as pessoas que compõe esse universo, logo temos que nos contentar em entrevistar uma pequena parcela dessa população que recebe o nome de amostra. Os principais fatores utilizados para definir a composição da amostra são: nível sócio econômico, grau de instrução, sexo e idade. Resumidamente, durante a realização de uma pesquisa existe uma proporção desconhecida de eleitores que pretendem votar num determinado candidato e após a conclusão das entrevistas, obtém-se a proporção de eleitores da amostra que manifestaram preferência por tal candidato. O problema agora é saber como usar essas informações para se obter uma estimativa para a proporção de eleitores na população. Uma forma de se mostrar esses resultados é utilizando intervalo de confiança para a proporção de eleitores da amostra, que pretendem votar num determinado candidato.

Palavras chaves: pesquisa eleitoral, estimação, tamanho da amostra

1 INTRODUÇÃO

Este artigo tem por finalidade discutir algumas idéias sobre técnicas de estimação de parâmetros utilizadas nas pesquisas de intenção de voto, no que se refere aos possíveis erros que poderão ocorrer, e o método estatístico de interpretá-los. O erro de previsão utilizado pelos Institutos de Pesquisas é um dos principais motivos que nos levou a escrever tal artigo.

Todo Instituto de Pesquisa deseja acertar os resultados de uma eleição através das pesquisas de campo, mas nem sempre isso será possível, como mostraremos a seguir.

2 TAMANHO DA AMOSTRA

Neste estudo serão consideradas apenas pesquisas de intenção de voto, ou seja, aquelas que serão realizadas antes das eleições. As pesquisas de Boca de Urna (realizada com o leitor que acabou de votar), não serão consideradas neste artigo, pois são proibitivas pelo TSE.

Numa pesquisa de campo considera-se a população alvo (conjunto de interesse) todo e qualquer eleitor, acima de 15 anos, apto a votar naquela eleição (eleitores que fizeram o título de eleitor). Numa pesquisa eleitoral o tempo e os custos tornam-se impraticáveis quando nos referimos a consulta a todos indivíduos da população, logo para resolver esse problema entrevistamos uma parte representativa dessa população, chamada de amostra.

Estatisticamente uma boa amostra é toda e qualquer parcela que tenha as mesmas características da população, para isso, deve-se levar em conta alguns fatores para a definição da amostra:

- Nível sócio-econômico;
- Grau de escolaridade;
- Sexo;
- Faixa de idade.

A escolha desses fatores é na maior parte em função de pesquisas realizadas no passado, podendo muitas vezes refletir na opinião pessoal do pesquisador que usa um determinado fator em suas pesquisas.

3 MÉDIA E PROPORÇÃO

A média é um dos conceitos estatísticos mais utilizados, aparecendo com certa frequência no nosso dia-a-dia, nas revistas, nos jornais, na internet e na televisão. Se abrirmos uma página de um jornal e nela encontrarmos uma manchete dizendo 3 em cada 5 brasileiros preferem jogar futebol a jogar voleibol, isso é uma média.

O uso massivo da média é talvez responsável pelas diversas interpretações errôneas que são dadas ao conceito de média, dando a ela poderes que ela não tem, como, por exemplo, como um homem de um metro e oitenta poderia ter morrido afogado num rio cuja profundidade média era de um metro e cinquenta?

É importante lembrar que a definição de média, está relacionada ao centro de gravidade, ou seja, é o valor central de uma distribuição, ou ainda é o ponto de equilíbrio de um conjunto de valores. O conhecimento de apenas a média de uma distribuição não nos dá uma informação precisa sobre ela, ou seja, não podemos saber como os demais valores se comportam em relação à média. Para medir essa variação ou dispersão, a medida utilizada para tal é a variância.

A variação de uma distribuição nunca será negativa e a determinação positiva da raiz quadrada da variância recebe o nome de desvio padrão. Embora existam infinitas distribuições com mesma média e mesma variância, a média e a variância nos permitirão tirar conclusões gerais sobre o comportamento da distribuição.

Quando dispomos de informações adicionais, como o comportamento da distribuição dos dados (por exemplo, normalmente distribuídos), a média mais

ou menos dois desvios padrões conterà no mínimo 95% dos valores da distribuição. Desta forma fica agora bem mais fácil entender o porque que uma pessoa de um metro e oitenta morreu afogado em um rio com profundidade média de um metro e cinqüenta.

Uma outra forma de expressar um valor em relação a um conjunto de dados é o uso de proporções, que vem a ser uma característica (k) em relação ao tamanho da amostra (n), $p = \frac{k}{n}$, e é essa proporção que é usada nas pesquisas eleitorais. Nos jornais que publicam as pesquisas eleitorais informam que de acordo com o instituto TAL, o candidato M tem 42% das intenções de voto e que a margem de erro é de 2,5 pontos percentuais para mais ou para menos.

Essa informação nos diz que o a amostra coletada pelo instituto que realizou a pesquisa, 42% dos entrevistados manifestaram sua preferência pelo candidato M, e que, com uma probabilidade conhecida, que quase nunca é declarada e que geralmente vale 95%, o valor real da proporção de eleitores de M está compreendido entre 39,5% e 44,5%.

4 INTERVALO DE CONFIANÇA

Para determinar o intervalo, precisam-se definir alguns elementos que compõem a fórmula estatística:

- p** – proporção de eleitores que pretendem votar no candidato M (valor estritamente positivo e menor que 1);
- n** – número de eleitores (tamanho da amostra);
- k** – número de eleitores que se manifestaram votar em M;

A proporção de eleitores da amostra que pretendem votar no candidato é dado por $p^* = \frac{k}{x}$, onde certamente uma outra amostra produzirá outro valor para p^* . Utilizando a distribuição binomial podemos mostrar que a média de p^* é igual a p e que a variância é dada por $\frac{p(1-p)}{n}$. Um resultado teórico importante nos permite mostrar que para uma amostra grande, p^* tem uma distribuição aproximadamente normal. Consultando uma tabela normal padronizada temos que: $P(-1,96 < z < +1,96) = 95\%$.

Segue-se então que a probabilidade de que o intervalo

$$p^* \pm 1,96 \sqrt{\frac{p(1-p)}{n}}$$

contenha o verdadeiro valor de p é aproximadamente igual a 95%. O problema que nos resta é que os extremos desse intervalo dependem do valor desconhecido de p . Uma solução para isso é aumentar o intervalo, substituindo $p(1-p)$ pelo seu valor máximo, ou seja, $p = 0,5$ e $q = 0,5$. Logo $p(1-p) = 0,25$.

Pode-se então afirmar que a probabilidade de que o intervalo

$$p^* \pm 1,96 \sqrt{\frac{p(1-p)}{n}} = p^* \pm \frac{1,96}{2\sqrt{n}}$$

contenha o verdadeiro valor de p é no mínimo de 95%.

Assim, se desejarmos uma margem de confiança de 95% e uma margem de erro de 2,5 pontos percentuais (para mais ou para menos) n deverá satisfazer:

$$\frac{1,96}{2\sqrt{n}} = \frac{2,5}{100}$$

E portanto n deverá ser igual a $1.536,64 = 1.537$ eleitores.

Para determinar o intervalo de confiança temos que inter-relacionar três unidades:

- 1) O tamanho da amostra n .
- 2) A precisão da estimativa (amplitude do intervalo ou margem de erro).
- 3) A Confiança depositada no intervalo que é definida pela probabilidade que o intervalo contenha o verdadeiro valor de p .

Assim, se o tamanho da amostra permanecer fixo, um aumento da precisão implica necessariamente numa diminuição da confiança. A única forma de melhorar a precisão sem alterar a confiança é aumentando o tamanho da amostra.

Finalmente, é importante observar que a confiança e a precisão estão relacionadas com o tamanho da amostra, logo se quisermos manter a confiança e reduzir o intervalo pela metade, precisaremos de uma amostra quatro vezes maior que a proposta inicialmente, logo o preço a ser pago pela melhoria da precisão nem sempre será compensado pelo tempo de coleta.

5 COLETA DA AMOSTRA

Nas pesquisas eleitorais um dos erros prováveis, mas impossível de ser evitado, é que um dos processos utilizados pela maioria dos Institutos de Pesquisas consiste em entrevistar pessoas escolhidas entre as que passam pelos pontos mais movimentados das grandes cidades.

Lembramos que existe um número maior de pessoas que raramente ou nunca passam por esses pontos. Se esses dois grupos tiverem opiniões diferentes sobre a eleição, os resultados finais serão totalmente distorcidos.

Infelizmente esse é um erro inevitável, onde a adoção de um plano de amostragem por domicílio, teria um custo proibitivo e um tempo bastante elevado, e isso os patrocinadores das pesquisas não iriam aceitar em hipóteses alguma.

6 PESQUISA IBOPE - 2002

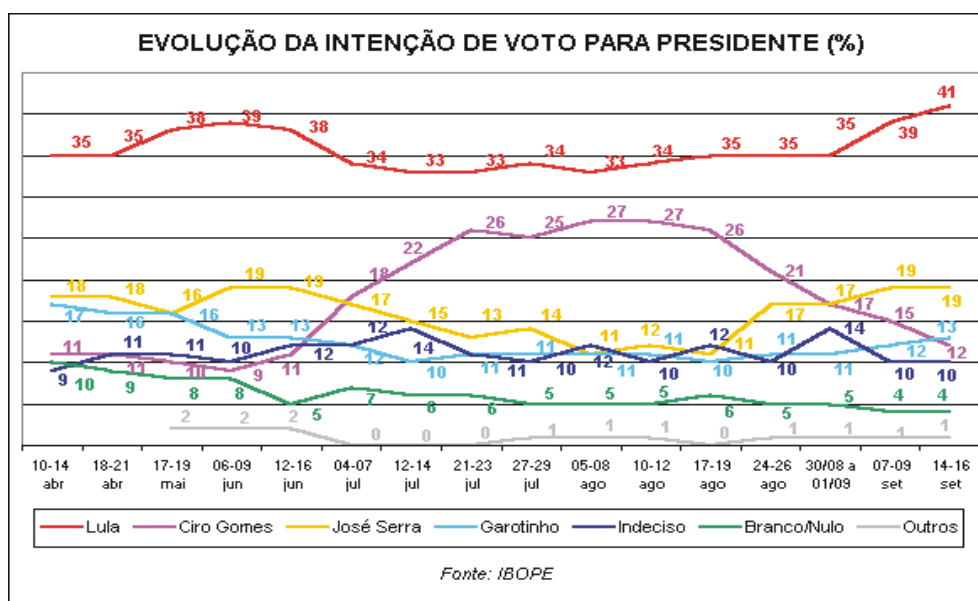


FIGURA 1 – Intenção de votos para presidente

BRASIL			
VOTOS VÁLIDOS PARA PRESIDENTE			
Data de campo	26/Out	27/Out	TSE
Número de entrevistas	2.000	52.170	
Candidatos	%	%	%
Lula	62	63	61
José Serra	38	37	39
	100	100	100

BRASIL			
VOTOS TOTAIS PARA PRESIDENTE			
Data de campo	26/Out	27/Out	TSE
Número de entrevistas	2.000	52.170	
Candidatos	%	%	%
Lula	58	59	58
José Serra	36	35	36
Indeciso/Branco/Nulo	6	6	6
	100	100	100

Fonte: IBOPE

FIGURA 2 – Resultados de uma pesquisa realizada pelo IBOPE

ÍNDICES DE ACERTO NO 2º TURNO DE 2002			
ÍNDICES DE ACERTO	GOVERNOS ESTADUAIS		TOTAL DE ESTIMATIVAS
	Pesquisas de Antevéspera e de Véspera (25 e 26/10)	Pesquisas de Boca de Urna (27/10)	
Total de Candidatos	26	16	42
Candidatos com percentuais dentro da margem de erro	22 85%	14 88%	36 86%
Candidatos com percentual fora da margem da erro, mas colocação correta	4 15%	2 12%	6 14%
Candidatos fora da margem da erro e com colocação errada	0 0%	0 0%	0 0%
% Total de Acerto, considerando a margem de erro ou a colocação do candidato	100%	100%	100%

Fonte: IBOPE

FIGURA 3 – Índices de acerto do IBOPE para o segundo turno de 2002

7 CONCLUSÃO

Neste trabalho procurou-se dar uma idéia dos problemas enfrentados pelos Institutos de Pesquisas, durante a realização de uma pesquisa de intensão de voto, mostrou-se os possíveis erros que poderão acontecer bem como as fórmulas utilizadas para determinar o tamanho da amostra.

Os defensores da idéia de que as pesquisas sejam proibidas, argumentam que a divulgação das pesquisas teria uma forte influência nos resultados finais da eleição.

Não há dúvida que deverá existir um controle rigoroso por parte da imprensa, pelo TSE e pela sociedade. É necessário que as informações sobre determinada pesquisa seja disponibilizada para todos os interessados, permitindo que a seriedade seja avaliada.

A idéia de proibir a realização das pesquisas ou impedir a sua divulgação nos parece não ser aceitável, pois de uma forma ou de outra as pesquisas continuam a ser realizadas e divulgadas de uma forma camuflada nos meios de comunicação, onde a sociedade não privilegiada seria a grande prejudicada.

8 REFERÊNCIAS BIBLIOGRÁFICAS

BARBETTA, P. A. **Estatística Aplicada às Ciências Sociais**. UFSC, Florianópolis, v.1, 1994.

_____, **Estatística Aplicada às Ciências Sociais**. UFSC, Florianópolis, v.2, 2000.

BUSSAB, W. O. & MORETTIN, P. A. **Estatística Básica**. 5. ed., Atual: São Paulo, 2002.

COSTA NETO, P. L. O. **Estatística**, 2. ed., São Paulo: Edgard Blücher, 2002.

FONSECA, J. S. & MARTINS, G. A. **Curso de Estatística**. 6. ed., Atlas: São Paulo, 1996.