

## O Tamanho da Amostra na Pesquisa de Mercado

*Luiz Fernando Kiehl \**

*1. Introdução. 2. Margem de Erro e Grau de Confiança. 3. A Fórmula. 4. Aplicações da Fórmula. 5. Número de Entrevistas — Margem de Erro. 6. Comentários.*

O número de livros de pesquisas de mercado disponíveis em língua portuguesa é bastante limitado e, na maioria dos casos, tratam-se de traduções pouco fiéis ao original.

Além das traduções deficientes, acrescenta-se o fato de que a maioria dos autores, ao tratarem do tamanho da amostra, aprofunda-se demais no assunto, supondo que o leitor tenha um conhecimento amplo de estatística. Por essa razão esse capítulo dos livros de pesquisas de mercado costuma tornar-se pouco claro, dando a impressão de que o cálculo do tamanho da amostra é coisa complicadíssima, só acessível a meia dúzia de iniciados.

Outros autores, sabendo dessas dificuldades, evitam o assunto apresentando apenas informações baseadas na experiência acumulada.

\* Bacharel em Administração de Empresas pela Escola de Administração de Empresas de São Paulo, da Fundação Getúlio Vargas e Gerente de Pesquisa de Mercado da Sandoz do Brasil S. A.

Meu trabalho, como contato e analista de uma agência de pesquisas de mercado e como gerente de pesquisas em duas grandes empresas, levou-me à observação de que as pessoas ligadas à aprovação de estudos de mercado tendem a evitar a discussão sobre esse tema, talvez com medo de revelar sua insegurança no assunto.

As relações entre o tamanho do universo e o tamanho da amostra assim como os conceitos de margem de erro e grau de confiança não costumam ser mencionados nos diálogos entre o cliente e a agência de pesquisa ou entre o gerente de pesquisas de mercado e seus superiores hierárquicos.

A conseqüência disso é a aprovação de pesquisas com amostras de tamanho exagerado e, portanto, desnecessariamente caras ou com amostras de tamanho insuficiente e, por conseguinte, com resultados pouco dignos de confiança.

O objetivo deste trabalho é esclarecer um pouco alguns aspectos desse tema, tornando-o compreensível para os não-especialistas em pesquisas de mercado. Não procurei inovar, mas apenas apresentar o assunto em forma mais didática.

## **2. Margem de Erro e Grau de Confiança**

Tomemos por exemplo, uma cidade que tenha apenas 20 quarteirões. Essa cidade é nosso **universo**. Para simplificar vamos admitir que todos os quarteirões tenham o mesmo número de residências.

Suponhamos que queríamos saber qual a porcentagem de lares que possuem aparelhos de TV. Um levantamento completo (visitas a todas as casas) revelou que naquela cidade 70% das residências tinha televisão e 30% não tinham. No gráfico 1 apresentamos um esboço daquela cidade. Dentro de cada quarteirão temos a porcentagem das casas que não tinham televisão. Observamos que em 19 dos 20 quarteirões a porcentagem das casas que tinham televisão estava entre 67% e 73%. Apenas em um quarteirão esta porcentagem era menor. Talvez se tratasse de um quarteirão mais pobre.

Se, em vez de fazer um levantamento completo em toda a cidade, visitando todas as casas, o pesquisador resolvesse sortear um só quarteirão e visitasse apenas as residências do quarteirão sorteado, ele teria 95% de probabilidade (19 em 20 quarteirões) de que o quarteirão sorteado fosse representativo da cidade. Em 19 quarteirões a porcentagem de lares que tinham TV estava entre 67% e 73%, ou seja, a porcentagem era de  $70\% \pm 3\%$ .

Se por azar o quarteirão sorteado fosse o quarteirão 4 do gráfico 1 a pesquisa revelaria que a porcentagem de casas com televisão naquela cidade era 60%, o que não representaria a realidade.

No nosso exemplo, o grau de confiança da amostra é 95% e a margem de erro é de 3%.

Grau de confiança pode ser definido como a probabilidade de que a amostra seja representativa do universo. No nosso exemplo, 19 dos 20 quarteirões (95% dos quarteirões) eram representativos do universo.

Margem de erro é a diferença entre o valor obtido através da amostra e o valor real do universo. No nosso exemplo o valor real do universo era de 70% de casas com TV. Em 19 dos 20 quarteirões a diferença entre o valor real do universo (70%) e o valor obtido com a amostra era igual ou menor do que 3%.

Observe-se, portanto, que margem de erro e grau de confiança estão intimamente relacionados. Falar em margem de erro sem mencionar o grau de confiança não tem nenhum significado. É preciso saber qual a probabilidade de que o valor levantado pela amostra seja igual ao valor real do universo com a margem de erro mencionada.

No nosso exemplo, se o quarteirão 2 tivesse servido de amostra o pesquisador diria que a porcentagem de lares com aparelhos de TV naquela cidade era de  $72\% \pm 3\%$ , com 95% de confiança. E teria acertado, porque  $72\% \pm 3\%$  significa que o valor real está entre 69% e 75% e, na realidade, o valor real era 70%.

Gráfico 1

Amostra					} A 73% 27%
1	2	3	4	5	
S = 70% N = 30%	S = 72% N = 28%	S = 68% N = 32%	S = 60% N = 40%	S = 68% N = 32%	
6	7	8	9	10	
S = 69% N = 31%	S = 71% N = 29%	S = 70% N = 30%	S = 73% N = 27%	S = 72% N = 28%	
11	12	13	14	15	
S = 73% N = 27%	S = 72% N = 28%	S = 67% N = 33%	S = 70% N = 30%	S = 71% N = 29%	
16	17	18	19	20	
S = 73% N = 27%	S = 70% N = 30%	S = 67% N = 33%	S = 69% N = 31%	S = 70% N = 30%	
Universo = Cidade					

### 3. A Fórmula

Embora existam diversas fórmulas mais perfeitas e mais sofisticadas, muitos pesquisadores de mercado costumam utilizar tanto para o cálculo do número de entrevistas como para o cálculo da margem de erro, a fórmula:

$$\sigma_{\bar{x}} = \sqrt{\frac{p \cdot q}{n}} \text{ onde:}$$

$p$  = porcentagem de sim

$q = 1 - p$  = porcentagem de não

$n$  = número de entrevistas

$\sigma_{\bar{x}}$  = margem de erro de  $p$ . com 68% de segurança

Quase tôdas as perguntas podem ser respondidas com sim ou não. No nosso exemplo a porcentagem de sim (o  $p$ . da fórmula) é 70%. Isto é, 70% de respostas sim à pergunta "O Sr. tem TV em casa?". Mesmo quando a resposta envolve escolha múltipla,

como no caso de se perguntar “Qual a marca da sua televisão?”, o índice de possuidores da marca M será a porcentagem de **sim — marca M** e os possuidores de qualquer outra marca de TV. podem ser agrupados como **não — marca M**.

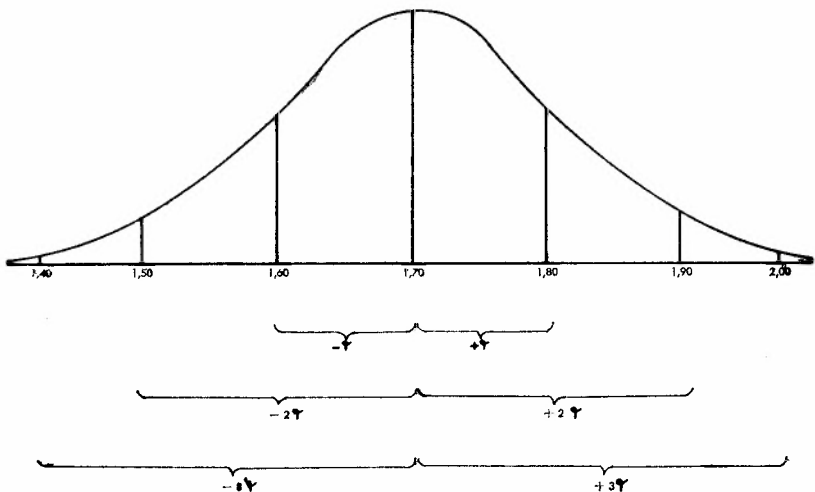
Tendo-se o valor de  $p$ , tem-se o valor de  $q$ , que é complemento de  $p$ , pois  $p + q = 1$  (isto é, porcentagem de sim + porcentagem de não = 100%).

Para se entender o que significa  $\sigma \bar{x}$  é preciso recorrer à Curva Normal:

A Teoria Estatística demonstra que, na maioria dos casos, os dados que compõem um universo se distribuem proporcionalmente em tórno da média do universo.

Tomemos, por exemplo, os homens de uma cidade e verifiquemos a altura de cada um. A altura média dos homens daquela cidade será, por exemplo, 1,70 m. A freqüência das alturas dos homens se distribuirá mais ou menos como apresentamos no gráfico 2.

Gráfico 2



A Lei de Gauss demonstra que 99%, (99,7% para ser exato) dos homens terá uma altura de  $1,70 \text{ m} \pm 3 \sigma$ ; que 95% dos ho-

mens terá uma altura de  $1,70 \text{ m} \pm 2 \sigma$  e que 68% deles terá uma altura de  $1,70 \text{ m} \pm \sigma$ , sendo  $\sigma$  o chamado desvio-padrão.<sup>1</sup>

Supondo que, neste exemplo, o desvio-padrão ( $\sigma$ ) seja 10 cm (0,10m), então 68% dos habitantes masculinos da cidade terão entre 1,60 m e 1,80 m de altura, 95% deles terão entre 1,50 m e 1,90 m de altura e quase todos (99%) estarão na faixa de 1,40 m a 2,00 m de altura. O 1% restante é composto de anões e gigantes.

Voltando à nossa fórmula, se o valor de  $\sigma \bar{x}$  fôr 3%, significa que, em 68% das amostras que forem sorteadas, a diferença entre o valor obtido através da amostra e o valor real do universo será igual ou menor do que 3%.

Em outras palavras, teremos 68% de confiança de que a amostra sorteada é representativa do universo.

Se acharmos que 68% de confiança é muito pouco, pois quase que uma em cada três amostras sorteadas não será representativa do universo, podemos usar dois desvios-padrões. Neste caso,  $2 \sigma \bar{x} = 6\%$  ( $2 \times 3\%$ ).

Isto significa que em 95% das amostras sorteadas a diferença entre o valor obtido através da amostra e o valor real do universo será igual ou menor do que 6%.

Se quisermos ter uma confiança quase absoluta usamos três desvios-padrões e, neste caso,  $3 \sigma \bar{x} = 9\%$  ( $3 \times 3\%$ ), e quase podemos jurar (99% de confiança) que o valor real do universo é o valor obtido através da amostra com uma margem de erro de 9%.

#### 4. Aplicações da Fórmula

Esta fórmula que apresentamos, aplica-se em duas situações práticas:

<sup>1</sup> O desvio-padrão é igual à raiz quadrada da variância. Por sua vez, a variância é igual à média dos quadrados das diferenças entre cada elemento da população e a média da população.

A primeira é a determinação do tamanho da amostra, ou seja, quantas entrevistas devem ser realizadas. Para isto, antes de tudo, é necessário estabelecer a margem de erro e o grau de confiança desejados para a pesquisa em pauta.

Digamos, por exemplo, que foi determinado em comum acordo com o gerente de Marketing da empresa que uma margem de erro de  $\pm 10\%$  com 95% de confiança é suficiente como base para a decisão que será tomada em função dos resultados da pesquisa.

Então temos que:  $2 \sigma_{\bar{x}} = 10\%$  e, portanto,  

$$\sigma_{\bar{x}} = 5\% \quad (10\% \div 2)$$

Como não temos nenhuma idéia de que valores de  $p$  e  $q$  a pesquisa revelará, adotamos por hipótese  $p = 50\%$  e  $q = 50\%$ . A razão desta escolha é que estes são os valores cujo produto ( $p \cdot q$ ) é maior, pois  $50 \times 50 = 2.500$ .

Substituindo-se na fórmula  $\sigma_{\bar{x}} = \sqrt{\frac{p \cdot q}{n}}$

$$\text{teremos} \quad 5 = \sqrt{\frac{50 \cdot 50}{n}} \quad e$$

$$25 = \frac{2.500}{n} \quad e$$

$$n = \frac{2.500}{25} = 100.$$

Descobrimos, assim, que com apenas 100 entrevistas teremos 95% de confiança de que o valor de  $p$  revelado com esta amostra será no máximo 10% diferente do valor real do universo.

Se por exemplo a pesquisa fôsse sobre hábitos de uso de gelatina e se a tabulação das respostas revelasse que 55% das donas-de-casa costumam servir gelatina como sobremesa, teríamos 95% de confiança que no total do universo pesquisado 55%  $\pm \pm 10\%$  das donas-de-casa costumam servir gelatina como sobremesa. Isto é, entre 45% e 65% das donas-de-casa.

A segunda aplicação prática desta fórmula é a determinação da margem de erro de uma pesquisa já feita. Suponhamos, por

exemplo, que foi feita uma pesquisa sobre o consumo de cigarros. O número de entrevistas realizadas, estabelecido por qualquer critério não científico, foi de 500 entrevistas. A tabulação dos questionários revelou que 80% das pessoas têm o hábito de fumar e 20% não fumam.

Então temos:

$$p = 80$$

$$q = 20$$

$$n = 500$$

Substituindo-se na fórmula:

$$\sigma_{\bar{x}} = \sqrt{\frac{p \cdot q}{n}} \quad \text{temos}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{80 \cdot 20}{500}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{1.600}{500}}$$

$$\sigma_{\bar{x}} = \sqrt{3,2}$$

$$\sigma_{\bar{x}} = 1,8$$

Se  $\sigma_{\bar{x}} = 1,8$ ;  $2\sigma_{\bar{x}} = 3,6$  e  $3\sigma_{\bar{x}} = 5,4$ .

O analista poderá então afirmar que naquele universo:

temos 68% de confiança de que a porcentagem de fumantes seja de  $80\% \pm 1,8\%$  (ou seja, entre 78,2% e 81,8%);

temos 95% de confiança de que a porcentagem de fumantes seja de  $80\% \pm 3,6\%$  (ou seja, entre 76,4% e 83,6%);

temos 99% de confiança de que a porcentagem de fumantes seja de  $80\% \pm 5,4\%$  (ou seja, entre 74,6% e 85,4%).

O que se faz na prática é combinar estas duas aplicações da fórmula. Por exemplo, suponhamos que desejamos determinar a porcentagem de consumidores de uísque numa certa cidade com uma margem de erro de 6%, com 95% de segurança. Como



não sabemos nada sobre os hábitos de consumo de tal cidade, adotamos  $p = q = 50\%$  para calcular o tamanho da amostra necessária.

Temos então que  $2\sigma_{\bar{x}} = 6$  e  $\sigma_{\bar{x}} = 3$ ; na fórmula  $\sigma_{\bar{x}} = \sqrt{\frac{p \cdot q}{n}}$  teremos:

$$3 = \sqrt{\frac{50 \cdot 50}{n}}$$

$$3 = \sqrt{\frac{2.500}{n}}$$

$$9 = \sqrt{\frac{2.500}{n}}$$

$$n = \frac{2.500}{9} = 277 \text{ entrevistas.}$$

Terminado o trabalho de campo a tabulação das respostas revelou que apenas 10% da população costuma tomar uísque.

Temos então que  $p = 10$  e  $q = 90$ . Substituindo na fórmula teremos:

$$\sigma_{\bar{x}} = \sqrt{\frac{10 \cdot 90}{277}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{900}{277}}$$

$$\sigma_{\bar{x}} = \sqrt{3,25}$$

$$\sigma_{\bar{x}} = 1,8 \quad \text{e, portanto,}$$

$$2 \sigma_{\bar{x}} = 3,6$$

Isto quer dizer que, embora nos satisfizesse uma margem de erro de 6% com 95% de confiança, a amostra de 277 entrevistas forneceu, como resultado, que a porcentagem de consumidores de uísque do universo é de  $10\% \pm 3,6\%$  com 95% de confiança ou seja, que no universo a porcentagem de consumidores de uísque está entre 6,4% e 13,6% da população.

E mais, com quase certeza absoluta (99% de confiança) a porcentagem de consumidores de uísque na cidade é de  $10\% \pm 5,4\%$ , porque  $3 \sigma_{\bar{x}} = 5,4$ .

### 5. Número de Entrevistas — Margem de Erro

Sabendo-se que a parte mais custosa de uma pesquisa de mercado é a entrevista, é óbvio que se deseje que o número de entrevistas seja tão pequeno quanto possível. O leigo em estatística, tende a acreditar que quanto maior o número de entrevistas, menor a margem de erro da pesquisa e mais exatos os resultados obtidos. Isto é certo apenas em parte porque a margem de erro não diminui em relação diretamente proporcional ao aumento do número de entrevistas.

Apenas a título de exemplo apresentamos na tabela 1 uma relação do número de entrevistas de uma amostra e sua margem de erro com 68%, 95% e 99% de confiança, admitindo  $p$  e  $q = 50\%$ . O leitor pode observar através desta tabela que, à medida em que aumentamos o número de entrevistas, a margem de erro do resultado da pesquisa tende a se manter constante, isto é, a diferença entre a margem de erro obtida com um certo número de entrevistas e a margem de erro obtida com o dobro de entrevistas tende a ser cada vez menor.

Tabela 1

N.º de Entrevistas	$\psi\bar{x}$ (68%)	$2\psi\bar{x}$ (95%)	$3\psi\bar{x}$ (99%)
50	7,0%	14,0%	21,0%
100	5,0%	10,0%	15, %
200	3,5%	7,0%	10,5%
300	2,8%	5,7%	8,5%
500	2,2%	4,5%	6,7%
800	1,7%	3,5%	5,2%
1.111	1,5%	3,0%	4,5%

## 6. Comentários

O leitor deve ter observado que, embora tenhamos mencionado o universo da pesquisa em vários parágrafos deste artigo, nenhuma vez falamos em **tamanho do universo**. Isto porque, ao contrário do que as pessoas tendem a supor, o tamanho da amostra não tem nada a ver com o tamanho do universo, desde que este seja suficientemente grande.

Aliás, na prática da pesquisa de mercado, o tamanho do universo geralmente não é conhecido pelo pesquisador. Tanto faz que o universo seja composto por donas-de-casa, por crianças em idade escolar ou por adultos fumantes. Desde que este universo seja razoavelmente grande, seu tamanho não influi no tamanho da amostra necessária e o conhecimento de sua dimensão pode ser ignorado.

Nos casos em que o universo é muito pequeno utiliza-se a fórmula

$$\sigma_{\bar{x}} = \sqrt{\frac{p \cdot q}{n} \frac{(N - n)}{n}}, \text{ sendo } N = \text{o tamanho do universo. Nes-}$$

ta fórmula,  $\frac{N - n}{N}$  funciona como um fator de correção.

Cumprе ainda ressaltar que, geralmente, o questionário de uma pesquisa de mercado contém perguntas-filtro que reduzem o tamanho da amostra da pergunta seguinte.

Se, por exemplo, numa pesquisa de 300 entrevistas a primeira pergunta do questionário é "A Sra. costuma usar batom?" e apenas 40% das entrevistadas respondem sim, a próxima pergunta: "Que marca de batom a Sra. costuma usar?" será respondida por apenas 120 entrevistadas.

Para o cálculo da margem de erro das respostas a esta pergunta o analista deve lembrar-se de que  $n$  (número de entrevistas), a ser usado na fórmula, é 120, e não as 300 iniciais.

É importante ter em mente que cada pergunta tem uma margem de erro que depende do número de respondentes a esta pergunta e não do número total de entrevistados. O que se faz na prática

é estabelecer uma margem de erro desejado para cada uma das perguntas mais importantes do questionário.

Gostaríamos ainda, de lembrar o leitor de que esta fórmula pressupõe que as amostras sejam sorteadas equiprobabilisticamente, isto é, que a cada componente do universo seja dada uma igual oportunidade de ser sorteado.

Na realidade, a fórmula apresentada neste trabalho só pode ser considerada rigorosamente correta quando aplicada à amostragem probabilística simples.

No caso de amostragem por conglomerados (amostragem por áreas), por exemplo, a teoria estatística recomenda o uso de uma fórmula um pouco mais sofisticada, que corrige os possíveis erros oriundos do fato de que cada unidade da amostra (quartel-rão) é composta, na verdade, por vários elementos (residências).

Entretanto, na prática diária, as diferenças entre os resultados obtidos através do uso de fórmulas mais simples ou mais complexas não costumam ser grandes e a tendência dos profissionais de pesquisas de mercado é usar a fórmula apresentada neste trabalho, principalmente pela sua facilidade de manejo.

## **Bibliografia**

- ADLER, Max K. *A Moderna Pesquisa de Mercado*. 2ª edição. São Paulo, Livraria Pioneira Editôra, 1964.
- BOYD, H. e WESTFALL, R. *A Pesquisa Mercadológica*. Rio de Janeiro, Fundação Getúlio Vargas, 1964.
- BOUQUEREL, Fernand. *L'Étude des Marchés au Service des Entreprises*. 2ª edição. Paris, Presses Universitaires de France, 1964.
- CHEVRY, Gabriel R. *Pratique des Enquêtes Statistiques*. Paris, Presses Universitaires de France, 1962.
- COCHRAN, William G. *Técnicas de Amostragem*. Rio de Janeiro, Editôra Fundo de Cultura S. A., 1965.
- FERBER, Blankertz & HOLLANDER JR. *Marketing Research*. Nova Iorque, The Ronald Press Company, 1964.
- LIOZOU, A. *Iniciación Práctica a la Estatística*. Barcelona, Francisco Casanovas Editor, 1959.