

DETERMINAÇÃO DO TAMANHO DE UMA AMOSTRA

INTRODUÇÃO

O pesquisador social procura tirar conclusões a respeito de um grande número de sujeitos. Por exemplo, ele poderia desejar estudar:

- os 170.000.000 de cidadãos que constituem a população brasileira.
- Os 1.000 membros de um sindicato.
- Os 45.000 estudantes de intercâmbio e assim sucessivamente.

Se o pesquisador trabalha com todo o grupo que ele tenta compreender, dizemos que está trabalhando com a POPULAÇÃO.

População *consiste em um conjunto de indivíduos que compartilham de, pelo menos, uma característica comum, seja ela cidadania, filiação a uma associação de voluntários, etnia, matrícula na universidade, etc.*

Entretanto, o pesquisador trabalha com tempo, energia e recursos econômicos limitados. Portanto, são raras as vezes em que pode trabalhar com todos os elementos da POPULAÇÃO. Geralmente, o pesquisador estuda um pequeno grupo de indivíduos retirados da população. Este grupo denomina-se AMOSTRA [Levin, 1987].

Amostra *é um subconjunto de indivíduos extraídos de uma população.*

O processo de escolha dos indivíduos que pertencerão a uma AMOSTRA, é denominado AMOSTRAGEM.

O pesquisador busca generalizar conclusões referentes à AMOSTRA, estendendo-as para toda a POPULAÇÃO da qual essa amostra foi extraída.

Há diversos MÉTODOS DE AMOSTRAGEM. Para o pesquisador social, interessam os métodos que permitem que qualquer indivíduo da POPULAÇÃO possa vir a fazer parte da AMOSTRA. Estes métodos de amostragem são denominados PROBALÍSTICOS.

Métodos de Amostragem Probabilística são os que selecionam os indivíduos da população de forma que todos tenham as mesmas chances de participar da amostra.

Não há dúvida de que uma amostra não representa perfeitamente uma população. Ou seja, a utilização de uma amostra implica na aceitação de uma margem de erro que denominaremos ERRO AMOSTRAL.

Erro Amostral é a diferença entre um resultado amostral e o verdadeiro resultado populacional; tais erros resultam de flutuações amostrais aleatórias

Ocorrem erros não-amostrais quando:

- Os dados amostrais são coletados, registrados ou analisados incorretamente.
- Há uma utilização de um instrumento defeituoso durante a realização de mensurações.
- Um questionário ou formulário possui questões formuladas de modo tendencioso [Triola, 1999].

Não podemos evitar a ocorrência do ERRO AMOSTRAL, porém podemos limitar seu valor através da escolha de uma amostra de tamanho adequado. Obviamente, o ERRO AMOSTRAL e o TAMANHO DA AMOSTRA seguem sentidos contrários (Figura 1). Quanto maior o tamanho da amostra, menor o erro cometido e vice-versa.

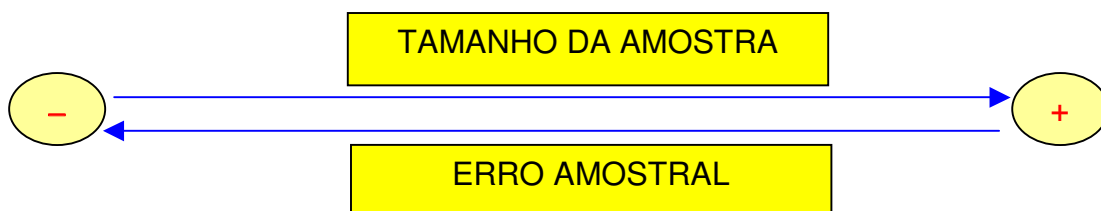


Figura 1 – Relação intuitiva entre o tamanho da amostra e o erro amostral.

DETERMINAÇÃO DO TAMANHO DE UMA AMOSTRA COM BASE NA ESTIMATIVA DA MÉDIA POPULACIONAL

Suponha, por exemplo, que queiramos estimar a renda média de pessoas que concluíram um curso superior, no primeiro ano após a formatura. QUANTAS rendas devemos incluir em nossa amostra? A determinação do tamanho de uma amostra é problema de grande importância, porque:

- amostras desnecessariamente grandes acarretam desperdício de tempo e de dinheiro;
- e amostras excessivamente pequenas podem levar a resultados não confiáveis.

Em muitos casos é possível determinar o tamanho mínimo de uma amostra para estimar um parâmetro estatístico, como por exemplo, a MÉDIA POPULACIONAL (μ) .

A fórmula para cálculo do tamanho da amostra para uma estimativa confiável da MÉDIA POPULACIONAL (μ) é dada por:

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

Equação 1

Onde:

n = Número de indivíduos na amostra

$Z_{\alpha/2}$ = Valor crítico que corresponde ao grau de confiança desejado.

σ = Desvio-padrão populacional da variável estudada (no exemplo, RENDA).

E = Margem de erro ou ERRO MÁXIMO DE ESTIMATIVA. Identifica a diferença máxima entre a MÉDIA AMOSTRAL ($\bar{X}$) e a verdadeira MÉDIA POPULACIONAL.

Os valores de confiança mais utilizados e os valores de Z correspondentes podem ser encontrados na Tabela 1:

Tabela 1 – Valores críticos associados ao grau de confiança na amostra

Grau de Confiança	α	Valor Crítico $Z_{\alpha/2}$
90%	0,10	1,645
95%	0,05	1,96
99%	0,01	2,575

EXEMPLO 1

Um economista deseja estimar a renda média para o primeiro ano de trabalho de um bacharel em direito. Quantos valores de renda devem ser tomados, se o economista deseja ter 95% de confiança em que a média amostral esteja a menos de R\$500,00 da verdadeira média populacional? Suponha que saibamos, por um estudo prévio, que para tais rendas, $\sigma = \text{R\$}6250,00$.

SOLUÇÃO

Queremos determinar o tamanho n da amostra, dado que $\alpha = 0,05$ (95% de confiança). Desejamos que a média amostral seja a menos de R\$ 500 da média populacional, de forma que $E = 500$. Supondo $\sigma = 6250$, aplicamos a Equação 1, obtendo:

$$n = \left(\frac{Z_{\alpha/2}}{E} \right)^2 = \left(\frac{1,96 \cdot 6250}{500} \right)^2 = 600,25 = 601 \quad (\text{Arredondado para cima})$$

Devemos, portanto, obter uma amostra de ao menos 601 rendas de primeiro ano, selecionadas aleatoriamente, de bacharéis de faculdades que tenham feito um curso de direito. Com tal amostra teremos 95% de confiança em que a média amostral $\bar{x}$ difira em menos de R\$500,00 da verdadeira média populacional μ .

EXPERIÊNCIA: *Baseado nos dados do EXEMPLO 1, utilize uma margem de erro maior, como R\$1.000,00 e determine qual seria o tamanho da amostra necessário nesta situação.*

E se σ não for conhecido?

A Equação 1 exige que se substitua por algum valor o desvio-padrão populacional σ , mas se este for desconhecido, devemos poder utilizar um valor preliminar obtido por processos como os que se seguem:

1. Utilizar a aproximação $\sigma \approx \text{amplitude}/4$.
2. Realizar um estudo piloto, iniciando o processo de amostragem. Com base na primeira coleção de pelo menos 31 valores amostrais selecionados aleatoriamente, calcular o desvio-padrão amostral S e utilizá-lo em lugar de σ . Este valor pode ser refinado com a obtenção de mais dados amostrais.

EXERCÍCIOS PROPOSTOS – Série 1

1. Uma pesquisa é planejada para determinar as despesas médicas anuais das famílias dos empregados de uma grande empresa. A gerência da empresa deseja ter 95% de confiança de que a média da amostra está no máximo com uma margem de erro de $\pm \$50$ da média real das despesas médicas familiares. Um estudo-piloto indica que o desvio-padrão pode ser calculado como sendo igual a \$400.
 - a. Qual o tamanho de amostra necessário?
 - b. Se a gerência deseja estar certa em uma margem de erro de $\pm \$25$, que tamanho de amostra será necessário?
2. O teste de QI padrão é planejado de modo que a média seja 100 e o desvio-padrão para adultos normais seja 15. Ache o tamanho da amostra necessária para estimar o QI médio dos instrutores de estatística. Queremos ter 99% de confiança em que nossa média amostral esteja a menos de 1,5 pontos de QI da verdadeira média. A média para esta população é obviamente superior a 100, e o desvio-padrão é provavelmente inferior a 15, porque se trata de um grupo com menor variação do que um grupo selecionado aleatoriamente da população geral; portanto, se tomamos $\sigma = 15$, estaremos sendo conservadores, por utilizarmos um valor

que dará um tamanho de amostra no mínimo tão grande quanto necessário.
Suponha $\sigma = 15$ e determine o tamanho da amostra necessário.

□

DETERMINAÇÃO DO TAMANHO DE UMA AMOSTRA COM BASE NA ESTIMATIVA DA PROPORÇÃO POPULACIONAL

Outro parâmetro estatístico cuja determinação afeta o tamanho da amostra é a proporção populacional. Tomemos, como exemplo, a necessidade de determinar a proporção de pessoas atendidas por uma Unidade de Saúde, originárias do município de Cariacica.

A fórmula para cálculo do tamanho da amostra para uma estimativa confiável da PROPORÇÃO POPULACIONAL (p) é dada por:

$$n = \frac{Z_{\alpha/2}^2 \cdot p \cdot q}{E^2}$$

Equação 2

Onde:

- | | | |
|----------------|---|--|
| n | = | Número de indivíduos na amostra |
| $Z_{\alpha/2}$ | = | Valor crítico que corresponde ao grau de confiança desejado. |
| p | = | Proporção populacional de indivíduos que pertence a categoria que estamos interessados em estudar. |
| q | = | Proporção populacional de indivíduos que NÃO pertence à categoria que estamos interessados em estudar ($q = 1 - p$). |
| E | = | Margem de erro ou ERRO MÁXIMO DE ESTIMATIVA. Identifica a diferença máxima entre a PROPORÇÃO AMOSTRAL e a verdadeira PROPORÇÃO POPULACIONAL (p). |

E se “p” e “q” não forem conhecidos?

A Equação 2 exige que se substituam os valores populacionais **p** e **q**, por valores amostrais $\hat{p}$ e $\hat{q}$. Mas se estes também forem desconhecidos, substituímos $\hat{p}$ e $\hat{q}$ por 0,5, obtendo a seguinte estimativa (Levine, 2000):

$$n = \frac{Z_{\alpha/2}^2 \cdot 0,25}{E^2}$$

Equação 3

EXEMPLO 2

Um assistente social deseja saber o tamanho da amostra (**n**) necessário para determinar a proporção da população atendida por uma Unidade de Saúde, que pertence ao município de Cariacica. Não foi feito um levantamento prévio da proporção amostral e, portanto, seu valor é desconhecido. Ela quer ter 90% de confiança que sua **erro máximo de estimativa (E)** seja de $\pm 5\%$ (ou 0,05). Quantas pessoas necessitam ser entrevistadas?

SOLUÇÃO

Considerando que o valor da proporção amostral de atendimentos para pessoas de Cariacica não é conhecida. Utilizamos a Equação 3 para determinar o tamanho da amostra. Sabemos que, para 90% de confiança teremos o valor crítico ($Z_{\alpha/2}$) = 1,645, conforme Tabela 1.

$$n = \frac{[Z_{\alpha/2}]^2 \cdot 0,25}{E^2} = \frac{1,645^2 \cdot 0,25}{0,05^2} = 270,6 = 271 \text{ (arredondado para cima)}$$

Devemos, portanto, obter uma amostra de 271 pessoas para determinar a proporção da população atendida na Unidade de Saúde, que se origina do município de Cariacica.

EXPERIÊNCIA: Baseado nos dados do EXEMPLO 2, utilize uma margem de erro maior, como $\pm 0,20$ (20%) e determine qual seria o tamanho da amostra necessário quando o nível de confiança é 90% e quando é 95%.

DETERMINAÇÃO DO TAMANHO DA AMOSTRA PARA POPULAÇÕES FINITAS

As fórmulas para determinação do tamanho da amostra que vimos até agora trabalhavam com a idéia de que a população de onde se retirava a amostra era tão grande, que poderíamos considerá-la infinita. Entretanto, a maior parte das populações não é tão grande em comparação com as amostras. Caso a amostra tenha um tamanho (**n**) maior ou igual a 5% do tamanho da população (**N**), considera-se que a população seja **FINITA**. Neste caso, aplica-se um fator de correção às fórmulas vistas anteriormente e teremos as seguintes fórmulas corrigidas:

- Fórmula para determinação do tamanho da amostra (**n**) com base na estimativa da média populacional.

$$n = \frac{N \cdot \sigma^2 \cdot (Z_{\alpha/2})^2}{(N-1) \cdot E^2 + \sigma^2 \cdot (Z_{\alpha/2})^2}$$

Equação 4

- Fórmula para determinação do tamanho da amostra (**n**) com base na estimativa da proporção populacional.

$$n = \frac{N \cdot \hat{p} \cdot \hat{q} \cdot (Z_{\alpha/2})^2}{\hat{p} \cdot \hat{q} \cdot (Z_{\alpha/2})^2 + (N-1) \cdot E^2}$$

Equação 5

REFERÊNCIA BIBLIOGRÁFICA

- LEVIN, Jack. **Estatística Aplicada a Ciências Humanas**. 2a. Ed. São Paulo: Editora Harbra Ltda, 1987.
- TRIOLA, Mário F. **Introdução à Estatística**. 7a. Ed. Rio de Janeiro: LTC, 1999.
- LEVINE, D. M. / BERENSON, M. L. / STEPHAN, David. **Estatística: Teoria e Aplicações usando Microsoft Excel em Português**. Rio de Janeiro: LTC, 2000.