ATUALIZAÇÃO/REVISÃO

METODOLOGIA DA PESQUISA: DIRETRIZES PARA O CÁLCULO DO TAMANHO DA AMOSTRA[1]

SCIENTIFIC RESEARCH METHODOLOGY: GUIDELINES FOR SIZE SAMPLE CALCULATION

Mauro José FONTELLES[2], Marilda Garcia SIMÕES[3], Jairo Cunha de ALMEIDA[4] e Renata Garcia Simões FONTELLES[5]

RESUMO

Objetivo: mostrar as diretrizes para a escolha do método de amostragem e para o cálculo do tamanho amostral, utilizados em projetos de pesquisa científica nas áreas das Ciências Biológicas e da Saúde. **Método:** realizado um criterioso levantamento bibliográfico na literatura científica, a partir da compilação de trabalhos publicados em revistas científicas e livros especializados e em bases de dados da rede BIREME. **Conclusão:** o entendimento dos diferentes tipos de amostragem, assim como o cálculo correto para o tamanho da amostra, são pontos fundamentais para o sucesso na realização de uma pesquisa científica.

Descritores: metodologia científica, amostragem, tamanho amostral.

INTRODUÇÃO

Tradicionalmente, o objetivo de todo projeto de pesquisa é, a partir do estudo de uma amostra, fazer inferências para uma determinada população. Logo, para que a inferência estatística seja válida, é necessário que a amostra selecionada seja representativa da população de onde foi retirada, de maneira que os resultados encontrados sejam os mais fidedignos possíveis.[1,2] Assim, é necessário que o autor do projeto atente para fatores importantes, tais como o método de amostragem e o cálculo do tamanho amostral, pois, amostras mal selecionadas e de tamanho inadequado, comprometem o resultado da pesquisa, uma vez que não representam fielmente a população.[3,4,5]

Neste sentido, é importante realçar que o cálculo do tamanho amostral tem sido um dos maiores desafios para aqueles que desejam conduzir um experimento científico, pois nem sempre os métodos para esse cálculo se apresentam de maneira simples e compreensíveis, o que traz ansiedade e dúvidas ao pesquisador.[6,7,8]

Assim, o objetivo deste estudo é mostrar, de maneira objetiva, as diretrizes para o cálculo do tamanho de amostras utilizadas em projetos de pesquisa científica nas áreas das Ciências Biológicas e da Saúde.

MÉTODO

Estudo desenvolvido no Núcleo de Bioestatística Aplicada à Pesquisa da Universidade da Amazônia – UNAMA. Para tanto, procedeu-se a um criterioso levantamento bibliográfico na literatura científica, a partir da compilação de trabalhos publicados em revistas científicas, livros especializados e em bases de dados da rede BIREME.

---

1 Trabalho realizado no Núcleo de Bioestatística Aplicado à pesquisa da Universidade da Amazônia – UNAMA.
2 TCBC – Doutor em Cirurgia do Trauma. Coordenador do Núcleo de Pesquisa em Saúde e Professor Titular da Disciplina de Bioestatística da Universidade da Amazônia - UNAMA.
3 M.Sc. Engenheira de Alimentos. Professora do Curso de Tecnologia de Alimentos da Universidade do Estado do Pará - UEPA.
4 Graduando do Programa de Mestrado em Doenças Tropicais, do Núcleo de Doenças Tropicais da Universidade Federal do Pará - UFPA.
5 Graduanda do Curso de Especialização em Próteses da Faculdade de Odontologia de Piracicaba – FOP-UNICAMP.

FATORES QUE AFETAM O TAMANHO DA AMOSTRA[2,5,7,9,10,11]

• Objetivo da amostra - Estudos descritivos costumam exigir amostra com menor número de participantes.

• Tipo de variável - As variáveis qualitativas exigem amostras maiores que as variáveis quantitativas, que, exigem amostras maiores quanto maior for a variação nos dados amostrais.

• Delineamento do estudo - Estudo pareado requer uma amostra com metade do número de sujeitos, quando comparados aos estudos não-pareados.

• Valor estimado para erro alfa (erro tipo I) - Corresponde ao erro máximo que o pesquisador aceita cometer ao aplicar o teste estatístico para aceitar ou rejeitar a hipótese nula. É o erro máximo que ele aceita para um erro falso-positivo. Na área das ciências da saúde é estipulado em 5%. Quanto menor o erro alfa estipulado pelo pesquisador, maior será o tamanho estimado para a amostra.

• Poder do teste estatístico (1–erro β) - Corresponde à probabilidade de que o estudo detecte uma diferença real entre os grupos estudados. Traduz a probabilidade de o pesquisador cometer um erro falso-negativo. Na área das ciências da saúde é arbitrado em 80%, 85% ou 90%, que corresponde a um erro beta de 20%, 15% e 10%, respectivamente. Quanto maior o tamanho da amostra, maior será o poder do estudo em detectar uma diferença ou um efeito real.

• O tamanho da diferença - Corresponde ao tamanho da verdadeira diferença que se deseja discriminar como significativa, entre as médias da variável considerada no estudo. Pequenas diferenças exigem amostras maiores.

• O tamanho da população – Para pequenas populações o tamanho da amostra é diretamente proporcional ao tamanho da população. Para grandes populações, o tamanho da amostra não é influenciado pelo tamanho da população, pois a mesma deverá ser considerada como ilimitada.

• Dos recursos e do tempo disponível – É outro fator limitante que, não menos importante, pode influenciar no tamanho da amostra.

ETAPAS PARA O CÁLCULO DO TAMANHO DA AMOSTRA[2,5,7,9]

a. Primeiro passo - Escolher a fórmula apropriada dependendo do tipo de estudo (analítico ou descritivo) e do tipo de erro (alfa ou beta).

b. Segundo passo - Especificar os valores dos parâmetros que serão utilizados. São eles:

• Variância esperada (s^2) - Deve ser obtida com base em conhecimentos prévios sobre o estudo a ser realizado. No caso de variáveis contínuas, esta pode ser estimada com base em estudos semelhantes publicados na literatura, ou pela realização de um estudo piloto previamente executado.

• Erro alfa ($z\alpha$) - Usualmente, na área das ciências da saúde, é estimado em 5% ou 1%.

• Erro beta ($z\beta$) - Usualmente, é considerado em 20%, 15% ou 10%. Quanto menor o erro beta estipulado, maior o poder do teste.

• Diferença estimada entre os grupos ($\bar{d}$) - Corresponde à diferença mínima a ser detectada entre a média da amostra ($\bar{x}$) e a verdadeira média da população (μ).

• Variância das proporções esperadas (p) - Se o parâmetro a ser estudado é uma proporção, digamos, a proporção de sucesso para um determinado tratamento, e assumindo-se que os grupos são iguais no tamanho, o pesquisador deve determinar a proporção média (p) no estudo, ou seja, no grupo inteiro. A fórmula para calcular a variância das proporções é $p = \bar{p}(1 - \bar{p})$.

Nas fórmulas, valores do erro alfa e do erro beta, arbitrados pelo pesquisador, devem ser introduzidos com base nos valores de $z\alpha$ e $z\beta$, determinados na tabela de valores críticos da distribuição normal gaussiana, conforme o Quadro 1, abaixo, sendo $z\alpha$, rotineiramente, bicaudal, e $z\beta$, unicaudal.

Quadro 1 – valores calculados de z_α e z_β conforme o valor do erro alfa e do erro beta estipulados para um estudo

Erro α	Bicaudal ($z_{\alpha/2}$)	Poder do teste ($1-\beta$)	Unicaudal (z_β)
0,01	2,58	0,95	2,33
0,05	1,96	0,90	1,64
0,10	1,64	0,85	1,28
0,20	1,28	0,80	0,84

CÁLCULOS PARA O TAMANHO DA AMOSTRA[2,5,7,9,12]

Fórmula para calcular o tamanho da amostra (n) para que se obtenha uma estimativa confiável da média populacional (μ):

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

Onde:

$z\alpha/2$ = Valor de z na curva normal segundo o valor α.

σ = Desvio padrão populacional da variável estudada.

E=Diferença máxima estimada entre a média amostral ($\bar{x}$) e a verdadeira média populacional (μ). É a margem de erro ou erro máximo de estimativa.

Nas populações ilimitadas deve-se realizar um estudo piloto com uma amostra aleatória de pelo menos 31 indivíduos da população; calcular o desvio padrão (s) dessa amostra e substituir, na fórmula, o valor σ pelo valor s; ou utilizar, para o desvio padrão, um valor extraído da literatura; ou ainda, utilizar um valor aproximado para o desvio padrão, o qual é dado por: σ   amplitude/4.

Por outro lado, uma variante desta fórmula é aquela que utiliza, também, o erro beta. Deste modo, temos dois tipos de fórmula: as que utilizam somente o erro alfa e as que utilizam o erro alfa e o erro beta, concomitantemente, como proposto por Snedecor & Cochran, 1967 e Steel & Torrie, 1980.[13,14]

• Com erro alfa: $$\boldsymbol{n} = \frac{(z_{\alpha/2})^2 . (\boldsymbol{s})^2}{(\bar{\boldsymbol{x}} - \mu)^2}$$

Onde: $(\bar{x} - \mu)$ = Diferença máxima estimada entre a média amostral e a verdadeira média populacional.

• Com erro alfa e beta: $$\boldsymbol{n} = \frac{(z_{\alpha/2} + z_{\beta})^2 . (\boldsymbol{s})^2}{(\bar{\boldsymbol{x}} - \mu)^2}$$

O Quadro 2 mostra os valores calculados para $(z\alpha+z\beta)^2$ e $(z\alpha/2+z\beta)^2$, segundo Snedecor & Cochran, 1967.[13,14]

Quadro 2 – valores calculados para $(z\alpha+z\beta)^2$ e $(z\alpha/2+z\beta)^2$

	Testes bicaudais $(z_{\alpha/2}+z_{\beta})^2$			Testes unicaudais $(z_{\alpha}+z_{\beta})^2$		
Poder do teste (P)	Nível alfa			Nível alfa		
	0,01	0,05	0,10	0,01	0,05	0,10
0,80	11,7	7,9	6,2	10,0	6,2	4,5
0,90	14,9	10,5	8,6	13,0	8,6	6,6
0.95	17,8	13,0	10,8	15,8	10,8	8,6

Note que, com base nas fórmulas apresentadas, ao calcular o tamanho da mostra, o pesquisador deverá considerar os seguintes pontos:

• Como a variância (s^2) encontra-se no numerador da fórmula, quanto maior for o seu valor, maior será o tamanho da amostra necessária para detectar uma grande variação na estimativa do parâmetro estudado;

• Para que o estudo seja confiável, o pesquisador necessita escolher um nível de significância (erro alfa) pequeno. Isto implica em um valor elevado para zα, pois, quanto menor o erro alfa, maior será o valor crítico na tabela normal. Como este valor encontra-se no numerador da fórmula, quanto menor o nível de significância, maior será o valor do ***n*** amostral;

• Se a diferença entre a média amostral dos dois grupos estudados, ou entre a média amostral e a verdadeira média da população, a ser detectada, for pequena, isto irá requerer um tamanho amostral maior.

• Ao erro beta deve ser dada especial atenção. Se em uma determinada pesquisa, para um nível alfa preestabelecido (5%, p. ex.), o investigador encontra uma diferença estatisticamente significativa entre as médias da variável estudada, não há necessidade de investigar o erro beta. Porém, se num experimento, o investigador esperava encontrar uma diferença clinicamente significativa entre as médias dos grupos estudados, mas essa diferença estatística não ocorreu, mesmo que os dados induzam o contrário, o erro beta pode ter ocorrido. Nesse caso, o pesquisador deverá aumentar o tamanho da amostra, a qual deve ser recalculada com a fórmula que utiliza ambos os erros, alfa e beta;

• As fórmulas apresentadas são aplicadas em estudos que utilizam o teste t de Student pareado. Em estudos com dois ou mais grupos (experimental e controle, p. ex.), o tamanho da amostra deve ser calculado para cada um dos grupos isoladamente. Porém, se o delineamento exige que os grupos sejam do mesmo tamanho, basta multiplicar o resultado calculado pelo número total de grupos.

CÁLCULO DO “n” AMOSTRAL PARA ESTUDOS DESCRITIVOS

• Cálculo de n para estimar a média da população (μ)

$$\boldsymbol{n} = \frac{\boldsymbol{s}^2}{(\bar{\boldsymbol{x}} - \mu)^2} \times \left(\boldsymbol{t}_{\alpha;gl}\right)^2$$

Onde:

$\bar{x}$ = Média da amostra e μ = média da população.

tα = valor crítico da tabela ***t*** de Student e ***gl*** = n – 1.

O valor da variância (s^2) deve ser obtida da literatura ou de um estudo piloto. É necessário estabelecer uma diferença máxima razoável entre a média obtida da amostra e a verdadeira média da população $(\bar{x} - \mu)$

O valor de $t_{\alpha;gl}$ não pode ser diretamente estimado, pois depende do erro alfa e dos graus de liberdade (gl = n –1). Como não se tem ***n*** (o qual se quer calcular), deve-se escolher um tamanho amostral provisório (n_0) para calcular *gl* e, então, obter *tα;gl*. Os valores provisórios de gl e tα;gl obtidos são, então, reintroduzidos na fórmula para que um novo valor de *n* seja encontrado, o qual é utilizado para se obter um novo valor de *tα;gl*. Este procedimento é repetido até que o valor de *n* se estabilize, tal como mostrado no Exemplo 1, a seguir.

Exemplo 1: A síndrome metabólica é caracterizada por um conjunto de fatores de risco cardiovasculares relacionados com resistência à insulina e com a obesidade abdominal. Quantos pacientes devem ser avaliados para se conhecer a média dos valores dos triglicerídeos séricos em pacientes portadores desta síndrome?

Para calcular $\boldsymbol{n}$, o pesquisador necessita da variância (s^2), a qual pode ser obtida da literatura ou de um projeto piloto. Suponhamos que o investigador realizou um projeto piloto com 30 pacientes e encontrou uma taxa de triglicerídeos igual a 170 ± 31 mg/dL (média ± desvio padrão). Esse desvio padrão (s) pode ser utilizado como uma estimativa provisória para o cálculo de $\boldsymbol{n}$.

Agora, é necessário estabelecer uma diferença máxima razoável, admitida entre a média obtida da amostra e a verdadeira média da população $(\bar{x} - \mu)$. Digamos que o pesquisador estabeleceu essa diferença em 10 mg/dL.

O erro alfa estipulado foi de 5% e o $\boldsymbol{n_0} = 30$. Logo, *gl(n – 1) = 29*. Assim, quando se busca, na tabela, os valores críticos da distribuição *t* de Student, temos: $t_{\alpha;gl} = t_{0,05;29} = 2{,}045$. Substituindo os valores na fórmula, vamos obter:

$$\boldsymbol{n}_1 = \frac{s^2}{(\bar{x}-\mu)^2} \times \left(t_{\alpha;gl}\right)^2 = \frac{(31)^2}{(10)^2} \times (2{,}045)^2$$

$$\boldsymbol{n}_1 = 9{,}61 \times 4{,}182 = 40{,}19$$

Com o valor obtido (40,19), considera-se a aproximação para o inteiro imediatamente superior, daí $\boldsymbol{n_1} = 41$. Logo, se $\boldsymbol{n_1} = 41$, temos $\boldsymbol{gl} = 40$ e $t_{0,05;40} = 2{,}021$.
Assim, para calcular $\boldsymbol{n_2}$, temos:

$$\boldsymbol{n}_2 = \frac{s^2}{(\bar{x}-\mu)^2} \times \left(t_{\alpha;gl}\right)^2 = \frac{(31)^2}{(10)^2} \times (2{,}021)^2$$

$$\boldsymbol{n}_2 = 9{,}61 \times 4{,}084 = 39{,}24$$

Com o valor de $\boldsymbol{n_2} = 40$, temos $t_{\alpha;gl}\neg = t_{0,05;39} = 2{,}021$, logo o valor de $\boldsymbol{n_3} = 40$. Ou seja, o valor de $\boldsymbol{n}$ fica estabilizado em 40. Portanto, o pesquisador necessitará de uma amostra de 40 pacientes para estimar, com 95% de confiança, a média dos valores de triglicerídeos séricos em pacientes com síndrome metabólica. Como o tamanho da amostra é calculado com base em suposições, o pesquisador poderá modificá-lo simplesmente alterando o valor escolhido para o erro alfa ou para a diferença entre os valores médios da amostra e a verdadeira média da população $(\bar{x} - \mu)$.

Uma maneira mais fácil de calcular o tamanho da amostra é usar o z_α em vez de $t_{\alpha;gl}$. Isto pode produzir uma discreta subestimação no tamanho amostral, mas, na prática, este método tem sido usado rotineiramente, como demonstrado a seguir:

$$\boldsymbol{n} = \frac{s^2}{(\bar{x}-\mu)^2} \times (z_{\alpha/2})^2 = \frac{(31)^2}{(10)^2} \times (1{,}96)^2$$

$$\boldsymbol{n} = 9{,}61 \times 3{,}84 = 36{,}90 \quad \therefore \quad \boldsymbol{n} = 37$$

• Cálculo de n para estimar o coeficiente de correlação

Fórmula para o cálculo do tamanho amostral de um estudo que correlaciona duas variáveis paramétricas (contínuas), como peso x estatura, por exemplo.

$$\boldsymbol{n} = \frac{(1-r^2) \times (t_{\alpha;gl})^2}{r^2} + 2, \quad \text{onde: } \boldsymbol{gl} = \boldsymbol{n} - 2$$

É necessário que o investigador suponha um dado valor para a correlação, o qual pode ser encontrado na literatura ou em um pequeno estudo piloto. O nível alfa deve ser previamente estipulado.

Exemplo 2: Um pesquisador deseja investigar a correlação entre o peso e a estatura de crianças, ao nascer. Uma pesquisa piloto revelou um coeficiente de correlação (r) igual a 0,7. Considere o erro alfa igual 0,05% e $\boldsymbol{n_0} = 10$. Então $\boldsymbol{gl} = 8$ e $\boldsymbol{t_{\alpha;gl}} = \boldsymbol{t_{0,05;8}} = 2{,}306$.

$$\boldsymbol{n_1} = \frac{(1-0{,}7^2) \times (2{,}306)^2}{0{,}7^2} + 2$$

$$\boldsymbol{n_1} = \frac{0{,}51 \times 5{,}317}{0{,}49} + 2 = 7{,}53$$

Logo $\boldsymbol{n_1} = 8$. Novos cálculos deverão ser efetuados até que o valor de n estabilize, como demonstrado no exemplo anterior. Se o erro beta for considerado em 20%, o valor encontrado para $\boldsymbol{n_1}$ será igual a 16.

• Cálculo de *n* para estimar uma proporção na população

Fórmula utilizada para estimar a proporção de uma variável dicotômica, como proporção de sobreviventes sim x não.

$$\boldsymbol{n} = \frac{P(1-P).z^2_{\alpha/2}}{(p-P)^2}$$

Onde: ***P***(1–***P***) = Variância das proporções na população.

(***p–P***) = Diferença mínima a ser detectada entre o valor da proporção esperada ***(p)*** e o verdadeiro valor na população ***(P)***.

Exemplo 3: Deseja-se conhecer, em uma população, qual a proporção de sobreviventes depois de transcorridos cinco anos do diagnóstico de uma determinada doença. Qual deve ser o tamanho amostral para se detectar uma diferença mínima entre a proporção esperada ***(p)*** e a verdadeira proporção na população ***(P)*** igual a 10%, considerando-se uma proporção de sobreviventes = 40% e α = 0,05.

$$\boldsymbol{n} = \frac{0,40(1-0,40)(1,96)^2}{(0,1)^2}$$

$$\boldsymbol{n} = \frac{0,24 \times 3,84}{0,01} = 92,16 \quad \therefore \quad \boldsymbol{n} = 93$$

O investigador necessita de uma amostra constituída por 93 indivíduos da população.

CÁLCULO DO "n" AMOSTRAL PARA ESTUDOS ANALÍTICOS

• Cálculo de n para o teste t de Student considerando os erros alfa e beta

a. Fórmula para o tamanho amostral de um estudo que compare as médias de dois grupos independentes, considerando as variâncias iguais nos dois grupos.

$$\boldsymbol{n} = \frac{2.(\boldsymbol{s})^2}{(\boldsymbol{\bar{d}})^2} \times (z_{\alpha/2} + z_{\beta})^2$$

Onde: $z\alpha/2$ = Valor do erro alfa (bicaudal).
$z\beta$ = Valor do erro beta.
s = desvio padrão.
$\boldsymbol{\bar{d}}$ = Diferença mínima a ser detectada.

Exemplo 4: Em um município, deseja-se avaliar a estatura média de crianças de escolas públicas e compará-las com a de crianças de escolas particulares. Qual deve ser o tamanho da amostra para que se possa identificar, com 95% de confiança (erro α = 0,05), uma diferença, se houver, de pelo menos 5 cm, entre as médias dos valores da estatura dos dois grupos de crianças? Um estudo piloto mostrou desvio padrão (***s***) = 12 cm. Considere o poder do teste de 80% (erro β = 0,20).

$$\boldsymbol{n} = \frac{2.(12)^2}{(5)^2} \times (1,96 + 0,84)^2$$

$$\boldsymbol{n} = \frac{2 \times 144}{25} \times (7,84) = \frac{2.257,92}{25} = 90,31 \ \therefore \ \boldsymbol{n} = 91$$

Logo o pesquisador irá precisar de uma amostra de 182 crianças, o que corresponde a 91 crianças para cada grupo estudado.

b. Fórmula para comparar as médias de dois grupos independentes, considerando as variâncias desiguais nos dois grupos. Em ambas as fórmulas, pode-se substituir $z_{\alpha/2}$ e z_{β} por $t_{\alpha;gl}$ e $\boldsymbol{t_{\beta}}$, respectivamente, e adotar o procedimento descrito no primeiro exemplo, que trata do cálculo do tamanho amostral para estimar a média da população ***(μ)***.

$$\boldsymbol{n} = \frac{\boldsymbol{s}_A^2 + \boldsymbol{s}_B^2}{(\boldsymbol{\bar{d}})^2} \times (z_{\alpha/2} + z_{\beta})^2 \qquad \boldsymbol{s}_A^2 \text{ e } \boldsymbol{s}_B^2 =$$

Onde: variâncias estimadas para as populações A e B.

• Cálculo de n para comparar duas proporções amostrais

$$\boldsymbol{n} = \frac{\left[z_{\alpha/2}\sqrt{2\boldsymbol{P_O Q_O}} + z_{\beta}\sqrt{\boldsymbol{P_A Q_A} + \boldsymbol{P_B Q_B}} \right]^2}{(\boldsymbol{P_A} - \boldsymbol{P_B})^2}$$

Onde: ***PA*** = Proporção no grupo A.
QA = Complemento de PA (1–PA).
$\boldsymbol{P_B}$ = Proporção no grupo B.
$\boldsymbol{Q_B}$ = Complemento de PB (1– PB).
$\boldsymbol{P_O}$ = (PA + PB)/2.
$\boldsymbol{Q_O}$ = Complemento de PO (1¬– PO).
$\boldsymbol{P_A – P_B}$ = Diferença mínima a ser detectada no estudo.

Exemplo 5: Um engenheiro químico deseja comparar dois tipos de embalagem para um determinado alimento, com o objetivo de testar a efetividade de ambas por um período de seis meses. Observou que a embalagem A conservou cerca de 60% do alimento testado, enquanto que a embalagem B conservou cerca de 80%, para o mesmo período de tempo. Quais devem ser os tamanhos amostrais para que o investigador possa demonstrar que esta diferença apresenta significância estatística, considerando α = 0,05 e poder do teste = 0,80.

Dados do problema:
$\boldsymbol{P_A}$ = 0,60 (proporção para a embalagem A)
$\boldsymbol{P_B}$ = 0,80 (proporção para a embalagem B)
$\boldsymbol{P_O}$ = (0,60 + 0,80)/2 = 0,70
α = 0,05. Logo z0,05 = 1,96
Poder do teste = 0,80. Logo z_{β} = 0,84

$$\boldsymbol{n} = \frac{\left[1,96\sqrt{0,42} + 0,84\sqrt{0,24+0,16} \right]^2}{(0,6-0,8)^2}$$

$$\boldsymbol{n} = \frac{(1,270 + 0,531)^2}{(-0,2)^2} = 81,09 \ \therefore \ \boldsymbol{n} = 82$$

O pesquisador precisará de uma amostra de 164 embalagens, o que corresponde a 82 embalagens do tipo A e 82 do tipo B.

Nota: Algumas vezes, o pesquisador é obrigado a calcular diferentes tamanhos amostrais em delineamentos experimentais que envolvam duas amostras distintas. Neste caso, devem-se determinar quantas vezes uma amostra será maior que a outra para que os tamanhos amostrais possam ser ajustados. Esse ajuste é feito com base no quadro de Kirkwood (1988),15 mostrado a seguir.

Quadro 1 – Fator de ajuste para o tamanho de duas amostras

c*	Fator de ajuste**	c*	Fator de ajuste**
2	3/4	7	4/7
3	2/3	8	9/16
4	5/8	9	5/9
5	3/5	10	11/20
6	7/12	--	--

* Indica quantas vezes uma amostra é maior que a outra
** Usado para calcular o tamanho da amostra menor

Exemplo 6 – No exemplo anterior, suponha que, para cada 4 embalagens do tipo A, é utilizada apenas 1 embalagem do tipo B, dando relação 4:1. Como deve proceder, o pesquisador, para que a amostra da embalagem A ***(n_A)*** seja o quádruplo da embalagem B ***(n_B)***?

Observe que, para a condição acima proposta, para c = 4 (amostra maior = 4 x amostra menor), o fator de ajuste corresponde a 5/8. Assim, o valor calculado para n será multiplicado pelo fator de ajuste 5/8 para determinar a amostra menor ***(n_B)***. A amostra maior ***(n_A)*** será calculada multiplicando-se o valor da amostra menor ***(n_B)*** pelo valor de 'c'. Assim, temos:

$$\text{Amostra calculada}(\boldsymbol{n}) = 82$$

$$\text{Amostra menor}(\boldsymbol{n}_B) = \boldsymbol{n} \times \frac{5}{8} = 82 \times \frac{5}{8} = 51{,}2 \therefore \boldsymbol{n}_B = 52$$

$$\text{Amostra maior}(\boldsymbol{n}_A) = \boldsymbol{n}_B \times c = 52 \times 4 = 208$$

O pesquisador precisará de uma amostra de 260 embalagens, o que corresponde a 208 do tipo A e 52 do tipo B. Note que, para um mesmo tipo de teste estatístico, com tamanhos amostrais diferentes, o número total de embalagens estudadas é maior que aquele necessário se as amostras fossem de tamanhos iguais.

• Cálculo de *n* para comparar uma proporção amostral com uma proporção populacional

$$\boldsymbol{n} = \frac{\left[z_{\alpha/2}\sqrt{\boldsymbol{P_O Q_O}} + z_{\beta}\sqrt{\boldsymbol{P_A Q_A}} \right]^2}{(\boldsymbol{P_A} - \boldsymbol{P_O})^2}$$

Onde: P_A = Proporção amostral.

Q_A = Complemento de PA ($_{1-PA}$).

P_O = Proporção na população.

Q_O = Complemento de PO ($_{1-PO}$).

$P_A - P_O$ = Diferença mínima detectada no estudo.

Exemplo 6: Em uma região de agropecuária, um pesquisador deseja estimar a proporção de bovinos acometidos por uma doença. Sabe-se, pela literatura, que a prevalência da doença em questão é algo em torno de 12%. Um estudo piloto realizado na mesma região mostrou uma proporção de 18% dos animais selecionados. Qual deve ser o tamanho da amostra para que ele possa testar se a diferença encontrada é estatisticamente significante, considerando α = 0,05 e o poder do teste = 0,90.

Dados do problema:

P_A = 0,18 (proporção do projeto piloto).

P_O = 0,12 (proporção da literatura).

α = 0,05. Logo $_{z0,05}$ = 1,96.

Poder do teste = 0,90. Logo $_{z\beta}$ = 1,28.

$$\boldsymbol{n} = \frac{\left[1{,}96\sqrt{0{,}12 \text{ x } 0{,}88} + 1{,}28\sqrt{0{,}18 \text{ x } 0{,}82} \right]^2}{(0{,}18 - 0{,}12)^2}$$

$$\boldsymbol{n} = \frac{(0{,}637 + 0{,}492)^2}{(0{,}06)^2} = 354{,}08 \quad \therefore \quad \boldsymbol{n} = 355$$

O pesquisador precisará de uma amostra de 355 animais. Caso considere a amostra demasiadamente grande, ele pode, a critérios bem definidos, recalculá-la utilizando um poder do teste menor, como o usual 80%, o que aumentará o erro beta para 20%. Então, zβ = 0,84. Neste caso o novo valor amostral será igual a 256 animais.

CÁLCULO DO "n" PARA POPULAÇÕES LIMITADAS (FINITAS)[5,7,9,16]

• Cálculo do "n" para uma amostra aleatória simples

O cálculo do tamanho de uma amostra aleatória simples impõe ao pesquisador que ele especifique um valor predefinido para o erro amostral (margem de erro), o qual deve ser pensado em termos de probabilidade, pois, mesmo que uma amostra seja suficientemente grande, ela não garante que suas características sejam exatamente iguais a da população de onde foi retirada, uma vez que sempre existe a probabilidade da randomização gerar uma amostra bem diferente da população. Ou seja, a margem de erro exprime o valor de quanto o pesquisador admite errar na avaliação dos parâmetros estudados. Assim, temos:

$$n = \frac{N.n_0}{N + n_0}, \quad \text{onde: } n_0 = \frac{1}{E_0^2}$$

Sendo: N = Tamanho da população estudada.

n_0 = Primeiro valor aproximado do tamanho da amostra.

E_0 = Erro amostral (margem de erro).

Exemplo 7: Um gerente de uma pequena fábrica de medicamentos deseja realizar uma pesquisa para conhecer uma determinada característica de seus funcionários. Supondo que a fábrica emprega 50 pessoas, qual deve ser o tamanho mínimo da amostra aleatória simples para que ele possa realizar a pesquisa, admitindo um erro amostral de 5%?

$$n_0 = \frac{1}{E_0^2} = \frac{1}{(0,05)^2} = 400 \text{ funcionários}$$

$$n = \frac{50 \times 400}{50 + 400} = \frac{20.000}{450} = 44,45 \;\therefore\; n = 45 \text{ funcionários}$$

Com base no exemplo dado, e supondo-se que a fábrica tivesse 300 funcionários. Qual seria o tamanho da amostra?

$$n = \frac{50 \times 400}{50 + 400} = \frac{20.000}{450} = 44,45 \;\therefore\; n = 45 \text{ funcionários}$$

• Cálculo do "n" para uma proporção populacional conhecida

$$n = \frac{n_0}{1 + \frac{n_0}{N}}, \quad \text{onde: } n_0 = \left[\frac{z_{\alpha/2}}{E}\right]^2 .P_0(1 - P_0)$$

Sendo: N = Tamanho da população estudada.

n_0 = Valor aproximado do tamanho amostral.

P_0 = Proporção amostral.

E_0 = Erro amostral (margem de erro).

Exemplo 8: Em uma comunidade rural, cuja população é estimada em 20.000 habitantes, um pesquisador deseja conhecer a taxa de prevalência de uma determinada doença endêmica. Um levantamento preliminar (estudo piloto) mostrou uma taxa de 12%. Qual deve ser o tamanho da amostra para que o referido pesquisador possa estimar a verdadeira taxa na população, admitindo um nível de confiança de 95% e um erro de amostragem de 5%?

Dados do problema:

P_0 = 0,08 (proporção do estudo piloto).

α = 0,01. Logo $z_{0,01}$ = 2,25

E = 0,05

N = 20.000 habitantes.

$$n_0 = \left[\frac{1,96}{0,05}\right]^2 \times 0,12(0,88) = 168,96 \;\therefore\; n_0 = 169$$

$$n = \frac{169}{1 + \frac{169}{20.000}} = \frac{169}{1,0085} = 167,57$$

$$\therefore\; n = 168 \text{ indivíduos.}$$

O pesquisador necessitará de 168 indivíduos para estimar a verdadeira taxa de prevalência da doença em questão.

CONCLUSÃO

O entendimento dos diferentes tipos de amostragem, assim como o cálculo correto para o tamanho da amostra, são pontos fundamentais para o sucesso na realização de uma pesquisa científica.

SUMMARY

SCIENTIFIC RESEARCH METHODOLOGY: GUIDELINES FOR SIZE SAMPLE CALCULATION

Mauro José FONTELLES, Marilda Garcia SIMÕES, Jairo Cunha de ALMEIDA e Renata Garcia Simões FONTELLES

Objective: to show the methodological aspects and guidelines for the choice of sampling method and sample size calculation used in scientific research projects of Life Sciences and Health areas. **Method:** for the organization of this present review study, a detailed bibliographic research of the scientific literature has been made, from the compilation of papers published in scientific magazines and specialized books, as well as from databases of Bireme's server. **Conclusion:** Understanding about the different types of sampling methods, as well as the correct sample size calculation, are important steps to success in conducting a scientific research.

Key-words: scientific methodology, sampling method, simple size.

REFERÊNCIAS

1. Callegari-Jacques S - Bioestatística - princípios e aplicações. 1 ed. Porto Alegre: Artmed Editora, 2003.
2. Jekel JF, Elmore JG, Katz DL - Epidemiologia, Bioestatística e medicina preventiva. 2 ed. Porto Alegre, RS: Artmed: Editora, 2004.
3. Dawson B, Trapp RG - Bioestatística básica e clínica. 3 ed. Rio de Janeiro, RJ: McGraw-Hill Interamericana do Brasil Ltda., 2003.
4. Triola MF - Introdução à Estatística. 7 ed. Rio de Janeiro, RJ: Livros Técnicos e Científicos Editora S.A., 1999.
5. Hulley SB, Cummings SR, Browner WS et al - Delineando a pesquisa clínica – uma abordagem epidemiológica. 2 ed. Porto Alegre, RS: Artmed Editora, 2001.
6. Levy P, Lemeshow S - Sampling for health professionals. Belmont, LLP, 1980.
7. Motta VT, Wagner MB - Bioestatística. 1 ed. Caxias do Sul, RG: Robe Editorial, 2003.
8. Beiguelman B - Curso prático de bioestatística. 5. ed. Ribeirão Preto: FUNPEC-Editora, 2002.
9. Fontelles MJ – Bioestatística aplicada à pesquisa experimental. 1 ed. Belém, Pará: Edição do autor, 2010.
10. Downing D, Clark J - Estatística aplicada. 1 ed. São Paulo, SP: Editora Saraiva, 2000.
11. Neter J, Wasserman W, Kutner MH - Planning sample sizes, nonprametric tests, and random ANOVA model. In: _____ Applied linear statistical models: regression, analysis of variance and experimental designs. 3. ed. Boston, Irwin, 1990. p. 633-42.
12. Lwanga SK, lemeshow S - Sample size determination in health studies: a practical manual. Geneva, World Health Organization, 1991.
13. Snedecor GW, Cochran WG - Statistical methods. 6th ed. Iowa: Iowa State University Press, 1967. 505 p.
14. Steel, RGD, Torrie JH - Principles and procedures of statistics. 2nd ed. New York: Mc Graw-Hill, 1980.
15. Kirkwood BR - Essentials of medical statistics. Oxford: Blackwell, 1988.
16. Dupont WD - Power and sample size calculations. Controlled Clinical Trials, 1990; 11: 116-128.

Endereço para correspondência
Dr. Mauro José Fontelles
Rua Antônio Barreto, 983/1502 – Umarizal
Belém – Pará.
CEP 66055-050
Fone : (091)3225 1850
E.mail – mikefox@uol.com.br

Recebido em 17.12.2010 – Aprovado em 28.01.2011