

Tamanho da amostra em estudos clínicos e experimentais

Hélio Amante Miot*

Estudos investigativos clínico-epidemiológicos ou experimentais objetivam descrever fenômenos ou comparar o comportamento de variáveis em subgrupos de uma população. Para tanto, não se realiza o estudo de todo o universo populacional, usualmente porque não é acessível ou viável, porém, principalmente, porque não é necessário quando se dispõe de uma amostra representativa para a realização de inferências à população-alvo^{1,2}.

O planejamento amostral da pesquisa determina o dimensionamento numérico e também a técnica de amostragem (coleta/seleção) dos elementos de um estudo. É fundamental na elaboração do projeto, e seus problemas podem comprometer a análise final dos dados e interpretação dos resultados. O planejamento amostral adequado depende do conhecimento básico da estatística do estudo e do conhecimento profundo do problema investigado, a fim de que se possa unir a significância estatística dos testes ao significado clínico dos resultados^{1,3,4}.

A maior parte dos testes bioestatísticos pressupõe que a amostra estudada seja probabilisticamente representativa da população. Algumas amostras coletadas por conveniência, como a escolha de pacientes consecutivos de um ambulatório específico, podem não representar adequadamente toda a população do estudo. O pesquisador deve estar atento a possíveis vieses de seleção oriundos da disponibilidade de pacientes em amostragens consecutivas, já que a ampliação do tamanho amostral não corrige o efeito de amostras enviesadas. Além disso, estratégias de amostragens estratificadas

não probabilísticas, por quotas, complexas (conglomerados, multiníveis), por resposta voluntária, por saturação de variáveis, tipo “bola de neve” ou com sistemática de coleta não aleatorizada devem ser desenhadas, dimensionadas e analisadas com suporte de estatístico experiente. Neste texto serão discutidos princípios para cálculos do tamanho de amostras aleatórias simples⁴.

A escolha da fração populacional que compõe a amostra do estudo implica que o pesquisador assumirá certo grau de erro relacionado à estimativa dos parâmetros populacionais de cada variável, tal erro amostral é possível de quantificação, sendo inversamente proporcional ao tamanho da amostra^{4,5}.

Para descrever a estimativa populacional representada por uma variável quantitativa (discreta ou contínua), deve-se dispor do desvio padrão populacional da variável, selecionar o nível de significância da estimativa e o erro amostral (em unidades da média) máximo tolerado (Quadro 1)².

Para descrever a estimativa populacional representada por uma variável qualitativa (nominal ou ordinal), deve-se dispor da frequência populacional de resultados da variável, selecionar o nível de significância da estimativa e o erro amostral (em percentual) máximo tolerado (Quadro 1). Quando uma variável qualitativa não for dicotômica, o dimensionamento amostral deve ser considerado para a proporção de cada categoria que compõe a variável⁴.

Quando não se conhecer o desvio padrão ou as frequências populacionais da variável, e não se dispuser de dados

Quadro 1. Fórmulas para cálculo do tamanho de amostras para descrição de variáveis quantitativas e qualitativas em uma população.

	Variável quantitativa	Variável qualitativa
População infinita	$n = \left(\frac{Z_{\alpha/2} \cdot \delta}{E} \right)^2$	$n = \left(\frac{Z_{\alpha/2} \cdot \sqrt{p \cdot q}}{E} \right)^2$
População finita (<10000)	$n = \frac{N \cdot \delta^2 \cdot (Z_{\alpha/2})^2}{(N - 1) \cdot (E)^2 + \delta^2 \cdot (Z_{\alpha/2})^2}$	$n = \frac{N \cdot p \cdot q \cdot (Z_{\alpha/2})^2}{(N - 1) \cdot (E)^2 + p \cdot q \cdot (Z_{\alpha/2})^2}$

n – tamanho da amostra; $Z_{\alpha/2}$ – valor crítico para o grau de confiança desejado, usualmente: 1,96 (95%); δ – desvio padrão populacional da variável; E – erro padrão, usualmente: $\pm 5\%$ da proporção dos casos (precisão absoluta), ou $\pm 5\%$ da média ($1,05 \times$ média); N – tamanho da população (finita); p – proporção de resultados favoráveis da variável na população; q – proporção de resultados desfavoráveis na população ($q=1-p$).

semelhantes na literatura, deve-se realizar um pré-teste com 30-40 indivíduos e considerar o comportamento desse subgrupo como a estimativa populacional².

Ainda, as fórmulas para cálculo do tamanho amostral pressupõem populações de tamanho ilimitado. Uma situação especial ocorre quando se estudam populações restritas (<10000 indivíduos), já que, nesses casos, cada unidade amostrada representa uma significativa fração do universo amostral finito. Nesses casos, as fórmulas podem ser ajustadas a partir de um fator de correção para populações finitas, minimizando a dimensão da amostra necessária (Quadro 1)².

Exemplo 1: Caso se objetivasse descrever as medidas de pressão arterial média de uma população de pacientes específica, que nunca tenha sido descrita anteriormente, e o erro tolerável fosse de ±5 mmHg, o tamanho amostral precisa se basear em uma expectativa do desvio padrão dos valores desse grupo. Caso um pré-teste com 30 pacientes evidenciasse um desvio padrão de 15 mmHg, o tamanho amostral baseado na fórmula do Quadro 1 seria:

$$n = (1,96 \times 15 / 5)^2 = 34,6 \text{ pacientes}$$

Exemplo 2: Caso se objetivasse descrever a prevalência de insuficiência venosa nos membros inferiores, com um erro tolerável de ±5%, na população de pacientes obesos mórbidos de um ambulatório específico de obesidade que possui um volume de 315 pacientes (630 membros), o cálculo do tamanho amostral do estudo poderia se basear nos resultados de Seidel et al.⁶ que estimaram a proporção de 69,3% de membros acometidos. O cálculo do tamanho amostral para uma população finita baseia-se na fórmula do Quadro 1:

$$n = [630 \times 0,693 \times 0,307 \times (1,96)^2] / \{ [(630-1) \times (0,05)^2] + [0,693 \times 0,307 \times (1,96)^2] \} = 215,5 \text{ membros}$$

O cálculo amostral para comparação de subgrupos (testes de hipóteses) dentro de uma amostra depende do teste estatístico escolhido, das diferenças entre os grupos,

da tolerância do pesquisador à detecção de diferenças quando elas não existem (erro tipo I) ou da falha na detecção de diferenças entre os subgrupos quando elas realmente existem (erro tipo II). As probabilidades associadas aos erros tipos I e II são convencionadas como α e β, e comumente, são adotados valores de 5% (bilateralmente) e 20%, mas outros valores podem ser utilizados de forma criteriosa (Quadro 2)^{1,2}.

Uma estratégia que permite a redução da variabilidade das medidas, aumentando a comparabilidade dos indivíduos em uma amostra, e, conseqüentemente, reduzindo a necessidade numérica amostral para a detecção de um fenômeno, é o pareamento (ou emparelhamento) das observações (Quadro 2). Isso ocorre quando um mesmo indivíduo é observado em diferentes momentos (estudo longitudinal), ou submetido a medidas em diferentes locais do corpo, como a comparação de um tratamento no membro inferior direito *versus* o esquerdo, desde que se respeitem os limites éticos dessa comparação. Outra forma de emparelhamento mais elaborada é a escolha de indivíduos com as mesmas características: idade, gênero, etnia, classe social, entre outras variáveis que possam controlar a variabilidade individual. Nesses casos, a medida ocorre entre os pares, ao invés da comparação direta dos subgrupos¹.

Exemplo 3: Caso se objetivasse comparar as medidas de fluxometria de dois membros de cães submetidos a dois diferentes procedimentos de revascularização arterial, e a diferença mínima tolerável para considerar um procedimento eficiente fosse de ±50 mL/min, seria necessário um estudo piloto que deveria indicar o desvio padrão das diferenças dos fluxos (p.ex.: 60 mL/min), o tamanho amostral baseado na fórmula do Quadro 2 seria:

$$n = [(1,96 + 0,84) \times 60 / 50]^2 = 11,3 \text{ animais}$$

Exemplo 4: Caso se objetivasse comparar as taxas de cura de dois procedimentos cirúrgicos, sendo que o tradicional oferece 70% de cura e o procedimento em estudo devesse ser ao menos 10% superior ao convencional, o cálculo

Quadro 2. Fórmulas para cálculo do tamanho de amostras para comparação de dois grupos segundo variáveis quantitativas e qualitativas e segundo pareamento dos casos.

	Variável quantitativa	Variável qualitativa
Amostra não pareada	$n = (S_a^2 + S_b^2) \cdot \left(\frac{Z_{\alpha/2} + Z_{\beta}}{d} \right)^2$	$n = \frac{(p_1 \cdot q_1 + p_2 \cdot q_2) \cdot (Z_{\alpha/2} + Z_{\beta})^2}{(p_1 - p_2)^2}$
Amostra pareada	$n_P = \left(\frac{(Z_{\alpha/2} + Z_{\beta}) \cdot S_d}{\bar{D}} \right)^2$	$n_P = \frac{(Z_{\alpha/2} + 2 \cdot Z_{\beta}) \cdot \sqrt{p_a \cdot q_a}}{4 \cdot p_d \cdot (p_a - 0,5)^2}$

n – tamanho da amostra (para cada subgrupo); nP – número de pares; Z_{α/2} – valor do erro α, usualmente: 1,96 (5%); Z_β – valor do erro β, usualmente: 0,84 (20%); d – diferença mínima entre as médias; S_a e S_b – desvio padrão da variável em cada grupo; S_d – desvio padrão da diferença entre os pares; \bar{D} – média da diferença entre os pares; p₁ e p₂ – proporção de resultados favoráveis no subgrupo 1 ou 2; q₁ e q₂ – proporção de resultados desfavoráveis no subgrupo 1 ou 2; p_a – proporção de pares discordantes para grupo 1; q_a – proporção de pares concordantes para o grupo 1; p_d – soma da proporção dos pares discordantes dos dois grupos.

do tamanho amostral mínimo de um ensaio clínico deve ser baseado na fórmula do Quadro 2:

$$n = \{[(0,7 \times 0,3) + (0,8 \times 0,2)] \times (1,96 + 0,84)^2\} / (0,7 - 0,8)^2 \\ = 290,4 \text{ pacientes (cada grupo)}$$

Em pesquisas onde diversas variáveis forem importantes para análise do desfecho estudado, ou seja, não são apenas controle ou variáveis de ajuste, faz-se necessário cálculo do tamanho amostral para cada variável importante estudada.

Testes de equivalência, de não-inferioridade e de concordância, requerem dimensionamentos amostrais próprios, distintos dos testes de diferenças de médias e de proporções comumente usados. Além disso, análises multivariadas, comparação de subgrupos com diferentes proporções numéricas, ou múltiplas comparações longitudinais, também envolvem maior complexidade do cálculo amostral. Todos esses itens ultrapassam o escopo deste texto^{1,5,7-10}.

O cálculo amostral para estudos que envolvam a estimativa da correlação linear entre duas variáveis quantitativas depende exclusivamente do coeficiente de correlação linear (Quadro 3).

Exemplo 5: Caso se objetivasse estabelecer a correlação entre a medida de força muscular dos quadríceps e a distância máxima percorrida por pacientes com história de claudicação intermitente, o tamanho amostral poderia ser baseado no estudo de Pereira et al.¹¹ que descreveu um coeficiente de correlação linear de 0,87. De acordo com a fórmula do Quadro 3:

$$n = 4 + \{[(1,96 + 0,84) / (0,5 \times \ln(1 + 0,87) / (1 - 0,87))]\}^2 \\ = 8,4 \text{ pacientes}$$

Estudos longitudinais (coortes prospectivas e ensaios clínicos), por necessitarem do seguimento dos pacientes em função do tempo, podem ser penitenciados pela saída, desistência, perda, morte ou exclusão do indivíduo do estudo. É recomendável a correção do cálculo da amostra inicial para prever essas ocorrências, usualmente, aumenta-se a amostra em até 30%. Os pacientes perdidos (*drop outs*) devem ser estudados criteriosamente quanto as suas razões para saída e se apresentam diferenças quanto às principais variáveis em relação aos remanescentes no estudo, a fim de identificar fatores ligados aos *drop outs*. Quando mais de

Quadro 3. Fórmula para cálculo do tamanho amostral para correlação linear entre variáveis quantitativas.

$$n = 4 + \left(\frac{(Z_{\alpha/2} + Z_{\beta})}{0,5 \cdot \ln\left(\frac{1+r}{1-r}\right)} \right)^2$$

n – tamanho da amostra; $Z_{\alpha/2}$ – valor do erro α , usualmente: 1,96 (5%); Z_{β} – valor do erro β , usualmente: 0,84 (20%); r – coeficiente de correlação linear (Pearson ou Spearman).

30% dos pacientes incluídos perdem o seguimento, pode haver comprometimento da representatividade da amostra, independentemente da suficiência numérica dos casos.

Desde que as conclusões de um estudo sejam generalizáveis apenas à população amostrada, é possível que a repetição do estudo em outros centros possa apresentar resultados diferentes que expressem a realidade da nova população pesquisada. Tais resultados podem, inclusive, extrapolar os limites do intervalo de confiança do parâmetro estimado primariamente, sem significar necessariamente falta de validade interna de nenhum dos dois estudos. Esse também é um dos riscos de se utilizar resultados de outros pesquisadores para o cálculo do tamanho amostral de uma população diferente. A análise preliminar da primeira fração dos casos (pré-teste) é extremamente recomendável, torna mais confortável a estimativa da amostra necessária para cada realidade e previne constrangimentos analíticos ao final do estudo¹².

Sempre que o tamanho da amostra do estudo for muito restrito (<30 medidas), é dificultada a análise de subgrupos e comprometido o desempenho dos testes estatísticos. Deve-se, contudo, ter cuidado adicional com o superdimensionamento amostral, que comumente ocorre quando se tem acesso a grandes bancos de dados informatizados. O aumento da amostra reduz os intervalos de confiança das estimativas e permite a detecção de diferenças entre subgrupos que, apesar de estatisticamente significantes, não possuem relevância clínica^{3,12-14}.

Por fim, há diferentes fórmulas para o cálculo do tamanho amostral para testes estatísticos específicos, além das aqui expostas, dependendo do modelo matemático pressuposto, que podem ser facilmente obtidas na literatura ou mesmo na Internet^{1,15,16}. Há *softwares* gratuitos em português, como o intuitivo BioEstat, que possuem módulos para cálculos de tamanho de amostra¹⁷. Entretanto, a suficiência amostral deve ser entendida como parte importante do planejamento metodológico do estudo, que precisa estar integrado à elaboração das hipóteses, desenho, técnicas de amostragem, análise e interpretação dos dados, para o sucesso da investigação.

Referências

1. Norman GR, Streiner DL. Biostatistics. The bare essentials. 3rd ed. Shelton, Connecticut: People's Medical Publishing House; 2008.
2. Fontelles MJ, Simões MG, Almeida JC, Fontelles RGS. Metodologia da pesquisa: diretrizes para o cálculo do tamanho da amostra. Rev Paran Med. 2010;24:57-64.
3. Paes AT. Itens essenciais em bioestatística. Arq Bras Cardiol. 1998;71:575-80.
4. Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Co; 1987.

5. Azevedo RS. Qual o tamanho da amostra ideal para se realizar um ensaio clínico? *Rev Assoc Med Bras.* 2008;54:289.
6. Seidel AC, Mangolim AS, Rossetti LP, Gomes JR, Jr FM. Prevalência de insuficiência venosa superficial dos membros inferiores em pacientes obesos e não obesos. *J Vasc Bras.* 2011;10:124-30.
7. Katz MH. *Multivariable analysis. A practical guide for clinicians.* 2nd ed. Cambridge, UK: Cambridge University Press; 2006.
8. Ortega Calvo M, Cayuela Dominguez A. Unconditioned logistic regression and sample size: a bibliographic review. *Rev Esp Salud Publica.* 2002;76:85-93.
9. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy.* 2005;85:257-68.
10. Pinto VF. Estudos clínicos de não-inferioridade: fundamentos e controvérsias. *J Vasc Bras.* 2010;9:141-4.
11. Pereira DAG, Faria BMA, Gonçalves RAM, Carvalho VBF, Prata KO, Saraiva PS, et al. Relação entre força muscular e capacidade funcional em pacientes com doença arterial obstrutiva periférica: um estudo piloto. *J Vasc Bras.* 2011;10:26-30.
12. Mourão Jr CA. Questões em bioestatística: o tamanho da amostra. *Rev Interdisc Est Experim.* 2009;1:26-8.
13. Coutinho ESF, da Cunha GM. Conceitos básicos de epidemiologia e estatística para a leitura de ensaios clínicos controlados. *Rev Bras Psiquiatr.* 2005;27:146-51.
14. Weyne GRS. Determinação do tamanho da amostra em pesquisas experimentais na área de saúde. *Arq Med ABC.* 2004;29:87-90.
15. Laboratório de Epidemiologia e Estatística - LEE - Pesquisa. 2000 [cited 2011 Sep 16]. Available from: <http://www.lee.dante.br/pesquisa.html>.
16. UCSF Biostatistics - Power and Sample Size Programs. 2006. [cited 2011 Sep 16]. Available from: <http://www.epibiostat.ucsf.edu/bios-tat/samplesize.html>.
17. BioEstat 5.3 - Instituto de desenvolvimento sustentável Mamirauá. 2011. [cited 2011 Sep 16]. Available from: <http://www.mamiraua.org.br/download/>.