

Determinação do tamanho da amostra em pesquisas experimentais na área de saúde

Gastão Rúbio de Sá Weyne*

Resumo

No presente trabalho são apresentados e discutidos alguns métodos para determinação do tamanho da amostra para atributos e para variáveis, operando-se com dados pareados e não-pareados, visando à validade estatística de trabalhos de pesquisa experimental na área da saúde. Os métodos conduzem a valores mínimos do tamanho da amostra, calculados por equações matemáticas, que devem ser coerentemente majorados de acordo com as condições dos experimentos. São apresentados e discutidos exemplos ilustrativos para melhor entendimento dos métodos sugeridos.

Unitermos

Tamanho de amostra; estatística.

Abstract

This paper introduces statistical methods for calculating the appropriate sample size for comparing two independent groups of subjects (parallel group design), or for comparing paired observations (paired or crossover group design). These methods includes categorical or qualitative data (attributes) and numerical or quantitative data (variables). The main idea behind the sample size calculations is to improve the reliability and credibility of the findings from health field research, giving some guarantee that the statistical aspects have been handled well or even validly.

Keywords

Sample size; statistics.

Introdução

A determinação do tamanho da amostra em trabalho de pesquisa experimental na área de saúde, ou seja, a seleção do número de pacientes a serem utilizados nesta tarefa, é de significativa importância, considerando-se que a correção desta determinação dá validade estatística ao experimento e é condição básica primordial para a aceitação de trabalhos na comunidade científica. Em razão da experiência vivida pelo autor, as perguntas iniciais mais frequentes, feitas, às vezes, até após a conclusão da coleta de dados, nesta área, estão, geralmente, entre as seguintes: a) O tamanho da amostra escolhido no meu trabalho está correto? b) Quando eu iniciar o meu trabalho, qual o tamanho da amostra que deverei utilizar?

c) Em razão do tamanho da amostra que utilizei, os resultados obtidos têm significância estatística? d) Com o tamanho da amostra utilizado, como devem ser correlacionados os dados disponíveis? d) Com o tamanho da amostra que utilizei, o meu trabalho pode ser aceito? O que tenho a corrigir para a aceitação?

Não resta dúvida, portanto, de que o problema da determinação do tamanho da amostra é considerado de fundamental relevância, basicamente, porque a área de saúde apresenta particularidades de tal especificidade que o tratamento a ser dispensado à coleta e à análise de dados requer um procedimento especial. Isto se justifica porque a complexidade do ser humano é incomparavelmente maior que as dificuldades que se apresentam, por exemplo, na área de ciências exatas, com máquinas e/ou computadores. Para caracterizar esta complexidade, veja-se que o ser humano é constituído, aproximadamente, por 30.000 genes e que, no córtex cerebral, existem cerca de $9,2 \times 10^9$ células nervosas interconectadas². Estes números, extremamente elevados, levam a concluir sobre a marcante complexidade do organismo humano. Por outro lado, verifica-se que, nos seres humanos, a tolerância bioestatística é, numericamente, muito ampla. Duas pessoas que têm, cada uma delas, por exemplo, índices glicêmicos de 85 e de 110 mg/dL (diferença de 25 mg/dL), são indivíduos considerados normais, o mesmo acontecendo com dois pacientes que, individualmente, apresentam pressões sistólicas de 100 mmHg e 120 mmHg (diferença de 20 mmHg). Estes valores representam uma ampla tolerância, aceita na área de saúde, que não é admitida, a título de ilustração, na área de engenharia, ciência na qual, por exemplo, uma peça somente poderia ser aceita se a diferença entre sua medida e a medida de um padrão (valor nominal) não excedesse a 1 micrômetro (10^{-6} m). Esta ampla tolerância em ciências da saúde concorre, substancialmente, como se verá adiante, para que as amostras nesta área sejam muito pequenas em relação a outras áreas, como a de engenharia. Por esta razão, enquanto a ordem de grandeza para uma amostra na área de saúde poderia ser, por exemplo, de 30 pacientes ($N = 30$), na área de engenharia poderia ser aceita uma amostra de 1.000 peças ($N = 1.000$).

* Livre-Docente pela Escola Politécnica da USP e Professor Titular do Curso de Ciências Farmacêuticas da Faculdade de Medicina do ABC.

Perguntas que antecedem os cálculos do tamanho da amostra

No desenvolvimento de um trabalho de pesquisa na área da saúde, particularmente na determinação do tamanho da amostra a ser usada, necessário se faz que sejam formuladas e respondidas algumas perguntas básicas, entre as quais citam-se as seguintes:

1) A pesquisa vai lidar com uma variável ou com um atributo?

Uma variável pode assumir diferentes valores, oscilando em torno de um valor médio (ou valor nominal), limitando-se superior e inferiormente por uma tolerância e sendo caracterizada por uma medida definida por um número. Logo, trata-se de um conceito quantitativo. Uma variável X , portanto, pode ser dada por $X = \text{valor nominal} \pm \text{tolerância}$. Como exemplo, admita-se que X é um índice glicêmico (expresso em mg/dL). Se $X = 110 \pm 15$, significa que $X_{\max} = 125 \text{ mg/dL}$ e $X_{\min} = 95 \text{ mg/dL}$.

Como exemplos de variáveis bioestatísticas, pode-se citar: 1) pressão sistólica de 120 mmHg; 2) peso do indivíduo igual a 78 kg; 3) índice glicêmico de 110 mg/dL; 4) temperatura corporal de 36,5° C; 5) altura da pessoa igual a 1,75 m; 6) diâmetro de vaso igual a 3,25 mm; 7) umidade de 0,016 g de água/g de ar seco; 8) pressão atmosférica de 101,35 kPa; 9) nível de colesterol total de 190 mg/dL.

O atributo, também chamado de variável categórica, é uma qualidade dicotômica, que define um “sim” ou um “não”, sendo, portanto, um conceito qualitativo. Ao contrário da variável, o atributo não assume diferentes valores que oscilam em torno de um valor médio. Citam-se, como exemplos de atributos, as seguintes questões que apresentam alternativas: 1) O paciente está ou não febril? 2) João está ou não imunizado? 3) A lâmpada acende ou não? Nestas situações, não há meio termo ou tolerância. Não seria razoável, portanto, dizer que um paciente estaria 20% imunizado. No entanto, um atributo pode ser quantificado por uma fração de sucesso (ou de insucesso). Exemplos: 1) uma vacina A, sendo considerada eficiente se imunizar 80% dos vacinados; 2) um lote B aceitável de lâmpadas quando contenha, no máximo, 0,01% de unidades queimadas. Um atributo pode também ser quantificado com atribuição de pesos ou escores, dando-se pesos maiores, obviamente, para os atributos de maior importância no trabalho experimental. Exemplo: na pesquisa de um paciente à tendência de acidentes vasculares cerebrais, possivelmente poder-se-ia atribuir peso 3 a fatores hereditários e peso 2 ao tipo de alimentação do paciente em determinadas circunstâncias.

2) Os dados são pareados ou não pareados?

Os dados pareados são dados sucessivos, como os resultados da situação física de um paciente antes e depois de

um tratamento. São caracterizados por se referirem a uma mesma amostra, testada antes e depois de um certo procedimento. Os dados não-pareados são simultâneos e correspondem a dois grupos de pesquisa: o grupo de controle e o grupo experimental. Neste caso, o tamanho da amostra, calculada matematicamente como será mostrado, deve ser multiplicado por 2, admitindo-se que o número de participantes dos dois grupos seja o mesmo, o que, em geral, pode ocorrer.

3) Qual o intervalo de confiança na pesquisa?

A precisão das pesquisas na área de saúde admite, usualmente, um erro de 5%, o que equivale a dizer que o intervalo de confiança dos testes é de 95%. Utilizando a distribuição normal com estes dados, o número de desvios-padrão é 1,96. Observe-se que, quanto maior o intervalo de confiança, maior o número de desvios-padrão e maior o tamanho da amostra, uma vez que o erro deverá ser menor. A título de exemplo, enquanto na área de engenharia trabalha-se com intervalo de confiança de 99,74% (três desvios-padrão e erro de 0,26%), na área de saúde utiliza-se, como já se fez menção, um intervalo de confiança de 95% (1,96 desvio-padrão e erro de 5%).

4) A medida a ser realizada tem precisão e exatidão?

A precisão e a exatidão (ou acurácia) do teste são fatores de significância no cálculo do tamanho da amostra. A precisão de uma medida é caracterizada pela reprodutibilidade, situação em que os testes resultam em medidas sempre muito próximas umas das outras e depende do processo usado, como o caso de uma balança precisa, que acusa sempre a mesma medida, mesmo estando viciada (descalibrada). A medida é dita exata (ou com acurácia) quando se aproxima satisfatoriamente do valor real dado por um padrão, logo, uma medida exata ou acurada tem um “compromisso” com a verdade.

5) O teste tem sensibilidade? É um teste específico?

A sensibilidade de um teste é a capacidade do teste para detectar uma doença quando ela está presente. Na triagem, todos os doentes e alguns não-doentes podem ser indicados. A especificidade de um teste mostra que, não ocorrendo a doença, o teste sempre acusará a não existência da enfermidade. Em outras palavras, a especificidade é a capacidade do teste em indicar a ausência de uma doença quando ela não está presente. Os não-doentes são, portanto, excluídos de suspeita.

Equações básicas e exemplos

Equação para variáveis

Sendo X um valor da variável, \bar{X} a média dos valores desta variável, s o desvio-padrão da amostra e N , o tamanho da amostra, tem-se, pela distribuição de *Student*, válida para pequenas amostras ($N = 30$):

$$t = \frac{|\bar{X} - \bar{X}|}{s/\sqrt{N}}$$

Como o número de desvios-padrão, t , na distribuição de *Student*, depende dos graus de liberdade, que pressupõem conhecido o tamanho da amostra, t é substituído pelo número de desvios-padrão Z da distribuição normal. Uma transformação algébrica simples dá o tamanho da amostra N .

$$N = \frac{Z^2 S^2}{d^2}$$

Veja-se que N é inversamente proporcional a d^2 (d = diferença) e, no caso das amostras da área de saúde, como a diferença d é muito grande (e d^2 , com mais forte razão), conforme se fez menção, resulta em valor pequeno para N , ao contrário das pesquisas em engenharia, em que a tolerância, caracterizada também por d , é muitíssimo menor, o que conduz a amostras de tamanho N muito elevadas.

Exemplo ilustrativo: Verificou-se que a população de um bairro de uma cidade litorânea apresentou um elevado valor da pressão arterial sistólica, na faixa média de 160 mmHg. Deseja-se aplicar um medicamento experimental, com o fim de abaixar o valor médio da pressão arterial sistólica até a faixa de 120 mmHg. Como estimar o tamanho da amostra para ensaio dos efeitos do medicamento antes e depois de corretamente aplicado, com um intervalo de confiança de 95% (usual na área de saúde), para o teste ser significativo?

Solução: A tabela da curva normal fornece, para intervalo de confiança de 95%, o número de desvios-padrão, indicando o valor $Z = 1,96$. A diferença será $d = 160 - 120 = 40$ mmHg. Como não há como calcular o desvio-padrão s , deve-se arbitrar uma condição logicamente mais desfavorável, escolhendo-se, por exemplo, um desvio-padrão $s = 80$ mmHg, valor que dificilmente ocorrerá na população estudada. Neste caso, ter-se-á:

$$N = 1,96^2 \cdot 80^2 / 40^2 \cong 16 \text{ indivíduos}$$

Como se trata de dados pareados (sucessivos), o tamanho da amostra é de, no mínimo, 16 indivíduos. Se o teste fosse conduzido com uso de um grupo de controle e um grupo experimental, cada um destes grupos deveria ter 16 pessoas, ou seja, um total de 32 indivíduos, no mínimo.

Equação para atributos

Para atributos, conforme estudos anteriores, considerando-se que p é a fração de sucesso (ou de insucesso), Z é o número de desvios-padrão e $X_{\text{máx}}$, o valor máximo do atributo quantificado, a estatística fornece:

$$X_{\text{máx}} = p + Z\sqrt{p(1-p)/N}, \text{ o que dá}$$

$$X_{\text{máx}} - p = Z\sqrt{p(1-p)/N}$$

$$(X_{\text{máx}} - p)^2 = Z^2 p(1-p)/N, \text{ o que resulta para } N \text{ o valor:}$$

$$N = \frac{Z^2 p(1-p)}{(X_{\text{máx}} - p)^2} \quad \text{ou} \quad N = \frac{Z^2 p(1-p)}{d^2}$$

Verifica-se que deve ser maximizado o valor de $p(1-p)$ e minimizado o valor da diferença $d = X_{\text{máx}} - p$, para maximizar o valor do tamanho da amostra N (situação mais desfavorável). Para dados não pareados (simultâneos), o valor de N deve ser multiplicado por 2.

Exemplo ilustrativo: Uma vacina está sendo testada em certa população, admitindo-se que um valor considerado satisfatório será a obtenção de uma fração de sucesso em torno de 70%. Valores da literatura indicam que, em condições razoáveis, pode-se obter uma fração de sucesso da ordem de 90%. Qual deverá ser o tamanho da amostra, para um intervalo de confiança de 95%?

Solução: Têm-se os seguintes valores:

$$Z = 1,96 \text{ (da curva normal)}$$

$$p = 0,705$$

$$d = X_{\text{máx}} - p = 0,90 - 0,70 = 0,20. \text{ Substituindo estes valores, obtém-se:}$$

$$N = 1,96^2 \cdot 0,70(1 - 0,70) / 0,20^2 \cong 21 \text{ indivíduos, no mínimo.}$$

Se forem usados dados não-pareados (simultâneos), o grupo experimental e o grupo de controle deverão ter, cada um deles, 21 pessoas, com um total de, no mínimo, 42 indivíduos.

Observações sobre o tamanho da amostra (N)

1) Para a determinação do tamanho da amostra, as condições mais desfavoráveis devem ser escolhidas, ou seja: maior s , no caso de variáveis, maior produto $p(1-p)$, no caso de atributos e menor d possível, para que N seja maximizado. 2) O valor de N calculado vale para dados pareados (sucessivos, como no caso de pacientes antes e depois do tratamento). Se os dados forem não-pareados (simultâneos, para um grupo de controle e um grupo experimental), o valor de N deve ser multiplicado por 2. 3) Em qualquer situação (variável ou atributo), o valor de N , calculado matematicamente, é um valor mínimo a ser utilizado na pesquisa e, na prática, deve ser majorado, usando-se um critério razoável. Exemplo: se o cálculo indicou um valor de $N = 13$ pessoas, será possível escolher 13, 15, 20 ou mais indivíduos, desde que o aumento da amostra não implique perda significativa de tempo gasto na pesquisa ou, ainda, que seja praticável nas condições do experimento. 4) No trabalho de

pesquisa na área de saúde, é recomendável apresentar sempre os detalhes que indiquem como foi calculado o tamanho da amostra. 5) Quando o tamanho N de pacientes, é muito grande, deve ser escolhida uma solução entre as seguintes: a) aumentar a taxa ou o tempo de recrutamento; b) reduzir o nível de rigor científico; c) abandonar a pesquisa. Esta última hipótese deve ser afastada a todo custo.

Programas específicos auxiliares de microcomputador

Alguns programas de microcomputador são de extrema importância no cálculo do tamanho da amostra e na determinação de outros parâmetros de interesse da bioestatística. Entre estes programas, estão o Microsoft Excel, o EPIINFO, o SPSS (Statistical Package of Social Science), o Statsoft Statistica 5.0 e o Matlab 6.

Bibliografia

Altman, Douglas G. Practical statistics for medical research. London: Chapman & Hall: 1999.
Dawson B, Trapp KG. Bioestatística básica e clínica. Rio de Janeiro: Mc Graw Hill: 2001.
Cox DR. Planning of experiments. New York: Wiley: 1992.
Herrick CJ. Brains of rats and man. New York: Hafner Publishing Co: 1993.
Jekel JF. Epidemiologia, bioestatística e medicina preventiva. Porto Alegre: Artmed: 1999.
Petrie A, Sabin C. Compêndio de estatística médica. Lisboa: Instituto Piaget: 2000.

Soares JF, Siqueira AL. Introdução à estatística médica. Belo Horizonte: Cooperativa Editora e de Cultura Médica: 2002.
Zar JH. Bioestatistical analysis. London: Prentice Hall: 1999.

Endereço para correspondência

Coordenação do Curso de Ciências Farmacêuticas da FMABC
Av. Príncipe de Gales, 821 – Vila Príncipe de Gales
Santo André, SP – CEP 09060-650
Tel.: (11) 4993-5411