

QUESTÕES EM BIOESTATÍSTICA: O TAMANHO DA AMOSTRA

BIOSTATISTICAL QUESTIONS: THE SAMPLE SIZE

Carlos Alberto Mourão Júnior*

RESUMO

O presente artigo discute de forma crítica a questão do cálculo do tamanho de amostras em pesquisas experimentais e sugere estratégias práticas para a abordagem do problema.

PALAVRAS-CHAVE

Bioestatística. Amostra. População. Teorema central do limite.

1 VISÃO GERAL DO PROBLEMA

Um dos principais desconfortos que aflige boa parte dos pesquisadores, quando elaboram seus projetos de pesquisa, é a dúvida sobre qual deve ser o tamanho da(s) amostra(s) que ele deve estudar, a fim de que seus resultados sejam confiáveis. As razões para tal preocupação são plenamente justificáveis, afinal, de um modo geral, quanto maiores são as amostras, mais demorado e caro é o estudo.

Existem técnicas estatísticas para calcular o tamanho “ideal” de amostras (para conhecer esses cálculos consulte COCHRAN, 1965). Entretanto, para que tais fórmulas sejam aplicadas, precisamos conhecer a variância da população, definir o poder estatístico que pretendemos que os testes apresentem e estabelecer a magnitude da diferença dos resultados que pretendemos encontrar entre os grupos estudados. Como esses dados na maior parte das vezes não estão disponíveis, apresentaremos nesse artigo algumas dicas de ordem prática para você lidar com a questão do tamanho das amostras.

Antes, porém, é muito importante que se compreenda o conceito de amostra e qual o seu papel em uma pesquisa. Vamos começar esclarecendo a diferença entre dois conceitos fundamentais: *população* e *amostra*. Talvez esses dois conceitos constituam o fundamento teórico mais importante da estatística, apesar de infelizmente serem tão mal compreendidos. É possível que a compreensão inadequada acerca desses fundamentos explique a ideia totalmente equivocada de que a estatística é uma ciência de difícil compreensão – uma verdadeira caixa-preta.

2 POPULAÇÃO E AMOSTRA

Imagine que você seja proprietário de uma grande indústria de parafusos, e em sua fábrica você tem uma máquina da marca A que produz 10.000 parafusos por dia. Com o objetivo de ampliar seus negócios, você acabou de importar de outro fabricante uma máquina da marca B, que produz 15.000 parafusos por dia. Você deseja saber se suas duas máquinas (A e B) estão produzindo parafusos com um grau semelhante de precisão, e para testar essa hipótese você pretende medir o diâmetro dos parafusos em micrômetros com um aparelho de altíssima precisão e efetuar um teste estatístico para comparar a média do diâmetro dos parafusos produzidos por A com a média daqueles produzidos por B.

Podemos dizer que, nesse caso, você pretende comparar duas populações: parafusos produzidos por A em um determinado dia, com 10.000 elementos ($N = 10.000$), e parafusos produzidos por B em um determinado dia ($N = 15.000$). Assim, definimos *população* como sendo o conjunto de elementos que apresenta uma determinada característica que pretendemos estudar.

Vejam outros dois exemplos de população: a) camundongos machos adultos criados no biotério da universidade X; b) pacientes hipertensos negros, do sexo masculino, com idade entre 40 e 50 anos atendidos pelo serviço público do município Y. É muito importante compreender que a população precisa ser bem descrita e delimitada (infelizmente isso raramente acontece nos trabalhos publicados na literatura científica), pois não podemos extrapolar os achados dos camundongos do item a para camundongos criados em uma universidade do Afeganistão, da mesma forma que não podemos extrapolar os resultados encontrados para a população do item b para todos os hipertensos do planeta Terra. Esse tipo de extrapolação é um equívoco gravíssimo. Por isso temos que ter um olhar crítico sobre os consensos, diretrizes e evidências científicas. Será que a população

* Correspondence Author: Carlos Alberto Mourão Júnior. Address: Campus UFJF, Bairro Martelos, s/n, CEP 36035360. Tel.: 55-2102-3211.
E-mail: camourao@uff.edu.br.
Departamento de Fisiologia – Universidade Federal de Juiz de Fora.
Recebido em: Agosto de 2008.
Aceito em: Dezembro de 2008.

que foi estudada por esses consensos tem as mesmas características da população onde pretendemos aplicar seus resultados?

É importante termos sempre em mente que os estudos experimentais, por mais bem desenhados que sejam, normalmente apresentam *validade externa* (capacidade de extrapolação dos resultados) limitada, já que se aplicam somente à população estudada, ainda que, repetimos, os estudos normalmente não explicitem as características dessa população. Por isso cabe a nós, leitores, examinarmos com olhos de linca as informações contidas nas pesquisas publicadas.

Vamos voltar aos parafusos. É óbvio que não é nada sensato pensar em medir o diâmetro de todos os parafusos produzidos por A e por B em um determinado dia. O ideal é você selecionar de forma aleatória um determinado número n de parafusos de A e B, medir seus diâmetros e comparar suas médias. A esse *subconjunto de uma determinada população* damos o nome de *amostra*.

Tendo em mente os conceitos fundamentais de população e amostra, podemos compreender os dois papéis que a estatística desempenha frente aos resultados encontrados: a) análise descritiva: descreve o resumo das medidas da amostra (média, mediana, desvio padrão, etc.); b) análise inferencial: através dos testes de hipóteses (valor de p e intervalos de confiança) diz qual a probabilidade de estarmos corretos ao extrapolar os resultados encontrados nas amostras para as populações estudadas.

Lembre-se sempre de que a estatística não trabalha com certezas, e sim com probabilidades. Além disso, a *única coisa* que a estatística e o tão propalado valor de p podem fazer por você é dizer com que grau de certeza você pode extrapolar seus achados para a população que você estudou. A estatística não pode dizer mais nada! Muitos pesquisadores esperam dessa ciência mais do que ela pode oferecer, muitas vezes generalizando resultados ou tomando o valor de p como dogma, dizendo estarem respaldados pela estatística, sem perceberem que a estatística *não se presta a extrapolar resultados de uma população delimitada para outras populações*.

Para que você consiga cumprir seu objetivo, ou seja, tirar conclusões a respeito das populações a partir da análise das amostras, é claro que essas amostras precisam de ter um tamanho (n) tal que permita que elas sejam representativas da população. A pergunta que não cala é: qual é o tamanho de n que garante que a amostra representa bem a população acerca da qual desejamos tirar conclusões? Uma coisa é certa: se o n amostral for suficiente para garantir que a amostra representa a população, então esse n é adequado. Dito isto, fica claro que o que importa é que a amostra seja representativa. Vamos discutir agora as características que a amostra precisa ter a fim de cumprir esse papel.

3 AMOSTRA REPRESENTATIVA

Ainda aproveitando o exemplo dos parafusos, a primeira preocupação que precisamos ter é: qual parafusos servem para compor

a amostra (critérios de inclusão) e quais não servem (critérios de exclusão). Um critério de inclusão poderia ser, por exemplo, parafusos fabricados pelas máquinas A e B em um determinado dia. Um exemplo de critério de exclusão seria: não entrarão na amostra parafusos que apresentem alguma alteração (defeito, etc.), percebida a olho nu.

Uma vez definidos os critérios de inclusão e exclusão, o segundo passo é tentar garantir que todos os elementos da população tenham a mesma probabilidade de serem escolhidos para compor a amostra. O ideal seria que essa escolha fosse aleatória, através algum tipo de sorteio (randomização), como por exemplo: a cada 10 parafusos produzidos, escolher sempre o décimo para compor a amostra, até completar o n desejado. Muitas vezes o pesquisador se vê obrigado a utilizar uma *amostragem por conveniência*, ou seja, utilizar os elementos ou dados que estão disponíveis (ex: entrevistar os alunos de um mesmo período e de um mesmo curso). A amostragem por conveniência, apesar de estar sujeita a vícios de seleção, não é necessariamente um procedimento condenável. Existem técnicas estatísticas para ajudar a controlar as possíveis variáveis intervenientes que possam surgir. Com certeza a fase de elaboração do projeto é o melhor momento para trocar ideias com algum colega que tenha boa experiência com estatística. O aconselhamento estatístico deve ser “preventivo”, pois o mais importante em qualquer projeto é saber escolher adequadamente a população, fazer uma boa amostragem e traçar estratégias para contornar as variáveis de confusão. Nenhum software de estatística é capaz de fazer isso. Os programas apenas fazem cálculos sem entrar no mérito da qualidade das informações neles digitadas.

Levando em consideração os dois passos iniciais (escolha dos critérios de inclusão e exclusão, e a técnica de amostragem), é finalmente chegada a hora de discutir qual deve ser o valor ideal do n amostral.

4 O TAMANHO DA AMOSTRA

Ao contrário do que muitos acreditam, talvez seja pouco viável e até certo ponto inútil tentar calcular, *a priori* (antes do experimento começar), o tamanho do n amostral suficiente para que a amostra seja representativa da população. Acreditamos que o mais sensato e factível seja, uma vez determinadas sua população e amostra, começar a coleta dos dados, e na medida em que estes vão sendo colhidos você já ir analisando sua distribuição e suas características. Nesse sentido, acreditamos que a determinação do n amostral seja um processo dinâmico, que pode mudar na medida em que os dados vão sendo colhidos e analisados. Vejamos algumas dicas sobre como proceder na prática.

Inicie sua coleta, a após ter colhido de 8 a 10 dados por grupo, calcule a média e o desvio padrão dos dados coletados (isso pode ser feito com uma calculadora ou com qualquer um dos muitos softwares e programas gerenciadores de planilhas eletrônicas disponíveis no mercado). Se os resultados sugerirem uma distribuição em que a

maior parte dos dados se concentra próxima da média, e a variação dos dados em relação à média não for muito grande, é possível que os dados de sua amostra apresentem uma distribuição simétrica, o que sugere que representam bem a população estudada. Você pode visualizar essa distribuição construindo um histograma ou um box-plot dos dados (através de software), ou então usando uma regra prática, que é extremamente simples e na maioria das vezes funciona muito bem: *se o desvio padrão for menor que a metade da média, é muito pouco provável que sua distribuição seja assimétrica*. Nesse caso, se em todos os grupos (amostras) estudados a distribuição for simétrica, talvez os 8 ou 10 sujeitos analisados sejam suficientes para que você possa aplicar testes estatísticos com segurança. É por isso que existem muitos bons estudos publicados com um $n = 8$ (ou até menos) em cada grupo. *Não é o tamanho da amostra que garante bons resultados, mas sim sua qualidade* (capacidade de representar a população). Nos experimentos com modelos animais, onde a variabilidade é muito pequena, geralmente um n de 8 a 10 animais por grupo costuma ser satisfatório. No caso dos parafusos, é provável também que você consiga inferir, com uma boa probabilidade de acerto, se a produção das máquinas A e B é diferente, usando em torno de 10 parafusos (ou menos) por amostra.

Em experimentos com seres humanos, onde a variabilidade é bem maior em função da individualidade biológica de cada um, e muitas vezes não se conseguem amostras tão homogêneas e bem controladas, um n amostral de 15 sujeitos por grupo já pode ser suficiente, desde que as amostras tenham distribuição simétrica e o desvio padrão das amostras não seja muito diferente.

No entanto se for possível conseguir cerca de 30 sujeitos por grupo, nem será necessário se preocupar com a distribuição das amostras, pois um importante fundamento matemático denominado *teorema central do limite*, garante que com um n em torno de 30 as médias amostrais apresentam uma distribuição que tende à distribuição normal, independentemente da forma da distribuição da amostra em si. Em outras palavras, se tomamos repetidamente amostras aleatórias independentes de tamanho n de uma população, então se n for grande, a distribuição das médias amostrais vai se aproximar de uma distribuição normal (para conhecer detalhes matemáticos sobre esse teorema, reporte-se a JAMES, 2004). Como a maior parte dos testes estatísticos trabalha com médias amostrais, logo a maioria dos testes são robustos o suficiente para fornecer resultados confiáveis, independentemente da distribuição das amostras, quando o n é maior que 30.

Até existem testes estatísticos para verificar a normalidade das amostras, bem como para verificar se as amostras são homocedásticas (apresentam variância semelhante), porém se a amostra for grande ou a distribuição for aparentemente simétrica esses testes se tornam desnecessários (uma discussão mais detalhada sobre isso pode ser encontrada nos livros sugeridos nas referências bibliográficas).

Para finalizar, vamos tentar derrubar um perigoso mito que existe acerca do n amostral. É muito comum ouvirmos sobre o tamanho da amostra a seguinte afirmativa: quanto maior, melhor. Isso definitivamente não é verdade. Com referência a essa questão se aplica perfeitamente o dito popular de que *"tamanho não é documento"*.

Na verdade, amostras muito grandes podem muitas vezes ser "perigosas", pois quanto maior o n amostral, maior a probabilidade de encontrarmos uma diferença estatisticamente significativa (o famigerado $p < 0,05$) entre os grupos estudados, ainda que tal diferença não tenha definitivamente significado prático algum. Infelizmente essa "artimanha estatística" de "esticar" o n até achar um p significativo é bem mais comum do que se imagina, e muito disso ocorre pelo fato de um grande número de periódicos se recusarem terminantemente a publicar resultados ditos "negativos" (com um $p > 0,05$). Seria muito saudável se os periódicos contassem com revisores e consultores que tivessem traquejo em estatística.

Para concluir, lembramos novamente que, a qualidade das amostras é muito mais importante que o tamanho das mesmas. Aqui, mais uma vez, se confirma a primazia da qualidade sobre a quantidade.

5 REFERÊNCIAS

- COCHRAN, W. G., **Técnicas de amostragem**. Rio de Janeiro: Editora Fundo de Cultura, 1965.
- JAMES, B. R., **Probabilidade**: um curso em nível intermediário. 3ª ed. Rio de Janeiro: IMPA, 2004.
- JEKEL, J. F.; ELMORE, J. G.; KATZ, D. L., **Epidemiologia, bioestatística e medicina preventiva**, Porto Alegre: Artmed, 2002.
- MOORE, D. S. A., **Estatística básica e sua prática**. 3ª ed. Rio de Janeiro: LTC, 2005.
- MOURAO-JUNIOR, C. A., Bioestatística: armadilhas e como evitá-las. **Boletim do Centro de Biologia da Reprodução**, v. 25, p. 105-111, 2006.
- SNEDECOR, G.W.; COCHRAN, W. G. **Statistical methods**. 8th ed. Ames: Blackwell, 1989.
- SOKAL, R. R.; ROHLF, F. J., **Biometry**. 3rd ed. New York: W. H. Freeman, 2003.
- TRIOLA, M. F., **Introdução à estatística**. 7ª ed. Rio de Janeiro: LTC, 1999.
- ZAR, J. H., **Biostatistical analysis**. 3rd ed. Upper Saddle River: Prentice Hall, 1996.