

## ARTIGOS

### A LÓGICA DA DETERMINAÇÃO DO TAMANHO DA AMOSTRA EM INVESTIGAÇÕES EPIDEMIOLÓGICAS

*The logic of sample size determination in epidemiological research*

Ronir Raggio Luiz<sup>1</sup>, Monica M. F. Magnanini<sup>2</sup>

#### RESUMO

Uma das perguntas mais freqüentes por parte dos pesquisadores da área da Saúde, seja para um estudo em laboratório, clínico ou epidemiológico, se refere ao número de elementos que deve ser investigado a fim de se ter um estudo “confiável” ou “significativo”. Ou seja, deseja-se saber qual o tamanho da amostra. Basicamente, o tamanho da amostra depende da precisão desejada, de arbitrio do pesquisador. Entender a lógica por trás da sua determinação é fundamental para o planejamento e suporte às conclusões de qualquer investigação epidemiológica. Este artigo procura explorar as idéias e os elementos influentes na determinação do tamanho da amostra.

#### PALAVRAS-CHAVE

Tamanho amostral, precisão, estatística, pesquisa epidemiológica

#### ABSTRACT

Researchers are very much concerned about the number of units that need to be investigated in order to generate “reliable” or “significant” results from laboratory, clinical, and epidemiologic studies. Briefly, they would like to know what an adequate sample size should be. In essence, sample size relies on the desired statistical precision, as established by investigator. Understanding of the logic construct which stands behind it is required for consistent research development. This paper presents the underlying principles involved in sample size determination.

#### Key words

Sample size, precision, statistics, epidemiological research

<sup>1</sup>Professor de Bioestatística do NESC/UFRJ e da FM/UFRJ,

<sup>2</sup>Estatística do NESC/UFRJ

## 1. INTRODUÇÃO

Uma das perguntas mais freqüentes por parte dos pesquisadores da área de saúde, seja para um estudo em laboratório, clínico ou epidemiológico, se refere ao número de elementos (cobaias, pacientes ou indivíduos) que deve ser investigado a fim de se ter um estudo ou amostra “confiável” (“representativo(a)” ou “significativo(a)”). Ou seja, deseja-se saber qual o tamanho da amostra.

As aspas nos adjetivos citados são para reproduzir as palavras utilizadas pelos pesquisadores, cabendo logo um primeiro esclarecimento a fim de se obter alguma uniformidade semântica. Pode-se considerar que um estudo é “confiável” (na linguagem utilizada, mas que se refere ao conceito de “validade”) se ele reproduz a verdade dos fatos. Entretanto, qualquer estudo pode falhar neste aspecto por dois tipos de erros: o erro sistemático ou viés e o erro aleatório, este natural de qualquer processo amostral. Admitindo-se que a amostra seja aleatória, premissa necessária para qualquer estudo onde o cálculo do tamanho da amostra constitua uma tarefa preliminar, espera-se que não haja nenhum erro sistemático. Isto é, espera-se obter uma amostra “representativa” da população da qual se pretende fazer alguma inferência. Qualquer erro seria então atribuível ao erro aleatório. Quanto ao termo “significativo(a)” utilizado pelos pesquisadores, pode-se atribuí-lo à expectativa deles em obter resultados (dados) que gerem testes estatisticamente “significativos” (p-valor pequeno).

Assim, diferente do conceito de validade, ao tamanho amostral associa-se o conceito de precisão. E a imprecisão é o preço a ser pago por qualquer investigação onde uma amostra aleatória esteja presente. No caso de uma amostra não-aleatória, situação muito comum na investigação em saúde, o conceito de validade é fundamental, dado que qualquer análise estatística feita com esses dados deve estar condicionada à hipótese de que a amostra estudada se “comporta” tal como se comportaria tivesse ela sido selecionada aleatoriamente. É freqüente o caso onde o pesquisador já tem a sua mão uma amostra não-aleatória (os casos de uma certa doença registrados em um hospital num certo período, por exemplo) e quer saber se esse tamanho amostral é “suficiente” para responder certa pergunta. Nesses casos, é mais prudente uma discussão con-

sistente de validade do que de tamanho amostral, mesmo porque qualquer eventual avaliação de tamanho amostral “pequeno” seria de difícil solução dada as freqüentes limitações operacionais de se coletar mais casos. Poderia se recomendar então, uma vez discutida a questão da validade, a realização do estudo com aquele tamanho amostral disponível, informando a precisão encontrada, mesmo que esta não tenha sido aquela que se desejaria. Estudos estatisticamente “não-significativos” também são importantes de serem relatados, seja como base para futuros estudos ou sistematizações, seja pela eventual significância clínica dos seus resultados. Diferente do conceito de significância estatística, significância clínica se refere a um resultado cuja magnitude seja de tal forma expressiva que mereça ser considerado pelo pesquisador.

Ainda sobre uma amostra não-aleatória, é comum a situação onde o pesquisador dispõe de toda uma população embora possa acreditar estar trabalhando com uma amostra. Todos os pacientes internados em certo hospital, por exemplo, constituem uma população. Se o interesse é só nesta população, não há razão para se fazer nenhum tipo de inferência (construção de intervalos de confiança ou cálculo de p-valores), bastando uma descrição dos dados. Por outro lado, pode-se imaginar estes pacientes como sendo uma amostra de outros pacientes (internados em outros hospitais) e, aí então, seria legítimo fazer inferências estatísticas. Mais uma vez a questão da validade deve ser considerada com cuidado, já que a população referenciada freqüentemente não é claramente definida.

Como último comentário introdutório, para se calcular um tamanho amostral é necessário se ter uma pergunta bastante específica. Quer-se uma amostra especificamente para quê? É comum, também, a situação onde o pesquisador tem um enorme questionário e quer saber quantos indivíduos ele precisa investigar para traçar um “perfil” de uma certa população. Neste caso, é difícil uma resposta satisfatória porque, na realidade, para cada pergunta daquele questionário haveria um tamanho amostral associado. Uma saída seria verificar qual pergunta exigiria, para uma precisão desejada, um tamanho amostral maior, então adotá-lo. Isto faria com que, para as outras perguntas teríamos uma amostra maior do que a necessária. Felizmente, na maioria dos estudos epidemiológicos, as perguntas não são tão abrangentes como “qual o perfil de uma popula-

ção?”, podendo freqüentemente ser respondidas pela estimação de um parâmetro populacional. “Com que freqüência surgem novos casos de AIDS na cidade X por ano?”, “há mais hipertensos na comunidade X que na comunidade Y?”, “exposição E está relacionada à doença D?” e “droga A é melhor que droga B no tratamento para a doença D?” são exemplos de perguntas que podem então ser respondidas através de amostras que estimem grandezas associadas a elas.

## 2. ELEMENTOS INFLUENTES NA DETERMINAÇÃO DO TAMANHO AMOSTRAL

Voltando à questão do tamanho da amostra, é intuitivo perceber ser ele dependente do erro aleatório mencionado acima. Há uma relação inversa entre o erro e o tamanho da amostra. Amostras “grandes” estão associadas a erros “pequenos” e amostras “pequenas” a erros “grandes”. E, assim, a grosso modo, parecem não haver amostras “grandes” nem amostras “pequenas”. Há amostras que são compatíveis com o erro que se “tolera” cometer em um particular estudo. Isto não melhora muito o problema de determinação do tamanho da amostra, mas pelo menos transfere para quem deve ser o responsável pela solução: o pesquisador. Isto porque a magnitude do erro tolerável deve ser atribuído por ele, de tal sorte que seu estudo tenha reconhecimento pela comunidade científica.

Entretanto, não é somente o erro tolerável de amostragem o único elemento que pode afetar o tamanho da amostra. Por exemplo, diante de um estudo de prevalência, é intuitivo imaginar que quanto mais rara for a doença, maior deverá ser o tamanho da amostra. No caso de um estudo comparativo, um estudo de coorte por exemplo, é também intuitivo considerar que, quanto maior for a diferença realmente existente entre as incidências nos grupos exposto e não-exposto, menor será a amostra total necessária para detectá-la. Em havendo, de fato, uma diferença entre estas incidências, mas de “pequena” magnitude, somente uma amostra “grande” será capaz de detectá-la.

No caso de um estudo onde se quer estimar a média de uma variável contínua (por exemplo, glicemia), é também intuitivo perceber que, quanto maior for a variabilidade desta variável, maior será o tamanho da amostra necessária para se estimar sua média, com uma precisão desejada.

Além destes elementos que, intuitivamente, percebemos influir no tamanho da amostra, há ainda outros que exigem uma maior reflexão. O processo de inferir a partir de uma amostra pode-se dar por estimação de um parâmetro populacional ou pelo teste de uma hipótese, sendo o cálculo do tamanho da amostra dependente de um desses objetivos. Sob um processo de estimação, é possível que seja selecionada uma amostra que resulte numa estimativa cujo valor possa estar fora dos limites estabelecidos pelo erro tolerável mencionado acima, incorrendo, assim, em um outro erro que se gostaria de poder minimizar. O tamanho da amostra mais uma vez é um regulador deste erro, e vice-versa. Esta idéia é a mesma daquela presente na construção de intervalos de confiança e será usada na seção seguinte para desenvolver um raciocínio geral para o cálculo do tamanho da amostra. Este último erro é conhecido como nível de significância ( $\alpha$ ), com valor complementar igual à confiança desejada ( $1 - \alpha$ ) (Soares & Siqueira, 1999).

Se o objetivo do estudo é testar uma hipótese colocada *a priori* (hipótese nula =  $H_0$ ), então um outro erro pode acontecer. É possível que esta hipótese não seja verdadeira e o estudo, em função do tamanho da amostra, não ser capaz de rejeitá-la. Entretanto, este erro não é somente dependente do tamanho da amostra, dependendo também de alternativas àquela hipótese. Para um tamanho de amostra e um erro  $\alpha$  fixados, valores alternativos muito distantes daquele estabelecido *a priori* terão um erro associado menor, e valores alternativos próximos, um erro maior. Este erro é conhecido como  $\beta$  e seu complementar,  $1 - \beta$ , como o poder (Soares & Siqueira, 1999). Assim, fixados um  $\alpha$  e uma hipótese alternativa, quanto maior a amostra, maior será também o poder do estudo em indicar a hipótese alternativa, quando de fato ela for verdadeira (Levy & Lemeshow, 1999). A fixação de uma hipótese alternativa deve respeitar condições impostas pelo pesquisador, de modo a se conseguir um tamanho amostral que tenha um poder “razoável” para se detectar a diferença entre as hipóteses nula e alternativa. Esta diferença deve ter alguma relevância “clínica”, ou seja, uma diferença que, se de fato existe, o estudo não deve deixar de captá-la.

Um outro elemento eventualmente importante na determinação do tamanho da amostra é o tamanho da população. Observa-se com relação a este elemento, entretanto, uma curiosidade que surpreende os pesquisa-

dores. A intuição dita que quanto maior a população maior deve ser a amostra para representá-la, com uma precisão desejada. Isto, entretanto, é “parcialmente” verdadeiro. É verdadeiro porque, de fato, quanto maior a população maior deverá ser a amostra. Porém, mostra-se que esta relação só é importante para populações “pequenas”. Para populações “grandes”, o tamanho da amostra não será substancialmente influenciado pelo tamanho da população, podendo esta ser considerada como infinita. Em outras palavras, o que importa é o tamanho da amostra e não a fração amostral (relação amostra/população). Uma fração amostral de 10%, por exemplo, pode gerar uma amostra gigantesca ou muito pequena, dependendo do tamanho da população. Esta propriedade será ilustrada mais à frente.

### 3. ASPECTOS TEÓRICOS

#### 3.1 DESENVOLVENDO UM RACIOCÍNIO GERAL

Para entendimento da lógica que está por trás do cálculo do tamanho de uma amostra, seja o problema de se estimar uma prevalência. Quando se estuda inferência estatística, aprende-se que, para um determinado tamanho amostral ( $n$ ), pode-se estimar a prevalência de uma doença, incorporando sua precisão, a partir da construção de um intervalo de confiança (Soares & Siqueira, 1999). Por outro lado, quando se aprende amostragem discute-se o conceito de erro-padrão ( $EP$ ), que se refere a uma medida de variabilidade de um estimador de um parâmetro populacional e depende do tamanho da amostra ( $n$ ) (Fleiss, 1981). Esses conceitos serão usados para se determinar o tamanho de uma amostra para um estudo de prevalência.

O erro-padrão para uma proporção  $EP_{\hat{p}}$  é dado por:

$$EP_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}} \quad (1)$$

onde  $P$  é a prevalência. Assim, os limites do intervalo de confiança de 95% para a prevalência  $P$ , aproximando-se pela distribuição normal, são dados então por:

$$\hat{P} \pm 1,96 \sqrt{\frac{P(1-P)}{n}}, \quad (2)$$

onde  $\hat{P}$  é a prevalência amostral (Fleiss, 1981) e 1,96 é o valor tabelado da distribuição normal padronizada ( $\mu = 0$  e  $\sigma = 1$ ) correspondente a um intervalo de 95%.

Associado ao intervalo de confiança de 95%, tem-se um  $\alpha = 0,05$ . Então, para um  $\alpha$  qualquer, tem-se um intervalo de confiança de  $100(1-\alpha)\%$ . E, lembrando da distribuição normal padronizada, para cada  $\alpha$  há um valor  $z_{\alpha/2}$  tal que o intervalo entre  $-z_{\alpha/2}$  e  $z_{\alpha/2}$  corresponde a uma probabilidade de  $100(1-\alpha)\%$ . Assim, de forma geral, os limites do intervalo de confiança de  $100(1-\alpha)\%$  para a prevalência  $P$  podem ser ilustrados pela desigualdade:

$$\hat{P} - z_{\alpha/2} \sqrt{\frac{P(1-P)}{n}} \leq P \leq \hat{P} + z_{\alpha/2} \sqrt{\frac{P(1-P)}{n}}. \quad (3)$$

Isto é, a prevalência amostral ( $\hat{P}$ ) dista do parâmetro de interesse, a prevalência populacional  $P$ , por um erro  $\varepsilon$  tal que

$$\varepsilon = z_{\alpha/2} \sqrt{\frac{P(1-P)}{n}}. \quad (4)$$

A partir desta expressão, fica fácil calcular um tamanho amostral para se estimar uma prevalência e reconhecer seus elementos influentes. Elevando ao quadrado ambos os lados da equação acima, tem-se:

$$\varepsilon^2 = z_{\alpha/2}^2 \frac{P(1-P)}{n}. \quad (5)$$

E, explicitando em  $n$ ,

$$n = \frac{z_{\alpha/2}^2 P(1-P)}{\varepsilon^2}. \quad (6)$$

Assim, no cálculo do tamanho amostral para estimar uma prevalência é necessário informar três elementos:

1) o erro tolerável de amostragem ( $\varepsilon$ ). Isto é, quanto de afastamento entre a prevalência populacional (parâmetro desconhecido) e a estimativa a ser obtida na amostra o pesquisador tolera. Exemplificando, se o pesquisador estabelece um erro absoluto de 2% ( $\varepsilon = 0,02$ ) para a estimação de uma prevalência que ele acredita ser de 30%, então ele estaria satisfeito

com qualquer valor entre 28% e 32% para a sua estimativa;

2) o erro  $\alpha$  (através do  $z_{\alpha/2}$ ). Isto é, a probabilidade de que a estimativa a ser obtida pela amostra esteja além do limite tolerável de  $\pm \epsilon$ . Este erro não apresenta muita dificuldade para sua determinação, sendo quase que uma regra o valor de 5% (com correspondente  $z_{\alpha/2}=1,96$ ), inclusive para qualquer outro tipo de investigação epidemiológica. Os valores de  $\alpha=10\%$  ( $z_{\alpha/2}=1,64$ ) e  $\alpha=1\%$  ( $z_{\alpha/2}=2,58$ ) são também referências clássicas. É importante observar a diferença entre os erros  $\epsilon$  e  $\alpha$  e entender porquê um é mais fácil de ser arbitrado que o outro.

3) a prevalência  $P$ . A dependência sobre este elemento costuma incomodar os pesquisadores já que esta é exatamente a informação que eles desejam conhecer, sendo necessário, então, arbitrar um valor (ou utilizar resultados de estudos similares, ou ainda, na hipótese de não haver nenhuma informação, fazer um estudo-piloto). Ao invés de se preocupar muito com o  $P$  isoladamente, pode-se olhar a dependência em função de  $P(1-P)$  e observar que este produto toma seu valor máximo quando  $P=0,5$ . Portanto, se não se quer arbitrar nenhum valor utiliza-se o valor máximo ( $0,5 \times 0,5 = 0,25$ ), maximizando-se assim o tamanho da amostra. Por outro lado, o que pode causar estranheza é o fato de para doenças raras o tamanho da amostra ser cada vez menor, já que  $P(1-P)$  diminui em função de  $P$ , diminuindo assim o tamanho da amostra para uma precisão desejada. A explicação é que, para  $\alpha$  e  $\epsilon$  fixados, uma diminuição em  $P$  implica em um erro relativo maior. Um estudo com um  $\epsilon=5\%$ , por exemplo, é bastante diferente quando se tem um  $P=50\%$  de quando se tem um  $P=10\%$ . No primeiro, o erro relativo é de 10% ( $0,05/0,5$ ) enquanto que no segundo o erro relativo é de 50% ( $0,05/0,10$ ). Imagine então um estudo com um  $\epsilon=5\%$  em uma população com prevalência esperada de 1%. De fato, com um erro tão grande quanto esse (em relação à prevalência), uma amostra pequena é suficiente.

### 3.2 O ERRO RELATIVO ( $\epsilon_r$ )

O problema acima poderia ser contornado pelo pesquisador informando um erro absoluto,  $\epsilon$ , que fosse compatível com a prevalência presumida (ou esperada). Entretanto, pode-se facilmente generalizar o problema do cálculo do tamanho da amostra quando se quer controlar o erro relativo ao invés do erro absoluto.



A partir da expressão (4), dividindo-se ambos os lados da equação por  $P$  de modo a relativizar o erro  $\varepsilon$ , tem-se:

$$\frac{\varepsilon}{P} = \frac{z_{\alpha/2}}{P} \sqrt{\frac{P(1-P)}{n}}. \quad (7)$$

E, fazendo-se  $\varepsilon/P = \varepsilon_r$ , tem-se

$$\varepsilon_r = \frac{z_{\alpha/2}}{P} \sqrt{\frac{P(1-P)}{n}}. \quad (8)$$

Assim, elevando ao quadrado e explicitando em  $n$ , obtém-se um tamanho amostral para prevalência a partir do erro relativo:

$$n = \frac{z_{\alpha/2}^2 (1-P)}{\varepsilon_r^2 P}. \quad (9)$$

Deve-se observar que as duas expressões desenvolvidas, (6) e (9), são equivalentes. O pesquisador ao arbitrar um erro tolerável absoluto e uma prevalência presumida, obtém imediatamente o erro relativo, e vice-versa. Basta, portanto, a identificação de apenas um deles.

Todo este desenvolvimento até aqui para estudos de prevalência se aplica diretamente a um estudo de incidência. Assim como a prevalência, a incidência (acumulada) também é uma proporção. A diferença obviamente está na interpretação epidemiológica destas medidas de frequência.

### 3.3 O TAMANHO DA POPULAÇÃO ( $N$ )

Pelas expressões (6) e (9), viu-se que o tamanho amostral não dependia do tamanho da população. A razão é que estas expressões foram obtidas supondo uma população infinita. Na realidade, quando a população não é tão grande, a expressão para o erro-padrão da prevalência tem uma correção, passando a depender também do tamanho da população, e é dado por

$$EP_{\hat{p}} = \sqrt{\frac{N-n}{N-1} \frac{P(1-P)}{n}}, \quad (10)$$

onde o termo

$$\frac{N-n}{N-1} \quad (11)$$

é conhecido como fator de correção para população finita (Levy & Lemeshow, 1999). Percebe-se pela Tabela 1 que quando  $N$  cresce muito

( $N \rightarrow \infty$ ), este fator converge para 1. Outra forma de entender o efeito desta correção é verificar que quando a população é finita, as probabilidades de seleção de cada unidade não são mais iguais (dadas por  $1/N$ ), caracterizando um desenho amostral “sem reposição”, diferente do desenho anterior que pode-se considerar “com reposição”. O termo acima é a razão entre as imprecisões desses dois desenhos, que como será discutido na seção seguinte, é conhecido como efeito de desenho (*deff*).

Tabela 1: Convergência do fator  $(N-n)/(N-1)$  em função do tamanho proporcional para três tamanhos amostrais

N	$(N-n)/(N-1)$		
	n=100	n=1000	n=3000
5.000	0,98020	0,80016	0,40008
10.000	0,99010	0,90009	0,70007
30.000	0,99670	0,96670	0,90003
50.000	0,99802	0,98002	0,94002
100.000	0,99901	0,99001	0,97001
1.000.000	0,99990	0,99900	0,99700
10.000.000	0,99999	0,99990	0,99970

Considerando-se agora o tamanho da população, o tamanho da amostra pode ser obtido de forma análoga ao caso anterior, utilizando-se o novo valor para  $EP_{\hat{p}}$ . Assim,

$$\varepsilon = z_{\alpha/2} \sqrt{\frac{N-n}{N-1} \frac{P(1-P)}{n}} \tag{12}$$

Elevando ao quadrado ambos os lados da equação acima e, após algumas manipulações algébricas, obtém-se o seguinte tamanho amostral para uma prevalência (ou incidência acumulada):

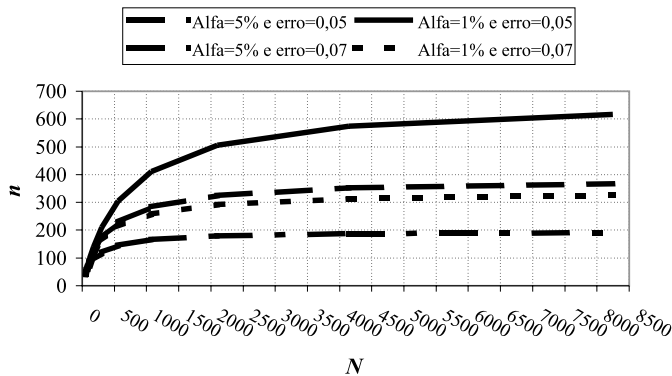
$$n = \frac{z_{\alpha/2}^2 NP(1-P)}{\varepsilon^2 (N-1) + z_{\alpha/2}^2 P(1-P)} \tag{13}$$

onde os elementos influentes  $\varepsilon$ ,  $P$  e  $z_{\alpha/2}^2$  são aqueles já definidos anteriormente e  $N$  é o tamanho da população.

Pode-se agora observar o papel do tamanho da população sobre o tamanho da amostra. Foi dito que o tamanho da população tem um

efeito importante no cálculo do tamanho da amostra apenas para populações “pequenas”. Para populações “grandes”, não é necessário se preocupar com ele. Uma vez fixados os valores de  $\epsilon$ ,  $P$  e  $z^2_{\alpha/2}$ , o tamanho amostral  $n$  pode ser escrito somente como função de  $N$ . A Figura 1, a seguir, ilustra esta propriedade, fixando uma prevalência presumida  $P=0,5$  (ou 50%) e combinando dois valores para  $\alpha$  (5% e 1%) e  $\epsilon$  (0,05 e 0,07). Percebe-se que o tamanho da amostra converge para um certo valor à medida que o tamanho da população aumenta. Quando  $\alpha = 5\%$  e  $\epsilon=0,05$ , este valor é igual a 384, obtido a partir da expressão (6). Para todas as curvas ilustradas, o tamanho amostral muda bastante quando se altera o tamanho populacional de 500 para 1000 indivíduos. Entretanto, quando se altera de 6500 para 7000, uma variação populacional também de 500 indivíduos, o tamanho amostral não se altera tanto.

Figura 1: Tamanho amostral segundo o tamanho populacional para se estimar uma prevalência, presumida em 50%



Embora este exemplo tenha sido construído para apenas dois valores de  $\epsilon$  e  $\alpha$  e para um valor fixado  $P$ , para quaisquer outras combinações de valores a forma da função será a mesma. O que difere é a “velocidade” com que ocorre a convergência para um tamanho amostral. Isto é, para algumas combinações de valores  $\epsilon$ ,  $P$  e  $\alpha$ , pode-se perceber mais “rapidamente” a convergência que para outras.

De forma análoga ao caso de uma população considerada infinita, pode-se chegar facilmente a um tamanho de amostra, considerando também o tamanho populacional, onde se considere o erro relativo  $\epsilon_r$ , e não

o erro absoluto  $\varepsilon$ . Ou seja, após algum algebrismo,

$$n = \frac{z_{\alpha/2}^2 N(1-P)}{\varepsilon_r^2 P(N-1) + z_{\alpha/2}^2 (1-P)} \quad (14)$$

### 3.4 O EFEITO DE DESENHO (*deff*)

Outra característica que exerce influência no tamanho amostral é o desenho amostral, isto é, de que forma as unidades amostrais são selecionadas. Os principais desenhos amostrais são amostragem aleatória simples (*AAS*), com ou sem reposição, amostragem estratificada (*strat*) e amostragem por conglomerados (*cluster*) (Silva, 1998).

As expressões colocadas até agora para o erro-padrão, tanto para população infinita quanto finita, são obtidas para um processo de amostragem aleatória simples, que do ponto de vista teórico é o mais simples. Para estimação de uma proporção populacional (ou de um parâmetro  $\theta$  qualquer), as expressões para o erro-padrão, tanto para uma amostragem estratificada quanto para conglomerados, são mais complexas. Entretanto, do ponto de vista prático (ou operacional), uma amostragem por conglomerados é mais simples e, freqüentemente, a mais usada (Levy & Lemeshow, 1999). Mas determinar um tamanho amostral sob este desenho é mais difícil devido à forma mais complexa de seu erro-padrão.

De forma geral, para a estimação de um parâmetro  $\theta$  tem-se a seguinte relação

$$EP_{strat} < EP_{AAS} < EP_{cluster} . \quad (15)$$

Isto é, usar conglomeração implica em uma maior imprecisão na estimativa que usar uma amostragem aleatória simples, que por sua vez, implica em mais imprecisão que uma amostra estratificada. Embora esta relação valha na prática, ela não é matematicamente verdadeira, podendo-se encontrar casos onde seja inversa (Bolfarine & Bussab, 1994). Em outras palavras, para uma mesma precisão, uma amostragem por conglomerados deve exigir uma amostra maior que uma amostragem aleatória simples, e uma amostragem estratificada seria a que exigiria o menor tamanho amostral.

Nestas circunstâncias, o ideal seria então trabalhar com amostras estratificadas. Entretanto, tal como a amostragem aleatória simples, ela é freqüentemente de difícil operacionalização, dependendo, além de um cadastro geral, da informação sobre qual estrato a unidade pertence. Além disso, se há muita heterogeneidade entre os estratos, o que seria bom no sentido de diminuir o erro-padrão, ter uma medida geral para a população pode ser de pouca utilidade para o pesquisador. Mais interessante seria manter a informação por estrato.

Em sendo então a amostragem por conglomerados a de mais fácil operacionalização, como calcular de forma simples um tamanho amostral para estimação de um parâmetro usando este desenho? Define-se uma quantidade conhecida por *deff* (*design effect*= efeito de desenho) como sendo a razão entre as imprecisões associadas à estimação de um parâmetro sob dois desenhos amostrais. Usando a amostragem aleatória simples como referência, tem-se:

$$deff = \frac{EP_{cluster}^2}{EP_{AAS}^2} \quad (16)$$

O *deff* funciona como um “preço” a ser pago pelo pesquisador por ter sua tarefa facilitada ao investigar apenas os *clusters* sorteados, aumentando sua imprecisão devido às possíveis correlações das unidades amostrais dentro e entre os *clusters*. Um valor de *deff*=1 é, obviamente, indicação de que a conglomeração não tem nenhum efeito sobre o tamanho amostral quando comparada com a amostragem aleatória simples.

Assim, pode-se simplificar o cálculo do tamanho amostral usando a teoria da Amostragem Aleatória Simples quando o estudo deve ser feito por Conglomeração, corrigindo a imprecisão maior devido a este desenho amostral através de algum valor para o *deff*. Surge, por outro lado, uma outra dificuldade que é saber qual o valor a ser atribuído ao *deff* para se corrigir adequadamente o tamanho amostral. Para se atribuir um valor razoável deveria se ter uma idéia da variabilidade das observações entre e dentro dos *clusters*, o que freqüentemente não se dispõe. No caso de um estudo de prevalência, quanto mais próximas as prevalências entre os *clusters*, menor seria o *deff* necessário. Na prática, salvo em situações especiais, um *deff* de 1,4 ou 1,5 (uma correção no tamanho amostral entre 40% e

50%) já deve ser suficiente para resguardar a precisão desejada do pesquisador.

### 3.5 GENERALIZANDO

De forma geral, o tamanho amostral para se estimar um parâmetro populacional  $\theta$  qualquer (como prevalência, incidência, risco relativo, razão de chances, média, diferença de duas proporções ou diferença de duas médias) pode ser obtido através da expressão seguinte, que relaciona o erro tolerável de amostragem à idéia de intervalo de confiança para estimação do parâmetro  $\theta$ . A premissa necessária é o conhecimento da expressão do erro-padrão associado a um estimador,  $\hat{\theta}$ , daquele parâmetro. A regra geral é:

$$\varepsilon = z_{\alpha/2} EP_{\hat{\theta}}, \quad (17)$$

onde  $EP_{\hat{\theta}}$  é o erro-padrão do estimador  $\hat{\theta}$ . O tamanho amostral estará sempre embutido em  $EP_{\hat{\theta}}$ . Uma vez identificada a expressão  $EP_{\hat{\theta}}$ , basta explicitá-la em  $n$  e, assim, obter-se uma expressão para o tamanho amostral para estimação de um parâmetro  $\hat{\theta}$ . Pela expressão acima, os elementos  $\varepsilon$  e  $\alpha$  (as especificações de erros) certamente exercem influência em  $n$  e eventuais outros elementos influentes aparecerão dependendo da expressão para  $EP_{\hat{\theta}}$ .

Poder-se-ia listar aqui os erros-padrões dos principais estimadores envolvidos numa investigação epidemiológica (Sahai & Khurshio, 1996). Entretanto, dada a assimetria das distribuições amostrais de alguns deles, suas expressões são mais complexas, apresentando-se numa escala logarítmica (freqüentemente de base  $e$ ). Mesmo assim, poder-se-ia a partir delas e algum algebrismo obter o tamanho amostral utilizando a equação geral acima.

### 4. TESTANDO UMA HIPÓTESE

Se, por outro lado, o pesquisador tem como objetivo testar uma hipótese previamente estabelecida, surge um outro elemento influente na determinação do tamanho amostral. Suponha que a situação de interesse seja investigar se um novo tratamento é melhor que um tratamento tradicional, cuja proporção de sucesso é conhecida ser de 70%. A pergunta

inicial é quantos doentes precisaria investigar para decidir ou não pelo tratamento alternativo com uma margem de erro aceitável. Isto é, se o tratamento novo de fato é melhor, por exemplo com um sucesso de 80%, gostaria de se poder evitar que o estudo não fosse capaz de chegar a esta conclusão devido a uma amostra pequena.

Ou seja, deseja-se controlar o erro  $\beta$ , sendo esse o elemento novo a ser considerado na determinação do tamanho amostral. Se, por outro lado o tratamento novo não é melhor, também não é desejável acidentalmente concluir que este é melhor. Ou seja, o interesse está em controlar o erro  $\alpha$ . Se o tratamento tradicional é satisfatório, seria conveniente um  $\alpha$  pequeno. Entretanto, se não, poderia se relaxar um pouco no  $\alpha$  e exigir mais do  $\beta$ . A determinação dos valores de  $\alpha$  e  $\beta$  depende de cada estudo. Quanto menor um, maior será o outro. A maneira de se ter ambos pequenos é aumentando o tamanho da amostra.

Para se chegar a um tamanho amostral com este novo elemento, considere a Figura 1 a seguir. As duas curvas representam a distribuição amostral para a proporção amostral ( $\hat{P}$ ) de sucesso no tratamento novo sob os dois valores hipotetizados no exemplo ( $P_0=70\%$ , compatível com o tradicional, e  $P_a=80\%$ ). A “largura” de cada uma dessas curvas é dada pelo seu erro-padrão que depende do tamanho da amostra e do valor hipotetizado. Para um “certo” tamanho amostral, há um valor limite  $p$  que um resultado a partir dele conduziria erroneamente à rejeição da hipótese  $P_0=70\%$  com uma probabilidade de até  $\alpha\%$ . Simultaneamente, em sendo verdadeira a hipótese  $P_a=80\%$ , um resultado até  $p$  conduziria à aceitação daquela hipótese, também erroneamente, com uma probabilidade de até  $\beta\%$ . O que se quer é determinar qual é este “certo” tamanho amostral. Utilizando a normal padronizada, sob  $P_0=70\%$ , tem-se

$$z_\alpha = \frac{p - 0,7}{\sqrt{\frac{0,7 \times 0,3}{n}}} \Rightarrow p = 0,7 + z_\alpha \sqrt{\frac{0,21}{n}}$$

E sob  $P_a=80\%$ ,

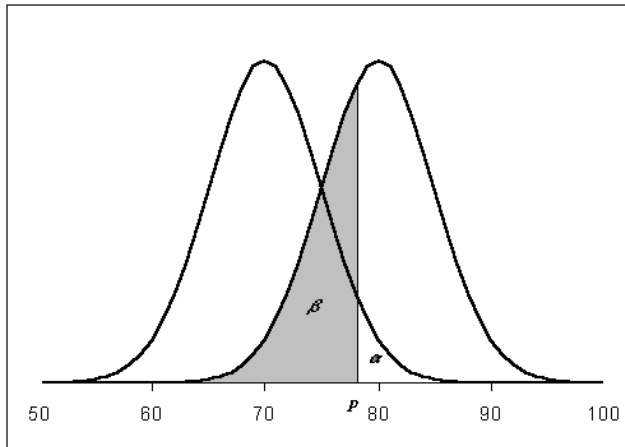
$$-z_\beta = \frac{p - 0,8}{\sqrt{\frac{0,8 \times 0,2}{n}}} \Rightarrow p = 0,8 - z_\beta \sqrt{\frac{0,16}{n}}$$

Igualando-se as duas últimas expressões para  $p$  e

$$n = \frac{(z_\alpha \sqrt{0,21} + z_\beta \sqrt{0,16})^2}{(0,8 - 0,7)^2}. \quad (18)$$

Estabelecendo-se  $\alpha=5\%$  (com  $z_\alpha = 1,64$ ) e  $\beta=20\%$  (com  $z_\beta = 0,84$ ), chega-se a um tamanho amostral de 119 unidades. Isto é, com 119 unidades o pesquisador teria um poder de 80% ( $1-\beta$ ) de detectar uma proporção de sucesso de 80% do tratamento novo com um nível de significância de 5%.

Figura 2



De forma geral tem-se então a seguinte expressão para determinação do tamanho amostral para se testar uma hipótese sobre uma proporção:

$$n = \frac{[z_\alpha \sqrt{P_0(1-P_0)} + z_\beta \sqrt{P_a(1-P_a)}]^2}{(P_0 - P_a)^2}. \quad (19)$$

Esta expressão se aplica para um teste unilateral. Isto é, no exemplo, apenas valores acima de 70% estariam sendo considerados. Se deseja-se um teste bilateral, poderia usar a expressão acima substituindo  $z_\alpha$  por  $z_{\alpha/2}$  (Levy & Lemeshow, 1999).

## 5. EXTENSÕES

De maior aplicação em investigações epidemiológicas são os estudos



comparativos - os ensaios clínicos e os estudos observacionais do tipo coorte e caso-controle - onde quer-se estudar a associação entre duas variáveis dicotômicas, classicamente denominadas exposição e doença. Em qualquer um desses três desenhos, pode-se calcular um tamanho amostral a partir da comparação de duas proporções, sendo que no caso-controle a comparação seria sobre as proporções de expostos. De forma geral, o tamanho amostral para a comparação entre duas proporções segue a mesma lógica já discutida para o caso de uma única amostra, tanto para estimação quanto para testes de hipóteses. A diferença básica é que agora tem-se duas amostras e, assim, será necessário ser estabelecido *a priori* pelo pesquisador uma razão desejada entre os dois tamanhos amostrais. As fórmulas também ficam mais complexas, principalmente quando se tem proporções pequenas.

Como em epidemiologia as medidas de associação conhecidas como Risco Relativo e Razão de Chances têm larga aplicação e interpretações imediatas, pode-se construir tamanhos amostrais diretamente para estimação destes parâmetros. As expressões para seus erros-padrões são mais complexas (Sahai & Khurshio, 1996), o que impede uma expressão simples para determinação do tamanho amostral. Mas a identificação dos elementos influentes e sua interpretação seguem a mesma lógica já apresentada.

Embora não tão freqüente numa investigação epidemiológica, um pesquisador pode estar interessado em conhecer qual o nível médio de uma variável numérica, por exemplo, glicemia ou colesterol em uma população. Tal como no caso da estimação de uma proporção, o tamanho amostral para a estimação de uma média populacional pode ser facilmente obtido a partir da expressão para o erro-padrão da média amostral ( $\bar{X}$ ), utilizando-se o raciocínio desenvolvido na seção 3.5. Considerando uma população infinita (amostragem aleatória simples “com reposição”), tem-se:

$$EP_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad (20)$$

E no caso de uma população finita de tamanho  $N$ ,

$$EP_{\bar{X}} = \sqrt{\frac{N-n}{N-1} \frac{\sigma^2}{n}} \quad , \quad (21)$$

onde  $\sigma$  é o desvio-padrão populacional da variável em estudo. A analogia com a estimação de uma proporção é imediata, observando-se que  $\sigma = \sqrt{P(1-P)}$ . Assim, os tamanhos amostrais para estimação de uma média podem ser obtidos tanto para um erro absoluto ( $\epsilon$ ) quanto para um erro relativo ( $\epsilon_r$ ) da mesma maneira como aqueles obtidos para a estimação de uma proporção.

Para o caso de um teste de hipótese a analogia também é imediata. Pela seção 4, pode-se calcular um tamanho amostral para um teste bilateral de uma média, supondo população infinita, através de:

$$n = \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\epsilon^2}, \quad (22)$$

onde  $\sigma$  é o desvio-padrão populacional da variável em estudo. Como esta informação não é conhecida (é populacional), pode-se usar no seu lugar alguma informação disponível em outros estudos similares. Um recurso que pode ser útil é observar que a diferença entre o máximo e o mínimo (a amplitude amostral) pode fornecer alguma informação sobre o desvio-padrão. Supondo que os dados são normais, a amplitude corresponde a aproximadamente seis vezes o desvio-padrão.

Em epidemiologia, a comparação de duas médias encontra mais aplicações nos estudos clínicos. Por exemplo, um pesquisador pode estar interessado em saber se há diferença entre dois tratamentos alternativos para controle de uma variável numérica, colesterol, por exemplo. Há uma diferença na redução média do colesterol para estes tratamentos? O tamanho amostral necessário para este tipo de investigação segue também a mesma lógica anterior. O erro-padrão associado à diferença entre duas médias é:

$$EP_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}. \quad (23)$$

Freqüentemente, assume-se que os desvios-padrões dos dois grupos são iguais. Assim, um tamanho amostral, em cada grupo, para um teste bilateral da diferença entre duas médias seria dado por:

$$n = \frac{2(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{d^2}, \quad (24)$$

onde  $d$  é o valor hipotetizado para a diferença entre as médias.

## 6. CONCLUSÃO

Em síntese, para a determinação do tamanho de uma amostra é necessária a colocação de uma pergunta específica que possa ser respondida pela estimação de um parâmetro populacional. Pela distribuição amostral de um estimador deste parâmetro e através de seu erro-padrão associado, pode-se obter um tamanho amostral seguindo o raciocínio geral desenvolvido na seção 3.5, permitindo-se assim identificar claramente seus elementos influentes.

Uma outra característica importante na determinação do tamanho amostral é que qualquer estudo estará sujeito à perda de dados por razões diversas e magnitudes diferenciadas, dependendo do estudo. O pesquisador obtém seu tamanho amostral em cima de especificações de erro que ele julgou pertinentes, mas que serão alteradas se houver perda de dados. Querendo se resguardar de sua precisão, o pesquisador deve majorar um pouco sua amostra a fim de compensar eventuais perdas. É difícil estabelecer um valor de compensação, pois cada estudo estará sujeito a diferentes níveis de perda. Mas, arriscando uma regra geral, um acréscimo de 20% poderia ser recomendado. Por outro lado, mais importante do que um eventual aumento na imprecisão devido a uma perda grande seria considerar a possibilidade das unidades perdidas serem “diferentes” das obtidas, implicando em potenciais vieses no estudo.

Como último comentário, e concretamente falando, qualquer tamanho amostral para uma investigação epidemiológica implicará em custos. Então, a determinação do tamanho amostral nem sempre poderá estar restrita a uma discussão puramente sobre imprecisão. Como objetivamente este conceito é determinante no tamanho da amostra, em havendo limitação de recursos, o pesquisador poderá chegar a um tamanho amostral “viável” se relaxar um pouco em suas especificações de erro.

## REFERÊNCIAS BIBLIOGRÁFICAS

BOLFARINE, H.; BUSSAB, W. O. *Elementos de Amostragem*. Belo Horizonte, XI SINAPE. 1994.

FLEISS, J. L. *Statistical Methods for Rates and Proportions*. 2 ed. New York: John Wiley & Sons. 1981.

LEVY, P. S.; LEMESHOW, S. *Sampling of Populations: methods and applications*. New York: John Wiley & Sons. 1999.

SILVA, N. N. *Amostragem probabilística: um curso introdutório*. São Paulo: EDUSP. 1998.

SOARES, J. F.; SIQUEIRA, A. L. *Introdução à Estatística Médica*. Belo Horizonte: Dep<sup>o</sup> de Estatística da UFMG. 1999.

SAHAI, H.; KHURSHIO, A. *Statistics in Epidemiology Methods, Techniques, and Applications*. Boca Raton: CRC Press. 1996.