

Amostragem

Tania Z. Guillén de Torres
Monica Maria Ferreira Magnanini
Ronir Raggio Luiz

INTRODUÇÃO

Tal como em qualquer área da saúde, a realização de pesquisas por amostragem surge da necessidade de fornecer informações que serão úteis para diversos fins, tais como investigação de relações de causa e efeito de doenças, planejamento e gestão de serviços, melhoria da qualidade de atendimento, caracterização de comportamentos de risco etc. Assim estas informações são constantemente solicitadas por políticos, serviços públicos, entidades não-governamentais, além dos pesquisadores.

Na operacionalização de qualquer estudo, geralmente não é razoável se examinar todos os elementos da população de interesse. Razões típicas para isso seriam limitações de recursos humanos ou materiais, devido à população ser infinita, porque o processo de observação é destrutivo, ou, ainda, por questões éticas. Nestes casos, observa-se somente uma pequena parte da população, chamada de amostra. Embora se associe à noção de amostra um subconjunto contável de elementos de uma população – tal como pacientes – uma gota de sangue também constitui uma amostra. Vale salientar que muitos acreditam que só pelo fato de se observar todos os elementos da população, as estimativas obtidas serão mais “precisas”. Entretanto, geralmente os erros que ocorrem na coleta e no manuseio de um grande número de dados são maiores do que as imprecisões a que se está sujeito quando, no processo de inferência, generalizam-se as conclusões de uma amostra bem selecionada.

O primeiro passo em qualquer pesquisa estatística é identificar os objetivos. Os mesmos nem sempre estão claros e podem até mesmo alterar-se, ao longo do estudo, daqueles inicialmente sugeridos. Os objetivos devem ser formulados de maneira clara, pois eles servem de base para as outras etapas da pesquisa.

A seguir, recomenda-se especificar melhor o que se deseja conhecer da população em estudo, formulando os objetivos específicos. Eles fornecerão uma idéia das características a serem observadas ou medidas. Nesta etapa é muito importante consultar também outras pessoas envolvidas no projeto, assim como especialistas na área de interesse, não somente para clarificar os objetivos, mas também para verificar a existência de alguma outra informação relevante que seja necessária à pesquisa.

Uma vez esclarecidos os objetivos, deve-se identificar a variável ou as variáveis apropriadas para poder atingi-los, assim como listar fatores que poderiam afetar a variável resposta. Por exemplo, para estudar o tempo de sobrevivência em pacientes com Aids, seria recomendável também ter informação sobre a idade do paciente, uso de medicamentos anti-retrovirais, adesão ao tratamento, doenças oportunistas etc. A obtenção da informação sobre a variável resposta e demais variáveis envolvidas no estudo pode-se dar através de fonte primária (ex.: questionários, entrevistas) ou secundária (ex.: anuário estatístico). Usando uma fonte primária, é essencial garantir que os dados coletados sejam confiáveis, acurados e representativos.

Pode-se dizer que a estatística consiste basicamente em coletar dados para sua posterior análise e interpretação. A seleção dos elementos que serão efetivamente observados deve ser feita a partir do uso de uma metodologia adequada, de modo que os resultados da amostra sejam informativos para avaliar características de toda a população. Para que isto seja possível, é necessário que a amostra seja representativa da população da qual ela foi extraída, de modo que o comportamento da variável em estudo na amostra seja o reflexo do seu comportamento na população, e que, assim, os valores das estatísticas avaliadas nesta amostra sejam uma "boa" aproximação dos parâmetros populacionais correspondentes. Caso contrário, corre-se o risco de ver o processo de inferência comprometido, ao se gerar estimativas enviesadas dos parâmetros. Uma das formas de se conseguir representatividade é fazer com que o processo de escolha da amostra seja, de alguma forma, aleatório. Esse processo permite o cálculo de estimativas dos erros que estarão envolvidos na inferência.

Resumidamente, amostragem se refere então ao processo de identificação de como, quantos e quais serão os elementos da população a serem estudados.

Neste capítulo serão apresentados então os conceitos de população e amostra, os métodos de seleção de amostras mais comumente usados na prática, as vantagens e desvantagens de usar amostragem nas pesquisas e quais os possíveis erros decorrentes do processo de amostragem.

A Figura 22.1 ilustra esquematicamente o processo de uma pesquisa por amostragem.

NOÇÕES PRELIMINARES

O que É Amostragem?

A amostragem é um campo da estatística bastante desenvolvido e sofisticado que estuda técnicas de planejamento de pesquisa que possibilitam a realização de inferências sobre um universo a partir do estudo de parte de seus componentes: a amostra. Como busca a inferência, isto é, de um ponto de vista estritamente estatístico, o procedimento somente será válido se a amostra for representativa da população. A amostragem visa diminuir os custos e o tempo do processo de estimação, mantendo uma precisão mínima segura. Pode-se reconhecer na amostragem a essência da teoria estatística. Alguns textos clássicos de amostragem são Cochran (1977), Kish (1965) e Levy e Lemeshow (1999).

População e Amostra

Como visto no capítulo Fundamentos da Pesquisa Epidemiológica, considera-se *população-alvo* ou *base populacional* o conjunto de elementos para os quais se deseja que as conclusões oriundas da pesquisa sejam válidas. A *população-fonte* é definida como o conjunto de elementos de onde serão selecionados aqueles que farão parte da

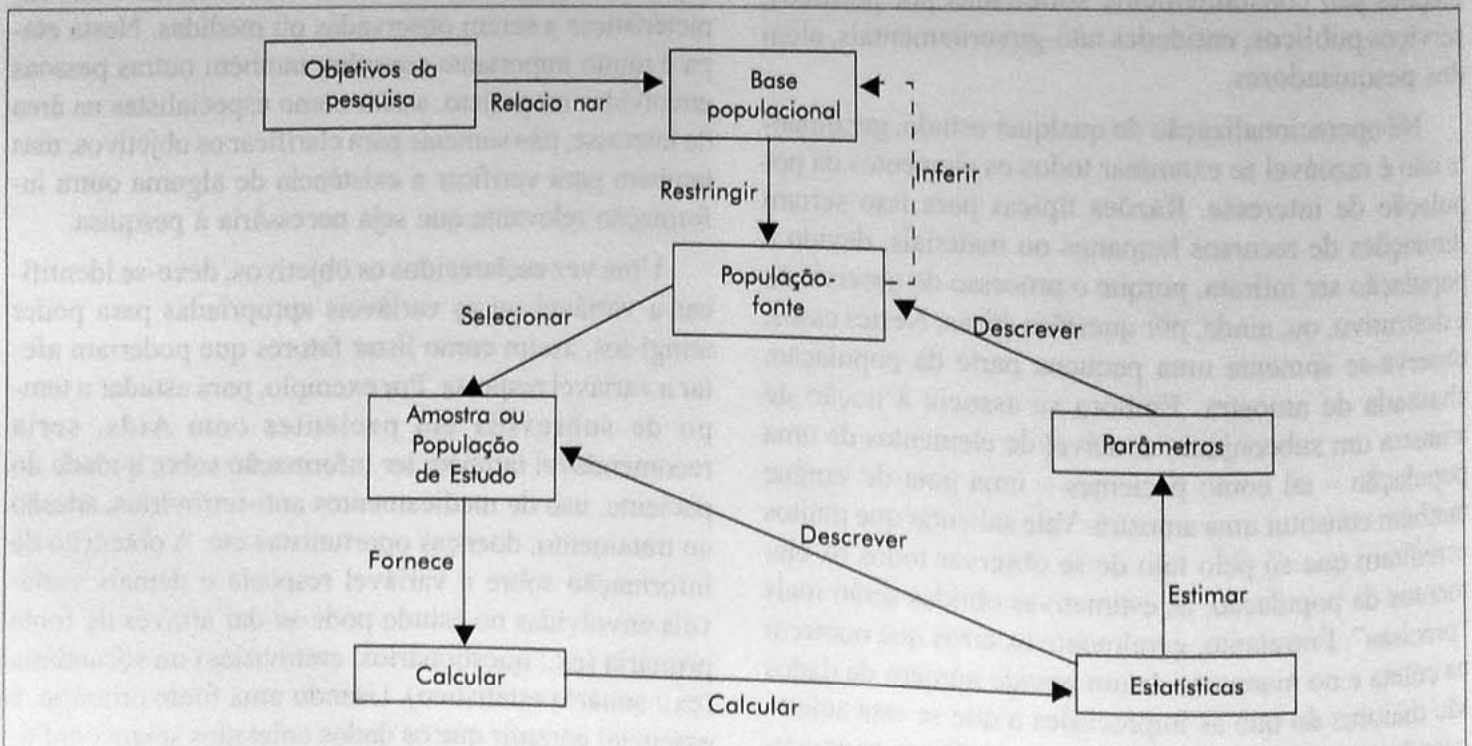


Fig. 22.1 - Esquemática do processo amostral.

investigação, representando, portanto, os potenciais elementos para constituir a *amostra* (ou *população de estudo*). Assim, amostra ou população de estudo seria o subgrupo de elementos da população-fonte sobre os quais se fazem as observações e coletam-se os dados. Na definição da população-fonte deve-se levar em conta além dos objetivos da pesquisa, as características a serem observadas e os recursos disponíveis. Os membros da população cuja característica será medida são chamados de *unidades elementares* ou *elementos* da população.

Por exemplo, se uma pesquisa tem como objetivo estimar a prevalência de obesidade entre jovens de 13 a 18 anos na Cidade do Rio de Janeiro, a base populacional, portanto, seria constituída por todos os jovens nessa faixa etária, residentes na Cidade do Rio de Janeiro. Imaginando que jovens nessa faixa etária devam estar na escola, a população de estudantes com essas idades cadastrados nas escolas (públicas e particulares) constituiria a população-fonte do estudo. Havendo dificuldades operacionais, a população-fonte poderia ser reduzida a estudantes da rede pública, mas, neste caso, talvez fosse mais conveniente a restrição da população-alvo apenas aos estudantes das escolas públicas. Cada jovem de 13 a 18 anos constituiria as unidades elementares.

Municípios também podem representar as unidades elementares, como é o caso, por exemplo, dos estudos ecológicos.

Por uma questão de notação, o número total de unidades elementares na população é comumente denotado por N e a cada unidade elementar será associado um número de 1 até N . As características da população ou variáveis aleatórias serão denotadas com letras maiúsculas, por exemplo Y , e o valor desta característica para a i -ésima unidade elementar será Y_i .

Um dos objetivos da pesquisa por amostragem é estimar certos valores como médias, proporções, totais, *odds* etc., obtidos através de funções matemáticas, relacionados com a distribuição de uma certa característica da população em estudo. Estes valores populacionais são chamados de *parâmetros*, e para uma dada população os valores dos parâmetros são constantes. Por outro lado, as funções matemáticas são chamadas de *estimadores* ou *estatísticas* dos parâmetros populacionais. A média amostral de uma certa variável, por exemplo, é um estimador da média populacional (o parâmetro), e uma vez obtida a amostra e calculado seu valor, obtém-se uma *estimativa*.

VANTAGENS E DESVANTAGENS DA AMOSTRAGEM

A seguir, são apresentadas algumas das razões para se preferir realizar uma pesquisa por amostragem:

- é mais econômica, e pode fornecer resultados mais acurados e abrangentes o levantamento da informação de uma parte ou amostra, do que observar a população toda, o que por vezes é impossível;
- a informação estará disponível em um prazo menor de tempo, já que o volume de dados a coletar é menor. Em consequência, será menor também o tempo necessário para criticar e apurar as informações. Este fator torna-se crucial, por exemplo, nas pesquisas eleitorais que precisam fornecer estimativas rápidas num curtíssimo prazo de tempo;
- uma amostra bem selecionada pode fornecer resultados mais acurados, ao diminuir a possibilidade de viés, incrementando a precisão dos resultados;
- pode-se diminuir os erros nas respostas, pois, ao observar-se um número menor de elementos, é possível dar mais atenção aos casos individuais. Neste sentido também é possível dedicar mais tempo e recursos para localizar e persuadir os não-respondentes, diminuindo assim o viés de não-resposta;
- a operacionalização e o controle da qualidade do trabalho em cada uma das etapas tendem a ser mais satisfatórios que nos censos (observação de toda a base populacional), principalmente quando a população é muito grande e dispersa como é, por exemplo, o caso do Brasil;
- quando se deseja estudar de forma detalhada certos subgrupos, pode-se desenhar uma amostra com maior representatividade, para que estes possam ser mais bem caracterizados e incrementar a precisão dos resultados;
- desenhos amostrais podem aumentar a precisão e diminuir os vieses em estudos observacionais.

Existem também situações nas quais não se justifica o esforço de selecionar uma amostra, como por exemplo:

- quando a população é muito pequena, não se justificaria todo o gasto em recursos e tempo necessários para selecionar uma amostra relativamente grande a fim de gerar resultados precisos;
- se a característica que se quer estudar é de fácil observação, pode não compensar a realização de um plano de amostragem, mesmo quando a população não seja muito pequena. Pode acontecer também que a informação necessária já esteja disponível como subproduto de alguma outra pesquisa;
- quando é preciso estudar características ou conhecer parâmetros que sejam fundamentais para o planejamento de um país ou para servir de referência para a realização de outras pesquisas, justifica-se a

realização de um levantamento de informação através de um censo.

Plano de Amostragem

Existem diferentes planos de amostragem e a escolha vai depender de vários fatores, tais como tamanho e tipo da população, perguntas a serem respondidas na pesquisa, tempo e recursos disponíveis (pessoal e material), acurácia desejada etc. Por exemplo, para obter o número de pacientes atendidos por dia, num determinado ambulatório de um Hospital Universitário, e para poder estimar a prevalência de hipertensão arterial na Cidade do Rio de Janeiro, são necessários planos de amostragem diferentes. Outras terminologias possíveis para plano amostral são desenho amostral, esquema amostral ou delineamento amostral.

A formulação das perguntas ajuda a definir os objetivos da pesquisa, os parâmetros que precisam ser estimados, as variáveis que irão ser medidas ou observadas, e os níveis de confiabilidade e validade para essas estimativas. Deve-se definir também o tamanho da amostra (número de elementos que devemos selecionar) e a forma de seleção ou critério de escolha dos elementos que farão parte dela. Também é conveniente avaliar a necessidade de dispor de planos de amostragem e coleta de dados alternativos.

No plano de amostragem, deve-se indicar também a definição da unidade amostral. Ou seja, a unidade a ser selecionada para se chegar aos elementos da população. Esta pode ser constituída do próprio elemento da população ou de outras unidades que sejam mais fáceis de serem selecionadas e que, de alguma forma, estejam associadas aos elementos da população. Por exemplo, para conhecer o perfil dos alunos de medicina de uma determinada região, as faculdades seriam as unidades de amostragem, que permitiriam chegar até os alunos, ou elementos da população.

As técnicas de inferência estatística usam a hipótese de a amostragem ser probabilística. Ou seja, a escolha dos elementos da amostra deveria ser realizada através de algum tipo de sorteio. Porém, nem sempre isto é possível de ser realizado. Pode acontecer de a população ser infinita ou não ser totalmente acessível, sendo necessário nestes casos o uso de amostragem não-aleatória. Assim, a coleta de dados numa amostragem pode ser classificada de duas formas, dependendo de como a amostra foi selecionada. Ela é *probabilística* se cada elemento da população tem associado uma probabilidade conhecida de ser incluído na amostra, ou ela é *não-probabilística*, se o processo de seleção da amostra não leva em conta as probabilidades de cada

elemento ser incluído na amostra. Como dito, nas amostragens probabilísticas, onde cada elemento da população tem uma probabilidade conhecida de ser selecionado, a confiabilidade das estimativas da população pode ser avaliada objetivamente usando a teoria de probabilidades, daí a possibilidade de realizar inferências.

AMOSTRAGENS ALEATÓRIAS OU PROBABILÍSTICAS

Nesse tipo de amostragem, a escolha dos elementos da amostra é realizada de forma aleatória (não depende do investigador), permitindo o cálculo das probabilidades na seleção. É esse fato que permite a utilização das técnicas clássicas de inferência estatística, facilitando a análise dos dados e permitindo generalizar resultados obtidos a partir da amostra para a população com um certo grau de segurança.

Os principais tipos de amostragem aleatória são: aleatória simples, sistemática, estratificada e por conglomerados.

Amostragem Aleatória Simples (AAS)

Suponha uma população de N elementos em que se deseja escolher uma amostra de n elementos, sem reposição. Usando a teoria de análise combinatória é possível mostrar que o número T de possíveis amostras é dado pela expressão a seguir:

$$T = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

onde:

$N! = N \times (N-1) \times \dots \times 1$ e, por convenção, $0! = 1$. Por exemplo, se $N = 10$ e $n = 3$ então o número de possíveis amostras será:

$$T = \binom{10}{3} = \frac{10!}{3!(10-3)!} = \frac{10 \times 9 \times 8 \times 7!}{3 \times 2 \times 1 \times 7!} = \frac{720}{6} = 120$$

A amostragem aleatória simples também é conhecida como amostragem ocasional, acidental, casual ou randômica. É um processo de seleção bastante fácil e muito usual, com a propriedade de que qualquer subconjunto da população, com o mesmo número de elementos, tem a mesma probabilidade de fazer parte da amostra, sendo esta probabilidade igual a:

$$\frac{1}{T} = \frac{1}{\binom{N}{n}}$$

Em particular, pode-se dizer que cada elemento da população tem a mesma probabilidade $\frac{n}{N}$ de pertencer à amostra.

Na seleção de uma amostra aleatória simples é necessário:

- dispor de um rol ou relação completa dos elementos da população ou das unidades de amostragem;
- numerar consecutivamente os elementos da população de 1 até N (tamanho da população);
- determinar n, o número de elementos a serem selecionados;
- selecionar cada um dos elementos da amostra por um mecanismo aleatório, que poderia ser um sorteio, como o empregado para escolher as dezenas ganhadoras da loteria. Nesse tipo de sorteio colocam-se todos os números misturados dentro de uma urna e os elementos são escolhidos um a um até completar a amostra. A Figura 22.2 ilustra esse esquema de amostragem. Quando cada elemento da população não pode ser sorteado mais que uma vez, o sorteio é dito sem reposição, e quando o elemento sorteado é colocado novamente na urna, antes de fazer uma nova extração, o sorteio é dito com reposição. Nesse caso, o número de possíveis amostras seria N^n . A escolha também pode ser feita através de algoritmos numéricos de geração de números aleatórios implementados na maioria dos pacotes estatísticos.

Um procedimento equivalente para escolher uma amostra aleatória simples é fazer uso da tabela de números aleatórios. A tabela é uma disposição retangular de dígitos, compreendidos entre 0 e 9, sorteados um a um de forma aleatória e com reposição. Para fazer uso

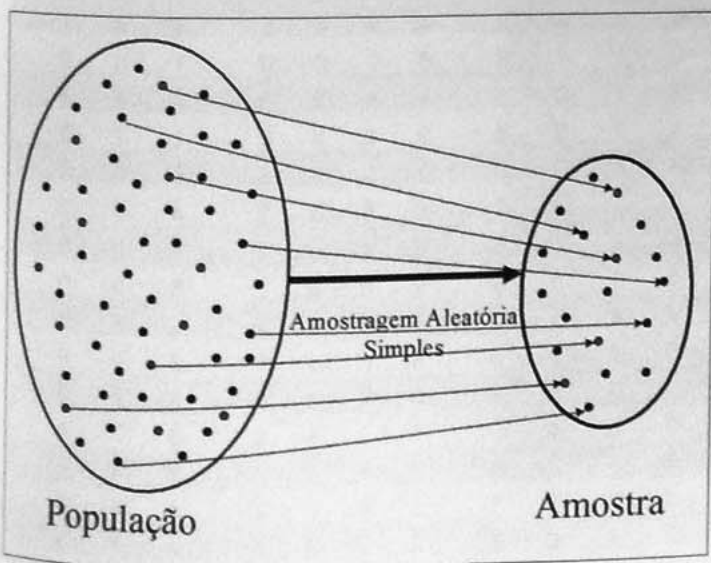


Fig. 22.2 – Escolha de elementos numa amostragem aleatória simples.

dessa tabela, devemos a princípio determinar o número de algarismos necessários para caracterizar N, escolher uma direção a seguir (à direita \rightarrow , à esquerda \leftarrow , para cima \uparrow , ou para baixo \downarrow) e um ponto de partida (uma linha e uma coluna, escolhidas de preferência de forma aleatória através de um sorteio). A partir desse ponto, começa-se a ler os números formados com aquela quantidade de algarismos seguindo a direção predeterminada. Por exemplo, se o tamanho da população for de 1.576.898 elementos, precisaremos de sete dígitos para caracterizar N. Na tabela ler-se-iam números de sete algarismos. Entretanto, se $N = 60$, são necessários somente dois dígitos para caracterizar N e ler-se-iam na tabela, números de dois algarismos.

Exemplo: Suponha que um pesquisador deseja escolher uma amostra de tamanho $n = 8$ elementos de uma população de 800 prontuários. Logo, precisa de três dígitos para caracterizar 800. Se ele escolhesse como ponto de partida a linha 1 e a coluna 1, e como direção à direita (\rightarrow), usando a tabela de números aleatórios apresentada na Tabela 22.1, leria números de três dígitos nas linhas começando da esquerda para a direita, obtendo os seguintes elementos: 184, 025, 660, 618, 483, 372, 392 na linha 1. Continuando a procura na linha 2 seguindo a mesma direção (\rightarrow), seria lido o elemento 645. Observe que o valor 830 não foi considerado por ser maior que 800. Nas amostragens sem reposição os números que sejam repetições daqueles lidos anteriormente, devem também ser descartados. O passo seguinte é usar os números assim escolhidos para identificar os elementos na lista de prontuários para compor a amostra.

A amostragem aleatória simples (AAS) é um procedimento de escolha da amostra que serve de base para o desenvolvimento da teoria de amostragem, sendo também usada como referência para comparação dos demais tipos de amostragem, como será visto mais adiante na discussão sobre o efeito do desenho. Uma desvantagem da AAS é que pode ser bastante cara e às vezes impossível de ser realizada. Uma das razões seria a necessidade de identificação de todos os elementos da população, e uma outra, a possibilidade de que os elementos sorteados se localizem em áreas geográficas muito dispersas.

Amostragem Sistemática

Trata-se de uma variação da amostragem aleatória simples. É geralmente usada quando a população está naturalmente ordenada, como, por exemplo: listas telefônicas, prontuários em um hospital, peças num processo contínuo de produção etc. Às vezes ela é preferida à

Tabela 22.1
Tabela de Números Aleatórios

Linha	Coluna																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1	8	4	0	2	5	6	6	0	6	1	8	8	3	0	4	8	3	3	7	2	3	9	2
2	6	4	5	4	6	7	6	9	2	3	8	2	5	4	4	0	8	6	4	3	1	9	8	6
3	5	9	8	6	5	8	8	1	1	8	8	1	2	7	5	8	1	8	2	1	9	0	7	1
4	6	9	5	9	0	6	2	3	3	3	9	1	5	9	1	3	6	0	4	1	7	3	1	8
5	6	6	5	7	7	4	1	2	6	5	3	0	9	4	1	4	7	7	8	4	2	5	8	6
6	1	4	1	8	5	4	7	0	2	7	9	5	2	6	7	5	5	8	5	7	5	2	0	5
7	6	1	8	9	0	4	5	5	2	9	1	4	3	1	7	8	0	3	8	8	0	2	2	9
8	0	1	3	3	3	1	2	2	8	1	8	1	0	7	3	1	3	2	2	1	8	9	4	2
9	5	1	2	3	3	9	2	8	3	5	4	4	4	7	0	3	0	1	9	7	7	0	7	0
10	8	5	8	2	7	5	5	9	9	5	7	8	6	2	0	0	8	6	9	0	3	9	3	0
11	2	5	2	9	2	8	1	9	3	0	9	9	5	5	7	0	4	5	3	2	6	0	4	4
12	0	0	4	0	8	1	9	2	8	0	9	8	4	5	2	1	5	0	8	5	8	5	2	9
13	4	6	8	6	0	9	0	2	3	8	4	6	8	5	3	7	3	8	8	6	6	4	0	0
14	8	8	0	5	4	6	1	7	3	3	0	6	0	0	8	5	2	1	2	4	3	6	3	1
15	5	6	0	2	6	5	0	1	2	9	7	2	9	7	4	5	4	7	1	3	5	9	5	7
16	8	0	1	1	6	6	2	3	7	3	4	8	8	2	7	4	1	1	0	9	2	7	8	5
17	2	1	0	9	1	4	7	8	7	7	0	1	9	1	6	2	9	4	0	0	2	7	8	7
18	5	2	0	8	2	6	7	8	4	3	7	4	4	2	7	0	1	1	2	4	0	5	4	6
19	2	7	2	4	2	4	0	7	7	6	3	1	1	4	6	2	2	8	7	8	1	0	9	1
20	9	6	1	3	6	8	6	7	3	3	8	4	7	2	5	2	2	2	0	4	9	6	5	7
21	2	8	1	8	8	2	5	3	4	3	3	9	4	2	9	0	6	0	3	8	7	6	5	5
22	1	7	1	6	6	7	2	4	3	8	7	7	8	1	7	0	0	0	8	9	3	5	2	8
23	4	0	9	0	5	5	0	8	7	4	5	3	8	3	0	0	5	3	2	9	7	5	2	3
24	7	7	2	0	7	7	1	4	5	8	6	4	6	9	2	8	7	4	8	6	3	7	7	6
25	8	6	3	5	9	4	1	2	2	4	4	8	7	8	7	4	2	2	2	3	3	7	2	3
26	4	2	1	2	9	9	3	2	8	3	5	9	2	4	0	0	2	6	9	2	7	1	3	3
27	4	6	9	2	7	8	9	6	1	5	5	2	1	4	1	5	4	6	5	7	9	6	6	4
28	8	1	6	0	9	5	1	4	5	2	5	3	8	4	1	1	6	6	4	0	1	2	5	5
29	0	5	5	9	0	6	9	5	8	2	5	4	3	0	2	3	1	2	6	5	9	4	3	1
30	6	9	1	5	0	4	1	1	3	9	2	3	4	1	1	0	8	6	1	6	6	8	4	9
31	7	7	7	2	4	9	8	5	8	1	6	3	1	0	1	6	7	5	3	2	2	5	0	5
32	6	5	9	4	7	6	9	6	3	4	1	8	6	8	6	8	3	8	9	3	0	0	7	6
33	2	7	7	3	5	6	4	4	8	2	7	7	7	0	1	6	3	1	1	4	2	9	0	8
34	0	2	0	8	8	7	9	3	4	2	5	8	8	0	6	4	2	4	9	4	5	8	8	1
35	5	9	8	6	7	2	9	0	1	8	8	8	7	0	1	7	8	8	7	0	0	1	8	2
36	3	7	0	3	9	7	7	6	9	9	8	5	8	3	1	0	9	1	8	3	6	3	6	4
37	8	4	7	7	9	4	8	9	3	4	2	0	0	4	7	3	6	9	8	5	0	1	1	2
38	1	9	0	5	7	7	7	8	2	8	0	8	8	3	6	9	4	1	5	3	6	9	8	1
39	5	7	9	9	7	6	5	0	6	9	9	2	5	3	1	4	1	8	6	3	9	4	9	2
40	1	5	6	4	7	5	9	1	1	5	5	2	7	7	7	7	5	8	4	6	1	6	3	5
41	9	0	7	1	6	9	3	4	5	8	9	7	0	6	5	5	5	9	5	5	7	8	5	0
42	2	8	2	9	6	7	6	1	4	4	9	9	9	9	6	3	2	1	2	9	7	3	8	5
43	2	4	2	7	0	7	2	5	3	3	4	9	0	0	9	9	6	6	9	6	0	3	0	8
44	1	1	7	3	8	1	2	4	8	3	7	8	3	6	0	9	8	0	9	3	9	5	8	3
45	7	4	5	3	4	5	2	6	2	5	7	1	4	0	8	4	0	1	6	4	2	5	7	6
46	3	7	7	0	2	4	1	1	1	8	2	2	5	7	5	4	1	0	7	3	9	9	4	6
47	3	7	4	6	8	8	6	9	7	2	4	2	3	5	7	5	1	3	9	7	2	1	4	3
48	9	7	3	1	6	8	8	2	7	1	2	2	5	7	2	8	1	3	2	0	1	4	7	3
49	5	7	8	4	5	2	6	7	9	9	1	9	0	5	3	7	5	1	9	0	2	6	5	0
50	7	9	1	1	6	8	2	7	1	4	3	7	8	3	3	5	5	0	7	9	5	4	1	1

amostragem aleatória simples, por ser de mais fácil execução, estando, portanto, menos sujeita a erros, fornecendo informação um tanto mais precisa e com um custo menor. Essa última afirmação somente é válida nos casos em que não existe periodicidade ou ciclos nos elementos da população. Caso contrário, o processo de amostragem sistemática está sujeito a erros, gerando vies de seleção.

Para a seleção de uma amostra sistemática é necessário:

- dispor de um rol ou relação completa dos elementos da população ou das unidades de amostragem;
- numerar os elementos da população de 1 até N (tamanho da população);
- conhecer n, o número de elementos a serem selecionados;
- definir o passo ou intervalo de seleção: $a = N/n$;
- escolher um número aleatório r entre 1 e a, através da tabela de números aleatórios;
- selecionar os números de ordem dos demais elementos da amostra. Sendo r o primeiro, o segundo número escolhido será: $r + a$, o terceiro será: $r + 2a$, e assim por diante. Observe que os números selecionados constituem uma progressão aritmética com valor inicial r e período a;

- usar os números assim escolhidos para identificar os elementos da lista selecionados para compor a amostra.

Exemplo: Se o número de elementos em uma população de pacientes de um hospital fosse 160 e o tamanho da amostra desejado fosse de 40 pacientes, o processo de seleção da amostra seria:

- calcular o passo ou intervalo de seleção: $a = 160/40 = 4$;
- definir o ponto de partida na tabela de números aleatórios: linha 5, coluna 5, direção (\downarrow) para escolher um número aleatório de um algarismo entre 1 e 4, obtendo $r = 3$, uma vez que os três primeiros algarismos da tabela estão compreendidos fora do intervalo;
- selecionar os números de ordem:
3, $3 + 4$, $3 + 2 \times 4$, $3 + 3 \times 4$, $3 + 4 \times 4$, ou seja,
3, 7, 11, 15, 19, 23, 27, 31,;
- identificar os elementos da lista para compor a amostra.

A Figura 22.3 ilustra o esquema de amostragem sistemática descrito.

Amostragem Estratificada

Este tipo de amostragem requer um conhecimento prévio da natureza da população, pois ela é dividida em

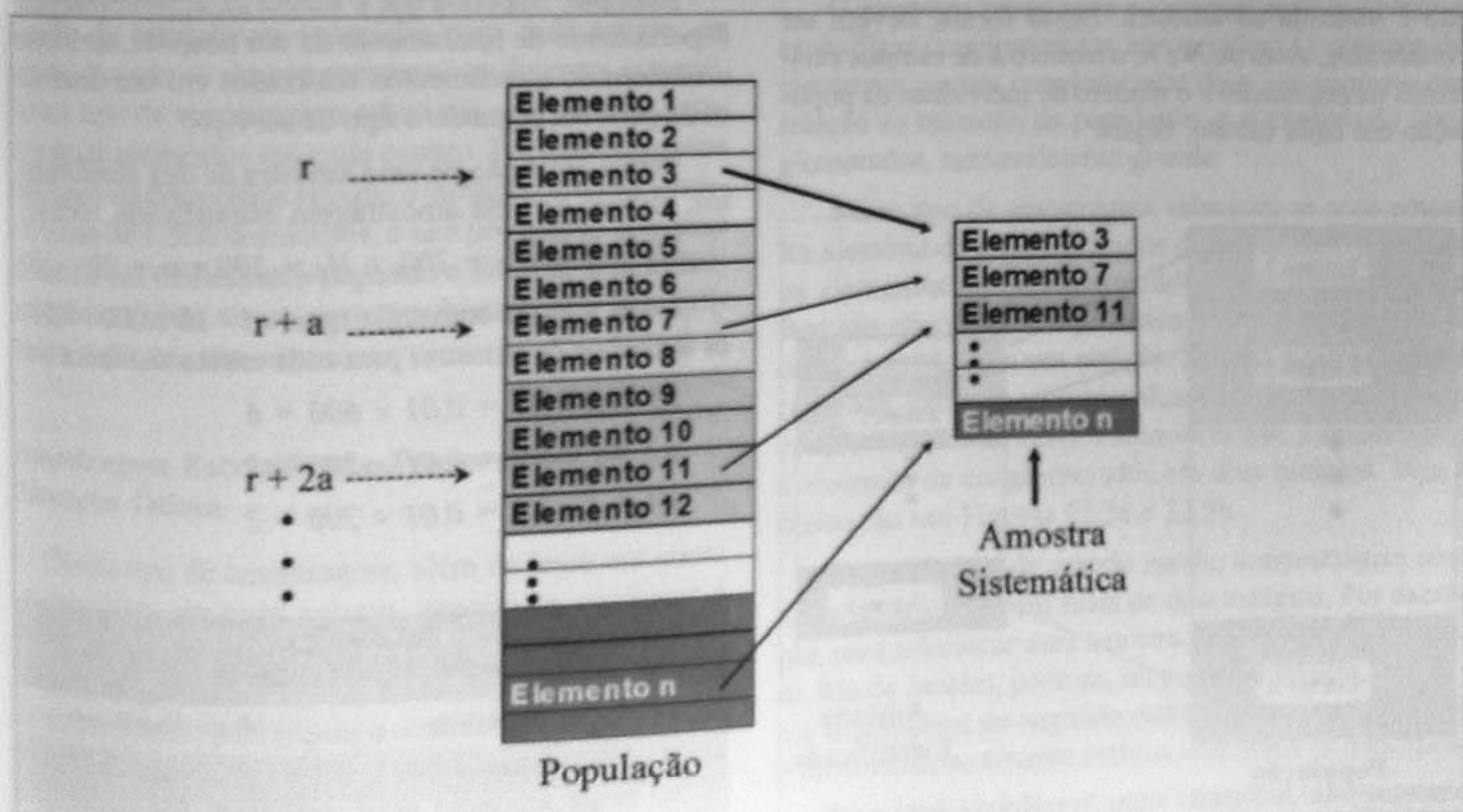


Fig. 22.3 - Esquema de uma amostragem sistemática.

subgrupos disjuntos e homogêneos, que são denominados de *estratos*. Deve-se escolher um critério de estratificação que forneça estratos internamente mais homogêneos, em relação à característica que está sendo estudada, comparativamente à população como um todo. Espera-se que a variabilidade nos estratos (intra) seja pequena, enquanto a variabilidade entre estratos seja maior. Por exemplo, numa pesquisa sobre usos de serviços hospitalares, a população poderia ser estratificada por níveis de renda, uma vez que se espera uma demanda diferenciada. As pessoas de um mesmo estrato (renda) tendem a ser mais "parecidas", enquanto se espera uma certa heterogeneidade entre os estratos de renda.

A amostragem estratificada também permite obter estimativas com certo nível de precisão para subgrupos da população em estudo, fixando tamanhos de amostra de forma separada para cada subgrupo.

Amostragens aleatórias devem ser realizadas de forma independente em cada estrato da população, e a amostra completa será obtida através da agregação das amostras de cada estrato. Na Figura 22.4 é apresentado um esquema de amostragem estratificada.

Amostragem Estratificada Proporcional

A amostragem estratificada proporcional é um caso particular da amostragem estratificada, no qual a proporcionalidade do tamanho de cada estrato da população é mantida na amostra. Dessa forma, devem ser conhecidos, além de N e n , o número h de estratos existentes na população e o número de indivíduos da população em cada estrato. Sejam:

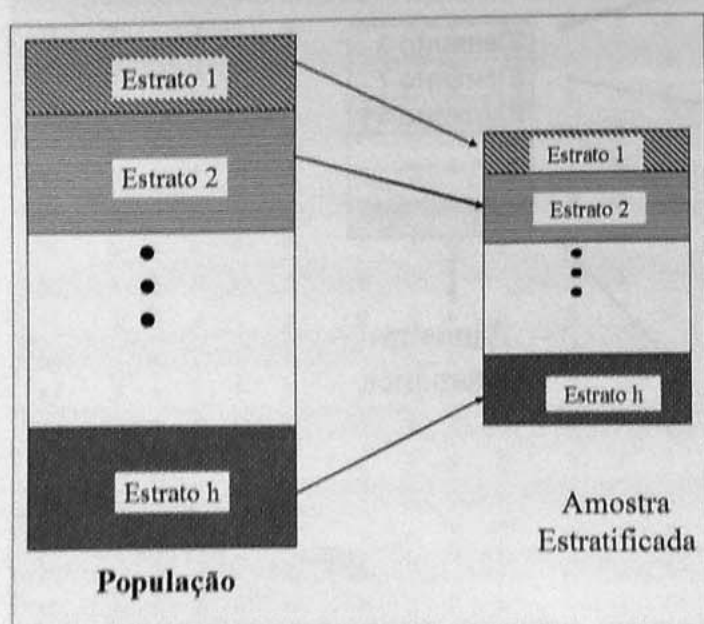


Fig. 22.4 - Amostragem estratificada.

N_1 = número de elementos no estrato 1
 N_2 = número de elementos no estrato 2
 N_h = número de elementos no estrato h .
 É óbvio que $N = N_1 + N_2 + \dots + N_h$.

Com o objetivo de que seja mantida na amostra a mesma distribuição dos estratos encontrada na população, define-se a fração amostral como $f = \frac{n}{N}$ e o número de elementos a observar em cada estrato é definido como:

$$n_1 = f \cdot N_1 = \frac{n \cdot N_1}{N}, \quad n_2 = f \cdot N_2 = \frac{n \cdot N_2}{N}, \dots,$$

$$e \quad n_h = f \cdot N_h = \frac{n \cdot N_h}{N}$$

A seguir, deve-se realizar amostragem aleatória simples em cada um dos estratos de onde serão escolhidos n_1 elementos de um total de N_1 , n_2 elementos de um total de N_2 , ..., n_h elementos de um total de N_h .

Na amostragem estratificada proporcional, cada elemento da população tem a mesma probabilidade de pertencer à amostra e, se nos estratos o comportamento da variável em estudo é mais homogêneo do que na população como um todo, a amostra estratificada proporcional também tende a gerar resultados mais precisos do que uma amostra aleatória simples.

Exemplo: Suponha que a Tabela 22.2, fornecida pelo departamento de faturamento de um hospital, apresenta o número de atendimentos realizados em um determinado período, segundo o tipo de serviço.

Para a obtenção de uma amostra de dez atendimentos, através de uma amostragem estratificada, tem-se:

$$N_1 = 600, N_2 = 200, e N_3 = 200 \text{ e } n = 10$$

Logo, a fração amostral é igual a $f = 10/1.000 = 0,01$ e os tamanhos de amostras para cada estrato são dados por:

$$n_1 = f \times N_1 = 0,01 \times 600 = 6$$

$$n_2 = f \times N_2 = 0,01 \times 200 = 2$$

$$n_3 = f \times N_3 = 0,01 \times 200 = 2$$

Tipo de Serviço	Nº de Atendimento
Ambulatorial	600
Emergência	200
Cirurgia	200

Dessa maneira, tem-se agora que realizar três amostragens aleatórias simples:

- escolher uma amostra de $n_1 = 6$ atendimentos de um total de $N_1 = 600$ atendimentos ambulatoriais;
- escolher uma amostra de $n_2 = 2$ atendimentos de um total de $N_2 = 200$ atendimentos de emergência;
- escolher uma amostra de $n_3 = 2$ atendimentos de um total de $N_3 = 200$ atendimentos cirúrgicos.

Pode-se realizar as três amostragens independentemente ou supor que os atendimentos estão numerados de forma seqüencial de 001 a 1.000, de acordo com o seguinte esquema:

1,2,3,4 ..., 600,	601,602, ..., 800,	801,802, ..., 1.000
Ambulatorial	Emergência	Cirúrgicas

Usando-se, por exemplo, uma tabela de números aleatórios, escolhem-se os elementos da amostra, identificando-os na categoria correspondente, até completar o número de elementos necessários em cada estrato. Neste caso, o número 1.000 poderia ser representado pelo 000.

Amostragem Estratificada Uniforme

A amostragem estratificada uniforme costuma ser usada nas situações em que o maior interesse é a obtenção de estimativas separadas para cada estrato, ou ainda, quando se deseja comparar os diversos estratos. Nesse tipo de amostragem seleciona-se a mesma quantidade de elementos em cada estrato. Por exemplo, uma pesquisa de domicílios em que o tamanho total da amostra fosse de 1.500 domicílios, e se a população estivesse dividida em três estratos segundo o nível de renda (baixa, média e alta), ter-se-iam que observar 500 domicílios em cada um dos estratos.

Amostragem Estratificada Proporcional de Alocação Ótima

Neste tipo de amostragem, além de levar em conta os tamanhos de cada estrato, precisa-se conhecer a variabilidade da característica a ser estudada em cada um dos estratos. Esse tipo de amostragem é mais difícil de ser realizado, pois requer o conhecimento prévio das variabilidades e dos custos. Porém, as estimativas dos parâmetros são melhores. Esse esquema aloca tamanhos de amostra maiores a estratos maiores e mais heterogêneos (com maior variabilidade) e aloca amostras

menores a estratos mais homogêneos ou nos quais seria mais caro e difícil realizar uma amostragem.

Amostragem por Conglomerados

Uma amostra por conglomerados é uma amostra aleatória simples na qual cada unidade de amostragem é um grupo ou conglomerado (*cluster*) de elementos. Por conglomerado entende-se um grupamento natural de elementos da população, os quais são bastante heterogêneos internamente em relação à característica estudada, porém de comportamento similar entre os conglomerados. Cada conglomerado deveria se comportar como uma população em miniatura. Por exemplo, numa pesquisa envolvendo a classe médica, os conselhos regionais serviriam como conglomerados. Outros exemplos de conglomerado seriam os Estados, Municípios, setores censitários.

A amostragem por conglomerado é útil quando a população é muito dispersa e é praticamente impossível ter uma lista de todos os elementos da população. Os resultados obtidos de uma amostragem por conglomerados tendem a ser menos precisos, diferentemente da amostragem estratificada, que tende a produzir resultados mais precisos e representativos, quando ambos esquemas são comparados com uma amostra aleatória simples do mesmo tamanho. Contudo, seu custo financeiro tende a ser bem menor.

O primeiro passo para se utilizar este processo é especificar conglomerados apropriados. O número de elementos em um conglomerado deve ser pequeno em relação ao tamanho da população, e o número de conglomerados, razoavelmente grande.

Nesse tipo de amostragem seleciona-se uma amostra aleatória de conglomerados de elementos. Se todos os elementos dos conglomerados selecionados nessa fase são observados, o processo é dito amostragem de conglomerados em um estágio. Se uma nova seleção é realizada, tomando amostras aleatórias de elementos nos conglomerados escolhidos anteriormente, a amostragem é chamada de conglomerados em dois estágios. Veja a ilustração nas Figuras 22.5a e 22.5b.

Em pesquisas de grande escala, a amostragem costuma ser realizada em mais de dois estágios. Por exemplo, para selecionar uma amostra de bancários do Estado de Rio de Janeiro, pode-se, no primeiro estágio, selecionar municípios; no segundo estágio, selecionar agências e finalmente, no terceiro estágio, selecionar os bancários.

Para uma abordagem mais completa, não somente em relação aos planos de amostragem como ao cálculo

das estimativas dos parâmetros, recomenda-se ler os livros já citados anteriormente: Cochram³, Kish⁷, entre outros.

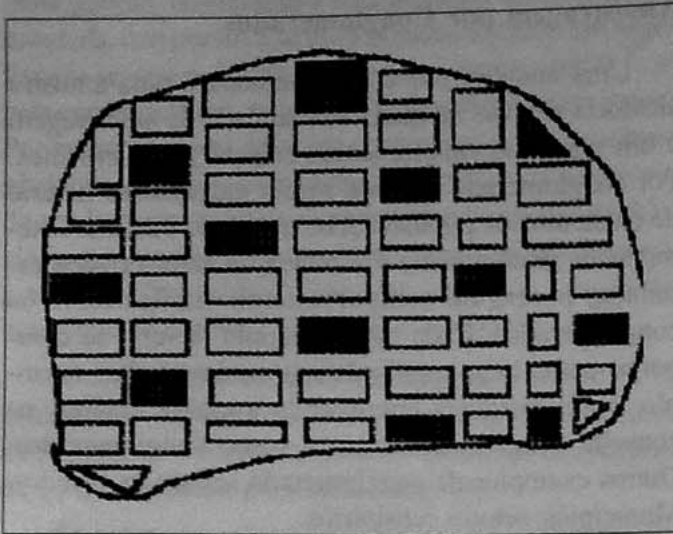


Fig. 22.5a – Amostragem por conglomerado em um estágio.

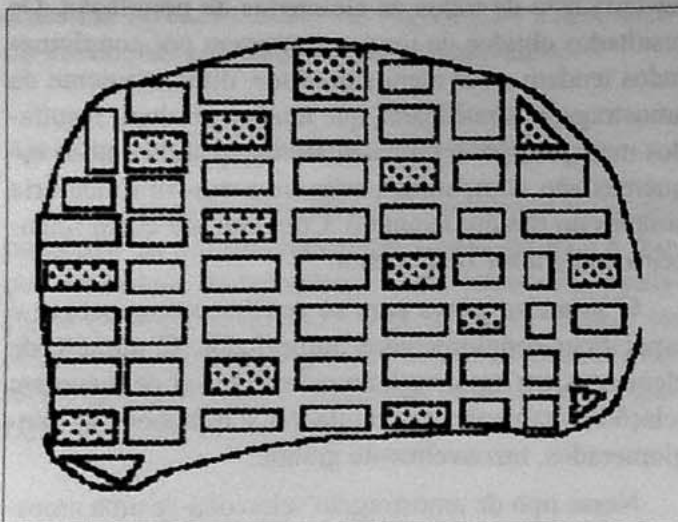


Fig. 22.5b – Amostragem por conglomerado em dois estágios.

Amostragem Complexa

Os esquemas amostrais probabilísticos apresentados até aqui são muito importantes não só para uma familiarização com a teoria da amostragem, como também por serem efetivamente utilizados na prática em muitos estudos clínicos e epidemiológicos de pequeno ou médio portes. Entretanto, os grandes planejamentos amostrais, como os de base nacional, são bem mais complexos, envolvendo, por exemplo, estratificação e conglomeração simultaneamente. Ou seja, para se conhecer uma certa característica populacional em um país como o Brasil, devido à diversidade das situações observadas, faz-se necessário um delineamento amostral que envolva probabilidades distintas de

seleção, estratificação e conglomeração das unidades, e ajustes para compensar não-resposta.

A Pesquisa Nacional por Amostragem Domiciliar (PNAD) é realizada anualmente pelo IBGE, investigando a situação familiar, de moradia, renda e inserção no mercado de trabalho (ocupação) da população. Sem a intenção de apresentar detalhes, pode-se dizer que o planejamento amostral dessa pesquisa é bastante complexo, pois inicialmente são identificados estratos (regiões ou municípios) para sorteio, sendo que algumas unidades entram no estudo com probabilidade 1 (estirão no estudo com certeza). Em um segundo momento identificam-se e sorteiam-se, dentro dos Municípios, conglomerados dados pelos setores censitários. Dentro desses são identificados e sorteados domicílios etc., de modo que cada uma das unidades da população (os indivíduos) apresente probabilidades de seleção bastante diferentes, além de se introduzir uma estrutura de dependência (correlação) entre as observações que tornam o processo de inferência uma tarefa para especialistas auxiliados por programas estatísticos específicos.

AMOSTRAGEM NÃO-PROBABILÍSTICA

Na apresentação das amostragens aleatória simples, sistemática e estratificada, observa-se que um dos requisitos é dispor de uma lista ou rol dos elementos da população. Uma forma de contornar esse problema seria realizar amostragens por conglomerados, necessitando-se apenas da lista dos conglomerados. Porém, na prática, apresentam-se situações em que nem esse procedimento pode ser realizado. Possíveis razões seriam: as listas dos conglomerados naturais não estarem disponíveis, ou devido à premência dos resultados, como no caso das pesquisas de mercado e de opinião pública que precisam de resultados num curto período de tempo, cuja execução seria cara e principalmente demorada. Nesses casos, pode-se usar procedimentos não-aleatórios, os quais procuram gerar amostras que representam razoavelmente bem a população de onde foram extraídas. Nelas qualquer análise estaria baseada na hipótese de que a amostra se comportaria de forma similar a uma amostra aleatória.

Os procedimentos de seleção de amostras não-probabilísticas freqüentemente usados são a amostragem por cotas, amostragem por julgamento e amostragem por conveniência.

Amostragem por Cotas

Este tipo de amostragem costuma ser usado em pesquisas de mercado e de opinião pública por ser

mais fácil de implementar que as amostragens probabilísticas, ser de menor custo e por produzir resultados em menor tempo. Nessa amostragem são formados diversos estratos, geralmente baseados em características da população como sexo, idade, classe social etc., e a escolha dos elementos a serem entrevistados em cada estrato é feita pelo entrevistador. A cota de elementos geralmente é proporcional ao tamanho do estrato. Com a finalidade de controlar ou diminuir o viés de seleção que poderia acontecer se os entrevistadores tivessem liberdade total de escolher as pessoas a serem entrevistadas, são atribuídas a eles cotas de diferentes tipos de pessoas. Por exemplo, a cota de um entrevistador poderia ser: dez homens brancos acima de 45 anos e quinze de 45 anos ou menos; 12 homens não-brancos acima de 45 anos e oito de 45 anos ou menos etc.

Uma aparente vantagem nesse tipo de amostragem é que sempre será possível atingir o tamanho de amostra pré-especificado, contornando os problemas de os elementos selecionados não serem encontrados e os de recusa para responder o questionário, pois sempre se pode entrevistar uma outra pessoa com as mesmas características. O problema nessa abordagem está na credibilidade dos resultados, já que não seria possível avaliar os vieses de seleção e de não-resposta e nem a precisão das estimativas.

Amostragem por Julgamento

Neste segundo tipo de amostragem não-probabilística, os elementos são escolhidos por um perito na matéria. Não é aplicado nenhum tipo de sorteio na seleção dos elementos, senão o critério pessoal do estudioso. Os elementos selecionados são aqueles julgados como típicos ou representativos da população que se deseja estudar. A desvantagem nesse tipo de pesquisa é que os resultados não necessariamente são válidos para toda a população, e o viés dos estimadores tende a crescer com o tamanho de amostra.

Amostragem de Conveniência

Caracteriza-se por não existir a preocupação com o desenho de um plano particular de amostragem. O objetivo não seria generalizar conclusões, pelo potencial viés de seleção, senão descrever as características principais do grupo de estudo. Por exemplo, os elementos da amostra, dependendo do objetivo do estudo, poderiam ser pessoas voluntárias em participar

de determinada experiência, pacientes de um médico ou que frequentam determinado hospital, alunos de uma sala de aula, respondentes a um questionário de revista, televisão ou qualquer outro meio de comunicação etc.

Erros nas Pesquisas por Amostragem

Quando se avalia a qualidade de uma pesquisa por amostragem, é importante diferenciar duas possíveis fontes de erro: o erro amostral e o erro sistemático (não-amostral).

O erro amostral é a diferença entre o valor de uma estatística (estimativa do parâmetro, calculada a partir das observações da amostra) e o verdadeiro valor do parâmetro (que seria calculado levando em conta todos os elementos da população). É possível estimar esse tipo de erro somente nos casos em que a amostragem é aleatória, e supondo que as n observações da amostra foram obtidas sem erros, o que geralmente não acontece na prática. Por essa razão, o planejamento e a execução da pesquisa devem ser conduzidos de modo a evitar ou reduzir ao máximo os erros nos dados da amostra, que são chamados de erros não-amostrais. Infelizmente, não existe um método simples e direto para estimar esse tipo de erro.

A seguir, apresentam-se algumas situações onde costumam aparecer erros não-amostrais numa pesquisa:

- Quando ao estudar uma certa população, por limitações de recursos ou falta de informação disponível sobre a população, seleciona-se uma amostra de uma lista incompleta de elementos. Nesse caso, recomenda-se limitar a abrangência da pesquisa à população que foi efetivamente estudada;
- Em pesquisas realizadas onde a população em estudo é a humana, é comum não se conseguir informação de alguns dos elementos selecionados, pois nem todos se dispõem a responder um questionário ou a serem entrevistados. Isso pode levar a sérias distorções nos resultados, principalmente se a conduta é substituir indivíduos que se recusam a responder, ou que não são encontrados no momento da pesquisa. Esse problema pode ser contornado retornando várias vezes a esses indivíduos. Pode acontecer também que o tema de pesquisa seja difícil de abordar e que os respondentes se sintam constrangidos diante do entrevistador. Uma possível solução seria usar questionários auto-explicativos respondidos pelos

próprios indivíduos, sem a presença do entrevistador e de forma anônima;

- Defeitos no aparelho de medição de algum ente físico ou uma pergunta mal formulada no questionário podem acarretar a impossibilidade de medir exatamente aquilo que se deseja. O uso de aparelhos bem calibrados, a elaboração de questionários que tenham itens de controle capazes de prevenir más respostas, assim como um bom treinamento dos entrevistadores, ajudam a reduzir esses possíveis erros.
- Podem também ser introduzidos erros nos dados no momento de codificação, na digitação, na elaboração dos programas de processamento da informação ou na escolha errada da medida de resumo. É recomendável dispor de programas de controle de qualidade da informação em cada uma dessas etapas.

BIBLIOGRAFIA

1. Barbetta PA. Estatística Aplicada às Ciências Sociais. 1994.
2. Barnett V. Elements of sampling Theory. The English universities Press Ltda. 1974.
3. Cochran W. Sampling e técnicas. Wiley N.Y. 1977.
4. Daniel. Biostatistics: a Foundation for Analysis in the Health Sciences
5. Fonseca JS, Martins GA. Curso de Estatística 3 ed. Editora Atlas S.A. 1987.
6. Freedman D, Pisani R, Purves R. Statistics WW Norton & Company. 1978.
7. Kish L. Survey sampling. Wiley NY. 1965.
8. Klein CH, Silva NAS, Nogueira AR, Bloch KV, Campos LHS. Hipertensão Arterial na Ilha do Governador, Rio de Janeiro, Brasil. I. Metodologia. Rio de Janeiro, Cadernos de Saúde Pública 1995; 11(2): 187-201.
9. Lemeshow S, Hosmer Jr DW, Klar J, Lwanga SK. Adequacy of Sample Size in Health Studies. John Wiley & Sons. 1990.
10. Milton JS. Statistical Methods in the Biological and Health Sciences. McGraw Hill International Editions – Biological Sciences Series 1992.
11. Soares JF, Farias AA, Cesar CC. Introdução à Estatística. Editora Guanabara Koogan S.A. 1991.