



A palavra *estatística* está associada à idéia de “coleção de números”. A estatística trata, sem dúvida alguma, da organização e da apresentação de contagens e de medições. Dentro deste conceito — de que estatística é coleção de números — cabe lembrar que no Brasil existem contagens e medições sobre assuntos tão diversos como índices de audiência em programas de televisão e valores de taxa de câmbio do dólar no mercado paralelo.

Mas a estatística não deve ser vista como simples “coleção de números”. As informações numéricas são obtidas com a finalidade de acumular informação para a tomada de decisão. Assim, por exemplo, é com base em índices de audiência que uma emissora de televisão pode tirar um programa do ar ou modificar o final previsto de uma novela; é com base em pesquisas de opinião que um candidato a cargo eletivo pode modificar certas atitudes ou mesmo o modo de se vestir; é com base na quantidade vendida de um produto em determinado período que um supermercado pode aumentar ou diminuir o estoque desse produto, e assim por diante.

Muitas decisões são tomadas com base em informações numéricas. No entanto — é preciso considerar —, os números não “falam por si mesmos”. Precisam ser organizados, discutidos e interpretados. Então, a estatística pode ser vista como um conjunto de técnicas que ajudam na tomada de decisão, quando prevalecem condições de incerteza. Por esta razão, todo profissional — seja ele administrador, executivo, cientista ou professor — deve adquirir algum conhecimento de estatística para tomar uma decisão consciente quando tiver, à sua disposição, informações numéricas, ou seja, dados estatísticos. Mas o que são os dados estatísticos?

1.1 DADOS ESTATÍSTICOS

Toda informação devidamente coletada e registrada, quer seja na forma de contagem ou de medição, é um dado estatístico. Todo dado se refere a uma variável. Não existe interesse em coletar mais de um valor quando se trata de uma constante. Assim, não existe interesse em coletar informações sobre analfabetismo entre secundaristas, porque todos os estudantes do secundário são alfabetizados.

Existem três tipos de variável:

- nominal ou qualitativa;
- ordinal;
- cardinal ou quantitativa.

Tem-se uma *variável nominal* ou *variável qualitativa* quando os indivíduos ou elementos são classificados em determinado número de categorias mutuamente exclusivas. Por exemplo, alfabetização é uma variável nominal porque determinada pessoa só pode pertencer a uma de duas categorias mutuamente exclusivas: ou ela é alfabetizada ou é analfabeta. Outro exemplo de variável nominal é a classificação dos brasileiros de acordo com o Estado onde nasceram: um indivíduo é ou baiano, ou paulista, ou cearense etc.

Denomina-se *variável ordinal* aquela que pode ser distribuída em determinado número de categorias mutuamente exclusivas, mas tais categorias se apresentam segundo uma ordem lógica. Então, grau de instrução é uma variável ordinal porque uma pessoa pode ser classificada em uma das seguintes categorias: 1. sem instrução; 2. com primário incompleto; 3. com primário completo; 4. com secundário incompleto; 5. com secundário completo; 6. com superior incompleto; 7. com superior completo.

As categorias de uma variável ordinal podem ser indicadas por números, de acordo com a ordem. É preciso deixar claro, no entanto, que neste caso apenas a ordem do número tem interesse. Não podem ser feitas operações aritméticas com estes números. Assim, no exemplo dado, duas vezes primário completo não é igual a curso superior incompleto.

Denomina-se *variável cardinal* ou *variável quantitativa* aquela que é expressa através de um número, com todas as suas propriedades. Por exemplo, a renda de uma pessoa é uma variável cardinal porque é expressa numericamente, em cruzados e, nesse caso, não há dúvida de que uma renda de 12 mil é três vezes maior do que uma renda de 4 mil.

1.2 APURAÇÃO DE DADOS

Os dados são registrados em fichas. Essas fichas normalmente contêm, além dos dados de interesse, diversas outras informações, como nome

e endereço de quem forneceu os dados, nome do entrevistador etc. Então, terminada a fase de coleta de dados, é preciso proceder à *apuração*, isto é, reunir os dados.

Quando os dados se referem a uma variável nominal, ou a uma variável ordinal, a apuração se reduz a simples contagem, isto é, a apuração consiste em estabelecer o número de elementos que caem em cada categoria. Para entender o procedimento, imagine que um professor quer saber a proporção de alunos de cada sexo que se candidataram a determinado exame. O professor deve então tomar todas as fichas de inscrição e escrever, numa folha de papel:

Masculino

Feminino

Depois o professor deve examinar cada ficha de inscrição e fazer um traço, na linha correspondente a um dos sexos, toda vez que a ficha de inscrição indicar que o aluno é desse sexo. No exemplo dado em seguida, cada traço representa um aluno e cada quadrado, cortado pela diagonal, representa cinco alunos.

Masculino $\square \square \square \square \square$ = 22

Feminino $\square \square \square \square$ = 18

Também pode ser adotada outra forma de notação. No exemplo dado em seguida, cada traço representa um aluno e cada conjunto de traços, cortados em horizontal, representa cinco alunos.

Masculino $\equiv \equiv \equiv \equiv \equiv$ = 22

Feminino $\equiv \equiv \equiv \equiv$ = 18

Se os dados se referem a uma variável cardinal, a apuração exige que se anote cada valor observado. Para entender o procedimento, imagine que um professor de educação física quer apurar os dados de estatura dos alunos de uma classe. Para isso o professor precisa tomar todas as fichas biométricas e anotar, em uma folha de papel, o nome (ou o número) de cada aluno e a respectiva estatura, como mostra o exemplo dado em seguida.

Nome	Estatura
Antônio	1,68 m
Bráulio	1,72 m
.....	
.....	
Tadeu	1,59 m

1.3 POPULAÇÃO E AMOSTRA

Na linguagem comum, população significa o conjunto de habitantes de um território, de um país, de uma região, de uma cidade etc. Em Estatística, no entanto, o termo *população* tem significado bem mais geral. Entende-se por *população* o conjunto de elementos que têm, em comum, determinada característica.

As populações podem ser finitas como, por exemplo, os alunos matriculados em determinada escola, ou infinitas como, por exemplo, os resultados obtidos quando se joga um dado sucessivamente. Existem populações que, embora finitas, são consideradas infinitas para qualquer finalidade prática. Como exemplo, imagine o número de peixes que há no mar. Esse número é, em dado instante, matematicamente finito, mas tão grande que pode ser considerado infinito para qualquer finalidade prática.

Entende-se por *amostra* qualquer conjunto de elementos, retirado da população, desde que esse conjunto seja não-vazio e tenha menor número de elementos do que a população. Imagine uma população constituída por quatro elementos; isto é, por:

Ana, Pedro, Luís, Clara.

Dessa população podem ser retiradas:

a) quatro amostras de três elementos cada uma:

Ana; Pedro; Luís	Ana; Luís; Clara
Ana; Pedro; Clara	Pedro; Luís; Clara

b) seis amostras de dois elementos cada uma:

Ana; Pedro	Ana; Clara	Pedro; Clara
Ana; Luís	Pedro; Luís	Luís; Clara

c) quatro amostras de um elemento cada uma:

Ana	Luís
Pedro	Clara

Quando são coletados dados sobre todos os elementos da população, diz-se que foi feito um recenseamento. No Brasil, os recenseamentos gerais são feitos por um órgão oficial, a Fundação Instituto Brasileiro de Geografia e Estatística — IBGE —, que tem sede na cidade do Rio de Janeiro. Denomina-se *censo* o conjunto de dados obtidos através de recenseamento. Existem censos demográficos (que descrevem as caracterís-

ticas da população), censos agropecuários, censos industriais, censos comerciais e de serviços etc. No Brasil foram feitos nove recenseamentos gerais, isto é, em 1872, 1890, 1900, 1920, 1940, 1950, 1960, 1970 e 1980.

O Recenseamento Geral de 1980, a exemplo dos anteriores, atendeu às determinações da ONU (Organização das Nações Unidas) para se integrar ao Censo das Américas. Essas determinações visam garantir a uniformidade de conceitos utilizados nos censos das nações americanas.

É preciso deixar claro, porém, que nem sempre é possível coletar dados de toda a população. Um censo pode revelar-se impossível por razões de economia. Assim, por exemplo, para determinar o índice de audiência de um programa de rádio, que atinge todo o Brasil, seria absurdo realizar um censo, devido ao custo muito elevado. Então, os índices de audiência de programas de rádio e de televisão são calculados com base em amostras.

Também não se usam censos nos casos de estudos demorados, que exigem atenção cuidadosa do pesquisador. Considere, como exemplo, um estudo que tenha a finalidade de obter uma descrição pormenorizada de como uma família gasta sua renda, isto é, um estudo sobre orçamento familiar. Nesse tipo de estudo é preciso coletar grande número de dados de cada família. É impraticável coletar dados de todas as famílias brasileiras ou, até mesmo, de todas as famílias de uma cidade. Então, os estudos de orçamento familiar são, necessariamente, feitos com base em amostras.

Estes exemplos servem para lembrar que boa parte do conhecimento se baseia no estudo de amostras. No entanto, a validade de uma amostra depende da técnica utilizada para sua obtenção.

1.4 TÉCNICAS DE AMOSTRAGEM

Uma amostra só é representativa da população quando é constituída por elementos selecionados de acordo com uma técnica conhecida. Existem diversas técnicas de amostragem, mas as mais conhecidas são:

- amostragem casual simples;
- amostragem sistemática;
- amostragem estratificada.

1.4.1 Amostragem casual simples

No processo de *amostragem casual simples*, todos os elementos da população têm igual probabilidade de serem selecionados para constituir

a amostra. Como exemplo, imagine que o gerente de uma indústria pretenda obter uma amostra dos empregados. Para isso, o gerente pode conferir um número a cada empregado e sortear aqueles que irão constituir a amostra.

Para fazer o sorteio, o gerente pode usar uma urna que contenha fichas numeradas, de zero a nove, inclusive. Então o gerente retira uma ficha da urna, ao acaso, e anota o número, que dará o primeiro dígito do número do empregado que será selecionado. Depois o gerente recoloca a ficha na urna, mistura e retira outra ficha, que dará o segundo dígito do número do empregado que será selecionado. O gerente deve repetir esse processo até que sejam obtidos todos os dígitos da numeração utilizada.

Se a indústria tiver, por exemplo, 900 empregados, devem ser sorteados três dígitos para definir cada empregado que será selecionado para constituir a amostra. Então, no primeiro sorteio pode ocorrer, por exemplo:

Primeira ficha retirada:	4
Segunda ficha retirada:	7
Terceira ficha retirada:	4

O resultado do primeiro sorteio indica que foi selecionado, para a amostra, o empregado de número 474. O gerente deve repetir esse processo até completar a amostra. Evidentemente, serão desprezados números iguais a zero (0,0,0), números já sorteados (já estão na amostra) e números maiores do que 900 (não existem nessa população).

1.4.2 Amostragem sistemática

No processo de *amostragem sistemática*, os elementos são selecionados para a amostra por um sistema preestabelecido. Assim, por exemplo, imagine que um gerente de vendas quer obter uma amostra dos clientes cadastrados de sua firma. Se as fichas dos clientes estiverem organizadas por ordem alfabética, o gerente obterá uma amostra sistemática se escolher a quinta de cada cinco fichas, ou a décima, de cada dez etc.

O processo de obtenção de uma amostra sistemática é bem mais simples do que o processo de obtenção de uma amostra casual simples nos casos em que a população está organizada. Assim, para obter uma amostra dos assinantes de um jornal é mais fácil localizar os nomes dos assinantes no arquivo e, de cada 20, tirar, por exemplo, o vigésimo do que dar um número a cada assinante e fazer sorteio até completar uma amostra que corresponderia a $1/20$ da população.

Nas pesquisas domiciliares também se usa amostragem sistemática. Nesses casos é comum estabelecer um ponto de partida ao acaso em cada bairro e, a partir daí, percorrer as ruas usando determinado sistema como, por exemplo, de cada dez domicílios selecionar, para a amostra, o décimo.

1.4.3 Amostragem estratificada

Usa-se o processo de *amostragem estratificada* quando a população se apresenta dividida em estratos, isto é, quando a população está dividida em grupos distintos. Como exemplo, imagine que o diretor de uma faculdade quer obter uma amostra da comunidade acadêmica. O diretor precisa considerar que a faculdade é constituída por três grupos distintos de pessoas, isto é, professores, funcionários e alunos.

Então, para obter uma amostra mais representativa da comunidade acadêmica, o diretor deve selecionar uma amostra dentro de cada estrato, isto é, uma amostra dos professores, uma amostra dos funcionários e uma amostra dos alunos e depois reunir essas três amostras em uma só, constituindo então uma amostra estratificada.

1.5 TENDENCIOSIDADE DA AMOSTRA

Muitas vezes as amostras são constituídas por pessoas que devem executar algum tipo de tarefa, como responder perguntas, preencher um questionário ou até mesmo testar um produto. Algumas pessoas se recusam a cooperar. Nesses casos é preciso ter muito senso crítico para avaliar se a amostra efetivamente utilizada é representativa da população. Sempre é possível que a amostra obtida seja *tendenciosa* ou *viciada*, isto é, não representativa da população.

O senso crítico ainda é mais importante quando as amostras são constituídas por voluntários. Muitas vezes, o procedimento usado para solicitar voluntários conduz à formação de amostras tendenciosas. Como exemplo, imagine que um professor de educação física peça que três alunos da turma se apresentem como voluntários para apostar uma corrida. Ora, é bastante razoável imaginar que, neste caso, se apresentarão como voluntários apenas os alunos que sabem ser bons corredores. Então, os três alunos não constituirão uma amostra representativa da turma, mas uma amostra tendenciosa ou viciada.

1.6 EXERCÍCIOS

1. Classifique as seguintes variáveis como nominal, ordinal ou cardinal: idade, sexo, renda familiar, religião, cor, tempo disponível para estudo, atividades recreativas de interesse.
2. Dê dois exemplos de variável nominal, dois exemplos de variável ordinal e dois exemplos de variável cardinal.

3. Dada uma população com três elementos, A, B e C, identifique todas as amostras possíveis que podem ser obtidas dessa população.
4. Dada uma população com seis elementos, A, B, C, D, E e F, explique como você faria para obter, dessa população, uma amostra casual simples com três elementos.
5. Descreva uma forma de obter uma amostra sistemática com dez elementos de uma população com 100 elementos.
6. Explique a forma de obter uma amostra estratificada dos empregados de uma firma, considerando que existem empregados de escritório, empregados de oficina e representantes da firma.
7. Imagine que se pretenda fazer um levantamento de opinião pública para verificar se as pessoas são contra ou a favor do uso gratuito de ônibus pelos idosos. Pense em duas maneiras distintas de fazer a pergunta — uma que induza a resposta positiva e outra que induza a resposta negativa.
8. Para verificar que as pessoas tendem a dar algum tipo de resposta, mesmo quando desconhecem o assunto, pergunte para diversas pessoas se elas conhecem uma revista, ou um jornal, ou um ator que na realidade não existe. Pergunte, por exemplo: “Você já ouviu as músicas de Linda Junk?” Discuta a proporção de respostas positivas.
9. Muitas vezes as pessoas concordam ou discordam de uma frase em função não da frase em si, mas do autor da frase. Para verificar essa hipótese pergunte, para diferentes pessoas escolhidas ao acaso: Você concorda com Ruy Barbosa quando ele disse: “É preciso lutar por nossas idéias, até as últimas consequências.” Depois pergunte para outro grupo de pessoas: Você concorda com Hitler quando ele disse: “É preciso lutar por nossas idéias, até as últimas consequências.” Compare a proporção de respostas positivas.
10. Se uma moeda for jogada 20 vezes, os resultados constituirão uma amostra de tamanho 20 da população infinita de resultados que podem ocorrer quando se joga uma moeda. Obtenha três amostras desse tipo, isto é, jogue uma moeda 20 vezes e depois repita a experiência mais duas vezes. Compare a proporção de caras obtidas nas três amostras.