Estatística

O termo estatística foi empregado, de início, para designar o conjunto de dados referentes a assuntos de interesse do estado, geralmente com finalidade de controle fiscal ou de segurança nacional. Esses dados referiam-se, particularmente, à população, às transações comerciais internas ou com outros países, ao controle da mortalidade e aos problemas de taxação e de proporcionalidade de tarifas e impostos.

Técnica auxiliar do estudo dos fenômenos, coletivos, econômicos, sociais ou científicos, a estatística é um método de observação, descrição, mensuração e interpretação dos fenômenos coletivamente típicos e da indagação de suas uniformidades e relações. O campo de aplicação da estatística é o dos fenômenos que apresentam regularidade na observação de massas de casos, embora uma parte de seus processos também encontre aplicação no domínio dos fenômenos atípicos. A teoria estatística permite que sejam tomadas decisões com base em informações geralmente limitadas e incompletas sobre os mais variados fenômenos que ocorrem no mundo. É a ciência da indução lógica, isto é, das generalizações de características de um conjunto, de cujos elementos se conhece um subconjunto.

Os métodos estatísticos conduzem a conclusões sobre causa e efeito e permitem testar teorias relativas ao consumidor. O economista os usa para escolher, dentre as possíveis formas de uma função teórica de consumo, a que melhor explica os dados observados. O médico emprega a técnica estatística nos resultados de testes de avaliação de um novo medicamento e o agricultor para decidir qual das fórmulas de fertilizantes é preferível. O resultado de uma eleição ou o julgamento da qualidade de um produto industrial podem ser determinados estatisticamente.

Histórico. Os estudiosos da disciplina distinguem três grandes etapas na história da estatística:

1. O período mais antigo é caracterizado pela simples organização de informações de interesse do estado, do qual é típico o famoso Domesday Book, recenseamento de propriedades ordenado por Guilherme o Conquistador (1086). Esse enfoque perdurou até meados do século XVII.

2. O período que medeia entre o século XVII e o princípio do século XIX, caracterizado pelas inúmeras tentativas de analisar as tabelas e os conjuntos de dados com a finalidade de obter conclusões úteis à organização do estado ou às previsões do futuro. No século XVIII, a Universidade de Jena promoveu, pela primeira vez, um curso avançado de estatística. Em Die göttliche Ordnung in den Veränderugen des menschlichen Geschlechts, aus der Geburt, dem Tode un der Fortpflanzung desselben erwiesen (1741; A ordem divina das variações do ser humano, conforme os números dos nascimentos, da procriação e das mortes), o demógrafo alemão Johann Peter Süssmilch procurou dar estrutura científica aos problemas estatísticos. Pouco depois o economista Gottfried Achenwall, da Universidade de Göttingen, publicou Staatsverfassung der heutingen vornehmsten europäischen Reichen (1752; Constituições dos principais países da Europa), em que procurou apresentar os limites e os métodos da pesquisa estatística e definir os objetos material e formal da estatística, No século XIX, Adolphe Quételet realizou cuidadoso estudo estatístico dos fatos demográficos e sociais, com um tratamento predominantemente matemático.

3. O terceiro período, iniciado com o Congresso Internacional de Estatística (1853), estende-se até o presente. Caracteriza-se não só pelos avanços metodológicos da estatística em si mas, sobretudo, pelas múltiplas aplicações que ela vem tendo, sobretudo na área da pesquisa científica.

A estatística moderna é uma técnica especial que trata de problemas como: (1) planejamento de um programa de obtenção de dados, de maneira que as conclusões extraídas mereçam confiança; (2) métodos de análise dos dados; (3) avaliação das conclusões e estabelecimento do grau de confiança inspirada; (4) análise e crítica da validade dessas conclusões. Na teoria estatística, há uma série de conceitos e processos que, por possuírem ampla generalidade, merecem estudo especial.

Tabelas. A maneira mais simples de organizar uma coleção de dados numéricos consiste em dispô-los em forma de tabela. As tabelas estatísticas possuem, geralmente, três elementos essenciais: o tempo, ou seja, a época a que se refere; o espaço, isto é, o local ou região onde os dados foram anotados; e a natureza dos fatos colhidos. A variação desses elementos primordiais determina o tipo da tabela, podendo distinguir-se, principalmente, as tabelas de tipo cronológico, geográfico, de categorias, de freqüências e outras.

Gráficos. Em geral é vantajoso representar em forma de diagrama o conjunto de informações que constituem as tabelas. A representação gráfica oferece um meio rápido de comparação de duas ou mais séries de números e tem ampla aplicação.

No gráfico de barras, as grandezas são representadas por retângulos de mesma largura e de comprimentos proporcionais aos valores numéricos dessas grandezas, podendo ser dispostos horizontal ou verticalmente. Variedades desses gráficos substituem as barras retangulares por figuras mais sugestivas, como moedas empilhadas, silhuetas humanas etc.

O gráfico de curvas, particularmente indicado quando uma das grandezas é o tempo, consta de uma série de pontos horizontalmente separados pela mesma distância, mas que se situam a diferentes alturas, proporcionais aos valores numéricos da grandeza.

O gráfico em setores, especialmente empregado quando se deseja salientar a importância de um dado em relação ao todo, consta de um círculo ou retângulo dividido em setores, cujas áreas são proporcionais aos valores numéricos.

O gráfico polar, particularmente útil na representação de fenômenos periódicos ou de distribuições cíclicas, é construído a partir de um círculo que se divide em partes proporcionais aos períodos, tomando-se, sobre cada raio, comprimentos correspondentes aos valores numéricos e ligando-se os extremos desses segmentos numa linha poligonal fechada. Para a representação do fenômeno e sua grandeza em regiões diversas emprega-se o cartograma.

Distribuição de freqüências. Quando se tem um conjunto de dados referentes a um fato qualquer, pode ocorrer a repetição de alguns desses dados, que, por isso, figuram muitas vezes no conjunto. De modo geral, é importante destacar o número de vezes que um dado aparece, sendo esse número denominado freqüência. Muitas vezes reúnem-se vários dados, formando-se com eles uma classe, e organizam-se tabelas compostas pelas diversas classes em que o conjunto foi dividido, acompanhado da respectiva freqüência. Nesse caso, a soma das freqüências de todas as classes deve ser igual ao número total de dados.

Na escolha dos intervalos de classe é conveniente adotar certas regras para que ela seja adequada e permita confrontos imediatos, facilitando os trabalhos estatísticos. Em primeiro lugar, salvo condições excepcionais, os intervalos devem ser iguais. Quando não for possível, devem ser múltiplos, preferivelmente duplos, quíntuplos ou décuplos de um deles. Quando se puder prever que os dados se acumulam em determinado valor, esse valor deve ser escolhido para ponto médio de uma classe. Sempre que possível, o limite inferior da primeira classe deve ser o mais baixo valor encontrado. Uma vez obtida a distribuição de freqüências, constrói-se o histograma, que consta de uma sucessão de retângulos colocados lado a lado, cuja base é proporcional ao intervalo de classe e cuja altura é proporcional à freqüência.

O contorno do histograma é denominado "poligonal característica". Se forem marcados os pontos médios de cada intervalo, tomando pontos cuja altura é proporcional à freqüência de classe, ter-se-á, ligando os pontos obtidos, o "polígono de freqüência".

Média, moda e mediana. Quando se possui grande número de dados referentes a um mesmo fato, é necessário determinar um valor que represente o conjunto. Isso se consegue através de cálculos das médias, que são de diversos tipos. A mais usada, por ser de mais fácil compreensão e mais simples de calcular, é a média aritmética, que se obtém somando todos os dados e dividindo o total pelo número de dados do conjunto. Sempre que for possível atribuir diferentes importâncias aos dados, é preferível calcular a média aritmética ponderada. Para tanto, atribui-se o valor ou peso 1 ao dado de menor importância e, aos demais, pesos proporcionais às importâncias merecidas. A média é calculada multiplicando-se cada valor pelo respectivo peso e dividindo-se o total pela soma dos pesos.

Outra média muito usada é a média geométrica ou proporcional, que se calcula multiplicando os números e extraindo do produto a raiz cujo índice é o total de números. No caso de apenas dois números, sua média geométrica é a raiz quadrada de seu produto. No caso de vários fatores, faz-se necessário empregar logaritmos para calcular a média.

A média harmônica entre dois números é calculada dividindo-se o dobro do produto entre esses números por sua soma, ou, o que é equivalente, dividindo-se 2 pela soma dos inversos desses números. Analogamente, a média harmônica entre n números é obtida dividindo-se n pela soma dos inversos desses números. Em alguns casos, calcula-se a média harmônica ponderada.

Quando se empregam instrumentos de medida, especialmente os de precisão, o resultado das medidas está sempre sujeito a um erro imprevisível. Para atenuar os efeitos desse erro é costume repetir a medida muitas vezes e calcular a média dos valores obtidos. O trato estatístico é bastante conveniente em casos como esse. Quando o sentido do sinal de erro não interessa, pode-se calcular a média quadrática, que é igual à raiz quadrada da média aritmética dos quadrados dos números.

Em muitos casos, o valor da média não é muito significativo para representar o conjunto, sendo, então, substituído por outros valores, entre os quais os mais usados são a moda e a mediana.

Moda, norma ou dominante de um conjunto é o dado que possui maior freqüência. Várias expressões permitem calcular seu valor aproximado, quando a distribuição de freqüências está organizada por classes. Entre essas fórmulas as mais aplicadas são as de Czuber e a de King, respectivamente:

Moda =

onde P é o ponto inicial da classe de maior freqüência, i é o intervalo de classe, N é a maior freqüência, N' a freqüência imediatamente anterior a N e N" a freqüência imediatamente posterior a N.

Moda =

Mediana de um conjunto de dados é o dado que fica situado bem no centro do conjunto, isto é, aquele que tem tantos elementos acima dele como abaixo dele. Quando o conjunto tem número par de elementos haverá dois elementos medianos. A mediana é calculada pela expressão:

Mediana =

onde P é o ponto inicial, i é o intervalo, s é o semitotal de elementos, Na a freqüência acumulada, N a freqüência simples.

Erros e confiabilidade. As pesquisas estatísticas de caráter científico e social devem ser acompanhadas de um conjunto de indicações técnicas que garantem sua confiabilidade.

Em toda análise de medidas aparecem, geralmente, dois tipos de erros de observação: (1) Erros sistemáticos, produzidos por defeitos dos instrumentos de medida ou em virtude do método de observação do investigador; (2) Erros acidentais, advindos de perturbações imprevisíveis, mais difíceis de valorar devido a sua variabilidade.

Estatística metodológica. A aplicação de um método estatístico em um estudo social ou científico efetua-se quando a variabilidade, complexidade ou parcial desconhecimento das origens do fenômeno impedem completar experiências práticas perfeitamente concebidas e de resultados confiáveis. O resultado obtido pelas técnicas estatísticas permite alcançar uma série de comportamentos gerais, frente a casos isolados ou acidentais, com os quais se fabrica um modelo.

Serviços de estatística no Brasil. Os primeiros levantamentos sistemáticos de dados no Brasil datam de 1808, com a realização do censo da população. Nos anos que se seguiram, poucos progressos foram alcançados, mas em 1871 foi instituída a Diretoria Geral de Estatística, cuja primeira tarefa relevante foi o recenseamento geral efetivado em 1872. Os dois recenseamentos seguintes, o de 1890 e o de 1900, não apresentaram melhorias significativas, mas o de 1920 obedeceu a um plano mais minucioso e complexo.

O passo posterior de maior importância foi dado com a criação, em 1934, do Instituto Nacional de Estatística, transformado depois no Instituto Brasileiro de Geografia e Estatística, dividido em duas alas: o Conselho Nacional de Geografia e o Conselho Nacional de Estatística.

Atualmente a Fundação Instituto Brasileiro de Geografia e Estatística (IBGE), criada em 1973, tem a função de coordenar, na qualidade de órgão central, as atividades do sistema estatístico nacional, bem como as de natureza geográfica e cartográfica.

©Encyclopaedia Britannica do Brasil Publicações Ltda.