



1. O que é Estatística?

“ A essência da Ciência é a observação”

Toda pesquisa ou trabalho científico, nas mais variadas áreas, como sociologia, saúde, psicologia, etc., de um modo bem geral, em alguma fase de seu desenvolvimento, se depara com situações que envolvem uma grande quantidade de dados relevantes ao objeto de estudo. Esses dados têm que ser trabalhados e transformados em informações, para que possam ser comparados com outros resultados, ou ainda para julgar sua adequação a alguma teoria. Para isto se recorre a técnicas desenvolvidas com a finalidade de auxiliar a análise dessas informações.

A utilização dessas técnicas, destinadas à análise de situações complexas ou não, tem aumentado e faz parte do nosso cotidiano. Jornais, revistas técnicas artigos, etc., publicam freqüentemente tabelas, gráficos, porcentagens e outros dispositivos destinados a complementar a apresentação de um fato ou justificar um argumento.

A ciência que se dedica a esse trabalho é a **Estatística**.

Estatística: é o conjunto de técnicas que permite, de forma sistemática, coletar, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento.

No passado, tratar uma grande massa de números era tarefa custosa e cansativa, que exigia horas de trabalho. Recentemente, no entanto, grande quantidade de informações pode ser analisada rapidamente com um computador pessoal e programas adequados. Desta forma, o computador contribui, positivamente, na difusão e uso de métodos estatísticos. Por outro lado, o computador possibilita uma automação que pode levar um indivíduo sem preparo específico a utilizar técnicas inadequadas para resolver um dado problema. Assim, é necessário a compreensão dos conceitos básicos da Estatística, bem como as suposições necessárias para o seu uso de forma criteriosa.

Quando se aborda uma problemática envolvendo métodos estatísticos, deve-se planejar a experiência que nos vai permitir recolher os dados, de modo que, posteriormente, se possa extrair o máximo de informações relevantes para o problema em estudo, ou seja, para a população de onde os dados provêm. Quando de posse dos dados, procura-se agrupá-los e reduzi-los sob forma de amostra.

Seguidamente o objetivo do estudo estatístico pode ser o de estimar uma quantidade ou testar uma hipótese. Utilizamos então técnicas estatísticas convenientes que vão permitir tirar conclusões acerca da população, baseando-se numa pequena amostra, dando-nos ainda uma medida do erro cometido.

2. Estatística Descritiva e Estatística Indutiva

Numa análise estatística distinguem-se essencialmente duas fases:

1ª Fase - Estatística Descritiva - Procura-se descrever e resumir dados, afim de que se possam tirar conclusões a respeito das características de interesse.

Exemplos de características de interesse: idade, sexo, peso.

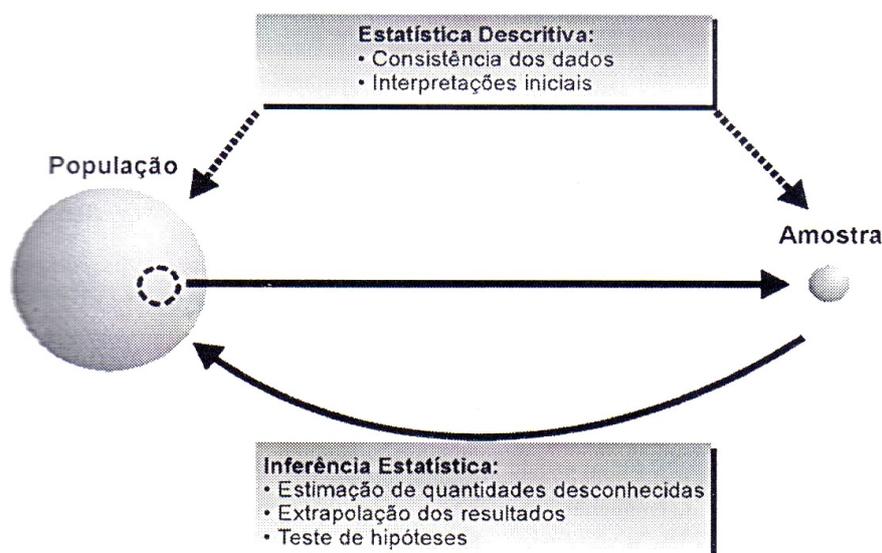
Exemplos de técnicas descritivas: gráficos, tabelas, de freqüência, parâmetros associados às freqüências.

2ª Fase - Estatística Indutiva (Inferência) - Conhecidas certas propriedades (obtidas a partir de uma análise descritiva de uma amostra), expressas por meio de proposições, imaginam-se proposições mais gerais (extrapolação), que exprimam conclusões para toda a população.

No entanto, ao contrário das proposições deduzidas, não podemos dizer se são falsas ou verdadeiras, já que foram verificadas sobre um conjunto restrito de indivíduos, e, portanto não são falsas, mas não foram verificadas para todos os indivíduos da População, pelo que também não podemos afirmar que são verdadeiras. Existe, assim, certo grau de incerteza.

Para que a inferência seja válida, é necessária uma boa amostragem, livre de erros, tais como falta de determinação correta da população, falta de aleatoriedade e erro no dimensionamento da amostra.

Para se estudar essa incerteza, oriunda das proposições mais gerais, recorreremos à teoria matemática das **Probabilidades**.



3. Planejamento de Experimentos

Os estudos que utilizam métodos estatísticos vão desde os que são concebidos e executados, dando resultados confiáveis, aos que são concebidos deficientemente e mal executados, levando a conclusões enganosas e sem qualquer valor real. Eis alguns pontos importantes para o planejamento de um estudo capaz de produzir resultados válidos:

1. Identificar com precisão a questão a ser respondida e definir com clareza a população de interesse.
2. Estabelecer um plano para coleta de dados. Esse plano deve descrever detalhadamente a realização de um estudo observacional ou de experimento e deve ser elaborado cuidadosamente, de modo que os dados coletados representem efetivamente a população em questão.
3. Coletar os dados. Devemos ser extremamente cautelosos, para minimizar os erros que podem resultar de uma coleta tendenciosa de dados.
4. Analisar os dados e tirar conclusões. Identificar também possíveis fontes de erros.

Os estudos que requerem métodos estatísticos decorrem tipicamente de duas fontes comuns: estudos observacionais e experimentais.

Estudo observacional – verificamos e medimos características específicas, mas não tentamos manipular ou modificar os elementos a serem estudados. Ex: plebiscito sobre porte de arma de fogo.

Estudo experimental – aplicamos determinado tratamento e passamos então a observar seus efeitos sobre os elementos a serem pesquisados. Ex: tratamento médico a um determinado grupo de pacientes a fim de determinar sua eficiência na cura.

4. População e Amostra

Qualquer estudo científico enfrenta o dilema de estudo da população ou da amostra. Obviamente teria-se uma precisão muito superior se fosse analisado o grupo inteiro, a população, do que uma pequena parcela representativa, denominada amostra. Observa-se que é impraticável na grande maioria dos casos, estudar a população em virtude de distâncias, custo, tempo, logística, entre outros motivos. A alternativa praticada nestes casos é o trabalho com uma amostra confiável.

População (N): Conjunto de todos os elementos relativos a um determinado fenômeno que possuem pelo menos uma característica em comum, a população é o conjunto Universo.

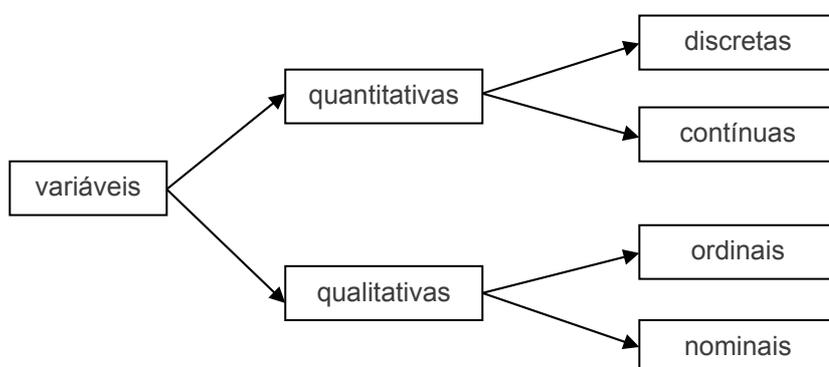
Amostra (n): É um subconjunto da população. A amostra deve ser selecionada seguindo certas regras e deve ser representativa, de modo que ela represente todas as características da população como se fosse uma fotografia desta.

Dados estatísticos: é qualquer característica que possa ser observada ou medida de alguma maneira. As matérias-primas da estatística são os dados observáveis.

Variável: é o que se deseja observar para se tirar algum tipo de conclusão. Geralmente as variáveis para estudo são selecionadas por processos de amostragem. Os símbolos utilizados para representá-las são letras maiúsculas do alfabeto, tais como X, Y, Z,... que podem assumir qualquer valor de um conjunto de dados.

Para podermos decidir como organizar os dados é preciso saber com que tipo de variáveis estamos trabalhando. Os tipos de variáveis são:

- quantitativas que podem ser discretas ou contínuas;
- qualitativas que podem ser ordinais ou nominais.



As variáveis **quantitativas discretas** assumem valores pontuais. Por exemplo, número de irmãos. Neste caso, representa valores bem definidos como 1, 2, 3, 0 (nenhum) irmãos.

As variáveis **quantitativas contínuas** assumem valores dentro de um intervalo. Por exemplo, podemos considerar a massa das pessoas em gramas. É claro que uma pessoa pode ter 60 235 gramas ou 60 236 gramas. Caberia a pergunta: não seria uma variável discreta? Neste caso, temos um conjunto muito grande de valores que essa variável pode assumir tornando-a contínua.

As variáveis **qualitativas ordinais** são aquelas que atribuem qualidades de modo que possam ser ordenadas de maneira hierárquica. Por exemplo, o grau de escolaridade: analfabeto, 1º grau incompleto, 1º grau completo, 2º grau incompleto e assim por diante.

Por fim, as variáveis **qualitativas nominais** são aquelas que atribuem qualidade mas que não é possível fazer uma ordenação. Por exemplo, matéria do colégio que mais gostava: Matemática, Física, Biologia, História...

É importante ressaltar que não existem regras fixas para se dizer que uma variável é discreta ou contínua. Muitas vezes, podemos dar tanto um tratamento contínuo à variável idade quanto um tratamento discreto. Tal decisão depende do que se quer analisar e da quantidade de dados envolvida. Por exemplo: se estivermos fazendo uma pesquisa numa festa e encontramos jovens de 18 a 25 anos, podemos considerar a variável idade como discreta, ou seja, podemos contar exatamente quantas pessoas há

com 18, 19, 20, 21, 22, 23, 24 e 25 anos. Porém, imaginemos que numa outra festa, com 1000 convidados, encontrássemos pessoas de 3 à 80 anos. É claro que poderíamos contar o número de indivíduos com 3,4,5,6,..., 79 e 80 anos. Porém, muitas vezes, nosso interesse está em analisar algumas faixas etárias. Por exemplo:

de 3 a 18 anos
de 18 a 25 anos
de 25 a 35 anos
de 35 a 50 anos
de 50 a 80 anos

Nesse caso, a variável idade passa a receber um tratamento contínuo.

Assim, é preciso tomarmos muito cuidado com o fato de que algumas pessoas defendem que a variável IDADE é discreta. Dependendo do tratamento dado a ela, podemos transformá-la de discreta para contínua.

Vejamos um outro caso: suponhamos um fabricante de tintas, que produz tintas coloridas fazendo o uso da tinta branca+pigmentos. Suponhamos, ainda, que ele trabalhe com as seguintes cores: branco, amarelo, vermelho, azul e preto. Aparentemente, a variável COR é qualitativa nominal. Porém, esse fabricante afirma que o pigmento amarelo é mais barato que o vermelho e que para se produzir tinta azul se usa muito corante (e mais corante ainda para tinta preta). Isso faz com que os custos sejam elevados para a tinta preta e reduzidos para a branca. Neste caso, podemos estabelecer uma ordem crescente para os custos:

1°) branco
2°) amarelo
3°) vermelho
4°) azul
5°) preto

Percebemos que foi estabelecida uma ordem. Assim, a variável COR é, agora, qualitativa ordinal.

Questionário

Para efeitos de análise, foi passado um questionário para uma amostra de 30 ouvintes de uma determinada palestra. Pediu-se para que respondessem com a maior exatidão possível. Um modelo do questionário é mostrado a seguir.

Questionário

Procure responder às questões com a maior exatidão possível.
Não deixe questões em branco!

1) Sexo: () masculino () feminino

2) Idade (em anos): _____

3) Altura (em metros): _____ m

4) Peso (em quilos): _____ kg

5) Número de irmãos (vivos): _____

6) Fuma atualmente? () SIM () NÃO

7) Qual a sua tolerância quanto à fumaça do cigarro?
() Muito tolerante () Pouco tolerante () Indiferente

8) Número de horas médias por semana que pratica exercícios e atividades físicas (academia, andar, correr, alongamento, esportes, etc): _____ horas

9) Qualidade da programação atual da Rede Globo:
() Boa () Regular () Péssima () Não sabe

É importante destacar alguns pontos importantes a respeito do questionário:

- ❑ Deve-se ter muito cuidado na elaboração das questões para que não gerem ambigüidades quanto à interpretação nem problemas de respostas diferentes que não possibilitem uma análise posterior.
- ❑ O resultado de um questionário nem sempre corresponde à realidade, visto que a pessoa pode não saber ao certo sua altura ou fazer muito tempo que não se pesa.
- ❑ Alguns cuidados especiais devem ser tomados na elaboração de questões “abertas”, ou seja, que não são do tipo teste, para que as respostas sejam padronizadas. Por exemplo, se não for especificado, uma pessoa pode responder que a sua altura é de 172 cm e outra de 1,72 m. Ou ainda, o que seria muito pior pois alteraria o resultado da pesquisa, é no caso de perguntar o número de irmãos: uma pessoa pode ter 4 irmãos vivos e 1 que faleceu. Qual valor ela deveria colocar no questionário: 4 ou 5? Daí a necessidade da especificação.

- Vale a pena, também, ficar atento a perguntas do tipo: “você gosta de carros brancos? () sim () não”. Aparentemente não há nenhum problema nessa pergunta, porém, uma análise mais cuidadosa faria perceber que o entrevistado poderia responder “não, não gosto de carros brancos, prefiro os vermelhos” como poderia responder “não, não gosto de carros, prefiro motos”. Porém, essa diferença de respostas não seria detectada com a pergunta (ambígua) acima. Neste caso, devemos reformular tal pergunta ou fazer outras confirmatórias. Embora isso não seja tratado neste texto, alertamos quanto ao fato na hora de elaborar e responder um questionário.

O resultado de tal questionário em uma amostra de tamanho 30 é mostrado na tabela a seguir. As variáveis em questão são:

Sexo – masculino (M) ou feminino (F)

Idade – em anos

Altura – em metros

Peso – em quilos

Irmãos – número de irmãos vivos

Fuma – é fumante (SIM) ou não é fumante (NÃO)

Tolerância – nível de tolerância à fumaça do cigarro: muito tolerante (M), pouco tolerante (P) ou indiferente (I)

Exercícios – número médio de horas que pratica atividades físicas por semana

Qualidade – qualidade da programação atual da Rede Globo: boa (B), regular (R), péssima (P) ou não sabe (N)

A partir da tabela a seguir, onde estão representados os dados brutos (ou seja, aqueles obtidos a partir do questionário), percebemos que há uma certa dificuldade de, por exemplo, dizer se a maioria das pessoas é muito ou pouco tolerante ao fumo, ou quanto ao número médio de horas que as pessoas praticam atividades físicas. Tal dificuldade já se apresenta com um pequeno conjunto de dados (apenas 30 entrevistados). Para conjuntos maiores, diria que é praticamente impossível tirar alguma conclusão apenas observando os dados brutos.

Daí a necessidade de reorganizarmos os dados em tabelas e gráficos. A organização em tabelas deve ser a mais simples possível, evitando-se utilizar tabelas muito incrementadas ou coloridas. A forma como esses dados serão organizados também pode variar, de acordo com os interesses e do que se quer analisar. Assim, daremos aqui, alguns exemplos de organização e tipos de gráficos.

Aliás, quanto aos gráficos, nem sempre há um gráfico correto e outro errado. Para representar um conjunto de dados, muitas vezes é possível usar mais de um tipo de gráfico. O melhor é aquele que mais enfatiza o resultado que você deseja apresentar, ou seja, que dá maior destaque às informações que você julga importantes.

Observaçã o	Sex o	Idad e	Altur a	Pes o	Irmão s	Fum a	Tolerânci a	Exercício s	Qualidad e
1	F	17	1,60	60	0	SIM	I	0	B
2	F	18	1,69	55	2	SIM	I	0	R
3	M	18	1,85	73	1	NÃO	M	5	R
4	M	23	1,85	80	0	NÃO	M	4	P

5	F	19	1,55	50	0	SIM	I	2	B
6	M	19	1,76	60	2	NÃO	M	2	P
7	F	20	1,64	47	1	NÃO	P	3	B
8	F	18	1,62	58	1	SIM	I	2	N
9	F	18	1,64	58	3	NÃO	P	10	R
10	F	17	1,72	70	0	NÃO	M	8	B
11	F	18	1,66	54	2	NÃO	P	5	B
12	F	18	1,70	58	0	NÃO	I	2	R
13	F	21	1,65	63	1	SIM	P	1	R
14	M	18	1,90	85	2	NÃO	P	0	B
15	M	18	1,65	70	2	NÃO	P	0	R
16	M	19	1,70	70	1	NÃO	I	3	P
17	M	20	1,75	68	3	SIM	I	2	N
18	M	22	1,78	65	4	NÃO	P	3	R
19	M	24	1,79	72	1	NÃO	M	5	B
20	M	23	1,84	81	5	NÃO	P	5	B
21	F	18	1,64	54	2	NÃO	I	10	B
22	F	19	1,70	59	1	NÃO	P	6	B
23	F	21	1,78	60	0	NÃO	M	2	R
24	F	24	1,69	62	1	NÃO	I	1	R
25	F	21	1,72	70	2	NÃO	P	7	P
26	F	19	1,74	65	4	NÃO	P	7	B
27	M	18	1,75	70	1	NÃO	P	6	P
28	F	20	1,67	54	1	NÃO	M	5	R
29	M	20	1,81	76	3	NÃO	P	7	B
30	M	24	1,79	65	0	NÃO	P	12	B

Baseado na classificação de variáveis que apresentamos, podemos dizer que são:

SEXO – nominal

IDADE – contínua

ALTURA – contínua (pois assume uma grande variedade de valores, embora possamos considerá-la discreta)

PESO – contínua

IRMÃOS – discreta

FUMA – nominal

TOLERÂNCIA – nominal

EXERCÍCIOS – discreta

QUALIDADE – ordinal

RESUMO

As variáveis podem ser classificadas dos seguintes modos:

1) Qualitativas (ou atributos): São características de uma população que não podem ser medidas, sendo classificadas em nominais ou ordinais.

- **Nominal:** são utilizados símbolos, ou números, para representar determinado tipo de dados, mostrando, assim, a qual grupo ou categoria eles pertencem, como sexo, nacionalidade, etc.

- **Ordinal:** quando uma classificação for dividida em categorias ordenadas em graus convencionados, havendo uma relação entre as categorias do tipo “maior do que”, “menor do que”, “igual a”, primeiro, segundo, terceiro e, assim, sucessivamente.

2) Quantitativas: São características populacionais que podem ser quantificadas, sendo classificadas em discretas e contínuas.

- **Discretas:** são aquelas variáveis que podem assumir somente valores inteiros num conjunto de valores. É gerada pelo processo de contagem, como o número de veículos que passa em um posto de gasolina, o número de estudantes nesta sala de aula.

- **Contínuas:** são aquelas variáveis que podem assumir um valor dentro de um intervalo de valores. É gerada pelo processo de medição. Neste caso serve como exemplo o volume de água em um reservatório ou o peso de um pacote de cereal.

Exercícios

1) Classifique as variáveis em qualitativas (nominais/ordinais) ou quantitativas (discretas/contínuas).

- cor dos cabelos dos alunos de uma escola.
- número de filhos de casais residentes em uma determinada rua.
- o ponto obtido em cada jogada de um dado.
- naturalidade das pessoas que vivem na cidade de São Paulo.
- escolaridade dos funcionários de uma empresa.

2) Diga quais das variáveis são discretas e quais são contínuas: salários, sexo dos filhos, número de peças defeituosas produzidas por uma máquina, altura de pessoas, grau de instrução, número de filhos, peso.

3) Classifique as variáveis em qualitativas (nominal ou ordinal) ou quantitativas (discreta ou contínua):

- número de ações negociadas na bolsa;

- b) número de filhos de um certo casal;
- c) comprimento dos pregos produzidos por uma máquina;
- d) número de volumes na biblioteca da UNIBAN;
- e) salário dos funcionários de uma empresa;
- f) cor predominante da parede externa de sua casa;
- g) grau de escolaridade;
- h) número de horas dormidas na última noite;
- i) tipo de comida preferido;
- j) cargo dos funcionários de uma empresa.