

IPESU
Instituição Pernambucana de Ensino Superior

Maria Cristiane Santana Andrade
Curso: Gestão Empreendedora

Título:

Introdução à Estatística

Recife - PE

2008

Sumário

01. Introdução.....	3
02. Conceito.....	3
03. Objetivos da Estatística	5
04. Conceitos Básicos (População, Amostra, Variáveis)	5
05. Ramos da Estatística.....	7
06. Dados Estatísticos.....	7
07. Formas Iniciais de Tratamento de Dados	10
08. Notações por Índices	12
09. Notação Sigma.....	12
10. Referências	15

01. Introdução

O Propósito deste trabalho é apresentar os conceitos básicos da estatística de uma forma clara e concisa. Assim cada tópico foi inscrito de uma forma aplicada e acessível, com o intuito de facilitar a compreensão e enfatizar o aprendizado.

02. Conceito

E palavra Estatística tem duas origens descritas, primeiro do alemão, Statistik = estatística. Palavra criada em 1748 pelo alemão Gottfried Achenwll.

Mas ela é cunhada na palavra latina, statisticum = assuntos relativos ao governo, ao Estado (em latim Statum = país), aparece como vocabulário na Enciclopédia Britânica em 1797 (Filho, 1999).

Estatística é Ciência ligada à matemática que através de dados agrupados relativos ao passado explica o presente e projeta o futuro.

Segundo Triola (1999), A Estatística é uma coleção de métodos para planejar experimentos, obter e organizar dados resumi-los, analisá-los e deles extrair conclusões.

Ao longo do século XX, os métodos estatísticos foram desenvolvidos como uma mistura de ciência, tecnologia e lógica para a solução e investigação de problemas em várias áreas do conhecimento humano (Stigler, 1986). Ela foi reconhecida como um campo da ciência neste período, mas sua história tem início bem anterior a 1900.

A estatística moderna é uma tecnologia quantitativa para a ciência experimental e observacional que permite avaliar e estudar as incertezas e os seus efeitos no planejamento e interpretação de experiências e de observações de fenômenos da natureza e da sociedade.

A estatística não é uma caixa-preta, nem bola de cristal, nem mágica. Nem tão pouco um conjunto de técnicas úteis para algumas áreas isoladas ou restritas da ciência. Por exemplo, ao contrário do que alguns imaginam, a estatística não é um ramo da matemática onde se investigam os processos de obtenção, organização e análise de dados sobre uma

determinada população. A estatística também não se limita a um conjunto de elementos numéricos relativos a um fato social, nem a números, tabelas e gráficos usados para o resumo, à organização e apresentação dos dados de uma pesquisa, embora este seja um aspecto da estatística que pode ser facilmente percebido no cotidiano (basta abrir os jornais e revistas para ver o "bombardeio" de estatísticas). Ela é uma ciência multidisciplinar: um mesmo programa de computador que permite a análise estatística de dados de um físico poderia também ser usado por um economista, agrônomo, químico, geólogo, matemático, biólogo, sociólogo, psicólogo e cientista político. Mesmo que as interpretações dessas análises sejam diferentes por causa das diferenças entre as áreas do conhecimento, os conceitos empregados, as limitações das técnicas e as consequências dessas interpretações são essencialmente as mesmas.

Segundo Rao (1999), a estatística é uma ciência que estuda e pesquisa sobre: o levantamento de dados com a máxima quantidade de informação possível para um dado custo; o processamento de dados para a quantificação da quantidade de incerteza existente na resposta para um determinado problema; a tomada de decisões sob condições de incerteza, sob o menor risco possível. Finalmente, a estatística tem sido utilizada na pesquisa científica, para a otimização de recursos econômicos, para o aumento da qualidade e produtividade, na otimização em análise de decisões, em questões judiciais, previsões e em muitas outras áreas.

De uma forma resumida e direta poderemos dizer que a estatística pode ser pensada como a ciência de aprendizagem a partir de dados.

Aprender a partir de dados é um dos desafios mais relevantes da era da informação que vivemos. O mundo está cheio de incerteza. Nós continuamente tomamos decisões sem saber como elas afetarão eventos futuros. Quando somos cuidadosos, tentamos pensar logicamente, avaliar a situação e determinar nossas prioridades.

03. Objetivos da Estatística

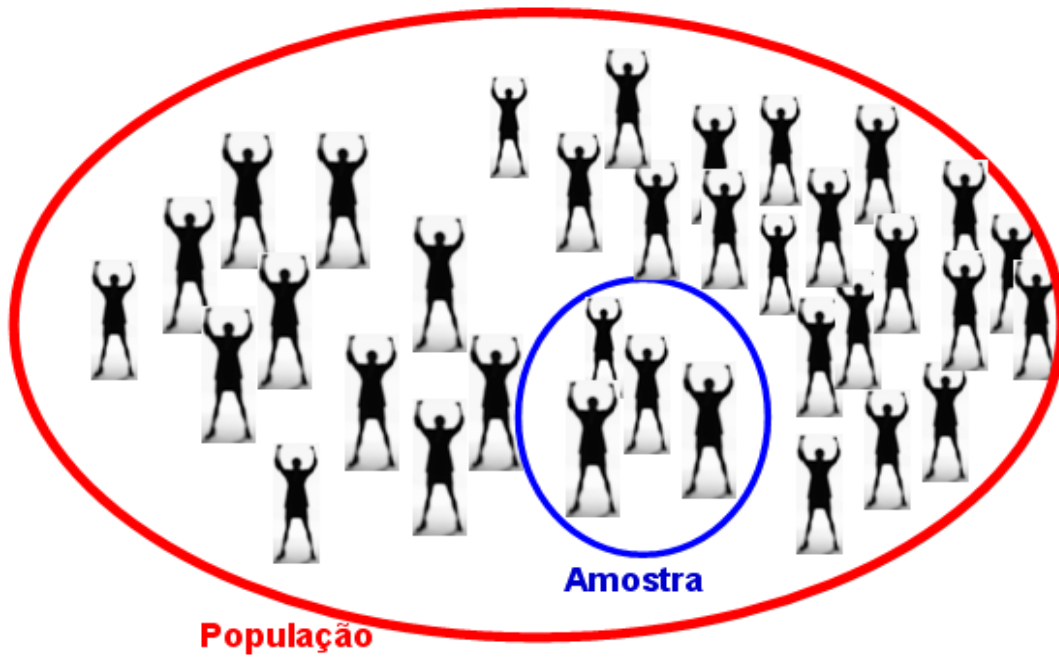
- Sumarizar a coleção de observações;
- Descrever o conhecimento de um dado assunto de forma matemática;
- Evitar manipulação de resultados;
- Dar “polimento” a publicações;
- Analisar a coleção de dados
- Determinar correlações
- Saber o grau de certeza das conclusões tiradas;

04. Conceitos Básicos (População, Amostra, Variável)

População: qualquer conjunto de informação que tenha entre si uma característica comum que delimite os elementos pertencentes a ela.

Amostra: é um subconjunto de elementos pertencentes a uma população.

Variável: Dados referentes a uma característica de interesse, coletados a partir de uma amostra.



A Amostra faz parte de uma população, ou seja, baseado na população e que podemos retirar uma amostra para servir como base estatística.

Variável



Variável é um dado referente a uma característica, exemplo: estatura, peso, cor e etc.

05. Ramos da Estatística

A Estatística possui três ramos principais: Estatística Descritiva, Teoria da Probabilidade e Teoria da Inferência.

Estatística Descritiva: Que envolve a organização e sumarização dos dados através de metodologia simples;

São exemplos de medidas da estatística descritiva a taxa de desemprego, os índices de inflação, a quilometragem média por litro de combustível, entre outros.

Teoria da Probabilidade: que proporciona uma base racional para lidar com situações influenciadas por fatores que envolvem o acaso;

Os jogos de azar, a maior parte dos jogos esportivos, a megasena, as decisões de marketing envolvem, em alguma medida, o acaso, e portanto, a probabilidade.

Teoria da Inferência: que envolve a análise e interpretação da amostra;

A inferência ou amostragem tem como idéia básica extrair determinadas informações de uma parcela pequena de uma população determinada, e a partir disto fazer inferência sobre toda a população. Ou seja, a amostra deverá necessariamente representar toda a população.

06. Dados Estatísticos

Dados estatísticos são todas as informações levantadas (coletadas) que servirão como base para o estudo e análise estatística e que chamaremos de **Dados**.

Esses dados precisarão ser organizados e sumarizados para sua correta interpretação. Ora, caso os dados ainda não foram numericamente organizados e processados, eles podem se apresentar a nós com quase nenhum sentido.

Estes seriam os chamados dados brutos. É o processamento e organização dos dados que os transforma em informação, enfatizando seus aspectos mais importantes. A informação, portanto, é resultado de um tratamento dos dados.

Este tratamento de dado é utilizado para mensurar esta informação, cada informação captada é variável, por exemplo: Quando fazemos uma pesquisa, os dados contidos nela são as variáveis.

Exemplo de uma pesquisa: Uma indústria automobilística que pretende lançar um novo modelo de carro faz uma pesquisa para sondar a preferência dos consumidores sobre o tipo de combustível, número de portas, potência do motor, preço, cor, tamanho, etc., cada uma dessas características é uma variável da pesquisa.

Na variável “tipo de combustível”, a escolha pode ser, por exemplo, entre o álcool e a gasolina. Dizemos que esses são “valores ou realizações” da variável “tipo de combustível”. Em estatística temos duas classificações de variáveis, que se desdobram formando quatro tipos de variáveis, como descritas no quadro abaixo:

Variáveis

Qualitativa	Nominal = pode nominar
	Ordinal = pode ordenar
Quantitativa	Discreta = pode contar
	Continua = pode medir

Variável Qualitativa

Em uma pesquisa que envolve pessoas, por exemplo, as variáveis consideradas podem ser: sexo, cor de cabelo, esporte favorito e grau de instrução.

Nesse caso dizemos que as variáveis são qualitativas, pois apresentam como possíveis valores uma qualidade (ou atributo) dos indivíduos pesquisados. Além disso, dizemos que as variáveis qualitativas podem ser ordinais, quando existe uma ordem nos seus valores, ou nominais, quando isso não ocorre.

Variáveis Nominais: são aquelas que existem com o objetivo de definir categorias, e as observações, mensurações e análises são feitas levando-se em conta estas mesmas categorias. Exemplos de categorias seriam a separação por sexo, idade, nível de escolaridade, etc.

Variáveis por Posto ou Ordinal: quando existe o desejo de dispor os elementos observados segundo uma ordem de preferência ou desempenho, atribui-se valores relativos atribuídos para indicar esta ordem. Exemplo: primeiro, segundo, terceiro.

Exemplo:

“Grau de instrução” é uma variável qualitativa ordinal já que seus valores podem ser ordenados (fundamental, médio, superior, etc.)

Variável Quantitativa

Quando as variáveis são, por exemplo, altura, peso, idade em anos e número de irmãos, dizemos que elas são quantitativas, pois seus possíveis valores são números.

As variáveis quantitativas podem ser discretas, quanto se trata de contagem (números inteiros) ou contínuas, quando se trata de medida (números reais).

Variáveis Contínuas: é a variável que pode assumir qualquer valor num intervalo contínuo (dado contínuo). Exemplos: altura, peso, velocidade, etc.

Variáveis Discretas: em geral originam-se da contagem de itens e só podem assumir valores inteiros. Exemplos: número de alunos em sala de aula, número de professores que trabalham na escola, etc.

Exemplo:

1º “Número de irmãos” é uma variável discreta, pois podemos contar (0, 1, 2, 3...);

2º “Altura” é uma variável quantitativa contínua, uma vez que pode ser medida (1,55m, 1,80m, 1,73m...).

07. Formas Iniciais de Tratamento de Dados

Em geral, quando nos propomos a buscar construir informações a partir de dados, nos deparamos inicialmente com um conjunto de dados brutos que pouco nos dizem. É preciso organizá-los minimamente para que eles comecem a fazer algum sentido, viabilizando sua análise.

Basicamente temos duas formas de organização dos dados, são elas: Rol e Distribuição de Frequência.

Rol

É um arranjo de dados brutos em ordem crescente ou decrescente.

Exemplo:

Consumo Mensal de Energia Elétrica, por 50 Usuários Particulares

KWH(quilowatts-hora)				
8	58	75	89	118
9	58	76	90	121
10	60	78	90	125
19	62	80	92	126
28	64	81	94	131
36	66	82	94	136
38	72	83	95	144
50	73	84	96	148
52	74	86	105	157
57	75	88	114	158

Essa classificação dos dados proporciona algumas vantagens concretas com relação à sua forma original:

- é possível visualizar de forma ampla as variações de consumo
- os valores extremos são percebidos de imediato
- é possível observar uma tendência de concentração dos valores na faixa de 50-90 kwh

Apesar de o rol propiciar ao analista mais informações e com menos esforço de concentração do que os dados brutos, ainda assim persiste o problema de a análise ter que se basear nas 50 observações. O problema se agravará quando o número de dados for muito grande.

Distribuição ou Tabelas de frequências

As distribuições ou tabelas de frequências são representações nas quais os valores se apresentam em correspondência com suas repetições, evitando-se assim que eles apareçam mais de uma vez na tabela como ocorre com o rol.

Exemplo: Uma empresa fabricante de instrumentos de precisão está interessada em saber o número de aparelhos defeituosos rejeitados pela seção encarregada do controle de qualidade.

Número Mensal de Aparelhos Defeituosos

Mês / Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2004	6	2	5	6	0	8	7	6	3	4	5	8
2006	10	9	7	6	3	4	6	4	5	4	5	1
2006	3	6	7	9	3	1	4	6	5	3	5	4
2007	7	2	5	8	6	4	2	5	1	6	5	2

Os dados brutos não informam muita coisa sobre o fenômeno “número de aparelhos defeituosos”.

Valores repetidos como o zero (que aparece duas vezes), esse fato irá sugerir, naturalmente, que se condensem todos os resultados em uma tabela, estabelecendo a correspondência entre o valor individual e o respectivo número de vezes que ele foi observado (frequência).

Através de uma tabela de frequências obtemos estatísticas (medidas baseadas na amostra) com menos cálculo e em menos tempo do que se esse trabalho fosse realizado a partir dos dados brutos.

As tabelas de frequências podem representar tanto valores individuais como valores agrupados em classes.

08. Notações por Índices

A notação por índices é bastante utilizada na estatística, sendo assim importante que esclareçamos seu significado. O símbolo x_i (onde se lê “x índice i”) irá representar qualquer um dos n valores assumidos pela variável x , $x_1, x_2, x_3, \dots, x_n$. “ n ” é denominada índice e poderá assumir qualquer dos números entre 1, 2, 3, 4, ..., n .

09. Notação Sigma

A maioria dos processos estatísticos vai exigir o cálculo da soma de um conjunto de números. A letra maiúscula grega sigma (Σ) é utilizada para representar estas somas.

Quando queremos representar uma soma de n valores do tipo $x_1 + x_2 + \dots + x_n$, podemos codificá-la através da expressão:

$$x_1 + x_2 + x_3 + \dots + x_n = \sum_{i=1}^n x_i$$

Onde:

Σ - é utilizada para representar as operações de adição entre as parcelas.

x_i - é a parcela genérica.

A parcela genérica é obtida tomando-se os tempos constantes em todas as parcelas, no caso x . Para representar a parte variável em cada parcela, no caso os índices, utilizamos a letra i e indicamos a variação de i .

Para que uma soma possa ser representada por essa notação é necessário que i assumam todos os valores interiores consecutivos entre os dois valores dados. Assim, a soma:

$$x_1 + x_2 + \dots + x_4 \neq \sum_{i=1}^4 x_i$$

Cabe destacar que, apesar de ser apenas um código e não uma operação, a notação sigma tem algumas propriedades que podem simplificar operações. Entre elas destacamos:

1. O somatório de uma soma é a soma dos somatórios.

$$\sum_{i=1}^n (x_i + y_i) = \sum_{i=1}^n x_i + \sum_{i=1}^n y_i$$

2. O somatório de uma diferença é a diferença do somatório.

$$\sum_{i=1}^n (x_i - y_i) = \sum_{i=1}^n x_i - \sum_{i=1}^n y_i$$

3. O somatório do produto de uma constante por uma variável é o produto da constante pelo somatório da variável.

$$\sum_{i=1}^n (a \times x_{ii}) = a \times \sum_{i=1}^n x_i$$

4. O somatório da divisão de uma variável por uma constante é a divisão do somatório da variável pela constante.

$$\sum_{i=1}^n \left(\frac{x_i}{a} \right) = \left(\frac{\sum_{i=1}^n x_i}{a} \right)$$

10. Referências

FILHO, UD. Introdução à Bioestatística. 9ªed. São Paulo: Elsevier; 1999.

TRIOLA, MF. Introdução à Estatística. LTC Livros Técnicos e Científicos Editora, 1999.

RAO, C.R. Statistics: A technology for the millennium Internal. J. Math. & Statist. Sci, Vol. 8, No. 1, June 1999, 5-25.

STIGLER, Stephen M. The History of Statistics: The Measurement of Uncertainty Before 1900. The Belknap Press of Harvard University Press. Cambridge, USA, 1986.

IPESU, Faculdade Ipesu: Apostila de ESTATÍSTICA APLICADA – EAD, 2008.

SIMIANO, Mauricio N. Apostila de Estatística Aplicada I. UNIFIN – União das Faculdades Integradas de Negócio, 2004.

