



Estatística Computacional

Pós- Graduação em Ciência dos Dados

Prof. Dr. Roberta A. de A. Fagundes
roberta.fagundes@upe.br

Homepage
robertafagundes.wix.com/raaf



Quem sou ?

- Roberta A. de A. Fagundes
 - Professora Adjunta UPE ;
 - Mestrado e Doutorado em Ciência da Computação no CIN/UFPE
 - Pós-Doutorado em Estatística no Departamento de Estatística da UFPE;
- Pesquisas sobre:
 - Mineração de Dados;
 - Aprendizagem de Máquina (Regressão, Cluster e Classificação);
 - Algoritmos de Otimização
 - Análise Dados Simbólicos do Tipo Intervalo.

Conhecendo a Turma ?



- Nome
- Formação
- Experiência com Estatística

Conhecendo a Turma ?

■ Conhecimentos Adquiridos

Introdução à Ciência dos Dados	Alexandre Magno	30
Tecnologias em Processo de Negócios	Rômulo César	30
Banco de Dados Não Convencionais	Andreza Leite de Alencar	30
Estatística Computacional	Roberta A. de A. Fagundes	30



Ementa

■ Teórico

- Estatística Descritiva ou Exploratória;
- Testes de Hipóteses e ANOVA;
- Modelos de Regressão Linear e Logístico;
- Análise de Conglomerados e Fatorial.

■ Prático

- Linguagem R



Metodologia

- Intersecção entre teoria e prática;
- Reflexões sobre a estatística no cotidiano;
- Relação com problemas reais ;
- Exercícios práticos e teóricos durante as aulas.

Metodologia

■ Avaliações Teóricas

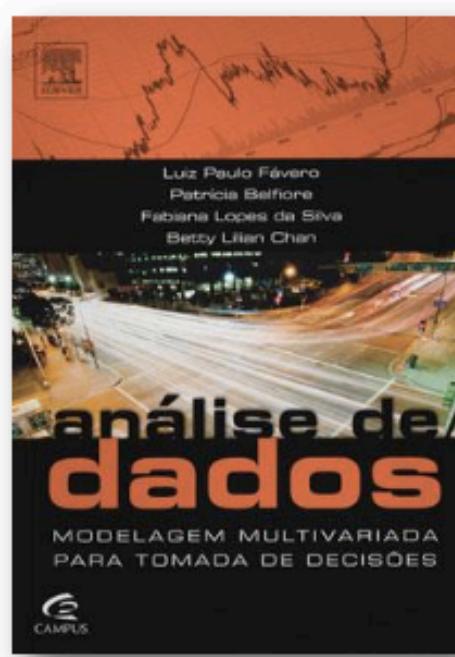
- Exercícios em sala de aula – 40%
- Participação em sala – 10%

■ Avaliações Práticas

- Realização de projetos práticos – 40%
- Apresentação dos resultados – 10%



Livros de Referência





O que veremos na aula

- Estatística
 - Definição
- Processos de amostragem;
- Estatística Descritiva;
- Natureza das variáveis;
- Medidas de Posição;
- Medidas de Dispersão;

Definição

O que é a estatística ?

Estatística é o estudo das populações, das variações e dos métodos de redução de dados.

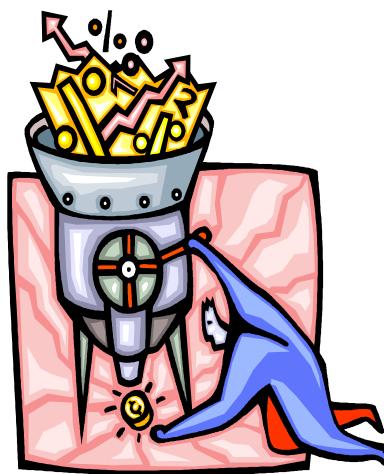


Sir Ronald A. Fisher (1890-1962).

Definição

O que é a estatística ?

“Eu gosto de pensar na Estatística como a ciência de aprendizagem a partir dos dados...”



**Jon Kettenring
Presidente da American Statistical Association, 1997**

Definição

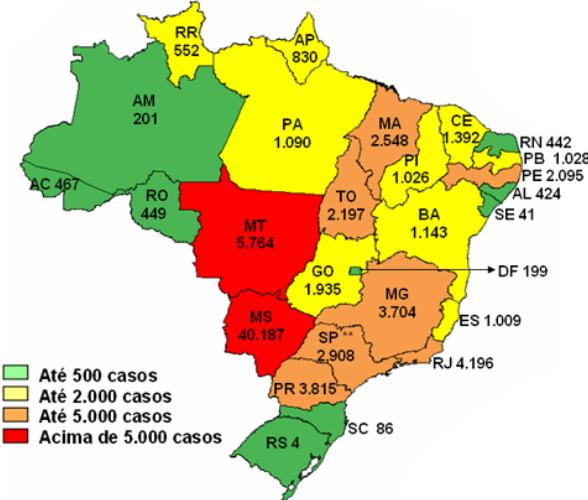
Onde acho a Estatística no meu dia-a-dia ?

- * Nos meios de comunicação...

Ex.: Dengue 2007 no Brasil

A dengue no País

Dados do Ministério da Saúde referentes às oito primeiras semanas de 2007



- * Na **média final** que você ficou em uma dada matéria da sua escola...
- * Quando se fala, por exemplo, que aproximadamente **51%** dos brasileiros são do sexo feminino...

Definição

Estatística no prática ?

- Levantamentos de dados dos alunos presentes para traçar um perfil

O que perguntar ?



Elaboração de Questionário !

Como resumir as respostas ?



Estatística Descritiva !

- Sorteio de Amostras de Populações

Populações: homogêneas ou heterogêneas

Definição

Como a Estatística pode ajudar ?

Fornecendo instrumentos que auxiliam a tomada de decisão em toda situação onde incertezas estão presentes.

Seu caráter interdisciplinar!

Pode trabalhar em qualquer área !

Definição

Alguns exemplos na área da saúde . . .

Institutos e laboratórios de pesquisas médicas e biomédicas, como: Anvisa, Fiocruz, etc.

- * Comparação entre medicamentos e procedimentos médicos;
- * Previsão de Epidemias;
- * Determinação de fatores de risco para certas doenças;
- * Determinação de associação entre certos genes e algumas doenças, etc.



Definição

Alguns exemplos na área social e econômica . . .

Institutos de pesquisas sociais, econômicas e de mercado, como: IBGE, IPEA, etc.

- * Determinação do perfil social do brasileiro,
- * Intenção de votos em Eleições,
- * Preferências do consumidor,
- * Previsão do comportamento da inflação e da produção industrial, etc.



Definição

Alguns exemplos na indústria . . .

Indústrias de vários ramos como: Petrobras, Coca-Cola, etc.

- * Logística;
- * Controle de qualidade;
- * Previsão de Demanda;
- * Satisfação dos empregados, etc.



Definição

Alguns exemplos na mercado financeiro . . .

- * Bancos, Seguradoras, Bolsa de valores, etc;
- * Previsão de ocorrência de sinistros (ex.: acidentes de carro, incêndio, etc);
- * Previsão de demanda por serviços;
- * Satisfação dos clientes, etc.



Definição

Ainda podemos citar o ensino . . .

- *Colégios, Escolas e Universidades (públicas ou privadas, para graduação ou pós-graduação);
- *Evasão escolar;
- *Desempenho dos alunos;
- *Identificação de perfil de aprendizagem;
- * A retenção dos alunos, devido ao alto índice de repetentes, etc.



Definição

Resumindo . . .

- * A Estatística utiliza **métodos matemáticos** para solucionar problemas reais de tomada de decisão quando há **incerteza**;
- * Em situações nas quais **poderíamos contar unicamente com a sorte**, temos um instrumento, que nos possibilita aumentar as chances de tomar a melhor decisão;
- * Na prática, a **Estatística** pode ser empregada como **ferramenta fundamental** em várias **outras ciências**.

Definição

Geração dos Dados . . .

- Existem diversas formas de coleta de dados
 - ✓ Coletas por instrumentos de coleta do tipo questionários;
 - ✓ Dados de redes sociais;
 - ✓ Dados por sensores;
- Através de questionários
 - ✓ Dados coletados por CENSO;
 - ✓ Pesquisas de opinião pública;
 - ✓ Pesquisas comportamentais;
 - ✓ Pesquisas políticas.

Definição

Geração dos Dados . . .

- **Através de plataformas web**
 - ✓ Facebook 968 milhões usuários ativos por dia;
 - ✓ Instagram + de 40 bilhões de fotos compartilhadas;
 - ✓ Instagram 3,5 bilhões de likes por dia;

facebook.



Google

YouTube

Processo de Amostragem

Amostragem

Uma área importante em muitas aplicações Estatísticas é a da Tecnologia de Amostragem.

Exemplos de Aplicação:

- Pesquisa de mercado,
- Pesquisa de opinião,
- Avaliação do processo de produção,
- Praticamente em todo experimento.

Processo de Amostragem

- **Por que ?**
 - Verificar a quantidade de poluentes do rio;
 - Entrega declarações a Recita Federal, só alguns caem em malha fina.
- **Vantagens**
 - Economia de tempo (**parte da população**);
 - Confiabilidade dos dados; (**menos dados**)
 - e operacionalidade (**menor escala**);



Processo de Amostragem

- O que devemos **evitar** no processo de amostragem?
 - Amostras por conveniência;
 - Amostra de voluntários;
 - Amostra intencionais;
 - Amostra com amigos ou conhecidos.



Processo de Amostragem

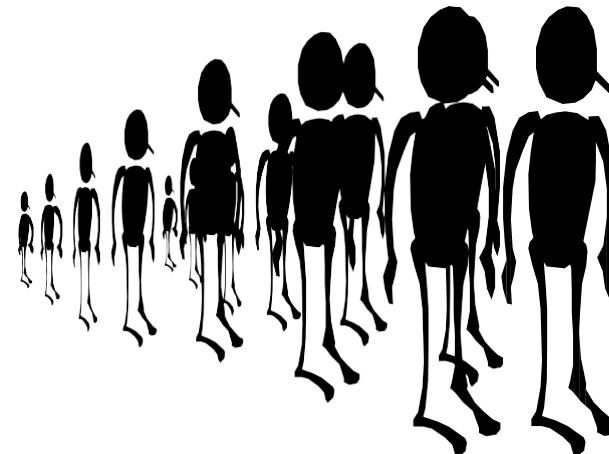
POPULAÇÃO:

- moradores de uma metrópole
- Valores computados são parâmetros



AMOSTRA:

- uma parte dos moradores
- Valores computados são estatísticas



Processo de Amostragem

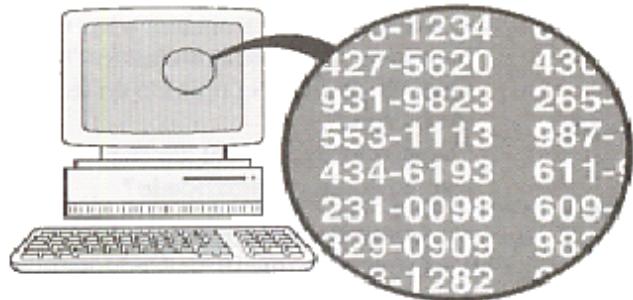
- A amostra deve ser representativa ?
 - Ex.: Empresa de 200 funcionários
 - $10\% = 20$ funcionário para compor a amostra

Mito ou Verdade ?



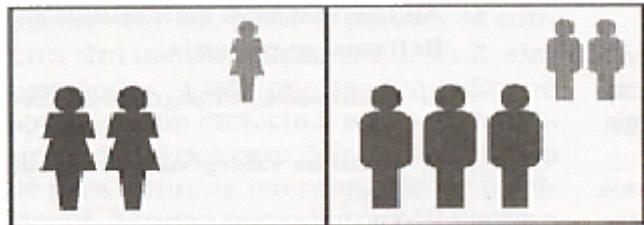
Processo de Amostragem

- **Tipos:**



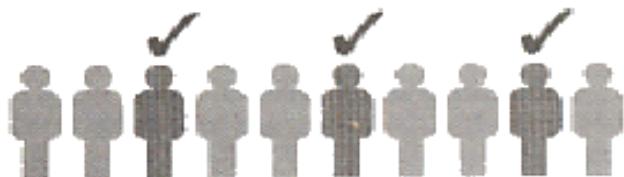
Amostragem Aleatória

Cada elemento da população tem a mesma chance de ser escolhido.



Amostragem Estratificada (Uniforme ou proporcional)

Classificar a população em, ao menos dois estratos e extrair uma amostra de cada um.



Amostragem Sistemática

Escolher cada elemento de ordem k .

Técnicas de Amostragem

- **Aleatória:** é aquela na qual todos os elementos da população têm chances iguais de serem selecionados.
 - Ex.: Contagem do número de pessoas que vivem em um domicílio.
 - Uma tabela de números aleatórios pode gerar os domicílios da amostra.
- **Estratificada:** Quando for importante que uma amostra tenha elementos de cada segmento da população
 - Ex.: Os domicílios da amostra podem ser divididos em níveis sócio-econômicos.

Técnicas de Amostragem

- **Ex.:** Uma empresa de Telemarketing conta com 480 funcionários, dos quais 288 são do sexo feminino e os 192 restantes do sexo masculino. Foi selecionada uma amostra de 50 funcionários, calcule a proporção de cada sexo contida na amostra.



Técnicas de Amostragem

- **Ex.:** Com o objetivo de levantar o estilo de comunicação, realiza-se o levantamento por amostragem. A população é composta por 200 chefes de seção. 4.400 operários especializados e 1.200 operários não especializados. Obtenha uma amostra com 5% dos operários da indústria, mantendo as mesmas relações de proporcionalidade em cada extrato.



Técnicas de Amostragem

- **Estratificada Uniforme:** Quando a população apresenta ocorrência natural de subgrupos, cada um deles com características similares. É selecionado uma amostra de grupos e todos os elementos dos grupos são usados no estudo.
Ex.: Coletar uma amostra de grupos de pessoas de acordo com o CEP.
- **Sistemática:** Os elementos da população são ordenados de acordo com uma numeração estabelecida e o número inicial é selecionado aleatoriamente e depois os demais elementos da amostra são selecionados segundo intervalos regulares.
- **Ex.:** Coletar domicílios na cidade de Recife selecionando cada domicílio a cada cem da população.

Técnicas de Amostragem

- **Amostragem Sistemática:**
- **Definir o tamanho da amostra "n";**

Calcular

$$k = \frac{N}{n}$$

... 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

- **Selecionar 1. item aleatoriamente**
- **Selecionar os demais a partir desse inicial**
 - **Posição primeiro = m**
 - **Posição segundo = m + K**
 - **Posição terceiro = m + 2*K**
 - **Posição Último elemento = primeiro + (n-1)*K**

Técnicas de Amostragem

- **Ex.:** Uma seguradora mantém uma carteira de 5.000 clientes, e pretende avaliar a satisfação de seus clientes mediante uma amostra sistemática de 200 segurados.
 - A) Determine os números dos 5 primeiros clientes selecionados, supondo que o primeiro segurado (obtido por sorteio) seja o de número 14.
 - B) Qual o número do último cliente selecionado?

Técnicas de Amostragem

- **Ex.:** Uma seguradora mantém uma carteira de 5.000 clientes, e pretende avaliar a satisfação de seus clientes mediante uma amostra sistemática de 200 segurados.
 - A) Determine os números dos 5 primeiros clientes selecionados, supondo que o primeiro segurado (obtido por sorteio) seja o de número 14.
 - B) Qual o número do último cliente selecionado?
 - **Resposta**
 - $N = 5000; n = 200; K = 5000/200 \Rightarrow K = 25$
 - A) $m = 14, m_1 = 14 + 25 = 39, m_2 = 14 + 50 = 64, m_3 = 89, m_4 = 114$
 - B) $m + (n-1)*k = 14 + (200-1)*25 = 4989$
 - **14,39,64,114, ..., 4989**

Técnicas de Amostragem

- **Ex.:** Uma empresa prestadora de serviço mantém um cadastro de 2.185 clientes, e pretende verificar o interesse de seus clientes por um novo produto; para isso, seleciona uma amostra sistemática de 70 clientes.
 - Supondo que o primeiro segurado (obtido por sorteio) seja o de número 21, determine os números dos próximos 5 clientes selecionados.
 - Qual o número do último cliente selecionado ?



Estatística Descritiva

Definição de Estatística

A estatística é um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento.

Estatística Descritiva

- A estatística lida com o planejamento, coleta, análise e interpretação dos dados como ferramentas auxiliares na tomada de decisões e resolução de problemas;
- Áreas da **computação** que utilizam **estatística**:
 - Inteligência Artificial;
 - Extração de informação;
 - Mineração de Dados/Texto;
 -

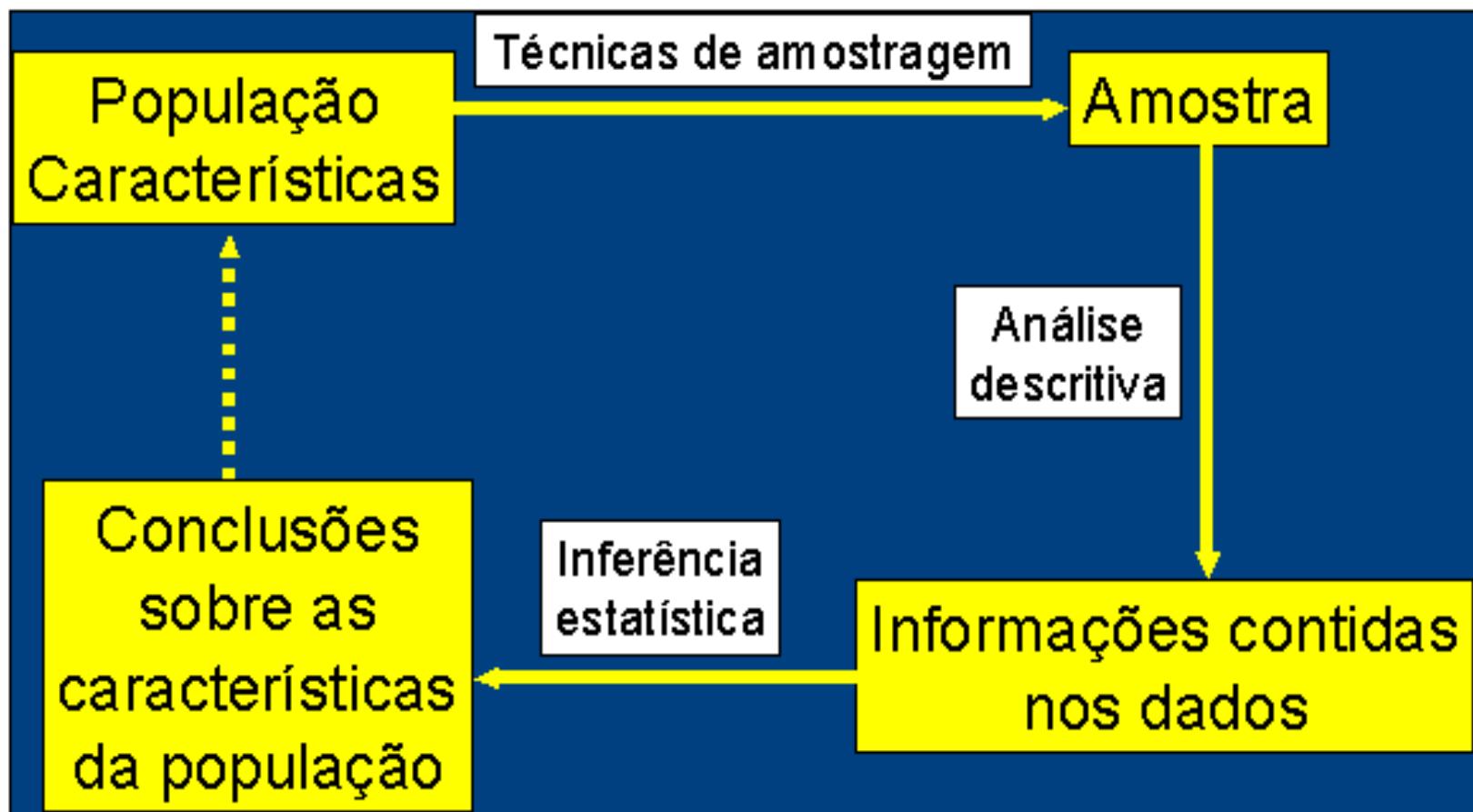


Estatística Descritiva

- A utilização da Estatística é cada vez mais acentuada em qualquer atividade profissional da vida moderna;
- A estatística está basicamente dividida em duas partes:
Estatística Descritiva e Estatística Inferencial:
 - ✓ **Descritiva ou Exploratória** é relacionada a descrição por tabelas, gráficos e medidas sobre a amostra;
 - ✓ **Inferencial** tem a capacidade de estimar parâmetros de uma população baseado em uma amostra representativa.

Estatística Descritiva

Etapas da Analise Estatística

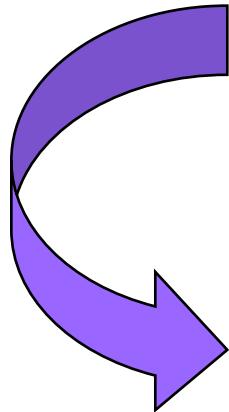
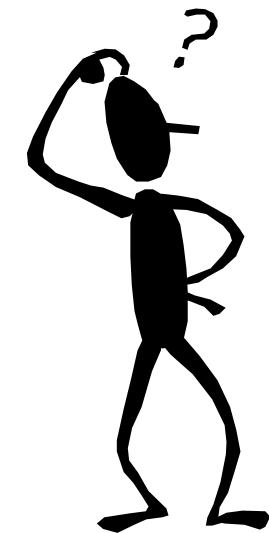


Estatística Descritiva

- A Análise Descritiva ou Exploratória de dados consiste em obter:
 - Medidas que resumem informações de um conjunto de dados quanto a posição central e dispersão (variabilidade);
 - Gráficos e tabelas que mostram a distribuição dos dados
 - Etapa inicial da análise utilizada para descrever, organizar e resumir os dados coletados.

Estatística Descritiva

O que fazer com as observações que coletamos?



Primeira Etapa:

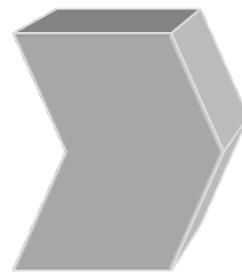
Resumo dos dados = Estatística descritiva

Natureza das Variáveis



Natureza das Variáveis

NOMINAL



Não pode ordenar:

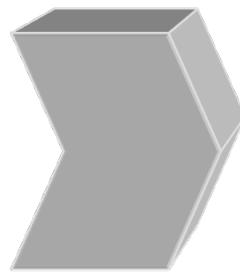
Sexo

Religião

Estado civil

Profissão

ORDINAL



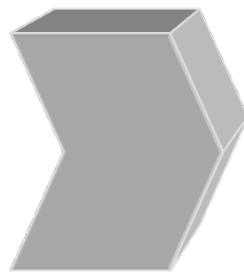
Pode ordenar:

Escolaridade

Nível sócio-econômico

Natureza das Variáveis

Discreta



Contáveis

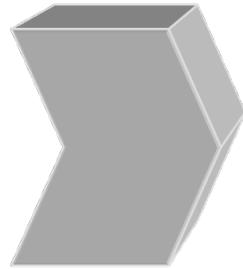
Nº de extrações

Nº de atendimentos

Idade

Nº de filhos

Contínua



Medidas

Altura

Peso

Temperatura

Pressão

Exercícios

1. Para as situações descritas a seguir, identifique a população e a amostra correspondente.
- Para avaliar a eficácia de uma campanha de vacinação no Estado de São Paulo, mães de recém-nascidos durante o primeiro semestre de 2005, foram perguntadas a respeito da última vez que vacinaram seus filhos;
 - Para verificar a audiência de um programa de TV no Brasil, indivíduos foram entrevistados com relação ao canal em que estavam sintonizados;
 - A fim de avaliar a intenção de voto para presidente do Brasil, pessoas foram entrevistadas em cidades brasileiras.

2. Classifique o tipo de variável para os itens abaixo.

- Marca de antitérmico preferida;
- Grau de satisfação com um produto alimentício;
- Peso de grãos exportados;
- Renda familiar;
- Grau de escolaridade;
- Número de computadores em um laboratório de informática.

3. Um questionário foi aplicado aos dez funcionários do setor de contabilidade de uma empresa, fornecendo os dados apresentados na tabela. **Pede-se:** Classifique cada uma das variáveis dada na tabela.

Funcionário	Curso (completo)	Idade (Anos)	Salário (R\$)	Anos de empresa
Abelardo	Superior	34	1100,52	5
Alberto	Superior	43	1452,12	8
Cristiano	Médio	31	965,57	6
Danilo	Médio	37	965,57	8
João	Fundamental	22	458,73	2
Maurício	Fundamental	21	458,73	3

4. Classifique cada uma das variáveis abaixo em qualitativa (nomial/ordinal) ou quantitativa (discreta/contínua).

- Intenção de voto para presidente (possíveis respostas são os nomes dos candidatos, além de não sei);
- Grau de satisfação de clientes com relação ao serviço prestado por uma empresa de consultoria (valores de 0 a 5, com 0 indicando totalmente insatisfeito e 5 totalmente satisfeito);
- Volume de petróleo extraído por hora de uma jazida.



Exercícios

Respostas

1. a)

População: Todas as mães de recém-nascidos, do Estado de São Paulo, que tiveram filhos no primeiro semestre de um dado ano.

Amostra: As mães de recém-nascidos entrevistadas na pesquisa.

1.b)

População: Todos os brasileiros que assistem TV.

Amostra: Os telespectadores entrevistados na pesquisa.

1.c)

População: Todos os eleitores votantes no Brasil.

Amostra: Os eleitores selecionados na pesquisa.

2. a) Qualitativa nominal; b) Qualitativa ordinal; c) Quantitativa contínua; d) Quantitativa contínua; e) Qualitativa ordinal; f) Quantitativa discreta

3. Funcionário: Qualitativa nominal; Curso: Qualitativa ordinal; Idade: Quantitativa contínua; Salário: Quantitativa contínua; Anos de empresa: Quantitativa contínua

4. a) Qualitativa nominal; b) Qualitativa ordinal; Quantitativa contínua

Natureza das Variáveis

Medidas de Posição

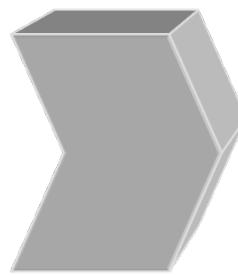
- Tendência Central: Média, Moda e Mediana
- Mínimo, Máximo
- Percentis, Quartis e Decis

Medidas de Dispersão

- Amplitude;
- Variância, Desvio Padrão;
- Coeficiente de Variação.

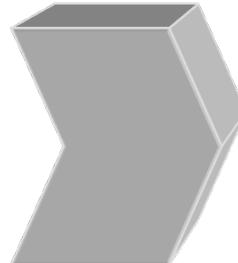
Natureza das Variáveis

**Numéricas ou
Quantitativas**
**Valores
expressos por
números**



Discreta

Contáveis
Nº de extrações
Nº de atendimentos
Idade
Nº de filhos



Natureza das Variáveis

Medidas de Posição

- **Máximo:** a maior observação
- **Mínimo:** a menor observação
- **Moda:** é o valor (ou atributo) que ocorre com maior frequência.

Ex,: Dados: 4, 5, 4, 6, 5, 8, 4

$$\max = 8 \quad \min = 4 \quad mo = 4$$

Natureza das Variáveis

Medidas de Posição: Média

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

Dados: 2, 5, 3, 7, 8

$$\bar{x} = \frac{2 + 5 + 3 + 7 + 8}{5} = 5$$

Natureza das Variáveis

Medidas de Posição: Mediana

A mediana é o valor da variável que ocupa a posição central de um conjunto de n dados ordenados.

Posição da mediana: $(n+1)/2$



Natureza das Variáveis

Medidas de Posição: Mediana

Dados: 2, 6, 3, 7, 8 $\Rightarrow n = 5$ (ímpar)

Dados ordenados: 2 3 6 7 8 \Rightarrow $5+1 = 3 \Rightarrow Md=6$

Posição da Mediana ↑

Dados: 4, 8, 2, 1, 9, 6 $\Rightarrow n = 6$ (par)

Dados ordenados: 1 2 4 6 8 9 $\Rightarrow \frac{6+1}{2} = 3,5$

$$Md = (4 + 6) / 2 = 5$$

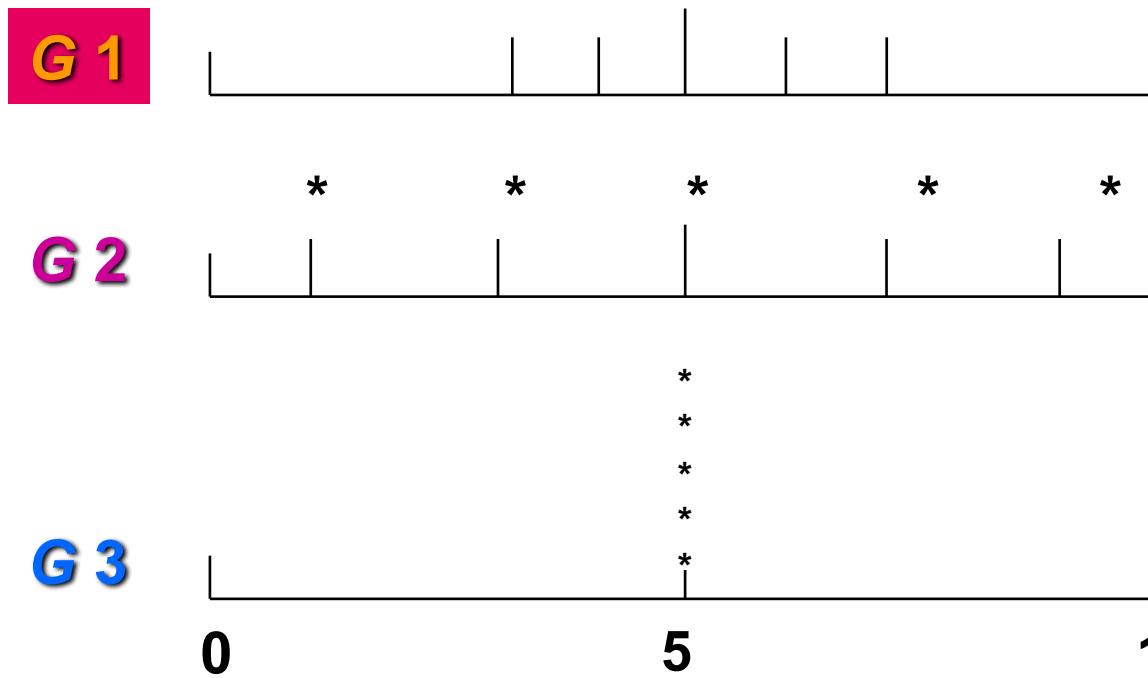
Natureza das Variáveis: Exemplo

Exemplo 2: Considere as notas de um teste em 3 grupos de alunos

Grupo 1: 3,4,5,6,7

Grupo 2: 1, 3, 5, 7, 9

Grupo 3: 5,5,5,5,5



Temos: $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 5$ e $md_1 = md_2 = md_3 = 5$

Natureza das Variáveis

Medidas de Dispersão : Separatrizes

A ideia de repartir o conjunto de dados pode ser levada adiante. Se ele for repartido em 4 partes tem-se os **QUARTIS**, se em 10 os **DECIS** e se em 100 os **PERCENTIS**.

Natureza das Variáveis

Por que Separatrizes ?

- A média e a moda podem não ser medidas adequadas para representar um conjunto de dados, pois:
 - São afetadas por valores extremos;
 - Apenas com esses dois valores não temos ideia da simetria da distribuição dos dados.

Para contornar esses fatos, consideramos outras medidas de posição.

Natureza das Variáveis

Quartis, Decis e Percentis

- São medidas de posição convenientes para comparar valores dentro de um mesmo conjunto de dados, ou entre conjuntos de dados diferentes.

Natureza das Variáveis

Quartis

- É uma medida de posição que divide as observações (ordenadas em ordem crescente) em quatro grupos;
- Dessa forma, temos três quartis denotados por Q_1 , Q_2 e Q_3 .

Natureza das variáveis

Quartis

- Q_1 - separa os 25% inferiores dos 75% superiores dos valores ordenados
- Q_2 - é a mediana
- Q_3 - separa os 75% inferiores dos 25% superiores dos valores ordenados

Natureza das Variáveis: Exemplo Quartis

Dados: 1,9 2,0 2,1 2,5 3,0 3,1 3,3 3,7 6,1 7,7

$$\Rightarrow n=10$$

Posição de Md : $0,5(n+1)= 0,5 \times 11 = 5,5 \Rightarrow Md = (3 + 3,1)/2 = 3,05$

Posição de $Q1$: $0,25 (11) = 2,75 \Rightarrow Q_1 = (2+2,1)/2 = 2,05$

Posição de $Q3$: $0,75 (11) = 8,25 \Rightarrow Q_3 = (3,7+6,1)/2 = 4,9$

$$Md = 3,05$$

$$Q_1 = 2,05$$

$$Q_3 = 4,9$$

Natureza das Variáveis: Exemplo Quartis

Dados: 0,9 1,0 1,7 2,9 3,1 5,3 5,5 12,2 12,9 14,0
33,6

Natureza das Variáveis: Exemplo Quartis

Dados: 0,9 1,0 1,7 2,9 3,1 5,3 5,5 12,2 12,9 14,0
33,6

$$\Rightarrow n=11$$

$$Md = 5,3$$

$$Q1 = 1,7$$

$$Q3 = 12,9$$

Natureza das variáveis

Decis

- É uma medida de posição que divide as observações em 10 grupos com cerca de 10% das observações em cada grupo;
- Denotamos os Decis por:

D_1 - 10%, D_2 - 20%, D_3 - 30%, D_4 - 40%, D_5 - 50%,
 D_6 - 60%, D_7 - 70%, D_8 - 80% e D_9 - 90%

Natureza das variáveis

Percentis

- Divide os dados em 100 grupos com cerca de 1% em cada grupo;
- Denotamos por:
 $P_1 - 1\%$, $P_2 - 2\%$, $P_3 - 3\%$, ... , $P_{99} - 99\%$.

Natureza das Variáveis

- As Medidas de Tendência Central:
 - representam de certa forma uma determinada distribuição de dados;
 - só elas não são suficientes para caracterizar a distribuição;

- Para uma análise estatística mais exata é necessária a verificação da flutuação dos valores em torno de sua média aritmética



Natureza das Variáveis

- Suponha as notas de 2 grupos de estudantes, cada qual com 5 alunos.

GRUPO “A” : 4, 5, 5, 6

GRUPO “B” : 0, 0, 10, 10

- Média do grupo “A”: 5
- Média do grupo “B”: 5

Natureza das Variáveis

- Os dois grupos apresentam a mesma média
- O comportamento dos 2 grupos são bem distintos

- **GRUPO “A”:** valores são mais homogêneo.
- **GRUPO “B”:** valores são dispersos em relação à média

Natureza das Variáveis

Medidas de Dispersão

- São utilizadas para caracterizar a variabilidade dos dados;
- ✓ Amplitude;
- ✓ Variância;
- ✓ Desvio padrão;
- ✓ Coeficiente de Variação (CV).

Quanto maior os valores maior é a dispersão dos dados.

Natureza das Variáveis

Medidas de Dispersão

Amplitude

- ✓ É a medida de dispersão ou variabilidade mais simples;
- ✓ É representada pela diferença entre o maior e o menor valor de um conjunto de observações:

Medidas de dispersão

Amplitude Total – At

É a diferença entre o maior e o menor valor observados.

$$At = X_{\max} - X_{\min}$$

■ Exemplo 5: Idade de 20 alunos:

X_i : 1, 1, 2, 2, 2, 3, 3, 3, 5, 5, 5, 5, 5, 6, 6, 6, 9

$$At = 9 - 1$$

$$At = 8$$

Natureza das Variáveis

Medidas de Dispersão

Amplitude Total – At

Tem o inconveniente de só levar em conta os dois valores extremos da série, descuidando do conjunto de valores intermediários, o que quase sempre invalida a idoneidade do resultado. Ela é uma indicação aproximada da dispersão ou variabilidade.

Natureza das Variáveis

Medidas de Dispersão

Variância:

A variância é a média aritmética do quadrado dos desvios de cada valor em relação à média.

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

(Para a população)

$$S^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n-1}$$

(Para a amostra)

Natureza das Variáveis

Medidas de Dispersão

Variância

Sendo a variância calculada a partir dos quadrados dos desvios, ela é um número em unidade quadrada em relação à variável em questão, o que sob o ponto de vista prático é um inconveniente.

Por isto imaginou-se uma nova medida que tem utilidade e interpretação prática, denominada desvio padrão.

Natureza das Variáveis

Medidas de Dispersão

Desvio Padrão

✓ Revela a dispersão do conjunto que se estuda

$$\sigma = \sqrt{\sigma^2}$$

(Para a população)

$$S = \sqrt{S^2}$$

(Para a amostra)

Natureza das Variáveis

Medidas de Dispersão

Variância e Desvio Padrão

Exemplo

12,4 22,6 18,9 9,7 14,5 22,5 26,3 17,7 31,2 20,4

$$S^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n-1}$$

$$S^2 = \frac{(12,4-19,62)^2 + (22,6-19,62)^2 + \dots + (20,4-19,62)^2}{10-1}$$

$$S^2 = 41,94$$

Logo:

$$\sqrt{41,94} = 6,476 \rightarrow \bar{x} = 22,6 \text{ e } S = 6,476$$

Medidas de dispersão

Desvio padrão:

- Se todos os valores forem iguais, o desvio padrão é nulo.
- quanto maior o desvio padrão mais heterogênea é a distribuição, significa que os valores são mais dispersos em torno da média
- $\text{MEDIA} \pm 1 \sigma \Rightarrow 68,26\%$ dos valores
- $\text{MEDIA} \pm 2 \sigma \Rightarrow 95,44\%$ dos valores
- $\text{MEDIA} \pm 3 \sigma \Rightarrow 99,74\%$ dos valores

Medidas de dispersão

Coeficiente de variação:

$$CV = \frac{S}{X}$$

S - desvio padrão
X - média aritmética

- o CV mede o grau de heterogeneidade da distribuição
- Valor máximo é $CV = 1$

$$0 \leq CV \leq 1$$



Medidas de dispersão

Coeficiente de variação:

Ao contrário do desvio padrão o coeficiente de variação não possui unidade, ou seja podemos comparar amostras medidas em unidades diferentes utilizando este parâmetro.

Medidas de dispersão

Exemplo:

Em um exame final de matemática a Média da nota de um grupo de 150 alunos foi de 7,8 e o desvio padrão 0,80. Em estatística, entretanto, a nota média foi 7,3 e o desvio padrão 0,76. Em que disciplina foi maior a dispersão. Justifique?

Exercícios

■ Considere os faturamentos mensais das seguintes filiais de uma grande empresa (em milhares de Reais)

- Filial A: 20 22 25 30 24 22 40 32 24 21 24 190
- Filial B: 35 39 37 38 42 40 45 41 33 37 40 47

Calcule: a média, mediana, moda, primeiro quartil e terceiro quartil, o desvio padrão e o coeficiente de variação.

Responda qual medida de posição é mais adequada para comparar o desempenho das filiais.

Dúvidas

