

UNIVERSIDADE FEDERAL DE PELOTAS
CENTRO DE DESENVOLVIMENTO TECNOLÓGICO

ESTATÍSTICA BÁSICA

Profa. Clause Fátima de Brum Piana

ANO 2020

ESTATÍSTICA BÁSICA

Conteúdo Programático

Unidade I. Introdução

Unidade II. Estatística Descritiva

Unidade III. Elementos de Probabilidade

Unidade IV. Inferência Estatística

Unidade I - Introdução

- ♦ Conceito, divisão e aplicações da Estatística
- ♦ Informações históricas
- ♦ Conceitos fundamentais
- ♦ Notação somatório
- ♦ Estatística na pesquisa científica

Questões relevantes do nosso cotidiano relacionadas à Estatística

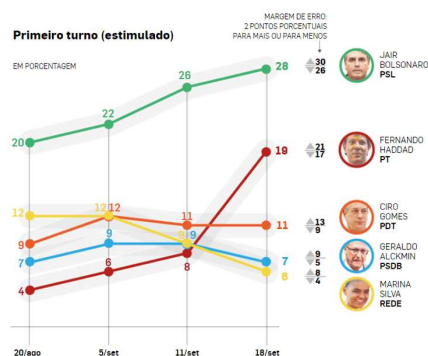
- ♦ Como verificar a eficácia de uma nova vacina?



Profa. Clause Piana

5

- ♦ Qual é a preferência eleitoral nas próximas eleições?



♦ Como saber se uma nova variedade de arroz é melhor do que as já existentes?



Profa. Clause Piana

7

♦ Como saber se um produto tem qualidade?



Linha de produção da Positivo



Linha de produção da Nestlé

Como verificar a qualidade da população de produtos resultantes de uma linha de produção?



Linha de produção de freios

Profa. Clause Piana

♦ Como estabelecer o valor do seguro?



Profa. Clause Piana

9

♦ Como estabelecer o grau de desenvolvimento de um país?



Indicadores econômicos e sociais dão idéia do grau de desenvolvimento de uma sociedade

Profa. Clause Piana

10

♦ Como se comportam os preços?



Profa. Clause Piana

11

♦ Como se caracterizam as pessoas que consomem drogas?



Profa. Clause Piana

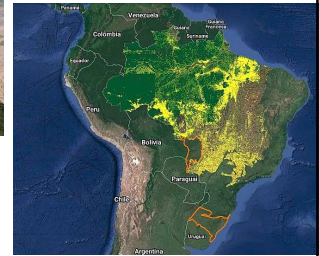
12

♦ Existem animais em extinção no Brasil?



www.geografiaparatos.com.br

♦ O desmatamento da Amazônia aumentou em 2020?



Profa. Clause Piana

14

Aspectos comuns a todas as questões

→ impossibilidade de examinar todos os casos, devido a diversos fatores:

- alto custo
- tempo excessivo
- desconhecimento dos efeitos (área médica)
- nem sempre todos os casos existem no momento da pesquisa
- coleta da informação por processo destrutivo

→ necessidade de uma metodologia que possibilite estudar alguns casos e generalizar as conclusões

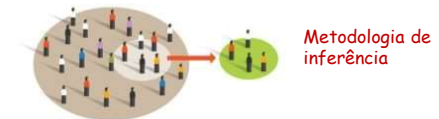
Metodologia de inferência

Aspectos comuns nas situações de pesquisa

→ Impossibilidade de examinar todos os casos

- alto custo
- tempo excessivo
- desconhecimento dos efeitos
- inexistência de todos os casos no momento da pesquisa
- coleta da informação por processo destrutivo

→ Necessidade de uma metodologia que possibilite estudar alguns casos e generalizar as conclusões



ESTATÍSTICA

- ♦ Sinônimo de dado numérico
- ♦ Planejamento, coleta, organização, resumo, apresentação e análise de dados
- ♦ Tomada de decisão na presença da incerteza provocada pela **aleatoriedade**
- ♦ Faz parte do **método científico**

Estratégia da Ciência para estudar problemas científicos

Compreender adequadamente a metodologia estatística implica em conhecer o contexto da pesquisa.

Matéria-prima da Estatística → variabilidade

Objetivo da Estatística → fornecer métodos para se conviver de modo racional com a variabilidade

De que forma?

Descobrir regularidade na variabilidade dos dados relativos às situações em estudo

⇒ A Estatística desenvolve métodos para descobrir e expor os padrões de comportamento (regularidade) que estão presentes nos dados, mas que não são facilmente identificáveis

Profa. Clause Piana

18

Duas idéias importantes:

- ⇒ Embora os dados sejam variáveis é sempre possível associar a eles a idéia de regularidade e expressar essa regularidade matematicamente
- ⇒ Devido à variabilidade inerente aos indivíduos, os pontos de interesse da Estatística são referentes aos grupos de indivíduos

Quando estudamos uma determinada característica, geralmente, queremos obter conclusões para o conjunto de todos os indivíduos que apresentam tal característica.

Profa. Clause Piana

19

Conceito

A Estatística é um corpo de conceitos e métodos úteis para coletar e interpretar dados relativos a uma área particular de pesquisa, e extrair conclusões em situações em que a incerteza e a variação estão presentes.

- ⇒ A Estatística parte de perguntas/desafios do mundo real
- ⇒ O que caracteriza a Estatística é a sua fundamentação nos dados de observação e, sendo assim, o contexto é essencial para dar significado aos resultados da aplicação dos procedimentos estatísticos.

Profa. Clause Piana

20

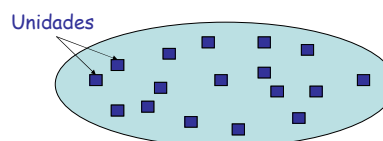
Alguns conceitos básicos:

- ◆ **População**
- ◆ **Amostra**
- ◆ **Amostragem**

Profa. Clause Piana

21

População: é o conjunto de todas as unidades (elementos) que têm pelo menos uma característica em comum. Pode ser finita ou infinita.



Exemplo: Pesquisas eleitorais no RS

Característica → votar no RS

População → conjunto de todos os eleitores que votam no RS

Unidade → o eleitor

Profa. Clause Piana

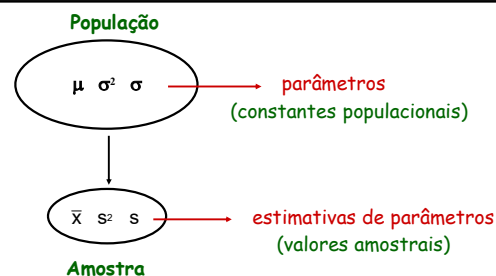
22

Amostra é parte de uma população, convenientemente escolhida, que tem a finalidade de representá-la. Deve apresentar as mesmas características da população.

Amostragem é a metodologia de obtenção das amostras.

Profa. Clause Piana

23

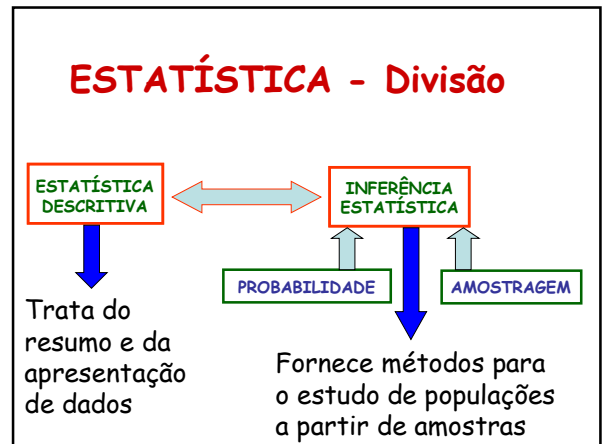
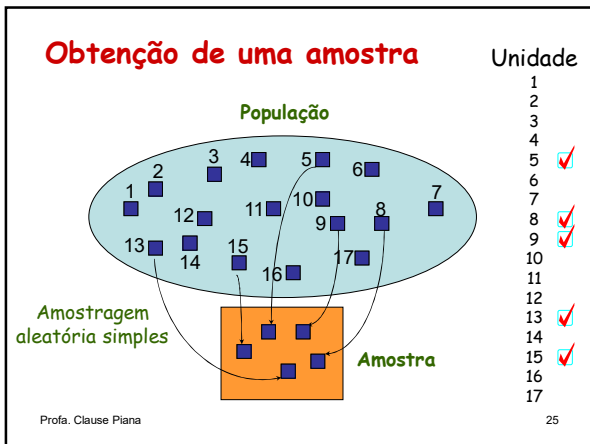


Parâmetros: são constantes que caracterizam uma população. Em geral, são valores desconhecidos. São representados por letras gregas.

Estimativas: são valores calculados na amostra para fornecer informação sobre os parâmetros, que estão sempre sujeitas a erro.

Profa. Clause Piana

24



Estatística Descritiva

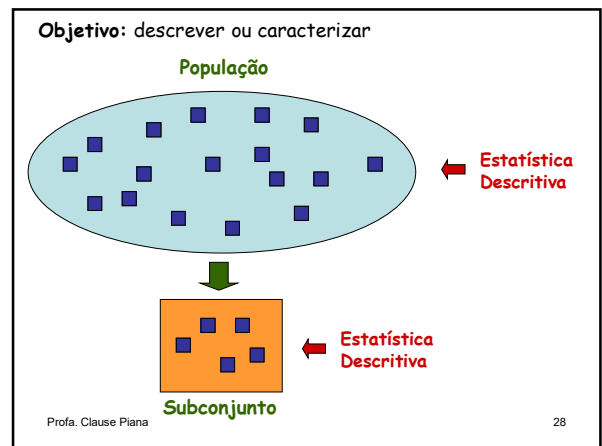
⇒ É a parte da estatística que trata:

- ♦ da apresentação de dados através de **tabelas e gráficos**;
- ♦ do resumo ou da descrição de dados através de **medidas descritivas** (de localização, de variação, de formato e separatrizes).

⇒ Em geral, não tem por objetivo tirar conclusões.

27

Profa. Clause Piana



Inferência Estatística ou Estatística Indutiva

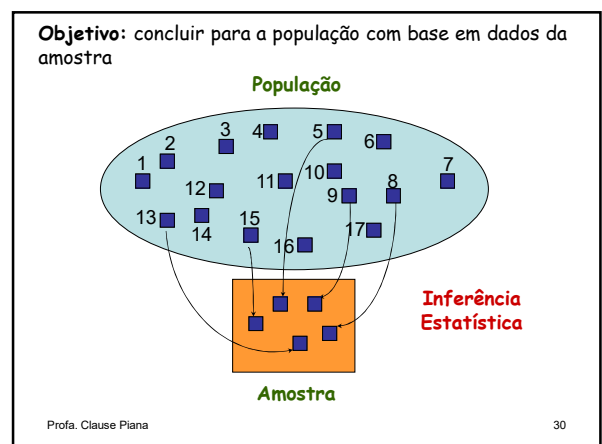
⇒ Provê métodos que possibilitam a realização de inferências sobre populações a partir de amostras delas retiradas, tendo por base o cálculo das probabilidades.

⇒ Trabalha basicamente com dois grandes tópicos:

- ♦ **estimação de parâmetros**
- ♦ **testes de hipóteses**

29

Profa. Clause Piana



Conjunto das temperaturas mínimas (em °C) do mês de setembro, no período 1990 a 1995, na cidade de Pelotas :

18,7 19,2 16,2 12,3 17,5 13,9

Se afirmamos que:

→ em metade **desses** anos a temperatura mínima do mês de setembro em Pelotas superou 17°C

→ **nesse** período de seis anos a temperatura mínima média do mês de setembro em Pelotas foi de 16,3°C

Estatística Descritiva

Se afirmamos que:

→ com base **nesses** dados, a temperatura mínima média do mês de setembro em Pelotas é de 16,3°C

Inferência Estatística

Profa. Clause Plana

31

Aplicações

⇒ em várias áreas da ciência e da tecnologia

Áreas mais específicas: Bioestatística, Psicometria, Estatística de negócios, Econometria, Demografia e Epidemiologia

⇒ em muitas atividades que afetam as nossas vidas

- decisões econômicas (públicas ou privadas)
- avaliação de controles de poluição
- análise de problemas de tráfego
- estudo dos efeitos de novos medicamentos
- estudo de novas técnicas e cultivares na agricultura

O argumento estatístico fornece critérios para determinar as conclusões que são realmente suportadas pelos dados e aquelas que não o são. Em todos os campos de estudo onde inferências são extraídas de análises de dados, a credibilidade das conclusões também depende grandemente do uso de métodos estatísticos adequados, especialmente no estágio da coleta de dados.

Informações históricas

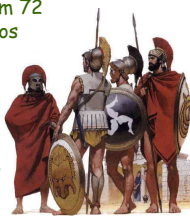
⇒ Existem indícios de que há mais de 2000 anos a.C. já se faziam censos na Babilônia, na China e no Egito.



⇒ Em 2238 a.C., o imperador Yao mandou realizar um censo da população e das lavouras cultivadas...

⇒ Os romanos fizeram 72 censos entre os anos 555 a.C. e 72 d.C.

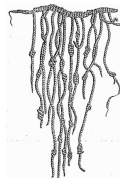
⇒ Objetivo do censo: saber o número de pessoas disponíveis para fazer a guerra e para a cobrança de impostos



⇒ O livro quarto do Velho Testamento, intitulado "Números", começa com a seguinte instrução a Moisés: "Fazer um levantamento dos homens de Israel que estivessem aptos para guerrear".

⇒ Na época do Imperador César Augusto, saiu um edito para que se fizesse o censo em todo o Império Romano

- ♦ José e Maria saíram de Nazaré, na Galiléia, para Belém, na Judéia



⇒ Continente Americano → muito antes de Cristóvão Colombo, os Incas já mantinham um registro numérico de dados da população em quipus (sistema de cordas com nós que representavam números no sistema decimal)

34

⇒ A Estatística teve origem na necessidade do Estado Político conhecer os seus domínios.

⇒ Em 1085, Guilherme I, o Conquistador, ordenou que se fizesse um levantamento estatístico da Inglaterra (com informações sobre terras, proprietários, empregados e animais). Esses dados originaram um livro chamado "Domesday Book" que serviu de base para o cálculo de impostos.



⇒ O termo *estatística* surge da expressão em latim *statisticum collegium* (palestra sobre os assuntos do Estado) de onde surge a palavra alemã *Statistik*, designando a análise de dados sobre o Estado.

Século XVII → a aritmética política (demografia) ganhou destaque na Inglaterra a partir das tábuas de mortalidade de **Graunt** e **Petty**. Não se limitaram a simples apresentação de dados e estabeleceram relações entre esses números.



Jonh Graunt
(1620 - 1674)

Tábuas de mortalidade:
exaustivas análises de nascimentos e mortes

♦ Morte é um fato imprevisível quando considera um indivíduo isolado, porém em um grupo é possível predizê-la com uma certa margem de segurança.



William Petty
(1623 - 1687)

Profa. Clause Plana

36

1708 → primeiro curso de Estatística na Universidade de Jena, na Alemanha

1740 → a palavra "Statistik" (Estatística) foi cunhada pelo acadêmico alemão **Gottfried Achenwall**, que estabeleceu as relações da Estatística com outras áreas.



1797 → o verbete "Statistics" (Estatística) apareceu na Enciclopédia Britânica

Século XIX → o avanço no estudo das probabilidades permitiu que a Estatística se estruturasse e ampliasse largamente o seu campo de ação.

Profa. Clause Piana

37

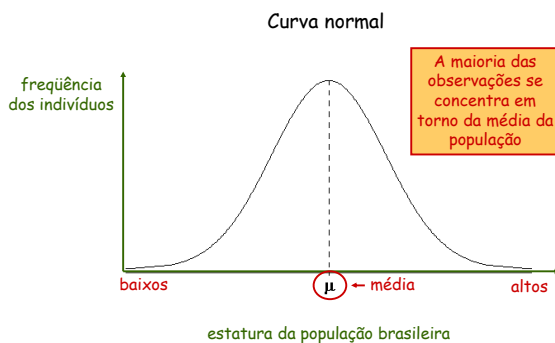
⇒ Esse avanço se deu a partir dos estudos dos matemáticos e astrônomos **Laplace** e **Gauss** → descreveram a famosa **distribuição normal**.



Pierre-Simon Laplace
(1749-1827)



Carl Gauss
(1777-1855)



Profa. Clause Piana

39



Profa. Clause Piana

40

Século XIX → astrônomo belga **Lambert Quételet** fez o casamento entre a Estatística e o Cálculo das probabilidades

- ♦ estudos na área social → demonstrou que muitos fenômenos vivos tinham um comportamento regular



Lambert Quételet
(1796-1855)

Século XX

- ⇒ Distribuições de probabilidade → criação de técnicas de amostragem e formas de relacionar amostras com populações
- ⇒ Desenvolvimento da **Estatística Experimental**
- ⇒ Marco decisivo → advento da **computação eletrônica**



Ronald Fisher
(1890 - 1962)

Conceitos fundamentais

- ♦ Características e variáveis
- ♦ Escalas de medida
- ♦ Classificação de variáveis
- ♦ Observação e conjunto de dados
- ♦ Notação somatório

Profa. Clause Piana

42

Características e Variáveis

- ♦ **Característica** é uma propriedade qualquer das unidades de uma população.
- ♦ Não há interesse nas unidades em si e sim nas suas características.
- ♦ **Variável** é uma característica para a qual foi definida uma **escala de medida**.
- ♦ Características são analisadas tendo por base os valores das variáveis a elas associados.

Profa. Clause Plana

43

Por exemplo, para um grupo de alunos de uma escola pode-se medir as seguintes características: **sexo**, **escolaridade dos pais**, **temperatura corporal** e **peso**.

A definição da forma de mensuração da característica gera uma variável. Por exemplo:

- a) **Sexo**, entre as opções {F e M}
- b) **Escolaridade dos pais**, entre as opções {EFI, EFC, EM, ES}
- c) **Temperatura**, em °C
- d) **Peso**, em kg

Variável é uma função que associa um número a cada alternativa de uma característica.

- a) **Sexo**, entre as opções {0-F e 1-M}
- b) **Escolaridade dos pais**, entre as opções {1-EFI, 2-EFC, 3-EM, 4-ES}
- c) **Temperatura**, em °C, [35; 40]
- d) **Peso**, em kg, [40; 60]

Esses números podem ser expressos em quatro **escalas de medida** distintas.

Escalas de medida

Nominal: os números são apenas rótulos ou nomes

Ordinal: os números possibilitam apenas ordenação

De intervalo (intervalar): algumas operações são possíveis, mas o conjunto de números não possui zero verdadeiro

De razão (racional): todas as operações são possíveis e o conjunto de números possui zero verdadeiro

Quantidade de informação das escalas de medida

- a) **Sexo** → pode-se apenas **distinguir** os alunos → **nominal**
- b) **Escolaridade dos pais** → pode-se **distinguir** e **ordenar** os alunos e dizer que os pais de um têm instrução superior aos pais de outro → **ordinal**
- c) **Temperatura corporal** → pode-se **distinguir** e **ordenar** os alunos e dizer que um aluno está com temperatura superior ao outro em tantos graus → **intervalar**
- d) **Peso** → pode-se **distinguir** e **ordenar** os alunos, dizer que o peso de um supera o do outro em tantos quilos e que o peso de um aluno é metade do peso do outro aluno → **de razão**

Escala nominal

Exemplos

- Sexo de estudantes (0 - Masculino, 1 - Feminino);
- Partido político de cada candidato (0 - Esquerda, 1 - Direita, 2 - Centro);
- Respostas do tipo sim ou não (0 - Sim, 1 - Não);
- Nomes de cidades (0 - Pelotas, 1 - Rio Grande, 2 - Porto Alegre).

Dados expressos em escala nominal



⇒ Não é possível estabelecer qualquer relação de ordem entre eles
⇒ Números não têm nenhum significado para efeito de cálculos. Apenas a contagem por categoria faz sentido.

Profa. Clause Plana

48

Escala ordinal

Exemplos

- Um editor de livros classifica alguns originais como: 1- excelente, 2- bom e 3- ruim;
- Classificação de alunos num teste de estatística: 1- ruim, 2- regular, 3- bom e 4- muito bom;
- Faixa etária: 0- criança, 1- jovem, 2- adulto e 3- idoso.

Dados expressos em escala ordinal

- ⇒ É possível associar a eles valores que representam as ordens
- ⇒ Diferenças entre valores não podem ser determinadas, ou não têm sentido

Profa. Clause Piana

49

Escala intervalar

Exemplos

- temperatura (graus Celsius, Fahrenheit)
O zero nesta escala é apenas uma convenção. O dobro da temperatura em °C não corresponde ao dobro da temperatura em °F.
- contagem do tempo
A contagem do tempo é feita a partir de uma escala em que o zero é arbitrário. O ocidente, a China, os judeus utilizam calendários diferentes.

Dados expressos em escala intervalar

- ⇒ Inclui a escala ordinal; permite determinar as diferenças entre os dados
- ⇒ Não possui um ponto de partida (zero verdadeiro) inerente ou natural

Profa. Clause Piana

50

Escala racional

Exemplos

- Pesos e estaturas
- Distância percorrida em km
- Duração de filmes

Dados expressos em escala racional

- ⇒ Inclui a escala intervalar, com a propriedade adicional de ter um zero verdadeiro (que significa ausência de quantidade)
- ⇒ Tanto as diferenças como as razões têm significado

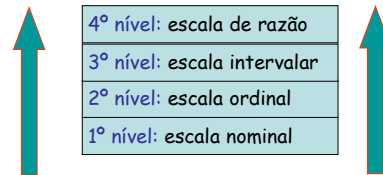
Profa. Clause Piana

51

Escalas de medida

Operações realizadas

Metodologia Estatística



A complexidade e a informação aumentam com o nível

Profa. Clause Piana

52

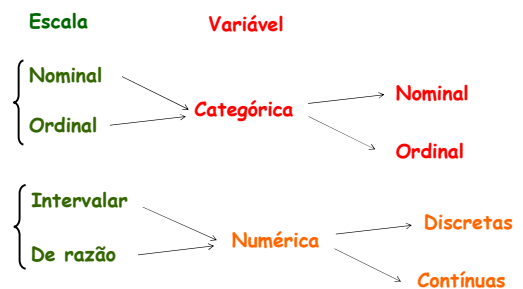
Classificação de variáveis



Profa. Clause Piana

53

As escalas nominal e ordinal geram variáveis catégoricas e as escalas intervalar e de razão geram variáveis numéricas.



Variáveis categóricas (fatores)

- seus valores representam categorias ou classes
- possuem um conjunto limitado de valores que se repetem para todas as observações
- cada valor da variável é denominado nível

Profa. Clause Piana

55

Variáveis categóricas qualitativas → descrevem qualidades, ou seja, as categorias se distinguem qualitativamente

Nominais → não há um sentido de ordem entre seus níveis

Exemplos: sexo (masculino e feminino)
raça de cavalos (manga-larga, crioulo e árabe)
região geográfica (norte, sul e sudeste)
estado civil (solteiro, casado e divorciado)
estado do tempo (bom, chuvoso, nublado)

Ordinais → há um sentido de ordem entre seus níveis

Exemplos: faixas de idade (criança, adolescente, adulto, idoso)
intensidade de cor (claro, médio, escuro)
nível de instrução (primário, secundário, universitário)
nível de armazenamento de um reservatório (muito alto, alto, médio, baixo e muito baixo).

Profa. Clause Piana

56

Variáveis categóricas quantitativas → descrevem quantidades (a cada nível está associado um valor numérico), ou seja, a diferença entre as categorias é quantitativa

Exemplo: Um experimento pretende comparar doses de um tranquilizante usado contra insônia

Variável: Dose de tranquilizante

Dose 1 - 0,5 mg
Dose 2 - 1,5 mg
Dose 3 - 2,5 mg

categoria
ou nível

valor
numérico

Profa. Clause Piana

57

Variáveis numéricas → seus valores são números reais

Discretas

→ descrevem dados discretos ou de enumeração (obtidos por **processo de contagem**)

→ assumem valores inteiros não negativos (0, 1, 2, 3, ...)

Exemplos: número de sementes germinadas
número de pacientes que se recuperaram
número de filhos de um casal

Contínuas

→ descrevem dados contínuos ou de mensuração (obtidos por **processo de medição**)

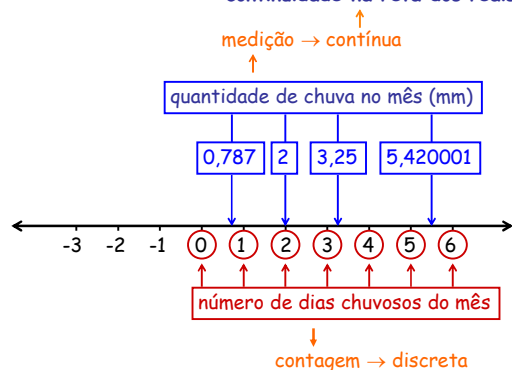
→ podem assumir qualquer valor dos reais (-10, 0, π)

Exemplos: peso, altura, tempo, teor de umidade, temperatura

Profa. Clause Piana

58

continuidade na reta dos reais



Profa. Clause Piana

59

Por que classificar as variáveis?

É de fundamental importância saber classificar corretamente uma variável porque esta discriminação é que irá indicar a possibilidade e a forma de utilização dos procedimentos estatísticos disponíveis.

Profa. Clause Piana

60

Exercício proposto:

Classifique as variáveis e identifique suas escalas de medida.

- > Número de vendas diárias de uma empresa
- > Vazão de uma bacia hidrográfica ($m^3.s^{-1}$)
- > Consumo de gasolina de um carro ($km.L^{-1}$)
- > Consumo de gasolina de um carro (baixo, médio, alto)
- > Grau de escolaridade
- > Número de dias sem chuva em um dado local
- > Temperatura de uma mistura (em $^{\circ}C$)
- > Modelo de um carro
- > Velocidade de um carro ($km.s^{-1}$)
- > Postos hierárquicos em um exército
- > Número de peças com defeito em um lote
- > Valor de venda diária de uma empresa (em R\$)

61

Unidade de observação → indivíduo onde são medidas as variáveis de interesse

Exemplos: aluno, planta, animal

Observação → conjunto dos valores de todas as variáveis medidas em uma unidade de observação

Exemplo: peso, estatura e número de irmãos de um estudante

Conjunto de dados → conjunto de todas as observações

Profa. Clause Piana

62

Um conjunto de dados com: 17 unidades ou observações, 1 variável identificadora, 1 variável categórica, 1 variável numérica discreta e 3 variáveis numéricas contínuas

i	Nome	Sexo	Número de irmãos	Idade	Estatura	Peso
1	Alfredo	M	2	14	1,75	51,03
2	Carol	F	2	14	1,60	46,49
3	Jane	F	1	12	1,52	38,33
4	João	M	0	12	1,50	45,13
5	Luísa	F	2	12	1,43	34,93
6	Roberto	M	4	12	1,65	58,06
7	William	M	1	15	1,69	50,80
8	Bárbara	F	2	13	1,66	44,45
9	Juca	M	3	12	1,46	37,65
10	Joca	M	3	13	1,59	38,10
11	Judite	F	1	14	1,63	40,82
12	Felipe	M	0	16	1,83	68,04
13	Tomas	M	5	11	1,46	38,56
14	Alice	F	4	13	1,44	38,10
15	Henrique	M	1	14	1,61	46,49
16	Janete	F	2	15	1,59	51,03
17	Joice	F	3	11	1,30	22,91

Profa. Clause Piana

63

Representação de variáveis

Variáveis → **letras maiúsculas** (X, Y, Z)

Valores da variável (dados) → **letras minúsculas** (x, y, z)

Exemplo: Se uma variável é representada por **X (xis maiúsculo)**, todos os seus valores serão representados por **x (xis minúsculo)**

Diferenciação dos valores da variável → **acrescenta-se um índice i ao x**

x_i → **conjunto de valores**

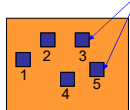
Se $i = 1, 2, \dots, n$, então, $x_i = x_1, x_2, x_3, \dots, x_n$

Profa. Clause Piana

64

Valores de variáveis nas amostras

Amostra



Unidades $i \in U = \{1, 2, 3, 4, 5\}$

x_i : valor da variável X na unidade i

y_i : valor da variável Y na unidade i

z_i : valor da variável Z na unidade i

Variáveis medidas (X, Y e Z)

→ $\left\{ \begin{array}{l} \{x_1, x_2, x_3, x_4, x_5\} \\ \{y_1, y_2, y_3, y_4, y_5\} \\ \{z_1, z_2, z_3, z_4, z_5\} \end{array} \right.$

Profa. Clause Piana

65

Notação somatório

i é o número da observação, tal que $i = 1, 2, \dots, n$

n é o número total de observações

x_i é o valor da variável X para a observação i, tal que $x_1 = 1, x_2 = 0, \dots, x_5 = 4$

y_i é o valor da variável Y para a observação i, tal que $y_1 = 2, y_2 = 1, \dots, y_5 = 0$

$x_{(i)}$ é o valor da variável X para a observação i, tal que $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

$x_{(1)}$ é o menor valor da variável X

$x_{(n)}$ é o maior valor da variável X

Profa. Clause Piana

66

Somatório (Σ)

⇒ indica a soma sequencial de um conjunto de valores

$$x_1 + x_2 + x_3 + x_4 + x_5 = \sum_{i=1}^5 x_i$$

limite superior da soma

$$\sum_{i=1}^n x_i$$

somatório →

← Valores que estão sendo somados

← número das observações

limite inferior da soma

Profa. Clause Piana

67

A notação $\sum_{i=1}^n$ inclui todos os valores do intervalo e pode ser simplificada por Σ

$$\sum_{i=1}^n = \Sigma$$

Outra forma de notação da soma:

$$\sum_{i=1}^n x_i = \Sigma x_i = X_+$$

$$\sum_{i=1}^n y_i = \Sigma y_i = Y_+$$

Profa. Clause Piana

68

Algumas quantidades de interesse:

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 = \sum_{i=1}^5 x_i^2$$

$$(x_1 + x_2 + x_3 + x_4 + x_5)^2 = \left(\sum_{i=1}^5 x_i \right)^2$$

$$x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4 + x_5 y_5 = \sum_{i=1}^5 x_i y_i$$

$$(x_1 + x_2 + x_3 + x_4 + x_5) \cdot (y_1 + y_2 + y_3 + y_4 + y_5) = \sum_{i=1}^5 x_i \cdot \sum_{i=1}^5 y_i$$

Profa. Clause Piana

69

11. Desenvolva os seguintes somatórios:

$$\sum_{j=6}^{10} f_j^2 =$$

$$\sum_{j=3}^6 (x_j + c) =$$

$$\sum_{i=1}^4 (x_i + 2)^2 f_i =$$

13. Represente, usando a notação Σ , as seguintes expressões:

$$5x_1^2 + 5x_2^2 + \dots + 5x_{10}^2 =$$

$$x_1 + y_1 + x_2 + y_2 + \dots + x_6 + y_6 =$$

$$w_1 + w_2 + w_3 + w_4 + k =$$

$$x_1 y_1 + x_2 y_2 + \dots + x_n y_n =$$

Profa. Clause Piana

70

Propriedades da soma

1ª propriedade: A soma é distributiva, ou seja, se cada termo da soma é multiplicado por uma constante c , os termos podem ser somados e a soma multiplicada pela constante.

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

$$= c(x_1 + x_2 + \dots + x_n)$$

$$= c \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n c x_i = c \sum_{i=1}^n x_i$$

Profa. Clause Piana

71

2ª propriedade: A soma de uma constante c sobre n termos é igual a n vezes esta constante.

$$\sum_{i=1}^4 c = c + c + c + c$$

$$= c(1+1+1+1)$$

$$= 4c$$

$$\sum_{i=1}^n c = nc$$

Profa. Clause Piana

72

3ª propriedade: A soma é associativa, ou seja, o somatório da soma (ou diferença) é igual a soma (ou diferença) de somatórios.

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

$$(x_1 + y_1) + (x_2 + y_2) + \dots + (x_n + y_n) = (x_1 + x_2 + \dots + x_n) + (y_1 + y_2 + \dots + y_n)$$

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

As propriedades devem ser usadas no sentido de simplificar as operações. Sempre que houver uma operação que precede a soma, devemos desenvolvê-la antes de aplicar as propriedades.

Exemplo:

$$\begin{aligned} \sum (x_i - c)^2 &= \sum (x_i^2 - 2x_i c + c^2) \leftarrow \text{3ª propriedade} \\ &= \sum x_i^2 - \sum 2x_i c + \sum c^2 \leftarrow \text{1ª propriedade} \\ &= \sum x_i^2 - 2c \sum x_i + \sum c^2 \leftarrow \text{2ª propriedade} \\ &= \sum x_i^2 - 2c \sum x_i + nc^2 \end{aligned}$$

Simplifique os seguintes somatórios, utilizando as propriedades da soma.

a) $\sum_{i=1}^5 2x_i =$

b) $\sum_{j=3}^6 (x_j + c) =$

c) $\sum_{i=1}^4 (x_i + 2)^2 f_i =$

d) $\sum_{i=3}^5 \frac{1}{4} x_i y_i =$

e) $\sum_{i=2}^4 (x_i + 3y_i)^2 + \sum_{i=2}^4 y_i =$

A partir do conjunto de dados, calcule os seguintes somatórios:

i	x _i	y _i
1	1	2
2	0	1
3	2	-2
4	-1	1
5	4	0

a) $\sum_{i=3}^5 x_i^2 =$

b) $\left(\sum_{i=3}^6 x_i \right)^2 =$

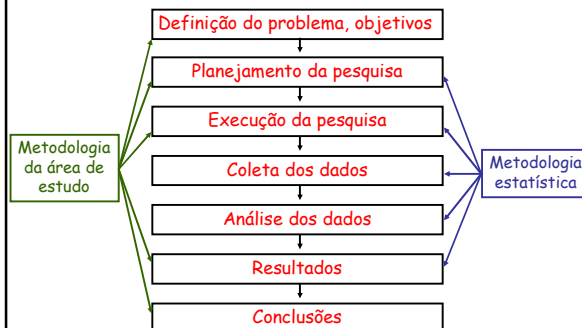
c) $\sum_{i=1}^3 x_i y_i =$

d) $\sum_{i=1}^3 x_i \sum_{i=1}^3 y_i =$

A estatística na pesquisa científica

- ⇒ Nas pesquisas científicas precisamos coletar dados que possam fornecer informações que respondam nossas indagações
- ⇒ Para que os resultados da pesquisa sejam confiáveis, tanto a coleta como análise dos dados devem ser feitas de forma criteriosa e objetiva
- ⇒ A metodologia estatística deve ser aplicada nas diversas etapas da pesquisa
- ⇒ A qualidade da informação depende da qualidade dos dados

Principais etapas de uma pesquisa científica



Métodos de pesquisa

⇒ Pesquisa em que as características de interesse são inerentes às unidades e se manifestam sem interferência do pesquisador.

Censo: todas as unidades da população são observadas.

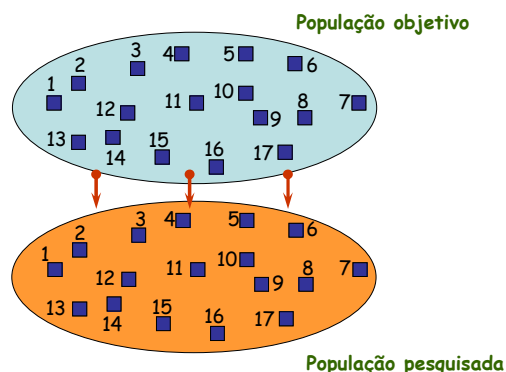
Levantamento por amostragem: as unidades da amostra podem ser escolhidas por processo aleatório (sorteio) ou não.

Estudo observacional: as unidades são incluídas no estudo segundo as circunstâncias. Comum na medicina e nas ciências sociais.

⇒ Pesquisa em que há intervenção do pesquisador na amostra.

Experimento: Os tratamentos são impostos às unidades da amostra com propósito de avaliar seus efeitos comparativamente.

Censo



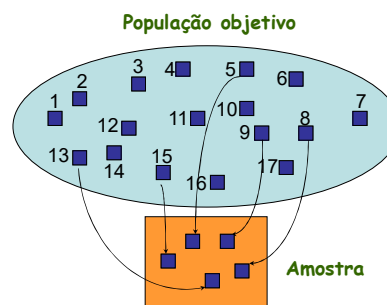
Censo

Tabela 1 . Evolução da população brasileira, de 1872 a 2010.

Ano	População
1872	9.930.478
1890	14.333.915
1900	17.438.434
1920	30.635.605
1940	41.165.289
1950	51.941.767
1960	70.070.457
1970	93.139.037
1980	119.002.706
1991	146.825.575
2000	169.799.710
2010	190.755.799

Fonte: Censos demográficos diversos

Levantamento por amostragem



IBGE → Pesquisa Nacional por Amostra de Domicílios

Levantamento por amostragem

Pesquisa Nacional por Amostra de Domicílios (PNAD)

- Entrevista 100.000 domicílios por ano (quase 500.000 pessoas)
- Obtém informações **anuais** sobre características demográficas e socioeconômicas da população, como **sexo, idade, educação, trabalho e rendimento, e características dos domicílios**,
- Com periodicidade variável, obtém informações sobre migração, fecundidade, nupcialidade, entre outras, tendo como unidade de coleta os domicílios.
- Fornece o retrato mais completo do Brasil.

Profa. Clause Piana

83

Quadro 1 - Principais características e variáveis levantadas pela Pnad

Características	Variáveis
Demográficas e sociais	Sexo, cor, condição na unidade domiciliar, posição na família e no domicílio, número na família e data de nascimento dos moradores.
Educacionais	Alfabetização, escolaridade (série e grau frequentados) e nível de instrução das pessoas que não são estudantes (última série concluída e grau correspondente).
Mão-de-obra	Para as pessoas de 10 anos de idade ou mais: condição de atividade. Para as pessoas ocupadas: ocupação, atividade e posição na ocupação no trabalho principal, horas normalmente trabalhadas por semana no trabalho principal e nos outros trabalhos, e se é contribuinte de instituto de previdência pelo trabalho. Para as pessoas desocupadas: tempo de procura de trabalho, ocupação, atividade, posição na ocupação e motivo da saída, se recebeu fundo de garantia, e tempo de permanência em relação ao último trabalho remunerado.
Rendimento	Rendimento mensal normalmente recebido do trabalho principal e dos outros trabalhos, aposentadoria, pensão, abono de permanência, aluguel e outros rendimentos.
Habitação	Espécie de domicílio. Para os domicílios particulares permanentes: tipo, estrutura, abastecimento de água, esgotamento sanitário, uso de instalação sanitária, destino do lixo, iluminação elétrica, número de cômodos, condição de ocupação, aluguel ou prestação mensal, filtro de água, fogão, geladeira, rádio e televisão.

Fonte: Fundação IBGE. Para Compreender a Pnad, 1991.

Estudo observacional

Coorte de nascidos em Pelotas

- Conduzida pelo **Centro de Pesquisas Epidemiológicas** da Universidade Federal de Pelotas.
- A investigação monitorou todas as crianças nascidas vivas na cidade de Pelotas, nos anos de 1982 (5.914 pessoas), 1993 (5.249 pessoas) e 2004 (4.231 pessoas).
- Nas três fases do estudo, foram realizadas visitas no terceiro e 12º mês de idade. Eram aplicados questionários às mães, além de medidas e avaliações do desenvolvimento das crianças.
- Os participantes continuam sendo acompanhados até o presente, e questões como mortalidade infantil, desnutrição, crescimento, sexualidade, gravidez na adolescência e hipertensão têm sido exploradas na pesquisa.
- Este estudo é a coorte de nascimentos com seguimento regular mais longo em um país em desenvolvimento.

Profa. Clause Piana

87

Experimento

Comparação de duas linguagens de programação

- Um pesquisador da área de computação está investigando a utilidade de duas diferentes linguagens de programação (A e B) na melhoria das tarefas computacionais.
- Trinta programadores experientes, familiarizados com ambas as linguagens, são divididos aleatoriamente em dois grupos.
- Cada grupo codificará uma função padrão em uma das linguagens e os tempos de codificação da função (em minutos) serão registrados.

Profa. Clause Piana

86

Limitações da Estatística

- ⇒ A Estatística não corrige erros grosseiros e técnicas defeituosas. Como toda informação está contida nos dados, se esses dados forem falsos ou viciados, será falsa qualquer conclusão retirada deles
- ⇒ Muitos procedimentos dependem da validade dos modelos escolhidos e de suas pressuposições
- ⇒ Alguns procedimentos possibilitam estabelecer associação entre variáveis, mas não relação de causa e efeito
- ⇒ A estatística não substitui o julgamento crítico do pesquisador

Profa. Clause Piana

87

Bibliografia

- FERREIRA, D.F. **Estatística básica**. Lavras: Editora UFLA, 2005.
- MONTGOMERY, D.C.; RUNGER, G.C.; HUBELE, N.F. **Estatística Aplicada à Engenharia**. 2 ed. Rio de Janeiro: Editora LTC. 2004. 335p.
- SILVEIRA JUNIOR, P.; MACHADO, A.A.; ZONTA, E.P.; SILVA, J.B. da. **Curso de Estatística** v.1. Pelotas: Universidade Federal de Pelotas, 1992, 135p.
- Sistema Galileu de Educação Estatística**. Disponível em: <http://www.galileu.esalq.usp.br/topico.html>

Profa. Clause Piana

88