



Universidade Federal do Oeste do Pará
Instituto de Biodiversidade e Florestas
Disciplina: Estatística Básica
Professora: Lia Melo

Capítulo 1

INTRODUÇÃO À ESTATÍSTICA

1. Introdução: Histórico

3.000 anos A.C. - Censos na Babilônia, China e Egito
CENSO - deriva de “CENSERE” - Taxar (latim)

ESTATÍSTICA - “ESTATUS” (Estado em latim),
- Descrições e dados relativos ao Estado
- Ferramenta administrativa

Século XVIII – A palavra Estatística foi cunhada pelo acadêmico alemão Gottfried Achenwall que determinou seus objetivos e relações com as outras ciências

O verbete “Statistics” (Estatística) apareceu na Enciclopédia Britânica em 1797

2. Alguns Conceitos Modernos:

- Estatística é uma parte da matemática aplicada que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados visando auxiliar a tomada de decisões.
- É o estudo de como chegar a conclusões sobre o todo (população), partindo da observação de partes desse todo (amostras);
- É a ciência que tem por objetivo o estudo quantitativo das populações;
- É um método científico empregado no estudo das variações de fenômenos aleatórios

Estatística

- *Método científico que determina questões de pesquisa;*
- *Projeta estudos e experimentos;*
- *Coleta, organiza, resume e analisa dados;*
- *Interpreta resultados*
- *Esboça conclusões*

Exemplos do uso da estatística no seu dia a dia

- Índices de audiência: uma emissora pode tirar um programa do ar ou modificar o fim da novela;

- Taxas de crescimento de vegetais ou animais: um agricultor pode utilizar este ou aquele adubo ou ração.
- Índices de acidentes de trânsito: leis restritivas, medidas de prevenção.
- Taxas de crescimento de câncer de mama
- Expectativa de vida de quem fuma

Procedimentos quantitativos chamados métodos estatísticos permitem a um pesquisador envolver-se em três importantes atividades

1. Descrever qualitativamente e resumir características de conjuntos de dados (população total ou amostras)
2. Extrair conclusões dos dados (de habitat, comunidades ou populações biológicas), tendo dados de somente pequenas porções (amostras) da população (o estatístico analisa também dados populacionais, por exemplo, dados de todos os clientes de um banco)
3. Avaliar objetivamente diferenças e relações entre conjuntos de dados

3. Confiabilidade da estatística

Exemplo: Propaganda do chiclet Trident

“Quatro em cada cinco dentistas recomendam Trident para seus pacientes”

- Quantos dentistas foram realmente entrevistados? A pesquisa não informou
- Saber o número de entrevistados faz diferença?
- Saber como foi feita a pesquisa (método) faz diferença?

Confiabilidade é um conceito importante. Relaciona-se à "representatividade" do resultado encontrado em uma amostra específica de toda a população. Em outras palavras, diz quão provável será encontrar uma relação similar se o experimento fosse feito com outras amostras retiradas da mesma população, lembrando que o maior interesse está na população.

- Com outro conjunto de amostra o resultado seria semelhante?

4. Maneiras de expressar a Estatística

1. **Razão:** Fração que divide duas quantidades. Ex: A razão de meninas para meninos é de 3 para 2 (razão é expressa no menor termo possível) – 300/200 ou 3000/2000
2. **Taxa:** Razão que reflete uma quantidade por certa unidade. Ex: 60km/h; 3 assaltos/1000 casas; Densidade demográfica em STM 12,82 hab/km², Taxa de mortalidade em STM: 17,9 óbitos/1000 nascidos vivos/ano, Taxa de natalidade em STM: 22,64/1000hab/ano.
3. **Porcentagens:** Número de 0 a 100 que reflete a proporção de um todo. Ex camiseta com 10% de desconto.

Atenção: números absolutos podem ser enganadores

➤ Exemplo

- Crimes na cidade (A) aumentaram de 50 para 60 em um ano; Crimes na cidade (B) aumentaram de 500 para 510 em um ano
- As duas cidades tiveram aumento absoluto de 10 crimes/ano.
- A criminalidade aumentou igual? Onde ocorreu maior aumento?

Cidade A

50 crimes----- 100%
10 crimes----- x

$$x = \frac{10}{50} x 100$$

$$x = 20\%$$

Cidade B

500 crimes----- 100%
10 crimes----- x

$$x = \frac{10}{500} x 100$$

$$x = 2\%$$

Exercício

Numero de crimes e tamanho estimado da população nos EUA (1987-1993)

Ano	Numero de crimes	Tamanho da população
1987	13.508.700	243.400.000
1993	14.144.800	257.908.000

O numero de crimes aumentou?

Taxa de criminalidade aumentou?

Calcular a taxa (unidade: crimes por 100.000 habitantes) - Resolver por regra de 3

$$13.508.700 \text{ ----- } 243.400.000$$

$$X \text{ ----- } 100.000$$

$$taxa \ 1987 = \frac{13508700}{243400000} x 100000$$

$$taxa \ 1987 = 0,0555x100.000 \ hab$$

$$taxa \ 1987 = 5550,0$$

$$14.144.800 \text{ ----- } 257.908.000$$

$$X \text{ ----- } 100.000$$

$$taxa \ 1993 = \frac{14144800}{257908000} x 100.000$$

$$taxa \ 1993 = 0,05484x 100.000 \ hab$$

$$taxa \ 1993 = 5484,0$$

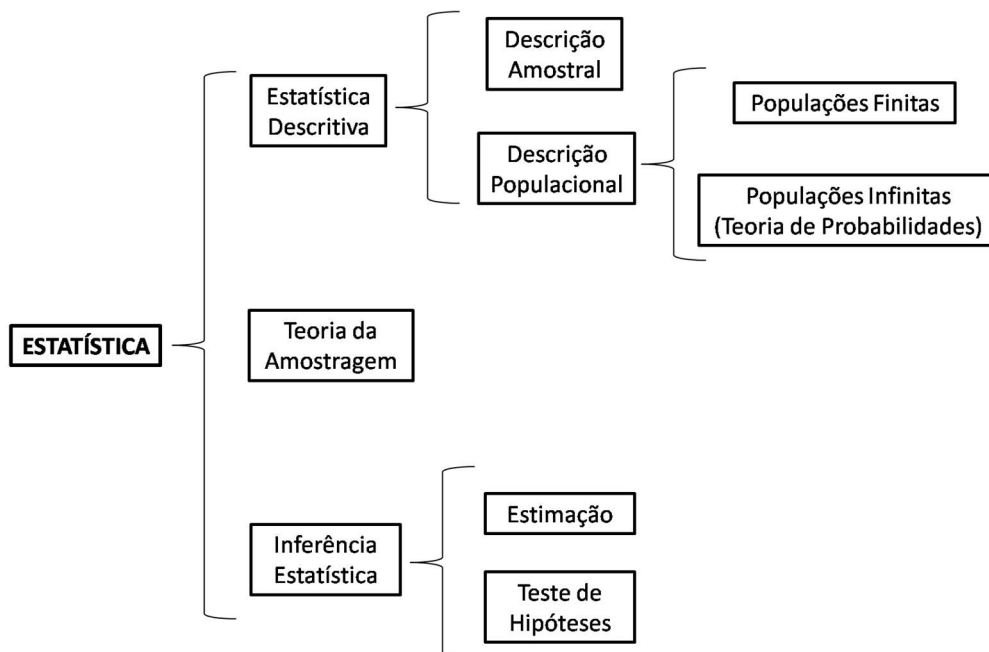
Em termos percentuais

$$mudança = \frac{5.484 - 5.550}{5550} \times 100$$

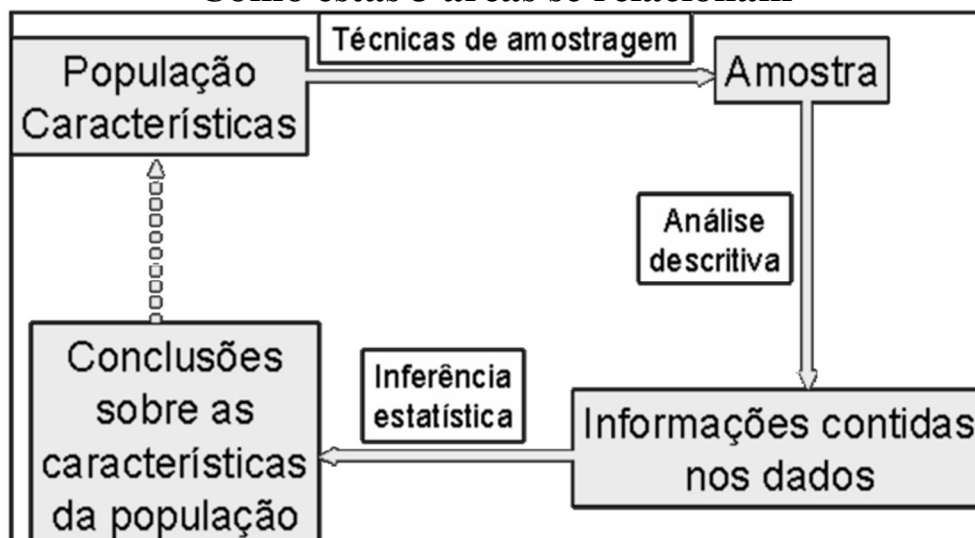
$$mudança = -1,2\%$$

5. Subdivisões da Estatística

Didaticamente, o estudo da estatística é dividido conforme a figura abaixo.



Como estas 3 áreas se relacionam



Estatística Descritiva

Objetiva sintetizar a informação contida em um conjunto de dados seja ele referente a uma população ou a uma amostra;

Alguns Termos Importantes

1. Variável:

É um termo que utilizamos em estatística para designar informação. Cada tipo de informação, objeto de estudo, corresponde um número de resultados possíveis.

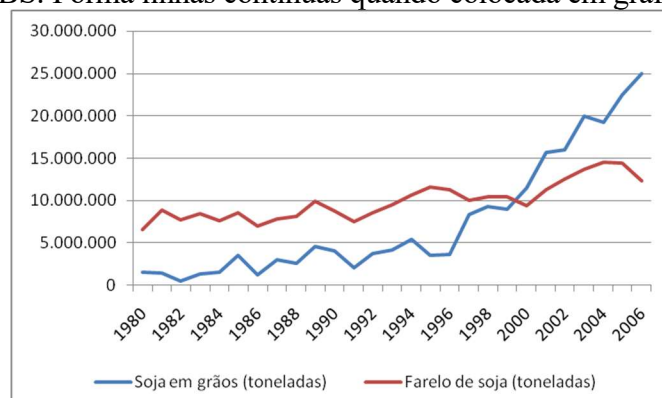
Exemplo:

- Informação “sexo” são dois os resultados possíveis, masculino e feminino
- Informação “número de filhos” há um número de resultados possíveis expressos através de números naturais (1, 2, 3, 4, ..., n)
- Informação “estatura” há um número infinito de valores numéricos dentro de um determinado intervalo

1.1 Tipos de Variáveis

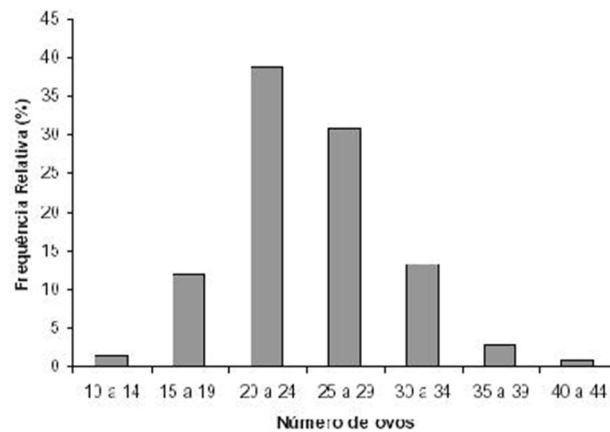
- Variável Qualitativa: quando seus valores são expressos por atributos. As variáveis qualitativas dividem-se em:
 - a) Variável Nominal: os indivíduos ou elementos são classificados em categorias mutuamente exclusivas. Não há relação clara entre os elementos Ex.: sexo (masculino e feminino), cor da pele (branca, preta, amarela, parda), clima, solo, etc.
 - b) Variável Ordinal: quando os elementos são classificados em categorias que se apresentam segundo uma ordem lógica crescente ou decrescente. Ex.: Grau de instrução - 1. sem instrução, 2. primário incompleto, 3. primário completo, Julgamentos do tipo: Bom/Médio/Ruim; Classes de fertilidade dos solos
- Variável Quantitativa: quando seus valores são expressos por números (salários, idade, peso, diâmetro, altura, etc.). As variáveis quantitativas dividem-se em:
 - a) Variável Contínua: Assume qualquer valor entre dois limites. Ex: altura, peso, volume,

OBS: Forma linhas contínuas quando colocada em gráficos



- b) Variável discreta: Só pode assumir valores inteiros. Ex: número de filhos, número de frutos, número de sementes germinadas, etc.

OBS: Não forma linhas contínuas quando colocada em gráficos



Resumindo

Variável Qualitativa (atributo)
Nominal (sexo, cor, tipo de solo)
Ordinal (grau de instrução, classe de fertilidade)

Variável Quantitativa (valor numérico)
Contínua (peso, altura, volume)
Discreta (número de frutos, número de sementes)

Valores provenientes de Medições --- Variáveis contínuas
Valores de Contagens --- Variáveis discretas

2. População ou Universo Estatístico:

- É a totalidade dos indivíduos (variáveis de mesma natureza) sobre o qual se faz uma inferência
- É qualquer conjunto de informações que tenham, entre si, pelo menos uma característica comum, cujo comportamento interessa analisar.

O que define a população é o objetivo da pesquisa ou a dúvida que se pretende responder

Exemplos:

- Controle de qualidade dos carros produzidos pela fábrica
 - População: todo o conjunto de carros produzidos por dia, ou por mês ou por ano.

- Volume de madeira em um talhão florestal de 10.000 ha
 - População: conjunto de todas as árvores com diâmetro maior que 10 cm presentes na área de 10000ha
- Taxas de crescimento e mortalidade dos frangos de um aviário
 - População: todos os frangos presentes no aviário
- Qual a taxa de crescimento de frangos na região do planalto em Santarém?
 - População: todos os frangos de todos os aviários presentes na região escolhida

As populações podem ser classificadas em:

- a. **Populações Finitas ou Reais**: Possuem existência real e todos os elementos ou indivíduos podem ser enumerados.

Ex.: alunos matriculados em determinada escola, árvores de um povoamento florestal

O número total de elementos de uma população finita é simbolizado pela letra maiúscula “N”

- b. **População Infinita ou Conceitual**: São aquelas sem existência real, mas de concepção bem definida. São de tamanho infinito porque nunca de disporá de todos os seus elementos na prática.

Ex: conjunto total de frangos que poderiam se alimentados com certa ração, conjunto total de mudas que poderão ser plantadas em determinado viveiro.

OBS: Quando não se pode determinar o tamanho e os limites da população, ela é considerada infinita para efeito estatístico.

Ex.: resultados obtidos quando se joga um dado sucessivamente.

3. Amostra

É qualquer conjunto de elementos, retirado da população, desde que esse conjunto seja não vazio e tenha menor número de elementos do que a população. É um subconjunto finito da população.

4. **Parâmetro**: Constitui o valor real, é obtido quando se mede todos os indivíduos da população.

5. **Estimativas**: São os dados obtidos quando se mede apenas parte de uma população (amostra).

Medidas de Posição	Parâmetro (população)	Estimativa (amostra)
Tamanho	N	N
Amplitude total	AT	AT
Média	M	X
Variância	σ^2	s^2
Desvio Padrão	σ	s
Coeficiente de Variação	CV	CV
Variância da média	σ_{μ}^2	sx^2
Erro padrão/desvio padrão da média	σ_{μ}	Sx

6. Medidas de Tendência Central

São medidas ou estatísticas cujos valores convergem para o centro de um conjunto de dados.

Por exemplo:

- Média Aritmética Simples
- Média Aritmética Ponderada
- Mediana
- Moda

a) Média Aritmética Simples:

É a de maior aplicação, dada a sua facilidade de determinação, mesmo que possa ser influenciada por valores discrepantes. A média aritmética geralmente é o valor que representa o conjunto de dados

A média aritmética simples é determinada pela somatória dos valores observados, divididos pelo número de elementos somados.


$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Exemplo:

Peso em gramas de 12 ratos machos da raça Wistar com 30 dias de idade

50	62	70
79	60	64
66	77	58
55	82	74

$$\bar{x} = \frac{\sum_{i=1}^{12} 50 + 79 + 66 + 55 + 62 + \dots + 74}{12}$$
$$\bar{x} = \frac{797}{12} \qquad \bar{x} = 66,4 \text{ g}$$

50	55	58	60	62	64	66		70	74	77	79	82
----	----	----	----	----	----	----	---	----	----	----	----	----

Normalmente a média fica situada no centro de gravidade


- Influência dos valores extremos discrepantes

50	62	70
86	60	64
66	77	58
55	82	74
230		

$$\bar{x} = \frac{\sum_{i=1}^{13} 50 + 86 + 66 + 55 + 62 + \dots + 230}{13}$$
$$\bar{x} = \frac{1027}{13} \qquad \bar{x} = 79,0 \text{ g}$$

Um novo rato com 230g entra para o conjunto

50	55	58	60	62	64	66	70	74	77	79	82	230
----	----	----	----	----	----	----	----	----	----	----	----	-----



O centro de gravidade foi deslocado devido ao valor discrepante

- A principal característica da média aritmética simples é que a somatória dos desvios dos valores das observações x_i em relação à média é igual a zero (0).

Exemplo: Número de sementes germinadas no quinto dia em 8 amostras de 10 sementes

7	5
5	10
3	6
9	3

$$\bar{x} = \frac{\sum_{i=1}^8 7 + 5 + \dots + 3}{8} \quad \bar{x} = \frac{48}{8} \quad \bar{x} = 6,0$$

Desvios ($x_i - \bar{x}$)

$$\begin{aligned} 7 - 6 &= 1 \\ 5 - 6 &= -1 \\ 3 - 6 &= -3 \\ 9 - 6 &= 3 \\ 5 - 6 &= -1 \\ 10 - 6 &= 4 \\ 6 - 6 &= 0 \\ 3 - 6 &= -3 \end{aligned}$$

$$\sum (x_i - \bar{x}) = 0$$

Propriedade da média aritmética

A soma algébrica dos desvios em relação à média é nula

$$\sum \text{desvios} = 0$$

b) Média Aritmética Ponderada

Quando os valores observados apresentam pesos (graus de importância) diferentes, a média é chamada de ponderada. É dada pelo quociente entre a soma dos produtos dos valores pelos respectivos pesos pela somatória dos pesos.

$$X_p = \frac{\sum X_i \cdot p_i}{\sum p_i}$$

Ex.: As notas obtidas por um candidato a um concurso vestibular nas diferentes disciplinas com seus respectivos pesos foram:

Disciplina	Notas	Pesos
Matemática	4,0	3
Português	6,0	2
Conhecimentos gerais	7,0	1
Química	8,0	1

$$X_p = \frac{(4 \times 3) + (6 \times 2) + (7 \times 1) + (8 \times 1)}{3 + 2 + 1 + 1} = 5,57$$

c) Moda

É o valor que ocorre com maior frequência em um conjunto de dados, ou seja, é o valor que mais se repete. É usado para variáveis quantitativas discretas ou variáveis qualitativas.

Exemplo: Pesquisa social em comunidades sobre a quantidade de filhos em cada família

Numero de Filhos/família	Num. de famílias
0	4
1	5
2	7
3	10
4	12
5	3
6	2

Moda = 4 filhos

- Série Amodal: Não existe valor modal, ou seja, nenhum valor aparece mais vezes que outros

Ex.: 5, 6, 3, 8, 9, 10, 2
2, 2, 3, 3, 4, 4, 5, 5.....

- Série Bimodal: Quando ocorrem dois ou mais valores de concentração

Opinião sobre sabor de uma marca de café	Num. de pessoas
Ruim	4
Razoável	15
Bom	15
Muito Bom	8
Excelente	3

Moda = Razoável e Bom (Bimodal)

Emprego da Moda:

- a) Quando desejamos obter uma medida rápida e aproximada de posição;
- b) Quando a medida de posição deve ser o valor mais típico da distribuição

d) Mediana

É o valor que ocupa a posição central de um conjunto de valores ordenados. A mediana separa o conjunto de dados em dois subconjuntos de mesmo número de elementos.

1º Caso: Quando (n) é impar

$$T = \frac{n + 1}{2}$$

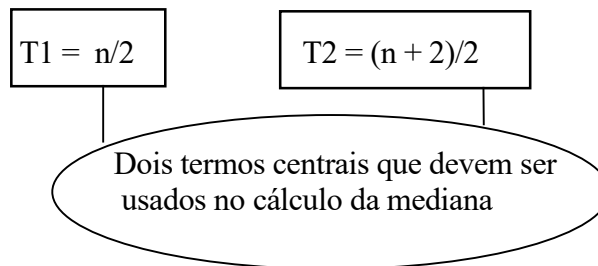
Termo que corresponde a mediana

Ex.: 9, 26, 15, 2, 5, 50, 50, 31, 44 ----- Ordenamos os valores
2, 5, 9, 15, 26, 31, 44, 50 -----n = 9

T = (9 + 1) / 2 ---- T = 5º termo ----- Md = 26

Obs.: A mediana é o valor que ocupa a posição central dos dados ordenados, constituindo um valor do próprio conjunto

2º Caso: Quando (n) é par:



Ex.: 9, 26, 15, 8, 12, 24, 6, 3 ---- ordenar --- 3, 6, 8, 9, 12, 15, 26, 24 n = 8

$T1 = 8/2$ $T2 = (8 + 2)/2$
 $T1 = 4$ termo $T2 = 5$ termo

$Md = (9 + 12)/2$ ----- $Md = 10,5$

Obs.: Nesse caso, a mediana é a média aritmética dos dois valores que ocupam a posição central dos dados ordenados, constituindo um valor teórico que não figura entre os dados originais.

Emprego da Mediana:

- a) Quando desejamos obter o ponto que divide a distribuição em partes iguais;
- b) Quando há valores extremos que afetam de uma maneira acentuada a média;

7. Medidas de Dispersão:

São utilizadas em estatística para quantificar variações, ou seja, medem o grau de dispersão dos valores de um conjunto de dados.

Ilustração

Caso do homem com um pé em um balde de água fervendo e outro pé em um balde de água com gelo – Na “média” ele deve estar se sentindo ótimo! Mas imagine a variabilidade da temperatura.

Ex.: Notas de alunos de uma classe: Calcular a média final desses alunos

A = 5, 5, 5, 5 X = 5 ----- Não tem dispersão
B = 4, 6, 4, 6 X = 5 ----- Dispersão pequena
C = 0, 4, 6, 10 X = 5 ----- Dispersão intermediária
D = 0, 0, 10, 10 X = 5 ----- Maior dispersão

a) Amplitude Total (AT)

É a diferença entre o maior e o menor valor observado em um conjunto de dados. A amplitude utiliza apenas com os valores extremos. Dois conjuntos de dados podem apresentar a mesma amplitude mesmo que tenham dispersão diferente.

$$AT = x (\text{máximo}) - x (\text{mínimo})$$

Ex.: idades de dois grupos de pessoas

$$A = 3, 4, 9, 11, 19, 20 \text{ ----- } AT = 20 - 3 \text{ ---- } AT = 17$$

$$B = 10, 3, 11, 11, 10, 20 \text{ ---- } AT = 20 - 3 \text{ --- } AT = 17$$

- Mesma amplitude, mas dispersão diferente

b) Variância (s^2)

É uma medida quadrática que indica a variação entre as observações

Para dizer que algo variou precisamos de um ponto de referencia. Este ponto é a média aritmética

Porque é quadrática?

Devido a propriedade da média aritmética

- Propriedades:

1. A soma dos desvios é igual a zero

$$\sum (X_i - \bar{X}) = 0$$

Exemplo anterior: Número de sementes germinadas em 8 amostras de 10 sementes

<hr/>		
7 5	$\bar{x} = 6,0$	Desvios ($x_i - \bar{x}$)
5 10		$7 - 6 = 1$
3 6		$5 - 6 = -1$
		$3 - 6 = -3$
9 3		$9 - 6 = 3$
<hr/>		$5 - 6 = -1$
		$10 - 6 = 4$
		$6 - 6 = 0$
	$3 - 6 = -3$	

$$\sum \text{desvios} = 0$$

Como quantificar a variação?

2. A soma dos quadrados dos desvios é diferente de zero, ou seja, quantifica a variação

$$\sum (X_i - \bar{X})^2 \neq 0 \text{ (variação)}$$

Quadrado dos desvios $(x_i - \bar{x})^2$

$$1^2 = 1$$

$$-1^2 = 1$$

$$-3^2 = 9$$

$$3^2 = 9$$

$$-1^2 = 1$$

$$4^2 = 16$$

$$0^2 = 0$$

$$-3^2 = 9$$

 $\sum x = 46$

Em média, quanto cada valor se desviou do valor central ?

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

$$S^2 = \frac{46}{8}$$

$$S^2 = 5,75$$

Fórmula da variância quando trabalhamos com toda a população de dados

Fórmulas para dados amostrais

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n - 1}$$

$$s^2 = \frac{\sum_{i=1}^n X_i^2 - \frac{(\sum X_i)^2}{n}}{n - 1}$$

Para dados amostrais a formula utiliza o denominador (n-1) porque perde-se um grau de liberdade.

O que isso quer dizer??

A formula da variância utiliza o valor da média (que já é uma estimativa) então estamos fazendo uma estimativa usando o valor de outra estimativa. Por isso nossa incerteza aumenta.

Somatório de n observações da variável X_i^2 (**soma de quadrados**):

$$\sum_{i=1}^n X_i^2 = X_1^2 + X_2^2 + X_3^2 + X_4^2 + \dots + X_n^2$$

Quadrado do Somatório de n observações da variável X (**quadrado da soma**):

$$\left(\sum_{i=1}^n X_i\right)^2 = (X_1 + X_2 + X_3 + \dots + X_n)^2$$

c) Desvio Padrão (s)

Serve para medir a variação de cada observação em relação à média. Quanto menor for o desvio, menor a variabilidade dos valores com relação à média.

O desvio padrão tal como o nome diz é um padrão, ou seja, algo típico de desvio (distancia mais comum da maioria dos valores em relação à média).

Em outras palavras

“É a distancia média dos dados em relação a média”

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

Quando somamos o quadrado de cada desvio $\sum(x_i - \bar{x})^2 \neq 0$ obtemos um valor diferente de zero que quantifica a variação, porém estamos também superestimando e variação.

Para corrigir esta superestimativa tiramos a raiz quadrada da variância (inverso da potencia) e obtemos o desvio padrão (medida linearizada)

d) Coeficiente de variação (CV)

É uma medida de dispersão relativa, porque estabelece relação entre o desvio padrão e a média, ou seja, é o desvio padrão expresso em porcentagem da média.

$$CV = \frac{s}{\bar{x}} \cdot 100$$

CV baixo --- inferior a 10 %

CV médio --- está entre 10 % e 20 %

CV alto --- está entre 20 % e 30 %

CV muito alto --- quando > 30 %

Ex.: Dispersão relativa

Grupo A

Idades: 1, 3, 5

$$x = 3$$

$$s^2 = 4$$

$$s = 2$$

Grupo B

Idades: 53, 55, 57

$$x = 55$$

$$s^2 = 4$$

$$s = 2$$

OBS.: A variância do primeiro grupo é igual à variância do segundo grupo, logo a dispersão dos dados em torno da média é exatamente a mesma nos dois grupos.

Grupo A

$$CV = \frac{s}{\bar{x}} \cdot 100$$

$$CV = \frac{2}{3} \cdot 100$$

$$CV = 66,7 \%$$

Grupo B

$$CV = \frac{s}{\bar{x}} \cdot 100$$

$$CV = \frac{2}{55} \cdot 100$$

$$CV = 3,6 \%$$

OBS.: A dispersão dos dados em relação à média é maior no primeiro grupo (CV = 66,7 %) e menor no segundo grupo (CV = 3,6 %).

8. Medidas de Precisão:

São utilizadas em estatística para avaliar a precisão de uma amostragem, ou seja, o quanto a amostragem foi eficiente em representar a população.

a) Variância da média (S_x^2)

Se em vez de uma amostra tivéssemos várias provenientes de uma mesma população, obteríamos também diversas estimativas da média, e, provavelmente distintas entre si. A amplitude de variação dessas médias é chamada variância da média.

$$S_x^2 = \frac{S^2}{n}$$

A variância da média é uma estatística que combina a variabilidade da amostra e o seu tamanho

b) Erro Padrão ou Desvio Padrão da Média (S_x)

É a raiz quadrada da variância da média, ou seja, é a variância da média, que é uma função quadrática, expressa de forma linear.

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

$$S_x = \frac{S}{\sqrt{n}}$$

REVISANDO

Medidas de tendência Central

Média Aritmética (mais usada para dados quantitativos)

Mediana (substitui a média quando existem valores discrepantes)

Moda (para dados qualitativos)

Medidas de Dispersão

Amplitude total (maior valor menos o menor)

Variância (medida quadrática, indica variação, superestimada)

Desvio Padrão (medida linear, indica a variação média dos valores em torno da média)

Coefficiente de variação (dispersão relativa, desvio expresso em % da média)

Medidas de Precisão

Variância da média (amplitude de variação das médias se varias amostragens fossem feitas em uma mesma população)

Erro Padrão (linearização da variância da média)

Exercícios de Aplicação

1) Os dados abaixo representam produções de milho em Kg /100 m² de diferentes fazendas situadas no Estado do Pará. Realizar a análise estatística dos dados.

Dados:

34,0	26,0	37,9	26,9	29,3	22,8
26,9	37,0	45,3	28,4	31,4	20,2
33,2	42,1	39,0	30,6	35,8	23,5
36,1	34,3	41,2	37,9	31,4	24,8
31,4	36,8	53,5	32,5	30,8	25,0

Pede-se:

- 1) Média Aritmética
- 2) Mediana
- 3) Moda
- 4) Amplitude Total
- 5) Variância
- 6) Desvio Padrão
- 7) Coeficiente de Variação
- 8) Variância da Média
- 9) Erro Padrão

2) Uma Empresa de Papel realizou testes para comparar três máquinas de produção de papel. Em cada máquina foram tomadas 5 amostras conforme o quadro abaixo. Qual das máquinas a empresa deve adquirir? Por quê?

GRAMATURA DO PAPEL PRODUZIDO EM DIFERENTES MÁQUINAS

Amostra	Máquinas		
	A	B	C
1	200	152	205
2	210	248	203
3	190	260	195
4	215	200	197
5	185	140	200
Média	200	200	200

3) Duas equipes de medição foram analisadas em termos do número de parcelas medidas por dia:

Equipe A: {24, 16, 26, 16, 15, 14, 13, 24, 18, 17}

Equipe B: {19, 16, 16, 18, 18, 19, 20, 17, 17, 15}

- Qual equipe é mais produtiva?
- Qual equipe é mais consistente?

Distribuição de Frequência de uma Variável

Para melhor estudar e descrever as variáveis usa-se agrupá-las e classificá-las em tabelas de frequência. Os objetivos das tabelas de frequência são:

- Realçar o que há de essencial nos dados (analisa o conjunto total de valores desinteressando-se pelos casos isolados);
- Tornar possível o uso de técnicas analíticas para a sua total descrição.

6.1 Elementos de uma distribuição de frequência

- Dados Brutos ou Tabela Primitiva: São observações no campo sem uma ordenação prévia
- Ordenação dos dados: Os dados devem ser organizados em forma crescente ou decrescente para permitir melhor manipulação

Ex.: Peso em Kg de suínos em um criadouro

Dados brutos

66,3	62,2	61,1	50,0	62,5	60,5
62,8	61,4	68,4	63,1	56,2	73,0
60,7	55,8	64,5	68,2	55,5	52,3
63,7	60,9	55,7	55,2	69,5	51,5
70,0	64,0	54,1	61,3	56,0	72,7

Dados ordenados

50,0	51,5	52,3	54,1	55,2	55,5
55,7	55,8	56,0	56,2	60,5	60,7
60,9	61,1	61,3	61,4	62,2	62,5
62,8	63,1	63,7	64,0	64,5	66,3
68,2	68,4	69,5	70,0	72,7	73,0

- Classes de frequência: são intervalos de variação da variável
As classes são representadas simbolicamente por i

$$i = 1, 2, 3, \dots, k$$

$$k = n^{\circ} \text{ total de classes de distribuição}$$

-- Maneiras de se determinar o número de classes de uma distribuição de frequência

- Regra de Sturges: o número de classes esta em função do número de valores da variável

$$i = 1 + 3,3 \cdot \log n$$

$$i = 1 + 3,3 \cdot \log 30$$

$$i = 1 + 3,3 \cdot 1,477$$

$$i = 5,8 \text{ --- } i = 6,0$$

Essa regra nos permite obter a seguinte tabela

n	classes (i)
6 -- 11	4
12 -- 22	5
23 -- 46	6
47 -- 90	7
91 -- 181	8
182 -- 362	9

- Fórmula de Yule

$$i = 2,5 \cdot \sqrt[4]{n}$$

n = n° de valores do rol de dados

$$i = 2,5 \cdot 2,34$$

$$i = 5,85 \text{ ---- } i = 6,0$$

OBSERVAÇÕES

- Os resultados obtidos por meio de fórmulas podem servir como referência - mas não devem ser entendidos como obrigatórios.
- O número de classes deve ser escolhido pelo pesquisador, em função do que ele quer mostrar (ligado à natureza dos dados, da unidade usada para expressá-los e, ainda, do objetivo que se tem em vista).
- Em geral, convém estabelecer de 5 a 20 classes.
- Se o número de classes for muito pequeno (por exemplo, 3), perde-se muita informação.
- Se o número de classes for muito grande (por exemplo, 30), têm-se pormenores desnecessários.
- Não existe um número "ideal" de classes para um conjunto de dados
- Evite classes com frequência nula ou com frequência muito alta

d) Limites de Classe

Denominamos limites de classe aos valores extremos de cada classe

Menor número --- Limite inferior (l)

Maior número --- Limite superior (L)

l | ——— L

Inclusão do limite inferior (l)

Exclusão do limite superior (L)

e) Amplitude Total da distribuição: É a diferença entre o maior e o menor valor do rol de dados

$$AT = X(\max) - X(\min)$$

$$AT = 73 - 50 \quad AT = 23$$

f) Intervalo de Classe

É a medida do intervalo que define cada classe

$$IC = \frac{AT}{N \text{ de classes}}$$

$$IC = 23/6 \text{ ----- } IC = 3,83 \text{ ----- } IC = 4,0 \text{ (sempre arredondar para +)}$$

g) Ponto Médio de uma Classe ou Centro de Classe (Xi)

É o ponto que divide o intervalo em duas partes iguais. O ponto médio de uma classe é o valor que a representa.

$$X_i = \frac{li + Li}{2}$$

OBS: A primeira classe geralmente inicia com o menor valor do rol. Depois basta adicionar o valor do intervalo de classe

Exemplo:

Nº	Classes	Centro classe (Xi)	Frequência (fi)	Freq. Acum. (fa)	Freq. relativa (fr %)	Freq. relat. acum. (fra%)	fi . xi	fi. xi ²
1	50,0 --- 54,0	52,0	3					
2	54,0--- 58,0	56,0	7					
3	58,0 --- 62,0	60,0	6					
4	62,0 --- 66,0	64,0	7					
5	66,0 --- 70,0	68,0	4					
6	70,0 ---74,0	72,0	3					
Σ	-----	----	30					

Centro de Classe

$$X_1 = \frac{50,0 + 54,0}{2} \text{ ---- } X_1 = 52,0$$

$$X_2 = \frac{54,0 + 58,0}{2} \text{ ---- } X_2 = 56,0$$

h) Frequência Simples ou Absoluta (fi)

É o número de observações cujos valores pertencem a determinada classe.

$$\sum Fi = N \text{ (número de valores na classe)}$$

i) Frequência Acumulada (fa)

Frequência acumulada de uma dada classe é a somatória de todas as frequências absolutas das classes que a antecedem.

$$Fa(i) = f_1 + f_2 + \dots + f_i$$

$$Fa(2) = 3 + 7$$

$$Fa(2) = 10$$

$$Fa(3) = 3 + 7 + 3$$

$$Fa(3) = 13$$

j) Frequência Relativa (fr)

São os valores das razões entre as frequências simples a frequência total

$$Fr(i) = \frac{f_i}{\sum f_i} \cdot 100$$

$$Fr(1) = 3/30 \quad \text{----} \quad Fr(1) = 10\%$$

$$Fr(2) = 7/30 \quad \text{----} \quad Fr(2) = 23,33\%$$

k) Frequência Relativa Acumulada (Fra)

Frequência relativa acumulada de uma classe é a somatória de todas as frequências relativas das classes que a antecedem.

$$Fra(i) = fra(1) + fra(2) + \dots + fra(i)$$

$$Fra(2) = 10,0 + 23,3 \quad \text{----} \quad Fra(2) = 33,3$$

O conhecimento dos vários tipos de frequência ajuda-nos a responder a muitas questões com relativa facilidade como:

- a) Quantos animais têm peso inferior a 58kg?
- b) Qual a percentagem de animais cujos pesos são superiores a 70 kg?
- c) Quantos animais têm peso situado entre 66,0 e 70 kg?
- d) Quantos animais têm peso superior a 62,0 kg?

Fórmulas estatísticas utilizadas para dados agrupados

- a) Média Aritmética: Quando os dados estão agrupados as frequências são números indicadores da intensidade de cada valor variável, funcionando assim como fatores de ponderação, por isso para dados agrupados só é válida a Média Aritmética Ponderada.

$$X_p = \frac{\sum X_i \cdot f_i}{\sum f_i}$$

$$X_p = \frac{1844,0}{30}$$

$$X_p = 61,47$$

b) Moda

- Moda Bruta: Dada pelo ponto médio da classe modal

$$Mo = \frac{l + L}{2} \quad \begin{array}{l} l = \text{Limite inferior da classe modal} \\ L = \text{Limite superior da classe modal} \end{array}$$

Bimodal

$$Mo1 = 56,0$$

$$Mo2 = 64,0$$

c) Mediana

$$Md = Li + \left(\frac{C \cdot (P - fa(ant))}{f} \right)$$

Onde:

l : limite inferior da classe que contem mediana

C : intervalo de classe

f ant : frequência acumulada da classe anterior à classe mediana

f : frequência simples da classe mediana

P: classe cuja frequência acumulada contem a mediana

$$P = \frac{\sum fi}{2}$$

$$P = \frac{\sum fi + 1}{2}$$

Se N for par

Se N for impar

$$P = 30/2 \text{ --- } P = 15 \text{ (3ª classe)}$$

$$Md = 58 + \frac{4 [15 - 10]}{6}$$

$$Md = 61,33 \text{ kg}$$

d) Variância e Desvio Padrão

$$S^2 = \frac{\sum fi \cdot xi^2 - \frac{(\sum Xi fi)^2}{\sum fi}}{\sum fi - 1}$$

$$S^2 = \frac{114384 - \frac{1844^2}{30}}{30 - 1}$$

$$S^2 = \frac{1039,47}{29}$$

$$S^2 = 35,84 \text{ kg}^2$$

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

$$S = \sqrt{35,84}$$

$$S = 5,99 \text{ kg}$$

e) Coeficiente de Variação

$$CV = \frac{S}{\bar{x}} \cdot 100$$

$$CV = \frac{5,99}{61,47}$$

$$CV = 9,74\%$$

a) Variância da Média e Erro Padrão

$$S_x^2 = \frac{S^2}{n} \quad S_x = \sqrt{S_x^2}$$

$$S_x^2 = \frac{35,84}{30} \quad S_x = 1,09 \text{ kg}$$

$$S_x^2 = 1,19 \text{ kg}^2$$

Obs.: O cálculo das medidas de tendência central e de dispersão utilizadas em dados agrupados em tabelas de frequência são menos precisos do que os realizados com os valores isolados.

Exercícios de Aplicação

1) Uma série de amostras de solo resultaram na seguinte distribuição para os valores de PH de um solo latossolo vermelho-amarelo.

Classes de PH	Frequência
4.6 - 5,0	7
5.0 - 5,4	14
5.4 - 5,8	23
5.8 - 6,2	32
6.2 - 6,6	19
6.6 - 7,0	8
7.0 - 7,2	2
Total	105

Encontre:

- Média
- Mediana
- Variância
- Desvio padrão
- Coefficiente de variação.

2) A distribuição dos diâmetros das árvores de uma floresta nativa segue a seguinte tabela:

Classes de Diâmetro	Frequência
10 - 20	351
20 - 30	160
30 - 40	86
40 - 50	40
50 - 60	20
60 - 70	4
70 - 80	4
80 - 90	3
90 - 100	1
Total	669

- Qual o diâmetro médio e o diâmetro mediano desta floresta? Qual dos dois é maior?
- Qual o desvio padrão dos diâmetros desta floresta?
- Um Engenheiro Florestal deseja fazer um corte seletivo retirando 20% das maiores árvores. Qual o diâmetro mínimo das árvores a serem removidas?
- Uma Engenheira Florestal deseja fazer um corte seletivo retirando 40% das menores árvores. Qual o diâmetro mínimo das árvores remanescentes?

Representação Gráfica de uma Distribuição de Freqüência

Conceito: é uma figura que mostra visualmente o banco de dados.

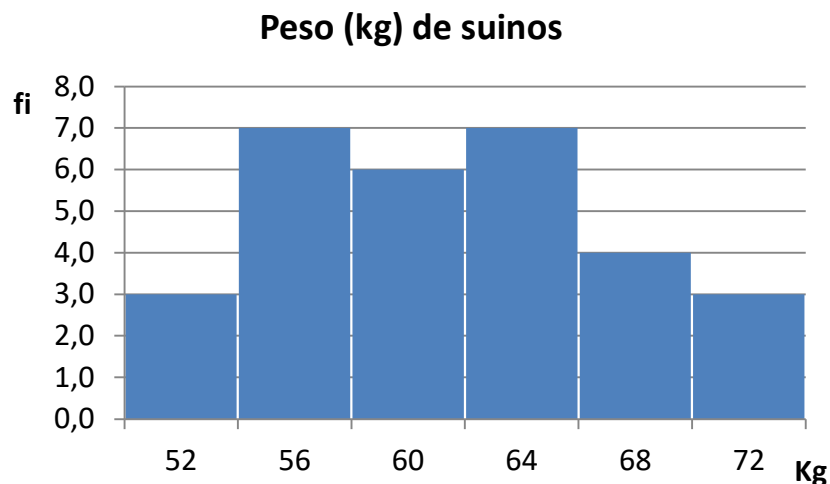
Finalidade: ilustrar um fenômeno, compreender resultados e destacar uma informação.

Uma distribuição de freqüência pode ser representada graficamente pelo:

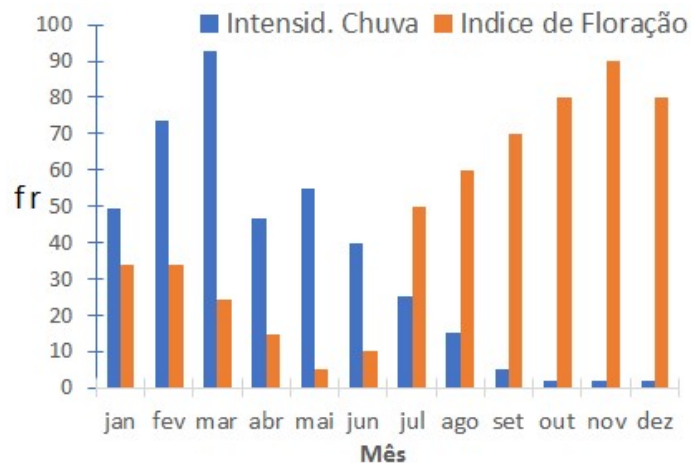
- Histograma
- Polígono de freqüência
- Polígono de freqüência acumulada

a) Histograma: É formado por um conjunto de retângulos justapostos tendo por base o intervalo de classe ou o centro de classe e para a altura a freqüência da classe.

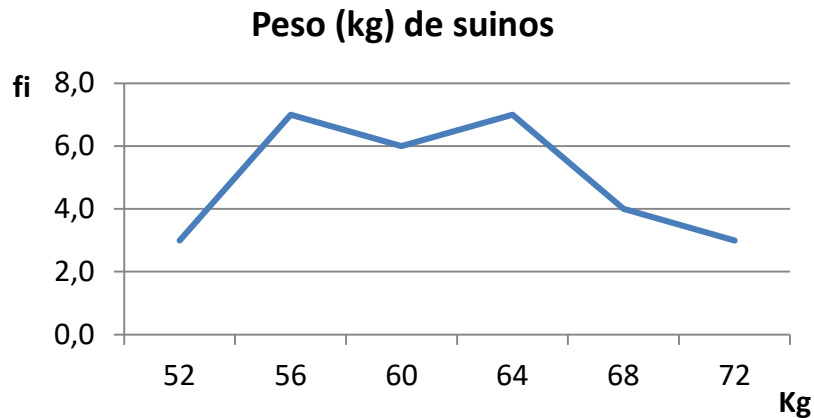
- A área de um histograma é proporcional à soma das freqüências



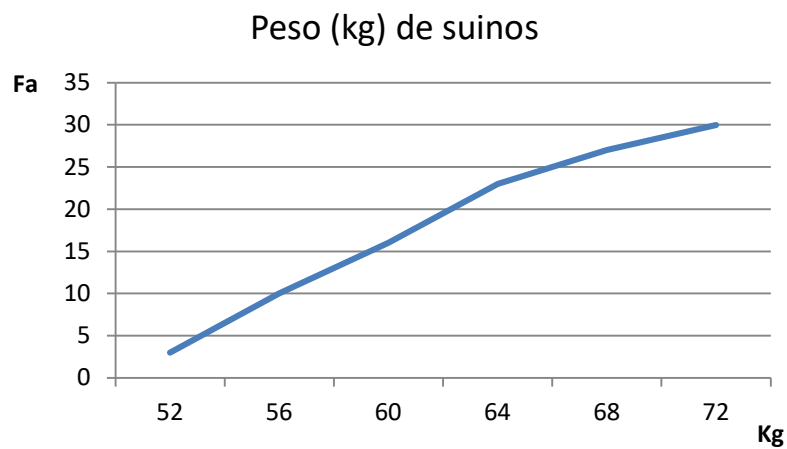
- Quando queremos comparar duas distribuições, o ideal é fazê-lo pelo histograma de freqüências relativas



b) Polígono de Frequência: É obtido pela união dos pontos correspondentes aos centros de classe



c) Polígono de Frequência Acumulada: É traçado marcando as frequências acumuladas sobre o eixo vertical e o centro de classe sobre o eixo horizontal.

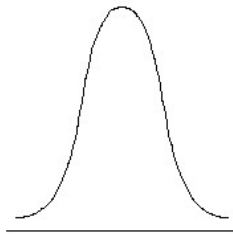


Curvas de Frequência

Mostram de modo mais evidente, a verdadeira natureza da distribuição, ou seja fornecem uma imagem tendencial do fenômeno.

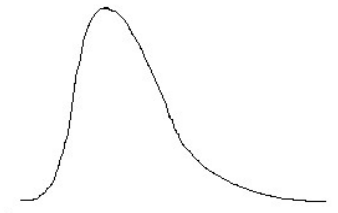
Tipos de curva

1. Curvas em forma de sino: apresentam um valor máximo na região central



Formade Sino
Unimodal e Simétrica

Simétrica: $\text{media} = \text{moda} = \text{mediana}$



Unimodal e Assimétrica
à Direita

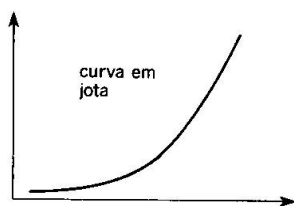
Assimétrica positiva: a calda se alonga a direita ($\text{media} > \text{mediana} > \text{moda}$)



Unimodal e Assimétrica
à Esquerda

Assimétrica negativa: a calda se alonga a esquerda ($\text{media} < \text{mediana} < \text{moda}$)

2. Curvas em forma de Jota: São relativas a distribuições extremamente assimétricas, caracterizadas por apresentarem o ponto de ordenada máxima em uma das extremidades. São curvas comuns aos fenômenos econômicos e financeiros e a distribuição de frequência em florestas primárias.



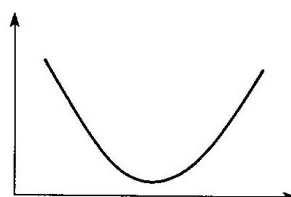
curva em
jota



curva em
jota invertido

3. Curvas em Forma de U: São caracterizadas por apresentarem ordenadas máximas em ambas a extremidades

Exemplo: Curva de mortalidade por idade



ESTATÍSTICA GRÁFICA

Em todo estudo estatístico existe a coleta de dados observados, estes dados necessitam ser organizados para que se faça a análise e a interpretação dos mesmos

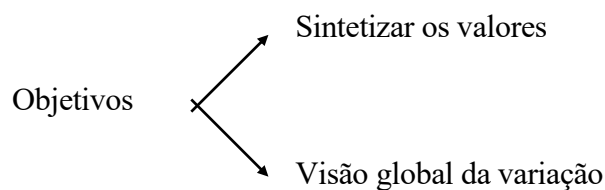
A estatística gráfica corresponde à apresentação dos dados sob diferentes formas, a fim de permitir uma visão rápida e global do fato estudado.

As informações podem ser apresentadas de duas formas: Tabelas e Gráficos

1. Organização de Dados em Tabelas

É a apresentação dos valores (variáveis) estatísticos em tabelas

As tabelas têm a vantagem de conseguir expor, sinteticamente em um só local, os resultados sobre determinado assunto



1.1 Tabela: É um quadro que resume um conjunto de observações

- A elaboração de tabelas obedece à Resolução nº 886, de 26 de outubro de 1966, do Conselho Nacional de Estatística.
- As normas de apresentação são editadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

Representação esquemática

Título

Cabeçalho

Corpo

Rodapé

a) Elementos de uma tabela

Título: O título deve responder as seguintes questões:

- O que? (Assunto a ser representado - Fato);
- Onde? (O lugar onde ocorreu o fenômeno - local);
- Quando? (A época em que se verificou o fenômeno - tempo).

Cabeçalho: parte da tabela na qual é designada a natureza do conteúdo de cada coluna.

Corpo: parte da tabela composta por linhas e colunas.

Linhas: parte do corpo que contém uma seqüência horizontal de informações.

Colunas: parte do corpo que contém uma seqüência vertical de informações.

Coluna Indicadora: coluna que contém as discriminações correspondentes aos valores distribuídos pelas colunas numéricas.

Casa ou célula: parte da tabela formada pelo cruzamento de uma linha com uma coluna.

Rodapé: É o espaço aproveitado em seguida ao fecho da tabela, onde são colocadas as notas de natureza informativa (fonte, notas e chamadas).

Fonte: refere-se à entidade que organizou ou forneceu os dados expostos.

Notas e Chamadas: são esclarecimentos contidos na tabela (**nota** - conceituação geral; **chamada** - esclarecer minúcias em relação a uma célula).

b) Normas para a apresentação de Tabelas:

- a) As tabelas devem ser delimitadas, no alto e em baixo, por traços horizontais. Esses traços podem ser mais fortes do que os traços feitos no interior da tabela;
- b) As tabelas não devem ser delimitadas, á direita e à esquerda, por traços verticais;
- c) O cabeçalho deve ser delimitado por traços horizontais;
- d) Podem ser feitos traços verticais no interior da tabela, separando as colunas;
- e) As tabelas devem ter significado próprio, isto é, devem ser entendidas mesmo quando não se lê o texto em que estão apresentadas;
- f) As tabelas devem ser numeradas por algarismos arábicos. Ex: Tabela 1, Tabela 1.2 etc...;
- g) Não devem ser apresentadas tabelas com um número de casas vazias maiores do que o com casas preenchidas;
- h) A tabela deve ser colocada no texto em posição tal que não exija, para a leitura, rotação da página. Caso isso não seja possível (tabelas com um número muito grande de colunas) a tabela deve ser alocada de modo que permita a leitura com rotação no sentido horário;
- i) Quando dois ou mais tipos de informação tiverem sido agrupados em um só conjunto, esse conjunto recebe o nome de “outros”.

1.2. Séries Estatísticas

Denominamos Série Estatística toda a tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do local ou da espécie.

- a) **Série Temporal ou Cronológica:** É a série cujos dados estão dispostos em correspondência com o tempo, ou seja, varia o tempo e permanece constante o fato e o local.

Produção de Petróleo Bruto no Brasil de 1976 a 1980 (x 1000 m³)

Anos	Produção
1976	9 702
1977	9 332
1978	9 304
1979	9 608
1980	10 562

Fonte: Conjuntura Econômica (fev. 1983)

- b) **Série Geográfica ou Territorial:** É a série cujos dados estão dispostos em correspondência com o local, ou seja, varia o local e permanece constante a época e o fato.

População Urbana do Brasil em 1980 (x 1000)

Região	População
Norte	3 037
Nordeste	17 568
Sudeste	42 810
Sul	11 878
Centro-Oeste	5 115
Total	80 408

Fonte: Anuário Estatístico (1984)

- c) **Série Específica ou Qualitativa:** É a série cujos dados estão dispostos em correspondência com a espécie ou qualidade, ou seja, varia o fato e permanece constante a época e o local

População Urbana e Rural do Brasil em 1980 (x 1000)

Localização	População
Urbana	80 408
Rural	38 566
Total	118 974

Fonte: Anuário Estatístico (1984)

- d) **Série Mista ou Composta:** A combinação entre duas ou mais séries constituem novas séries denominadas compostas e apresentadas em tabelas de dupla entrada. O nome da série mista surge de acordo com a combinação de pelo menos dois elementos.

Local + Época = Série Geográfica Temporal

População Urbana do Brasil por Região de 1940 a 1980 (x 1000)

Anos	REGIÕES				
	N	NE	SE	S	CO
1940	406	3 381	7 232	1 591	271
1950	581	4 745	10 721	2 313	424
1960	958	7 517	17 461	4 361	1 007
1970	1 624	11 753	28 965	7 303	2 437
1980	3 037	17 567	42 810	11 878	5 115

Fonte: Anuário Estatístico (1984)

2. Apresentação Gráfica de Dados

Os gráficos comunicam as mesmas idéias das tabelas, porém produzem uma impressão e compreensão mais rápida, mais viva, pois eliminam os detalhes desnecessários, visualizando somente as características mais importantes dos dados.

Requisitos para um bom gráfico

- **Simplicidade:** não deve ter detalhes de importância secundária ou traços desnecessários
- **Clareza:** deve possibilitar a correta interpretação dos valores
- **Veracidade:** deve expressar a verdade sobre o fenômeno em estudo

Regras de construção dos gráficos

- Legendas: todas iguais colocadas abaixo ou ao lado do gráfico.
- Leitura da esquerda para a direita.
- São incluídas no gráfico somente as coordenadas indispensáveis à leitura.
- Escala horizontal cresce para a direita e a escala vertical para cima.
- As linhas do gráfico devem ser individualizadas, usando cores diferentes ou espessuras diferentes.

2.1 Tipos de Gráfico

Diagramas:

- Gráfico em linhas ou curvas
- Gráfico em colunas
- Gráfico em barras
- Gráfico em setores (pizza)

Cartogramas

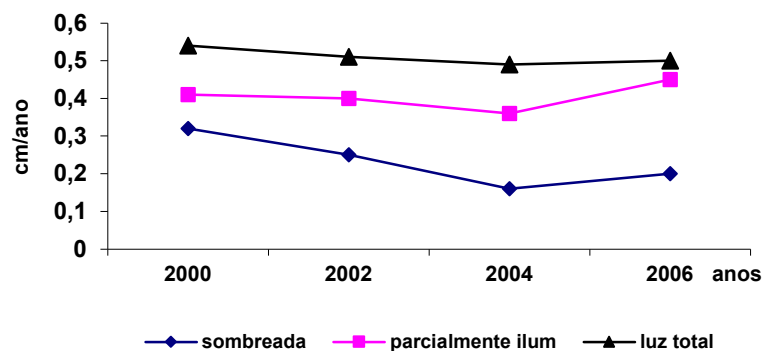
Pictogramas

a) Diagramas: gráficos geométricos de, no máximo, duas dimensões.

- **Gráfico em linha ou em curva:**

- Utiliza o sistema cartesiano
- É utilizado principalmente para séries temporais, no eixo x colocam-se as épocas e no eixo y as grandezas
- Obedece a ordem cronológica, independe do total.

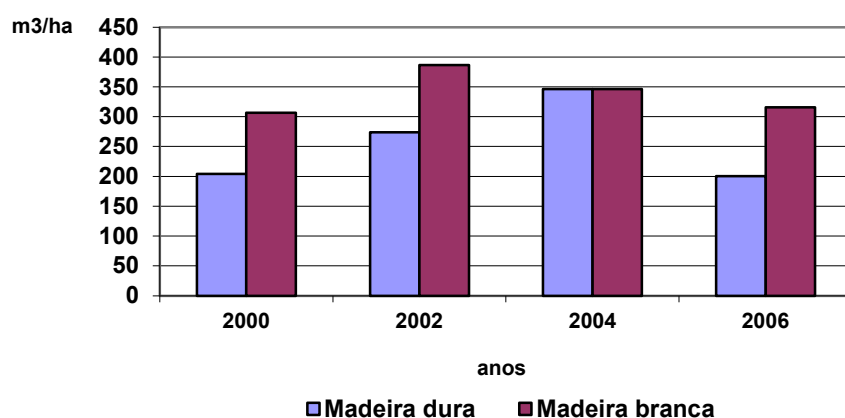
Ex: Taxas de crescimento em Diâmetro segundo categoria de iluminação em 20 ha da Flona Tapajós no período de 2000 a 2006.



- **Gráfico em colunas (simples e múltiplas):**

- Esse gráfico é formado por retângulos dispostos verticalmente de mesma largura, arbitrária, e altura proporcional às grandezas dos fenômenos
- Indicado para séries geográficas ou específicas com nomes curtos, independentes do total, obedecendo a ordem crescente ou decrescente.
- Se a série for temporal respeita-se a ordem cronológica.

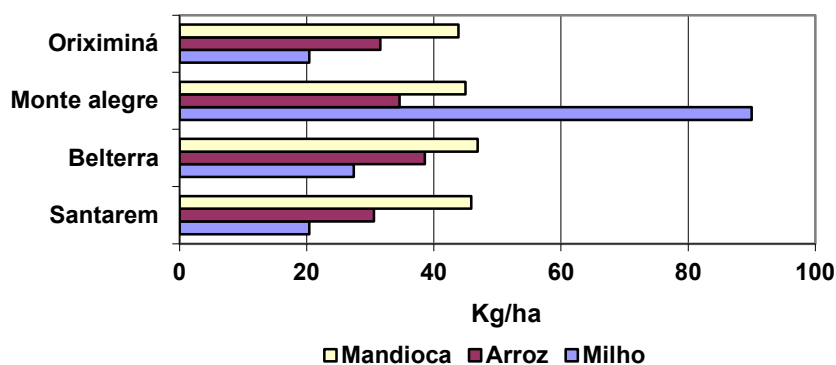
Volume de madeira comercializada na região de Santarém no período de 2000 a 2006



- **Gráfico em Barras:**

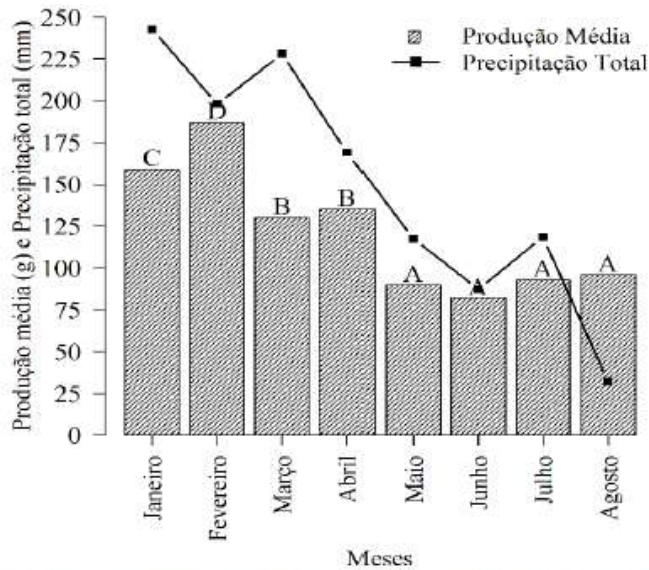
- Esse gráfico segue as mesmas normas do gráfico de colunas, porém os retângulos ocupam posição horizontal.
- Indicado para séries geográficas ou específicas com nomes extensos, independentes do total.

Produção média (kg/ha) de alimentos em três municípios do oeste do Pará no ano de 2005



- **Gráficos Misto**

- Mostram relação existente entre duas variáveis



Misto ⇒ Linha e Coluna

Figura 2. Produtividade média mensal e precipitação total, no período de janeiro a agosto (2012), Comunidade Jamaraguá, Flona Tapajós. Médias seguidas pela mesma letra não diferem entre si pelo teste SNK, a 5% de probabilidade. Dados de precipitação foram obtidos no Instituto Nacional de Meteorologia.

• **Gráfico em setores:**

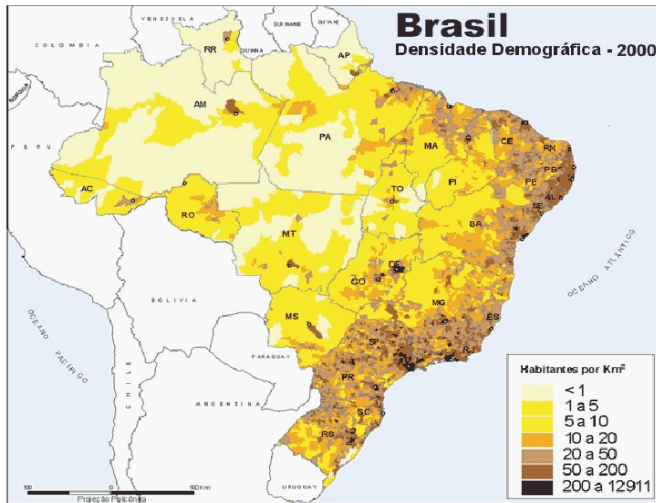
- É construído com base em um círculo e é empregado sempre que queremos ressaltar a participação do dado no total
- Cada dado (quantidade) representa um setor do círculo,
- Recomenda-se não inscrever no círculo, utilizar legenda

Rebanhos brasileiros (milhões de cabeças) em 1988

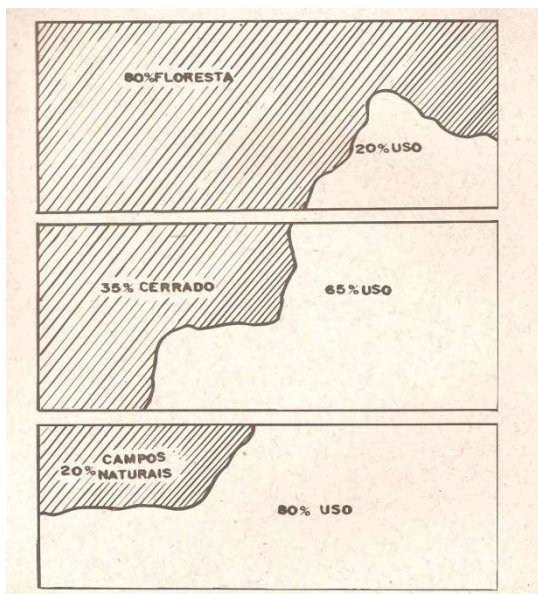


b) Cartogramas: É a representação sobre uma carta geográfica. É empregado quando o objetivo é o de figurar os dados diretamente com áreas geográficas.

Ex: mapas de solo, mapa de vegetação, mapa de densidade populacional, etc



c) Pictogramas: A representação gráfica é feita usando-se figuras. É muito utilizado, por exemplo, em propagandas.



TEORIA DA AMOSTRAGEM

Frequentemente não é possível ou viável a observação de todos os elementos de uma população devido a vários fatores:

- Populações geralmente são muito grandes
- Custo e demora dos censos
- Impossibilidade física de examinar toda a população
 - O estudo destrói as unidades (teste de qualidade dos fósforos em uma fábrica)
- Comprovado valor científico das informações coletadas por meio de amostras

Nesses casos deve-se fazer uso de amostras. A finalidade de uma amostra é a de descrever, indiretamente, a população.

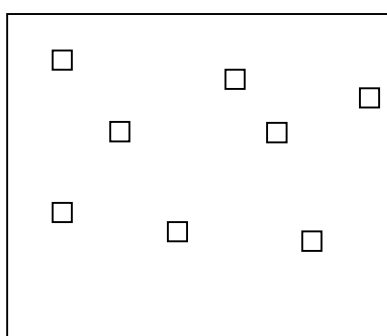
1. Unidade de Amostra e Amostra

Como na maioria dos casos a população objeto de estudo é demasiadamente grande para ser completamente enumerada ou mensurada, é possível estudar essa população por meio de uma seleção de elementos do total da população

- A seleção se chama amostra (conjunto das unidades de amostra).
- As unidades selecionadas se chamam unidades amostrais (área utilizada para a coleta de dados)

Amostra: É qualquer conjunto de elementos, retirado da população, desde que esse conjunto seja não vazio e tenha menor número de elementos do que a população.
É um subconjunto finito da população.

Ex.: Viveiro de mudas



Variável = altura média das mudas
Área do viveiro = 4,0 ha
Unidade de amostra = parcela de 10 x 10 m
Amostra = 8 unid. x 100 m² = 800 m²

O número total de elementos de uma amostra é simbolizado pela letra minúscula "n"

200 m

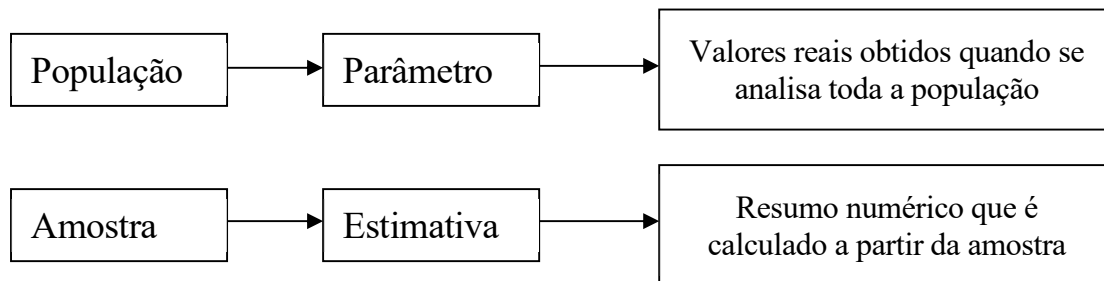
2. Características da Amostra

-- Deve ser representativa, ou seja, deve conter em proporção tudo o que a população possui qualitativa e quantitativamente (mesmas características básicas da população).

-- Deve ser imparcial, isto é, todos os elementos da população devem ter igual oportunidade de fazer parte da amostra.

Queremos	Fazemos
Representatividade	Análise da população para ver se seus elementos distribuem-se homogeneamente, ou se formam grupos com características peculiares. Se for esse o caso, temos que respeitar a proporção com que esses grupos integram a população
Imparcialidade	Sorteio dos elementos que farão parte da amostra

Obs.: Se não houver representatividade e imparcialidade, a amostragem torna-se tendenciosa ou viciada.



Exemplo:

Suponhamos que a nossa população seja apenas um grupo de 5 alunos de Bioestatística e a variável de interesse seja o número de horas por semana dedicadas ao estudo da Bioestatística (veja tabela abaixo).

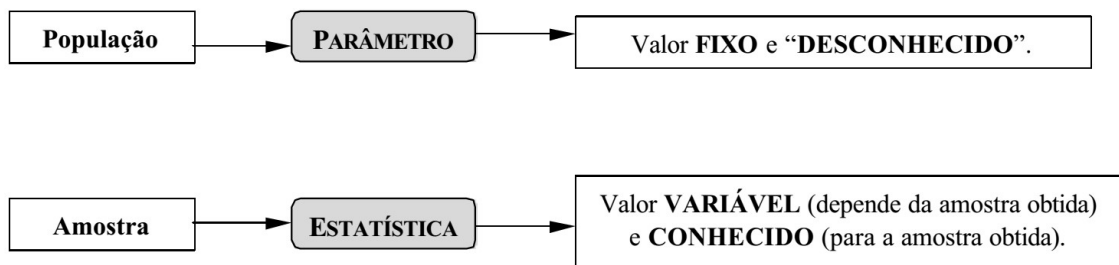
Aluno	Horas de Estudo
1	1
2	3
3	4
4	2
5	5

1. Calcule o tempo médio dedicado à Bioestatística nesta população (Parâmetro).

$$X = 3,0 \text{ horas}$$

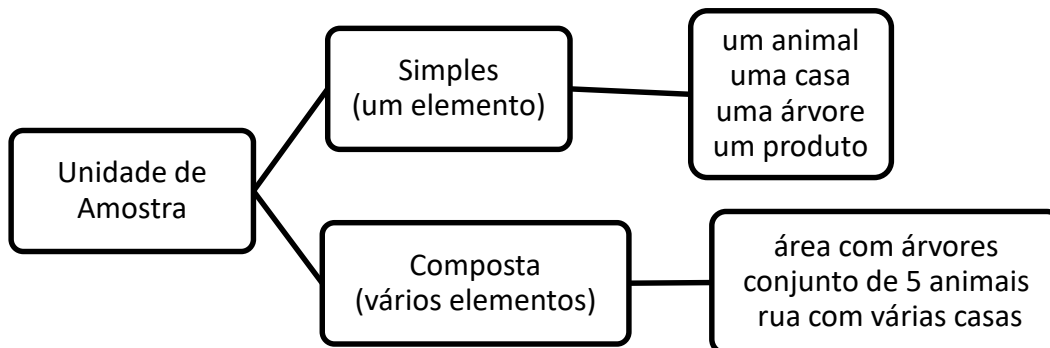
2. Tomemos uma amostra de tamanho 2 desta população, qual o valor da Estatística se:
- Os alunos 1 e 4 forem selecionados? $X = 1,5$ hs
 - Os alunos 2 e 3 forem selecionados? $X = 3,5$ hs
 - Os alunos 1 e 5 forem selecionados? $X = 3,0$ hs

O Valor do Parâmetro e da Estimativa são sempre iguais?



3. Unidade Amostral

É a unidade de área ou espaço físico utilizado para a coleta dos dados



- Cada unidade amostral gera uma única observação da variável de interesse.
- No caso de amostras compostas o valor da unidade amostral pode ser a média ou a somatória dos elementos que a compõe

4 Intensidade de Amostragem ou Fração Amostral

É a razão entre o tamanho da amostra e o tamanho total da população.

População = N (unidades de amostra de igual tamanho)

Amostra = n (seleção de unidades de amostra)

$$F = \frac{n}{N} \cdot 100$$

Exemplo 1:

População: Aviário com 3500 frangos
 Amostra: 150 animais (amostra simples)
 Variável: peso aos 20 dias (quantitativa contínua)
 Qual a fração amostral?

$$F = \frac{n}{N} \cdot 100 \qquad F = \frac{150}{3500} \cdot 100 \qquad F = 4,28\%$$

Exemplo 2

População: área de floresta com 4540 ha
 Amostra: 30 parc. de 20m x 100 m (amostra composta)
 Variável: volume das árvores (quantitativa contínua)
 Qual a fração amostral?

n= 30

N= ?

Tamanho da parcela: 20m x 100m = 2000m² (transformar em ha)
 2.000/10.000 = 0,2ha

N= quantas parc de 0,2ha cabem em 4540 ha?

$$N = \frac{\text{tam. da população (ha)}}{\text{tam. da parcela (ha)}} \qquad N = \frac{4540}{0,2} \qquad N = 22.700 \text{ parc.}$$

$$F = \frac{n}{N} \cdot 100$$

$$F = \frac{30}{22.700} \cdot 100 \qquad F = 0,13\%$$

Exemplo 3

Pesquisa sobre o uso de defensivo agrícola
 População: 2180 propriedades agrícolas de Belterra
 Amostra: 50 propriedades
 Variável qualitativa ordinal: (1) não usa defensivo, (2) usa raramente, (3) usa moderadamente, (4) usa frequentemente
 Qual a fração amostral?

Exemplo 4

População: área de floresta natural com 6230 ha

Amostra: 45 parc. de 20m x 250 m (amostra composta)

Variável: número árvores/ha

Qual a fração amostral?

Exercício

1. Analise a sentença e identifique o que se pede

Uma instituição de pesquisa deseja conhecer a composição florística e a estrutura de uma área de 5.750 ha de floresta natural. Para isso contratou um equipe de eng. Florestais que instalaram de modo totalmente aleatório 38 parcelas de 1ha onde foram identificadas e medidas o diâmetro de todas as árvores com Diâmetro > 10,0 cm. Pede-se:

População:

Unidade amostral:

Variáveis:

Classificação das Variáveis:

Amostra:

Fração Amostral:

2. O diretor de uma escola, na qual estão matriculados 50 alunos na 1^o série, 47 na 2^o série, 35 na 3^o, 29 na 4^o e 54 na 5^o série, pretende fazer uma pesquisa sobre as condições de vida extra escolar de seus alunos. Não dispondo de tempo para entrevistar todas as famílias, resolveu fazer uma amostragem de 15% dos alunos. Obtenha para esse diretor, os elementos componentes da amostra.

Serie	População	Amostra
1	50	
2	47	
3	35	
4	29	
Total		

5 Principais Técnicas de Amostragem

De acordo com a técnica usada para selecionar as unidades de amostra tem-se:

- Amostra aleatória, casual, ou probabilística
 - Aleatória Simples
 - Aleatória Estratificada
- Amostra semi-probabilística
 - Amostra sistemática
 - Amostra por conglomerados
 - Amostra por quotas.
- Amostra não probabilística ou de conveniência.

5.1 Amostra Aleatória, Casual ou Probabilística

- É constituída por n unidades retiradas ao acaso da população (sorteio)
- A população deve ser conhecida e cada unidade deve ser identificada por nome ou por número
- Todas as unidade da população tem a mesma probabilidade de pertencer à amostra.

5.1.1 Amostra Aleatória Simples

- É usada quando a população é constituída por **unidades homogêneas** para a variável que você quer estudar.
- Na prática a amostragem aleatória pode ser realizada por meio de sorteio direto ou utilizando-se uma tabela de números aleatórios.

Exemplo 1

Pesquisa sobre a produção de leite de vacas da Raça Holandesa

População: 850 vacas de idade e peso aproximados (população homogênea)

Amostra: 3% do rebanho

Variável: litros de leite /dia (quantitativa contínua)

Como obter esta amostragem?

1. Calcular o valor de n

$$\begin{array}{l} 850 \text{ -----}100\% \\ X \text{ -----}3\% \end{array}$$

$$n = 25,5$$

$$n = 26,0 \text{ vacas}$$

$$F = \frac{n}{N} \cdot 100$$

$$3 = \frac{n}{850} \cdot 100$$

$$3 \times 850 = n \times 100$$

$$n = \frac{2550}{100}$$

$$n = 25,5$$

$$n = 26,0 \text{ vacas}$$

2. Numerar as vacas da população (1 a 850)

3. Sorteio dos 26 animais que irão compor a amostra
 - Via manual
 - Pelo computador (exemplo utilizando o programa BioEstat 5.3 - gratuito)
4. Medir a produção de leite dos animais selecionados por algum tempo (15 dias)
5. Calcular as estatísticas com os valores da amostra (media, variância, desvio, etc....)
6. Concluir sobre a produção de leite de toda a população (850 animais) - **Inferência**

5.1.2 Amostra aleatória estratificada

- É usada quando a população é constituída por **unidades heterogêneas** para a variável que se quer estudar, ou seja, a população apresenta estratos ou grupos distintos
- A amostragem estratificada além de considerar a existência de estratos, obtém os elementos da amostra de maneira proporcional ao número total de elementos do mesmo

Exemplo 1

Pesquisa na comunidade da UFOPA

3 estratos: Professores N = 30
 Alunos N = 560
 Funcionários N = 40

N total = 630

Intensidade amostral = 8%

Seguir o exemplo e calcular o n por extrato

$$F = \frac{n}{N} \cdot 100$$

$$8 = \frac{n}{630} \cdot 100$$

$$8 \times 630 = n \times 100$$

$$n = 50,4$$

$$n = 60,0$$

Geral

Exemplo 2

Pesquisa sobre a produção de milho em Mojuí dos Campos

População: 1670 propriedades (população heterogênea)

- 436 com irrigação
- 1234 sem irrigação

Amostra: 5% das propriedades

Variável: sacas de milho/ha

Como obter esta amostragem?

1. Calcular o valor de n por estrato
2. Numerar as propriedades de 1 a n em cada estrato
 - Estrato 1 (irrigado) : 1 a 436 (sortear 22 propriedades)
 - Estrato 2 (sem irrigação): 1 a 1234 (sortear 62 propriedades)
3. Sorteio
 - Via manual
 - Pelo computador (**exemplo utilizando o programa BioEstat 5.3**)
4. Medir a produção de milho nas propriedades selecionadas em cada estrato
5. Calcular as estatísticas com os valores da amostra em cada estrato (média, variância, desvio, etc....)
6. Concluir sobre a produção de milho em cada estrato e comparar as médias (**Inferência**)

5.2 Amostra Semiprobabilística

- É constituída por n unidades retiradas da população por procedimento *parcialmente aleatório*
- Dentre as amostras semiprobabilísticas, temos:
 - Amostra sistemática
 - Amostra por conglomerados
 - Amostra por quotas.

5.2.1. Amostra Sistemática

- Os elementos são selecionados para a amostra por um sistema preestabelecido
- Não exige que a população seja conhecida, mas é preciso que esteja organizada de alguma forma (filas, arquivos, ordem de produção, ruas, etc.)

Ex.: Linhas de produção – A cada 20 itens retirar 1 para compor a amostra (5%)

Plantio de Pinus -- A cada 50 plantas retirar 1 para avaliação (2% de amostragem)

A cada 30 casas visitar uma para entrevista sobre condição socioeconômica (3,3%)

Exemplo 1

Pesquisa sobre conhecimento de educação ambiental com alunos do 9º ano do EF

População: 230 alunos concluintes de ensino fundamental de uma escola

Amostra: 12% dos alunos para fazer uma entrevista

Como obter esta amostragem?

1. Calcular o valor de n

$$\begin{array}{l} 230 \text{ ----}100\% \\ X \text{ -----}12\% \end{array}$$

$$n = \frac{230 \times 12}{100}$$

$$n = 27,6$$

$$n = 28,0$$

2. Numerar os alunos da população (1 a 230)

3. Calcular o intervalo sistemático: dividir a (N) população pelo (n) amostra

$$\text{Intervalo} = \frac{230}{28}$$

Inter. = 8,21

Inter. = 8,0 (arredondar)

4. Sortear um numero entre 1 e 8 (valor do intervalo) Ex: 5

Valores:

5

5+8 = 13

13+8 = 21

21+8 = 29

29+8 = 37

37+8 = 45 (assim sucessivamente)

5.2.2. Amostra por Conglomerados

- O conglomerado é um conjunto de unidades que estão agrupadas, qualquer que seja a razão.
- A amostra por conglomerados é constituída por n unidades de conglomerados selecionados ao acaso
- Todos os itens dentro dos conglomerados selecionados farão parte da amostra

Ex.: um asilo é um conglomerado de idosos

uma sala de aula é um conglomerado de estudantes

uma casa é um conglomerado de pessoas

uma comunidade: um conglomerado de propriedades agrícolas

Exemplo

Pesquisa sobre uso de adubação orgânica no município de Santarém

População: 15 comunidades agrícolas (15 conglomerados)

Amostra: 20%

Como obter esta amostragem?

1. Calcular o número de conglomerados da amostra

15 -----100%

X ----- 20%

$$n = \frac{15 \times 20}{100}$$

n = 3,0 conglomerados

2. Sortear 3 comunidades dentre as 15 possíveis

3. Realizar a pesquisa em todas as propriedades presentes nas 3 comunidades sorteadas

OBS: O controle é realizado por conglomerado, o número de propriedades não é conhecido previamente

5.2.3 Amostra por Quotas

- É constituída por n unidades retiradas da população segundo *quotas* estabelecidas de acordo com a distribuição desses elementos na população
- É semelhante a amostragem por estratos, porém não existem sorteios
- É muito usada em levantamentos de opinião e pesquisas de mercado

Para cada entrevistador é atribuída uma cota de entrevistas e este escolherá pessoas que estejam dentro do perfil da pesquisa

Exemplo:

Pesquisa sobre questões de gênero na produção agrícola

Pesquisas indicam: proporção de mulheres é de 47% e de homens é 53% na zona rural

Amostra com $n=50$

Como obter esta amostragem?

1. Calcular o n de acordo com a proporção na população

$$\begin{array}{l} 50 \text{ ----} 100\% \\ X \text{ ----} 47\% \end{array}$$

$$n_1 = \frac{50 \times 47}{100}$$

$$n_1 = 23,5$$

$$n_1 = 24 \text{ mulheres}$$

$$n_2 = 50 - 24$$

$$n_2 = 26 \text{ homens}$$

2. Realizar a pesquisa com entrevistas seguindo a quota de 24 mulheres e 26 homens

5.3 Amostra Não Probabilística ou de Conveniência

- É constituída por n unidades reunidas em uma amostra levando em conta somente a facilidade de acesso do pesquisador
- Os indivíduos são selecionados para compor a amostra porque estão prontamente disponíveis, sem seguir critério estatístico de aleatoriedade

Exemplos:

- Teste de medicamentos com ratos de um laboratório
- Teste de tratamento com pacientes internados com determinada doença

- Parar pessoas no supermercado e colher suas opiniões
- Solicitar a pessoas que voluntariamente testem um produto e que em seguida respondam a uma entrevista
- Colocar linhas de telefone adaptadas para que durante um programa de televisão os telespectadores possam dar suas opiniões (big brother)

Atenção

As amostras de conveniência não invalidam a pesquisa, mas precisam ser *muito bem descritas* porque representam apenas a população de indivíduos semelhantes àqueles incluídos na amostra, ou seja, não podem ser generalizadas para toda a população

Exercícios

1. Identifique na sentença abaixo o que se pede

Um pesquisador pretende estudar a produção de óleo de copaíba em uma área de 5.000 ha da Flona Tapajós. Para isso selecionou aleatoriamente 25 árvores de copaíba dentre as 82 presentes na área e observou o diâmetro da árvore e a produção do óleo em ml.

- a) População:
- b) Unidade Amostral:
- c) Amostra
- d) Variável
- e) Classificação da variável
- f) Intensidade amostral

Um produtor de porcos possui em sua pocilga 1200 animais divididos em 190 da raça Landrace, 120 da raça Canastra, 750 da raça Piau e 140 da raça Caruncho. O produtor precisa realizar exames de sangue para avaliar a presença de anemia. Não dispondo de recursos para realizar os exames em toda a população resolveu examinar somente 58 animais.

- a) População:
- b) Unidade Amostral:
- c) Amostra
- d) Variável
- e) Classificação da variável
- f) Intensidade amostral

INFERÊNCIA ESTATÍSTICA

Também conhecida como estatística indutiva, é aquela que partindo de uma amostra, estabelece hipóteses sobre a população de origem e formula previsões, fundamentando-se na teoria das probabilidades

As pesquisas são feitas com o objetivo de responder perguntas.

Ex:

- A Copaíba produz mais óleo no inverno (chuvas) ou no verão (seca)?
- Irrigar o plantio de feijão é viável economicamente?
- Ração com sais minerais provoca maior ganho de peso em suínos em relação a ração comum?
- Secar madeira de maçaranduba em estufa provoca rachaduras nas tábuas?
- Qual a dose de adubo NPK mais adequada para produzir legumes?

Para responder perguntas, são necessárias informações que, na área das ciências agrárias, são quase sempre obtidas por meio de **amostras**.

Mas os pesquisadores querem **generalizar** seus achados para **toda a população** de onde a amostra foi retirada. Isto pode ser feito, desde que a generalização seja fundamentada em um **teste de hipóteses**.

O cientista “conversa” com a natureza por meio de hipóteses e “obriga-a” a dar-lhe respostas por meio de experimentos

Hipótese: É uma conjectura, uma resposta provisória que, de acordo com certos critérios, será rejeitada ou não rejeitada

Obs.: A natureza não é determinista: as mesmas causas nem sempre produzem os mesmos efeitos. Por isso, em lugar de **aceitar**, o pesquisador prefere ser mais cauteloso e dizer **não rejeitar**, associando a essa última expressão uma probabilidade.

Exemplos de Hipóteses Estatísticas (*afirmativa provisória*)

- A produção de óleo pelas árvore da espécie copaíba é maior no inverno (chuvas)
- A irrigação não produz aumento na produtividade de feijão “fradinho” capaz se superar os custos
- Cães criados em cativeiro são mais agressivos do que os criados com contato social
- Milho produzido após método “roça e queima” produz mais que o método “roça, tritura e incorpora a matéria orgânica no solo”

OBS: A Hipótese é sempre formulada em relação a população

Experimento: É uma situação criada, construída pelo pesquisador com o objetivo de testar uma hipótese

Ex:

- Tipos e dosagens de inseticidas, adubos, etc
- Profundidade de semeadura
- Tipos de máquinas ou equipamentos mais adequados a extração de madeira
- Efeito do desbaste no crescimento das arvores remanescentes
- Efeito da luz no crescimento de mudas

A lógica da natureza não se confunde com a lógica formal

- Lógica formal: trabalha com certezas
- Lógica da natureza: trabalha com probabilidades

A prova de hipóteses é um recurso **lógico-probabilístico**, e onde há probabilidade há **risco de erro**

A observação faz parte do procedimento experimental, o pesquisador nunca **“acha nada”**: ele observa, mede, descreve, analisa e conclui. Ele jamais atribui ao que está vendo **Juízos de valor**

Os Juízos podem ser de valor ou de realidade

- **Juízos de Valor:** Acrescenta a coisa um atributo

Ex: madeira bonita – o atributo não está na madeira, mas na pessoa que a considera com tal.

- **Juízo de Realidade:** É **objetivo** e pode **ser comprovado**. Replicando a mensuração, testa-se objetivamente o juízo de realidade

Ex: madeira com densidade $0,72 \text{ g/cm}^3$

As hipóteses classificam-se em:

- **Científicas:** A sua veracidade ou a sua falsidade podem ser postas a prova, direta ou indiretamente, por procedimento experimental
- **Não Científicas:** A sua veracidade ou a sua falsidade não podem ser postas a prova por procedimento experimental

Ex: Hipóteses teológicas e filosóficas

Exemplo: Uma hipótese científica

Para verificar se meninos e meninas aprendem a falar na mesma idade, um pesquisador observou uma amostra de 20 crianças e anotou a idade em que cada uma delas começou a falar frases curtas de até 4 palavras.

Meninos	Meninas
1,5	1,8
2,2	2,4
3,0	2,0
3,1	2,5
3,6	3,2
1,9	2,8
1,8	1,9
1,5	1,7
2,7	1,4
2,6	2,0
$\bar{x} =$	

- As médias são matematicamente diferentes, ou seja, na amostra observada as meninas falaram mais cedo que os meninos
- Será que este resultado se repetiria se outra amostra de crianças fosse selecionada ?
- Será que a diferença de 0,2 entre as médias é suficiente para afirmar que meninas (população geral) falam mais cedo?
- Será que esta diferença pequena entre as médias não foi “por acaso”?

O teste estatístico de hipóteses pode responder a estas perguntas

Diante de um problema científico, o procedimento habitual do pesquisador é o seguinte:

- Formular uma hipótese que atribua ao acaso a ocorrência do fenômeno observado (Hipótese de trabalho H_0)
- Formular outra hipótese, que sirva de alternativa à primeira, se ficar demonstrado que o acaso não pode ser responsável pelo fenômeno observado (Hipótese alternativa H_1)

Os testes estatísticos permitem decidir, com base nos dados, se há evidência suficiente para rejeitar a hipótese da nulidade.

1. TESTES ESTATÍSTICOS DE COMPARAÇÃO DE POPULAÇÕES

Existem dois testes em estatística que são usados para avaliar diferenças entre médias

- 1 Teste T: Usado quando se compara 2 médias
- 2 Teste F: Usado quando se compara 2 ou mais médias

1.1. TESTE T

Utilização:

- A maioria das comparações que se fazem em estatística experimental são realizadas entre dois grupos de amostras. Ex.:

Ex: Testemunha x Tratados, Mulheres x Homens, Exploração tradicional x Exploração controlada

- Quando desejamos saber se a diferença numérica entre duas médias é suficiente para se dizer que as médias vêm de duas populações diferentes ou se as médias são da mesma população.

O teste T é usado para comparar as médias de dois grupos de dados e se baseia na razão entre a diferença entre esses dois grupos.

O teste T foi criado por um estatístico Irlandês chamado William Gosset, no início do século.

Testa duas hipóteses:

H₀: A diferença entre as médias de dois grupos de dados é igual a zero (ou desprezível segundo um determinado nível de probabilidade), ou seja, **os dois grupos são iguais**

$$\mu_1 - \mu_2 = 0$$

H₁: A diferença entre as médias de dois grupos de dados é diferente de zero (o valor da diferença é estatisticamente significativo), ou seja, **os dois grupos são diferentes.**

$$\mu_1 - \mu_2 \neq 0$$

Teste T (procedimento geral)

1. Com base nos dados amostrados se calcula o valor absoluto do “T calculado” (fórmula)
2. Comparar o T calculado com o “valor crítico” dado na Tabela de valores de t, no nível estabelecido de significância

$$T_{calc.} \times T_{tab.}$$

Tabela T

TABELA 6
Valores de t, segundo os graus de liberdade e o valor de α

<i>Graus de liberdade</i>	α		
	10%	5%	1%
1	6,31	12,71	63,66
2	2,92	4,30	9,92
3	2,35	3,18	5,84
4	2,13	2,78	4,60
5	2,02	2,57	4,03
6	1,94	2,45	3,71
7	1,90	2,36	3,50
8	1,86	2,31	3,36
9	1,83	2,26	3,25
10	1,81	2,23	3,17
11	1,80	2,20	3,11
12	1,78	2,18	3,06
13	1,77	2,16	3,01
14	1,76	2,14	2,98
15	1,75	2,13	2,95
16	1,75	2,12	2,92

Grau de liberdade: é sempre relativo ao tamanho da amostra utilizada

$$GL = n - 1$$

α = nível de significância (referente a probabilidade)

OBS.: As diferenças ou igualdades do ponto de vista estatístico são baseadas em probabilidades, ou seja, aceita-se ou rejeita-se hipóteses de acordo com um nível probabilístico

Em experimentação agrária, os níveis probabilísticos usados são 5% ou 1%

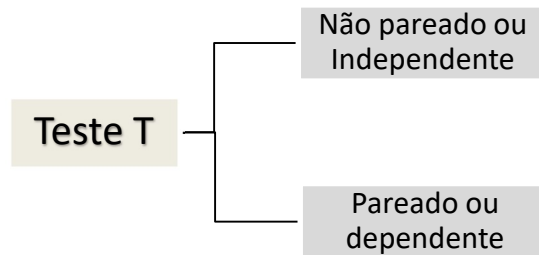
- $\alpha = 0,05$ ou 5% : Significa **95%** de probabilidade de **acerto** na afirmativa e **5%** de chance de **erro**
- $\alpha = 0,01$ ou 1% : Significa **99%** de probabilidade de **acerto** na afirmativa e **1%** de chance de **erro**

Representação

- * Quando a diferença é significativa ao nível de 5% de probabilidade (rejeita H0 e aceita-se H1);
- ** Quando a diferença é significativa ao nível de 1% de probabilidade (Rejeita H0 e aceita-se H1);
- NS Não significativo, quando não há diferença significativa (Aceita H0)

Teste T: Usado quando se compara 2 médias (duas situações ou duas populações)

Existem duas maneiras de realizar o teste T



A escolha vai depender de como as parcelas de amostragem foram planejadas

1.1.1 TESTE T NÃO PAREADO OU INDEPENDENTE

Usado quando não há restrição na casualização do experimento. Um fator de fundamental importância na aplicação do teste T é a casualização do material estudado. Se as amostras tomadas não representam a população, os resultados serão tendenciosos, ou seja, não serão válidos

Uma vantagem do teste T não pareado é a possibilidade de se poder utilizar um número diferente de observações em cada amostra.

Fórmula para o teste T independente

$$T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_p^2}}$$

Onde:

\bar{x}_1 = média do tratamento ou condição 1

\bar{x}_2 = média do tratamento ou condição 2

n_1 = número de repetições no tratamento 1

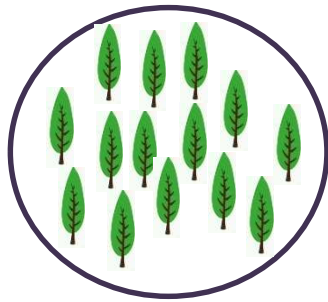
n_2 = número de repetições no tratamento 2

s_p^2 = Variância ponderada (envolve a variabilidade existente nas duas amostras)

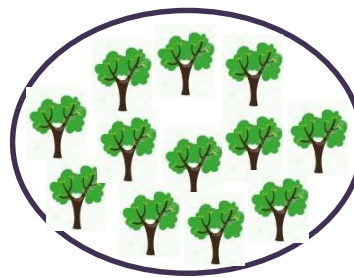
Quando utilizar a comparação de duas amostras independentes?

- Quando existem dois grupos cujas amostras foram coletadas de forma independente. Esses grupos podem ter sido formados de duas maneiras diferentes:

a) Extraíu-se uma amostra da população A e outra amostra da população B

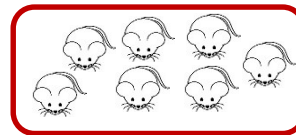
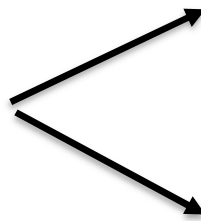
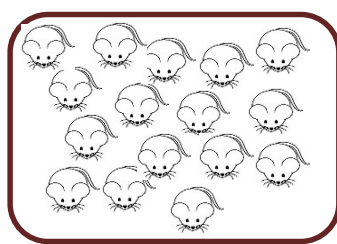


População A

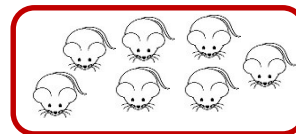


População B

b) Indivíduos da mesma população foram alocados aleatoriamente a um dos dois tratamentos em estudo.



Tratamento A



Tratamento B

Condições para a aplicação do teste T não pareado:

1. As amostras retiradas da de cada população devem ser aleatórias
2. As variâncias devem ser iguais (homogeneidade de variâncias)
3. As variáveis das populações de onde as amostras foram selecionadas devem apresentar distribuição aproximadamente normal (normalidade dos dados)

Exemplo

Um Engenheiro Florestal deseja saber se existe diferença entre duas procedências de Pinus (procedência A e procedência B) Os dados obtidos foram:

Procedência	Incremento Medio (st/ha.ano)	Média	Variância
A	45.6, 42.1, 44.9, 45.1, 47.6, 46.7, 45.5, 48.9	45.8	4.11
B	30.1, 21.6, 27.6, 27.3, 30.4, 31.4, 34.1, 30.6	29.1	13.85

Comparando as Médias

Hipóteses:

H0 : média A = média de B (as duas espécies de pinus produzem lenha de maneira igual)

H1 : média A \neq média de B (as duas espécies de pinus produzem lenha de maneira diferente)

Estatística:

Variância ponderada

$$s_p^2 = \frac{(n1 - 1)s_1^2 + (n2 - 1)s_2^2}{n1 + n2 - 2}$$

$$s_p^2 = \frac{(8 - 1)4,11 + (8 - 1)13,85}{8 + 8 - 2} = 8,98$$

Valor de T calculado

$$T = \frac{(x_1 - x_2)}{\sqrt{\left(\frac{1}{n1} + \frac{1}{n2}\right) s_p^2}}$$

$$T = \frac{(45,8 - 29,1)}{\sqrt{\left(\frac{1}{8} + \frac{1}{8}\right) 8,98}} = 11,15$$

O valor do T calculado é então comparado ao valor do T tabelado com os graus de liberdade correspondentes e nível de probabilidade.

Graus de Liberdade:

$$GL = (n1 - 1) + (n2 - 1)$$

$$GL = (8 - 1) + (8 - 1) = 14$$

$$T_{(14;0,05)} = 2,145$$

Se:

- T calculado < T tabelado: Aceita-se a hipótese H0, ou seja, as médias são estatisticamente iguais.
- T calculado \geq T tabelado: Rejeita-se a hipótese H0 e aceita-se H1, ou seja, as médias são estatisticamente diferentes.

Conclusão: Como T calculado > T tabelado, rejeita-se H0 ao nível de 5% de probabilidade, conclui-se que há evidências para se afirmar que as procedências têm produção diferente.

Exercícios

1. Um pesquisador deseja verificar se o melhoramento genético produziu redução marcante no grau de rachadura de topo de *Eucalyptus saligna*. Comparando dois grupos de árvores o pesquisador obteve os dados na tabela abaixo. Teste as hipóteses apropriadas e estabeleça as suas conclusões

Número de Rachaduras / cm²:

árvores não melhoradas: 0.2, 6.6, 8.0, 2.2, 1.9, 6.8, 1.7, 6.3, 1.4

árvores melhoradas: 2.6, 4.6, 4.8, 4.3, 4.4, 3.1, 3.8

2. Teste de duas procedências de *Eucalyptus grandis* em determinada região. Os resultados foram obtidos após 5 anos do plantio. A Variável estudada é o Volume

Procedências	Repetições
A	19, 14, 15, 17, 20
B	23, 19, 19, 21, 18

1.1.2 TESTE T PAREADO

- Usado quando há restrição na localização das parcelas no experimento. Essa restrição faz com que cada parcela de um tratamento corresponda a uma parcela de outro tratamento.
- Pares de parcelas a serem comparadas

Exemplos:

- Psicologia – estudos com paras de gêmeos, um dos gêmeos recebe o tratamento, enquanto o outro permanece como controle
- Para estudar o efeito de um tratamento para prevenção das cáries, o dentista pode aplicar o tratamento em um lado da arcada dentária de cada paciente e deixar o outro lado com controle.
- Observação de um mesmo indivíduo duas vezes (antes e após a aplicação do tratamento)
 - Para verificar o efeito de um tratamento sobre a pressão arterial, o médico pode obter a pressão de seus pacientes antes e após a aplicação do tratamento.
 - Estudos de crescimento de floresta, medições no tempo T0 (antes da aplicação do tratamento) e T1 (um ou mais anos após a aplicação do tratamento)

Fórmula para o teste T pareado

$$T = \frac{(\bar{x}_1 - \bar{x}_2) \cdot \sqrt{n}}{s_d}$$

Onde:

\bar{x}_1 = média do tratamento ou condição 1

\bar{x}_2 = média do tratamento ou condição 2

n = número de pares de amostras

s_d = Desvio padrão da diferença entre as duas amostras (envolve a variabilidade existente nas duas amostras)

$$s_d^2 = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n - 1}$$

Exemplo

Experimento para testar se a densidade da madeira de Pinus é afetada pela posição da árvore no plantio, se na borda ou no centro. A tabela abaixo representa os dados de parcelas contíguas onde:

T1 = árvores na borda do plantio

T2 = árvores a uma distância fixa de 30 m da borda

Dados da densidade básica da madeira de Pinus, expresso em Kg/m³ em dois tratamentos

Parcelas	T1 borda	T2 30 m da borda	di = T1 - T2	d ²
1	412	308	104	10816,0
2	408	401	7	49,0
3	396	381	15	225
4	422	322	100	10000,0
5	400	364	36	1296,0
6	381	310	71	5041,0
7	402	380	22	484,0
8	431	356	75	5625,0
9	436	390	46	2116,0
10	420	326	94	8836,0
TOTAL	4108	3538	570	44488,0

$$T = \frac{(\bar{x}_1 - \bar{x}_2) \cdot \sqrt{n}}{s_d}$$

A variância da diferença é calculada por:

$$s_d^2 = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n - 1}$$

$$s_d^2 = \frac{44488 - \frac{324900}{10}}{10 - 1} = 1333,11$$

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

$$s_d = \sqrt{1333,11} = 36,51$$

$$T = \frac{(410,8 - 353,8) \cdot \sqrt{10}}{36,51} = 4,94$$

O valor do T calculado é então comparado ao valor do T tabelado com os graus de liberdade correspondentes e nível de probabilidade.

Graus de Liberdade:

$$GL = (n - 1)$$

$$GL = 9$$

$$T_{(9;0,05)} = 2,262$$

Conclusão: Como T calculado > T tabelado, rejeita-se H0 ao nível de 5% de probabilidade, conclui-se que há evidências para se afirmar que a madeira de Pinus é afetada pela posição no plantio pois a densidade média da madeira é diferente.

Exercícios

- 1) Com o objetivo de verificar o efeito da exposição do solo sobre a microfauna do solo, uma pesquisadora fez levantamentos de microfauna antes e depois da exposição do solo em áreas desmatadas de vários tipos de ecossistemas (tabela abaixo). Teste as hipóteses apropriadas e estabeleça a sua conclusão

Ecossistema	No. de Micro-organismos/cm ³	
	Antes	Depois
Campo Limpo	1430	780
Cerrado	2500	1020
Restinga	732	640
Caatinga	640	680
Floresta Estacional	10530	2520
Floresta Pluvial	21883	2302