

Estatística Aplicada ao Serviço Social

Prof^a. Juliana Freitas Pires

Departamento de Estatística
Universidade Federal da Paraíba - UFPB
juliana@de.ufpb.br

O que é Estatística?

Coleção de métodos para planejar experimentos, obter e organizar dados, resumi-los, analisá-los interpretá-los e deles extrair conclusões.

Papel da estatística



Perguntas



Estudos



Dados

Exemplos:

- Qual a idade média das mulheres ao engravidarem pela primeira vez?
- O número de homicidas é maior entre os jovens?
- Investir em aperfeiçoamento dos funcionários contribui para o crescimento de uma empresa?
- Quem vencerá as próximas eleições?
- Participação de idosos em programas culturais melhora sua qualidade de vida?
- Qual o nível de escolaridade mais frequente dentre os paraibanos?

Os dois ramos da estatística:

Estatística Descritiva: Trata da organização, resumo e apresentação dos dados.

Estatística Inferencial: A partir de uma amostra, tirar conclusões sobre a população.

Objetivo da estatística descritiva

- Na estatística descritiva o objetivo é resumir os dados coletados de forma a extrair destes, conhecimento útil acerca do problema que gerou os dados.
- Nessa fase da pesquisa, estamos preocupados em apresentar os dados em forma de tabelas e gráficos e em obter medidas que quantifiquem os resultados do estudo.

População: é o conjunto de elementos, indivíduos ou objetos que se pretende estudar.

Amostra: é qualquer subconjunto de elementos de uma população. (n: número de indivíduos que constituem uma amostra)

Exemplo: Uma empresa de construção civil possui 500 funcionários. Desejando avaliar as condições de saúde, segurança e satisfação do trabalhador, a assistente social seleciona 50 funcionários aleatoriamente.

População: ?

Amostra: ?

População: 500 funcionários da empresa.

Amostra: 50 funcionários selecionados aleatoriamente.

Obs: a amostra deve ser selecionada seguindo certas regras e deve ser representativa, de modo que ela represente todas as características da população como se fosse uma fotografia desta.

Censo X Estudos por amostragem

Na pesquisa estatística a forma de coleta dos dados pode ser feita através de censo ou amostragem.

Censo: quando todos os indivíduos de uma população são pesquisados.

Amostragem: quando utilizamos uma amostra de indivíduos de uma determinada população. O processo de retirada de informações dos “n” elementos amostrais, deve seguir um método criterioso e adequado (tipos de amostragem).

Censo X Estudos por amostragem

Um estudo por amostragem é preferível a um censo por diversos motivos, dentre os quais:

- Menor custo;
- Rapidez;
- Resultados muito próximos aos do censo.

Parâmetro: Descrição numérica de uma característica da população.

Ex: Média populacional, Mediana populacional, Desvio Padrão populacional.

Estimador: Característica numérica estabelecida para uma amostra.

Ex: Média amostral, Mediana amostral, Desvio Padrão amostral.

Estimativa: Valor numérico assumido por um estimador numa determinada amostra.

Unidade: qualquer indivíduo, elemento ou objeto que faça parte do conjunto a ser estudado.

Obs: Podem ser pessoas, domicílios, escolas, creches, células ou qualquer outra unidade.

Variável Estatística: característica das unidades sobre as quais queremos obter informações.

Obs: As variáveis podem ser classificadas em quantitativas ou qualitativas.

Tipos de Variáveis

Qualitativa: consistem em atributos, classificações ou registros não numéricos.

Ex: Casse Social, Grau de Instrução, Tipo Sanguíneo, Porto de Embarque.

Quantitativa: constituem em medidas ou contagens numéricas.

Ex: Peso, Idade, Altura, Renda, Taxa de analfabetismo.

Nominal: Os indivíduos são classificados em categorias que não possuem ordem.

Ex: Gênero, Bairro, Curso, etc.

Ordinal: Os indivíduos são classificados em categorias que possuem algum tipo inerente de ordem.

Ex 1: Nível de Escolaridade (Analfabeto < Ensino Fundamental < Ensino Médio < Ensino Superior),

Ex 2: Escala de Qualidade (Péssimo < Ruim < Regular < Bom < Ótimo), etc.

Variáveis Quantitativa

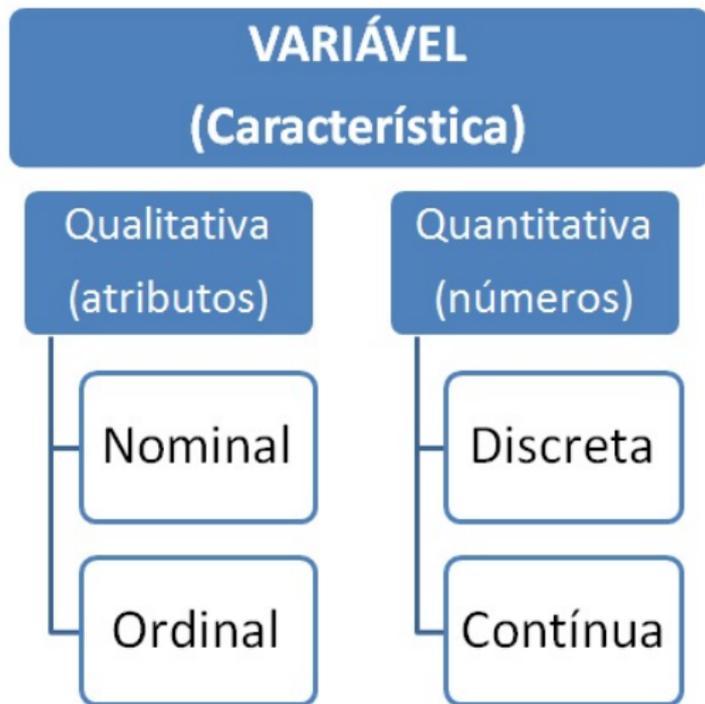
Discreta: é a variável que só pode assumir valores pertencentes a um conjunto enumerável. Normalmente resulta de enumerações ou contagens.

Ex: Número de Filhos, Número de Acidentes de Transito, etc.

Contínua: é a variável que pode assumir infinitos valores dentro de um intervalo finito. Resulta, em geral, de medições.

Ex: Peso, Altura, Renda, etc.

Tipos de Variáveis



Uma empresa realizou uma pesquisa junto a seus funcionários. Os funcionários responderam a um questionário que originou a seguinte tabela:

Tabela : Informações sobre estado civil, sexo, idade, número de filhos, salário e grau de instrução dos funcionários da empresa Z

| Nº | Estado Civil | Sexo | Idade (anos) | Nº de Filhos | Salário (R\$) | Grau de Instrução |
|----|--------------|------|--------------|--------------|---------------|--------------------|
| 1 | Solteiro | M | 25 | 0 | 799,00 | Ensino Fundamental |
| 2 | Solteiro | F | 35 | 3 | 843,57 | Ensino Médio |
| 3 | Casado | M | 43 | 2 | 1972,93 | Ensino Superior |
| 4 | Solteiro | M | 39 | 3 | 753,25 | Ensino Fundamental |
| 5 | Casado | F | 42 | 1 | 922,43 | Ensino Médio |
| 6 | Casado | F | 28 | 0 | 913,00 | Ensino Médio |
| 7 | Solteiro | F | 27 | 1 | 774,56 | Ensino Fundamental |
| 8 | Casado | M | 30 | 2 | 2043,78 | Ensino Superior |
| 9 | Solteiro | F | 33 | 0 | 1012,29 | Ensino Médio |
| 10 | Casado | M | 50 | 2 | 978,78 | Ensino Médio |

Etapas da Estatística Descritiva

- Definição do problema
- Planejamento
- Coleta dos dados
- Apuração dos dados
- Apresentação e resumo dos dados
- Análise e interpretação

1. Definição do problema

- Formulação completa do problema a ser estudado.
 - Qual a unidade amostral?
 - Quais perguntas que quero obter resposta?
 - Qual a meta a ser alcançada diante do problema em questão?
- Levantamento de outros trabalhos realizados na mesma área e trabalhos análogos.

2. Planejamento

Consiste em se determinar o procedimento necessário para resolver o problema e, em especial, como levantar informações sobre o objeto de estudo. Nesta fase, deve-se levar em consideração:

- A construção adequada de um questionário;
- O tipo de levantamento que será realizado:
 - (i) Censo
 - (ii) Amostragem;
- O cronograma de atividades;
- Os custos envolvidos;
- O exame das informações disponíveis;
- O delineamento da amostra.

3. Coleta dos dados

Fase de caráter operacional, compreende à coleta das informações propriamente ditas. Se refere à obtenção de dados, com um objetivo determinado.

Os dados podem ser classificados em:

Dados primários: quando são publicados pela própria pessoa ou instituição que os obteve.

Dados secundários: quando são publicados por outras pessoas ou instituições.

4. Apuração dos dados

- Após ser feita uma crítica aos questionários utilizados buscando observações incompletas e/ou incorretas, deve-se então realizar a condensação, processamento ou tabulação dos dados.

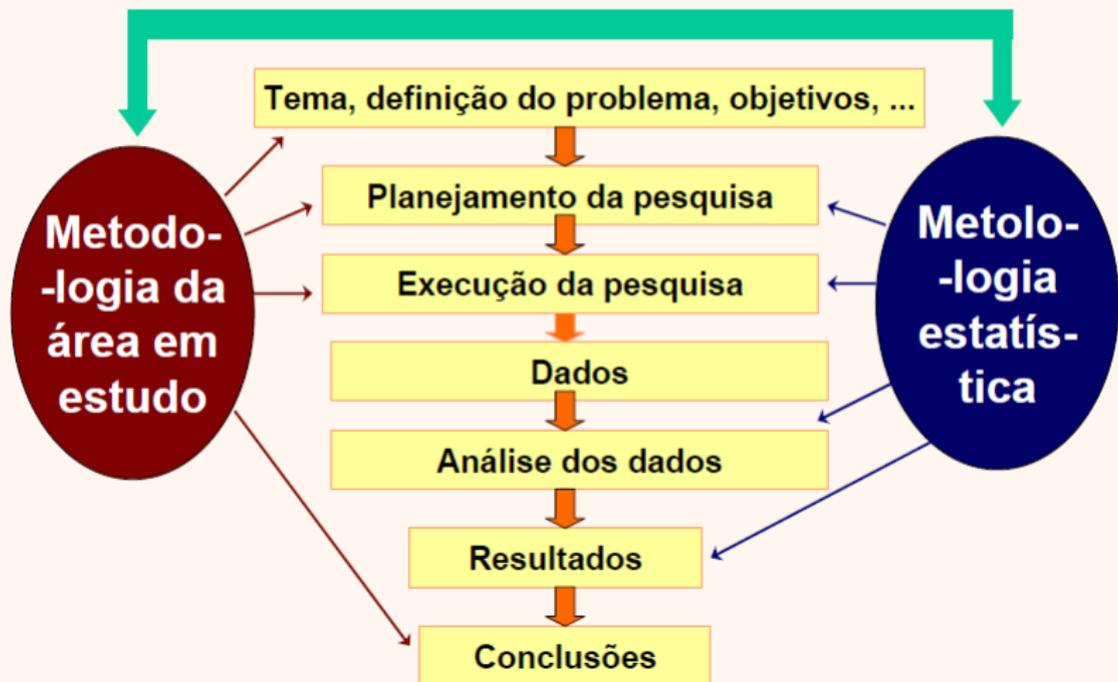
5. Apresentação e resumo dos dados

- A apresentação, consiste em apresentar os dados através de tabelas e gráficos, tornando mais fácil o exame do fenômeno sob estudo.
- No resumo, calculam-se medidas cuja finalidade principal é descrever o fenômeno que se está investigando.

6. Análise e interpretação

- O interesse maior reside em se tirar conclusões que auxiliem o pesquisador a resolver seu problema.
- As conclusões são baseadas na observação das tabelas, gráficos e medidas resumo.

Pesquisa, dados e estatística



Pesquisa

- Quem?
 - os elementos a serem pesquisados → POPULAÇÃO
- O quê?
 - características a serem observadas → VARIÁVEIS
- Como?
 - o instrumento de coleta de dados → QUESTIONÁRIO / ENTREVISTA ESTRUTURADA

Apresentação de dados

Apresentação tabular

| Pizza | Percentual |
|------------------------|------------|
| Pizza que eu já comi | 25.37% |
| Pizza que eu vou comer | 74.63% |

Apresentação gráfica



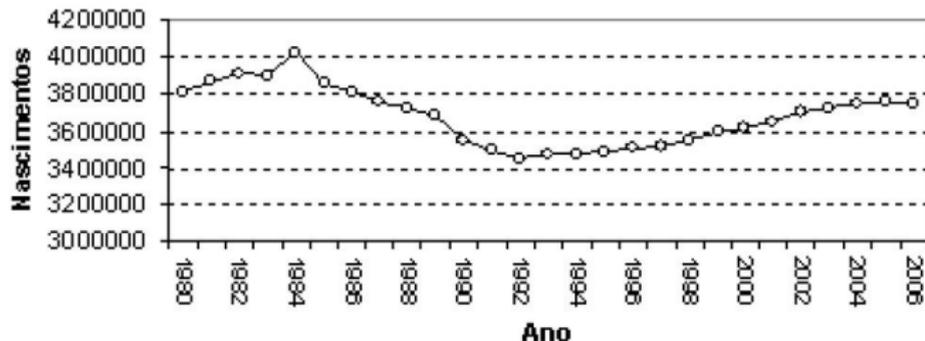
Apresentação tabular: a organização dos dados em tabelas proporciona um meio eficaz de estudo do comportamento de características de interesse. **Ex:** Distribuição de Frequências.

Apresentação gráfica: proporciona uma interpretação imediata dos resultados devido a sua simplicidade e clareza.

Representações Gráficas

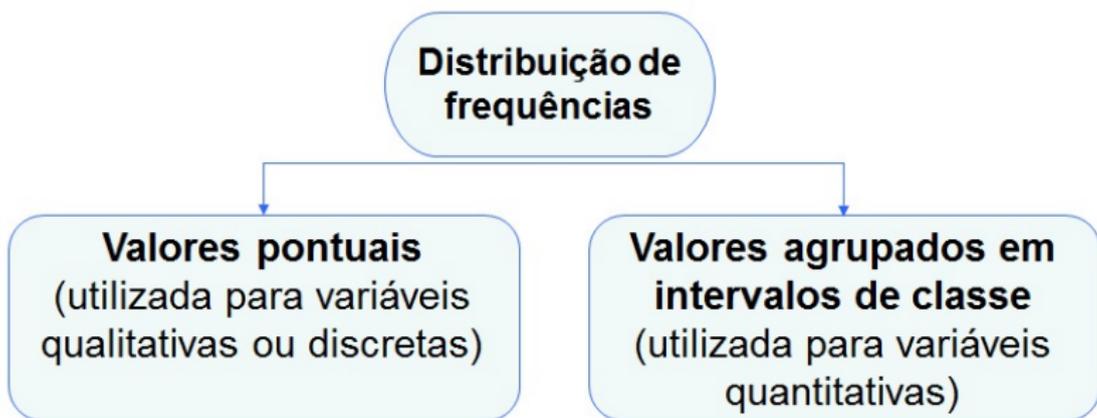
Gráficos em Linhas: São bastante utilizados na representação de séries de tempo. São eficientes na verificação de flutuações ou mudanças intensas na série ao longo do tempo. Para construir um gráfico em linhas, basta marcar os pontos correspondentes às grandezas e uní-los através de segmentos de reta.

**Número Total de Nascimentos no Brasil
(1980 - 2006)**



Distribuição de Frequências

- É uma tabela onde se preocupa em fazer corresponder os valores (categorias) observados da variável em estudo e as respectivas frequências.



Distribuição de Frequências

Dados Brutos: São os dados obtidos através de algum procedimento estatístico, que estão disponíveis logo após a coleta, mas não estão organizados.

Exemplo: idade de 50 alunos matriculados no ensino fundamental de uma escola pública.

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| 8 | 11 | 8 | 1 | 14 | 13 | 11 | 14 | 14 | 15 |
| 6 | 10 | 14 | 19 | 6 | 12 | 7 | 5 | 8 | 8 |
| 10 | 16 | 10 | 12 | 12 | 8 | 11 | 6 | 7 | 12 |
| 7 | 10 | 14 | 5 | 12 | 7 | 9 | 12 | 11 | 9 |
| 14 | 8 | 14 | 8 | 12 | 10 | 12 | 22 | 7 | 15 |

Como se pode observar, os valores estão dispostos de forma desordenada e pouca informação se consegue obter inspecionando os dados.

Distribuição de Frequências

Rol: São os dados ordenados, de forma crescente ou decrescente.

No exemplo anterior, em ordem crescente, temos:

| | | | | | | | | | |
|---|---|---|----|----|----|----|----|----|----|
| 5 | 7 | 8 | 8 | 10 | 11 | 12 | 12 | 14 | 15 |
| 5 | 7 | 8 | 8 | 10 | 11 | 12 | 12 | 14 | 15 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 16 |
| 6 | 7 | 8 | 9 | 10 | 12 | 12 | 14 | 14 | 19 |
| 6 | 7 | 8 | 10 | 11 | 12 | 12 | 14 | 14 | 22 |

Obs 1: Note que dessa forma fica fácil de verificar os valores extremos (máximo e mínimo).

Obs 2: Esse tipo de procedimento não é viável quando se tem um conjunto de dados muito grande.

Distribuição de Frequências

Frequência simples absoluta [f_i]: É o número de vezes que cada valor da variável se repete na amostra ou população.

Frequência simples relativa [fr_i]: É o número de vezes que esse valor ocorre relativamente ao total da amostra [n]; no fundo representa a parcela da amostra.

$$fr_i = \frac{f_i}{n}$$

Distribuição de Frequências por Valores

Definição: É uma tabela onde os valores da variável aparecem individualmente com suas respectivas frequências.

Teremos uma tabela assim:

| X_i | f_i |
|----------|--------------------------------------|
| X_1 | Número de valores iguais a $X_1=f_1$ |
| X_2 | Número de valores iguais a $X_2=f_2$ |
| X_3 | Número de valores iguais a $X_3=f_3$ |
| \vdots | \vdots |
| X_k | Número de valores iguais a $X_k=f_k$ |
| Σ | $f_1 + f_2 + \dots + f_k = n$ |

Note que para **cada** X_i existe uma frequência f_i associada.

Distribuição de Frequências por Valores

Exemplo: Construir a distribuição de frequências por valores, utilizando os dados do exemplo anterior.

| X_i (idade) | f_i (frequência) |
|---------------|--------------------|
| 05 | 2 |
| 06 | 3 |
| 07 | 5 |
| 08 | 7 |
| 09 | 2 |
| 10 | 5 |
| 11 | 4 |
| 12 | 9 |
| 13 | 1 |
| 14 | 7 |
| 15 | 2 |
| 16 | 1 |
| 19 | 1 |
| 22 | 1 |
| Σ | 50 |

A Secretária de Turismo da Paraíba, interessada em saber os estados com maior participação no número de turistas em João Pessoa, abordou 30 turistas que desembarcaram no aeroporto Castro Pinto. Os dados são apresentados a seguir:

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| SP | BA | BA | SP | RJ | BA | SP | SP | PR | PR |
| RJ | SP | SP | RJ | PR | RJ | SP | RJ | SP | BA |
| SP | RJ | PR | SP | BA | SP | PR | BA | SP | SP |

- 1 Que tipo de variável é essa?
- 2 Construa uma tabela de distribuição de frequências, usando f e fr .

| Estado de origem (X_i) | Frequência (f_i) | Frequência relativa (fr_i) |
|-------------------------------|-------------------------|-----------------------------------|
| BA | 6 | $6/30 = 0,2$ |
| PR | 5 | $5/30 = 0,17$ |
| RJ | 6 | $6/30 = 0,2$ |
| SP | 13 | $13/30 = 0,43$ |
| Total (Σ) | 30 | 1 |

Distribuição de Frequências por Classes

Definição: é uma tabela que mostra classes ou intervalos de dados, juntamente com as frequências correspondentes.

| Classes (idades) | f_i |
|------------------|-------|
| 5 † 11 | 24 |
| 11 † 15 | 21 |
| 15 † 18 | 3 |
| 18 † 23 | 2 |
| Σ | 50 |

Como construir uma Distribuição de Frequências por Classes

- 1 Determinar do número de classes;
- 2 Determinar quais são as classes;
- 3 Contar a frequência dos seus dados em cada classe.

Obs 1: A escolha do número de classes e da amplitude dos intervalos deve levar em consideração o conhecimento do pesquisador sobre o problema e a sua crítica sobre os dados.

Obs 2: Os intervalos de valores não precisam ter a mesma amplitude.

Distribuição de Frequências por Classes

Podemos expressar os limites das classes de várias formas:

- $LI \vdash LS$: considera valores entre LI e LS , incluindo LI e LS .
- $LI \vdash LS$: considera valores entre LI e LS , incluindo LI e excluindo LS .
- $LI \dashv LS$: considera valores entre LI e LS , excluindo LI e incluindo LS .

Distribuição de Frequências por Classes

- **Exemplo:** Distribuição de frequências por classes para as idades de 50 alunos na amostra.

| Classes (idades) | f_i | fr_i |
|------------------|-------|--------|
| 5 ┆ 11 | 24 | 48% |
| 11 ┆ 15 | 21 | 42% |
| 15 ┆ 18 | 3 | 6% |
| 18 ┆ 23 | 2 | 4% |
| Σ | 50 | 100% |

Distribuição de Frequências

Frequência acumulada absoluta [Fac_j]: É a soma do número de ocorrências para os valores iguais ou inferiores ao valor dado.

$$Fac_j = \sum_{i=1}^j f_i = f_1 + f_2 + \dots + f_j.$$

Frequência relativa acumulada [$Frac_j$]: É o número de vezes que a frequência acumulada absoluta ocorre relativamente ao total da amostra [n].

$$Frac_j = \sum_{i=1}^j fr_i = fr_1 + fr_2 + \dots + fr_j = \frac{f_1 + f_2 + \dots + f_j}{n}.$$

Distribuição de Frequências por Classes

- **Exemplo:** Distribuição de frequências por classes para as idades de 50 alunos na amostra.

| Classes (idades) | f_i | fr_i | fac_j | $frac_j$ |
|------------------|-------|--------|---------|----------|
| 5 † 11 | 24 | 48% | 24 | 48% |
| 11 † 15 | 21 | 42% | 45 | 90% |
| 15 † 18 | 3 | 6% | 48 | 96% |
| 18 † 23 | 2 | 4% | 50 | 100% |
| Σ | 50 | 100% | | |

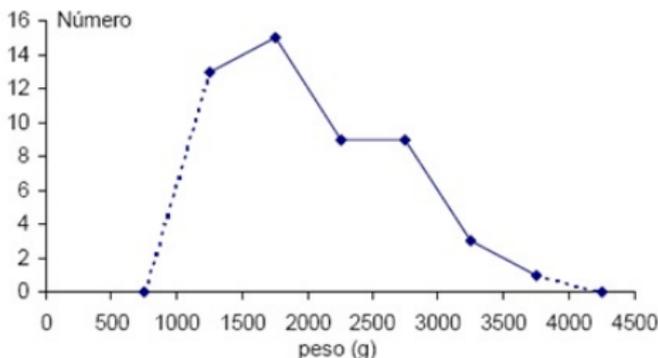
Representações Gráficas

Polígono de frequências: É a representação de uma distribuição de frequências por meio de um polígono.

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g).

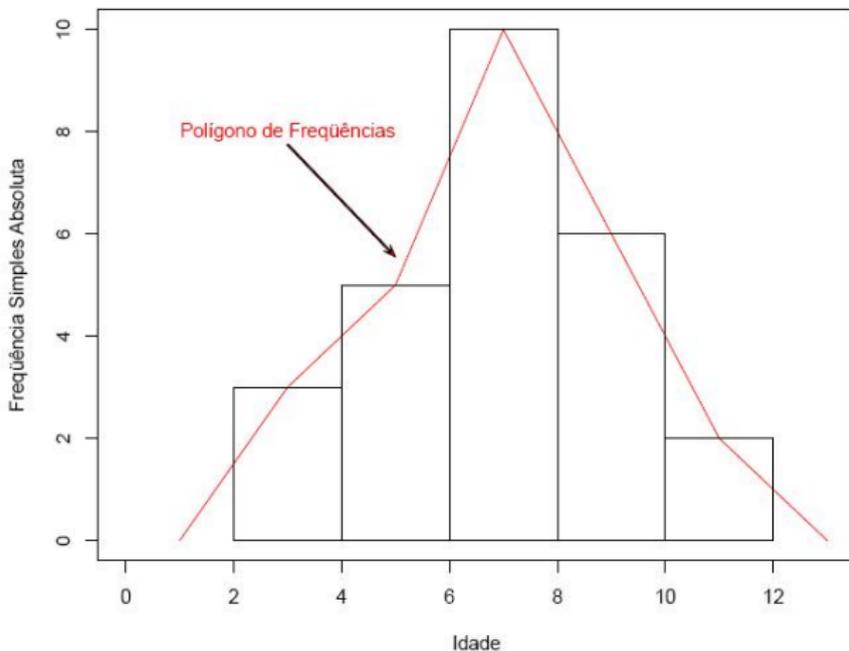
| Peso(g) | f_i | fr_i |
|---------------|-------|--------|
| 1000 -- 1500 | 13 | 26 |
| 1500 -- 2000 | 15 | 30 |
| 2000 -- 2500 | 9 | 18 |
| 2500 -- 3000 | 9 | 18 |
| 3000 -- 3500 | 3 | 6 |
| 3500 -- 4000 | 1 | 2 |
| Total | 50 | 100 |

Fonte: Hand DJ *et al.*, 1994.



Histograma e Polígono de Frequência

Histograma da Idade de Uma Amostra de Alunos da Escola XYZ



Tabulações cruzadas

- A relação entre duas variáveis qualitativas pode ser representada em uma tabulação cruzada.
- Conta-se quantos valores correspondem cada par de possíveis resultados, para as duas variáveis.
- O resultado pode ser apresentado como frequência absoluta ou relativa.

Dados brutos

| Obs | escola | trabalha |
|-----|---------|----------|
| 1 | pública | n |
| 2 | pública | n |
| 3 | privada | n |
| 4 | privada | s |
| 5 | privada | n |
| 6 | privada | s |
| 7 | pública | s |
| 8 | privada | n |
| 9 | pública | n |
| 10 | pública | n |
| 11 | pública | n |
| 12 | privada | n |
| 13 | pública | n |
| 14 | privada | s |
| 15 | privada | s |
| 16 | privada | n |
| 17 | privada | n |

- **Exemplo:** Tabela cruzada com relação as variáveis trabalha e tipo de escola, frequências absolutas.

| Trabalha \ tipo de escola | privada | publica | Total Geral |
|---------------------------|---------|---------|-------------|
| Não | 6 | 6 | 12 |
| Sim | 4 | 1 | 5 |
| Total Geral | 10 | 7 | 17 |

Tabela cruzada

- Exemplo:** Tabela cruzada com frequências relativas (divide cada valor pelo total geral e depois multiplica por 100).

| Trabalha \ tipo de escola | privada | publica | Total Geral |
|---------------------------|---------|---------|-------------|
| Não | 35,29% | 35,29% | 70,59% |
| Sim | 23,53% | 5,88% | 29,41% |
| Total Geral | 58,82% | 41,18% | 100,00% |

Vimos anteriormente a sintetização dos dados sob a forma de tabelas, gráficos e distribuições de frequências.

Medidas resumo: são medidas que possibilitam representar um conjunto de dados (valores de uma variável quantitativa, isto é, informações numéricas), relativo a observação de determinado fenômeno de forma resumida. São classificadas em medidas de posição, dispersão, separatrizes, assimetria e curtose.

Medidas de Posição (ou medidas de tendência central) : Essas medidas estabelecem valores em torno dos quais os dados se distribuem. Dizemos ainda que esse nome é dado pelo fato dos dados observados tenderem, em geral, a se concentrar em torno de valores centrais.

Ex: média, mediana, moda.

Média Aritmética Simples

Se dispomos de um conjunto de valores da amostra (ordenados ou não) podemos calcular sua média aritmética simples por

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + \dots + X_n}{n},$$

no caso amostral, em que n representa o número de indivíduos da amostra.

Exemplo: Abaixo, temos as quantidades mensais de lixo (em toneladas) produzidas em João Pessoa no 1º semestre do ano passado. Qual a média da quantidade de lixo produzida?

3,7 3,9 3,1 2,9 2,7 2,8

Temos que $n = 6$ e obtemos \bar{X} através de

$$\bar{X} = \frac{3,7 + 3,9 + 3,1 + 2,9 + 2,7 + 2,8}{6} = \frac{19,1}{6} = 3,18.$$

Ou seja, no primeiro semestre do ano passado, João Pessoa produziu em média 3,18 toneladas de lixo por mês.

Vantagens e desvantagens da média

- V1* É a medida mais conhecida e de maior uso;
- V2* É facilmente calculável;
- V3* Serve para compararmos conjuntos semelhantes;

- D1* É uma medida sensível a observações extremas. Ou seja, é grandemente influenciada pelos valores extremos (muito grandes ou muito pequenos) do conjunto;

Definição : o valor que divide a série **ordenada** em duas partes iguais.

Em outras palavras, é o valor que ocupa o centro da distribuição, ou seja, 50% dos elementos da série são menores do que ela e 50% dos elementos da série são maiores do que ela.

Exemplo: No Rol, temos:



Exemplo: Considere os dados sobre a produção mensal de lixo em João Pessoa. Temos $x_1 = 3, 7$, $x_2 = 3, 9$, $x_3 = 3, 1$, $x_4 = 2, 9$, $x_5 = 2, 7$ e $x_6 = 2, 8$.

A série ordenada é 2, 7, 2, 8, 2, 9, 3, 1, 3, 7, 3, 9 e o valor central está entre 2, 9 e 3, 1. Neste caso a mediana (denotaremos por Md) é:

$$Md = \frac{2,9 + 3,1}{2} = 3$$

Podemos encontrar a mediana de um conjunto de dados das seguintes formas:

- 1) **Se n é ímpar:** a mediana será o valor central da série ordenada.
- 2) **Se n é par:** teremos dois valores centrais e a mediana será a média entre esses dois valores centrais.

Vantagens e desvantagens da Mediana

- V1* A mediana não é influenciada por valores extremos (grandes) de uma série ou conjunto de dados;
- D2* Não é levada em consideração na maior parte dos testes estatísticos.

Definição: é o valor que ocorre com maior frequência (denotaremos por Mo).

Exemplo: 1, 2, 3, 3, 4, 4, 4, 4, 5, 5, 6, 6, 6, 6, 6, 6, 7, 7, 8.
Temos que o valor mais frequente é 6, logo, $Mo = 6$.

Obs 1: A moda pode não existir. Neste caso, dizemos que o conjunto de dados é amodal.

Exemplo: 1, 1, 2, 2, 3, 3, 4, 4, 5, 5

Obs 2: A moda pode não ser única.

Exemplo: 2, 2, 3, 3, 3, 4, 4, 5, 5, 5. Temos dois valores mais frequentes: 3 e 5.

Vantagens e desvantagens da moda

- V1* Não é influenciada por valores extremos (grandes) do conjunto de dados;
- D1* Não depende de todos os valores do conjunto de dados, podendo mesmo não se alterar com a modificação de alguns deles;

Um instrutor registra o número de faltas de seus alunos em determinado semestre. Em uma amostra aleatória, os dados são:

2 4 2 0 40 2 4 3 6

Calcule a média, a mediana e a moda.

Rol: 0 2 2 2 3 4 4 6 40

Média: $\bar{X} = \frac{0+2+2+2+3+4+4+6+40}{9} = \frac{63}{9} = 7$

Mediana: 0 2 2 2 3 4 4 6 40, $Md = 3$

Moda: $Mo = 2$, pois é o valor que ocorre mais vezes.

Suponha que o aluno com 40 faltas abandone o curso.
Agora temos:

2 4 2 0 2 4 3 6

Calcule a média, a mediana e a moda.

Rol: 0 2 2 2 3 4 4 6

Média: $\bar{X} = \frac{0+2+2+2+3+4+4+6}{8} = \frac{23}{8} = 2,875$

Mediana: 0 2 2 **2 3** 4 4 6, $Md = \frac{2+3}{2} = 2,5$

Moda: $Mo = 2$, pois é o valor que ocorre mais vezes.

Medidas de Dispersão

O número de pacientes atendidos em duas unidades de saúde em dez dias consecutivos. Calcule a média, a mediana e a moda de cada unidade.

| Unidade A | | | Unidade B |
|------------------|----|---|------------------|
| | 56 | | 33 |
| | 56 | | 42 |
| | 57 | | 48 |
| | 58 | | 52 |
| | 61 |  | 57 |
| | 63 | | 67 |
| | 63 | Iguais ?? | 67 |
| | 67 | | 77 |
| | 67 | | 82 |
| | 67 | | 90 |
| Média = 61,5 | | | Média = 61,5 |
| Mediana = 62 | | | Mediana = 62 |
| Moda = 67 | | | Moda = 67 |

- As medidas de posição apresentadas fornecem a informação dos dados apenas a nível pontual, sem ilustrar outros aspectos referentes à forma como os dados estão distribuídos na amostra.
- É preciso uma medida estatística complementar para melhor caracterizar os dados apresentados.
- **Medidas de Dispersão (ou medidas de variação)** servem para caracterizar o quanto os dados estão espalhados em torno de uma medida de posição como, por exemplo, a média.

As medidas de dispersão mais utilizadas são:

- Amplitude total (AT);
- Variância (S^2);
- Desvio padrão (S);
- Coeficiente de variação (CV).

Definição: é a diferença entre o maior e o menor valor da série, ou seja,

$$AT = X_{\text{máx}} - X_{\text{mín}}$$

A amplitude é útil para nos dar uma ideia do campo de variação da série. Verifica-se que é uma medida de dispersão limitada.

Definição: soma dos quadrados dos desvios com relação à média, dividida pelo número de elementos (ou pelo número de elementos menos um, no caso amostral). Ou seja, dada a amostra, temos que

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

É uma quantidade sempre não negativa e expressa em unidades quadradas do conjunto de dados.

| x_i | $(x_i - \bar{x})$ | $(x_i - \bar{x})^2$ |
|-------|-------------------|---------------------|
| 56 | -5,5 | 30,25 |
| 56 | -5,5 | 30,25 |
| 57 | -4,5 | 20,25 |
| 58 | -3,5 | 12,25 |
| 61 | -0,5 | 0,25 |
| 63 | 1,5 | 2,25 |
| 63 | 1,5 | 2,25 |
| 67 | 5,5 | 30,25 |
| 67 | 5,5 | 30,25 |
| 67 | 5,5 | 30,25 |
| | | <hr/> 188,50 |

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

$$S^2 = \frac{188,5}{9} = 20,94$$

—————> Soma dos quadrados

Desvantagem de uso da Variância

- É uma medida de difícil interpretação, devido ao fato de que a unidade de medida fica elevada ao quadrado. Em alguns casos, a unidade de medida ao quadrado nem fará sentido.
- O interessante é ter uma medida que descreva a variabilidade das informações na mesma escala em que estão os dados fornecidos.
- Esta medida se chama **Desvio Padrão**.

Definição: a raiz quadrada positiva da variância.

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

Prefere-se usar o desvio padrão porque este é expresso na mesma unidade dos dados.

Considerando o exemplo anterior, temos

$$\begin{aligned}\bar{X}_A &= 61,5 & S_A^2 &= 20,94 & S_A &= 4,57 \\ \bar{X}_B &= 61,5 & S_B^2 &= 335,38 & S_B &= 18,31\end{aligned}$$

O número de pacientes atendidos na unidade A varia menos do que na unidade B.

Coeficiente de Variação

- É uma medida de dispersão relativa que serve para comparar dois ou mais conjuntos de dados de unidades diferentes.
- Mede o grau de concentração dos dados em torno de sua média. É obtido através das expressões

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

- Quanto maior o coeficiente de variação, maior a dispersão em torno da média.
- Pode-se denotar CV também em termos percentuais, bastando fazer $CV \times 100\%$.

Exemplo:

Considere uma amostra de 10 recém-nascidos, dos quais são conhecidos os comprimentos (X em centímetros) e os pesos (Y em gramas).

| Bêbe | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------|------|------|------|------|------|------|------|------|------|------|
| Comp. (X) | 52 | 48 | 45 | 49 | 51 | 54 | 47 | 50 | 46 | 51 |
| Peso (Y) | 3300 | 3200 | 2950 | 3150 | 3350 | 3450 | 2900 | 3300 | 3150 | 3250 |

Neste caso,

$$\bar{X} = 49,3 \quad S_x = 2,69 \quad CV_x = 0,545$$

$$\bar{Y} = 3200 \quad S_y = 162,8 \quad CV_y = 0,051$$

Os recém-nascidos variam mais quanto ao comprimento do que quanto ao peso.