

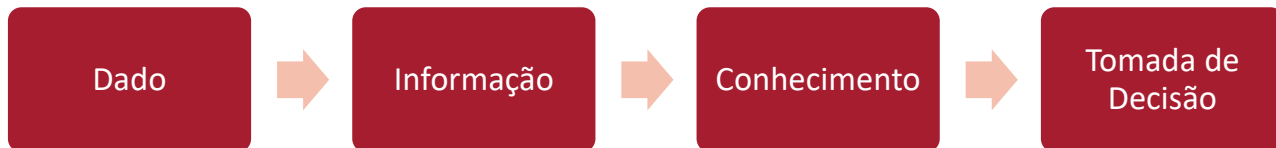


Métodos Quantitativos

ESTATÍSTICA DESCRITIVA

O que é Estatística?

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



Definições

Dados são observações (tais como medidas, sexos, respostas de pesquisas) que tenham sido coletados.

Uma **população** é uma coleção completa de todos os elementos (notas, pessoas, medidas e etc.) a serem estudados.

Um **censo** é o conjunto de dados obtidos de todos os membros da população.

Uma **amostra** é um subconjunto de membros selecionados de uma população.

Um **parâmetro** é uma medida numérica que descreve alguma característica de uma **população**.

Uma **estatística** é uma medida numérica que descreve alguma característica de uma **amostra**.

Exemplo 1

Uma pesquisa pergunta a 1.087 adultos: “Você tem a oportunidade de fazer uso de bebidas alcoólicas ou é totalmente abstinência?”

- ☐ Os 1.087 sujeitos da pesquisa constituem uma amostra.
- ☐ A população consiste na coleção inteira de todos os 140,7 milhões de adultos brasileiros.
- ☐ A cada 10 anos o governo brasileiro tenta obter um censo de todos os cidadãos, mas fracassa, pois é impossível atingir todos.

Variável Aleatória

Uma variável aleatória é uma variável quantitativa ou qualitativa, cujo resultado depende de fatores aleatórios.

- ☐ Resultado do lançamento de um dado que pode dar qualquer número entre 1 e 6.
- ☐ Cara ou coroa.
- ☐ Cor de um objeto.
- ☐ Temperatura média registrada.

Dizemos que uma variável aleatória (x) segue uma função $f(x)$ que associa um valor numérico a cada resultado possível de um experimento aleatório

Tipos de Variáveis



Exemplo 2

Dados Discretos: Os números de ovos que as galinhas botam por semana são dados discretos porque representam contagens.

Dados Contínuos: As quantidades de leite das vacas são dados contínuos, pois podem ser produzidos variando entre 0 a 20 litros por dia, tal como 14,273431 litros.

Ordinais: classe social, cargo na empresa, escolaridade e classificação de um filme.

Nominais: sexo, bairro, cor de pele e canal de TV preferido.

Estatística Descritiva

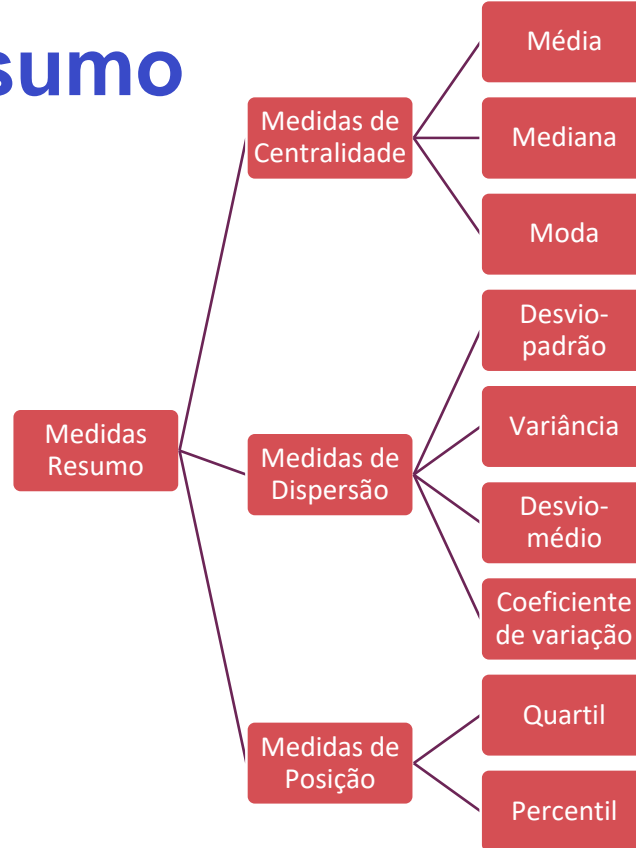
Medidas Resumo

Tabelas de Frequência

Gráficos e outras de Dados

Análise de *outliers*

Medidas Resumo



Medidas de Centralidade

Há várias maneiras de determinar o centro, de modo que temos diferentes definições de medidas de centro, incluindo a média, a mediana e a moda.

A **média aritmética** de um conjunto de valores é a medida de centro encontrada pela adição de valores e divisão do total pelo número de valores.

A **mediana** - representada por \tilde{x} (pronuncia-se “x til”) - de um conjunto de dados é a medida de centro que é o valor do meio quando os dados originais estão arranjados em ordem crescente de magnitude. Para encontrar a mediana:

1. Ordene os valores.
2. Se o nº de valores for ímpar, a mediana será o número localizado no meio exata da lista.
3. Se for par, a mediana será encontrada pelo cálculo da média dos dois números do meio.

Medidas de Centralidade

Notação	
x	É a variável , em geral, usada para representar os valores individuais dos dados.
n	Representa o número de valores de uma amostra .
N	Representa o número de valores de uma população .
$\bar{x} = \frac{\sum x}{n}$	É a média de um conjunto de valores amostrais .
$\mu = \frac{\sum x}{N}$	É a média de todos os valores em uma população .

Regra de Arredondamento: Use uma casa decimal a mais do que é apresentado no conjunto original de valores.

Medidas de Centralidade

A **moda** de um conjunto de dados, em geral representada por M , é o valor que ocorre mais frequentemente.

Quando dois valores ocorrem com a mesma maior frequência, cada um é uma moda, e o conjunto de dados é bimodal.

Quando mais de dois valores ocorrem com a mesma maior frequência, cada um é uma moda, e o conjunto de dados é multimodal.

Quando nenhum valor se repete, dizemos que não há moda.

Medidas de Dispersão

A **amplitude** em um conjunto de dados é a diferença entre o maior e o menor valor.

$$\textit{Amplitude} = \textit{maior valor} - \textit{menor valor}$$

O **desvio padrão (s)** de um conjunto de valores amostrais, é uma medida da variação dos valores em torno da média. É uma espécie de desvio médio dos valores em relação à média.

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Medidas de Dispersão

A **variância** (s^2) é um indicador importante para distribuições teóricas, como a distribuição normal.

O **coeficiente de variação** (CV) é um indicador adimensional para compara dispersão entre dados que possuem unidades de medida e ordem de grandeza diferentes. É obtido através do quociente entre o desvio-padrão e a média, tal que:

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

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Medidas de Dispersão

Notação	
s	Desvio padrão amostral
s^2	Variância amostral
σ	Desvio padrão populacional
σ^2	Variância populacional

Regra de Arredondamento:

Use uma casa decimal a mais do que é apresentado no conjunto original de valores.

Medidas de Posição

Um quartil é um dos três valores que dividem um conjunto de dados ordenado em quatro partes iguais. Cada parte contém aproximadamente 25% dos dados. Os três quartis são frequentemente denotados como Q_1 , Q_2 e Q_3 .

Primeiro Quartil (Q_1): Também conhecido como quartil inferior, é o valor abaixo do qual se encontra 25% dos dados e acima do qual se encontram 75% dos dados.

Segundo Quartil (Q_2): É a mediana do conjunto de dados, ou seja, o valor central onde 50% dos dados estão abaixo e 50% estão acima.

Terceiro Quartil (Q_3): Também conhecido como quartil superior, é o valor abaixo do qual se encontra 75% dos dados e acima do qual se encontram 25% dos dados.

Medidas de Posição

Um **percentil** é um valor que divide um conjunto de dados ordenados em 100 partes iguais, onde cada parte representa 1% dos dados. O k -ésimo percentil (P_k) indica o valor abaixo do qual se encontram $k\%$ dos dados. Por exemplo, o 25º percentil ($Q1$) é o valor abaixo do qual estão 25% dos dados, e o 75º percentil ($Q3$) é o valor abaixo do qual estão 75% dos dados.

Tabelas e Distribuições de Frequência

Tabelas de frequência são uma forma de organizar dados em uma tabela para mostrar a frequência de ocorrência de cada valor ou intervalo de valores em um conjunto de dados. Utiliza-se em estatística para resumir grandes conjuntos de dados e torná-los mais fáceis de entender e analisar.

Classes ou **categorias** são os diferentes valores ou categorias que estão sendo analisados na variável. Por exemplo, se você estiver analisando a altura das pessoas, os valores ou categorias podem ser “baixo”, “médio” e “alto”.

A **frequência absoluta** (f_i) de uma categoria i é o número de vezes que um valor específico aparece em um conjunto de dados, enquanto a **frequência relativa** (fr_i) é a proporção desse valor em relação ao total de dados, geralmente expressa como uma porcentagem (%).

Tabelas e Distribuições de Frequência

Frequência cumulativa trazem a soma cumulativa das frequências absolutas ou relativas à medida que você percorre os valores ou categorias da tabela. Pode ser crescente (somar do valor mais baixo a cada valor sucessivo) ou decrescente (somar do valor mais alto para baixo). Há tanto a **frequência absoluta acumulada (F_i)** como a **frequência relativa acumulada (FR_i)**.

Frequências acumuladas só são observadas e válidas para dados ordinais e dados quantitativos.

Tabelas e Distribuições de Frequência

Cada categoria i de um conjunto de dados possui os seguintes registros de frequência:

- ☐ Frequência Absoluta (f_i).
- ☐ Frequência Relativa (fr_i).
- ☐ Frequência Acumulada (F_i)
- ☐ Frequência Relativa Acumulada (FR_i).

Tabelas e Distribuições de Frequência

Tipo de cliente		
Classes	f	fr
Pessoa Física	79	56.8%
Pessoa Jurídica	60	43.2%
Total	139	100.0%

Status do pedido				
Classes	f	fr	F	FR
0) Em negociação	6	4.3%	6	4.3%
1) Planejamento	20	14.4%	26	18.7%
2) Em produção	52	37.4%	78	56.1%
3) Finalizado	50	36.0%	128	92.1%
4) Entregue	11	7.9%	139	100.0%
Total	139	100.0%		

Tabelas e Distribuições de Frequência

Total de produtos encomendados				
Classes	f	fr	F	FR
1	4	2.9%	4	2.9%
2	18	12.9%	22	15.8%
3	46	33.1%	68	48.9%
4	21	15.1%	89	64.0%
5	14	10.1%	103	74.1%
6	25	18.0%	128	92.1%
7	11	7.9%	139	100.0%
Total	139	100.0%		

Tabelas e Distribuições de Frequência

Intervalos do prazo de entrega dos pedidos				
Classes	f	fr	F	FR
1) Até 4 semanas	3	2.2%	3	2.2%
2) Entre 5 e 7 semanas	17	12.2%	20	14.4%
3) Entre 8 e 10 semanas	55	39.6%	75	54.0%
4) Entre 11 e 13 semanas	44	31.7%	119	85.6%
5) Entre 14 e 16 semanas	18	12.9%	137	98.6%
6) A partir de 17 semanas	2	1.4%	139	100.0%
Total	139	100.0%		

Tabelas e Distribuições de Frequência

Dados nominais:

- ☐ Ordenar as categorias de forma decrescente de acordo com a frequência absoluta.
- ☐ Ordenar as categorias em ordem alfanumérica.

Dados ordinais:

- ☐ Ordenar as categorias de forma crescente (preferência, cronologia).

Dados quantitativos:

- ☐ Ordenar as categorias de forma crescente.

Tabelas e Distribuições de Frequência

Exercício 1

Construir a tabela de frequências pro salário, considerando os seguintes valores:

1. Até 6 salários
2. Entre 6,01 e 9 salários
3. Entre 9,01 e 12 salários
4. Entre 12,01 e 15 salários
5. Entre 15,01 e 18 salários
6. Acima de 18 salários

OU

1. Menos que 6 salários
2. Entre 6 e 8,99 salários
3. Entre 9 e 11,99 salários
4. Entre 12 e 14,99 salários
5. Entre 15 e 17,99 salários
6. A partir de 18 salários

Tabelas e Distribuições de Frequência

Para dados contínuos, é necessário construir intervalos padrões onde se pode categorizar cada valor dentro de uma faixa de intervalo e assim construir-se uma tabela de frequências.

Muitos softwares nos permitem obter automaticamente distribuições de frequência.

Tabelas e Distribuições de Frequência

1. Decida sobre o **número de classes** desejado. O n^o de classes deve estar entre 5 e 20. Existem diversos critérios para se decidir o n^o de classes, como por exemplo:

– Regra de Sturges:

$$K = 1 + 3,3 \cdot \log_{10} n$$

– Regra da Raiz:

$$K = \sqrt{n}$$

Usa-se o valor truncado de K (arredondar pro valor inteiro abaixo).

2. Calcule:

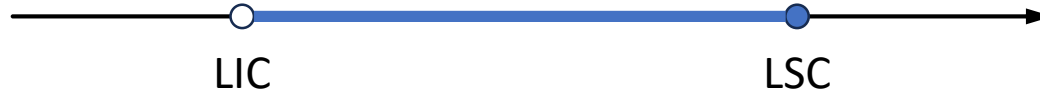
$$\text{Amplitude de classe} \approx \frac{\text{maior valor} - \text{menor valor}}{K}$$

Tabelas e Distribuições de Frequência

3. **Ponto inicial:** Comece escolhendo um n^o para limite inferior da 1ª classe. Escolha o valor mínimo ou convenientemente menor.
4. Usando o limite inferior da 1ª classe e amplitude de classe, prossiga e liste os **limites inferiores de classe (LIC)**.
5. Liste os LICs em uma coluna vertical e prossiga para preencher os **limites superiores de classe (LSC)** que são facilmente identificados.
6. Percorra o conjunto de dados associando cada valor a sua respectiva classe de modo a obter a frequência total das classes.

Tabelas e Distribuições de Frequência

Intervalo fechado no limite superior da classe: significa que o menor valor da classe (LIC) não será contabilizado na frequência absoluta, enquanto o maior valor da classe (LSC) sim. De maneira prática, quer fazer que os valores contados são os maiores que os limites inferiores e menores ou iguais aos limites superiores.



Representação:

$]LIC ; LSC]$

$(LIC ; LSC]$

Tabelas e Distribuições de Frequência

Intervalo fechado no limite inferior da classe: significa que o maior valor da classe (LSC) não será contabilizado na frequência absoluta, enquanto o menor valor da classe (LIC) sim. De maneira prática, quer dizer que os valores contados são os maiores ou iguais ao limite inferior e menores que os limites superiores.



Representação:

$[LIC ; LSC[$

$[LIC ; LSC)$

Tabelas e Distribuições de Frequência

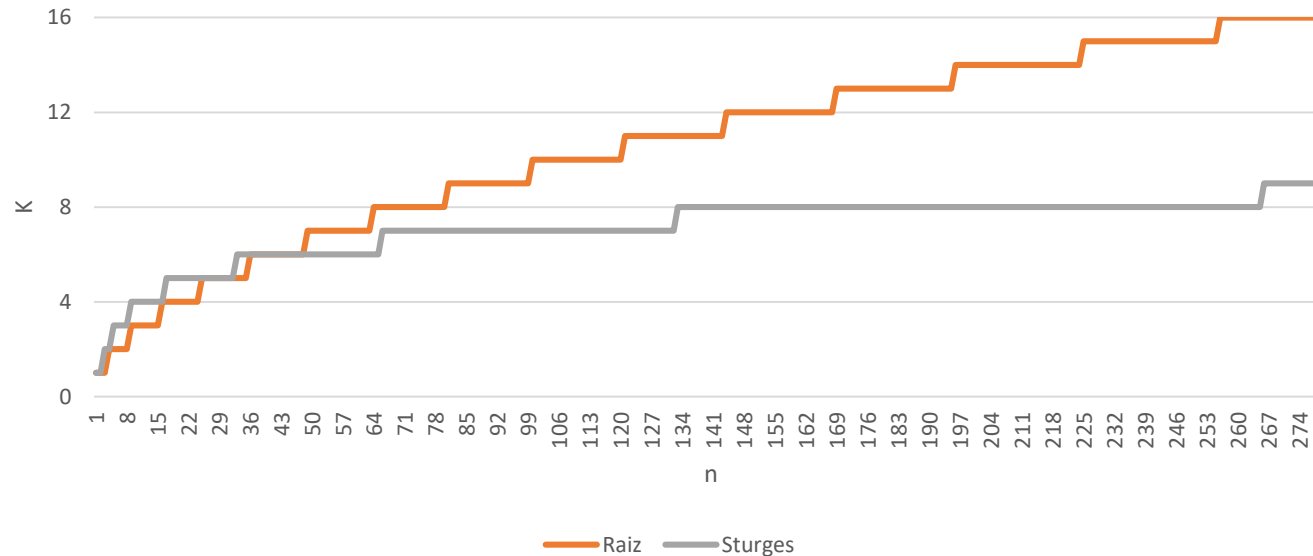
Intervalo fechado no limite inferior

Classes
[R\$ 2,500.00 ; R\$ 6,027.27 [
[R\$ 6,027.27 ; R\$ 9,554.55 [
[R\$ 9,554.55 ; R\$ 13,081.82 [
[R\$ 13,081.82 ; R\$ 16,609.09 [
[R\$ 16,609.09 ; R\$ 20,136.36 [
[R\$ 20,136.36 ; R\$ 23,663.64 [
[R\$ 23,663.64 ; R\$ 34,245.45 [
[R\$ 34,245.45 ; R\$ 41,300.00 [

Intervalo fechado no limite inferior

Classes
] R\$ 2,500.00 ; R\$ 6,027.27]
] R\$ 6,027.27 ; R\$ 9,554.55]
] R\$ 9,554.55 ; R\$ 13,081.82]
] R\$ 13,081.82 ; R\$ 16,609.09]
] R\$ 16,609.09 ; R\$ 20,136.36]
] R\$ 20,136.36 ; R\$ 23,663.64]
] R\$ 23,663.64 ; R\$ 34,245.45]
] R\$ 34,245.45 ; R\$ 41,300.00]

Tabelas e Distribuições de Frequência



Sturges
vs. Raiz

$n > 49$