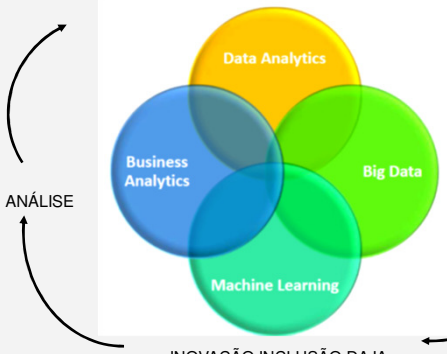
UNICARIOCA

CURSO DE PÓS-GRADUAÇÃO
ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS

CIÊNCIA DE DADOS

BASE ESTATÍSTICA E PROGRAMAÇÃO EM R



FOCO NOS DADOS

INOVAÇÃO INCLUSÃO DA IA

ANÁLISE

MÓDULOS	DISCIPLINAS	CH	Presencial	Atividade Prática Supervisionada
DATA ANALYTICS	Fundamentos de Estatística Aplicada	30	20	10
	Linguagem de Programação Aplicada R	30	20	10
	Análise Computacional e Qualitativa de Dados (R)	30	20	10
BIG DATA	Fundamentos de Big Data e Computação em Nuvem	30	20	10
	Data Mining	30	20	10
	Data Warehouse	30	20	10
MACHINE LEARNING	Fundamentos de Inteligência Artificial	30	20	10
	Linguagem de Programação Aplicada Python	30	20	10
	Machine Learning	30	20	10
BUSINESS ANALYTICS	Estatística Aplicada a Negócios	30	20	10
	Visualização de Dados	30	20	10
	Marketing, RH, Finanças e Redes Sociais Analytics	30	20	10
MÓDULO PRÁTICO	Projeto Integrador em Ciência de Dados	20	10	10
	Carga Horária Total	**380**	**250**	**130**

ESTATÍSTICA APLICADA A NEGÓCIOS

TEMAS....

- ANÁLISE EXPLORATÓRIA DE DADOS
- AMOSTRAGEM
- INTERVALO DE CONFIANÇA
- TESTE DE HIPÓTESES
- PCA - ANÁLISE DE COMPONENTES PRINCIPAIS (ANÁLISE FATORIAL)
- ANÁLISE DE VARIÂNCIA
- REGRESSÃO - REGRESSÃO LOGÍSTICA

CIÊNCIA DE DADOS

CIENTISTAS DE DADOS

*...geração de especialistas analíticos que possuem as **habilidades** técnicas para resolver problemas **complexos** - e a **curiosidade** para explorar quais problemas precisam ser resolvidos.*

PESQUISA OPERACIONAL

ENGENHARIA MATEMÁTICA

ESTATÍSTICA

PROBABILIDADE

+ TECNOLOGIA

FUNÇÕES TÍPICAS DOS CIENTISTAS DE DADOS...

Não há uma descrição definitiva!

- Coletar grandes quantidades de dados não-estruturados e transformá-los em um formato mais utilizável;
- Resolver problemas de negócios usando técnicas orientadas por dados;
- Trabalhar com uma variedade de linguagens de programação, incluindo R e Python;
- Ter uma sólida compreensão de ESTATÍSTICA, incluindo testes e distribuições;
- Estar sempre atualizado sobre técnicas analíticas, como *Machine Learning, Deep Learning* e Análise de Texto;
- Comunicar e colaborar tanto com TI quanto com a gerência;
- Por ordem e padrões nos dados, além de identificar tendências que podem ajudar no resultado financeiro da uma empresa.

CIÊNCIA DE DADOS

PREVISÃO
ANÁLISE DE DADOS
IDENTIFICAR PADRÕES
VISUALIZAÇÃO DE DADOS
BIG DATA
ANÁLISE DE SENTIMENTO
DATA MINING
TESTE DE HIPÓTESES
TÉCNICAS DE NLP
AMOSTRAGEM
DEEP LEARNING
MINERAÇÃO DE OPINIÃO
ANÁLISE FATORIAL
ANÁLISE EXPLORATÓRIA ***ANTES***
MACHINE LEARNING
REGRESSÃO
CORRELAÇÕES
ANALISAR RESULTADOS (INTERPRETAR) DEPOIS
INTELIGÊNCIA ARTIFICIAL
ANÁLISE GRÁFICA
MODELOS PROBABILÍSTICOS
ANÁLISE DE NEGÓCIO

Qual o impacto no nível de vendas do valor gasto com publicidade?

Qual a probabilidade de um e-mail ser SPAM?

Qual o efeito das comorbidades em pacientes com COVID19?

Glaucoma aumenta a pressão intraocular ?

Como detectar fraudes em declaração de importação?

Qual variável explica o nível de evasão escolar no 3º grau?
Renda - Nível Sócio Econômico - Sexo - Tempo fora da Escola

Como visualizar dados de maneira a facilitar a interpretação dos resultados?

Existe correlação entre anos de estudo e salário ?

Qual o tamanho de uma amostra para garantir com nível de confiança de 95% que o valor da média amostral não se afastará do verdadeiro valor da média na população por mais de 2%?

CIÊNCIA DE DADOS
ESTATÍSTICA

- *Quando desenvolvemos um modelo ou uma análise precisamos saber comunicar (explicar) o que o sistema está fazendo?*
- *Sem entender o que está sendo feito internamente, como convencer as pessoas (clientes) da utilidade do sistema, e que podem confiar nele?*
- *Sistemas de Machine Learning que realizam previsão de vendas utilizam técnicas estatísticas nos cálculos, dando peso aos valores mais recentes para aumentar a probabilidade de acerto do modelo.*
- *É fundamental entender como o algoritmo de previsão funciona para saber explicar os resultados.*

INTRODUÇÃO À ANÁLISE EXPLORATÓRIA DE DADOS

A análise exploratória de dados é uma filosofia que consiste no estudo dos dados a partir de todas as perspectivas e com todas as ferramentas possíveis, incluindo as já existentes. O propósito é extrair toda a informação possível, gerar novas hipóteses no sentido de construir conjecturas sobre as observações que dispomos.

***Antes** de modelar ou usar uma ferramenta é preciso conhecer os dados... Análise Exploratória!*

ESTATÍSTICA

PROBABILIDADE

MATEMÁTICA...

ALGORITMO E INFERÊNCIA

ANALISE ESTATÍSTICA

(a) ALGORÍTMICA

(b) INFERENCIAL

Exemplo: Considere o problema de **estimar** a **media** $\mu = E(X)$ de uma v.a. X, definida sobre uma população (P).

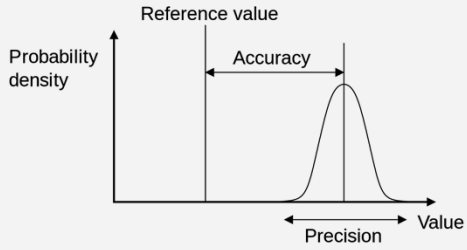
Para uma AAS $X_1;....;X_n$, considere o **estimador de μ** definido abaixo:

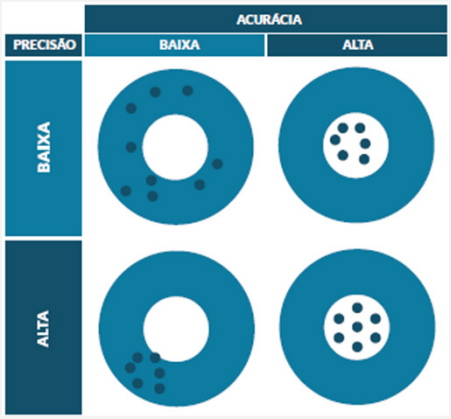
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$ ⟶ Este e o ALGORITMO!

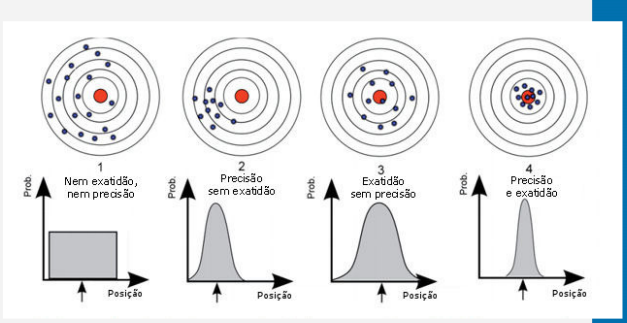
$Quão$ $\boldsymbol{acurado}$ e $\boldsymbol{preciso}$ é $\bar{X}?$ ⟶ Esta e a parte da INFERÊNCIA!

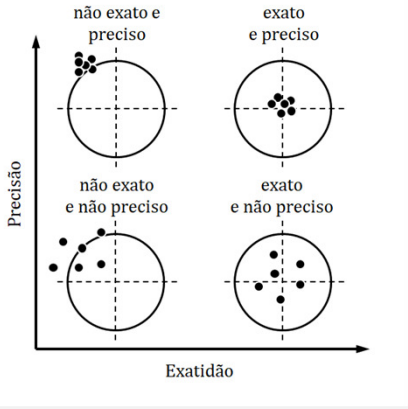
PRECISÃO ⇒ é o grau de variação gerado por diferentes medições. Dessa forma, quanto mais preciso for um processo, menor será a variação entre os valores obtidos

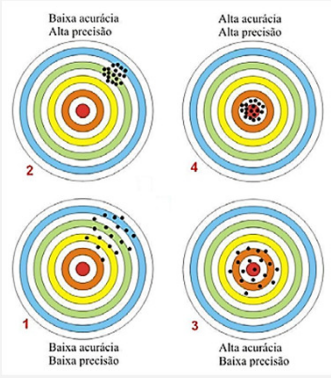
ACURÁCIA ⇒ **Exatidão** e **precisão** numa medição ou no resultado apresentado por um instrumento de medição. **PROXIMIDADE** entre o **resultado** de um instrumento de medida e o **verdadeiro** valor do que foi medido.











ALGORITMO E INFERÊNCIA

ALGORITMOS: e o que os Estatísticos fazem!

INFERÊNCIA: porque e para quê os estatísticos usam os algoritmos!

Grande conjuntos de dados (Big Data) requerem novas metodologias. Esta demanda esta sendo atendida por algoritmos estatísticos baseados em COMPUTAÇÃO INTENSIVA.

CD - PERSPECTIVA DA ESTATÍSTICA

- ESTATÍSTICA "serve" a Ciência guiando na coleta e analise de dados.
- Dados envolvem incertezas: como foram coletados, medidos ou como foram gerados. A modelagem estatística ajuda a quantificar e racionalizar incertezas de maneira sistemática.
- Conjuntos de dados são complexos: tipos diferentes de dependência (ao longo do tempo, entre variáveis diferentes)
- Dados de alta dimensão: medimos milhares de variáveis para cada unidade amostral.

INTRODUÇÃO À ANÁLISE EXPLORATÓRIA DE DADOS

CONCEITOS BÁSICOS

- Estatística
  - **Descritiva**
  - Inferencial

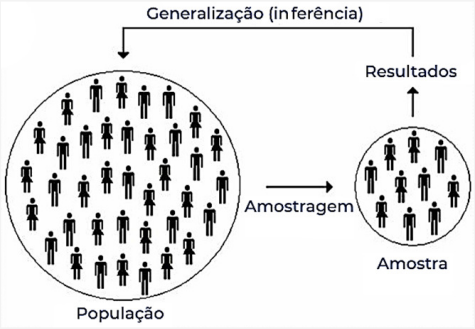
ESTATÍSTICA ⇒ é o estudo dos PROCESSOS de: coleta, organização e análise de um conjunto de dados e dos métodos de obtenção de conclusões ou de realização de previsões com base nos dados coletados.

✓ ESTATÍSTICA **DESCRITIVA** (DEDUTIVA)

✓ESTATÍSTICA **INDUTIVA** (INFERENCIAL)

ESTATÍSTICA **DESCRITIVA** ⇒ **POPULAÇÃO**

ESTATÍSTICA **INDUTIVA** ⇒ **AMOSTRA**



POPULAÇÃO ⇒ é o total do **grupo** a ser **observado** (universo)
Interesse do Pesquisador!

Características de uma População: **ATRIBUTOS** e **VARIÁVEIS**

✓ **VARIÁVEIS QUALITATIVAS - ATRIBUTOS**

✓ **VARIÁVEIS** ⇒ **QUANTITATIVAS**

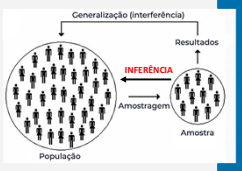
- DISCRETAS
- CONTÍNUAS

PESQUISA ⇒ **CENSO** ou **AMOSTRAGEM**.

CENSO ⇒ **contagem completa** na população

AMOSTRAGEM ⇒ é o **PROCESSO** de **dimensionamento** e **coleta** de informações de **parte** da população
Gera uma **AMOSTRA**!

AMOSTRA ⇒ **parte representativa da população** (*n*).



FRAÇÃO AMOSTRAL(FA) ⇒ $FA = \frac{n}{N}$

AMOSTRAGEM ⇒ **ESTIMA** os valores na **POPULAÇÃO** a partir de uma **AMOSTRA**.

✓ **AMOSTRA** ⇒ **ERRO** de **estimação**

✓ Quanto **menor** a **AMOSTRA** **maior** o **ERRO**, quanto **maior** a **AMOSTRA** **menor** o **ERRO**.

Se a **AMOSTRA** é a própria **POPULAÇÃO** o **ERRO** é **zero**.

✓Quanto **maior** a FRAÇÃO AMOSTRAL ⇒ **menor** o **ERRO** !

✓Quanto **menor** a FRAÇÃO AMOSTRAL ⇒ **maior** o **ERRO** !

$FA = \frac{n}{N}$ ↑ ↓ ERRO $\qquad$ $FA = \frac{n}{N}$ ↓ ↑ ERRO

MÉTODOS DE ANÁLISE DE DADOS

Simplesmente olhar para os dados não fornece um quadro claro do que pode estar acontecendo, especialmente quando a quantidade de dados for muito grande. A Estatística Descritiva possui uma grande quantidade de **instrumentos de resumo** que podem ser aplicados às diversas situações.

Existem dois tipos de métodos que podem ser utilizados, frequentemente de forma complementar:

✓ MÉTODOS GRÁFICOS OU TABULARES

✓ MÉTODOS NUMÉRICOS

RESUMOS ÚTEIS PARA A RESOLUÇÃO DE PROBLEMAS

MÉTODOS GRÁFICOS (TABULARES)

✓ Tabelas de Frequências

✓ Gráficos de Setores

✓ Gráficos de Barras

✓ Histogramas

✓ Ogiva (frequência acumulada)

✓ Ramos e Folhas

✓ Gráficos de Pontos

✓ Box-Plot

✓ Diagramas de Dispersão

RESUMOS ÚTEIS PARA A RESOLUÇÃO DE PROBLEMAS

MÉTODOS NUMÉRICOS

- ✓Média
- ✓Mediana
- ✓Moda
- ✓Quantis (Separatrizes)
- ✓Desvio Padrão, Variância
- ✓Coeficiente de Variação
- ✓ Coeficiente de Assimetria
- ✓Curtose
- ✓Coeficiente de Correlação Linear
- ✓Covariância

TIPOS DE VARIÁVEIS

- ✓QUALITATIVAS
- ✓QUANTITATIVAS

QUALITATIVAS (Categóricas)⇒ são as características que não podem ser medidas quantitativamente, como por exemplo: religião, estado civil, cor, etc.

Podem ser:

- NOMINAIS: quando as categorias **não** possuem uma ordem natural. Ex.: nomes, cores, sexo, naturalidade.
- ORDINAIS: quando as categorias **podem ser ordenadas**. Ex.: tamanho (pequeno, médio, grande), classe social (baixa, média, alta), grau de instrução (básico, médio, graduação, pós-graduação)...

TIPOS DE VARIÁVEIS

- ✓QUALITATIVAS
- ✓QUANTITATIVAS

QUANTITATIVAS ⇒ podem ser medidas quantitativamente, como por exemplo: peso, altura, taxas de inflação etc.

As quantitativas podem ser **DISCRETAS** ou **CONTÍNUAS**.

AMOSTRAGEM/AMOSTRA - APLICAÇÃO

IRPF - SIMULAR LEGISLAÇÃO ⇒ EFEITOS NA ARRECADAÇÃO

É POSSÍVEL SIMULAR EFEITOS NA ARRECADAÇÃO RECALCULANDO O VALOR DO IMPOSTO PARA OS 28 MILHÕES DE DECLARANTES ?

NÃO !

É UM INSTRUMENTO GERENCIAL DA SECRETARIA DA RECEITA FEDERAL !

SÃO DECISÕES QUE AFETAM:

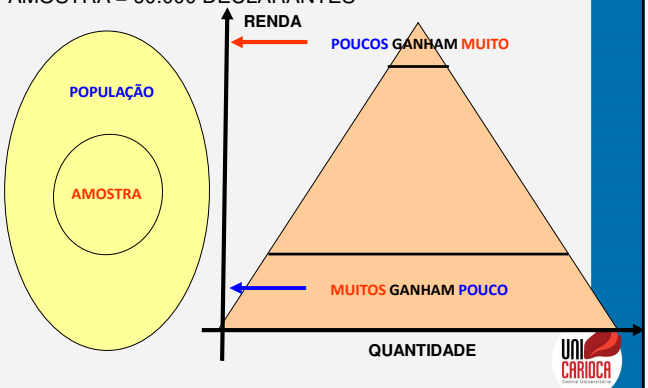
A DISTRIBUIÇÃO DE RENDA !

A ARRECADAÇÃO FEDERAL !

SOLUÇÃO ⇒ TRABALHAR COM UMA AMOSTRA DOS DECLARANTES !

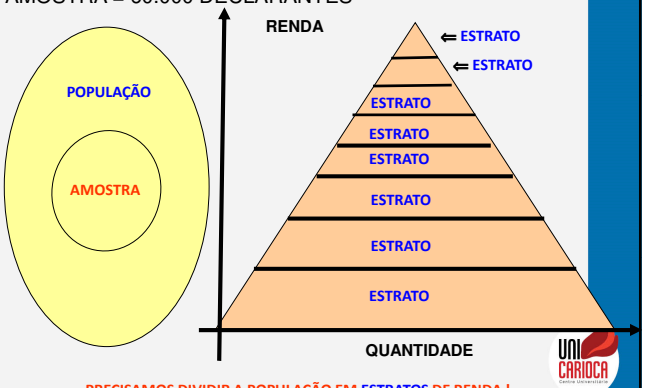
AMOSTRAGEM/AMOSTRA - APLICAÇÃO

AMOSTRA = 60.000 DECLARANTES



AMOSTRAGEM/AMOSTRA - APLICAÇÃO

AMOSTRA = 60.000 DECLARANTES



PRECISAMOS DIVIDIR A POPULAÇÃO EM ESTRATOS DE RENDA !

APRESENTAÇÃO - ORGANIZAÇÃO DE DADOS ESTATÍSTICOS

- Distribuição de Frequência por Intervalo e por Pontos.
- Histograma
- Polígono de Frequência.
- Tipos de Gráficos

ORGANIZAÇÃO DE DADOS ESTATÍSTICOS

SÉRIE ESTATÍSTICA ⇒ é toda tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do local ou da espécie.

SÉRIE DE DADOS NÃO GRUPADOS ⇒ é a série onde as variações do fenômeno são apresentadas de acordo com a época a que se referem, ao espaço onde se observa, ou a qualidade/espécie do fenômeno.

PRINCIPAIS ⇒

- TEMPORAL
- GEOGRÁFICA
- CATEGÓRICA

ORGANIZAÇÃO DE DADOS ESTATÍSTICOS

NOTAS DE **40** ALUNOS EM UMA PROVA DE **ESTATÍSTICA**

4 ; 2 ;10; 7 ; 3 ; 3 ; 4 ; 1 ; 9 ; 6 ; 1 ; 7 ; 8 ; 7 ; 6 ; 5 ; 9 ; 5 ; 6; 6
3 ; 9 ; 9 ; 6 ; 5 ;10; 1 ; 3 ;10; 2 ; 9 ; 3 ;10; 2 ; 1 ; 3 ;10; 3 ; 5; 4

PROBLEMA ⇒ COMO ANALISAR ESSES DADOS ?

O desempenho da turma foi bom?

Qual o indicador (a métrica) para medir desempenho?

Quantos alunos tiraram nota acima da média? E abaixo?

Quantos alunos tiraram 2 ou menos?

Quantos alunos tiraram mais de 9?

Como separar as 25% maiores notas?

E as 25% menores?

NOTAS...

4 ; 2 ;10; 7 ; 3 ; 3 ; 4 ; 1 ; 9 ; 6 ; 1 ; 7 ; 8 ; 7 ; 6 ; 5 ; 9 ; 5 ; 6; 6
3 ; 9 ; 9 ; 6 ; 5 ;10; 1 ; 3 ;10; 2 ; 9 ; 3 ;10; 2 ; 1 ; 3 ;10; 3 ; 5; 4

E AGORA? O QUE VOCÊ FARIA?
EXISTEM TÉCNICAS, MANEIRA DE RESPONDER AS PERGUNTAS QUE FORAM FEITAS?

O desempenho da turma foi bom?
Qual o indicador (a métrica) para medir desempenho?
Quantos alunos tiraram nota acima da média? E abaixo?
Quantos alunos tiraram 2 ou menos?
Quantos alunos tiraram mais de 9?
Como separar as 25% maiores notas?
E as 25% menores?

ORGANIZAÇÃO DE DADOS ESTATÍSTICOS

SÉRIE DE DADOS GRUPADOS ⇒ é a série onde o tempo, o espaço, a qualidade/espécie permanecem constantes e o fenômeno é agrupado em subintervalos do intervalo total. Estas séries serão estudadas a partir de **tabelas** chamadas de DISTRIBUIÇÃO DE FREQUÊNCIAS

DISTRIBUIÇÃO DE FREQUÊNCIA

DISTRIBUIÇÃO DE FREQUÊNCIA ⇒ é o método que consiste em agrupar dados em classes, categorias, ou intervalos. Existem **dois tipos** de Distribuição de Frequência: por **INTERVALO** e por **PONTOS**.

DISTRIBUIÇÃO DE FREQUÊNCIA por **INTERVALO** ⇒ as variações do fenômeno são agrupadas em **intervalos**. São características das **VARIÁVEIS CONTÍNUAS**.

DISTRIBUIÇÃO DE FREQUÊNCIA por **PONTOS** ⇒ é uma série de pontos grupados na qual o número de observações da variável, está relacionado com um ponto real. São características das **VARIÁVEIS DISCRETAS**.

DISTRIBUIÇÃO DE FREQUÊNCIA POR INTERVALO

Exemplo-1: Rendimento dos empregados de uma empresa do ABC paulista em Salários Mínimos (SM).

Classe	SM
1	1 ├ 3
2	3 ├ 5
3	5 ├ 7
4	7 ├ 9
5	9 ├ 11
Soma	Σ

DISTRIBUIÇÃO
Limite Inferior da distribuição = 1
Limite Superior da distribuição = 11
Amplitude da distribuição = 11-1 = 10

CLASSES ⇒ 5 classes
Limite Inferior da **classe.1** = 1
Limite Superior da **classe.1** = 3
Amplitude da **classe.1** = 2

DISTRIBUIÇÃO DE FREQUÊNCIA POR INTERVALO

Exemplo-1: Rendimento dos empregados de uma empresa do ABC paulista em Salários Mínimos (SM).

Frequência simples — DEPARTAMENTO DE PESSOAL

FREQUÊNCIA = CONTAGEM, QUANTIDADE !

Classe	SM	fi	Fi	fri	Fri	xi	Fi	Fri
1	1 ├ 3	90						
2	3 ├ 5	50						
3	5 ├ 7	30						
4	7 ├ 9	20						
5	9 ├ 11	10						
Soma	Σ	200						

DISTRIBUIÇÃO DE FREQUÊNCIA POR INTERVALO

Classe	SM	fi	Fi	fri	Fri	xi	Fi	Fri
1	1 ├ 3	90	90	0,45	0,45	2	200	1,00
2	3 ├ 5	50	140	0,25	0,70	4	110	0,55
3	5 ├ 7	30	170	0,15	0,85	6	60	0,30
4	7 ├ 9	20	190	0,10	0,95	8	30	0,15
5	9 ├ 11	10	200	0,05	1,00	10	10	0,05
	Σ	200		1,00				

Quantos funcionários ganham **entre** 1 e 3 SM ?
Quantos funcionários ganham **entre** 3 e 5 SM ?
Quantos funcionários ganham **entre** 7 e 9 SM ?
Qual o **percentual** de funcionários que ganha **entre** 1 e 3 SM ?
Qual o **percentual** de funcionários que ganha **entre** 3 e 5 SM ?
Qual o **percentual** de funcionários que ganha **entre** 7 e 9 SM ?

DISTRIBUIÇÃO DE FREQUÊNCIA POR INTERVALO

Classe	SM	fi	Fi	fri	Fri	xi	Fi	Fri
1	1 ├ 3	90	90	0,45	0,45	2	200	1,00
2	3 ├ 5	50	140	0,25	0,70	4	110	0,55
3	5 ├ 7	30	170	0,15	0,85	6	60	0,30
4	7 ├ 9	20	190	0,10	0,95	8	30	0,15
5	9 ├ 11	10	200	0,05	1,00	10	10	0,05
	Σ	200		1,00				

Quantos ganham **abaixo** de 3 SM ? 90
Quantos ganham **abaixo** de 7 SM ? 170
Quantos ganham **abaixo** de 11 SM ? 200
Qual o **percentual** que ganha **abaixo** de 3 SM ? 45%
Qual o **percentual** que ganha **abaixo** de 7 SM ? 85%
Qual o **percentual** que ganha **abaixo** de 11 SM ? 100%

DISTRIBUIÇÃO DE FREQUÊNCIA POR INTERVALO

Classe	SM	fi	Fi	fri	Fri	xi	Fi	Fri
1	1 ├ 3	90	90	0,45	0,45	2	200	1,00
2	3 ├ 5	50	140	0,25	0,70	4	110	0,55
3	5 ├ 7	30	170	0,15	0,85	6	60	0,30
4	7 ├ 9	20	190	0,10	0,95	8	30	0,15
5	9 ├ 11	10	200	0,05	1,00	10	10	0,05
	Σ	200		1,00				

Você vai visitar a empresa e encontra um funcionário.
Qual a **probabilidade** dele ganhar **entre** 7 e 9 SM ?
Qual a **probabilidade** dele ganhar **menos** de 5 SM ?
Qual a **probabilidade** dele ganhar 7 ou **mais** SM ?

TIPOS DE GRÁFICOS

HISTOGRAMA

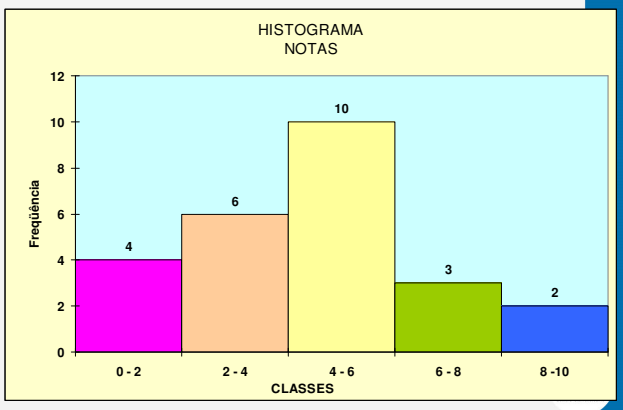
POLÍGONO DE FREQUÊNCIA

HISTOGRAMA

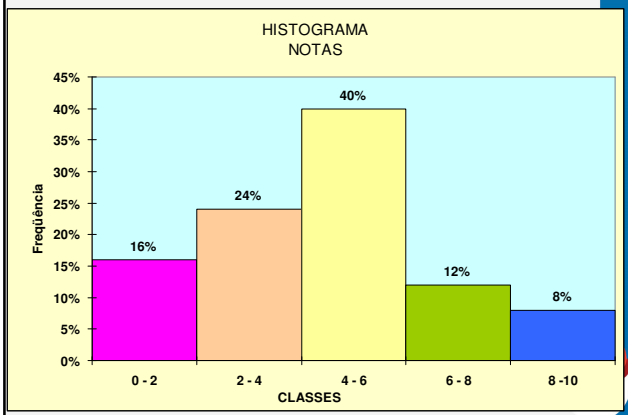
EXEMPLO ⇒ Notas de 25 alunos em uma prova de Estatística

Notas	Frequência(fi)	fri	fri
0 ├ 2	4	0,16	16%
2 ├ 4	6	0,24	24%
4 ├ 6	10	0,40	40%
6 ├ 8	3	0,12	12%
8 ├ 10	2	0,08	8%
	25	1,00	100%

HISTOGRAMA

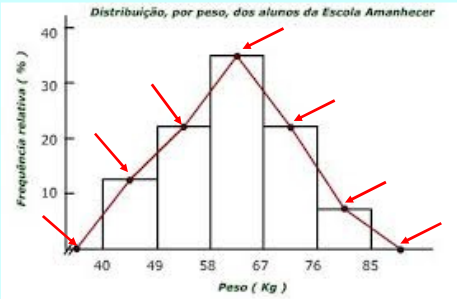


HISTOGRAMA



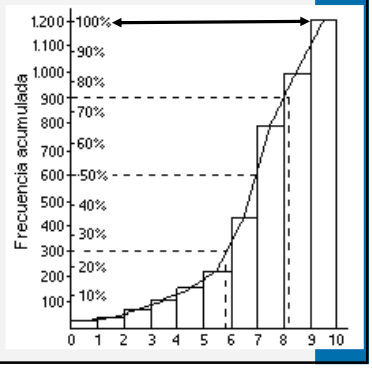
POLÍGONO DE FREQUÊNCIA

POLÍGONO DE FREQUÊNCIA ⇒ é construído ligando-se os pontos médios dos topos dos retângulos de um histograma.



POLÍGONO DE FREQUÊNCIA

OGIVA (Ogiva de Galton)
POLÍGONO DE FREQUÊNCIA ACUMULADA
É construído ligando-se os pontos médios dos topos dos retângulos de frequências acumuladas.



MEDIDAS DE POSIÇÃO

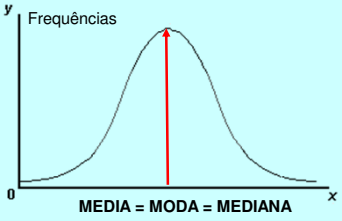
MEDIDAS DE POSIÇÃO

MEDIDAS DE POSIÇÃO ⇒ são aquelas que indicam a posição da distribuição no eixo das abcissas. Se dividem em Medidas de TENDÊNCIA CENTRAL e SEPARATRIZES.

MEDIDAS DE TENDÊNCIA CENTRAL

Tendem a se localizar no centro da distribuição

- MÉDIA ARITMÉTICA
- MEDIANA
- MODA



MEDIDAS DE POSIÇÃO

MÉDIA ARITMÉTICA ⇒ $\bar{x} = \frac{\sum x_i}{N} = \frac{Soma}{Quantidade}$

EXEMPLO: 6, 8, 0, 10 $\bar{x} = \frac{6+8+0+10}{4} = 6$

DESVIO EM RELAÇÃO A MÉDIA ⇒ $d_i = x_i$ - MÉDIA

d_1 = 6 - 6 = 0

d_2 = 8 - 6 = 2

d_3 = 0 - 6 = - 6

d_4 = 10 - 6 = 4

QUAL O PONTO MAIS DISTANTE DA MÉDIA ?

O PONTO 0 (zero) ! Está 6 unidades à ESQUERDA da MÉDIA !

SOMANDO OS DESVIOS VEM ⇒ **0 + 2 - 6 + 4 = 0 (ZERO) !**

PROPRIEDADE ⇒ **SOMA DOS DESVIOS** EM **RELAÇÃO** À **MÉDIA ARITMÉTICA É SEMPRE** IGUAL A **0** (**ZERO**) !

DESVIO EM RELAÇÃO A MÉDIA

Denominamos de desvio em relação à média, a diferença entre cada elemento de um conjunto de valores e a média aritmética desses valores. Assim, cada desvio é dado por:

$$d_i = (x_i - \bar{x}) \quad i = 1,2,\ldots.n$$

$\bar{x} = \frac{\sum x_i}{n} \rightarrow$ *Médi[illegible]ritmétic[illegible]*

Propriedade ⇒ soma dos desvios em relação à média aritmética é sempre igual a 0 (zero) !

$$\sum_{i=1}^{n} d_i = 0 \qquad \sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

MÉDIA ARITMÉTICA PARA DADOS DISTRIBUÍDOS POR FREQUÊNCIAS

$\bar{x} = \frac{\sum f_i x_i}{N}$ $N = \sum_{i=1}^{n} f_i$ *N = número de observações*

EXEMPLO-1 ⇒ Número de pessoas na família(**x**)

x	2	3	5	7	9
f	4	6	10	3	2

⇐ **PESSOAS** !

Qual o número médio de pessoas nas famílias ?

N = 4 + 6 + 10 + 3 + 2 = 25 ⇒ Soma das Frequências!

$$\bar{x} = \frac{4\times 2 + 6\times 3 + 10\times 5 + 3\times 7 + 2\times 9}{25} = \frac{115}{25} = 4,6$$

As famílias têm **em média**⇒ 4,6 pessoas !

MÉDIA PONDERADA ! **OS PONDERADORES SÃO AS FREQUÊNCIAS**

Média Aritmética para Dados Agrupados em Classes

$$\bar{x} = \frac{\sum f_i \times x_i}{\sum f_i} \quad i = 1,2 \ldots n$$

onde x_i é o ponto médio da classe i.

MEDIANA

MEDIANA ⇒ é o valor que ocupa a **POSIÇÃO CENTRAL** de um conjunto de **N** dados **ORDENADOS**. Assim se N for **ÍMPAR** a mediana é o **termo central**. Se **N** for **PAR**, a mediana será a **MÉDIA ARITMÉTICA** entre os **dois termos centrais**.

EXEMPLO-1

DADOS NÃO AGRUPADOS
NÚMERO DE OBSERVAÇOES ÍMPAR (N=9)

5	13	10	2	18	15	6	16	9

ORDENANDO OS DADOS VEM !

2	5	6	9	**10**	13	15	16	18

TERMO CENTRAL

Logo MEDIANA (Md) = **10**

REPRESENTATIVIDADE DA MEDIANA/MÉDIA

A Média Aritmética é sempre representativa de um conjunto de dados?

EXEMPLO ⇒ PESQUISA DE SALÁRIO

NÚMERO DE OBSERVAÇÕES = 5 pessoas entrevistadas

SALÁRIOS PESQUISADOS:

- ✓R$ 2.800,00
- ✓R$ 2.100,00
- ✓R$ 3.700,00
- ✓R$ 3.400,00
- ✓R$ 25.000,00

REPRESENTATIVIDADE DA MEDIANA/MÉDIA

EXEMPLO ⇒ PESQUISA DE SALÁRIO

R$ 2.800,00
R$ 2.100,00
R$ 3.700,00
R$ 3.400,00
R$ 25.000,00

MÉDIA = (2.800+2.100+3.700+3.400+25.000)/5
MÉDIA = 37.000/5
MÉDIA = **7.400**

MÉDIA DOS SALÁRIOS = **R$ 7.400,00**

PERGUNTA ⇒ **A MÉDIA É REPRESENTATIVA DOS SALÁRIOS ?**

REPRESENTATIVIDADE DA MEDIANA/MÉDIA

VAMOS CALCULAR A **MEDIANA** DOS SALÁRIOS

ORDENANDO OS SALÁRIOS TEMOS:

2.100	**2.800**	**3.400**	**3.400**	**25.000**

Md = 3.400
TERMO CENTRAL

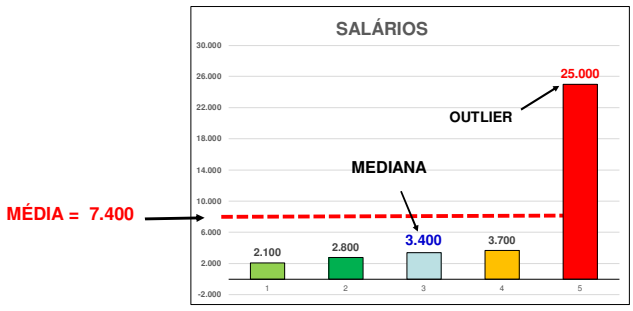
OUTLIER
Valor Extremo
Distorce o valor da Média

MÉDIA = **R$ 7.400,00**

MEDIANA = **R$ 3.400,00**

QUAL É MAIS REPRESENTATIVA ?
MÉDIA OU **MEDIANA** ?
POR QUÊ ?

A MEDIANA SÓ DEPENDE DA POSIÇÃO !

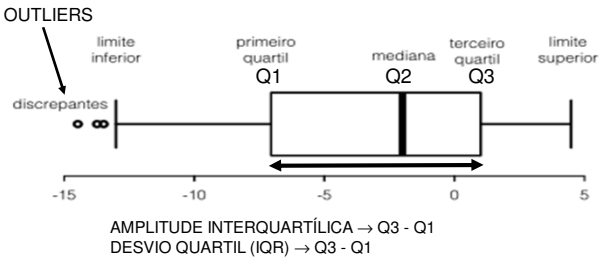


OUTLIERS → distorcem o valor da Média Aritmética!

MEDIA ARITMÉTICA → é fortemente impactadas por VALORES EXTREMOS!
ESTATÍSTICA ROBUSTA → para tratar OUTLIERS!

BOX-PLOT

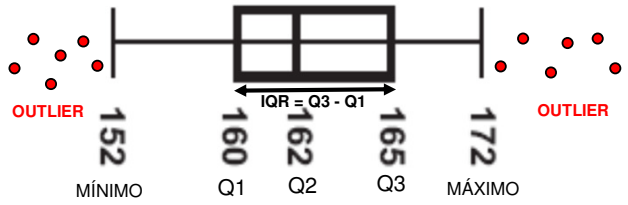
BOXPLOT (diagrama de caixa) **-** é uma ferramenta gráfica para representar a variação de dados observados de uma variável numérica por meio de QUARTIS.



AMPLITUDE INTERQUARTÍLICA → Q3 - Q1
DESVIO QUARTIL (IQR) → Q3 - Q1

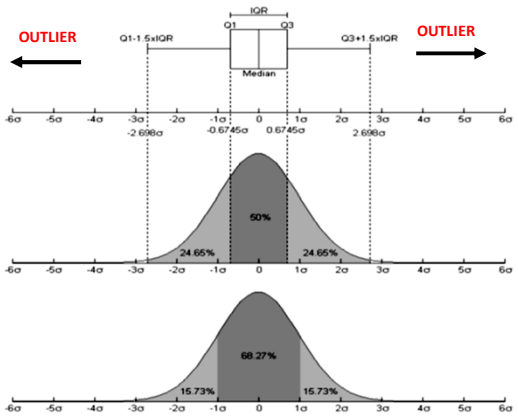
QUARTIS ⇒ SÃO 3 (Q1, Q2, Q3) - DIVIDEM A DISTRIBUIÇÃO EM 4 PARTES IGUAIS (**25%**)

BOX-PLOT



Valores INFERIORES a *Q*1-1.5×IQR e SUPERIORES a Q3+1.5×IQR são considerados ***outliers***.
Estão fora do intervalo: *Q*1 – 1.5×*IQR* < *valor* < *Q*3 + 1.5×*IQR*.
Limite inferior ⇒ ***Q*1 – 1.5×IQR**
Limite Superior ⇒ ***Q*3 + 1.5×IQR**

BOX-PLOT



MODA

MODA ⇒ é o valor que ocorre com MAIS FREQUÊNCIA em um conjunto de dados.

DADOS NÃO-AGRUPADOS

EXEMPLO-1 ⇒ 5, 6, 7, 7, 8, 9, **10**, **10**, **10**, 11

Nesta série a MODA é **10** !

UNIMODAL ⇒ **UMA** MODA !

EXEMPLO-2 ⇒ 1, 2, 3, **4**, **4**, 5, **6**, **6**, 7, 8, 9

Nesta série as MODAS são 4 e **6** !

BIMODAL⇒ **DUAS** MODAS !

MODA

DADOS AGRUPADOS - COM INTERVALO DE CLASSE

MODA DE CZUBER

$$Mo = l_i + \frac{D1}{D1 + D2} \times h$$

OUTRAS MODAS

- ✓ KING $Mo = l_i + \frac{f_{pos}}{f_{ant} + f_{pos}} \times h$
- ✓ PEARSON $Mo = 3 \times Md - 2 \times \bar{x}$
- ✓ MODA BRUTA *Ponto Médio da Classe Modal*

OBS ⇒ A MODA É EMPÍRICA ⇒ NÃO TEM RIGOR MATEMÁTICO COMO A MÉDIA E A MEDIANA !

MEDIDAS DE POSIÇÃO

OBSERVAÇÕES

- ✓ A **MÉDIA NEM SEMPRE** É UMA **MEDIDA REPRESENTATIVA** DE UM CONJUNTO DE DADOS.
- ✓ A **MEDIANA PODE SER MAIS REPRESENTATIVA QUE A MÉDIA**, EM ESPECIAL QUANDO OCORREM **VALORES EXTREMOS** (**OUTLIERS**).
- ✓ A MODA É UMA MEDIDA **EMPÍRICA**, NÃO TEM CONFIABILIDADE MATEMÁTICA COMO A **MÉDIA** E A **MEDIANA**.
- ✓ MEDIANA E MODA **NÃO SÃO** INFLUENCIADAS POR VALORES EXTREMOS (OUTLIERS)

PROPRIEDADES DA MÉDIA ARITMÉTICA

1. A **SOMA DOS DESVIOS** EM RELAÇÃO À MÉDIA É IGUAL A **ZERO** (0).
2. A MÉDIA ARITMÉTICA É UM VALOR CONTIDO ENTRE O MENOR VALOR (MIN) E O MAIOR VALOR (MAX).
3. **MULTIPLICANDO-SE** OU **DIVIDINDO-SE** TODOS OS VALORES DE UM CONJUNTO DE DADOS POR UMA **CONSTANTE**, A MÉDIA FICARÁ **MULTIPLICADA** OU **DIVIDIDA** POR ESTA **CONSTANTE**.
4. **SOMANDO-SE** OU **SUBTRAINDO-SE** A TODOS OS VALORES DE UM CONJUNTO DE DADOS UMA **CONSTANTE**, A MÉDIA FICARÁ **AUMENTADA** OU **SUBTRAÍDA** DESTA CONSTANTE.

Obs. As propriedades (3) e (4) valem para **TODAS** as **MEDIDAS DE POSIÇÃO**.

Exemplo: Suponha que a variável aleatória X tem média 4 e que são criadas as seguintes variáveis a partir de X:

Y = X + **3** ⇒ somando **3**

Z = **2**X + **1** ⇒ multiplicando por **2** e somando **1**

W = X/**2** ⇒ dividindo por **2**

Então as médias de Y, Z e W serão respectivamente:

$\bar{X} = 4 \quad \bar{Y} = \bar{X} + 3 \quad \bar{Y} = 4 + 3 \quad \bar{Y} = 7$

$\bar{Z} = 2\bar{X} + 1 \quad \bar{Z} = 2 \times 4 + 1 \quad \bar{Z} = 9$

$\bar{W} = \frac{\bar{X}}{2} \quad \bar{W} = \frac{4}{2} \quad \bar{W} = 2$

Exemplo - Suponha que a MÉDIA dos salários dos empregados de uma empresa pública é de **R$ 4.000,00** !

Vamos imaginar que o sindicato negociou um aumento **FIXO** de **R$ 200,00** reais para cada funcionário além de um aumento percentual de **20%** sobre o salário antigo.
Qual a nova MÉDIA salarial dessa empresa?

AUMENTO FIXO = **R$ 200,00** ⇒ cada salário será aumentado em **R$ 200,00**.
AUMENTO PERCENTUAL = **20%** ⇒ cada salário será multiplicado por **1,20** !

Assim, por exemplo, se um funcionário ganhava **R$ 2.000,00** o seu novo salário será:
Salário Novo = **2.000** + **200** + **2.000** × **0,20** =
200 + **2.000**×(1+**0,20**) = **200** + **2.000**×**1,20** = **200** + 2.400 = **2.600**
Ou seja, o novo salário será de **R$ 2.600,00 !**

MAS...QUANTO SERÁ A NOVA MÉDIA SALARIAL?

MÉDIA ANTIGA = R$ 4.000,00
AUMENTO FIXO = **200**
AUMENTO PERCENTUAL = **20%** (cada salário ficará multiplicado por **1,20** !

MÉDIA NOVA = 4.000 × **1,20** + **200**

MÉDIA NOVA = 4.800 + **200**

MÉDIA NOVA = R$ 5.000,00 !!!!

PERGUNTA: ESSA PROPRIEDADE VALE **SOMENTE** PARA A MÉDIA?

NÃO ! VALE PARA **TODAS** AS MEDIDAS DE POSIÇÃO !

FEROZ !

FÓRMULAS

$$\bar{x} = \frac{\sum f_i \times x_i}{\sum f_i}$$ MÉDIA

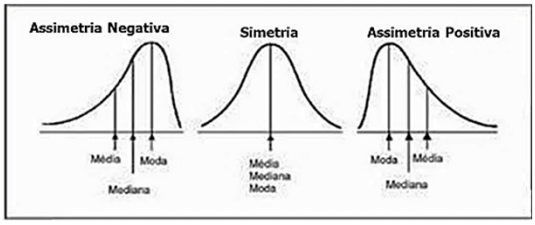
$$Md = l_i + \frac{[\frac{\sum f_i}{2} - Faa]}{f_{md}} \times h$$ MEDIANA

$$Mo = l_i + \frac{D1}{D1 + D2} \times h$$ MODA

POSIÇÃO RELATIVA DA MÉDIA E DA MEDIANA

AS DISTRIBUIÇÕES PODEM SER:

- SIMÉTRICAS
- ASSIMÉTRICAS **POSITIVA** (À **DIREITA**)
- ASSIMÉTRICAS NEGATIVA (À ESQUERDA)

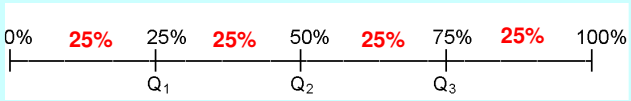


Assimetria negativa Simetria Assimetria positiva

SEPARATRIZES - QUANTIS

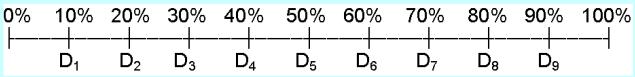
SEPARATRIZES

QUARTIS ⇒ Dividem os valores de uma série em **4** (**quatro**) partes iguais (frequências).



SEPARATRIZES

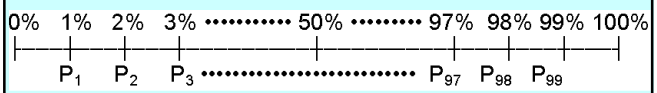
DECIS ⇒ separatrizes que dividem a série em **10** partes iguais.



OBS ⇒ OS **DECIS** SÃO **9** E DIVIDEM A DISTRIBUIÇÃO EM **10 PARTES** IGUAIS !

SEPARATRIZES

PERCENTIS ⇒ são os valores que dividem uma série em **100** partes iguais.



OBS ⇒ OS PERCENTIS SÃO **99** E DIVIDEM A DISTRIBUIÇÃO EM **100 PARTES** IGUAIS !

EXEMPLOS - RESULTADOS ENADE
NOTAS NO ENADE

Agrupamento	Ingressantes			
	Até P25	P25 a P50	P50 a P75	P75 a P100
Instituição	3,2	25,8	32,3	38,7
Brasil	25,2	25,0	25,0	24,9

PERCENTUAIS DE ACERTOS DE QUESTÕES

QUE INFORMAÇÕES PODE-SE TIRAR DESSA TABELA ?

A INSTITUIÇÃO ESTÁ ACIMA DA MÉDIA BRASILEIRA ?

QUAL O **PERCENTUAL** DE ALUNOS **INGRESSANTES NA INSTITUIÇÃO** QUE ACERTOU **MAIS DE 75%** DAS QUESTÕES?

QUAL O **PERCENTUAL** DE ALUNOS **INGRESSANTES NO BRASIL** QUE ACERTOU **MAIS DE 75%** DAS QUESTÕES?

EXEMPLOS - RESULTADOS ENADE
NOTAS NO ENADE

Agrupamento	Ingressantes			
	Até P25	P25 a P50	P50 a P75	P75 a P100
Instituição	3,2	25,8	32,3	38,7
Brasil	25,2	25,0	25,0	24,9

PERCENTUAIS DE ACERTOS DE QUESTÕES

QUAL O **PERCENTUAL** DE ALUNOS **INGRESSANTES NA INSTITUIÇÃO** QUE ACERTOU **MENOS DE 25%** DAS QUESTÕES?

QUAL O **PERCENTUAL** DE ALUNOS **INGRESSANTES NO BRASIL** QUE ACERTOU **MENOS DE 25%** DAS QUESTÕES?

EXEMPLOS - RESULTADOS ENADE
NOTAS NO ENADE

Agrupamento	Concluintes			
	Até P25	P25 a P50	P50 a P75	P75 a P100
Instituição	2,7	27,0	32,4	37,8
Brasil	25,2	25,1	24,7	25,0

QUAL O **PERCENTUAL** DE ALUNOS **CONCLUINTES NA INSTITUIÇÃO** QUE ACERTOU **MAIS DE 75%** DAS QUESTÕES?

QUAL O **PERCENTUAL** DE ALUNOS **CONCLUINTES NO BRASIL** QUE ACERTOU **MAIS DE 75%** DAS QUESTÕES?

EXEMPLOS - RESULTADOS ENADE
NOTAS NO ENADE

Agrupamento	Concluintes			
	Até P25	P25 a P50	P50 a P75	P75 a P100
Instituição	2,7	27,0	32,4	37,8
Brasil	25,2	25,1	24,7	25,0

QUAL O **PERCENTUAL** DE ALUNOS **CONCLUINTES NA INSTITUIÇÃO** QUE ACERTOU **MENOS DE 25%** DAS QUESTÕES?

QUAL O **PERCENTUAL** DE ALUNOS **CONCLUINTES NO BRASIL** QUE ACERTOU **MENOS DE 25%** DAS QUESTÕES?

EXEMPLOS - RESULTADOS ENADE

Agrupamento		Ingressantes			
		Até P25	P25 a P50	P50 a P75	P75 a P100
Região	Norte	31,3	27,2	25,1	16,3
	Nordeste	25,3	24,5	25,2	25,1
	Sudeste	23,9	24,7	25,1	26,3
	Sul	23,9	25,2	25,3	25,6
	Centro-Oeste	30,4	25,3	23,3	21,0

QUAL A REGIÃO COM **MELHOR** DESEMPENHO DOS INGRESSANTES ?

QUAL A REGIÃO COM **PIOR** DESEMPENHO DOS INGRESSANTES ?

EXEMPLOS - RESULTADOS ENADE

Agrupamento		Concluintes			
		Até P25	P25 a P50	P50 a P75	P75 a P100
Região	Norte	34,0	28,5	23,1	14,4
	Nordeste	25,1	25,0	24,0	25,9
	Sudeste	24,5	25,0	24,9	25,5
	Sul	20,9	24,6	26,3	28,2
	Centro-Oeste	34,2	24,7	21,8	19,3

QUAL A REGIÃO COM **MELHOR** DESEMPENHO DOS CONCLUINTES ?

QUAL A REGIÃO COM **PIOR** DESEMPENHO DOS CONCLUINTES ?

MEDIDAS DE DISPERSÃO

MOTIVAÇÃO

As medidas de posição (média, mediana e moda) não são suficientes para caracterizar perfeitamente um conjunto de dados.

Duas distribuições (dois conjuntos de dados) podem ter a mesma média, mediana e moda mas serem diferentes.

Em uma delas, os valores podem se concentrar fortemente em torno da média, na outra, podem se espalhar nos dois lados desse valor médio.

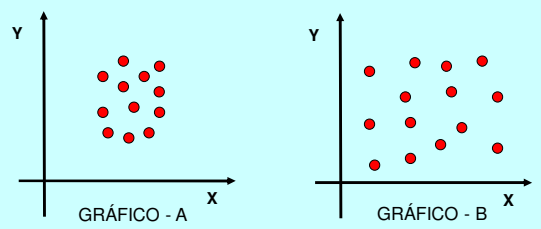
MEDIDAS DE DISPERSÃO

DISPERSÃO (variabilidade) ⇒ é a maior ou menor diversificação dos valores de uma variável, em torno de um valor de tendência central tomado como referência (MÉDIA ou MEDIANA).

Mais usadas são:

- **VARIÂNCIA**
- **DESVIO PADRÃO**
- **COEFICIENTE DE VARIAÇÃO**

MEDIDAS DE DISPERSÃO



ONDE A DISPERSÃO É MAIOR ⇒ A ou B ?

MEDIDAS DE DISPERSÃO

OBSERVE OS CONJUNTOS X E Y

X = 11; 9; 8; 12; 7; 10; 10; 13 ←

Y = 2; 18; 1; 5; 19; 5; 0; 30 ←

Calculando as médias dos conjuntos X e Y obtemos:

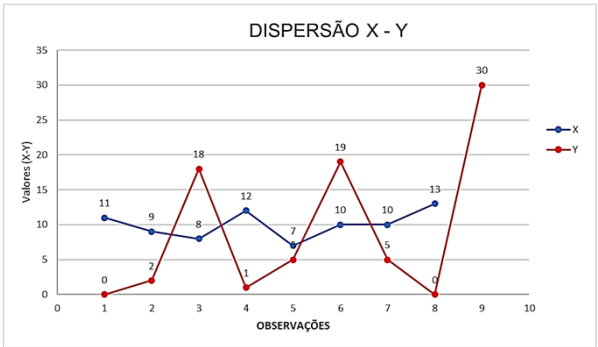
$$\bar{x} = \frac{\sum x_i}{n} = \frac{11+9+8+12+7+10+10+13}{8} = \frac{80}{8} \quad \bar{x} = 10$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{2+18+1+5+19+5+0+30}{8} = \frac{80}{8} \quad \bar{y} = 10$$

Os conjuntos X e Y são semelhantes ?

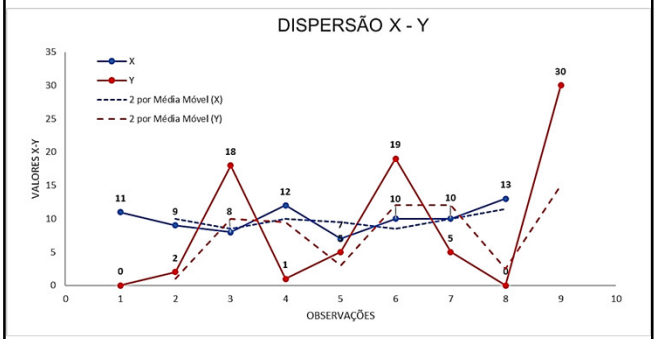
QUEM TEM A MAIOR DISPERSÃO X OU Y ?

CONJUNTOS X E Y

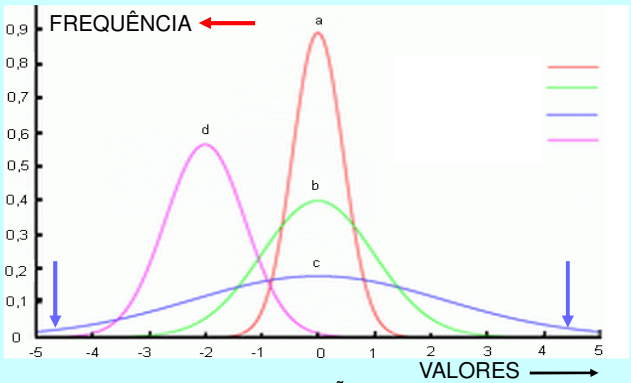


CONJUNTOS X E Y

PREVISÃO - MÉDIA MÓVEL (t = 2)

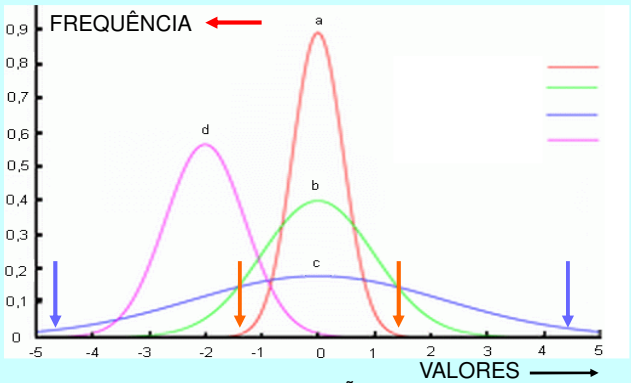


MEDIDAS DE DISPERSÃO



QUEM TEM A MAIOR DISPERSÃO ?

MEDIDAS DE DISPERSÃO



QUEM TEM A MENOR DISPERSÃO ?

MEDIDAS DE DISPERSÃO

PRINCIPAIS MEDIDAS DE DISPERSÃO

- DESVIO PADRÃO
- VARIÂNCIA
- COEFICIENTE DE VARIAÇÃO

MEDIDAS DE DISPERSÃO

VARIÂNCIA $\Rightarrow$ Símbolos $\Rightarrow$ S^2 ou σ^2

S^2 $\Rightarrow$ para **AMOSTRA**

$\sigma^2 \Rightarrow$ para **POPULAÇÃO**

$$d_i = x_i - \bar{x}$$

FÓRMULA (dados não agrupados) $\Rightarrow$ $$S^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

MÉDIA DOS QUADRADOS DOS DESVIOS !

PROPRIEDADES DA VARIÂNCIA

- SOMANDO-SE ou SUBTRAINDO-SE uma **CONSTANTE** a todos os elementos de um conjunto de dados, a **variância** deste conjunto **NÃO DE ALTERA**.
- MULTIPLICANDO-SE ou DIVIDINDO-SE todos os elementos de um conjunto de dados por uma CONSTANTE (diferente de zero), a variância deste conjunto fica MULTIPLICADA ou DIVIDIDA pelo QUADRADO desta constante.

PROPRIEDADES DA VARIÂNCIA

SOMAR/SUBTRAIR UMA CONSTANTE

X = 1 ; 4 ; 7 ; 10 $\Rightarrow$ Y = X **+ 2**

Y = 3 ; 6 ; 9 ; 12 $S^2y = S^2x \Rightarrow$ **NÃO SE ALTERA**

MULTIPLICAR/DIVIDIR POR UMA CONSTANTE

X = 1 ; 4 ; 7 ; 10 $\Rightarrow$ Y = **4**X

Y = 4 ; 16 ; 28 ; 40 $S^2y = \mathbf{4^2} \times S^2x$

$S^2y = 4^2 \times S^2x \Rightarrow$ fica multiplicada pelo QUADRADO da constante

PROPRIEDADES DA VARIÂNCIA

Exemplo: Sejam X, Y, Z e W variáveis aleatórias assim definidas:

Y = X + **2** ⇒ $S_Y^2 = S_X^2$

⇒ **SOMAR** uma **CONSTANTE não altera** a **variância**.

Z = **4**X + 6 ⇒ $S_Z^2 = 4^2 \times S_X^2$

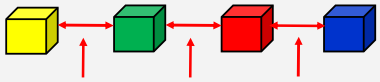
⇒ **MULTIPLICAR** por uma **CONSTANTE** a VARIÂNCIA fica **multiplicada** pelo **QUADRADO** da **constante**.

W = X/**4** ⇒ $S_W^2 = \frac{S_X^2}{4^2}$

⇒ **DIVIDIR** por uma **constante** a VARIÂNCIA fica **DIVIDIDA** pelo **QUADRADO** da **constante**.

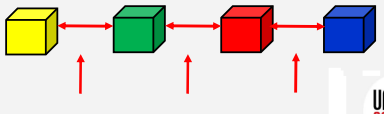
PROPRIEDADES DA VARIÂNCIA

SOMAR UMA CONSTANTE



AS DISTÂNCIAS ENTRE OS CUBOS NÃO MUDAM!

SUBTRAIR UMA CONSTANTE



AS DISTÂNCIAS ENTRE OS CUBOS TAMBÉM NÃO MUDAM!

CORREÇÃO DA FÓRMULA DA VARIÂNCIA PARA O CASO DE AMOSTRA - FATOR DE CORREÇÃO DE BESSEL

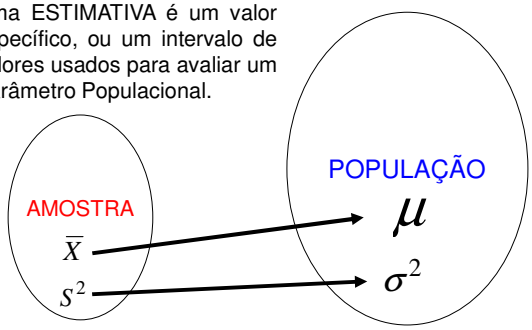
$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

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

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

FATOR DE CORREÇÃO DE BESSEL ⇒ $\frac{n}{n-1}$

Uma ESTIMATIVA é um valor específico, ou um intervalo de valores usados para avaliar um Parâmetro Populacional.



Um ESTIMADOR é uma CARACTERÍSTICA da **AMOSTRA**, utilizado para obter uma aproximação do parâmetro na **POPULAÇÃO**.

MEDIDAS DE DISPERSÃO

DESVIO PADRÃO ⇒ é a raiz quadrada da variância.

SÍMBOLO ⇒ S ou σ

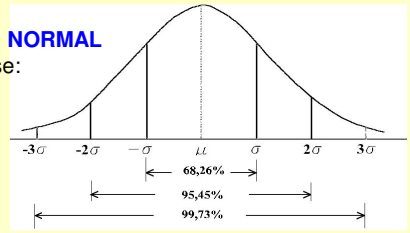
S ⇒ para **AMOSTRA**

σ ⇒ para **POPULAÇÃO**

FÓRMULA ⇒ $S = \sqrt{S^2}$

PROPRIEDADES DO DESVIO PADRÃO

Em uma distribuição **NORMAL** (forma de sino) tem-se:



INTERVALO	PERCENTUAL (%)
μ ± **1** σ	**68,26%** das observações
μ ± **2** σ	**95,45%** das observações
μ ± **3** σ	**99,73%** das observações

μ = Média

σ = Desvio Padrão

MEDIDAS DE DISPERSÃO

COEFICIENTE DE VARIAÇÃO = DESVIO PADRÃO / MÉDIA

$$CV = \frac{S}{\bar{x}}$$ ADIMENSIONAL (%) ⇒ NÃO TEM DIMENSÃO !

Exemplo: Para um conjunto de dados relativos a estaturas têm-se: Média =161 cm e S=5,57 cm. Achar o CV deste conjunto de dados.

$$CV = \frac{S}{\bar{x}} \quad CV = \frac{5{,}57\,cm}{161\,cm} = 0{,}0345 = 3{,}45\%$$

MEDIDAS DE DISPERSÃO

Exemplo: Consideremos os resultados das medidas de altura e peso de um mesmo grupo de indivíduos exibidos na tabela abaixo:

Medidas	Média	S
Estatura	175 cm	**5,0 cm**
Peso	68 Kg	**2,0 Kg**

DISPERSÃO ABSOLUTA

Qual apresenta maior grau de dispersão - **Estatura** ou **Peso** ?

Podemos comparar **cm** com **kg** ?

$$CVEstatura\ s = \frac{5\,cm}{175\,cm} = 0{,}0285 = 2{,}85\%$$

$$CVPeso = \frac{2\,kg}{68\,kg} = 0{,}0294 = 2{,}94\%$$

BIBLIOTECAS / PACOTES

NUMPY – https://numpy.org/

PANDAS https://pandas.pydata.org/

SCIPY https://scipy.org/

MATPLOTLIB.PYPLOT
https://matplotlib.org/

MÓDULOS MATEMÁTICOS E NUMÉRICOS
https://docs.python.org/pt-br/3/library/numeric.html

COLAB
Olá, este é o Colaboratory - Colaboratory (google.com)

Carpe Diem