



# Modelagem, Extração e Manipulação de Dados

---

BLOCO: B.I. E ANÁLISE DE DADOS

PROF. RODRIGO EIRAS, M.SC.

[ETAPA 5] AULAS 1 E 2 – ANALISAR DADOS NO POWER BI - PARTE 1



# Na última aula...

---

- Visualização de Dados
- Criar relatórios no Power BI
- Classificações e Ordenações
- Arredondamentos



# Agenda

---

- Dados agregados
- Escalas
- Análise Exploratória de Dados
- Continuar exercício da aula anterior

# Preambulo de fundamentos

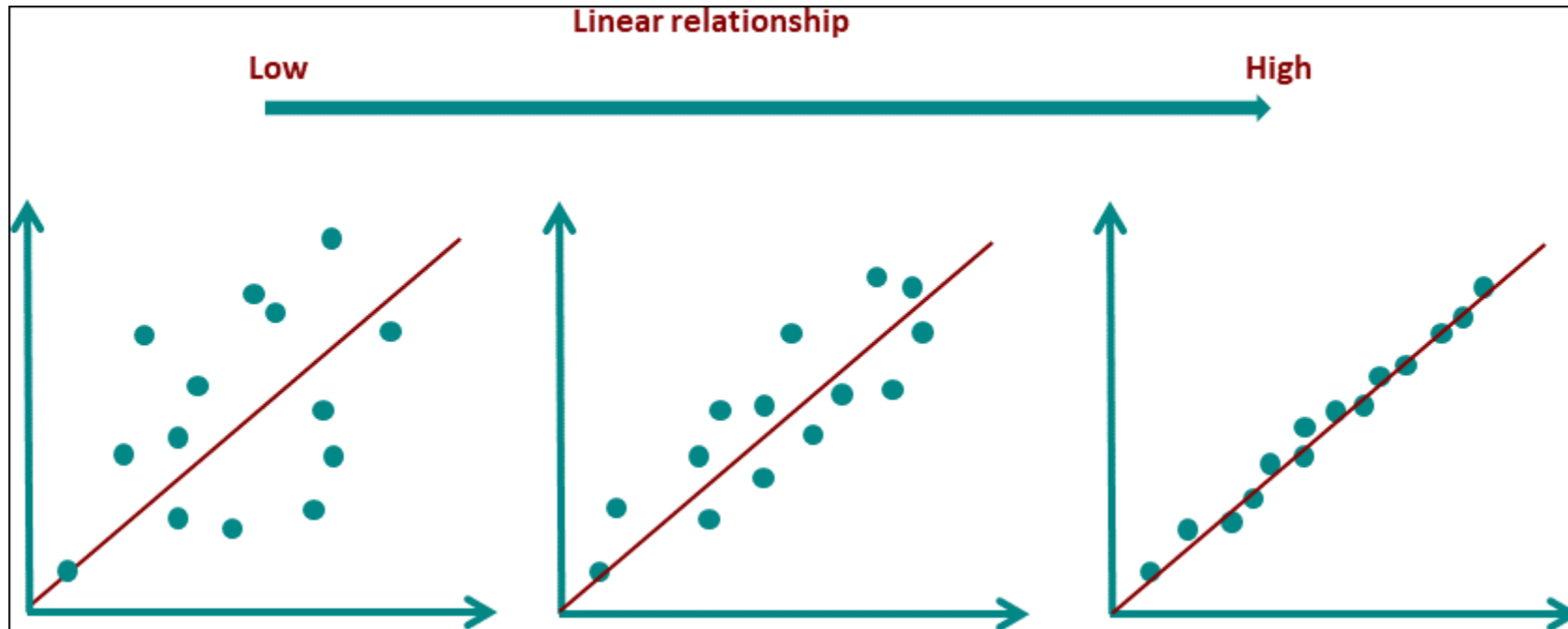
---

O que é Estatística?

- É fundamental na análise de dados provenientes de quaisquer processos onde exista **VARIABILIDADE**.
- Uso de informações na: **coleção, apresentação, análise e tomada de decisões**, para solucionar problemas.

# Preambulo de fundamentos

---



$$Y = X\beta + \varepsilon$$

# Estatística

- Uma estatística é uma quantidade que é calculada dos dados amostrados.
- Ela é usada para dar informações a respeito de valores desconhecidos da correspondente população.
- Por exemplo, a média dos dados amostrados é utilizada para dar informações sobre toda a média da população da qual a amostra foi retirada.

# Grandes áreas da estatística

## Amostragem e planejamento de experimentos

- coleção ou coleta de dados

## Estatística descritiva

- organização, apresentação e sintetização de dados

## Estatística inferencial

- métodos para tomada de decisões, nas situações onde existem incertezas e VARIAÇÕES.

É o processo de escolha da amostra. É a parte inicial de qualquer estudo estatístico. Consiste na escolha criteriosa dos elementos a serem submetidos ao estudo.

- Ex. Pesquisas sobre tendências de votação.
- Escolha da amostra, redação do questionário, a entrevista, a codificação dos dados, a apuração dos resultados são ETAPAS FUNDAMENTAIS deste tipo de pesquisa.

# Amostragem



# Estatística descritiva

É a parte mais conhecida.

Quem vê o noticiário, na televisão ou nos jornais, sabe quão frequente é o uso de média, índices e gráficos nas notícias.

- Exemplo:
  - O INPC, Índice Nacional de Preços ao Consumidor
    - Aumento dos produtos da cesta básica.
  - Anuário Estatístico Brasileiro
    - educação, saúde, transporte, economia, cultura etc.

# inferir

Deduzir, concluir por inferência ou por dedução, geralmente partindo de indícios, de fatos ou de raciocínios: com esse aparelho é possível (...)

[J] Dicio.com.br

## Estatística Inferencial

---

A estatística Inferencial faz uso das informações retiradas da amostra para conclusões (inferências), a respeito da população da qual a amostra foi retirada.

# População e amostra

---

O estudo de qualquer fenômeno, seja ele natural, social, econômico ou biológico, exige a **coleta e análise de dados** estatísticos.

- **População** é a coleção de **todas as observações** sobre determinado fenômeno.
- **Amostra** é o conjunto de **dados efetivamente observados**, ou **extraídos da população**.
- Exemplo: Determinação do consumo de óleo diesel em ônibus, avaliação de um programa de ensino, renda média per capita em diversas regiões do país etc.

# Inferência

---

A tomada de decisões sobre a população, com base nos dados da amostra, constitui o problema central da **INFERÊNCIA ESTATÍSTICA**.

A tais decisões estão sempre associados um **grau de incerteza** e, conseqüentemente, uma **probabilidade de erro**.

- Exemplo: Teste sobre medicamentos, experimentos agrícolas, análise financeira, consumo de energia etc.

# Apresentação dos dados

---

Técnicas que permitem detectar e corrigir erros e inconsistências ocorridos durante um processo de coleta de dados e determinar as principais características destes dados.

- Grupamento de dados;
- Construção de distribuições de frequência;
- Gráficos.

Salários (R\$)	Funcionários
500,00	10
1 000,00	5
1 500,00	6
2 000,00	15
5 000,00	8
10 000,00	2

# Tipos de Variáveis

---

## Qualitativa

- Nominal
  - Região de Procedência
- Ordinal
  - Educação, Classe Social

## Quantitativa

- Discreta
  - Número de Filhos
- Contínua
  - Peso de Indivíduos, Salários em R\$

# Grupamento de dados

---

Alturas, expressas em centímetros de 30 atletas de um clube.

168	172	170	181	169	173
164	175	182	177	176	173
170	186	183	170	168	166
169	180	175	164	181	179
172	169	174	171	178	166

# CONSTRUÇÃO DE DISTRIBUIÇÕES DE FREQUÊNCIA

---

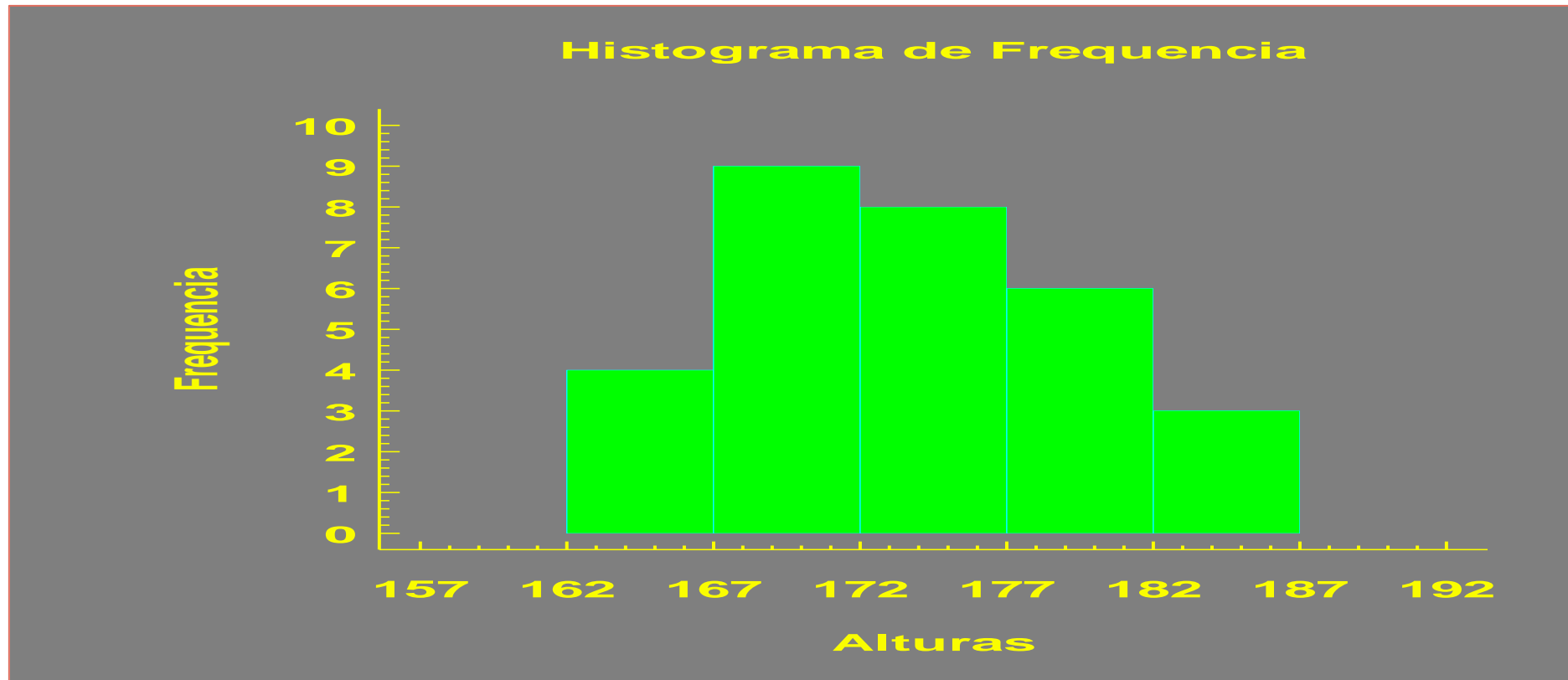
Alturas em cm. de 30 atletas

Tabulação de Frequências							
Classe	Limite Inferior	Limite Superior	Ponto Médio	Frequência	Frequência Relativa	Frequência Acumulada	Frequência Rel. Acum
1	162.000	167.000	164.500	4	0.133	4	0.133
2	167.000	172.000	169.500	9	0.300	13	0.433
3	172.000	177.000	174.500	8	0.267	21	0.700
4	177.000	182.000	179.500	6	0.200	27	0.900
5	182.000	187.000	184.500	3	0.100	30	1.000
-----							
Média = 173.367		Desvio Padrão = 5.89847		Mediana = 172.5			



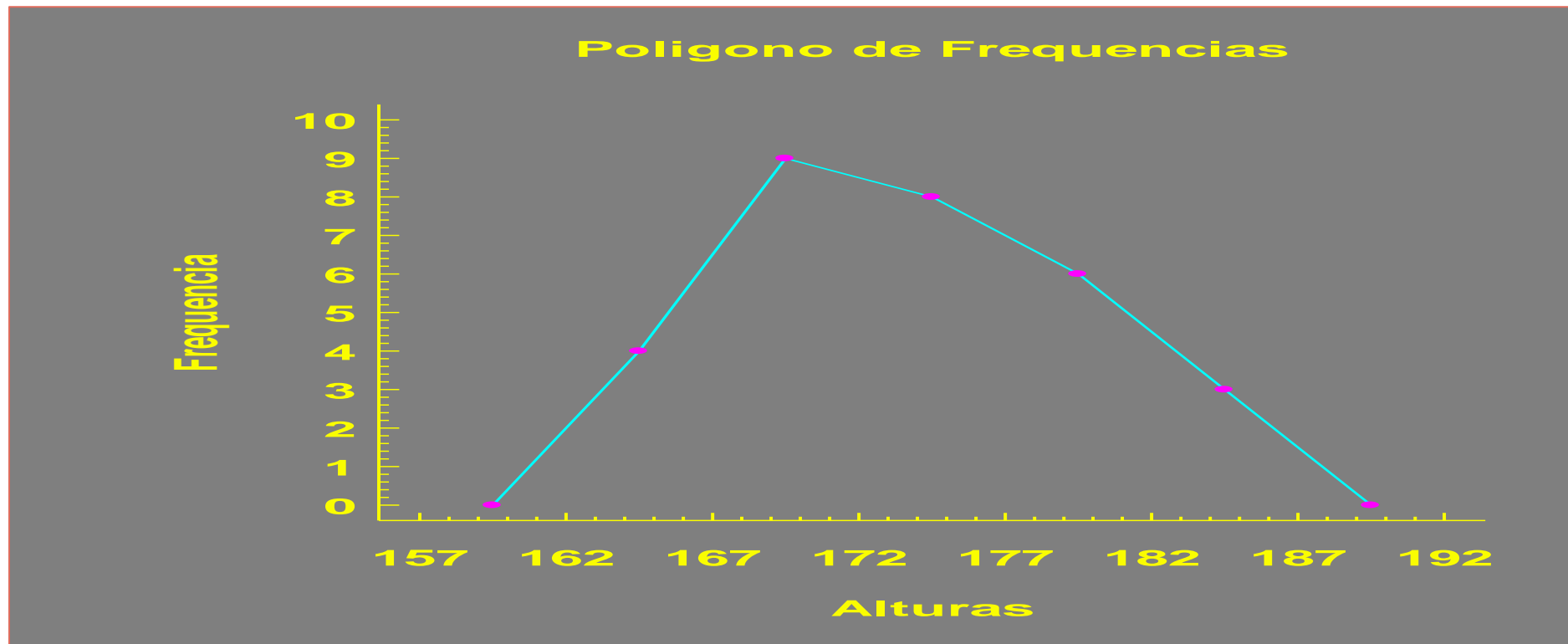
# GRÁFICOS

---



# GRÁFICOS

---



# GRUPAMENTO DE DADOS

---

Número de filhos em 25 famílias observadas

0	1	2	3	4
5	1	1	2	2
2	3	3	2	2
3	4	5	1	2
2	3	2	3	2

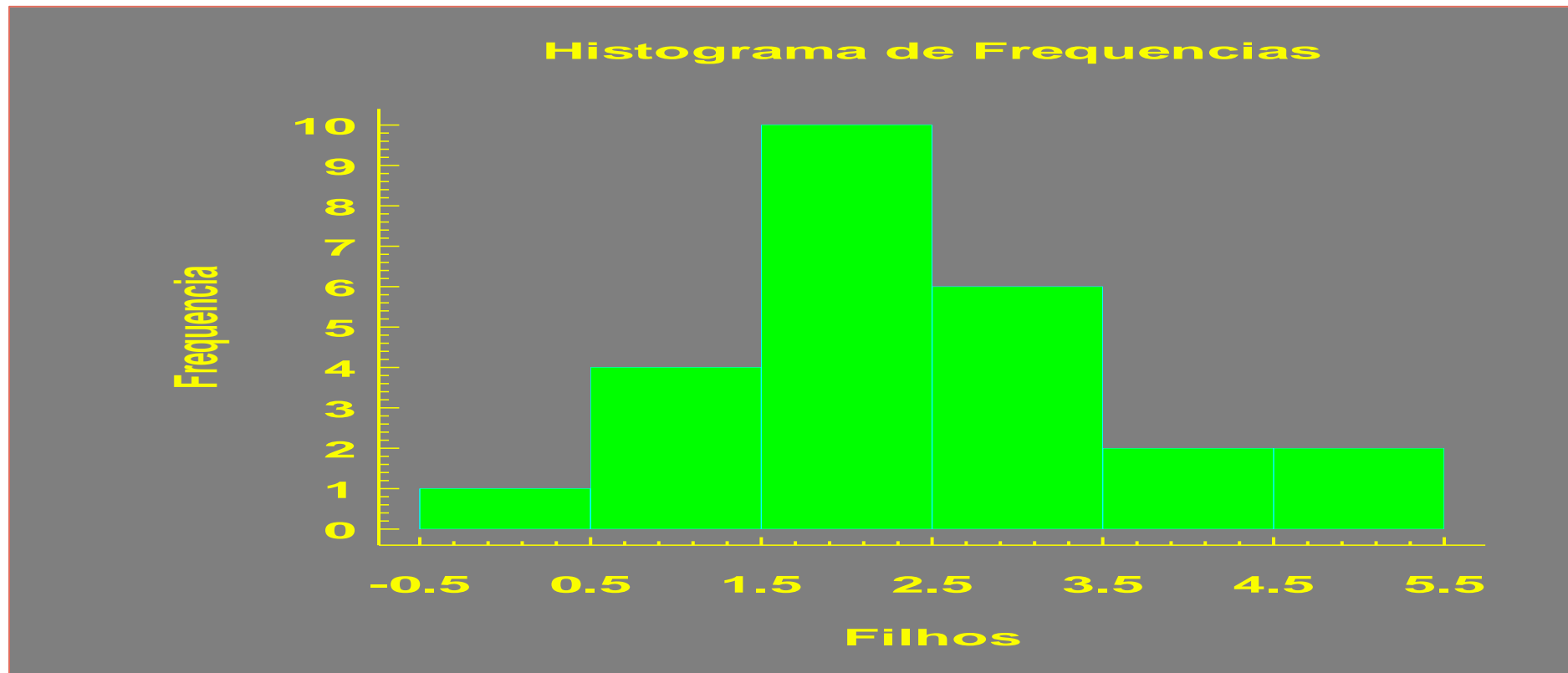
# CONSTRUÇÃO DE DISTRIBUIÇÕES DE FREQUÊNCIA

## Número de filhos em 25 famílias observadas

Tabulação de Freqüências							
Classe	Limite Inferior	Limite Superior	Ponto Médio	Freqüência	Freqüência Relativa	Freqüência Acumulada	Freqüência Rel. Acum.
1	-0.5000	0.5000	0.000	1	0.0400	1	0.0400
2	0.5000	1.5000	1.000	4	0.1600	5	0.2000
3	1.5000	2.5000	2.000	10	0.4000	15	0.6000
4	2.5000	3.5000	3.000	6	0.2400	21	0.8400
5	3.5000	4.5000	4.000	2	0.0800	23	0.9200
6	4.5000	5.5000	5.000	2	0.0800	25	1.0000
-----							
Média = 2.4		Desvio Padrão = 1.22474		Mediana = 2			

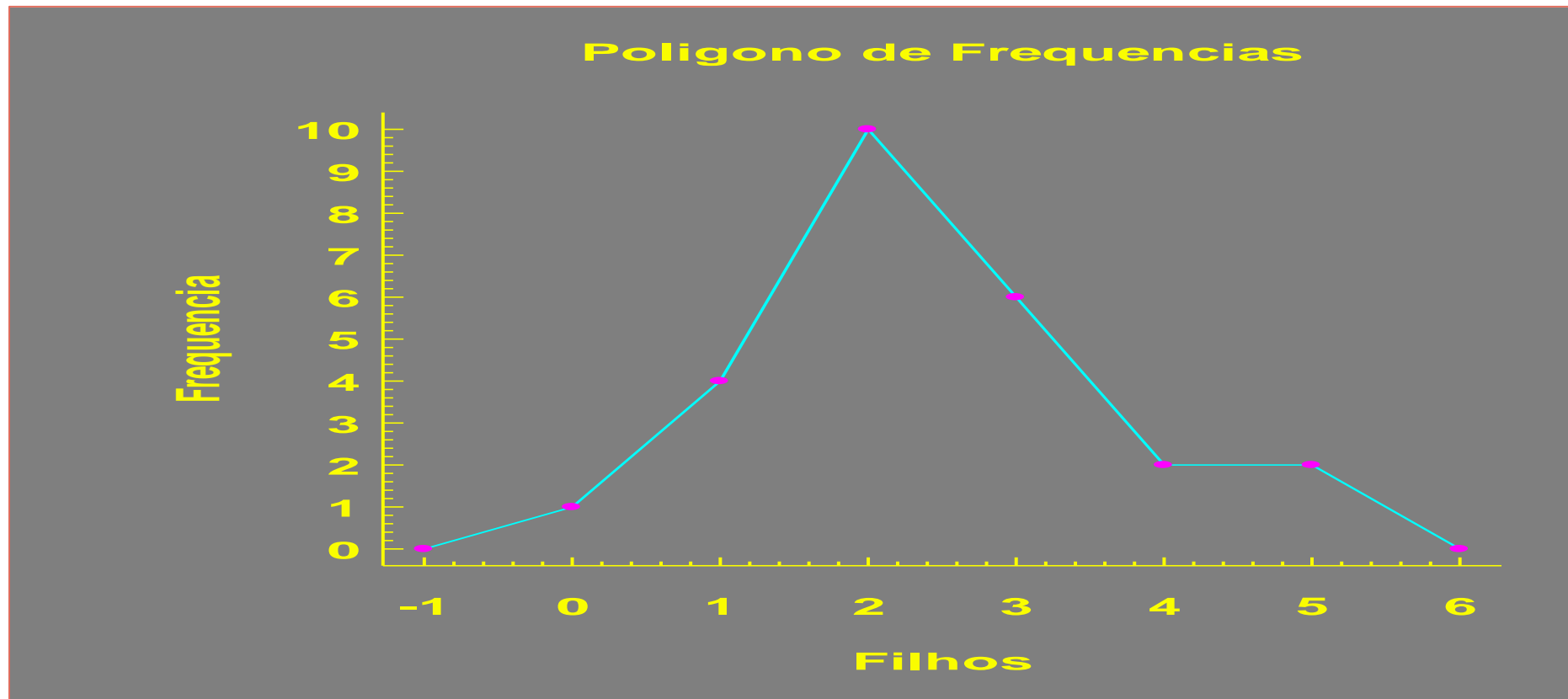
# GRÁFICOS

---



# GRÁFICOS

---



# Estatísticas Descritivas

---

▫ Tamanho da Amostra

▫ Média

▫ Mediana

▫ Moda

▫ Variância

▫ Desvio-padrão

▫ Erro-padrão

▫ Mínimo

▫ Máximo

▫ Amplitude

▫ Quartil Inferior

▫ Quartil Superior

▫ Intervalo Inter-quartil

▫ Somatório

# Estatística Clássica

---

Suposições Probabilísticas das Variáveis Envolvidas

Declarações sobre os Parâmetros ou Modelo Utilizado

“USE A ESTATÍSTICA COMO O BÊBADO USA OS POSTES - MAIS  
PELO APOIO QUE PELA ILUMINAÇÃO”

Andrew Lang



# Análise Exploratória de Dados

---

Tukey J. W. (1977)

- Técnicas Visuais

Dados = Modelo + Resíduos

Modelo = parte Suave

Resíduos = parte Grosseira

$$Y = X\beta + \varepsilon$$

# Análise Exploratória de Dados

---

## Ferramentas Principais

- Esquema de cinco números - “5-number summary”
- Desenho Esquemático - “Box-Plot”



# CONSTRUÇÃO DE DISTRIBUIÇÕES DE FREQUÊNCIA

Alturas em cm. de 30 atletas

Tabulação de Frequências							
Classe	Limite Inferior	Limite Superior	Ponto Médio	Frequência	Frequência Relativa	Frequência Acumulada	Frequência Rel. Acum
1	162.000	167.000	164.500	4	0.133	4	0.133
2	167.000	172.000	169.500	9	0.300	13	0.433
3	172.000	177.000	174.500	8	0.267	21	0.700
4	177.000	182.000	179.500	6	0.200	27	0.900
5	182.000	187.000	184.500	3	0.100	30	1.000
-----							
Média = 173.367		Desvio Padrão = 5.89847		Mediana = 172.5			

# Esquema ou Resumo de 5 Números

Sugestão (Tukey) - 1977

(i) a mediana

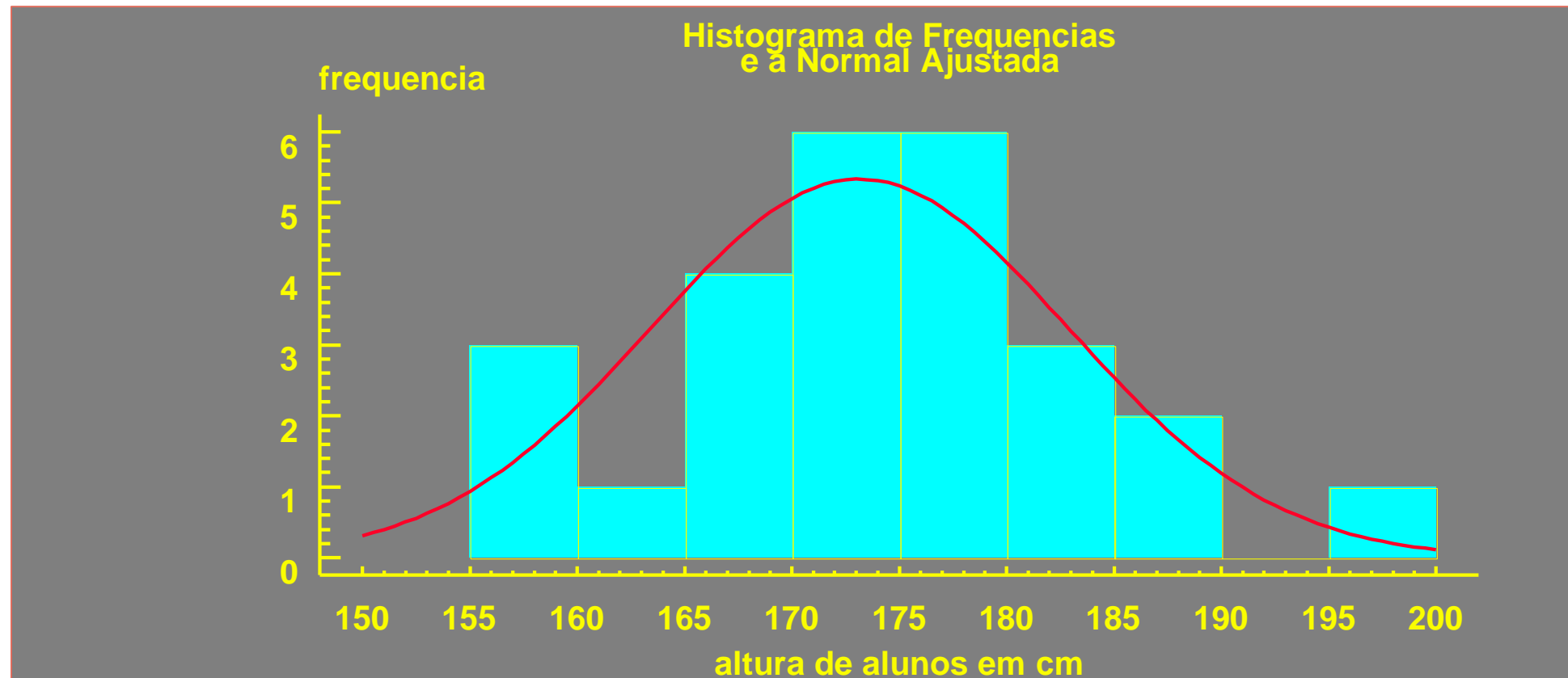
(ii) os extremos (mínimo e máximo)

(iii) os quartis ou juntas (inferior e superior)

- A Mediana é uma Medida Resistente, não é afetada por valores extremos.
- A Média amostral e o Desvio-padrão são afetados por valores extremos
- Não temos ideia quanto a simetria da distribuição dos dados

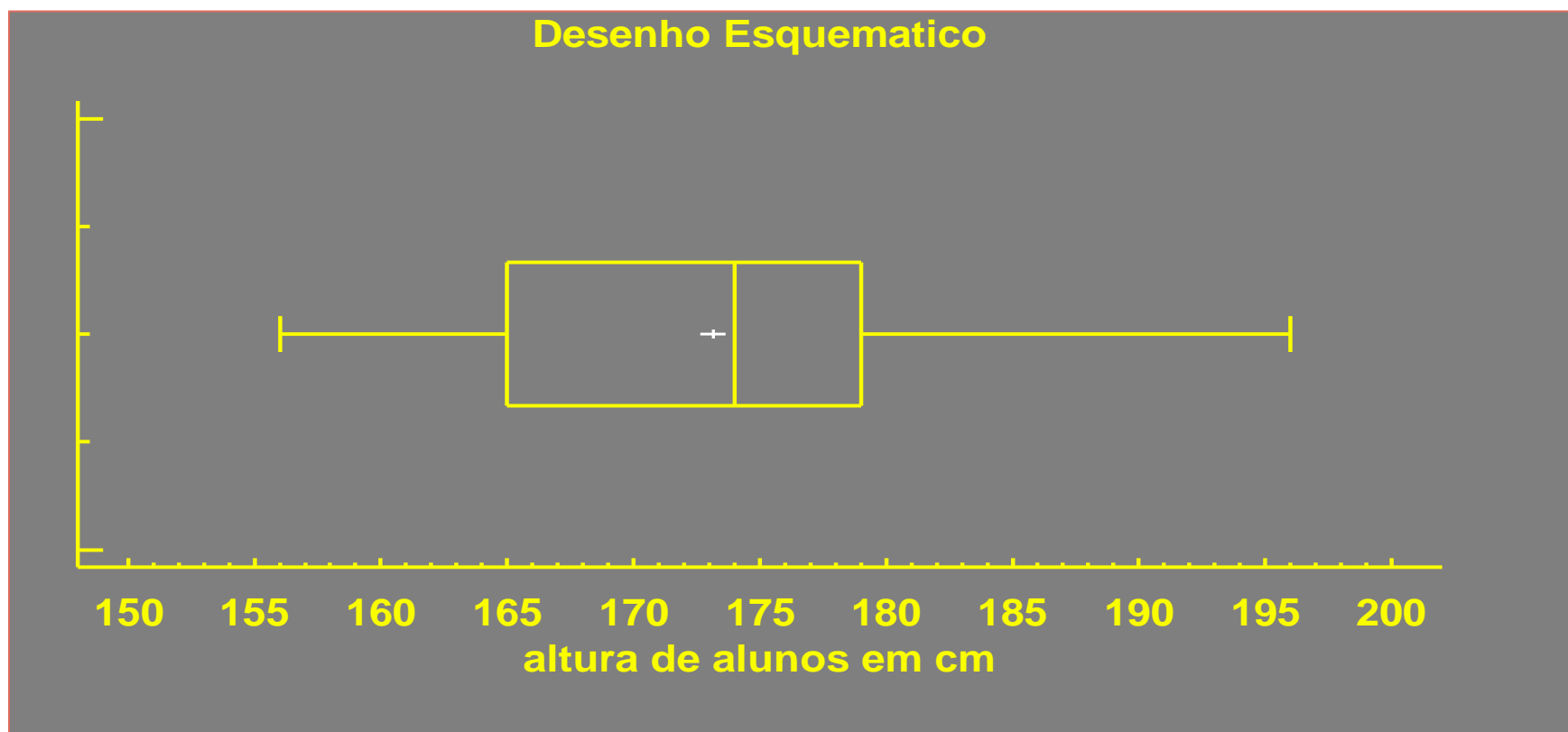
# Exemplo

---



# Exemplo

---



# Análise exploratória de dados (AED)

---

Análise exploratória de dados (AED), em estatística, é uma abordagem para analisar dados visando resumir as suas principais características, muitas vezes com métodos visuais. O objetivo da AED é observar o que os dados podem nos dizer para além da modelagem formal ou hipótese em testes.



## Analizando dados Estatísticos básicos no Power BI

► Exibir Audi

Podemos explorar os dados sob duas perspectivas: medidas de posição e medidas de dispersão. Medidas de posição tem o objetivo de descrever a tendência central e são: **média, mediana e moda**. Medidas de dispersão tem o objetivo de avaliar o quanto os dados são semelhantes, ou o quanto os dados distam do valor central e são: **desvio padrão e variância**.

Vamos ver medidas de posição e dispersão no Power BI.

Considerando o seguinte conjunto de dados sobre os Gastos Mensais em Reais de um estudante fazendo intercâmbio no EUA.

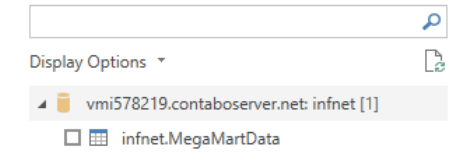
Meses	Gastos R\$
Jan	R\$ 68,40
Fev	R\$ 68,00
Mar	R\$ 67,80
Abr	R\$ 67,20
Mai	R\$ 68,00
Jun	R\$ 65,40
Jul	R\$ 64,40
Ago	
Set	
Out	
Nov	
Dez	

# Vamos acompanhar no moodle!



# Praticando!

## Navigator



Construir um dashboard utilizando boas práticas de visualização de dados, utilizando a tabela “infnet” com os dados “MegaMartData” armazenados no banco de dados MySQL (vmi578219.contaboserver.net).

- Faça ajustes ou modificações na base de dados necessários para atingir seu objetivo.
- Se atente ao tipo da informação e como ela pode ser melhor representada no Power BI.
- Use figuras e imagens para ilustrar seu dashboard fazendo uso de áreas de textos.

- Totalização do montante de compras por cidade
  - Crie também uma visão hierarquizada para podermos ver o montante de vendas por cada loja
- KPI do ticket médio do montante de compras
  - Por cidade
  - Por gênero
- Somatório e média de vendas de todos os departamentos da rede varejista para o último ano disponível
- Média de vendas do departamento de eletrônicos e vestuário para o último ano disponível, mostrando uma linha de tendencia das vendas para os próximos 2 meses.
- Ranking dos tops 10 clientes que mais gastaram nas lojas.
- Ranking das top 10 lojas em vendas e suas localizações, ordem decrescente.

# Na próxima aula...

---

Vamos continuar nosso estudo de visualização de dados no Power BI.

