

ESTATÍSTICA – Análise Exploratória

Prof. Eng. Rodolfo M. de Paiva

- **Um pouco da História da Matemática e da Estatística**

- A **Matemática** é uma ciência bem mais antiga do que a **Estatística** formal que conhecemos hoje.
- É fato que só houveram (na origem), descobertas e avanços matemáticos em regiões que tinham a necessidade da **agricultura**.
- Aos poucos, os profissionais das ciências exatas foram surgindo (corda, balança, astros...), e com eles novas necessidades de descobertas junto com o desenvolvimento das cidades.
- A **Matemática Primitiva** teve quatro alicerces: Clima, Agricultura, Comércio e Registro.



A Ciência Estatal

Esteve presente entre os hebreus, chineses, egípcios, maias, romanos, hindus, persas e babilônios, povos que se organizaram ao redor de um Estado, que necessitavam de informações censitárias, informações essas que eram colhidas pelo **estadista**.

A Estatística

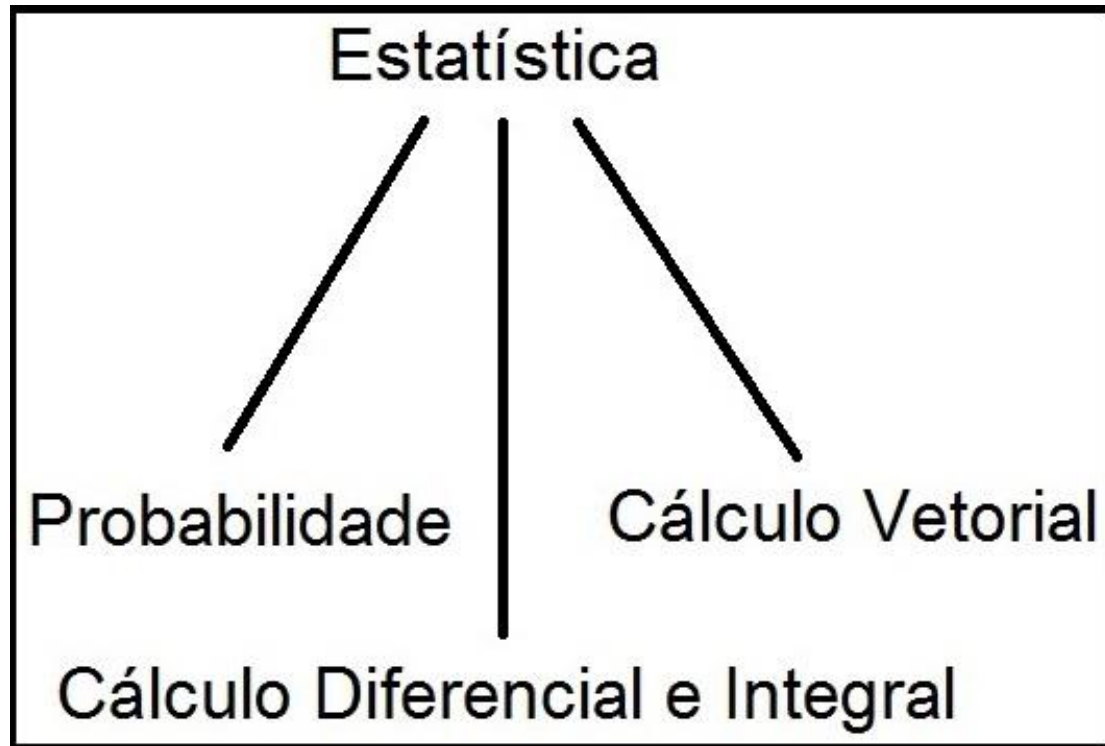
1858 – Florence Nightingale, com o trabalho:

“Notas sobre os Assuntos que Afetam a Eficiência da Saúde e Administração Hospitalar do Exército Britânico”

Nascida em 12 de Maio de 1820, Florence Nightingale era enfermeira britânica.

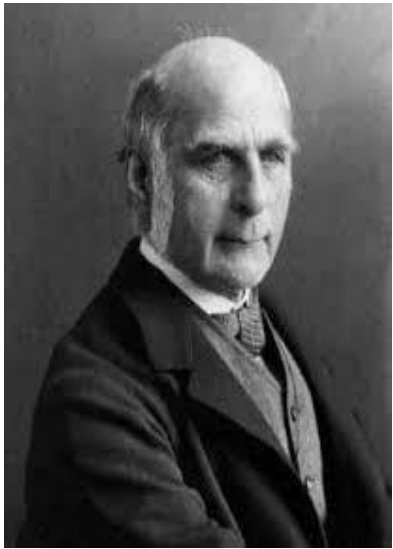


Com o passar do tempo a Estatística foi tomando forma, com os avanços de outros conteúdos matemáticos:



Tendo a atenção de estudiosos como: Christian Huygens, Pierre Fermat, Blaise Pascal, John Graunt, Jacques Bernoulli, Thomas Bayes, Poisson, Mary Somerville, entre outros...

No final do século XIX, com Francis Galton, Francis Ysidro Edgeworth, Karl Pearson e George Udny Yule, a Estatística ganhou sua aplicabilidade e operacionalidade, além do visual moderno.



Importância:

- Tomada de decisão;
- Quantificar incerteza;
- Levantamento de dados;
- Entendimento de um fenômeno;
- Etc...



Áreas de aplicação:

- Bioestatística;
- Ciência Atuarial;
- Demografia;
- Econometria;
- Epidemiologia;
- Geoestatística;
- Controle da Qualidade;
- Data Science;
- B.I.;
- Etc...





Embora a Estatística Clássica se fez por conta de conteúdos da Matemática, hoje em dia sabemos que são **ciências distintas**, com algumas coisas em interseção.



De volta para a Estatística...

- **Tipos de Dados**

Os dados (variáveis), são classificados em dois grupos:

Quantitativo X Qualitativo



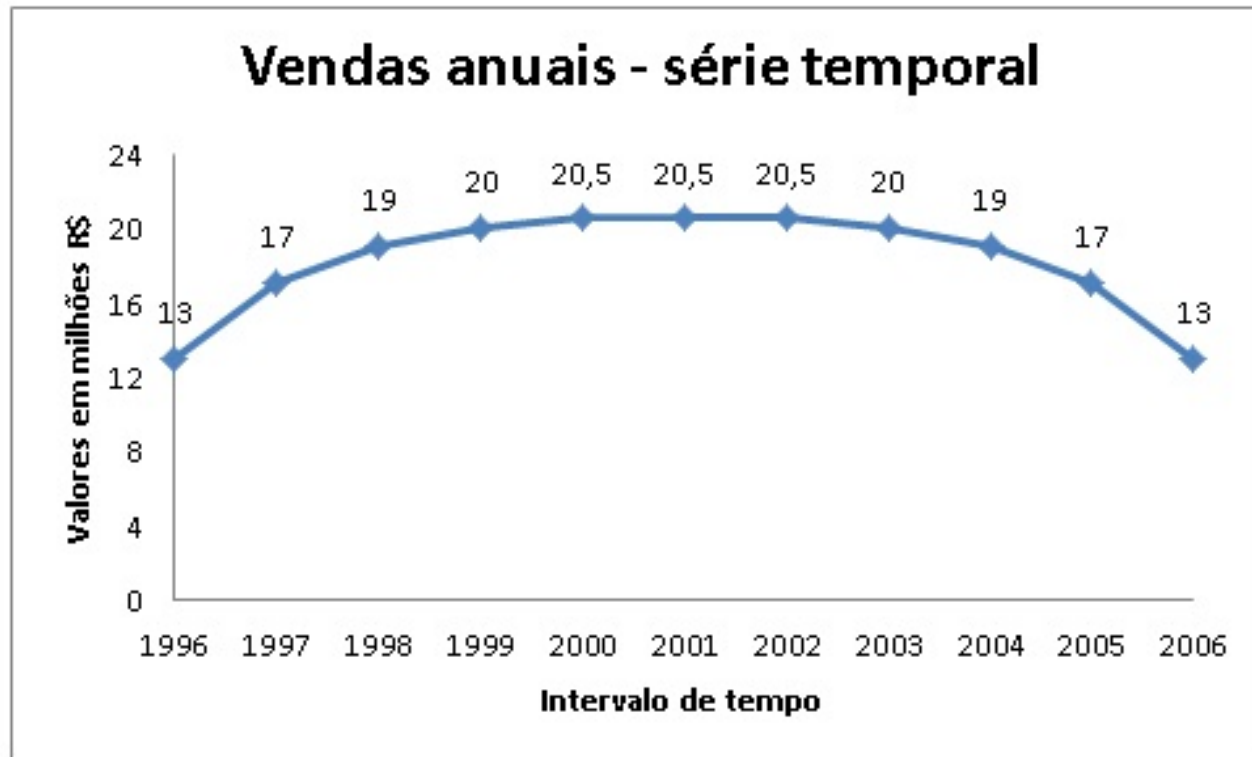
Quantitativo: Discreto = Números inteiros, contagem, é mensurável
Contínuo = Conjunto dos Reais, também mensurável



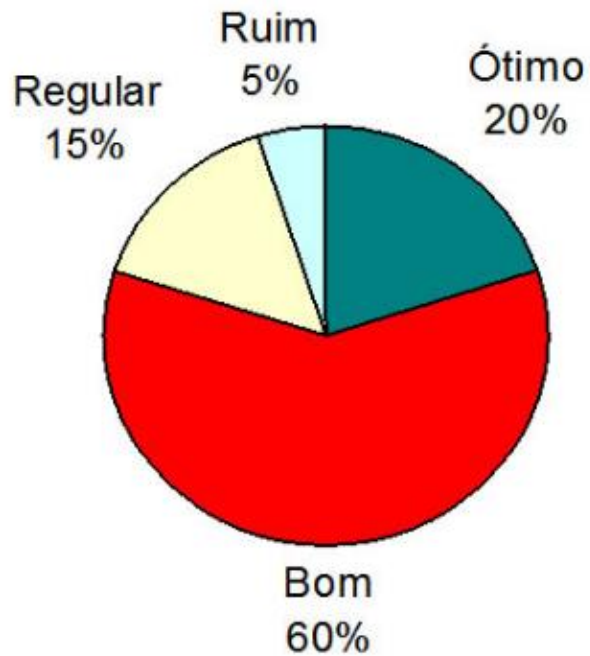
Qualitativo: Ordinal = Apresenta hierarquia
Nominal = Não apresenta hierarquia

- **Gráficos Estatísticos para Variáveis Qualitativas e Quantitativas**

Gráficos auxiliam de uma forma rápida e prática que entendamos o **comportamento de um fenômeno**.

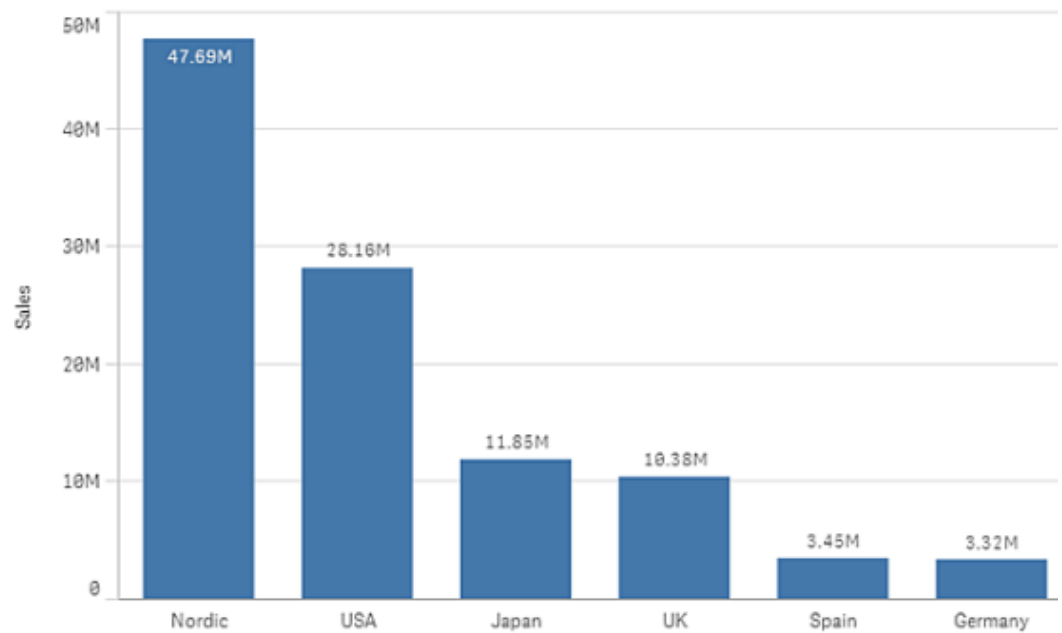


- Gráficos para Variáveis Qualitativas:



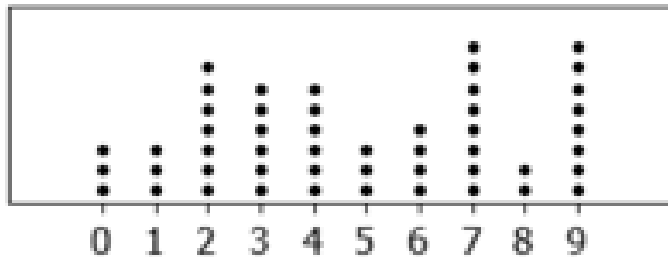
**Gráfico de Setores
ou Circular**

Gráfico de Barras

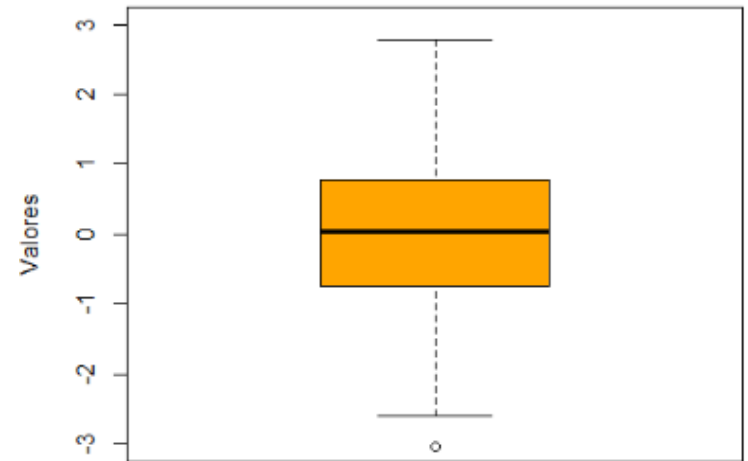


- Gráficos para Variáveis Quantitativas:

Dotplot of Random Values

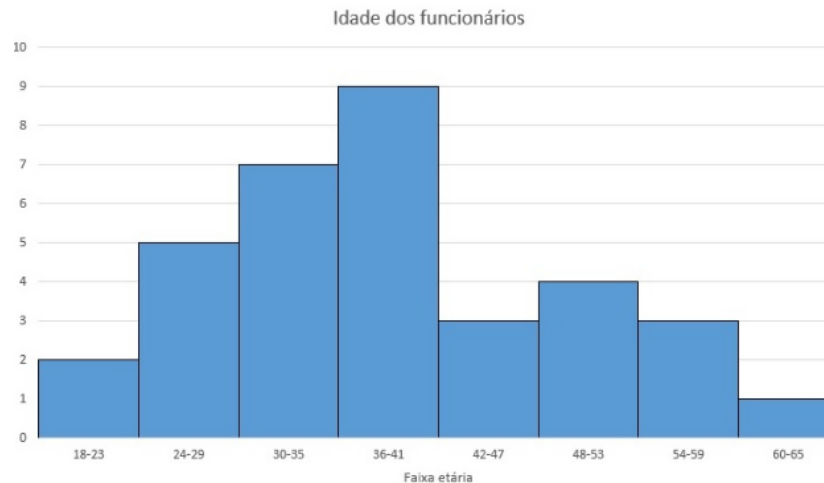


Dotplot ou Strip Chart



Variável

Boxplot

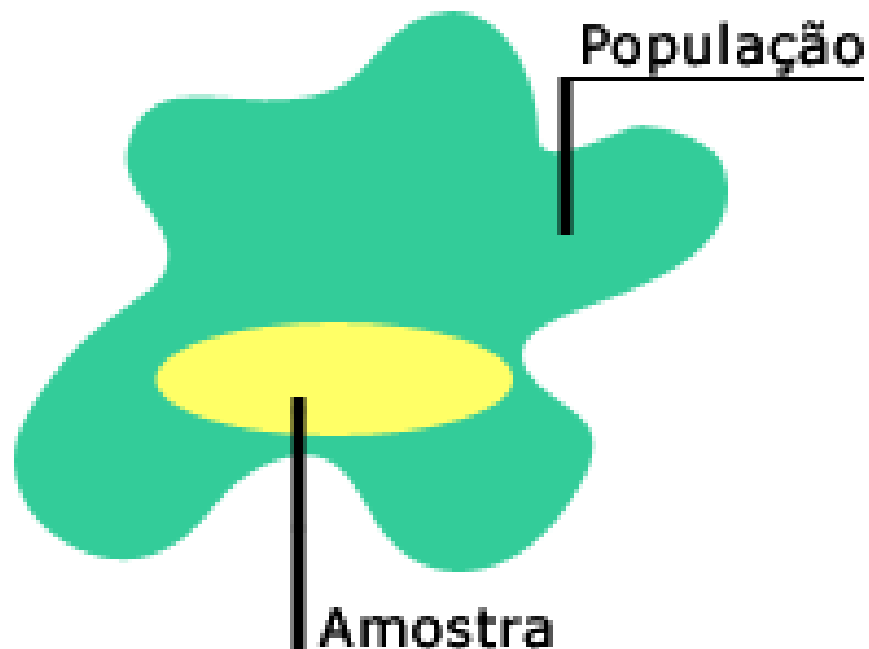


Histograma

- **Coleta e Organização de Dados**

A coleta de dados é uma parte extremamente importante, pois sem dados não existe a possibilidade de ser feito qualquer tipo de estudo estatístico.

Para obter dados, a Estatística utiliza amostragens, que podem ser **probabilísticas** e **não probabilísticas**.



Tipos de Amostragem

- **Probabilística:**

- Aleatória Simples;
- Sistemática;
- Estratificada;
- Conglomerados.

- **Não Probabilística:**

- Intencional;
- Cotas;
- Conveniência.

Uma vez os dados coletados, a forma mais prática de organizá-los é por meio de **tabelas**.

Aparelho	Potência (KW)	Tempo de uso diário (horas)
Ar condicionado	1,5	8
Chuveiro elétrico	3,3	1/3
Freezer	0,2	10
Geladeira	0,35	10
Lâmpadas	0,10	6

Nº de pessoas que doaram	Valor Doador
12	R\$ 5,00
10	R\$ 7,00
8	R\$ 10,00
7	R\$ 12,00
3	R\$ 15,00

Com os dados observados, é possível extrair vários tipos de informações!

- Frequência Relativa e Frequência Absoluta

Nº de veículos	Frequência Absoluta f_i	Frequência relativa f_{ri} (%)	Frequência acumulada F_{ac}	Frequência Acumulada Relativa F_{Ri} (%)
0	6	30%	6	30%
1	8	40%	14	70%
2	3	15%	17	85%
3	2	10%	19	95%
4	1	5%	20	100%
Total	20	100%	----	----

- **Medidas de Tendência Central**

- **Média (Esperança):**

$$\bar{X} = \mu = \frac{\sum x_i}{n}$$

- **Moda:**

Valor ou classe que mais se repete, podendo ser: Amodal, Unimodal, Bimodal, ... Representada por **Mo**

- **Mediana:**

Elemento central

$$M_e = \begin{cases} X_{(\frac{n+1}{2})}, & \text{se "n" é ímpar} \\ \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2}, & \text{se "n" é par} \end{cases}$$

- **Medidas de Dispersão**

- **Amplitude:**

$$R = X_{(n)} - X_{(1)}$$

- **Variância:**

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

$$\sigma^2 = \sum_{i=1}^N \frac{(x_i - \mu)^2}{N}$$

- **Desvio Padrão (Desvio Médio):**

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

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

- Coeficiente de Variação (CV):

$$CV = \frac{\sigma}{\mu} \text{ ou } \frac{s}{\bar{x}}$$

*Normalmente multiplicamos o resultado por **100** para expressá-lo em %.

- **Medidas Separatrizes**

Valores que dividem os dados em partes iguais:

- **Quartis:** Dividem os dados em 4 partes iguais
- **Decis:** Dividem os dados em 10 partes iguais
- **Centis:** Dividem os dados em 100 partes iguais

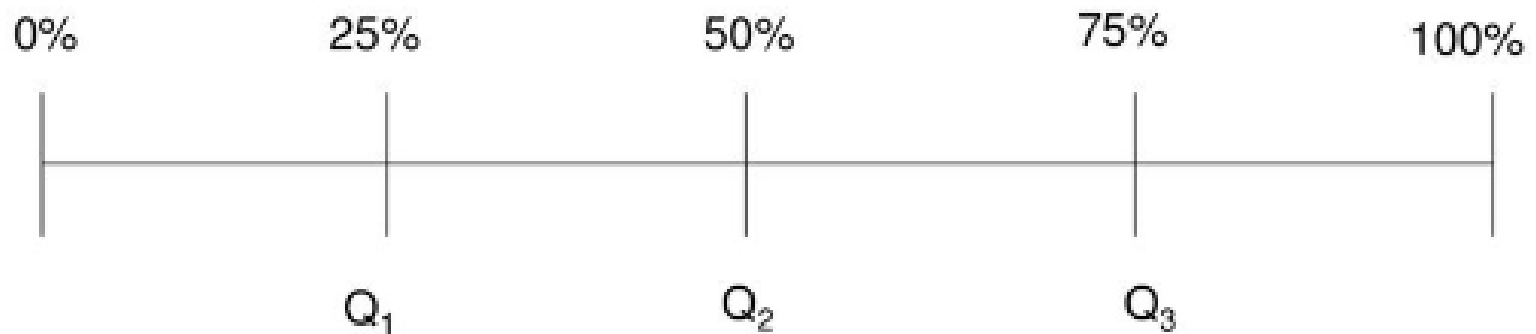
Normalmente os quartis são os mais usados.

- **Quartis:**

1° Quartil: $Q_1 = X_1 + \left(\frac{n+1}{4} - k \right) (X_{k+1} - X_k)$

2° Quartil: $Q_2 = X_3 + \left(\frac{2(n+1)}{4} - k \right) (X_{k+1} - X_k)$

3° Quartil: $Q_3 = X_5 + \left(\frac{3(n+1)}{4} - k \right) (X_{k+1} - X_k)$



Com essas ferramentas de análise exploratória de dados da **Estatística Descritiva** é possível entender o comportamento de diversos fenômenos do dia a dia, basta **ter** ou **iniciar** a coleta de dados e na sequência:



APLICAR!

