GRÁFICOS ESTATÍSTICOS

Definição:

O **gráfico estatístico** é uma forma de apresentação dos dados estatísticos, cujo objetivo é o de produzir, no investigador ou no público em geral, uma impressão mais rápida e viva do fenômeno em estudo, já que os gráficos falam mais rápido à compreensão que às séries.

A representação gráfica de um fenômeno deve obedecer a certos requisitos fundamentais para ser realmente útil:

- ✓ **Simplicidade** – o gráfico deve ser destituído de detalhes de importância secundária, assim como de traços desnecessários que possam levar o observador a uma análise morosa ou com erros.
- ✓ **Clareza** – o gráfico deve possibilitar uma correta interpretação dos valores representativos do fenômeno em estudo.
- ✓ **Veracidade** – o gráfico deve expressar a verdade sobre o fenômeno em estudo.

Os principais tipos de gráficos são os **diagramas**, **cartogramas, estereograma** e os **pictogramas**.

DIAGRAMAS

Utiliza elementos geométricos simples (pontos, linhas e retângulos**)** são gráficos geométricos de, no máximo, duas dimensões: para sua construção, em geral, fazemos uso do sistema cartesiano. Em *Pesquisa Científica* e *Trabalhos Acadêmicos*, o padrão é o uso de ***Diagramas***.

Dentre os principais diagramas, destacamos:

- ➢ **Gráfico em Linha ou em Curva.**

Este tipo de gráfico se utiliza da linha poligonal para representar a série estatística.

O gráfico em linha constitui uma aplicação do processo de representação das funções num sistema de coordenadas cartesianas.

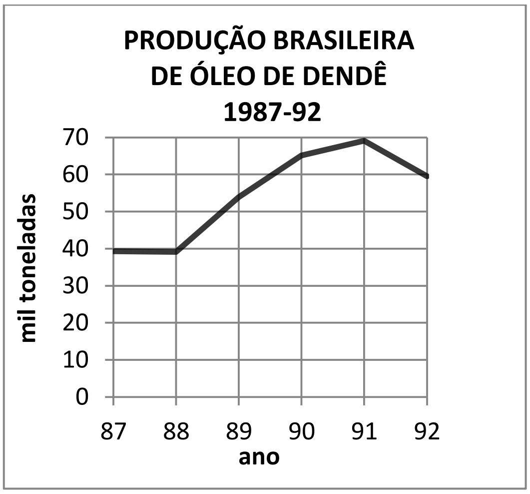
Exemplo: Para tornar bem clara a explanação, consideremos a seguinte série:

PRODUÇÃO BRASILEIRA DE ÓLEO DE DENDÊ 1987-92

ANOS	QUANTIDADE (1.000 t)
1987	39,3
1988	39,1
1989	53,9
1990	65,1
1991	69,1
1992	59,5

FONTE: Agropalma.

Vamos tomar os anos como abscissas e as quantidades como ordenadas. Assim, um ano dado (x) e a respectiva quantidade (y) formam um par ordenado (x, y), que pode ser representado num sistema cartesiano.



FONTE: Agropalma.

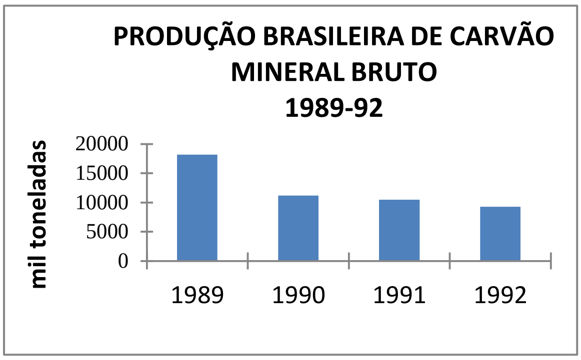
- **Gráfico em Colunas ou em Barras**

É a representação de uma série por meio de retângulos, dispostos **verticalmente** (em colunas) ou **horizontalmente** (em barras).
Exemplos: **Gráfico em colunas**

PRODUÇÃO BRASILEIRA DE CARVÃO MINERAL BRUTO 1987-92

ANOS	QUANTIDADE PRODUZIDA (1.000 t)
1989	18.196
1990	11.169
1991	10.468
1992	9.241

FONTE: Ministério da Agricultura.



FONTE: Ministério da Agricultura.

- **Gráfico em Colunas ou em Barras Múltiplas**

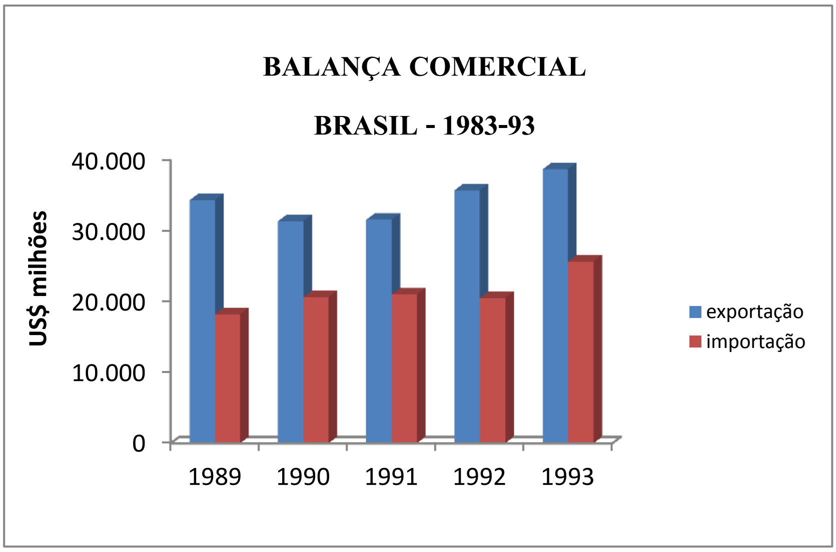
Este tipo de gráfico é geralmente empregado quando queremos representar, simultaneamente, dois ou mais fenômenos estudados com o propósito de comparação.

Exemplo:

BALANÇA COMERCIAL DO BRASIL
1989-93

ESPECIFICAÇÕES	VALOR (US$ 1.000.000)				
	1989	**1990**	**1991**	**1992**	**1993**
Exportação (FOB)	34.383	31.414	31.620	35.793	38.783
Importação	18.263	20.661	21.041	20.554	25.711

FONTE: Ministério da Fazenda.



FONTE: Ministério da Fazenda.

- **Gráfico em Setores**

Este gráfico é construído com base em um círculo, e é empregado sempre que desejamos ressaltar a participação do dado no total.
O total é representado pelo círculo, que fica dividido em tantos setores quantas são as partes.
Os setores são tais que suas áreas são respectivamente proporcionais aos dados da série.

Exemplo: Dada à série:

REBANHO SUÍNO DO SUDESTE DO BRASIL
1992

ESTADOS	QUANTIDADE (mil cabeças)
Minas Gerais	3.363,7
Espírito Santo	430,4
Rio de Janeiro	308,5
São Paulo	2.035,9
Total	6.138,5

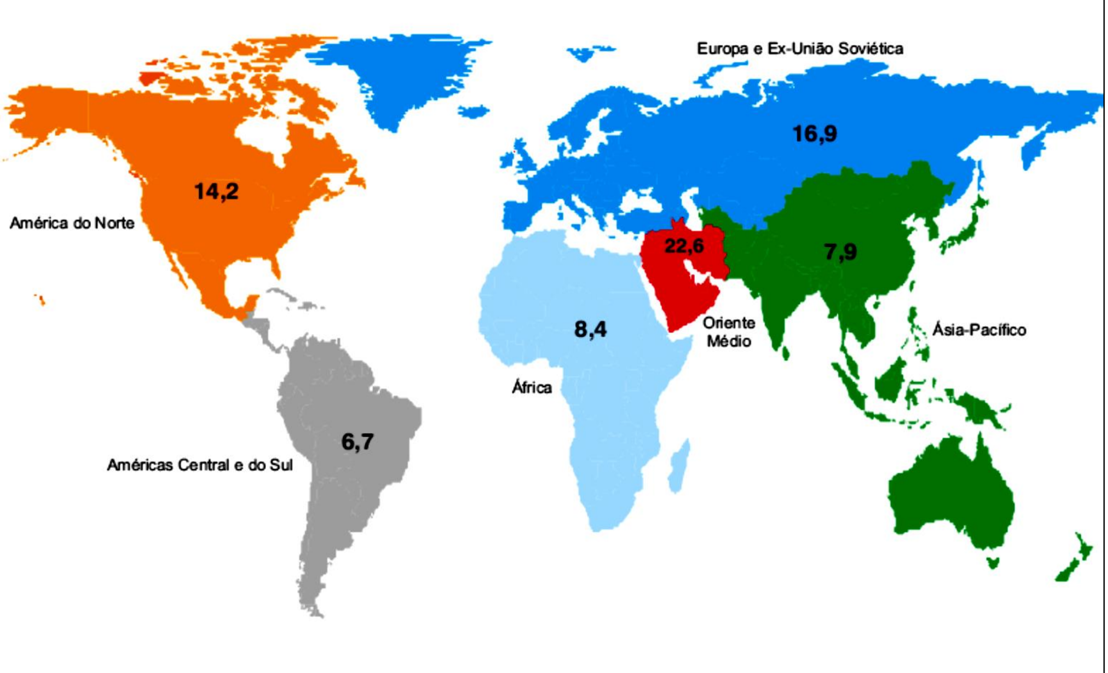
FONTE: IBGE.

FONTE: IBGE.

- **Cartograma**

O **cartograma** é a representação sobre uma carta geográfica (utiliza mapas).
Este gráfico é empregado quando o objetivo é o de figurar os dados estatísticos diretamente relacionados com áreas geográficas ou políticas.
Exemplos: Produção de Petróleo em 2003.

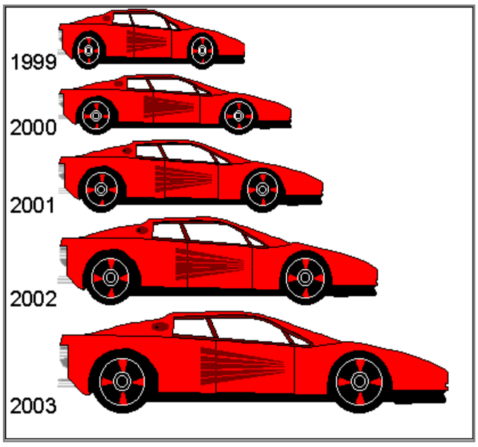
Cartograma 1.2 - Produção de petróleo, segundo regiões geográficas (milhões b/d) - 2003



Fontes: ANP/SDP; BP Amoco (Tabela 1.2).
Nota: Inclui óleo de xisto, óleo de areias betuminosas e LGN, exceto para o Brasil. Para o Brasil, inclui LGN e não inclui óleo de xisto e óleo de areias betuminosas.

- **Pictograma**

O **pictograma** constitui um dos processos gráficos que melhor fala ao público, pela sua forma ao mesmo tempo atraente e sugestiva. A representação gráfica consta de **figuras**.
Exemplo: Venda de carros no Brasil de 1999 a 2003.



Histograma: Um Histograma é um diagrama de barras de uma distribuição de frequência com uma diferença: não há espaços entre as barras. Os intervalos de classe são colocados no eixo horizontal enquanto as frequências são colocadas no eixo vertical.

São gráficos de superfícies utilizados para representar distribuições de frequências com dados agrupados em classes.

Polígonos de Frequência: O polígono de frequência é um gráfico de linha de uma distribuição de frequência. Os eixos de um Polígono de frequência são similares ao do Histograma, exceto que no eixo horizontal são colocados os pontos médios de cada intervalo de classe.

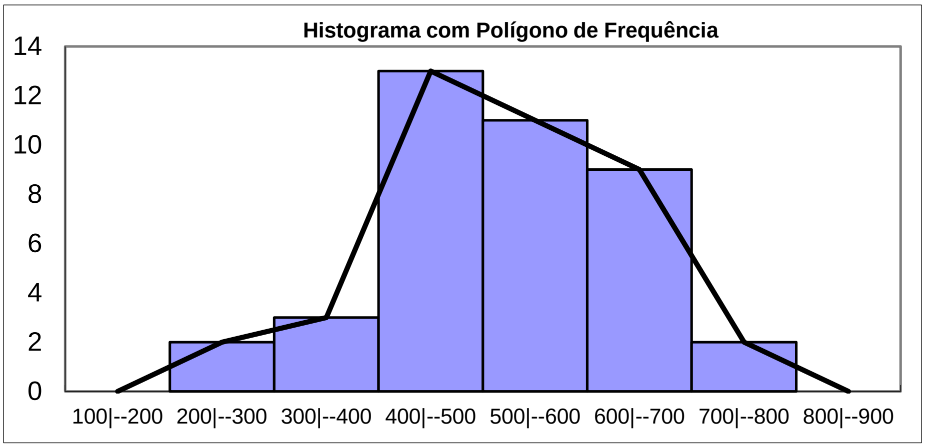


DIAGRAMA DE RAMO E FOLHA

As técnicas de análise exploratória dos dados consistem em aritmética simples e em gráficos fáceis de desenhar, que podem ser utilizados para sintetizar dados rapidamente. Uma técnica chamada de apresentação de **Ramo-e-folhas**, pode ser empregada para mostrar, simultaneamente, a ordem de classificação e o formato do conjunto de dados. Logo temos que:

_ Trata-se de uma ferramenta exploratória útil para descrever **pequenos conjuntos de dados**.
_ O método fornece uma boa **visão geral dos dados** sem que haja perda de informação.
_ É um procedimento alternativo para resumir um conjunto de valores, com o objetivo de se obter uma ideia da forma de sua distribuição.
_ A vantagem deste diagrama sobre o histograma é que não perdemos ou quase nada perdemos dos dados.

Exemplo:
Vamos imaginar uma loja com produtos nos seguintes preços (R$):

4,00	4,56	5,25	5,73	6,26	6,66	6,86	7,39
7,44	7,59	8,12	8,46	8,74	8,95	9,13	9,35
9,77	9,80	10,53	10,76	11,06	11,59	12,00	12,79
13,23	13,60	13,85	14,69	14,71	15,99	16,22	16,61
17,26	18,75	19,40	23,30				

Passos para obtenção do diagrama

1º Passo: organização e separação dos dados, combinando todos os valores que começam com 4, todos que começam com 5, todos que começam com 6 até o 23.

4 |
5 |
6 |
. |
. |
. |
21|
22|
23|

2º Passo: Vamos agora inserir a parte decimal de cada valor, para o 4 temos o 00 e assim por diante, fazendo para todos os valores mesmo que não apareçam na tabela original, temos que 21 e 22 aparecem mas não na nossa tabela original.

4 | 00 56
5 | 25 73
6 | 26 66 86
. |
. |
. |
21| não aparece
22| não aparece
23| 30

3º Passo: É bastante interessante que as folhas do diagrama sejam ordenadas facilitando ainda mais a interpretação.

4 | 00 56
5 | 25 73
6 | 26 66 86
7 | 39 44 59
8 | 12 46 74 95
9 | 13 35 77 80
10| 53 76
11| 06 59
12| 00 79
13| 23 60 85
14| 69 71
15| 99
16| 22 61
17| 26
18| 75
19| 40
20|
21|
22|
23| 30

Cada linha é denominada **ramo** e cada número à direita da linha vertical é denominado **folha.**
Para o caso de muitas folhas em um ramo, podemos duplicá-lo, ou seja, fazer um de 6,0 a 6,4 como (6*) e o outro de 6,5 a 6,9 de (6º).
6* |
6º |

Podemos tirar algumas conclusões a partir do diagrama:
- ✓ Destaque para o valor de R$ 23,30, veja que ele esta bem fora, sendo um valor isolado;
- ✓ Valores concentrados entre R$ 4,00 e R$ 19,40;
- ✓ Maior concentração entre R$ 8,00 e R$ 9,00;
- ✓ Leve simetria em direção aos valores mais elevados;
- ✓ Esta distribuição pode ser aproximada para uma normal, curva simétrica em forma de sino.

Exercício proposto: A seguir temos os tempos de resposta (em picosegundos 10^{-12}) de 30 circuitos integrados:
3,7 - 4,1 - 4,5 - 4,6 - 4,4 - 4,8 - 4,3 - 4,4 - 5,1 - 3,9
3,3 - 3,4 - 3,7 - 4,1 - 4,7 - 4,6 - 4,2 - 3,7 - 4,6 - 3,4
4,6 - 3,7 - 4,1 - 4,5 - 6,0 - 4,0 - 4,1 - 5,6 - 6,0 - 3,4
Construa um diagrama de ramos e folhas usando os dígitos de unidades como ramos e os de décimos como folhas.

Resolução:
3|3 4 4 4 7 7 7 7 9
4|0 1 1 1 1 2 3 4 4 5 5 6 6 6 6 7 8
5|1 6
6|0 0

DETERMINAÇÃO DO NÚMERO DE CLASSES - INTERVALOS DE CLASSE

A primeira preocupação que temos, na construção de uma distribuição de frequência, é a determinação do número de classes e, consequentemente, da amplitude e dos limites dos intervalos de classe.

Quando dispomos de uma tabela primitiva ou de um rol, precisamos estabelecer a quantidade e o intervalo das classes que vamos criar, de outro modo à distribuição de frequência pode não ser útil para a nossa análise.

Uma das maneiras de determinar o número de classes é usando a **Regra de Sturges** que determina *k* em função de *n*:

$$\boldsymbol{k \cong 1 + 3{,}3.\log (n)}$$

onde *k* é o número de classes e *n* o número de dados. Da mesma forma podemos usar outra regra que associa *k* e *n* de outra forma: $\boldsymbol{k \cong \sqrt{n}}$

Sabendo o número de classes (*k*) que vamos usar, podemos determinar o intervalo de classes através da amplitude total da distribuição (A_T)

Amplitude de classe

$$h \cong \frac{A_T}{k}$$

Nas equações acima foi usado o símbolo de aproximadamente (≅) ao invés de igualdade (=) porque estas fórmulas representam valores típicos a serem usados, mas que podem ser alterados ligeiramente de acordo com o objetivo da distribuição ou para evitar classes com frequências nulas enquanto outras tem valores muito altos.

Entretanto, a verdade é que essas fórmulas não nos levam a uma decisão final; esta vai depender, na realidade, de um julgamento pessoal, que deve estar ligado à natureza dos dados, e ainda, do objetivo que se tem em vista.

Abaixo temos uma tabela com o número de classes.

Regra de Sturges (Logaritmo)	
Quantidade de dados (n)	Quantidade de Classes (k)
1	1
2	2
3 a 5	3
6 a 11	4
12 a 23	5
24 a 46	6
47 a 93	7
94 a 187	8
188 a 376	9
377 a 756	10

Etapas para a construção de tabelas de frequência (dados contínuos ou discretos com valores muito distintos)

Definição das Classes

a) Determinar a amplitude da amostra (máximo - mínimo)
b) Dividir esta amplitude pelo número de classes, **k**.
c) Tomar para amplitude de classe, **h**, um valor aproximado por excesso do valor obtido em (b).
d) Construir as classes de modo que tenham toda a mesma amplitude e cuja união contenha todos os elementos da amostra.

Exemplo: Consideremos as idades dos alunos do Colégio Extensão:

18	18	19	19	20	21	21	21	22	22	22	22	22
22	23	23	23	23	23	23	23	24	24	24	24	24
24	24	24	24	25	25	25	25	26	26	26	26	26
26	26	26	26	26	26	27	27	27	27	27	27	27
27	27	27	27	27	27	27	27	27	27	29	29	29
29	29	29	29	29	30	30	30	30	30	30	30	30
30	30	31	31	31	31	31	31	31	32	32	32	32
32	33	33	33	34	34	34	34	34	35	36	36	36
37	37	37	37	37	38	38	38	38	38	39	39	39
40	40	40	40	40	40	41	41	41	42	42	42	42
43	43	43	44	44	44	45	45	45	46	46	47	47
47	47	48	48	48	48	48	48	49	49	50	50	50
51	51	52	52	53	53	53	53	56	61	62	63	63

3ª Lista de Exercícios

1- Conhecida as notas de 50 alunos:

84	68	33	52	47
73	68	61	73	77
74	71	81	91	65
55	57	35	85	88
59	80	41	50	53
65	76	85	73	60
67	41	78	56	94
35	45	55	64	74
65	94	66	48	39
69	89	98	42	54

Obtenha a distribuição de frequência, tendo 30 para limite inferior da primeira classe e 10 para intervalo de classe. Calcule a média ($\bar{x}$), moda (m_o) e a mediana (md).

2- A tabela abaixo apresenta uma distribuição de frequência das áreas de 400 lotes:

Áreas (m^2)	300|-- 400	|-- 500	|--600	|--700	|--800	|--900	|--1.000	|-- 1.100	|--1.200
Nº de Lotes	14	46	58	76	68	62	48	22	6

Com referência a essa tabela, determine:

a) a amplitude total;
b) o limite superior da quinta classe;
c) o limite inferior da oitava classe;
d) o ponto médio da sétima classe;
e) a amplitude do intervalo da segunda classe;
f) a frequência da quarta classe;
g) a frequência relativa da sexta classe;
h) a frequência acumulada da quinta classe;
i) o número de lotes cuja área não atinge 700 m^2;
j) o número de lotes cuja área atinge e ultrapassa 800 m^2;
k) a percentagem de lotes cuja área não atinge 600 m^2;
l) até que classe estão incluídos 60% dos lotes.

3- Os dados a seguir representam a duração da vida útil, em anos, medidos do décimo mais próximo, de 30 bombas de combustível:

2,0	3,0	0,3	3,3	1,3	0,4	0,2	6,0	5,5	6,5
0,2	2,3	1,5	4,0	5,9	1,8	4,7	0,7	4,5	0,3
1,5	0,5	2,5	5,0	1,0	6,0	5,6	6,0	1,2	0,2

a) Construa um diagrama de ramo-e-folhas para a vida, em anos, das bombas de combustível, usando o dígito à esquerda da vírgula decimal como ramo para cada observação;
b) Construa uma tabela com intervalo de classe contendo sete classes e estabeleça a distribuição de frequências relativas;
c) Calcule a média, moda e mediana.

4- Foi solicitado a uma classe de 40 alunos que escolhessem um número do conjunto de algarismos (0,1,2,3,4,5,6,7,8 e 9). Obteve-se o seguinte resultado:

8	0	2	3	3	5	7	7	7	9
8	4	1	9	6	6	6	8	3	3
7	7	6	0	1	3	3	3	7	7
6	5	5	1	2	5	2	5	3	2

a) Montar a distribuição de frequência;
b) Determinar a média;
c) Qual foi o número mais escolhido? O que ele representa?
d)Calcule a mediana.

5- Um pesquisador de rádio ZT aborda 30 transeuntes ao acaso e pergunta-lhes a idade. O resultado é dado pela tabela:

35	26	39	25	39	22
42	40	39	22	21	40
16	32	39	21	28	39
18	37	23	14	27	44
30	32	21	15	26	43

a) Construa uma tabela com intervalo de classe utilizando a fórmula de Sturges e estabeleça a distribuição de frequências absolutas simples;
b) A média aritmética; a mediana; a moda.

Mediana para tabela sem intervalo de classe

OBS: No caso de existir uma frequência acumulada (F_1), tal que: $F_1 = (\sum f_1) \div 2$

A mediana será dada por: $Md = (x_i + x_{(i+1)}) \div 2$. Isto é, a mediana será a média aritmética entre o valor da variável correspondente a essa frequência acumulada e o seguinte.

Exemplo:

x_i	f_i	Fa
12	1	1
14	2	3
15	1	**4**
16	2	6
17	1	7
20	1	8
	$\sum = 8$	

Temos: $8 \div 2 = 4 = F_3$.

Logo: $Md = (15 + 16) \div 2 = 31 \div 2 = 15,5$