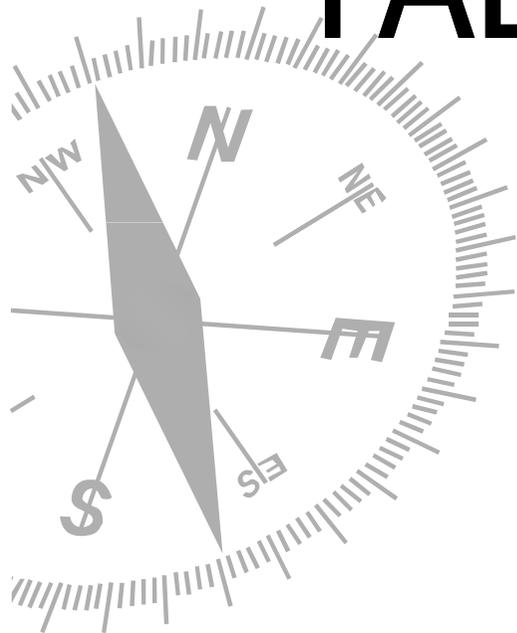


UNIVERSIDADE FEDERAL
DA PARAÍBA

TABELAS E GRÁFICOS

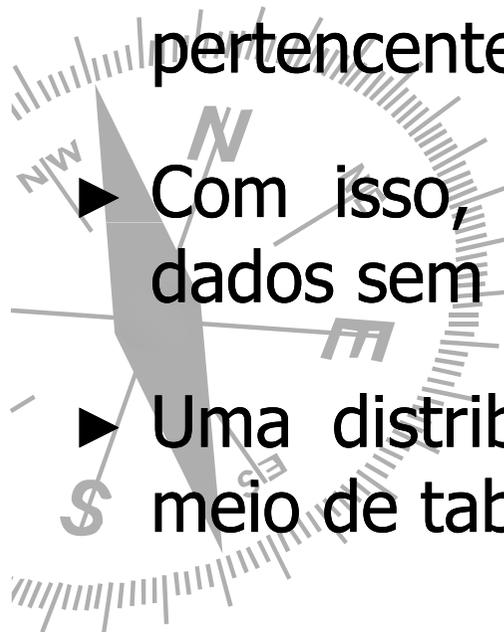


Departamento de Estatística

Luiz Medeiros

DISTRIBUIÇÃO DE FREQUÊNCIA

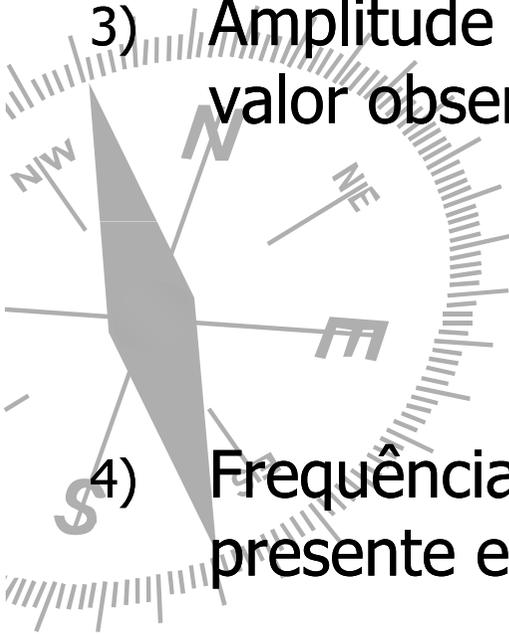
- ▶ Quando se estuda uma massa de dados é de frequente interesse resumir as informações de variáveis.
- ▶ Costuma-se, frequentemente, para uma melhor compreensão dos mesmos, distribuí-los em classes ou intervalos determinando-se o número de indivíduos pertencentes a cada classe ou intervalo.
- ▶ Com isso, podemos resumir e visualizar um conjunto de dados sem precisar levar em conta os valores individuais.
- ▶ Uma distribuição de frequência pode ser apresentada por meio de tabelas ou gráficos



Definições

- 1) Dados Brutos: Conjunto de dados que ainda não foram numericamente organizados, obtidos após a crítica dos valores.
- 2) Rol: É um arranjo dos dados brutos em ordem crescente.

3) Amplitude Total (AT): É a diferença entre o maior e o menor valor observado.


$$AT = \text{Maior valor} - \text{Menor valor}$$

4) Frequência (f_i): É o número de observações que se encontra presente em uma classe ou intervalo específico.

- 5) Frequência percentual ($f_i\%$): Representa o percentual de um certo valor na amostra.

$$f_i\% = (f_i / n) \times 100$$

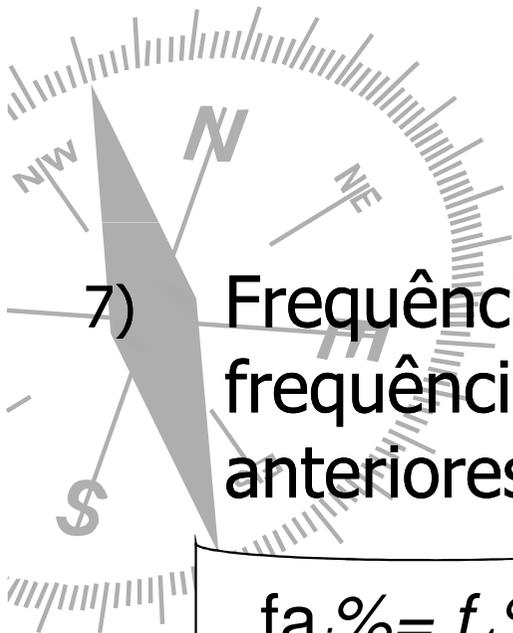
- 6) Frequência acumulada (fa_i): É a soma das frequências simples das classes ou dos valores anteriores.

$$fa_i = f_1 + f_2 + \dots + f_i$$

- 7) Frequência percentual acumulada ($fa_i\%$): É a soma das frequências relativas percentual das classes ou dos valores anteriores.

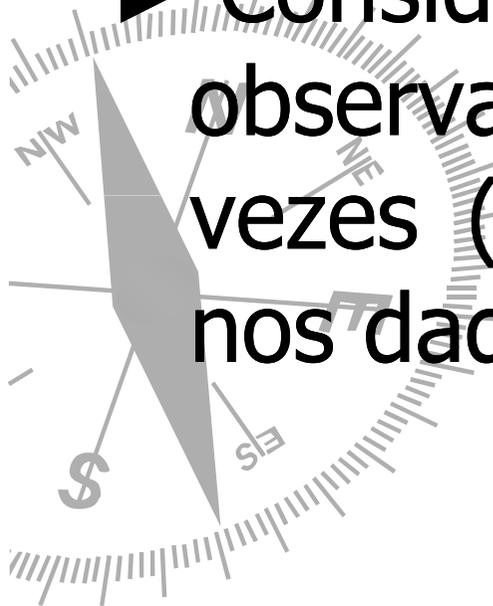
$$fa_i\% = f_1\% + f_2\% + \dots + f_i\%$$

$$fa_i\% = (fa_i / n) \times 100$$



DISTRIBUIÇÃO DE FREQUÊNCIA POR VALOR

- ▶ Utilizada para dados de uma variável qualitativa ou discreta.
- ▶ Considera-se os diferentes valores das observações ou categorias e o número de vezes (frequência) que cada valor aparece nos dados.

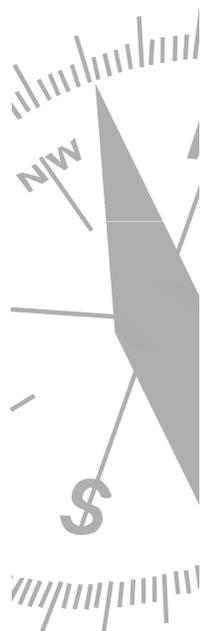


Exemplo:

Tabela 2. Tabela de frequência do Índice Antropométrico - CDC.

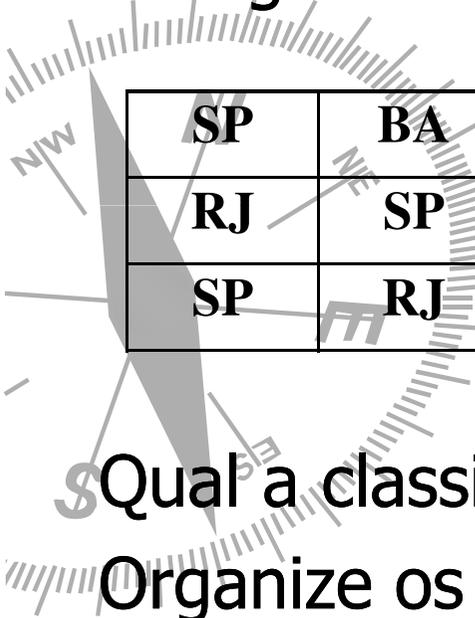
Índice Antropométrico - CDC

	Frequência	Percentual
magreza	57	10,3
normal	420	76,2
risco de sobrepeso	55	10,0
sobrepeso	19	3,4
Total	551	100,0



Exemplo

- ▶ De acordo com a Secretária de Turismo da Paraíba, os quatro estados com maior participação no número de turistas que desembarcaram no aeroporto de João Pessoa em 2009 foram: São Paulo (SP), Rio de Janeiro (RJ), Bahia (BA) e Paraná (PR). Os dados de uma amostra de 30 turistas abordados no aeroporto são apresentados a seguir:



SP	BA	BA	SP	RJ	BA	SP	SP	PR	PR
RJ	SP	SP	RJ	PR	RJ	SP	RJ	SP	BA
SP	RJ	PR	SP	BA	SP	PR	BA	SP	SP

Qual a classificação dessa variável?

Organize os dados em uma tabela com f_i , $f_i\%$.

DISTRIBUIÇÃO DE FREQUÊNCIA POR INTERVALO OU CLASSES

- ▶ Os dados são agrupados em classes
- ▶ Considera-se a frequência em cada classe (nº de observações pertencentes)



Classes	% da população	Renda Média Mensal Familiar (R\$ / pessoa)	
		de	até
Classe A 1	1%	9.733,47	acima
Classe A 2	4%	6.563,73	9.733,47
Classe B 1	9%	3.479,36	6.563,73
Classe B 2	15%	2.012,67	3.479,36
Classe C 1	21%	1.194,53	2.012,67
Classe C 2	22%	726,26	1.194,53
Classe D	25%	484,97	726,26
Classe E	3%	276,7	484,97
Classe Média (renda aprox.)		1.500,00	2.600,00

Para a construção de uma distribuição de frequência por intervalo é necessário algumas definições adicionais.

8) Número de classes (k): Não existe uma fórmula exata para o cálculo do número de classes. Seja n o tamanho da amostra, temos duas alternativas

1- Para $n < 25$, $K=5$ e para $n \geq 25$, temos que $K \approx \text{raiz}(n)$

2- Fórmula de Sturges:

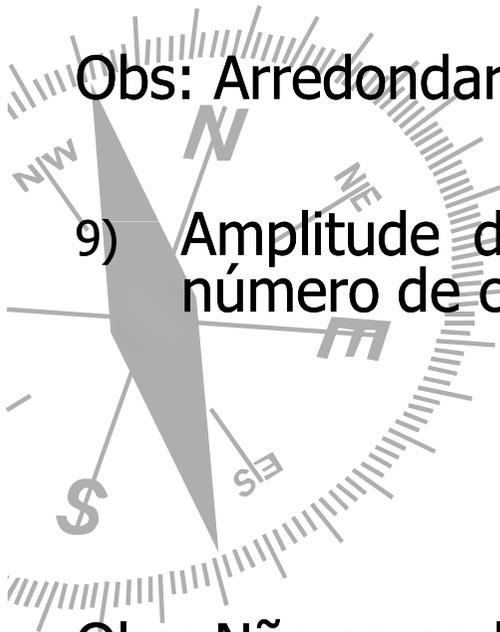
$$k = 1 + 3,3 \times \log_{10}(n)$$

Obs: Arredondar o número de classes para o inteiro mais próximo.

9) Amplitude das classes (h): É a divisão entre a amplitude total e o número de classes.

$$h = AT / k$$

Obs: Não se pode arredondar a amplitude das classes (h) para menos.



- 10) Limites das Classes: São os números extremos das classes. Utilizaremos a seguinte representação:

LI_i |— LS_i (Compreende todos os valores de LI_i a LS_i , excluindo LS_i)

1ª Classe:

Limite Inferior: $LI_1 = X_{\min}$

Limite Superior: $LS_1 = LI_1 + h$

2ª Classe:

Limite Inferior: $LI_2 = LS_1$

Limite Superior: $LS_2 = LI_2 + h$

⋮

i-ésima Classe:

Limite Inferior: $LI_i = LS_{i-1}$

Limite Superior: $LS_i = LI_i + h$

Continue estes cálculos até que seja obtido um intervalo que contenha o maior valor da amostra (X_{\max}) entre seus limites.

- 11) Ponto Médio (PM_i): É a média aritmética entre o limite superior, denotado por LS_i , e o limite inferior, denotado por LI_i . O ponto médio será denotado por PM_i , onde $PM_i = (LI_i + LS_i)/2$

Exemplo

Tabela 03 – Distribuição de Frequências das ALTURAS dos 46 Estudantes de CPE, Período: 97.1.

ALTURA (X_i)	Nº Estudantes (F_i)	Percentual ($f_i\%$)
153,0 ----- 157,3	4	8,7
157,3 ----- 161,6	8	17,4
161,6 ----- 165,9	7	15,2
165,9 ----- 170,2	10	21,7
170,2 ----- 174,5	3	6,5
174,5 ----- 178,8	6	13,0
178,8 ----- 183,1	8	17,4
Total ou Σ	46	100,0

FONTE: Quadro 1



Exemplo : Os dados abaixo representam o Preço (R\$) do produto A vendido em 25 diferentes estabelecimentos. A partir da distribuição de frequência que iremos construir podemos observar diversas características deste conjunto de dados como, por exemplo, identificar qual a faixa de preços com maior frequência.

Dados Brutos: 20,5 - 19,5 - 15,6 - 24,1 - 9,9 - 15,4 - 12,7 - 5,4 - 17,0 - 28,6 - 16,9 - 7,8 - 23,3 - 11,8 - 18,4 - 13,4 - 14,3 - 19,2 - 9,2 - 16,8 - 8,8 - 22,1 - 20,8 - 12,6 - 15,9

Rol: 5,4 - 7,8 - 8,8 - 9,2 - 9,4 - 9,9 - 11,8 - 12,6 - 12,7 - 13,4 - 14,3 - 15,4 - 15,6 - 15,9 - 16,8 - 16,9 - 17,0 - 18,4 - 19,2 - 19,5 - 20,5 - 20,8 - 22,1 - 24,1 - 28,6

Passos:

1. Ordenar as informações
2. Variável aleatória Preço é de natureza contínua implicando numa distribuição intervalar (Por classe).
3. Calcular a amplitude total.
4. Calcular o número de intervalos.
5. Calcular a amplitude do intervalo.

Exemplo

- ▶ Os dados abaixo representam o valor da hora de trabalho de 30 profissionais de Administração na cidade de João Pessoa.

20,5	9,5	15,3	17,2	24,1	19,9
15,4	12,7	7,4	16,5	15,3	26,2
14,9	7,8	23,3	15,9	11,8	18,4
13,4	14,3	16,2	16,7	9,2	16,8
9,8	20,1	17,8	17,1	12,6	15,9

Qual a classificação dessa variável?

Organize os dados em uma tabela completa.

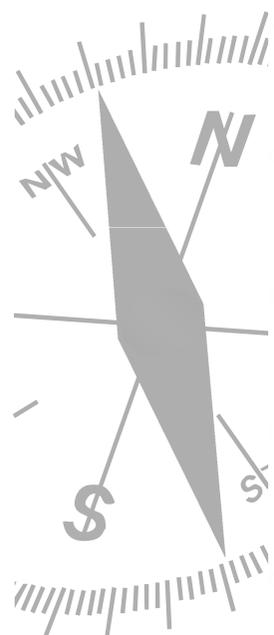


TABELA CRUZADA

Tabela 2

Tabulação Cruzada de Utilidade com os Valores Percentuais entre os Padrões

		Utilidade Percebida				Total	
		Nada Útil	Pouco Útil	Útil	Altamente útil		
Utilidade Social Percebida	Sem Utilidade Social	Frequência	6	9	5	2	22
		% em Utilidade Social Percebida	27,3	40,9	22,7	9,1	100,0
		% em Utilidade Percebida	75,0	42,9	18,5	3,4	19,1
	Pouca Utilidade Social	Frequência	2	5	7	5	19
		% em Utilidade Social Percebida	10,5	26,3	36,8	26,3	100,0
		% em Utilidade Percebida	25,0	23,8	25,9	8,5	16,5
	Certa Utilidade Social	Frequência	0	3	9	24	36
		% em Utilidade Social Percebida	0,0	8,3	25,0	66,7	100,0
		% em Utilidade Percebida	0,0	14,3	33,3	40,7	31,3
	Muita Utilidade Social	Frequência	0	4	6	28	38
		% em Utilidade Social Percebida	0,0	10,5	15,8	73,7	100,0
		% em Utilidade Percebida	0,0	19,0	22,2	47,5	33,0
Total	Frequência	8	21	27	59	115	
	% em Utilidade Social Percebida	7,0	18,3	23,5	51,3	100,0	
	% em Utilidade Percebida	100,0	100,0	100,0	100,0	100,0	



Gráficos

- ▶ A representação gráfica dos dados estatísticos tem por objetivo apresentar de forma rápida e concisa os resultados obtidos, permitindo-se chegar a conclusões sobre a evolução do fenômeno ou sobre como se relacionam os valores da série.



1) Gráfico em Colunas ou Barras: Os dados são representados através de retângulos dispostos verticalmente (colunas) ou horizontalmente (barras).

Evolução das Exportações Brasileiras
(1.000 U\$)

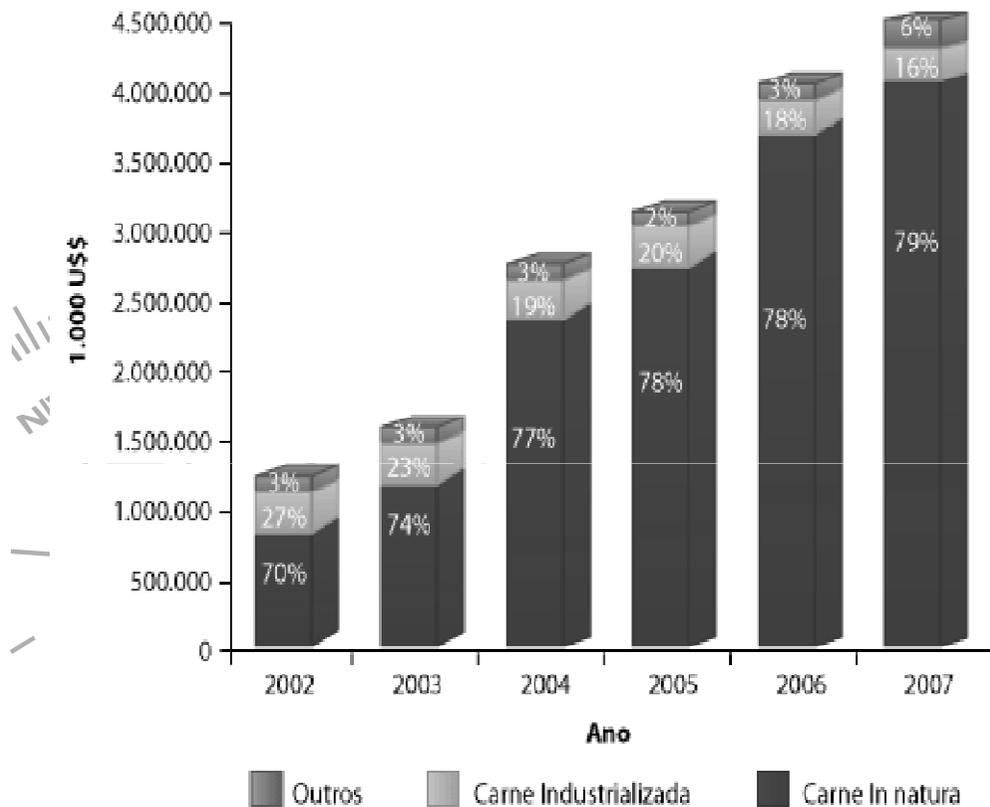
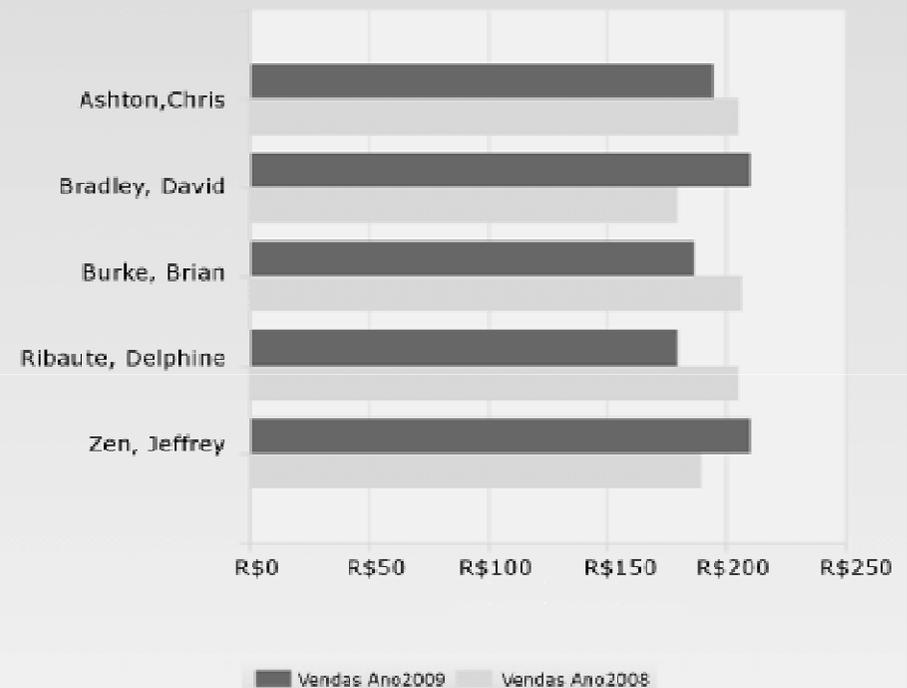


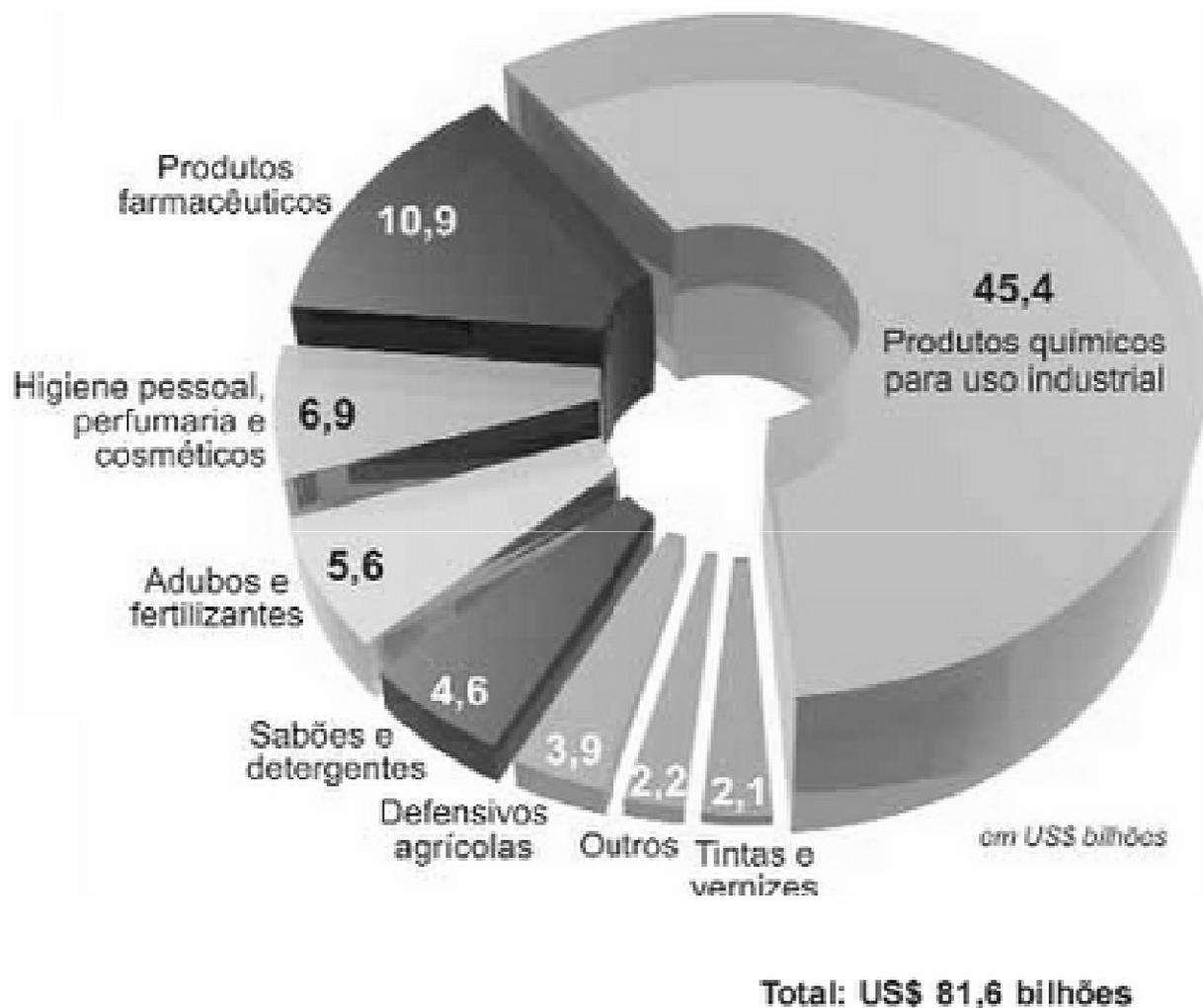
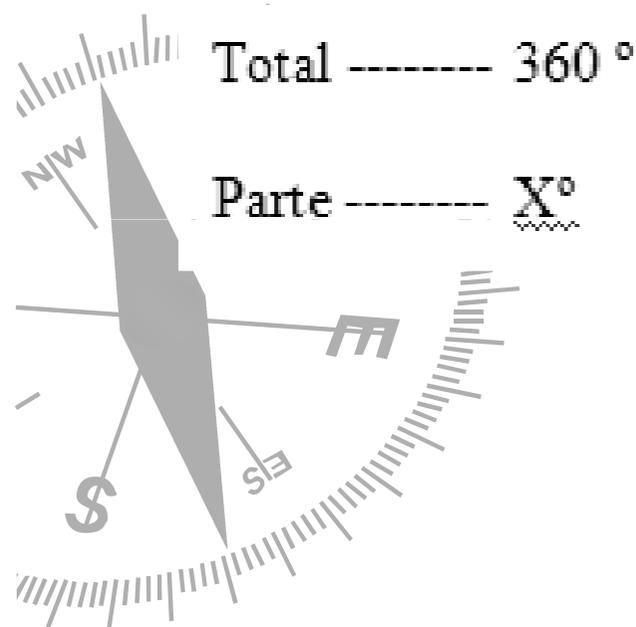
Gráfico de barras de vendas

Os cinco melhores vendedores de 2009

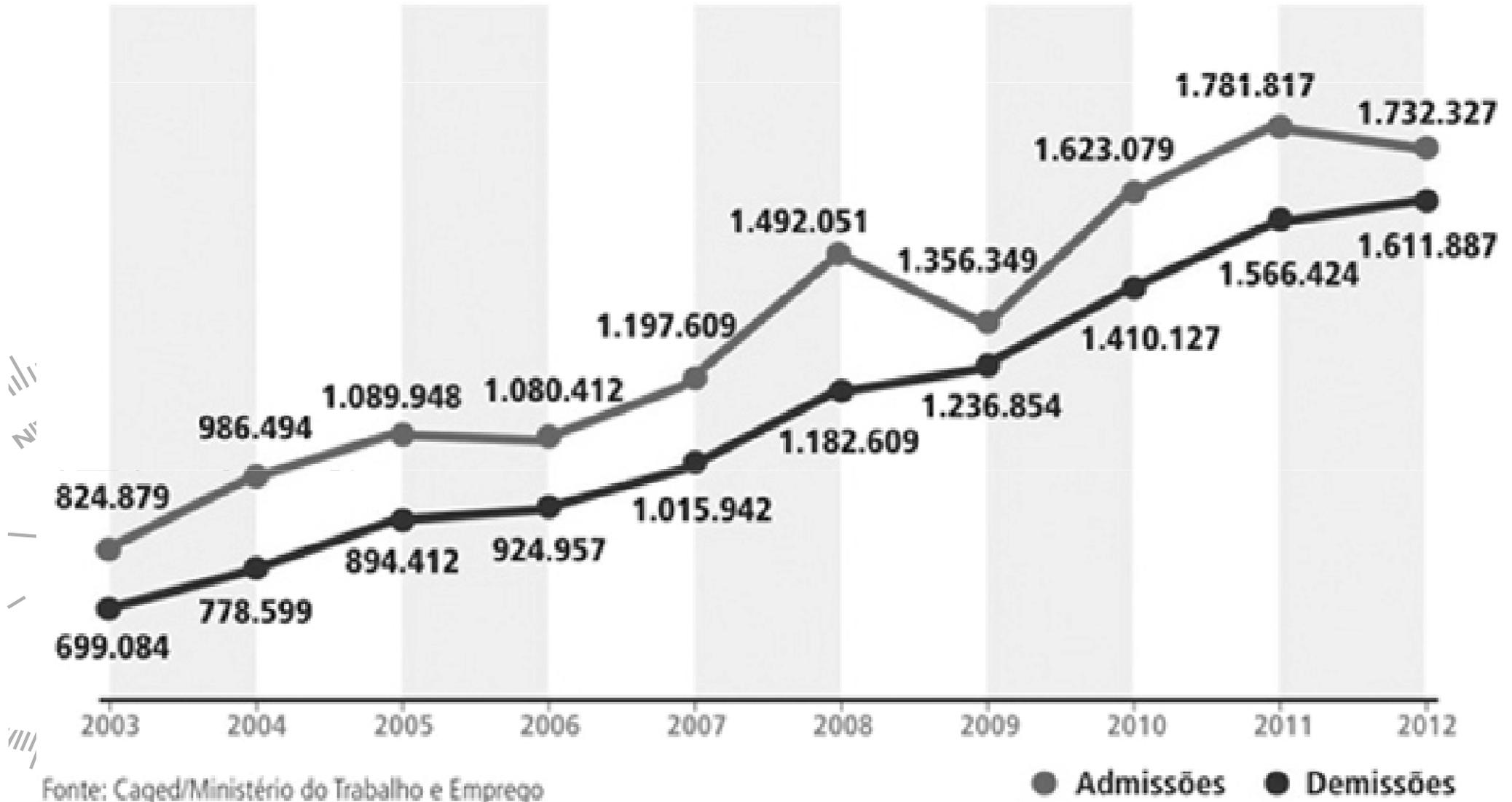
Vendas de 2008 e 2009



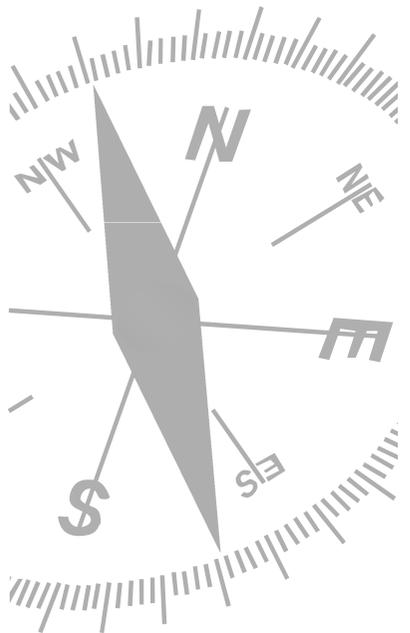
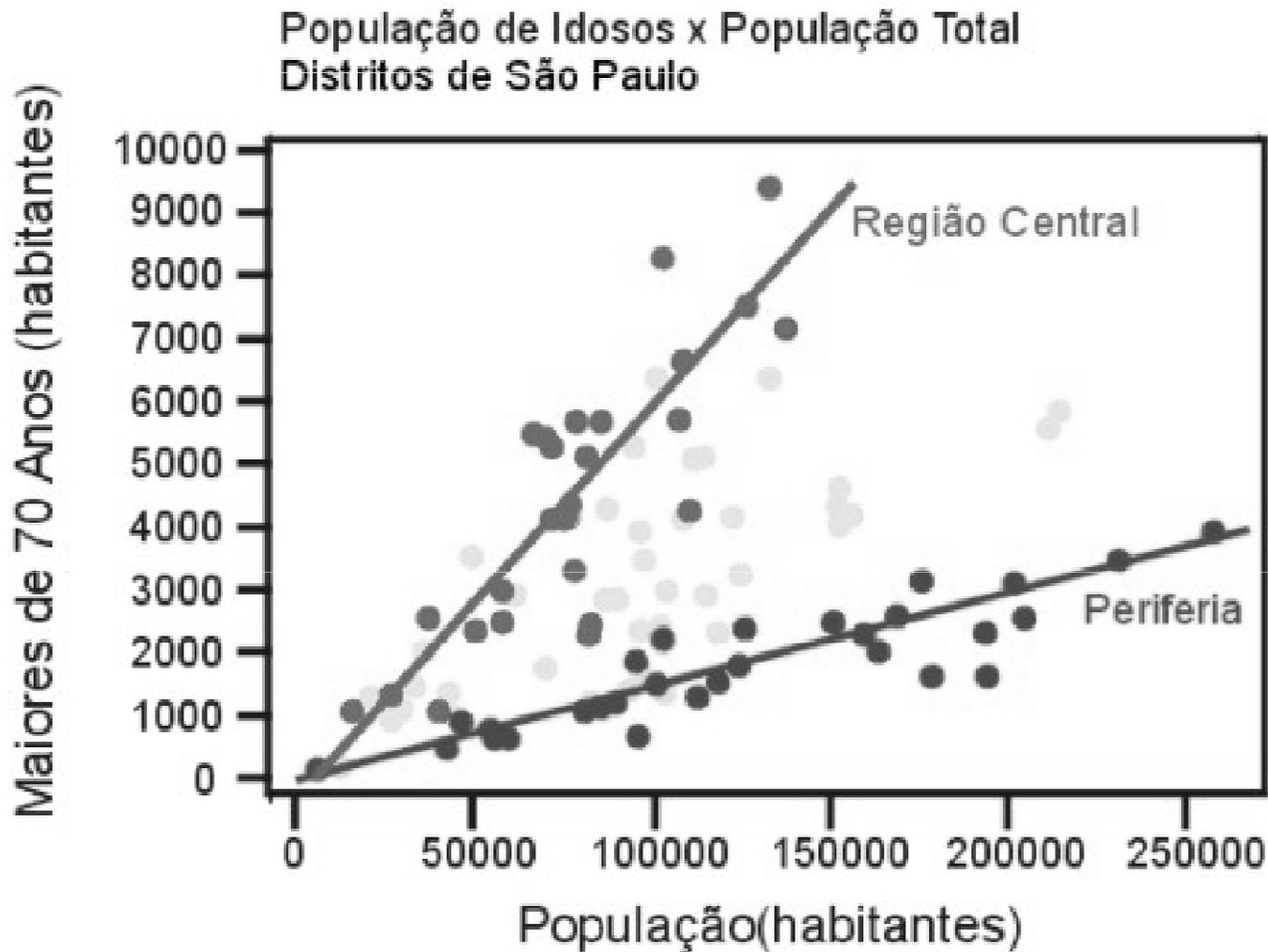
2) Gráfico de Setor: É a representação gráfica dos dados estatísticos em um círculo através de setores. As áreas são proporcionais aos valores da série. Utilizado principalmente para verificação de percentuais na amostra ou população em estudo.



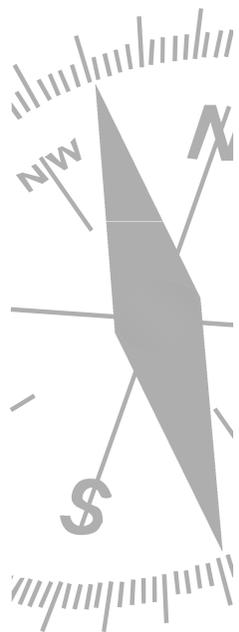
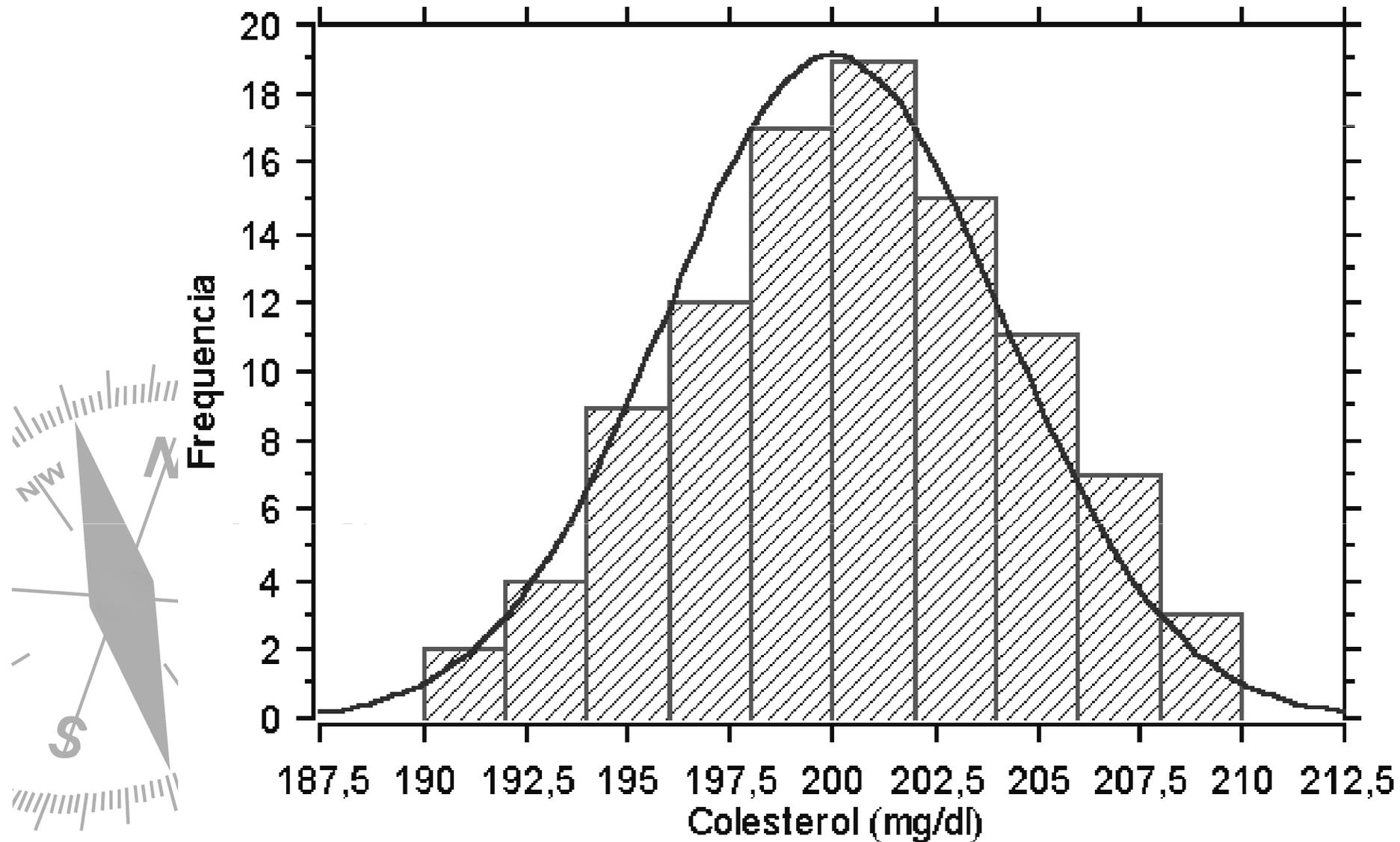
3) Gráfico de Linha: Utilizado usualmente para verificar o comportamento de uma determinada variável ao longo do tempo. O eixo x representa anos, meses, semestres, entre outros. Usado para uma série temporal.



4) Diagrama de Dispersão: Utilizado para verificar possíveis relações entre variáveis quantitativas.

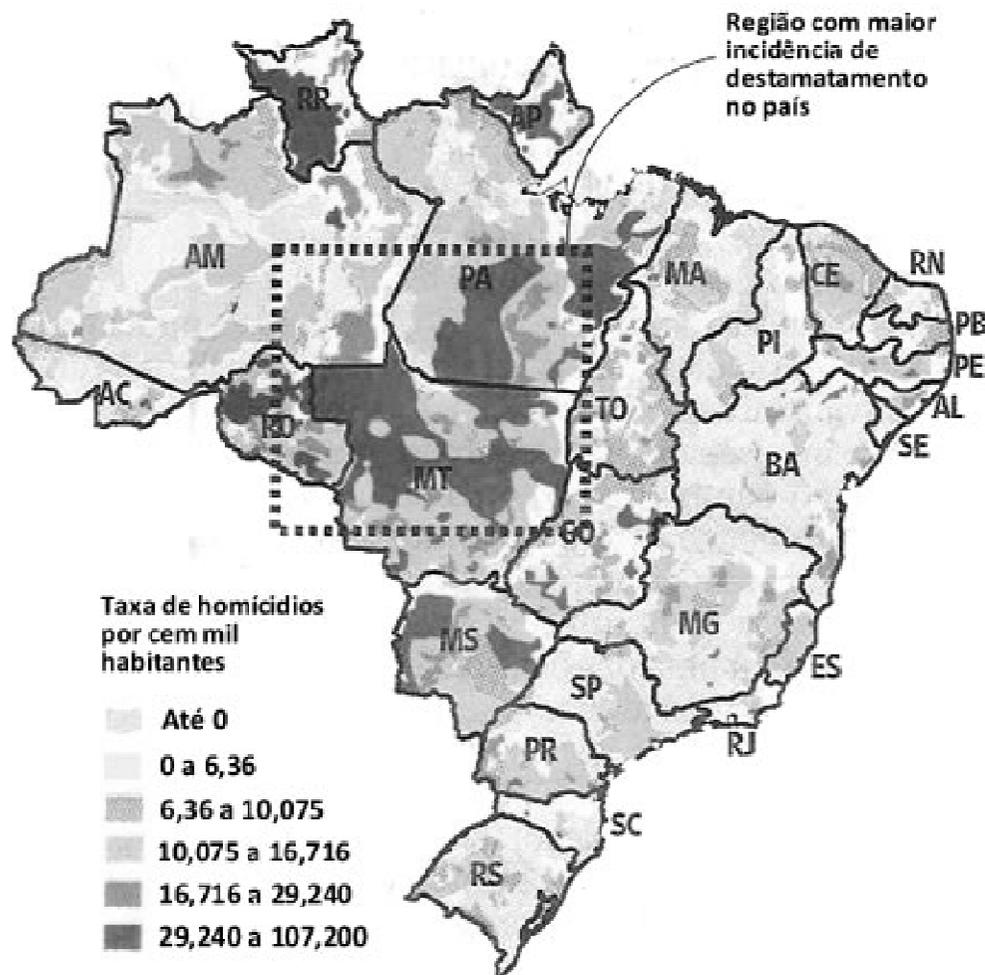


5) Histograma: É a representação gráfica de uma distribuição de frequência por meio de retângulos justapostos:



5) Mapas: É uma representação gráfica que permite explorar padrões geográficos nos dados.

Áreas com mais homicídios



Folha de S. Paulo, 01.02.2007.
Fonte: Datafolha.



EXEMPLO

- ▶ A Motorola esta interessada em comparar o orçamento projetado e o real a cada ano, entre 1990 e 1995. Os dados encontram-se na tabela abaixo:

Ano	Orçamento (em milhões de dólares)	
	Projetada	Real
1990	49,1	50,7
1991	53,7	54,2
1992	59,6	62,5
1993	67,4	67,2
1994	81,2	79,8
1995	78,9	80,1

Fonte: Departamento Financeiro – Motorola

Construa um título adequado para a tabela;

Construa um gráfico adequado.



EXEMPLO

Investimentos em Infra-Estrutura no Brasil, por Região — 2000

Região	Investimento (em milhões de US\$)
Norte	10,0
Nordeste	20,0
Centro-Oeste	5,7
Sudeste	9,3
Sul	10,0

Fonte: Ministério do Planejamento



Construa um gráfico adequado;