

# Probabilidade e estatística- Mat013

Professora - Hévilla Nobre Cezar

[hevilla@unifei.edu.br](mailto:hevilla@unifei.edu.br)

[hevilla@gmail.com](mailto:hevilla@gmail.com)

Mestre em Matemática Aplicada-USP

Doutoranda em Ensino de Ciências e Matemática-UNICAMP

Docente -Universidade Federal de Itajubá – UNIFEI - IMC

# Estatística

- Métodos que ajudam a transformar dados em informações úteis para tomada de decisões.

# Estatística e Uma forma de Raciocinar

- A estrutura **DCOVA**, consiste nas seguintes tarefas.
- **D**efinir
- **C**oletar
- **O**rganizar
- **V**isualizar
- **A**nalisar

# Objetivo

- Resumir e visualizar dados
- Tirar conclusões a partir desses dados
- Fazer prognósticos confiáveis
- Melhorar os processos

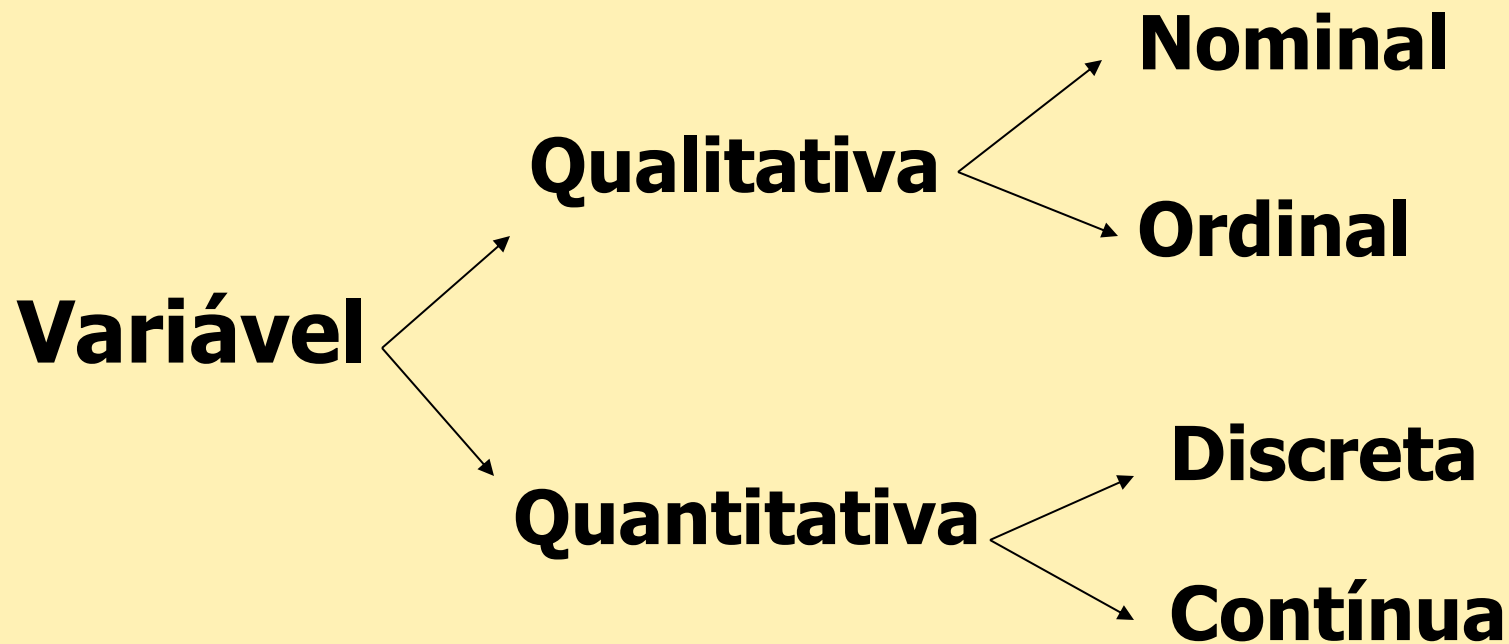
# Definir

- Objetivo da Pesquisa
- Que tipos de informações serão coletadas
- Definição das Variáveis

# Tipos de Variáveis

- Qualitativas – apresentam como possíveis realizações uma qualidade do indivíduo pesquisado
- Quantitativas – apresentam como possíveis realizações números resultantes de uma contagem ou mensuração

# Classificação de uma variável



# Coleta

- Fonte dos Dados
- População ou amostra
- Limpeza nos dados



# População e amostra

População (universo) = conjunto de todos os possíveis valores de uma variável ou característica.

Amostra = conjunto de observações extraída de uma população.

# Organização dos dados

- Após a coleta dos dados e classificação das variáveis, inicia-se o processo de organização.
- A organização se faz pelo desenvolvimento de **tabelas**.

# Visualização dos dados

- Para visualização dos dados utilizamos os recursos **gráficos**.

# Análise dos dados

- Com base nos dados coletados e organizados, tirar conclusões e apresentar os respectivos resultados.

# Para obter boas conclusões

- É necessário um bom planejamento:
  - Estabelecer a escala das variáveis
  - Escolher bem a fonte dos dados
  - Determinar o tipo de amostragem
  - Fazer limpeza dos dados
  - Estudar os possíveis erros de pesquisa
  - Pensar nas questões éticas da pesquisa

- Questionário:



# Representação Tabular

---

As tabelas são importantes para representação de dados estatísticos, pelo seu aspecto descritivo e maior facilidade na investigação.

Regras para confecção de tabelas:

- a) simplicidade e clareza
- b) auto-explicativas
- c) possuir título
- d) explicitar, em seu corpo, as variáveis e suas unidades
- e) Incluir rodapé quando necessário



# Elementos de uma tabela

---

- Título
- Cabeçalho
- Classificação (Histórica, Geográfica, Específica, Mista)
- Número de Linhas e Colunas
- Fonte
- Análise





# Exemplos de Tabela

---

Percentual de estudantes nos estágios de construção de competências		
Matemática – 8ª Série EF – Brasil – Saeb 2001 e 2003		
Estágio	2001	2003
Muito Crítico	6,7	7,3
Crítico	51,7	49,8
Intermediário	38,8	39,7
Adequado	2,8	3,3
Total	100,00	100,00

Fonte: SAEB 2001 e 2003



# Exemplos de Tabela

**Legenda: Construção de competências e desenvolvimento de habilidades na resolução de problemas em cada um dos estágios (resumo). 8ª Série – Matemática**

<b>Muito Crítico</b>	Não conseguem responder a comandos operacionais elementares compatíveis com a 8ª série. (Resolução de expressões algébricas com uma incógnita; características e elementos das figuras geométricas planas mais conhecidas).
<b>Crítico</b>	Desenvolveram algumas habilidades elementares de interpretação de problemas, mas não conseguem transpor o que está sendo pedido no enunciado para uma linguagem matemática específica, estando, portanto, muito aquém do exigido para a 8ª série. (Resolvem expressões com uma incógnita, mas não interpretam os dados de um problema fazendo uso de símbolos matemáticos específicos. Desconhecem as funções trigonométricas para resolução de problemas).
<b>Intermediário</b>	Adquiriram habilidades matemáticas mais compatíveis com oito anos de escolarização. Além das habilidades dos estágios anteriores, consolidaram habilidades que cabe destacar: identificam lados e ângulos de um quadrilátero (retângulo, losango, quadrado e trapézio); identificam o sistema de equações de primeiro grau, expressas em uma situação dada, lêem tabelas com números positivos e negativos e identificam o gráfico de colunas correspondente.
<b>Adequado</b>	Interpretam e sabem resolver problemas de forma competente; fazem uso correto da linguagem matemática específica. Apresentam habilidades compatíveis com a série em questão. (Interpretam e constroem gráficos; resolvem problema com duas incógnitas utilizando símbolos matemáticos específicos e reconhecem as funções trigonométricas elementares). Além disso, resolvem problemas simples envolvendo frações e porcentagens, equação de segundo grau, o conceito de proporcionalidade; resolvem expressão envolvendo as quatro operações, potências e raízes.



# Exemplos de Tabela

**Comparativos de custos do km rodado com diferentes combustíveis**

	Preço Combustível <sup>1</sup>	Consumo	Custo do Quilômetro Rodado <sup>2</sup> (R\$/km)
Diesel	1,848R\$/l	8km/l	0,2310
GNV	1,1R\$/m3	9km/m3	0,1222
Gasolina C	2,487R\$/l	11,4km/l	0,2182
Álcool Hidratado	1,509R\$/l	8km/l	0,1886
Hidrogênio C1	3,28R\$/m3	10,23km/m3	0,3206
Hidrogênio C2	2,61R\$/m3	10,23km/m3	0,2551
Hidrogênio C3	2,52R\$/m3	10,23km/m3	0,2463

**(Footnotes)**

<sup>1</sup> Preço médio para o estado de São Paulo em maio 2007.

<sup>2</sup> O custo do quilômetro rodado leva em consideração apenas o preço do combustível.

# Exemplos de Tabela

## indicadores

<b>Bovespa</b>	+0,71%	60,023	(14h32)
<b>Nasdaq</b>	+0,65%	3064,18	(16h05)
<b>Dólar com.</b>	+0,00%	R\$ 2,0350	(14h11)
<b>Euro</b>	+0,49%	R\$ 2,65250	(14h00)

Fonte: CMA

## INTERESSE POR POLÍTICA

Em %

■ 2008 ■ 2012

Muito interesse	12	14
Um pouco	36	40
Nenhum interesse	51	47

## RENDA POR DISTRITO

Em %

	Até R\$ 1.244	De R\$ 1.245 a R\$3.110	Mais de R\$ 3.111
Anhanguera	40	41	6
Brasilândia	44	41	9
Cachoeirinha	43	42	12
Casa Verde	24	44	23
Freguesia do Ó	30	33	19
Jaçanã	34	38	15
Jaraguá	38	39	11
Limão	26	42	23
Mandaqui	28	39	23
Perus	35	41	9
Pirituba	38	38	13
Santana	18	31	39
São Domingos	31	33	13
Tremembé	35	37	16
Tucuruvi	19	38	22
Vila Guilherme	22	38	22
Vila Maria	41	36	13



# frequência

---

frequência(  $n_i$  ) – número de vezes  
que ocorre as realizações.



# Tabela de frequência

---

Relaciona categorias (ou classes) de valores, juntamente com contagens (ou frequências) do número de valores que se enquadram em cada categoria.



# Tabela de frequência

Idade	Frequência
18	9
19	4
20	6
21	13
22	5
23	7
24	5
25	3
26	5
27	0
28	1
43	1

Variável discreta

Idade	frequência
18 + 20	13
20 + 22	19
22 + 24	12
24 + 26	8
26 + 28	5
28 + 44	2

Variável contínua



# Distribuição de frequência

---

Quando se estuda uma variável, o maior interesse é conhecer o comportamento dessa variável, analisando a ocorrência de suas possíveis realizações.





# Proporção ou frequência relativa

---

- Uma outra medida importante para análise de uma variável, é a proporção de cada realização em relação ao total

- Proporção( $f_i$ ) -  $f_i = \frac{n_i}{n}$

onde  $n$  = número total de realizações



# Porcentagem

---

A porcentagem é uma medida útil quando se quer comparar resultados de duas pesquisas distintas.

Definição:  $P_i = 100f_i$

Obs: a porcentagem é a frequência de uma variável em um total de 100 realizações.



# Tabela de frequência relativas

---

A tabela de frequências relativas tem as mesmas classes que a tabela de frequências, apenas, apresenta as frequências relativas em lugar das frequências absolutas.

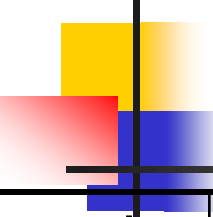


# Tabela de frequências acumuladas

---

A frequência acumulada de uma classe é a soma das frequências daquela classe e de todas as classes que a antecedem.

# Exemplo para variáveis discretas



Nº de filhos	frequências	frequências relativas	frequências acumuladas	Porcentagem
0	4	0,20	4	20
1	5	0,25	9	25
2	7	0,35	16	35
3	3	0,15	19	15
5	1	0,05	20	5
Total	20	1	-	100



# Tabela Dados Contínuos

---

Tabela de distribuição de frequência para variáveis contínuas e discretas quando não existem realizações iguais.

Neste caso, a solução é agrupar os dados por faixas de valores

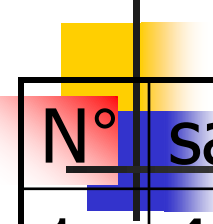


# Número de intervalos

---

- O número de intervalos depende do número de observações e da dispersão dos dados.
- Geralmente, de 5 a 20 intervalos são satisfatório.
- O número de intervalos pode ser aproximadamente a raiz quadrada do número de observações.

# Exemplo



N°	salário
1	4.00
2	4.56
3	5.25
4	5.75
5	6.26
6	6.66
7	6.86
8	7.39
9	7.59

N°	salário
10	7.44
11	8.12
12	8.46
13	8.74
14	8.95
15	9.13
16	9.35
17	9.77
18	9.80

N°	salário
19	10.53
20	10.76
21	11.06
22	11.59
23	12.00
24	12.79
25	13.23
26	13.60
27	13.85

N°	salário
28	14.69
29	14.71
30	15.99
31	16.22
32	16.61
33	17.26
34	18.75
35	19.40
36	23.30

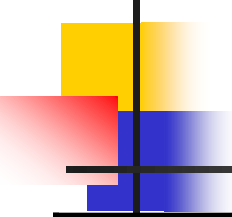




# Exemplo: Tabela distribuição de frequência

Classe de salários	frequência	porcentagem
[4.00 , 8.00)	10	27,78
[8.00 , 12.00)	12	33,33
[12.00 , 16.00)	8	22,22
[16.00 , 20.00)	5	13,89
[20.00 , 24.00)	1	2,78
Total	36	100,00

# Exemplo: Tabela distribuição de frequência



Classes de Salários	Ponto médio	frequência	Porcentagem
[4.00 , 8.00)	6.00	10	27,78
[8.00 , 12.00)	10.00	12	33,34
[12.00 , 16.00)	14.00	8	22,22
[16.00 , 20.00)	18.00	5	13,89
[20.00 , 24.00)	22.00	1	2,78
Total	-	36	100



# Exemplo para variável ordinal

Grau de Instrução	frequência	frequência Relativa	frequência Acumulada	Porcentagem
1°	650	0,325	650	32,50
2°	1.020	0,510	1.670	51,00
3°	330	0,165	2.000	16,50
Total	2.000	1,000	-	100



# Representação Gráfica

---

Os gráficos facilitam a visualização dos valores que são amplamente usados na apresentação de dados estatísticos.

Cabe destacar que os gráficos perdem informações, porque não mostram as observações originais.

Ao inferir sobre uma população a partir de gráficos, deve-se ter bastante cuidado.

# Elementos de um Gráfico

- Título
- Classificação (Barra ou coluna, Linha, Setor Circular, Pictograma, etc.)
- Fonte
- Análise

# Apresentação Gráfica

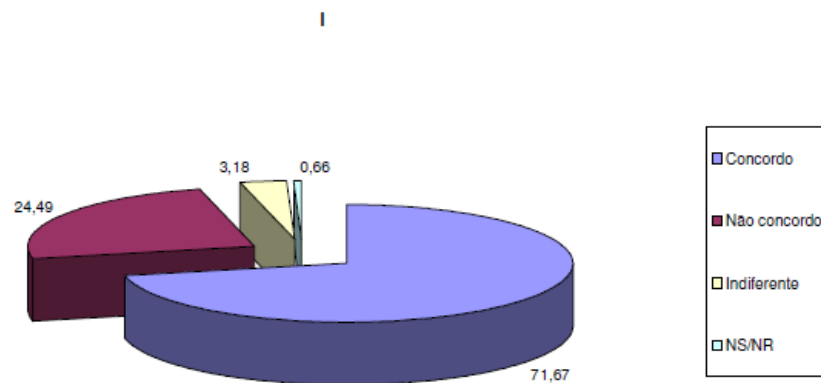
- Para variáveis qualitativas
  - Gráficos em barra
  - Gráficos em setores (“pizza”)

# Apresentação Gráfica

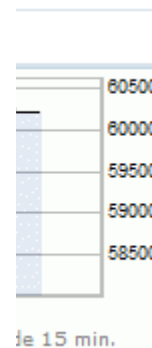
- Para variáveis quantitativas existe uma variedade de representações gráficas
  - Barras
  - Colunas
  - Gráficos de dispersão
  - Histogramas

# Exemplos de Gráficos

**GRÁFICO 2:** Distribuição dos responsáveis segundo a opinião sobre a sugestão de outros pais: "Proibir o trabalho de crianças que estão estudando". SUDESTE – Percentual.

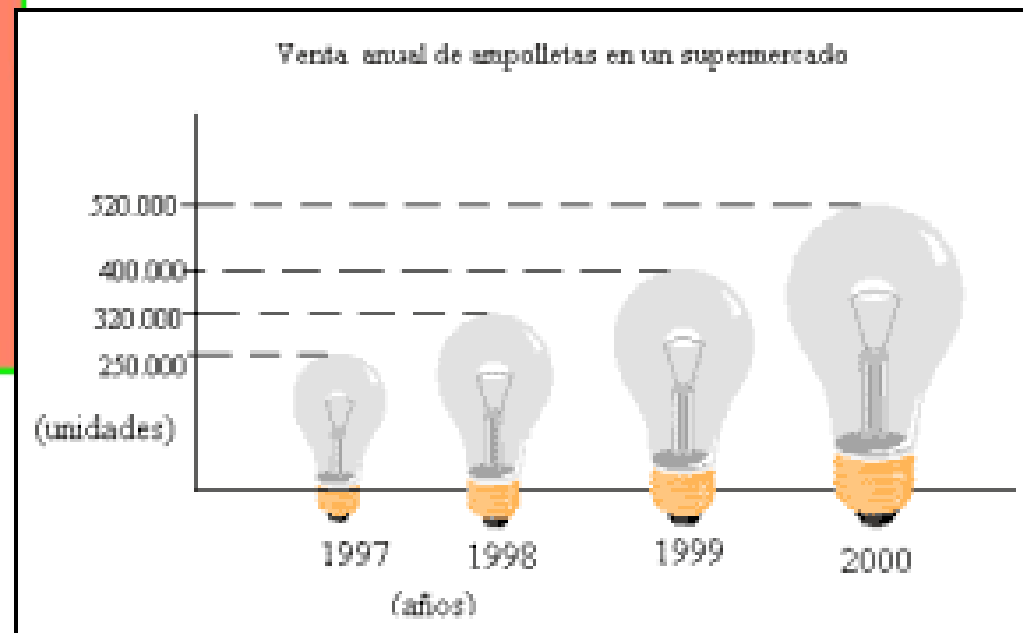
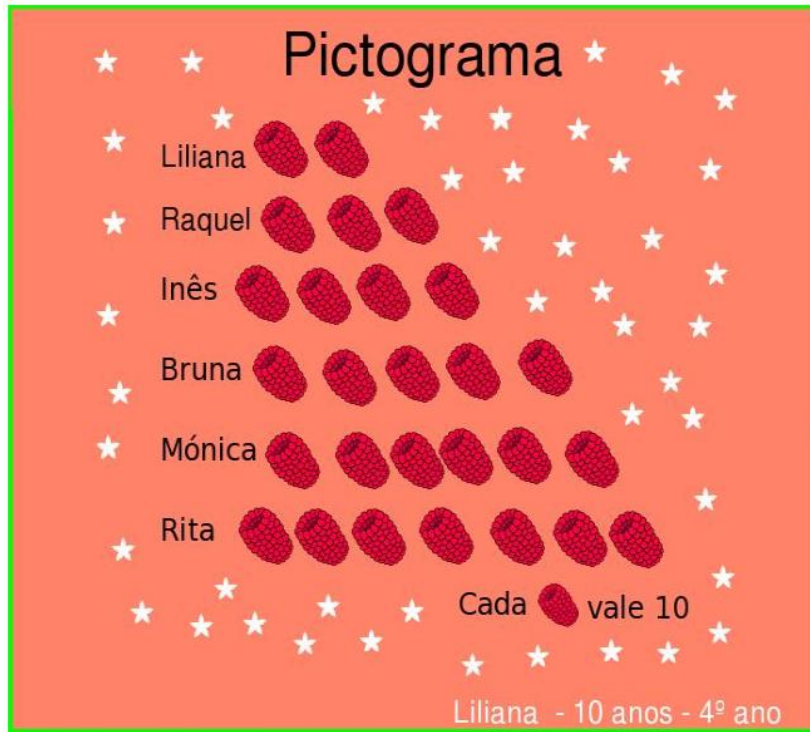


Fonte: MEC/INEP/DAEB

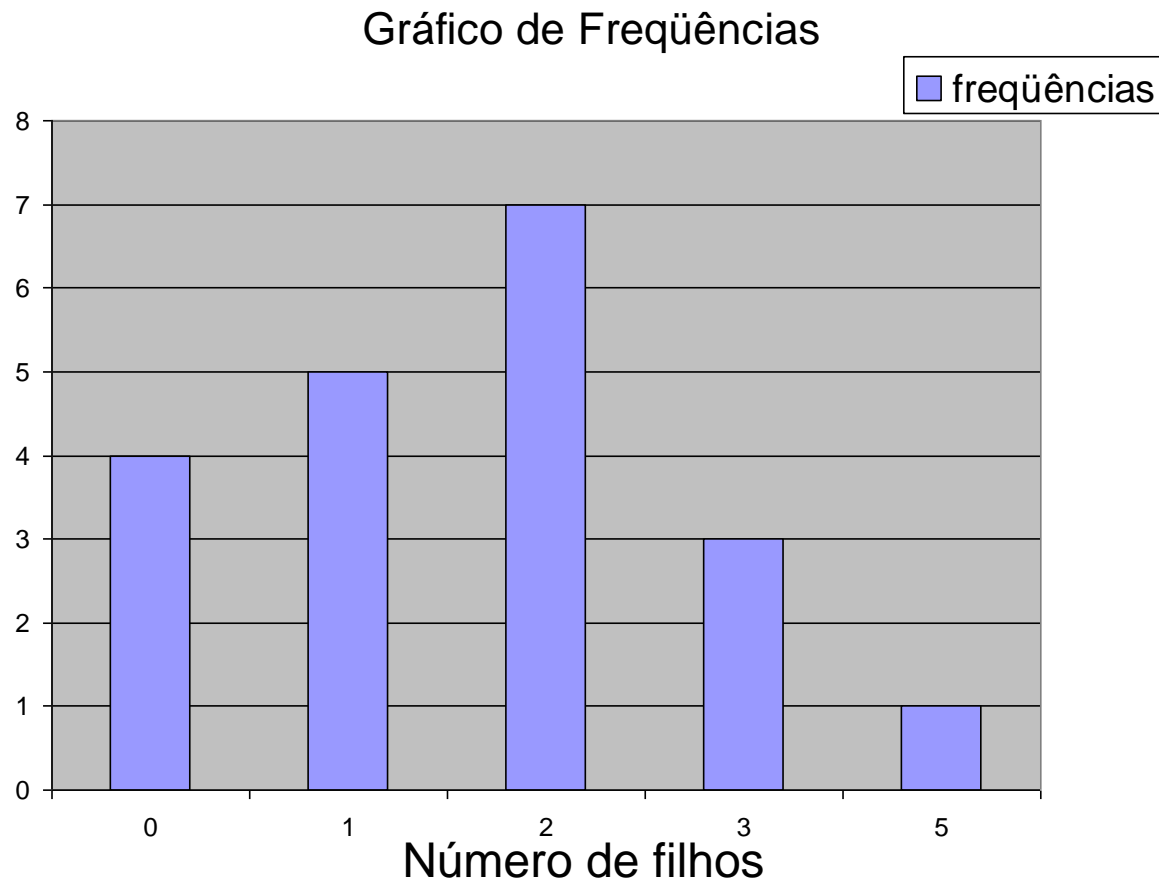




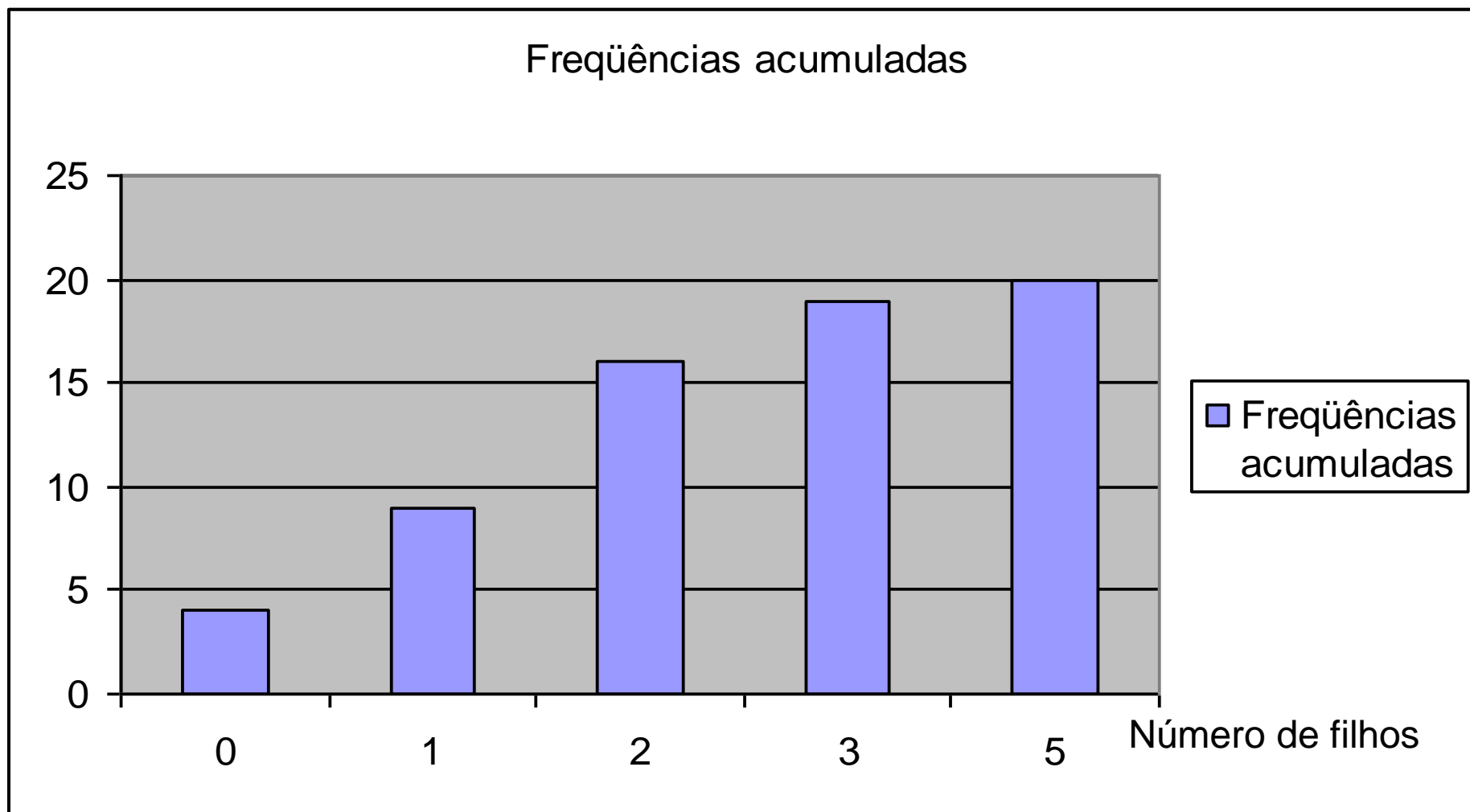
# Exemplos de Gráficos



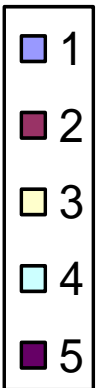
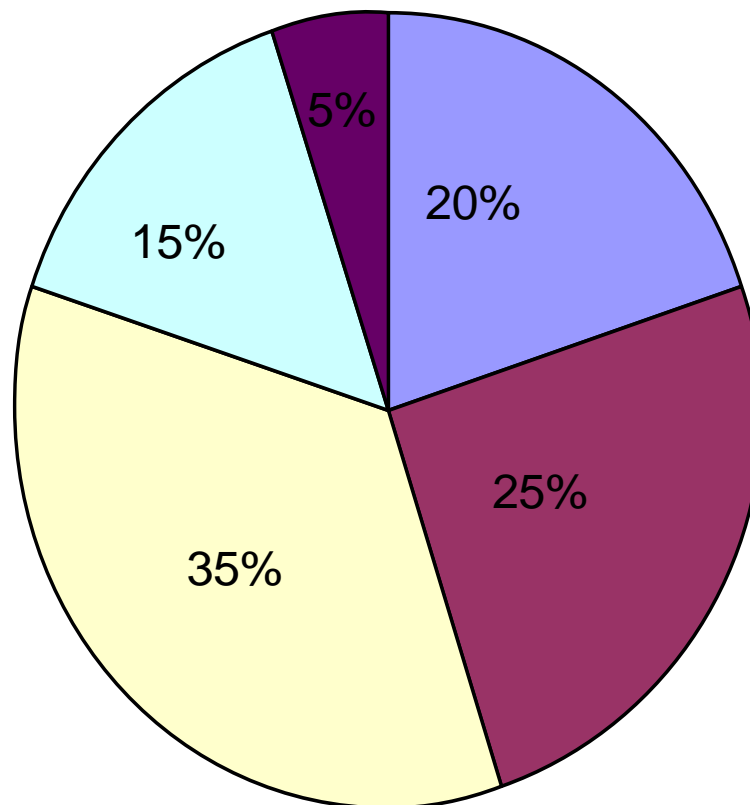
# Gráfico de Barras para variável discreta



# Gráfico de Barras para variável discreta



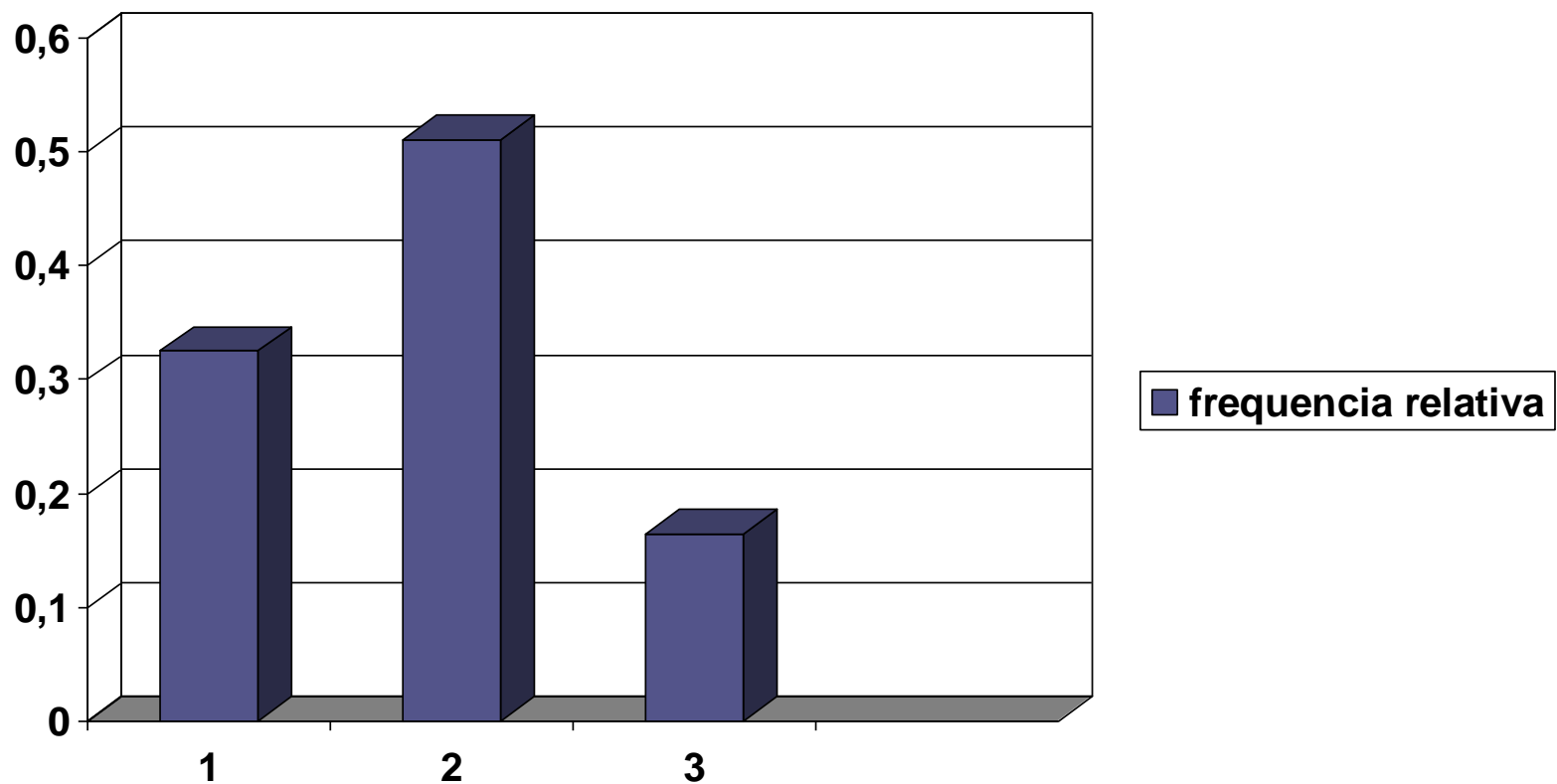
## Gráfico de Setores – para porcentagem



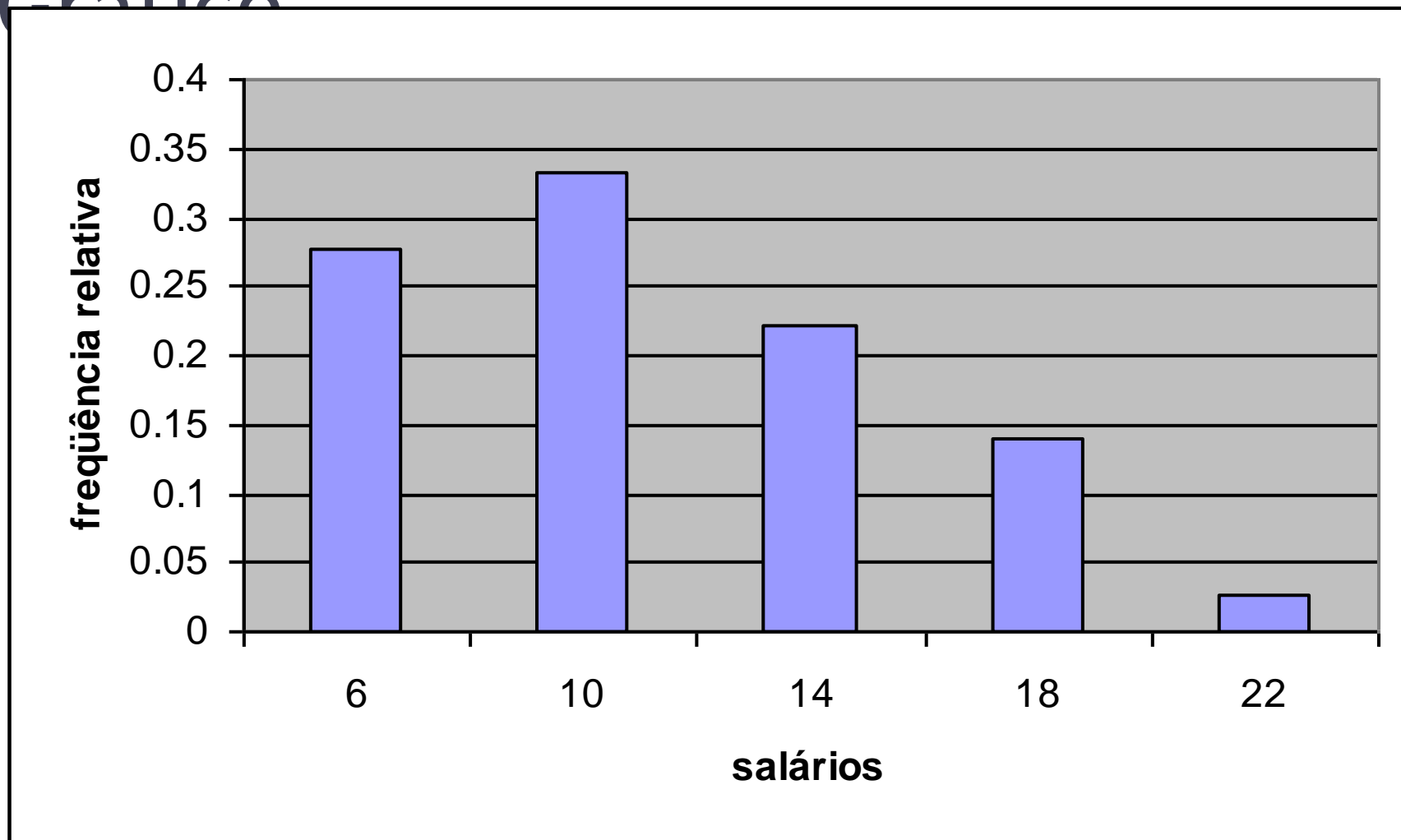
# Exemplo para variável ordinal

Grau de Instrução	frequência	frequência Relativa	frequência Acumulada	Porcentagem
1°	650	0,325	650	32,50
2°	1.020	0,510	1.670	51,00
3°	330	0,165	2.000	16,50
Total	2.000	1,000	-	100

# Gráfico de frequências



# Gráfico

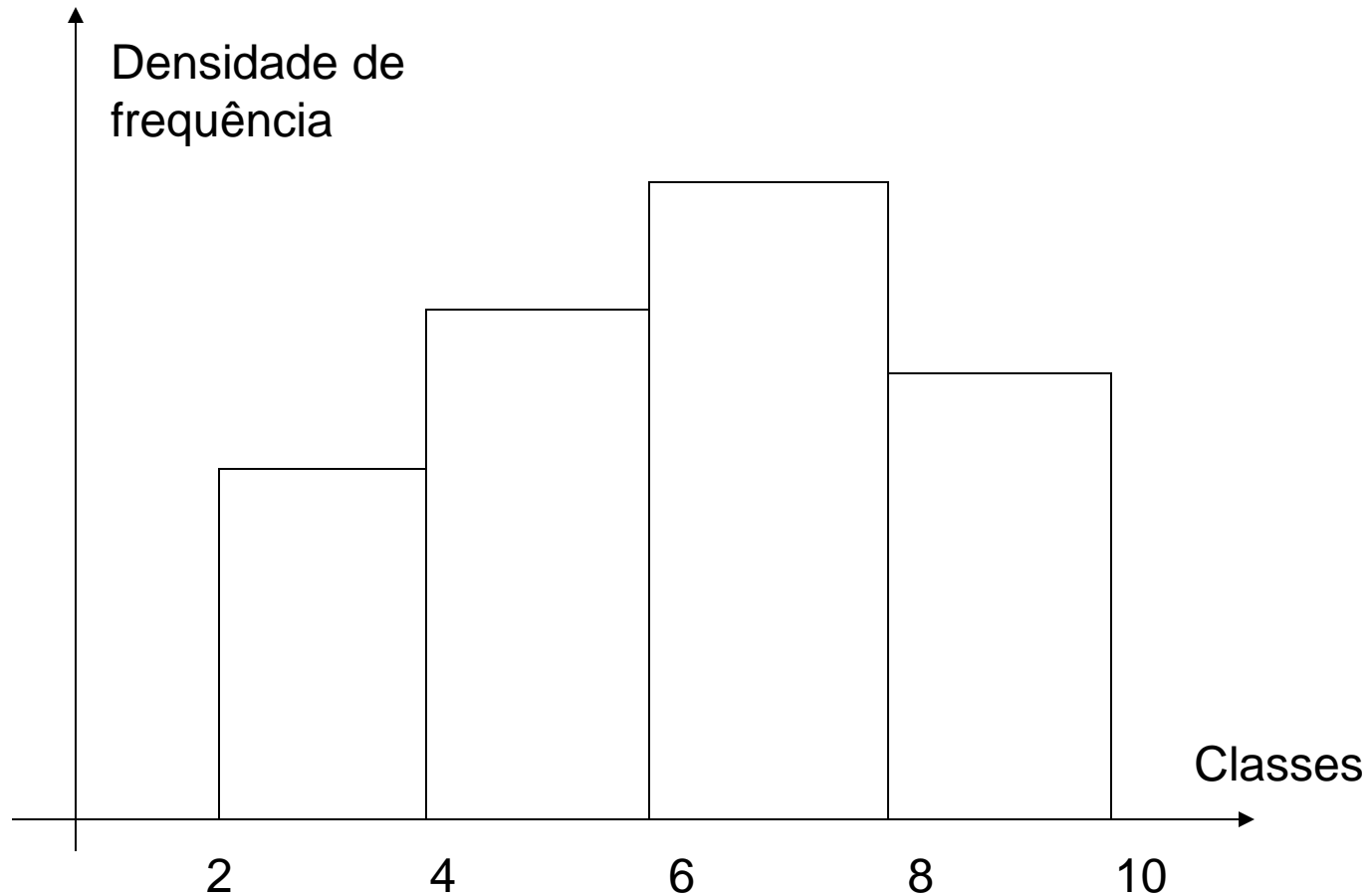


# Histograma

O histograma é um gráfico por setores contíguos, onde a altura é proporcional a frequência relativa e a base é constituída por um segmento cujos extremos representam os extremos das classes.

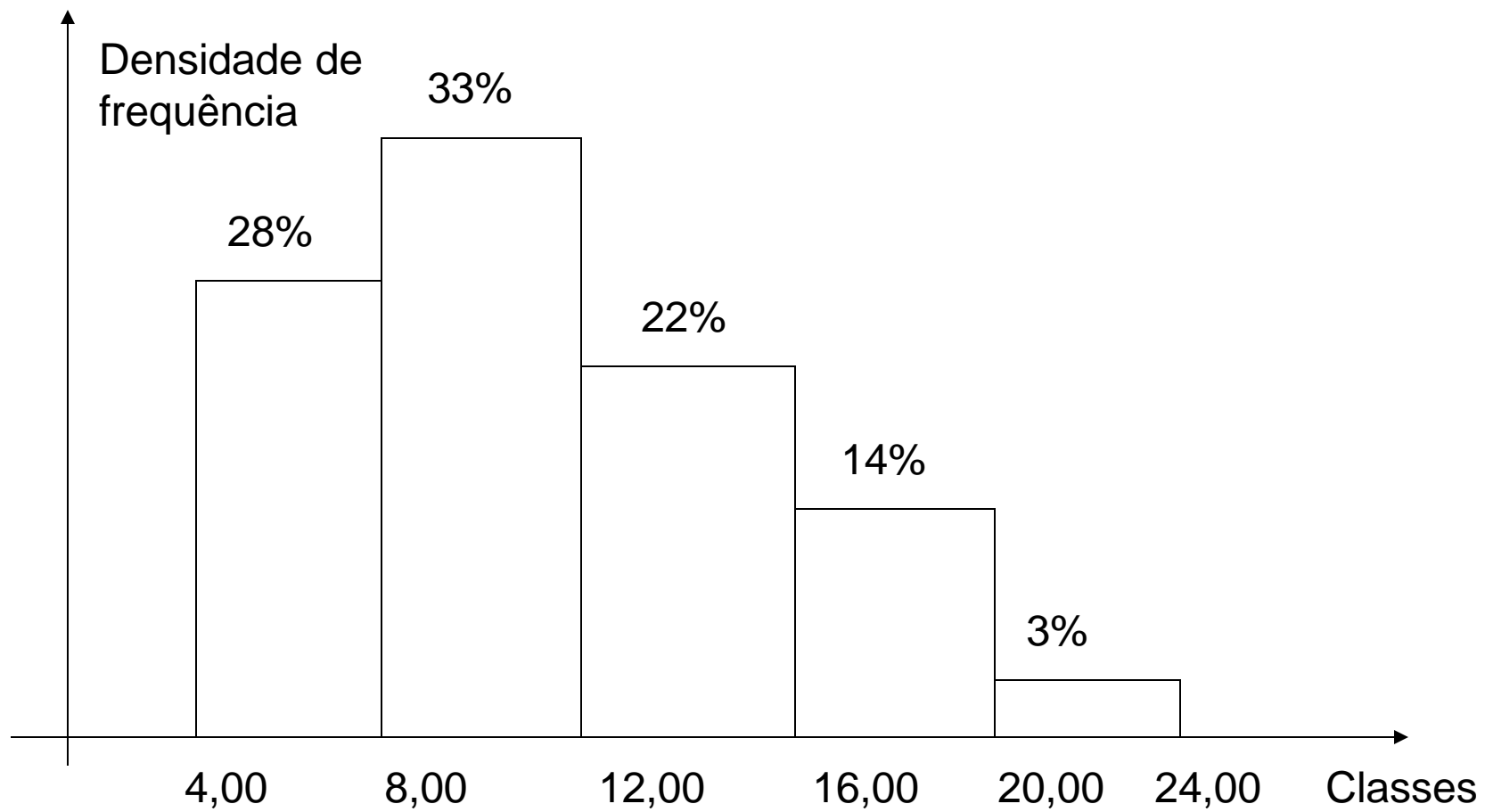


# Esboço de um histograma



# Tabela

Classe de salários	frequência relativa	Densidade de frequência
[4.00 , 8.00)	0,2778	0,06945
[8.00 , 12.00)	0,3334	0,08335
[12.00 , 16.00)	0,2222	0,05555
[16.00 , 20.00)	0,1389	0,034725
[20.00 , 24.00)	0,0278	0,00695
Total	1	-



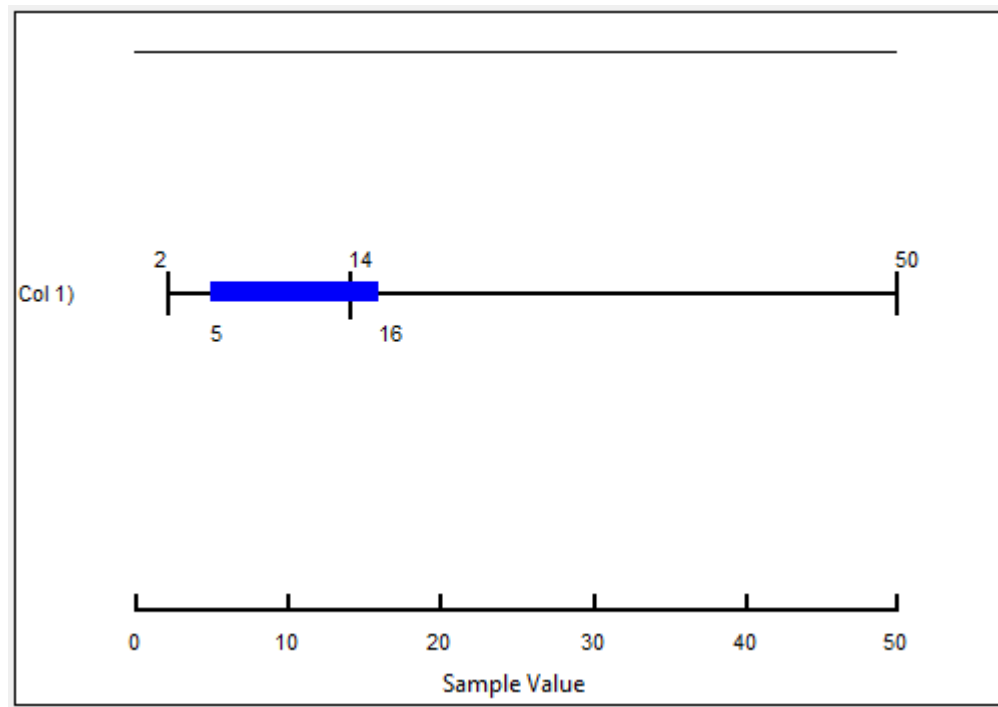
# Diagramas de Caixa (Boxplots)

- Um diagrama de caixa é um gráfico de um conjunto de dados que consiste em uma linha que se estende do valor mínimo ao valor máximo, em que uma caixa com linhas traçadas no primeiro quartil,  $Q_1$ , na mediana e no terceiro quartil,  $Q_3$ .

# Características

- Informa a variabilidade e simetria dos dados.
- São úteis para detectar diferenças nos comportamentos de grupos de variáveis.

# Diagramas de Caixa (Boxplots)



# Diagrama de *ramo-e-folhas*

- Dados estatísticos, gerados em grandes massas, podem se muito úteis para o estudo do comportamento da distribuição se apresentados em uma combinação de tabela e gráfico chamado diagrama de *ramos-e-folhas*.

# Diagrama de *ramo-e-folhas*

Vida útil das baterias de carro (anos)

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

Ramo	Folhas	Frequência
1	69	2
2	25669	5
3	0011112223334445567778899	25
4	11234577	8

Diagrama de ramo-e-folha da vida útil das baterias



# Diagrama de *ramo-e-folhas*

Vida útil das baterias de carro (anos)

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

Ramo	Folhas	Frequência
1●	69	2
2*	2	1
2●	5669	4
3*	001111222333444	15
3●	5567778899	10
4*	11234	5
4●	577	3

Diagrama de ramo-e-folha duplo da vida útil das baterias