

AULA1-AULA5

Estadística Descriptiva

Prof. Victor Hugo Lachos Davila

O que é a estatística ?

Para muitos, a estatística não passa de conjuntos de tabelas de dados numéricos. Os estatísticos são pessoas que coletam esses dados.

- A estatística originou-se com a coleta e construção de tabelas de dados para os governos
- A situação evoluiu e esta coleta de dados representa somente um dos aspectos da estatística.

Definição de Estatística

A estatística é um conjunto de técnicas que permite, de forma sistemática, organizar, descrever, analisar e interpretar dados oriundos de estudos ou experimentos, realizados em qualquer área do conhecimento.

Áreas da Estatística

- 1.- Estatística Descritiva
- 2.- Probabilidade
- 3.- Inferência estatística

ESTATÍSTICA DESCRITIVA

A estatística descritiva é a etapa inicial da análise utilizada para descrever e resumir os dados. A disponibilidade de uma grande quantidade de dados e de métodos computacionais muito eficientes revigorou esta área da estatística.

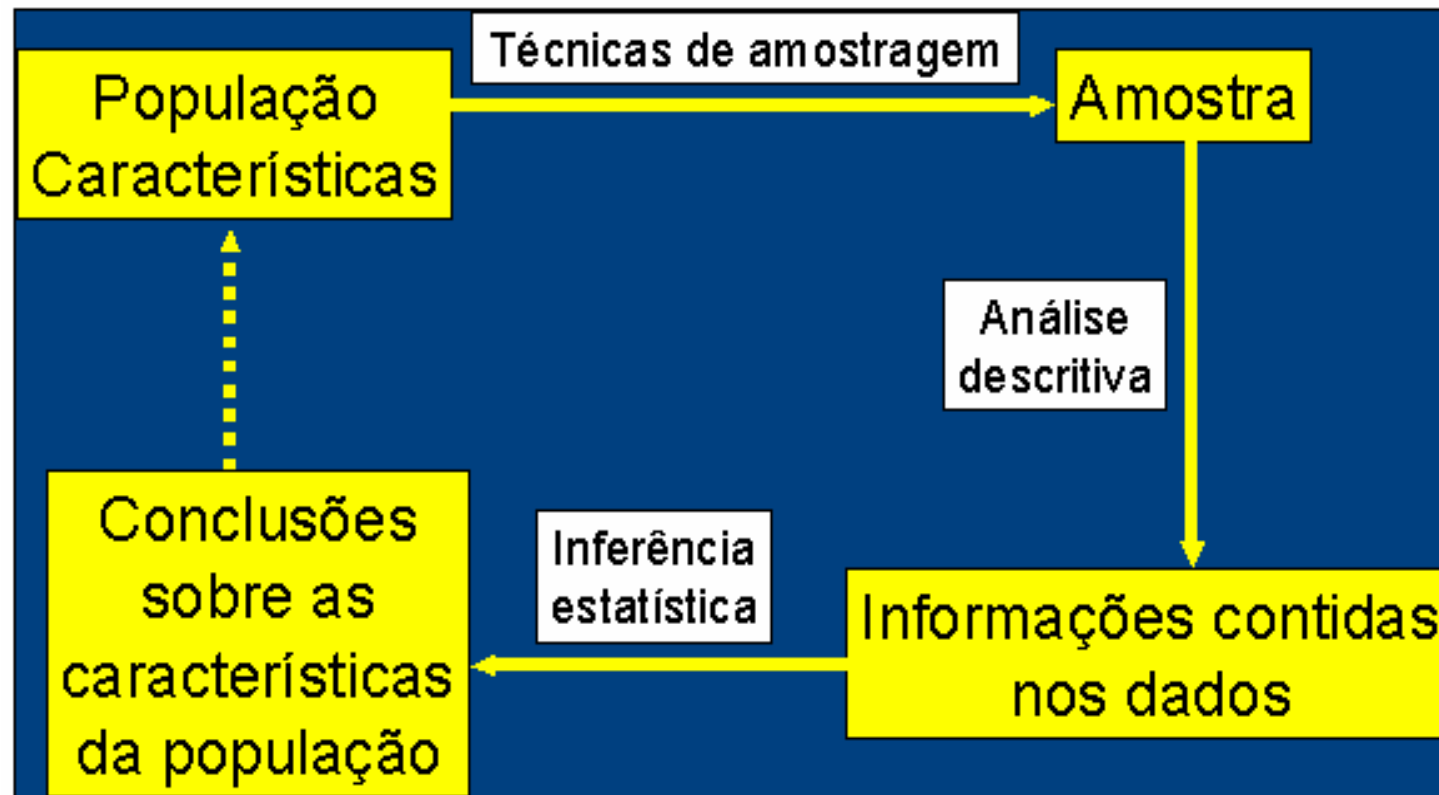
PROBABILIDADE

A teoria de probabilidades nos permite descrever os fenômenos aleatórios, ou seja, aqueles em que está presente a incerteza.

INFERENCIA ESTADÍSTICA

É o estudo de técnicas que possibilitam a extrapolação, a um grande conjunto de dados, das informações e conclusões obtidas a partir da amostra.

Etapas da Análise Estatística

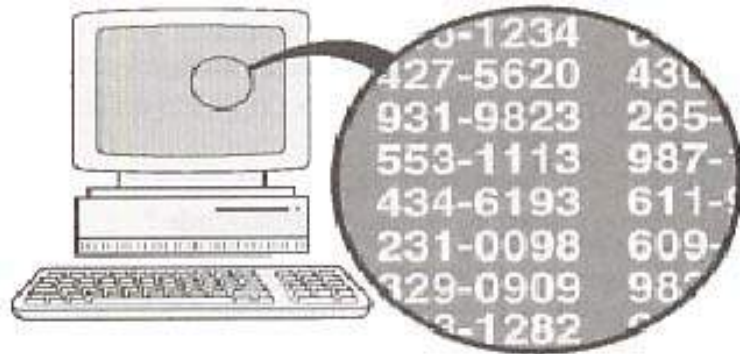


AMOSTRAGEM

Uma área importante em muitas aplicações Estatísticas é a da **Tecnologia de Amostragem**.

Exemplos de Aplicação:

- Pesquisa de mercado,
- Pesquisa de opinião,
- Avaliação do processo de produção,
- Praticamente em todo experimento.



Amostragem Aleatória

Cada elemento da população tem a mesma chance de ser escolhido.



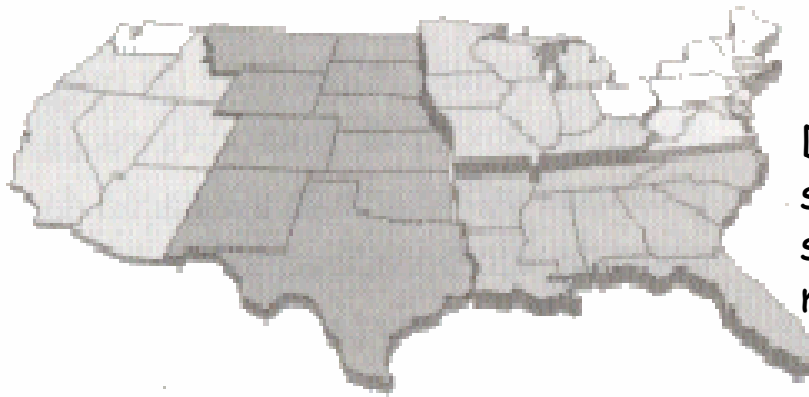
Amostragem Estratificada

Classificar a população em, ao menos dois estratos e extrair uma amostra de cada um.



Amostragem Sistemática

Escolher cada elemento de ordem k .

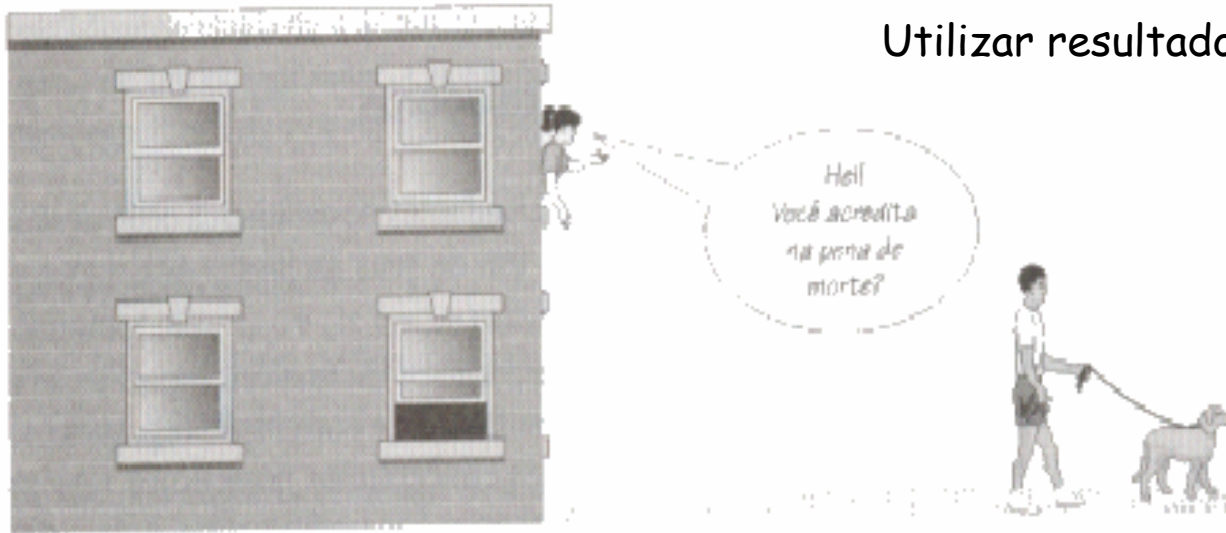


Amostragem por Conglomerados

Dividir em seções a área populacional, selecionar aleatoriamente algumas dessas seções e tomar todos os elementos das mesmas.

Amostragem de Conveniência

Utilizar resultados de fácil acesso.



Exemplo 1

Numa pesquisa eleitoral, um instituto de pesquisa procura, com base nos resultados de um levantamento aplicado a uma amostra da população, prever o resultado da eleição.

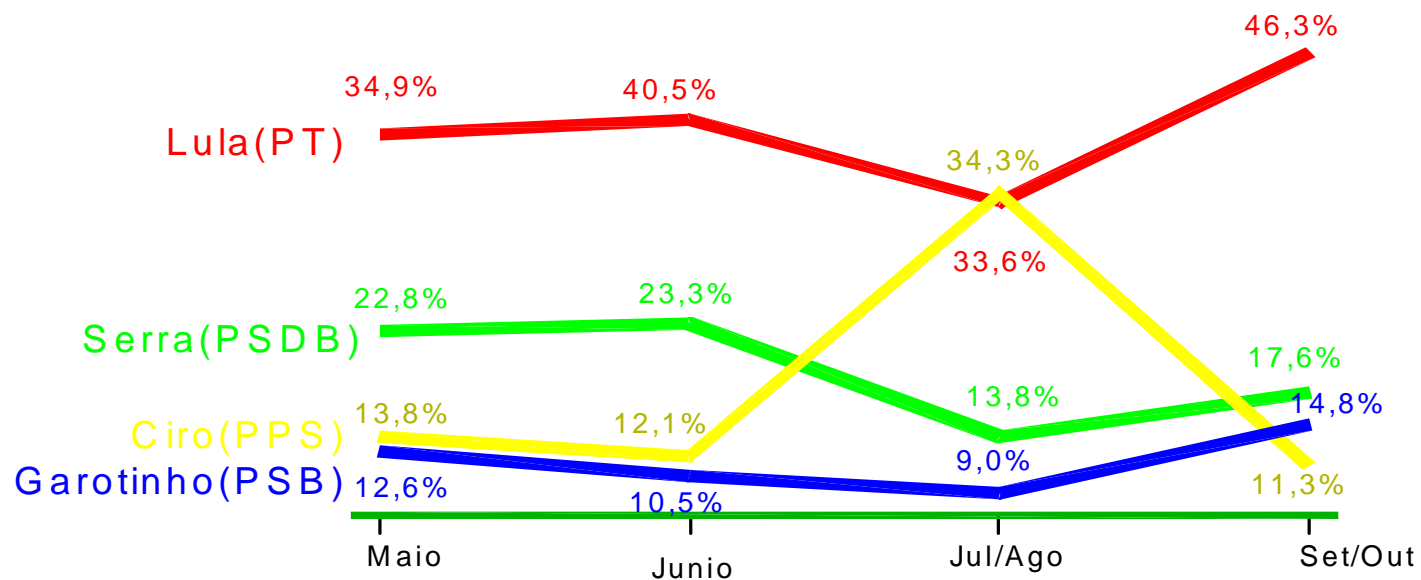
Na eleição Presidencial

Os Institutos de Pesquisa de opinião colhem periodicamente amostras de eleitores para obter as estimativas de intenção de voto da população. As estimativas são fornecidas com um valor e uma margem de erro.

O quadro do **Instituto Toledo & Associados**, a seguir refere-se à intenção de voto no 1º turno das eleições para o governo em 2002.

Intenção de voto para presidente do Brasil-2002

Voto estimulado, em % do total de votos. A última pesquisa ouviu 2.202 eleitores - Margem de erro de 2,09%



Fonte: Pesquisa toledo & Associados.

Confronto no segundo turno.

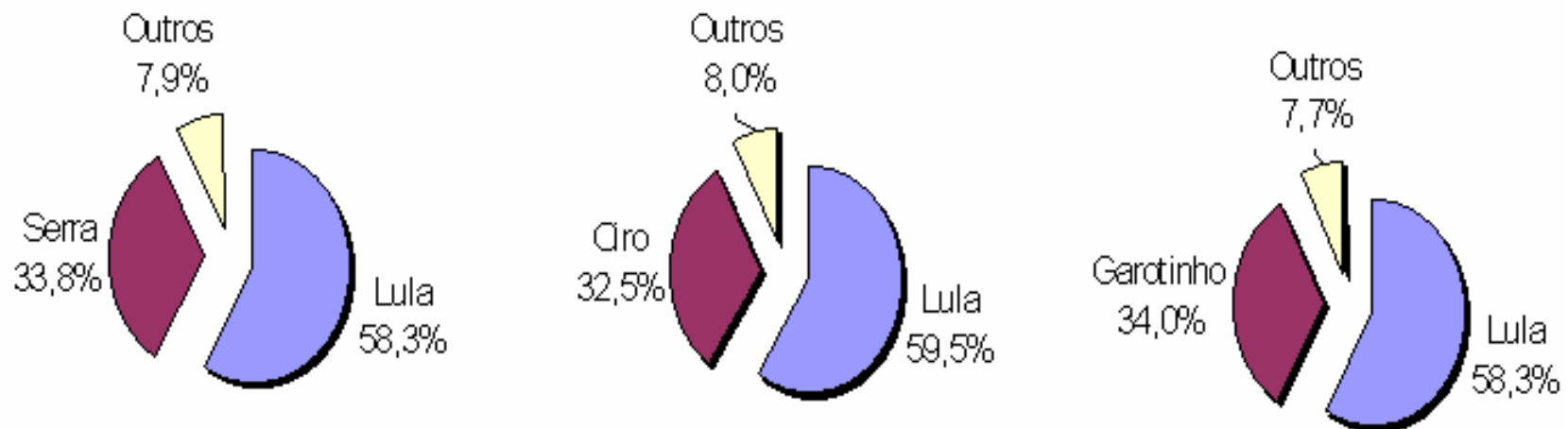
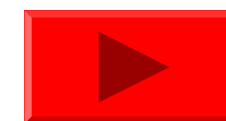


Gráfico de setores ou em forma de pizza

Tabela 1.1 Informação do estado civil, grau de instrução, número de filhos, idade e procedência de 36 funcionários sorteados ao acaso da empresa MB.(Bussab e Morettin)

Nº	Estado Civil	Grau de Instrução	No de filhos	Salário (X Sal. Min)	Idade anos meses	Região de procedência
1	Solteiro	1º grau	-	4,00	26 03	Interior
2	Casado	1º grau	1	4,56	32 10	Capital
3	Casado	1º grau	2	5,25	36 05	Capital
4	Solteiro	2º grau	-	5,73	20 10	Outro
5	Solteiro	1º grau	-	6,26	40 07	Outro
6	Casado	1º grau	0	6,66	28 00	Interior
7	Solteiro	1º grau	-	6,86	41 00	Interior
8	Solteiro	1º grau	-	7,39	43 04	Capital
9	Casado	2º grau	1	7,59	34 10	Capital
10	Solteiro	2º grau	-	7,44	23 06	Outro
11	Casado	2º grau	2	8,12	33 06	Interior
12	Solteiro	1º grau	-	8,46	27 11	Capital
13	Solteiro	2º grau	-	8,74	37 05	Outro
14	Casado	1º grau	3	8,95	44 02	Outro
15	Casado	2º grau	0	9,13	30 05	Interior
16	Solteiro	2º grau	-	9,35	38 08	Outro
17	Casado	2º grau	1	9,77	31 07	Capital
18	Casado	1º grau	2	9,80	39 07	Outro
19	Solteiro	Superior	-	10,53	25 08	Interior
20	Solteiro	2º grau	-	10,76	37 04	Interior
21	Casado	2º grau	1	11,06	30 09	Outro
22	Solteiro	2º grau	-	11,59	34 02	Capital
23	Solteiro	1º grau	-	12,00	41 00	Outro
24	Casado	Superior	0	12,79	26 01	Outro
25	Casado	2º grau	2	13,23	32 05	Interior
26	Casado	2º grau	2	13,60	35 00	Outro
27	Solteiro	1º grau	-	13,85	46 07	Outro
28	Casado	2º grau	0	14,69	29 08	Interior
29	Casado	2º grau	5	14,71	40 06	Interior
30	Casado	2º grau	2	15,99	35 10	Capital
31	Solteiro	Superior	-	16,22	31 05	Outro
32	Casado	2º grau	1	16,61	36 04	Interior
33	Casado	Superior	3	17,26	43 07	Capital
34	Solteiro	Superior	-	18,75	33 07	Capital
35	Casado	2º grau	2	19,40	48 11	Capital
36	Casado	Superior	3	23,30	42 02	Interior



Estatística Descritiva

O que fazer com as observações que coletamos?



Primeira Etapa:

Resumo dos dados = Estatística descritiva

Variável

Qualquer característica associada a uma população

Classificação de variáveis

Qualitativa

Nominal

sexo, cor dos olhos

Ordinal

Classe social, grau de instrução

Quantitativa

Contínua

Peso, altura, salario

Discreta

Número de filhos, numero de carros

Medidas Resumo

Variáveis Quantitativas

MEDIDAS DE POSIÇÃO: Moda, Média, Mediana, Percentís, Quartis.

MEDIDAS DE DISPERSÃO: Amplitude, Intervalo-Interquartil, Variância, Desvio Padrão, Coeficiente de Variação.

Medidas de Posição

Moda(mo): É o valor (ou atributo) que ocorre com maior frequência.

Ex: 4,5,4,6,5,8,4,4

Mo = 4



Variavel
qualitativa

Média

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Ex: 2, 5, 3, 7, 8

$$\text{Média} = [(2+5+3+7+8)/5]=5$$

Mediana

A mediana é o valor da variável que ocupa a posição central de um conjunto de n dados ordenados.

Posição da mediana: $(n+1)/2$

Ex: 2,5,3,7,8

Dados ordenados: 2,3,5,7,8 $\Rightarrow (5+1)/2=3$
 $\Rightarrow Md = 5$

Ex: 3,5,2,1,8,6

Dados ordenados: 1,2,3,5,6,8 \Rightarrow
 $(6+1)/2=3,5 \Rightarrow Md=(3+5)/2=4$

Percentis

O percentil de ordem $px100$ ($0 < p < 1$), em um conjunto de dados de tamanho n , é o valor da variável que ocupa a posição $px(n+1)$ do conjunto de dados ordenados.

O percentil de ordem p (ou p -quantil) deixa $px100\%$ das observações abaixo dele na amostra ordenada.

Casos Particulares:

Percentil 50=mediana, segundo quartil($md, Q2, q(0,5)$)

Percentil 25= primeiro quartil ($Q1$), $q(0,25)$

Percentil 75= terceiro quartil ($Q3$), $q(0,75)$

O p-quantil, $0 < p < 1$, pode ser calculado como:

$$q(p) = \begin{cases} x_{(i)}, & \text{se } p = p_i = \frac{i - 0.5}{n}, \quad i = 1, \dots, n \\ (1 - f_i)q(p_i) + f_i q(p_{i+1}), & \text{se } p_i < p < p_{i+1} \\ x_{(1)}, & \text{se } p < p_1 \\ x_{(n)}, & \text{se } p > p_n \end{cases}$$

Onde:

$$f_i = \frac{p_{i+1} - p_i}{p_{i+1} - p_i},$$

$$p_i = \frac{i - 0.5}{n}$$

Estatísticas de ordem

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)}$$

Exemplos

Ex(1): 15,5,3,8,10,2,7,11,12

=>n=9

=> ordenamos: 2<3<5<7<8<10<11<12<15

P1=1/18; p2=3/18; p3=5/18; p4=7/18; p5=1/2;

p6=11/18; p7=13/18; p8=15/18; p9=17/18

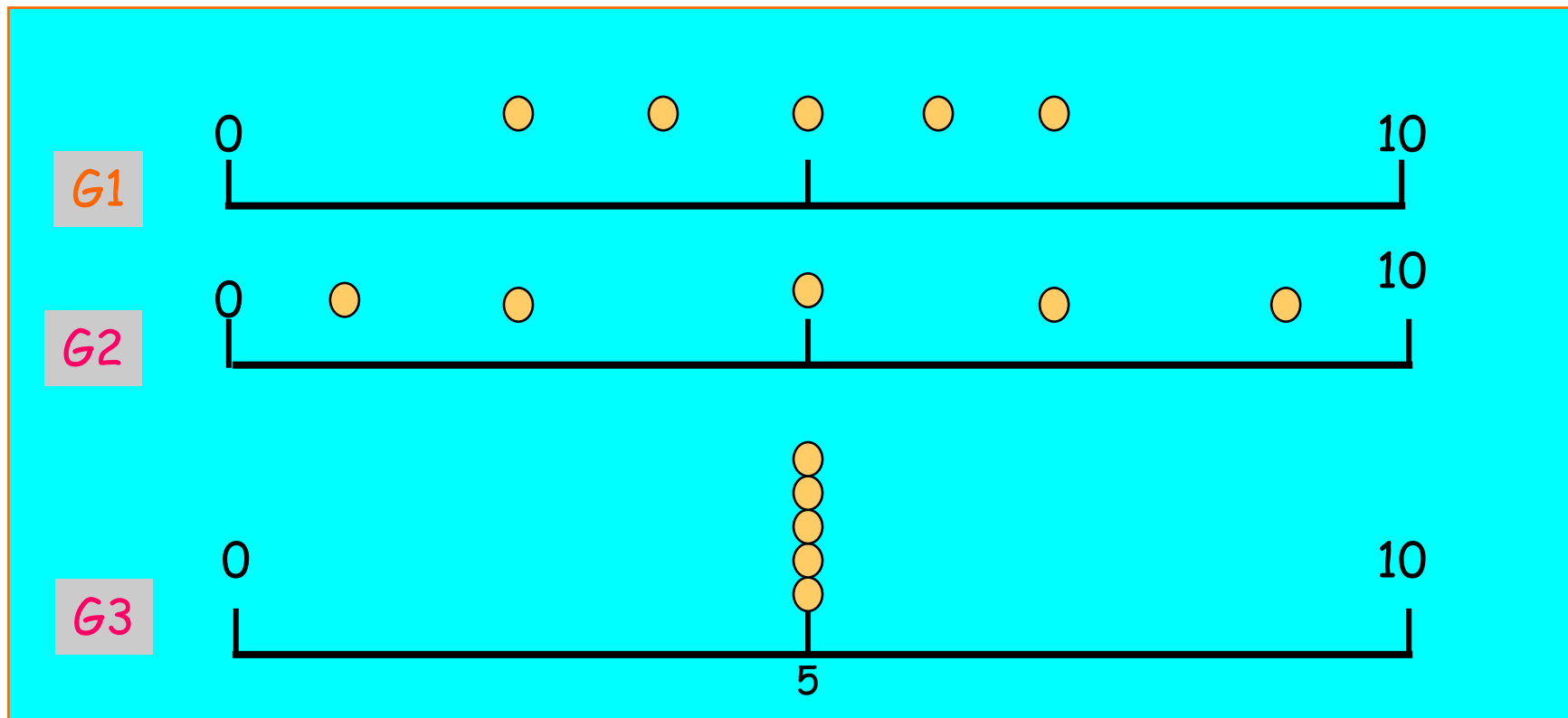
Posição Md : $q(0.5)=8$

Posição de Q1: $q(0.25)=4,5$

Posição de Q3: $q(0.75)=11,25$

Exemplo 2: Considere as notas de um teste de 3 grupos de alunos:

Grupo 1: 3, 4, 5, 6, 7; Grupo 2: 1, 3, 5, 7, 9; e Grupo 3: 5, 5, 5, 5, 5.



Temos : $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 5$ $Md_1 = Md_2 = Md_3 = 5$

Medidas de Dispersão

Finalidade: encontrar um valor que resuma a variabilidade de um conjunto de dados

Amplitude (A): $A = \text{máx} - \text{min}$

Para os grupos anteriores, temos:

Grupo 1, $A=4$

Grupo 2, $A=8$

Grupo 3, $A=0$

Intervalo-Interquartil (d)

É a diferença entre o terceiro quartil e o primeiro quartil, ou seja,

$$d = Q3 - Q1$$

Ex(1): 15,5,3,8,10,2,7,11,12

Q1=4,5 e Q3=11,25

$$d = Q3 - Q1 = 4,9 - 2,05 = 2,85$$

Max, Min, Q1, Q3, Q2: importantes para se ter uma boa ideia da forma dos dados (simétrica ou assimétrica) e construir box-plots

Variância

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Desvio padrão **S**

$$\text{Desvio Padrão : } S = \sqrt{\text{Variância}}$$

Cálculo da variância para o grupo 1:

G1: 3, 4, 5, 6, 7: Vimos que: $\bar{x} = 5$

$$S^2 = \frac{(3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2}{5-1} = \frac{10}{4} = 2,5$$

Desvio padrão $S = \sqrt{2,5} = 1,58$

$$G1 : S^2 = 2,5 \quad S = 1,58$$

$$G2 : S^2 = 10 \quad S = 3,16$$

$$G3 : S^2 = 0 \quad S = 0$$

Coeficiente de Variação (CV)

- É uma medida de dispersão relativa;
- Elimina o efeito da magnitude dos dados;
- Exprime a variabilidade em relação a média
- Útil Comparar duas ou mais variáveis

$$CV = \frac{S}{\bar{X}} \times 100 \%$$

Exemplo 4: Altura e peso de alunos

	Média	Desvio padrão	Coeficiente de variação
Altura	1,143m	0,063m	5,5%
Peso	50Kg	6kg	12%

Conclusão: Com relação as médias, os alunos são, aproximadamente, duas vezes mais dispersos quanto ao peso do que quanto a altura

ORGANIZAÇÃO E REPRESENTAÇÃO DOS DADOS

Uma das formas de organizar e resumir a informação contida em dados observados é por meio de tabela de freqüências e gráficos.

Tabela de freqüência relaciona categorias (ou classes) de valores, juntamente com contagem (ou freqüências) do número de valores que se enquadram em cada categoria ou classe.

1. Variáveis qualitativas: Podemos construir tabela de freqüência que os quantificam por categoria de classificação e sua representação gráfica é mediante gráfico de barras, gráfico setorial ou em forma de pizza.



Exemplo 1: Considere a variável grau de Instrução dos dados da tabela 1. (Variável qualitativa)

Tabela de frequência

Grau de instrução	Contagem	f_i	f_{r_i}	$f_{r_i} \%$
1o Grau		12	0,3333	33,3%
2o Grau	/ / /	18	0,5000	50 %
Superior		6	0,1667	16.7%
total		n=36	1,0000	100%

f_i : **Frequência absoluta** da categoria i (número de indivíduos que pertencem à categoria i)

$f_{r_i} = \frac{f_i}{n}$: **Frequência relativa** da categoria i

$f_{r_i} \% = f_{r_i} * 100\%$: **Frequência relativa percentual** da categoria i

Representação gráfica de variáveis qualitativas

- Gráfico de Barras
- Diagrama circular, de sectores ou em forma de "pizza"

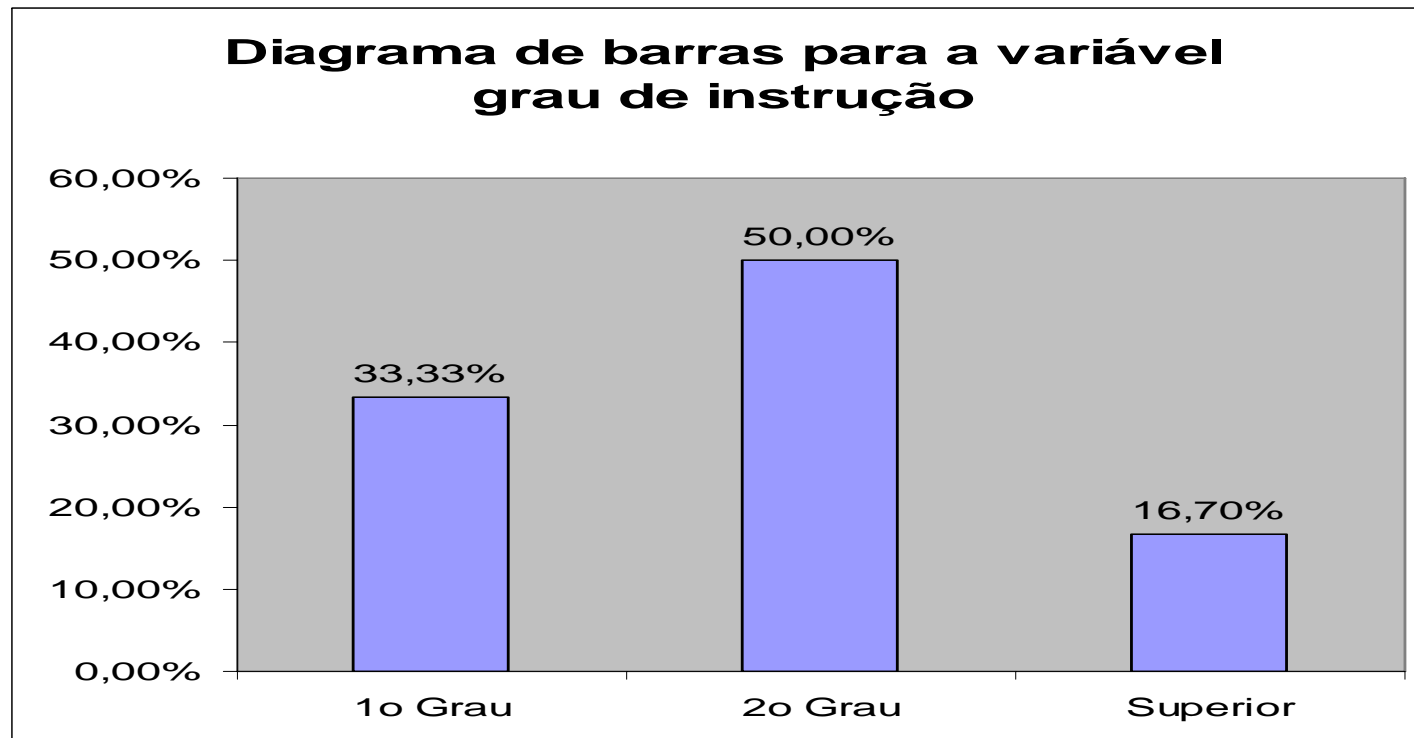


Diagrama circular para a variavel grau de instrução

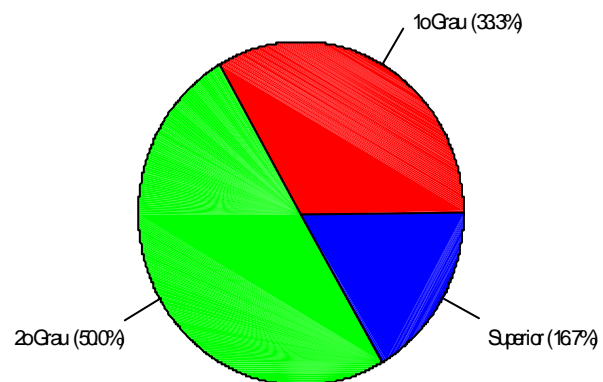
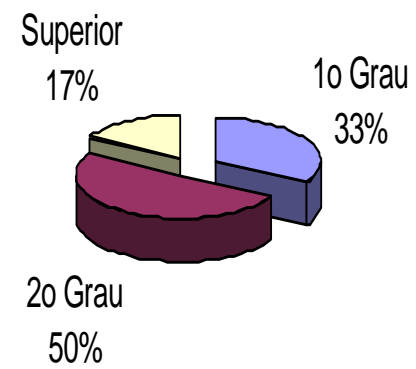


Diagrama circular para a variável grau de instrução



2. Organização e representação de variáveis quantitativas

2.1 Quantitativas discretos: Organizam-se mediante tabelas de frequências e a representação gráfica é mediante gráfico de barras

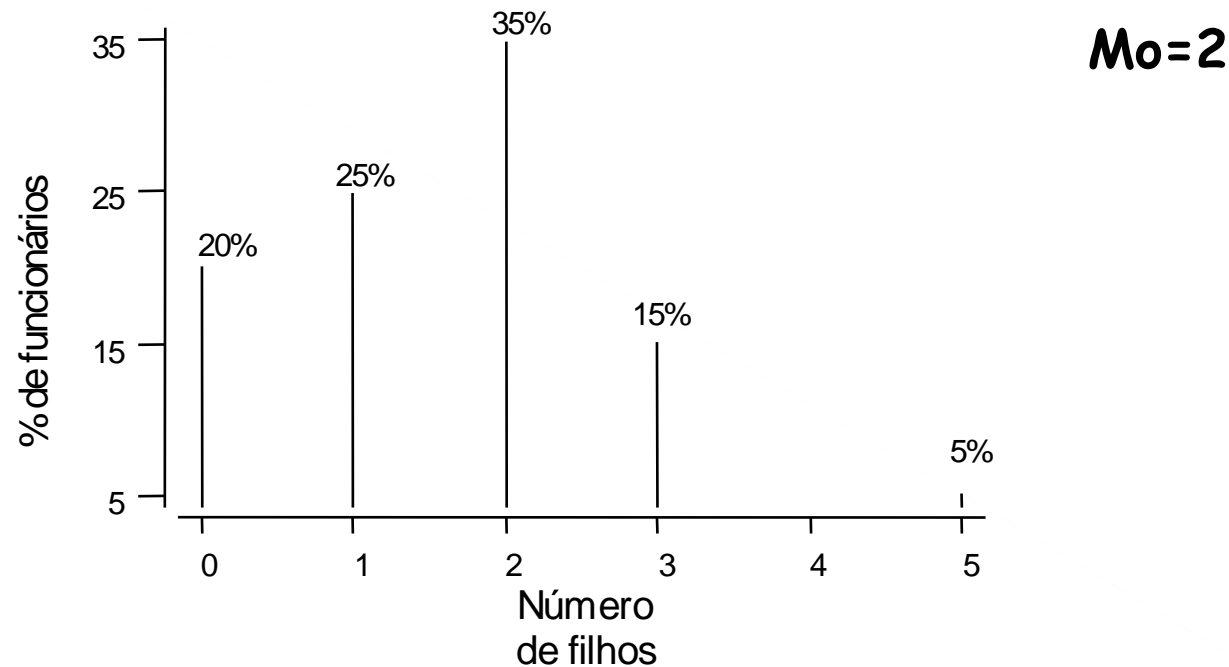
Exemplo: Considere a variável número de filhos dos dados da tabela 1.

Tabela 2.1: Distribuição de frequências de funcionários da empresa, segundo o número de filhos

i	Número de filhos (X_i)	Número de funcionários (f_i)	% de funcionários (f_{ri})
1	0	4	20%
2	1	5	25%
3	2	7	35%
4	3	3	15%
5	5	1	5%
total		20	100%

Observação 1: A partir da tabela 2.1 podemos recuperar as 20 observação da tabela 1.1, ou seja, aqui não temos perda de informação dos dados originais.

Representação gráfica : Diagrama de Barras



Determinação das **medidas de posição e medidas de dispersão** para variáveis quantitativas discretas agrupados em tabela de frequências:

• Média:

$$\bar{X} = \frac{X_1 f_1 + X_2 f_2 + \dots + X_k f_k}{n} = \frac{\sum_{i=1}^k X_i f_i}{n}$$

Exemplo: Considere a tabela 2.1 e determine a média de filhos dos funcionários.

$$\bar{X} = \frac{0 \times 4 + 1 \times 5 + 2 \times 7 + 3 \times 3 + 5 \times 1}{20} = \frac{33}{20} = 1,65$$

• **Mediana:**

Dados ordenados:

0 0 0 0 1 1 1 1 1 2 2 2 2 2 2 2 3 3 3 5 =>

$(20+1)/2=10,5 \Rightarrow$ **Md = (2+2) /2=2**

• Variância:

$$S^2 = \frac{(X_1 - \bar{X})^2 f_1 + (X_2 - \bar{X})^2 f_2 + \dots + (X_k - \bar{X})^2 f_k}{n-1} = \frac{\sum_{i=1}^k (X_i - \bar{X})^2 f_i}{n-1}$$

Cálculo da variância para os dados da tabela 2.1

$$S^2 = \frac{4(0 - 1,65)^2 + 5(1 - 1,65)^2 + 7(2 - 1,65)^2 + 3(3 - 1,65)^2 + (5 - 1,65)^2}{19}$$
$$= \frac{16,3125}{19} = 0,858553$$

Desvio padrão:

$$S = \sqrt{S^2} = \sqrt{0,858553} = 0,927$$

2.2 **Quantitativas contínuas:** Os seus valores podem ser qualquer número real e ainda geralmente existe um grande número de valores diferentes. Como proceder a construir uma tabela de frequência nestes casos?

A alternativa consiste em construir classes ou faixas de valores e contar o número de ocorrências em cada faixa

No caso da variável salário podemos considerar as seguintes faixas de valores: $[4,0; 7,0)$; $[7,0;10,0)$;.....

NOTAÇÃO: $4,0|----7,0$

2.2 Procedimento de construção de tabelas de freqüência para variáveis contínuas:

1. Escolha o número de intervalos de classe (k)
2. Identifique o menor valor (MIN) e o valor máximo (MAX) dos dados.
3. Calcule a amplitude dos dados (A): $A = \text{MAX} - \text{MIN}$
4. Calcule o comprimento de cada intervalo de classe (h): $h = \frac{A}{k}$
5. Arredonde o valor de h de forma que seja obtido um número conveniente.
6. Obtenha os limites de cada intervalo de classe.

PRIMEIRO INTERVALO :

Limite inferior : $LI_1 = \text{MIN}$

Limite superior : $LS_1 = LI_1 + h$

SEGUNDO INTERVALO :

Limite inferior : $LI_2 = LS_1$

Limite superior : $LS_2 = LI_2 + h$

k - ÉSIMO INTERVALO :

Limite inferior : $LI_k = LS_{k-1}$

Limite superior : $LS_k = LI_k + h$

7. Construa uma tabela de freqüências, constituída pelas seguintes colunas:

- Número de ordem de cada intervalo (i)
- Limites de cada intervalo. Os intervalos são fechados á esquerda e aberta à direita: NOTAÇÃO: |----

- **Ponto médio** (ou marca de classe) de cada intervalo de classe:

$$X'_i = \frac{LS_i + LI_i}{2}$$

- Contagem dos dados pertencentes a cada intervalo.
- **Freqüências absolutas** de cada intervalo de classe.
- **Freqüências relativas** de cada intervalo de classe.
- **Freqüências acumuladas absolutas** de cada intervalo de classe.

$$F_i = f_1 + f_2 + \dots + f_i = \sum_{j=1}^i f_j$$

- **Freqüências acumuladas relativa** de cada intervalo de classe.

$$F_{r_i} = f_{r_1} + f_{r_2} + \dots + f_{r_i} = \sum_{j=1}^i f_{r_j}; \text{ ou } F_{r_i} = \frac{F_i}{n}$$

Exemplo: Considere a variável salário da empresa comercializadora de produtos de informática.



Procedimento:

1. Considere $k=5$.
2. $MIN=4$; $MAX=23,30$.
3. $A=MAX-MIN=23,30-4=19,30$
4. $h=19,3/5=3,86$
5. $h\approx 3,9$
6. Cálculo dos limites de cada intervalo:

PRIMEIRO INTERVALO

$$LI_1 = 4$$

$$LS_1 = 4 + 3,9 = 7,9$$

SEGUNDO INTERVALO

$$LI_2 = 7,9$$

$$LS_2 = 7,9 + 3,9 = 11,8$$

Os demais limites dos intervalos foram gerados seguindo o procedimento anterior.

• **Ponto médio:** $X'_1 = \frac{(4+7,9)}{2} = 5,95$; $X'_2 = \frac{(7,9+11,8)}{2} = 9,85$

De forma similar obtém-se os outros pontos médios.

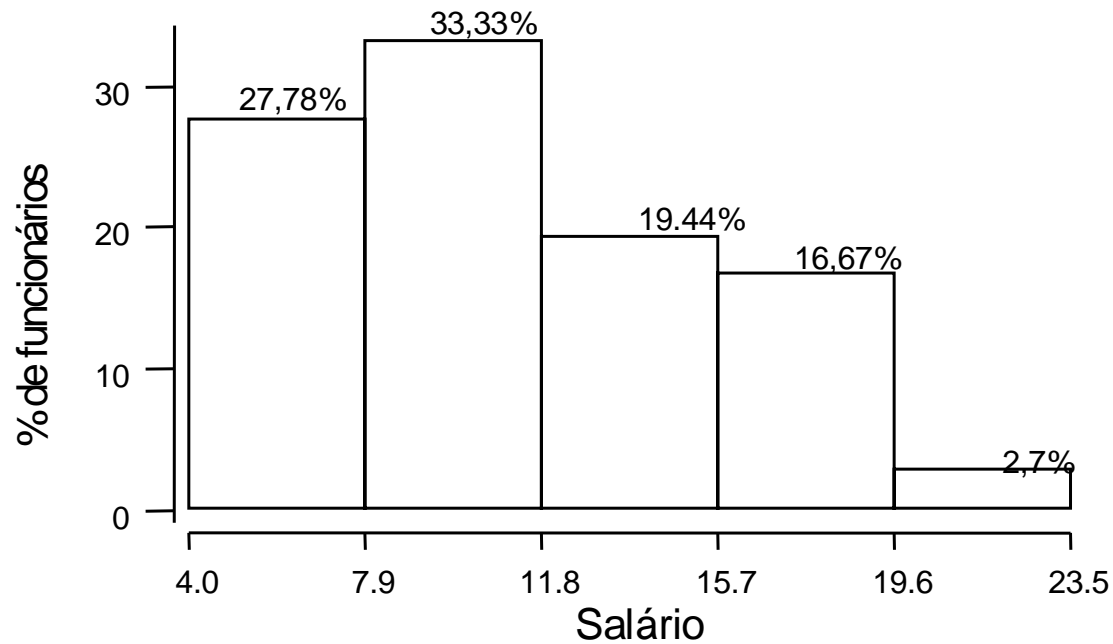
Tabela 2.2: Distribuição de freqüências da variável salário. 

<i>i</i>	<i>Intervalos de classe</i>	<i>Ponto médio (X'ᵢ)</i>	<i>Freqüência Absoluta (fᵢ)</i>	<i>Freqüência Relativa (fᵣᵢ)</i>	<i>Freqüência Acumulada Absoluta (Fᵢ)</i>	<i>Freqüência Acumulada Relativa (Fᵣᵢ)</i>
1	4,0 -- 7,9	5,95	10	0,277778	10	0,277778
2	7,9 -- 11,8	9,85	12	0,333333	22	0,611111
3	11,8 -- 15,7	13,75	7	0,194444	29	0,805556
4	15,7 -- 19,6	17,65	6	0,166667	35	0,972222
5	19,6 -- 23,5	21,55	1	0,027778	36	1
	Total		36	1,000000		

Nesta organização de dados, temos perda de informação dos dados originais

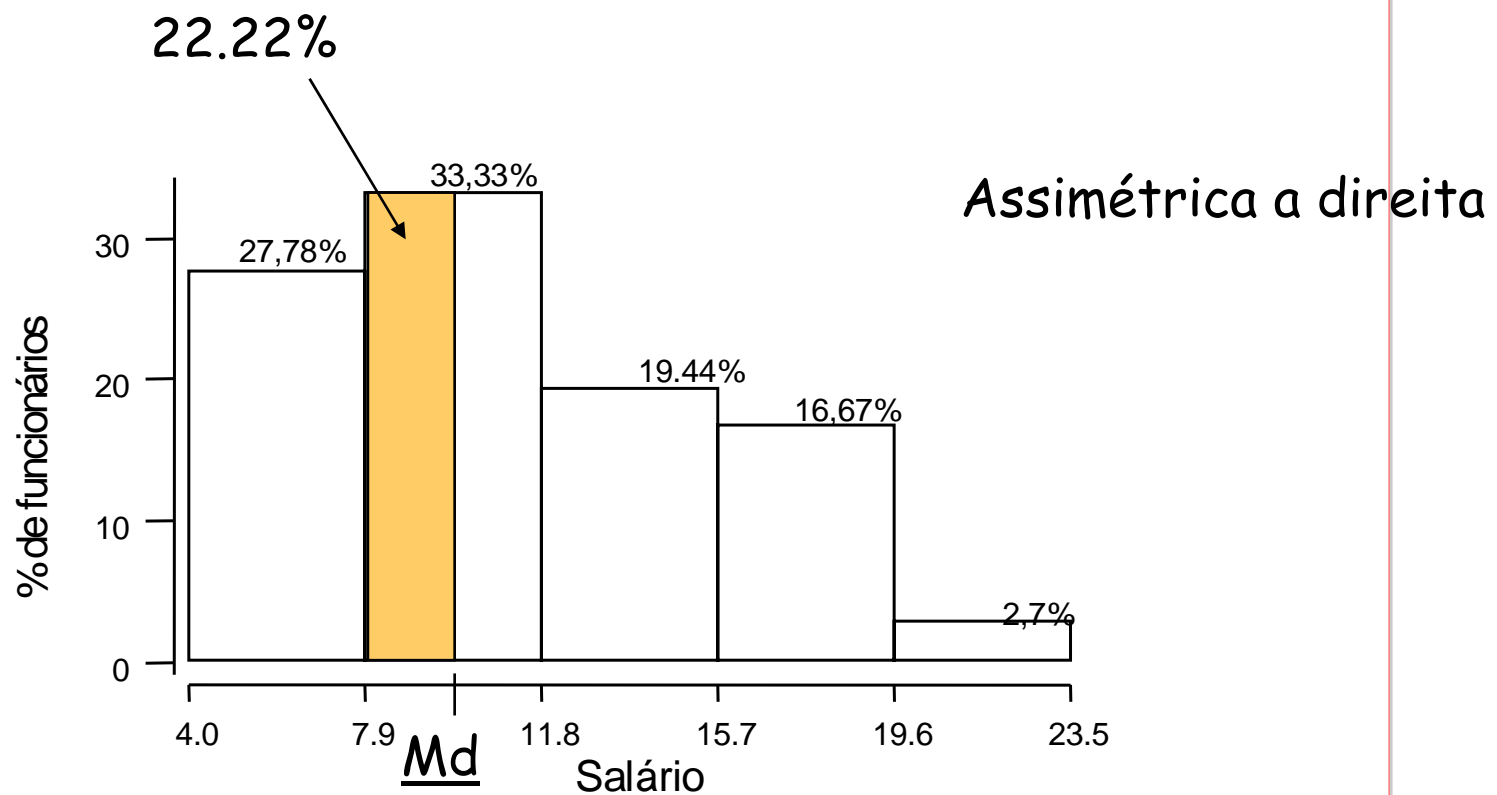
Representação gráfica:

- Histograma de freqüências relativas (em %) para a variável salário

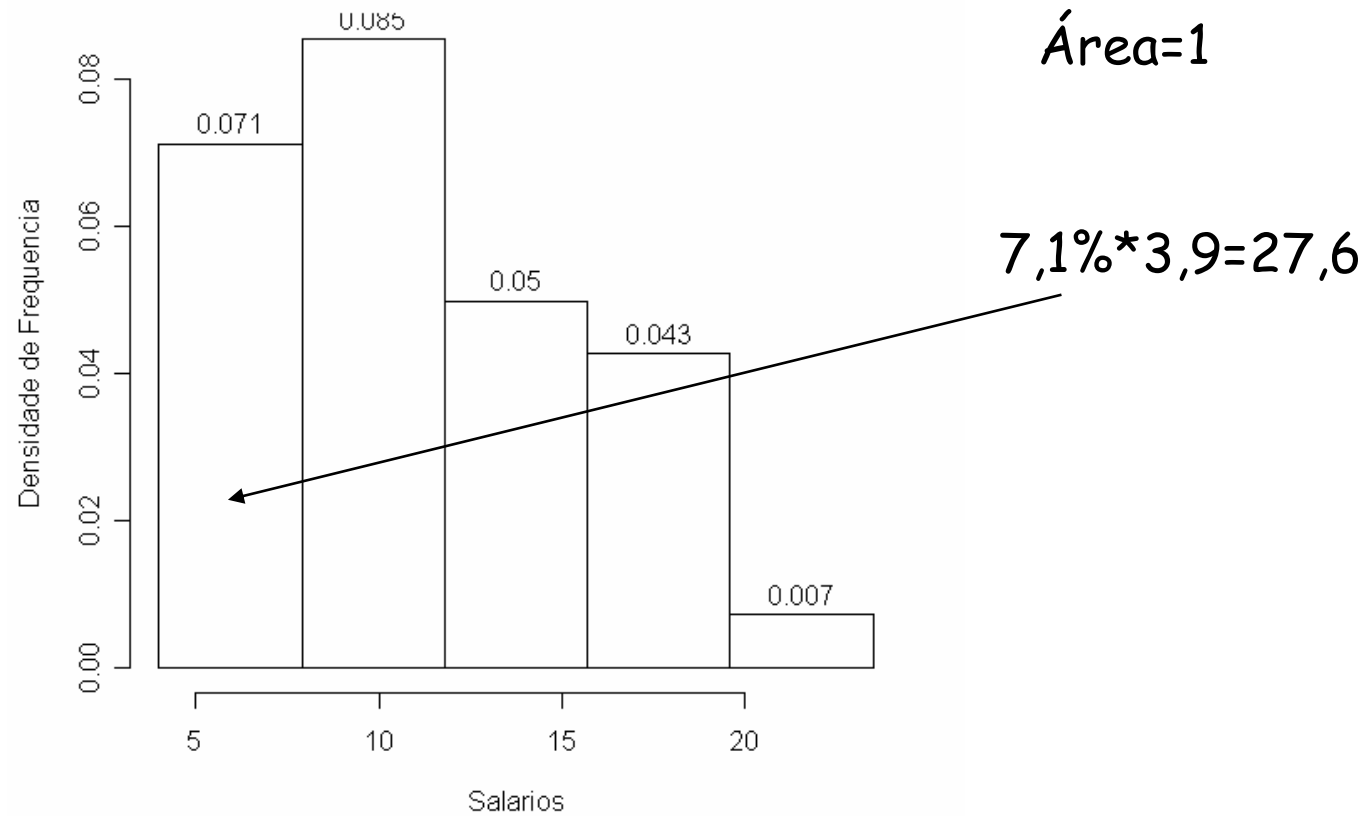


Útil para encontrar os percentis: Exemplo Q2 ou Md

$$\frac{11,8 - 7,9}{33,33\%} = \frac{Md - 7,9}{22,22} \Rightarrow Md = 10,5$$

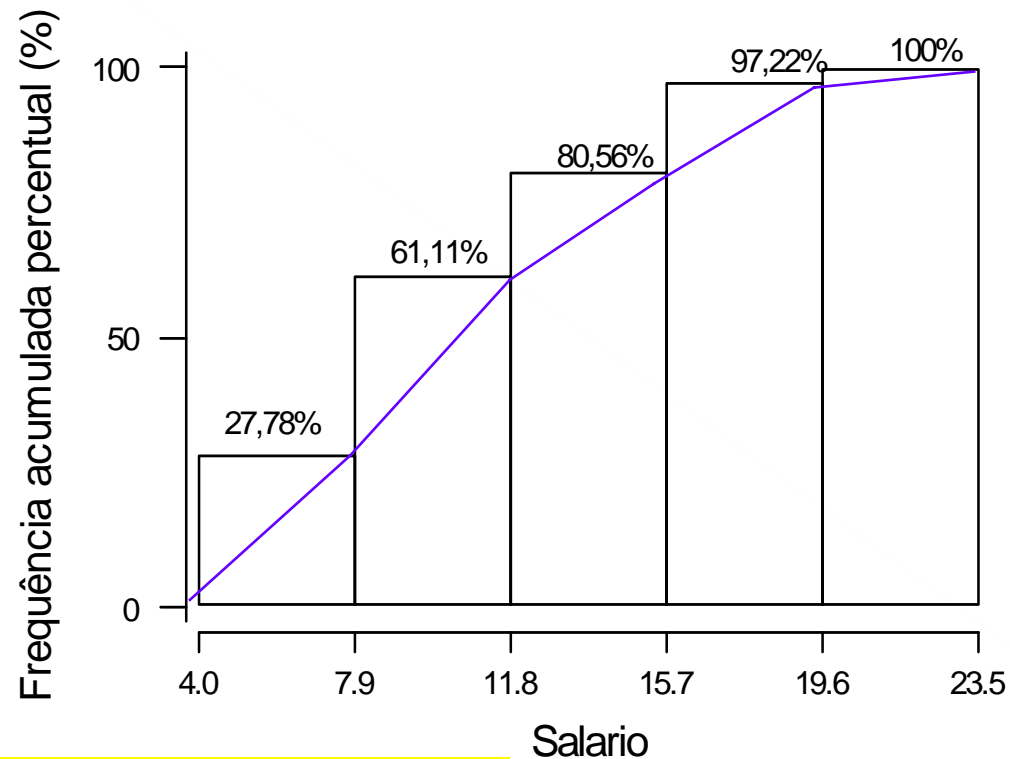


. Histograma usando densidade de frequência (mais comum!)



- Histograma de frequência acumulada relativa (em %)

61% dos empregados tem salário inferior a 12 salários mínimos



19% possuem salário superior a 16 salários mínimos

Gráfico de Ramo e Folhas: Variável salário

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

- Valores concentrados entre 4 e 19
- Leve assimetria na direção dos valores grandes(assimétrica à direita)
- Destaque do valor 23.30

Medidas de posição e medidas de dispersão para variáveis contínuas agrupadas em tabela de freqüências.

• Média:

$$\bar{X} = \frac{X'_1 f_1 + X'_2 f_2 + \dots + X'_k f_k}{n} = \frac{\sum_{i=1}^k X'_i f_i}{n}$$



Exemplo: Considere a tabela 2.2

$$\begin{aligned}\bar{X} &= \frac{5,95 \times 10 + 9,85 \times 12 + 13,75 \times 7 + 17,65 \times 6 + 21,55 \times 1}{36} \\ &= \frac{401,4}{36} = 11,15\end{aligned}$$

Se calculamos a média para dados não agrupados apresentadas anteriormente resulta:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_{36}}{36} = \frac{4 + 4,36 + \dots + 23,30}{36} = 11,122$$

Este resultado difere do valor obtido anteriormente. Porque?

• **Moda (mo):** $mo = LI_i + \left(\frac{d_1}{d_1 + d_2} \right) \times h$


i : Classe modal (é aquela classe que tem maior frequência absoluta (f_i))

LI_i : é o limite inferior da classe modal.

$$d_1 = f_i - f_{i-1}$$

$$d_2 = f_i - f_{i+1}$$

h : comprimento do intervalo de classe.

Exemplo: Considere a tabela 2.2. 

Já que, $f_2 = 12 > f_j \quad j \neq 2 \Rightarrow i = 2$, é a classe modal

$$mo = LI_2 + \left(\frac{d_1}{d_1 + d_2} \right) \times h = 7,9 + \left(\frac{12 - 10}{(12 - 10) + (12 - 7)} \right) \times 3,9 = 9,014$$

- **Mediana (Md)**

$$Md = LI_i + \left(\frac{0,5n - F_{i-1}}{f_i} \right) \times h$$

i : é a classe médiana (é o intervalo de classe onde a coluna dos F_i na TDF superou o 50% dos dados)

LI_i : Limite inferior da classe mediana.

F_{i-1} : é a frequência acumulada absoluta da classe anterior a classe mediana

f_i : frequência absoluta da classe mediana.

h : comprimento do intervalo de classe.

Exemplo: Considere a tabela 2.2



Já que, $F_2 = 22 > n / 2 \Rightarrow i=2$, é a classe mediana

$$Md = LI_2 + \left(\frac{0,5n - F_1}{f_1} \right) \times h = 7,9 + \left(\frac{18 - 10}{12} \right) \times 3,9 = 8,55$$

• Variância:

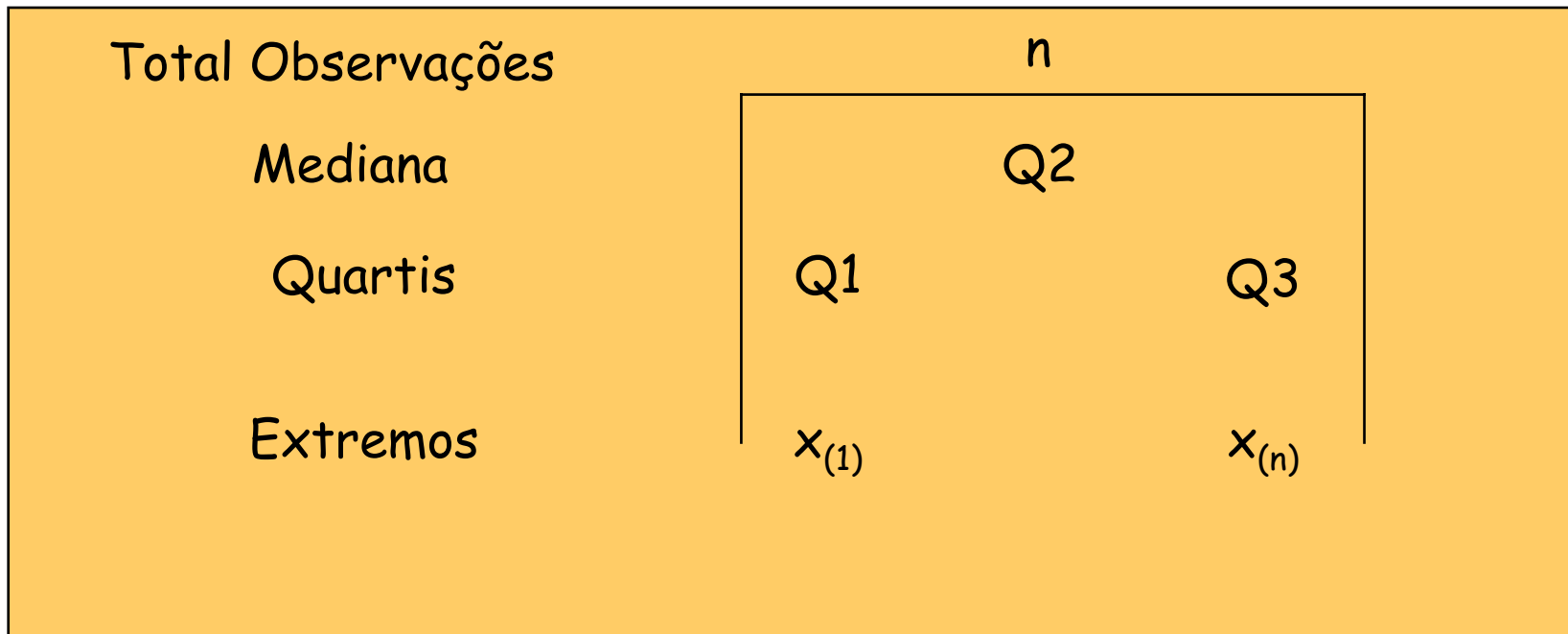
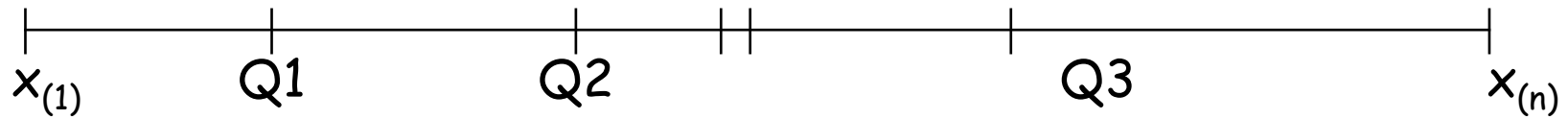
$$S^2 = \frac{\sum_{i=1}^k f_i (X'_i - \bar{X})^2}{n-1}$$

Exemplo: Considere a tabela 2.2. Vimos que $\bar{X} = 11,15$

<i>i</i>	<i>Intervalos de classe</i>	X'_i	f_i	$f_i (X'_i - \bar{X})^2$
1	4,0 -- 7,9	5,95	10	270,40
2	7,9 -- 11,8	9,85	12	20,28
3	11,8 -- 15,7	13,75	7	47,32
4	15,7 -- 19,6	17,65	6	253,50
5	19,6 -- 23,5	21,55	1	108,16
	Total		36	699,66

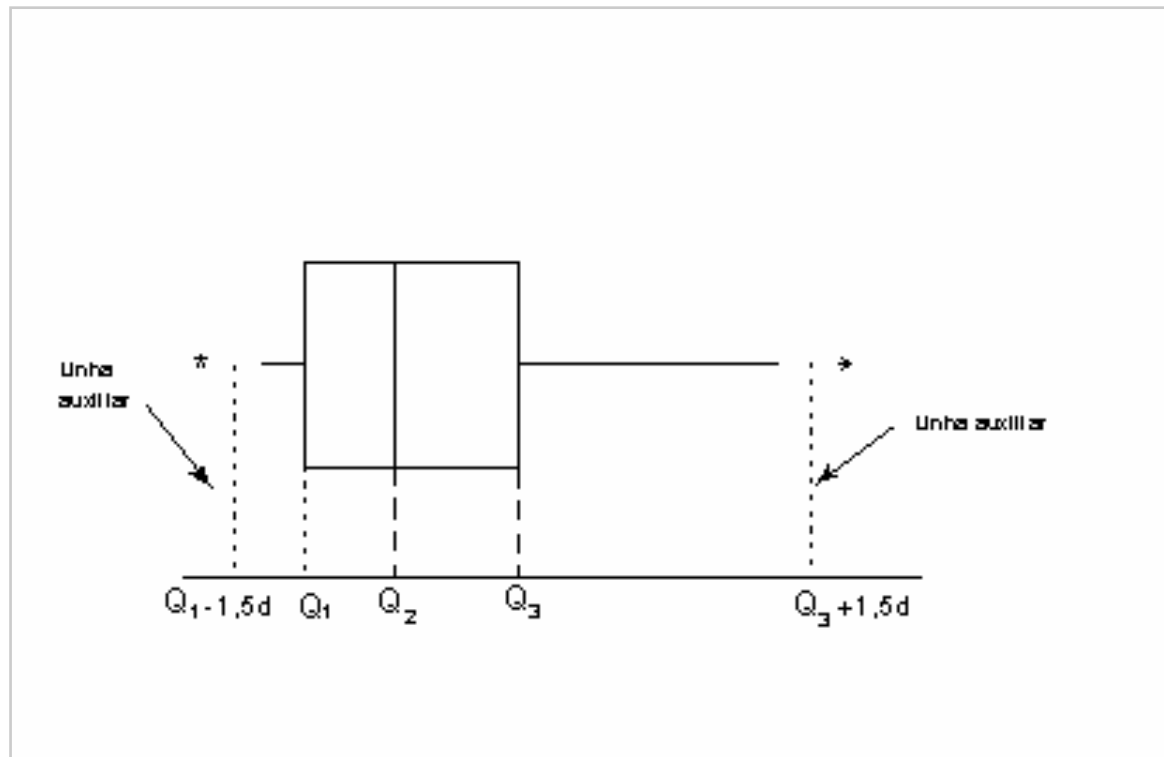
$$S^2 = \frac{\sum_{i=1}^5 f_i (X'_i - \bar{X})^2}{36-1} = \frac{699,66}{35} = 19,99029 \Rightarrow S = 4,47105 (\text{Desvio Padrão})$$

Esquema dos cinco números



Boxplot

O BOXPLOT representa os dados através de um retângulo construído com os quartis e fornece informação sobre valores extremos. (veja o esquema embaixo)



Exemplo de construção de um Boxplot. Com a finalidade de aumentar o peso (em Kg) um regime alimentar foi aplicado em 12 pessoas. Os resultados (ordenados) foram:

-0,7 2,5 3,0 3,6 4,6 5,3 5,9 6,0 6,2 6,3 7,8 11,2.

Calculando as medidas temos:

Mediana (md ou Q2) = 5,6kg

1°.quartil (Q1) = 3,3kg

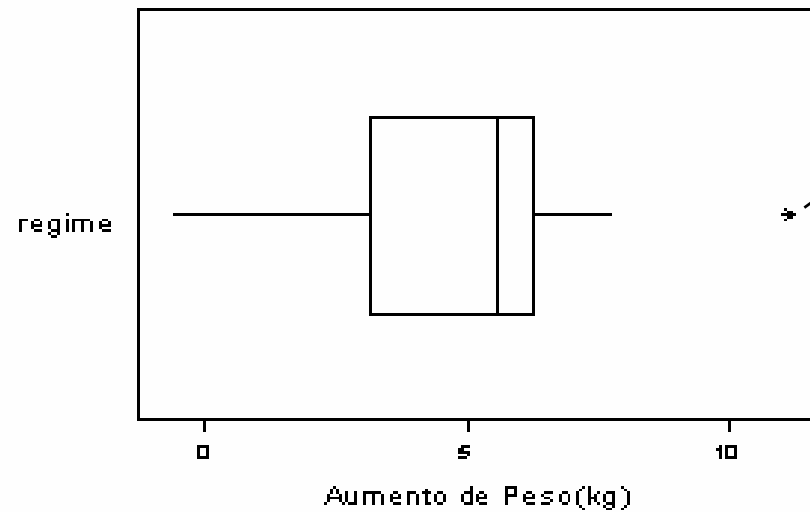
3°.quartil (Q3) = 6,25kg

$d = \text{intervalo interquartil} = Q3 - Q1 = 2,95\text{kg}$

Logo as linhas auxiliares correspondem aos pontos:

$Q1 - 1,5d = -1,25\text{kg}$

$Q3 + 1,5d = 10,675\text{kg}$



11.2
Observação
exterior
(discrepante
ou atípica)

Exemplo: Considere os dados da tabela 1.1, o boxplot para variável salário por educação e região de procedência dos funcionários da empresa.

