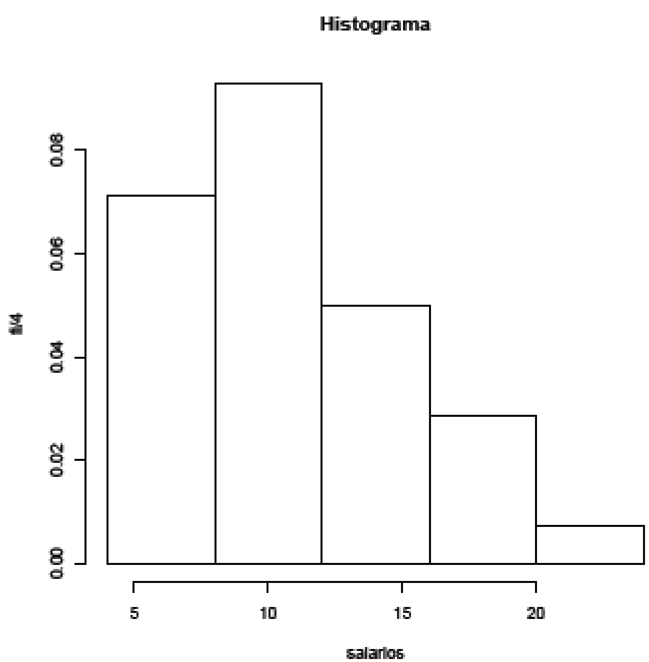
ESTATÍSTICA DESCRITIVA

- Frequência absoluta (f_a)
- Frequência relativa (f_r)
- Histograma
- Ramo e folha



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		

Medidas de Posição

- Média (Me)
- Mediana (Md)
- Moda (Mo)

Medidas de Dispersão

- Desvio Médio

$$DM(X) = \sum_{i=1}^{n} \frac{|x_i - \bar{x}|}{n}$$

- Variância

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

- Desvio Padrão

$$DP(X) = \sqrt{Var(X)}$$

Medidas Complementares

- Extremos: E_s e E_i
- Quartis (Q) ou Juntas (J)

- Intervalo interquartil

$$d_J = J_3 - J_1 = Q_3 - Q_1$$

- Dispersão Inferior

$$J_2 - E_i$$

- Dispersão superior

$$E_s - J_2$$

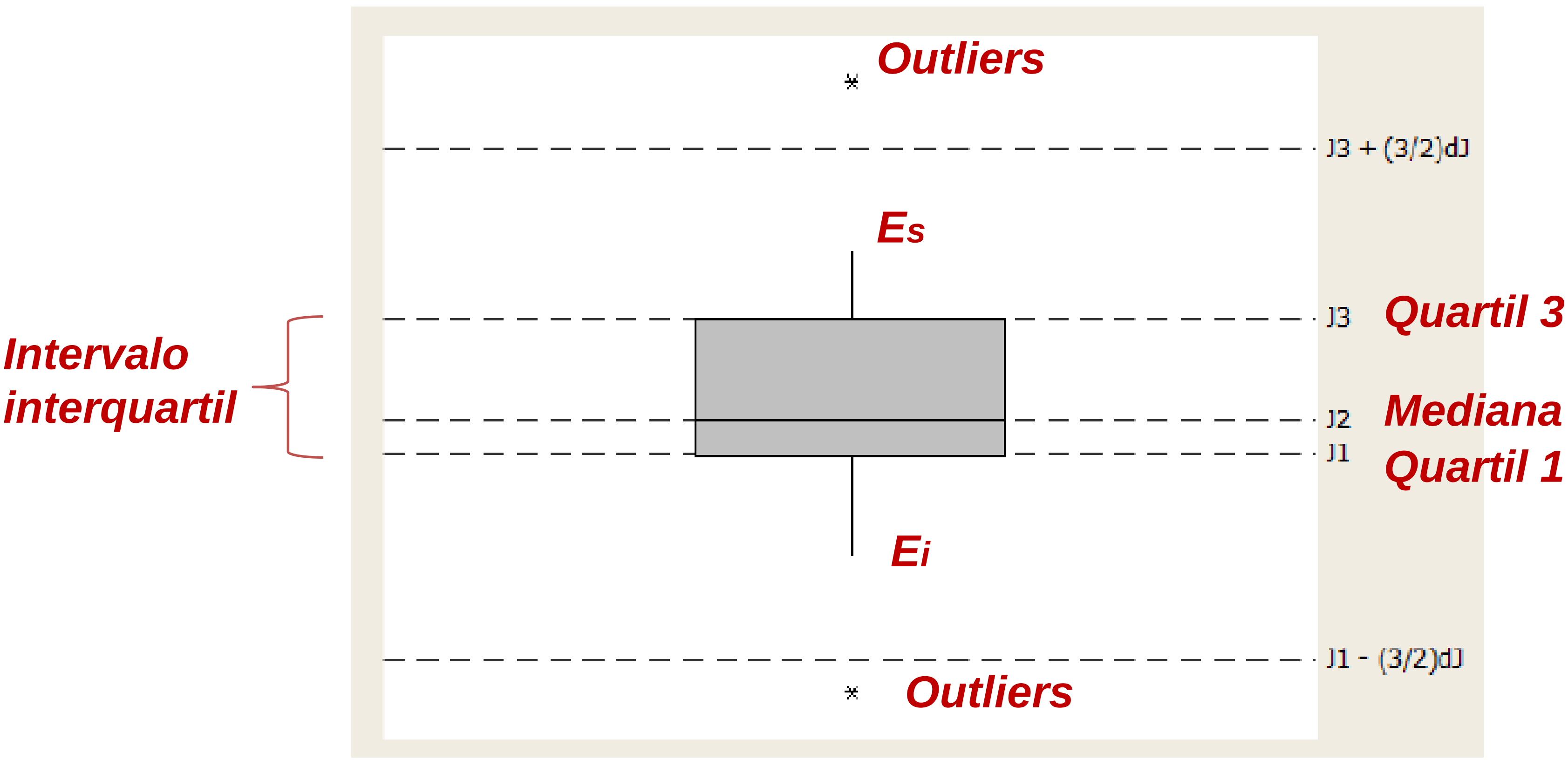
Esquema dos cinco números

	n	
Md	Mediana	
Q	Q1	Q3
E	Ei	Es

- Box Plot

Outliers (observação discrepante):

- observações menores que $J_1 - \frac{3}{2}d_J$
- observações maiores que $J_3 + \frac{3}{2}d_J$


Outliers
J3 + (3/2)dJ
Es
J3
Quartil 3
Intervalo interquartil
J2
Mediana
J1
Quartil 1
Ei
J1 - (3/2)dJ
Outliers

Exemplo 1

Calculando os quartis para a amostra:

{ 1, 1, 2, 3, 5, 5, 6, 7, 9, 9, 10, 13 }

- **Q2 = Md = (5+6)/2 = 5,5**
- O quartil 1 será a mediana da série à esquerda de Md : { 1, 1, 2, 3, 5, 5 }

Q1 = (2+3)/2 = 2,5

- O quartil 3 será a mediana da série à direita de Md : {6, 7, 9, 9, 10, 13 }

Q3 = (9+9)/2 = 9

Variáveis Multidimensionais

- Correlação

$$Corr(X,Y) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{DP(X)}\right)\left(\frac{y_i - \overline{y}}{DP(Y)}\right)$$

Propriedades:

- $-1 \leq Corr(X,Y) \leq 1$
- ***Corr*** *próxima de 1: X e Y estão positivamente associados*
- ***Corr*** *próxima de -1: X e Y estão negativamente associados*

Exemplo 2

Agente	Anos de Serviço (X)	N^o de Clientes (Y)
A	2	48
B	4	56
C	5	64
D	6	60
E	8	72
Total	25	300

Agente	X	Y	$x_i - \bar{x}$	$y_i - \bar{y}$	$\left(\frac{x_i - \bar{x}}{DP(X)}\right)$	$\left(\frac{y_i - \bar{y}}{DP(Y)}\right)$	$\left(\frac{x_i - \bar{x}}{DP(X)}\right)\left(\frac{y_i - \bar{y}}{DP(Y)}\right)$
A	2	48	-3	-12	-1.5	-1.5	2.25
B	4	56	-1	-4	-0.5	-0.5	0.25
C	5	64	0	4	0	0.5	0
D	6	60	1	0	0.5	0	0
E	8	72	3	12	1.5	1.5	2.25

Portanto: $Corr(X, Y) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{DP(X)}\right)\left(\frac{y_i - \bar{y}}{DP(Y)}\right) = \frac{4.75}{5} = 0.95$

Probabilidade

- Interseção de eventos

$$\boxed{A \cap B = \text{evento em que } A \text{ e } B \text{ ocorrem simultâneamente}}$$

- União de eventos

$$\boxed{A \cup B = \text{evento em que } A \text{ ou } B \text{ ocorrem}}$$

- Evento complementar
  - como $P(A) + P(B) = 1$, então $P(B) = 1 - P(A)$
  - B é denotado por $B = A^C$
- Propriedade fundamental

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

Exemplo 3

	Masculino (Ma)	Feminino (Fe)	Total
Mat. Pura (M)	70	40	110
Mat. Aplicada (A)	15	15	30
Estatística (E)	10	20	30
Computação (C)	20	10	30
Total	115	85	200

- Escolhendo um aluno ao acaso, definem-se os seguintes eventos:
  - M: estudante da Matemática Pura
  - A: estudante da Matemática Aplicada
  - E: estudante da Estatística
  - C: estudante da Computação
  - Ma: sexo Masculino
  - Fe: sexo Feminino

- $P(M) = \frac{110}{200} = 0.550$
- $P(A) = \frac{30}{200} = 0.150$
- $P(E) = \frac{30}{200} = 0.150$
- $P(C) = \frac{30}{200} = 0.150$
- $P(Ma) = \frac{115}{200} = 0.575$
- $P(Fe) = \frac{85}{200} = 0.425$

- Utilizando o último exemplo, vamos definir como evento (I), escolher ao acaso um aluno e ele ser estudante de estatística do sexo masculino, simultâneamente.
- $I = E \cap Ma$, o evento I é uma interseção dos eventos E e Ma.
- $P(E \cap Ma) = \frac{10}{200} = 0.05$

- Definimos agora como evento (U), escolher ao acaso um aluno e ele ser estudante de estatística ou sexo masculino.
- $U = E \cup Ma$, o evento U é uma união dos eventos E e Ma.
- $P(E \cup Ma) = P(E) + P(Ma) - P(E \cap Ma)$

  $P(E) = \frac{10+20}{200} = 0.150$

  $P(Ma) = \frac{70+15+10+20}{200} = 0.575$

  $P(E \cap MA) = \frac{10}{200} = 0.050$

  Então: $P(E \cup Ma) = 0.150 + 0.575 - 0.050 = 0.675$

Problemas de contagem

1. Com ordem e repetição → N^n
2. Com ordem e sem repetição → N x N-1 x N-2 x...
3. Sem ordem e com repetição → Combinação + N
4. Sem ordem e sem repetição → Combinação
5. Só repetição → N

- Com ordem: (1,2) ≠ (2,1) → mais possibilidades
- Sem ordem: (1,2) = (2,1) → menos possibilidades
- Com repetição: (1,1), (2,2), ...
- Sem repetição: (1,2), (2,1), ...

Probabilidade condicional

- Assumindo a potencial ocorrência de A, qual a probabilidade de B acontecer?

  $P(B|A) = P(\text{B ocorrer dado que A ocorre})$

Teorema de Bayes

$$\boxed{P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(A \cap B) = P(A)\,P(B|A)}$$

Independência

- Em geral, dois eventos A e B são considerados independentes, se:
  - $P(A|B) = P(A)$
  - ou $P(A \cap B) = P(A)\,P(B)$

Exemplo 4

	Masculino (Ma)	Feminino (Fe)	Total
Mat. Pura (M)	70	40	110
Mat. Aplicada (A)	15	15	30
Estatística (E)	10	20	30
Computação (C)	20	10	30
Total	115	85	200

$$P(E \cap Fe) = \frac{20}{200}$$

$$P(E) = \frac{30}{200}$$

$$\Rightarrow P(Fe|E) = \frac{P(E \cap Fe)}{P(E)} = \frac{(20/200)}{(30/200)} = \frac{20}{200} \times \frac{200}{30} = \frac{20}{30} = \frac{2}{3}$$