

ESTADÍSTICA

ÁREAS DE ATUAÇÃO

- Política
- Marketing
- Recursos Humanos
- Engenharia
- Medicina
- Agricultura

COM DEFINIR ESTATÍSTICA?

- Parte da matemática em que se investigam os processos de obtenção, organização e análise de dados sobre uma população ou sobre uma coleção de seres quaisquer, e os métodos de tirar conclusões e fazer ilações ou predições com base nesses dados.
- Qualquer parâmetro de uma amostra, como, p. ex., a sua média, o seu desvio-padrão, a sua variância.

- Conjunto de elementos numéricos respeitantes a um fato social.
- Representação e explicação sistemática, por observações quantitativas de massa, dos acontecimentos e das leis da vida social que deles se podem deduzir.
- Método que objetiva o estudo dos fenômenos de massa, i. e., os que dependem de uma multiplicidade de causas, e tem por fim representar, sob forma analítica ou gráfica, as tendências características limites desses fenômenos.

DEFINIÇÕES MAIS INTERESSANTES

- Ciência encarregada de toma decisões, face à incertezas
- Ciência encarregada de coletar, classificar e interpretar fatos numéricos
- Estudo de processos que apresentam variabilidade, com o objetivo de determinar suas grandezas, causas e consequências

PRELIMINARES DA ANÁLISE ESTATÍSTICA

- Obtenção dos dados:
 - Definição dos objetivos
 - Propósitos
 - Confiabilidade das Medidas
 - Registro das observações

DESCRIÇÃO DOS DADOS

- Planejamento do experimento
- Coleta dos dados

DEFINIÇÕES

- População: Conjunto de todas as medidas de interesse.
- Amostra: Sub-Conjunto da população usado para obter(Estimar), características da população.
- Estatística Descritiva: Explorar as informações da amostra, sobre a população.
- Inferência: Extrair informações da amostra

RESUMINDO

- Etapas da Análise Estatística
 - Definição do Problema
 - Planejamento do Experimento
 - Coleta de dados
 - Análise, Inferência
 - Confiabilidade

APRESENTAÇÃO DOS DADOS

- Métodos Gráficos:
 - Análise de Pareto:
 - **Origem: Economia- V.Pareto e M.C. Lorens**
 - **Finalidade: Classificação:**
 - » Poucos mas vitais
 - » Muitos mas triviais

- **Definição do Problema**
 - » **Itens Defeituosos**
 - » **Custos**
 - » **Ocorrência de Acidentes**
- **Classificação dos Dados**
 - » **Tipo de defeitos**
 - » **Processo**
 - » **Máquina**
 - » **Operário**

CONSTRUÇÃO DIAGRAMA DE PARETO

- Tabela Contendo:
 - Tipos de defeitos
 - Contagem
 - Total

EXEMPLO DEFEITOS EM SOM AUTOMOTIVO

Defeitos	Contagem(frequências)- f_i
Sintonia c/ folga	253
Parafuso solto	69
Display difuso	146
Material estranho	29
Outros	52
Total	549

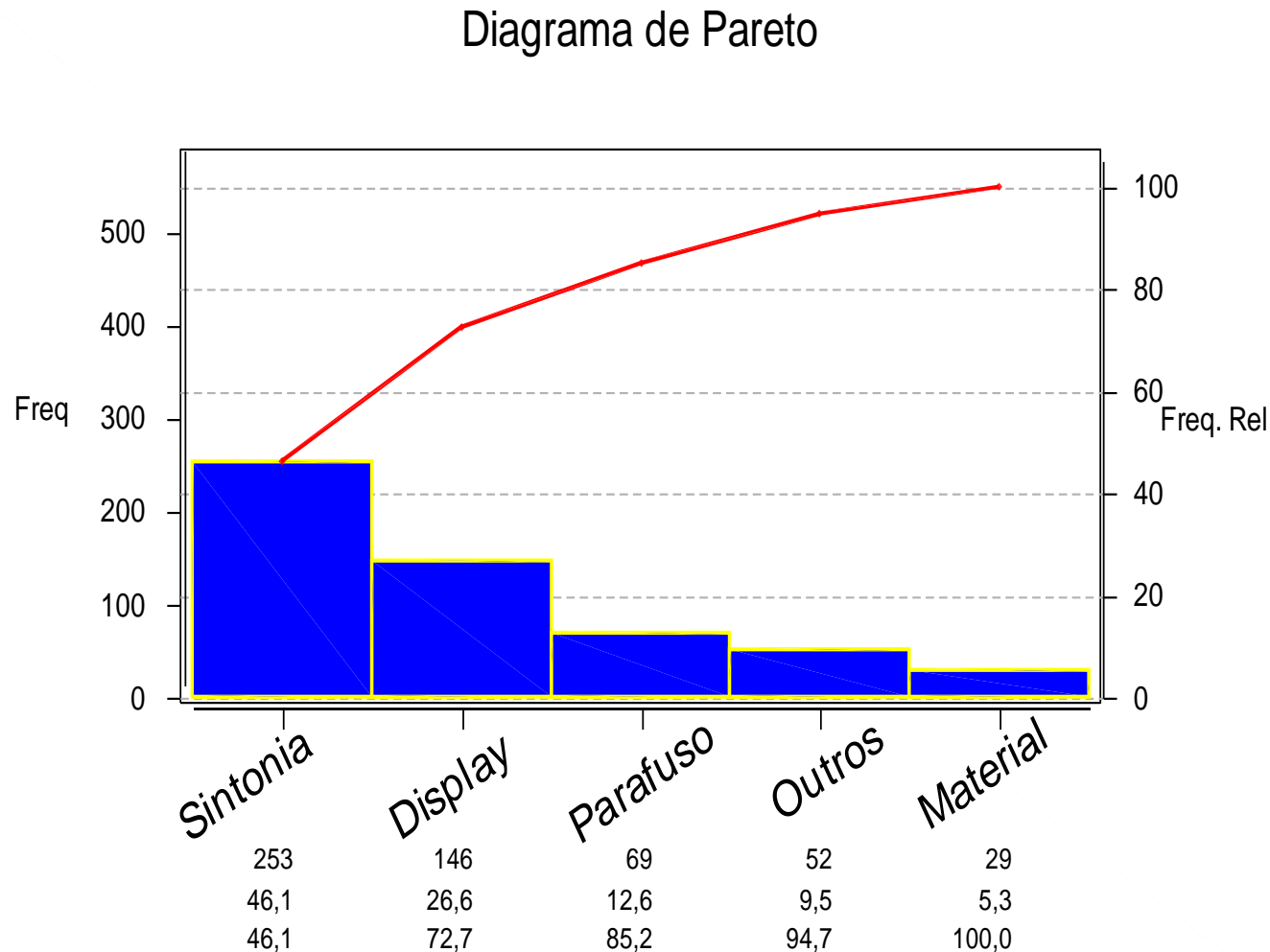
Rearranjar itens em ordem
decrecente de frequência,
calculando as frequências
relativas (f_i/n)

Defeitos	Contagem(frequências)-	
	f_i	f_i / n
Sintonia c/ folga	253	46,1%
Display difuso	146	26,6%
Parafuso solto	69	12,6%
Outros	52	9,4%
Material estranho	29	5,3%
Total	549	100%

Construir Diagrama de Barra

- Eixo vertical esquerdo f_i
- Eixo vertical direito freq. Relativa
- Unir frequências acumuladas

Construir Diagrama de Barras



ANÁLISE

Defeitos

	f_i	custo p/ unid.	total
Sintonia c/ folga	253	0,25	63,25
Display difuso	146	0,60	87,60
Parafuso solto	69	0,10	6,90
Outros	52	0,25	13,00
Material estranho	29	0,05	1,45
Total 549	100%		

CLASSIFICAÇÕES

- Segundo Fenômeno
 - Qualidade
 - Custo
 - Entrega, etc
- Segundo Causas
 - Operador, Máquina, Matéria Prima, etc

TIPOS DE VARIÁVEIS

- Quantitativa: Pode ser medida
 - » **Contínuas**
 - » **Discretas**
- Qualitativa: Definida por Categorias

Métodos Gráficos

– Diagrama por pontos (Dot-Plot)

- **Peso de tumores, retirados de 57 pacientes**

1,92	1,78	1,18	0,76	0,84	1,01	0,79	0,90
2,23	0,76	0,62	0,65	0,67	0,70	1,24	1,84
1,21	0,70	2,09	1,44	1,01	1,18	0,79	0,87
0,79	0,70	1,27	0,33	1,61	1,44	0,33	0,90
1,38	1,07	1,18	0,76	0,87	1,41	1,07	0,59
0,45	0,67	0,95	1,33	0,65	0,62	1,21	0,76
1,38	0,79	0,65	0,53	1,30	0,84	1,21	1,38
0,33							.

Gráfico por Pontos

Peso de Tumores



Métodos Gráficos

- Histogramas
 - Distribuições de Freqüências
 - **Conjunto de Observações Variável Contínua**
 - **Amplitude » $R = \text{Máx} - \text{Min}$**
 - **Classes » $C_i = \text{Sub-Intervalos}, i=1,2,3,\dots,k$**
 - **Amplitude da classe $a_i=R/k$**
 - **Freqüência da classe i » $f_i=\text{no. Obs em } C_i$**
 - **Freqüência Relativa de C_i » $f_{ri}= f_i/n, n =\text{total de obs.}$**
 - **Freqüência Acumulada de C_i**

$$F_i = \sum_{j=1} f_j$$

Histogramas

- **Frequência Acumulada Relativa da classe C_i :**
 - » $F_{ri} = F_i/n$
- **Ponto Médio da classe C_i :**
 - » $P_i = [\text{Lim. Sup} + \text{Lim. Inf.}]/2$

Regra Prática para determinar o número de classes, k:

- **k deve estar entre 5 e 20**
- **aproximação inicial $k = 1 + 3,322 \log(n)$**
- **amplitudes das classe:**
 - » **0,01 0,02 0,05**
 - » **0,1 0,2 0,5**
 - » **1 2 5**

Exemplo Tumores

- Distribuição de Frequências:
 - Encontrar R: $n = 57$, $\text{Max} = 2,23$, $\text{Min} = 0,33$
$$R = 2,23 - 0,33 = 1,9$$
 - Número de classes k
 - $1 + 3,322\log(57) = 6,83 \gg k = 6$, ou $k = 7$
 - $k = 6$ ---- $a = 1,9/6 = 0,31$ ----- $a = 0,35$
 - $k = 7$ ---- $a = 1,96/7 = 0,27$ ---- $a = 0,3$

Exemplo-1

- $k = 6$

<u>i</u>	Intervalos	Cont.	<u>f_i</u>	<u>f_{ri}</u>
1	0,3 ---0,65		8	0,14
2	0,65 ---1,00		23	0,40
3	1,00 ---1,35		14	0,25
4	1,35 ---1,70		7	0,12
5	1,70 ---2,05		3	0,05
6	2,05 ---2,40		2	0,04

Exemplo-1

- $k = 7$

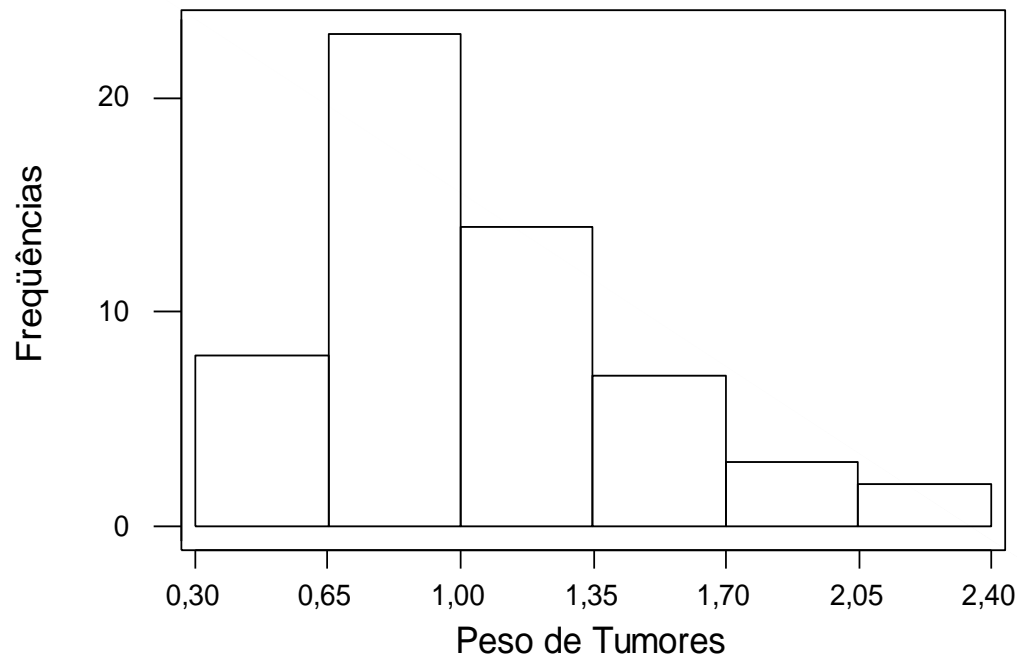
<u>i</u>	<u>Intervalos</u>	<u>Cont.</u>	<u>f_i</u>	<u>f_{ri}</u>
1	0,3 ---0,6		6	0,10
2	0,6 ---0,9		22	0,38
3	0,9 ---1,2		10	0,18
4	1,2 ---1,5		13	0,23
5	1,5 ---1,8		2	0,035
6	1,8 ---2,1		2	0,035
7	2,1 ---2,4		2	0,035

Construção do Histograma

- Representação Gráfica da Distribuição de Frequências
 - Gráfico consiste de Retângulos cujas bases são os intervalos de classe, e as alturas as frequências, ou frequências relativas da classe

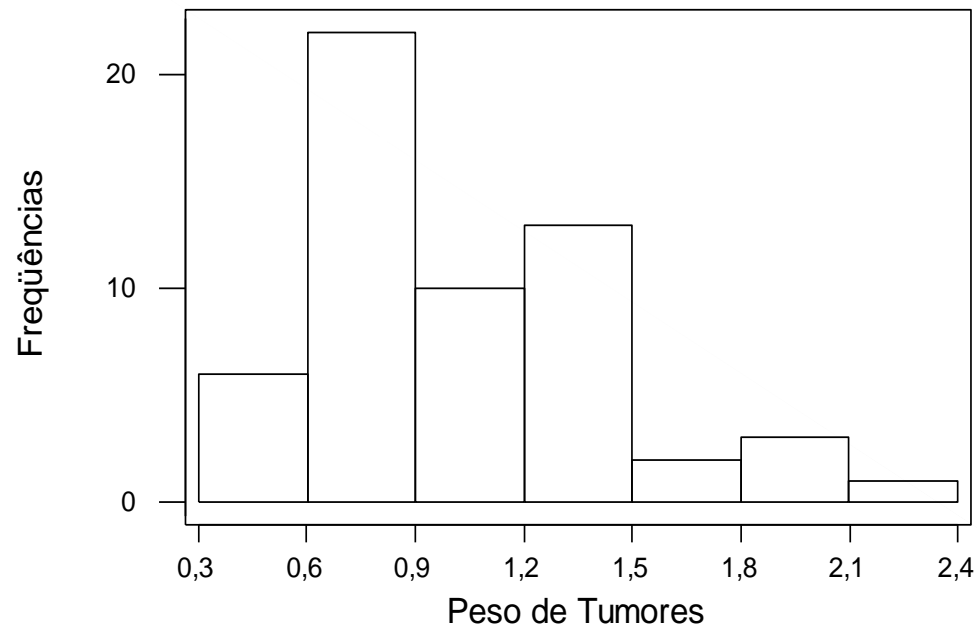
Histograma Exemplo-1

- $k=6$



Histograma Exemplo-1

- $k=7$



Histograma Exemplo-1

- k automático

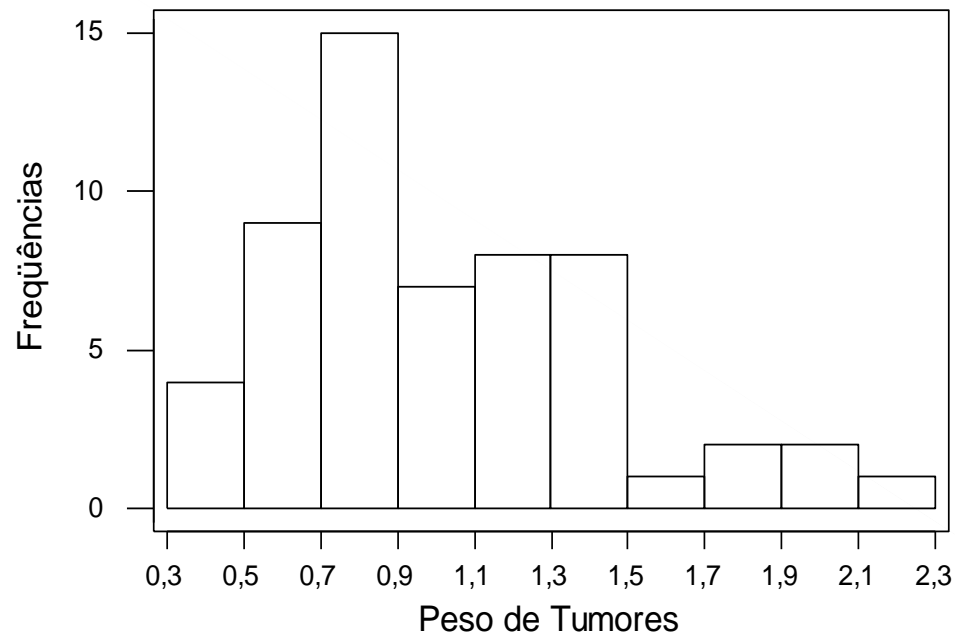
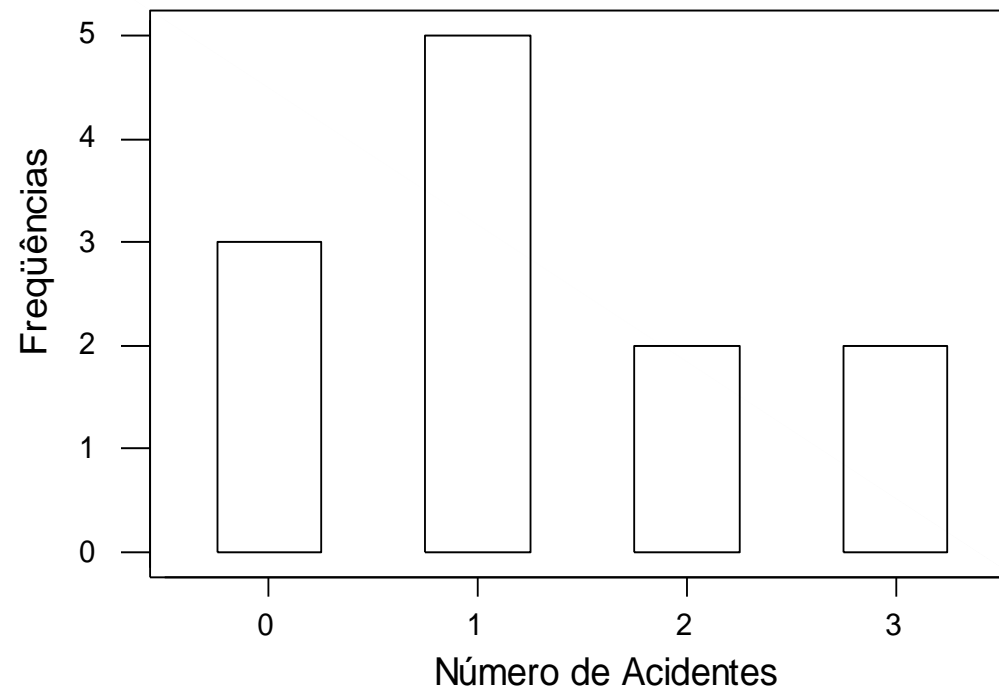


Diagrama de Barras

- Observações são contagens
- Exemplo: Número de acidentes de trabalho por mês, no período de um ano:

1 1 0 1 2 3 2 3 1 0 1 0

Diagrama de Barras



Ramos e Folhas

- Representar Distribuições de Frequências
 - Utilização: Conjunto pequeno de dados
 - Vantagem: Visualização completa das obs.
 - Construção: Cada obs. dividida em duas partes:
 - » **Ramo**
 - » **Folha**

Ramos e Folhas

- Exemplo 1:

2,5 2,6 2,5 2,4 5,4 8,8 12,3

Unidade das folhas = 0,1

RAMOS	FOLHAS
2	4 5 5 6
5	4
8	8
12	3

Ramos e Folhas

- Exemplo 2:

56 62 63 65 65 65 68 70 72

Unidade das folhas = 1

Ramos	Folhas
5	6
6	2 3 5 5 5 8
7	0 2

Ramos e Folhas

- Exemplo 3: Unidade das Folhas = 0,1

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

Métodos Numéricos

- Notação:
 - » **N : Tamanho da População**
 - » **n : Tamanho da Amostra**
- Parâmetros : Medidas da População
- Estatísticas: Medidas da Amostra
- Amostra Aleatória:
 - Variável X: X_1, X_2, \dots, X_n

Métodos Numéricos

- Medidas de Tendência Central:

- Média:

- Populacional:

$$\mu = \sum_{j=1}^N \frac{y_j}{N}$$

- Amostral:

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$$

- Mediana: x_M ---valor tal que:

- no. obs maiores ou iguais a x_M = no. obs. menores ou iguais a x_M

Se n é par mediana é média das duas obs. centrais

Se n é impar, mediana é a obs. central

Métodos Numéricos

– Moda : x_0 = observação com maior frequência

- Exemplo 4:

56 62 63 65 65 65 65 68 70 72

– Mediana $x_M = 65$

– Moda $x_0 = 65$

Métodos Numéricos

- Medidas de Variabilidade
 - Amplitude : $R = \text{Max} - \text{Min}$
 - Desvio: $X_i - \bar{X}$
 - Variância Populacional:

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

Métodos Numéricos

- **Variância Amostral**

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

- **Desvio Padrão:**

σ

S

Métodos Numéricos

- Aproximação para S

$$S \approx \frac{R}{4}$$

- Coeficiente de Variação:

$$C.V. = \frac{S}{\overline{X}}$$

Métodos Numéricos

- Exemplo 4:

- $R = 72 - 56 = 16$

- $n = 10$

- $S^2 = (1/9)[(56 - 65,1)^2 + (62 - 65,1)^2 + \dots + (72 - 65,1)^2] = 19,65$

- $S = 4,4328$

- Aproximação: $R/4 = 16/4 = 4$

- $C.V = 4,4328/58,6 = 0,0756$

Métodos Numéricos

- Medida de Posição Relativa

– “z-score”

$$z = \frac{(X_i - \bar{X})}{S}$$

Métodos Numéricos

- Ex. Notas Teste de Seleção

Nota do candidato $X_i = 71$, Média = 78 $S = 3,5$

$$z = (71 - 78) / 3,5 = -2,00$$

- Aplica-se o teste novamente

$$X_i = 68 \quad \text{Média} = 78 \quad S = 5$$

$$z = (68 - 78) / 5 = -2,00$$

Posições equivalentes

Métodos Numéricos

- Percentis
 - Conjunto com n observações
 - Definição:

p-ésimo percentil: é o valor, x_p , tal que pelo menos $p\%$ das observações, são menores ou iguais a, x_p , e pelo menos $(1-p)\%$ das obs. são maiores ou iguais a x_p

Métodos Numéricos

– Casos Particulares:

- **Primeiro Quartil, ou, Quartil Inferior(Q_1)**
 - É o percentil de ordem 25
 - » Posição no conjunto de dados: $(n+1)/4$
- **Segundo Quartil (Mediana) (Q_2)**
 - É o percentil de ordem 50
- **Terceiro Quartil, ou, Quartil Superior (Q_3)**
 - É o percentil de ordem 75
 - » Posição no conjunto de dados: $3(n+1)/4$

Métodos Numéricos

- Exemplo 5:

280 283 287 288 288 289 289 290 290

290 292 293 293 293

- **Primeiro Quartil: Posição---** $(14+1)/4 = 3.75$
 - 25% de 14 = 3,5 $Q_1 = 288$
- **Segundo Quartil (Mediana)~: Posição = $14/2=7$**
 - $Q_2 = 289,5 = (289+290)/2$
- **Terceiro Quartil: Posição---** $3(14+1)/4=11,25$
 - 75% de 14 = 10,5 $Q_3 = 292$

Diagrama dos Cinco Números (Box-Plot)

- Representação Gráfica dos Quartis

Q_1 , Q_2 e Q_3

– Utilização

- **Identificar Obs. Discrepantes (Outliers)**
- **Comparar dispersão de diferentes conjuntos de dados**

Diagrama dos Cinco Números (Box-Plot)

- Construção
 - Calcular o Intervalo Interquartis
 - $[Q_1 - 1,5(Q_3 - Q_1) ; Q_3 + 1,5(Q_3 - Q_1)]$
 - Obs. fora desse intervalos, são consideradas discrepantes
 - Construir um retângulo com base inferior igual a Q_1 , e base superior igual a Q_3 .
 - A partir do pto. médio da base superior, levantar haste indo até a maior obs, que esteja dentro do Intervalo Interquartis.

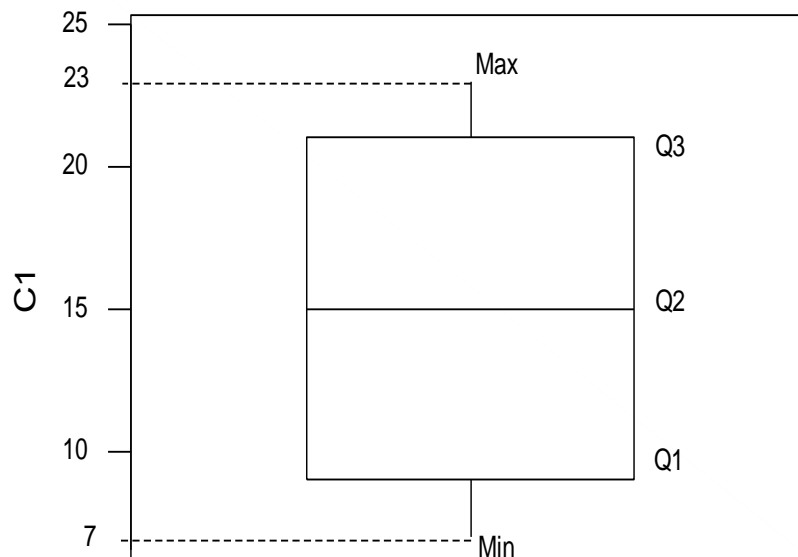
Diagrama dos Cinco Números (Box-Plot)

- A partir do pto. médio da base inferior, traçar haste indo até a menor obs, que esteja dentro do Intervalo Interquartis
- Assinalar com um traço, dentro do retângulo, a mediana
- Observações que estão fora do intervalo (Discrepante), assinalar com *, acima da haste superior, ou abaixo da inferior, conforme o caso

Box-Plot

- Exemplo 6: 7 9 12 15 17 21 23

$Q_1 = 9$ $Q_2 = 15$ $Q_3 = 21$ Inter. = (-9,39)



BOX-PLOT

- Exemplo Tumores:

1,92	1,78	1,18	0,76	0,84	1,01	0,79	0,90
2,23	0,76	0,62	0,65	0,67	0,70	1,24	1,84
1,21	0,70	2,09	1,44	1,01	1,18	0,79	0,87
0,79	0,70	1,27	0,33	1,61	1,44	0,33	0,90
1,38	1,07	1,18	0,76	0,87	1,41	1,07	0,59
0,45	0,67	0,95	1,33	0,65	0,62	1,21	0,76
1,38	0,79	0,65	0,53	1,30	0,84	1,21	1,38
<u>0,33</u>							

Peso de Tumores (Ordem Crescente)

0,33	0,33	0,33	0,45	0,53
0,59	0,62	0,62	0,65	0,65
0,65	0,67	0,67	0,70	0,70
0,70	0,76	0,76	0,76	0,76
0,79	0,79	0,79	0,79	0,84
0,84	0,87	0,87	0,90	0,90
0,95	1,01	1,01	1,07	1,07
1,18	1,18	1,18	1,21	1,21
1,21	1,24	1,27	1,30	1,33
1,38	1,38	1,38	1,41	1,44
1,44	1,61	1,78	1,84	1,92
2,09	2,23			

Posições

- Q1 posição $(57 + 1)/4 = 14,5$
- Q2 posição $57/2 = 28,5$
- Q3 posição $3(57+1)/4 = 43,5$
- $Q_1 = 0,70$ $Q_2 = 0,9$ $Q_3 = 1,27$
- $Q3 - Q1 = 1,27 - 0,70 = 0,57$
- $1,5(Q3 - Q1) = 0,855$ ----
- Intervalo $(-0,15 ; 2,12)$

BOX-PLOT

- $Q_1 = 0,70$ $Q_2 = 0,9$ $Q_3 = 1,27$

Intervalo $(-0,15 ; 2,12)$

