

Introdução a Bioestatística

Gleici Castro Perdoná

pgleici@fmrp.usp.br

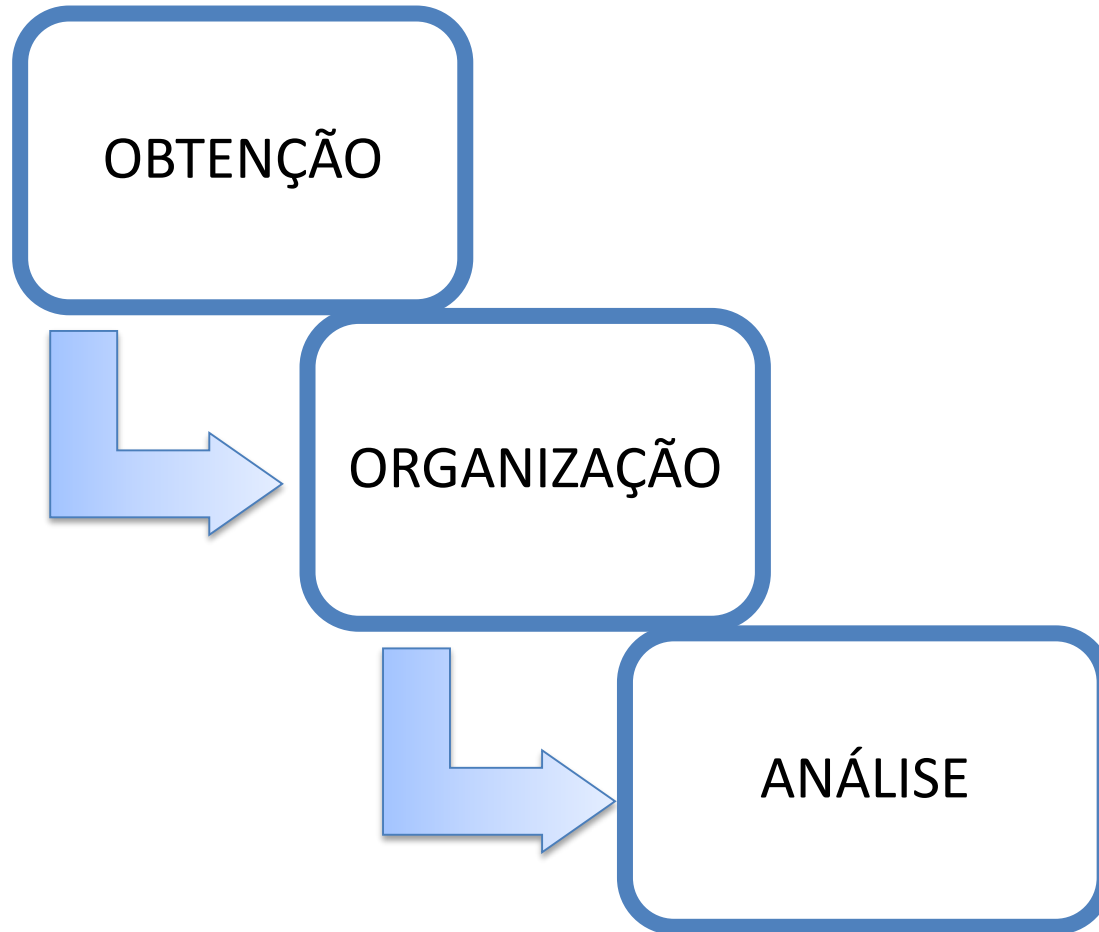
O que é estatística?

- A estatística é uma ciência que estuda a coleção, organização, análise e interpretação de dados. Envolve os conceitos básicos da teoria de probabilidade e inferência. Lida com a variabilidade da natureza dos dados.
- É fundamental à epidemiologia, à ecologia, à psicologia social e à medicina baseada em evidência.

Objetivos da estatística ?

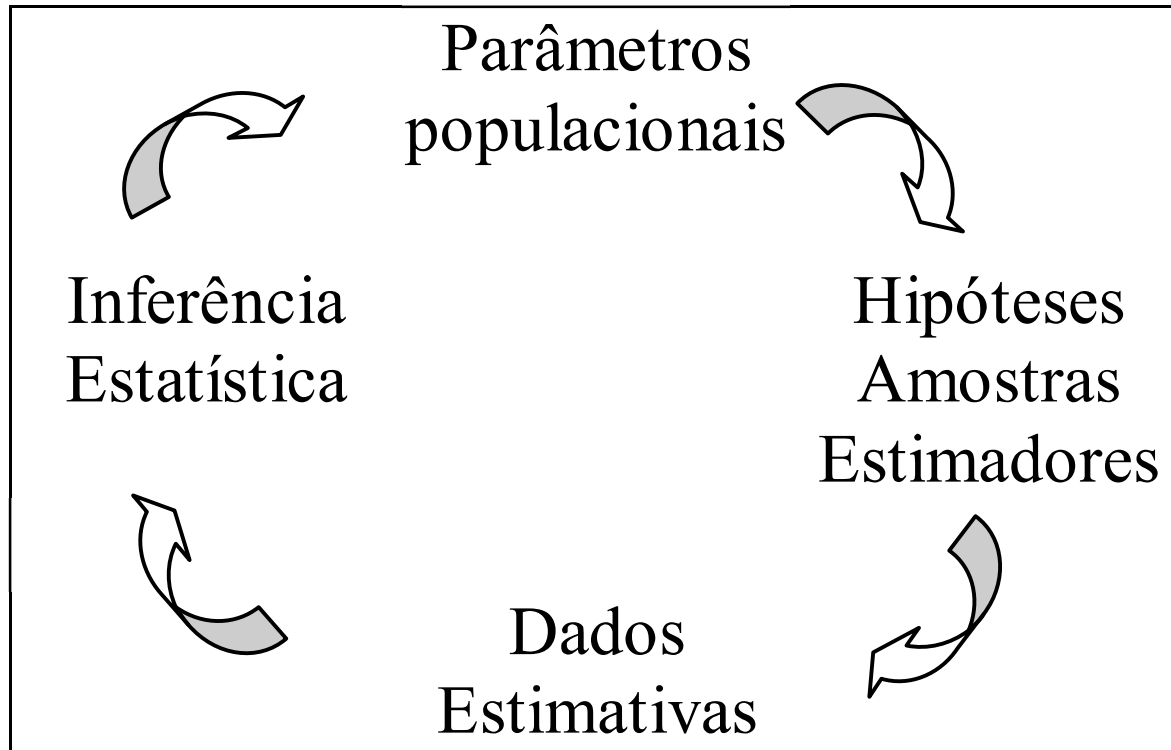
- Fazer inferências sobre a população com base em dados amostrais.
- Planejar o experimento e o tamanho da amostra para que tais inferências tenham uma base confiável.

Estatística trata da...



De dados

Processo de inferência



Definido a população e a amostra, precisamos operacionalizar as variáveis

Variável então, é uma característica da população

DEPENDENTE

São aquelas cujos efeitos são esperados de acordo com as causas

EX:
Cancer, Infarto

INDEPENDENTE

São aquelas cujos efeitos queremos medir.

Ex:
Fumo, alcool

Classificação das variáveis de uma amostra

QUALITATIVAS

NOMINAIS
(M,F) (A,B,C...)

ORDINAIS
(A1, B1) (fibroso,
+fibroso) (1º,2º)

QUANTITATIVAS

DISCRETAS
(1, 4, 7, 3, 8)

CONTÍNUAS
(134 cm, 136,5 cm)
(23 kg, 56,4 kg)

Operacionalizar uma variável pode não ser tão fácil

	Nominal	Ordinal	Discretas	Continuas
Idade		Criança, jovem, adulto	n. anos completos	Idade, A,M,D
Alcool	Sim, Não	Pouco, médio, muito	N de copos	Qte no sangue de alcool
Classe social		Baixa, media, alta	n. Sal. Min	Renda Fam. (R\$)

FORMULARIO DE COLETA DE DADOS

- 1) Qual sua idade? _____ anos
2) Qual seu sexo? () Masculino () Feminino
3) Qual seu estado civil? () Solteiro(a) () Casado(a) () Mora junto ()
Outro _____

4) Qual seu Estado de origem? _____

5) Qual a sua cor/etnia?

- () Branco(a)
() Preto(a)
() Amarelo(a)
() Pardo(a)
() Indígena

6) Qual sua renda familiar mensal ?

- () < 1 salário mínimo
() entre 1 e 2 salários mínimos
() entre 2 e 3 salários mínimos
() entre 3 e 5 salários mínimos
() entre 5 e 7 salários mínimos
() entre 7 e 10 salários mínimos
() entre 10 e 14 salários mínimos
() entre 14 e 20 salários mínimos
() > de 20 salários mínimos

7) Você cursou o ensino FUNDAMENTAL em escola

- () Pública
() Particular
() Particular com bolsa de estudos integral
() Particular com bolsa de estudos parcial
() Outra _____

8) Você cursou o ensino MÉDIO em escola

- () Pública
() Particular
() Particular com bolsa de estudos integral
() Particular com bolsa de estudos parcial
() Outra _____

9) Possui algum curso técnico? Se sim, qual?

- () Sim, _____
() Não

10) Possui algum curso SUPERIOR? Se sim, qual?

- () Sim, _____
() Não

11) Fez algum curso pré-vestibular ? Se sim por quanto tempo?

Atividade formulário

Classifique as variáveis (10 min)

- Num estudo sobre hábitos alimentares e risco de doenças cardíacas, foram obtidos os níveis de colesterol (em mg/dl) de 80 pacientes de uma região.
- O Nível de albumina no sangue (g/dl) é um indicador do estado nutricional
- Pressão intraocular (mmHg)
- Estadio do câncer no colo do útero
- Idade
- Tempo até ocorrência de displasia
- sexo

Ordenação dos dados

- Planejado e coletado um conjunto de dados, precisamos ordenar e resumir este conjunto.
 - Tabelas
 - Gráficos

TABELAS

Título

- precede a tabela; contém a designação do fato observado, o local e a época em que ocorreu.

Corpo

- Cabeçalho - parte superior da tabela, que especifica seu conteúdo.
- Restante: conjunto de dados numéricos que expressam as variações quantitativas e qualitativas associadas a determinado fenômeno.

Rodapé

- Fonte – especifica a entidade responsável pelo fornecimento dos dados ou pela elaboração da tabela.
- Notas – informações gerais sobre a tabela; podem indicar a metodologia utilizada.
- Chamadas – informações específicas sobre partes da tabela.

GRÁFICOS

- Gráficos são REPRESENTAÇÕES VISUAIS utilizadas para exibir dados, sejam eles, sobre determinada informação, ou valores numéricos.
- PODEM SER utilizados para demonstrar padrões, tendências e ainda, comparar INFORMAÇÕES qualitativas e quantitativas num determinado espaço de tempo, na descoberta de novos fenômenos, na aceitação ou rejeição de hipóteses, etc.
- SÃO UM MEIO RÁPIDO E FÁCIL de comunicação
- **Florence Nightingale** foi uma das pioneiras na utilização dos gráficos estatísticos. Em 1858, ela foi a primeira mulher eleita membro da Associação Inglesa de Estatística.

Colunas e barras: séries temporais e específicas, comparação de dados, diferença de volumes, dados acumulados, mais de uma variável



Histograma:
distribuição de frequências (dados contínuos)



GRÁFICOS



Linhas: correlações entre dados



Setores: comparação de parcelas em relação ao total (preferencialmente percentuais)



Diagrama de ordenadas:
distribuição de frequências (dados discretos)

Gráficos

- Linhas , em geral ligado a series temporais!
- Setores – Pizza, uso quando quero comparar uma parte com o todo
- Barras , uso em comparar dados, ou series geográficas
- Áreas, também em comparação de grupos ,semelhante ao de linhas
- Gráfico de pontos , relação de duas variáveis



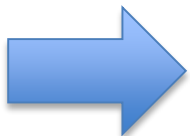
Indique qual o gráfico (10min).

- Para demonstrar o percentual de um todo.
- Para demonstrar a variação da taxa de infecção mensal.
- Para comparar três HC em percentual de internação por dengue em um mês.
- Para comparar relação de duas variáveis quantitativa (número de consultas e numero de crises asma)

Tabelas especiais e seus gráficos

- Tabela de distribuição de Frequências (discreto)

Numero de irmãos	F
0	3
1	7
2	15
3	5
4	4
6	2
Total	36



Número irmão da 6 série da escola X, Ribeirão Preto – SP, 2016

Diagrama de ordenada

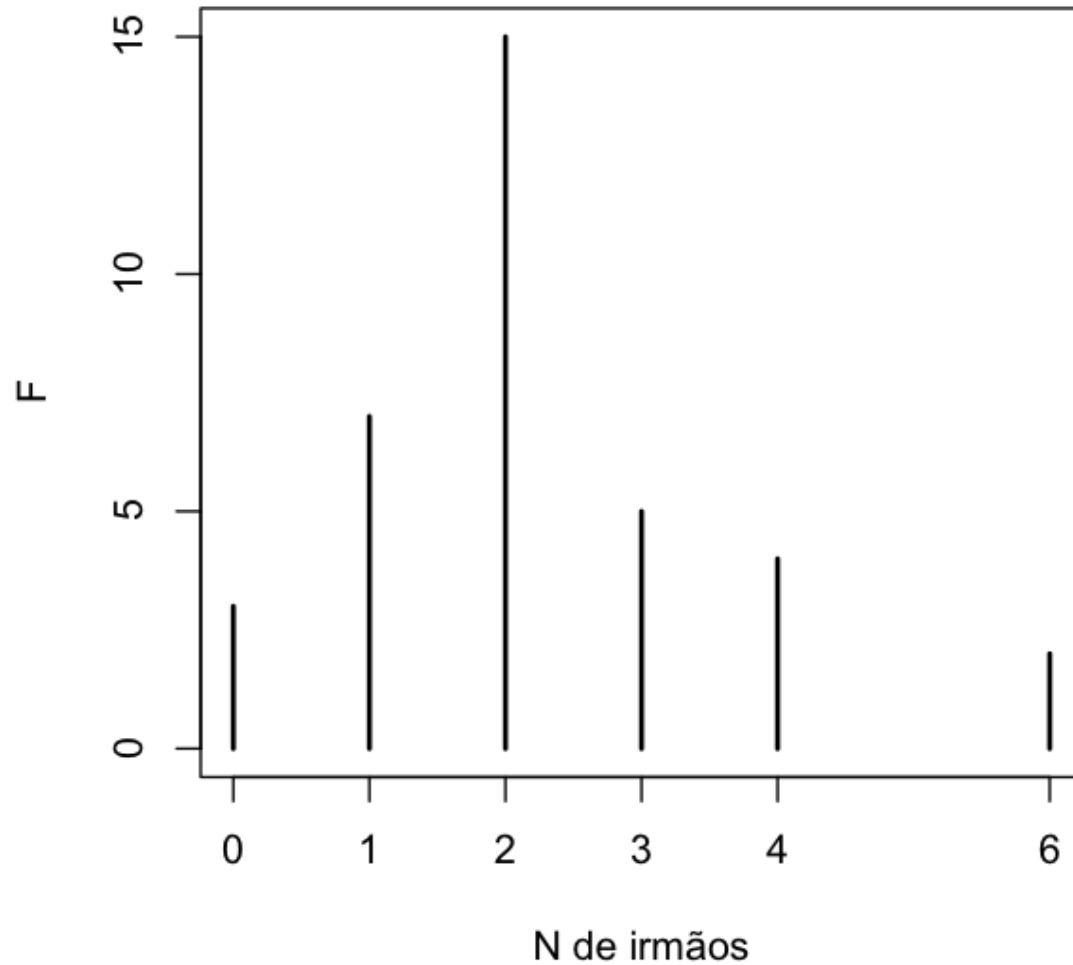
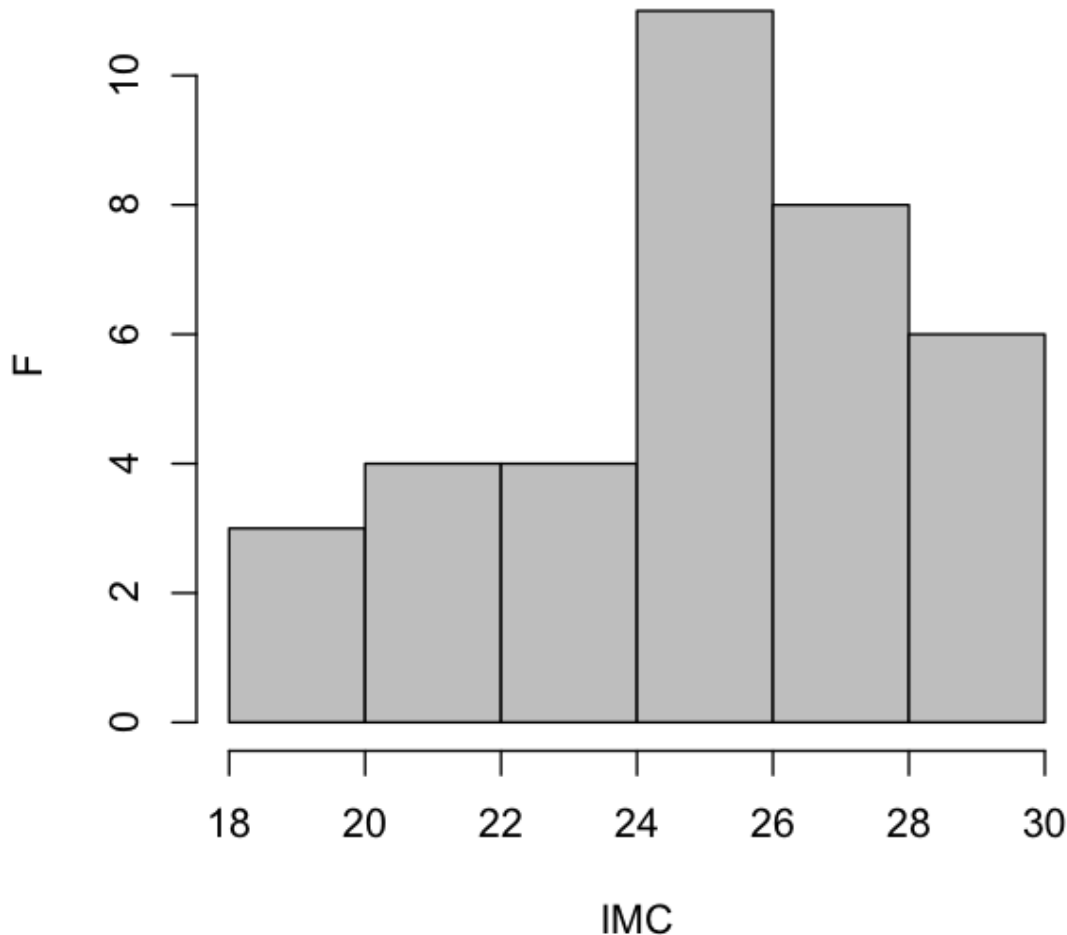


Tabela de Destruição de Frequências (contínuos)

IMC (kg/m ²)	f
18 20	3
20 22	4
22 24	4
24 26	9
26 28	10
28 30	6
Total	36

IMC dos alunos da 6 série da escola X, Ribeirão Preto – SP, 2016

Histograma



- O histograma é uma importante ferramenta de diagnóstico, pois dá uma ideia visual da variação dos dados.
 - Interpretação:
 - tendência central
 - dados parecem criar curva assimétrica/simétrica;
 - especificação para algum valor;
 - Existem alunos com maior nível.

CONSTRUÇÃO DA TABELA DE DISTRIBUIÇÃO DE FREQUÊNCIAS

1. Calcular a amplitude total dos dados, A (maior dado – menor dado).
2. Calcular o número de classes, k , por um dos seguintes métodos:
 - $k = \sqrt{n}$, sendo n = número de dados;
 - $k = 1 + 3,22 \log n$, sendo n = número de dados.
3. Dividir A por k , o que fornecerá a amplitude de cada classe, a .

Construa um diagrama de ordenadas e um histograma (20 min)

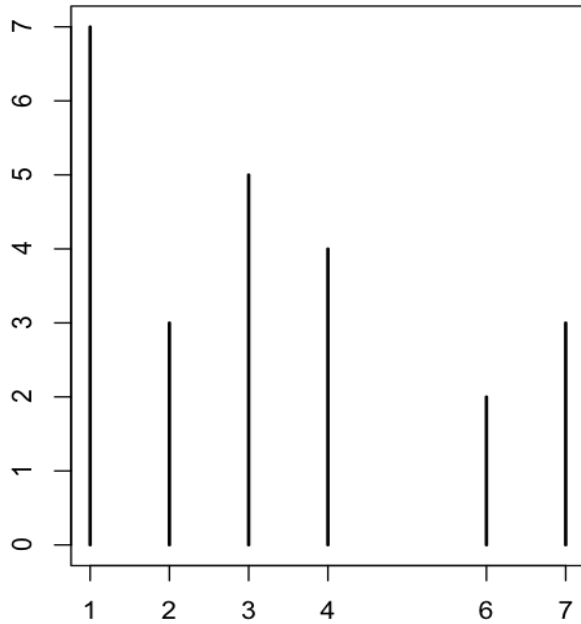
- Numero crises mensais em asmáticos

2, 6, 7, 4, 1, 1, 3, 1, 7, 7, 3, 2, 3, 3, 1, 1, 4, 3, 1, 4,
2, 1, 4, 6

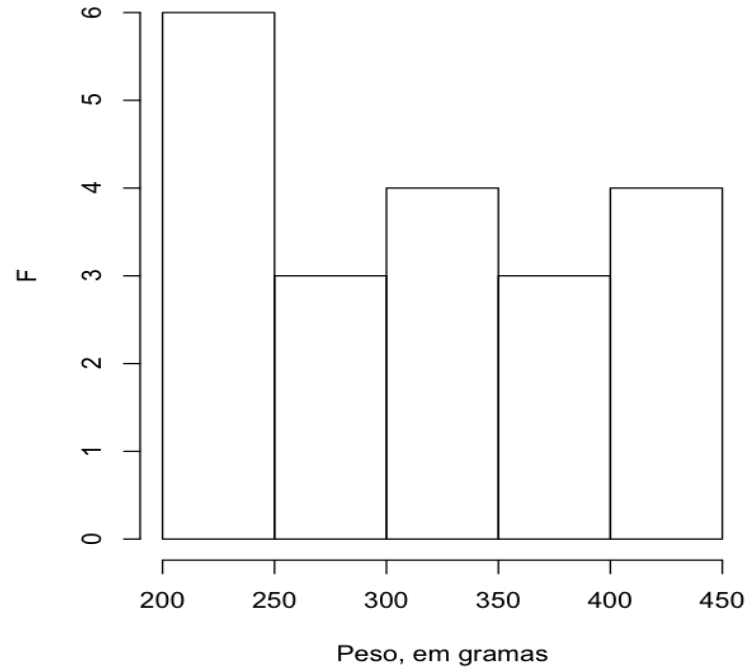
- Peso gestacional de fetos com 20 semanas(IG)

364 ,243, 408, 443, 232, 222, 368, 362, 346, 407,
342, 347, 266, 243 ,243, 254, 298, 437, 301, 240

Resposta



<u>N.</u>	<u>F</u>
1	7
2	3
3	5
4	4
6	2
7	3
<u>Tot</u>	<u>24</u>



<u>Peso (g)</u>	<u>F</u>
200 250	6
250 300	3
300 350	4
350 400	3
<u>400 450</u>	<u>4</u>
<u>Total</u>	<u>20</u>

Intervalo (20 minutos)



Ainda Ordenação

- Ordem crescente
- Ordem decrescente
- Percentis
- Quantis

QUANTIS



Q1: valor que tem $\frac{1}{4}$ dos dados abaixo dele e $\frac{3}{4}$ acima (percentil 25%)

Q2: valor correspondente à mediana, tem $\frac{1}{2}$ dos dados abaixo dele e $\frac{1}{2}$ acima (percentil 50%)

Q3: valor que tem $\frac{3}{4}$ dos dados abaixo dele e $\frac{1}{4}$ acima (percentil 75%)

Percentil – REGRA GERAL

Primeiro quantil = 25o. Percentil para $n=50$ ou $n=200$

1. Ordenar os dados
2. Calcular

$$l = \frac{nk}{100}$$

l é a posição ordenada do percentil procurado,
Para $n=50 \Rightarrow 50 \cdot 25 / 100 = 12,5$
Para $n=200 \Rightarrow 200 \cdot 25 / 100 = 50$

Se Inteiro (no caso 50º. posição)

Não inteiro (no caso 12,5º posição)

O k-ésimo percentil:

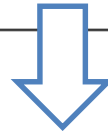
$$\frac{l - \text{ésima} + (l + 1) - \text{ésima}}{2}$$

O inteiro sucessor

EXEMPLO DE CÁLCULO DE MEDIDAS DESCRITIVAS

Tabela 1 – Volume Expiratório Forçado de 13 indivíduos do hospital X

Indivíduo	1	2	3	4	5	6	7	8	9	10	11	12	13
VEF 1 (L)	2.30	2.15	3.50	2.60	2.75	2.82	4.05	2.25	2.68	3.00	4.02	2.85	3.38



$$l=13*25/100=3,25$$

$$l=13*50/100=6,50$$

$$l=13*75/100=9,750$$

2,15	2,25	2,30	2,60	2,68	2,75	2,82	2,85	3,00	3,38	3,50	4,02	4,05
------	------	------	------	------	------	------	------	------	------	------	------	------

Quantil 1

Mediana

Quantil 3

$$k=25$$

$$l=(25*13)/100 = 3,25$$

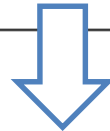
l é não inteiro

Q1 : 4 observação

EXEMPLO DE CÁLCULO DE MEDIDAS DESCRITIVAS

- O mesmo conjunto de dados, mas com 12 pacientes

Indivíduo	1	2	3	4	5	6	7	8	9	10	11	12
VEF 1 (L)	2.30	2.15	3.50	2.60	2.75	2.82	4.05	2.25	2.68	3.00	4.02	2.85



$$l=12*25/100=3$$

$$l=12*50/100=6$$

$$l=12*75/100=9$$

2,15	2,25	2,30	2,60	2,68	2,75	2,82	2,85	3,00	3,50	4,02	4,05
------	------	------	------	------	------	------	------	------	------	------	------

Quantil 1 = 2,45

$$k=25$$
$$l=(25*13)/100 = 3$$

l é inteiro

$$Q1 : (3^{\text{a}} + 4^{\text{a}} \text{ observação})/2 = (2,30+2,60)/2 = 2,45$$

Encontre os quantis (25%, 50% e 75%)

Encontre os percentis (5% e 95%)

10 min

- Numero crises mensais em asmáticos

1 1 1 1 1 1 1 2 2 2 3 3 3 3 3 4 4 4 4 6 6 7 7 7

- Peso gestacional de fetos com 20 semanas(IG)

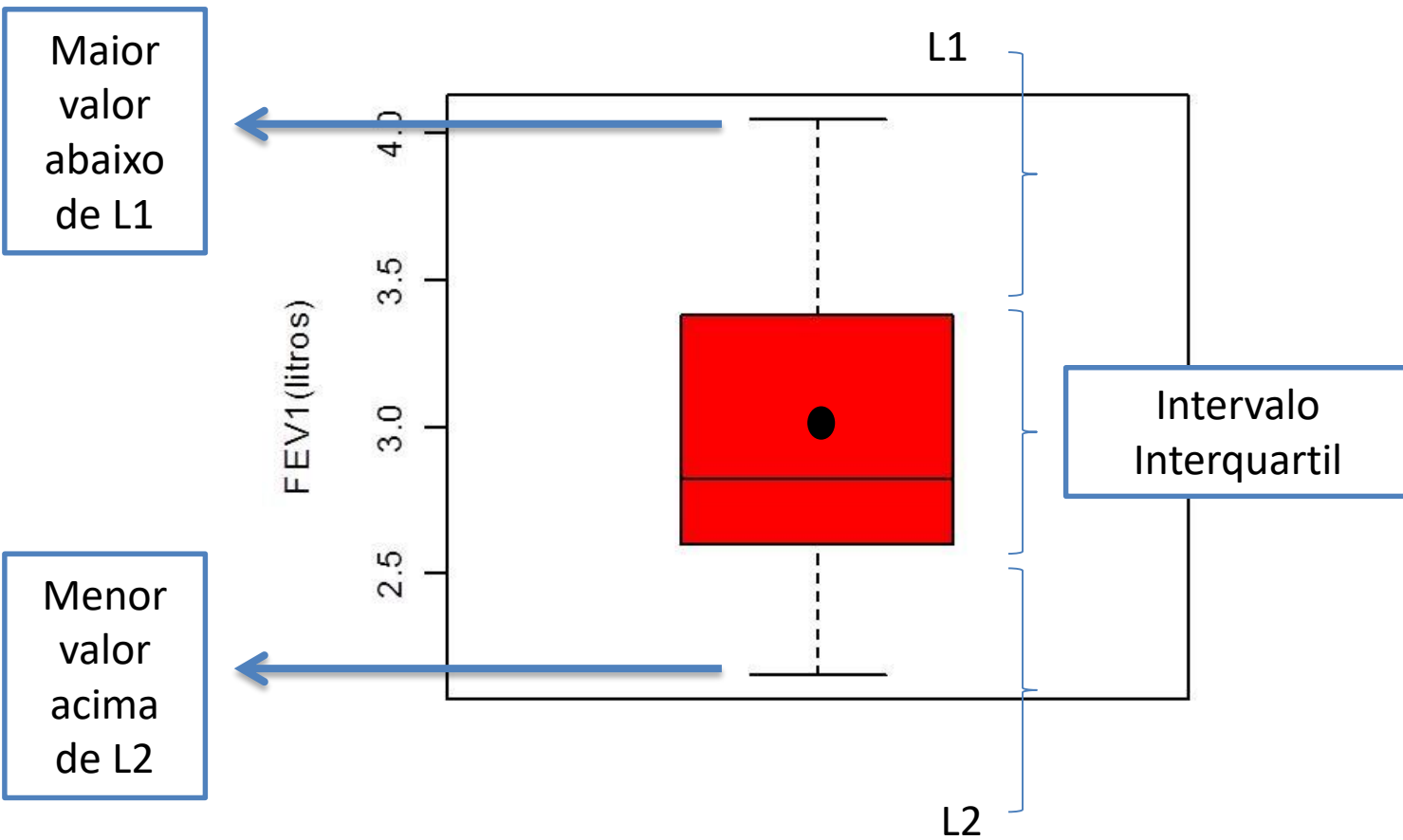
222 232 240 243 243 243 254 266 298 301 342

346 347 362 364 368 407 408 437 443

Gráficos associados

- Box plot (usa os quantis)
- Ramos e folhas (usa os próprios valores)

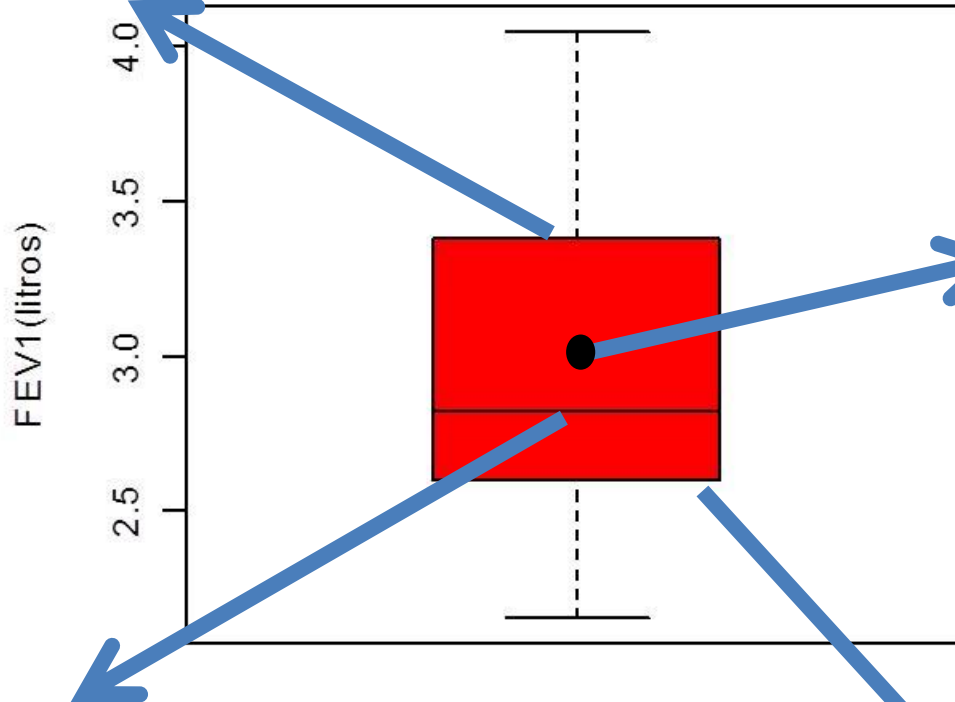
BOXPLOT



L1: $Q3 + 1,5 * \text{Intervalo Interquartil}$
L2: $Q1 - 1,5 * \text{Intervalo Interquartil}$

BOXPLOT

Q3 = 3,38



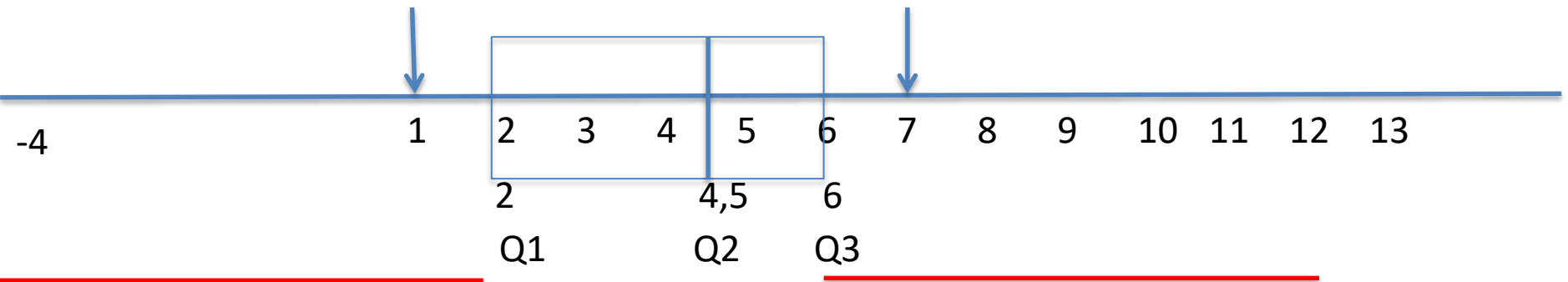
Média = 2,95

Mediana = 2,82

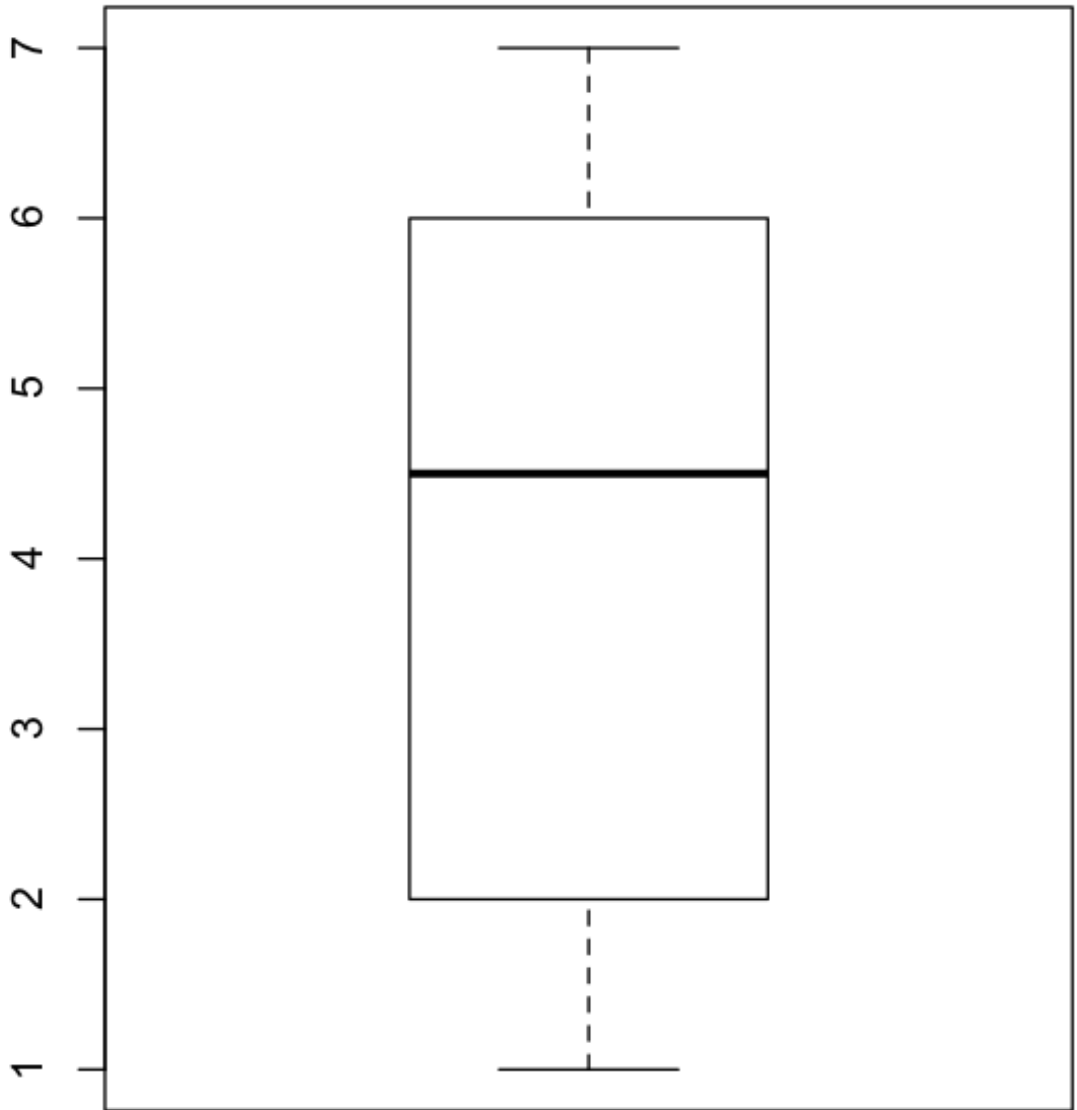
Q1 = 2,60

Exemplo: construção

1 1 1 1 2 2 2 3 3 3 4 4 5 5 5 5 5 6 6 7 7 7 7 7



Intervalo interquartil $Q3-Q1 = (6-2) = 4$, multiplicado por 1,5 = 6



Titulo e nome no eixo

Agora sua vez!!

- Numero crises mensais em asmáticos
- 1 1 1 1 1 1 1 2 2 2 3 3 3 3 3 4 4 4 4 6 6 7 7 13
- Peso gestacional de fetos com 20 semanas(IG)
222 232 240 243 243 243 254 266 298 301 342
346 347 362 364 368 407 408 437 443

BIBLIOGRAFIA RECOMENDADA

BUSSAB, W.O.; MORETTIN, P. **Estatística básica**. 4 ed. São Paulo, Atual, 1987.

PAGANO, M. e GAUVREAU, K. Princípios de Bioestatística - Tradução da 2ª Edição Norte Americana, Pioneira Thonpson Learning, São Paulo, SP,2004.

ROSNER, B.Fundamentos de bioestatística. 8ª Edição Norte Americana, Cengage Learning, 2016.

SOARES , J. F. e SIQUEIRA, A. L. Introdução à Estatística Médica. UFMG, Belo Horizonte,1999.

Aula usando R

- Digitando valores no R
- asma = c(2, 6, 7, 4, 1, 1, 3, 1, 7, 7, 3, 2, 3, 3, 1, 1, 4, 3, 1, 4, 2, 1, 4, 6)
- peso = c(364 ,243, 408, 443, 232, 222, 368, 362, 346, 407, 342, 347, 266, 243 ,243, 254, 298, 437, 301, 240)

Alguns comandos

- `quantile(asma)`
- `summary(asma)`
- `boxplot(asma)`
- `hist(asma)`
- `quantile(asma,seq(0.10,0.9,0.1))`

Comandos lógicos

Símbolo	Descrição
<	Menor
<=	Menor ou igual
>	Maior
>=	Maior ou igual
==	Igual (comparação)
!=	Diferente
&	AND
	OR
!	NOT
TRUE ou 1	Valor booleano verdadeiro (1)
FALSE ou 0	Valor booleano falso (0)

Gráficos

```
a= 1:20  
b = a^2  
plot(a,b)  
plot(a,b, type="l")  
lines(rev(a),b) #adição de linhas  
points(a, 400-b) #adição de pontos  
  
a <- seq(from=0, to=20, by=2); b <- a^2  
plot(a,b,type="n") #plota um gráfico vazio  
text(a,b,"G") #aplica a letra G
```

TEM MUITO MAIS

Veja o que acontece

- `x = c(2,3,4,5,6,7,8,9)` **#Representação das coordenadas**
- **"x" das cidades**
- `y = c(15,26,45,8,74,11,61,32)` **#Coordenadas y das cidades**
- **#Descrevendo o nome das cidades:**
- `nomes <- paste("cidade", LETTERS[1:8], sep= " ")`
- `idades <- data.frame(x,y,row.names=nomes)` **#Juntando os dados**
- `idades` **#visualizando o objeto cidades**
- `plot(idades)`
- `identify(x,y,idades,n=4)`

Voltamos ao gráficos vistos hoje

- `par(mfrow=c(1,2))`
- `hist(peso,nclass=12)`
- `hist(peso,nclass=6)`

- `boxplot(peso, col=3)`
- `boxplot(peso, boxwex=2)`

E o gráfico pie ou barra??

- `a=c(0.12, 0.3, 0.26, 0.16, 0.04, 0.12)`
- `names(a)=c("a","b","c","d","e","f")`
- `pie(a,col = c("red","blue","green","gray", "pink", "white"))`

- `x <- c(1,2,3,4,5,6,7)`
- `barplot(x)`
- `y=c("A","B","C","D","E","F","G")`
- `barplot(x,names.arg=y,col=2)`

PROXIMA AULA

- MEDIDAS DESCRIPTIVAS
- mean (peso)
- var(peso)
- median(peso)
- sd(peso)

OBRIGADA

