



Faculdade de Medicina de Ribeirão Preto  
Universidade de São Paulo

**RFM0010/RCG3023**  
**Bioestatística**

Cursos de Graduação em  
Fisioterapia

**2016**

Atualizado para o primeiro semestre de 2016

Edson Zangiacomi Martinez

*“Perhaps the most important contribution of statistics comes not from the direct use of statistical methods in science and technology, but rather in helping us learn about the world.”*

(Andrew Gelman, 2014)

**Estatística:** Ramo do conhecimento que explora a coleta, a organização, a análise e a interpretação de dados. Quando o foco está nas ciências biológicas e da saúde, usamos o termo *bioestatística*.

*“Ao nível da iniciação, a Estatística não deve ser apresentada como um ramo da Matemática. A boa Estatística não deve ser identificada com rigor ou pureza matemáticos mas ser mais estreitamente relacionada com pensamento cuidadoso. Em particular, os alunos devem apreciar **como a Estatística é associada com o método científico**: observamos a natureza e formulamos questões, coligimos dados que lançam luz sobre essas questões, analisamos os dados e comparamos os resultados com o que tínhamos pensado previamente, levantamos novas questões e assim sucessivamente.”* (Hogg, 1991)

**Método científico:** conjunto de estratégias, ferramentas e ideias resultantes da experiência humana e consequentes do acúmulo de saberes, que, estruturadas e sistematizadas, possibilitam alcançar um objetivo, que é responder a uma pergunta.

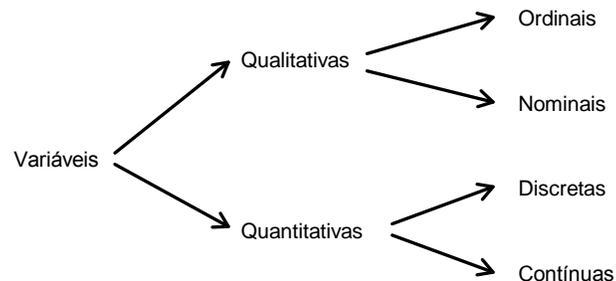
Toda pesquisa científica é baseada em uma **pergunta**. São exemplos:

- Qual é a prevalência da tuberculose na cidade do Rio de Janeiro?
- Qual é a incidência de aids na cidade de São Paulo?
- O consumo de alimentos transgênicos pode elevar o risco de doenças gástricas?



<p><b>Variável:</b> Uma variável é uma característica de interesse que pode assumir diferentes valores ou classificações para diferentes sujeitos, organismos ou objetos selecionados para nosso estudo.</p> <p>No planejamento da pesquisa, devemos definir quais são as nossas características de interesse, antes da coleta dos dados.</p>	<p><b>São exemplos:</b>  Sexo (M/F)  Idade (em anos completos)  Altura (em centímetros)  Hábito de fumar (sim/não)  Tempo de atividade física (em horas/semana)  Número de filhos na família  Classe econômica (A, B, C ou D)  Estádio da doença</p>
---	--

### Classificação de uma variável conforme sua natureza:



As variáveis **qualitativas** são aquelas cujos possíveis valores representam atributos e/ou qualidades.

<p><b>Qualitativa ordinal:</b> assumem classificações, atributos ou qualidades que podem ser descritas em uma ordem natural.</p>	<p><b>São exemplos:</b>  Classe econômica: A, B, C, D, E  Escolaridade: 1o grau, 2o grau, superior  Estadiamento: I, II, III, IV.</p>
<p><b>Qualitativa nominal:</b> não há uma ordem natural.</p>	<p><b>São exemplos:</b>  Tipo sanguíneo: A, B, AB, O  Sexo: M,F  Estado civil: solteiro, casado, divorciado, viúvo  Tabagismo: fumante, ex-fumante, nunca fumou  Cor dos olhos ....</p>

As variáveis “numéricas” são chamadas **quantitativas**.

<p><b>Quantitativa discreta:</b> seus possíveis valores pertencem a um conjunto finito ou contável. Em geral, expressa números inteiros, resultantes de um processo de contagem.</p>	<p><b>São exemplos:</b>  Número de filhos: 0, 1, 2, 3, 4, 5  Número de idas ao cinema, por mês: 0, 1, 2, 3  Número de pontos dolorosos: 0, 1, 2, ..., 16</p>
<p><b>Quantitativa contínua:</b> assumem valores em uma escala contínua (na reta real). Para essas variáveis, valores não inteiros fazem sentido. Seus resultados são geralmente provenientes de uma mensuração.</p>	<p><b>São exemplos:</b>  Peso (em kg)  Altura (em cm)  Nível sérico de colesterol, em mg/100 ml.</p>

Número	Idade	Estado civil	Tabagismo	Idade ao 1º filho	Partos	Peso	Altura	Estado de saúde
1	51	casada	não	26	3	74,6	1,59	bom
2	48	casada	não	20	2	53,3	1,51	bom
3	57	casada	não	20	3	64,0	1,63	bom
4	48	casada	sim	21	3	68,6	1,58	regular
5	49	casada	não	28	1	77,9	1,52	bom
6	47	casada	não	15	3	59,9	1,52	bom
7	49	casada	não	19	3	64,0	1,64	regular
8	52	casada	não	30	1	70,5	1,66	regular
9	45	casada	não	27	1	72,6	1,53	bom
10	64	casada	não	20	2	66,0	1,50	bom
11	55	casada	não	19	5	65,4	1,60	bom
12	45	solteira	não	29	1	55,0	1,56	ruim
13	54	casada	não	21	1	66,8	1,64	regular
14	51	casada	não	21	2	70,3	1,59	regular
15	59	viúva	sim	-	0	80,6	1,55	bom
16	56	viúva	sim	22	3	74,8	1,50	bom
17	49	divorciada	não	22	3	60,0	1,60	regular
18	52	casada	não	28	3	61,8	1,57	bom
19	64	casada	não	27	3	59,9	1,57	bom
20	47	divorciada	não	22	3	79,3	1,68	regular
21	50	casada	não	23	2	81,5	1,71	bom
22	64	casada	não	25	2	53,4	1,59	regular
23	52	casada	não	27	3	84,5	1,64	regular
24	56	casada	não	16	4	71,0	1,60	regular
25	59	viúva	sim	21	2	71,8	1,54	bom
26	48	casada	não	31	2	68,9	1,58	bom
27	51	divorciada	não	-	0	111,5	1,48	ruim
28	51	casada	não	22	3	66,7	1,53	bom
29	63	casada	não	22	4	72,5	1,56	bom
30	58	divorciada	não	15	5	79,9	1,53	ruim
31	52	divorciada	sim	-	0	47,9	1,53	bom
32	49	casada	não	19	2	54,6	1,58	bom
33	58	viúva	não	26	3	72,8	1,57	ruim
34	50	casada	não	25	1	89,6	1,54	bom
35	53	casada	sim	21	4	68,5	1,57	bom
36	54	casada	não	20	6	73,5	1,53	bom
37	65	casada	não	28	2	73,6	1,59	bom
38	57	casada	não	16	8	69,7	1,61	ruim
39	58	casada	não	20	4	64,3	1,52	regular
40	54	casada	não	18	4	56,4	1,64	bom

**Estatística descritiva:** é a parte da Estatística que abrange métodos destinados a resumir a informação contida nos dados, destacando os aspectos mais marcantes.

Exemplo:

**Tabela 2.2.** Tabela de frequências para o estado civil.

Estado civil	Frequências absolutas	Frequências relativas
Casada	30	75,0%
Divorciada	5	12,5%
Solteira	1	2,5%
Viúva	4	10,0%
Total	40	100%

**Medidas resumo:** são medidas que buscam sumarizar as informações disponíveis sobre o comportamento de uma variável. O interesse é caracterizar o conjunto de dados através de medidas que resumam a informação nele contida.

- (a) **Medidas de posição ou de tendência central:** São medidas ao redor das quais as observações tendem a se agrupar. Ex.: média, mediana, moda.
- (b) **Medidas de dispersão:** Medem a variabilidade dos dados. Ex. variância, desvio padrão.

### Medidas de posição ou de tendência central

**Média amostral:**  $\bar{x}$

Seja uma amostra de  $n$  indivíduos.  $\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$  ou  $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

**Ex.:**  $X$  : idade das pacientes (em anos)

$$\begin{aligned} x_1 &= 38 \\ x_5 &= 33 \end{aligned}$$

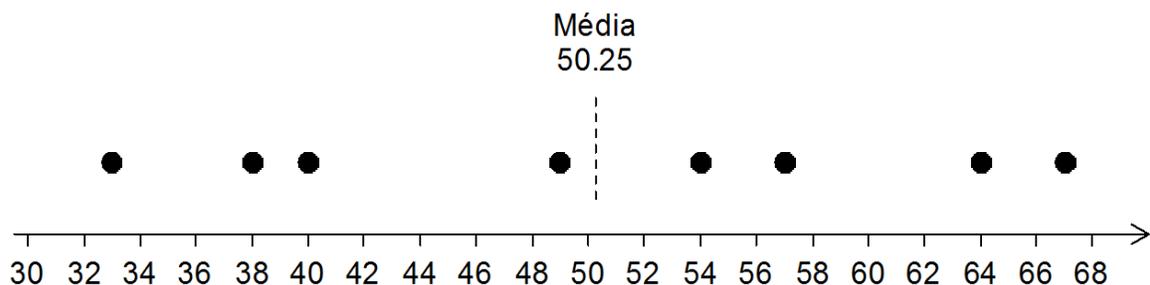
$$\begin{aligned} x_2 &= 40 \\ x_6 &= 57 \end{aligned}$$

$$\begin{aligned} x_3 &= 49 \\ x_7 &= 54 \end{aligned}$$

$$\begin{aligned} x_4 &= 67 \\ x_8 &= 64 \end{aligned}$$

$$\sum_{i=1}^8 x_i = 38 + 40 + 49 + 67 + 33 + 57 + 54 + 64 = 402 \text{ anos.}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{402}{8} = 50,25 \text{ anos ou } 50 \text{ anos e } 3 \text{ meses.}$$



**Ex.:**  $X$  : número de filhos

Seja uma amostra de  $n = 8$  famílias

1            1            1            2            2            2            3            4

Média = \_\_\_\_\_

Seja uma outra amostra de  $n = 8$  famílias

1            1            1            2            2            3            3            4

Média = \_\_\_\_\_

## Mediana

Se as observações são ordenadas da menor até a maior, metade dos valores é maior ou igual à mediana, enquanto a outra metade é menor ou igual a ela.

**Ex. 1:** seja  $X$  o volume expiratório forçado (VEF) em um segundo (em litros) para uma amostra de  $n = 13$  adolescentes que sofrem de asma.

$x_1 = 2,30$              $x_2 = 2,15$              $x_3 = 3,50$              $x_4 = 2,60$              $x_5 = 2,75$   
 $x_6 = 2,82$              $x_7 = 4,05$              $x_8 = 2,25$              $x_9 = 2,68$              $x_{10} = 3,00$   
 $x_{11} = 4,02$              $x_{12} = 2,85$              $x_{13} = 3,38$

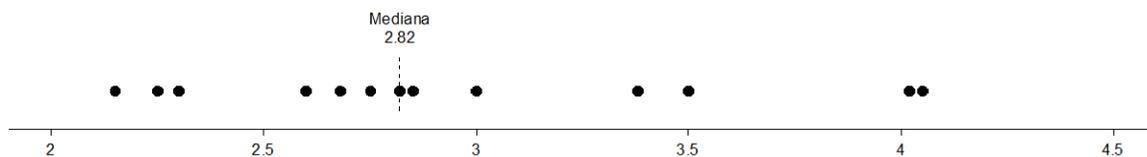
**1º passo:** ordenar as observações

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

**2º passo:** a mediana é o “número do meio”

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

mediana = \_\_\_\_\_ litros



**Ex. 2:**

$x_1 = 38$	$x_2 = 40$	$x_3 = 49$	$x_4 = 67$
$x_5 = 33$	$x_6 = 57$	$x_7 = 54$	$x_8 = 64$

**1º passo:** ordenar as observações

33            38            40            49            54            57            64            67

**2º passo:** a mediana é o “numero do meio”

33      38      40      49      54      57      64      67

Mediana = \_\_\_\_\_ anos

### Moda

É a observação que ocorre mais frequentemente.

**Ex.:** Imagine que em uma loja de sapatos femininos foram vendidos 20 pares de sapatos em um único dia.

34    34    35    35    36    36    36    36    36    36  
37    37    37    37    38    38    38    39    39    39

Moda = \_\_\_\_\_

**Ex.:** Em uma turma de primeiro ano de um curso de graduação, as idades dos 30 alunos (em anos completos) são

17    17    18    18    18    18    18    18    18    18  
18    18    18    18    18    18    18    18    18    18  
18    19    19    19    19    20    20    21    21    22

Moda = \_\_\_\_\_

**Ex.:** Níveis séricos de triglicérides (em mg/dl)

189      72      109      140      140      140      135

Moda = \_\_\_\_\_

**Ex.:** Tempo de aleitamento materno (em meses) de  $n = 7$  crianças usuárias de um serviço de saúde:

1      2      3      3      4      6      6

Moda = \_\_\_\_\_

**Ex.:** Solicitamos a  $n = 10$  pessoas de um bairro que classificassem seu próprio estado de saúde, como “ruim”, “regular” ou “bom”.

Ruim	Bom	Bom	Regular	Regular
Bom	Bom	Regular	Ruim	Ruim

Moda = \_\_\_\_\_

### Média ou mediana?

Ex.: idade, em anos completos, de oito indivíduos.

16 18 15 22 24 23 15 62

Média = 24,38 anos

Mediana = 20 anos

Qual medida descreve melhor a variável idade ?

**Cuidado!** Um pesquisador obteve dados de uma amostra de 18 indivíduos. Estão listadas abaixo as idades (em anos) dos indivíduos.

23	28	24	67	62	59
26	27	24	65	56	64
24	23	26	67	69	63

A média é 44,3 anos. Esta média descreve adequadamente estes dados?

## Medidas de Dispersão

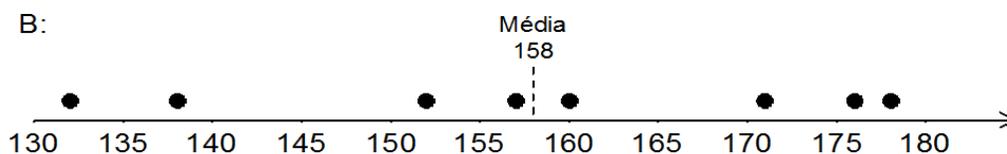
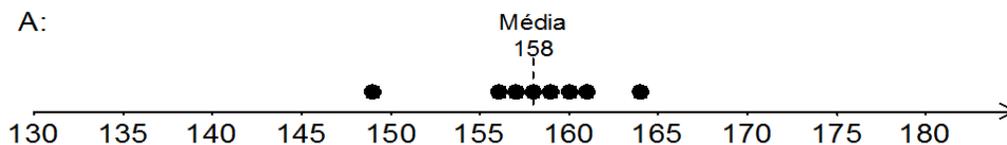
**Medidas de dispersão:** Medem a variabilidade dos dados.

Ex.: Estatura (em cm) observada em duas amostras de adolescentes saudáveis (A e B).

A: 149, 156, 157, 158, 159, 160, 161, 164

B: 132, 138, 152, 157, 160, 171, 176, 178

Amostra	<i>n</i>	média	mediana
A	8	158 cm	158,5 cm
B	8	158 cm	158,5 cm



**Amplitude amostral**

É a diferença entre o maior e o menor valor das observações.

**Ex.:** seja  $X$  o volume expiratório forçado (VEF) em um segundo (em litros) para uma amostra de  $n = 13$  adolescentes que sofrem de asma.

$$\begin{array}{ccccc} x_1 = 2,30 & x_1 = 2,15 & x_3 = 3,50 & x_4 = 2,60 & x_5 = 2,75 \\ x_6 = 2,82 & x_7 = 4,05 & x_8 = 2,25 & x_9 = 2,68 & x_{10} = 3,00 \\ x_{11} = 4,02 & x_{12} = 2,85 & x_{13} = 3,38 & & \end{array}$$

Amplitude amostral =  $4,05 - 2,15 = 1,9$  litros.

**Desvio médio**

**Ex.:** sejam as idades de oito indivíduos (média = 50,25 anos)

38 40 49 67 33 57 54 64

$x$	$x - \bar{x}$	$ x - \bar{x} $
38	$38 - 50,25 = -12,25$	12,25
40	$40 - 50,25 = -10,25$	10,25
49	$49 - 50,25 = -1,25$	1,25
67	$67 - 50,25 = 16,75$	16,75
33	$33 - 50,25 = -17,25$	17,25
57	$57 - 50,25 = 6,75$	6,75
54	$54 - 50,25 = 3,75$	3,75
64	$64 - 50,25 = 13,75$	13,75
Soma =		82

Desvio médio =  $82/8 = 10,25$  anos = 10 anos e 3 meses.

O desvio médio é a distância média das observações individuais a partir de  $\bar{x}$ . Apesar de ser uma intuitiva medida de variabilidade, o desvio médio não é usado na prática. A medida mais usada é o *desvio padrão*.

**Variância**

$x$	$x - \bar{x}$	$(x - \bar{x})^2$
38	$38 - 50,25 = -12,25$	150,06
40	$40 - 50,25 = -10,25$	105,06
49	$49 - 50,25 = -1,25$	1,56
67	$67 - 50,25 = 16,75$	280,56
33	$33 - 50,25 = -17,25$	297,56
57	$57 - 50,25 = 6,75$	45,56
54	$54 - 50,25 = 3,75$	14,06
64	$64 - 50,25 = 13,75$	189,06
Soma =		1083,5

$$\text{variância} = s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = 1083,5 / 7 = 154,8 \text{ anos}^2.$$

**Desvio padrão**

O desvio padrão é a raiz quadrada da variância:  $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = 12,44 \text{ anos.}$

Tem a mesma interpretação do desvio médio.

**Retornando ao exemplo:** Estatura (em cm) observada em duas amostras de adolescentes saudáveis (A e B).

**A:** 149, 156, 157, 158, 159, 160, 161, 164

**B:** 132, 138, 152, 157, 160, 171, 176, 178

Amostra	<i>n</i>	média	mediana	amplitude	desvio padrão	variância
A	8	158 cm	158,5 cm	15 cm	4,407 cm	19,429 cm <sup>2</sup>
B	8	158 cm	158,5 cm	46 cm	16,945 cm	287,142 cm <sup>2</sup>

**Exercícios**

**1. (Samuels, 1989)** Seis homens com altos níveis de colesterol participaram de um estudo que objetivou avaliar os efeitos de uma dieta sobre os níveis de colesterol. No início do estudo, as medidas de seus níveis séricos de colesterol (mg/dl) foram as seguintes:

366                      327                      274                      292                      274                      230

Determine a média amostral e a mediana.

**2. (Samuels, 1989)** Os ganhos de peso de bezerros foram medidos em um período de 140 dias. O ganho regular diário de peso (libras/dia) de 9 bezerros submetidos a uma única dieta são os seguintes:

3,89      3,51      3,97      3,31      3,21      3,36      3,67      3,24      3,27

Determine a média amostral e a mediana.

**3. (Samuels, 1989)** Um pesquisador aplicou um agente carcinogênico na pele de cinco camundongos e mediu a sua concentração em tecidos do fígado após 48 horas. Os resultados (nmoles/gm) são os seguintes:

6,3                      5,9                      7,0                      6,9                      5,9

Determine a média amostral e a mediana.

**4. (Samuels, 1989)** As medidas de pressão sanguínea sistólica (mmHg) de sete homens são:

151                      124                      132                      170                      146                      124                      113

Determine a amplitude, a média amostral e a mediana.

5. As notas de 15 estudantes inscritos no curso de estatística foram as seguintes:

2,0 3,5 3,5 4,5 5,0 5,5 6,5 7,5 7,5 7,5 8,5 8,5 9,5 9,5 10.

Após calcular a média, a mediana e a moda, descobriu-se um erro: uma das notas registradas como 7,5 é na realidade 8,5. As medidas de tendência central que vão mudar serão:

- somente a média.
- somente a moda.
- somente a mediana.
- a média e a moda.
- a média, a moda e a mediana.

6. Uma investigadora pesou cada um dos 30 peixes enviados para análise. Obteve uma média de 30 g e um desvio padrão de 2 g. Após ter completado as pesagens, verificou que a balança estava mal calibrada em menos 2 g, ou seja, um peixe que na realidade pesava 26 g foi registrado com um peso de 24 g (por exemplo). Respectivamente, qual é o valor da média e do desvio padrão após efetuar a devida correção nos dados? (Escolha a alternativa correta).

- 28 g, 2 g.
- 30 g, 4 g.
- 32 g, 2 g.
- 32 g, 4 g.
- 28 g, 4 g.

7. Oito pessoas estão em um elevador, sendo 2 mulheres e 6 homens. O peso médio das mulheres é de 65 kg e o peso médio dos homens é de 80 kg. Qual é o peso médio das 8 pessoas que estão no elevador ?

8. Classifique as variáveis abaixo utilizando cada um dos critérios:

(i) qualitativa ou quantitativa; (ii) discreta, contínua, nominal ou ordinal.

- O número de mortes de câncer no estado de São Paulo no ano de 1996,
- Os conceitos (A, B, C, D, ou E) recebidos em um curso de graduação,
- A pressão sanguínea de um adulto,
- O sexo de uma criança,
- A renda mensal de uma família,
- A idade de um homem,
- O estado civil de uma mulher,
- O número de pacientes que sobreviveram após um tratamento.

9. Os números abaixo se referem aos salários (em reais) de 10 indivíduos.

680,00	685,00	700,00	720,00	735,00
735,00	780,00	800,00	810,00	12.350,00

- Qual medida você usaria para descrever os salários destes indivíduos? Média ou mediana? Justifique a sua resposta.
- calcule a medida que você propôs no item (a) e interprete-a.

### Respostas:

1. A média amostral é 293,8 mg/dl e a mediana é 283 mg/dl.

2. A média amostral é 3,49 libras/dia e a mediana é 3,36 libras/dia.



**Ex.3:** Alturas (em cm) de meninos de uma amostra e alturas (em cm) de homens adultos de uma outra amostra.

Grupo	$n$	Média	Desvio padrão	CV
Meninos	150	50,0	6,0	12,0%
Homens adultos	150	160,0	16,0	10,0%

### Quartis

Os quartis são medidas que permitem dividir a distribuição dos dados em quatro partes iguais quanto ao número de elementos de cada uma. Dado um conjunto ordenado de valores, definimos então:

**Primeiro quartil ( $Q_1$ ):** é o valor que divide o conjunto em duas partes tal que um quarto (ou 25%) dos valores é menor ou igual a  $Q_1$  e três quartos (ou 75%) são maiores ou igual a  $Q_1$ .

**Segundo quartil ( $Q_2$ ):** é igual à mediana. Metade dos valores é menor ou igual a  $Q_2$ , enquanto outra metade é maior ou igual a  $Q_2$ .

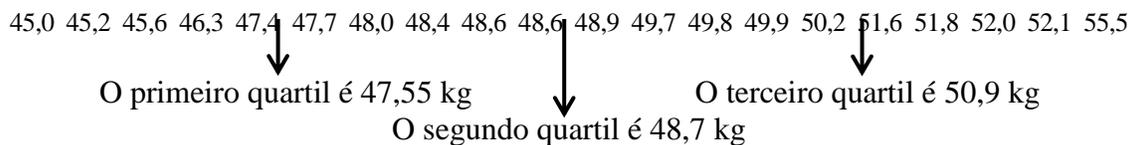
**Terceiro quartil ( $Q_3$ ):** é o valor que divide o conjunto em duas partes tal que três quartos (ou 75%) dos valores são menores ou iguais a  $Q_3$  e um quarto (ou 25%) é maior ou igual.

**Ex.:** Seja o peso de  $n = 20$  indivíduos (em kg)

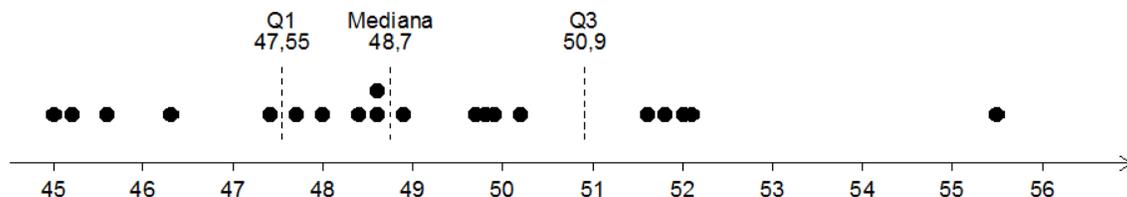
55,5	48,6	49,7	48,6	45,0
45,2	50,2	47,7	52,1	49,9
48,0	48,4	51,6	49,8	51,8
46,3	45,6	52,0	48,9	47,4

O primeiro passo é ordenar os dados...

45,0 45,2 45,6 46,3 47,4 47,7 48,0 48,4 48,6 48,6 48,9 49,7 49,8 49,9 50,2 51,6 51,8 52,0 52,1 55,5



O primeiro quartil é 47,55 kg      O segundo quartil é 48,7 kg      O terceiro quartil é 50,9 kg

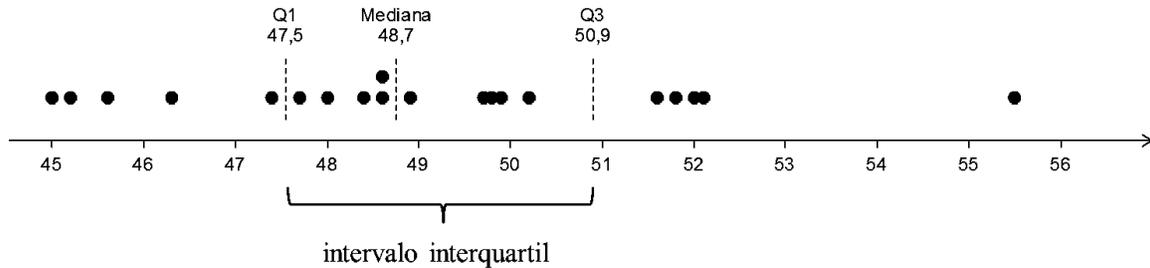


### Intervalo interquartil

O intervalo interquartil é  $(Q_1; Q_3)$ , e sua amplitude  $Q_3 - Q_1$  é uma medida de dispersão.

No exemplo anterior, a amplitude do intervalo interquartil é de  $50,9 \text{ kg} - 47,5 \text{ kg} = 3,4 \text{ kg}$ .

A vantagem desta medida é não ser facilmente influenciada por valores extremos.



### Quantis

**Tercis:** dividem a distribuição dos dados em três partes iguais quanto ao número de elementos de cada uma.

**Quartis:** dividem a distribuição dos dados em quatro partes iguais quanto ao número de elementos de cada uma.

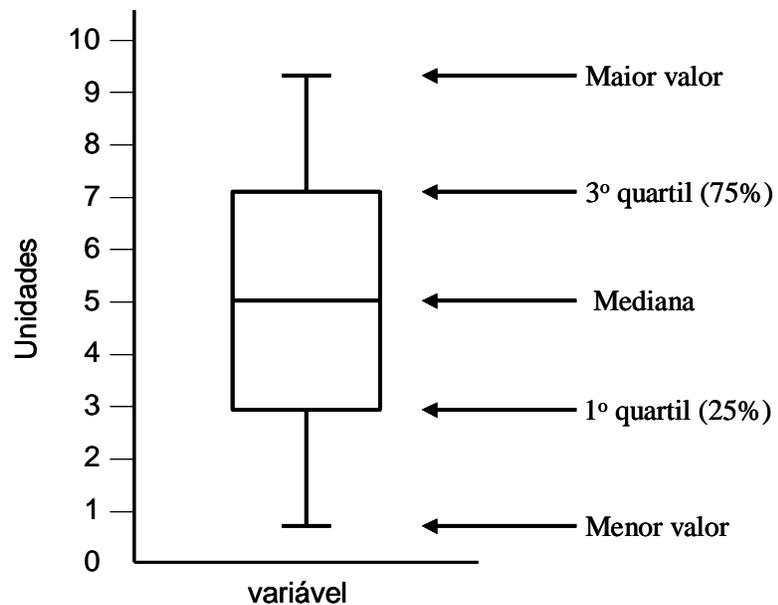
**Quintis:** dividem a distribuição dos dados em cinco partes iguais quanto ao número de elementos de cada uma.

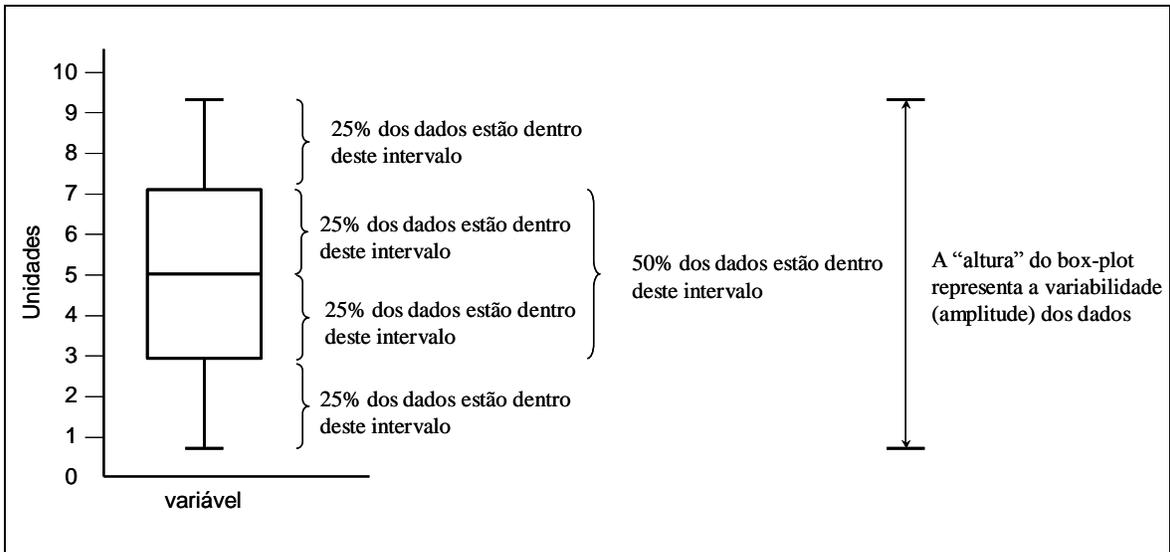
**Sextis:** dividem a distribuição dos dados em seis partes iguais quanto ao número de elementos de cada uma.

**Percentis:** dividem a distribuição dos dados em cem partes iguais quanto ao número de elementos de cada uma.

### Box-plot

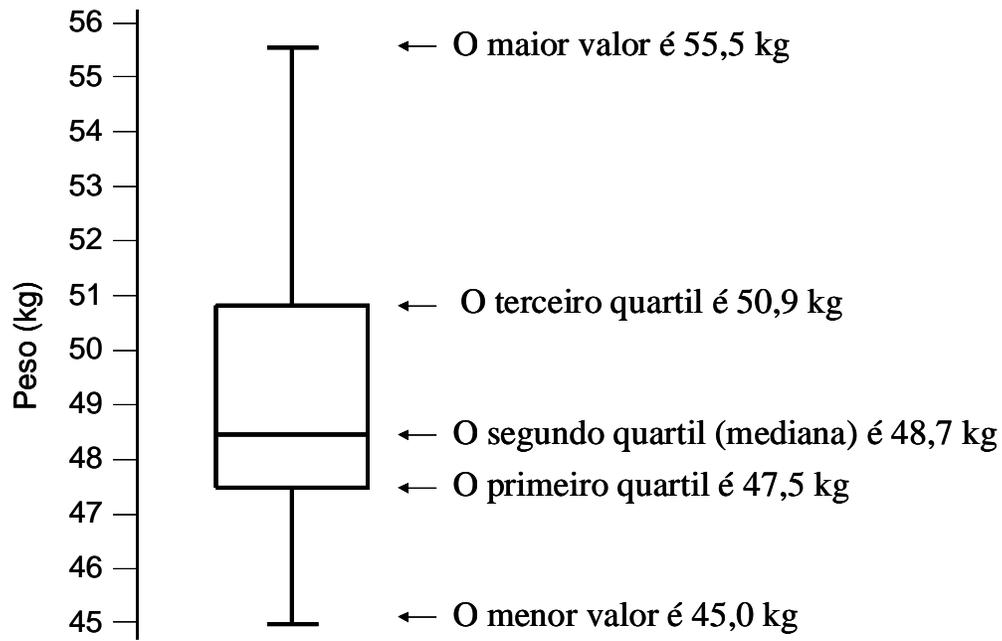
O *box-plot* representa os dados (contínuos) através de um retângulo construído com os quartis e fornece informações sobre os valores extremos.





Ex.: Seja o peso de  $n = 20$  indivíduos (em kg)

55,5	48,6	49,7	48,6
45,0	50,2	47,7	52,1
45,2	48,4	51,6	49,8
48,0	45,6	52,0	48,9
46,3	49,9	51,8	47,4

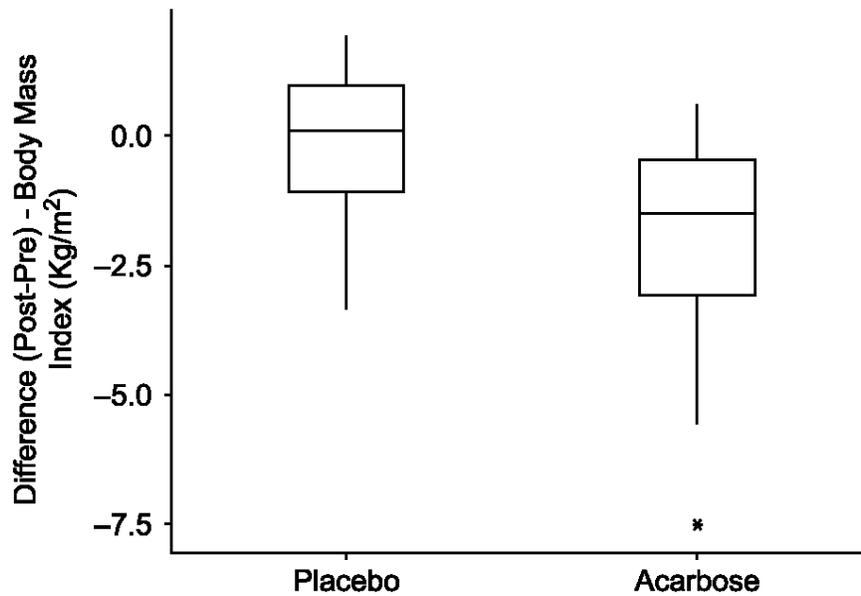


A média amostral é 49,12 kg.

**Exemplos:**

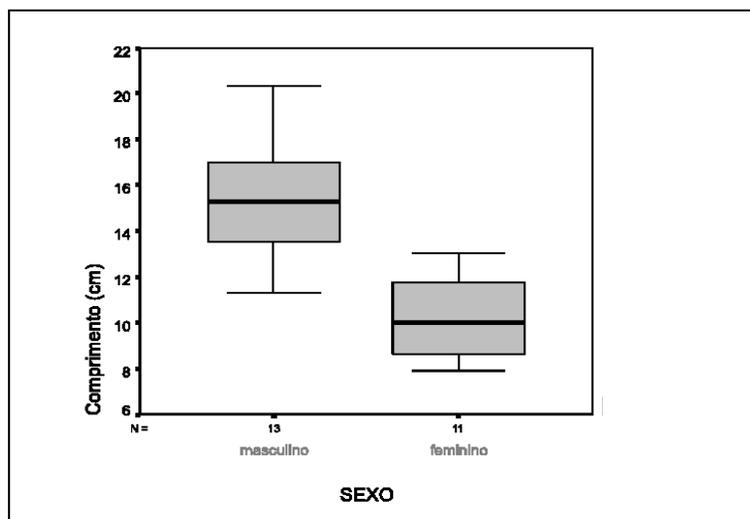
1)

Penna IA, Canella PR, Reis RM, Silva de Sá MF, Ferriani RA. Acarbose in obese patients with polycystic ovarian syndrome: a double-blind, randomized, placebo-controlled study. *Hum Reprod.* 2005 Sep;20(9):2396-401.



**Figure 2.** Box-plot of the difference in body mass index before and after treatment of the women in the acarbose and placebo groups.

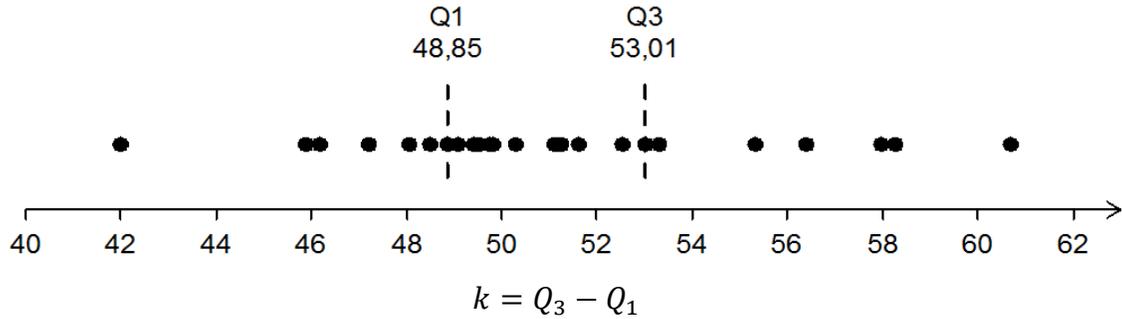
2)



**Gráfico 2.** Comparação entre os comprimentos das pregas vocais quanto ao sexo.

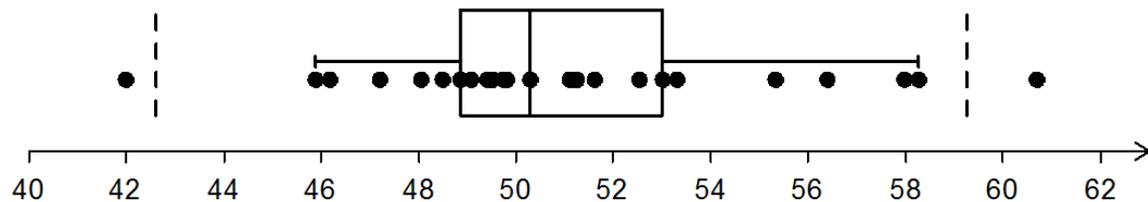
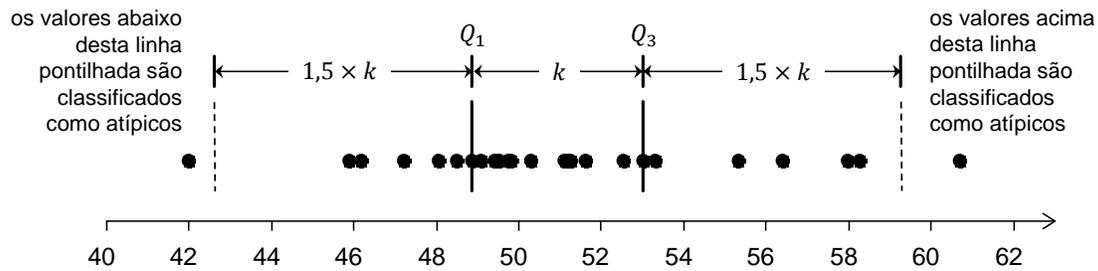
3) Sejam os pesos (em quilogramas) de  $n = 25$  pessoas mensurados por uma balança eletrônica.

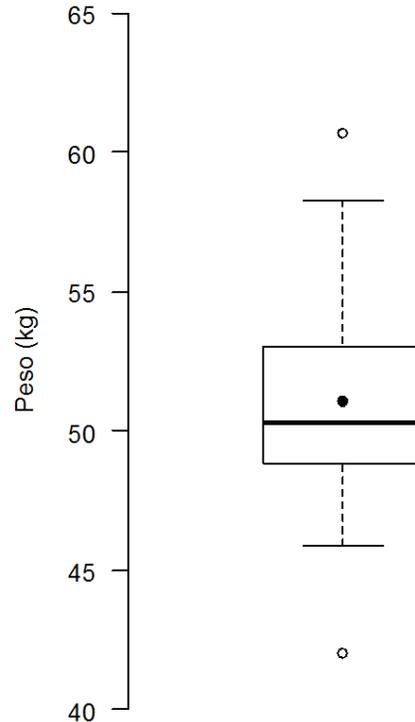
41,99	45,87	46,17	47,19	48,07	48,51	48,85	49,08	49,41
49,51	49,74	49,80	50,31	51,10	51,16	51,24	51,60	52,53
53,01	53,29	55,34	56,39	57,98	58,25	60,70		



$$Q_3 + 1,5 \times k = 53,01 + 1,5 \times 4,16 = 59,25 \text{ kg}$$

$$Q_1 - 1,5 \times k = 48,85 - 1,5 \times 4,16 = 42,61 \text{ kg.}$$

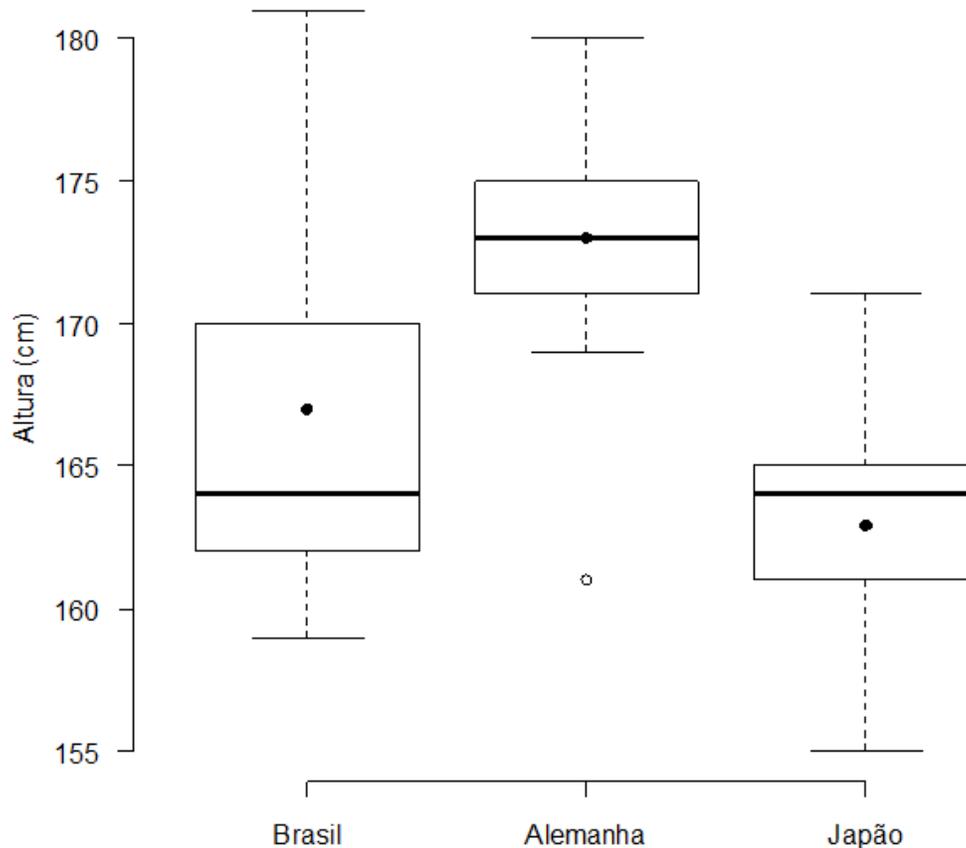




4) Em 2011 foi realizada na Alemanha a Copa do Mundo Feminina de futebol da FIFA. São listadas abaixo as alturas (em centímetros) das atletas que defenderam as seleções do Brasil, Alemanha e Japão.

	159	160	161	162	162	162	163
Brasil	164	164	164	164	167	168	169
	170	170	171	172	174	180	181
	161	169	170	170	170	171	171
Alemanha	172	172	173	173	173	174	174
	175	175	175	176	179	180	180
	155	155	157	157	157	161	162
Japão	163	163	164	164	164	164	164
	164	165	165	168	168	170	171

Estes dados foram obtidos da página eletrônica da FIFA. Os *box-plots* para as alturas das atletas, segundo a seleção que defenderam, encontram-se a seguir.



O que você pode dizer sobre as alturas das atletas, com base na observação da figura acima? Qual medida descreve melhor a altura das atletas, a média ou a mediana?

### Exercícios

**10.** A média do peso dos 100 alunos de uma determinada turma de alunos é 68,4 kg. Esta turma possui 25 homens e 75 mulheres. O peso médio das mulheres desta turma é 62,6 kg. Qual é o peso médio dos homens desta turma?

**11. (Samuels, 1989)** Em um estudo sobre a produção de leite em ovelhas, um pesquisador mediu o rendimento em três meses de cada uma de 11 ovelhas. Os rendimentos (em litros) foram os seguintes:

56,5	89,8	110,1	65,6	63,7	82,6
75,1	91,5	102,9	44,4	108,1	

Encontre a amplitude, os quartis e o intervalo interquartil.

**12. (Samuels, 1989)** Um botânico plantou 15 plantas de pimenta em uma mesma floreira de uma estufa. Após 21 dias, ele mediu o comprimento total (cm) da haste de cada planta, obtendo os seguintes valores:

12,4      12,2      13,4      10,9      12,2      12,1      11,8      13,5  
 12,0      14,1      12,7      13,2      12,6      11,9      13,1

Construa um box-plot para estes dados, indicando no gráfico onde está localizado o intervalo interquartil. Estes dados mostram evidências de valores atípicos?

**13.** Estão listados abaixo os comprimentos dos cascos de uma amostra aleatória de 48 tartarugas, em centímetros, sendo 24 machos e 24 fêmeas.

machos	35	35	35	37	37	38	38	39	39	40	40	40
	40	41	41	41	42	43	44	45	45	45	46	47
fêmeas	38	38	42	42	44	46	48	49	50	51	51	51
	51	51	53	55	56	57	60	61	62	63	63	67

(a) Complete a tabela de estatísticas descritivas abaixo para a variável “comprimento dos cascos”, segundo o sexo da tartaruga.

Sexo	média	desvio padrão	Variância	coeficiente de variação	mediana	primeiro quartil	terceiro quartil
macho	40,54	3,54					
fêmea	52,04	8,05					

- (b) Construa um *box-plot* para as medidas de comprimento dos cascos de cada sexo. Utilize a mesma escala nos dois *box-plots* para que eles sejam comparáveis. Os *box-plots* construídos indicam alguma evidência de que os cascos das tartarugas fêmeas tendem a ser diferentes dos cascos dos machos, em relação ao comprimento ?
- (c) Estes dados mostram evidências de valores atípicos?

**14.** Em um estudo sobre a influência do treinamento anaeróbico na massa de gordura corporal de adolescentes obesos, uma fisioterapeuta distribuiu ao acaso 54 indivíduos entre dois grupos: (1) grupo de treinamento: treinamento físico anaeróbico, com duração de três meses, com orientação nutricional, consulta à nutricionista a cada mês; e (2) grupo controle: não realizaram nenhum tipo de treinamento físico durante três meses, com orientação nutricional, com consulta à nutricionista a cada mês. A pesquisadora mediu a massa de gordura de membros inferiores, expressa em quilogramas, no período final do estudo. Estas medidas estão apresentadas na tabela a seguir, segundo o grupo.

Grupo	Massa de gordura de membros inferiores (em quilogramas)									média	desvio padrão
(1)	10,45	10,78	10,94	11,31	11,34	11,53	12,01	12,07	12,14	12,59	1,13
	12,15	12,17	12,38	12,39	12,40	12,76	12,84	12,84	12,89		
	13,07	13,21	13,23	13,44	13,80	14,09	14,40	14,56	14,66		
(2)	7,34	11,14	11,33	11,45	11,60	12,46	12,49	14,12	14,21	15,37	3,23
	14,33	14,63	14,81	14,95	15,39	15,44	16,53	16,73	16,79		
	17,69	17,81	18,04	18,39	18,40	19,01	19,25	19,62	21,05		

(a) Para a variável massa de gordura de membros inferiores, construa um *box-plot* para cada um dos dois grupos. Utilize a mesma escala nos dois *box-plots* para que eles sejam comparáveis (coloque os dois *box-plots* juntos, em uma única figura).

- (b) Interprete a figura que você obteve no item (a), e escreva se ela evidencia alguma relação entre os grupos e a massa de gordura de membros inferiores.
- (c) Estes dados mostram evidências de valores atípicos?

15. O espelho de Glatzel é um instrumento de avaliação da permeabilidade nasal extremamente simples. Consiste em uma placa de metal polida e graduada que, ao ser posicionada sob as narinas, condensa o vapor d'água do ar expirado. A área embaçada é então mensurada e a avaliação da função nasal pode ser realizada. O instrumento pode ser empregado como recurso para visualização e mensuração do escape aéreo nasal, sendo também empregado como auxiliar ao diagnóstico de obstrução mecânica ou desuso nasal. Foi realizado um estudo com uma amostra composta por 43 crianças, com idades variando entre quatro e 11 anos, que se encontravam em atendimento no Ambulatório do Respirador Oral de um grande hospital público. As crianças foram separadas em quatro grupos:

- casos cirúrgicos ( $n=7$ ),
- casos alérgicos com manifestação de obstrução ( $n=11$ ),
- casos cirúrgicos e concomitantemente alérgicos ( $n=10$ ),
- grupo controle ( $n=15$ ), sem obstrução.

A tabela a seguir mostra a média ( $\bar{x}$ ) e o desvio padrão ( $s$ ) para as medidas do escape aéreo nasal (em  $\text{cm}^2$ ), por grupo, sendo a média para o grupo controle denotada por  $a$ .

Grupo	$n$	$\bar{x}$	$s$
Alérgico	11	4,74	2,49
Alérgico e cirúrgico	10	3,07	2,71
Cirúrgico	7	5,91	3,13
Controle	15	$a$	2,13

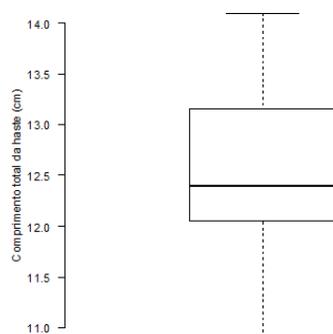
Considerando a média do escape aéreo nasal para toda a amostra de  $n = 43$  crianças igual a  $5,17 \text{ cm}^2$ , encontre o valor de  $a$ .

### Respostas:

10. Se  $\frac{25x + 75 \times 62,6}{100} = 68,4$ , então  $x = 85,8 \text{ kg}$ .

11. A amplitude é 65,7 litros, o primeiro quartil é 63,7 litros, o segundo quartil é 82,6 litros, o terceiro quartil é 102,9 litros e o intervalo interquartil é  $102,9 - 63,7 = 39,2$  litros.

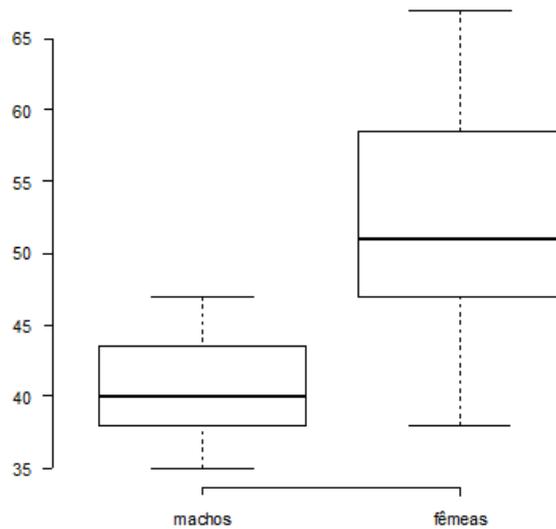
12.



(encontre no gráfico o intervalo interquartil).

13. (a)

Sexo	média	desvio padrão	Variância	coeficiente de variação	mediana	primeiro quartil	terceiro quartil
macho	40,54	3,54	12,52	8,7%	40	38	43,5
fêmea	52,04	8,05	64,74	15,5%	51	47	58,5



(b)

(escreva se os *box-plots* construídos indicam alguma evidência de que os cascos das tartarugas fêmeas tendem a ser diferentes dos cascos dos machos, em relação ao comprimento).



## Algumas considerações...

**Ex.1:**  $X$  : altura dos pacientes (em centímetros)

160    159    168    160    153    161    157    156    150    161

$$\text{A média amostral é } \bar{x} = \frac{160+159+\dots+161}{10} = 158,5 \text{ cm}$$

**Ex.2:**  $X$  : altura dos pacientes (em metros)

1,60    1,59    1,68    1,60    1,53    1,61    1,57    1,56    1,50    1,61

$$\text{A média amostral é } \bar{x} = \text{_____ m}$$

**Ex.3:**  $X$  : altura dos pacientes (em centímetros)

170    169    178    170    163    171    167    166    160    171

$$\text{A média amostral é _____ cm}$$

**Ex.4:**  $X$  : altura dos pacientes (em centímetros)

160    159    168    160    153    161    157    156    150    161

$$\text{Desvio padrão amostral: } s = 4,927 \text{ cm.}$$

**Ex.5:**  $X$  : altura dos pacientes (em metros)

1,60    1,59    1,68    1,60    1,53    1,61    1,57    1,56    1,50    1,61

$$\text{O desvio padrão amostral é _____ m}$$

**Ex.6:**  $X$  : altura dos pacientes (em centímetros)

170    169    178    170    163    171    167    166    160    171

$$\text{O desvio padrão amostral é _____ cm}$$

**Ex.7:**  $X$  : altura dos pacientes (em centímetros)

160    159    168    160    153    161    157    156    150    161

Média amostral = 158,5 cm  
Desvio padrão amostral = 4,927 cm

Coefficiente de variação:  $\frac{4,927 \text{ cm}}{158,5 \text{ cm}} = 3,11 \%$

**Ex.8:**  $X$  : altura dos pacientes (em metros)

1,60    1,59    1,68    1,60    1,53    1,61    1,57    1,56    1,50    1,61

O coeficiente de variação é: \_\_\_\_\_

### Uma nota sobre o cálculo do desvio padrão

O desvio padrão é dado por  $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$

Vamos supor que desejamos encontrar o desvio padrão utilizando uma calculadora...

Notar que  $\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}$ , e, portanto,  $s = \sqrt{\frac{\sum_{i=1}^n x_i^2 - n\bar{x}}{n-1}}$

A expressão pode ser mais conveniente quando usamos a calculadora.

1-) Usando  $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$ ,

$x$	$x - \bar{x}$	$(x - \bar{x})^2$
160	$160 - 158,5 = 1,5$	2,25
168	$168 - 158,5 = 9,5$	90,25
160	$160 - 158,5 = 1,5$	2,25
159	$159 - 158,5 = 0,5$	0,25
153	$153 - 158,5 = -5,5$	30,25
161	$161 - 158,5 = 2,5$	6,25
157	$157 - 158,5 = -1,5$	2,25
156	$156 - 158,5 = -2,5$	6,25
150	$150 - 158,5 = -8,5$	72,25
161	$161 - 158,5 = 2,5$	6,25

Temos  $\bar{x} = 158,5$  cm e

$$\sum_{i=1}^n (x_i - \bar{x})^2 = 218,50,$$

Portanto, o desvio padrão amostral é

$$s = \sqrt{\frac{218,5}{9}} = 4,93 \text{ cm.}$$

2-) Usando  $s = \sqrt{\frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1}}$ ,

$x$	$x^2$	
160	25.600	Temos $\bar{x} = 158,5$ cm e $n \bar{x}^2 = 10 \times 158,5^2 = 251.222,5$  Portanto,
168	28.224	
160	25.600	
159	25.281	
153	23.409	
161	25.921	
157	25.921	
156	24.649	
150	24.336	
161	22.500	
Soma = 251.441		$s = \sqrt{\frac{251.441 - 251.222,5}{9}} = \sqrt{\frac{218,5}{9}} = 4,93 \text{ cm.}$

### Exercícios

**16.** Para investigar a composição química e nutricional da alimentação oferecida em um restaurante universitário, foram coletadas amostras das refeições durante cinco dias consecutivos (Fausto e colaboradores, 2001). Para que as amostras correspondessem ao total de alimentos habitualmente porcionados pelos funcionários do restaurante, elas foram coletadas no utensílio próprio utilizado no local (bandeirão), durante o horário de funcionamento do restaurante (almoço e jantar). Para evitar modificações nos cardápios ou nas quantidades de alimentos porcionados, os funcionários do local não foram informados a respeito da pesquisa.

A Tabela a seguir apresenta a composição química das refeições (almoço ou jantar) oferecidas no Restaurante Universitário, em g/100 g da refeição. No entanto, algumas informações foram perdidas. Estas informações perdidas estão indicadas na Tabela com um ponto de interrogação (?).

Composição química <sup>(1)</sup>	Cardápio					média	dp <sup>(2)</sup>
	1	2	3	4	5		
Proteínas	4,6	5,0	4,9	5,2	4,0	4,74	? (a)
Lipídeos	7,9	12,5	9,5	? (b)	14,2	10,84	2,50
Carboidratos	26,6	27,4	28,0	22,8	? (c)	24,32	? (d)

<sup>(1)</sup> g/100 g da refeição.

<sup>(2)</sup> dp = desvio padrão.

Complete a tabela, encontrando as informações perdidas (a), (b), (c) e (d).

**17. (Samuels, 1989)** A dopamina é uma substância que exerce uma função importante na transmissão de sinais no cérebro. Um farmacologista mediu a quantidade de dopamina no cérebro de cada um de sete ratos. Os níveis de dopamina (nmoles/g) são os seguintes:

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

- (a) Encontre a média e o desvio padrão.  
 (b) Encontre a mediana e o intervalo interquartil.  
 (c) Calcule o coeficiente de variação  
 (d) Substitua a observação 7,4 por 10,4 e repita os itens (a) e (b). Quais medidas descritivas sofreram alteração ?

**18. (Samuels, 1989)** Em um estudo sobre o lagarto *Sceloporus occidentalis*, biólogos mediram a distância (m) percorrida em dois minutos por cada um de 15 animais. Os resultados (listados em ordem crescente) são listados a seguir:

18,4      22,2      24,5      26,4      27,5      28,7      30,6      32,9  
 32,9      34,0      34,8      37,5      42,1      45,5      45,5

- (a) Determine os quartis e o intervalo interquartil.  
 (b) Determine a amplitude.

**19. (Samuels, 1989)** Estão listados abaixo em ordem crescente os níveis séricos de creatina fosfoquinase (CK, medidos em U/l) de 36 homens saudáveis.

25    62    82    95    110    139  
 42    64    83    95    113    145  
 48    67    84    100    118    151  
 57    68    92    101    119    163  
 58    70    93    104    121    201  
 60    78    94    110    123    203

Construa o *box-plot* para estes dados.

**20. (Samuels, 1989)** A anfetamina é uma droga que reduz o apetite. Em um estudo sobre este efeito desta droga, uma farmacologista aleatoriamente alocou 24 ratos em três grupos de tratamento para receber uma injeção de anfetamina em um de dois diferentes níveis de dosagens, ou uma injeção de uma solução salina. Ela mediu a quantidade de alimento consumido por cada animal em um período de três horas, sendo os resultados (em g de alimento consumido por kg de peso do animal) mostrados na tabela ao lado. Construa um *box-plot* apropriado a estes dados, e escreva se esta figura evidencia alguma relação entre a dose de anfetamina e a quantidade de alimento consumido.

	Dose de anfetamina (mg/kg)		
	0	2,5	5,0
	112,6	73,3	38,5
	102,1	84,8	81,3
	90,2	67,3	57,1
	81,5	55,3	62,3
	105,6	80,7	51,5
	93,0	90,0	48,3
	106,6	75,5	42,7
	108,3	77,1	57,9

**21.** Quatro alunos matriculados em uma disciplina fizeram cinco provas. Suas notas são mostradas na Tabela abaixo.

Aluno	notas				
Antônio	5	5	5	5	5
João	6	4	5	4	6
José	10	5	5	5	0
Pedro	10	10	5	0	0

- (a) Calcule a média e a variância das notas de cada aluno, considerando que todas as provas têm o mesmo peso.  
 (b) Qual aluno teve maior variação em suas notas ?

**22.** A Tabela abaixo foi extraída de um estudo desenvolvido por Amorim Filho e colaboradores (2003) sobre as variáveis demográficas, ocupacionais e co-carcinogênicas associadas carcinoma espinocelular da base de língua em mulheres. No entanto, algumas informações da Tabela foram perdidas. Estas informações perdidas são indicadas por um ponto de interrogação (?).

**Tabela** – Pacientes portadores de carcinoma epidermóide de base de língua, segundo a idade, em anos.

grupo etário (classes)	frequências absolutas	frequências relativas (%)
30 a 39	3	? (a)
40 a 49	4	12,9
50 a 59	12	? (b)
60 a 69	6	19,4
70 a 79	? (c)	16,1
80 ou mais	? (d)	3,2
Total	? (e)	100,0

Complete a tabela, encontrando as informações perdidas (a), (b), (c), (d) e (e).

**23.** A Tabela abaixo mostra o peso (kg) e a altura (cm) de 10 pessoas de uma amostra.

Indivíduo	1	2	3	4	5	6	7	8	9	10
Peso (kg)	79	83	57	52	67	70	50	53	75	68
Altura (cm)	172	181	165	156	172	180	162	156	171	173

- (a) Calcular a média, a variância, o desvio padrão e o coeficiente de variação.  
 (b) Transformar os dados de altura para metros e recalculer as medidas de posição e dispersão. Compare os resultados e discuta.

**24. (Soares & Siqueira, 2002)** Consideremos 12 observações do tempo de internação (em dias) de pacientes acidentados no trabalho, em um certo hospital: 1, 4, 7, 9, 10, 13, 15, 17, 17, 18, 19, 21. Obtenha os quartis e interprete estes valores.

**25. (Soares & Siqueira, 2002)** O tempo (em meses) entre a remissão de uma doença e a recidiva de 48 pacientes de uma determinada clínica médica foi registrado. Os dados ordenados são apresentados a seguir, separadamente para os sexos masculino (M) e feminino (F).

M	2	2	3	4	4	4	4	7	7	7	8	9
	9	10	12	15	15	15	16	18	18	22	22	24
F	2	2	3	3	4	4	5	5	6	6	7	7
	7	7	8	8	8	8	10	10	11	11	12	18

- (a) Calcule a média, o desvio padrão, a mediana e o coeficiente de variação para cada sexo. Interprete.  
 (b) Repita os cálculos pedidos em (a) para todos os 48 pacientes. Compare com os resultados de (a).

**Respostas**

16. (a) 0,47 g/100 g      (b) 10,1 g/100 g      (c) 16,8 g/100 g      (d) 4,67 g/100 g

17. (a) a média é 6,34 nmoles/g e o desvio padrão é 0,70 nmoles/g.  
(b) a mediana é 6,2 nmoles/g e o intervalo interquartil é 0,85 nmoles/g  
(c) 11,1%

21. (a)

	Média	Desvio padrão
Antônio	5	0,0
João	5	1,0
José	5	3,5
Pedro	5	5,0

22. (a) 9,7%      (b) 38,7%      (c) 5      (d) 1      (e) 31

## Correlação

**Objetivo:** Estudar a relação entre duas variáveis quantitativas.

O coeficiente de correlação é uma medida numérica da “força” da relação ou associação entre duas variáveis quantitativas contínuas  $X$  e  $Y$ .

Para entendermos o coeficiente de correlação, vamos introduzir outra medida, chamada covariância. A covariância entre  $X$  e  $Y$  é uma medida do quanto uma das variáveis se modifica quando a outra se modifica. Se  $\bar{x}$  é a média amostral de  $X$  e  $\bar{y}$  é a média amostral de  $Y$ , a covariância amostral é dada pela expressão

$$cov(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Notar que  $cov(X, Y) = cov(Y, X)$ . Esta expressão também pode ser escrita na forma

$$cov(X, Y) = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{n - 1}$$

Se a covariância é igual a zero, entendemos que conforme as observações de uma das variáveis crescem, as observações da outra variável não tendem a crescer ou decrescer. Ou seja, não há então uma relação linear entre estas variáveis. Se a covariância é maior que zero, entendemos que conforme as observações de uma das variáveis crescem, as observações da outra variável tendem a crescer também. E, se a covariância é menor que zero, entendemos que conforme as observações de uma das variáveis crescem, as observações da outra variável tendem a decrescer.

Vamos considerar, como exemplo, estes dados:

	X	Y		X	Y
	Idade (anos)	Níveis séricos de triglicérides (mg/dl)		Idade (anos)	Níveis séricos de triglicérides (mg/dl)
1	51	189	14	48	54
2	48	72	15	59	319
3	57	109	16	56	267
4	48	140	17	49	94
5	49	88	18	52	218
6	47	135	19	47	56
7	49	229	20	50	100
8	52	125	21	52	169
9	45	87	22	56	145
10	55	163	23	48	122
11	45	136	24	51	98
12	54	255	25	51	188
13	51	231	26	58	179

A média da idade é  $\bar{x} = 51,1$  anos e a média dos níveis séricos de triglicérides é  $\bar{y} = 152,6$  mg/dl.

A covariância entre  $X$  e  $Y$  é dada por

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^{26} (x_i - 51,1)(y_i - 152,6)}{25} = 158,19.$$

Como a covariância é maior que zero, entendemos que pessoas mais idosas tendem a apresentar níveis séricos de triglicérides mais elevados. No entanto, como interpretamos a magnitude da covariância? Ou seja, este valor obtido, 158,19, pode nos indicar o “tamanho” da associação entre estas variáveis?

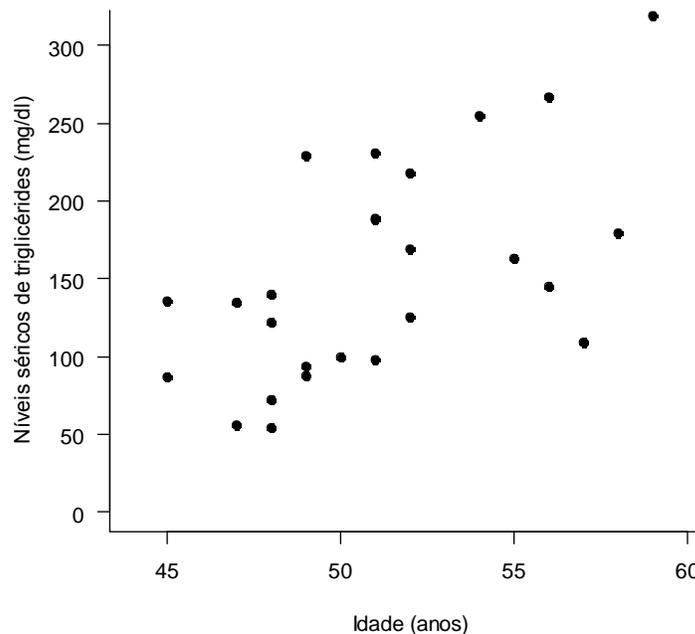
Não é simples “medir” a associação entre duas variáveis com base no valor da variância. Por isso, é bastante usual outra medida, dada pela divisão entre a covariância de  $X$  e  $Y$  e o produto dos desvios padrão amostrais destas variáveis. Esta medida é chamada de **coeficiente de correlação de Pearson**, dada matematicamente por

$$r = \frac{\text{cov}(X, Y)}{S_X S_Y}$$

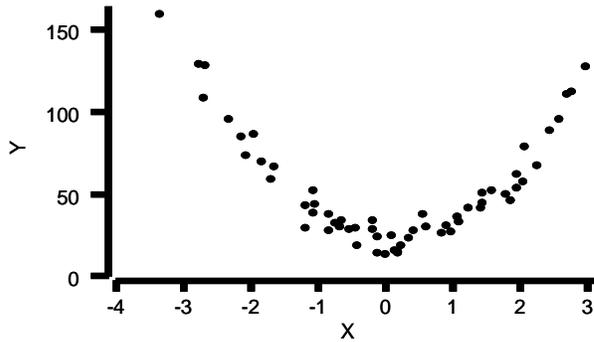
Esta quantidade  $r$  sempre assume valores entre -1 e 1.

O desvio padrão da idade é  $S_X = 3,9$  anos e o desvio padrão dos níveis séricos de triglicérides é  $S_Y = 68,93$  mg/dl. Portanto, o coeficiente de correlação entre estas variáveis é

$$r = \frac{158,19}{3,9 \times 68,93} = 0,588.$$



O coeficiente de correlação mede a associação linear entre duas variáveis contínuas. Portanto, antes de calcular  $r$ , é aconselhável construir um gráfico de dispersão entre as variáveis para verificar se o pressuposto de linearidade é satisfeito. Portanto, se construirmos o gráfico e encontrarmos, por exemplo, uma dispersão como a ilustrada na Figura ao lado, não é adequado o uso de  $r$ . Ainda assim, é possível calcular  $r$  para estes dados ( $r = -0,14$ ), mas o valor encontrado será enganoso e levará o pesquisador a conclusões equivocadas.



Idade (x)	Pressão sistólica (y)	$x_i y_i$
42	130	5.460
46	115	5.290
42	148	6.216
71	100	7.100
80	156	12.480
74	162	11.988
70	151	10.570
80	156	12.480
85	162	13.770
72	158	11.376
64	155	9.920
81	160	12.960
41	125	5.125
61	150	9.150
75	165	12.375
Soma =		146.260

Idade e pressão sistólica de 15 mulheres

Le CT, Boen JR (1995)

Temos  $\bar{x} = 65,6$ ,  $\bar{y} = 146,2$ ,  $s_X = 15,59$ ,  $s_Y = 19,48$

$$cov(X, Y) = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{n - 1}$$

$$= \frac{146.260 - 15 \times 65,6 \times 146,2}{14} = 171,37$$

$$r = \frac{171,37}{15,59 \times 19,48} = 0,564$$

**Interpretação (subjativa) do coeficiente de correlação**

Zou KH, Tuncali K, Silverman SG.

Correlation and simple linear regression.

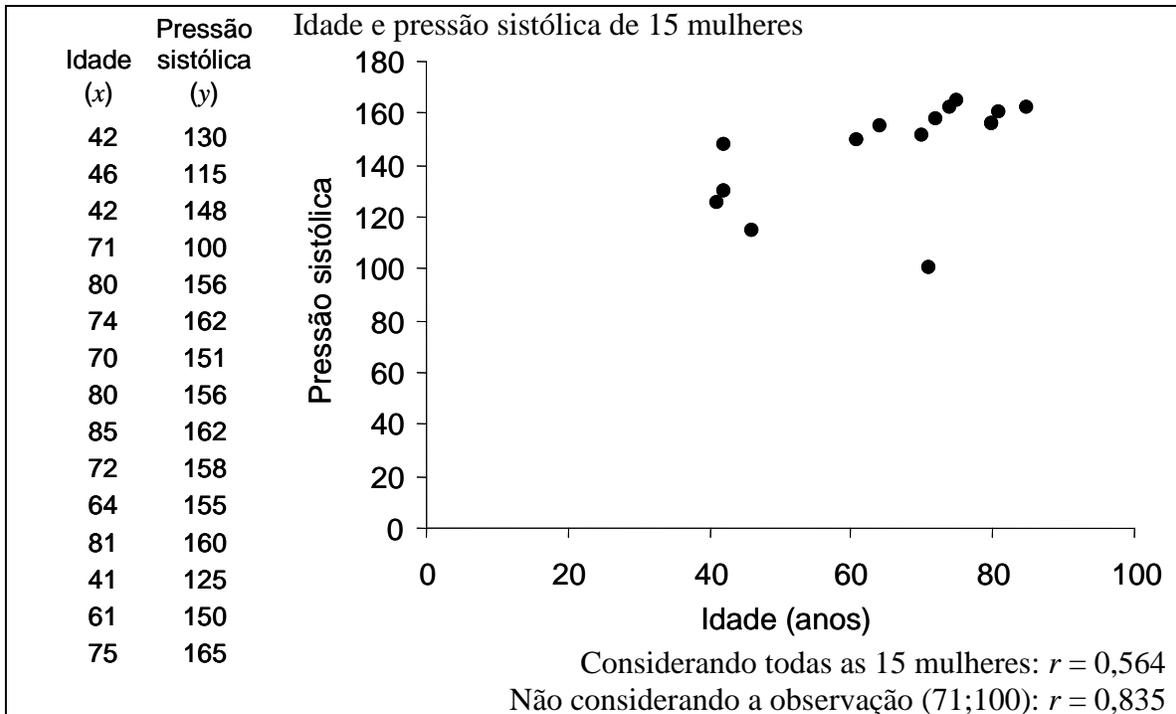
Radiology. 2003;227(3):617-22.

**TABLE 1**  
**Interpretation of Correlation Coefficient**

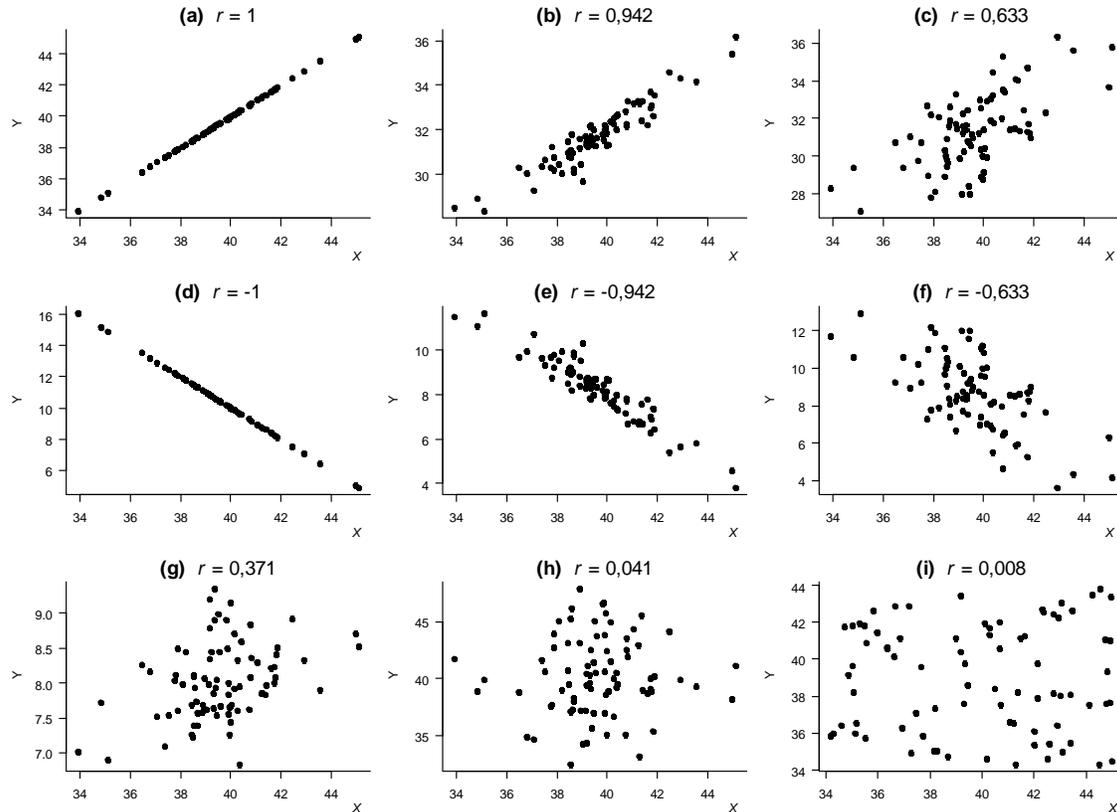
Correlation Coefficient Value	Direction and Strength of Correlation
-1.0	Perfectly negative
-0.8	Strongly negative
-0.5	Moderately negative
-0.2	Weakly negative
0.0	No association
+0.2	Weakly positive
+0.5	Moderately positive
+0.8	Strongly positive
+1.0	Perfectly positive

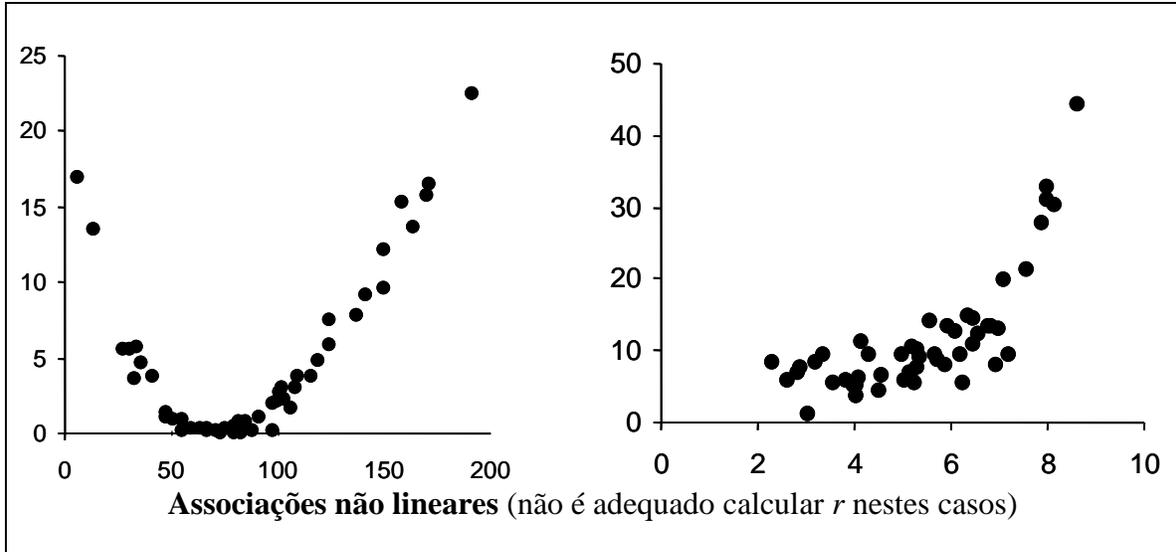
Note.—The sign of the correlation coefficient (ie, positive or negative) defines the direction of the relationship. The absolute value indicates the strength of the correlation.

	A	B	C	D	E	F
1						
2						
3						
4		42	130		0.564224	
5		46	115			
6		42	148			
7		71	100			
8		80	156			
9		74	162			
10		70	151			
11		80	156			
12		85	162			
13		72	158			
14		64	155			
15		81	160			
16		41	125			
17		61	150			
18		75	165			
19						
20						



Exemplos:





Associação e causa: Devemos ser cuidadosos para distinguirmos a diferença entre *associação* e *causa*. Duas variáveis têm associação se a distribuição de uma é afetada pelo conhecimento do valor da outra. Mas isso não significa que uma *causa* a outra. *Associação* não implica *causa*.

# Gráficos

## Referência

Normas para apresentação de documentos científicos  
Volume 10 – Gráficos  
Editora da UFPR, Curitiba, 2000.  
ISBN 85-7335-047-4



O gráfico, se bem construído, consegue transmitir uma idéia com muita rapidez e de forma simples e atraente, levando o leitor a poupar tempo e a despendar menor esforço na compreensão de uma série de dados, os quais são muitas vezes de difícil percepção na forma tabular.

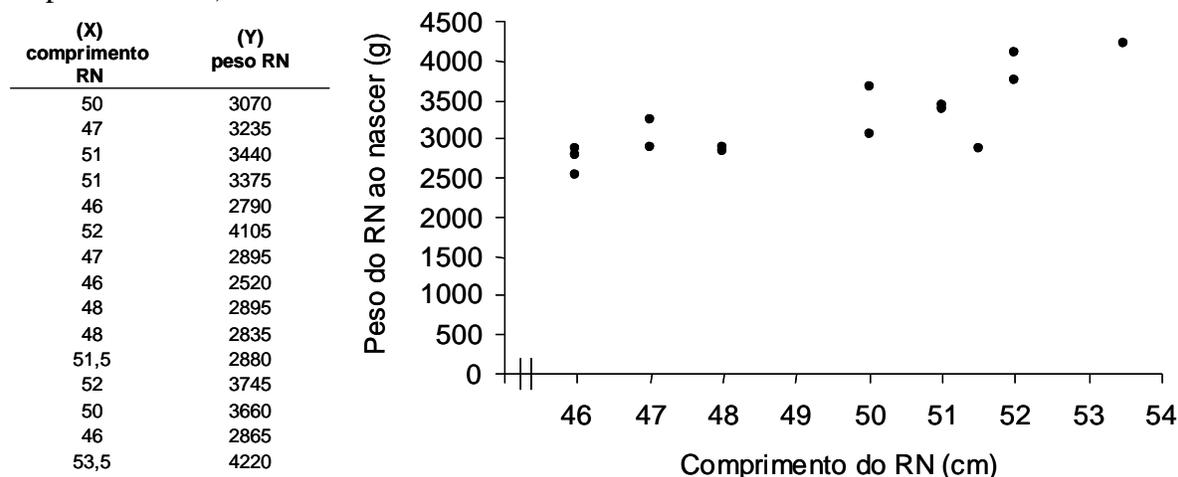
No entanto, se a relação entre os dados apresentados no gráfico não está clara, este deve ser descartado, pois não contribuirá na compreensão destes dados.



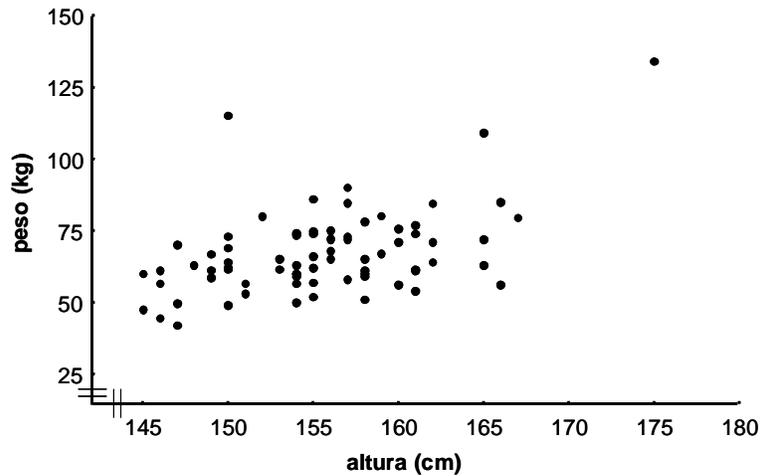
“Remember that graphs must tell their own story; they should be complete in themselves and require little or no additional explanation.”  
(Le e Boen, 1995.)

## Diagrama de pontos ou de dispersão

Cada par de observações é representado por um ponto no cruzamento do sistema de coordenadas cartesianas (abscissa e ordenada, normalmente intituladas como eixos X e Y, respectivamente).



É útil para estudar a correlação entre duas variáveis.

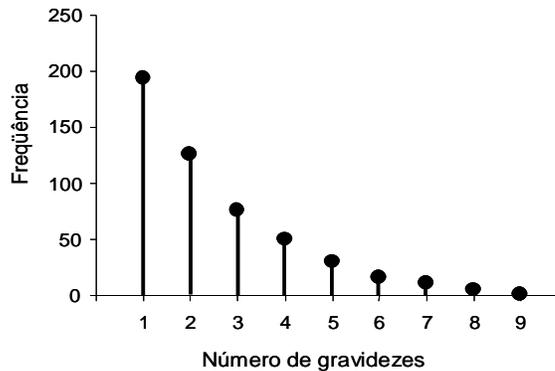


### Diagrama de bastões

A representação da variável é feita no eixo das abscissas (X) e sua frequência no eixo das ordenadas (Y), sendo os dados representados por um ponto e ligados por uma linha até o eixo X, formando os bastões.

gravidezes	freqüência
1	194
2	126
3	76
4	50
5	30
6	16
7	11
8	5
9	1

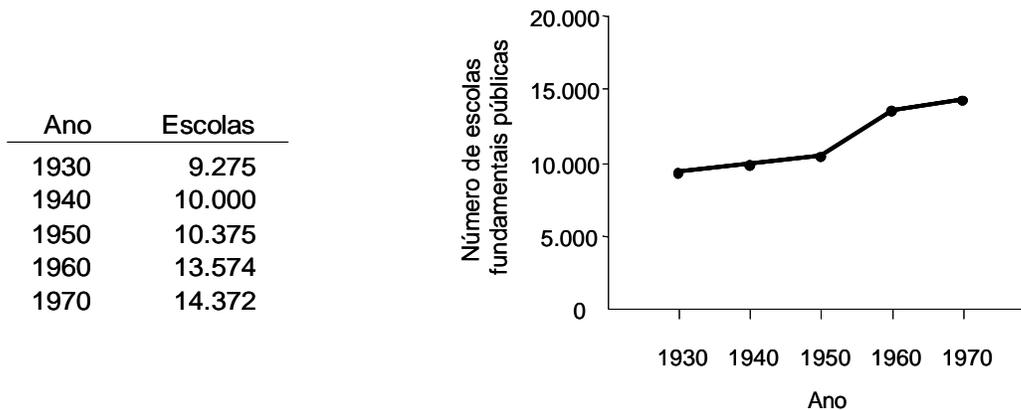
Total: 509 mulheres



### Diagrama de linhas

Possibilita identificar a variação dos dados em uma série cronológica.

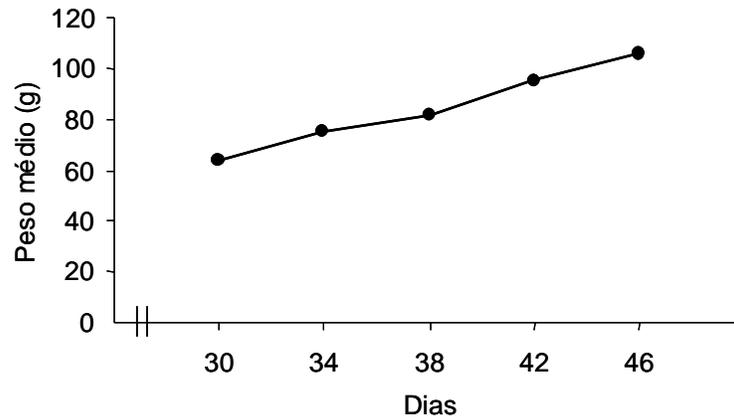
**Ex.1:** Número de escolas fundamentais públicas, de 1930 a 1970.



**Ex.2:** Um pesquisador anotou a idade, em dias, e o peso, em gramas, de 40 ratos da raça Wistar.

	30 dias	34 dias	38 dias	42 dias	46 dias
	76,2	95,5	99,2	122,7	134,6
	81,5	90,0	101,2	125,9	136,2
	50,0	60,0	62,3	72,2	85,3
	47,5	50,0	57,5	72,3	84,0
	63,5	79,2	82,1	94,7	110,0
	65,1	75,7	79,3	88,5	98,7
	63,2	74,8	79,0	88,1	100,0
	64,5	74,1	92,6	96,0	98,3
Média	63,94	74,91	81,65	95,05	105,89
Desvio padrão	11,50	14,71	15,95	20,17	20,02
CV	18,0%	19,6%	19,5%	21,2%	18,9%

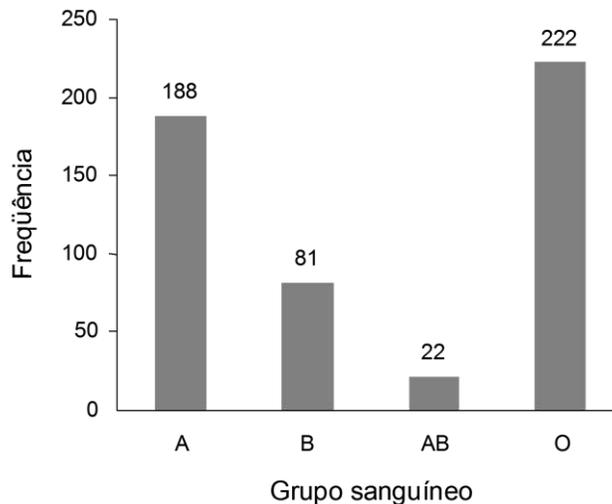
Dias	Peso médio (g)
30	63,94
34	74,91
38	81,65
42	95,05
46	105,89



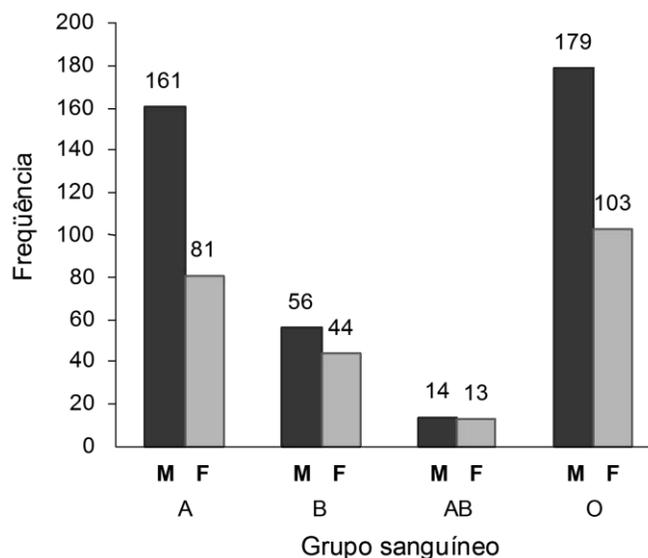
### Diagrama de colunas

Utiliza-se de colunas sucessivas com bases iguais e alturas proporcionais aos valores da série.

Grupo sanguíneo	freqüência
A	188
B	81
AB	22
O	222



Grupo	Sexo		Total
	M	F	
A	161	81	242
B	56	44	100
AB	14	13	27
O	179	103	282
Total	410	241	651

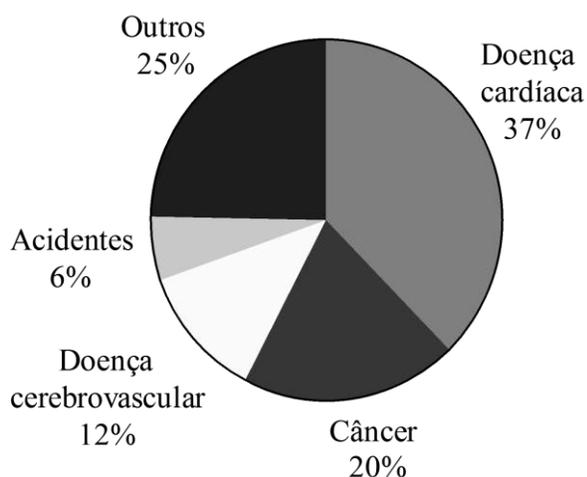


(porém...será que este gráfico está correto?)

### Gráfico de setores

**Ex.:** Número de óbitos devidos a diferentes causas entre os residentes de Minnesota, no ano de 1975.

Causa de óbito	Número de óbitos
Doença cardíaca	12.378
Câncer	6.448
Doença cerebrovascular	3.958
Acidentes	1.814
Outros	8.088
<b>Total</b>	<b>32.686</b>



## Histograma

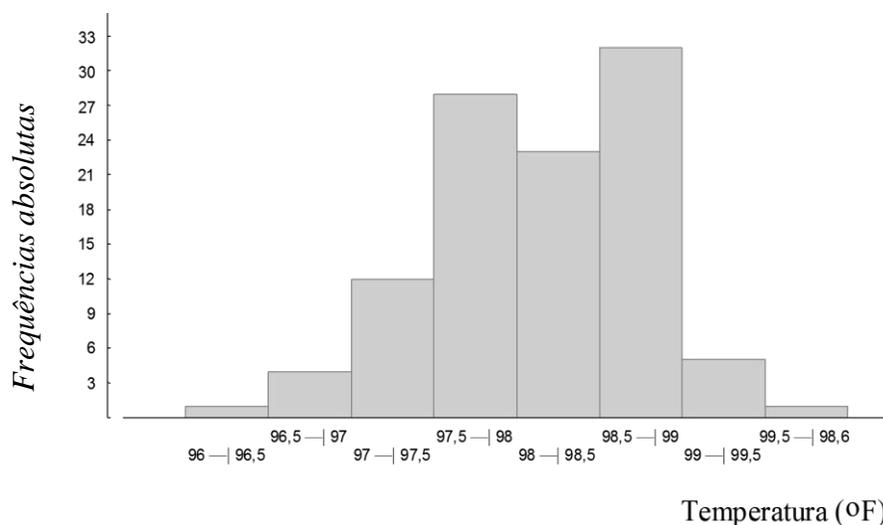


<https://www.youtube.com/watch?v=6R7TgA62ARM>

Sejam as temperaturas ( $^{\circ}\text{F}$ ) do corpo de 106 adultos sadios.

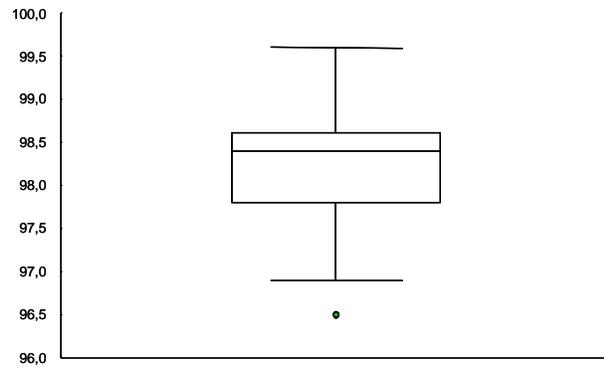
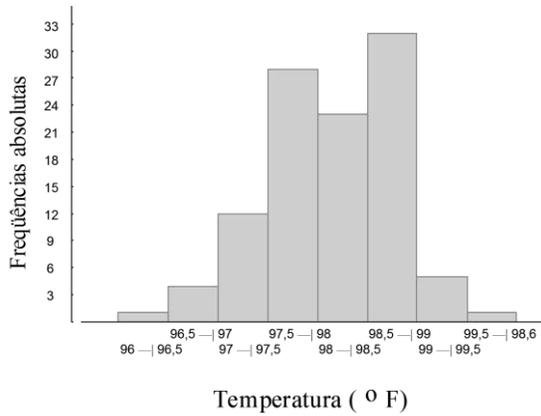
98,6	98,6	98,0	98,0	99,0	98,4	98,4	98,4	98,4	98,6	98,6	98,8
98,6	97,0	97,0	98,8	97,6	97,7	98,8	98,0	98,0	98,3	98,5	97,3
98,7	97,4	98,9	98,6	99,5	97,5	97,3	97,6	98,2	99,6	98,7	99,4
98,2	98,0	98,6	98,6	97,2	98,4	98,6	98,2	98,0	97,8	98,0	98,4
98,6	98,6	97,8	99,0	96,5	97,6	98,0	96,9	97,6	97,1	97,9	98,4
97,3	98,0	97,5	97,6	98,2	98,5	98,8	98,7	97,8	98,0	97,1	97,4
99,4	98,4	98,6	98,4	98,5	98,6	98,3	98,7	98,8	99,1	98,6	97,9
98,8	98,0	98,7	98,5	98,9	98,4	98,6	97,1	97,9	98,8	98,7	97,6
98,2	99,2	97,8	98,0	98,4	97,8	98,4	97,4	98,0	97,0		

Intervalo de classe	Frequência absoluta	Frequência relativa	Ponto médio	Amplitude do intervalo
96,0 — 96,5	1	0,94%	96,25	0,5
96,5 — 97,0	4	3,77%	96,75	0,5
97,0 — 97,5	12	11,32%	97,25	0,5
97,5 — 98,0	28	26,42%	97,75	0,5
98,0 — 98,5	23	21,70%	98,25	0,5
98,5 — 99,0	32	30,19%	98,75	0,5
99,0 — 99,5	5	4,72%	99,25	0,5
99,5 — 100,0	1	0,94%	99,75	0,5



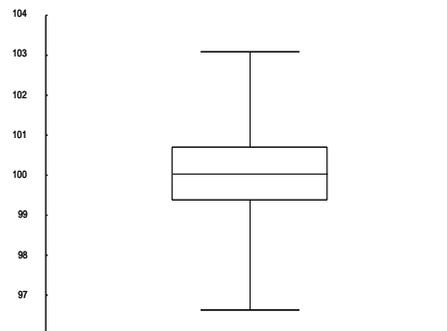
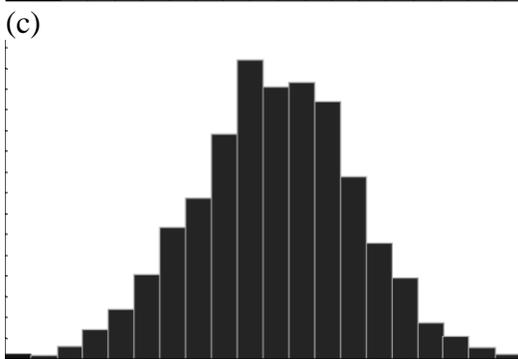
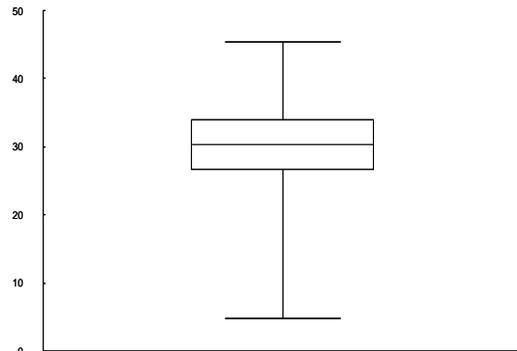
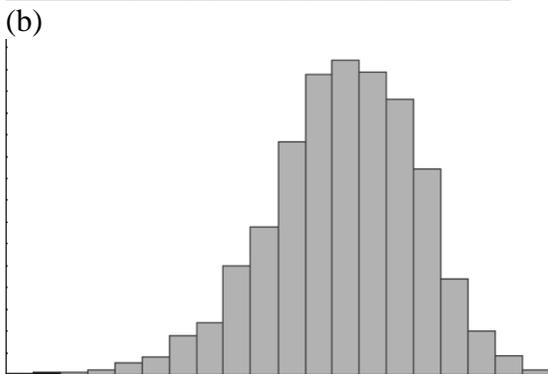
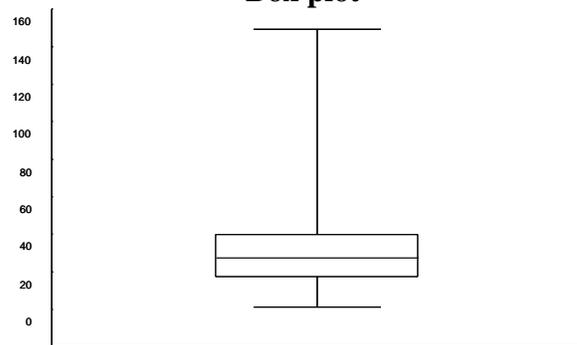
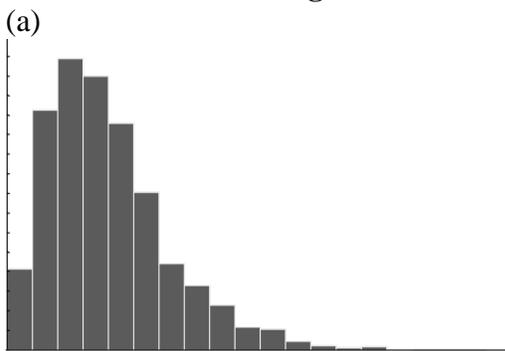
**Não confundir o histograma com o gráfico de colunas!** Não deve haver “espaços” entre os sucessivos retângulos que compõem o histograma.

**Histograma e Box-Plot:** Estas figuras permitem descrever a distribuição de uma variável contínua.



**Histograma**

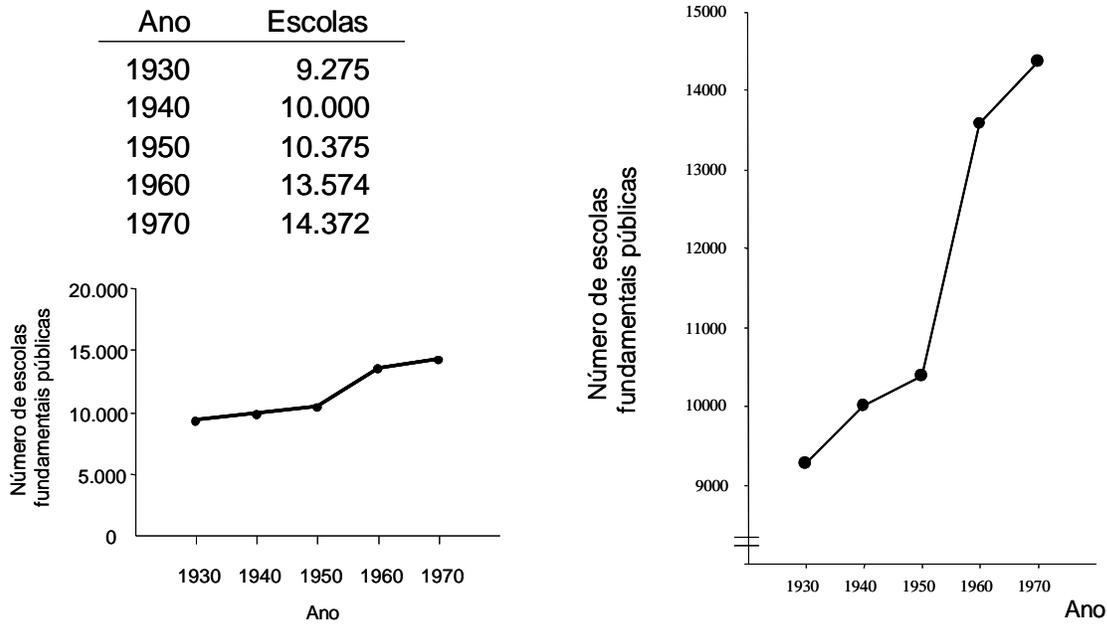
**Box plot**



- (a) Distribuição assimétrica, com cauda longa à direita. Média > Mediana.
- (b) Distribuição assimétrica, com cauda longa à esquerda. Média < Mediana.
- (c) Distribuição simétrica, aproximadamente normal. Média amostral aproximadamente igual à mediana amostral.

### Proporções das escalas

O uso de escalas com proporções corretas na elaboração do gráfico está diretamente associado à exatidão da informação nela contida.



### Proporções das escalas

De acordo com as proporções adotadas nas escalas vertical e horizontal, obtém-se um gráfico “mais alto” ou “mais largo”, o que, conforme o caso, pode distorcer o resultado.

**Ex.:** 100 pessoas são submetidas à dieta A e 100 pessoas são submetidas à dieta B.

Dieta	<i>n</i>	% com redução de peso
Dieta A	100	46 %
Dieta B	100	41 %

Gráfico A

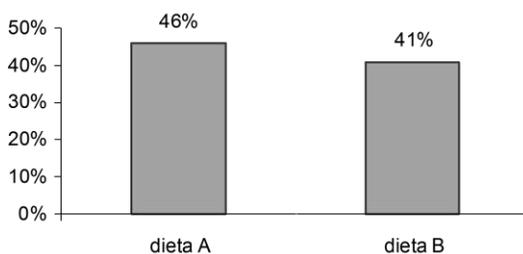
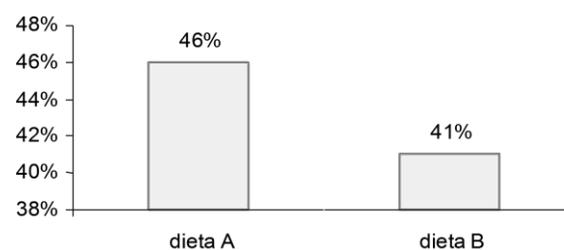
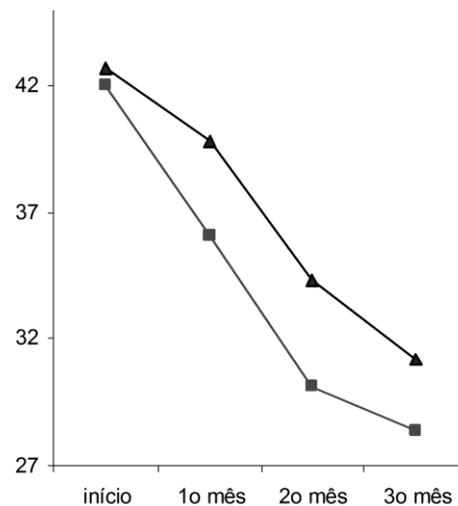
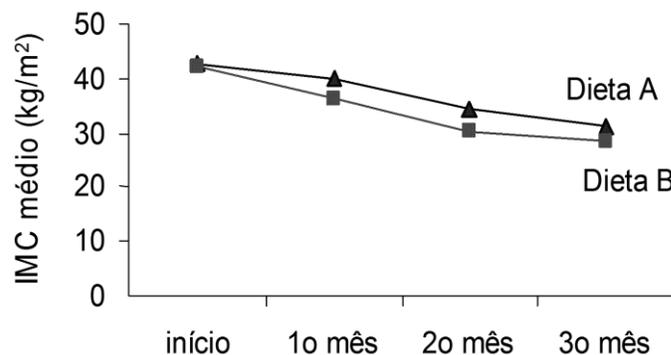


Gráfico B



Ex.:

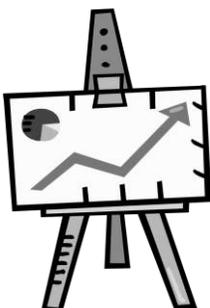
	<i>n</i>	Média do IMC (kg/m <sup>2</sup> )			
		início	1º mês	2º mês	3º mês
Dieta A	100	42,7	39,8	34,3	31,2
Dieta B	100	42,0	36,1	30,1	28,4



Considere-se que em trabalhos técnico-científicos, a finalidade principal dos gráficos não é apresentar uma composição artística e sim evidenciar informações.

### Simplicidade

O gráfico deve trazer ao observador uma percepção rápida do fenômeno descrito.



### Clareza

A apresentação do gráfico deve ser clara, de modo a proporcionar a interpretação correta dos valores nele apresentados.

O gráfico deve permitir uma única interpretação.

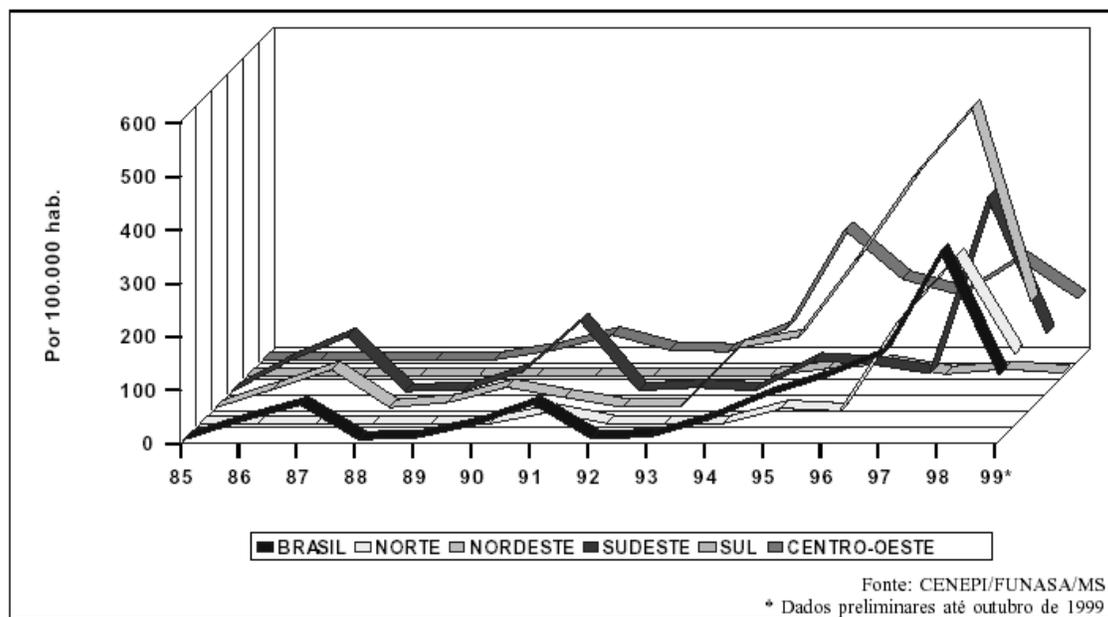
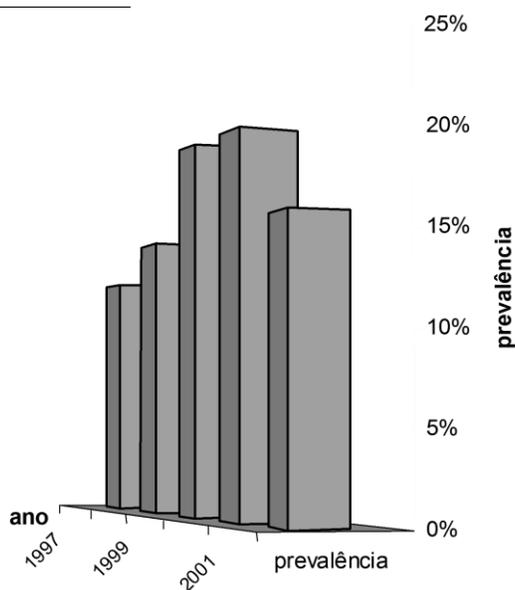
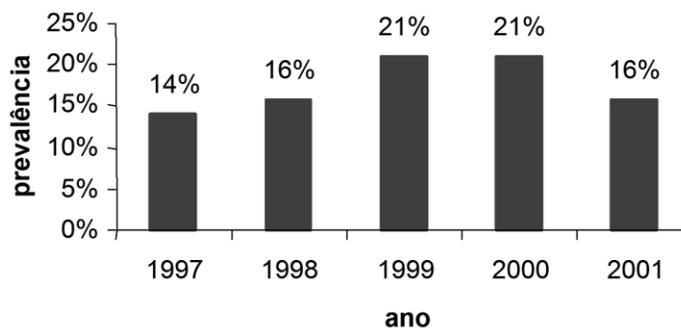


Figura 5 - Taxa de incidência de dengue por ano de ocorrência. Brasil, 1986 - 1999.

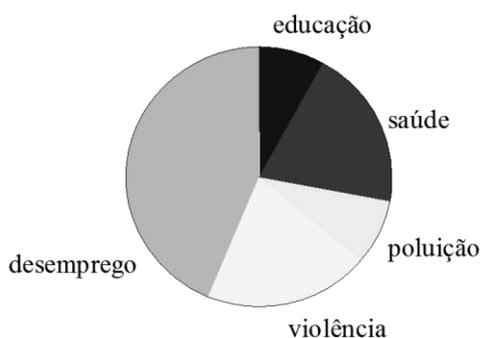
Ex.:

Ano	prevalência de hipertensos
1997	14%
1998	16%
1999	21%
2000	21%
2001	16%

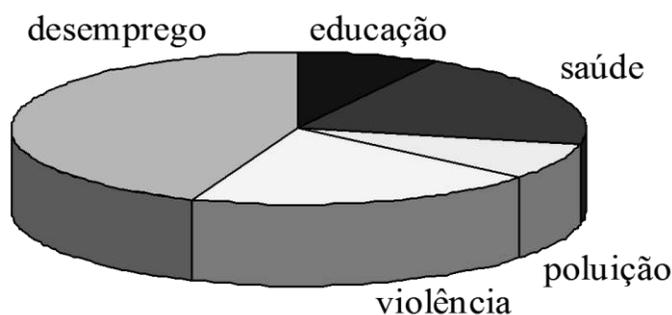


**Ex.:**

Foi perguntado a 500 habitantes qual é o principal problema de seu bairro.



	%
educação	8%
saúde	20%
poluição	8%
violência	20%
desemprego	44%
<b>total</b>	<b>100%</b>



### Exercícios

**26.** Em junho de 2006, o Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto contava com uma equipe de 4.306 funcionários. As tabelas abaixo mostram a distribuição destes funcionários por nível, tempo de serviço, sexo e idade.

Tabela 1 – Distribuição dos funcionários por nível

Nível	Funcionários
Nível Básico	1.032
Nível Médio	2.263
Nível Superior	1.011
<b>Total</b>	<b>4.306</b>

Tabela 2 – Distribuição dos funcionários por tempo de serviço

Tempo de serviço (anos)	Funcionários
0 —  5	947
5 —  10	1.015
10 —  15	553
15 —  20	842
20 —  25	355
25 —  30	559
30 —  35	32
35 —  40	3
<b>Total</b>	<b>4.306</b>

Tabela 3 – Distribuição dos funcionários por sexo e idade

Idade (anos)	Funcionários	
	Masculino	Feminino
18—  21	6	7
21—  25	32	103
25—  30	119	312
30—  40	369	925
40—  50	482	1.129
50—  60	210	540
60—  65	25	35
Mais de 65	8	4
Total	1.251	3.055

Estes dados foram obtidos do jornal *HC Notícias* de junho de 2006.

Represente as informações das tabelas acima em gráficos adequados.

**27.** Participaram de uma pesquisa 103 indivíduos do sexo masculino, com idades entre 20 a 50 anos, sem queixas vocais, que não faziam uso de tabaco nem de bebidas alcoólicas. Estes sujeitos foram submetidos a uma avaliação acústica da voz. Estão listados abaixo os valores da frequência fundamental (em Hz) para estes indivíduos.

87,3    87,7    91,7    93,7    93,9    97,7    98,8  
103,6    104,2    105,1    105,9    106,4    106,9    107,1  
107,4    107,7    108,3    108,4    109,5    109,9    110,7  
110,8    110,8    111,1    111,1    113,2    113,6    113,6  
113,7    114,0    114,1    114,2    114,2    114,4    114,6  
115,1    115,4    115,4    115,8    116,0    116,8    117,0  
117,7    118,0    118,2    118,4    119,1    119,2    119,6  
119,7    119,9    120,4    120,8    122,2    122,5    124,5  
126,3    126,7    127,8    128,4    129,5    129,7    129,7  
130,4    130,5    130,9    131,4    132,8    133,6    134,3  
134,6    134,9    135,5    137,5    137,9    138,0    138,9  
139,4    139,4    141,7    141,7    141,9    142,1    142,9  
143,1    143,4    143,8    143,9    145,2    145,3    145,5  
146,8    148,2    148,4    149,3    149,3    149,7    151,9  
152,4    152,8    153,5    156,6    158,1

- Encontre, para a frequência fundamental, a mediana e os quartis.
- Construa um *box-plot* para esta variável.
- Construa um histograma para esta variável.

**28.** O colesterol sérico foi medido em uma amostra aleatória de 100 homens, representativa da população de 20 a 74 anos de um município. Os valores encontrados, em mg/dl, são listados a seguir.

127	128	137	141	141	143	144	145	145	146
150	152	157	159	160	163	165	166	167	167
168	168	169	171	172	172	173	175	178	178
179	179	181	181	182	183	183	184	185	186
186	186	186	187	187	188	190	190	191	195
196	197	198	198	199	199	200	202	203	204
206	207	208	213	213	216	218	218	219	220
222	224	225	225	225	225	227	227	228	228
228	229	232	236	238	240	241	241	242	247
257	257	260	264	266	270	273	275	285	330

- Encontre os quartis para estes dados.
- Construa um *box-plot* para estes dados.
- Construa um histograma para estes dados.
- Com base nas figuras que você construiu nos itens (b) e (c), o que você pode dizer sobre a distribuição dos dados?
- Se você sortear um indivíduo desta amostra, ao acaso, qual a probabilidade dele possuir um valor de colesterol sérico entre 160 e 190?

**29.** Os números a seguir referem-se às concentrações séricas de hemoglobina (em g/dl) em uma amostra de 200 crianças.

10,6 10,6 10,8 10,9 10,9 11,1 11,1 11,3 11,3 11,3  
 11,3 11,3 11,3 11,4 11,4 11,4 11,4 11,4 11,4 11,4  
 11,4 11,5 11,5 11,5 11,5 11,5 11,5 11,5 11,6 11,6 11,6  
 11,6 11,7 11,7 11,7 11,7 11,7 11,7 11,7 11,7 11,7 11,8  
 11,8 11,8 11,8 11,8 11,8 11,8 11,8 11,8 11,8 11,9  
 11,9 11,9 11,9 11,9 11,9 11,9 11,9 11,9 11,9 11,9  
 11,9 11,9 11,9 11,9 12,0 12,0 12,0 12,0 12,0 12,0  
 12,0 12,0 12,0 12,0 12,0 12,0 12,1 12,1 12,1 12,1  
 12,1 12,1 12,1 12,1 12,1 12,1 12,1 12,2 12,2 12,2  
 12,2 12,2 12,2 12,2 12,2 12,2 12,2 12,2 12,3 12,3  
 12,3 12,3 12,3 12,3 12,3 12,3 12,4 12,4 12,4 12,4  
 12,4 12,4 12,4 12,4 12,4 12,4 12,4 12,4 12,5 12,5  
 12,5 12,5 12,6 12,6 12,6 12,6 12,6 12,6 12,6 12,7  
 12,7 12,7 12,7 12,7 12,7 12,7 12,8 12,8 12,8 12,8  
 12,8 12,8 12,8 12,8 12,8 12,8 12,9 12,9 12,9 12,9  
 12,9 12,9 13,0 13,0 13,0 13,0 13,0 13,1 13,1 13,1  
 13,2 13,2 13,2 13,2 13,3 13,3 13,3 13,3 13,3 13,4  
 13,4 13,4 13,4 13,4 13,5 13,5 13,5 13,5 13,6 13,6  
 13,6 13,7 13,7 13,7 13,8 13,8 13,8 13,8 13,9 13,9  
 13,9 13,9 14,0 14,1 14,2 14,2 14,2 14,3 14,4 14,8

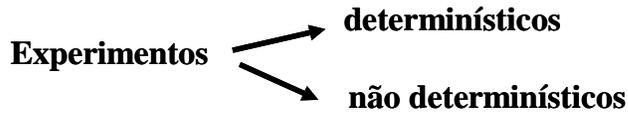
- Encontre a mediana e os outros quartis.
- Construa um histograma para esta variável, considerando amplitudes de 0,5 g/dl.

## Respostas

**29.** (a) A mediana é 12,3 g/dl, o primeiro quartil é 11,9 g/dl e o terceiro quartil é 12,9 g/dl.

# Noções de probabilidade

**Experimento:** ensaio científico objetivando a verificação de um fenômeno.



**Experimentos determinísticos:** as condições sob as quais um experimento é executado determinam o resultado do experimento.

**Experimentos não determinísticos ou probabilísticos:** o resultado do experimento é aleatório, ou seja, existe a incerteza do resultado.

**Fenômeno aleatório:** situação ou acontecimento cujos resultados não podem ser determinados com certeza.

## Exemplos:

1. Resultado do lançamento de um dado;
2. Hábito de fumar de um estudante selecionado ao acaso em sala de aula;
3. Condições climáticas do próximo domingo;
4. Taxa de inflação do próximo mês;
5. Resultado de um exame de sangue.

**Ex.:** lanço uma moeda uma vez.

Antes de lançar a moeda, não temos o conhecimento do resultado.



**Cara**

**Coroa**

**Possibilidades:** cara ou coroa.

**Evento:** resultado ou conjunto de resultados de um experimento. Representado por letras maiúsculas.

**Ex.:** A: cara B: coroa.

**Ex.:** lanço um dado uma vez e observo o número impresso na face voltada para cima.

Antes de lançar o dado, não temos o conhecimento do resultado.



**Possibilidades:** 1, 2, 3, 4, 5 ou 6.

**Espaço amostral:** o conjunto  $\Omega$  de todos os resultados possíveis de um experimento é chamado espaço amostral.

$\Omega =$  \_\_\_\_\_

**Exemplos de espaço amostral:**

1. Lançamento de um dado.

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

2. Exame de sangue (tipo sanguíneo).

$$\Omega = \{A, B, AB, O\}$$

3. Hábito de fumar.

$$\Omega = \{\text{Fumante, Não fumante}\}$$

4. Tempo de duração de uma lâmpada.

$$\Omega = \{t: t \geq 0\}$$

**Ex.:** lanço um dado uma vez e observo o número impresso na face voltada para cima.

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

**Evento:** resultado ou conjunto de resultados de um experimento. Representado por letras maiúsculas.

**Ex.:** A : observar um número par.

B: observar um número ímpar.

C: observar um número maior que 4.

D: observo o número 8.

A = \_\_\_\_\_ B = \_\_\_\_\_ C = \_\_\_\_\_ D = \_\_\_\_\_

$\emptyset$  (conjunto vazio): evento impossível

$P(A)$  : probabilidade de ocorrer o evento A

**Definição clássica de probabilidade:**

$$P(A) = \frac{\text{número de possibilidades de ocorrência do evento A}}{\text{número total de possibilidades do experimento}}$$

**Ex.:** A : observar um número par.

B: observar um número ímpar.

C: observar um número maior que 4.

D: observo o número 8.

$P(A) =$  \_\_\_\_\_  $P(B) =$  \_\_\_\_\_  $P(C) =$  \_\_\_\_\_  $P(D) =$  \_\_\_\_\_

**Definição frequentista de probabilidade:**

**Ex.:** lanço uma moeda uma vez.

**Possibilidades:** cara ou coroa.

$$\Omega = \{\text{cara, coroa}\}$$

**Eventos:** A: cara B: coroa.



**Cara**



**Coroa**

$$P(A) = \underline{\hspace{2cm}}$$

Se lançarmos a moeda um número grande de vezes,

$$P(A) = \frac{\text{número de vezes que o evento A ocorre}}{\text{número de vezes que o experimento é realizado}}$$

**Ex.:** em uma determinada população, qual a probabilidade de um indivíduo portar uma doença ?

$$\Omega = \{\text{doente, não doente}\}$$

**Eventos:** A: o indivíduo porta a doença.  
B: o indivíduo não porta a doença.

**Ou**

**Eventos:** A: o indivíduo porta a doença.  
 $A^C$ : o indivíduo não porta a doença.



**Notação:**

$A^C$ : o complemento do evento A

**Definições de probabilidade****1-) Definição clássica de probabilidade:**

Experimento: lanço um dado uma vez e observo o número impresso na face voltada para cima

A : observar o número 4                       $A = \{4\}$                        $\Omega = \{1,2,3,4,5,6\}$

$$P(A) = \frac{\text{número elementos em } A}{\text{número elementos em } \Omega}$$

**2-) Definição frequentista de probabilidade:**

Experimento: lanço uma moeda uma vez.

A : observar uma cara                       $A = \{\text{cara}\}$                        $\Omega = \{\text{cara,coroa}\}$

Realizo o experimento muitas vezes e

$$P(A) = \frac{\text{número de ocorrências de } A}{\text{número de realizações}}$$

Definições clássica e frequentista: independem do observador.

Definição frequentista: a probabilidade é o valor do qual se aproxima a frequência relativa de ocorrências do evento em um grande número de repetições do experimento.

As duas definições: são bastante intuitivas.

**Definição axiomática de probabilidade** (Kolmogorov, aprox. 1930)

Axiomas (regras) matemáticas que definem a probabilidade. Estas regras são mais formais e têm por consequências:

1.)  $0 \leq P(A) \leq 1$                       (A probabilidade é sempre um número entre 0 e 1).

2.)  $P(A^c) = 1 - P(A)$                       (A probabilidade de um evento não ocorrer é 1 menos a probabilidade dele ocorrer).

**Ex.:** em uma comunidade, 20% dos indivíduos adultos são hipertensos, 40% são diabéticos e 15% são hipertensos e diabéticos. Se um indivíduo for escolhido ao acaso, qual a probabilidade dele:

- (a) Ser diabético, mas não hipertenso ?
- (b) Ser hipertenso ou diabético (ou ambos) ?
- (c) Não ser hipertenso e nem diabético ?

Se um indivíduo for escolhido ao acaso, qual a probabilidade dele ser hipertenso, dado que ele é diabético?

Probabilidade condicional:

$$P(H | D) = ?$$

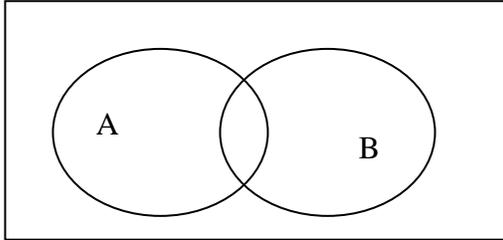
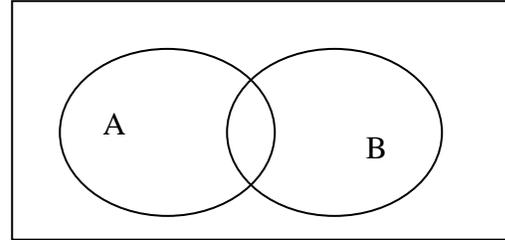
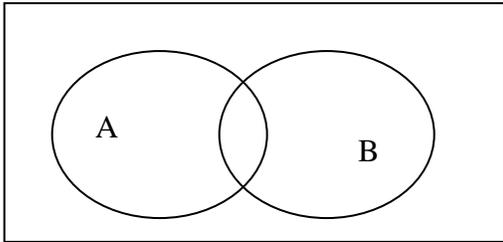
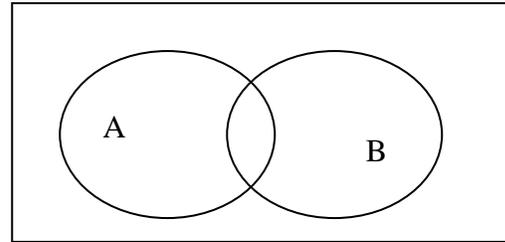
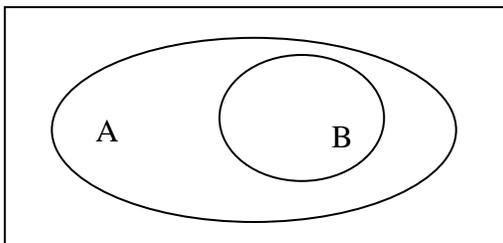
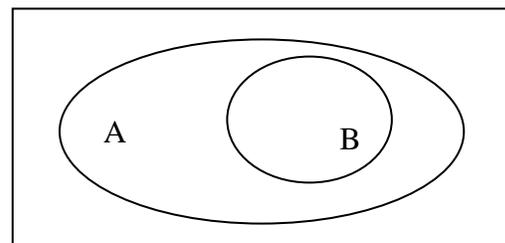
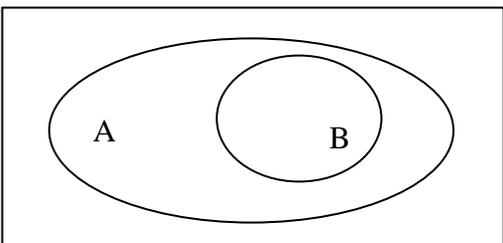
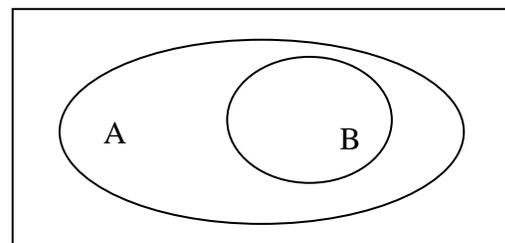


**Exercícios**

1. Sejam  $\Omega = \{a, b, c, d, e\}$ ,  $A = \{a, b, d\}$ ,  $B = \{b, d, e\}$ . Encontre:

- |                  |                    |                    |                  |
|------------------|--------------------|--------------------|------------------|
| (a) $A \cup B$   | (b) $B^C$          | (c) $(A \cap B)^C$ | (d) $A \cap B$   |
| (e) $A^C \cap B$ | (f) $A^C \cap B^C$ | (g) $(A \cup B)^C$ | (h) $A \cup B^C$ |

2. Nos diagramas de Venn, sombreie:

(a)  $B \cap A^C$ (b)  $A^C \cap B^C$ (c)  $(A \cup B) \cap (A \cap B)^C$ (d)  $(A \cap A^C)^C$ (e)  $B \cup A^C$ (f)  $A^C \cap B^C$ (g)  $(A \cup B) \cap (A \cap B)^C$ (h)  $(A \cap A^C)^C$

3. Numa classe de 150 alunos, 80 usam calça Lee e 30 usam tênis. Sabendo-se que 10 usam calça Lee e tênis, quantos não usam calça Lee nem tênis ?

4. Sejam  $\Omega = \{1,2,3,4,5,6,7,8,9\}$ ,  $A = \{1,2,3,4\}$ ,  $B = \{2,4,6,8\}$  e  $C = \{3,4,5,6\}$ . Encontre:

- (a)  $A^C$                       (b)  $A \cap C$                       (c)  $(A \cap C)^C$                       (d)  $A \cup B$

5. O questionário de uma pesquisa de mercado inquiria:

- (i) Eventualmente você toma a bebida A ?  
(ii) Eventualmente você toma a bebida B ?

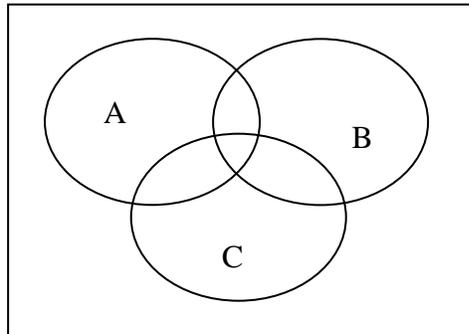
O resumo do resultado dos pesquisados que responderam a ambas as perguntas foi:

Bebida	A	B	ambas	nenhuma
no de consumidores	230	200	150	40

Qual o número de pesquisados ?

6. Em uma universidade são lidos dois jornais, A e B. Exatamente 80% dos alunos lêem o jornal A e 60% o jornal B. Sabendo que todo aluno é leitor de pelo menos um dos jornais, qual o percentual de alunos que lêem ambos ?

7. Dado o diagrama de Venn ao lado, hachure a parte que representa o conjunto  $(A \cup B) \cap (A \cup C)^C$ .



8. Defina um espaço amostral para cada um dos seguintes experimentos aleatórios:

- (a)  $E_1$ : Lançam-se dois dados e anota-se a configuração obtida.  
(b)  $E_2$ : Conta-se o número de peças defeituosas, no intervalo de uma hora, de uma linha de produção.  
(c)  $E_3$ : Investigam-se famílias com 4 crianças e anota-se a configuração obtida, segundo o sexo.  
(d)  $E_4$ : Três moedas são lançadas e anota-se a configuração obtida.  
(e)  $E_5$ : Três moedas são lançadas e anota-se o número de caras.  
(f)  $E_6$ : Joga-se um dado e observa-se o número mostrado na face de cima.  
(g)  $E_7$ : Joga-se uma moeda quatro vezes e observa-se o número de caras obtido.  
(h)  $E_8$ : Uma lâmpada é fabricada. Em seguida é testada quanto à sua duração, colocando-a em um soquete e anotando-se o tempo decorrido (em horas) até queimar.  
(i)  $E_9$ : De uma urna, que só contém bolas pretas, tira-se uma bola e verifica-se a sua cor.

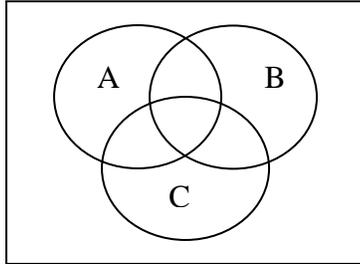
- (j)  $E_{10}$ : Verifica-se a variação de peso, em animais que recebem um novo tipo de ração, durante 2 meses.

9. Sejam os eventos A, B e C. Encontre uma expressão e indique o diagrama de Venn para o evento em que:

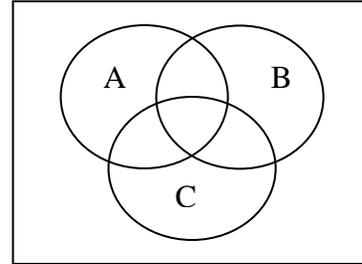
(a) A e B ocorrem, mas C não ocorre

(b) somente A ocorre.

(a)



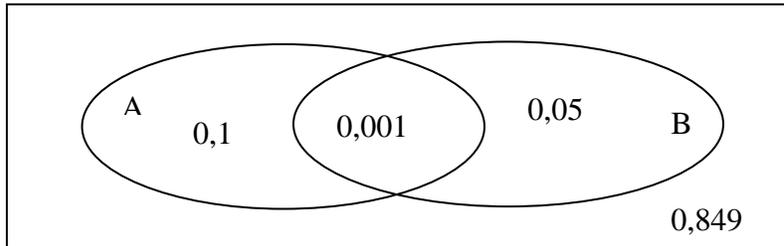
(b)



10. Considerando uma população de motoristas, os dados contidos no diagrama de Venn referem-se às probabilidades associadas aos eventos

A: o motorista treinou direção defensiva no último ano, e

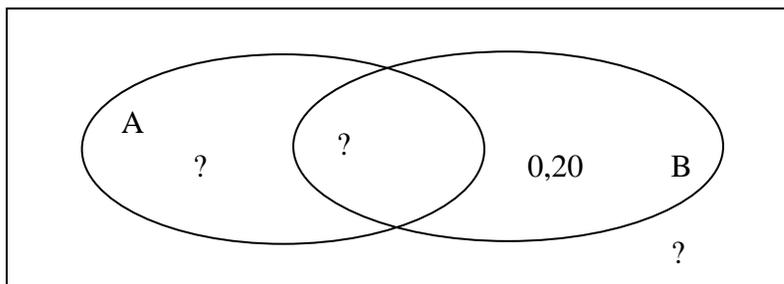
B: o motorista se acidentou no último ano.



(a) Qual é a probabilidade de um motorista desta população ter se acidentado?

(b) Qual é a probabilidade de um motorista desta população ter se acidentado, dado que ele treinou direção defensiva?

11. Suponha  $P(A) = 0,45$  e  $P(B) = 0,32$ . Determine as probabilidades necessárias para preencher as lacunas do diagrama de Venn:



12. Um estudo objetivou avaliar o estado nutricional dos indivíduos de uma população. A tabela abaixo mostra as frequências relativas dos indivíduos, segundo o sexo e o

diagnóstico nutricional. Este diagnóstico foi obtido conforme uma classificação baseada no índice de massa corpórea.

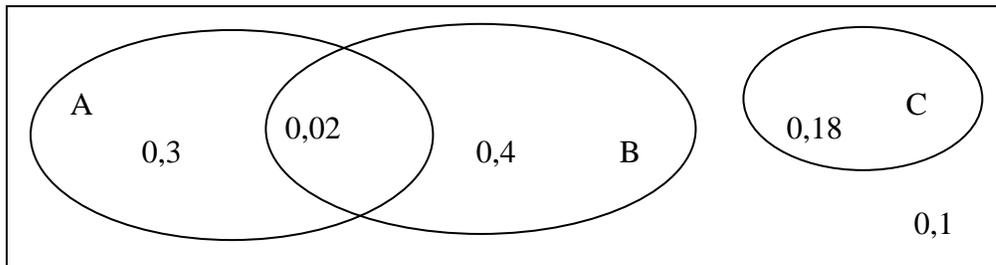
Diagnóstico nutricional	sexo	
	Masculino	Feminino
normal	0,16	0,19
sobrepeso	0,27	0,18
obesidade	0,12	0,08

Um indivíduo desta população é selecionado aleatoriamente. Sejam os eventos A : o indivíduo é homem, e B : o indivíduo é obeso. Encontre as probabilidades:

- (a)  $P(A)$                       (b)  $P(B)$                       (c)  $P(A \cap B)$                       (d)  $P(A \cup B)$   
 (e)  $P(A^C)$                       (f)  $P(A^C \cup B^C)$                       (g)  $P(A^C \cap B^C)$

13. Em uma dada região, 15% dos adultos são fumantes, 0,86% são fumantes com enfisema pulmonar, e 0,24% são não fumantes com enfisema. Qual é a probabilidade de que uma pessoa, selecionada ao acaso, tenha enfisema?

14. Considerando as probabilidades exibidas no diagrama de Venn abaixo, encontre  $P(A^C)$ ,  $P(A \cap B)$ ,  $P(B \cup C)$  e  $P(B \cap C)$ .



15. Sejam três eventos, A, B e C, em que  $P(A) = 0,51$ ,  $P(B) = 0,45$ ,  $P(C) = 0,5$ ,  $P(A \cap B) = 0,17$ ,  $P(B \cap C) = 0,20$ ,  $P(A \cap C) = 0,33$  e  $P(A \cap B \cap C) = 0,12$ .

- (a) Construa um diagrama de Venn, exibindo no diagrama as probabilidades associadas a cada região.  
 (b) Encontre  $P(B \cap C^C)$ .  
 (c) Considerando os eventos A, B e C, encontre a probabilidade de exatamente dois destes eventos ocorrerem.  
 (d) Encontre a probabilidade de nenhum dos eventos (A, B e C) ocorrerem.

16. Sejam dois eventos, A e B, sendo  $P(A \cup B) = 0,8$  e  $P(A) = 0,5$ . Encontre  $P(B)$  se:

- (a) A e B são mutuamente exclusivos  
 (b) A e B são independentes

17. Se A e B são eventos de um mesmo espaço amostral, com  $P(A) = 0,2$ ,  $P(A \cup B) = 0,5$  e  $P(A \cap B) = 0,1$ , encontre  $P(B)$ .

**18.** Uma fonoaudióloga está investigando a possível associação entre o tabagismo, o etilismo (alcoolismo) e um determinado distúrbio de voz. Em uma comunidade onde 6% dos indivíduos são portadores deste distúrbio de voz e 24% são tabagistas, todos os indivíduos etilistas são tabagistas (mas nem todos os tabagistas são etilistas). Nota-se que 3% dos indivíduos são, ao mesmo tempo, tabagistas, etilistas e portadores do distúrbio de voz. É conhecido também que 7% dos indivíduos são ao mesmo tempo tabagistas e etilistas; e que 5% dos indivíduos são ao mesmo tempo tabagistas e portadores do distúrbio de voz.

- (a) Se selecionado ao acaso um indivíduo desta comunidade, qual é a probabilidade deste indivíduo ser etilista?
- (b) Se selecionado ao acaso um indivíduo desta comunidade, qual é a probabilidade deste indivíduo ser portador do distúrbio de voz, dado que ele é tabagista?
- (c) Se selecionado ao acaso um indivíduo desta comunidade, qual é a probabilidade deste indivíduo ser portador do distúrbio de voz, dado que ele é etilista?

**19.** A importância da Fisioterapia na prevenção de quedas em idosos vem sendo estudada por muitos pesquisadores. Em uma população de idosos, foi implantado por fisioterapeutas um programa de prevenção de quedas que contou com a adesão de 30% dos integrantes. Seis meses após a implementação do programa, verificou-se que 53% dos idosos (dentre todos, que aderiram ou não ao programa) sofreram uma queda. Os fisioterapeutas observaram que, dentre todos os idosos da população, 7% sofreram uma queda e participaram do programa. Ao selecionarmos ao acaso um indivíduo desta população de idosos, encontre a probabilidade deste indivíduo:

- (a) não ter participado do programa e ter sofrido uma queda;
- (b) ter participado do programa e não ter sofrido uma queda;
- (c) não ter participado do programa e nem ter sofrido uma queda;
- (d) ter sofrido uma queda, dado que participou do programa;
- (e) ter sofrido uma queda, dado que não participou do programa.
- (f) Ao comparar os resultados encontrados nos itens (d) e (e), temos alguma evidência de que o programa cumpriu com o seu objetivo? Justifique sua resposta.

**20.** Em uma empresa trabalham 340 pessoas, sendo que 170 fazem com regularidade alguma atividade física e 90 fazem uso regular de algum tipo de bebida alcoólica. Foi observado que 70 pessoas fazem com regularidade alguma atividade física e também uso regular de algum tipo de bebida alcoólica. Se selecionarmos ao acaso uma pessoa dentre estas 340, qual a probabilidade desta pessoa:

- (a) Não ser uma praticante de atividades físicas?
- (b) Ser uma praticante de atividades físicas, mas não fazer uso regular de algum tipo de bebida alcoólica?
- (c) Fazer com regularidade alguma atividade física ou uso regular de algum tipo de bebida alcoólica (ou ambos)?
- (d) Considere que você sabe que a pessoa selecionada faz atividades físicas regularmente. Qual é a probabilidade dela usar regularmente alguma bebida alcoólica?

(e) Considere agora que você sabe que a pessoa selecionada não faz atividades físicas regularmente. Qual é a probabilidade dela usar regularmente alguma bebida alcoólica?

(f) Com base nas respostas aos itens (d) e (e), responda a esta questão: nesta turma de estudantes que participaram da pesquisa, o hábito de ingerir bebidas alcoólicas com regularidade está mais presente em quem pratica atividades físicas ou em quem não pratica atividades físicas regularmente?

**21.** O bruxismo é o hábito de ranger os dentes. É considerado uma patologia de ocorrência comum, podendo ser observado em todas as faixas etárias. Uma fonoaudióloga deseja estudar a relação entre o bruxismo e perdas auditivas. Em um grupo de 140 pessoas, ela observou que 57 eram portadoras de perdas auditivas e 20 eram portadoras de bruxismo. E ainda, observou que 5 pessoas eram portadoras de bruxismo mas **não** portadoras de perdas auditivas. Se uma pessoa é selecionada ao acaso deste grupo,

(a) qual é a probabilidade dela ser portadora de perda auditiva, dado que ela é portadora de bruxismo?

(b) qual é a probabilidade dela ser portadora de perda auditiva, dado que ela não é portadora de bruxismo?

(c) De acordo com as suas respostas aos itens (a) e (b), há evidências de uma relação entre bruxismo e perdas auditivas? Justifique a sua resposta.

**22.** Uma empresa possui 200 funcionários. Uma fonoaudióloga constatou que 80 funcionários desta empresa são portadores de perda auditiva. Foi também verificado que 100 funcionários frequentam bailes *funk*, o que levou a fonoaudióloga a acreditar que este hábito seria um fator de risco para as perdas auditivas. Considere que 75 funcionários frequentam bailes *funk* e possuem perda auditiva. Se selecionarmos ao acaso um funcionário desta empresa, encontre:

(a) A probabilidade do funcionário selecionado não frequentar bailes *funk* e não possuir perda auditiva;

(b) A probabilidade do funcionário possuir perda auditiva e não frequentar bailes *funk*;

(c) A probabilidade do funcionário frequentar bailes *funk* e não possuir perda auditiva;

(d) A probabilidade do funcionário possuir perda auditiva dado que ele frequenta bailes *funk*;

(e) A probabilidade do funcionário possuir perda auditiva dado que ele não frequenta bailes *funk*;

(f) Com base nas respostas aos itens anteriores, há alguma evidência de que, entre estes funcionários, a frequência a bailes *funk* está associada a perdas auditivas?

**23.** Em uma comunidade, dentre as crianças maiores de 2 anos, 22% são portadoras de anemia e 43% pertencem a famílias com renda familiar per capita (RFPC) **menor** que 1 salário mínimo (SM). Sabe-se ainda que 56% das crianças maiores de 2 anos pertencem a famílias com RFPC **maior ou igual** a 1 SM e **não são** portadoras de anemia. Ao escolhermos ao acaso uma criança maior de 2 anos desta população, encontre a probabilidade da criança assim selecionada:

(a) ser portadora de anemia e pertencer a família com RFPC igual ou superior a 1 SM;

(b) ser portadora de anemia e pertencer a família com RFPC inferior a 1 SM;

- (c) ser portadora de anemia dado que pertence a família com RFPC igual ou superior a 1 SM;
- (d) ser portadora de anemia dado que pertence a família com RFPC inferior a 1 SM.
- (e) Ao comparar os resultados encontrados nos itens (c) e (d), é possível constatar que a anemia é mais frequente em crianças de famílias com RFPC inferior a 1 SM?

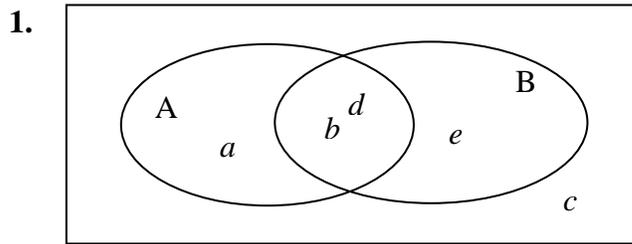
**24.** Dentre as crianças maiores de 2 anos de uma comunidade, 10% são portadoras de anemia ferropriva (AF) e 60% apresentam níveis adequados de aprendizagem escolar. Sabe-se ainda que 31% das crianças maiores de 2 anos não são portadoras de AF e não apresentam níveis adequados de aprendizagem escolar. Ao escolhermos ao acaso uma criança maior de 2 anos desta comunidade, encontre a probabilidade da criança assim selecionada:

- (a) ser portadora de AF e apresentar níveis adequados de aprendizagem escolar;
- (b) ser portadora de AF e não apresentar níveis adequados de aprendizagem escolar;
- (c) apresentar níveis adequados de aprendizagem escolar dado que é portadora de AF;
- (d) apresentar níveis adequados de aprendizagem escolar dado que não é portadora de AF.
- (e) Ao comparar os resultados encontrados nos itens (c) e (d), é possível constatar que há alguma relação entre AF e a adequação da aprendizagem escolar?

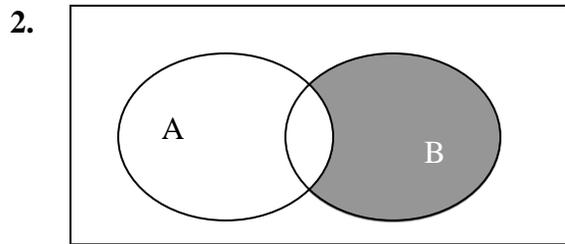
**25.** Um estudo avaliou a possível associação entre o sobrepeso e as condições de moradia das mulheres adultas de uma comunidade. As condições de moradia destas mulheres foram classificadas como adequadas ou inadequadas. Em adição, as mulheres foram classificadas como portadoras ou não portadoras de sobrepeso, conforme critérios pré-estabelecidos. Verificou-se que 15% das mulheres tinham condições inadequadas de moradia, e 20% eram portadoras de sobrepeso. Foi observado também que 70% das mulheres adultas desta comunidade tinham condições adequadas de moradia e **não** eram portadoras de sobrepeso. Quando escolhida ao acaso uma mulher adulta desta comunidade, encontre a probabilidade desta mulher:

- (a) ter condição inadequada de moradia e portar sobrepeso;
- (b) ter condição inadequada de moradia e não portar sobrepeso;
- (c) ser portadora de sobrepeso, dado que tem condições adequadas de moradia;
- (d) ser portadora de sobrepeso, dado que tem condições inadequadas de moradia.
- (e) Ao comparar os resultados encontrados nos itens (c) e (d), é possível constatar que há alguma relação entre as condições de moradia e o sobrepeso?

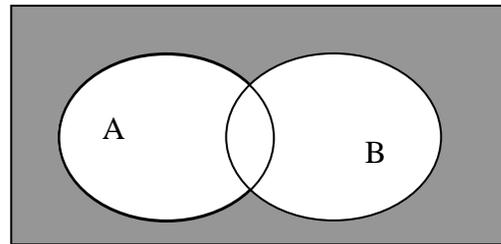
Respostas



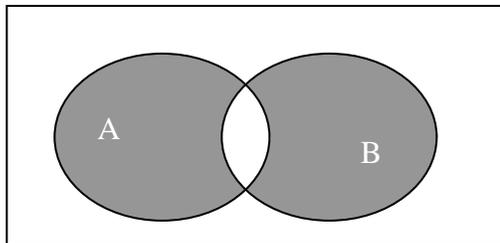
- (a) {a, b, d, e}    (b) {a, c}
- (c) {a, c, e}    (d) {b, d}
- (e) {e}    (f) {c}
- (g) {c}    (h) {a, b, c, d}



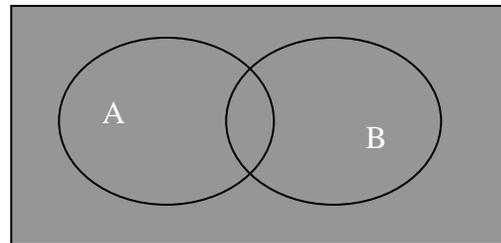
(a)  $B \cap A^c$



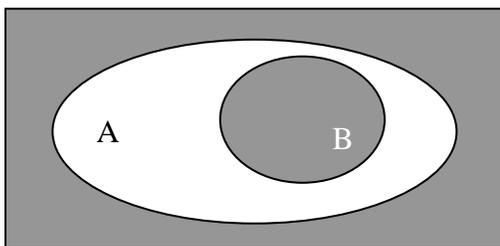
(b)  $A^c \cap B^c = (A \cup B)^c$



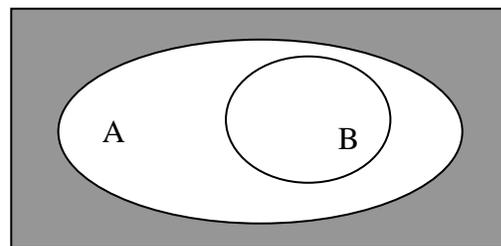
(c)  $(A \cup B) \cap (A \cap B)^c$



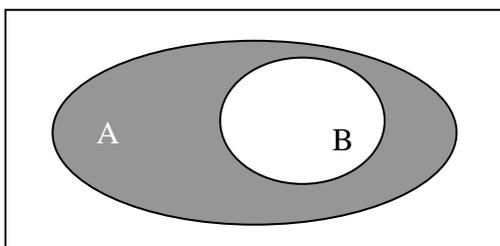
(d)  $(A \cap A^c)^c = \Omega$



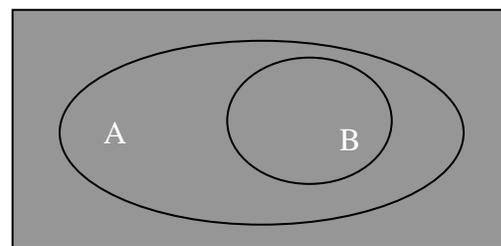
(e)  $B \cup A^c$



(f)  $A^c \cap B^c$

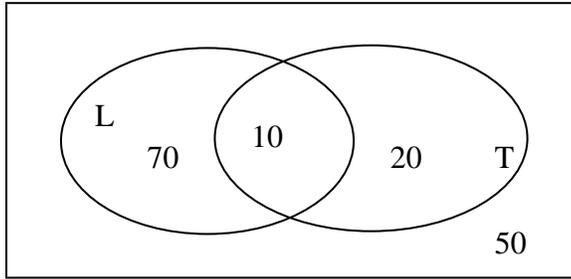


(g)  $(A \cup B) \cap (A \cap B)^c$



(h)  $(A \cap A^c)^c = \Omega$

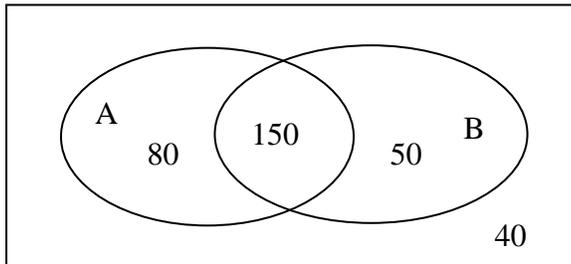
3.



50 alunos não usam calça Lee nem tênis.

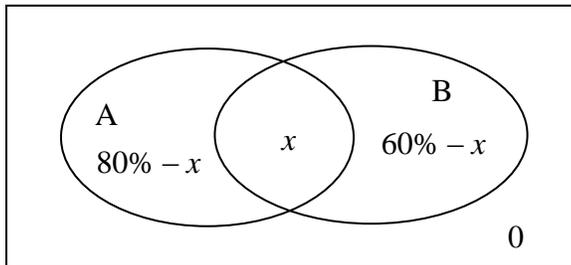
4. (a) {5,6,7,8,9} (b) {3,4} (c) {1,2,5,6,7,8,9} (d) {1,2,3,4,6,8}

5.



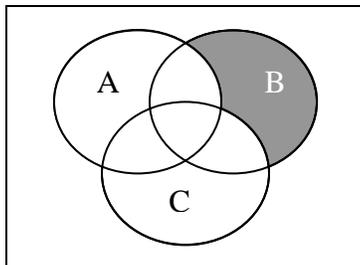
Responderam à pesquisa  
 $80 + 150 + 50 + 40 = 320$   
 pessoas.

6.



Temos que  
 $80\% - x + x + 60\% - x = 100\%$ .  
 Assim,  $x = 40\%$ .

7.



8.

- (a)  $\Omega_1 = \{ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \}$

Por simplicidade, poderíamos escrever:  $\Omega_1 = \{ (x,y) : x = 1, \dots, 6, y = 1, \dots, 6 \}$

- (a)  $\Omega_2 = \{ 0, 1, 2, 3, 4, \dots \}$

- (b) Se M:masculino e F:feminino, então  $\Omega_3 = \{ (M,M,M,M), (M,M,M,F), (M,M,F,M), (M,F,M,M), (F,M,M,M), (M,M,F,F), (M,F,M,F), (F,M,M,F) \}$

(M,F,F,M) , (F,M,F,M) , (F,F,M,M) , (M,F,F,F) , (F,M,F,F) , (F,F,M,F), (F,F,F,M), (F,F,F,F) }

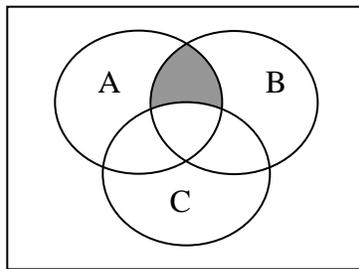
Por simplicidade, poderíamos escrever:  $\Omega_3 = \{ (x,y,w,z) : x,y,w,z = M \text{ ou } F \}$

- (c) Se C: cara e C<sup>c</sup>: coroa, então  $\Omega_4 = \{ (C,C,C) , (C^c,C,C) , (C,C^c,C) , (C,C,C^c) , (C^c,C^c,C) , (C^c,C,C^c) , (C,C^c,C^c) , (C^c,C^c,C^c) \}$
- (d)  $\Omega_5 = \{0,1,2,3\}$
- (e)  $\Omega_6 = \{1,2,3,4,5,6\}$
- (f)  $\Omega_7 = \{0,1,2,3,4\}$
- (g)  $\Omega_8 = \{t : t \geq 0\}$
- (h)  $\Omega_9 = \{\text{preta}\}$
- (i)  $\Omega_{10} = \mathfrak{R}$  (o conjunto dos números reais)

9.

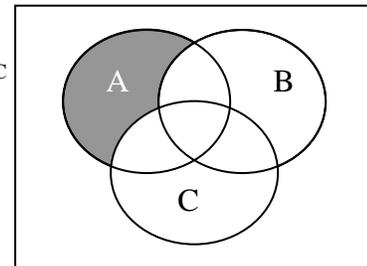
(a)

$(A \cap B) \cap C^c$



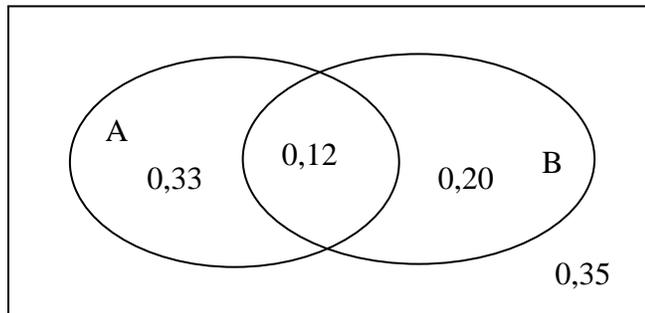
(b)

$A \cap (B \cup C)^c$



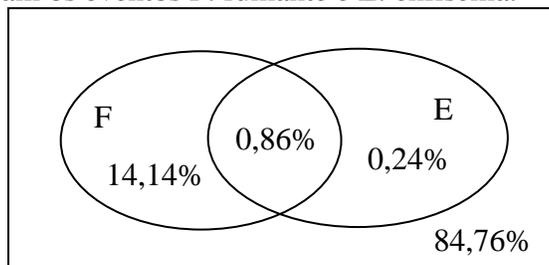
10. (a) 5,1 % (b)  $0,001/0,101 = 0,99\%$ .

11.



12. (a) 55% (b) 20% (c) 12% (d) 63%

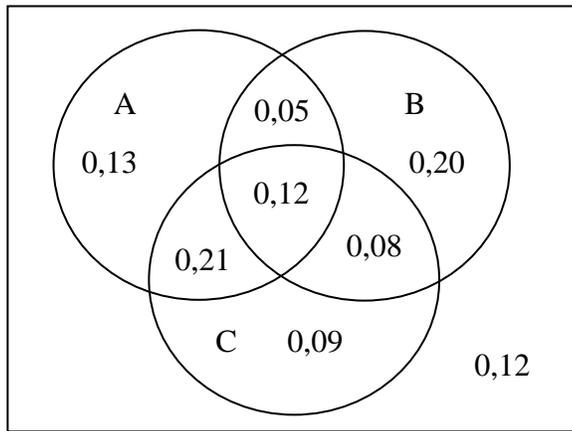
13. Sejam os eventos F: fumante e E: enfisema.



$P(E) = 1,1\%$

14.  $P(A^c) = 68\%$ ,  $P(A \cap B) = 2\%$ ,  $P(B \cup C) = 60\%$  e  $P(B \cap C) = 0$ .

15. (a)



(b) 25%

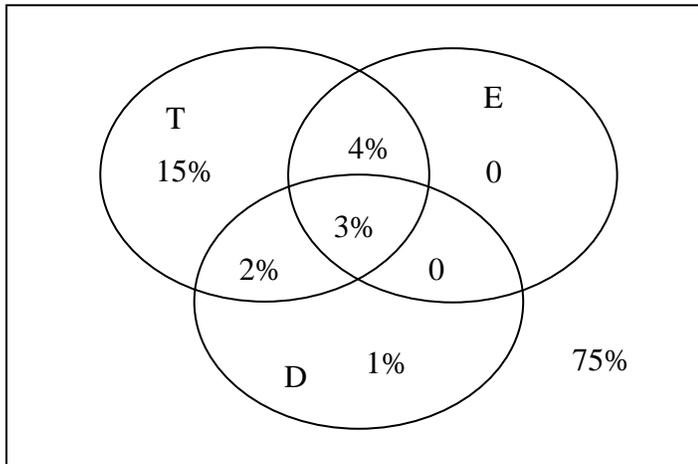
$$(c) 0,05 + 0,21 + 0,08 = 0,34 = 34\%$$

(d) 12%

16. (a) Se A e B são mutuamente exclusivos,  $P(A \cup B) = P(A) + P(B)$ , portanto  $P(B) = 0,8 - 0,5 = 0,3$ ; (b) Se A e B são independentes, então  $P(A \cap B) = P(A) P(B) = 0,5 P(B)$ . Considerando que  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ , temos  $0,8 = 0,5 + P(B) - 0,5 P(B)$ . Isolando  $P(B)$  desta expressão, temos  $P(B) - 0,5 P(B) = 0,8 - 0,5$ , ou seja,  $0,5 P(B) = 0,3$ , e, finalmente  $P(B) = 0,6$ .

17.  $P(B) = 0,4$ .

18.



Sejam os eventos:

T: ser tabagista

E: ser etilista

D: ser portador do distúrbio de voz

Temos:

$P(T) = 24\%$

$P(D) = 6\%$

$P(T \cap E \cap D) = 3\%$

$P(T \cap E) = 7\%$

$P(T \cap D) = 5\%$

$P(E \cap T^c) = 0$

(a)  $4\% + 3\% + 0 + 0 = 7\%$ .

(b) Dentre um total de 24% de tabagistas, o diagrama acima nos mostra que 5% destes indivíduos são portadores do distúrbio de voz. Assim, a probabilidade de um indivíduo ser portador do distúrbio de voz, dado que ele é tabagista, é de  $5\% / 24\% = 20,8\%$ .

Ou, utilizando as definições de probabilidade condicional, temos:

$$P(D|T) = \frac{P(D \cap T)}{P(T)} = \frac{5\%}{24\%} = 20,8\%.$$

(c) Dentre um total de 7% de etilistas, o diagrama acima nos mostra que 3% destes indivíduos são portadores do distúrbio de voz. Assim, a probabilidade de um indivíduo ser portador do distúrbio de voz, dado que ele é etilista, é de  $3\% / 7\% = 42,9\%$ .

Ou, utilizando as definições de probabilidade condicional, temos:

$$P(D|E) = \frac{P(D \cap E)}{P(E)} = \frac{3\%}{7\%} = 42,9\%.$$

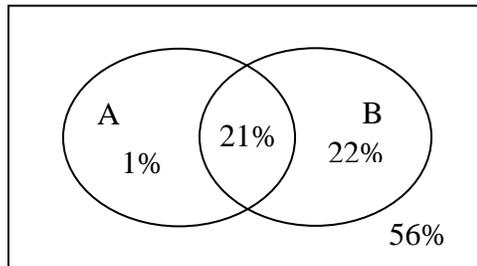
19. (a) 46% (b) 23% (c) 24% (d) 23,3% (e) 65,7%

20. (a) 50% (b) 29,4% (c) 55,9% (d) 41,2% (e) 11,8%  
(f) Quem pratica atividade física possui maior probabilidade de ser um consumidor de bebida alcoólica.

21. (a) 75% (b) 35%

22. (a) 47,5% (b) 2,5% (c) 12,5% (d) 75% (e) 5%

23.



Sejam os eventos:

A: ser portadora de anemia

B: RFPC menor que 1 SM

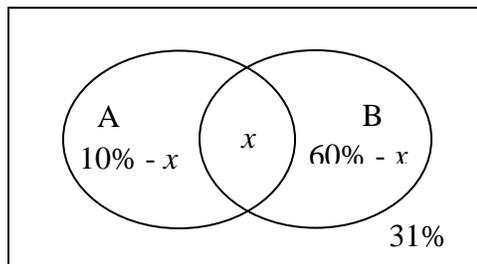
$$P(A) = 22\%$$

$$P(B) = 43\%$$

$$P(A^c \cap B^c) = 56\%$$

- (a) 1% (b) 21% (c)  $1\% / 57\% = 1,75\%$  (d)  $21\% / 43\% = 48,8\%$   
(e) Sim. A anemia é frequente em 48,8% das crianças com RFPC menor que 1 SM e é frequente em apenas 1,75% das crianças com RFPC maior ou igual a 1 SM.

24.



Sejam os eventos:

A: ser portadora de AF

B: níveis adequados de aprendizagem

$$10\% - x + x + 60\% - x + 31\% = 100\%$$

$$\text{Portanto, } x = 1\%$$

$$P(A) = 10\%$$

$$P(B) = 60\%$$

$$P(A^c \cap B^c) = 31\%$$

- (a) 1% (b) 9% (c)  $1\% / 10\% = 10\%$  (d)  $59\% / 90\% = 65,6\%$   
(e) Sim. Notar que apenas 10% das crianças com AF apresentam níveis adequados de aprendizagem, enquanto este percentual é de 65,6% para as crianças não portadoras de AF.

25. (a) 5% (b) 10% (c) 17,6% (d) 33,3%

# Distribuição normal

## Referências:

Capítulo 7 – “Distribuições teóricas de probabilidade” do livro: Pagano M, Gauvreau K. Princípios de Bioestatística. São Paulo: Pioneira Thomson Learning, 2004.

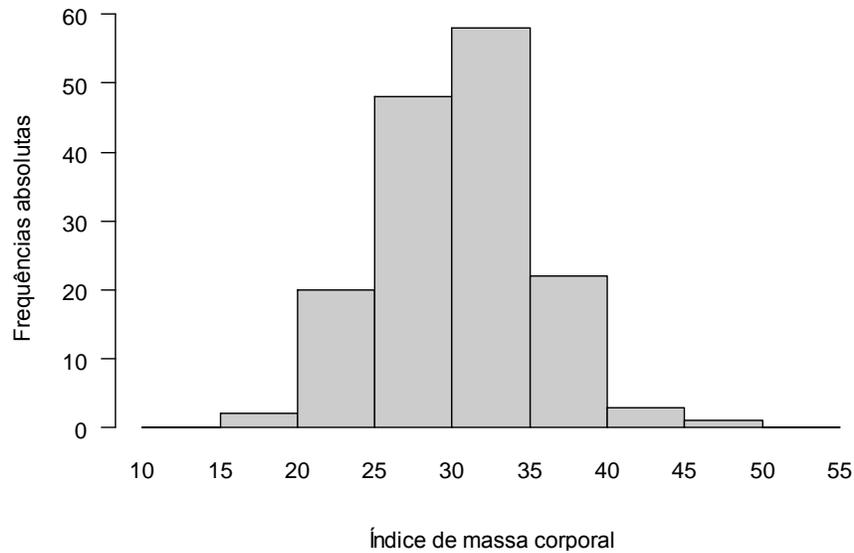
Capítulo 7 – “A distribuição normal” do livro: Martinez EZ. Bioestatística para os cursos de graduação da área da saúde.

**Ex.:** Índice de massa corpórea (IMC) de  $n=154$  mulheres adultas, medido em  $\text{kg/m}^2$ .

42,8 31,7 25,2 35,3 34,1 27,0 32,3 36,5 25,7 23,0 31,8 23,3  
 34,0 30,1 28,4 28,6 34,6 40,6 24,6 28,4 30,4 34,6 22,4 39,2  
 28,3 31,4 25,9 32,9 36,0 25,0 33,6 32,3 29,9 27,1 29,2 26,6  
 34,4 26,8 30,3 48,5 24,6 33,8 37,7 33,4 26,8 33,0 29,9 40,0  
 31,2 30,1 27,8 24,2 24,5 31,6 34,5 27,7 27,6 28,6 24,3 26,4  
 31,0 26,6 28,9 34,5 32,2 35,5 30,0 31,3 33,1 27,4 26,4 31,3  
 28,5 36,0 36,3 28,9 44,1 24,4 25,0 34,9 30,3 35,4 26,1 34,6  
 31,2 30,3 28,8 27,3 24,4 36,7 25,8 30,8 32,6 27,7 28,3 38,8  
 30,6 40,0 34,7 32,0 24,6 23,5 24,0 34,1 31,7 23,2 32,6 39,2  
 30,6 28,6 25,3 29,9 27,8 19,0 39,5 39,2 30,1 31,7 34,0 32,6  
 31,5 34,7 26,0 35,6 22,7 26,8 26,8 27,4 22,9 30,3 29,4 32,8  
 24,3 32,3 29,7 16,3 22,8 35,9 40,0 34,8 30,1 34,8 28,4 33,3  
 39,9 35,9 29,2 35,3 28,9 32,4 34,9 32,0 26,7 27,9

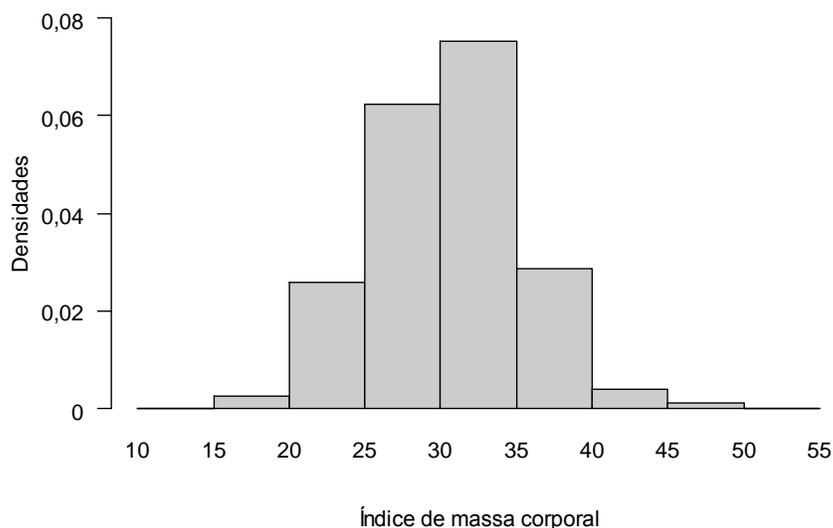
Tabela de frequências:

Intervalo de classe ( $\text{kg/m}^2$ )	Frequência absoluta	Frequência relativa	Ponto médio ( $\text{kg/m}^2$ )	Amplitude do intervalo ( $\text{kg/m}^2$ )
15,0 – 20,0	2	1,3%	17,5	5,0
20,0 – 25,0	20	13,0%	22,5	5,0
25,0 – 30,0	48	31,2%	27,5	5,0
30,0 – 35,0	58	37,7%	32,5	5,0
35,0 – 40,0	22	14,3%	37,5	5,0
40,0 – 45,0	3	1,9%	42,5	5,0
45,0 – 50,0	1	0,6%	47,5	5,0

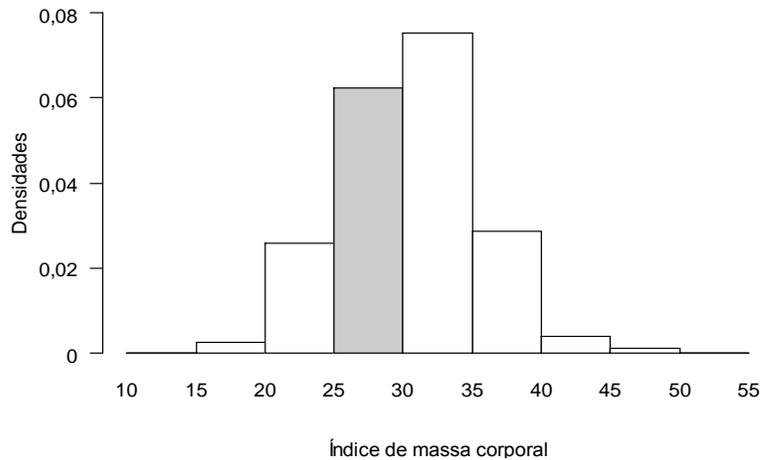


Seja 
$$\text{Densidade} = \frac{\text{Frequência relativa}}{\text{Amplitude do intervalo}}$$

Intervalo de classe (kg/m <sup>2</sup> )	Frequência absoluta	Frequência relativa	Ponto médio (kg/m <sup>2</sup> )	Amplitude do intervalo (kg/m <sup>2</sup> )	Densidade
15,0 – 20,0	2	1,3%	17,5	5,0	0,0026
20,0 – 25,0	20	13,0%	22,5	5,0	0,0260
25,0 – 30,0	48	31,2%	27,5	5,0	0,0624
30,0 – 35,0	58	37,7%	32,5	5,0	0,0754
35,0 – 40,0	22	14,3%	37,5	5,0	0,0286
40,0 – 45,0	3	1,9%	42,5	5,0	0,0038
45,0 – 50,0	1	0,6%	47,5	5,0	0,0012



Área do retângulo  
 = base x altura  
 = amplitude x densidade  
 = frequência relativa



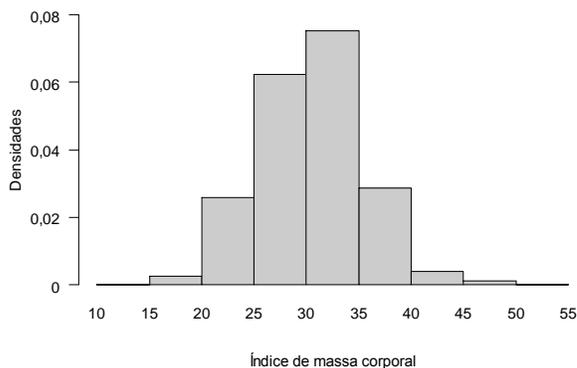
Área do retângulo  
sombreado =

$$5 \times 0,0624 = 31,2\%$$

1) Se eu sorteio casualmente um indivíduo, dentre estes 154, qual a probabilidade de seu IMC estar entre 25 e 30 kg/m<sup>2</sup> ?

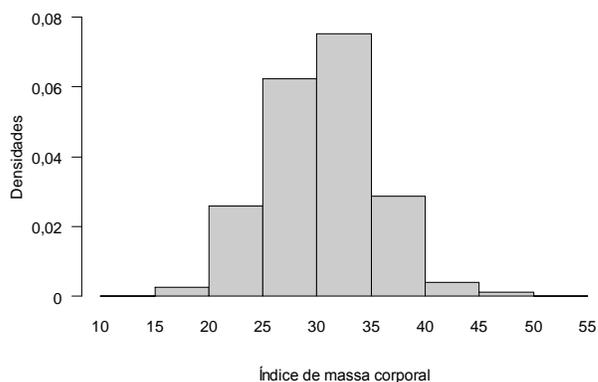
**R.:**  $P(25 < X \leq 30) = 31,2\%$

2) Se eu sorteio casualmente um indivíduo, dentre estes 154, qual a probabilidade de seu IMC estar entre 25 e 35 kg/m<sup>2</sup> ?

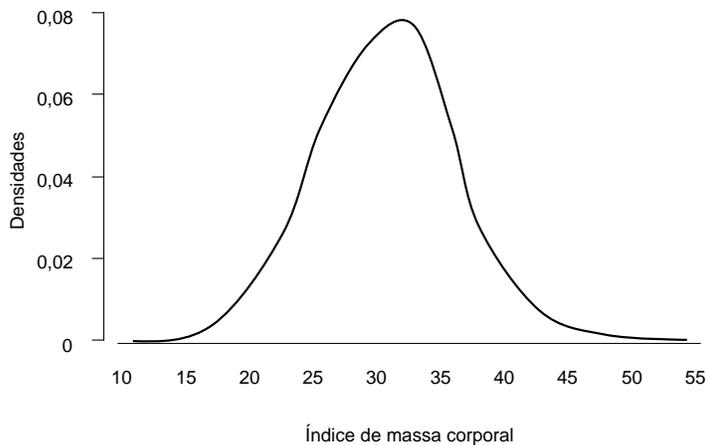
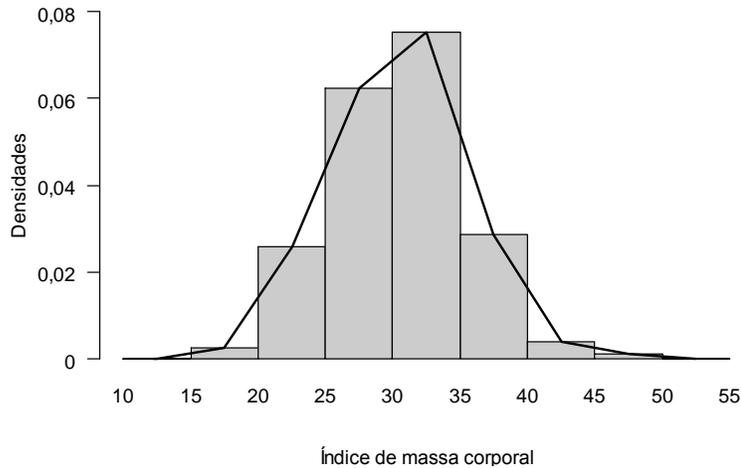


Resposta: \_\_\_\_\_

3) Se eu sorteio casualmente um indivíduo, dentre estes 154, qual a probabilidade de seu IMC estar entre 27 e 34 kg/m<sup>2</sup> ?



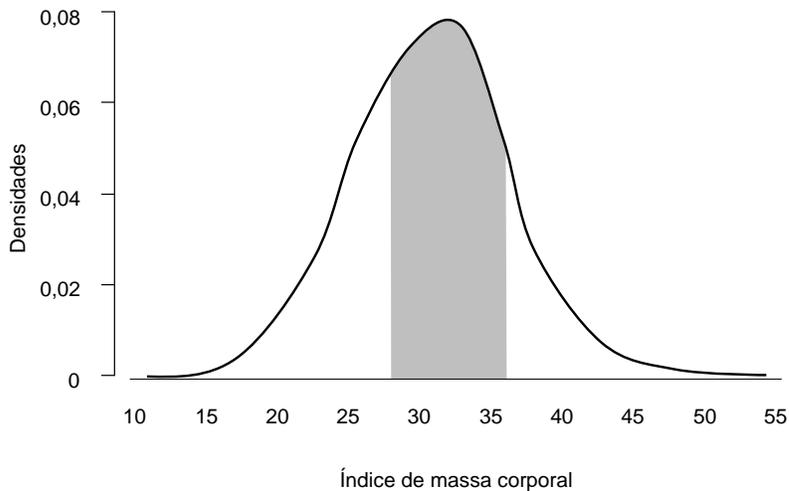
Resposta: \_\_\_\_\_



Se considerarmos amplitudes próximas a zero e um tamanho amostral bastante grande, o gráfico resultante da união dos pontos médios se aproxima de uma curva suave.

Esta curva é chamada **curva densidade de probabilidade**.

Se eu sorteio casualmente um indivíduo desta população, qual a probabilidade de seu IMC estar entre 28 e 36 kg/m<sup>2</sup> ?



A probabilidade  $P(28 < X \leq 36)$  equivale à área sombreada sob a curva.

Se esta curva é dada matematicamente por uma função  $f(x)$ , esta função é chamada **função densidade de probabilidade**. Os cálculos de probabilidades podem tornar-se bastante complexos...

$$P(28 < X \leq 36) = \int_{28}^{36} f(x) dx$$

# Distribuição normal

Dizemos que a variável aleatória contínua  $X$  segue uma distribuição normal se a sua **função densidade de probabilidade**  $f(x)$  é dada por

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]$$

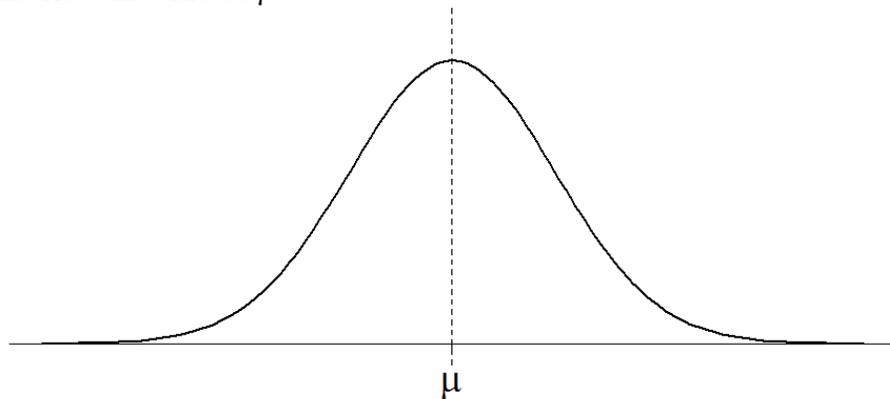
Em que :  $x$  é uma observação da variável  $X$

$\mu$  é a média (populacional) da variável  $X$

$\sigma$  é o desvio padrão (populacional) da variável  $X$  (sendo  $\sigma$  sempre maior que 0)

$\pi$  é a constante pi = 3,141592...

$f(x)$  tem a forma simétrica em torno de  $\mu$ :



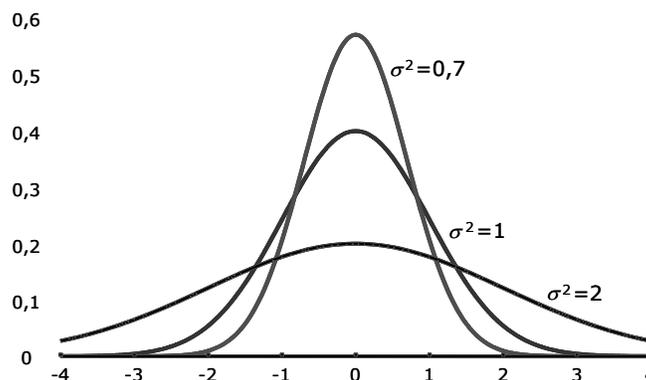
Seja uma variável aleatória  $X$  com distribuição normal com média  $\mu$  e variância  $\sigma^2$ .

Notação:

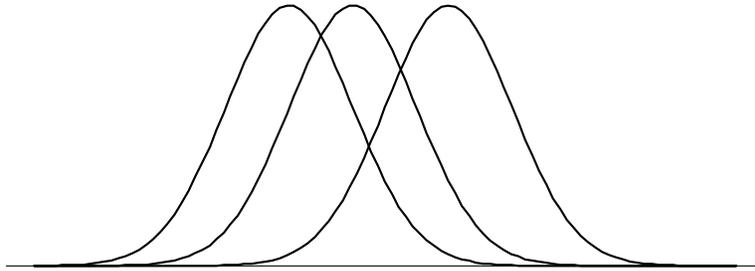
$$X \sim N(\mu ; \sigma^2)$$

$\swarrow$                        $\searrow$   
 média                      variância

A variância  $\sigma^2$  é responsável pela **forma** da curva:



A média  $\mu$  é responsável pela **locação (posição)** da curva:

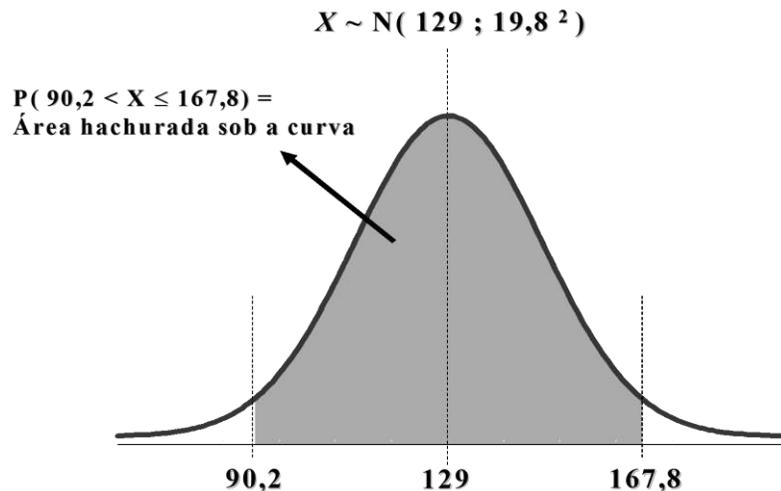


**Exemplo:** seja  $X$  uma variável aleatória que representa a pressão sanguínea sistólica. Vamos considerar que, para a população de homens de 18 a 74 anos de certo país, a pressão sistólica tem distribuição aproximadamente normal com média de 129 mmHg e desvio padrão de 19,8 mmHg.

$$X \sim N(129; 19,8^2)$$

Qual é a probabilidade de encontrarmos nesta população um indivíduo com pressão sanguínea sistólica entre 90,2 mmHg e 167,8 mmHg ?

$$P(90,2 < X \leq 167,8) = ?$$



Quanto ao uso das tabelas...

O problema é que temos infinitas possibilidades para  $\mu$  e infinitas possibilidades para  $\sigma^2$ .

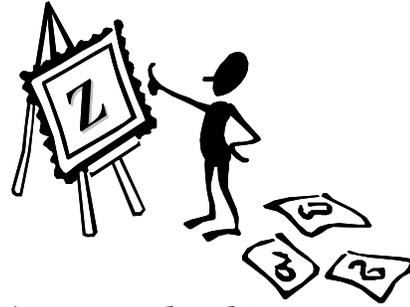
$$X \sim N(\mu; \sigma^2)$$

Geralmente, apenas temos disponíveis tabelas para o caso especial em que  $\mu = 0$  e  $\sigma^2 = 1$ .

**SOLUÇÃO:** Podemos transformar qualquer variável  $X$  com distribuição normal com média  $\mu$  e variância  $\sigma^2$  em uma variável  $Z$  com média  $0$  e variância  $1$ .

$$Z \sim N(0; 1)$$

Se  $X \sim N(\mu; \sigma^2)$  temos que  $Z = \frac{X - \mu}{\sigma}$



segue distribuição normal com média **0** e variância **1**.

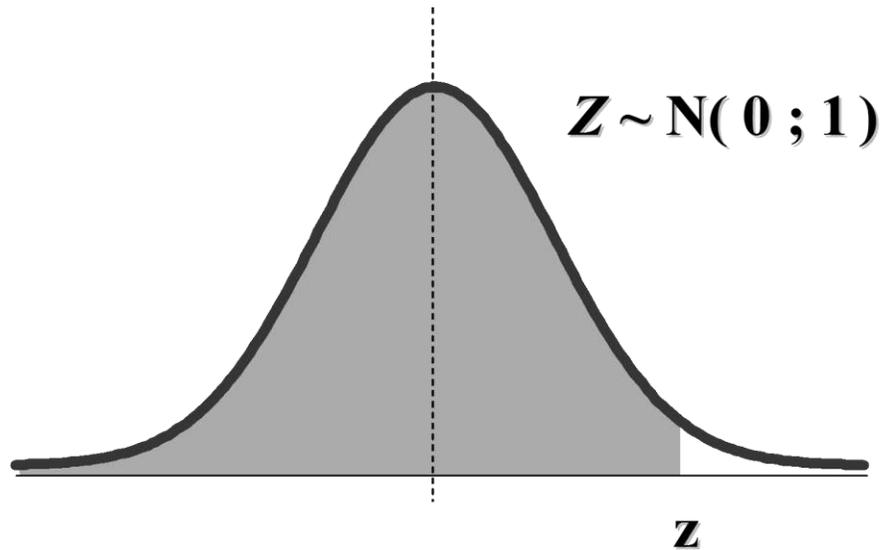
Quando  $Z \sim N(0; 1)$ , dizemos que **Z** segue distribuição *normal padrão* ou *normal standard*.

Um resultado da variável aleatória **Z** é conhecido como um *score z*.

GeoGebra

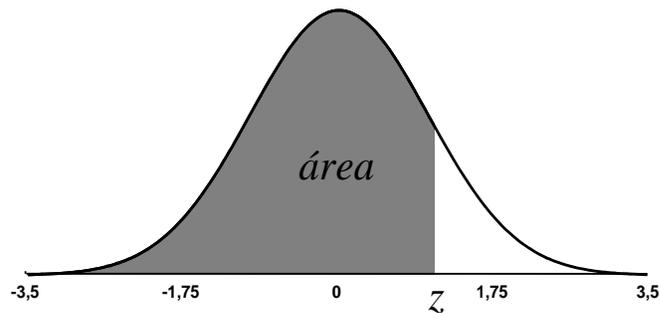
<https://app.geogebra.org/#probability>

## Tabelas da distribuição normal



Trazem  $P(Z < z)$

**TABELA 1A:**  
 Uso quando  $z$  é positivo



**TABELA 1B:**  
 Uso quando  $z$  é negativo

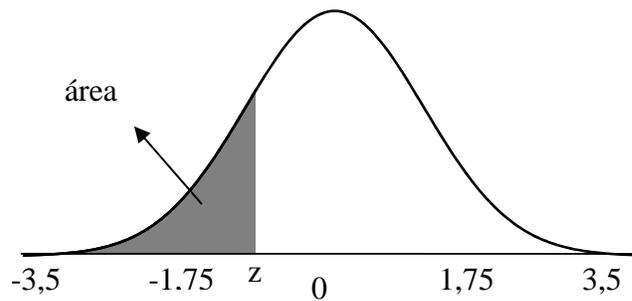


TABELA 1A

$z$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

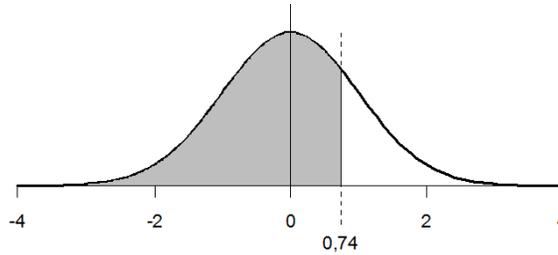
TABELA 1B

$z$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002

Por exemplo,

$$P(Z < 0,74) = ?$$

$$P(Z < 0,74) = 77,04\%$$

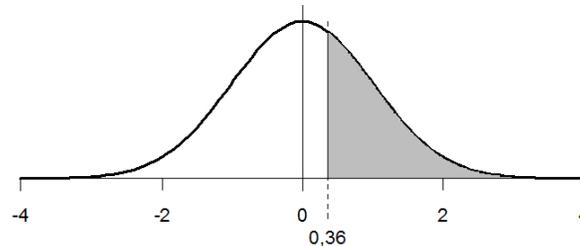


z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389

Por exemplo,

$$P(Z > 0,36) = ?$$

$$P(Z > 0,36) = 1 - 0,6406 \\ = 35,94\%$$

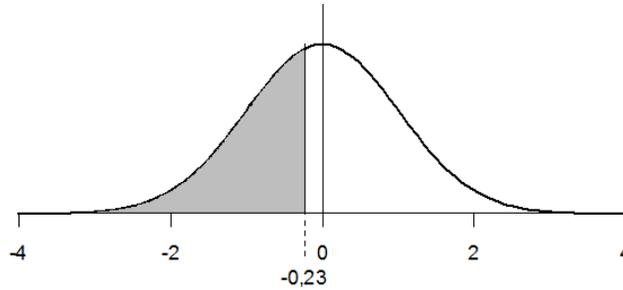


z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389

Por exemplo,

$$P(Z < -0,23) = ?$$

$$P(Z < -0,23) = 40,9\%$$



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611

Por exemplo,  $P(Z > -0,56) = ?$

**Voltando ao exemplo:**

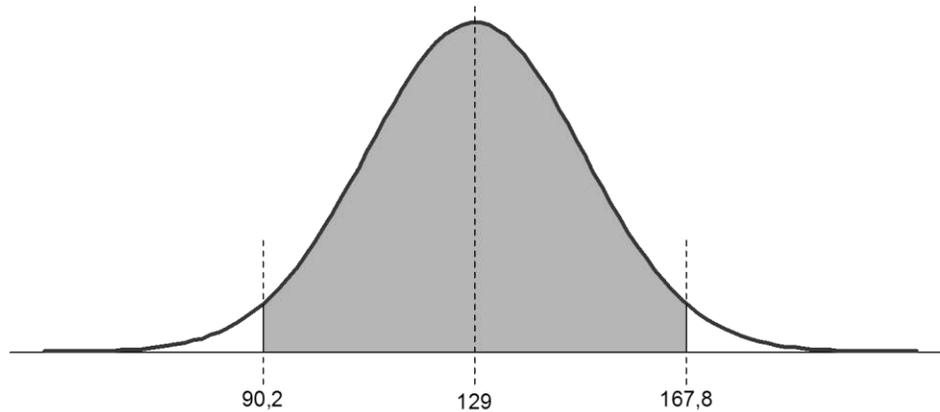
Seja  $X$  uma variável aleatória que representa a pressão sanguínea sistólica. Vamos considerar que, para a população de homens de 18 a 74 anos de um certo país, a pressão sistólica tem distribuição aproximadamente normal com média de 129 mmHg e desvio padrão de 19,8 mmHg.

$$X \sim N(129 ; 19,8^2)$$

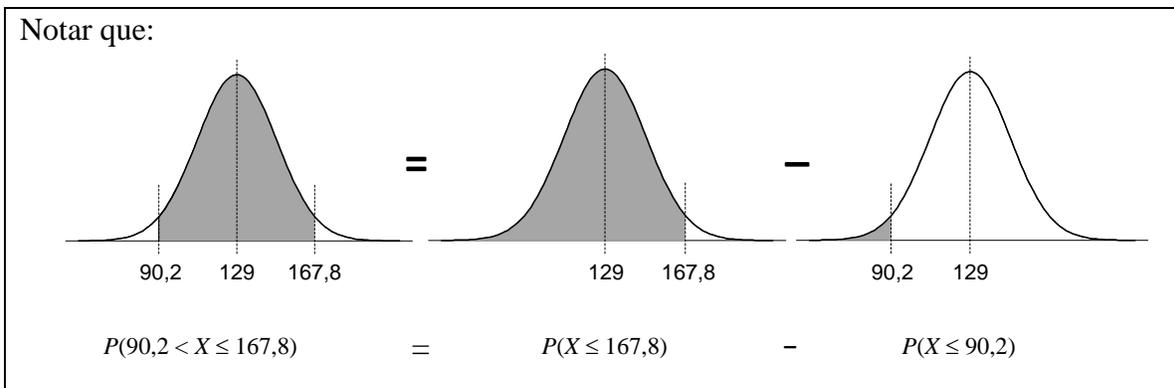
Qual é a probabilidade de encontrarmos nesta população um indivíduo com pressão sanguínea sistólica entre 90,2 mmHg e 167,8 mmHg ?

$$P(90,2 < X \leq 167,8) = ?$$

A probabilidade  $P(90,2 < X \leq 167,8)$  corresponde à área sombreada na curva a seguir.



Notar que:



$$X \sim N(129; 19,8^2)$$

$$P(X \leq 167,8) = ?$$

$$P(X \leq 90,2) = ?$$

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 129}{19,8}$$

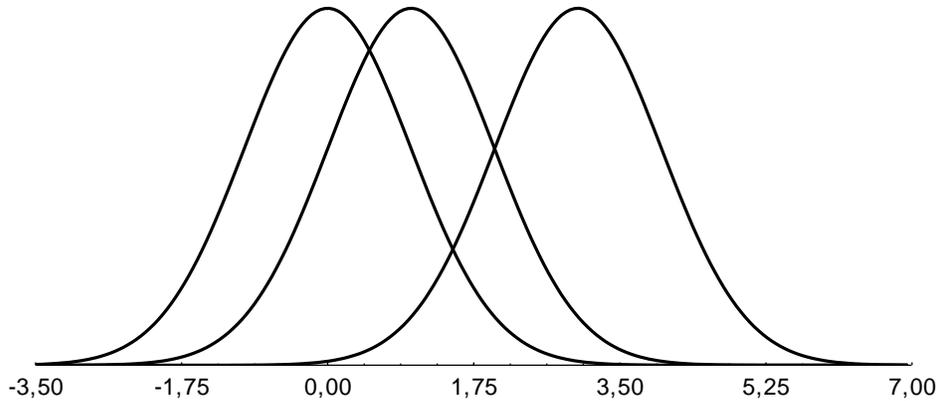
$$P(X \leq 167,8) = P\left(Z \leq \frac{167,8 - 129}{19,8}\right) = P(Z \leq 1,96) = \underline{\hspace{2cm}}$$

$$P(X \leq 90,2) = P\left(Z \leq \frac{90,2 - 129}{19,8}\right) = P(Z \leq -1,96) = \underline{\hspace{2cm}}$$

**Resposta:**  $P(90,2 < X \leq 167,8) = \underline{\hspace{4cm}}$

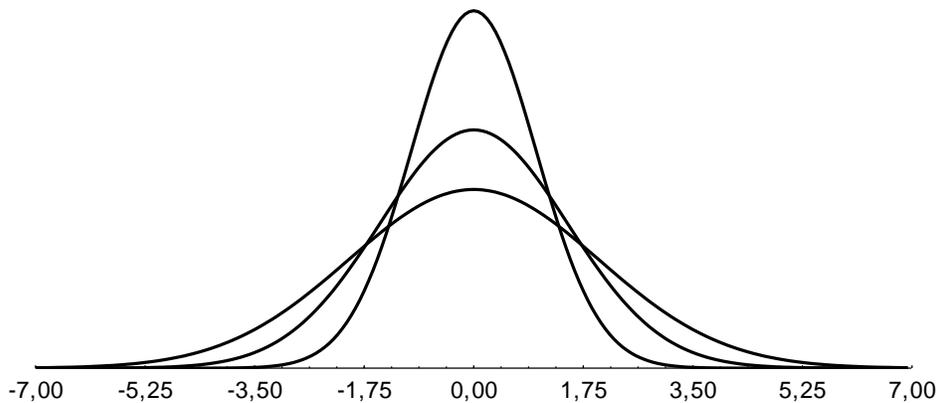
**Exercícios**

1. Observe a figura abaixo e escolha a alternativa mais adequada.



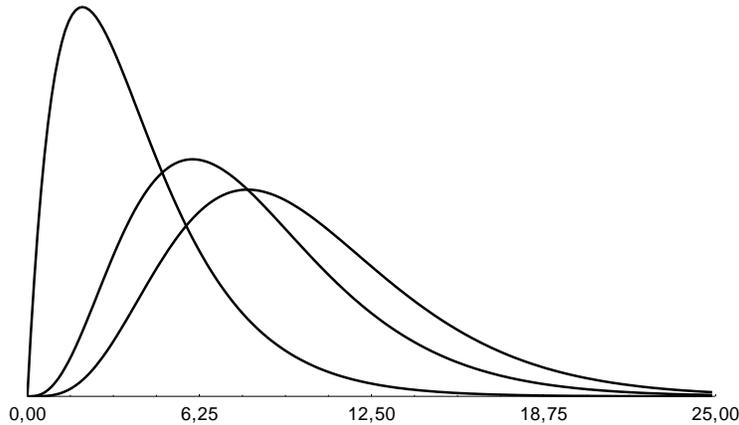
- (a) A figura representa as distribuições de três variáveis, todas com distribuição normal, médias iguais e variâncias iguais.
- (b) A figura representa as distribuições de três variáveis, todas com distribuição normal, médias diferentes e variâncias iguais.
- (c) A figura representa as distribuições de três variáveis, todas com distribuição normal, médias iguais e variâncias diferentes.
- (d) Nenhuma das alternativas acima é adequada.

2. Observe a figura abaixo e escolha a alternativa mais adequada.



- (a) A figura representa as distribuições de três variáveis, todas com distribuição normal, médias iguais e variâncias iguais.
- (b) A figura representa as distribuições de três variáveis, todas com distribuição normal, médias diferentes e variâncias iguais.
- (c) A figura representa as distribuições de três variáveis, todas com distribuição normal, médias iguais e variâncias diferentes.
- (d) Nenhuma das alternativas acima é adequada.

3. Observe a figura abaixo e escolha a alternativa mais adequada.



- (a) A figura representa as distribuições de três variáveis, todas com distribuição normal, médias iguais e variâncias iguais.  
 (b) A figura representa as distribuições de três variáveis, todas com distribuição normal, médias diferentes e variâncias iguais.  
 (c) A figura representa as distribuições de três variáveis, todas com distribuição normal, médias iguais e variâncias diferentes.  
 (d) Nenhuma das alternativas acima é adequada.

4. Seja  $Z$  uma variável aleatória com distribuição normal padrão. Encontre:

- (a)  $P(0 \leq Z \leq 1,42)$       (d)  $P(0,65 \leq Z \leq 1,26)$       (g)  $P(-0,5 \leq Z \leq 0,5)$   
 (b)  $P(-0,73 \leq Z \leq 0)$       (e)  $P(-1,79 \leq Z \leq -0,54)$   
 (c)  $P(-1,37 \leq Z \leq 2,01)$       (f)  $P(Z \geq 1,13)$

5. Seja  $Z$  uma variável aleatória com distribuição normal padrão. Encontre o valor de  $a$  se:

- (a)  $P(0 \leq Z \leq a) = 0,4236$       (b)  $P(Z \leq a) = 0,7967$       (c)  $P(a \leq Z \leq 2) = 0,1$

6. Seja  $X$  uma variável aleatória com distribuição normal, com média 8 e desvio padrão 4. Encontre:

- (a)  $P(5 \leq X \leq 10)$       (c)  $P(X \geq 15)$   
 (b)  $P(10 \leq X \leq 15)$       (d)  $P(X \leq 5)$

7. (Soares & Siqueira, 2002) A vida de um certo aparelho cirúrgico pode ser descrita pela distribuição normal com média de oito anos e desvio padrão de 1,4 anos. O contrato de garantia diz que o fabricante substituirá os aparelhos que apresentarem defeito dentro do prazo de garantia. Se ele deseja substituir no máximo 5% dos aparelhos que fabrica, qual deve ser o prazo de garantia a ser estabelecido ?

8. (Soares & Siqueira, 2002) Um teste de aptidão para o exercício de certa profissão exige uma sequência de operações a serem executadas rapidamente uma após a outra. Para passar no teste, o candidato deve completá-lo em 80 minutos no máximo. Admita que o tempo para completar o teste seja uma variável aleatória com distribuição  $N(90;20^2)$ .

- (a) Qual porcentagem de candidatos tem chance de ser aprovados ?

- (b) Os melhores 5% receberão um certificado especial. Qual o tempo máximo para fazer jus a tal certificado ?
9. (Soares & Siqueira, 2002) A distribuição da altura de 500 estudantes do sexo masculino de uma escola é aproximadamente normal, com média 1,70 m e desvio padrão 2,5 cm.
- (a) Quantos têm altura inferior a 1,75 m ?
- (b) Quantos têm altura entre 1,72 e 1,80 m ?
10. (Soares & Siqueira, 2002) É sabido que, para os adultos do sexo masculino, gozando de boa saúde, em uma certa população, a temperatura corporal segue distribuição normal com média de 36,8 graus e desvio padrão de 0,15 graus. Se considerarmos 1000 dessas pessoas, quantas se esperariam com temperatura entre 36,8 e 37,2 graus ?
11. (Soares & Siqueira, 2002) Para a população masculina dos Estados Unidos (1976-1980) com idade entre 18 e 74 anos, a pressão sistólica tem distribuição aproximadamente normal com média 129 mmHg e desvio padrão 19,8 mmHg. Recomendações do *Joint National Commitee of Hypertension* e *American Heart Association* consideram níveis pressóricos normais menores que 130/85 mmHg.
- (a) Qual a probabilidade de um homem dessa população possuir pressão sistólica normal ?
- (b) Selecionando-se ao acaso 1000 homens dessa população, quantos seriam diagnosticados com hipertensão moderada (pressão sistólica entre 160 e 179 mmHg) ?
12. (Soares & Siqueira, 2002) A fosfatase alcalina em uma população de pessoas gozando de boa saúde tem distribuição normal com média 42 mU/dL e desvio padrão 13 mU/dL. Calcule a porcentagem de pessoas com fosfatase alcalina entre 15 e 69 mU/dL.
13. (Soares & Siqueira, 2002) Desde o isolamento em 1983 do *Helicobacter pylori* na mucosa gástrica humana, inúmeros estudos têm sido realizados objetivando determinar uma possível relação causal entre sua presença e algumas entidades gastroduodenais: úlceras, gastrite crônica, etc. A presença do microrganismo tem sido diagnosticada através do exame de cultura. Um outro método, mais simples e rápido é o teste respiratório que emprega uréia marcada com carbono 14 ( $C_{14}$ ). Por possuir uma urease, enzima capaz de degradar uréia a gás carbônico  $CO_2$ , a presença da bactéria pode ser evidenciada pela detecção de carbono marcado no ar expirado após administração, por via oral, da uréia marcada. Suponha que a quantidade de  $C_{14}$ , liberada sob a forma de  $CO_2$  para pacientes não portadores da bactéria *H. pylori*, seja uma variável com distribuição aproximadamente normal com média 0,07 unidades de  $C_{14}$  e desvio padrão igual a 0,03 unidades de  $C_{14}$ . A partir dessas informações, calcular:
- (a) A probabilidade de uma pessoa não infectada liberar entre 0,04 e 0,10 unidade de  $C_{14}$ .
- (b) A probabilidade de uma pessoa não infectada liberar mais de 0,15 unidade de  $C_{14}$ .
14. O índice de massa corporal (IMC) é uma medida frequentemente utilizada para indicar se um indivíduo está abaixo ou acima de seu peso ideal. O IMC é calculado por  $IMC = \text{peso} / (\text{altura})^2$ , sendo o peso medido em quilogramas e a altura medida em metros. Alguns pesquisadores utilizam a tabela a seguir, para classificar indivíduos adultos como portadores de baixo peso, peso normal, pré-obeso, e portadores de obesidade classe I, II ou III.

classificação	IMC ( $\text{kg}/\text{m}^2$ )
baixo peso	$< 18,5$
peso saudável (normal)	$18,5 \text{ -- } 25,0$
pré-obeso	$25,0 \text{ -- } 30,0$
obesidade classe I	$30,0 \text{ -- } 35,0$
obesidade classe II	$35,0 \text{ -- } 40,0$
obesidade classe III	$\geq 40,0$

Em um grande município, o IMC dos indivíduos adultos da população tem média  $\mu = 25,2$   $\text{kg}/\text{m}^2$  e desvio padrão  $\sigma = 4,5$   $\text{kg}/\text{m}^2$ . Considerando que o IMC tem distribuição normal nesta população, encontre:

- a proporção de indivíduos de baixo peso;
- a proporção de indivíduos com peso saudável (normal);
- a proporção de indivíduos pré-obesos;
- a proporção de indivíduos com obesidade classe I;
- a proporção de indivíduos com obesidade classe II e
- a proporção de indivíduos com obesidade classe III.

**15.** Na análise antropométrica do estado nutricional, é frequentemente utilizado o índice de massa corporal (IMC), calculado a partir da divisão da massa corporal em quilogramas pela estatura em metro elevado ao quadrado ( $\text{kg}/\text{m}^2$ ). Os pontos de corte recomendados pela Organização Mundial de Saúde (OMS) para avaliação da população adulta e idosa são: magreza severa ( $\text{IMC} < 16,0$   $\text{kg}/\text{m}^2$ ), magreza moderada ( $16,0$   $\text{kg}/\text{m}^2 \leq \text{IMC} < 17,0$   $\text{kg}/\text{m}^2$ ), magreza leve ( $17,0$   $\text{kg}/\text{m}^2 \leq \text{IMC} < 18,5$   $\text{kg}/\text{m}^2$ ), adequado ( $18,5$   $\text{kg}/\text{m}^2 \leq \text{IMC} < 25,0$   $\text{kg}/\text{m}^2$ ), sobrepeso grau I ( $25,0$   $\text{kg}/\text{m}^2 \leq \text{IMC} < 30,0$   $\text{kg}/\text{m}^2$ ), sobrepeso grau II ( $30,0$   $\text{kg}/\text{m}^2 \leq \text{IMC} < 40,0$   $\text{kg}/\text{m}^2$ ) e sobrepeso grau III ( $\text{IMC} \geq 40,0$   $\text{kg}/\text{m}^2$ ). Em uma determinada população, o IMC tem média  $\mu = 26,4$   $\text{kg}/\text{m}^2$  e desvio padrão  $\sigma = 3,9$   $\text{kg}/\text{m}^2$ . Suponha que o IMC segue uma distribuição normal nesta população. Encontre:

- A proporção de indivíduos com magreza severa;
- A proporção de indivíduos com magreza moderada;
- A proporção de indivíduos com magreza leve;
- A proporção de indivíduos com peso adequado;
- A proporção de indivíduos com sobrepeso grau I;
- A proporção de indivíduos com sobrepeso grau II; e
- A proporção de indivíduos com sobrepeso grau III;



<https://www.youtube.com/watch?v=ENlygEM0CWs>

**16.** Em uma população de indivíduos a média da circunferência do braço (CB) é 26,4 cm e o desvio padrão é 2,3 cm. Se nesta população a CB segue uma distribuição normal, encontre a probabilidade de um indivíduo aleatoriamente selecionado desta população possuir:

- CB menor ou igual a 28,5 cm.
- CB maior ou igual a 28,5 cm.
- CB maior que 30 cm.
- CB menor que 25 cm.

17. Sabemos que as alterações específicas do desenvolvimento da linguagem devem ser identificadas precocemente, pois tais alterações podem interferir nos aspectos sociais e escolares da criança. Um pesquisador propôs um índice que mede o desenvolvimento da linguagem em crianças. Chamaremos este índice de IDL. Valores do IDL entre 68 e 85 pontos teriam implicações em termos de acompanhamento. Valores abaixo de 68 pontos indicariam atraso significativo em uma ou mais áreas do desenvolvimento da linguagem. Seja uma população de crianças de 3 a 6 anos, onde o IDL tem distribuição normal com média 90 e desvio padrão de 10 pontos. Encontre:

- (a) A proporção de crianças desta população com IDL abaixo de 68 pontos.
- (b) A proporção de crianças desta população com IDL entre 68 e 85 pontos.
- (c) A proporção de crianças desta população com IDL acima de 85 pontos.

18. Suponha que o tempo médio de permanência de pacientes com doenças crônicas em um hospital é de 50 dias, com desvio padrão igual a 10 dias. Admitindo que o tempo de permanência segue uma distribuição normal, qual é a probabilidade de um paciente permanecer no hospital:

- (a) mais de 30 dias ?
- (b) menos de 30 dias ?
- (c) mais de 50 dias ?
- (d) entre 40 e 60 dias ?
- (e) entre 35 e 70 dias ?

19. (Díaz e López, 2007) Entre os diabéticos, podemos supor que o nível de glicose no sangue, denotado por  $X$ , segue uma distribuição aproximadamente normal, com média 106 mg/100 ml e desvio padrão 8 mg/100 ml.

- (a) Encontre  $P(X \leq 120)$ .
- (b) Encontre a porcentagem de diabéticos com níveis compreendidos entre 90 e 120 mg/100 ml.
- (c) Encontre  $P(106 \leq X \leq 110)$ .
- (d) Encontre o ponto  $k$  caracterizado pela propriedade de que 25% de todos os diabéticos têm um nível de glicose em jejum inferior ou igual a  $k$ .

20. Uma nutricionista pretende conduzir um estudo de avaliação do estado nutricional de adultos saudáveis de um município. Ela obteve a informação que a concentração sérica de albumina na população dos indivíduos saudáveis de sexo masculino deste município, com idade entre 30 e 49 anos, tem média de 4,6 g/dl e desvio padrão de 0,5 g/dl. Se a distribuição desta variável nesta população aproxima-se de uma curva normal, encontre a probabilidade de um indivíduo escolhido ao acaso apresentar uma concentração sérica de albumina:

- (a) maior que 4,6 g/dl;
- (b) menor que 5,0 g/dl;
- (c) entre 3,7 g/dl e 5,5 g/dl;
- (d) maior que 6,5 g/dl.
- (e) Encontre um valor  $k$  tal que 25% dos indivíduos desta população possuam valores de concentração sérica de albumina menores que  $k$  g/dl.

21. Em uma população de crianças de sexo masculino que estão no período de dentição mista, o comprimento do lábio superior tem uma média de 21,5 mm e um desvio padrão de 1,8 mm. Suponha que esta variável segue uma distribuição aproximadamente normal nesta população. Se uma criança é aleatoriamente selecionada desta população, qual é a probabilidade dela possuir lábio superior com comprimento

- (a) maior que 21,5 mm?
- (b) entre 19,0 mm e 23,0 mm?
- (c) maior que 30,0 mm?

(d) Encontre um valor  $k$  tal que 95% das crianças desta população possuem comprimento do lábio superior menor que  $k$  mm.

22. A profissional responsável pela clínica ortodôntica de um grande hospital universitário observou que o tempo de permanência em tratamento ortodôntico de seus pacientes segue uma distribuição aproximadamente normal, com média de 28,5 meses e um desvio padrão de 6 meses.

(a) Qual é o percentual de pacientes com tempo de permanência em tratamento ortodôntico maior que 40 meses?

(b) Qual é o percentual de pacientes com tempo de permanência em tratamento ortodôntico maior que 6 meses?

(c) Encontre um valor  $k$  tal que 90% dos pacientes desta população tiveram um tempo de permanência em tratamento maior que  $k$  dias.

### RESPOSTAS:

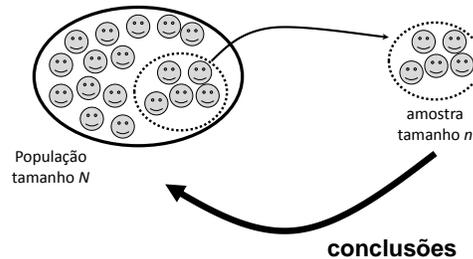
- | 1. (b)                         | 2. (c)                                  | 3. (d)  |                            |                      |           |          |
|--------------------------------|---|---|----------------------------|----------------------|-----------|----------|
| 4. (a) 0,4222<br>(b) 0,2673    | (c) 0,8925<br>(d) 0,1540                | (e) 0,2579<br>(f) 0,1292  | (g) 0,3830                 |                      |           |          |
| 5. (a) $a = 1,43$              | (b) $a = 0,83$                          | (c) $a = 1,16$  |                            |                      |           |          |
| 6. (a) 0,4649                  | (b) 0,2684                              | (c) 0,0401  | (d) 0,2266                 |                      |           |          |
| 7. 5,69 anos.                  | 8. (a) 30,85%<br>(b) 57 minutos.        | 9. (a) Aproximadamente 489 estudantes.<br>(b) Aproximadamente 106 estudantes. |                            |                      |           |          |
| 10. 496 graus.                 | 11. (a) Aproximadamente 52%<br>(b) 5,3% | 12. 96,2%   | 13. (a) 68,2%<br>(b) 0,4%  |                      |           |          |
| 14. (a) 6,8%                   | (b) 41,4%                               | (c) 37,5%   | (d) 12,8%                  | (e) 1,4%             | (f) 0,1%  |          |
| 15. (a) 0,4%                   | (b) 0,4%                                | (c) 1,3%  | (d) 33,9%                  | (e) 46,2%            | (f) 17,7% | (g) 0,1% |
| 16. (a) 0,8186                 | (b) 0,1814                              | (c) 0,0594  | (d) 0,2709                 |                      |           |          |
| 17. (a) 0,0139                 | (b) 0,2946                              | (c) 0,6915  |                            |                      |           |          |
| 18. (a) 0,9772                 | (b) 0,0228                              | (c) 0,5   | (d) 0,6826                 | (e) 0,9104           |           |          |
| 19. (a) 0,9599                 | (b) 0,937                               | (c) 0,191   | (d) $k = 100,6$ mg/100 ml. |                      |           |          |
| 20. (a) 0,50                   | (b) 0,788                               | (c) 0,928   | (d) 0                      | (e) $k = 4,26$ g/dl. |           |          |
| 21. (a) 0,50                   | (b) $0,7967 - 0,0823 = 0,7144$          | (c) 0   | (d) $k = 24,5$ mm.         |                      |           |          |
| 22. (a) $P(Z > 1,92) = 2,74\%$ | (b) $P(Z > -3,75) \approx 100\%$        | (c) $P(X > k) = 90\%$ , $k = 36,2$ dias.                                      |                            |                      |           |          |

## Inferência estatística

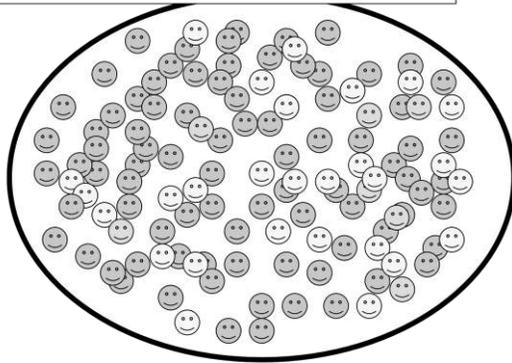


<https://www.youtube.com/watch?v=VIBMFNfFt5g>

**Objetivo:** obter conclusões sobre algumas características de um conjunto de interesse, denominado população, com base na informação oriunda de um conjunto de dados disponíveis, denominado amostra.



**POPULAÇÃO:** conjunto total de indivíduos sobre o qual se faz uma inferência



A **população** é o conjunto constituído por todos os indivíduos que apresentam pelo menos uma característica comum.

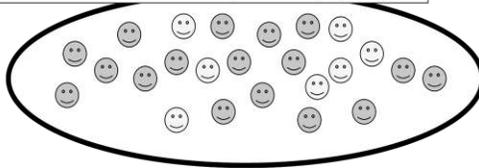
Um parâmetro é uma característica numérica de uma *população*.

Seja  $X$  uma variável de interesse.

A média populacional da variável  $X$ ,  $\mu$ , é um parâmetro.  
A variância populacional da variável  $X$ ,  $\sigma^2$ , é um parâmetro.

**Um parâmetro é um número fixo, mas geralmente não conhecemos seu valor.**

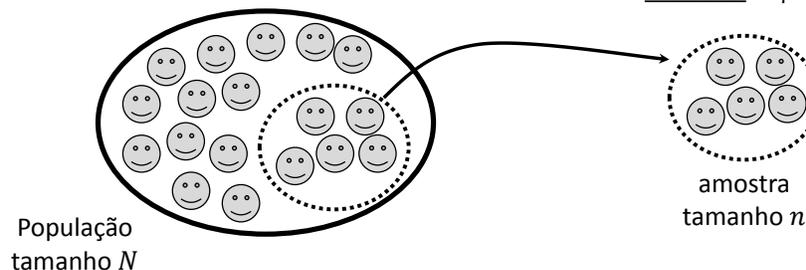
**AMOSTRA:** um subconjunto da população



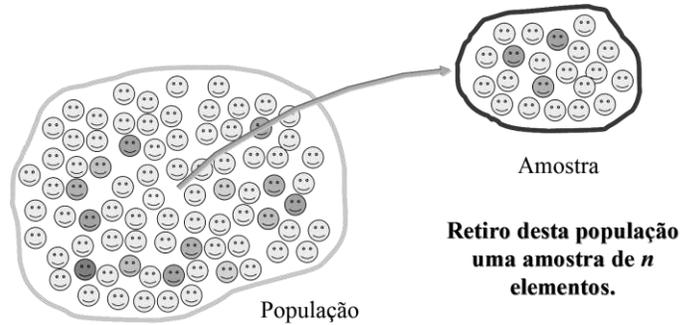
A **amostra** é um subconjunto, uma parte selecionada da totalidade de observações abrangidas pela população, através da qual se faz uma inferência sobre um ou mais **parâmetros** da população.

$\mu$ , a média populacional de  $X$ ,  
é um parâmetro

$\bar{x}$ , a média amostral de  $X$ ,  
é uma estimativa de  $\mu$



**Estimativas:** quantidades calculadas da amostra com a finalidade de representar um parâmetro de interesse.



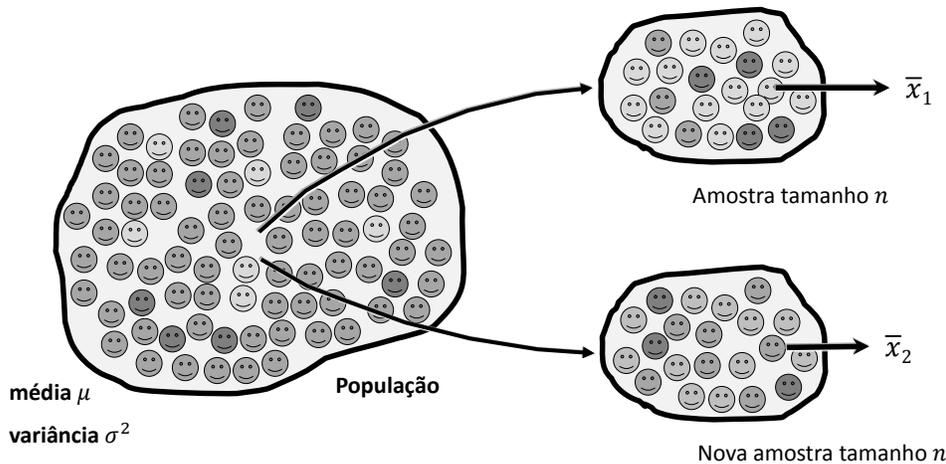
$X$  é uma variável (contínua) de interesse.

Na população,  $X$  tem média  $\mu$  e variância  $\sigma^2$  (são parâmetros!)

Baseado nesta amostra, encontro a média amostral  $\bar{x}$ .

$\bar{x}$  será igual a  $\mu$  ????

**Vamos retirar uma nova amostra desta mesma população. Esta nova amostra também terá tamanho  $n$ .**

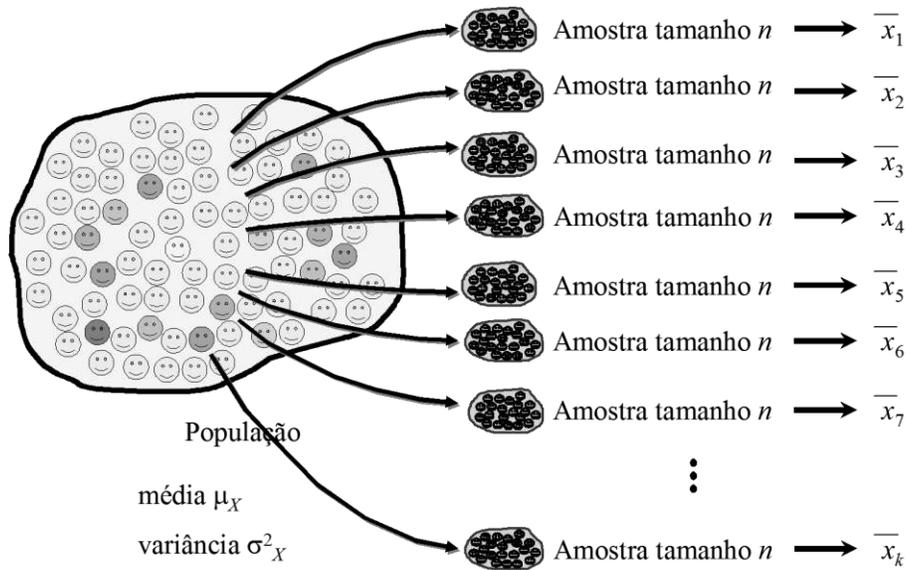


As médias amostrais  $\bar{x}_1$  e  $\bar{x}_2$  são iguais ???

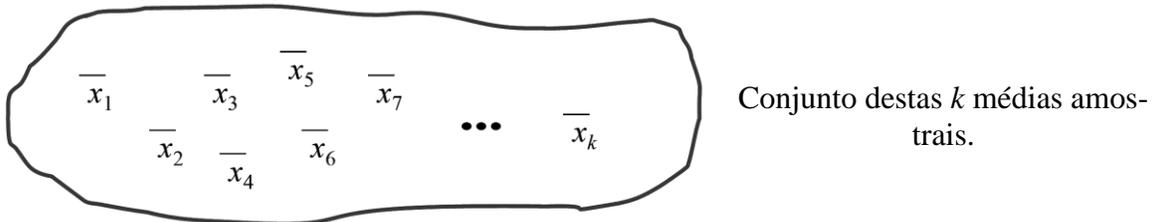
**Vamos agora imaginar a seguinte situação:**

**Estamos retirando  $k$  diferentes amostras tamanho  $n$  da mesma população. Em cada amostra, obtemos uma média amostral.**

**(É claro que, na realidade, ninguém tira  $k$  amostras tamanho  $n$  de uma população, mas uma amostra da população !!!)**

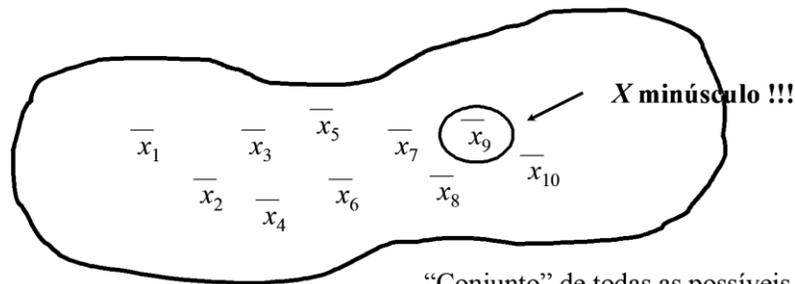


Temos  $k$  médias amostrais, baseadas em diferentes amostras tamanho  $n$  retiradas de uma única população.



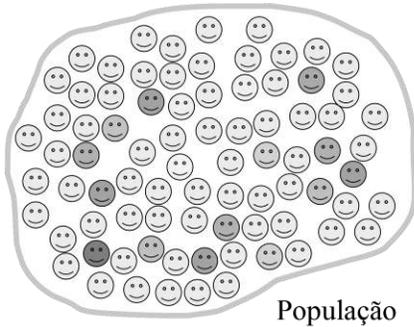
Vamos agora imaginar o seguinte:

**Estamos retirando todas as possíveis diferentes amostras tamanho  $n$  da mesma população. Em cada amostra, obtemos uma média amostral.**

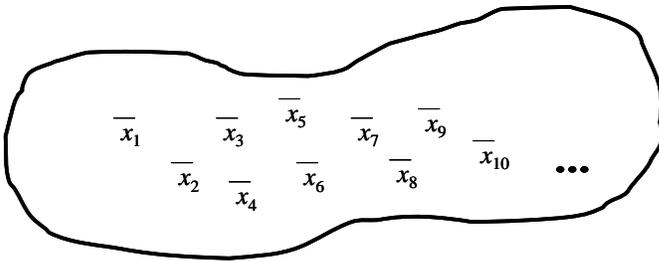


Este “conjunto” de todas as possibilidades para as médias amostrais é denotado por  $\bar{X}$ .

$X$  maiúsculo !!!



Lembremos que estas médias amostrais são baseadas em amostras de tamanho  $n$  retiradas de uma população onde  $X$ , a variável (contínua) de interesse, tem média  $\mu$  e variância  $\sigma^2$

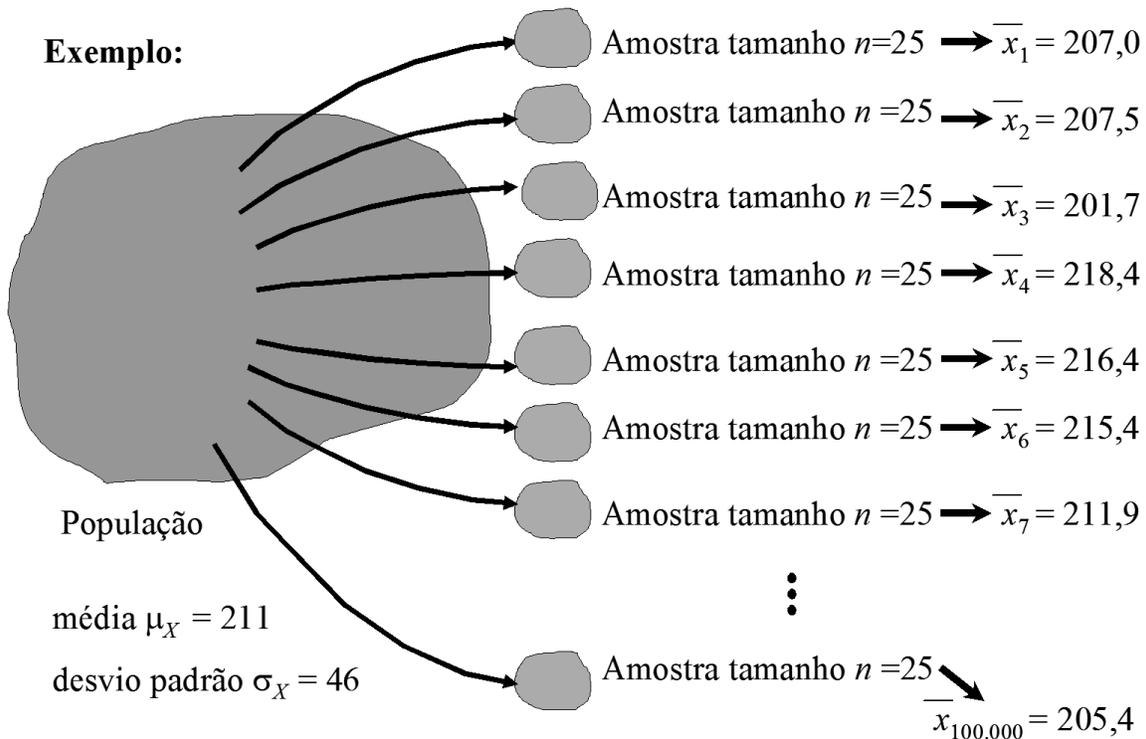


“Conjunto” de todas as possíveis médias amostrais baseadas em amostras tamanho  $n$ .

**PROPRIEDADE 1:** a média destas médias amostrais  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \dots$ , é  $\mu$ .

**PROPRIEDADE 2:** a variabilidade destas médias amostrais  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \dots$ , é  $\frac{\sigma}{\sqrt{n}}$

**Exemplo:**

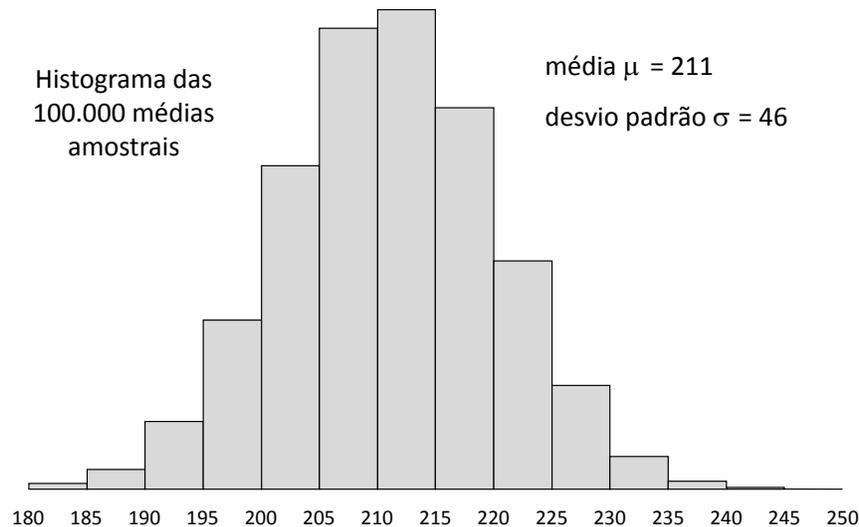


Neste exemplo, seja uma população cujo tamanho é tão grande que nós podemos considerá-lo “infinito”. Digamos que nós estamos interessados em uma variável contínua,  $X$ , que, nesta população, tem média  $\mu = 211$  e desvio padrão  $\sigma = 46$ . Usando um programa de

computador (um *software* estatístico) simulamos 100 mil amostras retiradas desta população, todas de tamanho  $n = 25$ . Em cada uma destas amostras, encontramos uma média amostral. Cada uma destas médias amostrais é uma possível observação do “conjunto”  $\bar{X}$ . Observe que cada amostra traz uma média amostral diferente.

$\bar{x}_1 = 207,0$ $\bar{x}_2 = 207,5$ $\bar{x}_3 = 201,7$ $\bar{x}_4 = 218,4$ $\bar{x}_5 = 216,4$ $\bar{x}_6 = 215,4$ $\bar{x}_7 = 211,9$ ... $\bar{x}_{100.000} = 205,4$	Média das 100.000 médias amostrais = 210,9968  Desvio padrão das 100.000 médias amostrais = 9,1764  Notar que $\frac{\sigma}{\sqrt{n}} = \frac{46}{\sqrt{25}} = 9,2$
---	--

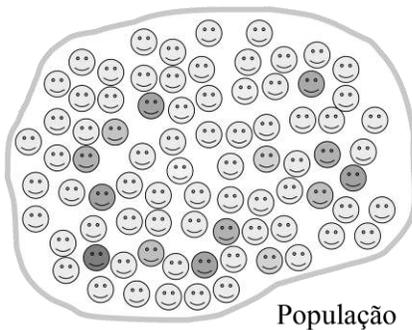
**PROPRIEDADE 3:** as médias amostrais  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \dots$ , seguem uma distribuição normal, como observamos na figura a seguir.



## Intervalos de confiança

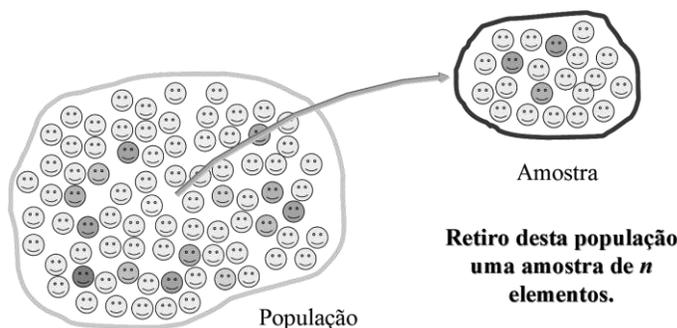
Exemplo:

The screenshot shows a Microsoft Internet Explorer browser window displaying a PubMed search result. The address bar shows the URL: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list\\_uids=16186473&query\\_hl=25](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=16186473&query_hl=25). The search result is for the article "Depression in older people in rural china." by Chen R, Wei L, Hu Z, Qin X, Copeland JR, Hemingway H. The abstract text is visible, starting with "BACKGROUND: In Western countries depression is the most common psychiatric condition in older persons and related to low socioeconomic status and low social supports. Along with social deprivation, the rural communities in China retain many Chinese traditions that involve high levels of social supports. Studying such a population might offer insights into the cause and prevention of depression that may be applicable in developing and developed countries. METHODS: Using a cross-sectional, household-based, community survey in rural China, we aimed to determine the prevalence of and risk factors for depression among older people who had a low income (mean annual per capita income of about US \$280) and high social support in 16 villages in Anhui Province. Participants included 754 men and 846 women aged 60 years or older. Depression was diagnosed using the Geriatric Mental State and the Automated Geriatric Examination for Computer Assisted Taxonomy. Risk factors, collected from the standard questionnaire and physical measurements, were examined in a stepwise multiple logistic regression model. RESULTS: The prevalence of depression (world age standardized) was 6.0% (95% confidence interval [CI], 4.8%-7.3%). Of all persons, 1374 older persons (85.9%) living



$X$  é uma variável (contínua) de interesse.

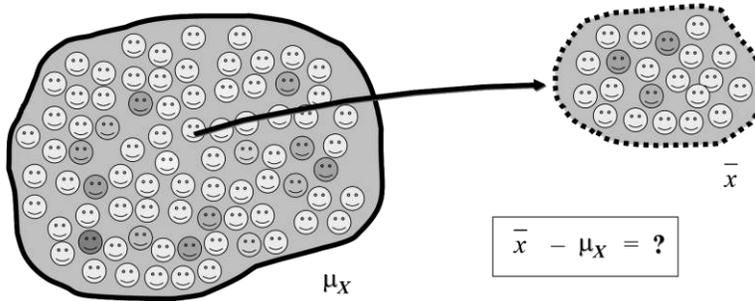
Na população,  $X$  tem média  $\mu$  e variância  $\sigma^2$  (são parâmetros)



Nosso objetivo é **estimar** a média  $\mu$  (populacional) através de uma **amostra**.

Com base nesta amostra, encontramos a média amostral  $\bar{x}$ .  
 $\bar{x}$  nos fornece uma **estimativa** de  $\mu$ .

**Problema:** o quanto  $\bar{x}$  está distante de  $\mu$  ?



A diferença  $\bar{x} - \mu$   
 é o nosso  
**erro de estimação.**

**Intervalo de confiança para a média:** Outro método de estimação, conhecido como **estimacão por intervalo**, fornece um intervalo de valores razoáveis no qual se presume que esteja o parâmetro de interesse (no nosso caso, a média  $\mu$ ) com certo grau de confiança. Este intervalo de valores é conhecido como **intervalo de confiança**.

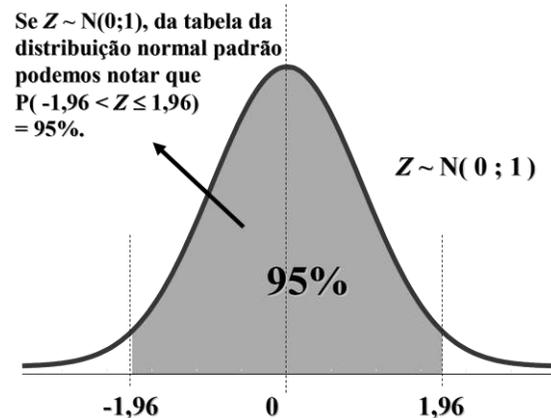
Ideia...

$$\text{se } X \sim N(\mu; \sigma^2)$$

$$\text{então } \bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right),$$

$$\text{e } Z = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$$

segue uma distribuição normal padrão.



$$P(-1,96 < Z \leq 1,96) = 0,95$$

$$P\left(-1,96 < \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} \leq 1,96\right) = 0,95$$

$$P(-1,96 \sigma < (\bar{X} - \mu)\sqrt{n} \leq 1,96 \sigma) = 0,95$$

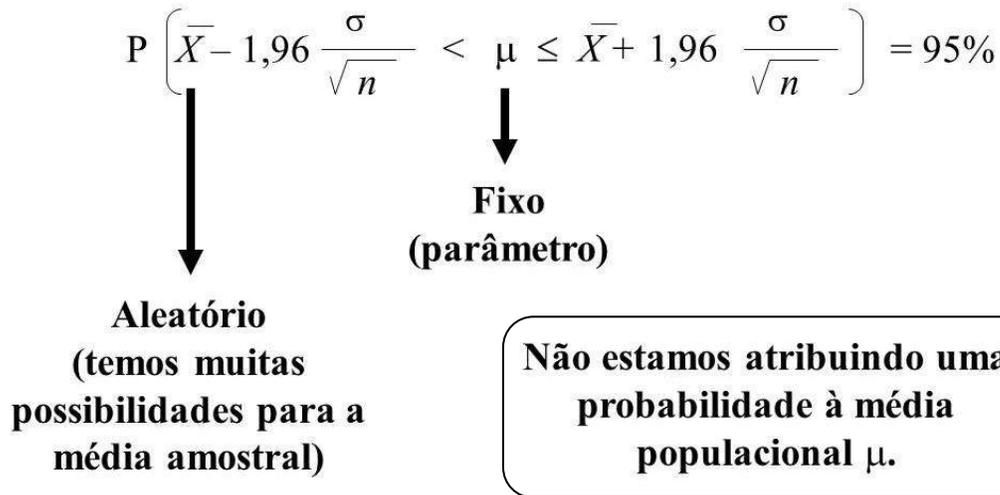
$$P\left(-1,96 \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu \leq 1,96 \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

$$\text{Temos } P\left(\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} < \mu \leq \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

$$\text{Assim, o intervalo } \left(\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}; \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}\right)$$

é chamado intervalo de confiança 95% para a média  $\mu$ .

Atenção...



**ERRADO :** a probabilidade da média populacional  $\mu$  pertencer ao intervalo

$$\left( \bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}; \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \right) \text{ é } 95\%.$$

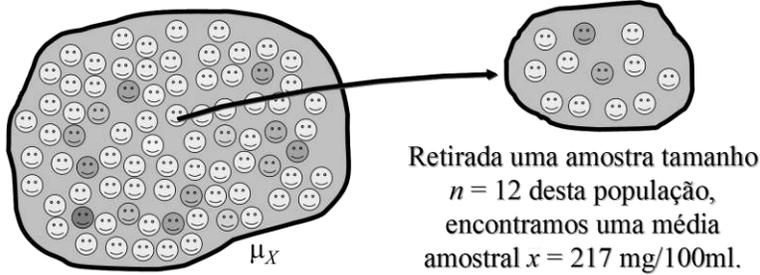
**CERTO:** a probabilidade de encontrarmos em uma amostra tamanho  $n$  um intervalo

$$\left( \bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}; \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \right) \text{ que contém a média populacional } \mu \text{ é } 95\%.$$

**Exemplo:** A distribuição dos níveis séricos de colesterol para todos os homens hipertensos e fumantes de um país tem distribuição aproximadamente normal com média  $\mu$  (desconhecida). Desejamos estimar  $\mu$ .

**Antes** que selecionemos uma amostra aleatória tamanho  $n$  da população em questão, sabemos que a probabilidade da nossa amostra gerar um intervalo

$$\left( \bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}; \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \right) \text{ que contém } \mu \text{ é } 95\%. \text{ Vamos supor que } \sigma = 46 \text{ mg/100ml.}$$



Intervalo de confiança 95% para a média  $\mu$  :  $\left( 217 - 1,96 \frac{46}{\sqrt{12}} ; 217 + 1,96 \frac{46}{\sqrt{12}} \right)$   
 ( 191 ; 243 )

Estamos 95% confiantes de que os limites de 191 e 243 contêm a verdadeira média  $\mu$ . **Não** estamos dizendo que há uma probabilidade de 95% de que  $\mu$  se encontre entre esses valores;  $\mu$  é fixo e pode estar entre 191 e 243 ou não.

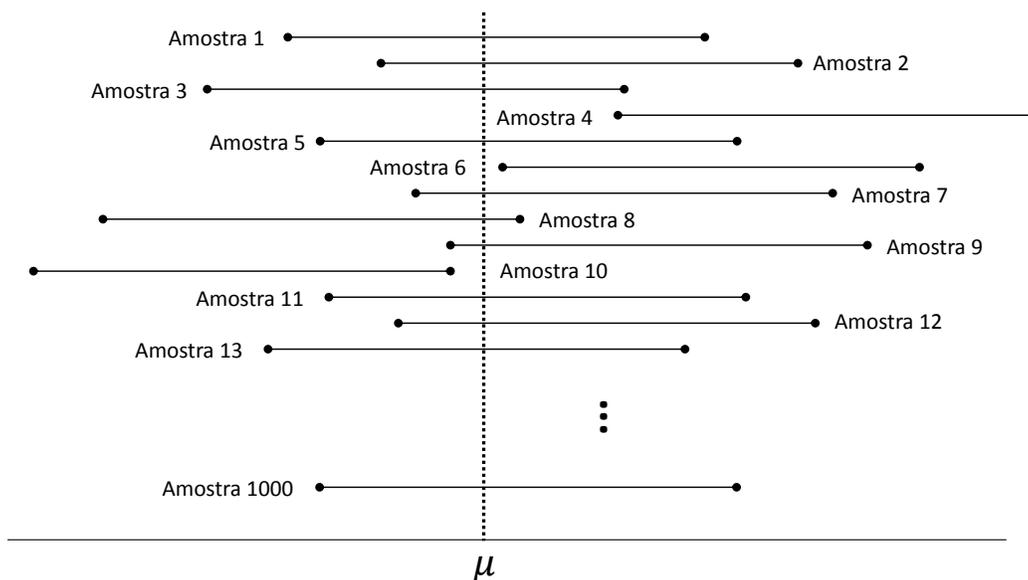
Intervalo de confiança 95% para a média  $\mu$  : ( 191 ; 243 )



Notação inadequada:  $P( 191 < \mu < 243 ) = 95\%$

Intervalo de confiança 99% para a média  $\mu$  :  $\left( 217 - 2,58 \frac{46}{\sqrt{12}} ; 217 + 2,58 \frac{46}{\sqrt{12}} \right)$   
 ( 183 ; 251 )

**Interpretação frequentista**: se retirássemos da população um número grande de amostras tamanho  $n$ , 95% destas amostras iriam gerar intervalos de confiança que contém  $\mu$ .



**Dizemos que:**  $\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}$  é o **limite inferior** do IC95%

$\bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}$  é o **limite superior** do IC95%

A diferença entre os limites superior e inferior é chamada de amplitude do IC95% para  $\mu_X$ .

Amplitude do IC95% para  $\mu$  :

$$\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} - \left( \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \right) = 2(1,96) \frac{\sigma}{\sqrt{n}}$$

**Notas:**

Quanto maior  $\sigma$ , maior é a amplitude do IC.

Quanto maior  $n$ , menor é a amplitude do IC.

**Encontrando um IC95% para a média populacional**

Podemos notar um certo problema quando utilizamos a expressão

$\left( \bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}; \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \right)$  para encontrarmos um IC95% para a média  $\mu_X$ .

O desvio padrão  $\sigma$  populacional é desconhecido !

Lembrando que, se  $X \sim N(\mu; \sigma^2)$ , usamos a propriedade de  $Z = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma}$  seguir uma distribuição normal padrão para obtermos o IC95%, podemos substituir  $\sigma$  por  $s$  (o desvio padrão amostral) na expressão de  $Z$  ?

Se a resposta for SIM, o IC95% para a média  $\mu$  será dado por:

$\left( \bar{X} - 1,96 \frac{S}{\sqrt{n}}; \bar{X} + 1,96 \frac{S}{\sqrt{n}} \right)$  ?

**Resposta:** sim, mas a quantidade  $T = \frac{(\bar{X} - \mu)\sqrt{n}}{S}$  não tem distribuição normal padrão.

Portanto, o intervalo  $\left( \bar{X} - 1,96 \frac{S}{\sqrt{n}}; \bar{X} + 1,96 \frac{S}{\sqrt{n}} \right)$  não é válido !!!

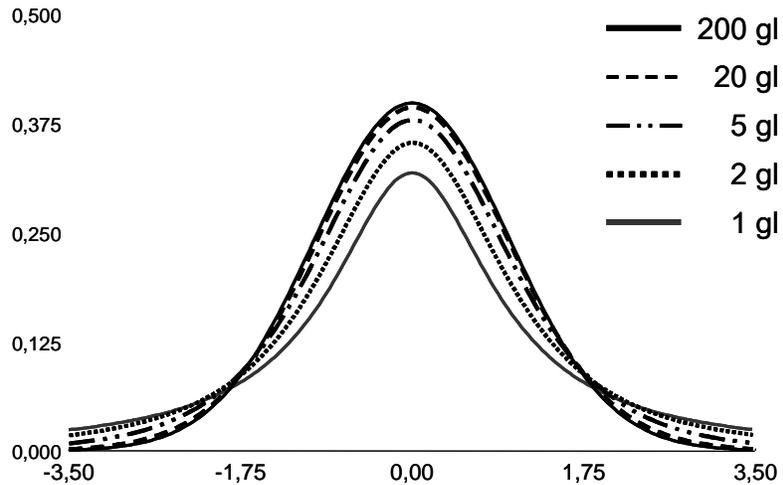
### Distribuição $t$ de Student

Se  $X \sim N(\mu; \sigma^2)$  e a amostra tamanho  $n$  foi aleatoriamente escolhida da população,

$T = \frac{(\bar{X} - \mu)\sqrt{n}}{S}$  segue distribuição  $t$  de Student com  $n - 1$  graus de liberdade (gl).

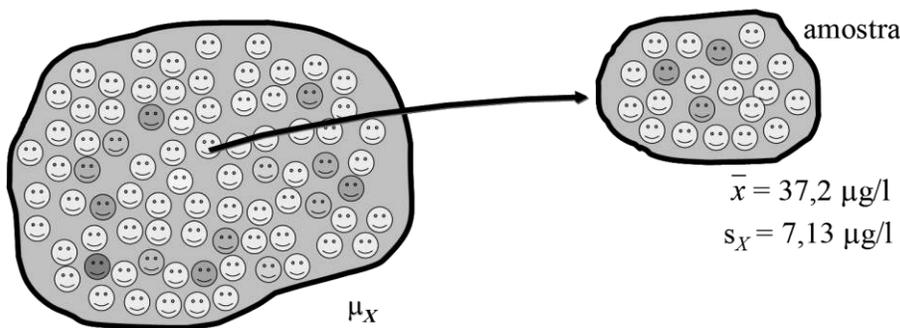
Notação:  $T \sim t_{(n-1)}$

A curva densidade de probabilidade da distribuição  $t$  tem **forma** semelhante à da distribuição normal padrão, sendo também simétrica ao redor de sua média 0.



**Exemplo:** Seja uma amostra de dez crianças selecionadas da população de bebês que recebe antiácidos que contêm alumínio e são frequentemente usados para tratar desarranjos pépticos e digestivos.

Seja a distribuição dos níveis de alumínio no plasma aproximadamente normal. O nível médio de alumínio para a amostra de dez bebês é  $\bar{x} = 37,2 \mu\text{g/l}$  e seu desvio padrão amostral é  $s = 7,13 \mu\text{g/l}$ . Calcular o IC95% para  $\mu$ .



Como  $T = \frac{(\bar{X} - \mu)\sqrt{n}}{S} \sim t_{(n-1)}$ , em uma amostra tamanho  $n = 10$ , temos que

$T = \frac{(\bar{X} - \mu)\sqrt{10}}{S}$  segue uma distribuição  $t$  de Student com  $n - 1 = 10 - 1 = 9$  graus de liberdade.

Se  $T \sim t_{(9)}$ , da tabela da distribuição  $t$  de Student podemos notar que  $P(-2,262 < T \leq 2,262) = 95\%$ .

Distribuição  $t$  de Student com 9 graus de liberdade

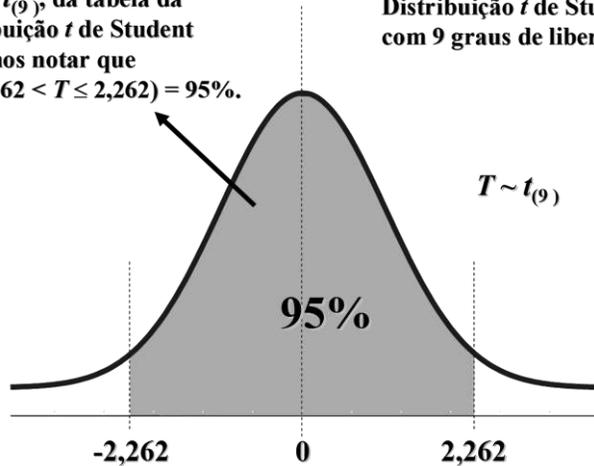
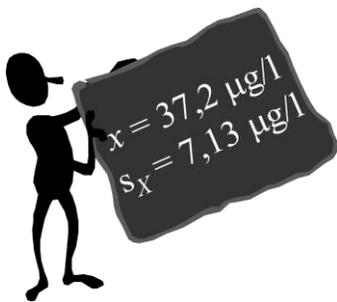


Tabela	g.l.	0,6	0,70	0,75	0,80	0,85	0,90	0,925	0,95	0,97	0,975
	1	1,376	1,963	2,414	3,078	4,165	6,314	8,449	12,706	21,205	25,452
	2	1,061	1,386	1,604	1,886	2,282	2,920	3,443	4,303	5,643	6,205
	3	0,978	1,250	1,423	1,638	1,924	2,353	2,681	3,182	3,896	4,177
	4	0,941	1,190	1,344	1,533	1,778	2,132	2,392	2,776	3,298	3,495
	5	0,920	1,156	1,301	1,476	1,699	2,015	2,242	2,571	3,003	3,163
	6	0,906	1,134	1,273	1,440	1,650	1,943	2,151	2,447	2,829	2,969
	7	0,896	1,119	1,254	1,415	1,617	1,895	2,090	2,365	2,715	2,841
	8	0,889	1,108	1,240	1,397	1,592	1,860	2,046	2,306	2,634	2,752
	9	0,883	1,100	1,230	1,383	1,574	1,833	2,013	2,262	2,574	2,685
	10	0,879	1,093	1,221	1,372	1,559	1,812	1,987	2,228	2,527	2,634
	11	0,876	1,088	1,214	1,363	1,548	1,796	1,966	2,201	2,491	2,593
	12	0,873	1,083	1,209	1,356	1,538	1,782	1,949	2,179	2,461	2,560

IC95% para  $\mu$  baseado em uma amostra tamanho  $n = 10$ :  $\left( \bar{X} - 2,262 \frac{s}{\sqrt{n}}; \bar{X} + 2,262 \frac{s}{\sqrt{n}} \right)$

Quando retirarmos uma amostra de tamanho  $n = 10$  da população em questão, temos uma chance de 95% de obtermos uma amostra que irá gerar um IC que “cobre” a média  $\mu$ .



**Agora** vamos usar as informações da nossa amostra.

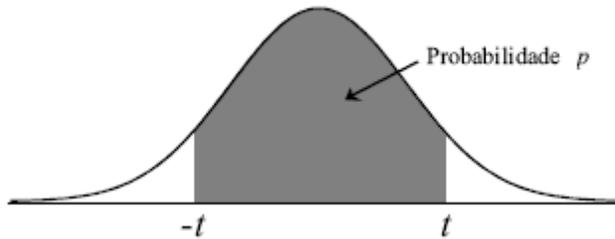
O nível médio de alumínio para a amostra de dez bebês é  $\bar{x} = 37,2 \mu\text{g/l}$  e seu desvio padrão amostral é  $s = 7,13 \mu\text{g/l}$ .

IC95% para  $\mu_x$  baseado em uma amostra tamanho  $n = 10$ :

$$\left( \bar{X} - 2,262 \frac{s}{\sqrt{n}}; \bar{X} + 2,262 \frac{s}{\sqrt{n}} \right)$$

$$\left( 37,2 - 2,262 \frac{7,13}{\sqrt{10}}; 37,2 + 2,262 \frac{7,13}{\sqrt{10}} \right) \text{ ou } (32,1; 42,3)$$

Assim, estamos 95% confiantes de que estes limites contêm o nível médio verdadeiro de alumínio no plasma para a população de bebês que recebe antiácidos.

Tabela. Distribuição  $t$  de Student

gl	Probabilidade $p$										
	0,60	0,70	0,75	0,80	0,85	0,90	0,925	0,95	0,97	0,975	0,99
1	1,376	1,963	2,414	3,078	4,165	6,314	8,449	12,706	21,205	25,452	63,656
2	1,061	1,386	1,604	1,886	2,282	2,920	3,443	4,303	5,643	6,205	9,925
3	0,978	1,250	1,423	1,638	1,924	2,353	2,681	3,182	3,896	4,177	5,841
4	0,941	1,190	1,344	1,533	1,778	2,132	2,392	2,776	3,298	3,495	4,604
5	0,920	1,156	1,301	1,476	1,699	2,015	2,242	2,571	3,003	3,163	4,032
6	0,906	1,134	1,273	1,440	1,650	1,943	2,151	2,447	2,829	2,969	3,707
7	0,896	1,119	1,254	1,415	1,617	1,895	2,090	2,365	2,715	2,841	3,499
8	0,889	1,108	1,240	1,397	1,592	1,860	2,046	2,306	2,634	2,752	3,355
9	0,883	1,100	1,230	1,383	1,574	1,833	2,013	2,262	2,574	2,685	3,250
10	0,879	1,093	1,221	1,372	1,559	1,812	1,987	2,228	2,527	2,634	3,169
11	0,876	1,088	1,214	1,363	1,548	1,796	1,966	2,201	2,491	2,593	3,106
12	0,873	1,083	1,209	1,356	1,538	1,782	1,949	2,179	2,461	2,560	3,055
13	0,870	1,079	1,204	1,350	1,530	1,771	1,935	2,160	2,436	2,533	3,012
14	0,868	1,076	1,200	1,345	1,523	1,761	1,923	2,145	2,415	2,510	2,977
15	0,866	1,074	1,197	1,341	1,517	1,753	1,913	2,131	2,397	2,490	2,947
16	0,865	1,071	1,194	1,337	1,512	1,746	1,904	2,120	2,382	2,473	2,921
17	0,863	1,069	1,191	1,333	1,508	1,740	1,897	2,110	2,368	2,458	2,898
18	0,862	1,067	1,189	1,330	1,504	1,734	1,890	2,101	2,356	2,445	2,878
19	0,861	1,066	1,187	1,328	1,500	1,729	1,884	2,093	2,346	2,433	2,861
20	0,860	1,064	1,185	1,325	1,497	1,725	1,878	2,086	2,336	2,423	2,845
21	0,859	1,063	1,183	1,323	1,494	1,721	1,873	2,080	2,328	2,414	2,831
22	0,858	1,061	1,182	1,321	1,492	1,717	1,869	2,074	2,320	2,405	2,819
23	0,858	1,060	1,180	1,319	1,489	1,714	1,865	2,069	2,313	2,398	2,807
24	0,857	1,059	1,179	1,318	1,487	1,711	1,861	2,064	2,307	2,391	2,797
25	0,856	1,058	1,178	1,316	1,485	1,708	1,858	2,060	2,301	2,385	2,787
26	0,856	1,058	1,177	1,315	1,483	1,706	1,855	2,056	2,296	2,379	2,779
27	0,855	1,057	1,176	1,314	1,482	1,703	1,852	2,052	2,291	2,373	2,771
28	0,855	1,056	1,175	1,313	1,480	1,701	1,849	2,048	2,286	2,368	2,763
29	0,854	1,055	1,174	1,311	1,479	1,699	1,847	2,045	2,282	2,364	2,756
30	0,854	1,055	1,173	1,310	1,477	1,697	1,845	2,042	2,278	2,360	2,750
31	0,853	1,054	1,172	1,309	1,476	1,696	1,842	2,040	2,275	2,356	2,744
32	0,853	1,054	1,172	1,309	1,475	1,694	1,840	2,037	2,271	2,352	2,738
33	0,853	1,053	1,171	1,308	1,474	1,692	1,839	2,035	2,268	2,348	2,733
34	0,852	1,052	1,170	1,307	1,473	1,691	1,837	2,032	2,265	2,345	2,728
35	0,852	1,052	1,170	1,306	1,472	1,690	1,835	2,030	2,262	2,342	2,724
36	0,852	1,052	1,169	1,306	1,471	1,688	1,834	2,028	2,260	2,339	2,719
37	0,851	1,051	1,169	1,305	1,470	1,687	1,832	2,026	2,257	2,336	2,715

gl	Probabilidade $p$										
	0,60	0,70	0,75	0,80	0,85	0,90	0,925	0,95	0,97	0,975	0,99
38	0,851	1,051	1,168	1,304	1,469	1,686	1,831	2,024	2,255	2,334	2,712
39	0,851	1,050	1,168	1,304	1,468	1,685	1,829	2,023	2,252	2,331	2,708
40	0,851	1,050	1,167	1,303	1,468	1,684	1,828	2,021	2,250	2,329	2,704
45	0,850	1,049	1,165	1,301	1,465	1,679	1,823	2,014	2,241	2,319	2,690
50	0,849	1,047	1,164	1,299	1,462	1,676	1,818	2,009	2,234	2,311	2,678
55	0,848	1,046	1,163	1,297	1,460	1,673	1,815	2,004	2,228	2,304	2,668
60	0,848	1,045	1,162	1,296	1,458	1,671	1,812	2,000	2,223	2,299	2,660
65	0,847	1,045	1,161	1,295	1,457	1,669	1,809	1,997	2,219	2,295	2,654
70	0,847	1,044	1,160	1,294	1,456	1,667	1,807	1,994	2,215	2,291	2,648
75	0,846	1,044	1,159	1,293	1,454	1,665	1,806	1,992	2,212	2,287	2,643
80	0,846	1,043	1,159	1,292	1,453	1,664	1,804	1,990	2,209	2,284	2,639
85	0,846	1,043	1,158	1,292	1,453	1,663	1,803	1,988	2,207	2,282	2,635
90	0,846	1,042	1,158	1,291	1,452	1,662	1,801	1,987	2,205	2,280	2,632
99	0,845	1,042	1,157	1,290	1,451	1,660	1,799	1,984	2,202	2,276	2,626
100	0,845	1,042	1,157	1,290	1,451	1,660	1,799	1,984	2,201	2,276	2,626
110	0,845	1,041	1,156	1,289	1,450	1,659	1,797	1,982	2,199	2,272	2,621
120	0,845	1,041	1,156	1,289	1,449	1,658	1,796	1,980	2,196	2,270	2,617
130	0,844	1,041	1,156	1,288	1,448	1,657	1,795	1,978	2,194	2,268	2,614
150	0,844	1,040	1,155	1,287	1,447	1,655	1,793	1,976	2,191	2,264	2,609
200	0,843	1,039	1,154	1,286	1,445	1,653	1,790	1,972	2,186	2,258	2,601
250	0,843	1,039	1,153	1,285	1,444	1,651	1,788	1,969	2,183	2,255	2,596
500	0,842	1,038	1,152	1,283	1,442	1,648	1,784	1,965	2,176	2,248	2,586
1000	0,842	1,037	1,151	1,282	1,441	1,646	1,782	1,962	2,173	2,245	2,581
5000	0,842	1,037	1,150	1,282	1,440	1,645	1,781	1,960	2,171	2,242	2,577

## Exercícios

1. Pesquisadores chineses publicaram um estudo no *International Journal of Hygiene and Environmental Health*, em 2004, objetivando determinar o nível médio de chumbo no sangue em crianças com idade entre 3 e 6 anos, frequentadoras das escolas infantis das comunidades rurais de Zhejiang, China. Os dados foram obtidos da análise do sangue de uma amostra de 217 crianças que frequentavam as escolas infantis selecionadas. Nesta amostra, foi observado um nível médio de chumbo de 95 microg/l, com um desvio padrão de 56 microg/l. Considere que os níveis de chumbo seguem uma distribuição aproximadamente normal na população.

- Encontre um intervalo de confiança 90% para  $\mu$  o nível médio de chumbo no sangue na população de crianças em questão.
- Encontre um intervalo de confiança 99% para  $\mu$  o nível médio de chumbo no sangue na população de crianças em questão.
- Interprete os intervalos de confiança encontrados nos itens (a) e (b) e explicitamente as diferenças que você observou entre estes intervalos.

2. Um estudo publicado em 2003 na revista *Environmental Health* alerta que a contaminação por chumbo em crianças não é incluída na lista das prioridades do sistema nacional de saúde pública do Líbano e indica a necessidade de um amplo programa de rastreamento deste problema neste país. Dentre outros objetivos, o estudo buscou determinar os níveis médios de chumbo no sangue em crianças com idade entre 1 e 3 anos. Os dados foram obtidos da análise do sangue de 281 crianças desta faixa etária usuárias do ambulatório de pediatria da *American University of Beirut Medical Center*, nos anos de 1997 e 1998. Nesta amostra, foi observado um nível médio de chumbo de 66 microg/l, com um desvio padrão de 26,3 microg/l. Considere que os níveis de chumbo seguem uma distribuição aproximadamente normal na população.
- Encontre um intervalo de confiança 95% para  $\mu$  o nível médio de chumbo no sangue na população de crianças em questão.
  - Encontre um intervalo de confiança 99% para  $\mu$  o nível médio de chumbo no sangue na população de crianças em questão.
  - Interprete os intervalos de confiança encontrados nos itens (a) e (b) e explicitamente as diferenças que você observou entre estes intervalos.
3. Foi verificado que a altura de 60 estudantes, em média, é igual a 175 cm, com um desvio padrão de 7 cm. Considere que estes 60 estudantes representam uma amostra aleatória de uma determinada população de estudantes. Construa um intervalo de confiança 95% para a altura média populacional.
4. Para estudar o efeito da merenda escolar, introduzida nas escolas de um grande município, foi acompanhada uma amostra de 100 crianças, que estão entrando na rede municipal de ensino. Dentre diversas características de interesse, pretende-se avaliar o parâmetro  $\mu$ , o ganho médio de peso dentre todas as crianças da rede municipal de ensino durante o primeiro ano letivo. Observando a amostra aleatória de  $n = 100$  crianças, encontraram-se as seguintes estatísticas relativas à variável ganho de peso ao longo do ano:
- ganho médio de peso das crianças da amostra:  $\bar{x} = 6$  kg,
  - desvio padrão dos pesos das crianças na amostra:  $s = 2$  kg.
- Encontre um IC 95% para o parâmetro  $\mu$ .
5. O Licopeno, um carotenóide da mesma família do betacaroteno, é o que dá aos tomates e a várias frutas a cor vermelha forte. Alguns pesquisadores acreditam que o Licopeno é capaz de combater doenças degenerativas. Em um estudo recente publicado na revista *Nutrition Research*, Lugasi e colaboradores estimaram o conteúdo de Licopeno em alguns alimentos. A tabela a seguir mostra alguns exemplos de alimentos analisados por estes pesquisadores, mostrando os seus conteúdos médios de Licopeno (em mg/100g) com os respectivos desvios padrão. Estas estatísticas descritivas mostradas na tabela foram calculadas utilizando amostras aleatórias independentes de cinco alimentos de cada tipo.

Tabela - Conteúdo de Licopeno (em mg/100g) encontrado em alguns alimentos

Alimento	Licopeno (mg/100g)	
	média	desvio padrão
Tomate fresco, vendido no inverno	0,85	0,05
Tomate fresco, vendido na primavera	1,10	0,07
Tomate fresco, vendido no verão	13,6	0,25
Pêssego	0,11	0,06
Melancia	4,77	1,18
Abóbora fresca	0,50	0,06

Suponha que o conteúdo de Licopeno encontrado em cada alimento tem distribuição populacional aproximadamente normal.

- Encontre um intervalo de confiança 95% para o conteúdo médio (populacional) de Licopeno encontrado nos pêssegos (desconsidere os demais alimentos).
- Encontre um intervalo de confiança 90% para o conteúdo médio (populacional) de Licopeno encontrado nos pêssegos (desconsidere os demais alimentos).
- Interprete os intervalos de confiança encontrados nos itens (a) e (b), justificando as diferenças observadas entre os intervalos.

6. Em uma amostra de 300 crianças com idade de 4 a 5 anos residentes em um município, a dosagem de hemoglobina em sangue periférico tem média 11,3 g/dl e desvio padrão 2,8 g/dl.

- Encontre um intervalo de confiança 95% para a média populacional desta variável.
- Escreva uma interpretação adequada para o resultado obtido no item (a).

### Respostas:

1) (a) (88,73 ; 101,27) (b) (85,12 ; 104,88)

2) (a) (62,91 ; 69,09) (b) (61,94 ; 70,06)

3) ( 173,1 ; 176,8 ). Com uma confiança de 95%, a altura média populacional é um valor entre 173,1 cm e 176,8 cm.

4) ( 5,6 ; 6,4 ). Com uma confiança de 95%, ganho médio de peso das crianças na população em questão é um valor entre 5,6 kg e 6,4 kg.

5) (a) Como foram utilizadas amostras de cinco alimentos de cada tipo, temos  $n = 5$ . O intervalo de confiança 95% é (0,04 ; 0,18). Com uma confiança de 95%, os pêssegos possuem um conteúdo médio (populacional) de Licopeno entre 0,04 e 0,18 mg/100g. (b) (0,05 ; 0,17). Com uma confiança de 90%, os pêssegos possuem um conteúdo médio (populacional) de Licopeno entre 0,05 e 0,17 mg/100g. (c) É conhecido que quanto maior o coeficiente de confiança, maior será a amplitude do intervalo; ao compararmos os intervalos obtidos nos itens (a) e (b), notamos que, para os dados apresentados, o intervalo de confiança 95% tem uma amplitude um pouco maior que o intervalo de confiança 90%.

6) (a) (10,98 ; 11,61)

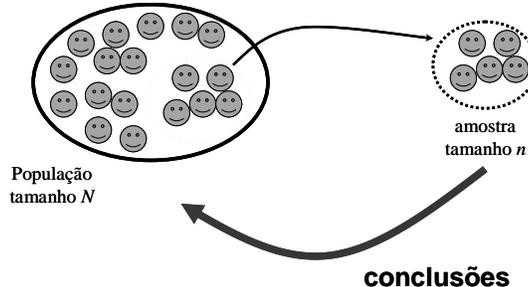
## Testes de Hipóteses



<https://www.youtube.com/watch?v=yXuEDTaw9FA>

**Já vimos que o objetivo da inferência estatística é:**

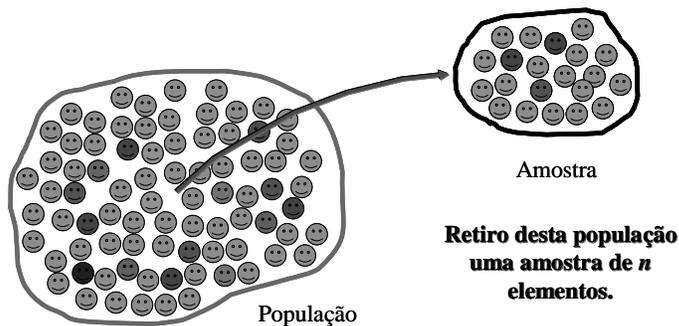
Obter conclusões sobre algumas características de um conjunto de interesse, denominado população, com base na informação oriunda de um conjunto de dados disponíveis, denominado amostra.



Tais conclusões são basicamente obtidas por duas formas:

1-) **Intervalos de confiança** – quando o objetivo é estimar um parâmetro, ou seja, uma característica numérica da população.

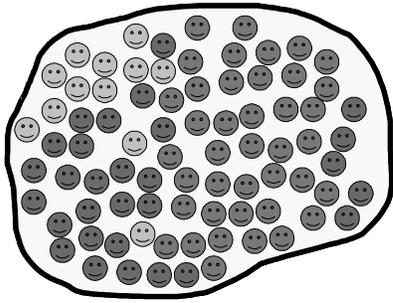
2-) **Testes de hipóteses** – quando há hipóteses sobre características numéricas da população.



Uma hipótese é uma suposição sobre um parâmetro populacional desconhecido.

### Exemplo:

- Em 1860, após analisar a temperatura da região axilar de aproximadamente 25 mil pessoas, Carl Wunderlich identificou a temperatura média de adultos saudáveis como 37,0° C ou 98,6° F.
- Determinou-se que 37,0° C ou 98,6° F seria uma “temperatura normal” para um indivíduo.
- Wunderlich também estabeleceu que uma temperatura superior a 38,0° C ou 100,4° F seria um “limite superior de normalidade” para a temperatura corporal, sendo que um indivíduo com temperatura maior que este limite seria classificado como portador de febre.



População geral dos adultos saudáveis.

média  $\mu = 37,0^{\circ}\text{C}$

Foi assim estabelecido que na população de adultos saudáveis, a temperatura média na região axilar é  $37,0^{\circ}\text{C}$ .

Em 1992, Mackowiak, Wasserman e Levine perguntaram...

**Será que a temperatura média de adultos saudáveis é mesmo  $37,0^{\circ}\text{C}$  ?**

*Referência: JAMA, 268(12):1578-80,1992*

**Pergunta:**  $\mu \neq 37,0^{\circ}\text{C} ???$

**Hipótese do pesquisador:** média  $\mu \neq 37,0^{\circ}\text{C}$

Um teste estatístico de hipóteses é uma **regra** utilizada para decidir quando rejeitar uma hipótese. Esta regra é sempre baseada em uma amostra.

Com base nos resultados de uma amostra aleatória de tamanho  $n$ , tomamos a decisão de rejeitar ou não rejeitar uma hipótese formulada sobre um parâmetro de interesse.

Na prática, consideramos duas hipóteses:

**Hipótese alternativa ( $H_A$ ):**

é a “hipótese do pesquisador”, aquilo que ele deseja verificar.

*Na população de adultos saudáveis, a temperatura média na região axilar é diferente de  $37,0^{\circ}\text{C}$*

**Hipótese nula ( $H_0$ ):** é o complemento da alternativa.

*Na população de adultos saudáveis, a temperatura média na região axilar é  $37,0^{\circ}\text{C}$*

**Hipótese alternativa ( $H_A$ ):**  $H_A: \mu \neq 37,0^{\circ}\text{C}$

**Hipótese nula ( $H_0$ ):**  $H_0: \mu = 37,0^{\circ}\text{C}$

**Agora uma definição mais completa...**

Com base nos resultados de uma amostra aleatória de tamanho  $n$ , tomamos a decisão de rejeitar ou não rejeitar a hipótese nula.





Estaremos assim sujeitos a dois tipos de erros...

A-) ERRO TIPO I : rejeito  $H_0$ , *mas  $H_0$  é verdadeira*

B-) ERRO TIPO II : não rejeito  $H_0$ , *mas  $H_0$  é falsa*

Decisão	$H_0$ é verdadeira	$H_0$ é falsa
Rejeito $H_0$	Erro tipo I	Sem erro
Não rejeito $H_0$	Sem erro	Erro tipo II

### Exemplo:

Certa droga (“droga 1”) vem sendo utilizada no tratamento de uma moléstia. Um pesquisador desenvolve uma nova droga (“droga 2”), que, se mais eficiente, substituirá a droga 1.

### Hipótese do pesquisador:

A droga 2 é mais eficiente.

### Hipóteses:

$H_0$ : A droga 1 é mais eficiente, \_\_\_\_\_

$H_A$ : A droga 2 é mais eficiente.



Qual é o erro mais grave ?

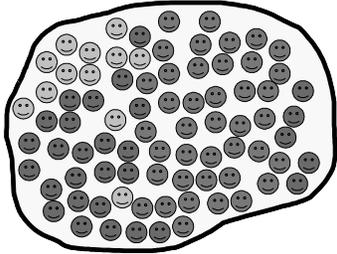
- A probabilidade de se cometer um erro tipo I é chamada *nível de significância do teste* e é denotada por  $\alpha$  .
- A probabilidade de se cometer um erro tipo II é denotada por  $\beta$  .
- Na área da saúde, a quantidade  $1 - \beta$  é geralmente chamada *poder* (ou *potência*) do teste.
- O *nível de significância do teste* ( $\alpha$ ) é fixado **antes** da coleta dos dados. Na área da saúde, é muito comum fixar  $\alpha = 5\%$ .
- A probabilidade de se cometer um erro tipo II ( $\beta$ ) é geralmente usada para o cálculo do tamanho amostral. Escolhas comuns:  $\beta = 5\%$ ,  $10\%$  ou  $20\%$ .

## Testes de hipóteses para uma média populacional



<https://www.youtube.com/watch?v=WfwNaExXvgw>

## Exemplo:



População geral dos adultos saudáveis.

média  $\mu = 37,0^{\circ}\text{C}$

Acredita-se que na população de adultos saudáveis, a temperatura média na região axilar é  $37,0^{\circ}\text{C}$ .

## Pergunta:

Será que a temperatura média de adultos saudáveis é mesmo  $37,0^{\circ}\text{C}$  ?

$\mu \neq 37,0^{\circ}\text{C} ???$

**Hipótese do pesquisador:** média  $\mu \neq 37,0^{\circ}\text{C}$

**Hipótese nula ( $H_0$ ):**  $H_0: \mu = 37,0^{\circ}\text{C}$

**Hipótese alternativa ( $H_A$ ):**  $H_A: \mu \neq 37,0^{\circ}\text{C}$

Antes de retirarmos a amostra da população, vamos fixar o nível de significância do teste.

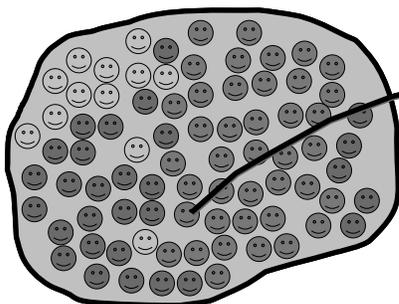
A maioria dos estudos da área da saúde adota  $\alpha = 0,05$ .

Se, na população, a temperatura média na região axilar segue uma distribuição normal

com média  $\mu$  e variância  $\sigma^2$ , ou seja,  $X \sim N(\mu; \sigma^2)$ , sabemos que  $T = \frac{(\bar{X} - \mu)\sqrt{n}}{s}$  segue

uma distribuição  $t$  de Student com  $n - 1$  graus de liberdade.

Se a hipótese nula for verdadeira,  $T_0 = \frac{(\bar{X} - 37)\sqrt{n}}{s}$  segue uma distribuição  $t$  de Student com  $n - 1$  graus de liberdade.



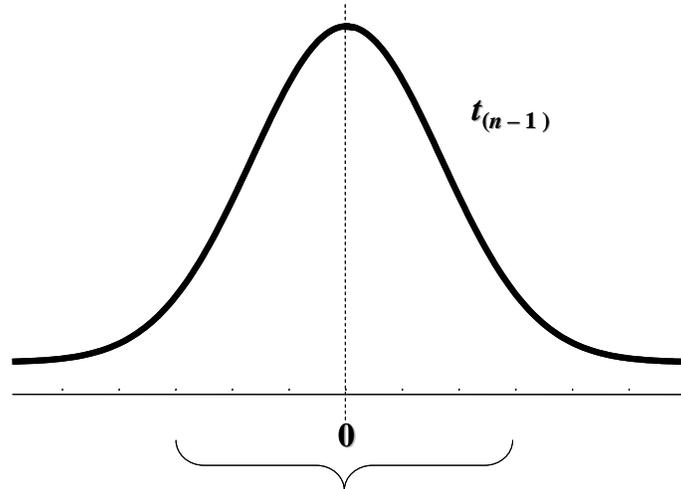
Amostra de tamanho  $n$

A partir de uma amostra tamanho  $n$ , vamos calcular

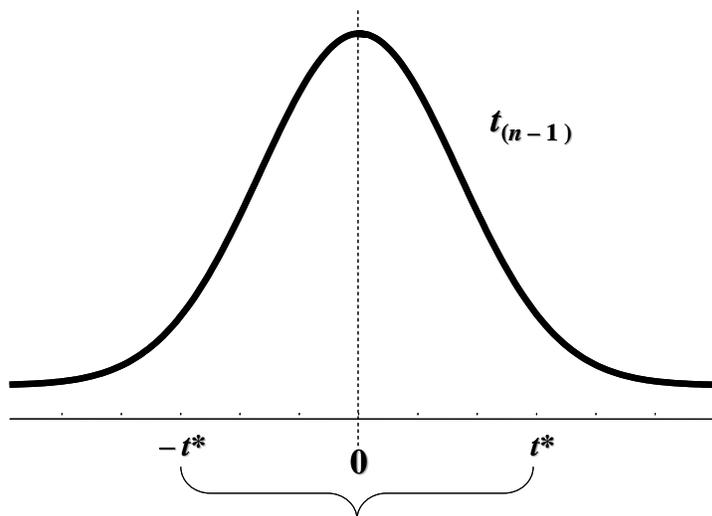
$$t_0 = \frac{(\bar{x} - 37)\sqrt{n}}{s}$$

Se a hipótese nula for verdadeira, este valor de  $t_0$  calculado da amostra será o resultado de uma distribuição  $t$  de Student com  $n - 1$  gl.

População geral dos adultos saudáveis.



**Se a hipótese nula é verdadeira, é mais provável que encontremos em uma amostra tamanho  $n$  valores de  $t_0$  situados nesta região “central” da curva, onde há maior densidade.**



Assim, vamos rejeitar  $H_0$  se encontrarmos em uma amostra tamanho  $n$  um valor de  $t_0$  maior que  $t^*$  ou menor que  $-t^*$ , de forma que  $t^*$  é determinado da seguinte maneira...

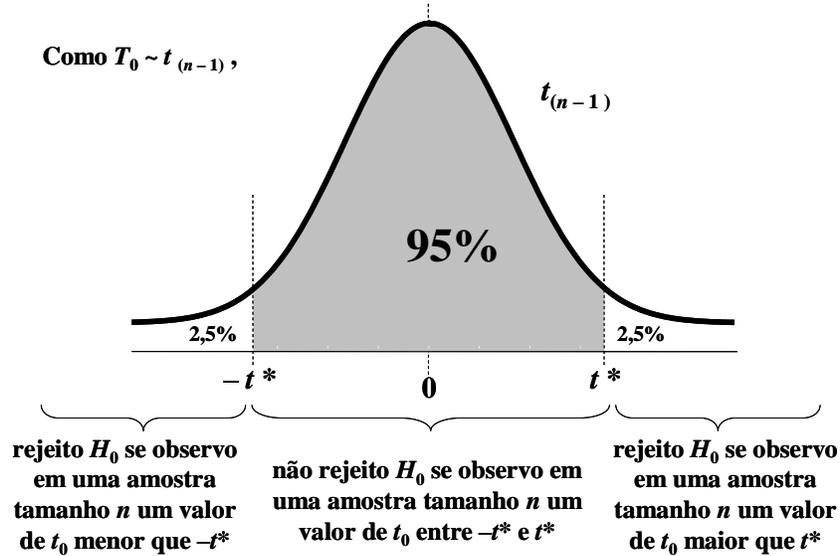
... se fixarmos o nível de significância em 5%, então

Nível de significância = P(erro tipo I) = P(rejeitar  $H_0$  |  $H_0$  é verdadeira) = 5%.

Decidimos rejeitar  $H_0$  nas situações onde  $T_0 > t^*$  e  $T_0 < -t^*$ .

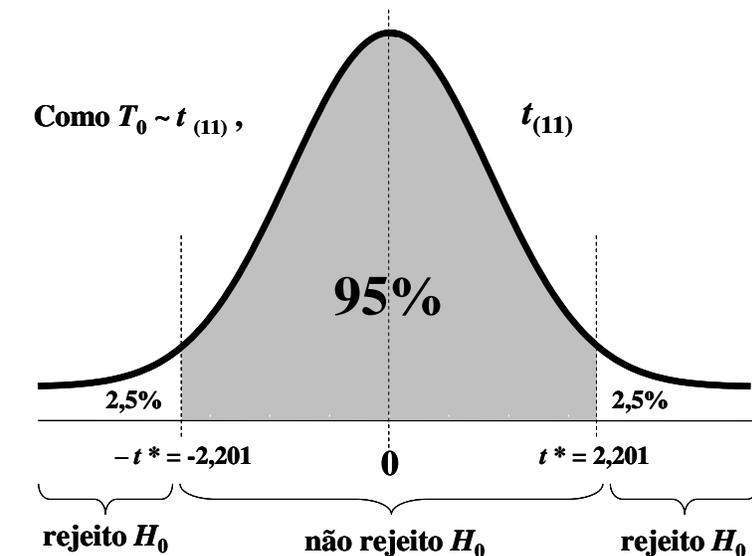
Assim, P(não rejeitar  $H_0$  |  $H_0$  é verdadeira) = 95%

$P(-t^* < T_0 < t^*) = 95\%$



$-t^*$  e  $t^*$  são chamados valores críticos de  $T_0$ .

Vamos retirar da população uma amostra tamanho  $n = 12$ .



Da tabela da distribuição  $t$  de Student, temos que  $t^* = 2,201$ .

Na amostra tamanho  $n = 12$ , observamos as seguintes temperaturas (em  $^{\circ}\text{C}$ ):

35,6	35,8	36,3	35,5	35,9	36,9
37,1	36,2	37,2	36,1	37,6	37,7

Temos então uma média amostral  $\bar{x} = 36,5^{\circ}\text{C}$  e um desvio padrão amostral  $s = 0,775^{\circ}\text{C}$ .

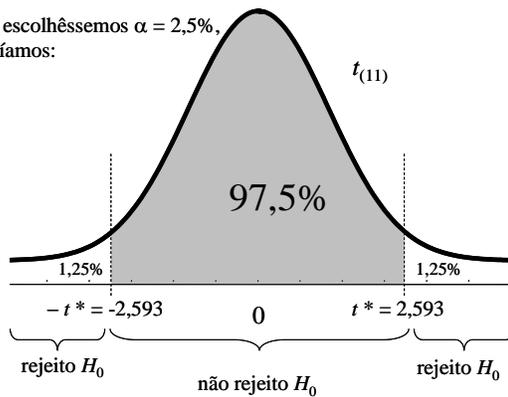
$$\text{Assim, } t_0 = \frac{(\bar{x} - 37)\sqrt{n}}{s} = \frac{(36,5 - 37)\sqrt{12}}{0,775} = -2,23.$$

Como  $t_0 < -2,201$ , rejeitamos  $H_0$ ; ou seja, temos evidências de que a temperatura média na região axilar na população de adultos saudáveis é diferente de  $37,0^\circ\text{C}$ .

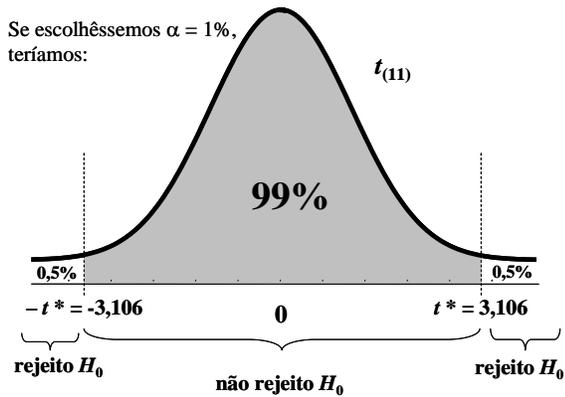
Esta conclusão é válida para um nível de significância  $\alpha = 5\%$ .

Uma nota...

Se escolhêssemos  $\alpha = 2,5\%$ ,  
teríamos:



Se escolhêssemos  $\alpha = 1\%$ ,  
teríamos:



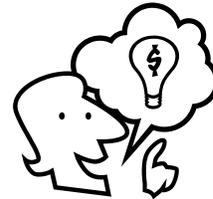
Como  $t_0 = -2,23$ , nós não rejeitaríamos  $H_0$   
para  $\alpha = 2,5\%$ .

Como  $t_0 = -2,23$ , também não rejeitaríamos  
 $H_0$  para  $\alpha = 1\%$ .

Temos  $t_0 = -2,23$ .

Escolha de $\alpha$	$-t^*$	$t^*$	Decisão
0,05	-2,2010	2,2010	Rejeito $H_0$
0,048	-2,2243	2,2243	Rejeito $H_0$
0,047	-2,2364	2,2364	Não rejeito $H_0$
0,04	-2,3281	2,3281	Não rejeito $H_0$
0,03	-2,4907	2,4907	Não rejeito $H_0$
0,025	-2,5931	2,5931	Não rejeito $H_0$
0,02	-2,7181	2,7181	Não rejeito $H_0$
0,01	-3,1058	3,1058	Não rejeito $H_0$

Nota-se que rejeitaríamos  $H_0$  se  
escolhêssemos um valor maior ou  
igual a 0,048 para  $\alpha$ .



A menor possível escolha de  $\alpha$  que nos levaria a rejeitar a hipótese nula é chamada “valor  $p$ ”.

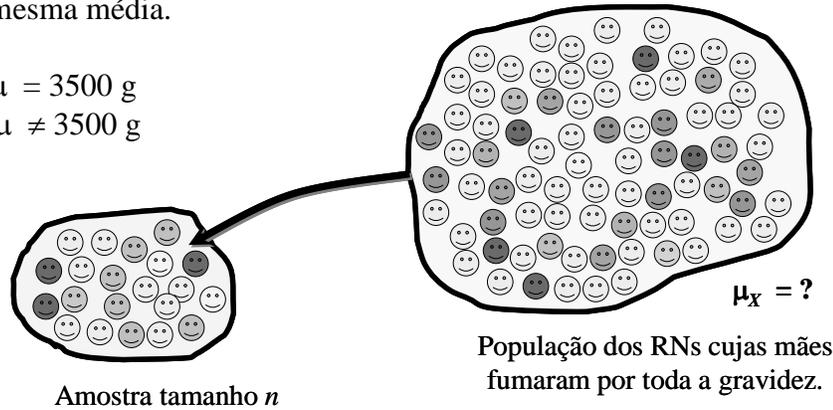
**Exemplo:** na Noruega, a distribuição dos pesos ao nascer, cuja idade gestacional é de 40 semanas, é aproximadamente normal, com média  $\mu_0 = 3500$  gramas. Um pesquisador planeja conduzir um estudo para determinar se o peso ao nascer, cujas mães fumaram por toda a gravidez, têm a mesma média.

**Hipóteses:**  $H_0: \mu = 3500$  g  
 $H_A: \mu \neq 3500$  g

**Regra de decisão:**

Rejeito  $H_0$  quando

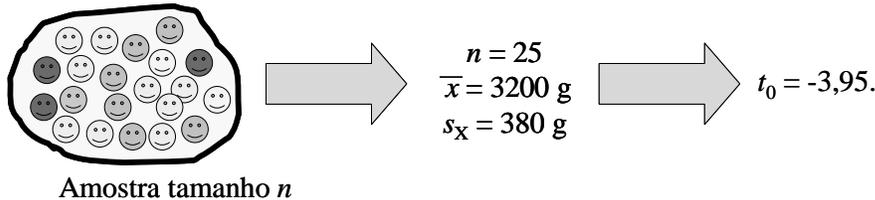
$$T_0 = \frac{(\bar{X} - 3500)\sqrt{n}}{s}$$



maior que  $t^*$  ou maior que  $-t^*$ .

Da amostra tamanho  $n = 25$ , temos  $\bar{x} = 3200$  g e  $s = 380$  g.

Como  $n = 25$ , rejeito  $H_0$  se  $t_0 > 2,064$  ou se  $t_0 < -2,064$ .



## Exercícios

1. Pesquisadores chineses publicaram um estudo no *International Journal of Hygiene and Environmental Health*, em 2004, objetivando determinar o nível médio de chumbo no sangue em crianças com idade entre 3 e 6 anos, freqüentadoras das escolas infantis das comunidades rurais de Zhejiang, China. Os dados foram obtidos da análise do sangue de uma amostra de 217 crianças que freqüentavam as escolas infantis selecionadas. Nesta amostra, foi observado um nível médio de chumbo de 95 microg/l, com um desvio padrão de 56 microg/l. Considere que os níveis de chumbo possuem distribuição aproximadamente normal na população. Sendo  $\mu$  o nível médio de chumbo no sangue na população de crianças em questão, testar a hipótese nula  $H_0: \mu = 80$  microg/l contra  $H_A: \mu \neq 80$  microg/l, considerando uma chance de erro tipo I de 5% (interpretar o resultado).

2. (Shott, 1990) Em um estudo sobre doença renal em estágio final, valores de creatinina sérica (em mg/dl) foram obtidos em uma amostra de 12 pacientes submetidos à hemodiálise, conforme tabela abaixo.

6,7	8,0	13,4	12,4	14,9	6,3
16,5	13,5	12,4	16,9	9,1	13,0

Um nefrologista acredita que o nível médio populacional da creatinina é 8,4 mg/dl. Os dados trazem evidências contra a hipótese de que esta média populacional é igual a 8,4 mg/dl ?

3. Um estudo publicado em 2003 na revista *Environmental Health* alerta que a contaminação por chumbo em crianças não é incluída na lista das prioridades do sistema nacional de saúde pública do Líbano e indica a necessidade de um amplo programa de rastreamento deste problema neste país. Dentre outros objetivos, o estudo buscou determinar os níveis médios de chumbo no sangue em crianças com idade entre 1 e 3 anos. Os dados foram obtidos da análise do sangue de 281 crianças desta faixa etária usuárias do ambulatório de pediatria da *American University of Beirut Medical Center*, nos anos de 1997 e 1998. Nesta amostra, foi observado um nível médio de chumbo de 66 microg/l, com um desvio padrão de 26,3 microg/l. Considere que os níveis de chumbo possuem distribuição aproximadamente normal na população.

- (a) Sendo  $\mu$  o nível médio de chumbo no sangue na população de crianças em questão, testar a hipótese nula  $H_0: \mu = 60$  microg/l contra  $H_A: \mu \neq 60$  microg/l considerando uma chance de erro tipo I de 5%. Interprete o resultado.
- (b) Sendo  $\mu$  o nível médio de chumbo no sangue na população de crianças em questão, testar a hipótese nula  $H_0: \mu = 60$  microg/l contra  $H_A: \mu \neq 60$  microg/l considerando um nível de significância de 1%. Interprete o resultado.
- (c) Escreva o que seria um erro tipo I e um erro tipo II considerando os testes apresentados nos itens anteriores.
4. Um fabricante de sabão líquido utilizado para fins hospitalares disse que, em média, o conteúdo de cada embalagem comercializada é de 10 litros. O responsável pela Seção de Compras de um hospital selecionou aleatoriamente 7 embalagens de um lote, encontrando os seguintes conteúdos:
- 10,2      9,8      10,4      9,8      10,0      10,2      e      9,6 litros.
- Sendo  $\mu$  o conteúdo médio de sabão de todo o lote (a população), testar a hipótese nula  $H_0: \mu = 10$  litros contra  $H_A: \mu \neq 10$  litros, considerando uma chance de erro tipo I de 5%. De acordo com o resultado do teste, o fabricante está dizendo a verdade quando afirma que o conteúdo de cada embalagem comercializada é de 10 litros, em média ?



<https://www.youtube.com/watch?v=aHYWOHxTsm0>

5. Foi medida a pressão diastólica de 30 crianças com idade entre 5 e 6 anos, escolhidas aleatoriamente em uma determinada comunidade. A média da pressão diastólica obtida com base nesta amostra foi de 56 mmHg, com desvio padrão de 17,9 mmHg. Sabendo que os valores obtidos a nível nacional dizem que a média da pressão diastólica é de 64 mmHg para as crianças com idade entre 5 e 6 anos, diga se há alguma evidência de que a pressão diastólica das crianças da comunidade em estudo seja diferente da média nacional, considerando um nível de significância de 5%.
6. Para estudar o efeito da merenda escolar, introduzida nas escolas de um grande município, foi acompanhada uma amostra de 100 crianças, que estão entrando na rede municipal de ensino. Dentre diversas características de interesse, pretende-se avaliar o parâmetro  $\mu$ , o ganho médio de peso dentre todas as crianças da rede municipal de ensino durante o primeiro ano letivo. Observando a amostra aleatória de  $n = 100$  crianças, encontraram-se as seguintes estatísticas relativas à variável ganho de peso ao longo do ano:
- ganho médio de peso das crianças da amostra:  $\bar{x} = 6$  kg,
  - desvio padrão dos pesos das crianças na amostra:  $s = 2$  kg.
- Testar a hipótese  $H_0: \mu = 7$  kg contra  $H_A: \mu \neq 7$  kg.
7. Foi verificado que a altura de 60 estudantes, em média, é igual a 175 cm, com um desvio padrão de 7 cm. Considere que estes 60 estudantes representam uma amostra aleatória de uma determinada população de estudantes. Sendo  $\mu$  a altura média populacional dos estudantes, testar a hipótese nula  $H_0: \mu = 170$  cm contra  $H_A: \mu \neq 170$  cm, considerando uma chance de erro tipo I de 5%. Interprete o resultado.

**Respostas (as interpretações dos resultados ficam por sua conta!):**

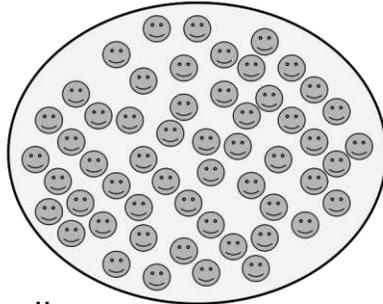
1. Temos  $n = 217$ ,  $t_0 = 3,95$  (interpretar!)
2. Temos  $n = 12$ ,  $\bar{x} = 11,9$  mg/dl,  $s = 3,61$  mg/dl e  $t_0 = 3,36$  (interpretar!)
3. (a) Temos  $t_0 = 3,8$  (interpretar!); (b) Rejeito  $H_0$  (interpretar!)
4. Temos  $t_0 = 0$  (interpretar!)
5.  $n = 30$ ,  $H_0: \mu = 64$  mmHg,  $H_A: \mu \neq 64$  mmHg,  $t_0 = -2,45$  e  $t^* = 2,04$  (interpretar!)
6. Temos  $t_0 = -5$  (interpretar!)
7. Temos  $t_0 = 5,5$  (interpretar!)

Testes de Hipóteses baseados na distribuição $t$ : comparações entre duas médias
--



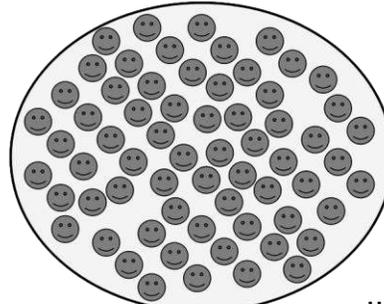
[https://www.youtube.com/watch?v=ZsAd2QND\\_wc](https://www.youtube.com/watch?v=ZsAd2QND_wc)

### População 1



$\mu_1$

### População 2



$\mu_2$

Sejam duas populações distintas, a população 1 e a população 2.

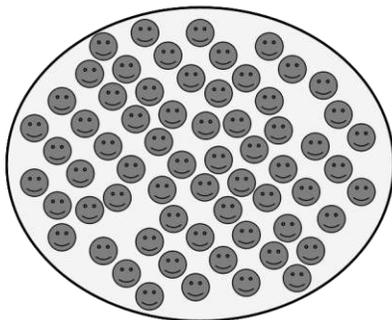
Seja  $X$  uma variável aleatória (contínua) de interesse. Esta variável tem média  $\mu_1$  na população 1 e  $\mu_2$  na população 2. O nosso objetivo é comparar as médias populacionais  $\mu_1$  e  $\mu_2$ . Assim, as hipóteses nula e alternativa são:

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$

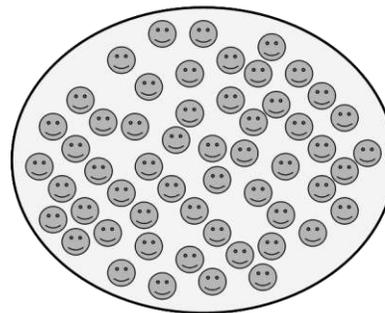
**Exemplo:** um estudo foi conduzido para determinar se a fumaça do cigarro de uma gestante tem algum efeito no conteúdo mineral ósseo da criança por ela gerada, sob outros aspectos saudáveis.

### População de RNs cujas mães fumaram durante a gravidez



$\mu_1 = ?$   
 $\sigma_1 = ?$

### População de RNs cujas mães não fumaram durante a gravidez



$\mu_2 = ?$   
 $\sigma_2 = ?$

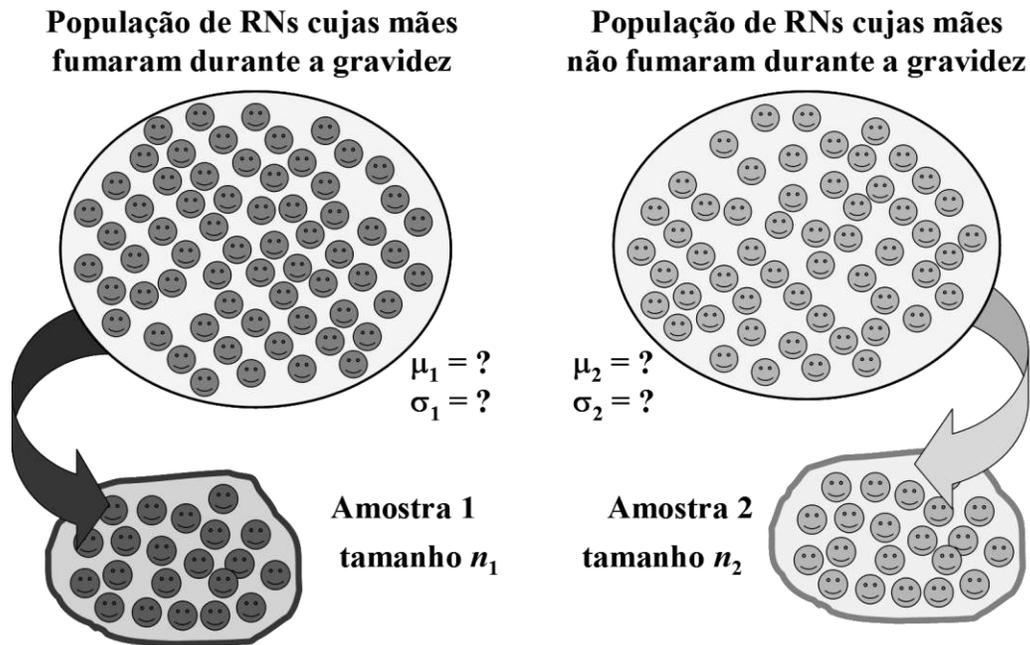
**X: conteúdo mineral ósseo**

Desejamos testar

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$

Inicialmente, vamos supor que os desvios padrão populacionais  $\sigma_1$  e  $\sigma_2$  são iguais. Assim,  $\sigma = \sigma_1 = \sigma_2$ .



Da população 1, retiramos uma amostra tamanho  $n_1$ , e da população 2, retiramos uma amostra tamanho  $n_2$ .

Observar que:  $H_0: \mu_1 = \mu_2$  equivale a  $H_0: \mu_1 - \mu_2 = 0$   
 $H_A: \mu_1 \neq \mu_2$   $H_A: \mu_1 - \mu_2 \neq 0$

Testamos  $H_0$  usando a quantidade

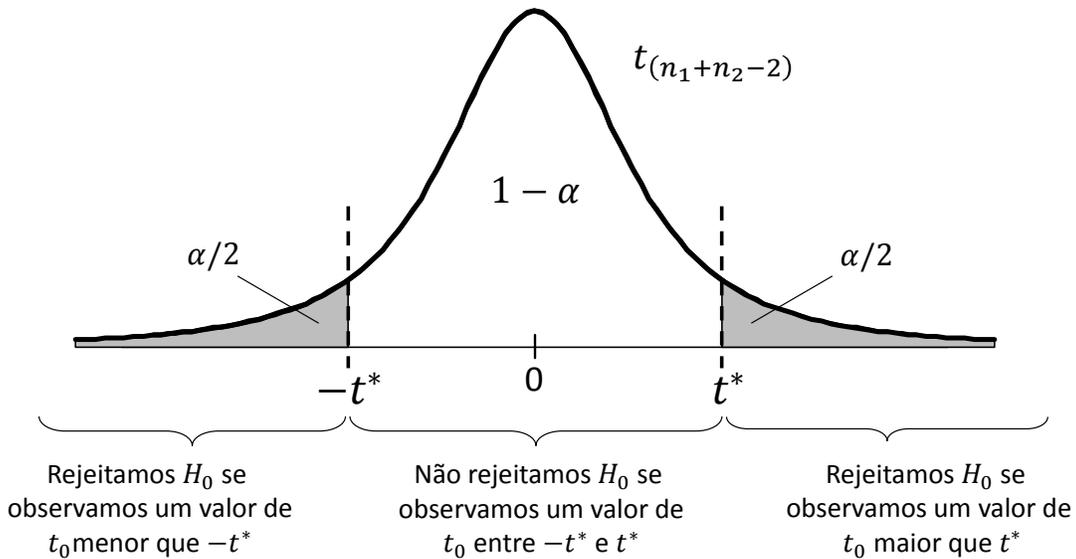
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}, \text{ em que } S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \text{ é a estimativa ponderada da}$$

variância populacional  $\sigma$ , comum às duas populações.

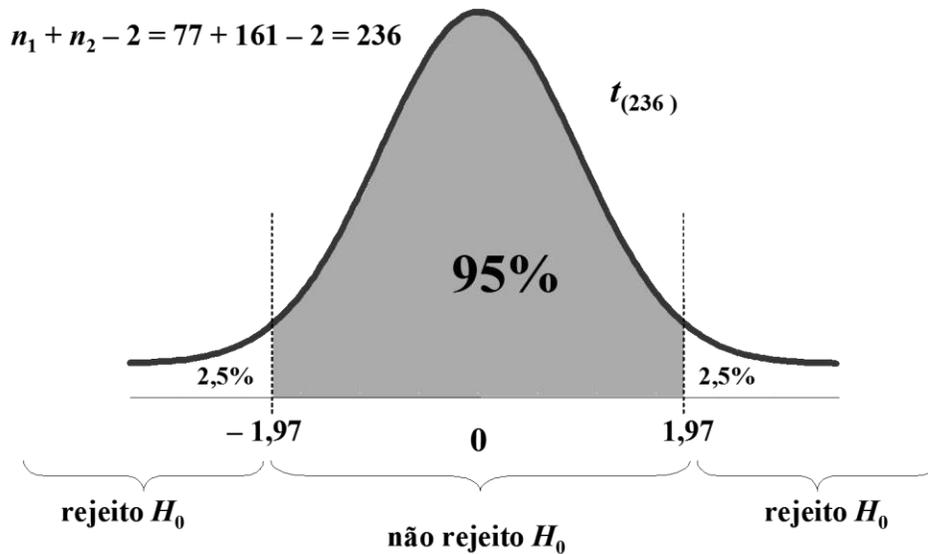
Se a hipótese nula  $H_0$  é verdadeira,  $t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$  segue uma distribuição  $t$  de

Student com  $n_1 + n_2 - 2$  graus de liberdade.

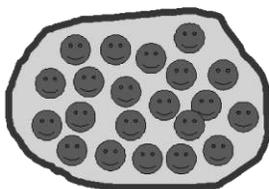
Seja um nível de significância  $\alpha$ .



Da tabela da distribuição  $t$  de Student, se  $\alpha = 0,05$ :



Sejam os dados amostrais:



Amostra 1

$$n_1 = 77$$

$$\bar{x}_1 = 0,098 \text{ g/cm}$$

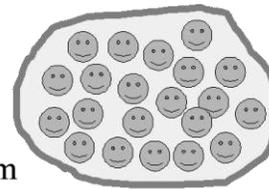
$$s_1 = 0,026 \text{ g/cm}$$

Amostra 2

$$n_2 = 161$$

$$\bar{x}_2 = 0,095 \text{ g/cm}$$

$$s_2 = 0,025 \text{ g/cm}$$



$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(77 - 1)0,026^2 + (161 - 1)0,025^2}{77 + 161 - 2} = 0,00064.$$

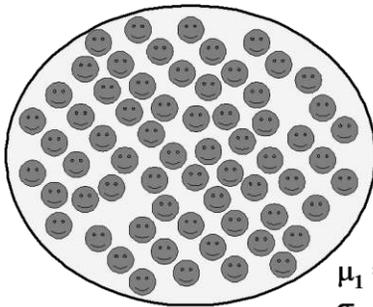
Notar que  $S_p = \sqrt{S_p^2} = \sqrt{0,00064} = 0,0253$  g/cm (aproximadamente)

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{0,098 - 0,095}{\sqrt{0,00064 \left( \frac{1}{77} + \frac{1}{161} \right)}} = 0,86.$$

Não rejeitamos  $H_0$ , ou seja, o estudo não fornece evidências de que fumar durante a gestação exerça algum efeito sobre o conteúdo mineral ósseo da criança por ela gerada (considerando  $\alpha = 5\%$ ). Valor  $p = 0,39$ .

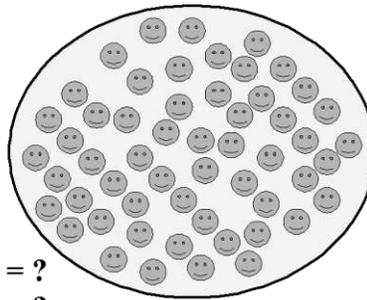
**Exemplo:** um estudo foi conduzido para investigar se, em média, o nível sérico de ferro é diferente entre crianças que sofrem de fibrose cística e crianças saudáveis.

**População de crianças que sofrem de fibrose cística**



$\mu_1 = ?$   
 $\sigma_1 = ?$

**População de crianças saudáveis**



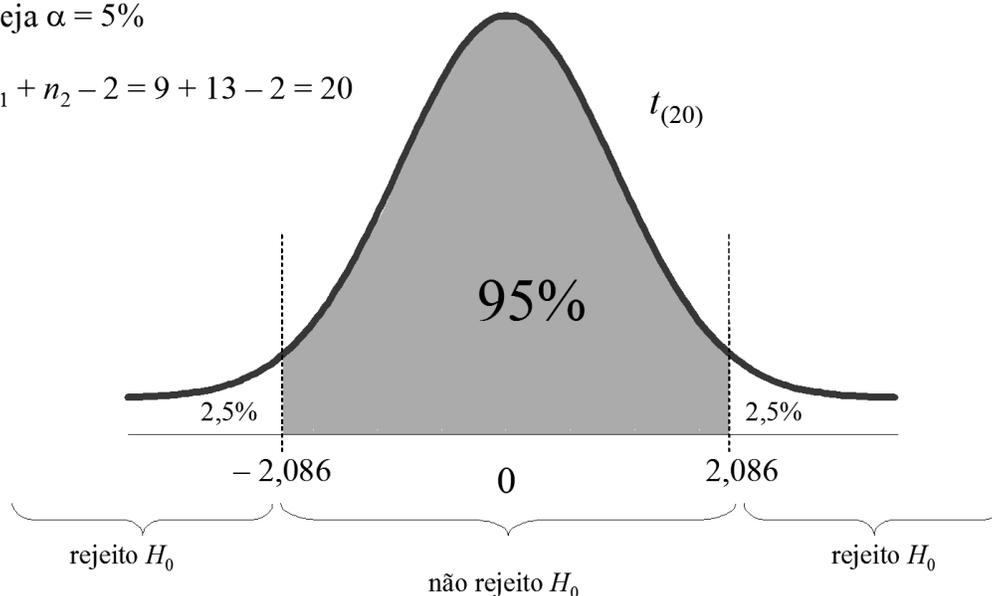
$\mu_2 = ?$   
 $\sigma_2 = ?$

$H_0: \mu_1 - \mu_2 = 0$

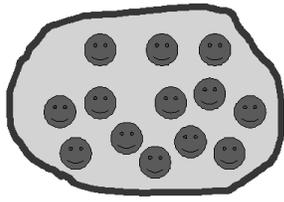
$H_A: \mu_1 - \mu_2 \neq 0$

Seja  $\alpha = 5\%$

$$n_1 + n_2 - 2 = 9 + 13 - 2 = 20$$

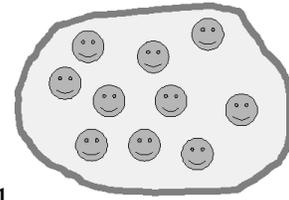


Suponha as amostras:



Amostra 1  
 $n_1 = 13$   
 $\bar{x}_1 = 11,9 \mu\text{mol/l}$   
 $s_1 = 6,3 \mu\text{mol/l}$

Amostra 2  
 $n_2 = 9$   
 $\bar{x}_2 = 18,9 \mu\text{mol/l}$   
 $s_2 = 5,9 \mu\text{mol/l}$



$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(13 - 1)6,3^2 + (9 - 1)5,9^2}{13 + 9 - 2} = 37,74 (\mu\text{mol/l})^2.$$

Observar que  $S_p = \sqrt{S_p^2} = \sqrt{37,74} = 6,143 \mu\text{mol/l}$  (aproximadamente)

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{11,9 - 18,9}{\sqrt{37,74 \left( \frac{1}{13} + \frac{1}{9} \right)}} = -2,64.$$

## Comparação entre duas médias, amostras independentes

### 1.) Variâncias iguais

$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$  segue uma distribuição  $t$  de Student com  $n_1 + n_2 - 2$  graus de

liberdade, em que  $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ .

### 2.) Variâncias desiguais

$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$  segue uma distribuição  $t$  de Student com  $\nu$  graus de liberdade,

em que  $\nu = \frac{\left( \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{\left( \frac{S_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left( \frac{S_2^2}{n_2} \right)^2}{n_2 - 1}}$ .

## Intervalo de confiança para $\mu_1 - \mu_2$

### 1.) Variâncias iguais

$$\left( \bar{x}_1 - \bar{x}_2 - t_{(n_1+n_2-2);\gamma} \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} ; \bar{x}_1 - \bar{x}_2 + t_{(n_1+n_2-2);\gamma} \sqrt{s_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \right)$$

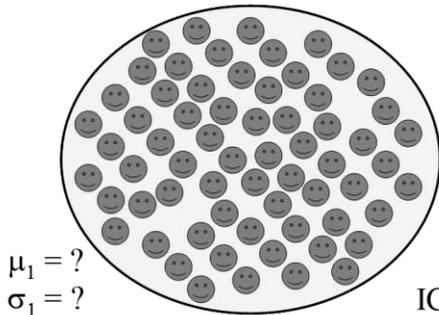
### 2.) Variâncias desiguais

$$\left( \bar{x}_1 - \bar{x}_2 - t_{(v);\gamma} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} ; \bar{x}_1 - \bar{x}_2 + t_{(v);\gamma} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

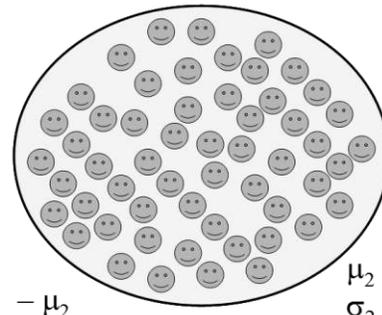
**Exemplo:** um estudo foi conduzido para determinar se a fumaça do cigarro de uma gestante tem algum efeito no conteúdo mineral ósseo da criança por ela gerada, sob outros aspectos saudáveis.

População de RNs cujas mães fumaram durante a gravidez

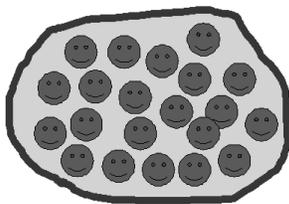
População de RNs cujas mães não fumaram durante a gravidez



IC para  $\mu_1 - \mu_2$



Supondo variâncias iguais:



Amostra 1

$$n_1 = 77$$

$$\bar{x}_1 = 0,098 \text{ g/cm}$$

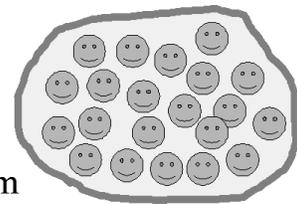
$$s_1 = 0,026 \text{ g/cm}$$

Amostra 2

$$n_2 = 161$$

$$\bar{x}_2 = 0,095 \text{ g/cm}$$

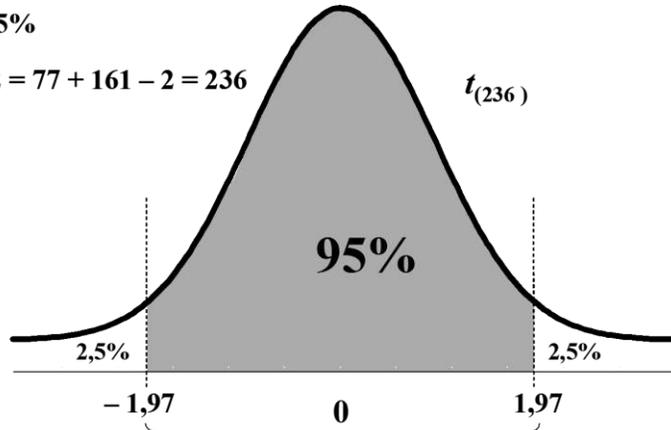
$$s_2 = 0,025 \text{ g/cm}$$



$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(77 - 1)0,026^2 + (161 - 1)0,025^2}{77 + 161 - 2} = 0,00064.$$

Seja  $\gamma = 95\%$

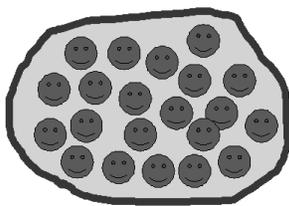
$$n_1 + n_2 - 2 = 77 + 161 - 2 = 236$$



Os valores  $-1,97$  e  $1,97$  delimitam uma área de tamanho 95% na região “central” da curva

$$\left( \bar{x}_1 - \bar{x}_2 - t_{(n_1+n_2-2;\gamma)} \sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} ; \bar{x}_1 - \bar{x}_2 + t_{(n_1+n_2-2;\gamma)} \sqrt{S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \right)$$

$$t_{(n_1+n_2-2;\gamma)} = 1,97$$

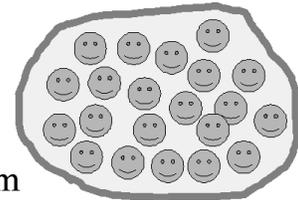


Amostra 1

$$\begin{aligned} n_1 &= 77 \\ \bar{x}_1 &= 0,098 \text{ g/cm} \\ s_1 &= 0,026 \text{ g/cm} \end{aligned}$$

Amostra 2

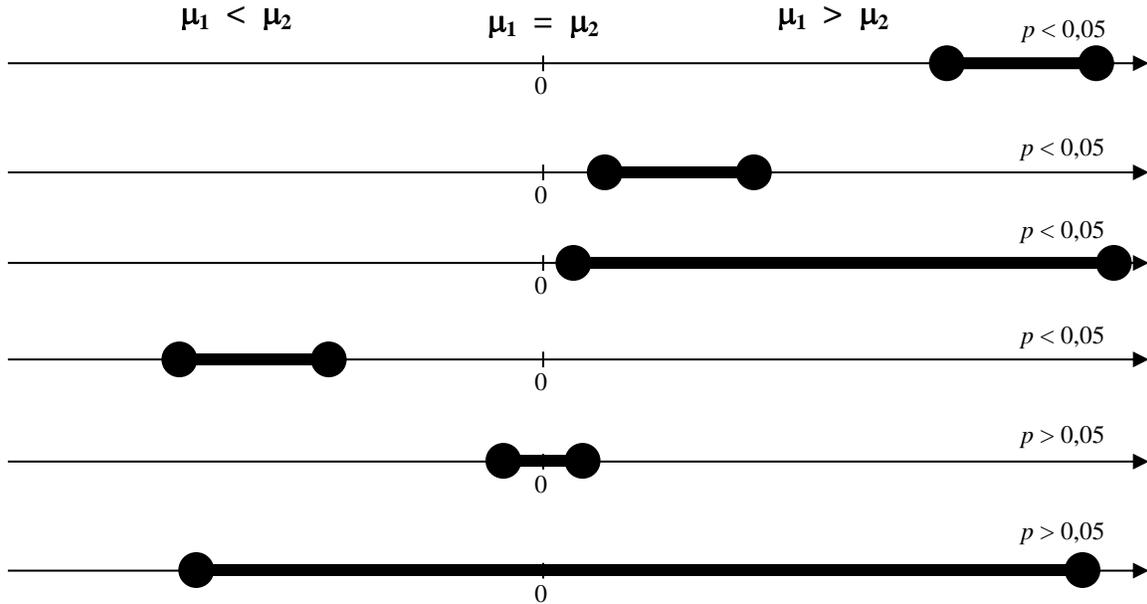
$$\begin{aligned} n_2 &= 161 \\ \bar{x}_2 &= 0,095 \text{ g/cm} \\ s_2 &= 0,025 \text{ g/cm} \end{aligned}$$



$$\left( 0,098 - 0,095 - 1,97 \sqrt{0,00064 \left( \frac{1}{77} + \frac{1}{161} \right)} ; 0,098 - 0,095 + 1,97 \sqrt{0,00064 \left( \frac{1}{77} + \frac{1}{161} \right)} \right)$$

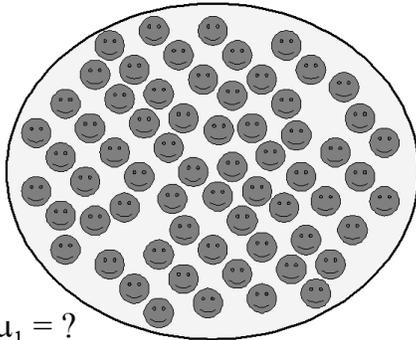
**O intervalo de confiança 95% para  $\mu_1 - \mu_2$  é:**  $(-0,004 ; 0,010 \text{ g/cm})$

Como interpretar um intervalo de confiança 95% para  $\mu_1 - \mu_2$  ?



**Exemplo:** um estudo foi conduzido para investigar os efeitos de um tratamento com medicamento anti-hipertensivo em pessoas acima de 60 anos que sofrem hipertensão sistólica isolada.

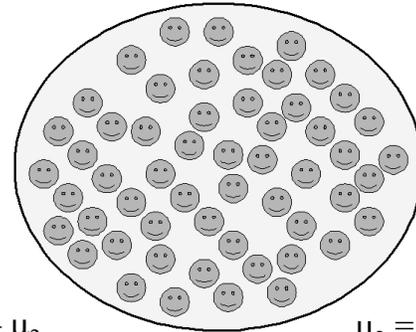
População de indivíduos com idade acima de 60 anos submetidos ao tratamento



$$\mu_1 = ?$$

$$\sigma_1 = ?$$

População de indivíduos com idade acima de 60 anos que não foram submetidos ao tratamento

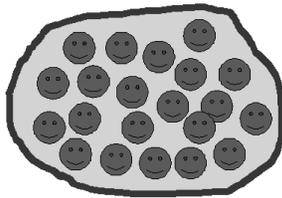


$$\mu_2 = ?$$

$$\sigma_2 = ?$$

X: pressão sangüínea sistólica

IC para  $\mu_1 - \mu_2$



Amostra 1

$$n_1 = 2308$$

$$\bar{x}_1 = 142,5 \text{ mmHg}$$

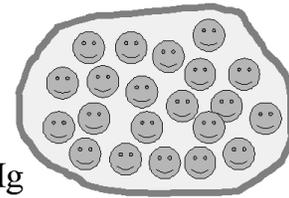
$$s_1 = 15,7 \text{ mmHg}$$

Amostra 2

$$n_2 = 2293$$

$$\bar{x}_2 = 156,5 \text{ mmHg}$$

$$s_2 = 17,3 \text{ mmHg}$$



Exercício: encontre IC 95% para  $\mu_1 - \mu_2$

## Amostras pareadas

As amostras são constituídas por um mesmo grupo de sujeitos, observados em dois momentos distintos.

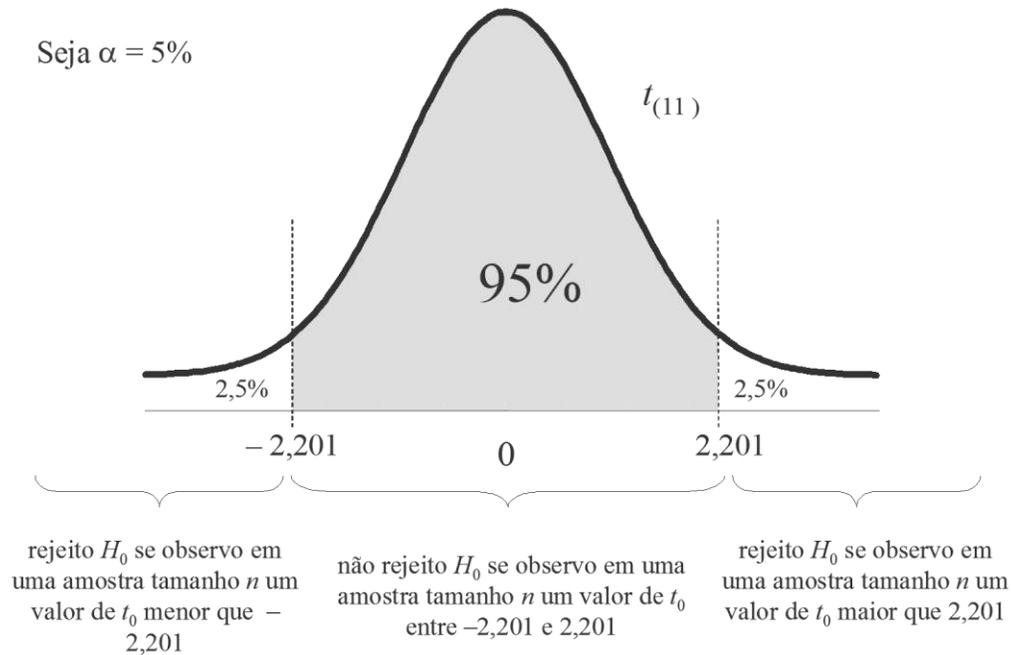
**Exemplo:** A tabela abaixo mostra as medidas da pressão arterial sistólica (em mmHg) para 12 mulheres com idade entre 20 e 35 anos, antes e após a administração de um novo contraceptivo oral.

Sujeito	Pressão arterial sistólica (mmHg)		Diferença
	Antes	Depois	
1	122	127	5
2	126	128	2
3	132	140	8
4	120	119	-1
5	142	145	3
6	130	130	0
7	142	148	6
8	137	135	-2
9	128	129	1
10	132	137	5
11	128	128	0
12	129	133	4
Média	130,67	133,25	2,58
Desvio padrão	6,933	8,226	3,088

Desejamos testar  $H_0: \mu_1 = \mu_2$   
 $H_A: \mu_1 \neq \mu_2$

Sejam  $\bar{d} = 2,58$  mmHg e  $s_d = 3,088$  mmHg.

Se  $H_0$  é verdadeira,  $t_0 = \frac{\bar{d}\sqrt{n}}{s_d}$  é resultado de uma distribuição  $t$  de Student com  $n - 1$  graus de liberdade.



Utilizando os dados de nossa amostra tamanho  $n = 12$ ,  $t_0 = \frac{\bar{d}\sqrt{n}}{S_d} = \frac{2,58\sqrt{12}}{3,088} = 2,89$ .

Portanto, \_\_\_\_\_

## Intervalo de confiança para $\mu_1 - \mu_2$

O intervalo de confiança 95% para  $\mu_1 - \mu_2$  é obtido pela expressão

$$\left( \bar{d} - t_{(n-1;0,95)} \frac{S_d}{\sqrt{n}} ; \bar{d} + t_{(n-1;0,95)} \frac{S_d}{\sqrt{n}} \right)$$

Assim, no último exemplo,

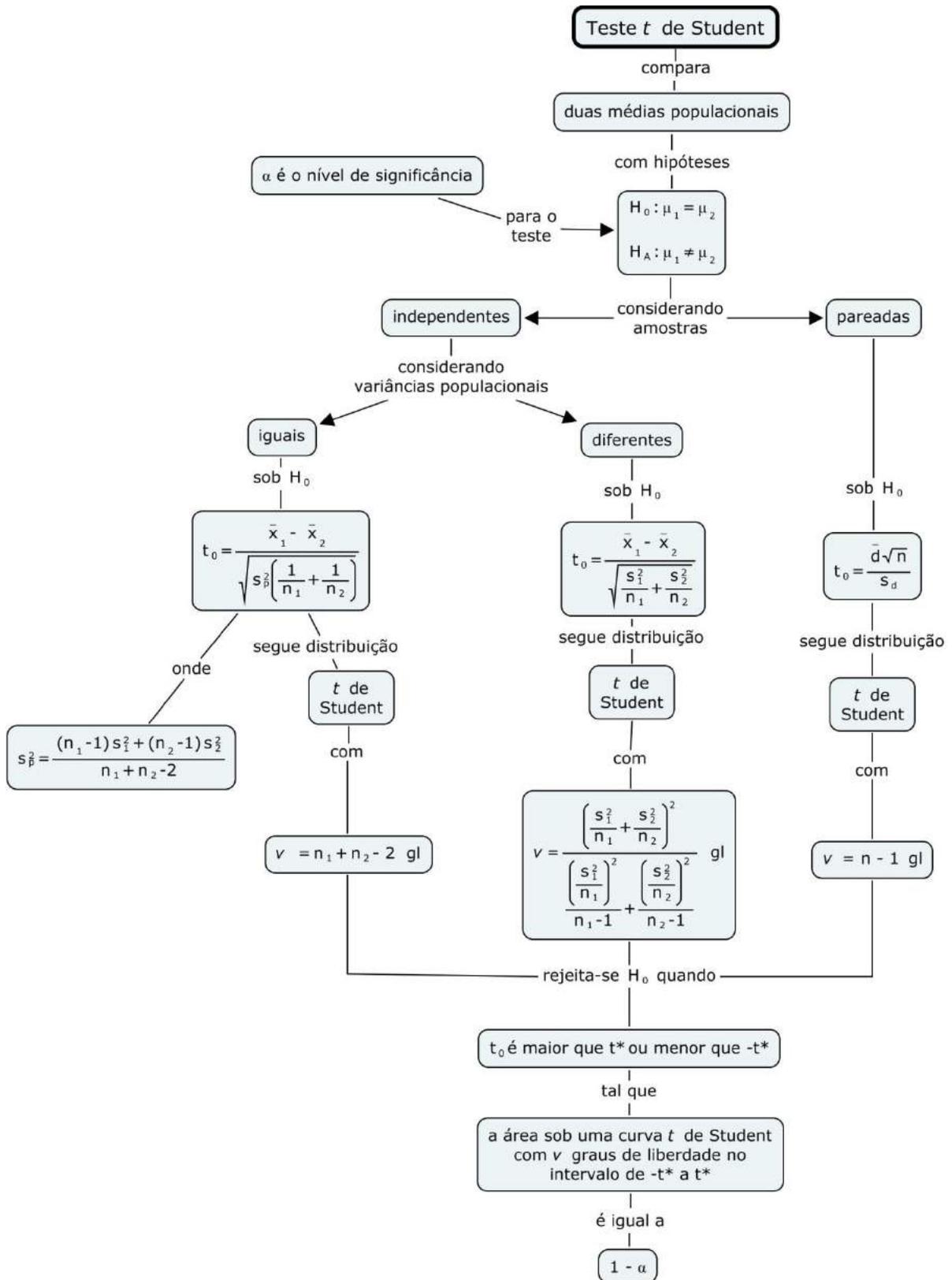
$$\left( 2,58 - 2,201 \frac{3,088}{\sqrt{12}} ; 2,58 + 2,201 \frac{3,088}{\sqrt{12}} \right)$$

$$(0,61 ; 4,54)$$

**Exercício**

Efeito do placebo e da Hidroclorotiazida sobre a pressão sistólica de 11 pacientes hipertensos (Guedes & Guedes, 1988).

Paciente	Placebo	Hidroclorotiazida
1	211	181
2	210	172
3	210	196
4	203	191
5	196	167
6	190	161
7	191	178
8	177	160
9	173	149
10	170	119
11	163	156



**Exercícios**

1. Em 1991 foi publicado na revista *Arquivos Brasileiros de Cardiologia* um estudo que discutia a importância dos fatores de riscos para doenças coronárias em atletas de elite. Neste estudo, os autores apresentaram as prevalências de fatores de risco em 88 atletas (62 homens e 26 mulheres) que representaram o Brasil nos Jogos Olímpicos de Seul em 1988. Foram observadas médias de 172 mg/dl para os níveis de colesterol total entre os homens (com um desvio padrão de 36 mg/dl) e de 187 mg/dl entre as mulheres (com um desvio padrão de 34 mg/dl). Os níveis médios de HDL colesterol foram 46 mg/dl entre os homens (com um desvio padrão de 10 mg/dl) e 60 mg/dl entre as mulheres (com um desvio padrão de 13 mg/dl). Considere que estes 62 homens e 26 mulheres podem ser tratados como amostras de populações, respectivamente, masculina e feminina, de atletas brasileiros de elite. Considere também que os níveis de colesterol total e de HDL colesterol são normalmente distribuídos nestas populações.

- (a) O problema trata de amostras independentes ou pareadas?
- (b) Sendo  $\mu_1$  o nível de colesterol total médio na população de mulheres brasileiras, atletas de elite, e  $\mu_2$  o nível de colesterol total médio na população de homens brasileiros, também atletas de elite, testar a hipótese nula  $H_0: \mu_1 - \mu_2 = 0$  contra a hipótese alternativa  $H_A: \mu_1 - \mu_2 \neq 0$  considerando um nível de significância de 5%.
- (c) Escreva, dentro do contexto do problema, o que seria um erro tipo I e um erro tipo II.

2. Pesquisadores do *King Hussein Medical Center*, em Amman, Jordânia, publicaram em 2002 no *Saudi Medical Journal* um estudo que objetivou descrever o perfil lipídico de indivíduos diabéticos e não diabéticos, com comprovação de doença da artéria coronária (DAC). Os pesquisadores compararam estes indivíduos com outros sem problemas cardíacos e não diabéticos, que compuseram um grupo de controles. Foi incluída no estudo uma amostra de 192 pacientes com DAC (incluindo diabéticos e não diabéticos) e outra de 162 indivíduos sem problemas cardíacos ou diabetes. O nível médio de colesterol no plasma para os indivíduos com DAC foi de 231,43 mg/dl, com um desvio padrão de 57,99 mg/dl. No grupo de controles, observou-se uma média para os níveis de colesterol no plasma de 202,8 mg/dl, com um desvio padrão de 36,58 mg/dl. Os níveis médios de HDL colesterol foram 35,98 mg/dl entre os indivíduos com DAC (com um desvio padrão de 9,37 mg/dl) e 44,43 mg/dl entre os indivíduos do grupo de controles (com um desvio padrão de 8,34 mg/dl). Considere que estes 192 pacientes com DAC e os 162 indivíduos sem problemas cardíacos ou diabetes podem ser tratados como amostras aleatórias retiradas das respectivas populações. Considere também que os níveis de colesterol no plasma e de HDL colesterol são normalmente distribuídos nestas populações.

- (a) O problema trata de amostras independentes ou pareadas?
- (b) Sendo  $\mu_1$  o nível de colesterol médio no plasma na população de jordanianos com DAC, e  $\mu_2$  o nível de colesterol médio no plasma na população de jordanianos sem

problemas cardíacos ou diabetes, testar a hipótese nula  $H_0: \mu_1 - \mu_2 = 0$  contra a hipótese alternativa  $H_A: \mu_1 - \mu_2 \neq 0$  considerando um nível de significância de 5%.

(c) Sendo  $\mu_1$  o nível médio de HDL colesterol na população de jordanianos com DAC, e  $\mu_2$  o nível médio de HDL colesterol população de jordanianos sem problemas cardíacos ou diabetes, testar a hipótese nula  $H_0: \mu_1 - \mu_2 = 0$  contra a hipótese alternativa  $H_A: \mu_1 - \mu_2 \neq 0$  considerando um nível de significância de 5%.

(d) Como você interpreta resultados obtidos nos itens (a) e (b) ?

**3.** Segundo dados do *U.S. National Center for Health Statistics*, os tempos (em dias) de internação dos pacientes de duas amostras independentes são mostrados abaixo. Uma primeira amostra é constituída por 40 pacientes de sexo masculino, e uma segunda amostra é constituída por 35 pacientes de sexo feminino.

Homens					mulheres				
4	4	12	18	9	14	7	15	1	12
6	12	10	3	6	1	3	7	21	4
15	7	3	55	1	1	5	4	4	3
2	10	13	5	7	5	18	12	5	1
1	23	9	2	1	7	7	2	15	4
17	2	24	11	14	9	10	7	3	6
6	2	1	8	1	5	9	6	2	14
3	19	3	1	13					

Seja  $\mu_1$  o tempo médio de internação na população masculina e  $\mu_2$  o tempo médio de internação na população feminina. Pressupondo que a variância dos tempos de internação na população masculina é igual à variância dos tempos de internação na

população feminina, testar a hipótese nula  $H_0: \mu_1 - \mu_2 = 0$  contra a hipótese alternativa  $H_A: \mu_1 - \mu_2 \neq 0$ . Considere que os tempos de internação possuem distribuição aproximadamente normal em ambas populações. Interprete o resultado.

**4.** A tabela abaixo mostra as quantidades médias de nicotina (mg) encontradas em amostras aleatórias de cigarros de uma determinada marca.

	tamanho das amostras	quantidade média de nicotina	desvio padrão
com filtro	$n_1 = 21$	0,94 mg	0,31 mg
sem filtro	$n_2 = 8$	1,65 mg	0,16 mg

Construa um teste de hipóteses com o objetivo de investigar se os cigarros com filtro e os sem filtro têm quantidades médias de nicotina diferentes. Considere um nível de significância de 5%. Escreva o que seria um erro tipo I e um erro tipo II.

**5.** Um estudo examinou a eficácia da cotinina na saliva como um indicador para a exposição à fumaça do tabaco. Foi solicitado a sete indivíduos fumar um cigarro. Estes indivíduos se abstiveram de fumar pelo menos uma semana antes do estudo. Foram tomadas amostras da saliva de todos os indivíduos 12 e 24 horas depois de terem fumado o cigarro.

indivíduo	níveis de cotinina (mmol/l)	
	depois de 12 horas	depois de 24 horas
1	73	24
2	58	27
3	67	49
4	93	59
5	33	0
6	18	11
7	147	43

(a) O problema trata de amostras independentes ou pareadas ?

(b) Sendo  $\mu_1$  e  $\mu_2$  os níveis médios de cotinina na população, respectivamente 12 e 24 horas após fumar um ci-garro, testar  $H_0: \mu_1 - \mu_2 = 0$  contra  $H_A: \mu_1 - \mu_2 \neq 0$ .

Interpretar o resultado.

6. Um pesquisador deseja comparar as medidas do nível sérico de ferro em duas populações de crianças: uma população é constituída por crianças saudáveis e a outra por crianças que sofrem de fibrose cística, uma doença congênita das glândulas da mucosa. As duas populações originais são independentes e normalmente distribuídas quanto a esta variável. O pesquisador obteve uma amostra de  $n_1 = 9$  crianças saudáveis, onde encontrou um nível médio sérico de ferro de  $\bar{x}_1 = 18,9 \mu\text{mol/l}$  e desvio padrão  $s_1 = 5,9 \mu\text{mol/l}$ . Em uma outra amostra de  $n_2 = 13$  crianças com fibrose cística, ele encontrou um nível médio sérico de ferro de  $\bar{x}_2 = 11,9 \mu\text{mol/l}$  e desvio padrão  $s_2 = 6,3 \mu\text{mol/l}$ . Sejam  $\mu_1$  e  $\mu_2$  os escores médios do nível sérico de ferro nas populações de crianças saudáveis e com fibrose cística, respectivamente.

(a) O problema trata de amostras independentes ou pareadas ?

(b) Testar  $H_0: \mu_1 - \mu_2 = 0$  contra  $H_A: \mu_1 - \mu_2 \neq 0$  e interpretar o resultado.

(c) Escreva, dentro do contexto do problema, o que seria um erro tipo I e um erro tipo II.

7. (Armitage & Berry, 1994) Em um ensaio clínico que objetivou avaliar o efeito de um novo tranqüilizante em pacientes neuróticos, a cada paciente foi administrado um tratamento com a droga e um tratamento com placebo. Um sorteio determinou se o paciente receberia primeiro a droga ou primeiro o placebo. Foi estabelecido um período entre os tratamentos suficientemente grande, de modo que a administração de um tratamento não exerça algum efeito sobre o segundo. Ao final de cada tratamento o

indivíduo	escore de ansiedade	
	droga	placebo
1	19	22
2	11	18
3	14	17
4	17	19
5	23	22
6	11	12
7	15	14
8	19	11
9	11	19
10	8	7

paciente preenchia um questionário, onde era obtido um “escore de ansiedade”, com possíveis valores entre 0 e 30. Escores altos correspondem a uma maior ansiedade. Os resultados são mostrados na tabela a seguir. Sendo  $\mu_1$  e  $\mu_2$  os escores médios de ansiedade nas populações que recebem ou não a droga, respectivamente.

(a) O problema trata de amostras independentes ou pareadas ?

(b) Testar  $H_0: \mu_1 - \mu_2 = 0$  contra  $H_A: \mu_1 - \mu_2 \neq 0$  e interpretar o resultado.

(c) Escreva, dentro do contexto do problema, o que seria um erro tipo I e um erro tipo II.

**8. (Armitage & Berry, 1994)** Um grupo de ratas foi submetido a uma dieta com alto conteúdo protéico, e outro grupo foi submetido a uma outra dieta com baixo conteúdo protéico. O ganho em peso entre o 28º e 84º dia de idade foi medido em cada rata. Os resultados são dados na tabela a seguir. Sejam  $\mu_1$  e  $\mu_2$  os ganhos médios de peso nas populações de ratas que recebem dietas com, respectivamente, alto e baixo conteúdo protéico.

ganho de peso (g)	
alto conteúdo protéico	baixo conteúdo protéico
134	70
146	118
104	101
119	85
124	107
161	132
107	94
83	
113	
129	
97	
123	

(a) O problema trata de amostras independentes ou pareadas ?

(b) Testar  $H_0: \mu_1 - \mu_2 = 0$  contra  $H_A: \mu_1 - \mu_2 \neq 0$  e interpretar o resultado.

(c) Escreva, dentro do contexto do problema, o que seria um erro tipo I e um erro tipo II.

**9. (Soares & Siqueira, 2002)** Dezenove crianças com diagnóstico de AIDS foram separadas em dois grupos de acordo com a susceptibilidade à droga. Considera-se susceptível quando o vírus HIV é inibido por concentração de zidovudina (AZT) menor do que 0,1  $\mu\text{g/L}$  e resistente quando a inibição exige nível acima de 10  $\mu\text{g/L}$ . A duração em meses da terapia com AZT relatada por Ogino (1993) é mostrada a seguir.

	Duração da terapia, em meses									
Susceptíveis ( $n = 10$ )	9	7	12	18	10	10	10	13	13	9
Resistentes ( $n = 9$ )	12	14	5	14	15	17	13	13	12	

(a) Para cada grupo, calcule a média e o desvio padrão do tempo de tratamento com AZT.

(b) Testar  $H_0: \mu_1 - \mu_2 = 0$  contra  $H_A: \mu_1 - \mu_2 \neq 0$  e interpretar o resultado.

(c) Escreva, dentro do contexto do problema, o que seria um erro tipo I e um erro tipo II.

**10. (Soares & Siqueira, 2002)** O Centro de Acompanhamento Pré-Natal, para dependentes de drogas químicas, da Escola de Medicina da Universidade de Northwestern – Chicago acompanhou a gravidez de 55 mulheres dependentes de cocaína. Destas, apesar de todo o esforço do centro, apenas 19 conseguiram parar de usar a droga durante o 1º trimestre. A tabela abaixo apresenta os resultados dos pesos dos recém-nascidos do grupo 1, filhos de mães que usaram cocaína apenas no 1º semestre de gravidez, e do grupo 2, filhos de mães que usaram cocaína durante toda a gravidez.

Informação	Grupo 1	Grupo 2
Tamanho da amostra	19	36
Média	3160 g	2829 g
Desvio padrão	453 g	708 g

- (a) Faça o teste estatístico adequado para verificar a hipótese de que os filhos de mães que usaram cocaína durante toda a gravidez têm peso ao nascer diferente dos filhos de mães que usaram cocaína apenas no primeiro trimestre de gravidez.
- (b) Estime o efeito da cocaína no peso dos recém-nascidos e construa o respectivo intervalo de confiança. Comente os resultados.

**11.** As complicações pulmonares pós-operatórias são uma fonte significativa de mortalidade e morbidade. A fisioterapia respiratória é freqüentemente utilizada na prevenção e tratamento de tais complicações, podendo ser iniciada no pré-operatório de forma a avaliar e orientar os pacientes. Uma pesquisa objetivou estabelecer a efetividade de um programa de orientação fisioterapêutica pré-operatória para pacientes submetidos à cirurgia de revascularização do miocárdio, em relação à redução do tempo de internação hospitalar, prevenção de complicações radiológicas pulmonares, alteração de volumes pulmonares e força muscular inspiratória. Foi utilizada uma amostra de pacientes submetidos à cirurgia eletiva de revascularização do miocárdio em um Instituto de Cardiologia. O grupo intervenção (54 pacientes) foi avaliado e recebeu orientação fisioterapêutica com material por escrito, no mínimo, 15 dias antes da cirurgia. Já o grupo controle (39 pacientes) recebeu cuidados de rotina no dia da internação hospitalar. O tempo médio de internação hospitalar foi de 14,65 dias no grupo controle (desvio padrão de 6,61 dias) e de 11,77 dias no grupo intervenção (desvio padrão de 6,26 dias).

- (a) Os grupos intervenção e controle representam amostras pareadas ou independentes?
- (b) Testar a hipótese de igualdade de médias populacionais do tempo de internação hospitalar entre os grupos intervenção e controle, considerando um nível de significância de 5%.
- (c) Escrever uma interpretação para o resultado que você encontrou no item (b), no contexto do problema apresentado.
- (d) Construir um intervalo de confiança 95% para a diferença entre as médias populacionais do tempo de internação hospitalar entre os grupos intervenção e controle.

**12.** A Reabilitação Pulmonar (RP) pode ser capaz de melhorar a qualidade de vida de pacientes portadores de doença pulmonar obstrutiva crônica (DPOC). Um estudo clínico foi conduzido com o objetivo de avaliar a influência da RP sobre algumas características de pacientes portadores de DPOC. Participaram deste estudo 11 pacientes, que foram submetidos a exames espirométricos, gasométricos, antropométricos e polissonográficos antes e depois de seis semanas de RP. O quadro a seguir mostra valores da pressão parcial de dióxido de carbono no sangue arterial ( $\text{PaCO}_2$ ) destes pacientes (em mmHg), antes e depois da RP.

Paciente	1	2	3	4	5	6	7	8	9	10	11
Antes	45,3	35,9	39,0	39,9	33,6	43,7	33,5	29,2	42,8	41,4	46,6
Depois	40,3	34,4	35,9	38,8	36,9	36,2	36,3	31,0	38,6	37,4	43,2

- Este estudo utiliza amostras pareadas ou independentes?
- Testar a hipótese de igualdade de médias populacionais da PaCO<sub>2</sub> antes e depois da RP, considerando um nível de significância de 5%.
- Escrever uma interpretação para o resultado que você encontrou no item (b), no contexto do problema apresentado.
- Escreva, dentro do contexto do problema, o que seria um erro tipo I e um erro tipo II.

**13.** Nas pesquisas em Fisioterapia, a avaliação do equilíbrio pode ser feita por meio do teste *Timed Up & Go* (TUG). O teste TUG faz uma monitoração rápida para detectar os problemas de equilíbrio que afetam as atividades da vida diária nos idosos. Quanto menor o tempo para a realização do teste, melhor o equilíbrio. Um estudo teve como objetivo avaliar a propensão a quedas em idosos que praticam atividade física e idosos sedentários através do teste TUG, mensurando o tempo de realização do teste em ambos os grupos. A amostra do estudo consistiu de 20 idosos que praticam atividades físicas (Grupo 1) e 20 idosos sedentários (Grupo 2), com idade entre 65 a 75 anos. Os tempos de realização (em segundos) do teste são listados a seguir, para os dois grupos.

Grupo 1	7,4	8,7	9,3	8,1	7,5	7,1	8,0	6,1	8,4	5,0	8,3	9,3	9,1	10,2
	6,0	7,6	6,0	10,0	7,7	8,6								
Grupo 2	9,9	9,7	7,6	15,1	8,2	21,2	11,8	16,0	10,8	12,8	16,9	9,6	15,3	
	17,5	15,6	16,6	14,6	5,9	16,4	18,2							

- Represente os dados deste estudo em *box-plots*. Os *box-plots* associados a cada grupo dever ser dispostos em uma única figura (e em uma única escala), de forma que seja possível compará-los.
- Escreva uma completa interpretação para os *box-plots* que você construiu no item anterior. Esta interpretação deve ser escrita de acordo com o problema apresentado.
- Considere  $\mu_1$  e  $\mu_2$  as médias populacionais para os tempos de realização do teste nas populações de idosos que praticam e não praticam atividades físicas (sedentários), respectivamente. Com base nestes dados amostrais, testar  $H_0: \mu_1 = \mu_2$  versus  $H_A: \mu_1 \neq \mu_2$ , considerando um nível de significância de 0,01.
- No contexto do problema apresentado, descreva o que seria um erro tipo I.
- No contexto do problema apresentado, escreva uma interpretação para o resultado do teste de hipóteses do item (c).

**14.** Realizou-se um estudo com o objetivo de avaliar a efetividade de uma dieta combinada com um programa de exercícios físicos na redução do nível de colesterol. A tabela abaixo mostra os níveis séricos de colesterol (em mg/dL) de 12 participantes no início e no final do programa, onde encontrou-se  $s_d = 23,13$  mg/dL.

Sujeito	1	2	3	4	5	6	7	8	9	10	11	12
Início	201	231	221	260	228	237	326	235	240	267	284	201
Final	200	236	216	233	224	216	296	195	207	247	210	209

- (a) Tratam-se de amostras pareadas ou independentes?
- (b) Sejam  $\mu_a$  e  $\mu_d$  respectivamente as médias populacionais dos níveis de colesterol antes e depois do programa. Testar a hipótese de que o programa não altera os níveis médios de colesterol ( $H_0: \mu_a = \mu_d$ ) contra a hipótese de que o programa altera estes níveis. Escreva qual foi o nível de significância adotado e uma conclusão para o resultado do teste.
- (c) Considerando o teste de hipóteses do item anterior, escreva o que seria um erro tipo I e um erro tipo II dentro do contexto do problema apresentado.

**15.** O “teste da caminhada de seis minutos” tem sido utilizado na avaliação de resultados de programas de reabilitação. Em um estudo, 9 pacientes de uma amostra aleatória foram submetidos a este teste. O quadro a seguir mostra valores da frequência cardíaca (FC) destes pacientes, em batimentos cardíacos por minuto (bpm), antes e após o teste.

Paciente	1	2	3	4	5	6	7	8	9
Antes	96	101	109	110	112	124	114	116	99
Após	109	113	117	117	127	133	127	127	104

- (a) Testar a hipótese de igualdade de médias populacionais da FC antes e após o teste da caminhada de seis minutos, considerando um nível de significância de 5%.
- (b) Escrever uma interpretação para o resultado obtido no item (a), dentro do contexto do problema apresentado.
- (c) Escreva, dentro do contexto do problema apresentado, o que seriam os erros tipo I e II.

**16.** Um estudo foi publicado na *Revista CEFAC* com o objetivo de verificar se a má oclusão Classe I de ANGLE interfere na atividade eletromiográfica dos músculos masséteres e supra-hióideos durante a fase oral da deglutição. Participaram da pesquisa vinte e seis mulheres com idade entre 20 anos e sete meses a 30 anos e oito meses, com dentição natural permanente, selecionadas por um protocolo específico, complementado com exame clínico miofuncional oral. Constituíram-se dois grupos: o controle (GC) composto por nove mulheres com oclusão clinicamente normal e o grupo I (GI) por 17 com má oclusão classe I de ANGLE. Foi investigada a atividade dos potenciais elétricos dos músculos em questão por meio de avaliação eletromiográfica de superfície. Os valores da atividade eletromiográfica (em mV) do músculo masseter esquerdo são listados a seguir.

Grupo	Atividade eletromiográfica (em mV) do músculo masseter esquerdo									dp
GC	3,7	3,6	6,5	4,3	3,0	4,0	9,2	2,0	5,6	2,16
GI	4,5	12,3	11,0	4,5	4,1	4,6	5,6	5,4	14,7	3,25
	7,6	5,6	9,4	5,6	4,0	5,9	6,1	3,2		

dp: desvio padrão

- (a) O problema envolve amostras pareadas ou independentes?
- (b) Seja  $\mu_1$  e  $\mu_2$  respectivamente as médias populacionais dos valores da atividade eletromiográfica (em mV) do músculo masseter esquerdo para as mulheres dos grupos GC e GI. Testar a hipótese  $H_0: \mu_1 = \mu_2$  contra  $H_A: \mu_1 \neq \mu_2$  considerando níveis de significância de 5% e de 1%.
- (c) Escreva uma conclusão para o resultado encontrado no item (b).
- (d) Dentro do contexto do problema apresentado, escreva o que seria um erro tipo I e um erro tipo II.
- (e) Encontre um intervalo de confiança 95% para  $\mu_2 - \mu_1$ .
- (f) Escreva uma interpretação para o intervalo encontrado no item (e).

**17.** A vitamina A é um nutriente essencial para as pessoas e a sua deficiência pode provocar alterações oculares, retardo de crescimento e aumento da susceptibilidade a infecções. Os pré-escolares estão sob maior risco para o desenvolvimento de hipovitaminose A devido ao seu rápido crescimento e desenvolvimento, com conseqüente aumento de necessidades da vitamina, além das múltiplas patologias a que estão expostos. Um estudo objetivou a comparação das médias de retinol sérico antes e após 30 dias de dose maciça de vitamina A em crianças com hipovitaminose A. Abaixo estão listados os níveis de retinol sérico (em  $\mu\text{mol/L}$ ) observados em uma amostra de 13 crianças, antes e após a intervenção.

Criança	1	2	3	4	5	6	7	8	9	10	11	12	13
Antes	1,3	2,1	1,1	0,2	0,6	1,7	1,6	0,4	1,2	2,1	1,4	1,7	1,6
Após	2,0	4,8	0,1	0,6	3,4	2,6	1,5	0,9	1,7	3,4	0,8	2,3	2,5

- (a) O problema envolve amostras pareadas ou independentes?
- (b) Seja  $\mu_1$  e  $\mu_2$  respectivamente as médias populacionais dos valores de retinol sérico para as crianças não tratadas com vitamina A e após 30 dias de dose maciça de vitamina A. Testar a hipótese  $H_0: \mu_1 = \mu_2$  contra  $H_A: \mu_1 \neq \mu_2$  considerando níveis de significância de 5% e de 1%. (Notar que o desvio padrão das diferenças é  $1,09 \mu\text{mol/L}$ ).
- (c) Escreva uma conclusão para o resultado encontrado no item (b).
- (d) Dentro do contexto do problema apresentado, escreva o que seria um erro tipo I e um erro tipo II.
- (e) Encontre um intervalo de confiança 95% para  $\mu_2 - \mu_1$ .
- (f) Escreva uma interpretação para o intervalo encontrado no item (e).

**18.** Um estudo publicado em 2008 na Revista de Neurociências objetivou avaliar os efeitos da fisioterapia sobre o equilíbrio e a qualidade de vida dos pacientes com esclerose múltipla (EM). Neste estudo, 10 pacientes com EM foram submetidos a um esquema de intervenções fisioterápicas específicas para EM. A tabela a seguir mostra os valores da escala de equilíbrio de Berg para estes pacientes, antes e após o esquema de intervenções (a escala de equilíbrio de Berg é utilizada para avaliar o equilíbrio funcional de indivíduos, a partir de tarefas que envolvem o equilíbrio estático e dinâmico).

---

Escala de Berg

---

Paciente	Antes	Depois	
1	56	58	(a) O delineamento do estudo usa amostras independentes ou pareadas?
2	42	46	(b) Escrever o que seria um erro tipo I e um erro tipo II dentro do contexto deste problema.
3	56	60	(c) Se $\mu_A$ é a média (populacional) dos valores da escala de Berg antes do esquema de intervenções e $\mu_D$ é a média populacional após o esquema de intervenções, testar a hipótese $H_0: \mu_A = \mu_D$ versus $H_A: \mu_A \neq \mu_D$ considerando um nível de significância de 5%.
4	52	58	
5	52	59	
6	54	62	
7	53	57	
8	53	60	(d) Interpretar o resultado encontrado no item anterior, dentro do contexto do problema apresentado.
9	39	41	
10	53	62	

(e) Encontrar um intervalo de confiança 95% para a diferença entre as médias populacionais e escrever uma interpretação para este intervalo dentro do contexto deste problema.

**19.** Um estudo objetivou descrever as modificações vocais acústicas e as sensações ocorridas após a técnica vocal de fonação reversa em mulheres adultas jovens, sem queixas vocais e com laringe normal. Nove mulheres submeteram-se à avaliação otorrinolaringológica e triagem fonoaudiológica para descartar possíveis alterações que pudessem interferir nos resultados da pesquisa. Elas tiveram amostras vocais coletadas antes e após realizarem três séries de 15 repetições de fonação reversa, em tempo máximo de fonação com tom e intensidade habituais, e 30 segundos de repouso passivo entre cada série. Uma das variáveis de interesse é a frequência fundamental ( $f_0$ ). São listadas a seguir as observações de  $f_0$  (em Hz), pré e pós-fonação reversa.

Paciente	1	2	3	4	5	6	7	8	9
Pré	213	213	215	208	207	216	217	212	213
Pós	212	219	220	221	207	220	227	214	224

Seja  $\mu_1$  a média de  $f_0$  pré-fonação reversa para a população de mulheres adultas jovens e seja  $\mu_2$  a média de  $f_0$  pós-fonação reversa para a população de mulheres adultas jovens.

- Testar  $H_0: \mu_1 = \mu_2$  versus  $H_A: \mu_1 \neq \mu_2$  a um nível de significância de 5% e de 1%.
- Escreva uma **interpretação** para o resultado encontrado no item anterior.
- Encontre um intervalo de confiança 95% para  $\mu_1 - \mu_2$ .
- Escreva uma **interpretação** para o intervalo de confiança encontrado no item anterior.

**20.** Um estudo objetivou verificar a ocorrência de alterações nos potenciais evocados auditivos de média e longa latências em indivíduos adultos portadores da síndrome da imunodeficiência adquirida (AIDS). Foram obtidos os potenciais evocados auditivos de média e longa latências em oito indivíduos com AIDS, que apresentavam audição normal ou até perda auditiva neurossensorial de grau moderado e resultados normais na Audiometria de Tronco Encefálico (Grupo Estudo), comparando os resultados com os obtidos no grupo controle constituído por 25 indivíduos, sem queixas auditivas e com audição dentro da normalidade, bem como com resultados normais na Audiometria de Tronco Encefálico (Grupo Controle). Foram observadas as latências e amplitudes da onda Pa, nas modalidades contralaterais C3/A2 e C4/A1, e da latência da onda P300. A tabela

a seguir mostra estatísticas descritivas para as latências da onda P300 entre os grupos controle e estudo, considerando a orelha esquerda.

Grupo	$n$	Média	Desvio padrão
Estudo	8	368,00	42,40
Controle	25	308,84	20,82

Seja  $\mu_1$  a média da latência da onda P300 a população de indivíduos com AIDS, que apresentam audição normal ou até perda auditiva neurosensorial de grau moderado e resultados normais na Audiometria de Tronco Encefálico, e seja  $\mu_2$  a média da latência da obda P300 para a população de indivíduos, sem queixas auditivas e com audição dentro da normalidade, bem como com resultados normais na Audiometria de Tronco Encefálico.

- Encontre um intervalo de confiança 95% para  $\mu_1$ .
- Encontre um intervalo de confiança 95% para  $\mu_2$ .
- Escreva uma **interpretação** para os intervalos de confiança encontrados nos itens (a) e (b).
- Testar  $H_0: \mu_1 = \mu_2$  versus  $H_A: \mu_1 \neq \mu_2$  a um nível de significância de 5% e de 1%.
- Escreva uma **interpretação** para o resultado encontrado no item anterior.
- Encontre um intervalo de confiança 95% para  $\mu_1 - \mu_2$ .
- Escreva uma **interpretação** para o intervalo de confiança encontrado no item anterior.

**21.** Um estudo avaliou o impacto do uso do leite em pó integral fortificado com 9 mg de ferro e 65 mg de vitamina C para cada 100 g de pó, sobre os níveis de hemoglobina de crianças menores de dois anos. Foi utilizada uma amostra de 9 crianças de creches municipais, por um período de 6 meses. São listados abaixo os níveis de hemoglobina (em g/dl) para as crianças, antes da realização do estudo e depois de 6 meses do uso do leite em pó.

Criança	1	2	3	4	5	6	7	8	9
Antes	9,5	10,4	10,0	8,8	10,2	11,4	9,3	9,9	8,2
Após 6 meses	11,5	12,7	12,1	9,9	12,9	13,2	10,1	12,6	9,8

Seja  $\mu_1$  a média da população de crianças menores de dois anos, antes da administração do leite em pó, e seja  $\mu_2$  a média da população de crianças menores desta faixa etária após a administração do leite em pó.

- Testar  $H_0: \mu_1 = \mu_2$  versus  $H_A: \mu_1 \neq \mu_2$  a um nível de significância de 5% e de 1%.
- Escreva uma **interpretação** para o resultado encontrado no item anterior.
- Encontre um intervalo de confiança 95% para  $\mu_1 - \mu_2$ .
- Escreva uma **interpretação** para o intervalo de confiança encontrado no item anterior.

**22.** Um estudo objetivou analisar a associação entre diversas variáveis com a síndrome metabólica (SM) em indivíduos de origem japonesa, com 30 anos de idade ou mais,

residentes em um município do interior de São Paulo. A tabela a seguir mostra estatísticas descritivas para a pressão arterial (PA) sistólica (em mmHg) encontradas em amostras de indivíduos desta população, portadores e não portadores da SM.

Grupo	$n$	Média	Desvio padrão
Portadores de SM	50	142,1	23,0
Não portadores de SM	50	121,6	21,3

Seja  $\mu_1$  a média da PA sistólica na população de indivíduos de origem japonesa residentes naquele município, portadores de SM, com idade de 30 anos ou mais, e seja  $\mu_2$  a média na população de indivíduos com estas mesmas características, mas não portadores de SM.

- Encontre um intervalo de confiança 95% para  $\mu_1$ .
- Encontre um intervalo de confiança 95% para  $\mu_2$ .
- Escreva uma **interpretação** para os intervalos de confiança encontrados nos itens (a) e (b).
- Testar  $H_0: \mu_1 = \mu_2$  versus  $H_A: \mu_1 \neq \mu_2$  a um nível de significância de 5% e de 1%.
- Escreva uma **interpretação** para o resultado encontrado no item anterior.
- Encontre um intervalo de confiança 95% para  $\mu_1 - \mu_2$ .
- Escreva uma **interpretação** para o intervalo de confiança encontrado no item anterior.

**23.** Um estudo objetivou comparar o consumo de nutrientes e alimentos em mulheres atendidas em um centro municipal de saúde do Rio de Janeiro, no período gestacional e no pós-parto. É listado abaixo o consumo de carboidrato (g por 1000 kcal) de 10 mulheres que compuseram a amostra, no período gestacional e no pós-parto.

Período	Mulheres									
	1	2	3	4	5	6	7	8	9	10
Gestacional	154	155	173	159	163	153	156	183	167	173
Pós-parto	148	149	168	151	157	146	153	178	164	167

- Testar a hipótese de igualdade de médias populacionais do consumo de carboidratos gestacional e pós-parto a um nível de significância de 5%.
- Escreva uma **interpretação** para o resultado encontrado no item anterior.
- Encontre um intervalo de confiança 95% para a diferença entre as médias populacionais.
- Escreva uma **interpretação** para o intervalo de confiança encontrado no item anterior.

**24.** O objetivo de um estudo foi avaliar a concentração de alfa-tocoferol no colostro humano em condições de suplementação com cápsulas de vitamina A acrescidas de vitamina E sintética. Para isto, foram recrutadas 10 parturientes saudáveis atendidas em uma maternidade pública. Após jejum noturno, foram coletadas amostras de sangue e de colostro das parturientes. Após a primeira coleta, as mesmas receberam suplemento na

forma de uma cápsula de palmitato de retinila acrescido de vitamina E sintética. Após 24 horas da suplementação, foi realizada nova coleta de 2 mL de colostro, também em jejum. A seguir, são exibidas as concentrações (em  $\mu\text{g/dL}$ ) de alfa-tocoferol no colostro antes e após a suplementação, observadas para as 10 parturientes.

Parturiente	1	2	3	4	5	6	7	8	9	10
Antes	1700	684	1326	791	2008	247	679	1626	1112	2135
Após 24h	2111	965	1475	1089	2272	508	818	1933	1346	2516

(a) Usando o teste de hipóteses adequado, verifique se estes dados trazem evidências de que a suplementação pode alterar, em média, as concentrações de alfa-tocoferol em parturientes saudáveis (considere um nível de significância de 5%).

(b) Encontre um intervalo de confiança 95% para a diferença média entre as concentrações de alfa-tocoferol na população de parturientes saudáveis antes e após 24 horas da suplementação. Escreva uma interpretação para este intervalo de confiança.

**25.** A fisioterapia neonatal consiste em procedimentos realizados pelo fisioterapeuta no período situado entre o clameamento do cordão umbilical até 28 dias após o parto, que compreendem o manuseio da parte motora e pulmonar do recém nascido. Um dos objetivos do manuseio pulmonar é a remoção das secreções brônquicas em excesso. Um estudo clínico utilizou uma amostra de 9 recém-nascidos pré-termos sob ventilação mecânica invasiva, hospitalizados em uma unidade de terapia intensiva neonatal. Os dados listados a seguir referem-se aos valores percentuais da saturação arterial periférica da hemoglobina pelo oxigênio ( $\text{SpO}_2$ ) observados três minutos antes e três minutos após um procedimento de fisioterapia neonatal.

RN	1	2	3	4	5	6	7	8	9
Antes	94,0	95,6	95,7	96,3	96,9	96,9	94,5	98,3	95,5
Após	92,2	97,3	97,6	96,3	98,3	93,2	92,8	99,0	94,4

(a) Com base nos dados expostos acima e nos procedimentos de inferência estatística discutidos na disciplina, utilize o teste de hipóteses adequado para testar a hipótese (nula) de igualdade entre as médias populacionais dos valores percentuais de  $\text{SpO}_2$  antes e após o procedimento (use  $\alpha = 0,05$ ).

(b) Escreva uma possível conclusão para o estudo, com base no resultado encontrado no item (a).

(c) Encontre um intervalo de confiança 95% para a diferença entre as médias populacionais dos valores percentuais de  $\text{SpO}_2$  antes e após o procedimento e escreva uma interpretação para o intervalo encontrado.

**26.** Com o intuito de comparar as dimensões do palato duro de crianças respiradoras nasais e respiradoras orais, uma fonoaudióloga usou um paquímetro para tomar algumas medidas de interesse. Em uma amostra de 10 crianças respiradoras nasais, ela encontrou as seguintes observações da distância transversal (em milímetros) entre os pontos da região dos caninos superiores:

25,0 25,6 26,3 28,9 29,4 26,5 24,9 27,7 29,8 29,9

Em outra amostra de 10 crianças respiradoras orais, ela encontrou estas observações para esta variável:

28,1 27,9 26,2 23,9 26,1 27,5 29,2 25,1 23,0 26,3

(a) Use o teste de hipóteses adequado para comparar, com base nestes dados, as médias populacionais da distância transversal entre os pontos da região dos caninos superiores, de crianças respiradoras nasais e orais. Use um nível de significância de 5%.

(b) Escreva quando seriam cometidos os erros tipos I e II para o teste de hipóteses proposto no item (a), no contexto do problema enunciado.

(c) Encontre um intervalo de confiança 95% para a diferença entre as médias populacionais da distância transversal entre os pontos da região dos caninos superiores entre crianças respiradoras nasais e orais, e escreva uma interpretação para o intervalo encontrado.

**27.** Um estudo objetivou analisar as respostas metabólicas e hemodinâmicas decorrentes da ingestão de bebidas energéticas. Uma amostra de 10 indivíduos de sexo masculino fisicamente ativos, não fumantes, com idade entre 20 e 25 anos, foi utilizada na pesquisa. Em um dia determinado, os indivíduos ingeriram 500 ml de uma bebida energética e foram submetidos a um teste ergoespirométrico. Durante o teste, algumas variáveis foram tomadas, incluindo a frequência cardíaca. Uma semana depois, os mesmos indivíduos ingeriram 500 ml de água e foram novamente submetidos ao teste ergoespirométrico. As variáveis de interesse foram tomadas mais uma vez. O quadro a seguir exhibe as medidas de frequência cardíaca (em batimentos por minutos) destes indivíduos, tomadas quando ingeriram a bebida energética e quando ingeriram água.

Indivíduo	1	2	3	4	5	6	7	8	9	10
Energético	180	188	177	187	193	188	185	188	199	164
Água	186	187	179	191	199	187	189	189	204	162

(a) Usando o teste de hipóteses adequado, testar a hipótese (nula) que a média da frequência cardíaca na população de indivíduos com as características descritas no enunciado do problema não se modifica ao ingerirem água ou energéticos na prática de exercícios. Considere um nível de significância de 5%.

(b) Considerar agora um nível de significância de 1% e comparar com o resultado encontrado no item anterior. O resultado do teste de hipóteses modificou-se?

**28.** Os dados a seguir referem-se às medidas do terço médio da face (da glabella ao subnasal, em milímetros) de crianças respiradoras nasais (RN) e orais (RO). Estas medidas foram obtidas sem pressionar as pontas do paquímetro contra a superfície da pele.

Grupo	Medidas do terço médio da face (mm)										<i>s</i>
RN	52,5	47,9	50,7	53,5	52,4	54,4	53,9	57,0	50,8	51,7	2,47
RO	49,9	52,5	50,4	53,4	57,1	56,3	51,1	45,3	54,2	52,8	3,40

*s*: desvio padrão amostral

(a) Considere estes dados amostrais e use o teste de hipóteses adequado para testar a hipótese (nula) de igualdade entre as médias populacionais das medidas do terço médio da face entre crianças respiradoras nasais e orais, considerando níveis de significância de 1% e 5%. Escreva uma conclusão para o resultado deste teste de hipóteses.

(b) Encontre o intervalo de confiança 95% para a diferença entre as médias das populações de crianças respiradoras nasais e orais, e escreva uma interpretação para o intervalo encontrado.

**29.** Um estudo teve por objetivo avaliar os efeitos de um programa de exercícios físicos (PEF) com frequência de duas sessões semanais e duração de doze semanas nos níveis pressóricos, na aptidão física e na capacidade funcional de idosos com hipertensão arterial (HA). Dentre algumas variáveis de interesse, a força muscular de membros inferiores foi avaliada pelo “teste de sentar e levantar da cadeira” (TSL), no qual o voluntário posicionava-se sentado em uma cadeira de 43 cm de altura, com os pés apoiados no chão e braços cruzados contra o tórax. Ao sinal, levantava-se e, em seguida, sentava-se completamente. Era então anotado o maior número de ciclos completos realizados durante 30 segundos. Os dados a seguir referem-se aos números de ciclos completos obtidos do TSL em uma amostra de voluntários, anotados antes e após o PEF.

<b>Voluntário</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
Antes do PEF	13	12	12	11	14	7	7	15	9	11	13	7
Após o PEF	15	11	15	15	14	6	9	18	9	12	13	8

(a) Utilize o teste de hipóteses adequado para avaliar se há (ou não) evidências de que o PEF traz alterações sobre o número médio de ciclos completos do TSL em uma população de idosos com HA. Para isso, utilize um nível de significância de 0,05.

(b) Escreva uma interpretação adequada para o resultado obtido no item (a).

(c) Construa um intervalo de confiança 95% para a diferença entre as médias do número de ciclos completos do TSL anteriores e posteriores ao PEF em uma população de idosos com HA. Não deixe de escrever uma interpretação para o intervalo de confiança obtido.

**30.** O teste “*timed up and go*” (TUG) mensura o tempo (em segundos) que um indivíduo demora para levantar-se de uma cadeira, caminhar 2,5m, retornar, caminhar de volta à cadeira e sentar-se. Um estudo objetivou comparar os tempos médios do teste TUG em populações de idosos praticantes e não praticantes de equoterapia. Em uma amostra de 22 idosos que praticavam equoterapia a pelo menos um ano, observou-se um tempo médio de 5,12 segundos e um desvio padrão de 2,12 segundos. Em outra amostra de 23 idosos não praticantes de equoterapia, observou-se um tempo médio de 5,98 segundos e um desvio padrão de 1,01 segundos. Considerando um nível de significância de 0,05, utilize o teste de hipóteses adequado para avaliar se há (ou não) evidências de que idosos praticantes e não praticantes de equoterapia têm, em média, desempenhos diferentes no teste TUG. Não deixe de escrever uma interpretação adequada para o resultado do teste de hipóteses.

**31.** A escoliose é caracterizada por uma curvatura lateral da coluna vertebral. Um estudo objetivou comparar as alterações posturais no pré e pós-operatório de artrodese da coluna vertebral em adolescentes com escoliose idiopática. São listados abaixo os ângulos (em

graus) formados pelo ponto mais alto do trapézio em relação ao pescoço e o manúbrio, tendo como vértice o acrômio, para o lado esquerdo, mensurados em uma amostra de 16 adolescentes, antes e após a cirurgia. Por simplicidade, chamaremos esta variável de TAM.

Adolescente	1	2	3	4	5	6	7	8
Pré-operatório	27,7	25,6	22,2	22,7	24,0	31,2	23,4	25,6
Pós-operatório	30,5	26,5	25,5	27,2	29,2	33,4	26,5	30,8

Adolescente	9	10	11	12	13	14	15	16
Pré-operatório	22,5	19,7	27,1	13,9	19,8	26,2	29,4	19,5
Pós-operatório	24,5	22,1	30,5	17,4	19,6	29,5	32,5	21,1

- Use um teste de hipóteses adequado para testar se há alguma diferença entre as médias populacionais do TAM antes e após a cirurgia. Considere um nível de significância de 5%. Notar que  $s_d = 1,438$  graus.
- Escreva uma interpretação adequada para o resultado obtido no item (a).
- Encontre um intervalo de confiança 95% para a diferença entre as médias populacionais do TAM antes e após a cirurgia.
- Escreva uma interpretação adequada para o intervalo de confiança obtido no item (c).

**32.** A postura da cabeça pode ser mensurada pelo ângulo craniovertebral (CV) formado por dois pontos anatômicos, tragus e sétima vértebra cervical, e a linha horizontal. Os valores deste ângulo indicam a posição da cabeça em relação ao tronco e, quando decrescentes, são indicativos de uma postura anteriorizada da cabeça. Em uma pesquisa, o grupo estudo (GE) foi composto por 50 mulheres na faixa entre 20 e 50 anos com queixas de dor cervical por mais de três meses, e o grupo controle (GC), por 50 mulheres assintomáticas. A tabela a seguir exibe estatísticas descritivas para o ângulo CV nos grupos GE e GC obtidas das respectivas amostras.

Grupo	$n$	Média	Desvio padrão
GE	50	53,33°	3,38°
GC	50	48,55°	5,58°

- Use um teste de hipóteses adequado para testar se há alguma diferença entre as médias populacionais do ângulo CV das mulheres com queixas de dor cervical e das mulheres assintomáticas. Considere um nível de significância de 5%.
- Escreva uma interpretação adequada para o resultado obtido no item (a).
- Encontre um intervalo de confiança 95% para a diferença entre as médias populacionais do ângulo CV das mulheres com queixas de dor cervical e das mulheres assintomáticas.
- Escreva uma interpretação adequada para o intervalo de confiança obtido no item (c).

**33.** A Doença Pulmonar Obstrutiva Crônica (DPOC) tem como característica principal uma alteração da função pulmonar associada à disfunção dos músculos esqueléticos

periféricos, acarretando intolerância ao exercício e piora progressiva do condicionamento físico. A fisioterapia respiratória busca melhorar a capacidade funcional do indivíduo portador de DPOC. Um estudo utilizou uma amostra de 11 voluntários, que foram submetidos a um programa de exercícios realizado duas vezes por semana, com duração de 60 minutos cada sessão. São mostrados abaixo valores do volume expiratório forçado (VEF) no 1º segundo (em litros), tomados para cada voluntário antes da intervenção e ao término do programa de exercícios, que teve a duração de 12 semanas.

Voluntário	1	2	3	4	5	6	7	8	9	10	11
Antes	2,00	1,48	0,74	1,24	1,48	2,12	0,91	1,46	0,69	1,20	1,34
Depois	2,19	1,61	0,79	1,23	1,45	1,94	1,21	1,59	0,73	1,11	1,41

- (a) Seja um teste de hipóteses que compara as médias populacionais do VEF antes e após a realização do programa de exercícios. A hipótese nula do teste assume igualdade entre as médias. Descreva, no contexto deste estudo, o que seriam os erros tipo I e II.
- (b) Considerando um nível de significância de 0,05, utilize um teste de hipóteses adequado para comparação entre médias populacionais do VEF antes e após realização do programa de exercícios.
- (c) Escreva uma interpretação adequada ao resultado que você encontrou no item (b), no contexto do estudo apresentado.
- (d) Considere agora um nível de significância de 0,01 e utilize novamente o teste de hipóteses para comparação entre médias populacionais do VEF antes e após realização do programa de exercícios. Neste caso, a conclusão se altera? Justifique sua resposta.

**34.** O teste clínico “*Timed Up and Go*” (TUG) mensura o tempo que um indivíduo demora para levantar-se de uma cadeira, caminhar 2,5m, retornar e caminhar de volta à cadeira e sentar-se. Esse teste avalia a agilidade e o equilíbrio dinâmico. Um estudo utilizou uma amostra de 20 idosos que praticam atividades físicas (Grupo 1) e outros 20 idosos sedentários (Grupo 2), com idade entre 65 a 75 anos. O tempo médio de realização do teste TUG foi de 7,75 segundos para os idosos do Grupo 1, com um desvio padrão de 1,46 segundos. Os idosos do Grupo 2 realizaram o teste TUG em uma média de 13,56 segundos, com um desvio padrão de 3,41 segundos.

- (a) Considerando um nível de significância de 0,05, utilize um teste de hipóteses adequado para comparar as médias populacionais do tempo de realização do teste TUG entre idosos praticantes de atividades físicas e sedentários.
- (b) Escreva uma interpretação adequada ao resultado que você encontrou no item (b), no contexto do estudo apresentado.
- (c) Encontre um intervalo de confiança 95% para a diferença entre as médias populacionais dos tempos de realização do teste TUG pelos idosos praticantes de atividades físicas e sedentários.
- (d) Escreva uma interpretação adequada ao intervalo de confiança que você encontrou no item (b), no contexto do estudo apresentado.

**35.** A fisioterapia neonatal consiste em procedimentos realizados pelo fisioterapeuta no período situado entre o clameamento do cordão umbilical até 28 dias após o parto, que compreendem o manuseio da parte motora e pulmonar do recém-nascido (RN). Um

estudo utilizou uma amostra de 12 recém-nascidos pré-termo sob ventilação mecânica invasiva internados na Unidade de Terapia Intensiva Neonatal de um hospital especializado. A seguir, estão listadas as frequências cardíacas (FC) dos recém-nascidos, antes e após a realização da fisioterapia neonatal, em batimentos por minuto (bpm).

RN	1	2	3	4	5	6	7	8	9	10	11	12
Antes	114	153	135	160	163	157	147	158	144	165	130	108
Após	104	144	121	144	148	147	139	150	139	158	121	97

(a) Utilize um teste de hipóteses adequado para testar se a fisioterapia neonatal tem algum efeito sobre as médias populacionais de FC em RN com estas características, considerando um nível de significância de 0,05 e que o desvio padrão das diferenças é 3,3 bpm.

(b) Escreva uma interpretação correta para o resultado encontrado no item (a), de acordo com o contexto do problema apresentado.

**36.** Em um estudo foram avaliados 12 pacientes com doença pulmonar obstrutiva crônica (DPOC) moderada a grave e sete indivíduos saudáveis. A partir da porcentagem de gordura obtida por análise de impedância bioelétrica, foi calculado o peso corporal magro (PCM) de cada indivíduo. A média do PCM encontrada nos pacientes com DPOC foi de 51 kg. Nos pacientes saudáveis, a média encontrada foi 57 kg. Nos dois grupos, o desvio padrão do PCM foi igual a 4 kg.

(a) Utilize um teste de hipóteses adequado para testar se as médias populacionais da PCM são iguais entre portadores e não portadores de DPOC.

(b) Escreva uma interpretação correta para o resultado encontrado no item (a), de acordo com o contexto do problema apresentado.

(c) Encontre um intervalo de confiança 95% para a diferença entre as médias populacionais da PCM entre portadores e não portadores de DPOC.

(d) Escreva uma interpretação correta para o resultado encontrado no item (c), de acordo com o contexto do problema apresentado.

### Respostas (a interpretação dos resultados fica por sua conta!):

- (a) independentes, (b)  $n_1 = 62$ ,  $n_2 = 26$ ,  $\bar{x}_1 = 172$  mg/dl,  $\bar{x}_2 = 187$  mg/dl,  $s_1 = 36$  mg/dl,  $s_2 = 34$  mg/dl,  $S_p^2 = 1255,302$  (mg/dl)<sup>2</sup>,  $t_0 = -1,812$ ,  $gl = 86$  (rej.  $H_0$ ) **(interpretar!)**
- (a) independentes, (b)  $n_1 = 192$ ,  $n_2 = 162$ ,  $S_p^2 = 2436,75$  (mg/dl)<sup>2</sup>,  $t_0 = 5,44$  (rej.  $H_0$ ) **(interpretar!)**, (c)  $S_p^2 = 79,45$  (mg/dl)<sup>2</sup>,  $t_0 = -8,89$  **(interpretar!)**
- $n_1 = 40$ ,  $n_2 = 35$ ,  $\bar{x}_1 = 9,075$ ,  $\bar{x}_2 = 7,114$ ,  $s_1 = 9,773$ ,  $s_2 = 5,161$ ,  $S_p^2 = 63,43$ ,  $gl = 73$ ,  $t_0 = 1,06$ . **(interpretar!)**
- $t_0 = -6,12$  **(interpretar!)**
- (a) pareadas, (b)  $\bar{d} = 39,4$  mmol/l,  $s_d = 31,4$  mmol/l,  $t_0 = 3,32$  **(interpretar!)**
- (a) independentes, (b)  $S_p^2 = 37,738$ ,  $t_0 = 2,63$  **(interpretar!)**

7. (a) pareadas, (b)  $\bar{d} = 1,3$  ponto,  $s_d = 4,55$  pontos,  $t_0 = 0,904$  (**interpretar!**)
8. (a) independentes, (b)  $n_1 = 12$ ,  $n_2 = 7$ ,  $\bar{x}_1 = 120$  g,  $\bar{x}_2 = 101$  g,  $s_1 = 21,39$  g,  $s_2 = 20,62$  g,  $S_p^2 = 446,12$  g,  $gl = 17$ ,  $t_0 = 1,89$  (**interpretar!**)
9.  $n_1 = 10$ ,  $n_2 = 9$ ,  $\bar{x}_1 = 11,1$   $\mu\text{g/L}$ ,  $\bar{x}_2 = 12,8$   $\mu\text{g/L}$ ,  $s_1 = 3,07$   $\mu\text{g/L}$ ,  $s_2 = 3,31$   $\mu\text{g/L}$ ,  $S_p^2 = 10,14$   $\mu\text{g/L}$ ,  $gl = 17$ ,  $t_0 = -1,16$  (**interpretar!**)
10. (a)  $S_p^2 = 400.717,02$  g<sup>2</sup>,  $t_0 = 1,84$ ,  $gl = 53$  (**interpretar!**) (b) IC95% (-29,04; 691,04) (**interpretar!**)
11. (a) independentes, (b)  $S_p^2 = 41,07$  dias,  $t_0 = -2,13$ ,  $gl = 91$  (**interpretar!**), (d) IC95% (-5,56; -0,21) (**interpretar!**)
12. (a) pareadas, (b)  $\bar{d} = 1,99$  mmHg,  $s_d = 3,43$  mmHg,  $t_0 = 1,92$  (**interpretar!**)
13. (c)  $n_1 = 20$ ,  $\bar{x}_1 = 7,92$  segundos,  $s_1 = 1,39$  segundos,  $n_2 = 20$ ,  $\bar{x}_2 = 13,49$  segundos,  $s_2 = 4,09$  segundos,  $S_p^2 = 9,33$  segundos<sup>2</sup>,  $t_0 = -5,77$  (**interpretar!**)
14. (a) pareadas, (b)  $n = 12$ ,  $\bar{d} = 20,2$  mg/dL,  $s_d = 23,13$  mg/dL,  $t_0 = 3,02$  (**interpretar!**)
15. (a)  $\bar{d} = -10,3$  bpm,  $s_d = 3,28$  bpm,  $t_0 = -9,45$  (**interpretar!**)
16. (a) independentes, (b) GC:  $n_1 = 9$ ,  $\bar{x}_1 = 4,66$  mV,  $s_1 = 2,16$  mV; GI:  $n_2 = 17$ ,  $\bar{x}_2 = 6,71$  mV,  $s_2 = 3,25$  mV;  $S_p^2 = 8,60$ ,  $t_0 = -1,696$  (**interpretar!**), (e)  $t^* = 2,06$ , IC95% para  $\mu_1 - \mu_2$ : (-0,44; 4,54) (**interpretar!**)
17. (a) pareadas, (b)  $\bar{d} = -0,7385$   $\mu\text{mol/L}$ ,  $s_d = 1,0889$ ,  $t_0 = -2,445$  (**interpretar!**), (e) IC95% (0,08; 1,40) ou (-1,40; -0,08) (**interpretar!**)
18. (a) pareadas, (b) um erro tipo I é cometido ao concluir que o esquema de intervenções fisioterápicas tem algum efeito sobre o escala de equilíbrio, quando na realidade, não tem, (c)  $\bar{d} = 5,3$ ,  $s_d = 2,45$ ,  $t_0 = 6,84$ ,  $t^* = 2,262$  (**interpretar!**), (e) IC95% (3,55; 7,05)
19. (a)  $\bar{d} = 50/9 = 5,56$  Hz,  $s_d = 4,93$  Hz,  $t_0 = 3,38$ ,  $t^* = 2,306$  se  $\alpha = 5\%$  (rejeita-se  $H_0$ ),  $t^* = 3,355$  se  $\alpha = 1\%$  (rejeita-se  $H_0$ ) (**interpretar!**), (c) IC95% (1,77; 9,35) (**interpretar!**)
20. (a) (332,55; 403,45) (**interpretar!**), (b) (302,61; 315,07) (**interpretar!**), (d)  $S_p^2 = 741,54$ ,  $t_0 = 6,14$ ,  $gl = 31$ ,  $t^* = 2,04$  se  $\alpha = 5\%$  (rejeita-se  $H_0$ ),  $t^* = 2,744$  se  $\alpha = 1\%$  (rejeita-se  $H_0$ ) (**interpretar!**), (f) IC95% para  $\mu_1 - \mu_2$ : (39,5; 78,8)
21. (a)  $\bar{d} = 17,1/9 = 1,9$  g/dl,  $s_d = 0,656$  g/dl,  $t_0 = 8,192$ ,  $t^* = 2,306$  se  $\alpha = 5\%$ ,  $t^* = 3,355$  se  $\alpha = 1\%$  (**interpretar!**), (c) (1,3 ; 2,4) (**interpretar!**)
22. (a) IC95% para  $\mu_1$ : (135,6; 148,6), (b) IC95% para  $\mu_2$ : (115,5; 127,6), (c) (**interpretar!**), (d)  $S_p^2 = 491,345$  mmHg,  $t_0 = 20,5/4,43 = 4,628$ ,  $gl = 98$ ,  $t^* = 1,98$  se  $\alpha = 5\%$ ,  $t^* = 2,63$  se  $\alpha = 1\%$  (**interpretar!**), (f) IC95% para  $\mu_1 - \mu_2$ : (11,7; 29,3) (**interpretar!**)
23. (a)  $\bar{d} = 5,5$  g por 1000 kcal,  $s_d = 1,581$  g por 1000 kcal,  $t_0 = 11,0$ ,  $t^* = 2,262$  (**interpretar!**), (c) IC95% para  $\mu_1 - \mu_2$ : (4,36 ; 6,64)
24. (a)  $\bar{d} = 272,5$   $\mu\text{g/dL}$ ,  $s_d = 86,69$   $\mu\text{g/dL}$ ,  $t_0 = 9,94$  (**interpretar!**), (b) IC 95% para  $\mu_1 - \mu_2$ : (210,5 ; 334,5) (**interpretar!**)

25. (a)  $\bar{d} = 0,289\%$ ,  $s_d = 1,911\%$ ,  $t_0 = 0,45$  (**interpretar!**), (c) IC 95% para  $\mu_1 - \mu_2$ :  $(-1,18 ; 1,76)$  (**interpretar!**)
26. Crianças respiradoras nasais:  $n_1 = 10$ ,  $\bar{x}_1 = 27,4$  mm,  $s_1 = 1,989$  mm. Respiradoras orais:  $n_2 = 10$ ,  $\bar{x}_2 = 26,3$  mm,  $s_2 = 1,937$  mm. (a) 18 graus de liberdade,  $S_p^2 = 3,854$ ,  $t_0 = 1,219$  (**interpretar!**), (c) IC 95% para  $\mu_1 - \mu_2$ :  $(-0,775 ; 2,915)$  (**interpretar!**)
27. (a)  $n = 10$ ,  $\bar{d} = -2,4$  bpm,  $s_d = 3,026$  bpm,  $t_0 = -2,51$  (**interpretar!**)
28. (a)  $n_1 = 10$ ,  $\bar{x}_1 = 52,48$  mm,  $s_1 = 2,47$  mm,  $n_2 = 10$ ,  $\bar{x}_2 = 52,30$  mm,  $s_1 = 3,40$  mm,  $S_p^2 = 8,83$ ,  $t_0 = 0,135$  (**interpretar!**), (b) IC95% para  $\mu_1 - \mu_2$ :  $(-2,61 ; 2,97)$  (**interpretar!**)
29. (a)  $n = 12$ ,  $\bar{d} = -1,17$  ciclos,  $s_d = 1,642$ ,  $t_0 = -2,46$  (**interpretar!**), (b) IC95% para  $\mu_1 - \mu_2$ :  $(-2,21 ; -0,13)$  (**interpretar!**)
30. Temos  $n_1 = 22$ ,  $\bar{x}_1 = 5,12$ s,  $s_1 = 2,12$ s,  $n_2 = 23$ ,  $\bar{x}_2 = 5,98$ s,  $s_1 = 1,01$ s,  $S_p^2 = 2,72$ ,  $t_0 = -1,75$  (**interpretar!**)
31. (a)  $\bar{d} = 2,89$  graus,  $s_d = 1,438$  graus,  $n = 16$ ,  $t_0 = 8,05$ ,  $t^* = 2,131$  (**interpretar!**), (c) IC95% para  $\mu_1 - \mu_2$ :  $(2,13 ; 3,65)$  (**interpretar!**)
32. (a)  $n_1 = n_2 = 50$ ,  $S_p^2 = 21,28$ ,  $t_0 = 5,18$  (**interpretar!**), (c) IC95% para  $\mu_1 - \mu_2$ :  $(2,94 ; 6,61)$  (**interpretar!**)
33. (b)  $\bar{d} = 0,05$  litros,  $s_d = 0,133$  litros,  $t_0 = -1,36$ ,  $t^* = 2,228$  (**interpretar!**), (d)  $t^* = 3,169$  (**interpretar!**)
34. (a)  $S_p^2 = 6,88$  segundos<sup>2</sup>,  $t_0 = -7$ s (**interpretar!**), (c) IC95%  $(-7,48 ; -4,13)$  (**interpretar!**)
35. (a)  $n = 12$ ,  $\bar{d} = 10,2$  bpm,  $s_d = 3,3$  bpm,  $t_0 = 10,6$  (**interpretar!**)
36.  $n_1 = 12$ ,  $\bar{x}_1 = 51$  kg,  $s_1 = 4$  kg,  $n_2 = 7$ ,  $\bar{x}_2 = 57$  kg,  $s_2 = 4$  kg,  $S_p^2 = 16$  kg<sup>2</sup>. (a)  $t_0 = -3,15$ ,  $t^* = 2,110$  (**interpretar!**) (c) IC95% para  $\mu_1 - \mu_2$ :  $(-10,01 ; -1,99)$  (**interpretar!**).

## Teste qui-quadrado



<https://www.youtube.com/watch?v=G4PtWdmMMnI>

**Objetivo:** testar a associação entre duas variáveis **qualitativas**, que chamaremos genericamente de  $A$  e  $B$ .

Hipóteses:  $H_0$ :  $A$  e  $B$  são independentes  
(não há associação entre  $A$  e  $B$ )

$H_A$ :  $A$  e  $B$  não são independentes  
(há algum tipo de associação entre  $A$  e  $B$ )

Vamos supor que  $A$  e  $B$  são variáveis qualitativas binárias. Exemplos:

- Sexo (masculino ou feminino)
- Tabagismo (fumante ou não fumante)
- Doença (portador ou não portador)
- Atividade física (prática ou não prática)

**Exemplo:** em uma pesquisa, foi perguntado ao entrevistado se ele tinha feito pelo menos uma consulta médica no último ano. Sejam as seguintes variáveis:

$$A : \text{consulta médica} \begin{cases} \text{realizou} \\ \text{não realizou} \end{cases} \quad B : \text{sexo} \begin{cases} \text{masculino} \\ \text{feminino} \end{cases}$$

Pergunta: há alguma associação entre a realização de pelo menos uma consulta médica no último ano ( $A$ ) e sexo ( $B$ )?

Hipóteses:  $H_0$ : não há associação entre  $A$  e  $B$   
 $H_A$ : há algum tipo de associação entre  $A$  e  $B$

Digamos que, em uma amostra de  $n = 100$  pessoas, 45 são homens, e 55 são mulheres.

Dentre os 45 homens, 19 realizaram ao menos uma consulta médica no último ano.

Dentre as 55 mulheres, 31 realizaram ao menos uma consulta médica no último ano.

Podemos então representar estes dados na tabela abaixo:

Sexo	Consulta médica no último ano		Total
	Realizou	Não realizou	
Homens	19	26	45
Mulheres	31	24	55
Total	50	50	100

De uma forma geral, vamos representar estes dados de acordo com esta notação:

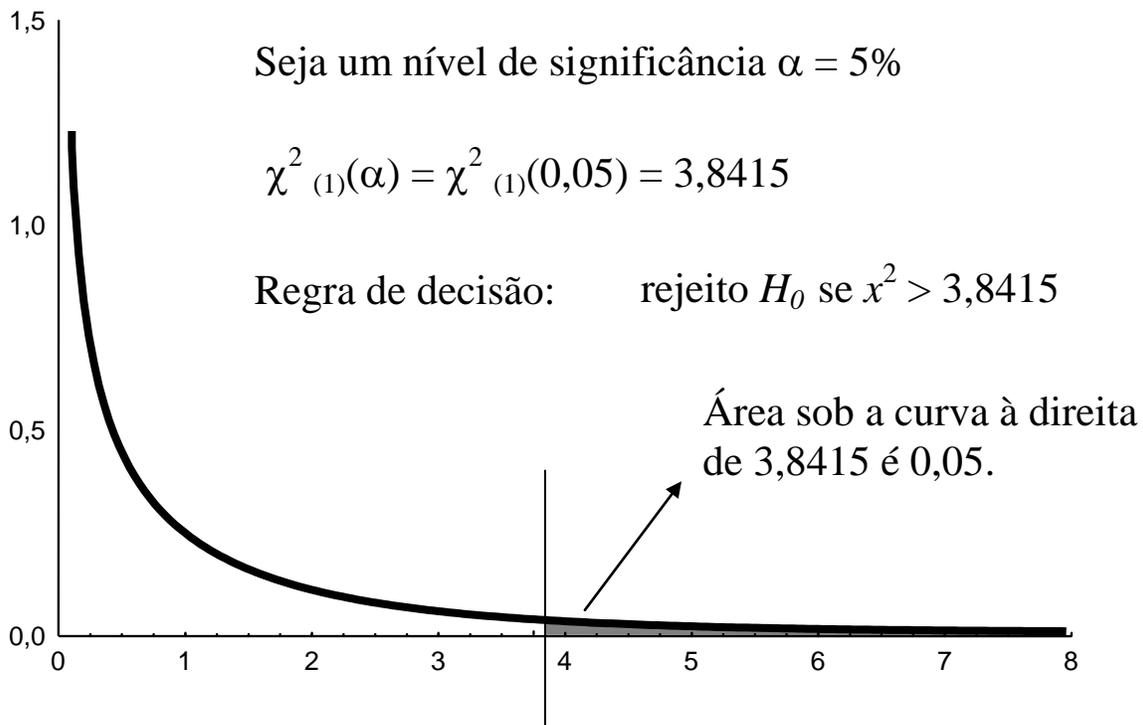
Variável B	Variável A		Total
	“categoria 1”	“categoria 2”	
“categoria 1”	$a$	$b$	$a + b$
“categoria 2”	$c$	$d$	$c + d$
Total	$a + c$	$b + d$	$n = a + b + c + d$

No nosso exemplo, temos  $a = 19$ ,  $b = 26$ ,  $c = 31$  e  $d = 24$ .

Para testarmos as hipóteses em questão, calculamos a *estatística qui-quadrado* dada por:

$$x^2 = \frac{n(a \times d - b \times c)^2}{(a + b) \times (c + d) \times (a + c) \times (b + d)}$$

Para um nível de significância  $\alpha$ , rejeitamos  $H_0$  se  $x^2$  for superior a um valor determinado de maneira que a área sob a *curva qui-quadrado com 1 grau de liberdade* à direita deste valor é  $\alpha$ .



Temos, portanto,  $x^2 = \frac{n(a \times d - b \times c)^2}{(a + b) \times (c + d) \times (a + c) \times (b + d)} = \frac{100(19 \times 24 - 26 \times 31)^2}{45 \times 55 \times 50 \times 50} = 1,98$ .

Não rejeitamos  $H_0$ , pois  $x^2$  não é maior que 3,8415. Portanto, não temos evidências de que a realização ou não de consultas médicas tem associação com o sexo.

**Exercícios**

1. Em uma amostra de 64 idosos, 42 participaram de um programa de prevenção de quedas. Seis meses após a implementação do programa, verificou-se que 8 idosos que participaram do programa sofreram uma queda, enquanto 10 idosos que não participaram do programa sofreram uma queda. Há evidências de que a participação no programa tem associação com a incidência de quedas neste intervalo de seis meses?

2. Uma fonoaudióloga está investigando a possível associação entre o tabagismo e um determinado distúrbio de voz. Os achados deste estudo são representados na tabela a seguir:

Tabagismo	Distúrbio de voz	
	Portador	Não portador
Presente	8	32
Ausente	5	70
<b>Total</b>	<b>13</b>	<b>102</b>

Há alguma evidência de que o tabagismo é associado ao distúrbio de voz?

3. Um estudo objetivou avaliar se profissionais de saúde têm um risco de infarto agudo do miocárdio (IAM) diferente daquele observado em outros profissionais. Em uma amostra de 90 trabalhadores, foi observado que 12 haviam apresentado pelo menos um episódio de IAM. Dentre estes 12 indivíduos, 8 eram profissionais de saúde, conforme tabela a seguir.

Profissionais de saúde	Infarto agudo do miocárdio	
	Sim	Não
Sim	8	52
Não	4	26
<b>Total</b>	<b>12</b>	<b>78</b>

Há alguma evidência de que o IAM tem alguma associação com o fato do indivíduo ser ou não um profissional de saúde?

4. Uma pesquisa teve por objetivo investigar a associação entre hábitos de vida e déficit cognitivo em indivíduos com 65 anos ou mais. Em uma amostra de 80 indivíduos, verificou-se que 31 eram praticantes de atividades físicas e 33 possuíam déficit cognitivo. Dentre os praticantes de atividades físicas, 9 possuíam déficit cognitivo. Há alguma evidência de que a prática de atividade física é associada ao déficit cognitivo?

5. Para estudar o estado nutricional e aspectos do estilo de vida de vegetarianos e onívoros residentes em um município, na faixa etária de 35 a 64 anos de idade, uma pesquisa utilizou uma amostra de 70 pessoas. Após uma entrevista, observou-se que dentre estas pessoas, 30 eram vegetarianas. Uma das variáveis de interesse do estudo é a razão entre a circunferência da cintura e do quadril (RCQ), sendo a medida das circunferências realizadas com o auxílio de uma fita métrica inextensível. A RCQ foi considerada inadequada quando maior que 0,80 em mulheres e maior que 0,95 em homens. Um total de 24 pessoas da amostra teve a RCQ classificada como inadequada. Testar a associação entre o tipo de alimentação (dieta vegetariana ou onívora) e a classificação da RCQ (adequada ou inadequada), considerando que 6 pessoas desta amostra são vegetarianas e têm a RCQ inadequada. Considere  $\alpha = 5\%$ .

6. Uma fonoaudióloga deseja estudar a relação entre o bruxismo (hábito de ranger os dentes) e perdas auditivas. Em um grupo de 140 pessoas, ela observou que 57 eram portadoras de perdas auditivas e 20 portavam de bruxismo. E ainda, observou que 5 pessoas eram portadoras de bruxismo mas **não** portadoras de perdas auditivas. Testar a associação entre o bruxismo e perdas auditivas, considerando um nível de significância de 5%. Escreva uma interpretação para o resultado encontrado.

7. Foi tomada uma amostra aleatória de 100 crianças nascidas no ano de 2010, em um dado município. Observou-se que 20 destas crianças eram filhas de mães adolescentes e 30 nasceram com peso inferior a 2.500g (ou seja, baixo peso). Dado que, destas 100 crianças, 59 tiveram peso ao nascer superior a 2.500g e eram filhas de mães adultas, use o teste qui-quadrado para testar a associação entre o peso ao nascer (classificado em baixo peso ou não) e a idade materna (classificada em adolescente ou adulta). Considere um nível de significância de 5%.

8. Para investigar a associação entre perda auditiva e diabetes mellitus (DM), foi conduzido um estudo com indivíduos adultos submetidos a uma avaliação audiológica em uma clínica especializada. Com base na anamnese e através de audiometria tonal limiar, estes indivíduos foram classificados como portadores ou não de perda auditiva. Foram selecionados 36 indivíduos portadores de DM e 36 indivíduos não portadores de DM. Observou-se que, dentre os indivíduos sem perda auditiva, 11 eram portadores de diabetes e 4 não possuíam esta doença. Considere um nível de significância de 5% e use o teste qui-quadrado para testar a associação entre perda auditiva e DM. Escreva uma conclusão para o resultado do teste.

9. Uma pesquisa teve por hipótese que idosos portadores de sintomas depressivos tinham mais propensão a quedas. Em uma amostra de 120 pacientes idosos ambulatoriais, 40% relataram ao menos uma queda nos últimos 12 meses e 20% apresentavam sintomas depressivos. Observou-se também que 14 destes 120 pacientes simultaneamente apresentavam sintomas depressivos e relataram uma queda ou mais nos últimos 12 meses. Utilizando o teste qui-quadrado e um nível de significância de 5%, avalie se há evidências de associação entre sintomas depressivos e propensão a quedas na respectiva população.

10. A osteoartrite (OA) é uma doença articular crônico-degenerativa que evidencia desgaste da cartilagem articular. Dentre as articulações de sustentação de peso, o joelho é o mais frequentemente afetado. Participaram de um estudo 15 voluntários com diagnóstico clínico e radiográfico de OA bilateral de joelho, e outros 15 voluntários sem OA. Dentre os 15 voluntários com diagnóstico de OA, sete eram sedentários (e os demais, ativos); e dentre os 15 voluntários sem diagnóstico de OA, quatro eram sedentários. Organize estes dados em uma tabela de contingência “2x2”, identificando os indivíduos ativos e sedentários, e os portadores e não portadores de OA. Com base nesta tabela, utilize o teste qui-quadrado de Pearson para testar a associação entre atividade física (sedentários/ativos) e a osteoartrite (portadores/não portadores). Utilize um nível de significância de 0,05. Não deixe de escrever uma interpretação para o resultado encontrado, no contexto do problema apresentado.

11. Na maioria dos casos de crianças com cardiopatias congênitas, é necessário tratamento cirúrgico com correção total ou paliativa. Um estudo incluiu 140 pacientes de zero a seis anos com cardiopatias congênitas, submetidos à cirurgia cardíaca. Foi feito um sorteio para determinar se cada paciente seria alocado em um grupo de intervenção (G1), que realizou fisioterapia pré e pós-operatória, ou para um grupo controle (G2), que realizou somente fisioterapia pós-operatória. Neste sorteio, 70 pacientes foram alocados no grupo G1 e a outra metade no grupo G2. Após a cirurgia, observou-se que 17 pacientes do grupo G1 e 29 pacientes do grupo G2 apresentaram complicação pulmonar.

- (a) Utilize um teste de hipóteses adequado para testar se há associação entre a complicação pulmonar (teve ou não teve) e o tipo de intervenção (G1 ou G2) utilizada. Considere um nível de significância de 0,05.
- (b) Escreva uma interpretação correta para o resultado encontrado no item (a), de acordo com o contexto do problema apresentado.

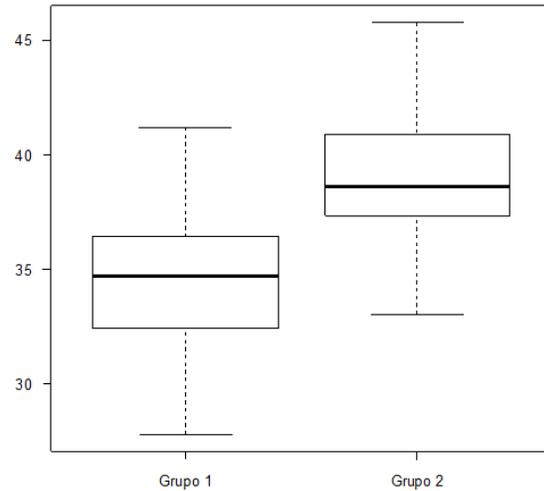
**Respostas:**

1.  $\chi^2 = 4,98$  (rejeita-se  $H_0$  dado que  $\chi^2 > 3,8415$ ).
2.  $\chi^2 = 4,62$  ( $\chi^2 > 3,8415$ , há evidências).
3.  $\chi^2 = 0$ , não há evidências.
4.  $\chi^2 = 3,12$  (não há evidências pois  $\chi^2 < 3,8415$ ).
5.  $\chi^2 = 4,76$  (rejeita-se  $H_0$  dado que  $\chi^2 > 3,8415$ ).
6.  $\chi^2 = 11,36$  (interpretar).
7.  $\chi^2 = 2,68$  (interpretar).
8.  $\chi^2 = 4,13$  (interpretar).
9.  $\chi^2 = 4,20$  (interpretar).
10.  $\chi^2 = 1,29$  (interpretar).
11.  $\chi^2 = 4,66$  (interpretar).

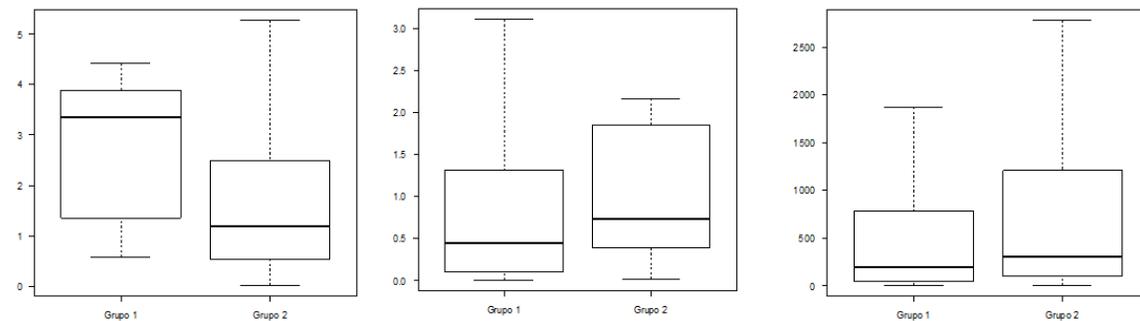
## Noções de técnicas não paramétricas

O teste  $t$  de Student para comparação entre duas médias, como vimos no Capítulo 8, pressupõe que a variável de interesse seja contínua e siga uma distribuição simétrica, aproximadamente normal, em cada um dos grupos considerados. Os *box-plots* ao lado exemplificam uma situação onde o teste de  $t$  de Student pode ser bem utilizado, dado que a distribuição dos dados é simétrica em ambos grupos.

Entretanto, em algumas situações, podemos ter dificuldades em utilizar o teste  $t$  de Student para comparação entre duas médias populacionais.



Os *box-plots* abaixo são exemplos destas situações. As distribuições dos dados são demasiadamente assimétricas, o que inviabiliza o uso do teste  $t$  de Student.



Por isso, alguns autores usam as chamadas **técnicas não paramétricas**, ou seja, ferramentas estatísticas que não fazem pressuposições sobre a distribuição de probabilidade dos dados.

Muitas **técnicas não paramétricas** são baseadas em *postos*. Isso significa que não utilizam em seus cálculos os valores observados da variável, mas sim a posição que cada valor ocupa em uma sequência ordenada.

Por exemplo, digamos que em uma amostra tamanho  $n = 10$ , encontramos as seguintes observações para a altura dos indivíduos (em centímetros):

164    184    165    180    181    159    168    167    169    170

O posto para o menor valor é 1, o posto para o segundo menor valor é 2, e assim por diante. Assim, os postos para os dados acima são respectivamente:

2        10        3        8        9        1        5        4        6        7

Quando há duas ou mais observações iguais em nossos dados, dizemos que ocorreu um *empate*. Por exemplo, imagine que nossos dados são estes:

164    184    165    180    181    159    168    165    169    170

Temos então, dois indivíduos com 165 cm de altura. Neste caso, estes indivíduos estão “empatados” em terceiro e quarto lugar. As posições das observações em uma sequência ordenada seriam, respectivamente:

2°	10°	3° e 4°	8°	9°	1°	5°	3° e 4°	6°	7°
----	-----	---------	----	----	----	----	---------	----	----

Para resolvermos o problema dos empates, nós consideramos os postos como uma média entre as posições. Assim, os postos para estes dados são:

2        10        3,5        8        9        1        5        3,5        6        7

Se tivéssemos três ou mais indivíduos “empatados”, resolveríamos o problema dos empates da mesma maneira. Por exemplo, sejam os dados:

164    184    165    180    181    159    165    165    169    170

As posições das observações em uma sequência ordenada seriam, respectivamente:

2°	10°	3°,4° e 5°	8°	9°	1°	3°,4° e 5°	3°,4° e 5°	6°	7°
----	-----	------------	----	----	----	------------	------------	----	----

Portanto, os postos para estes dados são:

2        10        4        8        9        1        4        4        6        7

### Teste de Mann-Whitney

Um teste de hipóteses não paramétrico bastante comum é o *teste de Mann-Whitney*. Este teste tem por objetivo comparar dois grupos independentes, mas **não compara médias populacionais**, como o teste *t* de Student.

Se obtivermos duas amostras distintas de indivíduos, o teste de Mann-Whitney tem por hipóteses:

$H_0$ : as amostras são provenientes de uma única população

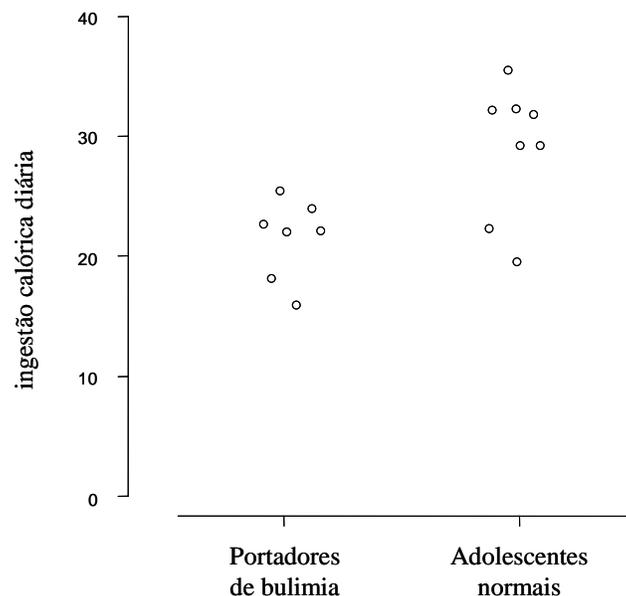
$H_A$ : as amostras são provenientes de populações diferentes

**Exemplo:** seja um estudo que objetiva comparar a ingestão calórica diária de adolescentes portadores de bulimia e adolescentes normais.

Portadores de bulimia: 22,7 22,2 24,0 18,2 25,5 22,1 e 16,0.

Adolescentes normais: 31,8 35,5 19,6 29,3 32,2 32,3 22,3 e 29,3.

A hipótese nula diz que portadores de bulimia e adolescentes normais são iguais em relação à ingestão calórica diária, ou seja, não são provenientes de populações diferentes em relação a esta característica. O gráfico abaixo ilustra estes dados:



O primeiro passo consiste em encontrar os postos para os dados (lembrando-se da possibilidade de empates), identificando quem pertence a cada um dos dois grupos. Podemos fazer isso, por exemplo, montando uma tabela como a que aparece a seguir.

Ingestão calórica diária (kcal/kg)	Postos para os portadores de bulimia	Postos para os não portado- res de bulimia
16,0	1	
18,2	2	
19,6		3
22,1	4	
22,2	5	
22,3		6
22,7	7	
24,0	8	
25,5	9	
29,3		10,5
29,3		10,5
31,8		12
32,2		13
32,3		14
35,5		15
Total	$W_1 = 36$	$W_2 = 84$

No segundo passo, encontramos  $W_1$  e  $W_2$ , as somas dos postos em cada um dos grupos. Temos  $W_1 = 36$  e  $W_2 = 84$ . Então definimos  $W$  como o menor valor entre  $W_1$  e  $W_2$ . Assim,  $W = 36$ , dado que  $36 < 84$ .

No terceiro passo, algumas outras quantidades são definidas:

- $n_1$  : é o tamanho amostral do grupo com menor soma de postos
- $n_2$  : é o tamanho amostral do grupo com maior soma de postos

Portanto,  $n_1 = 7$  indivíduos e  $n_2 = 8$  indivíduos.

Encontramos a quantidade  $m$ , definida por  $m = \frac{n_1(n_1 + n_2 + 1)}{2}$ .

$m$  é o valor médio que esperaríamos encontrar para  $W$  quando a hipótese nula é verdadeira. No nosso exemplo, temos  $m = \frac{7(7 + 8 + 1)}{2} = 56$ .

Encontramos também a quantidade  $v$ , definida por  $v = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$ .

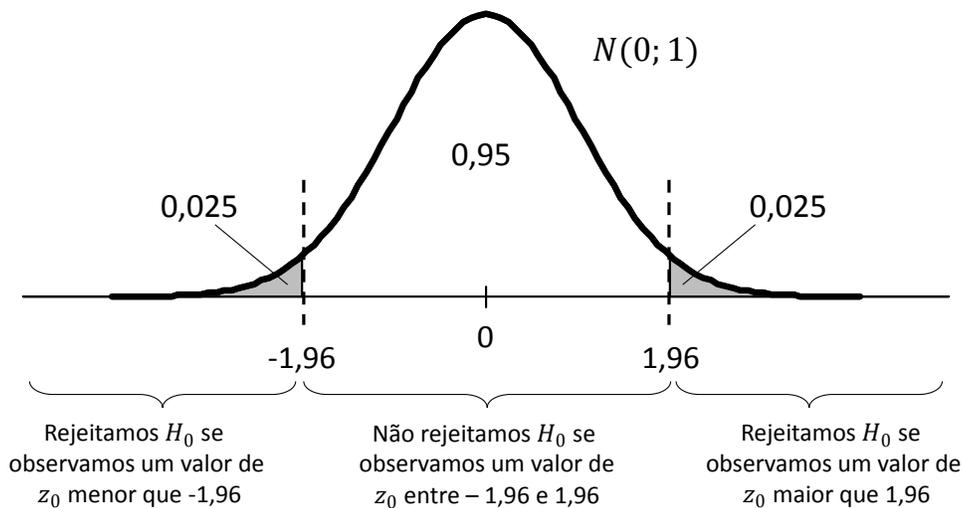
No nosso exemplo, temos  $v = \sqrt{\frac{7 \times 8(7 + 8 + 1)}{12}} = 8,64$ .

No passo seguinte, encontramos a estatística de teste

$$z_0 = \frac{W - m}{v}.$$

Sob o pressuposto de que a hipótese nula é verdadeira, a distribuição dos resultados de  $z_0$ , baseados em amostras aleatórias de  $n = n_1 + n_2$  indivíduos segue uma distribuição normal padrão.

Assim, considerando um nível de significância de 5%, temos:



No nosso exemplo, temos  $z_0 = \frac{W - m}{v} = \frac{36 - 56}{8,64} = -2,31$ .

Portanto: \_\_\_\_\_

### Notas:

- A expressão “não paramétrico” refere-se ao método estatístico, e não ao tipo ou a forma dos dados observados.
- Portanto, existem “testes não paramétricos”, mas a expressão “*dados não paramétricos*” é inadequada. Não são os dados que são paramétricos ou não paramétricos, mas o método empregado para sua análise.
- Existem outros testes não paramétricos: teste de Wilcoxon para dados pareados, teste de Kruskal-Wallis (para comparações entre 3 grupos ou mais), teste de Friedman, etc.

**Exercícios**

1. Uma nutricionista conduziu uma pesquisa com o objetivo de avaliar o consumo alimentar de gestantes adolescentes de um município. Foram selecionadas cinco adolescentes que receberam orientações sobre alimentação no pré-natal, e outras cinco que não receberam nenhuma orientação. Os dados que se encontram a seguir referem-se ao consumo de cálcio (mg) observado.

Receberam orientações	1105	613	689	772	1251
Não receberam orientações	395	784	789	457	1045

Use o teste de Mann-Whitney para avaliar se há evidências de que transmitir orientações sobre alimentação no pré-natal traz alguma modificação no consumo de cálcio de adolescentes gestantes.

2. Um estudo teve por objetivo avaliar se indivíduos com epilepsia controlada (mais de 150 dias sem crise) e não controlada possuem ou não os mesmos indicadores de qualidade de vida. Para isso, um pesquisador usou um instrumento apropriado, que mede a qualidade de vida (QV) em um escore de 0 a 100, tal que, quanto maior o valor, melhor a QV do indivíduo. A tabela a seguir mostra os dados de QV obtidos de uma amostra, onde o Grupo 1 refere-se aos indivíduos com epilepsia controlada e o Grupo 2 aos indivíduos com epilepsia não controlada.

Grupo 1	100	97	93	100	90	85	92	98
Grupo 2	69	70	88	90	91	75	77	

Use o teste de Mann-Whitney para avaliar se há evidências de que o controle da epilepsia traz algum efeito sobre a qualidade de vida.

3. Participaram de um estudo 14 sujeitos de sexo masculino, sendo 8 coralistas de coros evangélicos amadores e 6 que nunca realizaram a prática do canto coral. Ao avaliar as vozes dos sujeitos, contou-se o número total de semitons atingidos, que estão listados a seguir.

Coralistas	27	33	39	31	32	27	38	37
Não coralistas	29	29	31	25	24	24		

Use o teste de Mann-Whitney para avaliar se coralistas e não coralistas distinguem-se ou não em relação ao número de semitons atingidos (considere  $\alpha = 5\%$ ). Não deixe de escrever uma **interpretação** para o resultado do teste.

4. Os dados a seguir são relativos ao tempo de atividade física (em minutos por semana) de 10 adolescentes de sexo masculino e sete adolescentes de sexo feminino. Estas amostras foram tomadas da população de escolares de uma cidade do interior paulista.

Masculino	203	320	369	417	499	553	733	969	1243	1341
Feminino	158	175	341	480	592	615	1216			

Use o teste de Mann-Whitney e responda se há evidências de que as distribuições dos tempos de atividade física, nesta população, são diferentes entre os adolescentes de sexo masculino e feminino. Considere um nível de significância de 5%.

5. Uma pesquisadora da área de fonoaudiologia conduziu uma avaliação otorrinolaringológica e análises perceptivo-auditiva e acústica computadorizada das vozes de 18 mulheres adultas jovens, sem queixas vocais. Uma das variáveis de interesse é a frequência fundamental ( $f_0$ ), que corresponde ao número de ciclos que as pregas vocais fazem em um segundo. A seguir, são listados valores da  $f_0$  obtidos para mulheres fumantes e não fumantes.

Fumantes	160	177	184	186	190	190	199	204	206	228
Não fumantes	189	190	208	215	226	228	228	236	236	276

(a) Use um teste não-paramétrico adequado e, a partir de seus resultados, responda se há (ou não) evidências de que os valores de  $f_0$  são diferentes entre as populações de mulheres adultas jovens fumantes e não fumantes. Considere um nível de significância de 5%.

(b) No contexto deste problema, escreva o que seria cometer erros tipo I e II.

### Respostas

1.  $W_1 = 31$ ,  $W_2 = 24$ ,  $n_1 = n_2 = 5$ ,  $W = 24$ ,  $m = 27,5$ ,  $v = 4,79$ ,  $z_0 = -0,73$  (interpretar)
2.  $W_1 = 87,5$ ,  $W_2 = 32,5$ ,  $n_1 = 7$ ,  $n_2 = 8$ ,  $W = 32,5$ ,  $m = 56$ ,  $v = 8,64$ ,  $z_0 = -2,72$  (interpretar)
3.  $W_1 = 77,5$ ,  $W_2 = 27,5$ ,  $n_1 = 6$ ,  $n_2 = 8$ ,  $W = 27,5$ ,  $m = 45$ ,  $v = 7,75$ ,  $z_0 = -2,26$  (interpretar)
4.  $W_1 = 99$ ,  $W_2 = 54$ ,  $n_1 = 7$ ,  $n_2 = 10$ ,  $W = 54$ ,  $m = 63$ ,  $v = 10,25$ ,  $z_0 = -0,878$  (interpretar)
5. (a)  $W_1 = 69,5$ ,  $W_2 = 101,5$ ,  $n_1 = 10$ ,  $n_2 = 8$ ,  $W = 69,5$ ,  $m = 95$ ,  $v = 11,26$ ,  $z_0 = -2,265$  (interpretar)

(b) Cometer um erro tipo I significa concluir que as populações de mulheres adultas jovens fumantes e não fumantes são diferentes em relação à distribuição da variável frequência fundamental ( $f_0$ ), quando, na realidade, não há diferença. O erro tipo II é cometido se concluído que não há evidências de diferenças entre as populações de mulheres adultas jovens fumantes e não fumantes em relação à distribuição da variável frequência fundamental ( $f_0$ ), quando há na realidade alguma diferença.