

# **Introdução à Bioestatística**

**Faculdade de Saúde Pública  
Programa de Verão - 2005**

**Autores: Denise Pimentel Bergamaschi  
José Maria Pacheco de Souza  
Patrícia Emília Braga**

Obra protegida por direitos autorais. Proibida a reprodução total ou parcial sem a prévia autorização dos autores. Em caso de utilização é obrigatória a citação da fonte no texto, com respectiva referência bibliográfica.

Sugestão de referência: Bergamaschi DP, Souza JMP, Braga PE. **Introdução à Bioestatística**. São Paulo, 2005. [Material de apoio Didático de Curso de Atualização do Programa de Verão da Faculdade de Saúde Pública da Universidade de São Paulo]

## Índice

	página
<b>1. Elementos estatísticos para análise exploratória de dados</b>	3
1.1- Níveis de mensuração	3
1.2- Tópicos iniciais de amostragem	4
1.3- Apuração de dados	6
1.4- Apresentação tabular	7
1.5- Apresentação gráfica	12
1.5.1 – Diagrama de barras	12
1.5.2 – Diagrama de setores circulares	14
1.5.3 – Diagrama linear	15
1.5.4 - Histograma	15
1.5.5 – Polígono de frequência simples	17
1.5.6 – Representação gráfica de duas variáveis qualitativas	18
1.5.7 – Representação gráfica de duas variáveis quantitativas	19
1.5.8 – Escalas aritmética e logarítmica	21
1.6- Medidas de tendência central	26
1.6.1 – Média aritmética	26
1.6.2 - Mediana	27
1.7- Medidas de dispersão	28
1.7.1 – Valores mínimo e máximo	28
1.7.2 – Amplitude de variação	28
1.7.3 - Variância	28
1.7.4 – Desvio padrão	28
1.7.5 – Coeficiente de Variação de Pearson	28
1.7.6 – Quartil, percentil	30
1.8 – Box plot (Stata)	32
1.9 - Noções de correlação	35
<b>2. Medidas de associação</b>	38
2.1 – Razão de prevalências	38
2.2 – Razão de incidências	39
2.3 – Razão de odds (odds ratio)	39
2.4 – Qui quadrado de Pearson	40
<b>3. Distribuição de probabilidade</b>	44
3.1- Noções de probabilidade	44
3.2 – Distribuição Bernoulli	47
3.3 - Distribuição binomial	48
3.4 – Odds e probabilidade	52
3.5 - Distribuição normal	53
3.6 - Distribuição amostral da média	56
<b>4. Estimação de parâmetros populacionais por ponto e por intervalo</b>	57
4.1 – Estimação por ponto	57
4.2 – Estimação por intervalo – Intervalo de Confiança (IC)	58
4.2.1 - IC para a média populacional com variância populacional conhecida	59
4.2.2 - IC para a média populacional com variância populacional desconhecida	59
4.2.3 - IC aproximado para a proporção populacional	59
<b>5. Tomada de decisão</b>	62
5.1- Teste de hipóteses para uma proporção populacional	68
5.2- Teste de hipóteses para uma média populacional	71
5.2.1 - Revisão de conceitos básicos	71
5.2.2 - Teste de hipóteses para uma média populacional com variância conhecida	73
5.2.3 - Teste de hipóteses para uma média populacional com variância desconhecida	75
5.3- Teste de hipóteses de associação pelo qui-quadrado de <i>Pearson</i>	76
<b>6. Bibliografia</b>	79
<b>7. Resumos dos artigos utilizados nos exercícios</b>	81

# População, amostra, variável, coleta e apuração de dados

## 1- Elementos estatísticos para análise exploratória de dados

Estatística: é uma coleção de métodos para planejar experimentos, obter e organizar dados, resumi-los, analisá-los, interpretá-los e deles extrair conclusões (Triola, 1999).

Bioestatística – Estatística aplicada às ciências da vida.

### 1.1 - Níveis de mensuração

#### Escala nominal

Os indivíduos são classificados em categorias segundo uma característica.

Ex: sexo (masculino, feminino),  
hábito de fumar (fumante, não fumante),  
sobrepeso (sim, não).

Não existe ordem entre as categorias e suas representações, se numéricas, são destituídas de significado numérico.

Ex: sexo masculino=1, sexo feminino = 2.  
Os valores 1 e 2 são apenas rótulos.

#### Escala ordinal

Os indivíduos são classificados em categorias que possuem algum tipo inerente de ordem. Neste caso, uma categoria pode ser "maior" ou "menor" do que outra.

Ex: nível sócio-econômico (A, B, C e D; onde A representa maior poder aquisitivo);  
nível de retinol sérico (alto, aceitável, baixo, deficiente) onde alto: maior ou igual a 50,0 µg/dl;  
aceitável: 20,0 a 49,9 µg/dl, baixo: 10,0 a 19,9 µg/dl e deficiente: menor ou igual a 10,0 µg/dl.  
Estes critérios são do *Committee on Nutrition for National Defense ICNND/USA*, 1963 (in Prado MS *et al*, 1995).

Embora exista ordem entre as categorias, a diferença entre categorias adjacentes não tem o mesmo significado em toda a escala.

Escalas numéricas

#### Escala intervalar

Este nível de mensuração possui um valor zero arbitrário.

Ex: temperatura em graus Celsius.

Escala de razões – possui zero verdadeiro

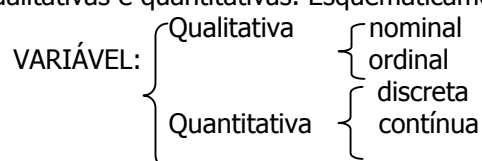
Escala de razões discreta: o resultado numérico da mensuração é um valor inteiro.

Ex: número de refeições em um dia (nenhuma, uma, duas, três, quatro, ...),  
frequência de consumo semanal de determinado alimento (1 vez, 2 vezes, 3 vezes, 4 vezes, 5 vezes, 6 vezes, 7 vezes) .

Escala de razões contínua: o resultado numérico é um valor pertencente ao conjunto dos números reais  $R = \{-\infty; \dots; -38; \dots; -0,3; \dots; 0; \dots; 0,2; \dots; 0,73; \dots; 1; \dots; 2,48; \dots; +\infty\}$ .

Ex: idade (anos), peso (g), altura (cm), nível de retinol sérico (µg/dl), circunferência da cintura (cm)

De acordo com o nível de mensuração, os fenômenos ao serem medidos, resultam em variáveis que podem ser classificadas em qualitativas e quantitativas. Esquemáticamente tem-se:



O tipo da variável irá indicar a melhor forma para o dado ser apresentado em tabelas e gráficos, em medidas de resumo e, a análise estatística mais adequada.

## 1.2 - Tópicos iniciais de amostragem

**Coleta de dados**

É a observação e registro da categoria ou medida de variáveis relacionadas ao objeto de estudo que ocorrem em unidades (indivíduos) de uma amostra ou população.

**População:** totalidade de elementos sob estudo. Apresentam uma ou mais características em comum.

Na prática, a população de estudo nem sempre corresponde à população alvo. Por exemplo, no estudo sobre a ocorrência de sobrepeso em crianças de 7 a 12 anos no Município de São Paulo, no ano de 2003, a população alvo é o conjunto de todas as crianças nesta faixa etária, deste município, neste ano. Porém, pode ser que algumas crianças estejam ausentes no momento da entrevista (crianças que estão internadas, que trabalham, que não permanecem no domicílio durante o dia ou estão viajando) fazendo com que a população de estudo represente um subconjunto da população alvo.

**Elementos:** são unidades de análise; podem ser pessoas, domicílios, escolas, creches, células ou qualquer outra unidade.

**Amostra:** é uma parte da população de estudo.

**Amostragem:** processo para obtenção de uma amostra. Tem como objetivo estimar parâmetros populacionais

**Parâmetro:** Quantidade fixa de uma população

Ex: peso médio ao nascer de crianças que nascem no município de São Paulo ( $\mu = 3100$  g);

proporção de crianças de 7 a 12 anos classificadas como obesas, no município de São Paulo ( $\pi = 12\%$ ).

**Estimador:** é uma fórmula matemática que permite calcular um valor (estimador por ponto) ou com um conjunto de valores (estimador por intervalo) para um parâmetro.

Ex: Média aritmética:  $\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$ ,

onde  $\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$  e  $n =$  número de observações.

**Estimativa:** Valor do estimador calculado em uma amostra. Estima o valor do parâmetro.

Ex: Peso médio ao nascer, calculado em uma amostra de 120.000 crianças nascidas no Município de São Paulo no ano de 2000: média amostral  $\bar{x} = 3000$ g

### Tipos de Amostragem

**Probabilística:** cada unidade amostral tem probabilidade conhecida e diferente de zero de pertencer à amostra. É usada alguma forma de sorteio para a obtenção da amostra

**Não probabilística:** não se conhece a probabilidade de cada unidade amostral pertencer à amostra. Algumas unidades terão probabilidade zero de pertencer à amostra.

Ex: amostragem intencional; por voluntários; acesso mais fácil; por quotas.

### Tipos de amostragem probabilística

- aleatória simples (com e sem reposição)
- sistemática
- com partilha proporcional ao tamanho do estrato
- por conglomerado

### Amostragem aleatória simples (AAS)

É o processo de amostragem onde qualquer subconjunto de  $n$  elementos diferentes de uma população de  $N$  elementos tem mesma probabilidade de ser sorteado (Kalton G. 1983, Silva, NN, 1998). Tamanho da população:  $N$ ; tamanho da amostra:  $n$ ; fração global de amostragem ou probabilidade de sortear um

$$\text{indivíduo} = \frac{n}{N}$$

- É necessário ter um sistema de referência que contenha todos os elementos da população da qual será retirada a amostra.
- Utilização da tabela de números aleatórios - mecânica
- Utilização de programas computacionais

### Amostragem sistemática

Utiliza-se a ordenação natural dos elementos da população (prontuários, casa, ordem de nascimento).

- Intervalo de amostragem  $k = \frac{N}{n}$ , onde  $N$ = tamanho da população e  $n$  = tamanho da amostra

- Início casual  $i$ , sorteado entre 1 e  $k$ , inclusive

- Amostra sorteada é composta pelos elementos:  $i, i+k, i+2k, \dots, i+(n-1)k$

OBS: É necessário ter cuidado com a periodicidade dos dados, por exemplo se for feito sorteio de dia no mês, pode cair sempre em um domingo onde o padrão de ocorrência do evento pode ser diferente.

Exemplo:  $N=80$ ;  $n=10$ ;  $k = \frac{N}{n} = \frac{80}{10} = 8$ ;

início casual:  $1 \leq i \leq 8$

Começo casual **sorteado**:  $i=4$

Amostra composta dos elementos:

$i$	4
$i+k$	12
$i+2k$	20
$\vdots$	$\vdots$
$\vdots$	$\vdots$
$\vdots$	$\vdots$
$i+(n-1)k$	76

Se o intervalo de amostragem não for inteiro proceder da seguinte forma:

$$N= 321 ; n=154; K = \frac{N}{n} = \frac{321}{154} = 2,084$$

$i$  deve ser um número sorteado entre 1 e 2,084

Sortear um número entre 1000 e 2084 e dividir o resultado por 1000

Número sorteado = 1941, portanto  $i=1,941$

Indivíduos:

		<u>elemento</u>
I	1,941	1
i+k	1,941+2,084 = 4,025	4
i+2k	1,941+4,1680 = 6,109	6
I+3k	1,941+6,252 = 8,193	8
.	.	
.	.	
.	.	
i+(n-1)k	1,941+318,852 = 320,793	320

#### Amostragem por conglomerado

É o processo no qual os elementos da população são reunidos em grupos que constituem a unidade amostral e, por sua vez, alguns destes são sorteados para comporem a amostra. Se o interesse residir no sorteio de escolares, em um processo de amostragem por conglomerados, seria possível sortear escolas (unidade amostral) e considerar todos os alunos destas para comporem a amostra (Silva, 1998).

### 1.3 - Apuração de dados

Processo no qual conta-se o número de vezes que a variável assumiu um determinado valor (frequência de ocorrência). Pode ser manual, mecânica ou eletrônica (programas estatísticos: Epi info, Stata, Excel, SPSS, SAS, R, S-Plus)

Distribuição de frequências - correspondência entre categorias (valores) e frequência de ocorrência.

Distribuição de frequências  $\left\{ \begin{array}{l} \text{valores pontuais} \\ \text{valores em intervalos de classe} \end{array} \right.$

Notação:

X : variável

$x_i$  : valor observado para o indivíduo i

#### **Apresentação pontual**

Ex: 10 indivíduos

X: número de refeições diárias

x: 2, 3, 3, 1, 5, 2, 3, 2, 3

#### **Apuração:**

<u>número de refeições</u>	<u>frequência absoluta</u>
1	1
2	3
3	4
5	1

X: idade (anos inteiros)

x: 5, 5, 15, 20, 20, 20, 21, 21, 22, 22

<u>idade</u>	<u>frequência</u>
5	2
15	1
20	3
21	2
22	2

X: peso ao nascer em gramas

X: 2250, 3025, 1600, 2725, 3750, 3950, 2400, 2180, 2520, 2530

<u>peso</u>	<u>freqüência</u>
1600	1
2180	1
2250	1
2400	1
2520	1
2530	1
2725	1
3025	1
3750	1
3950	1

Altura em metro

X: 1,63; 1,60; 1,59; 1,60; 1,45; 1,73; 2,05; 1,85

<u>altura</u>	<u>freqüência</u>
1,45	1
1,59	1
1,60	2
1,63	1
1,73	1
1,85	1
2,05	1

## 1.4 - Apresentação tabular

Elementos essenciais: título, corpo, cabeçalho e coluna indicadora.

Tabela 1 - Título: o que (natureza do fato estudado)? como (variáveis)? onde? quando?

<u>Variável</u>	<u>n°</u>	<u>%</u>
<hr/>		
<hr/>		
Total		

Fonte

notas, chamadas

OBS: nenhuma casela (intersecção entre linha e coluna) deve ficar em branco.

A tabela deve ser uniforme quanto ao número de casas decimais e conter os símbolos – ou **0** quando o valor numérico é nulo e ... quando não se dispõe do dado.

### Exemplo:

Distribuição de Municípios brasileiros segundo prevalência de bócio em escolares de 6 a 14 anos. Inquérito Nacional, Brasil, 1994 a 1996.

<u>Prevalência de bócio (%)</u>	<u>Freqüência</u>	<u>%</u>
< 5,0 (não endêmico)	326	76,18
5,0 a 19,9 (leve)	83	19,39
20,0 a 29,9 (moderado)	15	3,50
≥30 (grave)	4	0,93
Total	428	100

Fonte: Corrêa Filho HR *et al.*, 2002 (adaptado).

Como idade é variável quantitativa contínua, a melhor forma de apresentá-la em tabelas é utilizando intervalos de valores denominados intervalos de classe.

**Exemplo:**

x: 5, 5, 15, 20, 20, 20, 21, 21, 22, 22

Idade	freqüência	%
5  -- 10	2	20
10  -- 15	0	-
15  -- 20	1	10
20  -- 25	7	70
Total	10	100

Intervalos de classe: conjunto de observações contidas entre dois valores limite (limite inferior e limite superior)

**Representação:**

5   -- 10	intervalo fechado no limite inferior e aberto no limite superior (contém o valor 5 mas não contém o valor 10)
5 -- 10	intervalo aberto nos limites inferior e superior (não contém os valores 5 e 10)
5  --   10	intervalo fechado nos limites inferior e superior (contém os valores 5 e 10)

OBS: Representar o intervalo 0 |-- | 11 meses é equivalente a representá-lo como 0 |-- 12 meses

X: 2250, 3025, 1600, 2725, 3750, 3950, 2400, 2180, 2520, 2530

Peso	freqüência	%
1500 --2000	1	10
2000 --2500	3	30
2500 --3000	3	30
3000 --3500	1	10
3500 --4000	2	20
Total	10	100

Considerando-se a altura (m) de 7 indivíduos

X: 1,63; 1,60; 1,59; 1,60; 1,45; 1,73; 2,05; 1,85

Altura	freqüência	%
1,45 --1,55	1	14,3
1,55 --1,65	3	42,8
1,65 --1,75	1	14,3
1,75 --1,85	0	-
1,85 --1,95	1	14,3
1,95 --2,05	0	-
2,05 --2,15	1	14,3
Total	7	100

Os intervalos de classe devem ser **mutuamente exclusivos** (um indivíduo não pode ser classificado em dois intervalos ao mesmo tempo) e **exaustivos** (nenhum indivíduo pode ficar sem classificação).

A **amplitude do intervalo** é o tamanho do intervalo de classe. A amplitude do intervalo e o número de intervalos dependem basicamente do problema específico e da literatura existente sobre o assunto.

O **ponto médio do intervalo** é calculado somando-se o limite inferior e limite superior, dividindo-se o resultado por dois.



## Tabela de dupla entrada ou de contingência

Valores nacionais de dosagens de iodo salino e bócio clínico em crianças de 6 a 14 anos. Inquérito nacional, Brasil, 1994 a 1996.

Grau de Bócio	Iodação salina <sup>a</sup>									
	Suficiente		Leve		Moderada		Grave		Total	
	n	%	n	%	n	%	n	%	n	%
Zero <sup>b</sup>	1967	97,5	18293	97,4	61497	95,1	15450	95,8	97207	95,7
I e II	50	2,5	493	2,6	3140	4,9	669	4,2	4352	4,3
Total	2017	100	18786	100	64637	100	16119	100	101559	100

<sup>a</sup> suficiente  $\geq$  40 mg/kg; leve 20-39 mg/kg; moderada 10-19 mg/kg; grave < 10 mg/kg.

<sup>b</sup> normal

Fonte: Corrêa Filho HR *et al.*, 2002.

**Exemplo:** Tabela retirada do artigo Cartografia do retardo estatural em escolares do estado da Paraíba. Brasil. Carvalho AT *et al. Rev Saúde Pública* 2000; 34(1):3-8.

Tabela 1 - Dados demográficos e espaciais dos escolares estudados em estabelecimentos de ensino público do Estado da Paraíba, 1991.

Especificações	Mesorregiões				
	Sertão	Borborema	Agreste	Mata	Estado
Sexo					
Masculino	7.533	2.155	7.663	5.610	22.961
Feminino	8.899	2.724	9.186	6.269	27.078
Idade					
6 anos	711	116	434	217	1.478
7 anos	3.454	925	2.865	2.045	9.289
8 anos	5.782	1.693	5.935	4.506	17.916
9 anos	6.158	1.958	6.868	4.959	19.943
Área					
Urbana	7.120	1.995	8.788	9.236	27.139
Rural	9.252	2.862	8.027	2.369	22.510
Total	16.457	4.929	16.849	11.879	50.114

Fonte dos dados básicos: FAE/PMA<sup>2</sup>(1993)

### Questão 1

Classificar quanto a natureza, as seguintes variáveis

Variável	Tipo (natureza)
condição de saúde (doente, não doente)	
tipo de parto (normal, cesário)	
nível de colesterol sérico (mg/100cc)	
tempo de um procedimento cirúrgico (minutos)	
número de praias consideradas poluídas	
custo do procedimento (reais)	

### Questão 2

Os dados a seguir são de peso (kg) de 80 mulheres identificadas pela variável id (identificação).

Id	Peso	Id	peso	Id	Peso	Id	Peso	Id	Peso	Id	Peso
1	65	16	71	31	70	46	75	61	68	76	75
2	65	17	84	32	72	47	79	62	69	77	79
3	58	18	63	33	75	48	79	63	76	78	73
4	59	19	64	34	76	49	82	64	77	79	82
5	67	20	65	35	77	50	83	65	80	80	76
6	68	21	74	36	78	51	65	66	81		
7	74	22	81	37	80	52	68	67	59		
8	81	23	66	38	82	53	75	68	64		
9	66	24	69	39	63	54	76	69	70		
10	61	25	71	40	66	55	78	70	80		
11	64	26	71	41	72	56	78	71	85		
12	65	27	72	42	72	57	81	72	70		
13	67	28	73	43	72	58	85	73	71		
14	68	29	75	44	73	59	66	74	72		
15	70	30	77	45	73	60	68	75	72		

Fonte: Osborn JF. *Statistical Exercises in Medical Research*. John Wiley & Sons Inc., 1979. (adaptado)

Sorteie uma amostra sistemática de tamanho 20. Indique o intervalo de amostragem e o começo casual sorteado. Indique o número de identificação de cada elemento da amostra.

### Questão 3

Os dados a seguir são de altura de uma amostra de 351 mulheres idosas selecionadas aleatoriamente de uma comunidade para um estudo de osteoporose. Faça uma apuração dos dados e represente-os em uma tabela. Interprete os resultados.

142	152	154	156	157	158	160	161	163	164	165	169
145	152	154	156	157	158	160	161	163	164	165	169
145	152	154	156	157	158	160	161	163	164	165	169
145	152	154	156	157	158	160	161	163	164	165	169
146	152	155	156	157	158	160	161	163	164	166	169
147	152	155	156	157	158	160	161	163	164	166	169
147	153	155	156	158	158	160	161	163	164	166	169
147	153	155	156	158	159	160	162	163	164	166	170
148	153	155	156	158	159	160	162	163	164	166	170
148	153	155	156	158	159	160	162	163	164	166	170
149	153	155	156	158	159	160	162	163	164	166	170
150	153	155	156	158	159	160	162	163	164	166	170
150	153	155	156	158	159	160	162	163	164	166	170
150	153	155	156	158	159	160	162	163	164	166	170
150	153	155	156	158	159	160	162	163	164	166	170
150	153	155	157	158	159	160	162	163	165	167	170
150	153	155	157	158	159	160	162	163	165	167	170
150	153	155	157	158	159	161	162	163	165	167	170
151	153	155	157	158	159	161	162	163	165	167	171
151	153	155	157	158	159	161	162	163	165	167	171
151	153	155	157	158	159	161	162	163	165	167	171
151	153	155	157	158	159	161	162	163	165	167	173
151	153	155	157	158	159	161	162	163	165	168	173
151	154	155	157	158	159	161	162	163	165	168	173
152	154	155	157	158	159	161	162	163	165	168	174
152	154	156	157	158	160	161	162	163	165	168	176
152	154	156	157	158	160	161	163	163	165	168	177
152	154	156	157	158	160	161	163	164	165	168	178
152	154	156	157	158	160	161	163	164	165	169	178
152	154	156									

Fonte: Hand DJ *et al.*, 1994.

### Questão 4

Os dados a seguir são relativos ao peso ao nascer (g) de recém nascidos com síndrome de desconforto idiopático grave. Algumas crianças foram a óbito (\*) e outras sobreviveram.

1050*	2500*	1890*	1760	2830
1175*	1030*	1940*	1930	1410
1230*	1100*	2200*	2015	1715
1310*	1185*	2270*	2090	1720
1500*	1225*	2440*	2600	2040
1600*	1262*	2560*	2700	2200
1720*	1295*	2730*	2950	2400
1750*	1300*	1130	2550	3160
1770*	1550*	1575	2570	3400
2275*	1820*	1680	3005	3640

Fonte: Hand DJ *et al.*, 1994.

- Classifique a variável peso ao nascer em duas categorias: baixo peso (abaixo de 2500 g) e não baixo peso (2500 g e mais) e faça uma tabela bidimensional cruzando as variáveis: condição do recém-nascido (sobrevivente ou não sobrevivente) e peso ao nascer (baixo peso e não baixo peso).
- Interprete os resultados.

### Questão 5

Os dados abaixo são de um estudo de prevalência de doença cardíaca e investigação de fatores de risco associados. Calcular os valores relativos (percentuais).

Distribuição de pacientes segundo ronco noturno e doença cardíaca

Ronco noturno	Doença cardíaca				Total	
	Sim		Não		Nº	%
	Nº	%	Nº	%	Nº	%
Não	24		1355		1379	
Ocasional	35		603		638	
Quase todas as noites	21		192		213	
Sempre	30		224		254	
Total	110		2374		2484	

Fonte: Hand DJ *et al.*, 1994.

### Questão 6

Com base nos dados da tabela abaixo,

a) Calcular o percentual de mães com dieta boa, razoável e pobre entre os casos de spina bífida e entre os controles (fixando o 100% no total de casos e de controles);

b) Analise e discuta os resultados.

Distribuição de recém-nascidos casos (acometidos de spina bífida) e controles segundo dieta da mãe.

Dieta materna	Casos	Controles	Total
	Nº	Nº	Nº
Boa	34	43	77
Razoável	110	48	158
Pobre	100	32	132
Total	244	123	367

Fonte: Hand DJ *et al.*, 1994.

### Questão 7

A tabela abaixo foi extraída do artigo Tendência secular do peso ao nascer na cidade de São Paulo (1976-1998) de MONTEIRO CA *et al.* (*Rev. Saúde Pública*; 2000:34 (6, supl): 26-40).

Comente os resultados apresentados.

**Tabela 1** – Distribuição de nascidos vivos segundo intervalos do peso ao nascer nas cidades de São Paulo, SP (Brasil) e Gotemburgo (Suécia).

Peso ao nascer (gramas)	São Paulo -1998		Gotemburgo -1972/73	
	N	%	N	%
500 – 1.000	940	0,43	29	0,24
1.001 – 1.500	1.650	0,76	45	0,37
1.501 – 2.000	3.519	1,62	113	0,93
2.001 – 2.500	13.187	6,09	368	3,02
2.501 – 3.000	53.882	24,87	1.689	13,85
3.001 – 3.500	90.151	41,60	4.351	35,69
3.501 – 4.000	44.045	20,33	3.980	32,64
4.001 – 4.500	8.225	3,80	1.356	11,12
4.501 – 5.000	967	0,45	250	2,05
5.001 – 6.000	116	0,05	11	0,09
Total	216.682	100,00	12.192	100,00

## Questão 8

A tabela abaixo foi extraída do artigo Infecção pelo HIV durante a gestação: Estudo-Sentinela Parturiente, Brasil, 2002 de Souza Júnior *et al.* (*Rev. Saúde Pública*; 2004:38(6):764-72). Comente os resultados apresentados.

**Tabela 1** - Distribuição (%) das parturientes por trimestre de início e número de consultas no pré-natal segundo Grande Região, tamanho do município de realização do parto e grau de instrução da mãe. Estudo-Sentinela Parturiente, Brasil, 2002.

Fatores considerados	N. de consultas de pré-natal				Total N	Total %	Início do pré-natal			Total N	Total %
	Nenhuma	1-3	4-5	6+			1º Trimestre	2º	3º		
Região											
Norte	6,8	32,8	32,5	27,9	3.173	100,0	48,0	43,4	8,6	2.933	100,0
Nordeste	5,0	23,2	34,3	37,5	9.333	100,0	57,7	31,6	10,7	8.720	100,0
Sudeste	4,3	14,1	25,2	56,4	11.506	100,0	60,5	31,4	8,1	10.661	100,0
Sul	2,5	15,0	21,0	61,6	4.062	100,0	62,8	29,7	7,5	3.847	100,0
Centro-Oeste	2,2	20,4	27,1	50,3	2.078	100,0	59,6	24,6	15,8	1.960	100,0
Tamanho do município (hab)											
1 - 50.000	4,0	21,4	30,4	44,3	10.448	100,0	56,9	32,1	11,0	9.893	100,0
50.001 - 200.000	3,3	20,2	29,2	47,3	7.267	100,0	59,9	30,7	9,4	6.751	100,0
200.001 - 500.000	3,6	19,5	25,5	51,4	4.201	100,0	60,1	30,9	9,0	3.902	100,0
500.001 ou mais	6,1	16,3	26,6	51,0	8.237	100,0	58,9	33,6	7,5	7.574	100,0
Grau de instrução											
Analfabetas	15,7	33,1	30,2	21,0	1.264	100,0	39,3	39,3	21,4	1.008	100,0
1ª a 4ª série	6,7	24,2	28,8	40,4	8.470	100,0	52,8	34,0	13,2	7.712	100,0
5ª a 8ª série	3,3	18,7	29,8	48,2	11.825	100,0	57,8	33,7	8,6	11.148	100,0
2º grau/ superior	1,8	13,7	25,7	58,8	8.464	100,0	67,7	26,9	5,5	8.134	100,0
Brasil	4,4	19,4	28,4	47,8	30.152	100,0	58,6	32,0	9,4	28.121*	100,0

\*704 fizeram o pré-natal mas não tinham informações sobre o início do pré-natal

## 1.5 - Apresentação gráfica

### 1.5.1 - Diagrama de barras

Características: barras separadas e bases de mesmo tamanho.

Apropriados para representar as variáveis: qualitativa nominal, ordinal e quantitativa discreta.

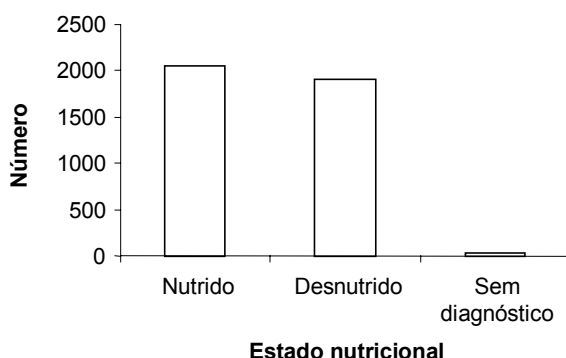
#### Variável qualitativa

O Inquérito Brasileiro de Nutrição (IBRANUTRI) foi um estudo de pacientes maiores de 18 anos, internados em hospitais da rede pública, conveniados, filantrópicos e universitários de 12 estados do Brasil e do Distrito Federal, realizado de maio a novembro de 1996 (Correia, 1997). Os dados da tabela são retirados deste estudo.

Distribuição de pacientes segundo estado nutricional. IBRANUTRI, maio a novembro, 1996

Estado nutricional	n	%
Nutrido	2061	51,5
Desnutrido	1905	47,6
Sem diagnóstico	34	0,9
Total	4000	100

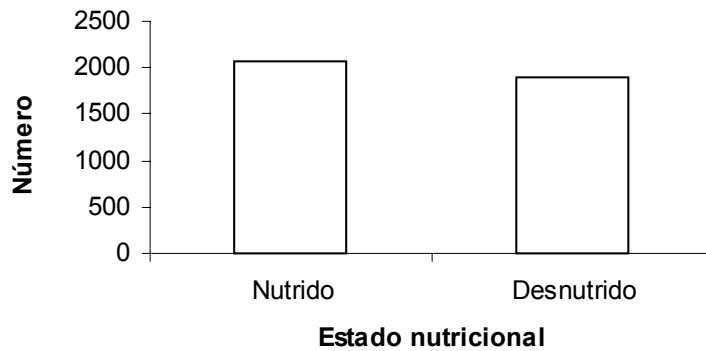
Fonte: adaptado de Soares JF *et al.*, 2002.



Fonte: adaptado de Soares JF *et al.*, 2002.

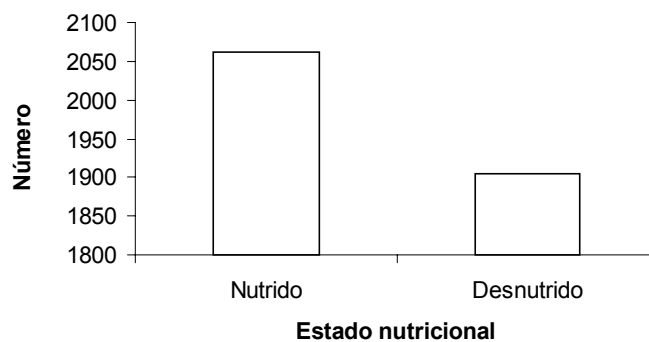
Distribuição de pacientes segundo estado nutricional. IBRANUTRI, maio a novembro, 1996

Excluindo-se os registros com informação ignorada



Fonte: adaptado de Soares JF *et al.*, 2002.  
 Distribuição de pacientes segundo estado nutricional. IBRANUTRI, maio a novembro, 1996

A representação gráfica abaixo está correta?

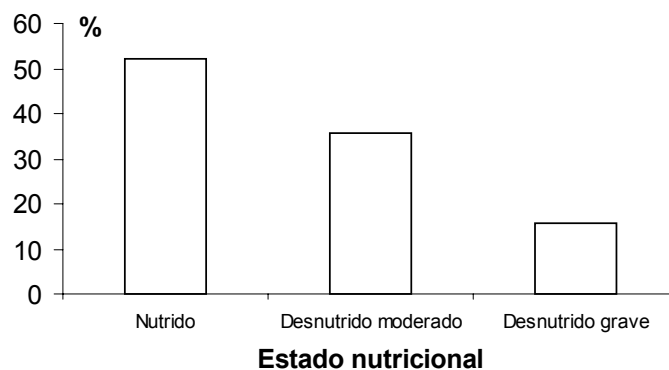


Variável qualitativa ordinal

Distribuição de pacientes segundo estado nutricional. IBRANUTRI, maio a novembro, 1996

Estado nutricional <sup>a</sup>	n	%
Nutrido	2061	52,0
Desnutrido moderado	1407	35,5
Desnutrido grave	498	12,5
Total	3966	100

<sup>a</sup> excluindo-se 34 (0,9%) de pacientes sem diagnóstico  
 Fonte: adaptado de Soares JF *et al.*, 2002.



<sup>a</sup> excluindo-se 34 (0,9%) de pacientes sem diagnóstico  
 Fonte: adaptado de Soares JF *et al.*, 2002.

Distribuição de pacientes segundo estado nutricional. IBRANUTRI, maio a novembro, 1996.

Variável quantitativa discreta

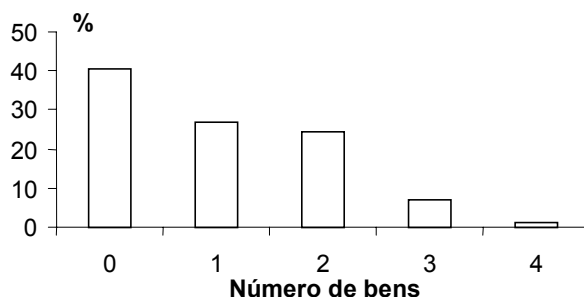
Foi realizada, no período de outubro de 1998 a outubro 1999, a pesquisa "Alimentação no primeiro ano de vida", onde estudou-se uma coorte de recém-nascidos da maternidade do Hospital Universitário (HU). Os dados a seguir são parte da caracterização sócio-econômica da amostra estudada.

Distribuição de famílias segundo número de bens\* que possuem.

Número de bens	n	%
0	146	40,6
1	97	26,9
2	87	24,2
3	26	7,2
4	4	1,1
Total	360	100

\* automóvel, telefone, TV a cabo e computador

Distribuição de famílias segundo número de bens\*



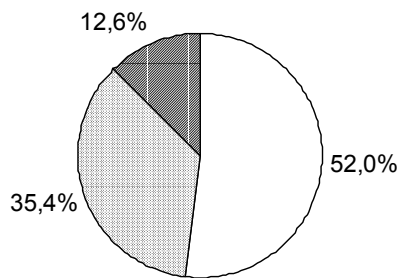
### 1.5.2 - Diagrama de setores circulares

Distribuição de pacientes segundo estado nutricional. IBRANUTRI, maio a novembro, 1996.

Estado nutricional <sup>a</sup>	n	%
Nutrido	2061	52,0
Desnutrido moderado	1407	35,4
Desnutrido grave	498	12,6
Total	3966	100

<sup>a</sup> excluindo-se 34 (0,9%) de pacientes sem diagnóstico

Fonte: adaptado de Soares JF *et al.*, 2002.



□ Nutrido    ▒ Desnutrido moderado    ■ Desnutrido grave

<sup>a</sup> excluindo-se 34 (0,9%) de pacientes sem diagnóstico

Fonte: adaptado de Soares JF *et al.*, 2002.

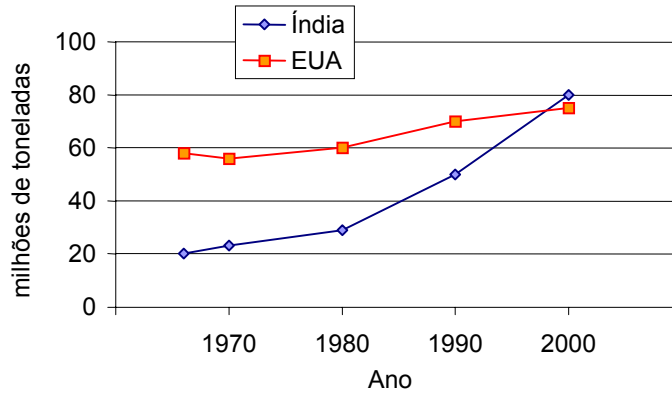
Distribuição de pacientes<sup>(a)</sup> segundo estado nutricional. IBRANUTRI, maio a novembro, 1996.

### 1.5.3 - Diagrama linear

Produção de leite (milhões de toneladas). Índia e Estados Unidos, 1966 – 2000.

Ano	Índia	Estados Unidos
1966	20	58
1970	23	56
1980	29	60
1990	50	70
2000	80	75

Fonte: *State of the World*, 2001. W W Norton&Company, N.Y.



Fonte: *State of the World*, 2001. W W Norton&Company, N.Y.

Produção de leite (milhões de toneladas). Índia e Estados Unidos, 1966 – 2000.

### 1.5.4 - Histograma

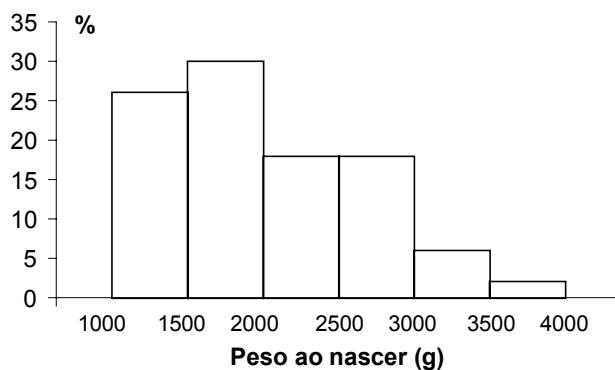
Adequado para representar uma variável quantitativa contínua

Intervalos de classe com mesma amplitude

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g)

Peso(g)	Nº	%
1000  -- 1500	13	26
1500  -- 2000	15	30
2000  -- 2500	9	18
2500  -- 3000	9	18
3000  -- 3500	3	6
3500  -- 4000	1	2
Total	50	100

Fonte: van Vliet PKJ *et al.*, 1973.

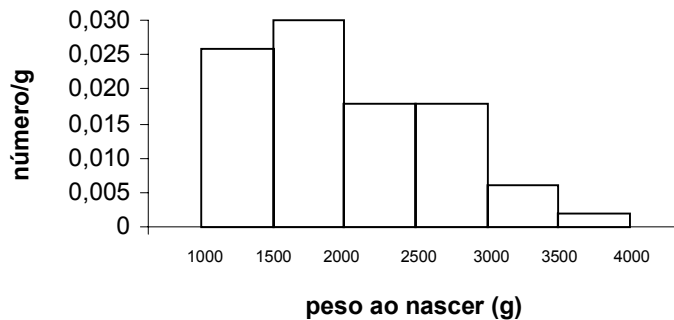


Fonte: van Vliet PKJ *et al.*, 1973.

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g).

**Atenção:** O gráfico é construído considerando-se pessoas por unidade de medida (densidade)

Peso(g)	Nº	%	amplitude	Nº/amplitude	(Nº/amplitude)x10000
1000  -- 1500	13	26	500	0,026	26
1500  -- 2000	15	30	500	0,030	30
2000  -- 2500	9	18	500	0,018	18
2500  -- 3000	9	18	500	0,018	18
3000  -- 3500	3	6	500	0,006	6
3500  -- 4000	1	2	500	0,002	2
Total	50	100			



**OBS:** notar que com intervalos iguais, não é necessário fazer ajuste na altura dos retângulos dado que as bases são de mesmo tamanho (mesma amplitude) e, portanto, com proporcionalidade assegurada.

Intervalos de classe com amplitudes diferentes

Distribuição de mulheres idosas segundo a altura.

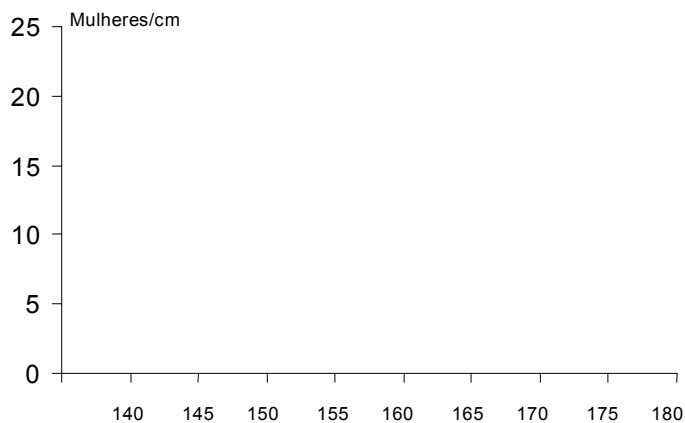
Altura (cm)	Nº	%
140 --150	12	3,4
150 --155	52	14,8
155 --160	109	31,1
160 --170	156	44,4
170 --180	22	6,3
Total	351	100

Fonte: Hand DJ *et al.*, 1994.

É necessário fazer o ajuste

Altura (cm)	Nº	Amplitude	Nº/amplitude
140 --150	12	10	1,2
150 --155	52	5	10,4
155 --160	109	5	21,8
160 --170	156	10	15,6
170 --180	22	10	2,2
Total	351		

Distribuição de mulheres idosas segundo a altura.

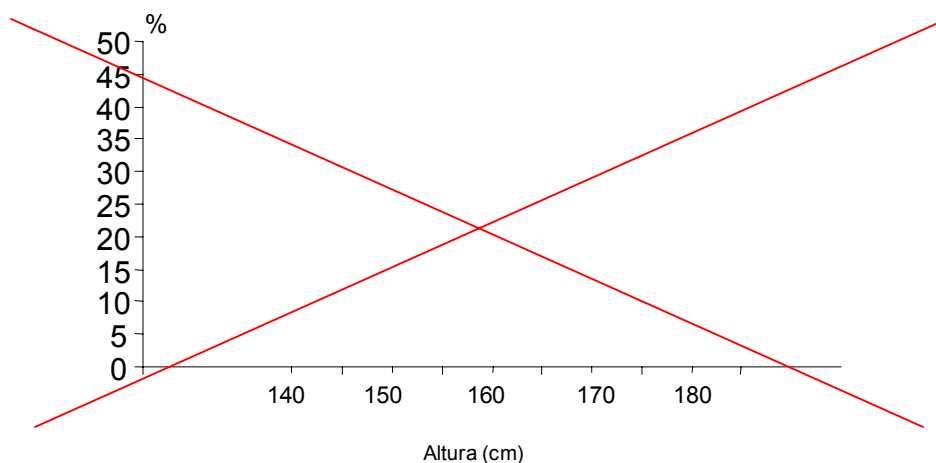


**Cuidado: S**

**altura (cm)**

**clusões erradas!**





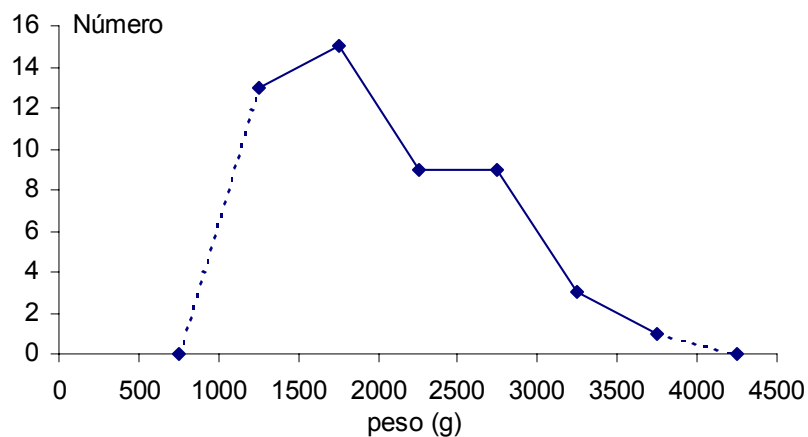
### 1.5.5- Polígono de freqüência simples

#### Intervalos de classe com mesma amplitude

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g).

Peso(g)	Nº	%
1000  -- 1500	13	26
1500  -- 2000	15	30
2000  -- 2500	9	18
2500  -- 3000	9	18
3000  -- 3500	3	6
3500  -- 4000	1	2
Total	50	100

Fonte: Hand DJ *et al.*, 1994.



Fonte: Hand DJ *et al.*, 1994.

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g).

### Intervalos de classe com amplitudes diferentes

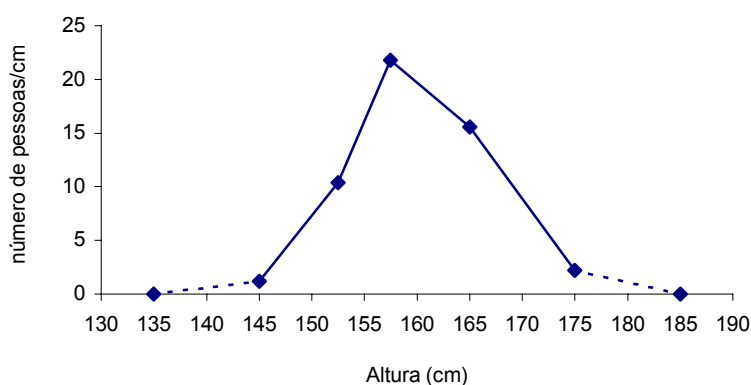
Distribuição de mulheres idosas segundo a altura.

Altura (cm)	Nº	%
140 --150	12	3,4
150 --155	52	14,8
155 --160	109	31,1
160 --170	156	44,4
170 --180	22	6,3
Total	351	100

Fonte: Hand DJ *et al.*, 1994.

Fazendo-se o ajuste:

Altura (cm)	Nº	Amplitude	Nº/amplitude
140 --150	12	10	1,2
150 --155	52	5	10,4
155 --160	109	5	21,8
160 --170	156	10	15,6
170 --180	22	10	2,2
Total	351		



Fonte: Hand DJ *et al.*, 1994.

Distribuição de mulheres idosas segundo a altura (cm).

### 1.5.6 - Representação gráfica de duas variáveis qualitativas

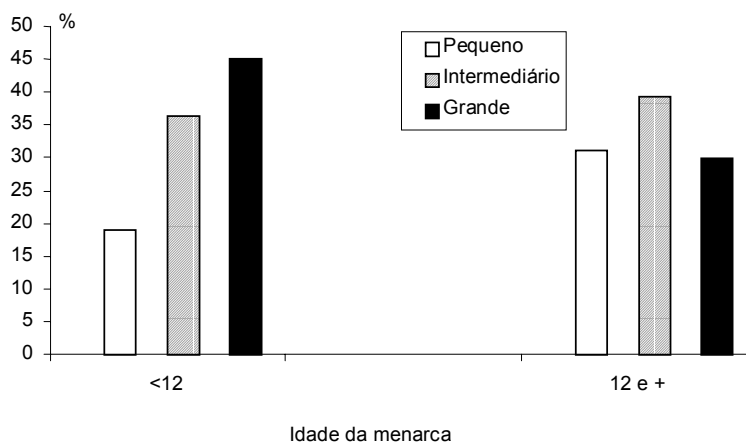
Os dados são de um estudo de obesidade em mulheres, que estuda a relação entre idade da menarca e a medida do tríceps.

Idade da menarca	Medida do tríceps		
	Pequeno	Intermediário	Grande
< 12 anos	15	29	36
12 anos e mais	156	197	150

Fonte: Hand DJ *et al.*, 1994.

Calculando-se as porcentagens tem-se:

Idade (anos)	Medida do tríceps						Total	
	Pequeno		Intermediário		Grande		n	%
	n	%	n	%	n	%		
<12	15	18,8	29	36,2	36	45,0	80	100
12 e +	156	31,0	197	39,2	150	29,8	503	100
Total	171	29,3	226	38,8	186	31,9	583	100



Distribuição de mulheres segundo idade da menarca e medida do tríceps.

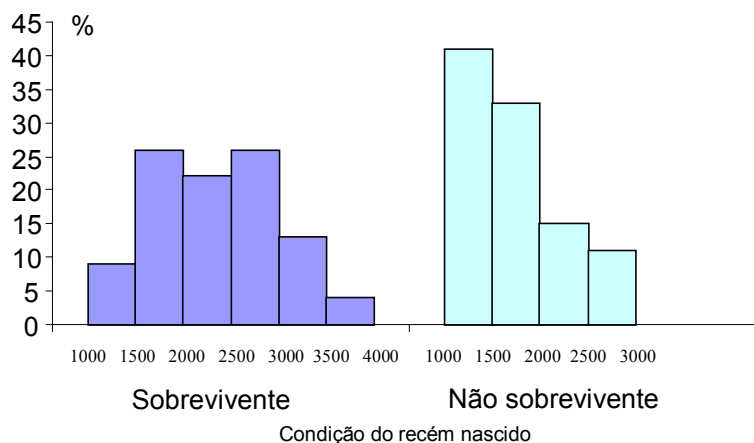
### 1.5.7 - Representação gráfica de duas variáveis quantitativas

Fixando-se os percentuais na condição do recém-nascido

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascido

Peso(g)	Sobrevivente		Não sobrevivente		Total	
	Nº	%	Nº	%	Nº	%
1000  -- 1500	2	9	11	41	13	26
1500  -- 2000	6	26	9	33	15	30
2000  -- 2500	5	22	4	15	9	18
2500  -- 3000	6	26	3	11	9	18
3000  -- 3500	3	13	0	-	3	6
3500  -- 4000	1	4	0	-	1	2
Total	23	100	27	100	50	100

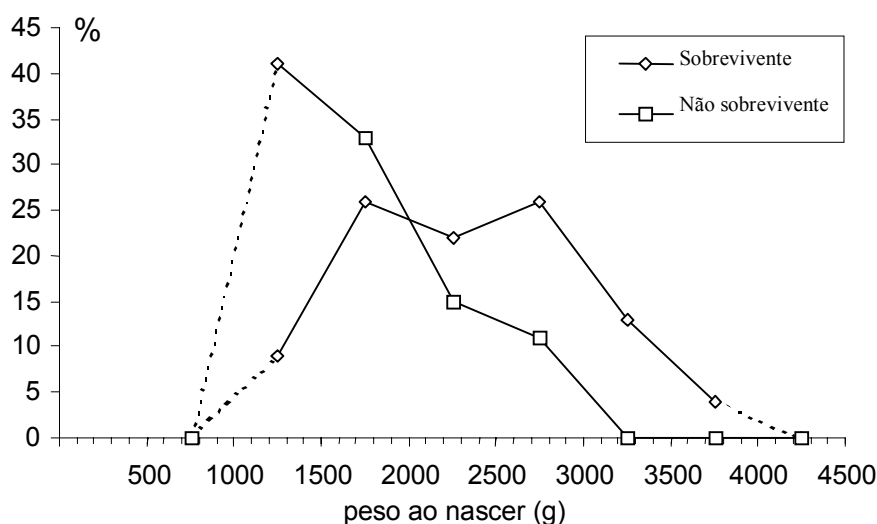
Fonte: Hand DJ *et al.*, 1994.



Fonte: Hand DJ *et al.*, 1994.

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascido.

## Polígono de frequências



Fonte: Hand DJ *et al.*, 1994.

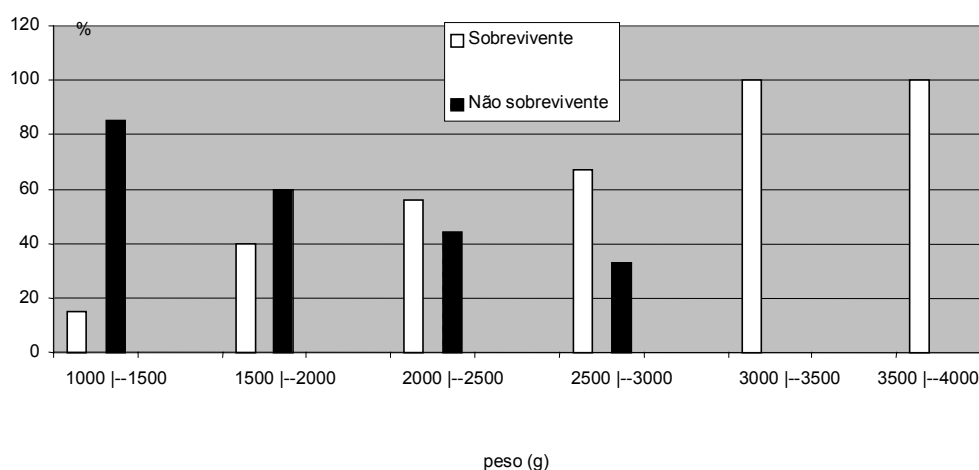
Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascido.

Fixando-se os percentuais no peso ao nascer:

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascido.

Peso(g)	Sobrevivente		Não sobrevivente		Total	
	Nº	%	Nº	%	Nº	%
1000  -- 1500	2	15	11	85	13	100
1500  -- 2000	6	40	9	60	15	100
2000  -- 2500	5	56	4	44	9	100
2500  -- 3000	6	67	3	33	9	100
3000  -- 3500	3	100	0	-	3	100
3500  -- 4000	1	100	0	-	1	100
<b>Total</b>	<b>23</b>	<b>46</b>	<b>27</b>	<b>54</b>	<b>50</b>	<b>100</b>

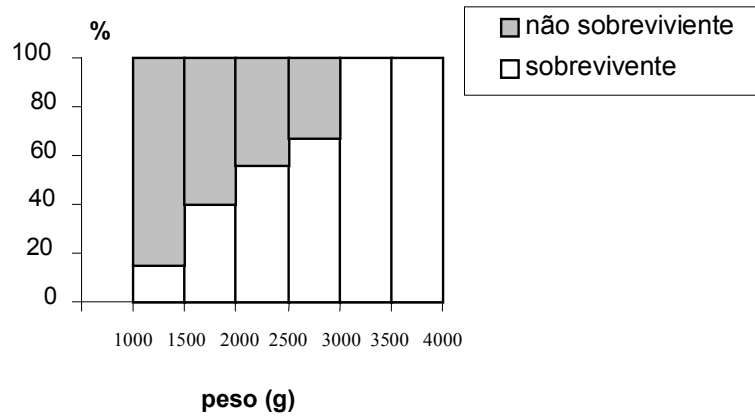
Fonte: Hand DJ *et al.*, 1994.



Fonte: Hand DJ *et al.*, 1994.

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascido.

Outro gráfico possível:



Fonte: Hand DJ *et al.*, 1994.

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascido.

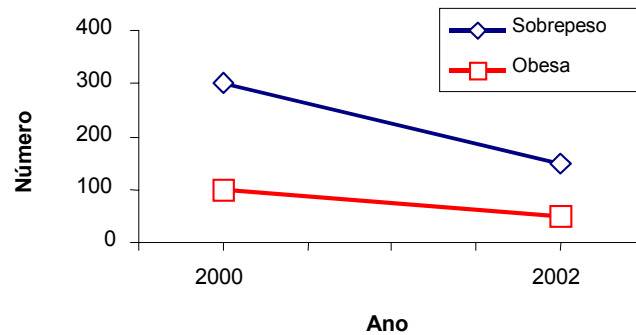
### 1.5.8 - Escalas aritmética e logarítmica

Número de crianças segundo massa corporal. Escola X, 2000 e 2002.

Ano	Sobrepeso	Obesas
2000	300	100
2002	150	50

Fonte: dados hipotéticos

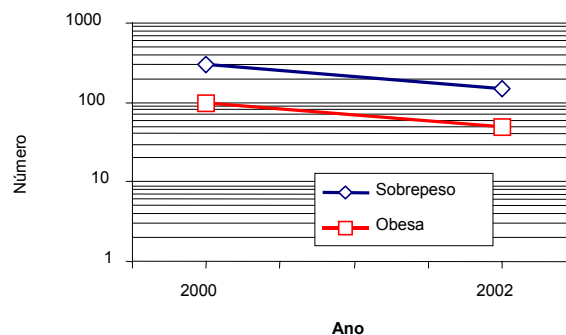
#### Gráfico em escala aritmética



Fonte: dados hipotéticos

Número de crianças segundo massa corporal. Escola X, 2000 e 2002.

#### Gráfico em escala logarítmica



Fonte: dados hipotéticos

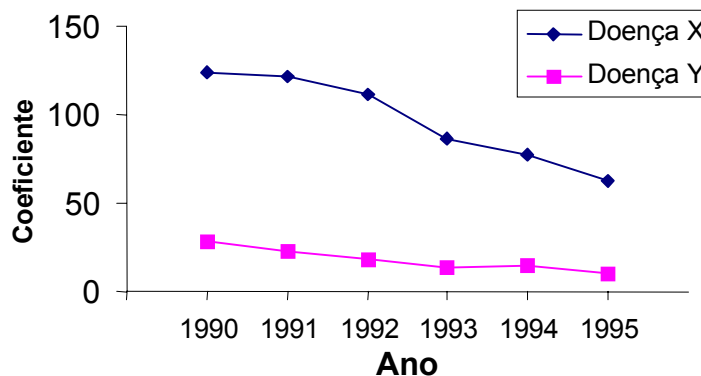
Número de crianças segundo massa corporal. Escola X, 2000 e 2002.

### Gráfico em escala aritmética

Coefficiente de mortalidade pela doença X e Y (100.000hab). Determinada localidade, 1990-1995.

Ano	Doença X	Doença Y
1990	123,5	28,7
1991	121,4	22,4
1992	111,9	17,7
1993	85,9	13,9
1994	77,1	14,8
1995	62,2	10,5

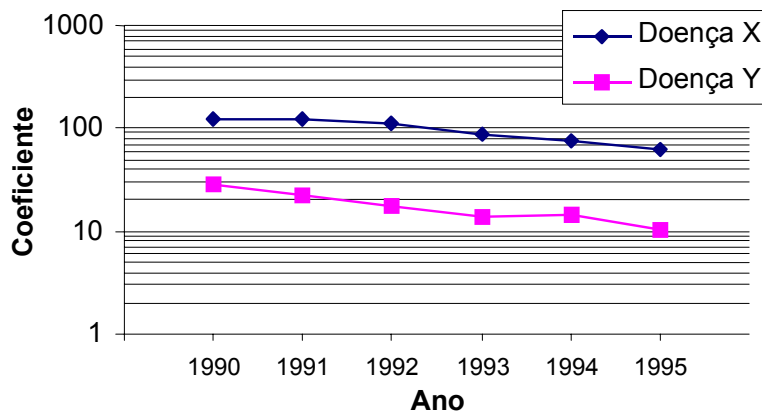
Fonte: Dados hipotéticos



Fonte: Dados hipotéticos

Coefficiente de mortalidade pela doença X e Y (100.000hab). Determinada localidade, 1990-1995.

### Gráfico em escala logarítmica



Fonte: dados hipotéticos

Coefficiente de mortalidade pela doença X e Y (100.000hab). Determinada localidade, 1990-1995.

### Questão 9

Apresente os dados da tabela em um gráfico apropriado.

Distribuição de crianças segundo nível de retinol sérico. Cansação, Bahia, 1992.

Retinol sérico ( $\mu\text{g}/\text{dl}$ )	Nº	%
Aceitável (20-49)	89	55,3
Baixo (10-19,9)	65	40,4
Deficiente ( $<10$ )	7	4,3
Total	161	100

Fonte: Prado MS *et al*, 1995.

### Questão 10

Apresente os dados da tabela em:

- um histograma.
- um polígono de freqüências.

Distribuição de homens adultos segundo comprimento do ante-braço

Comprimento (polegadas)	Nº	%
16,0  -- 16,5	4	2,9
16,5  -- 17,0	3	2,1
17,0  -- 17,5	12	8,6
17,5  -- 18,0	11	7,9
18,0  -- 18,5	21	15,0
18,5  -- 19,0	25	17,9
19,0  -- 19,5	23	16,4
19,5  -- 20,0	19	13,6
20,0  -- 20,5	10	7,1
20,5  -- 21,0	10	7,1
21,0  -- 21,5	2	1,4
Total	140	100

Fonte: Hand DJ *et al*, 1994.

### Questão 11

Apresente os dados da tabela em um gráfico apropriado.

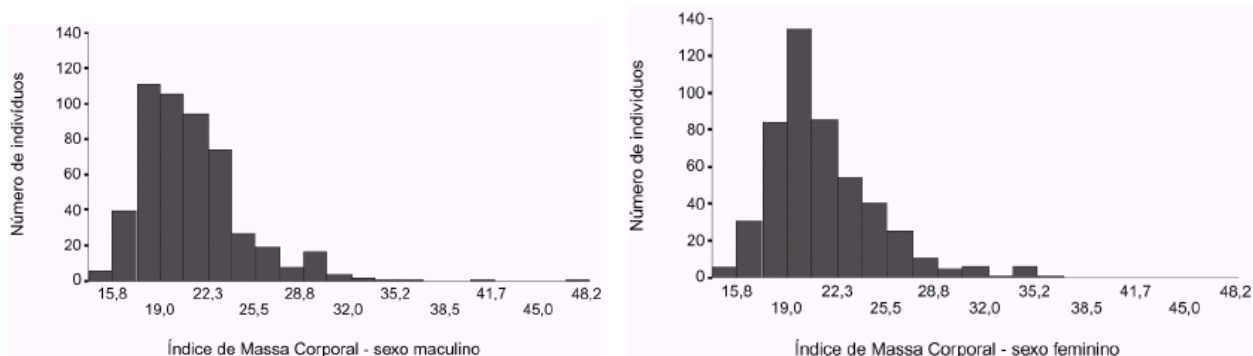
Distribuição de homens segundo presença de doença coronariana (CHD) e consumo de café

Condição	Alto consumo de café ( $\geq 100$ xícaras/mês)		Moderado consumo de café ( $<100$ xícaras/mês)		Total	
	n	%	n	%	n	%
Com CHD	38	4,8	39	4,2	77	4,5
Sem CHD	752	95,2	889	95,8	1641	95,5
Total	790	100	928	100	1718	100

### Questão 12

Interprete os gráficos dos itens a seguir

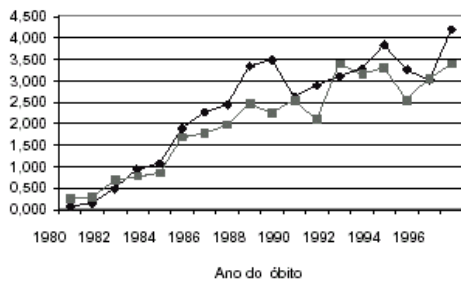
- artigo: Diagnóstico de sobrepeso em adolescentes: estudo do desempenho de diferentes critérios para o Índice de Massa Corporal de MONTEIRO POA *et al*. (*Rev. Saúde Pública*, 2000;34(5):506-13).



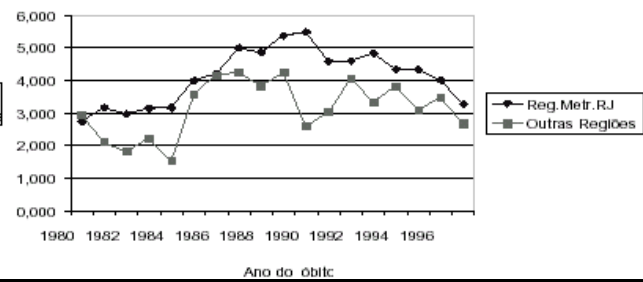
**Figura** - Histogramas para o Índice de Massa Corporal nos adolescentes nascidos em 1982, por sexo. Pelotas, RS, Brasil. 1998.

b) Artigo: Mortalidade por desnutrição em idosos, região Sudeste do Brasil, 1980-1997 de OTERO UB *et al.* (*Rev. Saúde Pública*, 2002; 36(2):141-48).

Taxa de mortalidade por desnutrição em MG (1980-1997)



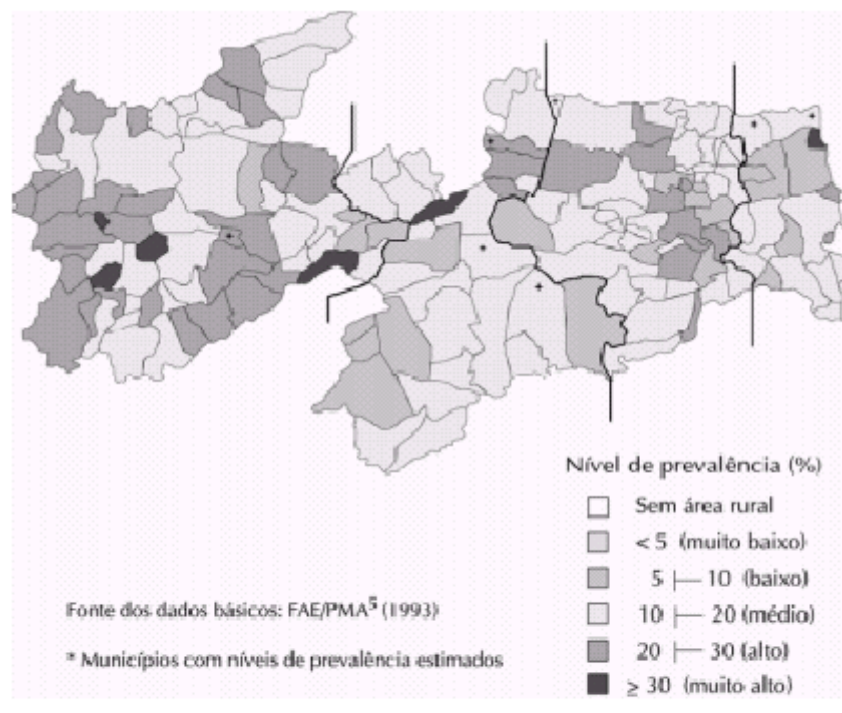
Taxa de mortalidade por desnutrição em idosos/10.000 no RJ (1980-1997)



**Figura 2** – Distribuição das taxas de mortalidade por desnutrição em idosos nas regiões metropolitanas e outras regiões da região Sudeste do País, entre 1980 a 1997.

(adaptado)

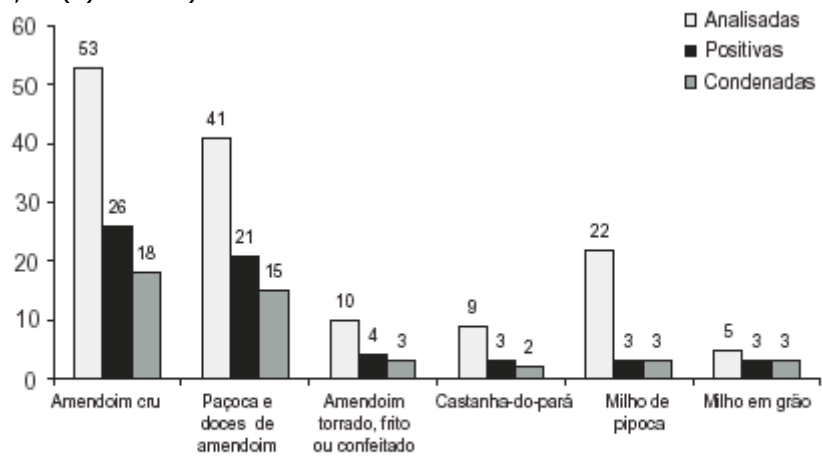
b) Artigo: Cartografia do retardo estatural em escolares do Estado da Paraíba, Brasil de CARVALHO AT *et al.* (*Rev. Saúde Pública*, 2000; 34(1):3-8).



**Figura 3** - Municípios classificados por níveis de prevalência do nanismo em escolares de 6 a 9 anos, da 1ª série do 1º grau. Zona Rural. Estado da Paraíba, 1991.

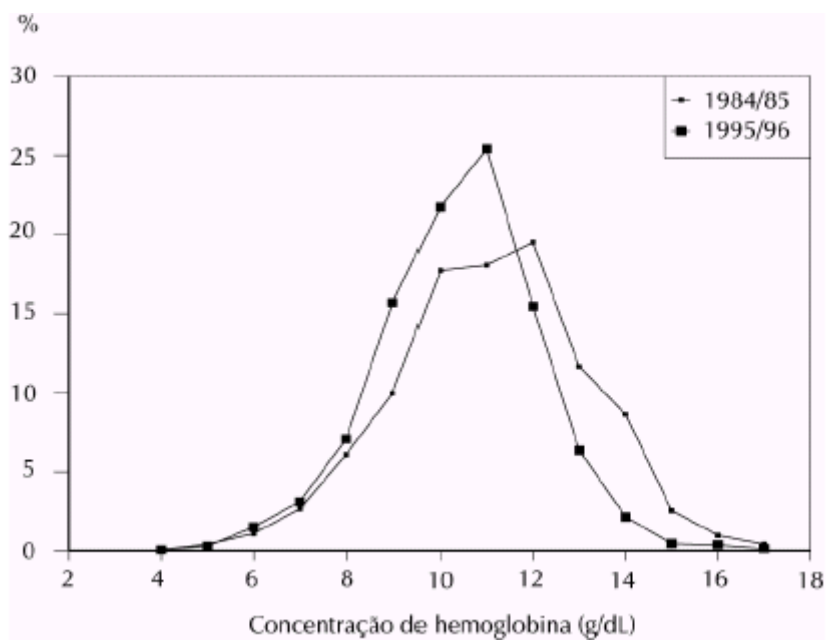


d) Artigo: Aflatoxinas e ocratoxina A em alimentos e riscos para a saúde humana de CALDAS ED *et al.* (*Rev. Saúde Pública*, 2002; 36(3):319-23).



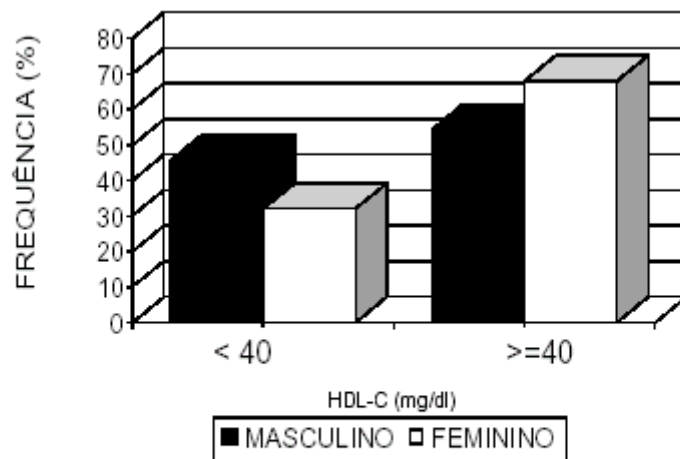
**Figura 1** - Aflatoxinas em alimentos analisados no Distrito Federal no período de julho de 1998 a dezembro de 2000.

e) Artigo: Tendência secular da anemia na cidade de São Paulo (1984-1996) de MONTEIRO CA *et al.* (*Rev. Saúde Pública*, 2000; 34(6, sup):62-72).



**Figura** - Distribuição (%) da concentração de hemoglobina (g/dL). Crianças entre zero e 59 meses de idade da cidade de São Paulo, SP, 1984/85 e 1995/96.

f) Artigo: FORNES NS *et al.* Escores de consumo alimentar e níveis lipêmicos em população de São Paulo, Brasil. *Rev. Saúde Pública*, 2002; 36(1):12-8.



**Figura 1** - Distribuição da população segundo níveis de lipídeos séricos (National Institutes of Health<sup>11</sup>) e sexo. Município de Cotia, São Paulo, Brasil. (adaptado)

## 1.6 - Medidas de tendência central

### 1.6.1 - Média aritmética

**Notação:**

$X$  → variável

$N$  → tamanho da população

$n$  → tamanho da amostra

$\mu$  → média populacional (parâmetro, geralmente desconhecido)

$\bar{X}$  → Estatística (fórmula)

$\bar{x}$  → média amostral (estimativa, valor calculado na amostra)

Média aritmética é o valor que indica o centro de equilíbrio de uma distribuição de freqüências de uma variável quantitativa.

Definição: é a soma dos valores de uma variável, dividida pelo número de valores.

Em uma amostra aleatória simples de tamanho  $n$ , composta das observações  $x_1, x_2, \dots, x_n$ , a média aritmética ( $\bar{x}$ ) é igual a:

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

**OBS:**

- só existe para variáveis quantitativas e seu valor é único;
- é da mesma natureza da variável considerada; e
- sofre influência dos valores aberrantes.

**Exemplo:**

Os dados a seguir são provenientes do grupo *Western Collaborative Group Study*, criado na Califórnia em 1960-61. Foram estudados 3154 homens de meia idade para investigar a relação entre padrões de comportamento e risco de doença coronariana. Os dados apresentados são de 40 homens para os quais foram medidos os níveis de colesterol (mg por 100ml) e realizada uma categorização segundo comportamento. O comportamento de tipo A é caracterizado pela urgência, agressividade e ambição. O de tipo B é relaxado, não competitivo e menos preocupado.

Tipo A: nível de colesterol

233	291	312	250	246	197	268	224	239	239
254	276	234	181	248	252	202	218	212	325

Tipo B: nível de colesterol

344	185	263	246	224	212	188	250	148	169
226	175	242	252	153	183	137	202	194	213

Colesterol médio:

$$\bar{x}_A = \frac{233 + 291 + \dots + 212 + 325}{20} = 245,05 \text{ mg} / 100 \text{ ml}$$

$$\bar{x}_B =$$

Se todos os homens do tipo A tivessem o mesmo nível de colesterol, este seria 245,1 mg/100ml e do tipo B, \_\_\_\_\_.

**1.6.2 - Mediana**

É o valor que ocupa a posição central de uma série de n observações, quando estas estão ordenadas de forma crescente ou decrescente.

Quando número de observações (n) for ímpar:

a mediana é o valor da variável que ocupa o posto  $\frac{n+1}{2}$

Quando o número de observações (n) for par:

a mediana é a média aritmética dos valores da variável que ocupam os postos  $\frac{n}{2}$  e  $\frac{n+2}{2}$

OBS:

- existe para variável quantitativa e qualitativa ordinal;
- é da mesma natureza da variável considerada;
- torna-se inadequada quando há muitos valores repetidos;
- não sofre influência de valores aberrantes;
- pode ser calculada mesmo quando os dados estão agrupados em intervalos de classe e os extremos de algum intervalo não esteja definido (a não ser que a mediana caia neste intervalo).

**Exemplo:**

Tipo A: nível de colesterol

233	291	312	250	246	197	268	224	239	239
254	276	234	181	248	252	202	218	212	325

Tipo B: nível de colesterol

344	185	263	246	224	212	188	250	148	169
226	175	242	252	153	183	137	202	194	213

Ordenando-se os valores:

Tipo A: nível de colesterol									
181	202	218	233	239	246	250	254	276	312
197	212	224	234	239	248	252	268	291	325

Mediana:  $(239+246)/2=242,5$  mg/100ml

Tipo B: nível de colesterol									

Mediana:  $( \quad + \quad )/2= \quad$  mg/100ml

## 1.7 - Medidas de dispersão

### 1.7.1 - Valores mínimo e máximo

Valores extremos da distribuição

### 1.7.2 - Amplitude de variação

É a diferença entre os 2 valores extremos da distribuição

### 1.7.3 – Variância

Indica o quanto, em média, os quadrados dos desvios de cada observação em relação à média aritmética estão afastados desta média.

$$\text{Variância populacional: } \sigma^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}$$

$$\text{Variância amostral: } S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

### 1.7.4 - Desvio padrão

É a raiz quadrada da variância, ou seja  $\sigma = \sqrt{\sigma^2}$

$$S = \sqrt{S^2}$$

### 1.7.5 - Coeficiente de Variação de Pearson (CV)

É o quociente entre o desvio padrão e a média, ou seja  $CV = \frac{S}{\bar{x}} \times 100$

**Exemplo:**

Tipo A: nível de colesterol

233	291	312	250	246	197	268	224	239	239
254	276	234	181	248	252	202	218	212	325

$$\text{Variância: } s^2 = \frac{(233-245,05)^2 + \dots + (325-245,05)^2}{19} = 1342,37(\text{mg}/100\text{ml})^2$$

$$\text{Desvio padrão: } S = \sqrt{1342,37} = 36,64\text{mg}/100\text{ml}$$

$$\text{Coeficiente de variação: } \frac{36,64}{245,05} \times 100 = 14,95\%$$

Tipo B: nível de colesterol

344	185	263	246	224	212	188	250	148	169
226	175	242	252	153	183	137	202	194	213

$$\text{Variância: } s^2 =$$

$$\text{Desvio padrão } S =$$

Coeficiente de variação:

**Questão 13**

São fornecidos valores de nível de triglicérides (mg/dL) de 9 pessoas

166	158	202	162	135	82	150	86	121
-----	-----	-----	-----	-----	----	-----	----	-----

Calcule, apresentando o desenvolvimento da fórmula:

- o nível médio de triglicérides;
- o nível mediano de triglicérides;
- o desvio padrão do nível de triglicérides e
- o coeficiente de variação do nível de triglicérides.

**Questão 14**

A tabela abaixo foi extraída do artigo: Diagnóstico de sobrepeso em adolescentes: estudo do desempenho de diferentes critérios para o Índice de Massa Corporal de MONTEIRO POA *et al.* (*Rev. Saúde Pública*, 2000; 34(5):506-13).

Discuta os resultados obtidos ignorando a coluna do valor de p ( este tópico será abordado posteriormente).

**Tabela 1** – Estatística descritiva da população em estudo, por sexo (n=493). Pelotas, RS, Brasil. 1998.

Variável	Meninos (n=242)		Meninas (n=251)		p - valor
	Média	DP	Média	DP	
Idade (anos)	16,1	0,2	16,1	0,2	0,6
Peso (kg)	65,2	12,3	57,5	10,5	<0,001
Altura (cm)	170,6	6,6	159,8	6,2	<0,001
IMC (kg/m <sup>2</sup> )	22,1	3,7	22,1	3,5	0,8
Dobra subescapular (mm)*	19,9	7,5	23,7	6,3	<0,001
Dobra tricipital (mm)*	19,6	6,3	26,3	5,4	<0,001

\*As dobras cutâneas foram medidas apenas nos 92 meninos e 96 meninas cujo IMC foi igual ou superior ao percentil 85 para idade e sexo conforme Nhanes I (OMS).<sup>8</sup>

**Questão 15**

A tabela abaixo foi extraída do artigo: Avaliação da capacidade preditiva da circunferência da cintura para obesidade global e hipertensão arterial em mulheres residentes na Região Metropolitana de Belo Horizonte, Brasil de VELASQUEZ-MELENDZ G *et al.* (*Cad. Saúde Pública*, 2002; 18(3): 765-771). Calcule e interprete os coeficientes de variação de Pearson para cada uma das variáveis apresentadas.

Table 1

Mean and standard deviation for selected variables in a sample of women.  
Belo Horizonte, Minas Gerais, Brazil, 2000.

Variables	Mean (n = 791)	Standard deviation
Age (years)	30.95	10.53
Weight (kg)	60.35	12.50
Height (cm)	158.49	6.13
Body mass index (kg/m <sup>2</sup> )	24.04	4.85
Waist circumference (cm)	78.60	13.39
Hips circumference (cm)	97.83	9.50
Waist-to-hips ratio	0.80	0.10
Systolic arterial pressure (mmHg)	116.26	15.66
Diastolic arterial pressure (mmHg)	76.55	12.80

### 1.7.6 - Quartil

Valores da variável que dividem a distribuição em quatro partes iguais.

1/4	1/2	3/4
25%	25%	25%

Q1: deixa abaixo 25% das observações

25%	75%
-----	-----

Q2: deixa abaixo 50% das observações

50%	50%
-----	-----

Q3: deixa abaixo 75% das observações

75%	25%
-----	-----

Primeiro quartil:  $Q1 = x_{\left(\frac{1}{4}(n+1)\right)}$  ; Terceiro quartil:  $Q3 = x_{\left(\frac{3}{4}(n+1)\right)}$

onde x é o valor da variável e  $\left(\frac{1}{4}(n+1)\right)$  e  $\left(\frac{3}{4}(n+1)\right)$  são índices que representam as posições ocupadas por x.

Os dados abaixo são referentes ao peso ao nascer de 50 recém-nascidos que tiveram síndrome de desconforto respiratório idiopático grave.

23 crianças sobreviveram e 27 foram a óbito (\*).

1.050*	2.500*	1.890*	1.760	2.830
1.175*	1.030*	1.940*	1.930	1.410
1.230*	1.100*	2.200*	2.015	1.715
1.310*	1.185*	2.270*	2.090	1.720
1.500*	1.225*	2.440*	2.600	2.040
1.600*	1.262*	2.560*	2.700	2.200
1.720*	1.295*	2.730*	2.950	2.400
1.750*	1.300*	1.130	2.550	3.160
1.770*	1.550*	1.575	2.570	3.400
2.275*	1.820*	1.680	3.005	3.640

Ordenando-se os dados, em cada grupo, obtém-se:

1.030*	1.310*	2.200*	1.680	2.550
1.050*	1.500*	2.270*	1.715	2.570
1.100*	1.550*	2.275*	1.720	2.600
1.175*	1.600*	2.440*	1.760	2.700
1.185*	1.720*	2.500*	1.930	2.830
1.225*	1.750*	2.560*	2.015	2.950
1.230*	1.770*	2.730*	2.040	3.005
1.262*	1.820*	1.130	2.090	3.160
1.295*	1.890*	1.410	2.200	3.400
1.300*	1.940*	1.575	2.400	3.640

Fonte: van Vliet PK *et al.* 1973.

Entre os recém-nascidos que sobreviveram:

$$Q1 = x_{\left(\frac{1}{4}(23+1)\right)} = x_6 = 1720g$$

$$Q3 = x_{\left(\frac{3}{4}(23+1)\right)} = x_{18} = 2830g$$

Observe que  $Q2 = x_{\left(\frac{1}{2}(23+1)\right)} = x_{12} = 2200g$

Entre os recém-nascidos que foram a óbito

$$Q1 = x_{\left(\frac{1}{4}(27+1)\right)} = x_7 = 1230g$$

$$Q3 = x_{\left(\frac{3}{4}(27+1)\right)} = x_{21} = 2200g \text{ e } Q2 = x_{\left(\frac{1}{2}(27+1)\right)} = x_{14} = 1600g$$

No cálculo dos quartis os índices 6, 7, 12, 14, 18 e 21 são valores inteiros. Isto irá acontecer sempre que o número de observações (n) for ímpar. Se n for par, o cálculo do índice resultará em um valor fracionário.

Por exemplo, para n=22

$$Q1 = x_{\left(\frac{1}{4}(22+1)\right)} = x_{\left(\frac{23}{4}\right)} = x_{\left(5\frac{3}{4}\right)}$$

que é  $\frac{3}{4}$  do caminho entre  $x_5=1715$  e  $x_6=1720$

$$Q1 = 1715 + \frac{3}{4}(1720 - 1715) = 1718,8g$$

$$Q3 = x_{\left(\frac{3}{4}(22+1)\right)} = x_{\left(17\frac{1}{4}\right)}$$

que é  $\frac{1}{4}$  do caminho entre  $x_{17}=2700$  e  $x_{18}=2830$

$$Q3 = 2700 + \frac{1}{4}(2830 - 2700) = 2732,5g$$

## Percentil

Valores da variável que dividem a distribuição em cem partes iguais.

Entre os recém-nascidos que sobreviveram

Percentil 5:

$$P5 = x_{\left(\frac{5}{100}(23+1)\right)} = x_{\left(\frac{120}{100}\right)} = x_{\left(1\frac{1}{5}\right)} \quad P5 = 1130 + \frac{1}{5}(1410 - 1130) = 1186g$$

que é  $\frac{1}{5}$  do caminho entre  $x_1=1130$  e  $x_2=1410$

Percentil 10:

$$P10 = x_{\left(\frac{10}{100}(23+1)\right)} = x_{\left(\frac{240}{100}\right)} = x_{\left(2\frac{2}{5}\right)}; P10 = 1410 + \frac{2}{5}(1575 - 1410) = 1476g$$

Percentil 50:

$$P50 = x_{\left(\frac{50}{100}(23+1)\right)} = x_{\left(\frac{1200}{100}\right)} = x_{(12)}; P50 = 2200g$$

Percentil 75:

$$P75 = x_{\left(\frac{75}{100}(23+1)\right)} = x_{\left(\frac{1800}{100}\right)} = x_{(18)}; P75 = 2830g$$

Percentil 90:

$$P90 = x_{\left(\frac{90}{100}(23+1)\right)} = x_{\left(\frac{2160}{100}\right)} = x_{\left(21\frac{3}{5}\right)}; P90 = 3160 + \frac{3}{5}(3400 - 3160) = 3304g$$

Percentil 95:

$$P95 = x_{\left(\frac{95}{100}(23+1)\right)} = x_{\left(\frac{2280}{100}\right)} = x_{\left(22\frac{4}{5}\right)}; P95 = 3400 + \frac{4}{5}(3640 - 3400) = 3592g$$

## 1.8 - Box plot e identificação de valores aberrantes (*outliers*)

O Box plot representa graficamente dados de forma resumida em um retângulo onde as linhas da base e do topo são o primeiro e o terceiro quartis, respectivamente. A linha entre estas é a mediana. Linhas verticais que iniciam no meio da base e do topo do retângulo, terminam em valores denominados adjacentes inferior e superior (Chambers *et al.*, 1983, pag 60).

O valor adjacente superior é o maior valor das observações que é menor ou igual a  $Q3 + 1,5(Q3 - Q1)$  e o valor adjacente inferior é definido como o menor valor que é maior ou igual a  $Q1 - 1,5(Q3 - Q1)$ , sendo a diferença  $Q3 - Q1$  denominada intervalo inter-quartil (IIQ).

Valores *outliers* (discrepantes ou aberrantes) são valores que "fogem" da distribuição dos dados. O box plot além de apresentar a dispersão dos dados torna-se útil também para identificar a ocorrência destes valores como sendo os que caem fora dos limites estabelecidos pelos valores adjacentes superior e inferior.

Exemplo:

Tipo A: nível de colesterol

181	202	218	233	239	246	250	254	276	312
197	212	224	234	239	248	252	268	291	325

Tipo B: nível de colesterol

137	153	175	185	194	212	224	242	250	263
148	169	183	188	202	213	226	246	252	344

**Tipo A:**

n=20;

$$Q1 = x_{\frac{1}{4}(n+1)} = x_{\frac{21}{4}} = x_{5\frac{1}{4}} = 218 + \frac{1}{4}(224 - 218) = 218 + 1,5 = 219,5$$

$$Q3 = x_{\frac{3}{4}(n+1)} = x_{\frac{3}{4}(21)} = x_{15\frac{3}{4}} = 254 + \frac{3}{4}(268 - 254) = 254 + 10,5 = 264,5$$

Intervalo Inter-Quartil (IIQ):  $Q3 - Q1 = 45$



325 é o valor adjacente superior. Este é o maior valor da distribuição, igual ou abaixo de 332, onde 332 é dado por:  $264,5 + 1,5 \times 45 = 332$ .

181 é o valor adjacente inferior. É o menor valor da distribuição, igual ou acima de 152, onde 152 é dado por:  $219,5 - 1,5 \times 45 = 152$ .

**Tipo B**

$n=20$

$$Q1 = x_{\frac{1}{4}(n+1)} = x_{\frac{21}{4}} = x_{5\frac{1}{4}} = 175 + \frac{1}{4}(183 - 175) = 175 + 2 = 177$$

$$Q3 = x_{\frac{3}{4}(n+1)} = x_{\frac{3}{4}(21)} = x_{15\frac{3}{4}} = 242 + \frac{3}{4}(246 - 242) = 242 + 3 = 245$$

Intervalo Inter-Quartil (IIQ):  $Q3 - Q1 = 68$

344 é o valor adjacente superior. Este é o maior valor da distribuição, igual ou abaixo de 347, onde 347 é dado por:  $245 + 1,5 \times 68 = 347$ .

137 é o valor adjacente inferior. É o menor valor da distribuição, igual ou acima de 75, onde 75 é dado por:  $177 - 1,5 \times 68 = 75$ .

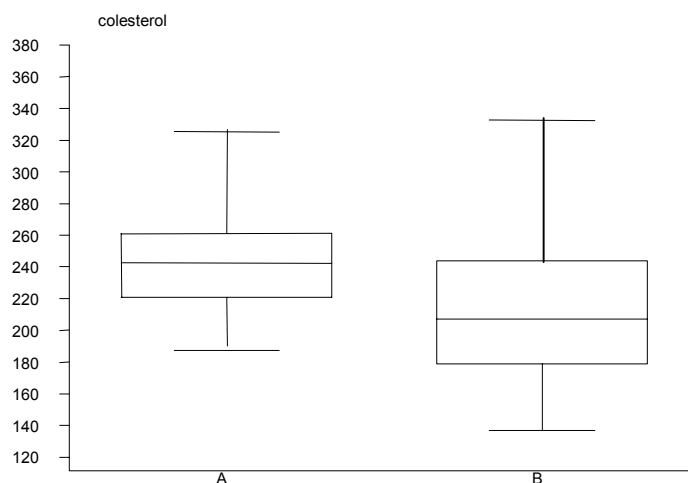


Gráfico - Box plot da variável nível de colesterol segundo tipo de personalidade.

OBS: A confecção do box plot utilizando um programa estatístico para microcomputador irá utilizar o algoritmo específico de cada programa. Por exemplo, o STATA, versão 7, para os dados acima, calcularia o primeiro e terceiro quartis como a média entre dois valores da variável que ocupam as posições 5 e 6; 15 e 16, respectivamente. Neste caso o gráfico pelo Stata identificaria os valores 325 e 344 como *outliers*.

Pelo Stata, versão 7, teríamos:

**Tipo A:**

$$n=20; \text{ para cálculo de } Q1: Q1 = \frac{x_5 + x_6}{2} = \frac{218 + 224}{2} = 221$$

$$\text{cálculo de } Q3: Q3 = \frac{x_{15} + x_{16}}{2} = \frac{254 + 268}{2} = 261$$

Intervalo Inter-Quartil (IIQ):  $Q3 - Q1 = 40$

Valor adjacente superior: maior valor igual ou abaixo de 321 ( $Q3 + 1,5 \times 40 = 321$ ) = 312

Valor adjacente inferior: menor valor igual ou acima de 161 ( $Q1 - 1,5 \times 40 = 161$ ) = 181

Outlier: valor 325

Importante: note como Q1 e Q3 são calculados.

### Tipo B

$n=20$ ; para cálculo de Q1:  $Q1 = \frac{x_5 + x_6}{2} = \frac{175 + 183}{2} = 179$

cálculo de Q3:  $Q3 = \frac{x_{15} + x_{16}}{2} = \frac{242 + 246}{2} = 244$

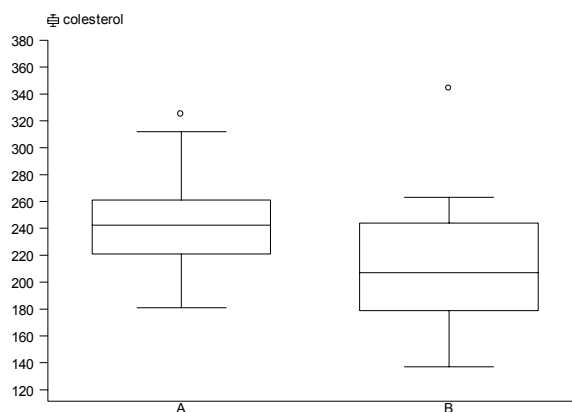
Intervalo Inter-Quartil (IIQ):  $Q3 - Q1 = 65$

Valor adjacente superior: maior valor abaixo de 341,5 ( $Q3 + 1,5 \times 65 = 341,5$ ) = 263

Valor adjacente inferior: menor valor acima de 81,5 ( $Q1 - 1,5 \times 65 = 81,5$ ) = 137

Outlier: valor 344

Box plot da variável nível de colesterol segundo tipo de personalidade (pelo Stata).



### Questão 16

Os dados a seguir são de uma pesquisa que investigou as concentrações de minerais no leite materno, no período de 1984 a 1985. Foram coletadas amostras de leite materno de 55 mulheres que tiveram seus filhos no Hospital Maternidade Odete Valadares, em Belo Horizonte. As mães foram divididas em período de lactação: colostro e leite maduro.

cálcio ( $\mu\text{g/mL}$  de leite) – grupo colostro

113	181	254	311	334	145	221	256	312	344
163	225	275	313	372	163	231	296	323	375
167	241	303	325	375	437				

cálcio ( $\mu\text{g/mL}$  de leite) – grupo maduro

159	175	181	188	200	206	213	214	217	231
238	238	242	244	256	259	260	263	264	275
277	279	281	293	302	303	314	344	394	

a) Calcule a quantidade média de cálcio ( $\mu\text{g/mL}$  de leite) em cada grupo.

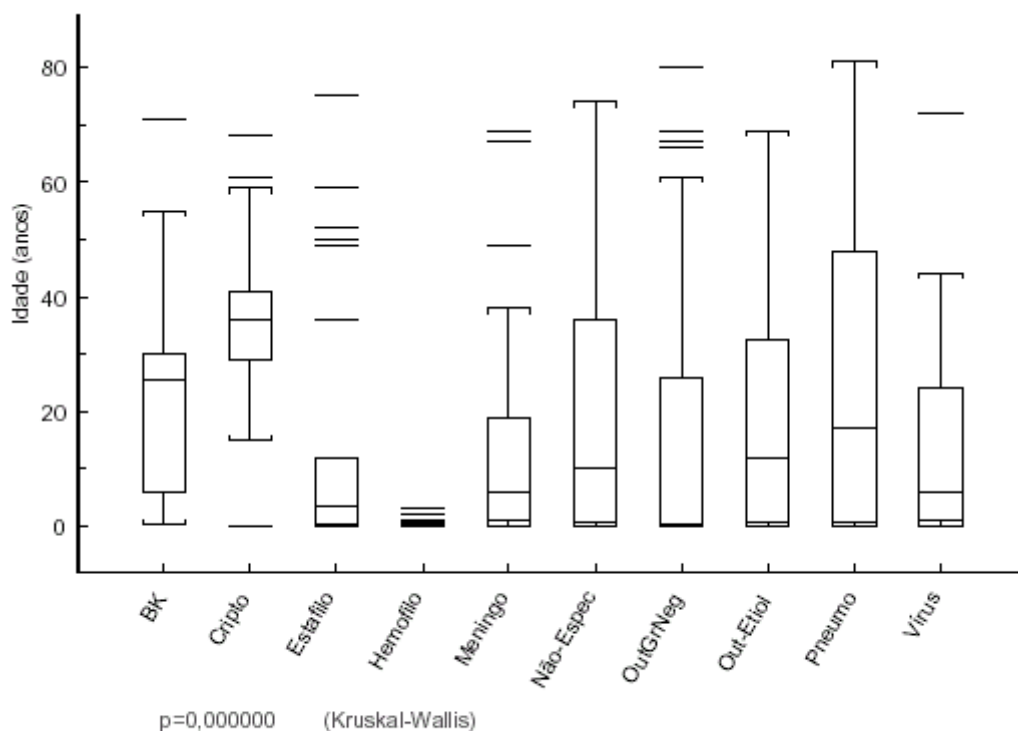
b) Calcule a quantidade mediana de cálcio ( $\mu\text{g/mL}$  de leite) em cada grupo.

c) Desenhe o *box plot* da concentração de cálcio ( $\mu\text{g/mL}$  de leite) representando os dois grupos em um só gráfico.

d) Comente o gráfico *box plot* quanto a dispersão dos dados, existência de valores aberrantes e igualdade de medianas.

### Questão 17

O boxplot abaixo foi extraído do artigo: Vigilância epidemiológica e avaliação da assistência às meningites de ESCOSTEGUY CC *et al.* (*Rev. Saúde Pública*, 2004;38(5):657-63). Avalie os resultados apresentados.

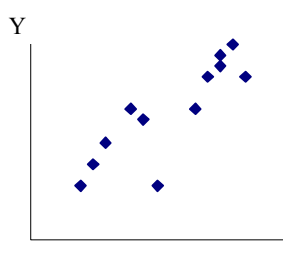


**Figura 1** - Box-plot da idade segundo etiologia da meningite, 1986 a 9/2002.

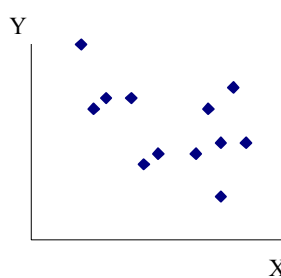
### 1.9 - Noções de correlação

X e Y são variáveis aleatórias quantitativas

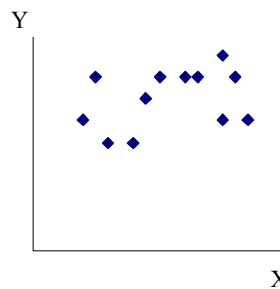
Gráficos de dispersão



correlação positiva



correlação negativa



correlação inexistente

Coefficiente de correlação de Pearson ( $\rho$ ): Mede o grau de associação entre 2 variáveis quantitativas X e Y.

**Definição:**

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

onde

$\sigma_{XY}$  é a covariância de X e Y (dispersão conjunta de X e Y)

$\sigma_X$  é o desvio padrão de X (dispersão de X)

$\sigma_Y$  é o desvio padrão de Y (dispersão de Y)

Covariância: É o valor médio do produto dos desvios de X e Y, em relação às suas respectivas médias.

$$\sigma_{XY} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n} = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n} = \sum X_i Y_i - n \bar{X} \bar{Y}$$

estimador (r)

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[ \sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2 \right]}}$$

### Propriedades

- a)  $-1 \leq \rho \leq +1$
- b)  $\rho$  não possui dimensão, isto é, não depende da unidade de medida das variáveis X e Y

### Exemplo:

Os dados a seguir são provenientes de um estudo que investiga a composição corporal e fornece o percentual de gordura corporal (%), idade e sexo para 18 adultos com idades entre 23 e 61 anos.

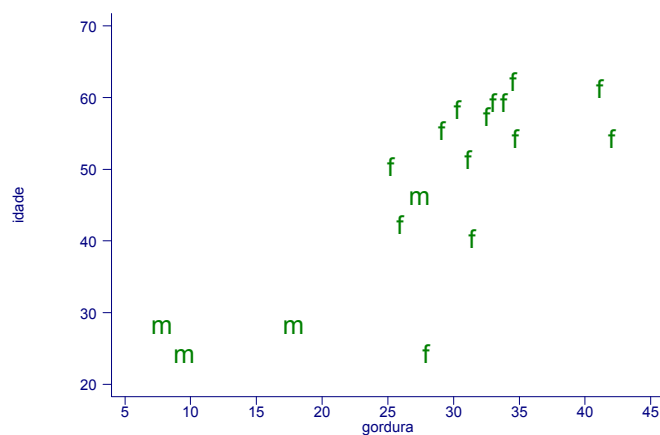
- a) Qual a relação entre a idade e o % de gordura? Existe alguma evidência de que a relação é diferente entre pessoas do sexo masculino e feminino? Explore os dados graficamente.
- b) Calcule o coeficiente de correlação de Pearson entre a idade e o % de gordura para homens e mulheres. Interprete os resultados.

Idade	% gordura	sexo	Idade	% gordura	sexo
23	9,5	M	53	34,7	F
23	27,9	F	53	42,0	F
27	7,8	M	54	29,1	F
27	17,8	M	56	32,5	F
39	31,4	F	57	30,3	F
41	25,9	F	58	33,0	F
45	27,4	M	58	33,8	F
49	25,2	F	60	41,1	F
50	31,1	F	61	34,5	F

M=masculino ; F= feminino

Fonte: Hand DJ *et al.*, 1994.

Dispersão entre % de gordura e idade



### Cálculo do coeficiente de correlação de Pearson

Sexo: masculino

Idade	% gordura	$(y - \bar{y})$	$(x - \bar{x})$	$(x - \bar{x})(y - \bar{y})$	$(y - \bar{y})^2$	$(x - \bar{x})^2$
23	9,5	-7,5	-6,13	45,94	56,25	37,52
27	7,8	-3,5	-7,83	27,39	12,25	61,23
27	17,8	-3,5	2,18	-7,61	12,25	4,73
45	27,4	14,5	11,78	170,74	210,25	138,65
$\bar{y} = 30,5$	$\bar{x} = 15,63$		<b>Total</b>	<b>236,45</b>	<b>291,00</b>	<b>242,13</b>

$$\text{Coeficiente de correlação (idade,%gordura) masculino: } r = \frac{236,45}{\sqrt{291 \times 242,13}} = 0,89$$

Sexo: feminino

Idade	% gordura	$(y - \bar{y})$	$(x - \bar{x})$	$(x - \bar{x})(y - \bar{y})$	$(y - \bar{y})^2$	$(x - \bar{x})^2$
23	27,9	-27,86	-4,42	123,17	776,02	19,55
39	31,4	-11,86	-0,92	10,93	140,59	0,85
41	25,9	-9,86	-6,42	63,30	97,16	41,23
49	25,2	-1,86	-7,12	13,23	3,45	50,71
50	31,1	-0,86	-1,22	1,05	0,73	1,49
53	34,7	2,14	2,38	5,10	4,59	5,66
53	42	2,14	9,68	20,74	4,59	93,67
54	29,1	3,14	-3,22	-10,12	9,88	10,38
56	32,5	5,14	0,18	0,92	26,45	0,03
57	30,3	6,14	-2,02	-12,42	37,73	4,09
58	33	7,14	0,68	4,85	51,02	0,46
58	33,8	7,14	1,48	10,56	51,02	2,19
60	41,1	9,14	8,78	80,26	83,59	77,06
61	34,5	10,14	2,18	22,10	102,88	4,75
$\bar{y} = 50,86$	$\bar{x} = 32,32$		<b>Total</b>	<b>333,64</b>	<b>1389,71</b>	<b>312,12</b>

$$\text{Coeficiente de correlação (idade,%gordura) feminino: } r = \frac{333,64}{\sqrt{1389,71 \times 312,12}} = 0,51$$

### Questão 18

Abaixo temos o peso (kg) e a altura (cm) de 30 meninas de 11 anos de idade atendidas na escola *Heaton Middlel* em *Bradford*, Inglaterra.

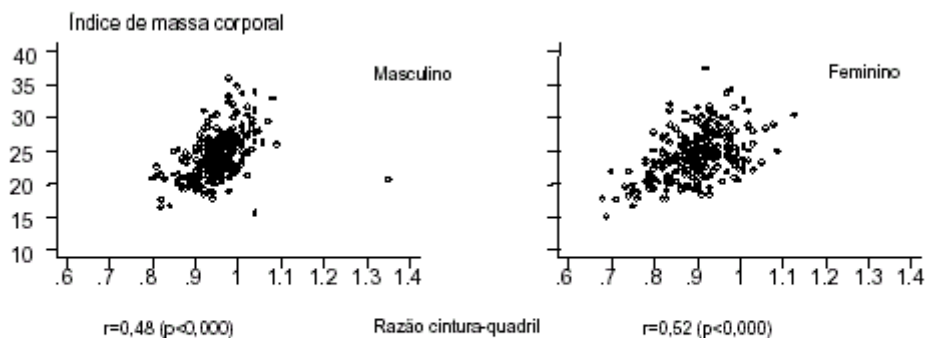
- Fazer o diagrama de dispersão e investigue como é a relação entre as variáveis.
- Calcular o coeficiente de correlação de Pearson.

Altura (cm)	Peso (kg)	Altura (cm)	Peso (kg)	Altura (cm)	Peso (kg)	Altura (cm)	Peso (kg)
135	26	133	31	136	28	146	35
146	33	149	34	154	36	143	42
153	55	141	32	151	48	148	32
154	50	164	47	155	36	143	36
139	32	146	37	149	32	140	33
131	25	149	46	141	29	141	28
149	44	147	36	137	34		
137	31	152	47	135	30		

$$\sum (x - \bar{x})(y - \bar{y}) = 1275; \quad \sum (x - \bar{x})^2 = 1716,80; \quad \sum (y - \bar{y})^2 = 1718,17$$

### Questão 19

Os gráficos abaixo foram extraídos do artigo: Excesso de peso e gordura abdominal para a síndrome metabólica em nipo-brasileiros de LERARIO DG *et al.* (*Rev. Saúde Pública*, 2002;36(1):4-11). Interprete as figuras apresentadas no artigo.



**Figura 1** - Correlação entre os valores do índice de massa corporal (Kg/m<sup>2</sup>) e da razão cintura-quadril de nipo-brasileiros segundo sexo.

## 2 – Medidas de associação

Análise da distribuição conjunta de duas variáveis qualitativas

### 2.1 – Razão de prevalências

Estudo de prevalência: n indivíduos são observados e classificados segundo duas variáveis X e Y, por exemplo X: ronco noturno (X1 – sim, X0 – não) e Y: presença de problema cardíaco (Y1 – sim, Y0 – não).

Variável X	Variável Y		Total (%)
	Y1 (sim)	Y0 (não)	
X1 (sim)	a	b	n1 (100)
X0 (não)	c	d	n0 (100)
Total	m1	m2	n (100)

$p =$  prevalência de Y1 =  $m1/n$

$p1 =$  prevalência de Y1|X1 =  $a/n1$

$p0 =$  prevalência de Y1|X0 =  $c/n0$

$rp =$  razão de prevalências =  $p1/p0$

$dp =$  diferença de prevalências =  $p1 - p0$

#### Exemplo:

São apresentados dados sobre o estado nutricional de 1226 crianças brasileiras de 2 anos de idade, segundo sexo.

Estado nutricional (Y)	Masculino (X1)	Feminino (X0)	Total
(Y1) Desnutridas	29	20	49
(Y0) Normais	574	603	1177
Total	603	623	1226

Prevalência de desnutrição:  $\frac{49}{1226} = 0,040$  ou 4%

Prevalência de desnutrição segundo sexo:

$$\text{Masculino: } \frac{29}{603} = 0,05 \text{ ou } 5,0\%; \text{ Feminino: } \frac{20}{623} = 0,032 \text{ ou } 3,2\%$$

$$\text{Razão de prevalências: } \frac{0,05}{0,032} = 1,56$$

Diferença de prevalências:  $0,05 - 0,032 = 0,018$  ou 1,8%

A prevalência de desnutrição parece ser maior entre as crianças do sexo masculino. Os meninos apresentam uma prevalência 56% maior do que as meninas.

A prevalência de desnutrição entre meninos é 1,56 vezes (uma vez e meia) a prevalência de desnutrição entre meninas

## 2.2 – Razão de incidências

Estudo de incidência

Variável X	Variável Y		Total
	Y1	Y0	
X1	a	b	n1 (100%)
X0	c	d	n0 (100%)
Total	m1	m2	n (100%)

$r =$  incidência de Y1 =  $m1/n$

$r_1 =$  incidência de Y1 entre os X1 =  $a/n_1$

$r_0 =$  incidência de Y1 entre os X0 =  $c/n_0$

$r_i =$  razão de incidências =  $r_1/r_0$

$d_i =$  diferença de incidências =  $r_1 - r_0$

incidência

$r_1$

$r_0$

risco

$r_1/r_0$

$r_1 - r_0$

$r_i = r_r =$  razão de riscos = risco relativo =  $r_1/r_0$

$d_i = r_a =$  risco atribuível =  $r_1 - r_0$

### Exemplo:

Fumar	Morte em 5 anos por DIC		Total
	Sim	Não	
Sim	208	850	1058
Não	264	1467	1731
Total	472	2317	2789

$$r = 472/2789 = 0,17 = 17\%$$

$$r_1 = 208/1058 = 0,20 = 20\%$$

$$r_0 = 264/1731 = 0,15 = 15\%$$

$$r_r = 0,20/0,15 = 1,33$$

$$r_a = 0,20 - 0,15 = 0,05 = 5\%$$

A incidência de mortes parece ser maior entre as pessoas que fumam. Os fumantes apresentam uma incidência 33% maior do que os não fumantes.

A mortalidade entre fumantes é 1,33 vezes a mortalidade entre não fumantes.

## 2.3 – Razão de odds (odds ratio)

## Estudo do tipo caso-controle

Variável X	Variável Y		Total
	Y1	Y0	
X1	a	b	n1 (100%)
X0	c	d	n0 (100%)
Total	m1	m2	n (100%)

*odds* a favor de Y1:

$$\begin{aligned} \text{na categoria X1} &= (a/n1) \div (b/n1) \\ \text{na categoria X0} &= (c/n0) \div (d/n0) \end{aligned}$$

$$\text{Razão de odds ou odds ratio: } [(a/n_1) \div (b/n_1)] \div [(c/n_0) \div (d/n_0)] = (a \div b) \div (c \div d) = \frac{a/b}{c/d} = \frac{a \cdot d}{b \cdot c}$$

### Exemplo:

Os dados a seguir são de um estudo sobre câncer de esôfago e consumo de álcool.

Condição	Consumo médio de álcool (g/dia)		Total
	80 e +	0-79	
Casos	96	104	200
Controles	109	666	775
Total	205	770	975

Fonte: Tuyns *et al.*, 1977.

(entre expostos) odds a favor de casos entre consumidores de 80 e + g/dia:  $\frac{96}{205} \div \frac{109}{770} = \frac{96}{205} \cdot \frac{770}{109} = 0,88$

(entre não expostos) odds a favor de casos entre consumidores de 0-79g/dia:  $\frac{104}{770} \div \frac{666}{770} = \frac{104}{666} = 0,16$

$$\text{odds ratio: } \frac{96}{109} \div \frac{104}{666} = \frac{96 \cdot 666}{109 \cdot 104} = 5,6$$

A força de morbidade de câncer de esôfago entre consumidores de 80 e + g/dias de bebida alcoólica é 5,6 a força de morbidade entre os que consomem de 0 a 79g/dia.

Odds ratio utilizando-se os dados de DIC:  $OR = (208 \times 1467) / (850 \times 265) = 1,36$

## 2.4 - Qui-quadrado de Pearson

Duas variáveis qualitativas:

- X - curso universitário e
- Y - sexo do aluno

Questão: sexo do indivíduo influi na escolha do curso?



Situação 1

Curso	Masculino	Feminino	Total
	n	n	n
Economia	24	36	60
Administração	16	24	40
Total	40	60	100

Curso	Masculino		Feminino		Total	
	n	proporção	n	proporção	n	proporção
Economia	24	0,6	36	0,6	60	0,6
Administração	16	0,4	24	0,4	40	0,4
Total	40	1	60	1	100	1

As proporções de escolha dos cursos não diferem segundo sexo do estudante

Situação 2

Curso	Masculino	Feminino	Total
	n	n	n
Física	100 (a)	20 (b)	120
Ciências Sociais	40 (c)	40 (d)	80
Total	140	60	200

Curso	Masculino		Feminino		Total	
	n	Proporção	n	proporção	n	proporção
Física	100	0,7	20	0,3	120	0,6
Ciências Sociais	40	0,3	40	0,7	80	0,4
Total	140	1	60	1	200	1

A distribuição de alunos em cada curso, segundo sexo não é a mesma. Sexo e curso podem estar associados.

Se a variável sexo **não fosse associada** à escolha do curso, quantos indivíduos **esperaríamos** em Física, entre os homens?

$$\text{Esperaríamos : } 0,6 \times 140 \text{ ou } \frac{120}{200} \times 140 = 84$$

Se a variável sexo **não fosse associada** à escolha do curso, quantos indivíduos **esperaríamos** em Ciências Sociais, entre os homens?

$$\text{Esperaríamos: } 0,4 \times 140 \text{ ou } \frac{80}{200} \times 140 = 56$$

Se a variável sexo **não fosse associada** à escolha do curso, quantos indivíduos **esperaríamos** em Física, entre as mulheres?

$$\text{Esperaríamos : } 0,6 \times 60 \text{ ou } \frac{120}{200} \times 60 = 36$$

Se a variável sexo **não fosse associada** à escolha do curso, quantos indivíduos **esperaríamos** em Ciências Sociais, entre as mulheres?

Esperaríamos:  $0,4 \times 60$  ou  $\frac{80}{200} \times 60 = 24$

Tabela esperada, sob a condição de independência

Curso	Masculino	Feminino	Total
	n	n	n
Física	84	36	120
Ciências Sociais	56	24	80
Total	140	60	200

Valores observados O	Valores esperados E	(O-E)	(O-E) <sup>2</sup>	$\frac{(O-E)^2}{E}$
100	84	16	256	3,048
40	56	-16	256	4,571
20	36	-16	256	7,11
40	24	16	256	10,667

Qui-quadrado=25,397

O Qui-quadrado é obtido somando-se a diferença ao quadrado entre as freqüências observadas e as esperadas, dividido pelas freqüências esperadas

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Se o Qui-quadrado for igual a zero, então não existe associação entre as variáveis. Quanto maior o valor do Qui-quadrado, maior a chance de existir associação entre as variáveis, entretanto, o Qui-quadrado não mede força de associação.

### Exemplo:

Com o objetivo de investigar a associação entre história de bronquite na infância e presença de tosse diurna ou noturna em idades mais velhas, foram estudados 1.319 adolescentes com 14 anos. Destes, 273 apresentaram história de bronquite até os 5 anos de idade sendo que 26 apresentaram tosse diurna ou noturna aos 14 anos.

Número de adolescentes segundo história de bronquite aos 5 anos e tosse diurna ou noturna aos 14 anos de idade. Local X, ano Y.

Tosse	Bronquite		Total
	Sim	Não	
Sim	26	44	70
Não	247	1002	1249
Total	273	1046	1319

Fonte: Holland WW *et al.*,1978.

Cálculo do qui-quadrado de Pearson

Valores observados O	Valores esperados E	(O-E)	(O-E) <sup>2</sup>	$\frac{(O-E)^2}{E}$
26	14,488	11,512	132,526	9,147
247	258,512	-11,512	132,526	0,513
44	55,512	-11,512	132,526	2,387
1002	990,488	11,512	132,526	0,134

Qui-quadrado= 12,181

### Questão 20

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo condição de sobrevivência e peso ao nascer (g).

Peso ao nascer	Óbito	Sobrevivida	Total
Baixo peso (<2500)	24	13	37
Não baixo peso (2500 e mais)	3	10	13
Total	27	23	50

Fonte: Hand DJ *et al.* *A handbook of small data sets.* Chapman&Hall, 1994.

Com base nos dados apresentados na tabela calcule:

- a incidência de óbitos entre crianças com baixo peso,
- a incidência de óbitos entre crianças sem baixo peso,
- a razão de incidências
- a diferença de incidências
- o qui-quadrado do Pearson
- interprete os resultados

### Questão 21

Defeitos do tubo neural são malformações congênitas que surgem durante o desenvolvimento fetal. É conhecida como spina bífida. Estes dados são de um estudo conduzido no país de Gales – Reino Unido, para investigar possível associação entre defeitos do tubo neural e dieta materna. O estudo é do tipo caso-controle: mães que tinham tido bebês com defeitos do tubo neural (casos) e suas irmãs que não tinham tido (controles) foram avaliadas segundo suas dietas e classificadas em boa, razoável e ruim.

Distribuição de recém-nascidos segundo condição caso - com defeitos do tubo neural; controle – recém-nascidos que não tinham defeitos do tubo neural e dieta materna.

Dieta	Casos		Controles		Total	
	n	%	n	%	n	%
Boa	34	13,9	43	35,0	77	21,0
Razoável	110	45,1	48	39,0	158	43,0
Pobre	100	41,0	32	26,0	132	36,0
Total	244	100	123	100	367	100

Considere a dieta boa como categoria de referência (basal) e calcule:

- o *odds ratio* de dieta razoável em relação a dieta boa;
- o *odds ratio* de dieta pobre em relação a boa;
- o qui-quadrado de Pearson e
- interprete os resultados.

### Questão 22

A tabela abaixo foi extraída do artigo: Excesso de peso e gordura abdominal para a síndrome metabólica em nipo-brasileiros de LERARIO DDG *et al.* (*Rev de Saúde Pública*, 2002; 36(1): 4-11). Analise as razões de prevalências obtidas no estudo.

**Tabela 3** - Razões de prevalências de hipertensão, dislipidemia e diabetes por ponto e por intervalo com 95% de confiança (IC 95%) segundo geração, sexo e presença de adiposidade central em indivíduos com excesso de peso.

Fator de risco	Razão de prevalência	IC 95%	p
<b>Hipertensão</b>			
Geração (1ª/2ª)	1,23	1,0-1,54	0,068
Sexo (masculino/feminino)	1,15	0,92-1,43	0,226
Adiposidade central (sim/não)	1,40	1,08-1,82	0,011
<b>Dislipidemia</b>			
Geração (1ª/2ª)	1,06	0,98-1,15	0,170
Sexo (masculino/feminino)	1,07	0,99-1,17	0,011
Adiposidade central (sim/não)	1,15	1,04-1,27	0,007
<b>Diabetes</b>			
Geração (1ª/2ª)	1,17	0,8 -1,69	0,388
Sexo (masculino/feminino)	1,92	1,30-2,86	0,001
Adiposidade central (sim/não)	1,85	1,17-2,91	0,008

### Questão 23

A tabela abaixo foi extraída do artigo: Evolução do padrão de aleitamento materno de Kummer SC *et al.* (*Rev de Saúde Pública*, 2000; 34(2): 143-8). Analise as razões de chances obtidas no estudo. Observe os intervalos de confiança apresentados e tire suas conclusões.

**Tabela 2** - Resultados da análise multivariada para avaliar os efeitos de algumas variáveis na interrupção precoce do aleitamento materno, 1987 e 1994.

Variável	Interrupção precoce	Razão de chances (Intervalo de Confiança de 95%)
Coorte		
1987	73 (35,6%)	1,00
1994	73 (39,0%)	1,16 (0,75-1,79)
Renda <i>per capita</i>		
metade superior	66 (33,2%)	1,00
metade inferior	80 (42,6%)	1,66 (1,06-2,62)
Escolaridade materna		
$\geq$ 4 anos	118 (36,2%)	1,00
< 4 anos	28 (44,4%)	1,18 (0,65-2,14)
Cor da mãe		
não branca	27 (26,5%)	1,00
branca	119 (41,5%)	2,26 (1,35-3,79)
Assistência pré-natal		
$\geq$ 5 consultas	120 (36,6%)	1,00
< 5 consultas	26 (42,6%)	1,18 (0,69-2,16)
Estado civil		
casados	75 (33,0%)	1,00
não casados	71 (43,8%)	1,15 (0,72-1,82)

## 3. Distribuições de probabilidade

### 3.1 - Noções de probabilidade

**Probabilidade** (*probability, chance, likelihood*)

- É uma afirmação numérica sobre a possibilidade de que algum evento ocorra.
- Quantifica o grau de incerteza de eventos, variando de 0 (0%) a 1 (100%).
- Um evento impossível de ocorrer tem probabilidade 0 (zero)
- Um evento certo tem probabilidade 1 (um)
- Quando se joga uma moeda, não se sabe se vai sair cara. Mas sabe-se que a probabilidade de sair cara é  $0,5 = 50\% = 1/2$ .
- Dizer que a eficácia de uma vacina é de 70% corresponde a dizer que cada indivíduo vacinado tem probabilidade 0,7 de ficar imune.

Probabilidade em espaços finitos contáveis

Espaço amostral (S)

- É o conjunto de todos os resultados possíveis de um experimento.
- Supor o experimento lançar uma moeda;  $S = \{\text{cara, coroa}\}$

Há dois pontos neste espaço amostral, sendo um favorável ao evento  $A = \{\text{cara}\}$ .

## Definição clássica de probabilidade

$$P(A) = \frac{\text{numero de elementos de } A}{\text{numero de elementos de } S} = \frac{1}{2} = 0,5$$

**Exemplo:** probabilidade de (ouros) em um baralho de 52 cartas  $\frac{13}{52} = \frac{1}{4}$

### Probabilidade de eventos mutuamente excludentes

- Diz-se que dois eventos são mutuamente excludentes (ou mutuamente exclusivos) quando não podem ocorrer simultaneamente

Exemplo:

A = {cara} ; B= {coroa}, no lançamento de uma moeda; A e B são mutuamente exclusivos

C = {sexo feminino}; D={sexo masculino}, no nascimento de uma criança; C e D são mutuamente exclusivos.

- A probabilidade da ocorrência de um evento A ou de um evento B é:

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

- A probabilidade da ocorrência simultânea de eventos mutuamente exclusivos é zero.

$P(\text{cara e coroa}) = P(\text{cara} \cap \text{coroa}) = 0$ , no lançamento de uma moeda.

- Se A e B forem mutuamente excludentes,  $P(A \cap B) = 0$  e

$$P(A \text{ ou } B) = P(A \cup B) = P(A) + P(B)$$

Exemplo 1:

P(Face 2 ou Face 3) no lançamento de um dado

$$P(2 \text{ ou } 3) = P(2) + P(3) = 1/6 + 1/6 = 2/6 = 1/3.$$

$$P(\text{Resultado ímpar}) = P(1 \text{ ou } 3 \text{ ou } 5) = P(1) + P(3) + P(5) = 3/6 = 1/2.$$

Regra da adição: $P(A \text{ ou } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$
---

### Probabilidade de eventos independentes

- Os eventos A e B são independentes quando o resultado de um não influi no resultado do outro.

Exemplo: no lançamento simultâneo de duas moedas, o resultado de uma não interfere no resultado da outra.

- A probabilidade da ocorrência de eventos independentes é o produto das probabilidades de cada evento.

$$P(A \text{ e } B) = P(A \cap B) = P(A) \times P(B)$$

- P(face 2 no primeiro dado e face 3 no segundo dado), no lançamento seqüencial de dois dados =  $P(2 \text{ e } 3) = P(2) \times P(3) = 1/6 \times 1/6 = 1/36 = 0,0278 = 2,78\%$ .

### Probabilidade condicional

A probabilidade condicional do evento A dado que ocorreu o evento B é

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ para } P(B) \neq 0$$

Lê-se  $P(A|B)$  como probabilidade de A dado B.

### Exemplo:

Número de adolescentes segundo história de bronquite aos 5 anos e tosse diurna ou noturna aos 14 anos de idade. Local X, ano Y.

Tosse	Bronquite		Total
	Sim	Não	
Sim	26	44	70
Não	247	1002	1249
Total	273	1046	1319

Fonte: Holland, WW *et al.*, 1978.

Sorteia-se um paciente. Qual é a probabilidade dele ter tosse aos 14 anos dado que teve bronquite aos 5 anos de idade?

$$P(\text{tosse}|\text{bronquite}) = P(\text{tosse e bronquite})/P(\text{bronquite}) = 26/1319 \div 273/1319 = 9,5\%$$

Regra da multiplicação

$$P(A \cap B) = P(A|B) \times P(B)$$

se A e B forem independentes,  $P(A|B) = P(A)$  e como consequência,  $P(A \cap B) = P(A) \times P(B)$

Os eventos tosse e bronquite não são independentes porque  $P(\text{tosse e bronquite}) \neq P(\text{tosse}) \times P(\text{bronquite})$ , pois  $26/1319 \neq (70/1319) \times (273/1319)$ , ou seja,  $0,02 \neq 0,011$ .

### Exemplo:

Considerar uma população de homens que foram classificados segundo o hábito de fumar e doença respiratória crônica. Nesta população sabe-se que 5% dos homens têm doença respiratória e são não fumantes, 15% têm doença e são fumantes, 50% não têm doença e são não fumantes e 30% não têm a doença e são fumantes.

Problema respiratório	Não fumante $\bar{F}$	Fumante $F$	Total
Não ( $\bar{R}$ )	0,50 = $P(\bar{F} \bar{R})$	0,30 = $P(F \bar{R})$	0,80 = $P(\bar{R})$
Sim ( $R$ )	0,05 = $P(\bar{F} R)$	0,15 = $P(F R)$	0,20 = $P(R)$
Total	0,55 = $P(\bar{F})$	0,45 = $P(F)$	

Escolhe-se um homem ao acaso, qual a probabilidade dele ter doença respiratória dado que era fumante?

$$P(R|F) = \frac{P(R \cap F)}{P(F)} = 0,15/0,45 = 0,33$$

Os eventos não são independentes porque  $P(\bar{F}\bar{R}) \neq P(\bar{F}) \times P(\bar{R})$

Relação entre eventos mutuamente exclusivos e independentes:

Os eventos mutuamente exclusivos A e B satisfazem a condição que  $P(A \text{ e } B) = 0$ , então dois eventos mutuamente exclusivos A e B são não independentes a menos que  $P(A)=0$  ou  $P(B)=0$ . Caso contrário, eles são claramente dependentes pois  $P(A).P(B)>0$  se ambos  $P(A)>0$  e  $P(B)>0$ , portanto  $P(A \cap B) \neq P(A)P(B)$  porque  $P(A \cap B) = 0$ .

Assim, dois eventos mutuamente exclusivos A e B são dependentes exceto nos casos onde  $P(A)=0$  ou  $P(B)=0$ .

Definição freqüentista de probabilidade:

n repetições do evento A; A ocorre m vezes, então a freqüência relativa de  $A = \frac{m}{n}$

Para n suficientemente grande,  $\frac{m}{n} \cong P(A)$  ou seja,  $\lim_{n \rightarrow \infty} \frac{m}{n} = P(A)$

Quando n cresce,  $\frac{m}{n}$  tende a se estabilizar em torno de uma constante,  $P(A)$

## Variável aleatória discreta

Variável aleatória é qualquer função de número real, definida no espaço amostral, com uma probabilidade de ocorrência associada.

### Exemplo:

No lançamento de 1 moeda, o número de caras é uma variável aleatória. Se esta variável for denominada  $X$ , tem-se que os valores possíveis para  $X$  são 0 e 1. Assim escreve-se  $X:0,1$ .

A probabilidade de cara é 0,5:  $P(\text{cara}) = P(X=1) = 0,5 = 1/2$ .

No lançamento de 10 moedas,  $X:0, 1, 2, \dots, 10$ ; e a probabilidade de cara = 0,5.

Sair cara (ou coroa) são eventos mutuamente exclusivos. Um particular resultado de cada lançamento exclui a ocorrência do outro.

É possível calcular a probabilidade da variável assumir cada valor  $x$ , ou seja,  $P(X=x)$ .

O conjunto de valores da variável aleatória e das probabilidades obtidas define uma **distribuição de probabilidades**. Se  $X$  assume valores inteiros, a variável é denominada discreta. Se  $X$  assume valores no conjunto dos números reais, a variável é denominada contínua.

## 3.2 - Distribuição Bernoulli

Estrutura básica: duas possibilidades de resultado (sucesso e fracasso).

### Exemplo 1:

Joga-se uma moeda uma vez. A moeda é equilibrada, ou seja, os lados possuem peso igual, não favorecendo nenhum dos lados, ao ser lançada. Tem-se como sucesso sair a face cara.

Define-se uma variável aleatória  $X$  que assume valor 1 se ocorrer sucesso e 0 se ocorrer fracasso.

$X: 0,1$

Parâmetro: probabilidade da variável assumir valor 1.

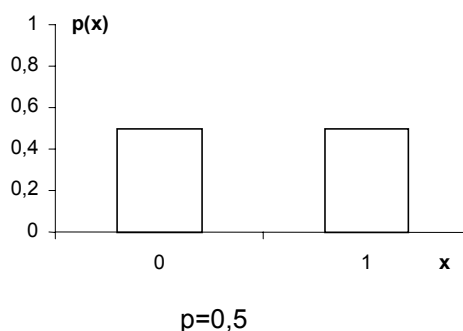
Notação:  $\pi$  ou  $p$ .

Se probabilidade de sucesso =  $p$ , a probabilidade de fracasso será igual a  $q=(1-p)$ , porque  $p+q=1$ .

Probabilidade de sair cara =  $P(X=1) = p(1) = p = 0,5$

Probabilidade de sair coroa =  $P(X=0) = p(0) = q = 1-p = 0,5$

Graficamente:

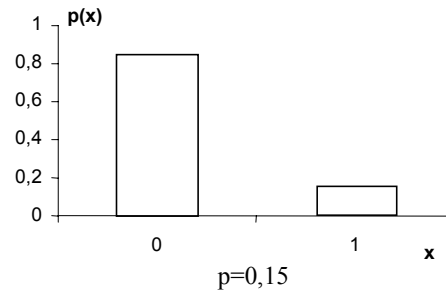


### Exemplo 2:

Uma droga cura 15% dos pacientes. Administra-se a droga a um paciente. Qual a probabilidade do paciente ficar curado? Qual a probabilidade do paciente não ficar curado?

$X: 0,1$  ( $X$  será 0 se o paciente não se curar e 1 se houver cura)

$$P(X=1) = p(1) = p = 0,15 ; P(X=0) = p(0) = q = 0,85$$



Os exemplos pertencem a mesma família de distribuições, mas têm parâmetros diferentes.

A distribuição de Bernoulli pode ser escrita como  $P(X=1) = p(1) = p$  e  $P(X=0) = p(0) = 1-p$ ; ou, de forma mais genérica,  $p(x) = p^x (1 - p)^{1-x}$ ,  $x=0,1$

Isto significa que para  $x=0$ ,  $p(0) = P(X = 0) = p^0 (1 - p)^{1-0} = 1 - p$  ,  
 para  $x=1$ ,  $p(1) = P(X = 1) = p^1 (1 - p)^{1-1} = p$

Média de uma variável aleatória discreta:  $\mu = E(X) = \sum_x xp(x)$

Na distribuição de Bernoulli:

$$\mu = E(X) = \sum_x xp(x) = 1p(x = 1) + 0p(x = 0) = p$$

Média da distribuição Bernoulli é  $p$  (probabilidade de ocorrer o sucesso)

Variância de uma variável aleatória discreta:

$$\sigma^2 = V(X) = E[(X - \mu)^2] = \sum_x (x - \mu)^2 p(x)$$

Desvio padrão:  $SD(X) = \sqrt{V(X)} = \sigma$

Desvio padrão da distribuição Bernoulli é

$$\sqrt{(0 - p)^2 \cdot p(x = 0) + (1 - p)^2 \cdot p(x = 1)} = \sqrt{(-p)^2 \cdot (1 - p) + (1 - p)^2 p} = \sqrt{(1 - p)p(p + (1 - p))} = \sqrt{p(1 - p)} = \sqrt{pq}$$

Resumindo,

**Modelo de probabilidade Bernoulli**

Uma variável aleatória discreta  $X$  que pode assumir valores 0 e 1, com função de probabilidade dada por

$$p(x) = p^x (1 - p)^{1-x} \text{ com } x=0,1$$

segue uma distribuição Bernoulli com parâmetro  $p$ ,  $0 < p < 1$ .

$p$  é a probabilidade de obter o resultado  $X=1$ . Isto pode ser escrito como  $X \sim \text{Bernoulli}(p)$  com média  $p$  e desvio padrão  $\sqrt{p(1 - p)}$ .

O símbolo  $\sim$  lê-se "tem distribuição" ou se "distribui segundo".

**3.3 - Distribuição binomial**

Distribuição Binomial é a soma de  $n$  distribuições Bernoulli



População: 2 categorias

Ex: sexo (masculino, feminino),  
 faces de uma moeda (cara, coroa),  
 desfecho de um tratamento (cura, não cura)

Lançamento de uma moeda  $\left\{ \begin{array}{l} \text{Cara (K)} \rightarrow \text{probabilidade(K)} = p \\ \text{Coroa (C)} \rightarrow \text{probabilidade (C)} = q \\ p + q = 1 \Rightarrow q = 1 - p \end{array} \right.$

$p$  = probabilidade de sucesso;  $q$  = probabilidade de fracasso

Realiza-se o experimento  $n$  vezes, onde cada ensaio é independente do outro e os resultados são mutuamente exclusivos.

$X$ : Número de vezes que sai cara

A moeda é lançada uma vez ( $n=1$ )  $\rightarrow X: 0,1 \quad X \sim \text{Bernoulli}(p)$

X	resultado	$P(X=x)$
0	C	$P(X=0) = q$
1	K	$P(X=1) = p$

A moeda é lançada duas vezes ( $n=2$ )  $\rightarrow X: 0,1,2 \quad X \sim B(n=2, p)$

X	resultado	$P(X=x)$
0	C,C	$P(X=0) = q \cdot q = q^2$
1	K,C ou C,K	$P(X=1) = p \cdot q + q \cdot p = 2 \cdot p \cdot q$
2	K,K	$P(X=2) = p \cdot p = p^2$

A moeda é lançada três vezes ( $n=3$ )  $\rightarrow X: 0,1,2,3 \quad X \sim B(n=3, p)$

X	resultado	$P(X=x)$
0	C,C,C	$P(X=0) = q \cdot q \cdot q = q^3$
1	K,C,C ou C,K,C ou C,C,K	$P(X=1) = p \cdot q \cdot q + q \cdot p \cdot q + q \cdot q \cdot p = 3 p \cdot q^2$
2	K,K,C ou K,C,K ou C,K,K	$P(X=2) = p \cdot p \cdot q + p \cdot q \cdot p + q \cdot p \cdot p = 3 p^2 \cdot q$
3	K,K,K	$P(X=3) = p \cdot p \cdot p = p^3$

Probabilidade ( $X=x$ ) é calculada pelo produto de 3 fatores:

1º - número (combinação de  $n$  elementos combinados  $x$  a  $x$ )

2º - probabilidade de sucesso elevado a um expoente (valor de  $x$ )

3º - probabilidade de fracasso elevado a um expoente (valor de  $n-x$ )

$$P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

Resumindo

### Modelo de probabilidade Binomial

Seja  $E$  um experimento com 2 resultados (mutuamente exclusivos):  $S$  (sucesso) e  $F$  (fracasso)

$p$  = probabilidade de ocorrência de  $S$  e  $q$  = probabilidade de ocorrência de  $F$

sendo que  $p+q=1$ .

Se E for repetido n vezes, de forma independente, mantendo-se p e q, a probabilidade da variável aleatória X= número de vezes que S ocorre é dada por

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$X \sim B(n,p)$  onde n e p são os parâmetros da

distribuição; a média =  $m = n.p$ , a variância =  $n.p.q$  e o desvio padrão =  $\sqrt{npq}$

### Exemplo 1:

Uma suspensão contendo organismos de Leishmania é preparada e quando uma determinada quantidade é inoculada em ratos, 30% deles se tornam infectados. Se 3 ratos forem inoculados independentemente, qual a probabilidade de:

a) Nenhum rato ficar infectado?

$$P(X=0) = \binom{3}{0} (0,3)^0 (0,7)^3 = \frac{3!}{0!(3-0)!} (0,7)^3 = 1 \times 0,343 = 0,343 = 34,3\%$$

b) Um rato ficar infectado?

$$P(X=1) = \binom{3}{1} (0,3)^1 (0,7)^{3-1} = \frac{3!}{1!(3-1)!} (0,3)^1 (0,7)^{3-1} = \frac{3 \times 2 \times 1}{1 \times 2 \times 1} 0,3 \times 0,49 = 0,441 = 44,1\%$$

c) Dois ratos ficarem infectados?

$$P(X=2) = \binom{3}{2} (0,3)^2 (0,7)^{3-2} = \frac{3!}{2!(3-2)!} (0,3)^2 (0,7)^{3-2} = \frac{3 \times 2 \times 1}{2 \times 1 \times 1} 0,09 \times 0,7 = 0,189 = 18,9\%$$

d) Todos os ratos ficarem infectados?

$$P(X=3) = \binom{3}{3} (0,3)^3 (0,7)^{3-3} = \frac{3!}{3!(3-3)!} (0,3)^3 (0,7)^0 = \frac{3 \times 2 \times 1}{3 \times 2 \times 1 \times 1} 0,027 \times 1 = 0,027 = 2,7\%$$

e) Pelo menos 2 fiquem infectados?

f) No máximo 1 fique infectado?

### Exemplo 2:

Lançamento de moedas.

- n= número de ensaios (nº de lançamentos)= 10
- X= variável aleatória (nº de caras)
- x= resultado particular de X (0, 1, 2, ...,10)
- p= probabilidade de ocorrer cara (sucesso); p=P(cara)= 0,5

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Distribuição de probabilidade B(n=10; p=0,5)

X= nº de caras	P(X=x)
0	0,0010
1	0,0098
2	0,0439
3	0,1172
4	0,2051
5	0,2461
6	0,2051
7	0,1172
8	0,0439
9	0,0098
10	0,0010

Média = np = 10x0,5 = 5

Variância = npq = 2,5

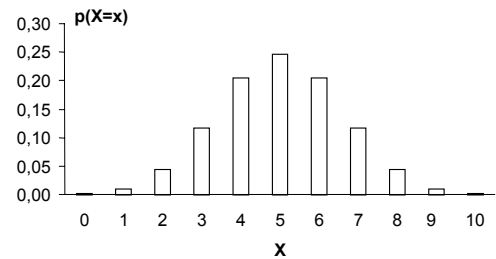
Desvio padrão =  $\sqrt{npq} = \sqrt{10x0,5x0,5} = \sqrt{2,5} = 1,58$

Se estivermos trabalhando com a proporção de sucessos,  $\frac{X}{n}$  :

Média =  $n \times \frac{p}{n} = p = 0,5$

Variância =  $n \times \frac{p}{n} \times \frac{q}{n} = \frac{pq}{n} = 0,025$

Desvio padrão =  $\sqrt{\frac{npq}{n}} = \sqrt{\frac{pq}{n}} = 0,158$

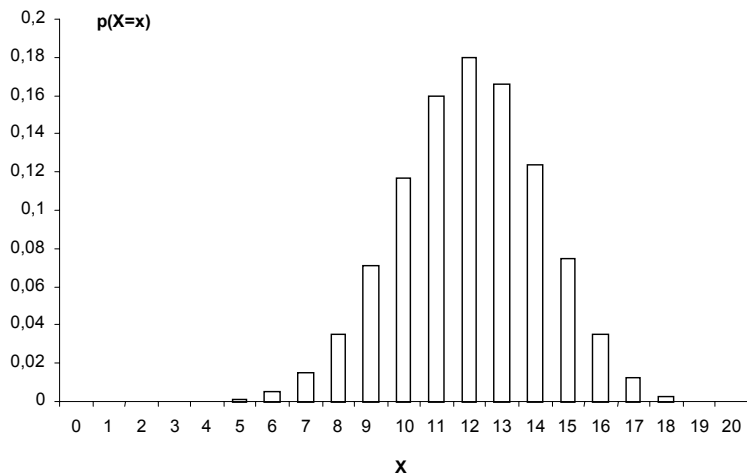


**Exemplo 3**

Um programa de incentivo à amamentação exclusiva ao seio nos primeiros 3 meses está sendo executado em um hospital universitário. Verificou-se que a eficácia do programa era de  $\pi = 60\%$ .

Para uma amostra de 20 mães que deram à luz neste hospital, a distribuição de probabilidade da variável aleatória número de mães amamentando exclusivamente ao seio é a seguinte:

X= nº de mães amamentando	P(X=x p=0,6)
0	0,000
1	0,000
2	0,000
3	0,000
4	0,000
5	0,001
6	0,005
7	0,015
8	0,035
9	0,071
10	0,117
11	0,160
12	0,180
13	0,166
14	0,124
15	0,075
16	0,035
17	0,012
18	0,003
19	0,000
20	0,000



### 3.4 - Odds e probabilidade

#### Probabilidade:

Supor o estudo de seguimento de 50 fumantes. No final do estudo observa-se que 30 pessoas desenvolveram bronquite crônica e 20 não desenvolveram.

Denominando-se  $p$  a probabilidade de desenvolver bronquite e  $q$  a probabilidade de não desenvolver bronquite, tem-se que  $\hat{p} = \frac{30}{50}$  e  $\hat{q} = \frac{20}{50}$  onde  $\hat{q} = 1 - \hat{p}$ .

#### Odds:

Define-se *odds* a favor de bronquite como a razão entre a probabilidade de desenvolver bronquite e a probabilidade de não desenvolver a doença. Portanto,  $odds = \frac{p}{1-p}$

Aplicando-se os dados do exemplo tem-se

$$odds \text{ a favor de bronquite} = \frac{\hat{p}}{1-\hat{p}} = \frac{\frac{30}{50}}{\frac{20}{50}} = \frac{30}{20} = 1,5 : 1 \quad \text{ou simplesmente } 1,5.$$

Lê-se o *odds* da seguinte forma: entre fumantes a probabilidade de desenvolver bronquite é 50% maior que a probabilidade de não desenvolver bronquite entre não fumantes ou a probabilidade de desenvolver bronquite entre fumantes é uma vez e meia a probabilidade de não desenvolver bronquite entre não fumantes.

#### Probabilidade escrita como função de Odds

Tem-se, ainda, que a probabilidade de doença pode ser escrita como *odds*, da seguinte forma:

$$p = \frac{odds}{1 + odds} \quad \text{e de fato, a probabilidade de bronquite é igual a } \hat{p} = \frac{odds}{1 + odds} = \frac{\frac{30}{20}}{1 + \frac{30}{20}} = \frac{\frac{30}{20}}{\frac{50}{20}} = \frac{30}{50}.$$

#### Exemplo

Com base no exemplo da tosse e bronquite tem-se que:

o odds de uma criança ter tido bronquite na infância é  $273/1046=0,261$ . A probabilidade de bronquite é igual a  $0,261/(1+0,261)=0,207=20,7\%$ .

ou

o odds de uma criança não ter tido bronquite na infância é  $1046/273=3,8$ . A probabilidade de não ter tido bronquite na infância é igual a  $3,8/(1+3,8)=0,792=79,2\%$ .

#### Questão 24

Considerando um estudo conduzido para comparar os efeitos adversos e a eficácia de duas drogas antidepressivas, 18 pacientes são alocados aleatoriamente em um de três grupos (6 pacientes em cada grupo). Os pacientes do primeiro grupo são tratados com a droga A, os do segundo, com a droga B e os do terceiro, com placebo. Estudos anteriores sugerem que a proporção de abandono de tratamento entre os pacientes do grupo placebo é 15%. Pressupondo-se que exista independência entre cada realização do experimento, calcule:

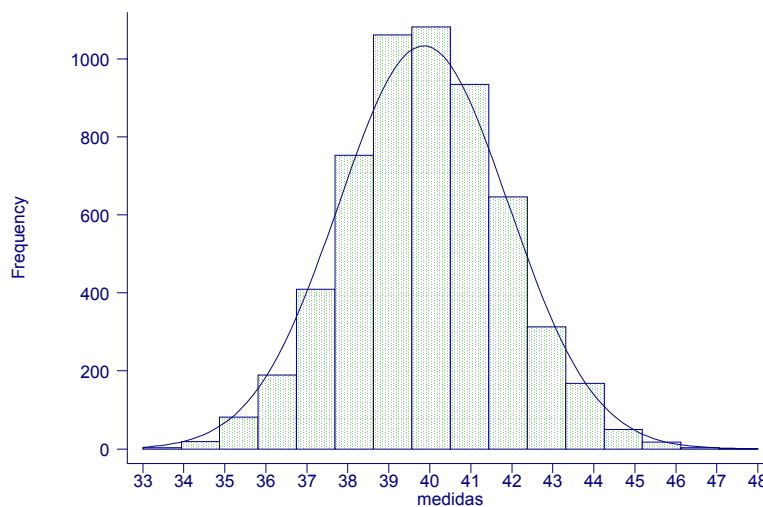
- a) a probabilidade que todos os 6 pacientes do grupo 3 irão abandonar o estudo;
- b) a probabilidade que nenhum paciente do grupo 3 abandonará o estudo;
- c) a probabilidade que 2 pacientes do grupo 3 abandonem o estudo;

### 3.5- Distribuição normal ou de Gauss

Os dados abaixo são medidas do tórax (polegadas) de 5732 soldados escoceses, tomadas pelo matemático belga, Adolphe Quetelet (1796-1874).

medidas	Freq.	Percent	Cum.
33	3	0.05	0.05
34	19	0.33	0.38
35	81	1.41	1.80
36	189	3.30	5.09
37	409	7.14	12.23
38	753	13.14	25.37
39	1062	18.53	43.89
40	1082	18.88	62.77
41	935	16.31	79.08
42	646	11.27	90.35
43	313	5.46	95.81
44	168	2.93	98.74
45	50	0.87	99.62
46	18	0.31	99.93
47	3	0.05	99.98
48	1	0.02	100.00
Total	5732	100.00	

Distribuição de medidas do tórax (polegadas) de soldados escoceses



Fonte: Daly F *et al. Elements of Statistics*, 1999

Função densidade de probabilidade da distribuição normal:

Se a variável aleatória  $X$  é normalmente distribuída com média  $\mu$  e desvio padrão  $\sigma$  (variância  $\sigma^2$ ), a

função densidade de probabilidade de  $X$  é dada por 
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]}, \quad -\infty < x < +\infty$$

onde

$\pi$  : constante  $\cong 3,1416$

$e$ : constante  $\cong 2,718$

$\mu$  : constante (média aritmética da população)

$\sigma$  : constante (desvio padrão populacional)

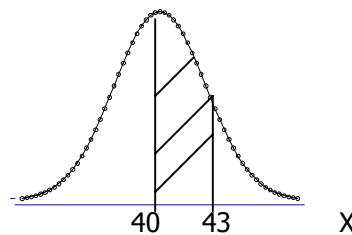
Propriedades:

- campo de variação :  $-\infty < X < +\infty$
- é simétrica em torno da média  $m$  (ou  $\mu$ )
- a média e a mediana são coincidentes
- a área total sob a curva é igual a 1 ou 100%
- a área sob a curva pode ser entendida como medida de probabilidade

$$\left\{ \begin{array}{l} \bar{X} \pm \sigma = 68,2\% \\ \bar{X} \pm 1,96 \sigma = 95,0\% \\ \bar{X} \pm 2,58 \sigma = 99,0\% \end{array} \right.$$

**Exemplo:**

Depois de tomarmos várias amostras, decidiu-se adotar um modelo para as medidas de perímetro do tórax de uma população de homens adultos com os parâmetros: média ( $\mu$ ) = 40 polegadas e desvio padrão ( $\sigma$ ) = 2 polegadas



Qual a probabilidade de um indivíduo, sorteado desta população, ter um perímetro de tórax entre 40 e 43 polegadas?

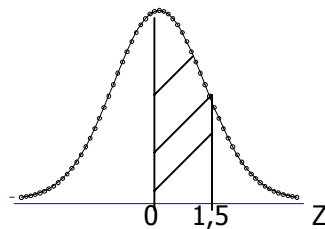
$$P(40 < X < 43) = \int_{40}^{43} \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-40)^2}{2 \cdot 4}} dx$$

Pode-se responder à pergunta sem calcular a integral e sim utilizando o número de desvios padrão que o valor 43 está em torno da média para calcular a probabilidade por meio da tabela da distribuição Normal reduzida  $N(0,1)$ .

**Normal reduzida:**

$$Z \sim N(0;1) \quad \text{onde } Z = \frac{X - \mu}{\sigma}$$

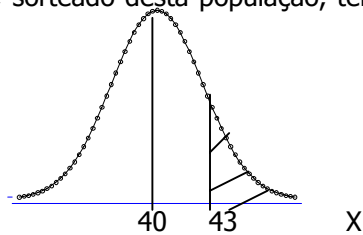
$$P(40 < X < 43) = P\left(\frac{40 - 40}{2} < \frac{X - \mu}{\sigma} < \frac{43 - 40}{2}\right) = P(0 < Z < 1,5)$$



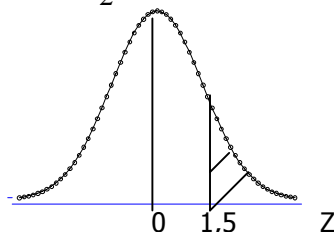
Utilizando a tabela da curva normal reduzida,  $P(0 < Z < 1,5) = 0,43319 = 43,3\%$

Com base na distribuição de  $X \sim N(\mu = 40, \sigma = 2)$ , do exercício 34, calcular:

a) a probabilidade de um indivíduo, sorteado desta população, ter um perímetro de tórax maior ou igual a 43 polegadas.



$$P(X > 43) = P\left(\frac{X - \mu}{\sigma} > \frac{43 - 40}{2}\right) = P(Z > 1,5)$$



Utilizando a tabela da curva normal reduzida,  $P(Z > 1,5) = 0,5 - 0,43319 = 0,06681 = 6,7\%$

b) a probabilidade de um indivíduo, sorteado desta população, ter um perímetro de tórax entre 35 e 40 polegadas;

c) a probabilidade de um indivíduo, sorteado desta população, ter um perímetro de tórax menor que 35;

d) Qual o valor do perímetro do tórax, que deixaria 75% da população abaixo dele?

### Questão 25

Considerar a altura de 351 mulheres idosas como seguindo uma distribuição normal com média 160cm e desvio padrão 6 cm. Sorteia-se uma mulher; qual a probabilidade de que ela tenha

- altura entre 160 cm e 165 cm?
- altura maior do que 170 cm?
- altura menor do que 150 cm?

### 3.6- Distribuição amostral da média

Apresentação do Teorema que fornece as bases teóricas necessárias para inferência estatística e estimação.

#### Teorema central do limite:

X é variável aleatória com média  $\mu$  e variância  $\sigma^2$ , então  $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

Supor a situação onde uma população é composta por 6 elementos, para os quais observou-se a característica X, cujos valores estão apresentados abaixo.

elementos	$X_i$
A	11
B	16
C	12
D	15
E	16
F	14

Fonte: Dixon WJ & Massey FJ, 1957.

Média populacional ( $\mu$ ) = 14

Variância populacional ( $\sigma^2$ ) = 3,667

Desvio padrão populacional ( $\sigma$ ) = 1,9149

Parâmetros população	valor	Estimador amostra	Valor (estimativa) Par(A,D)=(11,15)
Média ( $\mu$ )	14	$\bar{X}$	$\bar{x} = 13$
Variância ( $\sigma^2$ )	3,67	$S^2$	$s^2 = 8$
Desvio padrão ( $\sigma$ )	1,91	S	$s = 2,828$

Todas as possíveis amostras de tamanho 2, determinadas pelo processo de amostragem aleatório, com reposição (N=6, n=2)

Amostra	Elementos que compõem a amostra	valores	Média( $\bar{x}_i$ )	Amostra	Elementos que compõem a amostra	valores	Média( $\bar{x}_i$ )
1	A,A	(11,11)	11	19	D,A	(15,11)	13
2	A,B	(11,16)	13,5	20	D,B	(15,16)	15,5
3	A,C	(11,12)	11,5	21	D,C	(15,12)	13,5
4	A,D	(11,15)	13	22	D,D	(15,15)	15
5	A,E	(11,16)	13,5	23	D,E	(15,16)	15,5
6	A,F	(11,14)	12,5	24	D,F	(15,14)	14,5
7	B,A	(16,11)	13,5	25	E,A	(16,11)	13,5
8	B,B	(16,16)	16	26	E,B	(16,16)	16
9	B,C	(16,12)	14	27	E,C	(16,12)	14
10	B,D	(16,15)	15,5	28	E,D	(16,15)	15,5
11	B,E	(16,16)	16	29	E,E	(16,16)	16
12	B,F	(16,14)	15	30	E,F	(16,14)	15
13	C,A	(12,11)	11,5	31	F,A	(14,11)	12,5
14	C,B	(12,16)	14	32	F,B	(14,16)	15
15	C,C	(12,12)	12	33	F,C	(14,12)	13
16	C,D	(12,15)	13,5	34	F,D	(14,15)	14,5
17	C,E	(12,16)	14	35	F,E	(14,16)	15
18	C,F	(12,14)	13	36	F,F	(14,14)	14



Distribuição de frequência de todas as possíveis médias:

Distribuição amostral da média		
i	$\bar{x}_i$	freqüência
1	11	1
2	11,5	2
3	12	1
4	12,5	2
5	13	4
6	13,5	6
7	14	5
8	14,5	2
9	15	5
10	15,5	4
11	16	4
Total		36

$$\text{Média das médias } (\bar{\bar{x}}) = \frac{\sum_{i=1}^{11} \bar{x}_i f_i}{n} = 14$$

$$\text{Variância das médias } \sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{11} (\bar{x}_i - \bar{\bar{x}})^2 f_i}{n} = 1,833$$

$$\text{Desvio padrão das médias} = \text{erro padrão da média} = \sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2}$$

$$\text{Erro padrão da média} = \sqrt{1,833} = 1,354$$

No exemplo,  $X \sim N(\mu = 14, \sigma = 1,915)$ , portanto  $\bar{X} \sim N(\mu_{\bar{x}} = 14, \sigma_{\bar{x}} = \frac{1,915}{\sqrt{2}} = 1,354)$

### Questão 26

Entre homens adultos sadios, a concentração de ferro sérico segue uma distribuição normal com média 120 microgramas para 100ml e desvio padrão 15 microgramas para 100ml. Calcule a probabilidade que uma amostra de 50 homens resulte em nível médio de ferro sérico entre 115 e 125 microgramas por 100ml.

## 4. Estimação de parâmetros populacionais por ponto e intervalo

### 4.1 – Estimação por ponto

Seja  $X_1, X_2, X_3, \dots, X_n$  uma amostra aleatória de tamanho  $n$  de uma distribuição normal com média  $\mu$  e variância  $\sigma^2$  (desvio padrão  $\sigma$ ).

Os parâmetros ( $\mu$  e  $\sigma^2$ ) podem ser estimados com base na amostra.

Se o estimador for um único valor, a estimação é chamada de estimação por ponto.

Média aritmética

Populacional

Parâmetro  $\mu$                       estimador :  $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$

Variância

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

Parâmetro  $\sigma^2$

estimador :

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

Atenção: os estimadores são variáveis aleatórias. Pelo teorema central do limite,  $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ .

## 4.2 - Estimação por intervalo

Padronizando-se a média  $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ , obtém-se  $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ , que permite calcular

$P(-z \leq Z \leq z) = 1 - \alpha$ , ou seja,  $P(-z \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z) = 1 - \alpha$ . Para  $\alpha = 5\%$ , tem-se:

$P(-1,96 \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq +1,96) = 0,95$ , com manipulação algébrica tem-se:

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

obtendo-se  $P(\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}) = 0,95$ .

Tem-se um intervalo aleatório centrado na média amostral que possui 95% de probabilidade de conter a verdadeira média populacional.

Neste caso o parâmetro será estimado por um conjunto de valores provenientes de uma amostra. Quando isto é feito,  $\hat{X} = \bar{x}$  e o intervalo deixa de ser uma variável aleatória.

O intervalo  $\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}}$  contém ou não contém a verdadeira média. Diz-se então que a

confiança que se deposita neste intervalo é de 95% porque antes de tomada a **amostra** de tamanho  $n$ , existia, associada a ele, uma probabilidade de 95% de que contivesse a média populacional. Por isso chama-se intervalo de confiança para a média populacional

IC(95%) :  $(\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}})$  e  $(\bar{x} + 1,96 \frac{\sigma}{\sqrt{n}})$  ou  $\bar{x} \pm 1,96 \frac{\sigma}{\sqrt{n}}$

#### 4.2.1 - Intervalo de confiança para a média populacional com variância populacional conhecida

$$IC(\mu) = \bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}}$$

#### 4.2.2 - Intervalo de confiança para a média populacional com variância populacional desconhecida

Define-se a estatística

$$T = \frac{\bar{X} - \mu}{S_{\bar{X}}} = \frac{\bar{X} - \mu}{S_x / \sqrt{n}}, \text{ onde } T \sim t_{n-1}, \text{ n - 1: graus de liberdade}$$

obs : quando  $n \rightarrow \infty \Rightarrow T \sim N(0 ; 1)$

T é uma variável aleatória que segue uma distribuição *t* de "Student", com (n-1) graus de liberdade. *Student* é o pseudônimo de *W. S. Gosset* que, em 1908, propôs esta distribuição. T segue uma família de distribuições, dependendo do tamanho da amostra.

Assim como Z, T é simétrica em torno do zero. Quanto maior o número de graus de liberdade, mais a distribuição *t*(n-1) vai se aproximar da distribuição normal.

$$IC(\mu) = \bar{x} \pm t_{n-1, \alpha/2} \cdot \frac{S_x}{\sqrt{n}}$$

#### 4.2.3 - Intervalo de confiança aproximado para a proporção populacional

Seja X uma variável aleatória que segue uma distribuição binomial;  $X \sim B(n, \pi)$ .

A proporção de sucessos  $\pi = \frac{X}{N}$ , é estimada na amostra por  $p = \frac{X}{n}$

Para n grande,  $\pi \sim N(\mu = p, \sigma = \sqrt{\frac{pq}{n}})$ .

Intervalo de confiança para:  $IC(\pi) = p \pm z_{\alpha/2} \sqrt{\frac{pq}{n}}$ .

Onde p e q são proporções estimadas na amostra.

### Exemplo

Em uma pesquisa sobre propriedades teratogênicas de uma droga, foram observadas 85 camundongos fêmeas grávidas que não foram expostas a droga. Do total de 85 ninhadas, 12 tiveram pelo menos 1 filhote com malformação. Apresente o intervalo de confiança de 95% para a verdadeira proporção de malformação na população de camundongos de onde os 85 que participaram do estudo foram provenientes.

IC(95%) aproximado para a proporção populacional: 
$$\hat{p} \pm 1,96 \sqrt{\frac{\hat{p}\hat{q}}{n}} = \frac{12}{85} = 0,1412$$

IC(95%) 
$$0,1412 \pm 1,96 \sqrt{\frac{0,1412 \times 0,8588}{85}} ; (0,067 - 0,2153)$$

### Exemplo

Os dados a seguir são provenientes do grupo *Western Collaborative Group Study*, criado na Califórnia em 1960-61. Foram estudados 3154 homens de meia idade para investigar a relação entre padrões de comportamento e risco de doença coronariana. Os dados apresentados são de 40 homens para os quais foram medidos os níveis de colesterol (mg por 100ml) e realizada uma categorização segundo comportamento. O comportamento de tipo A é caracterizado pela urgência, agressividade e ambição. O de tipo B é relaxado, não competitivo e menos preocupado.

Tipo A: nível de colesterol

233	291	312	250	246	197	268	224	239	239
54	276	234	181	248	252	202	218	212	325

Tipo B: nível de colesterol

344	185	263	246	224	212	188	250	148	169
226	175	242	252	153	183	137	202	194	213

Medidas(mg/100ml)	Tipo A	Tipo B
Média	245,05	210,3
Variância (n-1)	1342,37	2336,747
Desvio padrão (n-1)	36,64	48,33991

Com base nos valores fornecidos, calcule o intervalo de confiança para o nível médio populacional de colesterol para cada um dos tipos. Compare os resultados e discuta.

IC(95%) para a média populacional com desvio padrão populacional desconhecido: 
$$\bar{x} \pm t_{n-1} \frac{S}{\sqrt{n}}$$

IC(95%) Tipo A:  $245,05 \pm 2,093 \frac{36,64}{\sqrt{20}} ; (227,9 - 261,2)$

IC(95%) Tipo B: ;

### Questão 27

Em uma pesquisa sobre saúde bucal na Inglaterra observou-se que de uma amostra de 262 mulheres com idade entre 25 e 34 anos, 7,6% não tinham dentes. Calcule o intervalo de confiança de 95% para a proporção de mulheres inglesas naquela faixa etária, que não possuem dentes.

### Questão 28

São apresentadas medidas de pressão arterial sistólica de uma amostra aleatória de 20 pacientes.

Construa o intervalo de confiança de 90% para a pressão sistólica média populacional.

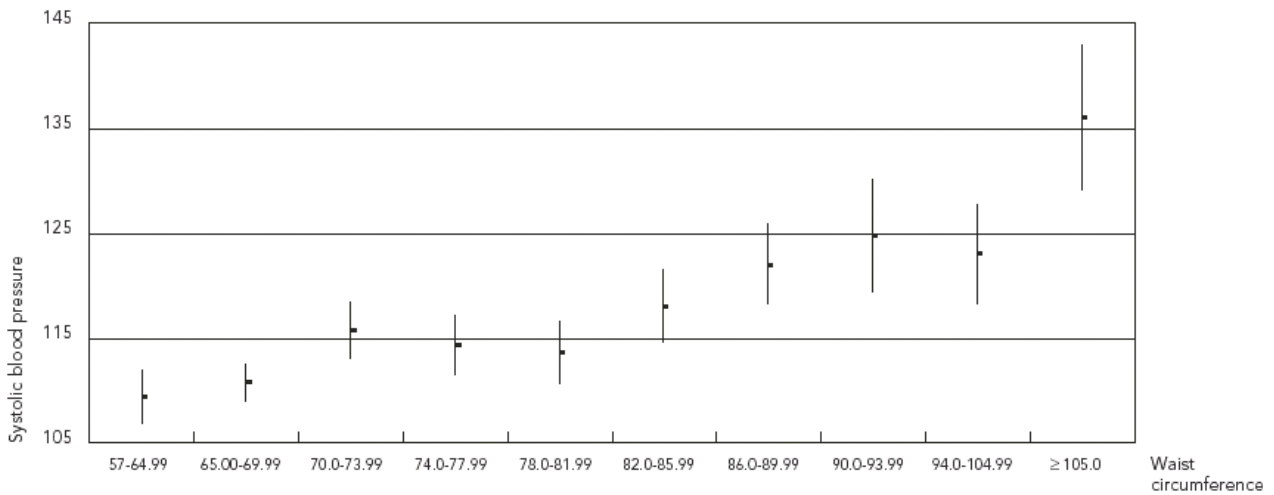
98	136	128	130	114	123	134	128	107	123
160	125	129	132	154	115	126	132	136	130

Valores de média e desvio padrão das observações:

Média ( $\bar{x}$ )	128
Desvio padrão ( $S_{n-1}$ )	13,91

### Questão 29

O gráfico abaixo foi extraído e adaptado do artigo: Avaliação da capacidade preditiva da circunferência da cintura para obesidade global e hipertensão arterial em mulheres residentes na Região Metropolitana de Belo Horizonte, Brasil de Velasquez-Melendez G *et al.* (*Cad. Saúde Pública*, 2002; 18 (3):765-71). Analise os resultados obtidos.



Mean values and 95% confidence intervals for systolic blood pressure in different waist circumference categories. Belo Horizonte, 2000.

## 5. Tomada de decisão

**Teste de hipóteses:** é um processo de tomada de decisão, com base probabilística.

Duas abordagens: de Neyman e Pearson (clássica) e de Fisher (valor de p)

Abordagem de Neyman e Pearson (Jerzy Neyman e Egon Pearson)

A decisão é tomada sobre hipóteses referentes a parâmetros da população de estudo, com probabilidades de erro de decisão fixadas a priori (antes da realização do estudo).

No caso mais simples, tem-se duas hipóteses:

$H_0$	Hipótese nula
$H_A$ ou $H_1$	Hipótese alternativa

Rejeita-se ou não rejeita-se  $H_0$  (aceita-se  $H_0$ ) com base nos resultados da amostra. Está, portanto, sujeita a erro, devido à variabilidade inerente ao processo de amostragem.

A possibilidade de cometer erro na decisão entre as hipóteses é medida probabilisticamente, usando-se um modelo adequado (binomial, normal, t de *Student*, qui-quadrado).

Erros possíveis na tomada de decisão e probabilidade a eles associada

DECISÃO (com base na amostra)	VERDADE estado da natureza (população)	
	$H_0$	$H_A$
$H_0$	Correta	errada (erro tipo II)
$H_A$	Errada (erro tipo I)	correta

$\alpha$  = Probabilidade (erro tipo I)  
= Probabilidade de rejeitar  $H_0$  (aceitar  $H_A$ ) quando a verdade é  $H_0$

$\alpha$  = Nível de significância do teste

$\beta$  = Probabilidade (erro tipo II)  
= Probabilidade de aceitar  $H_0$  quando a verdade é  $H_A$

$1-\beta$  = Probabilidade de rejeitar  $H_0$  quando  $H_0$  é falsa (aceitar  $H_A$  quando  $H_A$  é a verdade)

$1-\beta$  = Poder do teste

Conduta: Antes do experimento, fixa-se  $\alpha$  e trabalha-se com o menor  $\beta$  possível.

### Exemplo:

$\pi$  = Proporção de pacientes com efeitos adversos, no uso de uma droga a ser lançada.

$\pi_{\text{usual}}$  = Proporção de pacientes com efeitos adversos a uma droga usual.

Espera-se que a droga nova tenha menos efeitos adversos do que a droga usual.  $\pi < \pi_{\text{usual}}$ , ou seja,  $\pi < 0,6$ .

O experimento para verificar isto leva a um teste de hipóteses, da forma:

$H_0: \pi = 0,6$

$H_A: \pi < 0,6$

Podendo ser conduzido por exemplo, com 20 pacientes.

Modelo estatístico - probabilístico: Distribuição binomial de probabilidades, pois, para cada paciente, o resultado é da forma, tem efeito adverso/não tem efeito adverso e o resultado de um paciente não influencia os resultados nos outros.

Variável aleatória: Nº de pacientes com efeitos adversos = K.

Não se sabe qual das duas hipóteses é a verdadeira, mas pode-se construir a distribuição de probabilidades de K para o caso de  $H_0$  verdadeira.

Bin( $n=20; \pi=0,6$ )

K	P(K=k  $\pi=0,6$ )
0	0,000
1	0,000
2	0,000
3	0,000
4	0,000
5	0,001
6	0,005
7	0,015
8	0,035
9	0,071
10	0,117
11	0,160
12	0,180
13	0,166
14	0,124
15	0,075
16	0,035
17	0,012
18	0,003
19	0,000
20	0,000

Para se decidir entre  $H_0$  e  $H_A$ , há que se definir um conjunto de valores que levem a esta escolha. Isto é feito de forma probabilística, definindo-se um nível de significância  $\alpha$  (probabilidade de rejeitar  $H_0$  mas  $H_0$  é verdadeira)

Tomada de decisão:

Se, no experimento, o valor observado de K estiver entre aqueles da região de rejeição de  $H_0$ , a hipótese  $H_0$  será rejeitada.

$H_0: \pi=0,6$

$H_A: \pi < 0,6,$

Para

$\alpha = 5\%$

K	P(K=k  $\pi=0,6$ )	
0	0,000	<b>Área de rejeição de <math>H_0</math> resultados mais compatíveis com <math>H_A</math></b>  $\alpha$
1	0,000	
2	0,000	
3	0,000	
4	0,000	
5	0,001	
6	0,005	
7	0,015	
8	0,035	
9	0,071	Resultados mais compatíveis com $H_0$
10	0,117	
11	0,160	
12	0,180	
13	0,166	
14	0,124	
15	0,075	
16	0,035	
17	0,012	
18	0,003	
19	0,000	
20	0,000	

Realiza-se o estudo e verifica-se que **3** pacientes apresentaram efeitos adversos. QUAL A DECISÃO?

K	P(K=k  $\pi=0,6$ )
0	0,000
1	0,000
2	0,000
<b>3</b>	0,000
4	0,000
5	0,001
6	0,005
7	0,015
8	0,035
9	0,071
10	0,117
11	0,160
12	0,180
13	0,166
14	0,124
15	0,075
16	0,035
17	0,012
18	0,003
19	0,000
20	0,000

$\alpha = 5\%$

Qual a decisão se, entre os 20 pacientes, 14 tiverem efeitos adversos?

K	P(K=k  $\pi=0,6$ )
0	0,000
1	0,000
2	0,000
3	0,000
4	0,000
5	0,001
6	0,005
7	0,015
8	0,035
9	0,071
10	0,117
11	0,160
12	0,180
13	0,166
<b>14</b>	0,124
15	0,075
16	0,035
17	0,012
18	0,003
19	0,000
20	0,000

$\alpha = 5\%$

Supor agora que será lançada uma droga que pode reduzir o peso de pessoas em uma proporção maior do que a droga usual

Probabilidade da droga usual =  $\pi_{usual} = 0,6$

Hipóteses:  $\begin{cases} H_0 : \pi = 0,6 \\ H_A : \pi > 0,6 \end{cases}$  ; Teste monocaudal a direita

Supor agora que será lançada uma droga que pode, como efeito colateral, tanto reduzir como aumentar o peso de pessoas em uma proporção maior do que a droga usual

Hipóteses:  $\begin{cases} H_0 : \pi = 0,6 \\ H_A : \pi \neq 0,6 \end{cases}$  ; Teste bicaudal

Onde está  $\beta$  ?



Supor que se deseja comparar a eficácia de uma nova droga ( $E_N$ ) com eficácia de uma droga antiga ( $E_A$ ). A nova droga foi desenvolvida para ter maior eficácia.

Hipóteses:  $\begin{cases} H_0 : E_N = E_A \\ H_A : E_N > E_A \end{cases}$ ; Teste monocaudal a direita

Supor que a droga antiga tenha uma eficácia de 50% e que o estudo é realizado com uma amostra de 10 pacientes.

X: número de curas

X: 0, 1, 2, 3, ..., 10

$X \sim B(n=10; p=0,5)$

Hipóteses:  $\begin{cases} H_0 : E_N = 0,5 \\ H_A : E_N > 0,5 \end{cases}$

Valor esperado de curas =  $np = 10 \times 0,5 = 5$  curas

Distribuição de probabilidade (sob  $H_0$ ):  $B(n=10; p=0,5)$  e  $\alpha \cong 0,05$

X (número de curas)	P(X=x)	Região
0	0,001	
1	0,010	
2	0,044	aceitação de $H_0$
3	0,117	
4	0,205	
5	0,246	
6	0,205	
7	0,117	
8	0,044	rejeição de $H_0$
9	0,010	
10	0,001	

Supor que não é necessário uma diferença entre as eficácias  $E_N$  e  $E_A$  menor do que 10%, então supor  $\beta = 0,6$

X	p=0,5 P(X=x)	Região	p=0,6	Região
0	0,001		0,000	
1	0,010	$(1 - \alpha)$	0,002	$\beta$
2	0,044	aceitação de $H_0$	0,011	0,833
3	0,117		0,042	
4	0,205		0,111	
5	0,246		0,201	
6	0,205		0,251	
7	0,117		0,215	
8	0,044	rejeição de $H_0$	0,121	$(1 - \beta)$
9	0,010	$H_0$	0,040	0,167
10	0,001	$\alpha = 0,055$	0,006	

Valores de  $\beta$  e de  $1 - \beta$  para o teste de  $H_0: E_N = E_P = 50\%$  contra  $H_1: E_N > 50\%$ , quando  $n=10$ ;  $\alpha \cong 5\%$  (a rigor 5,47%) e  $\alpha \cong 1\%$  (a rigor 1,075%) segundo diferentes valores de  $E_N$ .

$E_N$	$\alpha \cong 5\%$		$\alpha \cong 1\%$	
	$\beta$ em %	$1 - \beta$ em %	$\beta$ em %	$1 - \beta$ em %
60%	83,27	16,73	95,36	4,64
70%	61,72	38,28	85,07	14,93
80%	32,22	67,78	62,42	37,58
90%	7,02	92,98	26,39	73,61

Valores de  $\beta$  e de  $1 - \beta$  para o teste de  $H_0: E_N = E_P = 50\%$  contra  $H_1: E_N = 60\%$ , quando  $\alpha \cong 5\%$  para diferentes valores de  $n$ .

Tamanho da amostra (n)	Valor de $\alpha$ mais próximo de 5%	Valor de $\beta$ (%)	Valor de $1 - \beta$ (%)
10	5,5	83,3	16,7
15	5,9	78,3	21,7
20	5,7	75,0	25,0
25	5,4	72,6	27,4
30	4,9	70,9	29,1
35	4,5	69,4	30,6
40	4,0	68,3	31,7

#### Abordagem de Fisher (Ronald Aylmer Fisher) (cálculo do valor de $p$ – valor descritivo do teste)

Fisher propôs um método de análise que também envolvia a tomada de decisão e consistia em verificar a probabilidade de ocorrência de um valor amostral observado ou de um valor mais extremo de uma estatística, sendo a probabilidade calculada com base em uma curva de probabilidade especificada.

Assim se na amostra de 10 pacientes fosse encontrada uma proporção de 65% de curas, Fisher recomendava que se calculasse a probabilidade desta proporção ou um valor mais extremo ocorrer em uma curva com proporção de cura conhecida (como por exemplo, 50%).

Se esta probabilidade for pequena conclui-se que possivelmente a droga cura mais do que 50% porque os dados não mostram evidência de pertencer a uma população com proporção de cura igual a 50%.

Se a probabilidade do valor observado ou de um valor mais extremo pertencer à curva especificada, for grande, então diz-se que os dados provavelmente vêm de uma população que possui como parâmetro 50% de curas.

Esta probabilidade é conhecida como valor de  $p$ ; ou nível descritivo do teste e é fornecido pelos pacotes estatísticos de análise de dados.

Valor de  $p$  = probabilidade de ocorrência do valor observado ou de um valor mais extremo, em uma curva de probabilidade especificada (conhecida, verdadeira).

Fisher dizia que antes de dar uma forma matemática a um problema, propondo hipóteses a serem testadas, era necessário um amplo conhecimento dos dados, o que poderia ser realizado com base no valor de  $p$ .

**Exemplo:** Em uma determinada comunidade, estudos têm apontado que a proporção de mães que amamentam até o 3<sup>o</sup> mês é de 60%. Deseja-se aumentar esta proporção e para tanto, desenvolve-se um programa educativo com a comunidade. Deseja-se saber se, após o término do programa, pode-se considerar que a proporção de mães que amamentam até o 3<sup>o</sup> mês foi maior do que 60%.

Proposição: proporção de mães que amamentam vem de uma população = 60%

A idéia é verificar se após a campanha, a proporção observada em uma amostra é compatível com 60% (proporção especificada na proposição). Se não for compatível, pode-se argumentar que a população que representa a proposição não é compatível com os resultados.

Para tanto, realiza-se o estudo que consiste em sortear, após uma campanha de incentivo ao aleitamento materno, um número de mães que deram à luz (nos últimos 12 meses), contar quantas amamentaram até o 3º mês e calcular a proporção. Supor que foram selecionadas 10 mães e destas, 9 amamentaram até o 3º mês.

Como vamos decidir se o programa teve efeito ou se este resultado foi somente um artifício da amostra selecionada? Ou seja, devido ao acaso?

Para decidir calcula-se a probabilidade de pelo menos 9 mães amamentarem se elas tivessem vindo de uma população onde 60% das mães amamentam.

Cálculo da probabilidade de ocorrer 9 ou mais mães em uma população com 60% de amamentação.

Definição da distribuição de probabilidade:

- Amamentar é uma variável dicotômica (amamenta ou não)
- A probabilidade de amamentar é conhecida  $\pi = 0,6$  ;
- Amamentar exclui não amamentar (eventos mutuamente exclusivos);
- O comportamento de uma mãe não interfere no comportamento de outra (eventos independentes);
- O tamanho da amostra é conhecido,  $n=10$ ;
- É possível definir a variável X: número de mães que amamentam, que assume os valores  $x: 0, 1, 2, \dots, 10$ .

Distribuição de probabilidade Binomial com parâmetros  $n=10$  e  $\pi = 0,6$ .

Cálculo da probabilidade:

Observou-se que 9 mães, do total de 10, amamentaram seus filhos até o 3º mês. Se com 9 mães amamentando concluímos que o programa foi eficiente, com 10 mães também seria. Portanto, calcula-se a probabilidade de 9 ou mais mães amamentarem, se elas tivessem vindo de uma população onde 60% amamenta.

Pela distribuição Binomial, ( $n=10$ ;  $\pi = 0,6$ ), obtém-se

$$P(X \geq 9) = P(X = 9) + P(X = 10) = 0,040 + 0,006 = 0,046 \text{ ou } 4,6\%$$

Valor de  $p=4,6\%$

A probabilidade de observar 9 ou mais mães amamentando, se elas tivessem vindo de uma população de mães que 60% amamentam é 4,6%.

Os resultados observados são compatíveis com a proposição de que as mães vêm de população que amamenta 60%? Em outras palavras, com base nos resultados, você diria que existe evidência de que estas mães vêm de uma população onde 60% amamenta o período recomendado?

Se  $p$  for considerado, pelo investigador, pequeno diz-se que os dados observados mostram evidência favorável à campanha. A proporção de mães que amamentaram, depois da campanha é mais compatível com a população de mães em uma população onde uma proporção maior do que 60% amamenta.

Se  $p$  for considerado grande, pelo investigador, diz-se que os dados não mostram evidência desfavorável à proposição e portanto, a campanha não surtiu efeito. A diferença observada na amostra não ocorreria na população.

O julgamento sobre o valor de  $p$ , se é grande ou pequeno, é arbitrário e quem decide é o investigador.

## 5.1- Teste de hipótese de uma proporção populacional

### Abordagem de Neyman e Pearson

Supor a situação onde um fabricante quer lançar um bolo mais doce do que está no mercado (código 15). Para tanto é desenvolvido um novo produto (código 23). O estudo consiste em oferecer uma fatia de cada produto a 20 degustadores e contar quantos acham o produto novo mais doce.

O investigador fixa o nível de significância em 5% ( $\alpha = 0,05$ ).

No teste "comparação de par, direcionada", o número K de degustadores que achou o bolo com código 23 mais doce foi  $k=8$ . Pode-se admitir que os bolos diferem quanto à doçura?

Se os bolos não forem diferentes quanto ao sabor, metade dos degustadores vai escolher um produto e a outra metade vai escolher o outro.

Probabilidade de escolher o código 15 = 0,5

$$\begin{cases} H_0 : \pi = 0,5 \\ H_A : \pi > 0,5 \end{cases}$$

### Distribuição de probabilidade:

X: número de degustadores que escolhem o produto novo.

X: 0, 1, 2, ..., 18, 19, 20

Cada degustador escolhe o produto independente do outro degustador e se for feita a escolha por um produto (eventos independentes), o outro fica eliminado (mutuamente exclusivos).  $X \sim B(n=20; p=0,50)$

X	P(X=x)	Região do teste
0		
1		
2		
3		
.		
.		
.		
13		
14	0,03696	
15	0,01479	Rejeição de $H_0$
16	0,00462	$\alpha = 0,058$
17	0,00109	
18	0,00018	
19	0,00002	
20	0,00000	

Como  $k=8$ , toma-se a decisão de não rejeitar  $H_0$ . Pode-se dizer que os produtos são semelhantes quanto ao sabor adocicado, para um nível de significância de 5%.

### Abordagem de Fisher

Proposição: Proporção de degustadores que escolhem código 15=0,5

$n=20$

K: número de degustadores que escolhem código 15

K: 0, 1, 2, ..., 20

$K \sim B(n=20; p=0,5)$

Observou-se que 8 degustadores escolheram código 23 ( $k=8$ ).

Valor de  $p$ =probabilidade de ocorrer 8 ou um valor mais extremo, se os dados vierem de uma população que escolhe os produtos em proporções iguais.

Valor de  $p = P(K \geq 8) = P(k=8) + P(k=9) + \dots + P(k=20) = 0,868$  ou 86,8%.

Conclui-se que os dados mostram evidência favorável à proposição sendo possivelmente provenientes de uma população que escolhe igualmente os códigos 15 e 23. O novo produto não foi considerado mais doce que o antigo.

OBS: Como  $np=10$  e  $nq=10$ , para o cálculo de  $p$  poderia ter sido utilizada a distribuição normal.

$$K \sim N(\mu = np = 20 \times 0,5 = 10; \sigma = \sqrt{npq} = \sqrt{20 \times 0,5 \times 0,5} = 2,24)$$

$$P(K \geq 8) = P\left(Z \geq \frac{7,5 - 10}{2,24}\right) = P(Z \geq -1,12) = 0,5 + 0,36864 = 0,86764 \text{ ou } 86,8\%$$

### Questão 30

Em um julgamento jurídico o júri tem que decidir sobre a culpa ou inocência de um réu. Considere dois fatos: 1) o sistema jurídico admite que toda pessoa é inocente até que se prove o contrário. 2) só vai a julgamento pessoas sobre as quais existe dúvida de sua inocência. Fazendo uma analogia com teste de hipóteses, responda

- Apresente as hipóteses nula e alternativa sobre a culpa ou inocência do réu;
- O júri pode errar se decidir que o réu é culpado quando na verdade ele é inocente; qual é o outro erro de decisão que o júri pode cometer?
- Qual dos dois erros é o mais sério?
- Na terminologia de teste de hipótese, qual tipo de erro (I ou II) pode-se vincular a cada uma das decisões do item b)?

### Questão 31

Quando um diagnóstico médico é fornecido, qual dos erros é geralmente mais sério: um resultado falso positivo que diz que a pessoa tem a doença quando na verdade ela não tem ou um resultado falso negativo, que diz que a pessoa não tem a doença quando na verdade ela tem? Imagine duas situações: 1) a pessoa está fazendo parte de um *screening* para câncer de mama. 2) a pessoa realiza o teste para detectar anticorpos anti-HIV.

### Questão 32

Com base na questão anterior

- Apresente as hipóteses nula e alternativa sobre a situação de saúde do paciente;
- Que tipo de erro (I ou II) seria cometido se o resultado do teste fosse falso positivo?
- Que tipo de erro (I ou II) seria cometido se o resultado do teste fosse falso negativo?

### Questão 33

Um estudo foi desenvolvido para investigar se aleitamento materno é um fator que protege a criança contra diabetes, em idades maiores. Considerando  $H_0$ : aleitamento materno não protege contra diabetes e  $H_a$ : aleitamento materno protege contra diabetes, responda

- Tomando qual decisão (aceitar ou rejeitar) sobre  $H_0$  você poderia estar cometendo o erro tipo I?
- Tomando qual decisão (aceitar ou rejeitar) sobre  $H_0$  você poderia estar cometendo o erro tipo II?
- Como é denominada a probabilidade de ocorrer o erro tipo I?
- Como é denominada a probabilidade de ocorrer o erro tipo II?
- O que é o poder do teste?
- Se você fosse fixar valores de probabilidades associadas à ocorrência dos erros tipo I e II para este estudo, qual deles seria menor? Justifique.

### Questão 34

Supor o experimento onde existe interesse em investigar se o odor de determinado alimento atrai camundongos. O experimento consiste em colocar um animal em um corredor que no final é dividido para a direita e para a esquerda. O alimento é colocado no final do corredor da esquerda, fora da visão do animal. Entretanto, antes da realização do experimento, decide-se eliminar a possibilidade de incluir no estudo

camundongos que têm predileção por um lado, independentemente do odor do alimento. Neste caso, decide-se investigar inicialmente se os camundongos escolhem os lados em proporções iguais. Para tanto, realiza-se o experimento com 12 camundongos sem a colocação do alimento e verifica-se que 7 viram para a esquerda.

a) realize um teste de hipóteses segundo o modelo clássico de Newman e Pearson, com nível de significância  $\alpha=5\%$ , para verificar se os camundongos vêm de uma população que escolhe mais um lado do que o outro;

b) realize um teste de hipóteses seguindo a proposta de Fisher, com cálculo do valor descritivo do teste, para verificar se os camundongos vêm de uma população que escolhe mais um lado do que o outro.

### **Questão 35**

Supor, agora, o experimento para investigar se o odor de determinado alimento atrai camundongos. Realiza-se o experimento colocando o alimento no final do corredor da esquerda, fora da visão dos camundongos. Observa-se que de 12 camundongos, 10 viram para a esquerda.

a) realize um teste de hipóteses segundo o modelo clássico de Neyman e Pearson, com nível de significância  $\alpha=5\%$ , para verificar se os camundongos vêm de uma população que escolhe mais o lado onde está o alimento;

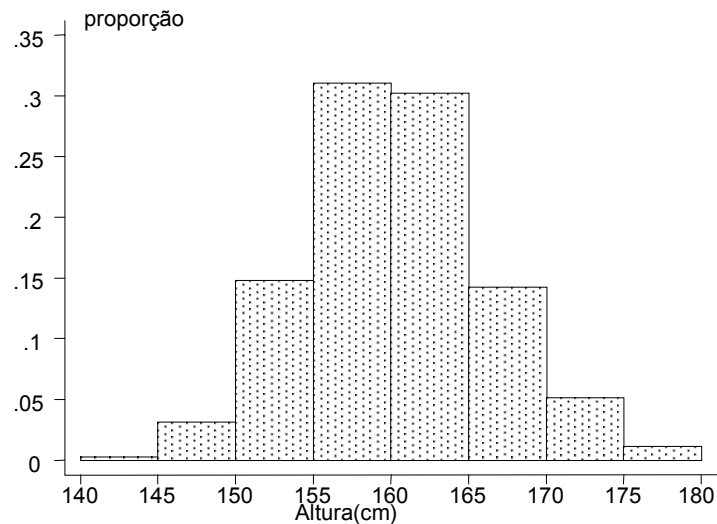
b) realize um teste de hipóteses seguindo a proposta de Fisher, com cálculo do valor descritivo do teste, para verificar se os camundongos vêm de uma população que escolhe mais o lado onde está o alimento.

## 5.2 – Teste de hipóteses de uma média populacional ( $\mu$ ) (com variância conhecida e desconhecida)

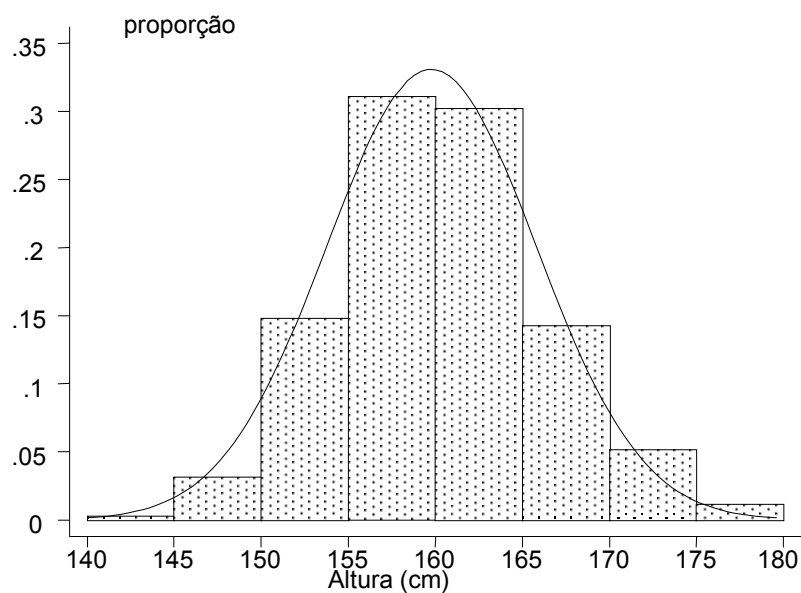
### 5.2.1 - Revisão de conceitos básicos

#### Distribuição Normal

Medindo-se a altura de muitas mulheres (população), obtém-se o gráfico a seguir.



Aos dados pode ser ajustada uma curva teórica



A curva ajustada aos dados é uma curva teórica (curva de Gauss) que representa a altura de mulheres idosas.

$$X: \text{altura}, X \sim N(m = 160, \sigma = 10)$$

A curva tem propriedades conhecidas:

- Soma da área sob a curva é igual a 1 ou 100%;
- Pode-se calcular probabilidade trabalhando-se com a área sob a curva;
- Sorteia-se uma mulher ao acaso;  $P(\text{desta mulher ter mais de } 160) = 50\%$ ;

$P(\text{esta mulher tenha mais de } 180) = P(X > 180) = P(Z > \frac{180-160}{10}) = P(Z > 2) = 0,5 - 0,47725 = 0,02275$  ou 2,3%

Sortear uma mulher com altura 180 ou mais desta população é uma coisa comum?

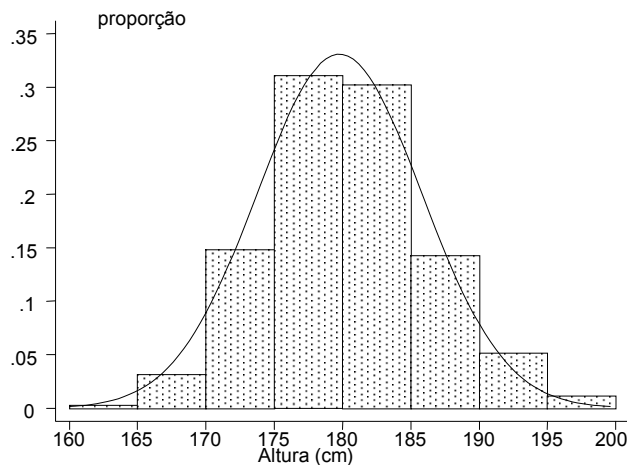
$P(\text{esta mulher tenha mais de } 170) =$

$P(X > 170) = P(Z > \frac{170-160}{10}) = P(Z > 1) = 0,5 - 0,34134 = 0,15866$  ou 15,9%

Sortear uma mulher com altura 170 ou mais desta população é uma coisa comum?

O limite para decidir o que é comum é arbitrário. O investigador é que decide, com base no problema que está sendo estudado.

Supor agora uma outra distribuição de altura, por exemplo, entre homens.



X: altura,  $X \sim N(m = 180, \sigma = 10)$

$P(\text{homem tenha mais de } 180) = P(X > 180) = 0,5$  ou 50%

Supor que as pessoas da primeira curva tenham uma marca vermelha e as pessoas da segunda curva tenham uma marca azul. Misturam-se todas as pessoas e sorteia-se uma pessoa ao acaso e ela tem altura 180.

De qual população seria este indivíduo?

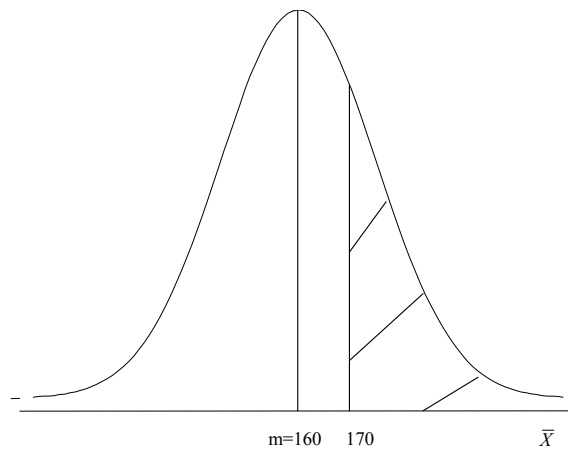
Agora não estamos mais interessados em um indivíduo e sim em vários indivíduos portanto, vamos trabalhar com a altura média.

Supor X: altura,  $X \sim N(m = 160, \sigma = 20)$ .

Toma-se uma amostra de tamanho  $n=30$  desta população e calcula-se a altura média ( $\bar{x} = 170$ ).

Deseja-se saber a probabilidade desta amostra vir de população com altura média maior ou igual a 170.

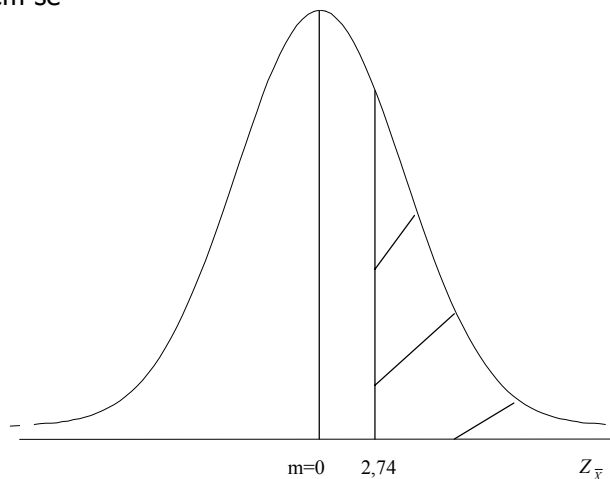




Pelo Teorema do Limite Central

$$P(\bar{X} > 170) = P\left(\frac{\bar{X} - m}{\sigma_{\bar{X}}} > \frac{170 - 160}{\frac{20}{\sqrt{30}}}\right) = P\left(Z_{\bar{X}} > \frac{170 - 160}{3,6515}\right) = P(Z > 2,74)$$

Na curva da Normal reduzida tem-se



$$= 0,5 - 0,49693 = 0,0031 \text{ ou } 0,31\%$$

Sortear uma amostra que apresenta altura média igual a 170 cm ou mais da população que tem altura média 160 cm é uma coisa comum?

Para decidir se a amostra representa uma população com altura média maior e que o resultado não é devido ao acaso, realiza-se o teste de hipóteses.

## 5.2.2 - Teste de hipóteses para uma média populacional com variância conhecida

Abordagem de Fisher:

Situação:

Estudos mostram que crianças sadias possuem peso médio ( $m$ ) ao nascer igual a 3100 gramas e desvio padrão  $\sigma = 610 \text{ gramas}$ .

Suspeita-se que crianças que nascem com síndrome de desconforto idiopático grave possuem peso ao nascer abaixo do peso ao nascer da população de crianças sadias.

Proposição: Crianças com síndrome vêm de uma população com peso médio = 3100 gramas

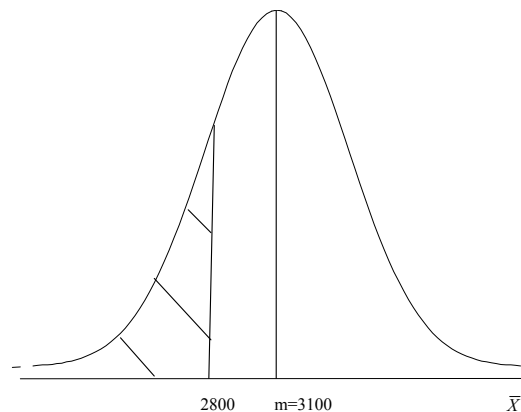
Realiza-se um estudo em uma amostra de  $n=50$  crianças que nasceram com esta síndrome, onde observou-se peso médio ( $\bar{x}$ ) igual a 2800 gramas.

Supondo-se que as crianças da amostra (com síndrome) vêm de uma população com mesma dispersão do peso ao nascer de crianças saudáveis, teste a hipótese de que crianças com síndrome idiopática grave possuem peso médio ao nascer igual ao peso médio ao nascer de crianças saudáveis.

Distribuição de probabilidade:

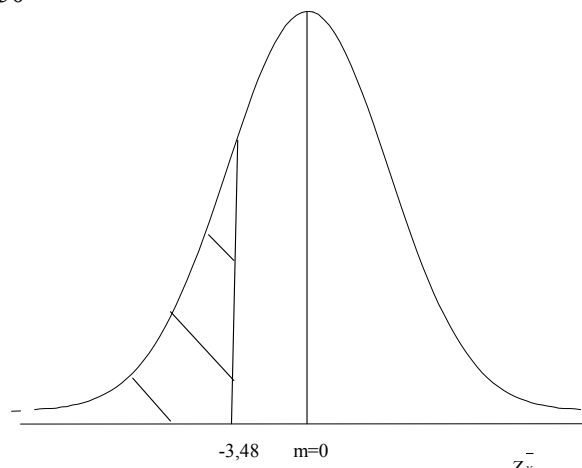
Distribuição do peso médio: segue uma distribuição normal com média  $m=3100$  gramas e desvio padrão

$$\frac{\sigma}{\sqrt{n}} = \frac{610}{\sqrt{50}} = 86,27 \text{ gramas}$$



Cálculo da probabilidade de observar um peso médio ao nascer igual ou menor que 2800 se  $H_0$  for verdadeira.

$$P(\bar{X} < 2800) = P\left(\frac{\bar{X} - m}{\sigma_{\bar{X}}} < \frac{2800 - 3100}{\frac{610}{\sqrt{50}}}\right) = P(Z_{\bar{X}} < \frac{-300}{86,27}) = P(Z_{\bar{X}} < -3,48)$$



Pela distribuição Normal reduzida tem-se que

$$P(Z < -3,48) = 0,5 - 0,49975 = 0,00025 \text{ ou } 0,025\%$$

Os resultados não são compatíveis com uma distribuição que tem peso médio igual a 3100. Possivelmente a amostra vem de uma população com média menor que 3100. Pode-se dizer que crianças com síndrome de desconforto idiopático grave possivelmente possuem peso ao nascer menor do que o peso médio de crianças saudáveis.

### 5.2.3 - Teste de hipóteses para uma média populacional com variância desconhecida

Abordagem de Fisher:

Supor a mesma situação anterior, só que neste caso somente a média populacional é conhecida. O peso médio de crianças saudias ( $m$ ) é igual a 3100 gramas.

Proposição: Crianças com síndrome de desconforto idiopático grave vêm de uma população com peso médio =3100 gramas

Seleciona-se uma amostra de 50 crianças com a síndrome e calcula-se o peso médio e o desvio padrão do peso, obtendo-se

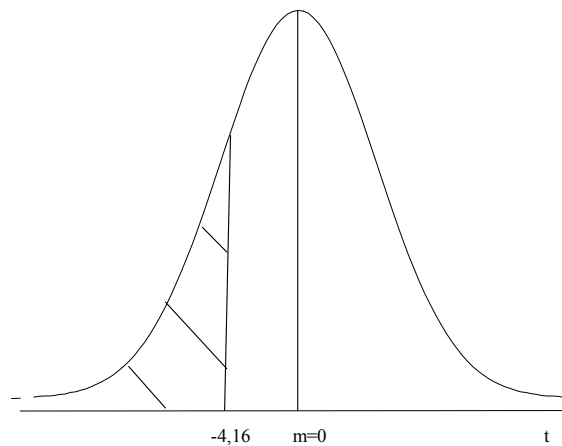
$n=50$ ;  $\bar{x} = 2800$  e  $s=510$

Distribuição de probabilidade:

Distribuição do peso médio ao nascer de crianças saudias: como não se sabe o desvio padrão populacional, este é estimado utilizando-se os dados da amostra.

Neste caso a variável segue uma distribuição *t de Student* com  $n-1=50-1=49$  graus de liberdade.

$$P(\bar{X} < 2800) = P\left(\frac{\bar{X} - m}{S_{\bar{X}}} < \frac{2800 - 3100}{\frac{510}{\sqrt{50}}}\right) = P\left(t_{\bar{X}} < \frac{-300}{72,12}\right) = P(t_{\bar{X}} < -4,159)$$



Pela distribuição *t de Student* com 49 graus de liberdade, tem-se  $P(t_{\bar{X}} < -4,159) < 0,05\%$

Os resultados não são compatíveis com uma distribuição que tem peso médio igual a 3100. Pode-se dizer que crianças com desconforto idiopático grave provavelmente vêm de uma população com peso médio ao nascer menor do que o peso médio ao nascer de crianças saudias.

#### Questão 36

Uma companhia de produtos alimentícios utiliza uma máquina para embalar salgadinhos cujas embalagens especificam 454gramas. Com o propósito de verificar se a máquina está trabalhando corretamente, selecionou-se 50 pacotes de salgadinhos, obtendo-se os seguintes valores de peso:

464	450	450	456	452	433	446	446	450	447
442	438	452	447	460	450	453	456	446	433
448	450	439	452	459	454	456	454	452	449
463	449	447	466	446	447	450	449	457	464
468	447	433	464	469	457	454	451	453	443

média da amostra,  $\bar{x} = 451,22$  gramas e desvio padrão amostral,  $s=8,40$  gramas

Testar a hipótese de que a máquina está trabalhando corretamente. Utilizar o valor de p para tomada de decisão

#### Questão 37

O nível médio de protombina em populações normais é 20 mg/100ml de sangue. Em uma amostra de 40 pacientes que tinham deficiência de vitamina K foram observadas as estatísticas: nível médio de protombina de 18,5mg/100ml e desvio padrão 4mg/100ml. Seria razoável concluir que a verdadeira média de pacientes com deficiência de vitamina K é a mesma que a da população normal? Realize um teste de hipóteses segundo a abordagem de Fisher para responder a pergunta.

### 5.3 - Teste de hipóteses de associação pelo Qui-quadrado de Pearson ( $\chi^2$ )

O Qui-quadrado é obtido somando-se a diferença ao quadrado entre as frequências observadas e as esperadas, dividido pelas frequências esperadas.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Se o Qui-quadrado for igual a zero, então não existe associação entre as variáveis. Quanto maior o valor do Qui-quadrado, maior a chance de existir associação entre as variáveis, entretanto, o Qui-quadrado não mede força de associação.

Distribuição Qui-quadrado ( $\chi^2_{(n-1)}$ ) com (n-1) graus de liberdade

Seja uma população com distribuição normal  $N(\mu, \sigma)$ . Se desta população se obtiver um número infinito de amostras de tamanho n, calculando-se as quantidades  $\bar{x}$  e  $S^2$  em cada amostra, a variável aleatória  $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ , onde  $\chi^2_{(n-1)}$  se lê "qui-quadrado com n-1 graus de liberdade" (Berquó, 1981).

A distribuição Qui quadrado é assimétrica e se torna menos assimétrica a medida que os graus de liberdade aumentam. Os valores da distribuição são sempre positivos (maior ou igual a zero). Existe uma família de distribuições qui quadrado, dependendo do número de graus de liberdade. Para grandes amostras, a distribuição qui quadrado tende para uma distribuição normal.

#### Teste de hipóteses de associação pelo Qui-quadrado de Pearson ( $\chi^2$ )

$H_0$  : Não existe associação

Estatística do teste:

$$\text{Qui-quadrado} = \sum \frac{(O - E)^2}{E} \sim \chi^2_{(r-1)(c-1)}$$

onde r e c representam o número de linhas e de colunas, respectivamente

Correção de continuidade:

$$\text{Qui-quadrado}_{\text{correcao de Yates}} = \sum \frac{(|O - E| - 0,5)^2}{E} \sim \chi^2_{(r-1)(c-1)}$$

**Limitações:**

Para  $n < 20$ , utilizar o teste exato de Fisher

Para  $20 \leq n \leq 40$ , utilizar o Qui-quadrado somente se os valores esperados forem maiores ou iguais a 5

**Exemplo:**

Com o objetivo de investigar a associação entre história de bronquite na infância e presença de tosse diurna ou noturna em idades mais velhas, foram estudados 1.319 adolescentes com 14 anos. Destes, 273 apresentaram história de bronquite até os 5 anos de idade sendo que 26 apresentaram tosse diurna ou noturna aos 14 anos.

Número de adolescentes segundo história de bronquite aos 5 anos e tosse diurna ou noturna aos 14 anos de idade. Local X, ano Y.

Tosse	Bronquite		Total
	Sim	Não	
Sim	26	44	70
Não	247	1002	1249
Total	273	1046	1319

Fonte: Holland, WW *et al*, 1978.

Valores observados (O)	Valores esperados (E)	(O-E)	(O-E) <sup>2</sup>	$\frac{(O - E)^2}{E}$
26	14,488	11,512	132,526	9,147
247	258,512	-11,512	132,526	0,513
44	55,512	-11,512	132,526	2,387
1002	990,488	11,512	132,526	0,134
			Qui-quadrado=	12,181

p não corrigido=0,0004829

**Questão 38**

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo condição de sobrevivência e peso ao nascer (g)

Peso ao nascer	Óbito	Sobrevida	Total
Baixo peso (<2500)	24	13	37
Não baixo peso (2500 e mais)	3	10	13
Total	27	23	50

Fonte: Hand DJ *et al*, 1994.

Com base nos dados apresentados na tabela realize o teste de hipótese para investigar a existência de associação entre o peso do recém-nascido e a condição de sobrevivência.

### Questão 39

Defeitos do tubo neural são malformações congênitas que surgem durante o desenvolvimento fetal. É conhecida como spina bífida. Estes dados são de um estudo conduzido no país de Gales – Reino Unido, para investigar possível associação entre defeitos do tubo neural e dieta materna. O estudo é do tipo caso-controle: mães que tinham tido bebês com defeitos do tubo neural (casos) e suas irmãs que não tinham tido (controles) foram avaliadas segundo suas dietas e classificadas em boa, razoável e ruim.

Distribuição de recém-nascidos segundo condição caso - com defeitos do tubo neural; controle – recém-nascidos que não tinham defeitos do tubo neural e dieta materna.

Dieta	Casos		Controles		Total	
	n	%	n	%	n	%
Boa	34	13,9	43	35,0	77	21,0
Razoável	110	45,1	48	39,0	158	43,0
Pobre	100	41,0	32	26,0	132	36,0
Total	244	100	123	100	367	100

Com base nos dados apresentados na tabela realize o teste de hipótese para investigar existência de associação entre as variáveis.

### Questão 40

A tabela apresenta dados de classificação de pessoas segundo doença coronariana (DIC) e hábito de consumo de café, para uma coorte de 1718 homens com idade 40-55 anos.

Distribuição de homens segundo consumo de café e incidência de doença coronariana (DIC).

condição	Consumo médio de café		total
	Alto ( $\geq 100$ xícaras/mês)	Moderado ( $< 100$ xícaras/mês)	
Com DIC	38	39	77
Sem DIC	752	889	1641
total	790	928	1718

Com base nos dados apresentados na tabela realize o teste de hipótese para associação entre consumo de café e doença coronariana.

## 6. Bibliografia

Berquó ES; Souza JMP. e Gotlieb SLD. *Bioestatística*. EPU, São Paulo, 1981.

Caldas ED, Silva SC, Oliveira JN. Aflatoxinas e ocratoxina A em alimentos e riscos para a saúde humana. *Rev. Saúde Pública*, 2002;36(3):319-23.

Carvalho AT, Costa MJC, Ferreira LOC, Batista Filho M *et al.* Cartografia do retardo estatural em escolares do estado da Paraíba. Brasil. *Rev Saúde Pública* 2000;34(1):3-8.

Chambers JM, Cleveland WS, Kleiner B, Tukey PA. *Graphical methods for data analysis*. Wadsworth & Brooks/Cole Publishing Company, Advanced Books & Software. Pacific Grove, California, 1983.

Corrêa Filho HR, Vieira JBF, Silva YS, Coelho GE, Cavalcante FAC, Pereira MP. Inquérito sobre a prevalência de bócio endêmico no Brasil em escolares de 6 a 14 anos: 1994 a 1996. *Rev Panam Salud Publica* 2002;12(5):317-325.

Daly F, Hand DJ, Jones MC, Lunn AD, McConway KJ. *Elements of Statistics*, 1999.

Dixon WJ e Massey FJ. *Introduction to Statistical Analysis*. 2nd edit. The Maple Press Company, York, 1957.

Escosteguy CC, Medronho RA, Madruga R *et al.* Vigilância epidemiológica e avaliação da assistência às meningites. *Rev. Saúde Pública*, 2004;38(5):657-63.

Fornes IS, Martins IS, Velasques-Melendez G *et al.* Escores de consumo alimentar e níveis lipêmicos em população de São Paulo, Brasil. *Rev. Saúde Pública*, 2002;36(1):12-8.

Hand DJ *et al.* *A handbook of small data sets*. Chapman&Hall, 1994.

Holland WW, Bailey R, Bland JM. Long-term consequences of respiratory disease in infancy. *Journal of Epidemiology and Community Health* 1978;32: 256-9.

Kummer SC, Giugliani ERJ, Susien LO *et al.* Evolução do padrão de aleitamento materno. *Rev de Saúde Pública*, 2000;34(2):143-8.

Lerario DG, Gimeno SG, Franco LJ *et al.* Excesso de peso e gordura abdominal para a síndrome metabólica em nipo-brasileiros. *Rev. Saúde Pública*, 2002;36(1):4-11.

Monteiro CA, Benicio MHDA, Ortiz LP. Tendência secular do peso ao nascer na cidade de São Paulo (1976-1998) *Rev. Saúde Pública*; 2000;34(6, supl):26-40.

Monteiro CA, Szarfarc SC, Mondini L. Tendência secular da anemia na cidade de São Paulo (1984-1996). *Rev. Saúde Pública*, 2000; 34(6, sup):62-72.

Monteiro POA, Victora CG, Barros F *et al.* Diagnóstico de sobrepeso em adolescentes: estudo do desempenho de diferentes critérios para o Índice de Massa Corporal. *Rev. Saúde Pública*, 2000;34(5):506-13.

Osborn JF. *Statistical Exercises in Medical Research*. John Wiley & Sons Inc., 1979.

Otero UB, Rozenfeld S, Gadelha AMJ *et al.* Mortalidade por desnutrição em idosos, região Sudeste do Brasil, 1980-1997. *Rev. Saúde Pública*, 2002;36(2):141-8.

Prado MS, Assis AMO, Martins MC, Nazaré MPA, Rezende IFB, Conceição MEP. Hipovitaminose A em crianças de áreas rurais do sertão semi-árido baiano. *Revista de Saúde Publica*, 1995;29(4):295-300.

Soares JF, Siqueira AL. *Introdução à Estatística Médica*, COOPMED, 2002.

Souza Júnior PRB, Szwarcwald CL, Barbosa Junior A *et al.* Infecção pelo HIV durante a gestação: Estudo-Sentinela Parturiente, Brasil, 2002. *Rev. Saúde Pública*, 2004;38(6):764-72.

Triola MF. *Introdução à Estatística*. LTC Livros Técnicos e Científicos Editora, 1999.

Tuyns AJ, Pequignot G, Jensen OM. Esophageal cancer in Ille-et-Vilaine in relation to levels of alcohol and tobacco consumption. Risks are multiplying. *Bull Cancer* 1977;64:45-60.

van Vliet PKJ e Gupta JM. THAM v. sodium bicarbonate in idiopathic respiratory distress syndrome. *Archives in Disease in Childhood*, 1973;48:249-55.

Velasquez-Melendez G, Kac G, Valente JG *et al.* Avaliação da capacidade preditiva da circunferência da cintura para obesidade global e hipertensão arterial em mulheres residentes na Região Metropolitana de Belo Horizonte, Brasil. *Cad. Saúde Pública*, 2002;18(3): 765-771.



## 7. Resumos dos artigos citados

CALDAS, Eloisa Dutra, SILVA, Saulo Cardoso e OLIVEIRA, João Nascimento. **Aflatoxinas e ocratoxina A em alimentos e riscos para a saúde humana.** *Rev. Saúde Pública*, jun. 2002, vol.36, no.3, p.319-323.

**OBJETIVOS:** A presença de micotoxinas em alimentos tem sido correlacionada a várias patologias humanas, e as autoridades de saúde no mundo todo têm implementado ações para diminuir a ingestão desses compostos pela dieta. Realizou-se pesquisa para analisar os níveis de aflatoxinas e ocratoxina A de alimentos para consumo e avaliar o potencial de risco da exposição humana a essas micotoxinas.

**MÉTODOS:** Foram analisadas 366 amostras de alimentos consumidos no Distrito Federal, no período de julho de 1998 a dezembro de 2001, como amendoim e derivados, castanhas, milho, produtos de trigo e/ou aveia, arroz e feijão. As amostras foram processadas, e as micotoxinas extraídas, detectadas e quantificadas por fluorescência após separação em cromatografia camada delgada. **RESULTADOS:** Foram detectadas aflatoxinas em 19,6% das amostras, em amendoim cru e derivados, milho de pipoca, milho em grão e castanha-do-pará (>2 µg/kg). Amendoim e derivados apresentaram maior incidência de contaminação por aflatoxinas (34,7%) com amostras contendo até 1.280 µg/kg de AFB1+AFG1 e 1.706 µg/kg de aflatoxinas totais. Das amostras positivas, AFB1 estava presente em 98,5%, AFB2 em 93%, AFG1 em 66,7% e AFG2 em 65,4%. A ocratoxina A não foi detectada (<25 µg/kg) em nenhuma amostra analisada. **CONCLUSÃO:** Os níveis de contaminação encontrados em amendoim e derivados ultrapassaram os níveis máximos permitidos pela legislação brasileira, podendo significar fator de risco para a população que os consome regularmente. A conscientização dos produtores de alimentos e as ações de vigilância sanitária permanentes são essenciais para diminuir a exposição humana a esses compostos e prevenir doenças crônicas advindas dessa exposição.

**Palavras-chave:** Aflatoxinas; Ocratoxinas; Análise de alimentos; Contaminação de alimentos; Vigilância sanitária; Amendoim; Milho.

CARVALHO, Alice T, COSTA, Maria JC, FERREIRA, Luiz OC *et al.* **Cartografia do retardo estatural em escolares do Estado da Paraíba, Brasil.** *Rev. Saúde Pública*, fev. 2000, vol.34, no.1, p.3-8. ISSN 0034-8910.

**OBJETIVO:** Parte de um projeto promovido pelas Nações Unidas (FAO) e pelo Ministério da Educação do Brasil (Fundação de Assistência ao Escolar) para estudar a condição nutricional de alunos de educandários públicos, visa a avaliar, com um enfoque cartográfico, a magnitude e a distribuição espacial do déficit estatural de alunos de escolas públicas do Estado da Paraíba. **MÉTODOS:** Fez-se levantamento de caráter censitário da estatura de 50.144 escolares com idade de 6 a 9 anos, matriculados na primeira série do primeiro grau de todos municípios do Estado da Paraíba, considerando-se como déficit estatural os valores classificados abaixo de -2 "scores" z do padrão do National Center of Health Statistics. **RESULTADOS:** Em todo o Estado, 14,5% dos escolares foram considerados casos de nanismo, com uma distribuição heterogênea para as quatro mesorregiões fisiográficas estudadas: 18,7% no Sertão, 13,8% no Agreste, 11,9% na Borborema e 10,9% na Zona da Mata. Nas escolas da área rural, 17,8% dos alunos apresentavam déficit estatural, enquanto no meio urbano a prevalência de nanismo foi de 11,8%.

**CONCLUSÕES:** Os dados encontrados são indicativos de um processo de interiorização da desnutrição, contrastando substancialmente com as descrições históricas sobre a cartografia do problema na região.

**Palavras-chave:** Estatura; Inquéritos nutricionais; Distribuição espacial; Transtornos nutricionais; Nanismo nutricional.

Correa Filho HR, Vieira JB, Silva YS, Coelho GE, Cavalcante Fdos A, Pereira Mda P. **Endemic goiter prevalence survey in Brazilian schoolchildren 6 to 14 years old, 1994-1996.** *Rev Panam Salud Publica*. 2002 Nov;12(5):317-26.

**OBJECTIVE:** To determine the national and state-level prevalence of endemic goiter associated with iodine deficiency in schoolchildren between 6 and 14 years of age in Brazil, based on data from a national survey carried out from 1994 to 1996. **MATERIALS:** Thyroid inspection and palpation were performed on 178,774 schoolchildren between 6 and 14 years of age, in all the states of Brazil. A subsample was selected to

measure the iodine content in their urine as well as the iodine content of the table salt used for cooking in their homes. RESULTS: The last survey before this one was carried out in 1975. We found an 86% decrease in the median prevalence of grade 1 and 2 grade 2 goiter, from 12% to 1%. In 21% of the municipalities, the observed prevalence was zero. From the 16,803 urine samples collected for iodine measurement, 7,702 were matched with the corresponding clinical record (a loss of 54%). The median urinary iodine level was 14.0 micrograms/dL, and without a significant correlation with the clinical data on goiter. The median iodine excretion level for the population in the states of Acre, Amapa, and Tocantins was equal to or below 9.0 micrograms/dL. In three municipalities, urinary iodine was below 2.5 micrograms/dL in all the samples: Parana (in the state of Tocantins), Conceicao (Paraiba), and Nova Roma (Goias). The supplementary iodine found in the salt collected from households (458 samples collected) was below recommended levels, with 50% of them below 20 mg/kg (20 ppm). In 7% of the samples, the level of iodine was below 10 mg/kg of salt, even in salt-producing states such as Rio Grande do Norte. CONCLUSIONS: There is evidence goiter is expanding to parts of Brazil that more recently have been opened up for agriculture, with children in those areas being exposed to low levels of supplementary iodine in the salt used for cooking. We found an increased risk of goiter in areas with deficient iodine supplementation ( $< 10$  mg/kg) (odds ratio = 1.85; Cornfield 95% confidence interval: 1.68-2.03).

ESCOSTEGUY, Claudia Caminha, MEDRONHO, Roberto de Andrade, MADRUGA, Roberto *et al.* **Vigilância epidemiológica e avaliação da assistência às meningites.** *Rev. Saúde Pública*, out. 2004, vol.38, no.5, p.657-663. ISSN 0034-8910.

**OBJETIVO:** Analisar o perfil clínico-epidemiológico dos casos de meningite internados em hospital público e os fatores associados à evolução hospitalar. **MÉTODOS:** Foram analisados 694 casos confirmados, notificados e investigados pelo serviço de epidemiologia de um hospital público de 1986 a 2002. Os dados foram coletados do Sistema de Informação de Agravos de Notificação (Sinan), como parte da rotina local de vigilância epidemiológica. Foi realizada análise multivariada por regressão logística. **RESULTADOS:** Etiologias mais frequentes: criptocócica (12,3%; letalidade =37,7%); meningocócica (8,7%; letalidade =13,3%); pneumocócica (7,2%; letalidade =46%); tuberculosa (6,1%; letalidade =40,5%); estafilocócica (5,2%; letalidade =38,9%), viral (5,5%; letalidade =7,9%); hemófilo (2,9%; letalidade =20%). 38,8% dos casos apresentavam etiologia não especificada (letalidade =36%) e 17,3% estavam associados à infecção pelo HIV. Observou-se meningite hospitalar em 27,1% e seqüelas em 9,2% dos casos com alta hospitalar. Variáveis associadas a uma maior chance de óbito: etiologia (referência viral) - tuberculose, criptococo, estafilococo, meningococo, não especificada, outros gram negativos, cândida e pneumococo; infecção pelo HIV; coma. A tríade febre, vômitos e rigidez de nuca associou-se a uma menor chance de óbito. **CONCLUSÕES:** A elevada proporção de etiologia não especificada e letalidade alta podem refletir problemas de processo de assistência e/ou seleção dos casos relacionada ao perfil do hospital. A vigilância epidemiológica operante no nível hospitalar foi capaz de retro-alimentar os serviços com indicadores da assistência, sendo pertinente o uso do Sinan neste nível.

**Palavras-chave:** Meningite [epidemiologia]; Sistemas de informação; Notificação de doenças; Vigilância epidemiológica; Meningite [etiologia]; Meningite [mortalidade]; Pacientes internados.

Holland WW, Bailey R, Bland JM. **Long-term consequences of respiratory disease in infancy.** Epidemiol Community Health. 1978 Dec;32(4):256-9.

In a study of Kent schoolchildren it has been shown that those who had a history of bronchitis under the age of five were more likely to have reported respiratory symptoms as the age of 11 (Bland et al., 1974). After this finding, it was necessary to test whether these differences would continue or diminish as the children grew older. One thousand three hundred schoolchildren in four areas of Kent were studied by physical examination and parental questionnaire at the ages of 5, 11, and 14. The relative risk of having reported respiratory symptoms for children with a history of early bronchitis, asthma, or pneumonia, compared with other children, was the same at the age of 14 as it was at the age of 11. These relationships could not be explained by social class effects, and were probably not due to parental bias in reporting.

KUMMER, Suzane C, GIUGLIANI, Elsa RJ, SUSIN, Lúlie O *et al.* **Evolução do padrão de aleitamento materno.** Rev. Saúde Pública, abr. 2000, vol.34, no.2, p.143-148. ISSN 0034-8910.

**OBJETIVO:** Acompanhar a evolução da prática do aleitamento materno na clientela de maternidade e avaliar as ações para sua promoção. **MÉTODOS:** Foram comparadas duas coortes de crianças nascidas em épocas diferentes em hospital de Porto Alegre, RS, quanto às prevalências do aleitamento materno nos primeiros seis meses de vida das crianças bem como as taxas de interrupção precoce da amamentação. Os dois estudos foram prospectivos, envolvendo 202 crianças na coorte de 1987 e 187 na de 1994, todas saudáveis, com peso de nascimento igual ou superior a 2.500 g, tendo iniciado o aleitamento materno e cujos pais morassem juntos. O acompanhamento foi realizado através de correspondência, no estudo de 1987, e de visitas domiciliares, no estudo de 1994. **RESULTADOS:** As curvas de sobrevivência do aleitamento materno revelaram que o percentual de crianças amamentadas ao longo dos primeiros seis meses foi semelhante nas duas coortes. A prevalência do aleitamento materno exclusivo, apesar de baixa nos dois grupos, foi superior na população de 1994, especialmente entre as crianças cujas mães tinham maior escolaridade. Não houve diferença no índice de interrupção precoce entre as duas coortes (36% na coorte de 1987 e 39% na de 1994). **CONCLUSÕES:** Os resultados apontam para uma apatia do serviço com relação à promoção do aleitamento materno no período estudado, justificando plenamente o investimento na promoção da amamentação, especialmente nas famílias menos privilegiadas.

**Palavras-chave:** Aleitamento materno; Desmame; Promoção da saúde; Fatores socioeconômicos; Maternidades; Leite humano; Lactação.

LERARIO, Daniel D G, GIMENO, Suely G, FRANCO, Laércio J *et al.* **Excesso de peso e gordura abdominal para a síndrome metabólica em nipo-brasileiros.** Rev. Saúde Pública, fev. 2002, vol.36, no.1, p.4-11. ISSN 0034-8910.

**OBJETIVO:** A obesidade, especialmente de distribuição abdominal, associa-se a fatores de risco cardiovasculares como a dislipidemia, a hipertensão arterial (HA) e o diabetes mellitus (DM). A importância desses fatores em nipo-brasileiros foi previamente demonstrada, apesar de a obesidade não ser característica marcante dos migrantes japoneses. Realizou-se estudo com o objetivo de avaliar a prevalência de excesso de peso e a adiposidade central (AC) em nipo-brasileiros e suas relações com distúrbios metabólicos. **MÉTODOS:** A amostra incluiu 530 nipo-brasileiros (40-79 anos) de primeira e segunda gerações, submetidos a medidas antropométricas de pressão arterial, perfil lipídico e teste oral de tolerância à glicose. A prevalência (por ponto e intervalo de confiança) de excesso de peso foi calculada pelo valor de corte  $\geq 26,4$  kg/m<sup>2</sup>. O diagnóstico de AC foi baseado na razão entre as circunferências da cintura e do quadril (RCQ), sendo que valores  $\leq 0,85$  e  $\leq 0,95$ , para mulheres e homens, respectivamente, firmavam esse diagnóstico. **RESULTADOS:** A prevalência de excesso de peso foi de 22,4% (IC95%  $\leq 20,6-28,1$ ), e a de AC, de 67,0% (IC95%  $\leq 63,1-70,9$ ). Além de maiores prevalências de DM, HA e dislipidemia, estratificando-se pelo índice de massa corporal (IMC) e RCQ, indivíduos com excesso de peso e adiposidade central apresentaram pior perfil metabólico: a pressão arterial foi significativamente maior naqueles com excesso de peso, sem e com AC; indivíduos com AC apresentaram maiores índices de glicemia,

triglicéridos, colesterol total e LDL e menor HDL quando comparados aos sem excesso de peso e sem AC; a insulinemia de jejum foi significativamente maior em indivíduos com excesso de peso (sem e com AC) do que naqueles sem excesso de peso e sem AC. **CONCLUSÃO:** A comparação de subgrupos com e sem adiposidade central foi compatível com a hipótese de que a deposição abdominal de gordura representa fator de risco para doenças interligadas pela resistência à insulina, inclusive em população de origem oriental. A alta prevalência de síndrome metabólica nos migrantes japoneses pode ser decorrente da deposição visceral de gordura, implicada na gênese da resistência à insulina.

**Palavras-chave:** Obesidade [epidemiologia]; Prevalência; Pressão arterial; Diabetes mellitus; Hiperlipidemia; Resistência à insulina; Fatores de risco; Índice de massa corporal; Pesos e medidas corporais; Migração; Antropometria; Abdome; Adiposidade central.

[FORNES, Nélida Schmid de](#), [MARTINS, Ignez Salas](#), [VELASQUEZ-MELENDZ, Gustavo](#) *et al.* **Escores de consumo alimentar e níveis lipêmicos em população de São Paulo, Brasil.** *Rev. Saúde Pública*, fev. 2002, vol.36, no.1, p.12-18. ISSN 0034-8910.

**OBJETIVO:** Analisar o padrão de consumo alimentar avaliado por meio de escores de consumo e relacionar esses escores com os níveis de colesterol total e de lipoproteínas de baixa e alta densidades em população da área metropolitana de São Paulo. **MÉTODOS:** Estudo transversal realizado no município de Cotia, São Paulo, em amostra representativa de 1.045 adultos, foram determinados níveis de lipídeos séricos e a ingestão de alimentos por meio da frequência de consumo alimentar. Foram utilizados escores de padrão de consumo, estabelecendo um peso para cada categoria de consumo baseado na frequência anual, obtendo-se, assim, a distribuição quintilar do escore I (alimentos considerados de risco para doenças cardiovasculares) e escore II (alimentos protetores). Foram comparados os valores médios das lipoproteínas para cada um dos quintis pela análise de variância, e foram verificadas possíveis relações entre os escores de consumo e as frações de lipídeos séricos, mediante modelos de regressão linear múltipla (*stepwise forward*). **RESULTADOS:** Observou-se aumento significativo dos níveis médios de lipídeos, segundo quintis de consumo do escore I para colesterol total e para lipoproteína de baixa densidade-colesterol, e constatou-se um comportamento inverso e significativo dos níveis desses lipídeos séricos em relação ao escore II. O escore I correlacionou-se positivamente e significativamente a esses lipídeos, e o escore II apresentou correlação inversa e significativa com esses constituintes sanguíneos. **CONCLUSÕES:** Em estudos populacionais, a análise da frequência de consumo de alimentos por meio de escores pode ser um método de escolha para avaliar qualidade de dieta e de seu potencial efeito nos níveis séricos de colesterol total e de lipoproteínas de baixa densidade.

**Palavras-chave:** Consumo de alimentos; Lipídios [sangue]; Colesterol [sangue]; Lipoproteínas; Gorduras na dieta; Hiperlipidemia; Lipoproteínas do colesterol HDL [sangue]; Lipoproteínas do colesterol LDL [sangue]; Doenças cardiovasculares; Escores.

[MONTEIRO, Carlos Augusto](#), [BENICIO, Maria Helena D'Aquino](#) e [ORTIZ, Luiz Patricio](#). **Tendência secular do peso ao nascer na cidade de São Paulo (1976-1998).** *Rev. Saúde Pública*, dez. 2000, vol.34, no.6, supl, p.26-40. ISSN 0034-8910.

**OBJETIVO:** Resgatar a tendência secular da distribuição do peso ao nascer na cidade de São Paulo, SP, bem como examinar suas possíveis causas, com base em dados coletados por dois inquéritos domiciliares sobre condições de saúde na infância realizados em 1984/85 e em 1995/96, complementados por informações procedentes de levantamento de prontuários de maternidades e por informações do Sistema Estadual de Declarações de Nascidos Vivos. **MÉTODOS:** Os inquéritos domiciliares estudaram amostras probabilísticas da população infantil de São Paulo com idade inferior a cinco anos (n=1.016 em 1984/85; n=1.280 em 1995/96). O levantamento de prontuários estudou uma amostra probabilística dos partos ocorridos nas maternidades da cidade no ano de 1976 (n=5.734). As declarações de nascidos vivos referem-se às coortes de crianças nascidas na cidade entre 1993 e 1998 (cerca de 200 mil crianças por ano). O estudo da distribuição social do peso ao nascer levou em conta a renda familiar *per capita* e a escolaridade materna. A estratégia analítica para estudar os determinantes da tendência secular do peso ao nascer empregou modelos hierárquicos de causalidade, análises multivariadas de regressão e procedimentos análogos aos utilizados para calcular riscos atribuíveis populacionais. **RESULTADOS/CONCLUSÕES:** A

distribuição do peso ao nascer na cidade de São Paulo (média de 3.160 g com 8,9% de pesos <2.500 g) é inferior àquela esperada quando são ótimas as condições do crescimento fetal (média de 3.400-3.500 g com cerca de 4-5% de pesos <2.500 g). Essa distribuição pouco se modificou nos últimos 22 anos (1976-1998). Entretanto, no período, há evidências de evolução desigual do peso ao nascer segundo o nível socioeconômico (NSE) da população. Nos estratos de baixo NSE, a evolução tem sido favorável e isso se deve, aparentemente, ao melhor desempenho do crescimento intra-uterino, o qual poderia decorrer de melhorias em condições econômicas, no peso e na altura das gestantes, na assistência pré-natal e, possivelmente, do declínio no hábito de fumar. Nos estratos de alto NSE, a evolução do peso ao nascer tem sido desfavorável devido, aparentemente, ao aumento na frequência de recém-nascidos prematuros, tendência provocada por fatores ainda não conhecidos.

**Palavras-chave:** Levantamentos epidemiológicos; Peso ao nascer [tendências]; Fatores socioeconômicos; Escolaridade; Renda familiar; Prematuro; Tabagismo; Séries de tempo.

MONTEIRO, Carlos Augusto, SZARFARC, Sophia Cornbluth e MONDINI, Lenise. **Tendência secular da anemia na infância na cidade de São Paulo (1984-1996)**. *Rev. Saúde Pública*, dez. 2000, vol.34, no.6, supl, p.62-72. ISSN 0034-8910.

**OBJETIVO:** Estimar a prevalência e a distribuição social da anemia na infância, estabelecer a tendência secular dessa enfermidade e analisar sua determinação, com base em dados coletados por dois inquéritos domiciliares realizados na cidade de São Paulo, SP, em 1984/85 e em 1995/96. **MÉTODOS:** Os inquéritos estudaram amostras probabilísticas da população residente na cidade com idade entre zero e 59 meses (1.016 em 1984/85 e 1.280 em 1995/96). Amostras de sangue capilar obtidas por punctura digital foram coletadas nos dois inquéritos e analisadas com relação à concentração de hemoglobina. O diagnóstico da anemia correspondeu a concentrações inferiores a 11 g/dL. O estudo da distribuição social da anemia levou em conta tercis da renda familiar *per capita* em cada um dos inquéritos. A estratégia analítica para estudar os determinantes da evolução da prevalência da anemia na população empregou modelos hierárquicos de causalidade, análises multivariadas de regressão e procedimentos análogos aos utilizados para calcular riscos atribuíveis populacionais. **RESULTADOS/CONCLUSÕES:** Houve entre os inquéritos redução significativa na concentração média de hemoglobina (de 11,6 g/dL para 11,0 g/dL) e aumento significativo na prevalência de anemia (de 35,6% para 46,9%). Essa evolução desfavorável foi observada em ambos os sexos, em todas as faixas etárias e em todos os estratos econômicos da população. A evolução tendeu a ser ainda mais desfavorável para o terço mais pobre das crianças da cidade, o que determinou o agravamento das desvantagens desse estrato frente aos demais. Determinantes distais (renda familiar e escolaridade materna) e proximais (tipo de aleitamento) da anemia evoluíram favoravelmente entre os inquéritos e, assim, não puderam explicar o aumento da enfermidade. A estabilidade apurada quanto à densidade de ferro na dieta, em valores inferiores às necessidades, justifica a elevada prevalência da enfermidade, mas não explica seu aumento.

**Palavras-chave:** Anemia [epidemiologia]; Levantamentos epidemiológicos; Inquéritos nutricionais; Anemia ferropriva; Hemoglobinas [deficiência]; Aleitamento materno; Fatores socioeconômicos; Séries de tempo; Criança.

MONTEIRO, Paulo Orlando Alves, VICTORA, Cesar G, BARROS, Fernando C *et al.* **Diagnóstico de sobrepeso em adolescentes: estudo do desempenho de diferentes critérios para o Índice de Massa Corporal**. *Rev. Saúde Pública*, out. 2000, vol.34, no.5, p.506-513. ISSN 0034-8910.

**OBJETIVO:** Pela simplificação da triagem para obesidade na adolescência, avaliou-se o desempenho de diferentes pontos de corte para o Índice de Massa Corporal (IMC) em uma coorte populacional nascida em 1982, em Pelotas, no sul do Brasil. **MÉTODOS:** Foram estudados 493 adolescentes, com idades de 15 a 16 anos, residentes na zona urbana de Pelotas, RS. A obesidade foi definida pelo percentil 85 de IMC mais o percentil 90 das dobras cutâneas tricipital e subescapular, conforme a Organização Mundial da Saúde (OMS). Diversos pontos de corte para sobrepeso tiveram sua sensibilidade e especificidade avaliadas. **RESULTADOS:** Nos meninos, o IMC  $\geq 25$  kg/m<sup>2</sup> apresentou sensibilidade superior a 90% e 5% de falso-positivos. O critério proposto para adolescentes brasileiros apresentou sensibilidade de 100%, mas os falso-

positivos chegaram a 23%. Nas meninas, os pontos de corte coincidiram, apresentando sensibilidade superior a 90%, com até 13% de falso-positivos. Pontos de corte mais altos foram testados, porém pouco melhoraram a especificidade, que foi acompanhada de redução na sensibilidade. **CONCLUSÕES:** O IMC  $\geq 25$  kg/m<sup>2</sup> mostrou o melhor desempenho na detecção de obesidade, parecendo adequado para triagem de adolescentes de ambos os sexos com 15 anos ou mais. Tem a vantagem de ser único, de fácil determinação e compatível com o ponto de corte recomendado pela OMS para adultos. Dispensa o uso de valores de IMC específicos para idade, sexo e medida de dobras cutâneas, sendo, portanto, recomendável para uso em serviços de saúde. **Palavras-chave:** Adolescência; Obesidade [diagnóstico]; Índice de Massa Corporal; Sensibilidade e especificidade; Dobras cutâneas; Peso corporal; Sobrepeso.

**Palavras-chave:** Adolescência; Obesidade [diagnóstico]; Índice de Massa Corporal; Sensibilidade e especificidade; Dobras cutâneas; Peso corporal; Sobrepeso.

OTERO, Ubirani Barros, ROZENFELD, Suely, GADELHA, Angela Maria Jourdan *et al.* **Mortalidade por desnutrição em idosos, região Sudeste do Brasil, 1980-1997.** *Rev. Saúde Pública*, abr. 2002, vol.36, no.2, p.141-148. ISSN 0034-8910.

**OBJETIVO:** Conhecer a freqüência da desnutrição como causa de morte na população idosa. **MÉTODOS:** Foram selecionados indivíduos com 60 anos ou mais de ambos os sexos, dos municípios da região Sudeste, entre 1980 e 1997. As fontes de dados foram o registro de óbitos do Sistema de Informação Sobre Mortalidade (1980-1998) e a população estimada pelo Centro de Desenvolvimento e Planejamento Regional (Cedeplar). Para classificação dos óbitos, empregou-se a CID-9 (260 a 263.9), para os anos de 1980 até 1995, e a CID-10 (E40 a E46), para os anos mais recentes. **RESULTADOS:** No Brasil, entre 1980 e 1997, ocorreram 36.955 óbitos por desnutrição em idosos. A região Sudeste concentrou o maior número  $\approx 23.968$  (64,9%)  $\square$  dentre as demais regiões brasileiras. No Estado de São Paulo, ocorreram 11.067 óbitos por desnutrição em idosos, e, no Rio de Janeiro, 7.763, obtendo, esses dois estados em conjunto, os maiores valores da região. É maior a proporção de óbitos e maiores os coeficientes de mortalidade em indivíduos de 70 anos ou mais do que em indivíduos da faixa etária de 60 a 69 anos, independentemente do sexo. **CONCLUSÃO:** Os resultados preliminares do estudo levantam algumas questões: o papel da desnutrição como causa associada; a tendência de aumento dos óbitos por desnutrição na velhice; o comportamento diferenciado entre estados da mesma região. Análises estatísticas do tipo séries-temporais possivelmente conseguiriam explicar melhor os fenômenos apontados. Será preciso aprofundar o estudo do papel da desnutrição na população com 60 anos ou mais para estabelecer estratégias de intervenção adequadas.

**Palavras-chave:** Mortalidade; Idoso; Transtornos nutricionais; Atestados de óbito; Causa básica da morte; Coeficiente de mortalidade; Distribuição espacial; Desnutrição.

PRADO, Matildes da Silva, ASSIS, Ana Marlúcia Oliveira, MARTINS, Maisa Cruz *et al.* **Hipovitaminose A em crianças de áreas rurais do semi-árido baiano.** *Rev. Saúde Pública*, ago. 1995, vol.29, no.4, p.295-300. ISSN 0034-8910.

Objetivou-se avaliar a distribuição e a magnitude da deficiência de vitamina A e o consumo dietético de 161 crianças de 6 a 72 meses de idade, de áreas rurais do Município de Cansanção-Bahia, Brasil. Os níveis de retinol sérico foram medidos pelo método espectrofotométrico (Bessey-Lowry modificado por Araújo e Flores, 1978). A média do retinol sérico distribuiu-se homogênea entre as diferentes faixas etárias. Níveis inadequados de retinol sérico (deficiente  $< 10,0$   $\mu\text{g/dl}$  e baixos  $< 20,0$   $\mu\text{g/dl}$ ) foram detectados em 44,7% das crianças, caracterizando a deficiência como problema de saúde pública. Os níveis de retinol sérico não mostraram associação estatisticamente significativa com sexo e idade das crianças, contudo as menores de 24 meses apresentaram prevalência mais alta de níveis inadequados. A principal fonte de vitamina A, disponível para essas crianças, é representada pelos carotenóides, em especial beta-caroteno. Foi observada maior diversificação no consumo dos alimentos de conteúdos moderado e baixo em vitamina A no grupo de 24 a 72 meses de idade, sem contudo assegurar níveis adequados de retinol sérico para este grupo etário.

**Palavras-chave:** Deficiência de Vitamina A [epidemiologia]; Inquéritos sobre dieta; População rural.

SOUZA JUNIOR, Paulo Roberto Borges de, SZWARCOWALD, Célia Landmann, BARBOSA JUNIOR, Aristides *et al.* **Infecção pelo HIV durante a gestação: estudo-Sentinela Parturiente, Brasil, 2002.** *Rev. Saúde Pública*, dez. 2004, vol.38, no.6, p.764-772. ISSN 0034-8910.

**OBJETIVO:** Avaliar a cobertura efetiva da detecção da infecção pelo HIV durante a gestação, em âmbito nacional. **MÉTODOS:** A cobertura efetiva do teste de HIV na gestação foi definida como a proporção de gestantes que teve atendimento pré-natal (pelo menos uma consulta), pedido de teste de HIV e conhecimento do resultado antes do parto, sendo estimada por processo de amostragem, utilizando-se as informações coletadas no Estudo-Sentinela Parturiente, 2002. As desigualdades da cobertura efetiva foram analisadas por: grande região; tamanho populacional do município de ocorrência do parto; e grau de instrução da mãe. **RESULTADOS:** A cobertura efetiva do teste de HIV durante a gestação foi estimada em 52%. As enormes desigualdades socioespaciais ficaram evidenciadas na comparação entre as regiões Nordeste (24%) e Sul (72%); entre parturientes analfabetas (19%) com as que têm o ensino fundamental completo (64%); entre as que realizaram o parto em municípios pequenos (36%) com as que o realizaram em municípios com mais de 500 mil habitantes (66%). As recomendações do Ministério da Saúde foram atendidas, completamente, por somente 27% parturientes. **CONCLUSÕES:** Os resultados estabelecem a necessidade de haver medidas voltadas para maior cobertura da detecção do HIV na gestação, e indicam que os programas do Programa Nacional de DST e Aids e os programas de saúde da mulher devem ser intensificados, com estratégias conjuntas entre eles.

**Palavras-chave:** Infecções por HIV, diagnóstico; Transmissão vertical de doença; Efetividade; Distribuição espacial; Fatores socioeconômicos; Cuidado pré-natal; Vigilância de evento sentinela.

TUYNIS AJ, Pequignot G, Jensen OM. **Esophageal cancer in Ille-et-Vilaine in relation to levels of alcohol and tobacco consumption. Risks are multiplying.** *Bull Cancer*. 1977;64(1):45-60.

A retrospective case-control study of 200 male cases of oesophageal cancer and 778 population controls has been carried out in Ille-et-Vilaine (France). The logarithms of the relative risks of developing the disease increase linearly with daily consumption of alcohol and tobacco independently. The combined effect of both fit with a multiplicative model which is proposed. This model could be applicable to other situations. It explains the sex ratio and the urban/rural differences observed in Ille-et-Vilaine. The practical implications for public health purposes are briefly discussed.

VELASQUEZ-MELENDREZ, Gustavo, KAC, Gilberto, VALENTE, Joaquim Gonçalves *et al.* **Evaluation of waist circumference to predict general obesity and arterial hypertension in women in Greater Metropolitan Belo Horizonte, Brazil.** *Cad. Saúde Pública*, May/June 2002, vol.18, no.3, p.765-771. ISSN 0102-311X.

*Avaliou-se a capacidade que a Circunferência da Cintura (CC) tem em identificar valores de IMC  $\geq 25$  (sobrepeso) e  $\geq 30$  (obesidade) a partir de níveis de ação recomendados internacionalmente. Setecentas e noventa e uma mulheres entre 15-59 anos foram recrutadas. Foram calculadas a sensibilidade e a especificidade no diagnóstico de sobrepeso e obesidade e a sensibilidade e especificidade na predição da hipertensão, a partir de valores de CC. As associações foram testadas por análise de regressão linear e regressão logística controlando para o efeito de fatores de confusão. A CC  $\geq 80$  e  $\geq 88$ cm discriminou corretamente 89,8% de mulheres com IMC  $\geq 25$  e 88,5% com IMC  $\geq 30$ . A obesidade abdominal (CC  $\geq 88$ cm) esteve associada significativamente com a hipertensão na análise multivariada (OR = 2,88, IC 95%: 1,77-4,67). A hipertensão foi identificada com sensibilidade de 63,8 e 42,8% e especificidade de 68,0 e 83,3% para CC  $\geq 80$  e  $\geq 88$ , respectivamente. A obesidade abdominal nos pontos de corte propostos pode discriminar adequadamente indivíduos em risco de se tornarem obesos, no entanto, apresenta um poder apenas moderado para discriminar indivíduos com níveis pressóricos.*

**Keywords:** Obesidade; Saúde da Mulher; Quadril; Hipertensão; Índice de Massa Corporal.