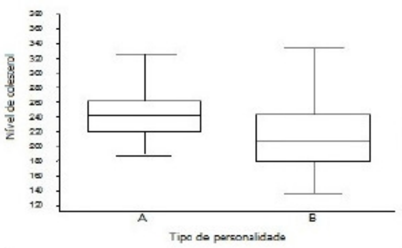
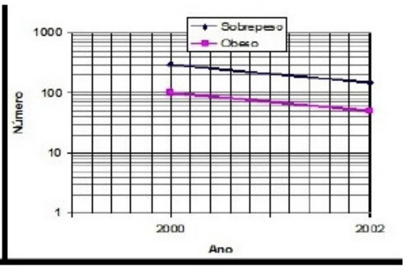
Universidade de São Paulo
Faculdade de Saúde Pública

Curso de Saúde Pública

Disciplina: HEP 146 – Bioestatística I





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

Professoras responsáveis:

Denise Pimentel Bergamaschi

Maria do Rosário Dias de Oliveira Latorre

Alunos do Programa de Aperfeiçoamento de Ensino (PAE):

Igor Conterato Gomes

Luana Fiengo Tanaka

Monitor:

Márcio Roberto Veneroso

São Paulo

2013

HEP 146 BIOESTATÍSTICA I – 2013
PROGRAMAÇÃO DAS AULAS

AULA	DIA	MÊS	CONTEÚDO
1	7	agosto	Apresentação do curso. População, amostra, variáveis, coleta e apuração de dados. Noções de como fazer um questionário.
2 e 3	14 e 21		Apresentação tabular e gráfica de dados.
4	28		Medidas de tendência central (média, mediana e moda)
5	11	setembro	Medidas de dispersão (variância, desvio padrão e percentis)
6	18		Inferência estatística. Noções de amostragem. Distribuição de probabilidades: distribuição Binomial
7	9	outubro	Distribuição Normal, distribuição amostral da média
8	16		Prova I
9	19		Estimação de parâmetros populacionais (média e proporção)
10	23		Testes de Hipóteses: teste de uma proporção.
11	30		Teste de uma média (com variância conhecida e com variância desconhecida)
12	1	novembro	Teste de associação pelo qui-quadrado.
13	6		Delineamento de estudos em epidemiologia Medidas de risco
14	8	novembro	Noções de correlação
15	13		Modelo de regressão linear simples.
16	27		Plantão de dúvidas, preparação dos trabalhos
17	4	dezembro	Prova II
18	11		Seminários - apresentação dos trabalhos
19	18		Prova de recuperação

Informações adicionais:

Local: sala Borges Vieira

Horário: das 14h00 às 18h00.

Avaliação:

- P: serão 2 provas, com consulta.

- E: exercícios em sala sobre o conteúdo que está sendo ministrado, trabalho sobre a análise do perfil dos alunos e apresentação do mesmo.

- $$\boxed{média\,final = \frac{2xP_1 + 3xP_2 + E}{6}}$$

Monitoria:

- das 13 às 14 hs, 2ª. feira, na sala José Maria Gomes.

População, amostra, variável, coleta de dados, apuração de dados e apresentação tabular.

A palavra estatística vem do latim *status* e significa estado. Inicialmente, era utilizada para compilar dados que descreviam características de países (Estados). Em 1662, *John Graunt* publicou estatísticas de nascimentos e mortes. A partir de então, o estudo dos eventos vitais e da ocorrência de doenças e óbitos impulsionou o desenvolvimento da Estatística nos campos teórico e aplicado (Triola, 1999).

Atualmente, índices e indicadores estatísticos fazem parte do dia a dia, tais como taxa de inflação, índice de desemprego, taxa de natalidade, taxa de crescimento populacional, índice de poluição atmosférica, índice de massa corporal, entre outros.

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.

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; deficiente: menor ou igual a 10,0 μg/dl. Estes critérios são do *Commitee on Nutrition for National Defense ICNND/USA*, 1963 (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.

Escala numérica intervalar
Este nível de mensuração possui um valor zero arbitrário.
Ex: temperatura em graus Celsius.

material	^{0}C	^{0}F	|dif^{0}C|	|dif ^{0}F|	dif^{0}C/dif^{0}F	razão^{0}C	razão^{0}F	Razão^{0}C/razão^{0}F
A	20	68	|A-B|=20	|A-B|=36	0,56	A/B=0,50	A/B=0,65	0,77
B	40	104	|B-C|=20	|B-C|=36	0,56	B/C=0,67	B/C=0,74	0,91
C	60	140	|A-C|=40	|A-C|=72	0,56	A/C=0,33	A/C=0,49	0,67

comprimento	cm	polegada	|difcm|	|dif pol|	Difcm/difpol	Razãocm	Razãopol	Razãocm/razãopol
A	20	50,8	|A-B|=15	|A-B|=38,1	0,394	A/B=0,571	A/B=0,571	1
B	35	88,9	|B-C|=5	|B-C|=12,7	0,394	B/C=0,875	B/C=0,875	1
C	40	101,6	|A-C|=20	|A-C|=50,8	0,394	A/C=0,5	A/C=0,5	1

Escala numérica de razões – possui zero inerente à natureza da característica sendo aferida.

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 ={-∞; ...; 0; 0,2; 0,73; 1; 2,48;...; +∞}.

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

De acordo com os níveis de mensuração, pode-se classificar a **natureza das variáveis** segundo a escala de mensuração em:

$$\text{VARIÁVEL:} \begin{cases} \text{qualitativa} \begin{cases} \text{nominal} \\ \text{ordinal} \end{cases} \\ \text{quantitativa} \begin{cases} \text{discreta} \\ \text{contínua} \end{cases} \end{cases}$$

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.

Exemplo 1 -Classificar quanto a natureza, as seguintes variáveis

Variável	Tipo (natureza)
Condição de saúde (doente, não doente)	
Tipo de parto (normal, cesáreo)	
Nível de colesterol sérico (mg/100cc)	
Tempo de um procedimento cirúrgico (minutos)	
Número de praias consideradas poluídas	
Peso (g)	
Estado nutricional (desnutrição, eutrofia, sobrepeso, obesidade)	
Consumo de energia (Kcal)	
Realização da refeição café da manhã (sim/não)	
Número de escolares por série	
Realização de atividade física diária (sim/não)	
Tempo assistindo TV/dia (< 2h, 2 a 4h, >4h)	
Percentual de gordura corporal (%)	

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.

Definições e notação

População: totalidade de elementos sob estudo. Apresentam uma ou mais características em comum. Supor o estudo sobre a ocorrência de sobrepeso em crianças de 7 a 12 anos no Município de São Paulo.

População alvo – todas as crianças nesta faixa etária deste município.
População de estudo – crianças matriculadas em escolas.

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 (μ = 3100 g);

Proporção de crianças de 7 a 12 anos classificadas como obesas, no município de São Paulo (π = 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: $\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}$,

onde $\sum_{i=1}^{N} X_i = X_1 + X_2 + ... + X_N$ 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 = $\bar{x} = 3000g$.

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 com dados pontuais e em intervalos de classe

Notação:

X : variável
x_i : valor observado para o indivíduo i

Apresentação pontual

Ex: 9 indivíduos

X: número de consultas por ano

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

Apuração:

número de consultas	frequência absoluta
1	1
2	3
3	4
5	1

X: idade (anos inteiros)

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

Idade	frequência
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

Peso	frequência
1600	1
2180	1
2250	1
2400	1
2520	1
2725	1
3025	1
3750	1
3950	1

Altura em metros

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

Altura	frequência
1,45	1
1,59	1
1,60	2
1,63	1
1,73	1
1,85	1
2,05	1

Apresentação tabular e gráfica de dados

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

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

Variável	n°	%
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 crianças[1] segundo níveis séricos de retinol. Cansação – Bahia, 1992

Nível de retinol sérico[2]	n	%
Aceitável	89	55,3
Baixo	65	40,4
Deficiente	7	4,3
Total	161	100

[1] 24 –72 meses

[2] aceitável: 20,0 – 49,9 μg/dl; baixo: 10,0 – 19,9 μg/dl; deficiente: <10,0 μg/dl

Fonte: Prado MS et al., 1995.

Apresentação tabular de variável quantitativa contínua

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

Ex:

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

Idade	frequência	%
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: peso (g)
X: 2250, 3025, 1600, 2725, 3750, 3950, 2400, 2180, 2520, 2530

Peso (g)	frequência	%
1500|--2000	1	10
2000|--2500	3	30
2500|--3000	3	30
3000|--3500	1	10
3500|--4000	2	20
Total	10	100

X: Altura (cm)
X: 1,63; 1,60; 1,59; 1,60; 1,45; 1,73; 2,05; 1,85

Altura (cm)	n	%
1,45|--1,55	1	12,5
1,55|--1,65	4	50,0
1,65|--1,75	1	12,5
1,75|--1,85	0	-
1,85|--1,95	1	12,5
1,95|--2,05	0	-
2,05|--2,15	1	12,5
Total	8	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.

Exemplo 2 – Os dados a seguir são de altura de uma amostra de 351 mulheres idosas selecionadas aleatoriamente de uma comunidade para um estudo de osteoporose.

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	158	160	161	163	164	166	170
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	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 alli. *A handbook of small data sets*. Chapman&Hall, 1994.

a) Faça uma apuração dos dados e represente-os em uma tabela.
b) Interprete os resultados.
c) Se entre as 351 mulheres não fossem conhecidas as alturas de 15 delas, como você representaria esses valores?
d) Ao apresentar os dados em uma tabela você iria incluir estes valores?
e) Estas 15 mulheres tinham, na verdade, altura maior que 180 e o investigador, por achar que eram "valores esquisitos" resolveu excluí-los. Você concorda com esta decisão? Justifique.

Tabela de dupla entrada

Distribuição de crianças[1] segundo níveis séricos de retinol[2] e idade. Cansação – Bahia, 1992.

Faixa etária (meses)	Aceitável		Inadequado		Total	
	n	%	n	%	n	%
<12	5	45,5	6	54,5	11	100
12|--24	10	43,5	13	56,5	23	100
24|--36	19	54,3	16	45,7	35	100
36|--48	21	65,6	11	34,5	32	100
48|--60	16	43,2	21	56,8	37	100
60|--73	18	78,3	5	21,7	23	100
Total	89	55,3	72	44,7	161	100

(1) 24 –72 meses.
(2) aceitável: 20,0 – 49,9 μg/dl; baixo: 10,0 – 19,9 μg/dl; deficiente: <10,0 μg/dl.
Fonte: Prado MS et al., 1995.

Exemplo 3
Os dados a seguir são de um estudo que investiga a relação entre níveis de β-caroteno (mg/l) e hábito de fumar em gestantes.

a) Calcule as frequências relativas. Fixando o 100% no total de fumantes e não fumantes.
b) Calcule as frequências relativas. Fixando o 100% no total do nível de B-caroteno (mg/l).
c) Interprete os resultados. Existe alguma indicação de existência de associação entre as variáveis? Justifique.

a)

Distribuição de gestantes segundo níveis de β-caroteno (mg/l) e hábito de fumar.

β-caroteno (mg/l)	Fumante		Não Fumante		Total	
	n	%	N	%	n	%
Baixo (0 – 0,213)	46		74		120	
Normal (0,214 – 1,00)	12		58		70	
Total	58		132		190	

Fonte: Silmara Silva. Tese de Mestrado/FSP/USP.

b)

Distribuição de gestantes segundo níveis de β-caroteno (mg/L) e hábito de fumar.

β-caroteno (mg/l)	Fumante		Não Fumante		Total	
	n	%	n	%	n	%
Baixo (0 – 0,213)	46		74		120	
Normal (0,214 – 1,00)	12		58		70	
Total	58		132		190	

Fonte: Silmara Silva. Tese de Mestrado/FSP/USP.

Exercícios suplementares

Exercício S1

Os dados a seguir são relativos ao número de refeições diária de 50 indivíduos.

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

Fonte X.

a) Apresente os dados em uma tabela.
b) Interprete os resultados.

Exercício S2

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/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

a) Quais variáveis que estão sendo estudadas? Identifique a natureza de cada variável.
b) Apure os dados e apresente a variável nível de colesterol em uma tabela bidimensional, considerando os níveis A e B.
c) Classifique a variável nível de colesterol em duas categorias: nível normal (abaixo de 200 mg/100ml) e nível elevado (200 mg/100ml e mais) e faça uma tabela bidimensional cruzando as variáveis: nível de colesterol (normal e alto) e tipo de comportamento (A e B). Interprete os resultados.

Exercício S3

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.

a) 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).
b) Interprete os resultados.

Exercício S4

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 N	%	Gotemburgo -1972/73 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

Exercício S5

São apresentados na tabela abaixo valores de consumo diário de calorias de uma amostra de mulheres com bulimia. A tabela contém erros.
a) Apresente uma nova tabela sem erros
b) Interprete os dados

Consumo de calorias por dia. Estados Unidos, 1980

Caloria diária (kcal/kg)	10 a 15 anos		16 a 20 anos		Total
	n^0	%	n^0	%	n^0
10,0|--15,0	5	16,166	12		
15,0|--20,0	12	40	42		
20,0|--25,0	7	23,3	19		
25,0|--30,0	4	13,3	15		
30,0|--35,0	0		7		
35,0|--40,0	2	6,7	5		
Total	30		100		

Fonte: Pagano & Gauvreau. Principles of Biostatistics. 2007 (adaptado)

Apresentação gráfica: diagrama de barras, diagramas de setores circulares, diagrama linear, histograma, polígono de frequência, ogiva de frequências acumuladas.

Diagrama de barras

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

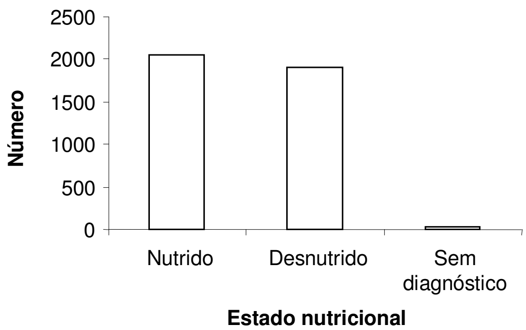
Características: figuras geométricas (barras) separadas e bases de mesmo tamanho. A altura das barras é proporcional às frequências.

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 (Soares JF, Siqueira AL. Introdução à Estatística Médica, COOPMED,Belo Horizonte, MG 2002). Os dados da tabela são retirados deste estudo.

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

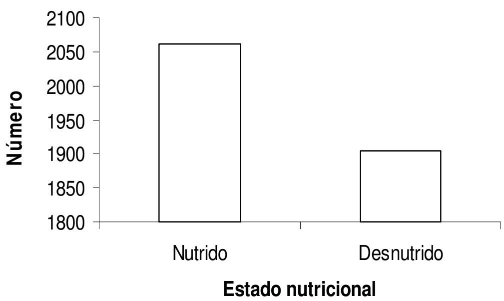
Estado nutricional	N	%
Nutrido	2061	51,5
Desnutrido	1905	47,6
Sem diagnóstico	34	0,9
Total	4000	100,0

Fonte: adaptado de Soares JF, Siqueira AL, 2002.



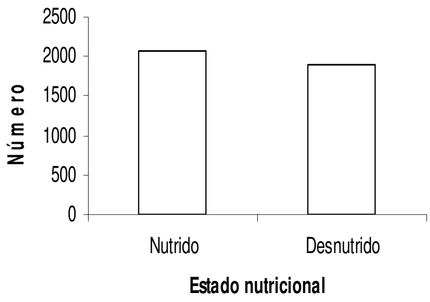
Fonte: adaptado de Soares JF, Siqueira AL, 2002.
Distribuição de pacientes segundo estado nutricional. IBRANUTRI, maio a novembro, 1996.

Esta representação gráfica está correta?



Atenção: cuidado com a origem!

Diagrama de barras da tabela anterior, excluindo-se os registros da categoria sem diagnóstico



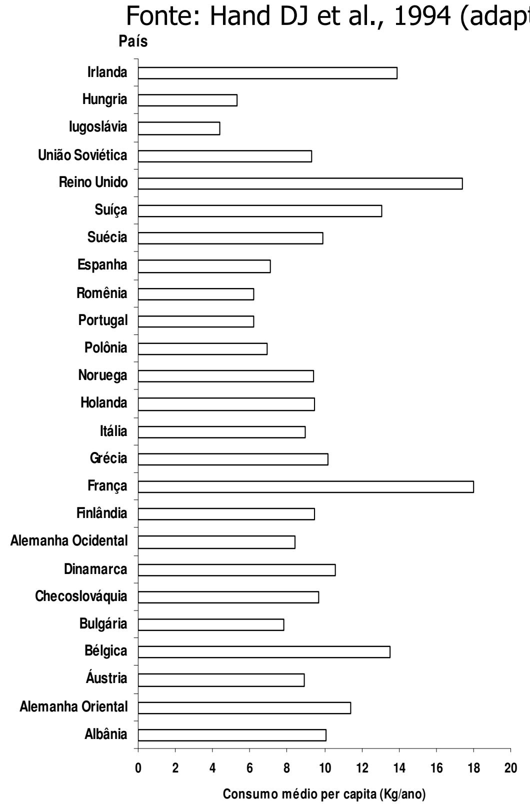
Fonte: adaptado de Soares JF, Siqueira AL, 2002.
Distribuição de pacientes segundo estado nutricional. IBRANUTRI, maio a novembro, 1996.

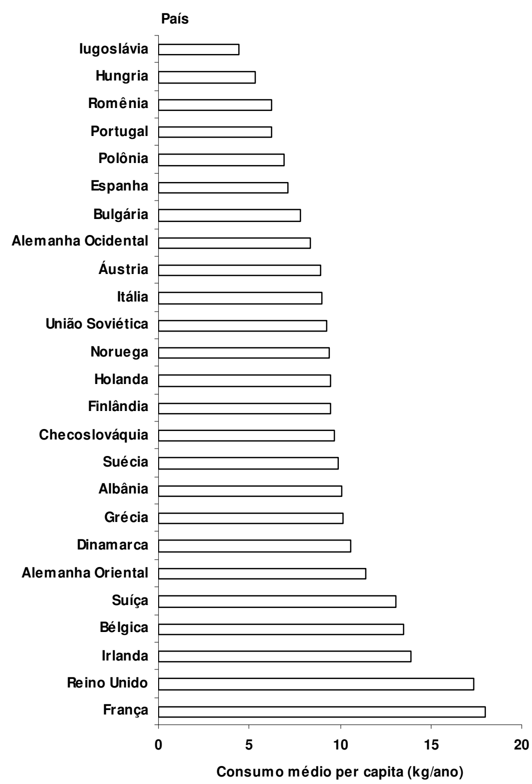
Variável qualitativa nominal

Distribuição do consumo médio per capita de carne vermelha (kg/ano) segundo país.

País	Consumo anual	País	Consumo anual
Albânia	10,1	Noruega	9,4
Alemanha Oriental	11,4	Polônia	6,9
Áustria	8,9	Portugal	6,2
Bélgica	13,5	Romênia	6,2
Bulgária	7,8	Espanha	7,1
Checoslováquia	9,7	Suécia	9,9
Dinamarca	10,6	Suíça	13,1
Alemanha Ocidental	8,4	Reino Unido	17,4
Finlândia	9,5	União Soviética	9,3
França	18,0	Iugoslávia	4,4
Grécia	10,2	Hungria	5,3
Itália	9,0	Irlanda	13,9
Holanda	9,5		

Fonte: Hand DJ et al., 1994 (adaptado).





Fonte: Hand DJ et al., 1994 (adaptado).
Distribuição do consumo médio (kg/ano) per capita de carne vermelha, segundo país.

Variável qualitativa ordinal

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

Estado nutricional[a]	n	%
Nutrido	2061	52,0
Desnutrido moderado	1407	35,4
Desnutrido grave	498	12,6
Total	3966	100

[a] excluindo-se 34 (0,9%) pacientes sem diagnóstico.
Fonte: adaptado de Soares JF, Siqueira AL, 2002.

[a] excluindo-se 34 (0,9%)pacientes sem diagnóstico.
Fonte: adaptado de Soares JF, Siqueira AL, 2002.
Distribuição de pacientes segundo estado nutricional. IBRANUTRI, maio a novembro, 1996.

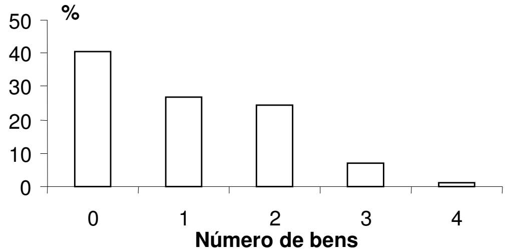
Variável quantitativa discreta: número de bens

Foi realizada, no período de outubro de 1998 a outubro 1999, a pesquisa "Alimentação no primeiro ano de vida", onde se estudou 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. Hospital Universitário/USP, São Paulo 1999.

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



*automóvel, telefone, TV a cabo e computador
Distribuição de famílias segundo número de bens*. Hospital Universitário/USP, São Paulo 1999.

Exemplo 4
Os dados a seguir são relativos ao número de refeições diárias de 50 indivíduos, utilizados no exercício S1.

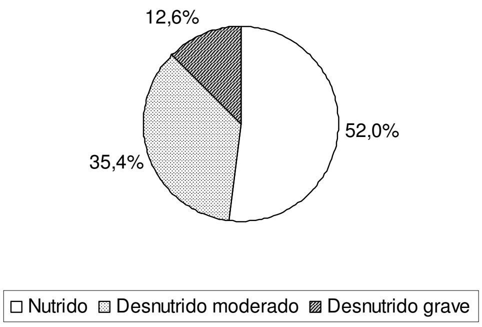
2	3	2	1	2	6	5	4	3
1	2	2	1	2	5	6	4	3
2	2	3	2	3	4	2	3	2
3	2	3	3	3	4	3	4	5
3	1	4	3	4	4	3		
3	1	6	4	4	2	4		

A) Apresente os dados em um gráfico.
B) Interprete o gráfico.

Diagrama de setores circulares
Variáveis: qualitativa nominal e qualitativa ordinal
Distribuição de pacientes segundo estado nutricional. IBRANUTRI, maio a novembro, 1996.

Estado nutricional[a]	n	%
Nutrido	2061	52,0
Desnutrido moderado	1407	35,4
Desnutrido grave	498	12,6
Total	3966	100

[a] excluindo-se 34 (0,9%)pacientes sem diagnóstico.
Fonte: adaptado de Soares JF, Siqueira AL, 2002.



[a] excluindo-se 34 (0,9%)pacientes sem diagnóstico.
Fonte: adaptado de Soares JF, Siqueira AL, 2002.
Distribuição de pacientes[(a)] segundo estado nutricional. IBRANUTRI, maio a novembro, 1996.

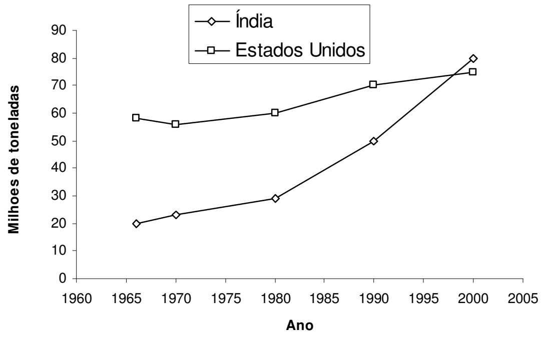
Diagrama linear

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.

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



Fonte: *State of the World*, 2001.

Exemplo 5

Os dados são referentes a produção (kg) de carne de peixes e de carne vermelha e de carneiro por pessoa, no mundo, no período de 1950 a 2000.

Ano	Pesca oceânica (kg)	Carne vermelha e de carneiro (kg)
1950	7,9	9,0
1960	12,0	10,0
1970	16,1	12,0
1980	15,5	11,9
1990	16,3	12,0
2000	15,0	11,7

Fonte: State of the World, 2001. The Worldwatch Institute.

a) Apresente os dados em um gráfico.
b) Interprete os resultados.

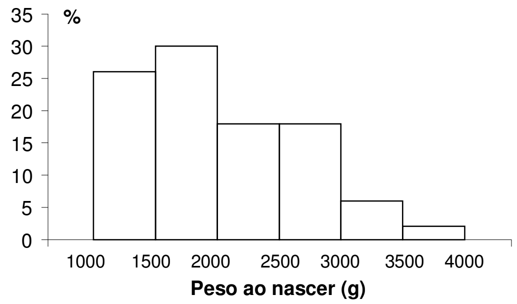
Histograma

Adequado para representar 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, Gupta JM. (1973).

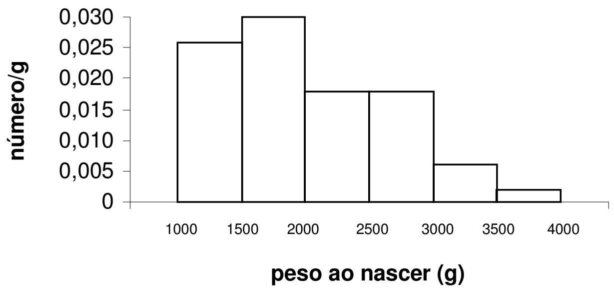


Fonte: van Vliet PKJ, Gupta JM. (1973)

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

Notar que o gráfico pode ser construído considerando-se pessoas por unidade de medida (densidade)

Peso(g)	Nº	Amplitude	Nº/amplitude	(Nº/amplitude)x1000
1000 |-- 1500	13	500	0,026	26
1500 |-- 2000	15	500	0,030	30
2000 |-- 2500	9	500	0,018	18
2500 |-- 3000	9	500	0,018	18
3000 |-- 3500	3	500	0,006	6
3500 |-- 4000	1	500	0,002	2
Total	50			



Fonte: van Vliet PKJ, Gupta JM. (1973).

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

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.

Exemplo 6

Os dados são referentes à distribuição de pacientes segundo taxa de albumina no sangue (g/dl).

Taxa de albumina (g/dl)	Nº	%
4,40|-4,60	6	10,0
4,60|-4,80	11	18,3
4,80|-5,00	14	23,3
5,00|-5,20	18	30,0
5,20|-5,40	8	13,3
5,40|-5,60	2	3,3
5,60|-5,80	0	-
5,80|-6,00	1	1,7
Total	60	100

Fonte: Soares JF, Siqueira AL. COOPMED, 2002.

a) Apresente os dados em um histograma.
b) Interprete os resultados.

Intervalos de classe com amplitudes diferentes

Distribuição de mulheres idosas segundo a altura. Local X, Ano Y.

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.

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. Local X, Ano Y.

Cuidado: Sem fazer o ajuste, o gráfico fica errado e pode levar a conclusões incorretas.

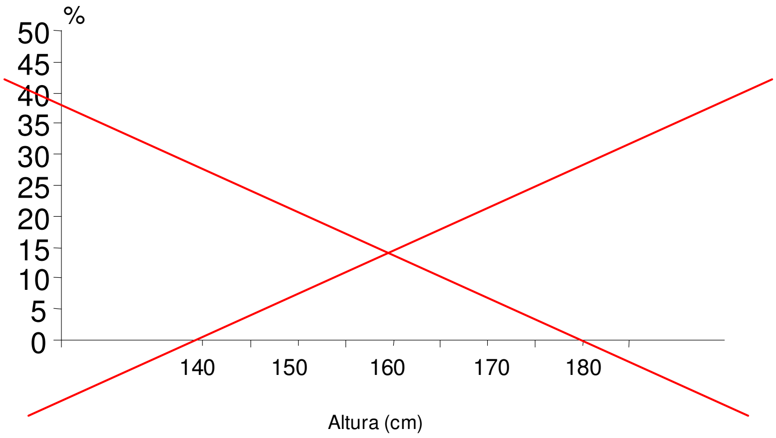
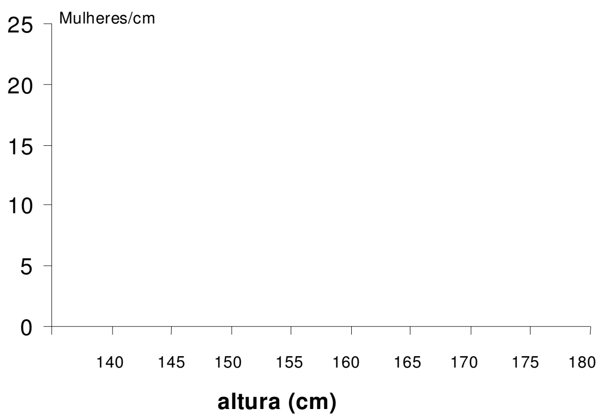


Gráfico correto, com o ajuste para intervalos de classe com amplitudes diferentes.



Fonte: Hand DJ et al., 1994.

Exemplo 7

Os dados a seguir são da altura (cm) de uma amostra de mulheres de Bangladesh.

Distribuição de mulheres segundo altura (cm). Bangladesh, Ano X.

Altura (cm)	número
137,0 |--140,0	71
140,0 |--143,0	137
143,0 |--145,0	154
145,0 |--147,0	199
147,0 |--150,0	279
150,0 |--153,0	221
153,0 |--155,0	94
155,0 |--157,0	51
157,0 |--160,0	37
Total	1243

Fonte: Hand DJ et al, 1994 (adaptado).

a) Represente os dados acima graficamente em um histograma.
b) Interprete os resultados.

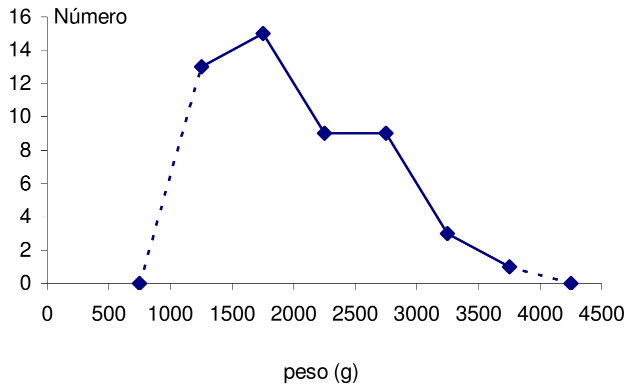
Polígono de frequência simples

Intervalos de classe com mesma amplitude

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

Peso(g)	N^o	%
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). Local X. Ano Y

Exemplo 8

Os dados a seguir são referentes à distribuição de usuárias do Serviço de Saúde X segundo idade (anos). Município de São Paulo, 2009.

Idade (anos)	n	%
15|-- 20	14	19,5
20|-- 25	24	33,3
25|-- 30	16	22,2
30|-- 35	9	12,5
35|-- 40	8	11,1
40|--45	1	1,4
Total	72	100

Fonte: Dados hipotéticos.

a) Apresente a variável em um polígono de frequências simples.
b) Interprete os resultados.

Intervalos de classe com amplitudes diferentes

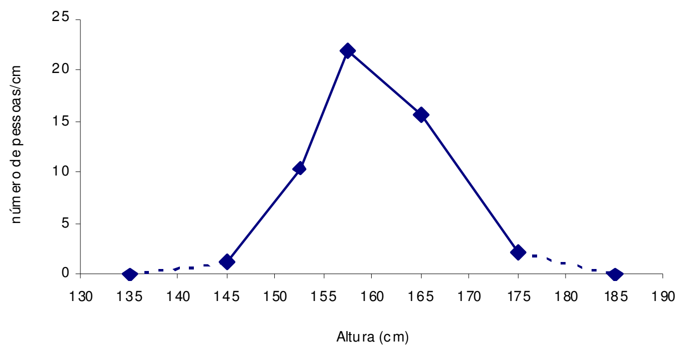
Distribuição de mulheres idosas segundo a altura. Local X, Ano Y.

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.

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). Local X, Ano Y.

Exemplo 9

Distribuição de homens segundo nível de glicose no sangue (mg%). Local X, Ano Y.

Nível de glicose no sangue (mg%)	n
50|-- 100	13
100|-- 150	45
150|-- 200	28
200|-- 250	10
250|-- 300	3
300|-- 450	1

Fonte: X.

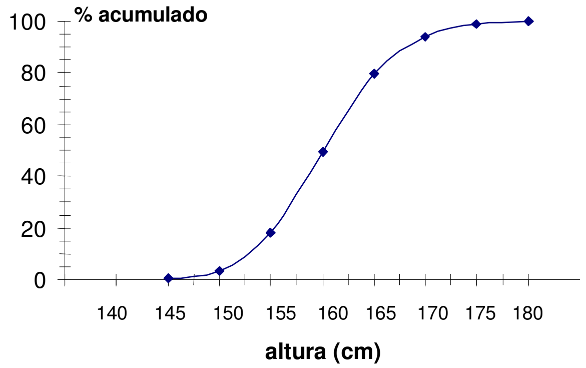
a) Apresente os dados acima graficamente utilizando o polígono de frequências simples.
b) Interprete os resultados.

Polígono (ogiva) de frequências acumuladas

Distribuição de mulheres idosas segundo a altura. Local X, Ano Y.

Altura (cm)	Nº	%	% acumulado
140|-145	1	0,29	0,29
145|-150	11	3,13	3,42
150|-155	52	14,81	18,23
155|-160	109	31,05	49,28
160|-165	106	30,20	79,48
165|-170	50	14,25	93,73
170|-175	18	5,13	98,86
175|-180	4	1,14	100
Total	351	100	

Fonte: Hand DJ et al., 1994.



Fonte: Hand DJ et al., 1994.

Distribuição acumulada de mulheres idosas segundo a altura. Local X, Ano Y.

Percentil	Valor da variável	Medidas estatísticas
25%	156 cm	Q1 – primeiro quartil
50%	160 cm	Q2 - segundo quartil ou mediana
75%	164 cm	Q3 – terceiro quartil

Exemplo 10

Os dados a seguir são medidas de circunferência do tórax (polegadas) de 5732 soldados escoceses apresentados pelo matemático belga *Adolphe Quetelet* (1796–1874).

Medida (polegada)	número	%	% acumulada
33,0 |– 34,0	3		
34,0 |– 35,0	19		
35,0 |– 36,0	81		
36,0 |– 37,0	189		
37,0 |– 38,0	409		
38,0 |– 39,0	753		
39,0 |– 40,0	1062		
40,0 |– 41,0	1082		
41,0 |– 42,0	935		
42,0 |– 43,0	646		
43,0 |– 44,0	313		
44,0 |– 45,0	168		
45,0 |– 46,0	50		
46,0 |– 47,0	18		
47,0 |– 48,0	3		
48,0 |– 49,0	1		
Total	5732		

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

a) Represente os dados em um polígono de frequências acumuladas.
b) Utilizando o gráfico, identifique o valor da circunferência de tórax que deixa 25% dos indivíduos abaixo.
c) Qual o valor de circunferência do tórax que divide a distribuição em 2 partes iguais, isto é, qual é o valor da variável que deixa 50% das observações abaixo dele?
d) Qual a proporção de soldados com circunferência do tórax entre 40 a 42 polegadas?
e) Qual é o valor de circunferência do tórax que deixa 95% dos soldados abaixo dele?

Representação gráfica de duas variáveis qualitativas

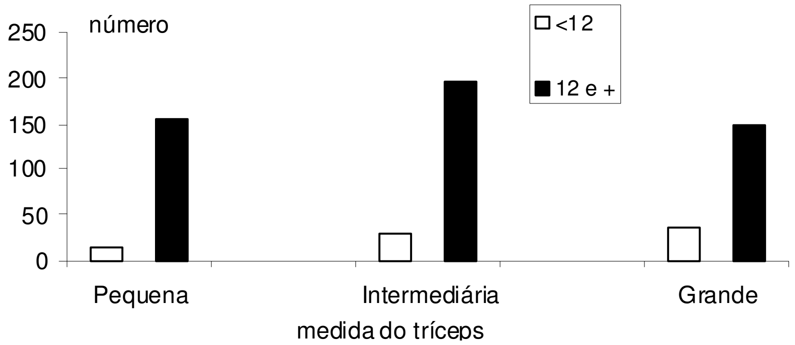
Os dados são de um estudo de obesidade em mulheres da zona urbana de Trinidade e Tobago, realizado em 1985, que estuda a relação entre idade da menarca e a medida do tríceps.

Distribuição de mulheres segundo idade da menarca e medida do tríceps. Trinidade e Tobago,1985.

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

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

Investigando-se a distribuição da idade segundo medida do tríceps tem-se:



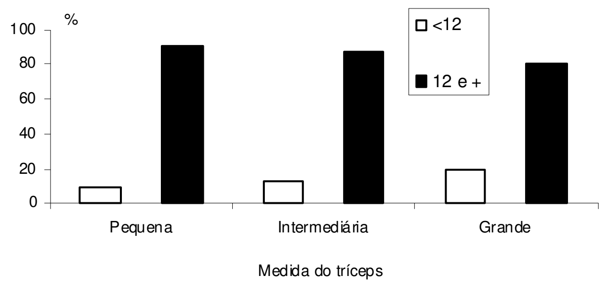
Fonte: Hand DJ et al. *A handbook of small data sets*. Chapman & Hall, 1994.
Distribuição de mulheres segundo idade da menarca e medida do tríceps. Trinidade e Tobago,1985.

Calculando-se as porcentagens, tomando-se as categorias da medida do tríceps como 100%, tem-se:

Distribuição de mulheres segundo idade da menarca e medida do tríceps. Trinidade e Tobago, 1985.

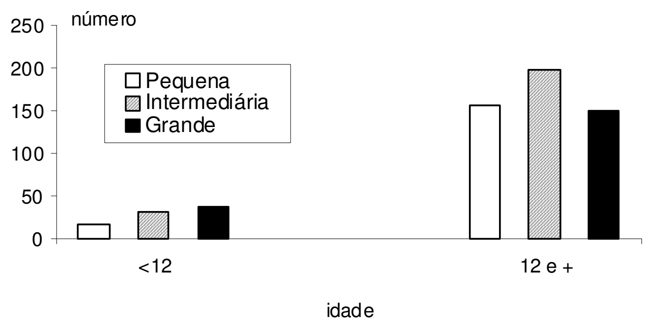
Idade (anos)	Medida do tríceps							
	Pequena		Intermediária		Grande		Total	
	n	%	n	%	N	%	n	%
<12	15	8,8	29	12,8	36	19,4	80	13,7
12 e +	156	91,2	197	87,2	150	80,6	503	86,3
Total	171	100	226	100	186	100	583	100

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



Fonte: Hand DJ et al. *A handbook of small data sets*. Chapman & Hall, 1994.
Distribuição de mulheres segundo idade da menarca e medida do tríceps. Trinidade e Tobago, 1985.

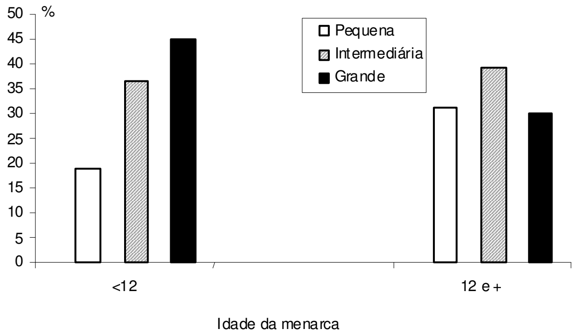
Investigando-se a distribuição da medida do tríceps segundo a idade:



Fonte: Hand DJ et al. *A handbook of small data sets*. Chapman&Hall, 1994.
Distribuição de mulheres segundo idade da menarca e medida do tríceps. Trinidade e Tobago, 1985.

Calculando-se as porcentagens tem-se:

Idade (anos)	Medida do tríceps							
	Pequena		Intermediária		Grande		Total	
	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



Exemplo 11

Considere os dados apresentados na tabela abaixo, coletados em estudo de incidência de óbitos entre crianças submetidas a uma intervenção com vitamina A.

Distribuição de crianças segundo condição de sobrevivência e suplementação. Indonésia, 1986

Grupo de tratamento	Condição de sobrevivência					
	Óbitos		Sobreviventes		total	
	n	%	n	%	n	%
Recebeu vitamina A	50		650		700	
Não recebeu vitamina A	100		515		615	
TOTAL	150		1165		1315	

Fonte: Moore Ds & McCabe G. Introdução à Prática da Estatística (adaptado)

a) Apresente os dados graficamente.
b) Interprete os resultados.

Representação gráfica de duas variáveis quantitativas

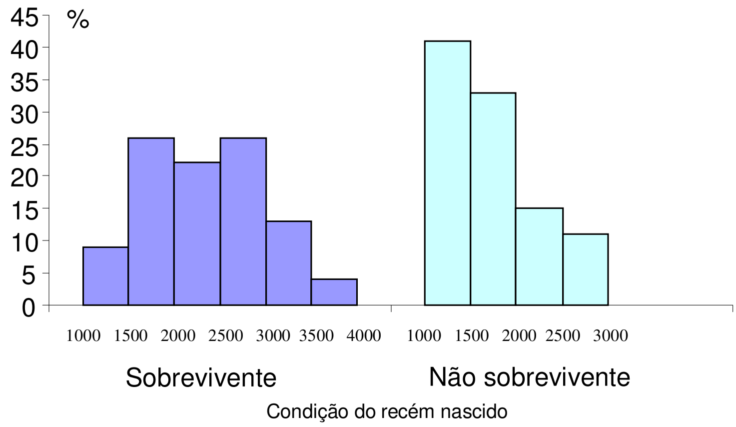
Histograma

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. Austrália, 1993.

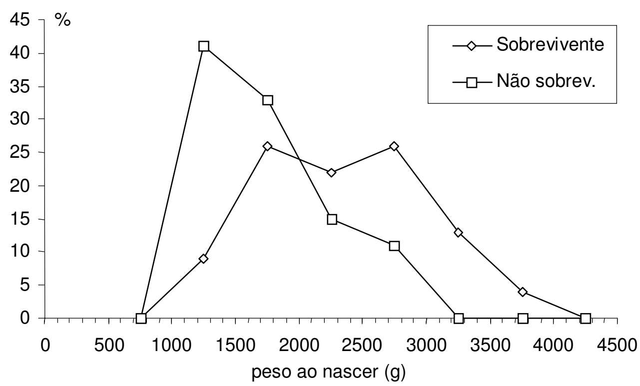
Peso(g)	Sobrevivente		Não sobrevivente		Total	
	nº	%	nº	%	nº	%
1000 |-- 1500	2	9	11	41	13	26
1500 |-- 2000	6	26	9	33	15	30
2000 |-- 2500	5	22	4	15	9	18
2500 |-- 3000	6	26	3	11	9	18
3000 |-- 3500	3	13	0	-	3	6
3500 |-- 4000	1	4	0	-	1	2
Total	23	100	27	100	50	100

Fonte: Hand DJ et al., 1994.



Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascidos. Austrália, 1993.
Fonte: Hand DJ et al. *A handbook of small data sets*. Chapman & Hall, 1994.

Polígono de frequências



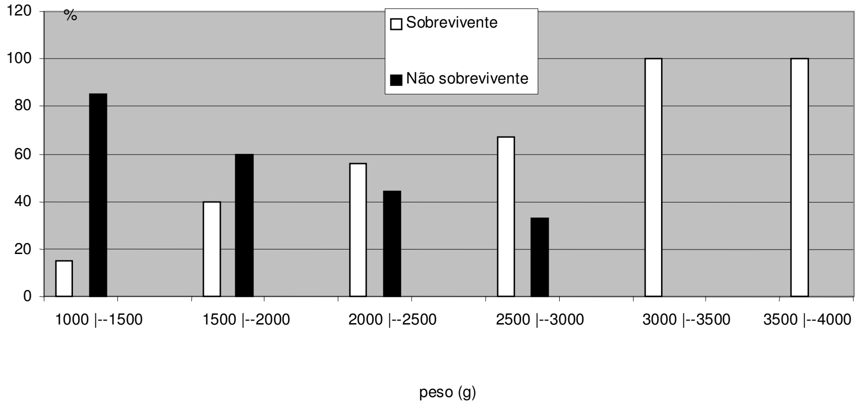
Fonte: Hand DJ et al., 1994.
Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascido. Austrália, 1993.

Fixando-se os percentuais no peso ao nascer:

Diagrama de barras
Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascido. Austrália, 1993.

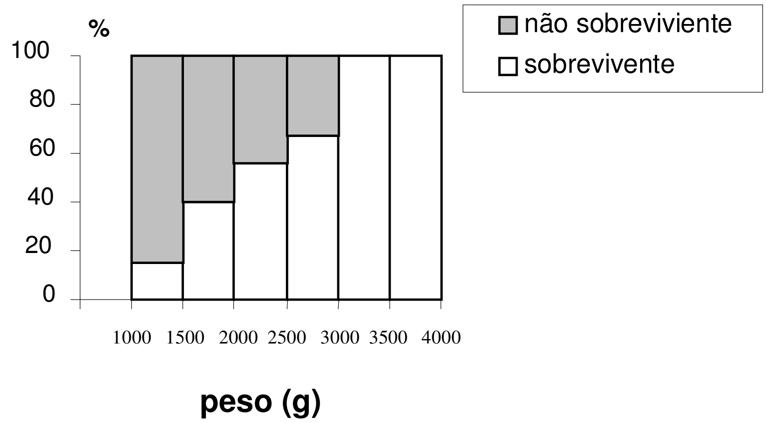
Peso(g)	Sobrevivente		Não sobrevivente		Total	
	n^o	%	n^o	%	n^o	%
1000 |-- 1500	2	15	11	85	13	100
1500 |-- 2000	6	40	9	60	15	100
2000 |-- 2500	5	56	4	44	9	100
2500 |-- 3000	6	67	3	33	9	100
3000 |-- 3500	3	100	0	-	3	100
3500 |-- 4000	1	100	0	-	1	100
Total	23	46	27	54	50	100

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



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

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascido. Austrália, 1993.



Fonte: Hand DJ et al. *A handbook of small data sets*. Chapman & Hall, 1994.
Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g) e condição do recém-nascido. Austrália, 1993.

Exemplo 12
Utilize os dados da tabela e apresente-os graficamente.
Distribuição de escolares de 7 a 10 anos segundo peso e sexo. Duas escolas do Município de São Paulo, 2005.

Peso (kg)	Sexo			
	Masculino		Feminino	
	n	%	n	%
15,0 |-- 25,0	52	18,5	68	24,7
25,0 |-- 35,0	146	51,9	132	48,0
35,0 |-- 45,0	59	21,0	53	19,3
45,0 |-- 55,0	11	3,9	18	6,5
55,0 |-- 65,0	10	3,6	2	0,7
65,0 |-- 75,0	3	1,1	1	0,4
75,0 |-- 85,0	0	-	0	-
85,0 |-- 95,0	0	-	1	0,4
Total	281	100	275	100

Fonte: Koga CR. Estado nutricional de escolares de 7 a 10 anos de idade: diagnóstico e comparação de métodos. São Paulo; 2005. [Dissertação de Mestrado-Faculdade de Saúde Pública da Universidade de São Paulo/USP].

Exemplo 13
Utilize os dados da tabela e apresente-os graficamente.
Distribuição de percentual (%) de escolares segundo estatura (cm), sexo e idade. Duas escolas do Município de São Paulo, 2005.

Estatura (cm)	Sexo			
	Masculino		Feminino	
	nº	%	nº	%
105,0 – 119,9	3	1,1	16	5,8
120,0 – 124,9	36	12,8	31	11,3
125,0 – 129,9	61	21,7	74	26,9
130,0 – 134,9	57	20,3	41	14,9
135,0 – 139,9	52	18,5	43	15,6
140,0 – 144,9	38	13,5	30	10,9
145,0 – 149,9	22	7,8	26	9,4
150,0 – 159,9	12	4,3	14	5,1
Total	281	100	275	100

Fonte: Koga CR. Estado nutricional de escolares de 7 a 10 anos de idade: diagnóstico e comparação de métodos. São Paulo; 2005. [Dissertação de Mestrado-Faculdade de Saúde Pública da Universidade de São Paulo/USP].

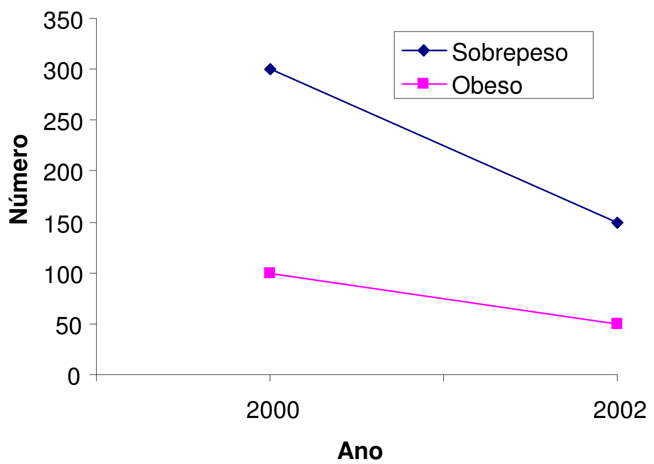
Escala aritmética e escala logarítmica

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

Ano	Sobrepeso	Obesas
2000	300	100
2002	150	50

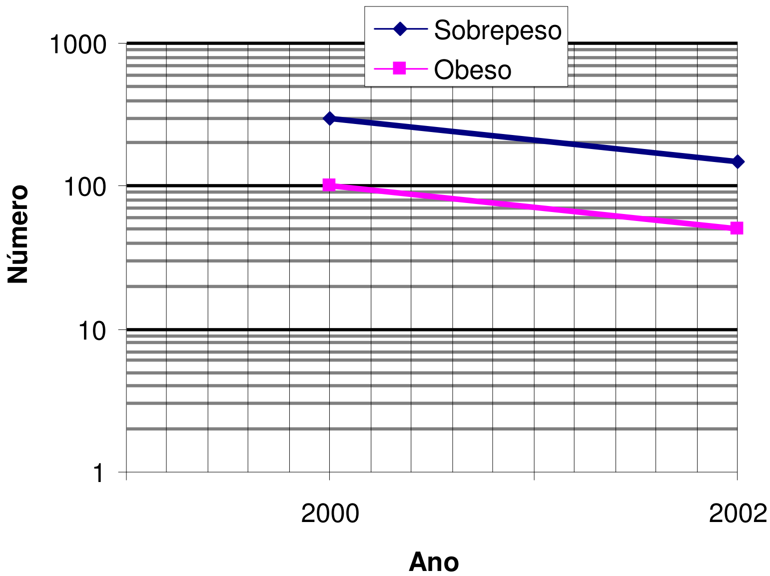
Fonte: dados hipotéticos.

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.

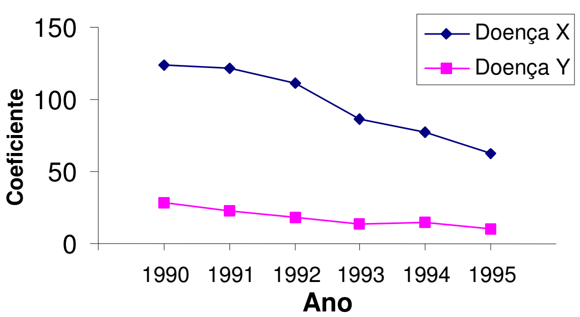
Número de crianças segundo massa corporal. Escola X, 2000 e 2002.

Gráfico em escala aritmética

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

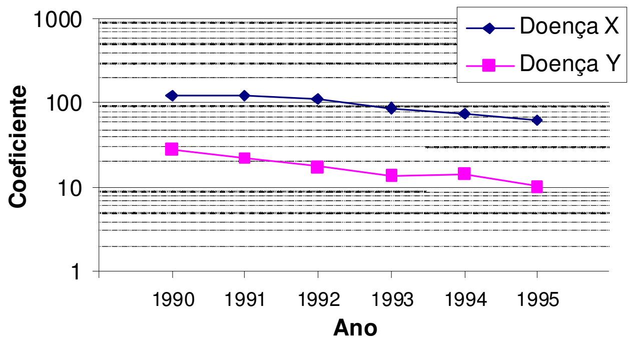
Ano	Doença X	Doença Y
1990	123,5	28,7
1991	121,4	22,4
1992	111,9	17,7
1993	85,9	13,9
1994	77,1	14,8
1995	62,2	10,5

Fonte: dados hipotéticos.



Fonte: dados hipotéticos.
Coeficiente de mortalidade pela doença X e Y (100.000 hab.). Determinada localidade, 1990- 1995.

Gráfico em escala logarítmica



Fonte: dados hipotéticos.
Coeficiente de mortalidade pela doença X e Y (100.000 hab.). Determinada localidade, 1990 - 1995.

Exemplo 14
Os dados a seguir são referentes à mortalidade por câncer de esôfago, segundo sexo, no município de São Paulo no período de 1968-1998.

Coeficientes de mortalidade por câncer de esôfago (por 100.000 hab.). Município de São Paulo, 1968-1998.

Ano	Masculino	Feminino
1968	8,81	2,00
1973	12,38	2,61
1978	10,93	1,98
1983	9,41	2,00
1988	8,60	1,67
1993	8,33	1,27
1998	8,37	1,12

Fonte: Incidência de câncer no Município de São Paulo, 1997-1998. Registro de Câncer de São Paulo. FSP/USP.

a) Represente os coeficientes de mortalidade por câncer de esôfago para o sexo masculino e feminino em um único gráfico, utilizando escala aritmética.

b) Represente os coeficientes de mortalidade por câncer de esôfago para o sexo masculino e feminino em um único gráfico utilizando escala logarítmica.

c) Comente os gráficos dos itens a e b. Qual a melhor representação para os dados?

Exercícios suplementares

Exercício S6

Apresente a variável peso ao nascer graficamente utilizando a variável definida em duas categorias, conforme tabela abaixo.

Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g). Austrália, 1993.

Peso(g)	Nº	%
Baixo peso (<2500 g)	37	74,0
Não baixo peso (2500 g e mais)	13	26,0
Total	50	100

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

Exercício S7

Apresentar a variável comprimento ao nascer em um histograma.

Distribuição de recém-nascidos segundo comprimento ao nascer (cm). Hospital X, 2009.

Comprimento (cm)	n	%
40|--43	1	1,2
43|--46	45	55,6
46|--49	25	30,9
49|--52	4	5
52|--55	3	3,7
55|--58	1	1,2
58|--61	2	2,5
Total	81	100

Fonte: dados hipotéticos.

Exercício S8

Defeitos do tubo neural são má formações congênitas que surgem durante o desenvolvimento fetal. É conhecida como spina bífida. Estes dados são de um estudo realizado no país de Gales – Reino Unido, para investigar possível associação entre defeito do tubo neural e dieta materna. O estudo é do tipo caso-controle: mães que tinham tido bebês com spina bífida (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 casos (acometidos de spina bífida) e controles segundo dieta da mãe. Reino Unido, ano X.

Dieta	Casos		Controles		Total	
	n	%	n	%	n	%
Boa	34		43		77	
Razoável	110		48		158	
Pobre	100		32		132	
Total	244		123		367	

Fonte: Hand DJ ET al., 1994.

a) Calcular percentuais tomando-se como 100% o grupo (caso, controle) e interprete os resultados.
b) Apresentar os dados em um gráfico.

Exercício S9

Represente os dados da tabela em um polígono de frequências e interprete os resultados. Trata-se de condenados por embriaguez em Londres, 1970.

Idade	Homens		Mulheres	
	Número	%	Número	%
0 |--30	185	20,5	4	9,1
30 |-- 40	207	22,9	13	29,5
40 |-- 50	260	28,8	10	22,7
50 |--60	180	19,9	7	15,9
60 |--80	71	7,9	10	22,7
total	903	100	44	100

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

Exercício S10

Analise o gráfico abaixo. Este artigo avaliou o dano causado por 20 drogas ao indivíduo que utiliza e às outras pessoas.

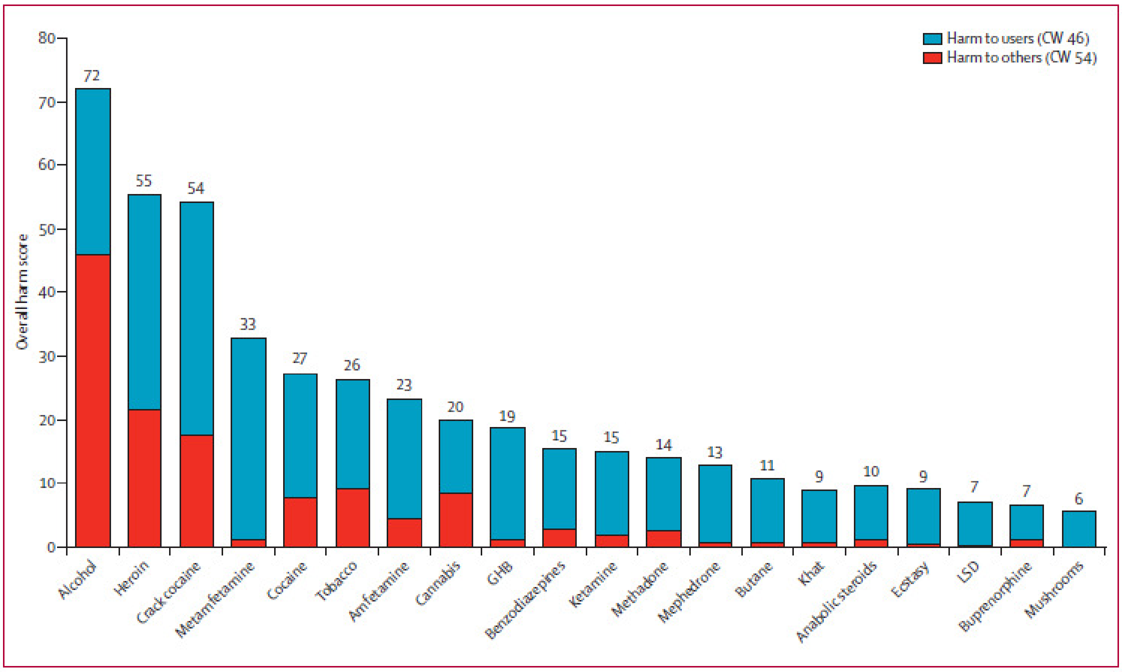


Figure 2: **Drugs ordered by their overall harm scores, showing the separate contributions to the overall scores of harms to users and harm to others**
The weights after normalisation (0–100) are shown in the key (cumulative in the sense of the sum of all the normalised weights for all the criteria to users, 46; and for all the criteria to others, 54). CW=cumulative weight. GHB=γ hydroxybutyric acid. LSD=lysergic acid diethylamide.

Fonte: Nutt D et al. Drug harms in the UK: a multicriteria decision analysis. The Lancet. 2010; v 376: 1558-1564.

Medidas de tendência central (média, mediana e moda)

Medidas de tendência central

<u>Média aritmética</u>
Notação:

X → variável

N → tamanho da população

n → tamanho da amostra

→ Média populacional (parâmetro, geralmente desconhecido)

$\overline{X}$ → Estatística (fórmula)

$\overline{x}$ → Média amostral (estimativa, valor calculado na amostra)

Média aritmética é o valor que indica o centro de equilíbrio de uma distribuição de frequências 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, ..., x_n, a média aritmética ($\bar{x}$) é igual a:

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

OBS:

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

Exemplo:

Os dados a seguir são provenientes do grupo *Western Collaborative Group Study.* Grupo tipo A: pessoas caracterizadas pela urgência, agressividade e ambição. Os participantes de tipo B são mais relaxados, não competitivos e menos preocupados.

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

Colesterol médio:

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

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

$$\bar{x}_B =$$

O nível médio de colesterol dos homens do grupo A é 245,1 mg/100ml e do tipo B _____________.

Exemplo 15

Os dados a seguir são provenientes de um estudo que avaliou o consumo alimentar de crianças de 7 a 10 anos de uma escola pública do município de São Paulo no ano de 2008. Os dados apresentados são de 15 meninos e 10 meninas para os quais foram investigados o consumo em energia (Kcal) de um dia alimentar. Calcule a média aritmética do consumo de energia para cada sexo:

Meninos

1976	3234	1405	1410	1782	2167	1917	2622	1824	3912
1412	1635	2230	1241	1866					

$$\bar{x}_{Meninos} =$$

Meninas

2002	2964	2203	1478	1151	1083	1362	1392	1637	1628

$$\bar{x}_{Meninas} =$$

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.

Ex:

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

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

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

Ordenando-se os valores:

Mediana:

Exemplo 16

Com os dados do exemplo 18, calcule a quantidade mediana de energia para os meninos e para as meninas:

Meninos

Mediana=

Meninas

Mediana=

Moda
Valor da variável que ocorre com maior frequeência

Medidas de dispersão (variância, desvio-padrão, coeficiente de variação e percentis)

Valores mínimo e máximo: valores extremos da distribuição.

Amplitude de variação: é a diferença entre os 2 valores extremos da distribuição.

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

Populacional Parâmetro σ^2 estimador :

$$S^2_{(N)} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N} \quad ou$$

$$S^2_{(N-1)} = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1}$$

Desvio padrão: é a raiz quadrada da variância , ou seja

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

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

Coeficiente de Variação de Pearson (CV):

é o quociente entre o desvio padrão e a média, ou seja $\mathrm{CV} = \frac{\mathrm{S}}{\overline{\mathrm{X}}} x100$

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

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

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

Coeficiente de Variação de Pearson $CV = \dfrac{36{,}64}{245{,}05} x100 = 15\%$

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

Variância: $s^2 =$

Desvio padrão $s =$

Coeficiente de Variação de Pearson CV=

Exemplo 17

Com os dados do exemplo 15, calcule a variância, o desvio-padrão e o coeficiente de variação de Pearson.

Meninos

Meninas

Quartil

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

¼	½	¾	
25%	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%

$$Q_1 = x_{(\frac{1}{4}(n+1))} \quad \text{e} \quad Q_3 = x_{(\frac{3}{4}(n+1))}$$

onde x é o valor da variável e $(\frac{1}{4}(n+1))$ e $(\frac{3}{4}(n+1))$ 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; Gupta JM. Sodium bicabornate in idiopatic respiratory distress syndrome. *Arch. Diseases in Childhood*,1973:48, 249-255.

Entre os recém-nascidos que sobreviveram:

$$Q_1 = x_{(\frac{1}{4}(23+1))} = x_6 = 1720g \; ; \quad Q_3 = x_{(\frac{3}{4}(23+1))} = x_{18} = 2830g$$

$$Q_2 = x_{(\frac{1}{2}(23+1))} = x_{12} = 2200g$$

Entre os recém-nascidos que foram a óbito

$$Q_1 = x_{(\frac{1}{4}(27+1))} = x_7 = 1230g \; ; \quad Q_3 = x_{(\frac{3}{4}(27+1))} = x_{21} = 2200g$$

$$Q_2 = x_{(\frac{1}{2}(27+1))} = x_{14} = 1600g$$

Se o resultado for um valor fracionário:

Por exemplo, para n=22

$$Q_1 = x_{(\frac{1}{4}(22+1))} = x_{(\frac{23}{4})} = x_{(5\frac{3}{4})}$$

que é ¾ do caminho entre x_5=1715 e x_6=1720

$$Q_1 = 1715 + \frac{3}{4}(1720 - 1715) = 1718{,}8g$$

$$Q_3 = x_{(\frac{3}{4}(22+1))} = x_{(17\frac{1}{4})}$$

que é ¼ do caminho entre x_{17}=2700 e x_{18}=2830

$$Q_3 = 2700 + \frac{1}{4}(2830 - 2700) = 2732{,}5g$$

Decil
Valores da variável que dividem a distribuição em dez partes iguais.

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

Entre os recém-nascidos que sobreviveram

Percentil 5:

$$P_5 = x_{(\frac{5}{100}(23+1))} = x_{(\frac{120}{100})} = x_{(1\frac{1}{5})}$$

que é 1/5 do caminho entre x_1=1130 e x_2=1410

$$P_5 = 1130 + \frac{1}{5}(1410 - 1130) = 1186g$$

Percentil 10:

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

Percentil 50:

$$P_{50} = x_{(\frac{50}{100}(23+1))} = x_{(\frac{1200}{100})} = x_{(12)} \ ; \ P_{50} = 2200\,g$$

Percentil 75:

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

Percentil 90:

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

Percentil 95:

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

Box plot

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).

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: $264{,}5 + 1{,}5x45 = 332$.

181 é o valor adjacente inferior. É o menor valor da distribuição, igual ou acima de 152, onde 152 é dado por: $219{,}5 - 1{,}5x45 = 152$.

Tipo B
n=20

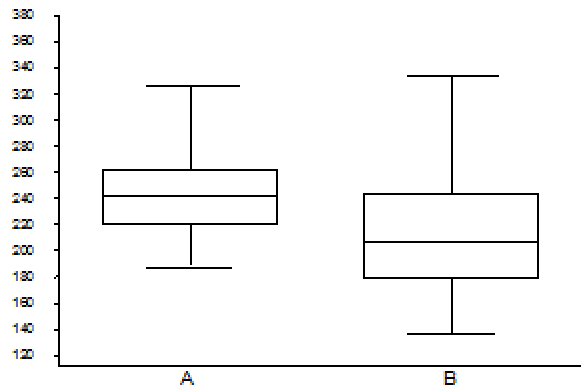
$$Q1 = x_{\frac{1}{4}(n+1)} = x_{\frac{21}{4}} = x_{5\frac{1}{4}} = 175 + \frac{1}{4}(183 - 175) = 175 + 2 = 177$$

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

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: $245 + 1{,}5x68 = 347$.

137 é o valor adjacente inferior. É o menor valor da distribuição, igual ou acima de 75, onde 75 é dado por: $177 - 1{,}5x68 = 75$.



Fonte: Fonte: Hand DJ et alli. *A handbook of small data sets*. Chapman & Hall, 1994.
Gráfico - Box plot da variável nível de colesterol segundo tipo de personalidade.

Exemplo 18
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 (µg/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 (μg/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 (μg/ml de leite) em cada grupo.
b) Calcule a quantidade mediana de cálcio (μg/ml de leite) em cada grupo.
c) Desenhe o *box plot* da concentração de cálcio (μg/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.

Exercícios suplementares
Exercício S11
Os dados a seguir são provenientes de um estudo que avalia o crescimento de crianças de 7 a 10 anos de uma escola pública do município de São Paulo no ano de 2008. Os dados apresentados são de 16 meninos e 16 meninas para os quais foram aferidos a circunferência do braço (CB) (cm):

Meninos									
18,3	19,3	20,9	19,0	20,5	16,3	21,0	17,8	21,6	22,6
27,3	26,7	29,0	22,0	25,2	19,5				
Meninas									
21,5	16,1	18,6	19,9	17,9	23,7	20,0	19,4	23,5	18,0
23,0	17,9	20,3	23,1	17,8	18,2				

a) Calcule a circunferência braquial (cm) média e mediana para cada sexo.
b) Calcule a variância, o desvio-padrão e o coeficiente de variação de Pearson da circunferência braquial (cm) para cada sexo.
c) Meninos e meninas são parecidos quanto a circunferência braquial (cm)?
d) E quanto à variabilidade?

Exercício S12

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.

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

Variável	Meninos (n=242)		Meninas (n=251)		
	Média	DP	Média	DP	p - valor
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)	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).[8]

Exercício S13

Os dados a seguir são provenientes de um estudo que avaliou o nível de colesterol sanguíneo (mg/dl) de 100 homens.

id	colesterol	id	colesterol	Id	colesterol	id	colesterol
1	134	26	189	51	216	76	239
2	147	27	189	52	217	77	239
3	157	28	190	53	217	78	240
4	161	29	190	54	218	79	240
5	162	30	192	55	218	80	240
6	164	31	194	56	219	81	243
7	165	32	195	57	219	82	246
8	166	33	196	58	219	83	248
9	171	34	198	59	221	84	251
10	173	35	199	60	221	85	255
11	176	36	199	61	223	86	255
12	176	37	199	62	223	87	256
13	178	38	201	63	224	88	259
14	179	39	203	64	225	89	261
15	179	40	204	65	228	90	267
16	180	41	205	66	230	91	268
17	181	42	206	67	230	92	272
18	181	43	209	68	231	93	279
19	183	44	210	69	231	94	286
20	184	45	211	70	231	95	287
21	185	46	211	71	232	96	289
22	186	47	212	72	234	97	290
23	186	48	213	73	234	98	296
24	186	49	215	74	238	99	298
25	187	50	216	75	238	100	382

a) Desenhe o *box plot* do colesterol (mg/dl).
b) Comente o gráfico *box plot* quanto a dispersão dos dados, existência de valores aberrantes e igualdade de medianas.

Tópicos iniciais de amostragem

População: totalidade de elementos sob estudo. Apresentam uma ou mais características em comum. Supor o estudo sobre a ocorrência de sobrepeso em crianças de 7 a 12 anos no Município de São Paulo.

População alvo – todas as crianças nesta faixa etária deste município.
População de estudo – crianças matriculadas em escolas.

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 (μ = 3100 g);

Proporção de crianças de 7 a 12 anos classificadas como obesas, no município de São Paulo (π = 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: $\overline{X} = \frac{\sum_{i=1}^{N} X_i}{N}$,

onde $\sum_{i=1}^{N} X_i = X_1 + X_2 + ... + X_N$ 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 = $\overline{x} = 3000g$.

Indicações para utilizar uma amostra

População muito grande
Processo destrutivo de investigação
Novas terapias

Vantagens de realizar um estudo com amostragem:

Menor custo
Menor tempo para obtenção dos resultados
Possibilidade de objetivos mais amplos
Dados possivelmente mais fidedignos

Desvantagens

Resultados sujeitos à variabilidade

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 (NN, 1998). Tamanho da população: N; tamanho da amostra: n; fração global de amostragem ou probabilidade de sortear um 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.

Exemplo 19 - Os dados a seguir são de peso (kg) de 80 mulheres identificadas pela variável id (i-dentificação).

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).

a) Sorteie uma amostra aleatória de tamanho 20 utilizando a tabela dos números equiprováveis.
b) Apresente os valores do peso dos indivíduos sorteados.
c) Some os valores e divida pelo tamanho da amostra (número de valores).
d) Este valor é o parâmetro, o estimador ou a estimativa do peso médio?

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, ..., 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: $1 \le i \le 8$

Começo casual **sorteado**: i=4

Amostra composta dos elementos:

i 4
i+k 12
i+2k 20
i+3k 28
i+4k 36
i+5k 44
i+6k 52
i+7k 60
i+8k 68
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:

		elemento
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

Exemplo 20 – Utilize os dados do Exemplo 19.

a) 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.
b) Some os valores e divida pelo tamanho da amostra (número de valores).
c) Compare com o peso médio obtido no exemplo 2. Você esperaria o mesmo resultado? Justifique.
d) Qual dos dois valores você diria que representa melhor o conjunto de dados? Justifique.

<u>Amostragem casual simples estratificada com partilha proporcional</u>

A população possui estratos com tamanhos:

N_1; N_2; N_3, onde a soma dos estratos é o tamanho da população, ou seja $\sum N_i = N$

A amostra deve conter os elementos da população nas mesmas proporções dos estratos. Tem-se que os tamanhos dos estratos amostrais são n_1, n_2 e n_3 tal que $\sum n_i = n$

Aplicando-se a proporção:

$$\frac{n_i}{n} = \frac{N_i}{N} \Rightarrow n_i = n\frac{N_i}{N}$$

Exemplo:

N=500; N_1=50; N_2=150; N_3=300 e n=40

Estrato i	Tamanho do estrato na população N_i	Tamanho do estrato na amostra n_i	$\frac{n_i}{n} = \frac{N_i}{N}$
1	50	4	0,1
2	150	12	0,3
3	300	24	0,6
Total	500	40	

$$n_1 = 40\frac{50}{500} = 4\ ;\ n_2 = 40\frac{150}{500} = 12\ ;\ n_3 = 40\frac{300}{500} = 24$$

Noções de probabilidade e distribuição Bernoulli e distribuição binomial

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.

<u>Probabilidade em espaços finitos contáveis</u>

Espaço amostral (S)

- É o conjunto de todos os resultados possíveis de um experimento.

- Supor o experimento lançar uma moeda: S= {cara, coroa}

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

<u>Definição clássica de probabilidade</u>

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

Exemplo: probabilidade de (ouros) = $\frac{13}{52} = \frac{1}{4}$

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 = {carta com naipe vermelho}; B={carta com naipe preto}, na retirada de uma carta de baralho.

Exemplo de eventos não mutuamente exclusivos
A= {naipe vermelho} ; B = {ás} .

- A probabilidade da ocorrência de um evento A ou de um evento B é:
P(A ou B) = $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Exemplo: P(naipe vermelho ou ás) = P(naipe vermelho) + P(ás) – P(naipe vermelho e ás) = (26/52) + (4/52) – (2/52) = 28/52 = 0,538.

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

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

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

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

P(2 ou 3)= P(2)+P(3)= 1/6 + 1/6 = 2/6 = 1/3.

P(Resultado ímpar)= P(1 ou 3 ou 5)= P(1)+P(3)+P(5)= 3/6 = 1/2.

Regra da adição: P(A ou B) = P(A u 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 e B)= P(A $\cap$ B) = P(A) x P(B)

- P(face 2 no primeiro dado e face 3 no segundo dado), no lançamento sequencial de dois dados = P(2 e 3) = P(2)xP(3)= 1/6 x 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:
Probabilidade de rei dado que ocorreu figura:
P(r|figura)= P(r e figura)/P(figura)= 4/52 ÷ 12/52= 4/12= 1/3

- Probabilidade de rei, dado que ocorreu copas:
  P(r|♥)= P(r e ♥)/P(♥)= 1/52÷13/52= 1/13

Regra da multiplicação

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

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

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 $\bar{S}$	Fumante S	
Não ($\bar{R}$)	0,5 = P($\bar{S}\bar{R}$)	0,30 = P($\bar{R}S$)	0,80 = P($\bar{R}$)
Sim (R)	0,05 = P($\bar{S}R$)	0,15 = P(*SR*)	0,20 = P(R)
	0,55 = P($\bar{S}$)	0,45 = P(*S*)	

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

$P(R \mid S) = \dfrac{P(R \cap S)}{P(S)}$ = 0,15/0,45 = 0,33

Os eventos não são independentes porque $P(\overline{SR}) \neq P(\bar{S})xP(\bar{R})$

Relação entre eventos mutuamente exclusivos e independentes:

Os eventos mutuamente exclusivos A e B satisfazem a condição que P(A 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 frequentista de probabilidade:

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

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

Quando n cresce, $\dfrac{m}{n}$ 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 e existe associado a este número uma probabilidade de ocorrência.

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(cara)= 0,5= 1/2.

No lançamento de 10 moedas, X:0, 1, 2,....,10; e a probabilidade de cara = 0,5.
Sair cara é mutuamente exclusivo de sair coroa e um particular resultado de cada lançamento independe dos demais.

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.

Distribuição de probabilidades

Modelo de probabilidade Bernoulli

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

Exemplo:
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.

Define-se como sucesso sair cara.

Define-se a variável aleatória X que assume valor 1 se ocorrer sucesso e 0 se ocorrer fracasso. X: 0,1

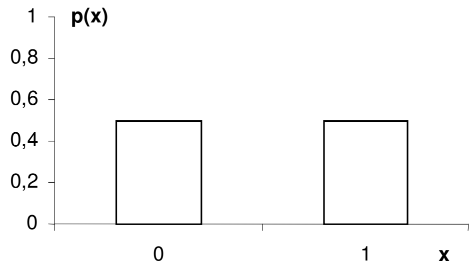
Parâmetro: probabilidade da variável assumir valor 1.
Notação: π ou p.

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:

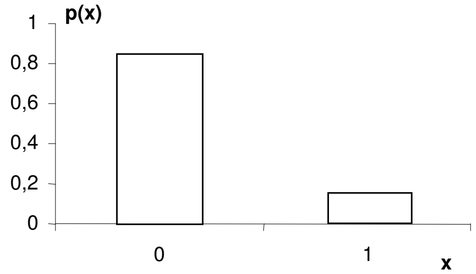


p=0,5

Exemplo:

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



p=0,15

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:

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

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

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

$$\sqrt{(-p)^2.(1-p) + (1-p)^2 p} = \sqrt{p(1-p)[p+(1-p)]} = \sqrt{pq}$$

Resumindo,

Modelo de probabilidade Bernoulli Uma variável aleatória discreta X que pode assumir valores 0 e 1, com função de probabilidade dada por $p(x) = p^x (1-p)^{1-x}$ 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~Bernoulli(p) com média p e desvio padrão $\sqrt{p(1-p)}$. **O símbolo ~ lê-se "tem distribuição" ou se "distribui segundo".**

Distribuição binomial: 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 $\begin{cases} \text{Cara (K)} \rightarrow \text{probabilidade(K)} = \text{p} \\ \text{Coroa (C)} \rightarrow \text{probabilidade (C)} = \text{q} \\ \text{p} + \text{q} = 1 \Rightarrow \text{q} = 1 - \text{p} \end{cases}$

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)→ X: 0,1 X~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) → X: 0,1,2 X~B(n=2, p)

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

A moeda é lançada três vezes (n=3) → X: 0,1,2,3 X~B(n=3, p)

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

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

X~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 = $\sqrt{npq}$

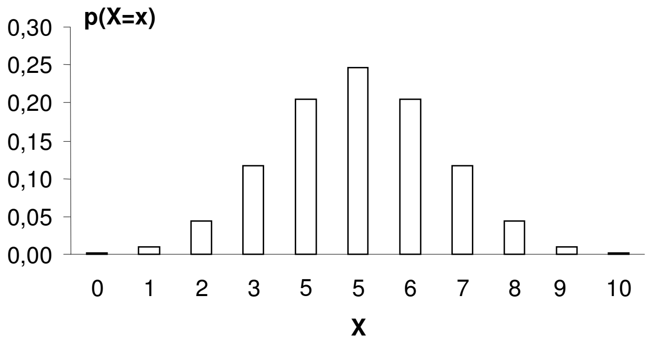
Exemplo
Lançamento de moedas.

- 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
	1



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

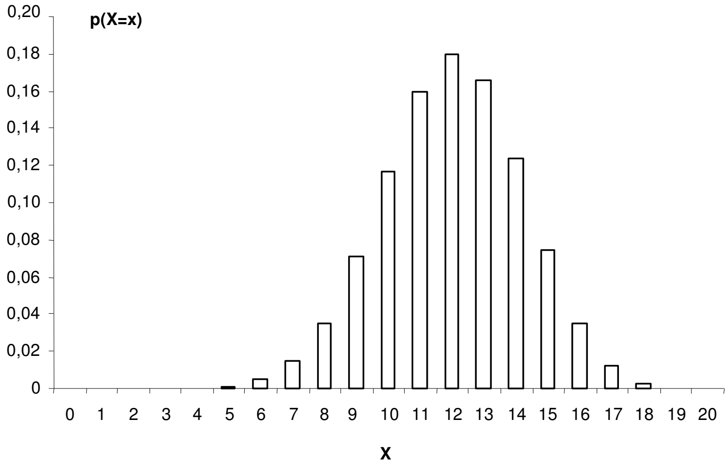
Variância = $n\frac{p}{n}x\frac{q}{n} = \frac{pq}{n}$ = 0,025

Desvio padrão = $\frac{\sqrt{npq}}{n} = \sqrt{\frac{npq}{n^2}} = \sqrt{\frac{pq}{n}}$ = 0,158

Exemplo 21

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 π= 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



Calcule a média, a variância e o desvio-padrão.

Exemplo 22
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) = $\binom{3}{0}(0,3)^0(0,7)^3 = \frac{3!}{0!(3-0)!}(0,7)^3 = 1x0,343 = 0,343$ = 34,3%

b) Um rato ficar infectado?

P(X=1) = $\binom{3}{1}(0,3)^1(0,7)^{3-1} = \frac{3!}{1!(3-1)!}(0,3)^1(0,7)^{3-1} = \frac{3x2x1}{1x2x1}0,3x0,49 = 0,441$ = 44,1%

c) Dois ratos ficarem infectado?

P(X=2) = $\binom{3}{2}(0,3)^2(0,7)^{3-2} = \frac{3!}{2!(3-2)!}(0,3)^2(0,7)^{3-2} = \frac{3x2x1}{2x1x1}0,09x0,7 = 0,189$ = 18,9%

d) Todos os ratos ficarem infectados?

P(X=3) = $\binom{3}{3}(0,3)^3(0,7)^{3-3} = \frac{3!}{3!(3-3)!}(0,3)^3(0,7)^0 = \frac{3x2x1}{3x2x1x1}0,027x1 = 0,027$ = 2,7%

e) Pelo menos 2 fiquem infectados?

f) No máximo 1 fique infectado?

Exemplo 23
Uma indústria de alimentos está realizando testes com um bolo que será comercializado. Durante a prova do bolo, 20% das pessoas selecionadas para tal tarefa acharam o sabor muito doce. Supondo que 5 pessoas provarão o bolo novamente, qual a probabilidade de:

a) Nenhuma pessoa achar o bolo muito doce?
b) Todos acharem o bolo muito doce ?
c) Pelo menos 4 pessoas acharem o bolo muito doce?
d) No máximo 2 pessoas acharem o bolo muito doce?

EXERCÍCIOS SUPLEMENTARES

Exercício S14
Supor um teste com questões com 5 respostas de múltipla escolha com somente uma alternativa correta.
a) Se o aluno escolher uma ao acaso ("chute"), qual a probabilidade dele escolher a resposta certa?
b) Supondo que o teste tenha 20 questões; definindo-se a variável aleatória T: número de questões certas, qual é a distribuição de probabilidade da variável T?
c) Calcular a probabilidade de um aluno acertar, no chute, 3 questões.
d) Se o escore mínimo para passar é 10, qual a probabilidade de um aluno passar no teste, somente chutando?

e) Qual o número médio de acertos esperado se o aluno somente chutar as respostas?

Exercício S15
Certa doença tem letalidade de 70%. Supondo-se que existam 20 pacientes com esta doença, calcular:

a) a probabilidade de que todos morram da doença.
b) a probabilidade de que nenhum paciente morra da doença.
c) a probabilidade de que 7 pacientes morram da doença.
d) a probabilidade de que, no máximo, 10 pacientes morram da doença.
e) a probabilidade de que, no mínimo, 5 pacientes sobrevivam.
f) o número esperado de óbitos e o respectivo desvio padrão.

Exercício S16
Em uma grande população, 20% das pessoas são canhotas. Assumindo que a variável X: número de pessoas canhotas segue uma distribuição Binomial, e sorteando-se uma amostra aleatória de 10 pessoas, encontre a probabilidade de:

a) encontrar 2 pessoas canhotas .
b) encontrar pelo menos 2 pessoas canhotas.
c) encontrar no máximo 1 pessoa canhota.
d) encontrar de 1 a 4 pessoas canhotas.

Exercício S17
Um caso de esquistossomíase é identificado pela detecção de ovo de xistossoma em amostra de fezes. Em pacientes com infecção baixa, uma técnica de exame de fezes tem probabilidade de 0,4 de detectar ovo. Se 5 amostras são examinadas para cada paciente, qual a probabilidade de um paciente com baixa infecção não ser identificado?

Exercício S18
Supor que 20% de certa população tem sangue tipo B. Para uma amostra de tamanho 18, retirada desta população, calcule a probabilidade de que sejam encontradas:

a) 3 pessoas com sangue tipo B.
b) 3 ou mais pessoas com sangue tipo B.
c) no máximo 3 pessoas com sangue tipo B.

Exercício S19
A probabilidade que uma pessoa que sofre de enxaqueca obter alívio utilizando certo medicamento é de 0,9. São selecionados 5 pacientes que sofrem de enxaqueca e recebem o medicamento. Quanto ao número de pessoas que vai ter alívio, encontre a probabilidade de:

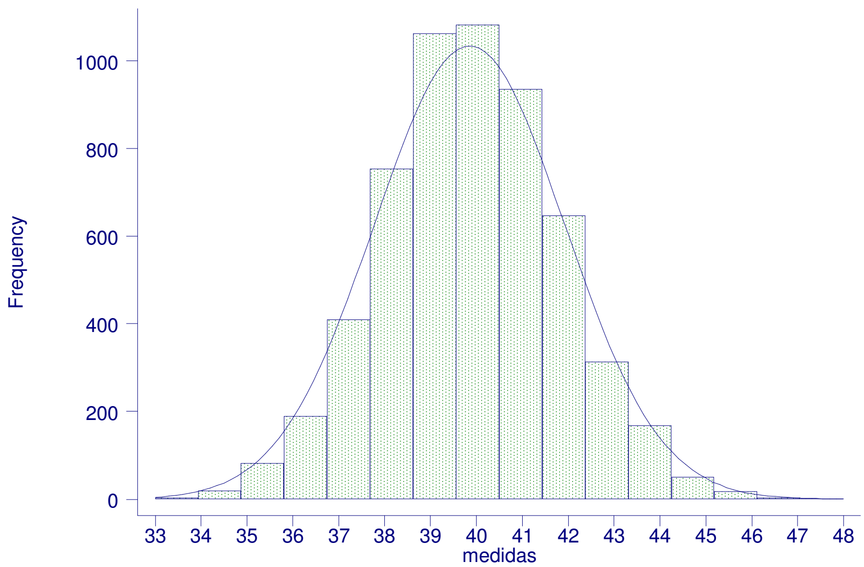
a) nenhuma pessoa ter alívio.
b) mais do que uma pessoa tenha alívio.
c) três ou mais pessoas tenha alívio.
d) no máximo duas pessoas tenham alívio.

Distribuição normal ou de Gauss; distribuição amostral da média

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 μ e desvio padrão σ (variância σ^2), a função densidade de probabilidade de X é dada por $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{[-\frac{(x-\mu)^2}{2\sigma^2}]}$, $-\infty < x < +\infty$;

onde

π : constante $\cong$ 3,1416; e: constante $\cong$ 2,718

μ : constante (média aritmética da população)

σ : constante (desvio padrão populacional)

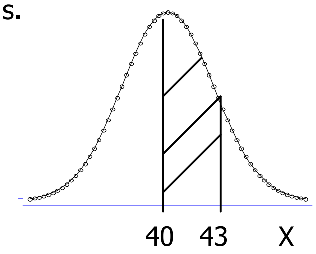
Propriedades:

- campo de variação : $-\infty < X < +\infty$;
- é simétrica em torno da média m (ou μ);
- 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.

$$\begin{cases} \mu \pm 1\sigma \ \ inclui \ \ 68{,}2\% \ das \ observações \\ \mu \pm 1{,}96\sigma \ inclui \ \ 95{,}0\% \ das \ observações \\ \mu \pm 2{,}58\sigma \ \ inclui \ \ 99{,}0\% \ das \ observações \end{cases}$$

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 (μ) = 40 polegadas e desvio padrão (σ) = 2 polegadas.



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

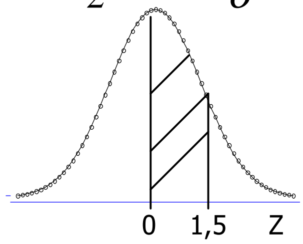
$$P(40 \leq X \leq 43) = \int_{40}^{43} \frac{1}{2\sqrt{2\pi}} e^{[-\frac{(x-40)^2}{2x4}]} dx$$

Quantos desvio padrão 43 está em torno da média?

Normal reduzida:

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

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



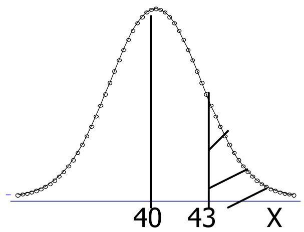
Utilizando a tabela da curva normal reduzida,

$P(0 \le Z \le 1{,}5)$ = 0,43319 =43,3%

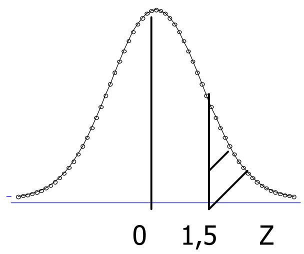
Exemplo 24

Com base na distribuição de X~N(μ =40, σ =2), calcular:

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



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



Utilizando a tabela da curva normal reduzida,

$P(Z \ge 1{,}5)$ = 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 seria ultrapassado por 25% da população?

Exemplo 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:

a) altura entre 160 cm e 165 cm?
b) altura menor do que 145 cm?
c) Altura maior do que 170 cm?

Distribuição amostral da média
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	Xi
A	11
B	16
C	12
D	15
E	16
F	14

Fonte: Dixon WJ e Massey FJ. Introduction to Statistical Analysis. 2nd edit. The Maple Press Company, York, 1957.

Média populacional (μ) = 14;

Variância populacional (σ^2) = 3,667;

Desvio padrão populacional (σ) = 1,9149.

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

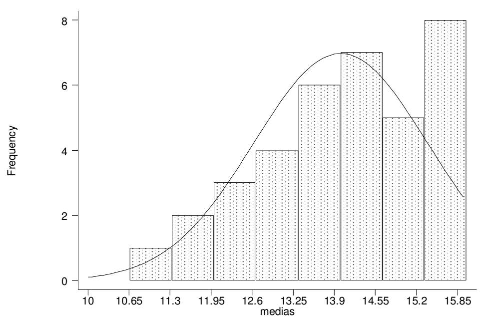
Todas as possíveis amostras de tamanho 2, determinadas pelo processo de amostragem aleatório, com reposição (N=6, n=2):

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

Distribuição de frequência de todas as possíveis médias:

Distribuição amostral da média

i	$\bar{x}_i$	frequência
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



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

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

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

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

Teorema central do limite: X é variável aleatória com média μ e variância σ^2, então

$$\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

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

Exemplo 26

Os valores de ácido úrico em homens adultos sadios seguem distribuição aproximadamente Normal com média 5,7mg% e desvio padrão 1mg%. Encontre a probabilidade de que uma amostra aleatória de tamanho 9, sorteada desta população, tenha média

a) maior do que 6 mg%.
b) menor do que 5,2 mg%.

X~N(μ =5,7; $\sigma = 1$)

a) $P(\overline{X} \geq 6) = P(Z_{\overline{X}} \geq \dfrac{6-5{,}7}{\frac{1}{\sqrt{9}}}) = P(Z_{\overline{X}} \geq 0{,}91) = 0{,}5 - 0{,}31859 = 0{,}18141$.

b) $P(\overline{X} \leq 5{,}2) = P(Z_{\overline{X}} \leq \dfrac{5{,}2-5{,}7}{\frac{1}{\sqrt{9}}}) = P(Z_{\overline{X}} \leq -1{,}52) = 0{,}5 - 0{,}43574 = 0{,}064$.

Exemplo 27

Suponha que o peso em gramas do conteúdo de pacotes de salgadinho siga uma distribuição normal com média 500g e desvio padrão 85g. Sorteia-se uma amostra de 50 pacotes. Calcule:

a) a probabilidade de obter peso médio entre 500 e 530 gramas.
b) a probabilidade de obter peso médio entre 450 e 500 gramas.

Gabarito

Exemplo 1 (pag 2)

Variável	Tipo (natureza)
Condição de saúde (doente, não doente)	Qualitativa nominal
Tipo de parto (normal, cesário)	Qualitativa nominal
Nível de colesterol sérico (mg/100cc)	Quantitativa contínua
Tempo de um procedimento cirúrgico (minutos)	Quantitativa contínua
Número de praias consideradas poluídas	Quantitativa discreta
Custo do procedimento (reais)	Quantitativa contínua
Peso (g)	Quantitativa contínua
Estado nutricional (desnutrição, eutrofia, sobrepeso, obesidade)	Qualitativa ordinal
Consumo de energia (Kcal)	Quantitativa contínua
Realização da refeição café da manhã (sim/não)	Qualitativa nominal
Número de escolares por série	Quantitativa discreta
Realização de atividade física diária (sim/não)	Qualitativa nominal
Tempo assistindo TV/dia (< 2h, 2 a 4h, >4h)	Quantitativa contínua
Porções consumidas por grupo de alimentos	Quantitativa discreta
Percentual de gordura corporal (%)	Quantitativa contínua

Exemplo 2 (pag 6)

a) Distribuição de mulheres idosas segundo a altura, Local X. Ano Y.

Altura (cm)	N^o	%
140|--145	1	0,3
145|--150	11	3,1
150|--155	52	14,8
155|--160	109	31,1
160|--165	106	30,2
165|--170	50	14,3
170|--175	18	5,1
175|--180	4	1,1
Total	351	100

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

b) Pode-se observar na tabela que aproximadamente 60% das mulheres idosas têm a estatura entre 155cm e 164,9cm.

c e d) Representaria com um número que indicasse valor desconhecido por exemplo, 999. Na tabela todos os valores 999 seriam contabilizados na categoria ignorados. 15 valores ignorados corresponderiam a 4,3% do total de valores. Uma possibilidade é citar, na tabela, como nota, que foram excluídos 15 (4,3%) valores ignorados. Também é possível apresentar os valores ignorados na tabela, o que revelaria a qualidade da informação.

e) Os valores "esquisitos" ou seja que se destacam dos demais, se não forem valores errados, fazem parte da distribuição não sendo a melhor conduta simplesmente excluí-los. A decisão do investigador que resolveu excluí-los pode não ser a melhor.

Exemplo 3 (pag 7)

a) Calculando-se o percentual "fixando" o hábito de fumar e investigando a distribuição dos níveis de β-caroteno entre fumantes e não fumantes; em outras palavras, comparando-se fumantes e não fumantes quanto aos níveis de β-caroteno.

Distribuição de gestantes segundo níveis de β-caroteno (mg/L) e hábito de fumar.

β-caroteno (mg/L)	Fumante		Não fumante		Total	
	n	%	n	%	n	%
Baixo (0 – 0,213)	46	79,3	74	56,1	120	63,2
Normal (0,214 – 1,00)	12	20,7	58	43,9	70	36,8
Total	58	100	132	100	190	100

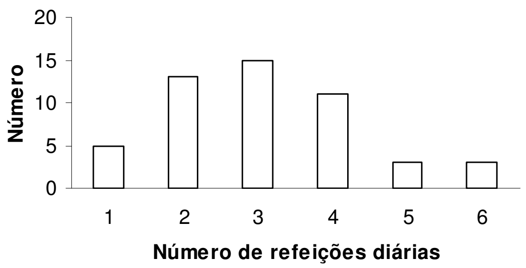
b) Calculando-se o percentual "fixando" o nível de β-caroteno e investigando a distribuição do hábito de fumar entre gestantes com nível baixo e normal de β-caroteno.

Distribuição de gestantes segundo níveis de β-caroteno (mg/L) e hábito de fumar.

β-caroteno (mg/L)	Fumante		Não fumante		Total	
	n	%	n	%	n	%
Baixo (0 – 0,213)	46	38,3	74	61,7	120	100
Normal (0,214 – 1,00)	12	17,1	58	82,9	70	100
Total	58	30,5	132	69,5	190	100

c)
interpretando-se a tabela do item **a**:
Do total de fumantes, 79,3% apresentam nível baixo de β-caroteno. Entre não fumantes este percentual é de 56,1%. Parece existir associação; a proporção de pessoas com nível baixo de β-caroteno parece maior entre fumantes.
interpretando-se a tabela do item **b**:
Entre as gestantes com nível baixo de β-caroteno, 38,3% são fumantes enquanto que entre as mulheres com nível normal de β-caroteno, este percentual era de 17,1%. Pode ser que exista associação; a proporção de fumantes parece maior entre gestantes com nível baixo de β-caroteno.

Exemplo 4 (pag 13)

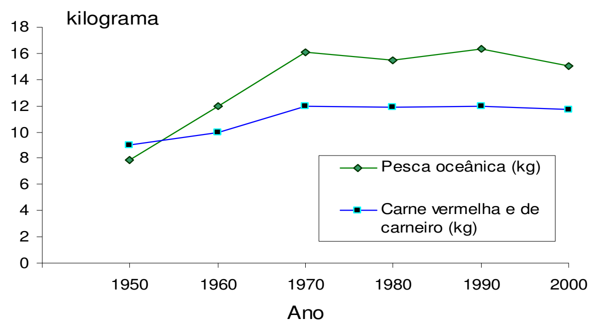


Fonte: X.
Distribuição de indivíduos segundo o número de refeições diárias. Local X, Ano Y.

Pode-se observar no gráfico que aproximadamente (78%) dos indivíduos realizam de 2 a 4 refeições por dia.

Exemplo 5 (pag 14)

a)

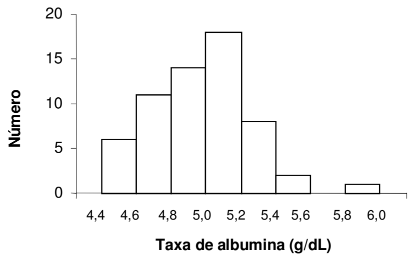


Fonte: *State of the World*, 2001. *The Worldwatch Institute*.
Distribuição da produção mundial de pesca oceânica e de carne vermelha e de carneiro por pessoa, 1950 – 2000.

b) Pode-se observar que tanto a distribuição de pesca oceânica como a de carne vermelha e de carneiro aumentaram entre os anos de 1950 a 1970, especialmente a pesca oceânica que ultrapassou a de carne vermelha e carneiro. Após este período, as duas produções apresentam estabilização com leve declínio.

Exemplo 6 (pag 15)

a)

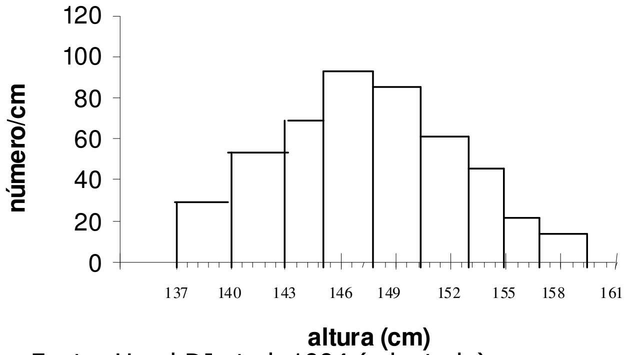


Fonte: Soares JF, Siqueira AL. COOPMED, 2002.
Distribuição de pacientes segundo taxa de albumina no sangue (g/dL). Local X, Ano Y.

b) O gráfico indica a presença de um maior número de pacientes com taxa de albumina no sangue entre 5,0 a 5,2 g/dl.

Exemplo 7 (pag 17)
a)

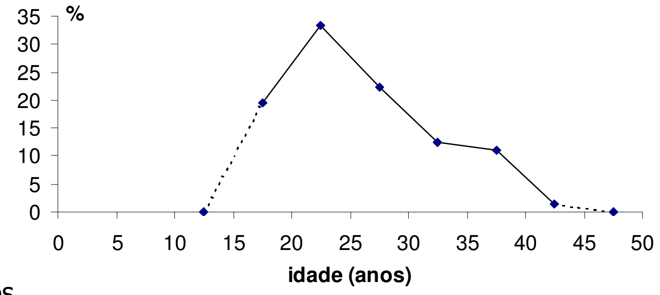
Altura (cm)	Número	Ponto médio	Amplitude	Número/Amplitude
137,0 |--140,0	71	138,5	3	23,7
140,0 |--143,0	137	141,5	3	45,7
143,0 |--145,0	154	144	2	77
145,0 |--147,0	199	146	2	99,5
147,0 |--150,0	279	148,5	3	93
150,0 |--153,0	221	151,5	3	73,7
153,0 |--155,0	94	154,0	2	47
155,0 |--157,0	51	156,0	2	25,5
157,0 |--160,0	37	158,5	3	12,3
Total	1243			



Fonte: Hand DJ et al, 1994 (adaptado).
Distribuição de mulheres segundo altura. Bangladesh, Índia. Ano Y.

b) Por meio do gráfico pode-se dizer que existe concentração de mulheres nos valores de altura de entre 145,0 e 150,0 cm.

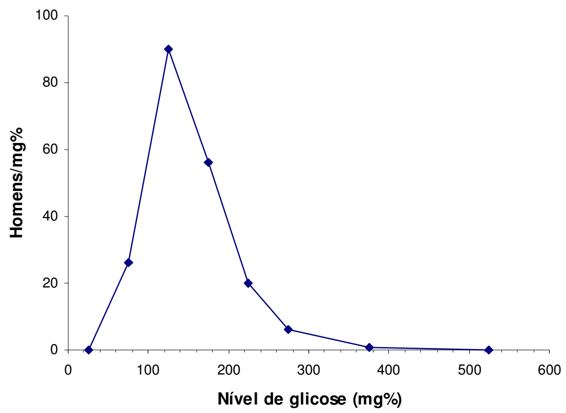
Exemplo 8 (pag 18)
a)



Fonte: Dados Hipotéticos.
Distribuição de usuárias do Serviço de Saúde X segundo idade (anos). Município de São Paulo, 2009.

b) Segundo o gráfico pode-se observar um maior número de usuárias com idade entre 20 e 25 anos.

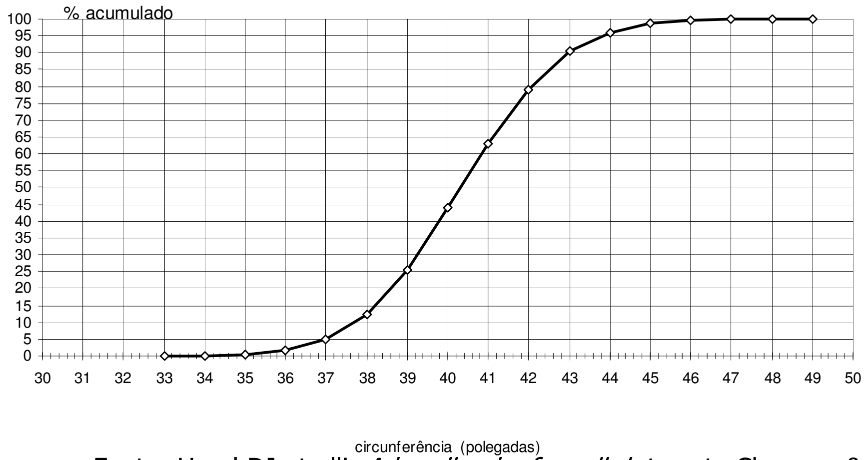
Exemplo 9 (pag 19)



Fonte: X.

Distribuição de homens segundo nível de glicose no sangue (mg%). Local X. Ano Y.
b) O gráfico indica que há uma maior concentração de homens com nível de glicose sanguínea de aproximadamente 150 mg%.

Exemplo 10 (pag 20)
a)



Fonte: Hand DJ et alli. *A handbook of small data sets*. Chapman&Hall, 1994.
Distribuição de soldados escoceses segundo circunferência do tórax (polegadas). Local X. 1976-1874.

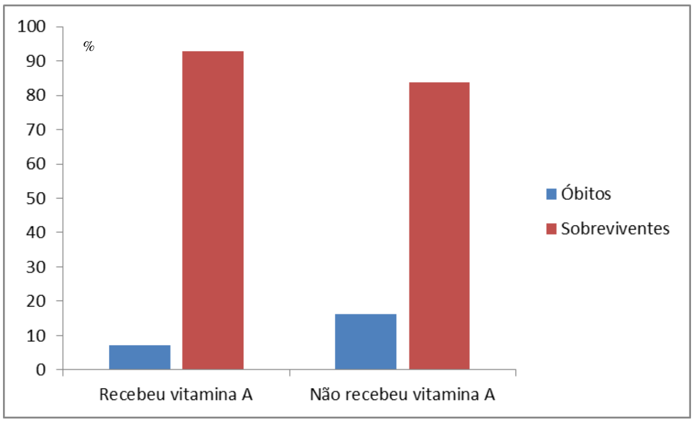
b) 39 polegadas.
c) 40,4 polegadas.
d) 35%.
e) 95% dos soldados têm circunferência do tórax até 44 polegadas.

Exemplo 11 –(pag 22)

Distribuição de crianças segundo condição de sobrevivência e suplementação. Indonésia, 1986

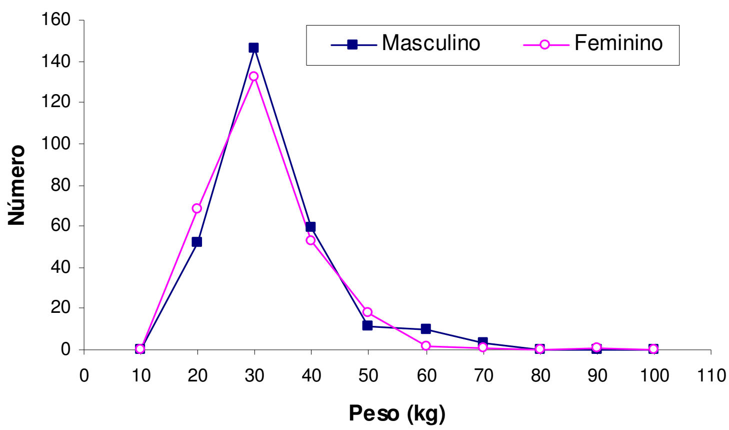
Grupo de tratamento	Condição de sobrevivência					
	Óbitos		Sobreviventes		total	
	n	%	n	%	n	%
Recebeu vitamina A	50	7,2	650	92,8	700	100
Não recebeu vitamina A	100	16,3	515	83,7	615	100
TOTAL	150	11,4	1165	88,6	1315	100

Fonte: Moore Ds & McCabe G. Introdução à Prática da Estatística (adaptado)



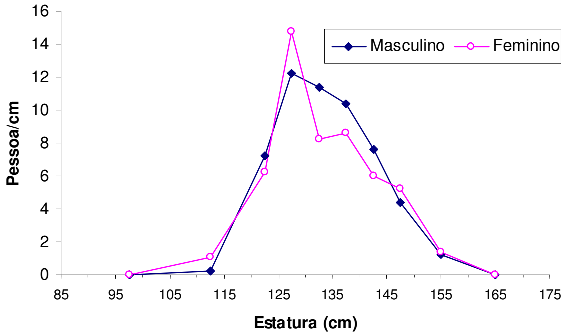
Fonte: Moore Ds & McCabe G. Introdução à Prática da Estatística (adaptado)
Distribuição de crianças segundo condição de sobrevivência e suplementação. Indonésia, 1986

Exemplo 12 - (pag 24)



Fonte: Koga C. 2005
Distribuição de escolares de 7 a 10 anos segundo peso (kg) e sexo. Duas escolas do Município de São Paulo, 2005.

Exemplo 13 – (pag 25)

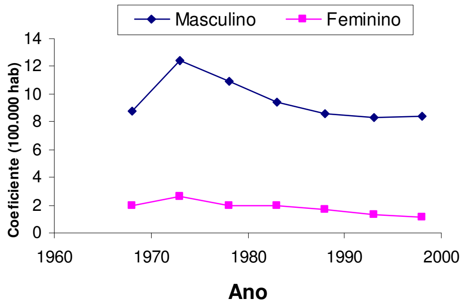


Fonte: Koga C. 2005
Distribuição de escolares de 7 a 10 anos segundo estatura (cm) e sexo. Duas escolas do Município de São Paulo, 2005.

Exemplo 14 (pag 27)

a)

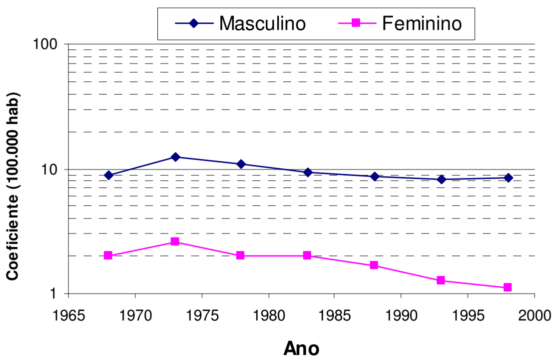
Coeficientes de mortalidade por câncer de esôfago (por 100.000 hab.) segundo sexo. Município de São Paulo, 1968-1998.



Fonte: Incidência de câncer no Município de São Paulo, 1997-1998. Registro de Câncer de São Paulo. FSP/USP.

b)

Coeficientes de mortalidade por câncer de esôfago (por 100.000 hab.) segundo sexo. Município de São Paulo, 1968-1998.



Fonte: Incidência de câncer no Município de São Paulo, 1997-1998. Registro de Câncer de São Paulo. FSP/USP.

c) Se os dados forem representados em um único gráfico de escala aritmética, pode-se dizer que existe tendência de decréscimo mais acentuada entre homens e que entre mulheres a tendência é estacionária.

Se for utilizada a escala logarítmica, representando-se sexo masculino e feminino em um único gráfico observa-se que em ambos os sexos existe tendência de decréscimo, sendo mais rápido entre mulheres. A escala logarítmica parece ser mais apropriada para representar as duas curvas em um único gráfico.

Exemplo 15 (pag 30)

Meninos	Meninas
1976	2002
3234	2964
1405	2203
1410	1478
1782	1151
2167	1083
1917	1362
2622	1392
1824	1637
3912	1628
1412	**Média:1690 kcal**
1635	
2230	
1241	
1866	
Média: 2042,2 kcal	

Exemplo 16 (pag 31)

Meninos	Meninas
1241	1083
1405	1151
1410	1362
1412	1392
1635	1478
1782	1628
1824	1637
1866	2002
1917	2203
1976	2964
2167	
2230	
2622	
3234	
3912	

Mediana <u>meninos</u>: posto = $\frac{15+1}{2}=8$. O valor da mediana é 1866 kcal.

Mediana <u>meninas</u>: posto = $\frac{10}{2}=5$ e $\frac{10+2}{2}=6$. O valor da mediana é $\frac{1478+1628}{2}=1553kcal$.

Exemplo 17 (pag 33)

Meninos

Valores	**(Valor-Média)²**
1241	641921,44
1405	406023,84
1410	399676,84
1412	397152,04
1635	165811,84
1782	67704,04
1824	47611,24
1866	31046,44
1917	15675,04
1976	4382,44
2167	15575,04
2230	35268,84
2622	336168,04
3234	1420387,24
3912	3496152,04
	7480556,4

7480556/14=**534325,45Kcal²**

Desvio Padrão (raiz quadrada da variância 534325,45Kcal)
DP: s = 730,98Kcal

CV de Pearson = (DP/Média) x 100)
CV: 35.80%

Meninas

Valores	**Variância**
2002	97344
2964	1623076
2203	263169
1478	44944
1151	290521
1083	368449
1362	107584
1392	88804
1637	2809
1628	3844
	2890544

2890544/9=**321171,5 kcal²**

DP = 566,7Kcal

CV de Pearson = 33,50%

Exemplo 18 (pag 37)

quantidade média de Cálcio (μg/mL de leite)

Grupo colostro: $\bar{x} = \frac{\sum_{i=1}^{26} x_i}{n} = \frac{7055}{26} = 271{,}35 \mu g / mL$;

Grupo maduro: $\bar{x} = \frac{\sum_{i=1}^{26} x_i}{n} = \frac{7310}{29} = 252{,}07 \mu g / mL$

quantidade mediana de Cálcio (μg/mL de leite)

Grupo colostro: n=26 (par) Mediana é a media dos valores que ocupam os postos 13 e 14.

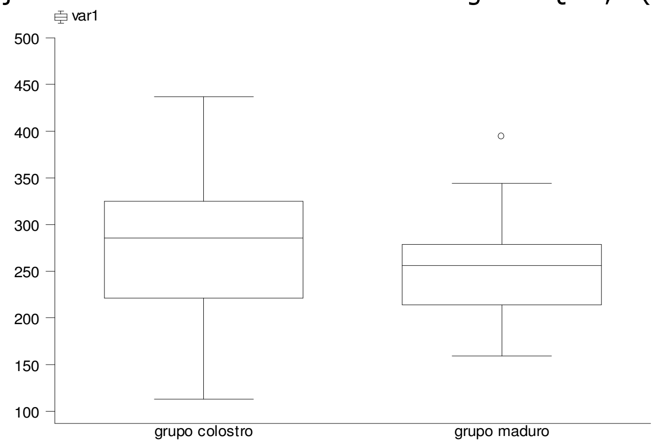
$$Med = \frac{275 + 296}{2} = 285{,}5 \mu g / mL \text{ ;}$$

Grupo maduro: n=29 (ímpar); a mediana é o valor da variável que ocupa o posto 15. Med= 256 μg/mL

Box plot da concentração de Cálcio (μg/mL de leite)

Medida	Grupo colostro	Grupo maduro
Q1	211	213,5
Q2	285,5	256
Q3	327,25	280
Valor adjacente inferior	113	159
Valor adjacente superior	437	344

Valor adjacente superior: maior valor abaixo ou igual a Q3+1,5x(IIQ)
Valor adjacente inferior: Menor valor acima ou igual a Q1-1,5x(IIQ)



Fonte: x.

"Box plot" da variável concentração de cálcio (μg/mL) segundo grupo de leite (colostro e maduro). Hospital Maternidade Odete Valadares, Belo Horizonte-MG, 1984 a 1985.

Valor aberrante apenas no grupo maduro.

Exemplo 19 – não tem gabarito porque são muitas possibilidades de sorteio

Exemplo 20 (pag 42)

a) Tamanho da população: N=80; tamanho da amostra: n=20; intervalo de amostragem: $k = \frac{80}{20} = 4$; começo aleatório (i): sorteia-se um número inteiro i no intervalo: $1 \leq i \leq 4$

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

b) Se o começo for 1, a média de peso será igual a 73,8 kg;
Se o começo for 2, a média de peso será igual a 73,9 kg;
Se o começo for 3, a média de peso será igual a 70,6 kg;
Se o começo for 4, a média de peso será igual a 71,4 kg.
c) e d) Não exatamente o mesmo. Esperaria valores semelhantes uma vez que as medidas de resumo dos dados são calculadas com base em amostras aleatórias da população. Assim, os dois valores poderiam representar o conjunto de dados.

Exemplo 21 (pag 50)

Média: 20 x 0,6 = 12 mães
Variância = 20 x 0,6 x 0,4 = 4,8 mães.
DP = 2,9 mães

Exemplo 22 (pag 51) feito em sala

Exemplo 23 (pag 51)

a) Nenhuma pessoa achar o bolo muito doce?
P (X=0)=32,8%
b) Todos acharem o bolo muito doce ?
P (X=5) $\cong$ 0
c) Pelo menos 4 pessoas acharem o bolo muito doce?
P (X ≥ 4) $\cong$ 6%
d) No máximo 2 acharem o bolo muito doce?
P (x ≤ 2) = 0,328 + 0,410 + 0,205 = 94,3%

Exemplo 24 (pag 55)

b) $P(35 < X < 40) = P(\frac{35-40}{2} < \frac{X-\mu}{\sigma} < \frac{40-40}{2}) = P(-2{,}5 < Z < 0)$

Utilizando a tabela da curva normal reduzida: P(-2,5<Z<0)=0,49379 = 49,4%

c) $P(X < 35) = P(\frac{X-\mu}{\sigma} < \frac{35-40}{2}) = P(Z < -2{,}5)$

Utilizando a tabela da curva normal reduzida: P(Z<-2,5)=0,5-0,49379=0,0062 ou 0,6%

d) 0,25 – olhar tabela da normal

Z = 0,675

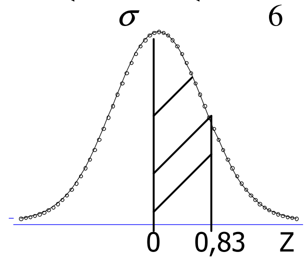
$$Z = \frac{x-m}{\sigma} = \frac{x-40}{2}$$

$0{,}675 = \frac{x-40}{2}$ = 1,35 = x - 40 = x = 41,35 polegadas

Exemplo 25 (pag 56)

a) X: altura; X~N(160,6)

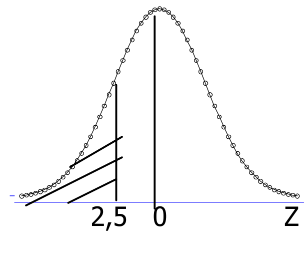
$$P(160 < X < 165) = P(\frac{160-160}{6} < \frac{X-\mu}{\sigma} < \frac{165-160}{6}) = P(0 < Z < 0{,}83)$$



Utilizando a tabela da curva normal reduzida, P(0<Z<0,83)=0,29673 ou 29,7%

b) X: altura; X~N(160,6)

$$P(X < 145) = P(\frac{X-\mu}{\sigma} < \frac{145-160}{6}) = P(Z < -2{,}5)$$



Utilizando a tabela da curva normal reduzida, P(Z<-2,5)=0,5-0,49379=0,0062 ou 0,6%

c) $P(X > 170) = P(\frac{170-160}{6} > \frac{X-\mu}{\sigma}) = P(1{,}66 > Z)$

Utilizando a tabela da curva normal reduzida, P(1,66 > Z)=0,5-0,45154=0,048 = 4,8%.

Exemplo 26 (pag 58) resolvido em sala

Exemplo 27 (pag 58)

$$\overline{X} \sim N(\mu_{\overline{X}} = 500; \sigma_{\overline{X}} = \frac{85}{\sqrt{50}} = 12,02)$$

$$P(500 \leq \overline{X} \leq 530) = P(\frac{500-500}{12,02} \leq Z \leq \frac{530-500}{12,02}) = P(0 \leq Z \leq 2,5) = 0,49379$$ ou 49,4%

$$P(450 \leq \overline{X} \leq 500) = P(\frac{450-500}{12,02} \leq Z \leq \frac{500-500}{12,02}) = P(-4,16 \leq Z \leq 0) = 0,49997$$

ou 50,0%

Gabarito - Exercícios suplementares

Exercício S1 (pag 8)

Distribuição de indivíduos segundo número de refeições. Local X, Ano Y

Número de refeições	n	%
1	5	10
2	13	26
3	15	30
4	11	22
5	3	6
6	3	6
Total	50	100

Fonte: X.

Pode-se observar na tabela que aproximadamente (78%) dos indivíduos realizam de 2 a 4 refeições por dia.

Exercício S2 (pag 8)

a) Variáveis: Tipo de comportamento (A e B) e nível de colesterol (mg/100ml).
Natureza: Tipo de comportamento (A e B) – qualitativa nominal; Nível de colesterol (mg/100ml) - quantitativa contínua.

b)

Distribuição de homens de meia idade segundo tipo de comportamento e nível de colesterol (mg/100ml). Califórnia, Estados Unidos, 1960-61.

Nível de colesterol (mg/100ml)	A		B		Total	
	n	%	n	%	n	%
120|--140	0	-	1	5	1	2,5
140|--160	0	-	2	10	2	5,0
160|--180	0	-	2	10	2	5,0
180|--200	2	10	4	20	6	15,0
200|--220	3	15	3	15	6	15,0
220|--240	5	25	2	5	7	17,5
240|--260	5	25	4	25	9	22,5
260|--280	2	10	1	5	3	7,5
280|--300	1	5	0	-	1	2,5
300|--320	1	5	0	-	1	2,5
320|--340	1	5	0	-	1	2,5
340|--360	0	-	1	5	1	2,5
Total	20	100	20	100	40	100

Fonte: *Western Collaborative Group Study, 1960-61.*

c)
Distribuição de homens de meia idade segundo nível de colesterol e tipo de comportamento Califórnia, Estados Unidos, 1960-61.

Nível de colesterol	Tipo A		Tipo B		Total	
(mg/100ml)	nº	%	nº	%	nº	%
Normal (<160)	0	-	3	15,0	3	7,5
Elevado (160 e +)	20	100,0	17	85,0	37	92,5
Total	20	100	20	100	40	100

Fonte: *Western Collaborative Group Study, 1960-61.*

Exercício S3 (pag 9)
a)
Tabela 3- Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo condição de sobrevivência e peso ao nascer (g). Local X, Ano Y.

Peso(g)	Não sobrevivente		Sobrevivente		Total	
	nº	%	nº	%	nº	%
Baixo peso (<2500)	24	64,9	13	35,1	37	100
Não baixo peso (2500 e mais)	3	23,1	10	76,9	13	100
Total	27	54,0	23	46,0	50	100

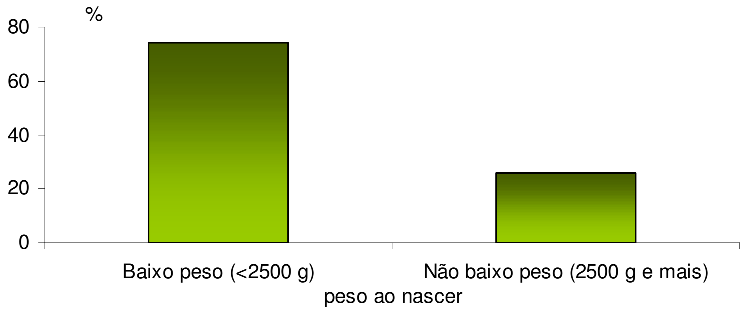
Fonte: Hand DJ et alli. *A handbook of small data sets*. Chapman&Hall, 1994.

b) Observa-se pela distribuição conjunta que a proporção de não sobreviventes é maior nos RN de baixo peso se comparado aos de não baixo peso. Os dados sugerem existência de associação entre nascer de baixo peso e não sobrevivência.

Exercício S4 (pag 9) discutido em sala

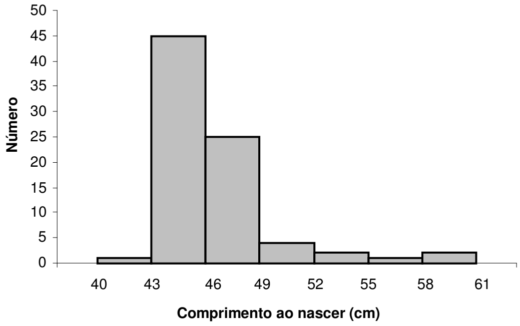
Exercício S5 (pag 9) corrigido em sala

Exercício S6 (pag 27)



Fonte: Hand DJ et al. *A handbook of small data sets*. Chapman&Hall, 1994.
Distribuição de recém-nascidos acometidos de síndrome de desconforto idiopático grave segundo peso ao nascer (g). Austrália, 1993.

Exercício S7 (pag 28)



Fonte: dados hipotéticos.
Distribuição de recém-nascidos segundo comprimento ao nascer (cm). Local X, Ano Y.

Exercício S8 (pag 28)

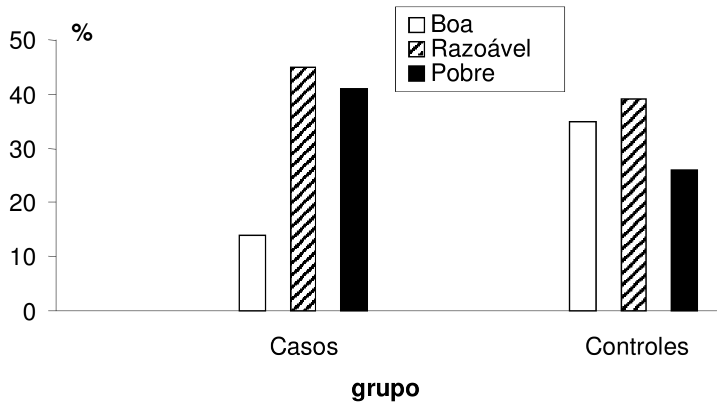
c) Calcular percentuais tomando-se como 100% o grupo (caso, controle) e interprete os resultados.

Distribuição de recém-nascidos segundo condição caso (acometidos de spina bífida) e controles segundo dieta da mãe e dieta materna. País de Gales-Reino Unido, Ano X.

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

Fonte: Hand DJ ET al., 1994.

d) Apresentar os dados em um gráfico



Fonte: Hand DJ ET al., 1994.
Distribuição de recém-nascidos segundo condição caso (acometidos de spina bífida) e controles segundo dieta da mãe e dieta materna. País de Gales-Reino Unido, Ano X.

Exercício S9 (pag 28)

Represente os dados da tabela em um polígono de frequências e interprete os resultados. Supor a última classe de etária contendo amplitude igual a 30 anos e a primeira como iniciando aos 10 anos.

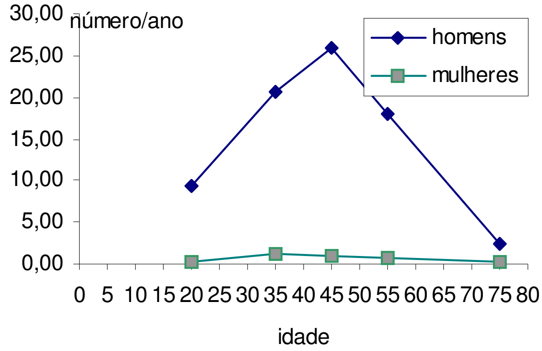
Distribuição de pessoas condenadas por embriaguez nos tribunais de *Tower Bridge* e *Lambeth* segundo idade e sexo. Londres. 1 de janeiro a 27 de junho de 1970.

Idade	Homens		Mulheres	
	Número	%	Número	%
0|-30	185	20,5	4	9,1
30|-40	207	22,9	13	29,5
40|-50	260	28,8	10	22,7
50|-60	180	19,9	7	15,9
60|-80	71	7,9	10	22,7
total	903	100	44	100

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

Corrigindo-se as frequências, obtém-se:

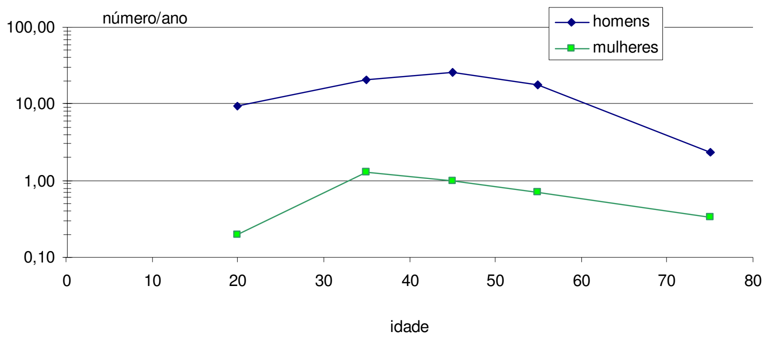
Idade	Ponto Médio	Amplitude	Número de homens	Número de homens/ano	Número de mulheres	Número de mulheres/ano
10 |- 30	20	20	185	9,25	4	0,20
30 |- 40	35	10	207	20,70	13	1,30
40 |- 49	45	10	260	26,00	10	1,00
50 |- 60	55	10	180	18,00	7	0,70
60 |- 90	65	30	71	2,37	10	0,33



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

Distribuição de pessoas condenadas por embriaguez nos tribunais de *Tower Bridge* e *Lambeth* segundo idade e sexo. Londres. 1 de janeiro a 27 de junho de 1970.

Como estão representados valores grandes e pequenos em um mesmo gráfico, os dados podem estar distorcidos devido a escala aritmética. Construindo-se o gráfico em escala logarítmica, obtém-se:



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

Distribuição de pessoas condenadas por embriaguez nos tribunais de *Tower Bridge* e *Lambeth*, Londres, 1º de janeiro a 27 de junho de 1970, segundo idade e sexo.

Exercício S11 (pag 38)

Meninos

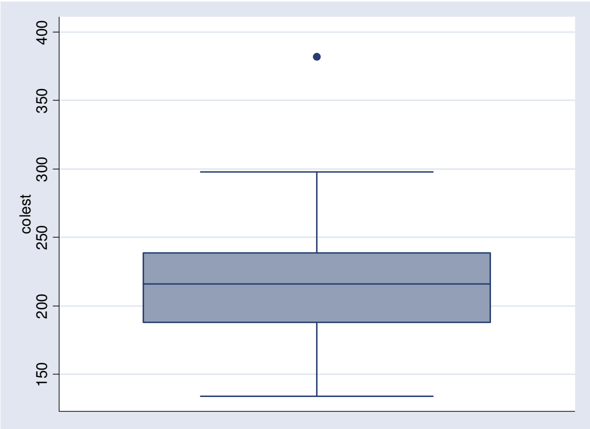
a) média = 21,7cm; mediana = 20,95cm.
b) variância = 13,3cm^2; DP=3,6cm; CV=16,6%.

Meninas

a) média = 19,9cm ; mediana = 19,65cm.
b) Variância: 5,7cm^2, DP= 2,4cm; CV=12%.
c) Não. Observando o valor da média, a CB dos meninos parece ser maior que a CB de meninas em quase 2 cm. Pelo valor da mediana esta diferença cai para 1 cm.
d) Não. A variabilidade da CB é maior em meninos quando comparado às meninas.

Exercício S12 (pag 38)

Medida	
Q1	187,5
Q2	216
Q3	238,75
Intervalo inter quartil	51,25
Valor adjacente inferior	298
Valor adjacente superior	134



Fonte: X.

Gráfico - *Box plot* da variável nível de colesterol em homens. Local X. Ano Y.

Exercício S14 (pag 51)

a) P(escolher resposta certa) = $\frac{1}{5} = 0,2$ ou 20%.

b) n=20; P(resposta certa) = 0,2; variável é discreta podendo assumir os valores T: 0,1,2,3,4,..., 20; cada questão contém somente uma resposta certa caracterizando eventos mutuamente exclusivos e o resultado de uma questão não interfere no resultado de outra. Portanto, a variável aleatória **T segue uma distribuição Binomial com parâmetros n=20 e π =0,2.**
c) P(T=3) = 0,205 ou 20,5%.
d) P(acertar 10 ou 11 ou 12 ou....ou 20) = P(acertar 10)+ P(acertar 11) +....+P(acertar20)= 0,002 ou 0,2%.
e) Resposta: 20x0,2 = 4 questões.

Exercício S15 (pag 52)

a) P(X=20) =0,001 ou 0,1%.
b) P(X=0) $\cong$ 0.
c) P(X=7) = 0,001 ou 0,1%.
d) probabilidade de que, no máximo, 10 pacientes morram da doença.
   P(X $\le$ 10) =P(X=0) + ...+ P(X=10) = 0+ = 0,031=0,048 ou 4,8%.
e) a probabilidade de que, no mínimo, 5 pacientes sobrevivam.
   =P(no máximo 15 morram) = P(X $\le$ 15) =P(X=0) + ...+ P(X=15) = 0+ = 0,179=0,762 ou 76,2%.
f) Número esperado = 20x0,7 =14 óbitos; Desvio padrão = $\sqrt{20x0,7x0,3} = 2,05$ óbitos.

Exercício S16 (pag 52)

X: número de canhotos; X: 0, 1, 2, ...,10; Probabilidade de canhoto=0,2; X~B(n=10; p=0,2)
a) P(X=2) = 0,302**.**
b) P(X≥2) = 0,302+0,201+0,088+0,026+0,006+0,001=0,624.
c) P(X≤1) = 0,107+0,268=0,375.
d) P(1 ≤ X ≤4) = 0,268+0,302+0,201+0,088 = 0,859.

Exercício S17 (pag 52)

Evento: detectar ovo (amostra ser positiva); n=5; $\pi = 0,4$ (probabilidade de detectar ovo);

X: número de amostras positivas; X: 0,1,2,3,4,5

$P(X = 0) = \binom{5}{0}(0,4)^0(0,6)^5 = 0,077$ ou **7,7%**

Exercício S18 (pag 52)
a) 0,23 ou 23%.
b) 0,729 ou 72,9%.
c) 0,501 ou 50,1%.

Exercício S19 (pag 52)
a) $\cong$ 0.
b) $\cong$ 1.
c) 0,991 ou 99,1%.
d) $\cong$ 0,008 ou 0,8%.

TÁBUA VII – DISTRIBUIÇÃO BINOMIAL B(n ; p)

Linha matriz: probabilidade do evento favorável: p

Blocos: valores do expoente do binômio $(q + p)^n$

Corpo da tábua: probabilidades P(x = r), r = 0, 1, 2, ..., n

Coluna matriz: valores da variável binomial: x

2 – 3 – 4
31 – 30 – 29

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=2
x 0	902	810	640	562	490	360	250	2
1	095	180	320	375	420	480	500	1
2	002	010	040	062	090	160	250	0 x
n=2	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=3
x 0	857	729	512	422	343	216	125	3
1	135	243	384	422	441	432	375	2
2	007	027	096	141	189	288	375	1
3	0*	001	008	016	027	064	125	0 x
n=3	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=4
x 0	815	656	410	316	240	130	062	4
1	171	292	410	422	412	346	250	3
2	014	049	154	211	265	346	375	2
3	0*	004	026	047	076	154	250	1
4	0*	0*	002	004	008	026	062	0 x
n=4	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=31
x 0	204	038	001	0*	0*	0*	0*	31
1	333	131	008	001	0*	0*	0*	30
2	263	219	029	007	001	0*	0*	29
3	134	235	070	022	006	0*	0*	28
4	049	183	122	052	017	001	0*	27
5	014	110	164	094	039	003	0*	26
6	003	053	178	135	072	009	0*	25
7	001	021	159	161	110	020	001	24
8	0*	007	119	161	142	041	004	23
9	0*	002	076	137	155	070	009	22
10	0*	0*	042	101	146	102	021	21
11	0*	0*	020	064	120	130	039	20
12	0*	0*	008	036	085	144	066	19
13	0*	0*	003	017	054	141	096	18
14	0*	0*	001	007	030	120	123	17
15	0*	0*	0*	003	014	091	140	16
16	0*	0*	0*	001	006	061	140	15
17	0*	0*	0*	0*	002	036	123	14
18	0*	0*	0*	0*	001	019	096	13
19	0*	0*	0*	0*	0*	008	066	12
20	0*	0*	0*	0*	0*	003	039	11
21	0*	0*	0*	0*	0*	001	021	10
22	0*	0*	0*	0*	0*	0*	009	9
23	0*	0*	0*	0*	0*	0*	004	8
24	0*	0*	0*	0*	0*	0*	001	7
25	0*	0*	0*	0*	0*	0*	0*	6
26	0*	0*	0*	0*	0*	0*	0*	5
27	0*	0*	0*	0*	0*	0*	0*	4
28	0*	0*	0*	0*	0*	0*	0*	3
29	0*	0*	0*	0*	0*	0*	0*	2
30	0*	0*	0*	0*	0*	0*	0*	1
31	0*	0*	0*	0*	0*	0*	0*	0 x
n=31	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=30
x 0	215	042	001	0*	0*	0*	0*	30
1	339	141	009	002	0*	0*	0*	29
2	259	228	034	009	002	0*	0*	28
3	127	236	079	027	007	0*	0*	27
4	045	177	133	060	021	001	0*	26
5	012	102	172	105	046	004	0*	25
6	003	047	179	145	083	012	001	24
7	0*	018	154	166	122	026	002	23
8	0*	006	111	159	150	050	005	22
9	0*	002	068	130	157	082	013	21
10	0*	0*	035	091	142	115	028	20
11	0*	0*	016	055	110	140	051	19
12	0*	0*	006	029	075	147	081	18
13	0*	0*	002	013	044	136	112	17
14	0*	0*	001	005	023	110	135	16
15	0*	0*	0*	002	011	078	144	15
16	0*	0*	0*	001	004	049	135	14
17	0*	0*	0*	0*	001	027	112	13
18	0*	0*	0*	0*	0*	013	081	12
19	0*	0*	0*	0*	0*	005	051	11
20	0*	0*	0*	0*	0*	002	028	10
21	0*	0*	0*	0*	0*	001	013	9
22	0*	0*	0*	0*	0*	0*	005	8
23	0*	0*	0*	0*	0*	0*	002	7
24	0*	0*	0*	0*	0*	0*	001	6
25	0*	0*	0*	0*	0*	0*	0*	5
26	0*	0*	0*	0*	0*	0*	0*	4
27	0*	0*	0*	0*	0*	0*	0*	3
28	0*	0*	0*	0*	0*	0*	0*	2
29	0*	0*	0*	0*	0*	0*	0*	1
30	0*	0*	0*	0*	0*	0*	0*	0 x
n=30	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=29
x 0	226	047	002	0*	0*	0*	0*	29
1	345	152	011	002	0*	0*	0*	28
2	254	236	039	011	002	0*	0*	27
3	120	236	088	032	009	0*	0*	26
4	041	171	144	070	026	002	0*	25
5	011	095	179	116	055	006	0*	24
6	002	042	179	155	095	015	001	23
7	0*	015	147	170	133	034	003	22
8	0*	005	101	156	157	062	008	21
9	0*	001	059	121	157	096	019	20
10	0*	0*	030	081	135	128	037	19
11	0*	0*	013	047	100	147	064	18
12	0*	0*	005	023	064	147	097	17
13	0*	0*	002	010	036	128	126	16
14	0*	0*	0*	004	018	098	144	15
15	0*	0*	0*	001	008	065	144	14
16	0*	0*	0*	0*	003	038	126	13
17	0*	0*	0*	0*	001	019	097	12
18	0*	0*	0*	0*	0*	009	064	11
19	0*	0*	0*	0*	0*	003	037	10
20	0*	0*	0*	0*	0*	001	019	9
21	0*	0*	0*	0*	0*	0*	008	8
22	0*	0*	0*	0*	0*	0*	003	7
23	0*	0*	0*	0*	0*	0*	001	6
24	0*	0*	0*	0*	0*	0*	0*	5
25	0*	0*	0*	0*	0*	0*	0*	4
26	0*	0*	0*	0*	0*	0*	0*	3
27	0*	0*	0*	0*	0*	0*	0*	2
28	0*	0*	0*	0*	0*	0*	0*	1
29	0*	0*	0*	0*	0*	0*	0*	0 x
n=29	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

TÁBUA VII – DISTRIBUIÇÃO BINOMIAL continuação – 2

5 – 6 – 7
28 – 27 – 26

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n 5
x 0	774	590	328	237	168	078	031	5
1	204	328	410	396	360	259	156	4
2	021	073	205	264	309	346	312	3
3	001	008	051	088	132	230	312	2
4	0*	0*	006	015	028	077	156	1
5	0*	0*	0*	001	002	010	031	0 x
n 5	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n 6
x 0	735	531	262	178	118	047	016	6
1	232	354	393	356	303	187	094	5
2	031	098	246	297	324	311	234	4
3	002	015	082	132	185	276	312	3
4	0*	001	015	033	060	138	234	2
5	0*	0*	002	004	010	037	094	1
6	0*	0*	0*	0*	001	004	016	0 x
n 6	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n 7
x 0	698	478	210	133	082	028	008	7
1	257	372	367	311	247	131	055	6
2	041	124	275	311	318	261	164	5
3	004	023	115	173	227	290	273	4
4	0*	003	029	058	097	194	273	3
5	0*	0*	004	012	025	077	164	2
6	0*	0*	0*	001	004	017	055	1
7	0*	0*	0*	0*	0*	002	008	0
n 7	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n 28
x 0	238	052	002	0*	0*	0*	0*	28
1	350	163	014	003	001	0*	0*	27
2	249	244	046	013	003	0*	0*	26
3	114	235	099	039	012	001	0*	25
4	037	163	155	080	032	002	0*	24
5	009	087	186	128	065	008	0*	23
6	002	037	178	164	107	020	001	22
7	0*	013	140	172	145	043	004	21
8	0*	004	092	150	163	074	012	20
9	0*	001	051	111	155	110	026	19
10	0*	0*	024	071	126	140	049	18
11	0*	0*	010	038	088	152	080	17
12	0*	0*	004	018	054	144	113	16
13	0*	0*	001	007	028	118	139	15
14	0*	0*	0*	003	013	084	149	14
15	0*	0*	0*	001	005	053	139	13
16	0*	0*	0*	0*	002	028	113	12
17	0*	0*	0*	0*	001	013	080	11
18	0*	0*	0*	0*	0*	005	049	10
19	0*	0*	0*	0*	0*	002	026	9
20	0*	0*	0*	0*	0*	001	012	8
21	0*	0*	0*	0*	0*	0*	004	7
22	0*	0*	0*	0*	0*	0*	001	6
23	0*	0*	0*	0*	0*	0*	0*	5
24	0*	0*	0*	0*	0*	0*	0*	4
25	0*	0*	0*	0*	0*	0*	0*	3
26	0*	0*	0*	0*	0*	0*	0*	2
27	0*	0*	0*	0*	0*	0*	0*	1
28	0*	0*	0*	0*	0*	0*	0*	0 x
n=28	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n 27
x 0	250	058	002	0*	0*	0*	0*	27
1	356	174	016	004	001	0*	0*	26
2	243	252	053	017	004	0*	0*	25
3	107	233	111	046	015	001	0*	24
4	034	156	166	092	039	004	0*	23
5	008	080	191	141	077	011	001	22
6	002	032	175	172	121	027	002	21
7	0*	011	131	172	155	053	007	20
8	0*	003	082	143	166	089	017	19
9	0*	001	043	101	150	125	035	18
10	0*	0*	019	060	116	150	063	17
11	0*	0*	008	031	077	154	097	16
12	0*	0*	003	014	044	137	130	15
13	0*	0*	001	005	022	105	149	14
14	0*	0*	0*	002	009	070	149	13
15	0*	0*	0*	001	003	041	130	12
16	0*	0*	0*	0*	001	020	097	11
17	0*	0*	0*	0*	0*	009	063	10
18	0*	0*	0*	0*	0*	003	035	9
19	0*	0*	0*	0*	0*	001	017	8
20	0*	0*	0*	0*	0*	0*	007	7
21	0*	0*	0*	0*	0*	0*	002	6
22	0*	0*	0*	0*	0*	0*	001	5
23	0*	0*	0*	0*	0*	0*	0*	4
24	0*	0*	0*	0*	0*	0*	0*	3
25	0*	0*	0*	0*	0*	0*	0*	2
26	0*	0*	0*	0*	0*	0*	0*	1
27	0*	0*	0*	0*	0*	0*	0*	0 x
n 27	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n 26
x 0	264	065	003	001	0*	0*	0*	26
1	361	187	020	005	001	0*	0*	25
2	237	259	061	020	006	0*	0*	24
3	100	230	123	054	019	001	0*	23
4	030	147	176	104	047	005	0*	22
5	007	072	194	153	089	015	001	21
6	001	028	170	178	134	034	003	20
7	0*	009	121	170	164	066	010	19
8	0*	002	072	134	167	104	023	18
9	0*	001	036	090	143	139	047	17
10	0*	0*	015	051	104	157	079	16
11	0*	0*	006	025	065	152	115	15
12	0*	0*	002	010	035	127	144	14
13	0*	0*	0*	004	016	091	155	13
14	0*	0*	0*	001	006	056	144	12
15	0*	0*	0*	0*	002	030	115	11
16	0*	0*	0*	0*	001	014	079	10
17	0*	0*	0*	0*	0*	005	047	9
18	0*	0*	0*	0*	0*	002	023	8
19	0*	0*	0*	0*	0*	001	010	7
20	0*	0*	0*	0*	0*	0*	003	6
21	0*	0*	0*	0*	0*	0*	001	5
22	0*	0*	0*	0*	0*	0*	0*	4
23	0*	0*	0*	0*	0*	0*	0*	3
24	0*	0*	0*	0*	0*	0*	0*	2
25	0*	0*	0*	0*	0*	0*	0*	1
26	0*	0*	0*	0*	0*	0*	0	0 x
n 26	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

TÁBUA VII – DISTRIBUIÇÃO BINOMIAL continuação – 3

3 – 9 – 10
25 – 24 – 23

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=8
x=0	663	430	168	100	058	017	004	8
1	279	383	336	267	198	090	031	7
2	051	149	294	311	296	209	109	6
3	005	033	147	208	254	279	219	5
4	0+	005	046	087	136	232	273	4
5	0+	0+	009	023	047	124	219	3
6	0+	0+	001	004	010	041	109	2
7	0+	0+	0+	0+	001	008	031	1
8	0+	0+	0+	0+	0+	001	004	0 x
n=8	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=9
x=0	630	387	134	075	040	010	002	9
1	299	387	302	225	156	060	018	8
2	063	172	302	300	267	161	070	7
3	008	045	176	234	267	251	164	6
4	001	007	066	117	172	251	246	5
5	0+	001	017	039	074	167	246	4
6	0+	0+	003	009	021	074	164	3
7	0+	0+	0+	001	004	021	070	2
8	0+	0+	0+	0+	0+	004	018	1
9	0+	0+	0+	0+	0+	0+	002	0 x
n=9	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=10
x 0	599	349	107	056	028	006	001	10
1	315	387	268	188	121	040	010	9
2	075	194	302	282	233	121	044	8
3	010	057	201	250	267	215	117	7
4	001	011	088	146	200	251	205	6
5	0+	001	026	058	103	201	246	5
6	0+	0+	006	016	037	111	205	4
7	0+	0+	001	003	009	042	117	3
8	0+	0+	0+	0+	001	011	044	2
9	0+	0+	0+	0+	0+	002	010	1
10	0+	0+	0+	0+	0+	0+	001	0 x
n=10	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=25
x=0	277	072	004	001	0+	0+	0+	25
1	365	199	024	006	001	0+	0+	24
2	231	266	071	025	007	0+	0+	23
3	093	226	136	064	024	002	0+	22
4	027	138	187	118	057	007	0+	21
5	006	065	196	165	103	020	002	20
6	001	024	163	183	147	044	005	19
7	0+	007	111	165	171	080	014	18
8	0+	002	062	124	165	120	032	17
9	0+	0+	029	078	134	151	061	16
10	0+	0+	012	042	092	161	097	15
11	0+	0+	004	019	054	147	133	14
12	0+	0+	001	007	027	114	155	13
13	0+	0+	0+	002	011	076	155	12
14	0+	0+	0+	001	004	043	133	11
15	0+	0+	0+	0+	001	021	097	10
16	0+	0+	0+	0+	0+	009	061	9
17	0+	0+	0+	0+	0+	003	032	8
18	0+	0+	0+	0+	0+	001	014	7
19	0+	0+	0+	0+	0+	0+	005	6
20	0+	0+	0+	0+	0+	0+	002	5
21	0+	0+	0+	0+	0+	0+	0+	4
22	0+	0+	0+	0+	0+	0+	0+	3
23	0+	0+	0+	0+	0+	0+	0+	2
24	0+	0+	0+	0+	0+	0+	0+	1
25	0+	0+	0+	0+	0+	0+	0+	0 x
n=25	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=24
x=0	292	080	005	001	0+	0+	0+	24
1	369	213	028	008	002	0+	0+	23
2	223	272	081	031	010	001	0+	22
3	086	221	149	075	031	003	0+	21
4	024	129	196	132	069	010	001	20
5	005	057	196	176	118	027	003	19
6	001	020	155	185	160	056	008	18
7	0+	006	100	159	176	096	021	17
8	0+	001	053	112	160	136	044	16
9	0+	0+	024	067	122	161	078	15
10	0+	0+	009	033	079	161	117	14
11	0+	0+	003	014	043	137	149	13
12	0+	0+	001	005	020	099	161	12
13	0+	0+	0+	002	008	061	149	11
14	0+	0+	0+	0+	003	032	117	10
15	0+	0+	0+	0+	001	014	078	9
16	0+	0+	0+	0+	0+	005	044	8
17	0+	0+	0+	0+	0+	002	021	7
18	0+	0+	0+	0+	0+	0+	008	6
19	0+	0+	0+	0+	0+	0+	003	5
20	0+	0+	0+	0+	0+	0+	001	4
21	0+	0+	0+	0+	0+	0+	0+	3
22	0+	0+	0+	0+	0+	0+	0+	2
23	0+	0+	0+	0+	0+	0+	0+	1
24	0+	0+	0+	0+	0+	0+	0+	0 x
n=24	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=23
x=0	307	089	006	001	0+	0+	0+	23
1	372	226	034	010	003	0+	0+	22
2	215	277	093	038	013	001	0+	21
3	079	215	163	088	038	004	0+	20
4	021	120	204	146	082	014	001	19
5	004	051	194	185	133	035	004	18
6	001	017	145	185	171	070	012	17
7	0+	005	088	150	178	113	029	16
8	0+	001	044	100	153	151	058	15
9	0+	0+	018	056	109	168	097	14
10	0+	0+	006	026	065	157	136	13
11	0+	0+	002	010	033	123	161	12
12	0+	0+	0+	003	014	082	161	11
13	0+	0+	0+	001	005	046	136	10
14	0+	0+	0+	0+	002	022	097	9
15	0+	0+	0+	0+	0+	009	058	8
16	0+	0+	0+	0+	0+	003	029	7
17	0+	0+	0+	0+	0+	001	012	6
18	0+	0+	0+	0+	0+	0+	004	5
19	0+	0+	0+	0+	0+	0+	001	4
20	0+	0+	0+	0+	0+	0+	0+	3
21	0+	0+	0+	0+	0+	0+	0+	2
22	0+	0+	0+	0+	0+	0+	0+	1
23	0+	0+	0+	0+	0+	0+	0+	0 x
n=23	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

TÁBUA VII – DISTRIBUIÇÃO BINOMIAL continuação – 4

11 – 12 – 13
22 – 21 – 20

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=11
x=0	569	314	086	042	020	004	0*	11
1	329	384	236	155	093	027	005	10
2	087	213	295	258	200	089	027	9
3	014	071	221	258	257	177	081	8
4	001	016	111	172	220	236	161	7
5	0*	002	039	080	132	221	226	6
6	0*	0*	010	027	057	147	226	5
7	0*	0*	002	006	017	070	161	4
8	0*	0*	0*	001	004	023	081	3
9	0*	0*	0*	0*	001	005	027	2
10	0*	0*	0*	0*	0*	001	005	1
11	0*	0*	0*	0*	0*	0*	0*	0=x
n=11	0,95	0,90	0,80	0,75	0,70	0,40	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=12
x=0	540	282	069	032	014	002	0*	12
1	341	377	206	127	071	017	003	11
2	099	230	283	232	168	064	016	10
3	017	085	236	258	240	142	054	9
4	002	021	133	194	231	213	121	8
5	0*	004	053	103	158	227	193	7
6	0*	0*	016	040	079	177	226	6
7	0*	0*	003	011	029	101	193	5
8	0*	0*	001	002	008	042	121	4
9	0*	0*	0*	0*	001	012	054	3
10	0*	0*	0*	0*	0*	002	016	2
11	0*	0*	0*	0*	0*	0*	003	1
12	0*	0*	0*	0*	0*	0*	0*	0=x
n=12	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=13
x=0	513	254	055	024	010	001	0*	13
1	351	367	179	103	054	011	002	12
2	111	245	268	206	139	045	010	11
3	021	100	246	252	218	111	035	10
4	003	028	154	210	234	184	087	9
5	0*	006	069	126	180	221	157	8
6	0*	001	023	056	103	197	209	7
7	0*	0*	006	019	044	131	209	6
8	0*	0*	001	005	014	066	157	5
9	0*	0*	0*	001	003	024	087	4
10	0*	0*	0*	0*	001	006	035	3
11	0*	0*	0*	0*	0*	001	010	2
12	0*	0*	0*	0*	0*	0*	002	1
13	0*	0*	0*	0*	0*	0*	0*	0=x
n=13	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=22
x=0	324	098	007	002	0*	0*	0*	22
1	375	241	041	013	004	0*	0*	21
2	207	281	107	046	017	001	0*	20
3	073	208	178	102	047	006	0*	19
4	018	110	211	161	096	019	002	18
5	003	044	190	193	149	046	006	17
6	001	014	134	183	181	086	018	16
7	0*	004	077	139	177	131	041	15
8	0*	001	036	087	142	164	076	14
9	0*	0	014	045	095	170	119	13
10	0*	0*	005	020	053	148	154	12
11	0*	0*	001	007	025	107	168	11
12	0*	0*	0*	002	010	066	154	10
13	0*	0*	0*	001	003	034	119	9
14	0*	0*	0*	0*	001	014	076	8
15	0*	0*	0*	0*	0*	005	041	7
16	0*	0*	0*	0*	0*	001	018	6
17	0*	0*	0*	0*	0*	0*	006	5
18	0*	0*	0*	0*	0*	0*	002	4
19	0*	0*	0*	0*	0*	0*	0*	3
20	0*	0*	0*	0*	0*	0*	0*	2
21	0*	0*	0*	0*	0*	0*	0*	1
22	0*	0*	0*	0*	0*	0*	0*	0=x
n=22	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=21
x=0	341	109	009	002	001	0*	0*	21
1	376	255	048	017	005	0*	0*	20
2	198	284	121	055	022	002	0*	19
3	066	200	192	117	058	009	001	18
4	016	100	216	176	113	026	003	17
5	003	038	183	199	164	059	010	16
6	0*	011	122	177	188	105	026	15
7	0*	003	065	126	172	149	055	14
8	0*	001	029	074	129	174	097	13
9	0*	0*	010	036	080	168	140	12
10	0*	0*	003	014	041	134	168	11
11	0*	0*	001	005	018	089	168	10
12	0*	0*	0*	001	006	050	140	9
13	0*	0*	0*	0*	002	023	097	8
14	0*	0*	0*	0*	0*	009	055	7
15	0*	0*	0*	0*	0*	003	026	6
16	0*	0*	0*	0*	0*	001	010	5
17	0*	0*	0*	0*	0*	0*	003	4
18	0*	0*	0*	0*	0*	0*	001	3
19	0*	0*	0*	0*	0*	0*	0*	2
20	0*	0*	0*	0*	0*	0*	0*	1
21	0*	0*	0*	0*	0*	0*	0*	0=x
n=21	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=20
x=0	358	122	012	003	001	0*	0*	20
1	377	270	058	021	007	0*	0*	19
2	189	285	137	067	028	003	0*	18
3	060	190	205	134	072	012	001	17
4	013	090	218	190	130	035	005	16
5	002	032	175	202	179	075	015	15
6	0*	009	109	169	192	124	037	14
7	0*	002	055	112	164	166	074	13
8	0*	0*	022	061	114	180	120	12
9	0*	0*	007	027	065	160	160	11
10	0*	0*	002	010	031	117	176	10
11	0*	0*	0*	003	012	071	160	9
12	0*	0*	0*	001	004	035	120	8
13	0*	0*	0*	0*	001	015	074	7
14	0*	0*	0*	0*	0*	005	037	6
15	0*	0*	0*	0*	0*	001	015	5
16	0*	0*	0*	0*	0*	0*	005	4
17	0*	0*	0*	0*	0*	0*	001	3
18	0*	0*	0*	0*	0*	0*	0*	2
19	0*	0*	0*	0*	0*	0*	0*	1
20	0*	0*	0*	0*	0*	0*	0*	0=x
n=20	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

TÁBUA VII — DISTRIBUIÇÃO BINOMIAL continuação — 5

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=14
x=0	488	229	044	018	007	001	0+	14
1	359	356	154	083	041	007	001	13
2	123	257	250	180	113	032	006	12
3	026	114	250	240	194	085	022	11
4	004	035	172	220	229	155	061	10
5	0+	008	086	147	196	207	122	9
6	0+	001	032	073	126	207	183	8
7	0+	0+	009	028	062	157	209	7
8	0+	0+	002	008	023	092	183	6
9	0+	0+	0+	002	007	041	122	5
10	0+	0+	0+	0+	001	014	061	4
11	0+	0+	0+	0+	0+	003	022	3
12	0+	0+	0+	0+	0+	001	006	2
13	0+	0+	0+	0+	0+	0+	001	1
14	0+	0+	0+	0+	0+	0+	0+	0=x
n=14	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=15
x=0	463	206	035	013	005	0+	0+	15
1	366	343	132	067	031	005	0+	14
2	135	267	231	156	092	022	003	13
3	031	129	250	225	170	063	014	12
4	005	043	188	225	219	127	042	11
5	001	010	103	165	206	186	092	10
6	0+	002	043	092	147	207	153	9
7	0+	0+	014	039	081	177	196	8
8	0+	0+	003	013	035	118	196	7
9	0+	0+	001	003	012	061	153	6
10	0+	0+	0+	001	003	024	092	5
11	0+	0+	0+	0+	001	007	042	4
12	0+	0+	0+	0+	0+	002	014	3
13	0+	0+	0+	0+	0+	0+	003	2
14	0+	0+	0+	0+	0+	0+	0+	1
15	0+	0+	0+	0+	0+	0+	0+	0=x
n=15	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=16
x=0	440	185	028	010	003	0+	0+	16
1	371	329	113	053	023	003	0+	15
2	146	275	211	134	073	015	002	14
3	036	142	246	208	146	047	009	13
4	006	051	200	225	204	101	028	12
5	001	014	120	180	210	162	067	11
6	0+	003	055	110	165	198	122	10
7	0+	0+	020	052	101	189	175	9
8	0+	0+	006	020	049	142	196	8
9	0+	0+	001	006	019	084	175	7
10	0+	0+	0+	001	006	039	122	6
11	0+	0+	0+	0+	001	014	067	5
12	0+	0+	0+	0+	0+	004	028	4
13	0+	0+	0+	0+	0+	001	009	3
14	0+	0+	0+	0+	0+	0+	002	2
15	0+	0+	0+	0+	0+	0+	0+	1
16	0+	0+	0+	0+	0+	0+	0+	0=x
n=16	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=19
x=0	377	135	014	004	001	0+	0+	19
1	377	285	068	027	009	001	0+	18
2	179	285	154	080	036	005	0+	17
3	053	180	218	152	087	017	002	16
4	011	080	218	202	149	047	007	15
5	002	027	164	202	192	093	022	14
6	0+	007	095	157	192	145	052	13
7	0+	001	044	097	153	180	096	12
8	0+	0+	017	049	098	180	144	11
9	0+	0+	005	020	051	146	176	10
10	0+	0+	001	007	022	098	176	9
11	0+	0+	0+	002	008	053	144	8
12	0+	0+	0+	0+	002	024	096	7
13	0+	0+	0+	0+	001	008	052	6
14	0+	0+	0+	0+	0+	002	022	5
15	0+	0+	0+	0+	0+	001	007	4
16	0+	0+	0+	0+	0+	0+	002	3
17	0+	0+	0+	0+	0+	0+	0+	2
18	0+	0+	0+	0+	0+	0+	0+	1
19	0+	0+	0+	0+	0+	0+	0+	0=x
n=19	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=18
x=0	397	150	018	006	002	0+	0+	18
1	376	300	081	034	013	001	0+	17
2	168	284	172	096	046	007	001	16
3	047	168	230	170	105	025	003	15
4	009	070	215	213	168	061	012	14
5	001	022	151	199	202	115	033	13
6	0+	005	082	144	187	166	071	12
7	0+	001	035	082	138	189	121	11
8	0+	0+	012	038	081	173	167	10
9	0+	0+	003	014	039	128	185	9
10	0	0+	001	004	015	077	167	8
11	0+	0+	0+	001	005	037	121	7
12	0+	0+	0+	0+	001	015	071	6
13	0+	0+	0+	0+	0+	004	033	5
14	0+	0+	0+	0+	0+	001	012	4
15	0+	0+	0+	0+	0+	0+	003	3
16	0+	0+	0+	0+	0+	0+	001	2
17	0+	0+	0+	0+	0+	0+	0+	1
18	0+	0+	0+	0+	0+	0+	0+	0=x
n=18	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

p →	0,05	0,10	0,20	0,25	0,30	0,40	0,50	n=17
x=0	418	167	023	008	002	0+	0+	17
1	374	315	096	043	017	002	0+	16
2	158	280	191	114	058	010	001	15
3	041	156	239	189	125	034	005	14
4	008	060	209	221	187	080	018	13
5	001	017	136	191	208	138	047	12
6	0+	004	068	128	178	184	094	11
7	0+	001	027	067	120	193	148	10
8	0+	0+	008	028	064	161	185	9
9	0+	0+	002	009	028	107	185	8
10	0+	0+	0+	002	009	057	148	7
11	0+	0+	0+	001	003	024	094	6
12	0+	0+	0+	0+	001	008	047	5
13	0+	0+	0+	0+	0+	002	018	4
14	0+	0+	0+	0+	0+	0+	005	3
15	0+	0+	0+	0+	0+	0+	001	2
16	0+	0+	0+	0+	0+	0+	0+	1
17	0+	0+	0+	0+	0+	0+	0+	0=x
n=17	0,95	0,90	0,80	0,75	0,70	0,60	0,50	← p

EXEMPLOS DE TABELAS BINOMIAIS - $P(r;\, n/p) = \binom{n}{r} p^r (1-p)^{n-r}$

n	r \ p	0,02	0,06	0,1	0,2	0,3	0,4	0,5		
	0	,90992	,73390	,59049	,32768	,16807	,07776	,03125	5	
	1	,09224	,23422	,32805	,40960	,36015	,25920	,15625	4	5
5	2	,00376	,02990	,07290	,20489	,30870	,34560	,31250	3	
	3	,00008	,00191	,00810	,05120	,13230	,23040	,31250	2	
	4		,00006	,00045	,00640	,02835	,07680	,15625	1	
	5			,00001	,00032	,00243	,01024	,03125	0	
	0	,81707	,53862	,34868	,10737	,02825	,00605	,00098	10	
	1	,16675	,34380	,38742	,26844	,12106	,04031	,00977	9	
	2	,01531	,09875	,19371	,30199	,23347	,12093	,04395	8	
	3	,00083	,01681	,05740	,20133	,26683	,21499	,11719	7	
	4	,00003	,00188	,01116	,08808	,20012	,25082	,20508	6	
10										10
	5		,00014	,00149	,02642	,10292	,20066	,24609	5	
	6		,00001	,00014	,00551	,03676	,11148	,20508	4	
	7			,00001	,00079	,00900	,04247	,11719	3	
	8				,00007	,00145	,01062	,04395	2	
	9					,00014	,00157	,00977	1	
	10					,00001	,00010	,00098	0	
	0	,73857	,39529	,20589	,03518	,00475	,00047	,00003	15	
	1	,22609	,37847	,34315	,13194	,03052	,00470	,00046	14	
	2	,03230	,16910	,26690	,23090	,09156	,02194	,00320	13	
	3	,00286	,04677	,12851	,25014	,17004	,06339	,01389	12	
	4	,00017	,00896	,04284	,18760	,21862	,12678	,04166	11	
15										15
	5	,00001	,00126	,01047	,10318	,20613	,18594	,09164	10	
	6		,00013	,00194	,04299	,14724	,20660	,15274	9	
	7		,00001	,00028	,01382	,08113	,17708	,19638	8	
	8			,00003	,00345	,03477	,11806	,19638	7	
	9				,00067	,01159	,06121	,15274	6	
	10				,00010	,00298	,02449	,09164	5	
	11				,00001	,00058	,00742	,04166	4	
	12					,00008	,00165	,01389	3	
	13					,00001	,00025	,00320	2	
	14						,00002	,00046	1	
	15							,00003	0	
	0	,66761	,29011	,12158	,01153	,00080	,00004		20	
	1	,27249	,37035	,27017	,05765	,00684	,00049	,00002	19	
	2	,05283	,22457	,28518	,13691	,02785	,00309	,00018	18	
	3	,00647	,08601	,19012	,20536	,07160	,01235	,00109	17	
	4	,00056	,02333	,08978	,21820	,13042	,03499	,00462	16	
	5	,00004	,00477	,03192	,17456	,17886	,07465	,01479	15	
	6		,00076	,00887	,10910	,19164	,12441	,03696	14	
	7		,00010	,00197	,05455	,16426	,16588	,07393	13	
	8		,00001	,00036	,02216	,11440	,17971	,12013	12	
	9			,00005	,00739	,06537	,15974	,16018	11	
20										20
	10			,00001	,00203	,03082	,11714	,17620	10	
	11				,00046	,01201	,07099	,16018	9	
	12				,00009	,00386	,03550	,12013	8	
	13				,00001	,00102	,01456	,07393	7	
	14					,00022	,00485	,03696	6	
	15					,00004	,00129	,01479	5	
	16					,00001	,00027	,00462	4	
	17						,00004	,00109	3	
	18							,00018	2	
	19							,00002	1	
	20								0	
		0,98	0,94	0,9	0,8	0,7	0,6	0,5	p \ r	n

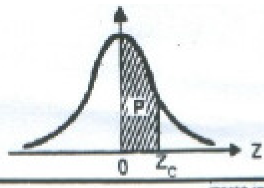
Distribuições Binomiais (continuação)

n	r \ p	0,02	0,06	0,1	0,2	0,3	0,4	0,5		
25	0	,60346	,21291	,07179	,00378	,00013			25	25
	1	,30789	,33975	,19942	,02361	,00144	,00005		24	
	2	,07540	,26023	,26589	,07084	,00739	,00038	,00001	23	
	3	,01180	,12735	,22650	,13577	,02328	,00194	,00007	22	
	4	,00132	,04471	,13842	,18668	,05723	,00710	,00038	21	
	5	,00011	,01199	,06459	,19602	,10302	,01989	,00158	20	
	6	,00001	,00255	,02392	,16335	,14717	,04420	,00528	19	
	7		,00044	,00722	,11084	,17119	,07999	,01453	18	
	8		,00006	,00180	,06235	,16508	,11998	,03223	17	
	9		,00001	,00038	,02944	,13364	,15109	,06089	16	
	10			,00007	,01178	,08164	,16116	,09742	15	
	11			,00001	,00401	,05355	,14651	,13284	14	
	12				,00117	,02678	,11395	,15498	13	
	13				,00029	,01148	,07597	,15498	12	
	14				,00006	,00422	,04341	,13284	11	
	15				,00001	,00132	,02122	,09742	10	
	16					,00035	,00884	,06089	9	
	17					,00008	,00312	,03223	8	
	18					,00002	,00092	,01433	7	
	19						,00023	,00528	6	
	20						,00005	,00158	5	
	21						,00001	,00038	4	
	22							,00007	3	
	23							,00001	2	
30	0	,54548	,15626	,04239	,00124	,00002			30	30
	1	,33397	,29921	,14130	,00928	,00029			29	
	2	,09883	,27693	,22766	,03366	,00180	,00004		28	
	3	,01882	,16498	,23609	,07853	,00720	,00027		27	
	4	,00259	,07108	,17707	,13252	,02084	,00120	,00003	26	
	5	,00028	,02359	,19230	,17228	,04644	,00415	,00013	25	
	6	,00002	,00627	,04736	,17946	,08293	,01152	,00055	24	
	7		,00137	,01804	,15382	,12165	,02634	,00190	23	
	8		,00025	,00576	,11056	,15014	,05049	,00545	22	
	9		,00004	,00157	,06756	,15729	,08228	,01332	21	
	10		,00001	,00037	,03547	,14156	,11519	,02798	20	
	11			,00007	,01612	,11031	,13962	,05088	19	
	12			,00001	,00638	,07485	,14738	,06055	18	
	13				,00221	,04442	,13604	,11154	17	
	14				,00067	,02312	,11013	,13544	16	
	15				,00018	,01057	,07831	,14446	15	
	16				,00004	,00425	,04895	,13544	14	
	17				,00001	,00150	,02687	,11154	13	
	18					,00046	,01294	,08055	12	
	19					,00013	,00545	,05088	11	
	20					,00003	,00200	,02798	10	
	21					,00001	,00063	,01332	9	
	22						,00017	,00545	8	
	23						,00004	,00190	7	
	24						,00001	,00055	6	
	25							,00013	5	
	26							,00003	4	
		0,98	0,94	0,9	0,8	0,7	0,6	0,5	p \ r	n

TÁBUA I

DISTRIBUIÇÃO NORMAL REDUZIDA: N(0 ; 1)

PROBABILIDADES p TAIS QUE p = $P(0 < Z < Z_c)$

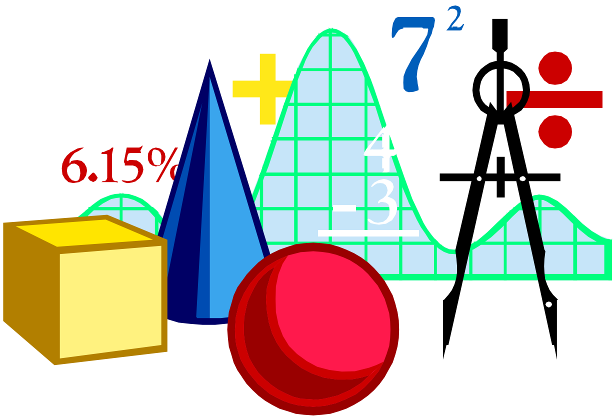


parte inteira e primeira decimal de Z_c	SEGUNDA DECIMAL DE Z_c 0	1	2	3	4	5	6	7	8	9	parte inteira e primeira decimal de Z_c
	p = 0,										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

parte inteira e primeira decimal de Z_c	SEGUNDA E TERCEIRA DECIMAIS DE Z_c 05	15	25	35	45	55	65	75	85	95	parte inteira e primeira decimal de Z_c
	p = 0,										
0,0	00199	00598	00997	01396	01795	02193	02591	02989	03387	03784	0,0
0,1	04181	04578	04974	05369	05764	06159	06553	06946	07339	07730	0,1
0,2	08121	08512	08901	09290	09677	10064	10450	10834	11218	11600	0,2
0,3	11982	12362	12741	13119	13495	13871	14244	14617	14988	15358	0,3
0,4	15726	16093	16458	16822	17184	17545	17903	18261	18600	18970	0,4
0,5	19322	19672	20021	20368	20712	21055	21396	21735	22073	22408	0,5
0,6	22741	23072	23401	23729	24054	24377	24697	25016	25333	25647	0,6
0,7	25959	26270	26577	26883	27186	27488	27786	28083	28377	28669	0,7
0,8	28959	29246	29531	29814	30094	30372	30648	30921	31192	31461	0,8
0,9	31727	31990	32252	32511	32767	33021	33273	33522	33769	34013	0,9
	05	15	25	35	45	55	65	75	85	95	

BIOESTATÍSTICA – parte 2

HEP- 146



Profa. Dra. Maria do Rosario D.O. Latorre

Titular de Bioestatística do Departamento de Epidemiologia

2013

Monitores:

Luana Fiengo Tanaka

Igor Conterato Gomes

Como fazer um levantamento de dados através de questionário?

1. Levantamento bibliográfico.
2. Definição das variáveis de estudo.
3. Elaboração das questões.

Exemplo:

IDENTIFICAÇÃO
1. nome do paciente: ____
2. número do caso (caso): ______________
3. instituição (instituto):
4. data de nascimento (dt_nasc): ______/______/______
5. data da 1ª consulta (dt_1cons): ______/______/______
6. data do diagnóstico (dt_diag): ______/______/______
7. sexo : masculino (1) feminino (2)
8. cor da pele : branco (1) negro/preto (2) pardo/mulato (3) amarelo (4) outros (5) ignorado (9)
9. renda (em reais): _____________
10. tempo de residência (em meses):__________
11. você é estrangeiro? não (0) sim (1)

CRIAÇÃO E GERENCIAMENTO DE BANCO DE DADOS

1) Elaboração da base de dados

Estabelecimento das regras para a entrada de dados (codificação das variáveis), **ANTES** de iniciar a coleta dos dados.

⇒ Nome da variável (não muito grande);

⇒ Codificação prévia dos formulários;

⇒ Livro de Códigos (manual);

⇒ Código para as categorias “não sabe”, “dado ausente” e “não se aplica”

2) Escolha do programa de entrada de dados

- ❑ Epi-Info
- ❑ Excel
- ❑ outros

Atenção!!!

- Não utilizar acentos e espaços!
- O microcomputador sabe fazer cálculos!
- Fazer cópias de segurança!

3) Estratégias para reduzir ao mínimo os erros na base de dados

- Verificações lógicas automáticas

- Entrada de dados duplicados

4) Sistema de gerenciamento de dados

ESTIMAÇÃO DE PARÂMETROS POPULACIONAIS

por ponto

$$\boxed{\bar{X} = \frac{\sum_i X_i}{n}}$$ ➔ média aritmética

$$\boxed{\hat{p} = \frac{\text{no. de sucessos}}{n}}$$ ➔ proporção

por intervalo (intervalo de confiança-IC)

$$IC = \begin{pmatrix} \text{estimativa por ponto} \\ \text{do parametro} \end{pmatrix} \pm \begin{pmatrix} \text{percentil critico da} \\ \text{distribuicao de probabilidades} \end{pmatrix} \cdot \begin{pmatrix} \text{erro padrão} \\ \text{da estimativa} \end{pmatrix}$$

$$\boxed{IC\left(\bar{X}\right) = \bar{X} \pm t_{n-1,1-\alpha} \cdot \frac{S_x}{\sqrt{n}}}$$

$$\boxed{IC\left(\hat{p}\right) = \hat{p} \pm z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}\,\hat{q}}{n}}}$$

$$|\text{-------------------}|\overbrace{\text{-----------}}^{erro}|\text{------------}|$$

$$\bar{X} - \bar{t}\,\frac{S}{\sqrt{n}} \qquad\qquad \bar{X} \qquad\qquad \hat{\bar{X}} \qquad\qquad \bar{X} + t\,\frac{S}{\sqrt{n}}$$

$$|\text{-------------------}|\overbrace{\text{-----------}}^{erro}|\text{------------}|$$

$$\hat{p} - z_{\alpha} \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}} \qquad\qquad \hat{p} \qquad\qquad p \qquad\qquad \hat{p} + z_{\alpha} \cdot \sqrt{\frac{\hat{p}.\hat{q}}{n}}$$

A DISTRIBUIÇÃO *t* DE STUDENT

A distribuição *t* de Student , a qual é simétrica em relação a 0, foi originalmente desenvolvida para descrever a variável aleatória:

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

$$\text{obs : quando } n \to \infty \quad \Rightarrow \quad T \sim N(0\,;1)$$

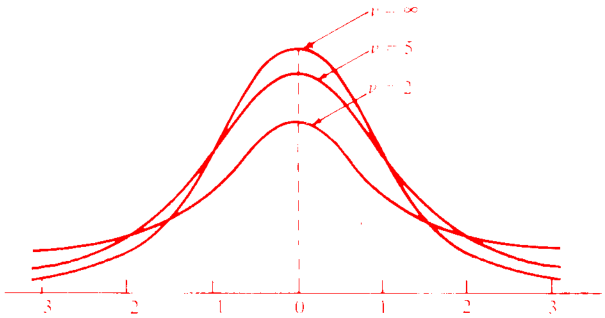


FIGURE 8.11 The *t*-distribution curves for $\nu = 2, 5$, and ∞.

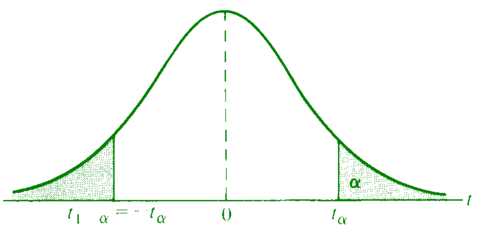
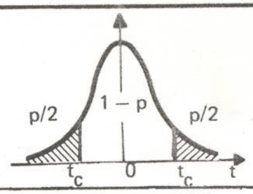


FIGURE 8.12 Symmetry property of the *t*-distribution.

TÁBUA III

DISTRIBUIÇÃO DE STUDENT: St (n)

VALORES CRÍTICOS DE t TAIS QUE $P(-t_c < t < t_c) = 1 - p$



Graus de liberdade	p = 90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	4%	2%	1%	0,2%	0,1%	Graus de liberdade
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	15,894	31,821	63,657	318,309	636,619	1
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	4,849	6,965	9,925	22,327	31,598	2
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	3,482	4,541	5,841	10,214	12,924	3
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	2,998	3,747	4,604	7,173	8,610	4
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	2,756	3,365	4,032	5,893	6,869	5
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	2,612	3,143	3,707	5,208	5,959	6
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,517	2,998	3,499	4,785	5,408	7
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,449	2,896	3,355	4,501	5,041	8
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,398	2,821	3,250	4,297	4,781	9
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,359	2,764	3,169	4,144	4,587	10
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,328	2,718	3,106	3,025	4,437	11
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,303	2,681	3,055	3,930	4,318	12
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,282	2,650	3,012	3,852	4,221	13
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,264	2,624	2,977	3,787	4,140	14
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,248	2,602	2,947	3,733	4,073	15
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,235	2,583	2,921	3,686	4,015	16
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,224	2,567	2,898	3,646	3,965	17
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,214	2,552	2,878	3,610	3,922	18
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,205	2,539	2,861	3,579	3,883	19
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,197	2,528	2,845	3,552	3,850	20
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,189	2,518	2,831	3,527	3,819	21
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,183	2,508	2,819	3,505	3,792	22
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,177	2,500	2,807	3,485	3,768	23
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,172	2,492	2,797	3,467	3,745	24
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,166	2,485	2,787	3,450	3,725	25
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,162	2,479	2,779	3,435	3,707	26
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,158	2,473	2,771	3,421	3,690	27
28	0,127	0,256	0,389	0,530	0,684	0,855	1,056	1,313	1,701	2,048	2,154	2,467	2,763	3,408	3,674	28
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,150	2,462	2,756	3,396	3,659	29
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,147	2,457	2,750	3,385	3,646	30
35	0,126	0,255	0,388	0,529	0,682	0,852	1,052	1,306	1,690	2,030	2,133	2,438	2,724	3,340	3,591	35
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,123	2,423	2,704	3,307	3,551	40
50	0,126	0,254	0,387	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,109	2,403	2,678	3,261	3,496	50
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,099	2,390	2,660	3,232	3,460	60
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,076	2,358	2,617	3,160	3,373	120
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,054	2,326	2,576	3,090	3,291	∞
	p = 90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	4%	2%	1%	0,2%	0,1%	

EXERCÍCIOS

1. Em uma amostra aleatória de 50 alunos encontrou-se que a altura média foi de 165 cm (desvio padrão de 15 cm). Construir o intervalo de confiança para essa média com 90% de confiança.

2. Desejando-se conhecer a média de consumo de carne em uma determinada população, selecionou-se uma amostra aleatória de 100 pessoas. Os resultados mostraram que, em média, os indivíduos consumiam 1000 g/mês (desvio padrão de 625 g). Determine o intervalo de confiança para essa média, com 95% de confiança.

3. Em um estudo sobre o número de atendimentos em um hospital, encontrou-se que, durante um ano (12 meses), o número médio de atendimentos por mês foi de 500 pacientes (desvio padrão de 100 pacientes). Determine o intervalo de confiança para essa média, com 98% de confiança.

4. Calcule o intervalo de 95% de confiança para a média das despesas totais com saúde/habitante, do número de nascidos vivos/habitante, do total da população/100.000, da porcentagem de despesas com pessoal/despesa total e da porcentagem de não alfabetizados. (página 37) desta apostila.

5. Desejando-se conhecer a prevalência de determinada doença na cidade A, selecionou-se uma amostra aleatória de 500 pessoas. Nesta amostra encontrou-se 20 doentes. Estimar a prevalência e calcular o respectivo intervalo de confiança (95%).

6. Para se determinar a letalidade da doença B, acompanhou-se uma amostra de 30 doentes durante um ano. Após esse período, 5 deles haviam morrido. Estimar a letalidade da doença B e calcular o respectivo intervalo de confiança (90%).

7. Desejando-se estimar a proporção de obesos em uma população, coletou-se uma amostra de 700 pessoas, sendo 350 homens e 350 mulheres. Nesta amostra, havia 130 homens e 70 mulheres obesos. Estime a proporção de obesos para cada um dos sexos e calcule os respectivos intervalos de confiança (80%).

8. Os neonatologistas da Maternidade A estão interessados em conhecer a prevalência de recém-nascidos com baixo peso ao nascer. Para tanto estudaram todos os nascimentos ocorridos durante o mês de janeiro. Analisando os dados encontraram que dos 1200 nascimentos, 200 eram de crianças com menos de 2500 gr. Estimar a prevalência de baixo peso ao nascer e calcular o respectivo intervalo de 95% de confiança.

9. Em uma amostra de 16 gestantes com diagnóstico clínico de pré-eclâmpsia, a taxa média de ácido úrico no plasma foi de 5,3 mg. Em gestantes normais a variabilidade a que está sujeita a taxa de ácido úrico no plasma é de 0,60 mg.

a) Estime, com 95% de confiança, a taxa média de ácido úrico no plasma da população de gestantes com diagnóstico de pré-eclâmpsia.
b) Que pressuposições foram necessárias para a estimação do item a)?

10. Em certa área, baseando-se na amostra de 100 recém-nascidos, a letalidade da diarréia do recém-nascido no verão e outono foi de 40%. Nestas condições, estime com 99% de confiança, a verdadeira letalidade da diarréia.

11) Desejando-se estimar a eficiência de uma droga, uma amostra de 100 pacientes foi sorteada. Supondo-se que tivessem sido observados:
a) 10 curados
b) 30 curados
c) 90 curados
Quais seriam, respectivamente, os intervalos com 95% de confiança para a verdadeira eficiência da droga utilizada.

12) Com o intuito de estudar o conteúdo de ácido lático no sangue de indivíduos com demência precoce amostra de 16 pacientes foi sorteada e os resultados foram: média ( ) = 13mg/100 cc e desvio padrão (s) = 4,6 mg/100 cc. Estime, através de um intervalo de 98% de confiança, a taxa média de ácido lático no universo dos pacientes com demência precoce.

13) Com a finalidade de estudar o efeito da aplicação de hormônios gonadotróficos, no tratamento de gestantes diabéticas, 60 receberam tal tratamento e, destas, 15 gestantes tiveram, como produto de concepção, natimortos. Estabeleça, a partir deste resultado, o intervalo com

95% de confiança para a verdadeira proporção de nascidos vivos, para medir a eficiência do tratamento.

GABARITO

1) $IC_{95\%}(\overline{X}) =$ (161,44 – 168,56)

2) $IC_{95\%}(\overline{X}) =$ (876,25 – 1123,75)

3) $IC_{95\%}(\overline{X}) =$ (421,54 – 578,46)

4) $IC_{95\%}(\overline{X}) =$ (84,56 - 103,85)

$IC_{95\%}(\overline{X}) =$ (65,31 - 170,19)

$IC_{95\%}(\overline{X}) =$ (31,33 - 94,33)

$IC_{95\%}(\overline{X}) =$ (32,78 - 39,84)

$IC_{95\%}(\overline{X}) =$ (25,74 - 32,56)

5) $IC_{95\%}(\hat{p}) =$ (0,023 – 0,057)

6) $IC_{95\%}(\hat{p}) =$ (0,057 – 0,282)

7) $IC_{95\%}(\hat{p}) =$ (0,17 – 0,23)

$IC_{95\%}(\hat{p}) =$ (0,34 – 0,40)

8) $IC_{95\%}(\hat{p}) =$ (0,146 - 0,188)

9) a)(5,006mg - 5,594mg);

b) A variabilidade da população de gestantes normais e de gestantes com diagnóstico de pré-eclâmpsia é a mesma.

10) (0,27 - 0,53)

11) a) (0,041 - 0,159) b) (0,21 - 0,39) c) (0,84 - 0,96)

12) (10,01mg - 15,99mg)

13) (0,64 - 0,86)

TESTE DE HIPÓTESES

Características de uma boa hipótese:

- hipótese simples (hipóteses complexas não podem ser testadas com um único teste estatístico e devem ser separadas em 2 ou mais hipóteses simples).

- hipótese específica (não leva à ambigüidade sobre o objeto de estudo e as variáveis ou sobre que teste estatístico deve ser aplicado).

- a hipótese deve ser estabelecida *a priori.*

HIPÓTESE NULA (H_0)

É a que estabelece a base formal para a construção do teste estatístico.

HIPÓTESE ALTERNATIVA (H_a)

Não é testada diretamente. Ela é aceita quando a hipótese nula é rejeitada.

obs: escolher, de preferência, a hipótese alternativa bicaudal. A hipótese monocaudal só deverá ser adotada sempre que apenas uma direção da associação for possível, ou importante.

ERROS

DECISÃO	VERDADE	
	H_0	H_a
H_0	não há erro	β
H_a	α	não há erro

α: erro tipo I = Prob (rejeitar H_0 , quando H_0 é verdade)

β: erro tipo II = Prob. (aceitar H_0 , quando H_0 é falsa)

1-β: poder do teste

obs: $\begin{cases} \alpha \text{ e } \beta \text{ entram no calculo do tamanho da amostra } (n) \\ \text{para um n fixo , tem-se que } \alpha \downarrow \ \beta \uparrow \end{cases}$

NÍVEL DESCRITIVO: valor de p

O nível descritivo do teste (p) é o menor nível de significância o qual o valor observado da estatística do teste é significativo.

Etapas para se fazer um teste de hipótese:

a) Determinar Ho;

b) determinar a estatística;

c) fixar alfa e achar a região crítica; e/ou trabalha-se com o nível descritivo do teste (p);

d) achar o valor da estatística;

e) verificar o posicionamento da estatística obtida com os valores observados em relação ao valor crítico.

TESTE DE UMA MÉDIA POPULACIONAL, com variância conhecida

$$\text{parametro}: \bar{\mathrm{X}}$$

$$\begin{cases} \mathrm{H_0}: \hat{\bar{X}} = \bar{\mathrm{X}} \\ \mathrm{H_a}: \hat{\bar{X}} \neq \bar{\mathrm{X}} \end{cases}$$

$$\text{estatistica}: Z_o = \frac{\hat{\bar{X}} - \bar{\mathrm{X}}}{\sigma_{\bar{\mathrm{X}}}}, \quad \text{onde } Z_c \sim N(0,1)$$

TESTE DE UMA MÉDIA POPULACIONAL, com variância desconhecida

$$\text{parametro}: \bar{\mathrm{X}}$$

$$\begin{cases} \mathrm{H_0}: \hat{\bar{X}} = \bar{\mathrm{X}} \\ \mathrm{H_a}: \hat{\bar{X}} \neq \bar{\mathrm{X}} \end{cases}$$

$$\text{estatistica}: t_o = \frac{\hat{\bar{X}} - \bar{\mathrm{X}}}{S_{\bar{\mathrm{X}}}} = \frac{\hat{\bar{X}} - \bar{\mathrm{X}}}{S_X / \sqrt{n}}, \quad \text{onde } t_c \sim t_{n-1}$$

EXERCÍCIOS

1. Em uma amostra aleatória de 50 alunos encontrou-se que a altura média foi de 165 cm (desvio padrão de 15 cm). Testar a hipótese de que essa média é igual ao esperado, sabendo-se que a média padrão é 170 cm (desvio padrão conhecido=20 cm) (α=5%)

2. Idem ao exercício anterior, porém supondo que a variância seja desconhecida. (α=5%)

3. Desejando-se conhecer a média de consumo de carne em uma determinada população, selecionou-se uma amostra aleatória de 100 pessoas. Os resultados mostraram que, em média, os indivíduos consumiam 1000 g/mês (desvio padrão de 625 g). Teste a hipótese de que o consumo médio dessa população está de acordo com o esperado, que é 1200 g/mês. (α=10%)

4. Deseja-se saber se o número médio de atendimentos diário no Posto de Saúde ZZ é igual à média diária dos postos da rede municipal que é de 40 atendimentos. Para isso coletou-se a informação dos últimos 20 dias de atendimento e verificou-se que a média é de 30 pacientes por dia (desvio padrão de 10 pacientes). Faça o teste estatístico para verificar se o número médio diário de atendimentos no Posto de Saúde ZZ é igual à média diária dos postos da rede municipais (α=2%).

5) Admite-se que a quantidade de carne ingerida por pessoa por semana (com renda familiar menor do que 3 salários mínimos e tamanho da família de 5 membros), na região Sudeste, possui distribuição normal com média 600g e desvio padrão 100g. Deseja-se saber se no subdis-

trito de Pirituba o consumo médio é menor do que esta quantidade. Para isto foi conduzida uma pesquisa, com nível de significância de 5%, cujos valores amostrais de consumo são apresentados a seguir. Elabore de forma completa o teste de hipótese.

Consumo médio semanal (em gramas):**300; 400; 350; 450; 100; 220; 150; 500; 900; 800; 600; 150; 50; 170; 370; 220**.

6) Em indivíduos sadios, o consumo renal médio de oxigênio tem distribuição normal com média igual a 12 cc/minuto e desvio padrão de 1,5 cc/minuto. Um pesquisador interessado em saber se indivíduos com insuficiência cardíaca tinham consumo maior, fixou α = 5%. Supondo que para o tamanho da amostra de 35 pessoas o consumo médio de oxigênio tivesse sido 19 cc/minuto, qual seria a conclusão?

7) Com dados da tabela abaixo, diga se o nível de proteína desses pacientes é significativamente menor (α=5%) que o nível de proteína no plasma em indivíduos sadios, cuja média é 7,0g/100cc.

Tabela 1 - Distribuição dos pacientes com endocardite sub-aguda bacteriana segundo nível de proteína no plasma sanguíneo (g/100cc). Local X, ano Y.

Proteína do plasma	Nº de pacientes
4,1	1
4,6	2
5,3	3
5,7	4
6,0	3
6,8	2
7,6	1
Total	16

GABARITO

1)Aceita Ho
2)Rejeita Ho
3)Rejeita Ho
4)Rejeita Ho

5)

X: quantidade de carne ingerida (gramas)
X~N(m=600, σ=100)

$$\begin{cases} H_0: m_{Pirituba} = m_{Sudeste} \\ H_a: m_{Pirituba} < m_{Sudeste} \end{cases}$$

$\alpha = 5\%$
$\overline{x}_{obs}$=358,13g
$z_{crítico}$ = -1,64
$z_{observado}$ = -9,68 $\Rightarrow$ **Decisão** : Rejeita-se H_0

Conclusão : O consumo médio de carne ingerida por pessoa por semana, no sub-distrito de Pirituba, é menor do que o consumo médio da região Sudeste, em um nível de significância de 5% (α = 0,05).

Caso o desvio padrão fosse desconhecido, usar-se-ia a distribuição *t-Student*
Para $\alpha = 5\% \rightarrow t_{crítico}$ = -1,753
$t_{observado}$ = -3,94 $\Rightarrow$ **Decisão** : Rejeita-se H_0
Conclusão : O consumo médio de carne ingerida por pessoa por semana no sub-distrito de Pirituba é menor do que o consumo médio da região Sudeste, em um nível de significância de 5% (α = 0,05).

6)

X: consumo renal de oxigênio (cc/minuto)
X~N(m=12; σ=1,5)
Para α = 5%, β=10% e d = 1,5

1) $\begin{cases} H_0: m_E = 12cc / minuto \\ H_a: m_E > 12cc / minuto \end{cases}$

2) para os valores de α, d e β especificados, obtém-se um tamanho de amostra n=4
$\overline{x}_{obs}$=19cc/minuto
$z_{observado}$ = 9,33; $z_{crítico}$ = 1,64 $\Rightarrow$ **Decisão** : Rejeita-se H_0
Conclusão : O consumo médio renal de O_2 em pacientes com insuficiência cardíaca é estatisticamente maior do que o consumo médio em indivíduos sadios, em um nível de significância de 5% (α=5%).

7)

X: nível de proteína (g/100cc)
X~N(m=7; σ=?)

$$\begin{cases} H_0: m_E = 7g / 100cc \\ H_a: m_E < 7g / 100cc \end{cases}$$

n=16; $\overline{x}_{obs}$ = 5,7g/100cc e S=0,885g/100cc

Para $\alpha = 5\%$, $t_{observado}$ = -5,876; $t_{crítico}$ = -1,753 $\Rightarrow$ **Decisão** : Rejeita-se H_0

Conclusão : O nível de proteína no plasma sangüíneo de pacientes com endocardite sub-aguda bacteriana é estatisticamente menor do que o de indivíduos sadios, em um nível de significância de 5% (α=5%).

TESTE DE UMA PROPORÇÃO

$$\text{parâmetro}: \mathrm{p}$$

$$\begin{cases} \mathrm{H_0} : \hat{\mathrm{p}} = \mathrm{p} \\ \mathrm{H_a} : \hat{\mathrm{p}} \neq \mathrm{p} \end{cases}$$

$$quando\ n\ é\ pequeno \Rightarrow utiliza - se\ a\ distribuição\ Binomial$$

$$quando\ p \cong \tfrac{1}{2} \Rightarrow distribuição\ Binomial\ é\ simétrica$$

$$quando\ n.p.(1-p) \geq 3 \Rightarrow B \approx N(np; np(1-p))$$

$$\text{estatística}: Z_o = \frac{\hat{\mathrm{p}} - \mathrm{p}}{S_p} = \frac{\hat{\mathrm{p}} - \mathrm{p}}{\sqrt{\frac{\mathrm{p.q.}}{\mathrm{n}}}} \text{ , onde } Z \sim N(0,1)$$

EXERCÍCIOS

1) Certa enfermidade, quando tratada pela terapia usual, apresenta 50% de curas, isto é, Ep=50%. Uma nova terapia foi proposta com o intuito de elevar esta percentagem de curas. Com o objetivo de testar a nova terapia contra a usual, 15 pacientes (com características semelhantes) foram tratados pela nova terapia. Nestas condições:

a) Formule as hipóteses correspondentes ao teste;

b) adotando a seguinte regra de decisão: "rejeitar H0 se na amostra de 15 pacientes houver 12, 13, 14, ou 15 curados", calcule:

b1) a probabilidade de rejeitar H0 se H0 fosse verdadeira;

b2) a probabilidade de aceitar H0 se H1: En = 60% fosse verdadeira;

b3) a probabilidade de aceitar H0 se H1: En = 70% fosse verdadeira;

b4) a probabilidade de aceitar H0 se H1: En = 80% fosse verdadeira.

c) Calcule as probabilidades indicadas no item b), adotando a seguinte regra de decisão: "rejeitar H0 se na amostra de 15 pacientes houver 13 ou mais curados".

2) Em certa área, os relatórios hospitalares informaram um total de 20.000 nascimentos dos quais 18.000 foram considerados partos normais, isto é, uma proporção de 90%. Uma nova maternidade foi instalada nesta área e durante um mês ocorreram 20 partos, dos quais apenas 16 foram considerados normais, isto é, uma proporção de 80%. Nestas condições, trabalhando a um nível de significância de 5% (ou mais próximo de 5%) diga se concorda com as autoridades sanitárias

que concluíram que na nova maternidade a proporção de partos normais é significantemente menor do que 90%.

3) Em certa comunidade durante um período de vários anos, a meningite meningocócica tem apresentado uma fatalidade de 20% para o grupo etário de 20 |-- 45 anos. Em 1999 nessa localidade, a meningite meningocócica se manifestou em 15 indivíduos desse grupo etário. Dos 15 casos (os quais eram semelhantes aos usualmente encontrados na referida localidade), a investigação de rotina verificou que 4 morreram. A fim de saber se a fatalidade por meningite meningocócica aumentou em 1999, faça um teste de hipóteses adotando um nível de significância de 1% (ou aproximadamente de 1%).

4) A resistência ao "resfriado comum" em uma dada indústria, durante o inverno, é de 0,60. Foi proposto um tratamento preventivo com a finalidade de aumentar para 0,70 a resistência ao "resfriado". Então:
a) formule as hipóteses.
b) fixando α= 0,05 (ou valor mais próximo) e admitindo ter sido sorteada uma amostra de tamanho n=20. Como resultado observou-se que 4 operários ficaram resfriados. Nestas condições, qual é a conclusão quanto à eficiência do medicamento?

5) Para se determinar a letalidade da doença B , acompanhou-se uma amostra de 30 doentes durante um ano. Após esse período, 5 deles haviam morrido. Testar a hipótese de que essa letalidade é igual a 20%.(α=10%)

6) Certa comunidade apresentou num período de vários anos incidência da doença X de 12 por 10.000 hab.. Em 1999, a incidência foi de 70 casos e a população estimada foi igual a 50.000 habitantes. Nestas condições, em um nível de significância de 1% (ou mais próximo) diga se concorda com as autoridades sanitárias que consideraram a situação dentro do esperado.

7) Desejando-se conhecer a prevalência de determinada doença na cidade A, selecionou-se uma amostra aleatória de 500 pessoas. Nesta amostra encontrou-se 20 doentes. Teste a hipótese de que a prevalência é semelhante ao que é descrito na literatura (p=10%).(α=5%)

8) Para se determinar a letalidade da doença B, acompanhou-se uma amostra de 30 doentes durante um ano. Após esse período, 5 deles haviam morrido. Testar a hipótese de que essa letalidade é igual a 20%.(α=10%)

9) Em uma amostra de 88 pacientes atendidos no ambulatório do Departamento de Oncologia Clínica, verificou-se que 38 eram fumantes. Teste a hipótese de que a porcentagem de fumantes atendidos neste ambulatório é igual ao referido na literatura (50%).(α=4%)

10) Estima-se que um medicamento A provoque efeitos colateral em 55% dos pacientes. Deseja-se testar se uma nova droga tem menos efeitos colaterais que A. Para tanto, tratou-se 50 pacientes com a nova droga e 30 deles apresentaram efeitos colaterais. Há diferença entre as proporções de pacientes com efeitos colaterais nos dois medicamentos? (α=1%)

11) Sabe-se que na cidade Y, 40% dos homens são obesos. Estudou-se uma amostra de 200 mulheres desta mesma cidade e verificou-se que havia 50 obesas. A prevalência de obesos entre os homens é igual à das mulheres? (α=2%)

GABARITO

1)

$$\begin{cases} H_0: E_N = 0{,}50 \\ H_a: E_N > 0{,}50 \end{cases}$$

n=15

b1) $\alpha \cong$ 1,76%	c1) $\alpha = 0{,}37\%$
b2) $\beta =$ 90,95%	c2) $\beta =$ 97,29%
b3) β = 70,31%	c2) $\beta =$ 87,32%
b4) β = 35,18%	c2) $\beta =$ 60,2%

2)

$$\hat{p} = \frac{16}{20} = 0{,}8$$

$$\begin{cases} H_0: \pi = 0{,}90 \\ H_a: \pi < 0{,}90 \end{cases}$$

$\alpha \cong$ 5% $\Rightarrow$ **Decisão** : Aceita-se H_0

Conclusão : Não há evidência estatística para concordar que a nova maternidade tenha uma menor proporção de partos normais, a um nível de significância de 5%.

3)

$$\begin{cases} H_0 : Fatalidade_{1999} = Fatalidade_{anterior} = 0{,}2 \\ H_a : Fatalidade_{1999} > Fatalidade_{anterior} \end{cases}$$

Decisão : Aceita-se H_0
$\alpha \cong 1\%$
Conclusão : Não há evidência estatística para confirmar que a fatalidade de 1999 é maior do que a fatalidade dos anos anteriores, com nível de significância próximo a 1%.

4)

1) $$\begin{cases} H_0: \text{resistencia} = 0{,}6 \\ H_a: \text{resistencia} > 0{,}6 \end{cases}$$

2) $\alpha \cong$ 5%; n=20
4 operários ficaram resfriados portanto 16 operários não ficaram doentes
Decisão : Rejeita-se H_0
Conclusão: A mediação permitiu que a resistência ao resfriado aumentasse, a um nível de significância $\alpha \cong 5\%$.

5)

$$\begin{cases} H_0: \text{letalidade} = 0{,}2 \\ H_a: \text{letalidade} \neq 0{,}2 \end{cases}$$

α=10%; n=30
5 doentes foram a óbito
Decisão : Aceita-se H_0

Conclusão: A letalidade da doença é igual a 0,20 a um nível de significância $\alpha \cong 10\%$.

6)

$$\begin{cases} H_0 : \pi_{85} = \pi_{\text{anterior}} \\ H_a : \pi_{85} \neq \pi_{\text{anterior}} \end{cases}$$

$\alpha \cong 1\%$; $z_{crítico} = \pm 2{,}58$; $z_{observado} = 1{,}29 \Rightarrow$ **Decisão** : Aceita-se H_0

Conclusão : Não há evidência estatística para discordar das autoridades sanitárias, com nível de significância próximo a 1%.

7)Rejeita Ho
8)Aceita Ho
9)Aceita Ho
10) Aceita Ho
11)Rejeita Ho

Teste de associação pelo χ^2

variável 1	variável 2			TOTAL
	1	...	j	
1 ... i				n_1 ... n_i
TOTAL	m_1	....	m_j	N

$$E_{ij} = \frac{n_i . m_j}{N}$$

$$\begin{cases} H_0\text{: ha indep e ndencia} \\ H_a\text{ : ha assoc i aca o} \end{cases}$$

$$\text{estatistica: } \chi_o^2 = \sum_k \frac{(O_k - E_k)^2}{E_k} \text{ , onde } \chi_c^2 \sim \chi_m^2$$

$$m = \begin{cases} \text{no. de graus de liberdade =} \\ (\text{no. de linhas - 1}).(\text{no. de colunas - 1}) \end{cases}$$

k: no. de caselas

limitações:

. N<20: Teste exato de Fisher

. 20≤N<40 → teste χ^2 somente se freqüências da tabela do esperado forem ≥ 5

Com correção de Yates:

$$\chi_c^2 = \sum_k \frac{\left(\left|O_k - E_k\right| - 0{,}5\right)^2}{E_k}$$

Coeficiente de associação de Yule (Y)

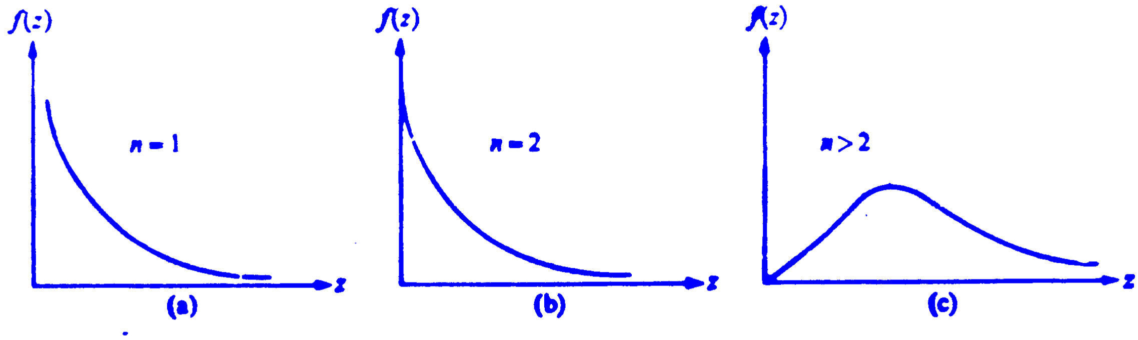
Variável 1	Variável 2		TOTAL
	sim	não	
sim	a	b	a+b
não	c	d	c+d
TOTAL	a+c	b+d	N=a+b+c+d

$$Y = \frac{a.d - b.c}{a.d + b.c} \quad \text{, onde:} \qquad -1 \leq Y \leq +1$$

DISTRIBUIÇÃO χ_m^2 (qui - quadrado com m graus de liberdade)

É a melhor medida para se avaliar as diferenças entre uma distribuição de freqüências teórica (E) e a obtida através de uma amostra (O).

A suposição básica é a de que os prováveis erros aleatórios existentes na amostra são constantes e pequenos em toda a distribuição.



$$\chi_m^2 = \sum_{i=1}^{k} \frac{(\mathrm{O_i} - \mathrm{E_i})^2}{\mathrm{E_i}} \qquad m: \text{ graus de liberdade}$$

aproximaç ão da Normal para $\chi^2 \rightarrow \; Z^2 \sim \chi_1^2$

$$\text{aproximaç ão } \chi^2 \text{ para Normal} \rightarrow \begin{cases} \text{quando n} \rightarrow \infty \Rightarrow \\ \chi_m^2 \sim N\left[(2n-1)^2 \; ; \; 1\right] \end{cases}$$

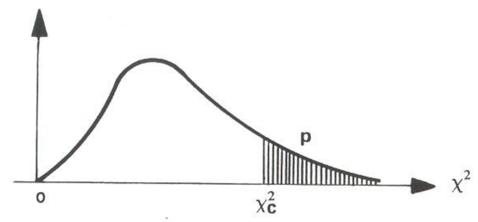
DISTRIBUIÇÃO DE QUIQUADRADO

TÁBUA II

DISTRIBUIÇÃO DE QUIQUADRADO: $\chi^2(n)$

VALORES CRÍTICOS DE QUIQUADRADO TAIS QUE

$P(\chi^2 > \chi_c^2) = p$



Graus de liberdade	p = 99%	98%	97,5%	95%	90%	80%	70%	50%	30%	20%	10%	5%	4%	2,5%	2%	1%	0,2%	0,1%	Graus de liberdade
1	$0,0^3 16$	$0,0^3 63$	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,706	3,841	4,218	5,024	5,412	6,635	9,550	10,827	1
2	0,020	0,040	0,051	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	6,438	7,378	7,824	9,210	12,429	13,815	2
3	0,115	0,185	0,216	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	8,311	9,348	9,837	11,345	14,796	16,266	3
4	0,297	0,429	0,484	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	10,026	11,143	11,668	13,277	16,924	18,467	4
5	0,554	0,752	0,831	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	11,644	12,832	13,388	15,086	18,907	20,515	5
6	0,872	1,134	1,237	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	13,198	14,449	15,033	16,812	20,791	22,457	6
7	1,239	1,564	1,690	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	14,703	16,013	16,622	18,475	22,601	24,322	7
8	1,646	2,032	2,180	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	16,171	17,534	18,168	20,090	24,352	26,125	8
9	2,088	2,532	2,700	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	17,608	19,023	19,679	21,666	26,056	27,877	9
10	2,558	3,059	3,247	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	19,021	20,483	21,161	23,209	27,722	29,588	10
11	3,053	3,609	3,816	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	20,412	21,920	22,618	24,725	29,354	31,264	11
12	3,571	4,178	4,404	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	21,785	23,337	24,054	26,217	30,957	32,909	12
13	4,107	4,765	5,009	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	23,142	24,736	25,472	27,688	32,535	34,528	13
14	4,660	5,368	5,629	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	24,485	26,119	26,873	29,141	34,091	36,123	14
15	5,229	5,985	6,262	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	25,816	27,488	28,259	30,578	35,628	37,697	15
16	5,812	6,614	6,908	7,962	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	27,136	28,845	29,633	32,000	37,146	39,252	16
17	6,408	7,255	7,564	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	28,445	30,191	30,995	33,409	38,648	40,790	17
18	7,015	7,906	8,231	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	29,745	31,526	32,346	34,805	40,136	42,312	18
19	7,633	8,567	8,906	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	31,037	32,852	33,687	36,191	41,610	43,820	19
20	8,260	9,237	9,591	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	32,321	34,170	35,020	37,566	43,072	45,315	20
21	8,897	9,915	10,283	11,591	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	33,597	35,479	36,343	38,932	44,522	46,797	21
22	9,542	10,600	10,982	12,338	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	34,867	36,781	37,659	40,289	45,962	48,268	22
23	10,196	11,293	11,688	13,091	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	36,131	38,076	38,968	41,638	47,391	49,728	23
24	10,856	11,992	12,401	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	37,389	39,364	40,270	42,980	48,812	51,179	24
25	11,524	12,697	13,120	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	38,642	40,646	41,566	44,314	50,223	52,620	25
26	12,198	13,409	13,844	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	39,889	41,923	42,856	45,642	51,627	54,052	26
27	12,879	14,125	14,573	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	41,132	43,194	44,140	46,963	53,022	55,476	27
28	13,565	14,847	15,308	16,928	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	42,370	44,461	45,419	48,278	54,411	56,893	28
29	14,256	15,574	16,047	17,708	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	43,604	45,722	46,693	49,588	55,792	[illegible]	[illegible]
30	14,953	16,306	16,791	18,493	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	44,834	46,979	47,962	50,892	57,1[illegible]	[illegible]	[illegible]
	p = 99%	98%	97,5%	95%	90%	80%	70%	50%	30%	20%	10%	5%	4%	2,5%	2%	1%	0,2[illegible]		

10

EXERCÍCIOS

1. Um estudo investigou alguns possíveis fatores de risco para câncer em criança e os resultados estão apresentados nas tabelas 2 e 3 (dados hipotéticos). Analise a possível associação entre a presença de câncer em criança menor de 12 anos e o fato de o pai ser trabalhador rural (tabela 2). Faça o mesmo em relação a antecedentes familiares de câncer (tabela 3).

Tabela 2- Número de crianças menores de 12 anos, segundo a presença de câncer e o pai ser trabalhador rural. Local X, 1990.

presença de	pai trabalhador rural		
câncer	Sim	Não	Total
Sim	4	6	10
Não	50	50	100
Total	54	56	110

Fonte: dados hipotéticos

Tabela 3- Número de crianças menores de 12 anos, segundo a presença de câncer e antecedentes familiares de câncer. Local X, 1990.

presença de	antecedentes de câncer		
câncer	Sim	Não	Total
Sim	8	2	10
Não	20	80	100
Total	28	82	110

Fonte: dados hipotéticos

2. Os dados a seguir referem-se ao trabalho “Análise de sobrevida em pacientes com lúpus eritematoso sistêmico”. (LATORRE, L.C. - Tese apresentada à Faculdade de Saúde Pública da USP para a obtenção do título de Doutor - 1997).

Faça o teste de associação para as tabelas a seguir:

a)
Tabela 4. Número de pacientes com LES, segundo ocorrência de óbito e sexo. Clínica de Reumatologia do Hospital Heliópolis. São Paulo - 1978 a 1995.

sexo	óbito		Total
	sim	não	
feminino	32	203	235
masculino	5	13	18
Total	37	216	253

p=0,1551 Y= -0,42

b)
Tabela 5. Número de pacientes com LES, segundo ocorrência de óbito e presença de envolvimento cardiopulmonar. Clínica de Reumatologia do Hospital Heliópolis. São Paulo - 1978 a 1995.

envolvimento cárdio-pulmonar	óbito		Total
	sim	não	
não	16	147	163
pulmão	10	38	48
coração	3	15	18
pulmão+coração	8	16	24
Total	37	216	253

p=0,0103

c)
Tabela 6. Número de pacientes com LES, segundo ocorrência de óbito e presença de insuficiência renal terminal (IRT). Clínica de Reumatologia do Hospital Heliópolis. São Paulo - 1978 a 1995.

IRT	óbito		Total
	sim	não	
sim	8	13	21
não	29	203	232
Total	37	216	253

p=0,0049 Y= 0,62

d)
Tabela 7. Número de pacientes com LES, segundo ocorrência de óbito e presença de hipertensão arterial (HA). Clínica de Reumatologia do Hospital Heliópolis. São Paulo - 1978 a 1995.

HA	óbito		Total
	sim	não	
sim	10	62	72
não	27	154	181
Total	37	216	253

p=0,9907 Y= -0,04

3) Analise, a possível associação (nível de significância de 5%), entre as variáveis das tabelas abaixo:

Tabela 8. Número de óbitos por suicídio segundo sexo e meio utilizado. Local X, 1999. (dados hipotéticos)

Meio utilizado	Sexo		
	Masculino	Feminino	Total
Envenenamento	31	44	75
Enforcamento	89	16	105
Total	120	60	180

Tabela 9. Número de pacientes segundo presença de câncer de esôfago e hábito de fumar. Local X, ano Y. (dados hipotéticos).

Hábito de fumar	Presença de câncer		
	Sim	Não	Total
Sim	65	115	180
Não	35	85	120
Total	100	200	300

Respostas:
Há associação estatisticamente significante entre sexo e meio utilizado no suicídio. A associação é negativa (Y=-0,77) concluindo-se que os homens utilizam mais enforcamento do que as mulheres, para um nível de significância α=5%.

Não existe associação estatisticamente significativa entre fumar e ter câncer de esôfago, em um nível de significância de 5%.

MEDIDAS DE RISCO

	DOENTE	NÃ0 DOENTE	TOTAL
EXPOSTO	a	b	a+b
NÃO EXPOSTO	c	d	c+d
TOTAL	a+c	b+d	N=a+b+c+d

Delineamento:

- estudo transversal: N é fixo
- estudo caso-controle: (a+c) e (b+d) fixos
- estudo de coorte: (a+b) e (c+d) fixos
- estudo ecológico: a, b, c e d desconhecidos

Medidas de risco:

RR: risco relativo ➔ $RR = \dfrac{\frac{a}{a+b}}{\frac{c}{c+d}}$

OR: *odds ratio* ➔ $OR = \dfrac{a/b}{c/d} = \dfrac{a.d}{b.c}$

RP: razão de prevalências ➔ $RP = \dfrac{\frac{a}{a+b}}{\frac{c}{c+d}}$

densidade de incidência, incidência acumulada.

NOÇÕES DE CORRELAÇÃO

COVARIÂNCIA

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

$$S_{XY} = \frac{\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{n - 1} = \frac{n\sum X_i Y_i - \left(\sum X_i\right)\left(\sum Y_i\right)}{n - 1} = \sum X_i Y_i - n\,\bar{X}\,\bar{Y}$$

COEFICIENTE DE CORRELAÇÃO (ρ)

Mede o grau de dependência entre 2 variáveis X e Y.

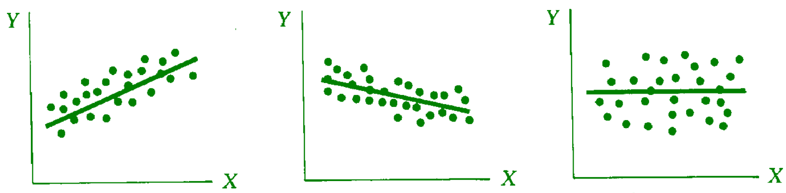
DEFINIÇÃO : $r = \dfrac{S_{XY}}{S_X\ S_Y} = \dfrac{\mathrm{Cov}(X, Y)}{S_X\ S_Y}$

$$r = \frac{\sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\left[\sum\left(X_i - \bar{X}\right)^2 \sum\left(Y_i - \bar{Y}\right)^2\right]}} = \frac{\sum X_i Y_i - \dfrac{\left(\sum X_i\right)\left(\sum Y_i\right)}{n}}{\sqrt{\left[\sum X_i^2 - \dfrac{\left(\sum X_i\right)^2}{n}\right]\left[\sum Y_i^2 - \dfrac{\left(\sum Y_i\right)^2}{n}\right]}}$$

PROPRIEDADES :

a) $-1 \le r \le +1$

b) r não possui dimensão, ié, não depende das unidades de X e Y



(a) r close to 1: strong positive association

(b) r close to -1: strong negative association

(c) r close to 0: no association

Exercício:

Vamos supor que a variável dependente (Y) seja DT e as demais as variáveis independentes. Calcular o coeficiente de correlação entre as despesas totais com saúde/habitante e o número de nascidos vivos/habitante, o total da população/100.000, a porcentagem de despesas com pessoal/despesa total e a porcentagem de não alfabetizados.

Indicadores Municipais, segundo UF. Brasil.2001

UF	1.DT	2.nasc.vivos	3. Pop.	4.Porc.DP	% Não alfab.
Rondônia	92,39	28,24	12,19	53,65	25,57
Acre	45,72	14,90	4,76	38,96	38,85
Amazonas	75,41	68,42	24,56	55,69	32,87
Roraima	80,51	9,61	3,13	36,54	28,84
Pará	86,44	134,74	54,46	30,56	32,77
Amapá	49,83	14,64	4,85	28,48	29,70
Tocantins	95,44	26,64	8,97	39,52	31,43
Maranhão	93,27	108,61	44,81	23,35	40,33
Piauí	95,16	58,65	25,53	25,54	40,19
Ceará	90,00	149,09	75,04	35,73	36,97
Rio Grande do Norte	104,76	53,53	25,88	28,13	35,09
Paraíba	84,83	64,95	33,43	30,81	37,73
Pernambuco	71,77	162,09	76,48	39,67	34,41
Alagoas	98,46	66,79	28,32	22,18	44,08
Sergipe	67,07	39,84	18,05	42,23	35,90
Bahia	62,85	234,75	120,66	28,23	33,33
Minas Gerais	124,59	297,76	176,17	36,41	21,83
Espírito Santo	81,89	57,12	29,57	47,36	21,99
Rio de Janeiro	134,51	242,36	144,22	33,81	16,97
São Paulo	126,72	632,54	374,68	51,26	17,35
Paraná	116,38	167,29	94,90	34,49	19,77
Santa Catarina	99,44	87,88	54,08	39,48	17,04
Rio Grande do Sul	118,40	160,60	103,10	32,96	16,99
Mato Grosso do Sul	127,67	40,07	21,11	35,85	22,10
Mato Grosso	122,57	47,58	24,94	41,86	23,59
Goiás	103,31	92,70	49,72	31,39	22,10
Total	107,74	3106,53	1633,61	38,98	24,76

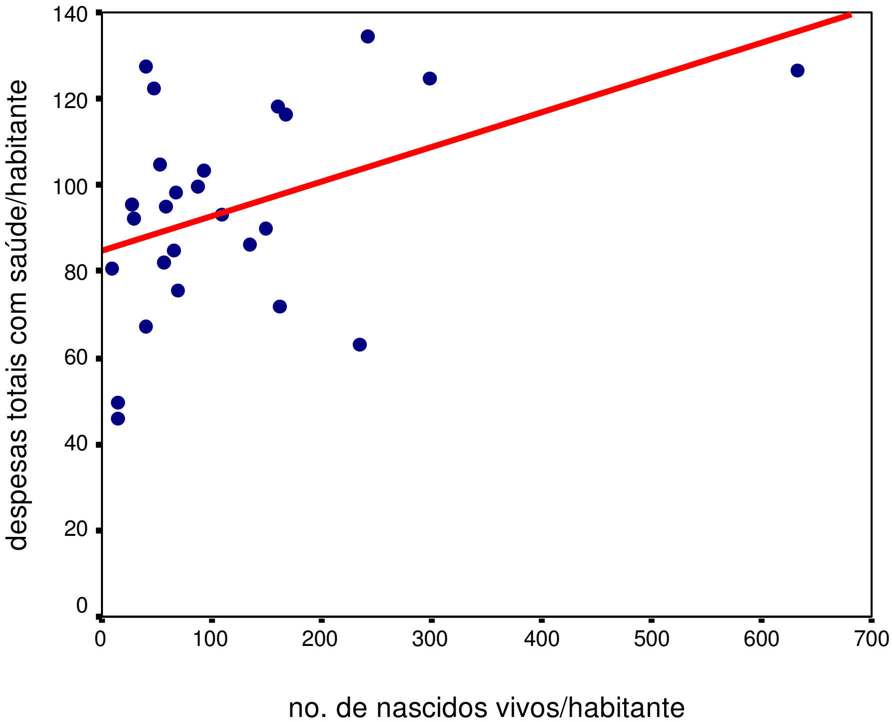
Fonte: http://tabnet.datasus.gov.br/cgi/sinasc/nvmap.htm ; 05/03/2004

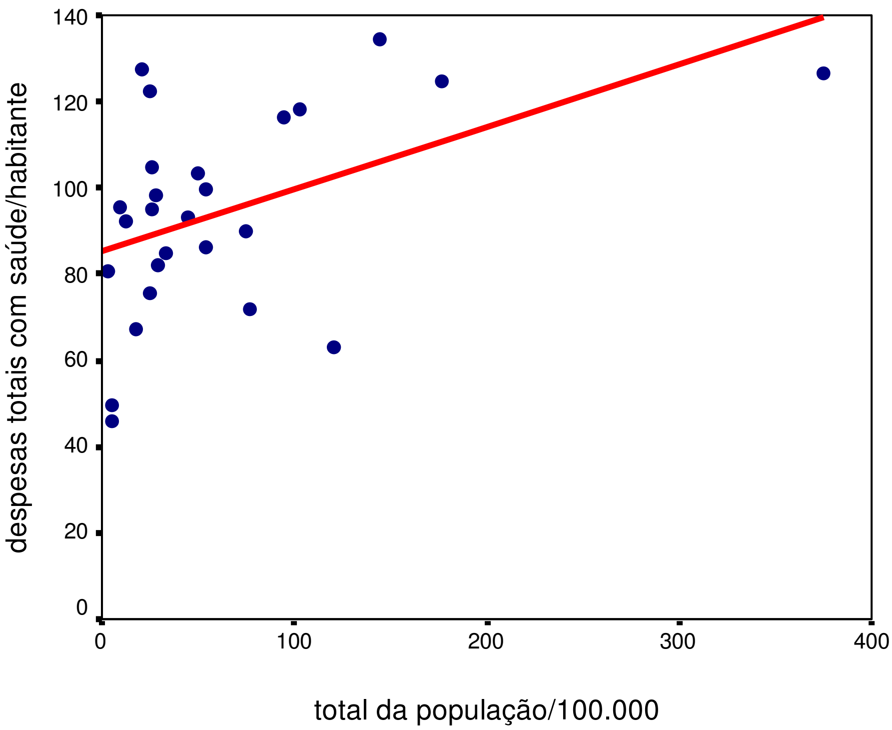
1. Despesas totais com saúde/habitante
2. Número de nascidos vivos (por 1000)
3. Total da população (por 100.000 habitantes)
4. % de despesas com pessoal/Despesas Totais
5. Porcentagem de não alfabetizados

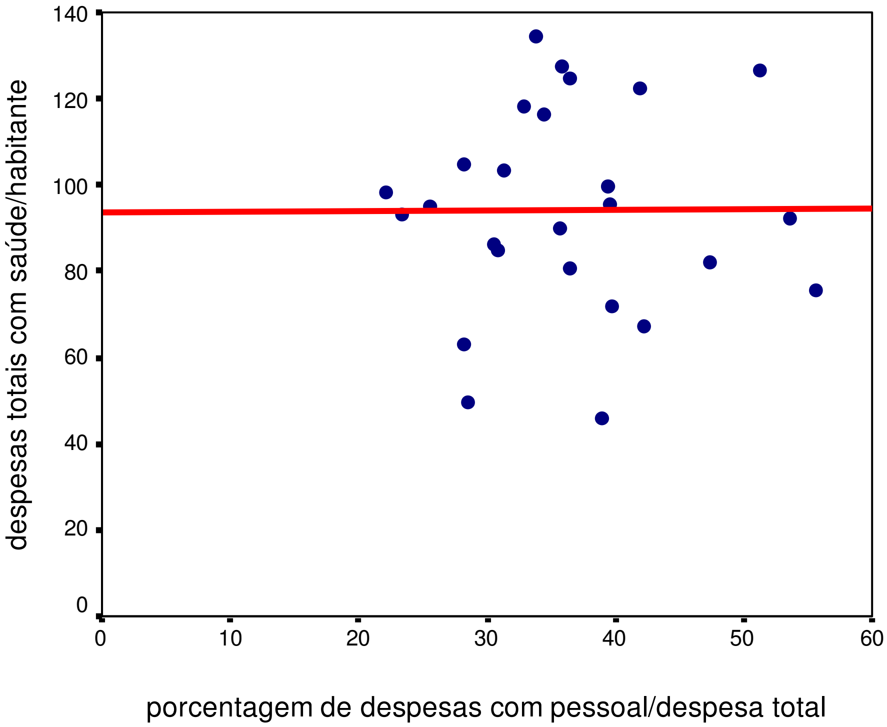
Gabarito

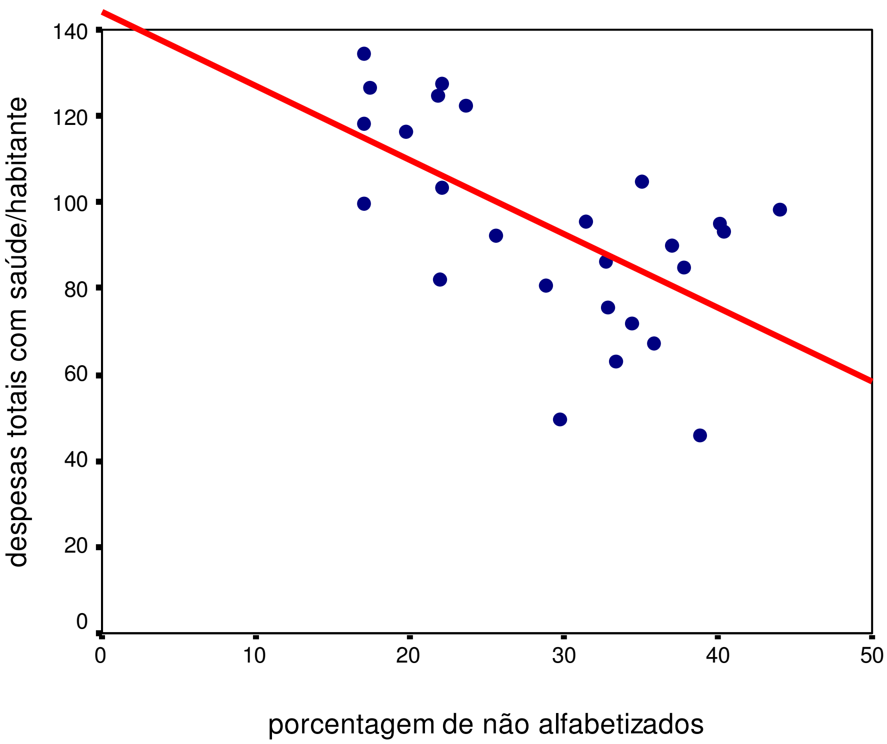
Tabela. Coeficientes de correlação de Pearson (r) entre as despesas totais com saúde/habitante e as demais variáveis de estudo.

variável	r (p)
número de nascidos vivos/habitante	0,44 (p=0,025)
total da população/100.000	0,48 (p=0,014)
porcentagem de despesas com pessoal/despesa total	0,01 (p=0,970)
porcentagem de não alfabetizados	-0,60 (p=0,001)


despesas totais com saúde/habitante
140
120
100
80
60
40
20
0
0
100
200
300
400
500
600
700
no. de nascidos vivos/habitante


despesas totais com saúde/habitante
140
120
100
80
60
40
20
0
0
100
200
300
400
total da população/100.000


despesas totais com saúde/habitante
140
120
100
80
60
40
20
0
0
10
20
30
40
50
60
porcentagem de despesas com pessoal/despesa total


despesas totais com saúde/habitante
140
120
100
80
60
40
20
0
0
10
20
30
40
50
porcentagem de não alfabetizados

O MODELO DE REGRESSÃO LINEAR SIMPLES

A função que determina uma reta é: $Y = \beta_0 + \beta_1 X$.

Porém, como se deseja fazer uma estimativa, a reta de regressão estimada pode ser escrita da seguinte maneira:

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$, e $Y = \beta_0 + \beta_1 X + \varepsilon$ ①, onde ε = erro $= Y - \hat{Y}$

$\hat{\beta}_0$ e $\hat{\beta}_1$ são estimados pelo Método dos Mínimos Quadrados da seguinte maneira:

Em uma amostra de tamanho n tem-se n pares de observações das v.a. X e Y: (X_1, Y_1), ... (X_n, Y_n) e n equações do tipo ①.

Somando-se todas as n equações, tem-se:

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} \left(\beta_0 + \beta_1 X_i + \varepsilon_i \right)$$

A soma (S) dos quadrados dos desvios (ε) é:

$$\sum_{i=1}^{n} \left(\varepsilon_i^2 \right) = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_i \right)^2$$

Para se encontrar os valores de β_0 e β_1 que minimizam a equação acima deve-se derivá-la em relação a β_0 e β_1, igualando as equações a zero. (Não se preocupem que não irei demonstrar isso nesse curso!!).

Dessa maneira os valores estimados para β_0 e β_1 são:

$$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2} \qquad ②$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \qquad ③$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \qquad ④$$

PRECISÃO DA RETA ESTIMADA

Considera-se a seguinte identidade:

$$Y_i - \hat{Y}_i = \left(Y_i - \bar{Y}\right) - \left(\hat{Y}_i - \bar{Y}\right).$$

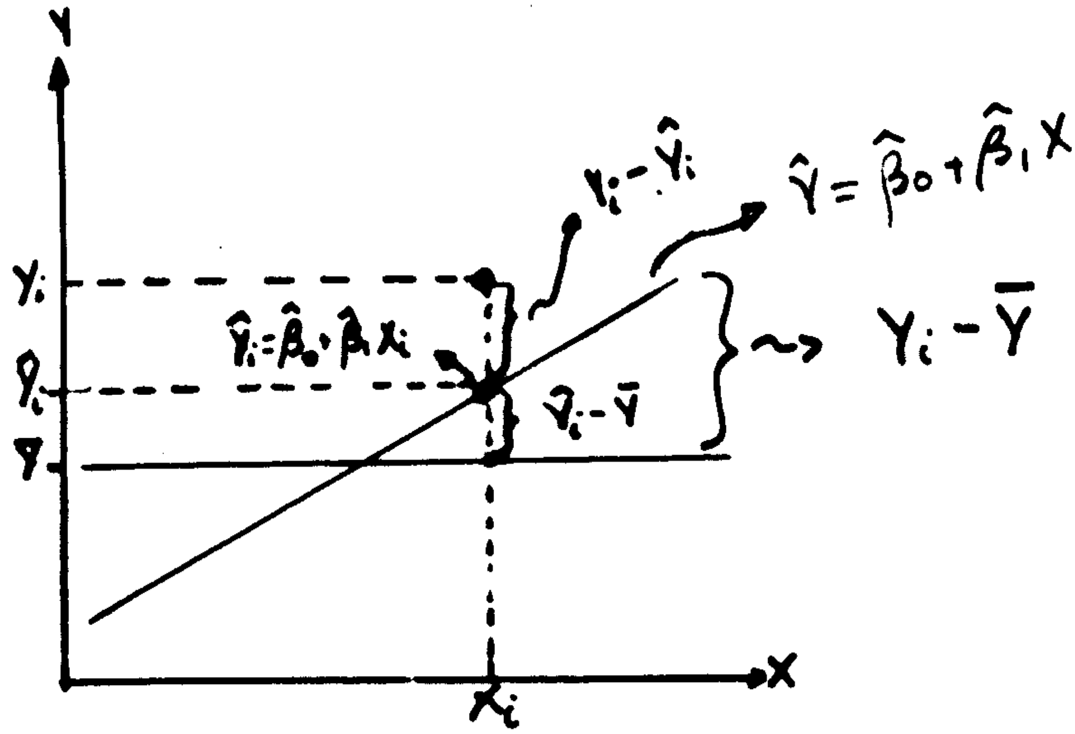
Elevando-se ao quadrado os 2 lados da igualdade acima e fazendo-se a soma de todas as n equações (i=1,2, ...,n), obtem-se:

$$\underbrace{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}_{\text{SQT}} = \underbrace{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}_{\text{SQR}} + \underbrace{\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}_{\text{SQM}} + 0 \qquad ⑤$$

- **SQT**: soma de quadrados total, ié, soma dos quadrados dos desvios do valor de Y da i-ésima observação em relação à média dos Y.

- **SQR**: soma dos quadrados devido aos resíduos, i-é,a soma dos quadrados dos desvios entre o valor de Y da i-ésima observação e seu valor estimado.

- **SQM:** soma dos quadrados devido à regressão, ié, a soma dos quadrados dos desvios do valor estimado de Y para a i-ésima observação e a média dos Y.



$$\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 + \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2$$

Isso significa que a variação total dos Y's sobre sua média pode ser explicada uma parte pela linha de regressão e outra pelos resíduos. Se todos os Y's caíssem sempre na linha de regressão a SQR seria zero!!

Portanto, quanto mais a SQM for próxima da SQT melhor.

Daí deriva-se uma medida quantitativa de precisão da reta estimada denominada r**2**.

$$r^2 = {}^{SQM}\!/_{SQT} \Rightarrow 0 \leq r^2 \leq 1$$

$\therefore$

$$\text{quanto} \quad \text{mais} \;\; r^2 \rightarrow 1, \text{melhor}$$

O Intervalo de Confianca (IC) :

$$IC = \hat{\beta}_1 \pm t_{n-2,1-\alpha} \frac{\sqrt{\dfrac{SQR}{n-2}}}{S_X \sqrt{n-1}}$$

O teste de hipotese :

$$\begin{cases} H_o : \hat{\beta}_1 = 0 \\ H_a : \hat{\beta}_1 \neq 0 \end{cases}$$

$$t_o = \frac{\hat{\beta}_1 \, S_X \sqrt{n-1}}{\sqrt{\dfrac{SQR}{n-2}}} \qquad \text{, onde} \quad t_c \sim t_{n-2}$$

EXERCÍCIO

Estimar a reta de regressão para modelos em que a variável dependente é a variável despesas totais com saúde/habitante e as demais são as variáveis independentes (número de nascidos vivos/habitante, o total da população/100.000, a porcentagem de despesas com pessoal/despesa total e a porcentagem de não alfabetizados).

GABARITO

Coefficients [a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	(Constant)	84,729	5,855		14,471	,000	72,645	96,813
	no. de nascidos vivos/habitante	,080	,034	,438	2,384	,025	,011	,150

a. Dependent Variable: despesas totais com saúde/habitante

Coefficients [a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	(Constant)	85,072	5,444		15,626	,000	73,836	96,308
	total da população/100.000	,145	,055	,475	2,643	,014	,032	,259

a. Dependent Variable: despesas totais com saúde/habitante

Coefficients [a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	(Constant)	93,445	20,793		4,494	,000	50,531	136,359
	porcentagem de despesas com pessoal/despesa total	,021	,557	,008	,038	,970	-1,129	1,171

a. Dependent Variable: despesas totais com saúde/habitante

Coefficients [a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
1	(Constant)	143,965	13,985		10,294	,000	115,101	172,829
	porcentagem de não alfabetizados	-1,707	,462	-,602	-3,698	,001	-2,660	-,754

a. Dependent Variable: despesas totais com saúde/habitante