ESTATÍSTICA É FÁCIL

ESTATÍSTICA PARA TODOS



FERNANDO RIBEIRO

Fernando Ribeiro

ESTATÍSTICA É FÁCIL

ESTATÍSTICA PARA ENGENHEIROS, ADMINISTRADORES E CONTADORES

Prof. MSc. Eng. Fernando Ribeiro

EDITORA CLUBE DE AUTORES
www.clubedeautores.com.br

ESTATÍSTICA É FÁCIL
ESTATÍSTICA PARA ENGENHEIROS, ADMINISTRADORES E CONTADORES

Copyright © 2010 por FERNANDO RIBEIRO

DIREITOS AUTORAIS RESERVADOS
www.peritofernando.cjb.net



RIBEIRO, FERNANDO
Estatística é fácil. Estatística para Engenheiros, Administradores e Contadores.
Fernando Ribeiro. São Paulo.
Ed. Clube de Autores 2010.

Aos meus pais que sempre
me apoiaram pela

educação e exemplo de vida.

In memorium: **João Fernando Ribeiro**
de eterno orgulho e lembrança.

À minha companheira Selene

Aos meus alunos com carinho.

INDICE

Capítulo I – Conceitos iniciais 7

1. Estatística: definições e conceitos básicos preliminares
2. Arredondamento
3. Razões e proporções
4. Porcentagem

Capítulo II – A divisão da Estatística 13

1. A Estatística
2. Variáveis
3. Qualitativas e quantitativas
4. Exercícios
5. Apresentação de dados estatísticos
6. Séries estatísticas
7. Exemplos e Exercícios

Capítulo III – Distribuição por freqüência 24

1. Conceitos

2. Organização de uma distribuição
3. Exercícios

Capítulo IV – Gráficos 28

1. Tipos de gráficos
2. Exercícios

Capítulo V – Técnicas de amostragem 44

1. Conceitos
2. Conceitos de População e amostra
3. Amostragem aleatória simples
4. Amostragem Proporcional estratificada
5. Amostragem sistemática
6. Distribuição por frequência
7. Exercícios

Capítulo VI – Medidas Estatísticas 64

1. Conceitos
2. Média aritmética
3. Mediana
4. Moda

Capítulo VI – Medidas de dispersão 68

1. Conceitos
2. Desvio padrão
3. Coeficiente de variação
4. Exercícios

Capítulo VII – Correlação e Regressão 72

1. Conceitos
2. Correlação linear
3. Correlação positiva e negativa
4. Análise da regressão
5. Regressão linear
6. Método dos mínimos quadrados

Capítulo VIII – Probabilidades 88

1. Conceitos
2. Exercícios
3. Testes de hipóteses
4. Tipos de erros
5. Exercícios

ESTATÍSTICA

Estatística e uma parte da matemática aplicada, que se preocupa em obter conclusões a partir de dados observados. Trata-se de levantamento de dados, apuração de dados. Os dados são colocados em forma de tabelas, gráficos para entendimento e posteriores conclusões.

A Estatística, também, propõe-se permitir a tomada de decisão, baseando-se nos dados observados.

A Estatística divide-se em Descritiva e Inferência.

A) Descritiva, ou seja, a coleta, organização e descrição dos dados. Nesse caso, a preocupação maior e apenas mostrar os dados, organizá-los, sintetizá-los sem tirar qualquer conclusão a respeito. Em sentido amplo, pode ser interpretada como uma função cujo objetivo e a observação de fenômenos, a coleta de dados numéricos referentes a esses fenômenos, a organização e a classificação desses dados observados e a sua representação através de gráficos e tabelas.

B) Inferência, ou seja, analise e interpretação dos dados. Vem a ser à parte da Estatística que tem por objetivo a tomada de decisões sobre os dados observados, permitindo inferência sobre o comportamento futuro e sobre os dados observados. A inferência estatística refere-se a um processo de generalização, a partir de resultados particulares. Consiste em obter e generalizar conclusões,

ou seja, inferir propriedades para o todo com base na parte, no particular.

A inferência estatística implica, pois, um raciocínio muito mais complexo
do que o que preside a Estatística Descritiva. Entretanto, bem compreendida e utilizada, pode converter-se em um instrumento muito importante para o desenvolvimento de uma disciplina cientifica.
Inferir significa concluir.

Exemplos Nos jornais há fatos descritos numericamente, pesquisas de opinião publica, relatórios, índices de mortalidade, crescimento de população, índice bovespa, etc. Tais informações estatísticas são de grande valia para que os interessados em curar, fazer ciência, fazer política etc.

Capítulo I

1. CONCEITOS BÁSICOS

- **População -** é o conjunto de elementos (pessoas, coisas, objetos) que têm em comum uma característica em estudo. A população pode ser:

  a) Finita: quando apresenta um número limitado de indivíduos.

  Ex.1 a população constituída por todos os parafusos produzidos em uma fábrica em um dia.

Ex. 2 nascimento de crianças em um dia em Novo Hamburgo.

b) Infinita: quando o número de observações for infinito.

Ex. a população constituída de todos os resultados (cara e coroa) em sucessivos lances de uma moeda.

- **Amostra -** é o conjunto de elementos retirados da população, suficientemente representativos dessa população. Através da análise dessa amostra estaremos aptos para analisar os resultados da mesma forma que se estudássemos toda a população.

Obs. A amostra é sempre finita. Quanto maior for a amostra mais significativa é o estudo.

- **Parâmetro -** é uma característica numérica estabelecida para toda uma população.

- **Estimador -** é uma característica numérica estabelecida para uma amostra.

- **Dado Estatístico -** é sempre um número real.

a) **Primitivo ou Bruto:** é aquele que não sofreu nenhuma transformação matemática. Número direto.

b) **Elaborado ou secundário:** é aquele que sofreu transformação matemática. Ex. porcentagem, média, etc.

2. ARREDONDAMENTO DE DADOS

- Quando o primeiro algarismo após aquele que vai ser arredondado for 0, 1, 2, 3 e 4 despreza-se este algarismo e conserva-se o anterior.

Exemplo: 5,733958 = 5,73; 78,846970 = 78,8.

- Quando o primeiro algarismo após aquele que vai ser arredondado for 5, 6, 7, 8 e 9 aumentamos uma unidade no algarismo anterior.

Exemplo: 5,735958 = 5,74; 78,886970 = 78,9.

3. RAZÕES E PROPORÇÕES.

Na Estatística deparamos com grandezas que devemos classificar em proporcionais ou não proporcionais.

Ex.: Idade, Faixa de renda, Grau de instrução, Número de filhos, Grandezas físicas, Velocidade, etc.

Grandezas diretamente proporcionais: Velocidade média e distância percorrida.

Grandezas inversamente proporcionais: Velocidade média e tempo.

Exercícios:

1) Determinar a distância que um automóvel percorrerá em 8 horas, sabendo-se que a mesma velocidade foi mantida durante 6 horas e o carro percorreu 900 km.
   R. 1200km.
2) Um automóvel com velocidade média de 90km/h, percorre um certo espaço durante 8 horas. Qual será o tempo

necessário para percorrer o mesmo espaço com velocidade média de 60km/h? R. 12 h.

4. PORCENTAGEM.

Na idade média os fenícios inventaram a porcentagem para negócios, navegação e finanças.

100 é o padrão de comparação.

Exercícios:

1) A gasolina vai aumentar 5%, sendo seu valor de R$2,10 qual será o novo valor?
2) Calcular 55% de 2000 e 217% de 4200.
3) Qual o número cujos 20% é 36?
4) Comprei um terreno cujo preço estava estipulado em R$200.000,00. Gastei 5% desse valor em impostos e 6% de corretagem. Quanto efetivamente tive de desembolsar?
5) Obtive 14% de desconto numa compra de R$24.000,00. Quanto paguei?
6) Numa prova de 50 questões quem errou 8 acertou qual porcentual?
7) A tabela abaixo traz a porcentagem da população mundial que vive nas grandes cidades desde 1980. Qual o número de habitantes em cada 1000 viviam nas grandes cidades?

1980	1985	1990	1995
39,9	41,6	43,6	45,9

8) No momento tenho disponível uma quantia de R$400,00. Se 35% do que tenho corresponde a 20% do que você tem, então, qual a quantia que você tem disponível?

9) Se 2/3 de uma obra foi realizada em 5 dias por 8 operários, trabalhando 6 h/dia o restante da obra será feito em quantos dias com 6 operários trabalhando 10h/dia.

Capítulo II

1. DIVISÃO DA ESTATÍSTICA

Podemos dividir a Estatística em duas áreas:

- Estatística Descritiva – é à parte da Estatística que tem por objetivo descrever os dados observados e na sua função dos dados, tem as seguintes atribuições.

    a) **A obtenção ou coleta de dados –** é normalmente feita através de um questionário ou de observação direta de uma população ou amostra.
    b) **A organização dos dados –** consiste na ordenação e crítica quanto à correção dos valores observados, falhas humanas, omissões, abandono de dados duvidosos.
    c) **A representação dos dados –** os dados estatísticos podem ser mais facilmente compreendidos quando apresentados através de tabelas e gráficos, que permite uma visualização instantânea de todos os dados.
- Estatística Inferencial – é à parte da Estatística que tem por objetivo obter e generalizar conclusões para a população a partir de uma amostra, através do cálculo de probabilidade. A tais conclusões estão sempre associados a um grau de incerteza e conseqüentemente, a uma probabilidade de erro.

5. VARIÁVEIS

Uma variável é qualquer característica de um elemento observado (pessoa, objeto ou animal).

Algumas variáveis, como sexo e designação de emprego, simplesmente enquadram os indivíduos em categorias. Outras,

como altura e renda anual, tomam valores numéricos com os quais podemos fazer cálculos.

Os exemplos acima nos dizem que uma variável pode ser:

I) Qualitativa: quando seus valores são expressos por atributos: sexo (masculino – feminino), cor da pele (branca, preta, amarela, vermelha);

II) Quantitativa: quando seus valores são expressos em números (salários dos operários, idade dos alunos de uma escola, número de filhos, etc.). Uma variável quantitativa que pode assumir, teoricamente, qualquer valor entre dois limites recebe o nome de **variável contínua** (altura, peso, etc.); uma variável que só pode assumir valores pertencentes a um conjunto enumerável recebe o nome de **variável discreta** (número de filhos, número de vitórias).

Exercícios

1) Classifique as variáveis abaixo:

a) Tempo para fazer um teste.
b) Número de alunos aprovados por turma.
c) Nível sócio-econômico
d) QI (Quociente de inteligência).
e) Sexo
f) Gastos com alimentação.
g) Opinião com relação à pena de morte
h) Religião
i) Valor de um imóvel
j) Conceitos em certa disciplina
k) Classificação em um concurso.

1. Identifique e classifique as variáveis:

a) Tabela de códigos de declaração de bens e direitos de imóveis: 11 – Apartamento; 12 - Casas; 13 – Terrenos; 14 – Terra nua; 15 – Salas ou lojas; 16 – Construção; 17 – Benfeitorias; 19 – Outras; (Declaração de Ajuste Anual, Instruções de Preenchimento, Imposto de Renda, Pessoa Física, 1999)

b) “O euro começa a circular com 13 bilhões de notas em sete valores(5, 10, 20, 50, 100, 200 e 500)...A cunhagem de 75 bilhões de moedas de 1 e 2 euros e de 1, 2, 5, 10, 20 e 50 centavos de euro implicará uma troca completa de máquinas e equipamentos de venda de jornais,café e refrigerantes.” (Revista Época, Ano 1, nº 33 , 4/1/1999)

APRESENTAÇÃO DE DADOS ESTATÍSTICOS

APRESENTAÇÃO TABULAR

A apresentação de dados estatísticos na forma tabular consiste na reunião ou grupamento dos dados em tabelas ou quadros com a finalidade de apresenta-los de modo ordenado, simples e de fácil percepção e com economia de espaço.

Componentes Básicos

Em termos genéricos, uma tabela se compõe dos seguintes elementos básicos:

Título

Cabeçalho

Indicadora de Coluna	Casa Linha Coluna

Rodapé

Exemplo:

Brasil - Estimativa de População
1970 – 76

Ano

População (1000 habitantes)	
1970	93.139
1971	95.993
1972	98.690
1973	101.433
1974	104.243
1975	107.145
1976	110.124

Fonte: Anuário Estatístico do Brasil

- **Principais Elementos de uma Tabela**

Título: Conjunto de informações, as mais completas possíveis, localizado no topo da tabela, respondendo às perguntas: **O quê? Onde? Quando?**

Cabeçalho: Parte superior da tabela que especifica o conteúdo das colunas.

Coluna Indicadora: Parte da tabela que especifica o conteúdo das linhas.

Linhas: Retas imaginárias que facilitam a leitura, no sentido horizontal, de dados que se inscrevem nos seus cruzamentos com as colunas.

Casa ou Célula: Espaço destinado a um só número.

Rodapé: são mencionadas a fonte se a série é extraída de alguma publicação e também as notas ou chamadas que são esclarecimentos gerais ou particulares relativos aos dados.

SÉRIES ESTATÍSTICAS

É toda tabela que apresenta a distribuição de um conjunto de dados estatísticos:

Exemplos:

Produção de petróleo bruto – Brasil
1966 – 1970.

Anos	Quantidade (cm^3)
1966	6.748.889
1967	8.508.848
1968	9.509.639
1969	10.169.531
1970	9.685.641

Fonte Brasil em dados.

Rebanhos bovinos – Brasil
1970.

Regiões	Bovinos (1000)
Norte	2.132
Nordeste	20.194
Sudeste	35.212

Sul	18.702
Centro-oeste	15.652

Fonte Brasil em dados.

Produção pesqueira – Brasil 1969.

Itens	Produção (ton.)
Peixes	314
Crustáceos	62
Moluscos	3
Mamíferos	12

Fonte Brasil em dados.

- Séries Composta ou Mista: é a combinação de dois ou mais elementos fundamentais de séries estatísticas.

Evolução do transporte de carga marítima nas 4 principais bacias brasileiras
Brasil -1968– 1970.

Bacias	Anos		

	1968	1969	1970
	233.768	324.350	316.557
Amazônica	16.873	20.272	20.246
	177.705	203.966	201.464
Nordeste	53.142	48.667	57.948
Prata			
São Francisco			

Fonte Brasil em dados.

* Os dados estão em toneladas.

A apresentação tabular de dados estatísticos é normalizada pela resolução nº 886 de 26-10-1966 do Conselho Nacional de Estatística a fim de uniformizar a apresentação de dados.

EXERCÍCIOS

1. : De acordo com o IBGE (1988), em 1986 ocorreram, em acidentes de trânsito, 27306 casos de vítimas fatais, assim distribuídos: 11712 pedestres, 7116 passageiros e 8478 condutores. Faça uma tabela para apresentar esses dados.

2. : De acordo com Ministério da Educação a quantidade e alunos matriculados no ensino de 1º grau no Brasil nos de 1990 a 1996 em milhares de alunos, são: 19.720 – 20.567 – 21.473 – 21.887 – 20.598 – 22.473 – 23.564. Faça uma tabela para apresentar esses dados.

3.: Estabelecimentos de ensino da região norte do Brasil em 1982. A região norte subdivide-se em: Rondônia, Acre, Amazonas, Roraima, Pará e Amapá e possuem um total de 29,

13, 78, 4, 10 e 9 estabelecimentos de ensino, respectivamente, segundo o MEC. Faça uma tabela para apresentar esses dados.

4.: De acordo com o IBGE(1988), a distribuição dos suicídios ocorridos no Brasil em 1986, segundo a causa atribuída, foi a seguinte: 263 por alcoolismo, 198 por dificuldade financeira, 700 por doença mental, 189 por outro tipo de doença, 416 por desilusão amorosa e 217 por outras causas. Apresente essa distribuição em uma tabela.

5.: Muitos sistemas escolares fornecem o acesso a Internet para seus estudantes hoje em dia. Desde 1996, o acesso à Internet foi facilitado a 1.133 escolas elementares, 886 escolas do nível médio e 482 escolas de nível superior. Existe em Portugal um total de 1.745 escolas elementares, 1.012 escolas do nível médio e 529 escolas do nível superior. Faça uma tabela com os dados.

DISTRIBUIÇÃO DE FREQUÊNCIA

É o tipo de série estatística na qual permanece constante o fato, o local e a época. Os dados são colocados em classes pré-estabelecidas, registrando freqüência.

Exemplos:

1) Sem intervalo de classes.

É uma série de dados agrupados na qual o número de observações está relacionados com um ponto real.

Ex.: Notas do Aluno "X" na Disciplina de Estatística – 1990

Nota	*Alunos*
6.3	2
8.4	3
5.3	2
9.5	3
6.5	5
Total	15

2) Com intervalo de classes.

Notas do curso de
Engenharia na disciplina de
Resistência dos Materiais de uma Faculdade

Notas	Nº de Estudantes
5 |-- 6	18
6 |-- 7	15
7 |-- 8	12
8 |-- 9	03
9 |--10	02

Elementos Principais:

a) Classe – é cada um dos intervalos em que os dados são agrupados.

b) Limites de classes são os valores extremos de cada classe.

l_i = limite inferior de uma classe;
L_i = limite superior de uma classe.

c) Amplitude – é a diferença entre o maior valor e o menor valor de certo conjunto de dados. Pode ser referida ao total de dados ou a uma das classes em particular.

- Amplitude Total (A_t) – é calculada pela seguinte expressão:

A_t = Max. (rol) – Min. (rol).

- Amplitude das classes (h) – é a relação entre a amplitude total e o número de classes, conforme mostra a expressão a seguir:

$h = \frac{Máx(rol) - Mín.(rol)}{n}$, onde n é o número de intervalos de classe.

d) Ponto médio de classe (x_i) - é calculado pela seguinte expressão:

$$x_i = \frac{L_i + l_i}{2}$$

e) Freqüência absoluta (f_i) - freqüência absoluta de uma classe de ordem i, é o número de dados que pertencem a essa classe.

f) Freqüência relativa (fr_i) - freqüência relativa de uma classe de ordem i, é o quociente da freqüência absoluta dessa classe (f_i), pelo total, ou seja,

$$fr = \frac{f_i}{Total}$$

Obs: a soma de todas as freqüências absolutas é igual ao total.

g) Freqüência acumulada (F_i) - freqüência acumulada de uma classe de ordem i, é a soma das freqüências até a classe de ordem i.

h) Freqüência relativa acumulada (Fr_i) - freqüência relativa acumulada de uma classe de ordem i, é a soma das freqüências relativas até a classe de ordem i.

ORGANIZAÇÃO DE UMA DISTRIBUIÇÃO DE FREQUÊNCIA:

Para organizar um conjunto de dados quantitativos em distribuição de freqüências, aconselha-se seguir a seguinte orientação:

1º Organizar o rol – colocar os dados em ordem crescente ou ordem decrescente.

2º Calcular (ou adotar) o número conveniente de classes – o número de classe deve ser escolhido pelo pesquisador, em geral, convém estabelecer de 5 a 15 classes. Existem algumas fórmulas para estabelecer quantas classes devem ser construídas. Nós usaremos,

$n \approx \sqrt{N}$ onde N é a quantidade total de observações.

3º Calcular (ou adotar) a amplitude do intervalo de classes conveniente - a amplitude do intervalo de classes deve ser o mesmo para todas as classes.

$h \approx \frac{Máx(rol) - Mín.(rol)}{n}$ onde n é o número de intervalos de classe.

4º Obter os limites das classes – Usualmente as classes são intervalos abertos á direita. Os limites são obtidos fazendo-se.

Limite inferior da 1ª classe é igual ao mínimo do rol, isto é,

$$l_1 = \text{Min.(rol)}$$

Encontram-se os limites das classes, adicionando-se sucessivamente a amplitude do intervalo de classes aos limites da 1ª classe.

5º Obter as f_i - contar o número de elementos do rol, que pertencem a cada classe.

6º Apresentar a distribuição – construir uma tabela com título, subtítulo, ...

Exercícios

1) Abaixo são relacionados os salários semanais (em Reais) de 60 operários de uma fábrica de sapatos.

110	120	125	136	145	150	165	172	180	185
110	120	125	140	145	155	165	172	180	190
115	120	130	140	145	158	168	175	180	190
115	120	130	140	147	158	168	175	180	195
117	120	130	140	150	160	170	175	180	195
117	123	135	142	150	163	170	178	185	198

1)Construir uma distribuição de freqüências adequada.
2)Interpretar os valores da terceira classe.

2) Abaixo são relacionados às estaturas e os pesos de 25 alunos de Estatística.

Estaturas Pesos

1.71	1.80	1.75	1.73	1.81		58	60	60	62	63
1.90	1.80	1.71	1.74	1.77		80	77	70	82	62
1.63	1.80	1.78	1.84	1.81		55	76	83	50	78
1.83	1.80	1.75	1.79	1.65		79	70	60	76	83
1.72	1.88	1.80	1.66	1.89		77	60	65	71	63

Construir uma distribuição de freqüências adequada para cada conjunto de dados.

3) Uma amostra de 20 operários de uma companhia apresentou os seguintes salários recebidos durante uma certa semana, arredondados para o valor mais próximo e apresentados em ordem crescente: 140, 140, 140, 140, 140, 140, 140, 140, 155, 155, 165, 165, 180, 180, 190, 200, 205, 225, 230, 240. Construir uma distribuição de freqüências adequada.

4) Complete os dados que faltam na distribuição de freqüência:

a)

Classes	x_i	f_i	F_i	f_i(%)
0 |-- 2	1	4	...	4
2 |-- 4	...	8	...	...
4 |-- 6	5	...	30	18
...	7	27	...	27
8 |-- 10	...	15	72	...
10 |-- 12	...	...	83	...
...	13	10	93	10
14 |-- 16	...	...	...	7
□		...		....

b)

Salários	x_i	f_i	F_i
500 |-- 700	600	8	8
...	800	20	...
900 |-- 1.100	...	...	35
1.100 |-- 1.300	...	5	40 ...
1.300 |-- 1.500	1.400	...	
...	...	1	43
1.700 |-- 1.900	1.800	...	...
Total		44	

GRÁFICOS ESTATÍSTICOS

O gráfico estatístico é uma forma de apresentação dos dados estatísticos, cujo objetivo é o de produzir, no investigador ou

no público em geral, uma impressão mais rápida e viva do fenômeno em estudo, já que os gráficos falam mais rápido à compreensão que as séries.

A representação gráfica de um fenômeno deve obedecer a certos requisitos fundamentais para ser realmente útil:

a) **Simplicidade** – o gráfico deve ser destituído de detalhes de importância secundária, assim como de traços desnecessários que possam levar o observador a uma análise com erros.
b) **Clareza** – o gráfico deve possibilitar uma correta interpretação dos valores representativos do fenômeno em estudo.
c) **Veracidade** – o gráfico deve expressar a verdade sobre o fenômeno em estudo.

Tipos de gráficos

Histograma, Polígono de Freqüência e Ogiva: São utilizados para representar a distribuição de freqüência.

Histograma é um gráfico formado por um conjunto de retângulos justapostos, sendo que as alturas dos retângulos são proporcionais às freqüências das classes.

Polígono de freqüência é um gráfico de linha cujo eixo das abscissas estão os pontos médios dos intervalos de classe.

Ogiva ou polígono de freqüência acumulada é o gráfico marcando-se no eixo das ordenadas as freqüências acumuladas.

Histograma e Polígono de Freqüência:

Exemplo:

Notas obtidas na disciplina de Estatística

Notas	fi	Fi
5 |-- 6	18	18
6 |-- 7	15	33
7 |-- 8	12	45
8 |-- 9	03	48
9 |--10	02	50

FONTE: Dados hipotéticos.

Ogiva ou polígono de freqüência acumulada:

Exemplo:

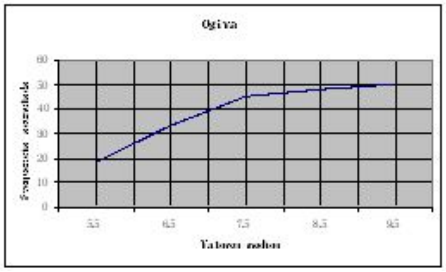
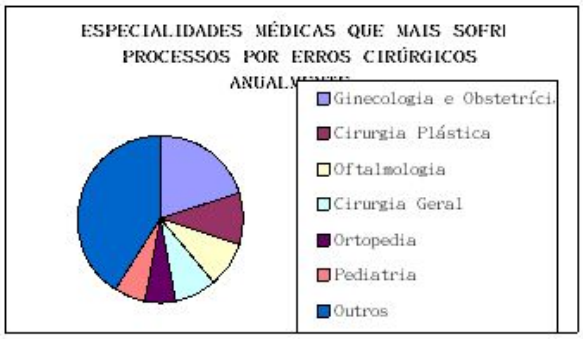


Gráfico em linha: é um dos mais importantes gráficos; representa observações feitas ao longo do tempo.



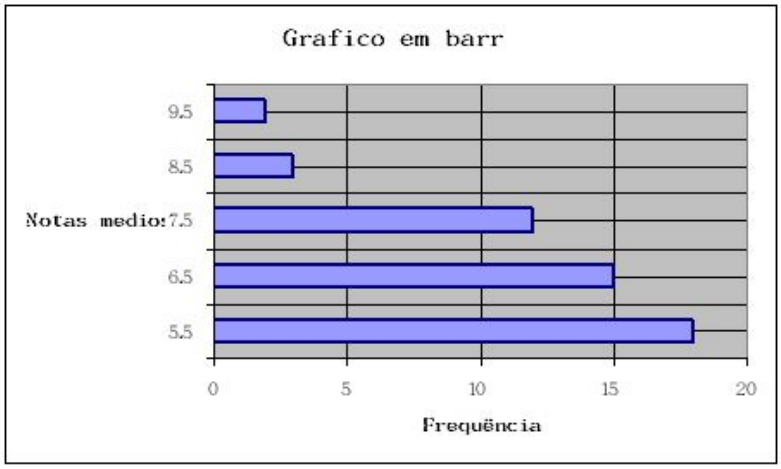
Gráfico em setores: É um gráfico construído no círculo, que é dividido em setores correspondentes aos termos da série e proporcionais aos valores numéricos dos termos da série.

Exemplo:



Gráficos em Barras (ou em colunas). É a representação de uma série por meio de retângulos, dispostos horizontalmente (em barras) ou verticalmente (em colunas).

Quando em *barras*, os retângulos têm a mesma altura e os comprimentos são proporcionais aos respectivos dados.



Quando em *colunas*, os retângulos têm a mesma base e as alturas são proporcionais aos respectivos dados.

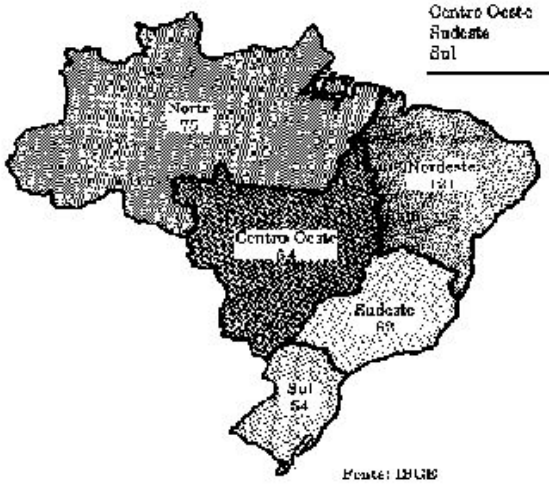


Cartograma. É representação sobre uma carta geográfica.

Este gráfico é empregado quando o objetivo é o de figurar os dados estatísticos diretamente relacionados com as áreas geográficas ou políticas.

Taxa de mortalidade infantil em grupo de 1.000 nascimentos nas regiões brasileiras em 1984

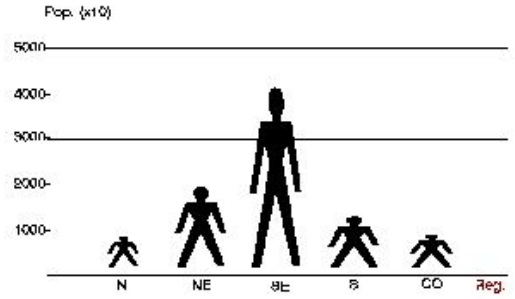
REGIÃO	TAXA
Nordeste	121
Norte	75
Centro Oeste	64
Sudeste	62
Sul	54



Fonte: IBGE

Pictograma. Constitui um dos processos gráficos que melhor fala ao público, pela sua forma ao mesmo tempo atraente e sugestiva. A representação gráfica consta de figuras.

Ex.: População Urbana do Brasil em 1980 (x 10)



Fonte: Anuário Estatístico (1984)

LISTA DE EXERCÍCIOS

1) Construir o Histograma, Polígono de Freqüência e a Ogiva das distribuições dos exercícios 1, 2 e 3 anteriores (pág. 10 e 11).

2) Faça os gráficos para representar os casos a seguir

a)

ACIDENTES DO TRABALHO					
Quantidade mensal de acidentes do trabalho registrados, por motivo - 2001/2003					
MESES	Anos	QUANTIDADE DE ACIDENTES DO TRABALHO REGISTRADOS			
		Total	Motivo		
			Típico	Trajeto	Doença do Trabalho
	2001	**340.251**	**282.965**	**38.799**	**18.487**
TOTAL.....	**2002**	**393.071**	**323.879**	**46.881**	**22.311**
	2003	**390.180**	**319.903**	**49.069**	**21.208**
	2001	28.930	24.277	3.088	1.565
Janeiro.....	2002	29.708	24.624	3.382	1.702
	2003	31.762	26.069	3.835	1.858
	2001	25.842	21.657	2.832	1.353
Fevereiro.....	2002	28.258	23.235	3.357	1.666
	2003	31.166	25.517	3.824	1.825
	2001	31.902	26.584	3.526	1.792
Março.....	2002	32.545	26.796	3.904	1.845
	2003	31.332	25.634	3.807	1.891

	2001	28.354	23.444	3.326	1.584	
Abril.....	2002	35.200	28.863	4.107	2.230	
	2003	30.516	24.847	3.912	1.757	
	2001	31.352	25.963	3.538	1.851	
Maio.....	2002	33.679	27.722	4.037	1.920	
	2003	33.224	27.248	4.221	1.755	
	2001	28.582	23.739	3.254	1.589	
Junho.....	2002	32.598	26.751	4.063	1.784	
	2003	32.173	26.479	4.016	1.678	
	2001	30.558	25.577	3.429	1.552	
Julho.....	2002	34.623	28.555	4.097	1.971	
	2003	31.995	26.308	4.063	1.624	
	2001	29.005	23.995	3.311	1.699	
Agosto.....	2002	36.016	29.747	4.178	2.091	
	2003	32.417	26.507	4.125	1.785	
	2001	24.401	20.317	2.780	1.304	
Setembro.....	2002	33.723	27.773	4.004	1.946	
	2003	35.629	29.307	4.328	1.994	
	2001	27.850	23.141	3.189	1.520	
Outubro.....	2002	36.690	30.234	4.393	2.063	
	2003	37.384	30.801	4.655	1.928	
	2001	27.615	23.005	3.241	1.369	
Novembro.....	2002	32.230	26.765	3.793	1.672	
	2003	33.730	27.605	4.331	1.794	
	2001	25.860	21.266	3.285	1.309	
Dezembro.....	2002	27.801	22.814	3.566	1.421	
	2003	28.852	23.581	3.952	1.319	
FONTE: DATAPREV, CAT.						
NOTA: Os dados são preliminares, estando sujeitos a correções.						

b)

ESTADOS	Nº DE TRANSPLANTES
DF	34
BA	38
ES	56
PE	56
CE	87

PR	181
RJ	181
RS	181
MG	231
SP	756

FONTE: Associação Brasileira de Transplante de Órgãos.

c)

ACIDENTES DO TRABALHO					
Quantidade de acidentes do trabalho registrados, por motivo, segundo as					
Grandes Regiões e Unidades da Federação - 2001/2003					
					(continua)
GRANDES REGIÕES E UNIDADES DA FEDERAÇÃO	Anos	QUANTIDADE DE ACIDENTES DO TRABALHO REGISTRADOS			
		Total	Motivo		
			Típico	Trajeto	Doença do Trabalho
	2001	**340.251**	**282.965**	**38.799**	**18.487**
BRASIL.....	**2002**	**393.071**	**323.879**	**46.881**	**22.311**
	2003	**390.180**	**319.903**	**49.069**	**21.208**
	2001	**11.027**	**9.065**	**1.342**	**620**
NORTE.....	**2002**	**13.002**	**10.603**	**1.521**	**878**
	2003	**13.493**	**11.013**	**1.596**	**884**
	2001	1.853	1.559	262	32
Rondônia.....	2002	1.996	1.665	263	68
	2003	2.064	1.704	286	74
	2001	224	180	42	2
Acre.....	2002	308	245	56	7
	2003	324	267	52	5
	2001	2.537	1.992	281	264
Amazonas.....	2002	3.086	2.443	324	319

	2003	3.473	2.678	393	402
	2001	95	51	40	4
Roraima.....	2002	111	76	31	4
	2003	113	76	28	9
	2001	5.070	4.219	556	295
Pará.....	2002	6.425	5.289	679	457
	2003	6.401	5.363	670	368
	2001	308	251	54	3
Amapá.....	2002	293	225	57	11
	2003	244	196	46	2
	2001	940	813	107	20
Tocantins.....	2002	783	660	111	12
	2003	874	729	121	24
	2001	**27.059**	**20.797**	**3.612**	**2.650**
NORDESTE.....	**2002**	**33.598**	**26.043**	**4.474**	**3.081**
	2003	**35.710**	**27.747**	**4.848**	**3.115**
	2001	1.101	876	171	54
Maranhão.....	2002	1.395	1.098	207	90
	2003	1.772	1.441	253	78
	2001	507	372	111	24
Piauí.....	2002	622	456	138	28
	2003	716	490	184	42
	2001	3.365	2.457	635	273
Ceará.....	2002	3.744	2.771	704	269
	2003	4.084	3.013	788	283
	2001	1.856	1.508	282	66
Rio Grande do Norte.....	2002	2.245	1.821	335	89
	2003	2.509	2.068	328	113
	2001	1.636	1.313	191	132
Paraíba.....	2002	1.789	1.419	248	122

	2003	1.825	1.426	251	148
	2001	5.914	4.583	836	495
Pernambuco.....	2002	7.066	5.462	1.111	493
	2003	7.370	5.628	1.259	483
	2001	2.600	2.328	205	67
Alagoas.....	2002	3.040	2.770	240	30
	2003	3.441	3.152	244	45
	2001	1.147	856	137	154
Sergipe.....	2002	1.614	1.226	184	204
	2003	1.449	1.076	203	170
	2001	8.933	6.504	1.044	1.385
Bahia.....	2002	12.083	9.020	1.307	1.756
	2003	12.544	9.453	1.338	1.753
	2001	**198.118**	**163.791**	**23.161**	**11.166**
SUDESTE.....	**2002**	**228.376**	**186.862**	**28.091**	**13.423**
	2003	**221.603**	**180.194**	**29.059**	**12.350**
	2001	37.423	32.266	3.569	1.588
Minas Gerais.....	2002	38.937	33.183	4.050	1.704
	2003	40.275	34.294	4.491	1.490
	2001	7.446	6.223	851	372
Espírito Santo.....	2002	8.553	7.179	946	428
	2003	8.335	7.074	997	264
	2001	19.313	14.577	3.538	1.198
Rio de Janeiro.....	2002	26.119	19.864	4.500	1.755
	2003	26.288	19.458	4.507	2.323
	2001	133.936	110.725	15.203	8.008
São Paulo.....	2002	154.767	126.636	18.595	9.536
	2003	146.705	119.368	19.064	8.273
	2001	**84.411**	**73.116**	**7.995**	**3.300**
SUL.....	**2002**	**95.007**	**81.534**	**9.430**	**4.043**
	2003	**94.370**	**80.618**	**9.899**	**3.853**

	2001	23.231	20.312	2.224	695
Paraná.....	2002	27.710	23.957	2.847	906
	2003	28.862	24.750	3.154	958
	2001	25.888	22.614	2.663	611
Santa Catarina.....	2002	27.663	23.830	3.117	716
	2003	25.882	22.316	2.907	659
	2001	35.292	30.190	3.108	1.994
Rio Grande do Sul.....	2002	39.634	33.747	3.466	2.421
	2003	39.626	33.552	3.838	2.236
	2001	**19.636**	**16.196**	**2.689**	**751**
CENTRO-OESTE.....	**2002**	**23.088**	**18.837**	**3.365**	**886**
	2003	**25.004**	**20.331**	**3.667**	**1.006**
	2001	4.017	3.441	466	110
Mato Grosso do Sul.....	2002	4.832	4.098	640	94
	2003	5.022	4.290	617	115
	2001	4.252	3.685	484	83
Mato Grosso.....	2002	4.933	4.265	558	110
	2003	5.706	4.786	690	230
	2001	7.724	6.316	1.133	275
Goiás.....	2002	8.834	7.130	1.416	288
	2003	10.034	8.166	1.600	268
	2001	3.643	2.754	606	283
Distrito Federal.....	2002	4.489	3.344	751	394
	2003	4.242	3.089	760	393
FONTE: DATAPREV, CAT.					

Fernando Ribeiro

POPULAÇÃO / AMOSTRA / TÉCNICAS DE AMOSTRAGEM

Muitas vezes é impraticável observar todos os elementos da população, portanto é preciso recorrer a uma parcela representativa dessa população (amostra).
Qualquer que seja a amostra há sempre o risco de chegar a conclusões erradas, mas esse risco diminui à medida que aumenta a quantidade de elementos da amostra. Essa quantidade não deve ser menos de 10% do total de elementos da população. Há métodos de escolha da amostra que devem garantir a representatividade do grupo (população).

I) Amostragem aleatória simples

Ex: Suponhamos uma população de 80 elementos (por exemplo, as alturas de 80 alunos). Numera-se esta população de 01 a 80 que serão submetidos a um sorteio.

Amostra = 10% da população = 8 elementos

Quando o número de elementos da amostra é grande, esse tipo de sorteio torna-se trabalhoso. Adotamos a TNA (tabela de números aleatórios). Para obtermos os elementos da amostra usando a tabela, sorteamos um algarismo qualquer da tabela, da qual iremos considerar números de 2, 3, ou mais algarismos, conforme a necessidade. Os números assim obtidos irão indicar os elementos da amostra.

Por exemplo: utilizamos a 18ª linha da esquerda para a direita

61 02 01 81 73 92 60 66 73 58 53 34

10% = 8 elementos = 61 02 01 73 60 66 58 53

Estudaremos então o 61°, 2°, 1°, 73°, 60°, 66°, 58°, 53°.

Exercício: Alunos de administração do ASSESC. População = X, amostra = 10%.

População de 90 alunos. Analisaremos renda per capita (01 a 90)

-escolher 1ª diagonal da esquerda para a direita na TNA.

-amostra = 10% = 9 elementos

57 88 59 81 16 03 04 20 08 85 05 68 49 22 33 01 73 35

06 98 90 80 63 41 52 46 76 93 22 28 38 97 26 39 26

Amostra = 57, 88, 59, 81,16,03, 04,20, 08.

Exercício: Uma população é formada por 56 notas resultantes de um teste de inteligência (já ordenados).

Determinar uma amostra com 12% (7 elementos). 1ª linha da direita para a esquerda da TNA. Amplitude: maior = 139 menor = 51

62 129 95 123 81 93 105 98 96 80 87 110 139 75

123 60 72 66 108 120 57 113 64 108 90 137 74 106

109 84 121 60 128 100 72 119 63 128 80 99 114 103

77 91 51 100 63 104 76 82 110 63 121 65 75 78

Amostra: 113, 99, 137, 123, 66,82, 91.

II) Amostragem proporcional estratificada : Quando há subdivisão na população = subpopulação = estratos. **Ex:** 90alunos (54 masculinos e 36 femininos). Dois estratos.

Sexo	População	%	Amostra
M	54	60	5
F	36	40	4
Total	90	100	9

1ª parte -

2ª parte - Se numerarmos os 90 alunos, sendo 01 a 54 masculinos, e 55 a 90 femininos. Olhar TNA da esquerda para a direita na última linha.

62 69 84 97 47 23 66 51 56 13 8 69 11 52 75 59 26 86 81

M	47 – 23 – 51 – 13 – 08 (5 elementos)
F	62 – 69- 84 – 66 (4 elementos)

Exercício: 120 alunos. Preencher a tabela. Amostra mínima = 10% = 12 elementos.

Classe	População	%	Amostra
A	20	16,67	2
B	15	12,50	2
C	35	29,16	3
D	30	25,00	3
E	20	16,67	2
Total	120	100,00	12

Exercício: 1831 funcionários. Amostra de 120 elementos.

Empresa	Funcionários Masc.Femin.	% M	% F	Amostra M F
A	80 95	9,13	9,95	5 6
B	102 120	11,64	12,57	7 8

C	110	92	12,56	9,63	7	6
D	134	228	15,30	23,67	9	15
E	150	130	17,12	13,61	10	9
F	300	290	34,25	30,37	19	19
Total	876	955	100,00	100,00	57	63

Amostra: homens = 57 e mulheres = 63. Total = 120 elementos.

III) Amostragem sistemática

Quando os elementos da população já estão ordenados. Utilizado para grandes amostras.
1º passo – ordenar a população
2º passo – arbitrar um número de 1 a 9
3º passo – a amostra serão todos os números terminados com o número arbitrado somando-se de dezena em dezena até o limite do número da população.

Exemplo: 120 alunos

1º passo – 01, 02, 03,04, 05, 06.....120.
2º passo – 3
3º passo – amostra = 3, 13, 23, 33, 43, 53, 63, 73, 83, 93, 103, 113. = 12 elementos.

Exercício: Linha de produção diária = 200 elementos. Fixando amostra de 10% = 20 elementos. Escolhe – se o número 7.

Amostra = 07, 17, 27, 37, 47, 57, 67, 77, 87, 97,107, 117, 127, 137, 147, 157, 167, 177,187,197.

DISTRIBUIÇÃO POR FREQUÊNCIA

Todo fenômeno estatístico pode ser representado por tabelas, quadro de distribuição de freqüência e gráficos.
Distribuição por freqüência: tabela contendo variável e freqüência dos elementos.

I)Distribuição de freqüência sem intervalos de classe

Nº de acidentes	Nº de motoristas	Fr (%)	Fa
0	20	28,57	20
1	10	14,29	30
2	16	22,86	46
3	9	12,85	55
4	6	8,57	61
5	5	7,14	66
6	3	4,29	69
7	1	1,43	70
Total	70	100,00	

Li (limite inferior) = 0 Ls (limite superior) = 7 H (amplitude) = Ls – Li = 7 – 0 = 7

Determinar:

a) Número de motoristas que não sofreram nenhum acidente? 20

b) Número de motoristas que sofreram pelo menos 4 acidentes. 15

c) Número de motoristas que sofreram menos de 3 acidentes? 46

d) Número de motoristas que sofreram no mínimo 3 e no máximo 5 acidentes? 20

e) Porcentagem dos motoristas que sofreram no máximo 2 acidentes. 65,72%.

Exercício: A tabela abaixo representa as vendas diárias de um aparelho elétrico apurado em um mês. 14 12 11 13 14 13

12 14 13 14 11 17
12 14 10 13 15 11
15 13 16 17 14 14

Li = 10 Ls = 17 H = 17 – 10 = 7

Faça o quadro de distribuição de freqüência:

Nº de vendas	F (freqüência)	Fr (%)	Fa
10	1	4,17	1
11	3	12,50	4

12	3	12,50	7
13	5	20,83	12
14	7	29,17	19
15	2	8,33	21
16	1	4,17	22
17	2	8,33	24
Total	24	100,00	

II)Distribuição de freqüência com intervalos de classe.

Exemplo: Empresa com 40 empregados.

S.M (salário mínimo)	F (frequência)
1	4
2	5
3	7
4	5
5	4
6	5
7	1
8	3
9	2
10	1
11	1
12	1
13	1
Total	40

Li = 1 Ls = 13
Número de classes: forma de Sturges

H = 13 – 1 = 12 n = 1 + 3,3 logN
n = 1 + 3,3 log40 = 6,29 => 7
h = H/n = 12/6 = 2

	F	Fr (%)	Fa	Fra
1 ⊢ 3	9	22,50	9	22,50
3 ⊢ 5	12	30,00	21	52,50
5 ⊢ 7	9	22,50	30	75,00
7 ⊢ 9	4	10,00	34	85,00
9 ⊢ 11	3	7,50	37	92,50
11 ⊢ 13	2	5,00	39	97,50
13 ⊢ 15	1	2,5	40	100,00
	40	100,00		

Quantos funcionários ganham até 5 s.m., exclusive? 21
Qual a % de funcionários que recebem até 6 s. m., exclusive?
75%

Exercício: Em uma empresa tomaram-se alturas (cm) de 16 funcionários:

160	152	155	154
163	156	162	161
155	151	158	166
163	167	157	152

Rol (ordenar de forma crescente)

151	155	158	163
152	155	160	163
152	156	161	166
154	157	162	167

Li = 151 Ls = 167
Número de classes: fórmula de Sturges
H = 16
n = 1 + 3,3 logN
n = 1 + 3,3 log16 = 4,97 => 5 classes
h = H/n = 16/5 = 3

	F	Fr (%)	Fa
|___ 151 154	3	18,75	3
___| 154 157	4	25,00	7
___| 157 160	2	12,50	9
|___ 160	3	18,75	12

163			
|___ 163 166	2	12,50	14
|___ 166 169	2	12,50	16
	16	100,00	

Determine:

a) Número de funcionários com altura até 157? 7

b) Número de funcionários com altura de 157 até 166? 7

c) Qual % de funcionários com estatura abaixo de 163? 75%

III)Distribuição por intervalo de classes.

A tabela abaixo indica uma distribuição de freqüência das áreas de lotes.

Áreas (m²)	Lotes (F)	Fr (%)	Fa	PM	Fra
____ | 300 400	14	3,50	14	350	3,50
____ | 400 500	46	11,50	60	450	15,00
____ |500 600	58	14,50	118	550	29,50

____ | 600 700	76	19,00	194	650	48,50
____ | 700 800	68	17,00	262	750	65,50
|____800 900	62	15,50	324	850	81,00
____ | 900 1000	48	12,00	372	950	93,00
___ | 1000 1100	22	5,50	394	1050	98,50
|___ 1100 1200	6	1,50	400	1150	100,00
Total	400	100,00			

Responda:

a) A amplitude total: 1200 – 300 = 900

b) O Ls da 5ª classe: 800

c) O Li da 8ª classe: 1000

d) O ponto médio da 7ª classe: 950

e) Amplitude da 2ª classe: 100

f) A freqüência da 4ª classe: 76

g) A Fr da 6ª classe: 15,50

h) A Fa da 5ª classe: 262

i) O número de lotes cuja área não atinja 700 m² : 194

j) O número de lotes cuja área atinja e ultrapasse 800 m²: 400 – 262 = 138

k) A percentagem dos lotes cuja Fra não atinja 600 m²: 29,50%

l) A percentagem dos lotes cuja área seja maior ou igual a 900 m² : 19,00%

m) A percentagem dos lotes cuja área é de 500 m², no mínimo, mas inferior a 1000 m²: 78%

n) A classe do 72º lote: 3ª classe

o) Até que classe estão incluídos 60% dos lotes? 5ª classe.

Histograma: gráfico de barras.

GRÁFICO

DISTRIBUIÇÃO DE CLASSES. GRÁFICOS.

Altura funcionários da empresa	F	Fr (%)	Fa	Fra %	Pm
|___ 150 155	6	15,00	6	15,00	152,50
___| 155 160	10	25,00	16	40,00	157,50
___| 160 165	15	37,50	31	77,50	162,50
|___	5	12,50	36	90,00	167,50

165 170					
|___ 170 175	3	7,50	39	97,50	172,50
|___ 175 180	1	2,50	40	100,00	177,50
	40	100,00			

Histograma: gráfico de áreas retangulares (barra) onde no plano cartesiano o eixo das abcissas (A é a amplitude de classes e B as ordenadas) é a frequência.

GRÁFICO

Ogiva: gráfico cujas ordenadas (y) é a Fa.
(ref. Ls e freqüência).

GRÁFICO

Distribuição de freqüências- Gráficos

A tabela contém uma amostra de 50 elementos de uma população de indústrias de S.C., representativa dos coeficientes de liquidez.

3,9	7,4	10,0	11,8	2,3	4,5	10,5
8,4	15,6	7,6				
18,8	2,9	2,3	0,4	5,0	9,0	5,5
9,2	12,4	8,7				
4,5	4,4	10,6	5,6	8,3	2,4	17,8
11,6	0,8	4,4				
7,1	3,2	2,7	16,2	2,7	9,5	13,1
3,8	6,3	7,9				

4,8 5,3 12,9 6,9 6,3 7,5 2,6
3,3 4,6 16,0

ROL
0,4 2,6 3,3 4,5 5,3 6,9 7,9
9,2 11,6 15,6
0,8 2,7 3,8 4,5 5,5 7,1 8,3
9,5 11,8 16,0
2,3 2,7 3,9 4,6 5,6 7,4 8,4
10,0 12,4 16,2
2,3 2,9 4,4 4,8 6,3 7,5 8,7
10,5 12,9 17,8
2,4 3,2 4,4 5,0 6,3 7,6 9,0
10,6 13,1 18,8

a) Forme uma distribuição com intervalos de classe iguais a 3 (h=3), tais que os limites inferiores sejam múltiplos de 3.

Classe	F	Fr (%)	Fa	Fra %	Pm
|___ 0 3	9	18,00	9	18,00	1,50
___ | 3 6	14	28,00	23	46,00	4,50
___| 6 9	11	22,00	34	68,00	7,50
|__ 9 12	8	16,00	42	84,00	10,50
|___ 12 15	3	6,00	45	90,00	13,50
|___ 15	4	8,00	49	98,00	16,50

18					
___ | 18 21	1	2,00	50	100,00	19,5
	50	100,00			

Li= 0,4
n=10 + 3,3 log 50 h= H/n = 2,63 ~= 3
Ls= 18,8
n=6,6 ~=7 classes
H= 18,4 (amplitude da amostra)

Em uma fábrica foram testadas 400 lâmpadas. A duração delas aparece na tabela abaixo.

Duração (horas)	Nº lâmpadas (F)	Fr (%)	Fa	Fra
____ | 300 400	14	3,50	14	3,50
____ | 400 500	46	11,50	60	15,00
____ |500 600	58	14,50	118	29,50

____ | 600 700	76	19,00	194	48,50
____ | 700 800	68	17,00	262	65,50
|____800 900	62	15,50	324	81,00
____ | 900 1000	48	12,00	372	93,00
___ | 1000 1100	22	5,50	394	98,50
|__ 1100 1200	6	1,50	400	100,00
Total	400	100,00		

Responda:

1) Qual a amplitude total da distribuição? 900

2) Qual o ponto médio da 5ª classe? 750 horas

3) Qual a freqüência relativa da 6ª classe? E o que significa esse valor? 15,50%

4) Qual a percentagem de lâmpadas com durabilidade máxima de 500 horas? 15%

5) Qual a percentagem de lâmpadas com durabilidade de 900 horas ou mais? 19,00%

6) Qual a percentagem de lâmpadas com durabilidade entre 500 (inclusive) e 800 (exclusive) horas? 50%

7) Indique graficamente a classe modal (moda=valor de maior freqüência) (histograma), e a mediana (valor do meio da posição da distribuição (Fra x classes).

GRÁFICOS

Exercício: Distribuição por frequência sem intervalo de classes.

Seja as notas de estatística da turma:

3 7 6 6 5 4 7 5
7 3 4 5 6 8 4 5
4 5 5 6 7 6 6 6
3 3 3 4 5 5 5 6

Montar a tabela de distribuição:

ROL:

3 3 3 3 3 4 4 4
4 4 5 5 5 5 5 5
5 5 5 6 6 6 6 6
6 6 6 7 7 7 7 8

Nota	F (freqüência)	Fr (%)	Fa	Fra (%)
3	5	15,63	5	15,03
4	5	15,63	10	31,26
5	9	28,12	19	59,38
6	8	25,00	27	84,38
7	4	12,5	31	96,88

8	1	3,12	32	100,00
Total	32	100,00		

1) Esboce graficamente a moda e mediana. Medidas de posição:

A - Média aritmética = valor de tendência da distribuição.
B – Moda= valor que aparece com a maior freqüência.
C – Mediana = valor que se encontra no meio da distribuição.

GRÁFICOS

2) Qual a percentagem de alunos com notas superiores a 5 e menores que 9? 40,62%

3) % de alunos com nota 7? 12,50%

4) % de alunos que tomaram pau (nota<=3)? 15,63%

5) % de alunos com notas entre 6 e 7, inclusive? 37,50%

MEDIDAS ESTATÍSTICAS

Estudaremos dois tipos fundamentais de medidas estatísticas: medidas de tendência central e medidas de dispersão.

As medidas de tendência central mostram o valor representativo em torno do qual os dados tendem a agrupar-se, com maior ou menor freqüência. São utilizadas para sintetizar em um único número o conjunto de dados observados.

As medidas de dispersão mostram o grau de afastamento dos valores observados em relação àquele valor representativo.

MEDIDAS DE TENDÊNCIA CENTRAL

A <u>média aritmética simples</u>

A média aritmética simples de um conjunto de valores é o valor obtido somando-se todos eles e dividindo-se o total pelo número de valores. É denotada por $\bar{x}$ (leia-se “x barra”)

$\bar{x} = \frac{\sum x}{n}$, onde x são os valores observados.

$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i}$, se os dados estiverem organizados em distribuição de freqüência.

Onde x_i e f_i são os valores do ponto médio e da freqüência absoluta da classe i-ésima respectivamente.

Exemplos:

1º) Calcule a média aritmética dos valores abaixo:

a. X = {0, 6, 8, 7, 4, 6}
b. Y = {25, 16, 29, 19, 17}
c. Z = {105, 123, 98, 140}

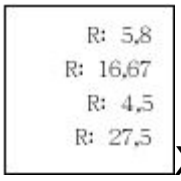
2º) Encontre a média para o salário destes funcionários. Preencher a tabela.

Salários semanais para 100 operários não especializados

Salários semanais	f_i	x_i	$x_i.f_i$
140 |-- 160	7		
160 |-- 180	20		
180 |-- 200	33		
200 |-- 220	25		
220 |-- 240	11		
240 |-- 260	4		
□	100		

Exercícios:

1) Encontre a média dos seguintes conjuntos de observações.



a) X = {2, 3, 7, 8, 9}.
b) Y = {10, 15, 22, 18, 25, 16}.
c) Z = {1, 3, 6, 8}.
d) T = {1, 3, 6, 100}.

2) Encontre a média das notas na disciplina de Estatística.
Notas obtidas

Notas	f_i
5 |-- 6	18
6 |-- 7	15
7 |-- 8	12

8 |-- 9	03
9 |--10	02

FONTE: Dados hipotéticos.

Resp 6,62.

A <u>mediana</u>

<u>A **mediana**</u> é um valor central de um rol, ou seja, a mediana de um conjunto de valores ordenados (crescente ou decrescente) é a medida que divide este conjunto em duas partes iguais.

Exemplo: Calcule a mediana dos conjuntos abaixo:

a- X={3, 7, 4, 12, 15, 10, 18, 14}

b- Y={29, 33, 42, 38, 31, 34, 45, 51, 95}

c- Z={29, 33, 42, 38, 31, 34, 45, 120, 95}

A moda

Seja X um conjunto de dados estatísticos. Define-se *Moda* de X, denotada por *Mo* como sendo o elemento mais frequente no conjunto.

Um conjunto de dados pode ter:

- Nenhuma moda (amodal);
- Uma moda (unimodal);
- Duas ou mais modas (multimodal).

Exercícios: Calcule a moda para os conjuntos abaixo:

a) X= {2, 3, 4, 3, 7, 8, 9, 14}.

b) Y= {2, 4, 6, 2, 8, 4, 10}.

c) Z= {32, 56, 76, 4, 8, 97}.

OBSERVAÇÕES:

Não há regra para se dizer qual a melhor medida de tendência central. Em cada situação específica o problema deve ser analisado pelo estatístico, que concluirá pela medida mais adequada a situação. Assim é que:

a) A **MA** é a medida mais adequada quando não há valores erráticos ou aberrantes.

b) A **mediana** deve ser usada sempre que possível como medida representativa de distribuições com valores dispersos, como distribuição de rendas, folhas de pagamentos, etc.

Exercícios:

1) Dados os conjuntos abaixo, calcule a média aritmética, mediana e moda.

$\bar{x}$	4,4	9,3	10,3	6,8
Md	4	8,5	10	6,5
Mo		6	5	

A = {3, 5, 2, 1, 4, 7, 9}.
B = {6, 12, 15, 7, 6, 10}.
C = {10, 5, 11, 8, 15, 4, 16, 5, 20, 6, 13}.
D = {4, 4, 10, 5, 8, 5, 10, 8}.

2) Calcule a média aritmética das distribuições de freqüências dos exercícios 1 e 2 das páginas 11. **Resp. 1) R$ 151,79; 2) 173,53 cm e 68,15 kg.**

MEDIDAS DE DISPERSÃO

Servem para verificarmos a representatividade das medidas de posição, pois é muito comum encontrarmos séries que, apesar de terem a mesma média, são compostas de maneira distinta.

Assim, para as séries:

a) 25, 28, 31, 34, 37
b) 17, 23, 30, 39, 46

temos $\bar{x}_a = \bar{x}_b = 31$.

Nota-se que os valores da série “a” estão mais concentrados em torno da média 31, do que a série “b”. Precisamos medir a dispersão dos dados em torno da média, para isto utilizaremos as medidas de dispersão:

Desvio Padrão
Coeficiente de Variação

Desvio Padrão:

É a raiz quadrada positiva da média aritmética dos quadrados das diferenças entre cada valor e a média aritmética do conjunto e é denotada por σ. Assim,

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$$, se os dados estiverem organizados em distribuição de freqüência.

Exemplo 1:

Encontre o desvio padrão para os dados das séries a), e b) acima.

Exemplo 2:

Salários semanais para 100 operários não especializados

Salários semanais	f_i	x_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
140 |-- 160	7			
160 |-- 180	20			
180 |-- 200	33			
200 |-- 220	25			
220 |-- 240	11			
240 |-- 260	4			
□	100			

Encontre o desvio padrão para o salário destes funcionários.

Exercício:
Calcule o desvio padrão das distribuições de frequências dos exercícios 1 e 2 das páginas anteriores.

Coeficiente de variação:

Trata-se de uma medida de dispersão, útil para a compreensão em termos relativos do grau de concentração em torno da média de séries distintas. É dado por:

$$C_v = \frac{\sigma}{\bar{x}}.100$$

Exemplo 4:

Para duas emissões de ações ordinárias da indústria eletrônica, o preço médio diário, no fechamento dos negócios, durante um período de um mês, para as ações **A**, foi de R$ 150,00 com um desvio padrão de R$ 5,00. Para as ações **B**, o preço médio foi de R$ 50,00 com um desvio padrão de R$ 3,00. Em relação ao nível do preço, qual dos tipos de ações é mais variável?

Exercícios.

1) Uma amostra de 20 operários de uma companhia apresentou os seguintes salários recebidos durante uma certa semana, arredondados para o valor mais próximo e apresentados em ordem crescente: 140, 140, 140, 140, 140, 140, 140, 140, 155, 155, 165, 165, 180, 180, 190, 200, 205, 225, 230, 240. Calcular (a) a média, (b) a mediana, (c) a moda, (d) o desvio padrão, (e) o coeficiente de variação, para este grupo de salários. **R:** a) 170,5; d) 33,12.

2) O número de carros vendidos por cada um dos vendedores de um negócio de automóveis durante um mês particular, em ordem crescente: 2, 4, 7, 10, 10, 10, 12, 12, 14, 15. Determinar (a) a média, (b) a mediana, (c) a moda, (d) o desvio padrão **R:** a) 9,6; d) 3,95.

3) Em conjunto com uma auditoria anual, uma firma de contabilidade pública anota o tempo necessário para realizar a auditoria de 50 balanços contábeis. Calcular (a) a média, (b) o desvio padrão, para o tempo de auditoria necessário para esta amostra de registro. **R:** a) 43,2; b)12,28.

Tempo necessário para a auditoria de balanços contábeis.

Tempo de auditoria. (min.)	Nº de balanços. (f_i)
10 |-- 20	3
20 |-- 30	5
30 |-- 40	10
40 |-- 50	12
50 |-- 60	20
Total	50

4) Os salários semanais de 50 funcionários de um hospital, em reais, foram os seguintes:

100	122	130	140	152	160	164	176	180	188	192	200	216
104	126	134	146	156	160	170	176	184	190	194	200	218
116	128	138	150	156	162	170	178	186	190	196	200	
120	128	140	150	156	162	176	180	186	192	196	210	

a) Construa uma distribuição de freqüências, com h = 20 e limite inferior para a primeira classe igual a 100.
b) Quantos funcionários tem um salário semanal situado entre R$ 120,00 (inclusive) e R$ 160,00 (exclusive)? **17 funcionários**
c) Que porcentagem de funcionários tem um salário semanal situado entre R$ 180,00 (inclusive) e R$ 200,00 (exclusive)?**26%**
d) Qual o salário médio semanal destes funcionários utilizando o item a)? **166,4**
e) Determine o desvio padrão e o coeficiente de variação da distribuição. **28,76; 17,28%**

5) A distribuição das alturas de um grupo de pessoas apresentou uma altura média de 182 cm e um desvio padrão de 15 cm, enquanto que a distribuição dos pesos, apresentou um peso médio de 78 kg, com um desvio padrão de 8 kg. Qual das duas distribuições apresentou maior dispersão? Por quê?

CORRELAÇÃO E REGRESSÃO

Introdução:

Já trabalhamos com a descrição de valores de uma única variável. Quando, porém, consideramos observações de duas ou mais variáveis surge um novo problema: as **relações** que podem existir entre as variáveis estudadas.

Assim, quando consideramos variáveis como peso e estatura de um grupo de pessoas, uso do cigarro e incidência do câncer, procuramos verificar se existe alguma relação entre as variáveis de cada um dos pares e qual dessa relação.

Uma vez caracterizada a relação, procuramos descrevê-la através de uma função matemática. A **regressão** é o instrumento adequado para determinação dos parâmetros dessa função. Se todos os valores das variáveis satisfazem exatamente uma equação, diz-se que elas estão *perfeitamente correlacionadas* ou que há *correlação perfeita* entre elas.

Quando estão em jogo somente duas variáveis, fala-se em correlação e regressão simples. Quando se trata de mais de duas variáveis, fala-se em correlação e regressão múltipla.

Diagrama de Dispersão

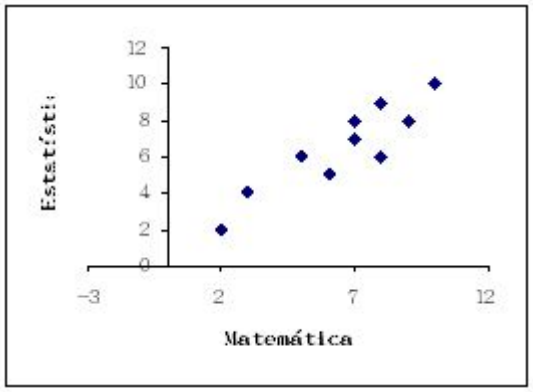
Para desenhar um diagrama de dispersão, primeiro traça-se o sistema de eixos cartesianos. Depois se representa uma das variáveis no eixo “x” e a outra no eixo “y” Colocam-se, então os valores das variáveis sobre os respectivos eixos e marca-se um ponto para cada par de valores. Consideremos uma amostra aleatória, formada por dez dos 98 alunos de uma classe da Universidade A e pelas notas obtidas por eles em Matemática e Estatística:

Fernando Ribeiro

Nº

Notas		
Matemática (X)	**Estatística (Y)**	
01	5,0	6,0
02	8,0	9,0
03	7,0	8,0
04	10,0	10,0
05	6,0	5,0
06	7,0	7,0
07	9,0	8,0
08	3,0	4,0
09	8,0	6,0
10	2,0	2,0

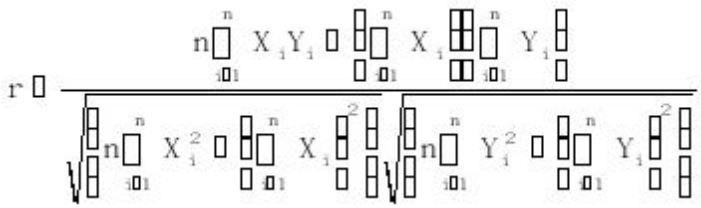
Representando, em um sistema cartesiano coordenado cartesiano ortogonal, os pares ordenados (x,y), obtemos uma nuvem de pontos que denominamos **diagrama de dispersão.** Esse diagrama nos fornece uma idéia grosseira, porém útil da correlação existente:



DEFINIÇÃO 1: **Correlação**

Dizemos que duas ou mais variáveis expressam a relação de causa e efeito ou se elas variam concomitantemente, são variáveis consideradas correlacionadas.

O grau de relacionamento para dados amostrais é dado pela seguinte expressão:



Onde: **n** é o número de observações;
r é o coeficiente de correlação linear para uma amostra.

EXEMPLO 1: Encontre o coeficiente de correlação para os dados da tabela anterior.

(X)	(Y)	XY	X^2	Y^2
5	6	30	25	36
8	9	72	64	81
7	8	56	49	64
10	10	100	100	100
6	5	30	36	25
7	7	49	49	49
9	8	72	81	64
3	4	12	9	16
8	6	48	64	36
2	2	4	4	4
65	65	473	481	475



PROPRIEDADE DO COEFICIENTE DE CORRELAÇAO LINEAR r.

1. O valor de r está sempre entre –1 e 1.
2. O valor de r não varia se todos os valores de qualquer uma das variáveis são convertidos para uma escala diferente.
3. O valor de r não é afetado pela escolha de x ou y.
4. r mede a intensidade, ou grau, de um relacionamento linear. Não serve para medir a intensidade de um relacionamento não-linear.

CORRELAÇÃO POSITIVA E CORRELAÇÃO NEGATIVA

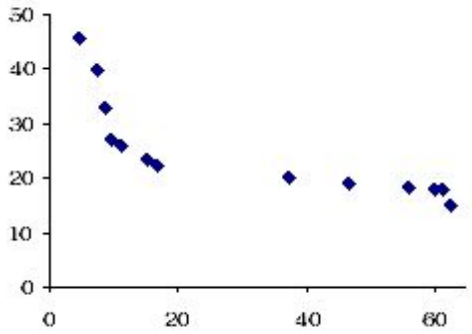
Se as variáveis x e y crescem no mesmo sentido, isto é, quando x cresce, y também cresce, diz-se que as duas variáveis têm correlação positiva.

Então, notas de matemática e notas de estatística dos alunos tem correlação positiva, porque quando uma das variáveis cresce, a outra , em média, também cresce.

Se as variáveis x e y variam em sentido contrário, isto é, quando x cresce, em média y decresce, diz-se que as duas variáveis têm correlação negativa. Observe os dados da Tabela abaixo:

Consumo individual de proteínas de origem animal, em gramas, e coeficiente de natalidade, em 14 países, 1961.

País	Consumo de proteínas	Coef. de natalidade
Formosa	4,7	45,6
Malásia	7,5	39,7
Índia	8,7	33,0
Japão	9,7	27,0
Iugoslávia	11,2	25,9
Grécia	15,2	23,5
Itália	15,2	23,4
Bulgária	16,8	22,2
Alemanha	37,3	20,0
Irlanda	46,7	19,1
Dinamarca	56,1	18,3
Austrália	59,9	18,0
Estados Unidos	61,4	17,9
Suécia	62,6	15,0



Eixo x = consumo de proteínas

Eixo y= coeficiente de natalidade

ANÁLISE DE REGRESSÃO

Muitas vezes é de interesse estudar-se um elemento em relação a dois ou mais atributos ou variáveis simultaneamente.

Nesses casos presume-se que pelo menos duas observações são feitas sobre cada elemento da amostra. A amostra consistirá, então, de pares de valores, um valor para cada uma das variáveis, designadas, X e Y. Um indivíduo “i” qualquer apresenta o par de valores (X_i; Y_i). O objetivo visado quando se registra pares de valores (observações) em uma amostra, é o estudo das relações entre as variáveis X e Y.

Para a análise de regressão interessam principalmente os casos em que a variação de um atributo é sensivelmente dependente do outro atributo.

O problema consiste em estabelecer a função matemática que melhor exprime a relação existente entre as duas variáveis. Simbolicamente a relação é expressa por uma equação de regressão e graficamente por uma curva de regressão.

REGRESSÃO LINEAR SIMPLES

Modelo: $Y_i = \square + \square x_i$

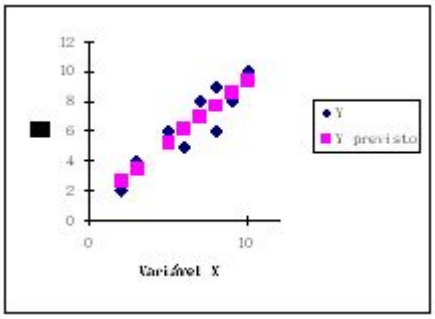
Pressuposições:

a) A relação entre X e Y é linear (os acréscimos em X produzem acréscimos proporcionais em Y e a razão de crescimento é constante).
b) Os valores de X são fixados arbitrariamente (X não é uma variável aleatória).
c) Y é uma variável aleatória que depende entre outras coisas dos valores de X.

Estimativas dos Parâmetros a e b

As estimativas dos parâmetros dadas por “a” e “b”, serão obtidas a partir de uma amostra de n pares de valores (x_i, y_i) que correspondem a n pontos no diagrama de dispersão.
Exemplo:

(X)	(Y)
5	6
8	9
7	8
10	10
6	5
7	7
9	8
3	4
8	6
2	2



Obtemos então: $\hat{y}_i = ax_i + b$
Para cada par de valores (x_i, y_i) podemos estabelecer o desvio: $e_i = y_i - \hat{y}_i$ = y_i-(ax_i + b)

Método dos Mínimos Quadrados

O método dos mínimos quadrados consiste em adotar como estimativa dos parâmetros os valores que minimizem a soma dos quadrados dos desvios.

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - ax_i - b]^2$$

S = f(a, b)

Essa soma, função de “a” e de “b”, terá mínimo quando suas derivadas parciais em relação a “a” e “b” forem nulas.

Para facilitar a escrita, considera-se $\sum_{i=1}^{n}$ $\equiv \sum$

$$\frac{\delta z}{\delta b} = -2\sum [y_i - ax_i - b](1) = 0$$

$$\frac{\delta z}{\delta a} = -2\sum [y_i - ax_i - b](x_i) = 0$$

$$\sum [y_i - ax_i - b] = 0$$

$$\sum [y_i - ax_i - b](x_i) = 0$$

$$\sum y_i - a\sum x_i - nb = 0$$

$$\sum x_i y_i - a\sum x_i^2 - b\sum x_i = 0$$

$$b = \frac{\sum y_i - a\sum x_i}{n}$$

$$\sum x_i y_i - b\sum x_i - a\sum x_i^2 = 0$$

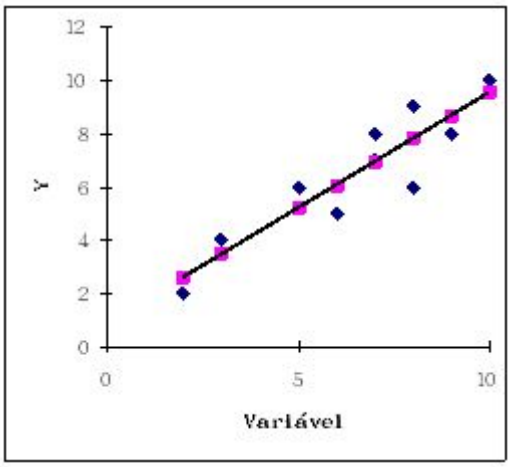
Resolvendo-se esse sistema, obtemos as estimativas para o cálculo de:

$$a = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}$$

e a partir da 1º equação $b = \bar{y} - a\bar{x}$

No exemplo:

(X)	(Y)	X.Y	X^2	Y^2
5	6	30	25	36
8	9	72	64	81
7	8	56	49	64
10	10	100	100	100
6	5	30	36	25
7	7	49	49	49
9	8	72	81	64
3	4	12	9	16
8	6	48	64	36
2	2	4	4	4
65	65	473	481	475



$$a = \frac{10.473 - 65.65}{10.481 - 65^2} = \frac{505}{585} = 0,8632$$

$$b = \frac{65}{10} - 0,8632.\frac{65}{10} = 0,8892$$

$$\hat{y}_i = 0,8632x_i + 0,8892$$

EXERCÍCIOS

Nos Exercícios 1-10,

a) Determine o coeficiente de correlação.
b) Determine a equação da reta de regressão.

1. A tabela apresenta dados de amostra referentes ao número de horas de estudo fora de classe para determinados alunos de um curso de estatística, bem como os graus obtidos em um exame aplicado no fim do curso.

Estudante	1	2	3	4	5	6	7	8
Horas de estudo	20	16	34	23	27	32	18	22
Grau no exame	64	61	84	70	88	92	72	77

c) Estimar o grau no exame obtido por um estudante que dedicou 30 horas fora de classe.

2. A tabela mostrada relaciona os números x de azulejos e os custos y (em dólares) de sua ajustagem e colocação.

x	1	2	3	5	6
y	5	8	11	17	20

c) Para x = 4, ache $\hat{y}$, o valor predito de y.

3. Os dados emparelhados que se seguem consistem no perímetro torácico (em cm) e dos pesos (em quilogramas) de uma amostra de ursos machos.

X Tórax	66	114	137	124	104	124	112	48
Y Peso	40	155	187	157	118	160	149	76

c) Para um urso com perímetro torácico de 132 cm, ache $\hat{y}$, o peso predito.

4. Os dados da tabela abaixo consistem nos pesos (em quilogramas) de plástico descartado e tamanhos de residências.

Plástico (Kg.)	0,12	0,63	0,99	1,27	0,98	0,81	0,38	1,37
Tam. da residência	2	3	3	6	4	2	1	5

c) Ache o tamanho predito de uma residência que descarta 1,12 Kg. de plástico.

5. A tabela abaixo apresenta os pesos totais (em Kg) de lixo descartado e tamanhos de residências.

Peso total	4,8	8,8	12,4	17,1	12,6	9,9	9,8	22,2	15,0	16,0
Tam da Residência	2	3	3	6	4	2	1	5	6	4

c) Ache o tamanho predito de uma residência que descarta 9,0 Kg. de lixo.

6. Os dados seguintes foram obtidos da altura (polegadas) e do peso (libras) de mulheres nadadoras.

Altura	68	64	62	65	66
Peso	132	108	102	115	128

c) Estimar o peso de uma mulher, que possui 67 polegadas.

7. Os dados seguintes mostram o gasto com mídia (milhões de dólares) e as vendas de caixas (milhões) para sete grandes marcas de refrigerantes.

Marca	**Gastos com mídia (US$)**	**Vendas de caixas**
Coca-Cola	131,3	1929,2
Pepsi-Cola	92,4	1384,6
Coca-Cola Light	60,4	811,4
Sprite	55,7	541,5
Dr. Pepper	40,2	536,9
Mountain Dew	29,0	535,6
7- Up	11,6	219,5

Fonte: Superbrands ’98, 20 de outubro de 1997

c) Estimar as vendas, sabendo que foi gasto US$ 80,0 com mídia.

8. Os dados a seguir são a média das notas x e salários mensais y de estudantes que obtiveram bacharelado em administração com ênfase em sistemas de informação.

Média das Notas	2,6	3,4	3,6	3,2	3,5	2,9
Salário Mensal (є)	2800	3100	3500	3000	3400	3100

c) Supondo que a nota de um estudante de bacharelado em administração com ênfase em sistemas de informação seja 8,0. Estime qual será seu salário mensal.

9. Um gerente de vendas reuniu os seguintes dados considerando os anos de experiência e as vendas anuais.

Vendedor	Anos de experiência	Vendas anuais (US$ 1.000)
1	1	80
2	3	97
3	4	92
4	4	102
5	6	103
6	8	111
7	10	119
8	10	123
9	11	117
10	13	136

c) Estimar as vendas anuais, supondo que um vendedor tenha 9 anos de experiência.

10. Dados sobre os gastos com publicidade e faturamento para a Joferi equipamentos, são apresentados a seguir.

Gastos com publicidade	*Faturamento*
1	19
2	32
4	44
6	40
10	52
14	53
20	54

c) Sabendo que os gastos com publicidade foi de ϵ 7.000,00. Quanto espera ganhar a Joferi equipamentos?

PROBABILIDADE

1.1. INTRODUÇÃO

Encontramos na natureza dois tipos de fenômenos: *determinísti aleatórios*.

Os fenômenos *determinísticos* são aqueles em que os resultado sempre os mesmos, qualquer que seja o número de ocorrências.

Nos fenômenos aleatórios, os resultados não serão previsíveis, mesm haja um grande número de repetições do mesmo fenômeno.

Nos experimentos aleatórios, mesmo que as condições iniciais seja mesmas, os resultados finais de cada tentativa do experimento, serão dife e não previsíveis, por isso, é conveniente dispormos de uma medida p estudo de tais situações. Esta medida é a ***probabilidade.***

1.2. EXPERIMENTO ALEATÓRIO. ESPAÇO AMOSTRAL. EVENT

Antes de passarmos à definição de probabilidade, é necessário fixarm conceitos de experimento, espaço amostral e evento.

Um *experimento aleatório* é o processo de coleta de dados relativos fenômeno que acusa variabilidade em seus resultados.

EXEMPLOS:

- **a)** lançamento de uma moeda honesta;
- **b)** lançamento de um dado;
- **c)** determinação da vida útil de um componente eletrônico;

Espaço amostral é o conjunto de todos os resultados possíveis d experimento. Vamos denotá-lo por □.

EXEMPLOS:

1) No caso do lançamento de um dado, □ =

2) Uma lâmpada é ligada e observada até queimar anotando-tempos decorridos, □ =

Quando o espaço amostral consiste em um número finito ou i numerável de eventos, é chamado espaço amostral discreto; e quando for os números reais de determinado intervalo, é um espaço amostral contínuo

Um *evento* é um subconjunto de um espaço amostral

EXEMPLO: Nos exemplos anteriores 1 e 2. Qual seria um possível e para cada um dos exemplos?

1.3. DEFINIÇÕES DE PROBABILIDADE

Seja "**A**" um evento de um experimento aleatório, definimos a *probabi* de "**A**", denotada por **P(A)**,

$P(A) = \frac{\text{Número de casos favoráveis}}{\text{Número de casos possíveis}}$ que é a definição clássica de probabilidade.

EXEMPLO: Na jogada de um dado, qual a probabilidade de aparecer f ou face 5?

Solução:

EXEMPLO: Consideremos o experimento que consiste em lançar moeda 15 vezes. Suponhamos que o número de caras obtido tenha sid Determine a probabilidade do evento cara:

Solução:

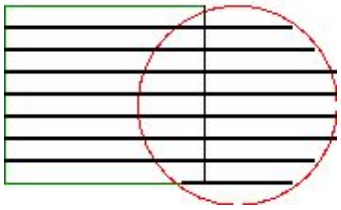
1.4. OPERAÇÕES COM EVENTOS ALEATÓRIOS

Consideremos um espaço amostral finito □. Sejam **A** e **B** dois event □. As seguintes operações são definidas.

a) UNIÃO

O evento união de **A** e **B** equivale à ocorrência de **A**, ou de **B**, ambos. Contém os elementos do espaço amostral em que estão em pelo r

um dos dois conjuntos. Denota-se por **A□B**. A área hachurada da figura a ilustra a situação.

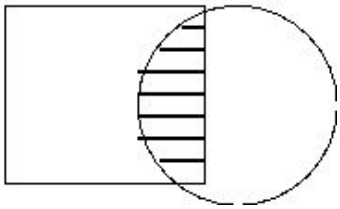


EXEMPLO: Se **A** é o conjunto dos alunos de um Estabelecimento frequentam o curso de Contabilidade e **B** é o conjunto de alunos do m estabelecimento que fazem Ciência da Computação, então:

A□B =

b) INTERSECÇÃO

O evento intersecção de dois eventos **A** e **B** equivale à ocorrênc ambos. Contém todos os pontos do espaço amostral comuns a **A** e a **B**. De se por **A□B**. A intersecção é ilustrada pela área hachurada do diagrama ab



EXEMPLO: Seja **A** o conjunto de alunos de uma Instituição que freqü o 2° grau, e **B** o conjunto dos que freqüentam um curso facultativ interpretação musical. A interseção **A□B** é dada por:

A□B =

c) EXCLUSÃO

Dois eventos **A** e **B** dizem-se mutuamente exclusivos ou mutuam excludentes quando a ocorrência de um deles impossibilita a ocorrência do Os dois eventos não têm nenhum elemento em comum. Exprime-se escrevendo **A□B = □**. O diagrama a seguir ilustra esta situação.

EXEMPLO: Na jogada de um dado, seja **A** o evento “aparece número
B o evento “aparece número ímpar”. Então A□B =

d) ___*NEGAÇÃO*

A negação do evento, denotada por A é chamada evento complen
de A.

EXEMPLO: Se, na jogada de um dado, o evento **A** consist
aparecimento de face par, seu complementar é dado por: $\overline{A}$ □

REGRAS BÁSICAS

Se A e B são dois eventos do espaço amostral □, então valem as seg
regras básicas:

- □ 0 □ P(A) □ 1

P(A) = 0 o evento é impossível e P(A) = 1 o evento é certo.

- □ P(□) = 1
- □ Se A e B são eventos mutuamente excludentes, **A**□**B =** □, e

P(A□B) = P(A) + P(B).

- □ Se **A**□**B** □ □, então: P(A□B) = P(A) + P(B) – P(A□B).
- □ _P(A) = 1- P(A).
- □ Se □ é o vazio, então P(□) =0.

EXERCÍCIO: Consideremos os alunos matriculados na disciplin
Estatística. Temos _____ homens com mais de 25 anos, _____ homens
menos de 25 anos, ____ mulheres com mais de 25 anos, ____ mulheres
menos de 25 anos. Uma pessoa é escolhida ao acaso dentre os ___
seguintes eventos são definidos:

A: a pessoa tem mais de 25 anos;
C: a pessoa é um homem;
B: a pessoa tem menos de 25 anos;
D: a pessoa é uma mulher.
Calcular: P(B□D) e P(A□C).

EXERCÍCIOS

1. Quais dos valores abaixo não podem ser probabilidades?
   0; $\sqrt{2}$; 0,001; -0,2; 3/2; 2/3.

2. Um estudo de 500 vôos da American Airlines selecionados aleatoriamente mostrou que 430 chegaram no horário (com base em dados do Ministério dos transportes). Qual é a probabilidade de um vôo da American Airlines chegar no horário?

3. Em uma pesquisa entre estudantes de uma faculdade, 1162 afirmaram que "colaram" nos exames, enquanto 2468 afirmaram não "colar". Selecionado aleatoriamente um desses estudantes, determine a probabilidade de ele ou ela ter "colado" em um exame.

4. A MasterCard International efetuou um estudo de fraudes em cartões de créditos; os resultados estão agrupados na tabela a seguir.

Tipo de fraude	**Nº de cartões**
Cartão roubado	243
Cartão falsificado	85
Pedidos por correio/telefone	52
Outros	46

   Selecionado aleatoriamente uma caso de fraude nos casos resumidos na tabela, qual a probabilidade de a fraude resultar de um cartão falsificado? . R: **0,2**.

5. Se IP (A)= 2/5, determine $\mathrm{IP}(\overline{\mathrm{A}})$.

6. Com base em dados do Centro Nacional de Estatística de Saúde dos EUA, a probabilidade de uma criança ser menino é 0,513. Determine a probabilidade de uma criança ser menina.

7. Determine $\mathrm{IP}(\overline{\mathrm{A}})$, dado que IP (A)= 0,228.

8. Com base em dados do Centro Nacional de Examinadores Forenses, se escolhermos aleatoriamente uma pessoa que se submete ao exame para exercício da advocacia, a probabilidade de obter alguém que seja aprovado é 0,57. Ache a probabilidade de alguém que seja reprovado.

9. Os pesquisadores estão preocupados com declínio do nível de cooperação por parte dos entrevistados em pesquisas. A tabela mostra o resultado de uma pesquisa feita com 359 pessoas.

Faixa etária	Respondem	Não respondem	Total
18-21	73	11	84
22-29	255	20	275
Total	328	31	359

a) Qual probabilidade de obter alguém que não queira responder? R: **0,086**.
b) Qual probabilidade de obter alguém na faixa etária 22-29? R: **0,766**.
c) Determine a probabilidade de obter alguém na faixa etária 18-21 ou alguém que recuse responder. R: **0,29**.
d) Determine a probabilidade de obter alguém na faixa etária 18-21 que não recuse responder. R: **0,203**.

Testes de Hipóteses

Nesta parte, vamos admitir um valor hipotético para o parâmetro desconhecido - as hipóteses estatísticas - e, depois utilizar a informação da amostra para aceitar ou rejeitar esse valor hipotético.

Por exemplo, com base na produtividade de uma hortaliça cultivada em uma área, onde for usado um novo fertilizante, e em outra área onde se utiliza o fertilizante padrão, temos de decidir se o novo fertilizante é, ou não, melhor. A dificuldade aqui - e daí a necessidade de dados estatísticos - é que a produtividade varia de planta para planta.

Os testes de hipóteses permitem-nos tomar decisões em presença da variabilidade, ou seja, verificar se estamos diante de uma diferença real (significativa) ou de uma diferença devida simplesmente à flutuação aleatória inerente ao processo.

Na realização de um teste, são feitas duas hipóteses: a ***hipótese nula (H_0)***, que será testada, e a ***hipótese alternativa (H_1)***, que será aceita caso nosso teste indique a rejeição da hipótese nula.

Exemplos :

1- Indique as hipóteses nula e alternativa para cada uma das situações:

a) Tubos galvanizados devem ter média de 2 polegadas para serem aceitáveis.

b) Um fabricante de conservas deseja evitar excesso no enchimento de potes de 12 oz. De geléia.

2- Para cada um dos casos seguintes, decida se é adequado um teste ***unilateral*** ou um teste ***bilateral***, trace a curva normal para ilustrar o teste.

a) H_0: □=10 , $H_{1:}$ □□10, □=0,02

b) H_0: □=0,037 , $H_{1:}$ □>0,037, □=0,05

c) H_0: □=3,2 , $H_{1:}$ □<3,2, □=0,01

Tipos de Erros

O esquema a seguir mostra os erros que podemos cometer:

Conclusão do teste	***H_0 verdadeira***	***H_0 falsa***
Não rejeitar H_0	*Correto*	*Erro tipo II*
Rejeitar H_0	*Erro tipo I*	*Correto*

- **Procedimento para se efetuar um teste de hipótese**
- **5 Passos:**

1º) Enunciar as hipóteses H_0 e H_1;

2º) Fixar-se o limite de erro □ e identificar-se a variável do teste;

3º) Determinar-se a região crítica em função da variável tabelada;

4º) Calcular o valor da variável do teste, obtido na amostra;

5º) Aceitar ou rejeitar a hipótese nula de acordo com a estimativa obtida no item 4º, em comparação com a região crítica estabelecida no 3º) passo.

Valores críticos de z em testes de hipóteses

Nível de significância

Tipo de teste		
unilateral	**bilateral**	
5%	**+1,65 ou -1,65**	**□1,96**
1%	**+2,33 ou -2,33**	**□2,58**

Teste para a média (□² conhecido)

1º) Enunciar as hipóteses:

H_0: □ = $□_0$

H_1: $□ \neq □_0$ (a) ou $□ > □_0$ (b) ou $□ < □_0$ (c)

2º) Fixar o nível de significância □.

Admitindo-se que conhecemos a variância populacional a variável do teste será a distribuição Normal (Z)

3º) Região crítica

4º) Calcular:

$$Z = \frac{\overline{X} - □_0}{\frac{□}{\sqrt{n}}}$$

onde:

$\overline{X}$ = média amostral

$□_0$ = valor da hipótese nula

□ = desvio padrão da população

n = tamanho da amostra

5º) **Conclusões:**

a) Se □ Z □ > z rejeita-se H_0 (para um teste ***bicaudal***)

b) Se Z > z rejeita-se H_0 (para um teste ***unicaudal*** a direita).

c) Se Z < -z rejeita-se H_0 (para um teste ***unicaudal*** a esquerda).

Exemplo 2: Uma máquina automática de encher pacotes de erva mate, enche-os segundo uma distribuição normal com média □ e desvio padrão de 20g. A máquina foi regulada para □=500g. Desejamos verificar se a produção esta sob controle, para isto analisamos uma amostra de 30 pacotes. Se uma amostra apresentar média $\overline{X}$=492g, você pararia ou não a produção para verificar se a máquina deve ser regulada? Use □=1%.

1º Passo:

2º Passo:

3º Passo:

4º Passo:

5º Passo:

Teste para a média (□² desconhecido; n < 30)

Neste caso usaremos a distribuição “t”de Student. Logo no 4º e 5º passo teremos:

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

4º) Calcular:

onde: $\overline{X}$ = média amostral

μ_0 = valor da hipótese nula

S = desvio padrão da amostra

n = tamanho da amostra

5º) **Conclusões:**

b) Se $|T| > t$ rejeita-se H_0 (para um teste ***bicaudal***)

b) Se $T > t$ rejeita-se H_0 (para um teste ***unicaudal*** a direita).

c) Se $T < -t$ rejeita-se H_0 (para um teste ***unicaudal*** a esquerda).

Exemplo 3: Um fabricante afirma que a média de vida útil das lâmpadas por ele fabricadas é de 4.200 horas. A média da vida útil para uma amostra de N=10 lâmpadas é de 4.000 horas com um desvio padrão de amostral de S=200 horas. A vida útil das lâmpadas segue uma distribuição normal. Teste a afirmação do fabricante a um nível de significância de 5%.

1º Passo:

2º Passo:

3º Passo:

4º Passo:

5º Passo:

EXERCÍCIOS

1) Uma amostra aleatória de 40 elementos retirados de uma população normal com desvio padrão □ = 3 apresentou um valor médio igual a 60. Teste, ao nível de significância de 5%, a hipótese de que a média populacional seja igual a 59, supondo a hipótese alternativa □ > 59.

2) Uma amostra aleatória de 100 mortes naturais, no Rio Grande do Sul, deu uma média de $\overline{X}$ =71,8 anos, com um desvio padrão de 8,9 anos. Isto indica que o tempo médio de vida no RS, atualmente, é maior do que 70 anos? (□= 5%)

3) Uma amostra aleatória de tamanho n = 18 de uma população normal tem média $\bar{x}$ = 31,5 e desvio padrão s = 4,2. Ao nível de significância de 5%, estes dados sugerem que a média populacional seja superior a 30?

4) A resistência dos cabos fabricados por determinada companhia acusa média de 1800 libras e desvio padrão de 100 libras. Adotando-se uma nova técnica de fabricação, espera-se aumentar esta resistência. Para testar tal hipótese, toma-se uma amostra de 50 cabos fabricados pelo novo processo, obtendo-se uma resistência média de 1850 libras. Pode-se aceitar a hipótese ao nível de significância de 0,01?

5) Doze latas de lubrificante de certa marca acusam os conteúdos médios seguintes (decilitros):10.2 9.7 10.1 10.3 10.1 9.8 9.9 10.4 10.3 9.8 10.4 10.2. Ao nível de 1% , testar a hipótese de que o conteúdo médio das latas daquele lubrificante é □ = 10 dl. Admitir a normalidade da distribuição.

Teste de diferença entre médias:

Amostras Dependentes.

Neste caso usaremos a distribuição “t”de Student. Logo no 4º e 5º passo teremos:

4º) Calcular:

$$T = \frac{\bar{d} - \mu_d}{\frac{S_d}{\sqrt{n}}}$$

amostrais emparelhados(dependentes)

onde: $\bar{d}$ = valor médio das diferenças *d* para os dados

μ_d = média das diferenças *d* para a população de dados emparelhados.

$\mathbf{S_d}$ = desvio padrão das diferenças *d* para os dados amostrais emparelhados

n = número de pares de dados.

Graus de liberdade = n-1.

5º) **Conclusões:**

c) Se $|T| > t$ rejeita-se H_0 (para um teste ***bicaudal***)

b) Se $T > t$ rejeita-se H_0 (para um teste ***unicaudal*** a direita).

c) Se $T < -t$ rejeita-se H_0 (para um teste ***unicaudal*** a esquerda).

Exemplo: Utilizando um cronometrador de reação foi obtido a tabela abaixo, no nível de 0,05 de significância, teste a afirmação de que há uma diferença entre a média dos tempos de reação da mão direita e da mão esquerda.

Pessoa	A	B	C	D	E	F	G	H	I	J	K	L	M	N
Direita	191	97	116	165	116	129	171	155	112	102	188	158	121	133
Esquerda	224	171	191	207	196	165	177	165	140	188	155	219	177	174
D	-33	-74	-75	-42	-80	-36	-6	-10	-28	-86	33	-61	-56	-41

1º Passo:

2º Passo:

3º Passo:

4º Passo:

5º Passo:

Amostras Grandes e Independentes.

Neste caso usaremos a distribuição Normal. Logo no 4º e 5º passo teremos:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

4º) Calcular:

OBS: Se não conhecemos os valores de σ_1 e σ_2, podemos substituí-los por s_1 e s_2, desde que ambas as amostras sejam grandes.

5º) **Conclusões:**

d) Se $|Z| > z$ rejeita-se H_0 (para um teste ***bicaudal***)

b) Se $Z > z$ rejeita-se H_0 (para um teste ***unicaudal*** a direita).

c) Se $Z < -z$ rejeita-se H_0 (para um teste ***unicaudal*** a esquerda).

Exemplo:

1) Os alunos de uma faculdade selecionaram aleatoriamente 217 carros de estudantes e constataram que a média de suas idades era

de 7,89 anos, com desvio padrão de 3,67 anos. Selecionaram também, aleatoriamente, 152 carros do corpo docente e do pessoal da administração, constatando uma média de 5,99 anos e um desvio padrão de 3,65 anos. No nível de significância de 0,05, teste a afirmação de que os carros dos estudantes são mais velhos do que os dos professores e demais funcionários.

Amostras Pequenas e Independentes.

Neste caso usaremos a distribuição Normal. Logo no 4º e 5º passo teremos:

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

4º) Calcular:

Onde $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$ e o grau de liberdade é gl = n_1+n_2-2.

5º) **Conclusões:**

a) Se $|Z| > z$ rejeita-se H_0 (para um teste ***bicaudal***)

b) Se $Z > z$ rejeita-se H_0 (para um teste ***unicaudal*** a direita).

c) Se $Z < -z$ rejeita-se H_0 (para um teste ***unicaudal*** a esquerda).

Exemplo:

Os dados amostrais a seguir apresentam os níveis de concentração de álcool no sangue por ocasião da prisão de criminosos selecionados aleatoriamente, e que foram condenados por dirigirem embriagados. Os dados são categorizados por tipo de bebida consumida.

Cerveja			Uísque		
0,129	0,154	0,187	0,185	0,225	0,247
0,146	0,155	0,19	0,19	0,226	0,253
0,148	0,164	0,203	0,22	0,227	0,257
			0,224	0,241	

Com o nível de 0,05 de significância, teste a hipótese de que os bebedores de cerveja e os de uísque e semelhantes têm os mesmos níveis concentração de álcool no sangue.

$\bar{x}_1 = 0{,}164$ e $s_1 = 0{,}02427$ e $\bar{x}_2 = 0{,}227$ e $s_2 = 0{,}02317$

Exercícios:

1. Costuma-se avaliar a inteligência das crianças dando-lhes blocos e pedindo-lhes que construam uma torre tão alta quanto possível. Repetiu-se um mês depois o mesmo experimento, com os tempos (em segundos) dados na tabela a seguir. No nível de 0,01 de significância, teste a afirmação de que não há diferença entre os dois tempos.

Criança	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1ª tentativa	30	19	19	23	29	178	42	20	12	39	14	81	17	31	52
2ª tentativa	30	6	14	8	14	52	14	22	17	8	11	30	14	17	15

2. Utilize o nível de significância de 0,05 para testar a alegação de que as duas amostras provêm de populações com a mesma média. As amostras são independentes.

Elementos Tratados	Elementos Não-Tratados

$n_1 = 60$	$n_2 = 75$
$\bar{x}_1 = 8{,}75$	$\bar{x}_2 = 9{,}66$
$S_1 = 2{,}05$	$S_2 = 2{,}88$

3. O estresse afeta a capacidade de memorização de testemunhas oculares? Este problema foi estudado em um experimento que testou a memória visual de uma testemunha uma semana após o interrogatório normal de um suspeito que cooperava, e um interrogatório exaustivo de um suspeito que não cooperava. Os números de detalhes lembrados uma semana após o incidente estão resumidos aqui. No nível de 0,01, teste a afirmação do artigo de que “o cansaço concorre para diminuir a quantidade de detalhes lembrados.”

Sem Estresse	Com Estresse
$n_1 = 40$	$n_2 = 40$
$\bar{x}_1 = 53{,}3$	$\bar{x}_2 = 45{,}3$
$S_1 = 11{,}6$	$S_2 = 13{,}2$

4. Os dados relativos às rendas mensais observadas em uma amostra de 12 engenheiros e 14 advogados estão na tabela abaixo:

Engenheiros		Advogados	
8	11,5	7	9
9	11,5	7	9
9,5	12	7,5	9
10	12,5	8	10,5
11	13	8	11
11	13	8,5	11
		8,5	12

Teste a afirmação que a renda mensal média de advogados e engenheiros são iguais. ($\alpha = 0{,}05$)

5. Os distúrbios psiquiátricos sérios estão relacionados com fatores biológicos que possam ser observados fisicamente? Em um estudo foi

utilizada a tomografia computadorizada de raios X para coletar dados sobre o tamanho do cérebro de um grupo de pacientes com distúrbios obsessivos-compulsivos, e um grupo de controle constituído de pessoas sadias. A lista apresenta os resultados amostrais (em milímetros) para volumes do cordato direito.

Pacientes obsessivos-compulsivos			Grupo de controle		
0,21	0,305	0,344	0,334	0,429	0,483
0,287	0,308	0,407	0,349	0,445	0,501
0,288	0,334	0,455	0,402	0,46	0,519
0,304	0,34	0,463	0,413	0,476	0,594

Com nível de 0,01 de significância, teste a afirmação de que os pacientes obsessivos-compulsivos e as pessoas sadias têm os mesmos volumes cerebrais.

Teste para uma proporção

Devemos calcular:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p(1-p)}{n}}}$$

onde: $\hat{p}$ = proporção amostral

p_0 = valor da hipótese nula

n = tamanho da amostra

Exemplo: Uma estação de televisão afirma que 60% dos televisores estavam ligados no seu programa especial do último sábado. Uma rede competidora deseja contestar essa afirmação e decide, para isso, usar uma amostra de 200 famílias. Destas 200 famílias 104 responderam afirmativamente. Ao nível de 5% de significância qual a sua conclusão?

Teste para duas proporções

Devemos calcular:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\overline{p}.\overline{q}}{n_1} + \frac{\overline{p}.\overline{q}}{n_2}}}$$

onde: $\hat{p}_1 = \frac{x_1}{n_1}$ (proporção amostral)

p_1 = proporção populacional

n_1 = tamanho da amostra

$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2}$ e $\overline{q} = 1 - \overline{p}$.

Exemplo: Pesquisadores fizeram um estudo de empregadas da IBM que estavam grávidas. De 30 empregadas que lidavam com éter-glicol, 10 (ou 33,33%) tiveram aborto (espontâneo), mas, de 750 que não estavam expostas ao éter-glicol, apenas 120 (ou 16%) abortaram. No nível de 0,01 de significância, teste a afirmação de que as mulheres expostas ao éter-glicol apresentam maior taxa de aborto.

1) A preocupação com ambiente entra freqüentemente em conflito com a tecnologia moderna, como no caso dos pássaros que representam perigo para a aviação durante a decolagem. Um grupo ambiental afirma que tais acidentes com pássaros são tão raros que não se justifica matá-los. Um grupo de pilotos alega que entre as decolagens interrompidas que levam um avião a ultrapassar o final da pista, 10% são devidas a colisão com pássaros. Teste esta afirmação ao nível de 0,05. os dados consistem em 74 decolagens interrompidas, destas 5 foram devidos à colisão com pássaros.

2) Um relatório do Ministério da Justiça dos EUA inclui a afirmação de que “em casos de crimes entre casais, as esposas acusadas têm menor probabilidade de ser condenadas do que os maridos acusados.” Os dados amostrais consistiram em 277 condenações entre 318 maridos acusados, e 155 condenações entre 222

esposas acusadas. Tese a afirmação feita com nível de 0,01 de significância.

3) Uma questão de teste é considerada boa se permite discriminar entre estudantes preparados e estudantes não-preparados. A primeira questão de um teste foi respondida corretamente por 62 dentre 80 alunos preparados, e por 23 dentre 50 alunos não-preparados. Com o nível de 0,05 de significância, teste a afirmação de que esta questão foi respondida corretamente por uma proporção maior de estudantes preparados.

REFERÊNCIAS BIBLIOGRÁFICAS

AZEVEDO, Amilcar Gomes de; CAMPOS, Paulo Henrique Borges De. **Estatística básica**. 5.ed. Rio de Janeiro :LTC,1987.

VIEIRA, Sonia. **Elementos de estatística**. 3. ed. São Paulo :Atlas,1999.

NAZARETH, Helenalda Resende de Souza. **Curso básico de estatística**. 9. ed. São Paulo Ática,1997.

MARTINS, Gilberto de Andrade; DONAIRE, Denis. **Princípios de estatística**. 4. ed. São Paulo :Atlas,1995.

CRESPO, Antônio Arnot. **Estatística fácil**. 17. ed. São Paulo Saraiva,2000.