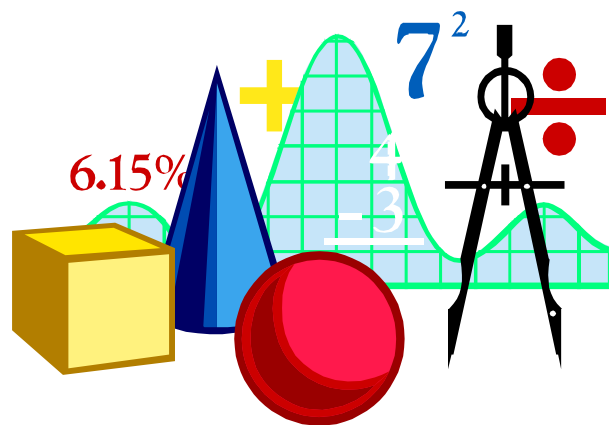


Estatística "II"



JAIR CROCE FILHO

SUMÁRIO

- **CAPÍTULO I - Modelo Probabilístico**
Página - 6

- **CAPÍTULO II - Variável Aleatória**
Página - 8

- **CAPÍTULO III - Distribuições Discretas de Probabilidade**
Página - 11

- **CAPÍTULO IV - Distribuições Contínuas de Probabilidade**
Página - 20

- **CAPÍTULO V - Testes de Hipóteses Aplicáveis à Pesquisa**
(Resumo) Página - 29

- **CAPÍTULO V - Testes de Hipóteses Aplicáveis à Pesquisa**
(Material completo) Página 34

- **CAPÍTULO VI - Teoria das Pequenas Amostras**
Página - 47

- **CAPÍTULO VII - Correlação e Regressão**
Página - 56

- **CAPÍTULO VIII - Séries Temporais**
Página - 65

- **CAPÍTULO IX - Números Índices**
Página - 68

ÍNDICE

- **CAPÍTULO I**
MODELO POBABILÍSTICO – 6
 - 1.1. Modelos Determinísticos - 6
 - 1.2. Modelos não Determinísticos - 6
 - 1.3. Aplicações - 7
 - 1.4. Experimento Aleatório - 7

- **CAPÍTULO II**
VARIÁVEL ALEATÓRIA – 8
 - 2.1. Conceito - 8
 - 2.2. Variável Aleatória Discreta Unidimensional - 8
 - 2.3. Variável Aleatória Contínua - 8
 - 2.4. Distribuições de Probabilidades - 8
 - 2.4.1. Distribuição Discreta - 9
 - 2.4.2. Distribuição Contínua - 9
 - Exercícios - 10

- **CAPÍTULO III**
DISTRIBUIÇÃO DISCRETAS DE PROBABILIDADE – 11
 - 3.1. Distribuição Bernoulli - 11
 - 3.1.1. Esperança Matemática, Variância e Desvio Padrão da Distribuição Bernoulli - 12
 - 3.2. Distribuição Binomial - 12
 - 3.2.1. Esperança Matemática, Variância e Desvio Padrão da Distribuição Binomial - 12
 - 3.3. Distribuição Multinomial - 13
 - 3.3.1. Esperança Matemática, Variância e Desvio Padrão da Distribuição Multinomial - 13
 - 3.4. Distribuição Geométrica - 14
 - 3.4.1. Esperança Matemática, Variância e Desvio Padrão da Distribuição Geométrica - 14
 - 3.5. Distribuição de Pascal - 14
 - 3.5.1. Esperança Matemática, Variância e Desvio Padrão da Distribuição de Pascal - 15
 - 3.6. Distribuição de Poisson - 15
 - 3.6.1. Esperança Matemática, Variância e Desvio Padrão da Distribuição Poisson - 16
 - Exercícios - 17

- **CAPÍTULO IV**
DISTRIBUIÇÕES CONTÍNUAS DE PROBABILIDADE – 20
 - 4.1. Distribuição Retangular ou Uniforme - 20
 - 4.1.1. Esperança Matemática, Variância e Desvio Padrão da Distribuição Uniforme - 20
 - 4.2. Distribuição Normal - 21
 - 4.2.1. Características (propriedades) da Curva Normal - 21
 - 4.2.2. Cálculo das Probabilidades - 22
 - 4.2.3. Distribuição Normal Padronizada - 22
 - 4.3. Distribuição Exponencial - 25
 - 4.3.1. Esperança Matemática, Variância e Desvio Padrão da Distribuição Binomial - 26

- CAPÍTULO V – (Resumo)
TESTES DE HIPÓTESES APLICÁVEIS À PESQUISA – 29

Exercícios - 33

- CAPÍTULO V – (Material completo)
TESTES DE HIPÓTESES APLICÁVEIS À PESQUISA – 34

- 5.1. Decisões Estatísticas - 34
- 5.2. Hipóteses Estatísticas. Hipótese Nulas - 34
- 5.3. Testes de Hipóteses e Significância - 34
- 5.4. Erro do Tipo I e II - 35
- 5.5. Nível de Significância - 35
- 5.6. Testes que Envolvem a Distribuição Normal - 35
- 5.7. Testes Unilaterais e Bilaterais - 36
- 5.8. Curvas Características de Operação. Potência de um Teste - 41
- 5.9. Cartas de Controle - 51
 - Exercícios - 46

- CAPÍTULO VI
TEORIA DAS PEQUENAS AMOSTRAS – 47

- 6.1. Distribuição t de Student - 47
- 6.2. Distribuição Qui-Quadrado - 51
 - Exercícios - 54

- CAPÍTULO VII
CORRELAÇÃO E REGRESSÃO – 56

- 7.1. Relação entre Variáveis - 56
- 7.2. Diagrama de Dispersão - 56
- 7.3. Tipos de Correlação Através do Diagrama de Dispersão - 56
 - 7.3.1. Correlação Linear Positiva - 56
 - 7.3.2. Correlação Linear Negativa - 57
 - 7.3.3. Correlação Nula - 57
 - 7.3.4. Correlação não Linear - 57
- 7.4. Coeficiente de Correlação Linear de Pearson - 57
- 7.5. Regressão Linear Simples - 59
 - 7.5.1. Objetivo - 59
 - 7.5.2. Equação da Reta - 59
 - 7.5.3. Fórmula para Calcular “a” e “b” - 59
- 7.6. Interpolação e Extrapolação - 60
- 7.7. Regressão Linear Múltipla - 61
 - Exercícios - 62

- CAPÍTULO VIII
SÉRIES TEMPORAIS – 65

- 8.1. Introdução - 65
- 8.2. Movimentos Característicos - 65
- 8.3. Análise das Séries Temporais - 66
- 8.4. Médias Móveis - Alisamento das Séries Temporais - 66

- CAPÍTULO IX NÚMEROS ÍNDICES – 68

- 9.1. Conceito e Objeto - 68
- 9.2. Conceito de Relativo - 68
- 9.3. Índices Relativos - 68
- 9.4. Tipos de Números Índices - 68
 - 9.4.1. Relativo Simples - 68
 - a) Relativo de Preços - 68
 - b) Relativo de Quantidades - 68
 - c) Valor Relativo - 69
 - d) Elos relativos - 70
 - e) Relativos em Cadeia - 70
 - 9.4.2. Agregativos Simples - 70
 - a) Agregativo Simples ou Índice de Bradstreet - 71
 - b) Médio-Aritmético - 71
 - c) Médio-Harmônico - 71
 - d) Médio-Geométrico - 72
 - e) Mediano - 72
 - **Notas e Conclusões - 73**
 - 9.4.3. Agregativos Ponderados - 73
 - a) Método ou Critério de Laspeyres - 73
 - b) Método ou Critério de Paasche - 74
 - Considerações sobre os Critérios de Laspeyres e Paasche - 74
 - c) Índice ideal de Fisher - 75
 - d) Método ou Critério das Médias Ponderadas de Relativos - 76
 - 1º) Média Aritmética Ponderada de Relativos - 77
 - 2º) Média Geométrica Ponderada de Relativos - 77
 - e) Método ou Critério de Sidgwick-Drobish - 77
 - f) Método ou Critério de Marshall-Edgeworth - 78
 - g) Método ou Critério das Médias das Quantidades dos Anos Típicos - 78
- 9.5. Testes, Provas e/ou Propriedades em Números Índices - 78
 - 1º) Identidade - 79
 - 2º) Reversão (Inversão) no Tempo - 79
 - 3º) Reversão de Fatores - 80
 - 4º) Cíclica ou Circular - 80
 - 5º) Cíclica ou Circular Modificada - 80
- 9.6. Mudança de Base em Séries de Números Índices - 81
 - 9.6.1. Construção - 81
 - a) Base Fixa - 81
 - b) Base Móvel em Cadeia - 81
 - 9.6.2. Vantagens e Desvantagens - 82
 - 9.6.3. Exemplo Prático - 82
- 9.7. Alguns Índices Especiais - 83
 - 1º) ICV – Índice de Custo de Vida (Índice de Preços ao Consumidor) - 83
 - 2º) Deflator Implícito de Renda (Produto) - 83
 - 3º) IGP – Índice Geral de Preços - 84
 - 4º) Índice “Relação de Troca” - 84
 - 5º) Índice da “Capacidade de Importar” - 84

- **ANEXOS:**
- * Anexo 1 – Tabela Valores “ λ ” (Poisson) – 85
 - * Anexo 2 – Tabela Valores da Curva Normal – 86
 - * Anexo 3 – Tabela Valores “t” de Student – 87
 - * Anexo 4 – Tabela Valores Qui-Quadrado – 88
 - * Anexo 5 – Gramática Grega – Símbolos Utilizados na Estatística – 89
 - * Glossário de Fórmulas – 90
 - * Bibliografia – 95

CAPÍTULO 1

Modelo Probabilístico

1.1- Modelos Determinísticos

Qualquer afirmativa fatural pode ser classificada, quanto à sua veracidade, em:

- 1º) **tautológica** – se sempre verdadeira;
- 2º) **contraditória** – quando sempre falsa;
- 3º) **contingente** (incerto) – quando for algumas vezes verdadeira e outras falsa.

A verdade e a falsidade podem ser afirmadas sempre que o fenômeno em estudo está sujeito a alguma lei que permita prever seus resultados a partir das informações disponíveis.

Exemplos:

1. um fio condutor de corrente elétrica partido pode ser “unido” se mergulhado em água com sal (sabe-se que ela é condutora de eletricidade);
2. se aquecida até 100° C, a água ferve (pois esse é o seu ponto de ebulição);
3. uma chapa metálica sem tratamento adequado oxida (em razão da combinação do oxigênio do ar com o metal de que ela é composta);
4. uma moeda lançada para o ar, com certa força, cai de volta ao solo depois de certo tempo (em razão da Lei da Gravitação Universal);
5. se meu salário passar de determinada faixa, terei que me entender com o imposto de renda;
6. se comprar a crédito, e não pagar as prestações nos respectivos vencimentos, serei importunado pelo credor, etc.

E veja que nem sempre as “leis” precisam ser assim complexas:

Exemplo:

Se for informado que determinada urna contém bolas pretas e brancas, não há como esperar que em alguma extração surja uma bola vermelha.

Os modelos que explicam as situações acima exemplificadas são chamados de **modelos determinísticos**. Em geral, são modelos científicos resultantes de muitos anos (e até mesmo séculos) de estudo apurado e contribuições de muitas gerações de grandes pensadores.

1.2- Modelos não Determinísticos (Probabilísticos)

Por outro lado, se o fenômeno não está sujeito a lei alguma (ou melhor, se as leis que o regem ainda não são conhecidas), nada pode, em princípio, ser afirmado sobre ele. Nesses casos, dizemos que ele está sujeito às leis do acaso e o único modelo possível para o seu estudo é o chamado modelo probabilístico. O modelo probabilístico é aquele aplicável aos fenômenos cujos resultados exatos em cada experimento futuro são desconhecidos e imprevisíveis. O estudo de tais modelos é o objetivo da teoria das probabilidades; sua aplicação só é possível nas situações onde há incerteza quanto a ocorrência ou não dos eventos estudados.

Acima, fez-se referência às “leis do acaso”, idéia um tanto quanto paradoxal, se levarmos em conta que “lei”, que pressupõe regularidade, aparenta ser o oposto de “acaso”, que é sinônimo de irregularidade. Destaque-se que o paradoxo se resolve no claro entendimento de que o que se busca é identificar as possíveis regularidades ou simetrias que os fenômenos possam apresentar quando repetidos muitas vezes.

1.3- Aplicações

A utilização da teoria das probabilidades tem crescido com o tempo. De seus primeiros usos nos jogos de azar, que ainda permanecem nos cassinos, loterias, corridas de cavalo etc., evoluiu até uma ampla utilização na sociedade em geral.

Exemplos:

1. os economistas às vezes a usam na previsão da demanda por produtos a serem lançados futuramente, na previsão das safras agrícolas e na avaliação dos impactos dos aumentos dos impostos sobre a inflação;
2. os professores avaliam as chances de determinados alunos serem aprovados ou não e, com isso, estimam o percentual de reprovados no período ou ano, conforme o caso;
3. os médicos determinam se (e quando) determinado doente deverá estar recuperado;
4. os sindicatos estabelecem a chance de determinada categoria entrar em greve;
5. o governo determina até quando será capaz de manter os preços congelados ou tabelados, quando ocorrerá uma paralisação de trabalhadores e quais as suas implicações.

Além dos usos (possíveis) acima relacionados, as probabilidades também nos ajudam no desenvolvimento e avaliação de estratégias de ação.

Exemplos:

1. numa auto-estrada, freamos (ou aceleramos) o carro dependendo da visibilidade, do número de carros ao nosso redor e do policiamento presente (ostensivo ou não);
2. de manhã, saímos de casa com roupas adequadas à nossa expectativa de frio ou calor;
3. as decisões sobre negócios ou aplicações financeiras são rápidas quando as chances de ganho ou perda são altas (especialmente, em termos de curto prazo);
4. os investimentos em pessoal e novos equipamentos são feitos se há razoável chance de aumento da produtividade.

1.4- Experimento Aleatório

O experimento aleatório é aquele que poderá ser repetido sob as mesmas condições indefinidamente. Tal experimento apresenta variações de resultados, não sendo possível afirmar *a priori* qual será sua determinação antes que o mesmo tenha sido realizado. É possível, porém, descrever todos os possíveis resultados – as probabilidades.

Exemplos:

1. O lançamento de um dado constitui um experimento aleatório, pois esse experimento poderá ser repetido quantas vezes desejarmos. Antes do lançamento, não poderemos dizer qual será o resultado, mas somos capazes de relatar os “possíveis” resultados: sair o número 1, 2, 3, 4, 5 ou 6;
2. o mesmo podemos dizer quando retiramos uma carta de baralho com 52 cartas e observamos seu naipe.

CAPÍTULO 2

Variável Aleatória

2.1- Conceito

Por variável aleatória se entende uma lista de valores ou uma função genérica que associa um **número** conveniente aos eventos componentes do espaço amostral de um dado experimento, sejam eles qualitativos ou quantitativos.

Ex.: Dado o experimento “lançamento de 3 moedas”, determine a probabilidade de dar “*pelos menos uma coroa*”.

X	N.º de Caras	Eventos Característicos	Frequência
0	Nenhuma	KKK	1/8
1	Uma	CKK, KCK, KKC	3/8
2	Duas	KCC, CKC, CCK	3/8
3	Três	CCC	1/8
C = Cara K = Coroa		Espaço Amostral	8/8

2.2- Variável Aleatória Discreta Unidimensional

Como “X” pode assumir **certos** valores com dadas probabilidades, ele é freqüentemente denominado variável aleatória discreta, ou seja, transforma espaço amostral não numérico em espaço amostral numérico.

O número de pontos do espaço amostral de um dado experimento pode ser finito ou infinito. Se o espaço amostral é finito, então os eventos são enumeráveis e a variável a ele associada deve permitir a contagem dos elementos nele contidos. Neste caso, a variável aleatória é dita “discreta” ou “enumerável”.

Ex.: Número de jogadores de um time de futebol.
 Número de torcedores no estádio.
 Número de gols marcados na partida.

2.3- Variável Aleatória Contínua

Como já foi dito, o número de pontos do espaço amostral de um dado experimento pode ser finito ou infinito. Se o espaço amostral é infinito, então a contagem do seu número de elementos não é possível por nenhum meio. Neste caso, a variável aleatória é dita “contínua”, indicando que ela pode assumir qualquer valor dentro do seu intervalo de validade.

Ex.: - tempo de duração de um jogo;
 - peso e velocidade da bola na hora do gol.

2.4- Distribuições de Probabilidades

Se associarmos a cada ponto do espaço amostral a correspondente probabilidade de ocorrência do evento, obteremos a distribuição (ou espaço) das probabilidades do experimento. Se, além disso, associarmos a esses mesmos pontos uma variável aleatória indicadora, obteremos uma distribuição que pode ser entendida como a distribuição das probabilidades das variáveis aleatórias associadas aos pontos do espaço amostral.

A distribuição assim obtida irá mostrar como a probabilidade total se divide pelos diversos resultados possíveis do experimento. Ela poderá ser totalmente definida tanto pelas frequências simples quanto pelas acumuladas. Essas “distribuições de probabilidades” são ditas “discretas” quando a variável é discreta (contável ou enumerável); se, por outro lado, a variável aleatória é contínua, então as distribuições a elas correspondentes são chamadas de “Distribuições Contínuas de Probabilidades”.

2.4.1. Distribuição Discreta de Probabilidade

As distribuições de probabilidades recebem o adjetivo “discretas” ou “descontínuas” quando a variável aleatória envolvida é enumerável ou contável.

Ex.: número de peças defeituosas por lote; o de nascimentos por ano; o de páginas de um jornal; o de número de pés de café por alqueire plantado; etc..

2.4.2. Distribuição Contínua de Probabilidade

As distribuições de probabilidades são ditas “contínuas” quando a variável aleatória associada ao evento pode assumir qualquer valor dentro de certo intervalo.

Ex.: peso de uma peça; altura de um aluno; tempo de duração de uma sonata; a área de uma quadra de vôlei; volume de refrigerante em um copo; etc..

Exercícios de aplicação:

01) No lançamento de um par de dados honestos onde ‘X’ indique a soma dos pontos obtidos, demonstre o espaço amostral do conjunto.

Então a distribuição de probabilidade é dada pela seguinte tabela.

X	2	3	4	5	6	7	8	9	10	11	12
P (X)											
$\sum P (X)$											

→ Função de Probabilidade

02) Em 900 lances de 2 dados, em quantos podemos esperar que dêem a soma 5?

03) No lançamento de duas moedas a variável aleatória X anota o número de caras obtidas. Determine os valores de X e a função de probabilidade associada.

X			
P (X)			

EXERCÍCIOS

- 1º) No lançamento de dois dados a variável aleatória X anota, em módulo, a diferença dos pontos das faces superiores. Determine os valores de X e a função de probabilidade associada.
- 2º) Uma carta é retirada aleatoriamente de um baralho de 52 cartas. A variável aleatória X anota o número de damas obtidas nesta retirada. Determine os valores de X e a função de probabilidade associada.
- 3º) Duas cartas são retiradas aleatoriamente, sem reposição, de um baralho comum de 52 cartas. A variável aleatória X anota o número de valetes obtidos. Determine os valores de X e a função de probabilidade associada.
- 4º) A urna A contém 3 bolas brancas e duas bolas pretas. A urna B contém 5 bolas brancas e 1 bola preta. Uma bola é retirada ao acaso de cada urna, a variável aleatória X anota o número de bolas brancas obtidas. Determine os valores de X e a função da probabilidade associada.

CAPÍTULO 3

Distribuições Discretas de Probabilidade

3.1- Distribuição de Bernoulli - (*James Bernoulli – fim do séc. XVII*)

É a mais simples das distribuições de probabilidades para variáveis aleatórias discretas e sua importância reside no fato de a mesma ser **geratriz** de outras distribuições mais complexas como a distribuição binomial e a distribuição de Pascal que serão estudadas oportunamente neste capítulo.

É um caso particular da **distribuição binomial**. Com ela é possível calcular as probabilidades associadas a uma **única** tentativa do experimento. Neste caso, só há duas possibilidades: fracasso (aí se atribui '0' à variável aleatória) ou sucesso (quando se atribui '1' à variável aleatória).

A fórmula adequada a este tipo de distribuição é:

$$P(x) = P^x (1 - P)^{1-x}$$

Ex.: Retira-se uma bola de uma urna contendo 30 bolas brancas e 20 verdes. Qual a probabilidade dessa bola ser verde?

$$P_{\text{(verde)}} = \frac{2}{5} \left(1 - \frac{2}{5}\right)^{1-1} = \frac{2}{5}$$

OBSERVAÇÃO: O que é esperança matemática?

Se uma tabela de mortalidade diz que uma mulher de 50 anos de idade pode esperar viver mais 31 anos, isto não quer dizer que realmente esperemos que uma mulher de 50 anos viva até completar 81 anos e morra no dia seguinte. Da mesma forma, se lemos que uma pessoa espera comer 104,4 lbs. de carne e beber 39,6 galões de refrigerante por ano, ou que uma criança de 6 a 16 anos espera visitar um dentista 2,2 vezes por ano, é óbvio que a palavra “espera” não está sendo empregada no sentido usual. Uma criança não pode ir ao dentista 2,2 vezes por ano, e seria realmente surpreendente se encontrássemos alguém que comesse 104,4 lbs. de carne e bebesse 39,6 galões de refrigerante em determinado ano. No que diz respeito às mulheres de 50 anos, umas vivem mais 12 anos, outras vivem mais 20 anos, outras viverão mais 33 anos,... e a esperança de vida de “31 anos ou mais” deve ser interpretada como uma média ou, como chamamos aqui, como uma **esperança matemática**.

Originalmente, o conceito de esperança matemática surgiu em relação aos jogos de azar e, em sua forma mais simples, é o produto da quantia que um jogador aposta pela respectiva probabilidade de ganho.

Exemplos:

- a) Qual é a nossa esperança matemática quando esperamos ganhar R\$ 10,00 se e somente se uma moeda equilibrada apresentar “cara”?
R.: $10,00 \times 1/2 = 5,00$
- b) Qual é a nossa esperança matemática se compramos um dos 2.000 bilhetes de rifa de um aparelho de TV avaliado em R\$ 640,00?
R.: Como a probabilidade de ganhar o aparelho de TV é $1/2000 = 0,0005$, nossa esperança matemática é $640,00 \times 0,0005 = 0,32$

Assim, do ponto de vista econômico, seria insensato pagar mais de R\$ 0,32 pelo bilhete.

3.1.1- Esperança Matemática, Variância e Desvio Padrão da Distribuição de Bernoulli

$$\mu = p \quad \sigma^2 = pq \quad \sigma = \sqrt{pq}$$

3.2- Distribuição Binomial

A distribuição Binomial constitui uma extensão da distribuição de Bernoulli, sendo aplicada em problemas nos quais um experimento de Bernoulli é realizado **um número de vezes preestabelecido**. Cada uma destas realizações é denominada prova.

É a distribuição adequada quando os pontos do espaço amostral podem ser agrupados em **duas classes** ou categorias mutuamente excludentes, às quais se atribuem os nomes genéricos de “sucesso” e “fracasso”, indistintamente. Outra forma de caracterizar a distribuição Binomial é dizer que ela é a distribuição aplicável quando o experimento atende às condições típicas do processo de amostragem de Bernoulli, que são as seguintes:

- há ‘*n*’ **provas** ou observações idênticas (repetidas sempre nas mesmas condições);
- o experimento é constituído de eventos independentes;
- só há dois resultados possíveis, que são mutuamente excludentes (um chamado “sucesso” e o outro falha ou “fracasso”);
- como os eventos são independentes e as provas repetidas sempre nas mesmas condições, as probabilidades de ocorrência dos resultados possíveis é constante em todas as tentativas.

A fórmula de cálculo da probabilidade de certo número ‘*x*’ de “sucessos” em ‘*n*’ provas de um experimento sujeito à distribuição binomial é dada por:

$$P(x) = C_n^r \cdot P^r \cdot q^{n-r}$$

onde: $C_n^r = \frac{n!}{r!(n-r)!}$

P = Probabilidade de sucesso

q = Probabilidade contrária ou insucesso

Ex.: Joga-se uma moeda seis vezes.

Qual é a probabilidade de se obter exatamente duas vezes a face cara?

3.2.1- Esperança Matemática, Variância e Desvio Padrão da Distribuição Binomial

$$\mu = np \quad \sigma^2 = npq \quad \sigma = \sqrt{npq}$$

Ex.: Considerando-se os dados do exemplo anterior, calcule a esperança matemática, a variância e o desvio padrão.

Exercícios de aplicação:

01) Joga-se um dado seis vezes.

Qual é a probabilidade de se obter quatro vezes **ou** mais a face 2 ?

02) Uma moeda não viciada é lançada 8 vezes. Encontre a probabilidade de sair pelo menos uma cara. (FONSECA: 65)

03) Utilizando os dados dos exercícios acima, calcule a média, a variância e o desvio padrão para cada situação.

3.3- Distribuição Multinomial (ou Polinomial)

A distribuição multinomial é uma **generalização** da binomial. Assim é a distribuição adequada quando o experimento tem **dois ou mais** resultados possíveis mutuamente excludentes.

Suas hipóteses características são:

- as provas são repetidas sempre nas mesmas condições;
- o experimento é constituído de eventos independentes;
- o experimento tem vários resultados possíveis, mas todos mutuamente excludentes;
- como os eventos são independentes e as provas repetidas sempre nas mesmas condições, as probabilidades de ocorrência dos resultados possíveis é constante em todas as tentativas.

Numa Distribuição Multinomial, a probabilidade de que os eventos x_1, \dots, x_k , cujas probabilidades de ocorrência são, respectivamente, p_1, \dots, p_k , ocorram n_1, \dots, n_k vezes, em 'n' tentativas, é dado por:

$$P(x) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$$

Onde: $n = n_1 + \dots + n_k$ é o número de tentativas ou repetições do experimento;
 n_1, \dots, n_k é o número / proporção / frequência dos eventos x_1, \dots, x_k ;
 p_1, \dots, p_k é a probabilidade / proporção / frequência dos eventos x_1, \dots, x_k .

Ex.: Dadas as máquinas A, B, C e D, que produzem respectivamente 10%, 20%, 30% e 40% da produção total de certa oficina, determine a probabilidade de um lote de 2 dúzias de peças aleatoriamente escolhidas terem sido produzidas do seguinte modo: 4 pela máquina A, 5 pela B, 7 pela C e 8 pela D.

$$P = \frac{24!}{4!5!7!8!} \cdot 0,10^4 \cdot 0,20^5 \cdot 0,30 \cdot 0,40^8 \quad P(A=4, B=5, C=7 \text{ e } D=8) = 0,49\%$$

3.3.1- Esperança Matemática, Variância e Desvio Padrão da Distribuição Multinomial

$$\mu_i = np_i \quad \sigma_i^2 = np_i q_i \quad \sigma = \sqrt{np_i q_i}$$

Ex.: Considerando-se os dados do exemplo anterior, calcule a esperança matemática, a variância e o desvio padrão.

Exercícios de aplicação:

- Se um dado honesto é lançado 12 vezes, qual a probabilidade de serem obtidos os pontos 1, 2, 3, 4, 5 e 6, exatamente duas vezes cada um? (SPIEGEL: 190)
- Um dado é lançado 10 vezes. Qual a probabilidade de terem aparecido 2 vezes o número 2, duas vezes o número 5, três vezes o número 1 e uma vez os demais resultados? (FONSECA:66)
- Utilizando os dados dos exercícios acima, calcule a média, a variância e o desvio padrão para cada situação.

3.4- Distribuição Geométrica

Destinada ao cálculo de probabilidades de situações em que são feitas sucessivas tentativas independentes de um mesmo experimento aleatório até que apareça o **1º sucesso**.

A probabilidade é calculada através da seguinte expressão:

$$P(x) = Pq^{x-1}$$

Onde: x = número de tentativas até o aparecimento do 1º sucesso;

P = probabilidade de sucesso;

$q = 1 - P$ é a probabilidade de fracasso.

Ex.: Qual é a probabilidade de que um dado deva ser lançado 15 vezes para que ocorra a primeira face 6?

$$P_{(\text{face } 6)} = \frac{1}{6} \cdot \frac{5}{6}^{15-1} = 1,298\%$$

3.4.1- Esperança Matemática, Variância e Desvio Padrão da Distribuição Geométrica

$$\mu = \frac{1}{p} \quad \sigma^2 = \frac{q}{p^2} \quad \sigma = \frac{\sqrt{q}}{p}$$

Ex.: Considerando-se os dados do exemplo anterior, calcule a esperança matemática, a variância e o desvio padrão.

Exercícios de aplicação:

- 01) Uma máquina produz peças das quais 80% são consideradas perfeitas e 20% defeituosas. Qual a probabilidade de a 5ª peça ser a 1ª a apresentar defeito? (APOSTILA UFJF)
- 02) A probabilidade de que haja alguma falha no lançamento de uma nave espacial é 10%. Qual é a probabilidade de que para lançar a nave seja necessário:
 - a) 2 tentativas?;
 - b) no máximo 3 tentativas? (APOSTILA JOSÉ ANTONIO)
- 03) Considerando-se os dados do exemplo 02, calcule o número esperado de tentativas de lançamento da nave espacial. Calcule também a variância e o desvio padrão do número de tentativas de lançamento.

3.5- Distribuição de Pascal (ou Binomial Negativa)

É o caso geral do modelo geométrico. A distribuição de Pascal calcula as probabilidades nas situações em que seja necessário **certo número de repetições** antes que ocorra um sucesso. O sucesso está na **última prova**.

A probabilidade é calculada através da seguinte expressão:

$$P(x) = C_{(n-1)}^{(x-1)} \cdot P^x \cdot q^{n-x}$$

Ex.: Dada uma máquina que produz 20% de peças defeituosas, qual a probabilidade de a 8ª peça fabricada ser a 5ª boa?

Sendo $n = 8$, $x = 5$, $P = 0,80$ e $q = 0,20$, temos:

$$P(5) = C_{8-1}^{5-1} \cdot 0,8^5 \cdot 0,2^{8-5} = C_7^4 \cdot 0,8^5 \cdot 0,2^3 = 35 \cdot 0,33 \cdot 0,008 = 9,18\%$$

3.5.1- Esperança Matemática, Variância e Desvio Padrão da Distribuição de Pascal

$$\mu = \frac{n}{p} \quad \sigma^2 = \frac{nq}{p^2} \quad \sigma = \frac{\sqrt{nq}}{p}$$

Ex.: Considerando-se os dados do exemplo anterior, calcule a esperança matemática, a variância e o desvio padrão.

Exercício de aplicação:

Uma máquina produz peças das quais 80% são consideradas perfeitas e 20% defeituosas. Qual a probabilidade de a 8ª peça fabricada ser a 3ª defeituosa? (APOSTILA UFJF)

3.6- Distribuição de Poisson – Siméon Denis Poisson – (1781-1840)

A distribuição de Poisson é uma aproximação da distribuição binomial, quando o número de provas ‘n’ tende para o infinito e a probabilidade ‘p’ do evento em uma prova tende para zero, mantendo-se finito e não nulo o produto ‘np’ (média da distribuição).

Na distribuição de Poisson se ‘n’ é um valor muito grande e ‘p’ um valor muito pequeno, temos o que se chama “acontecimento raro”. Como exemplo de acontecimento raro podemos citar: a ocorrência de gêmeos, fenômenos meteorológicos de rara apresentação, número diário de desastres de automóvel numa cidade, número de objetos defeituosos que aparecem num processo de fabricação, número de sementes de ervas daninhas no meio das sementes de uma planta, etc..

Na prática diz-se que um acontecimento é raro quando $n \geq 50$ e $p \leq 0,1$

Descreve a probabilidade de certo número de ocorrências num dado intervalo, espaço ou campo contínuo (tempo, comprimento, área, volume, peso, etc.). Determina, por exemplo, a probabilidade da ocorrência de certo número de chamadas telefônicas por minuto, de clientes por hora, de acidentes por dia, de defeitos por m² de tecido, de pés de café por alqueire e lactobacilos por ml de leite. **(Sempre a dupla ‘n’ e ‘p’)**.

A distribuição de Poisson, além da utilização para o cálculo aproximado da distribuição binomial tem vasta aplicação em problemas de fila de espera; controle de qualidade; programação de equipamentos; etc..

A fórmula que calcula as probabilidades numa distribuição de Poisson é:

$$P(x) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}$$

Onde: x = número de ocorrências;

e = 2,71828; é a base dos logaritmos neperianos;

λ = taxa média de ocorrências dos eventos por unidade de medida;

t = espaço de medida ou número de intervalos ou unidades.

OBS.: Para facilitar, os valores de λ foram tabelados e se encontram no anexo 1.

Ex.: Sabendo-se que os clientes chegam a uma loja Augusta à razão de 6 por hora, determine a probabilidade de, durante uma hora qualquer:

a-) não chegar nenhum cliente:

$$\lambda = 6 \text{ clientes por hora} = 6/1; \quad t = 1 \text{ hora}; \quad x = 0$$

$$P(0) = \frac{e^{(-6 \cdot 1)} \cdot (6 \cdot 1)^0}{0!} = 0,002479 = 0,25\%$$

b-) chegar ao menos um cliente:

$$P(X \geq 1) = 1 - P(0) = 100 - 0,2479 = 99,75\%$$

3.6.1- Esperança Matemática, Variância e Desvio Padrão da Distribuição de Poisson

$\mu = \lambda t$	$\sigma^2 = \lambda t$	$\sigma = \sqrt{\lambda t}$
-------------------	------------------------	-----------------------------

Exercícios de aplicação:

- 01) Em média há 2 chamadas por hora num certo telefone. Calcular a probabilidade de: (FONSECA: 68)
- nenhuma chamada em 90 minutos;
 - se receber no máximo 3 chamadas em 2 horas;
- 02) Suponha que haja em média 2 suicídios por ano numa população de 50.000 habitantes. Em uma cidade de 100.000 habitantes, encontre a probabilidade de que em um dado ano tenha havido: (FONSECA: 71)
- nenhum suicídio;
 - um suicídio;
 - dois suicídios;
 - dois ou mais suicídios.
- 03) Suponha 400 erros de impressão distribuídos aleatoriamente em um livro de 500 páginas. Encontre a probabilidade de que uma dada página contenha: (FONSECA: 71)
- nenhum erro;
 - exatamente 2 erros.
- 04) Certo tipo de máquina apresenta em média 18 falhas a cada 60 horas de trabalho. Calcule a média, a variância e o desvio padrão do número de falhas em 8 horas de trabalho desta máquina. (APOSTILA JOSÉ ANTÔNIO)
- 05) Num processo de fabricação de certo tipo de aparelho, ocorrem em média 1,6 defeitos por aparelho. Se o fabricante paga uma indenização de R\$20,00 por aparelho com mais de 2 defeitos, qual é o custo esperado desta indenização na venda de 500 aparelhos? (APOSTILA JOSÉ ANTÔNIO)

EXERCÍCIOS

DISTRIBUIÇÃO BINOMIAL

- 1º) Uma empresa produz 10% de peças defeituosas. As peças são embaladas em caixas que contêm 12 peças. Calcule a probabilidade de um cliente comprar uma caixa contendo:
- nenhuma peça defeituosa;
 - uma peça defeituosa. (ERMES: 37)
- R.: a) 28,24% b) 37,66%**
- 2º) Admitindo-se que os nascimentos de meninos e meninas sejam iguais calcular a probabilidade de um casal com 6 filhos ter 4 filhos homens e 2 mulheres. (FONSECA: 68)
- R.: 23,44%**
- 3º) Se 5% das lâmpadas de certa marca são defeituosas, achar a probabilidade de que, numa amostra de 100 lâmpadas, escolhidas ao acaso, nenhuma apresente defeito. (FONSECA: 69)
- R.: $0,95^{100}$**
- 4º) Uma VAD X tem distribuição binomial com média igual a 3 e variância igual a 2. Calcule $P(X = 2)$. (APOSTILA JOSÉ ANTONIO)
- R.: 23,41%**
- 5º) Uma empresa compra matrizes para gráfica em pacotes contendo 5 matrizes. Sabe-se que 20% das matrizes compradas têm defeitos de fabricação que implicam na sua utilização. Calcule a probabilidade de: (APOSTILA JOSÉ ANTONIO)
- um pacote conter mais da metade das matrizes oxidadas;
 - pelo menos 2 pacotes de um grupo de 6 pacotes conterem uma matriz inutilizada.
- R.: a) 5,79% b) 78,15%**
- 6º) Um produtor de sementes vende pacotes com 20 sementes cada. Os pacotes que apresentarem mais de uma semente sem germinar são indenizadas. A probabilidade de uma semente germinar é 0,98. (APOSTILA JOSÉ ANTONIO)
- Qual é a probabilidade de que um pacote não seja indenizado?;
 - se o produtor vende 1.000 pacotes, em quantos se espera indenização?;
 - quando o pacote é indenizado, o produtor tem um prejuízo de R\$1,20, e se o pacote não for indenizado, o produtor tem um lucro de R\$2,50. Qual o lucro líquido esperado por pacote?;
 - calcule a média, a variância e o desvio padrão do número de sementes por pacotes que germinam.
- R.: a) 94,01% b) 60 c) R\$2,28 d) $\mu = 19,6$; $\sigma^2 = 0,39$; $\sigma = 0,63$**

DISTRIBUIÇÃO MULTINOMIAL

- 7º) Uma caixa contém 5 bolas vermelhas, 4 brancas e 3 azuis. Uma bola é escolhida ao acaso da caixa, sua cor é observada, e a bola é então recolocada. Determinar a probabilidade de, entre 6 bolas assim escolhidas, 3 serem vermelhas, 2 brancas e 1 azul. (SPIEGEL: 204)
- R.: 12,06%**
- 8º) Sabem-se que as probabilidades referentes a declarações de imposto de renda são as seguintes:
- 0,6 de ser preenchida corretamente;
 - 0,2 de conter erros favoráveis ao contribuinte;
 - 0,1 de conter erros favoráveis ao fisco;
 - 0,1 de conter erros favoráveis ao contribuinte e ao fisco.

Em 10 declarações selecionadas ao acaso, qual a probabilidade de 6 estarem corretas, 2 com erros a favor do contribuinte, 1 com erro a favor do fisco e 1 com os 2 tipos de erro? (APOSTILA UFJF)

R.: 4,7%

9º) No estoque de certa loja existem refrigeradores das marcas A, B e C. Sabe-se que as probabilidades de que sejam vendidos os refrigeradores das marcas A, B e C são 35%, 25% e 40%.

a) Se forem vendidos 20 refrigeradores qual a probabilidade de que sejam 6 da marca A, 9 da marca B e 5 da marca C?;

b) Admitindo-se que sejam vendidos 200 refrigeradores calcule a média e o desvio padrão do número de refrigerantes de cada marca. (APOSTILA JOSÉ ANTONIO)

R.: a) 0,56% b) $\mu_A = 70$; $\mu_B = 50$; $\mu_C = 80$; $\sigma_A = 6,7$; $\sigma_B = 6,1$; $\sigma_C = 6,9$;

DISTRIBUIÇÃO GEOMÉTRICA

10º) Sabe-se que 20% das peças de um estoque são defeituosas. Um experimento consiste em retirar peças, uma de cada vez, com reposição, até que se obtenha uma peça defeituosa.

a) Qual a probabilidade de que a peça defeituosa seja retirada na 3ª tentativa?;

b) qual a probabilidade de serem necessárias mais de 3 tentativas para se retirar uma peça defeituosa?;

c) calcule a média, a variância e o desvio padrão do número de retiradas até que se obtenha uma peça defeituosa. (APOSTILA JOSÉ ANTONIO)

R.: a) 12,8% b) 51,2% c) $\mu = 5$ $\sigma^2 = 20$ $\sigma = 4,5$

11º) O custo operacional em cada tentativa de lançamento de um foguete é R\$100.000,00. Se houver falha, ocorrerá um custo extra de R\$ 50.000,00, em virtude de consertos na plataforma de lançamento. A probabilidade de um lançamento bem sucedido em cada tentativa é 0,8. (APOSTILA JOSÉ ANTONIO)

a) Qual a probabilidade de serem necessários mais de 3 tentativas para o lançamento do foguete?;

b) qual é o custo esperado do processo?

R.: a) 0,8% b) R\$137.500,00

DISTRIBUIÇÃO DE PASCAL (Binomial negativa)

12º) Uma empresa recebeu uma encomenda para fundir 3 peças complicadas. A probabilidade de se conseguir um molde adequado é 0,4, sendo o molde destruído quando a peça é retirada. Para atender à encomenda, qual a probabilidade de ser necessário fundir:

a) exatamente 4 peças?;

b) no máximo 6 peças? (APOSTILA JOSÉ ANTONIO)

R.: a) 11,52% b) 45,57%

13º) Uma empresa recebe uma encomenda para produzir 5 unidades de certo artigo. A probabilidade de se obter uma unidade sem defeitos é 0,8. O custo de cada unidade é R\$50,00 e se a unidade apresentar algum defeito ocorre um custo adicional de R\$10,00.

a) Qual a probabilidade de que seja necessário produzir pelo menos 8 unidades para atender a encomenda?;

b) qual o preço a ser cobrado pelo serviço para que se tenha um lucro esperado de R\$100,00? (APOSTILA JOSÉ ANTONIO)

R.: a) 14,8% b) R\$425,00

DISTRIBUIÇÃO DE POISSON

14°) Se a probabilidade de um indivíduo sofrer uma reação nociva, resultante da injeção de um determinado soro, é 0,001, determinar a probabilidade de, entre 2.000 indivíduos:

- a- exatamente 3 sofrerem aquela reação;
- b- mais do que 2 sofrerem aquela reação.

R.: a) 18,0% b) 32,3%

OBS.: Demonstrar a resolução do exercício acima através da distribuição binomial.

(SPIEGEL: 203)

15°) Uma máquina produz nove peças defeituosas a cada 1.000 peças produzidas. Calcule a probabilidade de que em um lote que contém: (ERMES: 43)

- a) 200 peças, sejam encontradas oito peças defeituosas;
- b) 500 peças, não haja nenhuma peça defeituosas.

R.: a) 0,045% b) 1,11%

16°) Supondo que o número de carros que chegam a uma fila do guichê de um pedágio tem distribuição de Poisson a uma taxa de três por minuto, calcule a probabilidade de que cheguem cinco carros nos próximos dois minutos. (ERMES: 44)

R.: 16,06%

17°) Um tear produz um defeito a cada 200 m de tecido produzido. Se o número de defeitos admite distribuição de Poisson, calcule a probabilidade de: (ERMES: 44)

- a) uma peça com 20 m não apresentar defeitos;
- b) um lote de 10 peças de 20 m cada, apresentar exatamente um defeito

R.: a) 90,48% b) 36,79%

18°) Em certo tipo de tecido ocorrem em média 3 defeitos a cada 2 m². Calcule a probabilidade de que numa peça de 1,5 m² deste tecido ocorram: (APOSTILA JOSÉ ANTONIO)

- a) nenhum defeito;
- b) menos de 4 defeitos;
- c) pelo menos 3 defeitos.

R.: a) 10,54% b) 80,94% c) 39,07%

19°) Certo tipo de equipamento elétrico consome 750 metros de fio. Sabe-se que o fio rompe em média duas vezes a cada 1.000 metros. O lucro e a qualidade deste equipamento são relacionados de acordo com o quadro a seguir

Qualidade	N.º de emendas	Lucro
1 ^a	Nenhuma	R\$100,00
2 ^a	Uma ou duas	R\$60,00
3 ^a	Mais de duas	R\$40,00

Se a produção anual deste equipamento é de 10.000 unidades, qual o lucro esperado em um ano? (APOSTILA JOSÉ ANTONIO)

R.: R\$651.000,00

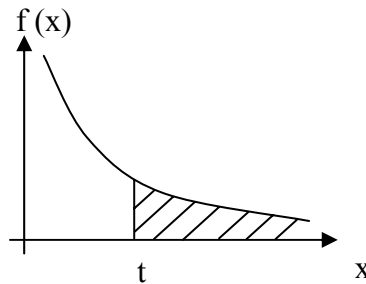
CAPÍTULO 4

Distribuições Contínuas de Probabilidade

4.1- Distribuição Exponencial

A distribuição exponencial é a distribuição de probabilidade do intervalo 't' entre dois sucessos consecutivos de uma lei de Poisson.

A distribuição exponencial é a adequada ao estudo das probabilidades de ocorrências de certos intervalos ou espaços contínuos entre eventos. É aplicável a situações como o tempo (ou distância) entre erros de operação de um equipamento, o tempo entre os gols num jogo de futebol, entre a chegada de navios, trens, ônibus, etc..



Uma v. a. X terá distribuição exponencial se sua função de densidade de probabilidade for da seguinte forma:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x > 0, \lambda > 0 \\ 0 & \text{para outros valores de } x \end{cases}$$

A determinação de probabilidades para a distribuição exponencial envolve o cálculo de áreas sob curva de densidade, problema que exige recursos do cálculo integral.

Pode-se mostrar que, se T tem distribuição exponencial com parâmetro λ , então:

$$\begin{aligned} P(T > t) &= e^{-\lambda t} \\ P(T < t) &= 1 - e^{-\lambda t} \end{aligned}$$

Onde: t = período de avaliação;
 λ = 1 dividido pela média das ocorrências.

Ex.: Se os clientes chegam a uma loja da Augusta à razão de 6 por hora, qual a probabilidade de que o primeiro cliente:

a-) só chegue 40 minutos depois de ela ter sido aberta?

$\lambda = 6/60 = 1/10 = 0,1$, ou seja, média de 1 cliente a cada 10 minutos.

$$P(T > 40) = e^{-0,1 \cdot 40} = e^{-4} = 1,83\%$$

b-) chegue no primeiro quarto de hora após ela ter sido aberta?

$$P(T < 15) = 1 - e^{-0,1 \cdot 15} = 1 - e^{-1,5} = 77,69\%$$

4.1.1- Esperança Matemática, Variância e Desvio Padrão da Distribuição Exponencial

$$\mu_1 = 1/\lambda \quad \sigma^2 = 1/\lambda^2 \quad \sigma = 1/\lambda$$

Ex.: Considerando-se os dados do exemplo anterior, calcule a esperança matemática, a variância e o desvio padrão.

Exercícios de aplicação:

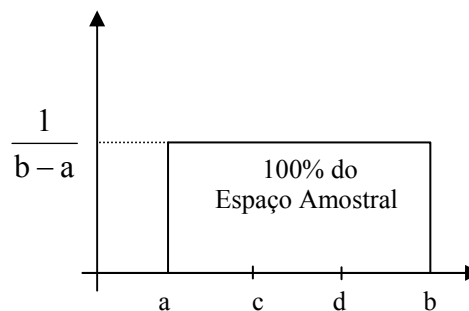
- 01) Os defeitos de um tecido seguem a distribuição de Poisson com média de um defeito a cada 400 m. Qual a probabilidade de que um intervalo entre dois defeitos consecutivos seja: (FONSECA: 84)
- no mínimo de 1.000 m;
 - entre 800 e 1.000 m;
 - calcule a média e a variância da distribuição.
- 02) Se as interrupções no suprimento de energia elétrica ocorrem segundo uma distribuição de Poisson com a média de uma interrupção por mês (quatro semanas), qual a probabilidade de que entre duas interrupções consecutivas haja um intervalo de: (FONSECA: 89)
- menos de uma semana;
 - exatamente um mês;
 - mais de três semanas;
 - entre 10 e 12 semanas.

4.2- Distribuição Retangular ou Uniforme

A distribuição uniforme é adequada quando a variável aleatória se distribui por igual dentro da sua escala de possíveis valores. Em outras palavras, ela deve ser adotada quando nada nos indicar que um valor em particular é preferível, mais freqüente ou mais provável que outro qualquer, fazendo supor que os eventos ocorrem de modo constante ou uniforme.

É a mais simples das distribuições de probabilidades de variável aleatória contínua, sendo utilizada em condições bastante restritas.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } a < x < b \\ 0 & \text{para } x \leq a \text{ e } x \geq b \end{cases}$$



No gráfico acima, sabendo-se com toda certeza que dado evento ocorre no intervalo $[a, b]$, a probabilidade de que ele ocorra no subintervalo $[c, d]$ é dada por:

$$P(c < x < d) = \frac{\overline{CD}}{\overline{AB}}$$

Onde: $\overline{CD} = d - c$ indica a amplitude do intervalo onde **espera-se encontrar** a v.a..

$\overline{AB} = b - a$ indica a amplitude do intervalo em que **certamente** se encontrará a v.a..

Ex.: Dado que o volume de refrigerante num copo enchido por uma máquina post-mix, varia uniformemente entre 270 ml e 300 ml, pede-se a probabilidade de se obter um copo com menos de 290 ml.

$$P(270 < X < 290) = \frac{290 - 270}{300 - 270} = \frac{20}{30} = \frac{2}{3} = 66,67\%$$

4.2.1- Esperança Matemática, Variância e Desvio Padrão da Distribuição Uniforme.

$$\mu = \frac{a + b}{2} \quad \sigma^2 = \frac{(b - a)^2}{12} \quad \sigma = \frac{b - a}{\sqrt{12}}$$

Ex.: Considerando-se os dados do exemplo anterior, calcule a esperança matemática, a variância e o desvio padrão do volume de refrigerante num copo enchido por uma máquina post-mix.

Exercícios de aplicação:

- 01) Um ponto é escolhido ao acaso no segmento de reta $[0, 2]$. Qual será a probabilidade de que o ponto esteja entre 1 e $3/2$? (FONSECA: 73)
- 02) Um ponto é escolhido ao acaso no segmento de reta $[1, 4]$. Qual será a probabilidade de que o ponto esteja entre: (FONSECA: 84)
 - a- 2 e 3;
 - b- 0,5 e 2,5;
 - c- seja exatamente o 2.
- 03) Calcule a média, a variância e o desvio padrão com os dados dos exercícios acima.

4.3- Distribuição Normal ou Curva Normal ou Distribuição de Gauss – Carl Friedrich. Gauss – astrônomo – (1777-1855)

A distribuição normal tem significativa importância no estudo da distribuição das médias e das proporções amostrais de grandes amostras.

A distribuição normal é a mais importante das distribuições contínuas de probabilidade porque um grande número de variáveis aleatórias associadas a experimentos reais tem distribuições de probabilidades que se aproximam da distribuição normal.

Também conhecida, como já foi dito, “curva em forma de sino” ou de “montanha”, tem sua origem associada aos erros de mensuração normalmente observados pelos cientistas nas medições efetivas das grandezas físicas (como distâncias, pesos, volumes, etc.).

É sabido que, quando se efetuam repetidas mensurações de determinada grandeza com um aparelho equilibrado, não se chega ao mesmo resultado todas as vezes; obtém-se, ao contrário, um conjunto de valores que oscilam, de modo aproximadamente simétrico, em torno do

verdadeiro valor. Construindo-se o histograma desses valores e o correspondente polígono de frequência, obtém-se uma poligonal aproximadamente simétrica.

Inicialmente se supunha que todos os fenômenos da vida real deveriam ajustar-se a uma curva em forma de sino; em caso contrário, suspeitava-se de alguma anormalidade no processo de coleta de dados. Daí a designação de **curva normal**.

A observação cuidadosa subsequente mostrou, entretanto, que essa pretensa universalidade da curva, ou distribuição normal, não correspondia à realidade. De fato, não são poucos os exemplos de fenômenos da vida real representados por distribuições não-normais (curvas assimétricas).

A distribuição normal desempenha, não obstante, papel preponderante na estatística, e os processos de inferência nela baseados têm larga aplicação.

4.3.1- Características (propriedades) da Curva Normal

- 1ª) É um modelo teórico e tem forma de sino ou de montanha;
- 2ª) é simétrica em relação ao ponto central (o de máxima frequência);
- 3ª) dada a simetria da distribuição, 50% dos valores são inferiores à média e 50% superiores a ela;
- 4ª) é unimodal, ou seja, no seu pico coincidem a média, a mediana e a moda;
- 5ª) por ser o padrão de enviesamento e afilamento, tem-se que todos os momentos e os coeficientes de assimetria são iguais a zero e o coeficiente de curtose igual a 0,263;
- 6ª) é assintótica em relação ao eixo horizontal, ou seja, prolonga-se indefinidamente para a esquerda e para a direita sem jamais tocá-lo;
- 7ª) fica completamente especificada pela média e pelo desvio padrão da variável;
- 8ª) há uma curva normal para cada combinação de média e desvio padrão;
- 9ª) qualquer distância medida em desvios padrão, **acima ou abaixo da média**, tem a mesma área sob a curva - **ela é simétrica**. Com isto a área sob a curva é igual a 1.

4.3.2- Cálculo das Probabilidades

As probabilidades associadas à distribuição normal são dadas pela integral da sua função de densidade, dada a seguir:

$$f(x) = \frac{e^{-\frac{1}{2}z^2}}{\sigma\sqrt{2\pi}} \quad \text{ou} \quad P(a < x < b) = \frac{1}{\sqrt{2\pi}} \int_a^b \frac{e^{-\frac{1}{2}z^2}}{\sigma} dx$$

$$\text{Onde: } z = \frac{X_i - \mu}{\sigma}$$

Como se observa, calcular as áreas (e as probabilidades) pela fórmula acima não é tarefa fácil. Por essa razão, os principais valores foram previamente calculados e os resultados encontram-se na Tabela de Distribuição Normal (ANEXO 2 – Tabela Curva Normal, página 84).

4.3.3- Distribuição Normal Padronizada ou Normal Reduzida

A montagem da tabela de probabilidades associada à distribuição normal só é possível com uma padronização. Isso porque na prática há uma série infinita de distribuições normais, cada uma produzida por um par de valores (média e desvio-padrão). A “padronização” é feita adotando-se uma distribuição particular como referência e transformando os valores reais em relativos. A distribuição de referência – chamada de **distribuição padronizada** ou **variável normal padronizada** ou **variável normal reduzida** ou **escala ‘Z’** ou **escore ‘Z’** – mede o

afastamento das variáveis em relação à média, em número de desvios-padrões. É aquela na qual a **média da distribuição** é igual a '0' e o **desvio-padrão da distribuição** é igual a '1'.

O cálculo do escore 'Z' é feito pela expressão:

$$\boxed{Z = \frac{X_i - \mu}{\sigma}} \Rightarrow \text{Fórmula para População}$$

$$\boxed{Z = \frac{X_i - \bar{X}}{S}} \Rightarrow \text{Fórmula para Amostra}$$

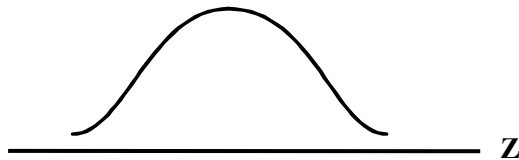
Onde: **Z** = número de desvios-padrões, a contar da média;

X = valor qualquer da variável aleatória;

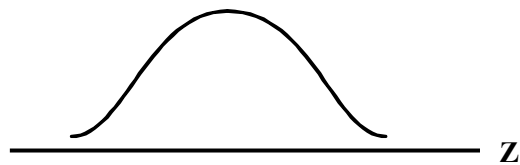
μ = média da distribuição;

σ = desvio-padrão da distribuição.

Ex.: 1) Determinar a área sob a curva normal padronizada à esquerda de 1,72.

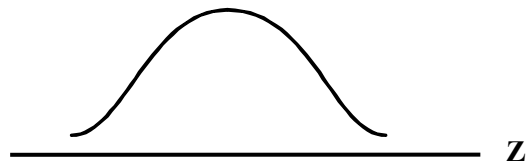


2) Determinar a área sob a curva normal padronizada abaixo de $Z = -0,53$.

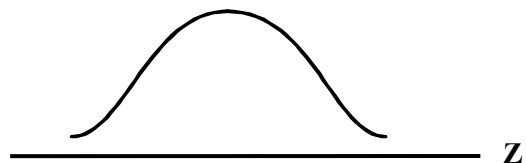


Exercícios de aplicação:

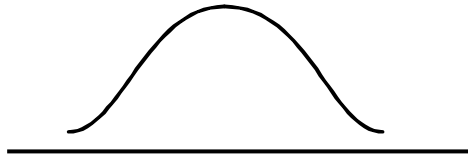
1) Determinar a probabilidade da v.a. padronizada Z estar entre $Z_0 = 0,70$ e $Z_1 = 1,35$.



2) Determinar a probabilidade da v.a. padronizada Z estar entre $Z_0 = -0,60$ e $Z_1 = 1,30$



- 3) Determinar a probabilidade de a v.a. reduzida Z tomar valores maiores do que 1,80.



- 4) Calcular, usando a tabela da distribuição normal padronizada, as seguintes probabilidades:
 a) $P(0 < Z < 1,34)$ b) $P(-2,58 < Z < 0)$ c) $P(Z > 1,54)$
- 5) Determine os valores da variável normal padronizada correspondente às seguintes probabilidades:
 a) $P(Z < z) = 0,0901$ b) $P(Z > z) = 0,1401$ c) $P(|Z| > z) = 0,0308$

OBSERVAÇÃO: Os problemas da vida real não se apresentam já na forma reduzida; ao contrário, são formulados em termos de variável normal original “ X ”, com média μ e desvio padrão σ . É preciso então, antes de passarmos à sua resolução, transformá-la na variável normal reduzida, que é o *Escore Padronizado* “ Z ”, visto anteriormente.

Exemplo:

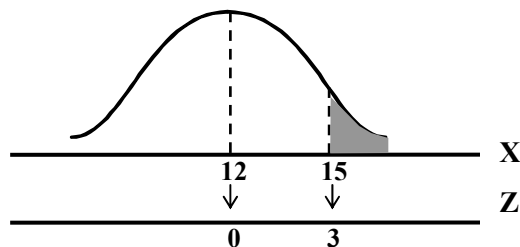
A taxa de hemoglobina no sangue de pessoas que gozam de boa saúde segue uma distribuição normal com média 12 e desvio padrão 1.

Qual a probabilidade de se encontrar uma pessoa normal com taxa de hemoglobina:

- a) Superior a 15?
 b) Inferior a 10?
 c) Entre 10 e 13?
 d) Num grupo de 500 pessoas, em quantas devemos esperar as características acima?

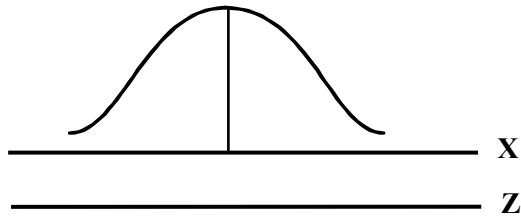
Solução: letra a) Padronizando $X = 15 \Rightarrow Z = \frac{15 - 12}{1} = 3$

$$P(X > 15) = P(Z > 3) = 0,5000 - 0,4987 = 0,0013 = 0,13\%.$$



Exercícios de aplicação:

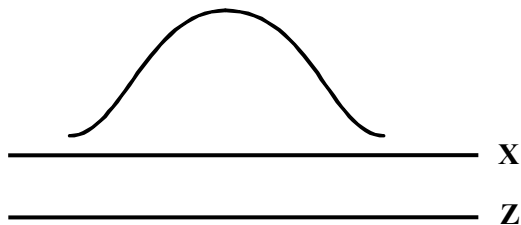
- 1) Suponha-se que a renda anual de São Paulo tenha uma média de R\$5.000,00 com desvio padrão de R\$1.500,00. Admitindo-se uma distribuição normal, que podemos dizer de uma renda de R\$7.000,00?



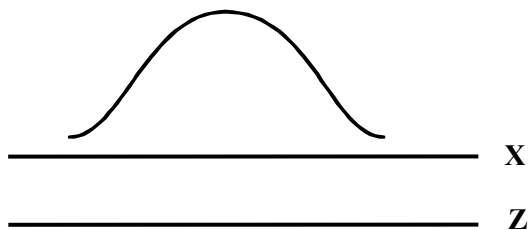
Conclusões: _____

2) Sabendo-se que o diâmetro de um rolamento que segue uma distribuição normal, tem média igual a 100 mm e desvio padrão igual a 5 mm, determine:

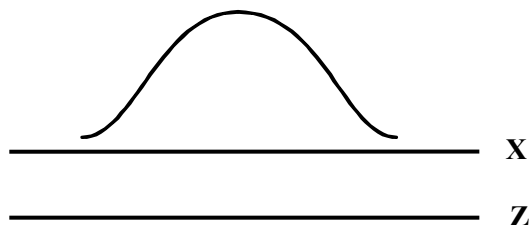
A) A porcentagem de diâmetros abaixo de 98 mm.



B) A medida do diâmetro, acima da qual, tem-se 10% de diâmetros.



3) Uma distribuição normal tem média igual a 62,4. Determinar o *desvio padrão* se 0,33 da área sob a curva estão à direita de 79,2.



4) Certa máquina de empacotar arroz produz pacotes cujo peso tem distribuição normal com desvio padrão de 40 g. para quanto deve ser regulado o peso médio dos pacotes para que menos de 9,51% dos pacotes tenham menos de 4.972 g?

EXERCÍCIOS

DISTRIBUIÇÃO EXPONENCIAL

- 1º) Se o tempo médio entre um pedido e o atendimento em um restaurante é uma v.a. com distribuição exponencial com média 10 min, determine: (APOSTILA UFJF)
- probabilidade de espera superior a 10 minutos;
 - probabilidade de espera inferior a 10 minutos;
 - probabilidade de espera inferior a 3 minutos.
- R.: a) 36,8% b) 63,2% c) 25,9%**
- 2º) A duração de certa marca de condensador tem distribuição exponencial com média de 200 horas. Qual é a proporção de condensadores que duram: (ZA)
- menos de 100 horas;
 - mais de 500 horas;
 - entre 200 e 400 horas.
- R.: a) 39,35% b) 8,21% c) 23,25%**
- 3º) Certa marca de lâmpada tem distribuição aproximadamente exponencial, com vida média de 1.000 horas. Determinar a percentagem das lâmpadas que queimarão antes de 1.000 horas. (SOARES: 133)
- R.: 63,2%**
- 4º) Um montanhista está perdido no flanco de uma montanha, fazendo sinal com sua lanterna para atrair a atenção de uma patrulha de salvamento. Se o tempo, em horas, que a lanterna é capaz de emitir um feixe de luz visível embaixo é uma v. a. exponencial com parâmetro $\lambda = 1/6$, qual a probabilidade de o sinal luminoso ser visto por ao menos 3 horas?
- R.: 60,65%**

DISTRIBUIÇÃO UNIFORME

- 5º) Um entreposto comercializa diariamente entre 100 e 200 toneladas de um cereal, com distribuição uniforme de probabilidades. Sabendo-se que o ponto de equilíbrio para esta operação corresponde a uma venda de 130 toneladas por dia, calcule a probabilidade de o comerciante realizar prejuízo em determinado dia. (ERMES SILVA: 68)
- R.: 30%**
- 6º) Uma VAC tem distribuição uniforme entre 2 e 7. Qual é a média e o desvio padrão desta variável? (APOSTILA JOSÉ ANTONIO)
- R.: $\mu = 4,5$; $\sigma = 1,44$**
- 7º) Uma variável aleatória x tem distribuição uniforme no intervalo $[3,5]$. (ERMES: 69)
- Determine a função de densidade de probabilidade associada;
 - construa o gráfico desta função.
- 8º) Uma variável aleatória x tem distribuição uniforme no intervalo $[0,10]$. Determine:
- a média da distribuição;
 - o desvio padrão. (ERMES: 69)
- R.: a) 5 b) 2,89**

DISTRIBUIÇÃO NORMAL

9º) Calcular, usando a tabela da distribuição normal padronizada, as seguintes probabilidades:

- a) $P(0 < Z < 1,24)$ b) $P(-2,63 < Z < 0)$ c) $P(Z > 1,24)$ d) $P(Z < -2,63)$
 e) $P(Z < 1,24)$ f) $P(-2,63 < Z < 1,24)$ g) $P(Z > -2,63)$ h) $P(1,24 < Z < 2,63)$

R.: a) 39,25% b) 49,57% c) 10,75% d) 0,43%
 e) 89,25% f) 88,82% g) 99,57% h) 10,32%

10º) Determine os valores da variável normal padronizada correspondente às seguintes probabilidades:

- a) $P(Z < z) = 0,9922$ b) $P(Z > z) = 0,0764$ c) $P(Z > z) = 0,9099$ d) $P(|Z| < z) = 0,95$

R.: a) 2,42 b) 1,43 c) - 1,34 d) 1,96

11º) Um cereal é ensacado automaticamente, sendo o peso médio da saca igual a 50 kg com desvio padrão de 1,6 kg. Haverá uma indenização de R\$5,00 para cada saca fornecida com menos de 48 kg. Admitindo-se que o peso das sacas tem distribuição normal, responda aos seguintes quesitos:

- a) qual será o custo médio no caso de serem fornecidas 200 sacas?
 b) para quanto se deve regular o peso médio das sacas para que o custo de indenização de 200 sacas caia para R\$40,00?
 c) como o administrador da produção acha desvantajoso para o custo total regular a máquina para aumentar o peso médio, ele pretende que a empresa adquira uma nova máquina para aumentar o peso médio, ele pretende que a empresa adquira uma nova máquina que mantém o peso médio das sacas em 50 kg. Qual deve ser o desvio padrão para que se pague indenização apenas de 3% das sacas produzidas?
 d) supondo que o custo do produto ensacado seja R\$40,00 por saca, qual é o custo esperado de 500 sacas ao incluir as indenizações, considerando-se a máquina antiga e a nova máquina?

R.: a) R\$105,60 b) 50,8 kg c) 1,06 kg d) antiga: R\$10.264,00; nova R\$20.075,00

12º) Os fios elétricos produzidos por certa empresa são cortados automaticamente e vendidos em rolos de 100 m. Devido aos erros da máquina, os comprimentos nos rolos se distribuem normalmente com desvio padrão igual a 1 m. O custo de cada metro de fio para a empresa é R\$3,00 por metro. Para reduzir as reclamações dos consumidores, estabeleceu-se que a empresa pagará multa de R\$ 21,00 para cada rolo que apresentar comprimento menor que 99 m. Sabendo-se que a máquina pode ser regulada para produzir rolos com comprimento médio de 100, 100,5 m ou 101 m e considerando-se o comprimento do fio uma VAC normalmente distribuída, determine a regulação mais econômica para a empresa.

R.: A regulação da máquina para o comprimento médio de 100,5 m é a mais econômica pois o custo total esperado é R\$302,90 enquanto os custos para as regulações do comprimento médio para 100 e 101 m acarretam custos totais esperados de R\$303,33 e R\$ 303,48, respectivamente.

CAPÍTULO 5 – (RESUMO)

Testes de Hipóteses Aplicáveis à Pesquisa

Fundamentos do Teste de Hipóteses

Metodologia de Teste de Hipóteses

Exemplo

O Gerente de produção está preocupado em avaliar se o processo está funcionando de modo a assegurar que, na média, a quantidade apropriada de cereal (isto é, 368 gramas) está sendo colocada em cada caixa.

Metodologia de Teste de Hipóteses

- Se o conteúdo médio não é igual a 368 gramas.

Decisão

- Ações corretivas são necessárias.

Metodologia de Teste de Hipóteses

- Hipóteses formuladas pelo Gerente de Produção:

O processo está sob controle, isto é, a média de conteúdo de cereais é igual a 368 gramas.

A hipótese de que o parâmetro da população seja igual à especificação da empresa é identificada como

Hipótese nula

Metodologia de Teste de Hipóteses

A metodologia de teste de hipótese é projetada de modo que nossa rejeição à hipótese nula se baseie em evidência a partir da amostra, e que nossa hipótese alternativa seja bem mais provável de ser verdadeira.

Objetivos

- Desenvolver métodos de testes de hipóteses para analisar diferenças;
- Tomar decisões

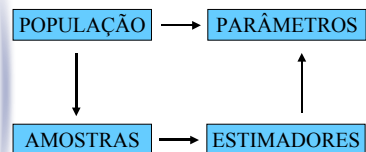
Metodologia de Teste de Hipóteses

- Se o conteúdo médio, no processo como um todo, é de 368 gramas.

Decisão

- Nenhuma ação corretiva é necessária.

Metodologia de Teste de Hipóteses



Metodologia de Teste de Hipóteses

$H_0: \mu = 368$ Hipótese nula

$H_1: \mu \neq 368$ Hipótese alternativa

Metodologia de Teste de Hipóteses

Aceitar a hipótese nula (H_0) não é prova de que ela seja verdadeira.

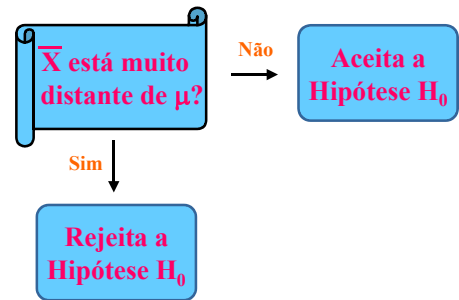
A decisão de aceitar H_0 , nos leva a concluir que não existem evidências suficientes para garantir a sua rejeição.

O Valor Crítico da Estatística do Teste

A quantidade de cereal por caixa é igual a 368 gramas?

Coletou uma amostra de caixas, pesou cada caixa e calculou a média aritmética da amostra.

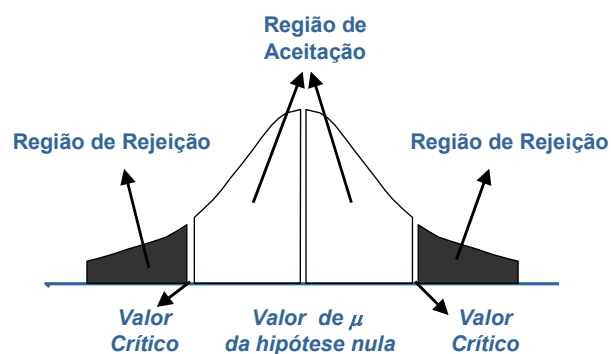
O Valor Crítico da Estatística do Teste



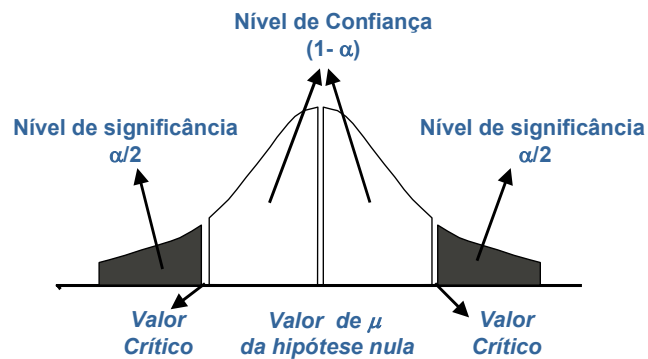
Muito próximo? Muito longe? ...

A metodologia de teste de hipóteses avalia tais diferenças e nos possibilita quantificar nosso processo de tomada de decisão.

Regiões de Rejeição e de Não-Rejeição



Regiões de Rejeição e de Não-Rejeição



Riscos na Tomada de Decisão

Erro do tipo I (α)

Quando uma hipótese for rejeitada quando deveria ser aceita.

Erro do tipo II (β)

Quando uma hipótese for aceita quando deveria ser rejeitada.

Riscos na Tomada de Decisão

Decisão Estatística	Situação Verdica	
	H_0 é verdadeiro	H_0 é falso
Aceitar H_0	Confiança $(1 - \alpha)$	Erro do tipo II β
Rejeitar H_0	Erro do tipo I α	Poder do Teste $(1 - \beta)$

ESTATÍSTICA DE TESTE

O teste Z de Hipóteses para a média populacional (σ conhecido)

$$Z_{\text{teste}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

\bar{x} = Média das amostras

μ = Média da população

ONDE: σ = Desvio Padrão da população

n = Número de elementos da amostra

O teste Z de Hipóteses para a média populacional (σ conhecido)

$H_0: \mu = 368$ Hipótese nula

$H_1: \mu \neq 368$ Hipótese alternativa

Exercício de Aplicação

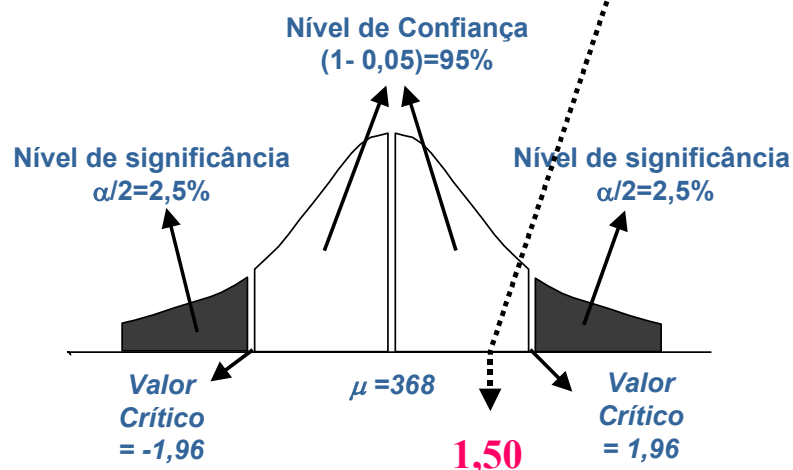
Testar estas hipóteses, sendo:

- a) nível de significância de 5%;
- b) amostra de tamanho 25;
- c) média dessa amostra = 372,5;
- d) σ (desvio padrão) = 15;

O teste Z de Hipóteses para a média populacional (σ conhecido)

$$Z_{\text{teste}} = \frac{372,5 - 368}{\frac{15}{\sqrt{25}}} = 1,5$$

Não rejeitar H_0



EXERCÍCIOS

1º Exercício

Suponha que o gerente de uma loja de comércio de tintas queira calcular a verdadeira quantidade de tinta contida nas latas de um galão, compradas de um fabricante nacionalmente conhecido.

Sabe-se, pelas especificações do fabricante, que o desvio padrão da quantidade de tinta é igual a 0,02 galão. Uma amostra aleatória de 50 latas é selecionada, e a quantidade média de tinta por lata de 1 galão é igual a 0,995 galão.

Existem evidências de que a quantidade média seja diferente de 1,0 galão? (utilize $\alpha = 0,01$)

2º Exercício

O gerente de controle de qualidade de uma fábrica de lâmpadas de filamento precisa calcular a vida útil média de uma grande remessa de lâmpadas.

Sabe-se que o desvio padrão do processo é de 100 horas.

Uma amostra aleatória de 50 lâmpadas indicou uma vida útil média da amostra é igual a 350 horas.

Há evidências de que a vida útil média seja diferente de 375 horas? (utilize $\alpha = 0,05$)

3º Exercício

A divisão de inspeção do Departamento de Pesos e Medidas de São Paulo está interessada em calcular a real quantidade de refrigerante que é colocada em garrafas de 2 litros, no setor de engarrafamento de uma grande empresa de refrigerantes.

O setor de engarrafamento informou à divisão de inspeção que o desvio padrão para garrafas de 2 litros é 0,05 litro.

Uma amostra aleatória de 100 garrafas de 2 litros, obtida deste setor de engarrafamento, indica uma média da amostra de 1,99 litro.

Há evidências de que a quantidade média nas garrafas seja diferente de 2,0 litros? (utilize $\alpha = 0,05$)

CAPÍTULO 5 – (MATERIAL COMPLETO)

Testes de Hipóteses Aplicáveis à Pesquisa

5.1- Decisões estatísticas

Na prática, somos chamados com muita frequência a tomar decisões acerca de populações, baseadas nas informações das amostras. Essas decisões são denominadas *decisões estatísticas*.

Ex.:

Pode-se desejar decidir, com base em dados amostrais, se um novo soro é realmente eficaz na cura de uma doença, se um processo educacional é melhor do que outro, se uma certa moeda é viciada etc..

5.2- Hipóteses estatísticas. Hipóteses nulas

Ao se tentar chegar às decisões, é conveniente a formulação de hipóteses ou de conjecturas acerca das populações interessadas. Essas suposições, que podem ser ou não verdadeiras, são denominadas *hipóteses estatísticas* e, em geral, são afirmações acerca das distribuições de probabilidade das populações.

Em alguns casos, formula-se uma hipótese estatística com o único propósito de rejeitá-la ou invalidá-la.

Ex.:

Se se deseja decidir se uma moeda é viciada, formula-se a hipótese de que ela não o seja, isto é $p = 0,5$, em que p é a probabilidade de caras. De modo semelhante, se se deseja decidir se um processo é melhor do que outro, formula-se a hipótese de que *não há diferença* entre eles, isto é, que quaisquer diferenças observadas sejam devidas meramente a flutuações das amostras provenientes da *mesma* população. Essas hipóteses são denominadas *hipóteses nulas* e representamo-las por H_0 .

Qualquer hipótese que difira de uma prefixada é denominada *hipótese alternativa*.

Ex.:

Se se admite que $p = 0,5$, são hipóteses alternativas: $p = 0,7$, $p \neq 0,5$ ou $p > 0,5$. Uma hipótese alternativa da nula é representada por H_1 .

5.3- Testes de hipóteses e significância

Admita uma hipótese particular como verdadeira, se se verificar que os resultados observados em uma amostra aleatória diferem acentuadamente dos esperados para aquela hipótese, com base na probabilidade simples mediante a utilização da teoria da amostragem, poder-se-á concluir que as diferenças observadas são *significativas* e ficar inclinado a rejeitar a hipótese (ou, pelo menos, a não aceitá-la com base nas provas obtidas).

Ex.:

Se 20 lances de uma moeda apresentarem 16 caras, ficamos inclinados a rejeitar a hipótese de que a moeda é honesta, embora seja concebível que se esteja incorrendo em erro.

Os processos que habilitam a decidir se se aceitam ou rejeitam as hipóteses, ou a determinar se as amostras observadas diferem, de modo significativo, dos resultados esperados, são denominados *testes de hipótese* ou de *significância*, ou *regras de decisão*.

5.4- Erro do Tipo I e II

Se uma hipótese for rejeitada quando deveria ser aceita, diz-se que foi cometido um *erro do Tipo I*. Se, por outro lado, for aceita uma hipótese que deveria ser rejeitada, diz-se que foi cometido um *erro do Tipo II*. Em ambos os casos ocorreu uma decisão errada ou um erro de julgamento.

Para que quaisquer testes de hipóteses ou regras de decisão sejam bons, eles devem ser planejados de modo que os erros de decisão sejam reduzidos ao mínimo. Isso não é tarefa simples, porquanto para um dado tamanho de amostra, a tentativa de diminuir um certo tipo de erro é acompanhada, em geral, pelo acréscimo de outro tipo. Na prática, um tipo de erro pode ser mais importante do que outro, de modo que se deve procurar uma acomodação que favoreça a limitação do erro mais sério. O único caminho para a redução de ambos os tipos de erros consiste em aumentar o tamanho da amostra, o que pode ou não ser possível.

5.5- Nível de significância

Ao testar uma hipótese estabelecida, a probabilidade máxima com a qual estaremos dispostos a correr o risco de um erro do Tipo I é denominada de *nível de significância do teste*. Essa probabilidade, representada freqüentemente por α , é geralmente especificada antes da extração de quaisquer amostras, de modo que os resultados obtidos não influenciem a escolha.

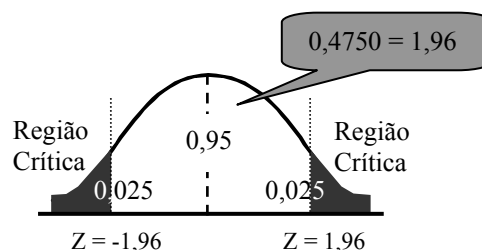
Na prática, é usual a adição de um nível de significância 0,05, ou 0,01 (5% ou 1%), embora possam ser usados outros valores.

Ex.:

Se é escolhido um nível de significância 0,05 ou 5%, no planejamento de um teste de hipótese, há então cerca de 5 chances em 100, da hipótese ser rejeitada, quando deveria ser aceita, isto é, há uma *confiança* de cerca de 95% de que se tome uma decisão acertada. Nesses casos, diz-se que a hipótese é *rejeitada no nível de significância 0,05*, o que significa que a probabilidade de erro seria de 0,05 ou 5%.

5.6- Testes que envolvem a Distribuição Normal

Para exemplificar as idéias apresentadas, admita-se que, sob uma certa hipótese, a distribuição amostral de uma estatística S é normal, com a média μ_s e desvio padrão σ_s . Então, a distribuição da variável reduzida (ou escore Z), dado por $Z = (S - \mu_s) / \sigma_s$ é a distribuição normal reduzida (com média 0 e variância 1) que esta representada abaixo.



Como está indicado na figura, pode-se estar 95% confiante de que, se a hipótese for verdadeira, o escore Z de uma estatística amostral real, S , estará compreendido entre $-1,96$ e $1,96$ (visto que a área subtendida pela curva normal, entre esses valores, é $0,95$, que dividido por $2 = 0,4750$).

Entretanto, se, ao escolher uma única amostra aleatória, fosse verificado que o escore Z dessa estatística cai *fora* do intervalo de $-1,96$ e $1,96$, concluir-se-ia que esse evento poderia ocorrer com a probabilidade de apenas $0,05$ (área total sombreada na figura) se a hipótese estabelecida fosse verdadeira. Dir-se-ia, então, que esse escore Z difere de modo significativo do que seria esperado daquela hipótese, e se estaria propenso a rejeitá-la.

A área total sombreada, de $0,05$ ou 5% , é o nível de significância do teste. Ela representa a probabilidade de incorrer-se em erro na rejeição da hipótese, isto é, a probabilidade de ser cometido um erro do Tipo I. Por essa razão diz-se que a hipótese é *rejeitada no nível de significância* $0,05$, ou que o escore Z da estatística amostral dada é *significativo* naquele nível.

O conjunto dos escores Z , situados fora do intervalo de $-1,96$ a $1,96$, constitui a denominada *região crítica de rejeição da hipótese* ou *de região de significância*. O conjunto dos escores Z , compreendidos no intervalo de $-1,96$ a $1,96$ poderia, então, ser denominado *região de aceitação da hipótese* ou *de região de não-significância*.

Com base nas observações apresentadas, pode ser formulada a seguinte *regra de decisão, teste de hipóteses* ou *significância*.

- (a) Rejeição da hipótese no nível de significância $0,05$, quando o escore Z da estatística S situar-se fora do intervalo de $-1,96$ a $1,96$ (isto é, $Z > 1,96$ ou $Z < -1,96$). Isso equivale a dizer que a estatística amostral observada é significativa no nível $0,05$.
- (b) Aceitação da hipótese (ou, se for desejado, não tomar nenhuma decisão) no caso contrário.

Como o escore Z representa papel tão importante nos testes de hipóteses e na significância, ele é também denominado *teste estatístico*.

Ex.:

Se for adotado o nível $0,01$ (1%), substituir-se-á, em toda a explanação anterior, $1,96$ por $2,58$ (veja tabela de valores críticos de Z da página seguinte).

OBS.: Deve-se assinalar que poderiam ser utilizados outros níveis de significância.

5.7- Testes unilaterais e bilaterais

Nos testes anteriores, manifestou-se interesse nos valores extremos da estatística S , ou nos escores Z correspondentes de ambos os lados da média, isto é, em ambas as “extremidades” da distribuição. Por esta razão, esses testes são denominados *bilaterais* ou *dos dois lados*.

Muitas vezes, entretanto, pode-se ter interesse apenas nos valores extremos de um único lado da média, isto é, em uma “extremidade” da distribuição, como por exemplo, quando se está testando a hipótese de um processo ser **melhor** do que o outro (o que é diferente de testar se um processo é **melhor** ou **pior** do que outro). Esses testes são denominados *unilaterais* ou de *um lado*. Nesses casos, a região crítica está situada de um só lado da distribuição e sua área é igual ao nível de significância.

A tabela da página seguinte, que nos dá os valores críticos de Z para ambos os testes, unilateral e bilateral, em vários níveis de significância, pode revelar-se útil como referência. Os valores críticos de Z , para outros níveis de significância, são determinados mediante o emprego das tabelas das áreas da curva normal.

ALGUNS VALORES CRÍTICOS DE Z

Nível de Significância α	0,10	0,05	0,01	0,005	0,002
Valores críticos de z para testes unilaterais	-1,28 ou 1,28	-1,645 ou 1,645	-2,33 ou 2,33	-2,575 ou 2,575	-2,88 ou 2,88
Valores críticos de z para testes bilaterais	-1,645 e 1,645	-1,96 e 1,96	-2,575 e 2,575	-2,81 e 2,81	-3,08 e 3,08

Exercícios de aplicação:

01) Determinar a probabilidade de obter-se entre 40 e 60 caras, inclusive, em 100 lances de uma moeda honesta.

Solução:

De acordo com a distribuição binomial, a probabilidade desejada é:

$$P(\text{entre } 40 \text{ e } 60) = C_{100}^{40} \cdot 0,5^{40} \cdot 0,5^{60} + C_{100}^{41} \cdot 0,5^{41} \cdot 0,5^{59} + \dots + C_{100}^{60} \cdot 0,5^{60} \cdot 0,5^{40} \quad (\text{muito longo})$$

Como $np = 100 \cdot 0,5$ e $nq = 100 \cdot 0,5$ são ambos maiores do que 5, pode-se empregar, para o cálculo dessa soma, o ajustamento normal à distribuição binomial.

A média, que é igual a esperança matemática e o desvio padrão do número de caras, em 100 lances, são dados por:

$$\mu = np = 100 \cdot 0,5 = 50 \quad \text{e} \quad \sigma = \sqrt{npq} = \sqrt{100 \cdot 0,5 \cdot 0,5} = 5$$

Considerando-se a escala contínua, o intervalo entre 40 e 60 caras, inclusive, é o mesmo que entre 39,5 e 60,5 caras.

$$\text{Calculando } Z \text{ temos: } Z = \frac{X_i - \mu}{\sigma} = \frac{39,5 - 50}{5} = -2,10 \quad \text{e} \quad Z = \frac{60,5 - 50}{5} = 2,10$$

A probabilidade desejada é igual a área subtendida pela curva normal entre $Z = -2,10$ e $Z = 2,10$. Portanto a resposta será $0,4821 + 0,4821 = 0,9642 = 96,42\%$.

02) Para testar a hipótese de que uma moeda é honesta, adotou-se a seguinte regra de decisão:

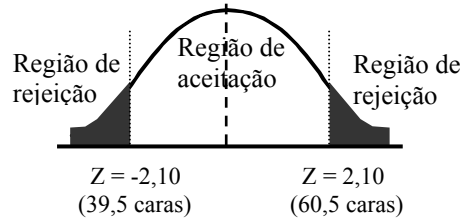
- (1) aceitar a hipótese, se o número de caras, em uma única amostra de 100 lances, estiver entre 40 e 60, inclusive;
- (2) rejeitá-la em caso contrário.

Pede-se:

- a- determinar a probabilidade de ser rejeitada a hipótese, quando ela for realmente correta;
- b- interpretar graficamente a regra de decisão e o resultado do item a;
- c- que conclusões se poderiam tirar do fato de uma amostra de 100 lances apresentar 53 caras? 60 caras?
- d- poder-se-ia estar errado nas conclusões do item c? Explicar.

Solução:

- a- De acordo com o exercício 01, a probabilidade de não se obter entre 40 e 60 caras, inclusive, quando a moeda é honesta, é igual a $1 - 0,9642 = 0,0358$. Então, a probabilidade da hipótese ser rejeitada quando ela é correta é igual a 0,0358 ou 3,58%.
- b- A regra de decisão é ilustrada pela figura abaixo, que mostra a distribuição de probabilidades das caras em 100 lances de uma moeda honesta.



Se uma única amostra de 100 lances resultar num escore Z , compreendido entre $-2,10$ e $2,10$, aceitar-se-á a hipótese; no caso contrário, ela será rejeitada e decidir-se-á que a moeda é viciada.

O erro cometido ao rejeitar a hipótese, quando deveria ser aceita, é do Tipo I da regra de decisão e a probabilidade de cometê-lo é igual a 0,0358, conforme o item (a), e é representada pela área total sombreada da figura.

Se em uma única amostra de 100 lances ocorrer um número de caras, cujo escore Z (ou estatística Z) caia na região sombreada, dir-se-á que o escore Z difere de maneira *significativa* do que seria esperado se a hipótese fosse verdadeira. Por esta razão, a área total sombreada (isto é, a probabilidade de um erro Tipo I) é denominada *nível de significância* da regra de decisão e é igual a 0,0358, neste caso. Portanto, fala-se em rejeição da hipótese do nível de significância 0,0358 ou 3,58%.

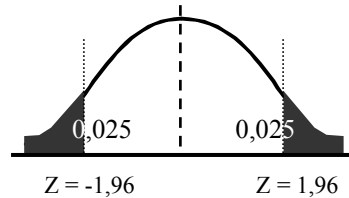
- c- De acordo com a regra de decisão, deve-se aceitar a hipótese da moeda ser honesta, em ambos os casos. Pode-se argumentar que, se apenas mais uma cara fosse obtida, teríamos rejeitado a hipótese. Isso é o que se tem de enfrentar, quando é utilizada uma linha tênue de divisão na tomada de decisões.
- d- Sim. Poder-se-ia aceitar a hipótese quando deveria ser rejeitada, e esse seria o caso, por exemplo, se a probabilidade de caras fosse, realmente, de 0,7 em vez de 0,5. O erro cometido ao aceitar a hipótese, quando deveria ser rejeitada, é do Tipo II da decisão.
- 03) Planejar uma regra de decisão para testar a hipótese de uma moeda ser honesta, quando é considerada uma amostra de 64 lances e é adotado o nível de significância:
- a- 0,05;
- b- 0,01.

Solução:

- a- **Primeiro método:** se o nível de significância é 0,05, cada área sombreada da figura a seguir é de 0,025, por simetria. Então, a área entre 0 e $Z_1 = 0,5000 - 0,0250 = 0,4750$ e $Z_1 = 1,96$.

Então, uma regra de decisão possível será:

- 1º) Aceitar a hipótese de que a moeda é honesta quando Z estiver compreendido entre $-1,96$ e $1,96$.
- 2º) Rejeitá-la no caso contrário.



OBS.: Os valores críticos $-1,96$ e $1,96$, podem também ser retirados da Tabela de Valores da Curva Normal, ANEXO 2.

Para exprimir essa regra de decisão em relação ao número de caras a serem obtidas em 64 lances da moeda, note-se que a média e o desvio padrão da distribuição das caras são dadas por:

$$\mu = np = 64 \cdot 0,5 = 32 \text{ e } \sigma = \sqrt{npq} = \sqrt{64 \cdot 0,5 \cdot 0,5} = 4, \text{ para a hipótese de que a moeda é honesta.}$$

$$\text{Então, } Z = \frac{X_i - \mu}{\sigma} \Rightarrow -1,96 = \frac{X_i - 32}{4} = 24,16 \text{ e } 1,96 = \frac{X_i - 32}{4} = 39,84.$$

Portanto, a regra de decisão torna-se:

- (1) Aceitar a hipótese de que a moeda é honesta, quando o número de caras estiver compreendido entre 24,16 e 39,84, isto é, entre 25 e 39, inclusive.
- (2) Rejeitá-la, no caso contrário.

Segundo método: Com a probabilidade de 0,95, o número de caras estará situado entre $\mu - 1,96\sigma$ e $\mu + 1,96\sigma$, isto é, $np - 1,96\sqrt{npq}$ e $np + 1,96\sqrt{npq}$, ou entre $32 - 1,96 \cdot (4) = 24,16$ e $32 + 1,96 \cdot (4) = 39,84$, o que conduz à regra de decisão anterior.

b- Se o nível de significância é 0,01, cada área sombreada da figura anterior é 0,005. Então, a área entre 0 e $Z_1 = 0,5000 - 0,0050 = 0,4950$ e $Z_1 = 2,575$.

Adotando-se o processo do segundo método, para o item (a), vê-se que, na probabilidade de 0,99, o número de caras estará situado entre $\mu - 2,575\sigma$ e $\mu + 2,575\sigma$, isto é, $32 - 2,575 \cdot (4) = 21,70$ e $32 + 2,575 \cdot (4) = 42,30$.

Em conseqüência, a regra de decisão torna-se:

- (1) Aceitar a hipótese se a quantidade de caras estiver compreendido entre 22 e 42, inclusive.
- (2) Rejeitá-la, no caso contrário.

04) Em uma experiência sobre a Percepção Extra-Sensorial (PES) um indivíduo (sujeito), em uma sala, é solicitado a declarar a cor **vermelha** ou **preta** de uma carta escolhida, de um baralho bem embaralhado de 50 cartas, por outro indivíduo colocado em outra sala.

O sujeito desconhece quantas cartas vermelhas ou pretas há no baralho. Se o sujeito identifica corretamente 32 cartas, determinar se os resultados são significativos, nos níveis de significância:

- a- 0,05;
- b- 0,01.

Solução:

Se p é a probabilidade do sujeito declarar corretamente a cor de uma carta, deve-se decidir, então, entre as duas hipóteses seguintes:

$H_0 : p = 0,5 \rightarrow$ o sujeito está simplesmente adivinhando, isto é, os resultados são devidos ao acaso.

$H_1 : p > 0,5 \rightarrow$ o sujeito tem faculdades de PES.

Escolhe-se um teste unilateral, visto que não há interesse na aptidão de obter escores extremamente baixos mas, ao contrário, na de obter escores altos.

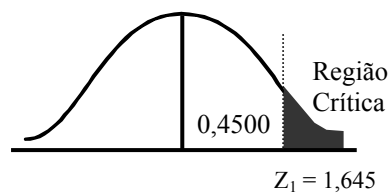
Se a hipótese H_0 for verdadeira, a média e o desvio padrão do número de cartas corretamente identificadas são dados por:

$$\mu = np = 50 \cdot 0,5 = 25 \quad \text{e} \quad \sigma = \sqrt{npq} = \sqrt{50 \cdot 0,5 \cdot 0,5} = \sqrt{12,5} = 3,54.$$

(a) Para um teste unilateral no nível de significância 0,05, deve-se escolher Z_1 , na figura abaixo, de modo que a área sombreada da região crítica dos escores altos seja 0,05. Então, a área entre 0 e $Z_1 = 0,4500$ e $Z_1 = 1,645$.

Portanto, a regra de decisão ou teste de significância será:

- (1) se o escore Z observado for maior do que 1,645, os resultados serão significativos no nível 0,05, e o indivíduo tem faculdades de PES;
- (2) se o escore Z for menor do que 1,645, os resultados são devidos ao acaso, isto é, não serão significativos no nível 0,05.



Como 32, em unidades reduzidas, é igual a $Z = \frac{X_i - \mu}{\sigma} = \frac{32 - 25}{3,54} = 1,98$, superior a 1,645, a decisão (1) é válida, isto é, conclui-se que, no nível 0,05, o indivíduo tem faculdades de PES.

(b) Se o nível de significância é 0,01, então a área entre 0 e $Z = 0,4900$ e $Z_1 = 2,33$. Como 32, em unidades reduzidas, é 1,98, inferior a 2,33, conclui-se que os resultados *não são significativos* no nível 0,01.

OBS.: - Alguns estatísticos adotam a seguinte terminologia:

- 1º) **Altamente Significativos** \rightarrow resultados significativos no nível 0,01;
- 2º) **Provavelmente Significativos** \rightarrow significativos no nível 0,05, mas não no 0,01;
- 3º) **Não-significativos** \rightarrow resultados significativos em níveis superiores a 0,05.

- De acordo com esta terminologia, concluir-se-ia que os resultados da experiência são *provavelmente significativos*, de modo que se justificam investigações ulteriores do fenômeno.

- Como os níveis de significância servem de guia para a tomada de decisões, alguns estatísticos citam as probabilidades reais pertinentes.

Por exemplo, no problema anterior, como $\Pr(Z \geq 1,98) = 0,0239$, os estatísticos diriam que, baseados na experiência, as probabilidades de incidir em erro, ao concluir-se que o indivíduo tem faculdades de PES é, aproximadamente de 3 em 100.

A probabilidade citada neste caso, 0,0239, é, às vezes, denominada *nível de significância experimental ou descritiva*.

- 05) O fabricante de uma droga medicinal reivindicou que ela era 80% eficaz em curar uma alergia, em um período de 8 horas. Em uma amostra de 200 pessoas que tinham alergia, a droga, em 8 horas, curou 150 pessoas. Determinar se a pretensão do fabricante é legítima no nível de significância de 0,01.

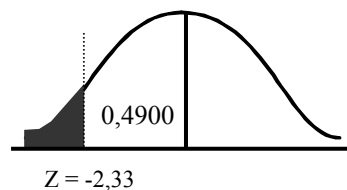
Solução:

Seja p a probabilidade de obter-se a cura da alergia, mediante o uso da droga. Então, deve-se decidir entre duas hipóteses:

$H_0 : p = 0,8 \rightarrow$ a pretensão é correta.

$H_1 : p < 0,8 \rightarrow$ ela é falsa.

Escolhe-se um teste unilateral, porque não há interesse em determinar se a proporção de pessoas curadas pela droga é muito baixa.



Como o nível de significância considerado é 0,01, isto é, se a área sombreada da figura acima é 0,01, então $Z = -2,33$. Toma-se a regra de decisão:

- (1) A pretensão não será legítima quando Z for inferior a $-2,33$ e, nesse caso, rejeita-se H_0 ;
- (2) No caso contrário, a pretensão será legítima e, então, aceita-se H_0 .

Se H_0 é verdadeira, $\mu = np = 200 \cdot 0,8 = 160$ e $\sigma = \sqrt{npq} = \sqrt{200 \cdot 0,8 \cdot 0,2} = \sqrt{32} = 5,66$.

Isso posto, 150, em unidades reduzidas $= (150 - 160) / 5,66 = -1,77$, que é maior do que $-2,33$. Portanto, de acordo com a regra de decisão, conclui-se que a pretensão é legítima e que os resultados da amostra são *altamente significativos*.

5.8- Curvas características de operação. Potência de um teste

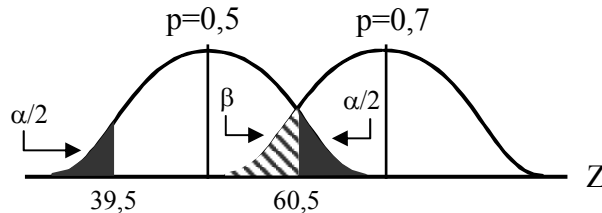
Viu-se como o erro de Tipo I pode ser adequadamente limitado, mediante a escolha do nível de significância. É possível evitar inteiramente o risco de erros do Tipo II, simplesmente pelo fato de não cometê-los, o que importa em nunca aceitar as hipóteses. Em muitos casos práticos, entretanto, isto não pode ser feito. Nesses casos, empregam-se frequentemente as *curvas características de operação* ou *curvas CO*, que são gráficos que indicam as probabilidades de erros do Tipo II, sob várias hipóteses. Elas proporcionam indicações de como testes bem aplicados podem possibilitar a redução ao mínimo de erros do Tipo II, isto é, elas indicam a *potência do teste*, para evitar que sejam tomadas decisões erradas. São úteis no planejamento de experiências por amostragem. Por exemplo: que tamanhos de amostras devem ser usados.

Exercícios de aplicação:

- 06) Com referência ao exercício 02, qual é a probabilidade de aceitação da hipótese da moeda ser honesta, quando a probabilidade real de caras for $p = 0,7$?

Solução:

A hipótese H_0 de que a moeda é honesta, isto é $p = 0,5$, é aceita quando o número de caras em 100 lances está compreendido entre 39,5 e 60,5. A probabilidade de rejeição de H_0 , quando deveria ser aceita (isto é, a probabilidade de ser cometido um erro do Tipo I) é representada pela área total α da região sombreada subentendida pela curva normal à esquerda da figura a seguir.



Como foi calculada no exercício 02 (a), essa área α , que representa o nível de significância do teste de H_0 é igual a 0,0358.

Se a probabilidade de caras for de $p = 0,7$, então a distribuição das caras nos 100 lances será representada pela curva normal à direita da figura acima. Nesse diagrama, é evidente que a probabilidade de aceitação de H_0 , quando p for realmente igual a 0,7 (isto é, a probabilidade de ser cometido um erro do Tipo II), é dada pela área hachurada β da figura. Para calcular essa área, observe-se que a distribuição, no caso da hipótese $p = 0,7$, tem a média e o desvio padrão dados por:

$$\mu = np = 100 \cdot 0,7 = 70 \quad \text{e} \quad \sigma = \sqrt{npq} = \sqrt{100 \cdot 0,7 \cdot 0,3} = 4,58.$$

$$60,5 \text{ em unidades reduzidas} = (60,5 - 70) / 4,58 = -2,07$$

$$39,5 \text{ em unidades reduzidas} = (39,5 - 70) / 4,58 = -6,66$$

$$\text{Então } \beta = (\text{área subentendida pela curva normal entre } Z = -6,66 \text{ e } Z = -2,07) = 0,0192$$

Em consequência, de acordo com a regra de decisão estabelecida, há uma probabilidade muito pequena de ser aceita a hipótese da moeda ser honesta, quando p for realmente igual a 0,7.

Note-se que, neste problema, estabeleceu-se a regra de decisão e, de acordo com ela, foram calculados α e β . Na prática, duas outras possibilidades podem surgir:

- (1) Optar por um valor de α (como 0,05 ou 0,01), chegar a uma regra de decisão e depois calcular o valor de β .
- (2) Optar por valores de α e de β e depois chegar a uma regra de decisão.

07) Resolver o exercício anterior para:

- a- $p = 0,6$;
- b- $p = 0,8$;
- c- $p = 0,9$;
- d- $p = 0,4$

Solução:

a- Se $p = 0,6$, a distribuição das caras tem a média e o desvio padrão dados por:

$$\mu = np = 100 \cdot 0,6 = 60 \quad \text{e} \quad \sigma = \sqrt{npq} = \sqrt{100 \cdot 0,6 \cdot 0,4} = \sqrt{24} = 4,90.$$

$$60,5 \text{ em unidades reduzidas} = (60,5 - 60) / 4,90 = 0,10$$

$$39,5 \text{ em unidades reduzidas} = (39,5 - 60) / 4,90 = -4,18$$

Então, $\beta = (\text{área subentendida pela curva normal entre } Z = -4,18 \text{ e } Z = 0,10) = 0,5398$

Por conseguinte, de acordo com a regra de decisão estabelecida, há uma grande possibilidade de ser aceita a hipótese da moeda ser honesta, quando $p = 0,6$.

b- Se $p = 0,8$, então $\mu = np = 100 \cdot 0,8 = 80$ e $\sigma = \sqrt{npq} = \sqrt{100 \cdot 0,8 \cdot 0,2} = \sqrt{16} = 4$.

$$60,5 \text{ em unidades reduzidas} = (60,5 - 80) / 4 = -4,88$$

$$39,5 \text{ em unidades reduzidas} = (39,5 - 80) / 4 = -10,12$$

Então, $\beta = (\text{área subentendida pela curva normal entre } Z = -10,12 \text{ e } Z = -4,88) = 0,0000$.

c- Da comparação com b, ou por meio de cálculo, verifica-se que, para $p = 0,9$, $\beta = 0$ para todas as finalidades práticas.

d- Por simetria, para $p = 0,4$, obtém-se os mesmos valores de β que para $p = 0,6$, isto é, $\beta = 0,5398$

08) Representar graficamente os resultados dos exercícios 06 e 07, mediante a construção de um gráfico de:

a- β em função de p ;

b- $(1 - \beta)$ em função de p .

Interpretar os gráficos obtidos.

Solução:

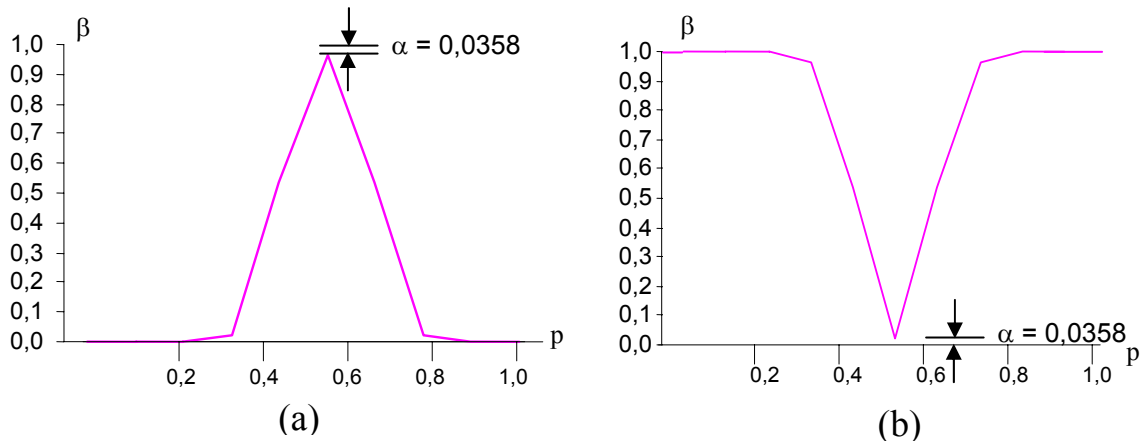
O quadro abaixo apresenta os valores de β correspondentes aos dados de p , obtidos nos exercícios 06 e 07.

p	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
β	0,0000	0,0000	0,0192	0,5398	0,9642	0,5398	0,0192	0,0000	0,0000

Note-se que β representa a probabilidade de aceitação da hipótese $p = 0,5$, quando p é realmente um valor diferente desse. Entretanto, se é realmente certo que $p = 0,5$, pode-se interpretar β como a probabilidade de aceitação de $p = 0,5$, quando ele deveria ser aceito. Essa probabilidade é igual a $1 - 0,0358 = 0,9642$ e foi anotado no quadro acima.

a- O gráfico de β em função de p , representado no gráfico a seguir (a), é denominado *curva característica de operação*, ou *curva CO* da regra de decisão ou do teste de hipótese.

A distância do ponto máximo da *curva CO* à linha $\beta = 1$ é igual a $\alpha = 0,0358$, e é o nível de significância do teste.



Em geral, quanto mais agudo for o pique da *curva CO*, mais adequada será a regra de decisão para a rejeição de hipóteses que não são válidas.

b-O gráfico de $(1 - \beta)$ em função de p , representado na figura acima (b), é denominado *curva de potência da regra de decisão* ou do teste da hipótese. Essa curva é obtida simplesmente, mediante a inversão da *CO*, de modo que, na realidade, os dois gráficos são equivalentes.

A quantidade $(1 - \beta)$ é denominada *função de potência*, porque indica a capacidade, ou *potência de um teste*, de rejeitar hipóteses falsas, isto é, que deveriam ser refugadas. A quantidade β é também denominada *função característica de operação* de um teste.

5.9- Cartas de controle

Na prática, é muitas vezes importante saber quando um processo se modifica consideravelmente, de maneira que devem ser tomadas algumas medidas para remediar a situação. Esses problemas surgem, por exemplo, no *controle de qualidade*, em que se deve, muitas vezes rapidamente decidir se as variações observadas são devidas simplesmente a flutuações ocasionais ou a variações reais do processo de fabricação, resultantes da avaria de elementos da máquina, erros de empregados etc.. As *cartas de controle* proporcionam um método simples e útil para tratar desses problemas.

Exercícios de aplicação:

- 09) Construiu-se uma máquina para produzir mancais de esfera que têm o diâmetro médio de 0,574 polegada e o desvio padrão de 0,008 polegada. Para determinar se a máquina está funcionando adequadamente, é retirada uma amostra de 6 mancais a cada 2 horas, por exemplo, e é calculado o diâmetro médio da amostra.
- Planejar uma regra de decisão, por meio da qual se poderá ter suficiente certeza de que a qualidade dos produtos está de acordo com as normas exigidas.
 - Mostrar como pode ser representada graficamente a regra de decisão do item a.

Solução:

- a- Com um grau de confiança de 99,73%, pode-se dizer que a média amostral \bar{X} deve estar compreendida no intervalo entre $(\mu_{\bar{X}} - 3\sigma_{\bar{X}})$ e $(\mu_{\bar{X}} + 3\sigma_{\bar{X}})$, ou entre $(\mu - 3\sigma/\sqrt{n})$ e $(\mu + 3\sigma/\sqrt{n})$. Como $\mu = 0,574$, $\sigma = 0,008$ e $n = 6$, segue-se que para aquele

grau de confiança, a média amostral deve estar compreendida entre $(0,574 - 0,024 / \sqrt{6})$ e $(0,574 + 0,024 / \sqrt{6})$, ou entre 0,564 e 0,584 polegada. Portanto, a regra de decisão será a seguinte:

- (1) Se a média amostral cair no intervalo entre 0,564 e 0,584 polegada, admitir-se-á que a máquina está em funcionamento normal.
- (2) No caso contrário, concluir-se-á que seu funcionamento é irregular e será preciso pesquisar a causa.

b-Pode ser mantido um registro das médias amostrais por meio de uma carta como a representada na figura abaixo, denominada *carta de controle de qualidade*. Cada vez que for calculada uma média amostral, ela será representada por um ponto particular. Enquanto eles caírem entre o limite inferior, 0,564 polegada, e o superior, 0,584 polegada, o processo está sob controle. Quando um ponto for para fora desses limites de controle (como ocorreu com a terceira amostra tomada na quinta-feira), há a possibilidade de haver alguma coisa errada, o que justifica uma investigação.

Média Amostral (cm)	Segunda-feira	Terça-feira	Quarta-feira	Quinta-feira	Sexta-feira
0,584					
	•	•	•	•	•
0,574					
	•	•	•	•	•
0,564					
	•	•	•	•	•

Os limites de controle especificados são denominados limites de confiança de 99,73% ou, abreviadamente, os limites de 3σ . Entretanto, podem ser também determinados outros limites de confiança, como os de 68% ou 95%. A escolha, em cada caso, depende de circunstâncias particulares.

EXERCÍCIOS

- 1º) Determinar a probabilidade de obter-se entre 40 e 60 caras, exclusive, em 100 lances de uma moeda honesta.
- 2º) Para testar a hipótese de que uma moeda é honesta, adotou-se a seguinte regra de decisão:
- (1) aceitar a hipótese, se o número de caras, em uma única amostra de 100 lances, estiver entre 40 e 60, exclusive;
 - (2) rejeitá-la em caso contrário.

Pede-se:

- a- determinar a probabilidade de ser rejeitada a hipótese, quando ela for realmente correta;
 - b- interpretar graficamente a regra de decisão e o resultado do item a;
 - c- que conclusões se poderiam tirar do fato de uma amostra de 100 lances apresentar 53 caras? 60 caras?
 - d- poder-se-ia estar errado nas conclusões do item c? Explicar.
- 3º) Planejar uma regra de decisão para testar a hipótese de uma moeda ser honesta, quando é considerada uma amostra de 36 lances e é adotado o nível de significância:
- a- 0,06;
 - b- 0,03.
- 4º) Em uma experiência sobre a Percepção Extra-Sensorial (PES) um indivíduo (sujeito), em uma sala, é solicitado a declarar a cor **vermelha** ou **preta** de uma carta escolhida, de um baralho bem embaralhado de 50 cartas, por outro indivíduo colocado em outra sala. O sujeito desconhece quantas cartas vermelhas ou pretas há no baralho. Se o sujeito identifica corretamente 30 cartas, determinar se os resultados são significativos, nos níveis de significância:
- a- 0,06;
 - b- 0,02.
- 5º) O fabricante de uma droga medicinal reivindicou que ela era 85% eficaz em curar uma alergia, em um período de 10 horas. Em uma amostra de 300 pessoas que tinham alergia, a droga, em 10 horas, curou 250 pessoas. Determinar se a pretensão do fabricante é legítima no nível de significância de 0,05
- 6º) Com referência ao exercício 02, qual é a probabilidade de aceitação da hipótese da moeda ser honesta, quando a probabilidade real de caras for $p = 0,55$ e $p = 0,42$?
- 7º) Construiu-se uma máquina para produzir mancais de esfera que têm o diâmetro médio de 0,555 polegada e o desvio padrão de 0,006 polegada. Para determinar se a máquina está funcionando adequadamente, é retirada uma amostra de 5 mancais a cada 1 hora, por exemplo, e é calculado o diâmetro médio da amostra.
- a- Planejar uma regra de decisão, por meio da qual se poderá ter suficiente certeza de que a qualidade dos produtos está de acordo com as normas exigidas.
 - b- Mostrar como pode ser representada graficamente a regra de decisão do item a.

CAPÍTULO 6

Teoria das Pequenas Amostras

Em capítulos anteriores lançou-se mão, freqüentemente, do fato de, para amostras de tamanho $N > 30$, denominadas **grandes amostras**, as distribuições amostrais de muitas estatísticas serem aproximadamente normais, tornando-se a aproximação melhor com o crescimento de N , de modo que devem ser introduzidas as modificações convenientes.

O estudo das distribuições amostrais de estatísticas de pequenas amostras é denominado **teoria das pequenas amostras**. Entretanto, o nome mais apropriado seria **teoria exata de amostragem**, visto que os resultados obtidos são válidos tanto para as grandes como para as pequenas amostras. A seguir, as duas principais distribuições de pequenas amostras:

6.1- Distribuição “t” de Student

Trata-se de um modelo de distribuição contínua que se assemelha à distribuição normal padrão. É utilizada para inferências estatísticas, particularmente, como já foi dito, quando se tem amostras com tamanhos inferiores a 30 elementos.

OBS.: O nome ‘t’ de Student, tem origem no seu descobridor, *William Gosset (1876 - 1937)*, que era empregado da cervejaria Guinness e precisava de uma distribuição que pudesse ser utilizada com pequenas amostras. Como a cervejaria irlandesa para a qual ele trabalhava não permitia a publicação de resultados de pesquisa, Gosset publicou-os com o pseudônimo de “*Student*”, durante a primeira parte do século XX.

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

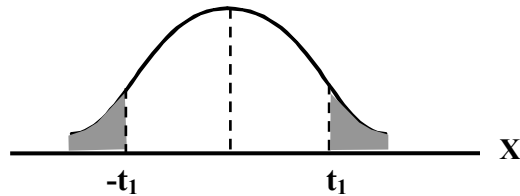
A tabela de valores de Student (ANEXO 3 – página 85), relaciona valores da distribuição ‘t’ juntamente com áreas denotadas por α .

Obtém-se valores de $t_{\alpha/2}$ na tabela, localizando o número adequado de **graus de liberdade** ($\nu = \text{letra grega ni}$) na coluna a esquerda e percorrendo a linha correspondente até atingir o número diretamente abaixo do valor aplicável (bilateral) de α .

O número de **graus de liberdade** para um conjunto de dados corresponde ao número de valores que podem variar após terem sido impostas certas restrições a todos os valores. Por exemplo, se 10 estudantes têm em um teste notas com média 80, podemos atribuir valores arbitrários a 9 delas, mas a décima fica determinada univocamente. A soma das 10 notas deve ser 800, de modo que a décima nota deve ser igual a 800 menos a soma das 9 primeiras. Como as 9 primeiras notas podem ser escolhidas arbitrariamente, dizemos que há nove **graus de liberdade**.

Para as aplicações da distribuição, o número de graus de liberdade é simplesmente o tamanho da amostra menos 1. **Graus de liberdade = $n - 1$** .

Ex.: O gráfico da distribuição de “Student” t, com 9 graus de liberdade, está representado na figura abaixo. (SPIEGEL: 289)



Determinar os valores de t_1 para os quais:

- a) a área sombreada à direita = 0,05;
- b) a área sombreada total = 0,05;
- c) a área não-sombreada total (ou em branco) = 0,99;
- d) a área sombreada à esquerda = 0,01;
- e) a área à esquerda de $t_1 = 0,90$.

Solução:

- Se a área sombreada à direita é 0,05, então a área à esquerda de t_1 é $(1 - 0,05) = 0,95$ e t_1 representa o percentil 95°, $t_{0,95}$.

Reportando-se à tabela do anexo 3, percorre-se a coluna encabeçada por v para baixo até encontrar a casa 9. Segue-se, então, para a direita até encontrar a coluna encabeçada por $t_{0,95}$. O resultado, 1,83, é o valor desejado de t .

- Se a área total sombreada é 0,05, então, por simetria, a sombreada à direita é 0,025 ($0,05 / 2$). Portanto, a área à esquerda de t_1 é $(1 - 0,025) = 0,975$, e t_1 representa o percentil 97,5%. Na tabela do anexo 3, encontra-se 2,26 para o valor desejado de t .
- Resolva → **R.: 3,25**
- Resolva → **R.: - 2,82**
- Resolva → **R.: 1,38**

Aplicação: Um ensaio das tensões de ruptura de 6 cabos produzidos por uma companhia mostrou a tensão média de ruptura de 7.750 kg e o desvio padrão de 145 kg, ao passo que o fabricante declara que aquela tensão média é de 8.000 kg.

Descubra se é verdadeira a declaração do fabricante, nos níveis de significância:

- 0,05
- 0,01 (SPIEGEL: 291)

Solução:

Deve-se primeiro decidir entre as hipóteses:

$H_0: \mu = 8.000$ kg, e a declaração do fabricante é justificada.

$H_1: \mu < 8.000$ kg, ela não o é, de modo que um teste unilateral é necessário.

Para a hipótese H_0 tem-se:

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \longrightarrow t = \frac{7.750 - 8.000}{\frac{145}{\sqrt{6}}} \longrightarrow t = -4,22$$

a) Para um teste unilateral, no nível de significância 0,05, adota-se a regra de decisão:

- (1) Aceitar H_0 , quando t é superior a $-t_{0,95}$, o qual, para $6 - 1 = 5$ graus de liberdade, significa $t > -2,02$.
- (2) Caso contrário, rejeitar H_0 .

Como $t = -4,22$, rejeita-se H_0 .

b) Para um teste unilateral, no nível de significância 0,01, adota-se, novamente a regra de decisão:

- (1) Aceitar H_0 , quando t for superior a $-t_{0,99}$, o qual, para 5 graus de liberdade, significa $t > -3,36$.
- (2) Caso contrário, rejeitar H_0 .

Como $t = -4,22$, rejeita-se H_0 .

CONCLUSÃO: é extremamente improvável que a declaração do fabricante seja justificada.

Podemos agora determinar valores da margem de erro E ao estimar μ quando se aplica uma distribuição t . Pode-se utilizar essa margem de erro para construir intervalos de confiança.

MARGEM DE ERRO para estimativa de μ

(Com base em uma amostra pequena ($n \leq 30$) e σ desconhecido)

$$E = t_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

INTERVALO DE CONFIANÇA para estimativa de μ

(Com base em uma amostra pequena ($n \leq 30$) e σ desconhecido)

$$\bar{X} - E < \mu < \bar{X} + E$$

Ex.: Com um teste destrutivo, obviamente, as amostras são destruídas nesse processo. O teste de colisão de carros é um exemplo muito dispendioso de teste destrutivo. Se você fosse o responsável por tais testes de colisão, dificilmente convenceria seu chefe da necessidade de fazer colidir e destruir mais de 30 carros, a fim de poder utilizar uma **Distribuição Normal**.

Suponha que tenhamos feito teste de colisão em 12 carros esporte Dodge Viper (preço de venda + ou - \$60.000) sob uma diversidade de condições que simulam colisões típicas. A análise dos 12 carros danificados resulta em custos de conserto que parecem ter distribuição em forma de sino com média = \$26.227 e desvio padrão = \$15.873.

Determine a estimativa intervalar de 95% de μ .

Solução:

Passaremos à construção de um intervalo de 95% de confiança utilizando a distribuição 't', porque são verificadas as condições seguintes:

- 1) a amostra é pequena ($n \leq 30$);
- 2) o desvio padrão **populacional** (σ) é desconhecido e;
- 3) a população parece ter distribuição normal, porque os dados amostrais têm distribuição em forma de sino.

Começamos determinando o valor da margem de erro conforme a seguir. Note que o valor crítico $t_{\alpha/2} = 2,20$ é obtido na tabela da Distribuição 't' na interseção da coluna rotulada 0,975 ($100\% - 95\%$ de confiança = 5% \longrightarrow $5\% / 2 = 2,5\% = 0,025 \longrightarrow 1 - 0,025$) com a linha correspondente a 11 graus de liberdade ($n - 1 = 11$).

$$E = t_{\alpha/2} \frac{S}{\sqrt{n}} = 2,20 \frac{15.873}{\sqrt{12}} = 10.080,716789$$

Podemos agora construir a estimativa intervalar de 95% para μ utilizando $E = 10.080,71$ e média igual a 26.227, fazendo:

$$\bar{X} - E < \mu < \bar{X} + E$$

$$26.227 - 10.080,71 < \mu < 26.227 + 10.080,71$$

$$\$16.146,29 < \mu < \$36.307,71$$

Com base nos resultados apresentados, temos 95% de confiança de que os limites \$16.146 e \$36.307 efetivamente contêm o valor da média populacional μ . Esses custos de conserto parecem ser bastante elevados. Na verdade, o Dodge Viper é o carro mais caro para consertar após uma colisão.

Esta informação é de grande importância para companhias que aceitam seguro de Dodge Vipers contra colisão.

Exercícios de aplicação:

- 01) Determinar os valores críticos de t para os quais a área da extremidade direita da distribuição t é 0,05, quando o número de graus de liberdade ν for igual a: (SPIEGEL:290)
 - a) 16
 - b) 27
 - c) 200
- 02) Ache o valor crítico $t_{\alpha/2}$ que corresponde ao grau de confiança e ao tamanho 'n' da amostra de dados: (TRIOLA: 155)
 - a) 99%; $n = 10$
 - b) 95%; $n = 16$
 - c) 98%; $n = 21$
 - d) 90% $n = 8$
- 03) Dados o grau de confiança e os elementos amostrais, determine: (TRIOLA: 155)
 - a) a margem de erro e;
 - b) o intervalo de confiança para a média populacional μ .

Em ambos os casos, admita que a população tenha distribuição normal.

 - a) alturas de mulheres: 95% de confiança; $n = 10$; média = 63,4 in.; $S = 2,4$ in.
 - b) médias de notas: 99% de confiança; $n = 15$; média = 2,76; $S = 0,88$
 - c) notas de testes: 90% de confiança; $n = 16$; média = 77,6; $S = 14,2$
 - d) salários: 98% de confiança; $n = 19$; média = \$23.228; $S = \$8.779$

6.2- Distribuição Qui-Quadrado (χ^2)

Em uma **população** distribuída normalmente com variância σ^2 , escolhamos aleatoriamente **amostras** independentes de tamanho **n** e calculamos a variância amostral S^2 para cada amostra. A estatística amostral tem uma distribuição chamada **distribuição Qui-quadrado**.

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

Onde: n = tamanho da amostra S^2 = Variância amostral σ^2 = Variância populacional

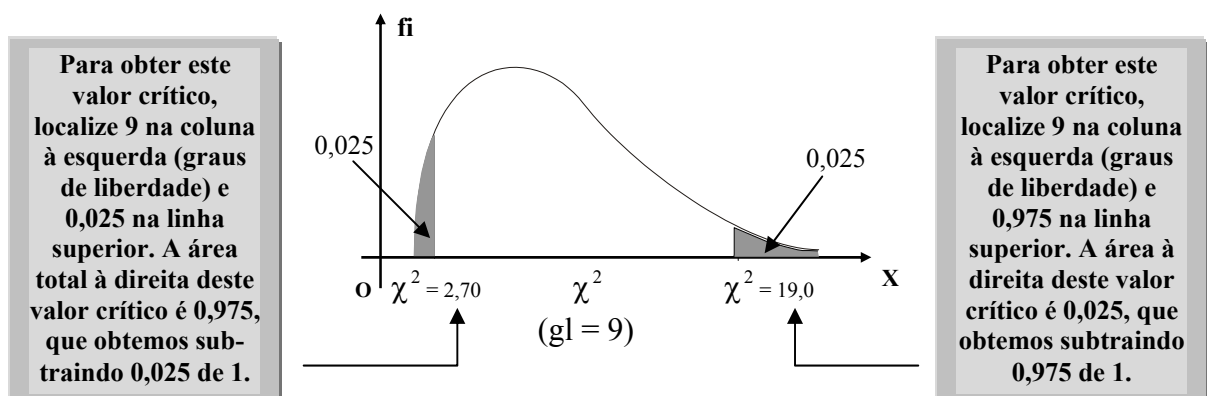
Propriedades da Distribuição da Estatística Qui-Quadrado

1. A distribuição qui-quadrado não é simétrica, ao contrário das distribuições **normal** e '**t**' de **Student**. (Na medida em que o número de graus de liberdade aumenta, a distribuição vai se tornando menos assimétrica).
2. Os valores de qui-quadrado podem ser zero ou positivos; **nunca negativos**.
3. Há uma distribuição qui-quadrado diferente para cada número de graus de liberdade.
4. À medida que o número de graus de liberdade aumenta, a distribuição qui-quadrado tende para uma distribuição normal.

Ex.: Determine os valores críticos de χ^2 que definem regiões críticas contendo uma área de 0,025 em cada cauda. Suponha que o tamanho da amostra seja 10, de modo que o número de graus de liberdade será 9, ou seja, $10 - 1$.

Solução:

Veja a figura abaixo e consulte a tabela do qui-quadrado (Anexo 4). Obtém-se o valor crítico à direita ($\chi^2 = 19,0$) diretamente, localizando 9 na coluna graus de liberdade à esquerda e 0,975 ($1 - 0,025$) na parte superior. O valor crítico $\chi^2 = 2,70$ à esquerda mais uma vez corresponde a 9 na coluna de graus de liberdade e 0,025 ($1 - 0,975$) na parte superior, porque os valores no topo são sempre **áreas à direita** do valor crítico. Verifique na figura abaixo, que a área total à direita de $\chi^2 = 2,70$ é 0,975.



**INTERVALO DE CONFIANÇA
(ou Estimativa Intervalar) para a Variância Populacional σ^2**

$$\frac{(n-1)S^2}{\chi_D^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_E^2}$$

Ex.: Uma confeitaria fabrica sonhos que são embalados em pacotes com a indicação de que há 12 sonhos pesando um total de 42 oz. Se a variação de peso entre os sonhos é muito grande, algumas caixas terão peso a menos (prejudicando o consumidor) e outras terão peso a mais (diminuindo o lucro da confeitaria). O Supervisor de Controle de Qualidade constatou que esses problemas podem ser evitados se os sonhos tiverem um peso médio de 3,50 oz e um desvio-padrão de 0,06 oz ou menos. Seleccionam-se aleatoriamente, na linha de produção, 12 sonhos, que são pesados, dando os resultados a seguir.

Construa dois intervalos de confiança de 95%, um para σ^2 e outro para σ e, determinar se o supervisor de controle está com problemas.

3,58 3,50 3,68 3,61 3,40 3,39 3,43 3,52 3,66 3,50 3,36 3,42

Solução:

Com base nos dados amostrais, a média = 3,504 parece satisfatória, pois está muito próxima do valor desejado de 3,50 oz. Os valores dados acusam um desvio-padrão = 0,109, superior ao valor desejado de 0,06 oz ou menos.

Passemos à construção do intervalo de confiança para σ^2 .

Com uma amostra de 12 valores, temos 11 graus de liberdade. Com um grau de confiança de 95%, dividimos $\alpha = 0,05$ igualmente entre as duas caudas da distribuição χ^2 e localizamos os valores 0,975 e 0,025 na linha superior. Os valores críticos de χ^2 são $\chi_E^2 = 3,82$ e $\chi_D^2 = 21,9$.

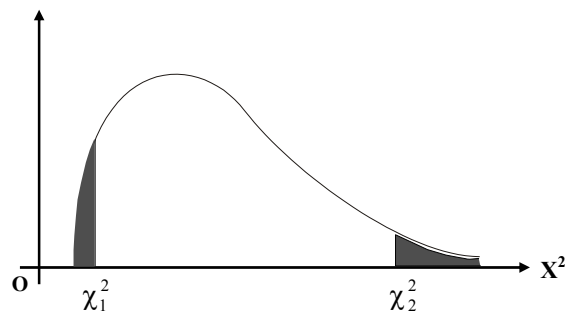
Com esses valores críticos, o desvio-padrão amostral $S = 0,109$ e o tamanho 12 da amostra, construímos como segue o intervalo de 95% de confiança.

$$\frac{(12-1)(0,109)^2}{21,9} < \sigma^2 < \frac{(12-1)(0,109)^2}{3,82}$$

Esta expressão se reduz a $0,006 < \sigma^2 < 0,034$. Tomando a raiz quadrada de cada membro, vem $0,077 < \sigma < 0,184$. Com base no intervalo de confiança de 95%, parece que o desvio-padrão é superior ao valor desejado de 0,06 oz; surge assim um problema para o Supervisor de Controle de Qualidade, que deve tomar providências corretivas para que os pesos dos sonhos sejam mais consistentes.

Exercícios de aplicação:

- 01) Ache os valores críticos χ_E^2 e χ_D^2 que correspondem ao grau de confiança e ao tamanho da amostra de dados: (TRIOLA: 165)
- a) 95%; n = 26 b) 99%; n = 17 c) 90%; n = 60 d) 95% n = 50
- 02) O gráfico da distribuição de qui-quadrado com 5 graus de liberdade está representado na figura da página seguinte. Determinar os valores críticos de χ^2 , para os quais: (SPIEGEL: 293)
- a) a área sombreada à direita = 0,05;
 b) a área total sombreada = 0,05;
 c) a área sombreada à esquerda = 0,10;
 d) a área sombreada à direita = 0,01.



- 03) Determinar os valores medianos de χ^2 correspondentes aos graus de liberdade:
- a) 9 b) 28 c) 40 (SPIEGEL: 295)

OBS.: é interessante notar que os valores medianos são aproximadamente muito iguais aos números de graus de liberdade.

De fato, para $\nu > 10$, os valores medianos são iguais a $(\nu - 0,7)$, como se pode verificar na tabela.

EXERCÍCIOS

DISTRIBUIÇÃO 't' DE STUDENT

- 1º) Os coeficientes de confiança de 95% (bilateral), para a distribuição normal, são dados por $\pm 1,96$. Quais serão os coeficientes correspondentes para a distribuição t , quando: (SPIEGEL: 290)
- a) $v = 9$ b) $v = 20$ c) $v = 30$ d) $v = 60$
R.: a) 2,26 b) 2,09 c) 2,04 d) 2,00
- 2º) Uma amostra constituída de 12 medidas de tensão de ruptura de um fio de algodão apresentou a média de 7,38 kg e o desvio padrão de 1,24 kg. Determinar os limites de confiança, para a tensão de ruptura real, de: (SPIEGEL: 299)
- a) 95%; b) 99%.
R.: a) 7,38 + ou - 0,79 kg b) 7,38 + ou - 1,11 kg
- 3º) Registraram-se os valores 0,28; 0,30; 0,27; 0,33 e 0,31 segundos, obtidos em cinco medições do tempo de reação de um indivíduo a certo estímulo. Determinar os limites de confiança, para o tempo real de reação, de: (SPIEGEL: 300)
- a) 95%; b) 99%;
R.: a) 0,298 + ou - 0,030 b) 0,298 + ou - 0,049
- 4º) As especificações para a produção de certa liga exige 23,2% de cobre. Uma amostra, constante de 10 análises do produto, apresentou o teor médio de cobre de 23,5% e o desvio padrão de 0,24%. Pode-se concluir, que o produto satisfaz às especificações exigidas, nos níveis de significância: (SPIEGEL: 300)
- a) 0,01?; b) 0,05?.
R.: Um teste bilateral, em ambos os níveis, indica que o produto não satisfaz às especificações exigidas

DISTRIBUIÇÃO QUI-QUADRADO

- 5º) Use o grau de confiança e os dados amostrais indicados para achar o intervalo de confiança para o desvio-padrão populacional σ . Em cada caso, admita que a população tenha distribuição normal. (TRIOLA: 165)
- a) alturas de mulheres: 95% de confiança; $n = 10$; média = 63,4 in.; $S = 2,4$ in.
b) médias de notas: 99% de confiança; $n = 15$; média = 2,76; $S = 0,88$
c) notas de testes: 90% de confiança; $n = 16$; média = 77,6; $S = 14,2$
d) salários: 98% de confiança; $n = 19$; média = \$23.228; $S = \$8.779$
R.: a) 1,7 in. < σ < 4,4in. b) 0,6 < σ < 1,6
c) 11,0 < σ < 20,4 d) \$6.313,81 < σ < 14.067.67
- 6º) Os valores, abaixo relacionados, são tempos de espera (em minutos) de clientes de um grande banco, onde os clientes entram em uma fila única que é atendida por três guichês. Construa um intervalo de 95% de confiança para o desvio-padrão populacional σ . (TRIOLA: 166)
- 6,5 6,6 6,7 6,8 7,1 7,3 7,4 7,7 7,7 7,7
- R.: Média = 7,15 S = 0,47667832 0,33min < σ < 0,87 min**

7º) O desvio-padrão das alturas de 16 estudantes do sexo masculino, escolhidos aleatoriamente em uma escola de 1.000 estudantes desse sexo, é 2,4 cm. Determinar os limites de confiança de: (a) 95%; e (b) 99%, do desvio-padrão para todos os estudantes do sexo masculino da escola. (SPIEGEL: 295)

R.: a) Entre 1,83 e 3,84 cm b) Entre 1,68 e 4,49 cm

8º) Considere uma distribuição qui-quadrado com parâmetro 18. Encontre: (FONSECA: 91)

a) a mediana;

b) o 1º quartil;

c) o 90º percentil

R.: a) 17,3 b) 13,7 c) 26,0

CAPÍTULO 7

Correlação e Regressão

7.1- Relação entre Variáveis

Freqüentemente procura-se verificar se existe relação entre duas ou mais variáveis. O peso pode estar relacionado com a idade das pessoas; o consumo das famílias pode estar relacionado com sua renda; as vendas de uma empresa e os gastos promocionais podem relacionar-se, bem com a demanda de um determinado produto e seu preço. A verificação da existência e do grau de relação entre variáveis é objeto do estudo da correlação.

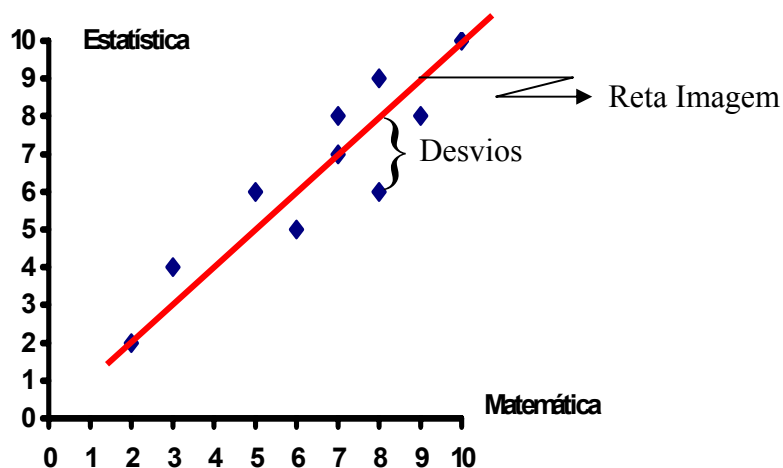
Uma vez caracterizada, procura-se descrever uma relação sob forma matemática, através de uma função. A estimação dos parâmetros dessa função matemática é o objeto da regressão

7.2- Diagrama de Dispersão

É um gráfico no qual cada ponto plotado representa um par ordenado de valores para as variáveis dependente e independente.

Exemplo:

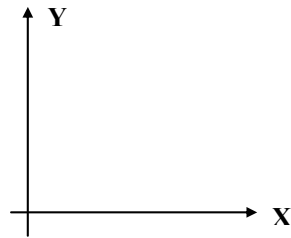
NÚMERO DO ALUNO	NOTAS	
	(X)MATEMÁTICA	(Y)ESTATÍSTICA
1	5	6
3	8	9
6	7	8
8	10	10
11	6	5
15	7	7
21	9	8
28	3	4
35	8	6
47	2	2



7.3- Tipos de Correlação Através do Diagrama de Dispersão

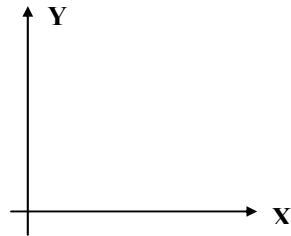
7.3.1- Correlação Linear Positiva: Quando as variáveis X e Y crescem no mesmo sentido.

Exemplo:



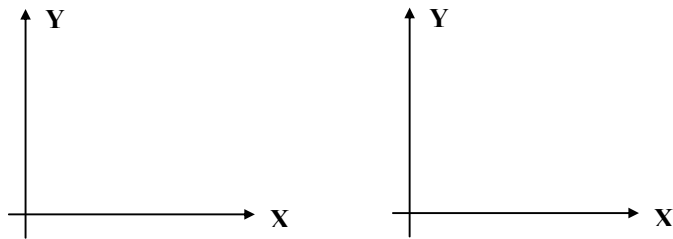
7.3.2- Correlação Linear Negativa: Quando as variáveis X e Y crescem em sentido contrário.

Exemplo:



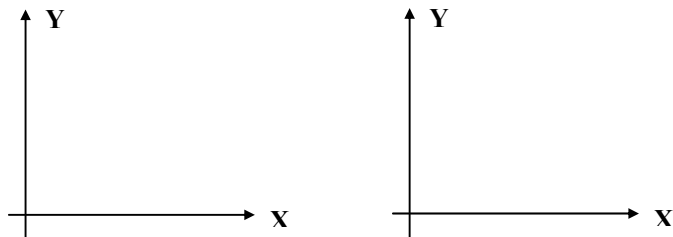
7.3.3- Correlação Nula: Quando uma variável cresce e a outra varia ao acaso

Exemplo:



7.3.4- Correlação Não Linear: Quando os pontos das variáveis se distribuem em torno de uma curva.

Exemplo:



7.4- Coeficiente de Correlação Linear de Pearson

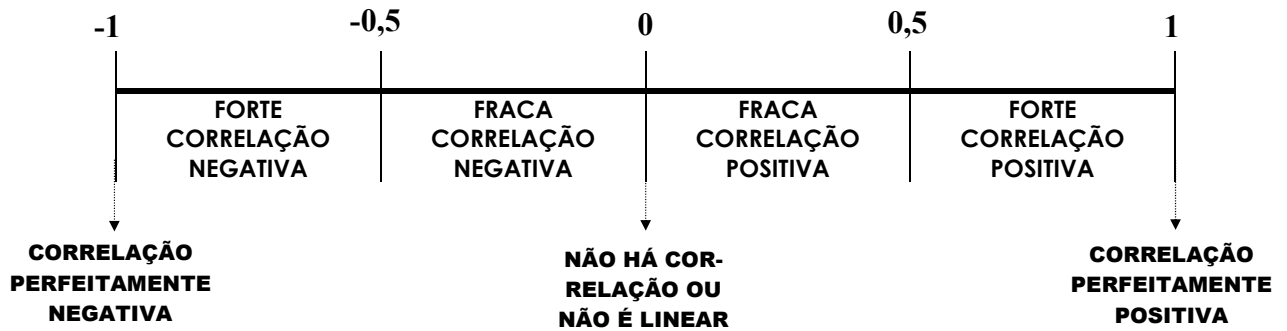
O estudo da correlação tem por objetivo medir e avaliar o grau de relação existente entre duas variáveis aleatórias. Assim, por exemplo, podemos medir se a relação entre o número de filhos de uma família e sua renda é forte, fraca ou nula.

A correlação linear procura medir a relação entre as variáveis X e Y através da disposição dos pontos (X, Y) em torno de uma reta.

Este coeficiente deve indicar o grau de intensidade da correlação entre as duas variáveis e, ainda, o sentido dessa correlação (positiva ou negativa).

$$r = \frac{n \cdot \Sigma xy - (\Sigma x) \cdot (\Sigma y)}{\sqrt{[n \cdot \Sigma x^2 - (\Sigma x)^2] \cdot [n \cdot \Sigma y^2 - (\Sigma y)^2]}}$$

Varição de r



Portanto, o campo de variação do Coeficiente 'r' situa-se entre - 1 e + 1

$$-1 \leq r \leq +1$$

Sua interpretação dependerá do valor numérico e do sinal.

Exemplo:

NÚMERO DO ALUNO	NOTAS				
	(X) MATEMÁTICA	(Y) ESTATÍSTICA			
01	5	6			
03	8	9			
06	7	8			
08	10	10			
11	6	5			
15	7	7			
21	9	8			
28	3	4			
35	8	6			
47	2	2			
Σ	65	65			

Exercício: Com base nos dados da tabela abaixo calcule r:

CARREGAMENTO AMOSTRADO	DISTÂNCIA EM KM (X)	TEMPO DE ENTREGA-DIAS (Y)			
1	825	3,5			
2	215	1,0			
3	1.070	4,0			
4	550	2,0			
5	480	1,0			
6	920	3,0			
7	1.350	4,5			
8	325	1,5			
9	670	3,0			
10	1.215	5,0			
Σ	7.620	28,5			

Observação: *Correlação não implica em causa e efeito.*

Exemplo: Quando aumenta o número de interações por desidratação, aumenta também o número de refrigerantes vendidos.

Não posso concluir, portanto, que os refrigerantes causam desidratação.

7.5- Regressão Linear Simples (Ajustamento da Reta)

Sempre que desejarmos estudar determinada variável em função de outra, deveremos fazer uma análise de regressão.

A variável sobre a qual desejamos fazer uma estimativa, recebe o nome de variável dependente (**Y**), e a outra recebe o nome de variável independente (**X**)

7.5.1- Objetivo: Descrever através de um modelo matemático a relação existente entre duas variáveis partindo de “n” observações das mesmas.

7.5.2- Equação da Reta (Y em função de X = equação que ajusta a reta)

$$\hat{Y} = a + bx$$

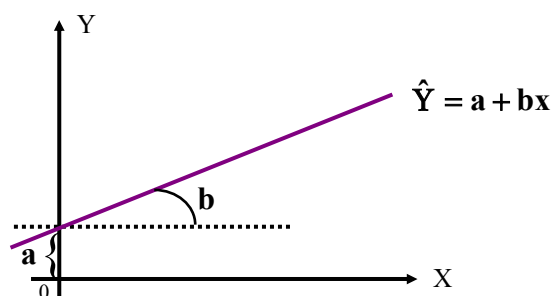
onde: \hat{Y} = valor estimado da variável dependente (Y)

a = Coeficiente linear.

É a altura onde a reta corta o eixo Y

b = Coeficiente angular.

Dá a inclinação da reta



7.5.3- Fórmulas para Calcular “a” e “b” (Método dos “mínimos quadrados” para ajustar uma linha de regressão.)

$$b = \frac{\sum xy - n \cdot \bar{x} \cdot \bar{y}}{\sum x^2 - n \cdot \bar{x}^2} \quad \text{e} \quad a = \bar{y} - b\bar{x}$$

onde: n = número de observações

$$\bar{X} = \frac{\sum X}{n}$$

$$\bar{Y} = \frac{\sum Y}{n}$$

Exemplo:

FACULDADE DE MIRANDÓPOLIS - 2ª SÉRIE - FLORIANÓPOLIS - 200X

NÚMERO DO ALUNO	NOTAS			
	(X) MATEMÁTICA	(Y) ESTATÍSTICA		
01	5	6		
03	8	9		
06	7	8		
08	10	10		
11	6	5		
15	7	7		
21	9	8		
28	3	4		
35	8	6		
47	2	2		
Σ	65	65		

7.6- Interpolação e Extrapolação

Voltando à tabela anterior, vemos que 4 não figura entre as notas de matemática, entretanto, podemos estimar a nota correspondente em estatística fazendo $X = 4$ na equação.

$$\hat{Y} = a + bx$$

Assim: $\hat{Y} = 0,91 + 0,86.(4)$

$$\hat{Y} = 4,35$$

O mesmo acontece com a nota 1. Repetindo o processo teremos:

$$\hat{Y} = a + bx$$

Assim: $\hat{Y} = 0,91 + 0,86 (1)$

$$\hat{Y} = 1,77$$

Como 4 pertence ao intervalo $\{ 2 ; 10 \}$, dizemos que foi feita uma interpolação, e como 1 não pertence a este intervalo, dizemos que foi feita uma extrapolação.

OBSERVAÇÃO: Uma norma fundamental no uso de equações de regressão é a de nunca extrapolar, exceto quando considerações teóricas ou experimentais demonstrem a possibilidade de extrapolação. Portanto, evite estimar valores de Y fora do intervalo estudado de X.

Exercício: Com base nos dados da página seguinte calcule:

- O coeficiente de correlação linear de Pearson e interprete o resultado
- A regressão linear simples através da equação da reta
- Através da equação de regressão estime a quantidade de componentes que serão rejeitados para um funcionário com 5 semanas de experiência
- Faça o gráfico com a linha de regressão.

**SEMANAS DE EXPERIÊNCIA E NÚMERO DE COMPONENTES REJEITADOS
PARA 12 TRABALHADORES ALEATORIAMENTE SELECIONADOS**

TRABALHADOR AMOSTRADO	1	2	3	4	5	6	7	8	9	10	11	12
SEMANAS DE EXPERIÊNCIA (X)	7	9	6	10	8	12	10	4	6	11	7	8
QTDE PEÇAS REJEITADAS (Y)	26	20	28	34	23	36	30	26	38	22	32	25

TRABALHADOR AMOSTRADO	SEMANAS EXPERIÊNCIA(X)	QTDE PEÇAS REJEITADAS(Y)			
1	7	26			
2	9	20			
3	6	28			
4	10	34			
5	8	23			
6	12	36			
7	10	30			
8	4	26			
9	6	38			
10	11	22			
11	7	32			
12	8	25			
Σ	98	340			

7.7- Regressão Linear Múltipla

Vamos, agora, considerar o caso em que a variável dependente (Y) seja função de duas variáveis: X_1 e X_2 . Teremos, então, o seguinte modelo de regressão:

$$\hat{Y} = a + b_1x_1 + b_2x_2$$

Esta equação, denominada de *estimador*, é a forma geral de regressão, denominada de “*linear múltipla*”.

Na regressão linear simples, em que $Y = f(x)$, tínhamos duas equações, chamadas “normais” para a determinação dos coeficientes ‘a’ e ‘b’ da função estimadora.

Na regressão linear múltipla, como temos $Y = f(x_n)$, ou seja: ‘n’ variáveis ‘x’ para a formação do estimador, teremos (n + 1) equações normais. No caso presente, teremos (2 + 1) = 3 equações normais, através das quais determinaremos os coeficientes ‘a’, ‘b₁’ e ‘b₂’:

$$\begin{aligned} \Sigma y &= na + b_1 \Sigma x_1 + b_2 \Sigma x_2 \\ \Sigma yx_1 &= a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 \\ \Sigma yx_2 &= a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 \end{aligned}$$

Exemplo: Vamos estimar o plano de regressão, considerando as seguintes variáveis:

Vendas. (y)	6	7	15	18	20	23
Gastos com TV. (x ₁)	3	4	8	8	10	11
Gastos com Jornal. (x ₂)	1	2	3	5	8	6

Elaboraremos a tabela auxiliar de cálculos:

y	x ₁	x ₂	Yx ₁	Yx ₂	x ₁ x ₂	x ₁ ²	x ₂ ²
6	3	1	18	6	3	9	1
7	4	2	28	14	8	16	4
15	8	3	120	45	24	64	9
18	8	5	144	90	40	64	25
20	10	8	200	160	80	100	64
23	11	6	253	138	66	121	36
89	44	25	763	453	221	374	139

Substituindo os valores respectivos no sistema de equações:

$$\begin{aligned} 89 &= 6a + 44b_1 + 25b_2 \\ 763 &= 44a + 374b_1 + 221b_2 \\ 453 &= 25a + 221b_1 + 139b_2 \end{aligned}$$

Aplicando o cálculo matricial, teremos:

1º) Matriz dos coeficientes:

$$\begin{bmatrix} 6 & 44 & 25 & 6 & 44 \\ 44 & 374 & 221 & 44 & 374 \\ 25 & 221 & 139 & 25 & 221 \end{bmatrix} = 311916 + 243100 + 243100 - \\ - 233750 - 293046 - 269104 = \\ = 798116 - 795900 = \mathbf{2216} = \Delta$$

2º) Matriz para encontrar-se o valor de 'a':

$$\begin{bmatrix} 89 & 44 & 25 & 89 & 44 \\ 763 & 374 & 221 & 763 & 374 \\ 453 & 221 & 139 & 453 & 221 \end{bmatrix} = 4626754 + 4404972 + 4215575 - \\ - 4235550 - 4346849 - 4666508 = \\ = 13247301 - 13248907 = \mathbf{-1606} = \Delta a$$

$$a = \frac{\Delta a}{\Delta} \longrightarrow \frac{-1606}{2216} \therefore a = -0,72$$

3º) Matriz para encontrar-se o valor de 'b₁':

$$\begin{bmatrix} 6 & 89 & 25 & 6 & 89 \\ 44 & 763 & 221 & 44 & 763 \\ 25 & 453 & 139 & 25 & 453 \end{bmatrix} = 636342 + 491725 + 498300 - \\ - 476875 - 600678 - 544324 = \\ = 1626367 - 1621877 = \mathbf{4490} = \Delta b_1$$

$$b_1 = \frac{\Delta b_1}{\Delta} \longrightarrow \frac{4490}{2216} \therefore b_1 = 2,03$$

4º) Matriz para encontrar-se o valor de ' b_2 ':

$$\begin{bmatrix} 6 & 44 & 89 & 6 & 44 \\ 44 & 374 & 763 & 44 & 374 \\ 25 & 221 & 453 & 25 & 221 \end{bmatrix} = \begin{matrix} 1016532 & + & 839300 & + & 865436 & - \\ - & 832150 & - & 1011738 & - & 877008 & = \\ 2721268 & - & 2720896 & = & \mathbf{372} & = & \Delta b_2 \end{matrix}$$

$$b_2 = \frac{\Delta b_2}{\Delta} \longrightarrow \frac{372}{2216} \therefore b_2 = 0,17$$

Como estimador é da forma $\hat{y} = a + b_1x_1 + b_2x_2$, vem que:

$$\hat{y} = -0,72 + 2,03x_1 + 0,17x_2$$

É claro que, para simplificar, tomamos um exemplo a apenas 3 variáveis, incluída a variável dependente ' y '. Todas as conclusões podem ser iguais para ' n ' variáveis.

Se o modelo tiver x_1, x_2, x_3 e x_4 , teremos que calcular ' a ', ' b_1 ', ' b_2 ', ' b_3 ' e ' b_4 ' e, para tanto, teremos um sistemas com 5 equações para calcularmos as 5 incógnitas.

EXERCÍCIO

Com base nos dados da tabela abaixo, pede-se:

- a-) determinar a equação de regressão de mínimos quadrados de y para x_1 e x_2 ;
 b-) determinar o valor de y para $x_1 = 5,4$ e $x_2 = 0,9$.

Y	6,4	7,1	5,3	6,7	5,5	5,8	7,7	5,7	5,6	6,8
x₁	5,7	5,9	4,9	6,2	5,1	5,0	5,5	4,8	5,2	5,7
x₂	0,8	1,0	0,6	1,1	0,8	0,7	1,0	0,9	1,0	0,9

y	x₁	x₂					
6,4	5,7	0,8					
7,1	5,9	1,0					
5,3	4,9	0,6					
6,7	6,2	1,1					
5,5	5,1	0,8					
5,8	5,0	0,7					
7,7	5,5	1,0					
5,7	4,8	0,9					
5,6	5,2	1,0					
6,8	5,7	0,9					
75,3	64,3	10,6					

CAPÍTULO 8

Séries Temporais

8.1- Introdução

Uma série temporal é um conjunto de observações tomadas em tempos determinados, comumente em intervalos iguais.

Ex.: Produção total anual de aço no Brasil, durante um certo número de anos;
valor diário de fechamento de uma determinada ação na Bolsa de Valores;
temperaturas horárias anunciadas pelo serviço meteorológico;
total mensal de vendas de uma loja de departamentos.

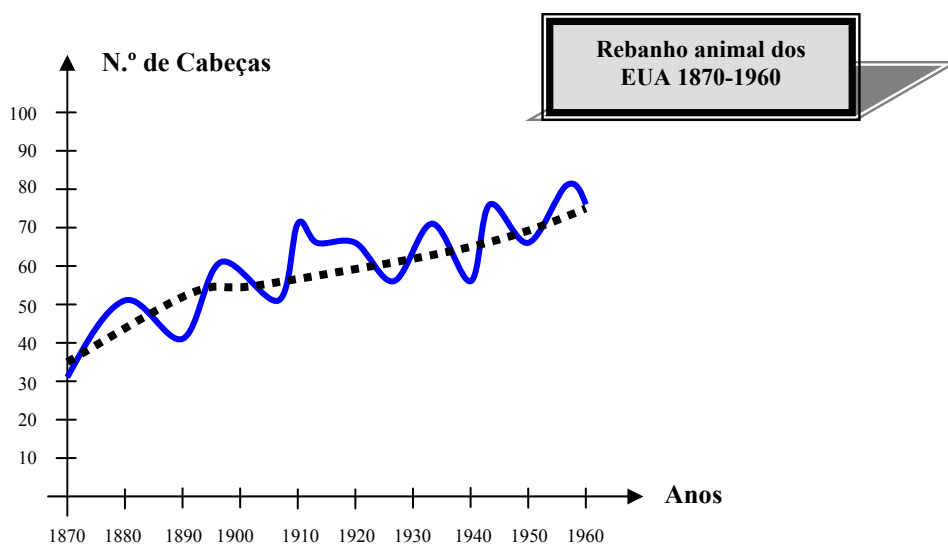
Matematicamente, uma série temporal é definida pelos valores $Y_1, Y_2 \dots$ de uma variável Y (temperatura, valor de fechamento de uma ação, etc.), nos tempos $t_1, t_2 \dots$. Portanto, Y é uma função de t simbolizada por $Y = F(t)$.

Portanto, uma série de dados estatísticos é denominada *temporal*, cronológica ou histórica, quando o elemento *tempo* varia, permanecendo fixos os elementos *fato* e seu *local de ocorrência*.

8.2- Movimentos Característicos

É interessante imaginar que o gráfico de uma série temporal, como representado abaixo, é descrito por um ponto que se move com o decorrer do tempo, de alguma forma análogo à trajetória de uma partícula material que se desloca sob a influência de forças físicas. Entretanto, o movimento pode ser provocado, em vez de forças físicas, por uma combinação de forças econômicas, sociológicas, psicológicas e outras.

Experiências realizadas com vários exemplos de séries temporais revelaram certos *movimentos* ou *variações características*, alguns dos quais, ou todos, estão presentes em graus diversos. A análise desses movimentos é de grande valor, especialmente quando se trata de *prever* movimentos futuros. Em consequência, não deve constituir surpresa o fato de muitas indústrias e setores governamentais estarem profundamente interessados nesse importante assunto.

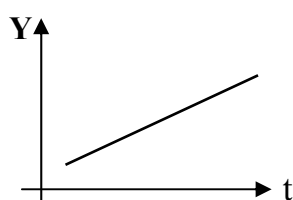


8.3- Análise de Séries Temporais

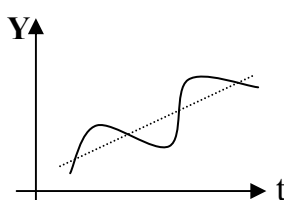
Os objetivos da análise das séries temporais, em primeiro lugar, são o de descrever seu comportamento no passado e, em seguida, analisar tal comportamento. A análise foi desenvolvida principalmente, como resultado de investigações sobre a natureza e a causa das flutuações da atividade econômica, ao que se chamou de “ciclos econômicos”. Um grande número de séries temporais revela flutuações cíclicas e, portanto, a relação entre essas séries deve ser investigada.

Para examinar qualquer série temporal, é essencial que se construa, em primeiro lugar, o seu gráfico, objetivando obter-se uma visão geral do seu comportamento.

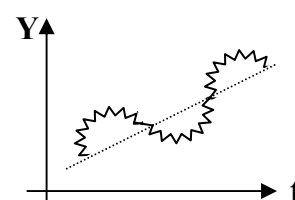
Classificam-se em quatro tipos principais os movimentos característicos das séries temporais, são eles:



(a) Tendência a Longo Prazo ou Movimentos Seculares – (T)



(b) Tendência a Longo Prazo e Movimento Cíclico – (C)



(c) Tendência a Longo Prazo, Movimentos Cíclicos e Estacionais, Sazonais – (S)

Ex.: (a) **Movimentos seculares ou de longo prazo – (T)**

Qualquer movimento de longo prazo. Indica a direção geral segundo a qual parece que o gráfico se desenvolve.

(b) **Movimentos ou variações cíclicas – (C)**

Ciclos de negócios, que representam intervalos de prosperidade, recesso, depressão e recuperação (nas atividades econômicas – cíclicos, após 01 ano).

(c) **Movimentos ou variações estacionais, sazonais – (S)**

Eventos periódicos que ocorrem anualmente – súbito aumento das vendas de uma loja, antes do Natal. (Semanal, quinzenal, mensal, bimensal, etc. < de 1 ano).

(d) **Movimentos irregulares ou aleatórios – (I)**

são os que se referem aos deslocamentos eventuais das séries temporais, provocados por eventos tais como greves, secas, enchentes, epidemias, etc.

8.4- Médias Móveis – Alisamento das Séries Temporais

Define-se uma *média móvel de ordem N*, a que é obtida pela seqüência das médias aritméticas.

Ex.: Dados os números 2, 6, 1, 5, 3, 7 e 2, uma média móvel de ordem 3 será dada pela seqüência:

$$\frac{2+6+1}{3}, \frac{6+1+5}{3}, \frac{1+5+3}{3}, \frac{5+3+7}{3}, \frac{3+7+2}{3}, \text{ isto é } 3, 4, 3, 5, 4.$$

É comum localizar-se cada número da média móvel em sua posição apropriada em relação aos dados originais. Assim, para o exemplo, escrever-se-ia:

Dados originais	2	6	1	5	3	7	2
Médias Móveis	-	3	4	3	5	4	-

As médias móveis têm a propriedade principal de tenderem a reduzir o total da variação que se apresenta em um conjunto de dados. No caso das séries temporais, essa propriedade é freqüentemente usada para eliminar flutuações indesejáveis regularizando as séries. Este processo é denominado *alisamento das séries temporais*.

Exercícios:

- 1) A que movimento característico de uma série temporal está principalmente associada cada uma das seguintes ocorrências? (SPIEGEL: 434)
- Um incêndio em uma fábrica, atrasando a produção em 3 semanas.
 - Uma era de prosperidade.
 - Uma venda posterior à páscoa, em uma loja de departamentos.
 - A necessidade de aumentar a produção de trigo devido ao acréscimo constante da população.
 - Número mensal de *cm* da precipitação da chuva, em uma cidade, durante um período de 5 anos.
- 2) Para os dados da tabela abaixo, construir uma: (SPIEGEL: 434)
- média móvel de 5 anos;
 - média móvel de 4 anos;
 - média móvel centrada de 4 anos.

Anos	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Variável	50,0	36,5	43,0	44,5	38,9	38,1	32,6	38,7	41,7	41,1	33,8

Solução da letra a:

Anos	Dados	Total móvel de 5 anos	Média móvel de 5 anos
1989	50,0	-	-
1990	36,5	-	-
1991	43,0	212,9	42,6
1992	44,5	201,0	40,2
1993	38,9	197,1	39,4
1994	38,1	192,8	39,6
1995	32,6	190,0	38,0
1996	38,7	192,2	38,4
1997	41,7	187,9	37,6
1998	41,1	-	-
1999	33,8	-	-

O primeiro total móvel, 212,9, da coluna 3 é a soma da 1ª até 5ª casa da coluna 2. O segundo, 201, é a soma da 2ª até a 6ª casa da coluna 2 e etc..

Dividindo-se cada total móvel por 5, obtém-se a média móvel desejada.

Calcule as letras “b” e “c”.

CAPÍTULO 9

Números Índices

9.1- Conceito e Objeto

Número Índice é um conceito que nos permite comparar o nível geral de magnitude de um grupo de variáveis distintas, porém interrelacionadas, em duas ou mais situações. Logo, vem a ser a designação dada a um número, destinado a representar as diferenças de uma variável, ou de um grupo de variáveis relacionadas entre si.

Os *Números Índices* podem ser usados para vários propósitos tais como para medir as variações de preços de bens de consumo, a quantidade física de mercadorias produzidas ou vendidas; ou podem relacionar-se a conceitos tais como produtividade, eficiência, inteligência, etc..

9.2- Conceito de Relativo

Quando queremos analisar a variação no preço ou na quantidade ou no valor de um só bem, basta representar um só valor em termos percentuais, obtendo o que denominamos relativo de preço de quantidade ou de valor.

9.3- Índices Relativos

É a relação entre o preço de uma única utilidade, em um *período determinado*, e o de outro período, denominado *básico* ou de *referência*.

9.4- Tipos de Números Índices

9.4.1- Relativos Simples:

São aqueles calculados através de uma “regra de três”.

Representando por **0** o período base e por **n** o período atual, temos:

a) Relativo de Preços

$$I_p = \frac{p_n}{p_o} \cdot 100$$

b) Relativo de Quantidades

$$I_q = \frac{q_n}{q_o} \cdot 100$$

onde:

p_o = preço no período base

p_n = preço no período considerado (aquele para o qual se deseja o índice)

q_o = quantidade no período base

q_n = quantidade no período considerado (aquele para o qual se deseja o índice)

Exemplo: 1) O preço de certo produto, em 1998, foi de R\$ 12,00 e, em 2000, de R\$ 13,80. Tomando-se como base o ano de 1998, calcular o preço relativo de 2000.

$$I_{00/98} = \frac{13,80}{12,00} \cdot 100 = 115$$

Logo: em 2000 houve um aumento de 15% no preço do produto, em relação ao seu preço em 1998.

2) Uma empresa produziu 45 ton de aço em 1998 e 68 ton em 1999. Qual é a quantidade relativa em 1999.

$$I_{99/98} = \frac{68}{45} \cdot 100 = 151$$

Logo: em 1999 a empresa aumentou sua produção em 51%, relativamente a 1998.

Observação:

Como a base é 100 (o ano escolhido, ou o período, ou a época), torna-se evidente que os números índices acima de 100 são acréscimos havido em relação à base. Aqueles abaixo de 100, são diminuições, através da diferença, isto é: se um número índice deu 93%, isto significa que houve uma redução de 7% em relação à base.

Exercício: Uma concessionária de veículos deseja acompanhar o desempenho de um de seus vendedores, recentemente contratado, estabelecendo um controle semanal das unidades vendidas. Os dados do primeiro mês encontram na tabela abaixo. Pede-se: (TOLEDO: 314)

- construir um índice de quantidade (relativo), tomando a primeira semana como base;
- construir um índice de quantidade, com base nas vendas da quarta semana;
- admitindo que a média semanal de vendas dos vendedores da concessionária seja 12 unidades, determinar os valores de um índice de venda com base nessa média;
- se você fosse o vendedor e, recebesse comissão pelo desempenho, medido pelo índice de quantidade, qual dos índices escolheria para receber suas comissões.

SEMANA	N.º de Veículos Vendidos	RELATIVOS DE VALOR		
		Venda 1ª Semana	Venda 4ª Semana	Venda Média 12 Unidades
1	6			
2	9			
3	9			
4	15			

c) Valor Relativo:

se “p” é o preço de uma utilidade durante um período e “q” é a quantidade ou volume produzido, vendido, etc., durante este período, “p . q” é denominado de “valor total”.

Exemplo: Se 1.000 peças de automóveis são vendidas a R\$30,00 cada uma, o valor total será de $30 \times 1000 = 30.000$.

Se “p_o” e “q_o” representam, respectivamente, o preço e a quantidade de uma utilidade durante um período ou ano-base e “p_n” e “q_n” representam, respectivamente, o preço e a quantidade num período ou ano considerado, os valores totais serão dados por “v_o” e “v_n”, respectivamente, e definidos por:

$$\text{Valor relativo} = \frac{v_n}{v_o} = \frac{p_n q_n}{p_o q_o} = \text{Preço Relativo} \times \text{Quantidade Relativa}$$

Exemplo: Uma indústria espera que suas vendas de peças de automóveis aumentem em 60% no próximo ano. De quanto deverá ser aumentado o preço de venda, para que a venda bruta duplique?

Temos: valor relativo = preço relativo x quantidade relativa
ou: preço relativo x 160 = 200%

Então: preço relativo $200/160 = 125$

Logo: os preços de venda devem ser aumentados de $125 - 100 = 25\%$

PROVA: Supondo que tivéssemos vendido 100 peças a R\$1,00 cada, teríamos uma receita de R\$100,00.
Atendendo ao exercício, teremos que vender 160 peças, obtendo uma receita de R\$200,00.
Pela resposta (aumentar o preço em 25%), nosso *novo* preço é R\$1,25.
Multiplicando R\$1,25 por 160 peças, chegamos ao pedido que é dobrar a venda bruta (R\$200,00).

d) Elos relativos:

consideremos a seqüência de preços relativos do tipo:

$I_{p1/0}, I_{p2/1}, I_{p3/2}, I_{p4/3}, \dots, I_{pn/(n-1)}$, em que tais relativos sejam considerados em intervalos sucessivos de tempo. A isto se denomina de “elos relativos” pois que cada relativo tem sua base no período imediatamente anterior.

Exemplo: Suponha-se que certa utilidade apresentou os seguintes preços, no período de 1997 a 2000: \$80, \$120, \$150 e \$180, respectivamente.

Pede-se os elos relativos de preços.

$$I_{p98/97} = (120/80) \times 100 = 150 \rightarrow \text{aumentou } 50\% \text{ (1998 em relação a 1997)}$$

$$I_{p99/98} = (150/120) \times 100 = 125 \rightarrow \text{1999 aumentou } 25\% \text{ relativamente a 1998}$$

$$I_{p00/99} = (180/150) \times 100 = 120 \rightarrow \text{2000 cresceu } 20\% \text{ em relação a 1999}$$

e) Relativos em Cadeia:

consideremos a seguinte seqüência de elos relativos de preços:

$I_{p1/0}, I_{p2/1}, I_{p3/2}, I_{p4/3}, \dots, I_{pn/(n-1)}$. Se desejarmos saber qual o incremento ocorrido – não entre os anos sucessivos, mas – entre, por exemplo, o último ano com base no primeiro, isto é $I_{pn/0}$, multiplicam-se, entre si, os elos relativos:

$$I_{pn/0} = I_{p1/0} \cdot I_{p2/1} \cdot I_{p3/2} \cdot I_{p4/3} \cdot \dots \cdot I_{pn/(n-1)}$$

Exemplo: Um I_q mostrou, relativamente aos períodos imediatamente anteriores, os seguintes resultados:

$$1997 = 130; \quad 1998 = 125; \quad 1999 = 140 \quad \text{e} \quad 2000 = 150$$

Pede-se o incremento havido em 2000, com base em 1996.

$$I_{q00/96} = I_{q97/96} \cdot I_{q98/97} \cdot I_{q99/98} \cdot I_{q00/99}$$

$$I_{q00/96} = 1,30 \cdot 1,25 \cdot 1,40 \cdot 1,50$$

$$I_{q00/96} = 3,4125 = 341,25 \text{ portanto, } 341,25 - 100,00 = 241,25$$

Logo: houve um aumento da quantidade, em 2000, de 241,25% relativamente a 1996.

9.4.2- Agregativos Simples:

- Agregativo Simples ou Índice de Bradstreet
- Médio-artmético Simples ou Índice de Sauerbeck
- Médio-harmônico Simples
- Médio Geométrico Simples
- Mediano

Nota: a tabela abaixo servirá de base para o cálculo dos índices acima especificados e os “Agregativos Ponderados”, a seguir.

MERCADORIAS	Ano Base = 1999		Ano Considerado = 2000		Anos Típicos Quantidade
	Preço = p_0	Quantidade = q_0	Preço = p_n	Quantidade = q_n	
A	1,50	30	2,00	20	25
B	2,00	20	1,50	20	25
C	1,30	10	1,50	20	10
Σ	4,80	60	5,00	60	60

a) Agregativo Simples ou Índice de Bradstreet:

É definido pela soma das mercadorias referentes a uma época determinada, medindo-se as alterações gerais de preços, comparando-se os resultados obtidos para as diferentes épocas.

Sua fórmula é dada por:

$$I_p = \frac{\sum p_n}{\sum p_0} \cdot 100 \quad \text{e} \quad I_q = \frac{\sum q_n}{\sum q_0} \cdot 100$$

Exemplo: Com os dados da tabela, fazendo 1999 = 100%, temos:

$$I_p = 5,00/4,80 = 1,042 \cdot 100 = 104,2$$

Conclusão: os preços das mercadorias A, B e C, em 2000 foram maiores em 4,2%, relativamente aos preços destas mesmas mercadorias em 1999.

b) Médio-Aritmético:

Consiste na redução de cada preço cotado à forma de um relativo, referido ao preço da mesma mercadoria, em certa época, tomando-se depois a média aritmética desses números relativos. O critério ou método de cálculo é também chamado de “Índice de Sauerbeck” e sua fórmula é dada por:

$$I_p = \frac{\sum \left(\frac{p_n}{p_0} \right)}{N} \quad \text{e} \quad I_q = \frac{\sum \left(\frac{q_n}{q_0} \right)}{N}$$

Onde: “N” = número de relativos (ou número de observações ou número de itens)

MERCADORIAS	1999		2000	
	p_0	Relativo = p_0/p_0	p_n	Relativo = p_n/p_0
A	1,50	100	2,00	133,3
B	2,00	100	1,50	75,0
C	1,30	100	1,50	115,4
Σ	- - - - -	- - - - -	- - - - -	323,7

Então: $I_p = 323,7 / 3 = 107,9$.

Conclusão: houve um acréscimo de 7,9%

c) Médio-Harmônico:

a média harmônica de um conjunto de valores x_1, x_2, \dots, x_n é o recíproco da média aritmética dos recíprocos desses valores. Assim, o recíproco de x_1 é $1/x_1$ e, assim, sucessivamente. Então: a média harmônica do conjunto é:

$$H = \frac{N}{\sum \left(\frac{1}{x_1} \right)}$$

Os termos focalizados no presente caso são preços relativos, da forma p_n/p_o . Logo, para o método, a fórmula é dada por:

$$I_p = \frac{N}{\sum \left(\frac{p_o}{p_n} \right)} \quad \text{e} \quad I_q = \frac{N}{\sum \left(\frac{q_o}{q_n} \right)}$$

Exemplo: Com os dados da tabela, temos:

MERCADORIAS	1999		2000	
	Relativo	Recíproco do Relativo	Relativo	Recíproco do Relativo
A	100	0,01 = (1/100)	133,3	0,0075 = 1/(133,3)
B	100	0,01	75,0	0,0133
C	100	0,01	115,4	0,0087
Σ	-----			0,0295

Então: $I_p = 3 / 0,0295 = 101,7$. Portanto, houve um acréscimo de 1,70%.

d) Médio-Geométrico:

é um processo muito utilizado e sua fórmula é dada por:

$$I_p = \sqrt[N]{\pi \cdot \left(\frac{p_n}{p_o} \right)}$$

Torna-se evidente que, em casos de formulações deste tipo, o emprego de logaritmos vem simplificar os cálculos; neste caso, teríamos:

$$\log I_p = \frac{\sum \log \left(\frac{p_n}{p_o} \right)}{N}$$

Exemplo: com os dados da tabela, temos:

MERCADORIAS	1999		2000	
	Relativo	log do Relativo	Relativo	log do Relativo
A	100	2,00000	133,3	2,12483
B	100	2,00000	75,0	1,87506
C	100	2,00000	115,4	2,06221
Σ	-----			6,06210

Então: $\log I_p = 6,06210 / 3 = 2,02070$

ant $\log I_p = \text{ant } \log 2,02070$, portanto, $I_p = 104,88$, daí, ocorreu um acréscimo de 4,88%.

10^{2,02070}

e) Mediano:

observemos os relativos, para o ano de 2000 (relativos de preços), que foram os seguintes:

133,3; 75,0 e 115,4. Com tais números, podemos organizar o seguinte rol crescente:

$$75,0 < 115,4 < 133,3$$

onde o valor “mediano” é “115,4”. Este valor mediano é o número índice de preços para 2000, significando aumento de 15,4% nos preços das mercadorias A, B e C, relativamente a 1999.

Notas e Conclusões:

Na construção dos cinco números índices, observa-se que não houve um critério lógico de ponderação (pesos) . Todos eles são designados, por isto, como “médias não-ponderadas” - o que, portanto, leva a uma falsa idéia da realidade.

O primeiro índice focalizado (agregativo simples) é um índice de forte ponderação, muito embora os pesos não sejam muito lógicos. Nos outros quatro tipos, os pesos são as diferentes quantidades que podiam ser adquiridas por um preço fixo de 100.

Quanto aos demais números índices, observa-se que as médias aritmética, harmônica e geométrica, apresentam, entre si, uma relação constante. Exceto quanto ao ano-base, a média harmônica é sempre menor do que a média geométrica e esta, por sua vez, é sempre menor do que a média aritmética - sendo que o valor dessas diferenças cresce na medida em que a dispersão dos preços torna-se maior.

Logo, nenhum desses números índices é perfeito, de vez que os pesos não medem, de modo lógico, a importância das diversas mercadorias incluídas nos referidos números índices.

9.4.3- Agregativos Ponderados

Na construção de números índices, representativos das variações de preços e/ou quantidades, os pesos a serem utilizados devem ser lógicos e devem refletir, ainda, com veracidade, a importância relativa das mercadorias levadas em consideração.

Não sendo considerado o problema da ponderação, não se poderá evitar a presença, reconhecida ou *não*, de pesos ilógicos e desordenados. Os pesos usados na construção de números índices, poderão ser os preços e/ou as quantidades ou os valores totais, segundo o tipo, ou critério, ou método de índice escolhido.

Os métodos e/ou critérios de cálculo de números índices agregativos ponderados são os seguintes:

- a) de Laspeyres
- b) de Paasche
- c) Índice Ideal de Fisher
- d) das médias ponderadas de relativos
 - 1º) média aritmética ponderada de relativos
 - 2º) média geométrica ponderada de relativos
- e) de Sidgwick-Drobisch
- f) de Marshall-Edgeworth
- g) das médias das quantidades dos anos típicos

a) Método ou Critério de Laspeyres:

para o índice ponderado de preços, usam-se como pesos as quantidades do ano-base:

$$I_p = \frac{\sum(p_n q_o)}{\sum(p_o q_o)}$$

Para o índice ponderado de quantidades, usam-se como pesos os preços do ano-base:

$$I_q = \frac{\sum(q_n p_o)}{\sum(q_o p_o)}$$

Exemplo: dada a tabela abaixo, calcular o I_p e o I_q , para 2000, relativamente a 1999, por Laspeyres.

MERCADORIAS	1999		2000		CÁLCULOS		
	Preço p_o	Qtde q_o	Preço p_n	Qtde q_n	$p_o q_o$	$p_n q_o$	$q_n p_o$
A	1,50	30	2,00	20	45	60	30
B	2,00	20	1,50	20	40	30	40
C	1,30	10	1,50	20	13	15	26
Σ					98	105	96

$$I_p = \frac{105}{98} = 107,10 \qquad I_q = \frac{96}{98} = 97,00$$

Os preços dos produtos A, B e C, em 2000, aumentaram de 7,1% relativamente aos preços destes produtos em 1999. Quanto às quantidades de 2000, estas diminuíram de 3%, em relação àquelas de 1999.

b) Método ou Critério de Paasche:

para o índice ponderado de preços, usam-se como pesos as quantidades do ano considerado (aquele para o qual se deseja o índice):

$$I_p = \frac{\sum(p_n q_n)}{\sum(p_o q_n)}$$

Para o índice ponderado de quantidades, usam-se como pesos os preços do ano considerado:

$$I_q = \frac{\sum(q_n p_n)}{\sum(q_o p_n)}$$

Exemplo: com os dados da tabela, calcular I_p e I_q , para 2000, com base em 1999, por Paasche.

MERCADORIAS	1999		2000		CÁLCULOS		
	Preço p_o	Qtde q_o	Preço p_n	Qtde q_n	$p_o q_n$	$p_n q_n$	$q_o p_n$
A	1,50	30	2,00	20	30	30	60
B	2,00	20	1,50	20	40	40	30
C	1,30	10	1,50	20	26	30	15
Σ					96	100	105

$$I_p = \frac{100}{96} = 104,20 \qquad I_q = \frac{100}{105} = 95,00$$

Os preços dos produtos A, B e C, em 2000, aumentaram de 4,2% em relação aos preços destes produtos em 1999, ao passo que quantidades diminuíram 5%.

Considerações sobre os critérios de Laspeyres e Paasche

1º) Os índices obtidos pelos métodos de Laspeyres e Paasche, ainda que visando medir a mesma variação, levam-nos a resultados diferentes, tendo em vista serem diferentes os critérios de ponderação. Assim o índice de Laspeyres é uma média aritmética ponderada de relativos de preços, sendo a ponderação feita com base na participação relativa (percentual)

de cada bem no valor dos bens, considerados na época-base. O índice de Paasche é calculado através de uma média harmônica dos relativos de preços, sendo o peso de cada bem considerado como sua participação relativa no valor dos bens na época atual.

- 2º) Os resultados de cada um dos índices, de Laspeyres e de Paasche, são, em geral, diferentes, quando aplicados aos mesmos dados (veja exemplos anteriores). Os resultados seriam iguais se os preços ou as quantidades de todos os bens que compõem o índice variassem na mesma proporção. Porém, na prática, nenhum desses elementos varia na mesma proporção e, assim, a relação entre os dois índices irá depender da correlação entre as suas variações. Logo, supondo que as variações dos preços e das quantidades, no tempo, resultassem de mudanças das condições de oferta, esperar-se-ia uma correlação negativa entre essas variáveis. Havendo uma elevação no nível geral de preços, os bens cujos preços tivessem sido elevados mais do que a média tenderiam a ter “ q_n ” relativamente menor do que “ q_0 ” pesando, conseqüentemente, menos no índice de Paasche do que no de Laspeyres.
- 3º) Os índices de Laspeyres e Paasche são iguais quando não houver correlação linear entre os relativos de preços e de quantidades; ou quando todos os preços ou todas as quantidades se alterarem na mesma proporção - inexistindo, conseqüentemente, dispersão de variação dos preços ou das quantidades.
- 4º) A discrepância relativa entre os índices de Laspeyres e Paasche será maior, quanto maior for o grau de correlação entre as variações de preços e de quantidades; ou, quando a dispersão dos dois grupos de variações for alta. Na prática, se os períodos comparados não forem muito distanciados, esses índices são muito próximos.

c) Índice Ideal de Fisher:

normalmente existe uma discrepância - denominada, comumente, de “erro” - entre a medida que se pretende obter, através dos números índices, e aquela que realmente se obtém. Tais discrepâncias, ou erros são, basicamente, três:

- 1º) erro de fórmula;
2º) erro de amostragem e;
3º) erro de homogeneidade.

Na prática, a medida do “erro de fórmula” é igual à diferença entre os índices de Laspeyres e Paasche.

Logo, existem dificuldades a serem sanadas e, para isto, “retificam-se” as fórmulas, mediante um processo cruzado, que consiste em se tomar a média geométrica dos erros em sentidos opostos.

Fisher realizou várias experiências com todas as fórmulas possíveis (chamadas de testes), encontrando apenas 13, em que as provas ou testes foram satisfeitos. Dentre elas, escolheu uma como sendo “ideal”, tanto do ponto de vista da exatidão como da simplicidade de cálculo.

O “índice ideal de Fisher” é expresso por uma fórmula que consiste na média geométrica dos critérios de Laspeyres e Paasche:

$$I_p = \sqrt{\frac{\Sigma(p_n q_0) \cdot \Sigma(p_n q_n)}{\Sigma(p_0 q_0) \cdot \Sigma(p_0 q_n)}}$$

NOTA: para o de quantidades, basta que se façam as modificações na simbologia.

Exemplo: com os dados da tabela e, com base em 1999, calcular o índice de preços para 2000, pelo método ideal de Fisher.

Substituindo na fórmula acima, os valores já encontrados anteriormente, para os números índices, calculados pelos critérios de Laspeyres e Paasche, obtemos:

$$I_p = \sqrt{107,10 \cdot 104,20} = 105,64$$

Logo: os preços das mercadorias A, B e C, em 2000, cresceram 5,64% relativamente aos preços destas mesmas mercadorias, em 1999.

d) Método ou Critério das Médias Ponderadas de Relativos:

para este tipo de cálculo, multiplica-se cada relativo pelo peso correspondente e divide-se a soma dos produtos pela soma dos pesos.

Nos casos anteriormente vistos, os pesos usados foram as quantidades, pois “preços x quantidades” resultam em agregados expressos em unidades monetárias. Porém, as quantidades não servem para ponderar relativos de preços; estes deverão ser ponderados por valores, a fim de possibilitar a comparação dos produtos resultantes. Isto por que os valores se exprimem por uma unidade comum, que é a monetária, ao passo que as quantidades podem ser expressas em unidades as mais diversas.

Fisher sugere os quatro processos seguintes, como fatores de ponderação:

1º) de tendência ascensional: a- $p_n q_o$: preço do ano considerado x quantidade do ano-base

b- $p_n q_n$: preço do ano considerado x quantidade do ano considerado.

2º) de tendência descensional: c- $p_o q_o$: preço do ano-base x quantidade do ano-base

d- $p_o q_n$: preço do ano-base x quantidade do ano considerado.

Os índices ponderados pelo “tipo a” devem exceder aos que tiverem como fator de ponderação o “tipo c”. Ponderando-se o relativo de um dado produto com o fator do “tipo a”, teremos:

$$\frac{p_n}{p_o} \cdot p_n q_o$$

Utilizando o sistema “c”, teremos:

$$\frac{p_n}{p_o} \cdot p_o q_o$$

Se $p_n > p_o$, isto quer dizer que $p_n / p_o > 100$: o peso indicado no “tipo a” será maior do que o fornecido pelo “tipo c”. Logo, os relativos superiores a 100 acham-se mais fortemente ponderados pelo sistema “a” do que pelo “c”.

Porém, se ocorrer que $p_n < p_o$, o fator de ponderação do sistema “a” será menor do que o do sistema “c”. Todos os relativos menores do que 100 são mais fracamente ponderados pelo “tipo a”, do que pelo “c”.

Portanto, o efeito de todos os aumentos de preços é exagerado para mais e o de todas as baixas de preços é exagerado para menos quando se usa o “tipo a”, fornecendo sempre um resultado líquido superior ao do “tipo c”.

Analogamente, podem-se comparar os tipos “b” e “d” obtendo-se os mesmos resultados.

No que se refere aos tipos “b” e “c” - embora entre eles não haja uma relação necessária - os índices ponderados pelo “tipo b” excederão, em geral, aos que o tiverem sido pelo “tipo c”.

1º) Média aritmética ponderada de relativos:

A média aritmética de relativos, ponderada com os valores do ano base, equivale ao índice agregativo ponderado com as quantidades do ano base. Se ponderada com o produto dos preços do ano base pelas quantidades do ano considerado, equivale ao índice agregativo ponderado com as quantidades do ano considerado.

Sua fórmula é a seguinte:

$$I = \frac{\sum \left(\frac{p_n}{p_o} \cdot p_o q_o \right)}{\sum (p_o q_o)}$$

Exemplo:

MERCADORIAS	Relativo de 1999	Peso $p_o q_o$	Relativo de 1999 x peso	Relativo de 2000 = p_n / p_o	Relativo de 2000 x peso
A	100	45	4500	133,3	5.998,5
B	100	40	4000	75,0	3.000,0
C	100	13	1300	115,4	1.500,2
Σ	-	98	-	-	10.498,7

$$I_p = \frac{10.498,7}{98} = 107,1$$

2º) Média Geométrica Ponderada de Relativos:

O processo seria o mesmo da média geométrica simples, se não se tivesse que multiplicar o “log” de cada relativo pelo respectivo peso e, dividir a soma destes “logs” ponderados pela soma dos pesos - resultando, daí, o “log” do índice procurado. Fazendo:

$p_o q_o = Q_i$ teremos:

$$I = \sqrt[{\sum Q_i}]{\pi \left(\frac{p_n}{p_o} \right)^{Q_i}}$$

ou

$$\log I = \frac{\sum Q_i \log \frac{p_n}{p_o}}{\sum Q_i}$$

Exemplo:

MERCADORIAS	Relativo de 1999	Peso $p_o q_o = Q_i$	Relativo de 2000 = p_n / p_o	$\log p_n / p_o$	$Q_i \cdot \log p_n / p_o$
A	100	45	133,3	2,12483	95,61735
B	100	40	75,0	1,87506	75,00240
C	100	13	115,4	2,06221	26,80873
Σ	-	98	-	-	197,42848

$$\log I = \frac{197,42848}{98} = 2,0145763 \quad \therefore \quad \text{o resultado será o anti log de I, que é } 103,41$$

e) Método ou Critério de Sidgwick-Drobish:

o método consiste na extração de uma média simples entre os critérios de Laspeyres e Paasche, sendo sua fórmula a seguinte:

$$I = \frac{\frac{\sum (p_n q_o)}{\sum (p_o q_o)} + \frac{\sum (p_n q_n)}{\sum (p_o q_n)}}{2}$$

Exemplo: com os dados da tabela e já tendo calculado os valores de Laspeyres e Paasche, substituindo-os na fórmula, vem:

$$I = \frac{107,10 + 104,20}{2} = 105,65$$

f) Método ou Critério de Marshall-Edgeworth:

Fisher considerou esta fórmula como a mais prática e compreensiva, por apresentar as seguintes características: exatidão, rapidez, mínimo de dispersão e simplicidade. Com o emprego dessa fórmula, os resultados diferem, comumente, em menos de 1/4% dos que são obtidos com o emprego da fórmula ideal de Fisher. Sua fórmula é dada por:

$$I = \frac{\sum p_n (q_o + q_n)}{\sum p_o (q_o + q_n)}$$

Exemplo:

MERCADORIAS	1999		2000		CÁLCULOS		
	Preço p _o	Qtde q _o	Preço p _n	Qtde q _n	q _o + q _n	p _o (q _o + q _n)	p _n (q _o + q _n)
A	1,50	30	2,00	20	50	75	100
B	2,00	20	1,50	20	40	80	60
C	1,30	10	1,50	20	30	39	45
Σ					-	194	205

$$I = \frac{205}{194} = 105,7$$

g) Método ou Critério das Médias das Quantidades dos Anos Típicos:

As fórmulas de Laspeyres e Paasche envolvem fatores de ponderação que variam quando as épocas a serem comparadas mudarem. Os pesos do índice de Paasche mudarão quando mudarem as épocas atuais; os do índice de Laspeyres, quando mudarem as épocas básicas. Dessa forma, esses dois índices são considerados como índices com ponderações variáveis, em contraposição com os índices agregativos com ponderação constante (fixa), sendo a fórmula do índice de preços dada por:

$$I_p = \frac{\sum (p_n q_t)}{\sum (p_o q_t)}$$

Exemplo:

MERCADORIAS	1999		2000		Qtde dos Anos Típicos q _t	CÁLCULOS	
	Preço p _o	Qtde q _n	Preço p _n	Qtde q _n		p _o q _t	p _n q _t
A	1,50	30	2,00	20	25	37,50	50,00
B	2,00	20	1,50	20	25	50,00	37,50
C	1,30	10	1,50	20	10	13,00	15,00
Σ						100,50	102,50

$$I = \frac{102,50}{100,50} = 102,00$$

8.5- Testes, Provas e/ou Propriedades em Números Índices

Quando tratamos do “Índice ideal de Fisher”, fizemos menção a testes em números índices. Não existe um índice considerado perfeito, ou uma fórmula universalmente aceita para quantificar, de

modo inequívoco e exato, as variações de preços e de quantidades, especialmente quando os índices se referem não a um, mas a um conjunto de bens. A literatura fornece uma variedade de métodos de cálculo de índice. A escolha do método será facilitada se houver algum critério que possibilite salientar as vantagens e as limitações de cada um deles.

Irving Fisher desenvolveu alguns testes ou critérios matemáticos muito úteis para comparar as várias fórmulas propostas de números índices.

Tais testes (propriedades) são em número de cinco e têm a finalidade de indicar-nos se o tipo de número índice empregado em um certo cálculo, é o melhor, ou o mais compatível. É evidente que aquele critério cujo índice responder a todas as provas, será o melhor e, portanto, o que deverá ser utilizado.

Ressalva-se, entretanto, que, na maioria das vezes, apenas o índice ideal de Fisher, o critério de Sidgwick-Drobish e o método de Marshall-Edgeworth satisfazem aos dois testes mais importantes, que são os de “reversão no tempo” e “reversão de fatores”.

Os testes são os seguintes:

1º) *Identidade*

O preço e/ou a quantidade relativo(a) deve ser igual a unidade quando a época dada (n) coincidir com a época básica (0). Por essa razão, quando se pretende identificar determinada data como básica costuma-se igualá-la a 100.

Exemplo: Suponhamos que um índice de preços calculado para o ano de 1999 tenha dado 200. Aplicando a propriedade:

$$\frac{I_{p99}}{I_{p99}} = \frac{200}{200} = 1,00 = 100\%$$

2º) *Reversão (Inversão) do Tempo*

Permutando-se dois períodos, os preços e/ou as quantidades relativos(as) correspondentes serão recíprocos, ou seja o inverso um do outro. Poucos índices satisfazem a essa condição.

$$I_{pn/o} = \frac{1}{I_{po/n}} \quad \therefore \quad I_{pn/o} \cdot I_{po/n} = 1$$

(n/o = ano considerado/ano-base; o/n = ano-base/ano considerado)

1º) Exemplo: Comprovar a propriedade de inversão do tempo usando os dados do exemplo da “concessionária de veículos”. (p. 45)

Os valores dos relativos necessários para a comprovação do teste são os seguintes:

$$q_{1,4} = \frac{q_4}{q_1} = \frac{15}{6} = 2,50$$

$$q_{4,1} = \frac{q_1}{q_4} = \frac{6}{15} = 0,40$$

Portanto $q_{1,4} \times q_{4,1} = 2,50 \times 0,40 = 1,00$

Os relativos satisfazem ao teste de inversão de tempo

2º) Exemplo: Suponhamos que tenhamos os seguintes índices de preços: $I_{p99} = 110$ e $I_{p00} = 120$
Aplicando a propriedade, vem:

$$\frac{120}{110} = \frac{1}{\frac{110}{120}} \quad \therefore \quad \frac{120}{110} \cdot \frac{110}{120} = 1$$

3º) *Reversão de Fatores*

Se, com um certo número de mercadorias, construir-se um índice de preços e um índice de quantidade, é de se esperar que: o seu produto seja igual a razão dos valores, ou seja:

$$I_p \cdot I_q = \frac{\sum(p_n q_n)}{\sum(p_0 q_0)} = \frac{\sum v_n}{\sum v_0}$$

Exemplo: Suponha-se que um certo método de cálculo de número índice tenha apresentado os seguintes resultados:

$$I_p = 105,64; \quad I_q = 96,60; \quad \sum v_n = 100,00 \quad \text{e} \quad \sum v_0 = 98,00$$

Um outro método aplicado apresentou os seguintes resultados:

$$I_p = 104,17; \quad I_q = 95,24; \quad \sum v_n = 102,00 \quad \text{e} \quad \sum v_0 = 99,00$$

Verificar qual dos dois critérios aplicados é o melhor.

Como se tomam os índices em razão de valores, significa que eles serão tomados divididos por 100. Então:

$$1,0564 \cdot 0,9660 = 100 / 98 \rightarrow 1,0204 = 1,0204, \text{ portanto, há compatibilidade.}$$

$$1,0417 \cdot 0,9524 = 102 / 99 \rightarrow 0,9921 \neq 1,0303, \text{ portanto, há incompatibilidade.}$$

Logo, o primeiro método de cálculo de número índice é o melhor.

4º) *Cíclica ou Circular*

O produto dos elos relativos de preços e quantidades, é igual à unidade - isto é:

$$I_{pb/a} \cdot I_{pc/b} \cdot I_{pa/c} = 1$$

Exemplo: Suponhamos que tenhamos os seguintes índices de preços:

$$I_{p97} = 110; \quad I_{p98} = 120; \quad I_{p99} = 150; \quad I_{p00} = 200$$

Aplicando a propriedade, vem:

$$I_{p98/97} \cdot I_{p99/98} \cdot I_{p00/99} \cdot I_{p97/00} = \frac{120}{110} \cdot \frac{150}{120} \cdot \frac{200}{150} \cdot \frac{110}{200} = \frac{110}{110} = 1$$

5º) *Cíclica ou Circular Modificada*

$$I_{pb/a} \cdot I_{pc/b} \cdot I_{pd/c} = I_{pd/a}$$

Exemplo: com parte dos dados acima, teremos:

$$I_{p98/97} \cdot I_{p99/98} = I_{p99/97}, \text{ portanto, } \frac{120}{110} \cdot \frac{150}{120} = \frac{150}{110} \therefore \frac{150}{110} = \frac{150}{110} \therefore 1,364 = 1,364$$

9.6- Mudança de Base em Séries de Números Índices

9.6.1- Construção

Há necessidade, muitas vezes, de se construir séries de números índices. Porém, a escolha da época base, a partir da qual os números índices são construídos, podem gerar dúvidas.

Dois aspectos importantes devem ser considerados para a construção de séries de números índices: a seleção ou escolha da época base e, a possibilidade de sua mudança, sem implicar em erro ou na independência do índice em relação à base.

A escolha da base de um número índice deve recair no período para o qual a atividade econômica não se acha sob a influência de variações estruturais de momento (variações ocasionais), sejam relativas a uma situação expansionista ou a uma retracionista. Logo, a base deve ser escolhida naqueles períodos em que as condições econômicas sejam normais.

Alem disso, pode surgir um outro problema ao se fixar o período base: “quanto maior for o lapso de tempo decorrido entre a época de comparação e a básica, maior será o perigo de que a importância relativa dos itens se tenha alterado e que, portanto, o sistema de ponderação tenha perdido sua validade”.

Dois são os critérios ou métodos:

a) Base Fixa:

Ao se construir uma série de índices e se comparar períodos afastados, ou, eventualmente, mudar a base, a fixação desta é importante de vez que os índices geralmente usados não respondem ao teste circular (página anterior). No índice de Laspeyres, por exemplo, quando se fixa a base comparativa, está se fixando, também, a base de ponderação.

Se o número índice fosse independente de sua base, esta poderia deslocar-se sem erro. Todavia, no caso de o índice depender da base, o único procedimento para deslocá-la será por meio dos dados originais; isto por que, para que o método seja correto do ponto de vista matemático, a propriedade circular (ou cíclica) deve ser satisfeita.

Os índices não ponderados, os agregativos com ponderação constante e a média geométrica de relativos são os que respondem ao teste circular.

b) Base Móvel em Cadeia:

Nesse critério ou método, a base seria modificada, de período a período. O método consiste em:

1º) construir os índices:

$$I_{1/0}, I_{2/1}, I_{3/2}, \dots, I_{n/(n-1)}$$

2º) construir um índice em cadeia, a fim de se fixar determinado período como básico:

$$I_{1/0} = I_{1/0}$$

$$I_{2/0} = I_{1/0} \cdot I_{2/1}$$

$$I_{3/0} = I_{1/0} \cdot I_{2/1} \cdot I_{3/2}$$

Os índices construídos em cadeia só coincidem com os de base fixa quando a fórmula usada responder ao teste circular, tal como acontece com a média geométrica e a média aritmética de ponderação constante.

9.6.2- Vantagens e Desvantagens

MÉTODO	VANTAGENS	DESVANTAGENS
Base Fixa	Cálculos mais simples.	Possibilidade de erros de fórmula e homogeneidade em séries longas.
Base Móvel	Fornecer medida mais correta da variação de preços, e/ou quantidades, de período a período, permitindo introduzir novos bens, eliminando outros de menor importância.	Maior volume de cálculo, sendo que as fórmulas geralmente utilizadas não satisfazem ao teste circular.

9.6.3- Exemplo Prático

A tabela abaixo apresenta o índice geral de exportação por tipo de bens (Índice de Quantum), do Brasil, de 1993 a 2000 com o ano de 1993 tomado como básico (dados fictícios):

A N O S	1993	1994	1995	1996	1997	1998	1999	2000
Índices	100	106	134	155	158	174	176	177

Pede-se obter um novo índice, adotando como base:

1º) o ano de 1996

2º) o período de 1995 a 1997

Solução do 1º item:

Dividir-se-á cada índice da tabela original por “155”, que é o índice correspondente ao novo ano-base, exprimindo-se o resultado em percentual. A tabela abaixo apresenta os resultados das divisões já multiplicadas por 100, que são os novos índices de quantum, com base em 1996:

A N O S	Índice de Quantum
1993	64,52
1994	68,39
1995	86,45
1996	100,00
1997	101,94
1998	112,26
1999	113,55
2000	114,19

Solução do 2º item:

Dividir-se-á cada índice da tabela original pelo índice médio do período 1995/97, isto é:

$$\frac{134 + 155 + 158}{3} = \frac{447}{3} = 149$$

A tabela abaixo apresenta o resultado das divisões já multiplicadas por 100, que são os novos índices de quantum com base no período 1995/1997:

A N O S	Índice de Quantum
1993	67,11
1994	71,14
1995	89,93
1996	104,03
1997	106,04
1998	116,78
1999	118,12
2000	118,79

Note-se que o índice médio do novo período base, de 1995/97, e:

$$\frac{89,93 + 104,03 + 106,04}{3} = \frac{300,00}{3} = 100, \text{ como deveria ser.}$$

9.7- Alguns Índices Especiais

1º) ICV – Índice do Custo de Vida (Índice de Preços ao Consumidor)

É um número índice, cujo objetivo é o medir o efeito das variações de preços sobre as despesas normais de uma unidade consumidora padrão.

É uma ponderação de subíndices homogêneos. Assim, o aumento percentual no preço de determinado bem, não redundará num mesmo aumento percentual do custo de vida. Tal aumento deverá ser ponderado por um fator dentro do seu respectivo subíndice e esse subítem, por sua vez, o será pelo seu peso dentro do ICV. Simbolicamente, o acréscimo no custo de vida, proveniente de um acréscimo no preço do bem “i”, é dado por:

$$\frac{\Delta \text{ICV}}{\text{ICV}} = \Delta \left[\frac{p_n^i}{p_0^i} \right] \cdot w_j^i \cdot k_j$$

Onde:

p_0^i = preço, no ano base, de “i” bens

p_n^i = preço, no ano atual, de “i” bens

w_j^i = índice de alimentação, ponderado pelos bens que o compõem

k_j = pesos dos subíndices dos bens

2º) Deflator implícito da Renda (Produto)

Considerando que:

Q = índice que descreve o comportamento da produção física de bens e serviços finais de uma economia - entre dois ou mais períodos de tempo (índice de quantum);

V = índice de valor desses mesmos bens e serviços.

O deflator (D) implícito da renda é dado por:

$$D = \frac{V}{Q} \cdot 100$$

Onde: Q = índice de quantidade de Laspeyres

D = Índice de preços de Paasche

$$\text{já que: valor relativo total} = \frac{\sum (p_n q_n)}{\sum (p_0 q_0)}$$

O “deflator implícito da renda” procura refletir o comportamento dos preços de todos os bens e serviços produzidos em uma economia, entre dois ou mais períodos. Será utilizado como deflator sempre que se procurar eliminar, de uma série, aquelas variações associadas ao nível geral de preços.

3º) IGP – Índice Geral de Preços

É o índice que indica a evolução dos preços durante determinado período. No caso do Brasil, ele é encontrado na “coluna 2” da revista “Conjuntura Econômica”, publicada pela Fundação Getúlio Vargas, constituindo um indicador da taxa de desvalorização monetária. Na construção deste índice, são considerados os seguintes itens, cada qual com o peso respectivo:

- a) preço no atacado = 0,60
- b) custo de vida = 0,30
- c) custo de construção = 0,10

Sua utilidade está, também, no possibilitar o deflacionamento ou o inflacionamento de dados, para ajustá-los à realidade econômica.

4º) Índice “Relação de Troca”

Seu objetivo é o de descrever o comportamento dos preços de exportação relativamente aos preços de importação, entre dois ou mais períodos. Sendo:

- R = índice relação de troca
- P_x = preço das exportações
- P_m = preço das importações

O índice é dado pela fórmula:

$$R = \frac{P_x}{P_m}$$

5º) Índice da “Capacidade de Importar”

Este índice - também denominado de “de poder de compra das exportações” - descreve a variação do volume das importações que um país pode realizar, tendo como base as exportações efetuadas, entre dois ou mais períodos.

Seja:

- Q_x = índice de quantum das exportações
- Q_m = índice de quantum das importações
- C = índice da capacidade de importar

Teremos:

$$C = R \cdot Q_x = \frac{P_x}{P_m} \cdot Q_x = \frac{V_x}{P_m}$$

ANEXO 1 – Tabela Valores “ λ ” (Poisson)

Valores de $e^{-\lambda t}$ quando λt é menor do que 1 e varia de 1 a 10

μ	0	1	2	3	4	5	6	7	8	9
0,0	1,0000	0,9900	0,9802	0,9704	0,9608	0,9512	0,9418	0,9324	0,9231	0,9139
0,1	0,9048	0,8958	0,8869	0,8781	0,8694	0,8607	0,8521	0,8437	0,8353	0,8270
0,2	0,8187	0,8106	0,8025	0,7945	0,7866	0,7788	0,7711	0,7634	0,7558	0,7483
0,3	0,7408	0,7334	0,7261	0,7189	0,7118	0,7047	0,6977	0,6907	0,6839	0,6771
0,4	0,6703	0,6636	0,6570	0,6505	0,6440	0,6376	0,6313	0,6250	0,6188	0,6126
0,5	0,6065	0,6005	0,5945	0,5886	0,5827	0,5770	0,5712	0,5655	0,5599	0,5513
0,6	0,5488	0,5434	0,5379	0,5326	0,5273	0,5220	0,5169	0,5117	0,5066	0,5016
0,7	0,4966	0,4916	0,4868	0,4819	0,4771	0,4724	0,4677	0,4630	0,4584	0,4538
0,8	0,4493	0,4449	0,4404	0,4360	0,4317	0,4274	0,4232	0,4190	0,4148	0,4107
0,9	0,4066	0,4025	0,3985	0,3946	0,3906	0,3867	0,3829	0,3791	0,3753	0,3716

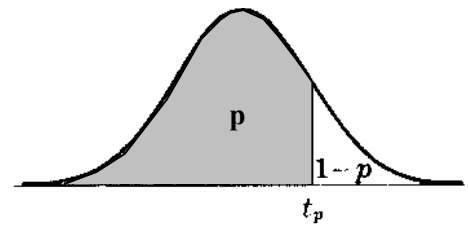
μ	1	2	3	4	5	6	7	8	9	10
$e^{-\mu}$	0,36788	0,13534	0,04979	0,01832	0,006738	0,002479	0,000912	0,000335	0,000123	0,000045

OBS.: Note-se que para ter $e^{-\lambda t}$ para valores de λt não dados nesta tabela, deve-se usar as leis dos expoentes.

$$\text{Ex.: } e^{-2,52} = (e^{-2,00}) \cdot (e^{-0,52}) = (0,13534) \cdot (0,5945) = 0,0805.$$

ANEXO 3 – Tabela Valores ‘t’ de Student

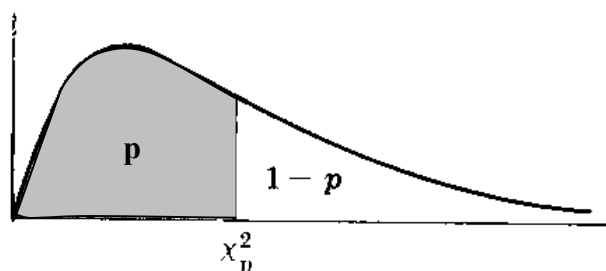
Valores de percentis (t_p)
para a distribuição t de Student
com ν graus de liberdade



ν	$t_{0,55}$	$t_{0,60}$	$t_{0,70}$	$t_{0,75}$	$t_{0,80}$	$t_{0,90}$	$t_{0,95}$	$t_{0,975}$	$t_{0,99}$	$t_{0,995}$
1	,158	,325	,727	1,000	1,376	3,08	6,31	12,71	31,82	63,66
2	,142	,289	,617	,816	1,061	1,89	2,92	4,30	6,96	9,92
3	,137	,277	,584	,765	,978	1,64	2,35	3,18	4,54	5,84
4	,134	,271	,569	,741	,941	1,53	2,13	2,78	3,75	4,60
5	,132	,267	,559	,727	,920	1,48	2,02	2,57	3,36	4,03
6	,131	,265	,553	,718	,906	1,44	1,94	2,45	3,14	3,71
7	,130	,263	,549	,711	,896	1,42	1,90	2,36	3,00	3,50
8	,130	,262	,546	,706	,889	1,40	1,86	2,31	2,90	3,36
9	,129	,261	,543	,703	,883	1,38	1,83	2,26	2,82	3,25
10	,129	,260	,542	,700	,879	1,37	1,81	2,23	2,76	3,17
11	,129	,260	,540	,697	,876	1,36	1,80	2,20	2,72	3,11
12	,128	,259	,539	,695	,873	1,36	1,78	2,18	2,68	3,06
13	,128	,259	,538	,694	,870	1,35	1,77	2,16	2,65	3,01
14	,128	,258	,537	,692	,868	1,34	1,76	2,14	2,62	2,98
15	,128	,258	,536	,691	,866	1,34	1,75	2,13	2,60	2,95
16	,128	,258	,535	,690	,865	1,34	1,75	2,12	2,58	2,92
17	,128	,257	,534	,689	,863	1,33	1,74	2,11	2,57	2,90
18	,127	,257	,534	,688	,862	1,33	1,73	2,10	2,55	2,88
19	,127	,257	,533	,688	,861	1,33	1,73	2,09	2,54	2,86
20	,127	,257	,533	,687	,860	1,32	1,72	2,09	2,53	2,84
21	,127	,257	,532	,686	,859	1,32	1,72	2,08	2,52	2,83
22	,127	,256	,532	,686	,858	1,32	1,72	2,07	2,51	2,82
23	,127	,256	,532	,685	,858	1,32	1,71	2,07	2,50	2,81
24	,127	,256	,531	,685	,857	1,32	1,71	2,06	2,49	2,80
25	,127	,256	,531	,684	,856	1,32	1,71	2,06	2,48	2,79
26	,127	,256	,531	,684	,856	1,32	1,71	2,06	2,48	2,78
27	,127	,256	,531	,684	,855	1,31	1,70	2,05	2,47	2,77
28	,127	,256	,530	,683	,855	1,31	1,70	2,05	2,47	2,76
29	,127	,256	,530	,683	,854	1,31	1,70	2,04	2,46	2,76
30	,127	,256	,530	,683	,854	1,31	1,70	2,04	2,46	2,75
40	,126	,255	,529	,681	,851	1,30	1,68	2,02	2,42	2,70
60	,126	,254	,527	,679	,848	1,30	1,67	2,00	2,39	2,66
120	,126	,254	,526	,677	,845	1,29	1,66	1,98	2,36	2,62
∞	,126	,253	,524	,674	,842	1,28	1,645	1,96	2,33	2,58

ANEXO 4 – Tabela Valores Qui-Quadrado

**Valores dos Percentis (χ_p^2)
para a Distribuição Qui-Quadrado
com ν graus de liberdade**



ν	$\chi^2_{.005}$	$\chi^2_{.01}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.10}$	$\chi^2_{.25}$	$\chi^2_{.50}$	$\chi^2_{.75}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$	$\chi^2_{.999}$
1	,0000	,0002	,0010	,0039	,0158	,102	,455	1,32	2,71	3,84	5,02	6,63	7,88	10,8
2	,0100	,0201	,0506	,103	,211	,575	1,39	2,77	4,61	5,99	7,38	9,21	10,6	13,8
3	,0717	,115	,216	,352	,584	1,21	2,37	4,11	6,25	7,81	9,35	11,3	12,8	16,3
4	,207	,297	,484	,711	1,06	1,92	3,36	5,39	7,78	9,49	11,1	13,3	14,9	18,5
5	,412	,554	,831	1,15	1,61	2,67	4,35	6,63	9,24	11,1	12,8	15,1	16,7	20,5
6	,676	,872	1,24	1,64	2,20	3,45	5,35	7,84	10,6	12,6	14,4	16,8	18,5	22,5
7	,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,0	14,1	16,0	18,5	20,3	24,3
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,2	13,4	15,5	17,5	20,1	22,0	26,1
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,4	14,7	16,9	19,0	21,7	23,6	27,9
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,5	16,0	18,3	20,5	23,2	25,2	29,6
11	2,60	3,05	3,82	4,57	5,58	7,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8	31,3
12	3,07	3,57	4,40	5,23	6,30	8,44	11,3	14,8	18,5	21,0	23,3	26,2	28,3	32,9
13	3,57	4,11	5,01	5,89	7,04	9,30	12,3	16,0	19,8	22,4	24,7	27,7	29,8	34,5
14	4,07	4,66	5,63	6,57	7,79	10,2	13,3	17,1	21,1	23,7	26,1	29,1	31,3	36,1
15	4,60	5,23	6,26	7,26	8,55	11,0	14,3	18,2	22,3	25,0	27,5	30,6	32,8	37,7
16	5,14	5,81	6,91	7,96	9,31	11,9	15,3	19,4	23,5	26,3	28,8	32,0	34,3	39,3
17	5,70	6,41	7,56	8,67	10,1	12,8	16,3	20,5	24,8	27,6	30,2	33,4	35,7	40,8
18	6,26	7,01	8,23	9,39	10,9	13,7	17,3	21,6	26,0	28,9	31,5	34,8	37,2	42,3
19	6,84	7,63	8,91	10,1	11,7	14,6	18,3	22,7	27,2	30,1	32,9	36,2	38,6	43,8
20	7,43	8,26	9,59	10,9	12,4	15,5	19,3	23,8	28,4	31,4	34,2	37,6	40,0	45,3
21	8,03	8,90	10,3	11,6	13,2	16,3	20,3	24,9	29,6	32,7	35,5	38,9	41,4	46,8
22	8,64	9,54	11,0	12,3	14,0	17,2	21,3	26,0	30,8	33,9	36,8	40,3	42,8	48,3
23	9,26	10,2	11,7	13,1	14,8	18,1	22,3	27,1	32,0	35,2	38,1	41,6	44,2	49,7
24	9,89	10,9	12,4	13,8	15,7	19,0	23,3	28,2	33,2	36,4	39,4	43,0	45,6	51,2
25	10,5	11,5	13,1	14,6	16,5	19,9	24,3	29,3	34,4	37,7	40,6	44,3	46,9	52,6
26	11,2	12,2	13,8	15,4	17,3	20,8	25,3	30,4	35,6	38,9	41,9	45,6	48,3	54,1
27	11,8	12,9	14,6	16,2	18,1	21,7	26,3	31,5	36,7	40,1	43,2	47,0	49,6	55,5
28	12,5	13,6	15,3	16,9	18,9	22,7	27,3	32,6	37,9	41,3	44,5	48,3	51,0	56,9
29	13,1	14,3	16,0	17,7	19,8	23,6	28,3	33,7	39,1	42,6	45,7	49,6	52,3	58,3
30	13,8	15,0	16,8	18,5	20,6	24,5	29,3	34,8	40,3	43,8	47,0	50,9	53,7	59,7
40	20,7	22,2	24,4	26,5	29,1	33,7	39,3	45,6	51,8	55,8	59,3	63,7	66,8	73,4
50	28,0	29,7	32,4	34,8	37,7	42,9	49,3	56,3	63,2	67,5	71,4	76,2	79,5	86,7
60	35,5	37,5	40,5	43,2	46,5	52,3	59,3	67,0	74,4	79,1	83,3	88,4	92,0	99,6
70	43,3	45,4	48,8	51,7	55,3	61,7	69,3	77,6	85,5	90,5	95,0	100	104	112
80	51,2	53,5	57,2	60,4	64,3	71,1	79,3	88,1	96,6	102	107	112	116	125
90	59,2	61,8	65,6	69,1	73,3	80,6	89,3	98,6	108	113	118	124	128	137
100	67,3	70,1	74,2	77,9	82,4	90,1	99,3	109	118	124	130	136	140	149

ANEXO 5 – Gramática Grega

SÍMBOLOS UTILIZADOS NA ESTATÍSTICA

ALFABETO ⁽¹⁾ - O Alfabeto grego consta de 24 (vinte e quatro) letras:

Maiúsculas	Minúsculas	Nome clássico	Designação em grego	Pronúncia	Nome moderno
Α	α	alfa	ἄλφα	<i>a</i>	alfa
Β	β	beta	βῆτα	<i>b</i>	vita
Γ	γ	gama	γάμμα	<i>g</i>	gama
Δ	δ	delta	δέλτα	<i>d</i>	delta
Ε	ε	èpsilon	ἒ ψιλόν	<i>ē</i>	épsilon
Ζ	ζ	dzeta	ζῆτα	<i>dz</i>	zita
Η	η	eta	ἦτα	<i>ē</i>	ita
Θ	θ, θ	theta ⁽²⁾	θῆτα	<i>th</i>	thita
Ι	ι	iota	ἰῶτα	<i>i</i>	iota
Κ	κ	capa	κάππα	<i>k</i>	capa
Λ	λ	lambda	λάμβδα	<i>l</i>	landa
Μ	μ	mü ⁽²⁾	μῦ	<i>m</i>	mi
Ν	ν	nü ⁽²⁾	νῦ	<i>n</i>	ni
Ξ	ξ	xi (csi)	ξῖ	<i>x (cs)</i>	xi (csi)
Ο	ο	òmicrón	ὀ μικρόν	<i>ō</i>	ómicron
Π	π	pi	πῖ	<i>p</i>	pi
Ρ	ρ	ró	ῥῶ	<i>rh</i>	ró
Σ	σ, ς	sigma	σῖγμα	<i>s</i>	sigma
Τ	τ	tau	ταῦ	<i>t</i>	taf
Υ	υ	üpsilon ⁽²⁾	ὑ ψιλόν	<i>ü</i>	ípsilon
Φ	φ	fi	φῖ	<i>ph</i>	fi
Χ	χ	khi ⁽²⁾	χῖ	<i>kh</i>	khi
Ψ	ψ	psi	ψῖ	<i>ps</i>	psi
Ω	ω	omega	ὦ μέγα	<i>ō</i>	omega

⁽¹⁾ A palavra *Alfabeto* deriva das 2 (duas) primeiras letras gregas: *alfa* e *beta*.

GLOSSÁRIO DE FÓRMULAS

ESTATÍSTICA

01 – DISTRIBUIÇÃO DISCRETA DE PROBABILIDADE – BINOMIAL

$$P(x) = C_n^r \cdot P^r \cdot q^{n-r}$$

02 – DISTRIBUIÇÃO MULTINOMIAL

$$P(x) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$$

03 – DISTRIBUIÇÃO DE BERNOULLI

$$P(x) = P^x (1 - P)^{1-x}$$

04 – DISTRIBUIÇÃO GEOMÉTRICA

$$P(x) = Pq^{x-1}$$

05 – DISTRIBUIÇÃO DE PASCAL

$$P(x) = C_{(n-1)}^{(x-1)} \cdot P^x \cdot q^{n-x}$$

06 – DISTRIBUIÇÃO DE POISSON

$$P(x) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}$$

07 – DISTRIBUIÇÃO NORMAL PADRONIZADA

$$Z = \frac{X_i - \bar{X}}{S}$$

08 – DISTRIBUIÇÃO EXPONENCIAL

$$P(T > t) = e^{-\lambda t}$$

$$P(T < t) = 1 - e^{-\lambda t}$$

09 – DISTRIBUIÇÃO 't' DE STUDENT

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

10 – MARGEM DE ERRO ('t' de Student)

$$E = t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$$

(Com base em uma amostra pequena ($n \leq 30$) e σ desconhecido)

11 – INTERVALO DE CONFIANÇA

$$\bar{X} - E < \mu < \bar{X} + E$$

(Com base em uma amostra pequena ($n \leq 30$) e σ desconhecido)

12 – DISTRIBUIÇÃO QUI-QUADRADO

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

13 – INTERVALO DE CONFIANÇA (ou Estimativa Intervalar) para a Variância Populacional σ^2

$$\frac{(n-1)S^2}{\chi_D^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_E^2}$$

14 – COEFICIENTE DE CORRELAÇÃO LINEAR DE KARL PEARSON

$$r = \frac{n \cdot \Sigma xy - (\Sigma x) \cdot (\Sigma y)}{\sqrt{[n \cdot \Sigma x^2 - (\Sigma x)^2] \cdot [n \cdot \Sigma y^2 - (\Sigma y)^2]}}$$

15 – REGRESSÃO LINEAR SIMPLES – EQUAÇÃO DA RETA (Y em função de X)

$$\hat{Y} = a + bx$$

16 – FÓRMULAS PARA CALCULAR “a” e “b” (Método dos “mínimos quadrados”)

$$b = \frac{\Sigma xy - n \cdot \bar{x} \cdot \bar{y}}{\Sigma x^2 - n \cdot \bar{x}^2} \quad \text{e} \quad a = \bar{y} - b\bar{x}$$

17 – REGRESSÃO LINEAR MÚLTIPLA

$$\hat{Y} = a + b_1x_1 + b_2x_2$$

18 – EQUAÇÕES NORMAIS PARA A FORMAÇÃO DO ESTIMADOR (n + 1)

$$\begin{aligned} \Sigma y &= na + b_1 \Sigma x_1 + b_2 \Sigma x_2 \\ \Sigma yx_1 &= a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 \\ \Sigma yx_2 &= a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 \end{aligned}$$

19 – RELATIVO DE PREÇOS

$$I_p = \frac{P_n}{P_o} \cdot 100$$

20 – RELATIVO DE QUANTIDADES

$$I_q = \frac{q_n}{q_o} \cdot 100$$

21 – VALOR RELATIVO

$$Vr = \frac{v_n}{v_o} = \frac{p_n q_n}{p_o q_o}$$

22 – AGREGATIVO SIMPLES OU ÍNDICE DE BRADSTREET

$$I_p = \frac{\sum p_n}{\sum p_o} \cdot 100 \quad \text{e} \quad I_q = \frac{\sum q_n}{\sum q_o} \cdot 100$$

23 – MÉDIO-ARITMÉTICO

$$I_p = \frac{\sum \left(\frac{p_n}{p_o} \right)}{N} \quad \text{e} \quad I_q = \frac{\sum \left(\frac{q_n}{q_o} \right)}{N}$$

24 – MÉDIO-HARMÔNICO

$$H = \frac{N}{\sum \left(\frac{1}{x_i} \right)}$$

25 – MÉDIO-HARMÔNICO PARA PREÇOS RELATIVOS

$$I_p = \frac{N}{\sum \left(\frac{p_o}{p_n} \right)} \quad \text{e} \quad I_q = \frac{N}{\sum \left(\frac{q_o}{q_n} \right)}$$

26 – MÉDIO-GEOMÉTRICO

$$I_p = \sqrt[N]{\pi \cdot \left(\frac{p_n}{p_o} \right)} \quad \text{e} \quad \log I_p = \frac{\sum \log \left(\frac{p_n}{p_o} \right)}{N}$$

27 – MÉTODO OU CRITÉRIO DE LASPEYRES – Para Preços

$$I_p = \frac{\sum (p_n q_o)}{\sum (p_o q_o)}$$

28 – MÉTODO OU CRITÉRIO DE LASPEYRES – Para Quantidades

$$I_q = \frac{\sum(q_n p_o)}{\sum(q_o p_o)}$$

29 – MÉTODO OU CRITÉRIO DE PAASCHE – Para Preços

$$I_p = \frac{\sum(p_n q_n)}{\sum(p_o q_n)}$$

30 - MÉTODO OU CRITÉRIO DE PAASCHE – Para Quantidades

$$I_q = \frac{\sum(q_n p_n)}{\sum(q_o p_n)}$$

31 – ÍNDICE IDEAL DE FISHER

$$I_p = \sqrt{\frac{\sum(p_n q_o) \cdot \sum(p_n q_n)}{\sum(p_o q_o) \cdot \sum(p_o q_n)}}$$

32 – MÉDIA ARITMÉTICA PONDERADA DE RELATIVOS

$$I = \frac{\sum\left(\frac{p_n}{p_o} \cdot p_o q_o\right)}{\sum(p_o q_o)}$$

33 – MÉDIA GEOMÉTRICA PONDERADA DE RELATIVOS

$$I = \sqrt[\sum Q_i]{\pi \left(\frac{p_n}{p_o}\right)^{Q_i}} \quad \text{ou} \quad \log I = \frac{\sum Q_i \log \frac{p_n}{p_o}}{\sum Q_i}$$

34 – MÉTODO OU CRITÉRIO DE SIDGWICK-DROBISH

$$I = \frac{\sum(p_n q_o) + \sum(p_n q_n)}{\sum(p_o q_o) + \sum(p_o q_n)} \cdot \frac{1}{2}$$

35 – MÉTODO OU CRITÉRIO DE MARSHALL-EDQEWORTH

$$I = \frac{\sum p_n (q_o + q_n)}{\sum p_o (q_o + q_n)}$$

36 – MÉTODO OU CRITÉRIO DAS MÉDIAS DAS QUANTIDADES DOS ANOS TÍPICOS

$$I_p = \frac{\sum(p_n q_t)}{\sum(p_o q_t)}$$

37 – TESTES E PROVAS DE AVALIAÇÃO DE ADEQUAÇÃO DA FÓRMULA DE UM ÍNDICE

2º) Reversão (Inversão) do Tempo

$$I_{pn/o} = \frac{1}{I_{po/n}} \quad \therefore \quad I_{pn/o} \cdot I_{po/n} = 1$$

3º) Reversão de Fatores

$$I_p \cdot I_q = \frac{\sum(p_n q_n)}{\sum(p_o q_o)} = \frac{\sum v_n}{\sum v_o}$$

4º) Cíclica ou Circular

$$I_{pb/a} \cdot I_{pc/b} \cdot I_{pa/c} = 1$$

5º) Cíclica ou Circular Modificada

$$I_{pb/a} \cdot I_{pc/b} \cdot I_{pd/c} = I_{pd/a}$$

38 – ALGUNS ÍNDICES ESPECIAIS

1º) ICV – Índice do Custo de Vida (Índice de Preços ao Consumidor)

$$\frac{\Delta \text{ICV}}{\text{ICV}} = \Delta \left[\frac{p_n^i}{p_o^i} \right] \cdot w_j^i \cdot k_j$$

2º) Deflator implícito da Renda (Produto)

$$D = \frac{V}{Q} \cdot 100$$

4º) Índice “Relação de Troca”

$$R = \frac{P_x}{P_m}$$

5º) Índice da “Capacidade de Importar”

$$C = R \cdot Q_x = \frac{P_x}{P_m} \cdot Q_x = \frac{V_x}{P_m}$$

BIBLIOGRAFIA

BUSSAB, Wilton O.; MORETTIN, Pedro A.. **Estatística básica métodos quantitativos**. 4. ed. São Paulo: Saraiva, 1999.

CRESPO, Antônio Arnot. **Estatística fácil**. 17. ed. São Paulo: Saraiva, 1999.

DOWNING, Douglas; CLARK, Jeffrey. **Estatística aplicada**. São Paulo: Saraiva, 1998.

FONSECA, Jairo Simon da; MARTINS, Gilberto de Andrade. **Curso de estatística**. 6. ed. São Paulo: Atlas, 1996.

FONSECA, Jairo Simon da; MARTINS, Gilberto de Andrade; TOLEDO, Geraldo de Andrade. **Estatística aplicada**. 2. ed. São Paulo: Atlas, 1985.

HOFFMANN, Rodolfo. **Estatística para economistas**. 3. ed. São Paulo: Pioneira, 1998.

KASMIER, Leonardo J.. **Estatística aplicada a economia e administração**. São Paulo: McGraw-Hill, 1982.

LOPES, Paulo Afonso. **Probabilidades e estatística**. Rio de Janeiro: Reichmann & Affonso Editores, 1999.

MARTINS, Gilberto de Andrade; DONAIRE, Denis. **Princípios de estatística**. 4. ed. São Paulo: Atlas, 1979.

MILONE, Giuseppe; ANGELINI, Flávio. **Estatística geral**. São Paulo: Atlas, 1993.

SILVA, Elio Medeiros da et al. **Estatística (Vol. 1, 2 e 3)**. São Paulo: Atlas, 1995.

SPIEGEL, Murray Ralph. **Estatística**. 3. ed. São Paulo: Makron Books, 1993. (Coleção Schaum)

SPIEGEL, Murray Ralph. **Probabilidade e estatística**. São Paulo: McGraw-Hill, 1978. (Coleção Schaum)

SPINELLI, Walter; SOUZA, Maria Helena S. de. **Introdução à estatística**. 3. ed. São Paulo: Ática, 1997.

TOLEDO, Geraldo Luciano; OVALLE, Ivo Izidoro. **Estatística básica**. 2. ed. São Paulo: Atlas, 1995.

TRIOLA, Mário F. **Introdução à estatística**. Rio de Janeiro: LTC Livros Técnicos e Científicos Editora, 1999.

VIEIRA, Sonia. **Princípios de estatística**. São Paulo: Pioneira, 1999.