

## Disciplina de Estatística Econômica

### Índice

#### Capítulo 1

1.1 Desemprego: “ a bomba que ameaça o mundo ”	3
1.2 A situação no brasil (1998)	4
1.3 A trajetória do desemprego estrutural no brasil	5
1.4 Brasil - 3º país do mundo em desemprego	7

#### Capítulo 2

Números índices	9
2.1 Introdução	9
2.2 Conceito de relativo	10
2.2.1 Relativo de preço	10
2.2.2 Relativo de quantidade	11
2.2.3 Relativo de valor	12
2.3 Índices agregativos simples e ponderados	13
2.3.1 Índice laspeyres ou método da época-base	14
2.3.2 Índice de paasche ou método da época-atual	14
2.3.3 Relativos em cadeia	15
2.3.4 Índice de valor calculados com base nos relativos em cadeia	16
2.3.5 Mudança do período-base	17
2.3.6 União de duas séries de números-índices	18
2.3.7 Conceito de deflator e de poder aquisitivo	19
2.4 Índice de preços ao consumidor (ipc)	21
2.5 Outros índices publicados	22

#### Capítulo 3\*

Regressão linear nos parâmetros (tratamento matricial)	23
3.1 Introdução	23
3.2 Sistema de equações normais e estimativa dos parâmetros	25
3.4 Análise de variância	27
3.5 Esperanças matemáticas das somas de quadrados	29
3.6 Matriz de dispersão	31

#### Capítulo 4 \*

O modelo $Y_i = A + BX_i + e_i$	
4.1 Estimativa de parâmetros e a equação estimada	33
4.2 Exercício	34
4.3. Variância de Y	37
4.4. Intervalo de confiança para Y	37
4.5. Intervalo de confiança para a	38
4.6. Intervalo de confiança para b	38
O modelo $Y_i = A + BX_i + CX_i^2 + e_i$	
4.7 Sistema de equações normais	39
4.7.1 Análise da variância	40

**Capítulo 5 \***

O modelo  $Y_i = A + B_1X_{1i} + B_2X_{2i} + \dots + B_pX_{pi} + e_i$

5.1 Sistema de equações normais	42
5.2 Análise da variância	43

**Capítulo 6 \***

Outros modelos	62
6.1. O modelo exponencial $Y = A B^X$	62
6.2 Função geométrica $Y = A X^B$	65
6.3 Curva de gompertz: $Y = A B^{C^X}$	67
6.4 Curva logística $Y^* = 1 / (A + B C^X)$	70

**Capítulo 7**

Séries temporais	72
7.1. Definição	72
7.2. Objetivo	72
7.3. Previsão de séries temporais	77
7.4. Métodos de previsão de séries temporais	79
7.4.1. Métodos de decomposição de séries temporais	82
7.4.2. Métodos simples de previsão de séries temporais	83
7.4.2.1. Média móvel	84
7.4.2.2. Alisamento exponencial simples	84
7.4.2.3. Alisamento exponencial linear	85
7.4.2.4. Alisamento exponencial sazonal e linear de winter	86
7.4.3 Modelos Arima	86
7.4.3.1. Modelos Autoregressivo e de médias móveis	86
7.4.3.2. Modelo de Box e Jenkins	87
7.5. Considerações sobre os métodos de previsão de séries temporais	93
7.6. Medidas de acuidade dos métodos de previsão de séries temporais	95

<b>Bibliografia</b>	97
---------------------	----

**Observação**

\* O conteúdo dos capítulos, 3 até 6, foi gentilmente doado pelo professor Dr Manuel Luiz Figueirôa

**Disciplina**

# **Estatística Econômica**

**Prof. Msc. [Daniel Francisco Neyra](#)**

## Capítulo 1

### 1.1 DESEMPREGO: “ A BOMBA QUE AMEAÇA O MUNDO ”

- A situação do emprego, no mundo, é alarmante. Em todo o planeta, o já combalido mercado de trabalho passa por profundas transformações - muitas delas traumáticas - , causadas pela globalização e pelo fantástico progresso tecnológico dos últimos 20 anos.
- O atual processo de acumulação capitalista prega o uso intensivo de informação, a horizontalização das grandes unidades produtivas e o sistema de produção flexível, provocando mudanças relevantes nas relações com fornecedores. Exige, ainda, transformações rápidas no mix, demandando reduzido mercado de trabalho. dos trabalhadores que ficam, exige-se grande variedade de habilidades: flexibilidade, soluções criativas, alto grau de engajamento na empresa e capacidade para tomar conhecimento de todo o processo produtivo.
- Segundo a OIT-Organização Internacional do Trabalho, existem hoje(1997), no mundo, perto de 800 milhões de desempregados, o nível mais alto desde a Grande Depressão, nos anos 30. Além do desemprego, também causa preocupação o processo de crescente precarização dos postos de trabalho. Somando-se os contingentes de desempregados e de subempregados, em todo o mundo, chega-se perto de 1 bilhão de pessoas. Ou seja: aproximadamente, 30 % de toda a força mundial de trabalho, adverte a OIT.
- O conceito clássico de pleno emprego, defendido por Willian Beveridge em 1944, admite como tolerável uma taxa de desemprego conjuntural aceitável em torno de 3% - desde que essa margem abranja pessoas desempregadas por breve período de tempo, indivíduos que possam manter-se graças a um seguro desemprego.
- A OIT revelou que a taxa média de desemprego, na União Européia (UE), foi de 11,3 % em julho/96, com ligeiro aumento em relação ao ano anterior. De 1995 a julho/96, o índice manteve-se acima de 22% na Espanha, de 14,5 % na Bélgica e em mais de 12 % na França e na Itália.

- Na Austrália a taxa subiu para 8,5 % em 1996 e no Japão 3,4 %. O desemprego não variou no Canadá, que registrou taxa de 9,8 % em julho/96, e continuou baixando nos EUA, de 5,7 % para 5,4 %, no mesmo período.
- Na América Latina, estima-se que, entre julho/95 e julho/96, a taxa de desemprego passou de 8 % para 10 % na Colômbia.

## 1.2 A SITUAÇÃO NO BRASIL (1998)

No Brasil dos anos 90, foram abertos, anualmente, 951.400 postos de trabalho em média. Em contrapartida, 1.417.100 pessoas em média ingressaram no mercado de trabalho a cada ano durante este período. Ou seja, o desemprego atingiu, em média, 465.700 pessoas ao ano.

No Brasil, de acordo com o IBGE, existem aproximadamente 3 milhões de desempregados, entretanto 3 milhões é o número de desempregados que o DIEESE considera que exista apenas nas setes regiões metropolitanas onde o órgão faz suas pesquisas. Só em São Paulo, hoje existem em torno de 1.500.000 trabalhadores desempregados. Segundo o Economista Márcio Pochmann (UNICAMP), as taxas de investimentos na década de 90, em relação ao PIB, são inferiores às registradas nos anos 80. A qualidade do emprego no país está pior, assim como as relações de trabalho. Nos anos 90, ocorreram uma ampliação das ocupações não-assalariadas e das assalariadas sem carteira de trabalho. Ao mesmo tempo, os empregos com registro em carteira diminuíram.

Para Pochmann, “ o trabalhador não-assalariado adota uma estratégia de sobrevivência e isso não pode ser desejado como futuro, pois a pessoa não tem acesso à previdência e aos direitos sociais”.

Para que o desemprego acabasse, a economia teria que crescer à taxa de 6% ao ano - mais que o triplo do que prevê o Ipea. Só assim seria possível absorver o contingente de de 1 milhão a 1,5 milhão de jovens que a cada ano ingressam no mercado de trabalho, segundo o professor Carlos Ivan Simonsen Leal.

"Nos próximos dez anos, o Brasil ainda vai ter uma taxa muito alta de pessoas entrando no mercado de trabalho. Isso significa que o país terá que gerar muitos empregos novos. Quando o PIB cresce a 3% ao ano, não gera o número de empregos novos suficientes", disse Simonsen Leal. O professor refere-se à criação líquida de emprego, isto é, empregos novos menos os empregos que deixaram de existir.

José Márcio Camargo, do Departamento de Economia da Pontifícia Universidade Católica (PUC), concorda que o corte de gastos públicos reduz o crescimento econômico, a curto prazo. Mas ressalta que o ajuste fiscal, aliado à execução das reformas da legislação trabalhista e da Previdência Social, permitirá a retomada do crescimento econômico a partir de 2000.

Fonte: Revista Rumos do Desenvolvimento, Jornal do Brasil, Jornal Sindibancários

### **1.3 A Trajetória do Desemprego Estrutural no Brasil**

O Desemprego é um problema mundial, mas suas razões são agravadas por problemas nacionais, diz Márcio Pochmann. O desemprego estrutural vem piorar um quadro de distorções sociais que remontam ao final do século passado no País, época da Segunda Revolução Industrial. Entre 1890 e 1980, o Brasil foi o país que mais cresceu no mundo, sem resolver os problemas tradicionais de mercado de trabalho, afirma Pochmann.

Entre os fatores que agravam a questão estão a baixa escolaridade da mão-de-obra, a informalidade do trabalho e a precária rede de proteção social que cerca a população. Além disso, o País passou a adotar políticas públicas de proteção ao trabalhador com enorme atraso. O seguro-desemprego, por exemplo, só foi implementado em meados da década de 80. O Brasil enfrenta, simultaneamente, problemas de atraso e da modernidade.

Os países desenvolvidos resolveram o problema do emprego no século XIX com um conjunto de medidas como a reforma agrária, a reforma tributária e a reforma social. Por não termos feito a reforma agrária, em três décadas o Brasil deixou de ser uma economia agrária para se tornar uma economia urbana: cerca de 100 milhões de pessoas deslocaram-se do campo para a cidade. Para fazer o mesmo movimento, a França levou 100 anos, compara Pochmann. O resultado, como se sabe, foi esse mercado de trabalho urbano abundante de mão-de-obra, sem alternativas ocupacionais.

O Brasil também não seguiu o exemplo das nações que hoje são desenvolvidas ao deixar de fazer a reforma tributária. Convivemos com uma estrutura tributária regressiva, onde os mais pobres pagam muitos impostos. Outra diferença marcante do Brasil em relação ao Primeiro Mundo, a má distribuição de renda, exclui grande parte da população do mercado de consumo. Nos Estados Unidos, por exemplo, 3 em cada 4 pessoas possuem um automóvel. No Brasil, a proporção é de 1 em cada 11 pessoas. Essa exclusão constrange a

expansão do emprego. A exclusão do mercado de consumo foi um dos fatores que não nos permitiram resolver os problemas estruturais da Segunda Revolução Industrial.

As famílias pobres, por exemplo, não utilizam os serviços pessoais, como cabeleireiros, tinturarias, e tampouco freqüentam restaurantes, três setores de atividade terciárias fortemente empregadores de mão-de-obra. O setor de serviços pessoais no Brasil não representa mais do que 5% na estrutura de emprego, ao passo que nos Estados Unidos esse percentual atinge 24%. Também não fizemos a reforma social para implantar no País o Estado do Bem-Estar Social, com garantia de Saúde, Educação, Transporte e Habitação, o que nos permitiria, além de generalizar a cidadania, resolver o problema do emprego.

Não é compromisso do capital gerar empregos. Não podemos exigir isso do setor privado, diz Pochmann. Cabe ao Estado implementar as reformas e garantir a empregabilidade, ainda que isso implique aumento dos gastos públicos. Estamos falando de um país em construção como o Brasil. Nos países mais desenvolvidos, não há mais muito o que fazer. Aqui, ao contrário, há muito o que fazer.

Fonte: Revista Inovação empresarial set/98 - Accor Brasil

#### INDÚSTRIA BRASILEIRA

<b>A N O</b>	<b>Nº DE INDÚSTRIAS</b>	<b>PESSOAL OCUPADO</b>
1920	13.569	293.673
1940	49.418	781.185
1950	92.350	1.279.184
1960	110.771	1.799.376
1970	164.793	2.699.969
1980	226.306	5.720.006
1985	207.157	5.608.704
1990	191.315	5.464.436
1995	208.806	4.906.524
1996	211.778	4.751.422

FONTE: 1920-1985 -IBGE (CENSO INDUSTRIAL); 1990-1996 - MINISTÉRIO DO TRABALHO

## 1.4 BRASIL - 3º PAÍS DO MUNDO EM DESEMPREGO

- O Brasil é o terceiro país em desemprego no mundo, em números absolutos, segundo pesquisa realizada pelo economista Márcio Pochmann, da Unicamp, com base em dados oficiais de 141 países. Em 1999, o volume de desemprego aberto em todo mundo foi de 138 milhões de pessoas. O Brasil, de acordo com os dados IBGE, com 7,7 milhões de pessoas sem trabalho, concentrou 5,61% desse total. O país só fica atrás da Rússia, com 9,1 milhões de pessoas sem emprego, e da Índia, com quase 40 milhões.
- De acordo com o estudo feito por Pochmann, no início dos anos 90, o país ocupava o oitavo lugar no ranking mundial do desemprego, em 95, subiu para quinto e, em 98 atingiu a terceira posição. Nos últimos 24 anos, o desemprego mundial aumentou de 2,3% da PEA (População Economicamente Ativa) para 5,5%. Nos países desenvolvidos, as taxas cresceram, em média, 53%. Em outros, o aumento chegou a 200%. No Brasil, nesse período, o índice cresceu 369,4% passando de 1,73% da PEA, em 1975, e para 9,85% em 1999.
- O economista avalia que a globalização está aumentando a concentração do desemprego em países pobres, incluindo o Brasil. Considera também que a pesquisa questiona a tese de que os avanços tecnológicos seriam responsáveis pelo desemprego, já que os países mais afetados pelo problema são justamente os mais atrasados (*Correio Bancário*, 01/02/2000).

### Cotação do Dólar Comercial venda

#### Em Reais ( último dia útil do mês )

Ano	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
1998	1,1237	1,1304	1,1374	1,1443	1,1505	1,1569	1,1634	1,1769	1,1856	1,1932	1,2019	1,2087
1999	1,9832	2,0648	1,7220	1,6607	1,7240	1,7695	1,7892	1,9159	1,9223	1,9530	1,9227	1,7890
2000	1,8024	1,7685	1,7473	1,8067	1,8266	1,8000	1,7748	1,8234	1,8437	1,9090	1,9596	1,9554
2001	1,9711	2,0452	2,1616	2,1847	2,3600	2,3049						

Fonte: Gazeta Mercantil e Momento BICBANCO



**Classes Sociais do Brasil**  
**( Mobilidade Social 1973/1996 )**

<b>Classes</b>	<b>Composição</b>	<b>(1973)</b>	<b>( 1996)</b>	<b>Evolução</b>
<b>Elite</b>	Profissionais pós-graduados, empresários e altos administradores.	3,5%	4,9%	cresceram 40%
<b>Classe média alta</b>	Pequenos proprietários, técnicos com especialização e gerentes de grande empresa.	6,3%	7,4%	cresceram 17%
<b>Classe média média</b>	Pequenos fazendeiros, auxiliares de escritório e profissionais com pouca especialização.	18,4%	13,3%	encolheram 13%
<b>Classe média baixa</b>	Motoristas, pedreiros, pintores, auxiliares de serviços gerais, mecânicos, etc.	23,7%	26,9%	cresceram 13%
<b>Pobres</b>	Vigias, serventes de pedreiros, ambulantes e outros trabalhadores sem qualificação.	16,1%	23,4%	cresceram 46%
<b>Muito pobres</b>	Trabalhadores rurais, bóias-frias, pescadores, peões de fazendas, catadores urbanos, etc.	32%	24%	encolheram 25%

Fonte: Revista Veja, 13 mai. 1999.

OBS: Existem no Brasil (1996) 1.894.000 de domicílios pertencentes à famílias da chamada classe A (menos de 5% da população total) que ganham acima de 20 salários mínimos. No sudeste concentram-se 970 mil desses domicílios, no sul 278 mil, no centro-oeste 333 mil e no norte/nordeste 313 mil (IBGE :1999).

## Capítulo 2

### Números Índices

#### 2.1 INTRODUÇÃO

Freqüentemente, profissionais de várias áreas de atuação, tais como: economistas, administradores, engenheiros, sociólogos e psicólogos envolvem-se em situações de análise, onde o interesse predominante concentra-se em medir possíveis diferenças entre grupos de dados. Nesse sentido, o número-índice constitui um instrumento de análise poderoso, mormente quando se procura estabelecer comparações entre grupos de variáveis distintas, mas relacionadas entre si.

Em termos gerais, um número-índice pode ser concebido como uma medida estatística destinada a comparar, através de uma expressão quantitativa global, grupos de variáveis relacionadas e com diferentes graus de importância. Através dele obtém-se um quadro resumido das mudanças ocorridas em áreas afins como preços dos insumos básicos adquiridos pelo produtor, preços dos produtos acabados, volume físico de produção etc.

As comparações decorrentes do emprego de números-índices podem ser consideradas sob três aspectos ou categorias:

- a) variações ocorridas ao longo do tempo;
- b) diferenças entre lugares;
- c) diferenças entre categorias semelhantes, como pessoas, produtos ou coisas.

Em qualquer dos casos, a obtenção de um número-índice resulta da combinação de variáveis mediante um total ou uma medida de tendência central, especialmente a média.

Um número-índice é um relativo percentual pelo qual uma medida em um dado período é expressa por meio de uma razão com a medida em um período-base fixado. As medidas podem referir-se a quantidade, preço ou valor.

EXEMPLO 1. O índice Nacional de Preços ao Consumidor (INPC) elaborado pela Fundação IBGE do Brasil é exemplo de um índice de preços. O índice de Quantum de Exportações elaborado pela Fundação Getúlio Vargas é exemplo de um índice de quantidades.

EXEMPLO 2. O índice de preços para a manteiga de, digamos, 130, é um índice de preços simples, ou relativo, e indica que o preço da manteiga no período dado foi 30% maior do que no período-base. O mesmo valor de 130 para o Índice de Preços ao Consumidor (um índice de preços agregados) indica que o preço médio de uma "cesta de mercado" com um certo número de bens e serviços foi 30% maior no período dado, comparativamente ao período-base.

## 2.2 CONCEITO DE RELATIVO

O montante de dinheiro despendido por determinada unidade de compra em um período, comparativamente a outro período tomado como referência, pode variar em virtude de variações no número de unidades compradas dos diferentes artigos e mudanças nos preços unitários dos mesmos. São três, portanto, as variáveis consideradas: preço, quantidade e valor, sendo este último o resultado do produto do preço pela quantidade.

### 2.2.1 Relativo de Preço

Consideremos duas épocas e os respectivos preços de um artigo:

0 = época básica, base ou época de referência

t = época atual, época dada ou época a ser comparada

$p_0$  = preço do artigo na época básica

$p_t$  = preço do artigo na época dada (atual)

O relativo de preço será definido pela seguinte expressão:

$$p_{0,t} = \frac{p_t}{p_0} \quad (1)$$

ou, em termos percentuais,

$$p_{0,t} = \frac{p_t}{p_0} \times 100$$

### Exercício 1

Um artigo foi adquirido por Cr\$ 1000,00 em 1979 e por Cr\$ 2.500,00 no ano seguinte. Calcular o relativo de preço em 80, com base em 1979.

De acordo com (1),

$$p_{79} = 2000$$

$$p_{80} = 2500$$

$$p_{79,80} = 2500 / 2000 = 1,25 \text{ ou } 125\% \text{ (multiplicando-se por 100)}$$

Pode-se afirmar, então, que, em 1980, o preço do artigo corresponde a 125% de seu preço em 1979, tendo uma variação de 25%.

### 2.2.2 Relativo de Quantidade

É possível também comparar quantidades, quer sejam elas quantidades produzidas, vendidas ou consumidas.

Considerando

$q_t$  = quantidade de um produto na época dada, e

$q_0$  = quantidade desse mesmo produto na época básica (base),

a quantidade relativa será fornecida pelo seguinte quociente

$$q_{0,t} = \frac{q_t}{q_0} \tag{2}$$

ou, em termos percentuais,

$$q_{0,t} = \frac{q_t}{q_0} \times 100$$

que descreve quanto a quantidade na época dada (atual) representa da época básica.

### Exercício 2

Um vendedor de automóveis vendeu 600 veículos em 1980 contra 400 em 1978. Calcular o relativo de quantidade em 80, com base em 1979.

De acordo com (2),

$$q_{79} = 400$$

$$q_{80} = 600$$

$$q_{79,80} = 600 / 400 = 150 \text{ ou } 150\% \text{ (multiplicando-se por 100)}$$

Com base neste resultado pode-se afirmar que o vendedor apresentou em 1980 um desempenho 50% superior ao de 1979.

### 2.2.3 Relativo de Valor

Sendo  $p$ , o preço unitário de um bem e  $q$ , a quantidade adquirida, o produto  $p \times q$  será denominado valor da transação.

Supondo  $p_t$  e  $q_t$ , respectivamente o preço e a quantidade de um bem na época dada e  $p_0$  e  $q_0$ , o preço e a quantidade do mesmo bem na época básica, o valor total relativo será definido pelo quociente:

$$v_{0,t} = \frac{v_t}{v_0} = \frac{p_t \cdot q_t}{p_0 \cdot q_0} = \frac{p_t}{p_0} \times \frac{q_t}{q_0} \quad (3)$$

ou, em termos percentuais,

$$v_{0,t} = \frac{v_t}{v_0} = \frac{p_t \cdot q_t}{p_0 \cdot q_0} = \frac{p_t}{p_0} \times \frac{q_t}{q_0} \times 100$$

O relativo de valor pode, então, ser decomposto em um relativo de preço e um relativo de quantidade.

**Exercício 3**

Uma empresa vendeu, em 1979, 12.000 unidades de um artigo ao preço unitário de Cr\$ 500,00. Em 1980 vendeu 15.000 unidades do mesmo artigo ao preço unitário Cr\$ 600,00. Com base em 1979, calcular o relativo de valor em 1980.

$$v_{79,80} = \frac{v_{80}}{v_{79}} = \frac{p_{80} \cdot q_{80}}{p_{79} \cdot q_{79}} = \frac{600}{500} \times \frac{15000}{12000} = \frac{9.000.000}{6.000.000} = 1,50 \quad \text{ou } 150\%$$

Em 1980, o faturamento da empresa, com a venda do artigo, foi 50% superior ao de 1979.

**2.3 ÍNDICES AGREGATIVOS SIMPLES E PONDERADOS**

Até o presente, as referências aos índices e os exemplos apresentados circunscreveram-se às comparações entre preços, quantidades ou valores de um único item (relativo). Entretanto, tais comparações elementares, embora úteis para a compreensão de conceitos básicos e para averiguação da validade dos critérios de avaliação de um índice, são incompletas para o desenvolvimento do assunto. A quase totalidade dos índices econômicos envolvem a avaliação simultânea de variações de preços ou de quantidades de mais de um item.

Quando pretendermos avaliar a variação de preços, entre duas épocas, digamos, de dez artigos, através de um número-resumo (índice), o cálculo dos relativos representa apenas o primeiro passo para a solução do problema.

O método mais comum de cálculo de um índice envolvendo vários índices é o da média aritmética simples (não ponderada). Basta calcular o valor dos relativos de todos os itens considerados e, em seguida, aplicar a fórmula da média aritmética.

Os índices simples apresentam uma séria limitação na prática, pois não levam em consideração a importância relativa de cada um dos vários bens ou serviços que os integram. Além da restrição decorrente da não ponderação dos itens, o índice agregativo simples, embora de fácil aplicação, é afetado por unidades particulares de medidas.

Nesse caso, pode-se dizer, por exemplo, que um item de elevado preço unitário tende a exercer maior influência sobre o índice agregativo de preços do que um de baixo preço unitário. Além disso, determinadas unidades de medida díspares impedem que se somem as quantidades ou trazem distorções quando a elas se associam os preços unitários.

### 2.3.1 Índice Laspeyres ou Método da Época-Base

Um dos mais populares índices agregados de preços é o índice de Laspeyres, no qual os preços são ponderados pelas quantidades associadas com o ano-base antes de serem somados. A fórmula é:

$$I(L) = \frac{\sum p_n \cdot q_0}{\sum p_0 \cdot q_0} \times 100$$

### 2.3.2 Índice de Paasche ou Método da Época-Atual

Quando os preços são ponderados pelas quantidades associadas com o ano dado ou ano atual antes de serem somados, temos o Índice de Paasche. A fórmula é:

$$I(P) = \frac{\sum p_n \cdot q_n}{\sum p_0 \cdot q_n} \times 100$$

## Exercício 4

Determinar os índices de preço de Laspeyres e Paasche, para 1976, dos 3(três) produtos constantes da tabela abaixo, usando 1970 como ano-base.

Produto	Unidade de Medida	Preço Médio (CR\$)		Consumo per capita (mensal)	
		1970 (p <sub>0</sub> )	1976 (p <sub>n</sub> )	1970 (q <sub>0</sub> )	1976 (q <sub>n</sub> )
Leite	Litro	0,30	0,38	30	35
Pão	500 g	0,25	0,35	3,8	3,7
Ovos	dúzia	0,60	0,90	1,5	1,0

## Cálculo dos índices

Produto	Unidade de Medida	Laspeyres		Paasche	
		$p_n \cdot q_0$	$p_0 \cdot q_0$	$p_n \cdot q_n$	$p_0 \cdot q_n$
Leite	Litro	11,40	9,00	13,30	10,50
Pão	500 g	1,33	0,85	1,30	0,92
Ovos	dúzia	1,35	0,90	0,90	0,60
T O T A L		14,08	10,85	15,50	12,02

$$I(L) = (14,08 / 10,85) \cdot 100 = 129,8\%$$

$$I(P) = (15,50 / 12,02) \cdot 100 = 129,0\%$$

**2.3.3 Relativos em Cadeia**

Relativos em cadeia são índices para os quais a base é sempre o período precedente. Portanto, para um conjunto de relativos em cadeia para valores de vendas anuais, cada número-índice representa uma comparação percentual com o ano anterior. Tais relativos são úteis para evidenciar comparações ano a ano, mas não são convenientes como base para fazer comparações de longo prazo.

$$L_n = \frac{p_n}{p_{n-1}} \cdot 100$$

**Exercício 5**

As vendas (em milhões de dólares) de uma empresa automobilística entre 1970 e 1975 apresentaram os valores contidos na tabela abaixo. Determinar os relativos em cadeia para estes valores.



Ano	Vendas (milhões US\$)	Relativos em Cadeia
1970	14.980	-
1971	16.433	<b>109,7</b>
1972	20.194	<b>122,9</b>
1973	23.015	<b>114,0</b>
1974	23.621	<b>102,6</b>
1975	24.009	<b>101,6</b>

### 2.3.4 Índice de Valor calculados com base nos Relativos em Cadeia

Quando dispomos somente dos Relativos em Cadeia e necessitamos do Índices de Valor, os mesmos podem ser obtidos da seguinte forma:

$$I_{n-1} = \frac{I_n}{L_n} \cdot 100 \quad \text{e} \quad I_n = \frac{L_n \cdot I_{n-1}}{100}$$

### Exercício 6

Considerando os relativos em cadeia do exercício anterior, calcular o índice de valor para aqueles 6 anos, utilizando 1972 como ano-base.

Ano	Relativos em Cadeia	Índice de Valor
1970	-	<b>74,2</b>
1971	109,7	<b>81,4</b>
1972	122,9	100,0 (base)
1973	114,0	<b>114,0</b>
1974	102,6	<b>117,0</b>
1975	101,6	<b>118,9</b>

$$I_{71} = \frac{I_{72}}{L_{72}} \cdot 100 = \frac{100,0}{122,9} \cdot 100 = 81,4$$

$$I_{70} = \frac{I_{71}}{L_{71}} \cdot 100 = \frac{81,4}{109,7} \cdot 100 = 74,2$$

$$I_{73} = \frac{L_{73} \cdot I_{72}}{100} = \frac{114,0 \times 100,0}{100} = 114,0$$

$$I_{74} = \frac{L_{74} \cdot I_{73}}{100} = \frac{102,6 \times 114,0}{100} = 117,0$$

$$I_{75} = \frac{L_{75} \cdot I_{74}}{100} = \frac{101,6 \times 117,0}{100} = 118,9$$

### 2.3.5 Mudança do Período-Base

Freqüentemente muda-se a base de uma série de um número-índice para um ano mais recente, de tal forma que comparações correntes possam ser mais significativas. Isto é obtido dividindo-se cada índice (original) pelo índice da nova base (fixada), multiplicando o resultado por 100.

$$I_{n(\text{alterado})} = \frac{I_{n(\text{antigo})}}{I_{n(\text{fixado})}} \cdot 100$$

**Exercício 7**

Dada a tabela abaixo com índices de valor base 1972, mudar para base ano=1977

Ano	Índice de Valor (base 1972)	Índice de Valor (base 1977)
1970	74,2	$74,2 / 129,9 = 57,1$
1971	81,4	$81,4 / 129,9 = 62,7$
1972	100,0 (base)	77,0
1973	114,0	87,8
1974	117,0	90,0
1975	118,9	91,5
1976	124,8	96,1
1977	129,9	100,0 (base)
1978	137,7	106,0
1979	148,3	114,2
1980	160,6	123,6
1981	172,7	132,9

**2.3.6 União de Duas Séries de Números-Índices**

Um número-índice pode, freqüentemente, experimentar alterações pela introdução de novos produtos ou pela exclusão de determinados produtos antigos ou, ainda, por mudanças no ano-base. Contudo, para fins de continuidade histórica, é desejável ter disponível uma série unificada do número-índice. Para que se possa unir duas séries distintas para formar uma única série contínua de números-índices, é necessário haver um ano de superposição, das duas séries, de tal forma que ambos os tipos de índices tenham sido calculados para aquele ano. Tal ano de superposição, geralmente também é a nova base, uma vez que é o ano no qual foram adicionados novos produtos e/ou removidos do índice agregado. O número-índice que deve ser alterado no processo de união é o índice da série antiga. Esta mudança é conseguida, dividindo-se o novo número-índice para o ano de superposição (100,0 se for a nova base) pelo índice antigo para aquele ano, multiplicando-se então cada valor da série do índice antigo por este quociente.

**Exercício 8**

Ano	Índice de Preço Antigo (base 1972)	Índice de Preço Previsto (base 1977)	Índice de Preço Unido (base 1977)
1970	74,2		57,1
1971	81,4		62,7
1972	100,0		77,0
1973	114,0		87,8
1974	117,0		90,1
1975	118,9		91,5
1976	124,8		$124,8 \times 0,7698 =$ 96,1
1977	129,9	100,0	100,0
1978		106,0	106,0
1979		114,2	114,2
1980		123,6	123,6
1981		132,9	132,9

A série unida é calculada multiplicando os índices de preço antigos pelo fator resultante da divisão de **100,0** (valor do ano de superposição) pelo valor antigo **129,9**, que é igual a **0,7698**

**2.3.7 Conceito de Deflator e de poder Aquisitivo**

Deflator é qualquer índice de preços utilizado para equiparar, por redução, valores monetários de diversas épocas ao valor monetário de uma determinada época tomada como base. O processo de redução é denominado deflacionamento. O uso de deflatores permite, ao analisar uma série de valores monetários em termos de suas variações, eliminar uma de suas causas, a variação de preços.

O valor deflacionado é calculado dividindo-se o valor corrente pelo índice da data corresponde, multiplica do por 100.

$$ValorDeflacionado = \frac{ValorCorrente}{Índice} \cdot 100$$

### Exercício 9

Considere os dados constantes na tabela abaixo, referentes ao faturamento de uma empresa e índice de preço no período de 1975 a 1981. Calcular o faturamento real em cada ano considerando 1977 como ano base.

Ano	Índice de Preço (base 1977)	Faturamento a Preços Correntes	Faturamento a Preços de 1977
1975	91,5	1.600	$1.600 / 91,5 \times 100 = 1.748,6$
1976	96,1	1.800	1.873,0
1977	100,0	2.400	2.400,0
1978	106,0	2.800	2.641,5
1979	114,2	3.000	2.627,0
1980	123,6	3.200	2.589,0
1981	132,9	3.900	2.934,5

### Poder Aquisitivo

O poder aquisitivo de um determinado volume de unidades monetárias, com relação a uma determinada época, é seu valor deflacionado com referência a essa época.

Para calcular o poder aquisitivo de uma unidade monetária basta calcular o inverso do índice de preço escolhido.

### Exercício 10

Considere os dados constantes na tabela abaixo, referentes a um índice de preço no período de 1975 a 1981. Calcule o poder aquisitivo de CR\$ 1,00, considerando 1977 como base..

Ano	Índice de Preço (base 1977)	Poder Aquisitivo CR\$ 1,00 (base 1977)
1975	<b>91,5</b>	<b><math>1 / 91,5 \times 100 = 1,09</math></b>
1976	<b>96,1</b>	<b>1,04</b>
1977	<b>100,0</b>	<b>1,00</b>
1978	<b>106,0</b>	<b>0,94</b>
1979	<b>114,2</b>	<b>0,88</b>
1980	<b>123,6</b>	<b>0,81</b>
1981	<b>132,9</b>	<b>0,75</b>

### Exercício 11

O salário de um indivíduo foi majorado em 80% num dado período, enquanto a inflação acusou uma elevação de 92% em igual período. Qual a perda percentual de poder aquisitivo do salário desse indivíduo?

Os valores já estão expressos em percentagens. Os índices correspondentes serão:

Inflação:  **$100 + 92 = 192$**

Salário:  **$100 + 80 = 180$**

Índice de salário deflacionado:  **$180 / 192 = 0,9375$**

Esse valor indica que o poder aquisitivo do salário no fim do período considerado é igual a 93,75% do poder aquisitivo no início do período.

A perda percentual é portanto:  **$100 - 93,75 = 6,25\%$**

## 2.4 ÍNDICE DE PREÇOS AO CONSUMIDOR (IPC)

O Índice de Preços ao Consumidor é um dos índices publicados mais amplamente conhecidos por causa de sua utilização como indicador do custo de vida. No Brasil, várias instituições calculam tais índices com abrangência regional. Dentre tais instituições, pode-se citar: no Rio de Janeiro, a Fundação Getúlio Vargas; em São Paulo, a Fundação Instituto de

Pesquisas Econômicas (FIPE) da Universidade de São Paulo; em Porto Alegre, o Centro de Estudos e Pesquisas Econômicas da Universidade Federal do Rio Grande do Sul. A nível nacional, existe o Índice Nacional de Preços ao Consumidor (INPC), elaborado pela Fundação IBGE, o qual tem servido de base para o cálculo dos reajustes salariais.

Enquanto o Índice de Preços ao Consumidor indica os preços relativos comparados ao ano-base, o recíproco do IPC indica o valor da moeda com relação ao ano-base:

## **2.5 OUTROS ÍNDICES PUBLICADOS**

O Índice de Preços por Atacado (IPA) é publicado mensalmente pela Fundação Getúlio Vargas, do Rio de Janeiro. Este índice (com peso 6) juntamente com o Índice de Custo da Construção Civil no Rio de Janeiro (com peso 1) e com o IPC do Rio de Janeiro (com peso 3) resultam no Índice Geral de Preços (IGP) também publicado mensalmente pela Fundação Getúlio Vargas. O IGP, em seu conceito de disponibilidade interna, tem sido tradicionalmente utilizado no Brasil como o indicador do nível de inflação.

Também mensalmente são publicados, pela Fundação IBGE, indicadores da Produção Industrial e do Emprego Industrial.

## Capítulo 3

### Regressão Linear nos Parâmetros (tratamento matricial)

#### 3.1 Introdução

O estudo da regressão se inicia com a questão do modelo adequado para a representação do fenômeno em estudo.

Se o fenômeno for conhecido este problema já tem solução conhecida.

Ao estudarmos a diferença de potencial em função da intensidade da corrente, experimento este bem conhecido, o modelo adequado é o linear. Entretanto, se estamos interessado na trajetória de um projétil de canhão, o modelo deve ser parabólico ou quadrático.

Quando o fenômeno não é conhecido o procedimento para a escolha do modelo adequado passa por um gráfico cartesiano denominado pelos estatísticos, nestas questões, de gráfico de dispersão.

Consideremos alguns modelos estatísticos onde se supõe:

$$e_i \sim N(0; \sigma^2); \quad E(e_i e_j) = 0; \quad E(e_i^2) = \sigma^2$$

a)  $Y = A + BX + e$

b)  $Y = A + BX + CX^2 + e$

c)  $Y = A + B_1X_1 + B_2X_2 + \dots + B_pX_p + e$





Aqui, também, percebe-se que ( II ) pode ser escrito  $Y = X\beta + \varepsilon$

c) Seja o modelo  $Y_i = A + B_1X_{1i} + B_2X_{2i} + \dots + B_pX_{pi} + e_i$  e construíamos:

$$\begin{aligned} Y_1 &= A + B_1X_{11} + B_2X_{21} + \dots + B_pX_{p1} + e_1 \\ Y_2 &= A + B_1X_{12} + B_2X_{22} + \dots + B_pX_{p2}+ e_2 \\ ..... \\ Y_n &= A + B_1X_{1n} + B_2X_{2n} + \dots + B_pX_{pn}+ e_n \end{aligned}$$

( III )

Defina-se:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{p1} \\ 1 & X_{12} & X_{22} & \dots & X_{p2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix}; \quad \beta = \begin{bmatrix} A \\ B_1 \\ \dots \\ B_p \end{bmatrix}; \quad \square = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}$$

Pelas mesmas razões anteriores (III) pode ser estudado através de  $Y = X\beta + \varepsilon$

### 3.2 Sistema de equações normais e estimativa dos parâmetros

Seja:  $\varepsilon = Y - XB$

$$\varepsilon' = (Y - X\beta)' = Y' - \beta'X'$$

Aqui o símbolo ( ` ) significa transposto.

Considerare  $\varepsilon'\varepsilon = (Y' - \beta'X')(Y - X\beta)$

$$Z = \varepsilon' \varepsilon = Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta$$

$$\text{Quem é } Z = \varepsilon' \varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix} = e_1^2 + e_2^2 + \dots + e_n^2$$

A questão que estamos resolvendo é a determinação do vetor  $\beta$  sob as condições de minimização de  $Z$ .

Ocorre que:  $Y'X\beta = \beta'X'Y$  pois uma é transposta da outra e são matrizes de ordem  $1 \times 1$ .

$$\text{Logo: } Z = Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

$$dz = \phi - 2(d\beta')X'Y + (d\beta')X'X\beta + \beta'X'X(d\beta)$$

Ocorre que:  $(d\beta')X'X\beta = \beta'X'X(d\beta)$  pelas mesmas razões acima.

$$\text{Então: } dz = -2(d\beta')X'Y + 2(d\beta')X'X\beta$$

$$dz = -2(d\beta')[X'Y - X'X\beta]$$

Faça  $dz = \phi$  entendendo-se  $\phi$  como um vetor de zeros.

$$\text{Resulta } X'Y - X'X\beta = \phi$$

$$X'X\hat{\beta} = X'Y$$

$$\text{Faça } X'X = S \text{ logo } S\hat{\beta} = X'Y. \quad (\text{IV})$$

O sistema  $S\hat{\beta} = X'Y$  é conhecido pelo nome de equações normais.

$$\text{Como, neste contexto, } S \text{ geralmente é não-singular então } \hat{\beta} = S^{-1}X'Y \quad (\text{V})$$

que é a solução do sistema.

A condição de segunda ordem é observada através dos menores principais do determinante Hesiano, que são todos positivos.

### 3.3 Teorema

Se  $S$  é não-singular, então  $\hat{\beta}$  é uma estimativa imparcial de  $\beta$ .

Hipótese:  $\exists S^{-1}$

Tese:  $E(\hat{\beta}) = \beta$  ou  $\hat{\beta} \doteq \beta$

Demonstração:

$$\hat{\beta} = S^{-1} X'Y$$

$$\hat{\beta} = S^{-1} X' (X\beta + \varepsilon)$$

$$\hat{\beta} = S^{-1} X'X\beta + S^{-1} X'\varepsilon$$

$$\hat{\beta} = S^{-1} S\beta + S^{-1} X'\varepsilon$$

$$\hat{\beta} = \beta + S^{-1} X'\varepsilon$$

$$E(\hat{\beta}) = \beta + S^{-1} X'\phi$$

$$E(\hat{\beta}) = \beta$$

### 3.4 Análise de Variância

Consideremos:  $\varepsilon'\varepsilon = \hat{e}_1^2 + \hat{e}_2^2 + \dots + \hat{e}_n^2$  que é a soma dos quadrados dos desvios da regressão.

$$\varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$$

$$= (Y - \beta'X') (Y - X\beta)$$

$$= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta$$

Ocorre que  $X'X\beta = X'Y$  logo

$$\varepsilon'\varepsilon = Y'Y - Y'X\beta$$

Como  $Y'X\beta = \beta'X'Y$  por razões já explicitadas anteriormente, resulta:

$$\varepsilon'\varepsilon = Y'Y - \beta'X'Y$$

Substituindo  $\varepsilon'\varepsilon$  por SQDR que significa soma de quadrados de desvios da regressão e introduzindo o fator  $C = (\sum Y)^2 / N$  denominado de correção, tem-se:

$$SQDR = (Y'Y - C) - (\beta'X'Y - C) \quad (VI)$$

SQRD = Soma de quadrado de desvios da regressão

SQT = Soma de quadrado total =  $Y'Y - C$

SQReg = Soma de quadrado da regressão =  $\beta'X'Y - C$

Quadro da análise de variância

Causas da Variação	G. L.	S. Q.	Q. M.	F
Regressão	p - 1	S. Q. Reg	SQReg / (p - 1)	QMR / QMDR
Desvio da regressão	N - p	S. Q. D. R.	SQDR / (N - p)	
Total	N - 1	SQT		

Onde:

G.L. = graus de liberdade

P = número de parâmetros

N = total de dados observados

S.Q. = soma de quadrados

Q.M. = quadrado médio

Q.M.R. =  $SQR / (p - 1)$  = quadrado médio da regressão

Q.M.D.R. =  $SQDR / (N - p)$  = quadrado médio do desvio da regressão

### 3.5 Esperanças matemáticas das somas de quadrados

Sabe-se que:  $SQDR = (Y'Y - \beta'X'Y)$ , logo:

$$E[ SQDR ] = E( Y'Y ) - E( \beta'X'Y ) \quad (VII)$$

$$E( Y'Y ) = E[ (X\beta + \varepsilon)' (X\beta + \varepsilon) ]$$

$$= E[ (\beta'X' + \varepsilon') (X\beta + \varepsilon) ]$$

$$= E[ \beta'X'X\beta + \beta'X'\varepsilon + \varepsilon'X\beta + \varepsilon'\varepsilon ]$$

$$E( Y'Y ) = \beta'X'X\beta + \beta'X'E(\varepsilon) + E(\varepsilon')X\beta + E(\varepsilon'\varepsilon)$$

$$= \beta'X' \quad X\beta + E(\varepsilon'\varepsilon)$$

Pois  $E(\varepsilon)$  e  $E(\varepsilon')$  são dois vetores de zeros sendo o primeiro vetor coluna e o segundo vetor linha.

$$E( Y'Y ) = \beta'S\beta + E(e_1^2 + e_2^2 + \dots + e_n^2)$$

Admitindo homocedasticia, isto é,  $E(e_i^2) = \sigma^2$ , tem-se:

$$E( Y'Y ) = \beta'S\beta + N\sigma^2 \quad (VIII)$$

$$E(\beta'X'Y) = E[(S^{-1}X'Y)' X'Y] = E[Y'XS^{-1}X'Y] = E[(\beta'X' + \varepsilon')XS^{-1}X'(X\beta + \varepsilon)]$$

$$= E[\beta'X'XS^{-1}X'X\beta + \beta'X'XS^{-1}X'\varepsilon + \varepsilon'XS^{-1}X'X\beta + \varepsilon'XS^{-1}X'\varepsilon] =$$

$$= \beta' S \beta + E(\epsilon' X S^{-1} X' \epsilon)$$

Ocorre que:

$$E[\epsilon' A \epsilon] =$$

$$\text{Seja: } \epsilon = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} ; \quad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} ; \quad \epsilon' = [e_1 \quad e_2 \quad e_3]$$

$$\epsilon' A = [a_{11} e_1 + a_{21} e_2 + a_{31} e_3 ; a_{12} e_1 + a_{22} e_2 + a_{32} e_3 ; a_{13} e_1 + a_{23} e_2 + a_{33} e_3]$$

$$\epsilon' A \epsilon = [a_{11} e_1^2 + a_{21} e_1 e_2 + a_{31} e_1 e_3 + a_{12} e_1 e_2 + a_{22} e_2^2 + a_{32} e_2 e_3 + a_{13} e_1 e_3 + a_{23} e_2 e_3 + a_{33} e_3^2]$$

$$E(\epsilon' A \epsilon) = a_{11} \sigma^2 + a_{22} \sigma^2 + a_{33} \sigma^2 = \sigma^2 (a_{11} + a_{22} + a_{33}) = \sigma^2 \text{ Traço de } A$$

Admitiu-se homocedasticidade e independência de erros.

**Teorema:** Se  $A \times B$  e  $B \times A$  existem então  $\text{traço}(A \times B) = \text{traço}(B \times A)$

Logo

$$\text{Traço}(X S^{-1} X') = \text{traço}(S^{-1} X' X) = \text{traço } I_p = p$$

$\underbrace{\quad\quad\quad}_{A \quad B}$

A      B

$$= \beta' S \beta + \sigma^2 \text{traço} (X S^{-1} X') = \beta' S \beta + p \sigma^2 \quad (\text{IX})$$

$$= \beta' S + \sigma^2 \text{traço} I_p$$

Substituindo (VIII) e (IX) em (VII) tem-se:

$$E(\text{SQDR}) = \beta' S \beta + N \sigma^2 - \beta' S \beta - p \sigma^2$$

$$E(\text{SQDR}) = (N - p) \sigma^2 \quad \text{e} \quad E(\text{QMDR}) = [1 / (N - p)] \cdot E(\text{SQDR})$$

$$E(\text{QMDR}) = [1 / (n - p)] \cdot (N - p) \sigma^2$$

$$E(\text{QMDR}) = \sigma^2$$

### 3.6 Matriz de Dispersão

Por definição:

$$D = \{ [\beta - E(\hat{\beta})] [\beta - E(\hat{\beta})]' \}$$

Ocorre que:

$$\hat{\beta} = S^{-1} X' Y$$

$$\hat{\beta} = S^{-1} X' (X \beta + \varepsilon)$$

$$\hat{\beta} = S^{-1} X' X \beta + S^{-1} X' \varepsilon$$

$$\hat{\beta} = \beta + S^{-1} X' \varepsilon \quad (\text{VIII})$$



$$E(\hat{\beta}) = \beta \quad (\text{IX})$$

Faça (VIII) – (IX) e obtenha

$$\beta - E(\hat{\beta}) = S^{-1}X'\varepsilon \quad \text{logo:}$$

$$D = E[S^{-1}X'\varepsilon (S^{-1}X'\varepsilon)']$$

$$= E[S^{-1}X'\varepsilon\varepsilon'XS^{-1}]$$

$$= S^{-1}X' E(\varepsilon\varepsilon')XS^{-1}$$

$$= S^{-1}X' I \sigma^2 X S^{-1} \quad \text{aqui I é matriz identidade}$$

$$= S^{-1}X'X S^{-1}\sigma^2$$

$$D = S^{-1}\sigma^2 \quad (\text{X})$$

## Capítulo 4

O modelo  $Y_i = A + BX_i + e_i$

### 4.1 Estimativa de parâmetros e a equação estimada

De acordo com ( IV ):

$$S\hat{\beta} = X'Y$$

$$\begin{bmatrix} N & \sum X \\ \sum X & \sum X^2 \end{bmatrix} \begin{bmatrix} \hat{A} \\ \hat{B} \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

Faça  $X = X - \bar{X}$  e substitua no sistema acima  $X$  por  $\bar{X}$  o que resulta:

$$\begin{bmatrix} N & \sum X \\ \sum X & \sum X^2 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

$$\begin{bmatrix} N & 0 \\ 0 & \sum X^2 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 1/N & 0 \\ 0 & 1/\sum X^2 \end{bmatrix} \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \sum Y/N \\ \sum XY / \sum X^2 \end{bmatrix}$$

Logo:

$$Y = \frac{\sum Y}{N} + \frac{\sum XY}{\sum X^2} \cdot X$$

Substitua  $X$  por  $X - \bar{X}$  resulta

$$Y = \frac{\sum Y}{N} + \frac{\sum XY}{\sum X^2} \cdot (X - \bar{X})$$

$$Y = \frac{\sum Y}{N} - \bar{X} \cdot \frac{\sum XY}{\sum X^2} + \frac{\sum XY}{\sum X^2} \cdot X \quad (XI)$$

que é a equação procurada.

**4.2 Exercício ( Extraído de J. Johnston, Métodos Econométricos, pág 29 ed. Atlas)**

Acidentes em estradas e veículos licenciados na Grã-Bretanha

ANO	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	Total
Acidentes	166	153	177	201	216	208	227	238	268	268	274	2396
V. Licenc.	352	373	411	441	462	490	529	577	641	692	743	5711

Admita que o número de acidentes é uma função linear do número de veículos licenciados e estime a equação que representará o fenômeno?

Solução:

A equação procurada de acordo com ( XI ) é:

$$\hat{Y} = [ ( \sum Y / N ) - \bar{X} \cdot ( \sum XY / \sum X^2 ) ] + ( \sum XY / \sum X^2 ) \cdot X$$

Ocorre que:

$$N = 11 ; \quad \sum X = 5711 ; \quad \sum Y = 2396 ; \quad \bar{Y} = 217,82 ; \quad \bar{X} = 519,18$$

$X = X - \bar{X}$	-167,18	-146,18	-108,18	-78,18	-57,18	-29,18	9,82	57,82	121,82	172,82	223,82
-------------------	---------	---------	---------	--------	--------	--------	------	-------	--------	--------	--------

$$\sum X^2 = 169495,64 ; \quad \sum XY = 52880,72$$

logo:

$$\hat{Y} = 217,82 - [ 519,18 \cdot ( 52880,72 / 169495,64 ) ] + ( 52880,72 / 169495,64 ) \cdot X$$

$$\hat{Y} = 55,84 + 0,3120 X$$

#### 4.2.1 Interpolação

Calcule o número de acidentes se o número de veículos licenciados é 500?

$$\hat{Y} = 55,84 + ( 0,3120 \cdot 500 ) = 218,84$$

#### 4.2.2 Extrapolação

Calcule o número de acidentes para 750 veículos licenciados?

$$\hat{Y} = 55,84 + ( 0,3120 \cdot 750 ) = 523,84$$

#### 4.2.3 Análise da variância

Cálculo das somas de quadrados:

$$C = ( \sum Y )^2 / N = 521892,36$$

$$SQT = \sum Y^2 - C = 539512,36 - C = 17619,64$$

$$\beta'X'Y = [55,84 \quad 0,3120] \begin{bmatrix} 2396 \\ 1296836 \end{bmatrix} = 538405,47$$

$$SQReglinear = \beta X'Y - C = 16513,11$$

Causas de variação	G. L.	S. Q.	Q. M.	F
Regressão linear	1	16513,11	16513,11	134,32
Desvio da regressão	9	1106,53	122,94	
Total	10	17619,64	-	

Conclusão: Como  $F = 134,32$  é maior que 10,56 rejeita-se a hipótese de nulidade e portanto ao nível de significância de 1% tem-se uma boa regressão.

Coefficiente de determinação  $= r^2 = 0,9372$ ;

Coefficiente de correlação  $= r = + 0,9681$

#### 4.2.4 Matriz de dispersão

Como visto em ( X )  $D = S^{-1}\sigma^2$

$$D = \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum X^2} \end{bmatrix} \cdot \sigma^2$$

OBS: Uma estimativa de  $\sigma^2$  é o quadrado médio do Desvio da Regressão.

$$D = \begin{bmatrix} \frac{1}{11} & 0 \\ 0 & \frac{1}{169495,64} \end{bmatrix} * 122,94$$

$$D = \begin{bmatrix} 11,17636 & 0 \\ 0 & 0,00072 \end{bmatrix}$$

$$\text{logo } V(\hat{A}) = 11,7636 ; \quad V(\hat{B}) = 0,0072 \quad \text{e} \quad \text{Cov}(\hat{A}, \hat{B}) = 0$$

#### 4.3. Variância de Y

$$\text{Considere-se } \hat{Y} = \hat{a} + \hat{b} X$$

$$V(Y) = V(\hat{a}) + V(\hat{b} X) + 2 \text{Cov}(\hat{a}, \hat{b} X)$$

$$V(Y) = (\sigma^2 / N) + X^2 V(\hat{b}) + 2X \text{cov}(\hat{a}, \hat{b})$$

$$V(Y) = (\sigma^2 / N) + [(X^2 / \sum X^2) \cdot \sigma^2] + 0$$

$$V(Y) = [(1/N) + (X_i^2 / (\sum X_i^2))] \cdot \sigma^2$$

#### 4.4. Intervalo de Confiança para Y

$$I. C. = \hat{Y}_i \pm t \cdot \sqrt{[(1/N) + (X_i^2 / \sum X_i^2)] S^2}$$

Onde  $S^2$  é uma estimativa de  $\sigma^2$  e  $t$  depende do nível de significância.

$$I. C. = \hat{Y}_i \pm t \cdot S \cdot \sqrt{[(1/N) + (X_i^2 / \sum X_i^2)]}$$

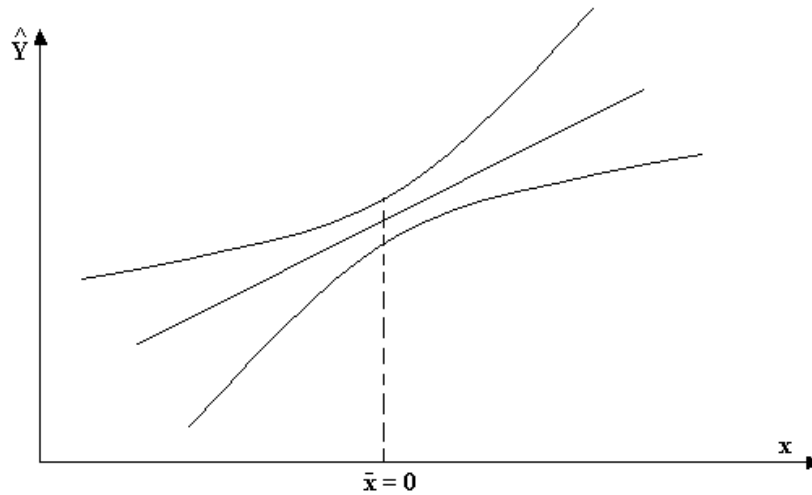
$$\text{Faça } d = t S \sqrt{[(1/N) + (X_i^2 / \sum X_i^2)]}$$

$$d^2 = t^2 S^2 [(1/N) + (X_i^2 / \sum X_i^2)]$$

$$d^2 / (t^2 S^2) - (X_i^2 / \sum X_i^2) = 1 / N$$

$$\{ d^2 / [ ( t^2 S^2 ) / N ] \} - [ X_i^2 / ( \sum X_i^2 / N ) ] = 1$$

Que é a equação de uma hipérbole.



#### 4.5. Intervalo de confiança para a

$$I. C. = \hat{a} \pm t S(\hat{a})$$

$$I. C. = \hat{a} \pm t \sqrt{S^2 / N}$$

$$I. C. = \hat{a} \pm t S \sqrt{1 / N}$$

#### 4.6. Intervalo de confiança para b

$$I. C. = \hat{b} \pm t S(\hat{b})$$

$$I. C. = \hat{b} \pm t \sqrt{S^2 / \sum X^2}$$

$$I. C. = \hat{b} \pm t S \sqrt{1 / \sum X^2}$$

OBS: “t” é o valor da tabela de Student ao nível de significância  $\alpha$  e com os graus de liberdade de  $S^2$ .

**O modelo  $Y_i = A + BX_i + CX_i^2 + e_i$**

#### 4.7 Sistema de equações normais

De acordo com ( IV ):

$$S\hat{\beta} = X'Y$$

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \\ X_1^2 & X_2^2 & \dots & X_n^2 \end{bmatrix} \cdot \begin{bmatrix} 1 & X_1 & \dots & X_1^2 \\ 1 & X_2 & \dots & X_2^2 \\ 1 & X_n & \dots & X_n^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \\ X_1^2 & X_2^2 & \dots & X_n^2 \end{bmatrix} \cdot \begin{bmatrix} Y_1 \\ Y_2 \\ Y_n \end{bmatrix}$$

$$\begin{bmatrix} N & \sum X & \sum X^2 \\ \sum X & \sum X^2 & \sum X^3 \\ \sum X^2 & \sum X^3 & \sum X^4 \end{bmatrix} \cdot \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum XY \\ \sum X^2Y \end{bmatrix}$$

Cuja solução fornece  $\hat{A}$ ,  $\hat{B}$  e  $\hat{C}$ .

Exercício: Sabe-se que Y depende de X de acordo com a equação de uma parábola.

Determine a equação de mínimos quadrados correspondentes aos dados da tabela a seguir:

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

Solução:

$$\sum X = 42,2 ; \quad \sum X^2 = 291,2 ; \quad \sum X^3 = 2275,35 ; \quad \sum X^4 = 18971,93$$



$$\sum Y = 46,4 ; \quad \sum XY = 230,42 ; \quad \sum X^2Y = 1448,99$$

$$\begin{bmatrix} 8 & 42,2 & 291,2 \\ 42,2 & 291,2 & 2275,35 \\ 291,2 & 2275,35 & 18971,93 \end{bmatrix} \cdot \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} 46,4 \\ 230,42 \\ 1448,99 \end{bmatrix}$$

$$\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \frac{1}{222.194,85} \cdot \begin{bmatrix} 347.408,39 & -138.033,53 & 11.222,33 \\ -138.033,53 & 66.978 & -5.914,16 \\ 11.222,33 & -5.914,16 & 548,76 \end{bmatrix} \cdot \begin{bmatrix} 46,4 \\ 230,42 \\ 1.448,99 \end{bmatrix}$$

$$\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} 2,5883 \\ 2,0646 \\ -0,2109 \end{bmatrix}$$

A equação de regressão procurada é

$$Y = 2,5883 + 2,0646 X - 0,2109 X^2$$

#### 4.7.1 Análise da Variância

Calculo das somas de quadrados

$$C = (\sum Y)^2 / N = 269,12$$

$$SQT = \sum Y^2 - C = 290,52 - C = 21,4$$

$$SQ \text{ Reg. Quad} = \hat{\beta}' X' Y - C = 21,11$$

$$\hat{\beta}' X' Y = \begin{bmatrix} 2,5883 & 2,0646 & -0,2109 \end{bmatrix} \begin{bmatrix} 46,40 \\ 230,42 \\ 1448,99 \end{bmatrix} = 290,23$$

Causas da Variação	G. L.	S. Q.	Q. M.	F	F( 5% ) = 5,79
--------------------	-------	-------	-------	---	----------------

Regressão quadrática	2	21,11	10,555	181,98
Desvio da regressão	5	0,29	0,058	
Total	7	21,40		

Conclusão: Como  $F = 181,98$  situa-se na região de rejeição, a hipótese de nulidade será rejeitada e conclui-se que existe uma boa regressão ao nível de significância de 5%.

Coeficiente de determinação  $= r^2 = 0,9864$

$$\text{O modelo } Y_i = A + B_1 X_{1i} + B_2 X_{2i} + \dots + B_p X_{pi} + e_i$$

### 5.1 Sistema de equações normais

De acordo com ( IV ):

$$S\hat{\beta} = X'Y$$

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{12} & \dots & X_{1n} \\ \dots & \dots & \dots & \dots \\ X_{p1} & X_{p2} & \dots & X_{pn} \end{bmatrix} \cdot \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{p1} \\ 1 & X_{12} & X_{22} & \dots & X_{p2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} \cdot \begin{bmatrix} \hat{A} \\ \hat{B}_1 \\ \dots \\ \hat{B}_p \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{12} & \dots & X_{1n} \\ \dots & \dots & \dots & \dots \\ X_{p1} & X_{p2} & \dots & X_{pn} \end{bmatrix} \cdot \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix}$$

$$\begin{bmatrix} N & \sum X_{1i} & \sum X_{2i} & \dots & \sum X_{pi} \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i} \cdot X_{2i} & \dots & \sum X_{1i} \cdot X_{pi} \\ \sum X_{2i} & \sum X_{1i} \cdot X_{2i} & \sum X_{2i}^2 & \dots & \sum X_{2i} \cdot X_{pi} \\ \dots & \dots & \dots & \dots & \dots \\ \sum X_{pi} & \sum X_{1i} \cdot X_{pi} & \sum X_{2i} \cdot X_{pi} & \dots & \sum X_{pi}^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{A} \\ \hat{B}_1 \\ \hat{B}_2 \\ \dots \\ \hat{B}_p \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_{1i} Y_i \\ \sum X_{2i} Y_i \\ \dots \\ \sum X_{pi} Y_i \end{bmatrix}$$

Cuja solução fornece  $\hat{A}, \hat{B}_1, \hat{B}_2, \dots, \hat{B}_p$ .

Exercício: A tabela abaixo indica o peso em kg (  $Y$  ) em função da altura em cm (  $X_1$  ) e idade em meses (  $X_2$  ) de 12 porcos mestiços. Estabeleça a equação de regressão do tipo  $Y = A + BX_1 + CX_2$ .

Peso	64	71	53	67	55	58	77	57	56	51	76	68
Altura	57	59	49	62	51	50	55	48	52	42	61	57
Idade	8	10	6	11	8	7	10	9	10	6	12	9

Solução:

$$N = 12; \quad \sum X_{1i} = 643; \quad \sum X_{2i} = 106; \quad \sum X_{1i}^2 = 34843; \quad \sum X_{2i}^2 = 976;$$

$$\sum X_{1i} X_{2i} = 5779; \quad \sum Y = 753; \quad \sum X_{1i} Y_i = 40830; \quad \sum X_{2i} Y_i = 6796.$$

$$\begin{bmatrix} 12 & 643 & 106 \\ 643 & 34.843 & 5.779 \\ 106 & 5.779 & 976 \end{bmatrix} \cdot \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} 753 \\ 40.830 \\ 6.796 \end{bmatrix}$$

$$\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \frac{1}{67.116} \cdot \begin{bmatrix} 609.927 & -14.994 & 22.539 \\ -14.994 & 476 & -1.190 \\ 22.539 & -1.190 & 4.667 \end{bmatrix} \cdot \begin{bmatrix} 753 \\ 40.830 \\ 6.796 \end{bmatrix}$$

$$\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} 3,6512 \\ 0,8546 \\ 1,5063 \end{bmatrix}$$

A equação de regressão procurada é:

$$Y = 3,6512 + 0,8546 X_1 + 1,5063 X_2$$

## 5.2 Análise da Variância

Cálculo das somas de quadrados

$$C = \frac{(\sum Y)^2}{N} = 47.250,75$$

$$SQT = \sum Y^2 - C = 48139, - C = 888,25$$

$$\beta'X'Y = [3,6512 \quad 0,8546 \quad 1,5063] \cdot \begin{bmatrix} 753 \\ 40.830 \\ 6.796 \end{bmatrix} = 47.879,49$$

$$SQ_{\text{Reg. Múltipla}} = \beta'X'Y - C = 628,74$$

Causas da Variação	G. L.	S. Q.	Q. M.	F	F(5%) = 4,26
Regressão Múltipla	2	628,74	314,37	10,90	
Desvio da Regressão	9	259,51	28,83		
Total	11	888,25			

Conclusão: Como  $F = 10,90$  situa-se na região de rejeição, a hipótese de nulidade será rejeitada e conclui-se por uma boa regressão ao nível de significância de 5%.

Coeficiente de determinação  $= r^2 = 0,7078$ .

### Exercícios:

#### Regressão Linear

1. Considere os dados da tabela abaixo e construa a reta de mínimos quadrados, admitindo que a altura do filho é função da altura do pai. As medidas estão em metros.

Altura do filho	y	1,65	1,63	1,67	1,64	1,68	1,62	1,70	1,66	1,68	1,71
Altura do pai	x	1,68	1,66	1,68	1,65	1,69	1,66	1,68	1,65	1,71	1,70

Solução:

a) A equação procurada de acordo com ( XI ) é

$$Y = \frac{\sum Y}{N} - \bar{X} \cdot \frac{\sum XY}{\sum X^2} + \frac{\sum XY}{\sum X^2} \cdot \bar{X}$$

Ocorre que:

$$N = 10; \quad \sum X = 16,76; \quad \sum Y = 16,64; \quad \bar{Y} = 1,664; \quad \bar{X} = 1,676$$

$x = X - \bar{X}$	0,004	-0,016	0,004	-0,026	0,014	-0,016	0,004	-0,026	0,034	0,024
-------------------	-------	--------	-------	--------	-------	--------	-------	--------	-------	-------

$$\sum x^2 = 0,00384; \quad \sum xY = 0,00396$$

Logo:

$$Y = 1,664 - \left( 1,676 \cdot \frac{0,00396}{0,00384} \right) + \left( \frac{0,00396}{0,00384} \right) \cdot X$$

$$Y = -0,064375 + 1,03125 X$$

Análise da variância

$$C = \frac{(\sum Y)^2}{N} = 27,69$$

$$SQ_{Tot} = Y'Y - C = 27,70 - 27,69 = 0,01$$

$$SQ_{RegLin} = \beta' X'Y - C = 0,00036$$

$$\hat{\beta}' X'Y = [-0,064375 \quad 1,03125] \cdot \begin{bmatrix} 16,64 \\ 27,89 \end{bmatrix} = 27,69036$$

Causa de variância	G. L.	S. Q.	Q. M.	F	F( 5% ) = 0,0010
Regressão Linear	1	0,000036	0,00036	0,2987	
Desvio da Regressão	8	0,00964	0,001205		
TOTAL	9	0,001			

$$\alpha = 5\%; \quad F_{1,8} = 0,0010$$

Conclusão: Rejeita-se  $H_0$

$$\text{Coeficiente de determinação} = r^2 = 0,0037$$

### Regressão quadrática

A tabela abaixo indica a população de uma cidade dos anos de 1950 a 1959, dados fictícios. Determine a equação representativa do fenômeno admitindo o modelo parabólico.

ANO	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
População em 1.000.000 hab	23,2	31,4	39,8	50,2	62,9	76,0	92,0	105,7	122,8	131,7

Solução:

É claro que adotaremos a variável ano como independente e a população como variável dependente.

Observe que os anos formam uma progressão aritmética de números par e quando isto ocorre é permitido substituí-los por outra série obedecendo a seguinte correspondência:

ANO ( X )	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
ANO ( X )	-9	-7	-5	-3	-1	1	3	5	7	9

Esta transformação acima tem por objetivo diminuir o trabalho aritmético.

Desta forma de acordo com ( IV )

$$S\hat{\beta} = X'Y$$

$$\begin{bmatrix} N & \sum X & \sum X^2 \\ \sum X & \sum X^2 & \sum X^3 \\ \sum X^2 & \sum X^3 & \sum X^4 \end{bmatrix} \cdot \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} Y \\ \sum XY \\ \sum X^2 Y \end{bmatrix}$$

Como:

$$N = 10; \quad \sum X = 0; \quad \sum X^2 = 330; \quad \sum X^3 = 0; \quad \sum X^4 = 19338; \quad \sum Y = 735,7$$

$$\sum XY = 2084,3; \quad \sum X^2 Y = 25158,9$$

Logo:

$$\begin{bmatrix} 10 & 0 & 330 \\ 0 & 330 & 0 \\ 330 & 0 & 19.338 \end{bmatrix} \cdot \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} 735,7 \\ 2.084,3 \\ 25.158,9 \end{bmatrix}$$

$$\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \frac{1}{27.878.400} \cdot \begin{bmatrix} 6.381.540 & 0 & -108.900 \\ 0 & 84.480 & 0 \\ -108.900 & 0 & 3300 \end{bmatrix} \cdot \begin{bmatrix} 735,7 \\ 2.084,3 \\ 25.158,9 \end{bmatrix}$$

$$\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} 70,1294 \\ 6,3161 \\ 0,1043 \end{bmatrix}$$

e a equação procurada é  $\hat{Y} = 70,1294 + 6,3161 X + 0,1043 X^2$ .



## Regressão múltipla

A tabela abaixo indica o peso correspondente a altura e a idade de 10 crianças. Estabeleça a equação de regressão múltipla que explica o fenômeno:

Peso ( y )(kg)	50	55	30	60	45	70	65	80	90	75
Altura ( m ) $x_1$	1,52	1,55	1,25	1,55	1,50	1,75	1,70	1,82	1,80	1,70
Idade ( anos ) $x_2$	14	16	8	18	15	20	18	21	22	17

Solução:

O modelo é  $Y = A + BX_1 + CX_2$

Desta forma de acordo com ( IV )

$$S\hat{\beta} = X'Y$$

$$\begin{bmatrix} N & \sum X_{1i} & \sum X_{2i} \\ \sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i}X_{2i} \\ \sum X_{2i} & \sum X_{1i}X_{2i} & \sum X_{2i}^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_{1i}Y_i \\ \sum X_{2i}Y_i \end{bmatrix}$$

Como:

$$N = 10; \quad \sum X_{1i} = 16,14; \quad \sum X_{1i}^2 = 26,32; \quad \sum X_{2i} = 169; \quad \sum X_{2i}^2 = 3003;$$

$$\sum X_{1i} X_{2i} = 278,8; \quad \sum Y_i = 620; \quad \sum X_{1i} Y_i = 1027,35; \quad \sum X_{2i} Y_i = 11080$$

$$\begin{bmatrix} 10 & 16,14 & 169 \\ 16,14 & 26,32 & 278,8 \\ 169 & 278,8 & 3003 \end{bmatrix} \cdot \begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} 620 \\ 1.027,35 \\ 11.080 \end{bmatrix}$$

$$\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \frac{1}{32,2592} \cdot \begin{bmatrix} 1.309,52 & -1.351,22 & 51,75 \\ -1.351,22 & 1.469 & -60,34 \\ 51,75 & -60,34 & 2,7004 \end{bmatrix} \cdot \begin{bmatrix} 620 \\ 1.027,35 \\ 11.080 \end{bmatrix}$$

$$\begin{bmatrix} \hat{A} \\ \hat{B} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} -89,38 \\ 88,46 \\ 0,47 \end{bmatrix}$$

E a equação procurada é  $\hat{Y} = -89,38 + 88,46 X_1 + 0,47 X_2$

**Teorema:** Se  $Y_i = \hat{a} + bx_i$  é a estimativa de  $Y = a + bx_i$ , onde  $x_i = X_i - \bar{X}$ , então  $E(\hat{a}) = a$ ;  $V(\hat{a}) = \sigma^2 / n$ ;  $E(b) = b$ ;  $V(b) = \sigma^2 / \sum x_i^2$ .

Demonstração:  
que:

$$\hat{a} = \bar{Y}$$

$$\hat{a} = 1 / n ( Y_1 + Y_2 + \dots + Y_n )$$

$$E(\hat{a}) = ( 1 / n ) \cdot [ E( Y_1 ) + E( Y_2 ) + \dots + E( Y_n ) ]$$

$$E(\hat{a}) = ( 1 / n ) \cdot \sum E( Y_i )$$

$$E(\hat{a}) = ( 1 / n ) \cdot \sum ( a + bx_i )$$

$$E(\hat{a}) = ( 1 / n ) \cdot ( na + b \sum x_i )$$

$$E(\hat{a}) = a$$

$$V(\hat{a}) = ( 1 / n^2 ) \cdot \{ [ V( Y_1 ) + V( Y_2 ) + \dots + V( Y_n ) ] + \sum D.P. \}$$

Admitindo homocedasticia e independência dos  $Y_i$

$$V(\hat{a}) = ( 1 / n^2 ) \cdot n \sigma^2 \quad ; \quad V(\hat{a}) = \sigma^2 / n$$

Sabe-se que:

$$\hat{b} = (\sum x_i \cdot Y_i) / (\sum x_i^2)$$

$$\text{Faça: } k = \sum x_i^2 \quad ; \quad \hat{b} = (\sum x_i \cdot Y_i) / k \quad ; \quad \hat{b} = \sum [ (x_i / k) \cdot Y_i ] \quad ;$$

$$\text{Faça: } w_i = x_i / k \quad ; \quad \hat{b} = \sum w_i \cdot Y_i \quad ; \quad \hat{b} = w_1 \cdot Y_1 + w_2 \cdot Y_2 + \dots + w_n \cdot Y_n$$

$$E(\hat{b}) = w_1 E(Y_1) + w_2 E(Y_2) + \dots + w_n E(Y_n)$$

$$E(\hat{b}) = \sum w_i E(Y_i)$$

$$E(\hat{b}) = \sum w_i (a + bx_i)$$

$$E(\hat{b}) = a \sum w_i + b \sum w_i x_i$$

$$E(\hat{b}) = a \sum (x_i / k) + b \sum (x_i / k) x_i$$

$$E(\hat{b}) = (a / k) \sum x_i + (b / k) \sum x_i^2$$

$$E(\hat{b}) = (b / k) \cdot k \quad ; \quad E(\hat{b}) = b$$

$$V(\hat{b}) = w_1^2 V(Y_1) + w_2^2 V(Y_2) + \dots + w_n^2 V(Y_n) + \sum D.P.$$

Admitindo homocedasticia e independência de erros.

$$V(\hat{b}) = \sigma^2 \sum w_i^2$$

$$V(\hat{b}) = \sigma^2 \sum (x_i^2 / k^2)$$

$$V(\hat{b}) = (\sigma^2 / k^2) \sum x_i^2$$

$$V(\hat{b}) = [ \sigma^2 / (\sum x_i^2)^2 ] \cdot \sum x_i^2$$

$$V(\hat{b}) = \sigma^2 / \sum x_i^2$$

Seja o modelo  $Y = a + bx$

a) Estimativa de parâmetros

$$\begin{cases} \sum Y = Na + b \sum x \\ \sum xY = a \sum x + b \sum x^2 \end{cases} \quad ; \quad \hat{a} = \sum Y / N \quad ; \quad \hat{a} = \bar{Y}$$

$$\quad ; \quad \hat{b} = (\sum xY) / \sum x^2$$

b) Matriz de dispersão :  $D = S^{-1} \sigma^2$

$$S = \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \end{bmatrix} \cdot \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} = \begin{bmatrix} N & \sum x \\ \sum x & \sum x^2 \end{bmatrix} = \begin{bmatrix} N & 0 \\ 0 & \sum x^2 \end{bmatrix}$$

$$S^{-1} = \frac{1}{N \cdot \sum x^2} \cdot \begin{bmatrix} \sum x^2 & 0 \\ 0 & N \end{bmatrix} = \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{\sum x^2} \end{bmatrix}$$

$$D = \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{\sum x^2} \end{bmatrix} \cdot \sigma^2$$

$$V(\hat{a}) = \frac{\sigma^2}{N} \quad ; \quad V(\hat{b}) = \frac{\sigma^2}{\sum x^2} \quad ; \quad \text{Cov}(\hat{a}, \hat{b}) = 0$$

c) Intervalo de confiança para Y

$$\hat{Y} = \hat{a} + \hat{b}x$$

$$V(\hat{Y}) = V(\hat{a}) + x^2 V(\hat{b}) + 2x \text{Cov}(\hat{a}, \hat{b})$$

$$V(\hat{Y}) = (\sigma^2 / N) + [(x^2 \sigma^2) / \sum x^2] = [(1 / N) + (x^2 / \sum x^2)] \sigma^2$$

$$S(\hat{Y}) = \sigma \sqrt{(1 / N) + (x^2 / \sum x^2)}$$

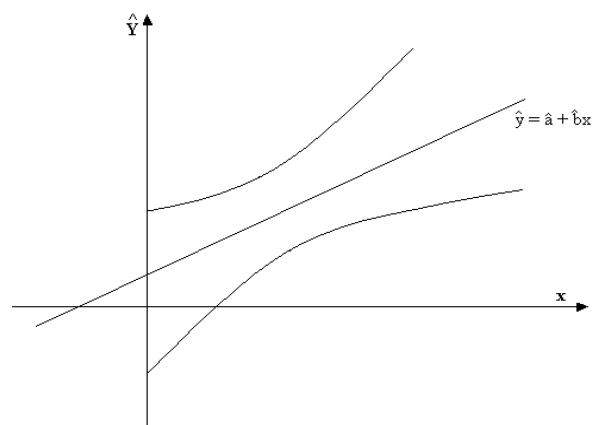
$$\text{I.C.: } \hat{Y} \pm t S(\hat{Y})$$

$$\hat{Y} \pm t \sigma \sqrt{(1 / N) + (x^2 / \sum x^2)}$$

$$\text{Faça: } d = t \sigma \sqrt{(1 / N) + (x^2 / \sum x^2)}$$

$$d^2 = t^2 \sigma^2 [(1 / N) + (x^2 / \sum x^2)]$$

$$d^2 / (t^2 \sigma^2) = (1 / N) + (x^2 / \sum x^2)$$



$$d^2 / [ (t^2 \sigma^2) / N ] = 1 + [ x^2 / ( \sum x^2 / N ) ]$$

$$\{ d^2 / [ (t^2 \sigma^2) / N ] \} - [ x^2 / ( \sum x^2 / N ) ] = 1; \quad \text{Equação de uma hipérbole}$$

Exercício: Considere os dados abaixo:

Y	X	x = X - 4,5	x . Y	x <sup>2</sup>
17	6	1,5	25,5	2,25
11	4	-0,5	-5,5	0,25
18	8	3,5	63,0	12,25
8	3	-1,5	-12,0	2,25
5	1	-3,5	-17,5	12,25
14	5	0,5	7,0	0,25
73	27		60,5	29,50

a) Estimativa dos parâmetros

$$\hat{Y} = \hat{a} + \hat{b}x$$

$$\begin{cases} \sum Y = Na + b \sum x & \Rightarrow \hat{a} = \bar{Y} = 12,1667 \\ \sum x Y = a \sum x + b \sum x^2 & \Rightarrow \hat{b} = ( \sum x Y ) / \sum x^2 = 60,5 / 29,5 = 2,0508 \end{cases}$$

$$\hat{Y} = 12,1667 + 2,0508x$$

$$\hat{Y} = 12,1667 + 2,0508 ( X - 4,5 )$$

$$\hat{Y} = 12,1667 - 9,2286 + 2,0508X$$

$$\hat{Y} = 2,9381 + 2,0508X$$

Causas da variância	G.L.	S.Q.	Q.M.	F	F(5%)
Regressão Linear	1	124,0758	124,0758	73,44	7,71
Desvio da regressão	4	6,7575	1,6894		
Total	5	130,8333			

$$C = \frac{(\sum Y)^2}{n} = \frac{73^2}{6} = 888,1667$$

$$SQ_{Total} = \sum Y^2 - C = 1019 - C = 130,8333$$

$$SQ_{Reg.Linear} = \beta' X' Y - C = [12,1667 \quad 2,0508] \cdot \begin{bmatrix} 73 \\ 60,5 \end{bmatrix} - C = 1.012,2425 - C = 124,0758$$

Conclusão: A regressão linear é significativa para  $\alpha = 5\%$

c) Matriz de dispersão :  $D = S^{-1} \sigma^2$

$$D = \frac{1}{177} \cdot \begin{bmatrix} 29,5 & 0 \\ 0 & 6 \end{bmatrix} \cdot 1,6894 = \begin{bmatrix} 0,2816 & 0 \\ 0 & 0,0573 \end{bmatrix}$$

$$V(\hat{a}) = 0,2816 \quad ; \quad V(\hat{b}) = 0,0573 \quad ; \quad \text{Cov}(\hat{a}, \hat{b}) = 0$$

d) Teste de hipóteses

$$d.1) H_0: a = 0 \quad ; \quad H_a: a \neq 0$$

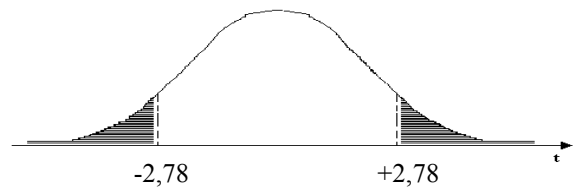
$$t = (\hat{a} - 0) / S(\hat{a}) = 12,1667 / 0,5307 = 22,9258$$

Conclusão: Aceita-se  $H_a$ , i. é, rejeita-se  $H_0$

$$d.2) H_0: b = 0 \quad ; \quad H_a: b \neq 0$$

$$t = (\hat{b} - 0) / S(\hat{b}) = 2,0508 / 0,2394 = 8,5664$$

Conclusão: Rejeita-se  $H_0$



e) Variância de  $\hat{Y}$

$$\hat{Y} = \hat{a} + \hat{b} x$$

$$V(\hat{Y}) = V(\hat{a}) + x^2 V(\hat{b}) + 2x \text{Cov}(\hat{a}, \hat{b})$$

$$V(\hat{Y}) = 0,2816 + 0,0573 x^2$$

f) Estimativa por intervalo de  $Y$  p/  $X = 5$

$$\text{IC: } \hat{Y} \pm t S(\hat{Y})$$

$$\text{IC: } 13,1921 \pm 2,78 \cdot 0,5440$$

$$13,1921 \pm 1,5122$$

$$11,6799 < Y < 14,7043$$

$$\hat{Y} = 2,9381 + 2,0508X$$

$$V(\hat{Y}) = V(\hat{a}) + x^2 V(\hat{b}) + 2x \text{Cov}(\hat{a}, \hat{b})$$

$$V(\hat{Y}) = 0,2816 + 0,0573x^2$$

$$V(\hat{Y}) = 0,2816 + 0,0573(X - 4,5)^2$$

$$V(\hat{Y}) = 0,2816 + 0,0573 \cdot 0,25$$

$$V(\hat{Y}) = 0,2959$$

Para os dados abaixo

Consumo ( Y )	1,99	2,04	2,16	2,18	2,24	2,35	2,38	2,56	2,64	2,70
Renda ( X )	2,12	2,14	2,31	2,37	2,44	2,55	2,57	2,73	2,84	2,90

Determine:

- A equação linear correspondente;
- A análise de variância;
- A matriz de dispersão;
- Os testes de hipóteses;
- Intervalo de confiança de Y para

$$\sum x Y = 0,6014$$

$$\sum x^2 = 0,6664$$

Solução:

$$a) \hat{A} = 0,0705 \quad ; \quad \hat{B} = 0,9025 \quad \hat{Y} = 0,0705 + 0,9025X$$

$$\hat{Y} = 0,0705 + 0,9025 (X + \bar{X})$$

$$\hat{Y} = 0,0705 + 0,9025 (x + 2,497)$$

$$\hat{a} = 2,3240 \quad ; \quad \hat{b} = 0,9025 \quad \hat{Y} = 2,3240 + 0,9025x$$

b) Análise da variância

Causas da variância	G.L.	S.Q.	Q.M.	F	F(5%)
Regressão Linear	1	0,5427	0,5427	904,50	5,12
Desvio da regressão	8	0,0049	0,0006		
Total	9	0,5476	-		

$$C = \frac{(\sum Y)^2}{n} = \frac{23,24^2}{10} = 54,0098$$

$$SQ_{Total} = \sum Y^2 - C = 0,5476$$

$$SQ_{Reg.Linear} = \beta' X' Y - C = [2,3240 \quad 0,9025] \cdot \begin{bmatrix} 23,24 \\ 0,6014 \end{bmatrix} - C = 0,5427$$

Conclusão: A regressão é significativa.

c) Matriz de dispersão :  $D = S^{-1} \sigma^2$

$$D = \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{\sum x^2} \end{bmatrix} \cdot \sigma^2 = \begin{bmatrix} 0,00006 & 0 \\ 0 & 0,0009 \end{bmatrix}$$

$$V(\hat{a}) = 0,00006 \quad ; \quad V(\hat{b}) = 0,0009 \quad ; \quad Cov(\hat{a}, \hat{b}) = 0$$

d) Teste de hipóteses

$$d.1) H_0: a = 0 \quad ; \quad H_a: a \neq 0 \quad ; \quad \alpha = 5\%$$

$$t = (\hat{a} - 0) / S(\hat{a}) = 2,324 / 0,0077 = 301,82$$

Conclusão: Rejeita-se  $H_0$

$$d.2) H_0: b = 0 \quad ; \quad H_a: b \neq 0 \quad ; \quad \alpha = 5\%$$

$$t = (\hat{b} - 0) / S(\hat{b}) = 0,9025 / 0,03 = 30,08$$

Conclusão: Rejeita-se  $H_0$



e) Intervalo de confiança de Y para  $X = 2,30$

$$IC: \hat{Y} \pm t s(\hat{Y}) \quad V(\hat{Y}) = 0,00006 + 0,0009 (2,3 - 2,497)^2$$

$$Y = 0,0705 + 0,9025 X \quad V(\hat{Y}) = 0,00006 + 0,0009 \cdot 0,0388$$

$$Y = 0,0705 + 0,9025 \cdot 2,3 \quad V(\hat{Y}) = 0,000095$$

$$Y = 2,1463 \quad V(\hat{Y}) = 0,00974$$

$$IC: 2,1463 \pm 2,31$$

$$V(\hat{Y}) = V(\hat{a}) + x^2 V(\hat{b})$$

$$V(\hat{Y}) = 0,00006 + 0,0009 x^2$$

Exercícios:



A tabela abaixo indica o aumento dos preços em função dos aumentos salariais.

Indústria	Aumento percentual dos preços ( y )	Aumento percentual dos salários ( X )
Bens duráveis	13,0	19,3
Artigos elétricos	21,9	18,2
Produtos de borracha	20,5	20,2
Construção Civil	9,8	21,0
Produtos de aço	30,7	26,4
Mineração de carvão	13,4	22,6
Automóveis	14,1	19,2
Produtos químicos	3,5	22,4

Pede-se:

- A equação de regressão linear;
- Análise de variância;
- Matriz de dispersão;
- Testes de hipóteses;
- Intervalo de confiança de Y para  $x = 24,0$ .

Solução:

- Equação de regressão linear

$$\hat{Y} = -4,5165 + 0,9630 X$$

$$\hat{Y} = -4,5165 + 0,9630 (x + \bar{X})$$

$$\hat{Y} = -4,5165 + 0,9630 x + 0,9630 \bar{X}$$

$$\hat{Y} = -4,5165 + 0,9630 x + 0,9630 (21,1625)$$

$$\hat{Y} = 15,8630 + 0,9630 x$$

- Análise de variância

Causas da variância	G.L.	S.Q.	Q.M.	F	F(5%)
Regressão Linear	1	44,6504	44,6504	0,6083	0,0011
Desvio da regressão	6	440,4096	73,4016		
Total	7	485,06			

$$C = \frac{(\sum Y)^2}{n} = \frac{126,9^2}{8} = 2.012,95$$

$$SQ_{Total} = \sum Y^2 - C = 2498,01 - C = 485,06$$

$$SQReg. = \beta'X'Y - C = [15,863 \quad 0,963] \cdot \begin{bmatrix} 126,9 \\ 46,2988 \end{bmatrix} - C = 44,6504$$

Conclusão: Não existe diferenças significativas para  $\alpha = 5\%$

c) Matriz de dispersão :  $D = S^{-1} \sigma^2$

$$D = \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{\sum x^2} \end{bmatrix} \cdot \sigma^2 = \begin{bmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{48,0788} \end{bmatrix} \cdot 73,4016 = \begin{bmatrix} 9,1752 & 0 \\ 0 & 1,5267 \end{bmatrix}$$

$$V(\hat{a}) = 9,1752 \quad ; \quad V(\hat{b}) = 1,5267 \quad ; \quad Cov(\hat{a}, \hat{b}) = 0$$

d) Teste de hipóteses

$$H_0: a = 0 \quad ; \quad H_a: a \neq 0 \quad ; \quad \alpha = 5\%$$

$$t = (\hat{a} - 0) / s(\hat{a}) = 15,8630 / 3,0291 = 5,2369$$

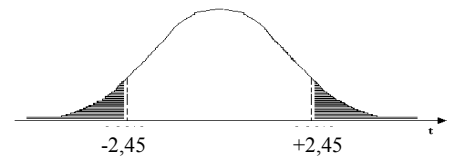
Conclusão: Rejeita-se  $H_0$

$$H_0: b = 0 \quad ; \quad H_a: b \neq 0$$

$$t = (\hat{b} - 0) / s(\hat{b}) = 0,9630 / 1,2356 = 0,7794$$

Conclusão: Aceita-se  $H_0$ . Este teste confirma a não significância na análise da variância.

e) Não há mais que se falar em estimativas para uma regressão reprovada.



## Exercícios Propostos

### I - Regressão Linear:

1. Uma empresa desejando estudar a influência das despesas com publicidade no total das vendas apresentou ao seu serviço de estatística as seguintes informações:

Vendas em R\$ 1000,00	120	190	240	140	180	280	150	115	215	220	320
Despesas em R\$ 1000,00 com publicidade	2,5	6,5	11,0	4,0	8,5	14,0	6,0	5,0	10,0	13,5	16,0

Determine a equação de regressão linear correspondente.

2. Admitindo que o peso dos filhos aos 25 anos depende do peso que o pai tinha na mesma idade estabeleça a equação de regressão linear para os dados.

Peso dos pais (kg)	63	61	65	63	66	60	68	66	67	68	71
Peso dos filhos (kg)	66	64	66	64	67	64	66	65	70	68	70

3. Em uma Prefeitura a sua receitas em milhões de reais de 1950 a 1959 encontra-se na tabela baixo:

Ano	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
Receita	21,8	31,6	41,8	53,2	64,5	76,8	90,2	92,4	104,6	112,0

Determine a equação de regressão linear.

4. Repita o exercício anterior acrescentando a receita de 1960 que foi de 125,2 milhões de reais.

5. A tabela indica a demanda de um determinado bem a um determinado preço:

Demanda (μ. f.)	180	220	160	320	230	245	180	300	305
Preço (R\$)	100	100	120	120	125	130	160	170	180

6. Uma pesquisa no mercado de carros usados do Shopping Rio Mar apresentou os preços dos veículos vendidos com os respectivos anos de fabricação conforme tabela abaixo:

1.000 km	35	10	25	50	30	15	70	40	55	20	45	65	60
R\$ 1.000,00	5,0	7,1	5,9	3,7	5,7	6,7	2,3	4,4	3,6	6,5	4,3	2,8	3,0

Obtenha o modelo que permita estimar o preço do veículo com base na quilometragem.

7. A tabela abaixo indica por ano as vendas em R\$ 1.000,00 de uma pequena empresa. Determine o provável faturamento para o ano de 1994.

Ano	84	85	86	87	88	89	90	91	92	93
Vendas	4	10	9	20	25	22	31	32	40	42

8. Repita o exercício anterior para os dados da tabela:

Ano	1	2	3	4	5	6	7	8	9	10
Vendas	4	10	9	20	25	22	31	32	40	42

9. Repita o exercício 2 Para os dados da tabela:

Ano	-9	-7	-5	-3	-1	1	3	5	7	9
Vendas	4	10	9	20	25	22	31	32	40	42

10. A tabela abaixo indica o PIB<sub>a.p.m.</sub> do Estado de Sergipe em US\$ 1.000.000 de 1970 a 1993.

Ano	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
PIB <sub>a.p.m.</sub>	766,3	914,2	1013,6	1232,5	1238,6	1344,0	1465,6	1658,2	1767,7	1956,2

Ano	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
PIB <sub>a.p.m.</sub>	2027,1	2092,0	2341,4	2286,6	2410,3	2724,5	3190,7	3016,4	3091,4	3290,1

Ano	1990	1991	1992	1993
PIB <sub>a.p.m.</sub>	3105,6	3091,4	3044,8	2955,6

Fonte: SEDENE/DPO/Contas Regionais

Estabeleça a equação de regressão e estime o PIB para o ano de 1994.

11. Um professor de educação física da UFS anotou o tempo de treinamento dos seus atletas e o número de pontos obtidos em um teste de aptidão física, conforme tabela abaixo:

Tempo (hs)	5	3	4	1	6	8	10	9	7	2	11
Pontos	30	25	27	15	38	45	51	47	41	19	55

Estabeleça a equação de regressão e estime o número de pontos que um atleta atingiria se ele treinasse 12 horas?

12. A tabela abaixo indica o tempo médio de estudo de um grupo de alunos e as respectivas notas obtidas em uma prova de estatística.

Horas	2	4	6	7	8	9	10	12
Notas	24	32	52	60	68	80	84	100

Estabeleça a equação de regressão e estime o número de pontos que um aluno alcançaria se estudasse 11 horas.

12. Um aluno do curso de Estatística da UFS registrou as horas de estudos por disciplina e as notas obtidas, que constam da tabela abaixo:

Horas de estudo	14	13	22	19	23	35	26	21	11	15	17	20
Nota	61	56	74	70	75	96	82	75	50	65	65	70

Estabeleça a equação de regressão correspondente.

## II – Regressão quadrática

1. A população de uma cidade em mil habitantes dos anos de 1950 a 1957 encontra-se na tabela abaixo:

Anos	1950	1951	1952	1953	1954	1955	1956	1957
População (1000 hab.)	232	314	502	629	760	920	1057	1228

Admita uma relação quadrática e estabeleça a equação correspondente.

2. Repita o exercício anterior acrescentando a população de 1958 que foi de 1300 mil habitantes.

3. A tabela abaixo indica a produção de milho em uma determinada lavoura em função da presença de nitrogênio.

Produção (kg)	245	287	173	1154	1297	1276	1100
Nitrogênio (kg)	0	8	16	8	20	15	18

Admita uma relação quadrática e estabeleça a equação de regressão correspondente.

4. Com os dados da tabela abaixo estabeleça as equações do espaço percorrido, da velocidade e da aceleração do móvel que a desenvolveu?

Tempo (Seg)	1	2	3	4	5	6	7	8	9	10
Distância (m)	13	20	30	42	57	76	97	122	150	180

5. Os dados da tabela abaixo indicam as taxas mensais de desempenho do ano de 1999. Estabeleça uma equação que represente o fenômeno?

Mês	Jan	fev	mar	abril	maio	junho	julho	ago	set	out	nov	dez
Taxa	15,4	14,8	14,3	14,0	13,8	13,5	13,3	13,6	14,1	14,9	15,0	15,2

Fonte: Dados fictícios

6. Os dados da tabela abaixo apresenta preço, demanda e oferta de um determinado produto. Estabeleça as equações de demanda e ofertas e o preço de equilíbrio desse mercado.

Preço (R\$)	8	7	6	5	4	3	2	1	0
Demanda (kg)	0,0	3,1	4,8	5,8	6,6	7,5	8,3	8,9	9,5
Oferta (kg)	10,1	9,6	8,8	8,0	6,8	5,0	2,3	0,0	0,0

7. Os dados da tabela abaixo indica o ano e a respectiva produção de uma empresa. Estabeleça a equação da produção em função do ano?

Ano	88	89	90	91	92	93	94
Produção (ton)	40	45	51	50	59	69	76

8. Um fabricante de “Pinga Boa Morte” estabeleça a seguinte tabela:

Cana (kg)	52,0	50,2	50,3	49,2	48,0	48,0	45,4	42,6	40,0	43,2
Pinga (l)	1,0	1,5	1,7	1,2	1,8	2,1	2,3	1,9	2,1	2,4

Determine a equação de transformação de cana para pinga?

9. Um fabricante estabeleceu a seguinte meta de produção para a sua empresa:

Ano	1	2	3	4	5	6
Produção (ton)	4,5	5,2	5,0	5,8	6,7	7,0

Determinar a equação de produção projetada?

### III – Regressão múltipla

1. A tabela abaixo indica uma variável Y dependente de duas outras  $X_1$  e  $X_2$ . Estabeleça a equação de regressão linear múltipla  $Y = a + bX_1 + cX_2$ .

Y	6,4	7,1	5,3	6,7	5,5	5,8	7,7	5,7	5,6	5,1	7,6	6,8
$X_1$	5,7	5,9	4,9	6,2	5,1	5,0	5,5	4,8	5,2	4,2	6,1	5,7
$X_2$	0,8	1,0	0,6	1,1	0,8	0,7	1,0	0,9	1,0	0,6	1,2	0,9

2. Estabeleça a equação do consumo em função do preço e renda para os dados da tabela abaixo:

Consumo	5	8	8	9	9	13	6	9	4	3
Preço	2	3	5	4	6	2	3	4	5	6
Renda	3	4	6	5	7	6	4	5	4	3

3. Estabeleça a equação de regressão múltipla para os dados da tabela abaixo:

Y	15	65	100	110	115	165
$X_1$	0	1	1	2	2	3
$X_2$	0	2	4	2	4	6

4. Estabeleça a equação de regressão múltipla para os dados da tabela abaixo:

Y	10	10	40	50	40
$X_1$	0	1	1	1	1
$X_2$	1	0	1	2	3

5. Idem para a tabela abaixo.

Y	40	50	40	120
$X_1$	0	1	2	3
$X_2$	3	2	1	0

## Capítulo 6

### Outros modelos

#### 6.1. O modelo exponencial $Y = A B^x$

A primeira observação é de que o modelo acima não é linear nos parâmetros. Considere logaritmos em ambos os membros da função acima.

$$\ln Y = \ln A + x \ln B$$

Faça – se  $y = \ln Y$ ;  $a = \ln A$  e  $b = \ln B$ , então tem-se

$$y = a + bx$$

que é uma função linear nos parâmetros cuja solução é dada por:  $S\hat{\beta} = X'Y$

Aplicação:

A tabela abaixo indica a produção anual de automóveis no Brasil em milhares de unidades.

Ano ( X )	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
Produção ( Y )	133,0	145,6	191,2	174,2	183,7	185,2	224,6	225,4	278,5	349,5	416	516	609

Admita que a produção em função do ano pode ser estudada através do modelo  $y = A B^x$ . Determine a equação correspondente ao fenômeno.

Solução:

Observe que os anos formam uma progressão aritmética de número ímpar e quando isto ocorre é permitido substituí-lo por outra série obedecendo a seguinte correspondência.

Ano ( X )	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
X	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6

Y = ln Y	4,8903	4,9809	5,2533	5,1602	5,2133	5,2214	5,4143	5,4179	5,6294	5,8565	6,0307	6,2461	6,4118
----------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

Desta forma:

$$\begin{bmatrix} N & \sum X \\ \sum X & \sum X^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{a} \\ b \end{bmatrix} = \begin{bmatrix} \sum Y \\ \sum XY \end{bmatrix}$$

$$\text{Como } N = 13 \quad \sum X = 0 \quad \sum X^2 = 182 \quad \sum Y = 71,73 \quad \sum XY = 21,68$$

$$\begin{bmatrix} 13 & 0 \\ 0 & 182 \end{bmatrix} \cdot \begin{bmatrix} \hat{a} \\ b \end{bmatrix} = \begin{bmatrix} 71,73 \\ 21,68 \end{bmatrix}$$

$$\begin{bmatrix} \hat{a} \\ b \end{bmatrix} = \frac{1}{2.366} \cdot \begin{bmatrix} 182 & 0 \\ 0 & 13 \end{bmatrix} \cdot \begin{bmatrix} 71,73 \\ 21,68 \end{bmatrix}$$

$$\begin{bmatrix} \hat{a} \\ b \end{bmatrix} = \begin{bmatrix} 5,5174 \\ 0,1181 \end{bmatrix}$$

$$\text{Mas: } \ln \hat{A} = 5,5174 \quad ; \quad \hat{A} = 248,98$$

$$\ln \hat{B} = 0,1181 \quad ; \quad \hat{B} = 1,13$$

A equação procurada é

$$\hat{Y} = \hat{A} \cdot (\hat{B})^X$$



**- Análise da Variância**

	Y	$\hat{Y}$	$(Y - \hat{Y})^2$
	133,0	119,59	179,8360
	145,6	135,14	109,4876
	191,2	152,70	1.481,9346
	174,2	172,56	2,7040
	183,7	194,99	127,4158
	185,2	220,34	1.234,5584
	224,6	248,98	594,3844
	225,4	281,35	3.130,1116
	278,5	317,92	1.554,1384
	349,5	359,25	95,1112
	416	405,96	100,8956
	516	458,73	3.279,9089
	609	518,36	8.214,8215
	3.631,90		20.105,3080

Causa de variância	G. L.	S. Q.	Q. M.	F	F( 5% )
Regressão ( $Y = AB^x$ )	1	250217,63	250217,63	136,89	4,84
Desvio da Regressão	11	20105,31	1827,75		
Total	12	270322,94	-		

Conclusão: Existe significância para a regressão.

Coeficiente de determinação:  $r^2 = 0,9256$  ou 92,56%

**6.2 Função geométrica  $Y = AX^B$**

A primeira observação é que o modelo acima não é linear nos parâmetros. Considere o logaritmo de ambos os membros da função.

$$\ln Y = \ln A + B \ln X$$

Faça-se:  $y = \ln Y$ ;  $a = \ln A$ ;  $x = \ln X$  e obtenha

$$y = a + Bx$$

que é uma função de solução conhecida,  $S\hat{\beta} = X'Y$ .

Aplicação: A tabela abaixo indica o preço em R\$ 1.000,00 do carro usado da marca Fiat Uno Sx, em função do ano de fabricação.

ANO	1998	1997	1996	1995	1994	1993	1992	1991
Preço	8,2	8,0	7,5	5,0	4,5	4,0	3,0	2,0

Admitindo que o modelo geométrico pode explicar a variação do preço em decorrência do ano de fabricação estabeleça a equação correspondente.

Solução:

Substitua a série 1998, 1997, ..., 1991 por 8, 7, 6, 5, 4, 3, 2, 1.

Cálculo de y e x.

y	2,1041	2,0794	2,0149	1,6094	1,5041	1,3863	1,0986	0,6931
x	2,0794	1,9459	1,7917	1,6094	1,3863	1,0986	0,6931	0,0000

Desta forma:

$$\begin{bmatrix} N & \sum x \\ \sum x & \sum x^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{a} \\ \hat{B} \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$

$$\begin{bmatrix} 8 & 10,6044 \\ 10,6044 & 17,5199 \end{bmatrix} \cdot \begin{bmatrix} \hat{a} \\ \hat{B} \end{bmatrix} = \begin{bmatrix} 12,4899 \\ 18,9914 \end{bmatrix}$$

$$\begin{bmatrix} \hat{a} \\ \hat{B} \end{bmatrix} = \frac{1}{27,705} \cdot \begin{bmatrix} 17,5199 & -10,6044 \\ -10,6044 & 8 \end{bmatrix} \cdot \begin{bmatrix} 12,4899 \\ 18,9914 \end{bmatrix}$$

$$\begin{bmatrix} \hat{a} \\ \hat{B} \end{bmatrix} = \begin{bmatrix} 0,6291 \\ 0,7032 \end{bmatrix}$$

Como  $a = \ln A$  ;  $A = e^a$  ;  $A = e^{0,6291} = 1,8759$

A equação procurada é  $\hat{Y} = 1,8759 X^{0,7032}$

#### - Análise da Variância

Y	$\hat{Y}$	$(Y - \hat{Y})^2$
8,2	8,10	0,0108
8,0	7,37	0,3966
7,5	6,61	0,7866
5,0	5,82	0,6681
4,5	4,97	0,2233
4,0	4,06	0,0038
3,0	3,05	0,0029
2,0	1,86	0,0154
42,2		2,1075

Causa de variância	G. L.	S. Q.	Q. M.	F	F( 5% )
Regressão ( $Y = AX^B$ )	1	37,0275	37,0275	105,40	5,99
Desvio da Regressão	6	2,1075	0,3513		
Total	7	39,1350	10		

$$C = \frac{(\sum Y)^2}{8} = \frac{42,2^2}{8} = 222,6050$$

$$SQ_{Total} = 261,7400 - C = 39,1350$$

Conclusão: Existe significância para a regressão.

$$r^2 = 0,9461 \text{ ou } 94,61\%$$

### 6.3 Curva de Gompertz: $Y = AB^{C^x}$

A primeira observação é que o modelo acima não é linear nos parâmetros. Verifique que a aplicação de logaritmos não consegue linearizar o modelo, pois

$$\ln y = \ln A + C^x \ln B.$$

A fim de superar esta dificuldade admita a possibilidade de estimar A de tal forma que  $(y / A) > 0$ .

Desta forma seja  $A = a$  e considere:

$$(Y / A) = B^{C^x}$$

$$\ln(Y / A) = C^x \ln B$$

Faça  $y = \ln (Y / A)$ ;  $b = \ln B$  então

$$Y = b C^x$$

que aplicando logaritmos resulta:

$$\ln Y = \ln b + X \ln c$$

Faça  $y^* = \ln Y$  ;  $b^* = \ln b$  e  $c^* = \ln c$  então

tem-se  $y^* = b^* + c^* X$  que é uma função linear nos parâmetros de solução conhecida.

Aplicação: Os dados abaixo foram simulados para a curva de Gompertz, admitindo que o parâmetro A é igual a 10.

Y	100,201	99,468	100,174	101,960	108,565	100,213	157,744	312,523	1726,667	123884,629	5401862922
X	-5	-4	-3	-2	-1	0	1	2	3	4	5

Determine a equação correspondente?

Solução:

$y^*$	0,8349	0,8317	0,8348	0,8424	0,8691	0,8349	1,0146	1,2361	1,6393	2,2433	3,1095
-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------

$$\begin{bmatrix} 11 & 0 \\ 0 & 110 \end{bmatrix} \cdot \begin{bmatrix} \hat{b}^* \\ \hat{c}^* \end{bmatrix} = \begin{bmatrix} \sum y^* \\ \sum Xy^* \end{bmatrix}$$

$$\begin{bmatrix} 11 & 0 \\ 0 & 110 \end{bmatrix} \cdot \begin{bmatrix} \hat{b}^* \\ \hat{c}^* \end{bmatrix} = \begin{bmatrix} 14,182 \\ 19,83 \end{bmatrix}$$

Mas:

$$\hat{b}^* = 1,2883 \quad ; \quad \hat{c}^* = 0,1802$$

$$\hat{b} = e^{1,2883} = 3,6266 \quad ; \quad \ln \hat{c} = 0,1802 \quad ; \quad \hat{c} = e^{0,1802} = 1,197$$

$$\ln \hat{B} = 3,6660 \quad ; \quad \hat{B} = e^{3,6660} = 39,0952$$

A equação procurada é

$$\hat{Y} = 10 (39,0952)^{1,2033} X$$

### - Análise da Variância

Y	Y	y*	X	X y*
100,201	2,304	0,835	-5	-4,173
99,468	2,297	0,831	-4	-3,326
100,174	2,304	0,834	-3	-2,502
101,960	2,322	0,842	-2	-1,684
108,565	2,385	0,869	-1	-0,869
100,213	2,305	0,835	0	0,000
157,744	2,758	1,014	1	1,014
312,526	3,442	1,236	2	2,472
1.726,667	5,151	1,639	3	4,917
123.884,629	2,435	2,244	4	8,976
5.401.989.922,	20,107	3,001	5	15,005
5.401.989.614		14,182		19,83

$$b^* = 1,2883$$

$$C^* = 19,83 / 110 = 0,1802$$

$$b = e^{1,2883}$$

$$C = e^{0,1802} = 1,197$$

$$B = e^{3,6266} = 37,5848 \quad Y = 10 (37,736)^x$$

#### 6.4 Curva Logística $y^* = 1 / (A + B C^x)$

O modelo acima não é linear nos parâmetros.

Consideramos o inverso da função tem-se:

$$1 / y^* = A + B C^x$$

$$(1 / Y) - A = B C^x$$

$$\text{Faça } (1 / y) - A = y$$

$y = B C^x$  que é a função exponencial já trabalhada.

$$\ln y = \ln B + X \ln C$$

$$\text{Faça } y^* = \ln y \quad ; \quad B^* = \ln B \quad ; \quad C^* = \ln C$$

$$\text{Resulta } y^* = B^* + C^* \ln C$$

Aplicação: Considere os dados da tabela da aplicação do exercício 4 e ajuste os dados a curva logística admitindo  $A = 10$ .

ANO ( X )	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972
Produção ( Y )	133,0	145,6	191,2	174,2	183,7	185,2	224,6	225,4	278,5	349,5	416	516	609

Solução:

Y	X	y	y*	X y*
133,0	-6	0,0065	-5,0359	30,2154
145,6	-5	0,0059	-5,1328	25,664
191,2	-4	0,0042	-5,4727	21,8908
174,2	-3	0,0047	-5,3602	16,0806
183,7	-2	0,0044	-5,4262	10,8524
185,2	-1	0,0044	-5,4262	5,4262
224,6	0	0,0035	-5,6550	0,0000
225,4	1	0,0034	-5,6840	-5,6840
278,5	2	0,0026	-5,9522	-11,9044
349,5	3	0,0019	-6,2660	-18,7980
416	4	0,0014	-6,5712	-26,2848
516	5	0,0009	-7,0131	-35,0655
609	6	0,0006	-7,4186	-44,5116
			-76,414	-32,1189

$$B^* = -5,878 \quad ; \quad B = e^{-5,878} = 0,0028$$

$$C^* = -\frac{32,1189}{182} = -0,1765$$

$$C = e^{-0,1765} = 0,8382$$

$$Y = \frac{1}{0,001 + (0,0028 \cdot 0,8382^x)}$$



## Capítulo 7

### SÉRIES TEMPORAIS

Este curso é dedicado ao estudo de séries temporais e sua aplicabilidade na previsão de valores futuros dessas séries. São introduzidos os conceitos de série temporal e previsão, além da classificação e descrição de alguns métodos de previsão. Também são feitas determinadas considerações a respeito desses métodos e sobre as medidas de acuidade usualmente adotadas na avaliação de tais procedimentos.

#### 7.1. Definição

A classe de fenômenos cujo processo observacional e conseqüente quantificação numérica gera uma seqüência de dados distribuídos no tempo é denominada *série temporal* [SOUZA1996].

A natureza de uma série temporal e a estrutura de seu mecanismo gerador estão relacionadas com o intervalo de ocorrência das observações no tempo. Caso o levantamento das observações da série possa ser feito a qualquer momento do tempo, a série temporal é dita *contínua*, sendo denotada por  $x(t)$  (1.1)

Entretanto, de acordo com e , na maioria das séries, as observações são tomadas em intervalos de tempo discretos e eqüidistantes.

Uma série temporal *discreta* pode ser representada por  $X^T = \{x_1, x_2, \dots, x_T\}$ , sendo que cada observação discreta  $x_t$  está associada a um instante de tempo distinto, existindo uma relação de dependência serial entre essas observações [SOUZA1989].

#### 7.2. Objetivo

O objetivo inicial da análise de séries temporais a realização de inferências sobre as propriedades ou características básicas do mecanismo gerador do processo estocástico das observações da série. Assim, através da abstração de regularidades contidas nos fenômenos observáveis de uma série temporal existe a possibilidade de se construir um modelo matemático como uma representação simplificada da realidade.

Após a formulação do modelo matemático, obtido pela seleção entre as alternativas de classes de modelos identificadas como apropriadas para essa representação e subsequente

estimação de seus parâmetros, é possível utilizá-lo para testar alguma hipótese ou teoria a respeito do mecanismo gerador do processo estocástico e realizar a previsão de valores futuros da série temporal.

### Fundamento Probabilístico

Definição. Seja  $T$  um conjunto arbitrário. Um processo estocástico é uma família  $\{Y(t), t \in T\}$  tal que,  $\forall t \in T, Y(t)$  é uma variável aleatória (v.a.)

Série temporal é um processo aleatório.

Uma Forma alternativa de definição de processo estocástico é uma família de variáveis aleatórias  $\{Y(t), t \in T\}$  é um processo estocástico se as v.a  $Y(t_1), Y(t_2), \dots, Y(t_n)$  tem função finito dimensionais,  $F(y_1, y_2, \dots, y_n; t_1, t_2, \dots, t_n) = P(Y(t_1) \leq y_1, Y(t_2) \leq y_2, \dots, Y(t_n) \leq y_n)$ , conhecidas para todo  $n \geq 1$  satisfazendo as condições de:

- simetria: para qualquer permutação  $j_1, j_2, \dots, j_n$ , dos índices  $1, 2, \dots, n$  temos:

$$F(y_{j_1}, y_{j_2}, \dots, y_{j_n}; t_{j_1}, t_{j_2}, \dots, t_{j_n}) = F(y_1, y_2, \dots, y_n; t_1, t_2, \dots, t_n)$$

exemplo ( $n=3$ )  $F(y_2, y_1, y_3; t_2, t_1, t_3) = F(y_1, y_2, y_3; t_1, t_2, t_3)$

- consistência:  $\lim_{y_n \rightarrow \infty} F(y_1, y_2, \dots, y_n; t_1, t_2, \dots, t_n) = F(y_1, y_2, \dots, y_{n-1}; t_1, t_2, \dots, t_{n-1})$

Essa definição não é muito útil na prática pois é muito difícil a especificação de todas as distribuições finito-dimensionais.

Normalmente o que se faz é concentrar nos primeiro momentos. Estes são:

i) Função média:  $\mu(t) = E\{Y(t)\}$

ii) Função de auto- covariância (facv):

$$\gamma(t_1, t_2) = E\{[Y(t_1) - \mu(t_1)][Y(t_2) - \mu(t_2)]\} = E\{Y(t_1)Y(t_2)\} - \mu(t_1)\mu(t_2)$$

Em particular se  $t_1 = t_2 = t$ ,

$\gamma(t, t) = V\{Y(t)\}$  é a variância de  $Y(t)$  denotado por  $V(t)$  ou  $\sigma^2(t)$ .

A facv fornece a forma de dependência temporal do processo  $Y(t)$  Ela não traduz a força dessa dependência pois depende da unidade de medição de  $Y$ .

Para contornar esse problema, a facv é comumente substituída pela:

iii) Função de auto correlação:  $\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sigma(t_1)\sigma(t_2)}$

A importância da função média e da facv deve-se ao fato de que se as distribuições finito dimensionais de  $Y(t)$  são normais então basta conhecer  $\mu$  e  $\gamma$  para conhecer todo o processo.

De agora em diante consideraremos apenas processos estacionários ou possíveis de estacionarização por transformações.

Após observar a série temporal discreta  $Y_1, \dots, Y_n$ ,  $\gamma_k$  será estimado por:

$$C_k = \frac{1}{n} \sum_{i=1}^{n-k} (Y_i - \bar{Y})(Y_{i+k} - \bar{Y}) \text{ onde, } \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, k=0,1,\dots,n-1$$

$C_k$  é a facv amostral

$\rho_k$  pode então ser estimado por  $r_k = \frac{C_k}{C_0}$ , função de autocorrelação amostral.

Modelos de séries temporais pode ser dividido em duas classes:

- i) Paramétricos – número finito de parâmetros, análise é feita no domínio do tempo.
- ii) Não paramétricos – número infinito de parâmetros, análise é feita no domínio da frequência.

Curso será basicamente sobre modelos paramétricos

Os modelos dessa classe podem ser escritos genericamente como:  $Y_t = S_t + \varepsilon_t$  ou seja:

Observação = sinal + ruído.

Assim temos:

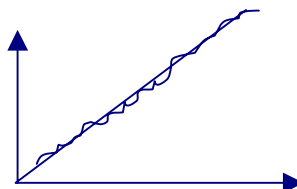
Modelos de regressão.

- Ruídos são não correlacionados
- $S_t = \underline{X}_t \underline{\beta}$

Exemplos

1. Modelos de tendência linear

$$S_t = (1 \quad t) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha + \beta t$$



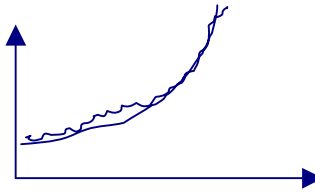
## 2. Modelo de curva de crescimento

$$Y_t = \alpha e^{\beta t + \varepsilon_t}$$

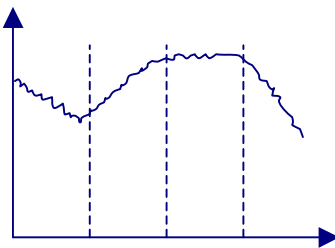
ou seja

$$\log Y_t = \log \alpha + \beta t + \varepsilon_t$$

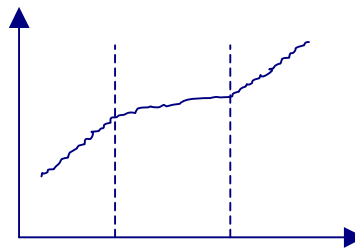
onde  $S_t = \log \alpha + \beta t$



Parâmetros sujeitos a pequenas variações temporais descreve a série da forma:



Localmente constante



Localmente Linear

## Modelos Autorregressivos (AR)

Como em (a) com

$$S_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p}$$

Nível (Sinal) atual depende dos níveis passados

Modelos Lineares estacionários

$$S_t = \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots + \psi_p \varepsilon_{t-p} = \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}$$

Este processo é estacionário se  $\sum_{i=1}^{\infty} \psi_i^2 < \infty$

Inclui os modelos ARMA, que inclui os modelos AR.

Os modelos ARIMA são uma generalização dos modelos ARMA, que visam basicamente tornar o processo estacionário através de operações de diferença.

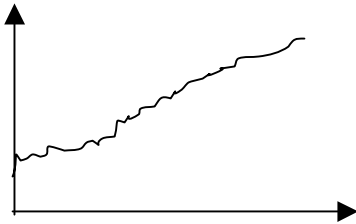
**Fenômenos típicos de Séries Temporais.**

Boa parte das séries tem características típicas

As principais são;

**i) Tendência**

É o efeito de longo prazo na média. Especificação de longo prazo é difícil.



A série acima apresenta tendência de crescimento linear .

**ii) Sazonalidade.**

Efeitos ligados a variações periódicas (semanal, mensal, anual, ...) exemplo: Medidas de temperatura: inverno, verão.

**iii) Ciclos**

Variações que apesar de periódicas não são associáveis automaticamente a nenhuma medida temporal.

Exemplos: Ciclos econômicos (5 a 7 anos), ciclos de epidemias.

Uma das tarefas mais importantes em séries temporais é identificar essas componentes visando a decomposição da série estudada.

**Modelos de regressão.**

Conforme visto nos capítulos anteriores podemos pensar a série temporal como uma coleção de observações determinadas por um sinal dependendo de uma forma determinista do tempo às quais são super impostas erros não correlacionados.

Procura-se nesse caso, usar modelos de regressão para caracterizar o sinal que controla a série.

Exemplos:

1) Modelos de tendência linear:  $S_t = \alpha + \beta t$

2) Modelos de crescimento exponencial:  $Y_t = \alpha e^{\beta t + \epsilon_t}$

3) Modelos de regressão linear simples :  $S_t = \alpha + \beta X_t$

4) Modelos de regressão não-linear simples :  $S_t = 1/(\alpha + \beta X_t)$

Os modelos 2 e 4 não são lineares nos parâmetros embora possa ser linearizado através das transformações logarítmica dando

$$\log Y_t = \log \alpha + \beta t + \varepsilon_t$$

Os modelos a serem considerados serão lineares nos parâmetros.

Exemplo: Um modelo que pode ser usado para séries temporais descrevendo uma curva S é

$\log S_t = \alpha + \beta_t$  que embora não seja linear em  $t$  é linear em  $\alpha$  e  $\beta$  após transformações logarítmicas.

Revisão de Modelos Lineares

Variável independente:  $Y$

Variável explicativa:  $X_1, X_2, \dots, X_p$

Relacionadas através de  $y_t = \beta_0 + \sum_{i=1}^p \beta_i x_{it} + \varepsilon_t \quad t = 1, \dots, n$

$$\text{Onde: } S_t = \beta_0 + \sum_{i=1}^p \beta_i x_{it},$$

podemos escrever,  $S_t = \underline{x}_t^T \underline{\beta}$ ,  $\underline{x}_t^T = (1, x_{1t}, x_{2t}, \dots, x_{pt})$ ,  $\underline{\beta}^T = (\beta_0, \beta_1, \dots, \beta_p)$

os erros  $\varepsilon_t$  são não correlacionados com  $E\{\varepsilon_t\} = 0$ ,  $V\{\varepsilon_t\} = \sigma^2$ , comumente se assume também que  $\varepsilon_t$  são normais.

Usando notação matricial pode-se escrever  $\underline{y} = \underline{x}\underline{\beta} + \underline{\varepsilon}$ , onde

$$\underline{y}^T = (y_1, y_2, \dots, y_n), \quad \underline{x}^T = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n) \text{ e } E\{\underline{\varepsilon}\} = 0, \quad V\{\underline{\varepsilon}\} = \sigma^2 I_n.$$

Critério de estimação para  $\underline{\beta}$  é a minimização de  $S(\underline{\beta}) = \sum_{t=1}^n (y_t - \underline{x}_t^T \underline{\beta})^2 = (\underline{y} - \underline{x}\underline{\beta})^T (\underline{y} - \underline{x}\underline{\beta})$

Para  $\hat{\underline{\beta}}$  minimizar  $S(\underline{\beta})$  é preciso que  $\underline{x}^T \underline{x} \hat{\underline{\beta}} = \underline{x}^T \underline{y}$  se  $\underline{x}$  tem posto máximo. (variáveis explicativas são linearmente independentes)

Então  $\underline{x}^T \underline{x}$  pode ser invertido fornecendo  $\hat{\underline{\beta}} = (\underline{x}^T \underline{x})^{-1} \underline{x}^T \underline{y}$

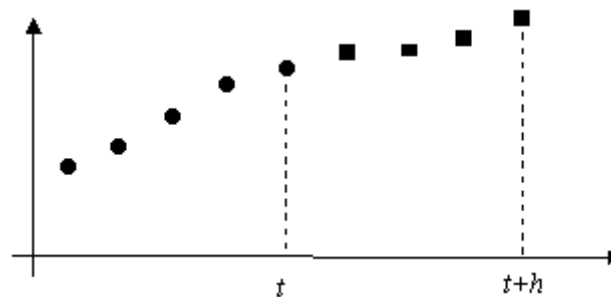
### 7.3. Previsão de Séries Temporais

Uma previsão é uma manifestação relativa a sucessos desconhecidos em um futuro determinado. A previsão não constitui um fim em si, mas um meio de fornecer informações

e subsídios para uma conseqüente tomada de decisão, visando atingir determinados objetivos.

Considerando um conjunto de observações de uma série temporal coletadas até o instante  $t$  e de um modelo que represente esses fenômenos, a previsão do valor da série no tempo  $t+h$  ( figura 1 ) pode ser obtida.

[Figura 1](#). Observações de uma série temporal com previsões de origem  $t$  e horizonte  $h$ .



Dependendo do valor assumido pelo horizonte de previsão, [SOUZA1989] classifica as previsões dos valores futuros de uma série temporal como de curto, médio ou longo prazo. Desse modo, diante da possibilidade de existência de diferentes horizontes de previsão, se especifica técnicas distintas para prognosticar os valores futuros de uma série temporal :

- *previsão múltiplos passos* : esta abordagem, adotada para longos horizontes de previsão, procura identificar as tendências gerais e os pontos de inflexão mais relevantes na série temporal. Na *previsão múltiplos passos*, o conjunto de valores correntes é empregado na realização da previsão para determinado instante; esta previsão é, então, introduzida entre as observações passadas, compondo, desta forma, um novo conjunto de dados, sobre o qual será obtida a previsão do tempo subsequente.
- *previsão simples passo* : nesta técnica não há incorporação de previsões aos dados utilizados para encontrar a próxima previsão, sendo esta, independente dos valores anteriormente previstos. A previsão é feita apenas para o período de tempo imediatamente posterior ao atual, a partir das observações da série temporal.

Para [SOUZA1989], a garantia da otimalidade das previsões de uma série temporal somente é alcançada adotando como horizonte de previsão o instante de tempo imediatamente subsequente à origem  $t$ .

Naturalmente, a investigação do poder preditivo do modelo especificado como o mais adequado para explicar o mecanismo gerador das observações de uma série temporal é um processo empírico de verificação, visto que são feitas comparações entre as previsões e as observações, a fim de confirmar a habilidade do modelo matemático em descrever a estrutura definida pelos dados da série temporal analisada. É importante salientar que o caráter previsional de um determinado modelo não pode ser sustentado quando forem identificadas mudanças estruturais nas observações obtidas a partir da origem, em relação àquelas utilizadas para a elaboração do modelo.

#### 7.4. Métodos de Previsão de Séries Temporais

Um método de previsão considera-se como um conjunto de procedimentos usados no desenvolvimento de uma determinada previsão.

Os métodos de previsão de séries temporais, classificados como métodos quantitativos, baseiam suas previsões na extrapolação de características de observações passadas e no inter-relacionamento entre essas observações, fornecendo previsões acuradas se o futuro apresentar comportamento similar ao passado .

A maioria dos métodos de previsão de séries temporais se baseia na suposição de que observações passadas contém todas as informações sobre o padrão de comportamento da série temporal e esse padrão é recorrente no tempo. O propósito dos métodos de previsão consiste em distinguir o padrão de qualquer ruído que possa estar contido nas observações e então usar esse padrão para prever os valores futuros da série temporal. Assim, pela identificação desse componente, a previsão para períodos de tempo subseqüentes ao observado pode ser desenvolvida.

Apesar de quase a totalidade dos métodos de previsão de séries temporais estar fundamentada apenas na análise das observações da série de interesse para a especificação de algum modelo que descreva essas observações, alguns procedimentos de previsão tentam explicar o comportamento de uma série temporal pela evolução dos fenômenos observacionais de outras séries. Desta forma, dependendo do número de séries temporais envolvidas na modelagem, [SOUZA1989] classifica os métodos de previsão em *univariados*, *funções de transferência* e *multivariados*.

Os *métodos univariados*, que compreendem a maior parte dos métodos de previsão de séries temporais, consideram somente uma única série para a realização dos prognósticos. As previsões decorrentes da aplicação de métodos univariados podem estar



relacionadas apenas com as informações contidas na série histórica de interesse (métodos baseados na estatística clássica) ou também, além de incorporarem essas informações, consideram outras supostamente relevantes e que não estão contidas na série analisada (métodos baseados na estatística bayesiana).

Aquelas metodologias nas quais a série de interesse é explicada não só pelo seu passado histórico, como também por outras séries temporais não-correlatas entre si, são conhecidas como *funções de transferência*. Esta classe de métodos de previsão envolve, portanto, mais de uma série temporal, com a ressalva de que a relação de causalidade entre estas séries é perfeitamente conhecida.

Os *métodos multivariados* abrangem os procedimentos de previsão que associam mais de uma série temporal na efetivação de prognósticos sem, no entanto, qualquer imposição com relação à causalidade entre essas séries.

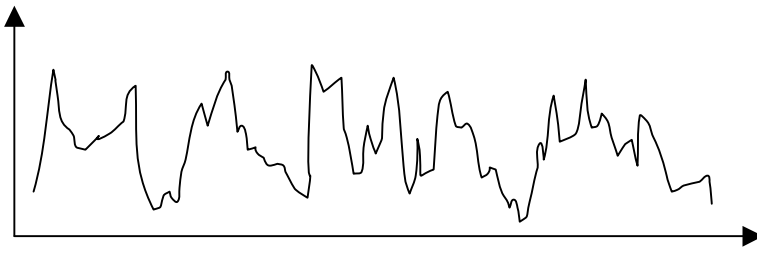
Diante disso, há uma variedade enorme de métodos de previsão de séries temporais, cada qual com suas capacidades e limitações. Qualquer que seja a classificação desses métodos, é possível utilizar um número muito grande de métodos diferentes para descrever o comportamento de uma série particular. A seleção do método de previsão adequado depende de vários fatores, tais como o comportamento do fenômeno observável ou o conhecimento a priori que se tenha sobre a sua natureza e do objetivo da análise.

[SOUZA1989] relaciona os métodos de previsão baseados exclusivamente em uma única série histórica de dados em métodos de *decomposição*, métodos *simples* de previsão e métodos *avançados* de previsão de séries temporais.

### Processos Estacionários

Como a quantidade de parâmetros é usualmente maior que o número de observações são necessárias hipóteses simplificadas.

A mais comum em séries temporais é a de estacionariedade. Basicamente isto significa que o comportamento da série não se altera com o passar do tempo, ou seja, média e variância não mudam se caminhamos no tempo.  $\Delta^n y_t = \Delta(\Delta^{n-1} y_t)$



Como normalmente as séries não apresentam esse comportamento estável, recorre-se a transformações dos dados.

As transformações mais comuns são

i) **Transformação Box Cox**

Considere a função da forma

$$g(y) = \frac{y^\lambda - 1}{\lambda}$$

se  $\lambda = 1$ ,  $g$  é a identidade

se  $\lambda = -1$ ,  $g$  é a transformação inversa

se  $\lambda = 1/2$ ,  $g$  transforma  $y$  em  $\sqrt{y}$

$$\text{e } \lim_{\lambda \rightarrow 0} g(y) = \log y \quad \lim_{\lambda \rightarrow 0} g(y) = \log y.$$

Essas transformações são usadas principalmente para estabilizar a variância

**Operação diferença ( $\Delta$ )**

Define-se o operador diferença através de  $\Delta y_t = y_t - y_{t-1}$

Aplicando-se o operador novamente obtém-se a 2ª diferença  $\Delta^2 y_t = \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$

$$\Delta^2 y_t = \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

A n-ésima diferença de  $y_t$  é obtida recursivamente por

$$\Delta^n y_t = \Delta(\Delta^{n-1} y_t)$$

Normalmente uma ou duas diferenças são suficientes para tornar a série estacionária.

- Estrita (ou Forte)
- Fraca (ou ampla ou de 2ª ordem).

Um processo estocástico  $Y(t)$  é estritamente estacionário sse suas distribuições finito dimensionais são invariantes por translações no tempo, isto é;

$$F(y_1, y_2, \dots, y_n; t_1 + \tau, t_2 + \tau, \dots, t_n + \tau) = F(y_1, y_2, \dots, y_n; t_1, t_2, \dots, t_n), \forall t_1, \dots, t_n, \tau$$

O processo deslocado  $\tau$  unidades no tempo permanece com as mesmas características.

Em particular com  $n=1$  e  $t_2 = t_1 + \tau$  temos que,

$$F(y_1; t_2) = F(y_2; t_1) \text{ e por tanto } \mu = \mu(t) \text{ e } \sigma^2 = \sigma^2(t) \text{ são constantes}$$

$$\text{Alem disso, } F(y_1, y_2; t_1 + \tau, t_2 + \tau) = F(y_1, y_2; t_1, t_2)$$

Logo

$$\text{Cov}\{Y(t_1), Y(t_2)\} = \gamma(t_1, t_2) = \gamma(t_1 - t_2) \quad \gamma(t_1, t_2) = \gamma(t_1 + \tau, t_2 + \tau), \text{ fazendo } \tau = -t_2 \text{ e } t = t_1 - t_2, \text{ temos que: } \gamma(t_1, t_2) = \gamma(t_1 - t_2, 0) = \gamma(t, 0) = \gamma(t)$$

A facv depende apenas da distancia dos pontos considerados.

Um processo estocástico  $Y(t)$  é fracamente estacionário sse:

- i)  $\mu(t) = E\{Y(t)\} = \mu$
- ii)  $V\{Y(t)\} = \sigma^2(t) = \sigma$
- iii)  $\text{Cov}\{Y(t_1), Y(t_2)\} = \gamma(t_1, t_2) = \gamma(t_1 - t_2)$

Se os momentos existem então: Estacionariedade Forte  $\longrightarrow$  Estacionariedade Fraca

A volta só vale se:

Distribuição finito dimensionais de  $Y(t)$  são normais (Processo Gaussiano).

Observe que a função de autocorrelação de processo estacionário é

$$\rho(t_1, t_2) = \frac{\sigma(t_1, t_2)}{\sigma(t_1)\sigma(t_2)} = \frac{\gamma(t_1 - t_2)}{\sigma^2(t_1)} = \frac{\gamma(t)}{\gamma(0)} = \rho(t) \quad \rho(t_1, t_2) = \frac{\sigma(t_1, t_2)}{\sigma(t_1)\sigma(t_2)} = \frac{\gamma(t_1 - t_2)}{\sigma^2(t_1)} = \frac{\gamma(t)}{\gamma(0)} = \rho(t)$$

#### 7.4.1. Métodos de Decomposição de Séries Temporais

Os métodos de decomposição assumem que uma série temporal é constituída por um conjunto de componentes não-observáveis [SOUZA1989]. Dessa forma, pela identificação das componentes individuais presentes no padrão básico da série histórica de dados (tendência, ciclo, sazonalidade e aleatoriedade), a extrapolação para o futuro pode ser realizada .

Expressa-se o relacionamento entre as componentes não-observáveis da série temporal através da equação (1.1).

$$x_t = f(S_t, T_t, C_t, E_t) \quad (1.1)$$

onde  $S_t$  corresponde à componente sazonal para o período  $t$ ;

$T_t$  é a componente de tendência no período  $t$ ;

$C_t$  é a componente de ciclo no período  $t$  e

$E_t$  é a componente aleatória no período  $t$ .

A componente sazonal representa as flutuações da série de acordo com algum fator de sazonalidade. O ciclo apresenta um comportamento similar à componente sazonal, embora tenha normalmente comprimento maior que aquela. Justamente pelo fato de não apresentar duração uniforme, a identificação da componente ciclo é mais problemática. A tendência representa o aumento ou declínio gradual nos valores das observações de uma série temporal. Com a remoção das componentes de sazonalidade, ciclo e tendência, a componente aleatória fica determinada.

Vários procedimentos para a decomposição de séries temporais foram desenvolvidos, cada qual tentando isolar as componentes não-observáveis da série o mais acuradamente possível. O objetivo desses procedimentos consiste em remover cada uma das componentes, permitindo que o comportamento da série temporal seja melhor compreendido e, conseqüentemente, prognosticar valores futuros mais apropriados.

#### 7.4.2. Métodos Simples de Previsão de Séries Temporais

Os métodos assim classificados efetuam a previsão do valor futuro da série temporal pelo alisamento das observações passadas da série de interesse. Assumindo que os valores extremos da série representam flutuações aleatórias, o propósito desses métodos consiste em identificar o padrão básico presente nos dados históricos e, então, usar esse padrão para prever valores futuros.

Associa-se a grande popularidade desses métodos à simplicidade, à eficiência computacional e à razoável previsão obtida. Entre os métodos simples de previsão destacam-se o da *Média Móvel*, o *Alisamento Exponencial Simples*, o *Alisamento Exponencial Linear* e o *Alisamento Exponencial Sazonal e Linear de Winter*, os quais são apresentados sucintamente na seqüência.

#### 7.4.2.1. Média Móvel

Esse método considera como previsão para o período futuro a média das observações passadas recentes.

A média móvel para o período de tempo  $t$  é definida por

$$x_t = \frac{x_{t-1} + x_{t-2} + \dots + x_{t-n}}{n} \quad (1.2)$$

onde  $n$  representa o número de observações incluídas na média  $x_t$ .

O termo *média móvel* é utilizado porque à medida que a próxima observação se torna disponível, a média das observações é recalculada, incluindo essa observação no conjunto de observações e desprezando a observação mais antiga.

Coloca que quanto maior o número de observações incluídas na média móvel, maior o efeito de alisamento na previsão. Assim, caso a série temporal apresente muita aleatoriedade ou pequenas mudanças nos padrões dessa série, um número maior de valores podem ser utilizados no cálculo da média móvel, obtendo-se uma previsão mais alisada. Entretanto, se houver pouca flutuação aleatória nos dados ou mudança significativa no padrão da série, um número menor de observações deve ser incluído no conjunto de valores empregado na determinação da média para que se possa reagir a essas alterações mais rapidamente.

#### 7.4.2.2. Alisamento Exponencial Simples

A princípio, o método conhecido como Alisamento Exponencial Simples se assemelha ao da Média Móvel por extrair das observações da série temporal o comportamento aleatório pelo alisamento dos dados históricos. Entretanto, a inovação introduzida pelo Alisamento Exponencial Simples advém do fato de este método atribuir pesos diferentes a cada observação da série. Enquanto que na Média Móvel as observações usadas para encontrar a previsão do valor futuro contribuem em igual proporção para o cálculo dessa previsão, no Alisamento Exponencial Simples as informações mais recentes são evidenciadas pela aplicação de um fator que determina essa importância.

Segundo Wheelwright, o argumento para o tratamento diferenciado das observações da série temporal é fundamentado na suposição de que as últimas observações contém mais informações sobre o futuro e, portanto, são mais relevantes para a previsão.

especifica o método Alisamento Exponencial Simples através da equação (1.3).

$$\hat{F}_{t+1} = \alpha x_t + (1 - \alpha) \hat{F}_t \quad (1.3)$$

onde  $\hat{F}_{t+1}$  representa a previsão no tempo  $t + 1$  e

$\alpha$  é o peso atribuído à observação  $x_t$ ,  $0 < \alpha < 1$ .

O valor assumido determina o ajuste aplicado aos dados. Quanto menor o valor da constante, mais estáveis serão as previsões, visto que a utilização de baixo valor de  $\alpha$  implica na atribuição de peso maior às observações passadas e, conseqüentemente, qualquer flutuação aleatória no presente contribui com menor importância para a obtenção da previsão. Contudo, não há metodologia que oriente quanto à seleção de um valor apropriado para  $\alpha$ , sendo normalmente encontrado por tentativa e erro. Um procedimento mais objetivo seria a seleção do valor de  $\alpha$  que forneça a "*melhor previsão das observações contidas na série temporal*".

#### 7.4.2.3. Alisamento Exponencial Linear

Quando o método Alisamento Exponencial Simples é aplicado na previsão de séries temporais que apresentam tendência entre as observações passadas, os valores prognosticados superestimam (ou subestimam) os valores reais. Desta forma, a acuidade das previsões fica prejudicada.

Para evitar esse erro sistemático, o método Alisamento Exponencial Linear foi desenvolvido procurando reconhecer a presença de tendência na série de dados. O valor da previsão obtido através deste método é alcançado pela aplicação da equação (1.4).

$$\hat{F}_{t+m} = S_t + mT_t \quad (1.4)$$

onde  $S_t$  corresponde à previsão no tempo  $t$ , conforme equação (1.5);

$T_t$  representa a componente de tendência, obtida pela equação (1.6) e

$m$  é o horizonte de previsão.

$$S_t = \alpha x_t + (1 - \alpha)(S_{t-1} + T_{t-1}) \quad (1.5)$$

Onde  $\alpha$  é o peso atribuído à observação  $x_t$ ,  $0 < \alpha < 1$ .

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \quad (1.6)$$

onde  $\beta$  é o coeficiente de alisamento, análogo a  $\alpha$ .

#### 7.4.2.4. Alisamento Exponencial Sazonal e Linear de Winter

Este método produz resultados similares ao Alisamento Exponencial Linear, sendo, no entanto, capaz de manipular séries temporais que além de apresentarem tendência nos dados, apresentam também sazonalidade.

As equações (1.7), (1.8), (1.9) e (1.10) definem o referido método de previsão.

$$S_t = \alpha \frac{x_t}{I_{t-l}} + (1 - \alpha)(S_{t-1} + T_{t-1}) \quad (1.7)$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1} \quad (1.8)$$

$$I_t = \gamma \frac{x_t}{S_t} + (1 - \gamma)I_{t-l} \quad (1.9)$$

$$F_{t+m} = (S_t + mT_t)I_{t-l+m} \quad (1.10)$$

onde  $I_t$  corresponde ao alisamento do fator de sazonalidade  $\frac{x_t}{S_t}$  ;

$l$  é o intervalo da sazonalidade e  $\gamma$  corresponde ao peso atribuído ao fator de sazonalidade.

### 7.4.3 MODELOS ARIMA

#### 7.4.3.1. Modelos Autoregressivo e de Médias Móveis

Descreve três procedimentos capazes de representar as observações de uma série temporal estacionária : modelo *Autoregressivo* (AR), modelo de *Médias Móveis* (MA) e o modelo *Autoregressivo e de Médias Móveis* (ARMA).

##### a) Modelo Autoregressivo

A especificação de um modelo Autoregressivo (AR) é dada pela equação (1.11).

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t \quad (1.11)$$

onde  $x_t$  corresponde à observação da série temporal no tempo  $t$ ;

$\phi_p$  corresponde ao parâmetro do modelo AR de ordem  $p$  e

$\varepsilon_t$  representa o erro de eventos aleatórios que não podem ser explicados pelo modelo.

Caso as observações da série temporal possam ser representadas pela equação (1.11), a ordem do modelo puder ser determinada e os parâmetros estimados, é possível prever o valor futuro da série em análise.

### b) Modelo de Médias Móveis

Um modelo de Médias Móveis (MA) fica definido conforme equação (1.12) .

$$x_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (1.12)$$

onde  $e_t$  representa o erro de eventos aleatórios que não podem ser explicados pelo modelo e  $\theta_q$  corresponde ao parâmetro do modelo MA de ordem  $q$ .

A equação (1.12) é similar à equação (1.11), exceto pelo fato de que o valor previsto para a observação  $\alpha$  depende dos valores dos erros observados em cada período passado, ao invés das observações propriamente ditas.

### c) Modelo Autoregressivo e de Médias Móveis

Wheelwright e Makridakis especificam o modelo misto Autoregressivo e de Médias Móveis (ARMA) através da equação (1.13), como sendo a combinação dos modelos AR e MA.

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (1.13)$$

Analisando a equação (1.13) é possível verificar que os modelos ARMA relacionam os valores futuros com as observações passadas, assim como também com os erros passados apurados entre os valores reais e os previstos.

#### 7.4.3.2. Modelo de Box e Jenkins

O destaque atribuído ao modelo de George Box e Gwilyn Jenkins, que também pode ser incluído nesta classificação, é devido principalmente a sua fundamentação teórica, sendo a princípio capaz de manipular séries temporais de qualquer natureza.

O método de Box e Jenkins consiste na busca de um modelo ARIMA (AutoRegressive Integrate Moving Average) que represente o processo estocástico gerador da série temporal,



a partir de um modelo ARMA aplicável na descrição de séries temporais estacionárias, estendendo esse conceito para séries temporais não-estacionárias.

Genericamente, um processo ARIMA( $p, d, q$ ) pode ser representado pela equação (1.14).

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (1.14)$$

sendo  $w_t = x_t - x_{t-d}$ ,

onde  $\phi_p$  e  $\theta_q$  são os parâmetros dos processos Autoregressivo e de Média Móvel de ordem  $p$  e  $q$  (ARMA( $p, q$ ));

$e_t$  corresponde ao erro de eventos aleatórios que não podem ser explicados pelo modelo e  $d$  equivale ao grau de homogeneidade não-estacionária.

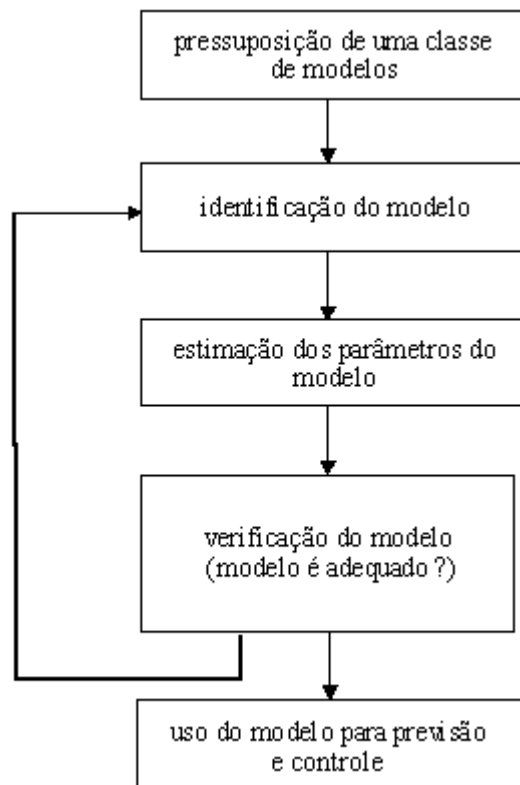
A estratégia para construção de um modelo ARIMA envolve uma abordagem iterativa que pode ser sumarizada conforme ilustra a figura 2.

O objetivo da *identificação* é determinar os valores de  $p$ ,  $d$  e  $q$  do modelo ARIMA( $p, d, q$ ). Inicialmente, a série temporal  $X^T$  é diferenciada para se obter uma série estacionária. Com isso, o processo fica reduzido a um modelo ARMA( $p, q$ ). Em seguida, a ordem do processo ARMA é identificada pela análise dos coeficientes de autocorrelação e autocorrelação parcial. Ainda nesta etapa são efetuadas estimativas preliminares dos parâmetros do modelo identificado.

Após a identificação do modelo que seja uma representação adequada do mecanismo gerador da série, a *estimação* dos parâmetros desse modelo é efetuada. Os parâmetros do processo AR são estimados através de métodos de regressão; caso o processo MA esteja envolvido, a estimação dos parâmetros deste modelo é obtida pela aplicação de algum algoritmo de otimização não-linear.

[Figura 2](#). Estágios da construção de um modelo ARIMA.

Fonte : [BOX1976]



Estimado o modelo, a *verificação* de sua habilidade em representar os fenômenos observáveis da série temporal é confirmada pela análise dos erros do modelo proposto. Caso a inadequação fique evidenciada, o ciclo de identificação, estimação e verificação é novamente aplicado, até que a representação apropriada seja encontrada.

Após a validação do modelo, a *previsão* dos valores futuros da série temporal modelada pode, enfim, ser obtida.

Análise de séries temporais – é uma técnica diferente, útil para previsão de curto prazo.

A técnica de séries temporais relaciona os valores correntes de uma variável econômica com seus valores passados, eventualmente, com valores passados de outras variáveis e valores de erros aleatórios passados e presentes.

Os modelos que relacionam os valores correntes de uma variável com seus valores passados e com erros passados e correntes são chamados de modelos de séries temporais univariados.

Se os modelos incluem várias variáveis cujos seus valores correntes são explicados pelo conjunto dessas variáveis defasadas e de erros correntes e defasados são chamados de modelos de séries temporais multivariados.

Exemplo de um processo de séries temporais:

**Processo autoregressivo de 1ª ordem AR(1):**

$$Y_t = \alpha + \theta Y_{t-1} + e_t \text{ onde } -1 < \theta < 1$$

**Processo autoregressivo de 2ª ordem AR(2):**

$$Y_t = \alpha + \theta Y_{t-1} + \theta_2 Y_{t-2} + e_t$$

**Processo de autoregressivo de ordem p AR(p):**

$$Y_t = \alpha + \theta Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} + e_t$$

Os parâmetros deste tipo de modelo podem ser estimados por mínimos quadrados.

Considere a série de desemprego no Brasil, mostrada no gráfico abaixo.



Estimou-se um modelo AR(2)

Dependent Variable: DESEMP

Method: Least Squares

Date: 06/06/00 Time: 17:34

Sample: 1990:07 2000:03

Included observations: 117

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.516913	0.259776	1.989843	0.0490
DESEMPS(-1)	0.830804	0.093123	8.921606	0.0000
DESEMPS(-2)	0.083809	0.094775	0.884295	0.3784
R-squared	0.795160	Mean dependent var		5.736581
Adjusted R-squared	0.791566	S.D. dependent var		1.229159
S.E. of regression	0.561168	Akaike info criterion		1.707713
Sum squared resid	35.89965	Schwarz criterion		1.778538
Log likelihood	-96.90120	F-statistic		221.2653
Durbin-Watson stat	1.999695	Prob(F-statistic)		0.000000

Como o coeficiente correspondente à segunda defasagem não é estatisticamente significativo, estimou-se um AR(1)

Dependent Variable: DESEMPS

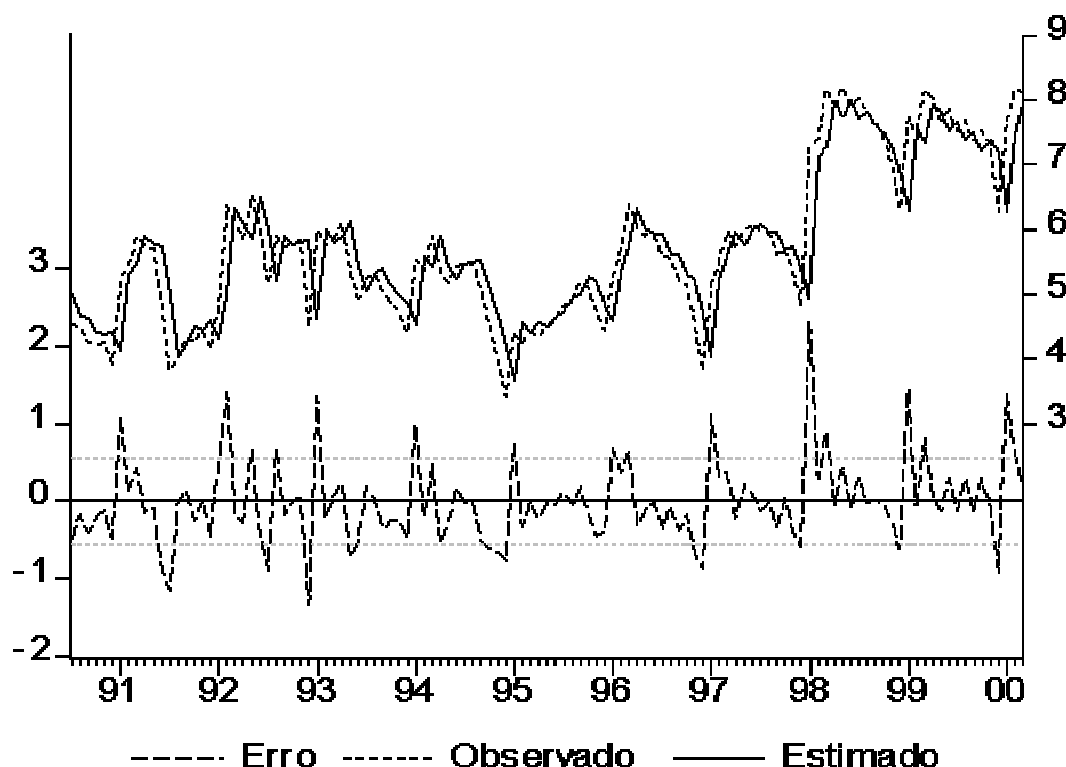
Method: Least Squares

Date: 06/06/00 Time: 17:36

Sample: 1990:07 2000:03

Included observations: 117

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.576317	0.250702	2.298811	0.0233
DESEMPS(-1)	0.903846	0.042963	21.03777	0.0000
R-squared	0.793754	Mean dependent var		5.736581
Adjusted R-squared	0.791961	S.D. dependent var		1.229159
S.E. of regression	0.560636	Akaike info criterion		1.697455
Sum squared resid	36.14590	Schwarz criterion		1.744671
Log likelihood	-97.30110	F-statistic		442.5877
Durbin-Watson stat	2.151648	Prob(F-statistic)		0.000000



Pelo  $R^2$  ajustado, verifica-se que o segundo modelo (com uma só defasagem) apresenta um ajustamento melhor. Em geral, a comparação desses modelos são realizadas utilizando os critérios de Akaike e Schwarz. Nesse caso deve-se escolher o modelo que apresente o menor valor.

No gráfico acima, são mostrados os valores observados da variável desemprego, e os valores estimados e os erros obtidos utilizando o modelo AR(1).

Uma maneira alternativa para verificar qual a ordem do processo autoregressivo é por meio da função de autocorrelação parcial da variável em questão.

$FAP(Y_t, Y_{t-2}) = \text{cor}(Y_t, Y_{t-2})$  controlando o efeito de  $Y_{t-1}$  sobre  $Y_t$ .

A ordem do modelo será definido pelo coeficiente de autocorrelação parcial de maior ordem que é estatisticamente diferente de zero. Tal verificação pode ser realizada pelo gráfico das autocorrelações chamado de correlograma, que pode ser visto na última página. Pela função de autocorrelação parcial, o modelo será de ordem um. A correlação parcial

### 7.5. Considerações sobre os Métodos de Previsão de Séries Temporais

Makridakis e Wheelwright, investigaram o poder preditivo de vários métodos comumente utilizados na previsão de séries temporais. Pela comparação dos resultados alcançados, identificaram situações onde esses métodos apresentam melhor desempenho e definiram alguns critérios para a seleção dos procedimentos de previsão, pela confrontação dos objetivos a atingir.

O estudo desenvolvido por tais pesquisadores constatou que o incremento da complexidade e da sofisticação estatística dos métodos de previsão de séries temporais não implica, necessariamente, em uma melhora na acuidade da previsão. *"Métodos simples de previsão podem apresentar desempenho extremamente satisfatório sob certas condições"*. Além disso, métodos de previsão menos complexos normalmente permitem alcançar total compreensão de suas suposições e limitações, e de interpretação de seus resultados. Assim, antes de se adotar um método de previsão mais complexo, é necessário avaliar os benefícios que um método dessa natureza pode gerar em relação ao custo de sua aplicação.

Segundo, como a acuidade de uma previsão fica determinada não apenas pelo horizonte de previsão especificado, mas também pelas características das observações da série temporal investigada, a otimização desse critério pode ser alcançada com a aplicação de mais de um método de previsão. A combinação de previsões ou a verificação da consistência desses valores permite aumentar a confiabilidade da previsão e reduzir a possibilidade de grandes desvios.

Mentzer e Cox também analisaram a frequência de aplicação de diferentes métodos de previsão de séries temporais considerando como critérios o horizonte de previsão, familiaridade com os procedimentos preditivos e nível de satisfação decorrente do emprego desses procedimentos. Em relação ao horizonte de previsão (tabela 1), o Alisamento Exponencial Simples e a Média Móvel são mais freqüentemente usados para períodos curtos, menos para médio prazo e menos ainda para longo prazo, comprovando evidências empíricas que indicam melhor desempenho destes métodos para horizonte de previsão menores. Independente do horizonte de previsão adotado, pouca aplicabilidade do método de Box e Jenkins pode ser observada.

**Tabela 1.** Frequência relativa de uso de técnicas de previsão em relação ao horizonte de previsão (em %).

Método	Horizonte de Previsão		
	Curto	Médio	Longo
Alisamento Exponencial Simples	24	17	6
Média Móvel	24	22	5
Decomposição	9	13	5
Box e Jenkins	5	6	2

Fonte : [MENTZER1984].

A tabela 2 sintetiza os estudos de Mentzer quanto à familiaridade do grupo entrevistado com diferentes métodos de previsão. Analisando os resultados dessa pesquisa, fica evidente que métodos mais simples de previsão são mais conhecidos que métodos avançados de séries temporais.

**Tabela 2.** Frequência relativa dos graus de familiaridade com os diferentes métodos de previsão (em %).

Método	Familiaridade		
	Muito familiar	Vagamente familiar	Completamente não familiar
Média Móvel	85	7	8
Alisamento Exponencial Simples	73	12	15
Decomposição	42	9	49
Box e Jenkins	26	9	65

Fonte : [MENTZER1984].

A satisfação ao se adotar algum desses métodos de previsão também foi estudada e os resultados descritos na tabela 3.

**Tabela 3.** Frequência relativa dos graus de satisfação com os diferentes métodos de previsão (em %).

Método	Satisfação		
	Satisfatório	Indiferente	Insatisfatório
Alisamento Exponencial Simples	60	19	21
Média Móvel	58	21	21
Decomposição Clássica	55	14	31
Box e Jenkins	30	13	57

Fonte : [MENTZER1984].

Analisando a tabela 3, pode ser observado que o Alisamento Exponencial Simples foi considerado o mais satisfatório entre os métodos de previsão avaliados. Esse resultado é consistente com estudos empíricos, que constataram ser esse método capaz de alcançar considerável acuidade, além de ser de fácil compreensão e utilização. O método de Médias Móveis também apresentou alto nível de satisfação, enquanto que o método de Box e Jenkins ficou caracterizado como o de menor satisfação e maior insatisfação.

## 7.6. Medidas de Acuidade dos Métodos de Previsão de Séries Temporais

De acordo com Makridakis, a suposição básica de qualquer técnica de previsão de séries temporais é que o valor observado na série fica determinado por um padrão que se repete no tempo e por alguma influência aleatória. Isto significa dizer que mesmo quando o padrão exato que caracteriza o comportamento da série temporal tenha sido isolado, algum desvio ainda existirá entre os valores da previsão e os valores realmente observados. Essa aleatoriedade não pode ser prevista; entretanto, se isolada, sua magnitude pode ser estimada e usada para determinar a variação ou erro entre as observações e previsões realizadas.



A acuidade de um método de previsão pode ser mensurada através de muitas medidas de erro, dentre as quais pode-se citar :

erro médio =

$$\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n} \quad (1.15)$$

onde  $x_i$  é o valor observado no instante  $i$ ;

$\hat{x}_i$  é o valor previsto no instante  $i$  e

$n$  corresponde ao número de previsões efetuadas.

$$\text{erro absoluto médio} = \frac{\sum_{i=1}^n |(x_i - \hat{x}_i)|}{n} \quad (1.16)$$

erro quadrado médio =

$$\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n} \quad (1.17)$$

erro percentual absoluto =

$$\left| \frac{(x_i - \hat{x}_i)}{x_i} \right| (100) \quad (1.18)$$

erro percentual médio =

$$\frac{\sum_{i=1}^n \frac{(x_i - \hat{x}_i)}{x_i}}{n} (100) \quad (1.19)$$

erro percentual absoluto médio =

$$\frac{\sum_{i=1}^n \left| \frac{(x_i - \hat{x}_i)}{x_i} \right|}{n} (100) \quad (1.20)$$

Dessa forma, a verificação da adequação de um determinado modelo supostamente representativo da série histórica de dados é dependente da medida de erro adotada para efetuar essa validação.

### Bibliografia

MAKRIDAKIS, WHEELWRIGTH, MCGEE :Forecasting. Methods and Aplications. Wiley 1983.

THOMAS H. WONNACOTT, RONALD J. WONNACOTT. “ Estatística aplicada à economia e à administração.Edit. Livros técnicos e cinetíficos. 1ª edição. Rio de Janeiro. 1981

S.SOUZA, R.C. CAMARGO, M.E. Análise e Previsão de Séries Temporais. SEDIGRAF, 1996.