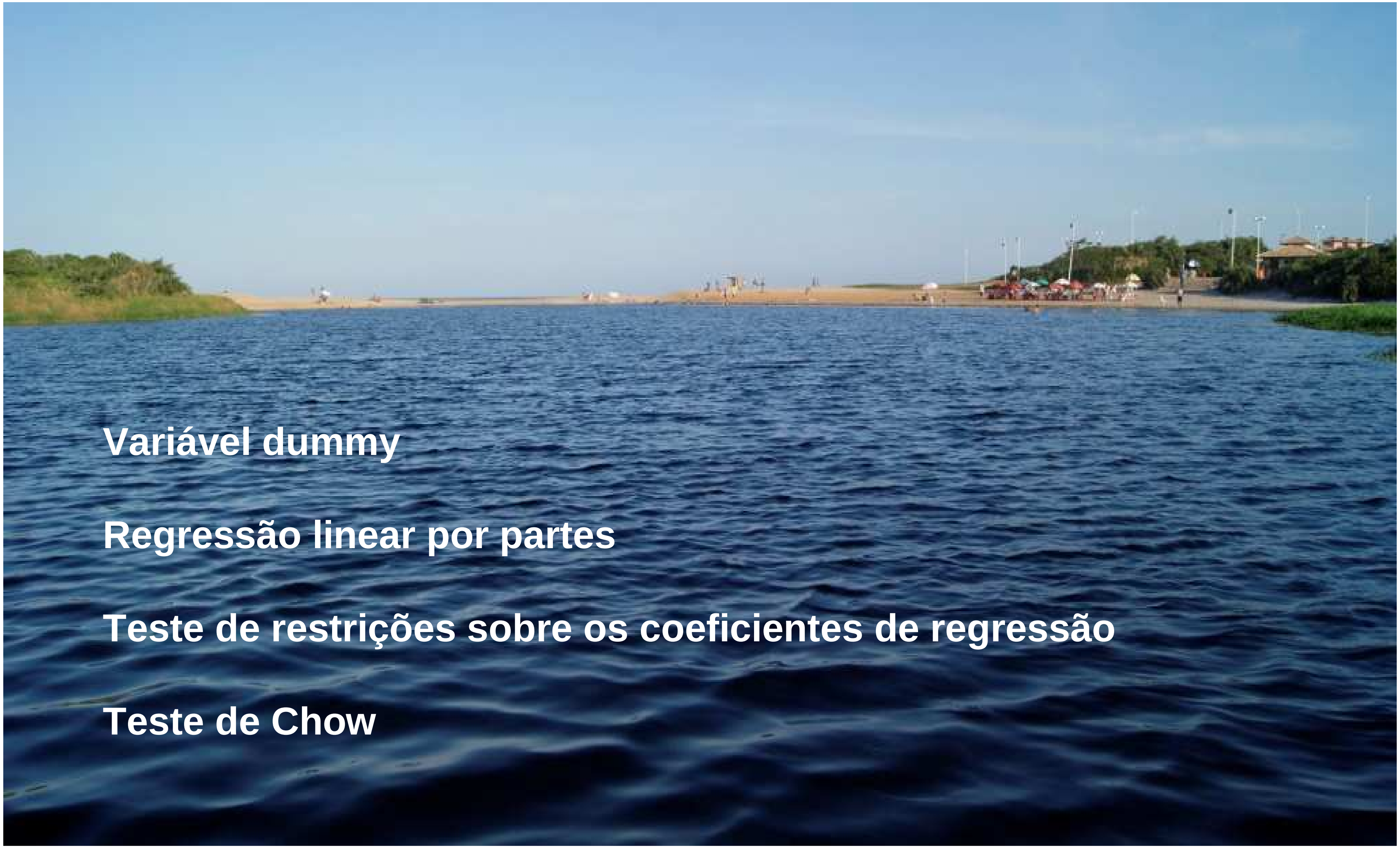
Econometria

Prof. José Francisco
professorjfmp@hotmail.com



Variável dummy

Regressão linear por partes

Teste de restrições sobre os coeficientes de regressão

Teste de Chow

Variável dummy

Variável explicativa (X) que assume apenas dois valores: 0 e 1 (variável indicadora). Indica a presença (1) ou ausência (0) de um atributo.

Finalidade: Permitir a inserção de variáveis qualitativas em um modelo de regressão, por exemplo, estado civil e sexo.

Por exemplo, considere um modelo de regressão linear em que o rendimento anual do trabalho (Y) é explicado por variáveis que caracterizam o perfil do trabalhador: escolaridade (anos de estudo), idade e sexo.

$$\text{rendimento anual } (Y_i) = \beta 0 + \beta 1 \text{ escolaridade}_i + \beta 2 \text{ idade}_i + \beta 3 \text{ sexo}_i + \varepsilon_i$$

A variável sexo é qualitativa com duas categorias: feminino e masculino.

Para inserir a variável sexo no modelo devemos criar uma variável dummy que atribui os valores 0 e 1:

$Sexo_i$ = 1 se o sexo do trabalhador *i* é masculino
$Sexo_i$ = 0 se o sexo do trabalhador *i* é feminino

A inversão na atribuição dos valores 0 e 1 aos sexos não muda as conclusões obtidas a partir do modelo.

Variável dummy

Se a variável qualitativa tem *K* categorias, então devemos incluir *K-1* variáveis dummies no modelo.

Exemplo: Uma forma usual de mensurar o nível de escolaridade consiste em pedir ao entrevistado que marque uma das opções abaixo:

- ☐ **Analfabeto**
- ☐ **Ensino básico incompleto**
- ☐ **Ensino básico completo**
- ☐ **Ensino fundamental completo**
- ☐ **Ensino fundamental incompleto**
- ☐ **Ensino Superior completo**
- ☐ **Ensino Superior incompleto**

Esta variável qualitativa tem sete categorias (*K*=7), logo *K-1*=6 variáveis dummies devem ser incluídas no lado direito do modelo de regressão, por exemplo:

Categorias	Variáveis dummies					
	Dummy(1)	Dummy(2)	Dummy(3)	Dummy(4)	Dummy(5)	Dummy(6)
Analfabeto	0	0	0	0	0	0
Ensino básico incompleto	**1**	0	0	0	0	0
Ensino básico completo	0	**1**	0	0	0	0
Ensino fundamental incompleto	0	0	**1**	0	0	0
Ensino fundamental completo	0	0	0	**1**	0	0
Ensino superior incompleto	0	0	0	0	**1**	0
Ensino superior completo	0	0	0	0	0	**1**

Para um trabalhador analfabeto todas as variáveis dummies são iguais a 0.

Para um trabalhador com nível superior completo, apenas a variável dummy 6 assume valor igual a 1

Variável dummy

Modelo de regressão linear múltipla com as variáveis dummies

$$\text{rendimento anual}_i = \beta_0 + \underbrace{\sum_{j=1}^{6} \beta_j . Dummy(j)_i}_{\text{Escolaridade}} + \beta_2 \,\text{idade}_i + \beta_3 \,\underbrace{\text{dummy}_i}_{} + \varepsilon_i$$

***Dummy* = 1 se trabalhador i é masculino**
***Dummy* = 0 se trabalhador i é feminino**

Variável dummy

Três formas de inserção das variáveis dummies em um modelo de regressão linear:

- **Forma aditiva**
- **Forma multiplicativa**
- **Forma mista**

Variável dummy – Forma aditiva

A variável dummy altera o termo constante (intercepto) do modelo de regressão linear.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \varepsilon_i$$

Y = salário do indivíduo i
X = anos de estudo do indivíduo i
D = variável dummy:
1 para indivíduo do sexo masculino
0 para indivíduo do sexo feminino

Por hipótese $E(\varepsilon_i) = 0$ **logo:**

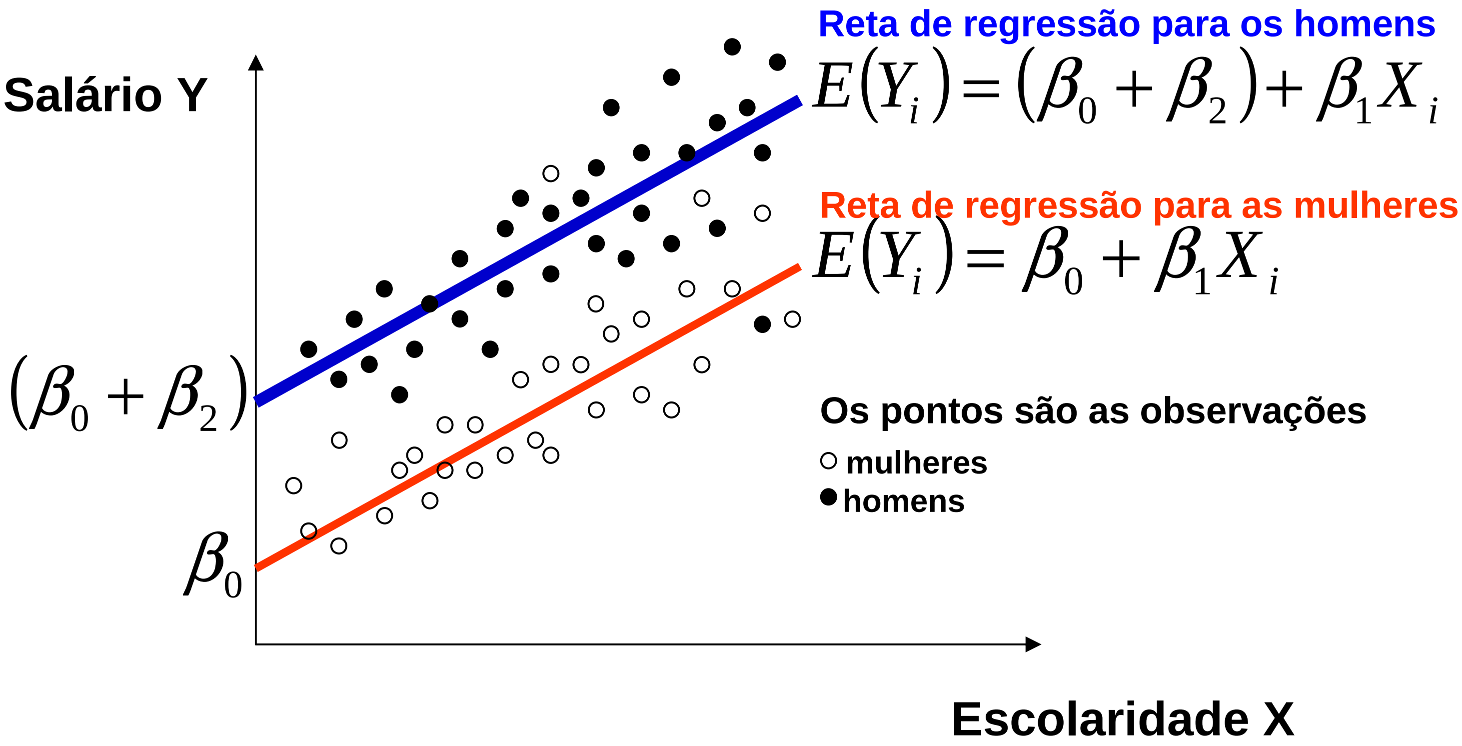
$$E(Y_i) = (\beta_0 + \beta_2) + \beta_1 X_i \Longrightarrow$$

Salário esperado, em função da escolaridade, para indivíduos do sexo masculino (D_t=1)

$$E(Y_i) = \beta_0 + \beta_1 X_i \Longrightarrow$$

Salário esperado, em função da escolaridade, para indivíduos do sexo feminino (D_t=0)

Variável dummy – Forma aditiva



β_2 é o diferencial entre os salários médios de homens e mulheres

O modelo aditivo admite que este diferencial é constante e independe do nível de escolaridade do trabalhador. Esta hipótese parece pouco plausível para este problema.

Variável dummy – Forma multiplicativa

A variável dummy altera o coeficiente de uma variável explicativa do modelo de regressão linear.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 \underbrace{D_i X_i}_{\text{Interação entre escolaridade e sexo}} + \varepsilon_i$$

Y = salário do indivíduo i
X = anos de estudo do indivíduo i
D = variável dummy:
1 para indivíduo do sexo masculino
0 para indivíduo do sexo feminino

Por hipótese $E(\varepsilon_i) = 0$ **logo:**

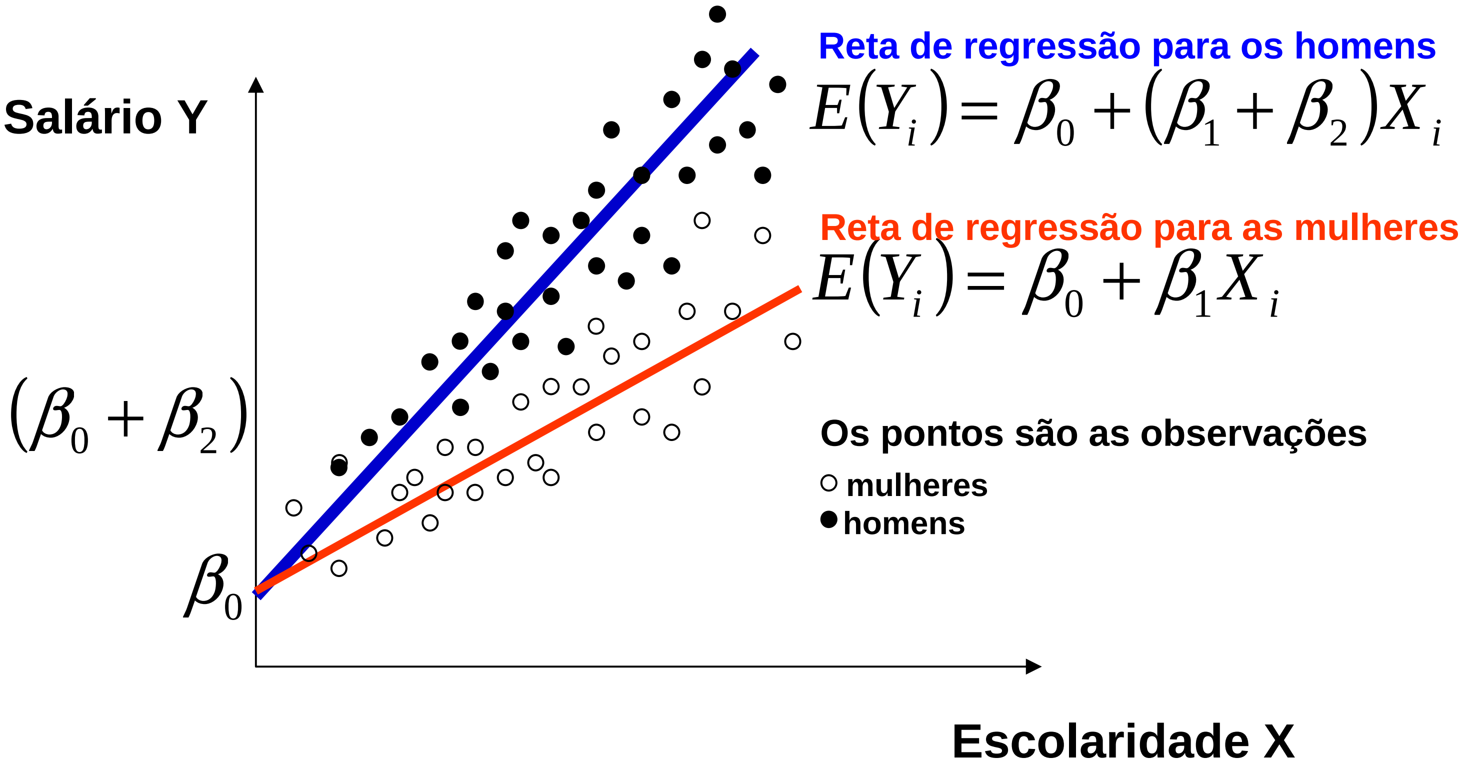
$$E(Y_i) = \beta_0 + (\beta_1 + \beta_2) X_i \Rightarrow$$

Salário esperado, em função da escolaridade, para indivíduos do sexo masculino (D_t=1)

$$E(Y_i) = \beta_0 + \beta_1 X_i \Rightarrow$$

Salário esperado, em função da escolaridade, para indivíduos do sexo feminino (D_t=0)

Variável dummy – Forma multiplicativa



Para as mulheres cada ano de estudo adicional acrescenta β_1 \$ ao salário médio.

Para os homens cada ano de estudo adicional acrescenta $\beta_1+\beta_2$ \$ ao salário médio.

O efeito de cada ano de estudo adicional sobre o salário médio depende do sexo do indivíduo (EFEITO DE INTERAÇÃO ENTRE VARIÁVEIS EXPLICATIVAS)

Variável dummy – Forma mista

A variável dummy altera o intercepto e o coeficiente de uma variável explicativa do modelo de regressão linear.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 D_i X_i + \varepsilon_i$$

Y = salário do indivíduo i
X = anos de estudo do indivíduo i
D = variável dummy:
1 para indivíduo do sexo masculino
0 para indivíduo do sexo feminino

Por hipótese $E(\varepsilon_i) = 0$ **logo:**

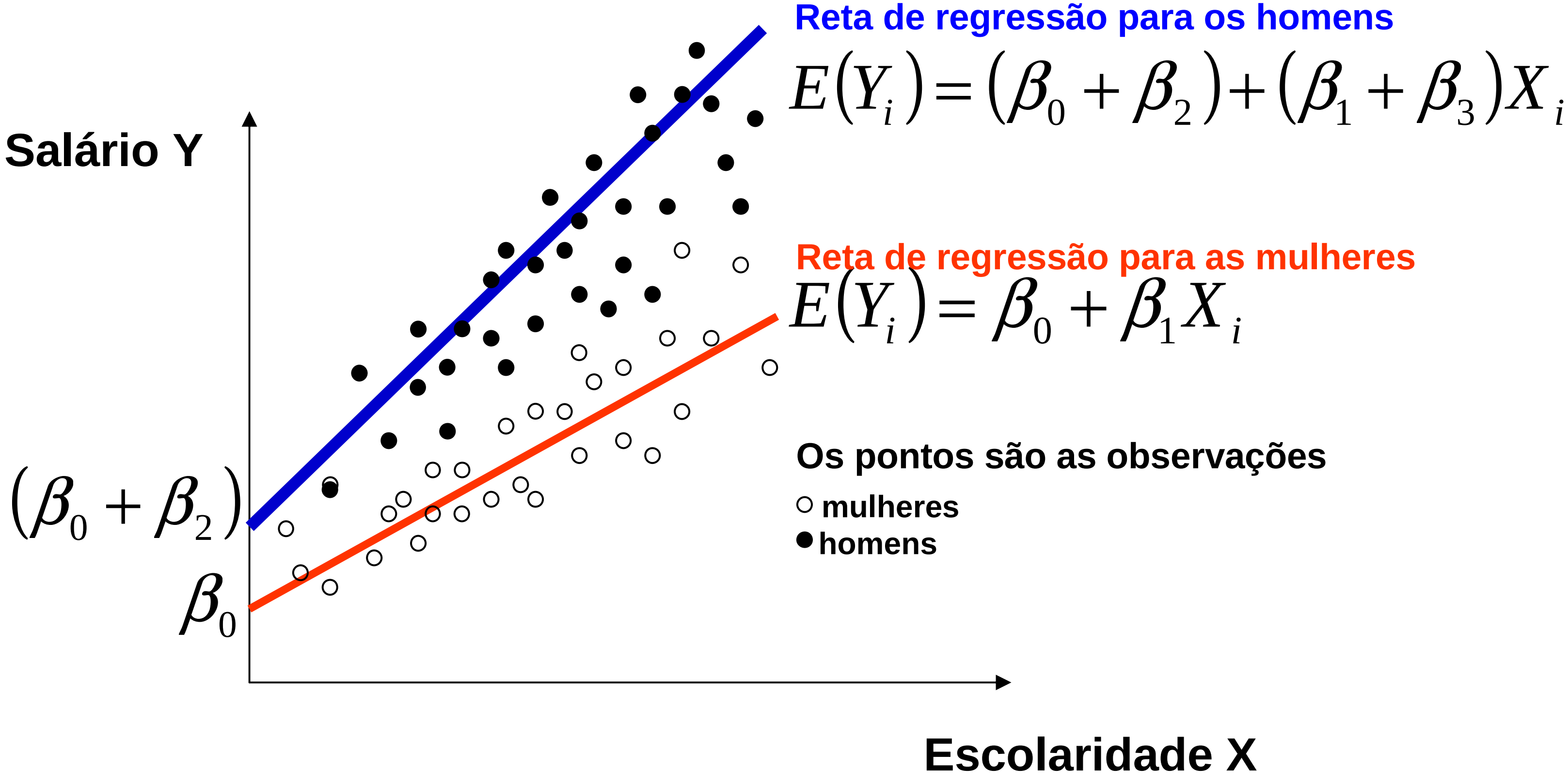
$$E(Y_i) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_i$$

Salário esperado, em função da escolaridade, para indivíduos do sexo masculino (D_t=1)

$$E(Y_i) = \beta_0 + \beta_1 X_i$$

Salário esperado, em função da escolaridade, para indivíduos do sexo feminino (D_t=0)

Variável dummy – Forma mista



Variável dummy em séries de tempo

Também permite distinguir o comportamento de um fenômeno em períodos de tempo com características diversas, por exemplo:

- **Sazonalidade: para dados mensais usamos 11 dummies, enquanto para dados trimestrais usamos 3 dummies. Por exemplo, para representar as quatro estações do ano podemos usar três variáveis dummies. Note que no verão todas as dummies valem zero (categoria de referência).**

Estações	Variáveis dummies		
	Dummy(1)	Dummy(2)	Dummy(3)
Verão	0	0	0
Outono	**1**	0	0
Inverno	0	**1**	0
Primavera	0	0	**1**

- **Períodos anterior e posterior a uma medida econômica: uma dummy que assume valor 0 para o período anterior e o valor 1 para o período posterior.**

- **Indicador de períodos com racionamento e sem racionamento de energia: uma dummy que assume valor 0 para o período anterior e o valor 1 para o período posterior.**

Exemplo 1 (Mattos, 1997)

A redução de consumo provocada pelo horário de verão tem efeito significativo no consumo anual de energia elétrica?

Para responder esta pergunta vamos estimar o seguinte modelo de regressão linear múltipla a partir de dados anuais do Sistema Elétrico Brasileiro:

$$Q_t = \beta_0 + \beta_1 T_t + \beta_2 P_t + \beta_3 D_t + u_t$$

Q_t = **demanda de energia elétrica no ano t**
T_t = **tarifa média no ano t**
P_t = **PIB no ano t**
D_t = **variável dummy:**
0 – se ano t não tem horário de verão
1 – se ano t tem horário de verão

Com base na teoria econômica esperamos que as estimativas dos coeficientes de regressão pertençam aos seguintes intervalos:

$$\beta_1 < 0 \quad \beta_2 > 0 \quad \beta_3 < 0$$

Exemplo 1 (Mattos, 1997)

As séries de demanda (Q), tarifa média (T) e PIB (P) estão expressas em índices (ano base = 1986).

O horário de verão foi introduzido em 1985.

Ano	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
Q	69	76	81	90	94	100	103	108	113	115
T	143	134	117	111	109	100	137	122	85	90
P	84	85	82	86	93	100	104	104	107	102
D	0	0	0	0	1	1	1	1	1	1

$$Y = \begin{bmatrix} 69 \\ 76 \\ 81 \\ 90 \\ 94 \\ 100 \\ 103 \\ 108 \\ 113 \\ 115 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 143 & 84 & 0 \\ 1 & 134 & 85 & 0 \\ 1 & 117 & 82 & 0 \\ 1 & 111 & 86 & 0 \\ 1 & 109 & 93 & 1 \\ 1 & 100 & 100 & 1 \\ 1 & 137 & 104 & 1 \\ 1 & 122 & 104 & 1 \\ 1 & 85 & 107 & 1 \\ 1 & 90 & 102 & 1 \end{bmatrix}$$

Estimador de mínimos quadrados

$$\boxed{\hat{\beta} = \left(X^T X\right)^{-1} X^T Y}$$

$$\hat{\beta}_0 = 5,7319$$

$$\hat{\beta}_1 = -0,2645$$

$$\hat{\beta}_2 = 1,2660$$

$$\hat{\beta}_3 = -0,5958$$

Exemplo 1 (Mattos, 1997)

Saída do Excel

P-valor > 5%
Teste t não rejeita a hipótese nula β_3=0

Intervalo de confiança contém o zero, logo não rejeitamos a hipótese β_3=0

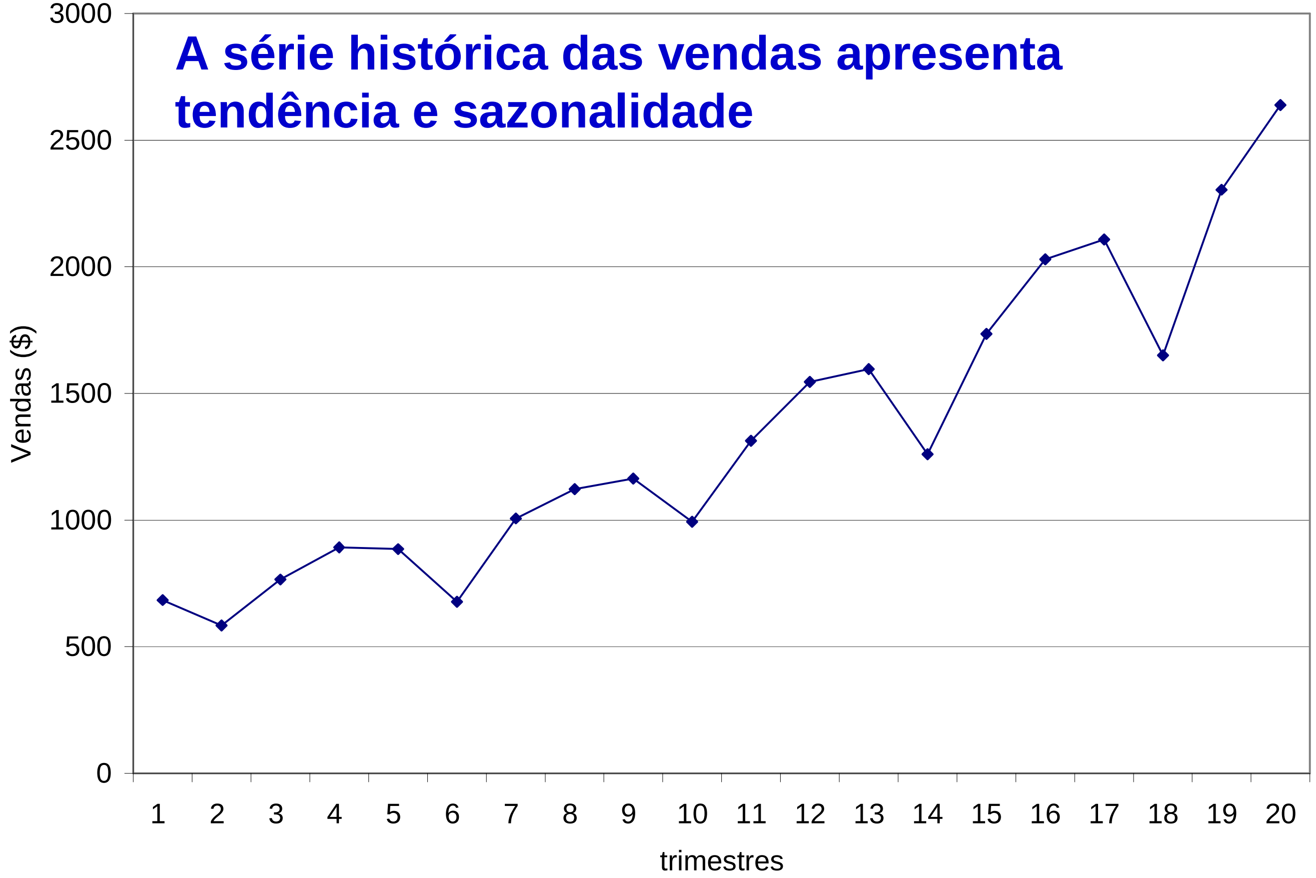
	A	B	C	D	E	F	
1	SUMMARY OUTPUT						
2							
3	*Regression Statistics*						
4	Multiple R	0,97					
5	R Square	0,93					
6	Adjusted R Square	0,90					
7	Standard Error	5,04					
8	Observations	10,00					
9							
10	ANOVA						
11		*df*	*SS*	*MS*	*F*	*Significance F*	
12	Regression	3	2088,69	696,23	27,44	6,68E-04	
13	Residual	6	152,21	25,37			
14	Total	9	2240,90				
15							
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
17	Intercept	5,73	40,49	0,14	0,89	-93,33	104,80
18	T	-0,26	0,10	-2,68	0,04	-0,51	-0,02
19	P	1,27	0,45	2,82	0,03	0,17	2,36
20	D	-0,60	8,59	-0,07	0,95	-21,63	20,43

$$\hat{Q}_t = 5{,}73 - 0{,}26T_t + 1{,}27P_t - 0{,}60D_t$$

- **Apesar da pequena amostra, os sinais dos coeficientes de regressão estão de acordo com o esperado.**
- **O coeficiente da variável dummy não é estatisticamente significativo, logo pode-se inferir que a economia de energia promovida pelo horário de verão não significativa em relação ao consumo anual.**

Exemplo 2 (Ragsdale, 2004)

Previsão de vendas trimestrais com modelo de regressão linear



Exemplo 2 (Ragsdale, 2004)

Modelo de regressão linear a ser ajustado

$$\boldsymbol{Vendas_t} = \underbrace{\beta_0 + \beta_1 \boldsymbol{t} + \beta_2 \boldsymbol{t}^2}_{\text{tendência}} + \underbrace{\beta_3 \boldsymbol{D}_{1\boldsymbol{t}} + \beta_4 \boldsymbol{D}_{2\boldsymbol{t}} + \beta_5 \boldsymbol{D}_{3\boldsymbol{t}}}_{\text{sazonalidade}} + \boldsymbol{u_t}$$

t = contador de trimestres

No histórico t vai de 1 até 20

4 trimestres, logo a sazonalidade é representada por 3 dummies

$$\boldsymbol{D}_{1\boldsymbol{t}} \begin{cases} 1 & \text{Se primeiro trimestre} \\ 0 & \text{Se não é primeiro trimestre} \end{cases}$$

$$\boldsymbol{D}_{2\boldsymbol{t}} \begin{cases} 1 & \text{Se segundo trimestre} \\ 0 & \text{Se não é segundo trimestre} \end{cases}$$

$$\boldsymbol{D}_{3\boldsymbol{t}} \begin{cases} 1 & \text{Se terceiro trimestre} \\ 0 & \text{Se não é terceiro trimestre} \end{cases}$$

Exemplo 2 (Ragsdale, 2004)

$$\boldsymbol{Vendas_t} = \beta_0 + \beta_1 \boldsymbol{t} + \beta_2 \boldsymbol{t}^2 + \beta_3 \boldsymbol{D}_{1\boldsymbol{t}} + \beta_4 \boldsymbol{D}_{2\boldsymbol{t}} + \beta_5 \boldsymbol{D}_{3\boldsymbol{t}} + \boldsymbol{u_t}$$

Vendas esperadas

$$\boldsymbol{E}\left(\boldsymbol{Vendas_t}\right) = \beta_0 + \beta_1 \boldsymbol{t} + \beta_2 \boldsymbol{t}^2 + \beta_3$$ **primeiro trimestre**

$$\boldsymbol{E}\left(\boldsymbol{Vendas_t}\right) = \beta_0 + \beta_1 \boldsymbol{t} + \beta_2 \boldsymbol{t}^2 + \beta_4$$ **segundo trimestre**

$$\boldsymbol{E}\left(\boldsymbol{Vendas_t}\right) = \beta_0 + \beta_1 \boldsymbol{t} + \beta_2 \boldsymbol{t}^2 + \beta_5$$ **terceiro trimestre**

$$\boldsymbol{E}\left(\boldsymbol{Vendas_t}\right) = \beta_0 + \beta_1 \boldsymbol{t} + \beta_2 \boldsymbol{t}^2$$ **quarto trimestre**

Exemplo 2 (Ragsdale, 2004)

	A	B	C	D	E	F	G	H
1			Time		Indicator for Qtr			Actual
2	Year	Qtr	Period	Time^2	1	2	3	Sales
3	1998	1	1	1	1	0	0	$684,2
4		2	2	4	0	1	0	$584,1
5		3	3	9	0	0	1	$765,4
6		4	4	16	0	0	0	$892,3
7	1999	1	5	25	1	0	0	$885,4
8		2	6	36	0	1	0	$677,0
9		3	7	49	0	0	1	$1.006,6
10		4	8	64	0	0	0	$1.122,1
11	2000	1	9	81	1	0	0	$1.163,4
12		2	10	100	0	1	0	$993,2
13		3	11	121	0	0	1	$1.312,5
14		4	12	144	0	0	0	$1.545,3
15	2001	1	13	169	1	0	0	$1.596,2
16		2	14	196	0	1	0	$1.260,4
17		3	15	225	0	0	1	$1.735,2
18		4	16	256	0	0	0	$2.029,7
19	2002	1	17	289	1	0	0	$2.107,8
20		2	18	324	0	1	0	$1.650,3
21		3	19	361	0	0	1	$2.304,4
22		4	20	400	0	0	0	$2.639,4
23	2003	1	21	441	1	0	0	--
24		2	22	484	0	1	0	--
25		3	23	529	0	0	1	--
26		4	24	576	0	0	0	--

Histórico

Objetivo: com base no histórico 1998 – 2002 gerar previsões trimestrais para 2003

Exemplo 2 (Ragsdale, 2004)

Valores menores que o nível de significância usual 5%, logo aceito as hipótese nulas β_1 =0 e β_3 =0

RESUMO DOS RESULTADOS

Estatística de regressão	
R múltiplo	0,992741
R-Quadrado	0,985534
R-quadrado ajustad	0,980368
Erro padrão	82,19265
Observações	20

$\leftarrow$ **R^2**

Menor que os nívei de significância usual 5%, logo rejeito a hipótese nula $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

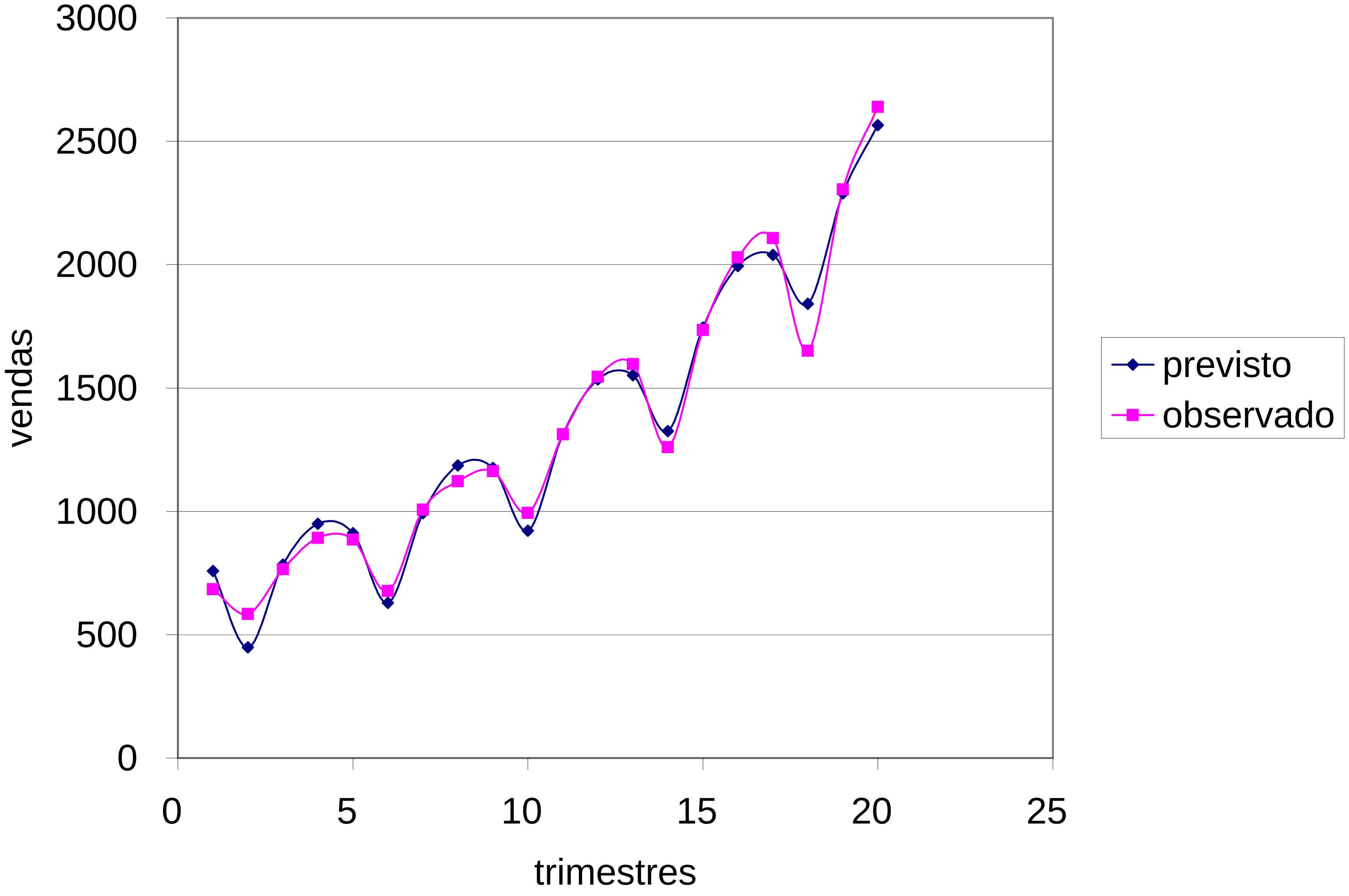
ANOVA

	gl	*SQ*	*MQ*	*F*	*F de significação*
Regressão	5	6443613,818	1288723	190,7628	2,31527E-12
Resíduo	14	94578,83513	6755,631		
Total	19	6538192,653			

	Coeficientes	*Erro padrão*	*Stat t*	*valor-P*	*95% inferiores*	*95% superiores*
Interseção	824,4727	71,38844455	11,54911	1,53E-08	671,3595927	977,5858361
Period	17,31886	13,43309658	1,289268	0,2182	-11,49229666	46,13000806
Time^2	3,485476	0,620679918	5,615577	6,37E-05	2,154248511	4,81670293
1	-86,805	52,88906781	-1,64127	0,123007	-200,2408838	26,63085515
2	-424,737	52,40244365	-8,10528	1,18E-06	-537,1289039	-312,3445769
3	-123,453	52,09941535	-2,36957	0,032719	-235,1955882	-11,71112443

$$\boldsymbol{Vendas_t} = 824{,}47 + 17{,}31\boldsymbol{t} + 3{,}46\boldsymbol{t}^2 - 86{,}81\boldsymbol{D}_{1t} - 424{,}74\boldsymbol{D}_{2t} - 123{,}45\boldsymbol{D}_{3t}$$

Exemplo 2 (Ragsdale, 2004)



Exemplo 2 (Ragsdale, 2004)

Previsão de vendas para o trimestre t

$$\boldsymbol{Vendas_t} = 824{,}47 + 17{,}31\boldsymbol{t} + 3{,}46\boldsymbol{t}^2 - 86{,}81\boldsymbol{D}_{1\boldsymbol{t}} - 424{,}74\boldsymbol{D}_{2\boldsymbol{t}} - 123{,}45\boldsymbol{D}_{3t}$$

Previsão de vendas para o primeiro trimestre de 2003 (t=21)

$$\boldsymbol{Vendas}_{21} = 824{,}47 + 17{,}31 \cdot 21 + 3{,}46 \cdot 21^2 - 86{,}81 = 2527{,}03$$

Previsão de vendas para o segundo trimestre de 2003 (t=22)

$$\boldsymbol{Vendas}_{22} = 824{,}47 + 17{,}31 \cdot 22 + 3{,}46 \cdot 22^2 - 424{,}74 = 2455{,}19$$

Previsão de vendas para o terceiro trimestre de 2003 (t=23)

$$\boldsymbol{Vendas}_{23} = 824{,}47 + 17{,}31 \cdot 23 + 3{,}46 \cdot 23^2 - 123{,}45 = 2929{,}49$$

Previsão de vendas para o quarto trimestre de 2003 (t=24)

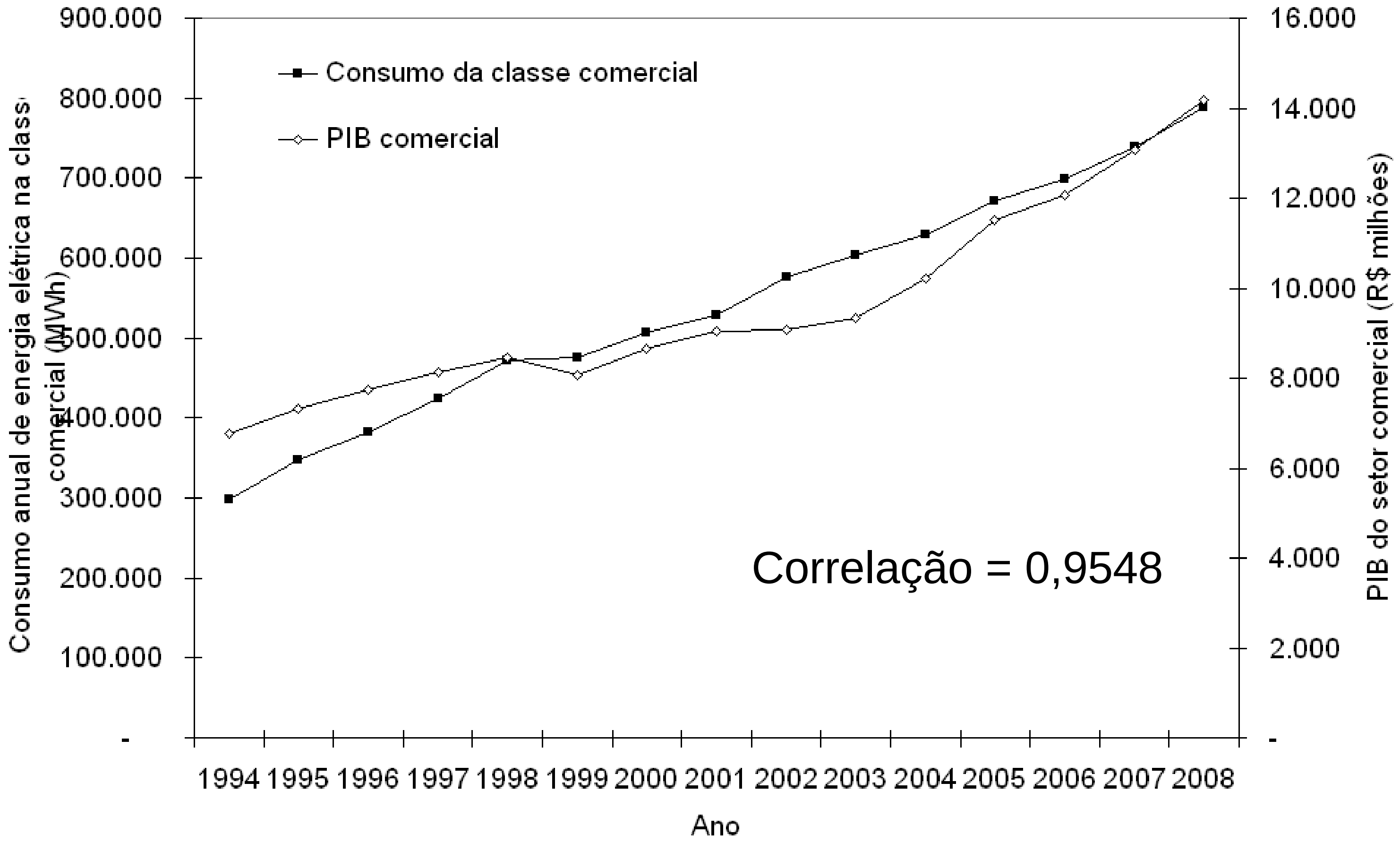
$$\boldsymbol{Vendas}_{24} = 824{,}47 + 17{,}31 \cdot 24 + 3{,}46 \cdot 24^2 = 3109{,}42$$

Regressão linear por partes
piece-wise regression

Regressão linear por partes (piecewise regression)

Uso de efeitos de interação com variáveis dummy

Exemplo: No estado do Amazonas, o consumo de energia elétrica na classe residencial (MWh) guarda uma associação com o PIB do setor comercial (R$ milhões)



Regressão linear por partes (piecewise regression)

Na modelagem da demanda por energia elétrica é bastante comum o uso da seguinte especificação:

Elasticidade

$$E_t = \beta_0 PIB_t^{\beta_1} e^{\varepsilon_t}$$

Erro

Consumo de energia elétrica da classe residencial

PIB do setor comercial

Transformação logarítmica

Equação de regressão linear

$$\ln E_t = \ln \beta_0 + \beta_1 \ln PIB_t + \varepsilon_t$$

Regressão linear por partes (piecewise regression)

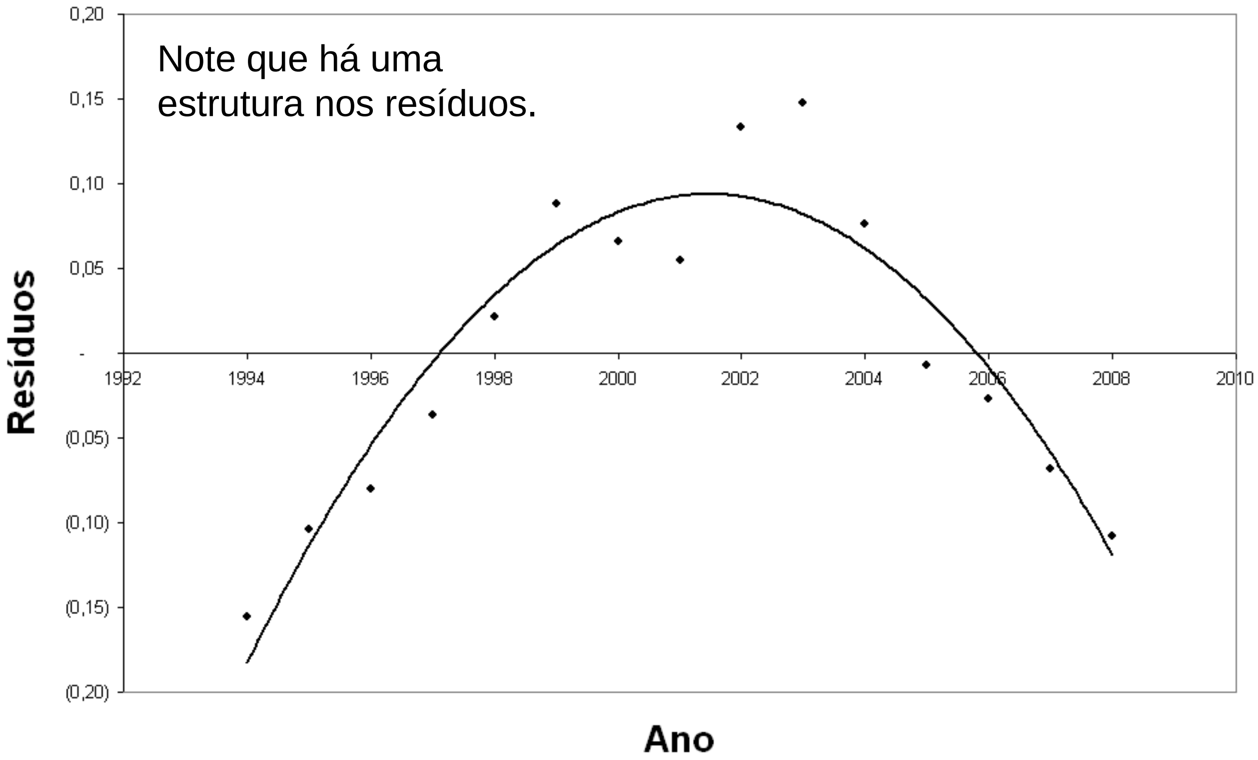
Dados

Ano	LN(E)	LN(PIB)
1994	12,61	8,82
1995	12,76	8,90
1996	12,85	8,96
1997	12,96	9,01
1998	13,06	9,04
1999	13,07	9,00
2000	13,14	9,07
2001	13,18	9,11
2002	13,26	9,11
2003	13,31	9,14
2004	13,35	9,23
2005	13,42	9,35
2006	13,46	9,40
2007	13,51	9,48
2008	13,58	9,56

Estimação por MQO

	Coeficientes	*Erro padrão*	*Stat t*	*valor-P*
Interseção	1,81	1,08	1,66	0,12
LN(PIB)	1,24	0,12	10,48	0,00

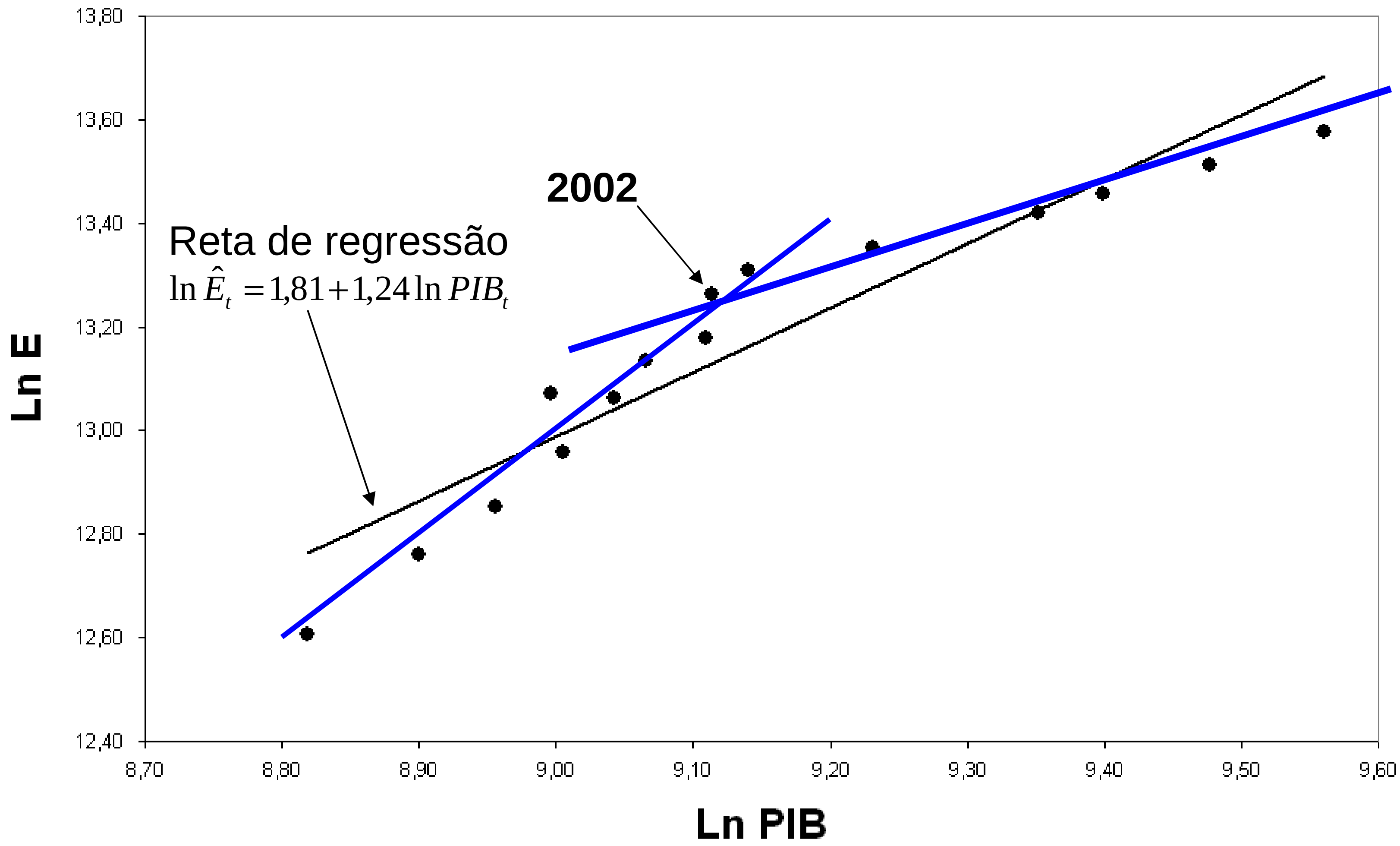
$$\ln \hat{E}_t = 1{,}81 + 1{,}24 \ln PIB_t \qquad R^2 = 0{,}89$$



Regressão linear por partes (piecewise regression)

O gráfico do Ln E contra Ln PIB sugere uma mudança de tendência.

A mudança de tendência foi provocada por uma modificação na metodologia do cálculo do PIB em 2002.



Regressão linear por partes (piecewise regression)

Por meio da regressão linear por partes pode-se ajustar um modelo que considere a mudança de tendência, mas sem descontinuidade em 2002.

Especificação com adição de um termo de interação (piecewise).

$$\ln E_t = \ln \beta_0 + \beta_1 \ln PIB_t + \beta_2 \overbrace{\left(\ln PIB_t - \ln PIB_{2002}\right) D_t}^{\text{termo de interação entre Ln PIB e uma dummy}} + \varepsilon_t$$

D_t é uma variável dummy $\begin{cases} D_t = 1 \text{ para } t \geq 2002 \\ D_t = 0 \text{ para } t \leq 2001 \end{cases}$

Observe que com uma única equação de regressão podemos ajustar equações diferentes para os períodos anterior e posterior ao ano de 2002. No ano de 2002 as duas equações fornecem o mesmo valor esperado da variável dependente.

$$\ln E_t = \ln \beta_0 + \beta_1 \ln PIB_t + \varepsilon_t \qquad \text{Para } t \leq 2001$$

$$\ln E_t = \left(\ln \beta_0 - \beta_2 \ln PIB_{2002}\right) + \left(\beta_1 + \beta_2\right) \ln PIB_t + \varepsilon_t \qquad \text{Para } t \geq 2002$$

Regressão linear por partes (piecewise regression)

Dados

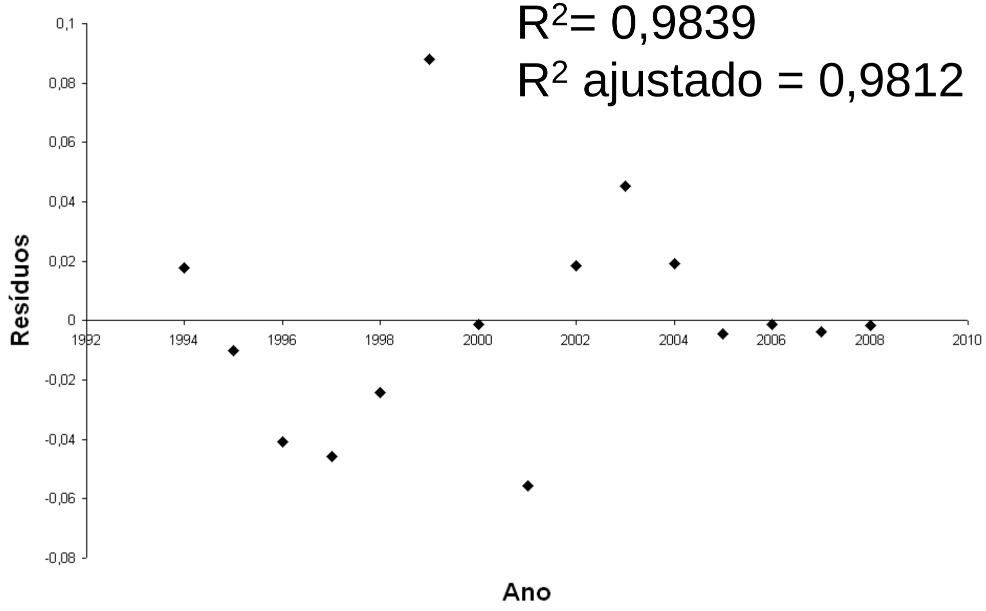
Ano	LN(E)	LN(PIB)	piece-wise
1994	12,61	8,82	-
1995	12,76	8,90	-
1996	12,85	8,96	-
1997	12,96	9,01	-
1998	13,06	9,04	-
1999	13,07	9,00	-
2000	13,14	9,07	-
2001	13,18	9,11	-
2002	13,26	9,11	-
2003	13,31	9,14	0,03
2004	13,35	9,23	0,12
2005	13,42	9,35	0,24
2006	13,46	9,40	0,29
2007	13,51	9,48	0,36
2008	13,58	9,56	0,45

Estimação por MQO

$$\ln \hat{E}_t = -6{,}98 + 2{,}22 \ln PIB_t - 1{,}47\left(\ln PIB_t - \ln PIB_{2002}\right) D_t$$

	Coeficientes	Erro padrão	Stat t	valor-P
Interseção	-6,98	1,16	-6,02	0,00
LN(PIB)	2,22	0,13	17,25	0,00
piece-wise	-1,47	0,18	-8,18	0,00

ANOVA					
	gl	SQ	MQ	F	F de significação
Regressão	2	1,12	0,56	367,08	1,73007E-11
Resíduo	12	0,02	0,00		
Total	14	1,14			

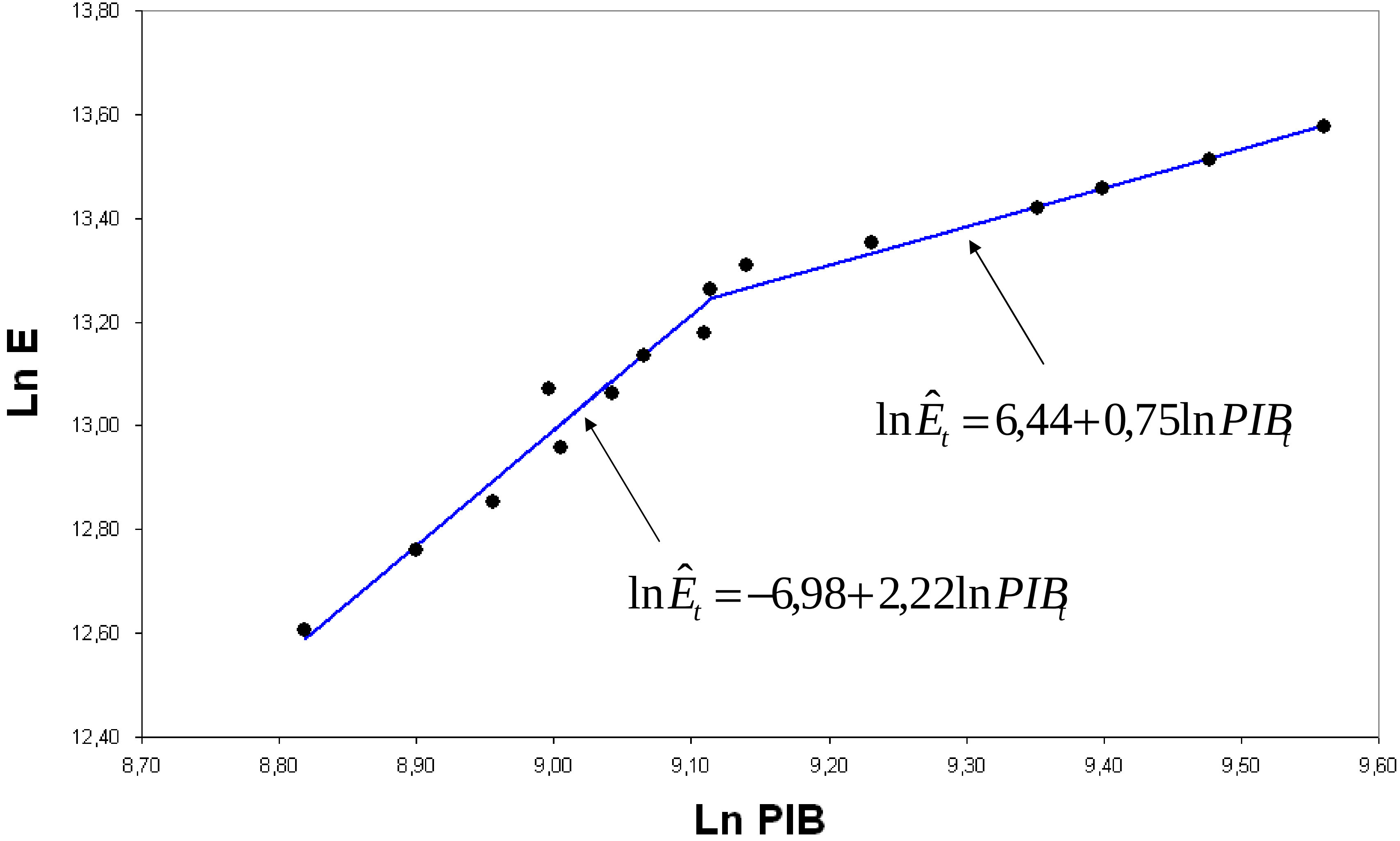


$R^2 = 0{,}9839$

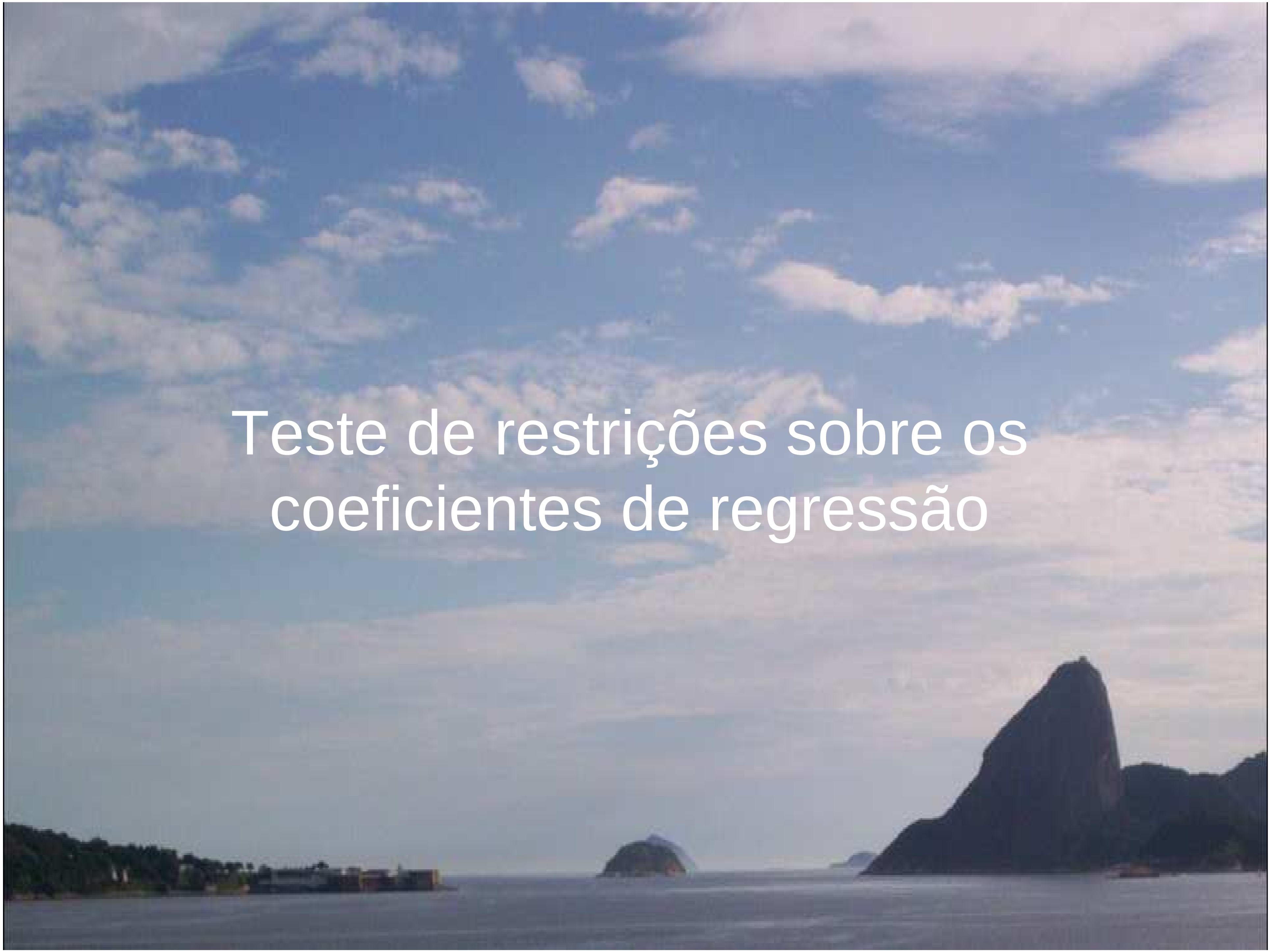
R^2 ajustado = 0,9812

Regressão linear por partes (piecewise regression)

$$\ln \hat{E}_t = -6{,}98 + 2{,}22 \ln PIB_t - 1{,}47 (\ln PIB_t - \ln PIB_{2002}) D_t$$



Teste de restrições sobre os coeficientes de regressão



Teste de hipóteses simultâneas sobre coeficientes

Considere o modelo de regressão linear múltipla

$$\mathbf{Y_i = \beta_0 + \beta_1 X_{i,1} + \ldots + \beta_m X_{i,m} + \beta_{m+1} X_{i,m+1} + \ldots + \beta_k X_{i,k} + \varepsilon_i}$$

Estamos interessados em avaliar a significância estatística de um subconjunto dos coeficientes de regressão.

Por exemplo, avaliar a significância dos m primeiros coeficientes simultâneamente.

Para este fim, vamos testar a hipótese nula de que nennhum dos m componentes sejam significativos, contra a hipótese alternativa de que pelo menos um deles seja:

$\mathbf{H_0: \beta_1 = \ldots = \beta_m = 0}$ $\Rightarrow$ $\mathbf{Y_i = \beta_0 + \beta_{m+1} X_{i,m+1} + \ldots + \beta_k X_{i,k}\ \varepsilon_i}$ **Modelo restrito**

$\mathbf{H_1: \beta_1 \neq 0 \text{ ou } \ldots \text{ ou } \beta_m \neq 0}$ $\mathbf{Y_i = \beta_0 + \beta_1 X_{i,1} + \ldots + \beta_k X_{i,k} + \varepsilon_i}$ **Modelo irrestrito**

Estatística teste sob H0

$$\frac{\dfrac{SQE_{\mathrm{Restrito}} - SQE_{Irrestrito}}{J}}{\dfrac{SQE_{Irrestrito}}{n-(k+1)}} \sim F_{J,n-(k+1)}$$

Rejeita H_0 se $F_{calculado} > F_{crítico}$

***SQE**$_{Restrito}$* **= Soma dos quadrados dos resíduos do modelo restrito**

***SQE**$_{Irrestrito}$* **= Soma dos quadrados dos resíduos do modelo irrestrito**

n = número de observações

K = número de variáveis explicativas

J = número de restrições em H0 (neste caso, *J*=*m*)

Exemplo 3 (Oliveira et al, 1997)

Exemplo:

Uma análise das importações portuguesas, para o período 1981 a 1999, forneceu o seguinte conjunto de resultados, estimados a partir de observações trimestrais:

$\ln(IMP_t) = 4{,}9 + 0{,}0491t \quad (R^2 = 0{,}9814)$

$\ln(IMP_t) = 4{,}92 + 0{,}0487t - 0{,}365D_t + 0{,}0122D_t.t \quad (R^2 = 0{,}9869)$

onde

t é um contador de trimestres.

IMP_t é o total de importações no trimestre t.

D_t é uma variável dummy que vale 1 para observações após a entrada de Portugal na CEE, em 1986, e 0 caso contrário.

Teste a hipótese que as importações após a entrada de Portugal na CEE tiveram comportamento semelhante ao período antes da adesão.

Exemplo 3 (Oliveira et al, 1997)

Exemplo:

Uma análise das importações portuguesas, para o período 1981 a 1999, forneceu o seguinte conjunto de resultados, estimados a partir de observações trimestrais:

$\ln(IMP_t) = 4{,}9 + 0{,}0491t \quad (R^2 = 0{,}9814)$

$\beta_0 \qquad \beta_1$

Variáveis associadas com a entrada de Portugal no CEE

$\ln(IMP_t) = 4{,}92 + 0{,}0487t - 0{,}365D_t + 0{,}0122D_t.t \quad (R^2 = 0{,}9869)$

$\beta_0 \qquad \beta_1 \qquad \beta_2 \qquad \beta_3$

Teste a hipótese que as importações após a entrada de Portugal na CEE tiveram comportamento semelhante ao período antes da adesão.

H_0: $\beta_2 = \beta_3 = 0$ (modelo restrito)

H_1: $\beta_2 \neq 0$ ou $\beta_3 \neq 0$ (modelo irrestrito)

Estatística teste sob H0

$$\frac{\dfrac{SQE_{\mathrm{Re}strito} - SQE_{Irrestrito}}{J}}{\dfrac{SQE_{Irrestrito}}{n-(k+1)}} \sim F_{J,n-(k+1)}$$

Exemplo 3 (Oliveira et al, 1997)

Exemplo:

$$\frac{\dfrac{SQE_{Restrito} - SQE_{Irrestrito}}{J}}{\dfrac{SQE_{Irrestrito}}{n-(k+1)}} \sim F_{J,n-(k+1)}$$

No lugar das somas dos quadrados dos resíduos (SQE) foram fornecidos os coeficientes de determinação (R^2)

variação total da variável dependente Y

$$R^2 = \frac{SQR}{SQT} = \frac{SQT - SQE}{SQT} = 1 - \frac{SQE}{SQT} \Rightarrow SQE = \underbrace{(1 - R^2)}_{\text{Parcela não explicada da variação total da variável dependente Y}} SQT$$

Substituindo SQE por (1-R^2)SQT tem-se que

$$\frac{\dfrac{R^2_{Irrestrito} - R^2_{Restrito}}{J}}{\dfrac{R^2_{Irrestrito}}{n-(k+1)}} \sim F_{J,n-(k+1)}$$

Exemplo 3 (Oliveira et al, 1997)

Exemplo:

Uma análise das importações portuguesas, para o período 1981 a 1999, forneceu o seguinte conjunto de resultados, estimados a partir de observações trimestrais:

n = 19 x 4 = 76 trimestres

$\ln(IMP_t) = 4{,}9 + 0{,}0491t$ modelo restrito ($R^2_{restrito} = 0{,}9814$)

$\ln(IMP_t) = 4{,}92 + 0{,}0487t - 0{,}365D_t + 0{,}0122D_t.t$ ($R^2_{irrestrito} = 0{,}9869$)

$$\frac{\dfrac{R^2_{Irrestrito} - R^2_{\mathrm{Re}\,strito}}{J}}{\dfrac{R^2_{Irrestrito}}{n-(k+1)}} \Rightarrow \frac{\dfrac{0{,}9869 - 0{,}9814}{2}}{\dfrac{0{,}9814}{76-(3+1)}} = 0{,}20175$$

$F_{calculado} = 0{,}20175$

$F_{crítico}$ = INVF(0.05;2;72) no Excel = 3,12

qf(0.95,2,72) no R

Conclusão: não rejeitamos H0

Teste de Chow

Teste de quebra estrutural (Gregory Chow, 1960)

Testa a igualdade dos coeficientes de duas equações de regressão.

Útil na avaliação de quebra estrutural.

Permite determinar se as variáveis independentes têm impactos distintos em diferentes subrupos da população.

Considere o modelo de regressão linear:

$$\mathbf{Y_i = \beta_0 + \beta_1 X_{i,1} + \ldots + \beta_k X_{i,k} + \varepsilon_i}$$

Se os dados são divididos em dois grupos, podemos ter duas estimativas para o mesmo modelo:

$\mathbf{Y_i = \beta_{0,1} + \beta_{1,1} X_{i,1} + \ldots + \beta_{k,1} X_{i,k} + \varepsilon_i}$ Modelo a ser estimado com n_1 obs. do grupo 1

$\mathbf{Y_i = \beta_{0,2} + \beta_{1,2} X_{i,1} + \ldots + \beta_{k,2} X_{i,k} + \varepsilon_i}$ Modelo a ser estimado com n_2 obs. do grupo 2

O teste de Chow envolve as seguintes hipóteses: H_0: $\mathbf{\beta_{0,1}=\beta_{0,2}\ ,\ \beta_{1,1}=\beta_{1,2}\ ,\ldots,\ \beta_{k,1}=\beta_{k,2}}$

H_1: $\mathbf{\beta_{0,1}=\beta_{0,2}\ ,\ \beta_{1,1}=\beta_{1,2}\ ,\ldots,\ \beta_{k,1}=\beta_{k,2}}$

Teste de quebra estrutural

H$_0$: $\beta_{0,1}=\beta_{0,2}$, $\beta_{1,1}=\beta_{1,2}$,..., $\beta_{k,1}=\beta_{k,2}$
H$_1$: $\beta_{0,1}=\beta_{0,2}$, $\beta_{1,1}=\beta_{1,2}$,..., $\beta_{k,1}=\beta_{k,2}$

Estatística teste

$$\frac{\dfrac{SQE_{restrito}-\left(SQE_{Modelo\_1}+SQE_{Modelo\_2}\right)}{k+1}}{\dfrac{SQE_{Modelo\_1}+SQE_{Modelo\_2}}{n_1+n_2-2(k+1)}}$$

$SQE_{restrito}$ é obtida na ANOVA do modelo estimado com toda a amostra

SQE_{Modelo_1} é obtida na ANOVA do modelo estimado com a amostra do grupo 1

SQE_{Modelo_2} é obtida na ANOVA do modelo estimado com a amostra do grupo 2

Exemplo 4 (OLIVEIRA et al, 1997)

Com o objetivo de avaliar o impacto ambiental de diferentes políticas de gestão de Parques Naturais, foi efetuado um estudo, cobrindo os anos de 1960 a 1987, referente ao Parque Natural em Portugal (situado numa região fronteiriça), que entre outras, produziu a seguinte estimação, pelo método dos mínimos quadrados:

1960 - 1987
S_Y^2=54.8293,18

$$\begin{array}{llll} Y_t = & -12{,}67 & +3{,}74X_{1t} & -9{,}61X_{2t} \\ & (-1{,}09) & (17{,}30) & (-2{,}16) \end{array}$$

R^2=0,984

Onde

Y_t = número de espécies animais e vegetais em perigo de extinção, no ano t

X_{2t} = número de visitantes no parque (em milhares), no ano t

X_{3t} = despesa com pessoal especializado na conservação da natureza, a preços de 1980 (em milhões de escudos), no ano t.

No início de 1972, foi aberto à circulação de automóveis em um das fronteiras do parque. As seguintes equações de regressão foram estimadas:

1960 - 1971
S_Y^2=143,33

$$\begin{array}{llll} Y_t = & 34{,}92 & +0{,}51X_{1t} & -0{,}51X_{2t} \\ & (6{,}32) & (0{,}98) & (9{,}92) \end{array}$$

R^2=0,344

1972 - 1987
S_Y^2=28.956,48

$$\begin{array}{llll} Y_t = & 31{,}11 & +3{,}42X_{1t} & -8{,}31X_{2t} \\ & (9{,}34) & (0{,}07) & (1{,}73) \end{array}$$

R^2=0,997

Exemplo 4 (OLIVEIRA et al, 1997)

Em 1972, durante a polêmica gerada, um autarca defensor da abertura da fronteira do parque afirmou que, com base na experiência anterior, o aumento do número de visitantes não iria provocar efeitos significativos ao nível das espécies existentes no parque. Poder-se-ia refutar, na altura, tal afirmação?
Para responder, precisamos tomar os resultados da regressão estimada com dados período 1960-1971.

1960 - 1971

S_Y^2=143,33

$$\begin{array}{lccc} & \beta_0 & \beta_1 & \beta_2 \\ Y_t = & 34{,}92 & +0{,}51X_{1t} & -0{,}51X_{2t} \\ & (6{,}32) & (0{,}98) & (9{,}92) \end{array}$$

R^2=0,344

H_0: β_1=0 (número de visitantes tem efeito nulo sobre conservação das espécias)
H_1: $\beta_1 \neq 0$ (número de visitantes tem efeito não nulo sobre conservação das espécias)

Estatística teste

$$\frac{\hat{\beta}_1}{\sqrt{s^2_{\hat{\beta}_1}}} \sim t_{n-3}$$

t calculado

$$\frac{0{,}51}{9{,}98} = 0{,}52$$

t crítico ao nível alfa de 5%
= INVT(0,05;12) = 2,8

t calculado < t crítico, não rejeitamos a hipótese nula.

A um nível de significância de 5% e face à evidência estatística disponível em 1972, um aumento do número de visitantes não influencia significativamente o número de espécies em extinção, a afirmação não podia ser refutada.

Exemplo 4 (OLIVEIRA et al, 1997)

O então diretor do parque contrapôs, na altura, que o passado não podia ser validamente utilizado como guia do que se passaria após a abertura da fronteira, uma vez que era de recear uma alteração do tipo de visitantes, menos preocupados com o ambiente, com repercussões que se alargariam ao impacto de outras variáveis. Verifique, se o temop veio dar ou não razão ao diretor.

A argumentação do diretor do parque baseia-se na possibilidade de alteração dos impactos ambientais, uma mudança estrutural. Então é apropriado aplicar o teste de Chow e testar a hipótese de permanência da equação de regressão antes e após a abertura da fronteira do parque.

	β_0	β_1	β_2	
Modelo I 1960 - 1971 S_Y^2=143,33	$Y_t = 34{,}92$ $(6{,}32)$	$+0{,}51X_{1t}$ $(0{,}98)$	$-0{,}51X_{2t}$ $(9{,}92)$	R^2=0,344
Modelo II 1972 - 1987 S_Y^2=28.956,48	$Y_t = 31{,}11$ $(9{,}34)$	$+3{,}42X_{1t}$ $(0{,}07)$	$-8{,}31X_{2t}$ $(1{,}73)$	R^2=0,997

H_0: $\beta_{0,I}=\beta_{0,II}$ e $\beta_{1,I}=\beta_{1,II}$ e $\beta_{2,I}=\beta_{2,II}$ (não houve mudança estrutural)

H_1: $\beta_{0,I}\neq\beta_{0,II}$ ou $\beta_{1,I}\neq \beta_{1,II}$ ou $\beta_{2,I}\neq \beta_{2,II}$ (houve mudança estrutural)

Exemplo 4 (OLIVEIRA et al, 1997)

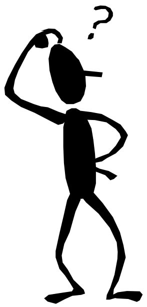
Teste de Chow

H_0: $\beta_{0,I}=\beta_{0,II}$ e $\beta_{1,I}=\beta_{1,II}$ e $\beta_{2,I}=\beta_{2,II}$ (não houve mudança estrutural)
H_1: $\beta_{0,I}\neq\beta_{0,II}$ ou $\beta_{1,I}\neq \beta_{1,II}$ ou $\beta_{2,I}\neq \beta_{2,II}$ (houve mudança estrutural)

Estatística teste

$$\frac{\dfrac{SQE_{restrito}-\left(SQE_{Modelo\_I}+SQE_{Modelo\_II}\right)}{k+1}}{\dfrac{SQE_{Modelo\_I}+SQE_{Modelo\_II}}{n_1+n_2-2(k+1)}} \sim F_{k+1,n_1+n_2-2(k+1)}$$

$k+1 = 3$
$n_1 = 12$
$n_2 = 16$



No lugar das somas dos quadrados dos resíduos (SQE)
foram fornecidos os coeficientes de determinação (R^2) e S_Y^2

Exemplo 4 (OLIVEIRA et al, 1997)

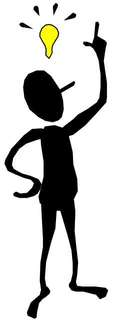
Teste de Chow

H_0: $\beta_{0,I}=\beta_{0,II}$ e $\beta_{1,I}=\beta_{1,II}$ e $\beta_{2,I}=\beta_{2,II}$ (não houve mudança estrutural)
H_1: $\beta_{0,I}\neq\beta_{0,II}$ ou $\beta_{1,I}\neq \beta_{1,II}$ ou $\beta_{2,I}\neq \beta_{2,II}$ (houve mudança estrutural)

Estatística teste

$$\frac{\dfrac{SQE_{restrito}-\left(SQE_{Modelo\_I}+SQE_{Modelo\_II}\right)}{3}}{\dfrac{SQE_{Modelo\_I}+SQE_{Modelo\_II}}{22}}$$

$$R^2=\frac{SQR}{SQT}=\frac{SQT-SQE}{SQT}=1-\frac{SQE}{SQT}\Rightarrow SQE=\left(1-R^2\right)SQT$$



$$SQT=\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2\Rightarrow SQT=(n-1)S_Y^2$$

$$SQE=\left(1-R^2\right)(n-1)S_Y^2$$

Exemplo 4 (OLIVEIRA et al, 1997)

Teste de Chow

H_0: $\beta_{0,I}=\beta_{0,II}$ e $\beta_{1,I}=\beta_{1,II}$ e $\beta_{2,I}=\beta_{2,II}$ (não houve mudança estrutural)
H_1: $\beta_{0,I}\neq\beta_{0,II}$ ou $\beta_{1,I}\neq \beta_{1,II}$ ou $\beta_{2,I}\neq \beta_{2,II}$ (houve mudança estrutural)

$$SQE_{restrito} = (1-0{,}984)(28-1)54.829{,}175 = 23686{,}2$$

$$SQE_{Modelo\_I} = (1-0{,}344)(12-1)143{,}33 = 1.034{,}27$$

$$SQE_{Modelo\_II} = (1-0{,}997)(16-1)28.956{,}48 = 1.303{,}04$$

F calculado

$$\frac{\dfrac{23.686{,}2-(1.034{,}27+1.303{,}04)}{3}}{\dfrac{1.034{,}27+1.303{,}04}{22}} = 66{,}98$$

F calculado = 66,98

F crítico = INVF(0,05;3;22) = 3,04

F calculado > F crítico, logo rejeitamos H_0

A evidência estatística disponível posteriormente (até 1987) veio dar razão aos argumentos apresentados em 1972, pelo diretor do parque

Referências Bibliográficas

Matos, O.C. Econometria básica: teoria e aplicações. Editora Atlas, São Paulo, 1997.

Oliveira, M.M., Aguiar, A., Carvalho, A., Mendes, V., Portugal, P. Econometria: Exercícios, Editora McGraw Hill de Portugal, Alfragide, 1997.

Ragsdale, C.T. Spreadsheet Modeling & Decision Analysis: A practical introduction to management science, 4e, Thompson, 2004.