

Painel: Análise e soluções de problemas típicos em modelos de regressão linear

Msc Arq. e Urb. Ana Maria de Biazzi
Dias de Oliveira

anabiazzi@uol.com.br

(055) 11 98109-7733



11 A 15

SETEMBRO | 2023

SÃO PAULO - SP

REALIZAÇÃO



IBAPE NACIONAL
Instituto Brasileiro de Avaliações
e Perícias de Engenharia

Variáveis Dicotômicas (*binárias/dummy*): especificações de modelos com parâmetros variáveis

Nos modelos de regressão de regressão para estimativa do valor do bem, é comum a influência das variáveis quantitativas e qualitativas.

Uma das formas de "quantificar" os atributos qualitativos é construir variáveis artificiais que assumam valores de 0 ou 1 (indicando ausência ou presença de um atributo) conhecidas pela literatura de "variável dummy".

modelos
de preços
hedônicos

Especificamente para expressar comportamento de preços do mercado imobiliário é comum envolvendo as variáveis do **entorno/ localização** (próximos ou distantes de um polo de atração); com vista ou sem; **e do imóvel em si** (se há piscina ou não, etc...)

Estas variáveis, trazem em uma mesma equação múltipla **diferenças** de atributos agrupados – possibilita projeções nas duas situações distintas, lembrando que **as demais variáveis do modelo permanecem constantes**

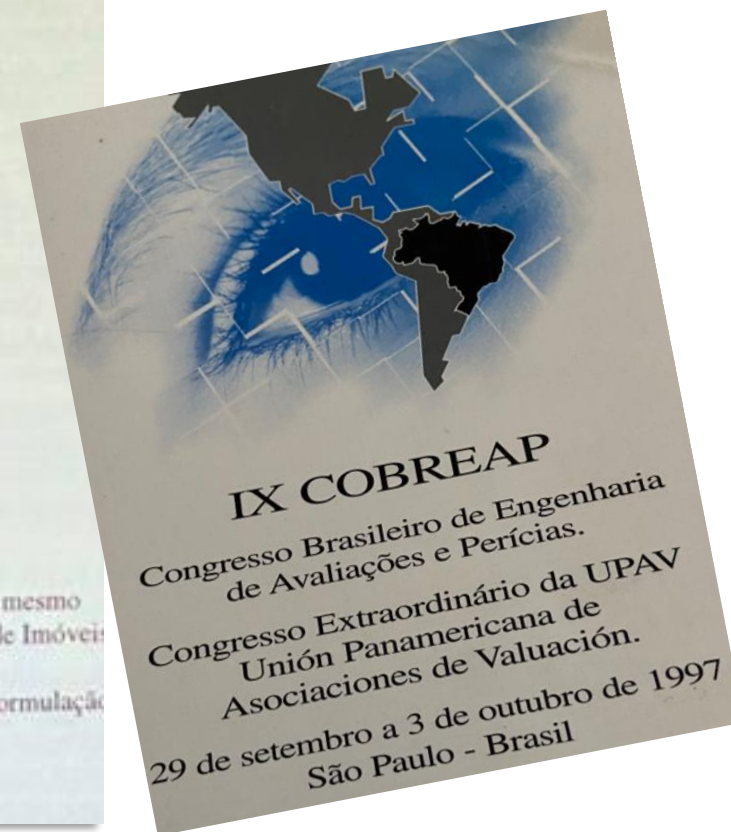
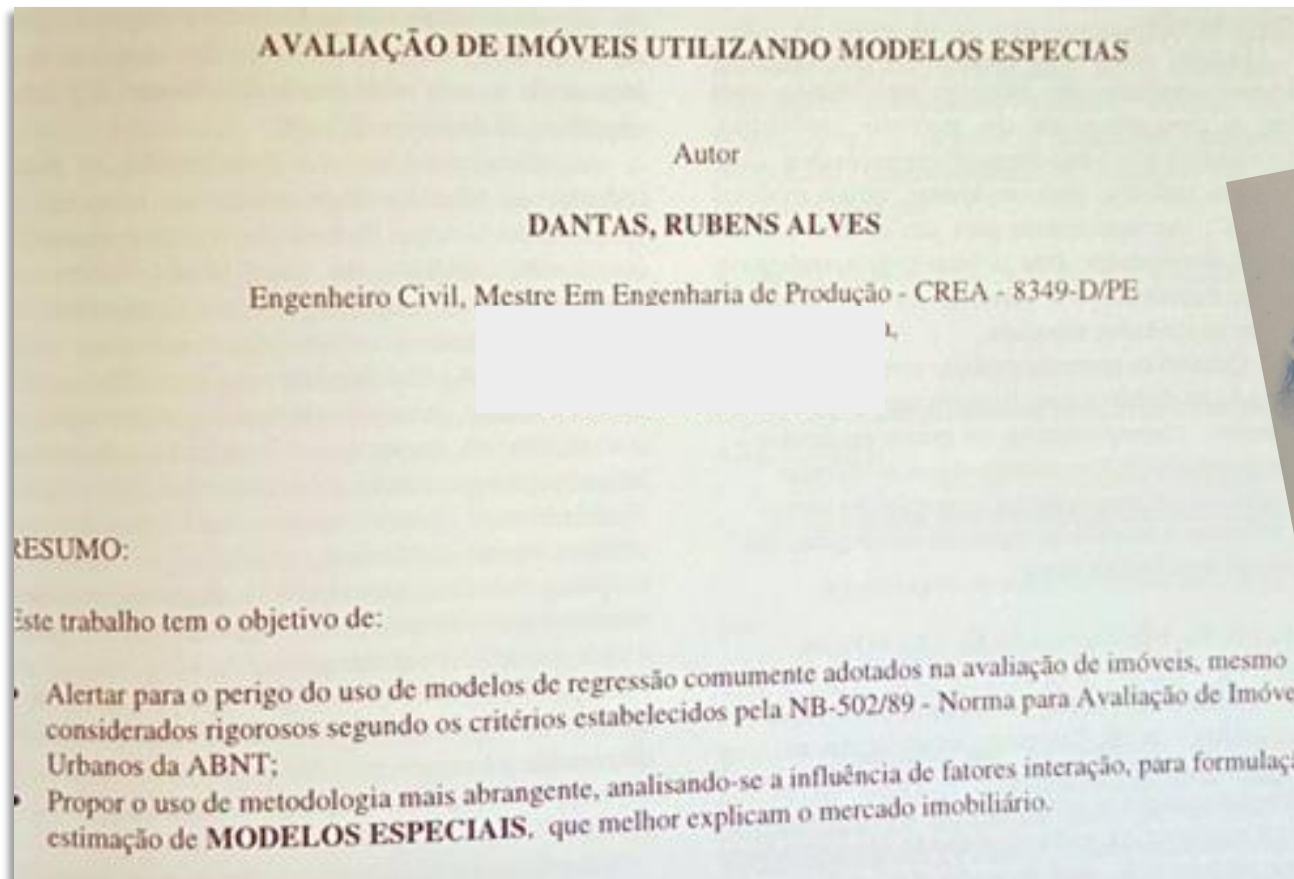
?

Inspiração



11^A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



Inspiração

Variáveis Dicotômicas (*binárias/dummy*): especificações de modelos com parâmetros variáveis

ABNT NBR	14653-2	Anexo A.8 Diferentes agrupamentos
		No caso de utilização no mesmo modelo de regressão de diferentes agrupamentos (tipologia, mercados, localização, usos etc.), recomenda-se verificar a independência entre os agrupamentos, entre as variáveis utilizadas e possíveis interações entre elas



Na especificação de modelos que combinam *dummies* e variáveis quantitativas importante verificar necessidade de incorporar possíveis mudanças no intercepto e/ou na declividade de uma função

Duas Categorias

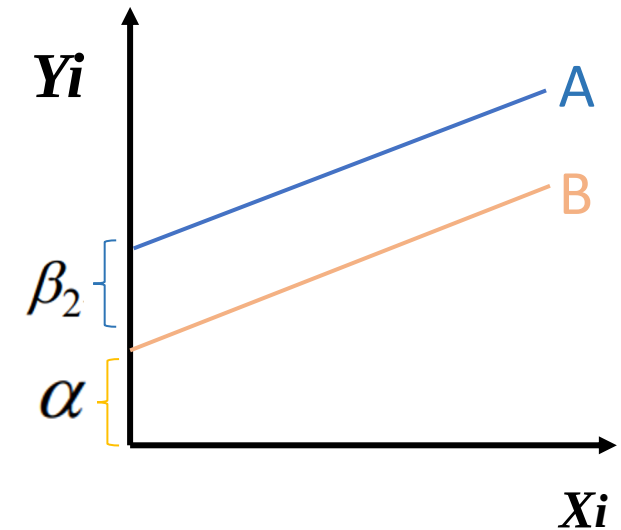
Para representar duas categorias (A e B) em um modelo de regressão, é necessária apenas uma variável binária (D_i)

A categoria de referência é aquela em que $D_i=0$

Categoria	D_i
A	1
B	0

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + e_i$$

O coeficiente B_2 indica quanto será Y em média, maior (ou menor) para a categoria **A** ($D=1$) que a categoria de referência B ($D=0$), **independentemente do valor de X_i .**



A $Y_i = (\alpha + \beta_2) + \beta_1 X_i + e_i$

B $Y_i = \alpha + \beta_1 X_i + e_i$

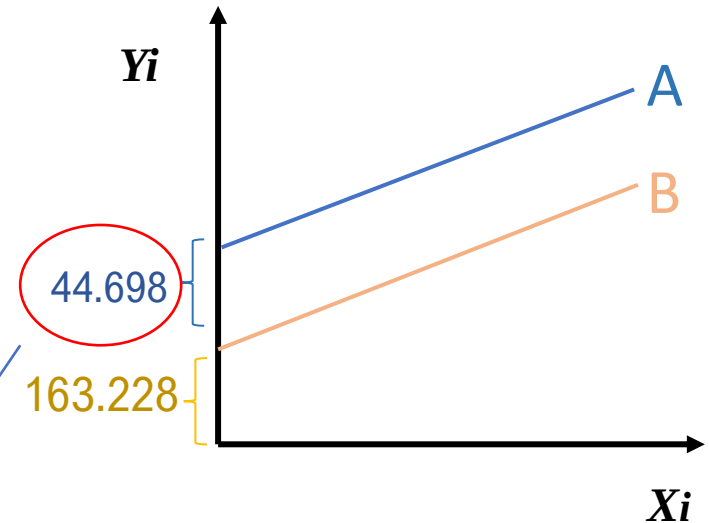
Duas Categorias

Aplicação: Modelo de estimativa de Valor de Terreno, com a área e oferta (Sim=1; Transação=0)

$$\text{Valor} = 163.228,19 + 1.170,89 * \text{Area} + 44.698,96 * \text{Oferta}$$

O coeficiente **44698,96** indica quanto será **Valor do imóvel** (em média) maior para a categoria **A (Oferta =1)** que a categoria de referência **B (Transação=0)**, mantida **Xi (Área)** constante

A diferença entre oferta e transação (\$44.698) ou **13%**, não depende do tamanho da área, por isso as retas são paralelas



$$\text{Oferta} = (163228, + 44698 + 1170,89 * 150 = \text{R\$}338.862$$

$$\text{Transação} = 1632289 + 1170,89 * 150 = \text{R\$}383.561$$

RESUMO DOS RESULTADOS

<i>Estatística de regressão</i>	
R múltiplo	0,86485545
R-Quadrado	0,74797495
R-quadrado aju	0,72984965
Erro padrão	21053,3893
Observações	26

ANOVA

	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>de significação</i>
Regressão	2	2,81E+10	1,4E+10	5,38846059	0,01204313
Resíduo	23	5,99E+10	2,61E+09		
Total	25	8,8E+10			

	<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>
Interseção	163228,198	82155,81	1,986813	5,8982%
Area total	1170,89553	608,2123	1,925143	6,6666%
Oferta	44698,9697	21228,97	2,105564	4,6365%

$$\text{Valor} = 163228,19 + 1170,89 * \text{Area} + 44698,96 * \text{Oferta}$$

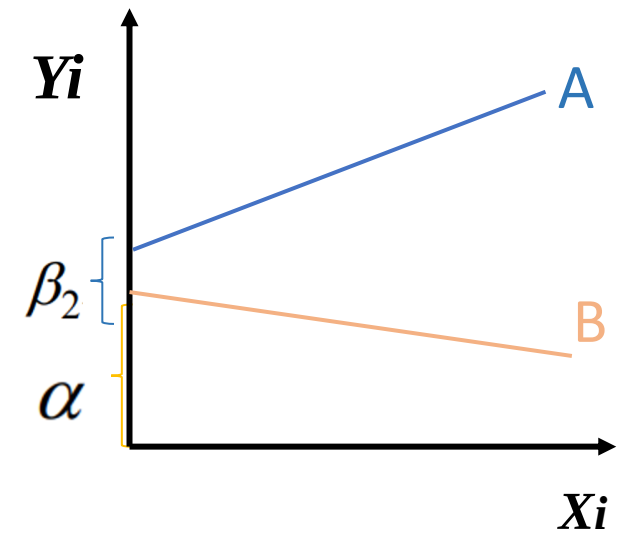
HIPÓTESE NULA E HIPÓTESE ALTERNATIVA
– A hipótese nula de não-existência de diferença entre oferta e transação':
H0: b2 = 0.
– A hipótese alternativa de que existe diferença
H0: b2 ≠ 0.

Variável *Dummy* de Inclinação

Às vezes, é possível que os efeitos de fatores qualitativos não sejam independentes, isto é, o modelo pode ser multiplicativo. Em outras palavras, pode haver interação entre as variáveis qualitativas. As retas de regressão podem ter o mesmo intercepto com coeficientes de inclinação distintos.

Dessa forma, as retas de regressão que representam essas especificações, para cada um dos diferentes grupos, precisam ser especificadas

O produto de uma variável *dummy* por uma variável contínua resulta no que se pode chamar de "variável dummy de inclinação e/ou variável de interação", e é recomendável, para em um único modelo, produzir uma estimativa para o termo independente e coeficientes de inclinação distintos.



Observa-se que esta análise pode ser combinada. Pode-se estimar uma equação de regressão com diferentes interceptos e diferentes coeficientes de inclinação.

Variável Dicotômica

Aplicação: Modelo de valor de casas com parâmetro qualitativo de localização para descrever se está em condomínio fechado (**sim=1**) ou em outras partes (**não =0**) e área

RESUMO DOS RESULTADOS

<i>Estatística de regressão</i>	
R múltiplo	0,836214
R-Quadrado	0,699254
R-quadrado ajus	0,665838
Erro padrão	718,0949
Observações	21

ANOVA

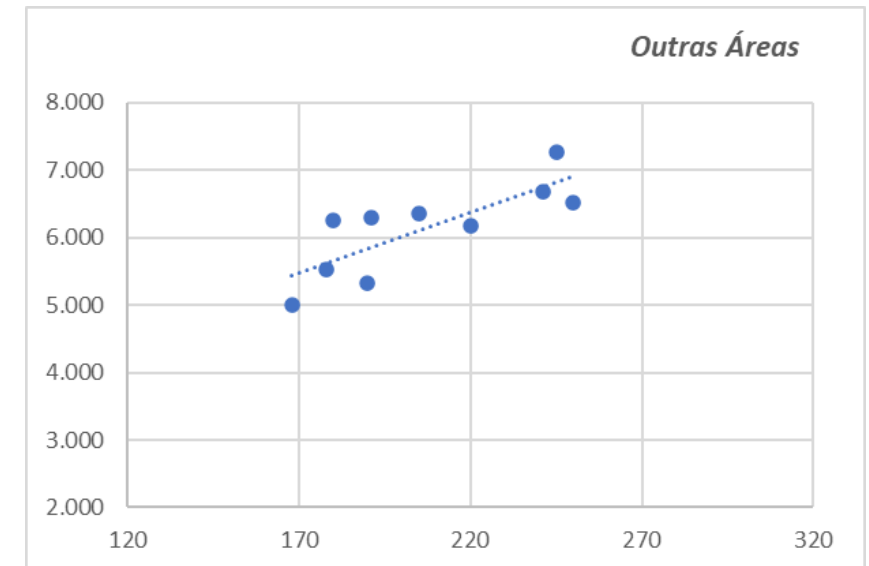
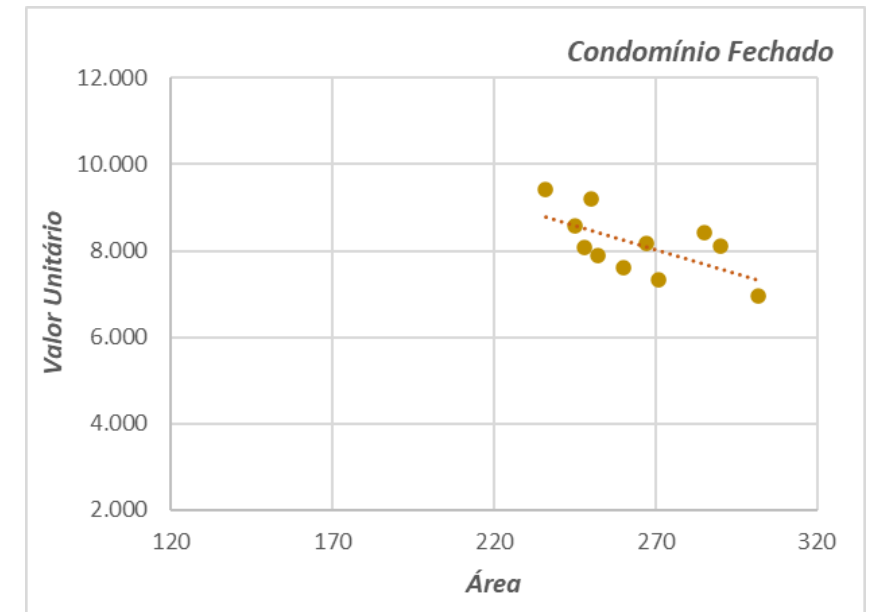
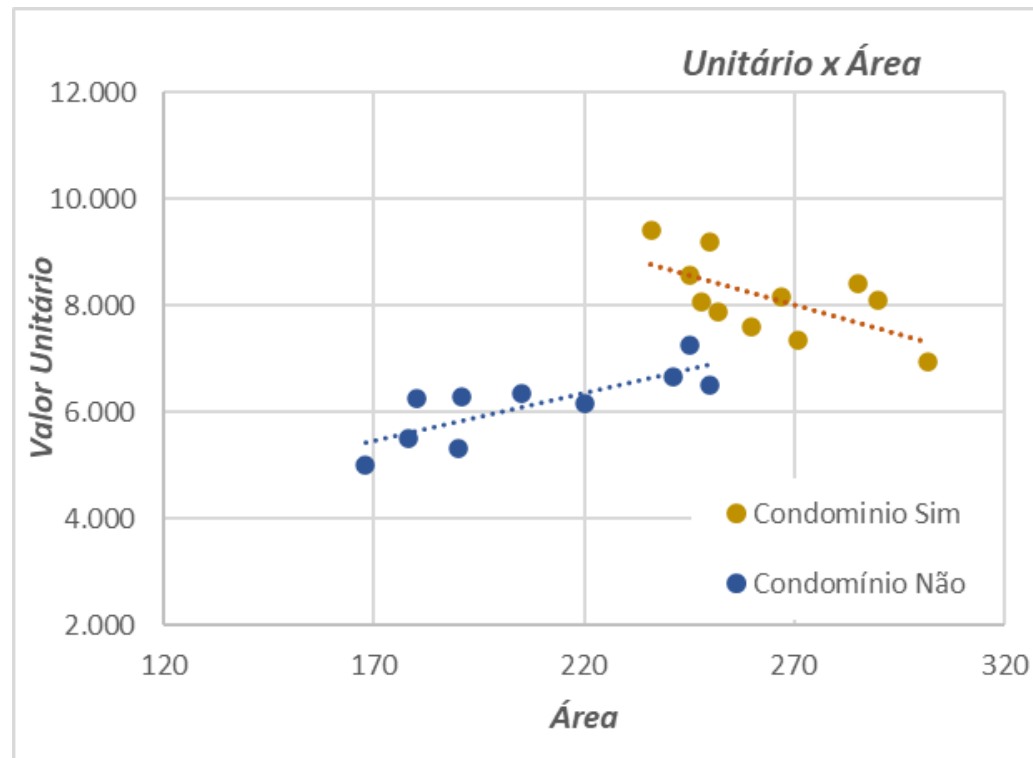
	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>e significação</i>
Regressão	2	21580975	10790488	20,9255714	2,01E-05
Resíduo	18	9281886	515660,4		
Total	20	30862861			

	<i>Coefficiente</i>	<i>erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>
Interseção	5325,094	1340,592	3,972194	0,00089349
Condominio (sin	1795,11	482,5389	3,720135	0,00156679
Area	3,92939	1,388877	2,615036	0,05462285

Projeção de valores

		Condominio		outras partes	
		Sim		não	
<i>b0</i>	5325,09				
<i>b1</i>	1795,11	1	1	0	0
<i>b2</i>	3,92939	245	302	168	245
		R\$ 8.082,90	R\$ 8.306,88	R\$ 5.985,23	R\$ 6.287,79

Entretanto *Análises gráficas – indicam comportamentos distintos para área nos dois grupos (para casas em condomínio e fora do condomínio)*



Modelo com variável de interação

		Condominio		outras partes	
		Sim		não	
b0	5325,09				
b1	1795,11	1	1	0	0
b2	3,92939	245	302	168	245
		R\$ 8.082,90	R\$ 8.306,88	R\$ 5.985,23	R\$ 6.287,79

RESUMO DOS RESULTADOS

<i>Estatística de regressão</i>	
R múltiplo	0,92166
R-Quadrado	0,849458
R-quadrado ajustado	0,822892
Erro padrão	522,7845
Observações	21

		Condominio		outras partes	
		Sim		não	
b0	2441,42				
b1	11630,80	1	1	0	0
b2	17,87	245	302	168	245
b3	-40,26	245	302	0	0
		R\$ 8.587,68	R\$ 7.311,68	R\$ 5.444,19	R\$ 6.820,46

ANOVA

	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>e significação</i>
Regressão	3	26216699	8738900	31,97505518	3,27E-07
Resíduo	17	4646162	273303,7		
Total	20	30862861			

	<i>Coefficiente</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>
Interseção	2441,423	1201,155	2,032562	5,80%
Condominio (sim/não)	11630,8	2413,888	4,818284	0,02%
Area	17,87364	5,753019	3,106828	0,64%
Intereção Area x Condomínio	-40,2595	9,775354	-4,11847	0,07%

Múltiplas Categorias

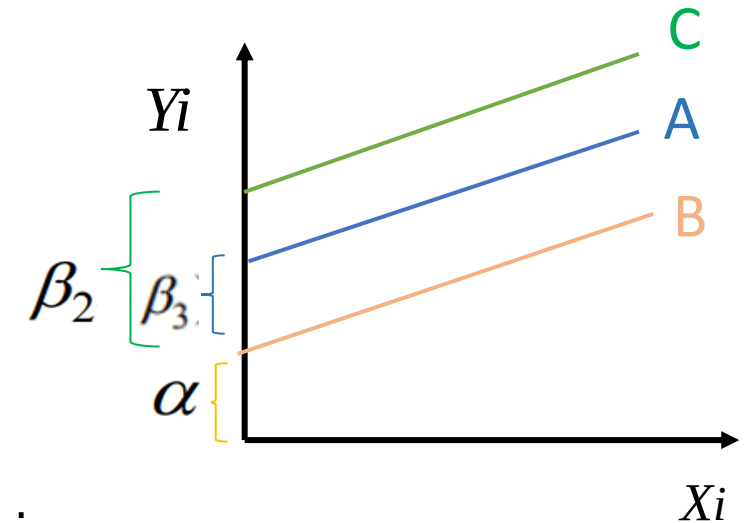
Para representar k categorias nominais, são **necessárias k-1 variáveis binárias D's**. A referência é a categoria em que todas as variáveis binárias **assumem 0 simultaneamente**.

Por exemplo, para k=3:

Categoria	D_{1i}	D_{2i}
A	1	0
B	0	1
C	0	0

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_{1i} + \beta_3 D_{2i} + e_i$$

O coeficiente β_2 indica quanto Y será, em média, maior para a categoria A ($D_1=1$) que a categoria de referência C ($D_1=0$ e $D_2=0$), *independentemente do valor de X*. O coeficiente β_3 indica quanto Y seria, em média, maior para a categoria B ($D_2=1$) que a categoria de referência C.



A $Y_i = (\alpha + \beta_2) + \beta_1 X_i + e_i$

B $Y_i = (\alpha + \beta_3) + \beta_1 X_i + e_i$

C $Y_i = \alpha + \beta_1 X_i + e_i$

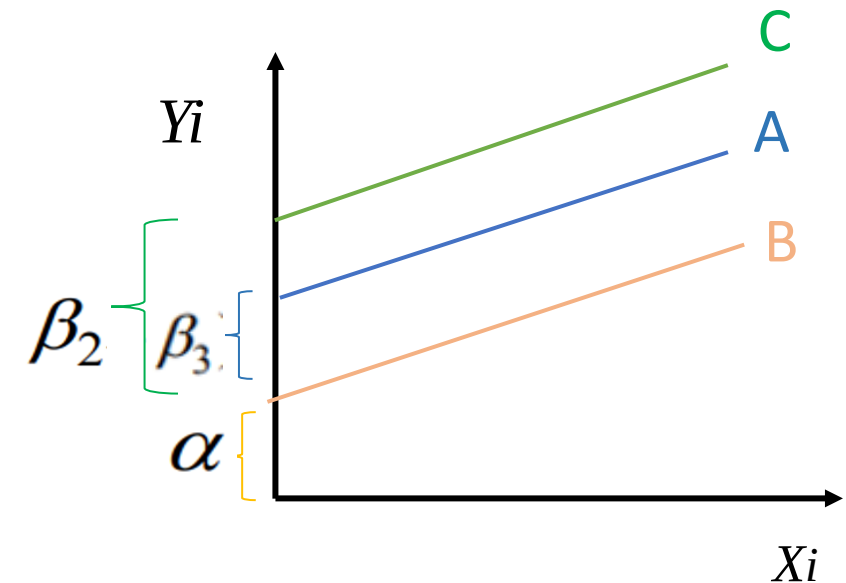
Múltiplas Categorias

Valor unitário = $+7825,51 - 23,49 * \text{Area} + 523,16 * \text{Bairro B} + 1659,72 * \text{Bairro C}$

O coeficiente **523,16** indica Valor Unitário (Y), em média, maior para a categoria **A** ($D1=1$) que a categoria de referência **C** ($D1=0$ e $D2=0$), *independentemente do valor de X*. O coeficiente 1.559 indica Valor Unitário (Y), em média, maior para a categoria **B** ($D2i=1$) que a categoria de referência **C**.

Exemplo, para $k=3$ (três bairros)

Bairro	D_{1i}	D_{2i}
A	1	0
B	0	1
C	0	0



Múltiplas Categorias - modelo

RESUMO DOS RESULTADOS

estatística de regressão

R múltiplo 0,881483

R-Quadrac 0,758505

R-quadrac 0,736708

Erro padrão 476,9138

Observaçô 51

ANOVA

	gl	SQ	MQ	F	de significação
Regressão	3	20613541,87	6871181	20,21006047	0,00493%
Resíduo	47	10689998,11	227446,8		
Total	50	31303539,98			

	Coefficientes	Erro padrão	Stat t	valor-P
Interseção	7825,515	944,0411878	8,289378	0,945%
Área	-23,4901	5,626708902	-4,17475	0,013%
Bairro B	523,1553	150,5258306	3,475518	0,111%
Bairro C	1259,718	186,5855284	8,895214	0,012%

Equação Valores de Y

b0	7825,51				Diferenças
b1	-23,49	200	200	200	
b2	523,16	0	1	0	17%
b3	1259,72	0	0	1	40%
		R\$ 3.127	R\$ 3.651	R\$ 4.387	



O coeficiente de inclinação é semelhante aos 3 bairros

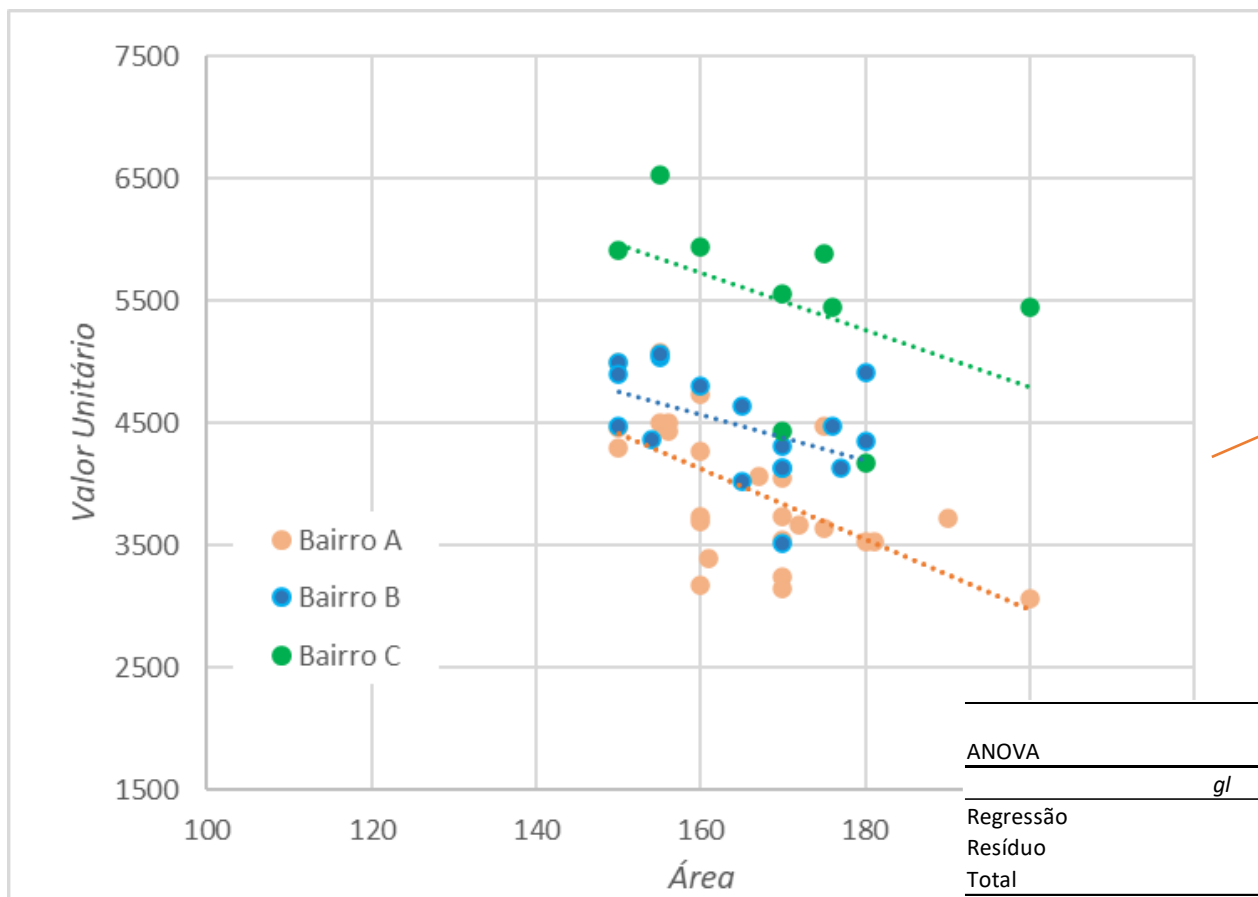
Neste caso, a hipótese implícita, é de que as variáveis de intercepto sejam aditivas.

O efeito de cada fator qualitativo é somado ao intercepto de regressão, e o efeito de qualquer variável binária é independente da variação da área

Análise Gráfica –

11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



*bairro 2 inclinação
levemente distinta*

ANOVA

	gl	SQ	MQ	F	F de significação
Regressão	4	20666945,17	5166736,29	22,34454484	2,67251E-10
Resíduo	46	10636594,8	231230,322		
Total	50	31303539,98			

	Coeficientes	Erro padrão	Stat t	valor-P
Interseção	8091,799246	1101,390262	7,34689558	0%
Área	-25,08539687	6,573108956	-3,8163671	0%
Bairro B	-509,7898406	2154,743613	-0,2365896	81%
Bairro C	1665,695154	188,5416989	8,83462472	0%
Interação Bairro B	6,255000058	13,01564356	0,48057555	63%

11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

Obrigada

“ Uma resposta aproximada para o problema certo vale muito mais do que uma resposta exata para um problema aproximado. ”

(John Wilder Tukey)

REALIZAÇÃO



