

ANALISE E SOLUÇÕES DE PROBLEMAS TÍPICOS EM MODELOS DE REGRESSÃO LINEAR

Rubens Alves Dantas, DSc.



11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



IBAPE NACIONAL
Instituto Brasileiro de Avaliações
e Perícias de Engenharia

ETAPAS DA MODELAGEM

- Especificação do Modelo: Escolha de variáveis e escalas mais adequadas
- Variável Explicada - *Preço Unitário ou Preço Total*
- Codificação Variáveis Qualitativas - *Dummy, Proxy, Código Alocado ou Ajustado*
- Análise Exploratória dos Dados – *Amplitudes, Gráficos de Dispersão, Matriz de Correlações*
- Ajustamento do Modelo
 - $Y_i = \beta_0 + \beta_1 * X_{i1} + \beta_2 * X_{i2} + \dots + \beta_k * X_{ik} + \epsilon_i$, $i=1, \dots, n$
- Análise dos resultados:
 - * *Poder de Explicação*
 - * *Testes de significância Global e Individual dos Parâmetros*
 - * *Verificação dos Pressupostos Básicos* (*Normalidade, Homocedasticidade, Não autocorrelação, Não Multicolinearidade, ...*)



11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



Dois Problemas Típicos

- Exclusão de Variáveis Relevantes
- Multicolinearidade

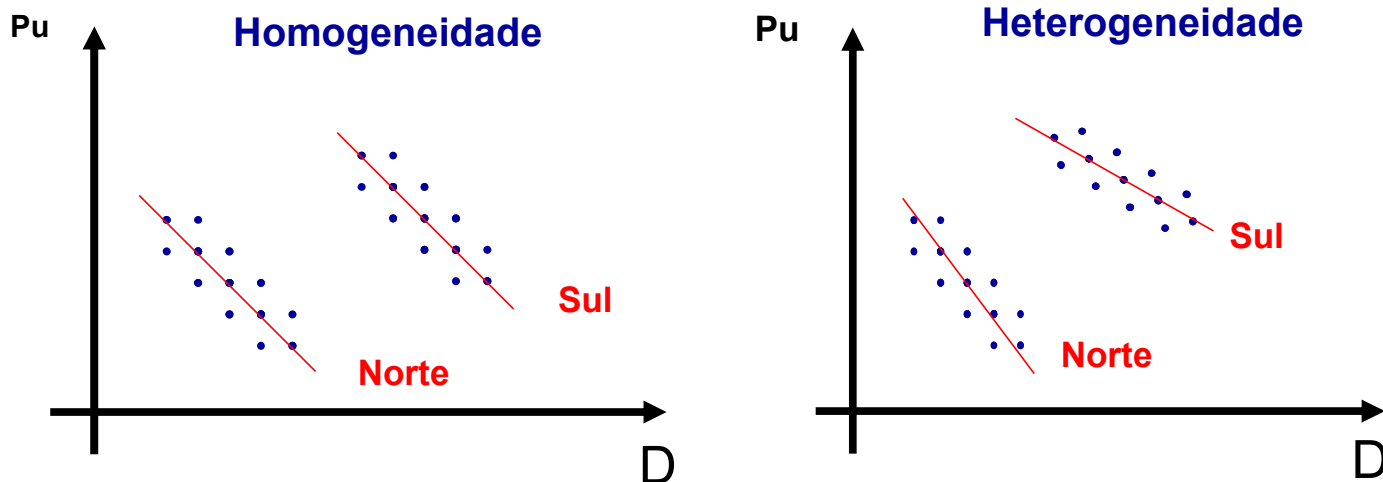


Item 8.2.1.4.1 da NBR 14.653-2 (obrigatório)

Deve-se levar em conta que qualquer modelo é uma representação simplificada do mercado, uma vez que não considera todas as suas informações. Por isso, precisam ser tomados cuidados científicos na sua elaboração, desde a preparação da pesquisa e o trabalho de campo, até o exame final dos resultados. O poder de predição do modelo **deve** ser verificado a partir do gráfico de preços observados na abscissa versus valores estimados pelo modelo na ordenada, que **deve** apresentar pontos próximos da bissetriz do primeiro quadrante. Alternativamente, podem ser utilizados procedimentos de validação. A qualidade da amostra **deve** estar assegurada quanto a:

- a) **correta identificação dos dados de mercado, com especificação e quantificação das principais variáveis levantadas, mesmo aquelas não utilizadas no modelo;**
- b) **isenção das fontes de informação;**
- c) **identificação das fontes de informação, observada a exceção contida em 8.2.1.3.3;**
- d) **número de dados de mercado efetivamente utilizados, de acordo com o grau de fundamentação;**
- e) **sua semelhança com o imóvel objeto da avaliação, no que diz respeito à sua situação, à destinação, ao grau de aproveitamento e às características físicas; diferenças relevantes perante o avaliando, devem ser tratadas adequadamente nos modelos adotados;**
- f) **inserção de mais de um tipo de agrupamento no mesmo modelo. Nestes casos, o engenheiro de avaliações deve se certificar de ter contemplado as diferenças significativas entre esses grupos, sendo obrigatória a verificação da influência das interações entre as variáveis. Recomenda-se a inclusão dos endereços completos dos dados de mercado.**

Avaliação com dados de duas Regiões



- Variáveis Distância ao Centro (D) e Região (R): Norte - 0 e Sul - 1

$$Pu = b_0 + b_1 D + b_2 \cdot R$$

$$R = 0 \Rightarrow Pu = b_0 + b_1 D$$

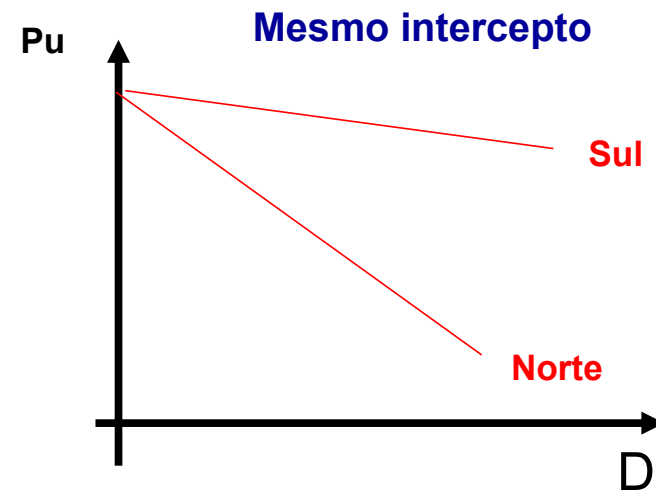
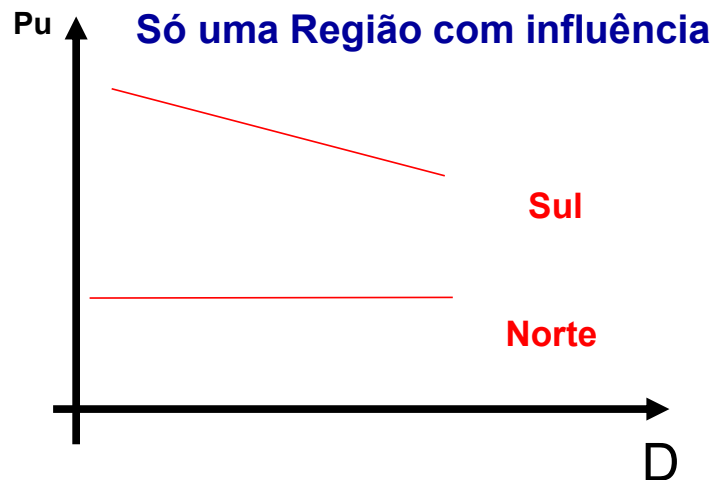
$$R = 1 \Rightarrow Pu = (b_0 + b_2) + b_1 D$$

$$Pu = b_0 + b_1 \cdot D + b_2 \cdot R + b_3 D \cdot R$$

$$R = 0 \Rightarrow Pu = b_0 + b_1 \cdot D$$

$$R = 1 \Rightarrow Pu = (b_0 + b_2) + (b_1 + b_3) D$$

Outras Situações Possíveis



- Variáveis Distância ao Centro (D) e Região (R): **Norte - 0 e Sul - 1**

$$Pu = b_0 + b_1 \cdot D \cdot R$$

$$R = 0 \Rightarrow Pu = b_0$$

$$R = 1 \Rightarrow Pu = b_0 + b_1 D$$

$$Pu = b_0 + b_1 \cdot D + b_2 D \cdot R$$

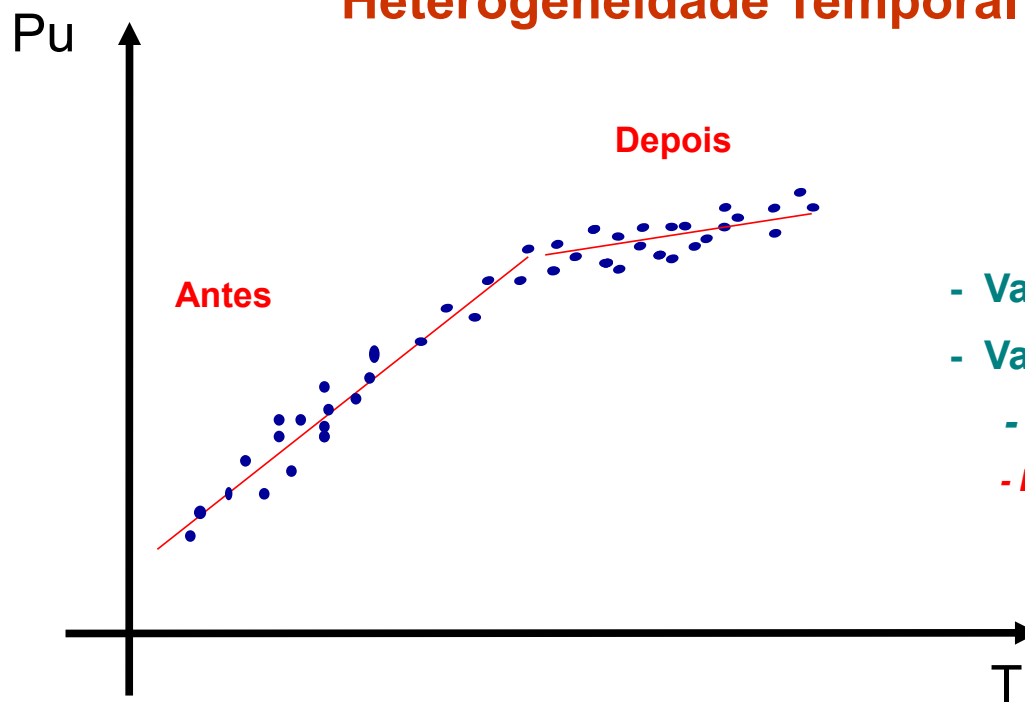
$$R = 0 \Rightarrow Pu = b_0 + b_1 \cdot D$$

$$R = 1 \Rightarrow Pu = b_0 + (b_1 + b_2) D$$

11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

Heterogeneidade Temporal



- Variável Temporal - T

- Variável Época - E :

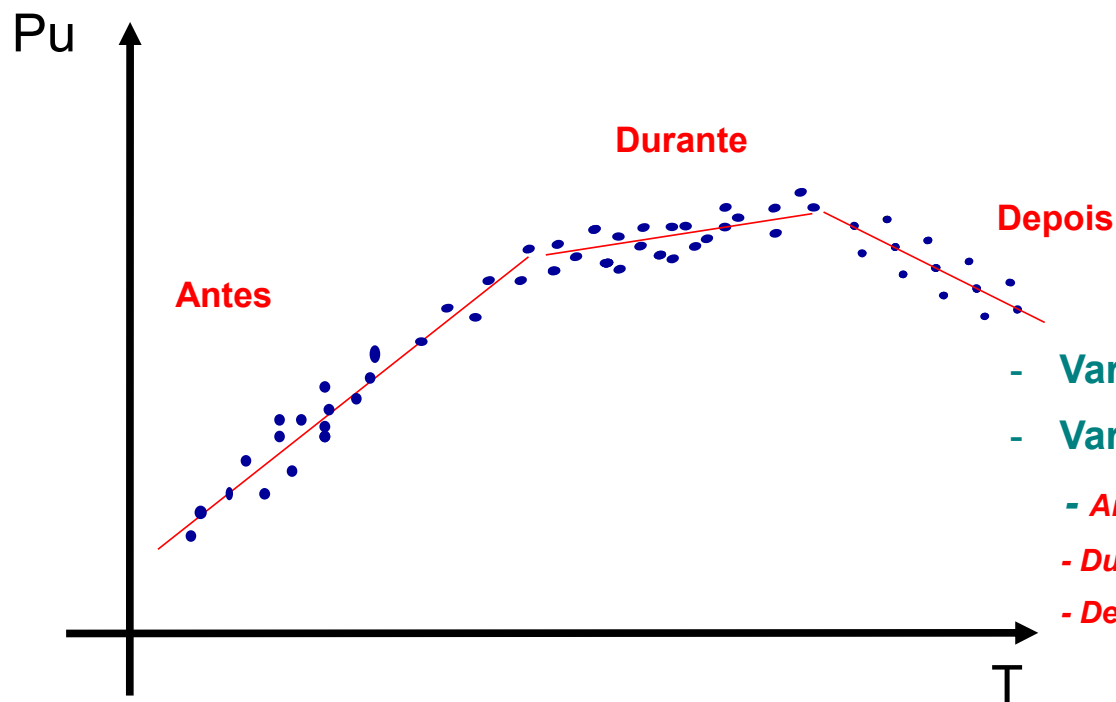
- Antes da Pandemia - 0

- Depois da Pandemia - 1

Especificação do Modelo:

$$- Pu = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_T * T + \beta_N * E + \beta_{TE} * T * E$$

Situação de duas quebras estruturais



- Variável Tempotal - T

- Variável Época - E :

- Antes da Pandemia ($E0$) : 0 - 0

- Durante a Pandemia ($E1$): 1 - 0

- Depois da Pandemia ($E2$): 0 - 1

Especificação do Modelo:

$$- P_u = \beta_0 + \beta_1 * X_1 \dots + \beta_T * T + \beta_{E1} * E1 + \beta_{E2} * E2 + \beta_{T E1} * T * E1 + \beta_{T E2} * T * E2$$

Aplicação para dados Antes, Durante e Depois da Pandemia

Amostra:

Dado	Época (E)	DC(km)	PU(R\$/m2)
1	Antes	18	10,00
2	Antes	8	18,00
3	Antes	1	24,00
4	Antes	12	16,00
5	Antes	13	14,00
6	Antes	11	17,00
7	Durante	20	8,50
8	Durante	10	10,00
9	Durante	1	13,00
10	Durante	15	9,00
11	Durante	14	10,00
12	Durante	5	12,00
13	Depois	18	18,00
14	Depois	10	25,00
15	Depois	1	38,00
16	Depois	5	21,00
17	Depois	20	18,00
18	Depois	19	19,00

Codificação

Dado	Época	E1	E2	E1*DC	E2*DC	DC	PU
1	Antes (E0)	0	0	0	0	18	10,00
2	Antes (E0)	0	0	0	0	8	18,00
3	Antes (E0)	0	0	0	0	1	24,00
4	Antes (E0)	0	0	0	0	12	16,00
5	Antes (E0)	0	0	0	0	13	14,00
6	Antes (E0)	0	0	0	0	11	17,00
7	Durante (E1)	1	0	20	0	20	8,50
8	Durante (E1)	1	0	10	0	10	10,00
9	Durante (E1)	1	0	1	0	1	13,00
10	Durante (E1)	1	0	15	0	15	9,00
11	Durante (E1)	1	0	14	0	14	10,00
12	Durante (E1)	1	0	5	0	5	12,00
13	Depois (E2)	0	1	0	18	18	18,00
14	Depois (E2)	0	1	0	10	10	25,00
15	Depois (E2)	0	1	0	1	1	38,00
16	Depois (E2)	0	1	0	5	5	21,00
17	Depois (E2)	0	1	0	20	20	18,00
18	Depois (E2)	0	1	0	19	19	19,00

Ajustamento do Modelos de Regressão pelo SAB

Relatório Geral Resíduos Grandes Resíduos Pontos Influenciantes Matriz de Correlações Dados não Utilizados

Resultados Gerais

Modelo	Clássico de Regressão
Desvio Padrão	0.1135
Estatística Fc	40,6560
Nível de Significância do Modelo	0.0001
Coefficiente de determinação	0.9443
Coefficiente de determinação ajustado	0.9210

Resultados Por Variável

Variável	Escala	Coefficientes	Desvio Padrão	Estatística t	Nível de Significância
Interseção		3.2878	0.1046	31,4336	0.0000
E1	x	-0.7072	0.1390	-5.0886	0.0003
E2	x	0.1055	0.1378	1,3464	0.2031
E1*DC	x	0.0264	0.0115	2.2938	0.0407
E2*DC	x	0.0191	0.0109	1,7417	0.1071
DC(km)	x	-0.0494	0.0089	-5.5292	0.0001

Equação do Modelo:

$$PU(R\$/m^2) = 26,7840 + 0,4930 * E1 + 1,2039 * E2 + 1,0268 * E1*DC + 1,0192 * E2*DC + 0,9518 * DC(km)$$

Interpretação do Modelo

$$PU = 26,78 * 0,49^{E1} * 1,20^{E2} * 1,03^{E1*DC} * 1,02^{E2*DC} * 0,95^{DC}$$

Modelo Antes da Pandemia (E1= 0 e E2 = 0)

$$PU = 26,78 * 0,95^{DC}$$

Modelo Durante a Pandemia (E1= 1 e E2 = 0)

$$PU = 26,78 * 0,49 * 1,03^{DC} * 0,95^{DC} = 13,12 * 0,98^{DC}$$

Modelo Depois da Pandemia (E1= 0 e E2 = 1)

$$PU = 26,78 * 1,20 * 1,02^{DC} * 0,95^{DC} = 32,10 * 0,97^{DC}$$

Conclusão: Durante a pandemia houve queda acentuada nos preços e também uma redução na importância distância ao centro, de 5% para 2%, a cada km. Depois houve uma retomada, com média preços em patamares superiores aos praticados antes da pandemia, e também um aumento da importância de estar mais próximo ao centro, de 1% para cada km .

Multicolinearidade

Algumas recomendações da NBR 14.653-2

*Em caso de correlação linear elevada entre quaisquer subconjuntos de variáveis independentes, isto é, multicolinearidade, **deve-se** examinar a coerência das características do imóvel avaliando com a estrutura de multicolinearidade inferida, **vedada a utilização do modelo em caso de incoerência.***

item A.2, alinea g)

*Para verificação da multicolinearidade **deve-se**, em primeiro lugar, analisar a matriz das correlações, que espelha as dependências lineares de primeira ordem entre as variáveis independentes, com atenção especial para resultados superiores a 0,80...*

Item A.2.1.5.2

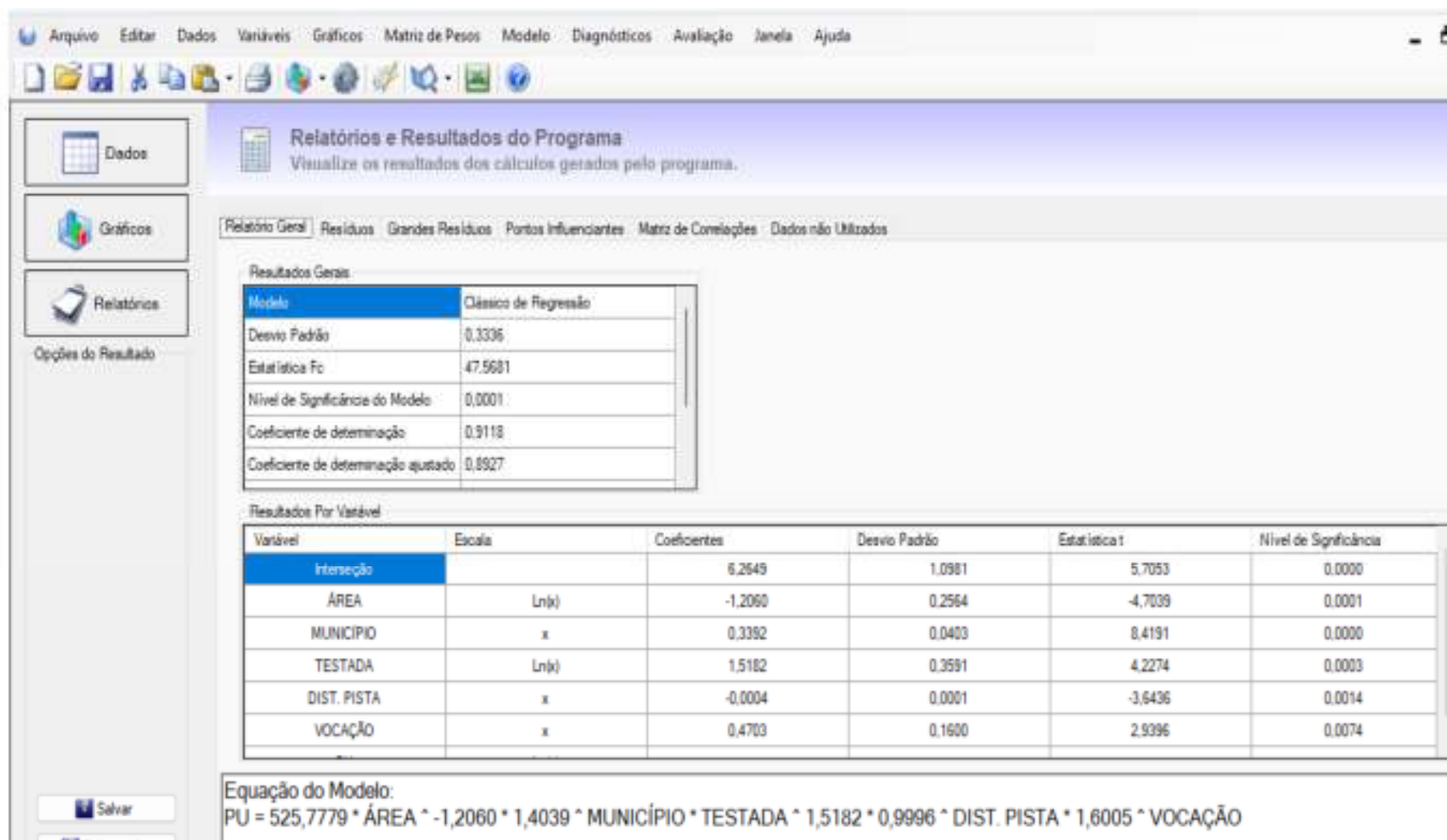


11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



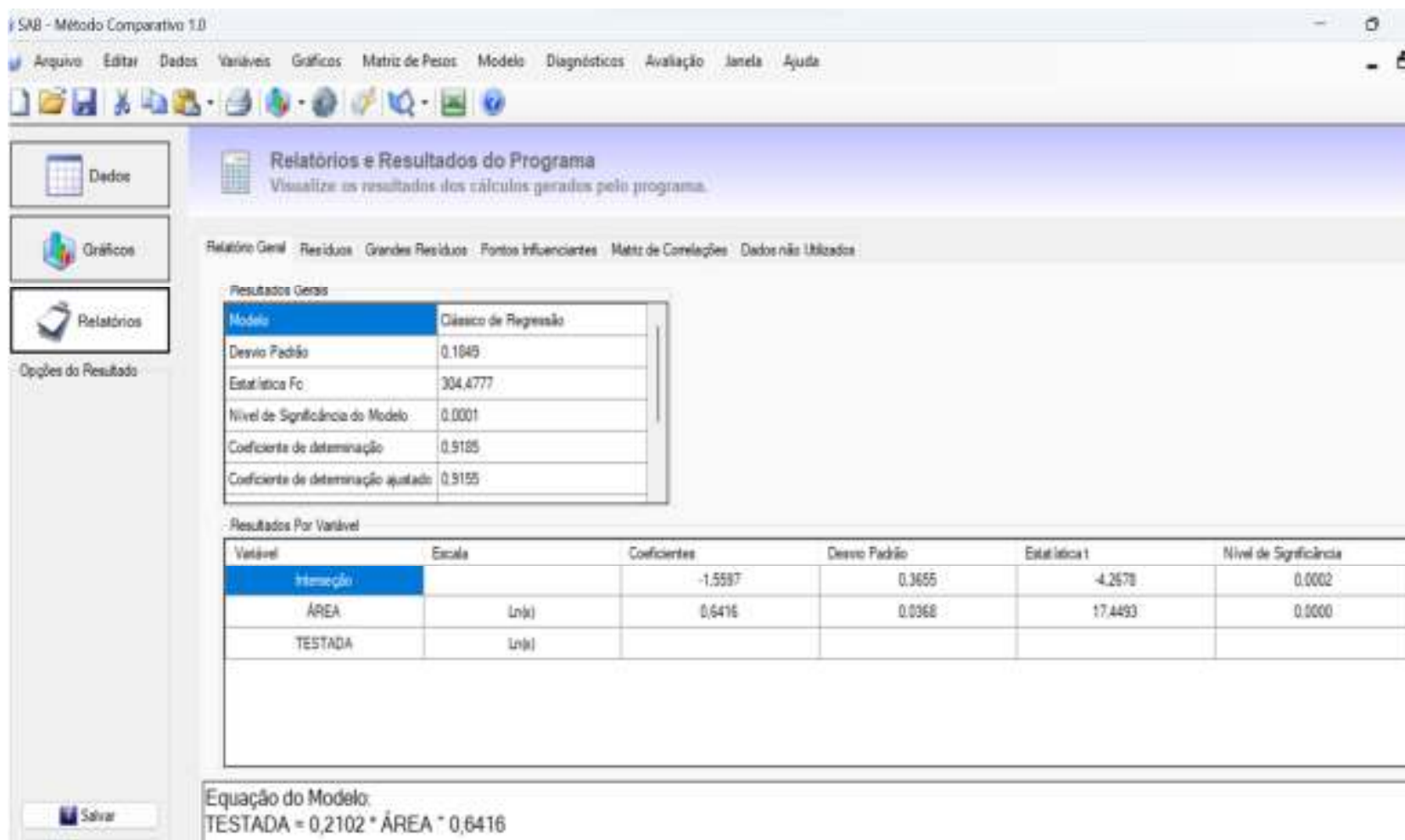
Caso Prático: Modelo de um Perito Judicial



Caso Prático: Matriz das Correlações



Caso Prático: Modelo de Frente x Área



Ex: Para Area = 40.000,00 => FE = 0,2102 * 40.000 ^ 0,6416 = 190,00m

SIMULAÇÕES

ÁREA(m2)	MUN	FE(m)	DP(m)	VO	PU(R\$/m2)	PT (R\$)
20.000,00	4	200	500	0	34,03	680.630,57
30.000,00	4	200	500	0	20,87	626.090,83
40.000,00	4	200	500	0	14,75	590.065,81
50.000,00	4	200	500	0	11,27	563.556,38
60.000,00	4	200	500	0	9,05	542.783,13
80.000,00	4	200	500	0	6,39	511.551,61

Conclusão: Na presença da multicolinearidade é fundamental observar as restrições de uso do modelo. O não atendimento desse preceito normativo pode gerar resultados completamente distorcidos da realidade do mercado. No caso, pode-se observar as incoerências: Metade da área com a mesma frente (20mil) vale mais que a área toda (40mil) e também, o dobro da área (80mil), com a mesma frente, vale menos que a área toda.



11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



IBAPE NACIONAL
Instituto Brasileiro de Avaliações
e Perícias de Engenharia

Obrigado!

rubens@dantas.eng.br

www.dantas.eng.br

