



IBAPE NACIONAL
Instituto Brasileiro de Avaliações
e Perícias de Engenharia



IBAPE BAHIA
Instituto Brasileiro de
Avaliações e Perícias de
Engenharia da Bahia

Toda Micronumerosidade é Prejudicial? Depende uma Proposta para Alteração da ABNT NBR 14.653-2/2011

**Radegaz Nasser Junior; Luiz Alberto
Pretti**



O conteúdo dos trabalhos técnicos apresentados no COBREAP é de inteira responsabilidade de seus autores.

TODA MICRONUMEROSIDADE, É PREJUDICIAL? DEPENDE UMA PROPOSTA PARA ALTERAÇÃO DA ABNT NBR 14.653-2/2011

Resumo.

O objetivo do trabalho é questionar a imposição restritiva constante na ABNT NBR 16.543-2:2011 / anexo A – Procedimentos para a utilização de modelos de regressão linear – A.2 - Pressupostos básicos, Subitem (a) e propor em uma nova atualização da citada Norma ABNT, a adoção de outros testes num modelo estatístico, para verificar a importância da existência de micronumerosidade, não se baseando somente nos preceitos subjetivos normativos atuais. Visa principalmente verificar se esta restrição afeta ou não, de forma negativa, a definição do valor adequado ao mercado imobiliário. A proposta sugere a utilização de técnicas estatísticas, tais como o teste *t*, verificação das correlações entre as variáveis independentes (multicolinearidade) e outros demonstrados ao longo deste trabalho a fim de mostrar que cabe ao profissional Avaliador decidir se após os testes e análises, a micronumerosidade é prejudicial ou não na definição do valor provável de transação para o imóvel avaliando. Como exemplo prático, foi desenvolvido um caso real que reforça a necessidade da adoção e verificação de outros testes estatísticos para a comprovação técnica da presença da micronumerosidade no modelo.

Palavras-chave: micronumerosidade, multicolinearidade, teste de hipótese

1 INTRODUÇÃO

Atendendo a boa técnica estatística direcionada a avaliação de imóveis que qualquer natureza, a avaliação imobiliária deve atender quatro fundamentos:

- O objetivo da avaliação,
- Os dados relativos ao imóvel avaliando,
- Os dados de mercado para compor a amostra representativa da população
- O tratamento estatístico aplicável à característica do avaliando

Com esses fundamentos definidos, torna-se possível definir a metodologia de cálculo dentre as diversas possíveis prescritas na ABNT 14.653/1:2019.

Partindo desses fundamentos, de acordo com a teoria estatística tradicionalmente aceita faz-se necessário ainda o atendimento aos pressupostos estatísticos mínimos, situação que possibilita a obtenção de boas inferências quando todos esses pressupostos ou premissas do modelo proposto são alcançados. Esses pressupostos são:

O modelo é linear nos parâmetros

$$Y_i = a_i + b_1 X_i + e_i$$

Onde,

A variável independente Y_i é aleatória

As variáveis independentes não são aleatórias (são controladas)

Os erros são aleatórios

Isto é, todo erro que não é captado por $y_i = a_i + b_1 x_i$ está em e_i aleatório

A esperança dos erros e_i , dadas as variáveis explicativas, x_i é nula.

Isto é a média dos erros é zero representada por $E(e_i / x_i) = \mu_i = 0$.

A variância dos termos de erros e_i dadas as variáveis explicativas x_i é constante

Isto é, a variância dos erros é constante = Homocedasticidade = $\text{Var}(e_i / x_i) = \sigma^2$,

A covariância entre os termos de erros é nula

Isto é $\text{Cov}(e_i / x_i) = 0$ para $i \neq j$

Normalidade dos resíduos

Isto é, os termos de erros seguem uma distribuição normal com média zero e variância ou desvio padrão constante representados por $e_i \sim N(0, \sigma^2)$.

Isto posto conclui-se que para cada pressuposto que não for verificado, implicará na obtenção de um modelo estatístico regressivo inferencial sujeito a consequências que o tornarão impreciso.

Seguindo, é comum, durante o processo de elaboração de um modelo avaliatório inferencial, o profissional de avaliações ser confrontado com a presença de micronumerosidade nos termos prescritos pela ABNT 14.653-2:2011 ao buscar atender ANEXO A, em seus pressupostos básicos.

Recomenda-se que as características específicas do imóvel avaliando estejam contempladas na amostra utilizada em número representativo de dados de mercado;

Os efeitos deste mau condicionamento causam o aumento do multicolinearidade. Os valores projetados de avaliação para modelos multicolineares podem ser inadequados, em relação ao grau de precisão mínimo permitido pela NBR 14653-2:2011, restringindo a aplicabilidade e generalidade do modelo estimado.

Existem várias formas de se detectar a existência de micronumerosidade em um modelo de regressão e outras soluções que se atendidos, passam pela adoção do modelo, que serão apresentadas neste trabalho.

O presente estudo apresenta um caso real de avaliação de terreno, utilizando um modelo estatístico para a determinação do valor de mercado, contendo a “famigerada” micronumerosidade.

2 EXPOSIÇÃO DE MOTIVOS

2.1 A Micronumerosidade

A ABNT 14.653-2:2011 trata o fenômeno da micronumerosidade em um formato tanto quanto empírico como demonstra a transcrição abaixo:

- a) para evitar a micronumerosidade, o número mínimo de dados efetivamente utilizados (n) no modelo deve obedecer aos seguintes critérios, com respeito ao número de variáveis independentes (k):

Quantidade de dados	Restrição em relação ao tamanho da amostra
$n \geq 3(k + 1)$	-
$n_i \geq 3$	para $n \leq 30$,
$n_i \geq 10\% n$	para $30 < n \leq 100$
$n_i \geq 10$	para $n > 100$,
Atuais recomendações normativas	

onde:

n_i é o número de dados de mesma característica, no caso de utilização de variáveis dicotômicas e variáveis qualitativas expressas por códigos alocados ou códigos ajustados;

O fato real é que essa abordagem não encontra amparo na literatura e é motivo de comentário em Gujarati & Porter (5ªEd / 2011) onde textualmente lê-se:

“O termo micronumerosidade deve-se a Goldberger que criou o termo **micronumerosidade**, para contrapor à polissílabo multicolinearidade.

De acordo com Goldberger, a **micronumerosidade exata** (a contraparte da multicolinearidade exata) surge quando n , o tamanho da mostra, é zero, caso em que qualquer tipo de estimação é impossível.

A quase micronumerosidade, como a quase multicolinearidade, surge quando o número de observações mal excede o número de parâmetros a serem estimados, originalmente, a existência de uma relação linear “perfeita” ou exata entre algumas ou todas as variáveis explanatórias de um modelo de regressão”.

A partir dessa conceituação teórica existente na literatura atual e com base nas possibilidades teóricas de se verificar a presença da multicolinearidade, seja em presença da micronumerosidade ou não, como definida na ABNT NBR 14.653-2-2011, resta que as consequências da micronumerosidade sempre poderão estar presentes em um modelo de avaliação e da mesma forma podem ser analisados exaustivamente apenas com base na teoria disponível, sem necessidade de qualquer tipo de empirismo como o atualmente recomendado na ABNT NBR 14.653-2:2011.

Assim, as consequências da micronumerosidade podem ser tratadas como fenômenos que levam a multicolinearidade e assim devem ser tecnicamente analisados como relacionadas por (Gujarati e Porter, 5ª Ed / 2011) e transcritas a seguir:

1	Embora sejam os melhores estimadores lineares não viesados, os estimadores de MQO têm grandes variâncias e covariâncias, tornando difícil uma estimação precisa.
2	Devido à consequência 1, os intervalos de confiança tendem a ser muito mais amplos, levando à aceitação imediata da “hipótese nula igual a zero” (isto é, o verdadeiro coeficiente populacional igual a zero).
3	Também, devido à consequência 1, a razão <i>t</i> de um ou mais coeficientes tende a ser estatisticamente insignificante.
4	Embora a razão <i>t</i> de um ou mais coeficientes seja estatisticamente insignificante, R ² , a medida geral da qualidade do ajustamento, pode ser muito alto.
5	Os estimadores de MQO e seus erros padrão podem ser sensíveis a pequenas alterações nos dados.
Consequências da micronumerosidade	

2.2 A Multicolinearidade

Conceitualmente a multicolinearidade em uma RLM (Regressão Linear Múltipla) acontece quando pelo menos dois de seus termos independentes relacionam-se entre si de forma disjunta. Pode ser explicada, observando-se o efeito isolado de duas variáveis independentes X_1 e X_2 sobre a variável dependente Y (Multicolinearidade / Econometria / Alexandre Goya) através de duas regressões simples como segue:

$Y = \alpha_1 + \beta_1 X_{1i} + e_{1i} \Rightarrow SQRe\ g_1$ $Y = \alpha_2 + \beta_2 X_{2i} + e_{2i} \Rightarrow SQRe\ g_2$	Eq.2.1
Regressões Simples de Y sobre cada variável independente X_i	

Nestas condições, quando X_1 e X_2 são independentes com algum grau de colinearidade entre si.

A ocorrência da multicolinearidade entre as variáveis independentes X_1 e X_2 é observada pela relação linear exata entre ambas, ou seja, uma relação sem o termo de erro e_i , dada pela expressão seguinte:

$X_{1i} = \lambda_2 X_{2i}$	Eq.2.2
Relação exata entre as variáveis independentes X_1 e X_2	

Nesta situação é impossível estimar os efeitos isolados de X_1 e X_2 sobre a variável dependente Y . Em casos assim, tem-se que o problema dividido em três partes, ou seja, o efeito isolado de X_1 sobre Y , O efeito isolado de X_2 sobre Y e finalmente o efeito conjunto de X_1 e X_2 sobre Y .

Constitui-se o problema da multicolinearidade o quando efeito conjunto de X_1 e X_2 sobre Y é muito superior aos efeitos isolados de cada variável independente X_1 e X_2 , caso em que ele é representado pela expressão seguinte, agora com o termo de erro v_i .

$X_{1_i} = \lambda_2 X_{2_i} + v_i$	Eq.2.3
Relação entre as variáveis independentes X_1 e X_2 com termo de erro e_i	

A multicolinearidade é uma característica da amostra e não existe um teste específico para seu diagnóstico. Por essa razão ela é medida em grau de influência em uma amostra específica e o profissional avaliador deve estar atento sempre para as seguintes análises no modelos estatísticos:

- 1 – Coeficientes de determinação altos e estatísticas t dos parâmetros pouco significativas ocorrendo simultaneamente constituem-se em sintomas diretos e clássicos da presença da multicolinearidade.
- 2 – A presença de altas correlações entre pares de variáveis independentes é outro sintoma que não deve ser menosprezado na análise da multicolinearidade. Correlações acima de 0,8 é motivo suficiente para o profissional já ficar alerta quanto a precisão do seu modelo.
- 3 – Análises gráficas dos resíduos constituem-se em uma boa sinalização inicial.

2.3 A Análise da Presença da Micronumerosidade frente a Multicolinearidade

A análise da Multicolinearidade é uma etapa importante na verificação do pressuposto VI na análise da Regressão Linear Múltipla, por se constituir como um sério problema de ajuste em modelo RLM de avaliação, que causa fortes impactos na estimativa dos respectivos parâmetros.

Pode ser diagnosticada principalmente pela análise da diagonal da Matriz de Covariâncias definida por $(X'X)^{-1}$, procedimento que permite calcular o chamado Fator de Inflação da Variância (VIF).

Os principais sinais da existência de Multicolinearidade durante o ajuste de um modelo RLM, em princípio, são os seguintes:

$\hat{\beta}_i$	O iésimo parâmetro apresenta sinal de crescimento oposto ao esperado e definido para o modelo.
$\hat{\beta}_i$	O iésimo parâmetro apresenta mudança significativa quando outras variáveis independentes são adicionadas ou excluídas.
$\hat{\beta}_i$	O iésimo parâmetro não é significativo para uma variável independente X_i esperada e definida como importante para a regressão.
Sinais da presença de Multicolinearidade no modelo RLM	

As principais consequências que podem ocorrer quando o pressuposto da Multicolinearidade é violado, pode estar presente quando se percebe que ainda que sejam os melhores estimadores lineares não viesados, isto é, não viciados, os estimadores de MQO têm grandes variâncias e covariâncias, tornando difícil uma estimação precisa.

O que isto significa:

Consequência 1 - As variâncias e covariâncias da RLM são calculadas através das seguintes formulações algébricas:

$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}$	$\text{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)}$	Eq.2.4
Variâncias dos Parâmetros		
$\text{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{(1 - r_{23}^2)\sqrt{\sum x_{2i}^2 \sum x_{3i}^2}}$		Eq.2.5
Covariância entre os Parâmetros		

Observe na formulação das variâncias que quando se tem uma correlação exata entre as variáveis X_2 e X_3 , o Coeficiente de Determinação r_{23}^2 será igual a 1 anulando o denominador nas fórmulas do cálculo das variâncias, fazendo com que as variâncias dos seus parâmetros, no caso limite, vá ao infinito. Da mesma forma ocorre com a covariância entre os parâmetros das variáveis independentes X_2 e X_3 .

O desvio padrão presente nas fórmulas refere-se aos resíduos u_i presentes na regressão entre as duas variáveis independentes X_2 e X_3 dado pelas seguintes expressões:

$X_2 = E(X_2 X_3) + u_i \quad u_i = X_2 - E(X_2 X_3)$	Eq.2.6
Origem dos erros entre variáveis independentes X_2 e X_3	

A partir do cálculo dos erros originados na regressão entre as variáveis em que se pretende examinar a presença da colinearidade, obtém-se a Variância σ^2 constante ou homocedástica desses erros u_i para aplicação nas expressões das Variâncias e Covariâncias dos parâmetros em análise. Daí tem-se:

$\text{var}(u_i X_i) = \sigma_i^2$	Eq.2.7
$\sigma^2 = \sum_{i=1}^N \frac{(u_i - \mu)^2}{N}$	Eq.2.8
Representação simbólica da Variância dos resíduos dos parâmetros de X_3 e X_2 σ^2	

Ao contrário, quando o Coeficiente de Determinação entre as variáveis é 0, inexistente qualquer possibilidade de Colinearidade em uma análise meramente teórica, visto a impossibilidade de tal ocorrência.

Na prática, esses extremos não ocorrem ou raramente ocorrerão dois entre parâmetros de uma RLM formulada com dados aleatórios, mas nada impede que a correlação entre as variáveis independentes alcance valores altos e próximos de 1, comprometendo a estimativa pela RLM com a presença desses regressores simultaneamente, ou no caso contrário valores muito próximo de 0, sinalizando sua pouca importância no modelo.

Consequência 2 - Devido à consequência 1, os intervalos de confiança tendem a ser muito mais amplos, levando à aceitação imediata da “hipótese nula igual a zero” (isto é, o verdadeiro coeficiente populacional igual a zero).

Essa constatação pode ser verificada através da formulação usualmente empregada para o cálculo do intervalo de confiança, como é mostrada abaixo para um intervalo de confiança de 95%, habitualmente adotado em avaliações de imóveis.

$\hat{\beta}_2 \pm 1,96 \sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$	Eq.2.9
O coeficiente t = 1,96 é válido apenas para Nível de Confiança IC = 95%	

Mas de acordo com a narrativa desta consequência, os intervalos de confiança tendem a ser muito amplos narrativa que pode ser demonstrada pela análise do Fator de Inflação da Variância (FIV) dado pela relação seguinte:

$FIV = \frac{1}{(1 - r_{23}^2)}$	Eq.2.10
FIV – Fator de inflação da Variância	

Examinando-se esta formulação, observa-se que quando a Correlação r_{23} entre os parâmetros das variáveis X_2 e X_3 tende ao valor máximo 1, o FIV tende ao infinito e no caso inverso quando esta mesma Correlação tende a 0 o FIV tende a 1 e o Intervalo de Confiança do parâmetro assume o valor dado pela equação 2.9 acima.

Com isso em mente, consegue-se estabelecer, pelo menos para valores mais representativos para r_{23} uma sequência de fatores em função do FIV que permite estabelecer a amplitude do Intervalo de Confiança dos parâmetros, tornando possível que o analista se decida por qual nível de precisão prefere adotar para a sua estimativa.

Assim se obtém os fatores que aplicados à equação 2.9 estabelecerão os limites do Intervalo de Confiança desejável, conforme demonstrado na Tabela abaixo:

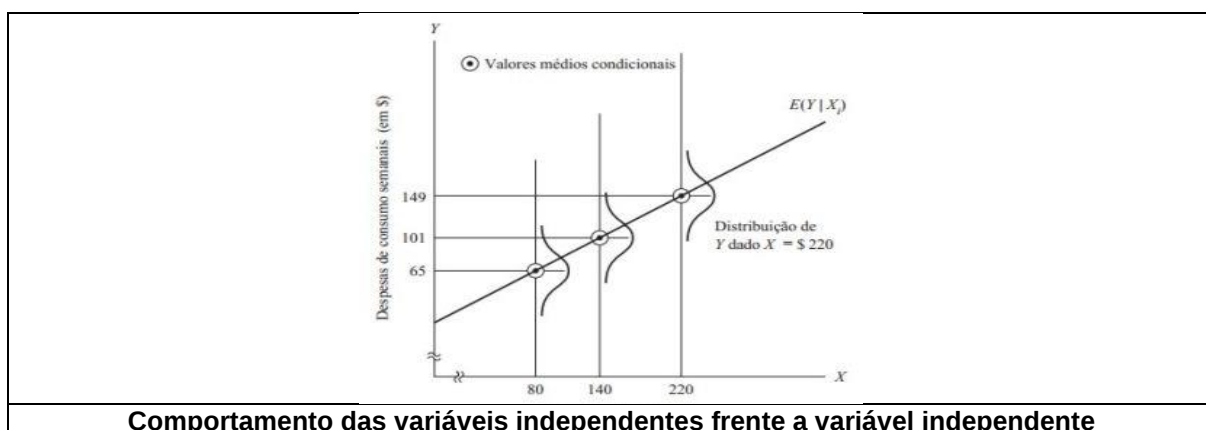
r_{23}	FIV	Tolerância ao erro	Cálculo do IC do Parâmetro	
0,000	1,00	0%	$\hat{\beta}_2 \pm 1,96 \sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$	1,00
0,500	1,33	25%		1,33
0,600	1,56	36%		1,56
0,700	1,96	49%		1,96
0,800	2,78	65%		2,78
0,900	5,76	83%		5,76
0,950	10,26	90%		10,26
0,990	50,25	98%		50,25
0,995	100,00	99%		100,00
0,999	500,00	100%		500,00

Pela tabela anterior, é perceptível o quanto a Correlação entre duas variáveis independentes influencia na estimativa da pelo parâmetro envolvido, notável pelo crescimento acelerado do Intervalo de Confiança do seu respectivo parâmetro estimativo, expondo-o a variável a imprecisões importantes em sua estimativa regressiva.

A análise desta consequência conclui que cabe ao analista a decisão pelo grau de precisão da estimativa regredida por seu modelo conforme mostra a tabela acima expondo os níveis percentuais de tolerância ao erro a que se expõe por cada escolha.

Consequência 3 - Também, devido à consequência 1, a razão t de um ou mais coeficientes tende a ser estatisticamente insignificante.

No relacionamento entre duas variáveis quaisquer existe uma distribuição de valores para cada variável X_3 em relação à variável X_2 como mostra a figura genérica abaixo, onde X_3 estaria no eixo de ordenadas Y, e X_2 no eixo de abcissas X.



No caso em que se estuda a correlação entre duas variáveis independentes com a finalidade de especificar ou não a existência de multicolinearidade em uma RLM, faz-se necessário testar a hipótese em que um dos parâmetros não seja nulo.

Consequência 4 - Embora a razão t de um ou mais coeficientes seja estatisticamente insignificante, R^2 , a medida geral da qualidade do ajustamento, pode ser muito alto.

Para que essa consequência seja verificada, é necessário empreender o teste de hipóteses ou restrições a fim de confirmar ou não a possibilidade dos parâmetros em exame serem iguais ou não. Ver (Gujarati & Porter, 5ª Ed, 2011).

$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2 \text{cov}(\hat{\beta}_3, \hat{\beta}_4)}}$	Eq.2.11
Razão t entre os parâmetros em análise de Multicolinearidade	

Consequência 5 - Os estimadores de MQO e seus erros padrão podem ser sensíveis a pequenas alterações nos dados.
(Gujarati & Porter, 5ª Ed, 2011)

3 O ESTUDO DE CASO REAL

3.1 Discussão do Modelo Estatístico

Na sequência, será apresentado um caso prático de avaliação de terreno, utilizando-se um mesmo modelo estatístico com existência de micronumerosidade. O software utilizado foi o SisDea.

Foi analisada uma amostra que, após tratamento estatístico, resultou na consideração de 46 elementos amostrais efetivos.

REF	ÁREA	VIAS	POLO	TOPO	V UNIT
1	300,00	0	1	0,85	316,00
2	500,00	1	1	0,90	500,00
3	300,00	1	1	0,90	500,00
4	252,00	0	0	0,63	150,79
5	255,00	1	0	1,00	376,03
6	280,00	1	0	0,80	210,71
7	252,00	0	0	0,75	178,57
8	300,00	1	0	0,65	130,00
9	300,00	1	0	0,75	183,33
10	300,00	1	0	0,70	150,00
11	252,00	0	0	0,80	178,57
12	300,00	0	0	0,65	116,66
13	300,00	0	0	0,70	160,00
14	300,00	0	0	0,60	83,33
15	300,00	0	0	0,70	116,67
16	300,00	0	0	0,85	266,66
17	300,00	1	0	0,80	216,66
18	300,00	0	0	0,65	166,33
19	316,00	0	0	0,75	221,52
20	386,00	0	0	1,00	333,19
21	387,00	0	0	0,60	80,10
22	250,00	1	0	0,85	232,00
23	600,00	0	0	0,60	66,67

REF	ÁREA	VIAS	POLO	TOPO	V UNIT
24	225,00	0	0	0,90	253,33
25	780,00	0	0	0,75	102,56
26	225,00	1	0	0,70	200,00
27	200,00	0	0	0,60	100,00
28	1.020,00	0	0	0,80	176,47
29	1.100,00	0	0	0,70	86,36
30	1.200,00	1	0	0,90	329,16
31	1.344,00	0	0	0,80	223,21
32	186,00	1	0	0,90	338,79
33	168,00	0	0	0,80	148,80
34	125,00	0	0	0,70	168,00
35	3.540,00	1	0	0,90	268,36
36	101,50	0	0	1,00	418,40
37	5.000,00	1	0	0,80	150,00
38	5.176,00	1	0	0,80	202,85
39	300,00	1	0	0,80	219,17
40	100,00	0	0	0,60	120,00
41	12.000,00	0	0	0,85	125,00
42	13.228,00	0	0	0,60	80,00
43	16.460,44	0	0	0,60	40,95
44	20.000,00	0	0	0,70	60,00
45	32.000,00	0	0	0,70	55,00
46	34.000,00	1	0	1,00	200,00

Amostra de dados de mercado

Segundo o teste t de Student para o cálculo da significância para rejeição da hipótese nula aplicada em 4 variáveis independentes, estas se mostraram significativas com limite máximo de 10%, máximo admitido para o grau III de fundamentação, segundo a ABNT NBR 14653-2:2011.

Súdea Windows - [100] - COBRAP 2019 - Rev Radula

Modelo Editor Exibir Variáveis Dados Ferramentas Gráficos

Novo Abrir Salvar Salvar como Propriedades de Impressão Configuração de Impressão Visualização de Impressão

Modelo	Transf.	Elastic...	t Calculado	Sig.%	Coef. R ₁	Coef. R ₂	Média	Mínimo	Máximo
ÁREA	x/y	-12,09%	-7,74	0,01%	-0,030917		3.387,15	100,00	34.000,00
VIAS	x	19,72%	2,86	0,67%	1,126274		0,37	0,00	1,00
POLO	x	84,81%	5,84	0,01%	4,298254		0,07	0,00	1,00
TOPO	x	15,11%	13,36	0,01%	13,809887		0,77	0,60	1,00
V UNIT	y/y		9,04	0,01%	5,587703		195,66	40,95	500,00

Nível de significância dos parâmetros dos regressores inferior a 5% conferindo confiabilidade estatística aos mesmos

Resultados RIL

Dados e Variáveis

- Dados: 46
- Dados Considerados: 46
- Variáveis: 5
- Variáveis Consideradas: 5

Coefficientes

- Constante: 4 - 0,9549 / 0,9817
- Determinação: 4 - 0,9118 / 0,9248
- R² Ajustado: 4 - 0,9032 / 0,9175

Testes de Hipóteses

- F Calculado: 106
- Significância Modelo: 0,01
- D Calculado: 0,01
- Durbin Watson: 0,91

Normalidade dos Resíduos

- 1 e +1 desvios padrões: 71%
- 1,64 e +1,64 desvios padrões: 89%
- 1,96 e +1,96 desvios padrões: 97%

Diversos

- Desvio Padrão: 1,1832
- Outliers do Modelo: 0 (0,00%)
- Método de Cálculo: Geral

Estadística F de Snedecor alta indicando confiabilidade do modelo ao nível de 0,01%

Normalidade dos resíduos aceitáveis para a quantidade de variáveis consideradas

Aceitação inicial do modelo em exame

A seguir são apresentados os principais resultados estatísticos encontrados e os respectivos comentários relativos à aceitação de cada um.

3.2 Metodologias Utilizadas

O Método Comparativo Direto de Dados de Mercado foi adotado para a análise avaliatória por permitir que as diversas características dos imóveis semelhantes disponíveis com preços conhecidos e características conhecidas conduzam a determinação do valor procurado.

Neste método, os atributos do imóvel avaliando, portanto, necessários para comparação necessários precisam estar presentes nos dados pesquisados e possam ser trabalhados por meio de técnicas de inferência estatística com o auxílio de um modelo matemático de regressão linear múltipla.

Entre outras peculiaridades, ainda é necessário considerar que os dados coletados devem ser comparáveis entre si e com o avaliando, que estejam sujeitos as mesmas condições de venda, que sejam recolhidos em período compatível com a data de referência da avaliação, que sejam obtidos em fontes com qualidade e credibilidade e finalmente que não estejam sujeitos a influências de fatores externos e subjetivos. (IST / DECivil / Pedro Gameiro Henriques)

Assim definida a metodologia, no caso presente, foram estabelecidas as seguintes premissas para o modelo, conforme segue:

Variável	Transfor...	Restrição	I	D
ÁREA	Auto	(-)	☒	☐
VIAS	Auto	(+)	☒	☐
POLO	Auto	(+)	☒	☐
TOPO	Auto	(+)	☒	☐
V UNIT	Auto	(+)	☒	☐

Comportamento das variáveis quanto a valorização dos imóveis no mercado pesquisado

Opções

Tipo	Todas as regressões
Transformações	1/x; Ln(x); x ² ; x ^{1/2}
Seleção do Modelo	Estimativa
Coefficientes	Mínimos Quadrados
Número de modelos	100
Classificação	R2

Variável Dependente

Categoria	Unitário
Variável base	ÁREA

Premissas atribuídas ao comportamento das variáveis

3.3 Dados de Mercado Pesquisados

A pesquisa de mercado resultante e de acordo com a metodologia aplicada, anteriormente descrita, resultou na seguinte base de dados:

REF	ÁREA	VIAS	POLO	TOPO	V UNIT	REF	ÁREA	VIAS	POLO	TOPO	V UNIT
1	300,00	0	1	0,85	316,00	24	225,00	0	0	0,90	253,33
2	500,00	1	1	0,90	500,00	25	780,00	0	0	0,75	102,56
3	300,00	1	1	0,90	500,00	26	225,00	1	0	0,70	200,00
4	252,00	0	0	0,63	150,79	27	200,00	0	0	0,60	100,00
5	255,00	1	0	1,00	376,03	28	1.020,00	0	0	0,80	176,47
6	280,00	1	0	0,80	210,71	29	1.100,00	0	0	0,70	86,36
7	252,00	0	0	0,75	178,57	30	1.200,00	1	0	0,90	329,16
8	300,00	1	0	0,65	130,00	31	1.344,00	0	0	0,80	223,21
9	300,00	1	0	0,75	183,33	32	186,00	1	0	0,90	338,79
10	300,00	1	0	0,70	150,00	33	168,00	0	0	0,80	148,80
11	252,00	0	0	0,80	178,57	34	125,00	0	0	0,70	168,00
12	300,00	0	0	0,65	116,66	35	3.540,00	1	0	0,90	268,36
13	300,00	0	0	0,70	160,00	36	101,50	0	0	1,00	418,40
14	300,00	0	0	0,60	83,33	37	5.000,00	1	0	0,80	150,00
15	300,00	0	0	0,70	116,67	38	5.176,00	1	0	0,80	202,85
16	300,00	0	0	0,85	266,66	39	300,00	1	0	0,80	219,17
17	300,00	1	0	0,80	216,66	40	100,00	0	0	0,60	120,00
18	300,00	0	0	0,65	166,33	41	12.000,00	0	0	0,85	125,00
19	316,00	0	0	0,75	221,52	42	13.228,00	0	0	0,60	80,00
20	386,00	0	0	1,00	333,19	43	16.460,44	0	0	0,60	40,95
21	387,00	0	0	0,60	80,10	44	20.000,00	0	0	0,70	60,00
22	250,00	1	0	0,85	232,00	45	32.000,00	0	0	0,70	55,00
23	600,00	0	0	0,60	66,67	46	34.000,00	1	0	1,00	200,00

Base de dados para a avaliação

3.4 Regressão Linear Múltipla

Após a análise dos dados brutos de mercado e exame dos resultados dos testes de hipóteses fornecidos pelo aplicativo SisDea, o modelo resultante mostrou-se consistente utilizando-se de 4 (quatro) variáveis independentes efetivas mais representativas para a obtenção do resultado desejado nesta análise.

Daí resultou o Modelo de Regressão Linear por tratamento estatístico inferencial representativo da amostra de 46 (quarenta e oito) dados de mercado efetivamente utilizados, cujos resultados primeiros elementares foram:

Coeficiente de Determinação	R^2	0,9118024 / 0,9248063
Coeficiente de Determinação Ajustado	$R_{ajustado}^2$	0,9031977 / 0,9174703
Coeficiente de Correlação	R	0,9548834 / 0,9616685
Coeficientes de relacionamento entre as variáveis		

3.4.1 Coeficiente de Determinação

Medida de ajuste utilizada na inferência estatística e que se pode obter da soma dos quadrados dos resíduos (SQR), sendo, também, a proporção da melhoria que pode ser debitada ao modelo em relação à Soma dos Quadrados Totais (SQT).

O coeficiente de determinação representa o poder de explicação das variáveis independentes, sobre a variável dependente.

Na análise de regressão linear múltipla, o coeficiente de determinação R^2 , na maioria das vezes, aumenta (e nunca diminui) quando é adicionada nova variável independente, exceto quando este estiver perfeitamente correlacionado com as demais variáveis independentes, visto que, ao se acrescentar novas variáveis ao modelo, estas diminuem o SQR.

O modelo adotado responde por $R^2 = 91,18\%$ no Modelo de regressão não linear e por $R^2 = 92,48\%$ no Modelo da Estimativa linear da formação dos preços com base na amostra pesquisada.

3.4.2 Coeficiente de Correlação

O Coeficiente de Correlação R é obtido do Coeficiente de Determinação R^2 e nos fornece uma medida de força de correlação entre as variáveis do modelo estatístico.

Para efeitos de classificação, quanto à intensidade de Correlação entre as variáveis, a literatura fornece os seguintes intervalos para análise: (Avaliação de Bens / AVALIEN / 2Ed / 1993)

Coeficiente de Determinação	Intensidade do Relacionamento
$R = 0$	Nulo
$0 < R \leq 0,30$	Fraco
$0,30 < R \leq 0,60$	Médio
$0,60 < R \leq 0,90$	Forte
$0,90 < R \leq 0,99$	Fortíssimo
$R = 1$	Perfeito
Graus de relacionamento entre variáveis	

O modelo adotado responde por $R = 95,49\%$ no Modelo de Regressão não linear e por $R = 96,17\%$ no Modelo da Estimativa linear da formação dos preços, portanto fortíssimo a partir base amostral pesquisada.

3.5 Análise da Variância

Uma parte importante da análise de variância é a estatística F (Fisher-Snedecor) e sua significância estatística, trata-se de uma medida da qualidade de ajuste do modelo.

A estatística F mede quanto o modelo matemático melhorou na previsão dos valores, comparando-se com o seu nível de não precisão. Testada a hipótese nula da não representatividade do modelo, obteve-se:

A estatística F calculada alcançou o valor 106 correspondente ao Nível de Significância de 0,01% e Nível de Confiabilidade de 99,99%.

3.6 Normalidade dos Resíduos

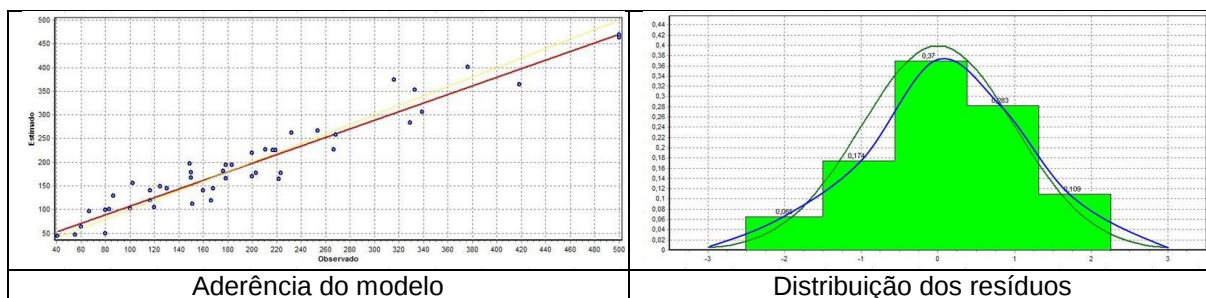
A verificação da normalidade na distribuição dos resíduos é um pressuposto essencial e que não deve ser violado para que os resultados do ajuste do modelo de regressão linear sejam estatisticamente confiáveis.

Na Distribuição Normal de resíduos, a probabilidade dos valores críticos da estatística z se localizar em alguns pontos críticos notáveis ou normalizados deve apresentar resultados únicos e os valores obtidos na Distribuição dos Resíduos do modelo devem ser confrontados com os primeiros para a verificação e aceitação ou não da normalidade dos resíduos.

Com este objetivo o quadro seguinte demonstra esta análise:

Pontos críticos z	Probabilidade Normalizada	Probabilidade Normalizada
1,00 DP	68%	71%
1,64 DP	90%	89%
1,96 DP	95%	97%
Análise da normalidade dos resíduos		

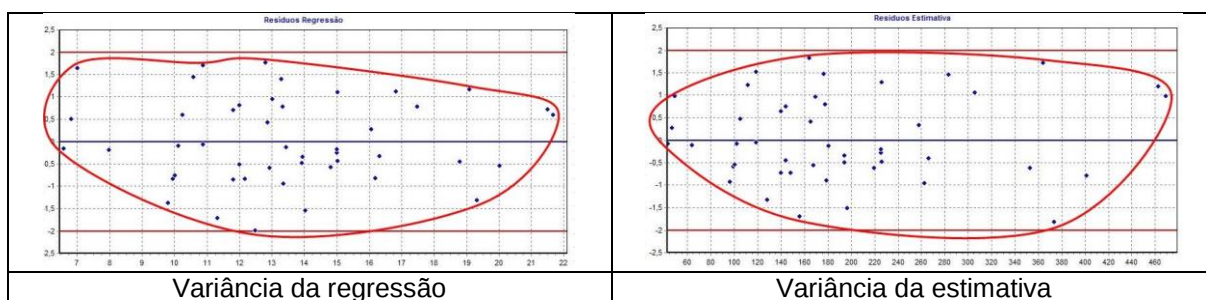
Em consequência, comparando esses percentuais, o software SIsDea disponibiliza gráficos que permitem a verificação da Aderência e Normalidade dos resíduos do modelo:



Conforme os gráficos acima, conclui-se que a distribuição dos resíduos no modelo se aproxima bastante da Curva Normal Padronizada, com suficiente semelhança que permite a aceitação da normalidade dos resíduos para o presente caso, sendo a diferença perceptível atribuída ao número de dados disponibilizados para a análise.

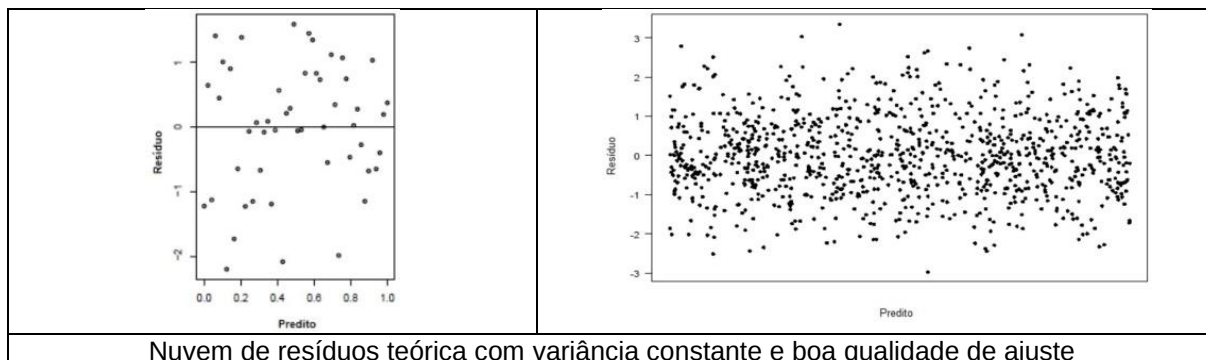
3.7 Outliers

São os resíduos extremos que apresentam distantes da média dois ou mais desvios padrões considerados pontos atípicos capazes de influencia no valor da regressão e/ou estimativa. No presente caso, frente aos elementos efetivamente utilizados, conforme figuras abaixo, não foram constatadas a presença dessas anomalias.

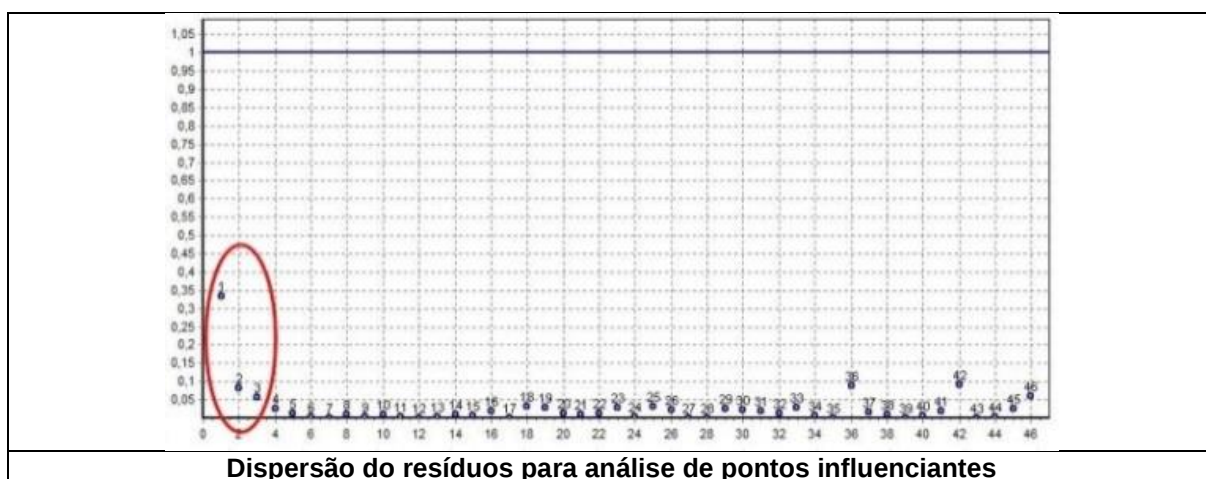


Nas figuras acima é notável a distribuição preponderantemente simétrica em relação ao eixo médio com a distribuição dos resíduos apresentando uma dispersão em nuvem regular, sinalizando visualmente que a variância dos resíduos é constante e em consequência o modelo é homocedástico, como estabelece e impõe um dos pressupostos estatísticos a ser obrigatoriamente observado.

Esta nuvem referenciada na teoria apresenta o seguinte formato, tradicionalmente exposto na teoria.



Concomitantemente, não ocorre no modelo nenhuma presença de pontos influenciantes, condição que pode ser confirmada na figura abaixo que segue com especial atenção para os pontos realçados, que de acordo com a ABNT NBR 14.653-2:2011 seria pontos possíveis de exclusão por conta do conceito de micronumerosidade atualmente normatizado.



A presença desses pontos seria notada, na hipótese em que pela Análise da Distância de Cook, um ou mais pontos estivesse além de $D > 1$ (um) no gráfico da figura acima. Esta estatística não será objeto de discussão neste trabalho, neste momento.

3.8 O Modelo de Regressão

Para o desenvolvimento da presente análise utilizou-se um modelo estatístico baseado na coleta de 46 (quarenta e seis) dados reais de mercado e quatro variáveis que permaneceram como consistentes segundo as hipóteses adotadas a seguir descritas:

Variável	Característica	Crescimento	Variação admitida
Área	Quantitativa	Negativo	Áreas => Min = 252 / Max 34.000
Vias	Qualitativa	Positivo	Via Coletora = 1 / Outras vias = 0
Polo	Dicotômica	Positivo	J América / CEASA = 1 / Outros = 0
Topografia	Quantitativa	Positivo	Percentual de área plana
Pressupostos admitidos para as variáveis			

Processadas essas informações com o auxílio do software customizado para a avaliação de imóveis SisDea, o programa retornou as funções estimativas seguintes, aceitas como confiáveis após a realização e verificação das análises anteriores.

Para a regressão não linearizada foi aceito o modelo seguinte:

$V \text{ UNIT}^{\frac{1}{2}} = + 5,587703393 - 0,03091722077 * \text{ÁREA}^{\frac{1}{2}} + 1,126274216 * \text{VIAS} \\ + 4,29825441 * \text{POLO} + 13,80988711 * \text{TOPO}^2$	Eq.3.1
Função Regressiva não linear	

Para a estimativa linearizada foi aceito o modelo seguinte:

$V \text{ UNIT} = (+ 5,587703393 - 0,03091722077 * \text{ÁREA}^{\frac{1}{2}} + 1,126274216 * \text{VIAS} \\ + 4,29825441 * \text{POLO} + 13,80988711 * \text{TOPO}^2)^2$	Eq.3.2
Função Estimativa linear	

3.9 Micronumerosidade

De acordo com o ABNT NBR 14.653-2/2011, ANEXO A, a micronumerosidade deve ser considerada “no caso de utilização de variáveis dicotômicas e variáveis qualitativas expressas por códigos alocados ou códigos ajustados”.

Entretanto, podemos observar na pesquisa a seguir, que a variável dicotômica **polo**, não atende ao preceito da norma - **para $30 < n \leq 100$, $n_i \geq 10\% n$** . Se observa, que no modelo, só existe 3 (três) dados de mesma característica, em desacordo com o preceito normativo acima.

REF	ÁREA	VIAS	POLO	TOPO	V UNIT
1	300,00	0	1	0,85	316,00
2	500,00	1	1	0,90	500,00
3	300,00	1	1	0,90	500,00
4	252,00	0	0		
5	255,00	1	0		
6	280,00	1	0		
7	252,00	0	0		
8	300,00	1	0	0,65	130,00
9	300,00	1	0	0,75	183,33
10	300,00	1	0	0,70	150,00
11	252,00	0	0	0,80	178,57
12	300,00	0	0	0,65	116,66
13	300,00	0	0	0,70	160,00
14	300,00	0	0	0,60	83,33
15	300,00	0	0	0,70	116,67
16	300,00	0	0	0,85	266,66
17	300,00	1	0	0,80	216,66
18	300,00	0	0	0,65	166,33
19	316,00	0	0	0,75	221,52
20	386,00	0	0	1,00	333,19
21	387,00	0	0	0,60	80,10
22	250,00	1	0	0,85	232,00
23	600,00	0	0	0,60	66,67

A variável POLO com apenas tres valores 1 deve é afetada pela micronumerosidade?

REF	ÁREA	VIAS	POLO	TOPO	V UNIT
24	225,00	0	0	0,90	253,33
25	780,00	0	0	0,75	102,56
26	225,00	1	0	0,70	200,00
27	200,00	0	0	0,60	100,00
28	1.020,00	0	0	0,80	176,47
29	1.100,00	0	0	0,70	86,36
30	1.200,00	1	0	0,90	329,16
31	1.344,00	0	0	0,80	223,21
32	186,00	1	0	0,90	338,79
33	168,00	0	0	0,80	148,80
34	125,00	0	0	0,70	168,00
35	3.540,00	1	0	0,90	268,36
36	101,50	0	0	1,00	418,40
37	5.000,00	1	0	0,80	150,00
38	5.176,00	1	0	0,80	202,85
39	300,00	1	0	0,80	219,17
40	100,00	0	0	0,60	120,00
41	12.000,00	0	0	0,85	125,00
42	13.228,00	0	0	0,60	80,00
43	16.460,44	0	0	0,60	40,95
44	20.000,00	0	0	0,70	60,00
45	32.000,00	0	0	0,70	55,00
46	34.000,00	1	0	1,00	200,00

Dados de mercado pesquisados

Vamos então fazer a verificação através dos testes citados anteriormente, se podemos aceitar a equação, apesar da micronumerosidade.

Razões t “insignificantes”

Os **t** calculados para os parâmetros do modelo mostraram-se maiores que a estatística $t_{90\%} = 1,68$ e estão demonstrados na tabela seguinte:

Variável	Estatística t calculada
Área	-7,74
Vias	2,86
Polo	5,84
Topografia	13,36

Utilizando-se da função PRO.JLIN do EXCEL, os resultados calculados mostram os coeficientes (parâmetros) e seus respectivos erros padrão e comprova-se o relatado anteriormente, conforme o quadro seguinte:

	Coefficientes =	13,8099	4,2983	1,1263	-0,0309	5,5877	
	ep dos coeficientes =	1,0340	0,7364	0,3943	0,0040	0,6179	
	R2 =	0,9118	1,1832				
	F =	105,9663	41,0000				
	SQ	593,4212	57,4010	t dos parâmetros > t tabelado			
	t =	13,3554	5,8372	2,8561	-7,7402	9,0432	t tab = 1,6829
	Resultado da divisão dos coeficientes (parâmetros) pelos erros padrão dos mesmos						Valor P < Nível de significância adotado $\alpha = 10\%$ bicaudal
	valor p =	1,60E-16	7,37E-07	6,70E-03	1,52E-09	2,58E-11	
	Resultado da aplicação da função DIST						
Análise da Regressão no EXCEL / Proj.Lin							

3.9.1 Questões que surgem

Quais as medidas corretivas desenvolver, ao se deparar com as seguintes indagações mais comuns:

- Quando ocorrer casos de micronumerosidade, deve-se eliminar a variável independente que ocasiona tal problema?
R: Nem sempre, a solução dependerá de outros testes.
- Quais são as orientações da norma?
R: Para evitar a micronumerosidade, o número mínimo de dados efetivamente utilizados (n) no modelo deve obedecer aos seguintes critérios, com respeito ao número de variáveis independentes (k):

Quantidade de dados da amostra	Quantidade de dados de mesma característica
$n \geq 3(k + 1)$	-
$n \leq 30$,	$n_i \geq 3$
$30 < n \leq 100$	$n_i \geq 10\% n$
$n > 100$	$n_i \geq 10$
Nota:	n_i refere-se aos casos de utilização de variáveis dicotômicas e variáveis qualitativas expressas por códigos alocados.
Recomendações da ABNT NBR 14.653-2-2011	

- c) O critério adotado para analisar o problema fornece um subsídio seguro para adoção da equação?
R: Se a equação passar pelos critérios propostos (testes) a utilização do modelo é confiável e segura.
- d) Quais os testes de hipóteses (métodos de aceitação dos parâmetros) devem ser utilizados para detectar se a influência da micronumerosidade sobre o modelo é danosa?
R: Os sugeridos seriam os seguintes:
 I) Teste t para validação dos parâmetros.
 II) Verificação de multicolinearidade entre as variáveis independentes.
 III) Mudança de sinais nas relações entre as variáveis dependentes e independentes em desacordo com as premissas iniciais.

3.9.2 Verificação de multicolinearidade

A multicolinearidade é um problema no ajuste do modelo de regressão e pode causar sérios impactos nas estimativas dos parâmetros e degenerações em seu comportamento.

Através das matrizes da colinearidade, isoladas e com influência, no presente estudo verifica-se que a mesma não se manifesta a níveis preocupantes e capazes de restringir o uso da variável independente no modelo.

Variável	ÁREA	VIAS	POLO	TOPO	V UNIT	Variável	ÁREA	VIAS	POLO	TOPO	V UNIT
ÁREA		-2	-11	2	-38	ÁREA		31	47	71	78
VIAS	-2		17	40	47	VIAS	31		23	23	41
POLO	-11	17		25	52	POLO	47	23		55	68
TOPO	2	40	25		81	TOPO	71	23	55		91
V UNIT	-38	47	52	81		V UNIT	78	41	68	91	
Correlações isoladas						Correlações com influência					

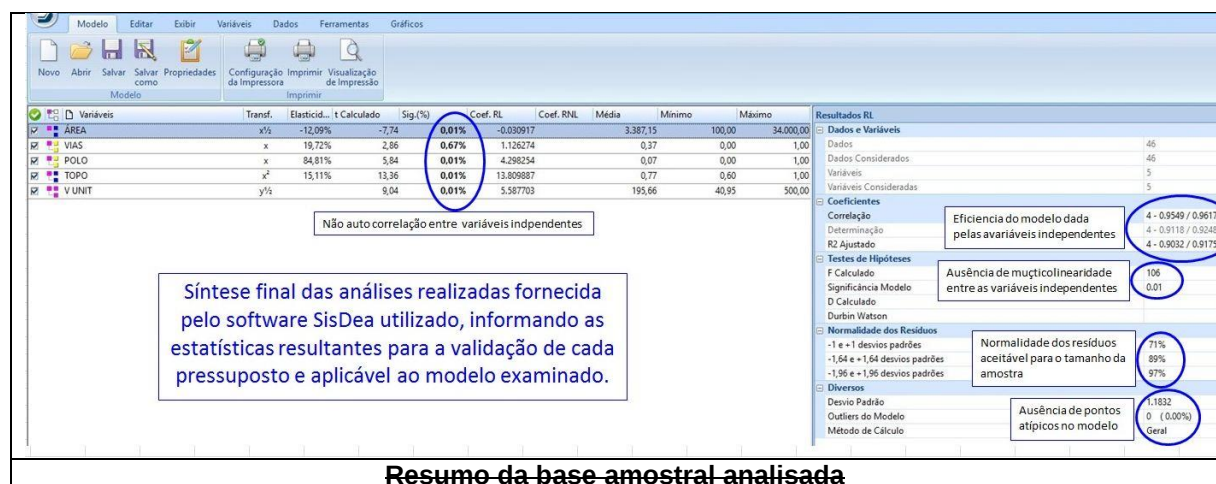
3.9.3 Análise dos sinais das variáveis dependentes x independentes

Os crescimentos das variáveis independentes dados pelos sinais a priori esperados para cada uma, em suas relações com a variável dependente de acordo com os quadros acima mantiveram a coerência com a premissa inicialmente estabelecida, sinalizando que a multicolinearidade tem a sua presença afastada por esta análise.

Em resumo tem-se o seguinte:

Variável	Crescimento pressuposto inicial	Correlação final calculada
Área	Negativo	-
Vias	Positivo	+2%
Polo	Positivo	+17%
Topografia	Positivo	+25%
Comparativo do crescimento Pressuposto x Calculado para as variáveis independentes		

De fato o mercado imobiliário admite tradicionalmente que as áreas urbanas são mais valorizadas quando mais próximas de vias coletoras de polos valorizantes e com topografias planas ou percentualmente mais favoráveis confirmando o pressuposto inicial estabelecido, ou seja, vias coletoras, polos mais atrativos e topografia mais próxima da plana são melhores valorizadas unitariamente e este resultado confirma a premissa inicial estabelecida no item 3.8 anterior.



4 RECOMENDAÇÕES

Do estudo de caso prático apresentado, conclui-se pela importância que assume a análise da micronumerosidade em que a simples recomendação aparentemente subjetiva e constante na NBR 14.653-2:2011 / Anexo A pode e deve ser melhor especificada, considerando a existência de testes complementares para validação das hipóteses, que aplicados podem evitar a rejeição de dados em um modelo estatístico inferencial para avaliação imobiliária e amparado em procedimentos técnicos reconhecidos, superando análise da micronumerosidade não atendida de acordo com os atuais preceitos normativos recomendados.

A literatura disponível comprova que modelos de avaliação na presença de micronumerosidade real, limita consideravelmente a utilização dos mesmos e impõe cuidados especiais quanto à verificação das suas características convenientemente descritas nos dados amostrais a utilizar.

No caso da equação estudada, os exames gráficos e demais testes levaram a concluir que de acordo com as atuais recomendações da ABNT NBR 14.653-2:2011, a micronumerosidade presente na amostra em exame, independente dos conceitos normativos vigentes, pode ser aceita sem perda de qualidade e conteúdo para o processo avaliatório.

Desta forma, a evidência dos resultados obtidos nas projeções dos valores para o estimador pontual, o intervalo de confiança e o campo de arbítrio pode ser utilizada para a determinação de valores de mercado.

5 CONCLUSÕES

Daí, há a considerar que o conceito assumido pela ABNT NBR 14.653-1:2011 mostra-se controvertido, à medida que para amostras acima de 100 (cem) dados pesquisados, a micronumerosidade normativa para quantidades abaixo de 10% dos dados não é **“válida”**, ao passo que vale o contrário para o caso de amostras menores.

Note-se que com quantidades de dados entre 30 (trinta) e 100 (cem) na ABNT NBR 14.653-2:2011 exige o limite de uma quantidade mínima representativa na variável de 10% dos dados utilizados, imposto como condição para validação normativa do modelo.

Para o caso de amostras maiores, como por exemplo, no caso de 200 dados, aceitar-se-ia uma quantidade acima de tão somente 5% do total dos dados utilizados (10 conforme a norma vigente = 5% de 200 dados) e no caso mais extremo de 1000 dados essa quantidade seria reduzida até 1% (10 conforme a norma vigente = 1% de 1.000 dados), conflitando com a regra para as amostras menores impositiva de um número mínimo representativo aceitável equivalente a 10% dos dados úteis.

Portanto, é enfaticamente recomendável a adoção na ABNT NBR 14,653-2:2011 de testes para comprovação do efeito negativo da micronumerosidade eventualmente presente, independente de um número mínimo representativo, sob o amparo de recomendações normativas tecnicamente comprováveis e justificáveis a fim de validar o modelo em exame.

Note-se que uma iniciativa dessa natureza poderia inclusive se estender, eventualmente, até a alteração no grau de fundamentação e precisão atualmente recomendados e adotados.

Resta que assim procedendo, o atual conceito normativo restritivo a análise da micronumerosidade nos dados amostrais, estaria fundamentada em princípios técnicos largamente aceitos na literatura disponível.

BIBLIOGRAFIA

ABNT NBR 14.653-2:2019 - Avaliação de bens – Parte 1: Procedimentos Gerais

ABNT NBR 14.653-2:2011 - Avaliação de bens – Parte 2: Imóveis Urbanos

Avaliações de Bens por Estatística Inferencial e Regressões Múltiplas – Teoria e Aplicações – 2ª Ed. / 1993 - AVALIEN – Engenharia de Avaliações Ltda.

Avaliação de Bens – Princípios Básicos e Aplicações – Nasser Junior Radegaz – 2ª Ed / 2013 - Editora LEUD

Econometria Básica – Damodar N. Gujarati / Dawn C. Porter / 5ª Ed / 2011 – Mc Graw Hill / AMGH Editora