Aplicação do Teste de Farrar-Glauber para Análise de Multicolinearidade Em Regressões Lineares

Carlos Augusto Zilli; Murilo Damian Ribeiro; Norberto Hochheim

APLICAÇÃO DO TESTE DE FARRAR-GLAUBER PARA ANÁLISE DE MULTICOLINEARIDADE EM REGRESSÕES LINEARES

RESUMO

Conforme a NBR 14.653-2, ao se usar modelos de regressão linear em avaliação de imóveis, há a necessidade de se observar seus pressupostos básicos, dentre eles o de não-multicolinearidade, com o objetivo de se obter avaliações não tendenciosas, eficientes e consistentes. A literatura estatística aponta a multicolinearidade como um dos problemas mais incômodos e intratáveis em toda a análise de regressão. A multicolinearidade pode causar diversos efeitos indesejáveis nos coeficientes estimados em uma regressão linear. É importante que se saiba como detectar sua presença e algumas formas de contorna-la. Dependendo das variáveis explicativas consideradas no modelo de regressão, ocorre o aparecimento de multicolinearidade em um grau de gravidade que varia de modelo para modelo. O teste de Farrar-Glauber é utilizado para detectar a presença, a gravidade, localizar o regressor e o padrão da multicolinearidade em um modelo de regressão. Esse estudo tem como objetivo aplicar o teste de Farrar-Glauber em uma amostra com 50 dados de mercado coletados em Florianópolis (SC) afim de analisar a presença de multicolinearidade. Contatou-se que é possível identificar não apenas a existência de multicolinearidade, como também sua gravidade, quais são as variáveis afetadas por ela e quais as variáveis que causam a multicolinearidade.

Palavras-Chave: *Multicolinearidade; Teste de Farrar-Glauber; Avaliação de Imóveis.*

1 INTRODUÇÃO

Segundo o engenheiro Lélio Moreira (2001) já se foi o tempo em que o "olho clínico" do avaliador, ou seja, a sua experiência, era a melhor técnica admitida para a avaliação de um bem; não há dúvida que a experiência do avaliador muito influi para uma boa aplicação das técnicas hoje conhecidas, mas os métodos científicos desenvolvidos atualmente fazem com que o avaliador, cada vez mais, se paute por dados estatísticos, tecnicamente analisados, do que por sentimento pessoal.

Entre os métodos utilizados para a avaliação de imóveis previstos na NBR 14.653-2, o mais comum é o método comparativo direto de dados de mercado, que além de apresentar uma ótima estimativa dos valores dos bens envolvidos, oferecem qualidade e transparência ao processo avaliativo do bem.

O método comparativo direto de dados de mercado comumente usa de inferência estatística com modelos de regressão linear, técnica bastante conhecida por profissionais da área de avaliação de imóveis e pela academia em geral.

Com base em uma amostra de dados de mercado, os parâmetros populacionais são estimados por inferência estatística. Entretanto, quando se adota técnicas de análise por regressão linear, certos pressupostos devem ser atendidos, dentre eles o pressuposto de não-multicolinearidade dos regressores do modelo.

Segundo a NBR 14653-2 (2011), uma forte dependência linear entre duas ou mais variáveis independentes provoca degenerações no modelo e limita a sua utilização. As variâncias das estimativas dos parâmetros podem ser muito grandes e acarretar a aceitação da hipótese nula e a eliminação de variáveis fundamentais.

O problema de multicolinearidade pode aparecer em diversos níveis de gravidade, indo de fraca a forte, e em diversas situações práticas, estando, segundo

Gujarati e Porter (2011), quase sempre presente. Dificilmente se encontrarão variáveis explicativas perfeitamente ortogonais, fazendo com que sempre haja ao menos um pequeno grau de linearização entre elas.

Se a multicolinearidade é algo sempre presente, deve-se avaliar a gravidade dela, uma vez que as consequências podem levar à estimativas de parâmetros instáveis, testes de hipóteses não confiáveis e erros de especificação, fazendo com que haja a degeneração do modelo e a impossibilidade de sua utilização.

Ponderando que a literatura estatística considera a multicolinearidade como um dos problemas mais indesejáveis ao se realizar uma análise de regressão, várias são as técnicas e ferramentas disponíveis para sua detecção, dentre elas o teste de Farrar-Glauber, que se apoia no teste qui-quadrado, no teste F e no teste t.

O teste de Farrar e Glauber (1967) é útil para detectar a presença, localizar, classificar e verificar o padrão da multicolinearidade. Busca-se nesse estudo aplicar o teste criado por Farrar-Glauber em uma amostra de 50 dados de oferta de imóveis em Florianópolis (SC) afim de se verificar a presença, localizar e classificar a multicolinearidade, propondo na sequência algumas formas de contorna-la.

2 REVISÃO DA LITERATURA

2.1. REGRESSÃO LINEAR MÚLTIPLA

É comum em engenharia de avaliações a utilização de modelos de regressão linear múltipla, caso no qual Y é uma função linear de duas ou mais variáveis independentes X_1, X_2, X_3, ..., X_m. Para esse caso, o modelo de regressão seria

$$Y_i = f(X_1, X_2, X_3, ..., X_m) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_m X_{mi} + e_i \tag{2.0}$$

Sendo Y a variável dependente, X_1, X_2, X_3, ..., X_m as variáveis explanatórias, ou regressoras, $\boldsymbol{e}$ o termo de erro estocástico e $\boldsymbol{i}$, o indicador da i-ésima observação.

Na equação (2.0), β_0 é o intercepto. Como de costume, ele dá o efeito médio sobre Y de todas as variáveis excluídas do modelo, embora sua interpretação mecânica seja do valor médio de Y quando X_1, X_2, X_3, ..., X_m são iguais a zero. Os coeficientes β_1, β_2, β_3, ..., β_m são denominados coeficientes parciais de regressão.

Segundo Gujarati e Porter (2011), coeficientes parciais de regressão ou coeficientes parciais angulares medem as variações que ocorrem no modelo quando determinados parâmetros se mantém constantes. Temos que β_1 mede a variação no valor médio de Y, E(Y), por unidade de variação em X_1, mantendo-se o valor de X_2, X_3, ..., X_m constantes. Em outras palavras, ele nos dá o efeito "direto" ou "líquido" de uma unidade de variação em X_1 sobre o valor médio de Y, excluídos os efeitos que X_2, X_3, ..., X_m possam ter sobre a média da variável explicada Y.

Esse modelo de regressão é particularmente útil quando desejamos fazer o ajuste de dados experimentais em que a variável explicada é uma função de duas ou mais variáveis explicativas. Nesse caso, a reta de regressão se torna um plano ou hiperplano, dependendo da quantidade de variáveis explicativas do modelo.

Quanto às hipóteses básicas, tem-se a normalidade, não autocorretação e homocedasticidade dos erros, e mais uma: não deve existir nenhuma relação linear exata entre quaisquer variáveis independentes (DANTAS, 2012).

2.1.1. Pressuposto de Inexistência de Relação Linear Exata

Conforme Gujarati e Porter (2011), a hipótese de ausência de relação linear exata entre X_1, X_2, X_3, ..., X_m, é conhecida tecnicamente como ausência de colinearidade ou ausência de multicolinearidade, se estiverem envolvidas mais de uma relação linear exata. Informalmente, a ausência de colinearidade significa que nenhum dos regressores pode ser expresso como uma combinação linear exata dos demais regressores do modelo. Formalmente, a ausência de colinearidade significa que não existe um conjunto de números λ_1, λ_2, λ_3, ..., λ_m, que não sejam, considerando os regressores dois a dois, iguais a zero. Por exemplo, se estivéssemos analisando a relação entre os regressores X_1 e X_2:

$$\lambda_1 X_{1i} + \lambda_2 X_{2i} = 0 \tag{2.1}$$

Se essa relação linear exata existe, diz-se que X_1 e X_2 são colineares ou linearmente dependentes. Por outro lado, se a equação (2.1) só é verdadeira quando $\lambda_1 = \lambda_2 = 0$, diz-se que X_1 e X_2 são linearmente independentes.

2.2. SOBRE A MULTICOLINEARIDADE

Segundo Fonseca (2003), em modelos de regressão com duas ou mais variáveis explicativas é usual que tais variáveis apresentem algum tipo de interdependência e a essa relação de interdependência chamamos multicolinearidade. A presença de multicolinearidade independe da existência de relação de dependência entre a variável dependente e termos independentes (FARRAR E GLAUBER, 1967).

Gujarati e Porter (2011) nos ensinam que, originalmente, a expressão multicolinearidade significava a existência de uma relação linear “perfeita” ou exata entre algumas ou todas as variáveis explanatórias do modelo de regressão. Hoje, no entanto, o termo multicolinearidade é usado em um sentido mais amplo, para incluir o caso de multicolinearidade perfeita, bem como o caso em que as variáveis X do modelo de regressão estão fortemente intercorrelacionadas, mas não perfeitamente.

O aparecimento da multicolinearidade pode ocorrer por diversos motivos, entre eles o método como se deu a coleta dos dados empregados no modelo; as restrições ao modelo ou à população ao qual pertence essa amostra; à especificação do modelo ou a utilização de um modelo sobredeterminado.

2.2.1. Consequências da Multicolinearidade

Segundo Hoffmann (2016 p. 184), as principais consequências desse fato são:
1) As variâncias e covariâncias das estimativas dos parâmetros serão muito elevadas, isto é, as estimativas obtidas podem ter erros muito grandes e esses erros podem estar altamente correlacionados entre si. A baixa precisão das estimativas torna difícil distinguir as influências das diversas variáveis explanatórias;
2) Pode-se eliminar variáveis da análise porque os coeficientes não se mostraram estatisticamente diferentes de zero; essas variáveis podem, na realidade, ser importantes e a amostra disponível é que não permite detectar sua influência;
3) As estimativas dos coeficientes variam muito de amostra para amostra. A adição de algumas observações à amostra pode alterar muito o valor da estimativa obtida.

E o que acontece se a multicolinearidade for severa? Primeiramente conforme observado por Farrar e Glauber (1967) em um artigo sobre o assunto, se o conjunto de variáveis independentes for totalmente interdependente os coeficientes da regressão não poderão ser estimados uma vez que a matriz resultante da multiplicação da matriz transposta das variáveis independentes pela matriz das variáveis independentes será singular e não será possível a inversão dessa matriz necessária para cálculo dos coeficientes da regressão, ou seja não existe $(X^TX)^{-1}$. Nesse caso a multicolinearidade é, obviamente, severa e o modelo deve ser revisto.

2.2.2. Detectando a Multicolinearidade

Há alguns indicadores da presença de multicolinearidade que são calculados pela maioria dos softwares que fazem regressão linear múltipla (NEWBOLD, 1994), entretanto é importante que se fique atento a elementos que indicam sua presença.

Segundo Montgomery *et al* (*apud* Silva, 2008), alguns indicativos da presença de multicolinearidade são mostrados abaixo:
1) valores altos do coeficiente de correlação;
2) grandes alterações nas estimativas dos coeficientes de regressão, quando uma variável independente for adicionada ou retirada do modelo, ou quando uma observação for alterada ou eliminada;
3) a rejeição da hipótese H_0: $\beta_1 = \beta_2 = ... = \beta_m = 0$, por meio da realização do teste *F*, mas nenhuma rejeição das hipóteses H_0: $\beta_i = 0$, $i = 1, 2, 3, ..., m$, por meio da realização dos testes *t*, sobre os coeficientes individuais da regressão;
4) obtenção de estimativas para os coeficientes de regressão com sinais algébricos contrários àqueles que seriam esperados a partir de conhecimentos teóricos disponíveis ou de experiências anteriores sobre o fenômeno estudado, e;
5) obtenção de intervalos de confiança com elevadas amplitudes para os coeficientes de regressão, associados a variáveis independentes importantes.

Segundo Miloca e Conejo (2014), a presença de multicolinearidade pode ser detectada de várias maneiras. Entre as formas conhecidas, dois testes formais mais comumente utilizados são o valor de tolerância (TOL) ou seu inverso, chamado fator de inflação da variância (VIF) definido pelas equações mostradas abaixo:

$$VIF\ (\beta_i) = \frac{1}{1 - r_i^2} \qquad\qquad TOL\ (\beta_i) = \frac{1}{VIF\ (\beta_i)} = 1 - r_i^2 \tag{2.2}$$

É uma medida do grau em que cada variável independente é explicada pelas demais variáveis independentes. Quanto maior for o fator de inflação da variância, mais severa será a multicolinearidade. Segundo Gujarati e Porter (2011), quanto maior for o valor de VIF_i, mais "problemática" ou colinear será a variável X_i. Os autores em geral dizem que, se o VIF_i de uma variável for maior que 10, o que acontecerá se R_i^2 for maior que 0,90, essa variável será tida como altamente colinear.

Conforme Miloca e Conejo (2014), outros autores sugerem que os fatores de inflação da variância não devem exceder 4 ou 5, isso dependerá do conhecimento teórico do pesquisador sobre o assunto que está sendo estudado.´

Nesse caso, pode-se perceber que o TOL_i também pode ser utilizado como teste para detecção de multicolinearidade, haja vista a estreita relação com o VIF_i.

Nesse caso, quanto mais próximo TOL_i for de zero, maior o grau de colinearidade daquela variável com os outros regressores e quanto mais próximo TOL_i for da unidade, maior a evidência de que X_i não é colinear com os outros regressores.

Finalmente Ferrar e Glauber (1967) indicam um teste que, segundo eles, gera uma medida objetiva para a multicolinearidade, permitindo localizar, detectar e compreender o seu padrão. Esse teste será abordado em um tópico seguinte.

2.2.3. Contornando a Multicolinearidade

No momento que a multicolinearidade é detectada e considerada prejudicial, o pesquisador deve procurar maneiras de reduzir seus efeitos indesejáveis.

Diversas são as medidas que têm sido propostas para resolver o problema de multicolinearidade. Os autores Hair *et al* (2005) destacam os seguintes métodos:
1) excluir uma ou mais variáveis independentes altamente correlacionadas e identificar outras variáveis independentes para ajudar na previsão. Deve-se ter cautela com esse método pois há descarte de informações contidas nas variáveis removidas;
2) usar o modelo com variáveis independentes altamente correlacionadas apenas para previsão, ou seja, não interpretar os coeficientes de regressão;
3) usar as correlações simples entre cada variável independente e a dependente para compreender a relação entre variáveis independentes e dependente, e;
4) usar um método mais sofisticado de análise como a regressão Bayesiana, ou regressão *ridge*, ou a regressão sobre componentes principais para obter um modelo que reflita mais claramente os efeitos simples das variáveis independentes.

2.3. O TESTE DE FARRAR-GLAUBER

Devido ao problema de contar-se com correlações de ordem zero, Farrar e Glauber sugeriram que se devem examinar os coeficientes de correlação parcial. Desta forma, por exemplo, na regressão de Y sobre X_2, X_3 e X_4, um resultado em que $r^2_{1,\,2\,3\,4}$ é muito elevado, mas $r^2_{12,\,3\,4}$, $r^2_{13,\,2\,4}$ e $r^2_{14,\,2\,3}$ são comparativamente baixos, pode sugerir que as variáveis X_2, X_3 e X_4 são estreitamente intercorrelacionadas e que pelo menos uma dessas variáveis é supérflua (GUJARATI E PORTER, 2011).

Sobre multicolinearidade, os autores dizem ainda que a relação de interdependência quase sempre presente pode ser facilmente identificada, medi-la, entretanto, não é tarefa tão simples. De modo bastante pragmático identificar a presença de multicolinearidade em pouco contribui no sentido de saber se isso representa um problema ou não para o modelo. É preciso calcular a magnitude da multicolinearidade para saber se ela é severa, caso que exige tratamento, ou não.

Na sequência são apresentadas as etapas para aplicação do Teste de Farrar-Glauber, extraídas do artigo original “multicollinearity in regression analysis: the problem revisited” escrito pelos mesmos, no ano de 1964.

Segundo os autores, é possível percorrer três etapas para aplicação do respectivo teste, sendo que na etapa 1 é realizado o teste χ^2 para determinar a presença e a gravidade da multicolinearidade; na etapa 2 é realizado o teste F para localizar as variáveis inter-correlacionadas caso o teste χ^2 seja positivo e finalmente a etapa 3 é realizado o teste t para se determinar quais variáveis que são responsáveis pela multicolinearidade, caso o teste F seja positivo.

Para as etapas seguintes está se supondo o teste para um modelo que possui uma variável explicada Y e três variáveis explicativas X_1, X_2 e X_3, entretanto o modelo pode ser generalizado para tantas variáveis explicativas quanto se fizer necessário.

Etapa 01 – Teste χ^2 para determinar a presença e a gravidade da multicolinearidade

Obtém-se inicialmente o valor de cada r_{xixj} através da equação 2.3. Na sequência monta-se a matriz de correlações R através da equação 2.4. Finalmente calcula-se χ^2_{Cal} usando para isso a equação apresentada em 2.5.

$$r_{x_i x_j} = \frac{n \Sigma x_i x_j - \Sigma x_i \, \Sigma x_j}{\sqrt{[n \Sigma x_i^2 - (\Sigma x_i)^2] \cdot [n \Sigma x_j^2 - (\Sigma x_j)^2]}} \tag{2.3}$$

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix} \tag{2.4}$$

$$\chi^2_{Cal} = -\left[(n-1) - \frac{1}{6}(2k+5) \right] \cdot \ln |R| \tag{2.5}$$

Nas equações, n é o número de dados da amostra, k é o número de variáveis independentes do modelo e ln $|R|$ é logaritmo neperiano do determinante da matriz R.

Com um nível de significância de α = 0,05, busca-se o valor crítico tabelado $\chi^2_{Tab} = \chi^2_{\alpha\,;\,k(k-1)/2}$ e aplica-se o teste de hipóteses. Quanto maior o valor do qui-quadrado, maior a gravidade da multicolinearidade. Nesse teste, consideram-se:

Hipóteses de Teste:
H_0: $x_{i's}$ são ortogonais (não apresentam multicolinearidade)
H_1: $x_{i's}$ não são ortogonais (apresentam multicolinearidade)

Decisão do Teste: Se $\chi^2_{Cal} > \chi^2_{Tab}$ rejeita-se H_0, caso contrário, não se rejeita.

Etapa 02 – Teste F para determinar quais os regressores multicolineares do modelo

Para se determinar quais os regressores do modelo que são multicolineares, utiliza-se a equação 2.6 em que se deve realizar o teste F. Sendo assim, tem-se:

$$F_{i\,Cal} = \frac{R^2_{x_i, x_1 x_2 \ldots x_k} / (k-1)}{(1 - R^2_{x_i, x_1 x_2 \ldots x_k}) / (n-k)} \tag{2.6}$$

Com um nível de significância de α = 0,05, busca-se o valor crítico tabelado $F_{Tab} = F_{\alpha, k-1, n-k}$ e aplica-se o teste de hipóteses. Nessas condições, consideram-se:

Hipóteses de Teste:

H_0: $R_{x_i,x_1x_2...x_k}$ = 0 (x_i não é afetado por multicolinearidade)

H_1: $R_{x_i,x_1x_2...x_k} \neq 0$ (x_i é afetado por multicolinearidade)

Decisão do Teste: Se $F_{Cal} > F_{Tab}$ rejeita-se H_0, caso contrário, não se rejeita.

Etapa 03 – Teste t para determinar o regressor responsável pela multicolinearidade

Por fim, para se determinar qual o padrão (responsável) da multicolinearidade, aplica-se o teste t de correlações parciais conforme a equação 2.7. Vejamos:

$$t_{ij\,Cal} = \frac{r_{x_i x_j, x_1 x_2 ... x_k} \cdot \sqrt{n-k}}{\sqrt{1 - r^2_{x_i x_j, x_1 x_2 ... x_k}}} \qquad (2.7)$$

Com um nível de significância de α = 0,05, busca-se o valor crítico tabelado $t_{Tab} = t_{\alpha, n-k}$ e aplica-se o teste de hipóteses. Nessas condições, consideram-se:

Hipóteses de Teste:

H_0: $r_{x_i x_j, x_1 x_2 ... x_k}$ = 0 (x_i e x_j não são variáveis multicolineares)

H_1: $r_{x_i x_j, x_1 x_2 ... x_k} \neq 0$ (x_i e x_j são variáveis multicolineares)

Decisão do Teste: Se $t_{Cal} > t_{Tab}$ rejeita-se H_0, caso contrário, não se rejeita.

3 ESTUDO DE CASO

Para esse estudo foram utilizados os 50 (cinquenta) dados de mercado de imóveis coletados em Florianópolis (2015). Os dados são mostrados na Tabela 01, em que ID é a identificação, VT refere-se ao Valor Total do Imóvel, AT é a área total, DM é a distância ao mar e PC refere-se ao padrão construtivo do imóvel coletado.

ID	VT	AT	DM	PC	ID	VT	AT	DM	PC
AP_01	1.060.000,00	350,00	720	2	AP_26	420.000,00	76,00	700	1
AP_02	510.000,00	136,56	665	2	AP_27	195.000,00	48,00	730	1
AP_03	780.000,00	164,77	415	2	AP_28	290.000,00	66,00	745	1
AP_04	550.000,00	174,58	320	2	AP_29	272.000,00	50,00	1430	1
AP_05	850.000,00	123,01	895	3	AP_30	430.000,00	61,00	170	1
AP_06	300.000,00	89,83	645	1	AP_31	895.000,00	109,00	530	2
AP_07	750.000,00	174,00	860	3	AP_32	450.000,00	89,00	745	2
AP_08	650.000,00	123,00	745	3	AP_33	1.950.000,00	393	550	3
AP_09	620.000,00	121,00	745	3	AP_34	2.150.000,00	578	260	3
AP_10	740.000,00	109,00	300	2	AP_35	940.000,00	182	200	2
AP_11	770.000,00	170,00	590	2	AP_36	1.400.000,00	262	60	3
AP_12	680.000,00	141,00	290	2	AP_37	1.090.000,00	205,00	465	2
AP_13	850.000,00	174,00	465	2	AP_38	1.272.000,00	196	610	3

AP_14	420.000,00	105,00	60	1	AP_39	2.800.000,00	463	590	3
AP_15	547.000,00	128,00	745	3	AP_40	1.796.000,00	273	140	2
AP_16	1.600.000,00	163,00	90	3	AP_41	1.400.000,00	330	655	3
AP_17	1.320.000,00	230,00	215	3	AP_42	3.000.000,00	533	427	3
AP_18	615.000,00	108,00	745	3	AP_43	1.200.000,00	221	607	3
AP_19	705.000,00	174,00	900	3	AP_44	800.000,00	220	1000	2
AP_20	418.000,00	85,00	620	3	AP_45	950.000,00	127	60	2
AP_21	270.000,00	71,00	1380	1	AP_46	2.061.000,00	362,00	310	3
AP_22	418.000,00	100,00	620	3	AP_47	1.326.000,00	315,00	600	3
AP_23	650.000,00	90,00	215	3	AP_48	850.000,00	151,00	660	2
AP_24	700.000,00	161,00	500	3	AP_49	1.650.000,00	246,00	307	3
AP_25	680.000,00	174,00	860	3	AP_50	650.000,00	159,72	120	2

Tabela 01 – Dados de Mercado no Centro de Florianópolis (2015)

Nos dados originais coletados no centro de Florianópolis há outras variáveis explicativas, como número de quartos, número de garagens e número de suítes, entretanto, para tornar o método de aplicação do teste mais didático e as matrizes mais simples, optou-se por, nesse estudo de caso, utilizar somente as variáveis valor total (VT), área total (AT), distância ao mar (DT) e padrão construtivo (PC).

Todos os testes foram realizados em planilha eletrônica *Excel*, previamente formatada pelos autores, usando funções básicas e funções estatísticas específicas.

Com os 50 (cinquenta) dados de mercado de imóveis de Florianópolis/SC, realizou-se os testes, seguindo a sequência seguinte para obtenção dos resultados:

1. Aplicou-se a Etapa 01 do Teste de Farrar-Glauber (Teste χ^2) para detecção da existência e da gravidade da multicolinearidade;
2. Aplicou-se a Etapa 02 do Teste de Farrar-Glauber (Teste F) para se identificar quais são as variáveis que são multicolineares;
3. Aplicou-se a Etapa 03 do Teste de Farrar-Glauber (Teste t) para detecção do padrão da multicolinearidade e na determinação de quais variáveis são responsáveis pelo aparecimento de variáveis multicolineares no modelo.

Finalmente, discutiu-se os resultados obtidos ao se aplicar o Teste de Farrar-Glauber, comparando-os com o que era esperado para o mercado imobiliário.

4 ANÁLISE DE RESULTADOS

Nos tópicos seguintes será apresentada a análise dos resultados obtidos ao se aplicar aos 50 (cinquenta) dados de mercado o Teste de Farrar-Glauber (FG).

Etapa 01 – Teste χ^2 para determinar a presença e a gravidade da multicolinearidade

Considerando-se n = 50 dados, k = 3 variáveis explicativas, α = 0,05 de significância e gl = 3.(3 – 1)/2 = 3 graus de liberdade, e que $\chi^2_{Tab} = \chi^2_{\alpha\,;\,k(k-1)/2}$, tem-se

χ^2_{Cal} = ***15,1705*** $\chi^2_{Tab} = \chi^2_{0,05\,;\,3}$ = ***7,8147***

Conclusão: Como o $\chi^2_{Cal} > \chi^2_{Tab}$ rejeita-se H_0, portanto há correlação entre variáveis do modelo (existe multicolinearidade), considerada de baixa gravidade.

Etapa 02 – Teste F para determinar quais os regressores multicolineares do modelo

Considerando-se n = 50 dados, k = 3 variáveis explicativas, α = 0,05 de significância e que $F_{Tab} = F_{\alpha, k-1, n-k}$, tem-se os seguintes resultados:

$$F_{Tab} = F_{0,05\,;\,2\,;\,47} = 3{,}1951$$

A) Correlação da área total AT (x_1) com distância ao mar DM (x_2) e padrão PC (x_3):

$R^2_{1,23}$ = **0,2637** $F_{1\,Cal}$ = **8,4157**

Conclusão: Como o $F_{Cal} > F_{Tab}$ rejeita-se H_0, portanto AT (x_1) é uma variável afetada por multicolinearidade e está intercorrelacionada com DM (x_2) e PC (x_3).

B) Correlação da distância ao mar DM (x_2) com área total AT (x_1) e padrão PC (x_3):

$R^2_{2,13}$ = **0,05581** $F_{2\,Cal}$ = **1,3900**

Conclusão: Como $F_{Cal} < F_{Tab}$ aceita-se H_0, portanto DM (x_2) não é uma variável afetada por multicolinearidade e não é intercorrelacionada com AT (x_1) e PC (x_3).

C) Correlação do padrão PC (x_3) com área total AT (x_1) e distância ao mar DM (x_2):

$R^2_{3,12}$ = **0,2323** $F_{3\,Cal}$ = **7,1098**

Conclusão: Como o $F_{Cal} > F_{Tab}$ rejeita-se H_0, portanto PC (x_3) é uma variável afetada por multicolinearidade e está intercorrelacionada com AT (x_1) e DM (x_2).

Etapa 03 – Teste t para determinar o regressor responsável pela multicolinearidade

Considerando-se n = 50 dados, k = 3 variáveis explicativas, α = 0,05 de significância e que $t_{Tab} = t_{\alpha, n-k}$, tem-se os seguintes resultados:

$$t_{Tab} = t_{0,05\,;\,47} = 1{,}6779$$

A) Verificação da correlação parcial entre área total AT (x_1) e distância mar DM (x_2):

$r_{12,3}$ = **0,2026** $t_{12,3\,Cal}$ = **1,4186**

Conclusão: Como o $t_{Cal} < t_{Tab}$ aceita-se H_0, portanto área total AT (x_1) e distância ao mar DM (x_2) não são variáveis responsáveis pela multicolinearidade.

B) Verificação da correlação parcial de área total AT (x_1) e padrão construtivo PC (x_3):

$r_{13,2}$ = **0,4693** $t_{13,2\,Cal}$ = **3,6436**

Conclusão: Como o $t_{Cal} > t_{Tab}$ rejeita-se H_0, portanto área total AT (x_1) e padrão construtivo PC (x_3) são variáveis responsáveis pela multicolinearidade.

C) Verificação da correlação parcial entre a distância ao mar DM (x_2) e o padrão (x_3):

$r_{23,1}$ =	***0,0123***	$t_{23,1\ Cal}$ =	***0,0842***

Conclusão: Como o $t_{Cal} < t_{Tab}$ aceita-se H_0, portanto distância ao mar DM (x_2) e padrão construtivo PC (x_3) não são variáveis responsáveis pela multicolinearidade.

5 CONSIDERAÇÕES FINAIS

Esse estudo teve como objetivo aplicar o Teste de Farrar-Glauber para detecção da existência, mensurar a gravidade e identificar as variáveis que poderiam gerar multicolinearidade em um modelo de regressão que fosse construído com base em uma amostra de 50 dados de mercado coletados em Florianópolis/SC.

As variáveis explicativas da amostra que foram consideradas para esse estudo são a área total (AT), distância ao mar (DM) e padrão construtivo (PC). Outras variáveis que compunham os dados coletados foram excluídas para fins didáticos.

Ao se realizar o teste χ^2 para Farrar-Glauber, constatou-se a presença de multicolinearidade de baixa gravidade nos dados do modelo.

Na sequência, com a aplicação do teste *F*, constatou-se que área total e padrão construtivo eram duas variáveis afetadas por multicolinearidade.

Por fim, com a aplicação do teste *t*, pode-se verificar que eram essas duas variáveis juntas as responsáveis pela multicolinearidade detectada nos dados.

É de se esperar que apareça um certo grau de correlação entre as variáveis explicativas área total e padrão construtivo, pois no mercado imobiliário, é normal que imóveis de padrão construtivo mais alto tenham áreas totais maiores.

Uma alternativa para se contornar essa situação é excluir uma das variáveis independentes que foi afetada por multicolinearidade e identificar outras variáveis que não foram consideradas no modelo mas que possam auxiliar nas estimativas.

Para dados de natureza imobiliária, a norma de avaliação de imóveis sugere examinar a coerência das características do imóvel avaliando com a estrutura de multicolinearidade inferida nos casos em que correlação linear for elevada entre quaisquer subconjuntos de variáveis independentes.

Vale destacar que, em engenharia de avaliações, se a multicolinearidade não afeta seriamente as estimativas dos coeficientes, pode-se tolerar a sua presença.

REFERÊNCIAS BIBLIOGRÁFICAS

ABNT. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-2:** Avaliação de Bens. Parte 2: Imóveis Urbanos. Rio de Janeiro, 2011.

DANTAS, Rubens Alves. **Engenharia de avaliações:** uma introdução à metodologia científica. 3. ed. São Paulo: Pini, 2012. 255 p.

FARRAR D. E. & GLAUBER, R. R. **Multicollinearity in regression analysis** – the problem revisited. The Review of Economics and Statistics, v. 49, n. 1, feb. 1967.

FONSECA, M. A. R. **Álgebra linear aplicada a finanças, economia e econometria.** Barueri: Manole, 2003.

GUJARATI, D. N; PORTER, D. C. **Econometria básica.** 5. ed. Porto Alegre: AMGH Bookman, 2011.

HAIR, Jr., J. H.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. Trad. Adonai Schlup Sant'Ana e Anselmo Chaves Neto. **Analise multivariada de dados**. 5 ed. Porto Alegre: Bookman. 2005

HOCHHEIM, Norberto. **Engenharia de avaliações:** módulo básico. Florianópolis: IBAPE - SC, 2015.

HOFFMANN, Rodolfo. **Análise de regressão:** uma introdução à econometria. São Paulo: Portal de Livros Abertos – USP, 2016.

MILOCA, S. A; CONEJO, P. D. **Multicolinearidade em modelos de regressão.** In XXII Semana Acadêmica da Matemática. Universidade Estadual do Oeste do Parana. Cascavel: UEOP, 2014.

MOREIRA, Alberto Lélio. **Princípios de engenharia de avaliações.** 5. ed. São Paulo: Pini, 2001. 512 p.

NEWBOLD, P. **Statistics for business and economics.** 4 ed. Upper Saddle River: PrenticeHall, 1994.

SILVA, Adriano Victor Lopes da. **Alternativas e comparações de modelos lineares para estimação da biomassa verde de bambusa vulgaris na existência de multicolinearidade.** Dissertação (Mestrado em Biometria e Estatística) – Programa de Pós Graduação Estatística e Informática, UFRPE, Pernambuco, 2008.