XIX COBREAP | Foz do Iguaçu

INOVAÇÕES CIENTÍFICAS E TECNOLÓGICAS

CONGRESSO BRASILEIRO DE ENGENHARIA DE AVALIAÇÕES E PERÍCIAS

21 a 25 agosto de 2017

Hotel Mabu Thermas Grand Resort

Foz do Iguaçu / PR / Brasil

FATOR DE INFLACÃO DA VARIÂNCIA E REGRESSÕES AUXILIARES PARA DIAGNÓSTICO DO PROBLEMA DE MULTICOLINEARIDADE NOS MODELOS DE REGRESSÃO

MARIA LUCIA SABEDOTTI DE BIAGGI

MARCELO MEDVID

CYNTHIA MARÍLIA CARRARO DE ASSIS

IBAPE

IBAPE PR

O Conteúdo dos trabalhos técnicos apresentados no COBREAP é de inteira resposabilidade dos seus autores.

Resumo

O objetivo do trabalho é analisar, num modelo estatístico, a importância da existência de multicolinearidade entre as variáveis independentes e, principalmente, se esta afeta ou não, de forma negativa, a definição do valor adequado ao mercado imobiliário. 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? Utilizando-se de técnicas estatísticas, como o cálculo do Fator de Inflação de Variância e por meio de modelos de regressões auxiliares, pode-se, após a detecção da existência de multicolinearidade, verificar a intensidade e as variáveis independentes que se encontram mais afetadas. Logo, cabe ao avaliador decidir se a multicolinearidade é somente uma característica da população amostral e, portanto, não prejudicial na definição do valor de mercado para o imóvel avaliando. Caso verifique-se que a multicolinearidade poderá distorcer o resultado da inferência, propõem-se medidas corretivas a serem implementadas pelo avaliador. Como exemplos práticos foram detalhados dois casos que reforçaram a necessidade da verificação e influência da multicolinearidade no valor de avaliação e alertam para as consequências da aplicação pura e simples do modelo estatístico sem esta análise, ocasionando distorção do valor de avaliação projetado.

Palavras-chave: Multicolinearidade, Fator de Inflação de Variância, Regressões Auxiliares.

1 Introdução

É comum, no processo de elaboração de um modelo estatístico, após coleta de dados no mercado imobiliário e modelagem, o engenheiro de avaliações se deparar com a existência de forte correlação ou multicolinearidade entre duas ou mais variáveis independentes, segundo Reynaldo (1997, p.22): “em muitas análises de modelos de regressão deparamo-nos com o mau condicionamento da matriz de delineamento”, ou seja, nada mais é do que a forte correlação, muitas vezes existente - mas não desejável - entre as variáveis independentes de um modelo de regressão, que são ditas variáveis explanatórias, independentes ou simplesmente regressores.

Uma estatística pouco usual para detecção da presença de multicolinearidade, mas de fácil cálculo e interpretação, é o Fator de Inflação de Variância (FIV), que mede o quanto da variância de cada coeficiente de regressão do modelo estatístico se encontra inflado em relação à situação em que as variáveis independentes não estão correlacionadas.

Juntamente com a utilização do FIV para identificar quais as variáveis independentes são mais multicolineares, pode ser efetuado o cálculo dos modelos de regressão auxiliares, através da regressão de cada uma das variáveis independentes contra as demais, computando-se o coeficiente de determinação (R^2_j) e correlação r_j, valores das estatísticas t de Student de cada regressor, com o objetivo de medir qual é o grau de multicolinearidade e hierarquização da importância de cada variável independe sobre aquela tomada como dependente.

Os efeitos deste mau condicionamento causam o aumento do Fator de Inflação da Variância do estimador de mínimos quadrados das variáveis, que varia de FIV = 1,0 (ausência de multicolinearidade) até infinito (multicolinearidade perfeita). 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 multicolinearidade entre variáveis independentes em um modelo de regressão e diversas soluções que passam pela adoção de medidas corretivas para eliminá-la ou ao menos minimizá-la, que serão apresentadas neste trabalho.

O presente estudo apresenta dois casos propostos de avaliação de terreno, utilizando um mesmo modelo estatístico para a determinação do valor de mercado de dois terrenos (imóveis de mesma tipologia), de elementos amostrais reais coletados, mas com características físicas diferentes com relação a testada, profundidade e área, mostrando quando a multicolinearidade entre as variáveis se torna prejudicial e quando não interfere significativamente, com o objetivo de auxiliar o profissional avaliador na tomada de decisão sobre a necessidade ou não de intervir aplicando medidas corretivas.

Outra ocorrência detectada, em modelos com estas características, é que ao se projetar avaliando com características fora da proporção existente nos demais dados de mercado do modelo, pode-se ter resultados que não traduzam no valor projetado pelo modelo para o avaliando a desvalorização que uma testada muito estreita em relação a sua profundidade e a área total, o que fatalmente ocorre.

Como medidas corretivas, as indagações mais comuns são:

a) Quando ocorrer casos de multicolinearidade, devemos eliminar a variável independente que ocasiona tal problema?
b) Quais são as orientações da norma?
c) O critério adotado para analisar o problema fornece um subsídio seguro para adoção das medidas corretivas?
d) Qual a metodologia a ser utilizada?

Com o objetivo de responder a estas e outras indagações sobre o assunto, busca-se explicitar a metodologia empregada na análise dos casos apresentados de multicolinearidade, determinar o Fator de Inflação de Variância e efetuar a análise através de regressões auxiliares, concluindo-se que se trata de técnica eficiente, principalmente para identificar as variáveis independentes colineares e o padrão das correlações entre as variáveis independentes, através dos modelos de regressão auxiliares.

2 Exposição

2.1 Multicolinearidade

De acordo com Gujarati e Porter (2011, p.330): “O termo multicolinearidade deve-se a Ragnar Frisch. Originalmente, significava a existência de uma relação linear “perfeita” ou exata entre algumas ou todas as variáveis explanatórias de um modelo de regressão”.

Numa regressão com k variáveis independentes (X_1, X_2, …, X_k), há uma relação colinear exata, quando se satisfaz a seguinte equação:

$$\lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_k X_k = 0$$

Equação 1 – Relação colinear exata
Fonte: Gujarati e Poter

Em que **λ_1, λ_2, ..., λ_k**, são constantes tais que nem todas são simultaneamente iguais a zero.

No caso específico do mercado imobiliário, a possibilidade de se encontrar uma amostra em que todas as variáveis independentes sejam perfeitamente correlacionadas, da forma descrita na equação 1, é muito pequena na prática avaliatória, salvo quando criada propositalmente, como por exemplo quando se utiliza poucos dados para muitas variáveis independentes (micro numerosidade geral do modelo) ou variáveis que são operações matemáticas de outras variáveis independentes.

Atualmente, porém, inclui não somente o caso de multicolinearidade perfeita, mas também para o caso em que as variáveis X estão Inter correlacionadas, mas não de forma perfeita, como segue:

$$\lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_k X_k + \upsilon_i = 0$$

Equação 2 – Colinearidade menos perfeita
Fonte: Gujarati e Poter (2011)

Em que υ_i é um termo de erro aleatório.

Assim, visando ilustrar a diferença entre a multicolinearidade perfeita e a menos perfeita, é necessário supor que **λ_2** ≠ 0, assim a equação 1 pode ser reescrita como segue:

$$X_2 = -\frac{\lambda_1}{\lambda_2} X_1 - \frac{\lambda_3}{\lambda_2} X_3 - \ldots - \frac{\lambda_k}{\lambda_2} X_k$$

Equação 3 – Diferença entre colinearidade perfeita e a menos perfeita
Fonte: Gujarati e Poter (2011)

A equação 3 mostra que a variável independente X_2 possui relação linear exata com das demais variáveis independentes, ou seja, X_2 é função de X_1, X_3, ... e X_k, em se tratando de combinação linear das outras variáveis independentes, quando na presença da multicolinearidade perfeita. Nesta situação, o coeficiente de correlação entre a variável X_2 e a combinação linear das outras variáveis X é igual à unidade.

Analogamente, considerando **λ_2** ≠ 0, também é possível reescrever a equação 2 como segue:

$$X_2 = -\frac{\lambda_1}{\lambda_2} X_1 - \frac{\lambda_3}{\lambda_2} X_3 - \ldots - \frac{\lambda_k}{\lambda_2} X_k - \frac{1}{\lambda_2} \upsilon_i$$

Equação 4 – Colinearidade menos perfeita
Fonte: Gujarati e Poter (2011)

A equação 4 mostra que a variável X_2 não é combinação linear exata das demais variáveis X, pois é determinada também pelo termo de erro aleatório $\upsilon_{i;}$ este termo de erro representa que a relação entre a variável X_2 com as demais X deixa de ser exata ou perfeitamente correlacionada, indicando que nem todas as demais variáveis independes explicam as variações X_2. Nesta situação, o coeficiente de correlação da equação 4, não é mais igual a unidade, sendo que pode assumir valores inferiores a unidade, e quanto menor for o coeficiente de correlação menor é a intensidade da multicolinearidade entre a variável X_2 com as demais X.

Na situação de ausência de multicolinearidade, o coeficiente de correlação da equação 4 é nulo, entretanto, para dados reais do mercado imobiliário esta situação é de rara ocorrência, sendo na prática encontrados valores intermediários de zero à unidade, consistente com a situação de colinearidade menos que perfeita.

Diante do exposto, é necessário se utilizar medidas estatísticas que indiquem o quanto próximo da situação extrema de colinearidade perfeita, entre as variáveis independentes X, se encontra o modelo de regressão linear a ser utilizado para avaliações imobiliárias.

A ABNT NBR 14653-2:2011, em seu anexo A, item 2.1.5.2, indica que as dependências lineares de primeira ordem devem possuir atenção especial para resultados superiores a 0,80, sendo que também pode ocorrer multicolinearidade quando a matriz de correlações isoladas apresenta valores baixos; é necessário que se efetuem análises mais completas, quando identificada a presença de correlações prejudiciais a estimativa do valor de mercado pelo modelo inferido.

A leitura do texto normativo recomenda a verificação do correlacionamento de cada variável independente com subconjuntos de outras variáveis independentes por meio de regressões auxiliares, conforme segue:

> 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. Como também é possível ocorrer multicolinearidade, mesmo quando a matriz de correlação apresenta coeficientes de valor baixo, recomenda-se, também, verificar o correlacionamento de cada variável com subconjuntos de outras variáveis independentes, por meio de regressões auxiliares. (Anexo A – ABNT NBR 14653-2:2011 – item 2.1.5.2)

As regressões auxiliares se tratam de procedimentos comuns, aplicados a dados econômicos e explorados através da disciplina de Econometria, sendo considerado um método pouco usual para dados de mercado imobiliário. Trata-se de uma forma de descobrir qual variável independente X se encontra mais correlacionada as outras X, sendo que a lógica do teste se encontra em regredir cada uma das variáveis X em relação as demais, de forma alternada, em um novo modelo; se o coeficiente de correlação deste modelo for elevado, acima de 0,80, indica forte indício da existência da multicolinearidade no modelo.

No caso de relação exata, tem-se a multicolinearidade perfeita, ou seja, variáveis independentes que são perfeitamente relacionadas, que é muito raro de se encontrar na prática. Neste caso, poderia ser facilmente detectado e também resolvido, tanto com a eliminação de uma ou mais variáveis independentes do modelo, ou com a criação de outra variável que sintetizasse as variáveis independentes correlacionadas em uma única variável.

Segundo Kmenta (1977) *apud* Gazola (2002, p. 46):

> O interesse no que se refere a multicolinearidade está nos casos em que ela ocorre com alto grau, isto é, quando duas variáveis independentes estão altamente correlacionadas ou quando há uma combinação quase linear entre um conjunto de variáveis independentes. Assim, a multicolinearidade é mais uma questão de grau do que de natureza.

Conforme Miloca e Conejo (2011, p.1):

> Uma teoria abordada na análise multivariada de dados é a construção e validação de modelos de regressão linear. Tais modelos surgem em problemas em que o interesse de estudo está em saber qual o comportamento das variáveis em questão e qual relação existente entre elas. Deseja-se construir um modelo matemático que melhor represente tal relacionamento. Quando se tem pares de observações de duas variáveis, é possível avaliar o relacionamento entre elas fazendo-se um gráfico de dispersão e assim, indicar como seria o modelo matemático. Em muitos problemas, os modelos a serem construídos são lineares.

Ainda, de acordo com Miloca e Conejo (2011, p.1): "A teoria de Regressão Linear é importante quando se tem duas ou mais variáveis envolvidas no problema (tanto na variável resposta quanto nas covariáveis). Um modelo de regressão linear tem por equação":

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times p}\,\mathbf{B}_{p \times 1} + \epsilon_{n \times 1}$$

Equação 5 – Regressão linear
Fonte: Miloca e Conejo (2011)

> Onde
> - **Y** é o vetor das *n* observações (variável dependente).
> - **X** ´e a matriz das variáveis independentes.
> - ϵ é a matriz dos erros aleatórios e representa a influência de outros fatores não considerados no modelo, bem como os erros de medição da variável resposta Y. Em geral é suposto que $\epsilon \approx N(0; \sigma^2)$, ou seja, os erros experimentais são independentes e normalmente distribuídos.
> - **B** é a matriz dos coeficientes desconhecidos do modelo que devem ser estimados.

Neste modelo, espera-se encontrar dependência entre a variável resposta Y_i e os regressores X_j, para isto faz-se necessário uma análise das suposições do modelo de regressão.

De acordo Miloca e Conejo (2011), há alguns fatores que podem influenciar na busca do melhor modelo de regressão:

a) Especificação dos objetivos da na análise de regressão, que inclui a seleção de variáveis dependentes e independentes;
b) Testar as suposições referentes à análise de regressão (normalidade, linearidade, homocedasticidade e independência dos termos de erro), testando as variáveis individualmente;
c) Fazer análise diagnóstica do modelo, conferindo se as suposições de regressão se confirmam;
d) Interpretar a variável estatística de regressão, examinando o papel das variáveis independentes, com relação à previsão da variável dependente;
e) Resultados são confirmados na generalização da população.

Segundo Elian (1988) *apud* Gazola (2002, p. 47):

> A multicolinearidade geralmente é causada pela própria natureza dos dados, principalmente nas áreas de economia com variáveis que representam valores de mercado. Algumas vezes a multicolinearidade pode também ocorrer devido à amostragem inadequada.

De acordo com Miloca e Conejo (2011): “assim, deve-se investigar se existem dependências entre os regressores *Xj.* Em situações onde essas dependências forem fortes, dizemos que existe multicolinearidade. A multicolinearidade refere-se à correlação entre três ou mais variáveis independentes”. Miloca e Conejo ainda orientam a procurar variáveis independentes que tenham baixa colinearidade com outras variáveis independentes, porém tenham fortes correlações com a variável dependente.

Conforme Montgomery (1997) *apud* Gujarati e Poter (2011, p.332), ocorre a multicolinearidade devido aos seguintes fatores:

> 1. O método de coleta de dados empregado. Por exemplo, a amostragem de uma faixa limitada de valores pelos regressores da população.
> 2. Restrições ao modelo ou à população que está sendo amostrada. Por exemplo, na regressão do consumo de eletricidade contra renda (X_2) e o tamanho da casa (X_3), há uma restrição física na população, no sentido de que famílias com rendas mais altas em geral têm casas maiores que as com rendas mais baixas.
> 3. Especificação do modelo. Por exemplo, adicionando termos polinomiais a um modelo de regressão, especialmente quando a amplitude da variável X é pequena.
> 4. Um modelo sobre determinado. Isto acontece quando o modelo tem mais variáveis explanatórias que o número de observações. Poderia ocorrer em pesquisa médica na qual pode haver um número pequeno de pacientes sobre os quais são coletadas informações a respeito de um grande número de variáveis.
>
> Outra razão para a multicolinearidade, principalmente nos dados de séries temporais, pode ser que os regressores incluídos no modelo tenham uma tendência comum: todos aumentam ou diminuem ao longo do tempo.

2.2 Detectando a Multicolinearidade

Existem muitas sugestões, ou métodos propostos, para detectar a multicolinearidade segundo Gazola (2002, p.49-51), os mais comumente são:

> - **Coeficiente de correlação simples:** é uma medida comumente usada no caso de duas variáveis independentes, sendo suficiente para detectar a colinearidade. Considera-se que um coeficiente de correlação maior que 0,80 ou 0,90 é indicativo de um problema sério de colinearidade. Porém, para mais de duas variáveis independentes, mesmo os coeficientes de correlação sendo baixos ainda pode existir a multicolinearidade, pois pares de correlações podem não dar visão de intercorrelacionamentos mais complexos entre três ou mais variáveis (Judge *et al*., 1980, p.458-459).
> - **Determinante de (X’X):** Se as variáveis independentes estão padronizadas, tal que (**X’X**) contém elementos que são os coeficientes 50 de correlação linear entre as variáveis independentes, então o determinante de (**X’X**) é um valor no intervalo [0;1]. Caso as variáveis independentes não estejam padronizadas, é melhor analisar o determinante da matriz de correlações entre as variáveis independentes, **R**x, que sempre assume

valores no intervalo [0,1]. Um valor deste determinante próximo de zero é indicativo de multicolinearidade (Elian, 1988, p.125; Neter e Wasserman, 1974, p.347).

- **Coeficiente de explicação do modelo, R_2**: o R_2 tendo um valor alto, mas o coeficiente de correlação parcial tendo valores baixos tem-se a indicação de multicolinearidade.
- **Regressões auxiliares**: se o valor de R_2, calculado da regressão de cada variável independente sobre as outras (k-1) variáveis independentes é alto, então há indicativo de multicolinearidade.
- **Raízes características:** Sejam λi , i=1,...,p, as raízes características de **R**x, tem-se que det(**R**x)=Πλi. Baixos valores de uma ou mais raízes características, comparado com o maior valor, são indicativos de multicolinearidade. Um critério, bem eficiente, na quantificação da multicolinearidade, é a análise do valor de L, dado por L=λmáx/λmín, onde λmáx é o maior valor das raízes características e λmín é o menor valor das raízes características. Se L < 100, considera-se não existir multicolinearidade, se 100 < L < 1000 existe multicolinearidade moderada e se L > 1000 há indicativo de multicolinearidade séria (Elian, 1998, p.131).
- **O gráfico dos resíduos**: o gráfico dos resíduos versus cada variável independente, inclusive as variáveis que não fazem parte da equação de regressão, indicam a inexistência de correlacionamento quando os pontos estiverem dispostos aleatoriamente, sem qualquer padrão. Caso contrário, se apresentarem algum tipo de tendência há indicativo de correlacionamento.

2.3 Consequências da Multicolinearidade

Em casos de quase ou de alta multicolinearidade, as seguintes consequências podem ocorrer:

- **Grandes Variâncias e Covariâncias dos estimadores do MQO:** A velocidade com a qual as variâncias e covariâncias aumentam pode ser vista com o Fator de Inflação da Variância (FIV), definido como:

$$FIV_j = -\frac{1}{(1-R_j^2)}$$

Equação 4: Fator de Inflação da Variância
Fonte: Gujarati e Poter (2011)

Onde FIV_j representa o Fator de Inflação da Variância, R^2_j representa o coeficiente de determinação parcial de Xj em relação as demais variáveis Xj, (com j = 1, 2, ..., k) e mostra como a variação de um estimador é *inflada* pela presença da multicolinearidade", de acordo com Gujarati e Poter (2011, p. 337).

Para os casos em que a multicolinearidade é elevada, ou seja, quando Xj se encontra altamente correlacionado com as demais variáveis X, o valor do R^2_j assume valores próximos a unidade, fazendo com que os resultados de FIV também sejam elevados:

$$R_j^2 \approx 1{,}0 \Rightarrow FIV_j\ grande$$

Equação 5: Caso de multicolinearidade elevada
Fonte: Gujarati e Poter (2011)

A Multicolinearidade: o que acontece se os regressores estiverem correlacionados? Segundo Guajarati e Poter (2011, p. 337): “Quando R^2 aproxima-se de 1, o FIV aproxima-se do infinito. Ou seja, quando a colinearidade aumenta, a variância de um estimador aumenta e no limite, pode tornar-se infinita. Caso não haja colinearidade entre as variáveis, o FIV será 1”, conforme segue:

$$R_j^2 \approx 0{,}0 \Rightarrow FIV_j = 1{,}0$$

Equação 6: Sem colinearidade entre as variáveis
Fonte: Gujarati e Poter (2011)

- **Intervalos de Confiança mais amplos:** Conforme Gurajati e Poter (2011, p. 338):

> Quando os erros padrão são grandes, os intervalos de confiança dos parâmetros populacionais relevantes tendem a ser maiores, portanto, em casos de alta multicolinearidade, os dados da amostra podem ser compatíveis com um conjunto diverso de hipóteses, logo, a probabilidade de se aceitar uma hipótese falsa aumenta.

- **Razões t “insignificantes”:** De acordo com Gurajati e Poter (2011, p. 338):

> para testar a hipótese nula de que o regressor não seja significativo, compara-se o valor de *t* estimado com o valor crítico de *t* na tabela *t de Student*, porém em casos de alta colinearidade, os erros padrão aumentam acentuadamente, tornando os valores *t* menores, aceita-se cada vez mais a hipótese nula de que o verdadeiro valor populacional relevante é zero.

- **Alto valor de R^2, mas poucas razões t significativas:** Segundo Gurajati e Poter (2011, p. 345):

> Sintoma “clássico” da multicolinearidade. Se R^2 for alto (superior a 0,8) o teste *F* na maioria dos casos rejeitará a hipótese de que os coeficientes angulares parciais são simultaneamente iguais a zero, mas os testes *t* individuais mostrarão que nenhum dos coeficientes angulares parciais ou poucos deles são estatisticamente diferentes de zero”. [...] “sua desvantagem está no fato de ser “forte demais, no sentido de que a colinearidade é considerada prejudicial somente quando todas as influências das variáveis explanatórias sobre *Y* não puderem ser distintas.

Logo de acordo com, Gurajati e Poter (2011, p. 339): “como forte indício de multicolinearidade temos valores *t* insignificantes, mas um R^2 geral alto e um valor F significativo. O R^2 tendo um valor alto, mas os coeficientes de correlação parcial tendo valores baixos, tem-se a indicação de multicolinearidade”.

- **Sensibilidade dos estimadores de MQO e de seus erros padrão a pequenas alterações nos dados:** Segundo Gurajati e Poter (2011, p. 338):

> Se a multicolinearidade não for perfeita é possível estimar os coeficientes de regressão, mas as estimativas e seus erros padrão tornam-se muito sensíveis a qualquer pequena alteração nos dados.

- **Altas correlações entre pares de regressores:** Conforme Gurajati e Poter (2011, p. 345):

> Se o coeficiente de relação entre dois regressores for alto, maior que 0,8, a colinearidade será um problema sério. O problema desse critério, embora altas correlações de ordem zero possam sugerir colinearidade, não é necessário que sejam altas para que exista colinearidade em qualquer caso específico [...], altas correlações de ordem zero são condição suficiente, mas não necessária, para a existência da multicolinearidade, porque ela pode existir embora as correlações de ordem zero ou simples sejam comparativamente baixas (por exemplo, menores que 0,50).

2.4 Efeito da Multicolinearidade nos Testes de Regressão

De acordo com Neter e Wasseman (1996) *apud* Gazola (2002, p. 48):

> A alta multicolinearidade detectada entre variáveis independentes, apesar do citado anteriormente, não é, geralmente, um problema quando o propósito da análise de regressão é fazer inferências sobre a função resposta ou predições de novas observações, contanto que estas inferências sejam feitas dentro do âmbito das observações.

Ainda de acordo com Judge (1988) *apud* Gazola (2002, p. 49). "Quando o relacionamento entre as variáveis independentes se mantém, no período previsto e no período da amostra, as previsões são corretas até mesmo fora da amostra".

Então, pode-se observar, em alguns casos, como nos modelos econômicos onde a estrutura permanece de amostra para amostra, a multicolinearidade não é um problema, não sendo o caso de uma amostra infectada, logo, o modelo de regressão poderá ser empregado.

Na maioria das vezes a principal razão da ocorrência da multicolinearidade é a micronumerosidade ou deficiência de dados, ou seja, independe da vontade ou empenho do engenheiro de avaliações para ser corrigida.

2.5 Soluções para resolver o problema de Multicolinearidade

Quando a existência de multicolinearidade detectada for prejudicial, Gazola (2002, p. 51) propõe:

> - **Remoção de variáveis -** uma medida simples é remover uma ou várias variáveis independentes, pouco importantes no contexto geral, que venham a diminuir a multicolinearidade. Porém, esta ação não ajuda a avaliar os efeitos da variável independente, pois nenhuma informação é obtida sobre a variável removida, e também porque o valor do coeficiente de regressão para a variável independente remanescente no modelo é afetado pelas variáveis independentes correlacionadas não incluídas no modelo.
> - **Ampliação do tamanho da amostra -** algumas vezes é possível adicionar algumas observações na amostra que elimina o padrão de multicolinearidade. Esta medida é usada quando o problema é causado por informação amostral inadequada. Porém, muitas variáveis independentes não podem ser controladas de forma que novas observações tenderão a mostrar o mesmo padrão de intercorrelação.
> - ***Ridge Regression* -** outro critério, para o qual se tem dado atenção, é *Ridge Regression* (Regressão em Cumeeira) que consiste no uso de estimadores tendenciosos para os coeficientes. A função de regressão estimada pela *Ridge Regression* produz predições de novas observações

que tendem a ser mais precisas doque as predições feitas pela função de regressão estimada pelo método de mínimos quadrados, quando as variáveis independentes são correlacionadas e a nova observação segue o mesmo padrão de multicolinearidade. Esta precisão na predição de novas observações é favorecida pela *Ridge Regression*, especialmente quando a multicolinearidade é forte (Neter et al., 1996, p.411-416).

- **Componentes principais –** outra forma que pode ser utilizada para tratar o problema causado pela multicolinearidade é a técnica de componentes principais. Esta técnica permite que todas as variáveis independentes participam de certa forma do modelo. Através desta técnica, é possível reduzir um grande número de variáveis independentes em um número razoavelmente pequeno de novas variáveis independentes, que são chamadas de componentes e são determinadas pela combinação linear das variáveis originais. Estas novas variáveis (ou componentes principais) são não correlacionadas e são usadas para determinar o modelo de regressão.

O objetivo da análise de componentes principais é representar ou descrever um conjunto de variáveis por outro conjunto menor de novas variáveis, sem perda significativa da informação original (Reis, 1997, p.255). A redução da dimensionalidade das variáveis consiste no fato de que as primeiras componentes principais possam explicar a maior parte da variabilidade total dos dados originais.

A análise de componentes principais permite um estudo detalhado da importância de cada variável, fornecendo a quantidade de explicação na componente principal e seu relacionamento com as demais variáveis.

Outra forma de resolver o problema de colinearidade, de acordo com Gurajati e Poter (2011, p. 346):

- **Regressões auxiliares –** uma forma de descobrir qual variável *X* está relacionada a outras variáveis *X* é fazer a regressão de cada X_i contra as demais variáveis *X* é calcular o R^2 correspondente (R^2_i). Cada uma dessas regressões é chamada de regressão auxiliar, pois é auxiliar em relação à principal regressão de *Y* contra *X*. Seguindo a relação entre *F* e R^2, a variável segue a distribuição *F* com $k-2$ e $n-k+1$ graus de liberdade. Na equação abaixo n representa o tamanho da amostra, *k* representa o número de variáveis explanatórias que incluem o termo do intercepto e $R^2\ x1, x2, x3... xk$ é o coeficiente de determinação da variável X_i contra as variáveis *X* remanescentes.

$$F_i = \frac{R^2_{x_i, x_2 x_3 \cdots x_k} / (k-2)}{\left(1 - R^2_{x_i, x_2 x_3 \cdots x_k}\right) / (n - k + 1)}$$

Equação 5: Equação Regressões Auxiliares
Fonte: Gujarati e Poter (2011)

Para Gurajati e Poter (2011, p. 347):

Se o *F* calculado excede o F_i crítico no nível de significância escolhido, considera-se que o Xi é colinear com os outros X; se não exceder o F_i crítico, diremos que não é colinear aos outros *X* e neste caso mantemos a variável no modelo. Se F_i for estatisticamente significativo, teremos que decidir se o X_i em questão deve ser excluído do modelo.

Quando R^2 - coeficiente de determinação na regressão do regressor X_j contra os regressores remanescentes do modelo aumenta no sentido da unidade, a colinearidade de X_j com os outros regressores também aumenta. Então o FIV também aumenta e, no limite, pode ser infinito.

Logo, o FIV pode ser usado como indicador de multicolinearidade, pois quanto maior for o FIV, mais colinear será a variável X_j. Como regra prática, se o FIV for maior que 10 (o que acontecerá se R^2_j for maior que 0,90), esta variável será considerada altamente colinear. Porém, conforme recomendado pela NBR 14653 -2 item A.2.1.5.2 vigente, onde a correlação múltipla superior a 0,80 deve ser verificada especialmente, logo, aconselha-se limitar a aceitação da variável a um FIV até 5. Logo, FIV = 1 significa a não existência de colinearidade entre as variáveis e FIV = infinito significa a existência de colinearidade.

Abaixo, tabela que indica a relação entre o aumento do grau de correlação entre as variáveis e o aumento do FIV, ou seja, quanto maior a correlação entre as variáveis dependentes maior será o FIV e o nível dessa correlação:

Correlação	**FIV**	**Níveis**
0,40	1,19	Fraca
0,60	1,56	Média
0,75	2,29	Forte
0,85	3,60	Muito Forte
> 0,85	3,60	Fortíssima

Tabela 1 - Níveis de Multicolinearidade
Fonte: Dados produzidos pelos autores (2017)

3 Discussão do Modelo Estatístico

Na sequência, serão apresentados dois casos práticos de avaliação de terrenos, utilizando-se um mesmo modelo estatístico com existência de multicolinearidade entre variáveis independentes.

O software utilizado é o CastleR que permite testar regressões auxiliares e calcula os FIV referentes a cada uma dessas regressões, permitindo a hierarquização da significância de cada uma das variáveis independentes em relação a que apresenta maior correlação.

Foi analisada uma base com 167 dados de terrenos a venda que, após tratamento estatístico, resultou na consideração de 156 elementos amostrais efetivos. Segundo o teste t de Student para o cálculo da significância para rejeição da hipótese nula aplicada em 10 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.

Abaixo são apresentados os principais resultados estatísticos encontrados.

3.1 Metodologias Utilizadas

Utilizou-se o Método Comparativo Direto de Dados de Mercado, que permite a determinação do valor levando-se em consideração as diversas características e comportamentos do mercado imobiliário de Curitiba.

Neste método, a determinação do valor do imóvel avaliando resulta da comparação deste com amostra de natureza e características intrínsecas e extrínsecas diversas, a partir de dados pesquisados no próprio mercado. Os atributos dos dados obtidos são trabalhados por meio de técnicas de inferência estatística, valendo-se de modelo matemático de regressão linear.

3.2 Regressão Linear

Após execução de todos os testes e cálculos necessários, utilizando-se de 10(dez) variáveis independentes efetivas, que se mostraram as mais representativas, em conjunto, nesta análise, elaborou-se modelo de regressão linear e tratamento estatístico aplicável, baseado na amostra de 167(cento e sessenta e sete) dados de mercado, dos quais 156 (cento e cinquenta e seis) foram efetivamente utilizados, cujos resultados elementares foram:

- Coeficiente de Determinação (R^2): 0,736564428413132(er) / 0,772069119982883(fe)
- Coeficiente de Determinação Ajustado (R ajustado²): 0,719677532798589(er) / 0,757458166135632(fe)
- Coeficiente de Correlação (R): 0,858233318167695(er) / 0,878674638295019(fe)

3.3 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 73,66%(er) e por 77,21%(fe) da formação dos preços pertencentes à amostra.

3.4 Coeficiente de Correlação

O coeficiente de correlação (ou R) é obtido do 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, utiliza-se a seguinte escala:

- Correlação Perfeita: R = 1,00
- Correlação Forte: 1,00 > R >= 0,75
- Correlação Média: 0,75 > R >= 0,50
- Correlação Fraca: 0,50 > R > 0,00
- Correlação Inexistente: R = 0

O modelo de regressão, aqui ajustado, indica um coeficiente de correlação Forte para a Equação de Regressão (85,82%) e também para a Função Estimativa (87,87%).

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:

- Equação de Regressão: F calculado em 43,6175153341275, correspondendo ao nível de significância de 0,01% e confiabilidade mínima de 99,99%.
- Função Estimativa: F calculado em 52,8418012988345, correspondendo ao nível de significância de 0,01% e confiabilidade mínima de 99,99%.

3.6 Normalidade dos Resíduos

A existência da normalidade dos resíduos é uma suposição essencial para que os resultados do ajuste do modelo de regressão linear sejam confiáveis.

Em uma distribuição normal de resíduos, a probabilidade de o valor estar no intervalo que dista um desvio-padrão da média é cerca de 68%; a probabilidade de estar no intervalo que dista 1,64 desvio padrões da média é de 90% e a probabilidade de estar no intervalo que dista dois desvios-padrão da média é cerca de 95%.

Os valores, aqui obtidos, são:

- -1dp à 1dp: 68%(er) / 74%(fe)
- -1,64dp à 1,64dp: 89%(er) / 89%(fe)
- -1,96dp à 1,96dp: 97%(er) / 95%(fe)

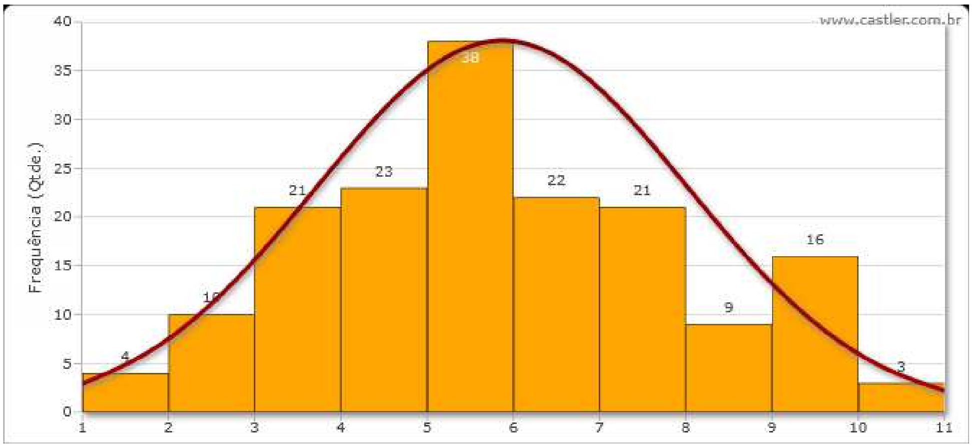

Figura 1 **-** Histograma dos Resíduos
Fonte: Histograma produzido pelos Autores (2017)

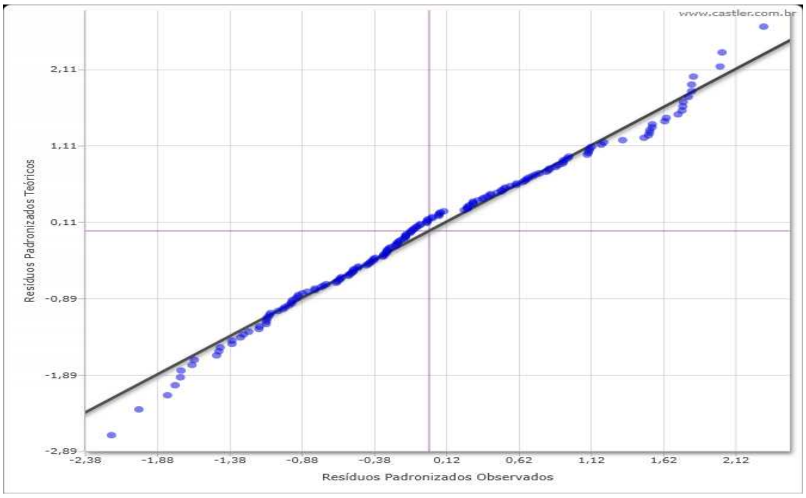

Figura 2 - Gráfico de normalidade QQ Plot[1]
Fonte: Gráfico produzido pelos Autores (2017)

Conforme os gráficos apresentados na figura 01 e 02, histograma dos resíduos e QQ plot, indicam que a distribuição dos resíduos se assemelha a curva normal ou gaussiana de probabilidades.

3.7 Outliers

São os resíduos extremos que apresentam alto afastamento dos restantes, sendo considerados, também, pontos atípicos à massa de dados.

Identifica-se, no presente modelo de regressão estatística, 5 (2,99%) outlier (s) acima de +/-2DP para a Equação de Regressão e 8 (4,79%) para a Função Estimativa.

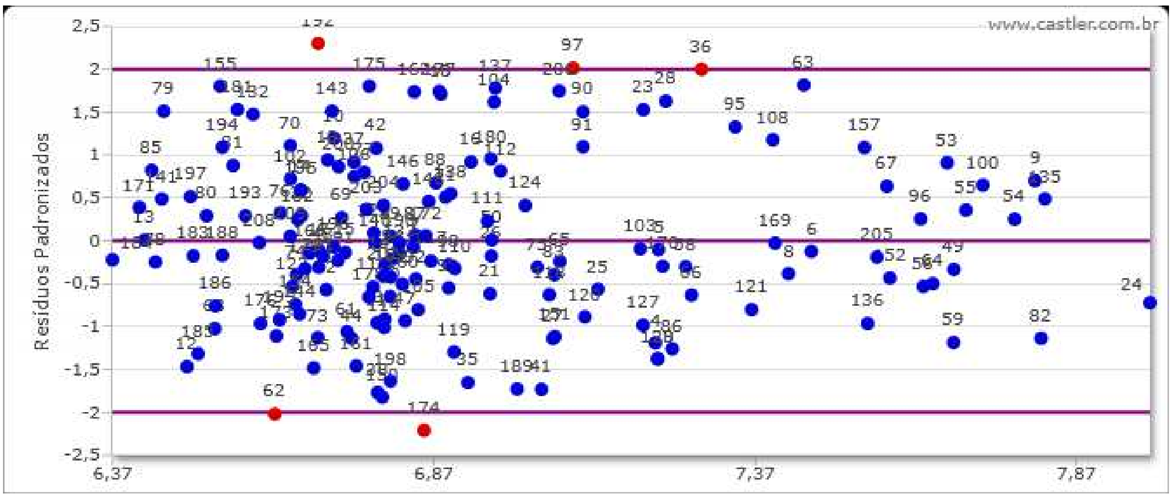

Figura 3 - Gráfico dos Resíduos versus Preços Calculados
Fonte: Gráfico produzido pelos Autores (2017)

[1] Se os pontos se dispuserem em torno de uma linha, os dados podem ser considerados como se tivessem vindo de uma população Gaussiana, ou seguem a distribuição normal.

Os pontos representados no gráfico correspondente a figura 3 apresenta os pontos dispersos aleatoriamente, sem nenhum padrão definido garantindo a hipótese de homocedasticidade e ausência de autocorrelação dos resíduos.

3.8 Variáveis

3.8.1 Variáveis Aplicadas

Mostraram-se significativas e estão presentes, no modelo estatístico, as seguintes variáveis:

Nome		Descrição
Preço Unitário	**Y**	**Preço de Venda / Área Privativa**
Data do Evento	**X1**	**Período de contagem: mensal. Tem como referência inicial o dado mais antigo.**
SHOPPING PATIO BATEL#	**X2**	**Quantitativa. Distância à: 22J, x:671861m, y:7184882m. DISTÂNCIA DO PONTO AO SHOPPING PATIO BATEL**
Renda IBGE 2010	**X3**	**Qualitativa - Variável proxy para localização. Kriging - Determinada por método de regressão usando geoestatística espacial na interpolação de valores (Renda, por setor censitário, IBGE 2010). Parâmetros Base de interpolação espacial: Método Ordinário, Modelo Esférico de Semi-Variograma e 12 metros de Tamanho de Célula de saída.**
Serviços Presentes no Entorno	**X4**	**Quantitativa. Assume o valor da somatória das características presentes dentre a gama das disponíveis: 'Coleta de Lixo', 'Comércio', 'Esporte/Lazer', 'Transporte Coletivo', 'Unidade de Saúde'e 'Unidade de Segurança'.**
Tipo de Uso Predominante no Trecho	**X5**	**Qualitativa. Valores assumidos: 1(Residencial Unifamiliar), 2(Residencial Multifamiliar) e 3(Comercial, Industrial).**
Tipo de Pavimentação no Endereço	**X6**	**Qualitativa. Valores assumidos: 1(Antipó, Saibro) e 2(Pavimentação Asfáltica, Pavimentação em Concreto, Articulada).**
Área Privativa	**X7**	**Quantitativa. Área (privativa) terrestre da unidade.**
Frente	**X8**	**Quantitativa. Medida da frente da unidade terrestre.**
Coeficiente de Aproveitamento	**X9**	**Quantitativa. Índice que, multiplicado pela área de uma unidade terrestre, indica a quantidade total de metros quadrados passíveis de serem construídos.**
Topografia	**X10**	**Qualitativa. Valores assumidos: 1(declive, acidentado), 2(aclive) e 3(plano, semi-plano).**

Tabela 2 - Descrição das variáveis do Modelo
Fonte: Dados produzidos pelos Autores (2017)

3.8.2 Parâmetros dos Regressores

As variáveis aplicadas no modelo estatístico apresentam as seguintes características:

Variável	IDENTIFICAÇÃO	TRANSFORMAÇÃO	COEFICIENTE	T OBSERVADO	SIGNIFICÂNCIA	CRESCIMENTO	FIV
Preço Unitário	Y	ln(y)					
Data do Evento	X1	ln(x)	0,10053122	4,972500651	0,00%	1,79%	1,085326
SHOPPING PATIO BATEL#	X2	$1/x^2$	1132126,165	3,058787015	0,26%	-0,85%	1,495408
Renda IBGE 2010	X3	1/?x	-33,45859459	-5,960978884	0,00%	4,06%	1,344346
Serviços Presentes no Entomo	X4	ln(x)	0,278829194	2,779870415	0,61%	1,67%	1,156975
Tipo de Uso Predominante no Trecho	X5	$1/x^2$	-0,486846221	-8,445893383	0,00%	5,83%	1,57676
Tipo de Pavimentação no Endereço	X6	$1/x^2$	-0,0943505	-1,796196173	7,44%	0,48%	1,364164
Área Privativa	X7	ln(x)	-0,193240163	-5,384240498	0,00%	-20,85%	3,76514
Frente	X8	ln(x)	0,193680269	3,373674406	0,09%	8,15%	3,446008
Coeficiente de Aproveitamento	X9	x^2	0,058348138	2,966608492	0,35%	1,42%	1,528578
Topografia	X10	x^2	0,020169555	3,334001748	0,11%	2,23%	1,08555

Tabela 3 - Modelo de Regressão Linear Múltipla
Fonte: Autores (2017)

3.9 Equações

As funções de ajuste calculadas são as seguintes:

Tipo	Função
Função Estimativa	$Y = e\text{^}(7{,}57613489883726 + 0{,}100531219580206 * \ln(X1) + 1132126{,}16517739 / X2^2 - 33{,}4585945905453 / \sqrt{X3} + 0{,}278829194124601 * \ln(X4) - 0{,}486846220530466 / X5^2 - 0{,}0943504995609494 / X6^2 - 0{,}193240163254748 * \ln(X7) + 0{,}193680269091028 * \ln(X8) + 0{,}0583481378295041 * X9^2 + 0{,}0201695550172749 * X10^2)$
Equação de Regressão	$\ln(Y) = 7{,}57613489883726 + 0{,}100531219580206 * \ln(X1) + 1132126{,}16517739 / X2^2 - 33{,}4585945905453 / \sqrt{X3} + 0{,}278829194124601 * \ln(X4) - 0{,}486846220530466 / X5^2 - 0{,}0943504995609494 / X6^2 - 0{,}193240163254748 * \ln(X7) + 0{,}193680269091028 * \ln(X8) + 0{,}0583481378295041 * X9^2 + 0{,}0201695550172749 * X10^2$

Tabela 4 - Equação de Regressão e Função Estimativa
Fonte: Autores (2017)

3.10 Multicolinearidade

3.10.1 Fator de Inflação de Variância (FIV)

Deve-se investigar se há dependências entre os regressores, pois existem situações em que essas dependências são significativas, causando efeitos nocivos de multicolinearidade.

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

A correlação prejudicial pode ocorrer da forma isolada ou de forma múltipla.

Diagnostica-se a correlação múltipla, de forma preliminar, por meio do FIV (Fator de Inflação de Variância), que é uma medida do grau em que cada variável independente é explicada pelas demais variáveis também independentes.

Variável	IDENTIFICAÇÃO	TRANSFORMAÇÃO	FIV
Preço Unitário	Y	ln (y)	
Data do Evento	X1	ln (x)	1,085326
SHOPPING PATIO BATEL#	X2	1/x²	1,495408
Renda IBGE 2010	X3	1/?x	1,344346
Serviços Presentes no Entorno	X4	ln (x)	1,156975
Tipo de Uso Predominante no Trecho	X5	1/x²	1,57676
Tipo de Pavimentação no Endereço	X6	1/x²	1,364164
Área Privativa	X7	ln (x)	**3,76514**
Frente	X8	ln (x)	**3,446008**
Coeficiente de Aproveitamento	X9	x²	1,528578
Topografia	X10	x²	1,08555

Tabela 5 - Fator de Inflação da Variância
Fonte: Autores (2017)

Conforme os resultados do FIV para cada uma das variáveis independentes, aquelas que superam o valor de 2,77 e representadas em negrito e sublinhado na tabela 04, sendo indicadas as variáveis Área Privativa como FIV = 3,765 e Frente que possui FIV = 3,446, pois possuem coeficiente de correlação múltipla superiores a 0,80, sendo classificada multicolinearidade como muito forte.

Assim, pode-se identificar que as variáveis selecionadas "Área privativa" e "Frente" se apresentam muito correlacionadas, sendo que no caso da utilização do modelo para avaliação de imobiliária deve-se ter atenção especial quanto a esta característica identificada no padrão estrutural do modelo em relação a estas características do imóvel avaliando.

3.10.2 Regressões Auxiliares

Pode ocorrer a multicolinearidade nociva mesmo quando os coeficientes de correlação isolada são baixos, no momento em que existir uma ou mais variáveis independentes altamente correlacionadas entre si, de forma múltipla. Verifica-se este comportamento por intermédio de regressões auxiliares de cada Xi contra as demais X's, obtendo-se os respectivos coeficientes de correlação (r). Cada uma dessas regressões é chamada de regressão auxiliar em relação à equação de regressão principal, que tem Y como variável dependente dos X's.

A seguir, mostram-se os parâmetros de multicolinearidade obtidos através de duas regressões auxiliares, quando a Área Privativa e Frente se encontram na situação de dependente contra as demais X's:

Variável dependente em Análise		r	FIV	F	Significância	
Área Privativa	**X7**	**85,7%**	**3,7651**	**48,236**	**0,01%**	
Acessória		Isolada	Coef.	t	Signif.	
Data do Evento	**X1**	**8,64%**	**0,107**	**2,423**	**1,65%**	**+**

Variável dependente em Análise		r	FIV	F	Significância	
SHOPPING PATIO BATEL#	**X2**	**6,09%**	**-6,061e+005**	**-0,738**	**46,18%**	**+**
Renda IBGE 2010	**X3**	**-29,35%**	**-34,145**	**-2,803**	**0,57%**	**+**
Serviços Presentes no Entorno	**X4**	**5,42%**	**0,101**	**0,452**	**65,18%**	**+**
Tipo de Uso Predominante no Trecho	**X5**	**-26%**	**-0,201**	**-1,578**	**11,65%**	**+**
Tipo de Pavimentação no Endereço	**X6**	**-22,62%**	**-0,12**	**-1,028**	**30,56%**	**+**
Frente	**X8**	**83,49%**	**1,322**	**18,378**	**0,01%**	**+**
Coeficiente de Aproveitamento	**X9**	**15,53%**	**-0,03**	**-0,677**	**49,92%**	**-**
Topografia	**X10**	**-4,61%**	**-0,012**	**-0,863**	**38,97%**	**-**

Tabela 6 - Regressão auxiliar com área privativa como variável dependente em análise
Fonte: Dados produzidos pelos Autores (2017)

Na tabela 06 são apresentados os resultados estatísticos da regressão auxiliar em que a variável "Área Privativa" se encontra colocada como dependente, sendo que todas as demais variáveis independentes foram consideradas na situação e explicativas. O coeficiente de correlação múltiplo (r) desta regressão é de 85,7 % (ou 0,857) indicando um grau de multicolinearidade forte, correspondendo a um valor de FIV de 3,765.

Diante dos valores do teste t de Student e das significâncias estatísticas, os resultados da Tabela 05 indicam que as variáveis que são significativas à um nível de 5% são Frente (significância = 0,01%), Renda IBGE 2010 (significância = 0,57%) e Data do Evento (significância = 1,65%), nesta ordem de relevância estas três variáveis são, de todas as variáveis incluídas na regressão auxiliar, aquelas que mais explicam as variações da variável em análise "Área Privativa".

Variável dependente em Análise		r	FIV	F	Significância	
Frente	**X8**	**84,25%**	**3,446**	**42,669**	**0,01%**	
Acessória		Isolada	Coef.	t	Signif.	
Data do Evento	**X1**	**-1,34%**	**-0,056**	**-2,03**	**4,41%**	**-**
SHOPPING PATIO BATEL#	**X2**	**1,57%**	**-1,661e+005**	**-0,323**	**74,71%**	**+**
Renda IBGE 2010	**X3**	**-19,56%**	**8,739**	**1,124**	**26,25%**	**-**
Serviços Presentes no Entorno	**X4**	**2,97%**	**-0,045**	**-0,324**	**74,61%**	**-**
Tipo de Uso Predominante no Trecho	**X5**	**-18,68%**	**0,036**	**0,446**	**65,59%**	**-**
Tipo de Pavimentação no Endereço	**X6**	**-18,6%**	**0,007**	**0,09**	**92,81%**	**-**
Área Privativa	**X7**	**83,49%**	**0,517**	**18,378**	**0,01%**	**+**
Coeficiente de Aproveitamento	**X9**	**13,08%**	**0,019**	**0,707**	**48,06%**	**+**
Topografia	**X10**	**-1,48%**	**0,005**	**0,624**	**53,36%**	**+**

Tabela 7 - Regressão Auxiliar com Frente como variável dependente em análise
Fonte: Dados produzidos pelos Autores (2017)

Na tabela 7 são apresentados os resultados estatísticos da segunda regressão auxiliar em que a variável “Frente” se encontra colocada como dependente, sendo que todas as demais variáveis independentes foram consideradas na situação e explicativas. O coeficiente de correlação múltiplo (r) desta regressão é de 84,25 % (ou 0,8425) indicando um grau de multicolinearidade forte, correspondendo a um valor de FIV de 3,446.

Diante dos valores do teste t de Student e das significâncias estatísticas, os resultados da Tabela 6 indicam que as variáveis significativas à um nível de 5% são Área Privativa (significância = 0,01%) e Data do Evento (significância = 4,41%), nesta ordem de relevância estas duas variáveis dentre as demais variáveis incluídas na regressão auxiliar, aquelas que mais explicam as variações da variável em análise “Frente”.

4 Casos Práticos

Para exemplificar a aplicação do modelo de regressão que apresenta multicolinearidade muito forte, foram selecionados dois imóveis avaliandos, sendo que a diferença entre os terrenos reside principalmente na proporção entre testada, profundidade e a área total. As demais características como localização, parâmetros como taxa de ocupação e coeficiente de aproveitamento, número máximo de pavimentos passíveis de edificação, topografia, recuos obrigatórios, etc., são os mesmos para os dois.

O Avaliando A, possui frente principal de 34,00 m e área privativa de 1.054,00 m² (profundidade estimada de 31,00 m), o Avaliando B possui uma frente principal de 10m para uma área privativa de 5.000,00m² (profundidade estimada de 500,00 m).

No caso do Avaliando A existe uma proporcionalidade adequada do ponto de vista de máximo aproveitamento para incorporação; no caso do avaliando B, o terreno é de difícil aproveitamento, trata-se da chamada “tripa”, ou seja, frente muito pequena em relação a profundidade e área, de dificílima comercialização, face a impossibilidade de implantação de projeto arquitetônico adequado, devido a exigência de cumprimento dos recuos obrigatórios, áreas de estacionamento, manobras e acesso, etc....

Considerando os resultados obtidos no modelo de regressão foi diagnosticado que as variáveis: Área privativa e Frente apresentaram um alto grau de multicolinearidade, com coeficientes superiores a 0,80, sendo que neste caso, a NBR 14653-2:2011 recomenda atenção especial, devendo o engenheiro de avaliações verificar se o imóvel avaliando segue os padrões estruturais do modelo, quando a existência da multicolinearidade pode ser negligenciada (Anexo A – item A.2.1.5.4 da ABNT NBR 14653-2:2011)

Os resultados de avaliação, referente ao estimador pontual, intervalo de confiança e campo de arbítrio do avaliador, são os que seguem:

4.1 Valores de Avaliação do Avaliando A

Lote com Área Privativa de 1.054,00 m² e Frente de 34,00 m.

O Terreno possui 1054m² de área (privativa), 34m de frente (principal) e tem como posicionamento: "esquina".

Atributos de Cálculo do Avaliando: Data do Evento (X1)=49 , SHOPPING PATIO BATEL# (X2)= 6641 , Renda IBGE 2010 (X3)= 5155 , Serviços Presentes no Entorno (X4)= 6 , Tipo de Uso Predominante no Trecho (X5)=1 , Tipo de Pavimentação no Endereço (X6)=2 , Área Privativa (X7)=1054 , Frente (X8)=34 , Coeficiente de Aproveitamento (X9)=1 e Topografia (X10)=3.

Endereço completo: Rua Holanda, 1501 - esq R David Geronasso. CEP: 82510-190. Bairro: Bacacheri. Curitiba/Paraná.

Coordenadas (Sirgas2000 / WGS84)	Latitude - y	Longitude - x	Fuso
UTM	7190298 m	675705 m	22J
Geodésica Decimal	-25,393795°	-49,253353°	-
Geodésica Sexagesimal	25° 23' 37,66" S	49° 15' 12,07" O	-

	Inferido	Nível de Confiança	Estimador Pontual	Precisão
Venda	Valor	80%	Mediana	grau III

	Valor Unitário
Estimador pontual - Valor Mediano (Unitário)	R$ 1.204,78 / m² (0%)
Intervalo de Confiança (Unitário)	R$ 1.121,18 / m² (-6,94%) à R$ 1.294,61 / m² (7,46%)
Campo de Arbítrio (Unitário)	R$ 1.024,06 / m² (-15%) à R$ 1.385,49 / m² (15%)
Amplitude (Unitário)	R$ 173,43 / m² (14,39%)
Valor Definido (Unitário)	R$ 1.204,78 / m² (0%)

	Valor Total
Estimador pontual - Valor Mediano (Total)	R$ 1.269.834,41 (0%)
Intervalo de Confiança (Total)	R$ 1.181.723,28 (-6,94%) à R$ 1.364.515,23 (7,46%)
Campo de Arbítrio (Total)	R$ 1.079.359,24 (-15%) à R$ 1.460.309,57 (15%)
Amplitude (Total)	R$ 182.791,95 (14,39%)
Valor Definido (Total)	R$ 1.269.834,41 (0%) → (R$ 1.270.000,00)

Tabela 8 – Valores calculados para o avaliando A
Fonte: Autores (2017)

4.2 Valores de Avaliação do Avaliando B

Lote com Área Privativa de 5.000,00 m² e Frente de 10,00 m.

Atributos de Cálculo do Avaliando: Data do Evento (X1)=49, SHOPPING PATIO BATEL# (X2)=6641, Renda IBGE 2010 (X3)=5155, Serviços Presentes no Entorno (X4)=6, Tipo de Uso Predominante no Trecho (X5)=1, Tipo de

Pavimentação no Endereço (X6)=2, Área Privativa (X7)=5000, Frente (X8)=10, Coeficiente de Aproveitamento (X9)=1 e Topografia (X10)=3.

Endereço completo: Rua Holanda, 1501 - esq R David Geronasso. CEP: 82510-190. Bairro: Bacacheri. Curitiba/Paraná.

Coordenadas (Sirgas2000 / WGS84)	Latitude - y	Longitude - x	Fuso
UTM	7190298 m	675705 m	22J
Geodésica Decimal	-25,393795°	-49,253353°	-
Geodésica Sexagesimal	25° 23' 37,66" S	49° 15' 12,07" O	-

	Inferido	Nível de Confiança	Estimador Pontual	Precisão
Venda	Valor	80%	Mediana	grau III

	Valor Unitário
Estimador pontual - Valor Mediano (Unitário)	R$ 703,58 / m² (0%)
Intervalo de Confiança (Unitário)	R$ 613,89 / m² (-12,75%) à R$ 806,37 / m² (14,61%)
Campo de Arbítrio (Unitário)	R$ 598,04 / m² (-15%) à R$ 809,12 / m² (15%)
Amplitude (Unitário)	R$ 192,48 / m² (27,36%)
Valor Definido (Unitário)	R$ 703,58 / m² (0%)

	Valor Total
Estimador pontual - Valor Mediano (Total)	R$ 3.517.905,34 (0%)
Intervalo de Confiança (Total)	R$ 3.069.459,26 (-12,75%) à R$ 4.031.869,12 (14,61%)
Campo de Arbítrio (Total)	R$ 2.990.219,54 (-15%) à R$ 4.045.591,14 (15%)
Amplitude (Total)	R$ 962.409,86 (27,36%)
Valor Definido (Total)	R$ 3.517.905,34 (0%) → (R$ 3.500.000,00

Tabela 9 - Valores calculados para o avaliando B
Fonte: Autores (2017)

4.3 Verificação dos imóveis avaliandos quanto ao padrão estrutural do Modelo

Conforme o resultado dos modelos de regressão auxiliar para a utilização do modelo existe a **necessidade de atenção** especial quanto a Multicolinearidade, entre as seguintes variáveis independentes:

- Área Privativa relacionada fortemente com Frente, Renda IBGE 2010 e Data do Evento
- Frente relacionada fortemente com Área Privativa e Data do Evento.

A melhor maneira para se verificar se os imóveis avaliandos seguem o padrão estrutural do modelo é a através dos gráficos de dispersão, sendo que no eixo vertical serão incluídas as variáveis mais afetadas pela multicolinerirade (Área e Frente), sendo nele representados os pontos com as características dos avaliandos.

4.3.1 Área Construída Privativa Averbada versus Frente

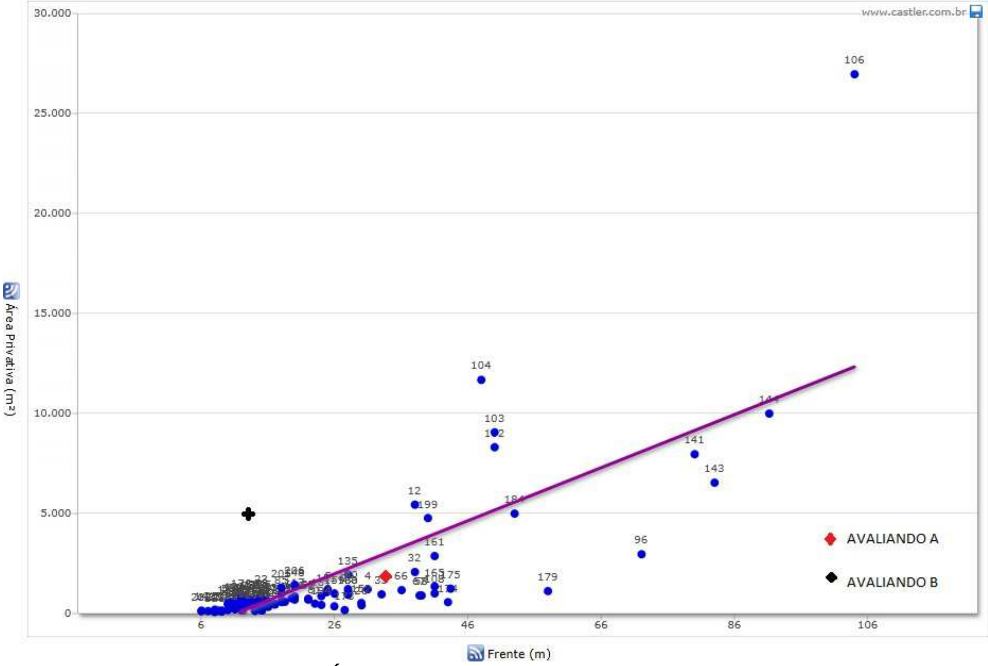

Figura 4 – Área construída averbada versus frente
Fonte: Gráfico produzido pelos Autores (2017)

4.3.2 Área Construída Privativa Averbada versus Renda IBGE 2010

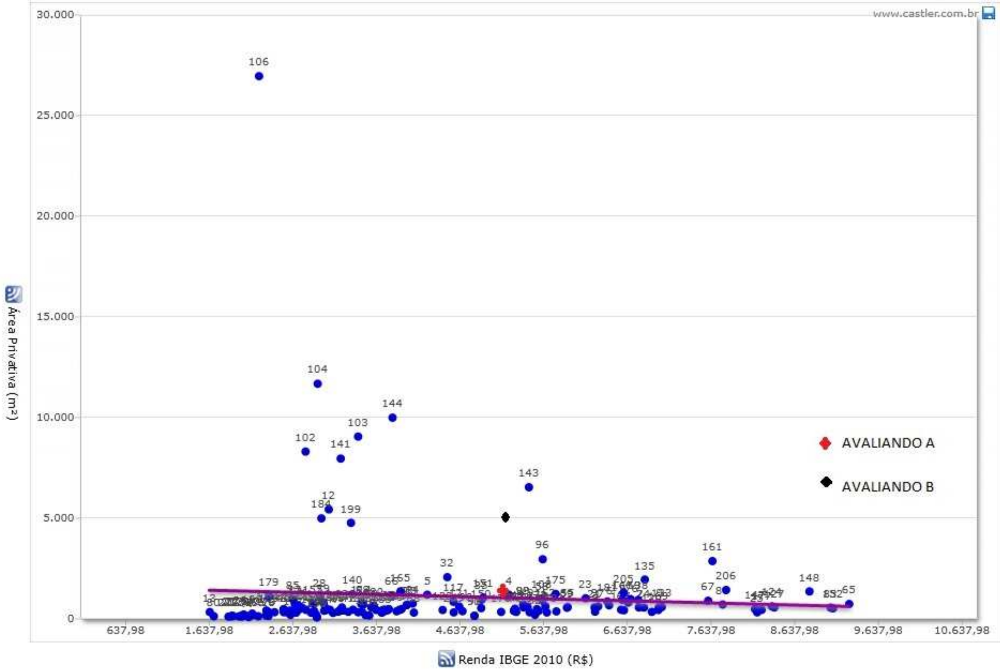

Figura 5 - Área construída privativa versus Renda IBGE 2010
Fonte: Gráfico produzido pelos Autores (2017)

4.3.3 Área Construída Privativa Averbada versus Data

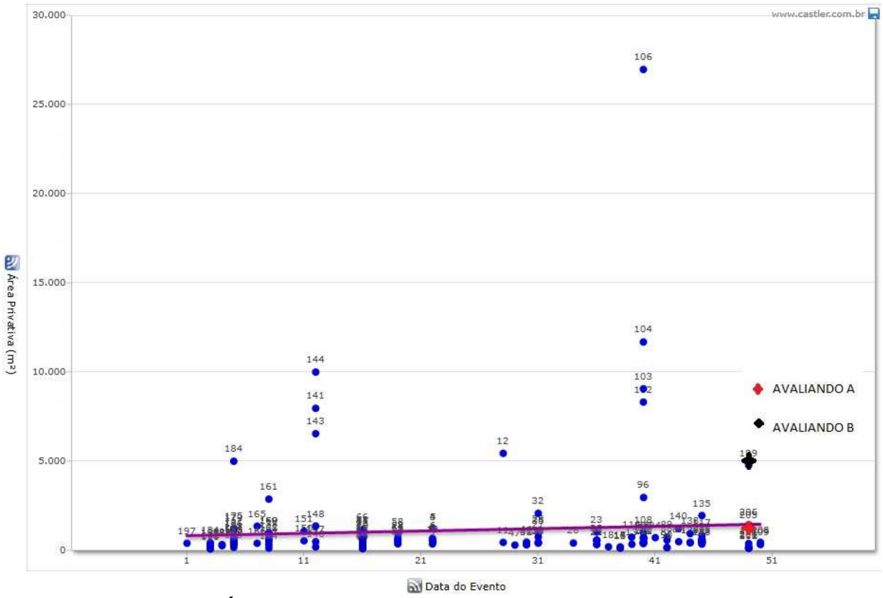

Figura 6 - Área construída privativa averbada versus Data
Fonte: Gráfico produzido pelos Autores (2017)

4.3.4 Frente versus Data do Evento

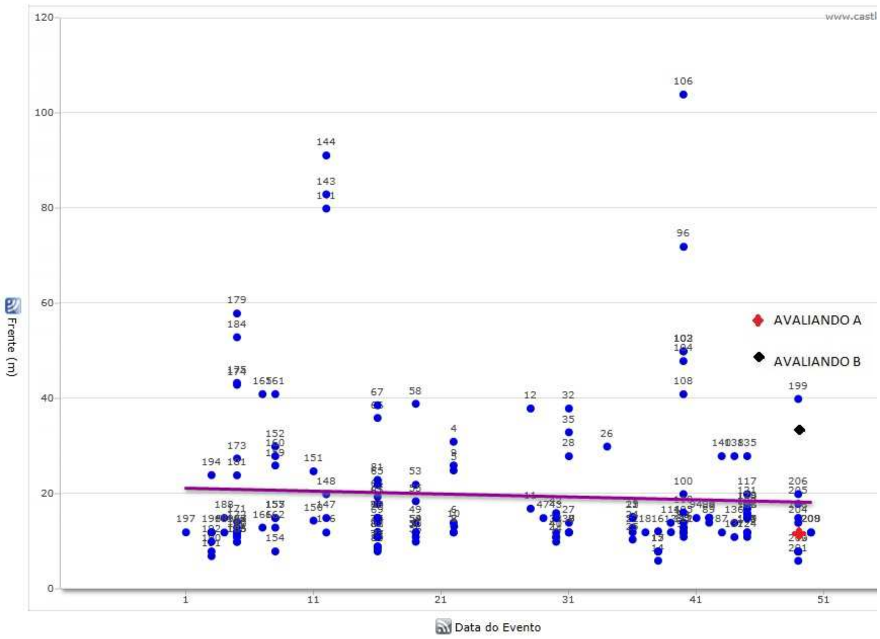

Figura 7 - Área construída privativa averbada versus Data
Fonte: Gráfico produzido pelos Autores (2017)

Verifica-se que para se avaliar o Avaliando A pode ser utilizado o modelo estatístico em questão, o que não ocorre com os Avaliando B, em função de que suas características estão fora da curva de normalidade**.**

5 Conclusões e Recomendações

Do estudo de caso prático apresentado, pode-se concluir a importância da consideração da análise dos valores do Fator de Inflação da Variância (FIV) e dos modelos de regressão auxiliares para o diagnóstico da presença de multicolinearidade elevada, na elaboração de modelos estatísticos de regressão múltipla.

Sabe-se que os modelos multicolineares tem sua utilização limitada e depende de cuidado especial quanto à verificação das características do imóvel avaliando em relação ao padrão de colinearidade presente nos dados amostrais.

No caso do Avaliando A, os exames gráficos nos levam a concluir que, as características físicas, no tocante a área e frente, seguem os padrões estruturais da multicolinearidade presente na amostra estudada. Desta forma, os resultados obtidos nas projeções de valor (estimador pontual, intervalo de confiança e campo de arbítrio) podem ser utilizados para determinar o valor de mercado.

Quanto ao avaliando B isto não ocorre, ficando evidente que suas características físicas são discrepantes em relação ao padrão amostral, sendo vedada pela norma de avaliações vigente ABNT 14653-2 sua utilização.

Dificilmente se conseguirá a absorção pelo mercado de um terreno com as características do Avaliando B, em função do seu aproveitamento ineficiente, ou seja, o valor unitário estimado na projeção pode não ser fidedigno com o mercado imobiliário (vide valores totais para imóveis com as características do Avaliando B).

Com relação aos resultados unitários das projeções, verifica-se que frequentemente é necessária a inserção de mais dados de mercado com características de mercado semelhantes ao Avaliando B (principalmente terrenos com frentes pequenas e áreas grandes) e/ou uma ou mais variáveis que contemplem ou busquem contemplar a proporcionalidade adequada entre frente, profundidade e área. Caso contrário, existe o risco de superavaliação do Avaliando B, pois a desvalorização do avaliando em função da sua forma nem sempre é mensurada adequadamente, conseqüentemente não se traduzindo na projeção de valores.

Há uma outra alternativa a ser considerada é a de adotar o valor mínimo do campo de arbítrio em função da ausência de variável que traduza este mau aproveitamento em função da forma inadequada, caso os 15% de deflação sejam suficientes para depreciar de forma adequada o imóvel avaliando.

Logo, a metodologia proposta neste trabalho se mostra bastante eficiente, rápida e facilitará a identificação dentre o conjunto de variáveis independentes aquelas que se encontram mais correlacionadas entre si, através da identificação das variáveis independentes que possuem maior valor do FIV.

Complementa o cálculo e análise das regressões auxiliares e possui o objetivo de identificar o grau de importância, em função da significância estatística obtida através do teste t de Student, de cada variável mais significativa para cada regressão auxiliar, com o intuito de suprimir ou alterar a variável, ou as variáveis causadoras da multicolinearidade.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653:** Avaliação de Bens - Parte 1 - Procedimentos Gerais e Parte 2 – Imóveis Urbanos. Rio de Janeiro, 2011.

DANIEL, Lindomar P. **O Modelo Clássico de Regressão Linear (MCRL), o Teorema de Gauss Markov e a Violação dos Pressupostos Notas de Aula** - UFMT.

ELIAN, Silvia N. **Análise de Regressão.** São Paulo: IME,1988.

GAZOLA, Sebastião et al. **Construção de um Modelo de Regressão para Avaliação de Imóveis**, Dissertação de Mestrado, UFSC - Florianópolis: 2002.

GUJARATI, Damodar N.; PORTER, Dawn C. **Econometria Básica.** 5a Ed., Porto Alegre: Bookman, 2011.

JUDGE, G. G.; HILL, R. C.; GRIFFITHS, E. W.; LUTKEPOHL, H.; LEE, T. **Introduction to the Theory and Practice of Econometrics.** 2ª ed. USA: John Wiley & Sons, 1988.

KMENTA, Jan; DURÁN, Lluis Barbé. **Elementos de econometria**. Vicens Vives, 1977.

MILOCA, Simone A.; CONEJO, Paulo D. Multicolinearidade em modelos de regressão. **SEMANA ACADÊMICA DA MATEMÁTICA**, v. 22, 2013.

NETER, J., WASSERMAN, W., KUTNER, M. H.; NACHTSHELM, C. J. **Applied Linear Regression Models.** 3ª ed., Times Mirror Hiher Group, Inc., Boston, 1996.

MONTGOMERY, Douglas C. **Design and analysis of experiments**. 4ª ed. John Wiley & Sons, 1997.

REIS, Elizabeth. **Estatística Multivariada Aplicada.** Lisboa: Silabo, 1997.

REYNALDO, Cristiane et al. **Regressão" Ridge": um metodo alternativo para o mal condicionamento da matriz das regressoras.** 1997.

SALVIAN, Mayara. **Multicolinearidade.** USP - Piracicaba: 2016.