

ARBITRAMENTO DO VALOR VENAL DE IMÓVEL URBANO UTILIZANDO REGRESSÃO LINEAR MÚLTIPLA

Vilker Zucolotto Pessin¹; Ana Carolina D. T. Contarato²; Poline Fernandes Fialho³

1. Administrador de empresas pela Universidade Federal do Espírito Santo, acadêmico de Engenharia Civil Multivix Vitória;
2. Acadêmico de Engenharia Civil Multivix Vitória;
3. Mestra em Engenharia Civil, docente Multivix Vitória.

RESUMO

O mercado de avaliações imobiliárias é expressivo em processos que envolvam transações financeiras, especialmente no que tange alienações, garantias reais, questões judiciais, espólio, ou qualquer outra situação em que pese o valor de mercado. É neste contexto que se situa o engenheiro de avaliações enquanto profissional habilitado para a emissão de laudos de avaliação e perícias.

O presente trabalho teve como escopo o desenvolvimento de um simulador, na plataforma Excel®, para avaliação de imóveis urbanos pelo método comparativo de dados de mercado. A metodologia utilizada pautou-se em tratamento científico dos dados, mediante fundamentação em sistemas de regressão linear. A validação científica dos modelos foi embasada em processos de inferência estatística. Buscou-se descrever os principais testes estatísticos, com destaque para análise de *outliers* – pontos influenciadores, correlação, determinação, significância do modelo, micronumerosidade, homocedasticidade, linearidade, aderência.

O resultado do trabalho mostrou a possibilidade de desenvolvimento de uma ferramenta, capaz de verificar a aplicabilidade do conjunto de normas técnicas NBR 14653 no que tange ao arbitramento do valor venal de imóveis em conformidade com as boas práticas da engenharia.

Palavras chaves: Simulador, avaliação de imóveis, NBR 14653, método científico.

Introdução

A avaliação de imóveis busca atender quatro objetivos principais, quais sejam: o objetivo da avaliação, os informes sobre o imóvel avaliado, os dados coletados no mercado e o tratamento científico aplicado a esses informes (SÁ, 2013). Nesta perspectiva, a avaliação de bens, realizada por engenheiro de avaliações, trata-se de uma análise técnica para identificar o valor de um bem, seus custos, frutos e direitos, assim como determinar indicadores de viabilidade de sua utilização econômica, para uma determinada finalidade, situação e data (ABNT, 2005). Nas mesmas palavras, Sá (2013) afirma que a Engenharia de Avaliações almeja a determinação técnica do valor de um bem, seus frutos, custos ou direitos sobre ele.

Arbitrar o valor de um bem exige procedimentos lógicos, baseados em metodologias apropriadas e, sobretudo, que o valor quantificado seja elaborado de forma imparcial. As avaliações de bens imóveis são fundamentais em processos que envolvam transações financeiras, especialmente no que tange alienações, garantias reais, questões judiciais, espólio, ou qualquer outra transação em que pese o valor de mercado.

O valor de mercado corresponde ao preço em que vendedores e compradores concordariam entre si, considerando as condições de perfeita concorrência, cujos envolvidos na negociação conhecem o imóvel e o comportamento do mercado, possuindo condições de avaliar sua tendência de valorização ou de desvalorização (STERTZ, 2016).

Esta pesquisa buscou fazer um estudo dos principais critérios de avaliação contidos nas normas técnicas, bem como análise de parâmetros estatísticos utilizados no arbitramento de valor venal do imóvel através de modelo de regressão em consonância com os requisitos das normas técnicas brasileiras. Como resultado foi desenvolvido um simulador para avaliação de imóveis urbanos em Excel® para fins de análise estatística e validação do modelo. Espera-se que tal simulador possa ser utilizado na elaboração de laudos de avaliação de imóveis urbanos.

Sistemas de regressão

Os sistemas de regressão são métodos estatísticos e econométricos utilizados na determinação de modelos matemáticos preditivos com a finalidade de se observar e descrever o relacionamento entre as variáveis utilizadas na modelagem. Ao mesmo tempo possibilita-se a predição de valores para um objeto de estudo baseado em uma amostra representativa. Temos a relação de uma variável dependente (Y), a qual é determinada em função do conjunto de variáveis independentes (Xi). Em termos práticos, tal situação pode ser observada no arbitramento do valor venal dos imóveis em que temos uma relação de dependência da variável preço em função das variáveis: área construída, área do terreno, situação de conservação, número de quartos, vagas de garagem, localização, dentre outras.

Os sistemas de regressão múltipla são métodos apropriados para estudo de determinada variável dependente métrica quando relacionada a duas ou mais variáveis independentes métricas (HAIR, 2005).

A natureza da relação entre duas variáveis pode ser expressa de variadas formas, incluindo representação de funções matemáticas simples (regressão linear), ou funções mais complexas, em que se utilizem modelos curvilíneos (LEVINE, 2014).

Em suma, as variáveis independentes serão explicadas pelas variáveis dependentes, através de uma equação matemática formulada e validada mediante procedimentos de inferência estatística. O objetivo é permitir previsões coerentes sobre a variável dependente (valor venal de imóveis) fundamentadas em modelo científico.

Método comparativo direto de dados de mercado

A avaliação de bens ao valor de mercado caracteriza-se enquanto a quantia mais provável pela qual se negociaria voluntariamente e conscientemente um bem, numa data de referência, dentro das condições do mercado vigente (ABNT, 2005). Seguindo tal linha de pensamento, a identificação do valor de mercado, sempre que possível deve preferir o método comparativo de dados de mercado (ABNT, 2005).

O método comparativo direto de dados de mercado identifica o valor de um bem por meio de tratamento técnico dos atributos dos elementos comparáveis, constituintes da amostra (ABNT, 2005). A engenharia de avaliações permite a utilização de dois tipos de tratamentos para a determinação do valor de um imóvel, o tratamento por fatores e o tratamento científico, sendo que este último utiliza a estatística inferencial, permitindo assim, estimar e explicar o valor de mercado por meio das variáveis que

influenciam sua formação (COELHO, 2011). Nesta metodologia o valor do bem é estimado através da comparação com dados de mercado assemelhados em função das características, ou seja, por comparação de imóveis similares os em comercialização ou oferta (DANTAS, 2005).

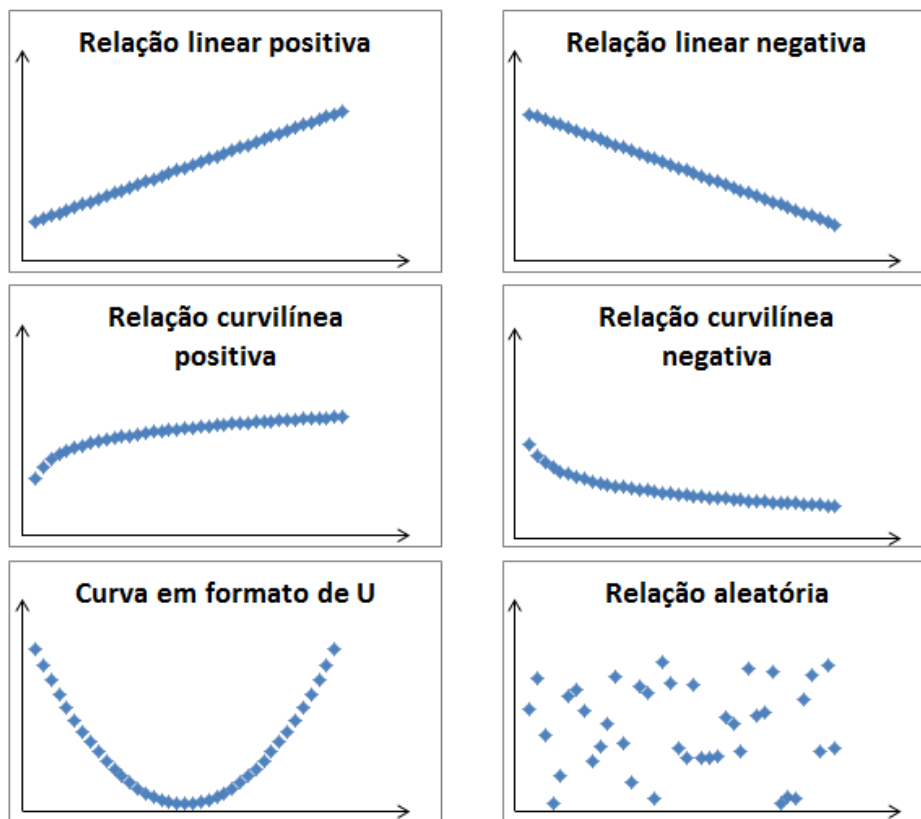
Com relação à utilização do modelo de regressão linear, para representar o mercado, temos a variável dependente que será explicitada por uma combinação linear das variáveis independentes (ABNT, 2011). O método comparativo de dados de mercado, desenvolvido por meio de análise de regressão múltipla, visa obter uma equação que explique a relação da variável dependente (valor venal) em função de variáveis independentes (características dos imóveis avaliados) que podem ser quantitativas e qualitativas (PRUNZEL, 2016).

Radegaz (2011) propõe um roteiro para a análise da regressão através da análise dos seguintes passos: análise do coeficiente de determinação, testes de significância das variáveis individuais, teste t “*student*”, verificação da equação do modelo de regressão, análise gráfica e de resíduos, verificação da autocorrelação para séries temporais, verificação da homocedasticidade, multicolineariedade, normalidade dos resíduos, determinação do intervalo de confiança.

Análise da regressão

A análise gráfica é instrumento importante na validação estatística dos dados utilizados, ao mesmo tempo em que explica alguns comportamentos e relações entre os elementos amostrais. Permite-se verificar a relação entre as variáveis utilizadas na modelagem, seja linear ou curvilínea, positiva ou negativa, bem como validar os pressupostos da avaliação estatística. Conforme descrito por Levine (2014), os gráficos de dispersão permitem verificar as relações entre as variáveis. A seguir, Figura 1 com dados gerados para exemplificar algumas relações possíveis:

Figura 1 - Relações encontradas entre os dados



Fonte: Elaborado pelos autores (2017).

Há necessidade de se observar, no uso de modelos de regressão, os seguintes pressupostos básicos: especificação, normalidade, homocedasticidade, não-multicolineariedade, não-autocorrelação, independência e inexistência de pontos atípicos. O atendimento a tais pressupostos objetiva garantir que as avaliações sejam não tendenciosas, eficientes e consistentes (ABNT, 2011).

Levine (2014) define os pressupostos:

- Linearidade - relação entre variáveis linear.
- Independência - requer que os erros (ϵ_j) sejam independentes entre si (importância maior, principalmente quando a coleta de dados se dá ao longo de determinado período de tempo).
- Normalidade - requer que os erros sejam distribuídos de forma normal para cada valor de X . Os resíduos padronizados podem ser verificados pela proporção de frequência relativa dos resíduos amostrais padronizados, na proporção de 68% para intervalos $[-1; 1]$, 90% para intervalos $[-1,64; 1,64]$, 95% para intervalos $[-1,96; 1,96]$ (ABNT, 2011). Tal atributo pode ser verificado através de gráfico de barras com distribuição das frequências (situação melhor verificada para números amostrais suficientes).
- Igualdade de variância (homocedasticidade) - requer que a variância dos erros seja constante em relação a todos os valores de X (situação de aleatoriedade).

A análise dos resíduos, seja através de testes estatísticos ou análise gráfica, fornece um conjunto simples, mas poderoso de ferramentas analíticas para exame da adequação do modelo de regressão (HAIR, 2005).

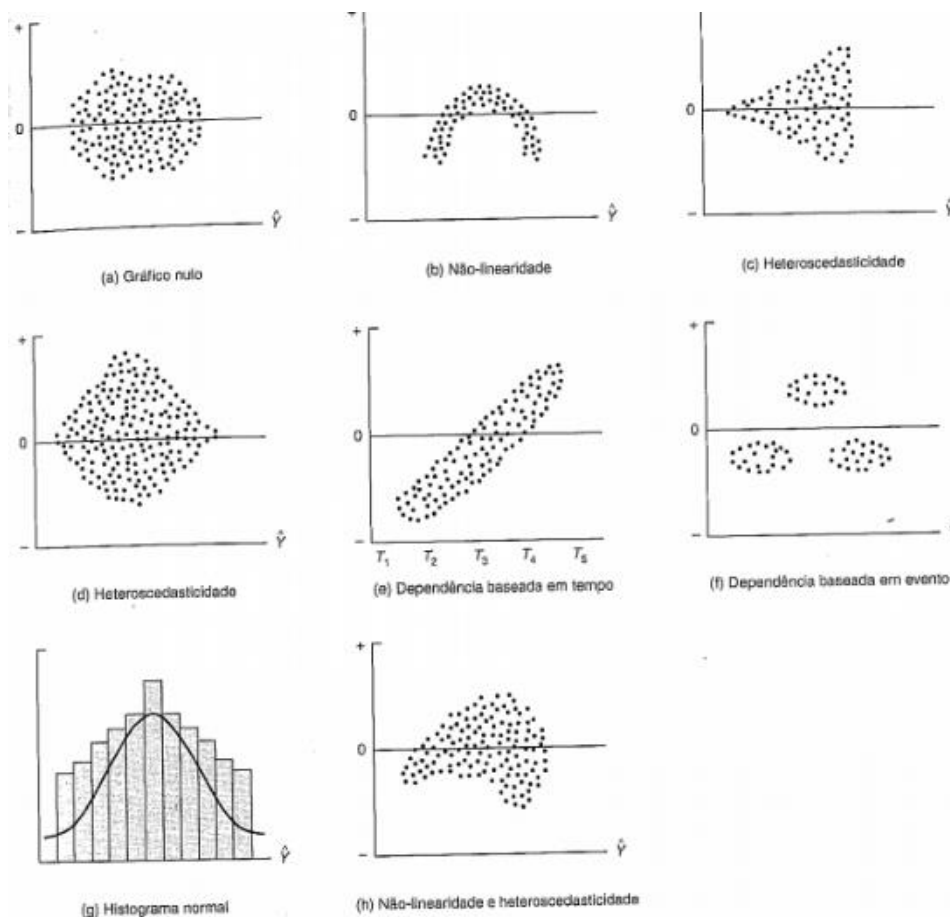
Os gráficos de dispersão dos resíduos são fundamentais na análise da regressão. As informações visuais permitem uma análise fidedigna, de tal modo que tais métodos devem ser incluídos como parte da análise da regressão (LEVINE, 2014).

O resíduo é expresso pela diferença entre cada variável observada (variável amostral dependente) em relação às previsões feitas pela equação definida no modelo de regressão. Ou seja $\varepsilon = Y - \hat{Y}$.

A principal medida de erro de previsão é o resíduo que é medido pela diferença entre valores observados e previstos para a variável dependente. A representação gráfica dos resíduos em função das variáveis previstas é um método apropriado para verificar violações de suposições para relações gerais do modelo de regressão (HAIR, 2005).

De acordo com HAIR (2005), o gráfico dos resíduos serve de estratégia para verificação dos pressupostos para a regressão ao permitir examinar as seguintes suposições: linearidade do fenômeno, variância constante dos termos de erro, independência dos termos de erro, normalidade da distribuição dos termos de erro, conforme Figura 2.

Figura 2 – Gráfico de resíduos



Fonte: HAIR (2005).

A homocedasticidade significa que a variância é constante. A análise gráfica dos resíduos pelos valores ajustados serve de verificação da deste parâmetro em que os pontos devem estar dispostos de forma aleatória (ABNT, 2011).

Nesta perspectiva Hair (2005), ao analisar a variância constante do termo de erro, afirma que a presença de variâncias desiguais (heteroscedasticidade) representa uma das violações mais comuns de suposições nos sistemas de regressão.

De forma análoga, a verificação da autocorrelação pode ser feita, dentre outros procedimentos, pela análise dos resíduos cotejados em relação aos valores ajustados, em que, nessa situação os pontos deverão se apresentar de forma aleatória, sem padrão definido (ABNT, 2011).

A verificação da normalidade, dentre outras formas, pode ser feita pelo exame do histograma dos resíduos amostrais padronizados com a verificação de sua forma em consonância com a curva normal (ABNT, 2011).

O diagnóstico mais simples da normalidade pode ser verificado através de gráfico dos resíduos através de histograma normal com verificação da distribuição que se aproxima da curva normal. Embora este método seja atraente por sua simplicidade, para pequenas amostras a distribuição normal será mal formada (HAIR, 2005).

Por sua vez, a existência de pontos atípicos (“*outliers*”) pode ser verificada pelo gráfico de resíduos em relação a cada variável independente, como também em função dos valores ajustados para se detectar pontos influenciantes (ABNT, 2011).

Para a validação da modelagem é necessário verificar que os resíduos sigam uma distribuição normal, com variância constante no tempo (STERTZ, 2016).

Em suma, para que o modelo seja aceitável, precisamos verificar a aleatoriedade da distribuição dos erros nos gráficos de dispersão. Levine (2014) aconselha a utilização de métodos alternativos de regressão quando não se verifique tal característica gráfica, como por exemplo, a regressão pelo método dos mínimos quadrados.

A condição de colinearidade envolve a possibilidade de duas variáveis independentes estarem fortemente correlacionadas entre si. Nesta situação, as variáveis utilizadas não oferecem informações individualizadas, ou seja, não são isentas de influências. Nestas situações, Levine (2014) apresenta o método que permite mensurar a colinearidade pelo fator inflacionário da variância (FIV), em que:

$$FIV = \frac{1}{1-R^2_j} \quad (1)$$

De acordo com o autor supra, se os valores de FIV forem iguais a 1 (um) indica que um conjunto de variáveis independentes não estão correlacionadas. Por sua vez, valores de FIV superiores a 10 (dez) sugerem que existe uma correlação demasiadamente grande entre as variáveis independentes. Assim, o valor aceitável deve estar no intervalo: $1 < FIV < 10$.

O conjunto de estatísticas são essenciais para perceber a precisão e grau de confiança do modelo de previsão, em que se destaca o coeficiente de determinação (R^2); erro padrão dos coeficientes de regressão (S_b), estatística t e erro da estimativa (S_e).

Como estatísticas adicionais na regressão múltipla, destacam-se a estatística F do modelo, estudo da multicolineariedade e autocorrelação.

O estudo da heterocedasticidade é outro problema que precisa ser verificado, uma vez que o modelo homocedástico (com variância constante dos resíduos) representa uma propriedade fundamental no sentido de que os erros sejam aleatórios, ou seja, não se relacionam com as características dos imóveis. A coerência do modelo pode ser verificada através dos sinais dos coeficientes da equação (RADEGAZ, 2011).

A normalidade é outra propriedade que se verifica, no Excel, através da curva da probabilidade normal. A aderência do modelo pode ser constatada através de visualização gráfica em que se plote valores observados versus valores de predição. E quanto mais próximos os pontos se encontrem da reta de referência, maior o ajuste do modelo aos dados.

GRAU DE FUNDAMENTAÇÃO

No caso de utilização de modelos de regressão linear, deve-se determinar o grau de fundamentação, descrito na Figura 3 (ABNT, 2011):

Figura 3 - Grau de fundamentação no caso de utilização de modelos de regressão linear

Item	Descrição	Grau		
		III	II	I
1	Caracterização do imóvel avaliando	Completa quanto a todas as variáveis analisadas	Completa quanto às variáveis utilizadas no modelo	Adoção de situação paradigma
2	Quantidade mínima de dados de mercado, efetivamente utilizados	6 (k + 1), onde k é o número de variáveis independentes	4 (k + 1), onde k é o número de variáveis independentes	3 (k + 1), onde k é o número de variáveis independentes
3	Identificação dos dados de mercado	Apresentação de informações relativas a todos os dados e variáveis analisados na modelagem, com foto e características observadas no local pelo autor do laudo	Apresentação de informações relativas a todos os dados e variáveis analisados na modelagem	Apresentação de informações relativas aos dados e variáveis efetivamente utilizados no modelo
4	Extrapolação	Não admitida	Admitida para apenas uma variável, desde que: a) as medidas das características do imóvel avaliando não sejam superiores a 100 % do limite amostral superior, nem inferiores à metade do limite amostral inferior; b) o valor estimado não ultrapasse 15 % do valor calculado no limite da fronteira amostral, para a referida variável, em módulo	Admitida, desde que: a) as medidas das características do imóvel avaliando não sejam superiores a 100 % do limite amostral superior, nem inferiores à metade do limite amostral inferior; b) o valor estimado não ultrapasse 20 % do valor calculado no limite da fronteira amostral, para as referidas variáveis, de per si e simultaneamente, e em módulo
5	Nível de significância (somatório do valor das duas caudas) máximo para a rejeição da hipótese nula de cada regressor (teste bicaudal)	10 %	20 %	30 %
6	Nível de significância máximo admitido para a rejeição da hipótese nula do modelo através do teste F de Snedecor	1 %	2 %	5 %

Fonte: Associação Brasileira de Normas Técnicas (2014).

A caracterização do imóvel deverá constar de todos os elementos necessários para a individualização do objeto ora em processo de avaliação. Documentos do imóvel são imprescindíveis no sentido de se comprovar suas dimensões, localização, regularidade, dentre outros elementos importantes.

Com base na amostra dos elementos de mercado, os parâmetros populacionais serão estimados através de inferência estatística. Faz-se a recomendação de que as características do imóvel avaliando estejam contempladas na amostra para atender critérios de representatividade (ABNT, 2011).

A quantidade mínima de dados de mercado é um parâmetro utilizado na classificação do laudo quanto seu grau de fundamentação. Busca-se evitar a micronumerosidade. O número de elementos amostrais servirá de critério de classificação do grau de fundamentação. Temos que a quantidade mínima de dados de mercado, para fins de enquadramento, será obtida pelas expressões: $n = 6(k + 1)$; $n = 4(k + 1)$ e $n = 3(k + 1)$ em que k representa o número de variáveis independentes, ao mesmo tempo em que

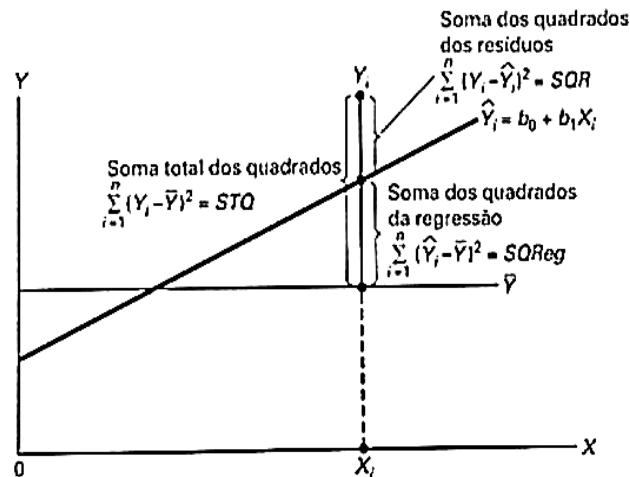
determinará a pontuação atingida no grau de fundamentação (respectivamente 3, 2 ou 1 pontos). Em uma avaliação imobiliária, em que foram utilizadas duas variáveis independentes (tamanho do terreno e área construída, por exemplo), temos que: $n = 6(2+1) = 18$; $n = 4(2+1) = 12$; $n = 3(2+1) = 9$. Logo, a depender do número de elementos amostrais ($n = 18$, 3 pontos; $n = 12$, 2 pontos; $n = 9$, 1 ponto), teremos um “score” que será somado para fins de enquadramento em determinado grau de fundamentação.

A identificação dos dados de mercado passa pelo detalhamento dos dados e variáveis utilizadas na modelagem. Neste quesito pode-se elaborar relatório fotográfico com a comprovação das características do imóvel alvo da avaliação.

O modelo de regressão linear considera um intervalo de valores para as variáveis independentes. Assim, ao se elaborar a previsão, deve-se verificar que o caso em específico não extrapole o intervalo relevante da variável independente. Conforme Levine (2014), a extrapolação é uma armadilha da regressão linear e deve ser evitada. De acordo com este mesmo autor, ao utilizar o modelo de regressão para fins de previsão, o intervalo relevante inclui todos os valores, desde o menor X até o maior.

A utilidade geral do modelo de regressão pode ser testada a partir de alguns métodos estatísticos: coeficiente de determinação múltipla R^2 , R^2 ajustado e teste F geral. As medidas de dispersão são elementos utilizados na verificação da adequabilidade dos modelos, conforme se verifica na Figura 4:

Figura 4 - Medidas de dispersão soma dos quadrados da regressão



Fonte: Levine (2014).

O coeficiente de determinação mede a variação Y que é explicada pela variável independente. O coeficiente de determinação é igual à soma dos quadrados da regressão (SQReg) dividida pela soma total dos quadrados. Temos assim, a expressão:

$$R^2 = \frac{\text{Soma dos quadrados da regressão}}{\text{soma total dos quadrados}} \quad (2)$$

É o índice que indica a porcentagem de explicação da variável dependente em função das variáveis independentes, reflete uma medida da proporção da variância da

variável dependente em torno de sua média e que é explicada pelas variáveis de predição. O coeficiente poderá apresentar variação de 0 a 1. Quanto maior o R^2 , maior o poder de explicação da equação de regressão (HAIR, 2005).

Por sua vez o R^2 ajustado é a alteração no coeficiente de determinação que considera o número de variáveis independentes que foram incluídas na equação de regressão, bem como o número de elementos amostrais utilizados. Essa estatística mostra-se útil para comparar diferentes modelos de regressão, para previsão da mesma variável dependente, indicando qual tipo de modelo apresenta-se mais apropriado (HAIR, 2005).

A equação para R^2 ajustado, conforme Levine (2014) pode ser expressa por:

$$R^2_{aj} = 1 - \left[(1 - R^2) * \frac{n-1}{n-k-1} \right] \quad (3)$$

Em que k é número de variáveis independentes na equação da regressão.

Neste sentido, tal variável pode ser utilizada como parâmetro para estimar a adequabilidade dos diferentes modelos de regressão, seja linear, exponencial, logarítmico ou outro método em que se efetue transformação das variáveis.

A distribuição de Snedcor (teste F) é muito usada em engenharia de avaliações, servindo principalmente para determinar a significância global do modelo.

Nas palavras de Levine (2014), o teste F geral é utilizado para se determinar se existe uma relação significativa entre a variável dependente e o conjunto inteiro de variáveis independentes.

Tendo em vista que existe mais de uma variável independente, você utiliza a hipótese nula e a hipótese alternativa apresentada a seguir: $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ (Não existe nenhuma relação linear entre variável dependente e variável independente. H_1 = pelo menos uma $\beta_j \neq 0$, $j = 1, 2, \dots, k$ (Existe uma relação linear entre a variável dependente e pelo menos uma das variáveis independentes)... A regra da decisão é rejeitar H_0 , no nível de significância de α , se $F_{est} > F_{\alpha}$; caso contrário, não rejeitar H_0 (LEVINE, 2014, p. 543).

Em outras palavras, caso o valor do F calculado seja superior ao F tabelado, aceitamos o modelo de regressão.

Por sua vez, A norma NBR 14653-2 (ABNT, 2011) pontua o enquadramento do grau de fundamentação em virtude do nível de significância máximo para rejeição da hipótese nula do modelo através do teste F, respectivamente: 1% Grau III, 2% Grau II e 5% Grau I.

O teste t, com distribuição de t de Student, é utilizado para verificar a significância de individual de cada variável utilizada no modelo de regressão linear. A norma NBR 14653-2 (ABNT, 2011) descreve o enquadramento do nível de atingimento do grau de fundamentação em virtude do nível de significância (somatório do valor das caudas) máximo para rejeição da hipótese nula de cada regressor (teste t), respectivamente: 10% Grau III, 20% Grau II e 30% Grau I.

Finalmente, a classificação do laudo quanto à fundamentação se dará pela soma de pontos obtidos, conforme tabela 1:

Tabela 1 - Enquadramento do laudo segundo grau de fundamentação no caso de utilização de modelos de regressão linear

Graus	III	II	I
Pontos mínimos	16	10	6
Itens obrigatórios	2, 4, 5 e 6 no Grau III e os demais no mínimo no Grau II	2, 4, 5 e 6 no mínimo no Grau II e os demais no mínimo no Grau I	Todos, no mínimo no Grau I

Fonte: Associação Brasileira de Normas Técnicas (2011).

GRAU DE PRECISÃO

As normas técnicas ainda explicitam que o grau de precisão depende exclusivamente das características do mercado e da amostra coletada e, por isso, não é passível de fixação a priori (ABNT, 2011). O grau de precisão é determinado pela amplitude do intervalo de confiança de 80% em torno da estimativa de tendência central. Assim, ao se efetuar o enquadramento no grau de precisão faz-se necessário determinar o intervalo de confiança ilustrado na tabela 2.

Tabela 2 - Enquadramento do laudo segundo grau de fundamentação no caso de utilização de modelos de regressão linear ou do tratamento de fatores

Graus	III	II	I
Amplitude do intervalo de confiança de 80% em torno da estimativa de tendência central	≤30%	≤40%	≤50%

Fonte: Associação Brasileira de Normas Técnicas (2011).

Conforme Radegaz (2011), o intervalo de confiança, é obtido no Excel, através de operações de matrizes e decorre de uma inferência sobre um parâmetro a partir de uma amostra e procura conter o valor deste parâmetro. O intervalo de confiança contém o erro da estimativa paramétrica, no caso, a média de uma distribuição.

É representado pela expressão algébrica no caso de regressões lineares simples:

$$Y = \bar{y} \pm t_{crit} * Se * \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2}} \quad (4)$$

Em que Se = variação explicada;

Da mesma forma, Radegaz (2011) afirma que o intervalo de predição, é obtido no Excel, através de operações de matrizes. Refere-se ao estudo dos prováveis valores de uma nova e independente observação. É representado pela expressão algébrica:

$$Y = \bar{y} \pm t_{crit} * Se * \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2}} \quad (5)$$

Em se tratando de regressões lineares múltiplas a parte da equação supra $\frac{(x_i - \bar{x})^2}{\sum(x_i - \bar{x})^2}$ será calculada mediante operações com matrizes.

Ferramenta de análise da regressão

No Excel há disponível uma ferramenta de análise da regressão linear múltipla que inclui a regressão linear simples. Ao efetuar o procedimento de regressão, informando todos os parâmetros (variáveis dependentes, variáveis independentes, nível de confiança e intervalo de saída), permitirão que o Excel processe todas as estatísticas da regressão, parâmetros estes para análise do modelo.

Como resultado da regressão, o Excel fornece os seguintes parâmetros estatísticos:

- a) R múltiplo: é o coeficiente de correlação – mostra a força de relação entre as variáveis independentes em relação à variável dependente (Y), também conhecido como coeficiente de correlação;
- b) R²: é o coeficiente de determinação da regressão e representa o percentual da variação do valor explicada pela equação;
- c) R² ajustado: é o coeficiente de determinação que tem utilidade na escolha de equações para regressões múltiplas, por isso, serve de parâmetro para escolha de diferentes modelos considerando transformações nas variáveis, por exemplo;
- d) Coluna gl: graus de liberdade. Que expressa-se pelo número de elementos da amostra deduzidos o número k de variáveis menos 1 ;
- e) Coluna F: o valor F é utilizado para realizar testes de hipóteses da equação da reta da regressão. A distribuição F testa hipóteses em que se determine a significância da regressão e serve de parâmetro para uma análise global do modelo. Mede a variabilidade e permite concluir sobre significância ou incerteza do modelo na presença de todos regressores. É o resultado da divisão da variação explicada (SQ) pela variação não explicada (MQ). Se o seu valor for superior ao F tabelado, pode-se inferir que as variáveis independentes utilizadas no modelo têm representatividade para o cálculo da variação do valor;
- f) F de significação: Teste que avalia a significação global do modelo. Representa a densidade da probabilidade do valor calculado do F crítico, *p-value*.
- g) Coluna dos coeficientes: são os coeficientes de cada regressor que determinam a função de predição do modelo de regressão;

h) Coluna Stat t – informa o teste t observado dos coeficientes, que serão utilizados no teste t student;

i) Coluna Valor-P – mede a probabilidade $P(t \geq t_{\text{observado}})$. O valor P, deverá ser menor que o nível de significância escolhido (no caso de grau de fundamentação III, 10%) (Radegaz, 2011);

h) Coluna “inferiores e superiores”: determina intervalos “superiores e inferiores” de confiança ao nível pré-estabelecido;

i) Análise dos resíduos - procedimento gráfico que permite analisar o ajuste da reta de regressão.

O coeficiente de determinação – R^2 pode assumir valores que variam no intervalo de 0 a 1. Representa uma proporção total de variação que é explicada pela variável independente. Desta forma, quanto maior for o coeficiente de determinação, maior confiança poder-se-á inferir da qualidade do modelo preditivo. Coeficientes muito altos (próximos a 1) podem expressar que as variáveis adotadas estão com excelente poder de explicação. Conforme Radegaz (2011), este fenômeno também pode ser observado naquelas situações em que existe uma variação total muito grande, fato que exige uma interpretação adequada.

Por sua vez, a ocorrência de coeficientes muito baixos (abaixo de 0,6), pode indicar que as variáveis selecionadas para o modelo preditivo não explicam a variação do valor em torno da média. Isso ocorre quando as variáveis independentes estão mal definidas, e por isso, devem ser analisadas uma a uma. De forma análoga, ocorrência de coeficiente reduzido pode indicar que os dados são homogêneos, portanto refletindo em variação total pequena.

A análise destes parâmetros, em conjunto, serve para verificação da pertinência do modelo de regressão no que tange à estimativa dos valores do imóvel objeto de avaliação.

Processamento DOs dados via simulador

Para comprovar os resultados da análise no simulador realizamos, na região de Venda Nova do Imigrante, Espírito Santo, Brasil, pesquisa de mercado de imóveis (casas) com característica de uso residencial. Conforme demonstrado na Figura 5, as variáveis estudadas foram: valor unitário - representa o valor do imóvel avaliando ou o valor de venda negociado para cada amostra dividido pela respectiva área construída; área construída - variável quantitativa que representa a área construída do avaliando e dos dados de mercado; conservação - variável quantitativa que representa a área construída do avaliando e dos dados de mercado; local - variável proxy que representa o local do imóvel avaliando e dados de mercado (1 = Distritos e outras cidades da região, 2 = Outros bairros da cidade na zona urbana, 3 = Centro da cidade); sanitários - variável quantitativa que representa o número de sanitários das amostras.

Figura 5 – Pesquisa de mercado

Nº	Valor unitário	A. Const.	Conservação	Local	Sanitários
1	2.692,31	260,00	1,00	3,00	2,00
2	2.500,00	220,00	2,00	3,00	1,00
3	4.333,33	300,00	3,00	3,00	2,00
4	1.388,89	360,00	1,00	1,00	2,00
5	2.857,14	210,00	3,00	2,00	2,00
6	1.368,42	95,00	1,00	1,00	1,00
7	2.915,38	130,00	2,00	2,00	2,00
8	3.095,24	210,00	3,00	2,00	3,00
9	3.750,00	200,00	3,00	2,00	2,00
10	3.235,29	170,00	2,00	2,00	2,00
11	4.062,50	160,00	3,00	2,00	2,00
12	3.421,05	380,00	3,00	2,00	5,00
13	3.666,67	150,00	3,00	1,00	2,00
14	3.513,51	370,00	3,00	2,00	2,00
15	1.062,50	80,00	1,00	1,00	1,00
16	1.100,00	100,00	1,00	1,00	1,00

Fonte: Elaborado pelos autores, 2017.

O processamento dos dados gerou o resultado verificado na Figura 6.

Figura 6 – Estatística da regressão e resultados

Estatística de regressão	
R múltiplo	0,922574645
R-Quadrado	0,851143975
R-quadrado ajustado	0,797014511
Erro padrão	477,3188223
Observações	16

ANOVA					
	gl	SQ	MQ	F	F de significação
Regressão	4	14330007,5	3582501,876	15,724227	0,0001602
Resíduo	11	2506165,84	227833,2581		
Total	15	16836173,3			

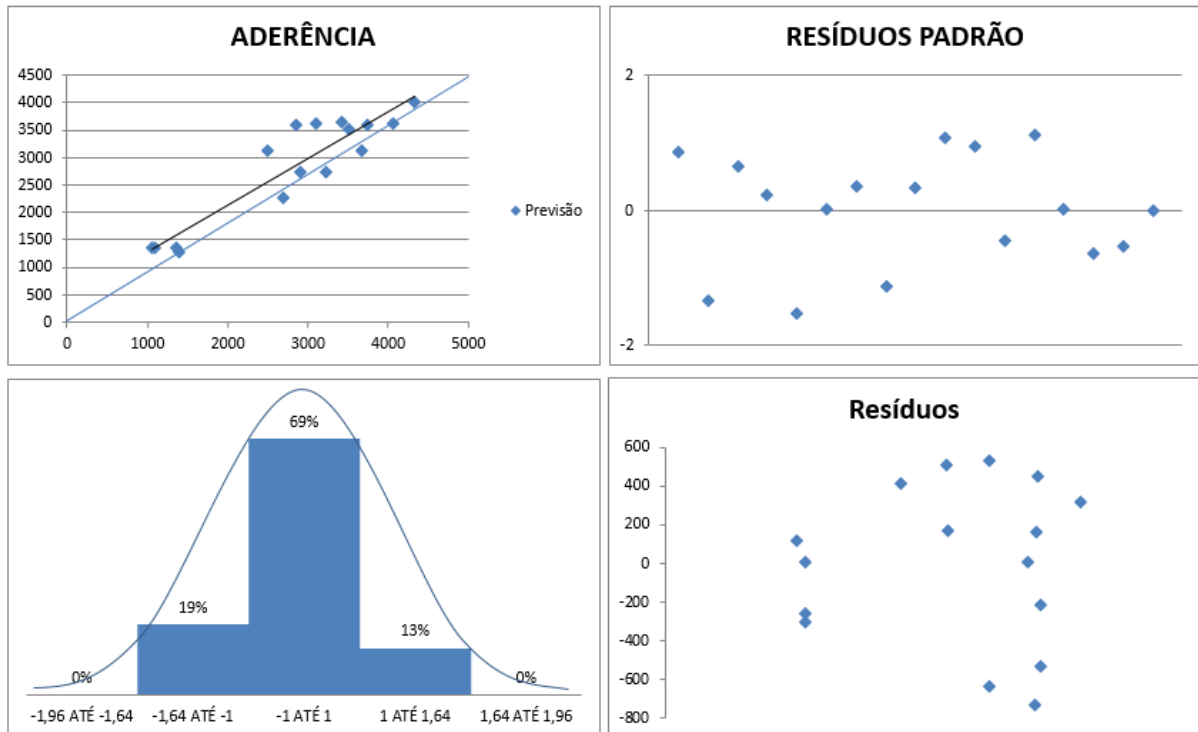
Observação	Amostra	Previsão	Resíduos	Resíduos padrão	Erro padrão
1	2692,307692	2278,860895	413,4467976	1,011487125	0,866185824
2	2500	3134,191103	-634,1911027	-1,551532479	-1,328653037
3	4333,333333	4020,435233	312,8981003	0,765497282	0,655532708
4	1388,888889	1275,359882	113,5290066	0,277745841	0,237847328
5	2857,142857	3586,255918	-729,1130611	-1,783756648	-1,527517934
6	1368,421053	1358,846451	9,574601149	0,023424019	0,020059132
7	2915,384615	2744,427831	170,9567842	0,418241445	0,358160576
8	3095,238095	3630,671962	-535,4338663	-1,3099254	-1,121753095
9	3750	3591,082432	158,9175681	0,388787808	0,332937987
10	3235,294118	2725,121776	510,1723412	1,248123718	1,068829297
11	4062,5	3610,388487	452,1115134	1,106079373	0,947189786
12	3421,052632	3637,453316	-216,400684	-0,529418795	-0,453367171
13	3666,666667	3137,597062	529,0696042	1,294355394	1,108419738
14	3513,513514	3509,031699	4,481814067	0,010964645	0,009389561
15	1062,5	1366,086222	-303,586222	-0,742716007	-0,63602399
16	1100	1356,433195	-256,4331946	-0,627357319	-0,537236712

Fonte: Elaborado pelos autores, 2017.

A análise gráfica permite visualizar a aderência e pertinência do modelo. Os resíduos padrão apresentaram distribuição no intervalo de 2 e -2, demonstrando que não há presença de *outlier*. A homocedasticidade é verificada na distribuição aleatória dos

resíduos. O gráfico da normalidade demonstra adequação à curva de Gauss. A Figura 7 exprime tais relações.

Figura 7 – Gráficos dos resultados



Fonte: Elaborado pelos autores, 2017.

Finalmente, a validação estatística forneceu coeficiente de determinação de 85%, correlação de 92% (muito forte), significância do modelo (probabilidade de erro) aceitável.

Conclusão

Os modelos de regressão linear são indicados, nas normas técnicas, enquanto metodologia preferencial à avaliação imobiliária. Neste tipo de método, podemos projetar previsões para uma população a partir de uma amostra representativa, através de processos de inferência estatística. A avaliação de imóveis pauta-se em um processo de validação dos dados, pela sistemática elaboração de testes estatísticos, no sentido de se verificar a conveniência dos modelos de regressão.

Além da validação estatística, as normas técnicas exigem que se atinja grau de fundamentação satisfatório, expressos pelo atendimento dos critérios prescritos no processo de avaliação. Em consonância com as recomendações técnicas o laudo de avaliação deverá conter número satisfatório de elementos amostrais, caracterização adequada do imóvel, variáveis representativas na identificação dos dados de mercado, testes estatísticos apropriados, intervalos de confiança para arbitramento do valor venal.

Todas essas medidas, em conjunto, buscam atingir a finalidade de se garantir a isonomia da avaliação imobiliária, expressa pela não tendenciosidade. Neste sentido, se expressa o valor venal do imóvel, isento de interesses particulares em consonância com aquilo que se pratica no mercado.

O Excel é uma solução versátil e adequada à manipulação de dados, permitindo a elaboração de modelos aptos ao arbitramento do valor venal dos imóveis sob a ótica científica e em conformidade com as prescrições estabelecidas nas normas técnicas. Fornece suporte satisfatório ao desenvolvimento de sistemas de previsão, incluindo a possibilidade de análise avançada de dados. Permite a visualização gráfica de dispersão dos dados de forma a agregar confiabilidade ao processo de avaliação e validação do modelo de regressão. Tal ferramenta mostrou-se um poderoso instrumento para emissão e elaboração de laudos de avaliação de imóveis, de acordo com as prescrições normativas e em conformidade com as boas práticas da engenharia.

REFERÊNCIAS

Associação Brasileira de Normas Técnicas. **NBR 14653**. Avaliação de bens - parte 1: procedimentos gerais. Rio de Janeiro: 2005.

_____. **NBR 14653-2**. Avaliação de bens - parte 2: imóveis urbanos. Rio de Janeiro: 2011.

COELHO, J.; ZANCAN, E. C. **Modelo de regressão linear múltipla para avaliação de aluguéis de salas comerciais na cidade de Araranguá-SC**. Engenharia Civil - Universidade do Extremo Sul Catarinense - Unesc, Santa Catarina, fevereiro 2011.

HAIR, Joseph Jr. et al. **Análise Multivariada de Dados**. 5ª ed. Porto Alegre: Bookman, 2005.

LAPPONI, Juan Carlos. **Estatística usando o Excel**. 4ª edição. Rio de Janeiro: Campus, 2005. 476p.

LEVINE, D. M. et al. **Estatística: teoria e aplicações usando o Microsoft Excel em português**. 6.ed. Rio de Janeiro: LTC, 2014.

PRUNZEL, J.; TOEBE, M.; LOPES, A. B.; MOREIRA, V. S. Modelos de regressão linear múltipla aplicados à avaliação de terrenos urbanos - caso do município de Itaquí-RS. **BCG - Boletim de Ciências Geodésicas** - On-Line version, Curitiba, v. 22, n. 4, p.651-664, out - dez/2016.

RADEGAZ, Násser Júnior. **Avaliação de bens: princípios básicos e aplicações**. São Paulo: Liv. e Ed. Universitária de Direito, 2011.

SÁ, A. R. da S. Avaliação imobiliária: método comparativo de dados do mercado - tratamento científico. **Revista Especialize on-line IPOG**, Goiânia, v.1, n. 5, julho/2013.

STERTZ, E. da S.; AMORIN, A. L. W.; FLORES, S. A.; WEISE, A. D. Mercado imobiliário: uma análise sobre o comportamento dos preços dos imóveis na cidade de Porto Alegre/RS. **Revista Gestão.Org, Porto Alegre**, v. 14, n. 1, p. 158-166, 2016.