

INTRODUÇÃO

1.1 - CONTEXTUALIZAÇÃO

Os primeiros trabalhos de avaliação de imóveis no Brasil que se tem conhecimento foram publicados em revistas técnicas de engenharia, em São Paulo e datam de 1918. Em 1923 foram introduzidos novos métodos de avaliação de terrenos, que a partir de 1929 começaram a ser sistematicamente aplicados. A partir daí a Engenharia de Avaliação no Brasil vem crescendo e evoluindo nas técnicas de avaliação. Atualmente um grande número de profissionais vem desenvolvendo estudos nesse campo, visando dar à matéria o suporte científico necessário como apoio aos métodos técnicos até então utilizados.

O desenvolvimento da Engenharia de Avaliação, o crescimento do número de profissionais atuando nesse campo e as necessidades do uso das técnicas de avaliação pelo mercado privado e também pelos órgãos públicos, levaram a Associação Brasileira de Normas Técnicas a elaborar a Norma para Avaliação de Imóveis Urbanos. Atualizada pela NRB 14653/2001, sob o título geral “Avaliação de bens”, em suas partes 1(procedimentos gerais) e parte 2 (imóveis urbanos).

O mercado imobiliário é um dos setores mais complexos da economia. As principais dificuldades de análise estão vinculadas a algumas características especiais dos imóveis, que são bem heterogêneos, sendo composto por um conjunto diversificado de atributos, o que dificulta ou impede a comparação direta das unidades.

É nesse ambiente que atua a Engenharia de Avaliações, com a finalidade básica de obter valores para imóveis, ou seja, buscar representações e interpretações numéricas para fenômenos do mercado imobiliário, é uma especialidade da engenharia que reúne um conjunto amplo de conhecimentos na área de engenharia, arquitetura e economia, com o objetivo de determinar tecnicamente o valor de um bem.

A Engenharia de Avaliações serve para subsidiar tomada de decisão a respeito de valores envolvendo bens de qualquer natureza, tais como: imóveis, máquinas e equipamentos, automóveis, móveis e utensílios, além de outros. O presente trabalho restringirá sua pesquisa no estudo de avaliações de terrenos urbanos no município de Betim - Minas Gerais.

Atualmente a Engenharia de Avaliação está totalmente integrada com os sistemas computacionais, sem os quais, a qualidade técnica desejada nestes trabalhos não poderia ser alcançada. Com o desenvolvimento de softwares, as técnicas para a avaliação de um bem foram se consolidando, principalmente a técnica com base na inferência estatística, tema principal desse trabalho.

A representação do funcionamento do mercado imobiliário através de análises empíricas, utilizando a inferência estatística, é geralmente baseada na análise de regressão, que permite a verificação de teorias ou a obtenção de valores para imóveis específicos ou de indicadores do funcionamento geral deste setor da economia.

1.2 - AVALIAÇÃO DE IMÓVEIS

1.2.1 - Conceito

Os bens do mercado imobiliário são singulares. Não existem, a rigor, dois imóveis iguais. Em geral, a localização e as diferentes características físicas impedem a comparação direta. Os imóveis podem ser vistos como um conjunto indissociável de atributos. Para comparar diversos imóveis, deve-se recorrer à comparação simultânea de suas características. As diferenciações aparecem em alguns ou em todos dos vários aspectos que compõem os imóveis (González, 2003).

Do ponto de vista geral e pela definição contida na **NBR-14.653 – PARTE 1: PROCEDIMENTOS GERAIS**, a avaliação de um bem consiste na *“análise técnica, realizada por Engenheiro de Avaliações, para identificar o valor de um bem, de seus custos, frutos e direitos, assim como determinar indicadores da viabilidade de sua utilização econômica, para uma determinada finalidade, situação e data”*.

Independentemente do conceito de avaliação que tomemos como adequada, é necessário definir o valor do bem considerado e o próprio mercado imobiliário.

1.2.2 - Valor

Existem várias definições para valor, valor de mercado e preço. O ângulo de análise, ou mesmo a área científica de quem os estuda, tem influência nos termos e nos sentidos em que são empregados, mas interessa conceituar do ponto de vista específico da avaliação de bens imóveis (González, 2003).

Assim, surge um aspecto bastante complexo que é o Conceito de Valor. Existem diversos tipos de valor que podem ser atribuídos a um bem, entre eles: Valor Venal, Valor Potencial, Valor Comercial, Valor do Mercado, etc. No entanto, numa avaliação, o valor a ser determinado é o valor de mercado (Dantas, 1998, p.7). Estas atribuições são impostas pelo mercado que determina o valor pela lei da oferta e da procura. Assim, o valor de mercado é o preço consciente determinado por um vendedor e pago por um comprador a um bem, sem coação de ambos os lados (Ayres, 1996, p.21).

No entanto, o valor do bem difere e não deve ser confundido com o preço do bem, que representa a quantidade de dinheiro paga pelo mesmo. Assim, a necessidade de venda ou compra imediata e/ou a não existência de um livre comércio podem alterar o preço de um bem, tornando-o superior ou inferior ao valor avaliado (Moreira Filho, 1993).

O valor de mercado é normatizado pela NBR 14653/2001, como um valor único num dado instante, independente da finalidade da avaliação e subjugado a um mercado de concorrência perfeita. Obviamente, o mercado imobiliário não é, pela sua própria natureza, de concorrência perfeita (Dantas, 1998, p.8). Na verdade, o mercado imobiliário é um dos segmentos de mercado que mais se ajusta ao mercado teórico da

concorrência imperfeita. Isto faz com que o preço de um bem seja desviado daquele determinado teoricamente pelo mercado de concorrência perfeita (Barbosa Filho, 1988). Portanto, o que realmente se paga numa negociação imobiliária é o preço e não o valor (Dantas, 1998, p.8).

Assim na prática o máximo que se consegue é a estimação do preço médio de mercado, através de uma amostragem de preços que carregam todas as imperfeições do mercado (González, 2003). O que faz existir, portanto a necessidade da busca por técnicas que torne mais precisas as formas de se estimar o valor de um bem o aproximando ao máximo de seu valor de mercado.

1.2.3 - Mercado Imobiliário

O mercado imobiliário tem comportamento muito distinto dos mercados de outros bens. Os imóveis são considerados “bens compostos”, pois existem múltiplos atributos que despertam interesses nos agentes, impedindo a comparação direta das unidades.

A avaliação de imóveis, qualquer que seja sua finalidade, envolve a consideração de características econômicas. Avaliar é buscar o valor, e o valor de um bem é representado essencialmente pelo valor atribuído pelo mercado onde ele é transacionado.

A existência do mercado imobiliário depende da presença de três componentes: os bens imóveis disponíveis, os vendedores e os compradores. Assim sendo, o fator determinante na formação dos preços será a relação quantitativa dentre os três, onde a

situação ideal será aquela onde haja uma abundância equilibrada dentre os mesmos. Isto determinará num dado momento, um preço de equilíbrio de mercado que podemos considerar como sendo um preço justo. Este mercado, considerado como sendo de concorrência perfeita, é inatingível. O extremo a esta situação, ou seja, um mercado de concorrência imperfeita cria um desbalanço que faz os preços se afastem do ideal ou do justo. Obviamente, somente no mercado de concorrência perfeita, a construção do valor de um bem pode seguir a lei da oferta e procura (Dantas, 1998, p.9).

Dessa forma, cada bem imobiliário acabará por gerar em torno de si um micro-mercado que guardará caracteres tão intimamente relacionados que dificultará a relação deste com o macro-mercado que o circunda. Isto dificulta a avaliação do bem porque condiciona a avaliação à coleta de dados do micro-mercado considerado. Se os elementos amostrais forem insuficientes dentro do micro-mercado, a coleta de elementos do macro-mercado circundante gerará tendências de mercado que invalidariam a avaliação (Trivelloni, 1998).

1.2.4 - Métodos de Avaliação

Os métodos de avaliação podem ser divididos em dois grupos: métodos diretos e indiretos. Os métodos diretos baseiam-se na comparação de dados de transação semelhantes no mercado ou na determinação do custo para reprodução de bem, enquanto que os indiretos são baseados na renda que o imóvel pode proporcionar ou em sua capacidade de aproveitamento (González, 2003). Esquemáticamente:

Métodos diretos

- a) Comparativos de dados de mercado;
- b) Evolutivo (custo de reprodução).

Métodos indiretos

- a) Renda;
- b) Involutivo.

A utilização dos métodos diretos tem preferência e sempre que existem dados do mercado suficientes para utilização do método comparativo ele deve ser escolhido (Dantas, 1998, p.15). Considerando isso o método adotado nesse trabalho é o direto de comparação de dados do mercado e por isso restringiremos a conceitualizar esse método.

No método comparativo de dados de mercado, o valor do bem é avaliado por comparação com dados do mercado similares quanto às características intrínsecas e extrínsecas; para isto exige a presença de um conjunto atual de dados que represente estatisticamente o mercado. Portanto, qualquer bem pode ser avaliado por este método, desde que existam dados suficientes e atuais no mercado imobiliário que possam ser utilizados para representá-los estatisticamente (Trivelloni, 1998, p.2000).

Quando a questão é avaliação de imóveis, o método mais utilizado e recomendado é o método comparativo de dados de mercado, já que este método permite que a estimativa considere as diferentes tendências do mercado imobiliário que, por sua vez, diferenciam-se das tendências de outros ramos da economia. Este método estima valores baseado na comparação com outros semelhantes, partindo-se de um grupo de dados somado às informações sobre transações e ofertas do mercado,

e originando com isto uma amostra de dados do mercado imobiliário. Na prática, de modo geral, a semelhança entre o imóvel avaliado e os componentes da amostra é imperfeita e incompleta, por faltar algum atributo que tenha influenciado no valor ou por apresentá-lo de forma parcial. Portanto, os atributos dos dados pesquisados que influenciam o valor devem ser ponderados por homogeneização ou inferência estatística, respeitando os níveis de rigor definidos na NBR-14653/2001.

1.2.5 - Amostragem em Mercado Imobiliário

Em geral, as técnicas de amostragem tradicional não se aplicam ao mercado imobiliário. Não existem listas completas nem formas de enumerar todos os elementos participantes do universo em análise. As amostragens por área ou zona também não são de maior interesse, pois as avaliações focam um local específico (o da situação do imóvel avaliado). Assim, busca-se a maior quantidade de elementos possível, situados na mesma região e com as mesmas características do avaliado. As amostras são sempre do tipo “julgadas”, ou seja, o profissional escolhe os dados, dentro de seus critérios, buscando obter a melhor amostra possível, mas sem utilizar critérios científicos rigorosos. Não há aleatoriedade ou chance igual de seleção para todos os elementos do universo, por exemplo, (González, 2003).

1.3 - JUSTIFICATIVA

Os indivíduos valorizam os diversos atributos da habitação de forma diferenciada e, de modo geral, os preços implícitos de cada característica ou serviço da habitação tendem a variar diretamente com a oferta dos mesmos. Contudo, não existe consenso entre os pesquisadores sobre quais variáveis ou característica dos imóveis devem ser incluídas nos modelos, nem sobre a forma destes modelos (formato da equação). Ao contrário, existe uma grande diversidade de abordagens e, na verdade, os dados e formatos empregados dependem mais da disponibilidade de informações do que das decisões técnicas.

Para que haja objetividade na avaliação, ou ainda, para que a estimativa não seja feita com base no “eu acho” do avaliador deve se construir um procedimento estatístico adequado, buscando um modelo matemático através da estatística inferencial de forma que possa prever o valor de um imóvel em função de suas características.

A principal justificativa para a realização deste trabalho é a possibilidade de construir, através de uma metodologia científica um modelo que possa fornecer uma estimativa aceitável, não distante da realidade, do valor de um terreno urbano na região centro - norte do município de Betim. A importância desse trabalho é a procura pelo melhor modelo para esses dados e a determinação das variáveis que mais influenciam no preço unitário de um imóvel dessa região.

Os métodos de desenvolvimento do trabalho seguem as etapas que constam na ABNT 14653-2/2001. A primeira etapa consiste no conhecimento do bem avaliado,

nessa etapa é definido que tipo de imóvel vai ser avaliado, ou seja, delimitar o campo de pesquisa.

A segunda etapa é a preparação da pesquisa, planejar a execução, indicar formas de coletas dos dados, julgar a melhor, e só então na terceira etapa compor uma amostra representativa de dados do mercado de imóveis com características, tanto quanto possível, semelhantes as do avaliado, usando-se todas as evidências definíveis.

Seguindo o planejamento após a coleta defini-se o quarto passo que é a sumarização e a descrição dos dados, que compreende a organização do banco de dados o resumo do mesmo, bem como a análise e interpretações numéricas e gráficas, envolvendo cálculo de medidas de resumo e de posição. O refinamento do modelo empírico é realizado através do estudo da inferência estatística aplicada à análise de regressão, testes de hipóteses e intervalos de confiança.

Na quinta etapa é feita à análise da adequação do modelo e sua interpretação.

E por fim, como sexta etapa tem a explicação e discussão dos resultados.

1.4 – OBJETIVOS

1.4.1 – Objetivo Geral

O objetivo geral é descrever a construção de um ou mais modelos de regressão, que possa predizer de forma adequada o valor de um terreno urbano em função de

suas principais características. O procedimento é ilustrado com um banco de dados de imóveis provenientes da cidade de Betim, MG.

1.4.2 – Objetivos Específicos

- 1 - Identificar as variáveis importantes para o modelo;
- 2 – identificar um ou mais modelos adequados para o caso prático de interesse;
- 3- Avaliar e comparar os modelos com relação ao seu poder de predição.

1.4.3 – Delimitações da Pesquisa

Existem várias técnicas para a avaliação de imóveis, restringiu-se nesta pesquisa, ao uso dos modelos clássicos de regressão.

A aplicação da estratégia de construção de um modelo de regressão, proposto nesta monografia, restringe-se aos terrenos urbanos da região centro-norte do município de Betim - Minas Gerais, os dados foram obtidos pela própria pesquisadora, usando o banco de dados imobiliário de Betim-MG, em confronto com o cadastro urbano do município.

1.5 – ESTRUTURA DO TRABALHO

A monografia está estruturada em quatro capítulos, construídos de forma a facilitar o entendimento e compreensão do leitor desde os objetivos até a conclusão.

O Capítulo 2 apresenta a metodologia estatística necessária, de forma detalhada, da construção até o diagnóstico do modelo, para realização da análise de dados com a finalidade de atingir os objetivos do trabalho.

O Capítulo 3 descreve um roteiro estratégico para a construção do modelo e faz a análise dos dados, do caso dos terrenos urbanos da cidade de Betim-MG. Descreve-se e verifica-se detalhadamente os resultados e a utilização para novas predições.

O Capítulo 4 apresenta uma conclusão com base na metodologia e resultados obtidos nos capítulos anteriores.

METODOLOGIA ESTATÍSTICA

2.1- CONCEITUAÇÃO

Em muitas situações, duas ou mais variáveis estão inerentemente relacionadas, sendo necessário explorar a natureza dessa relação. Análise de regressão é uma técnica estatística para modelar e investigar a relação entre uma resposta e uma ou mais variáveis (Montgomery, 2003)

O modelo Clássico de Regressão teve origem nos trabalhos de astronomia elaborados por Gauss no período de 1809 a 1821. Atualmente a análise de regressão é uma das metodologias estatísticas mais utilizadas nas pesquisas científicas.

Em Engenharia de Avaliações considera-se geralmente como variável dependente o preço a vista dos dados de mercado em oferta ou efetivamente transacionados, e como variáveis independentes as respectivas características decorrentes dos seus aspectos físicos e de localização, bem como de aspectos econômicos (dado de oferta ou transação, à época de ocorrência do evento, etc.). Observa-se que as variáveis independentes podem ser tanto de natureza quantitativa (área, testada etc.), como de ordem qualitativa (infra-estrutura urbana, localização, etc.) (Dantas, 1998, p.95).

2.2 - MODELO DE REGRESSÃO LINEAR-NORMAL

O modelo de regressão linear múltipla descreve uma variável dependente Y como função de várias variáveis regressoras ou independentes. O modelo geral, com p variáveis regressoras para uma amostra de tamanho n , é dado por:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i \quad (i=1, \dots, n),$$

em que Y_i representa as observações da chamada variável dependente ou variável resposta, X_{ik} são chamadas de variáveis independentes ou variáveis explicativas, β_i são os parâmetros da população a serem estimados e ε_i são os erros aleatórios do modelo.

A representação do modelo na forma matricial é $Y = X\beta + \varepsilon$, em que

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_p \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} .$$

Em geral, y é um vetor ($n \times 1$) das observações, X é uma matriz ($n \times p+1$) dos valores das variáveis independentes, β é um vetor ($p+1 \times 1$) dos coeficientes de regressão e ε é um vetor ($n \times 1$) dos erros aleatórios, com media zero e variância σ^2 .

Um dos objetivos da análise de regressão é desenvolver uma equação que permita ao investigador prever respostas para valores dados de variáveis explicativas.

Para descrever a equação é necessário estimar os coeficientes de regressão β e a variância σ^2 do erro a partir dos dados observados.

Existem alguns métodos para estimação destas quantidades, no entanto os mais utilizados são o método dos mínimos quadrados e o método da máxima verossimilhança (ver Draper e Smith, 98). O método dos mínimos quadrados consiste em encontrar uma estimativa para os parâmetros de forma que a soma do quadrado dos erros seja mínima.

O estimador de mínimos quadrados para o vetor de parâmetros β é dado por (Draper e Smith, 98):

$$\hat{\beta} = (X'X)^{-1}X'Y .$$

As suposições exigidas para o modelo de regressão linear múltipla, são as seguintes:

1) os erros são variáveis aleatórias com as seguintes propriedades:

- valor esperado zero – $E(\varepsilon_i) = 0$, $i = 1, \dots, n$

- variância constante - $\text{Var}(\varepsilon_i) = \sigma^2$, $i = 1, \dots, n$

- não correlacionados – $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$

2) a distribuição dos erros é normal, $\varepsilon_i \approx N(0, \sigma^2)$. Como os erros são não correlacionados, pode-se afirmar, sob a hipótese de normalidade, que estes são independentes.

2.3 -TESTE DE HIPÓTESES PARA A REGRESSÃO LINEAR MÚLTIPLA

O teste de hipótese é uma regra usada para decidir se uma hipótese estatística deve ser rejeitada ou não com base nos dados. O objetivo desse teste é decidir se uma hipótese sobre determinada característica da população é ou não apoiada pela evidência obtida de dados amostrais.

2.3.1 - Teste de Hipótese Para a Significância do Modelo

O teste para a significância do modelo de regressão é um teste para determinar se existe uma relação linear entre a variável de resposta y e um subconjunto de repressores x_1, x_2, \dots, x_k . As hipóteses apropriadas são (Montgomery, 2003):

$$\begin{aligned} H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1: \beta_k \neq 0 \text{ para pelo menos um } k \end{aligned}$$

A estatística de teste para H_0 é:

$$F_0 = \frac{SQ_R/p}{SQ_E/(n-p-1)} = \frac{QM_R}{QM_E}$$

em que, SQ_R é a soma quadrática da regressão, dada por $SQ_R = SQ_T - SQ_E$, SQ_E é a soma quadrática do erro, dada por $SQ_E = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ e SQ_T é a soma quadrática total, dada por $SQ_T = \sum_{i=1}^n (Y_i - \bar{Y})^2$.

A estatística do teste F_o , sob H_0 , tem a distribuição F com p e $(n-p-1)$ graus de liberdade. A rejeição da hipótese H_0 indica que pelo menos um regressor está associado com a resposta de forma linear.

2.3.2 - Teste de Hipótese Para o Parâmetro β_k

Após a verificação de que pelo menos um dos parâmetros β_k é significativo, deve-se testar a significância de cada um deles, isto é, para cada parâmetro β_k ($k=1, \dots, p$), testam-se as hipóteses:

$$\begin{aligned} H_0: \beta_k &= 0, \\ H_1: \beta_k &\neq 0. \end{aligned}$$

A estatística de teste para H_0 é:

$$t = \frac{\hat{\beta}_k}{\hat{s}(\hat{\beta}_k)}$$

em que $\hat{s}(\hat{\beta}_k)$ é o erro padrão estimado do coeficiente $\hat{\beta}_k$, obtido a partir do resultado que $\text{Var}(\hat{\beta}) = \sigma^2 (X'X)^{-1}$, e que $\sigma^2 = \text{QM}_E$.

Se $|t| > t_{(\alpha/2; n-p-1)}$, o teste rejeita H_0 . A rejeição de H_0 indica uma contribuição significativa da variável independente X_k no modelo.

2.4 - INTERVALOS DE CONFIANÇA

Os intervalos de confiança mais utilizados em uma análise de regressão são descritos brevemente a seguir.

2.4.1-Intervalo de Confiança Para o Parâmetro β_k

O intervalo de confiança para β_k , é dado por:

$$\hat{\beta}_k - t_{(1-\alpha/2; n-p-1)} \hat{S}(\hat{\beta}_k) ; \hat{\beta}_k + t_{(1-\alpha/2; n-p-1)} \hat{S}(\hat{\beta}_k)$$

2.4.2 - Intervalo de Confiança Para Valores Médios Preditos

O valor médio estimado para um caso (imóvel) i representado pelas suas variáveis explicativas x_i é dado por $\hat{y}_i = X_i' \hat{\beta}$. O intervalo de confiança para o valor médio estimado é calculado por:

$$(\hat{y}_i - t_{(1-\alpha/2; n-p-1)} \hat{S}(\hat{y}_i) ; \hat{y}_i + t_{(1-\alpha/2; n-p-1)} \hat{S}(\hat{y}_i)),$$

em que $\hat{S}(\hat{y}_i)$ é o desvio-padrão de \hat{y}_i , que é dada por:

$$\hat{S}^2(\hat{y}_i) = (QM_E) X'_i (X'X)^{-1} X_i$$

2.4.3 - Intervalo de Confiança Para Predição de Novas Observações

O valor predito para um novo caso observado h é dado por $\hat{y}_h = X'_h \beta$. E o intervalo de confiança para este novo caso observado é dado por:

$$(\hat{y}_h - t_{(1-\alpha/2; n-p-1)} \cdot \hat{S}(\hat{pred}); \hat{y}_h + t_{(1-\alpha/2; n-p-1)} \cdot \hat{S}(\hat{pred})),$$

em que $\hat{S}(\hat{pred})$ é o desvio-padrão do valor predito, que é dado por:

$$\hat{S}^2(\hat{pred}) = QM_E (1 + X'_h (X'X)^{-1} X_h).$$

2.5 - MEDIDAS DE ADEQUAÇÃO DO MODELO

Nesta seção será apresentada uma medida de adequação global de ajuste, o R^2 , e técnicas de verificação, através dos resíduos, de validação das suposições do modelo.

2.5.1 - Coeficiente de Determinação (R^2)

O coeficiente de determinação múltipla R^2 é definido como

$$R^2 = \frac{SQ_R}{SQ_T} = \frac{1 - SQ_E}{SQ_T} .$$

R^2 é uma medida de porcentagem da variabilidade da resposta, explicada pelos regressores. Entretanto, um valor grande de R^2 ($0 \leq R^2 \leq 1$) não implica necessariamente que o modelo de regressão seja bom. A adição de uma variável ao modelo sempre aumentará R^2 , independente da variável adicional ser ou não estatisticamente significativa (Montgomery, 2003). Quando se deseja comparar diferentes modelos, muitos autores preferem usar o chamado coeficiente de determinação ajustado, com um ajuste realizado para os correspondentes graus de liberdade de SQ_E e SQ_T , como definido abaixo (Draper e Smith, 1981, p.92):

$$R^2_a = 1 - (1 - R^2) (n - 1 / n - p) .$$

2.5.2 - Análise de Resíduo-Verificação das Suposições do Modelo

Com a definição do modelo, devem-se realizar a análise dos resíduos, a fim de procurar evidências sobre eventuais violações das suposições estabelecidos na seção 2.2. Apresenta-se aqui a metodologia de interpretação dos gráficos dos resíduos.

2.5.2.1 – Independência

Verifica-se a independência dos erros quando os resíduos se distribuem aleatoriamente, em torno de zero. Esta situação é ilustrada na Figura 2.1.

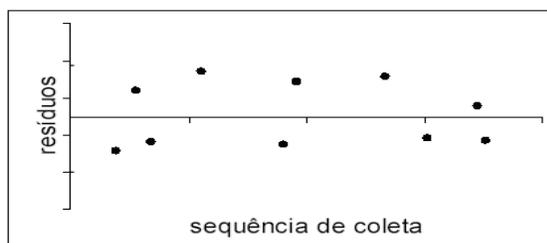


Figura 2.1 Gráfico de independência dos erros pela plotagem dos resíduos (Dados Fictícios)

Uma outra forma de verificação da existência de autocorrelação dos erros pode ser feita pela estatística de Durbin-Watson (Triola, 2005).

2.5.2.2- Variância Constante - Homocedasticidade

A suposição de variância constante ou homogeneidade de variância é verificada através dos gráficos dos resíduos versus valores preditos. Quando o gráfico produz forma de megafone, implica que a variância não é constante, conforme mostra a Figura 2.2, por outro lado, quando estão distribuídos aleatoriamente em torno de uma reta horizontal que passa pela origem, sem qualquer padrão, há indicação de variância constante.

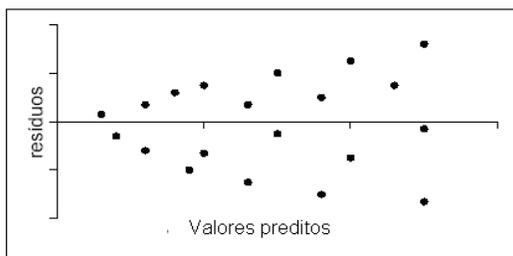


Figura 2.2 Gráfico da variância não constante. (Dados Fictícios)

2.5.2.3 - Distribuição Normal

O gráfico construído pelos resíduos ordenados versus os respectivos valores teóricos da distribuição normal, quanto mais próximo dessa reta estiverem os valores dos resíduos, maior indicação de normalidade dos dados, conforme mostra a Figura 2.3.

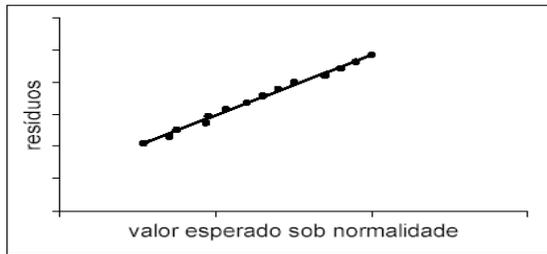


Figura 2.3: Gráfico de normalidade dos erros. (Dados Fictícios)

2.5.2.4 - Adequação do Modelo

A verificação de que o modelo de regressão ajustado é adequado aos dados pode ser detectado através dos gráficos dos valores preditos versus os valores observados. Quando estes gráficos apresentarem os pontos seguindo a reta da bissetriz, Figura 2.4, há indicação de que o modelo ajustado é adequado.

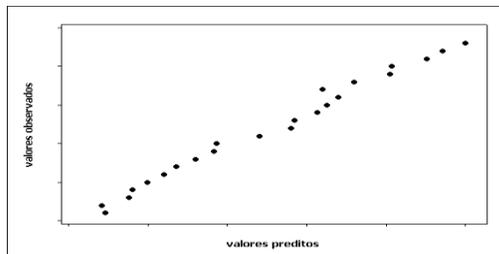


Figura 2.4: Gráfico de não adequação do modelo. (Dados Fictícios)

2.5.2.5 - Valores Discrepantes

Os gráficos dos resíduos versus variáveis independentes ou resíduos versus valores ajustados, podem indicar a presença de pontos discrepantes ou “*outliers*”. Um ponto é discrepante quando se comporta muito diferente do restante das observações e é muito maior que o resto em valor absoluto, e ainda, se estiver afastado de zero por três ou mais desvios padrões, conforme mostra a Figura 2.5.

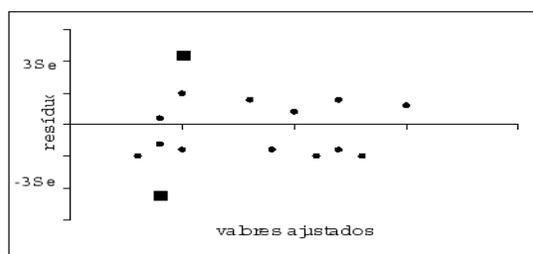


Figura 2.5: Gráfico de valores discrepantes. (Dados Fictícios)

O não atendimento de algumas ou de todas as suposições de existência do modelo, indica que o modelo não está bem ajustado, inviabilizando seu uso.

Para resolver esse problema pode se tentar transformar a variável dependente, ou tentar adequar o modelo a distribuições diferentes da normal. As distribuições log-normal, log logística e weibull, são outras possibilidades. A seguir, são apresentadas as transformações de Box Cox que indica uma possível transformação na variável resposta e também outros modelos.

2.6 - TRANSFORMAÇÃO BOX - COX

Para valores não negativos da variável y a transformação Box–Cox é dada pelas seguintes igualdades:

$$Y^\lambda = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{para } y \neq 0, \\ \ln y & \text{para } \lambda = 0. \end{cases}$$

Essa transformação pode ser particularmente útil quando uma variável resposta não cumpre com os pressupostos de normalidade e/ou homocedasticidade. A hipótese de trabalho por trás da transformação é equivalente à afirmativa de que existe algum valor de λ para o qual $y^{(\lambda)}$ tem distribuição (aproximadamente) normal com média dada pelo produto vetorial $\beta'X$ e variância constante.

Nas fórmulas anteriores observa-se que, quando $\lambda = 1$, a transformação deixa a variável praticamente inalterada (a não ser pela subtração da unidade de cada observação, cuja conseqüência é o deslocamento de toda a distribuição para esquerda nessa quantia, mas mantendo a mesma variância). Por sua vez, como já foi mencionado, $\lambda = 0$ implica na transformação logarítmica da variável resposta.

A estimativa do λ é feita pelo método da máxima verossimilhança. Como as expressões resultantes são bastante complicadas, encontra-se λ através de método gráfico, plotando o logaritmo da função de verossimilhança estimada versus valores de λ ajustados. O valor correspondente ao L_{\max} fornece a solução procurada, ou seja, o

expoente da transformação potencia indicada para a variável Y (ver Draper e Smith, 98).

2.7 - OUTROS MODELOS PARAMÉTRICOS

Muitas vezes o modelo de regressão linear-normal não é apropriado para os dados em estudo, é necessário, portanto identificar um modelo alternativo. Para utilização de métodos paramétricos é necessária a especificação de uma distribuição de probabilidade adequada para os dados em estudo.

O modelo de regressão bastante utilizado quando a variável reposta é não-negativa é dado por:

$$W = \ln (y) = X\beta + \sigma\varepsilon ,$$

onde y pode seguir as seguintes distribuições,

Log-normal: A função de densidade de uma variável aleatória y com distribuição log-normal é dada por:

$$f(y) = \frac{1}{\sqrt{2\pi y \sigma}} \exp \left\{ -\frac{1}{2} \left[\frac{\log(y) - \mu}{\sigma} \right]^2 \right\} , \quad y > 0,$$

em que μ é a media do logaritmo de y e σ é o seu desvio padrão.

Existe uma relação entre as distribuições log-normal e normal, que facilita a apresentação e análise de dados provenientes da distribuição log-normal. Esta relação significa que dados vindos de uma distribuição log-normal podem ser analisados segundo uma distribuição normal, desde que, se considere o logaritmo dos dados em vez dos valores originais (Colosimo e Giolo, 2006).

Weibull: Para uma variável aleatória y com distribuição de Weibull, tem-se a função de densidade de probabilidade dado por :

$$f(y) = \frac{t^{\gamma-1}}{\alpha^\gamma} \exp\left\{-\left[\frac{t}{\alpha}\right]^\gamma\right\}, \quad t > 0,$$

em que γ , o parâmetro de forma, e α , o de escala, são ambos positivos. O parâmetro α tem mesma unidade de medida de y e γ não tem unidade.

Log-logística: A função de densidade, para uma variável aleatória y com distribuição log-logística, é expressa por:

$$f(t) = \frac{\gamma}{\alpha\gamma} y^{\gamma-1} (1 + (y/\alpha)^\gamma)^{-2}, \quad t > 0,$$

sendo $\alpha > 0$ o parâmetro de forma e $\gamma > 0$ o de escala, com variância dado por $\pi^2 / 3 \sigma^2$.

2.7.1 - Estimando os Parâmetros do Modelo - O Método de Máxima

Verossimilhança

O método apropriado para estimar os parâmetros dos modelos descritos anteriormente é o Método da Verossimilhança. Suponha, inicialmente, uma amostra de observações y_1, \dots, y_n de uma certa população de interesse. Suponha ainda, que a população é caracterizada pela função de densidade $f(y)$. A função de verossimilhança para um parâmetro genérico θ desta população é, então, expressa por:

$$L(\theta) = \prod_{i=1}^n f(y_i; \theta) .$$

Observa-se que L é função de θ . Nesta expressão, θ pode estar representando um único parâmetro ou um conjunto de parâmetros. No modelo log-normal, por exemplo, $\theta = (\mu, \sigma)$. Busca-se o valor de θ que maximize a função $L(\theta)$. Isto é, encontrar o valor de θ que maximize a probabilidade da amostra observada ocorrer.

É conveniente trabalhar com o logaritmo da função de verossimilhança. Os estimadores de máxima verossimilhança são os valores de θ que maximizam $L(\theta)$, isto é, $\text{Log}(L(\theta))$. Eles são encontrados resolvendo-se o sistema de equações:

$$U(\theta) = \partial \text{Log}(\theta) / \partial \theta = 0 .$$

2.7.2 – Teste de Significância

Para testar as hipóteses de significância dos parâmetros da equação de regressão, além de outros, se usa o teste Wald. Considere-se a hipótese nula:

$$H_0: \theta = \theta_0$$

a estatística para esse teste é dada por:

$$W = (\hat{\theta} - \theta_0)' F(\hat{\theta}_0) (\hat{\theta} - \theta_0),$$

que, sob H_0 , tem aproximadamente uma distribuição qui-quadrado com p graus de liberdade (χ^2_p) e $F(\hat{\theta}_0)$ é a matriz de informações. A um nível 100α % de significância, valores de W superiores ao valor tabelado da distribuição $\chi^2_{1,1-\alpha}$ indicam a rejeição de H_0 .

O teste da razão de verossimilhança envolve a comparação dos valores do logaritmo da função de verossimilhança maximizada sem restrição e sob H_0 , ou seja, a comparação de $\text{Log}(\hat{\theta})$ e $\text{Log} L(\hat{\theta}_0)$. A estatística para este teste é dada por:

$$\text{TRV} = -2 [\log L(\hat{\theta}) - \log L(\hat{\theta}_0)] ,$$

que, sob $H_0: \theta = \theta_0$, segue aproximadamente uma distribuição qui-quadrado com p graus de liberdade.

CONSTRUÇÃO DO MODELO PARA AVALIAÇÃO DE IMÓVEIS – UM ESTUDO DE CASO

3.1 - ETAPAS PARA CONSTRUÇÃO DO MODELO

A construção de um modelo de regressão deve seguir alguns passos, ou etapas. Essas etapas fazem parte de um roteiro que será seguido nesse trabalho da forma seguinte.

Identificação das variáveis Independentes: Antes de realizar a pesquisa deve-se escolher as variáveis que vão compor o banco de dados, ou seja, as variáveis que vão explicar a formação do preço do imóvel (variável dependente) é fundamental para a realização de um bom modelo. Esta é a fase que se investiga o mercado imobiliário, obtendo informações importantes para a escolha dessas variáveis. De forma geral, pode se dizer que as características do imóvel e do mercado imobiliário que influenciam na formação do preço são: área total, padrão de acabamento, infra-estrutura urbana, zoneamento, acesso ao transporte urbano, transação, etc. Esta lista de variáveis explicativas tende a variar de município para município, dependendo das características de cada um.

Levantamento dos dados: O levantamento dos dados deve ser feito com muito cuidado, pois dele dependerá o sucesso da análise estatística. É importante a realização de um planejamento para a realização da coleta dos dados. Neste planejamento, o pesquisador deve definir o espaço físico, local onde está inserida a população a ser estudada, e o número de imóveis a serem pesquisados. Na maioria dos estudos é inviável utilizar todos os imóveis da população, isto por diversas razões, entre elas o tempo, o alto custo, a dificuldade na obtenção das informações e outras. No mercado de imóveis, é freqüente a entrada de novos dados, por isso, deve se fazer um novo levantamento a cada nova avaliação para garantir a representação dos novos dados na amostra (Dantas, 1998, p.50).

Descrição e Análise exploratórias das variáveis: Descrever detalhadamente a composição da variável, sua utilização na formação do modelo. Fazer uma análise exploratória das características da amostra, de maneira que possa ser medida e utilizada para inferir as características da população da qual foi extraída. Na exploração começa se o estudo do mercado e tem-se alguma idéia a respeito dos fatores que influenciam sobre a formação do preço.

Transformação das variáveis: As variáveis que são definidas em função das características e da localização de um imóvel são do tipo quantitativas e qualitativas. Geralmente, estas variáveis necessitam de transformações para que possam ser realizadas as análises. As variáveis qualitativas com mais de dois níveis devem ser representadas através de uma codificação adequada, transformadas em variáveis

indicadoras. As variáveis quantitativas podem sofrer transformações variadas, como: logarítmica, raiz quadrada, entre outras.

Construção do modelo: Um ponto fundamental no processo de modelagem é que não se deve ficar restrito a um único modelo e desprezar outros. É prudente considerar um conjunto amplo de modelos, levando em consideração a facilidade de interpretação, boas previsões e conhecimento profundo da estrutura dos dados.

Análise de resíduos: A investigação da adequação do modelo é uma etapa, do procedimento necessário na análise dos dados, tão importante quanto à sua construção. O gráfico dos resíduos é o instrumento usado para examinar o modelo. A análise gráfica dos resíduos é necessária para examinar a adequação do modelo. Ou seja, para confirmar se ele tem uma boa aproximação ao verdadeiro sistema e para verificar se as suposições inerentes ao modelo de regressão por mínimos quadrados não foram violadas (Montgomery, 1997, p.563-565).

Avaliação prática do modelo construído: Um último passo que deve ser realizado antes de adotar o modelo para avaliação de imóveis, é verificar sua aplicabilidade e sua capacidade de predição de novas observações.

3.2 – UM ESTUDO DE CASO

O presente estudo constituiu sua base de dados com terrenos urbanos da região centro-norte da cidade de Betim-MG. As variáveis são do tipo quantitativas e qualitativas e representam as características do imóvel.

O software utilizado para as todas as análises estatísticas foi o *Minitab 14*.

O banco de dados foi composto através de formulário adequado, e a pesquisa feita nas imobiliárias de Betim, classificados do jornal local e consulta na Prefeitura ao cadastro imobiliário do município.

Na Engenharia de Avaliação considera-se como variável dependente o preço praticado no mercado em reais e como variáveis independentes as respectivas características físicas e econômicas do imóvel. As variáveis independentes estão relacionadas e descritas nas Tabelas 3.1 e 3.2.

Tabela 3.1: Descrição das variáveis independentes quantitativas.

VARIÁVEIS	UNIDADE DE MEDIDA	DESCRIÇÃO
Área total (Área)	m ²	Refere-se à área total do imóvel em m ² .
Testada	m ²	Parte da via pública (estrada, rua, passeio) que fica à frente de um imóvel.

Tabela 3.2: Descrição das variáveis independentes qualitativas.

VARIÁVEIS	CATEGORIAS	DESCRIÇÃO
Localização(L)	1-Senhora das Graças	Variável classificada de acordo com o bairro em escala do pior para o melhor.
	2-Novo Horizonte	
	3-Novo Guarujá	
	4-Angola	
	5-Ingá Baixo	
	6-Ingá Alto	
	7-Jardim da Cidade	
Via Pública (VP)	1-Principal	Variável composta pelo logradouro, dividida em: Via publica principal: Avenidas, praças e corredores comerciais. Via publica secundaria: outras.
	2-Secundaria	
Topografia (TOP)	1-Declive	Refere-se ao relevo do terreno
	2- Aclive	
	3-Plana	
Infra-Estrutura (IF)	1-Completa	Composta e classificada de acordo com a presença ou não de melhorias urbanas como: água luz, esgoto e pavimentação. O terreno que possui todas as melhorias urbanas tem uma infra-estrutura completa se não possuir qualquer um dos itens tem uma infra-estrutura parcial.
	2- parcial	
Zoneamento(Z)	1-comercial	Zoneamento: regulamentação que organiza a divisão de um território (área urbana, cidade, etc.) em zonas (comercial, residencial), fixando para cada uma delas o gênero e condições de uso do solo, com o objetivo de obter um desenvolvimento ordenado ou para melhor gestão de problemas específicos.
	2- Residencial	
Transação(T)	1- oferta	Operação ou ato comercial que se encontrava o terreno no momento da pesquisa.
	2- Venda	
Data(D)	1- coletas de dados entre maio/2000 e novembro/2004	Mês e ano em que foi feita a coleta dos dados
	2- coletas de dados entre janeiro/2005 e setembro/2006	
	3- coletas de dados entre março/2007 e outubro/2007	

A variável data era composta por 17 datas diferentes, o que levaria a um número muito grande de variáveis indicadoras, assim esta variável foi categorizada novamente. Para isso foi feito um gráfico de série temporal com os dados do valor /m² e suas

respectivas datas, usando um suavizador de série chamado *Lowess* (Morettin, 2006), (ver Fgura 3.1). Esse gráfico indicou três grandes mudanças no valor do imóvel com o tempo, e a divisão ficou conforme se descreveu na Tabela 3.2.

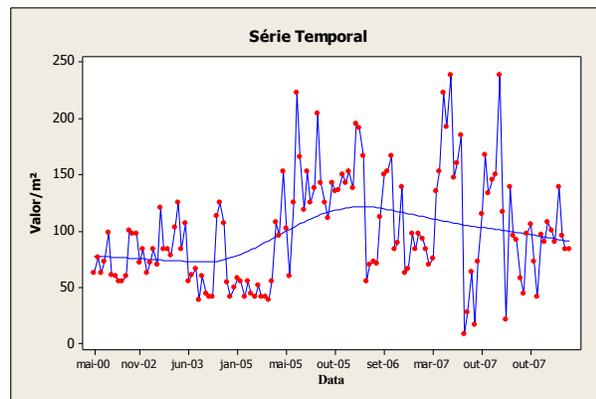


Figura 3.1: Série temporal da variável valor /m²

3.3 - ANÁLISE EXPLORATÓRIA E TRANSFORMAÇÃO DAS VARIÁVEIS

Antes de fazer a análise exploratória dos dados foi feita primeiramente a transformação das variáveis qualitativas em variáveis indicadoras, uma vez que variáveis qualitativas dizem sobre características não numéricas do imóvel sendo necessário uma categorização adequada dessas para análise. Assim as variáveis que precisaram sofrer essa transformação estão descritas na Tabela 3.3

Tabela 3.3 Transformação das variáveis qualitativas em variáveis Indicadoras (d).

VARIÁVEIS	CATEGORIAS					
	d1-L	d2-L	d3-L	d4-L	d5-L	d6-L
Localização 1	1	0	0	0	0	0
Localização 2	0	1	0	0	0	0
Localização 3	0	0	1	0	0	0
Localização 4	0	0	0	1	0	0
Localização 5	0	0	0	0	1	0
Localização 6	0	0	0	0	0	1
Localização 7	-1	-1	-1	-1	-1	-1
VARIÁVEIS	d7-VP					
Via pública 1	1					
Via pública 2	-1					
VARIÁVEIS	d8-TOP			d9-TOP		
Topografia 1	1			0		
Topografia 2	0			1		
Topografia 3	-1			-1		
VARIÁVEIS	d10-IF					
Infra Estrutura 1	1					
Infra Estrutura 2	-1					
VARIÁVEIS	d11-Z					
Zoneamento 1	1					
Zoneamento 2	-1					
VARIÁVEIS	d12-T					
Transação 1	1					
Transação 2	-1					
VARIÁVEIS	d13-D			d14-D		
Data 1	1			0		
Data 2	0			1		
Data 3	-1			-1		

Após a transformação, ou categorização das variáveis qualitativas em variáveis indicadoras, passamos a ter um banco de dados que antes eram de 9 agora com 16 variáveis.

Para a análise exploratória das variáveis quantitativas foram feitas medidas resumo como, média mediana, máximo, mínimo e desvio padrão, Tabela 3.4. E gráficos

boxplot, Figuras 3.2 e 3.3. Para as variáveis qualitativas foi feita à proporção de cada item por variável, tabela 3.5.

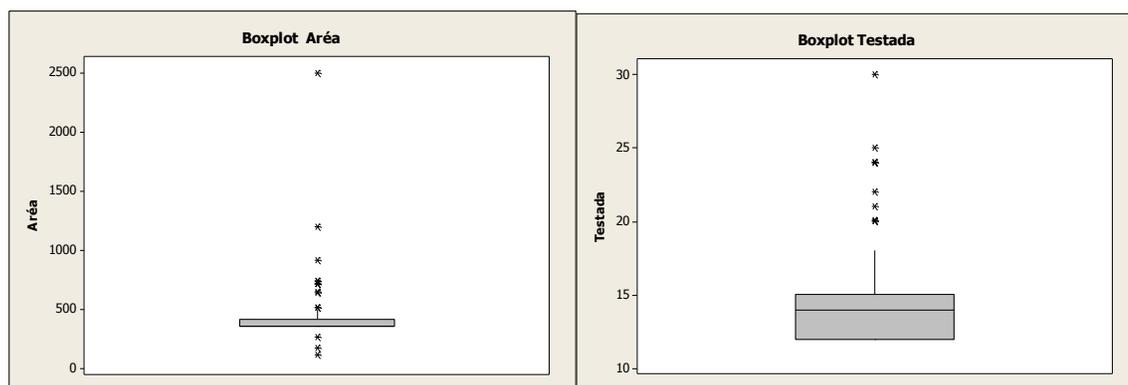


Figura 3.2: box plot para as variáveis área e testada.

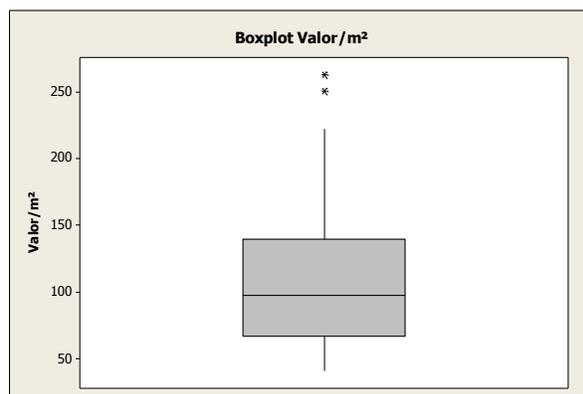


Figura 3.3 box lot pra a variável resposta (preço do imóvel).

Tabela 3.4: Tabela resumo de medidas estatísticas para as variáveis quantitativas.

	Média	Desvio Padrão	Mínimo	Máximo	Mediana
Área	452,600	280,100	120,00	2500,000	360,000
Testada	14,897	3,800	12,000	30,000	14,000
Valor/m²	107,190	47,850	41,670	262,290	97,740

Tabela 3.5: Proporção de participação de cada variável qualitativa por nível.

VARIAVEIS	Quantidade	Porcentagem
Localização 1	16	20,50%
Localização 2	6	7,70%
Localização 3	8	10,26%
Localização 4	16	20,50%
Localização 5	17	21,79%
Localização 6	8	10,26%
Localização 7	7	9,00%
VARIAVEIS		
Via publica 1	23	29,49%
Via publica 2	55	70,41%
VARIAVEIS		
Topografia 1	4	5,13%
Topografia 2	5	6,40%
Topografia 3	69	88,46%
VARIAVEIS		
Infra Estrutura 1	72	92,31%
Infra Estrutura 2	6	7,69%
VARIAVEIS		
Zoneamento 1	17	21,79%
Zoneamento 2	61	78,21%
VARIAVEIS		
Transação 1	59	75,64%
Transação 2	19	24,36%
VARIAVEIS		
Data 1	29	37,18%
Data 2	29	37,18%
Data 3	20	25,64%

3.4 - CONSTRUÇÃO DO MODELO COM A RESPOSTA ORIGINAL

A análise de regressão foi realizada com um conjunto de 78 unidades amostrais, com as variáveis qualitativas transformadas em variáveis indicadoras. A escolha das variáveis que iriam compor o modelo final foi feita manualmente, baseada no método

passo-a-passo pra trás (ver Montgomery, 2003). O modelo foi gerado com todas as dezesseis variáveis, e foi sendo tirada a variável com p-valor do teste *F* maior, usando nível de significância igual 5%. Em cada etapa calculava-se o p-valor do teste *F* para as variáveis e assim foram sendo retiradas do modelo a variáveis, uma de cada vez, até que o modelo tivesse somente variáveis com p-valor inferior a 0, 05 (ver exemplo no Anexo A).

O primeiro modelo a ser ajustado é o modelo com as variáveis originais, sem sofrer nenhum tipo de transformação. As variáveis significativas e seus respectivos valores são apresentados no Tabela 3.6.

Foram testadas também todas as interações de segunda ordem possíveis, através de forma gráfica e incluídas no modelo, juntamente com as variáveis selecionadas anteriormente, para verificar quais seriam realmente significativas ao mesmo. Nenhuma interação foi significativa ao modelo.

Tabela 3.6: Resultado final da seleção de variáveis dependentes usando as variáveis originais

Modelo de Regressão					
Preditores	coeficientes	Desvio padrão dos coeficientes	T	P	P
Constante	126,57	13,1	9,66	0,000	
d1-L	-57,323	6,025	-9,51	0,000	0,000
d2-L	-43,49	8,736	-4,98	0,000	
d3-L	5,947	7,815	0,76	0,449	
d4-L	71,262	6,353	11,22	0,000	
d5-L	17,92	8,098	2,21	0,030	
d6-L	-17,082	7,852	-2,18	0,033	
d7-VP	-14,68	3,427	-4,28	0,000	
d12-T	9,606	3,508	2,74	0,008	
d13-D	-9,525	5,044	-1,89	0,063	0,000
d14-D	-7,885	4,415	-1,79	0,079	
Área	0,05064	0,01602	3,16	0,002	
Testada	-3,737	1,207	-3,1	0,003	
R² =80,8%		R-Sq(adj) =77,3%			

A adequação do modelo foi verificada através da análise dos gráficos de resíduos, conforme Figuras 3.4 e 3.5.

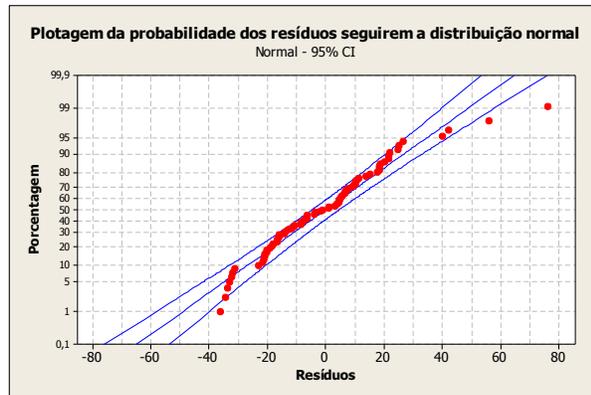


Figura 3.4: Gráfico de probabilidade dos resíduos seguirem a distribuição normal

Na Figura 3.4, o gráfico mostra que os pontos estão dispostos sob uma linha reta, indicando normalidade dos erros.

A Figura 3.5 apresenta os pontos seguindo uma forma de “megafone”, a variância começa pequena e vai aumentando. Esta disposição dos pontos indica que a violação da suposição de variância constante.

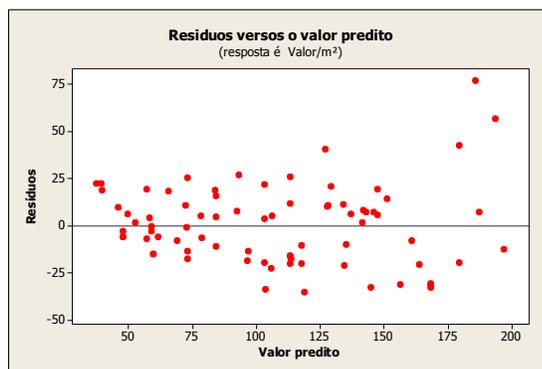


Figura 3.5: Gráfico de dispersão dos dados

O não atendimento a essa suposição importante, variância não constante, pode invalidar a inferência estatística.

3.5 - CONSTRUÇÃO DO MODELO COM A RESPOSTA TRANSFORMADA

Para tentar resolver esse problema podemos realizar transformações na variável dependente ou utilizar outros modelos. As transformações são necessárias, principalmente quando ocorrem dois problemas: afastamento da normalidade e variância não constante para o erro. A transformação logarítmica é preferida quando se procuram ajustar modelos à dados de valores imobiliários. Outro aspecto importante é que a transformação logarítmica na variável explicada torna o modelo multiplicativo, característica esta sugerida pelas próprias normas brasileiras que versam sobre avaliações.

Para verificar se é mesmo a transformação logarítmica mais indicada também para o banco de dados em estudo, foi realizada a transformação Box-Cox.

A Figura 3.6, mostra o gráfico da transformação Box Cox correspondente para a variável dependente valor /m², observa que $\lambda = -0,204956$, pode se dizer que seja aproximadamente igual à zero, portanto a transformação indicada seria a logarítmica.

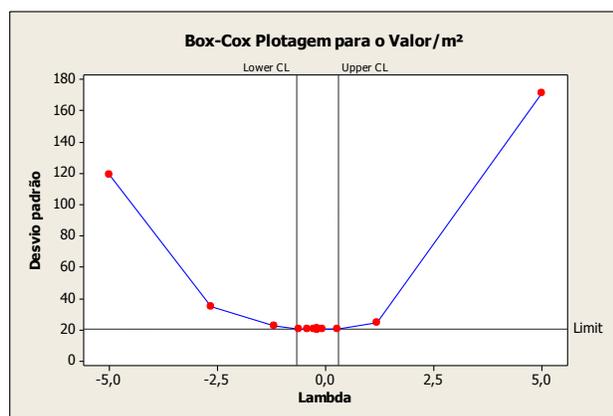


Figura 3.6: Gráfico da transformação Box Cox da variável valor /m²

A transformação logarítmica foi utilizada na variável dependente valor/m² e o resultado da seleção das variáveis para composição do modelo encontram-se na Tabela 3.7.

Tabela 3.7 :Resultado final da seleção de variáveis dependentes com a transformação logarítmica na variável dependente.

Modelo de Regressão					
Preditores	coeficientes	Desvio padrão dos coeficientes	T	P	P
constante	4,8354	0,1046	46,23	0,000	
d1-L	-0,600	0,048	-12,460	0,000	0,000
d2-L	-0,498	0,070	-7,140	0,000	
d3-L	0,123	0,062	1,980	0,052	
d4-L	0,598	0,051	11,790	0,000	
d5-L	0,182	0,065	2,810	0,006	
d6-L	-0,071	0,063	-1,140	0,259	
d7-VP	-0,108	0,027	-3,930	0,000	
d12-T	0,068	0,028	2,410	0,019	
d13-D	-0,073	0,040	-1,810	0,074	0,000
d14-D	-0,076	0,035	-2,160	0,035	
Área	0,000	0,000	3,450	0,001	
Testada	-0,037	0,010	-3,830	0,000	
R² =86,0%			R-Sq(adj) =83,4%		

A transformação logarítmica na variável dependente valor /m², aumentou o R², e provocou melhoras nos gráficos de resíduos (Figuras 3.7 e 3.8).

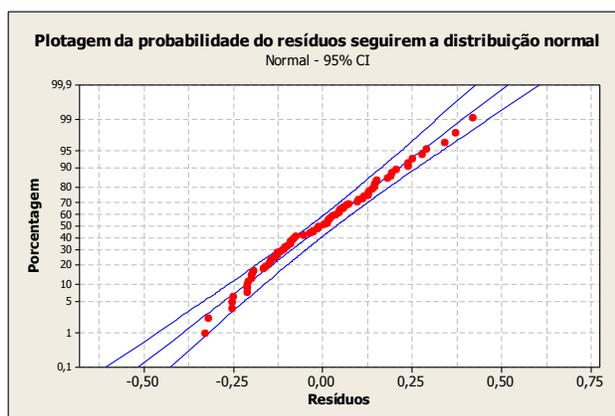


Figura 3.7: Gráfico de probabilidade da distribuição normal.

Na Figura 3.7 o gráfico mostra que os pontos estão dispostos sob uma linha reta, indicando normalidade dos erros. A Figura 3.8 apresenta os pontos aleatoriamente distribuídos em torno da linha que passa pelo ponto zero. Esta disposição dos pontos indica que a suposição de variância constante foi validada.

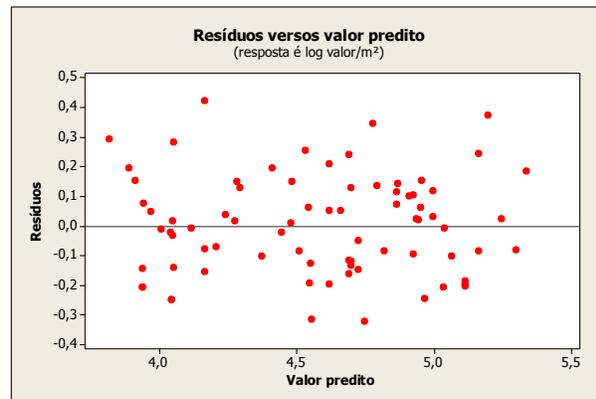


Figura 3.8 : Dispersão dos dados

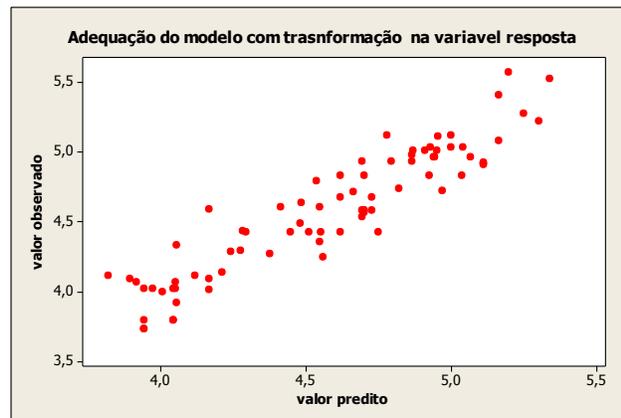


Figura 3.9: Adequação do modelo com transformação na variável resposta

A Figura 3.9, mostra que os dados estão bem alinhados, indicando que o modelo é adequado.

3.6 – OUTROS MODELOS

Pode-se também avaliar a possibilidade dos dados originais não seguirem uma distribuição normal, mais outras como: Log-normal, Weibull, Log-logística, entre outras.

3.6.1 - Modelo Utilizando a Distribuição Log - normal

É muito freqüente o uso da distribuição log-normal em Engenharia de Avaliações. Isso acontece porque os preços considerados em escala logarítmica, na maioria das vezes apresentam características mais aderentes à distribuição normal que em escala original. Uma vez que os preços observados abrangem apenas campo dos reais positivos, que coincide com o mesmo intervalo de variação desta distribuição.

Foi feito o mesmo processo de escolha das variáveis utilizado no modelo com transformação logarítmica, observou-se que as mesmas variáveis que foram significativas aos modelos anteriores foi também significativas pra o modelo Log-normal, o resultado final encontra-se na Tabela 3.8.

Tabela 3.8: Resultado da seleção das variáveis para o modelo com distribuição Log-normal

Modelo de Regressão					
Preditores	coeficientes	erro	Z	P	P
Intercepto	4,83544	0,095478	50,64	0	
d1-L	-0,600	0,0439	-13,650	0,000	0,000
d2-L	-0,498	0,0637	-7,820	0,000	
d3-L	0,123	0,0570	2,170	0,030	
d4-L	0,598	0,0463	12,910	0,000	
d5-L	0,182	0,0590	3,080	0,002	
d6-L	-0,071	0,0572	-1,250	0,212	
d7-VP	-0,108	0,0250	-4,310	0,000	
d12-T	0,068	0,0256	2,640	0,008	
d13-D	-0,073	0,0368	-1,990	0,047	0,000
d14-D	-0,076	0,0322	-2,360	0,018	
Área	0,0003351	0,0001	3,780	0,000	
Testada	-0,037	0,0088	-4,200	0,000	
Escala	0,166	0,0133			

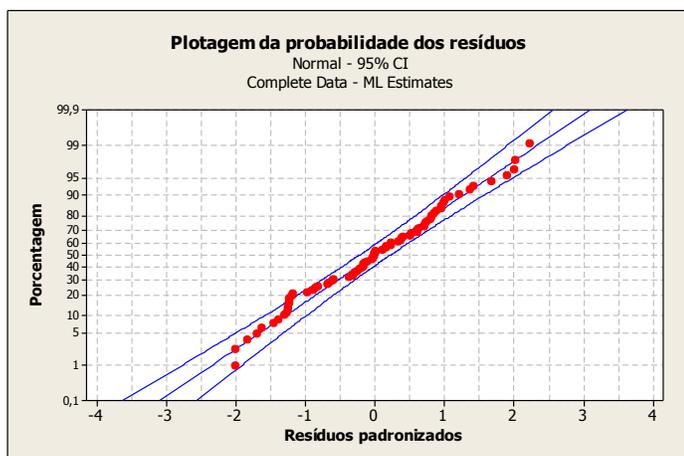


Figura 3.10: Gráfico de resíduos do modelo com distribuição log-normal

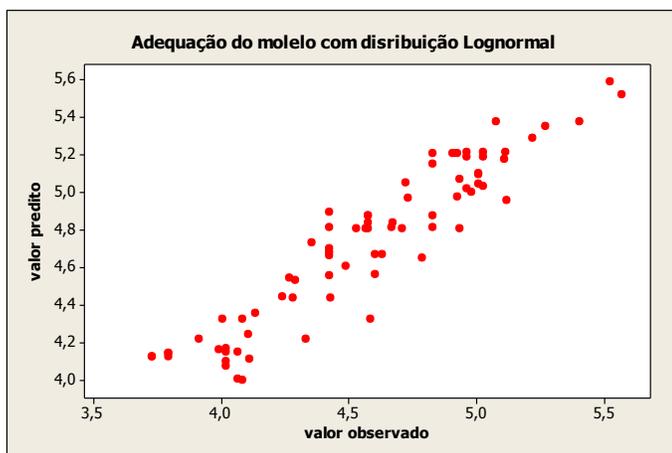


Figura 3.11: Gráfico de adequação do modelo com distribuição log-normal

A Figura 3.10 mostra a adequação do modelo log-normal. A Figura 3.11 indica que o modelo parece adequado.

3.6.2. - Modelo Utilizando Distribuição Log-logística

A distribuição log-logística, em muitas situações prática, tem se apresentado como uma alternativa a de Weibull e a log-normal. O resultado do modelo ajustado com essa distribuição encontra-se na Tabela 3.9.

Tabela 3.9: Resultado da seleção das variáveis para o modelo com distribuição log-logística

Modelo de Regressão					
Preditores	coeficientes	erro	Z	P	P
Intercepto	4,86254	0,099544	48,85	0,000	
d1-L	-0,606	0,046	-13,310	0,000	0,00
d2-L	-0,491	0,069	-7,150	0,000	
d3-L	0,131	0,053	2,470	0,013	
d4-L	0,581	0,047	12,430	0,000	
d5-L	0,179	0,064	2,800	0,005	
d6-L	-0,062	0,061	-1,020	0,310	
d7-VP	-0,096	0,025	-3,800	0,000	
d12-T	0,061	0,026	2,330	0,020	
d13-D	-0,080	0,039	-2,040	0,041	0,00
d14-D	-0,070	0,032	-2,190	0,029	
Área	0,0002971	0,000	4,040	0,000	
Testada	-0,039	0,009	-4,190	0,000	
Escala	0,096	0,009			

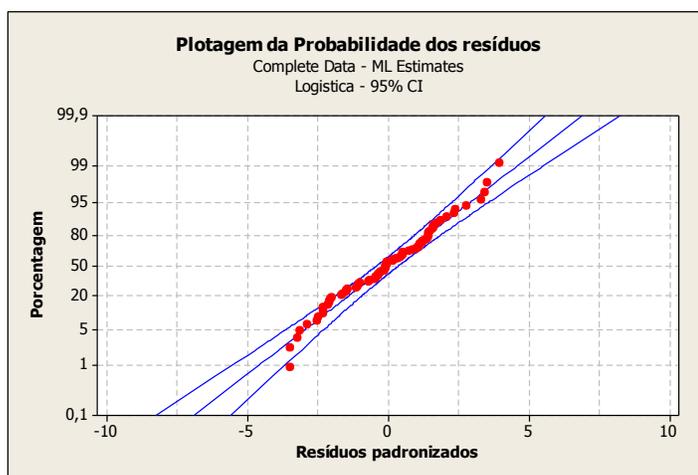


Figura 3.12: Gráfico de resíduos do modelo com distribuição log-logística.

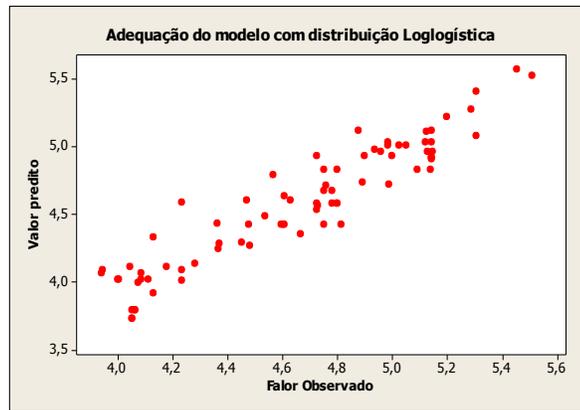


Figura 3.13: Gráfico de adequação do modelo com distribuição log-logística.

Os dados, segundo a Figura 3.12, seguem a distribuição log-logística. A Figura 3.13 indica que o modelo parece adequado.

3.6.3 - Modelo Utilizando Distribuição Weibull

A distribuição de Weibull é amplamente aplicada na área biomédica e industrial. Na Tabela 3.10 o resultado da seleção das variáveis para este modelo.

Tabela 3.10: Resultado final da seleção de variáveis com distribuição Weibull

Modelo de Regressão					
Preditores	coeficientes	erro	Z	P	P
Intercepto	4,837	0,098	49,550	0,000	
d1-L	-0,587	0,041	-14,160	0,000	0,000
d2-L	-0,470	0,063	-7,400	0,000	
d3-L	0,054	0,057	0,940	0,348	
d4-L	0,632	0,048	13,180	0,000	
d5-L	0,227	0,060	3,790	0,000	
d6-L	-0,084	0,054	-1,540	0,124	
d7-VP	-0,131	0,022	-5,870	0,000	
d12-T	0,099	0,023	4,320	0,000	
d13-D	-0,084	0,041	-2,040	0,042	0,000
d14-D	-0,096	0,034	-2,830	0,005	
Área	0,000	0,000	3,310	0,001	
Testada	-0,033	0,009	-3,640	0,000	
Forma	6,374	0,559			

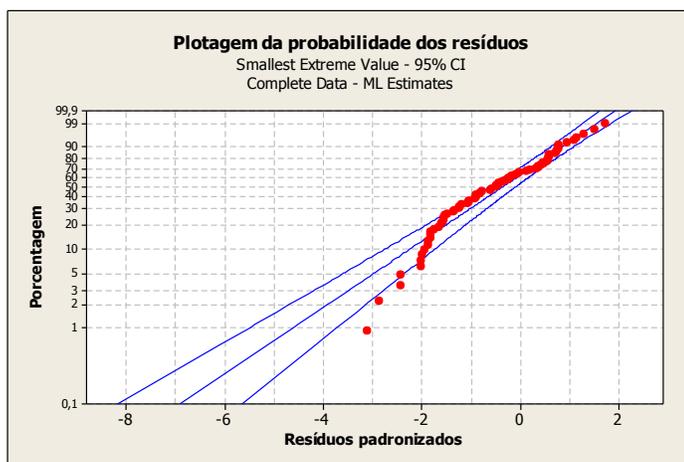


Figura 3.14: Gráfico de resíduos do modelo com distribuição Weibull

O gráfico de resíduo, Figura 3.14, parece mostrar a não adequação do modelo weibull. Por isso a modelo com a distribuição Weibul, não será testado para predição de novas observações.

Os três modelos, normal com transformação logarítmica, log-normal e log-logística foram constantes na escolha das variáveis explicativas. As variáveis importantes foram: localização via pública, transação, data, área e testada.

Para comparar esses modelos podemos utilizar também critérios como AIC (Critério de informação) e o BIC (Critério de informação Bayesiano). O modelo com transformação logarítmica não pode ser comparado nesse caso porque utiliza a variável repostada com transformação. Os valores de AIC e BIC quanto menores melhor. A Tabela 3.11 apresenta os resultados encontrados, notemos que o melhor modelo é o modelo com distribuição log-normal.

Tabela 3.11: Medidas de AIC e BIC

MODELO	AIC	BIC
Com distribuição Log-normal	681,538	768,700
Com distribuição Log-logística	684,600	771,800

3.7 - COMPARAÇÃO DOS MODELOS PARA USO PRÁTICO

3.7.1 – Interpretação do Modelo

As variáveis que foram importantes para os modelos são: Área, Testada, Via Pública (VP), Transação (T), Data (D) e Localização (L).

A seguir serão apresentadas as equações dos modelos com transformação logarítmica na resposta, com distribuição Log - Normal e distribuição Log – Logística, respectivamente.

Valor = Exp {126,57 – 57,323 d1-L – 43,49 d2-L + 5,947 d3-L + 71,262 d4-L +17,92 d5-L -17,082 d6 – L -14,68 VP + 9,606 d12T – 9,525 d13-D – 7,885 d14-D + 0,05064 área -3,737 testada}.

Valor = Exp {4,83544 – 0,600 d1-L – 0,498 d2-L+ 0,123 d3-L + 0,598 d4-L + 0,182 d5-L - 0,071d6-L + 0,108 d7-VP + 0,068 d12-T - 0,073 d13-D - 0,076 d14-D + 0,0003351 área-0,037 testada + 0,166}.

Valor = Exp {4,86254 – 0,606 d1-L - 0, 491d2-L +0,131 d3-L + 0,581 d4-L + 0,179 d5-L - 0,062 d6-L - 0,096 d7-VP +0,061d12-T - 0,080 d13-D - 0,070 d14-D + área 0,0002971 -0,039 testada + 0,098}.

Conforme se podem observar os coeficientes de todos os modelos possuem os mesmos sinais, portanto são interpretados na mesma direção. Ou seja, podemos fazer as seguintes interpretações:

- Como o coeficiente da área é positivo á medida que a ela aumenta o valor do imóvel aumenta;
- Aumentando a testada o valor do imóvel diminui;

- Quando a via pública é secundária o valor do imóvel aumenta, enquanto se a via pública é principal o mesmo diminui;
- Se o imóvel está em oferta ele tem o valor maior do que quando ele já foi transacionado;
- Os imóveis ofertados ou vendidos entre as datas maio/2004 e novembro/2004 e janeiro/2005 e setembro/2006 tem valor menor que os vendidos ou ofertados entre março/2007 e outubro/2007.

3.7.2 – Avaliação da Qualidade de Predição dos Modelos

Um modelo que atende os pressupostos para sua existência, não necessariamente é um bom modelo para predições de novas observações, ponto mais importante para Engenharia de Avaliação. Uma avaliação prática do modelo mostrará a sua capacidade de predição e sua qualidade de ajuste.

Para verificar a qualidade do modelo para predição de novas observações, foram separadas aleatoriamente da amostra oito imóveis, conforme mostrada na Tabela 3.12. Os resultados das predições para cada modelo encontram-se na Tabela 3.13.

Tabela 3.12: Novas observações para testar qualidade de predição

VALOR OBSERVADO	LOCALIZAÇÃO						VIA PUBLICA	TRANSAÇÃO	DATA		ÁREA	TESTADA
173,6427	0	0	0	1	0	0	1	-1	-1	-1	368	13
186,4196	0	0	0	1	0	0	-1	1	0	1	480	14
121,3283	0	0	0	0	1	0	-1	1	0	1	360	12
89,59504	0	0	0	0	0	1	-1	1	1	0	420	14
175,7039	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	400	13
54,54358	1	0	0	0	0	0	-1	1	0	1	360	12
55,19101	1	0	0	0	0	0	-1	1	1	0	420	14
140,3445	0	0	1	0	0	0	-1	1	-1	-1	420	14

Tabela 3. 13: Predição de novas observações pelas equações ajustadas.

Modelo com Transformação Logarítmica						Modelo com Distribuição Log-normal					Modelo com Distribuição Log-logística				
Valor observado	Valor predito	Intervalo de predição		desvio	% erro	Valor predito	Intervalo de predição		desvio	% erro	Valor predito	Intervalo de predição		desvio	% erro
173,64	165,23	119,73;	251,84	8,41	0,05	202,29	142,55;	287,06	28,64	16,50	190,50	132,24;	274,42	16,86	9,71
186,42	196,14	130,8;	265,63	9,72	0,05	217,18	156,14;	302,09	30,76	16,50	204,09	144,40;	288,46	17,67	9,48
121,33	125,00	82,72;	178,00	3,67	0,03	141,34	98,61;	202,59	20,01	16,49	133,71	91,44;	195,52	12,38	10,20
89,60	88,89	61,95;	129,60	0,71	0,01	104,37	74,29;	146,63	14,78	16,50	96,08	66,50;	138,82	6,49	7,24
175,70	165,97	121,13;	254,88	9,73	0,06	204,70	142,81;	293,40	28,99	16,50	189,58	130,95;	274,46	13,88	7,90
54,54	55,56	38,29;	77,80	1,02	0,02	63,60	45,72;	88,47	9,05	16,60	59,46	41,94;	84,29	4,91	9,01
55,19	58,33	38,30;	78,82	3,14	0,06	64,30	45,31;	91,25	9,11	16,50	58,78	40,52;	85,27	3,59	6,51
140,34	150,00	97,79;	201,42	9,66	0,07	163,50	116,37;	229,71	23,16	16,50	152,21	107,15;	216,21	11,87	8,46

Os intervalos de predição foram calculados usando nível de significância igual 5%. Observa-se, na Tabela 3.13, que os valores preditos pelo modelo com transformação logarítmica na variável dependente valor/m², para valores dos terrenos urbanos de Betim, apresentam desvios individuais baixos, e, em média o erro ficou em torno de 4%, ficando o modelo com distribuição log-normal, com erro médio de 16,51% e modelo com distribuição log-logística com erro médio de 8,56%. Isto permite concluir que as equações estão bem ajustadas para predições de novas observações, assim sugere também as Figura 3.15, 3.16 e 3.17, os dados seguem quase uma reta, o que indica que são bem próximos e possuem erros pequenos.

Por fim o melhor modelo é aquele que trás menor erro de predição, por isso o melhor modelo para os dados em estudo é o modelo linear com transformação logarítmica na variável resposta.

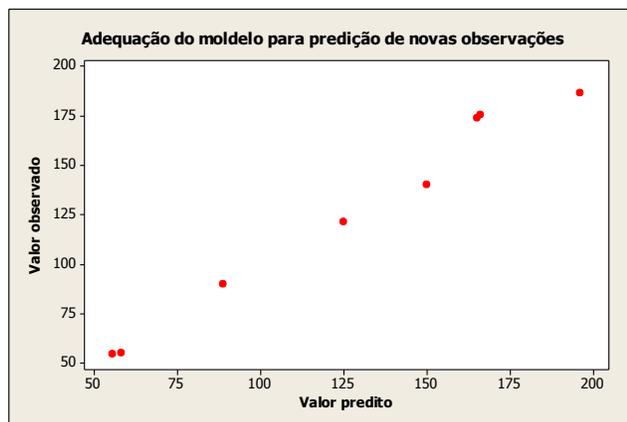


Figura 3.15 Gráfico de adequação do modelo com transformação logarítmica para novas predições.

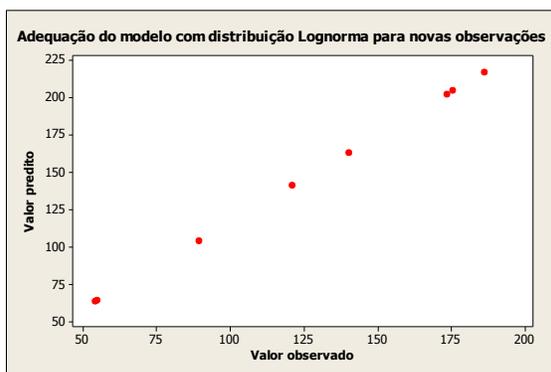


Figura 3.16; Gráfico de adequação do modelo com Distribuição log-normal para novas predições.

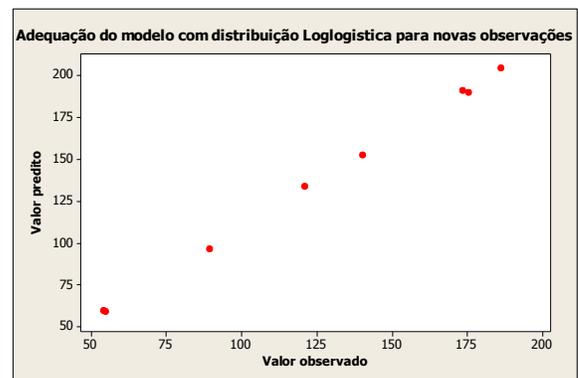


Figura 3.17; Gráfico de adequação do modelo com distribuição log-logística para novas predições.

CONSIDERAÇÕES FINAIS

4.1 CONCLUSÕES

A estratégia proposta foi seguida assim como, as investigações relativas às suposições do modelo como análise de resíduo e inferência sobre as equações ajustadas. O valor de um imóvel pode ser representado por uma equação de regressão linear múltipla. Reunindo um bom conjunto de dados, o avaliador tem condições de estudar modelos complexos dos fenômenos de mercado.

Comparando os modelos concluímos que o melhor modelo encontrado foi o modelo com transformação logarítmica na variável resposta, apresentou o valor de R^2 igual a 86,0%, o que mostra que o modelo consegue explicar 86,0% da variação dos preços dos terrenos urbanos estudados, e menor erro médio para previsão, 4%. A análise de resíduos atende as suposições teóricas, possui boa capacidade de predição de preços para outros terrenos da região estudada na cidade de Betim, considerando as mesmas condições dos terrenos analisados.

A amostra foi retirada de uma região bem localizada, próxima a pontos de referências como shopping, universidade, parque ecológico e centro comercial. Em quase sua totalidade possui infra-estrutura completa, o que talvez possa explicar o fato

da infra-estrutura não ter sido uma variável importante para o modelo. A variável via publica foi apontada pelo modelo como significativa, podemos explicar isso pensando que como se trata de bairros considerados bons pelo município, o que poderia diferenciar o preço dos terrenos seria o fato deles localizarem-se em vias de melhor acesso, vias que de ligação ou vias principais.

O banco de dados foi formado entre os meses de setembro e outubro de 2007, porém contando com dados desde maio de 2000. Por isso a amostra utilizada para testar a capacidade de predição do modelo pertence a essas datas. A extrapolação de dados, ou seja, dados que não pertence às características da amostra utilizada na modelagem da equação, pode provocar avaliações perturbadas ou distorcidas da realidade. Para sua aplicação nos dias de hoje faz-se necessário à atualização do banco de dados.

O uso do modelo para cálculo de impostos municipais é uma importante aplicação, uma vez que sabendo do preço mercadológico, os impostos podem ser mais justos e trazerem maior arrecadação ao município. Essa aplicação pode ser feita com esse modelo obedecendo ao que se disse acima e as especificações do município.

Tudo que se pretende fazer na vida merece um momento de reflexão e análise, uma avaliação é uma tomada de decisão muito séria, para isso mesmo tendo o conhecimento estatístico necessário ou superior, o avaliador precisa ser sempre um grande conhecedor do mercado imobiliário, conhecendo os fenômenos da economia urbana, e usando a estatística como ferramenta de trabalho.

..... REFERÊNCIAS BIBLIOGRÁFICAS

ABNT (Associação Brasileira de Normas técnicas) NRB 14653/2001, sob o título geral “*Avaliação de Bens*”, em suas partes 1(procedimentos gerais) e parte 2 (imóveis urbanos).

AYRES, Antônio. **Como Avaliar Imóveis**. São Paulo. Editora Imobiliária S/C Ltda, 1996.

BARBOSA FILHO, D.S. In: *Técnicas Avançadas de Engenharia de Avaliações*. Caixa Econômica Federal, 1988.

DANTAS, Rubens A. **Engenharia de Avaliações – Introdução à Metodologia Científica**. São Paulo: Pini, 1998.

GONZÁLEZ, M. A. S. **Metodologia de Avaliação de Imóveis**. Novo Hamburgo - Editora SGE,2003.

FILKER, José. **Avaliação de imóveis Urbanos**. 5º ed., São Paulo: Pini, 1997.

COLOSIMO, A. Enrico; GIOLO, R. Suely, **Análise de Sobrevivência Aplicada**, Ed.: Edgard Blucher,2006.

MATOS, Orlando C. de. **Econometria Básica**. São Paulo: Atlas S/A, 2000.

MONTGOMERY, Douglas C. ; RUNGER, C. George. **Estatística Aplicada e Probabilidade para Engenheiros**. 2º Ed., Rio e Janeiro, LTC, 2003.

MOREIRA FILHO, I. I., FRAINER, J. I., MOREIRA, R. M. I., MOREIRA, R. M.I. **Avaliação de Bens por Estatística Inferencial e Regressão Múltiplas**. Porto Alegre: Avalien, 1993.

TRIVELLONI, Carlos A. P.. *Metodologia para Avaliação em Massa de Apartamentos por Inferência Estatística e técnicas de Análise Multivariada*. Dissertação de Mestrado. Curso de Pós Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina. Florianópolis, 1998.

ZACAN, Evelise C. *Metodologia para Avaliação em Massa de imóveis para Efeito de Cobrança de Tributos Municipais - Caso de Apartamentos da Cidade de Criciúma, Santa Catarina*. Dissertação de Mestrado. Curso de Pós Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina. Florianópolis, 1995.

GAZOLA,,Sebastião. *Construção de um Modelo de regressão Para Avaliação de Imóveis*. Dissertação de Mestrado. Curso de Pós Graduação em Engenharia de Produção. Universidade Federal de Santa Catarina. Florianópolis, 2002.

PEREIRA, Júlio C. R. **Análise de Dados Qualitativos**. 3º ed. São Paulo: Edusp, 2001.

DRAPER, N. R. & SMITH, H. **Applied Regression Analysis**. New York: Jhon Wiley & Sons, Inc, 1981.

TRIOLA, F. Marcos, **Introdução a Estatística**. 9^o ed.LTC

MORETIN, A. Pedro, **Econometria Financeira, Um Curso em Séries Temporais Financeiras**. São Paulo, 2006

Anexo A: Exemplo de seleção das variáveis utilizando o passo-a-passo para trás.

Modelo de regressão linear múltipla com a variável dependente (valor/m²) com transformação logarítmica.

Passo 0

Preditores	Coef	D.P	Coef	T	P	P
Constant	4,8026	0,1240		38,74	0,000	
d1-L	-0,58925	0,05940		-9,92	0,000	
d2-L	-0,50475	0,07242		-6,97	0,000	
d3-L	0,12898	0,06604		1,95	0,055	0,00000
d4-L	0,59160	0,05308		11,15	0,000	
d5-L	0,18825	0,06773		2,78	0,007	
d6-L	-0,06677	0,06863		-0,97	0,334	
d7-VP	-0,11193	0,03056		-3,66	0,001	
d8-TP	-0,05248	0,07568		-0,69	0,491	
d9-TP	0,03102	0,07040		0,44	0,661	0,800754
d10-IF	0,03017	0,04941		0,61	0,544	
d11-Z	-0,00383	0,03412		-0,11	0,911	
d12-T	0,07455	0,02805		2,66	0,010	
d13-D	-0,07559	0,04243		-1,78	0,080	0,000557
d14-D	-0,07889	0,03630		-2,17	0,034	
Área	0,0004525	0,0001340		3,38	0,001	
Testada	-0,038648	0,009582		-4,03	0,000	

R² = 86,1% R-Sq(adj) = 82,5%

Passo 1

Preditores	Coef	D.P	Coef	T	P	P
Constant	4,8051		0,1209	39,75	0,000	
d1-L	-0,58891		0,05886	-10,01	0,000	
d2-L	-0,50409		0,07161	-7,04	0,000	
d3-L	0,12909		0,06551	1,97	0,053	
d4-L	0,59076		0,05213	11,33	0,000	0,00000
d5-L	0,18629		0,06494	2,87	0,006	
d6-L	-0,06619		0,06789	-0,97	0,333	
d7-VP	-0,11343		0,02726	-4,16	0,000	
d8-TP	-0,05259		0,07508	-0,70	0,486	
d9-TP	0,03170		0,06959	0,46	0,650	0,79757
d10-IF	0,03015		0,04902	0,62	0,541	
d12-T	0,07472		0,02778	2,69	0,009	
d13-D	-0,07508		0,04185	-1,79	0,078	
d14-D	-0,07894		0,03601	-2,19	0,032	0,000494
Área	0,0004491		0,0001295	3,47	0,001	
Testada	-0,038572		0,009483	-4,07	0,000	

R² = 86,1% R-Sq(adj) = 82,8%

Passo 2

Preditores	Coef	D.PCoef	T	P	P
Constant	4,8176	0,1128	42,72	0,000	
d1-L	-0,58187	0,05733	-10,15	0,000	
d2-L	-0,50118	0,07010	-7,15	0,000	
d3-L	0,11771	0,06244	1,89	0,064	0,0000
d4-L	0,59429	0,05055	11,76	0,000	
d5-L	0,18459	0,06398	2,89	0,005	
d6-L	-0,07645	0,06320	-1,21	0,231	
d7-VP	-0,11170	0,02679	-4,17	0,000	
d10-IF	0,03060	0,04804	0,64	0,526	
d12-T	0,07264	0,02720	2,67	0,010	
d13-D	-0,07147	0,04035	-1,77	0,081	0,000284
d14-D	-0,07988	0,03542	-2,26	0,027	
Área	0,0004417	0,0001275	3,46	0,001	
Testada	-0,037964	0,009332	-4,07	0,000	

R² = 86,1% R-Sq(adj) = 83,2%

	Valor	Localidade
1	100,00	Angola
2	125,00	Angola
3	282,29	Angola
4	145,38	Angola
5	152,78	Angola
6	95,24	Angola
7	222,22	Angola
8	138,38	Angola
9	150,00	Angola

Passo 3

Preditores	Coef	D.P	Coef	T	P	P
Constant	4,8472	0,1022		47,41	0,000	
d1-L	-0,60190	0,04772		-12,61	0,000	
d2-L	-0,49632	0,06937		-7,16	0,000	
d3-L	0,12086	0,06197		1,95	0,055	0,0000
d4-L	0,59532	0,05029		11,84	0,000	
d5-L	0,18872	0,06337		2,98	0,004	
d6-L	-0,07136	0,06241		-1,14	0,257	
d7-VP	-0,11107	0,02666		-4,17	0,000	
d12-T	0,07171	0,02703		2,65	0,010	
d13-D	-0,07302	0,04009		-1,82	0,073	
d14-D	-0,07732	0,03503		-2,21	0,031	0,000316
Área	0,0004540	0,0001255		3,62	0,001	
Testada	-0,038441	0,009259		-4,15	0,000	

R² = 86,0% R-Sq(adj) = 83,4%