

MARCELO SUAREZ SALDANHA
ANDRÉ MONTENEGRO DUARTE

APLICAÇÃO DE VARIÁVEIS CATEGÓRICAS COM REGRESSÃO LOGÍSTICA COMO ALTERNATIVA NA AVALIAÇÃO DE IMÓVEIS

Trabalho de Avaliação

Goiânia/GO
2021

TRABALHO DO XXI COBREAP – GOIÂNIA-GO

09 A 12 DE NOVEMBRO DE 2021

**APLICAÇÃO DE VARIÁVEIS CATEGÓRICAS COM REGRESSÃO LOGÍSTICA
COMO ALTERNATIVA NA AVALIAÇÃO DE IMÓVEIS.**

Palavras Chaves: Avaliação Imobiliária, Análise de Mercado, Regressão Linear, Regressão Logística e Variáveis Categóricas

RESUMO

NA AVALIAÇÃO DE IMÓVEIS, VÁRIAS TÉCNICAS PODEM SER UTILIZADAS QUANDO ADOTADO O MÉTODO COMPARATIVA DE DADOS DE MERCADO. ESTE ESTUDO PROPÕE, DE FORMA INTRODUTÓRIA, A UTILIZAÇÃO DE UMA METODOLOGIA ALTERNATIVA PARA ANÁLISE E TRATAMENTO DE DADOS CONSIDERANDO A REGRESSÃO LOGÍSTICA OU SEMI-LOGÍSTICA, A PARTIR DA APLICAÇÃO DE VARIÁVEIS CATEGÓRICAS COMO FERRAMENTA EFICAZ DE PROBABILIDADE DE ANÁLISE DO VALOR NO PROCESSO DE AVALIAÇÃO IMOBILIÁRIA. EM ESTUDO DE CASO DE APARTAMENTOS RESIDÊNCIAS DE ALTO PADRÃO, O MODELO DESENVOLVIDO POR ESTA METODOLOGIA SE MOSTROU EFICAZ E CONSISTENTE, INDICANDO QUE ESTA ABORDAGEM METODOLÓGICA DE TRATAR OS DADOS DE MERCADO PODE SER MUITO ÚTIL, INCLUSIVE POR ANULAR A CONDIÇÃO DE SOBREAJUSTE EM MODELOS DE REGRESSÃO MÚLTIPLA NO DIAGNÓSTICO DO MERCADO IMOBILIÁRIO.

1. INTRODUÇÃO:

Inicialmente, quando se utiliza a inferência estatística na engenharia de avaliações, se faz uma análise de regressão para modelar o mercado imobiliário, por meio de amostragem (pesquisa de dados) e tratamento de variáveis, que podem ser quantitativas, contínuas e discretas, cujas unidades ou quantidades são originadas de medição ou contagem ou qualitativas, que, de alguma forma, devem ser expressas em seus atributos de maneira quantitativa, seja por variáveis-chave que, a priori e tradicionalmente, são importantes para a formação do valor do imóvel, por variável dicotômica ou código alocado, analogia (proxy), variáveis independentes.

As técnicas de inferência estatística usam a hipótese de uma amostragem ser probabilística; no entanto, nos casos em que a população não é finita ou não é totalmente acessível, utiliza-se a amostragem não aleatória, e nesses casos o bom senso deverá indicar a possibilidade de utilizar ou não as técnicas de indução para esse tipo de amostragem.

A proposição deste estudo é a utilização de uma técnica alternativa para aplicação da regressão logística, que tem como objetivo produzir modelos matemáticos a partir de uma variante do modelo tradicional de regressão logística, substituindo analisar a variável resposta como uma variável binária por uma variável contínua, onde a variável resposta corresponderia, por exemplo, se o cliente comprou ou não comprou um imóvel de alto padrão, ou se o imóvel é alugado ou próprio. Na prática, este artigo de uma maneira mais simples, pretende buscar uma alternativa inversa de análise logística ao codificar devidamente as variáveis preditoras como categóricas para explicar a variável resposta como variável contínua.

Para uma boa análise de regressão, o avaliador possui a sua disposição o uso de grandes quantidades de dados confiáveis e um pequeno número de preditores que possuem relações causais bem estabelecidas, valendo-se de um raciocínio lógico para incluir variáveis no modelo de regressão garantindo de uma maneira clara a interpretação adequada das informações amostrais, evitando assim à condição reduzida da capacidade de fazer previsões, já com o questionamento de uma ideia de quais são as variáveis importantes junto com suas relações, sinais de coeficiente e magnitudes de efeito com base em pesquisas para formação do modelo de regressão.

Diante deste fato, a proposição deste estudo é a utilização alternativa de aplicação da regressão logística, técnica estatística que tem como objetivo produzir modelos matemáticos a partir de uma série de variáveis independentes, tipo contínuas e/ou discretas.

É evidente que o modelo deva ser firmemente fundamentado na realidade (JEFFERIEIS, 2009), realidade esta que, no caso da avaliação de um bem, é o “mercado imobiliário”, o qual deve ser observado e consequentemente modelado com base no método científico, nas ciências e na boa técnica. Esse é, fundamentalmente, a finalidade da engenharia de avaliações quando o objetivo da avaliação for a identificação do valor de mercado.

Mas, como brilhantemente colocado por BLACKARAD (1989), avaliação não é uma ciência, mas sim uma arte e a precisão não é garantida, pois os valores imobiliários não são cristais nos quais cortamos arestas vivas.

Então, coletados e sistematizados os dados, propõem-se estudar os atributos dos imóveis como se fossem variáveis categóricas, assumindo um mínimo de dois atributos; por exemplo: podem assumir os valores “0” ou “1”, aplicados num conjunto de observações que permita uma predição de valores tomados por uma variável categórica, frequentemente binária.

A regressão logística (CABRAL, 2013) é a técnica estatística que tem como objetivo modelar, a partir de um conjunto de observações, a relação “logística” entre uma variável resposta dicotômica e uma série de variáveis explicativas numéricas (contínuas, discretas) e/ou categóricas.

Os resultados práticos deste estudo serão analisados como uma ferramenta de análise estatística do valor imobiliário em processos de avaliação de imóveis, através de metodologia científica e considerando as normas brasileiras e internacionais.

2. REGRESSÃO LOGÍSTICA:

A regressão logística é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, frequentemente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias. (PORTAL ACTION – Estatcamp, 2020).

A regressão logística é amplamente usada em ciências médicas e sociais, e tem outras denominações, como modelo logístico, modelo logit, e classificador de máxima entropia. A regressão logística é utilizada em áreas como as seguintes: Em medicina, permite, por exemplo, determinar os fatores que caracterizam um grupo de indivíduos doentes em relação a indivíduos sãos.

No domínio dos seguros, permite encontrar frações da clientela que sejam sensíveis a determinada política securitária em relação a um dado risco particular. Em instituições financeiras, pode detectar os grupos de risco para a subscrição de um crédito.

Em econometria, a regressão logística é amplamente utilizada, e permite explicar muito bem as variáveis discretas, cujo êxito está alicerçado sobretudo nas numerosas ferramentas de testes que permitem interpretar de modo aprofundado os resultados obtidos.

Em comparação com as técnicas conhecidas em regressão, em especial a regressão linear, a regressão logística distingue-se essencialmente pelo fato de a variável resposta ser categórica.

Enquanto método de predição para variáveis categóricas, a regressão logística é comparável às técnicas supervisionadas propostas em aprendizagem automática (árvores de decisão, redes neurais, etc.), ou ainda a análise discriminante preditiva em estatística exploratória. É possível coloca-las em concorrência para a escolha do modelo mais adaptado para um certo problema preditivo a resolver.

Trata-se de um modelo de regressão para variáveis dependentes ou de resposta binomialmente distribuídas. É útil para modelar a probabilidade de um evento ocorrer como função de outros fatores. É um modelo linear generalizado, que usa como função de ligação a função logarítmica.

Então, resumidamente, temos a probabilidade de rejeitar uma hipótese verdadeira ou aceitar uma falsa, de uma modelagem econométrica são verificadas através das seguintes análises de variância:

- a) (r) grau de correlação entre as variáveis independentes;
- b) o valor-p, significância dos regressores;
- c) a ANOVA, significância do modelo de regressão;
- d) problemas de econometria (multicolinearidade, heteroscedasticidade, autocorrelação, endogeneidade).

2.1. Modelo de Regressão Logística Simples

O modelo de regressão logística simples é usado para o caso de regressão com uma variável explicativa. Suponha-se uma amostra de n observações independentes da terna (x_i, m_i, y_i) , $i=1,2,\dots,n$, sendo que:

- ✓ x_i é o valor da variável explicativa;
- ✓ m_i é a quantidade de itens verificados na amostra (número de ensaios);
- ✓ y_i número de ocorrência de um evento (exemplo: número de dormitórios) em m_i ensaios; e n é o tamanho da amostra.

Com isso, assumimos que a variável resposta tem distribuição de probabilidade binomial, tal que:

$$P[Y_i = y_i] = \binom{m_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{m_i - y_i}.$$

Para adequarmos a resposta média ao modelo linear usamos a função de ligação

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}, \quad i = 1, \dots, n,$$

que pode ser escrita como:

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_i.$$

A interpretação dos parâmetros de um modelo de regressão logística é obtida comparando a probabilidade de sucesso com a probabilidade de fracasso, usando a função odds ratio - OR (razão de chances). Essa função é obtida a partir da função odds.

$$g(x) = \frac{\pi(x)}{[1 - \pi(x)]} = \frac{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}}{1 - \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}} = \frac{\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}}{\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}}} = e^{\beta_0 + \beta_1 x_i}.$$

Assim, ao tomarmos dois valores distintos da variável explicativa, x_j e x_{j+1} , obtemos

$$OR = \frac{g(x_{j+1})}{g(x_j)} = \frac{e^{\beta_0 + \beta_1 x_{j+1}}}{e^{\beta_0 + \beta_1 x_j}}.$$

Temos ainda que:

$$\begin{aligned}\ln(OR) &= \ln\left[\frac{g(x_{j+1})}{g(x_j)}\right] = \ln[g(x_{j+1})] - \ln[g(x_j)] \\ &= \beta_0 + \beta_1 x_{j+1} - \beta_0 - \beta_1 x_j = \beta_1(x_{j+1} - x_j).\end{aligned}$$

Fazendo $x_{j+1} - x_j = 1$ unidade, então:

$$\ln(OR) = \ln(e^{\beta_1}) = \beta_1.$$

Assim, temos o quão provável o resultado ocorrerá entre os atributos x_{j+1} em relação aos atributos x_j , fazendo, portanto, nossas análises.

Na inferência de uma modelagem logística, após estimar os coeficientes, temos interesse em assegurar a significância das variáveis no modelo. Isto geralmente envolve formulação e teste de uma hipótese estatística para determinar se a variável independente no modelo é significativamente relacionada com a variável resposta.

Na regressão logística a ideia está no interesse do valor da SQR. Um valor alto da SQR sugere que a variável independente é importante, caso contrário, a variável independente não é útil na predição da variável resposta.

Ao comparar os valores observados da variável resposta com os valores preditos obtidos dos modelos de regressão com e sem a variável em questão. A comparação dos observados com os valores preditos é baseado no log da verossimilhança. Para entender melhor essa comparação, é útil pensar em um valor observado da variável resposta também como sendo um valor predito resultante de um modelo de regressão, isto é, aquele que contém tantos parâmetros quanto observações se quer analisar.

Para isso, temos a análise de variância a ANOVA, significância do modelo de regressão, F de Snedecor, os testes de hipóteses, o valor-p, significância dos regressores.

Quanto aos parâmetros base para construção das estimativas do intervalo de confiança é a mesma teoria estatística que usamos para os testes de significância do modelo. Em particular, um intervalo de confiança de $100(1-g)\%$ para o parâmetro β_1 , para a inclinação e intercepto são baseados em seus respectivos referidos testes.

2.2. Modelo de Regressão Logística Múltipla.

O modelo de regressão logística múltipla, assim como no modelo de regressão linear simples, podemos ajustar um modelo para a variável resposta levando em conta mais de uma variável explicativa (covariável) é usado para o caso de regressão com mais de uma variável explicativa.

Considere um conjunto de p variáveis independentes denotadas como um vetor $X=(X_1, X_2, X_3, \dots, X_p)$, neste caso a função de ligação é:

$$g(X) = \ln\left(\frac{\pi(X)}{1 - \pi(X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

$$E[Y] = \pi(X) = \frac{e^{g(X)}}{1 + e^{g(X)}}.$$

Assim como na Inferência em um modelo de regressão logística simples, para avaliar a significância dos coeficientes as seguintes hipóteses, onde podemos testar

a significância dos parâmetros pelo teste de hipótese, o valor-p, significância dos regressores.

Vamos considerar a seguinte hipótese:

$$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases} \quad j = 0, 1, 2, 3, 4, 5$$

Para testar a hipótese acima, temos a seguinte análise, se não rejeitarmos H_0 temos que a variável X_j não explica a variável resposta.

Desta forma, vamos testar a significância dos parâmetros pelo teste de hipótese, o valor-p, cujo modelo de probabilidade proposto é:

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2}},$$

Os intervalos de confiança para os parâmetros modelo de regressão logística múltipla, valores ajustados e odds ratio, são os mesmos os mesmos no caso da regressão simples.

2.2.1. Variáveis Independentes Categóricas:

As variáveis independentes categóricas, variáveis explicativas assumindo valores discretos, ou seja, a variável independente é codificada em 0 ou 1.

Por exemplo, considere uma única variável explicativa X com níveis A ou B. Para estimar os parâmetros do modelo, estas variáveis são substituídas por valores numéricos, por exemplo 0 para o nível A e 1 para o nível B. Esses valores são chamados de codificação das categorias da variável qualitativa. Assim, a mesma teoria desenvolvida para o modelo logístico considerando variáveis explicativas contínuas pode ser aplicado para esses casos, através dos valores do modelo de regressão logística quando a variável independente é dicotômica, será interpretado por:

$$OR = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} = \frac{\frac{\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}}{1}}{\frac{\frac{e^{\beta_0}}{1 + e^{\beta_0}}}{1}} = \exp(\beta_1)$$

	x=1	x=0
y=1	$\pi(1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$	$\pi(0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$
y=0	$1 - \pi(1) = \frac{1}{1 + \exp(\beta_0 + \beta_1)}$	$1 - \pi(0) = \frac{1}{1 + \exp(\beta_0)}$
Total	1	1

Portanto podemos fazer algumas análises, antes de fazer conclusões a respeito da odds ratio (OR) encontrada, é importante considerar o efeito que a codificação da variável tem no valor da estimativa da odds. Vimos que $OR^{\wedge} = \exp(\beta^{\wedge}1)$, esse resultado é válido apenas quando a variável independente é codificada em 0 ou 1.

2.2.2. Seleção de Variáveis Independentes:

A seleção de variáveis é feita pelo método Stepwise, muito usado em regressão linear, ou seja, qualquer procedimento para seleção ou exclusão de variáveis de um modelo é baseado em um algoritmo que checa a importância das variáveis, incluindo ou excluindo-as do modelo se baseando em uma regra de decisão.

A importância da variável é definida em termos de uma medida de significância estatística do coeficiente associado à variável para o modelo. Essa estatística depende das suposições do modelo de análise de regressão.

No método Stepwise da regressão linear um teste F é usado desde que os erros tenham distribuição normal. Na regressão logística os erros seguem distribuição binomial e a significância é assegurada via teste de hipótese e da razão de verossimilhança. Assim, em cada passo do procedimento a variável mais importante, em termos estatísticos, é aquela que produz a maior mudança no logaritmo da verossimilhança em relação ao modelo que não contém a variável.

Temos a seguir o passo a passo do algoritmo do método de Stepwise:

- 1º Passo - ajustamos o modelo contendo as variáveis [$x_1, x_2, \dots, x_n$], para verificar se o valor-p das variáveis são importantes para o modelo, ou seja, o valor-p do teste associado à variável x_2 é o menor dentre todos os outros. Se esse valor-p for menor que α de x_2 entra no modelo e prosseguimos para o passo 2, caso contrário o algoritmo termina e apenas a variável x_1 foi incluída no modelo.
- 2º Passo - começa com o ajuste do modelo contendo as variáveis x_1 e x_2 . É possível que, com a inclusão de x_2 , x_1 passa a ser não significativa para o modelo. Por isso, nesse passo testamos a significância de uma das variáveis dado que a outra está no modelo. Para decidir se removemos ou não alguma das variáveis, selecionamos o maior valor-p. Vamos supor que o valor-p da variável x_2 seja o maior, a variável x_2 sai do modelo, comparamos o seu p-valor com um nível de significância de saída, ou seja, no método Stepwise em modelos de regressão logística, é recomendado $\alpha = 0,15$ e $\alpha = 0,2$. Após a verificação da eliminação da variável x_2 , ajustamos as outras variáveis, $x_3, x_4, \dots, x_n$. Seja x_3 a variável associada ao menor valor-p, x_3 entra no modelo e prosseguimos para o passo 3, caso contrário paramos por aqui e o modelo é explicado por x_1 e x_2 .
- 3º Passo - o procedimento é o mesmo do 2º passo, ajusta o modelo incluindo a variável selecionada durante os passos anteriores, checando se a inclusão

dessa variável fez com que as outras variáveis do modelo perdessem a significância. O processo continua dessa mesma maneira até o último passo, o passo seguinte (s).

- Passo Seguinte (s)- esse passo ocorre quando: todas as “ p ” variáveis entraram no modelo ou se todas variáveis no modelo tem valor-p para sair menor que α e as variáveis não incluídas no modelo tem valor-p para entrar maior que α .

2.2.3. Medidas de Qualidade do Ajuste do Modelo:

Para avaliar a qualidade do ajuste do modelo fazendo um gráfico do modelo ajustado, mas existem medidas, como a estatística de Pearson e a análise de variância ANOVA, os testes de significância do modelo de regressão, F de Snedecor, e de hipóteses, o valor-p, significância dos regressores, que ajudam a verificar a qualidade do modelo de regressão.

Então, precisamos identificar, dentre as observações, aquelas com os mesmos valores entre si para todas as covariáveis. Assim, seja $X=(X_1, X_2, \dots, X_p)$ o conjunto de variáveis explicativas do modelo. Seja também J o número distinto de valores que x assume na amostra. O número de indivíduos na amostra com valores iguais $x=x_j$ é denotado por: m_j . Já y_j é o número de indivíduos positivos, isto é, $y=1$, em m_j . Consequentemente, $\sum y_j = n$ que representa o total de indivíduos com $y=1$.

2.2.4. Diagnóstico de Influência dos Dados:

É importante diagnosticar a influências dos dados observados, isto é, detectar OUTLIERS, pontos influentes, ou seja, análise dos resíduos, pontos que afetam de forma significativa o ajuste do modelo de regressão.

Assim como no modelo de regressão linear, podemos utilizar a distância de Cook para avaliar a influência geral da observação i nas estimativas dos coeficientes da regressão. A métrica da distância de Cook para a observação i é:

$$D_i = \frac{r_{SPi}^2 h_i}{(p + 1)(1 - h_i)}.$$

Se, D_i é grande quando o resíduo e_i e/ou outliers é grande, desta forma, todas as observações de distâncias maiores que 1 são consideradas influentes.

3. ESTUDO DE CASO:

No estudo de caso prático a seguir, como exemplo de aplicação da metodologia alternativa para aplicação da regressão logística¹, vamos analisar os preços praticados no seguimento de compra e venda de apartamentos de alto padrão e de grande porte com áreas de 250m² a 800m², comercializados em zonas de incorporação do mercado imobiliário.

Para utilização dos conceitos econométricos, é considerado como o método comparativo por excelência. Com aplicação em qualquer estudo mercadológico de bases imobiliárias de alienação de imóveis, urbana ou até mesmo de imóveis em áreas não urbanizadas na cidade.

3.1. Análise dos Preços dos Apartamentos:

Buscando entender como os preços dos apartamentos se relacionam com as suas características construtivas, é utilizada uma amostra de dados da tabela de venda de lançamentos imobiliários, que possui informação sobre os preços médios dos apartamentos de 3 dormitórios, tipo uma planta e de cobertura duas plantas. As características adicionais utilizadas no modelo de estimação foram obtidas por meio de pesquisa e análise individual das fichas técnicas das tipologias de apartamentos selecionados. Os dados estão disponíveis em lançamentos.xls (planilha Excel).

¹ Em verdade, se procedeu uma regressão semi-logística, pois a variável resposta, a explicada (y) não foi trabalhada como categórica, mas sim como variável quantitativa contínua na sua unidade real (\$/m²), sendo, porém, todas as variáveis explicativas (X_n) categóricas, o que, no stricto sensu, não se caracterizaria como um modelo logístico. Mas, como um trabalho introdutório, os princípios da regressão logística estão todos presentes.

Na avaliação qualitativa dos bens imóveis, as variáveis categóricas assumem dois valores distintos, geralmente (0) zero e (1) um, e são usadas para representar variáveis qualitativas, mudanças súbitas no processo em análise, extratos de uma variável qualquer, etc, sendo muito úteis como atributos de homogeneização fundamentada, na mensuração das diferenças existentes entre os dados amostrais.

Variáveis qualitativas com duas categorias indicam se uma observação de interesse pertence a uma dada categoria ou se possui alguma dada característica. São exemplos de variáveis qualitativas:

- Características dos apartamentos: área, padrão de construção, possui garagem número de dormitórios.
- Condição dos apartamentos: locação compra e venda, próximo de uma via importante, estado de conservação.
- Condição de um local, ou região: zona de incorporação imobiliária, faz parte ou não de um aglomerado urbano.
- Condição temporal: data do negócio, mês de locação, ano, ou qualquer outra frequência temporal.

Transformar essas variáveis qualitativas em binárias é umas das formas usadas para analisá-las. Assim, para utilizar uma variável binária e tecer interpretações adequadas é preciso definir claramente o que será medido por 1 ou 0. Veja o exemplo da informação de ter vaga de garagem:

Vaga de Garagem = 1, se tem vaga de garagem e 0 se não possui vaga.

Utiliza-se 1 e 0 para ter interpretações naturais do modelo. O objetivo é demonstrar a facilidade que o uso de variáveis binárias gera.

Supondo que se está interessado em analisar se existe gap (hiato ou diferença) padrão construtivo entre ter ou não ter o salão de festas por meio do seguinte modelo

$$\text{preço} = B_0 + B_1\text{padrão} + B_2\text{vaga} + u$$

Supondo $E[u/\text{padrão}, \text{vaga}] = 0$, o modelo indicará existência de diferencial preço se $E[\text{preço}/\text{padrão}/\text{vaga} = 0] = B_0 + B_1\text{padrão}$ é diferente de $E[\text{preço}/\text{padrão}/\text{vaga} = 1] = (B_0 + B_2) + B_1\text{padrão}$

Ou seja, esse diferencial ocorrerá se B_2 for diferente de 0. Pode-se dizer que existem evidências empíricas sobre a existência desta desigualdade, mas seria preciso testar essa hipótese para a comprovação (lembrando que foi considerada válida a seguinte hipótese: $E[\text{preço}/\text{padrão}/\text{vaga} = 0]$

3.2. Análise Qualitativa:

O objetivo do estudo é aplicar a metodologia alternativa da regressão logística para análise dos preços dos praticados no mercado imobiliário e, para tal, se faz necessário transformar todas as variáveis quantitativas em variáveis qualitativas em variáveis categóricas, e proceder nossa análise de regressão logística.

Então, procedemos a nossa análise qualitativa do valor unitário praticados nos lançamentos de apartamentos residenciais em zona de incorporação imobiliária, utilizamos as seguintes variáveis categóricas na construção do modelo de análise de regressão:

Variável Resposta (contínua):

- ✓ **Unitário (PU):** representa o preço unitário dos apartamentos em dólares por m² (\$/m²);

Variáveis Preditora (categórica):

- ✓ **Área Global (AG):** representa a área global (área privativa mais a área comum) dos apartamentos, incluindo os boxes de estacionamento em m², assumindo valores 1 (sim) para metragens superiores a 500m² e 0 (não) para áreas menores; (definidos os limites de porte dos apartamentos da amostra que se quer analisar)

- ✓ **Tipo de Unidade (TU):** representa se o apartamento analisado é de cobertura 3D, assumindo os valores 1 (sim) e 0 (não);
- ✓ **Bairro Local (BL):** representa o bairro onde encontra-se localizado o apartamento, assumindo os valores 1 (sim) e 0 (não);
- ✓ **Vaga de garagem (VG):** representa se o apartamento possui poucas ou muitas vagas de garagem, assumindo os valores 1 (sim) e 0 (não);
- ✓ **Padrão Construtivo (PC):** representa o padrão de construção H8-3A definido pela NB 140/91, assumindo os valores 1 (sim) e 0 (não);
- ✓ **Estágio da Construção (EC):** representa se o estágio de construção do edifício se está pronto ou em fase de construção, assumindo os valores 1 (sim) e 0 (não);

Tabela VIII - Tabulação das informações categóricas dos dados de mercado pesquisados.

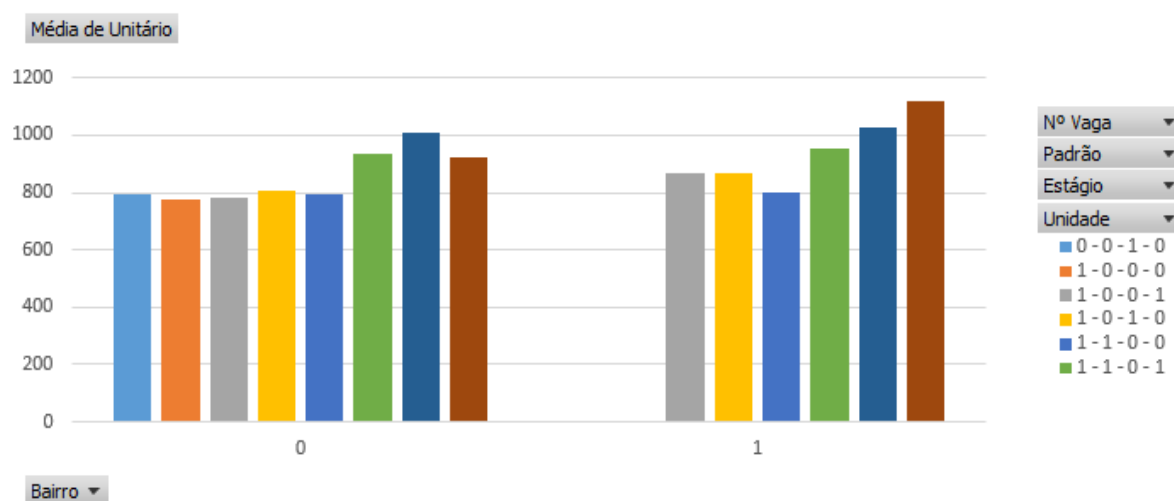
Unitário	Área	Unidade	Bairro	Vaga	Padrão	Estágio
941,17	1	0	1	1	1	1
797,27	0	0	1	1	1	0
841,35	1	1	1	1	1	0
946,05	1	0	1	1	1	1
946,05	1	1	1	1	1	0
948,76	1	0	1	1	0	1
914,20	0	1	1	1	0	0
1.209,90	0	0	1	1	1	1
1.119,85	0	1	1	1	1	1
791,90	1	0	1	1	0	1
816,60	0	1	1	1	0	0
1.023,25	1	0	1	1	1	1
978,30	0	1	1	1	1	0
869,91	1	1	1	1	0	0
1.013,96	0	1	1	1	1	0
1.000,00	1	1	1	1	1	0
857,27	1	0	0	1	0	0
857,60	0	1	0	1	0	0
907,70	0	0	0	1	0	1

Unitário	Área	Unidade	Bairro	Vaga	Padrão	Estágio
1.033,33	0	1	0	1	1	0
1.127,06	0	1	0	1	1	0
954,36	0	0	0	1	1	1
1.071,35	0	1	0	1	1	0
777,50	0	0	0	1	0	1
1.132,11	0	0	0	1	1	1
785,99	0	0	0	1	0	1
751,33	0	0	0	1	0	1
814,78	0	1	0	1	0	0
983,80	0	0	0	1	1	1
883,15	0	1	0	1	1	0
798,95	0	0	0	1	0	1
827,77	0	0	0	1	0	1
1.031,25	0	0	0	1	1	1
780,00	0	1	0	1	1	0
697,08	1	0	0	1	0	0
686,21	0	1	0	1	0	0
778,31	0	1	0	1	0	0
926,08	0	1	0	1	1	0
1.041,05	1	0	0	1	1	1
872,06	0	0	0	1	1	1
946,07	0	1	0	1	1	0
887,45	1	1	0	1	1	0
795,39	0	0	0	1	1	0
964,98	0	1	0	1	1	0
810,76	1	0	0	1	0	1
734,48	0	1	0	1	0	0
746,78	0	0	0	0	0	1
1.028,65	0	0	0	1	1	1
826,26	0	0	0	1	0	1
749,64	0	0	0	1	0	1
925,53	0	1	0	1	1	1
837,00	0	0	0	0	0	1
846,49	0	1	0	1	1	0
833,73	1	1	0	1	1	0
837,66	1	1	0	1	0	0

Analogamente, como no modelo da regressão clássica, a nossa análise alternativa de regressão logística, também tem como ponto de partida a tabela dinâmica, para após proceder a análise de correlação entre as variáveis e obter o modelo de regressão linear múltipla.

Tabela Dinâmica:

Partimos da análise comparativa da coluna preço “Unitário” com as colunas contendo as variáveis categóricas: “Área Global, Tipo de Unidade, Bairro Local, Número de Vaga, Padrão Construtivo e Estágio de Construção”, obtendo a média dos preços unitários frente as tipologias e características físicas dos imóveis analisadas, otimizando assim suas informações dos valores unitários estão dispostas no gráfico a seguir:



Valores Médios do Unitário (\$/m²)

Rótulo	Unitário \$/m²		Total Geral
0	\$ 791,89	\$ 880,09	\$ 875,56
1		\$ 947,41	\$ 947,41
Total	\$ 791,89	\$ 900,41	\$ 896,46

Análise de Correlação:

Selecionar as variáveis a serem estudadas como formadoras do preço unitário (\$/m²) e proceder na análise das correlações existentes entre elas, para verificar a dependência linear observada frente as restrições de utilização do modelo de regressão logística.

CORRELAÇÃO

	Área Glo- bal	Uni- dade	Bairro	Nº Vaga	Pa- drão	Está- gio	Unitá- rio
Área Global	100%						
Unidade	-7%	100%					
Bairro	38%	9%	100%				
Nº Vaga	12%	19%	12%	100%			
Padrão	0%	20%	16%	22%	100%		
Estágio	-5%	-78%	-5%	-21%	-12%	100%	
Unitário	-2%	7%	28%	17%	65%	14%	100%

Interpretação – foi verificada baixa correlação entre as variáveis explicativas utilizadas, exceto uma correlação forte entre área global com vaga de garagem e moderada entre cobertura com obra em construção. Pode-se concluir, conjuntamente com o teste t (valor-P), pelo indicativo da inexistência de qualquer relação ou tendência linear entre as variáveis independentes e dependentes. Logo, ficou caracterizado que os prepostos da regressão foram todos atendidos, ou seja, normalidade, ausência de multicolinearidade, independência de erros e linearidade.

Em relação ao modelo de regressão clássica, verifica-se no modelo alternativo de regressão logística que das variáveis explicativas analisadas, confirma-se que somente as variáveis do tipo de Unidade (TU) e o Estágio (EC) de construção não atendem os prepostos da regressão de normalidade, ausência de multicolinearidade, independência de erros e linearidade, devido a dependência de correlação forte entre as variáveis explicativas utilizadas.

Modelo de Alternativo de Regressão Logística

$$(PU) = B_0 + B_1(AG) + B_2(TU) + B_3(BL) + B_4(VG) + B_5(PC) + B_6(EC)$$

Parâmetros Estatísticos:

<i>Estatística de regressão</i>	
R múltiplo	73,66%
R-Quadrado	54,26%
R-quadrado ajustado	48,54%
Erro padrão	85,14
Observações	55

Interpretação – o poder de explicação do modelo de regressão logística é aferido pelo seu coeficiente de determinação é de 54,26%. Devido ao fato de que este coeficiente sempre cresce com o aumento do número de variáveis independentes e não leva em conta o número de graus de liberdade perdidos a cada parâmetro estimado, é recomendável considerar também o coeficiente de determinação ajustado de 54,26%.

ANOVA

	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>
Regressão	6	412705,13	68784,19	9,49	0,000072%
Resíduo	48	347968,64	7249,35		
Total	54	760673,78			

Interpretação - como $F_{cal} > F_{tab}$, rejeita-se a hipótese nula de regressão, logo, existe diferença significativa (F de significação $> 5\%$), pois confirma-se que os erros são independentes e provenientes de uma distribuição normal com média igual a zero e variância constantes, ou seja, aceita-se com 95,0% de confiabilidade mínima que o modelo de regressão seja representativo dos dados na formação do valor de mercado.

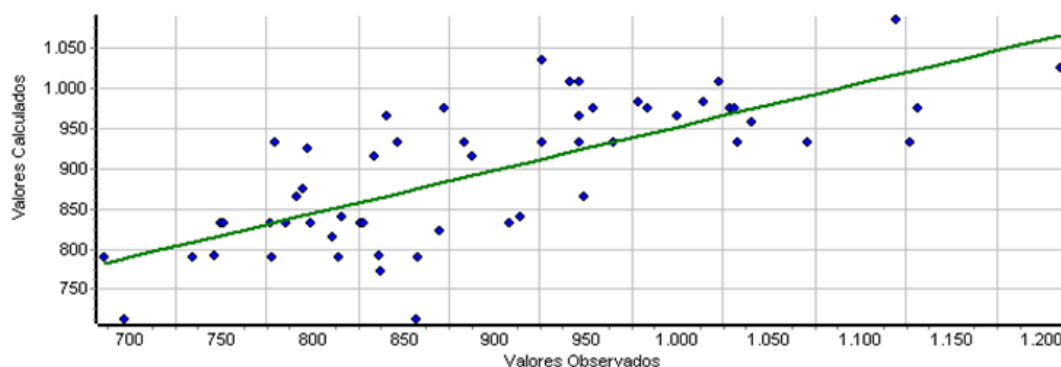
TESTE DE HIPÓTESES

	<i>Coeficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>
Interseção	689,90	71,19	9,69	0,00%
Área Global (AG)	-17,81	28,22	-0,63	53,10%
Tipo Unidade (TU)	59,26	38,56	1,54	13,09%
Bairro Local (BL)	50,53	28,10	1,80	7,85%
Vaga Garagem (VG)	39,98	64,55	0,62	53,86%
Padrão (PC)	144,08	24,39	5,91	0,00%
Estágio Constr. (EC)	101,99	37,99	2,68	0,99%

Interpretação – Os regressores área global T(AG) e vaga de garagem T(VG), não passaram no teste de significância, pois o valor-P > 5%, portanto não são importantes na formação do preço unitário. Os demais regressores das variáveis T(TU), T(BV), T(EC) e T(PC) > t_c são significativos na formação do valor de mercado, uma vez que rejeita-se a hipótese nula de regressão, logo o valor da probabilidade de se obter o efeito desejado é verdadeira com aceite de 95,0% de confiabilidade mínima.

Gráfico de Aderência:

O gráfico de resíduos versus ajustes representa graficamente os resíduos no eixo Y e os valores ajustados no eixo X, é utilizado para verificar a pressuposição de que os resíduos são aleatoriamente distribuídos e têm variância constante.



Histograma de Resíduos Padronizados

O histograma dos resíduos mostra a distribuição dos resíduos padronizados para todas as observações. O histograma dos resíduos determina se os dados são assimétricos ou se incluem outliers. Os padrões de distribuição de frequência indicam que o modelo atende às suposições da modelagem, mas possuindo alguns outliers.

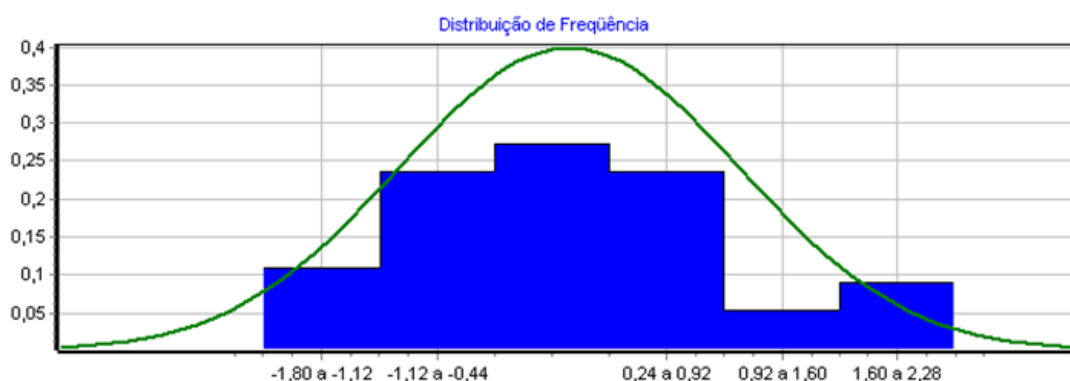
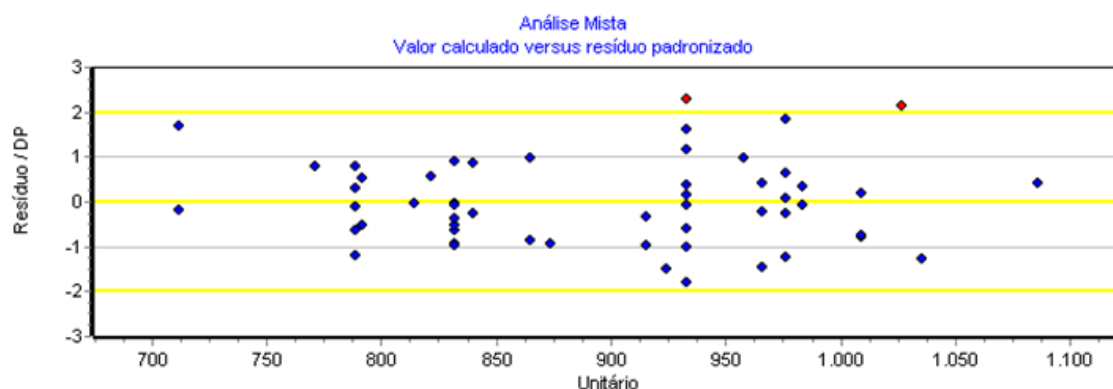


Gráfico dos Resíduos Padronizados versus Valores Ajustados

O gráfico dos resíduos versus valores ajustados (valores preditos) é uma das principais técnicas utilizadas para verificar as suposições dos resíduos. Além da detecção de heteroscedasticidade, esse gráfico pode indicar que não existe uma relação linear entre as variáveis explicativas com a variável resposta por meio de alguma tendência nos pontos.



Interpolação das Variáveis Independentes

Para calcular o preço unitário (PU) inicia-se pela interpolação dos atributos das variáveis independentes, que são resultado da modelagem de regressão linear pelo método dos mínimos quadrados, chegando aos seguintes resultados para as características dos apartamentos estudados.

Então, para avaliar um apartamento do tipo cobertura (duplex) com área global real de 527m², situado no bairro Bela Local, possuindo duas ou mais vagas de garagem de padrão construtivo alto e pronto para morar, o seu preço unitário calculado é:

Modelo de Regressão

$$PU = 689,90 - 17,81*(AG) + 59,26*(TU) + 50,53*(BL) - 39,98*(VG) + 144,08*(PC) - 101,99*(EC)$$

INTERPOLAÇÃO:

<i>Cálculo</i>	<i>Coeficientes</i>	<i>Atributo</i>	<i>Resultado</i>
Interseção (Constante)	689,90	1	\$ 689,90
Área Global (AG)	-17,81	1	\$ -17,81
Tipo de Unidade (TU)	59,26	1	\$ 59,26
Bairro Local (BL)	50,53	1	\$ 50,53
Vaga de Garagem (VG)	39,98	1	\$ 39,98
Padrão Construtivo (PC)	144,08	1	\$ 144,08
Estágio de Constr. (EC)	101,99	1	\$ 101,99
Preço Unitário (PU)			\$ 1.067,94

Feita a análise da interpolação das variáveis, verifica-se que a área global dos apartamentos de alto padrão não é significativa (valor acrescido e/ou descontado de \$ -17,81/m²) na formação do preço unitário obtido de \$ 1.067,94/m², por simples razão, as tipologias das unidades de apartamentos em questão possuem as suas áreas globais de porte, similar na totalidade das amostras pesquisadas.

3.3. Mensuração de Valores:

Para a mensuração de valores unitários, avaliar as características de especificações particulares de cada projeto de lançamento imobiliário, que no caso em estudo, iniciamos a nossa análise a partir da interpolação de todas variáveis categóricas explicativas do imóvel no modelo ajustado, observando as questões mercadológicas estudadas, tais como: “Área Global, Tipo de Unidade, Bairro Local, Número de Vaga, Padrão Construtivo e Estágio de Construção”.

Seleção de Variáveis:

Incorporar as variáveis categóricas e depois, por etapas, cada uma pode ser ou não eliminada, analisamos todas as variáveis estudadas. A decisão de retirada da variável é tomada baseando-se em testes do valor-p, que são calculados para cada variável como se ela fosse a última a entrar no modelo, ou seja, para cada variável explicativa proceder o cálculo do valor-p, então verifica-se a significância do regressor.

A variável resposta será baseada na diferença do padrão construtivo que o correspondente comprador está a adquirir, sendo esta categorizada como produto de interesse ou de mercado.

Como já se tinha referido, será o diagnóstico das variáveis categorizadas em 0 e 1, em que o resultado será a aquisição ou a desistência da compra do imóvel.

Assim feito, proceder a interpolação das variáveis categóricas explicativas, resultando na formação fundamentada dos preços unitários dos registros dos imóveis no mercado imobiliário, a partir das seguintes análises:

Diagnóstico do Padrão Construtivo:

Considerar a influência de formação de preços de apartamentos de alto padrão, com aplicação de todas as variáveis categóricas estudadas:

MENSURAÇÃO DE VARIÁVEIS:

Cálculo	Coeficientes	Atributo	Resultado
Interseção (Constante)	689,90	1	\$ 689,90
Área Global (AG)	-17,81	1	\$ -17,81
Tipo de Unidade (TU)	59,26	1	\$ 59,26
Bairro Local (BL)	50,53	1	\$ 50,53
Vaga de Garagem (VG)	39,98	1	\$ 39,98
Padrão Construtivo (PC)	144,08	0	\$ 0,00
Estágio de Constr. (EC)	101,99	1	\$ 101,99
Preço Unitário (PU)			\$ 923,86

O preço unitário estimado, calculado para o apartamento de padrão NORMAL, considerando as especificações do projeto de lançamento do imóvel é de \$ 923,86/m², que comparado com o imóvel de alto padrão apresenta um deságio de \$ 144,08/m².

Analogamente, considera-se somente a influência de formação de preços de apartamentos de padrão normal, com aplicação das variáveis categóricas mais utilizadas nas modelagens dos avaliadores:

Diagnóstico do Estágio da Construção:

A diferença de preço de mercado obtida por um imóvel de padrão NORMAL, em estágio de construção, em negociação no lançamento do projeto e das informações dos apartamentos ofertados nos plantões de venda nas modelagens realizadas pelos avaliadores, obtidas no modelo de regressão logística, são:

MENSURAÇÃO DE VARIÁVEIS:

Cálculo	Coeficientes	Atributo	Resultado
Interseção (Constante)	689,90	1	\$ 689,90
Área Global (AG)	-17,81	1	\$ -17,81
Tipo de Unidade (TU)	59,26	1	\$ 59,26
Bairro Local (BL)	50,53	1	\$ 50,53
Vaga de Garagem (VG)	39,98	1	\$ 39,98
Padrão Construtivo (PC)	144,08	0	\$ 0,00
Estágio de Constr. (EC)	101,99	0	\$ 101,99
Preço Unitário (PU)			\$ 821,87

Então o valor comparativo de um apartamento de alto padrão pronto para morar em relação de uma unidade de padrão normal em construção é de (\$ 821,87/m²), corresponde a uma variação por metro quadrado de \$ 246,07/m² em termos absolutos e 23,04% em termos relativos.

Tabela IX - Quadro Resumo – Resultado da Regressão

Apto	B	Erro	Stat t	valor-P	F. Sig.	Y	OR 90%.
Constante	689,90	71,19	9,69	0,00%	0,000072%	689,90	64,60%
Alto Padrão	144,08	24,39	5,91	0,00%		923,86	86,51%
Estágio Constr.	101,99	37,99	2,68	0,99%		821,87	76,96%
Preço Total						1067,94	100,00%

Vimos que o modelo alternativo de regressão logística se mostra um instrumento consistente e confiável para a análise dos preços praticados no mercado, podendo, deste modo, ser utilizada com assertividade na área da engenharia de avaliações quando da adoção do Método Comparativo Direto de Dados de Mercado para estimar valores de imóveis.

Os resultados do estudo da pesquisa apontaram que:

- os testes estatísticos evidenciaram que o modelo proposto é significativo para classificação do padrão construtivo dos imóveis;
- o modelo tem um percentual de acerto de formação dos preços de 64,60%;
- a maior margem de acerto está na classificação dos apartamentos de alto padrão (86,51%);
- as variáveis Área Global (AG) e Vaga Garagem (VG) não reduzem a probabilidade de o apartamento não ser de alto padrão;

Análise das Variáveis Isoladamente:

Avançamos na nossa análise das variáveis isoladamente, com a retirada das variáveis; vaga de garagem, salão de festas e piscina, pois não se mostraram significativas, por uma simples razão, todos os apartamentos de alto padrão, possuem estas variáveis analisadas, e que são intrínsecas a formação do padrão construtivo.

Como processo de análise a partir da interpolação das variáveis estudadas, fixamos o padrão construtivo e passamos analisar isoladamente cada variável explicativa, retirando uma por uma e verificando o resultado.

Então, comparativamente, o modelo de regressão logística identifica o peso de relevância de cada variável explicativa na análise de formação dos preços praticados no mercado, a ser francamente utilizada pelo engenheiro avaliador, conforme apresentado no quadro a seguir:

Tabela X - Quadro A1 – Interpolação das Variáveis

Variáveis	Coeficientes	Cálculo 1	Resultado	Cálculo 2	Resultado	Cálculo 3	Resultado
Interseção	\$689,90	1	689,89	1	689,89	1	689,89
Área Global	\$-17,81	1	-17,81	1	-17,81	1	-17,81
Unidade	\$59,26	1	59,26	1	59,26	1	59,26
Bairro	\$50,53	1	50,53	1	50,53	1	50,53
Nº Vaga	\$39,98	1	39,98	1	39,98	1	39,98
Padrão	\$144,08	1	144,08	1	144,08	0	0,00
Estágio	\$101,99	1	101,99	0	0,00	0	0,00
Unitário	Y		\$1.067,94		\$965,95		\$821,87

Variáveis	Coeficientes	Cálculo 4	Resultado	Cálculo 5	Resultado	Cálculo 6	Resultado
Interseção	\$689,90	1	689,89	1	689,89	1	689,89
Área Global	\$-17,81	1	-17,81	1	-17,81	1	-17,81
Unidade	\$59,26	1	59,26	1	59,26	0	0,00
Bairro	\$50,53	1	50,53	0	0,00	0	0,00
Nº Vaga	\$39,98	0	0,00	0	0,00	0	0,00
Padrão	\$144,08	0	0,00	0	0,00	0	0,00
Estágio	\$101,99	0	0,00	0	0,00	0	0,00
Unitário	Y		\$781,88		\$731,35		\$672,09

O modelo de regressão logística é composto pelas variáveis padrão construtivo, área global, apartamento tipo, bairro e se estágio de construção do prédio se está pronto para ocupação ou em construção.

Para a variável padrão construtivo, onde o padrão de referência são os imóveis de luxo (alto padrão), como conclusão, depois de todas as interações realizadas (Quadro A1 - Interpolação das Variáveis), com inserção e retiradas de variáveis, correspondem há uma configuração aceitável, onde obtém a probabilidade de 86,51% de que o modelo se encontra bem ajustado.

Diante desses resultados, conclui-se que a modelagem estatística elaborada, é fonte de estudos comparativos das variáveis independentes que o influenciam na tomada de decisão, como um instrumento estatístico de apoio à avaliação de imóveis.

A regressão logística pode ser uma técnica mais adequada e até mesmo mais consistente e assertiva do que a regressão clássica para estimar valores de imóveis ao eliminar a subjetividade das variáveis qualitativas e minimizar a imprecisão das variáveis quantitativas.

Tabela XI - Quadro Comparativo de Aplicação das Variáveis Categóricas

Variável	Regressão Clássica	Regressão Logística
X1 = Área Global	Valor real da unidade (em m ²)	0 qd < 500 m ² e 1 qd > 500 m ²
X2 = Tipo de Unidade	1 qd apto-tipo e 2 qd cobertura	0 qd tipo e 1 qd cobertura
X3 = Bairro	1, 2 e 3	0 qd outro e 1 qd Bela Vista
X4 = N. vagas	0, 1, 2, 3 e 4	0 qd poucas (até 1 vaga) e 1 qd muitas (2 ou mais vagas)
X5 = Padrão	1 qd normal e 2 qd alto	0 qd normal e 1 qd alto
X6 = Estágio Construtivo	1 qd inacabado e 2 qd concluído	0 qd inacabado e 1 qd concluído
Y = Unitário (\$/m ²)	O valor unitário em dólares americanos	O valor unitário (em \$/m ²)

4. CONSIDERAÇÕES FINAIS:

No estudo desenvolvido, constatou-se que existe relação significativa de atributos formadores do valor dos imóveis com as variáveis praticadas no mercado imobiliário analisado.

Na avaliação de bens em geral, dentre as diversas técnicas e abordagens, a aplicação de critérios econométricos é, por excelência, a mais praticado no mercado imobiliário.

Então, nesse sentido, se busca sistematizar a alternativa para aplicação da regressão logística ou semi-logística, o uso da variável categórica (*dummy*) como atributo de comparação fundamentada na mensuração das diferenças existentes entre os dados amostrais, através do emprego do método econométrico no questionamento dos preços praticados no mercado de avaliações de bens, em busca das variáveis formadoras dos preços, dos quais se determinar o valor de mercado com base nas evidências dos dados amostrais, respeitados os níveis de rigor definidos pelas normas técnicas de avaliação, especificados quanto aos graus de fundamentação e precisão.

Devido à facilidade do emprego da metodologia de análise da variável categórica, recomendada para uso em quaisquer tipos de avaliações, seja de: terrenos, glebas urbanas e rurais, depósitos, pavilhões, prédios, apartamentos, lojas, escritórios, veículos, máquinas, equipamentos, etc.

Na prática este estudo estabelece estatisticamente de como fazer o uso das variáveis categóricas (dicotômicas) que tecnicamente mostraram-se eficazes como ferramenta de análise de valor no processo de avaliação imobiliária, devendo este ser um instrumento corriqueiro da aplicação da técnica avaliatória pelo engenheiro avaliador.

Porém, como resultado final, é importante verificar-se que todas as variáveis (resposta ou preditoras) forem efetivamente analisadas e medidas, independentemente de que forma, se qualitativamente ou quantitativamente, desde que atendem aos níveis de confiança e os graus de fundamentação e de precisão estabelecidos pelos normativos da ABNT (NBR 14653), isto quer dizer, que se faz relevante antes de qualquer avaliação fazer análise de quais variáveis que realmente são importantes formadoras dos preços praticados no mercado imobiliário a ser estudado.

Finalmente, a regressão logística é o algoritmo de classificação mais conhecido e utilizado (MATHEUS, 2017), assim como a de regressão linear múltipla, mas ainda assim é um total desconhecido, não explorado ou utilizado na engenharia de avaliações podendo ser, entretanto, uma abordagem metodológica a ser utilizada como alternativa à regressão recorrentemente utilizada, como ficou demonstrado no presente estudo, já que a regressão semi-logística desenvolvida apresentou alta acurácia, além de ser extremamente simples, portanto, recomendável o seu uso em trabalhos profissionais e acadêmicos.

BIBLIOGRAFIA:

- ABNT. Associação Brasileira de Normas Técnicas 1989, *NBR 5676 – Norma Brasileira de Avaliação de Imóveis Urbanos*, Rio de Janeiro, RJ, 1989.
- Barbosa F^o, D. S., 1992, *Examinando os Modelos de Regressões*, CBAP, Avalien Ed., Porto Alegre, RS, 1992.
- Blackadar, C.G, 1989, *Dynamic Capitalization: An Income Approach in Real Dollars ar Real Interest*”, *The Appraisal Journal* July 338-371, 1989.
- Cabral, C.I.S, 2013, *Aplicação do Modelo de Regressão Logística num Estudo de Mercado*, Projeto de Mestrado da Universidade de Lisboa, Faculdade de Ciências, Lisboa, PT, 2013
- Donaire, D.; Guerra, M., 1982, *Estatística Indutiva – Teoria e Aplicações*, LTC Ed., São Paulo, SP, 1982.
- Duarte, A. M. e Gabbay, A., 2001, *Busca da Eliminação ou Isenção da Subjetividade na Modelagem de Mercados Imobiliários*, Anais do XI COBREAP, Guarapari, ES, 2001.
- Duarte, A. M.; Rottmann, E.; Duarte, A.R.C. de L. M.; Costa Jr., C.T., Kawashima, M. e Galvão Neto, O., 2007, *Um Novo enfoque na Engenharia de Avaliações: O Método do Entendimento Racional*, Anais do XIV COBREAP, Salvador-BA, 2007.
- Hoffmann, R.; Vieira, S., 1983, *Análise de Regressão – Uma Introdução a Econometria*, Ed. Hucitec, São Paulo, RS, 1983.

- Levine, D.M.; Stephan, D.F.; Krehbiel, T.C.; Berenson, M.L., 2015, *Estatística – Teoria e Aplicações, Usando o Microsoft Excel em Português*, 6ª edição. Ed. LTC, Rio de Janeiro, RJ, 2015.
- Missio, F. e Jacobi, L.F., Variáveis Dummy: Especificações de Modelos com Parâmetros Variáveis, *Ciência e Natura*, UFSM, Artigo, 2007.
- Cardoso, C.E., *Normas IVS – Normas Internacionais de Avaliação*, Internacional Valuation Standards Council, Versão 2011, IVCS e IBAPE.
- Pereda, P.C.; Alves, D., 2018, *Econometria Aplicada*, Ed. Elsevier, Rio de Janeiro, RJ, 2018.
- Pereira, J.C. R., 1999, *Análise de Dados Qualitativos*, Ed. EDUSP, São Paulo, SP, 1999.
- Portal Action é mantido pela Estatcamp - *Consultoria Estatística e Qualidade*, São Carlos - SP
- Saldanha, M., 1998, *Aplicação da Variável Dummy no Questionamento do Valor de Mercado*, CBAP, Avalien Ed., Porto Alegre, RS, 1998.
- Tognetti, G, 2015, *Qual é a valorização dos imóveis no Brasil? Uma análise pelo Índice Fipe-Zap*, Portal RExperts, São Paulo, SP, 2020.