

MARCELO SUAREZ SALDANHA

# ANÁLISE QUALITATIVA DE DADOS Variáveis Categóricas

ABNT NBR 14.653



São Paulo – SP

2023

# Sumário

Apresentação.....	13
Prefácio.....	15
1. Introdução .....	17
1.2 Estatística Aplicada.....	20
1.3 Aspectos Fundamentais da Avaliação .....	21
2. Avaliação no Brasil.....	23
2.1 Procedimentos Gerais .....	25
2.2 Utilização de Modelos de Regressão.....	29
2.3 Pressupostos Normativos (ABNT) .....	30
2.3.1 Verificação dos Pressupostos do Modelo.....	33
3. Mercado imobiliário.....	35
3.1 Pesquisa de Mercado Imobiliário .....	38
3.2 Validação da Informação de Mercado .....	39
3.3 Pesquisa do Evento Qualitativo .....	41
4. Abordagem de comparação.....	45
4.1 Processo de Amostragem .....	46
4.2 Tipo de Variáveis .....	46
4.3 Questões Éticas.....	48
5. Regressão múltipla .....	51

5.1	Modelagem de Análise de Regressão .....	52
5.2	Construção de Modelos de Regressão .....	53
5.3	Armadilhas da Inferência.....	57
5.4	Pressupostos para Validação.....	58
6.	Análise estatística .....	59
6.1	Presença do Erro .....	59
6.2	Teste de Hipóteses – (significância do Valor-p) .....	61
6.3	Análise de Variância – ANOVA (F de significação).....	64
6.4	Análise de Correlação (r) .....	67
6.5	Análise de Resíduos ( <i>outliers</i> ).....	70
6.6	Aspectos Econométricos .....	71
6.7	Problemas comuns na análise econométrica .....	74
6.7.1	Multicolinearidade.....	74
6.7.2	Heteroscedasticidade .....	75
6.7.3	Endogeneidade .....	75
6.7.3.1	Covariância ou Variância Conjunta .....	76
6.7.3.2	Interpretação da endogeneidade .....	76
7.	Regressão logística.....	79
7.1	Modelo de Regressão Logística Simples .....	81
7.2	Modelo de Regressão Logística Múltipla .....	84
7.2.1	Variáveis Independentes Categóricas .....	85
7.2.2	Seleção de Variáveis Independentes.....	86
7.2.3	Medidas de Qualidade do Ajuste do Modelo:.....	88
7.2.4	Diagnóstico de Influência dos Dados .....	89
8.	Software Microsoft-Excel.....	91
8.1	Uso da Planilha Excel – Análise de Dados.....	92
8.2	Análise de Dados por Regressão Linear .....	94
8.3	Tabela Dinâmica – Análise da Base de Dados .....	94
8.4	Análise de Correlação – Independência das Variáveis.....	96
8.5	Parâmetros da Estatística de Regressão .....	97

<b>9. Avaliação imobiliária urbana.....</b>	<b>99</b>
<b>9.1 Lançamentos Imobiliários.....</b>	<b>101</b>
<b>9.2 Análise Exploratória de Dados.....</b>	<b>105</b>
<b>9.3 Emprego das Variáveis Categóricas.....</b>	<b>106</b>
<b>9.4 Análise dos Preços dos Apartamentos .....</b>	<b>109</b>
<b>9.5 Espectro Amostral Analisado .....</b>	<b>113</b>
<b>10. Resultados e discussões .....</b>	<b>119</b>
<b>10.1 Medidas de Tendência Central .....</b>	<b>119</b>
<b>10.2 Medidas de Dispersão .....</b>	<b>120</b>
<b>10.3 Análise Quantitativa.....</b>	<b>121</b>
<b>10.3 Análise Qualitativa .....</b>	<b>133</b>
<b>10.4 Mensuração de Valores .....</b>	<b>146</b>
<b>11. Mais aplicações.....</b>	<b>155</b>
<b>11.1 Avaliação Imobiliária Rural .....</b>	<b>155</b>
<b>12. Reflexão .....</b>	<b>161</b>
<b>Bibliografia .....</b>	<b>165</b>



# Prefácio

A engenharia é tecnologia, ciência aplicada, ofício, arte e muito mais coisas que se possa definir, conceituar ou denominar. Mas, fundamentalmente, ela é a atividade que materializa projetos, estudos, pesquisas e obras que atendem necessidades humanas e resolvem questões das mais diversas ao elaborar uma peça técnica digital, um Laudo Pericial, uma estrutura física em madeira, aço ou concreto, uma tese acadêmica, etc.

As necessidades humanas e as questões que a sociedade demanda para a engenharia são virtualmente ilimitadas e infinitas, sempre crescentes e cada vez mais diversas, o que faz com que o campo de atuação dos engenheiros e das engenheiras também assim o seja. Isso requer que aqueles e aquelas que abraçam essa profissão sejam ao mesmo tempo holísticos/generalistas e especialistas.

Dentre as suas especialidades a Engenharia de Avaliações é uma das mais empolgantes, importantes e dinâmicas e o Engenheiro Marcelo Suarez Saldanha, um dos grandes profissionais tanto da área de Avaliações quanto de Perícias, apresenta para a comunidade técnica esta obra que se constitui em uma enorme contribuição ao trazer uma abordagem inovadora quando propõe o tratamento dos dados pesquisados no mercado imobiliário por

meio da regressão logística ou semi-logística, sem negligenciar dos procedimentos e tratamentos tradicionalmente adotados.

Ao adotar a abordagem inovadora proposta nesta obra, o(a) profissional ou acadêmico terá uma ferramenta conceitual e operacional poderosa e eficiente para identificar as relações de causa e efeito entre as variáveis explicativas ( $X$ 's) e a explicada ( $Y$ ), ou seja, poderá mensurar e ranquear de forma metodológica, racional e objetiva as forças ou intensidades da “causação” ou da “causalidade”, indo além da simples identificação de coeficientes estatísticos das relações, correlações ou determinações, que são inquestionavelmente importantes, mas muitas vezes são esdrúxulos e, com isso, “enganadores”.

Desta forma, essa nova abordagem poderá fazer com que o diagnóstico do mercado e a aplicação do Método Comparativo Direto de Dados de Mercado para realizar a identificação de bens imóveis sejam mais consistentes e assertivos e, desta forma, os trabalhos de Engenharia de Avaliações que a adotarem tenham ainda mais robustez e precisão, o que trará maior segurança para as demandas da sociedade, demandas essas cada vez maiores, mais diversas e mais exigentes.

Esta é uma obra que já nasce clássica por sua originalidade, profundidade e aplicabilidade e certamente deverá estar nas pastas, estantes ou bibliotecas, sejam digitais ou nas prateleiras físicas, de todos e de todas que trabalham, se dedicam, gostam um pouco ou até mesmo tenham paixão por esse ofício, como muitos de nós, Engenheiros e Engenheiras de Avaliações, temos.

**André Montenegro Duarte**

Professor da Faculdade de Engenharia Civil – FEC  
Instituto de Tecnologia da Universidade Federal do Pará – ITEC/UFPA  
Membro do Instituto Brasileiro de Avaliações e  
Perícias de Engenharia – IBAPE



# Introdução

Na avaliação de imóveis, várias técnicas avançadas podem ser utilizadas para a análise comparativa de dados de mercado, com o objetivo de obter uma apreciação mais precisa da formação de preços, com utilização da metodologia da regressão logística, cuja aplicação de variáveis categóricas (dicotômicas) mostrou-se eficaz como ferramenta de probabilidade de análise de valor no processo de avaliação imobiliária.

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, tais como áreas, idade, número de vagas de garagem, número de dormitórios etc., e também de variáveis 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, tais como padrão construtivo, localização, estado de conservação, etc., as quais podem assumir valores diversos em qualquer quantidade, finita ou não.

Então, independente das variáveis que se deseja estudar, a definição da amostragem exige muita experiência do analista do assunto, além de bom senso.

Fundamentalmente, existem dois tipos de amostragem: a probabilística e a não-probabilística. A amostragem será probabilística se todos os elementos da população tiverem probabilidade conhecida e não nula de pertencer a amostra; caso contrário, ela será não-probabilística.

### 1.3 Aspectos Fundamentais da Avaliação

De modo geral, é sabido que a informação é um aspecto fundamental de qualquer processo de avaliação, é resultado de uma série de propriedades comparáveis permitem uma coleta mais fiel do comportamento de mercado. As técnicas de inferência estatística usam a hipótese de 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. Portanto, sempre que possível, para obter uma amostra que seja representativa da população, deve-se optar pela amostragem probabilística.

A coleta da amostragem no mercado imobiliário é a esmo (LOPES, 1998), sem norma, uma vez que não utiliza nenhum critério para escolher os elementos da amostra. O avaliador deve ser aleatório na busca em uma amostra homogênea, que não seja influenciada por qualquer característica dos elementos da população; só assim esse processo será equivalente ao de uma amostragem probabilística (SALDANHA, 1999).

Em estudos anteriores, DUARTE e GABBAY (2001) abordaram, num experimento proposto, o estudo do mercado imobiliário da Cidade de Belém-PA, restringindo-se ao segmento de apartamentos residenciais, modelando, através da análise de regressão, o valor de venda de apartamentos de forma absolutamente objetiva, sem qualquer “opinião” originária da subjetividade intrínseca do ser humano.

O principal objetivo são as modelagens de análise de regressão, que constituem uma das ferramentas estatísticas mais importantes na análise estatística de dados quando se pretende modelar relações entre variáveis. A ideia destes modelos é explorar a relação entre uma ou mais variáveis explicativas (ou independentes) e uma variável resposta (ou dependente). Um dos casos particulares dos modelos lineares generalizados são os modelos onde a variável resposta apresenta apenas duas categorias ou que de alguma forma foi dicotomizada assumindo valores 0 ou 1 sendo o modelo de regressão logística o mais popular desses modelos.

Ao propor CABRAL (2013) uma abordagem objetiva quando trabalhando com variáveis qualitativas na análise de formação dos preços com a adoção de regressão logística nas avaliações de imóveis, buscamos alcançar explicação de origem do valor de mercado e seus fundamentos da 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.



## Avaliação no Brasil

A avaliação de imóveis é um trabalho de engenharia especializado, que passa ao largo de mera opinião de valor. Deve obrigatoriamente, conforme artigo 39 do Código de Defesa do Consumidor (CDC), atender as normas técnicas pertinentes, como por exemplo a ABNT NBR 14653 partes 1 – Procedimentos Gerais e 2 – Avaliação de Imóveis Urbanos. Portanto, deve ser feita por profissional registrado no Conselho de Engenharia e Arquitetura (CREA), pois contém atividades exclusivas da classe como, Vistoria e Avaliação, além de que para cada trabalho deve ser recolhida uma anotação de responsabilidade técnica (ART).

Os procedimentos avaliatórios usuais no Brasil são estabelecidos pelas normas técnicas da ABNT (NBR 14.653), e têm por finalidade identificar o valor de um bem, de seus frutos e direitos e o seu custo, bem como determinar indicadores de viabilidade, etc.

A metodologia aplicável é função, basicamente, da natureza do bem avaliando, da finalidade da avaliação e da disponibilidade, qualidade e quantidade de informações colhidas no mercado. O objetivo é retratar o comportamento do mercado por meio de modelos que suportem racionalmente o convencimento do valor, sendo facultado ao Engenheiro de Avaliações, em situações atípicas onde ficar comprovada a impossibilidade de utilizar as metodologias

### 2.3.1 Verificação dos Pressupostos do Modelo

Os pressupostos de análise do modelo de regressão (Anexo A do Normativo), que deverão ser verificados são:

- a) linearidade: o comportamento da variável dependente em relação a cada variável independente, em escala original.
- b) transformações utilizadas para linearizar o modelo devem, tanto quanto possível, refletir o comportamento do mercado, com preferência pelas transformações mais simples de variáveis, que resultem em modelo satisfatório.
- c) testes de significância: o nível de significância máximo admitido nos demais testes estatísticos não devem ser superiores a 10%.
- d) poder de explicação do modelo pode ser aferido pelo seu coeficiente de determinação. 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.
- e) códigos alocados: seus critérios para construção da escala deverão ser explicitados. A escala será composta por números naturais consecutivos em ordem crescente, em função da importância das características possíveis na formação do valor, com valor inicial igual a 1. Não é necessário que a amostra contenha dados de mercado em cada uma das posições da escala construída.
- f) códigos ajustados: admite-se que os códigos sejam extraídos da amostra por meio de modelo de regressão com a utilização de variáveis dicotômicas, desde que haja pelo menos três dados por característica.

- g) no caso de utilização no mesmo modelo de regressão de diferentes agrupamentos (tipologia, mercados, localização, usos etc.), recomenda-se verificar a independência entre os agrupamentos, entre as variáveis utilizadas, e possíveis interações entre elas.
- h) a variável dependente no modelo de regressão deve ser apresentada no laudo na forma não transformada.



## Mercado imobiliário

O mercado imobiliário funciona comparativamente sobre dois pontos de vista diferentes (MATOS e BARTKIW, 2011): de quem vende imóveis e de quem compra imóveis. Primeiramente, de quem vende imóveis, pode se dizer que a atividade é baseada em oferta e demanda, e, para quem compra, o ponto de vista é outro; é para morar ou para investir. Então, o mercado imobiliário pode ser dividido em duas categorias: o mercado de quem compra imóveis para morar e de quem compra para investir.

Especificamente, o mercado imobiliário em momento de expansão, se torna ainda mais promissor quando o assunto é investimento. Por isso, essa pode ser uma ótima opção para quem pretende proteger o seu patrimônio ou aumentar os lucros. Para isso, é necessário aprofundar seus conhecimentos mercadológicos, estudar como andam as ofertas e demandas na sua região e analisar aos atributos e formações dos preços; pergunta-se, por exemplo, se vale mais a pena comprar um imóvel na planta ou pronto para morar.

Entender como funciona é um processo que exige muito estudo e atenção. O mercado imobiliário é caracterizado em fases de sazonalidades, como, por exemplo, a fase onde a demanda é superior às ofertas em uma determinada região; em casos assim, as

interpretação das variáveis serem utilizadas para análise e tratamento dos dados, como por exemplo:

- Terrenos e Glebas: área superficial, topografia, uso do solo urbano ou rural, regime urbanístico, localização na malha urbana, distância a pólos valorizantes ou desvalorizantes, vista, valor unitário, etc.
- Imóveis Residenciais e Comerciais: área privativa, área global, frente, pé-direito, padrão construtivo, vagas de garagens, localização, idade da construção, instalações especiais, valor unitário, etc.
- Casas: áreas do terreno e construída, relação entre áreas (edificada, jardim e pátio), frente(s) para logradouro(s), padrão construtivo, dependências, concepção de projeto, idade da construção, localização na malha urbana (unidade isolada ou em condomínio), regime urbanístico, infraestrutura urbana, valor unitário, etc.
- Apartamentos: área privativa, números de cômodos, padrão da construção, tipologia das unidades, instalações especiais, vagas de garagem, números de pavimentos e de unidades por andar, vista e orientação solar, vizinhança, infraestrutura urbana e de lazer, valor unitário, etc.
- Fazendas: área total, áreas reservadas, topografia, classe de uso e exploração da terra, nota agrônômica, localização, acesso de escoamento da produção, benfeitorias reprodutivas (culturas, árvores, etc.) e não reprodutivas (cercas, currais, casas de moradia, galpões, silos, etc.), valor do hectare, etc.





## Abordagem de comparação

Uma possível abordagem para o problema de comparação de um modelo de regressão múltipla é fazer cada comparação independentemente, usando um procedimento estatístico adequado.

A dificuldade com a abordagem “por comparação” para comparações múltiplas é a possibilidade do aumento da probabilidade do Erro Tipo I ou, equivalentemente, a possibilidade de diminuição do nível de confiança global.

Uma análise estatística de dados enganosos produz conclusões enganosas. A questão da qualidade dos dados pode ser mais sutil. Em previsões, não há concordância sobre a precisão de uma medida de previsão.

Os testes de hipótese enfatizam a rejeição, que é baseada na probabilidade em vez da aceitação. Isso requer mais etapas de lógica. De acordo com David Bakan, a probabilidade de rejeitar a hipótese nula é uma função de cinco fatores: se o teste tem uma ou duas caudas, o nível de significância, o desvio padrão, a quantidade da hipótese nula e o número de observações.

Estes fatores são fontes de críticas, uma vez que os fatores sob controle do avaliador conferem aos resultados a aparência de subjetividade. Um teste de hipóteses de sucesso está associado à

O erro de amostragem, a seleção da amostra é simples, e se for obtida intencionalmente, sem que seja feitas referências ao tamanho da amostra e a margem de erro de modo a defender tal resultado sem qualquer significado de variação com base na probabilidade.

O erro de medição passa ser uma questão de ética diante da seguinte situação a pesquisa é escolhida e orientada em uma direção específica e intencionalmente com informações falsas.

As questões éticas surgem também quando os resultados de amostras não são probabilísticos para forma as conclusões sobre a população estuda. Os resultados não podem ser generalizados além dos limites da amostra.

Quanto à análise das hipóteses estudadas, as questões éticas surgem quando o processo relativo ao teste de hipótese é manipulado. Incluem o método de coleta de dados, o tipo de teste, a escolha de significância, o filtro e o descarte dos dados, bem como deixar de divulgar as descobertas pertinentes da pesquisa, ou seja, deixar de fazer a distinção entre uma metodologia de pesquisa precária e um comportamento fora dos padrões de ética.

E, por último, quando se deseja realizar previsões manipula-se o processo de desenvolvimento do modelo de regressão múltipla, pois a questão chave é a intenção, isto é, há um comportamento fora dos padrões éticos quando o avaliador utiliza a análise de regressão intencionalmente e deixa de remover da avaliação as variáveis que representam uma elevada colinearidade entre as outras variáveis independentes, ou intencionalmente deixa de utilizar métodos outros que não a regressão dos mínimos quadrados quando os pressupostos necessários para a regressão são seriamente violados.





## Regressão múltipla

Partimos o estudo da análise de regressão linear múltipla, utilizada quando se quer medir a relação de uma variável dependente binária com uma ou mais variáveis independentes, sendo que as independentes tanto podem ser categóricas ou não.

Preliminarmente, referenciamos conceitualmente o processo de avaliação estabelecido pela norma IVSC (International Valuation Standards Council), em harmonização e adequação às atividades básicas da avaliação no Brasil, representadas pelas normas brasileiras estabelecidas pela ABNT e pelo IBAPE, se resumem nos seguintes princípios básicos:

- Princípio da Lei de Mercado “Supply and Demand”, que trata dos preços praticados no mercado imobiliário, ou seja: é a lei da oferta e da procura, que é conhecida por todos, onde, havendo equilíbrio dos preços médios no mercado, haverá equilíbrio na oferta e na procura;
- Princípio da Comparação “Comparison Principle”, metodologia que define o método comparativo, segundo o qual bens semelhantes, em mercados semelhantes, tem preços semelhantes, enquanto as eventuais diferenças de preços são proporcionais às diferenças entre as características dos bens e dos mercados.

- Escolher, com base nos modelos apresentados, o que melhor explica o valor de mercado;
- Realizar uma análise completa do modelo escolhido, incluindo a análise de resíduos;
- Acrescentar os termos de interação, eliminar dados, transformar as variáveis, tabular os dados e analisar novamente o modelo e os dados;
- Utilizar o modelo selecionado para fins de previsão e inferências.

### 5.3 Armadilhas da Inferência

Com a ampla disponibilidade de *softwares* estatísticos e planilhas eletrônicas, tornou-se bem mais fácil fazer a análise de regressão; entretanto, para muitos usuários, essa maior disponibilidade de pacotes estatísticos não vem acompanhada pelo entendimento de como utilizar a análise de regressão de maneira apropriada. Não se pode esperar que alguém que não esteja familiarizado com os pressupostos da regressão com o modo de avaliar esses pressupostos e conheça quais são as alternativas válidas à regressão dos mínimos quadrados caso um determinado pressuposto seja violado (LEVI-DE, STEPHAN, KRENBIEL, BERENSON, 2012),

Algumas armadilhas envolvidas na utilização da análise de regressão devem ser evitadas:

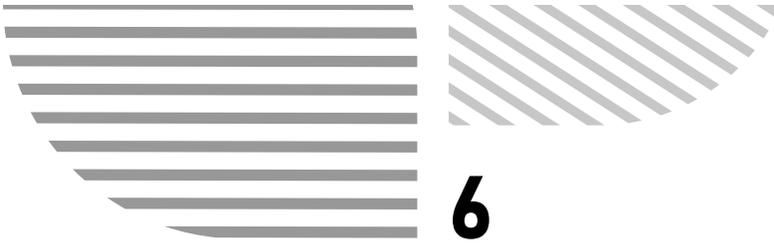
- Deixar de atentar para os pressupostos da regressão do MMQ;
- Não saber como avaliar os pressupostos da regressão do MMQ;
- Desconhecer quais são as alternativas à regressão MMQ caso um pressuposto seja violado;

- Utilizar um modelo de regressão sem conhecimento do assunto;
- Extrapolar além do intervalo de análise;
- Concluir que uma relação significativa identificada em uma observação é decorrente de uma relação de endogeneidade;
- Modelar o ruído aleatório nos dados, essa condição é o “overfitting” do modelo e produz valores R-quadrados enganosamente tendenciosos para fazer previsões.

## 5.4 Pressupostos para Validação

Os pressupostos de validação do modelo de regressão estão relacionados com o teste de hipóteses e análise de variância, de modo que o teste t e o teste F de ANOVA confirmem a análise de significância da regressão em relação aos pressupostos, que são os seguintes:

- Linearidade – confirmar que a relação entre as variáveis é linear;
- Independência de Erros – requerer que os erros sejam independentes entre si;
- Normalidade de Erros – requerer que os erros sejam normalmente distribuídos para cada um dos valores de X;
- Igualdade de Variâncias – homoscedasticidade, requerer que a variância dos erros seja constante em relação a todos os valores de X;



# Análise estatística

A técnica estatística usual para determinar a presença do erro é conhecida como teste t. Com esse teste, assim como com qualquer outro teste de hipóteses, existem chances de cometermos erros.

## 6.1 Presença do Erro

Um dos possíveis erros é rejeitar a hipótese nula, quando esta é verdadeira (Erro Tipo I) ou então aceitar a hipótese nula, quando esta é falsa (Erro Tipo II). Qualquer regra para decidir entre as hipóteses  $H_0$  e  $H_1$  é avaliada em termos das probabilidades dos dois tipos de erros.

O nível de significância para o teste t é especificado pelo valor-p, onde o avaliador controla a probabilidade de encontrar diferenças errôneas. Quando cada um dos vários testes de hipóteses é feito ao mesmo nível de significância, o valor-p é chamado de nível de significância por comparação.

Em geral, o erro acontece quando funções, fórmulas e estatísticas não conseguem explicar ou modelar totalmente um valor verdadeiro ou teórico. Em outras palavras, o erro é a diferença entre um valor real e o predito. Embora possa existir algum grau de erro

sobre  $Y$ , pode ter várias razões para a invalidez do teste de hipóteses, nas quais se destacam:

- a) Problemas de especificação: a forma funcional da relação entre  $Y$  e  $X$  está errada;
- b) Omissão de variáveis: a variável omitida é incorporada ao erro do modelo estimado, que por sua vez será correlacionada com as demais variáveis explicativas;
- c) Erro de medida da variável explicativa: quando a variável explicativa for medida com um erro não aleatório, ou seja, correlacionado com covariadas;
- d) Problema de simultaneidade: quando o sentido da causalidade entre  $Y$  e  $X$  não é claro. Nestes casos, a variável  $X$  também será endogenamente determinada, e, portanto, isso levará o viés quando MMQ for usado para estimar o efeito de  $X$  em  $Y$ .





## 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

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,  $\sum y_j = n_1$  que representa o total de indivíduos com  $y = 1$ .

### 7.2.4 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.





## Software Microsoft-Excel

A disponibilidade e simplicidade do Software Microsoft-Excel é a principal característica do uso da planilha Excel, que pode ser encontrada em praticamente qualquer computador ou lugar: em casa, no trabalho ou no cliente, cuja portabilidade da planilha pode ser executada em desktops, laptops, notebooks e handhelds, sem maiores dificuldade. (MOURA, 2013).

### **Vantagens da Ferramenta de Análise**

- Facilidade de uso – execução de cálculos sofisticados como modelagens de regressões lineares e não lineares, análises de variância, etc. podem ser facilmente executados por meio de suas funções ou de acesso as macros em VBA.
- Variáveis nomeadas – o Excel oferece a possibilidade de o usuário trabalhar com variáveis nomeadas como “preço”, “área”, “padrão”, “localização”, etc.
- Biblioteca de funções – o Excel possui internamente uma extensa gama de funções de engenharia (estatísticas, financeiras, etc.), acessíveis ao engenheiro avaliador, sem a necessidade de uma linha de código sequer.

- **Integração** – o Excel se comunica com facilidade com outros programas da suíte Office e do Windows, podendo executá-los ou ser executados a partir deles.
- **Abrangência** – o uso do Excel não tem limites, estendendo-se em redes locais e à nuvem. A quantidade de planilhas, programas escritos em Excel é muito grande, basta acessar no GOOGLE.

## 8.1 Uso da Planilha Excel – Análise de Dados

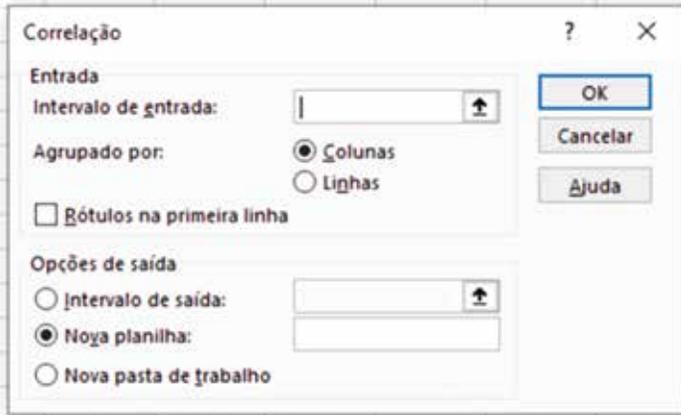
Para incluir funções de VBA (Visual Basic for Application) nas Ferramentas de Análise, se pode carregar o Suplemento de Ferramentas de Análise – VBA da mesma forma que carrega as Ferramentas de Análise. Na caixa Suplementos disponíveis, marcar a caixa de seleção Ferramentas de Análise – VBA.

E para acessar as ferramentas de análise VBA, clicar em Análise de Dados no grupo Análise na guia Dados. Se o comando Análise de Dados não estiver disponível, deve carregar o programa suplementar Ferramentas de Análise.

### **Carregar e ativar as Ferramentas de Análise:**

1. Clique na guia Arquivo, em Opções do Excel e depois na categoria Suplementos.
2. Na caixa Gerenciar, selecione Suplementos do Excel e clique em Ir.
3. Se estiver usando o Excel para Mac, no menu arquivo, acesse Ferramentas > Suplementos do Excel.
4. Na caixa Suplementos, marque a caixa Ferramentas de Análise e clique em OK, se as Ferramentas de Análise não

acima de 0,70 entre as variáveis independentes caracteriza indícios de multicolinearidade e gera restrições de utilização do modelo, na mesma direção da dependência linear observada.



## 8.5 Parâmetros da Estatística de Regressão

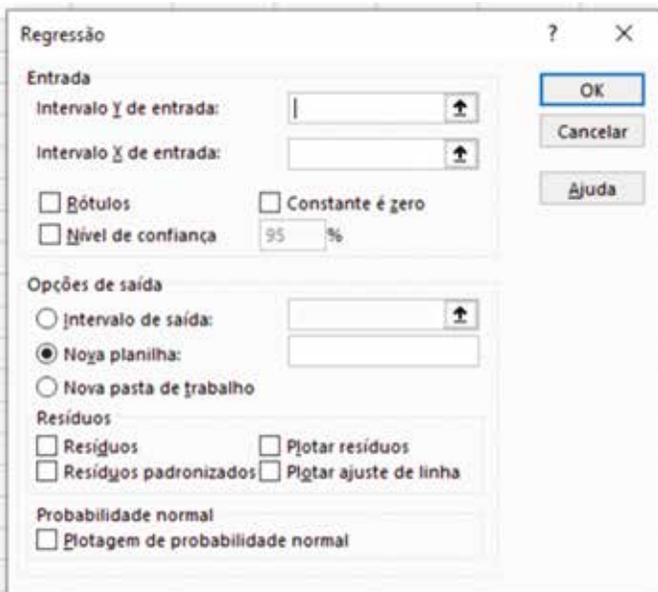
A ferramenta de análise de regressão executa uma análise de regressão linear usando o método dos mínimos quadrados para ajustar uma linha em um conjunto de observações, ou seja, analisar como uma única variável dependente é afetada pelos valores de uma ou mais variáveis independentes.

Na opção intervalo de entrada de  $Y$ , selecionar a coluna dos preços unitários (variável dependente) e no intervalo de entrada de  $X$ , selecionar o bloco de variáveis independentes que se deseja testar. Habilitar rótulo, resíduos e resíduos padronizados.

Observar nos resultados da estatística de regressão o poder de explicação do modelo dado por  $R$ -quadrado (quanto maior, maior é o poder de explicação do modelo), a significância global do modelo de regressão é dado pelo  $F$  de Significação (quanto menor for este valor, menor será a probabilidade de erro quando se afirma que

o conjunto de variáveis considerado é importante para explicar a variabilidade observada nos preços unitários analisados. Os coeficientes obtidos para montagem do modelo (interseção e as variáveis independentes analisadas) de regressão.

A análise de cada variável independente no modelo será medida pelo *valor-P* que indica o nível de significância de cada parâmetro a ela associado (quanto menor for o valor-P, menor será a probabilidade de erro quando afirmar que a respectiva variável é importante na formação do modelo de regressão. Quanto aos resíduos, se pode utilizar o assistente gráfico para construir o gráfico dos resíduos padronizados *versus* valores ajustados (previsto), para verificar as hipóteses de homoscedasticidade, normalidade, não correlação e *outliers* (a situação ideal é um gráfico com pontos dispostos aleatoriamente sem nenhuma tendência) ou verificar se a maioria dos pontos (Resíduos padrão) encontram-se no intervalo  $[-2;+2]$ , os que não pertencem ao intervalo são considerados *outliers*.





## Avaliação imobiliária urbana

O estudo caso proposto para análise dos preços dos imóveis praticados na cidade de Porto Alegre, Rio Grande do Sul, Brasil, é modelar a formação de preços dos imóveis, com a análise de técnicas avançadas a ser utilizadas na avaliação direta de dados extraídos do mercado imobiliário, com utilização da metodologia da regressão logística, obtendo uma apreciação mais precisa da origem do valor, com aplicação de variáveis categóricas (dicotômicas) como ferramenta eficaz de análise de probabilidade do valor de mercado no processo de avaliação imobiliária.

Análise qualitativa deu-se a partir de uma pesquisa realizada em zona de incorporação imobiliária de imóveis de alto padrão, cujas informações foram obtidas junto aos plantões de venda das empresas imobiliárias, através das informações constantes do material de publicidade e tabelas dos valores de comercialização dos imóveis.

Na venda de um lançamento de classe média ou classe média alta, a demonstração da tabela deve ser levada a sério, em qualquer venda, os aspectos do imóvel, como localização, arquitetura, áreas coletivas, garagens e tudo o mais que você puder apontar como qualidades do empreendimento, pois existem quesitos que podem ser qualificados como se tendo uma maneira certa do momento da

apresentação e, de repente, em uma venda de imóvel de luxo, você nem precisa mostrar esse comparativo – isso tudo também vai depender da avaliação qualitativa do imóvel.

### Bairro Bela Vista – Porto Alegre-RS, Brasil



O bairro Bela Vista de Porto Alegre é sem dúvida um dos bairros que mais abriga apartamentos de luxo, atrás apenas do tradicional bairro Moinhos de Vento. Muito requisitado pelos porto alegrenses, o bairro Bela Vista se destaca pela beleza natural e pela boa localização dentro da cidade.

Dentre os empreendimentos analisados em Porto Alegre, buscamos analisar o valor do m<sup>2</sup> de imóveis de alto padrão, construídos ou em construção nas regiões, cujos lançamentos imobiliários possuem os preços mais caros comercializados nos principais bairros de incorporação da cidade, como no bairro Bela Vista e imediações, no sentido de posicionar os vendedores e compradores a respeito dessas informações mercadológicas.

<b>Endereço Logradouro</b>	<b>Bairro</b>	<b>Preço US\$</b>	<b>Apto Tipo</b>	<b>Nº Vaga</b>	<b>Área m<sup>2</sup></b>	<b>Unitário US\$/m<sup>2</sup></b>
Casemiro De Abreu, 900	Rio Branco	302.818,76	C03	2	412	734,48
Pedro Weingartner, 29	Rio Branco	197.336,62	A3D	0	264	746,78
Fonte da Saúde, 47	Rio Branco	290.243,88	A3D	2	282	1.028,65
Santa Cecília, 2001	Rio Branco	295.264,01	A3D	3	357	826,26
Casemiro de Abreu, 851	Rio Branco	208.999,63	A3D	2	279	749,64
Castro Alves, 1205	Rio Branco	434.999,10	C03	2	470	925,53
Santa Cecília, 2096	Rio Branco	215.284,77	A3D	1	257	837
Fonte da Saúde, 47	Rio Branco	431.709,90	C03	3	510	846,49
Santa Cecília, 2001	Rio Branco	500.238,00	C03		600	833,73
Casemiro de Abreu, 851	Rio Branco	408.778,08	C03	2	488	837,66

Fonte (SALDANHA, 2005)





## Resultados e discussões

Inicialmente, como efeito de resultados práticos, pode-se entender que a regressão logística é como se fosse o análogo de regressão linear para discussões de problemas de classificação de tipologia dos imóveis, como o caso em estudo de apartamentos tipo e de cobertura de alto padrão. Esse tipo de problema surge quando se busca categorizar alguma variável por classes, como, por exemplo, padrão construtivo de apartamentos, ou seja, a probabilidade de obter o valor do lançamento de um empreendimento só depende da lei oferta e da procura dos preços práticos dos apartamentos no mercado imobiliário.

### 10.1 Medidas de Tendência Central

Em estatística, uma tendência central (ou, normalmente, uma medida de tendência central) é um valor central ou valor típico para uma distribuição de probabilidade.

As medidas de tendência central são valores típicos para uma distribuição de probabilidade, as mais comuns são a média aritmética, a mediana e moda.

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.





## Mais aplicações

### 11.1 Avaliação Imobiliária Rural

Por similitude, partindo para análise qualitativa do mercado imobiliário rural, com a mensuração das informações dos dados de mercado, busca-se também a validação de resultados dos preços praticados através da modelagem de regressão, a partir do uso de variáveis categóricas, considerando a elaboração de modelo de regressão linear representativo do valor do hectare praticados no mercado imobiliário de áreas rurais localizadas no pampa gaúcho, mais precisamente nos Municípios de Bagé e Dom Pedrito, Estado do Rio Grande do Sul, sendo a variável dependente, preço do hectare expressa por uma combinação linear das diversas variáveis independentes, configuradas pelas características mercadológicas das propriedades mensuradas em escala original, representada pela aplicação das variáveis *dummy*, em observação as informações relevantes a serem incluídas numa modelagem estatística.

Sendo a seguinte amostragem de mercado de terras rurais no início do século XXI, temos:

**Imóvel Avaliado:**

Denominação do Imóvel: Estância Santa Ana

Município de Dom Pedrito-RS

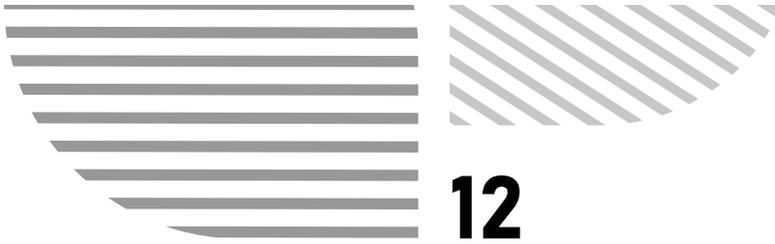
Localidade Serrilhada, 5º Subdistrito Fontouras

Área Avaliada: 1.383,32 hectares

Exploração predominante: pecuária de corte e, em menor proporção, orizicultura

Valor de Avaliação do Imóvel Rural: \$ 1.432.663,00





## Reflexão

Este trabalho permitiu concluir que existe relação significativa de atributos formadores do valor dos imóveis com as variáveis praticadas no mercado imobiliário de Porto Alegre.

Na avaliação de imóveis e bens em geral, dentre as técnicas avançadas, a aplicação de critérios econométricos na valoração imobiliária é por excelência a técnica alternativa avançada para análise da formação dos preços praticados no mercado imobiliário.

Então, nesse sentido, se busca sistematizar com a regressão 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,