SILVIA NEIDE BRAULIO

PROPOSTA DE UMA METODOLOGIA PARA A AVALIAÇÃO DE IMÓVEIS URBANOS BASEADO EM MÉTODOS ESTATÍSTICOS MULTIVARIADOS

CURITIBA

2005

SILVIA NEIDE BRAULIO

PROPOSTA DE UMA METODOLOGIA PARA AVALIAÇÃO DE IMÓVEIS URBANOS BASEADO EM MÉTODOS ESTATÍSTICOS MULTIVARIADOS

Dissertação de Mestrado apresentada como requisito parcial à obtenção do grau de Mestre em Ciências no Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Área de Concentração em Programação Matemática dos setores de Ciências Exatas e de Tecnologia da Universidade Federal do Paraná.

Orientação: Profº. Dr. Anselmo Chaves Neto.

CURITIBA

2005

Não se deve ir atrás de objetivos fáceis.
É preciso buscar o que só pode ser
alcançado por meio dos maiores esforços.

Albert Einstein

A Deus, Nossa Senhora, à minha mãe
Maria Lúcia e à minha irmã Ana Marilsa.

AGRADECIMENTOS

A Deus por ter me dado o dom da vida e a Nossa Senhora, que estiveram sempre ao meu lado, e muitas vezes à frente para derrubar as barreiras e tornar o caminho mais fácil e ainda, por me oportunizar conhecer, conviver e aprender com pessoas muitos especiais que este programa me proporcionou.

Ao meu amigo André Luís, que de forma ímpar esteve ao meu lado, me apoiando, me incentivando e sempre pronto para me ouvir. André, muito obrigada.

Aos corretores de imóveis das imobiliárias, pela gentileza com que forneceram as informações que sem eles esta pesquisa não se realizaria.

Aos professores, que estiveram dispostos em sanar dúvidas que surgiram, e ensinar com muita clareza e objetividade.

A todos os meus colegas de curso pelo apoio e amizade e em especial aos meus amigos: Douglas, pela sua ajuda essencial, Rogério que sempre me incentivou, Amauri e Flávia pela sua prestabilidade.

Ao meu orientador, professor Anselmo, pela compreensão, sempre me apoiando no desenvolvimento do trabalho e oferecendo todas as contribuições necessárias para a realização desta pesquisa.

A todos que acreditaram e torceram por mim e que direta ou indiretamente contribuíram para a que realização desta fosse possível.

E finalmente e especialmente, à minha família que se privou de minha presença, me apoiou e compreendeu, sempre incentivando o meu crescimento. Em especial à minha mãe Maria Lúcia, á minha irmã Ana Marilsa e meu irmão Arnaldo, Mãe, Fia e Arnaldo, muito obrigada.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

RESUMO

O presente trabalho tem por finalidade apresentar um procedimento de construção de um modelo estatístico para avaliação de imóveis em função de suas características, utilizando técnicas de Análise Multivariada em apoio às técnicas tradicionais de estatística. Foi aplicada a Análise de Agrupamento (*Clusters Analysis*) nos dados de imóveis para obtenção de classes homogêneas. Com a técnica de Componentes Principais reduziu-se e transformou-se as variáveis (características dos imóveis), visando facilitar a análise e interpretação. Essas novas variáveis foram utilizadas para a determinação dos modelos de Regressão Linear Múltipla de cada classe ou grupo homogêneo de itens para cada tipo de imóveis (apartamento, casa e terreno). A proposta de avaliação foi aplicada a um conjunto de dados referentes a 44 apartamentos, 51 casas e 24 terrenos da cidade de Campo Mourão – PR. O modelo de cada classe dos três tipos de imóveis avaliados apresentou um bom ajuste aos dados e uma boa capacidade preditiva, atendendo todas as suposições teóricas de regressão linear múltipla.

ABSTRACT

The present study aims to introduce a building procedure of a statistic model to evaluate real estates based on their characteristics, using Multivariate Analysis techniques added to the traditional statistic ones. The Cluster Analysis was used on the real estates data in order to obtain homogeneous classes. By using the Main Components technique, the variables (he characteristics of the real estates) were reduced and changed with the purpose of easing the interpretation analysis. Such new variables were used in order to determine the Multiple Linear Regression models of either each class or homogeneous group of items for each type of real estate (apartment, house and plot). The suggestion of evaluation was applied to a data set referring to 44 apartments, 51 houses and 24 real estates in Campo Mourão city, state of Paraná. Each class model of the three types of real estates evaluated showed a good adjustment to the data, as well as a good predictive capacity, attending all the theoretical suppositions of multiple linear regression.

1 INTRODUÇÃO

1.1 TEMA DO ESTUDO

O mercado imobiliário é uma das áreas mais dinâmicas do setor econômico terciário e as principais dificuldades em uma análise para avaliação de bens provêm das características especiais dos imóveis, que são muito heterogêneas.

Em muitos mercados imobiliários a base de cálculo das estimativas dos valores dos imóveis, quer seja para cobrança de impostos ou para venda, o nível de rigor na apuração é do tipo expedido, ou seja, a avaliação é feita de forma subjetiva, não utilizando qualquer procedimento matemático de suporte para a estimação do valor do imóvel. Muitas dessas estimativas são obtidas com base na experiência ou opinião pessoal.

Então, para que haja objetividade na avaliação, ou ainda, para que a estimativa não seja feita com base no "sentimento" do avaliador, deve-se construir um procedimento estatístico adequado, de forma que possa predizer o valor de um imóvel em função de suas características. Mas, como construí-la? Quais as variáveis (características dos imóveis) que devem ser levadas em consideração?

1.2 OBJETIVOS DO TRABALHO

1.2.1 Objetivo Geral

O que se propõe nesse trabalho é mostrar a aplicação de uma metodologia de forma eminentemente objetiva no processo avaliatório de imóveis urbanos. Desta forma trabalhou-se

com técnicas estatísticas multivariadas, uma vez que todo imóvel em avaliação pode ser considerado como um vetor de características. E considerou-se como estudo de caso o mercado imobiliário da cidade de Campo Mourão nos segmentos de: apartamentos, casas e terrenos. Portanto, o objetivo geral do trabalho é o de construir um modelo estatístico que melhor represente o mercado analisado em um período de tempo e, então, predizer o valor de mercado do imóvel, com máxima precisão.

1.2.2 Objetivos Específicos

Buscando alcançar o objetivo geral foram estabelecidos os seguintes objetivos específicos:

1. Construir classes homogêneas de itens em cada tipo de imóveis.
2. Construir regras de reconhecimento e classificação de itens nos grupos homogêneos definidos, através do método de Fisher.
3. Reduzir o número de variáveis para efeito de análise e interpretação, sem perda significativa de informação.
4. Construir modelos de regressão linear múltipla para a variável valor do item, para cada classe homogênea, usando um número reduzido de variáveis transformadas.

1.3 IMPORTÂNCIA DO TRABALHO

Uma considerável parcela de bens públicos, de pessoas físicas ou jurídicas consiste de bens imóveis. A amplitude desse recurso primordial da sociedade cria a necessidade de informes avaliatórios como suporte para tomadas de decisões referentes ao uso e disposições desses bens (Abunahman, 1998). Principalmente, porque os imóveis se constituem em garantias reais para empréstimos de capital para investimento, feitos por bancos de desenvolvimento ou comerciais.

O homem, desde os primórdios da história, tem procurado critérios para estabelecer, fixar, estimar ou arbitrar preços dos bens que satisfaçam às noções de valor de cada mercado. Sendo essa uma das atividades mais freqüentes e complexas para o homem.

A avaliação de imóveis é utilizada geralmente em negócios, discussões e pendências interpessoais e sociais nas comunidades, quer seja para transações de compra ou de venda, operações de garantias, instalações industriais, transações de locação, decisões jurídicas, operações de seguro, na reavaliação de ativos de empresas, no lançamento de impostos, nas hipotecas imobiliárias, nas divergências que originam ações democráticas, possessoras, nas indenizações, nas desapropriações e servidões, ou seja, em muitas ações oriundas de problemas inerentes aos relacionamentos do homem, em que o valor de um bem é de caráter fundamental. Porém, o que é o valor? Só existe um único valor para um mesmo bem? Esse conceito permite a aplicação do termo a qualquer estimativa, seja ela uma conclusão fundamentada ou, simplesmente, uma opinião pessoal.

O conceito de valor de um imóvel é, de modo geral, intuitivo e subjetivo (opinião pessoal), podendo assim diferir muito entre os avaliadores. O preço, por outro lado, é uma característica objetiva relacionada ao imóvel.

Uma avaliação deve ser objetiva e clara, identificando o bem a ser avaliado e o método a ser utilizado, objetivando minimizar qualquer subjetividade, inerente a todas as atividades humanas.

A avaliação, como uma ciência de mensuração do valor, não é exata, no entanto, pode ser altamente precisa. Basicamente, se trata, do juízo de valor sobre um bem e, quando realizada de forma científica, ou seja, baseada em teorias e métodos adequados, utiliza instrumental tecnológico pertinente. Assim, prima pela objetividade. Os resultados das estimativas realizadas por diferentes avaliadores, ou grupos de avaliadores, deverão ser próximos uns dos outros.

Mas como avaliar um imóvel? Existem vários métodos para se estimar o valor de mercado de um bem imóvel. O melhor caminho é a comparação de dados de transação de imóveis semelhantes, efetuadas no mesmo período em que se necessita fazer a avaliação. Mas nem sempre isto é possível. Logo, deve-se buscar e aplicar outros métodos.

Para que uma avaliação tenha características científicas é necessário o uso da

estatística avançada, com equação de regressão múltipla, definida a partir de amostras da população de imóveis em estudo. A norma de avaliação de imóveis urbanos, NBR/5676, exige essa técnica para os níveis mais altos de confiança: “o tratamento para alcançar a convicção do valor deve ser baseado em processos de inferência estatística que permitam calcular estimativas não tendenciosas e eficientes de valor.”

Na cidade de Campo Mourão a avaliação de imóveis, por imobiliárias, é feita de forma subjetiva e assim, a principal importância desse trabalho é a possibilidade de se construir um modelo estatístico para a melhoria da qualidade das avaliações, usando método científico, determinado através de inferência estatística por regressão múltipla com apoio de técnicas da Análise Multivariada.

1.4 ESTRUTURA DO TRABALHO

Esta pesquisa está estruturada em cinco capítulos, escritos de forma a facilitar o entendimento do leitor. Os objetivos, a justificativa, o problema e a estrutura desta pesquisa estão descritos nesta introdução. O segundo capítulo trata da revisão de literatura necessária, de forma clara, para a compreensão do método de avaliação de imóveis urbanos com técnicas multivariadas. O material e a metodologia desenvolvida para realização do trabalho se encontra no terceiro capítulo, onde é descrita a construção dos modelos e faz-se a análise dos dados. No quarto capítulo são descritos de forma detalhada os resultados e faz-se a análise para cada um deles. E finalmente, as conclusões com base no estudo realizado e sugestão para trabalho futuro são encontradas no quinto capítulo.

2 REVISÃO DE LITERATURA

2.1 INTRODUÇÃO

O objetivo principal desse capítulo é fornecer um texto claro e de forma acessível sobre avaliação de imóveis, contemplando de forma abrangente a Engenharia de Avaliação e dar embasamento teórico sobre a estatística multivariada, necessária para compreensão do desenvolvimento do modelo estatístico proposto.

2.2 ENGENHARIA DE AVALIAÇÃO

2.2.1 Introdução

A Engenharia de Avaliação surgiu no Brasil no início do século XX. Os primeiros trabalhos nesse sentido, das quais se tem conhecimento no Brasil, foram publicados em revistas técnicas de engenharia, em São Paulo entre 1918 e 1919. Em 1923 foram introduzidos novos métodos de avaliação de terrenos, que a partir de 1929 começaram a ter uso sistematizado. (Fiker, 1997, p17).

A introdução de métodos científicos na Engenharia de Avaliação teve um progresso importante na última década no Brasil.

Dantas (1998), define a Engenharia de Avaliação como uma parte da engenharia que reúne um conjunto de conhecimentos da área de engenharia e arquitetura, e de outras áreas (ciências sociais, exatas e da natureza), com o propósito de determinar de uma forma técnica o valor de um bem, de seus direitos, frutos e custos de reprodução, subsidiando tomadas de

decisões a respeito de valores, envolvendo bens de qualquer natureza. A Engenharia de Avaliações é de interesse para imobiliárias, bancos de comerciais e de desenvolvimento, compradores e vendedores de imóveis, entre outros.

A Engenharia de Avaliações pode ser praticada por engenheiros, arquitetos, agrônomos, sendo que cada um dentro de sua habilitação profissional, em conforme as leis do CONFEA (Conselho Federal de Engenharia e Arquitetura).

2.2.2 Normas Técnicas

A ABNT - Associação Brasileira de Normas Técnicas é o Fórum Nacional de Normalização. As Normas Brasileiras, cujos conteúdos são de responsabilidade dos Comitês Brasileiros (ABNT/CB) e dos Organismos de Normalização Setorial (ONS), são elaboradas por Comissões de Estudo (ABNT/CE), formadas por representantes dos setores envolvidos, delas fazendo parte: produtores, consumidores e neutros (universidades, laboratórios e outros).

Em meados de 1950 surgiram as primeiras normas de avaliação de imóveis organizadas por entidades públicas e institutos. Devido a ocorrência de grande quantidade de desapropriações em São Paulo, ocasionado pela expansão da cidade e construção de metrô e outras obras, por volta de 1960, as normas ganharam maior relevância, apesar de que o primeiro anteprojeto de normas da ABNT na Engenharia de Avaliação data de 1957. Em 1977, estudos feitos por comissões de profissionais dedicados a perícias e avaliações judiciais, em essência, deram origem a primeira Norma Brasileira para Avaliação de Imóveis Urbanos, a NB-502/77 da ABNT (Dantas, 1998).

No decorrer do tempo, a ABNT produziu outras normas para avaliações, com tipologia: Imóveis Rurais, Unidades Padronizadas, Máquinas, Equipamentos, Complexos Industriais e Glebas Urbanizáveis.

Revista em 1989, a Norma Brasileira para Avaliação de Imóveis Urbanos foi registrada no INMETRO como NBR 5676. Os níveis de precisão foram transformados em níveis de rigor. Segue-se a ela a Norma para Avaliação de Servidões. Alguns institutos, com

base na NBR 5676, produziram, paralelamente, normas específicas de forma mais detalhada observando as características de cada região.

2.2.3 Avaliação de Imóveis Urbanos

Alguns conceitos pertinentes à técnica de avaliar são apresentados a seguir.

Avaliação

A Norma NB-502/89 (NBR-5676) da ABNT, de avaliação de imóveis urbanos, define avaliação como a determinação técnica do valor de um imóvel ou de um direito sobre o imóvel.

Avaliar significa determinar a valia ou valor de apreciar ou estimar o merecimento e reconhecer a grandeza. Portanto, avaliar pelo preço de mercado significa analisar o mercado com a finalidade de obter um padrão de comparação e classificar o imóvel em questão segundo o padrão de comparação obtido. Segundo Moreira (1994), avaliar é a arte de estimar valores apropriados específicos, em que o conhecimento técnico e o bom-senso são condições fundamentais.

Abunahman (1998), define avaliação como sendo uma aferição de vários fatores econômicos definidos em relação a propriedades descritas com data prevista, tendo como base a análise de dados relevantes.

Valor

O conceito de valor de um bem, de modo geral, é intuitivo e subjetivo, podendo variar muito entre os participantes de um mercado. Quer seja vendedor ou comprador. O preço, no entanto, é uma característica objetiva relacionada ao bem, que representa a quantidade de dinheiro pago pelo bem.

Desde os primórdios da história, o homem tem buscado critérios para estabelecer

preços dos bens que satisfaçam as noções de valor de cada participante envolvido na transação, de maneira a se efetivarem trocas, sejam elas diretas (escambo), ou indiretas (usando um elemento comparativo, como a moeda).

Diversas medidas de valor podem ser associadas a um bem, dentre elas o custo de produção, ao qual são agregados outros custos, formando o preço, e o valor de mercado, não havendo necessariamente uma relação matemática entre eles. É natural que os produtores esperem uma remuneração pelos seus produtos, na maioria das vezes constituída por uma margem sobre os custos incidentes na sua produção, estocagem e comercialização. No entanto, em mercados que se aproximam daquele de concorrência perfeita, os preços são estabelecidos pela lei da oferta e demanda, independentemente dos custos de produção, estocagem e comercialização. Portanto, no mercado considerado, o valor do bem poderá não apresentar nenhuma relação com os custos citados (podendo mesmo ser inferior). Quando o mercado permanece estável por um tempo suficientemente longo, a oferta e a demanda acabam determinando o preço e a quantidade negociada. Pode-se dizer que haverá disposição para fornecer determinadas quantidades a determinados preços, e disposição para comprar determinadas quantidades a determinados preços. Toda a subjetividade que leva os participantes do mercado a tentar maximizar sua satisfação acaba materializando-se em quantidades vendidas e/ou ofertadas e seus respectivos preços. São essas quantidades e preços que constituem os dados sobre os quais o avaliador irá tirar conclusões. Portanto, o preço estabelecido pelo mercado é considerado uma representação justa do valor do bem em questão.

Considerando os mercados onde são efetuadas trocas indiretas, os preços (valores) são expressos em moeda corrente, podendo ou não ser transformados em outras moedas. As avaliações pelo valor de mercado são instantâneas, ou seja, são válidas, apenas, por um intervalo curto de tempo.

O conceito de valor, valor de mercado e preço apesar de terem várias definições e interpretações, sujeito e suscetível a mudanças filosóficas, é importante no relacionamento humano e social determinar alguns critérios para que se exerça um caráter de justiça em sua aplicação prática. Então, um trabalho de avaliação imobiliária constitui-se de uma série de operações até que se chega em uma definição de valor. Dentre os diversos conceitos de valor, a Norma NB-502/89 (NBR-5676) da ABNT, de Avaliação de Imóveis Urbanos, define valor como sendo aquele fornecido para um bem em um dado instante, único, não importando qual

a finalidade da avaliação. Esse valor corresponde ao valor real que se definiria em um mercado de concorrência perfeita caracterizado pelas seguintes premissas:

a) homogeneidade dos bens levados ao mercado;
b) número elevado de compradores e vendedores, de tal forma que o mercado não possa ser alterado por eles;
c) não haja influências externas;
d) conhecimento pleno e absoluto entre os participantes sobre o bem, o mercado e as tendências deste;
e) os participantes oferecendo liquidez com plena liberdade de entrada e saída do mercado.

Quando a necessidade de estimar o valor ocorre a nível particular, o problema se reduz a que as partes interessadas (vendedora, a que oferta o bem e a compradora) estejam de acordo com a quantidade necessária (expressa em unidades monetárias) em um dado instante (Molina, 1999). No entanto, quando se necessita estimar um valor além do nível particular, de uma maneira mais ampla, isto é, quando os interesses, que podem ser de ordem privada ou pública, são extensivos a outras pessoas além daquelas que estão diretamente envolvidas, procura-se uma perspectiva técnica, surgindo, então, a "Ciência da Avaliação", ou seja, a Engenharia de Avaliações, que conclui sobre o valor de um bem de forma fundamentada.

Segundo Barbosa Filho (1998) o valor de um bem antes de tudo é um fenômeno social, e pode ser associado a um vetor composto por um conjunto de variáveis que abrange todas as suas características físicas, do seu entorno, da sua utilidade e dos fatores subjetivos que a própria coletividade cria no contexto em que está situado a cada instante. Para Moreira (1994), a palavra valor é usada correntemente em muitos sentidos diferentes. Quando aplicada à propriedade, a palavra valor traz consigo um sentido de desejo de posse, domínio ou troca de propriedade, medida em termos de uma unidade monetária.

Fiker (1997) afirma que valor é a relação entre a intensidade das necessidades econômicas do homem e a quantidade de bens disponíveis para satisfazê-las, sendo determinado dependendo da oferta e da demanda do bem.

Recentemente o valor de mercado de um imóvel é referido como sendo o preço fixado pelo vendedor desejoso e um comprador igualmente desejoso, mas sem que ambos

sejam forçados, sem estar sujeitos a pressões anormais e ambos tenham pleno conhecimento das condições de compra e venda e da utilidade do imóvel. No entanto, o mercado imobiliário não é, por natureza, de concorrência perfeita. Dessa forma, o que se pode conseguir é a estimação do preço médio de mercado, através de uma amostragem de preços que carregam todas as imperfeições do mercado.

2.2.4 Classificação dos Imóveis Urbanos

De acordo com a Norma NB-502/89 (NBR-5676) da ABNT, os imóveis são classificados em:

a) *Quanto ao uso*

O imóvel urbano pode ser: residencial, comercial, industrial, institucional e misto.

b) *Quanto ao tipo do imóvel*

O imóvel urbano pode ser: terreno (lote ou gleba), apartamento, casa, escritório (sala ou andar corrido), loja, galpão, vaga de garagem, misto, hotéis, hospitais, cinemas e teatros, clubes e recreativos.

c) *Quanto ao agrupamento*

Os imóveis urbanos se agrupam da seguinte forma: loteamento, condomínio de casas, prédio de apartamentos, conjunto habitacional (casas, prédios ou mistos), conjunto de salas comerciais, prédio comercial, conjunto de prédios comerciais, conjunto de unidades comerciais, *shopping-centers* e complexo industrial.

Este trabalho utilizou dados correspondentes apenas aos imóveis dos tipos: apartamentos, casas residenciais e terrenos.

2.2.5 O Mercado

O mercado pode ser definido como o local onde são efetuadas as transações comerciais envolvendo troca de bens, tangíveis ou intangíveis, ou direitos sobre os mesmos. Aqui o termo mercado refere-se àquele de concorrência perfeita e contendo, em geral, as características dos bens. Todos os que participam do mercado o fazem voluntariamente e têm conhecimento pleno das condições vigentes; nenhum participante, sozinho, é capaz de alterar as condições estabelecidas; cada transação é feita de maneira independente das demais; o número de ofertas e/ou transações é suficientemente grande, de maneira que a retirada de uma amostra não afeta o mercado.

2.2.5.1 Mercado Imobiliário

O mercado imobiliário é a instância de determinação dos preços de imóveis urbanos que, como quaisquer outras mercadorias, passa pelo crivo da oferta e da demanda (Moscovitch, 1997).

Ele é formado por três segmentos: o dos imóveis a serem vendidos, o das partes que desejam vendê-los (vendedores) e o das partes interessadas em adquiri-los (compradores). Pode ser subdivido em várias especialidades, entre outras de terrenos, de apartamentos, de casas, que foram especificidades analisadas neste trabalho (Dantas, 1998).

O mercado imobiliário tem comportamento muito diferente dos mercados de outros bens devido às características especiais dos imóveis. Existem inúmeras fontes de divergências e de desigualdade entre os imóveis. Por sua localização fixa, qualquer alteração no ambiente provoca modificações no valor do imóvel. Como as influências não são homogêneas, as variações provocadas são claramente distintas, causando progressivamente as diferenças.

Por outro lado, como todo bem econômico, a escassez relativa à lei da oferta e procura define o preço. Os preços dos imóveis estão sujeitos às influências de governos e das economias local, regional, nacional e global. E, por sua importância e significado social, as leis propiciam tratamentos especiais.

Quanto ao mercado em que esses bens são transacionados, os principais fenômenos identificados são: o dinamismo da atividade imobiliária e o processo de estruturação interna das áreas urbanas. Existem, também, influências externas, que alteram continuamente os valores e usos do solo. O estudo desses fatores constitui o processo de formação de valores, ou seja, como os valores dos imóveis são compostos.

A situação simultânea e não coordenada de empreendedores, intermediários, poder público e população, em geral, provoca transformações nas condições do mercado e nos valores praticados. Devido à natureza dos bens econômicos, todas as modificações que provocam alterações de disponibilidades são refletidas em alterações de valor. Em termos nacionais, a oferta de crédito, a inflação, a condução da economia, as políticas fiscais, o crescimento demográfico e a confiança no governo são importantes na flutuação de preços.

As variações nas condições do mercado são absorvidas, internalizadas, pelos imóveis, através de um aumento ou diminuição de seus valores, que variam no tempo e no espaço e que, em última análise, são resultados da oferta e demanda pelo bem. Em um dado momento, há um "equilíbrio instantâneo", do qual resulta um valor de mercado. Mudanças na oferta ou na demanda provocarão novo equilíbrio, em outros níveis de preço.

Finalmente, é interessante verificar que, por existirem inúmeras influências, uma parte das variações dos valores imobiliários pode ser considerada aleatória, ou seja, pode-se pensar no preço final como baseado em um valor mais provável que é aumentado ou diminuído por uma parcela imprevisível, de acordo com as influências pontuais do acaso. Deste modo o valor de um imóvel segue o modelo estatístico,

$$Y = \mu + \varepsilon \qquad (2.1)$$

onde: Y é o valor negociado;

μ é o valor mais provável, ou seja, $E(Y) = \mu$;

ε é a perturbação estocástica.

Dantas (1998), informa que a relação quantitativa de três componentes (imóvel, vendedor e comprador) formadores do mercado é uma relação determinante na formação dos preços, sendo a situação ideal aquela onde a oferta e a procura é equilibrada, ou seja, um mercado de concorrência perfeita. Por outro lado, pode ocorrer o mercado de concorrência imperfeita, onde ocorrem casos de monopólio (raro). Neste caso o mercado é comandado por

um único vendedor. Já no caso do oligopólio (mais comum) ocorre a situação de mercado em que a oferta é controlada por um pequeno número de vendedores, e em que a competição tem por base, não as variações de preços, mas sim a propaganda e as diferenças de qualidade. Isto faz os preços serem "puxados" para cima. No caso de monopsônio (raro) onde há apenas um comprador e oligopsônio (mais comum) alguns compradores, os preços têm uma tendência para baixo.

Em muitas análises teóricas, se supõe a concorrência perfeita. Aceitar a perfeição de um mercado significa, simplificadamente, admitir que os bens podem ser considerados idênticos, que a entrada no mercado é livre, que as pessoas têm informação perfeita, decidem livre e prudentemente, sem pressões de qualquer ordem, e que ações individuais não afetam os preços.

Nessas condições, o valor do bem é igual ao preço que ele atinge no mercado, e é rigidamente proporcional à qualidade adquirida.

Tal situação claramente não é do mercado imobiliário. Existem falta de informações, que é um problema. A heterogeneidade dos imóveis e de suas localizações dificulta, assim, a comparação. Há diversos sub-mercados para cada tipo de imóvel. Não há liberdade para negociar, mas ao contrário, as partes sofrem diversas pressões e existem muitos fatores não monetários (psicológicos e culturais) que afetam a avaliação subjetiva da qualidade do bem. O custo elevado dificulta ou impede a participação de expressiva parcela da população, relegada à locação, enquanto outra parte depende de financiamentos, geralmente indisponíveis.

Diante dessas afirmativas, pode-se concluir que o mercado imobiliário é de concorrência imperfeita, acarretando diversas implicações na análise. O preço não coincide com o valor, necessariamente, existindo uma faixa de preços razoáveis, dentro da qual está o valor, de mercado (valor esperado) para o bem. E assim, o mercado imobiliário se torna um dos setores mais complexos da economia.

2.2.6 Característica dos Imóveis

De acordo com González (2000), o mercado imobiliário, por ter um comportamento diferenciado dos mercados de outros bens, devido às características especiais dos imóveis e

do mercado imobiliário, possui inúmeras fontes de divergências e desigualdades entre os imóveis, impossibilitando, assim, a comparação direta. Entre os fatores que diferenciam os imóveis, em si, pode-se citar a grande vida útil, a fixação espacial, a singularidade, o elevado prazo de maturação e o alto custo das unidades. A combinação desses elementos permite explicar grande parcela das diferenças de valores entre os imóveis, em um dado momento.

Os preços dos imóveis podem ser compreendidos como a soma dos produtos das quantidades de cada um desses serviços pelos seus preços implícitos. Inicialmente, não são conhecidas as importâncias relativas (participação no preço) das características contidas no imóvel. É conhecido apenas o preço integral do imóvel. Os preços implícitos são os preços relacionados, indiretamente, com cada um dos atributos dos imóveis, tais como área, idade e localização.

Por fim, é importante verificar que, por existirem inúmeras influências, uma parte das variações dos valores imobiliários pode ser considerada aleatória, ou seja, pode-se pensar no preço final como baseado em um "valor mais provável" que é sujeito a variações, de acordo com as influências pontuais do caso.

2.2.7 Métodos de Avaliação

Pode-se definir as metodologias de avaliação como sendo as várias e diferentes vias decorridas com o objetivo de atribuir valor a um imóvel. Cada via utilizada é caracterizada como um método de avaliação diferente. No entanto, independentemente da metodologia aplicada, essa deverá apoiar-se em pesquisa de mercado e considerar os preços comercializados e/ou ofertados, bem como outros elementos e atributos que influenciam o valor (NBR-5676/90). A escolha da metodologia mais apropriada para uma dada avaliação depende das condições atuais do mercado, do tipo de serviço a que se presta e da precisão que se deseja.

Segundo Montenegro Duarte, 1999, uma avaliação deve ser objetiva e clara, identificando o bem a ser avaliado e o método a ser utilizado, objetivando minimizar qualquer subjetividade, inerente a todas as atividades humanas.

A avaliação, como uma ciência de mensuração do valor, não é exata, no entanto, pode ser altamente precisa. Basicamente, se trata, do juízo de valor sobre um bem e, quando realizada de forma científica, ou seja, baseada em teorias e métodos adequados, utiliza instrumental tecnológico pertinente. Assim, prima pela objetividade. Os resultados das estimativas realizadas por diferentes avaliadores, ou grupos de avaliadores, deverão ser próximos uns dos outros (Montenegro Duarte, 1999).

A ABNT (NBR-5676/90) divide os métodos de avaliação em dois grandes grupos: métodos diretos e métodos indiretos.

a) Métodos Diretos

Considera-se um método como sendo direto quando o valor do resultado da avaliação independe de outros (Dantas, 1998, p.15). Os métodos diretos subdividem-se em método comparativo de dados de mercado e método comparativo de custo de reprodução de benfeitorias.

Método comparativo de dados: aquele que define o valor de comparação com dados de mercado assemelhados quanto a características intrínsecas e extrínsecas. As características e os atributos dos dados pesquisados que exercem influência na formação dos preços conseqüente, no valor, devem ser ponderados por homogeneização ou por inferência estatística, respeitados os níveis de rigor definidos nessa norma. É condição fundamental para a aplicação desse método a existência de um conjunto de dados que possa a ser tomado, estatisticamente, como amostra de mercado imobiliário.

Esse método é o mais indicado para trabalhos de avaliação.

Método comparativo de custo de reprodução de benfeitorias: aquele que apropria o valor das benfeitorias, através da reprodução dos custos de seus componentes. A composição dos custos é feita com base em orçamento detalhado ou sumário em função do rigor do trabalho avaliatório. Devem ser justificados e quantificados os efeitos do desgaste físico e ou do obsoletismo funcional das benfeitorias. A composição dos custos é feita baseada em orçamento detalhado ou sumário, dependendo do rigor do trabalho.

A utilização dos métodos diretos tem preferência e sempre que existirem dados de mercado suficientes para utilização do método comparativo ele deve ser escolhido (Dantas, 1998, p.15).

b) Métodos Indiretos

O método é considerado indireto quando necessita de resultados de algum método direto. Os métodos indiretos são os seguintes:

O método da renda: avalia o valor do imóvel ou de suas partes componentes em função de um rendimento já existente ou previsto pelo bem no mercado, ou seja, o valor econômico do bem (Ayres, 1996, p.23; NBR-5676/90).

Método involutivo: o valor do terreno é estimado por estudos da viabilidade técnica-econômica do seu aproveitamento, considerando como aproveitamento eficiente à realização de um empreendimento imobiliário hipotético compatível com as características do imóvel e com as condições do mercado (Moreira Filho et al., 1993, p.5; NBR-5676/90).

Método residual: obtém-se o valor do terreno a partir da diferença entre o valor total do imóvel e o valor das benfeitorias, levando-se em conta o fator de comercialização (Fiker, 1997, p.27; NBR-5676/90).

Quando se analisam os vários métodos de avaliação, descritos anteriormente, pode-se observar que de uma forma, ou de outra, todos são comparativos. Sendo assim, tem-se que no método comparativo comparam-se bens semelhantes; no método de custo, comparam-se os próprios custos no mercado; nos métodos da renda e involutivo compara-se à possibilidade de renda do bem; e no método residual, compara-se o nível de comercialização do mercado (Dantas, 1998).

No entanto, na avaliação de imóveis, o método mais utilizado e recomendado é o método comparativo de dados de mercado, já que esse método permite que a estimativa seja obtida considerando-se as diferentes tendências do mercado imobiliário. Esse método estima o valor 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 isso uma amostragem estatística de dados do mercado imobiliário. Na prática, de um modo geral, a semelhança entre o imóvel avaliado e os componentes da amostra é imperfeita e incompleta devido à falta de algum atributo que possa ter influenciado no valor ou, ainda, por apresentá-lo de forma incompleta. Assim, 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-5676/89. Com o uso de técnicas estatísticas obtém-se uma avaliação

isenta de subjetividade e de grande confiabilidade (Moreira Filho et al., 1993; González e Formoso, 2000).

Tradicionalmente, usavam-se tabelas na comparação de vendas para justificar o estado real e/ou estimar valores aproximados. Mais recentemente, os modelos de preços hedônicos (regressão linear múltipla) têm sido utilizados para completar o método de comparação de vendas. No entanto, ambos os métodos tem sofrido críticas tanto das comunidades acadêmica quanto das comunidades profissional. O primeiro método é, freqüentemente, criticado por utilizar julgamentos subjetivos para determinar os ajustes necessários e também, por ser impreciso, tornando, assim, difícil para o avaliador obter dados confiáveis. A regressão linear múltipla tem produzido, freqüentemente, sérios problemas para a avaliação de imóvel devido a problemas de multicolinearidade nas variáveis explicativas e também de inclusão de *outlier* na amostra. Além disso, a colinearidade dentro dos dados pode tornar a regressão linear múltipla um modelo inadequado para um mercado que requer respostas rápidas e de alta precisão. A regressão é um método padrão aceitável para a avaliação de imóveis (Worzala et al., 1995), mas deve ser feita por profissional capacitado.

2.2.8 Nível de Precisão da Avaliação

Os níveis de precisão que caracterizam uma determinada avaliação são normatizados pela NBR-5676/90. De acordo com NBR-5676, em seu item 7, o nível de rigor almejado numa dada avaliação relaciona-se diretamente com as informações extraídas do mercado, ou seja, a precisão do mercado será determinada por esse nível que será, por sua vez, tanto maior quanto menor for a subjetividade presente na avaliação. O rigor de uma avaliação está condicionado a abrangência da pesquisa, à confiabilidade e adequação dos dados coletados, à qualidade do processo avaliatório e ao menor grau de subjetividade empregado pelo avaliador. Assim, os trabalhos avaliatórios podem, de acordo com a norma, ser classificados como de nível de rigor expedito, normal, rigoroso e rigoroso especial.

No caso de rigor expedito, o valor é obtido sem a utilização de qualquer instrumento matemático. Dessa forma, a ausência de método científico determina que o valor seja atribuído através de escolha arbitrária, não caracterizando o aspecto técnico da avaliação.

Assim, basta que o avaliador tenha bom nível de conhecimento de mercado. Esse tipo de avaliação é muito freqüente entre os corretores e não é o objetivo desse trabalho.

Para a avaliação em rigor normal utiliza-se métodos estatísticos e existem exigências com relação à coleta e tratamento dos dados.

Nas avaliações rigorosas, o trabalho deverá apresentar, através de metodologia adequada, isenção de subjetividade. O tratamento dos dados deve se basear em processos estatísticos que permitam calcular estimativas não tendenciosas do valor. O valor final da avaliação, resultado do tratamento estatístico adotado, deve estar contido em um intervalo de confiança fechado e com um nível de confiança máximo de 80%, desde que as hipóteses nulas sejam testadas ao nível de significância máximo de 5%.

A avaliação rigorosa especial caracteriza-se pelo encontro de um modelo estatístico o mais abrangente possível, ou seja, que incorpore o maior número de características que contribuem para a formação do valor.

A função estimada da formação de valor deve ser eficiente e não tendenciosa. Portanto, a hipótese nula da equação de regressão deve ser rejeitada ao nível de significância máximo de 1% (ANOVA). Já as hipóteses nulas sobre os parâmetros do modelo de regressão ao nível de significância máximo de 10% para o teste unicaudal (teste "*t*") ou 5% em cada ramo do teste bicaudal. Devem ser analisadas as seguintes condições básicas referentes aos resíduos do modelo ajustado aos dados: Gaussianidade, homogeneidade da variância e independência. Desta forma os resíduos devem ser Gaussianos, independentes e identicamente distribuídos, ou seja, $\varepsilon_i \sim N(0, \sigma^2)$.

2.3 ANÁLISE MULTIVARIADA

2.3.1 Introdução

Em qualquer decisão tomada, sempre deve ser levado em conta o grande número de fatores envolvidos. Obviamente nem todos os fatores têm a mesma importância e, portanto,

devem ser escolhidos os principais. Muitas vezes, quando se toma uma decisão usando a intuição, não são identificados de maneira sistemática estes fatores. Ou seja, não são definidas as variáveis que afetam a decisão.

Quando se analisa o mundo à volta de todos, verifica-se que todos os acontecimentos sejam eles culturais ou naturais, envolvem um grande número de variáveis. As diversas ciências têm a pretensão de conhecer a realidade e de interpretar os acontecimentos (ciências humanas) e os fenômenos (ciências naturais), baseados no conhecimento das variáveis intervenientes consideradas importantes nestes eventos.

Estabelecer relações, encontrar ou propor leis explicativas é o papel próprio da ciência. Para isso é necessário controlar, manipular, medir as variáveis que são consideradas relevantes ao entendimento do fenômeno analisado. Muitas são as dificuldades em traduzir as informações obtidas em conhecimento, mas a maior delas é de natureza epistemológica: a ciência, que tenta representar a realidade através de modelos e teorias dos diversos ramos do conhecimento. Outra dificuldade é a aspiração de universalidade das explicações científicas, implicando, desta forma, ao condicionamento da pesquisa a uma "padronização" metodológica. Um aspecto essencial desta padronização é a avaliação estatística das informações.

Cientificamente, na maior parte das vezes, as análises são feitas considerando-se os dados das variáveis isoladamente. E, a partir desta análise univariada faz-se inferências sobre a realidade. Esta simplificação tem vantagens e desvantagens. Quando um fenômeno depende de muitas variáveis, geralmente, este tipo de análise falha, pois não basta conhecer informações estatísticas isoladas, mas é necessário também conhecer a totalidade destas informações fornecida pelo conjunto das variáveis. As relações existentes entre as variáveis não são percebidas e assim efeitos antagônicos ou sinergéticos entre as variáveis complicam a interpretação do fenômeno, quando se faz a análise a partir das variáveis consideradas. Porém, no caso restrito de variáveis independentes entre si é possível, com razoável segurança, interpretar um fenômeno complexo usando as informações estatísticas de poucas variáveis.

O desenvolvimento científico ampliou em muito a capacidade de obter informações de acontecimentos e fenômenos que estão sendo analisados. Uma massa de dados, grande ou pequena, deve ser processada antes de ser transformada em conhecimento, ou seja, em informações. Portanto, cada vez mais há necessidade de ferramentas estatísticas que

apresentem uma visão global do fenômeno e que melhore as informações obtidas com uma abordagem univariada. A denominação "Análise Multivariada" corresponde um conjunto de técnicas que utiliza simultaneamente todas as variáveis que caracterizam um item na análise estatística. Assim, ao invés de trabalhar com uma variável explicativa X ela considera o vetor $\underline{X}$ cujas componentes são variáveis explicativas.

A Análise Multivariada se preocupa com métodos estatísticos para descrever e analisar dados de muitas variáveis simultaneamente. A necessidade de entender o relacionamento entre as diversas variáveis aleatórias faz da Análise Multivariada uma metodologia com muito potencial de aplicação, principalmente no momento atual em que a tecnologia é veloz e barata.

Quando se aborda as técnicas multivariadas, fala-se em técnicas que não abordam unicamente uma dimensão da análise de dados, e sim, uma gama de cruzamento entre variáveis dependentes e independentes, ou ainda um cruzamento de dados envolvendo informações de várias questões de ordem dependente, oferecendo desta forma ao pesquisador uma segunda dimensão, mais rica que normalmente em abordagem univariada.

Bouroche e Saporta (1982) comentam que a estatística univariada clássica fixou-se no estudo de uma única característica (ou variável) medida para um conjunto pequeno de indivíduos. Desenvolveu as noções de estimativa e de testes fundamentados em hipóteses muito restritivas. Entretanto, na prática, os indivíduos observados são freqüentemente caracterizados por um grande número de características (ou variáveis). Os métodos de análise de dados permitem um estudo global dessas variáveis, pondo em evidência ligações, semelhanças ou diferenças. Por isso, mergulham-se indivíduos e variáveis em espaços geométricos, fazendo-se a máxima economia de hipóteses, e transformam-se os dados para visualizá-los num plano ou classificá-los em grupos homogêneos, e isso perdendo o mínimo de informação.

Segundo Cuadras (1981) a Análise Multivariada é a parte da estatística que estuda, interpreta e elabora o material estatístico sobre a base de um conjunto de $n > 1$ variáveis (quantitativa e/ou qualitativa). Os dados onde cabe uma Análise Multivariada são, portanto, de caráter multidimensional.

De acordo com Johnson e Wichern (1998), a Análise Multivariada pode ser usada para:

◊ Redução ou simplificação de dados.

◊ Distribuição e agrupamentos.

◊ Investigação da dependência entre variáveis.

◊ Predição.

◊ Teste de hipótese, e muitas outras.

Alguns dos objetivos mais importantes dos métodos multivariados, de acordo com Pla (1986), são:

a) A simplificação da estrutura dos dados, encontrando uma maneira adequada de representar o universo em estudo. Isto pode ser obtido mediante a transformação (combinação linear ou não linear) de um conjunto de variáveis interdependentes em outro conjunto independente e/ou em um conjunto de menor dimensão.

b) Classificação. Este tipo de análise permite situar as observações dentro de grupos ou, então, concluir que os indivíduos estão dispersos aleatoriamente no multiespaço. Também é possível alocar novos itens em grupos identificados.

c) Análise da interdependência. O objetivo é examinar a interdependência entre as variáveis, a qual abarca desde a independência total até a colinearidade quando uma delas é combinação linear de outras ou, em termos mais gerais, é uma função $f(x)$ qualquer das outras.

Segundo Johnson e Wichern (1998), em problemas que envolvem p variáveis ($p > 1$), tomando-se n observações de cada vetor aleatório $\underline{X}$ tem-se que as medidas observadas x_{ij}, com $i = 1, 2, \ldots, n$ e $j = 1, 2, \ldots, p$, podem ser agrupadas em uma matriz de dados genérica de ordem $n \times p$, ${}_nX_p$.

$$ {}_nX_p = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \qquad (2.2) $$

A representação da matriz de dados correspondente a n observações do vetor $\underline{X}'$ =

$[X_1, X_2, ..., X_p]$ de dimensão p, composto por p varáveis aleatórias, pode ser ${}_nX_p = (X_{ij})$. Por outro lado, essa matriz corresponde a uma amostra aleatória de tamanho n do vetor p-dimensional $\underline{X}$, ou seja, $[\underline{x}_1, \underline{x}_2, \underline{x}_3, ..., \underline{x}_n]$.

Em quase todas as áreas de pesquisas e aplicação de técnicas estatísticas, não apenas uma característica do objeto estudado é observada. Na maioria das vezes, várias características (variáveis) são estudadas. Essas variáveis, em geral, não são independentes e por isso devem ser analisadas conjuntamente. Análise Multivariada é a área da Estatística que trata desse tipo de análise e visa trabalhar, conjuntamente, com mais do que uma variável. As técnicas multivariadas mais utilizadas são aquelas relacionadas ao estudo da estrutura de covariância do vetor observado, ao estudo do agrupamento de itens e ao estudo do reconhecimento de padrões e classificação. Especificamente pode-se citar: Análise de Componentes Principais, Análise Fatorial, Análise de Correlação Cônica, Análise de Agrupamento (*Cluster Analysis*), Reconhecimento e Classificação de Padrões e, entre estas, Análise Discriminante. Várias são as técnicas que podem ser aplicadas aos dados. Sua utilização depende do tipo de dados que se deseja analisar e dos objetivos do estudo. Nesta pesquisa trabalhou-se com as seguintes técnicas multivariadas: Análise de Agrupamentos, Análise de Componentes Principais e Reconhecimento de Padrões e Classificação.

2.3.2 Estatísticas Descritivas Multivariadas

Geralmente, na Ciência Estatística, trabalha-se com uma parte da população denominada amostra. As informações amostrais contidas nas observações multivariadas, $[\underline{x}_1, \underline{x}_2, \underline{x}_3, ..., \underline{x}_n]$ podem ser resumidas em números sumários conhecidos como estatísticas descritivas. As estatísticas são usadas na inferência sobre os parâmetros, ou seja, na estimação do vetor médio $\underline{\mu}$, da matriz de covariância Σ ou da matriz de correlação ρ, entre outros. De maneira que o vetor médio populacional $\underline{\mu}$ deve ser estimado pelo vetor médio amostral, $\overline{\underline{X}}$, definido pela expressão,

$$\overline{\underline{X}} = \frac{\sum_{i=1}^{n} \underline{x}_i}{n}, \qquad (2.3)$$

onde $\underline{x}_i$ com $i = 1, 2, ..., n$ corresponde às observações amostrais do vetor $\underline{X}$ e n é o tamanho da amostra observada. Evidentemente, como se afirmou, outros parâmetros de uma população multivariada $f(\underline{x})$ podem ser avaliados, tais como a matriz de covariância Σ, definida por:

$$\sum = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix} \tag{2.4}$$

onde tem-se na diagonal principal as variâncias das variáveis aleatórias e fora da diagonal principal as covariâncias entre elas. E, a matriz de correlação ρ, definida por:

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix} \tag{2.5}$$

com as correlações entre as variáveis fora da diagonal principal. Então estes parâmetros, Σ e ρ, são estimados, respectivamente, pela matriz de covariância amostral S e pela matriz de correlação amostral R ou $\hat{\rho}$, cujas expressões são:

$$S = \frac{\sum_{i=1}^{n} (\underline{x}_i - \overline{\underline{x}})(\underline{x}_i - \overline{\underline{x}})'}{n-1} = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{bmatrix} \tag{2.6}$$

sendo s_j^2 a variância amostral da variável aleatória X_j,

$$\boldsymbol{s}_j^2 = \frac{\sum_{i=1}^{n} \left(X_{ij} - \overline{X}_j \right)^2}{n-1} \tag{2.7}$$

e s_{jk} a covariância amostral entre as variáveis aleatórias X_j e X_k, ou seja,

$$s_{jk} = \frac{\sum_{i=1}^{n} \left(X_{ij} - \overline{X}_j \right)\left(X_{i1} - \overline{X}_{1j} \right)}{n-1} \tag{2.8}$$

e

$$R = \hat{\rho} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \tag{2.9}$$

com as correlações amostrais fora da diagonal principal e dadas pelo quociente entre a covariância amostral e o produto dos desvios padrões amostrais, ou seja:

$$r_{jk} = \frac{s_{jk}}{s_j s_k} \text{ para } j \neq k. \tag{2.10}$$

Esses estimadores são muito bons para estimar os parâmetros. Os primeiros são UMVU (Estimador Uniformemente de Mínima Variância) e o último é EMV (Estimador Máxima Verossimilhança).

2.3.3 Métodos Multivariados

Existem vários métodos de Análise Multivariada, com finalidades bem diversas; é necessário saber que tipo de conhecimento se pretende gerar, ou melhor, o que se pretende afirmar a respeito dos dados.

Métodos Multivariados são técnicas estatísticas importantes cujo uso é particularmente difundido nas ciências físicas, sociais e médicas. Em muitos destes estudos científicos há necessidade de entender as relações entre várias variáveis.

A Análise Multivariada também é complexa porque é difícil de identificar técnicas que são projetadas para estudar relações dependentes e interdependentes.

Os métodos estatísticos são escolhidos de acordo com os objetivos da pesquisa. Aplicou-se neste trabalho alguns destes métodos e em razão disto procurou-se detalhá-los, o que é feito nas próximas seções.

2.3.3.1 Análise de Componentes Principais

Para investigar as relações em um conjunto de dados de p variáveis correlacionadas pode ser interessante transformar o conjunto de variáveis originais em um novo conjunto de variáveis não-correlacionadas chamadas Componentes Principais, tendo propriedades especiais em termos de variâncias.

As novas variáveis (componentes Principais), são combinações lineares das variáveis originais e são derivadas em ordem decrescente de importância.

Vale lembrar que a Análise de Componentes Principais não depende da suposição inicial de Gaussianidade e é um método muito útil para auxiliar em Regressão, Análise Fatorial e Análise de Agrupamentos.

O método das Componentes Principais é um dos mais usados para resolver o problema clássico da Análise Fatorial a partir da matriz de correlação amostral R. Uma Análise de Componentes Principais diz respeito a explicar a estrutura da variância e covariância de um vetor através de poucas combinações lineares das variáveis originais. Seu objetivo geral consiste tanto em reduzir os dados como em facilitar a interpretação, pois consiste numa transformação, de eixos, tornando as novas variáveis (combinações lineares) não correlacionadas.

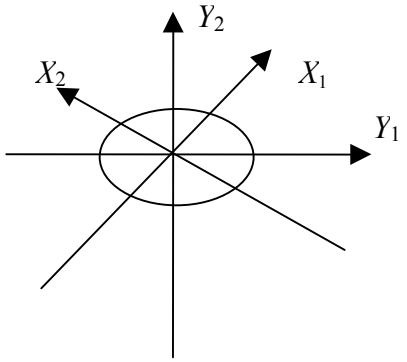
Embora as p componentes sejam necessárias para reproduzir toda a variabilidade presente na estrutura de covariância do vetor $\underline{X}$ de dimensão p, freqüentemente, uma grande parte desta variabilidade poderá ser explicada por um número $m < p$ de Componentes Principais. Neste caso existe praticamente a mesma quantidade de informações nas m Componentes Principais do que nas p variáveis originais. A Análise das Componentes Principais freqüentemente revela relações que não eram previamente consideradas e assim permitem interpretações que não iriam, de outro modo, aparecer (Johnson e Wichern,1998).

a) Componentes Principais Populacionais

Segundo Johnson e Wichern, algebricamente as componentes principais são combinações lineares das p variáveis originais X_1, X_2, ..., X_p que compõem o vetor aleatório $\underline{X}$. Geometricamente, as combinações lineares representam a seleção de um novo sistema de coordenadas, obtido por rotação do sistema original, sendo que os novos eixos representam as

direções com variabilidade máxima. Como exemplo, tem-se a representação da estrutura de componentes principais para $p = 2$:

Figura 2.1: Representação da estrutura de componentes principais



onde:

X_1 e X_2 são eixos originais.

Y_1 e Y_2 são novos eixos (eixos originais rotacionados: centrado na média amostral).

As Componentes Principais são obtidas a partir da matriz de covariância Σ ou da matriz de correlação ρ, que resumem a estrutura de relacionamento das p variáveis originais que compõem o vetor $\underline{X}$. Então, da matriz de covariância $\sum$ ou da matriz de correlação ρ, obtém-se os autovalores $\lambda_1 \geq \lambda_2 \geq ..... \geq \lambda_p$ e os respectivos autovetores $\underline{e}_1, \underline{e}_2, ... , \underline{e}_p$. E, com estes entes algébricos se constrói as combinações lineares que definem as componentes principais, ou seja, $Y_i = \underline{e}_i' \underline{X} \quad i = 1, 2, ... , p.$ (2.11)

As Componentes Principais são combinações lineares, Y_i i = 1, 2, ..., p, não correlacionadas, uma vez que a matriz dos autovetores P, abaixo, é ortogonal,

$$P = \begin{bmatrix} e_{11} & e_{12} & ... & e_{1p} \\ e_{21} & e_{22} & ... & e_{2p} \\ ... & ... & ... & ... \\ e_{p1} & e_{p2} & ... & e_{pp} \end{bmatrix}. \qquad (2.12)$$

A variância da Componente Principal $Y_i = \underline{e}_i' \underline{X} \quad i = 1, 2, ... , p$ é dada por,

$$V(Y_i) = V(\underline{e}_i' \underline{X}) = \underline{e}_i' V(\underline{X}) \underline{e}_i' = \underline{e}_i' \Sigma \underline{e}_i' \qquad (2.13)$$

e a covariância é entre Y_j e Y_k é nula, ou melhor, cov (Y_j, Y_k) = 0.

Logo, pode-se definir:

- A primeira componente principal como a combinação linear $Y_1 = \underline{e}_1' \underline{X}$ que maximiza a variância de Y_1, sob a restrição $\underline{e}_1' \underline{e}_1 = 1$.

- A segunda componente principal como a combinação linear $Y_2 = \underline{e}_2' \underline{X}$ que maximiza $V(\underline{e}_2' \underline{X})$ sujeita a restrição $\underline{e}_2' \underline{e}_2 = 1$ e $\text{cov}(\underline{e}_1' \underline{X} \underline{e}_2' \underline{X}) = 0$.

- A i-ésima componente principal como a combinação linear $Y_i = \underline{e}_i' \underline{X}$ que maximiza $V(\underline{e}_i' \underline{X})$ sujeita a restrição $\underline{e}_i' \underline{e}_i = 1$ e $\text{cov}(\underline{e}_k' \underline{X} \underline{e}_i' \underline{X}) = 0$ para todo $i < k$.

b) Componentes Principais da Amostra

Geralmente os parâmetros da estrutura de covariância, Σ ou ρ, são desconhecidos, então a obtenção das componentes principais é feita a partir de seus estimadores, que são a matriz de covariância amostral S ou a matriz de correlação amostral R. Estas estatísticas são definidas por:

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (\underline{x}_i - \overline{\underline{x}})(\underline{x}_i - \overline{\underline{x}})' \qquad (2.15)$$

$$R = D^{-1} S D^{-1} \qquad (2.16)$$

onde D é a matriz desvio padrão amostral e $\overline{\underline{x}}$ é o vetor médio amostral, dados respectivamente por:

$$D = \begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & s_p \end{pmatrix} \qquad (2.17)$$

$$\overline{\underline{x}} = \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_p \end{bmatrix} \qquad (2.18)$$

Então, obtém-se as estimativas dos elementos da estrutura de covariância do vetor aleatório $\underline{X}$, ou seja, os autovalores $\hat{\lambda}_i$ $i = 1, 2, ..., p$ e os correspondentes autovetores $\underline{\hat{e}}_i$ e constrói-se as componentes principais amostrais $\hat{Y}_i = \underline{\hat{e}}_i' \underline{X}$ $i = 1, 2, ..., p$. As propriedades das componentes principais se mantêm e são obtidas com base em estimadores.

A obtenção das componentes principais com base nas informações da matriz de correlação é preferida, devido ao fato de se conseguir eliminar o efeito de escala nos valores das componentes do vetor de variáveis originais $\underline{X}$. Como é bem conhecida, a matriz de correlação é uma matriz de covariância, mas de variáveis padronizadas. Assim, consegue-se eliminar a influência da escala na magnitude das variâncias.

Os autovalores e os autovetores da matriz de correlação são a essência do método das componentes principais. Os autovetores definem as direções da máxima variabilidade e os autovalores especificam as variâncias. Quando os primeiros autovalores são muito maiores que os demais, a maior parte da variância total pode ser explicada por um número menor do que as p dimensões do vetor $\underline{X}$. Cabe também realçar que o desenvolvimento da Ciência Estatística ensejou, com o passar do tempo, o aparecimento de outro método de extração dos fatores, que é o da máxima verossimilhança. Os dois métodos estão disponíveis nos modernos programas computacionais.

c) Critérios Para Definição do Número de Componentes Principais Extraídas

Segundo Johnson e Wichern (1998) não existe resposta definitiva para a questão de quantos componentes reter; o que se pode considerar é a quantidade de variância total da amostragem explicada, os tamanhos relativos dos autovalores e as interpretações em questão de componentes. Se uma componente associada com um autovalor próximo de zero é, então, considerado não importante, e ainda pode indicar uma dependência linear não suspeita nos dados.

Um critério para a determinação do número de fatores a ser extraído foi sugerido por Kaiser em 1960 segundo Johnson e Wichern, 1988, que propõe escolher-se somente os fatores correspondentes aos autovalores (raízes latentes) maiores do que um. Outra maneira de se definir o número de fatores é através da percentagem de variação explicada pelos fatores. O pesquisador, neste caso, deve julgar se m fatores explicam suficientemente o relacionamento. Geralmente, um bom grau de explicação é superior a 75% para um m pequeno.

Um procedimento que visualiza muito bem o Critério de Kaiser é grafar os autovalores contra o número de fatores na ordem de extração (*Scree Plot*). Fixando-se um nível de corte fica fácil decidir a definição de *m*.

2.3.3.2 Análise de Agrupamento (*Clusters*)

A Análise de Agrupamento consiste em uma técnica exploratória cuja aplicação tem por objetivo a formação de grupos homogêneos de objetos (ou variáveis). Os grupos são formados calculando-se as distâncias entre os itens, representados por vetores compostos pelas suas características, construindo-se uma matriz de distâncias e juntando os itens em grupos conforme suas proximidades. A reunião de itens similares em determinados grupos é importante, pois é freqüente a ocorrência de situações em que se deseja separar objetos por uma determinada característica qualquer. Porém, quando se trabalha com um número grande de variáveis, o agrupamento torna-se difícil.

A utilização de Análise de Agrupamento teve seu início com o surgimento da informática, contudo, seu desenvolvimento se deu na década de 30. A dificuldade de cálculos era bastante grande e a manipulação de matrizes era bastante dispendiosa em termos de trabalho e tempo.

A Análise de Agrupamento possui vasto campo de aplicação (medicina, psiquiatria, entre outros), sendo uma técnica distinta dos Métodos de Reconhecimento de Padrões e Classificação (Análise Discriminante, Regressão Logística). No Reconhecimento de Padrões tem-se um número de grupos conhecidos, e o objetivo é alocar uma nova observação em um destes grupos. Agrupar é uma técnica mais primitiva, no sentido de que nenhuma suposição é feita quanto ao número de grupos ou estrutura de agrupamento. O agrupamento é feito com base na similaridade ou distância.

Segundo Crivisqui (1993) os chamados Métodos de Agrupamento, ou Métodos de Classificação, *Cluster Analysis*, ou Métodos de Classificação Automática, são métodos estatísticos destinados a dividir em subconjuntos (classes) um conjunto de dados observados. Aplicar um método de classificação a um conjunto de observações significa definir nesse conjunto as classes em que se distribuem os elementos do conjunto.

Se *n* indivíduos sobre os quais se observaram *p* características estão representados num espaço de *p* dimensões, chamam-se classes aos subconjuntos de indivíduos desse espaço de representação que são identificáveis.

Não se pode postular a existência de classes num conjunto de observações. Só é possível verificar a existência de níveis de síntese significativos correspondentes à organização em classes e subclasses dos elementos, de modo que os elementos de uma matriz de dados qualquer não são necessariamente classificáveis. Por isso, é necessário explorar previamente a estrutura da informação disponível, antes de orientar-se em direção a um algoritmo de classificação.

Para Lebart et al. (1995) a utilização conjunta da Análise Fatorial e da classificação automática permite pronunciar não somente sobre a composição das classes, mas também sobre suas posições relativas. E correntemente, nas aproximações exploratórias, as partições ou árvores de classificação vêm a completar e matizar as Análises Fatoriais prévias.

a) Medidas de similaridade e dissimilaridade

Quando itens (unidades ou casos) são agrupados, a proximidade é usualmente indicada por uma espécie de distância. Por outro lado, as variáveis são usualmente agrupadas com base nos coeficientes de correlação ou outras medidas de associação. Quanto maior o valor do índice de similaridade, mais parecidos são os objetos. E quanto menor esse índice, tem-se uma maior dissimilaridade e menos parecidos são os objetos.

Existem vários índices de similaridades, sendo que a sua principal medida é o coeficiente de correlação.

Mas, os itens podem sem comparados por uma distância. Existem várias métricas que podem ser usadas, sendo que a principal é Euclidiana.

Distância Euclidiana: Essa é, provavelmente, a mais conhecida e usada medida de distância. Ela simplesmente é a distância geométrica no espaço Multidimensional. A distância entre os itens $\underline{x}$ e $\underline{y}$ é definida por:

$$d\left(\underline{x}, \underline{y}\right) = \sqrt{\sum_{i=1}^{p} \left(x_i - y_i\right)^2} \qquad (2.19)$$

Outras medidas para a distância entre $\underline{x}$ e $\underline{y}$ são definidas por:

Quadrado da distância Euclidiana:

$$d\left(\underline{x}, \underline{y}\right) = \sum_{i=1}^{p} \left(x_i - y_i\right)^2 \tag{2.20}$$

Distância city-block (*Manhattan*):

$$d\left(\underline{x}, \underline{y}\right) = \sum_{i=1}^{p} \left|x_i - y_i\right| \tag{2.21}$$

Distância de Mahalanobis (*distância estatística*):

$$d\left(\underline{x}, \underline{y}\right) = \sqrt{\left(\underline{x} - \underline{y}\right)' s^{-1} \left(\underline{x} - \underline{y}\right)} = \sqrt{\frac{\left(x_1 - y_1\right)^2}{s_1^2} + \cdots + \frac{\left(x_p - y_p\right)^2}{s_p^2}} \tag{2.22}$$

Distância de Minkowski:

$$d\left(\underline{x}, \underline{y}\right) = \sqrt[n]{\left|x_1 - y_1\right|^n + \left|x_2 - y_2\right|^n + \cdots + \left|x_p - y_p\right|^n} = \sqrt[n]{\sum_{i=1}^{p} \left|x_i - y_i\right|^n} \tag{2.23}$$

b) Método de agrupamento

b.1) Método de agrupamento hierárquico

Neste método, no início existem tantos grupos quanto objetos (itens). Diversos objetos semelhantes são agrupados primeiro, e estes grupos iniciais são fundidos de acordo com as suas similaridades, eventualmente, relaxando no critério de similaridade os sub-grupos vão se unindo a outros sub-grupos até formar um grupo único. O procedimento é o seguinte (Johnson e Wichern, 1998):

1. No início tem-se g grupos, sendo que cada um é formado por um único objeto; calcula-se a matriz simétrica de distâncias $n \times n$, $D = (d_{ij})$, onde d_{ij} é a distância ou similaridade entre o objeto i e o objeto j, onde: $d_{11} = d_{22} = ... = d_{nn} = 0$.

2. Na matriz D, acha-se o par de grupos mais próximo (menor distância) e junta-se a estes grupos.

3. O novo grupo formado (*AB*). Nova matriz de distâncias é construída, simplesmente apagando-se as linhas e colunas correspondentes aos grupos *A* e *B* e adicionando-se a linha e a coluna dadas pelas distâncias entre (*AB*) e os grupos remanescentes.

4. Repetem-se os passos 2 e 3, num total de (g – 1), vezes observando-se as identidades dos grupos que são agrupados.

A função de um item ou grupo a outro grupo é feita usando-se ligação. Os tipos de ligações mais comuns são: Ligação Simples (vizinho mais próximo), Ligação Completas (vizinho mais distante), Método de Ward, Método das Médias das Distâncias e Método do Centróide. A seguir serão detalhados os três primeiros tipos de ligações.

Ligações Simples (vizinho mais próximo)

Na ligação simples o agrupamento é feito juntando-se dois grupos com menor distância ou maior similaridade. Uma vez formado o novo grupo, por exemplo, (*AB*), na ligação simples, a distância entre (*AB*) e algum outro grupo *C* é calculado:

$$d_{(AB)C} = \min\{d_{AC}, d_{BC}\} \qquad (2.24)$$

Os resultados obtidos são dispostos graficamente em um diagrama em árvore ou dendograma que possui uma escala para se observar os níveis.

Ligação Completa (vizinho mais longe)

Na ligação completa o procedimento é muito semelhante ao da ligação simples, com uma única exceção. O algoritmo aglomerativo começa determinando a menor distância d_{ik}, constrói-se a matriz de distâncias $D = (d_{ik})$ e os grupos vão se juntando. Se *A* e *B* são dois grupos de um único elemento, tem-se (*AB*) como novo grupo. A distância entre (*AB*) e outro grupo *C* é dada por:

$$d_{(AB)C} = \max\{d_{AC}, d_{BC}\} \qquad (2.25)$$

Método de Ward

De acordo com Johnson e Wichern, o método de Ward é um método hierárquico, baseado em minimizar a “perda das informações” ao juntar grupos. Este método é usado,

geralmente, avaliando a perda de informações utilizando o critério da soma ao quadrado dos erros, *SQE*, entre dois agrupamentos, para todas as amostras. Quando se tem a distância mínima unem –se os grupos próximos, e volta-se a iterar nos grupos. Então tem-se:

$$SQE = \sum_{i=1}^{n} (x_i - \bar{x})'(x_i - \bar{x}) \qquad (2.26)$$

Em que:

$\bar{x}$ é média das amostra;

x_i é uma medida multivariada associada com o *i*-ésimo valor.

Os resultados são fornecidos na forma de dendrograma. O eixo vertical dá os valores de *SQE*.

Por fim, a decisão do número de classes ou tipos para análise é tomada, geralmente, a partir do exame do dendrograma ou árvore hierárquica, onde podem ser lidos os índices de nível (ou índice de similaridade), que são as distâncias euclidianas em que ocorrem as junções dos pontos observados para formar grupos. Um grande salto nesses índices (o que equivale a uma grande distância no dendograma) indica que a agregação reuniu dois grupos muito dissimilares e, em razão disso, deve-se definir o número de grupos anterior a este salto.

2.3.3.3 Discriminação, Classificação e Reconhecimento de Padrões

Na área de Estatística, a construção e a avaliação de regras de reconhecimento e classificação de padrões podem ser baseadas em três métodos principais: Função Discriminante Linear de Fisher, Regressão Logística e vizinhanças mais próximas. Desde a década de 80, surgiu a tecnologia de Redes Neurais (tecnologia emergente) e também métodos de Programação Matemática e muitos outros métodos para formação do conjunto de procedimentos usados no reconhecimento e classificação de objetos e indivíduos.

Existem três questões importantes em Reconhecimento de Padrões:

i) São estas técnicas adequadas ou mesmo aplicáveis para resolver problemas de reconhecimento e classificação?

ii) É possível desenvolver ou modificar modelos úteis para determinados problemas, determinando os parâmetros do modelo?

iii)Existem algoritmos que podem ser aplicados e que são computacionalmente práticos nos procedimentos de solução do problema?

É tratado aqui o método de Análise Multivariada, Análise Discriminante.

Análise Discriminante

A Análise Discriminante é uma técnica multivariada que tem por objetivo tratar dos problemas relacionados com separar conjuntos distintos de objetos (itens ou observações) e alocar novos objetos (itens ou observações) em conjuntos previamente definidos. A Análise Discriminante quando empregada como procedimento de classificação não é uma técnica exploratória, uma vez que ela conduz a regras bem definidas, as quais podem ser utilizadas para classificação de outros objetos.

Os objetivos imediatos da técnica, quando usada para discriminação e classificação são, respectivamente, os seguintes:

1. Descrever algebricamente ou graficamente as características diferenciais dos objetos (observações) de várias populações conhecidas afim de achar "discriminantes" cujos valores numéricos sejam tais que as populações possam ser separadas tanto quanto possível.

2. Agrupar os objetos (observações) dentro de duas ou mais classes determinadas. Tenta-se encontrar uma regra que possa ser usada na alocação ótima de um novo objeto (observação) nas classes consideradas.

Uma função que separa pode servir para alocar, e da mesma forma uma regra alocadora pode sugerir um procedimento discriminatório. Na prática, os objetivos 1 e 2, freqüentemente, sobrepõem-se e a distinção entre separação e alocação torna-se confusa.

A denominação "discriminar" e "classificar" foi introduzida por R. A. Fisher (Johnson e Wichern, 1988) no primeiro tratamento moderno dos problemas de separação.

Discriminação e Classificação: Método de Fisher

a) Método de Fisher para Duas Populações

A idéia de Fisher foi transformar as observações multivariadas $\underline{X}$'*s* em observações univariadas *Y*'*s* tal que os *Y*'*s* das populações π_1 e π_2 sejam separados tanto quanto possível. Como fazer? Fisher teve a idéia de tomar combinações lineares de $\underline{X}$ para criar os *Y*'*s*, dado que as combinações lineares são funções de $\underline{X}$ e por outro lado são de fácil cálculo.

Seja μ_{1y} a média dos *Y*'*s* obtidos dos $\underline{X}$'*s* pertencentes a π_1 (população 1) e μ_{2y} a média dos *Y*'*s* obtidos dos $\underline{X}$'*s* pertencentes a π_2 (população 2), então Fisher selecionou a combinação linear que maximiza a distância quadrática entre μ_{1y} e μ_{2y} relativamente à variabilidade dos *Y*'*s*. Assim, seja:

$\underline{\mu}_1 = E(\underline{X}|\pi_1) =$ valor esperado de uma observação multivariada de π_1. (2.27)

$\underline{\mu}_2 = E(\underline{X}|\pi_2) =$ valor esperado de uma observação multivariada de π_2. (2.28)

e supondo a matriz de covariância

$$\Sigma = E[(\underline{X} - \underline{\mu}_i)(\underline{X} - \underline{\mu}_i)'] \quad i = 1, 2 \tag{2.29}$$

como sendo a mesma para ambas as populações, então, considerando a combinação linear,

$$\underset{1x1}{Y} = \underset{1xp}{\underline{C}'}\,\underset{px1}{\underline{X}} \tag{2.30}$$

tem-se:

$$\mu_{1y} = E(Y|\pi_1) = E(\underline{C}'\underline{X}|\pi_1) = \underline{C}'E(\underline{X}|\pi_1) = \underline{c}'\underline{\mu}_1, \tag{2.31}$$

$$\mu_{2y} = E(Y|\pi_2) = E(\underline{C}'\underline{X}|\pi_2) = \underline{C}'E(\underline{X}|\pi_2) = \underline{C}'\underline{\mu}_2 \tag{2.32}$$

e

$$V(Y) = \sigma_y^2 = V(\underline{C}'\underline{X}) = \underline{C}'\,V(\underline{X})\,\underline{C} = \underline{C}'\Sigma\underline{C}, \tag{2.33}$$

que é a mesma para ambas as populações. Segundo Fisher, a melhor combinação linear é a derivada da razão entre o "quadrado da distância entre as médias" e a "variância de *Y*".

$$\frac{(\mu_{1y} - \mu_{2y})^2}{\sigma_y^2} = \frac{(\underline{c}'\underline{\mu}_1 - \underline{c}'\underline{\mu}_2)^2}{\underline{c}'\sum\underline{c}} = \frac{\underline{c}'(\underline{\mu}_1 - \underline{\mu}_2)(\underline{\mu}_1 - \underline{\mu}_2)'\underline{c}}{\underline{c}'\sum\underline{c}} = \frac{(\underline{c}'\delta)^2}{\underline{c}'\sum\underline{c}} \tag{2.34}$$

onde

$$\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2. \tag{2.35}$$

Seja $\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2$ e $Y = \underline{C}'\underline{X}$, então $\frac{(\underline{c}'\underline{\delta})^2}{\underline{c}'\sum\underline{c}}$ é maximizada por:

$$\underline{C} = k\,\Sigma^{-1}\,\underline{\delta} = k\,\Sigma^{-1}\,(\underline{\mu}_1 - \underline{\mu}_2) \text{ para qualquer } k \neq 0. \tag{2.36}$$

Para $k = 1$ tem-se:

$$\underline{C} = \Sigma^{-1}\,(\underline{\mu}_1 - \underline{\mu}_2) \text{ e } Y = \underline{C}'\underline{X} = (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{X}, \tag{2.37}$$

que é conhecida como Função Discriminante Linear de Fisher.

A Função Discriminante Linear de Fisher transforma as populações multivariadas π_1 e π_2 em populações univariadas, tais que as médias das populações univariadas correspondentes são separadas tanto quanto possível relativamente a variância populacional, considerada comum. Assim tomando-se,

$$Y_0 = (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{X}_0 \tag{2.38}$$

como o valor da Função Discriminante de Fisher para uma nova observação $\underline{X}_0$, e considerando o ponto médio entre as médias das duas populações univariadas,

$$m = \frac{1}{2}(\mu_{1y} + \mu_{2y}), \tag{2.39}$$

como

$$m = \frac{1}{2}(\underline{c}'_1\underline{\mu}_1 + \underline{c}'_2\underline{\mu}_2)$$

$$m = \frac{1}{2}\left[(\underline{\mu}_1 - \underline{\mu}_2)'\sum{}^{-1}\underline{\mu}_1 + (\underline{\mu}_1 - \underline{\mu}_2)'\sum{}^{-1}\underline{\mu}_2\right]$$

$$m = \frac{1}{2}\left[(\underline{\mu}_1 - \underline{\mu}_2)'\sum{}^{-1}(\underline{\mu}_1 + \underline{\mu}_2)\right] \tag{2.40}$$

e tem-se que:

$$E\,(Y_0\,|\pi_1) - m \geq 0 \tag{2.41}$$

e

$$E\,(Y_0 \,|\pi_2) - m < 0, \tag{2.42}$$

ou seja, se $\underline{X}_0$ pertence a π_1, se espera que Y_0 seja igual ou maior do que o ponto médio. Por outro lado se $\underline{X}_0$ pertence a π_2, o valor esperado de Y_0 será menor que o ponto médio. Desta forma, a regra de classificação é:

alocar $\underline{x}_0$ em π_1 se $y_0 - m \geq 0$

alocar $\underline{x}_0$ em π_2 se $y_0 - m < 0$

Geralmente, os parâmetros $\underline{\mu}_1$, $\underline{\mu}_2$ e Σ são desconhecidos, então supondo que se tem n_1 observações da v.a. multivariada,

$$\underline{X}_1{}' = \left[\underline{X}_{11} \vdots \underline{X}_{21} \vdots \ldots \vdots \underline{X}_{p1}\right] \tag{2.43}$$

da população π_1 e n_2 observações da v.a. multivariada,

$$\underline{X}_2{}' = \left[\underline{X}_{12} \vdots \underline{X}_{22} \vdots \ldots \vdots \underline{X}_{p2}\right] \tag{2.44}$$

da população π_2, então os resultados amostrais para aquelas quantidades são:

$$\bar{\underline{x}}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\underline{x}_{i1};\; S_1 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(\underline{x}_{i1} - \bar{\underline{x}}_1)(\underline{x}_{i1} - \bar{\underline{x}}_1)' \tag{2.45}$$

$$\bar{\underline{x}}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}\underline{x}_{i2};\; S_2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(\underline{x}_{i2} - \bar{\underline{x}}_2)(\underline{x}_{i2} - \bar{\underline{x}}_2)' \tag{2.46}$$

Mas uma vez que se assuma que as populações sejam assemelhadas é natural considerar a variância como a mesma daí estima-se a matriz de covariância comum Σ por:

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)} \tag{2.47}$$

que é um estimador não-viciado daquele parâmetro.

Consequentemente, a Função Discriminante Linear de Fisher Amostral é dada por:

$$Y = \underline{\hat{c}}' \underline{X} = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S_p^{-1} \underline{x} \tag{2.48}$$

a estimativa do ponto médio entre as duas médias amostrais univariadas,

$$\bar{y}_1 = \underline{\hat{c}}' \underline{\bar{x}}_1 \tag{2.49}$$

e

$$\bar{y}_2 = \underline{\hat{c}}' \underline{\bar{x}}_2 \tag{2.50}$$

é dado por:

$$\begin{aligned} \hat{m} &= \frac{1}{2}(\bar{y}_1 + \bar{y}_2) = \frac{1}{2}\left[(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S_p^{-1} \underline{\bar{x}}_1 + (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S_p^{-1} \underline{\bar{x}}_2\right] \\ \hat{m} &= \frac{1}{2}(\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S_p^{-1} (\underline{\bar{x}}_1 + \underline{\bar{x}}_2) \end{aligned} \tag{2.51}$$

e finalmente a regra de classificação é a seguinte:

$y_0 - \hat{m} \geq 0$ $\quad$ x_0 é alocado em π_1

$y_0 - \hat{m} < 0$ $\quad$ x_0 é alocado em π_2

A combinação linear particular $Y = \underline{\hat{c}}' \underline{x} = (\underline{\bar{x}}_1 - \underline{\bar{x}}_2)' S_p^{-1} \underline{x}$ maximiza a razão:

$$\frac{(\bar{y}_1 - \bar{y}_2)^2}{S_y^2} = \frac{(\underline{\hat{c}}_1 \underline{\bar{x}}_1 - \underline{\hat{c}}_2 \underline{\bar{x}}_2)^2}{\underline{\hat{c}}' S_p \underline{\hat{c}}} = \frac{(\underline{\hat{c}}' \underline{d})^2}{\underline{\hat{c}}' S_p \underline{\hat{c}}} \tag{2.52}$$

onde:

$$\underline{d} = \underline{\bar{x}}_1 - \underline{\bar{x}}_2 \tag{2.53}$$

e

$$S_y^2 = \frac{\sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{i2} - \bar{y}_2)^2}{n_1 - n_2 - 2} \tag{2.54}$$

b) Discriminação entre Diversas Populações

O método de discriminação exposto para duas populações pode ser estendido para diversas populações. O primeiro objetivo de Fisher com a Análise Discriminante foi o de separar populações. Ele pode, contudo, ser usado também para classificar. Este método não necessita da suposição de que as diversas populações sejam normais multivariadas, entretanto assumindo que as matrizes de covariâncias populacionais Σ'*s* são, $\Sigma_1 = \Sigma_2 = ... = \Sigma_g = \Sigma$. Assim, seja $\underline{\overline{\mu}}$ o vetor médio dos diversos grupos (populações),

$$\underline{\overline{\mu}} = \frac{1}{g}\sum_{i=1}^{g} \underline{\mu}_i \tag{2.55}$$

e B_0 a matriz "Soma de produtos cruzados entre grupos populacionais" tal que,

$$B_0 = \sum_{i=1}^{g} (\underline{\mu}_i - \underline{\overline{\mu}})(\underline{\mu}_i - \underline{\overline{\mu}})' \tag{2.56}$$

A combinação linear $Y = \underline{C}'\underline{X}$ tem por esperança:

$$E(Y) = \underline{c}'E(\underline{x}\,|\pi_i) = \underline{c}'\,\underline{\mu}_i \tag{2.57}$$

para a população π_i e variância:

$$V(Y) = \sigma_y^2 = \underline{c}'V(\underline{X})\underline{c} = \underline{c}'\Sigma\underline{c} \tag{2.58}$$

para todas as populações. Desta forma, o valor esperado $\mu_{iy} = \underline{C}'\,\underline{\mu}_i$ muda quando a população da qual $\underline{X}$ é selecionado é outra. Tem-se então uma média global:

$$\overline{\mu}_y = \frac{1}{g}\sum_{i=1}^{g} \mu_{iy} = \underline{c}'\underline{\overline{\mu}} \tag{2.59}$$

e a razão entre a "soma dos quadrados das distâncias das populações para a média global e a variância de Y' é $\dfrac{\underline{c}'B_0\underline{c}}{\underline{c}'\Sigma\underline{c}}$ que é uma generalização multigrupal do caso de duas populações. Esta razão mede a variabilidade entre grupos de valores (escores) Y relativamente a variabilidade comum dentro dos grupos. Da mesma forma do que no caso de duas populações,

pode-se selecionar $\underline{C}$ que maximiza esta razão. É conveniente normalizar $\underline{C}$ tal que $\underline{c}^{'}\Sigma\underline{c}=1$.

Sejam $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_s > 0$ os $s \leq \min(g-1, p)$ autovalores não-nulos de $\Sigma^{-1}B_0$ e $\underline{e}_1$, $\underline{e}_2$, ..., $\underline{e}_s$ os correspondentes autovetores (escalonados tal que $\underline{e}'\Sigma\underline{e} = 1$). Então o vetor de coeficientes $\underline{c}$ que maximiza a razão $\frac{\underline{c}' B_0 \underline{c}}{\underline{c}'\Sigma\underline{c}}$ é dado por $\underline{c}_1 = \underline{e}_1$. A combinação linear $\underline{c}_1'\underline{X}$ é chamada primeiro discriminante. E de forma análoga, pode-se generalizar para o k-ésimo discriminante com $\underline{c}_k = \underline{e}_k$ com $k = 1, 2, ..., s$.

Como, geralmente, Σ e $\underline{\mu}_i$ não são disponíveis, toma-se amostras aleatórias de tamanhos n_i das populações π_i, $i = 1, 2, ..., g$ e denotando o conjunto de dados (a.a) da população π_i, $i = 1, 2, ..., g$, por ${}_{ni}X_p$ tem-se na j-ésima linha o vetor $\underline{x}_{ij}$ e os estimadores dos parâmetros $\underline{\mu}_i$ e $\overline{\underline{\mu}}$ são:

$$\overline{\underline{x}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\underline{x}_{ij} \tag{2.60}$$

$$\overline{\underline{x}} = \frac{\sum_{i=1}^{g} n_i \overline{\underline{x}}_i}{\sum_{i=1}^{g} n_i} = \frac{\sum_{i=1}^{g}\sum_{j=1}^{n_i}\underline{x}_{ij}}{\sum_{i=1}^{g} n_i} \tag{2.61}$$

A matriz "Soma de produtos cruzados entre grupos", $B_{0,}$ é estimada por:

$$\hat{B}_0 = \sum_{i=1}^{g}(\overline{\underline{x}}_i - \overline{\underline{x}})(\overline{\underline{x}}_i - \overline{\underline{x}})' \tag{2.62}$$

e um estimador para Σ pode ser conseguido com base na matriz W:

$$W = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\underline{x}_{ij} - \overline{\underline{x}}_i)(\underline{x}_{ij} - \overline{\underline{x}})' = \sum_{i=1}^{g}(n_i - 1)S_i \tag{2.63}$$

Conseqüentemente,

$$\frac{W}{n_1 + n_2 + ... + n_g - g} = \frac{(n_1-1)S_1 + (n_2-1)S_2 + ... + (n_g-1)S_g}{n_1 + n_2 + ... + n_g - g} = S_p \tag{2.64}$$

Assim, o mesmo $\underline{\hat{c}}$ que maximiza a razão $\frac{\underline{\hat{c}}'\hat{B}_0\underline{\hat{c}}}{\underline{\hat{c}}'S_p\underline{\hat{c}}}$ também maximiza $\frac{\underline{\hat{c}}'\hat{B}_0\underline{\hat{c}}}{\underline{\hat{c}}'W\underline{\hat{c}}}$. Então, apresentar-se-á o otimizante $\underline{\hat{c}}$ na forma mais usual, que é o autovetor $\underline{\hat{e}}_i$ da matriz $W^{-1}B_0$, porque se $W^{-1}B_0\underline{\hat{e}} = \hat{\lambda}\underline{\hat{e}}$ então $S_p^{-}1\,\hat{B}_0\underline{\hat{e}} = \hat{\lambda}(n_1 + n_2 + ... + n_g - g)\underline{\hat{e}}$, portanto, concluindo que sejam $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_g > 0$ os autovalores não nulos de $W^{-1}B_0$ e $\underline{\hat{e}}_1, \underline{\hat{e}}_2, ..., \underline{\hat{e}}_s$ os correspondentes autovetores, sendo $s \leq \min(g - 1, p)$ e $\underline{\hat{e}}_i$ normalizado tal que $\underline{\hat{e}}_i S_p \underline{\hat{e}}_i = 1$; então o vetor de coeficientes que maximiza a razão citada acima é $\underline{\hat{c}}_1 = \underline{\hat{e}}_1$ e a combinação linear $\underline{\hat{e}}_1'\underline{x}$ é chamada primeiro discriminante amostral. Generalizando, tem-se no passo k o k-ésimo discriminante amostral $\underline{\hat{e}}_k'\underline{x}$, $k \leq s$.

Avaliação de Funções de Reconhecimento e Classificação

a) Critério *TPM* (*Total Probability of Misclassification*)

Uma maneira de julgar a performance de um procedimento de reconhecimento de padrões é calcular a sua taxa de erro de reconhecimento. A *TPM* é dada por:

$$TPM = p_1 \int_{R_2} f_1(\underline{x}) d\underline{x} - p_2 \int_{R_1} f_2(\underline{x}) d\underline{x} \qquad (2.65)$$

onde: p_1 e p_2 são as probabilidades de uma observação pertencer a π_1 ou a π_2, respectivamente. E o menor valor para esta quantidade, obtido pela escolha adequada das regiões R_1 e R_2, é chamado de taxa ótima de erro (*optimum error rate*), *OER*,

$$OER = p_1 \int_{R_2} f_1(\underline{x}) d\underline{x} - p_2 \int_{R_1} f_2(\underline{x}) d\underline{x} \qquad (2.66)$$

com R_1 e R_2 determinados por $R_1 : \frac{f_1(\underline{x})}{f_1(\underline{x})} \geq \frac{p_2}{p_1}$ e R_2 em caso contrário.

Uma medida da performance que não depende da forma da distribuição e que pode ser calculada para qualquer procedimento de classificação é a taxa aparente de erro que é

definida como a fração das observações no treinamento amostral correspondente a reconhecimento equivocado pela função. A taxa aparente de erro é calculada da matriz de confusão que mostra a situação real das observações nos grupos versus o reconhecimento. Para n_1 observações de Π_1 e n_2 de Π_2, a matriz de confusão tem a forma:

Figura 2.2: Matriz de confusão

		Predito		
		Π_1	Π_2	
Classificação atual	Π_1	n_{1C}	n_{1M}	n_1
	Π_2	$n_{2M} = n_2 - n_{2C}$	n_{2C}	n_2

Onde:

n_{1C}: número de itens de Π_1 corretamente reconhecido como de Π_1;

n_{1M}: número de itens Π_1 misturados com de Π_2;

n_{2C}: número de itens Π_2 corretamente reconhecido como de Π_2;

n_{2M}: número de itens Π_2 misturados com de Π_1.

A taxa aparente de erro (*APER*) é dada por:

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \tag{2.67}$$

e é interpretada como a proporção de itens ou observações no conjunto de treinamento que são reconhecidos erroneamente.

b) Abordagem de Lachenbruch

Uma abordagem que costuma funcionar bem, neste caso, é a de Lachenbruch, Lachenbruch (1975), é uma técnica para avaliar a eficiência da regra de classificação, e segue os passos:

1. Comece com o grupo da população Π_1. Omita uma observação deste grupo e construa uma função baseada nas $n_1 - 1$ e n_2 observações.

2. Reconheça (classifique), usando a função, a observação não incorporada.

3. Repita os passos 1 e 2 até que todas as n_1 observações de Π_1 sejam classificadas. Seja $n_{1M}^{(H)}$ o número de observações reconhecidas erroneamente neste grupo.

4. Repita os passos de 1 a 3 para as n_2 observações de Π_2. Seja $n_{2M}^{(H)}$ o número de observações reconhecidas erroneamente neste grupo.

Então,

$$\hat{P}(2|1) = \frac{n_{1M}^{(H)}}{n_1} \tag{2.68}$$

e

$$\hat{P}(1|2) = \frac{n_{2M}^{(H)}}{n_2} \tag{2.69}$$

e a proporção total esperada de erro é:

$$\hat{E}(AER) = \frac{n_{1M}^{(H)} + n_{2M}^{(H)}}{n_1 + n_2} \tag{2.70}$$

Desta forma obtém-se uma regra de reconhecimento e classificação construída com as n observações amostrais e testadas com todas referidas observações.

2.4 ANÁLISE DE REGRESSÃO LINEAR

2.4.1 Introdução

A Análise de Regressão é, entre as técnicas estatísticas, a mais utilizada na prática quando se deseja modelar o relacionamento entre as variáveis (dependente e independentes).

A origem do termo “regressão” deu-se pelo estatístico inglês Francis Galton quando estudou o relacionamento das alturas de pais e filhos. As aplicações desta técnica são numerosas e ocorrem em quase todos os campos científicos, sendo que a seguir, é apresentada de forma resumida esta técnica.

Em Engenharia de Avaliação, considera-se, geralmente, como variável dependente os preços à vista de mercado em oferta e efetivamente transacionado e como variáveis independentes. as características do imóvel.

No modelo linear que representa o mercado, a variável resposta (dependente) é expressa por uma combinação linear das variáveis independentes, de forma original ou transformada, e respectivas estimativas dos parâmetros populacionais, acrescidas de erro aleatório, oriundo de variações do comportamento humano (NBR 14653:2004).

2.4.2 Modelo Linear Geral de Regressão

Seja o modelo:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{2.71}$$

onde $\underline{Y}$ é o vetor aleatório de resposta, $\underline{\beta}$ é o vetor de parâmetros de dimensão p, X é a matriz do modelo de ordem n x p e $\underline{\varepsilon}$ é o vetor aleatório de erros de dimensão n. Assim tem-se:

$$\underset{n\,x\,1}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix} \qquad \underset{n\,x\,p}{X} = \begin{bmatrix} 1 & X_{11} & . & . & X_{1,p-1} \\ 1 & X_{21} & . & . & X_{2,p-1} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & X_{n1} & . & . & X_{n,p-1} \end{bmatrix} \qquad \underset{p\,x\,1}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_{p-1} \end{bmatrix} \quad \underset{n\,x\,1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix} \tag{2.72}$$

Admite-se para o modelo as seguintes suposições:

1) o vetor de erros $\underline{\varepsilon}' = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_n]$ é aleatório, ou seja, as componentes ε_i $i = 1, 2, ..., n$ são variáveis aleatórias;

2) a esperança de cada componente de $\underline{\varepsilon}$ é zero, ou seja, $E(\underline{\varepsilon}) = \underline{0}$;

3) as componentes do vetor $\underline{\varepsilon}$ não são correlacionadas ou melhor cov $(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$ e possuem variância constante, σ^2. Assim, a matriz de covariâncias de $\underline{\varepsilon}$ é a matriz diagonal $\sigma^2 I_n$, onde I_n é a matriz identidade de ordem n, $V(\underline{\varepsilon}) = \sigma^2 I_n$.

O modelo (2.72) com as três suposições anteriores é conhecido como Modelo Linear de Gauss Markov e o Teorema de Gauss-Markov garantem que sob as três suposições e com $X'X$ não singular, o estimador não viciado uniformemente de mínima variância (UMVU) do vetor $\underline{\beta}$ e para a variância σ^2 são, respectivamente:

$$\underline{\hat{\beta}} = (X'X)^{-1}(X'\underline{Y}) \tag{2.73}$$

$$S^2 = \frac{1}{n-p}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \tag{2.74}$$

Uma suposição exigida, além das três já citadas, para o modelo de regressão é a seguinte:

4) a distribuição de ε_i i =1,2, ..., n é a Normal (Gaussiana).

Considerando esta suposição, tem-se o modelo de Gauss-Markov Normal e $Y_i \sim N\left(\sum_{i=1}^{p}\beta_i x_i, \sigma^2\right)$.

2.4.3 Análise da Variância da Regressão

A Análise da Variância é uma das técnicas estatísticas cujas bases foram lançadas por Fisher. Esta é a técnica geralmente usada para verificar se o ajuste de regressão existe. É comum construir-se um quadro que resume as informações da Análise da Variância, para um modelo geral:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_{p\text{-}1} X_{p\text{-}1,\,i} + \varepsilon \tag{2.75}$$

para $p > 2$ parâmetros, tem-se o quadro 2. A seguir.

Quadro 2.1: Análise de variância

Fonte de variação	Soma de quadrados	G.L.	Quadrado médio	F
Regressão	$SQ\text{Regr} = \underline{\hat{\beta}}' X' \underline{Y} - n\bar{y}^2$	$p-1$	$\frac{SQ\,\text{Re}\,gr}{p-1}$	$\frac{SQ\,\text{Re}\,gr}{p-1} / \frac{SQR}{n-p}$
Residual	$SQR = \underline{Y}'\underline{Y} - \underline{\hat{\beta}}' X' \underline{Y}$	$n-\text{p}$	$\frac{SQR}{n-p}$	
Total	$SQT = \underline{Y}'\underline{Y} - n\bar{y}^2$	$n-1$	$\frac{QMT}{n-1}$	

O teste feito com a estatística F (última coluna do quadro 2.1) é o da hipótese nula H_0: $\beta_1 = \beta_2 = ... = \beta_{p-1} = 0$, ou seja, se existe regressão dos X's para Y, ou melhor, se existe relação linear entre a variável resposta Y e as covariáveis X_i $i = 1, 2, ... , p-1$.

2.4.4 Verificação dos Pressupostos do Modelo

a) Homocedasticidade

Homocedasticidade é a variância constante dos resíduos. Esta é uma propriedade essencial, que deve ser garantida, sob pena de invalidar toda a análise estatística. Deseja-se que os erros sejam aleatórios, ou seja, não devem ser relacionados com as características dos imóveis. Se isto não ocorre, há heterocedasticidade. Significa dizer que as chances de ocorrerem erros grandes (ou pequenos) variam conforme o tipo de imóvel. Há tendências nos erros. As conseqüências da heterocedasticidade são que as estimativas dos parâmetros da regressão (β_0, β_1, β_2, ..., β_p) não são tendenciosas mas são ineficientes e as estimativas das variâncias são tendenciosas. Os testes t e F tendem a dar resultados incorretos. Neste caso, os resultados não são confiáveis, ou seja, o modelo pode parecer bom, mas ele não é adequado aos dados, na verdade.

A homocedasticidade pode ser verificada, entre outros, através de gráficos de resíduos (erros). Os gráficos dos erros contra os valores reais e contra os valores calculados pela equação são importantes. Se os pontos estão distribuídos aleatoriamente, sem demonstrar um comportamento definido, há homocedasticidade. Mas se existe alguma tendência

(crescimento, decrescimento ou oscilação), então á heterocedasticidade. Havendo heterocedasticidade, podem ser feitas transformações nas variáveis (geralmente logarítmicas) ou outras soluções mais complexas. O modelo deve ser modificado.

b) Independência serial dos resíduos (não-autocorrelação)

Existe autocorrelação quando os erros são correlacionados com os valores anteriores ou posteriores na série. Este problema também é chamado de correlação serial. Pode surgir por especificação incorreta do modelo da regressão, por causa de erros na forma do modelo ou por exclusão de variáveis independentes importantes para a análise. Existindo autocorrelação, os estimadores ordinários de mínimos quadrados não são mais os melhores estimadores lineares não-tendenciosos (as variâncias amostrais dos coeficientes estimados para a equação serão excessivamente grandes, essas variâncias serão subestimadas, as fórmulas perderão a validade e serão obtidas predições ineficientes). Neste caso existirão outros métodos que produzem menor variância amostral nos estimadores. Além disso, em presença de correlação serial, os testes de significância (t e F) e de construção de intervalos de confiança dos coeficientes da regressão também oferecem conclusões incorretas, isto é, as regiões de aceitação e os intervalos de confiança podem ser mais largos ou mais estreitos do que os calculados, dependendo da tendência ser positiva ou negativa.

A verificação da autocorrelação pode ser feira pela análise do gráfico dos resíduos cotejados com os valores preditos, onde este deve apresentar pontos dispersos aleatoriamente, sem nenhum padrão definido ou pelo teste de Durbin-Watson.

c) Normalidade dos resíduos

A Análise de Regressão baseia-se na hipótese de que os erros seguem uma distribuição Normal (distribuição de Gauss). A condição de normalidade dos resíduos é fundamental para a definição de intervalos de confiança e testes de significância. Ou seja, em presença de falta de normalidade, os estimadores são não-tendenciosos, mas os testes não têm validade, principalmente em amostras pequenas. Entretanto, pequenas fugas da normalidade não causam grandes problemas.

A não-normalidade dos resíduos pode ser causada por violações de outras condições básicas, tais como a heterocedasticidade (variância não constante dos erros) ou a escolha de um modelo incorreto para a equação.

A verificação da normalidade pode ser feita pelos testes de aderência não-paramétricos, como por exemplo, o de Kmogorov-Sminorv.

d) *Outliers*

Denomina-se *outlier* um dado que contém grande resíduo em relação aos demais que compõem a amostra e assim tem comportamento muito diferente dos demais (Dantas, 1998, p.112).

É extremamente importante controlar os *outliers,* porque em virtude da forma de estimação da equação, um erro grande modifica significativamente os somatórios, alterando os coeficientes da equação. Assim, um imóvel apenas pode modificar a equação.

Não existem limites fixos, mas geralmente se adota o intervalo de 2 desvios-padrão em torno da média dos erros. Como a média tem de ser zero, os resíduos padronizados $\left(\frac{\varepsilon_i}{dp_y} \right)$ devem estar entre –3 e 3. Os imóveis com erros que ultrapassam estes limites são elementos suspeitos e devem ser analisados cuidadosamente. A existência de *outliers* deve sempre ser interpretada como um sinal de problemas na amostra.

e) Colinearidade ou multicolinearidade

Define-se multicolinearidade como sendo o problema geral, a existência de relações lineares entre as variáveis independentes, de tal forma correlacionadas umas às outras, tornando-se difícil, se não impossível isolar suas influências separadas e obter uma estimativa precisa de seus efeitos relativos (Johnston, 1986). Quando a relação é exata tem-se o caso da multicolinearidade perfeita. Na prática atual, raramente encontram-se variáveis independentes que são perfeitamente relacionadas. Esse caso não traz problemas, pois é facilmente detectado e pode ser resolvido simplesmente eliminando uma ou mais variáveis independentes do modelo. O interesse no que se refere a multicolinearide está nos casos em que ela ocorre com alto grau, isto é, quando duas variáveis independentes estão significativamente correlacionadas ou quando há uma combinação linear entre um conjunto de variáveis independentes. Assim, a multicolinearidade é mais uma questão de grau do que de natureza (Kmenta, 1978, p.411-423).

O fato de muitas funções e regressões diferentes proporcionarem bons ajustes para

um mesmo conjunto de dados é porque os coeficientes de regressão atendem várias amostras onde as variáveis independentes são altamente correlacionadas. “Assim, os coeficientes de regressão estimados variam de uma amostra para outra quando as variáveis independentes estão altamente correlacionadas. Isso leva a informações imprecisas a respeito dos coeficientes verdadeiros” (Neter e Wasserman, 1974, p.344).

A multicolinearidade geralmente é causada pela própria natureza dos dados, principalmente nas áreas de economia com variáveis que representam valores de mercado. Algumas vezes a multicolinearidade pode também ocorrer devido à amostragem inadequada (Elian, 1998).

Em Análise de Regressão Linear Múltipla, existe um freqüente interesse com relação à natureza e significância das relações entre as variáveis independentes e a variável dependente. “Em muitas aplicações de administração e economia, freqüentemente encontram-se variáveis independentes que estão correlacionadas entre elas mesmas e, também, com outras variáveis que não estão incluídas no modelo, mas estão relacionadas à variável dependente” (Neter e Wasserman, 1974, p.339).

A existência de multicolinearidade tendo sido detectada e considerada prejudicial indica que o pesquisador deve procurar soluções para suavizar seus efeitos ruins. Várias medidas corretivas têm sido repostas, desde simples às mais complexas, para suavizar os efeitos provocados pela multicolinearidade (Elian,1988, p.131-134; Judge et al., 1980, p.464-468).

Algumas soluções para o problema de multicolinearidade são através de: remoção de variáveis, ampliação do tamanho da amostra, adoção de técnicas estatísticas como Análise de Componentes Principais, entre outras.

2.4.5 Poder de Explicação do Modelo

Para se medir o quanto a variabilidade total dos dados é explicada pelo modelo de regressão, compara-se a Soma de Quadrados da Regressão com a Soma de Quadrados Total e tem-se o coeficiente de determinação ou de correlação múltipla ao quadrado R^2,

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad 0 < R^2 < 1 \tag{2.76}$$

Quando o ajuste é bom o modelo explica boa parte da variação total e consequentemente o valor de R^2 é próximo de 1. O coeficiente de determinação é uma medida da qualidade do ajuste.

2.4.6 Relação Entre Variáveis

O coeficiente de correlação é uma medida estatística importante na análise em um modelo de regressão. O grau de relação entre as variáveis, que expressa quão bem essas variáveis estão relacionadas entre si é definido numericamente pelo Coeficiente de Correlação, parâmetro representado por ρ. Com base em n observações do par (X, Y) este parâmetro é estimado pela estatística,

$$\hat{\rho} = r = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\left(X_i - \overline{X}\right)^2 \left(Y_i - \overline{Y}\right)^2}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{2.77}$$

Onde:

$\overline{X}$ é a média da variável independente X;

$\overline{Y}$ é a média da variável dependente Y;

σ_{xy} é a covariância amostral de X e Y;

σ_x é o desvio padrão de *X*;

σ_y é o desvio padrão de *Y*.

O coeficiente de correlação varia entre os limites –1 e 1 podendo, portanto, ser positivo ou negativo ($-1 \leq \rho \leq 1$). Quando o coeficiente de correlação é nulo ($\rho = 0$), significa que não existe nenhum relacionamento entre as variáveis. E quando o coeficiente de correlação é igual à unidade, –1 ou + 1, tem-se um relacionamento perfeito entre elas. O sinal (+) ou (–) das variáveis indica a relação direta ou indireta existente entre as variáveis. O grau de relacionamento entre as variáveis, definido numericamente pelo valor $\hat{\rho}$, em Pereira (1970), pode ser assim interpretado:

Coeficiente		**Correlação**
$\lvert\hat{\rho}\rvert = 0$	.....	relação nula
$0 < \lvert\hat{\rho}\rvert \leq 0{,}30$	.....	relação fraca
$0{,}30 < \lvert\hat{\rho}\rvert \leq 0{,}70$	.....	relação média
$0{,}70 < \lvert\hat{\rho}\rvert \leq 0{,}90$	.....	relação forte
$0{,}90 < \lvert\hat{\rho}\rvert \leq 0{,}99$	.....	relação fortíssima
$\lvert\hat{\rho}\rvert = 1$	.....	relação perfeita

Deve ser observado também que nem sempre uma elevada correlação entre duas variáveis representa a existência de relação de causa e efeito entre as mesmas; é necessário analisar se a correlação é absurda. Esses casos dão origem as chamadas de influência no caso.

O estudo do relacionamento entre um conjunto de variáveis pode ser realizado aplicando diversas técnicas, desde os coeficientes de correlação de Pearson, de Spearman, Análise Fatorial e a Análise de Componentes Principais.

A estatística 2.7.7 é conhecida como o coeficiente de correlação linear de Pearson e é uma medida usada no estudo da relação linear existente entre duas varáveis *X* e *Y*.

2.4.7 Seleção de Variáveis Regressoras

Um dos problemas mais freqüentes em Análise de Regressão é a seleção do conjunto de variáveis independentes a serem incluídas no modelo (Neter e Wasserman, 1974, p.371).

O pesquisador deve especificar o conjunto de variáveis independentes a ser empregado para descrever, controlar ou predizer a variável dependente. Um problema muito difícil de relacionamento que aparece na seleção de variáveis é quando uma equação de regressão é construída com o objetivo de predição e envolve muitas variáveis. Talvez, muitas delas contribuam pouco ou nada para precisão da predição. A escolha apropriada de algumas delas fornece a melhor predição, porém quais e quantas devem ser selecionadas? (Snedecor e Cochram, 1972, p.412-413).

Em algumas áreas, a teoria pode ajudar na seleção das variáveis independentes a serem empregadas e na especificação da forma funcional da relação de regressão. Em tais áreas, os experimentos podem ser controlados para fornecer dados sobre a base de que os parâmetros de regressão podem ser estimados e a forma teórica da regressão testada. Em muitos outros campos, entretanto, modelos teóricos são raros. Assim, os investigadores são freqüentemente forçados a explorar as variáveis independentes para que possam realizar estudos sobre a variável dependente. Obviamente, tais conjuntos de variáveis independentes são grandes. Algumas das variáveis independentes podem ser removidas seletivamente. Uma variável independente pode não ser fundamental ao problema; pode estar sujeita a grandes erros de medidas; e pode efetivamente duplicar outra variável independente da lista. Outras variáveis independentes, que não podem ser medidas, podem ser excluídas ou substituídas por variáveis que estão altamente correlacionadas com estas.

Normalmente, após uma seleção inicial, o número de variáveis independentes que permanece ainda é grande. E assim, muitas destas variáveis estarão altamente intercorrelacionadas. Portanto, o investigador geralmente desejará reduzir o número de variáveis independentes a serem usadas no modelo final. Existem várias razões para isto. Um modelo de regressão com um número grande de variáveis independentes é caro para se utilizar. Dessa forma, modelos de regressão com um número limitado de variáveis independentes são fáceis para se avaliar e estudar. Finalmente, a presença de muitas variáveis

independentes altamente intercorrelacionadas, pode adicionar pouco ao poder de predição do modelo, enquanto retira suas habilidades descritivas e aumenta os erros de predição.

O problema, então, é como reduzir a lista de variáveis independentes de forma a obter a melhor seleção de variáveis independentes. Este conjunto precisa ser suficientemente pequeno para que a manutenção dos custos de atualização do modelo sejam manuseáveis e a análise facilitada, e ainda, deve ser grande o suficiente de forma que seja possível uma descrição, um controle e uma predição adequados.

Os procedimentos de procura para se encontrar o melhor conjunto de variáveis independentes que deve ser empregado após o investigador ter estabelecido a forma funcional da relação de regressão, ou seja, se as variáveis dadas estão na forma linear, quadrática, etc.; se as variáveis independentes são primeiramente transformadas, como por exemplo por transformação logarítmica; e se algum termo de interação foi incluído. Neste ponto, os procedimentos de procura são empregados para reduzir o número de variáveis independentes.

Existem muitos procedimentos de seleção, mas nenhum deles pode, comprovadamente, produzir o melhor conjunto de variáveis independentes. Não existe um conjunto ótimo de variáveis independentes, pois o processo de seleção das variáveis possui julgamentos subjetivos. Dentre os procedimentos, pode-se citar como os mais comumente usados: todas as regressões possíveis, *backward, forward* e *stepwise*.

Todas as regressões possíveis: este procedimento consiste em ajustar todas as possíveis equações de regressão. Após a obtenção de todas as regressões, deve-se utilizar os critérios para comparação dos modelos ajustados. Alguns critérios que podem ser usados são o R^2 (coeficiente de explicação), *MSE* (quadrado médio dos resíduos) e C_p (estatística de Mallows). Para alguns conjuntos de variáveis, os três critérios podem levar para o mesmo "melhor" conjunto de variáveis independentes. Este não é o caso geral, pois diferentes critérios podem sugerir diferentes conjuntos de variáveis independentes. Daniel e Wood (1971, p.86) recomendam, no caso de um grande número de equações alternativas, o critério do erro quadrado total para caracterizar a equação. A principal desvantagem do procedimento de procura de todas as regressões possíveis é a quantidade de esforço computacional necessária, já que cada variável independente potencial pode ser incluída ou excluída, gerando $(2p-1)$ regressões possíveis quando existem p variáveis independentes potenciais (Elian, 1998, p.139; Draper e Smith, 1981, p.296).

***Stepwise* (passo a passo)**: é, provavelmente, o mais amplamente usado dos métodos de pesquisa que não requerem a computação de todas as regressões possíveis. Ele foi desenvolvido para economizar esforços computacionais, quando comparado com a abordagem de todas as regressões possíveis, enquanto atinge um conjunto de variáveis independentes razoavelmente bom. Essencialmente, este método de pesquisa computa uma seqüência de equações de regressão, adicionando ou excluindo uma variável independente em cada passo. A rotina de regressão *stepwise* permite que uma variável independente, trazida para dentro do modelo em um estágio anterior, seja removida subseqüentemente se ela não ajudar na conjunção com variáveis adicionadas nos últimos estágios. Esta rotina empregada conduz a um teste para rastrear alguma variável independente que seja altamente correlacionada com variáveis independentes já incluídas no modelo. A limitação da procura da regressão *stepwise* é que ela presume a existência de um único conjunto ótimo de variáveis independentes e busca identificá-lo. Como notado anteriormente, não existe freqüentemente um único conjunto ótimo. Outra limitação da rotina de regressão *stepwise*, é que ela algumas vezes surge com um conjunto de variáveis independentes razoavelmente fraco para predições, quando as variáveis independentes estão altamente correlacionadas (Draper e Smith, 1981, p.307-312).

Seleção *forward*: este procedimento de procura é uma versão simplificada da regressão *stepwise*, omitindo o teste, se uma variável uma vez que tenha entrado no modelo deva ser retirada. Este procedimento considera, inicialmente, um modelo simples usando a variável de maior coeficiente de correlação com a variável dependente. Uma variável por vez é incorporada até que não haja mais inclusão, e as variáveis selecionadas definem o modelo.

Eliminação *backward*: este procedimento de procura é oposto à seleção *forward*. Ele começa com o modelo contendo todas as variáveis independentes potenciais. O procedimento de eliminação *backward* requer mais cálculos do que o método de seleção *forward*, já que ela começa com o maior modelo possível. Entretanto, ela tem uma vantagem de mostrar ao analista as implicações do modelo com muitas variáveis.

3 MATERIAL E MÉTODO

3.1 MATERIAL

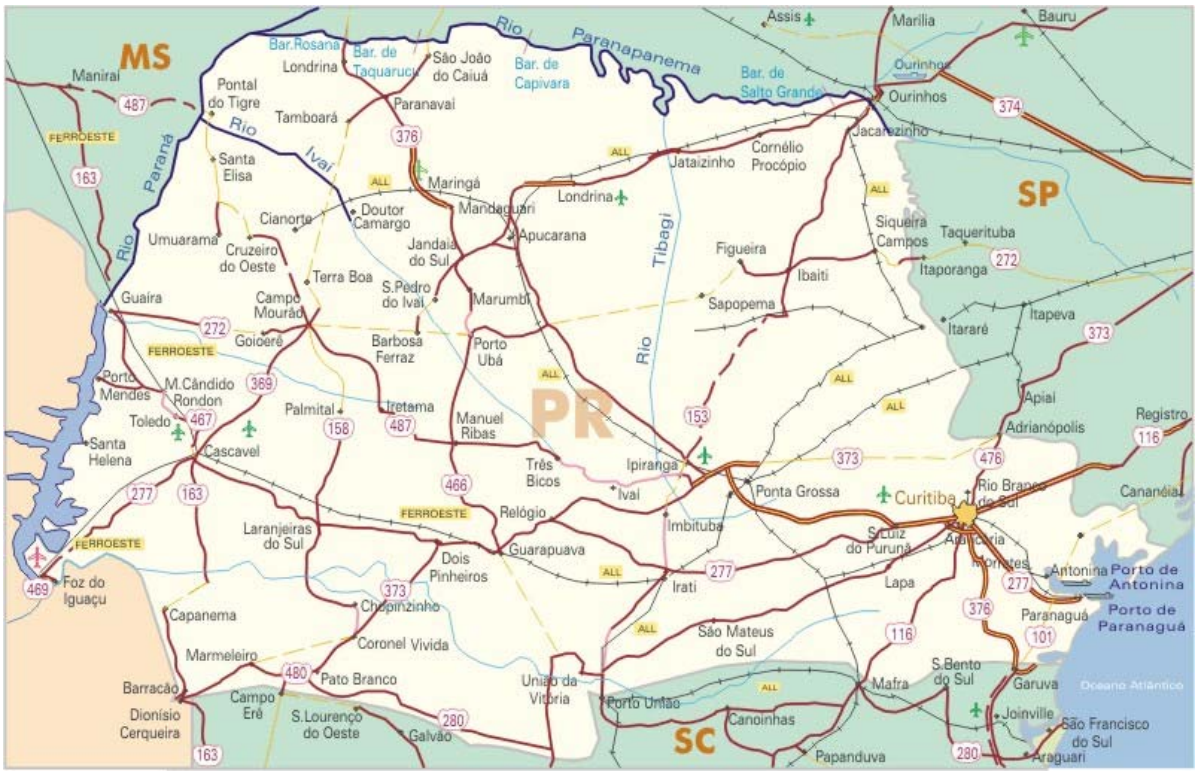
3.1.1 Área de Estudo

A área de estudo é a cidade de Campo Mourão, situada ao Noroeste do Estado do Paraná. A região dos "Campos" bordejados pelas matas Atlântica e das Araucárias, sede da Nação Guarani, começou a ser visitada pelos espanhóis entre 1524 e 1541 e pelos bandeirantes portugueses a partir de 1628. A região pertenceu a Província Del Guaíra com capital em Assuncion. Em 1765 começou a ser guardada por milícias do governo da Província de Piratininga (hoje São Paulo), que denominaram o vale descampado entre os rios Ivaí e Piquiri, de "Campos do Mourão", em homenagem ao governador provincial, Dom Luiz Antônio de Souza Botelho e Mourão. Por volta de 1890, o pasto natural e o cerrado nativo dos "Campos do Mourão" serviam de ponto de descanso dos tropeiros que por ali passavam, tocando boiadas para negociar no Mato Grosso do Sul. Em 1903 chegou e fixou-se nos "Campos do Mourão" a família do paulista José Luiz Pereira, seguida das famílias dos Teodoro, dos Custódio, dos Oliveira, dos Mendonça, dos Mendes e dos guarapuavanos Guilherme de Paulo Xavier, João Bento, Norberto Marcondes, Jorge Walter, dentre outros pioneiros que se fixaram em grandes áreas no território de Campo do Mourão. Até 1943, "Campos do Mourão" pertencia ao município de Guarapuava. A partir desse ano passou a distrito do município de Pitanga e no dia 10 de outubro de 1947 foi emancipado política e economicamente pela Lei 02/47, sancionada pelo governador Moysés Lupion. Até a década de 60, o município de Campo Mourão compreendia toda a Microrregião 12 e os municípios que hoje a integram eram seus distritos administrativos. Na década de 80, foram desmembrados dois dos seus distritos administrativos: Luiziana e Farol do Oeste. Ficou, então, sob sua tutela apenas o distrito de Piquirivaí.

Campo Mourão é cidade Pólo do Centro-Oeste paranaense. Situa-se a 80 km de Maringá, 320 km de Foz do Iguaçu, 460 km de Curitiba e 750 km de São Paulo e se constitui no maior entroncamento rodoviário do Sul do Brasil, sendo ponto estratégico na rota do anel de integração do Estado do Paraná. Com isto integra-se toda malha rodoviária paranaense ao Mercosul com fácil acesso aos Estados de São Paulo, Mato Grosso do Sul, Santa Catarina, Rio Grande do Sul e também para a Argentina, Paraguai e Uruguai. Campo Mourão tem um solo fértil, uma vez que está localizada numa região com as melhores e mais produtivas terras do Estado. Sede da Microrregião 12 (divisão administrativa estadual) Campo Mourão agrega 24 municípios com economia baseada inicialmente no setor primário e hoje realiza investimentos na área industrial, já em avançado estágio de implementação do setor secundário e desenvolvimento do terciário.

As Coordenadas geográficas do Município são 24º02'38" de Latitude Sul e 52º22'40" de Longitude Oeste do Meridiano de *Greenwich*, a uma altitude média de 630 metros sobre o nível do mar. A seguir, na figura 3.1, tem-se a localização da cidade no mapa do Estado do Paraná.

Figura 3.1: Mapa rodoviário do Paraná



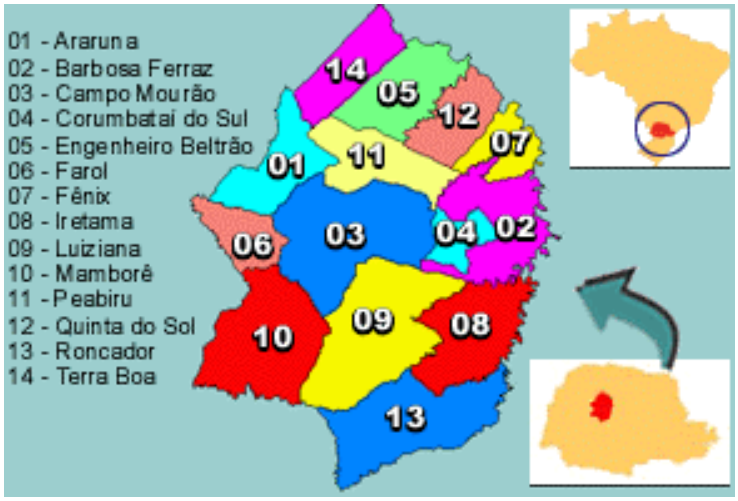
Fonte: Prefeitura Municipal de Campo Mourão, 2005.

Campo Mourão limita-se com os seguintes Municípios:

- Norte: Peabiru
- Nordeste: Barbosa Ferraz
- Sul: Luiziana
- Leste: Corumbataí do Sul
- Oeste: Farol e Mamborê
- Noroeste: Araruna

A seguir, na figura 3.2, tem-se o mapa dos municípios vizinhos de Campo Mourão.

Figura 3.2: Mapa dos municípios vizinhos à Campo Mourão



Fonte: Prefeitura Municipal de Campo Mourão, 2005.

A fertilidade da terra permite uma grande produtividade no campo. A área cultivada de Campo Mourão ultrapassa os 50 mil hectares. As principais culturas são: soja, trigo, milho, algodão e cana-de-açúcar. Paralelamente à agricultura, destaca-se o parque de revendas e assistência técnica de equipamentos e insumos. Campo Mourão é sede da cooperativa singular da América Latina, a Coamo - Cooperativa Agropecuária Mourãoense. Também é dessa região um dos maiores rebanhos bovinos do Paraná. Todo esse potencial agropecuário viabiliza a instalação de novas agro-industriais.

Atualmente cerca de trezentas indústrias operam na cidade. São 26 empresas atuando diretamente na área extrativa e várias outras em diversos setores como embalagens de papelão, tinturaria, fiação de algodão, óleo vegetal, álcool, vestuário e confecção, calçados, materiais eletro-eletrônicos, metalúrgicas, industrialização do milho e café. No setor terciário,

Campo Mourão é pólo atacadista com cerca de 70 estabelecimentos em operação. O comércio varejista tem aproximadamente 900 lojas.

O clima do Município de Campo Mourão é subtropical úmido mesotérmico, com verões quentes e geadas pouco freqüentes, com tendência de concentração das chuvas nos meses de verão, sem estação seca definida. A média das temperaturas dos meses mais quentes é superior a 22ºC e a dos meses mais frios é inferior a 18ºC. A temperatura média anual está entre 20ºC e 21ºC. Os índices pluviométricos apresentam-se em média entre 1.400mm e 1.500mm por ano, tendo nos meses de verão as maiores concentrações de chuvas e nos meses de inverno as menores. Os ventos predominantes na região são os de quadrante nordeste, apresentando probabilidade de geadas nos meses de inverno, quando os ventos sopram de sul e sudoeste.

O solo predominante é o roxo, de textura argilosa, profundo, muito fértil, de grande aptidão para sustentar intensa atividade agrícola.

O Município de Campo Mourão pertence à bacia hidrográfica do Rio Ivaí, sendo seu rio mais importante o Rio Mourão, que atravessa o Município de sul a norte. A vazão deste rio, associada à topografia de seu vale, oferece o maior potencial hidrodinâmica do Município. Outros rios, importantes por serem condicionantes físico-naturais à expansão urbana de Campo Mourão, são o Rio Km 119 e Rio do Campo.

A seguir tem-se outras informações:

Área da unidade territorial: 757,11 km²;
População estimada em 2004: 81.259 habitantes;
Pessoas Residentes na Área Urbana: 74.754 habitantes;
Domicílios particulares permanentes em 2004: 22.829 domicílios;
Atividades imobiliárias (aluguéis e serviços prestados às empresas): 36 empresas.

3.1.2 Limitação da Pesquisa

Essa pesquisa, dentre outros métodos para avaliação de imóveis, se limita a usar uma técnica de Análise de Regressão Múltipla, com o apoio de técnicas de Análise Multivariada.

A aplicação dessa técnica se restringe aos imóveis urbanos nos segmentos de casas, apartamentos e terrenos da cidade de Campo Mourão – Paraná, no ano de 2004.

Outra limitação do trabalho é o espaço de tempo. Como qualquer alteração na economia provoca modificações nos valores do imóvel, estes estão sujeitos às influências dos governantes e das economias local, regional, nacional e global. Assim, no decorrer do tempo, existe uma flutuação dos valores dos imóveis.

3.1.3 Levantamento dos Dados

O levantamento dos dados foi feito de forma cautelosa, pois dele depende o sucesso da Análise Estatística. Foi realizado um planejamento antes da coleta dos dados. Nesse planejamento contemplou-se o espaço físico, local onde está inserido o total de imóveis, população a estudar e o número de imóveis a serem pesquisados. No mercado de imóveis, é freqüente a entrada de dados novos, 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).

Na determinação da oferta imobiliária existem aspectos de extensa variação e combinação de atributos constituindo a heterogeneidade do produto habitação. Essa dispersão deve estar presente na descrição completa do mercado, incluída nas faixas de preços, tamanhos dos imóveis e, ainda, nas diferentes localizações. Assim, faz-se necessário obter o maior número de dados e atributos possível.

3.1.3.1 As Variáveis Utilizadas

As variáveis explicativas (independentes) são do tipo quantitativas e qualitativas, representando as características do imóvel, e estão a seguir detalhados.

A variável resposta (dependente) é o preço, que representa o valor de venda do imóvel em reais. As varáveis originais, independentes estão relacionadas, classificadas e descritas nos quadros 3.1, 3.2 e 3.3 para os tipos de imóveis, apartamentos, casas e terrenos respectivamente.

Quadro 3.1: Variáveis independentes para apartamento

Variáveis	*Categorias*	*Descrição*
Revestimento do prédio	1 a 4	Identifica o revestimento do prédio.
Andar	1 a 4	Identifica o andar que o apartamento está localizado. Sabe-se que dependendo do andar que se localiza o apartamento ele é mais ou menos valorizado.
Dependência de empregado	Sem = 0 Com = 1	Identifica a existência ou não de dependência de empregados.
Estado de conservação	1 a 4	Identifica o nível de conservação do imóvel.
Suíte	Sem = 0 Com = 1	Identifica a presença ou não de suíte, atribuindo o valor 1 mesmo quando há presença de mais de uma suíte.
Idade aparente Idade real	Até 1 ano = 6 2-5 anos =5 6-10 anos = 4 11-15 =3 15-20 anos =2 mais de 20 anos = 1	Idade real: idade cronológica do edifício, reflete o estágio tecnológico. Idade aparente: idade aparente do edifício. Por ser uma variável contínua, a idade do imóvel dividiu-se em períodos.
Proximidade	1 a 3	Identifica a quanto o imóvel se localiza próximo de escolas, supermercados, hospitais e do centro comercial.
Lavanderia	0 a 1	Identifica a existência ou não de lavanderia.
Posição do apartamento	1 a 3	Identifica a posição do apartamento em relação ao prédio (frente , lateral ou fundo).
Padrão de acabamento	1 a 3	Identifica os vários níveis de acabamento.
Sala	Unidade	Indica o número de salas existentes no apartamento.
Pavimento	Unidade	Indica o número de pavimentos do prédio.
Garagem	Unidade	Quantifica o número de vagas para carros disponível para cada apartamento.
Dormitório	Unidades	Quantifica o número de dormitórios.

Elevador	Unidades	Identifica a quantidade de elevadores no prédio.
Área privativa	m^2	Corresponde à superfície ou área do apartamento expressa em metros quadrados, obtida do registro de imóveis.
Peças	Unidades	Quantifica as peças constituinte do imóvel.
Banheiro	Unidades	Identifica o número de banheiro social.

Fonte: Imobiliária Tapowik, 2004.

Quadro 3.2: Variáveis independentes: casas residenciais

Variáveis	*Categorias*	*Descrição*
Localização	1 a 5	Naturalmente um local é "melhor" ou "pior" do que um outro em função de diversas características, entre as quais sua infra-estrutura urbana.
Dependência de empregado	Completa = 1 Incompleta = 0,5 Inexistente = 0	Identifica a existência ou não de dependência de empregado, completa ou incompleta.
Nível de conservação	1 a 4	Identifica o nível de conservação do imóvel.
Suíte	0 a 1	Identifica a presença ou não de suíte, atribuindo o valor 1 mesmo quando há presença de mais que uma suite.
Idade aparente	Até 1 ano = 6 2-5 anos =5 6-10 anos = 4 11-15 =3 15-20 anos =2 mais de 20 anos = 1	Por ser uma variável continua a idade do imóvel dividiu-se em períodos.
Garagem	0 a 1	Identifica a presença de garagem, onde é atribuído o valor mesmo quando a mais que uma vaga.
Distância de supermercados	1 a 3	Identifica a proximidade do imóvel de grandes mercados.
Presença de lavanderia	0 a 1	Identifica a existência ou não de lavanderia.
Edícula	0 a 1	Identifica a presença (1) ou não (0) de edícula.
Padrão de acabamento	1 a 3	Identifica os vários níveis de acabamento.
Piscina	0 a 1	Identifica a existência ou não de piscina.
Cobertura	1 a 4	Identifica o tipo de cobertura do imóvel.
Estrutura	1 a 5	Identifica o material de construção do imóvel.
Dormitório	Unidades	Quantifica o número de dormitórios.
Área do terreno	m^2	Identifica a área do terreno.
Área construída	m^2	Identifica a área total construída.
Peças	Unidades	Quantifica as peças constituinte do imóvel.

Banheiro	Unidades	Identifica o número de banheiro social.

Fonte: Imobiliária Tapowik, 2005.

Quadro 3.3: Variáveis independentes: terrenos

Variáveis	*Categorias*	*Descrição*
Localização	1 a 6	Variável que qualifica a localização do imóvel.
Pólo de influência	–1 a 1	Indica se móvel se localiza próximo a um local que influência o valor do imóvel.
Plano	0 a 3	Identifica se o terreno está acima, abaixo ou ao nível da rua.
Inclinado	0 a 3	Indica o nível de inclinação do terreno.
Pavimentação	0 a 1	Identifica a presença ou não de pavimentação na rua ou avenida onde está inserido o terreno.
Proteção	0 a 1	Indica se o terreno possui ou não proteção (muro ou cerca).
Posição	1 a 2	Identifica a posição do terreno na quadra (meio ou esquina).
Frente	1 a 3	Identifica a largura do terreno. Sabendo que um terreno de frente com maior metragem possui uma melhor valorização.
Ponto Comercial	0 a 3	Sabendo que os terrenos localizados em zona de comércio ou de moradia, o terreno é mais ou menos valorizado. Esta variável identifica os vários níveis de localização.
Área do terreno	m^2	Quantifica a área do terreno.

Fonte: Imobiliária Tapowik, 2005.

3.1.3.2 Questionário Proposto Para Coleta de Dados

Na formulação do questionário, anexos IV e V, para obtenção das características dos imóveis, inicialmente visitou-se uma imobiliária (a maior de Campo Mourão) para conseguir subsídios na elaboração do questionário. Isto foi feito com finalidade de apurar quais características dos imóveis (casas, apartamentos e terrenos) da cidade de Campo Mourão. Na composição do questionário procurou-se agregar o maior número possível de informações sobre as características dos imóveis. A partir daí idealizou-se o documento, contendo informações dos imóveis tanto quantitativo (área de construção, quantidade de quartos, etc.) quanto qualitativo (conservação, padrão de construção, etc.), vide nos anexos I, II e III. Após elaborado o questionário, visitou-se as imobiliárias de Campo Mourão para o preenchimento

do mesmo, já que dificilmente o preço total, as condições de pagamento e as características dos imóveis estão claramente estampadas nos anúncios, o que pode estar caracterizando uma atuação especulativa ou estratégica das imobiliárias diante dos compradores potenciais e da concorrência. As informações relativas às vendas efetuadas são de domínio dos gerentes de venda e dos corretores e consegui-se total apoio das imobiliárias perante as informações coletadas junto a esses profissionais.

3.1.3.3 Amostra

A amostra foi constituída por 119 imóveis. Sendo 44 apartamentos, 51 casas e 24 terrenos localizados na área urbano da Cidade de Campo Mourão – PR, dos quais 80 estão localizados na área central.

3.2 METODOLOGIA PARA O DESENVOLVIMENTO DA PESQUISA

A metodologia aqui proposta procura determinar classes homogêneas de apartamentos, casas e terrenos através de uma Análise de Agrupamento aplicada à amostra considerada. Utilizou-se as técnicas de inferência, ao nível de avaliação rigorosa, de acordo com NB-502/89 (avaliação de imóveis urbanos), com o auxílio do *Software Statgraphics Plus* 5.0 para o processamento dos dados e o *Software Excel* ®, por ser um aplicativo de uso quase universal.

A Análise de Agrupamento foi aplicada para juntar imóveis semelhantes. Utilizou-se a Distância Euclidiana e a ligação pelo método de Ward. Então, a partir dos grupos formados (*clusters*), aplicou-se o método de Componentes Principais, procurando obter uma redução da dimensionalidade dos dados. Assim, conservou-se as primeiras componentes e as que constituem um resumo de informação mais importante da estrutura de covariância. Com a finalidade de alcançar um dos objetivos deste, que é a obtenção do modelo de precisão, foi desenvolvido um estudo com a técnica da Regressão Linear Múltipla, para prever dentro de cada agrupamento o valor de um novo imóvel. As primeiras ferramentas descritas são para

atender o objetivo de estudo das variáveis que participam da construção do modelo de Regressão Linear Múltipla.

3.2.1 Considerações Para a Construção do Modelo

As etapas, ou roteiro, que são necessárias para construir um modelo matemático através de critérios multivariados usando a Regressão Linear Múltipla com a finalidade de estimar valores de imóveis urbanos em Campo Mourão são apresentadas a seguir.

3.2.1.1 Identificação das Variáveis Independentes

Uma das dificuldades existentes na avaliação de imóveis é a determinação das variáveis que influenciam no seu valor. São muitos os fatores que devem ser levados em consideração, mas nem sempre se pode desenvolver um único modelo representativo da realidade do conjunto do mercado de imóveis. Um dos aspectos mais importantes na avaliação de imóveis é a seleção das variáveis independentes que possam ser utilizadas na regressão, que são aquelas que tem influência na formação do preço, pois várias importantes na formação de valor de uma determinada categoria ou subconjuntos de imóveis não necessariamente são as mesmas que para outro subconjunto, inclusive dentro de uma mesma região. As possíveis variáveis independentes (ou explicativas) que podem influenciar no preço de um imóvel devem ser listadas a priori. A definição das variáveis explicativas preliminarmente economiza tempo e diminui o custo de execução da pesquisa. Resulta necessário em ocasiões desestimar alguns dos elementos da amostra coletada pelo fato de serem elementos diferenciados do resto, razão pela qual sua presença afeta fortemente os valores globais da equação de regressão, não permitindo então a sua consideração no modelo de avaliação. As variáveis explicativas para a avaliação de imóveis são aquelas referentes a todas as características físicas e locacionais do imóvel. No entanto, dentre todas as características físicas e locacionais relacionadas a um imóvel, nem todas são relevantes à formação de seu preço.

De forma geral e preliminar, pode-se citar como relevantes à formação do preço, as

seguintes variáveis explicativas.

Apartamentos: Área total, área útil, número de dormitórios, número de suíte, número de carros na garagem, dependências de empregados, idade do imóvel, elevador, estado de conservação, padrão de acabamento, região de valorização imobiliária, distância à escola, etc.

Casas residenciais: Estado de conservação, área construída, área do terreno, localização, número de suítes, dependência de empregados, estrutura, padrão de acabamento, entre outras.

Terrenos: Área total, comprimento frontal (frente), localização, área comercial, etc.

Essa lista de variáveis explicativas tende a variar de município para município, dependendo das características de cada um. Para a cidade de Campo Mourão – PR, a variável explicativa mais relevante é a de proximidade do centro comercial.

3.2.1.2 Transformações de Variáveis

As variáveis que são definidas para a caracterização e localização de um imóvel são do tipo quantitativas ou qualitativas. Geralmente, estas variáveis precisam sofrer transformações para que então possam ser realizadas as análises. As variáveis qualitativas devem ser quantificadas através de uma codificação adequada. Em muitas situações são atribuídas para as variáveis qualitativas o valor 0 (zero) quando não tem a característica e 1 (um) caso contrário. Então, tem-se uma variável do tipo *dummy*, pronta para ser utilizada para análise. E outras variáveis que se referem às características qualitativas dos imóveis, como a conservação do imóvel (péssimo, regular, bom e ótimo); classificação do imóvel (baixo, normal e alto) e outras, são casos que são resolvidos dando pesos para a característica. Geralmente esses pesos são na ordem crescente, da situação menos favorável para a mais favorável. E ainda, quando uma variável pode vir a gerar um número muito grande de modalidades, algumas vezes, ela pode ser definida por uma escala numérica, atribuindo-se também pesos às modalidades, (por exemplo, a idade do imóvel).

As variáveis originais, quadros 3.1, 3.2 e 3.3, foram transformadas utilizando a

técnica multivariada Análise de componentes Principais.

3.2.1.3 Análise Exploratória

Para o estudo de relacionamento entre as variáveis pode ser utilizado, entre outros, o coeficiente de correlação de linear de Pearson para as variáveis quantitativas e as qualitativas, sendo que, no segundo caso, elas devem ser transformadas em *dummy*. O coeficiente de correlação indica a existência ou não de relação linear significativa entre as variáveis independentes e a variável dependente, informação necessária para uso da regressão linear. Esses coeficientes, quando apresentam valores altos entre as variáveis independentes, indicam a possível existência de multicolinearidade, e ainda, o valor do determinante da matriz ($X'X$), quando é próximo de zero, também indica a existência de multicolinearidade. Apesar do coeficiente de correlação linear de Pearson e do determinante da matriz ($X'X$) indicarem a existência da multicolinearidade, eles não a quantificam.

3.2.1.4 Análise dos Resíduos

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

3.2.1.5 Verificação da Adequação do Modelo

Um último passo que deve ser realizado antes de adotar o modelo para avaliação de imóveis, é verificar sua aplicabilidade. Inicialmente, deve-se fazer a Análise de Variância para

testar a significância do modelo ajustado, no entanto, isto por si só não garante a qualidade das predições. A qualidade do ajuste pode ser testada comparando os valores preditos com os valores observados. O ajuste é tão bom, quanto maior for a quantidade de valores preditos próximos dos valores observados, isto é, com pequeno erro de predição. O valor do coeficiente de determinação R^2 é importante para definir a qualidade do modelo adotado.

3.3 ESTUDO DE CASO

Neste trabalho, efetuou-se um estudo no mercado imobiliário da Cidade de Campo Mourão – PR, restringindo-se ao segmento de imóveis urbanos, cujo objetivo será modelar este mercado através da análise de regressão, pautando-se da análise multivariada e estimar ou calcular o valor de venda de apartamentos de forma absolutamente objetiva, sem qualquer "opinião" originária da subjetividade intrínseca do ser humano.

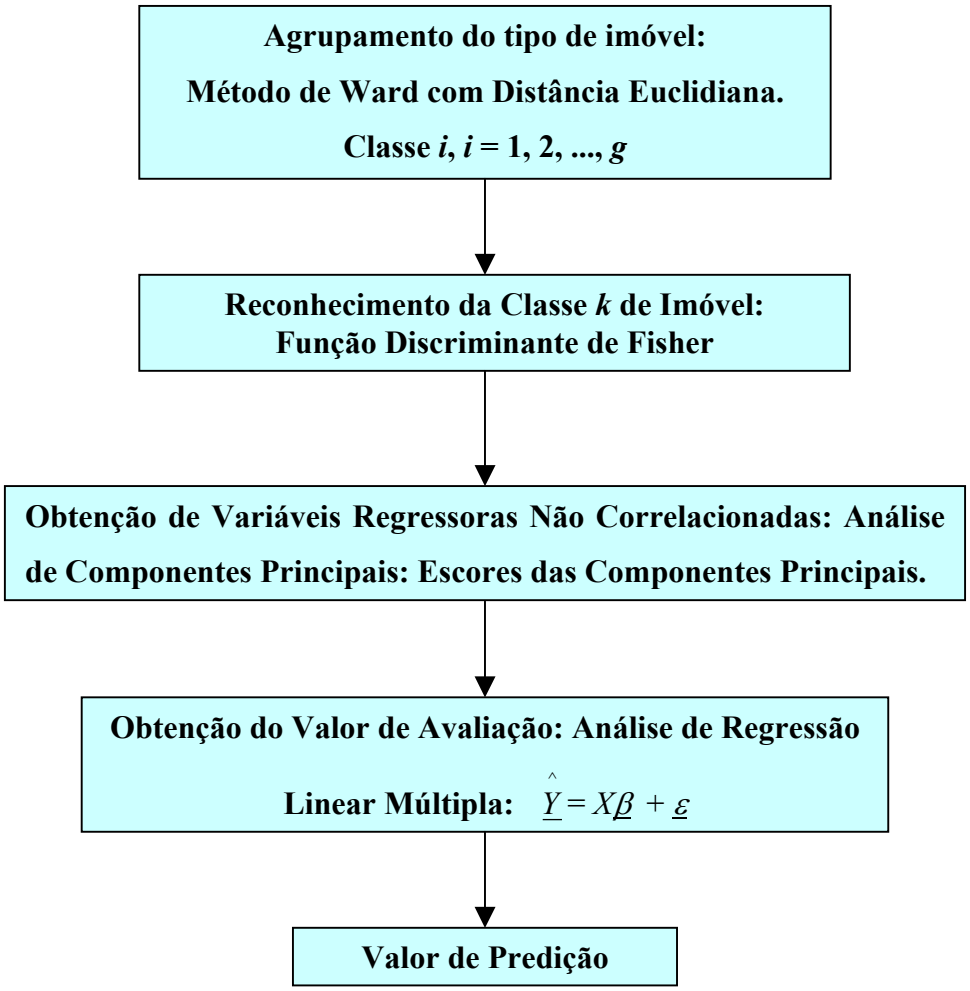
O *Software* utilizado para a construção da tabela de dados e as devidas transformações foi o *Excel*.

As matrizes de dados resultantes (apartamentos, casas e terrenos) foram submetidas aos tratamentos estatísticos descritos no segundo capítulo. Os resultados estatísticos foram obtidos através do *Software Statgraphics Plus* 5.0.

Em primeiro lugar as matrizes de dados foram submetidas à Análise de Agrupamentos (*clusters*) hierárquicos, utilizando-se a Distância Euclidiana, sendo que os agrupamentos foram feitos por meio da ligação de Ward para formar as classes homogêneas. Após várias simulações, ficou claro que o número ótimo de classes a considerar para o caso de residências seria quatro e para os apartamentos três, enquanto para os terrenos apenas duas classes. Após a formação das classes homogêneas realizou-se uma Análise Discriminante para avaliar a consistência das classes obtidas. Em seguida realizou-se uma Análise de Componentes Principais para cada classe formada para os tipos de imóveis. Com a obtenção dos escores das Componentes Principais para explicar a variação total, substituiu-se as variáveis explicativas originais. Por fim foi desenvolvido um modelo de Regressão Linear Múltipla para cada uma das classes de cada tipo de imóvel. Considerou-se como variável resposta o preço de venda à vista, que denominou-se valor.

O fluxograma do método utilizado é apresentado na figura 3.3.

Figura 3.3: Fluxograma do método de avaliação proposto



Fonte: A Autora, 2005.

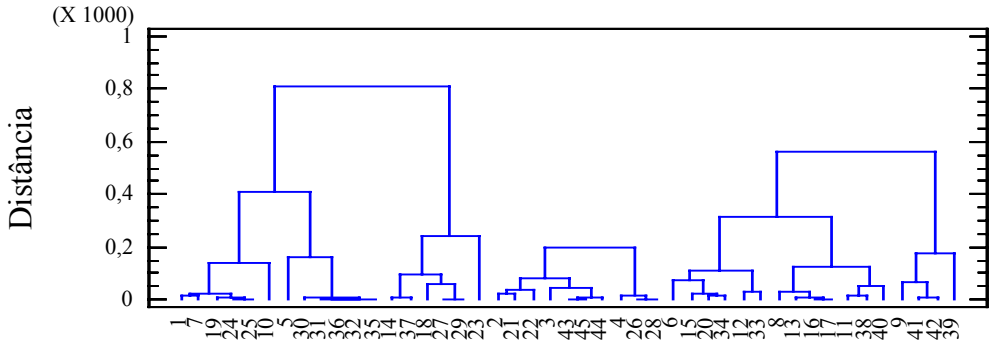
Os resultados e discussão obtidos para cada tipo de imóvel são apresentados no capítulo 4 a seguir.

4 RESULTADOS E DISCUSSÕES

4.1 APARTAMENTOS

Para formar as classes, a técnica multivariada de Análise de Agrupamento, foi aplicada começando-se com cada observação como sendo um grupo individual. Juntou-se, então, as duas observações que eram mais próximas para formar um grupo novo. Depois de recalcular as distâncias entre os grupos, tem-se nova matriz de distâncias onde, os dois grupos mais próximos foram combinados. Esse processo foi repetido até que três grupos permaneceram como mostra a figura 4.1.

Figura 4.1: Dendrograma das classes formadas de apartamentos



Fonte: A Autora, Pesquisa de Campo, 2004.

A classe 1 contém 38,64% dos apartamentos, a classe 2 contém 22,73% e a classe 3 contém 38,64% dos apartamentos analisados, conforme a tabela 4.1 adiante.

Tabela 4.1: Resultado das classes formadas de apartamentos

Cluster	Membros	Percentagem (%)
1	17	38,64
2	10	22,73
3	17	38,64

Fonte: A Autora, Pesquisa de Campo, 2004.

Após a Análise de Agrupamento, realizou-se a Análise Discriminante e os resultados estão na tabela 4.2 a seguir.

Tabela 4.2: Classificação de apartamentos

Número de Classses	Quantidade de imóveis	classe 1	Classe 2	Classe 3
1	17	17 (100,00%)	0 (0,00%)	0 (0,00%)
2	10	0 (0,00%)	10 (100,00%)	0 (0,00%)
3	17	0 (0,00%)	1 (5,88%)	16 (94,12%)

Fonte: A Autora, Pesquisa de Campo, 2004.

Os resultados apresentaram-se muito consistentes. Entre as 44 observações que ajustam o modelo 97,73% foram classificados corretamente. A interpretação das classes obtidas foi realizada de acordo com as características de cada classe. A seguir apresentam-se os aspectos mais determinantes de cada uma das classes:

Classe 1: Todos os apartamentos estão localizados no centro; possuem pelo menos um elevador; área privada acima de 220 m^2; prédio acima de treze pavimentos; mais de dois dormitórios; todos possuem suíte; dependência de empregada completa; revestimento do prédio é pastilhado e seu valor é superior à R$ 115.000,00.

Classe 2: Todos possuem mais de um elevador; localizados no centro; prédio com pelo menos sete pavimentos; mais de dois quartos; todos possuem suíte; apartamentos com menos de quinze anos; possui área no mínimo de 160 m^2; com valor de mercado acima R$ 175.000,00.

Classe 3: Todos possuem apenas uma vaga na garagem; área privativa menor que

132 m^2; prédios baixos e seus valores estão entre R$ 30000,00 e R$ 110000,00.

Foi ajustado, então um modelo de Regressão Linear Múltipla para cada uma das três classes de apartamentos obtidas pela Análise de Agrupamento (*Cluster Analysis*). Para tanto, foi realizado primeiramente uma Análise de Componentes Principais com os dados das variáveis explicativas originais. Em seguida, com a obtenção dos escores das componentes principais, substituiu-se as variáveis originais pelas componentes principais (escores) e realizou-se a regressão. Considerou-se como variável resposta (dependente) o preço total de venda à vista, que foi denominada valor e como variáveis explicativas as componentes principais.

Os resultados obtidos para cada uma das classes são apresentados a seguir.

4.1.1 Classe 1 de Apartamentos

A tabela 4.3 mostra os autovalores e a percentagem da variância explicada e acumulada das seis componentes principais extraídas. As seis primeiras componentes explicam 87,64% da variabilidade dos dados originais. Sendo que a primeira componente explica 25,775 %, a segunda componente explica 21,685%, a terceira componente explica 14,925%, a quarta componente explica 11,401%, a quinta componente explica 8,229 % e a sexta componente explica 5,626%. Adotou-se como definição do número de componentes a serem tomadas, o Critério de Kaiser (Johnson e Wichern, 1998), porém incluiu-se a sexta componente devido ter o autovalor 0,956367 muito próximo de um.

Tabela 4.3: Análise das componentes principais na classe 1 de apartamentos

Componente	autovalor	% de variâcia	% Acumulada
1	4,38168	25,775	25,775
2	3,68641	21,685	47,459
3	2,5373	14,925	62,385
4	1,93819	11,401	73,786
5	1,3989	8,229	82,015
6	0,956367	5,626	87,640
7	0,644585	3,792	91,432
8	0,548348	3,226	94,658
9	0,397581	2,339	96,996
10	0,375483	2,209	99,205
11	0,118514	0,697	99,902
12	0,0117594	0,069	99,971
13	0,00487836	0,029	100,000
14	1,18656E-16	0,000	100,000
15	0,0	0,000	100,000
16	0,0	0,000	100,000
17	0,0	0,000	100,000

Fonte: A Autora, Pesquisa de Campo, 2004.

A tabela 4.4, adiante, mostra os pesos das variáveis em cada uma das seis componentes.

Tabela 4.4: Pesos das componentes principais na classe 1 de apartamentos

variável	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
ppredio	0,0893909	0,246161	-0,0526886	-0,0979668	0,416273	-0,570934
elevador	-0,0526883	-0,413666	0,0726574	-0,152862	0,35089	-0,0682263
vgaragem	0,0292966	-0,115928	-0,142546	-0,344625	0,513443	0,376004
área const.	0,289871	-0,133004	0,088613	-0,431775	-0,000990526	-0,0850487
pavimento	0,133738	-0,0357628	0,243017	0,47084	0,391287	0,066217
andar	0,208736	0,285383	0,0221086	0,102197	0,246664	-0,27849
pecas	0,225747	0,063101	-0,531922	-0,0212474	0,0193336	0,0672196
sala	0,223168	-0,107456	-0,355162	0,311851	-0,0248622	0,274322
quarto	0,127221	0,400858	-0,25558	-0,00234591	-0,1004	0,11388
banheiro	0,308159	-0,0270833	-0,397666	0,147821	0,00518783	-0,255486
descola	-0,343604	0,278393	-0,136115	-0,045763	0,233369	0,169827
dhospital	-0,363872	0,279642	-0,118547	-0,0527298	-0,0264049	0,15829
dsmercado	-0,381969	0,262674	-0,0967804	-0,0107717	0,212156	-0,0572527
acabamento	0,0963664	0,245521	0,347094	0,306414	-0,0709979	0,0243955
conservação	0,327714	0,14195	0,174912	0,0723854	0,268025	0,455252
idreal	0,263027	0,240884	0,114627	-0,410822	-0,194047	0,064233
idaparente	0,220789	0,338124	0,255992	-0,193006	-0,0148376	0,108

Fonte: A Autora, Pesquisa de Campo, 2004.

Através dos resultados na tabela 4.4, pode-se observar que a primeira componente possui pesos mais altos nas variáveis: distâncias à escola, distância ao supermercado, distância ao hospital, nível de conservação e quantidade de banheiro. A segunda componente possui pesos mais altos nas variáveis: número de elevador, quantidade de quartos e idade aparente do imóvel. A terceira componente tem pesos maiores nas variáveis: número de peças do imóvel, número de sala, número de banheiros, padrão de acabamento. A quarta componente possui pesos mais altos nas variáveis: vaga de carro na garagem, área útil, número de pavimentos, número de salas, padrão de acabamento e idade real. A quinta possui pesos mais altos nas variáveis: componente posição do apartamento no prédio, quantidade de elevador, número de vaga de carro na garagem e quantidade de pavimentos do prédio. A sexta componente possui pesos mais elevados na varáveis: posição do apartamento no prédio, número de vagas de carro na garagem e o nível de conservação.

Os escores fornecidos pelas seis componentes para os 17 apartamentos compõem a tabela 4.5 adiante. Esses são os valores das variáveis explicativas considerados para a Análise de Regressão.

Tabela 4.5: Escores das componentes principais na classe 1 de apartamentos

imóvel	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
1	-2,1512	-0,741581	1,27721	0,867893	-0,379761	-2,77916
2	-0,215068	0,281233	3,1519	1,28665	2,81409	-0,021318
3	0,554301	0,0739132	-0,793701	2,64601	-1,49977	0,403009
4	-0,896709	0,0629922	-0,31083	0,478471	1,86133	0,619036
5	2,97263	-1,50975	2,21074	0,225224	-0,129701	0,859525
6	2,93197	-1,36201	-3,42417	-1,01255	1,87297	0,0124523
7	-0,142195	1,89695	-0,454084	-1,25077	0,675696	0,381225
8	4,01045	-1,83611	-0,735231	-0,644179	-0,672452	-1,73787
9	0,663582	2,92859	-1,21238	1,96973	0,00939094	-0,404564
10	0,28557	2,2559	-1,17425	1,96289	-0,827946	0,66776
11	0,115707	3,10054	0,897485	-2,17611	-0,685172	-0,0879383
12	0,115707	3,10054	0,897485	-2,17611	-0,685172	-0,0879383
13	-2,30919	-1,28958	-0,548849	-0,223296	-0,267799	0,384237
14	-2,46238	-1,34315	-0,643239	-0,47653	-0,0581804	0,0206226
15	-2,61556	-1,39672	-0,737628	-0,729764	0,151438	-0,342991
16	-2,84039	-2,01583	-0,605112	-0,48337	-0,895518	1,09295
17	1,98277	-2,20593	2,20466	-0,264193	-1,28344	1,02096

Fonte: A Autora, Pesquisa de Campo, 2004.

No ajuste do modelo $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ verificou-se que a quinta e a sexta componentes não são significativamente importantes devido os seus valores-p serem maiores que 0,05, como mostra a tabela 4.6. Assim, estas variáveis não foram incluídas no modelo.

Tabela 4.6: Ajuste do primeiro modelo de regressão múltipla na classe 1 de apartamentos

Parâmetro	Estimativa	Erro Padrão	Estatística T	Valor-p
CONSTANTE	174412,0	3375,41	51,6712	0,0000
PCOMP_1	18607,3	1662,15	11,1947	0,0000
PCOMP_2	4386,74	1812,13	2,42077	0,0360
PCOMP_3	7100,19	2184,26	3,25061	0,0087
PCOMP_4	-23492,7	2499,16	-9,40024	0,0000
PCOMP_5	972,152	2941,69	0,330474	0,7479
PCOMP_6	-2079,22	3557,77	-0,584415	0,5719

Fonte: A Autora, Pesquisa de Campo, 2004.

As variáveis utilizadas e os respectivos valores dos coeficientes estão apresentados na tabela 4.7.

Tabela 4.7: Ajuste do modelo final de regressão múltipla da classe 1 de apartamentos

Parâmetro	Estimativa	Erro Padrão	Estatistica T	Valor-p
CONSTANTE	174412,0	3150,0	55,3689	0,0000
PCOMP_1	18607,3	1551,15	11,9958	0,0000
PCOMP_2	4386,74	1691,11	2,594	0,0235
PCOMP_3	7100,19	2038,4	3,48323	0,0045
PCOMP_4	-23492,7	2332,26	-10,0729	0,0000

Fonte: A Autora, Pesquisa de Campo, 2004.

O coeficiente de determinação do ajuste foi $R^2 = 0{,}956557$. A estatística R^2 indica que o modelo ajustado explica 95,6557% da variabilidade do valor de mercado. Portanto, a equação de Regressão Linear Múltipla para a primeira classe dos apartamentos e que descreve a relação entre valor e as quatro variáveis independentes é:

$$\textbf{Valor} = 174412{,}0 + 18607{,}3\,X_1 + 4386{,}74\,X_2 + 7100{,}19\,X_3 - 23492{,}7\,X_4$$

Em que:

X_1 = 1ª componente

X_2 = 2ª componente

X_3 =3ª componente

X_4 = 4ª componente

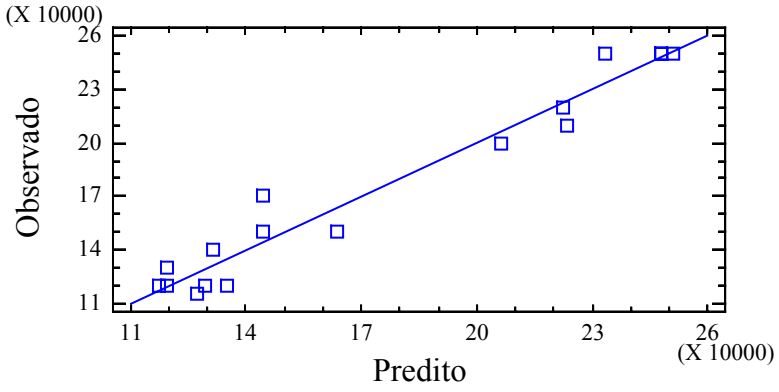
A Análise de Variância, contida na tabela 4.8, mostra que rejeita-se a hipótese de não haver regressão. Isto é, o modelo é significativo.

Tabela 4.8: Análise de variância do ajuste do modelo de regressão na classe 1 de apartamentos

Fonte	Soma dos Quadrados	G.L.	Quadrado Médio	F	Valor-p
Modelo	4,45699E10	4	1,11425E10	66,06	0,0000
Resíduo	2,02419E9	12	1,68682E8		
Total	4,65941E10	16			

Fonte: A Autora, Pesquisa de Campo, 2004.

Figura 4.2: Valores preditos *versus* valores observados na classe 1 de apartamentos



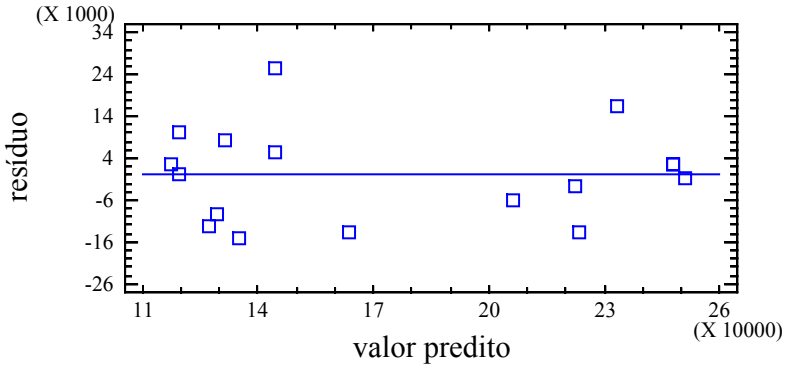
Fonte: A Autora, Pesquisa de Campo, 2004.

Como pode-se observar na figura 4.2 os pontos estão disposto em linha diagonal, indicando uma boa linearidade. Assim, as previsões se aproximam dos valores reais. A análise dos resíduos é apresentada a seguir.

A Independência Serial dos resíduos (não existir autocorrelação dos erros) foi comprovada pelo teste Durbin-Watson (DW), pois o valor-p é 0,1193 que é maior que 0,05. Assim, não há nenhuma indicação de autocorrelação entre os resíduos.

A homocedasticidade de variância (variância constante) foi verificada e a figura 4.3 apresenta os pontos distribuídos aleatoriamente em torno da linha que passa pela origem. Esta disposição dos pontos indica que a suposição de variância constante está correta.

Figura 4.3: Resíduos *versus* valores preditos na classe 1 de apartamentos



Fonte: A Autora, Pesquisa de Campo, 2004.

A Gaussianidade foi testada e o teste de *Kolmogorov* forneceu valor-p de 0,987074, maior que 0,05, o que indica que a distribuição dos resíduos é Gaussiana.

Os valores preditos pela equação ajustada e os valores observados e a percentagem de erro da predição são mostrados no quadro 4.1.

Quadro 4.1: Quadro de valores da classe 1 de apartamentos

Valor Observado (R$) (Y)	Valor Predito (R$) $\left(\hat{Y}\right)$	Resíduo (R$) $\left(\varepsilon\right)$	Erro Percentual (%)
130000	128715	1285	0,98846
150000	149891	109	0,07267
120000	117273	2727	2,2725
170000	170992	992	0,583529
250000	244826	5174	2,0696
220000	219705	295	0,13409
200000	198514	1486	0,743
250000	249772	228	0,0912
150000	146357	3643	2,42867
120000	127476	7476	6,23
250000	249481	519	0,2076
250000	249481	519	0,2076
115000	114222	778	0,67652
120000	128543	8543	7,119167
140000	142864	2864	2,045714
120000	109661	10339	8,61583
210000	217228	7228	3,441905

Fonte: A Autora, Pesquisa de Campo, 2004.

4.1.2 Classe 2 de Apartamentos

A tabela 4.9 mostra os autovalores e a percentagem variância explicada e acumulada. As seis primeiras componentes explicam 9,244% da variabilidade dos dados originais. Sendo que a primeira componente explica 40,382%, a segunda componente explica 21,955%, a terceira componente explica 16,570%, quarta componente explica 10,498%, a quinta componente explica 5,551% e a sexta componente explica 4,286%. Adotou-se como definição do número de componentes a serem tomadas, o Critério de Kaiser (Johnson e Wichern, 1998), mas incluiu-se a sexta componente devido ter o autovalor a 0,814385 próximo de um.

Tabela 4.9: Análise das componentes principais na classe 2 de apartamentos

Componente	Autovalor	% de Variância	Percentagem Acumulada
1	7,67267	40,382	40,382
2	4,17153	21,955	62,338
3	3,14831	16,570	78,908
4	1,99471	10,498	89,406
5	1,05474	5,551	94,958
6	0,814385	4,286	99,244
7	0,122288	0,644	99,888
8	0,0213724	0,112	100,000
9	7,25359E-16	0,000	100,000
10	5,17316E-16	0,000	100,000
11	2,60668E-16	0,000	100,000
12	1,92897E-16	0,000	100,000
13	8,70947E-17	0,000	100,000
14	0,0	0,000	100,000
15	0,0	0,000	100,000
16	0,0	0,000	100,000
17	0,0	0,000	100,000
18	0,0	0,000	100,000
19	0,0	0,000	100,000

Fonte: A Autora, Pesquisa de Campo, 2004.

A tabela 4.10, adiante, mostra os pesos das variáveis em cada uma das seis componentes.

Tabela 4.10: Pesos das componentes principais da classe 2 de apartamentos

variáveis	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
ppredio	0,121506	-0,159884	-0,223506	-0,157613	0,415464	-0,669934
elevador	0,240302	0,0847117	0,000145723	-0,463114	0,139136	0,302634
vgaragem	0,172032	-0,387555	0,109816	-0,0868311	0,133602	0,29816
acm2	0,254548	0,304926	-0,172346	0,0935485	0,0100245	0,00392387
pavimento	0,327342	-0,134889	-0,00770765	-0,156651	0,0636842	0,239366
andar	0,264228	-0,134445	0,105401	-0,0560174	0,484473	-0,23645
pecas	0,333358	-0,0736112	-0,0558262	0,231813	-0,0814017	0,0182491
sala	0,274362	0,203643	0,227794	0,12502	0,184269	0,134375
quarto	0,176043	-0,370227	0,160414	0,0789146	-0,281724	-0,127203
banheiro	0,282975	0,243096	-0,0452525	-0,18415	-0,225469	-0,117677
dempregados	0,187415	-0,231873	-0,363074	-0,171614	-0,132969	0,118611
descola	-0,248924	-0,322116	0,162516	-0,052034	0,0452559	0,0322505
dhospital	-0,155428	0,123612	0,417322	-0,286132	0,0656826	-0,0862526
dsmercado	-0,248924	-0,322116	0,162516	-0,052034	0,0452559	0,0322505
padraoc	-0,0412622	-0,330561	-0,372321	0,109761	0,13972	0,240059
revestimento	0,182678	0,0325761	0,428564	0,129288	0,282022	0,205096
conservação	-0,169002	0,141163	-0,280164	-0,45102	0,133757	0,178019
idreal	-0,106127	0,125374	-0,228446	0,497113	0,390723	0,186667
idaparente	-0,317108	0,149061	-0,0505335	-0,105682	0,291242	0,138737

Fonte: A Autora, Pesquisa de Campo, 2004.

A primeira componente possui pesos mais altos nas variáveis: quantidade de pavimentos, número de peças e idade aparente. A segunda componente possui pesos maiores nas variáveis: número de vagas para carro na garagem, área do apartamento, quantidade de quartos, distância à escola, distância ao supermercado e padrão de acabamento. A terceira componente possui pesos mais elevados nas variáveis: distância aos hospitais, revestimento do prédio e dependências de empregado. A quarta componente possui peso mais altos nas variáveis: número de elevadores, nível de conservação do apartamento e idade real. A quinta componente possui pesos mais altos nas variáveis: posição do apartamento (frente, fundo e lateral), andar do apartamento, nível de conservação, idade real e idade aparente do prédio. A sexta componente possui maiores pesos nas variáveis: posição do apartamento (frente, fundo e lateral) e número de elevadores.

Os escores fornecidos pelas seis componentes para os 10 apartamentos compõem a tabela 4.11 adiante. Esses são os valores da varáveis explicativas considerados para a Análise de Regressão.

Tabela 4.11: Componentes principais na classe 2 de apartamentos

imóvel	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
1	-2,80524	-0,257401	-2,82461	-0,292811	-0,896241	-0,164107
2	-2,51356	-0,258273	-2,29545	-0,685822	-0,232534	-0,469715
3	0,262396	-0,799519	-0,588756	-1,48931	1,94672	1,57887
4	2,31747	-1,48849	2,04693	-2,82213	-1,17288	-0,432653
5	5,4357	3,82429	-1,45618	0,295399	-0,13585	-0,0747495
6	1,77734	-2,61285	0,0743284	1,75323	-0,0594618	-0,00885276
7	1,77734	-2,61285	0,0743284	1,75323	-0,0594618	-0,00885276
8	-1,95574	1,23317	1,42087	0,329931	0,641174	-0,846152
9	-2,65331	1,8982	2,00267	0,894774	-1,24715	1,55276
10	-1,6424	1,07374	1,54586	0,263502	1,21569	-1,12655

Fonte: A Autora, Pesquisa de Campo, 2004.

As variáveis utilizadas e os respectivos valores dos coeficientes estão apresentados na tabela 4.12.

Tabela 4.12: Ajuste do modelo de regressão múltipla na classe 2 de apartamentos

Parâmetro	Estimativa	Erro Padrão	Estatística T	Valor-p
CONSTANTE	132389,0	1011,86	130,838	0,0000
PCOMP_1	17772,3	365,745	48,5922	0,0000
PCOMP_2	4451,33	525,658	8,4681	0,0035
PCOMP_3	-3115,71	580,034	-5,3716	0,0126
PCOMP_4	10597,3	769,832	13,7657	0,0008
PCOMP_5	-1079,08	1023,02	-1,0548	0,3690
PCOMP_6	-3433,46	1122,56	-3,05861	0,0551

Fonte: A Autora, Pesquisa de Campo, 2004.

O coeficiente de determinação do ajuste foi R^2 = 0,998942. A estatística R^2 indica que o modelo ajustado explica 99,8942% da variabilidade do valor de mercado. Portanto, a equação de Regressão Linear Múltipla, para a segunda classe dos apartamentos, para descrever a relação entre valor e as seis variáveis independentes é:

Valor = 133036,0 + 18814,5 X_1 + 2245,63 X_2 – 1733,73 X_3 + 11164,2 X_4 – 515,169 X_5 – 2907,38 X_6

Em que:

X_1 = 1ª componente

X_2 = 2ª componente

X_3 =3ª componente

X_4 = 4ª componente

X_5 = 5ª componente

X_6 = 6ª componente

A Análise de Variância, na tabela 4.13, mostra que rejeita-se a hipótese de não haver regressão. Isto é, o modelo é significativo.

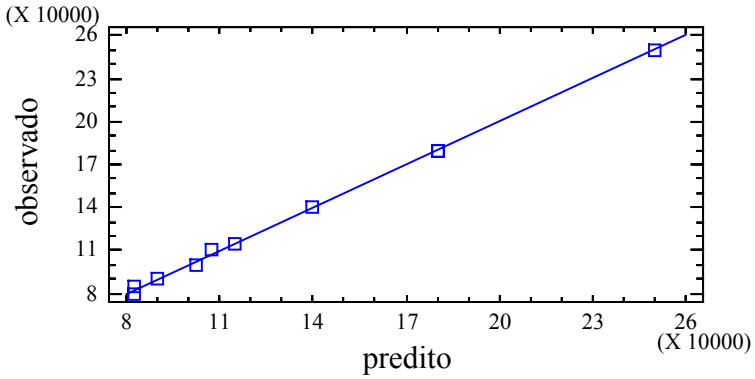
Tabela 4.13: Análise de variância do ajuste do modelo de regressão na classe 2 de apartamentos

Fonte	Soma dos Quadrados	G.L.	Quadrado Médio	F	Valor-p
Modelo	2,70314E10	6	4,50523E9	472,13	0,0002
Residual	2,86269E7	3	9,5423E6		

Fonte: A Autora, Pesquisa de Campo, 2004.

Observar-se na figura 4.4 pontos em linha diagonal, indicando uma boa linearidade, então, dessa forma as previsões se aproximam dos valores reais.

Figura 4.4: Valores preditos e valores observados da classe 2 de apartamentos



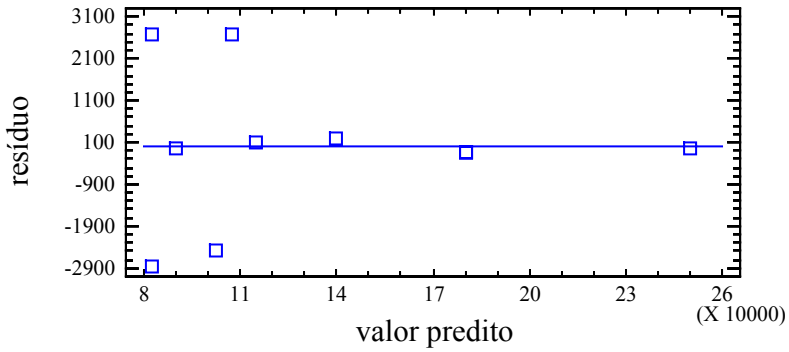
Fonte: A Autora, Pesquisa de Campo, 2004.

A seguir apresenta-se uma análise dos resíduos.

A independência serial dos resíduos (não existir autocorrelação dos erros) foi comprovada pelo teste de Durbin-Watson (DW), pois o valor-p é 0,1470 que é maior que 0,05. Assim, não há nenhuma indicação de autocorrelação entre os resíduos.

A homogeneidade de variância foi verificada e a figura 4.5 apresenta os pontos distribuídos aleatoriamente em torno da linha que passa pela origem. Esta disposição dos pontos indica que a suposição de variância constate está correta.

Figura 4.5: Resíduos versus valores preditos da classe 2 de apartamentos



Fonte: A Autora, Pesquisa de Campo, 2004.

Foi testado a Gaussianidade e o teste *Kolmogorov* forneceu valor-p de 0,899229 que é maior que 0,05 o que indica que a distribuição dos resíduos é Gaussiana.

Os valores preditos pela equação ajustada e os valores observados e os erros da predição são mostrados no quadro 4.2.

Quadro 4.2: Quadro de valores da classe 2 de apartamentos

Valor Observado (R$) (Y)	Valor Predito (R$) $\left(\hat{Y}\right)$	Resíduo (R$) $\left(\varepsilon\right)$	Erro Percentual (%)
85000	81807,1	3192,9	3,756353
80000	82836,1	2836,1	3,54513
115000	115891	891	0,77478
140000	139220	780	0,557143
250000	249969	31	0,0124
180000	180089	89	0,04944
180000	180089	89	0,04944
100000	102653	2653	2,653
90000	89414,4	585,6	0,650667
110000	108032	1968	1,789091

Fonte: A Autora, Pesquisa de Campo, 2004.

4.1.3 Classe 3 de Apartamentos

A tabela 4.14 mostra os autovalores e a percentagem de variância explicada e acumulada. As sete primeiras componentes explicam 88,949 % da variabilidade dos dados originais. Sendo que a primeira componente explicar 28,811%, a segunda componente explica 18,309%, a terceira componente explicar 15,285%, a quarta componente explica 8,414%, a quinta componente explica 6,794%, a sexta componente explica 6,058% e a sétima componente explica 5,279% da variabilidade das variáveis originais.

Tabela 4.14: Análise de componentes principais na classe 3 de apartamentos

Componente	Autovalor	% de Variância	Percentagem Acumulada
1	5,47402	28,811	28,811
2	3,47873	18,309	47,120
3	2,90424	15,285	62,405
4	1,59863	8,414	70,819
5	1,29088	6,794	77,613
6	1,15093	6,058	83,671
7	1,00296	5,279	88,949
8	0,723919	3,810	92,759
9	0,565309	2,975	95,735
10	0,322144	1,695	97,430
11	0,234006	1,232	98,662
12	0,146358	0,770	99,432
13	0,0674311	0,355	99,787
14	0,027785	0,146	99,933
15	0,00869023	0,046	99,979
16	0,0039856	0,021	100,000
17	2,19357E-16	0,000	100,000
18	0,0	0,000	100,000
19	0,0	0,000	100,000

Fonte: A Autora, Pesquisa de Campo, 2004.

A tabela 4.15, adiante, mostra os pesos das variáveis em cada uma das sete componentes.

Tabela 4.15: Pesos das componentes principais na classe 3 de apartamentos

Variável	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6	Componente 7
ppredio	0,0355161	0,146945	0,0922227	-0,160926	-0,603673	0,26609	0,491997
elevador	0,30127	-0,221678	-0,0112887	-0,229396	0,14114	0,385131	0,0959099
local	0,113992	-0,391631	0,082682	0,173764	-0,256599	-0,11936	-0,26283
acostruída	0,276261	0,167324	-0,151167	-0,0881305	-0,0398314	-0,515897	-0,076287
pavimento	0,327907	-0,110138	-0,0278141	0,219803	-0,0660019	0,427343	-0,110663
andar	0,163413	-0,119032	0,388078	-0,168676	0,356453	0,108438	0,097377
peças	0,358012	0,206807	0,133924	-0,088151	-0,0207633	-0,0712288	-0,114467
sala	0,267096	0,169209	-0,136379	-0,278656	0,245042	0,228425	-0,254274
quarto	0,1395	0,388767	0,226482	-0,145857	-0,0250215	-0,171251	-0,0497328
suite	0,389228	-0,0404886	-0,173169	-0,0613339	0,00793009	-0,194723	-0,0663769
banheiro	-0,092734	0,390924	0,267335	0,0808445	-0,0627842	0,0478683	-0,174613
dempregados	0,268066	0,107101	0,271189	0,0868636	-0,0755425	-0,086849	-0,0376019
descola	-0,0710845	0,00329606	0,348914	0,184936	0,479676	-0,118802	0,383484
dhospital	0,000977593	-0,365756	0,355174	-0,21724	0,0108902	-0,0622485	-0,14143
dsmercado	0,0316045	-0,345995	0,31039	0,159114	-0,229425	-0,215805	-0,139091
p.construção	0,198085	0,190392	0,197855	0,488175	-0,0797887	0,255137	-0,194445
revestimento	0,294519	-0,135735	0,06724	-0,267858	-0,174454	-0,15606	0,357968
conservação	0,247986	0,0393493	-0,0418319	0,464282	0,0786929	-0,119294	0,411767
idreal	0,199937	-0,152973	-0,39727	0,223349	0,141994	-0,0593419	0,116876

Fonte: A Autora, Pesquisa de Campo, 2004.

A primeira componente possui pesos mais altos nas variáveis: número de elevadores, número de pavimentos, número de peças, existência de suíte. A segunda componente possui pesos mais elevados nas variáveis: localização, número de quartos, número de banheiros, distância ao hospital e distância ao supermercado. A terceira componente possui pesos mais altos nas varáveis: andar do apartamento, distância ao supermercado, distância ao hospital, distância à escola e idade real. A quarta componente possui pesos maiores nas varáveis: padrão de acabamento e nível de conservação do apartamento. A quinta componente possui pesos mais altos nas seguintes variáveis: posição do apartamento no prédio (frente, fundo, lateral), andar do apartamento e distância à escola. A sexta componente possui pesos mais altos nas variáveis: número de elevadores, localização, número de pavimentos do prédio. A sétima componente possui pesos mais elevados nas variáveis: posição do apartamento no prédio, distância à escola, revestimento do prédio e nível de conservação.

Os escores fornecidos pelas sete componentes para os 17 apartamentos compõem a tabela 4.16. Esses são os valores das variáveis explicativas considerados para a Análise de Regressão.

Tabela 4.16: Componentes principais na classe 3 de apartamentos

imóvel	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6	Componente 7
1	4,20779	2,57019	2,22985	3,02843	-0,399689	1,1395	-0,756789
2	0,935506	0,897342	-2,97585	-1,13859	-1,21834	1,25397	-1,92766
3	-2,70094	-1,41964	-1,09475	2,08868	0,937125	-0,314036	0,270862
4	-2,42144	5,28681	-0,931834	-1,07796	1,28539	0,533092	1,02295
5	0,210679	0,175515	-1,66252	1,82722	-0,907468	-2,10561	1,19412
6	2,27063	-0,565144	1,25474	-1,22279	-0,168489	-0,439692	0,11332
7	1,00426	0,869945	-2,52735	0,251311	0,949092	0,202999	0,706466
8	2,02584	-1,43436	0,318826	-1,00396	0,334902	0,493558	0,969784
9	2,23134	-1,58405	0,806848	-1,21608	0,783154	0,629922	1,09224
10	2,74746	0,337432	1,46723	-0,932581	-0,0449293	-1,32826	0,406358
11	1,13215	-0,95827	-2,40867	-0,432365	-2,07192	-0,527416	-0,116081
12	0,528336	-1,45185	-0,58524	0,0435521	2,82677	-1,13804	-2,22212
13	-2,11076	1,14772	2,47339	-0,446867	-0,497754	-0,796154	-0,526714
14	-3,0197	-0,522425	1,02965	-0,11439	-0,680555	-0,413447	0,0431369
15	-3,22685	0,509006	1,66016	-0,59947	-0,954935	-0,491023	-0,964612
16	-2,20982	-2,2267	0,0163215	0,721693	-0,283903	1,75512	0,353981
17	-1,60447	-1,63152	0,929193	0,224159	0,11154	1,54552	0,34076

Fonte: A Autora, Pesquisa de Campo, 2004.

As variáveis utilizadas e os respectivos valores dos coeficientes estão apresentados na tabela 4.17.

Tabela 4.17: Ajuste do modelo de regressão múltipla na classe 3 de apartamentos

Parâmetro	Estimativa	Erro Padrão	Estatística T	valor-p
CONSTANTE	70705,9	2423,94	29,1698	0,0000
COMP_1	3470,01	1040,72	3,33425	0,0087
COMP_2	6673,2	1330,36	5,0161	0,0007
COMP_4	-1917,89	1933,92	-0,991712	0,3473
COMP_3	4096,52	1361,92	3,00791	0,0148
COMP_5	-7571,59	2198,29	-3,44431	0,0073
COMP_6	-6981,03	2319,38	-3,00987	0,0147
COMP_7	-7104,54	2492,96	-2,84984	0,0191

Fonte: A Autora, Pesquisa de Campo, 2004.

O coeficiente de determinação de ajuste foi $R^2 = 0{,}893306$. A estatística R^2 indica que o modelo como provido explica 89,3306% da variabilidade do valor. Assim, a equação de Regressão Linear Múltipla para a terceira classe de apartamentos para descrever a relação entre valor e seis variáveis independentes é:

Valor = 70705,9 + 3470,01 X_1 + 6673,2 X_2 – 1917,89 X_3 + 4096,52 X_4 – 7571,59 X_5 – 6981,03 X_6 – 7104,54 X_7

Em que:

X_1 = 1ª componente

X_2 = 2ª componente

X_3 = 3ª componente

X_4 = 4ª componente

X_5 = 5ª componente

X_6 = 6ª componente

X_7 = 7ª componente

A Análise de Variância, na tabela 4.18, mostra que rejeita-se a hipótese de não haver regressão. Isto é, o modelo é significativo.

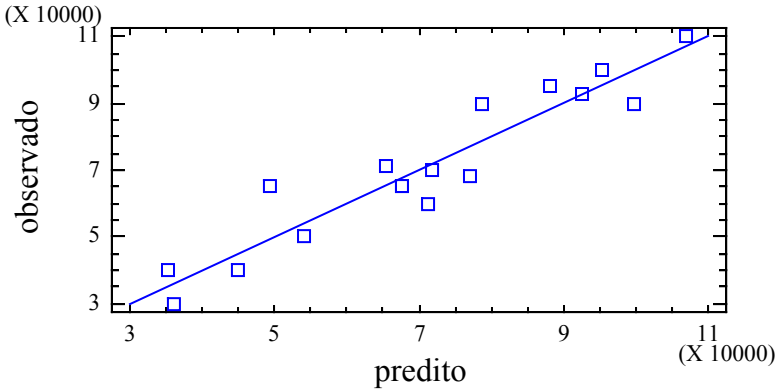
Tabela 4.18: Análise de variância do ajuste do modelo de regressão na classe 3 de apartamentos

Fonte	Soma dos Quadrados	G.L.	Quadrado Médio	F	Valor-p
Modelo	7,52658E9	7	1,07523E9	10,76	0,0010
Residual	8,98951E8	9	9,98835E7		
Total (Corr.)	8,42553E9	16			

Fonte: A Autora, Pesquisa de Campo, 2004.

A figura 4.6, adiante, apresenta pontos dispostos em linha diagonal, indicando uma boa linearidade, ou seja, as previsões se aproximam dos valores reais.

Figura 4.6: Valores preditos *versus* observados na classe 3 de apartamentos



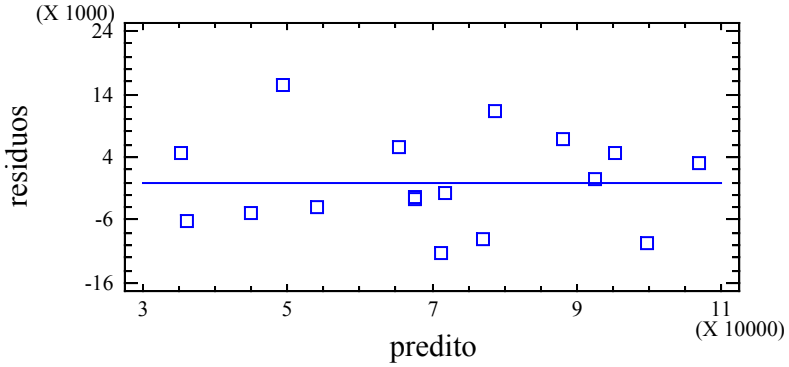
Fonte: A Autora, Pesquisa de Campo, 2004.

A análise dos resíduos é apresentada adiante.

A independência serial dos resíduos (não existir autocorrelação dos erros) foi comprovada pelo teste de Durbin-Watson (DW), pois o valor-p é 0,0692, que é maior que 0,05. Assim, não há nenhuma indicação de autocorrelação entre os resíduos.

A homogeneidade de variância foi verificada e a figura 4.7 apresenta os pontos distribuídos aleatoriamente em torno da linha que passa pela origem. Esta disposição dos pontos indica que a suposição de variância constate está correta.

Figura 4.7: Resíduos *versus* valores preditos na classe 3 de apartamentos



Fonte: A Autora, Pesquisa de Campo, 2004.

Foi testada a Gaussianidade e o teste de *Kolmogorov* forneceu valor-p de 0,110277 que é maior que 0,05, o que indica que a distribuição dos resíduos é Gaussiana.

Os valores preditos pela equação ajustada e os valores observados e o erro da predição são mostrados no quadro 4.3.

Quadro 4.3: Quadro de valores na classe 3 de apartamentos

Valor Observado (R$) (Y)	Valor Predito (R$) $\left(\hat{Y}\right)$	Resíduo (R$) $\left(\varepsilon\right)$	Erro Percentual (%)
110000	106931	3069	2,79
68000	77103,6	9103,6	13,3876
40000	35267,7	4732,3	11,83075
50000	54040,3	4040,3	8,0806
60000	71329,7	11329,7	18,8828
93000	92388,2	611,8	0,657849
65000	49436,3	15563,7	23,94415
65000	67527,4	2527,4	3,88831
71000	65530,7	5469,3	7,703239
90000	99614,9	9614,9	10,6832
90000	78691,3	11308,7	12,56522
65000	67623,4	2623,4	4,036
100000	95406,5	4593,5	4,5935
70000	71746,9	1746,9	2,49557
95000	88195,8	6804,2	7,162316
30000	36158,4	6158,4	20,528
40000	45008,1	5008,1	12,5203

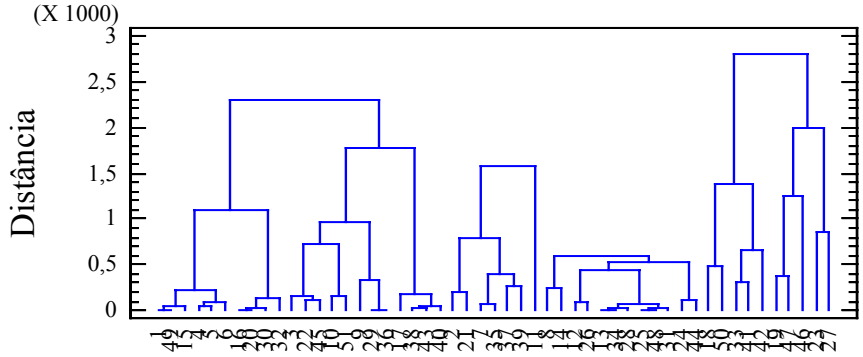
Fonte: A Autora, Pesquisa de Campo, 2004.

4.2 CASAS RESIDENCIAIS

Para formar as classes de casas residenciais homogêneas, a técnica de Análise de Agrupamento foi aplicada, começando-se com cada observação formando um grupo. Combinou-se os dois grupos com as observações que eram mais semelhantes para formar um grupo novo. Depois de recalculada a distância entre os grupos, os dois grupos então mais semelhantes foram combinados. Esse processo foi repetido até que quatro grupos permaneceram.

O dendrograma da figura 4.8 mostra como cada um dos agrupamentos foram formados.

Figura 4.8: Dendrograma de classes de casas residenciais



Fonte: A Autora, Pesquisa de Campo, 2004.

A primeira classe contém 43,14% das casas, a segunda classe contém 13,73%, a terceira classe contém 23,53% e a quarta classe contém 19,615% das residências analisadas, conforme a tabela 4.19.

Tabela 4.19: Resultado das classes formadas de residências

Cluster	Membros	Percentagem
1	22	43,14
2	7	13,73
3	12	23,53
4	10	19,61

Fonte: A Autora, Pesquisa de Campo, 2004.

Após a Análise de Agrupamento, realizou-se a Análise Discriminante e os resultados estão apresentados na tabela 4.20.

Tabela 4.20: Tabela de classificação para residências

clusters	Número de Imóvel		Clusters Preditos 1	2	3	4
1	22	20	0 (90,91%)	1 (0,00%)	1 (4,55%)	(4,55%)
2	7	0	7 (0,00%)	0 (100,00%)	0 (0,00%)	(0,00%)
3	12	1	0 (8,33%)	11 (0,00%)	0 (91,67%)	(0,00%)
4	10	1	0 (10,00%)	0 (0,00%)	9 (0,00%)	(90,00%)

Fonte: A Autora, Pesquisa de Campo, 2004.

Os resultados se apresentam muito consistentes. Entre as 51 observações que ajustam o modelo, 47 foram classificados corretamente, ou seja, 92,16% das casas foram classificadas corretamente. A interpretação das classes obtidas foi realizada levando-se em consideração a característica de cada classe. A seguir apresentam-se as características mais determinantes de cada uma das classes:

Classe 1: Residência com área construída entre 100 m^2 à 242 m^2 e com terreno entre (350 – 500) m^2, com valores acima de R$ 30000,00.

Classe 2: Residências que possuem garagem e lavanderia. Com área construída acima de 100 m^2 e com área do terreno acima de 446 m^2, valor entre R$ 60.000,00 e 180.000,00.

Classe 3: Residências com áreas construídas acima de 250m^2, com valor entre R$

30.000,00 e 150.000,00.

***Classe 4**:* Todos os imóveis são de alvenaria com cobertura de telha, com área construída entre $70m^2$ à $400m^2$, com valores acima de R$ 30000,00.

Com o objetivo de determinar um modelo de Regressão Linear Múltipla adequado a cada uma das quatro classes de casas residenciais foi realizado primeiramente a Análise de Componentes Principais com os dados das variáveis explicativas originais. Em seguida, com a obtenção dos escores das componentes principais, substituíram-se as variáveis originais pelas componentes principais (escores) realizou-se a regressão. Foi ajustado em seguida um modelo de Regressão Linear Múltipla para cada uma das quatro classes de apartamentos obtidas pela Análise de Agrupamento (*Cluster Analysis*). Considerou-se como variável resposta (dependente) o preço total de venda à vista, que foi denominada valor e como varáveis explicativas as componentes principais (escores).

Os resultados obtidos para cada uma das classes são apresentados a seguir.

4.2.1 Classe 1 de Casas Residenciais

A tabela 4.21 mostra os autovalores e a percentagem variância explicada e acumulada das seis componentes principais extraídas. As seis primeiras componentes explicam 81,561% da variabilidade dos dados originais. A primeira componente explica 34,434%, a segunda componente explica 13,481, a terceira componente explica 11,022%, a quarta componente explica 10,213%, a quinta componente explica 7,505% e a sexta componente explica 4,905%. Adotou-se como definição do número de componentes a serem tomadas, o Critério de Kaiser (Johnson e Wichern, 1998), porém incluiu-se a sexta componente devido ter o autovalor, 0,882831o próximo de um.

Tabela 4.21: Análise das componentes principais na classe 1 de residências

Componente	Autovalor	% de Variância	Percentagem Acumulada
1	6,19816	34,434	34,434
2	2,42653	13,481	47,915
3	1,98401	11,022	58,937
4	1,83839	10,213	69,151
5	1,35098	7,505	76,656
6	0,882831	4,905	81,561
7	0,729046	4,050	85,611
8	0,622797	3,460	89,071
9	0,516564	2,870	91,941
10	0,446232	2,479	94,420
11	0,307089	1,706	96,126
12	0,250013	1,389	97,515
13	0,19718	1,095	98,610
14	0,113167	0,629	99,239
15	0,0841657	0,468	99,706
16	0,0321228	0,178	99,885
17	0,0132998	0,074	99,959
18	0,00742567	0,041	100,000

Fonte: A Autora, Pesquisa de Campo, 2004.

A tabela 4.22, adiante, mostra os pesos das variáveis em cada uma das seis componentes.

Tabela 4.22: Tabela de pesos das componentes principais na classe 1 de residências

variável	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
localização	0,18402	0,323946	0,265447	0,278936	-0,138594	0,283284
garagem	0,149367	0,44763	0,0717797	0,143562	-0,0402839	-0,145844
suíte	0,344845	-0,155134	0,0774941	-0,020109	-0,24014	-0,108086
banheiro	0,139523	-0,126192	0,456053	-0,14664	0,188031	-0,274159
edícula	0,29732	0,116019	-0,00455696	0,0718954	-0,258986	0,307287
smercado	-0,104064	0,123382	0,45277	0,268462	-0,297913	0,100159
área const.	0,286589	-0,0376579	-0,141976	0,331362	-0,148964	-0,100998
área do ter.	0,22538	0,0817627	0,135006	-0,337851	0,176318	0,572052
p de contrução	0,264964	0,0610122	-0,268291	0,0718392	0,323453	0,114408
cobertura	0,337659	-0,217798	0,0978793	-0,0830461	-0,0236547	-0,114697
estrutura	0,303517	-0,179914	0,146365	-0,118055	-0,0879624	-0,33778
econs	0,0576373	-0,435626	-0,0810838	0,253058	-0,147396	0,335389
piscina	0,175548	0,0949346	-0,502539	-0,0781235	-0,0192746	0,0153232
quarto	0,0126595	0,214667	0,135013	0,392244	0,592833	-0,111485
dempr	0,313162	0,16018	-0,11429	0,0286577	-0,175147	-0,268903
lavand	0,189501	0,194593	0,179698	-0,519606	0,0855429	0,0858198
peças	0,341514	-0,00542479	-0,0743738	0,146295	0,296004	0,0158858
idapa	0,109682	-0,475356	0,189143	0,190492	0,254876	0,155938

Fonte: A Autora, Pesquisa de Campo, 2004.

Através dos resultados obtidos na tabela 4.22 pode-se observar que a primeira componente possui pesos mais altos nas variáveis: presença de suíte, tipo de cobertura, estrutura, dependências de empregado e número de peças. A segunda componente possui pesos mais altos nas varáveis: localização, presença de garagem, nível de conservação e idade aparente. A terceira componente tem pesos maiores nas variáveis: quantidade de banheiro, distância ao supermercado e presença de piscina. A quarta componente possui pesos mais altos nas variáveis: área construída, área do terreno, quantidade de quarto e presença de lavanderia. A quinta componente tem pesos maiores nas variáveis: padrão de construção e quantidade de quarto. A sexta componente tem pesos mais alto nas variáveis: presença de edícula, área do terreno, estrutura e nível de conservação.

Os escores fornecidos pelas seis componentes das 22 residências compõem a tabela 4.23 adiante. Esses são os valores das variáveis explicativas considerados para a Análise de Regressão.

Tabela 4.23: Escores das componentes principais na classe 1 de residências

imóvel	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
1	1,1195	1,36622	1,81638	-0,663497	-2,08947	-0,128076
2	1,02301	0,780476	1,28479	-0,231682	0,848372	-2,5159
3	-0,22076	-0,71108	-0,0522917	-2,17947	0,424518	-1,15144
4	1,04312	-0,928564	0,229272	-0,691072	-0,556219	0,80852
5	0,995388	0,0893067	1,8554	0,671277	-0,419405	1,39255
6	-4,66171	-2,02479	-1,67147	1,88563	-1,37953	0,0317062
7	-3,50872	0,515953	-1,28424	1,81979	0,709338	-0,230405
8	2,16289	0,199196	0,0577154	0,353459	-0,770028	0,846527
9	2,75542	0,544711	-1,04278	0,649919	0,697833	0,235513
10	-1,22808	-2,1061	0,347781	-1,31896	2,28644	0,235642
11	2,90972	0,769056	-2,764	-0,468245	-0,719443	0,356949
12	-2,68358	1,0615	0,648711	0,0721607	0,709106	-0,0712971
13	-0,493412	-1,97412	0,325907	3,34067	-0,182259	-0,360158
14	2,85294	-0,643921	0,435001	0,63009	1,03945	0,0765788
15	4,22624	1,0795	-1,86294	0,50014	1,56511	0,50212
16	1,40978	0,499633	1,1804	0,41472	-2,44538	-1,22378
17	-3,75898	2,20471	0,0233164	-1,5831	-0,467005	1,0893
18	-2,76882	2,90907	0,613496	-0,0192542	0,617695	1,18746
19	-0,939124	-3,88111	0,273226	-2,51343	-0,50548	0,682386
20	0,889937	-0,149351	-2,72748	-1,09129	-1,03774	-0,759284
21	1,66366	-0,80862	2,53063	0,908304	0,902225	0,282195
22	-2,78842	1,20832	-0,21683	-0,486162	0,771871	-1,2871

Fonte: A Autora, Pesquisa de Campo, 2004.

As variáveis utilizadas e os respectivos valores dos coeficientes estão apresentados na tabela 4.24.

Tabela 4.24: Ajuste do modelo de regressão múltipla na classe 1 de residências

Parâmetro	Estimativa	Erro Padrão	Estatística T	Valor-p
CONSTANTE	90954,5	2448,38	37,1489	0,0000
PCOMP_1	17279,3	1006,58	17,1663	0,0000
PCOMP_2	3478,72	1608,75	2,16238	0,0472
PCOMP_3	-3362,4	1779,13	-1,88991	0,0783
PCOMP_4	5162,11	1848,25	2,79297	0,0137
PCOMP_5	-6686,66	2156,04	-3,10136	0,0073
PCOMP_6	-4568,88	2667,11	-1,71304	0,1073

Fonte: A Autora, Pesquisa de Campo, 2004.

O coeficiente de determinação do ajuste foi R^2 = 0,91937, o que significa que 91,937% do valor de mercado foi explicado pelo modelo.

Portanto, a equação de Regressão Linear Múltipla para a primeira classe de casas residenciais para descrever a relação entre valor e as seis variáveis independentes é:

Valor = 91181,8 + 17686,6 X_1 + 3605,89 X_2 – 3878,05 X_3 + 5047,61 X_4 – 4,17 X_5 – 4627,55 X_6

Em que:

X_1 = 1ª componente

X_2 = 2ª componente

X_3 =3ª componente

X_4 = 4ª componente

X_5 = 5ª componente

X_6 = 6ª componente

A Análise de Variância, na tabela 4.25, mostra que rejeita-se a hipótese de não haver regressão. Isto é, o modelo é significativo.

Tabela 4.25: Análise de variância do ajuste do modelo de regressão na classe 1 de residências

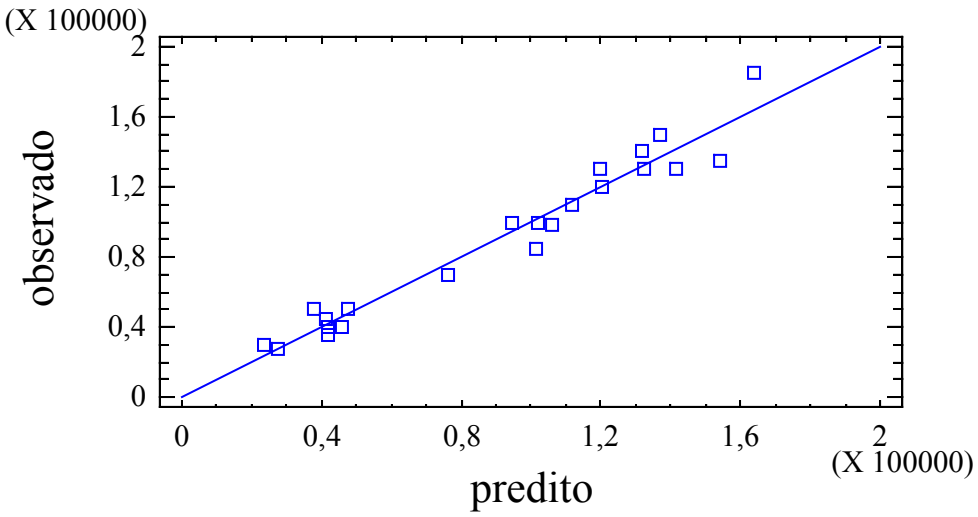
Análise de Variância

Fonte	Soma dos Quadrados	Graus de Liberdade	Quadrado Médio	F	P
Regressão	4,40629E10	6	7,34382E9	28,51	0,0000
Resíduo	3,86437E9	15	2,57625E8		
Total	4,79273E10	21			

Fonte: A Autora, Pesquisa de Campo, 2004.

Como pode se observar na figura 4.9, adiante, apresenta os pontos em linha diagonal, indicando uma boa linearidade, indicando dessa forma previsão que se aproximam dos valores reais.

Figura 4.9: Valores preditos *versus* valores observados na classe 1 de residências



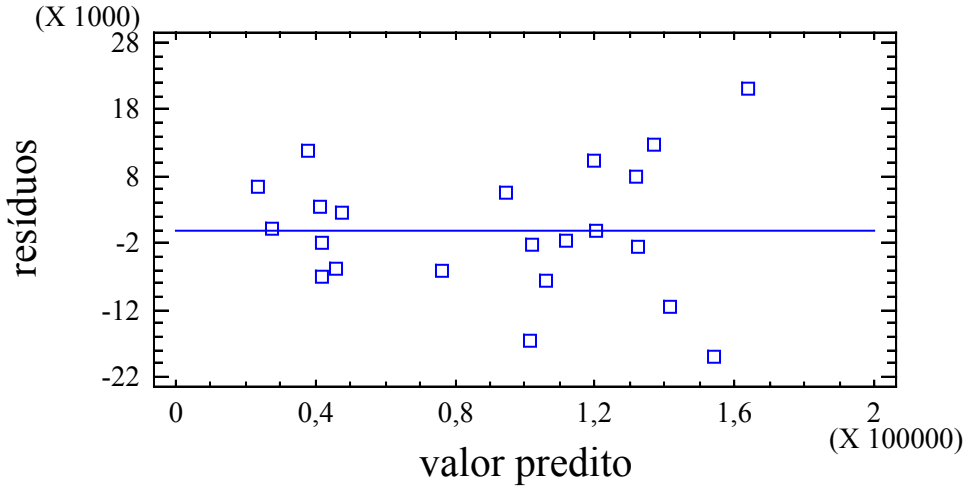
Fonte: A Autora, Pesquisa de Campo, 2004.

É apresentada adiante a análise dos resíduos.

A independência serial dos resíduos (não existir autocorrelação dos erros) foi comprovada pelo teste estatístico Durbin-Watson (DW), pois o valor-p é 0,4250 que é maior que 0,05. Assim, não há nenhuma indicação de autocorrelação consecutivo entre os resíduos.

A homogeneidade de variância foi verificada e a figura 4.10 apresenta os pontos distribuídos aleatoriamente em torno da linha que passa pela origem. Esta disposição dos pontos indica que a suposição de variância constate está correta.

Figura 4.10: Resíduos versus valores preditos na classe 1 de residências



Fonte: A Autora, Pesquisa de Campo, 2004.

A Gaussianidade foi testada e o teste *Kolmogorov– Smirnov* forneceu valor-p de 0,987074, que é maior que 0,05 o que indica que a distribuição dos resíduos é Gaussiana.

Os valores preditos pela equação ajustada e os valores observados e os erros de predição estão no quadro 4.4.

Quadro 4.4: Quadro de valores na classe 1 de residências

Valor Observado (R$) (Y)	Valor Predito (R$) $\left(\hat{Y}\right)$	Resíduo (R$) $\left(\varepsilon\right)$	Erro Percentual (%)
120000	120076	76	0,06333
110000	111653	1653	1,50273
70000	76013,7	6013,7	8,591
85000	101436	16436	19,3365
100000	102133	2133	2,133
28000	27793,4	206,6	0,737857
35000	42142,9	7142,9	20,4083
140000	131932	8068	5,762857
130000	141580	11580	8,90769
50000	38064,4	11935,6	23,8712

135000	153964	18964	14,0474
40000	42052,3	2052,3	5,13075
100000	94574,6	5425,4	5,4254
130000	132501	2501	1,92385
185000	163823	21177	11,44703
150000	137167	12833	8,555333
30000	23566,9	6433,1	21,44367
45000	41513,2	3486,8	7,748444
50000	47594,8	2405,2	4,8104
130000	119758	10242	7,878462
98000	105746	7746	7,90408
40000	45914,8	5914,8	14,787

Fonte: A Autora, Pesquisa de Campo, 2004.

4.2.2 Classe 2 de Casas Residenciais

A tabela 4.26 mostra os autovalores e a percentagem variância explicada e acumulada das cinco componentes extraídas. As cinco primeiras componentes explicam 99,451% da variabilidade nos dados originais. A primeira componente explica 50,040%, a segunda componente explica 22,141%, a terceira componente explica 14,946%, a quarta componente explica 9,621% e a quinta componente explica 2,703%. Adotou-se como definição do número de componentes a serem tomadas, o Critério de Kaiser (Johnson e Wichern, 1998). No entanto incluiu-se a quinta componente devido obtenção de melhores resultados de regressão.

Tabela 4.26: Análise das componentes principais na classe 2 de residências

Componente	Autovalor	% de Variância	Percentagem Acumulada
1	8,00642	50,040	50,040
2	3,54259	22,141	72,181
3	2,39143	14,946	87,128
4	1,5393	9,621	96,748
5	0,43242	2,703	99,451
6	0,0878467	0,549	100,000
7	7,89321E-16	0,000	100,000
8	4,71004E-16	0,000	100,000
9	2,75845E-16	0,000	100,000
10	1,72561E-16	0,000	100,000
11	6,05346E-17	0,000	100,000
12	0,0	0,000	100,000
13	0,0	0,000	100,000
14	0,0	0,000	100,000
15	0,0	0,000	100,000
16	0,0	0,000	100,000

Fonte: A Autora, Pesquisa de Campo, 2004.

A tabela 4.27, adiante, mostra os pesos das variáveis em cada uma das cinco componentes.

Tabela 4.27: Pesos das componentes principais na classe 2 de residências

Variável	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5
localização	0,273473	0,309819	0,0833471	-0,155982	-0,106903
suíte	-0,299599	-0,121542	0,0792291	-0,358987	-0,188782
banheiro	-0,172411	0,158024	0,519089	-0,116013	-0,00208855
edícula	-0,243542	0,277872	-0,303758	0,0218007	-0,261705
smercado	0,266821	0,309658	-0,153315	-0,108718	-0,160088
m2c	-0,234792	0,145806	0,257802	0,445539	-0,0417591
m2t	-0,0559348	0,396756	-0,00678435	0,49992	-0,244163
p de contrução	-0,14639	0,265509	-0,389595	-0,106225	0,676891
cobertura	-0,325089	0,175897	0,0952425	-0,0202968	0,0819694
estrutura	-0,328146	0,168557	0,0988402	-0,00930642	-0,126782
econs	-0,273452	0,0709998	-0,384792	0,103368	-0,170174
piscina	-0,224816	-0,245172	-0,37143	0,184347	0,0158622
quarto	-0,0930335	0,463859	0,13933	-0,223519	0,313487
dempr	-0,266821	-0,309658	0,153315	0,108718	0,160088
peças	-0,337617	-0,0793574	0,132224	-0,0244275	0,156128
idapa	-0,251039	0,0120396	-0,134124	-0,503776	-0,372811

Fonte: A Autora, Pesquisa de Campo, 2004.

De acordo com a tabela 4.27, pode-se observar que a primeira componente possui pesos mais altos nas variáveis: tipo de cobertura, estrutura (alvenaria, madeira) e número de

peças. A segunda componente possui pesos mais altos nas varáveis: localização, quantidade de quarto e dependência de empregados, distância ao supermercado e área do terreno. A terceira componente tem pesos maiores nas variáveis: padrão de construção, nível de conservação, presença de edícula, quantidade de banheiro e presença de piscina A quarta componente possui pesos mais altos nas variáveis: quantidade de banheiro, área do terreno, área construída e idade aparente. A quinta componente possui pesos mais altos nas variáveis: padrão de construção, quantidade de quarto e idade aparente.

Os escores fornecidos pelas cinco componentes para as sete residências compõem a tabela 4.28 adiante. Esses são os valores das variáveis explicativas considerados para a Análise de Regressão.

Tabela 4.28: Componentes Principais na classe 2 de residências

Imóvel	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5
1	-2,17244	-1,24199	3,08778	-0,153567	0,187116
2	-1,4386	1,83075	-0,465762	-2,26217	-0,46448
3	-0,733039	3,29598	0,15041	1,72628	0,0859842
4	-4,08194	-1,96967	-2,01436	0,643518	0,0155551
5	2,62226	-0,777984	0,1484	0,527854	-0,499642
6	3,30913	-0,99663	-0,248129	0,150119	-0,626207
7	2,49464	-0,140456	-0,65834	-0,63203	1,30167

Fonte: A Autora, Pesquisa de Campo, 2004.

As variáveis utilizadas e os respectivos valores dos coeficientes estão apresentados na tabela 4.29.

Tabela 4.29: Primeiro ajuste do modelo de regressão múltipla da classe 2 de residências

Parâmetro	Estimativa	Erro Padrão	Estatística T	Valor-p
CONSTANTE	142857,0	18830,2	7,58661	0,0169
PCOMP_1	-8663,42	7188,01	-1,20526	0,3514
PCOMP_2	-1439,41	10806,1	-0,133204	0,9062
PCOMP_3	1918,05	13152,2	0,145834	0,8974
PCOMP_4	5471,07	16393,3	0,333738	0,7703

Fonte: A Autora, Pesquisa de Campo, 2004.

O coeficiente de determinação foi $R^2 = 0{,}576526$. Significando que apenas 57,6526% do valor de mercado está sendo explicado pelo modelo. Com o propósito de melhores resultados, substitui-se a segunda componente, por não ser significativa, pela quinta componente.

Tabela 4.30: Ajuste final do modelo de regressão da classe 2 de residências

Parâmetro	Estimativa	Erro Padrão	Estatística T	Valo-p
CONSTANTE	142857,0	2228,4	64,1074	0,0002
PCOMP_1	-8663,43	850,645	-10,1846	0,0095
PCOMP_3	1918,05	1556,46	1,23232	0,3430
PCOMP_4	5471,08	1940,02	2,82012	0,1061
PCOMP_5	-43628,8	3660,28	-11,9195	0,0070

Fonte: A Autora, Pesquisa de Campo, 2004.

O coeficiente de determinação foi $R^2 = 0{,}99226$, cujo significado é que 99,226% do valor de mercado está sendo explicado pelo modelo. Portanto, a equação de Regressão Linear Múltipla, para a classe 2 das residenciais, para descrever a relação entre valor e as quatro variáveis independentes é:

$$\textbf{Valor} = 142857{,}0 - 8663{,}43\, X_1 + 1918{,}05\, X_2 + 5471{,}08\, X_3 - 43628{,}8\, X_4$$

Em que:

$X_1 =$ 1ª componente

$X_2 =$ 3ª componente

$X_3 =$ 4ª componente

$X_4 =$ 5ª componente

A Análise de variância, tabela 4.31, mostra que rejeita-se a hipótese de não haver regressão. Isto é, o modelo é significativo.

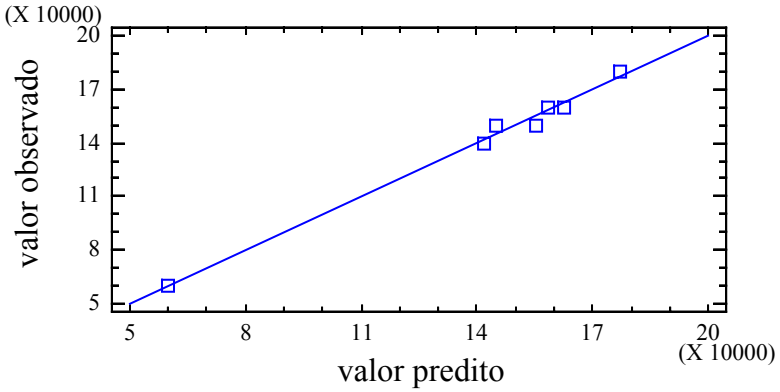
Tabela 4.31: Análise de variância do ajuste do modelo de regressão na classe 2 de residências

Fonte	Soma dos Quadrados	G.L.	Quadrado Médio	F	Valor-p
Modelo	8,87334E9	4	2,21833E9	63,82	0,0155
Resíduo	6,95209E7	2	3,47604E7		
Total	8,94286E9	6			

Fonte: A Autora, Pesquisa de Campo, 2004.

A figura 4.11 adiante mostra que os pontos estão dispostos em linha diagonal, indicando, assim, uma boa linearidade. Logo, as previsões se aproximam dos valores reais.

Figura 4.11: Valores preditos versus valores observados da classe 2 de residências



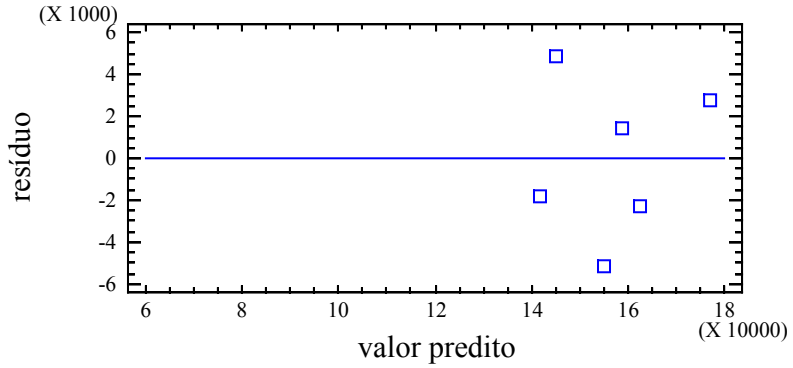
Fonte: A Autora, Pesquisa de Campo, 2004.

A seguir tem-se a análise dos resíduos.

A independência serial dos resíduos (não existir autocorrelação dos erros) foi comprovada pelo teste de Durbin-Watson (DW), pois o valor-p é 0,1367 que é maior que 0,05. Assim, não há nenhuma indicação de autocorrelação entre os resíduos.

A homogeneidade de variância foi verificada e a figura 4.12 apresentou os pontos distribuídos aleatoriamente em torno da linha que passa pela origem. Esta disposição dos pontos indica que a suposição de variância constate está correta.

Figura 4.12: Resíduos *versus* valor predito na classe 2 de residências



Fonte: A Autora, Pesquisa de Campo, 2004.

Foi testada a Gaussianidade e o teste de *Kolmogorov- Smirnov* forneceu valor-p de 0,999516 que é maior que 0,05 o que indica que a distribuição dos resíduos é Gaussiana.

Os valores preditos pela equação ajustada e os valores observados e o erro da predição são mostrados na quadro 4.5.

Quadro 4.5: Quadro de valores na classe 2 de residências

Valor Observado (R$) (Y)	Valor Predito (R$) $\left(\hat{Y}\right)$	Resíduo (R$) $\left(\varepsilon\right)$	Erro Percentual (%)
160000	158597	1.403,00	0,876875
160000	162315	2.315,00	1,44688
150000	155189	5.189,00	3,45933
180000	177199	2.801,00	1,556111
150000	145111	4.889,00	3,259333
140000	141855	1.855,00	1,325
60000	59734,1	265,90	0,443167

Fonte: A Autora, Pesquisa de Campo, 2004.

4.2.3 Classe 3 de Casas Residenciais

A tabela 4.32 mostra os autovalores e a percentagem variância explicada e acumulada das seis componentes principais extraídas. As seis primeiras componentes explicam 89,181% da variabilidade nos dados originais. Sendo que a primeira componente explica 30,529%, a segunda componente explica 19,079%, a terceira componente explica 15,297%, a quarta componente explica 12,012%, a quinta componente explica 6,952 % e a sexta componente explica 5,311%. Adotou-se como definição do número de componentes a serem tomadas, o Critério de Kaiser. Porém, incluiu-se a sexta componente devido seu autovalor 0,902825 ser muito próximo de um.

Tabela 4.32: Análise das componentes principais na classe 3 de residências

Componente	Autovalor	% de Variância	Percentagem Acumulada
1	5,19	30,529	30,529
2	3,2435	19,079	49,609
3	2,60043	15,297	64,905
4	2,04211	12,012	76,918
5	1,1819	6,952	83,870
6	0,902825	5,311	89,181
7	0,709032	4,171	93,352
8	0,446268	2,625	95,977
9	0,348611	2,051	98,028
10	0,291148	1,713	99,740
11	0,044175	0,260	100,000
12	2,52235E-16	0,000	100,000
13	3,84988E-17	0,000	100,000
14	1,61874E-17	0,000	100,000
15	0,0	0,000	100,000
16	0,0	0,000	100,000
17	0,0	0,000	100,000

Fonte: A Autora, Pesquisa de Campo, 2004.

A tabela 4.33, adiante, mostra os pesos das variáveis em cada uma das seis componentes.

Tabela 4.33: Pesos das componentes principais na classe 3 de residências

Variável	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
localização	0,0298988	-0,140115	0,438094	0,200778	-0,272926	0,514726
garagem	0,094771	0,181138	0,137755	0,144618	0,69589	0,328312
suíte	0,361336	0,0345366	0,164003	0,111941	-0,00184351	0,0966769
banheiro	-0,124205	0,50282	-0,0148528	-0,0535591	-0,0999089	0,129793
edícula	0,263148	0,0355433	0,398508	-0,259574	-0,0763141	-0,18009
smercado	-0,281975	-0,262098	-0,146321	0,153609	-0,157732	0,143978
acostruída	-0,0473542	0,382193	0,0641632	0,332712	-0,234552	-0,0519501
aterreno	-0,192042	-0,0265296	0,332111	0,411074	0,0430758	-0,367991
p de contrução	0,236081	0,138261	-0,267254	-0,157422	-0,463918	0,296058
cobertura	0,354361	-0,17077	-0,191895	0,0943447	-0,0120738	-0,268745
estrutura	0,358521	-0,0444914	-0,260894	0,0447077	0,0359642	-0,291783
conservação	0,305037	0,18165	-0,193841	0,156453	0,271793	0,174719
quarto	-0,214085	0,319751	-0,208384	-0,306137	0,0904683	-0,0313843
dempregada	0,263148	0,0355433	0,398508	-0,259574	-0,0763141	-0,18009
lavandeiria	0,0408309	0,186533	-0,147268	0,549599	-0,154269	-0,174587
peças	0,0700327	0,49916	0,140177	-0,0250207	-0,101246	-0,132932
idaparente	0,362924	-0,0780252	-0,0994425	0,169644	-0,0714291	0,229541

Fonte: A Autora, Pesquisa de Campo, 2004.

De acordo com a tabela 4.33 pode-se observar que a primeira componente possui pesos mais altos nas variáveis: presença de suíte, tipo de cobertura, estrutura, nível de conservação e idade aparente. A segunda componente possui pesos mais altos nas varáveis: quantidade de banheiro, área construída, quantidade de quarto e quantidade de peças. A terceira componente tem pesos maiores nas variáveis: localização, presença de edícula, área do terreno e dependência de empregado. A quarta componente possui pesos mais altos nas variáveis: área total do terreno, área construída, presença de lavanderia e quantidade de quarto. A quinta componente possui pesos mais altos nas variáveis: presença de garagem e padrão de construção. A sexta componente possui pesos mais elevados nas variáveis: localização, área do terreno e vaga para garagem.

Os escores fornecidos pelas seis componentes das 12 residências compõem a tabela 4.34 adiante. Esses são os valores das variáveis explicativas considerados para a Análise de Regressão.

Tabela 4.34: Escores das componentes principais na classe 3 de residências

imóvel	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Component 6
1	-2,46874	-1,65073	0,361759	-1,507	1,99923	-0,0645304
2	-2,04696	5,1788	-0,122646	-0,347307	-0,374963	0,372098
3	-0,142077	0,190293	-1,07182	-0,701539	0,514516	-0,432815
4	-1,79145	-1,81337	0,557637	-1,06285	-1,86204	0,552594
5	1,3635	-0,922768	-1,43082	1,29238	0,3631	0,673587
6	-0,314579	-0,702248	-2,09146	-0,201655	-1,65959	-1,82174
7	-1,4549	-0,627888	0,432111	-0,483436	-0,351472	1,71189
8	4,33679	0,366078	3,29066	-1,68322	-0,286411	-0,516292
9	2,03685	-0,553154	-0,294257	1,83523	-0,499463	1,0179
10	-2,7775	-0,269423	2,61394	2,88347	0,348473	-1,07993
11	2,95603	0,744365	-0,874708	0,985235	0,775034	-0,0673845
12	0,303027	0,0600426	-1,37041	-1,0093	1,03358	-0,345379

Fonte: A Autora, Pesquisa de Campo, 2004.

As variáveis utilizadas e os respectivos valores dos coeficientes estão apresentados na tabela 4.35.

Tabela 4.35: Ajuste do modelo de regressão múltipla na classe 3 de residências

Parâmetro	Estimativa	Erro Padrão	Estatística T	Valor-p
CONSTANTE	59000,0	3411,69	17,2935	0,0000
COMPP_1	3359,31	1564,16	2,14769	0,0845
COMPP_2	1233,0	1978,59	0,623173	0,5605
COMPP_3	15364,2	2209,74	6,95295	0,0009
COMPP_4	17691,4	2493,59	7,09476	0,0009
COMPP_5	5654,93	3277,72	1,72526	0,1451
COMPP_6	-3422,73	3750,26	-0,912666	0,4033

Fonte: A Autora, Pesquisa de Campo, 2004.

O coeficiente de determinação foi R^2 = 0,955551. A estatística R^2 indica que o modelo ajustado explica 95,5551% da variabilidade do valor de mercado. Portanto, a equação de Regressão Linear Múltipla para a terceira classe de residências para descrever a relação entre valor e as seis variáveis independentes é:

Valor = 59000,0 + 3359,31 X_1 + 1233 X_2 + 15364,2 X_3 + 17691,4 X_4 + 5654,93 X_5 – 3422,73 X_6

Em que:

X_1 = 1ª componente

X_2 = 2ª componente

X_3 = 3ª componente

X_4 = 4ª componente

X_5 = 5ª componente

X_6 = 6ª componente

A Análise de Variância, na tabela 4.36, mostra que rejeita-se a hipótese de não haver regressão. Isto é, o modelo é significativo.

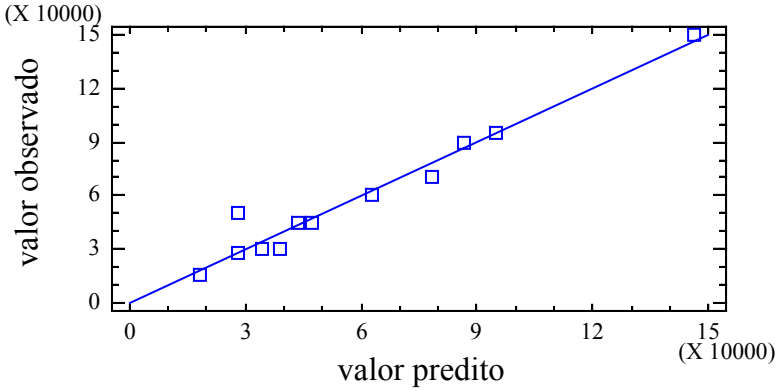
Tabela 4.36: Análise de variância do ajuste do modelo de regressão na classe 3 de residências

Fonte	Soma dos Quadrados	G.L.	Quadrado Médio	F	Valor-p
Modelo	1,50136E10	6	2,50227E9	17,91	0,0031
Resíduo	6,98377E8	5	1,39675E8		
Total	1,5712E10	11			

Fonte: A Autora, Pesquisa de Campo, 2004.

A figura 4.13, adiante, apresenta pontos distribuídos em linha diagonal, indicando uma boa linearidade. Dessa forma as previsões se aproximam dos valores reais.

Figura 4.13: Valores preditos *versus* valores observados na classe 3 de residências



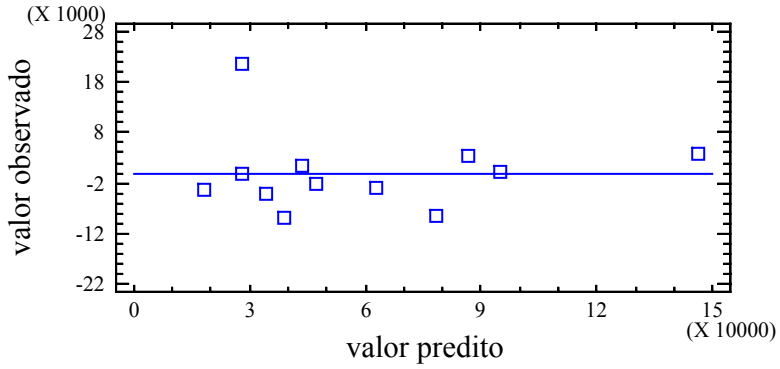
Fonte: A Autora, Pesquisa de Campo, 2004.

A análise dos resíduos é apresentada a seguir.

A independência serial dos resíduos (não existir autocorrelação dos erros) foi comprovada pelo teste de Durbin-Watson (DW), pois o valor-p é 0,3915 que é maior que 0,05. Assim, não há nenhuma indicação de autocorrelação entre os resíduos.

A homogeneidade de variância foi verificada e a figura 4.14 apresenta os pontos distribuídos aleatoriamente em torno da linha que passa pela origem. Esta disposição dos pontos indica que a suposição de variância constate está correta.

Figura 4.14: Resíduos versus valores preditos na classe 3 de residências



Fonte: A Autora, Pesquisa de Campo, 2004.

Foi testada a Gaussianidade e o teste de *Kolmogorov* forneceu valor-p de 0,552759 que é maior que 0,05 o que indica que a distribuição dos resíduos é Gaussiana.

Os valores preditos pela equação ajustada e os valores observados e o de erro de predição são mostrados no quadro 4.6.

Quadro 4.6: Quadro de valores da classe 3 de residências

Valor Observado (R$) (Y)	Valor Predito (R$) $\left(\hat{Y}\right)$	Resíduo (R$) $\left(\varepsilon\right)$	Erro Percentual (%)
30000	39095	9094,97	30,3167
45000	47086,4	2086,43	4,63644
30000	34269,5	4269,46	14,2317
28000	28089,3	89,3417	0,31893
60000	63071	3071,04	5,11833
15000	18226,7	3226,66	21,5113
45000	43577,8	1422,15	3,160444
95000	94947,4	52,606	0,055368
90000	86798,71	3201,29	3,557
150000	146178	3822,08	2,548
70000	78452,46	8452,44	12,0749
50000	28207,8	21792,2	43,5844

Fonte: A Autora, Pesquisa de Campo, 2004.

4.2.4 Classe 4 de Casas Residenciais

A tabela 4.37 apresenta os autovalores e a percentagem variância explicada e acumulada das seis componentes principais extraídas. As seis primeiras componentes explicam 90,8% da variabilidade nos dados originais. Sendo que a primeira componente explica 34,484%, a segunda componente explica 16,016%, a terceira componente explica 14,124%, a quarta componente explica 10,724%, a quinta componente explica 8,610% e a sexta componente explica 6,841%. Adotou-se como definição do número de componentes a serem tomadas, o Critério de Kaiser.

Tabela 4.37: Análise das componentes principais na classe 4 de residências

Componente	Autovalor	% de Variância	Percentagem Acumulada
1	5,86236	34,484	34,484
2	2,72264	16,016	50,500
3	2,40114	14,124	64,624
4	1,82308	10,724	75,348
5	1,46377	8,610	83,959
6	1,16303	6,841	90,800
7	0,974142	5,730	96,530
8	0,326857	1,923	98,453
9	0,262992	1,547	100,000
10	5,0014E-16	0,000	100,000
11	3,04424E-16	0,000	100,000
12	1,49798E-16	0,000	100,000
13	0,0	0,000	100,000
14	0,0	0,000	100,000
15	0,0	0,000	100,000
16	0,0	0,000	100,000
17	0,0	0,000	100,000

Fonte: A Autora, Pesquisa de Campo, 2004.

A tabela 4.38, adiante, mostra os pesos das variáveis em cada uma das seis componentes.

Tabela 4.38: Pesos das componentes principais na classe 4 de residências

Variável	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
local	0,214988	0,0548955	-0,306678	0,193162	0,364254	0,295502
garagem	0,276184	-0,235289	-0,0496349	0,00533634	0,158592	-0,0887286
suíte	0,358834	-0,139239	-0,0134131	-0,15182	0,273172	0,0602231
banheiro	0,231365	0,384722	-0,171708	0,0684011	0,213946	-0,259821
edícula	0,175579	0,346158	0,039604	-0,273241	0,155258	0,292288
smercado	0,0516159	-0,0819704	0,146812	0,514963	-0,140112	0,540941
area const	0,337167	0,0405617	-0,26755	0,176054	-0,0850243	-0,108817
área ter	-0,246698	-0,391058	-0,241311	-0,0269492	0,198993	-0,0561792
acabamento	0,276792	-0,149382	0,275969	0,082672	0,102782	-0,411449
cobertura	0,213562	-0,433222	0,244103	0,109952	0,0840168	0,193219
estrutura	0,216097	0,186716	0,409834	0,322895	-0,0613574	-0,153984
conservação	0,00186308	-0,0792463	0,450057	-0,0448755	0,357905	0,125007
quarto	0,260944	0,000720961	-0,223285	0,0464551	-0,511898	0,231104
d empregado	0,172492	-0,285192	0,174259	-0,345919	-0,436047	-0,0934758
lavanderia	0,258168	-0,164827	-0,139487	-0,458329	0,0766971	0,239429
peças	0,392114	0,0813916	-0,0359162	0,0039213	-0,144348	-0,147271
id aparente	0,00210101	0,356152	0,337128	-0,325983	-0,0531898	0,231307

Fonte: A Autora, Pesquisa de Campo, 2004.

De acordo com a tabela 4.38, observou-se que a primeira componente possui pesos mais altos nas variáveis: presença de suíte, área construída e número de peças. A segunda componente possui pesos mais altos nas varáveis: quantidade de banheiro, presença de edícula, área do terreno, tipo de cobertura e idade aparente. A terceira componente tem pesos maiores nas variáveis: local do imóvel, estrutura, nível de conservação e idade aparente. A quarta componente possui pesos mais altos nas variáveis: distância ao supermercado, estrutura, dependência de empregada, presença de lavanderia e idade aparente. A quinta componente possui pesos mais altos nas variáveis: local, nível de conservação, quantidade de quartos e dependência de empregado. A sexta componente possui pesos mais altos nas variáveis: presença de edícula, distância ao supermercado e o padrão de acabamento.

Os escores fornecidos pelas seis componentes para as dez residências compõem a tabela 4.39 adiante. Esses são os valores das variáveis explicativas considerados para o ajustamento do modelo regressão.

Tabela 4.39: Escores das componentes principais na classe 4 de residências

imóvel	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5	Componente 6
1	-0,178996	-2,42095	0,376732	-1,40361	0,536502	-1,75053
2	1,50744	2,87986	-2,4722	-1,27789	0,618713	-0,873337
3	1,81973	-1,55724	-2,00885	0,555566	-2,0396	0,800763
4	0,809522	-1,02282	-1,4137	1,11555	1,09638	0,503211
5	-0,436962	-0,19928	0,485193	-0,53929	1,7394	2,08598
6	-4,60801	1,8232	0,339193	-0,027688	-0,660687	0,293695
7	-0,345381	-1,04724	1,17634	-1,73445	-0,0932097	0,179294
8	3,40433	1,4167	2,40042	-0,348447	-1,37011	0,265213
9	1,40297	0,512388	1,33385	2,58227	1,02929	-0,944804
10	-3,37465	-0,384623	-0,216977	1,07799	-0,856677	-0,559482

Fonte: A Autora, Pesquisa de Campo, 2004.

No ajuste do modelo $Y = X\underline{\beta} + \underline{\varepsilon}$ verificou-se que a segunda componente não é significativamente importante devido ao valor-p ser muito maior que 0,05, como mostra a tabela 4.40. Assim, esta variável não foi incluída no modelo.

Tabela 4.40: Primeiro ajuste do modelo de regressão múltipla da classe 4 de residências

Parâmetro	Estimativa	Erro Padrão	Estatística T	Valor-p
CONSTANTiva	137000,0	11016,9	12,4354	0,0011
PCOMP_1	24375,8	4796,27	5,08224	0,0147
PCOMP_2	565,154	7037,92	0,0803012	0,9411
PCOMP_3	7678,47	7494,3	1,02457	0,3810
PCOMP_4	28744,7	8600,75	3,34212	0,0443
PCOMP_5	16018,0	9598,51	1,6688	0,1937
PCOMP_6	-25110,7	10768,2	-2,33192	0,1020

Fonte: A Autora, Pesquisa de Campo, 2004.

As variáveis utilizadas e os respectivos valores dos coeficientes estão apresentados na tabela 4.41.

Tabela 4.41: Ajuste final de regressão múltipla na classe 4 de residências

Parâmetro	Estimativa	Erro Padrão	Estatística T	Valor-p
CONSTANTE	137000,0	9551,19	14,3438	0,0001
PCOMP_1	24375,8	4158,15	5,86217	0,0042
PCOMP_3	7678,47	6497,23	1,18181	0,3027
PCOMP_4	28744,7	7456,47	3,855	0,0182
PCOMP_5	16018,0	8321,48	1,9249	0,1266
PCOMP_6	-25110,7	9335,59	-2,68978	0,0547

Fonte: A Autora, Pesquisa de Campo, 2004.

O coeficiente de determinação do ajuste foi R^2 = 0,969181%. A estatística R^2 indica que o modelo como provido explica 96,9181% da variabilidade do valor. Portanto, a equação de Regressão Linear Múltipla para a quarta classe de casas para descrever a relação entre valor e cinco variáveis independentes é:

$$\textbf{Valor} = 136000{,}0 + 22695{,}4\ X_1 + 5349{,}07\ X_2 + 25348{,}7\ X_3 + 13114{,}7\ X_4 - 22964{,}2\ X_5$$

Em que:

X_1 = 1ª componente

X_2 = 3ª componente

X_3 = 4ª componente

X_4 = 5ª componente

X_5 = 6ª componente

A Análise de Variância, tabela 4.42, mostra que rejeita-se a hipótese de não haver regressão. Isto é, o modelo é significativo.

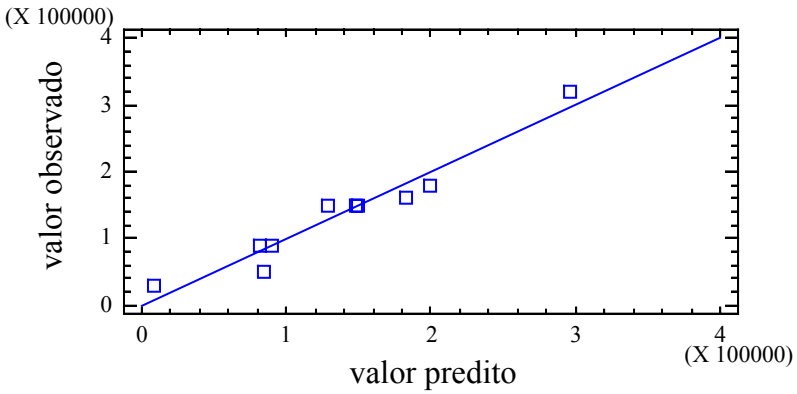
Tabela 4.42: Análise de variância no ajuste do modelo de regressão na classe 4 de residências

Fonte	Soma dos Quadrados	G.L.	Quadrado Médio	F	Valor-p
Modelo	5,6161E10	5	1,12322E10	12,31	0,0153
Resíduo	3,64901E9	4	9,12253E8		
Total	5,981E10	9			

Fonte: A Autora, Pesquisa de Campo, 2004.

Pode-se observar na figura 4.15 apresenta pontos distribuídos em linha diagonal, indicando uma boa linearidade, ou seja, as previsões que se aproximam dos valores reais.

Figura 4.15: Valores preditos versus valores observados na classe 4 de residências



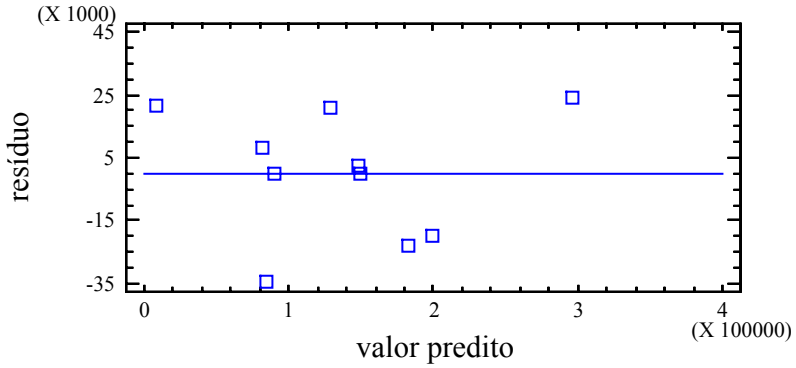
Fonte: A Autora, Pesquisa de Campo, 2004.

A análise dos resíduos é apresentada a seguir.

A independência serial dos resíduos (não existir autocorrelação dos erros) foi comprovada pelo teste de Durbin-Watson (DW), pois o valor-p é 0,0973634 que é maior que 0,05. Assim, não há nenhuma indicação de autocorrelação entre os resíduos.

A homogeneidade de variância foi verificada e a figura 4.16 apresentou os pontos distribuídos aleatoriamente em torno da linha que passa pela origem. Esta disposição dos pontos indica que a suposição de variância constante está correta.

Figura 4.16: Resíduos *versus* valores preditos da classe 4 de residências



Fonte: A Autora, Pesquisa de Campo, 2004.

Foi testada a Gaussianidade e o teste de *Kolmogorov- Smirnov* forneceu valor-p de 0,985558 que é maior que 0,05, o que indica que a distribuição residual é Gaussiana.

Os valores preditos pela equação ajustada e os valores observados e os resíduos são mostrados no quadro 4.7.

Quadro 4.7: Quadro de valores da classe 4 de residências

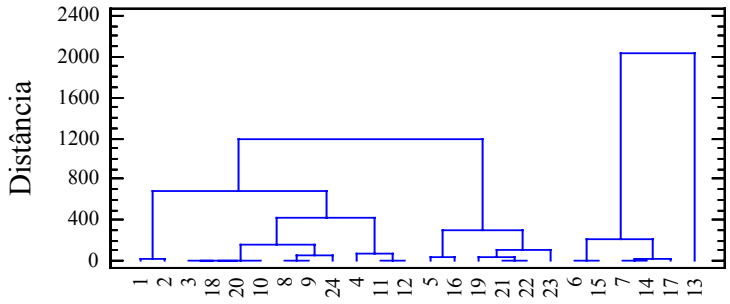
Valor Observado (R$) (Y)	Valor Predito (R$) $\left(\hat{Y}\right)$	Resíduo (R$) $\left(\varepsilon\right)$	Erro Percentual (%)
150000	145609	4391,32	2,927333
150000	152765	2764,82	1,84333
150000	135499	14500,8	9,667333
165000	177911	12911,1	7,82485
90000	89916,9	83,0967	0,092333
30000	17122,8	12877,2	42,924
90000	85147,9	4852,08	5,391222
180000	193211	13210,9	7,33944
290000	275629	14371,4	4,955517
65000	87189,1	22189,1	34,1371

Fonte: A Autora, Pesquisa de Campo, 2004.

4.3 TERRENOS

Para formar as classes de terrenos com características semelhantes, o método multivariado de Análise de Agrupamento foi aplicado, começando-se com cada observação em um grupo separado. Juntou-se os dois grupos em que as observações eram mais semelhantes para formar um grupo novo. Depois de recalculada a distância entre os grupos, os dois grupos então semelhantes foram combinados. Esse processo foi repetido até que dois grupos permaneceram. O dendrograma, figura 4.17, mostra como cada um dos agrupamentos foi formado.

Figura 4.17: Dendrograma das classes de terrenos formadas



Fonte: A Autora, Pesquisa de Campo, 2004.

A classe 1 contém 75% e a classe 2 contém 25% dos terrenos analisadas, conforme a tabela 4.43.

Tabela 4.43: Resultado das classes formadas de terrenos

Classe	Membros	Percentagem
1	18	75,00
2	6	25,00

Fonte: A Autora, Pesquisa de Campo, 2004.

Após a Análise de Agrupamento, realizou-se a Análise Discriminante e os resultados estão apresentados na tabela 4.44 adiante.

Tabela 4.44: Classificação para terrenos

cluster atual	quantidade de imóveis	Cluster predito 1	2
1	18	18 (100,00%)	0 (0,00%)
2	6	0 (0,00%)	6 (100,00%)

Percentagem de casos classificados corretamente: 100,00%

Fonte: A Autora, Pesquisa de Campo, 2004.

Os resultados se apresentam consistentes ao agrupamento, 100% dos terrenos foram classificados corretamente. A interpretação das classes obtidas foi realizada levando-se em consideração a característica de cada classe. A seguir apresentam-se as características mais determinantes de cada uma das classes:

Classe 1: todos os terrenos possuem área menor que 700*m* e os seus valores variam entre R$ 12000,00 e R$ 145000,00.

Classe 2: todo os terrenos possui frente superior à 15m, possui área superior à 900*m*, são todos planos e os seus valores variam R$ 90000,00 e R$ 300000,00.

Com o objetivo de determinar um modelo de Regressão Linear Múltipla adequado a cada uma das duas classes de terrenos foi realizado primeiramente a Análise de Componentes Principais com os dados das variáveis explicativas originais. Em seguida, com a obtenção dos escores das componentes principais, substituiu-se as variáveis originais pelas componentes principais (escores) e realizou-se a regressão. Foi ajustado em seguida um modelo de Regressão Linear Múltipla para cada uma das duas classes de terrenos obtidas pela Análise de Agrupamento (*Cluster*). Considerou-se como variável resposta (dependente) o preço total de venda à vista, que foi denominada valor e como variáveis explicativas as componentes principais (escores). Os resultados obtidos para cada uma das classes são apresentados a seguir.

4.3.1 Classe 1 de Terrenos

A tabela 4.45 mostra os autovalores e a percentagem variância explicada e acumulada das cinco componentes principais extraídas. As cinco primeiras componentes explicam 89,221% da variabilidade dos dados originais. Sendo que a primeira componente explica 32,368%, a segunda componente explica 18,203%, a terceira componente explica 14,337%, a quarta componente explica 13,689% e a quinta componente explica 10,625 %. Adotou-se como definição do número de componentes a serem tomadas, o Critério de Kaiser (Johnson e Wichern, 1998).

Tabela 4.45: Análise das componentes principais na classe 1 de terrenos

Componente	autovalor	% de variância	% Acumulada
1	3,23679	32,368	32,368
2	1,82035	18,203	50,571
3	1,43365	14,337	64,908
4	1,36889	13,689	78,597
5	1,06247	10,625	89,221
6	0,449579	4,496	93,717
7	0,334287	3,343	97,060
8	0,152748	1,527	98,588
9	0,119135	1,191	99,779
10	0,0221061	0,221	100,000

Fonte: A Autora, Pesquisa de Campo, 2004.

A tabela 4.46, adiante, mostra os pesos das variáveis em cada uma das cinco componentes.

Tabela 4.46: Pesos das componentes principais na classe 1 de terrenos

variável	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5
localização	0,41786	0,177003	0,333343	-0,036582	-0,295612
ponto com.	0,213762	0,196754	0,236597	-0,590969	-0,375902
polo	0,438963	0,0477081	0,368336	0,193581	0,132636
frente	-0,258231	0,492805	0,00776802	0,316699	-0,308742
área	0,160302	0,303966	-0,476754	-0,513236	0,0294181
proteção	0,163916	0,47532	-0,24219	0,00877233	0,542305
plano	-0,393541	0,0000184093	0,477734	-0,254653	0,148276
inclinado	0,313106	-0,203386	-0,370926	0,27352	-0,449515
posição	-0,265613	0,539121	0,0344659	0,222118	-0,253869
pavimento	0,37987	0,185699	0,210281	0,243227	0,273748

Fonte: A Autora, Pesquisa de Campo, 2004.

Com os resultados obtidos na tabela 4.46, pode-se observar que a primeira componente possui pesos mais altos nas variáveis: localização, pólo valorizante, inclinação e existência de pavimentação. A segunda componente possui pesos mais altos nas varáveis: comprimento frontal (frente), área total do terreno, existência de proteção (muro) e posição na quadra (meio ou esquina). A terceira componente tem pesos maiores nas variáveis: localização, pólo valorizante, área total do terreno e inclinação. A quarta componente possui pesos mais altos nas variáveis: ponto comercial, comprimento frontal (frente) e área do

terreno. A quinta componente tem pesos maiores nas variáveis: ponto comercial, comprimento frontal (frente), existência de proteção e inclinação.

Os escores fornecidos pelas cinco componentes dos 18 terrenos compõem a tabela 4.47 adiante. Esses são os valores das variáveis explicativas considerados para a Análise de Regressão.

Tabela 4.47: Escores das componentes principais da classe 1 de terrenos

imóvel	Componente 1	Componente 2	Componente 3	Componente 4	Componente 5
1	2,91759	1,32568	-1,15365	-1,36729	-1,07184
2	-1,53888	2,38977	-2,31731	-0,0695958	0,273227
3	1,37487	0,0656474	-0,289017	-0,351298	1,28221
4	0,00143538	0,16029	0,120141	0,385298	1,38413
5	-0,515911	1,85711	1,1905	1,79485	0,00648569
6	1,92518	-0,47488	-1,44966	0,471298	0,571883
7	-0,293815	2,644	0,998322	-0,0512669	-0,265436
8	1,44699	-0,0599322	2,05123	-2,92285	-1,09841
9	-3,66788	-0,344302	-0,79127	-0,674226	-1,07488
10	-3,09158	-1,51404	-0,866052	-1,15616	-0,524056
11	0,362523	-1,56812	1,66246	0,5792	0,256003
12	1,26587	-0,25043	-0,127966	-0,357132	0,921593
13	-0,719597	1,14271	1,30691	1,38254	-1,05275
14	1,91666	-1,23661	-0,844575	0,316819	-0,704078
15	-2,03461	-1,163	0,134286	-0,0271534	0,300861
16	-0,72504	-0,931414	1,22142	0,400543	0,410381
17	1,56936	-1,61508	-0,872922	2,118	-1,73026
18	-0,193164	-0,427388	0,0271553	-0,47158	2,11494

Fonte: A Autora, Pesquisa de Campo, 2004.

As variáveis significativas e os respectivos valores dos coeficientes estão apresentados na tabela 4.48.

Tabela 4.48: Ajuste do modelo de regressão múltipla da classe 1 de terrenos

Parâmetro	Estimativa	Erro Padrão	Estatística T	Valor-p
CONSTANTE	45388,9	1376,92	32,9641	0,0000
COMPP_1	10288,2	787,522	13,064	0,0000
COMPP_2	7916,81	1050,13	7,5389	0,0000
COMPP_3	10912,0	1183,31	9,22162	0,0000
COMPP_4	-15008,9	1210,98	-12,394	0,0000
COMPP_5	-10333,7	1374,55	-7,51786	0,0000

Fonte: A Autora, Pesquisa de Campo, 2004.

O coeficiente de determinação do ajuste foi $R^2 = 0, 977556$. Indicando, assim, que o modelo como provido explica 97,7556% da variabilidade do valor. Portanto, a equação de Regressão Linear Múltipla para a primeira classe de terrenos para descrever a relação entre valor e cinco variáveis independentes é:

Valor = 46222,2 + 10708,5 X_1 + 8440,44 X_2 + 9641,95 X_3 – 15023,2 X_4 – 0036,3 X_5

Em que:

X_1 = 1ª componente

X_2 = 2ª componente

X_3 = 3ª componente

X_4 = 4ª componente

X_5 = 5ª componente

A Análise de Variância, tabela 4.49, mostra que rejeita-se a hipótese de não haver regressão. Isto é, o modelo é significativo.

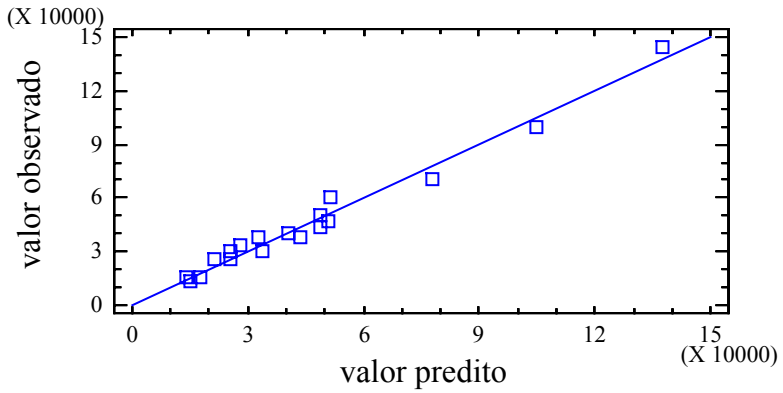
Tabela 4.49: Análise de variância do ajuste do modelo de regressão na classe 1 de terrenos

Fonte	soma dos quadrados	G.L.	Quadrado Médio	F	Valor-p
Modelo	1,78368E10	5	3,56735E9	104,53	0,0000
Residual	4,09514E8	12	3,41262E7		
Total	1,82463E10	17			

Fonte: A Autora, Pesquisa de Campo, 2004.

A figura 4.18 adiante apresenta os pontos em linha diagonal, indicando uma boa linearidade, ou seja, as previsões se aproximam dos valores reais.

Figura 4.18: Valores observados *versus* valores preditos na classe 1 de terrenos



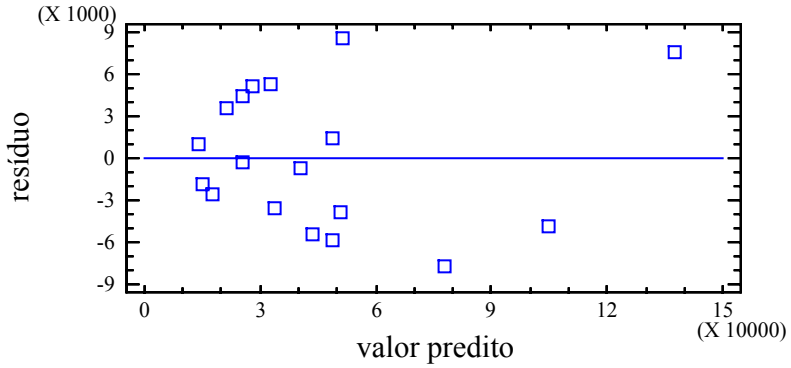
Fonte: A Autora, Pesquisa de Campo, 2004.

A seguir é apresentada a análise dos resíduos.

A independência serial dos resíduos (não existir autocorrelação dos erros) foi comprovada pelo teste de Durbin-Watson (DW), pois o valor-p é 0,315 que é maior que 0,05. Assim, não há nenhuma indicação de autocorrelação entre os resíduos.

A homogeneidade de variância foi verificada e a figura 4.19 apresenta os pontos distribuídos aleatoriamente em torno da linha que passa pela origem. Esta disposição dos pontos indica que a suposição de variância constate está correta.

Figura 4.19: Resíduos *versus* valores preditos na classe 1 de terrenos



Fonte: A Autora, Pesquisa de Campo, 2004.

Foi testada a Gaussianidade e o teste de *Kolmogorov- Smirnov* forneceu valor-p de 0,803714, maior que 0,05, o que indica que a distribuição dos resíduos é Gaussiana.

Os valores preditos pela equação ajustada e os valores observados e o erro da predição são mostrados no quadro 4.8.

Quadro 4.8: Quadro de valores na classe 1 de terrenos

Valor Observado (R$) (Y)	Valor Predito (R$) $\left(\hat{Y}\right)$	Resíduo (R$) $\left(\varepsilon\right)$	Erro Percentual (%)
100000	104910	4910,00	4,91
25000	21410,6	3589,40	14,3576
43000	48922,4	5922,40	13,773
33000	27897,6	5102,40	15,46182
40000	40768,6	768,60	1,9215
38000	32633,9	5366,10	14,12132
70000	77704,2	7704,20	11,006
145000	137404	7596,00	5,238621
15000	17519,8	2519,80	16,7987
13000	14913,5	1913,50	14,7192
38000	43506,3	5506,30	14,4903
47000	50870,1	3870,10	8,23426
60000	51421,6	8578,40	14,29733
50000	48622,4	1377,60	2,7552
15000	14013,1	986,90	6,579333
30000	33631,4	3631,40	12,1047
25000	25314,3	314,30	1,2572
30000	25537,1	4462,90	14,87633

Fonte: A Autora, Pesquisa de Campo, 2004.

4.3.2 Classe 2 de Terrenos

A tabela 4.50 mostra os autovalores e a percentagem variância explicada e acumulada das duas componentes principais extraídas. As duas primeiras componentes explicam 89,068% da variabilidade dos dados originais. Sendo que a primeira componente explica 45,844% e a segunda componente explica 43,224%. Adotou-se como definição do número de componentes a serem tomadas, o Critério de Kaiser (Johnson e Wichern, 1998).

Tabela 4.50: Análise das componentes principais na classe 2 de terrenos

Componente	Autovalor	% de Variância	Percentagem Acumulada
1	2,75066	45,844	45,844
2	2,59342	43,224	89,068
3	0,576617	9,610	98,678
4	0,079307	1,322	100,000
5	1,81172E-16	0,000	100,000
6	0,0	0,000	100,000

Fonte: A Autora, Pesquisa de Campo, 2004.

A tabela 4.51, adiante, mostra os pesos das variáveis em cada uma das duas componentes.

Tabela 4.51: Tabela de pesos das componentes principais na classe 2 de terrenos

variável independente	Componente 1	Componente 2
localização	0,437385	0,412552
área comercial	0,575146	-0,00652354
área do terreno	-0,017062	-0,615392
proteção	-0,13465	0,594668
plano	0,455555	0,0931079
posição	0,501945	-0,297913

Fonte: A Autora, Pesquisa de Campo, 2004.

Com os dados da tabela 4.51, pode-se observar que a primeira componente possui pesos mais altos nas variáveis: localização, pertence à área comercial, plano e a posição do

terreno na quadra (meio ou esquina). A segunda componente possui pesos mais elevados nas varáveis: presença de proteção (muro), localização e a área do terreno.

Os escores fornecidos pelas duas componentes para os seis terrenos compõem a tabela 4.52 adiante. Esses são os valores das variáveis explicativas considerados para a Análise de Regressão.

Tabela 4.52: Escores das componentes principais na classe 2 de terrenos

imóvel	Componente 1	Componente 2
1	-0,592268	0,878942
2	-1,59444	0,618472
3	-0,105032	-3,27717
4	-1,59444	0,618472
5	1,32835	0,66839
6	2,55783	0,492894

Fonte: A Autora, Pesquisa de Campo, 2004.

As variáveis significativas e os respectivos valores dos coeficientes estão apresentados na tabela 4.53.

Tabela 4.53: Ajuste do modelo de regressão múltipla na classe 2 de terrenos

Parâmetro	Estimativa	Erro Padrão	Estatística T	Valor-p
CONSTANTE	161667,0	2421,74	66,7564	0,0000
PCOMP_1	52279,6	1599,56	32,6838	0,0001
PCOMP_2	17091,4	1647,33	10,3752	0,0019

Fonte: A Autora, Pesquisa de Campo, 2004.

O coeficiente de determinação de ajuste R^2 = 0,997455. A estatística R^2 indica que o modelo como provido explica 99,7455% da variabilidade do valor. Assim, a equação de regressão Linear Múltipla para a segunda classe de terrenos para descrever a relação entre valor e duas variáveis independentes é:

Valor = 165000,0 + 51529,4 X_1 + 18528,3 X_2

Em que:

X_1 = 1ª componente

X_2 = 2ª componente

A Análise de Variância, tabela 4.54, mostra que rejeita-se a hipótese de não haver regressão. Isto é, o modelo é significativo.

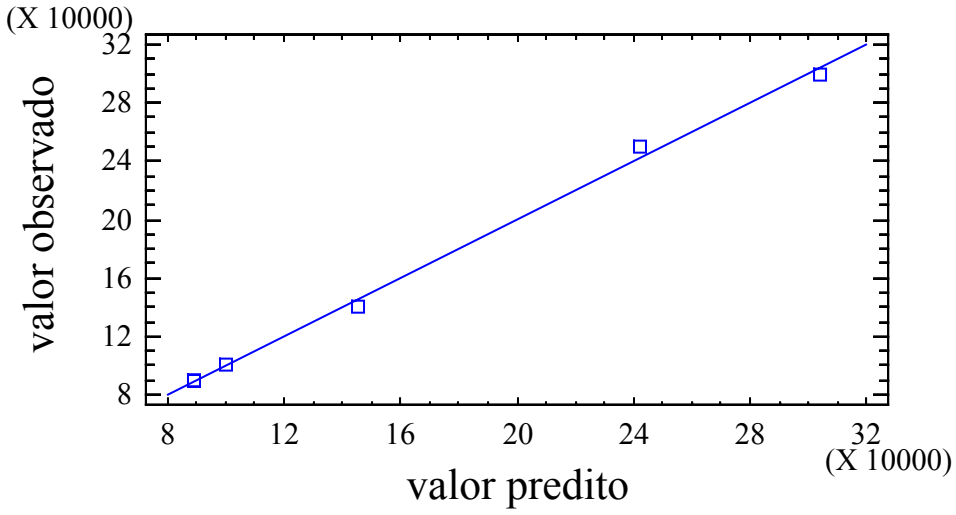
Tabela 4.54: Análise de variância do ajuste do modelo de regressão da classe 2 de terrenos

Fonte	Soma dos Quadrados	G.L.	Quadrado Médio	F	Valor-p
Modelo	4,13778E10	2	2,06889E10	587,94	0,0001
Resíduo	1,05567E8	3	3,51889E7		
Total	4,14833E10	5			

Fonte: A Autora, Pesquisa de Campo, 2004.

A figura 4.20 adiante apresenta pontos em linha diagonal, indicando uma boa linearidade, ou seja, as previsões se aproximam dos valores reais.

Figura 4.20: Valor predito *versus* valor observado na classe 2 de terrenos



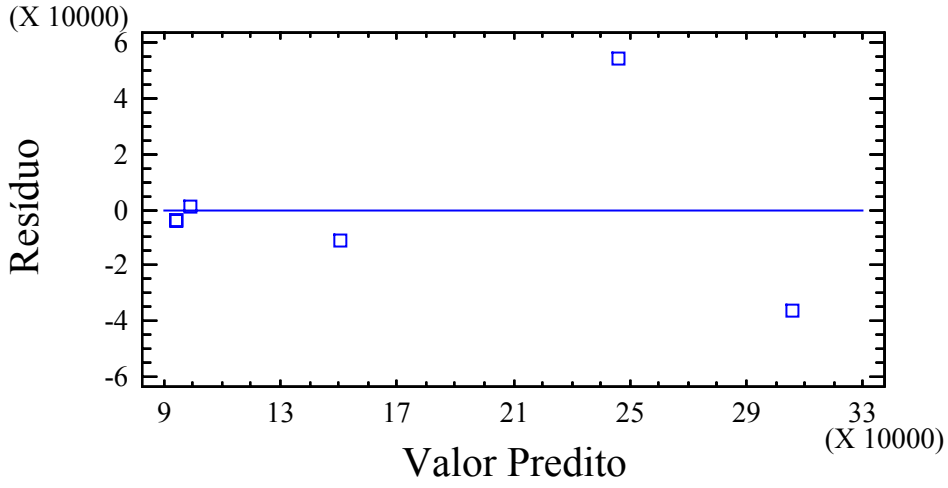
Fonte: A Autora, Pesquisa de Campo, 2004.

A análise dos resíduos é apresentada a seguir.

A independência serial dos resíduos (não existir autocorrelação dos erros) foi comprovada pelo teste Durbin-Watson (DW), pois o valor-p é 0,3462, que é maior que 0,05. Assim, não há nenhuma indicação de autocorrelação entre os resíduos.

A homogeneidade de variância foi verificada e a figura 4.21 apresenta os pontos distribuídos aleatoriamente em torno da linha que passa pela origem. Esta disposição dos pontos indica que a suposição de variância constate está correta.

Figura 4.21: Resíduos *versus* valores preditos na classe 2 de terrenos



Fonte: A Autora, Pesquisa de Campo, 2004.

Foi testada a Gaussianidade e o teste de *Kolmogorov- Smirnov* forneceu valor-p de 0, 888692, que é maior que 0,05, o que indica que a distribuição dos resíduos é Gaussiana.

Os valores preditos pela equação ajustada e os valores observados e o de erro da predição são mostrados no quadro 4.9.

Quadro 4.9: Quadro de valores na classe 2 de terrenos

Valor Observado (R$) (Y)	Valor Predito (R$) $\left(\hat{Y}\right)$	Resíduo (R$) $\left(\varepsilon\right)$	Erro Percentual (%)
140000	145725	5725,00	4,08929
90000	88880,6	1119,40	1,243778
100000	100164	164,00	0,164
90000	88880,6	1119,40	1,243778
250000	242536	7464,00	2,9856
300000	303813	3813,00	1,271

Fonte: A Autora, Pesquisa de Campo, 2004.

5 CONSIDERAÇÕES FINAIS

5.1 CONCLUSÕES

O valor de um imóvel pode ser estimado por meio de uma equação de regressão linear múltipla, desde que se disponha de um banco de dados formado por uma amostra do tipo aleatório com informações de preço e das principais características dos imóveis.

A construção de grupos homogêneos de itens para cada tipo de imóveis (apartamentos, casas e terrenos) pode ser feita pelo método multivariado de Análise de Agrupamento. A transformação e redução do número de variáveis pode ser obtida pelo método da Análise de Componentes Principais, sem perda significativa de informação e obtendo , assim, variáveis não correlacionadas.

Com os resultados obtidos da Análise de Componentes Principais permitiu-se a construção de modelos de regressão linear múltipla para cada classe homogênea sem problema de multicolinearidade que, geralmente, está presente nos estudos referente a avaliação de imóveis.

Os resultados obtidos com os modelos de regressão linear múltipla, que melhor representa o mercado de imóveis urbanos de Campo Mourão – PR, para cada uma das classes homogêneas de cada tipo de imóveis (apartamentos, casas e terrenos), são altamente precisos, com ótimos valores nos testes realizados, cumprindo todas as condições exigidas para serem consideradas avaliações rigorosa, dessa forma, em um período de tempo determinado, prediz o valor de mercado do imóvel, com alta precisão e de forma objetiva.

A conclusão geral obtida dos resultados é que a metodologia multivariada aplicada se mostrou viável e altamente apropriada para avaliação dos conjuntos de: apartamentos, casas e terrenos da área de estudo, permitindo a consideração e correto tratamento das diferenças e variância interna do próprio conjunto. Assim, atingiu-se resultados com alto nível de precisão

e a metodologia pode ser aplicada de modo geral em predições dos preços de imóveis da cidade de Campo Mourão, levando em conta as mesmas condições dos imóveis analisados.

Assim, o presente trabalho procurou oferecer uma contribuição na área de Engenharia de Avaliação.

5.2 SUGESTÃO PARA FUTURA PESQUISA

Construir um modelo de avaliação utilizando Redes Neurais e comparar a eficiência dessa metodologia com a estatística aqui aplicada.

REFERÊNCIAS

ABNT (Associação Brasileira de Normas Técnicas). **Avaliação de imóveis urbanos** (NBR5676 e NBR 502). Rio de Janeiro: ABNT, 1989.

ABNT (Associação Brasileira de Normas Técnicas). **Avaliação de imóveis urbanos** (NBR 5676 e NBR 502). Rio de Janeiro: ABNT, 1990.

ABNT (Associação Brasileira de Normas Técnicas). **Avaliação de imóveis urbanos** (NBR 5676 e NBR 502). Rio de Janeiro: ABNT, 2004.

ABUHNAMAN, S. A. **Curso básico de engenharia legal e de avaliações**. São Paulo: Pini, 1998.

BARBOSA FILHO, D. S. **Técnicas avançadas de engenharia de avaliações**. Caixa Econômica Federal, 1998.

BOUROCHE, J. M. SAPORTA, G. **Análise de dados**. Rio de Janeiro: Zahar, 1982.

CUADRAS, C. M. **Métodos de análisis multivariante**. Universidade de Barcelona, 1981.

CRIVISQUI, E. M. **Análisis factorial de correspondencias**:.un instrumento de investigación en ciencias sociales. Laboratoire de Méthodologie du Traitement des Données, Université Libre de Bruxelles. Edición: Universidad Católica de Asunción, Asunción, Paraguay, 1993.

DANIEL, C.; WOOD, T. E. **Fitting equations to data**. New York: John Wiley & Sons, Inc, 1971.

DANTAS, R. A. **Engenharia de avaliações**: introdução à metodologia científica. São Paulo: Pini, 1998.

DANTAS, R.A. **Engenharia de avaliações uma introdução à metodologia científica**. [S.l.] São Paulo: Pini, 2000.

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

ELIAN, S. N. **Análise de regressão**. São Paulo: IME,1998.

FIKER, J. **Avaliação de imóveis urbanos**. 5. ed. São Paulo: Pini, 1997.

GONZÁLEZ, M. A. S.; FORMOSO, C. T. **Análise conceitual das dificuldades na determinação de modelos de formação de preços através da análise de regressão**. Engenharia Civil – UM, 8: 65-75, 2000.

GONZÁLEZ, M. A. S.. **A engenharia de avaliações na visão inferencial**. São Leopoldo: Unisinos,1998.

GONZALÉZ, M. A. S. **A engenharia de avaliações na visão inferencial**. São Leopoldo: Unisinos, 1997.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. Englewood Cliffs, New Jersey: Prentice Hall, 1982.

JOHNSON, R. A. & WICHERN, D. W. **Applied multivariate statistical analysis**. 2. ed. New Jersey: Prentice Hall International,Inc., 1988.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4. ed. Nova Jersey: Prentice Hall, Inc., 1998.

JOHNSTON, J. **Métodos econométricos**. São Paulo: Atlas, 1974.

JOHNSTON, J. **Métodos econométricos**. São Paulo: Atlas, 1986.

JUDGE, G. G.; GRIFFITHS, E. W.; HILL, R. C.; LEE, T. **The theory and practice of econometrics**. New York: John Wiley & Sons,1980.

KMENTA, J. **Elementos de econometria**. São Paulo: Atlas, 1978.

LACHENBRUCH, P. A. **Discriminant analysis**. New York: Hafner Press, 1975.

LEBART, L.; MORINEAU, A.; FÉNELON, J. **Tratamiento estadístico de datos**. Barcelona: Marcombo Boixareu, 1995.

MOLINA, M. A. **El catastro en España**. Valência: UPV, 1999.

MONTENEGRO DUARTE, A. **Modelo geral de valores isento de subjetividade**: caso de apartamentos na cidade de Belém. Dissertação de Mestrado. Porto Alegre e Valência, 1999.

MONTGOMERY, D. C. **Design and analysis of experiments**. 4. ed. USA: John Wiley, 1997.

MOREIRA, A. L. **Princípios de engenharia de avaliações**. São Paulo: Pini, 1990.

MOREIRA, A. L. **Princípios de engenharia de avaliação**. São Paulo: Pini, 1994.

MOREIRA, A. L. **Princípios de engenharia de avaliações**. São Paulo: Pini, 1997.

MOREIRA FILHO, I. I.; FRAINER, J. I.; MOREIRA, R. M. I. **Avaliação de bens por estatística inferencial e regressões múltiplas**. Porto Alegre: Avalien, 1993.

MOSCOVITCH, S. K. **Qualidade de vida urbana e valores de imóveis**: um estudo de caso para Belo Horizonte. Nova Economia, número especial: 247-279, 1997.

NETER, J.; WASSERMAN, W. **Applied linear statistical models**. Richard D. Irwin, Inc, Illinois, 1974.

PEREIRA, R. S. **Estatística e suas aplicações**. São Paulo: Grafosul, 1970.

PLA, L. E. **Análisis multivariado**: método de componentes principales. Secretaria general de la organización de estados americanos. Washington, 1986.

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical methods**. 6. ed., Iowa: Ames, 1972.

WORZALA, E.; LENK M.; SILVA A. **An exploration of neural networks and iténs application to real estate valuation**. The Journal of Real Estate Research, 10 (2): 185-201, 1995.

ANEXO I - Matriz de dados referente aos apartamentos

apartamento	posição do apto	elevador	garagem	localização	área privativa	pavimento	andar	pecas	sala	dormitório	suíte	banheiro	dep. de emp.
1	3	2	1	1	222	15	3	9	1	2	1	1	1
2	3	1	1	1	162,44	7	1	9	1	2	1	1	1
3	3	1	1	1	176	8	2	9	1	2	1	1	1
4	3	2	2	1	179,2	14	3	10	2	2	1	1	1
5	3	2	2	1	279,4	20	3	8	1	2	1	1	1
6	3	1	1	1	120	12	3	12	2	3	1	2	1
7	2	1	1	1	220	16	2	12	3	3	1	2	1
8	3	1	1	1	107	8	1	9	3	2	1	1	0
9	2	0	1	1	50	4	2	4	1	1	0	1	0
10	3	2	3	1	240	15	3	12	3	2	1	1	1
11	3	0	1	0	100	3	2	7	2	3	0	2	0
12	3	0	1	1	147	6	1	7	1	2	1	1	0
13	3	1	1	1	108	7	3	11	2	2	1	1	1
14	2	2	2	1	311	16	3	10	3	2	1	1	1
15	2	1	1	1	132	7	2	8	2	2	1	1	0
16	3	2	1	1	107	8	3	8	2	2	1	1	0
17	3	2	1	1	107	8	4	8	2	2	1	1	0
18	3	2	4	1	330	15	3	17	3	3	1	3	1
19	3	2	2	1	220	15	3	13	2	3	1	1	1
20	3	1	1	1	130	6	3	11	2	3	1	1	1
21	3	2	2	1	164	10	3	11	2	3	1	1	1
22	3	2	2	1	160	15	3	11	2	3	1	2	1
23	3	2	1	1	374	15	3	14	3	2	1	3	1
24	3	1	1	1	220	16	4	14	3	3	1	2	1
25	2	1	1	1	220	16	3	14	3	3	1	2	1
26	3	1	2	1	180	13	3	13	2	3	1	1	1
27	3	1	2	1	320	13	3	11	1	3	1	1	1

28	3	1	2	1	180	13	3	13	2	3	1	1	1
29	3	1	2	1	320	13	3	11	1	3	1	1	1
30	1	2	2	1	260	14	3	11	2	2	1	1	1
31	2	2	2	1	260	14	2	11	2	2	1	1	1
32	3	2	2	1	260	14	1	11	2	2	1	1	1
33	3	1	1	1	140	8	2	6	1	2	1	1	0
34	1	1	1	1	130	7	3	7	2	2	1	1	0
35	1	2	2	1	260	14	1	11	2	2	1	1	1
36	1	2	2	1	310	16	2	11	2	2	1	1	1
37	3	0	1	1	100	4	3	7	1	3	0	2	0
38	3	0	1	1	70	3	2	5	1	2	0	1	0
39	3	0	1	1	90	3	2	6	1	2	0	2	0
40	3	1	1	1	38	8	2	4	1	1	0	1	0
41	3	1	1	1	40	8	3	5	1	2	0	1	0
42	3	1	1	1	170	7	2	9	2	2	1	1	0
43	1	1	1	1	170	7	1	9	2	2	1	1	0
44	3	1	1	1	170	7	3	9	2	2	1	1	0

apartamento	dist. de escola	dist. de hospital	dist. de supermercado	acabamento	revestimento do prédio	conservação	idade real	idade aparente	valor (R$)
1	2	2	3	3	4	2	2	2	130000
2	3	1	3	3	1	5	4	6	85000
3	3	2	3	3	1	5	4	6	80000
4	3	2	3	3	4	5	4	6	115000
5	3	2	3	3	4	5	2	4	150000
6	3	2	3	3	2	5	3	4	110000
7	2	2	2	3	4	4	2	2	120000
8	2	2	2	2	1	3	3	5	68000
9	3	2	3	2	1	4	4	4	40000
10	3	3	3	3	4	4	2	3	170000
11	3	1	1	2	1	3	2	3	50000
12	3	2	3	2	2	5	4	5	60000
13	3	3	3	2	4	3	3	3	93000
14	1	1	1	3	4	5	4	4	250000

15	3	2	1	2	1	4	5	5	65000
16	3	3	3	2	4	4	3	3	65000
17	3	3	3	2	4	4	3	3	71000
18	2	1	2	2	4	4	3	2	220000
19	3	3	3	2	4	4	5	5	200000
20	3	3	3	2	3,5	4	3	4	90000
21	3	3	3	2	3,5	4	3	4	140000
22	3	3	3	2	4	4	3	4	250000
23	1	1	1	2	4	4	4	4	250000
24	3	3	3	3	4	4	3	4	150000
25	3	3	3	3	4	4	3	4	120000
26	3	1	3	3	4	3	4	4	180000
27	3	3	3	3	4	4	5	5	250000
28	3	1	3	3	4	3	4	4	180000
29	3	3	3	3	4	4	5	5	250000
30	3	3	3	2	4	3	2	2	115000
31	3	3	3	2	4	3	2	2	120000
32	3	3	3	2	4	3	2	2	140000
33	2	2	3	2	4	3	5	5	90000
34	3	3	3	2	1	3	4	4	65000
35	3	3	3	2	4	3	2	2	120000
36	1	1	1	3	4	4	4	4	210000
37	3	3	3	2	1	3	1	1	100000
38	3	3	3	2	1	3	1	1	70000
39	3	3	3	2	1	2	1	1	95000
40	3	3	3	2	1	3	3	4	30000
41	3	3	3	2	1	3	3	4	40000
42	3	3	3	2	4	4	4	6	100000
43	3	3	3	2	4	4	4	6	90000
44	3	3	3	2	4	4	4	6	110000

ANEXO II - Matriz de dados referente a casas residenciais

casa	bairro	garagem	suíte	banheiro	edícula	dist. supermercado	área construída	área do terreno	acabamento	cobertura	estrutura	conservação	piscina
1	5	1	1	2	1	3	183	500	1	4	3	1	0
2	4	1	1	3	0	2	216	977	1	4	3	1	0
3	4	1	1	3	0	2	155	420	2	4	3	1	0
4	2	1	1	2	0	1	170	490	2	4	3	1	0
5	4	1	1	2	1	1	160	480	2	4	3	3	0
6	5	1	1	2	1	3	160	500	2	4	3	3	0
7	5	1	1	2	1	3	134	1000	2	4	3	2	0
8	3	1	0	1	0	3	70	300	1	2	1	1	0
9	3	0	0	1	0	2	198	350	1	1	1	3	0
10	3	1	0	1	0	2	113	400	2	2	1	2	0
11	5	1	0	2	1	3	238	1200	2	4	3	2	0
12	3	1	0	2	0	2	158	315	2	1	1	2	0
13	3	1	0	1	0	3	124	300	2	4	3	2	0
14	5	0	0	1	0	3	95	340	2	2	1	1	0
15	4	1	1	1	1	3	187	490	3	4	3	2	0
16	4	1	1	1	1	1	242	490	3	4	3	2	0
17	3	0	0	2	0	1	100	480	2	4	3	2	0
18	3	1	1	1	0	1	180	786	3	4	3	3	0
19	5	1	1	3	1	1	380	480	2	3	3	2	0
20	4	1	1	1	1	1	240	490	3	4	3	2	1
21	3	1	1	1	1	2	184	1000	2	4	3	3	1
22	3	1	0	2	0	3	140	446	2	1	2	2	0
23	5	1	1	1	0	3	400	600	2	4	3	2	0
24	4	1	0	1	0	3	110	270	2	4	3	3	0
25	2	0	0	1	0	3	120	300	2	4	3	1	0

26	2	1	0	1	0	3	150	300	2	4	3	3	0
27	5	1	1	2	0	3	400	750	2	4	3	3	0
28	5	1	0	1	0	3	130	300	2	2	1	1	0
29	4	1	1	1	0	3	200	400	2	4	3	3	0
30	4	1	1	2	1	1	244	500	2	4	3	2	0
31	5	1	1	1	1	1	113	300	2	4	3	2	0
32	5	1	1	2	1	1	220	500	4	4	3	2	1
33	5	1	1	1	1	3	92	650	2	4	3	3	0
34	5	1	1	1	0	3	123	300	2	4	3	2	0
35	5	1	0	1	0	3	150	1000	1	2	1	1	0
36	5	1	1	2	1	3	200	400	2	4	4	2	0
37	5	1	0	1	0	3	100	1000	1	1	1	1	0
38	4	1	0	1	0	2	70	490	1	1	1	1	0
39	5	1	0	1	0	3	100	950	2	2	1	1	0
40	5	1	0	1	0	3	80	475	2	2	1	1	0
41	3	0	0	1	0	2	70	600	1	3	3	3	0
42	3	1	1	1	0	2	180	640	2	4	3	3	0
43	2	0	1	2	0	2	70	480	2	4	3	3	0
44	3	1	1	1	0	1	115	250	2	4	3	3	0
45	3	1	1	1	0	1	160	450	2	4	3	3	1
46	3	1	1	2	1	3	325	225	3	4	4	3	0
47	5	1	1	2	0	3	320	450	3	4	4	3	0
48	2	1	0	1	0	2	116	300	2	3	3	2	0
49	5	1	1	3	0	3	180	500	2	4	3	3	0
50	2	1	0	1	0	3	100	800	2	3,5	3	2	0
51	3	1	0	1	0	2	80	390	2	2	3	1	0

casa	dormitório	dep.de empregados	lavanderia	peças	idade aparente	valor (R$)
1	2	1	1	8	2	120.000
2	3	1	1	13	2	160.000
3	3	1	1	12	2	110.000
4	2	0	1	7	3	70.000

5	2	0	1	9	3	85.000
6	3	0	1	8	3	100.000
7	4	0	1	10	4	160.000
8	3	0	0	5	1	30.000
9	2	0	0	5	3	28.000
10	3	0	0	8	2	35.000
11	4	0	1	9	1	150.000
12	4	0	1	12	1	45.000
13	3	0	0	8	1	30.000
14	2	0	0	6	1	28.000
15	2	1	1	12	4	140.000
16	3	1	1	13	3	130.000
17	3	0	1	8	5	50.000
18	2	1	1	10	3	150.000
19	3	0	1	14	5	150.000
20	2	1	1	11	2	135.000
21	2	1	1	13	3	180.000
22	3	0	1	6	2	40.000
23	5	1	1	14	3	150.000
24	2	0	1	5	4	60.000
25	3	0	1	6	3	15.000
26	3	0	1	6	5	45.000
27	3	0	1	11	3	165.000
28	3	0	0	5	3	95.000
29	3	0	0	9	6	100.000
30	3	1	1	13	6	130.000
31	2	1	0	9	4	90.000
32	3	1	1	15	3	185.000
33	2	0	1	7	5	90.000
34	2	0	1	7	5	15.000
35	2	0	1	7	1	150.000
36	2	1	1	7	2	150.000
37	2	0	1	5	1	140.000

38	2	0	1	5	1	30.000
39	3	0	1	7	1	60.000
40	3	0	1	7	1	45.000
41	2	0	0	5	5	30.000
42	2	1	1	10	6	90.000
43	1	0	1	8	5	50.000
44	2	0	1	8	4	70.000
45	2	1	1	9	1	130.000
46	4	1	1	16	6	180.000
47	2	0	0	13	4	290.000
48	3	0	0	6	3	50.000
49	3	0,5	1	12	5	98.000
50	2	0	0	5	4	65.000
51	3	0	1	6	2	40.000

ANEXO III - Matriz de dados referente aos terrenos

terreno	localização	setor comercial	pólo	frente	área do terreno	proteção	plano	inclinado	posição	pavimentação	valor (R$)
1	6	2	1	2	650	3	0	1	1	1	100000
2	2	0	0	3	640	3	1	0	2	1	25000
3	5	0	1	1	500	3	1	0	1	1	43000
4	3	0	1	2	390	3	2	0	1	1	33000
5	5	0	1	3	262	3	2	0	2	1	40000
6	5	2	1	3	1000	3	2	0	1	1	140000
7	4	1	1	3	950	3	2	0	1	1	90000
8	4	0	1	1	475	3	0	1	1	1	38000
9	5	1	1	3	470	3	3	0	2	1	70000
10	6	3	1	1	500	0	3	0	1	1	145000
11	2	0	0	2	420	0	3	0	2	0	15000
12	2	0	0	2	420	0	3	0	1	0	13000
13	3	2	1	3	2000	0	2	0	2	1	100000
14	4	1	1	3	950	3	2	0	1	1	90000
15	6	3	1	3	1000	3	2	0	2	1	250000
16	5	0	1	1	242	0	2	0	1	1	38000
17	6	3	1	3	940	2	3	0	2	1	300000
18	5	0	1	1	500	2	1	0	1	1	47000
19	5	0	1	3	350	0	2	0	2	1	60000
20	5	0	1	1	500	0	0	1	1	1	50000
21	3	0	0	2	336	0	3	0	1	1	15000
22	3	0	1	2	336	0	3	0	1	1	30000
23	4	0	1	2	300	0	2	0	1	1	25000
24	3	0	1	1	450	3	3	0	1	1	30000

Anexo IV - Questionário para apartamentos e residências

Ficha			
Classificação do imóvel:			
Rua:			
Bairro:			
Valor (reais)			
Data da venda imobiliária:			

pólos de influência : (0) não existente (1) valorizante (-1) desvalorizante Qual:

Classificação do bairro: (5) centro (4) ótimo (3)bom (2) razoável (1)inferior

Peças do imóvel (unidade) total = quantidade mesmo			
Sala de star=	Dep. Empregado (1) completa (0.5) incompleta (0) inexistente		
SalaJantar=	Cozinha=	Varanda=	Área de seviço=
Salatv=	Escritório=	Lavabo=	Despensa=
Garagens=	Suíte=	Sacada=	Churrasqueira=
Quarto social	Banheiros=	1.1 EDÍCULA (M^2)	Piscina

Relação a proximidade			
(3) próximo até 500m (2) distante de 500m – 800m (1) muito distante mais de 800m			
Tipos de escolas	( ) próximo	( ) distante	() muito distante
De hospitais	( ) próximo	( ) distante	() muito distante
Centro comercial	( ) próximo	( ) distante	() muito distante
De supermercados	( ) próximo	( ) distante	() muito distante

Sobre o terreno Área (m^2)

Sobre a edificação				
Padrão de construção	Alto (3)	Normal (2)	Baixo (1)	
Estrutura	Alvenaria (3)	Madeira (2)	Mista (1)	Sobrado (4)
Cobertura	Laje+ Telha de barro (4)	Laje+ Eternit (3)	madeira+telha(2)	Eternit+madeira (1)
Revest. Fachada	Rebco/pintura (2)	Madeira(1)		
Est. Conservação	Otima(4)	3 bom	Regular 2	Péssima(1)
Idade real	Idade aparente			
0-1ano (6) 1-5 (5) 5-10(4) 10-15(3) 10-20amos (2) mais de 20 (1)				
Área construída (m^2)				

Anexo V - Questionário para terrenos

Localização
Rua:
Bairro classificação
Valor:

Pólo valorizante: não existe (-1) desvalorizante(0) valorizante (1)

Relação a proximidade

(3) próximo até 500m (2) distante de 500m – 800m (1) muito distante mais de 800m

Tipos de escolas	( ) próximo	( ) distante	() muito distante
De hospitais	( ) próximo	( ) distante	() muito distante
Centro comercial	( ) próximo	( ) distante	() muito distante
De supermercados	( ) próximo	( ) distante	() muito distante

Ponto Comercial:

(3) centro (2) periferia central (1) bairro (0) moradia

Sobre o terreno

Dimensões (m) : frente lateral fundo

Proteção: muro ( ) não tem ( ) outro ( ) qual?

Plano ( ) nível da rua () abaixo da rua ( ) acima da rua

Inclinado () requer aterro() () pouco inclinado ( ) muito inclinado

Firme ( ) brejoso () inundáveis ( ) () rochoso() outro().....

2 Posição na quadra: Meio() esquina()

Serviços urbanos disponível: Pavimentação () rede elétrica() outros (_) qual

Classificação vizinhança: Ótima boa razoável péssima

Outros:

Ficha

Classificação do imóvel:

Rua:
Bairro:
Valor (reais)
Data da venda imobiliária:

pólos de influência : (0) não existente (1) valorizante (-1) desvalorizante Qual:

Classificação do bairro: (5) centro (4) ótimo (3)bom (2) razoável (1)inferior

Peças do imóvel (unidade) total = quantidade mesmo

Sala de star=	Dep. Empregado (1) completa (0.5) incompleta (0) inexistente		
SalaJantar=	Cozinha=	Varanda=	Área de seviço=
Salatv=	Escritório=	Lavabo=	Despensa=
Garagens=	Suíte=	Sacada=	Churrasqueira=
Quarto social	Banheiros=	Edícula (m^2)	Piscina

Relação a proximidade

(3) próximo até 500m (2) distante de 500m – 800m (1) muito distante mais de 800m

Tipos de escolas	( ) próximo	( ) distante	() muito distante
De hospitais	( ) próximo	( ) distante	() muito distante
Centro comercial	( ) próximo	( ) distante	() muito distante
De supermercados	( ) próximo	( ) distante	() muito distante

Sobre o terreno Área (m^2)

Sobre a edificação				
Padrão de construção	Alto (3)	Normal (2)	Baixo (1)	
Estrutura	Alvenaria (3)	Madeira (2)	Mista (1)	Sobrado (4)
Cobertura	Laje+ Telha de barro (4)	Laje+ Eternit(3)	madeira+telha(2)	Eternit+madeira (1)
Revest. Fachada	Rebco/pintura (2	Madeira(1)		
Est. Conservação	Otima(4)	3 bom	Regular 2	Péssima(1)
Idade real	Idade aparente			
0-1ano (6) 1-5 (5) 5-10(4) 10-15(3) 10-20amos (2) mais de 20 (1)				
Área construída (m^2)				

Ficha de avaliação de imóveis comerciais

Localização: centro comercial ( ) periferia do centro ( ) bairro ( )
Rua:
Bairro classificação () ótimo () bom () regular () inferior
Valor:
Pólo valorizante: não existe () desvalorizante( ) valorizante ( ) Qual?_____

Relação a proximidade
(4) 0 (3) próximo de 50 até 300m (2) distante de 300m – 600m (1) muito distante mais de 600m

Tipos de escolas	( ) Até 50m	( ) próximo	( ) distante	() muito distante
De hospitais	( ) Até 50m	( ) próximo	( ) distante	() muito distante
Centro comercial	( ) Até 50m	( ) próximo	( ) distante	() muito distante
Banco	( ) Até 50m	( ) próximo	( ) distante	() muito distante

Sobre o terreno: Dimensões m2 frente(M)

Posição na quadra: Meio() esquina()

Serviços urbanos disponível: Pavimentação () rede elétrica() outros (_) qual
Telefone ( ) ar condicionado ( )

Área total da loja: sobreloja área
Frente (m) :

Posição se caso loja de shopping
Interior ( ) frente para rua ( )

Padrão de acabamento
Alto médio baixo

Estado de conservação
Ótima boa regular péssima

Idade aparente Idade real
Possui: banheiro ( ) cozinha ()
Utilização:

Comercio: luxo fino normal inferior

Existência de estacionamento

Cobertura: laje telha de barro Eternit madeira gesso
Estrutura: madeira concreto alvenaria mista