UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

VALÉRIA DA CRUZ RIBEIRO

ANÁLISE DE DEMANDA POR TRANSPORTES DE PASSAGEIROS VIA MODELOS DE REGRESSÃO GEOREFERENCIADOS

Vitória
2012

VALÉRIA DA CRUZ RIBEIRO

ANÁLISE DE DEMANDA POR TRANSPORTES DE PASSAGEIROS VIA MODELOS DE REGRESSÃO GEOREFERENCIADOS

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Civil Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Mestre em Engenharia Civil, Área de Concentração em Transportes.

Orientador: Prof. Dr. Adelmo Inácio Bertolde
Co-orientador: Prof. Dr. Gregório Coelho de Morais Neto

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Vitória – ES, Junho de 2012

ANÁLISE DE DEMANDA POR TRANSPORTES DE PASSAGEIROS VIA MODELOS DE REGRESSÃO GEOREFERENCIADOS

Valéria da Cruz Ribeiro

Dissertação apresentada ao Curso de Engenharia Civil do Programa de Pós Graduação em Engenharia Civil do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para Obtenção do Grau de Mestre em Engenharia Civil, na área de Concentração em Transportes.

Aprovada em 29 /06/2012, por:

Adelmo Inácio Bertolde – Prof.
Doutor em Estatística
Deptº Estatística/ UFES
Orientador

Gregório Coelho de Morais Neto – Prof.
Doutor em Engenharia de Transportes
Deptº Eng. Produção/ UFES
Co – Orientador

Marta Monteiro da Costa Cruz – Prof. Drª.
Doutora em Engenharia de Transportes
Deptº Eng. Produção/ UFES
Examinadora Interna

Vânia Barcellos Gouvêa Campos – Prof.ª
Doutora em Ciências em Engenharia de Produção
Instituto Militar de Engenharia - IME/RJ
Examinadora Externa

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

VITÓRIA - ES, Junho de 2012

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Central da Universidade Federal do Espírito Santo, ES, Brasil)

R484a

Ribeiro, Valéria da Cruz, 1980-
Análise de demanda por transportes de passageiros via modelos de regressão georeferenciados / Valéria da Cruz Ribeiro. – 2012.
81 f. : il.

Orientador: Adelmo Inácio Bertolde.
Coorientador: Gregório Coelho de Morais Neto.
Dissertação (Mestrado em Engenharia Civil) – Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Transportes - Planejamento. 2. Análise de regressão. 3. Análise espacial (Estatística). I. Bertolde, Adelmo Inácio. II. Morais Neto, Gregório Coelho de. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 624

Aprender é a única coisa que a mente nunca se cansa, não teme e nem se arrepende.

Leonardo da Vinci

DEDICATÓRIA

Aos meus pais Antonio e Edna, que mesmo com os poucos anos de estudo que tiveram souberam me mostrar o valor que o aprendizado pode trazer para vida. Obrigada por serem meus exemplos de vida, fonte de inspiração, apoio e ensino diário.

AGRADECIMENTOS

Primeiro a Deus, que me capacitou para superar todos os obstáculos e me fortaleceu nos momentos mais difíceis.

À minha querida família e em especial aos meus pais Antonio e Edna, exemplos de vida para mim, por me incentivarem a prosseguir nos estudos e me motivarem a lutar pelos meus sonhos.

Aos colegas de Mestrado pelas horas de estudo em grupo, trocas de conhecimentos, apoio para prosseguir e pelos momentos de descontração, em especial a: Anne Francine, Belcristi Amorim, Josiane Baldo e Patrícia Rodrigues.

À todos aqueles que me apoiaram e de alguma forma contribuíram para a realização deste trabalho assim como minha irmã e grande motivadora Eliane da Cruz Ribeiro Calisto e minhas amigas de todas as horas: Silvana Nascimento e Regiane Teodoro, que acreditaram em mim e me incentivaram a todo momento.

Ao Programa de Pós Graduação em Engenharia Civil da UFES que me possibilitou um ambiente acadêmico propício à confecção deste trabalho.

Ao meu orientador, Adelmo Inácio Bertolde, que me ensinou a enfrentar meus medos e a acreditar em mim, obrigada pela paciência, pelas valiosas orientações e apoio nos momentos mais difíceis, obrigada por contribuir para meu crescimento pessoal e intelectual.

Ao meu Co – Orientador, Gregório Coelho de Morais Neto, pelo apoio e incentivo indispensáveis ao longo de todo o curso.

À minha chefe, a Engenheira Civil Cecília Carvalho, e demais colegas de trabalho por compreenderem a minha ausência no trabalho para frequentar a minhas aulas e reuniões de orientação que foram fundamentais para a execução dessa dissertação de mestrado.

RESUMO

A presente dissertação – Análise de Demanda por Transportes de Passageiros via Modelos de Regressão Georeferenciados – apresenta, além de uma metodologia para a construção de modelos de regressão espacial e geograficamente ponderados, uma avaliação dos mesmos quando comparados aos modelos de regressão tradicionais e modelos de regressão com variáveis dummies, no sentido de prever a demanda de viagem para o município de Vitória, capital do Espírito Santo, com o intuito de obter informações que possam subsidiar o planejamento de transportes de maneira mais eficaz. Para isto, utilizou-se dados da pesquisa domiciliar de origem e destino (OD) realizada no ano de 1998 na região metropolitana da grande Vitória, foram calibrados quatro modelos de regressão de modelagem de demanda de viagem: Modelo de Regressão Tradicional, Modelo de Regressão Dummy, Modelo de Regressão Espacial e Modelo de Regressão Geograficamente Ponderada. Após a calibração, os modelos foram testados a partir da aplicação nos dados da pesquisa domiciliar de origem e destino realizada em 2007 na mesma cidade, para comparar e validar a estimativa. Conclui-se que a hipótese principal, ou parte dela, considerada neste trabalho foi confirmada, de que um modelo de regressão espacial ou geograficamente ponderada por distâncias pode ser mais explicativo do que modelos de regressão convencionais, pois a calibração de modelos de demanda de viagem pelo modelo de regressão ponderada apresentou valores das estatísticas de ajustes menores que os outros modelos.

Palavras-Chave: Demanda por Transportes, Regressão Espacial, Regressão Ponderada Geograficamente, Estatística Espacial.

ABSTRACT

This dissertation - Analysis of Demand for Passenger Transport via Regression Models georeferenced - presents, and a methodology for the construction of spatial regression models and geographically weighted, a risk assessment when compared to traditional regression models and regression models with dummies variables in order to forecast demand for travel to the city of Vitoria, capital of Espirito Santo, in order to obtain information that can subsidize the transportation planning more effectively. For this, we used data from the household survey of origin and destination (OD) held in 1998 in the metropolitan region of Vitoria, four models were calibrated regression modeling of travel demand: Traditional Model Regression, Regression Model dummy Regression Model Space and Geographically Weighted Regression Model. After calibration, the models were tested from the application data in the household survey of origin and destination conducted in 2007 in the same city, to compare and validate the estimate. We conclude that the main hypothesis, or part thereof, considered in this work was confirmed that a regression model spatial or geographically weighted distances can be more explanatory than conventional regression models, since the calibration of travel demand models by weighted regression model showed values of statistical adjustments smaller than the other models.

Keywords: Demand for Transport, Spatial Regression, Weighted Regression Geographically, Spatial Statistics.

LISTA DE QUADROS

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

CAPÍTULO 1

1 - INTRODUÇÃO

As pesquisas de demanda por transporte são de grande importância para o planejamento e gerenciamento urbano, onde há a necessidade de realização de estudos específicos com informações socioeconômicas e também sobre os deslocamentos da população, de maneira a embasar as projeções das demandas para o futuro. Então, modelos matemáticos são alimentados por essas informações, gerando prognósticos com certa credibilidade. Nesse contexto, a dependência espacial presente nos dados pode ser um importante fator a ser considerado no planejamento dos transportes, podendo levar a resultados de previsão de demanda mais eficientes que os usuais.

A metodologia aplicada nesse trabalho contempla a implementação de dois modelos de regressão: o espacial e o geograficamente ponderada, utilizando o conjunto de dados da pesquisa de origem e destino (OD) da região metropolitana da grande Vitória realizada no município de Vitória – ES, calibrando os modelos de demanda de viagem aos dados do ano de 1998, e testando-os aos dados do ano de 2007. Cabe ressaltar que tal banco também foi objeto de estudo de BRAGA (2009), que propôs um modelo direto de previsão de demanda de viagens utilizando modelo com variáveis Dummy.

Assim, pretende-se mostrar o ganho de adequabilidade do modelo de regressão espacial e do modelo geograficamente ponderada quando comparado a outros modelos de regressão que desconsideram a dependência espacial e a distância entre macrozonas. Espera-se que os resultados finais obtidos neste estudo possam ser usados pelos órgãos e entidades responsáveis pelo planejamento de transportes no gerenciamento e na administração do trânsito da cidade a fim de estabelecer situações de maximização de bem-estar social, o que vem a ser uma contribuição importante no processo de conhecimento da demanda por transporte.

1.1 - FORMULAÇÃO DO PROBLEMA

No estudo de planejamento de transportes um dos objetivos é modelar a demanda de viagem para que se possa ter o maior grau possível de conhecimento dos movimentos atuais e futuros de viagens. Esse conhecimento permite ao planejador de transportes tomar decisões mais acertadas, evitando assim o excesso ou escassez de oferta de transportes.

Assim, o problema de pesquisa é: “Como modelar a quantidade demandada de viagens por transportes nos deslocamentos entre os pares de Origem e Destino (O/D), onde os dados sofrem influências espaciais e geográficas, de maneira a auxiliar e melhor entender o planejamento de transportes?”

1.2 - HIPÓTESES

A hipótese principal a ser considerada neste trabalho é:

- Um modelo de regressão espacial ou geograficamente ponderada por distâncias pode ser mais explicativo do que modelos de regressão convencionais.

Como hipótese secundária tem-se que:

- As zonas de tráfego mais próximas ou que guardem relação direta de realização de viagens tendem a apresentar algum grau de correlação no que tange ao processo gerador de viagens.

1.3 - JUSTIFICATIVAS

Um dos fatores importantes ao se estudar a quantidade de viagens demandada entre origem e destino é a estrutura espacial das variáveis envolvidas. Nos modelos de regressão espaciais e modelos ponderadas pela distância é possível modelar o número de viagens de uma particular área considerando a estrutura espacial dos dados, possibilitando a redução dos erros nas estimativas obtidas.

SILVA (2006) afirma que a utilização de um modelo matemático que não incorpore o fator “espaço” em sua estrutura poderá subestimar ou superestimar as verdadeiras relações entre as variáveis, pois a contribuição do fator aleatório “espaço” estará distribuída entre as variáveis do modelo. Caso isso ocorra, o modelo perderá sua principal função que é tentar representar um fenômeno real. O transporte, em geral, depende de fatores geográficos, físicos, políticos e socioeconômicos sendo que o fator geográfico é muitas vezes não observado.

Assim este trabalho tem a intenção de repercutir de maneira positiva após a análise dos dados modelados com os modelos de regressão espacial e de regressão geograficamente ponderada para previsão de demanda de viagens, servindo para melhorar o entendimento do processo espacial que ocasiona a demanda nos diversos pares de origem e destino.

1.4 – OBJETIVOS

O objetivo geral dessa pesquisa é analisar a demanda por transportes, para tanto serão utilizados os dados do município de Vitória/ES, mediante o uso de ferramentas de planejamento de transportes e do desenvolvimento de modelos de regressão georeferenciados para previsão de demanda por viagem, de modo a obter previsões mais próximas da realidade de maneira que venha auxiliar o planejamento estratégico desse serviço.

Os objetivos específicos são:

- Realizar uma análise descritiva das variáveis do estudo;
- Comparar os resultados de um modelo de regressão convencional com os dos modelos de regressão espacial e ponderada.
- Definir quais variáveis influenciam na demanda por transporte.

1.5 - ESCOPO DO TRABALHO

O Trabalho está assim organizado:

O Capitulo 1 inicia-se com uma introdução, mostrando o problema de pesquisa, os objetivos de estudo e as justificativas do trabalho.

No Capitulo 2 são apresentados resumidamente os modelos de previsão de demanda por transportes, o referencial teórico sobre a análise de regressão espacial, apontando os principais conceitos da análise de regressão, regressão espacial, regressão geograficamente ponderada e regressão com variáveis dummies.

O Capítulo 3 discorre sobre os procedimentos metodológicos que delineiam a pesquisa, e explica os modelos a serem utilizados ao longo do estudo.

No Capítulo 4 são apresentados os resultados das análises dos modelos de regressão calibrados.

O capítulo 5 apresenta as considerações finais e algumas sugestões para trabalhos futuros de modo a contribuir para realização de mais pesquisas neste assunto.

Finalmente, no Capítulo 6, têm-se as referências bibliográficas.

Por fim, agrupam-se ao trabalho os anexos.

CAPÍTULO 2

2 – REFERENCIAL TEÓRICO

O presente capítulo apresenta as teorias e referências que serviram de base ao estudo.

2.1 - TRANSPORTE URBANO

O termo transporte em Física e Geografia está associado à mudança de entes físicos no espaço. Na área de Engenharia a denominação é dada ao deslocamento de pessoas e produtos. O deslocamento de pessoas é referido como transporte de passageiros, ao passo que o de produtos é referido como transporte de cargas. Quando os deslocamentos ocorrem no interior das cidades é empregado o termo transporte urbano (FERRAZ e TORRES, 2004).

Existem diversos modos empregados na realização do transporte. Segundo FERRAZ e TORRES (2004, p. 2), “a palavra modo é empregada para caracterizar a maneira como o transporte é realizado”.

O Quadro 1 apresenta classes, características e modos comumente utilizados nos deslocamentos diários da população nos centros urbanos das cidades.

Classes	Características	Modos
Privado ou Individual	Os veículos são conduzidos por um dos usuários, que pode escolher livremente o caminho e o horário de partida. Há, portanto, total flexibilidade de uso no espaço e no tempo. A capacidade do veículo é pequena e a posse pode ser momentânea.	A pé, bicicleta, motocicleta, carro (incluindo automóvel, perua/van ou camioneta/caminhonete).
Público, Coletivo ou de massa	Os veículos pertencem, em geral, a uma empresa e operam em rotas predefinidas e horários fixos. Não há flexibilidade de uso no espaço e no tempo. A capacidade do veículo é grande.	Ônibus, bonde, pré-metrô, metrô e trem suburbano.
Semipúblico	O veículo pertence a uma empresa ou indivíduo e pode ser utilizado por determinado grupo de indivíduos ou por qualquer pessoa, tendo rota e horários adaptáveis aos desejos dos usuários em vários graus.	Táxi, mototáxi, carona programada, lotação, veículo fretado ou alugado.

Quadro 1 – Classificação dos modos de transporte urbano de passageiros.
Fonte: FERRAZ e TORRES (2004), adaptado pela autora

2.2 - A DEMANDA POR TRANSPORTES

O Transporte tornou-se parte integral da vida urbana. A necessidade de movimentação de pessoas e/ou mercadorias entre os diferenciais locais gera demanda por transportes.

A demanda por viagens é derivada das atividades das pessoas: atividades de produção e de consumo de bens. Quanto maior o desenvolvimento da sociedade, maior a atividade econômica e, por consequência, a necessidade de deslocamentos. O padrão atual de uso do solo urbano em grandes cidades, onde se verifica um crescimento populacional desordenado e uma especialização das diferentes zonas em residenciais, comerciais e industriais, leva à necessidade de transporte motorizado para cobrir as grandes distâncias que separam as pessoas da maioria dos seus destinos (FERRONATTO, 2002)

A demanda de transporte tem como característica ser:

- Altamente diferenciada: Ela pode variar com a hora do dia, com o dia da semana, propósito da viagem, tipo de carga, com o tipo de transporte oferecido.
- Derivada, isto é, as pessoas viajam para satisfazer uma necessidade em seu destino.
- Concentrada em poucas horas do dia nas áreas urbanas, particularmente nas horas de pico.

A previsão do número de viagens produzidas e atraídas é determinada por meio de relações matemáticas estabelecidas, principalmente, entre o uso do solo e os padrões de deslocamentos para as condições presentes. Essas informações são adquiridas através da coleta de dados. Antes da coleta é necessário, porém, que se defina a área de estudo. Como a pesquisa é no âmbito de transportes, divide-se a área de estudo em zonas de tráfego.

Segundo MANHEIN (1979), demanda básica tem como características os motivos das decisões de um indivíduo (ou domicílio) e o desejo de ter um padrão de atividades, que pode ser definido pelas escolhas que ele faz quanto a emprego, residência, padrões de consumo e atividades sociais. O estilo de vida desejado determina o padrão de atividades adotado, que origina as escolhas de localização, as quais, por sua vez, levam às decisões de viagens. Para adotar um determinado padrão de atividades, o indivíduo necessita estar em determinados lugares em determinados momentos, o que leva às escolhas.

Conhecer e compreender a demanda de transporte da região sob estudo é de fundamental importância para se obter o máximo de satisfação na demanda de transporte, pois dessa pode dimensionar a oferta, implantar novos sistemas e prever melhores formas de atender a demanda, ou seja, tomar uma decisão mais eficiente.

2.3 - PREVISÃO DE DEMANDA

A estimação da demanda por transporte, seja de passageiros ou de cargas, é um dos principais objetos de estudo do planejamento dos transportes MEYER e MILLER (2001). O objetivo principal na modelagem da demanda de viagens é produzir estimativas do volume de tráfego futuro. Isso é feito substituindo os fatores (variáveis) projetados em uma data futura no modelo de estimativa. Ter uma estimativa adequada da demanda existente é um apoio importante aos que precisam tomar as decisões e também uma forma de prevenir a possibilidade de não alcançar boas soluções para os problemas existentes.

Uma das maneiras de se obter informações sobre a demanda de viagem é através de pesquisas de origem e destino (O-D), e a análise da demanda de transporte é feita utilizando-se modelos de demanda, que procuram compreender os determinantes da demanda e a maneira como eles interagem e afetam a evolução do tráfego. De acordo com NOVAES (1986), três níveis de previsão de análise são em geral encontrados nos estudos da demanda de transportes:

Previsão a curto prazo: são previsões feitas através de análise marginal com base no quadro atual. Não são feitas projeções desagregadas das variáveis socioeconômicas e uso do solo. Sendo assim, as projeções se baseiam fundamentalmente na hipótese de que a distribuição espacial de atividades e os valores das variáveis socioeconômicas e uso do solo permanecerão a mesma.

Previsão a médio e longo prazo, se envolver efeitos nas atividades socioeconômicas: são previsões que exigem projeções detalhadas das variáveis socioeconômicas e atividades, tornando-se necessário estudar a evolução de todas as zonas.

Previsão a longo prazo, com avaliação dos efeitos nas atividades socioeconômicas e no seu assentamento (uso do solo): são previsões que fazem projeções detalhadas das variáveis socioeconômicas e de atividades, tornando-se necessário estudar a evolução de todas as zonas, estabelecem relações diretas de “feedback” entre os fluxos de transportes projetados e seus efeitos nas atividades socioeconômicas.

2.4 - MODELOS DE DEMANDA DE VIAGEM

MANHEIN (1979) classifica os modelos baseados em redes de transportes em dois grandes grupos:

- Modelos de demanda sequenciais.
- Modelos de demanda diretos ou simultâneos;

A identificação do ponto de equilíbrio entre a demanda e a oferta de viagens em uma rede de transportes é um objetivo comum aos modelos de demanda. A diferença entre os modelos de demanda diretos e sequenciais está no processo de modelagem.

2.4.1 - MODELOS SEQUENCIAIS

O modelo sequencial adota submodelos que implicam no seccionamento da modelagem da demanda de viagens em vários estágios sucessivos, pois considera que o viajante adota uma determinada sequência de decisões, sem que haja, a “priori’, uma razão que justifique tal escolha dentre as sequências alternativas (BEN-AKIVA *et al.* 1985).

O modelo sequencial (ou quatro etapas) recebe este nome por seguir etapas ou sequências e tem sido amplamente empregado no planejamento de transporte. De posse dos dados referentes ao zoneamento e ao sistema de redes de transportes, este modelo estima viagens entre as diversas zonas de tráfego. A divisão do problema em um modelo sequencial analisa a sequência de decisão que se acredita que um indivíduo tome antes de efetuar uma viagem. É baseado na hipótese de que o processo de decisão de viagem de um indivíduo é desenvolvido em etapas, ou seja, supõe-se primeiramente que o indivíduo decide exercer uma atividade e o local onde irá exercê-la, depois escolhe o modo de viagem e, por último, a rota. Dessa forma o modelo de quatro etapas é composto de submodelos, apresentados a seguir.

2.4.2 - MODELOS DE GERAÇÃO DE VIAGENS

Antes de começar a falar sobre os modelos diretos, é muito importante que se compreendam, primeiramente, alguns conceitos utilizados no planejamento de transportes. Diversos autores, dentre eles MEYER e MILLER (2001), adotam as seguintes terminologias:

- **Viagem**: é o movimento entre uma origem e um destino por algum motivo.
- **Viagem de base domiciliar**: são viagens que iniciam ou terminam no domicílio. **Viagem de base não domiciliar**: são as viagens que nem a origem nem o destino é o domicílio.
- **Produção de viagens**: refere-se à extremidade domiciliar (origem ou destino) de uma viagem de base domiciliar ou à origem de uma viagem de base não domiciliar.
- **Atração de viagens**: são viagens com destino não domiciliar, de viagens de base domiciliar ou destino de viagens de base não domiciliar.
- **Geração de viagens**: É a determinação do número de viagens, associada com uma zona de tráfego, domicílios ou outra unidade de geração, consistindo em viagens produzidas e atraídas para a unidade de geração.

Segundo PAPACOSTA e PREVEDOUROS (2000), uma viagem pode ser classificada de duas maneiras: origem e destino (O-D) ou produção e atração (P-A). Estes termos não são idênticos, sendo que origem e destino estão relacionados a ponto de saída e ponto de chegada, sem se preocupar com o uso do solo; já produção e atração se preocupam com o uso do solo. Essa distinção é feita por se considerar que produção de viagens é mais facilmente estimada a partir das características e necessidades de viagens das zonas, e atração de viagens depende de oportunidades não residenciais disponíveis nas zonas.

No modelo sequencial, o modelo de geração de viagens é o ponto de partida de todo o processo, as etapas seguintes se baseiam no seu resultado. Assim, é importante que o resultado desta etapa seja a mais precisa possível. Um cuidado que se deve tomar ao se fazer um estudo da geração de viagens é na definição das zonas de tráfego.

Uma série de características existentes em uma zona influencia o número de viagens. Deste modo, torna-se muito importante a elaboração de um zoneamento que agrupe regiões vizinhas com características semelhantes, formando zonas ou macrozonas vizinhas de tal forma que as características intrazonais sejam homogêneas, e as características interzonais sejam heterogêneas. O objetivo da geração de viagens é a previsão do número de viagens de pessoas que são produzidas e/ou atraídas para cada zona de tráfego da área em estudo.

A geração de viagens pode ser individual, familiar ou valores médios zonais. Os dados com nível de desagregação maior permitem uma melhor precisão na determinação do número de viagens geradas.

As viagens também podem ser classificadas por motivos (propósitos) que refletem as atividades desenvolvidas pelas pessoas para uma melhor análise. No estudo de geração de viagens é importante que as viagens sejam agrupadas em um número de categorias ou motivos, de acordo com o interesse do estudo e dos dados disponíveis. Os estudos mostram que as categorias mais aplicadas para o caso de viagens com base domiciliar são:

- Viagens para trabalho;
- Viagens para estudo;
- Viagens para compras;
- Viagens para recreação;
- Outras viagens.

Segundo NOVAES (1981) os modelos de geração de viagens são dois: os modelos de produção de viagens e os modelos de atração de viagens.

- **Modelos de produção de viagens** explicam o total de viagens produzidas numa zona em função das características socioeconômicas e do uso do solo encontrados nessa zona.

- **Modelos de atração de viagens** procuram explicar o influxo de pessoas ou mercadorias numa determinada zona em função das características socioeconômicas e do uso do solo da mesma.

De acordo com MELLO (1975), as variáveis consideradas de maior importância nos modelos de produção e atração de viagem são:

- Na produção: Renda; Propriedade de veículos; Número de residências; Números de pessoas empregadas; Número de pessoas em idade escolar; População.

- Na atração: Área destinada à indústria, comércio e outros; Número de empregos; Matrículas escolares.

De acordo com ORTÚZAR e WILLUMSEN (2001), várias técnicas foram propostas para modelar a geração de viagens desde o início da década de 50. Entre essas técnicas, podem ser citados os modelos de fator de crescimento, taxas de viagens, de classificação cruzada, escolha discreta e os de regressão linear.

Fator de crescimento: Determina o número de viagens futuras por zona de tráfego em função de variáveis que têm influência na geração das mesmas, tais como: população, renda, propriedade de veículos, densidade residencial ou comercial etc. Portanto, é um modelo que trabalha com dados agregados.

Taxas de viagens: Esse método consiste em relacionar dados do estudo do tráfego (viagens) com dados do uso do solo, assim estabelece uma taxa média de geração (produção e atração) de viagens para os principais usos do solo para cada zona de viagem. Segundo BRUTON (1975), esse método foi aplicado nos primeiros estudos feitos na área de transportes para estimar o crescimento de viagens em uma determinada zona de tráfego.

Classificação cruzada: Esse método consiste em classificar os dados das unidades domiciliares em subgrupos homogêneos formados por mais de uma variável, cada uma delas subdivididas em níveis. Para cada subgrupo será estimada uma taxa média de geração de viagens. Esse método é baseado na hipótese de que as taxas de geração de viagens para os diversos subgrupos permanecerão constantes no futuro BRUTON (1975). Conhecendo-se o número de domicílios e a taxa média de geração para cada subgrupo, podem ser obtidas estimativas da geração de viagens futura multiplicando-se a taxa média de geração de cada subgrupo pelo seu respectivo número de domicílios. Segundo BRUTON (1975), a deficiência do método está na ausência de meios para testar a significância estatística das variáveis escolhidas para representar as viagens.

2.4.3 - MODELOS DE DEMANDA DIRETOS OU SIMULTÂNEOS

Nos modelos de demanda diretos uma única equação explica mais de uma das etapas do modelo sequencial. Um exemplo é o modelo de Quandt e Baumol apresentado por PAPACOSTAS (2000), o qual modela a demanda de viagens interurbanas empregando variáveis do uso do solo variáveis socioeconômicas e variáveis do sistema de transportes. Esse modelo estima a demanda de viagem da origem i para o destino j pelo modo m, ou seja, esse modelo é um modelo direto que modela a geração de viagem, distribuição de viagens e escolha modal. É expresso por:

$$V_{ijm} = a_0 (P_i)^{a_1} (P_j)^{a_2} (C_{ij})^{a_3} \left(\frac{C_{ijm}}{C_{ij^*}}\right)^{a_4} (H_{ij^*})^{a_5} \left(\frac{H_{ijm}}{H_{ij^*}}\right)^{a_6} \left(\frac{D_{ijm}}{D_{ij^*}}\right)^{a_7} (Y_{ij})^{a_8} \quad , \qquad (1)$$

onde:

- V_{ijm}: quantidade de viagem da zona i, para zona j, pelo modo *m*;
- P_i e P_j: população de *i* e *j*;

- C_{ij*}: menor custo de viagem entre a zona *i* e *j*;
- C_{ijm}: custo do modo *m*;
- H_{ij*}: menor tempo de viagem entre a zona *i* e *j*;
- H_{ijm}: tempo de viagem via modo *m*;
- D_{ij*}: freqüência de partida do modo mais freqüente;
- D_{ijm}: frequência de partida do modo *m;*
- Y_{ij}: renda média ponderada de *i* e *j*;
- $a_{0}, \dots, a_{8}$: parâmetros a serem calibrados.

2. 5 – ANÁLISE DE REGRESSÃO

A análise de regressão linear é uma técnica estatística que pode ser usada para analisar a relação entre uma variável, dita dependente, e uma ou mais variáveis, ditas independentes ou preditoras. O objetivo da análise de regressão é prever as mudanças na variável dependente como resposta a mudanças nas variáveis independentes por meio da regra estatística dos mínimos quadrados. A regressão múltipla também pode ser usada para comparar dois ou mais conjuntos de variáveis para avaliar seu poder preditivo, comparando assim dois ou mais modelos alternativos ou concorrentes. Esta técnica pode ser aplicada em duas classes de problema de pesquisa: previsão e explicação (HAIR *et al.*, 2005).

Principais Objetivos do Emprego da Análise de Regressão:

De maneira geral, a análise de regressão pode ser utilizada com vários objetivos, dentre os quais é possível destacar:

- Descrição
- Predição
- Controle
- Estimação

2.5.1 – MODELO DE REGRESSÃO LINEAR SIMPLES

Análise de regressão linear simples é um método estatístico que utiliza a *relação* entre duas variáveis quantitativas (ou qualitativas) de tal forma que uma variável pode ser predita a partir da outra.

Considere o modelo com uma variável preditora e que a função de regressão é linear. O modelo é dado por:

$$\begin{aligned} &Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \\ &\mathrm{i} = 1,2,...,\mathrm{n}, \end{aligned} \tag{2}$$

onde:

- Y_i é o i-ésimo valor da variável resposta;
- β_0 e β_1 são os parâmetros (coeficientes de regressão);
- X_i é o i-ésimo valor da variável preditora (é uma constante conhecida, fixo).
- ε_i é o termo do erro aleatório com $E(\varepsilon_i) = 0$ e $\sigma^2(\varepsilon_i) = \sigma^2$, sendo σ^2 a variância.
- ε_i e ε_j não são correlacionados $\Rightarrow \sigma(\varepsilon_i, \varepsilon_j)=0$ para todo *i,j; i≠ j;* (covariância é nula).
- *i = 1,2,...,n.*

Os dados são usados para estimar β_0 e β_1, isto é, ajustar o modelo aos dados, para:

- Quantificar a relação entre Y e X;
- Usar a relação para predizer uma nova resposta Y_0 para um dado valor de X_0 (não incluído no estudo);
- Calibração – ou capacidade de predição de novas observações, pode ser feita usando uma nova amostra e comparando os valores estimados com os observados, dado um valor de Y_0, para o qual o correspondente valor de X_0 é desconhecido, estimar o valor de X_0.

Segundo NETER *et al.* (2005), a equação (2) é dita simples, pois apresenta a relação entre uma característica de qualidade e uma variável de controle, é linear quanto aos parâmetros, pois nenhum dos parâmetros aparece como expoente ou está sendo multiplicado ou dividido por outros parâmetros.

2.5.2 – MODELO DE REGRESSÃO MÚLTIPLA

As variáveis deste trabalho apresentam uma relação de dependência estatística e não de dependência determinística ou funcional como no caso das leis da física. Esse tipo de dependência é estudado na análise de regressão, que segundo GUJARATI (1995) "é um estudo da dependência de uma variável dependente em relação a uma variável, ou mais, explicativa com o objetivo de estimar ou prever a média ou valor médio da variável dependente provável conforme o valor assumido pelas variáveis explicativas.".

Quando está em questão apenas duas variáveis a análise de regressão será linear simples, enquanto que quando estão mais de duas variáveis, no caso deste trabalho, a análise de regressão será linear múltipla. Na análise de regressão simples é possível colocar em um gráfico os pontos de combinação entre a variável dependente e explicativa e traçar a reta de regressão, que representa um padrão de pontos.

Conforme DOWNING e CLARK (2000) o método de cálculo da reta de regressão busca encontrar uma reta em que o somatório dos erros - distância vertical entre o ponto e a reta de regressão - seja minimizado.

Supondo que temos $X_1, X_2, \ldots, X_{p-1}$ variáveis preditoras. Defini-se o modelo de regressão, com erros normais, em termos das variáveis preditoras:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \,, \quad (3)$$

onde:

- $\beta_0, \beta_1, \ldots, \beta_{p-1}$, são os parâmetros;
- $X_{i1}, \ldots, X_{i,p-1}$ são constantes conhecidas;
- ε_i são independentes com distribuição $N(0, \sigma^2)$
- $I = 1,2,\ldots,n.$

A função resposta para o modelo, como $E(\varepsilon_i)=0$, é dada por:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1} \,. \qquad (4)$$

2.5.2.1 *STEPWISE*

As técnicas de seleção de variáveis buscam determinar qual o melhor subconjunto de variáveis de entrada para compor um modelo. A técnica Stepwise (passo a passo) utiliza uma técnica de regressão linear múltipla para escolha de variáveis. O modelo começa com todas as variáveis do conjunto e remove de forma gradativa as que são estatisticamente menos significantes.

Esse processo ocorre até que as variáveis restantes sejam todas importantes (estatisticamente relevantes), ou seja, até que não haja melhora no desempenho do modelo ou não haja variáveis a serem retiradas. Essa técnica supõe que algumas variáveis não contribuem de forma significativa para a resposta de todo o conjunto (DEMUTH *et al.*, 2008). Após a retirada de uma variável, esta não poderá mais compor o modelo.
Segundo JUNIOR (2004) a aplicação dessas técnicas pode facilitar o trabalho de modelagem e melhorar os resultados obtidos.

2.5.3 – MÉTODO DOS MÍNIMOS QUADRADOS

O método mais usado para ajustar uma linha reta a um conjunto de pontos é conhecido como técnica dos mínimos quadrados. A reta resultante tem duas características importantes:

(1) A soma dos desvios verticais dos pontos em relação a reta é zero, e
(2) A soma dos quadrados desses desvios é mínima (isto é nenhuma outra reta daria menor soma de quadrados de tais desvios).

Simbolicamente o valor que é minimizado é

$$\sum(y_i - y_c)^2 \, , \qquad (5)$$

onde:

- y_i = um valor observado de y
- y_c = o valor calculado de y utilizando-se a equação de mínimos quadrados com o valor de x correspondente a y_i

Os valores de *a* e *b* para a reta $y_c = a + bx$ que minimiza a soma dos quadrados dos desvios são as soluções das chamadas “equações normais”:

$$\sum y = na + b\left(\sum x\right) \; e \qquad (6)$$

$$\sum xy = a\left(\sum x\right) + b\left(\sum x^2\right), \qquad (7)$$

onde n é o número de pares de observações. Assim, obtendo-se as quantidades $\sum x, \sum xy, etc.,$ pode-se resolver essas equações simultâneas em relação a *a* e *b*. Todavia, as equações podem ser resolvidas algebricamente em relação a a e b, e isto proporciona uma forma muito mais simples. O resultado consiste em duas fórmulas, uma para a e uma para b, usadas para fins de cálculo:

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \qquad (8)$$

$$a = \frac{\sum y - b\sum x}{n}. \qquad (9)$$

O Método dos Mínimos Quadrados, ou Mínimos Quadrados Ordinários (MQO) ou OLS (do inglês *Ordinary Least Squares*) é uma técnica de otimização matemática que procura encontrar o melhor ajustamento para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (tais diferenças são chamadas de resíduos).

É a forma de estimação mais amplamente utilizada na econometria. Consiste em um estimador que minimiza a soma dos quadrados dos resíduos da regressão, de forma a maximizar o grau de ajuste do modelo aos dados observados.

2.5.4 – COEFICIENTE DE DETERMINAÇÃO R^2

O coeficiente de determinação R^2 pode ser interpretado como a proporção da variabilidade que pode ser estimada pela equação da regressão. Todavia, quando multiplicado por 100, o coeficiente de determinação múltipla pode ser interpretado como a porcentagem da variabilidade em y que pode ser explicada através da equação da regressão. (ANDERSON *et al.*, 2002; SUBRAMANIAN *et al.*, 2007).

2.5.5 – CORRELAÇÃO DE PEARSON

O coeficiente de correlação de Pearson mede o grau da correlação (e a direção dessa correlação - se positiva ou negativa) entre duas variáveis de escala métrica (intervalar ou de razão).

Este coeficiente, normalmente representado por ρ assume apenas valores entre -1 e 1.

• ρ = 1 Significa uma correlação perfeita positiva entre as duas variáveis.

• ρ = – 1 Significa uma correlação negativa perfeita entre as duas variáveis - Isto é, se uma aumenta, a outra sempre diminui.

• ρ = 0 Significa que as duas variáveis não dependem linearmente uma da outra. No entanto, pode existir uma dependência não linear. Assim, o resultado ρ = 0 deve ser investigado por outros meios.

Calcula-se o coeficiente de correlação de Pearson segundo a seguinte fórmula:

$$\rho = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}.\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{cov(X,Y)}{\sqrt{var(X).var(Y)}}\,, \qquad (10)$$

onde $x_1, x_2, \dots, x_n$ e $y_1, y_2, \dots, y_n$ são os valores medidos de ambas as variáveis.

sendo que:

$$\bar{x} = \frac{1}{n}.\sum_{i=1}^{n} x_i \qquad (11)$$

e

$$\bar{y} = \frac{1}{n}.\sum_{i=1}^{n} y_i \qquad (12)$$

são as médias aritméticas de ambas as variáveis.

A análise correlacional indica a relação entre duas variáveis lineares e os valores sempre serão entre +1 e -1. O sinal indica a direção, se a correlação é positiva ou negativa, e o tamanho da variável indica a força da correlação.

Interpretando ρ (FRANZBLAU, 1958)

- Se $| \rho | < 0,20$, a correlação é negligenciável.
- Se $0,20 < | \rho | < 0,40$, a correlação é fraca.
- Se $0,40 < | \rho | < 0,60$, a correlação é moderada.
- Se $0,60 < | \rho | < 0,80$, a correlação é forte.
- Se $| \rho | > 0,80$, a correlação é muito forte.

2.6 – MODELO DE REGRESSÃO ESPACIAL

Dados espaciais por sua vez são caracterizados no espaço em função de um sistema de coordenadas absolutas ou relativas. Segundo TEIXEIRA (2003), pode-se definir analise espacial como qualquer sistema que torna possível a apresentação, manipulação, análise, inferência e estimação de dados espaciais. Os princípios básicos da análise espacial consistem em entender de que forma os dados procedentes dos fenômenos ocorridos no espaço se organizam e qual a relação existente entre eles (HENRIQUE, 2004).

Devido à relação entre transporte e espaço, as técnicas que podem ser as mais adequadas para a previsão da demanda de passageiros são as que utilizam análise espacial. O primeiro passo é escolher o modelo inferencial a ser utilizado. A hipótese mais comum é supor que as áreas são diferenciadas e que cada uma delas possui uma “identidade” própria.

Alguns trabalhos que são exemplo da utilização da análise espacial podem ser encontrados nos seguintes pesquisas: QUEIROZ (2003), onde o autor analisa geograficamente os dados de acidentes de trânsito em Fortaleza – CE; SANTOS (2006), que realiza um estudo com os dados de acidentes de trânsito na cidade de São Carlos através de SIG e estatística espacial; KREMPI (2004) analisou a acessibilidade da cidade de Bauru – SP e PERINI (2008), que por sua vez realizou um diagnóstico espacial de acessibilidade da cidade de Vitória – ES.

Segundo CÂMARA *et al.* (2002), um aspecto fundamental na análise exploratória espacial é a caracterização da dependência espacial, que mostra como os valores estão correlacionados no espaço.

2.6.1 – MODELOS DE REGRESSÃO ESPACIAL

Dados espaciais agregados são caracterizados pela dependência (autocorrelação espacial) e pela heterogeneidade ou estrutura espacial (ANSELIN, 1988). Esses efeitos espaciais são importantes, pois, em alguns casos, são os principais responsáveis pela realização dos eventos. Entretanto, invalidam os resultados dos modelos tradicionais de regressão, por violarem alguns pressupostos como a independência e a homocedasticidade. Assim, pela necessidade de se incorporar tais fenômenos à estrutura de um modelo é que foram desenvolvidos os modelos de regressão espacial ou *spatial econometric models* como são conhecidos na literatura.

A construção de modelos ou modelagem envolve a formulação, o ajuste e o diagnóstico do modelo de uma maneira iterativa e interativa (CHATFIELD, 1995). A formulação envolve considerações do problema em estudo, hipóteses, teorias. Isto indicará as possíveis variáveis que entrarão no modelo e, também, indicará restrições nos parâmetros e variáveis. O diagnóstico do modelo é uma etapa fundamental da modelagem. Nesta etapa verifica-se o ajuste do modelo e se as suposições acerca do modelo são satisfeitas. Técnicas gráficas são as indicadas ou preferenciais. Se necessário o modelo é modificado e um novo modelo é ajustado, isto indica que o processo é iterativo. Como existe a participação ativa do analista, o processo também é interativo.

No caso de dados espaciais, onde está presente a dependência espacial, é pouco provável que o pressuposto padrão de observações não correlacionadas seja verdadeiro. No caso mais comum, os resíduos continuam apresentando a autocorrelação espacial presente nos dados, que pode se manifestar por diferenças regionais sistemáticas nas relações do modelo, ou ainda por uma tendência espacial contínua (CÂMARA *et al.*, 2002).

CÂMARA *et al.* (2002) comentam que essa dependência é uma característica inerente à representação dos dados através de subdivisões territoriais, ou seja, os dados de uma determinada área tendem a ser mais parecidos com os de seus vizinhos do que com os de áreas mais distantes. Segundo TOBLER (1979): "tudo está relacionado a tudo, mas as coisas mais próximas estão mais relacionadas que as coisas mais distantes". Vale ressaltar que o termo "vizinho" está baseado no padrão espacial adotado: geográfico (fronteiras, distância, etc.) ou conectividade (tempo de viagem, trocas comerciais, etc.).

Os modelos de regressão espacial também necessitam dos três principais pressupostos do modelo de regressão convencional, porém ao incorporar em sua estrutura o fator "espaço", eliminam, na maioria das vezes, os problemas de autocorrelação e heterocedasticidade mencionados anteriormente. Isso porque esse último ocorre, dentre outros motivos, devido à ausência de variáveis, sejam observáveis ou não observáveis como é o caso do "espaço".
As três hipóteses (ou pressupostos) necessárias para validação de um modelo de regressão são:

- os erros são normais com média zero;
- os erros têm variância constante (homocedasticidade);
- os erros são não correlacionados.

- **Modelo Espacial Autoregressivo de Primeira Ordem**

Esse modelo também conhecido como *First-order spatial AR model* (FAR) é utilizado quando se deseja explicar a variável dependente *y* a partir de seus vizinhos. Note que a dependência espacial está na própria variável *y*. Ou seja no caso dos modelos espaciais, a variável *y* é explicada pelos seus "vizinhos". CÂMARA *et al.* (2002)

- **Modelo Espacial Autoregressivo Misto**

Também conhecido como *Spatial Autoregressive Model* (SAR) ou *Spatial Lag Models* é utilizado quando se deseja explicar a variável dependente *y* a partir dela mesma e de outras variáveis explicativas.

Verifique que se o parâmetro espacial ρ for zero, então o modelo resultante é exatamente igual a um modelo de regressão convencional.

Quando o valor de ρ estiver próximo de zero (baixa dependência espacial), pouca informação será agregada a β , enquanto que se estiver próximo de +1 ou -1 (alta dependência espacial), um valor significativo será agregado a β . Esse fato explica o porquê da regressão espacial "corrigir" os parâmetros do modelo, quando comparada à regressão convencional. CÂMARA *et al.* (2002)

- **Modelo com Erro Espacial Autoregressivo**

Esse modelo também conhecido como *Spatial Error Model* (SEM) tem a mesma função do modelo SAR, porém a estrutura espacial está no erro aleatório.

Na prática, a distinção entre esses dois modelos é difícil. No entanto, o modelo SEM tem mais indícios de ser utilizado quando o resíduo resultante de um modelo de regressão convencional possui dependência espacial, constatada pelo *I* de Moran. Outros fatores que corroboram a utilização desse modelo são os erros de medida, ausência de variáveis explicativas ou variáveis não observáveis, além da heterocedasticidade.

2.6.1.1 - SAR (SPATIAL AUTO REGRESSIVE) OU SPATIAL LAG MODEL

No modelo SAR (ou LAG, como é denominado neste estudo) a autocorrelação espacial ignorada é atribuída à variável dependente Y. Considera-se a dependência espacial através da adição, ao modelo de regressão, de um novo termo na forma de uma relação espacial para a variável dependente. Formalmente, ANSELIN (2002) apresenta o modelo SAR pela Equação 13.

$$y = \rho WY + X\beta + \varepsilon , \qquad (13)$$

onde:

- Y = variável dependente;
- X = variáveis independentes;
- ρ = coeficientes de regressão;

- ε = erros aleatórios com média zero e variância σ^2;
- W = matriz de vizinhança espacial ou matriz de ponderação espacial;
- β = coeficiente espacial autorregressivo.

A hipótese nula para a não existência de autocorrelação é que λ = 0. A idéia básica é incorporar a autocorrelação espacial como componente do modelo.

2.6.1. 2 - CAR (CONDITIONAL AUTO REGRESSIVE) OU SPATIAL ERROR MODEL

Outro tipo de modelo de regressão espacial com parâmetros globais, também referido como *Spatial Error Model*, considera que os efeitos espaciais são um ruído, ou perturbação, ou seja, fator que precisa ser removido. Neste caso, os efeitos da autocorrelação espacial são associados ao termo de erro e o modelo pode ser expresso pela Equação (14).

$$Y = X\beta + \varepsilon, \quad \varepsilon = \lambda W\varepsilon + \xi, \quad (14)$$

onde:

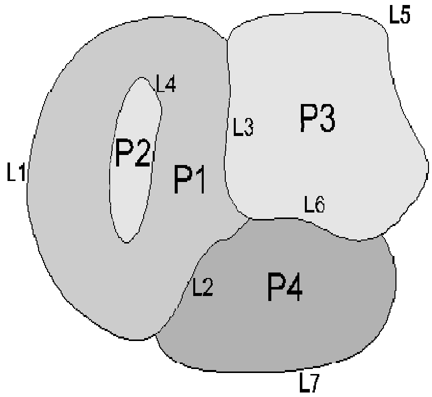
- W_ε = erros com efeito espacial;
- ε = erros aleatórios com média zero e variância σ^2;
- λ = coeficiente autoregressivo.

A hipótese nula para a não existência de autocorrelação é que $\rho = 0$, ou seja, o termo de erro não é espacialmente correlacionado. CÂMARA *et al.* (2002a) salientam que, na prática, a distinção entre os dois tipos de modelos de regressão espacial com parâmetros globais é difícil, pois, apesar da diferença nas suas motivações, eles são muito próximos em termos formais.

2.6.2 - MATRIZ DE PROXIMIDADES

A matriz de proximidade é utilizada em cálculos de indicadores na fase de análise exploratória dos dados espaciais. A matriz de proximidade espacial é a responsável pela estrutura espacial, usualmente denominada “matriz W” e também chamada

Matriz de Vizinhança. Dado um conjunto de *n* áreas {*A1*,..,*An*}, construímos a matriz *W*(1) (*n* x *n*), onde cada um dos elementos *wij* representa uma medida de proximidade espacial entre o polígono *Ai* e o polígono *Aj,* sendo ***n,*** o número total de objetos. Na Figura a seguir pode-se observar um exemplo ilustrativo de matriz de proximidade.



$$W = \begin{pmatrix} W_{11} & W_{12} & W_{13} & W_{14} \\ W_{21} & W_{22} & W_{23} & W_{24} \\ W_{31} & W_{32} & W_{33} & W_{34} \\ W_{41} & W_{42} & W_{43} & W_{44} \end{pmatrix}$$

w_{ij} : “distância” do objeto ***i*** ao objeto ***j.***

FIGURA 1: Exemplo de uma Matriz de Proximidade.

Costuma-se padronizar as linhas da matriz W, criando uma nova matriz assimétrica, a fim de facilitar a derivação de fórmulas e as propriedades estatísticas envolvidas. A padronização consiste em fazer com que a soma da linha *i* seja igual a 1.

O vetor de médias ponderadas (*Wz*) é obtido pela multiplicação do vetor transposto dos desvios, pela matriz de proximidade espacial com linhas normalizadas, onde cada elemento de uma linha i qualquer originariamente com valor 1, é dividido pelo número de elementos não nulos da mesma linha. Desta forma, como resultado, cada elemento wz_i, contém um valor correspondente à média dos desvios dos vizinhos ao objeto i. A matriz W, por si própria, não fornece informações que indiquem a presença de dependência espacial. Serve apenas para indicar a estrutura espacial existente.

2.7 – MODELO DE REGRESSÃO GEOGRAFICAMENTE PONDERADA

A regressão linear geograficamente ponderada (RGP) foi inicialmente proposta por BRUNSDON *et al.* (1996) como um método para explorar a não estacionariedade espacial, sendo esta última uma condição na qual um modelo de regressão espacial global não pode explicar adequadamente as relações entre alguns conjuntos de variáveis definidas numa região geográfica. Deste modo, uma solução mais adequada seria modificar a natureza do modelo para refletir, ao longo do espaço, a estrutura intrínseca aos dados.

A idéia básica da RGP é ajustar um modelo de regressão para cada ponto no conjunto de dados, ponderando as observações por uma função de distância a este ponto. Isto corresponde a considerar que pontos mais próximos ao ponto em estudo tenham maior influência nos parâmetros estimados da regressão do que observações obtidas em pontos mais distantes.

Como resultado, obtêm-se, portanto, um conjunto de parâmetros ajustados para cada ponto na região geográfica analisada. A proposta inicial tinha como objetivo fornecer uma ferramenta útil de análise exploratória através da identificação das variações entre variáveis ao longo do espaço, o que corresponde a não estacionaridade espacial; entretanto, por auxiliar na explicação e previsão de fenômenos geográficos, este tipo de modelo pode ser classificado como confirmatório, desde que, conforme discutido no trabalho de LOUREIRO *et al.* (2006), seja avaliada com o devido cuidado sua premissa básica de influência do comportamento de uma determinada unidade geográfica sobre as suas vizinhas, e vice-versa. A seguir, apresenta-se uma descrição mais detalhada deste modelo, bem como considerações sobre os diversos tipos de ponderações geográficas passíveis de utilização na RGP.

2.7.1 - CONCEPÇÃO DO MODELO RGP

Considerando inicialmente um modelo convencional de regressão linear multivariável não-espacial, no qual uma variável dependente y é representada como uma combinação linear de variáveis explicativas x_k, pode-se obter uma estimativa de y para um ponto i a partir da equação de regressão:

$$y_i = a_0 + \sum a_k * X_{k_i} + \varepsilon_i , \quad (15)$$

em que,

X_{k_i} : k-ésima variável explicativa na localização i;

a_0 : intercepto global (valor constante);

a_k : k-ésimo coeficiente angular global referente a X_k (valor constante);

ε_i : erros independentes.

Para estimativa dos coeficientes angulares e do intercepto em um modelo de regressão linear convencional, normalmente utiliza-se o método dos mínimos quadrados simples, resultando na seguinte equação dos estimadores em notação matricial:

$$\hat{A} = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y, \qquad (16)$$

em que,

- Â: vetor dos coeficientes (intercepto e coeficientes angulares);
- X: matriz de observações das variáveis explicativas;
- Y: vetor de observações da variável dependente (explicada).

Vale observar, portanto, que para um conjunto de observações é obtido o mesmo conjunto de estimativas para os coeficientes na Equação 15. Logo, se as observações forem coletadas para um conjunto de pontos, os estimadores obtidos serão constantes ao longo do espaço. A regressão geograficamente ponderada (RGP) consiste numa técnica simples de extensão da regressão convencional permitindo que parâmetros locais – em vez de globais – sejam estimados (FOTHERINGHAM *et al.*, 1997). Deste modo, as estimativas tornam- se específicas para cada localização i, e o modelo pode então ser reescrito como:

$$y_i = a_{0,i} + \sum_{k=1}^{p} a_{k,i} \cdot x_{k,i} + \varepsilon_i, \quad (17)$$

em que,

- $a_{0,i}$: intercepto local específico da localização i;
- $a_{k,i}$: valor do k-ésimo coeficiente angular local específico da localização i.

Pode-se notar, a partir das Equações 15 e 17, que a principal diferença entre os modelos está na extensão e generalidade obtida para cada localização *i*, passando de observação para um modelo específico, particular. Assumindo que os parâmetros exibam algum grau de consistência espacial, pode-se considerar, por exemplo, um subconjunto de localizações próximas a que está sendo estudada como fornecedor de informações adicionais. Como este tipo de distribuição espacial de valores é o mais comumente encontrado, com valores pertencentes a localizações próximas tendo grandezas e sinais semelhantes, então este será usado ao longo da discussão.

Uma abordagem geral, ao se tratar este tipo de modelo, é notar que, apesar da obtenção de estimativas não tendenciosas ser impossível — já que os coeficientes mostrarão um desvio ao longo do subconjunto local de calibração — pode-se ainda conseguir uma tendenciosidade pequena. De fato, como FOTHERINGHAM *et al.* (1997) comentam, o processo de modelagem em RGP pode ser visto como um equilíbrio de troca entre tendenciosidade e erro padrão.

Quanto maior for o subconjunto local de calibração, menores serão os erros padrões das estimativas dos coeficientes; no entanto, deve-se atentar para o fato de que quanto maior o subconjunto local, maior a probabilidade de que os desvios introduzidos pela região espacial acrescentem tendenciosidade. Uma forma de reduzir tal efeito é considerar uma função de ponderação espacial, de modo que localizações mais distantes tenham menor influência na calibração do modelo em estudo.

De um modo geral, no método de mínimos quadrados ponderados, fatores de ponderação são aplicados em cada diferença ao quadrado antes da minimização, de forma que a imprecisão de alguns preditores receba maior penalidade que outros. Os estimadores ponderados para a Equação 15 podem ser então escritos como:

$$\hat{A} = \left(X^T \cdot W \cdot X\right)^{-1} \cdot X^T \cdot W \cdot Y, \quad (18)$$

em que W é uma matriz quadrada cujos elementos da diagonal principal correspondem aos pesos de cada observação e os outros elementos são nulos, ou seja:

$$W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix} \quad (19)$$

Vale notar que, no modelo de regressão não espacial (Equação 15), os pesos atribuídos a cada observação são também constantes, e logo a estimativa dos parâmetros é a mesma para todo o espaço — estimativa global. Na RGP, os quadrados mínimos ponderados são utilizados para ponderação de cada observação de acordo com sua proximidade ao ponto i, de forma que, como no modelo representado pela Equação 17, os estimadores variem de acordo com a localização do ponto i em estudo pela variação de W.

Desta forma, o estimador para os parâmetros da Equação 17 é semelhante ao da Equação 18, levando em consideração, contudo, que a matriz de ponderação depende de i. Logo, tem-se:

$$\hat{A} = \left(X^T \cdot W_i \cdot X\right)^{-1} \cdot X^T \cdot W_i \cdot Y, \quad (20)$$

em que, analogamente, W_i é uma matriz quadrada cujos elementos da diagonal principal correspondem aos pesos de cada observação com relação ao ponto i e os elementos restantes são nulos. De modo semelhante a $\hat{A}_i$, considerando-se agora pesos para cada ponto i, deve- se estender também a notação. Por $W_{i,j}$ denota-se o peso atribuído à observação (ponto) j em relação ao ponto i. Logo,

$$W = \begin{bmatrix} w_{i,1} & 0 & \ldots & 0 \\ 0 & w_{i,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{i,n} \end{bmatrix} \quad (21)$$

2.7.1.1 - FUNÇÕES DE PONDERAÇÃO ESPACIAL E PROCESSO DE CALIBRAÇÃO DA RGP

Pode-se notar pela Equação 20 que, uma vez fornecidos os parâmetros de entrada X e Y, basta a definição das matrizes W_i para completar o modelo de regressão geoponderada. Isto pode ser feito através da escolha adequada de funções de ponderação espacial, como já anteriormente mencionado. Considerando inicialmente o modelo de regressão não-espacial estabelecido pelas Equações 15 e 16, este pode ser visto como um caso particular da Equação 20, na qual:

$$w_{i,j} = 1, \qquad \forall i, j, \qquad (22)$$

onde *j* representa um ponto específico no espaço no qual os dados são observados e *i* um ponto no qual os parâmetros são estimados. Aqui todas as observações têm peso unitário, o que corresponde a uma estimativa constante dos parâmetros ao longo do espaço, ou seja, ao modelo de regressão global já discutido anteriormente.
Uma forma de considerar pesos consistentes com a proximidade e, portanto, um modelo de características locais, seria excluir do processo de calibração local observações que estejam mais distantes do ponto em estudo do que uma determinada distância de influência ($D_{influencia}$). Isto seria equivalente a fornecer valores nulos para tais observações, o que corresponde à função de ponderação:

$$w_{i,j} = \begin{cases} 1 & se\, d_{i,j} \langle \quad D_{\inf luencia} \\ 0 & caso \quad contrário \end{cases}, \ \forall i, j, \qquad (23)$$

em que $d_{i,j}$ representa a distância entre os pontos *i* e *j* (FOTHERINGHAM *et al.*, 1997). Embora a consideração de $w_{i,j}$, de acordo com a Equação 23, forneça um modelo de regressão local e geograficamente ponderada, este sofre do problema de descontinuidade. Como FOTHERINGHAM *et al.* (1997) observam, à medida que *i* varia ao longo da região em estudo, os coeficientes das regressões podem mudar drasticamente, já que os pontos de observação podem estar incluídos ou não na região circular em torno de *i*, ou seja, o modelo é fortemente dependente do parâmetro $D_{influencia}$. Uma forma de contornar tal problema é especificar $w_{i,j}$ como uma função contínua da distância $d_{i,j}$. Uma escolha razoável seria considerar uma curva gaussiana:

$$w_{i,j} = e^{\left(-\beta \cdot d_{ij}^2\right)}, \qquad (24)$$

em que β é um parâmetro de decaimento de acordo com a distância. A vantagem da consideração de funções contínuas está na inclusão de contribuições fracionárias das observações, de modo que pontos mais distantes tenham contribuições menos significativas.

Vale observar que pesos fracionários não alteram o valor da observação, mas sim sua influência no processo de calibração do modelo. Para observações em pontos muito distantes de i, os pesos são praticamente nulos, efetivamente excluindo tais observações. Uma forma de conciliar as funções de ponderação, representadas nas Equações 23 e 24, é estimar β com base no valor crítico $D_{influência}$ e um valor de tolerância e.

$$\beta = \frac{-\ln(e)}{d_{\inf luencia}^2}, \qquad (25)$$

onde e representa a proximidade a zero desejada. Um valor razoável seria e = 0,0001 que corresponde a atribuição de peso virtualmente igual a zero para as observações do ponto *j* quando este dista linearmente de $D_{influência}$ do ponto *i* em análise. O problema agora da calibração do modelo RGP reside na escolha adequada de valores para β ou, alternativamente, para $D_{influência}$. Na medida em que β tende a zero — ou, equivalentemente, $D_{influência}$ tende a maior distância entre *i* e *j*

— os pesos tendem ao valor unitário para todos os pares *i,j* de pontos, de forma que os parâmetros estimados tornam-se uniformes, resultando em um modelo global.

Por outro lado, à medida que *β* torna-se maior, as estimativas dos parâmetros dependerão cada vez mais das observações mais próximas de *i*, e, conseqüentemente, terão maior variância. FOTHERINGHAM *et al.* (2000) sugerem um processo de validação cruzada, no qual são utilizados os valores,

$$SQE = \sum_{i=1}^{n} \left[y_i - \hat{y}_{\neq i} \cdot (\beta) \right]^2, \quad (26)$$

em que,

- SQE : soma dos quadrados dos erros;
- y_i : valor observado da variável explicativa no ponto *i*;
- $\hat{y}_{\neq i}$: valor ajustado de y_i de acordo com (β ou $D_{influência}$) com o ponto *i* omitido do processo de calibração.

A escolha final destes parâmetros (*β* ou $D_{influência}$) geralmente será baseada na busca da minimização de SQE do fenômeno em estudo. É importante ainda chamar a atenção que SQE, embora utilizado no modelo local RGP, é um parâmetro global que, em função desta característica, servirá para comparar globalmente o nível de desempenho da RGP em relação aos modelos convencionais e espaciais globais.

Além disso, a associação deste parâmetro com a variabilidade total da variável dependente *y* permite avaliar coeficientes globais de determinação (R^2) que também servirão para comparações.

Vale também destacar que, por questões de convergência do modelo RGP, o ponto *i* em análise é excluído da calibração de sua própria reta de regressão. A não consideração desta medida resultará na minimização de SQE para uma faixa de valores demasiadamente elevados para *β* (ou, equivalentemente, para uma faixa de valores de $D_{influência}$ demasiadamente pequenos), produzindo uma função de ponderação (Equação 24) bastante acentuada, de tal forma que os pesos atribuídos aos pontos observados j's serão virtualmente iguais a zero, não exercendo influência sobre o ponto em análise i e fazendo com que o valor predito y_i dependa exclusivamente da observação do ponto i, o que obviamente não é desejado.

2.8 – MODELO DE REGRESSÃO DUMMY

Para introduzir variáveis categóricas no modelo é preciso que sejam criadas uma ou mais variáveis assumindo valores numéricos, que representem as categorias da variável categórica considerada. Essas variáveis criadas são chamadas variáveis *dummies*.

Para uma variável nominal A com k categorias (A1, A2,..., AK) ser representada em um modelo, é preciso serem criada (K-1) variáveis D1, D2, ..., DK-1, assumindo quaisquer dois valores numéricos distintos, que neste capitulo será considerada por conveniência os valores 0 e 1, de forma que, para i = 1, 2, ..., k-1.

tenha-se;

D_i = 1, se a unidade amostral considerada pertence a categoria A_i;

D_i = 0, se a unidade amostral considerada pertence a categoria A_j.

2.8.1 - USO DE VARIÁVEL *DUMMY* EM MODELOS DE DEMANDA DE VIAGEM

Segundo MELO (1975), podem-se utilizar variáveis *dummies* na elaboração de funções de regressão para a determinação do número de viagens produzidas ou atraídas em uma zona de tráfego. Como exemplo, tem-se o modelo de produção de viagem apresentado por Heathington e Isibor, com a seguinte equação de regressão:

$$Y = b_0 + b_1x_1 + b_{1i}Z_{1i} + b_{2i}Z_{2i} + b_{3i}Z_{3i} + b_{1r}Z_{1r} + b_{2r}Z_{2r} + b_{1c}Z_{1c} + b_{2i}Z_{2i}, \quad (27)$$

onde:

Y: Número de viagens geradas por unidade residencial;

X1: Tamanho da família;

Z_{ji}: Classe de renda j, j = 1 a 3;

Z_{1r}: Tipo de residência 1;

Z_{2r}: Tipo de residência 2;

Z_{1c}: Número de carros próprios por residência, tipo 1;

Z_{2c}: Número de carros próprios por residência, tipo 2.

Observa-se que a renda é classificada em três categorias, o tipo de residência em duas categorias e o número de carros próprios por residência em duas categorias. As variáveis *dummies* assumem valores um, se pertencer à categoria e zero se não pertencer à categoria.

CAPÍTULO 3

3 – METODOLOGIA

Este capítulo apresenta a metodologia do trabalho, destacando os métodos propostos para identificação de dois modelos: modelo de regressão espacial (veja seção 2.6) e modelo de regressão geograficamente ponderada (veja seção 2.7).

O método de procedimento utilizado neste trabalho é o que se fundamenta em teorias estatísticas e na técnica de pesquisa descritiva, visto que se observa a realidade (os modelos de regressão na previsão de demanda por viagens) sem manipulá-las. O método de abordagem do trabalho é hipotético dedutivo.

A metodologia para testar a adequação do modelo proposto consiste em utilizar os dados da pesquisa O-D, realizada no município de Vitória-ES em 1998, em seguida estima para 2007, matrizes O-D, matrizes estas que, através de estatísticas apropriadas, serão comparadas com as matrizes O-D observadas na pesquisa realizada no município de Vitória-ES em 2007.

As variáveis consideradas neste estudo relacionadas às macrozonas são: população, área, densidade populacional, número de automóveis, renda média, oferta de empregos, oferta de matrículas escolares, população de estudantes e população de ocupados. Os resultados foram obtidos com base em informações nas pesquisas O-D 1998 e 2007, respectivamente.

3.1 - APLICAÇÃO DOS MODELOS

A aplicação dos modelos será realizada em 2 etapas: Para validar o modelo proposto foram calibrados os modelos de regressão espacial e regressão geograficamente ponderada, com os dados do município de Vitória-ES, da pesquisa origem/destino (O-D) realizada na região da Grande Vitória em 1998.

Dessa forma, este capítulo apresenta o método proposto para a identificação, construção e análise de um modelo de regressão espacial e de regressão geograficamente ponderada. A metodologia utilizada na construção dos modelos está nessa seção. A seção posterior descreve com detalhes cada etapa necessária para atingir os objetivos propostos.

A metodologia padrão para a construção de modelos de regressão, esquematizada por OGLIARI (2004), é apresentada na Figura 2. Verifica-se a necessidade de 4 etapas até a consolidação do modelo final coleta e preparação dos dados, redução do número de variáveis regressoras, refinamento e seleção de modelos e validação do modelo.

As Etapas são as seguintes:

- ETAPA 1: Coleta e preparação dos dados
- ETAPA 2: Redução do número de Variáveis Regressoras
- ETAPA 3: Refinamento e seleção de modelos
- ETAPA 4: Validação do modelo

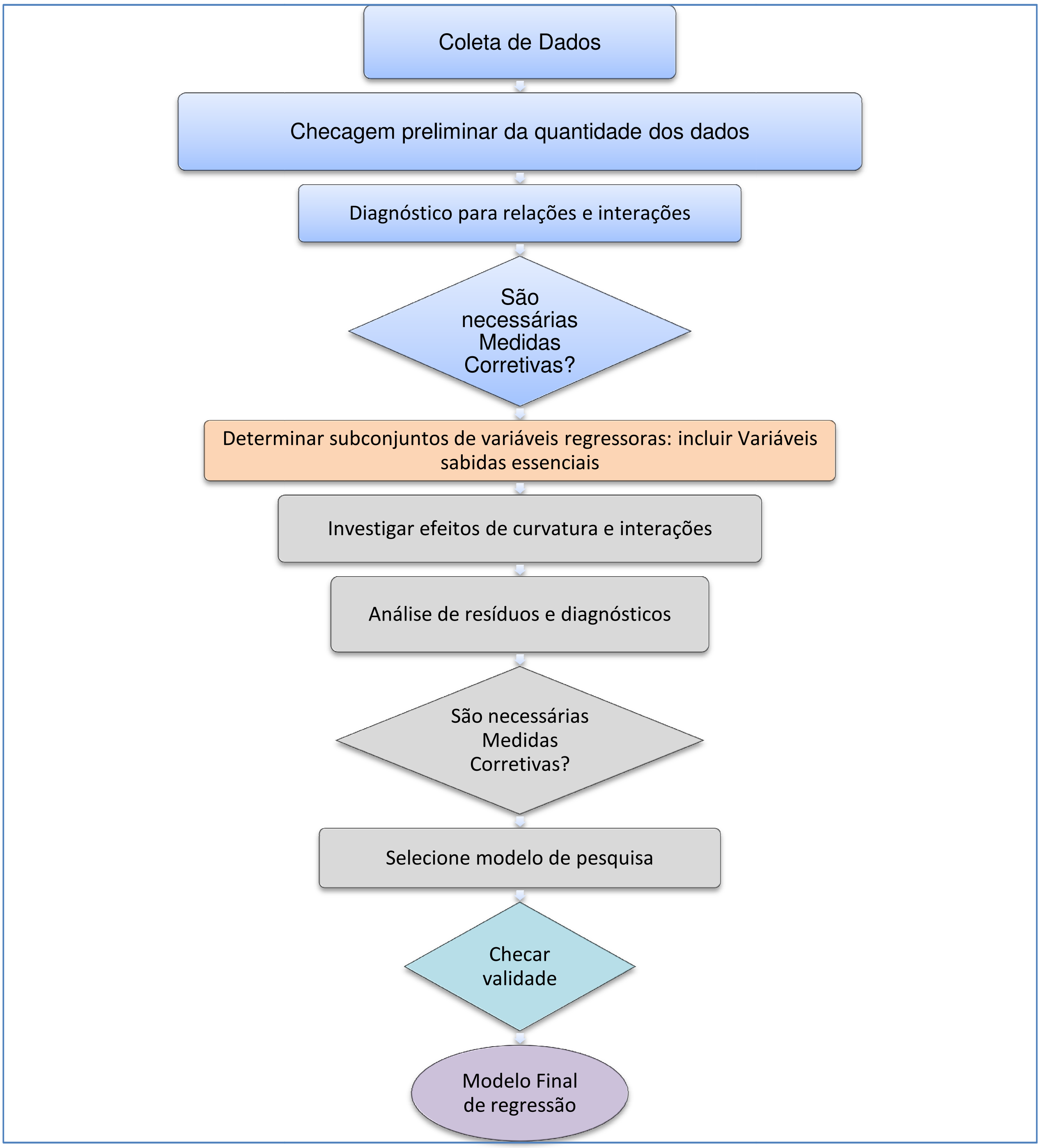


Figura 2: Fluxograma para Construção de Modelos de Regressão.

Fonte: OGLIARI (2004).

3.2 - ÁREA DE ESTUDO

Para a realização deste estudo, foi utilizada a pesquisa origem-destino desenvolvida em 1998 e 2007, no município de Vitória-ES. Nessa pesquisa elaborada em 1998 houve a divisão em 39 zonas, enquanto que na pesquisa do ano de 2007 dividiu-se em 85 zonas, as quais foram agrupadas em 13 macrozonas. Pelo fato das zonas de estudo em 1998 e 2007 não serem as mesmas, houve a necessidade de tornar compatível os dados de ambas as pesquisas, de forma que as macrozonas de 1998 ficassem iguais as de 2007.

A Figura 3 mostra como o município fica dividido em macrozonas.

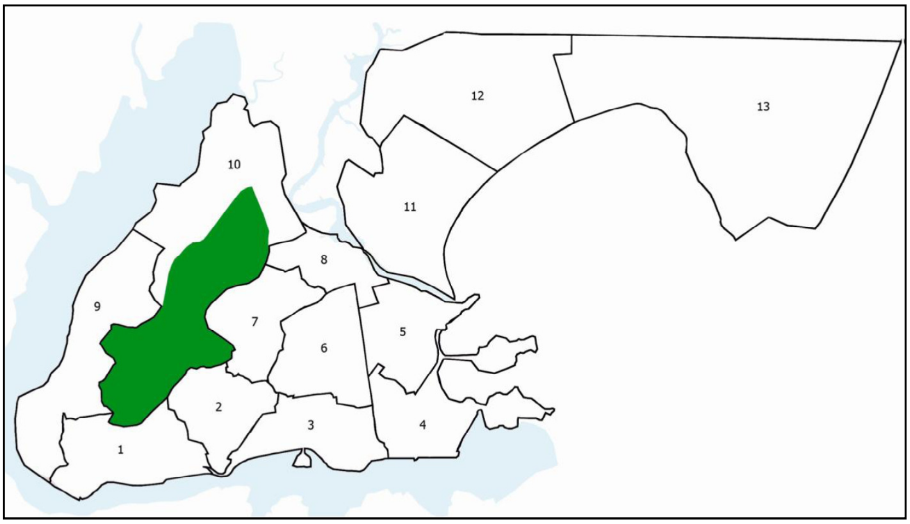


Figura 3: Mapa - Área de estudo município de Vitória-ES.

As viagens, a serem estudadas consistem apenas naquelas realizadas por pessoas ao longo do dia com origem em residências e destino qualquer no município de Vitória-ES, pelo modo motorizado, que se classificam em modo coletivo: (viagens feitas a ônibus, van, barco) e modo individual: (viagens feitas de automóvel, caminhão e motocicleta) não sendo consideradas as viagens feitas a pé. E os propósitos de viagens a serem considerados são: residência para trabalho, residência para estudo e residência para outros.

Para uma melhor visualização das variáveis, nos Quadros 2 e 3 estão expostos os valores das variáveis socioeconômicas do município de Vitória-ES, para o ano de 1998.

Macrozona	População	Área em Km^2	Densidade Pessoas/km^2	Automóvel	Renda	Oferta de Emprego	Oferta de Matriculas Escolares	População de Estudantes	População de Ocupados
1	31.729	4,31	7.362,06	11.272	4,01	49.141	19.988	9.090	12.028
2	10.934	2,87	3.805,49	2.838	3,24	7.742	11.394	2.569	4.050
3	17.916	4,31	4.157,01	5.163	4,03	17.697	9.576	5.102	6.873
4	6.475	5,75	1.126,74	4.887	5,79	11.008	3.757	1.854	2.720
5	16.722	4,31	3.879,92	6.523	6,71	20.936	11.664	5.104	6.722
6	35.309	4,31	8.192,58	3.390	2,41	6.505	3.708	9.435	13.463
7	7.131	4,31	1.654,69	532	3,39	685	810	2.205	2.721
8	7.367	2,87	2.563,86	808	1,95	2.791	1.326	2.072	2.734
9	24.209	7,18	3.370,31	1.379	1,89	2.528	3.585	7.105	8.069
10	26.111	7,18	3.635,04	808	0,98	2.317	4.690	9.115	8.580
11	37.299	7,18	5.192,68	7.370	4,79	10.745	8.668	13.245	13.491
12	16.670	10,06	1.657,62	296	2,24	2.844	3.880	4.057	5.703
13	32.039	20,11	1.593,00	3.035	3,17	8.743	6.360	8.971	11.534
Total	269.911	84,76		48.301		143.683	89.405	79.925	98.688

Quadro 2 – Variáveis Socioeconômicas do Município de Vitória – ES no ano de 1998.

Fonte: BRAGA (2009)

Macrozona	População	Área em Km^2	Densidade Pessoas/km^2	Automóvel	Renda (S. M)	Oferta de Emprego	Oferta de Matriculas Escolares	População de Estudantes	População de Ocupados
1	37.434	4,31	8.685,78	7.568	3,19	44.652	6.022	6.300	10.266
2	12.900	2,87	4.489,73	7.703	3,12	7.356	6.471	5.305	12.140
3	21.137	4,31	4.904,46	9.189	3,73	27.746	17.523	8.295	8.668
4	7.639	5,75	1.329,33	9.595	5,58	21.681	4.236	3.938	8.813
5	19.728	4,31	4.577,53	10.000	5,04	28.263	6.807	3.072	8.887
6	41.657	4,31	9.665,63	7.433	2,78	12.002	3.684	5.808	11.808
7	8.414	4,31	1.952,21	6.622	2,64	9.464	3.909	6.672	10.255
8	8.691	2,87	3.024,84	4.865	2,68	6.195	2.459	5.133	12.190
9	28.562	7,18	3.976,30	5.135	1,86	5.420	4.769	5.887	8.924
10	30.806	7,18	4.288,63	12.703	1,81	5.420	8.481	6.750	9.028
11	44.006	7,18	6.126,34	12.298	4,87	20.132	17.411	6.068	8.394
12	19.667	10,06	1.955,67	10.946	2,71	8.130	5.910	6.157	8.794
13	37.800	20,11	1.879,43	20.136	2,92	18.584	5.910	6.489	10.656
Total	318.442	84,76		124.193		215.045	93.592	75.875	128.823

Quadro 3 – Variáveis Socioeconômicas do Município de Vitória – ES no ano de 2007.

Fonte: BRAGA (2009)

3.3 – MONTAGEM DO BANCO DE DADOS

O primeiro passo antes de iniciar a coleta de dados, é definir qual a unidade espacial de análise que será utilizada. Os casos mais comuns de unidades espaciais de análise (geográficas) são zonas de tráfego, setor censitário, município, microrregião, macrorregião ou mesorregião.

O próximo passo é a aquisição dos dados propriamente ditos, que podem ser obtidos através de fontes primarias (coleta de dados) ou secundários (pesquisa documental). Como a regressão espacial utiliza dados agregados, ou seja, os dados na forma *cross-sectional* são agrupados segundo uma unidade maior, uma coleta de dados pode se tornar bastante cara dependendo da definição do “espaço”. Por isso é recomendável utilizar as bases de dados do IBGE ou de outra fonte de pesquisa conceituada, devido ao baixo custo de aquisição.

Neste estudo como já foi citado foi utilizado a pesquisa de origem-destino realizada pela Prefeitura do Município de Vitória capital do Espírito Santo nos anos de 1998 e 2007, estas matrizes O-D consideram cada modo e propósito de viagem separadamente para cada ano, as matrizes O-D motorizadas (coletivo e individual) com viagens de origem nas residência e um destino qualquer, ao longo do dia, no município de Vitória – ES, com propósitos de viagens a trabalho, estudo e outros.

3.3.1 - MONTAGEM DO BANCO DE DADOS PARA MODELO DE REGRESSÃO ESPACIAL

Na montagem do banco de dados utilizado para calibração do modelo de regressão espacial, foi necessária uma agregação dos dados, sendo que para isto foram desconsiderados os modos *i* e propósitos *j* de viagens.

Foi efetuado o cálculo de média aritmética simples a cada 6 grupos (3 modos multiplicado por 2 propósitos) de valores, obtendo-se assim uma matriz composta de valores médios totalizando 169 pares de origem e destino (veja no anexo A uma ilustração do banco de dados de média das matrizes O-D). Isso se faz necessário para atender ao pressuposto do modelo de regressão espacial, de que a matriz W de vizinhança espacial deve ser simétrica, bem como a soma dos elementos desta mesma matriz deve ser igual a 1 (um).

3.3.2 - MONTAGEM DO BANCO DE DADOS PARA MODELO DE REGRESSÃO GEOGRAFICAMENTE PONDERADA

Para a calibração do modelo de regressão geograficamente ponderada foi utilizado o banco de dados considerando todas as variáveis da pesquisa de origem e destino de Vitória em 1998, levando em consideração os modos *i* e os propósitos *j* de viagens, totalizando 1014 linhas. Para fins de ajuste do modelo foi considerado um vetor Ω de distancias entre as 13 macrozonas.

3.4 - VARIÁVEIS UTILIZADAS

As variáveis que fazem parte deste estudo são descritas abaixo:

- V^m_{ij}: é a demanda de transporte pelo modo m com origem na zona i e destino na zona j;
- A_m: constante numérica: porcentagem de mudança na demanda pelo meio de transporte m, quando se modifica em 1% o seu preço de viagem;
- $b_{m,n:}$ porcentagem de mudança na demanda pelo meio de transporte m, quando se modifica em 1% o preço de viagem pelo meio n;
- $c_{m,m:}$ porcentagem de mudança na demanda pelo meio de transporte m, quando se modifica 1% no tempo de viagem pelo modo m;
- $c_{m,n}$: porcentagem de mudança na demanda pelo modo m, quando se modifica 1% no tempo de viagem pelo modo n;
- d_m : porcentagem de modificação na demanda pelo meio de transporte m, em relação à modificação de 1% no produto das populações das cidades de origem e destino (i e j);
- I_m: porcentagem de modificação na demanda pelo meio de transporte m, em relação à modificação de 1% na renda da localidade de origem i. V m a m m b , n m b , m m c , n m c , m d m I

- V_{ijm}: quantidade de viagem da zona i para zona j pelo modo m;
- P_iP_j : população de i e j ;
- C_{ij*} : menor custo de viagem entre a zona i e j;
- C_{ijm} : custo via modo m;
- H_{ij*} : menor tempo de viagem entre a zona i e j;
- H_{ijm} : tempo de viagem via modo m.
- D_{ij*}: frequência de partida do modo mais frequente;
- D_{ijm}: frequência de partida do modo;
- Y_{ij}: Renda média ponderada de i e j;
- a_0, ..., a_8 : parâmetros a serem calibrados.
- t_{ijm}: tempo em minutos de viagem da macrozona i para a macrozona j via modo m;
- $V^p{}_{ijm}$: viagens com origem na macrozona i destino em j pelo modo m e propósito p;
- pop_i: população da macrozona i;
- pop_j: população da macrozona j;
- t_{ijm}: tempo de viagem de i para j via modo m;
- pop_{pi}: população da macrozona i que realiza a atividade p;
- OE_j: Oferta de emprego na macrozona j;
- OM_j: Oferta de matrículas escolares na macrozona j;
- Aut_i: número de automóveis na macrozona i.
- $dens_i$ (densidade populacional da macrozona i, habitante por km^2);
- $dens_j$ (densidade populacional da macrozona j, habitante por km^2);
- Poppi/popi (a razão entre a população que realiza a atividade p na macrozona i e a população da macrozona i).

3.5 - MATRIZ O/D

No contexto deste trabalho a matriz O/D está sendo tratada como sendo a representação do número de passageiros que viajam entre cada par de zonas ou em uma direção particular, ou seja, neste trabalho a matriz O/D é um método descritivo da demanda de passageiros que mostra a distribuição de viagens que se originam em uma zona (ponto) *i* (O_i), entre vários destinos (pontos) *j* (D_j). Esquematicamente,

	Zonas de Destino				
	1	2	.	n	
1	V_{11}	V_{12}	...	V_{1j}	O1
2	V_{21}	V_{22}	...	V_{2j}	O2
Zonas de .	.	.	.	.	. Viagens
origem .	.	.	.	.	.
					originadas
n	V_{i1}	V_{i2}	...	V_{ij}	O_n
	D_1 D_2 . D_n Viagens Destinadas				

Figura 4: Matriz de Origem-Destino

A somatória na linha i representa as viagens que se iniciam no ponto i, portanto as viagens originadas nesta zona:

$$\sum_j v_{ij= O_i}. \qquad (28)$$

A somatória na coluna j representa as viagens cujos destinos são a zona j, portanto as viagens atraídas neste ponto:

$$\sum_j v_{ij= D_i}. \qquad (29)$$

A matriz de distância foi obtida com o auxilio do *software* Google Earth com a ferramenta rota, pelo caminho mínimo de automóvel entre os centroides das macrozonas, sendo possível que a rota de origem *i* para *j* possa ser diferente da rota de *j* para *i*, por motivo de mão e contramão.

3.6 – CALIBRAÇÃO E VALIDAÇÃO DO MODELO

A calibração e a validação dos modelos foram realizadas em duas etapas:

i) Na primeira etapa fez-se para o modelo de regressão espacial, englobando os dados da média, onde se desconsiderou os 2 modos e 3 propósitos.
ii) Na segunda etapa, para o modelo de regressão geograficamente ponderada, onde adotou-se todos os dados disponíveis e considerando modos e propósitos.

Em ambas as etapas utilizou-se sempre processo de *stepwise*. As análises foram realizadas utilizando o *software* estatístico *Minitab* 13. Para tais modelos foram adotados o nível de significância de 95% ou $\alpha<0,05$.

A validação do modelo fica então condicionada à verificação dos valores das estatísticas R^2 e $R^2_{ajustado}$, bem como da estatística F. As estatísticas R^2 e $R^2_{ajustado}$ indicam a qualidade do ajuste. Quanto mais próximo os valores destas estatísticas estiverem de 1, mais ajustado está o modelo; enquanto que quanto mais próximo de 0, pior este ajuste. O valor de $R^2_{ajustado}$ é mais aconselhável do que R^2 por levar em consideração a quantidade de parâmetros do modelo, pois à medida que novas variáveis são inseridas, mais próximo de 1 tende a ser o valor de R^2 (quando existem poucas variáveis essas estatísticas tendem a ser muito próximas). Os modelos sem o intercepto, ou seja, com a reta passando pela origem apresentam maiores valores de R^2 e $R^2_{ajustado}$, no entanto, essas medidas perdem o sentido por poderem assumir valores negativos, conforme OGLIARI (2004).

CAPÍTULO 4

4 – RESULTADOS

O presente capítulo tem por objetivo verificar a aplicabilidade da metodologia proposta no Capítulo 3.

4.1 - APLICAÇÃO DOS MODELOS

4.1.1 - INTRODUÇÃO

Foram calibrados modelos de regressão Georeferenciados para previsão de demanda por viagem, com os dados do município de Vitória-ES, da pesquisa origem/destino (O-D) realizada na região da Grande Vitória em 1998, em seguida foram aplicados a esses modelos calibrados os dados da pesquisa O-D realizada na região da Grande Vitória-ES em 2007, a fim de estimar matrizes O-D para 2007, as quais serão avaliadas com as respectivas matrizes observadas pela pesquisa O-D 2007 por medidas (estatísticas) de similaridade, ou seja, por medidas que medem a semelhança entre duas matrizes, podendo dessa forma analisar o desempenho dos modelos propostos em relação aos demais modelos.

As viagens, a serem estudadas nesta dissertação, consistem somente naquelas realizadas por pessoas ao longo do dia com origem em residências e destino qualquer no município de Vitória-ES, pelo modo motorizado, que se classifica em:

i) modo coletivo: viagens feitas a ônibus, van, barco, e,

ii) modo individual: viagens feitas de automóvel, caminhão e motocicleta, não sendo consideradas as viagens feitas a pé.

E os propósitos de viagens a serem considerados são: residência para trabalho, residência para estudo e residência para outros.

4.1.2 - CALIBRAÇÃO DO MODELO DE REGRESSÃO ESPACIAL

Como citado na seção 3.3.1, além do pressuposto necessário de simetria da matriz W, outro pressuposto é de que a soma dos valores dos elementos desta matriz seja igual a um. Assim, para este trabalho, considerou-se uma matriz W com o seguinte formato para cada elemento origem *i* e destino *j*:

O elemento *ij* da matriz W será igual a "1" para uma área *j* que seja destino de todas as outras áreas. E será "0" para os demais. Dessa forma, a matriz W fica composta de 0's e 1's.

Para fins de calibração do modelo de regressão espacial, foram utilizados, após aplicação do logaritmo, os dados médios para todas as variáveis existentes, como explicado anteriormente. E para que a soma dos seus elementos seja igual a 1 é multiplicado à matriz, o produto do valor inverso da soma dos elementos de W, ou seja, (1/2197). Em resumo, V_{ij} está sendo explicado, nessa parte do modelo, por todas as viagens que chegam à macrozona *j* (para o cálculo do produto de matrizes W^*V_{ij}, relativo ao modelo expresso em (13), ver Anexo C).

4.1.3 - CALIBRAÇÃO DO MODELO DE REGRESSÃO PONDERADA

Um vetor de distância é utilizado para a calibração do modelo de regressão geograficamente ponderada, porém para isto é realizado um calculo do valor inverso deste vetor, onde o vetor Ω de distâncias foi obtida com o auxilio do *software* Google Earth com a ferramenta rota, pelo caminho mínimo de automóvel entre os centroides das macrozonas, sendo possível que a rota de origem i para j seja diferente da rota de j para i, por motivo de mão e contramão. E assim cria-se um banco de dados utilizando o *software* estatístico *minitab* 13, desenvolve-se uma análise de regressão pelo método *stepwise* para calibração do modelo, afim de que se expliquem o número de viagens. Gerados os possíveis modelos, o critério de escolha do melhor modelo será o que tenha o melhor índice de determinação.

4.1.4 - VARIÁVEIS UTILIZADAS NA CALIBRAÇÃO DOS MODELOS

Poderá ser usada qualquer variável, como:

- Variáveis socioeconômicas: renda, oferta de matriculas escolares, oferta de empregos, número de automóveis, etc;

- Variáveis do uso do solo: população, área, densidade, número de domicílios, número de empregados no comércio e número de empregados na indústria etc;

- Variáveis do sistema: tempo de viagem, custo da viagem, conforto, etc;

- Variáveis derivadas: variáveis originadas de outras variáveis e ou variáveis que representam a interação entre as variáveis. Como por exemplo: Densidade populacional, tempo de viagem x modo de transporte, etc.

A seguir são apresentadas as tabelas com os resultados dos modelos calibrados. Vale ressaltar que os dados são gerados utilizando-se uma matriz de distâncias (O/D).

A Tabela 1 apresenta os resultados da média, mediana, máximos e mínimos das variáveis selecionadas a fim de caracterizá-las.

Variável	Média	Mediana	Mínimo	Máximo
vij	1,927	1,950	2,303	7,167
WY	25,06	26,2	3,940	44,65
tij	3,1803	3,1937	2,576	3,6439
pop j	9,7765	9,7222	9,4237	10,2087
of emp j	8,764	8,7829	7,880	9,645
of mat j	8,526	8,5032	7,9684	9,1274
automove	7,7495	8,0179	5,689	9,3301
Pop pi/p	1,1086	1,0954	1,448	0,8675

Tabela 1: Estatísticas Descritivas das Variáveis Estudadas.

O Quadro 4 mostra a correlação entre as variáveis socioeconômicas de Vitória para o ano de 1998.

	Pop.	Área	Dens.	Automóvel	Renda	Oferta Emp.	Oferta Matr.	Pop. Ocup.	Pop. Est.
Pop.	1,00	0,395 0,182	0,655 0,015	0,415 0,159	-0,133 0,665	0,291 0,334	0,341 0,255	0,972** 0,000	0,993** 0,000
Área	0,395 0,182	1,00	-0,371 0,212	-0,156 0,661	-0,142 0,643	-0,165 0,59	-0,134 0,662	0,348 0,244	0,352 0,238
Dens.	0,655 0,015	-0,371 0,212	1,00	0,575 0,040	-0,006 0,985	0,527 0,064	0,535 0,060	0,598 0,031	0,703** 0,007
Automóvel	0,415 0,159	-0,156 0,661	0,575 0,040	1,00	0,659 0,014	0,907 0,020	0,851 0,015	0,420 0,153	0,487 0,092
Renda	-0,133 0,665	-0,142 0,643	-0,006 0,985	0,659 0,014	1,00	0,475 0,101	0,432 0,140	-0,089 0,773	-0,051 0,868
Oferta Emp.	0,291 0,334	-0,165 0,59	0,527 0,064	0,907 0,020	0,475 0,101	1,00	0,902 0,011	0,250 0,411	0,357 0,232
Oferta Matr.	0,341 0,255	-0,134 0,662	0,535 0,060	0,851 0,015	0,432 0,140	0,902 0,011	1,00	0,312 0,299	0,389 0,189
Pop. Ocup	0,972** 0,000	0,348 0,244	0,598 0,031	0,420 0,153	-0,089 0,773	0,250 0,411	0,312 0,299	1,00	0,955** 0,000
Pop. Est.	0,993** 0,000	0,352 0,238	0,703** 0,007	0,487 0,092	-0,051 0,868	0,357 0,232	0,389 0,189	0,955** 0,000	1,00

Quadro 4: Matriz de correlação das Variáveis Socioeconômicas do município de Vitória-ES em 1998.

4.1.5 - RESULTADOS DOS MODELOS CALIBRADOS

Resumindo as equações dos modelos calibrados são as seguintes:

Regressão Tradicional (Modelo Direto):

V_{ij} = exp (0,7858V_{ijm}/v_{ij} + 1,1112 Automóvel)

Regressão Espacial:

V_{ij} = exp (- 22,30 + 1,82 P_i*O_{ej} + 0,790 V_{ijm}/v_{ij} + 1,40 Automóvel -1,61 P_i*r_i + 1,75 Pop p_i/p)

Regressão Dummy:

V_{ij} = exp (-1,57 t_{ij} -1,73 R*p_2 + 0,00014 OM_J*P_1)

Abaixo na Tabela 2 tem-se uma comparação entre os modelos de regressão calibrados para dados da média.

Variável	Regressão Tradicional		Regressão Espacial		Regressão Dummy	
	Parâmetro	P - Valor	Parâmetro	P - Valor	Parâmetro	P - Valor
Intercepto			-22,30	0,00*		
Pi*Oej			1,82	0,00*		
Vijm/vij	0,7858	0,00*	0,790	0,00*		
Automovel	1,1112	0,00*	1,40	0,00*		
Pi*ri			-1,61	0,00*		
Pop pi/p			1,75	0,00*		
tij					-1,57	0,00*
R*p2					-1,73	0,008*
OMJ*P1					0,00014	0,008*
R^2	66,70		61,74		77,97	
$R^2_{Ajustado}$	58,80		60,57		76,73	
F	42,32		8,56		35,60	

(p-valor), (*) = muito significativo (p<0,01), (**) = pouco significativo (p<0,05), NS = não significativo

Tabela 2: Resultados do Modelo de Regressão - para dados de média (N = 169).

Resumindo as equações dos seguintes modelos abaixo:

<u>Regressão Tradicional (Modelo Direto):</u>

V_{ij} = exp (-17,844 + 0,6943 Pop_i + 0,82966Automóvel - 0,6072 T_{ij}*M -0,6618 Pop_j + 2,4147 $Poppi_i/pop_i$ + 1,2088 Oe_j + 0,8255 OM_j)

<u>Regressão Ponderada:</u>

V_{ij} = exp (0,9164 Pop_i + 1,220Automóvel + 0,9230 $Renda_j$ + 0,663 V_{ijm}/v_{ijmp})

<u>Regressão Dummy:</u>

V_{ij} = exp (- 22,310 + 1,3825 Pop_i - 2,4285 t_{ij} + 1,9275 T_{ij}*M + 1,2823 O_{ej} + 0,3495 OM_j + 0,46443 O_{ej}*P_2 + 0,0001 OM_j*P_1 + 2,2137Renda $_i$ -4,5133 R*M + 2,4740 R*M*P_1 + 0,6431 P_1), onde:

- Intercepto = Valor constante;
- Pop_i = (população da macrozona i);
- T_{ij} = Tempo de viagem em minutos da macrozona i para macrozona j;
- Aut_i = Número de automóveis na macrozona i;
- T_{ij}*M = Tempo de viagem da zona i para zona j x M: variável dummy = 2 se for modo Coletivo,1 se for modo Individual;
- $Renda_j$ = Renda média em salário mínimo da macrozona j;
- V_{ijm}/v_{ijmp} = A razão entre Número de viagens por dia de i para j e Número de viagens por dia de i para j pelo modo m e propósito p;
- Pop_j = População da macrozona j;
- Pop_{pi}/pop_i = A razão entre a população que realiza a atividade p na macrozona i e a população da macrozona i;
- Oe_j = Oferta de emprego na macrozona j;
- OM_j = Oferta de matrículas escolares na macrozona j;
- Oe_j*P2 = Oferta de emprego na macrozona j x P2: variável dummy = 1 se o propósito da viagem for a trabalho,0 caso contrário;
- OM_j*P1 = Oferta de matrículas escolares na macrozona i x P1: variável dummy = 1 se o propósito da viagem for a estudo, 0 caso contrário;
- $Renda_i$ = Renda média em salário mínimo da macrozona i;
- R*M = Renda média em salário mínimo x M: variável dummy = 2 se for modo Coletivo,1 se for modo Individual;
- R*M*P1 = Renda média em salário mínimo x M: variável dummy = 2 se for modo Coletivo,1 se for modo Individual x Variável dummy = 2 se o propósito da viagem for a estudo,1 se for outro qualquer;
- P1 = Variável dummy = 2 se o propósito da viagem for a estudo,1 se for outro qualquer

Variável	Regressão Tradicional		Regressão Ponderada		Regressão Dummy	
	Parâmetro	P - Valor	Parâmetro	P - Valor	Parâmetro	P - Valor
Intercepto	-17,844	0,00*			- 22,310	0,00*
Pop i	0,6943	0,00*	0,9164	0,00*	1,3825	0,00*
tij					- 2,4285	0,00*
Automovel	0,82966	0,00*	1,220	0,00*		0,00*
Tij*M	-0,6072	0,00*			1,9275	0,00*
Renda j			0,9230	0,00*		
Vijm/vijmp			0,663	0,00*		
Pop j	-0,6618	0,001*				
Poppi/popi	2,4147	0,00*				
Oej	1,2088	0,00*			1,2823	0,00*
OMj	0,8255	0,001*			0,3495	0,101
Oej*P2					0,46443	0,00*
OMj*P1					0,0001	0,141
Renda i					2,2137	0,00*
R*M					-4,5133	0,00*
R*M*P1					2,4740	0,00*
P1					0,6431	0,261
R^2	35,00		56,40		50,10	
R^2_{Ajustado}	34,60		55,00		49,50	
F	77,46		78,99		80,96	

Tabela 3: Parâmetros dos Modelos de Regressão - para Todos os dados (N = 1.014).

4.2 - AVALIAÇÃO DO DESEMPENHO DOS MODELOS

4.2.1 - INTRODUÇÃO

Existem diversos métodos estatísticos para avaliar o desempenho de modelos, os testes estatísticos propostos são medidas de similaridade e medem a proximidade da matriz observada com a matriz estimada. Os testes estatísticos utilizados neste trabalho são: Phi – normalizado e Índice de Dissimilaridade.

Considere a seguinte notação:

- $\mathbf{\widehat{V}_{ij}}$ = Matriz O-D Estimada
- $\mathbf{V_{ij}}$ = Matriz O-D Observada
- $\mathbf{V}$ = Número total de viagens

4.2.2 - ESTATÍSTICA PHI-NORMALIZADA

Esta Estatística é baseada na teoria da informação e segundo SMITH e HUTCHINSON (1981), a estatística phi-normalizada mostra-se adequada para avaliação de modelos de distribuição de viagens. GONÇALVES e ULYSSÉA NETO (1993) definem estatística phi-normalizada por:

$$\phi = \sum_{ij} \left(\frac{\mathrm{V_{ij}}}{\mathrm{V}}\right) . \left|\ln \left(\frac{\mathrm{V_{ij}}}{\hat{\mathrm{V}}_{\mathrm{ij}}}\right)\right| \qquad (30)$$

O valor de ϕ igual a zero indica que as matrizes observadas e estimadas são idênticas. Assim, quanto menor o valor de ϕ tanto melhor será a matriz O/D de viagens estimadas.

4.2.3 - ÍNDICE DE DISSIMILARIDADE

Segundo GONÇALVES e ULYSSÉA NETO (1993), o índice de dissimilaridade mede a porcentagem de viagens que podem ser realocadas entre pares (*i, j*), para que a matriz observada coincida com a matriz estimada. O índice de dissimilaridade é dado por:

$$ID = \left(\frac{50}{V}\right) \sum_{ij} \left| V_{ij} - \hat{V}_{ij} \right| \qquad (31)$$

O valor de ID varia entre 0 e 100. Assim, quanto mais próximo de zero for o valor de ID, tanto melhor será a estimativa da matriz.

4.2.4 – MEDIDAS DE AVALIAÇÃO DOS MODELOS E RESULTADOS

Nas tabelas 4 e 5 são apresentados os valores das estatísticas de dissimilaridade PHI e ID das comparações das matrizes O-D estimadas pelos modelos de regressão tradicional, espacial, dummy e ponderada, com as matrizes O-D obervadas, para cada combinação de modo e propósito de viagens.

Estatística	Modelo Regressão		
	Tradicional	Espacial	Dummy
PHI	3,56	5,23	**1,49**
ID	41,45	**40,09**	55,90

Tabela 4 . Estatísticas de Ajuste dos modelos de Regressão para dados da média (n = 169).

A análise dos resultados consiste na comparação das matrizes O-D estimadas pelos respectivos métodos de modelagem com a matriz O-D obtida na pesquisa de 2007, por medidas de similaridade que medem a proximidade da matriz estimada com a matriz observada.

Observando a Tabela 4, que contém as medidas de ajuste dos modelos calibrados e, seguindo o critério de que quanto menor o valor de Phi-normalizado melhor o ajuste, concluímos que o modelo Dummy para dados da média foi o que melhor estimou a demanda. E para todos os dados, a regressão tradicional foi o modelo de melhor ajuste. Por outro lado a estatística ID para o modelo de regressão espacial apresentou um melhor desempenho em relação aos demais modelos.

Estatística	Modelo Regressão				
	Modo	Propósito	Tradicional	Ponderada	Dummy
PHI	Coletivo	Trabalho	0,30	**0,20**	0,25
	Coletivo	Estudo	2,99	**0,74**	2,25
	Coletivo	Outros	1,92	**1,02**	1,57
	Individual	Trabalho	1,04	0,85	**0,75**
	Individual	Estudo	2,76	**0,98**	2,45
	Individual	Outros	2,11	**0,55**	1,26
ID	Coletivo	Trabalho	46,31	**44,45**	47,89
	Coletivo	Estudo	50,29	**38,07**	53,46
	Coletivo	Outros	66,52	**46,60**	60,21
	Individual	Trabalho	65,01	**31,05**	34,02
	Individual	Estudo	55,67	**42,12**	48,03
	Individual	Outros	56,71	55,09	**49,19**

Tabela 5. Estatísticas de Ajuste dos modelos de Regressão para Todos os dados (N= 1.014).

Na Tabela 5 pode-se observar que a estatística Phi - normalizado para o modelo de regressão ponderada apresentou resultados um pouco melhores do que os outros dois métodos, sendo que estes apresentaram desempenhos similares.

No caso da estatística ID o modelo de regressão ponderada também apresentou um melhor desempenho em relação aos demais modelos. Concluímos que o modelo Dummy para dados da média foi o que melhor estimou a demanda e para todos os temos a regressão ponderada como o modelo de melhor ajuste.

CAPÍTULO 5

5 – CONCLUSÕES E CONSIDERAÇÕES FINAIS

Este capítulo apresenta as conclusões obtidas pela presente pesquisa, bem como algumas recomendações para trabalhos futuros.

5.1 - CONCLUSÕES DE CARÁTER GERAL

O objetivo deste trabalho era utilizar os modelos de regressão espacial e o modelo de regressão geograficamente ponderada para as previsões de demanda por viagem utilizando técnicas de estatística espacial, juntamente com ferramentas de planejamento de transporte. Os dados utilizados para calibração são os da pesquisa domiciliar de origem e destino (O-D) realizada em 1998 no município Vitória-ES. Após a calibração os modelos para os dados de 1998, os mesmos foram testados a partir da aplicação nos dados da pesquisa domiciliar de origem e destino realizada em 2007, da mesma cidade.

Neste trabalho, apresentou-se a calibração de quatro modelos de regressão de modelagem de demanda de viagem: Modelo de Regressão Tradicional, Modelo de Regressão Dummy, Modelo de Regressão Espacial, Modelo de Regressão Geograficamente Ponderada.

Quando se avalia os resultados do ajuste para os modelos aplicados aos dados médios, conclui-se que o modelo de regressão espacial apresenta certa vantagem em relação aos demais, porém, ligeiramente próximo dos resultados obtidos para o modelo tradicional. Pela simplicidade e aplicabilidade talvez fosse adequado considerar que é melhor manter o uso do modelo tradicional.

No entanto, ao se analisar os resultados dos ajustes obtidos quando se compara o modelo de regressão geograficamente ponderada pelas distâncias, entende-se que este apresenta um grau de ajuste melhor aos dados, sendo, portanto, mais adequado para as realizações de previsões do que os modelos tradicionais e de uso de variáveis dummies.

Finalmente, pode-se concluir que a hipótese principal, ou parte dela, considerada neste trabalho foi confirmada, a de que um modelo de regressão espacial ou de regressão geograficamente ponderada por distâncias pode ser mais explicativo do que aqueles modelos de regressão convencionais, pois, a calibração de modelos de demanda de viagem pelo modelo de regressão ponderada apresentou valores das estatísticas de ajustes PHI e ID menores que os outros modelos para praticamente todos os modos e propósitos, indicando assim uma melhor aproximação com as matrizes observadas de viagens do que os demais modelos.

5.2 - RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Espera-se que a elaboração desta pesquisa abra caminhos para que novos pesquisadores, de posse de outros recursos tecnológicos disponíveis ou modelos, encontrem melhores soluções para problemas semelhantes ao pesquisado.

Como recomendações para trabalhos futuros sugere-se o uso de modelagem via dados de fluxo entre as áreas geradoras, e que tais áreas sejam desagregadas em outras de menor tamanho e em maior quantidade, permitindo assim um melhor detalhamento das previsões dos modelos.

CAPÍTULO 6

6 – REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, D. R.; SWEENEY, D. J.; WILLIAMS, T. **Statistics for business and economics**. 3 ed, South-Western, 2002.

ANSELIN, L. and GRIFFITH, D. A. **Do spatial effects really matter in regression analysis?** Papers of the Regional Science Association, 65, 1988, p. 11-34.

ANSELIN, L. **Under the hood: issues in the specification and interpretation of spatial regression models** (forthcoming: Agricultural Economics), obtido em: <http://agec221.agecon.uiuc.edu/users/anselin/papers/ hood.pdf> Acesso em 03/11/2009. 2002.

BEN AKIVA, M. E.; e LERMAN, S. R. **Discrete Choice Analysis; Theory and Application to Travel Demand**. The MIT Press, Cambridge, Massachussets. 1985.

BRAGA, M. F., **Modelo Direto de Previsão de Demanda de Viagens Utilizando Variável Dummy**. Dissertação de Mestrado. PPGEC/UFES. 2009.

BRAGA, M. F.; NETO, G. C. de M.; BERTOLDE, A. I. **Modelagem da demanda por transportes utilizando variáveis dummies.** XXIII ANPET - Associação Nacional de Pesquisa e Ensino em Transportes, Vitória, ES, (2009)

BRUNSDON, C., FOTHERINGHAM, A. S., & CHARLTON, M. **Geographically weighted regression: A method for exploring spatial nonstationarity. Geographical Analysis**, 28 (4), 281-289. - 1996

BRUTON, M.J. **Introdução ao Planejamento dos Transportes**. Editora Interciência, Rio de Janeiro. 1975

CÂMARA, G.; Carvalho, M. S.; Cruz, O. G.; Correa, V. . **Análise Espacial de áreas, Em: Análise Espacial de Dados Geográficos**, Eds. Fuks, S.D. 2002ª.

CÂMARA, G.; CARVALHO, M.S.; CRUZ, O. G.; CORREA, **V. Análise espacial de áreas, In: Fuks, S.D.; Carvalho, M.S.; Câmara, G; Monteiro, A.M.V. (Eds) Análise espacial de dados geográficos** – Divisão de Processamento de Imagens, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brasil. 2002

CHATFIELD, C. **Model uncertainty, data mining and statistical inference (with Discussion**). J. R. Stat. Soc. A, Cambridge, v.158, p.419-66, 1995.

CRESSIE, NOEL A.C. **Statistics for spatial data**. Revised edition - New York: J. Wiley, c1993. 900p. - (Wiley series in probability and mathematical statistics).

DEMUTH H., Beale M., and Hagan M. Neural Network Toolbox 6. **The MathWorks, Natic,** MA, USA, 2008.

DOWNING, DOUGLAS; CLARK, Jeffrey. **Estatística Aplicada**. São Paulo: Saraiva, 455p. 2000.

FERRAZ, Antonio Clovis pinto; TORRES, Isaac Guilhermo Espinosa. **Transporte Público Urbano.** 2. ed. São Carlos: Rima, 2004.

FERRONATTO, L. G. **Potencial de medidas de gerenciamento da demanda no transporte público urbano por ônibus**. 119 f. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Sul, Escola de Engenharia. 2002.

FRANZBLAU, A. (1958). ***A primer of statistics for non-statisticians.*** New York: Harcourt, Brace & World.

FOTHERINGHAM, A.S.; Charlton, M.; Brusdon, C. (1997) Recent developments in spatial analysis, Chapter measuring spatial variations in relationships with geographically weighted regression, pages 60- 82. Springer, New York, EUA.

GONÇALVES, M. B.; ULYSSÉA NETO, I. (1993). **Análise comparativa do desempenho de alguns modelos de distribuição de viagens usados para estimar fluxos intermunicipais de passageiros.** In: Anais do VII Encontro Nacional da ANPET. São Paulo - SP, Vol. I, p. 337 – 338.

GUJARATI, DAMODAR N., **Basic Econometrics**, 3rd Edition. New York: McGraw-Hill Book Company. 1995

HAIR, J. F. *et al.* **Análise muitivariada de dados**. 5ª ed. Porto Alegre: Bookman, 2005.

HENRIQUE, C. S. **Diagnostico Espacial da Mobilidade e da Acessibilidade dos Usuários do Sistema Integrado de Transportes de Fortaleza.** Dissertação de Mestrado, Programa de Mestrado de Transportes, Universidade Federal do Ceará, Fortaleza, CE, 165 fl. 2004.

JUNIOR D.C.M. **Redução de dimensionalidade utilizando entropia condicional média aplicada a problemas de bioinformática e de processamento de imagens**. Master's Thesis, USP, 2004.

KREMPI, A.P. **Explorando recursos de estatística espacial para análise da acessibilidade da cidade de Bauru. 94 f. Dissertação** (Mestrado em Engenharia

Civil) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos. 2004.

LOUREIRO, C.F.G.; SILVA H. N.; CARVALHO, L. E. X. **Metodologia de análise de regressão geograficamente ponderada aplicada ao fenômeno das viagens intermunicipais. Anais do XX Congresso de Pesquisa e Ensino em Transportes**. ANPET - Associação Nacional de Pesquisa e Ensino em Transportes, Brasília, DF, 479-491. (2006)

MANHEIN, M. L. **Fundamentals of Transportation System Analysis**: Volume 1 - Basic concepts. The MIT Press, Cambridge, Mass. USA. 1979.

MEYER, D.M.; MILLER E.J**. Urban Transportation Planning.** Editora McGraw-Hill, New York, USA. 2001.

MELLO, J. C. **Planejamento dos transportes**. São Paulo, Mcgraw Hill do Brasil. 1975.

NETER, J.; KUTNER, M H.; NACHTSHEIM, C. J.; LI, W. **Applied Linear Statistical Models.** 5. Ed. New York: Mc Graw-Hill/ Irwin, 2005, 1396p.

NOVAES, A. G., **Modelos em Planejamento Urbano, Regional e de Transportes**. Editora Edgard Blücher Ltda. São Paulo, SP. 1981.

NOVAES A. G., **Sistemas de Transporte – Volume 1: Análise da Demanda**. Edgard Blücher, São Paulo. 1986.

OGLIARI, P. **Construção de Modelos: Seleção de Variáveis Regressoras. Nota de aula, UFSC.** 2004. Disponível em < WWW.inf.ufsc.br/~ogliariarquivos/construção de modelos de regressao.ppt> acesso em 05/10/2009.

PAPACOSTAS, C. S.; PREVEDOUROS P.D. **Fundamentals of Transportation Engineering** (3ª ed) Prentice- Hall, New Jersey. 2000.

PERINI, F. R. C. . **Diagnóstico Espacial da Acessibilidade dos Usuários do Sistema Municipal de Transporte Coletivo da Cidade de Vitória.** Dissertação de Mestrado. PPGEC/UFES. 2008.

QUEIROZ, M. P. **Análise espacial dos acidentes de trânsito do município de Fortaleza.** Dissertação de Mestrado, Programa de Mestrado em Engenharia de Transportes, Universidade Federal do Ceará, Fortaleza, CE, 124 fl. 2003.

SANTOS, L. **Análise dos acidentes de trânsito no município de São Carlos utilizando SIG e ferramentas de análise espacial.** Dissertação (Mestrado). Programa de Pós-Graduação em Engenharia Urbana. UFSCar. São Carlos. 2006.

SILVA, A. R. **Avaliação de Modelos de Regressão Espacial para Análise de Cenários do Transporte Rodoviário de Carga.** Dissertação de Mestrado, Publicação T.DM – 013ª/2006, Departamento de Engenharia Civil e Ambiental, Faculdade de Tecnologia, Universidade de Brasília, DF, 125p. 2006.

SMITH, D. P., HUTCHINSON, B. G. **Goodness of fit statistics for trip distribution models**. Transportation Research, 15A, p.295-303. 1981.

SUBRAMANIAN, A.; COUTINHO, A. S.; da SILVA, L. B. Aplicação de método e técnica multivariados para previsão de variáveis termoambientais e perceptivas. Produção, v. 17, n. 1, p. 052-070, 2007.

TEIXEIRA, G. L. **Uso de dados Censitários para Identificação de zonas Homogêneas para planejamento de Transportes Utilizando Estatística Espacial** - Dissertação de Mestrado, Publicação T. DM.-010A/03 – Departamento de Engenharia Civil e Ambiental – Faculdade de Tecnologia – Universidade de Brasília – DF- 155p. 2003.

TOBLER, W. R., **Cellular Geography Philosophy in Geography.** Edited By S. Gale and G. Olsson. Eds, Amsterdam. 1979.

7. ANEXOS

ANEXO A: Arrumação dos dados para Calibração

Modo(M)	Propósito (P)	O-D	Variáveis												
			Var. Dep.	Variáveis Explicativas				Variáveis Dummies			Variáveis Interativas				
			$\ln V_{ijmp}$	$\ln X_1$	$\ln X_2$	...	$\ln X_k$	$\ln M_1$	$\ln P_2$	$\ln P_2$	$\ln M \ \ln X_1$	$\ln M \ \ln X_2$	...	$\ln M \ \ln P_2 \ \ln X_k$	
		1-1.													
		1-2.													
		:													
		1-n													
	Trabalho	2-1.													
		2-2.													
		:													
		2-n													
		n-n													
		1-1.													
		1-2.													
		:													
		1-n													
Coletivo	Estudo	2-1.													
		2-2.													
		:													
		2-n													
		n-n													
		1-1.													
		1-2.													
		:													
		1-n													
	Outros	2-1.													
		2-2.													
		:													
		2-n													
		n-n													
		1-1.													
		1-2.													
		:													
		1-n													
	Trabalho	2-1.													
		2-2.													
		:													
		2-n													
		n-n													
		1-1.													
		1-2.													
		:													
		1-n													
Individual	Estudo	2-1.													
		2-2.													
		:													
		2-n													
		n-n													
		1-1.													
		1-2.													
		:													
		1-n													
	Outros	2-1.													
		2-2.													
		:													
		2-n													
		n-n													

ANEXO B: Arrumação dos dados da média para Calibração

O-D	Variáveis								
	Var. Dep.	Variáveis Explicativas				Variáveis Interativas			
	ln V_{ijmp}	ln X_1	ln X_2	...	ln X_k	ln M ln X_1	ln M ln X_2	...	ln M ln P_2 ln X_k
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)
1-1.									
1-2.									
:									
2-n									
n-6									
Média	Média(ln Vijmp)	Média(ln X_1)	Média(ln X_2)	Média(...)	Média(ln X_k)	Média(ln M ln X_1)	Média(ln M ln X2)	Média(...)	Média(ln M ln P2 ln X_k)

ANEXO C - RELATÓRIO DO SOFTWARE R

Cálculo do produto das matrizes W e Y que resultaram na matriz wy, utilizando o software estatístico R:

```
> w=read.table("F:w.txt")

> y=read.table("F:y.txt")

> y1=as.vector(y)

> w1=as.matrix(w)

> viz=w1%*%y1

> viz=w1%*%y

> w=read.table("F:w.txt")

> y1=as.vector(y,[,1])

> w=read.table(F:w.txt")

+ w=read.table("F:w.txt")

"w=read.table(F:w.txt")

w=read.table(""

> w1=as.matrix(w)

> viz=w1%*%y[,1]

> viz
```

ANEXO D – Relatório do *software* Minitab 13.2 relativo á calibração do modelo tradicional.

$\ln V_{ij}$ = - 17,8 + 0,694 ln (pop i) - 0,662 ln (pop j) + 1,21 ln (of emp j) + 0,825 ln (of mat j) + 0,830 ln (automoveis i) + 2,41 ln (Pop pi/popi)

Predictor	Coef	SE Coef	T	P
Constant	-17,844	2,482	-7,19	0,000
tij	-0,6072	0,3095	-1,96	0,050
pop i	0,6943	0,1680	04,13	0,000
pop j	-0,6618	0,1921	-3,44	0,001
of emp j	1,2088	0,1831	6,60	0,000
of mat j	0,8255	0,2591	3,19	0,001
automove	0,82966	0,09532	8,70	0,000
Pop pi/p	2,4147	0,6620	3,65	0,000

S = 3,008 R-Sq = 35,0% R-Sq(adj) = 34,6%

Análise de Variância

Source	DF	SS	MS	F	P
Regression	7	4907,21	701,03	77,46	0,000
Residual Error	1006	9104,15	9,05		
Total	1013	14011,36			

ANEXO E - Relatório do *software* Minitab 13.2 relativo á calibração do modelo regressão espacial

vij = - 13,0 + 0,0724 WY + 1,10 tij + 0,565 pop i - 0,25 pop j - 0,13 of emp j + 0,34 of mat j + 0,935 automoveis + 2,27 Pop pi/popi

Predictor	Coef	SE Coef	T	P
Constant	-12,99	14,41	-0,9	0,369
WY	0,07236	0,03626	2	0,048
tij	1,1045	0,7369	1,5	0,136
pop i	0,5646	0,2721	2,07	0,04
pop j	-0,245	1,599	-0,15	0,878
of emp j	-0,133	1,747	-0,08	0,94
of mat j	0,341	2,074	0,16	0,87
automove	0,9353	0,1564	5,98	0
Pop pi/p	2,274	1,044	2,18	0,031

S = 1,855 R-Sq = 44,2% R-Sq(adj) = 41,4%

Análise de Variância

Source	DF	SS	MS	F	P
Regression	8	436,515	54,564	15,86	0
Residual Error	160	550,452	3,44		
Total	168	986,967			