**VERIFICAÇÃO DA APLICABILIDADE DA TÉCNICA DE
MINERAÇÃO DE DADOS NA PREVISÃO DA DEMANDA POR
TRANSPORTE DE PASSAGEIROS URBANOS USANDO DADOS
DA REGIÃO METROPOLITANA DE SÃO PAULO**

Mateus Araújo e Silva

Dissertação apresentada à Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Engenharia Civil: Planejamento e Operação de Sistemas de Transportes.

Orientador: Prof. Dr. Eiji Kawamoto

São Carlos
2006

Ficha catalográfica preparada pela Seção de Tratamento
da Informação do Serviço de Biblioteca – EESC/USP

S586v Silva, Mateus Araújo e
Verificação da aplicabilidade da técnica de
mineração de dados na previsão da demanda por
transporte de passageiros urbanos usando dados da
Região Metropolitana de São Paulo / Mateus Araújo e
Silva. -- São Carlos, 2006.

Dissertação (Mestrado) -- Escola de Engenharia de
São Carlos-Universidade de São Paulo, 2006.
Área: Planejamento e Operação de Sistema de
Transportes.
Orientador: Prof. Dr. Eiji Kawamoto.

1. Planejamento de transportes. 2. Demanda por
transportes. 3. Demanda por transportes - modelos
baseados em atividades. 4. Previsão da demanda por
transportes. 5. Elaboração de cenários futuros.
6. Planejamento urbano. I. Título.

*À memória dos meus avós,
Severiano, Maria Amália,
Firmino e Joana*

*Aos meus pais, Severiano e Margarida
pelo amor, carinho, amizade, conselho e apoio
durante toda a minha vida.
As minhas irmãs Andréa e Mariana
e ao meu irmão Rodrigo
pelo amor, carinho, amizade, apoio e companheirismo.
O amor de vocês, minha família, é o maior bem que tenho na vida.*

AGRADECIMENTOS

Inicialmente agradeço a Universidade de São Paulo, pela minha aceitação no programa de Pós-graduação, e concessão de espaço físico e material para o desenvolvimento desta pesquisa.

Agradeço ao Governo Federal, que através do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq me laureou com uma bolsa de auxílio para desenvolvimento do Mestrado.

Meus sinceros e prestimosos agradecimentos ao Prof. Eiji Kawamoto da Universidade de São Paulo, que me aceitou como seu orientado, conduziu sua orientação com espírito educador e formador de recursos.

Meus agradecimentos aos Professores Antônio Ferraz, Ernesto Schaal e Manoel Sória, da Universidade de São Paulo pelo apoio, conselhos e ensinamentos.

Agradeço também ao Prof. Antônio Nelson Rodrigues da Silva da Universidade de São Paulo e a Profa. Suely Sanches da Universidade Federal de São Carlos pelas sugestões ao trabalho no exame de qualificação. Meus agradecimentos também se estendem ao Prof. Orlando Strambi da Universidade de São Paulo pelas suas sugestões tecidas ao trabalho.

Aos professores da Universidade Federal da Bahia, Emerson Ferreira, que me possibilitou o primeiro contato com a pesquisa científica, por seus conselhos e amizade ao longo desses anos; Wellington Figueiredo pelos conselhos, apoio e primeiros ensinamentos na área de Engenharia de Transportes; Prof. Adailton Gomes, Profa. Ana Helena Almeida, Profa. Anita Martinelli, Profa. Ilce Marília, Prof. João Augusto, Prof. Juan Delgado, Profa. Sônia Medeiros e ao Prof. Sobral, onde sempre encontrei conhecimento, e palavras de incentivo.

Meus agradecimentos ao Dr. Masoud Motavalli, Christoph Czaderski e Andréa Bergamini da Eidgenössische Materialprüfung & Forschungsanstalt - EMPA pelos conselhos e orientações na época em que fui estagiário desta empresa.

Ao colega e amigo Moacyr Cardoso da Construtora Odebrecht pelos ensinamentos na época em que fui estagiário desta empresa. "Grampão" agradece!

Gostaria de deixar registrados meus agradecimentos às prestativas e competentes secretárias do Departamento de Transportes, Heloísa, Beth e D. Magaly que sempre

estiveram voluntárias e atenciosas as minhas dúvidas e solicitações. À Magaly meu carinho especial.

Meus agradecimentos aos amigos e funcionários do Departamento de Transportes: Gigante, Paulo, Toco, Vicente, Alexandre, João e Walter pela amizade e por estarem sempre prestativos em me auxiliar. “Barriga” agradece!

Aos amigos, Adalberto, Dinato, Eduardo Caramuru, Eimair, Francis, Igor Bária, Jesner, Marcelo, Rômulo, Wesley e as amigas, Adriana, Ana Larocca, Ana Furlan, Andréa Júlia, Camilla Carrara, Cynthia, Danielle, Karênina e Nívea pela amizade e apoio ao longo deste trabalho.

À Cida, ao Mário Garrido e ao Dioguinho pelo carinho e amizade.

Ao amigo André Cunha pelo auxílio nos desenvolvimentos de macros para processamento de parte dos dados, pelo apoio ao longo deste trabalho e pela amizade.

Ao amigo Pablo pela recepção na sala dos orientados do Prof. Eiji, as sugestões e apoio constante no Mestrado além da amizade.

Aos amigos Bruno e Marcus pelo companheirismo, pelo clima de harmonia, humor e trabalho na sala dos orientados do Prof. Eiji, apoio constante no Mestrado além da grande amizade formada.

A minha adorável amiga Cira, minha “Tutora”, pela recepção na sala dos orientados do Prof. Eiji, que além de me ajudar no desenvolvimento do Mestrado, sempre esteve ao meu lado, com seu carinho, apoio e inestimável amizade. Seu “Pupilo” agradece!

Aos amigos Alípio Jorge, Daniel Vêras, Rodrigo Albuquerque, Tiago Freire, à amiga Priscila Lapa, pelo apoio e amizade e a Carol pelo grande apoio nos momentos finais.

Aos amigos Martin Arraigada e Thaís Bocca, que estando a milhares de quilômetros sempre estiveram presentes com suas amizades, conselhos e apoios.

Aos amigos inseparáveis desde os oito anos de idade: José Leiro, Jaime Barreiros, Antônio Quadros e Marcos Peixoto, pelo apoio e amizade.

Aos demais amigos que sempre me acompanham meus sinceros agradecimentos!

“A palavra, esse dom celeste que Deus deu ao homem e recusou ao animal é a mais sublime expressão de natureza: ela revela o poder do criador e reflete toda a grandeza de sua obra divina.”

(José de Alencar)

RESUMO

SILVA, M. A. Verificação da aplicabilidade da técnica de mineração de dados na previsão da demanda por transporte de passageiros urbanos usando dados da Região Metropolitana de São Paulo. 2006. 176 f. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2006.

O objetivo desta pesquisa é validar a hipótese de que o minerador de dados pode ser utilizado como uma ferramenta para previsão dos padrões de viagens, inclusive sob mudança comportamental dos indivíduos. Para o desenvolvimento deste trabalho foi adotada uma postura científica indutiva, utilizando como dados as informações contidas nas duas Pesquisas Origem-Destino realizadas em 1987 e 1997 pelo METRÔ-SP na Região Metropolitana de São Paulo (RMSP). Os dados da primeira pesquisa e as mudanças comportamentais dos indivíduos observadas no período de 1987 a 1997 forneceram as condições para elaboração do cenário futuro a ser usado para projeção da demanda por transporte para 1997. Aplicando a Árvore de Decisão e Classificação, uma das técnicas de mineração de dados disponível no software S-PLUS 6.1, foram obtidas as distribuições das probabilidades das distribuições dos padrões de viagens encadeadas relacionadas a cada grupo homogêneo de viajantes urbanos. Estas probabilidades foram aplicadas aos indivíduos da amostra de 1997, estimando o número de viajantes urbanos por padrão de viagem encadeada em cada uma das 361 zonas de tráfego da RMSP. Os valores estimados pelo modelo foram comparados com os dados observados em 1997 pelo teste estatístico não-paramétrico de Kolmogorov-Smirnov (K-S). Ao final, concluiu-se que a hipótese proposta foi confirmada em 92,2% das zonas de tráfego analisadas.

Palavras-chave: Planejamento de transportes. Demanda por transportes. Modelos baseados em atividades. Previsão da demanda por transportes. Elaboração de cenários futuros. Planejamento urbano.

ABSTRACT

SILVA, M. A. An evaluation process of the data mining technique for forecasting urban passengers' transportation demand using São Paulo Metropolitan Area data. 2006. 176 f. Dissertation (Master's Thesis) - Engineering School of São Carlos, University of São Paulo, São Carlos, 2006.

The main aim of this work is to submit to a validation test the hypothesis that data mining technique can be used as a tool for forecasting the trip-chaining patterns including individual behavioral changes. For the project development, an inductive scientific method point of view was adopted, using as a data resource the information contained in two origin-destination surveys carried out by METRÔ-SP in the São Paulo Metropolitan Area (SPMA) in 1987 and 1997. The first survey data and the individual behavioral changes observed in the period from 1987 to 1997 provided the information to build a future scenario to predict the 1997's travel demand. Applying the Classification and Regression Tree which is a data mining technique available in S-PLUS 6.1 software package, it were obtained the probabilities distribution of the trip-chaining patterns and set the contextual socioeconomics characteristics of the urban travelers. These probabilities were applied to the 1997 individuals belonging to the sample collected in 1997 to estimate the number of urban travelers by trip-chaining patterns for each of the 361 traffic zones of SMPA. The estimated results from the model were compared with the real values of 1997 data set using the Kolmogorov-Smirnov (K-S) non-parametric statistic test. The main conclusion is that the hypothesis was confirmed in 92.2% of the SPMA traffic zones.

Keywords: Transportation Planning. Transportation Demand. Activity-based models. Forecasting transportation demand. Building future scenarios. Urban planning.

SUMÁRIO

Capítulo 1 - Introdução	14
1.1 Definição do Problema	14
1.2 Justificativa para se Estudar o Problema	19
1.3 Idéia Diretriz	21
1.4 Discurso do Método	21
1.4.1 Postura científica adotada	21
1.4.2 O experimento	23
1.5 Estrutura do Trabalho	25
Capítulo 2 - Viagens Baseadas em Atividades	26
2.1 Análise de Demanda por Transporte: Novo Enfoque	26
2.2 Fundamentos da Teoria de Viagens Baseada em Atividades	28
2.2.1 A viagem é uma demanda derivada	28
2.2.2 Disponibilidade para desempenho das atividades	32
2.2.3 Ênfase no domicílio como unidade de tomada de decisão	33
2.2.4 Padrão de atividade	34
2.2.5 Processo de elaboração da decisão e programação de atividades	34
2.3 Modelos de Projeção da Demanda por Transporte Baseados em Atividades	35
2.3.1 Modelos de escolha discreta	36
2.3.2 Modelos baseados em regras	47
2.4 Considerações Adicionais	55
Capítulo 3 - Considerações sobre Construção de Cenários Futuros	56
3.1 Definições	56
3.2 Abordagem Tradicional para Construção de Cenários Futuros	58
3.2.1 Abordagem proposta por Kahn	59

3.3 Abordagem Perspectiva para Construção de Cenários	60
3.4 Comportamento do Indivíduo na Construção de Cenários Futuros	61
Capítulo 4 - Delineamento do Experimento	63
4.1 Ferramenta de Mineração de Dados	63
4.1.1 Características gerais	63
4.1.2 S-Plus	65
4.2 Etapas do Experimento	67
4.2.1 Limpeza dos dados	67
4.2.2 Codificação dos dados	67
4.2.3 Seleção da variável dependente	68
4.2.4 Seleção das variáveis independentes	68
4.2.5 Construção do cenário para projeção da demanda	69
4.2.6 Adaptação dos dados à mudança de comportamento	73
4.2.7 Mineração dos dados: construção do modelo de previsão	76
4.2.8 Previsão dos padrões	77
4.2.9 Levantamento dos padrões de viagens encadeadas dos indivíduos	78
4.2.10 Comparação dos resultados	78
4.3 Considerações Adicionais	82
4.3.1 Sobre o método proposto para a construção do cenário futuro	82
4.3.1 Sobre o método proposto para a construção do cenário futuro	84
Capítulo 5 - Tratamento dos Dados	85
5.1 Região Metropolitana de São Paulo: Caracterização Urbana	85
5.2 Dados: Pesquisa Origem-Destino da Região Metropolitana de São Paulo	88
5.2.1 Pesquisa Origem-Destino de 1987 da Região Metropolitana de São Paulo	88
5.2.2 Pesquisa Origem-Destino de 1997 da Região Metropolitana de São Paulo	89
5.3 Etapas do Tratamento dos Dados	89

5.3.1 Limpeza dos dados	89
5.3.2 Codificação dos dados	90
5.3.3 Variáveis dependentes	93
5.3.4 Variáveis independentes	93
5.3.5 Verificação da mudança de comportamento de indivíduos	101
5.4 Considerações Adicionais	114
Capítulo 6 - Respostas ao Primeiro Problema Proposto	116
6.1 Considerações Iniciais	116
6.2 Incorporando o Caso da Mudança Comportamental dos Idosos	119
6.3 Incorporando o Caso da Mudança Comportamental das Crianças	124
6.4 Incorporando o Caso da Mudança Comportamental das Mulheres	127
6.5 Atributos do Cenário Construído	131
6.5.1 Tamanho da família	131
6.5.2 Posse de automóvel	132
6.5.3 Sexo	133
6.5.4 Situação familiar	134
6.5.5 Grau de instrução	134
6.5.6 Estudo	135
6.5.7 Trabalho	136
6.5.8 Renda individual	136
6.5.9 Renda familiar	138
6.6 Considerações Adicionais	140
Capítulo 7 - Respostas ao Segundo Problema Proposto	142
7.1 Resultados Obtidos pela Mineração dos Dados: Construção do Modelo de Previsão Baseado no Banco de Dados Sintético de 1987	142
7.2 Resultados Obtidos pela Previsão dos Padrões	144
7.3 Levantamento dos Padrões de Viagens Encadeadas Observadas para os Indivíduos do Banco de Dados de 1997	146

7.4 Comparação dos Resultados: Aplicação da Prova de Kolmogorov-Smirnov (K-S)	147
7.4.1 Seqüência dos padrões gerada pelo S-Plus: ordem alfabética	148
7.4.2 Seqüência aleatória dos padrões	156
7.5 Considerações Adicionais	159
Capítulo 8 - Conclusões e Recomendações	160
8.1 Introdução	160
8.2 Resultados Alcançados	162
8.3 Considerações sobre o Método Adotado	163
8.3.1 Quanto à postura científica adotada	163
8.3.2 Quanto à ferramenta de processamento dos dados	164
8.3.3 Limpeza dos dados	165
8.3.4 Codificação dos dados	166
8.3.5 Seleção das variáveis dependentes e independentes	166
8.3.6 Construção do cenário para projeção da demanda	167
8.3.7 Adaptação dos dados à mudança de comportamento	167
8.3.8 Construção do modelo e previsão dos padrões	168
8.3.9 Levantamento dos padrões de viagens encadeadas dos indivíduos	169
8.3.10 Comparação dos resultados	169
8.3.11 Considerações adicionais	170
8.4 Caráter Científico do Trabalho	170
8.5 Recomendações a Trabalhos Futuros	170
Referências	172

Capítulo 1

Introdução

1.1 Definição do Problema

A abordagem de atividades para o estudo do comportamento de viagens urbanas é baseada no princípio que a demanda por viagem é derivada da necessidade de participação em atividades fora do domicílio (BHAT; KOPPELMAN, 2000; ETTEMA, 1996; JONES et al., 1990). Esta abordagem tem conduzido ao que foi denominado como uma mudança no paradigma nas pesquisas sobre comportamento dos viajantes (KITAMURA, 1997; KUPPAM; PENDYALA, 2001; PAS, 1997), caracterizado pela mudança no foco da realização de viagens discretas para padrões de viagens encadeadas, dirigidas por decisões interdependentes, realizadas dentro do contexto domiciliar (JONES et al., 1990).

Esta nova estrutura de entendimento proposta para viagens urbanas preenche a deficiência dos modelos tradicionais ao considerar aspectos como necessidade dos indivíduos, estrutura familiar (CHAPIN, 1974), limites físicos e restrições (dependência, autoridade e capacidade (HÄGERSTRAND, 1970)), na elaboração da estrutura de decisão das atividades nos modelos (BOWMAN, 1998).

Vários foram os modelos desenvolvidos (STARCHILD, AMOS, SMASH, ALBATROSS etc.), diferindo entre si (ETTEMA, 1996):

- Quanto ao fundamento técnico: teoria das escolhas discreta, teoria microeconômica do consumidor e técnicas de inteligência artificial;
- Quanto ao foco ou direcionamento do foco: foco na programação da atividade ou foco nos padrões de encadeamento de atividades;
- Quanto à estratégia de decisão: por exemplo, a maximização da utilidade supõe que os indivíduos agem racionalmente possuindo todas as informações a respeito das escolhas alternativas e possuem a capacidade de maximizar as suas escolhas; em contraste, outras técnicas consideram que o processo de decisão do indivíduo é o de satisfazer-se sem a necessidade de maximizar a satisfação;
- Quanto ao método estatístico;
- Quanto à escolha das variáveis no processo da modelagem.

A todos os modelos propostos existe uma característica em comum: os mesmos foram elaborados para simular cenários de uma realidade levantada sob o efeito de alterações no sistema de transporte e nas práticas políticas, analisando como a demanda se comporta sob o efeito destas transformações.

Em particular, a elaboração de cenários para projeções futuras depende da habilidade do planejador em entender como fatores exógenos afetam a demanda por transporte ao longo do tempo. O potencial para previsão da demanda não está no modelo ou nos procedimentos utilizados, mas na metodologia aplicada e na lógica usada para projetar os fatores exógenos (KANAFANI, 1983).

Da escola de relações internacionais, sob o enfoque de previsões, Bernstein et al. (2000) afirmam que as práticas de planejamento devem ser direcionadas a situações que ainda não ocorreram, com a construção de cenários contendo internamente hipóteses consistentes de como o futuro provavelmente será em um encadeamento lógico de caminhos que levem aos resultados.

Particularmente, em nível do indivíduo, Heijden (1996) sugere que a ação do homem não é em resposta a realidade, mas a uma inferência perceptiva do intelecto humano da realidade, uma construção interna.

Migrando as idéias básicas propostas por Bernstein et al. (2000), sobre eventos que ainda não ocorreram, e Heijden (1996), sobre a inferência perceptiva do intelecto humano sobre a realidade, para o entendimento do comportamento do indivíduo, especificamente em relação ao desempenho de atividades, pode ser assumido como válido que estas novas situações podem produzir alterações cognitivas no encadeamento de decisões para desempenho de atividades, produzindo o que pode ser denominado de mudança comportamental em indivíduos com características socioeconômicas homogêneas em tempos distintos.

Desta forma, fica definido neste trabalho o termo “mudança comportamental” para indivíduos como sendo uma alteração na tomada de decisão em relação a atividades por indivíduos com características socioeconômicas homogêneas, que em épocas distintas, em decorrência de alterações no ambiente e conseqüentemente mudança na percepção da realidade por estes indivíduos, mudam de atitude em relação ao desempenho de atividades.

Em face da inexistência de modelos baseados em atividades que contemplem mudanças comportamentais dos indivíduos em previsões de demanda, emergem as questões a serem propostas na presente pesquisa:

- É possível elaborar um cenário para um modelo de previsão de demanda por transporte que reflita a mudança no comportamento humano, em um período de tempo?
- É possível utilizar uma técnica de modelagem capaz de realizar previsões de padrões de viagens encadeadas no meio urbano, que contemplem estas mudanças comportamentais dos indivíduos?

Uma forma para melhor compreensão da análise de viagens baseada em atividades é a segmentação da demanda, definido por Van de Bilt (1997) como um recurso que simplifica a análise, facilitando a compreensão e interpretação dos comportamentos em relação às viagens, ao identificar grupos mais homogêneos e diminuir a complexidade e variabilidade associada a uma abordagem desagregada.

Segundo Van de Bilt (2002, p. 7):

[...] Somente através de da utilização do enfoque desagregado é possível, num contexto de crescente complexidade, formular e analisar políticas de transportes, para tentar sanar ou minorar os problemas relacionados à demanda crescente por transportes e avaliar os impactos das políticas de transportes sobre os diferentes segmentos da população, tarefa particularmente importante devido a um aumento da preocupação com os critérios de equidade na distribuição dos recursos [...].

Uma técnica de análise desagregada é a segmentação da demanda através da modelagem de padrões de viagens encadeadas com a árvore de decisão (conjunto de regras de classificação que separa, exaustivamente e sem sobreposição, um conjunto de dados em subconjuntos) realizada por minerador de dados.

A partir de um conjunto de dados contendo valores de variável resposta e de variáveis explicativas é obtido um conjunto de regras para classificação ou predição da variável resposta dentro de regras definidas pela divisão hierárquica dos conjuntos de dados em subconjuntos. Estas regras são definidas de acordo com os valores de um grupo de variáveis explicativas em relação a uma variável de interesse para formar grupos homogêneos (BORGONI; BERRINGTON, 2004).

Uma das peculiaridades da aplicação da árvore de classificação e regressão à técnica de modelagem de padrões de viagens encadeadas é a capacidade de representar a natureza probabilística do objeto analisado. Esta técnica foi aplicada nos trabalhos de Ichikawa (2002), Pitombo (2003), Souza (2004), Aguiar (2005), utilizando o modelo de Classificação e Regressão de Árvores (CART), contido nos programas S-Plus 2000 e S-Plus 6.1.

Ichikawa (2002) demonstrou a aplicabilidade do minerador de dados para obter as relações entre padrões de viagens encadeadas e as características socioeconômicas dos viajantes urbanos. Analisando a relação entre variáveis socioeconômicas e padrões de viagens encadeadas, utilizando a mineração de dados, Pitombo (2003) concluiu que há relação entre as variáveis analisadas e os padrões de encadeamento de viagens, indicando que o minerador de dados pode ter um papel importante à análise do comportamento relacionado às viagens.

Ao comparar os resultados da extração de padrões de uma base de informações utilizando a mineração de dados nas regiões urbanas de Bauru, Belém e São Paulo, Souza (2004) identificou que o comportamento de viajantes urbanos sofre influência de políticas urbanas regionais, características socioeconômicas e espaciais.

Aguiar (2005) analisou a possibilidade de reduzir a quantidade de indivíduos na amostra de Pesquisa de Entrevistas Domiciliares, através da mineração de dados.

Os autores supracitados demonstraram a eficiência da aplicação do modelo de mineração de dados para identificar e explicar padrões de viagens encadeadas. O objetivo deste projeto, seguindo a mesma linha de pesquisa, é testar uma aplicação da técnica de mineração de dados (apoiada na análise de viagens baseada em atividades e sob o efeito de mudanças comportamentais) como uma ferramenta de auxílio ao planejamento de transportes pela previsão dos padrões de viagens (Figura 1.1).

Figura 1.1 - Cerne do aparelho conceitual do trabalho

As próximas seções discorrem sobre o desenvolvimento do projeto de pesquisa proposto.

1.2 Justificativa para se Estudar o Problema

Eco (1989) sugere três pré-requisitos como princípios para se justificar o estudo de um problema e defini-lo como tema central de um estudo científico. São eles:

- O estudo deve debruçar-se sobre um objeto reconhecível e definido de tal maneira que seja reconhecido igualmente pelos outros;
- O estudo deve dizer do objeto algo que ainda não foi dito ou rever sob uma ótica diferente o que já se disse;
- O estudo deve ser útil aos demais.

Os três princípios estabelecidos por Eco (1989) foram utilizados para compor a idéia diretriz do presente projeto, pautando-se da seguinte forma: o objeto de estudo é reconhecível por outros pesquisadores da área. A abordagem de atividades para viagens tem sido largamente estudada desde o início da década de 70. Novas técnicas para modelagem de viagens encadeadas vêm sendo empregadas e Pas (1997) cita em seu trabalho que o progresso na área vem sendo creditado ao desenvolvimento de novos métodos para coleta de dados, angariando novos tipos de informações que enriquecem o modelo na produção dos resultados.

Outra abordagem do objeto de estudo, que é o aspecto da previsão, é foco de estudo da comunidade científica desde a II Guerra Mundial, quando havia a preocupação política e econômica do cenário internacional (NEUMANN; ØVERLAND, 2004).

Para a área de transportes, a previsão de viagens, com abordagem em atividades, sempre foi tratada sob a influência de mudanças no sistema de transporte e na política praticada, investigando o impacto das novas condições na demanda.

Em nenhuma das referências que constituem o aparelho conceitual deste trabalho o fenômeno de mudança do comportamento associada à previsão não foi considerado.

Além do mais, este trabalho tem o intuito de ser útil aos demais com a proposição de uma nova abordagem para previsão de viagens, com uma técnica recente, que é a mineração de dados, sob um novo enfoque para previsão da demanda.

Finalizando, o trabalho se justifica na medida em que há um questionamento rigoroso com aparelho conceitual, com reflexão teórica e com conhecimento das diferentes abordagens, das diferentes interpretações teóricas já produzidas, e reflexão crítica sobre elas, princípios estabelecidos por Beaud (1996) para um trabalho de caráter científico.

1.3 Idéia Diretriz

Segundo Beaud (1996), a idéia diretriz deve responder aos problemas propostos e subentender o conjunto do movimento do projeto.

Os problemas propostos a serem respondidos neste projeto de pesquisa relacionam dois aspectos. O primeiro diz respeito à mudança de comportamento dos indivíduos na previsão de demanda por viagens. O segundo, relacionado ao primeiro, refere-se à utilização de uma técnica de modelagem para realizar previsões de padrões de viagens encadeadas no meio urbano, contemplando mudanças comportamentais dos indivíduos.

Para responder a esses problemas, formula-se a hipótese de que o minerador de dados pode ser utilizado como uma ferramenta para previsão dos padrões de viagens, inclusive sob mudança comportamental. Esta hipótese será submetida a um teste de verificação para confirmá-la ou refutá-la.

1.4 Discurso do Método

1.4.1 Postura científica adotada

As principais linhas de pensamento sobre o método científico são: a visão baconiana de Francis Bacon (BACON, 1979), visão popperiana de Sir Karl Popper (POPPER, 1975) e a visão intermediária entre os pensamentos baconiano e popperiano de Hempel (1981).

A visão baconiana (BACON, 1979) do método científico consiste na idéia de que a ciência se desenvolve por meio de processos indutivos, do particular para o geral, e que “a confirmação das leis obtidas depende apenas de constante verificação em casos concretos, de certo “acordo” entre o que a generalização afirma e o que efetivamente acontece”, Kawamoto (1987, p. 90).

Para Popper (1975), a confirmação de uma hipótese é impossível, contudo ela pode ser refutada, portanto, pela visão popperiana, a ciência avança propondo teorias, desenvolvendo esforços para refutá-las e só aceitando provisoriamente aquelas teorias que sobrevivem a esse processo.

Para Hempel (1981) as inferências indutivas levam de premissas sobre casos particulares a uma conclusão que tem o caráter de lei geral ou princípio, acrescentando que as premissas de uma inferência indutiva implicam a conclusão apenas com maior ou menor probabilidade, enquanto as premissas de uma inferência dedutiva implicam a conclusão certamente. A investigação científica é indutiva num sentido mais amplo, na medida em que aceita hipóteses baseadas em dados que não fornecem para ela evidência conclusiva, mas lhe conferem apenas um “suporte indutivo” ou confirmação mais ou menos forte.

Para o presente projeto de pesquisa será adotada a postura indutiva de Hempel, na medida em que o estudo é baseado em um estudo de caso, e não se associa a uma “evidência conclusiva”, mas apenas uma indicação de que a hipótese é verdadeira.

As outras posturas científicas são indeferidas pelos seguintes motivos:

- A visão baconiana é insustentável na medida em que “fatos” ou dados empíricos só podem ser qualificados como logicamente relevantes ou irrelevantes relativamente a uma dada hipótese, e não relativamente a um dado problema (HEMPEL, 1981).
- Há quatro pontos passíveis de discussão sobre o esquema apresentado por Popper, citados por Kawamoto (1987, p. 90) apoiado em Beveridge (1981):

[...] a) Popper não se ocupa da origem da hipótese; b) a política de concentrar-se na refutação de hipótese, de demolir teorias é negativa em muitos aspectos; c) a tentativa de refutação da hipótese, tema central da filosofia popperiana, depara, na prática, com a dificuldade de projetar um experimento realmente crucial; d) a ênfase na afirmação de que as únicas hipóteses válidas são aquelas suscetíveis de teste e de possível refutação é exagerada, visto que muitas hipóteses valiosas são impossíveis de serem testadas [...].

1.4.2 O experimento

- Etapas de verificação

A verificação a que a hipótese será submetida é o experimento, que é constituído pelas seguintes etapas: limpeza dos dados; codificação dos dados, seleção da variável dependente, seleção das variáveis independentes, verificação da mudança de comportamento de indivíduos, ajuste dos dados à mudança de comportamento, construção do modelo, previsão dos padrões, levantamento dos padrões de viagens encadeadas dos indivíduos e análise dos resultados (Figura 1.2).

Figura 1.2 - Esquema das etapas do experimento

▪ Síntese do experimento

Para realização do projeto, será realizado um estudo de caso baseado nos dados oriundos da Pesquisa Origem - Destino realizada pelo METRÔ-SP na Região Metropolitana de São Paulo (RMSP) nos anos de 1987 e 1997. Os dados da primeira pesquisa irão fornecer as probabilidades das distribuições dos padrões de viagens encadeadas relacionadas às características socioeconômicas dos viajantes urbanos, obtidos através da mineração de dados pela técnica da Árvore de Decisão e Classificação, ferramenta disponível no software S-PLUS 6.1. As probabilidades relacionadas às características socioeconômicas serão apli-

casas aos indivíduos da amostra de 1997, estimando as quantidades de viajantes urbanos por padrão de viagem encadeada. Os valores estimados pelo modelo serão comparados com os dados observados em 1997.

1.5 Estrutura do Trabalho

Além deste capítulo, o trabalho é estruturado em mais seis partes. No capítulo seguinte é apresentada uma revisão das referências que constituem o aparelho conceitual deste trabalho, com ênfase aos modelos baseados em atividades desenvolvidos. Uma abordagem sobre estudos da área de demografia com atenção à construção de cenários futuros e a mudanças comportamentais é realizada no capítulo 3. No capítulo 4 é apresentado o delineamento do experimento e as etapas que o envolvem. A apresentação dos dados e a forma como eles foram tratados são descritos no capítulo 5. A análise dos resultados é apresentada nos capítulos 6 e 7 e na última parte deste trabalho são tecidas considerações a respeito do trabalho.

Capítulo 2

Viagens Baseadas em Atividades

2.1 Análise de Demanda por Transportes: Novo Enfoque

Caracterizando o estudo de demanda por transporte, Kanafani (1983, p. 1) destaca que:

[...] a necessidade por transporte emerge da interação entre atividades econômicas e sociais difundidas no espaço. A diversidade destas atividades e a complexidade dos padrões de interação resultam em uma infinidade de determinantes para as necessidades de transporte. As razões para necessidades de deslocamentos das pessoas são infinitas; da indispensável busca por alimento e por agasalho ao exercício voluntário de deslocamento para lazer. Bens materiais são também transportados de um lugar a o outro por uma infinidade de motivos que emergem das necessidades de mercado [...].

Embora a concepção da idéia de que a demanda por transporte é uma demanda derivada da necessidade por realização de atividades e considerando as atividades como fomentador das viagens, os estudos nesta área foram devotados inicialmente às viagens discretas como unidade de análise a exemplo do modelo de quatro etapas, constituído de quatro passos seqüenciais: geração de viagens, distribuição de viagens, escolha modal e alocação das viagens às rotas.

Na primeira etapa, geração de viagens, medidas das frequências dos deslocamentos são realizadas. As viagens são representadas como deslocamentos terminais com distinção para as viagens produzidas e atraídas. Na segunda etapa, distribuição de viagens, os deslocamentos produzidos e atraídos são distribuídos (número de viagens por par de zonas). Na terceira etapa, escolha modal, as viagens de cada par de zonas são divididas entre os modos de transporte. Finalmente na última etapa, alocação das viagens às rotas, a cada arco da rede é atribuído um número de viagens para obtenção dos volumes.

Embora seja largamente utilizado nos pacotes computacionais de auxílio ao planejamento de transportes, a exemplo do TRANPLAN, MINUTP, TMODEL, TRANSCAD e EMME/2, o modelo de quatro etapas apresenta uma deficiência quanto ao método ao considerar a viagem como unidade de análise. Esta deficiência é considerada por Bhat e Koppelman (2000) como “problema conceitual fundamental”.

A dependência entre viagens não é considerada no modelo de quatro etapas. Distintos modelos são elaborados para viagens com origem residencial e para origem não residencial, desta forma a modelagem desconsidera a sequência de atividades que envolveria as viagens. Além do mais, a programação das viagens não é levada em conta e similarmente não há distinção entre os períodos em que são realizadas as viagens. Em síntese, a organização das viagens e a consequência da inter-relação entre os atributos das múltiplas viagens são ignoradas.

É incompreensível que indivíduos determinem a quantidade de viagens com origem na residência e fora dela separadamente. Pelo contrário, é conveniente que as necessidades e desejos dos indivíduos sejam traduzidas em uma quantidade de atividades seguida por um conjunto de decisões que determinam a quantidade de viagens necessárias ao cumprimento destas atividades.

A este ponto é que emerge a abordagem de atividades para viagens na análise de demanda por transporte, mudando o foco da realização de viagens discretas para padrões de viagens encadeadas, dirigidas por decisões interdependentes, realizadas dentro do contexto domiciliar.

2.2 Fundamentos da Teoria de Viagens Baseada em Atividades

Podem ser citados como fundamentos da teoria de viagens baseadas em atividades: viagem é uma demanda derivada; disponibilidade para desempenho das atividades; ênfase no domicílio como unidade de tomada de decisão; padrão de atividades; processo de elaboração da decisão.

2.2.1 A viagem é uma demanda derivada

Na maioria dos casos a viagem não é uma demanda independente, ela advém da necessidade de participação em atividades que ocorrem em locais distintos. Como consequência as características das atividades influenciam o comportamento individual subjacente à viagem. As atividades a serem desempenhadas dependem das características individuais dos indivíduos (psicológicas, econômicas e sociais), de suas necessidades e de seu estilo de vida (definido como a combinação de papéis desempenhados pelo indivíduo na residência, no trabalho etc). Além disso, geralmente as atividades possuem diferentes prioridades. Atividades como trabalho e estudo têm relevância maior que atividades como lazer. A interação entre tempo e custo requeridos para as atividades e suas prioridades determinam quais as atividades serão desempenhadas.

Em um dos estudos pioneiros nesta área, Chapin (1974) argumenta que o indivíduo realiza atividades para preenchimento das necessidades individuais, correlacionando

as viagens para o preenchimento destas necessidades através do desempenho de um padrão diário de atividades. Chapin (1974) sugere que há quatro forças que determinam as atividades, são elas: propensão, oportunidade, horário e circunstâncias apropriadas, e contexto.

- Propensão

O aspecto propensão diz respeito a fatores relacionados à satisfação de necessidades fisiológicas (repouso, alimentação, etc), de segurança (estar seguro), social (relacionamento com outras pessoas), estima (desenvolvimento na carreira profissional, obtenção de bens materiais) e de auto-realização (realizar atividades que dão prazer ao indivíduo).

Segundo Maslow (BANDURA, 1969; SKINNER, 1965), parece existir uma hierarquia em que se enquadram as necessidades humanas. Em ordem hierárquica a necessidade fisiológica é a mais alta, porque tende a ser mais intensa enquanto não for de alguma forma satisfeita. Enquanto não forem satisfeitas no nível exigido para a operação suficiente do corpo, a atratividade para o desempenho das atividades ligadas às outras necessidades oferece pouca motivação.

Uma vez atendidas as necessidades fisiológicas, seguindo a escala hierárquica proposta por Maslow, a subsequente necessidade é a de segurança. Esta necessidade é essencialmente o anseio por estar livre do medo de perigo físico e da privação das necessidades fisiológicas básicas.

Satisfeitas as necessidades fisiológicas e de segurança, começará a emergir como dominante a necessidade social. Como as pessoas são seres sociais, sentem necessidade de pertencerem a vários grupos e de serem aceitas por estes.

Depois que uma pessoa começou a satisfazer à sua necessidade de participar, ela almeja ser mais que um simples membro do grupo. Passa a sentir a necessidade de estima, tanto de auto-estima como de reconhecimento por parte dos outros. A satisfação dessa necessidade de estima produz sentimentos de confiança em si mesmo, de prestígio, de poder e de controle.

Estando adequadamente satisfeita a necessidade de estima, começa a predominar a necessidade de auto-realização, que é a necessidade que as pessoas sentem de maximizar o seu próprio potencial, sofrendo alterações ao longo vida quando seus atributos físicos mudarem ou seus horizontes se ampliarem.

Esta hierarquia não segue necessariamente o padrão descrito por Maslow em todos os indivíduos, mas segundo o próprio Maslow este é um padrão típico que se verifica na maioria das vezes.

Além destes fatores, há os que dizem respeito às características individuais como sexo, estágio no ciclo de vida etc. e os papéis ocupados por cada indivíduo na sociedade.

- Oportunidade

Oportunidade está associada a variáveis físicas e espaciais que afetam a probabilidade da escolha de uma atividade, reflete principalmente a disponibilidade de acesso aos locais para o desempenho de atividades específicas.

- Horário e circunstâncias apropriadas

Estes fatores referem-se às oportunidades de uma atividade ocorrer em certo horário (sua realização completa) considerando os horários de realização de outras atividades. Circunstâncias apropriadas dependem de vários fatores como a disponibilidade de acessórios (uso de trajes de banho se for praticar natação) ou condições do tempo.

- Contexto

Por último, atividades são influenciadas pelo contexto que engloba aspectos não-psicológicos divididos em fontes internas e externas que influenciam as atividades.

As fontes internas da mudança são representadas por quatro situações interligadas em um sistema de ação-resposta e readaptação-reação, representando como os indivíduos reagem ao resultado das atividades sob alterações do contexto da seguinte forma:

- Ação: a execução das atividades pode conduzir a uma mudança nas características individuais;
- Resposta e readaptação: os resultados produzidos pelas atividades podem conduzir a uma mudança na atitude e motivação em relação às atividades;
- Reação: os resultados alcançados na realização de atividades motivam as pessoas a modificarem a oportunidade de desempenhar as atividades, por

exemplo, a um local de grande densidade de acesso a serviços; e indivíduos podem adaptar seus horários às circunstâncias de ocorrência das atividades.

As fontes externas que modificam constantemente o meio ambiente dizem respeito à evolução cultural, econômica, política e social do meio em que o indivíduo pertence.

2.2.2 Disponibilidade para desempenho das atividades

Uma segunda importante noção sobre a teoria de viagens baseada em atividades é de que o desempenho das atividades depende da disponibilidade de acesso à execução de determinadas atividades ou em função de compromissos agendados que impõe determinadas limitações de horário ao possível desempenho de atividades.

Atividades como compras, trabalho e estudo usualmente ocorrem em um número limitado de lugares que obedecem a um período de funcionamento. De outra forma, certas atividades, como compromissos sociais, submetem-se informalmente a um horário específico.

Outras restrições são abordadas por Hägerstrand (1970). Além das restrições já mencionadas que dizem respeito à autoridade e que são institucionalmente impostas dentro de uma sociedade, como horário de funcionamento de serviços e regulamentações como restrições ao barulho, por exemplo, há restrições que correspondem à dependência e a capacidade.

A restrição de dependência é uma limitação imposta pela natureza ou pela tecnologia. Um exemplo deste tipo de imposição é a necessidade de retorno do ser humano ao domicílio para descanso ao final do dia.

A restrição de capacidade significa que o indivíduo necessita da presença de outra pessoa ou algum outro recurso para conseguir realizar uma atividade.

Além disso, a necessidade de deslocamento para realização das várias atividades em locais distintos de acesso a serviços determina a quantidade necessária de viagens para ir a estes locais sob a imposição de outras restrições como disponibilidade de tempo, custo de viagem etc. Hägerstrand (1970) denomina esta restrição como prisma espaço-tempo. Pessoas se deslocam de um ponto a outro baseado na experiência do tempo e no custo que a pessoa percebe desse deslocamento.

2.2.3 Ênfase no domicílio como unidade de tomada de decisão

O domicílio afeta as atividades dos indivíduos e sua decisão de viagem. Normalmente, os indivíduos agem dentro de um contexto familiar, dividindo tarefas e compartilhando recursos com outros membros do domicílio. Algumas decisões tomadas no domicílio podem ser consideradas como uma unidade (a decisão do chefe de família pela participação de um curso de informática).

A decisão de um indivíduo também pode ser influenciada pelos outros membros (a presença de crianças em idade de amamentação força a presença da mãe na residência). A composição do domicílio e a idade dos membros também afetam as decisões individuais e domiciliares. Estas condicionantes influenciam os padrões de atividades de cada indivíduo no domicílio.

2.2.4 Padrão de atividade

A viagem deve ser considerada dentro do contexto do padrão de atividade que representa as decisões básicas da participação em atividades e suas prioridades e os locais de realização de cada atividade em uma configuração de viagens com episódios no domicílio (BOWMAN, 1998). Isto implica interdependência existente entre eventos ao longo do dia.

2.2.5 Processo de elaboração da decisão e programação de atividades

As formas como os indivíduos tomam decisões relativas ao comportamento das viagens e a atividades é uma questão a ser suscitada. Sem especificamente direcionar o enfoque a questão dos mecanismos cognitivos subjacentes ao processo decisório, diferentes estágios no processo decisório são direcionados em função de variáveis independentes e dependentes que são relevantes a cada estágio. Bowman (1998) sugere os estágios esquematizados na Figura 2.1 no processo de elaboração da decisão:

Figura 2.1 - Esquema do processo decisório para viagens e atividades (BOWMAN, 1998)

- Desenvolvimento urbano: melhorias e implementações no transporte público, mudanças nas localizações de oportunidades de emprego ou novos centros urbanos influenciam diretamente as decisões de viagens individuais e domiciliares;
- Decisões de mobilidade e estilo de vida: decisões como escolha do local da residência, do local de trabalho, decisão de comprar ou não um automóvel e a escolha de atividades a serem desempenhadas em um intervalo regular de tempo, que são realizadas para longos períodos, determinam em geral as condições para realizações de atividades e viagens que se manterão inalteradas por um período de tempo;
- Programação das atividades e viagens: refere-se à implementação de atividades e viagens em uma situação específica englobando decisões sobre quais atividades e viagens serão realizadas em um período específico, a seqüência, o horário e a duração das viagens e atividades;
- Implementação e reprogramação: esta fase refere-se a do processo de monitoramento da execução do plano estabelecido na fase de programação da atividade e adaptação da programação para situações imprevistas;
- Desempenho do sistema de transporte: as decisões individuais, juntamente com o desenvolvimento urbano, afetam o desempenho do sistema de transporte. Isto se manifesta de várias maneiras, inclusive volumes de viagens, velocidades, congestionamentos e impactos ambientais. Estas manifestações dos sistemas de transporte afetam simultaneamente o desenvolvimento urbano e as suas decisões individuais.

2.3 Modelos de Projeção da Demanda por Transporte Baseados em Atividades

Ao final da década de 70 pesquisadores iniciaram a operacionalização das teorias propostas por Hägerstrand (1970) e Chapin (1974), através de modelos matemáticos em um esforço visando à descrição dos padrões de viagens encadeadas a partir de grupos

de características sócio-demográficas para identificar como as características de cada grupo afetam a viagem e o comportamento da atividade.

Basicamente, os modelos devem simular cenários para análise do comportamento da demanda, calibrados com dados coletados a partir de uma realidade apurada (ex. diários de atividades dos viajantes urbanos, pesquisa de entrevista domiciliar, informações sobre sistema de transporte, práticas políticas vigentes e etc).

Na estrutura dos modelos as simulações de tomadas das decisões pelos indivíduos relativas às viagens devem representar a programação das atividades diárias, preservando restrições de desempenho das atividades e a formulação do processo decisório, fundamentado matematicamente em uma estrutura de encadeamento lógico.

Os modelos devem simular cenário futuros (alterações no sistema de transporte, práticas políticas, incluindo a possibilidade de mudanças comportamentais sofridas pelos indivíduos nas previsões), analisando o efeito da demanda sob a influência destas transformações.

2.3.1 Modelos de escolha discreta

A principal abordagem dos modelos desta natureza é que no processo decisório o indivíduo tentará maximizar a utilidade com sua escolha, sendo o resultado traduzido em termos de probabilidade de relacionar uma alternativa a partir de características individuais e características do conjunto de alternativas que poderiam ser escolhidas.

Em geral os modelos de escolha postulam que “a probabilidade de um indivíduo escolher uma determinada opção é em função de suas características socioeconômicas e da

relativa atratividade da opção” (ORTÚZAR e WILLUMSEM, 1994, p. 208) ou como define Arruda (2005, p.31) “a probabilidade de um indivíduo escolher uma determinada alternativa é função da atratividade da alternativa escolhida em relação à atratividade das outras alternativas disponíveis”.

Modelos desta natureza são utilizados para descrever e modelar uma grande variedade de escolhas relacionadas a viagens como modo, rota e destino. Também são aplicados na escolha por um indivíduo de uma programação diária de atividades, de padrão de atividades e viagens, assumindo diferentes abordagens quanto ao processo de programação de atividades. Os mais simples e antigos dividem a programação de atividades em viagens (do local de uma atividade a outro local para realização de outra atividade). Os mais recentes tratam a programação em jornadas (uma sequência de viagens com origem e destino final no mesmo local).

Os modelos de escolha discreta podem ser distinguidos segundo a forma de distribuição do erro na percepção da utilidade de cada alternativa. Os modelos *probit* assumem que o erro para a escolha de cada alternativa segue uma distribuição normal. A maior desvantagem deste tipo de modelo é que limita sua aplicabilidade para previsão da demanda é o cálculo das probabilidades de escolhas das alternativas que envolvem integrais complexas que não podem ser solucionadas analiticamente. O segundo grupo, e o mais utilizado, correspondem aos modelos *logit*, que não assumem a distribuição normal para o erro.

Os modelos *logit* assumem a distribuição de Gumbel para o erro. Uma das limitações deste modelo é a propriedade de independência entre alternativas irrelevantes que assume que erros de alternativas separadas são identicamente e independentemente distribuídos, ou seja, os desvios de todos os erros são iguais e não há correlação entre eles.

Quanto a este problema duas outras abordagens são sugeridas: modelos *universal logit* e modelos aninhado *logit*.

Os modelos denominados de *universal logit* incluem características de alternativas concorrentes nas funções de utilidade. Apesar de permanecerem identicamente e independentemente distribuídos, o efeito causado pela introdução de características de alternativas concorrentes permite uma modelagem de comportamento do tipo não independente da alternativa irrelevante.

Modelos aninhado *logit* assumem que a distribuição do erro a cada dimensão ou a combinação de dimensões pode ter um valor escalar diferente. Conseqüentemente, se todos os componentes de erro são independentes, a covariância entre duas alternativas que dividem uma dimensão comum pode ser especificada de maneira simples.

Os modelos *logit* englobam a maior parte das variáveis do processo decisório, são flexíveis (habilidade de respostas às alterações introduzidas) e oferecem a oportunidade de uso de ferramentas estatísticas para calibração e projeção.

Segundo Ettema (1996) os modelos são baseados em dados de atividades declaradas não podendo ser utilizados como fontes de informações em casos de reprogramação das atividades. Isto é, devido à natureza da estrutura dos modelos *logit* ignora o fato que diferentes decisões subjacentes às atividades podem ser tomadas de diferentes maneiras.

Contudo, a maior crítica aos modelos *logit* é a propriedade do tratamento da duração das atividades como uma variável exógena, sendo esta dimensão uma variável relevante para a modelagem, por que ela é que determina o tempo das viagens.

Os modelos *logit* apresentam limitações quanto à aplicação em projeções. Foram elaborados basicamente para simular os efeitos de práticas políticas e mudanças no sistema de transporte no comportamento da demanda por transporte. Devido à base de sua formulação, alimentados com dados que revelam as atividades declaradas, os modelos *logit* são insensíveis quanto às alterações futuras no cenário como mudança no comportamento humano (devido a fatores externos ou até mesmo pela mudança da percepção da realidade por parte do indivíduo na evolução da faixa etária).

Exemplos de modelos de escolha discreta são: Simulação de Respostas de Viagem e Atividade para Decisões Logísticas Interativas do Complexo do Domicílio (*Simulation of Travel / Activity Responses to Complex Household Interactive Logistic Decisions* - STARCHILD, RECKER et al., 1986a; RECKER et al., 1986b), o modelo desenvolvido com dados da cidade de Boston (BOWMAN, 1995) e o Sistema de Programação Diária da Atividade para Portland (*Portland Day Activity Schedule System*, BOWMAN, 1998).

O STARCHILD, o modelo desenvolvido para a cidade de Boston e o *Portland Day Activity Schedule System* são exemplos de modelos que fazem parte da terceira etapa do desenvolvimento de modelos econômicos desagregados baseados em atividades.

Integram a primeira etapa modelos de escolha discreta baseados em viagens que usam interligações no modelo representado restrições de tempo e espaço e interações dos membros do domicílio. Um exemplo de modelo desta etapa é o sistema desenvolvido para a Comissão de Transporte Metropolitana (*Metropolitan Transportation Commission - MTC*) na década de 70 para a área metropolitana da Baía de São Francisco.

A segunda etapa de desenvolvimento surgiu com os modelos baseados em jornadas na Holanda. Estes modelos capturam o efeito das viagens encadeadas utilizando a

jornada como unidade básica de decisão. O modelo desenvolvido para a cidade de Estocolmo desenvolvido na década de 90 é um exemplo.

Na cadeia evolutiva, a terceira etapa é caracterizada por modelos que introduzem uma maior integração, representando as escolhas diárias dos indivíduos (atividades e viagens). Os aspectos que envolvem os modelos baseados em viagens e jornadas são prolongados, incluindo aspectos como uso do solo.

- Simulação de Respostas de Viagem e Atividade para Decisões Logísticas Iterativas do Complexo do Domicílio - STARCHILD

STARCHILD (RECKER et al., 1986a; RECKER et al., 1986b) inicia com a introdução de dados ao modelo contendo informações detalhadas do programa de atividade. Através destas informações é possível identificar muitos detalhes da programação, incluindo propósito, duração e localização, assim como restrições das atividades. A modelagem é iniciada em um processo de quatro etapas que gera o horário e a sequência das atividades. Nas duas primeiras etapas é gerado o conjunto de escolhas. As possíveis alternativas são enumeradas exaustivamente baseadas nas restrições. As duas etapas restantes compreendem o processo da escolha. Um modelo *logit* (*multinomial logit*) é utilizado para representar a melhor escolha do conjunto. Os criadores do STARCHILD conceberam a programação de atividades como um plano, que é seguido pela sua execução e reprogramação, Figura 2.2.

A operação do modelo inclui cinco modelos internos: TROOPER é o módulo responsável pela preparação dos dados, criando programa de atividades em nível do indivíduo baseado nas informações provenientes dos diários de atividades. SNOOPER é o responsável pela programação operacional, identificando possíveis padrões de atividades e o preenchi-

mento de lacunas na programação por atividades. GROOPER, SMOOPER e CHOOSER gerenciam juntamente o processamento até a escolha do padrão de atividades. SMOOPER é responsável pela redução inicial das alternativas utilizando um algoritmo de decisão multi-objetivo. CHOOSER é o módulo responsável pela escolha do padrão de atividades, baseado em um modelo multinomial *logit*.

Figura 2.2 - Esquema do modelo STARCHILD (BOWMAN, 1998)

De acordo com Bowman (1998), o STARCHILD apresenta duas falhas no processo de previsão. Primeiro, ele recai em fontes externas para realizar previsão de importantes dimensões do programa de atividade e de viagens como participação em atividades, propósito, localização e modo. Segundo, a classificação e a utilização de regras podem inadequadamente representar o verdadeiro conjunto de escolhas. As regras utilizadas geram um pequeno conjunto com somente uma alternativa para cada classe distinta, enquanto as pessoas freqüentemente realizam escolhas de um conjunto similar de alternativas concorrentes.

Analisando o STARCHILD, Buliung (2005, p. 5), cita:

[...] A versão operacional do STARCHILD não captura efeitos das interações do domicílio e do meio abordados na teoria devido a limitações dos dados. Apesar do papel dos membros do domicílio não ser explorado totalmente, as suas informações são utilizadas para construir os programas de atividades (TROOPER), permitindo programações avançadas, escolhas dos padrões e formação do conjunto de atividades. De fato, o indivíduo assume o papel primário da tomada de decisão em contextos operacionais e teóricos. Um dos mais interessantes aspectos do modelo é a sensibilidade do indivíduo quanto à ocorrência de eventos imprevistos (atividades não planejadas). É difícil generalizar este aspecto pela pequena amostra utilizada no modelo. Interessantemente os desenvolvedores sugerem que o STARCHILD propicia uma plataforma para experiências com especificações alternativas do padrão de formação das atividades [...].

- Modelo desenvolvido com dados de Boston

O modelo de Boston (BOWMAN, 1995) foi desenvolvido para simular a programação diária de atividades e viagens para previsão de viagens de passageiros urbanos com dados de viagem e do sistema de transporte da cidade de Boston numa pesquisa realizada em 1991. Basicamente o modelo é dividido em cinco partes: padrão de atividades, estrutura de jornada, horário da jornada, escolha do destino e modalidade da jornada.

O modelo do padrão de atividades é baseado em um modelo aninhado *logit*, representando a escolha entre um padrão de atividades sem viagem e um entre os 54 padrões possíveis de atividade contendo o motivo, modo e período da viagem.

Cada padrão de viagem encadeada é composto por uma jornada primária e jornadas secundárias. A jornada primária é definida como a jornada que expressa a atividade mais importante do dia.

O cálculo da probabilidade de cada jornada consiste na probabilidade conjunta de todas as dimensões (modo, destino, e etc) modeladas para todas as jornadas na programação. As jornadas secundárias são condicionadas ao resultado da jornada primária. Desta forma a probabilidade das jornadas é o produto da probabilidade da jornada primária.

ria e da probabilidade condicional dos resultados das jornadas secundárias (consideradas mutuamente independentes), dada a jornada primária.

Dois modelos *multinomial logit* estimam a escolha do horário da jornada, um para a jornada primária e outro para as jornadas secundárias. Os tempos de partida são alocados em períodos do dia estabelecidos. Todos os períodos são considerados disponíveis para todas as pessoas para jornadas primárias e nas jornadas secundárias os horários que se sobrepõem ao horário da primeira jornada são excluídos.

A escolha do modo e do destino envolve a escolha do modo para a jornada ao invés da usual escolha do modo para a viagem. Apesar de os dados que alimentaram o modelo conterem o modo em cada viagem, foi adotado um único modo para toda a jornada. Regras são usadas para transladar os vários modos utilizados para um conjunto de escolha de modos alternativos para toda a jornada. O resultado final do modelo é a previsão da demanda por viagem por par de origem-destino por período.

Avaliando este modelo, Bowman e Ben-Akiva (2000) citam como falhas: a modelagem do horário que não inclui a influência do nível de serviço do sistema de transporte, as paradas nas jornadas secundárias que foram omitidas; a baixa variabilidade dos padrões de viagem em resposta às práticas políticas; algumas relações da hierarquia, como a relação entre a decisão do horário para a escolha do modo e destino, não são totalmente explicadas pelo modelo; a consideração de que as jornadas secundárias são mutuamente independentes é irrealista, violando restrições de tempo e de prioridade e a ausência da modelagem das atividades realizadas no domicílio devido à limitação dos dados.

A quantidade de padrões de viagem encadeados e a consideração que em cada jornada o indivíduo utiliza um único modo de transporte são outras falhas do modelo. Os

54 padrões de viagens encadeadas limitam a análise do comportamento individual e restringe a projeção para a demanda futura. É aceitável que indivíduos se desloquem de sua residência para exercer uma atividade através de um modo e retornar a sua residência com outro modo.

- Sistema de Programação Diária da Atividade para Portland

Há uma grande compatibilidade de características entre o modelo desenvolvido para Boston e o Sistema de Programação Diária da Atividade para Portland, incorporando muito da teoria desenvolvida e o aprendizado das falhas no modelo de Boston.

O modelo para Portland é um sistema de modelos desagregados logit e modelos logit aninhados com cinco tipos de modelos na hierarquia (Figura 2.3): modelo de padrão de atividades, modelo para horários das jornadas com origem e destino no domicílio, modelo para escolha de destino e modo da jornada primária com origem e destino no domicílio e modelo para jornadas com origem e destino no local de trabalho e paradas intermediárias.

Sobre a estrutura do modelo de Portland, Bowman et al. (1998, p. 3) consideram que:

[...] Um modelo ideal de atividade inclui todas as informações da cadeia de atividades em que cada pessoa no domicílio é envolvida durante um dia, incluindo hora, duração, tipo de atividade, localização, modo, e tempo de viagem para cada atividade. A estrutura do modelo é desenhada com o intuito de chegar o mais próximo o possível do modelo ideal, considerando a necessidade imediata de implantação pelo órgão de planejamento metropolitano, Metro, da cidade de Portland. Embora concessões tenham sido feitas, as características principais do modelo baseado em atividades foram mantidas [...].

Figura 2.3 - Estrutura do modelo de programação da atividade para Portland (BOWMAN, 1998)

O modelo de padrões de atividade determina simultaneamente para a jornada primária e as jornadas secundárias os propósitos das atividades, entre uma das categorias subsistência, manutenção e lazer enquadrando em um dos 114 padrões possíveis. A função utilidade para cada alternativa é contida dentro de um modelo *multinomial logit*.

Definido o padrão de atividades, o modelo para o horário determina a sequência e duração de cada jornada. São estabelecidos cinco períodos do dia. Para cada jornada, o modelo de horário prevê a combinação do horário de partida do domicílio e o horário de partida da atividade primária, totalizando 25 combinações possíveis.

Modelados o padrão e o período de realização das atividades, o modelo realiza a previsão do modo e destino para cada jornada. Corrigindo a falha do modelo aplicado aos dados de Boston, neste modelo há a possibilidade de viagens em uma mesma jornada se-

rem realizadas por modos diferentes (são nove tipos alternativos de modos). A escolha do destino é uma entre as 21 zonas da amostra utilizada no modelo.

Na última etapa nenhum modelo é utilizado para realizar previsões dos horários das jornadas com origem e destino no local de trabalho, ao contrário das etapas anteriores frações baseadas nos dados coletados da pesquisa são utilizadas. A modelagem do modo e a do destino são similares à modelagem para as jornadas com base na residência. O resultado final do sistema é a produção de matrizes de viagens contendo o propósito, horário, modo e faixa de rendimento.

O modelo da programação de atividade diária possui algumas deficiências. A primeira é relacionada à falta de interação entre as atividades realizadas dentro e fora do domicílio. A segunda é causada quando são agregadas várias alternativas discretas reduzindo o tamanho do conjunto de alternativas e quando são modificadas variáveis inicialmente contínuas em variáveis discretas, impedindo ao modelo capturar variações no comportamento pelo mascaramento que pode ocorrer pela junção de grupos de variáveis discretas.

A falta de especificação das funções de utilidade ocorrem quando variáveis importantes estão ausentes ou a forma das funções está incorreta, causando erros na previsão e insensibilidade ao modelo para responder a alterações no sistema. Outras ausências de especificações dizem respeito à estrutura do modelo. Primeiro, no modelo podem ser agrupados duas ou mais dimensões (como a escolha do modo e destino) em um nível do modelo, quando eles devem ser aninhados por causa dos atributos compartilhados não observados em uma dimensão. Segundo, são tratadas incorretamente as jornadas primárias e secundárias como condicionalmente independentes quando suas utilidades são correlacionadas. Terceiro, o modelo utiliza uma estrutura aninhada quando atributos não observa-

dos, compartilhados em uma ou mais dimensões, poderiam ser tratados de uma forma mais complexa do que um modelo *logit* aninhado.

2.3.2 Modelos baseados em regras

Os modelos baseados em regras destinam atenção ao processo de geração do conjunto de escolhas com emprego de regras até atingir um pequeno conjunto de escolhas. Ao contrário dos modelos de escolha discreta que utilizam funções complexas, os modelos baseados em regras aplicam funções simples de utilidade ou de satisfação. Outra diferença com relação aos modelos de escolha discreta é a presença da etapa de interação entre busca e escolha.

O objetivo é avaliar como os indivíduos, dentro de um domicílio, tomam suas decisões de realização de atividades, dadas as condições do ambiente de viagens (sistemas de transporte disponíveis e padrões de uso do solo) e das restrições que podem ser encontradas. Exemplos de modelos desta categoria são: Simulador de Mobilidade e Atividade (*Activity - Mobility Simulator* - AMOS, RDC, 1995), Modelo de Simulação da Heurística Programação de Atividade (*Simulation Model of Activity Scheduling Heuristics* - SMASH), (ET-TEMA, 1996) e Sistema de Simulação Orientado de Transporte Baseado em Aprendizagem (*A Learning-Based Transportation Oriented Simulation System* - ALBATROSS), (ARENTZE e TIMMERMANS, 2004).

- Simulador de Mobilidade e Atividade - AMOS

AMOS (RDC, 1995) é um modelo de viagens baseadas em atividades com foco na adaptação e processos de aprendizado que pessoas exibem quando se deparam com alterações no sistema de transporte. AMOS simula um novo padrão de atividades em resposta a

medidas de gerenciamento de demanda por transporte (TDM - *Travel Demand Management*).

O sistema é composto pelos seguintes módulos: analisador do padrão de viagens encadeadas (*baseline activity-travel analyzer*), gerador de opções de respostas ao TDM (*TDM response option generator*), modificador de padrões de atividade (*activity-travel pattern modifier*), módulo de avaliação e aceitação de rotinas (*evaluation module and acceptance routines*) e registro estatístico (*statistics accumulator*).

O analisador do padrão de viagens encadeadas realiza a leitura dos registros das viagens, gerando o padrão de viagem encadeada para cada indivíduo. Em seguida o gerador de opções de respostas ao TDM cria a resposta básica para cada indivíduo a partir do eixo de raciocínio para padrões de viagem, características socioeconômicas e demográficas, e atributos do TDM. Os resultados gerados são respostas comportamentais.

O módulo modificador de padrões de atividade consiste no algoritmo de reprogramação de viagens e atividades. Fornece uma ou mais de uma alternativa de padrão de viagens encadeadas baseado na resposta promovida pelo gerador de opções de respostas ao TDM. A partir do analisador de padrões de viagem e atividade, dados da rede de transporte, dados de uso e ocupação do solo, características socioeconômicas e demográficas e as opções de resposta ao TDM são modificados os padrões de viagens encadeadas.

No módulo de avaliação e aceitação de rotinas é calculada a utilidade, através de um modelo *multinomial logit*, associado ao padrão de viagens encadeadas modificado. Finalmente no último módulo, registro estatístico, todos os padrões de viagens encadeadas são lidos e geradas estatísticas para um dia de rotina.

- Modelo de Simulação da Heurística Programação de Atividade - SMASH

Assim como é requerido pelo STARCHILD, SMASH (ETTEMA, 1996) inicia o modelo requisitando entrada de dados contendo informações detalhadas sobre o programa de atividades. Através de um processo iterativo o modelo gradualmente constrói uma programação a partir das informações das atividades fornecidas inicialmente. Em cada interação ele inicia com uma programação (a primeira está em branco) e conduz a um processo de preenchimento de atividades na programação produzindo programações ajustadas. Estas programações são inseridas, apagadas e substituídas da agenda de atividades até a obtenção da melhor programação, como mostra a Figura 2.4. O processo entre a programação ajustada e a programação aceita é implementada através de um modelo *logit* aninhado.

Uma programação é tida como aceitável quando a utilidade da programação alternativa é maior que de todas as programações consideradas na interação. Uma programação contendo tempos de atividades programados, curtos tempos de viagem, prioridades das atividades e ausência de conflitos tem maior chance de ser aceita.

Figura 2.4 - Esquema do modelo SMASH (BOWMAN, 1998)

A principal crítica ao SMASH é que, além de necessitar de muitos dados para entrada, o que requer uma pesquisa detalhada, a exaustiva busca heurística utilizada pelo modelo pode ser inadequada além de necessitar validação. Seu método de restringir a busca a domínios limitados pode sistematicamente excluir alternativas que poderiam ser escolhidas pelos indivíduos (BOWMAN e BEN-AKIVA, 1997; BOWMAN, 1998). Decisões relacionadas a escolha ao modo de transporte e escolha da rota foram omitidas no SMASH, dimensões importantes no processo-decisório. Ettema (1996) cita que um outro aspecto suprimido é relativo às decisões relacionadas a restrições de capacidade, definido por Hägerstrand (1970) como o indivíduo que necessita da presença de outra pessoa para realização de uma atividade.

- Sistema de Simulação Orientado de Transporte Baseado em Aprendizagem - ALBATROSS

O ALBATROSS (ARENTZE e TIMMERMAN, 2004) é um modelo baseado em regras que simula decisões individuais relacionadas às várias peculiaridades da programação de atividades geralmente consideradas na análise de viagens baseadas em atividades. No modelo as árvores de decisão são utilizadas para representar as escolhas heurísticas dos indivíduos na elaboração de suas atividades utilizando um conjunto de dados sobre viagens e atividades.

Segundo Arruda (2005, p. 44) no ALBRATROSS:

[...] a demanda por viagem é considerada como derivada do processo pelo qual indivíduos programam suas atividades e viagens em um período de tempo. As decisões são tomadas em nível domiciliar, sendo consideradas restrições espaço-temporais e institucionais no processo de programação das atividades [...].

As atividades no modelo são classificadas em fixas (compulsórias) e flexíveis (não compulsórias), considerando que a geração de programações de atividades depende da natureza e de suas prioridades. Na programação é estabelecido que indivíduos buscam realizar objetivos particulares, sujeitos a restrições espaço-temporais e o contexto domiciliar.

Uma vez estabelecidas as restrições que definem o espaço de ação dos indivíduos, a próxima etapa é a definição de como os indivíduos agrupam as atividades em padrão de atividades. O ALBATROSS admite que os padrões de atividades sejam construídos a partir de mecanismos em que o indivíduo determina um padrão de atividades através da experiência adquirida.

A ênfase dada à abordagem de aprendizado no ALBATROSS é uma alternativa proposta em relação à abordagem convencional da modelagem econométrica dos padrões

de atividade governados pela maximização da utilidade. No processo de aprendizagem padrões de busca são transformados, gradativamente, em preferências individuais (localizações, tempo e modo de transporte) para realização de atividades, gerando regras que guiam o processo de tomada de decisão na construção de programações de atividades possíveis de serem realizadas. O resultado final são regras que guiam o processo.

Como o ALBATROSS é um sistema baseado em regras lógicas, para cada etapa de decisão o modelo gera uma árvore de decisão. Essas árvores consistem em algumas variáveis independentes (preditoras) e variáveis dependentes (que são as alternativas de escolha). O algoritmo usado para gerar as árvores de decisão no modelo é o CHAID (*Chi-square Automatic Interaction Detection*). Uma das características do CHAID é a divisão dos nós em múltiplos ramos. Na Figura 2.5 é apresentado o esquema do básico do modelo ALBATROSS.

Figura 2.5 - Esquema do processo de programação do ALBATROSS (ARENTZE e TIMMERMAN, 2004)

No primeiro passo o modelo decide o modo de transporte para a atividade com motivo trabalho. A escolha do modo de transporte para a atividade trabalho representa a prioridade mais elevada, atividade compulsória, por que esta escolha determina qual pessoa usa o carro para uma parte considerável do dia em casos onde há apenas um carro e mais de um indivíduo habilitado para usá-lo.

O segundo passo determina o restante da composição das atividades na programação. Para cada atividade de categoria flexível, a decisão de se adicionar ou não na atividade programação é feita baseada nas regras estabelecidas. Se uma atividade for adicionada, o passo seguinte é a determinação da viagem e a duração da atividade. A duração é determinada de maneira qualitativa como uma escolha entre um episódio longo, médio e curto. As restrições temporais definem a adequação das decisões de seleção e de duração.

O terceiro passo consiste na determinação do horário para realização das atividades flexíveis em ordem de prioridade. A modelagem ocorre considerando uma subdivisão do dia em alternativas discretas.

O quarto passo organiza as viagens entre as atividades levando em consideração a ordem de prioridade e a presença do domicílio entre as atividades, ou seja, se a atividade é conduzida antes de outra atividade fora do domicílio, depois de outra atividade fora do domicílio ou apenas uma jornada com uma parada somente.

O passo cinco envolve a escolha do modo de transporte para cada jornada assumindo que não há troca de modos entre as viagens dentro de uma jornada. Finalmente, o último passo determina a localização das atividades flexíveis em ordem de prioridade.

O modelo constrói simultaneamente uma programação para cada indivíduo adulto do domicílio. Em cada passo da programação, restrições dinâmicas determinam quais escolhas alternativas são adequadas a cada estado da programação.

Um dos pontos fracos do ALBATROSS é a grande quantidade de dados solicitados para gerar a programação da atividade individual. Fundamentalmente, um dos princípios básicos que devem nortear a construção dos modelos de demanda por transporte é a redução do custo e do tempo de análises que auxiliem na adoção de medidas de planejamento.

A quantidade de dados solicitada pelo ALBATROSS demanda a elaboração de um diário de atividade detalhado que deve ser submetido a membros do domicílio através de pesquisa de entrevista domiciliar, o que implica custos para elaboração do diário, realização das entrevistas (acarretando tempo para preenchimento do questionário) e tabulação dos dados. Muito embora ainda não tenha sido comparado com entrevistas domiciliares tidas como convencionais, o questionário elaborado para a entrevista visando alimentação do ALBATROSS é mais extenso do que o questionário de entrevista domiciliar aplicado na Região Metropolitana de São Paulo que detém a experiência de quatro pesquisas decenais de origem-destino.

Este fato pode ser muito bem comprovado no trabalho desenvolvido por Arruda (2005) que tentou aplicar o modelo ALBATROSS na cidade de São Carlos/SP. Além da dificuldade com a busca por dados para alimentar o modelo, há a dificuldade de implementação deste modelo em outra região, neste caso em outro país.

O modelo ALBATROSS foi concebido para uma realidade holandesa, sendo calibrado para as características dos indivíduos daquele país. Alguns parâmetros para calibração do modelo não podem ser alterados, pois fazem parte de uma “caixa-preta”. Assim

sendo, mesmo de posse de todos os dados necessários para calibrar o modelo ainda persiste a barreira da programação do modelo.

2.4 Considerações Adicionais

Neste capítulo foram abordados três aspectos que integram a fundamentação teórica deste trabalho. O primeiro aspecto apresentado introduz o novo enfoque dado por diversos pesquisadores à análise de demanda por transportes, apresentando as razões que motivaram o surgimento desta nova abordagem. O segundo aspecto descreve os principais fundamentos desta nova abordagem em um encadeamento lógico de considerações que permeia a periferia cognitiva do indivíduo ao ambiente que interage com o ser.

Em seguida foi realizada uma leitura dos principais modelos de análise de demanda sob a nova abordagem, mencionados na literatura. Muito das linhas escritas traduzem pontos de vistas de diversos pesquisadores da área, principalmente as falhas dos modelos, e desta forma há a possibilidade de recair em falácias *ad hominem*. Outras partes, contudo, foram inferidas da análise dos trabalhos mencionados que dissecam a estrutura do modelo e seu funcionamento. O intuito com a ênfase à falha dos modelos não é depreciar os modelos. Conforme apresentado na postura científica deste trabalho é assumido neste trabalho que a ciência se desenvolve por meio de etapas, e estes modelos fazem parte das etapas evolutivas no desenvolvimento desta nova abordagem que é proposta para a análise de demanda por transporte. Assim sendo, a ênfase às questões não abordadas nos modelos é para dirigir atenção que novas idéias são imperiosas ao prosseguimento do desenvolvimento do novo enfoque.

Capítulo 3

Considerações sobre Construção de Cenários Futuros

A partir da literatura das Escolas de Relações Internacionais e das Escolas de Políticas e Ciências Sociais são abordados de forma periférica no presente capítulo alguns métodos para a construção de cenários futuros. Não constitui interesse deste trabalho uma abordagem aprofundada deste tema que, como já mencionado, é de competência de pesquisadores das Escolas de Relações Internacionais e de Políticas e Ciências Sociais. Contudo o aparelho conceitual relatado neste capítulo está diretamente relacionado às fases seguintes do trabalho.

3.1 Definições

A projeção da demanda de transportes para anos horizontes requer a construção de cenários. Esta criação servirá para visualizar a maneira como a demanda se comportará sob as mudanças que poderão ocorrer ao longo do tempo. Desta forma, a principal função dos cenários em um modelo de demanda é atuar como uma ferramenta auxiliar impor-

tante. Dos cenários serão extraídas as demandas futuras a serem incorporadas à análise de demanda para planejamento de ações.

Em síntese, o que se espera de cenários futuros para modelos de demanda por transporte é a conversão, através de variáveis, das características que dizem respeito ao ambiente, das características socioeconômicas que se referem aos indivíduos e aos aspectos inerentes ao sistema de transporte em formas possíveis.

A construção de cenários é uma das principais áreas de estudo da escola de Relações Internacionais. Bernstein et al. (2000) consideram que cenários futuros são hipóteses internamente consistentes de como o futuro pode ser e é um encadeamento lógico que conecta os fatores a resultados. Para Heijden (1996) o cenário futuro é um conjunto de futuros razoavelmente plausíveis, mas estruturalmente diferentes entre si devido a diversidade dos fatores existentes. Fahey e Randall (1998) assumem que cenários são narrativas descritivas de projeções alternativas plausíveis de uma específica parte do futuro. Godet e Roubelat (1996) postulam que a construção de cenários é uma descrição de uma situação futura e do curso dos eventos que permite que um cenário se desloque da situação original para a situação futura.

Newman e Øverland (2004) fazem uma distinção entre os termos utilizados para descrever o processo de elaboração de eventos futuros: planejamento de cenários e construção de cenários. Segundo estes pesquisadores a expressão que deve ser utilizada é construção de cenários. Esta expressão comunica que o desenvolvimento de cenários é um processo ativo que envolve construção de blocos ou componentes (variáveis) e a organização de todos estes blocos resultando em um cenário.

3.2 Abordagem Tradicional para Construção de Cenários Futuros

Segundo Neuman e Øverland (2004), muito embora exemplos de tentativas de construção de cenários sejam encontradas na área de planejamento urbano no século XVII, foi após a Segunda Guerra Mundial que estudos para construção de cenários futuros emergiram em importância para Governos.

A Segunda Guerra Mundial introduziu significativas mudanças na economia (capitalismo versus economia planificada), na política (liberalismo versus socialismo), na sociedade (anti-semitismo, emergência do *American Way of Life* etc) entre outros. Em decorrência destas mudanças introduzidas no panorama mundial, incluindo recessões econômicas mundiais, organismos governamentais buscaram formas para desenvolvimento de projeções como tentativa de evitar consequências infelizes, principalmente à economia que não podem ser diretamente gerenciadas.

Surge em 1946, nos Estados Unidos, a RAND Corporation. Uma iniciativa de órgãos de defesa do Governo dos Estados Unidos dedicada especialmente ao estudo de elaboração de métodos de projeções de cenários, principalmente na área militar, com ramificação na área civil através do Hudson Institute (presente até os dias atuais) e Institute for the Future.

Na década de 1960 o estudo sobre construção de cenários passa a também ser explorado por grandes organizações civis, a exemplo da Shell Oil Company e o Stanford Research Institute (SRI) desenvolvendo métodos para elaboração de cenários com propósitos civis.

Emergem neste período duas abordagens tradicionais propostas para construção de cenários futuros. A primeira é proposta pelo matemático Olaf Helmer e a segunda pelo cientista social Herman Kahn. Ambas as propostas são pragmáticas, sendo a de Helmer positivista enquanto a de Kahn é mais próxima da realidade. Contudo, Helmer não enfatiza aspectos da previsão e sim da dificuldade em realizá-la. Já a proposta de Kahn, do ponto de vista metodológico possui uma contribuição maior. Sua principal contribuição é a abordagem de uso de dados históricos para estabelecimento de um “eixo básico”.

3.2.1 Abordagem proposta por Kahn

Kahn (KAHN e WIENER, 1967) define cenário como uma sequência de eventos hipotéticos construídos com o propósito de dar ênfase aos processos e às decisões. A principal contribuição metodológica de Kahn é a proposta de um “eixo básico” (Figura 3.1).

Figura 3.1 – Esquema do “eixo básico” da abordagem de Kahn para construção de cenários

A idéia principal do “eixo básico” proposto por Kahn é que para construir um cenário em um período $T + \Delta T$, primeiro deve ser realizado um retrospecto no período $T - \Delta T$, examinadas as mudanças ocorridas neste intervalo e finalmente, usando o “senso” derivado da análise das alterações ocorridas no passado, especificar o grau da mudança que ocorrerá, incorporado à construção do cenário futuro.

Usando este método o primeiro aspecto a ser identificado é o que Kahn chama de *clusters* de eventos, que se refere às relações entre cursos de desenvolvimento das diferentes áreas implicando que um evento não se realizaria se não houvesse o exercício de um outro evento. O outro aspecto refere-se às variações canônicas que ocorrem ao longo do tempo, pensar o impensável para construção de cenários futuros.

3.3 Abordagem Perspectiva para Construção de Cenários

O termo perspectiva utilizado para esta abordagem deriva originalmente do Latin *perspicicere* que significa “olhar de perto”. Nas artes o termo perspectiva se refere à falsa impressão de profundidade de uma pintura ou desenho. Nietzsche usou o termo perspectiva para ressaltar que cada visão é uma visão de algum lugar, e que cada visão, portanto possui sua utilidade e seus limites.

Esta nova abordagem foi proposta em função da incorporação de grandes organizações ao estudo de cenários futuros, deixando de ser uma ferramenta de planejamento de Governos e passando a ser também uma ferramenta de auxílio aos negócios com avaliação de tendências de mercados.

A partir da proposta tradicional de Kahn, duas principais características do método da abordagem perspectiva para construção de cenário são o processo de desenvolvimento orientado e a elaboração de grupos de discussão.

A técnica de aprendizado é a parte histórica da construção de cenários. História não se limita somente ao estudo e comentários sobre fatos passados, mas atenção ao fluxo da história, conhecimento das mudanças históricas e das idéias sobre as forças que direcionaram o passado, que direcionam o presente e que direcionarão o futuro.

Os grupos de discussão, denominados de Formação de Atores-Formação de Fatores (*Shaping Actors-Shaping Factors*), são formados por membros de áreas multidisciplinares com o propósito de que ocorra o debate sobre os mais diversos aspectos (sociais, políticos, técnicos, econômicos e etc) das variáveis que envolvem a construção de cenários futuros. O esquema da nova abordagem é apresentado na Figura 3.2.

Figura 3.2 - Esquema da abordagem perspectiva para construção de cenários futuros (adaptado de NEUMAN e ØVERLAND, 2004)

Na abordagem perspectiva da construção de cenários o tempo é tratado como variável dinâmica, sendo classificado em passado, presente e futuro. Há uma interligação entre os três tempos, ou seja, na nova abordagem, presente, passado e futuro são analisados conjuntamente.

3.4 O Comportamento do Indivíduo na Construção de Cenários Futuros

Uma das variáveis de importância na construção de cenários futuros é o comportamento do indivíduo. Heijden (1996) sugere que a ação do homem não é em resposta a realidade, mas a uma inferência perceptiva do intelecto humano da realidade, uma cons-

trução interna. Sendo a realidade que o indivíduo vivencia passível de sofrer alterações ao longo do tempo, naturalmente esta inferência perceptiva é alterada e consecutivamente o comportamento individual é passível de sofrer mudança.

Além do mais, segundo Raudenbush (2005) as trajetórias do comportamento individual refletem a interação entre as intervenções no meio ambiente em que o indivíduo vive e as ações individuais.

Nagin e Tremblay (2005) relatam um método para abordagem da mudança do comportamento individual ao longo do tempo. Os indivíduos podem ser agrupados em função das suas características socioeconômicas. Da mesma forma que uma alteração do cenário é percebida por um indivíduo alterando o seu comportamento, essa mesma alteração pode provocar alterações em outros indivíduos (pertencentes a um mesmo grupo) com resposta semelhante ao comportamento.

As mudanças introduzidas no meio em que o indivíduo vivencia podem introduzir diferentes respostas para cada agrupamento de indivíduos, como descrito em Nagin e Tremblay (2005, p. 84):

[...] A modelagem da trajetória do comportamento individual baseada em grupo assume uma diferente abordagem da modelagem de diferenças individuais. Ao invés de seguir que a distribuição das trajetórias varia continuamente pelos indivíduos, é assumido que pode haver *clusters* ou agrupamentos com diferentes desenvolvimentos de trajetórias de comportamento [...].

Referente a esta abordagem um aspecto deve ser ressaltado: cada indivíduo possui uma probabilidade de pertencer a um grupo, ou seja, o indivíduo será associado a aquele grupo que ele possui maior probabilidade de pertencer, contudo não pode ser criada a idéia de imutabilidade para indivíduos.

Capítulo 4

Delineamento do Experimento

Este capítulo aborda aspectos da ferramenta utilizada para realização do projeto e a descrição das etapas do teste de verificação da hipótese proposta.

4.1 Ferramenta de Mineração de Dados

4.1.1 Características gerais

A mineração de dados é um processo analítico desenvolvido para extrair padrões de uma base de dados, que vem sendo aplicada cada vez mais em análise de demanda por transporte.

Existe uma grande variedade de métodos de mineração de dados que são utilizados. Entre as técnicas mais utilizadas estão às técnicas de inteligência artificial (do inglês *Artificial Intelligence - AI*) com o emprego das redes neurais, árvores de decisão, algoritmos genéticos, vizinho mais próximo (do inglês *nearest neighbor*), lógica nebulosa ou difusa (*fuzzy logic*), regras de indução e visualização (WESTPHAL; BLAXTON, 1998).

Em particular, a técnica de árvore de decisão se constitui em uma forma simples de representação, classificando exemplos de uma base de dados em um número finito de classes (QUINLAN, 1983).

A árvore é estruturada em uma sequência de questões simples, e as respostas para essas questões traçam o caminho da árvore. O produto final é um conjunto de regras hierárquicas que dividem os dados dentro de grupos onde uma divisão (classificação ou predição) é feita para cada grupo, formando uma árvore, com um nó raiz (conjunto original que contém os dados), nós intermediários (que representam grupos de dados) e o nós finais, também chamado de folhas (quando os dados do nó não podem ser mais subdivididos em um outro subconjunto), conforme exemplo apresentado na Figura 4.1.

Figura 4.1 - Exemplo de divisão pela árvore de decisão na escolha para compra de um carro

4.1.2 S-Plus

Entre os programas computacionais que disponibilizam o processo analítico de mineração de dados, será utilizado o S-Plus (versão 6.1). As principais razões que levam a adoção deste programa são:

- Este programa contém as ferramentas necessárias para aplicação da mineração de dados pela técnica de árvore de decisão;
- A aplicação do S-Plus obteve êxito nos trabalhos de análise de demanda por transportes de Ichikawa (2002), Pitombo (2003), Souza (2004) e Aguiar (2005). Esses autores relacionaram em seus trabalhos aspectos de encadeamento de viagens às características socioeconômicas dos viajantes urbanos.

A técnica de árvore de decisão utilizada pelo S-Plus 6.1 é baseada no algoritmo CART (do inglês, *Classification and Regression Trees*), considerado por Bell (1996) como um modelo de regressão não paramétrico que têm por objetivo estabelecer uma relação entre um vetor de variáveis preditoras e uma única variável resposta. O algoritmo é ajustado mediante sucessivas divisões binárias no conjunto de dados, de modo a tornar os subconjuntos resultantes cada vez mais homogêneos, em relação à variável resposta.

Ao escolher e apresentar as variáveis em ordem de importância, as árvores de decisão permitem aos usuários visualizar imediatamente quais fatores mais influenciam as suas classificações ou previsões. Oferecem assim uma estrutura clara e maior facilidade de compreensão dos resultados.

Um dos principais motivos que levaram à escolha da árvore de classificação e regressão é a sua capacidade de representar a natureza probabilística do objeto analisado, que no caso do presente trabalho representará a probabilidade de ocorrência dos padrões

de viagens encadeadas por grupos de indivíduos classificados segundo suas características particulares.

Na CART, as variáveis explicativas e as variáveis respostas podem assumir valores contínuos ou categóricos. O modelo recebe o nome de árvore de regressão para as respostas contínuas e árvore de classificação para variáveis categóricas.

O S-Plus possui restrições quanto ao processamento de dados. Há um limite máximo de 128 tipos de variáveis respostas e 32 categorias em relação ao nível das variáveis explicativas discretas.

Quanto à seleção do tipo de árvore, este é automaticamente selecionado pelo programa quando se define a variável resposta, que assume uma distribuição multinomial. A divisão dos dados dentro de cada nó é progressiva, tendendo à homogeneidade até que os nós contenham poucas observações. O critério quanto ao mínimo número de observações antes da divisão e no nó final é definido pelo usuário, sendo o limite mínimo estabelecido de 5 observações.

Nas árvores de classificação, a heterogeneidade de cada nó é medida pelo desvio (D) da seguinte forma (equação 4.1):

$$D = -2 \times \sum_t \sum_{j \in t} n(j|t) \log p(j|t) \quad (4.1)$$

Em que: $n(j|t)$ é o número de casos no nó t com a resposta categórica j ;

$P(j|t)$ denota a proporção dos casos na categoria j em relação a todos os casos pertencentes ao nó t .

Para o S-Plus 6.1 o desvio foi definido como recíproco da função verossimilhança elevada ao quadrado, isto significa que o desvio é uma medida da heterogeneidade dos grupos resultantes. Quanto menor o valor do desvio (D), maior é a complexidade da árvore, ocorrendo ramificações e um maior número de nós terminais.

4.2 Etapas do Experimento

4.2.1 Limpeza dos dados

A primeira etapa consiste na retirada dos dados inconsistentes ou incompletos, eliminando da amostra os indivíduos que realizaram uma ou mais de quatro viagens e que não tiveram a primeira e a última viagem do dia como motivo a residência. A opção de limitar o máximo de viagens em quatro deve-se à limitação do S-Plus (128 variáveis resposta). A inclusão de indivíduos que tenham realizado mais de quatro viagens aumentaria o número de padrões, que além de considerar o motivo da viagem, contabiliza o modo (3 classes) e o período em que foi realizada a viagem (5 classes).

4.2.2 Codificação dos dados

Serão codificados os dados relacionados à viagem: motivo da viagem, modo de transporte e período do dia em que foi realizada a viagem, adotando o sistema apresentado na Tabela 4.1.

Tabela 4.1 - Agrupamento proposto para representação dos padrões de viagem

<i>Motivo de viagem</i>	H - Residência W - Trabalho S - Escola A - Outras Atividades				
<i>Modo de transporte</i>	P - Particular T - Coletivo N - Não motorizado				
<i>Período do dia</i>	1 0:01 - 8:00	2 8:01 - 11:30	3 11:31 - 14:30	4 14:31 - 17:30	5 17:31 - 24:00

A combinação dos códigos da Tabela 4.1 forma o padrão de viagem encadeada, ou simplesmente padrão. Assim os padrões são representados por uma sequência de letras (H, W, S, A) que indicam as atividades realizadas pelos indivíduos ao longo do dia e a ordem cronológica em que as atividades são desenvolvidas (exemplo HWHWH, casa - trabalho - casa - trabalho). Analogamente, as letras (P, T e N) e os números (1, 2, 3, 4 e 5) representam respectivamente o modo de transporte e o período do dia em que é realizada a viagem.

4.2.3 Seleção da variável dependente

A variável dependente escolhida para simbolizar o padrão do banco de dados é o padrão de viagem encadeada, onde são considerados o motivo da viagem, modo de transporte utilizado e período em que a viagem foi realizada.

4.2.4 Seleção das variáveis independentes

As variáveis independentes, ou variáveis preditoras, escolhidas para explicar o comportamento do viajante urbano são relacionadas às características socioeconômicas.

Esta escolha é baseada em estudos das escolas de ciências humanas e dos trabalhos de análise de demanda por transporte com abordagem em atividades, que mostram que o comportamento humano é influenciado diretamente pelas suas características sociais (idade, sexo, situação familiar etc) e características econômicas (posse de automóvel, renda individual, renda familiar etc).

Para viabilizar um modelo de previsão é necessário que as variáveis dos bancos de dados que serão utilizados sejam compatibilizadas (as variáveis categóricas devem pos-

suir o mesmo número de classes e as variáveis contínuas devem possuir o mesmo sistema de medida).

4.2.5 Construção do cenário para projeção da demanda

Deve ser ressaltado que a proposta deste trabalho é submeter a hipótese de que o minerador de dados pode ser utilizado como uma ferramenta para previsão dos padrões de viagens a um teste de verificação para refutá-la ou confirmá-la. Não constitui intenção neste trabalho avaliar a construção do cenário futuro e sim testar a aplicabilidade da mineração de dados como ferramenta de auxílio à previsão de demanda das viagens urbanas. Para que a utilidade do modelo seja avaliada é necessário que o cenário projetado retrate com fidelidade as mudanças que irão ocorrer.

São utilizados nesta pesquisa os dados extraídos de cenários que retratam situações reais sobre o escopo do objeto em análise de dois períodos distintos. O primeiro cenário tem como base de informações o ano T e o segundo cenário o ano $T+\Delta T$, em que ΔT representa um intervalo de tempo, distinguindo os cenários quanto ao período.

Como mencionado anteriormente, é necessário que o cenário que será construído para projeção da demanda com base no ano T , contemple as mudanças ocorridas no intervalo de tempo que separa o ano do modelo elaborado para previsão da demanda para o ano horizonte de projeção, ou seja, no intervalo de tempo ΔT .

As mudanças que serão incorporadas ao cenário do ano T são provenientes da análise das informações provenientes da comparação entre os cenários, como mostra a Figura 4.2.

Figura 4.2 - Esquema da construção de cenário que será adotada no trabalho

Como neste trabalho apenas os atributos socioeconômicos dos indivíduos estão sendo considerados como variáveis preditoras da demanda baseada na teoria das atividades, só serão levadas em conta nas mudanças ocorridas as alterações relacionadas a estas variáveis, ou seja, as mudanças comportamentais. Não são consideradas na construção do cenário no ano $T+\Delta T$ as variáveis que descrevem o sistema de transporte (variáveis relacionadas às linhas de transporte etc.) e uso e ocupação do solo (variáveis que caracterizem áreas densamente povoadas, áreas comerciais etc.).

Na forma de avaliação das mudanças comportamentais é adotada a abordagem descrita por Nagin e Tremblay (2005), descrita no capítulo 3, que assume que agrupamentos de indivíduos segundo suas características socioeconômicas possuem diferentes trajetórias comportamentais.

O processo da verificação da mudança do comportamento de indivíduos consiste na percepção das alterações dos contextos dos cenários no ano T e $T+\Delta T$, ocasionada pela mudança de comportamento de grupos de indivíduos e seus reflexos nos padrões de atividades desempenhados. Portanto o enfoque dado à verificação da mudança comportamental em indivíduos está relacionado à atividade.

A Figura 4.3 ilustra o exemplo de um grupo de indivíduos com características particulares que optam pelo desempenho de certo tipo de comportamento em relação a uma atividade específica e devido a algum fator externo (momento político, situação econômica, política social etc.) ou interno (mudança na estrutura familiar) parte desse grupo passa a desempenhar um outro comportamento que modifica as decisões em relação às viagens.

Figura 4.3 - Exemplo de como a mudança de comportamento pode influenciar na previsão

Em particular, as mudanças de comportamento que se destacam na análise das informações estão relacionadas aos indivíduos que só exerciam atividade no domicílio e passam a exercer atividade extra-domiciliar e aos indivíduos que exerciam atividade extra-domiciliar e passam a exercer atividades no domicílio.

Para o caso de indivíduos com características particulares semelhantes (idade, sexo, situação familiar, trabalho, estudo etc.), que em duas épocas distintas deixam a residência e passam a desempenhar atividades extra-domiciliares, isso pode ser resultado da inserção desse grupo no mercado de trabalho (passam a desempenhar padrões contendo a atividade trabalho), do aumento da escolaridade desse grupo (passam a desempenhar padrões contendo a atividade estudo), aumento do poder aquisitivo desse grupo (engajamento em diversas atividades) etc.

No caso em que indivíduos de características particulares semelhantes (idade, sexo, situação familiar, trabalho, estudo etc.), que em duas épocas distintas deixam de exercer atividades extra-domiciliares, passado a exercer atividades no domicílio, isso pode ser decorrente de fatores de ordem econômica (desemprego, redução da renda familiar etc) e fatores de ordem social (ex. alterações nos papéis familiares).

Alguns estudos, a exemplo de pesquisas na área de previsão demográfica e econômica, contemplam os efeitos de mudanças na hierarquia social, política e econômica e seus efeitos no comportamento do indivíduo.

Kozu et al. (2003) analisaram os efeitos de mudanças demográficas na macroeconomia do Japão, em um trabalho realizado para o *Bank of Japan* e concluíram que o aumento da expectativa de vida da população e a baixa taxa de natalidade terão as seguintes

consequências: dificuldade em manter o atual sistema previdenciário; declínio no número de trabalhadores, e a diminuição da poupança interna do país.

Em um outro estudo realizado pelo *Deutsche Bank* (Deutsche Bank Research, 2002), foi concluído que mudanças demográficas, particularmente o envelhecimento progressivo da população alemã, implicarão aumento do consumo interno, diminuição da poupança interna e dificuldade em manter o sistema previdenciário atual, ocasionando um impacto negativo na macroeconomia alemã.

4.2.6 Adaptação dos dados à mudança de comportamento

Pelas referências consultadas não foi encontrado um método pré-estabelecido para este tipo de ajuste em casos que fazem uso do minerador de dados pela árvore de decisão para previsão.

Nos casos de modelos paramétricos, o ajuste para o caso de uma previsão poderia ser realizado mediante ajustes dos coeficientes, mas como o modelo de árvore de decisão é considerado um modelo não paramétrico, não permite a utilização de pesos na composição do modelo.

No modelo paramétrico de regressão linear simples, a partir de uma tabela de dados onde são apresentados a relação entre valores de uma variável independente e as respostas da variável dependente, é obtido uma função em que podem ser representados, através de uma função linear, a relação entre a variável dependente e a variável resposta. Graficamente esta relação pode ser representada por uma nuvem de pontos e uma reta, Figura 4.4.

Figura 4.4 - Ilustração de um modelo de regressão linear

Uma forma encontrada para ajustar modelo de árvore de decisão as mudanças, é simulando, através da inserção de dados, o efeito que é produzido no modelo paramétrico quando são ajustados os coeficientes, como apresentado na Figura 4.5. Os dados inseridos para ajuste do modelo recebem neste trabalho a denominação de “dados sintéticos”.

Figura 4.5 - Ilustração de um modelo de regressão linear ajustado com a inserção de novos dados

No caso do banco de dados, que contém as informações do cenário do ano T e que será utilizado para construir o cenário em que se baseia o modelo de projeção para o ano horizonte $T+\Delta T$, as alterações de comportamento dos indivíduos identificadas no perí-

odo ΔT , serão contempladas através de inserções de indivíduos sintéticos, os chamados clones, gerando um banco de dados sintético.

O indivíduo sintético é um indivíduo selecionado do banco de dados do cenário atual que representa as características socioeconômicas do indivíduo com a mudança de comportamento no cenário futuro. A quantidade e as características dos indivíduos sintéticos são determinadas pelo tipo de mudança e o grupo de ocorrência da mudança. O ajuste pela inserção de indivíduos sintéticos produzirá o efeito dos coeficientes em um modelo paramétrico, Figura 4.6.

Figura 4.6 - Esquema da introdução do indivíduo sintético no banco de dados

Desta forma, espera-se induzir no modelo de árvore de decisão o efeito da mudança comportamental com o desdobramento de novas folhas que retratem os grupos de indivíduos afetados pela mudança, como pode ser apresentado pela Figura 4.7.

Figura 4.7 - Desdobramento da folha devido à introdução de indivíduos sintéticos no banco de dados

A quantidade de indivíduos a ser replicada pelo processo de clonagem e introduzida no banco de dados ou quantidade de indivíduos a serem retiradas é dado pela equação 4.2.

$$(\text{Indivíduos sem mudança}) + \alpha \times (\text{Indivíduos com mudança}) = \beta(\text{Total de indivíduos}) \quad (4.2)$$

Em que:

Indivíduo sem mudança é a quantidade de indivíduos sem mudança comportamental;

α é o coeficiente de ajuste da quantidade de indivíduos com mudança;

β é o coeficiente de ajuste do total de indivíduos;

Total de indivíduos é o total de indivíduos do banco de dados.

4.2.7 Mineração dos dados: construção do modelo de previsão

A fase de construção do modelo de previsão, utilizando as informações do banco de dados (ajustado em relação à mudança comportamental dos indivíduos), é realizada

com a entrada dos dados no S-Plus 6.1 e seu processamento pela árvore de classificação. Selecionam-se os campos correspondentes às variáveis dependentes e independentes, além dos critérios adotados para classificação da árvore: desvio mínimo, número mínimo de observações antes da divisão do nó e número mínimo de observações no nó.

O resultado final é a produção de um relatório onde constam todas as classes de indivíduos, com suas características particulares, apresentadas através da árvore, seguidas de todos os padrões contidos na análise e a frequência de ocorrência de cada um deles.

4.2.8 Previsão dos padrões

Para realizar a previsão dos padrões de viagens a partir do modelo construído, utiliza-se o S-Plus 6.1. A opção de previsão do modelo de árvore do S-Plus associa cada indivíduo de um novo banco de dados a uma das classes de indivíduos agrupados pelo modelo de árvore do banco de dados anterior (que foi utilizado para gerar o modelo) e associa a probabilidade de ocorrência de cada um dos 128 padrões dessa classe a esse novo indivíduo.

O resultado final é um relatório onde são apresentados todos os indivíduos do banco de dados em que se deseja realizar a previsão e a probabilidade de ocorrência de cada um dos 128 padrões de viagem.

4.2.9 Levantamento dos padrões de viagens encadeadas dos indivíduos

Os valores observados da quantidade de indivíduos por padrões de viagens encadeadas são obtidos a partir do segundo banco de dados, que retrata as condições reais

para o ano horizonte em que foi realizada a projeção, $T+\Delta T$. Estes valores serão comparados aos valores estimados pelo modelo de previsão.

4.2.10 Comparação dos resultados

Realizada a previsão e de posse dos valores estimados e observados, cria-se a necessidade de se utilizar um método para verificar se os valores obtidos com o modelo condizem com os valores observados.

Seguindo a disposição geográfica dos dados que são agrupados por região metropolitana e por zona de tráfego, a comparação poderá ser efetuada de duas formas: uma em nível global (pela região metropolitana) e a outra ao nível da zona de tráfego.

Na comparação global são somados os valores estimados por indivíduos para cada um dos 128 padrões de viagens encadeadas (probabilidades de ocorrência de cada padrão), resultando ao final os valores estimados de indivíduos para cada um dos 128 valores de padrões de viagens encadeadas (equação 4.3). O resultado obtido é então comparado ao resultado obtido pelo mesmo procedimento do levantamento do banco de dados para o ano em que se realizou a previsão.

$$\text{Global } X_i = \sum_{j=1}^n P_{X_i}^j \quad (4.3)$$

Em que:

Global X_i é o valor estimado do padrão de viagem encadeada X_i , com i variando de 1 a 128;

$P_{X_i}^j$ é a probabilidade de ocorrência do padrão X_i do indivíduo j do banco de dados.

Já para o resultado por zona de tráfego, as probabilidades de ocorrências dos padrões dos indivíduos são somadas para cada uma das zonas de tráfego resultando ao final em 128 valores de padrões de viagens encadeada por zona, equação 4.4.

$$\text{Zona}_k X_i = \sum_{j=1}^n P_{X_i}^j \quad (4.4)$$

Em que:

$\text{Zona}_k X_i$ é o valor estimado do padrão de viagem encadeada X_i , com i variando de 1 a 128 para a zona de tráfego k ;

$P_{X_i}^j$ a probabilidade de ocorrência do padrão X_i do indivíduo j do banco de dados.

A escala de comparação de dados adotada neste trabalho é em nível da zona de tráfego. A verificação em nível da zona de tráfego permite analisar o desempenho do modelo de previsão em cada zona e explorar as possíveis falhas que podem ocorrer, e que poderiam ser encobertas em uma análise global.

Uma vez definida a escala de comparação, o passo seguinte é a seleção do método estatístico para efetuar a verificação dos resultados.

A primeira consideração a ser feita em relação ao teste é que a prova estatística a ser considerada não faça nenhuma exigência em relação à distribuição dos dados. O principal fundamento dessa consideração é que os dados sofreram processos aleatórios de tratamento (exclusão de dados na etapa da limpeza) sem a preocupação com a forma de distribuição da amostra. Este processo foi imposto por uma condição da ferramenta de análise que foi aplicada, além da disposição de informações incompletas nos bancos de dados.

Desta forma, ficam eliminadas todas as provas estatísticas paramétricas, que fazem exigência quanto à distribuição dos dados, limitando o método a ser empregado entre as provas da estatística não-paramétrica.

A idéia preliminar sobre a prova estatística a ser utilizada é de que ela proporcione uma avaliação rigorosa da diferença entre cada resultado estimado e observado, e que essas diferenças sejam avaliadas para um dado nível de significância.

Particularmente, entre os testes estatísticos não-paramétricos está o teste Kolmogorov-Smirnov (K-S). Este teste verifica se duas amostras foram extraídas da mesma população (ou de populações com mesma distribuição).

Um outro teste que poderia ser utilizado seria o Qui-quadrado, mas a prova de K-S segundo, Siegel (1977) é mais rigorosa que a prova Qui-quadrado, que depende do tamanho da amostra para que as aproximações possam ser validadas.

Para aplicar a prova de K-S para duas amostras em um teste bilateral, constrói-se uma distribuição de frequência acumulada para cada uma das observações, utilizando o mesmo intervalo para ambas as distribuições. A prova focaliza o maior desses desvios observados (D), que é obtido através da equação 4.5.

$$D = \max | S_{n1}(X) - S_{n2}(X) | \quad (4.5)$$

Em que:

$S_{n1}(X)$ é a função acumulada observada para uma das amostras;

$S_{n2}(X)$ é a função acumulada observada para a outra amostra.

O maior valor desse desvio observado (D) é comparado a um valor crítico ($D_{\text{crítico}}$) que, para grandes amostras (N_1 e $N_2 > 40$) com nível de significância de 5%, é calculado pela equação 4.6.

$$D_{\text{crítico}} = 1,36 \times \sqrt{\frac{N_1 + N_2}{N_1 \times N_2}} \quad (4.6)$$

Em que:

N_1 é o total de indivíduos da primeira amostra;

N_2 é o total de indivíduos da segunda amostra.

Para a prova de K-S formula-se a hipótese nula (H_0) a ser testada pela prova. Caso $D > D_{\text{crítico}}$, rejeita-se a hipótese nula. Se $D < D_{\text{crítico}}$, então não se rejeita a hipótese nula.

A prova de K-S ao analisar a diferença da frequência acumulativa das amostras atende a idéia inicial do método pretendido para avaliar os valores estimados pela previsão e os observados.

Considerando que não se pode fazer uso de testes paramétricos e entre os testes não-paramétricos este ser um dos mais consistentes, a prova de K-S será adotada como o método estatístico para comparação dos resultados.

A hipótese nula formulada para a prova de K-S no trabalho é de que tanto a amostra contendo dados reais quanto a amostra contendo os dados da previsão tenham a mesma distribuição de frequência de ocorrência dos padrões de viagens encadeadas. Caso $D > D_{\text{crítico}}$, a hipótese será rejeitada. Se $D < D_{\text{crítico}}$, então não se rejeita a hipótese nula e assume-se que as frequências de ocorrência dos padrões de viagens encadeadas da previsão e do real podem ser consideradas semelhantes.

Um ponto importante antes da execução do teste de K-S é a ordem de disposição dos padrões. A prova de K-S é sensível quanto à sequência dos padrões, dependendo do arranjo dos padrões pode ocorrer a conjunância de padrões que aumente o desvio ou diminua o valor do desvio.

Neste trabalho a ordem de disposição inicial adotada será a sequência gerada pelo relatório do S-Plus (ordem alfabética). Adicionalmente serão realizados cinquenta testes sendo os padrões dispostos em ordens aleatórias, com o intuito de elaborar medidas estatísticas descritivas sobre os resultados gerados em relação ao número de zonas onde a hipótese é confirmada ou refutada.

4.3 Considerações Adicionais

4.3.1 Sobre o método proposto para construção do cenário futuro

Van de Bilt (2002) propôs um conjunto de procedimentos para projeção de uma população de forma desagregada, e com a população obtida estimar dos totais de viagens através de um modelo de geração de viagens baseado em análise de segmentação.

Para projeção da população Van de Bilt (2002) aplicou o método de ajuste proporcional iterativo - IPF (*Iterative Proportional Fitting*) aos dados do Censo de 1980 do IBGE da Região Metropolitana de São Paulo (RMSP) para projeção desta população para o ano de 1991.

A população projetada para 1991 foi comparada à população observada pelo Censo de 1991 (IBGE) através da distribuição de família segundo um conjunto de variáveis

socioeconômicas. Após a comparação, o IPF demonstrou ser um método eficaz para aplicação da projeção de populações.

Para a etapa de geração de viagens, novamente foi aplicado o IPF aos dados do Censo de 1980 para projeção da população para o ano de 1987. Aliado a um modelo de escolha discreta para estimar o nível de motorização das famílias projetadas foram estipulados, através de um modelo de desagregado, a quantidades de viagens para as famílias da RMSP e em um segundo momento para cada um dos municípios que compõem a RMSP.

Os valores estimados foram comparados aos valores observados, pela Pesquisa Origem-Destino da RMSP para o ano de 1987, demonstrando a aplicabilidade dos procedimentos.

Os principais pontos de divergência entre o método adotado para construção do cenário futuro neste trabalho e o método proposto por Van de Bilt (2002) são:

- Há um interesse em Van de Bilt (2002) na evolução dos indivíduos na projeção da população. Neste trabalho o intuito é a construção de cenários futuros que retratem alterações comportamentais individuais;
- No trabalho de Van de Bilt (2002) não é apresentado a mudança comportamental em grupos de indivíduos, que é tema central neste trabalho para construção do cenário futuro;
- Em Van de Bilt (2002) foi utilizada a população projetada para estimar os totais de viagens. Neste trabalho a proposta é utilizar a amostra sintética para calibrar um modelo que deverá ser aplicado na população em que se deseja estimar os padrões de viagens encadeadas;
- A unidade de projeção utilizada em Van de Bilt (2002) é a família, enquanto neste trabalho a unidade de análise é o indivíduo;

- A comparação dos resultados apresentados em Van de Bilt (2002) em relação às estimativas de viagens é dada em função do total de viagens. Neste trabalho a postura adotada é a análise baseada em grupos com características socioeconômicas homogêneas, que é mais pertinente aos fundamentos da teoria de viagens baseada em atividades, com resultados em padrões de viagens encadeadas;

4.3.2 Sobre os dados

Sobre os dados a serem utilizados no experimento, algumas considerações são necessárias:

- Escolha das amostras: o processo de amostragem dos elementos que constituem as duas fontes primárias de dados deste trabalho (anterior às etapas do experimento) deve ser considerado idêntico em ambos;
- Expansão das amostras: os dados das amostras não devem ser expandidos para o conjunto da população pela incerteza quanto à proporcionalidade de atributos individuais em relação ao total da população. Esta incerteza é creditada à dificuldade de obtenção das informações em parte dos domicílios selecionados para entrevista.
- Representação proporcional do padrão H na amostra: em função dos atributos considerados no padrão de encadeamento de viagem (atividade, modo e período da viagem), o que aumenta a diversidade dos padrões, na etapa de limpeza dos dados são retirados da amostra os indivíduos que realizaram mais de quatro viagens e selecionados os 128 padrões mais frequentes. Este procedimento aumenta proporcionalmente a representação dos indivíduos que realizam o padrão H. Este método foi adotado nos trabalhos de Ichikawa (2002), Pitombo (2003), Souza (2004) e Aguiar (2005).

Capítulo 5

Tratamento dos Dados

Este capítulo aborda a descrição e as etapas de tratamento dos dados envolvidos no estudo, as Pesquisas Origem-Destino de 1987 e 1997 realizadas pela Companhia do Metropolitano de São Paulo (METRÔ) na Região Metropolitana de São Paulo (RMSP).

5.1 Região Metropolitana de São Paulo: Caracterização Urbana

A Região Metropolitana de São Paulo (Figura 5.1), com 8.051 km², é formada por 39 municípios, sendo o de São Paulo, com 1.509 km², o seu centro. Nessa região reside uma população de 17,9 milhões de habitantes e três características se destacam: 95,75% da população reside em áreas urbanas, 10,4 milhões estão concentrados no município de São Paulo e apenas 8 dos 39 municípios metropolitanos não formam uma mancha urbana contínua com os demais municípios (SÃO PAULO, 2003).

Os dados da Tabela 5.1 e do gráfico da Figura 5.2 apontam que a participação da população do Município de São Paulo na RMSP vem reduzindo-se significativamente, com crescimento expressivo das periferias metropolitanas.

Os municípios que compõem a região metropolitana são os seguintes: ao centro, São Paulo; a oeste Osasco, Carapicuíba, Barueri, Jandira, Santana do Parnaíba, Itapevi, Pirapora do Bom Jesus, Cotia, Vargem Grande Paulista; a sudoeste, Taboão da Serra, Embu, Itapeçerica da Serra, São Lourenço da Serra, Embu-Guaçu e Juquitiba; a sudeste São Caetano do Sul, Diadema, Santo André, São Bernardo do Campo, Mauá, Ribeirão Pires e Rio Grande da Serra; a leste, Ferraz de Vasconcelos, Poá, Itaquaquecetuba, Suzano, Mogi das Cruzes, Guararema, Biritiba Mirim, e Salesópolis; a nordeste, Guarulhos, Arujá e Santa Isabel; ao norte, Cajamar, Caieiras, Franco da Rocha, Mairiporã, e Francisco Morato.

No decorrer da década de 80, o ritmo de crescimento demográfico na região metropolitana em média (1,89% a.a. para os anos da década de 80) se reduziu à metade dos valores verificados na década anterior (70). Entretanto, é relevante observar que esse processo de desaceleração não se deu de forma uniforme, e enquanto o ritmo de crescimento da capital se reduzia significativamente, alcançando 1,18% ao ano na década de 80 e 0,88% ao ano no decorrer da década de 90, nos demais municípios metropolitanos sucedia o inverso, ocorrendo um expressivo crescimento populacional com valores médios de 2% e 9% a.a. nos anos 90 (SÃO PAULO, 2003).

Esse movimento populacional revela uma situação paradoxal. Por um lado o crescimento expressivo das periferias metropolitanas com dispersão da população no território metropolitano. Os últimos dados censitários confirmam esse movimento: 53 dos 96 distritos do município de São Paulo estão perdendo moradores. Como parte do mesmo processo, verificou-se o acréscimo significativo de moradores nos 43 distritos periféricos restantes, que são em grande parte regiões carentes de infra-estrutura e serviços públicos.

5.2 Dados: Pesquisa Origem-Destino da Região Metropolitana de São Paulo

O presente estudo baseia-se nos dados referentes às Pesquisas Origem-Destino (O/D) da Região Metropolitana de São Paulo (RMSP), realizadas em 1987 e 1997 pela Companhia do Metropolitano de São Paulo (METRÔ). Ambas as pesquisas foram feitas por meio de entrevistas domiciliares e obtiveram informações detalhadas acerca das características socioeconômicas e dos deslocamentos da população no dia anterior ao da entrevista.

5.2.1 Pesquisa Origem-Destino da Região Metropolitana de São Paulo de 1987

Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), em 1987 a RMSP contava com uma população de 14,2 milhões de habitantes, correspondentes aos 38 municípios, divididos em 254 zonas identificadas pela homogeneidade no uso do solo.

O questionário informativo para a coleta dos dados da pesquisa domiciliar para a RMSP foi dividido em três tipos. O primeiro refere-se a informações relativas ao domicílio (tipo do domicílio, condição de moradia e valor do aluguel), às informações relativas à família (tempo de residência no bairro, tamanho e itens de conforto) e às informações relativas às características socioeconômicas de cada pessoa da família (posição na família, idade, sexo, nível de instrução, ocupação profissional e classe de atividade da empresa em que trabalha e faixa de renda mensal). O segundo tipo corresponde a informações sobre o local de trabalho e/ou estudo. O terceiro tipo de questionário corresponde a informações relativas às viagens realizadas no dia anterior ao da entrevista pelas pessoas do domicílio (origem, destino, hora de início e término da viagem, motivo, modo e duração), METRÔ (1997).

O período da coleta compreendeu os meses de outubro, novembro e dezembro de 1987, além dos meses de março, abril, maio e junho de 1988, com maior número de entrevistas domiciliares no ano de 1987. Os dados da pesquisa correspondem às informações sobre 110.629 indivíduos entrevistados em 26.056 domicílios.

5.2.2 Pesquisa Origem-Destino da Região Metropolitana de São Paulo de 1997

Segundo o IBGE, em 1997 a RMSP contava com uma população de aproximadamente 16,8 milhões de habitantes, distribuída em 39 municípios. A Pesquisa Origem-Destino na RMSP foi realizada entre os meses de fevereiro e setembro de 1997.

O questionário para a coleta dos dados da pesquisa domiciliar para a RMSP de 1997 também foi dividido em três tipos conforme o questionário da pesquisa domiciliar para RMSP de 1987, METRÔ (1997). Os dados da pesquisa de 1997 correspondem às informações sobre 98.780 indivíduos entrevistados em 26.278 domicílios.

5.3 Etapas do Tratamento dos Dados

5.3.1 Limpeza dos dados

Com a aplicação dos quatro tipos de filtragens, retirada dos registros com dados incompletos, eliminação dos indivíduos que não tiveram a origem ou destino final da viagem a residência, remoção dos indivíduos que realizaram uma ou mais de quatro viagens e o expurgo dos indivíduos que não desempenharam um dentre os 128 padrões de viagens encadeadas mais frequentes em cada um dos bancos de dados), o total resultante dos registros na amostra de 1987 são 80.764 indivíduos e 79.839 indivíduos para 1997, como pode ser apresentado pela Tabela 5.2.

Tabela 5.2 - Tratamento das amostras das Pesquisas O/D

Descrição	Pesquisa O/D 1987	Pesquisa O/D 1997
<i>Amostra inicial</i>	110.629 (100%)	98.780 (100%)
<i>Dados incompletos</i>	9.218 (8,33%)	417 (0,42%)
<i>Dados completos</i>	101.411 (91,67%)	98.363 (99,58%)
<i>Indivíduos eliminados devido à realização de uma ou mais de quatro viagens</i>	5.040 (4,56%)	6.159 (6,24%)
<i>Indivíduos eliminados devido ao fato da primeira e a última viagem não terem como motivo a residência</i>	2.156 (1,95%)	1.496 (1,51%)
<i>Indivíduos que não desempenharam um dentre os 128 padrões de viagens encadeadas mais frequentes em cada um dos bancos de dados</i>	13.451 (12,16%)	10.869 (11,0%)
<i>Amostra final</i>	80.764 (73,00%)	79.839 (80,83%)
<i>Totais de Viagens</i>	99.690	92.939

Após a limpeza dos dados a representação da amostra em relação aos dados completos é de 79,64% em 1987 (80.764 de uma amostra de 101.411 registros completos) e 81,17% (79.839 de uma amostra de 98.363 registros completos). Ao ser considerado os 128 padrões mais frequentes a análise dos dados será baseada em aproximadamente 80% da amostra.

5.3.2 Codificação dos dados

Codificando ambos os bancos de dados pelo sistema da Tabela 4.1, são obtidas as seguintes frequências dos 128 padrões utilizados na análise, conforme apresentado na Tabela 5.3 e Tabela 5.4.

Tabela 5.3 - Frequência dos 128 padrões utilizados na análise para o ano de 1987

Padrão	Freq.	%	Padrão	Freq.	%	Padrão	Freq.	%	Padrão	Freq.	%
<i>H</i>	34028	42,13	<i>HAHNN45</i>	290	0,36	<i>HAHPP23</i>	127	0,16	<i>HWHPP13</i>	78	0,10
<i>HWHTT15</i>	7063	8,75	<i>HSHP45</i>	288	0,36	<i>HSHP14</i>	120	0,15	<i>HWHWHPPPP1335</i>	76	0,09
<i>HSNN12</i>	3058	3,79	<i>HWHPP14</i>	284	0,35	<i>HAHNN24</i>	119	0,15	<i>HWAHTTTP1555</i>	72	0,09
<i>HWHNN15</i>	2374	2,94	<i>HAHNN55</i>	268	0,33	<i>HWAHTTT155</i>	119	0,15	<i>HWHSHPPPP1555</i>	72	0,09
<i>HWHPP15</i>	2269	2,81	<i>HSHP12</i>	260	0,32	<i>HAHNN23</i>	117	0,14	<i>HWHWHPPPP2345</i>	72	0,09
<i>HSNN45</i>	2085	2,58	<i>HAHPP44</i>	247	0,31	<i>HSNN22</i>	115	0,14	<i>HWHTP25</i>	71	0,09
<i>HWHTT25</i>	2006	2,48	<i>HSHTT35</i>	244	0,30	<i>HAHPP25</i>	110	0,14	<i>HSNSHNNNN3355</i>	70	0,09
<i>HSNN35</i>	1768	2,19	<i>HSHTT55</i>	241	0,30	<i>HSHTT12</i>	110	0,14	<i>HAHNN25</i>	69	0,09
<i>HSNN24</i>	1755	2,17	<i>HAHTT25</i>	232	0,29	<i>HSNN44</i>	108	0,13	<i>HWHPP12</i>	69	0,09
<i>HSNN13</i>	1693	2,10	<i>HAHTT55</i>	223	0,28	<i>HWAHPP155</i>	108	0,13	<i>HWHPP24</i>	69	0,09
<i>HWHTT14</i>	1293	1,60	<i>HWSHTTT155</i>	209	0,26	<i>HAHTT35</i>	107	0,13	<i>HSHPN12</i>	68	0,08
<i>HWHPP25</i>	898	1,11	<i>HSNN15</i>	199	0,25	<i>HWHPT15</i>	106	0,13	<i>HAHTT14</i>	67	0,08
<i>HSHP35</i>	723	0,90	<i>HWHWHNNNN1235</i>	199	0,25	<i>HAHPP24</i>	105	0,13	<i>HWHNN13</i>	66	0,08
<i>HSHP13</i>	711	0,88	<i>HAHTT34</i>	192	0,24	<i>HAHNN34</i>	104	0,13	<i>HSNN23</i>	65	0,08
<i>HAHTT45</i>	708	0,88	<i>HSNN14</i>	182	0,23	<i>HWHPP35</i>	104	0,13	<i>HWHNN24</i>	65	0,08
<i>HSNN55</i>	612	0,76	<i>HSHTT14</i>	176	0,22	<i>HWHNN35</i>	103	0,13	<i>HAHPP12</i>	63	0,08
<i>HSNN34</i>	582	0,72	<i>HAHPP22</i>	167	0,21	<i>HWHTP15</i>	102	0,13	<i>HWSHPP155</i>	62	0,08
<i>HWHNN14</i>	541	0,67	<i>HSHTT45</i>	166	0,21	<i>HSHP24</i>	101	0,13	<i>HAHTT13</i>	59	0,07
<i>HWHTT45</i>	486	0,60	<i>HWHTT24</i>	165	0,20	<i>HWSHTTNN2555</i>	92	0,11	<i>HWAHTTNN1555</i>	59	0,07
<i>HAHPP55</i>	462	0,57	<i>HWHNN45</i>	164	0,20	<i>HAHTT15</i>	89	0,11	<i>HWSHNNTT1555</i>	57	0,07
<i>HAHTT24</i>	454	0,56	<i>HWHWHNNNN1345</i>	161	0,20	<i>HSNN25</i>	88	0,11	<i>HWHTT13</i>	57	0,07
<i>HSHTT13</i>	426	0,53	<i>HSHP15</i>	160	0,20	<i>HSNSHNNNN4455</i>	87	0,11	<i>HWSHTTTP1555</i>	56	0,07
<i>HAHPP45</i>	415	0,51	<i>HSNSHNNNN1122</i>	149	0,18	<i>HAHPP34</i>	86	0,11	<i>HSHP34</i>	55	0,07
<i>HAHTT44</i>	391	0,48	<i>HAHTT12</i>	147	0,18	<i>HSNN11</i>	85	0,11	<i>HAAHPP445</i>	53	0,07
<i>HAHTT23</i>	358	0,44	<i>HSHTT13</i>	145	0,18	<i>HSHPN13</i>	84	0,10	<i>HWSHTTTT1455</i>	53	0,07
<i>HAHNN44</i>	345	0,43	<i>HSHTT15</i>	143	0,18	<i>HAHNN12</i>	83	0,10	<i>HSHP25</i>	52	0,06
<i>HWSHTTNN1555</i>	343	0,42	<i>HWHPP45</i>	139	0,17	<i>HSHTT24</i>	83	0,10	<i>HWAHTTT145</i>	52	0,06
<i>HWHNN25</i>	339	0,42	<i>HWSHNNNN1555</i>	137	0,17	<i>HAHPP15</i>	82	0,10	<i>HWAHTTT255</i>	52	0,06
<i>HWHWHNNNN1335</i>	338	0,42	<i>HWAHPPPP1555</i>	136	0,17	<i>HSHP55</i>	82	0,10	<i>HAHNN15</i>	49	0,06
<i>HAHTT22</i>	333	0,41	<i>HWSHTTT255</i>	136	0,17	<i>HSNSHNNNN2244</i>	82	0,10	<i>HWAHPP255</i>	48	0,06
<i>HAHNN22</i>	328	0,41	<i>HWHWHPPPP1345</i>	135	0,17	<i>HSWHPP115</i>	82	0,10	<i>HWHWHNNNN2345</i>	48	0,06
<i>HWHTT35</i>	328	0,41	<i>HWSHTTTT1555</i>	128	0,16	<i>HSHTT14</i>	79	0,10	<i>HAAHPP555</i>	46	0,06

Codificação:

(H) - Residência (W) - Trabalho (S) - Escola (A) - Outras Atividades

(P) - Particular (T) - Coletivo (N) - Não motorizado

(1) 0:01 - 8:00 (2) 8:01 - 11:30 (3) 11:31 - 14:30 (4) 14:31 - 17:30 (5) 17:31 - 24:00

Tabela 5.4 - Frequência dos 128 padrões utilizados na análise para o ano de 1997.

Padrão	Quant.	%	Padrão	Quant.	%	Padrão	Quant.	%	Padrão	Quant.	%
H	35964	45,05	HAHPP44	303	0,38	HWHPP35	130	0,16	HWHTP25	75	0,09
HWHTT15	4924	6,17	HAHTT44	302	0,38	HAHPP25	128	0,16	HWHPP13	74	0,09
HWHPP15	2587	3,24	HAHNN45	285	0,36	HWSHTTT155	124	0,16	HWHAHPPPP1555	73	0,09
HSNN13	2410	3,02	HAHNN44	283	0,35	HSHTT12	116	0,15	HWHWHPPPP2345	73	0,09
HSNN35	2347	2,94	HAHNN22	270	0,34	HWHWHPPPP1345	115	0,14	HWAWHTNNT1335	72	0,09
HWHTT25	2248	2,82	HAHTT55	269	0,34	HAHTT15	113	0,14	HWHTP15	71	0,09
HWHNN15	1768	2,21	HSHTT14	252	0,32	HWHSNNNN1555	107	0,13	HAHNN12	70	0,09
HWHPP25	1411	1,77	HSHP45	251	0,31	HSHP24	105	0,13	HWHWHNNNN2345	70	0,09
HSNN12	1271	1,59	HAHTT25	235	0,29	HSHSNNNN4455	104	0,13	HWHWHPPPP1335	68	0,09
HSNN45	1252	1,57	HWHPP45	227	0,28	HAHNN34	98	0,12	HAHTT14	65	0,08
HSNN55	1127	1,41	HSHP14	226	0,28	HWHNN24	97	0,12	HWHSHTTTT1555	64	0,08
HWHTT14	1030	1,29	HWHTT24	206	0,26	HAHTT35	93	0,12	HAHNN25	61	0,08
HSHP13	939	1,18	HAHPP22	204	0,26	HSHSNNNN1122	93	0,12	HWHSHTTNN2555	61	0,08
HSNN24	916	1,15	HAHTT34	201	0,25	HWSHTTT255	92	0,12	HSHP25	60	0,08
HSHP35	752	0,94	HSHP15	191	0,24	HSNN22	91	0,11	HAHPP35	59	0,07
HAHTT45	723	0,91	HWHNN45	187	0,23	HAHPP34	89	0,11	HWHAHPP255	58	0,07
HWHTT45	624	0,78	HSHP55	184	0,23	HSHT14	89	0,11	HWHTT55	58	0,07
HWHNN25	578	0,72	HSNN34	182	0,23	HSHSNNNN3455	88	0,11	HWHWHNNNN1235	58	0,07
HAHPP55	544	0,68	HSNN14	180	0,23	HSWHPPP115	88	0,11	HSHP255	57	0,07
HSHTT13	513	0,64	HAHPP23	172	0,22	HSNN25	86	0,11	HSNN23	56	0,07
HAHTT24	484	0,61	HSHP12	169	0,21	HWHNN13	86	0,11	HAHTT13	55	0,07
HSHTT55	444	0,56	HWHPP24	168	0,21	HAHNN24	85	0,11	HWHSHPPP1555	55	0,07
HSHTT35	431	0,54	HSHTT15	160	0,20	HSNN11	83	0,10	HWHTTTT155	54	0,07
HWHNN14	431	0,54	HWHSHTTNN1555	159	0,20	HSHP155	82	0,10	HAAHPP445	53	0,07
HAHPP45	403	0,50	HAHPP24	158	0,20	HAHPP15	81	0,10	HWHAHPP155	53	0,07
HAHNN55	370	0,46	HAHTT12	158	0,20	HWHTP15	81	0,10	HWHPP55	52	0,07
HWHTT35	369	0,46	HSHTT45	152	0,19	HAHPP12	78	0,10	HSNN33	51	0,06
HAHTT23	354	0,44	HWHWHNNNN1345	150	0,19	HSHSNNNN3355	78	0,10	HWHTT13	51	0,06
HSNN15	323	0,40	HSHPN13	149	0,19	HSHTT24	77	0,10	HSHP34	49	0,06
HAHTT22	319	0,40	HAHNN23	137	0,17	HSHSNNNN2244	76	0,10	HSHP33	47	0,06
HWHPP14	319	0,40	HSHTT13	135	0,17	HSNN44	75	0,09	HSHTT25	46	0,06
HWHWHNNNN1335	309	0,39	HWHNN35	132	0,17	HSHSNNNN1133	75	0,09	HWHSHTTNN1455	46	0,06

Codificação:

(H) - Residência (W) - Trabalho (S) - Escola (A) - Outras Atividades

(P) - Particular (T) - Coletivo (N) - Não motorizado

(1) 0:01 - 8:00 (2) 8:01 - 11:30 (3) 11:31 - 14:30 (4) 14:31 - 17:30 (5) 17:31 - 24:00

5.3.3 Variáveis dependentes

As variáveis dependentes selecionadas para o processamento dos dados de 1987 são os padrões de viagem encadeada listados na Tabela 5.3.

5.3.4 Variáveis independentes

O banco de dado da Pesquisa O/D de 1987 e de 1997 detém características sociais e características econômicas dos indivíduos entrevistados. Dentre as características socioeconômicas levantadas na entrevista domiciliar foram selecionadas as seguintes variáveis para elaboração do modelo de previsão: automóveis, grau de instrução, idade, renda familiar, renda individual, sexo e situação familiar, conforme apresentado na Tabela 5.5 e Tabela 5.6:

Tabela 5.5 - Variáveis selecionadas da Pesquisa O/D de 1987 para o modelo de Previsão.

Variável	Descrição
<i>Auto</i>	Quantidade de automóveis na família: variável contínua.
<i>Grau de Instrução</i>	Grau de instrução do indivíduo entrevistado: variável categórica (1 - analfabeto / 1º grau até 4ª. série; 2 - 1º grau até 4ª. série concluído; 3 - 1º grau concluído; 4 - colegial concluído; 5 - superior concluído).
<i>Idade</i>	Idade do indivíduo entrevistado em anos: variável contínua (0 a 99).
<i>Renda Familiar</i>	Renda da família em Salários Mínimos: variável contínua.
<i>Renda Individual</i>	Renda do indivíduo em Salários Mínimos: variável contínua.
<i>Sexo</i>	Sexo do indivíduo entrevistado: variável categórica (1 - homem; 2 - mulher).
<i>Situação Familiar</i>	Situação do Indivíduo na família: variável categórica (1 - chefe; 2 - cônjuge; 3 - filho(a); 4 - parente; 5 - agregado; 6 - empregada/residente; 7 - visitante não residente).

Tabela 5.6 - Variáveis selecionadas da Pesquisa O/D de 1997 para o modelo de Previsão.

Variável	Descrição
<i>Auto</i>	Quantidade de automóveis na família: variável contínua.
<i>Grau de Instrução</i>	Grau de instrução do indivíduo entrevistado: variável categórica (1 - não alfabetizado; 2 - pré-escola; 3 - 1º grau incompleto; 4 - 1º grau completo; 5 - 2º grau incompleto; 6 - 2º grau completo; 7 - superior incompleto; 8 - superior completo).
<i>Idade</i>	Idade do indivíduo entrevistado em anos: variável contínua (0 a 99).
<i>Renda Familiar</i>	Renda da família em Reais (R\$): variável contínua.
<i>Renda Individual</i>	Renda do indivíduo em Reais (R\$): variável contínua.
<i>Sexo</i>	Sexo do indivíduo entrevistado: variável categórica (1 - homem; 2 - mulher).
<i>Situação Familiar</i>	Situação do Indivíduo na família: variável categórica (1 - chefe; 2 - cônjuge; 3 - filho(a); 4 - parente/agregado; 5 - empregado residente; 6 - visitante não residente na RMSP).

Conforme pode ser observado nas Tabelas 5.5 e 5.6, as categorias das variáveis categóricas grau de instrução e situação familiar não apresentam o mesmo número de classes, sendo necessário a sua compatibilização. Este processo foi realizado tomando-se por parâmetro os valores de 1987 e alterando as categorias dessas variáveis para 1997.

Com relação as variáveis contínuas, às variáveis renda individual e renda familiar precisaram ser ajustadas em ambos os bancos de dados. Na pesquisa realizada em 1987 as informações sobre renda familiar e renda individual foram coletadas em múltiplos de salários mínimos do mês corrente em que se realizou a entrevista (a moeda vigente era cruzado - CZ\$), e para 1997 as informações obtidas foram valores em Reais (R\$) do mês corrente em que a entrevista foi realizada, havendo portanto diferença em relação a moeda e variação monetária de 1987 a 1997.

O procedimento de ajuste dos valores de renda de 1987 e 1997 foi denominado neste trabalho de “processo de transferência monetária” e tem por mês base de ajuste o último mês em que a Pesquisa O/D foi realizada, setembro de 1997, mês considerado para a previsão dos padrões de viagens encadeadas. É assumido que as demais variáveis independentes não sofram variações ao longo do tempo.

Conforme mencionado no Capítulo 4, um índice de reajuste monetário utilizado são os Índices de Preços aos Consumidores - IPC. No Brasil os IPC mais usados são: o Índice de Preço ao Consumidor Amplo do Instituto Brasileiro de Geografia e Estatística (IPCA-IBGE), o Índice Nacional de Preços ao Consumidor do Instituto Brasileiro de Geografia e Estatística (INPC-IBGE) e o Índice de Preços ao Consumidor da Fundação do Instituto de Pesquisas Econômicas (IPC-FIPE).

O IPCA-IBGE é baseado no critério plutocrático, onde é atribuído a cada grupo de consumidores em uma sociedade um peso equivalente à participação relativa de seus gastos de consumo, relativamente aos gastos totais da sociedade. O INPC-IBGE ilustra o critério democrático, em que são atribuídos os mesmos pesos a cada grupo independentemente de sua afluência econômica. O IPCA - IBGE e o INPC-IBGE utilizam as fórmulas de Laspeyres, na versão adotada pelo *BLS - Bureau of Labor Statistics* (BLS) dos Estados Unidos da América desde 1926 e utilizam dados nacionais.

O IPC-FIPE é baseado na fórmula proposta por Konüs e Byushgens (CARMO, 2004), sendo calculado exclusivamente com dados colhidos no Município de São Paulo. Por se tratar de um indicador local do custo de vida do Município de São Paulo, o IPC-FIPE será adotado como índice de reajuste para o processo de transferência temporal monetária.

Para o cálculo dos valores monetários foram considerados os planos e medidas econômicas implantados no período em que será realizada a previsão da demanda: o Plano Verão I de 15/01/1989; o Plano Collor I de 15/03/1990; a transição do Cruzeiro para Cruzeiro Real em 28/07/1993 e o Plano Real em 27/05/1994.

Para o processo de transferência temporal monetária será adotada a linha plutocrática, sendo calculado um índice de reajuste específico para cada mês em que foi realizada a entrevista até o mês base (setembro de 1997) por faixa de renda.

O cálculo dos índices é procedido da seguinte forma: para cada faixa de renda familiar há um percentual médio mensal de despesa familiar por tipo de categoria de consumo (habitação, alimentação, transportes, despesas pessoais, saúde, vestuário e educação) conforme pode ser demonstrado pelas Pesquisas de Orçamento Familiar (POF) conduzidas pelo IBGE nos anos de 1987 e 1996 (na busca pelo Sistema IBGE de Recuperação Au-

tomática - SIDRA, não foi encontrado a POF para 1997, considerando então o ano de 1996 como referência para 1997), conforme apresentado na Tabela 5.7 e na Tabela 5.8.

Tabela 5.7 - Percentual da despesa média mensal familiar por tipo de despesa e classes de recebimento mensal familiar de 1987

Faixa de Renda em Salários Mínimos	Categoria de Despesa de Consumo						
	Habitação	Alimentação	Transportes	Despesas Pessoais	Saúde	Vestuário	Educação
<i>Até 2</i>	21,64%	44,10%	7,52%	9,66%	6,92%	8,59%	1,57%
<i>Mais de 2 a 3</i>	21,27%	41,28%	8,53%	10,75%	5,92%	10,41%	1,85%
<i>Mais de 3 a 5</i>	21,62%	36,96%	9,46%	12,20%	5,89%	11,81%	2,06%
<i>Mais de 5 a 6</i>	21,16%	37,10%	9,86%	12,26%	5,34%	11,95%	2,33%
<i>Mais de 6 a 8</i>	21,12%	33,43%	11,10%	12,83%	5,98%	12,95%	2,59%
<i>Mais de 8 a 10</i>	19,92%	31,10%	12,51%	12,72%	6,83%	13,86%	3,06%
<i>Mais de 10 a 15</i>	20,98%	28,71%	13,12%	13,45%	6,59%	14,05%	3,10%
<i>Mais de 15 a 20</i>	19,39%	25,03%	15,93%	14,70%	7,64%	13,66%	3,66%
<i>Mais de 20 a 30</i>	20,32%	21,70%	17,07%	15,56%	7,99%	12,55%	4,81%
<i>Mais de 30</i>	22,60%	17,12%	18,29%	16,89%	7,86%	12,91%	4,33%

Fonte: IBGE

Tabela 5.8 - Percentual da despesa média mensal familiar por tipo de despesa e classes de recebimento mensal familiar de 1996

Faixa de Renda em Salário Mínimos	Categoria de Despesa de Consumo						
	Habitação	Alimentação	Transportes	Despesas Pessoais	Saúde	Vestuário	Educação
<i>Até 2</i>	27,55%	35,23%	9,51%	10,25%	10,31%	5,66%	1,49%
<i>Mais de 2 a 3</i>	26,37%	37,12%	10,33%	10,30%	8,10%	6,09%	1,68%
<i>Mais de 3 a 5</i>	27,61%	33,65%	12,15%	10,45%	7,60%	6,54%	2,00%
<i>Mais de 5 a 6</i>	29,00%	30,50%	11,88%	11,49%	7,66%	6,83%	2,64%
<i>Mais de 6 a 8</i>	27,90%	29,57%	13,04%	12,12%	7,99%	6,88%	2,49%
<i>Mais de 8 a 10</i>	28,16%	27,27%	12,99%	12,86%	8,78%	7,13%	2,81%
<i>Mais de 10 a 15</i>	29,06%	24,21%	13,48%	13,11%	8,75%	7,30%	4,09%
<i>Mais de 15 a 20</i>	28,98%	21,15%	14,61%	13,19%	9,86%	6,91%	5,29%
<i>Mais de 20 a 30</i>	27,90%	22,72%	14,35%	13,21%	9,47%	5,96%	6,39%
<i>Mais de 30</i>	29,88%	17,21%	14,85%	15,28%	9,73%	6,43%	6,63%

Fonte: IBGE

Mediante pressão inflacionária no período de 1987 a 1997, essas categorias de despesas sofreram variações de preços. Como cada item do orçamento familiar responde de forma diferente em relação às mudanças na economia, as variações para os mesmos

ocorrem de forma heterogênea, conforme pode ser observado pelos valores das variações no período de 1987 a 1997 apresentados pela Tabela 5.9.

Tabela 5.9 - Variação percentual por categoria de despesa de consumo até o mês base (setembro/1997)

Mês de origem	Habitação	Alimentação	Transportes	Despesas Pessoais	Saúde	Vestuário	Educação
<i>Agosto/1987</i>	201570443,95	108557242,66	129782278,93	92901486,26	198550357,84	47252879,33	645182801,19
<i>Setembro/1987</i>	190664438,03	104522667,65	126444153,26	90857199,26	194963037,92	45518619,87	430696128,63
<i>Outubro/1987</i>	178024685,31	100251935,17	116270485,67	87220120,20	173223489,82	38125990,18	404599463,19
<i>Novembro/1987</i>	158483651,01	92313015,73	103618648,56	81029468,72	159962590,95	32304685,65	375463495,84
<i>Dezembro/1987</i>	141465367,22	80377027,07	89403493,01	72483646,66	134365888,92	29464324,65	351294438,40
<i>Março/1988</i>	100642081,13	52118508,32	57790858,37	44990974,04	75829545,67	25209705,61	202789557,62
<i>Abril/1988</i>	88112485,55	43076707,26	49029318,91	38381653,19	64867019,25	21455068,46	140154507,69
<i>Mai/1988</i>	71006918,61	36065561,84	43535179,17	32166990,44	58213245,21	14453697,10	123986648,58
<i>Junho/1988</i>	61276249,96	31217486,10	36291412,94	27855031,42	50292220,35	11998752,19	116288359,14
<i>Julho/1988</i>	51749218,63	24926130,50	29454924,69	23185476,30	41420046,23	10161544,73	105716690,04
<i>Agosto/1988</i>	44120742,14	19662483,42	25427248,39	18681392,36	34502329,05	8991721,62	61498946,67
<i>Fevereiro/1997</i>	6,99	-0,22	3,25	0,20	5,78	-2,79	0,10
<i>Março/1997</i>	6,53	-0,97	3,48	0,10	5,51	0,87	0,28
<i>Abril/1997</i>	6,11	-2,52	3,53	0,25	5,11	5,44	0,06
<i>Mai/1997</i>	5,61	-2,16	3,60	0,35	4,03	-1,52	-0,15
<i>Junho/1997</i>	3,58	-0,80	3,58	0,76	2,78	-6,48	-0,27
<i>Julho/1997</i>	1,65	-1,72	0,67	0,88	1,82	-8,93	-0,28
<i>Agosto/1997</i>	0,89	-1,44	-0,52	1,01	1,31	-7,30	-0,15
<i>Fevereiro/1997</i>	6,99	-0,22	3,25	0,20	5,78	-2,79	0,10
<i>Março/1997</i>	6,53	-0,97	3,48	0,10	5,51	0,87	0,28
<i>Abril/1997</i>	6,11	-2,52	3,53	0,25	5,11	5,44	0,06
<i>Mai/1997</i>	5,61	-2,16	3,60	0,35	4,03	-1,52	-0,15

Fonte: FIPE

Os índices com valores positivos na Tabela 5.8 indicam ocorrência de inflação para o período analisado, enquanto os índices com valores negativos indicam a ocorrência de deflação no período.

Para obter o índice de reajuste para cada mês por faixa de renda, é necessário que se efetue o somatório dos pesos da categoria multiplicado pela sua variação no orçamento familiar, conforme apresentado pela equação 5.1.

$$\text{IPC}_{\text{Renda } j, \text{ Mês } i} = \sum_{n=1}^n \Delta \text{categoria}_n \times \text{peso}_n \quad (5.1)$$

Em que:

$\text{IPC}_{\text{Renda } j, \text{ Mês } i}$: é o índice de reajuste para a faixa de renda j no mês i ;

$\Delta \text{categoria}_n$: é a variação percentual da categoria de despesa familiar;

peso_n : é o peso da categoria na despesa familiar;

j : é a faixa de renda em salários mínimos;

i : é o mês de limite inferior do intervalo de tempo até setembro de 1997;

n : é a categoria da despesa de consumo (habitação, alimentação, transportes, despesas pessoais, saúde, vestuário e educação).

Os IPC são calculados pela equação (5.1) transformando em fatores multiplicativos para ajuste da renda, conforme apresentados na Tabela 5.10.

Tabela 5.10 - Fatores de IPC calculados por mês de entrevista e por faixa de renda em salários mínimos

Mês de Entrevista	Até 2	Mais de 2 a 3	Mais de 3 a 5	Mais de 5 a 6	Mais de 6 a 8	Mais de 8 a 10	Mais de 10 a 15	Mais de 15 a 20	Mais de 20 a 30	Mais de 30
Agosto/1987	0,0502	0,0499	0,0501	0,0503	0,0509	0,0516	0,0519	0,0530	0,0560	0,0557
Setembro/1987	0,0472	0,0467	0,0468	0,0468	0,0471	0,0475	0,0478	0,0485	0,0506	0,0507
Outubro/1987	0,0442	0,0437	0,0437	0,0437	0,0440	0,0443	0,0445	0,0452	0,0471	0,0471
Novembro/1987	0,0401	0,0397	0,0397	0,0397	0,0399	0,0402	0,0404	0,0410	0,0427	0,0426
Dezembro/1987	0,0353	0,0350	0,0350	0,0350	0,0352	0,0355	0,0357	0,0362	0,0378	0,0378
Março/1988	0,0233	0,0231	0,0231	0,0231	0,0232	0,0233	0,0235	0,0237	0,0246	0,0247
Abril/1988	0,0196	0,0195	0,0195	0,0194	0,0195	0,0195	0,0197	0,0198	0,0204	0,0206
Maio/1988	0,0163	0,0161	0,0162	0,0161	0,0162	0,0162	0,0164	0,0165	0,0171	0,0172
Junho/1988	0,0141	0,0140	0,0140	0,0140	0,0140	0,0141	0,0142	0,0143	0,0149	0,0149
Julho/1988	0,0117	0,0116	0,0116	0,0116	0,0116	0,0117	0,0118	0,0119	0,0124	0,0125
Agosto/1988	0,0095	0,0094	0,0094	0,0094	0,0094	0,0094	0,0095	0,0096	0,0099	0,0101
Fevereiro/1997	1,0262	1,0242	1,0253	1,0262	1,0261	1,0267	1,0275	1,0286	1,0278	1,0295
Março/1997	1,0242	1,0224	1,0239	1,0251	1,0251	1,0259	1,0270	1,0283	1,0271	1,0293
Abril/1997	1,0199	1,0181	1,0204	1,0222	1,0223	1,0236	1,0252	1,0267	1,0248	1,0280
Maio/1997	1,0149	1,0132	1,0150	1,0164	1,0165	1,0174	1,0187	1,0203	1,0192	1,0218
Junho/1997	1,0104	1,0092	1,0102	1,0107	1,0109	1,0113	1,0119	1,0130	1,0129	1,0142
Julho/1997	0,9968	0,9956	0,9960	0,9966	0,9967	0,9971	0,9976	0,9988	0,9990	1,0002
Agosto/1997	0,9951	0,9941	0,9942	0,9947	0,9948	0,9951	0,9955	0,9963	0,9966	0,9974

O gráfico da Figura 5.3 apresenta a distribuição da frequência para as rendas individuais dos anos de 1987 e 1997, corrigidas, depois do processo de transferência monetária.

Figura 5.3 - Gráfico das distribuições das frequências para as faixas de renda familiar dos indivíduos dos banco de dados de 1987 e 1997 depois do processo de transferência monetária

O gráfico da Figura 5.4 apresenta a distribuição da frequência acumulativa para as rendas individuais dos anos de 1987 e 1997 depois do processo de transferência monetária.

Figura 5.4 - Gráfico das distribuições das frequências acumulativas para as faixas de renda familiar dos indivíduos nos bancos de dados de 1987 e 1997 depois do processo de transferência monetária

Observa-se, visualmente, que o método de ajuste monetário adotado proporciona uma compatibilidade coerente entre as rendas de 1987 e 1997.

Além das variáveis socioeconômicas obtidas dos levantamentos das Pesquisas O/D foram construídas outros tipos de variáveis com o intuito de obter um maior número de variáveis explicativas para a variável independente, conforme apresentado na Tabela 5.11.

Tabela 5.11 - Variáveis construídas a partir das informações das Pesquisas O/D de 1987 e 1997

Variável	Descrição
<i>Aposentado</i>	Se o indivíduo é aposentado: variável categórica (1 - Sim; 0 - Não)
<i>Assalariado</i>	Se o indivíduo é assalariado: variável categórica (1 - Sim; 0 - Não)
<i>Autônomo</i>	Se o indivíduo exerce atividade autônoma: variável categórica (1 - Sim; 0 - Não)
<i>Criança em Casa</i>	Se há presença de criança com idade menor igual a 5 anos na residência: : variável categórica (1 - Sim; 0 - Não)
<i>Densidade de Auto</i>	Quantidade de automóveis por membro da família: variável contínua
<i>Dona de Casa</i>	Se a mulher é dona de casa: variável categórica (1 - Sim; 0 - Não)
<i>Estuda</i>	Se o indivíduo estuda: variável categórica (1 - Sim; 0 - Não)
<i>Idoso</i>	Se o indivíduo possui idade superior a 60 anos: variável categórica (1 - Sim; 0 - Não)
<i>Renda Familiar per Capita</i>	Renda Familiar per Capita: variável contínua
<i>Trabalha</i>	Se o indivíduo trabalha: variável categórica (1 - Sim; 0 - Não)
<i>Tamanho da Família</i>	Número de membros na família: variável contínua

5.3.5 Verificação da mudança de comportamento de indivíduos

Pela análise dos bancos de dados de 1987 e 1997 foram observadas mudanças de comportamento em três classes de indivíduos (idosos, crianças e mulheres) com influência nos padrões de encadeamento de atividades.

▪ Idosos

Pela classificação do IBGE são consideradas pessoas idosas os indivíduos que possuem 60 anos ou mais. Analisando os idosos (pessoas acima de 60 anos), observa-se que para o banco de dados da Pesquisa O/D de 1987, os indivíduos acima dessa idade representam 7,29% da amostra total (80.764 pessoas) dos quais 65,06% exercem atividades no domicílio. Pelo banco de dados de 1997 esses indivíduos representam 10,43% da amostra total (79.839 pessoas), dos quais 70,43% realizam atividades no domicílio. Conclui-se que em 1997 os idosos exercem mais atividades no domicílio, com um aumento de 5,37 pontos percentuais, fornecendo indícios de que para realizar a previsão de viagens encadeadas para este grupo o modelo deve ser ajustado a esta mudança de comportamento.

Esta observação é inferida comparando os grupos de idosos no banco de dados de 1987 e 1997 a partir do processamento dos dados pelo minerador utilizando um desvio de 0,1 e 50 e 25 para os critérios de mínimo número de observações antes da divisão e nas folhas de ambos os bancos de dados (1987 e 1997) para verificar a relação entre as atividades domiciliares e extra-domiciliares em função da idade dos indivíduos, como pode ser visualizado pela Figura 5.5. A opção por um desvio desta magnitude é dada em função da análise ser apenas de verificação da forma do arranjo dos grupos.

Figura 5.5 - Parte do resultado do processamento do modelo de árvore de decisão para a relação entre as atividades desempenhadas no domicílio (representada pela letra H) e a idade dos indivíduos (acima de 50 anos)

De acordo com a Figura 5.5, observa-se que há uma redução na idade para a frequência relativa de atividades desempenhadas no domicílio de aproximadamente 72%, de 68 anos em 1987 para 63 anos em 1997.

Identificada o grupo (idosos) e o tipo de mudança comportamental (desempenho de atividades no domicílio *versus* desempenho de atividades extra-domiciliares) procede-se com ajuste do modelo que consiste na adaptação do banco de dados utilizado para introduzir a mudança de comportamento dos indivíduos.

O cálculo é baseado na equação 4.2. O total de indivíduos da amostra de 1987 é 80.764, dos quais 5.882 são idosos (indivíduos com mudança), substituindo os valores na equação 4.2, obtém-se $74.882 + \alpha \times 5.882 = \beta \times 80.764$, onde α representa o fator de ajuste do total de idosos e β o fator de ajuste do total do banco de dados. Resolvendo a equação valores para $\alpha = 1,482$ e $\beta = 1,035$.

Substituindo os valores na equação são obtidos novos valores para os indivíduos com idade superior a 60 anos (8.717) e para o total do banco de dados (83.699).

Os valores de Estes valores de α e β apresentados são resultados aproximados. Para o cálculo da quantidade de indivíduos utilizando a planilha eletrônica do Microsoft Excel 2003 não foram utilizadas aproximações, mas todas as casas decimais disponíveis.

Para que dados dos indivíduos sintéticos de 1987 sejam inseridos, é necessário primeiro saber quais as características dos indivíduos que serão inseridos e quais serão retirados da amostra. Esses parâmetros podem ser obtidos em função das características particulares dos subgrupos de idosos para 1997.

A partir do banco de dados de 1997, foi selecionada uma sub-amostra contendo apenas os indivíduos idosos. Esta sub-amostra contempla as características particulares desse grupo e suas respostas ao desempenho do padrão de atividade: código (1) para o indivíduo que realiza padrão de atividade no domicílio e código (0) para o indivíduo que realiza padrão de atividade extra-domiciliar.

A utilização de apenas duas respostas ao padrão de atividade (0 ou outro 1) deve-se a natureza da mudança de comportamento observada: indivíduos que antes exerciam atividades extra-domiliares e que passaram a exercer atividades no domicílio.

Os dados da sub-amostra foram submetidos ao processo de mineração de dados listando as variáveis dependentes e as variáveis independentes (as características particulares dos indivíduos), utilizando um desvio de 0,1 com um número de observações antes da divisão de 50 e de 25 nos nós terminais.

O resultado desse processamento foi uma árvore contendo 13 folhas com os indivíduos agrupados pelas suas características particulares e a probabilidade de ocorrência dos padrões, conforme apresentado na Tabela 5.12.

Tabela 5.12 - Resultado dos 13 grupos de indivíduos para idosos em 1997

Folha /Grupo		Características Socioeconômicas					Indivíduos	Frequência relativa de indivíduos com atividades no domicílio
32	Não Trabalha	Densidade de Auto < 0,183	Pessoas na Família =< 2	Renda Familiar Per Capita < R\$ 615,97	Pessoas na Família = 2		512	0,72
33	Não Trabalha	Densidade de Auto < 0,183	Pessoas na Família =< 2	Renda Familiar Per Capita < R\$ 615,97	Pessoas na Família = 1		1199	0,79
17	Não Trabalha	Densidade de Auto < 0,183	Pessoas na Família =< 2	Renda Familiar Per Capita > R\$ 615,97			328	0,69
18	Não Trabalha	Densidade de Auto < 0,183	Pessoas na Família > 2	Renda Familiar Per Capita < R\$ 152,39			716	0,85
19	Não Trabalha	Densidade de Auto < 0,183	Pessoas na Família > 2	Renda Familiar Per Capita > R\$ 152,39			1176	0,83
10	Não Trabalha	Densidade de Auto > 0,183	Filho ou Parente				804	0,88
44	Não Trabalha	Densidade de Auto > 0,183	Chefe, Cônjuge ou Agregado.	Renda Familiar Per Capita < R\$ 840,94	Renda Familiar < R\$ 633,46		408	0,71
90	Não Trabalha	Densidade de Auto > 0,183	Chefe, Cônjuge ou Agregado.	Renda Familiar Per Capita < R\$ 840,94	Renda Familiar > R\$ 633,46	Idade =< 69	801	0,70
91	Não Trabalha	Densidade de Auto > 0,183	Chefe, Cônjuge ou Agregado.	Renda Familiar Per Capita < R\$ 840,94	Renda Familiar > R\$ 633,46	Idade > 69	411	0,75
23	Não Trabalha	Densidade de Auto > 0,183	Chefe, Cônjuge ou Agregado.	Renda Familiar Per Capita > R\$ 840,94			706	0,67
6	Trabalha	Densidade de Auto < 0,183					540	0,33
14	Trabalha	Densidade de Auto > 0,183	Analfabeto / 1º grau até 4ª. Série ou 1º grau até 4ª. série concluído				277	0,40
15	Trabalha	Densidade de Auto > 0,183	1º grau concluído, colegial concluído ou superior concluído.				447	0,26

O passo seguinte do processo de ajuste consiste no agrupamento dos idosos de 1987 seguindo a classificação estabelecida pela árvore de 1997. Desse agrupamento são obtidos os números de indivíduos por folha de classificação e calculados os fatores de expansão para os indivíduos alocados nesta folha.

Pela Tabela 5.13, os fatores de expansão (coluna [5]) são calculados pelo produto do percentual representativo em relação à amostra de 1997 (coluna [2]) e pela nova quantidade de idosos em 1987, 8.717 indivíduos. Em seguida, pela divisão da nova quantidade de indivíduos da amostra ajustada pela quantidade original de 1987 (coluna [4]) sob a condição de ocorrência de atividades no domicílio (coluna [7]), ou seja: Coluna [2] x 8.717 = Coluna [3]; Coluna [5] = Coluna [3] / Coluna [4]; Coluna [7] = Coluna [4] x Coluna [5] x Coluna [6].

Tabela 5.13 - Fator de expansão por grupo de indivíduo e a frequência de idosos que exercem atividades no domicílio

Folha/Grupo [1]	Peso do grupo em 1997 % [2]	Quantidade ajustada 1987 X Peso [3]	Consulta Dados 1987 [4]	Fator de Expansão do grupo [5]	Frequência relativa de indivíduos com atividades no domicílio [6]	Frequência de indivíduos com atividades no domicílio [7]
32	6,15	536	300	1,79	0,72	386
33	14,40	1255	690	1,82	0,79	985
17	3,94	343	102	3,37	0,69	236
18	8,60	750	795	0,94	0,85	636
19	14,13	1231	839	1,47	0,83	1.020
10	9,66	842	1.710	0,49	0,88	743
44	4,90	427	426	1,00	0,71	304
90	9,62	839	397	2,11	0,70	591
91	4,94	430	128	3,36	0,745	320
23	8,48	739	110	6,72	0,67	496
6	6,49	565	442	1,28	0,33	185
14	3,33	290	216	1,34	0,40	116
15	5,37	468	192	2,44	0,26	121
Total	100%	8.715*	-	-	-	6.139

* devido a aproximações nos cálculos finais da quantidade ajustada de idosos para 1987, o valor calculado de 8.717 foi reduzido para 8.715.

No banco de dados os idosos segundo as características particulares listadas na Tabela 5.12 serão multiplicados pelos fatores de expansão da Tabela 5.13 sob a condição de ocorrência de atividades no domicílio. Para a composição do restante dos indivíduos, aqueles que exercem atividades extra-domiciliares, será adotado o método aleatório para inserção dos indivíduos sintéticos e retirada de registros da amostra.

▪ Crianças

Analisando o banco de dados da Pesquisa O/D de 1987, os indivíduos com idade inferior a 7 anos representam 20,51% da amostra total (80.764 pessoas) dos quais 70,53% das crianças exerciam atividades no domicílio. Pelo banco de dados de 1997 os indivíduos com idade inferior a 7 anos representam 14,60% da amostra total (79.839 pessoas), dos quais 67,58% desempenhavam atividades no domicílio.

Da mesma forma que no caso dos idosos, a diminuição na proporção de crianças na amostra não caracteriza uma mudança comportamental, todavia comparando as análises em relação as atividades exercidas no domicílio e fora dele nota-se uma diminuição de aproximadamente 3 pontos percentuais, na comparação com 1987, fornecendo indícios de que o modelo de previsão de viagens encadeadas deve ser ajustado a essa mudança.

Essa mudança foi observada ao analisar os resultados da mineração dos dados (desvio de 0,1 e 50 e 25 para os critérios de mínimo número de observações antes da divisão e nas folhas) de 1987 e 1997, como apresentado nas Figuras 5.6 e 5.7.

Figura 5.6 - Parte do resultado do processamento do modelo de árvore de decisão para a relação entre o padrão de atividades no domicílio (representado pela letra H) e a idade de indivíduos inferior a 7 anos para a Pesquisa O/D de 1987

Os valores de α e β apresentados são resultados aproximados. Para o cálculo da quantidade de indivíduos utilizando a planilha eletrônica do Microsoft Excel 2003 não foram utilizadas aproximações, mas todas as casas decimais disponíveis.

Para que dados dos indivíduos sintéticos de 1987 sejam inseridos, é necessário primeiro saber quais as características dos indivíduos que serão inseridos e quais serão retirados da amostra. Esses parâmetros podem ser obtidos em função das características particulares dos subgrupos de idosos para 1997.

A partir do banco de dados de 1997, foi selecionada uma sub-amostra contendo apenas os indivíduos crianças. Esta sub-amostra contempla as características particulares desse grupo e suas respostas ao desempenho do padrão de atividade: código (1) para o indivíduo que realiza padrão de atividade no domicílio e código (0) para o indivíduo que realiza padrão de atividade extra-domiciliar.

A utilização de apenas duas respostas ao padrão de atividade (0 ou outro 1) deve-se a natureza da mudança de comportamento observada: indivíduos que antes exerciam atividades extra-domiliares e que passaram a exercer atividades no domicílio.

Os dados da sub-amostra foram submetidos ao processo de mineração de dados listando as variáveis dependentes e as variáveis independentes (as características particulares dos indivíduos), utilizando um desvio de 0,1 com um número de observações antes da divisão de 50 e de 25 nos nós terminais.

O resultado desse processamento foi uma árvore contendo 10 folhas com os indivíduos agrupados pelas suas características particulares e a probabilidade de ocorrência dos padrões, conforme apresentado na Tabela 5.14.

Tabela 5.14 - Resultado dos 10 grupos de indivíduos para crianças em 1997

Folha/Grupo		Características Socioeconômicas					Indivíduos	Frequência relativa de indivíduos com atividades no domicílio
4	Não Estuda	Densidade Auto < 0,095					4677	0,94
5	Não Estuda	Densidade Auto > 0,095					2687	0,91
12	Estuda	Renda Familiar per Capita < R\$ 342,23	Idade =< 4				703	0,26
104	Estuda	Renda Familiar per Capita < R\$ 342,23	Idade > 4	Analfabeto / 1º grau até 4ª. Série	Densidade Auto < 0,095	Renda Familiar Ind. =< R\$95,37	359	0,29
105	Estuda	Renda Familiar per Capita < R\$ 342,23	Idade > 4	Analfabeto / 1º grau até 4ª. Série	Densidade Auto < 0,095	Renda Familiar Ind. > R\$95,37	673	0,30
53	Estuda	Renda Familiar per Capita < R\$ 342,23	Idade > 4	Analfabeto / 1º grau até 4ª. Série	Densidade Auto > 0,095		493	0,25
27	Estuda	Renda Familiar per Capita < R\$ 342,23	Idade > 4	1º grau até 4ª. série concluído			724	0,21
28	Estuda	Renda Familiar per Capita > R\$ 342,23	Renda Familiar per Capita < R\$ 684,67	Idade =< 5			315	0,23
29	Estuda	Renda Familiar per Capita > R\$ 342,23	Renda Familiar per Capita < R\$ 684,67	Idade > 5			407	0,21
15	Estuda	Renda Familiar per Capita > R\$ 342,23	Renda Familiar per Capita > R\$ 684,67				619	0,16

O passo seguinte do processo de ajuste consiste no agrupamento das crianças de 1987 seguindo a classificação estabelecida pela árvore de 1997. Desse agrupamento são obtidos os números de indivíduos por folha de classificação e calculados os fatores de expansão para cada indivíduo alocado nesta folha.

Pela Tabela 5.15, os fatores de expansão (coluna [5]) são calculados pelo produto do percentual representativo em relação à amostra de 1997 (coluna [2]) e pela nova quantidade de idosos em 1987, 11.385 indivíduos. Em seguida, pela divisão da nova quantidade de indivíduos da amostra ajustada pela quantidade original de 1987 (coluna [4]) sob a condição de ocorrência de atividades no domicílio (coluna [7]), ou seja: Coluna [2] x 8.717 = Coluna [3]; Coluna [5] = Coluna [3] / Coluna [4]; Coluna [7] = Coluna [4] x Coluna [5] x Coluna [6].

Tabela 5.15 - Fator de expansão por grupo de indivíduo e ocorrência do padrão H para crianças

Folha/Grupo [1]	Peso do grupo em 1997 % [2]	Quantidade ajustada 1987 X Peso [3]	Consulta Dados 1987 [4]	Fator de Expansão do grupo [5]	Frequência relativa de indivíduos com atividades no domicílio [6]	Frequência de indivíduos com atividades no domicílio [7]
4	40,1	4.568	7.912	0,577	0,94	4.307
5	23,1	2.624	3.957	0,680	0,91	2.377
12	6,0	687	408	1,684	0,26	179
104	3,1	351	727	0,483	0,29	101
105	5,8	657	1.021	0,643	0,30	194
53	4,2	481	822	0,585	0,25	122
27	6,2	707	0	0,000	0,21	151
28	2,7	308	318	0,969	0,26	81
29	3,5	397	500	0,794	0,21	85
15	5,3	605	1.004	0,603	0,16	97
Total	100%	11.385	-	-	-	7.694

No banco de dados as crianças segundo as características particulares listadas na Tabela 5.14 serão multiplicados pelos fatores de expansão da Tabela 5.15 sob a condição de ocorrência de atividades no domicílio. Para a composição do restante dos indivíduos, aqueles que exercem atividades extra-domiciliares, será adotado o método aleatório para inserção dos indivíduos sintéticos e retirada de registros da amostra.

▪ Mulheres

Montali (2004) notou duas tendências gerais na relação papel do indivíduo na família e trabalho entre as décadas de 80 e 90 na Região Metropolitana de São Paulo (RMSP). A primeira tendência se expressa na crescente participação da mulher-cônjuge entre os ocupados da família, na redução do peso do chefe masculino e na diminuição da participação dos filhos.

A segunda tendência refere-se às famílias com chefia feminina sem cônjuge, nas quais a participação da chefe entre os ocupados da família aumenta em decorrência da menor absorção dos filhos e parentes jovens no mercado de trabalho.

Tais mudanças estruturais são verificadas pelo SEADE, que constata que as famílias chefiadas por mulheres na RMSP representavam 10,2% dos arranjos familiares em 1988/1989 e passaram a responder por 14,3% do total de famílias em 2000/2001.

Pela amostra inicial de 1987 (80.764 indivíduos) 43.130 são mulheres, dos quais 18.708 (43,38%) com situação familiar de chefe ou cônjuge. Deste último universo, 4.297 trabalhavam, representando 22,39 % das mulheres de hierarquia familiar de chefe ou cônjuge. Comparando os valores de 1987 aos de 1997 observa-se que as mulheres chefes ou cônjuges que trabalham representam um total de 33,65% da amostra de pessoas do sexo feminino de situação familiar de chefe ou cônjuge, ratificando o estudo de Montali, com o aumento da participação feminina no mercado de trabalho.

O ajuste do banco de dados de 1987 é realizado ajustando a proporção de mulheres cônjuges ou chefes de família que trabalham em relação ao total da amostra. Após a modificação realizada para contemplar a mudança de idosos e crianças, o banco de dados de 1987 apresenta 78.411 indivíduos dos quais 42.173 são mulheres, sendo 19.967 das mulheres que assumem o papel de cônjuge ou chefe na família e 4.381 dessas trabalham (22,39%). Aplicando o percentual de 33,65%, que corresponde aos indivíduos do sexo feminino que trabalham e possuem papel de chefe ou cônjuge na família ao grupo das mulheres é obtido um novo valor de 6.178 indivíduos para essa classe dentro do total de 19.967 mulheres.

Para que dados de 1987 sejam ajustados, é necessário primeiro entender de que forma os dados desse grupo de 1997 são agrupados e qual a ocorrência da probabilidade de atividades desempenhadas no domicílio e atividades extra-domiciliares nos subgrupos.

Dentro da amostra de 1997 foi selecionada uma sub-amostra contendo apenas as mulheres de situação familiar de chefe ou cônjuge. Esta sub-amostra contempla as características particulares desse grupo e suas respostas ao desempenho do padrão de atividade: código (1) para o indivíduo que realiza atividades no domicílio e código (0) para o indivíduo que realiza atividades extra-domiciliares.

Os dados da sub-amostra foram submetidos ao processo de mineração de dados listando as variáveis dependentes e as variáveis independentes (as características particulares dos indivíduos), utilizando um desvio de 0,1 com um número de observações antes da divisão de 25 e de 50 nas folhas.

O resultado desse processamento foi uma árvore contendo 10 folhas com os indivíduos agrupados pelas suas características particulares e a probabilidade de ocorrência dos padrões, conforme apresentado na Tabela 5.16.

Tabela 5.16 - Resultado dos 16 grupos de indivíduos para mulheres em 1997

Folha/ Grupo	Características Socioeconômicas						Indivíduos	Frequência relativa de indivíduos com atividades no domicílio	
32	Não Trabalha	Não Estuda	Densidade Auto <0,155	Idade =< 47	Dona de Casa: Sim		736	0,67	
33	Não Trabalha	Não Estuda	Densidade Auto <0,155	Idade =< 47	Dona de Casa: Não		3.085	0,75	
17	Não Trabalha	Não Estuda	Densidade Auto <0,155	Idade > 47			3.008	0,76	
36	Não Trabalha	Não Estuda	Densidade Auto >0,155	Densidade Auto < 0,392	Idade =< 46		2.210	0,7	
37	Não Trabalha	Não Estuda	Densidade Auto >0,155	Densidade Auto < 0,392	Idade > 46		1.298	0,76	
19	Não Trabalha	Não Estuda	Densidade Auto >0,155	Densidade Auto > 0,392			2.305	0,7	
5	Não Trabalha	Estuda					263	0,29	
96	Trabalha	Densidade Auto < 0,118	Sem Auto	Analfabeto / 1º grau até 4ª. Série ou 1º grau até 4ª. série concluído, ou 1º grau concluído		Não Estuda	Renda Familiar per Capita < 121,14	499	0,17
97	Trabalha	Densidade Auto < 0,118	Sem Auto	Analfabeto / 1º grau até 4ª. Série ou 1º grau até 4ª. série concluído, ou 1º grau concluído		Não Estuda	Renda Familiar per Capita > 121,14	1.629	0,16
49	Trabalha	Densidade Auto < 0,118	Sem Auto	Analfabeto / 1º grau até 4ª. Série ou 1º grau até 4ª. série concluído, ou 1º grau concluído		Estuda		26	0,12
25	Trabalha	Densidade Auto < 0,118	Sem Auto	Colegial concluído ou Superior concluído				600	0,11
13	Trabalha	Densidade Auto < 0,118	Auto = 1					561	0,48
60	Trabalha	Densidade Auto > 0,118	Assalariado	Auto = 1	Analfabeto / 1º grau até 4ª. Série ou 1º grau até 4ª. série concluído, ou 1º grau concluído			808	0,48
61	Trabalha	Densidade Auto > 0,118	Assalariado	Auto = 1	Colegial concluído ou Superior concluído			781	0,25
31	Trabalha	Densidade Auto > 0,118	Assalariado	Auto > 1				868	0,13
14	Trabalha	Densidade Auto > 0,118	Não Assalariado					772	0,21

O passo seguinte do processo de ajuste consiste no agrupamento das mulheres que ocupam chefia ou são cônjuges na família seguindo a classificação estabelecida pela árvore de 1997. Desse agrupamento são extraídos os números de indivíduos por folha de classificação e calculados os fatores de expansão para cada indivíduo alocado nesta folha.

Pela Tabela 5.17, os fatores de expansão (coluna [5]) são calculados pelo produto do percentual representativo em relação à amostra de 1997 (coluna [2]) e pela nova quantidade de mulheres em 1987, 19.967 indivíduos. Em seguida, pela divisão da nova quantidade de indivíduos da amostra ajustada (coluna [3]) pela quantidade original de 1987 (coluna [3]) sob a condição de ocorrência de indivíduos que desempenham atividades

no domicílio (coluna [7]) ou seja: Coluna [2] x 11.384 = Coluna [3]; Coluna [5] = Coluna [3] / Coluna [4]; Coluna [7] = Coluna [4] x Coluna [5] x Coluna [6].

No banco de dados às mulheres, segundo suas características particulares listadas na Tabela 5.16 serão multiplicados pelos fatores de expansão da Tabela 5.17 sob a ocorrência do padrão H no grupo.

Tabela 5.17 - Fator de expansão por grupo de indivíduo e ocorrência para mulheres

Folha/Grupo [1]	Peso do grupo em 1997 % [2]	Quantidade ajustada 1987 X Peso [3]	Consulta Dados 1987 [4]	Fator de Expansão do grupo [5]	Frequência relativa de indivíduos com atividades no domicílio [6]	Frequência de indivíduos com atividades no domicílio [7]
32	3,78	756	237	3,19	0,67	508
33	15,86	3.167	5.642	0,56	0,75	2.389
17	15,47	3.088	3.179	0,97	0,76	2.345
36	11,36	2.269	2.090	1,09	0,7	1.586
37	6,67	1.333	1.166	1,14	0,76	1.015
19	11,85	2.366	3.147	0,75	0,7	1.652
5	1,35	270	135	2,00	0,29	78
96	2,57	512	436	1,17	0,17	89
97	8,38	1.672	1.216	1,38	0,16	265
49	0,13	27	19	1,42	0,12	3
25	3,08	616	314	1,96	0,11	66
13	2,88	576	631	0,91	0,48	277
60	4,15	829	438	1,89	0,48	401
61	4,02	802	497	1,61	0,25	197
31	4,46	891	467	1,72	0,13	112
14	3,97	793	363	2,18	0,21	163
Total	100%	19.967	-	-	-	11.146

5.4 Considerações Adicionais

Ressalta-se que a verificação da mudança de comportamento refere-se à alteração da atividade desempenhada pelo indivíduo. Não obstante, uma das peculiaridades das alterações introduzidas no banco de dados de 1987 teve-se em representar o conjunto de indivíduos que sofreram mudança no comportamento à frequência relativa de ocorrência do desempenho de atividades no domicílio apresentado em 1997.

No caso de idosos e da criança a correspondência para representar a frequência relativa de ocorrência do desempenho de atividades no domicílio traduz a intenção de simular a alteração comportamental no modelo. Contudo no caso das mulheres esta justificativa não é mais válida, uma vez que a mudança identificada é direcionada a atividade trabalho.

Para o ajuste realizado para as mulheres, quando foi selecionado o conjunto de mulheres, foram incluídos indivíduos do sexo feminino acima de 60 anos, havendo, portanto interferência no primeiro ajuste realizado. Contudo, foi observado que no banco de dados original de 1987 (80.764) a quantidade de indivíduos do sexo feminino, com idade acima dos 60 anos, com situação familiar chefe ou cônjuge e que trabalham é 127.

Em relação ao total os 127 casos que são afetados por duas mudanças comportamentais representam 0,16% em relação ao total do banco de dados, 0,29% em relação a quantidade de 43.130 mulheres, 0,68 em relação ao número de casos de 18.708 mulheres com situação familiar de chefe ou cônjuge, 2,95% em relação ao total de mulheres com situação familiar de chefe ou cônjuge que trabalham e 3,64% em relação a quantidade de mulheres com idade acima de 60 anos.

Capítulo 6

Respostas ao Primeiro

Problema Proposto

Neste capítulo são derivadas respostas ao primeiro problema proposto: É possível elaborar um cenário para um modelo de previsão de demanda por transporte que reflita a mudança no comportamento humano, em um período de tempo?

6.1 Considerações Iniciais

Uma das etapas do método proposto teve por objetivo a averiguação da alteração de comportamento dos indivíduos, e se esta produz mudança nos padrões de atividades no período em que se deseja realizar a projeção da demanda. Analisando as fontes de informações disponíveis no trabalho, as Pesquisas Origem/Destino da Região Metropolitana de São Paulo, conduzidas pela Companhia do Metropolitano de São Paulo (RMSP) nos anos de 1987 e 1997, foram observadas influências das mudanças de comportamento nos padrões de viagens em três grupos socioeconômicos: idosos (acima de 60 anos), crianças (com idade inferior a 7 anos) e mulheres (maior inserção no mercado de trabalho). Essas mudan-

ças foram comprovadas pelo processo analítico de mineração de dados e incorporadas ao modelo de previsão.

O modelo de previsão erguido sob condições atuais (1987 sem alterações), tendo a deficiência em retratar as alterações no comportamento na previsão, não proporcionaria resultados satisfatórios. Além do mais, as conjecturas que admitem a transposição de cenários atuais ao longo do tempo, sem alterações, são criticadas por pesquisadores da Escola de Relações Internacionais pela não incorporação destas alterações na construção de cenários futuros.

Vieses opostos a previsões enraizadas no cenário atual são as propostas de cenários futuros formuladas por Kahn e a abordagem perspectiva para construção de cenário futuro descrita por Neuman e Øverland (2004) incorpora a base do conceito da proposta por Kahn.

Apesar de a construção do cenário futuro ser de natureza especulativa, neste trabalho há duas peculiaridades que desmistificam esta característica em relação ao cenário futuro.

Primeiro, é tido como premissa neste trabalho que na comparação entre dois cenários separados por um intervalo de tempo ΔT mudanças são observáveis, desde que este intervalo ΔT seja em uma escala temporal que contemple alterações sociais, culturais, econômicas etc. A característica da construção do cenário deixa de ter um caráter especulativo e passa a ter um caráter investigativo das mudanças. Segundo, o cenário futuro neste trabalho já é conhecido.

Por estas duas peculiaridades e pelas características dos aspectos (variáveis socioeconômicas dos indivíduos) dos bancos de dados que constituem os cenários é que neste trabalho foi adotada a proposta perspectiva para construção de cenário futuro aliada ao conceito de trajetória comportamental descrita por Nagin e Tremblay (2005), Figura 6.1.

Figura 6.1 – Apresentação esquemática do tratamento dos cenários adotado no trabalho

Nos tópicos relacionados abaixo são discutidas as mudanças comportamentais identificadas, auxiliada por informações provenientes do Instituto Brasileiro de Geografia e Estatística (IBGE) e da Fundação Sistema Estadual de Análise de Dados (SEADE).

6.2 Incorporando o Caso da Mudança Comportamental dos Idosos

A redução de mobilidade da população em 1997 comparada a 1987 é observada em quase toda a população. A mobilidade média na RMSP retraiu de 2,06 viagens/habitante em 1987 para 1,87 viagens/habitante em 1997 (BARAT, 2001).

Particularmente no grupo dos idosos, pela análise dos bancos de dados da Pesquisa O/D de 1987 e 1997, a perda da mobilidade é refletida no aumento de atividades domiciliares em 1997 (70,43%) em comparação com 1987 (65,06%). Esta diminuição generalizada da mobilidade pode ser atribuída principalmente ao nível de atividade econômica e a nova fase de expansão capitalista no Brasil, que através de uma modernização seletiva desemprega grandes massas de trabalhadores.

De acordo com dados da SEADE de 1987 a 1997, a taxa de desemprego no Município de São Paulo (MSP) para faixa etária acima de 40 anos saltou de 4% em 1987 para 8,3% em 1997 (Figura 6.2).

Figura 6.2 - Gráfico da evolução do percentual de desemprego por faixa etária no Município de São Paulo (SEADE)

A perda de mobilidade em 1997 toma proporções maiores para o grupo de pessoas acima de 60 anos, pelo seu aumento participativo na composição da população brasileira por faixa etária. Essa tendência vem sendo observada ao longo dos últimos quatro censos demográficos realizados pelo IBGE nos anos de 1970, 1980, 1991 e 2000 (Figuras 6.3a - 6.3d).

Figura 6.3a - Gráfico da participação da população por faixa etária na população brasileira nos censos de 1970 (IBGE, 1987)

Figura 6.3b - Gráfico participação da população por faixa etária na população brasileira nos censos de 1980 (IBGE, 1987)

Figura 6.3c - Gráfico da participação da população por faixa etária na população brasileira nos censos de 1991 (IBGE)

Figura 6.3d - Gráfico da participação da população por faixa etária na população brasileira nos censos de 2000 (IBGE)

Observa-se pelos gráficos das Figuras 6.3a - 6.3d uma modificação na pirâmide etária com o encurtamento da base e alargamento da parte superior. Esse alargamento da parte superior é reflexo do aumento da população com idade acima de 40 anos em decorrência da redução da taxa de mortalidade e melhoramentos na área de saúde nas últimas décadas.

Analisando o comportamento da alteração da pirâmide etária para a RMSP no período de 1987 e 1997, observa-se que participação dos idosos na população da RMSP saltou de 6,52% para 7,71%. Comparativamente, enquanto a RMSP experimentava um crescimento populacional de 18,81% no período, o grupo de idosos aumentou em 40,64% (930.667 para 1.308.922), como pode ser observado pelos gráficos das Figuras 6.4a e 6.4b.

Figura 6.4a - Gráfico da participação da população por faixa etária na RMSP 1987 (SEADE)

Figura 6.4b - Gráfico da participação da população por faixa etária na RMSP 1997 (SEADE)

Essa tendência na RMSP foi também observada através da comparação das pirâmides etárias das amostras das Pesquisas Origem-Destino de 1987 e 1997 com aumento da população idosa amostrada em 1987 de 7,29% para 10,43% em 1997, Figuras 6.5a e 6.5b.

Figura 6.5a – Gráfico da participação por faixa etária na amostras da Pesquisa OD de 1987 RMSP

Figura 6.5b – Gráfico da participação por faixa etária na amostras da Pesquisa OD de 1997 RMSP

Ajustando o banco de dados de 1987 para as mudanças verificadas, a pirâmide etária do banco de dados sintético da amostra de 1987 é dada pelo gráfico da Figura 6.6.

Nesta nova composição a participação da população idosa é de 10,82%, retratando a tendência observada para o período de 1997 incluindo a mudança de comportamento.

Figura 6.6 - Gráfico da participação da população por faixa etária da amostra sintética de 1987 da RMSP

6.3 Incorporando o Caso da Mudança Comportamental das Crianças

De 1988 a 1997 a população com idade até 19 anos saltou de 5.883.504 para 6.335.635 indivíduos, um crescimento de 7,68% no período (Figura 6.7).

Figura 6.7 - Gráfico da evolução da população jovem (até 19 anos) na RMSP: 1988 a 1997 (SEADE)

Ao comparar a evolução do crescimento populacional dos jovens com idade até 19 anos e o acompanhamento do número de matrículas iniciais (registros efetivados ao início de cada ano letivo) nos ensinos pré-escolar, fundamental e médio, pode ser observado uma inserção cada vez maior do jovem no ensino através do aumento gradativo, com exceção ao ano de 1996, que experimentou uma queda no número de matrículas, Figura 6.8.

Figura 6.8 - Gráfico comparativo do total de matrículas na pré-escola, fundamental e ensino médio na RMSP (SEADE)

Em 1988 foram realizadas um total 3.600.918 matrículas (pré-escola, fundamental e médio) o que representou 61,2% da participação da população jovem nas escolas. Em 1997 foram efetuadas 4.626.050 matrículas (crescimento de 28,47% no período) com um incremento da participação dos jovens em 11,8 pontos percentuais, respondendo por 73% engajados na atividade de estudo.

Este aumento da inserção do jovem na escola afeta diretamente o padrão de viagem encadeada deste grupo, principalmente para indivíduos que não exerciam atividades extra-domiciliar e passam a exercer.

Como a demanda por viagem é gerada pela demanda por atividade, a inserção de uma atividade extra-domiciliar, que no caso é a viagem se dirigindo à escola, modifica o tipo de padrão de viagem. Este processo foi particularmente observado para crianças até 7 anos, nos dados das amostras da Pesquisa O/D.

Em 1987, 70,53% da população abaixo de 7 anos só exerciam atividades domiciliares, enquanto em 1997 as crianças que só realizavam atividades domiciliares passaram a representar 67,58%, indicando que houve um aumento da mobilidade para esse grupo, em consequência dos deslocamentos para a atividade de estudo.

Este indício, de crianças exercendo a atividade extra-domiciliar (estudo), pode ser corroborado pelo aumento do número de crianças com idade até 6 anos matriculadas em escolas na RMSP. Em 1991 eram 2.124.546 registros de matrículas em escolas na RMSP e em 1997 foram 2.184.771 registros em 1997, um crescimento de 2,83%. No mesmo período a população com faixa etária até 9 anos sofreu uma redução de 0,19%.

Ajustando o banco de dados de 1987 para as mudanças verificadas, a nova composição para atividades realizadas no domicílio e fora do domicílio é dado pela Tabela 6.1.

Tabela 6.1 – Comparação dos percentuais representativos de crianças com idade igual e inferior a 7 anos e o padrão de atividade desempenhado

	1987		1987 Sintético		1997	
	Frequência	%	Frequência	%	Frequência	%
<i>Indivíduos</i>	80.764	100,00	78.411	100,00	79.839	100,00
<i>Crianças Inferiores a 7 anos</i>	16.569	20,52	11.385	14,52	11.657	14,60
Padrão Crianças	Frequência	%	Frequência	%	Frequência	%
<i>Atividade no Domicílio</i>	11.686	70,53	7.694	67,58	7.878	67,58
<i>Atividades Extra-Domiciliares</i>	4.883	29,47	3.691	32,42	3.779	32,42

Observa-se pelos valores da Tabela 6.1 que o banco de dados da amostra sintética de 1987 representa a mudança comportamental das crianças com idade igual e inferior a 7 anos para o padrão de atividades.

6.4 Incorporando o Caso da Mudança Comportamental das Mulheres

A deterioração das condições de inserção no mercado de trabalho ao final dos anos 80 e na década 90, incluindo o aumento do desemprego (crescimento taxa de desempregados de 8,9% em 1987 para 14,8% em 1997, Figura 6.9) e retração dos postos de trabalho (redução de 247.277 do total de empregos ocupados de 1987 para 1997, Figura 6.10) provocou a queda dos rendimentos do trabalho para os ocupados, com maior intensidade ao final dos anos 90, quando novamente ocorreu redução da atividade econômica em âmbito nacional.

Figura 6.9 - Gráfico do crescimento da taxa de desemprego na RMSP (SEADE)

Figura 6.10 - Gráfico da evolução no número de empregos formais na RMSP (SEADE)

Dada à precariedade do mercado de trabalho, o preenchimento das necessidades de manutenção do padrão de vida da família passou a depender cada vez mais do esforço familiar coletivo, com maior divisão entre seus membros para composição da renda familiar.

Neste contexto está a alteração do papel da cônjuge mulher no domicílio. As atribuições do cônjuge no domicílio que eram ligadas a trabalhos domésticos, devido à dificuldade de manutenção do domicílio, vem sendo alteradas, dentre outros motivos, pela inserção da mulher cônjuge no mercado de trabalho.

Entretanto, é necessário observar que a relação entre família e mercado de trabalho não é direta. Os valores culturais e sociais são definidores das possibilidades de seus diferentes componentes, homens, mulheres, jovens e adultos, estarem disponíveis e serem disponibilizados para o trabalho.

A família é, portanto, um espaço de socialização dos indivíduos: em seu interior articulam-se valores, práticas e tradições, convivem hierarquias e são definidos papéis e reafirmadas as identidades. As possibilidades de cada membro inserir-se no mercado de trabalho, além da conjuntura econômica que define a disponibilidade de vagas de postos criados, são determinadas, portanto, pelas relações de gênero e de idade, pela posição na família e pelas atribuições domésticas segundo a composição da família.

É na articulação de seus membros com o mercado de trabalho que é possível organizar a subsistência do grupo, através de um esforço coletivo que não se reduz apenas aos imperativos da sobrevivência imediata, mas a superação ou melhora das condições de vida, conformadas aos papéis familiares e aos mecanismos do mercado de trabalho.

Sob a influência deste ambiente é que se observou uma maior inserção da mulher no mercado de trabalho, com a redução gradativa do número de desempregados mulheres e aumento dos homens desempregados, Tabela 6.2.

Tabela 6.2 - Participação de homens e mulheres no desemprego

Ano	Homens (%)	Mulheres (%)
1987	49,00	51,00
1988	51,20	48,80
1989	52,27	47,73
1990	53,81	46,19
1991	54,60	45,40
1992	54,00	46,00
1993	53,70	46,30
1994	52,41	47,59
1995	51,84	48,16
1996	51,60	48,40
1997	50,85	49,15

Fonte: Fundação SEADE

Esta nova estrutura, em que há uma maior participação feminina no mercado de trabalho, foi notada na análise comparativa dos bancos de dados utilizados no experimento.

Em 1987 as mulheres que ocupavam posição de chefe ou cônjuge e que exerciam qualquer atividade de trabalho representavam 22,97% do total de mulheres do banco de dados. Em uma década este percentual sofreu um crescimento, passando para 33,65%.

Este fenômeno foi absorvido no ajuste do banco de dados de 1987 produzido no banco de dados sintético de 1987 utilizado para construção do modelo, Tabela 6.3.

Tabela 6.3 - Comparação dos percentuais representativos das mulheres chefes ou cônjuges e sua relação com a atividade trabalho

	1987		1987 Sintético		1997	
	Frequência	%	Frequência	%	Frequência	%
<i>Indivíduos</i>	80.764	100,00	78.411	100,00	79.839	100,00
<i>Mulher</i>	43.130	53,40	42.173	53,79	42.317	53,00
<i>Mulher</i>	Frequência	%	Frequência	%	Frequência	%
<i>Chefe</i>	2.948	6,84	3.965	9,40	4.427	10,46
<i>Cônjuge</i>	15.760	36,54	16.002	37,94	15.022	35,50
<i>Chefe ou Cônjuge</i>	18.708	43,38	19.967	47,34	19.449	45,96
<i>Mulher Chefe ou Cônjuge</i>	Frequência	%	Frequência	%	Frequência	%
<i>Trabalha</i>	4.297	22,97	6.718	33,65	6.544	33,65
<i>Não Trabalha</i>	14.411	77,03	13.249	66,35	12.905	66,35

6.5 Atributos do Cenário Construído

Em síntese, após comparação dos bancos de dados de 1987 (que retrata o cenário do ano T) com as informações do banco de dados de 1997 (que retrata o cenário futuro do ano $T+\Delta T$), foram identificadas mudanças comportamentais em três grupos: idosos, crianças e mulheres.

Estas mudanças foram incorporadas ao banco de dados de 1987 gerando o banco de dados sintético para 1987 que foi utilizado no modelo para projeção da demanda para 1997. Nas subseções a seguir são realizadas comparações entre os principais atributos socioeconômicos dos três bancos de dados. Não é intuito das subseções avaliar a construção do banco de dados sintético e sim apresentar de forma comparativa as características dos indivíduos que compõem os bancos de dados de 1987, 1987 adaptado e 1997.

6.5.1 Tamanho da família

Observa-se pelos dados da Tabela 6.4, onde são apresentados a média e o desvio padrão, que em média o tamanho da família no banco de dados gerado, 1987 sintético, foi reduzido se comparado ao banco de dados original (1987). A frequência de redução na quantidade de membros na família ocorreu em famílias com apenas um membro e acima de 5 membros, e de forma oposta houve aumento da frequência de famílias com 2, 3 e 4 membros (Figura 6.11).

Tabela 6.4 – Média e desvio padrão do atributo tamanho da família nos bancos de dados

	1987	1987 Sintético	1997
<i>Média</i>	4,3667	4,2866	4,5449
<i>Desvio Padrão</i>	2,53792	2,49243	1,89838

Figura 6.11 - Gráfico da representação percentual da frequência da característica tamanho da família nas amostras da RMSP

Visualmente a partir do gráfico da Figura 6.11 a distribuição das frequências relativas do banco de dados de 1987 sintético segue a tendência de distribuição da amostra de 1987 do que se comparada a amostra de 1997. Na amostra de 1997, 46,68% dos indivíduos são membros de famílias de 4 e 5 pessoas enquanto em 1987 este valor é de 29,77% e de 29,41% no banco de dados sintético.

6.5.2 Posse de automóvel

Comparando a média da posse de automóveis de 1987 e 1997, Tabela 6.5, observa-se que em 1997 houve um aumento da posse de auto com crescimento em todas as faixas de quantidade de veículos, Figura 6.11. O banco de dados sintético gerado reproduz uma parcela do aumento da posse de automóveis, mas não retrata o quadro esperado para 1997.

Tabela 6.5 - Média e desvio padrão do atributo posse de automóveis nos bancos de dados

	1987	1987 Sintético	1997
<i>Média</i>	0,5521	0,5805	0,7210
<i>Desvio Padrão</i>	0,76373	0,77310	0,91802

Figura 6.12 - Gráfico da representação percentual da frequência da característica posse de automóveis nas amostras da RMSP

6.5.3 Sexo

Pela Figura 6.13, nota-se que a proporção entre homens e mulheres no banco de dados manteve-se dentro de um intervalo de $\pm 0,8\%$, sem variações acima de 1%.

Figura 6.13 - Gráficos da representação percentual da frequência da característica sexo nos bancos de dados

6.5.4 Situação familiar

Em função das manipulações realizadas para ajuste as mudanças comportamentais verificadas, no banco de dados sintético houve um aumento de 3,4% da participação do chefe, aumento de 0,9% do cônjuge e redução de membros filhos em 4,7% comparado ao banco de dados original de 1987, Figura 6.14.

Acredita-se que o aumento da participação dos papéis familiares chefe e cônjuge pode ser atribuído à alteração imposta pelo caso mulher chefe ou cônjuge que trabalha e a redução do papel filho pode ser atribuída ao ajuste para o caso das crianças com idade igual e inferior a 7 anos. Devido a estas alterações a composição da situação familiar do banco de dados sintético se aproxima ao retrato do banco de dados de 1997.

Figura 6.14 – Gráficos da representação percentual da frequência da característica situação familiar nos bancos de dados

6.5.5 Grau de instrução

A composição do grau de instrução de 1987 e 1997 são distintos em tendências. Enquanto em 1987 é observada uma grande parcela de analfabetos ou de escolaridade incompleta até a 4ª. série, 1997 apresenta grande parcela de indivíduos de 1º grau completo, Figura 6.15.

O banco de dados sintético apresenta uma tendência de aumento de escolaridade dos indivíduos.

Figura 6.15 - Gráficos da representação percentual da frequência da característica grau de instrução nas amostras da RMSP

6.5.6 Estudo

Analisando os valores representativos da característica estudo nos banco de dados 1987 e 1987 sintético observa-se que em relação a este atributo os valores são quase idênticos (diferença de 0,2%).

Figura 6.16 - Gráficos da representação percentual da frequência da característica estudo nas amostras da RMSP

6.5.7 Trabalho

Analisando os gráficos da Figura 6.17, observa-se um aumento da quantidade de indivíduos que realizam a atividade trabalho. Um dos fenômenos que pode ser atribuído ao aumento da quantidade de indivíduos que na amostra sintética trabalham pode ser creditada a incorporação do caso da mudança de comportamento da mulher cônjuge ou chefe que trabalha.

Figura 6.17 – Gráficos da representação percentual da frequência da característica trabalho nas amostras da RMSP

6.5.8 Renda individual

Analisando as peculiaridades das médias e desvios da Tabela 6.6 para os valores de renda individual dos bancos de dados, nota-se que o banco de dados sintético apresenta em relação à renda individual uma média superior aos valores de 1987 e 1997.

Tabela 6.6 – Média e desvio padrão do atributo renda individual nos bancos de dados (em R\$)

	1987	1987 Sintético	1997
<i>Média</i>	281,66	337,85	279,26
<i>Desvio Padrão</i>	755,81	844,84	866,20

Pode ser atribuída à forma de manipulação dos dados dos indivíduos que foram clonados o aumento do valor médio da renda individual no banco de dados. Analisando as

distribuições das frequências para as distribuições das faixas de renda individual dos bancos de dados no gráfico da Figura 6.18, observa-se um aumento da quantidade de indivíduos nas faixas de renda superior a R\$ 0,01 e diminuição dos indivíduos sem renda.

Figura 6.18 – Gráficos das distribuições das frequências para as faixas de renda dos indivíduos nos bancos de dados

No gráfico da Figura 6.19, onde são apresentadas as frequências acumulativas das rendas individuais para os três bancos de dados, observam-se desvios superiores a 3% (Tabela 6.7) se comparar as amostras de 1987 e 1997 à amostra sintética de 1987 nas três primeiras faixas de renda. A maior razão para este comportamento é a frequência representativa dos indivíduos na faixa de renda igual a zero.

Tabela 6.7 – Desvios das frequências acumulativas por faixa de renda dos indivíduos das amostras de 1987 e 1997 em relação à amostra sintética de 1987

Desvio	Faixas de Renda em R\$										
	0	0,01 até 240,00	240,01 até 360,00	360,01 até 600,00	600,01 até 720,00	720,01 até 960,00	960,01 até 1.200,00	1.200,01 até 1.800,00	1.800,01 até 2.400,00	2.400,01 até 3.600,00	Mais de 3.600,01
$ D_{1987} $	6,06%	4,49%	3,70%	2,54%	2,13%	1,58%	1,13%	0,71%	0,50%	0,28%	0,00%
$ D_{1997} $	8,58%	5,95%	5,99%	3,31%	3,40%	1,99%	1,06%	0,40%	0,32%	0,13%	0,00%

Figura 6.19 - Gráficos das distribuições das frequências acumuladas para as faixas de renda dos indivíduos nos bancos de dados

Analisando visualmente o comportamento das curvas do gráfico da Figura 6.19 nota-se um amortecimento dos desvios para as faixas de rendas acima de R\$600 reais. Este amortecimento é dado em função da similaridade de frequências existentes para faixas superiores a este valor nos três bancos de dados.

6.5.9 Renda familiar

Pela análise dos valores da Tabela 6.8, onde são apresentados os valores de média e de desvio padrão da renda familiar nos três bancos de dados, observa-se que em 1997 há um aumento de 28,02% da renda familiar média. Na amostra sintética o valor da renda familiar média apresenta um aumento de 5,11% em relação à média de 1987 e 17,89% menor do que o valor aferido para 1997.

Tabela 6.8 - Média e desvio padrão do atributo renda familiar nos bancos de dados (em R\$)

	1987	1987 Sintético	1997
<i>Média</i>	1.392,74	1.464,02	1.783,00
<i>Desvio Padrão</i>	1.837,19	1.910,44	2.433,85

Observado a distribuição das frequências relativas das faixas de renda familiar pode ser notado que no banco de dados sintético para recebimentos familiares até R\$1.200,00 os números da frequência relativa da amostra sintética são inferiores a 1987, passando a ser maiores para valores acima de R\$ 1.200,01.

Em relação à distribuição de 1997 observa-se que as faixas com vencimentos familiares acima de R\$ 1.200,01 desta amostra são superiores a da amostra sintética e de 1987.

Figura 6.20 - Gráficos das distribuições das frequências para as faixas de renda das famílias nos bancos de dados

Analisando as curvas de distribuição acumulativa das faixas de renda familiar, gráfico da Figura 6.21, observa-se que as curvas de 1987 e 1987 sintético possuem a mesma tendência, com desvios não superiores a 2,3% (Tabela 6.9) e média de 1,31%.

Em relação a 1997, os desvios iniciais da amostra sintética são superiores a 5% sendo amortecidos ao final da distribuição.

Figura 6.21 - Gráficos das distribuições das frequências acumuladas para as faixas de renda das famílias nos bancos de dados

Tabela 6.9 - Desvios das frequências acumuladas por faixa de renda dos indivíduos das amostras de 1987 e 1997 em relação à amostra sintética de 1987

Desvio	Faixas de Renda em R\$										
	0	0,01 até 240,00	240,01 até 360,00	360,01 até 600,00	600,01 até 720,00	720,01 até 960,00	960,01 até 1.200,00	1.200,01 até 1.800,00	1.800,01 até 2.400,00	2.400,01 até 3.600,00	Mais de 3.600,01
$ D_{1987} $	0,12%	0,19%	0,55%	1,60%	1,86%	2,22%	2,28%	1,74%	1,31%	0,76%	0,00%
$ D_{1997} $	7,85%	8,53%	8,11%	7,73%	5,63%	4,92%	5,49%	4,72%	4,07%	3,39%	0,00%

6.6 Considerações Adicionais

Após uma análise das características do cenário elaborado para projeção do modelo conclui-se que em relação ao primeiro problema proposto (É possível elaborar um cenário para um modelo de previsão de demanda por transporte que reflita a mudança no comportamento humano, em um período de tempo?) foi possível elaborar um cenário que

retratasse as mudanças comportamentais identificadas nos grupos socioeconômicos, idosos, crianças e mulheres no período de tempo de 1987 a 1997.

Capítulo 7

Respostas ao Segundo

Problema Proposto

Neste capítulo são derivadas respostas ao segundo problema proposto: É possível utilizar uma técnica de modelagem capaz de realizar previsões de padrões de viagens encadeadas no meio urbano, que contemplem as mudanças comportamentais dos indivíduos?

7.1 Resultados Obtidos pela Mineração dos dados: Construção do Modelo de Previsão Baseado no Banco de Dados Sintético de 1987

A fase de processamento dos dados corresponde à aplicação do minerador de dados com o modelo de árvore de decisão ao banco de dados sintético de 1987 (78.411 indivíduos), adotando um desvio de 0,01 e os critérios de 10 e 5 para o número mínimo de observações antes da divisão e o número mínimo de observações nas folhas. Estes critérios foram adotados como tentativa em se obter um modelo que seja ao máximo detalhista, classificando os dados de 1997 aos agrupamentos estabelecidos para a amostra sintética de 1987.

Foi adotado como variável independente os 128 padrões de viagens encadeadas, apresentados na Tabela 5.2, e como variáveis dependentes o número de indivíduos na família (FAM), a renda familiar (RFAM), a quantidade de autos (AUTO), sexo (SEXO), situação familiar (SF), idade (IDADE), grau de instrução (GI), estuda (ESTUDA), trabalha (TRAB), aposentado (APOSE), dona de casa (DCASA), assalariado (ASSAL), autônomo (AUT), densidade de auto (DENSAUTO), criança na residência (CRIAN.H) e idoso (VELHO).

Na mineração dos dados o modelo de árvore selecionou as variáveis "TRAB", "ESTUDA", "SF", "DENSAUTO", "DCASA", "FAM", "IDADE", "RFAM", "GI", "AUT", "SEXO", "AUTO" e "VELHO".

O resultado do processamento é uma árvore com a divisão dos indivíduos em 110 grupos (folhas) distintos e as probabilidades de ocorrências dos 128 padrões de viagens apresentados em forma de relatório, Figura 7.1, e o esquema da árvore, Figura 7.2.

Figura 7.1 - Parte do relatório gerado pelo S-Plus após o processamento do banco de dados sintético de 1987

128 padrões de viagens encadeadas em ordem alfabética									
	H	HAAHPPP445	HAAHPPP555	HAHNN12	HAHNN15	HAHNN22	HAHNN23	HAHNN24	HAHN
Zona 1	40,45	0,09	0,06	0,15	0,05	0,86	0,30	0,15	0,0
Zona 2	47,11	0,06	0,04	0,16	0,13	0,96	0,41	0,26	0,1
Zona 3	37,47	0,10	0,04	0,12	0,09	0,60	0,26	0,25	0,0
Zona 4	19,78	0,07	0,03	0,08	0,02	0,33	0,19	0,14	0,0
Zona 5	19,92	0,07	0,05	0,08	0,02	0,37	0,10	0,11	0,0
Zona 6	28,44	0,05	0,06	0,08	0,06	0,57	0,19	0,16	0,0
Zona 7	56,98	0,22	0,07	0,14	0,08	0,67	0,23	0,27	0,1
Zona 8	38,20	0,06	0,05	0,11	0,05	0,63	0,31	0,19	0,0
Zona 9	41,32	0,13	0,08	0,14	0,05	0,77	0,23	0,18	0,0
Zona 10	53,60	0,20	0,08	0,15	0,06	0,71	0,27	0,16	0,0
Zona 11	73,68	0,39	0,31	0,29	0,09	1,02	0,45	0,40	0,1
Zona 12	67,25	0,38	0,30	0,27	0,06	1,00	0,34	0,30	0,1

Valor estimado de indivíduos na zona de tráfego por padrão

Figura 7.5 - Parte da planilha eletrônica do Microsoft Office Excel 2003 com o valor estimado de indivíduos por padrão na zona de tráfego

7.3 Levantamento dos Padrões de Viagens Encadeadas Observados para os Indivíduos do Banco de Dados de 1997

Para o levantamento dos padrões de viagens encadeadas do banco de dados de 1997 foi utilizado o programa de estatística SPSS 13.0 for Windows. A escolha do SPSS para manipulação do banco de dados de 1997 foi em função das ferramentas de análise estatística disponíveis por este programa além da simples interface gráfica.

Através do SPSS os indivíduos do banco de dados foram agrupados segundo as zonas a que eles pertencem, e em seguida foram extraídas as quantidades observadas de indivíduos por padrão na zona de tráfego.

O arquivo gerado pelo SPSS foi exportado para o Microsoft Office Excel 2003 (Figura 7.6) para a etapa seguinte de comparação dos resultados estimados e observados.

128 padrões de viagens encadeadas em ordem alfabética

	H	HAAHPPP445	HAAHPPP555	HAHNN12	HAHNN15	HAHNN22	HAHNN23	HAHNN24	HAHNN25
Zona 1	47			1		1	1		2
Zona 2	47					1	2		
Zona 3	32					1			
Zona 4	30						2		
Zona 5	16					1			
Zona 6	27					1			
Zona 7	70						1		
Zona 8	45					1			
Zona 9	50					1			
Zona 10	65					1	1		
Zona 11	74					1			
Zona 12	82								

Quantidade observada de indivíduos na zona de tráfego por padrão

Figura 7.6 - Parte da planilha eletrônica do Microsoft Office Excel 2003 com a quantidade observada de indivíduos por padrão na zona de tráfego

Associando os 128 padrões mais frequentes de 1987 aos 128 padrões mais frequentes de 1997, foram encontrados 117 tipos de padrões comuns a ambos os bancos de dados, restando 11 que aparecem entres os 128 mais frequentes de 1997 nas que não são contemplados em 1987 e 11 em 1987 que não existem no banco de dados de 1997.

Estes padrões não similares reúnem um grupo de 651 indivíduos (0,81%) da amostra de 1997. Isto implica a impossibilidade de ser realizada a previsão para esses indivíduos. Os padrões similares e não-similares totalizaram 139 padrões de viagens encadeadas, que foram considerados na comparação dos resultados nas 361 zonas de tráfego da RMSP.

7.4 Comparação dos Resultados: Aplicação da Prova de Kolmogorov-Smirnov (K-S)

Conforme salientado no capítulo onde foi apresentado o método do experimento, em virtude de a prova de Kolmogorov-Smirnov ser sensível à ordem em que é realizada a comparação dos valores foram adotadas nesta etapa dois estados comparativos dos resultados. O primeiro estado adotado foi a comparação dos valores na forma em que são dispostos os valores dos padrões de viagem encadeada pelo S-Plus, ordem alfabética. O se-

gundo estado adotado foi à comparação dos valores em ordem aleatória, sendo realizado cinquenta seqüências diferentes para ordenação dos padrões.

7.4.1 Seqüência dos padrões gerada pelo S-Plus: ordem alfabética

Os resultados da previsão (quantidade esperada de indivíduos por padrão na zona de tráfego) são dispostos em uma matriz, denominada MATRIZ PREVISÃO (Figura 7.5), onde as linhas representam as 361 zonas de tráfego e as colunas os 139 padrões de viagens encadeadas em ordem alfabética.

Uma outra matriz, denominada MATRIZ REAL (Figura 7.6), é montada apresentando a mesma disposição da MATRIZ PREVISÃO onde as linhas representam as zonas de tráfego e as colunas os 139 padrões de viagens encadeadas. Cada célula desta matriz é preenchida com os somatórios dos indivíduos que realizam aquele padrão de viagem na zona de tráfego.

Quatro outras matrizes são construídas, seguindo o mesmo critério de linha e coluna. Nas duas primeiras matrizes, as células são preenchidas com os valores das frequências acumuladas (para cada zona distinta) para os resultados da MATRIZ REAL e da MATRIZ PREVISÃO, estas duas matrizes recebem a denominação de MATRIZ F-PREVISÃO (Figura 7.7) e MATRIZ F-REAL (Figura 7.8).

128 padrões de viagens encadeadas em ordem alfabética									
	H	HAAHPPP445	HAAHPPP555	HAHNN12	HAHNN15	HAHNN22	HAHNN23	HAHNN24	HAHN
Zona 1	40,45	0,09	0,06	0,15	0,05	0,86	0,30	0,15	0,0
Zona 2	47,11	0,06	0,04	0,16	0,13	0,96	0,41	0,26	0,1
Zona 3	37,47	0,10	0,04	0,12	0,09	0,60	0,26	0,25	0,0
Zona 4	19,78	0,07	0,03	0,08	0,02	0,33	0,19	0,14	0,0
Zona 5	19,92	0,07	0,05	0,08	0,02	0,37	0,10	0,11	0,0
Zona 6	28,44	0,05	0,06	0,08	0,06	0,57	0,19	0,16	0,0
Zona 7	56,98	0,22	0,07	0,14	0,08	0,67	0,23	0,27	0,1
Zona 8	38,20	0,06	0,05	0,11	0,05	0,63	0,31	0,19	0,0
Zona 9	41,32	0,13	0,08	0,14	0,05	0,77	0,23	0,18	0,0
Zona 10	53,60	0,20	0,08	0,15	0,06	0,71	0,27	0,16	0,0
Zona 11	73,68	0,39	0,31	0,29	0,09	1,02	0,45	0,40	0,1
Zona 12	67,35	0,38	0,50	0,37	0,06	1,00	0,34	0,30	0,1

Valor estimado de indivíduos na zona de tráfego por padrão

Figura 7.7 - Parte da planilha eletrônica do Microsoft Office Excel 2003 contendo a MATRIZ F-PREVISÃO

128 padrões de viagens encadeadas em ordem alfabética									
	H	HAAHPPP445	HAAHPPP555	HAHNN12	HAHNN15	HAHNN22	HAHNN23	HAHNN24	HAHN
Zona 1	47,00	47,00	47,00	48,00	48,00	49,00	50,00	50,00	52,00
Zona 2	47,00	47,00	47,00	47,00	47,00	48,00	50,00	50,00	50,00
Zona 3	32,00	32,00	32,00	32,00	32,00	33,00	33,00	33,00	34,00
Zona 4	30,00	30,00	30,00	30,00	30,00	30,00	32,00	32,00	32,00
Zona 5	16,00	16,00	16,00	16,00	16,00	17,00	17,00	17,00	17,00
Zona 6	27,00	27,00	27,00	27,00	27,00	28,00	28,00	28,00	28,00
Zona 7	70,00	70,00	70,00	70,00	70,00	70,00	71,00	71,00	71,00
Zona 8	45,00	45,00	45,00	45,00	45,00	46,00	46,00	46,00	46,00
Zona 9	50,00	50,00	50,00	50,00	50,00	51,00	51,00	51,00	51,00
Zona 10	65,00	65,00	65,00	65,00	65,00	66,00	67,00	67,00	67,00
Zona 11	74,00	74,00	74,00	74,00	74,00	75,00	75,00	75,00	75,00
Zona 12	82,00	82,00	82,00	82,00	82,00	82,00	82,00	82,00	82,00

Frequência acumulada por zona

Figura 7.8 - Parte da planilha eletrônica do Microsoft Office Excel 2003 contendo a MATRIZ F-REAL

As células das duas matrizes seguintes são preenchidas com o resultado da frequência acumulada dividido pelo total da zona, o valor de cada célula estará compreendido no intervalo de 0 a 1, seguindo uma ordem crescente de distribuição. Essas matrizes são denominadas de MATRIZ P-PREVISÃO (Figura 7.9) e MATRIZ P-REAL (Figura 7.10).

128 padrões de viagens encadeadas em ordem alfabética									
	H	HAAHPPP445	HAAHPPP555	HAHNN12	HAHNN15	HAHNN22	HAHNN23	HAHNN24	HAHN
Zona 1	0,36	0,37	0,37	0,37	0,37	0,38	0,38	0,38	0,3
Zona 2	0,41	0,41	0,41	0,42	0,42	0,43	0,43	0,43	0,4
Zona 3	0,39	0,40	0,40	0,40	0,40	0,40	0,41	0,41	0,4
Zona 4	0,33	0,33	0,33	0,33	0,33	0,34	0,34	0,34	0,3
Zona 5	0,30	0,30	0,30	0,30	0,31	0,31	0,31	0,31	0,3
Zona 6	0,34	0,34	0,34	0,34	0,34	0,35	0,35	0,35	0,3
Zona 7	0,37	0,37	0,37	0,37	0,37	0,37	0,37	0,38	0,3
Zona 8	0,36	0,36	0,36	0,37	0,37	0,37	0,38	0,38	0,3
Zona 9	0,36	0,36	0,36	0,36	0,36	0,37	0,37	0,37	0,3
Zona 10	0,48	0,48	0,48	0,48	0,48	0,49	0,49	0,49	0,4
Zona 11	0,32	0,32	0,32	0,32	0,32	0,33	0,33	0,33	0,3
Zona 12	0,36	0,36	0,36	0,36	0,36	0,37	0,37	0,37	0,3

Frequência acumulada dividida pelo total da zona

Figura 7.9 - Parte da planilha eletrônica do Microsoft Office Excel 2003 contendo a MATRIZ P-PREVISÃO

128 padrões de viagens encadeadas em ordem alfabética									
	H	HAAHPPP445	HAAHPPP555	HAHNN12	HAHNN15	HAHNN22	HAHNN23	HAHNN24	HAHN
Zona 1	0,42	0,42	0,42	0,43	0,43	0,44	0,45	0,45	0,4
Zona 2	0,41	0,41	0,41	0,41	0,41	0,42	0,44	0,44	0,4
Zona 3	0,34	0,34	0,34	0,34	0,34	0,35	0,35	0,35	0,3
Zona 4	0,50	0,50	0,50	0,50	0,50	0,50	0,53	0,53	0,5
Zona 5	0,24	0,24	0,24	0,24	0,24	0,26	0,26	0,26	0,2
Zona 6	0,32	0,32	0,32	0,32	0,32	0,33	0,33	0,33	0,3
Zona 7	0,45	0,45	0,45	0,45	0,45	0,45	0,46	0,46	0,4
Zona 8	0,43	0,43	0,43	0,43	0,43	0,44	0,44	0,44	0,4
Zona 9	0,43	0,43	0,43	0,43	0,43	0,44	0,44	0,44	0,4
Zona 10	0,58	0,58	0,58	0,58	0,58	0,59	0,60	0,60	0,6
Zona 11	0,32	0,32	0,32	0,32	0,32	0,32	0,32	0,32	0,3
Zona 12	0,44	0,44	0,44	0,44	0,44	0,44	0,44	0,44	0,4

Frequência acumulada dividida pelo total da zona

Figura 7.10 - Parte da planilha eletrônica do Microsoft Office Excel 2003 contendo a MATRIZ F-REAL

A última matriz, denominada neste trabalho de MATRIZ DESVIO (Figura 7.11), é o resultado da diferença em módulo de cada célula da MATRIZ P-PREVISÃO pela célula na MATRIZ P-REAL. Cada célula da matriz desvio representa o desvio comparativo entre as duas distribuições (D_{ij}), o maior valor de D_{ij} da zona de tráfego que está sendo analisada ($D_{\max}$) é comparado com o valor de desvio crítico ($D_{\text{crítico}}$), calculado pela equação (2.3). Cada zona de tráfego terá um $D_{\text{crítico}}$ específico que será comparado ao $D_{\max}$ da zona, portanto são realizadas 361 comparações.

128 padrões de viagens encadeadas em ordem alfabética

	H	HAAHPPP445	HAAHPPP555	HAHNN12	HAHNN15	HAHNN22	HAHNN23	HAHNN24	HAHI
Zona 1	0,06	0,06	0,06	0,07	0,06	0,07	0,07	0,07	0,1
Zona 2	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,1
Zona 3	0,06	0,06	0,06	0,06	0,06	0,06	0,06	0,06	0,1
Zona 4	0,17	0,17	0,17	0,17	0,17	0,16	0,19	0,19	0,1
Zona 5	0,06	0,06	0,06	0,06	0,06	0,05	0,05	0,06	0,1
Zona 6	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,1
Zona 7	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,08	0,1
Zona 8	0,06	0,06	0,06	0,06	0,06	0,07	0,06	0,06	0,1
Zona 9	0,08	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,1
Zona 10	0,10	0,10	0,10	0,10	0,10	0,10	0,11	0,11	0,1
Zona 11	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,1
Zona 12	0,08	0,08	0,08	0,08	0,08	0,07	0,07	0,07	0,1

Desvios

Figura 7.11 - Parte da planilha eletrônica do Microsoft Office Excel 2003 contendo a MATRIZ DESVIO

Na Tabela 7.1 são apresentados os resultados nas zonas de tráfego onde o valor de $D_{\max}$ foi superior ao $D_{\text{crítico}}$ (hipótese nula é rejeitada) e na Tabela 7.2 os resultados onde o valor de $D_{\max}$ foi inferior ao $D_{\text{crítico}}$ (hipótese nula é comprovada).

Tabela 7.1 - Zonas de tráfego que apresentaram valores de $D_{\max}$ maiores que o $D_{\text{crítico}}$

Zona	Quant.	$D_{\max}$	$D_{\text{crítico}}$	Zona	Quant.	$D_{\max}$	$D_{\text{crítico}}$	Zona	Quant.	$D_{\max}$	$D_{\text{crítico}}$	Zona	Quant.	$D_{\max}$	$D_{\text{crítico}}$
1	111	0,2021	0,1826	72	121	0,1997	0,1748	178	252	0,1602	0,1212	311	400	0,1181	0,0962
16	209	0,1398	0,1330	73	79	0,2624	0,2164	196	150	0,1630	0,1570	315	547	0,0961	0,0822
17	237	0,1838	0,1249	83	233	0,1608	0,1260	200	150	0,1572	0,1570	322	219	0,1490	0,1300
31	347	0,1404	0,1032	116	90	0,2553	0,2027	220	206	0,1445	0,1340	326	226	0,1813	0,1279
32	169	0,2454	0,1479	120	143	0,3092	0,1608	221	138	0,2020	0,1637	331	516	0,1096	0,0847
33	65	0,3057	0,2386	123	63	0,3530	0,2423	222	202	0,1527	0,1353	332	267	0,1199	0,1177
34	288	0,1583	0,1133	124	131	0,1771	0,1680	251	270	0,1213	0,1171	344	194	0,1958	0,1381
36	73	0,2623	0,2251	126	63	0,3697	0,2423	254	289	0,1543	0,1131	347	233	0,1723	0,1260
37	100	0,1991	0,1923	127	125	0,2012	0,1720	258	257	0,1503	0,1200	369	117	0,2284	0,1778
64	207	0,1514	0,1337	128	148	0,1868	0,1581	269	298	0,2065	0,1114	370	200	0,1784	0,1360
66	111	0,2680	0,1826	156	221	0,1592	0,1294	276	245	0,2248	0,1229	372	256	0,1458	0,1202
67	116	0,2166	0,1786	158	289	0,1562	0,1131	281	493	0,1099	0,0866	377	213	0,2592	0,1318
68	124	0,2021	0,1727	161	244	0,1317	0,1231	290	413	0,1643	0,0946	382	207	0,2107	0,1337

Tabela 7.2 - Zonas de tráfego que apresentaram valores de $D_{\max}$ menores que o $D_{\text{crítico}}$

Zona	Quant.	D _{max}	D _{critico}	Zona	Quant.	D _{max}	D _{critico}	Zona	Quant.	D _{max}	D _{critico}	Zona	Quant.	D _{max}	D _{critico}
2	114	0,1756	0,1801	94	237	0,0499	0,1249	184	358	0,0342	0,1017	280	191	0,0877	0,1392
3	95	0,0992	0,1973	95	145	0,0608	0,1597	185	292	0,0563	0,1126	282	369	0,0772	0,1001
4	60	0,1952	0,2483	96	266	0,0246	0,1179	186	153	0,0455	0,1555	283	193	0,0773	0,1384
5	66	0,1003	0,2367	97	439	0,0523	0,0918	187	111	0,0563	0,1826	284	200	0,1260	0,1360
6	84	0,1159	0,2099	98	185	0,0681	0,1414	188	132	0,0995	0,1674	285	357	0,0721	0,1018
7	156	0,1532	0,1540	99	161	0,1012	0,1516	189	92	0,1285	0,2005	286	217	0,1058	0,1306
8	105	0,1013	0,1877	100	265	0,0704	0,1181	190	213	0,1127	0,1318	287	692	0,0659	0,0731
9	115	0,0754	0,1794	101	109	0,0945	0,1842	191	108	0,1286	0,1851	288	164	0,0645	0,1502
10	112	0,1081	0,1817	102	139	0,1220	0,1631	192	179	0,0500	0,1438	289	212	0,0599	0,1321
11	232	0,0483	0,1263	103	587	0,0310	0,0794	193	118	0,1063	0,1771	291	378	0,0660	0,0989
12	188	0,0838	0,1403	104	132	0,0589	0,1674	194	148	0,1396	0,1581	292	300	0,0490	0,1110
13	152	0,0673	0,1560	105	114	0,0862	0,1801	195	120	0,1364	0,1756	293	292	0,0775	0,1126
14	186	0,0669	0,1410	106	434	0,0858	0,0923	197	76	0,0873	0,2206	294	370	0,0355	0,1000
15	184	0,1207	0,1418	107	198	0,1315	0,1367	198	131	0,0577	0,1680	295	144	0,0568	0,1603
18	205	0,0974	0,1343	109	372	0,0430	0,0997	199	122	0,1267	0,1741	296	210	0,1170	0,1327
19	301	0,0758	0,1109	110	142	0,0725	0,1614	201	425	0,0640	0,0933	297	694	0,0508	0,0730
20	139	0,1539	0,1631	111	204	0,1301	0,1347	202	171	0,1302	0,1471	298	318	0,0997	0,1079
21	249	0,0770	0,1219	113	188	0,0627	0,1403	203	230	0,0739	0,1268	299	324	0,0929	0,1069
22	135	0,0951	0,1655	114	140	0,1580	0,1626	204	161	0,0912	0,1516	300	438	0,0653	0,0919
23	220	0,0919	0,1297	115	367	0,0422	0,1004	205	233	0,0615	0,1260	301	355	0,0903	0,1021
24	175	0,0426	0,1454	117	155	0,1149	0,1545	206	323	0,0785	0,1070	302	386	0,0474	0,0979
25	228	0,0553	0,1274	118	119	0,1248	0,1763	207	150	0,0541	0,1570	303	267	0,0446	0,1177
26	192	0,0604	0,1388	119	79	0,0879	0,2164	208	141	0,1246	0,1620	304	199	0,0801	0,1363
27	258	0,1022	0,1197	121	108	0,1335	0,1851	209	439	0,0613	0,0918	305	146	0,0916	0,1592
28	126	0,0906	0,1713	122	81	0,1350	0,2137	210	252	0,0965	0,1212	306	245	0,0641	0,1229
29	243	0,1091	0,1234	125	52	0,1869	0,2667	215	270	0,0350	0,1171	307	276	0,0523	0,1158
30	148	0,0507	0,1581	129	107	0,0733	0,1859	216	237	0,0665	0,1249	308	396	0,0665	0,0967
35	105	0,1319	0,1877	130	64	0,0753	0,2404	217	251	0,0302	0,1214	309	320	0,0787	0,1075
38	177	0,0882	0,1446	131	69	0,0629	0,2315	218	222	0,0728	0,1291	310	264	0,0878	0,1184
39	146	0,1143	0,1592	132	71	0,1100	0,2283	219	490	0,0548	0,0869	312	168	0,1428	0,1484
40	195	0,0446	0,1377	133	174	0,0931	0,1458	223	418	0,0706	0,0941	313	213	0,0814	0,1318
41	201	0,0871	0,1357	134	132	0,0747	0,1674	224	260	0,0890	0,1193	314	100	0,0857	0,1923
42	123	0,0988	0,1734	135	93	0,1002	0,1994	225	244	0,0477	0,1231	316	210	0,0573	0,1327
43	44	0,0952	0,2900	136	96	0,1037	0,1963	226	137	0,0732	0,1643	317	263	0,0815	0,1186
44	133	0,0426	0,1668	137	103	0,0760	0,1895	227	356	0,0683	0,1019	318	117	0,1543	0,1778
45	180	0,0529	0,1434	138	220	0,0732	0,1297	228	322	0,0969	0,1072	319	253	0,0742	0,1209
46	193	0,0555	0,1384	139	106	0,1057	0,1868	229	247	0,0630	0,1224	321	195	0,0453	0,1377
47	228	0,0272	0,1274	140	276	0,0615	0,1158	230	253	0,1107	0,1209	324	284	0,0571	0,1141
48	185	0,0634	0,1414	141	194	0,0544	0,1381	231	229	0,0927	0,1271	325	499	0,0751	0,0861
49	480	0,0603	0,0878	142	88	0,0831	0,2050	232	186	0,0418	0,1410	328	273	0,0728	0,1164
50	339	0,0720	0,1045	143	264	0,0512	0,1184	233	480	0,0451	0,0878	329	356	0,0711	0,1019
51	156	0,1185	0,1540	144	187	0,0343	0,1406	234	119	0,0614	0,1763	333	288	0,0387	0,1133
52	221	0,0808	0,1294	145	571	0,0402	0,0805	235	197	0,0782	0,1370	334	285	0,0497	0,1139
53	325	0,0418	0,1067	146	226	0,0963	0,1279	236	167	0,0951	0,1488	335	293	0,1069	0,1124
54	231	0,1064	0,1265	147	262	0,0836	0,1188	237	416	0,0747	0,0943	336	227	0,1188	0,1277
55	228	0,0627	0,1274	148	221	0,0328	0,1294	238	121	0,0491	0,1748	337	211	0,0628	0,1324
56	193	0,1377	0,1384	149	162	0,1269	0,1511	239	223	0,1222	0,1288	338	233	0,0804	0,1260
57	193	0,1056	0,1384	150	103	0,1004	0,1895	240	377	0,0918	0,0991	339	203	0,0978	0,1350
58	179	0,0626	0,1438	151	147	0,0533	0,1586	241	399	0,0402	0,0963	340	88	0,0703	0,2050
59	149	0,0654	0,1576	152	127	0,0546	0,1707	242	243	0,0994	0,1234	341	391	0,0876	0,0973
60	193	0,0464	0,1384	153	310	0,0994	0,1092	243	238	0,0678	0,1247	342	217	0,0871	0,1306
61	267	0,0459	0,1177	154	131	0,0457	0,1680	244	283	0,0299	0,1143	343	134	0,0614	0,1662
62	194	0,0906	0,1381	155	359	0,0462	0,1015	245	235	0,0595	0,1255	345	328	0,0604	0,1062
63	170	0,0260	0,1475	157	379	0,0787	0,0988	246	196	0,0562	0,1374	346	180	0,0564	0,1434
65	116	0,0457	0,1786	159	239	0,0854	0,1244	247	226	0,0430	0,1279	348	440	0,0602	0,0917
69	83	0,1577	0,2111	160	328	0,0723	0,1062	248	271	0,0589	0,1168	349	186	0,1270	0,1410
70	55	0,1482	0,2593	162	251	0,1000	0,1214	249	207	0,0685	0,1337	350	208	0,0890	0,1334
71	126	0,1598	0,1713	163	155	0,0443	0,1545	250	146	0,0852	0,1592	351	419	0,0721	0,0940
74	335	0,0973	0,1051	164	291	0,0716	0,1127	252	272	0,0593	0,1166	355	352	0,0424	0,1025
75	135	0,0826	0,1655	165	232	0,0463	0,1263	253	169	0,1324	0,1479	356	427	0,0350	0,0931
76	99	0,0794	0,1933	166	150	0,0425	0,1570	255	229	0,1073	0,1271	357	105	0,1244	0,1877
77	147	0,0875	0,1586	167	265	0,0521	0,1181	256	122	0,0518	0,1741	358	204	0,1069	0,1347
78	147	0,0443	0,1586	168	151	0,0496	0,1565	257	195	0,0399	0,1377	361	428	0,0468	0,0930
79	72	0,0809	0,2267	169	217	0,0662	0,1306	259	164	0,0554	0,1502	368	213	0,1094	0,1318
80	177	0,0475	0,1446	170	155	0,0249	0,1545	260	132	0,1115	0,1674	371	336	0,0348	0,1049
81	150	0,0868	0,1570	171	176	0,0431	0,1450	261	523	0,0316	0,0841	373	222	0,0625	0,1291
82	168	0,1057	0,1484	172	138	0,0447	0,1637	262	268	0,0417	0,1175	374	151	0,0963	0,1565
84	68	0,1148	0,2332	173	279	0,0463	0,1151	263	498	0,0270	0,0862	375	199	0,0965	0,1363
85	167	0,0777	0,1488	174	234	0,0779	0,1257	264	249	0,0746	0,1219	376	244	0,1204	0,1231
86	137	0,0527	0,1643	175	277	0,0516	0,1156	265	75	0,0893	0,2221	378	139	0,1078	0,1631
87	218	0,0904	0,1303	176	326	0,0890	0,1065	266	378	0,0358	0,0989	379	256	0,0459	0,1202
88	391	0,0488	0,0973	177	184	0,0953	0,1418	267	207	0,0605	0,1337	380	155	0,0581	0,1545
89	148	0,0723	0,1581	179	274	0,0500	0,1162	270	324	0,0621	0,1069	383	402	0,0349	0,0959
90	276	0,0722	0,1158	180	332	0,0453	0,1056	271	168	0,1130	0,1484	384	137	0,1487	0,1643
91	142	0,0898	0,1614	181	313	0,0330	0,1087	272	141	0,0314	0,1620	385	234	0,0997	0,1257
92	248	0,0409	0,1221	182	307	0,1038	0,1098	278	236	0,0615	0,1252	387	87	0,0862	0,2062
93	134	0,0762	0,1662	183	289	0,0524	0,1131	279	296	0,0427	0,1118	388	134	0,1002	0,1662
												389	122	0,1689	0,1741

Das 361 zonas que foram realizadas a previsão, em 309 zonas o critério utilizado para validação da hipótese foi aceito e em 52 zonas a hipótese nula foi rejeitada. Estas zonas rejeitadas representam 11.065 indivíduos, 13,86% do total de 79.839 indivíduos.

Em 8 zonas de tráfego o $D_{\text{máximo}}$ foi superior ao $D_{\text{crítico}}$ com uma diferença inferior a 0,01 contabilizando um total de 1.592 indivíduo (1,99%), restando 44 zonas com 9.473 indivíduos (11,87%). Em 47 zonas o $D_{\text{máximo}}$ foi superior ao $D_{\text{crítico}}$ com uma diferença inferior a 0,1 totalizando 10.338 indivíduos nas zonas (12,95%), restando 5 zonas com 727 indivíduos (0,91%) com uma diferença maior que 0,1.

Uma das razões para a diferença entre a prova de K-S aplicada nas amostras contendo 139 padrões e nas amostras contendo 117 padrões está justamente na diferença da quantidade de padrões. Quando é realizada a previsão para 1997 com 139 padrões, são previstos 651 indivíduos que não estarão enquadrados em nenhum dos padrões de 1997, da mesma forma em 1997 há 707 indivíduos para os quais não será possível realizar a previsão do padrão de encadeamento de viagem.

Caso fosse eliminando os 22 padrões incompatíveis e os indivíduos que desempenham estes padrões das matrizes e aplicando a prova de K-S, a hipótese nula é comprovada em 344 zonas sendo rejeitada em 17 zonas, representando 2.737 indivíduos (3,43%).

Contudo vale ressaltar que este procedimento de eliminar padrões incompatíveis é incorreto e é de consenso que no cenário futuro haverá ocorrências incapazes de serem previstas dada à dinâmica de interação entre aspectos políticos, econômicos e sociais e seus reflexos nas atividades desempenhadas pelos indivíduos.

Todavia, se houvesse a possibilidade de o programa disponibilizar um maior número de variáveis respostas (a limitação é de 128 variáveis dependentes) alguns dos padrões incompatíveis poderiam ser contemplados e um melhor resultado ser obtido.

O fenômeno subjacente aos padrões incompatíveis é explicado da seguinte forma. No banco de dados sintético de 1987 há um grupo de 128 padrões que foram selecionados por serem os mais frequentes e há outro grupo que, por uma limitação do programa, não foi incorporado ao modelo.

Devido a uma flutuação ocorrida nos tipos de atividades, modos utilizados e períodos em que são realizadas e até mesmo pelas peculiaridades que há na amostragem, em 1997, 11 padrões que faziam parte do grupo dos 128 padrões de 1987 sofrem uma redução da frequência e outros 11 padrões que ficaram fora do grupo de 128 de 1987 passam a integrar o conjunto dos 128 padrões mais frequentes, como apresentado na Tabela 7.3.

Tabela 7.3 – Frequência de ocorrência dos padrões incompatíveis

Padrão	Ano de presença no grupo dos 128 padrões	Frequência	
		1987	1997
<i>HAAHPPP555</i>	1987	46	41
<i>HAHNN15</i>	1987	50	45
<i>HAHPP35</i>	1997	28	59
<i>HSNHN33</i>	1997	40	51
<i>HSHPN12</i>	1987	68	33
<i>HSHP33</i>	1997	23	47
<i>HSNHN1133</i>	1997	34	75
<i>HSNHN113455</i>	1997	37	89
<i>HSHTT25</i>	1997	46	48
<i>HWAHTTT145</i>	1987	52	26
<i>HWAHTTT255</i>	1987	55	28
<i>HWAHTNNT1335</i>	1997	2	72
<i>HWAHTTNN1555</i>	1987	61	23
<i>HWAHTTPP1555</i>	1987	76	20
<i>HWHP12</i>	1987	44	32
<i>HWHP55</i>	1997	23	52
<i>HWHSNNTT1555</i>	1987	62	36
<i>HWHSHTTNN1455</i>	1997	43	46
<i>HWHSHTTPP1555</i>	1987	61	21
<i>HWHSHTTTT1455</i>	1987	56	32
<i>HWHTT55</i>	1997	40	58
<i>HWSHPPP255</i>	1997	26	63

Pela Tabela 7.3 observa-se que os padrões não contemplados em 1997 sofreram redução de sua ocorrência para abaixo da linha de corte dos 128 em 1987 (46 indivíduos) e que os não contemplados em 1987 estavam abaixo da linha de corte dos 128 padrões e tiveram frequências superiores a linha de corte em 1997 (46 indivíduos).

Foi verificado que nas zonas onde a hipótese nula foi rejeitada, a mediana dos 11 padrões incompatíveis (que ocorrem em 1997 e não são contemplados em 1987) é de 11,2% e média de 15,4% (Tabela 7.4).

Tabela 7.4 - Representatividade dos padrões que ocorrem em 1997 e não contemplados em 1987

Padrões	Indivíduos nas 361 zonas de tráfego	Ocorrência nas 52 zonas onde a hipótese foi rejeitada	Percentual (%)
<i>HAHPP35</i>	59	13	22,0%
<i>HSHNN33</i>	51	2	3,9%
<i>HSHPP33</i>	47	10	21,3%
<i>HSHSHNNNN1133</i>	75	6	8,0%
<i>HSHSHNNNN3455</i>	89	10	11,2%
<i>HSHTT25</i>	48	13	27,1%
<i>HWAWHTNNT1335</i>	72	6	8,3%
<i>HWHPP55</i>	52	14	26,9%
<i>HWHSHTTNN1455</i>	46	3	6,5%
<i>HWHTT55</i>	58	4	6,9%
<i>HWSHPPP255</i>	63	17	27,0%

Observa-se em 7 zonas de tráfego (valores em destaque na Tabela 7.5) que a ocorrência dos padrões incompatíveis (que não são contemplados em 1987) em termos percentuais, foi próximo a amplitude em relação ao $D_{\text{crítico}}$ da zona, reforçando a idéia de que se houvesse um maior número de variáveis dependentes no programa, o número de zonas em que a hipótese nula é rejeitada seria menor do que o atual número de 52.

Tabela 7.5 – Ocorrência dos padrões incompatíveis nas zonas onde a hipótese foi rejeitada

Zona	Quant.	D _{máximo}	D _{crítico}	Diferença	Ocorrência	%	Zona	Quant.	D _{máximo}	D _{crítico}	Diferença	Ocorrência	%
1	111	0,20	0,18	0,02	1	0,90%	161	252	0,16	0,12	0,04	1	0,40%
76	209	0,14	0,13	0,01	3	1,44%	178	150	0,16	0,16	0,01	1	0,67%
17	237	0,18	0,12	0,06	4	1,69%	196	150	0,16	0,16	0,00	1	0,67%
31	347	0,14	0,10	0,04	7	2,02%	200	206	0,14	0,13	0,01	0	0,00%
32	169	0,25	0,15	0,10	2	1,18%	220	138	0,20	0,16	0,04	3	2,17%
33	65	0,31	0,24	0,07	0	0,00%	227	202	0,15	0,14	0,02	3	1,49%
34	288	0,16	0,11	0,04	5	1,74%	222	270	0,12	0,12	0,00	2	0,74%
36	73	0,26	0,23	0,04	1	1,37%	251	289	0,15	0,11	0,04	0	0,00%
37	100	0,20	0,19	0,01	0	0,00%	254	257	0,15	0,12	0,03	1	0,39%
64	207	0,15	0,13	0,02	8	3,86%	258	298	0,21	0,11	0,10	1	0,34%
65	111	0,27	0,18	0,09	0	0,00%	269	245	0,22	0,12	0,10	3	1,22%
66	116	0,22	0,18	0,04	4	3,45%	281	493	0,11	0,09	0,02	1	0,20%
67	124	0,20	0,17	0,03	1	0,81%	290	413	0,16	0,09	0,07	1	0,24%
68	121	0,20	0,17	0,02	1	0,83%	311	400	0,12	0,10	0,02	3	0,75%
72	79	0,26	0,22	0,05	4	5,06%	315	547	0,10	0,08	0,01	0	0,00%
73	233	0,16	0,13	0,03	0	0,00%	322	219	0,15	0,13	0,02	2	0,91%
83	90	0,26	0,20	0,05	2	2,22%	326	226	0,18	0,13	0,05	2	0,88%
116	143	0,31	0,16	0,15	0	0,00%	331	516	0,11	0,08	0,02	3	0,58%
120	63	0,35	0,24	0,11	1	1,59%	332	267	0,12	0,12	0,00	1	0,37%
123	131	0,18	0,17	0,01	3	2,29%	344	194	0,20	0,14	0,06	2	1,03%
124	63	0,37	0,24	0,13	2	3,17%	347	233	0,17	0,13	0,05	1	0,43%
126	125	0,20	0,17	0,03	1	0,80%	369	117	0,23	0,18	0,05	3	2,56%
127	148	0,19	0,16	0,03	0	0,00%	370	200	0,18	0,14	0,04	1	0,50%
128	221	0,16	0,13	0,03	1	0,45%	372	256	0,15	0,12	0,03	4	1,56%
156	289	0,16	0,11	0,04	1	0,35%	377	213	0,26	0,13	0,13	2	0,94%
158	244	0,13	0,12	0,01	1	0,41%	382	207	0,21	0,13	0,08	3	1,45%

7.4.2 Sequência aleatória dos padrões

Utilizando a função *Aleatório* do Microsoft Office Excel 2003 foi atribuído a cada um dos padrões de viagem encadeada um número compreendido entre 0 e 1 e em seguida foi realizada a classificação dos padrões em ordem crescente tendo como base os valores dos números atribuídos.

As etapas seguintes que são a elaboração das matrizes PREVISÃO, REAL, F-PREVISÃO, F-REAL, P-PREVISÃO, P-REAL e DESVIO seguem a forma descrita para a sequência em ordem alfabética dos padrões de encadeamento de viagens.

Conforme mencionado no capítulo de método, onde foi proposta a realização de cinquenta testes adicionais, foram desempenhados 50 testes da prova de K-S com dife-

rentes ordenações dos padrões de viagem encadeada. Os resultados obtidos são apresentados na Tabela 7.6.

Tabela 7.6 - Resultado da prova de k-S em testes com ordem aleatória

Número do Teste	Quantidade de Zonas em que a Hipótese foi Rejeitada	Número do Teste	Quantidade de Zonas em que a Hipótese foi Rejeitada
1	26	26	9
2	45	27	39
3	15	28	10
4	18	29	19
5	9	30	35
6	21	31	45
7	22	32	31
8	55	33	13
9	9	34	32
10	21	35	36
11	13	36	39
12	21	37	10
13	60	38	51
14	22	39	40
15	24	40	20
16	25	41	29
17	20	42	45
18	51	43	31
19	10	44	17
20	24	45	8
21	30	46	53
22	18	47	18
23	70	48	43
24	29	49	51
25	12	50	17

Calculando a média das quantidades de zonas em que a hipótese nula é rejeitada obtém-se o valor de 28,22 e desvio padrão de 15,45. No gráfico da Figura 7.12 é apresentada a distribuição da frequência das zonas em que $D_{\max}$ foi maior que o $D_{\text{crítico}}$.

Figura 7.12 - Gráfico da distribuição da quantidade de zonas em que a hipótese nula é rejeitada

A primeira análise inferida sobre o teste com sequência em ordem aleatória é sobre o resultado do teste em ordem alfabética. A frequência de 52 zonas rejeitadas está pelo gráfico da Figura 7.12, na região em 14% dos resultados foram rejeitados, ou seja, a probabilidade de se realizar um teste com a prova de K-S com os resultados dispostos em ordem alfabética e que o resultado de que hipótese nula seja rejeitada em mais de 50 zonas é de 14%.

Sendo assim, dada conjuminância dos desvios na ordem alfabética dos padrões pode-se afirmar, que com os dados utilizados e resultados obtidos neste trabalho, esta ordem tende a ser mais severa e rígida quando à aplicabilidade da prova de K-S.

Segundo, analisando os resultados dos testes realizados em ordem aleatória, observa-se que em média 50% destes resultados se enquadram no intervalo com limite superior de 28 zonas rejeitadas de 361 no total (7,76% da amostra em média).

7.5 Considerações Adicionais

Conforme mencionado no capítulo introdutório deste trabalho e no início deste capítulo, um dos problemas propostos que é a idéia diretriz do trabalho foi: “é possível utilizar uma técnica de modelagem capaz de realizar previsões de padrões de viagens encadeadas no meio urbano, que contemplem estas mudanças comportamentais dos indivíduos?”.

A resposta a este segundo problema veio com a aplicação do processo analítico de mineração de dados aos bancos de dados de 1987 e 1997. Com o uso desta ferramenta foi possível gerar um modelo que permitiu a previsão de padrões de viagens encadeadas para um ano horizonte de dez anos.

Os resultados da previsão foram comparados com os dados reais através da prova estatística de Kolmogorov-Smirnov. Pela aplicação da prova de K-S com a ordem alfabética dos padrões obteve-se o resultado de que em 52 zonas de tráfego (14,4%) a hipótese nula é rejeitada e em 309 zonas ela é confirmada (85,6%). Realizando cinquenta testes adicionais variando aleatoriamente a ordem dos padrões foi observado que em média 28,22 zonas a hipótese nula é rejeitada.

Foi observado que se houvesse, no programa de mineração de dados, a possibilidade de inserção de um maior número de variáveis respostas, a quantidade de zonas onde a hipótese é comprovada seria maior do que as aqui encontradas.

Capítulo 8

Conclusões e Recomendações

8.1 Introdução

Juntamente com os trabalhos de Ichikawa (2002), Pitombo (2003), Souza (2004) e Aguiar (2005) esta pesquisa faz parte de uma das linhas de pesquisa da Escola de Engenharia de São Carlos (EESC) da Universidade de São Paulo (USP) que propõe a abordagem de atividades para o estudo do comportamento de viagens urbanas com a utilização da técnica de mineração de dados com o modelo de árvore de decisão.

Ichikawa (2002) demonstrou a aplicabilidade do minerador de dados para obter relações entre padrões de viagens encadeadas e as características socioeconômicas dos viajantes urbanos, utilizando dados da Região Metropolitana de São Paulo coletados na Pesquisa de Origem-Destino de 1987.

Pitombo (2003) estudou a relação entre variáveis socioeconômicas e padrões de viagens encadeadas, aplicando a técnica de mineração nos dados da Pesquisa Origem-Destino de 1987 da Região Metropolitana de São Paulo. Nesta pesquisa foi concluído que há relação entre as variáveis analisadas e os padrões de encadeamento de viagens, indicando

que o minerador de dados pode ter um papel importante na análise do comportamento relacionado às viagens.

Souza (2004) identificou que o comportamento de viajantes urbanos sofre influência de políticas urbanas, características socioeconômicas e espaciais ao comparar os resultados da extração de padrões de bases de informações utilizando a mineração de dados da Pesquisa Origem-Destino de 1997 de Bauru da Pesquisa de Entrevistas Domiciliares de 2000 da Região Metropolitana de Belém e da Pesquisa Origem-Destino de 1997 da Região Metropolitana de São Paulo.

Em Aguiar (2005) foi proposta a possibilidade de reduzir a quantidade de indivíduos na amostra de Pesquisa de Entrevistas Domiciliares, através da mineração de dados utilizando a base de informações da Região Metropolitana de Belém.

A proposta deste trabalho, inserida a esta linha de pesquisa, foi apresentar soluções a dois problemas identificados na área de análise de demanda por transporte:

- É possível elaborar um cenário para um modelo de previsão de demanda por transporte que reflita a mudança no comportamento humano, em um período de tempo?
- É possível utilizar uma técnica de modelagem capaz de realizar previsões de padrões de viagens encadeadas no meio urbano, que contemplem estas mudanças comportamentais dos indivíduos?

Para responder a esses problemas, formulou-se a hipótese que o minerador de dados pode ser utilizado como uma ferramenta para previsão dos padrões de viagens, produzindo o efeito da mudança comportamental quando ela existe.

8.2 Resultados Alcançados

A hipótese formulada foi submetida a um teste de verificação delineado no segundo capítulo deste trabalho. Para realização do experimento foram utilizados os dados disponíveis na Pesquisa Origem-Destino da Região Metropolitana de São Paulo realizada pela Companhia do Metropolitano de São Paulo (METRÔ) nos anos de 1987 e 1997.

Ao analisar estudos da área de demografia e análises dos bancos de dados das Pesquisas de 1987 e 1997 que confirmam tais mudanças de comportamento, foram identificadas mudanças de comportamento em três grupos socioeconômicos com influência nos padrões de viagens encadeadas.

As alterações comportamentais identificadas foram incorporadas ao cenário elaborado a partir de 1987 para projeção da demanda de 1997. Após uma análise de suas características concluiu-se que o cenário construído contemplou as mudanças comportamentais identificadas na sociedade no período de tempo de 1987 a 1997 atendendo ao primeiro problema proposto no trabalho.

A resposta ao segundo problema veio com a aplicação do processo analítico de mineração de dados nos bancos de dados de 1987 e 1997. Com o uso desta ferramenta foi possível gerar um modelo que permitiu a previsão de padrões de viagens encadeadas para um ano horizonte de dez anos, valendo-se apenas de informações socioeconômicas de 1997.

Os resultados da previsão foram comparados com os dados reais através da prova estatística de Kolmogorov-Smirnov de duas formas distintas, dada à peculiaridade desta

prova estatística que é sensível quanto à ordem de disposição dos parâmetros que estão sendo comparados.

Pela aplicação da prova de K-S com a disposição dos padrões na ordem gerada pelo relatório do programa S-Plus (ordem alfabética) obteve-se o resultado de que em 52 zonas de tráfego (14,4%) a hipótese nula é rejeitada e em 309 zonas ela é corroborada (85,6%).

Realizando cinquenta testes da prova de K-S com a disposição dos padrões em ordem aleatória foi obtida uma média de 28,22 zonas de tráfego onde a hipótese nula é rejeitada e em 332,78 zonas em que ela é aceita.

Foi observado que se houvesse no programa de mineração de dados utilizado à possibilidade de inserção de um maior número de variáveis respostas, a quantidade de zonas onde a hipótese é comprovada provavelmente seria maior do que o valor encontrado de 309.

8.3 Considerações sobre o Método Adotado

8.3.1 Quanto à postura científica adotada

Entre as três principais linhas de pensamento sobre o método científico (visão baconiana, visão popperiana e a visão Hempel), foi adotada no trabalho a postura científica de Hempel (1981).

Para este pensamento a investigação científica é indutiva num sentido mais amplo, aceitando hipóteses baseadas em dados que não fornecem para ela evidência con-

clusiva, mas lhe conferem apenas um “suporte indutivo” ou confirmação mais ou menos forte.

Como a hipótese proposta neste trabalho foi baseada em dados amostras da Região Metropolitana de São Paulo (RMSP) com representatividade de 0,78% e 0,59% respectivamente em relação ao total da população, o intuito desta pesquisa não é de creditar à hipótese um caráter de uma crença em uma lei e nem de refutação da hipótese proposta. A pretensão deste trabalho foi a de verificar se a hipótese poderia ser corroborada, o que de fato ocorreu em média em 332,78 das 361 zonas de tráfego da RMSP onde ela foi testada.

Todo o encadeamento lógico da idéia diretriz que subentendeu todo o movimento do projeto de pesquisa foi baseado na busca por corroborações ou refutações da hipótese nas zonas de tráfego onde ela foi testada. Não foi objetivo a busca pela comprovação de uma lei de um caso particular para o geral, como também não foi de somente refutar a hipótese.

8.3.2 Quanto à ferramenta de processamento dos dados

Neste trabalho a proposta foi testar a aplicação da técnica de mineração de dados com o modelo de árvore de decisão como ferramenta para previsão dos padrões de viagens encadeadas. A aplicabilidade desta técnica à análise de demanda por transporte já havia sido confirmada nos trabalhos de Ichikawa (2002), Pitombo (2003), Souza (2004) e Aguiar (2005).

A ferramenta foi aplicada e os resultados obtidos foram satisfatórios, demonstrando que a técnica de mineração de dados com o modelo de árvore de decisão pode ser utilizada para realizar previsões de padrões de encadeamento de viagens urbanas.

Contudo, com relação ao programa que foi utilizado para aplicação da técnica, S-Plus 6.1, há uma limitação em relação à quantidade de variáveis respostas, 128. Esta limitação, provavelmente, impede que a quantidade de zonas onde a hipótese nula é corroborada seja em maior número.

8.3.3 Limpeza dos dados

A opção de limitar o máximo de viagens em quatro é dada em função da limitação do S-Plus (128 variáveis respostas). A inclusão de indivíduos com mais de quatro viagens realizadas aumentaria o número de padrões, que além de considerar o motivo da viagem, contabiliza o modo e o período em que foi realizada a viagem.

Como são excluídos os indivíduos que realizam mais de quatro viagens e de jornadas com origem fora do domicílio, uma desvantagem desta etapa do método é o aumento da proporção das pessoas que realizam o padrão H, distanciando da realidade existente. Uma alternativa para solucionar este problema seria um ajuste proporcional entre os indivíduos que viajam e deixam de viajar.

De fato seria adequado numa análise de demanda de transporte completa que fossem incluídos indivíduos que realizam mais de quatro viagens bem como os indivíduos com jornadas com origem fora do domicílio.

8.3.4 Codificação dos dados

Novamente em função da limitação do programa em 128 variáveis respostas foram limitados em três os atributos da viagem: motivo, modo e período. Outras características da viagem poderiam ser incorporadas como duração da viagem, mas aumentaria a diversidade dos padrões o que reduziria a frequência de indivíduos nos 128 padrões selecionados.

8.3.5 Seleção das variáveis dependentes e independentes

Foi considerado como variável independente os 128 padrões de viagens encadeadas e como variáveis dependentes os atributos socioeconômicos do indivíduo: número de indivíduos na família (FAM), a renda familiar (RFAM), a quantidade de autos (AUTO), sexo (SEXO), situação familiar (SF), idade (IDADE), grau de instrução (GI), estuda (ESTUDA), trabalha (TRAB), aposentado (APOSE), dona de casa (DCASA), assalariado (ASSAL), autônomo (AUT), densidade de auto (DENS AUTO), criança na residência (CRIAN.H) e idoso (VELHO).

A seleção destes atributos foi baseada em estudos das escolas de ciências humanas e dos trabalhos de análise de demanda por transporte com abordagem em atividades, que mostram que o comportamento humano é influenciado diretamente pelas suas características sociais (idade, sexo, situação familiar etc) e características econômicas (posse de automóvel, renda individual, renda familiar etc).

Apesar de citadas na literatura como variáveis que influenciam na demanda por transportes, neste trabalho não foram acrescentados atributos relacionados ao uso e ocupação do solo e de oferta de transporte.

Na mineração dos dados e no âmbito das variáveis independentes utilizadas, o modelo de árvore mostrou que o comportamento do indivíduo representado pelo padrão de viagem encadeada foi influenciado pelas variáveis "TRAB", "ESTUDA", "SF", "DENSEAUTO", "DCASA", "FAM", "IDADE", "RFAM", "GI", "AUT", "SEXO", "AUTO" e "VELHO".

8.3.6 Construção do cenário para projeção da demanda

O método proposto para construção do cenário futuro para projeção da demanda foi baseado em dois princípios. Primeiro, alterações no cenário ocorrem, e segundo, estas alterações provocam mudanças comportamentais.

Estes princípios foram traduzidos na análise das mudanças ocorridas ao comparar os bancos de dados de 1987 (cenário do ano T) e 1997 (cenário do ano $T+\Delta T$). Estas mudanças identificadas foram recriadas no cenário no ano T através do banco de dados sintético de 1987 o que possibilitou a reprodução destas alterações para projeção da demanda de 1997.

Após uma análise das características dos cenários 1987, 1987 sintético e 1997 concluiu-se que o cenário elaborado contemplou as alterações estruturais identificadas na sociedade no período de tempo de 1987 a 1997, mas alguns valores ficaram mais próximos aos do ano de 1987 do que do ano de 1997.

8.3.7 Adaptação dos dados à mudança de comportamento

Como o modelo de árvore de decisão é considerado um modelo não paramétrico uma forma encontrada para se permitir o ajuste do modelo às mudanças foi simulando no banco de dados utilizado para construir o modelo às alterações de comportamento dos in-

divíduos, inserindo indivíduos sintéticos. Esta forma adotada teve como produto um banco de dados sintético que contemplou as alterações futuras na projeção para 1997.

Para o ajuste realizado para as mulheres, quando foi selecionado o conjunto de mulheres, foram incluídos indivíduos do sexo feminino acima de 60 anos, havendo, portanto interferência no primeiro ajuste realizado. Contudo, foi observado que no banco de dados original de 1987 (80.764) a quantidade de indivíduos do sexo feminino, com idade acima dos 60 anos, com situação familiar chefe ou cônjuge e que trabalham é 127.

Em relação ao total os 127 casos que são afetados por duas mudanças comportamentais representam 0,16% em relação ao total do banco de dados, 0,29% em relação a quantidade de 43.130 mulheres, 0,68 em relação ao número de casos de 18.708 mulheres com situação familiar de chefe ou cônjuge, 2,95% em relação ao total de mulheres com situação familiar de chefe ou cônjuge que trabalham e 3,64% em relação a quantidade de mulheres com idade acima de 60 anos.

8.3.8 Construção do modelo e previsão dos padrões

Esta fase correspondeu à aplicação do minerador de dados com o modelo de árvore de decisão ao banco de dados sintético de 1987 (78.412 indivíduos), adotando um desvio de 0,01 e os critérios de 5 e 10 para o número mínimo de observações antes da divisão e o número mínimo de observações nas folhas.

Adotando este desvio e os critérios mínimos de observações antes da divisão e o número mínimo de observações nas folhas, foi possível obter um modelo que fosse ao máximo detalhista, com a divisão dos indivíduos em 110 grupos (folhas) distintos que possibilitou a

classificação dos dados de 1997 aos agrupamentos estabelecidos para a amostra sintética de 1987.

8.3.9 Levantamento dos padrões de viagens encadeadas dos indivíduos

Esta fase correspondeu a simples contagem da frequência de indivíduos por padrão em cada zona de tráfego.

8.3.10 Comparação dos resultados

A escala de comparação de dados adotada neste trabalho foi em nível da zona de tráfego, permitindo, desta forma, analisar o desempenho do modelo de previsão em cada zona. Com relação ao teste a ser aplicado, a idéia inicial foi de que a prova estatística a ser considerada não poderia fazer nenhuma exigência em relação à distribuição dos dados e que proporcionasse uma avaliação rigorosa da diferença entre cada resultado previsto e o levantado no real, sendo estas diferenças avaliadas para um dado nível de significância.

Ao analisar a diferença da frequência acumulativa das amostras, prova de Kolmogorov-Smirnov (K-S) atendeu a idéia inicial do método pretendido para avaliar os valores esperados pela previsão e obtidos no levantamento dos dados reais.

Uma peculiaridade da execução do teste de K-S é a sensibilidade da ordem de disposição dos padrões. Dependendo do arranjo dos padrões poderiam ocorrer a conjunção de padrões que aumente o desvio ou diminua o valor do desvio.

Uma forma encontrada neste trabalho, como tentativa de se obter um melhor desempenho do teste, foi à adoção de duas formas comparativas: a ordem alfabética e a ordem aleatória. Os resultados obtidos no trabalho sugerem que a prova de K-S com os padrões dispostos na ordem alfabética demonstrou ser rigorosa.

8.3.11 Considerações adicionais

Ressalta-se neste trabalho que o processo de amostragem dos elementos que constituem as duas fontes primárias de dados (anterior às etapas do experimento) foi considerado idêntico em ambos e que os dados das amostras não foram expandidos para o conjunto da população pela incerteza quanto à proporcionalidade de atributos individuais em relação ao total da população. Esta incerteza é creditada à dificuldade de obtenção das informações em parte dos domicílios selecionados para entrevista.

8.4 Caráter Científico do Trabalho

Este trabalho é de caráter científico na medida em que documenta, de modo público e controlável, toda a experiência da pesquisa e permite que outros pesquisadores repitam a pesquisa, quer para replicar os experimentos, quer para constatar que os resultados aqui encontrados foram de natureza casual e que não se deveram à intenção do autor, mas a outros fatores que não foram considerados.

8.5 Recomendações a Trabalhos Futuros

Recomenda-se aos trabalhos seguintes nesta linha de pesquisa:

- A utilização de um programa que possibilite a ampliação do limite de 128 variáveis respondidas;
- Consideração dos indivíduos que realizam mais de quatro viagens e que não possuem origem da viagem no domicílio;
- Incorporar ao padrão de viagem encadeada atributos como duração e custo da viagem;
- No atributo modo do padrão de viagem considerar os diferentes tipos de transporte público (ônibus, metrô, etc);
- Na elaboração do cenário para modelagem considerar as características do uso e ocupação do solo e do sistema de transporte, e;
- Considerar estudos da área de ciências humanas a trabalhos de análise de comportamento do indivíduo e sua relação com a demanda por transportes.

Referências

- AGUIAR, M. F. M. (2005) *Possibilidade de redução no tamanho da amostra de pesquisas de entrevistas domiciliares para planejamento de transportes*. 180 p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2005.
- ARENTZE, T. A.; TIMMERMANS, H. J. P. (2004) A learning-based transportation oriented simulation system. *Transportation Research Part B*, vol. 38, p. 613-633.
- ARRUDA, F. S. (2005) *Aplicação de um modelo baseado em atividades para análise da relação uso do solo e transportes no contexto brasileiro*. 145 p. Tese (Doutorado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2005.
- BACON, F. (1979) *Novum organum ou verdadeiras indicações da interpretação da natureza*. Tradução e Notas de José Aluísio Reis de Andrade, 2ª ed. São Paulo: Nova Atlântida.
- BANDURA, A. (1969) *Principles of behavior modification*. New York: Holt, Rinehart & Winston. 677 p.
- BARAT, J. (2001) Transporte e Mobilidade em São Paulo. *Revista dos Transportes Públicos*, nº 93, 4º Trimestre, p. 51-74.
- BEAUD, M. (1996) *Arte da tese: como redigir uma tese de mestrado ou de doutorado, uma monografia ou qualquer outro trabalho universitário*. Tradução: Glória de Carvalho Lins - Rio de Janeiro: Betrand Brasil.
- BELL, J. B. (1996) Application of classification tree to the habitat preference of upland birds. *Journal of Applied Statistics*, v. 23, n. 2-3, p. 349-359.
- BERNSTEIN, S.; LEBOW, R. N.; STEIN, J. G.; WEBER, S. (2000) God gave physics the easy problems: adapting social science to an unpredictable world. *European Journal of International Relations*, vol. 6, n. 1, p. 43-76.
- BEVERIDGE, W. I. B. (1981) *Semente da descoberta científica*. Tradução de S. R. Barreto. São Paulo: T.A. Queiroz Editor Ltda e Editora da Universidade de São Paulo.
- BHAT, C. R.; KOPPELMAN, F. S. (2000) Activity-based travel demand analysis: History results and future directions. In: TRANSPORTATION RESEARCH BOARD ANNUAL MEETING, 79TH, Washington. *Proceedings...* Washington: Transportation Research Board. CD-ROM.
- BLS - BUREAU OF LABOR STATISTICS. Apresenta informações relativas à economia. Disponível em: <<http://www.bls.gov>> . Acesso em 29 Abr. 2006.
- BORGONI, R.; BERRINGTON, A. (2004) *Evaluating a sequential tree-based procedure for multivariate imputation of complex missing data structures*. Disponível em: <<http://www.ccsr.ac.uk/methods/events/attrition/documents/ANR14.pdf>>. Acesso em: 23 set. 2005.
- BOWMAN, J. L. (1998) *The day activity schedule approach to travel demand analysis*. 172 f. Tese - Massachusetts Institute of Technology, Boston, 1998.

- BOWMAN, J. L. (1995) *Activity based travel demand model system with daily activity schedules*. 94 f. Thesis (Master) - Massachusetts Institute of Technology, Boston, 1995.
- BOWMAN, J. L.; BRADLEY, M. A.; SHIFTAN, Y.; LAWTON, T. K.; BEN-AKIVA, M. E. (1998) Demonstration of an activity based model system for Portland. In: WORLD CONFERENCE ON TRANSPORTATION RESEARCH, 8th, Julho 12-17, Antwerp: *Proceedings...* Disponível em: <<http://jbowman.net/papers/bet99.pdf>>. Acesso em: 3 jan. 2006.
- BOWMAN, J. L.; BEN-AKIVA, M. E. (2000) Activity-based disaggregate travel demand system with daily activity schedules. *Transportation Research Part A*, vol. 35, p. 1-28.
- BOWMAN, J. L.; BEN-AKIVA, M. E. (1997) Activity-based travel forecasting. In: ACTIVITY-BASED TRAVEL FORECASTING CONFERENCE, Junho 2-5, New Orleans: *Proceedings...* Arlington: Texas Transportation Institute, pp. 1-33.
- BULIUNG, R. N. (2005) *Activity/travel behaviour research: approaches and findings with identification of research themes and emerging methods*. Disponível em: <<http://www.ccsr.ac.uk/methods/events/attrition/documents/ANR14.pdf>>. Acesso em: 23 set. 2005.
- CARMO, H. C. E. (2004) A teoria dos índices de preços e o sistema de metas de inflação no Brasil. In: PROGRAMA DE SEMINÁRIOS ACADÊMICOS 5^a. FEIRA, São Paulo, 2004: *Documento...* São Paulo: Instituto de Pesquisas Econômicas da Universidade de São Paulo. Disponível em: <<http://www.econ.fea.usp.br/seminarios/artigos2/ipeset04.pdf>>. Acesso em: 22 jun. 2005.
- CHAPIN, F.S. (1974) *Human activity patterns in the city: things people do in time and space*. New York: John Wiley & Sons. 296 p.
- DEUTSCHE BANK RESEARCH (2002) *The demographic challenge*. Deutsche Bank Research - Editor Stefan Schneider, Berlin.
- ECO, U. (1989) *Como se Faz uma Tese*. Rio de Janeiro: Perspectiva.
- ETTEMA, D. (1996) *Activity-based travel demand modeling*. 280 f. Tese - Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 1996.
- FAHEY, L.; RANDALL, R. M. (1998) *Learning from the future: competitive foresight scenarios*. New York: John Wiley & Sons. 464 p.
- FIPE - FUNDAÇÃO INSTITUTO DE PESQUISAS ECONÔMICAS. Índice acumulado IPC-FIPE. Apresenta informações sobre o índice acumulado por período do IPC-FIPE. Disponível em: <http://www.fipe.com.br/indices/ipc_macumulado.asp>. Acesso em: 22 jun. 2005.
- GODET, M.; ROUBELAT, F. (1996) Creating the future: the use and misuse of scenarios. *Long Range Planning*, vol. 29, n. 2, p. 164-171.
- HÄGERSTRAND, T. (1970) What about people in regional science? *Papers of the Regional Science Association*, n. 23, p. 7-23.
- HEIJDEN, K. (1996) *Scenarios: the art of strategic conversation*. Chichester: John Wiley & Sons. 356 p.

HEMPEL, C.G. (1981) *Filosofia da Ciência Natural*. Tradução P.S. Rocha, 3a. Ed. Rio de Janeiro: Zahar Editores. 142 p.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. SIDRA - Sistema IBGE de recuperação de dados. Apresenta informações sobre os recenseamentos da população dos anos 1991, 1996 e 2000. Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em: 22 set. 2005.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (1987) *Estatísticas históricas do Brasil: séries econômicas, demográficas e sociais*. Vol. 3. Rio de Janeiro: Fundação Instituto Brasileiro de geografia e Estatística. 596 p.

ICHIKAWA, S.M. (2002) *Aplicação de minerador de dados na obtenção de relações entre padrões de encadeamento de viagens codificados e características sócio-econômicas*. 2002. 136 p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2002.

JONES, P.; KOPPELMAN, F.; ORFUEIL, J.P. (1990) Activity analysis: state-of-the-art and future directions. In JONES, P. (ed), *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*, Avebury, Aldershot. p. 34-55.

KAHN, H.; WIENER, A. J. (1967) *The year 2000*. A framework for speculation on the next thirty-three years. New York: Macmillan. 452 p.

KANAFANI, A. (1983) *Transportation Demand Analysis*. New York: McGraw-Hill. 320 p.

KAWAMOTO, E. (1987) *Um novo enfoque do processo de escolha em transporte com tratamento baseado na psicofísica multidimensional*. 126 p. Tese (Doutorado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 1987.

KITAMURA, R. (1997) Applications of models of activity behavior for activity based demand forecasting. In: ACTIVITY-BASED TRAVEL FORECASTING CONFERENCE, Junho 2 - 5, New Orleans: *Proceedings...* Arlington: Texas Transportation Institute, p. 119-150.

KOZU, T.; SATO, Y.; INADA, M. (2003) Demographic changes in Japan and their macroeconomic effects. Disponível em: < <http://www.boj.or.jp/en/ronbun/04/data/wp04e06.pdf>> Acesso em: 16 set. 2005.

KUPPAM, A.,R.; PENDYALA, R.,M. (2001) A structural equations analysis of commuters' activity and travel patterns. *Transportation*, vol 1, n. 28, p. 33-54.

METRÔ - Companhia do Metropolitano de São Paulo (1997) Pesquisa origem-destino 1967-1977-1987-1997: Região Metropolitana de São Paulo, 1997. CD-ROM.

MONTALI, L. (2004) Rearranjos Familiares de Inserção, Precarização do Trabalho e Empobrecimento. In: XIV ENCONTRO NACIONAL DE ESTUDOS POPULACIONAIS, Setembro 20-24, Caxambu: *Anais...* Caxambu: ABEP, p. 1-24.

NAGIN, D.; TREMBLAY, R. (2005) What has been learned from group-based trajectory modeling? Examples from physical aggression and other problem behaviors. *Annals of the American Academy of Political and Social Science*, n. 602, pp. 82-117.

- NEUMANN, I.B. ØVERLAND E. F. (2004) International relations and policy planning: the method of perspectivist scenario building. *International Studies Perspectives*, vol. 5, p. 258-277.
- ORTÚZAR, J. D.; WILLUMSEN, L. G. (1997) *Modeling Transport*. Second Edition, Chichester: John Wiley & Sons Ltd. 439p.
- PAS, E.I. (1997) Recent advances in activity-based travel demand modeling. In: ACTIVITY-BASED TRAVEL FORECASTING CONFERENCE, Junho 2 - 5, New Orleans: *Proceedings...* Arlington: Texas Transportation Institute, p. 79-102.
- PITOMBO, C.S. (2003) *Análise do comportamento subjacente ao encadeamento de viagens através do uso de minerador de dados*. 149 p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2003.
- POPPER, K.R. (1975) *Conhecimento objetivo*. Tradução de Milton Amado. Belo Horizonte: Itatiaia; São Paulo, Editora da Universidade de São Paulo.
- QUINLAN, J.R. (1983) Learning efficient classification procedures and their application to chess end-games. *Machine Learning: An Artificial Intelligence Approach*, p. 463-482.
- RAUDENBUSH, S. W. (2005) How do we study "what happens next"? *Annals of the American Academy of Political and Social Science*, n. 602, pp. 131-144.
- RECKER, W. W.; McNALLY, M. G.; ROOT, G. S. (1986a) A model of complex travel behavior: part II - an operation model. *Transportation Research A*, vol. 4, n. 20, p. 319-330.
- RECKER, W. W.; McNALLY, M. G.; ROOT, G. S. (1986b) A model of complex travel behavior: part I - theoretical development. *Transportation Research A*, vol. 4, n. 20, p. 307-318.
- RDC Inc. (1995) *Activity-based modeling system for travel demand forecasting*. DOT-T-96-02. US Department of Transportation and US Environmental Protection Agency, Washington D.C.
- SÃO PAULO (2003) Lei nº 11.605, de 24 de dezembro de 2003. Dispõe sobre o Plano Plurianual para o período de 2004/2007. Disponível em <<http://www.planejamento.sp.gov.br/PlanOrca/pluri.asp>>. Acesso em 23 set. 2005.
- SEADE - SISTEMA ESTADUAL DE ANÁLISE DE DADOS. Informações dos municípios paulistas. Contém séries históricas de variáveis, reunidas em 25 temas, que podem ser pesquisadas por municípios e regiões do Estado de São Paulo. Disponível em <<http://www.seade.gov.br/produtos/imp/index.php>>. Acesso em 26 set. 2005.
- SIEGEL, S. (1977) *Estatística não-paramétrica para ciências do comportamento*. Tradução de Alfredo Alves de Freitas, Rio de Janeiro: Editora McGraw-Hill do Brasil Ltda.
- SKINNER, S. (1965) *Science and human behavior*. New York: Free Press. 461 p.
- SOUZA, P. B. (2004) *Análise comparativa do encadeamento de viagens através de três áreas urbanas*. 130 p. Dissertação (Mestrado) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2004.

VAN DE BILT, K.-A. (2002) *Desenvolvimento e validação de um procedimento de projeção desagregada a um modelo de geração de viagens baseado em análise de segmentação*. 164 p. Tese (Doutorado) - Escola Politécnica, Universidade de São Paulo, São Paulo, 2002.

VAN DE BILT, K.-A. (1997) *Análise de taxas de produção de viagens urbanas utilizando modelagem de segmentação*. 131 p. Dissertação (Mestrado) - Escola Politécnica, Universidade de São Paulo, São Paulo, 1997.

WESTPHAL, C.; BLAXTON, T. (1998) *Data mining solutions: methods and tools for solving real-world problems*. New York: John Wiley & Sons, Inc. 617 p.