UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA CIVIL

Fernando Schultz Peña Rodrigues

MODELOS COMPORTAMENTAIS DESAGREGADOS: UMA ANÁLISE CONCEITUAL

Porto Alegre
dezembro 2012

FERNANDO SCHULTZ PEÑA RODRIGUES

MODELOS COMPORTAMENTAIS DESAGREGADOS: UMA ANÁLISE CONCEITUAL

Trabalho de Diplomação apresentado ao Departamento de Engenharia Civil da Escola de Engenharia da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Engenheiro Civil

Orientador: Luiz Afonso dos Santos Senna

Porto Alegre

dezembro 2012

FERNANDO SCHULTZ PEÑA RODRIGUES

MODELOS COMPORTAMENTAIS DESAGREGADOS: UMA REVISÃO BIBLIOGRÁFICA SOBRE A ESTIMATIVA DE DEMANDA POR TRANSPORTES

Este Trabalho de Diplomação foi julgado adequado como pré-requisito para a obtenção do título de ENGENHEIRO CIVIL e aprovado em sua forma final pelo Professor Orientador e pela Coordenadora da disciplina Trabalho de Diplomação Engenharia Civil II (ENG01040) da Universidade Federal do Rio Grande do Sul.

Porto Alegre, dezembro de 2012.

Prof. Luiz Afonso dos Santos Senna
PhD University of Leeds/UK
Orientador

Profa. Carin Maria Schmitt
Coordenadora

BANCA EXAMINADORA

Prof. Luiz Afonso dos Santos Senna (UFRGS)
PhD University of Leeds/UK

Prof. Fernando Dutra Michel (UFRGS)
Msc pela Pontifícia Universidade Católica do Rio de Janeiro/RJ

Prof. Fernanda David Weber (UFRGS)
Msc pela Universidade Federal do Rio Grande do Sul/RS

Dedico este trabalho aos meus avós Conrado, Elsa, Wilmar e Edith, que contribuíram para formação do meu caráter e me inspiraram na busca por meus objetivos.

AGRADECIMENTOS

Agradeço aos meus pais, Wilmar e Vera, e a meu irmão Eduardo pelo apoio e compreensão que tiveram comigo durante toda essa jornada que aqui termina.

Agradeço ao Professor Orientador Luiz Afonso dos Santos Senna, pelo apoio e disponibilidade durante o período de elaboração do trabalho.

Agradeço à Professora Carin Maria Schmitt pela exemplar dedicação e seriedade com que encara o suas atividades, auxiliando os alunos na elaboração de seus trabalhos.

Agradeço ao Engº André Bresolin Pinto e à Engª Roberta Dal Sasso Meira, pela oportunidade de trabalho que me foi dada e pelo conhecimento a mim transmitido.

Agradeço à Roberta Machado Bertoi e à sua família, pela compreensão e carinho que sempre recebi e continuo recebendo.

Agradeço à minha família e amigos pela compreensão e apoio nos momentos em que estive ausente para dedicar-me à minha graduação.

Se você quer chegar aonde a maioria não chega, faça o que
a maioria não faz.

Bill Gates

RESUMO

Este trabalho versa sobre os modelos comportamentais desagregados e a sua utilização para estimar a demanda por transportes. Primeiramente é apresentada uma contextualização dos problemas de transportes e a tentativa de mitigá-los. Para isso são apresentadas as técnicas de modelagem, que visam representar a realidade de maneira simplificada, utilizando somente as variáveis mais importantes. A partir de uma revisão bibliográfica são apresentadas as variáveis a serem consideradas para a modelagem de um sistema de transportes, destacando a importância de entender as necessidades e desejos dos usuários. Com o objetivo de transformar o sentimento das pessoas em dados mensuráveis, é apresentada a base teórica dos modelos comportamentais desagregados, suas principais características, seus pressupostos e as restrições ao seu uso. Os modelos comportamentais desagregados utilizam as funções de utilidade, que são expressões matemáticas que representam a atratividade de cada modo de transporte disponível para os usuários. A teoria da utilidade aleatória é apresentada como um dos pressupostos mais importantes desses modelos, dividindo as funções de utilidade em uma parte representativa – onde estarão os atributos mais importantes de cada modo de transporte – e uma parte aleatória – que representa os atributos não observados e os erros cometidos durante a obtenção de dados. Para estimar a demanda por cada modo de transporte, são apresentados os modelos Logit Multinomial e Logit Hierárquico. Estes modelos utilizam ferramentas matemáticas que transformam as funções de utilidade em probabilidade de escolha de cada modal. Por fim, são abordadas as técnicas de obtenção de dados para geração desses modelos. São indicadas as pesquisas de preferência revelada e preferência declarada como as principais alternativas para obtenção desses dados.

Palavras-chave: Modelos Comportamentais Desagregados. Funções de Utilidade. Logit Multinomial. Logit Hierárquico. Preferência Declarada. Escolha modal em transportes.

LISTA DE FIGURAS

Figura 1 – Etapas do trabalho..... 18
Figura 2 – Representação de um modelo..... 20
Figura 3 – Exemplo de falácia ecológica..... 24
Figura 4 – Modelo Logit Hierárquico com dois nichos..... 39
Figura 5 – Diagrama de árvore para escolha modal..... 41
Figura 6 – Exemplo de cartão para uma pesquisa de preferência declarada..... 52

LISTA DE QUADROS

Quadro 1 – Atributos e níveis em um estudo experimental..... 49
Quadro 2 – Pesquisa piloto..... 50

LISTA DE TABELAS

Tabela 1 – Exemplo de independência de alternativas irrelevantes..... 37
Tabela 2 – Mudanças nas probabilidades quando uma alternativa é removida..... 40

LISTA DE SIGLAS

IAI.....- Independência de Alternativas Irrelevantes

IID.....- Independente e Identicamente Distribuída

LH.....- Logit Hierárquico

LM.....- Logit Multinomial

MCD.....- Modelos comportamentais desagregados

PR.....- preferência revelada

PD.....- preferência declarada

LISTA DE SÍMBOLOS

P_i.....- probabilidade da alternativa i ser escolhida

f.....- frequência das viagens

d.....- destino das viagens

m.....- modo de transporte utilizado

r.....- rota utilizada

U_i.....- utilidade da alternativa i

$X_1, \dots, X_n$ - atributos do produto ou serviço

$a_1, \dots, a_n$ - coeficientes relativos às funções de utilidade

a_0 .....- constantes específica de modo dos modelos

V_{iq}.....- parte representativa da Utilidade da alternativa i, associada ao indivíduo q

ε_{iq}.....- componente randômica da Utilidade da alternativa i associada ao indivíduo q

S_q.....- Vetor das características do indivíduo q

Z_{iq}.....- Vetor de valores atribuídos à alternativa i pelo indivíduo q

SUMÁRIO

1 INTRODUÇÃO 13
2 DIRETRIZES DA PESQUISA 16
2.1 QUESTÃO DE PESQUISA 16
2.2 OBJETIVOS DA PESQUISA 16
2.2.1 Objetivo Principal 16
2.2.2 Objetivo Secundário 16
2.3 DELIMITAÇÕES 16
2.4 DELINEAMENTO 17
3 ESTIMATIVA DA DEMANDA POR TRANSPORTES 19
3.1 TÉCNICAS DE MODELAGEM 20
3.2 MODELOS DE PREVISÃO DE DEMANDA 21
4 MODELOS COMPORTAMENTAIS DESAGREGADOS 23
4.1 FUNÇÕES DE UTILIDADE 25
4.2 TEORIA DA UTILIDADE ALEATÓRIA..... 27
5 ESTIMATIVA DAS PROBABILIDADES DE ESCOLHA..... 31
5.1 LOGIT MULTINOMIAL 32
5.2 LOGIT HIERÁRQUICO 37
6 TÉCNICAS DE OBTENÇÃO DE DADOS..... 45
6.1 PREFERÊNCIA REVELADA..... 45
6.2 PREFERÊNCIA DECLARADA..... 46
7 CONSIDERAÇÕES FINAIS..... 53
REFERÊNCIAS..... 55

1 INTRODUÇÃO

O aumento populacional e o desenvolvimento urbano dos grandes centros vêm acentuando-se ao longo do tempo. Observou-se, no último século, novas cidades surgindo, pequenas cidades desenvolvendo-se e – devido à forte industrialização do mundo moderno – uma crescente migração da população rural em direção aos centros urbanos. Estes fatores contribuíram para que muitas pessoas disputassem os mesmos espaços, gerando, nas principais cidades e localidades mais desenvolvidas, altas taxas de densidade demográfica.

O desenvolvimento tecnológico e econômico globalizou o acesso a bens de consumo antigamente muito restritos, como, por exemplo, o automóvel. Concomitante ao aumento da frota de veículos, observou-se a falta de planejamento e investimento financeiro no setor de transportes, gerando, assim, a saturação da infraestrutura viária das grandes cidades.

Constantes transformações estão ocorrendo em todos os setores da sociedade, tanto em termos das novas tecnologias que estão surgindo, quanto nos diferentes padrões comportamentais que foram apresentados nos últimos anos. Observa-se que "O mundo, incluindo transportes, está mudando rapidamente. [...] [Entretanto] ainda encontramos muitos dos mesmos problemas de transporte do passado: congestionamentos, poluição, acidentes, déficits financeiros e bolsões de difícil acesso." (ORTÚZAR; WILLUMSEN, 2011, tradução nossa).

A solução destes problemas, ou pelo menos a minimização deles, é um dos grandes desafios que a Engenharia tenta superar. Para enfrentar esta questão os planejadores de tráfego têm utilizado as técnicas de modelagem, que é a representação da realidade de uma maneira simplificada, utilizando apenas as variáveis que sejam consideradas mais importantes para cada situação.

Para representar o sistema viário na forma de um modelo é necessário, além de conhecer a estrutura física disponível e os veículos que podem ser utilizados no transporte de passageiros e cargas, entender as necessidades dos usuários. Tentar estimar o comportamento das pessoas diante de cenários hipotéticos é importante para avaliar a demanda por transportes gerada por elas. Conhecer essa demanda é fundamental para as tomadas de decisões em planejamento de transportes.

Antes de implantar uma nova rodovia ligando dois polos estaduais, ou um novo modal de transporte em uma cidade – por exemplo, o metrô –, é imprescindível estimar de que maneira a população reagiria a esse novo cenário. É preciso saber se os usuários estão dispostos a pagar pela implantação, por exemplo, da rodovia – seja na forma de impostos seja por um sistema de pedágios – ou quanto as pessoas estariam dispostas a pagar para se locomover através de uma viagem de metrô.

Para estimar o comportamento das pessoas, muitos planejadores utilizam as pesquisas de Preferência Declarada, nas quais – após definidas as variáveis mais importantes para o estudo – geram-se cenários hipotéticos com diferentes valores para cada variável, afim de que os entrevistados indiquem sua preferência em cada um deles.

A partir dos dados obtidos nas entrevistas, utilizam-se ferramentas estatísticas para gerar fórmulas matemáticas que atribuam pesos a cada variável do estudo, chamadas funções de Utilidade. No caso da escolha entre uma viagem de ônibus ou de carro, suas funções poderiam, de maneira simplificada, ter como variáveis o tempo e o custo de viagem, gerando, assim, de acordo com os valores atribuídos a eles, um valor de Utilidade para cada modal de transporte, que, na realidade, representa a satisfação do entrevistado ao escolher esta opção.

De posse da Utilidade de cada alternativa, aplicam-se modelos matemáticos que usam estes dados para estimar a probabilidade de um indivíduo escolher determinada alternativa no estudo em questão. Os dois modelos mais utilizados para este fim são o Logit Multinomial e o Logit Hierárquico.

Este trabalho tem como objetivo descrever os pressupostos teóricos utilizados nesses modelos, bem como suas principais características e limitações, e está divido em 7 capítulos.

O capítulo 2 descreve o objetivo do trabalho e identifica a questão de pesquisa, as delimitações e o delineamento do trabalho. O capítulo 3 é reservado à estimativa de demanda por transportes, incluindo as técnicas de modelagem e os modelos de previsão de demanda.

No capítulo 4 são abordados os modelos comportamentais desagregados, suas funções de utilidade e a teoria da utilidade aleatória. O capítulo 5 é destinado à estima das probabilidades de escolha, através dos modelos Logit Multinomial e Logit Hierárquico.

O capítulo 6 explica de que maneira os dados utilizados nos modelos são obtidos. As técnicas abordadas são as de pesquisa de preferência revelada e de preferência declarada. Por fim, no capítulo 7 são feitas as considerações finais a respeito do tema abordado.

2 DIRETRIZES DA PESQUISA

As diretrizes para desenvolvimento do trabalho são descritas nos próximos itens.

2.1 QUESTÃO DE PESQUISA

A questão de pesquisa do trabalho é: considerando-se a bibliografia especializada, quais são as principais características e restrições ao uso dos Modelos Comportamentais Desagregados?

2.2 OBJETIVOS DA PESQUISA

Os objetivos da pesquisa estão classificados em principal e secundário e são descritos a seguir.

2.2.1 Objetivo Principal

O objetivo principal do trabalho é a descrição das principais características e restrições ao uso dos Modelos Comportamentais Desagregados, com base na bibliografia.

2.2.2 Objetivo secundário

O objetivo secundário do trabalho é descrição dos procedimentos utilizados nas Pesquisas de Preferência Declarada.

2.3 DELIMITAÇÕES

O trabalho delimita-se ao estudo dos Modelos Comportamentais Desagregados Logit Multinomial e Logit Hierárquico.

.

2.4 DELINEAMENTO

O trabalho foi realizado através das etapas apresentadas a seguir, que estão representadas na figura 1, e são descritas nos próximos parágrafos:

a) revisão bibliográfica,
- técnicas de modelagem;
- variáveis envolvidas na modelagem de transportes;
- modelos comportamentais desagregados;
- funções de utilidade;
- estimativa das probabilidades de escolha;
- Logit Multinomial;
- Logit Hierárquico;
- técnicas de obtenção de dados;
- pesquisas de preferência declarada;

b) considerações finais.

O trabalho é limitado à realização de uma **pesquisa bibliográfica** sobre as técnicas utilizadas para estimar a demanda por transportes.

Primeiramente é apresentada a utilização das **técnicas de modelagem** para solução de problemas, representando a realidade através das variáveis mais importantes. Para os problemas de transporte são indicadas as **variáveis envolvidas na modelagem de transportes**.

Dentre as variáveis destacadas está o sentimento dos usuários, e para mensurá-los são apresentados os pressupostos teóricos dos **modelos comportamentais desagregados**, bem como suas principais características e limitações. Nesse contexto, o trabalho destaca os principais conceitos teóricos das **funções de utilidade**, que são expressões matemáticas importantes para definir a atratividade de cada modal de transporte.

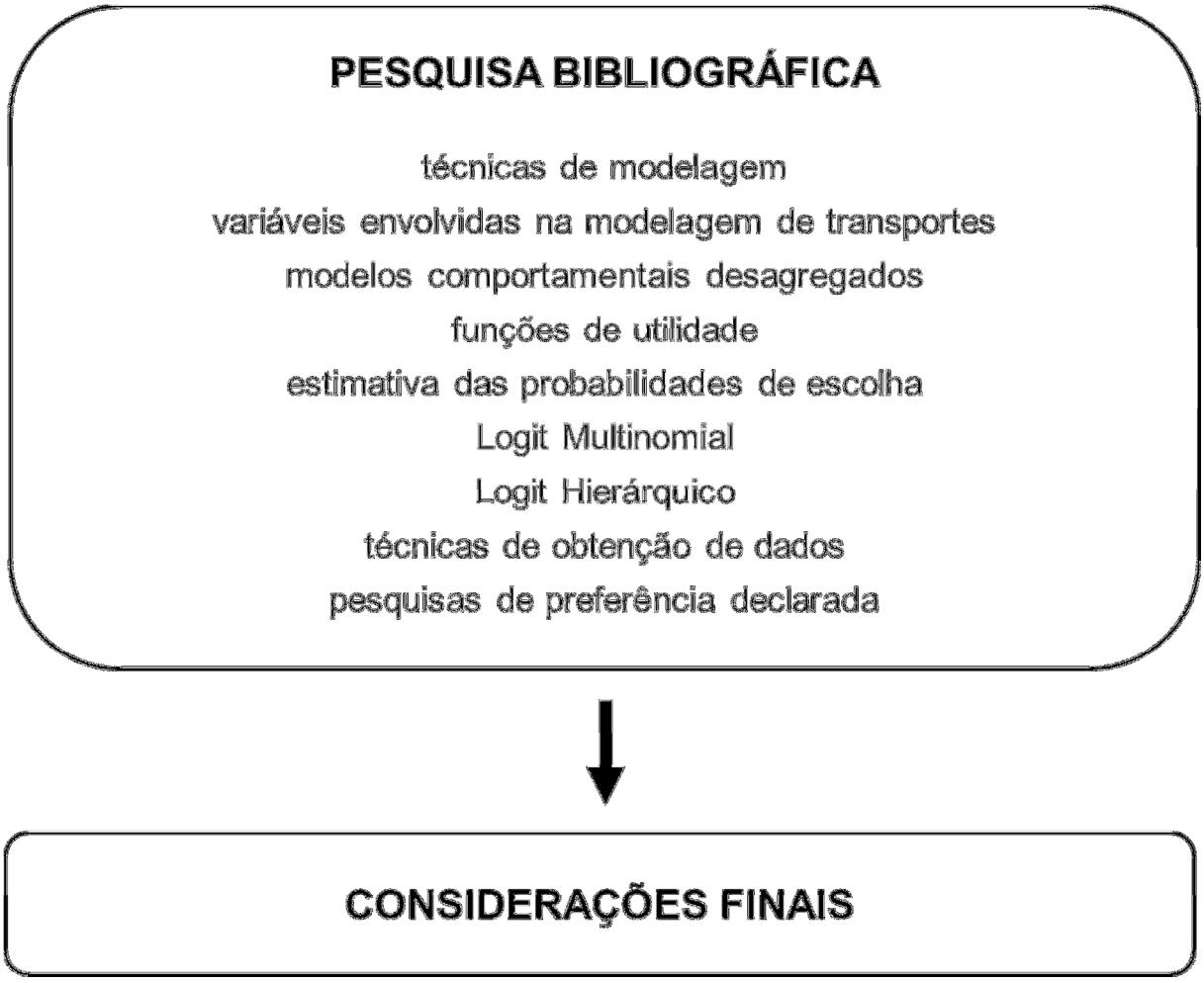
Após a descrição dessas funções é feito um estudo que visa obter a **estimativa das probabilidades de escolha** de cada modo de transporte disponível. Para isso são necessárias ferramentas matemáticas que irão transformar as **funções de utilidade** em dados percentuais.

A literatura especializada indica que para a **estimativa das probabilidades de escolha** os modelos mais utilizados são os modelos **Logit Multinomial** e **Logit Hierárquico**. O trabalho apresenta uma revisão sobre as características, os pressupostos teóricos e as limitações desses modelos.

Após a apresentação desses modelos, são descritas as **técnicas de obtenção de dados** para eles, destacando-se as **pesquisas de preferência declarada**, muito utilizadas para formulação de modelos.

Por fim, são apresentadas as **considerações finais** a respeito do tema, destacando as ideias mais importantes e as principais dificuldades encontradas para formulação dos modelos comportamentais desagregados.

Figura 1 – Etapas do trabalho



(fonte: elaborada pelo autor)

3 ESTIMATIVA DA DEMANDA POR TRANSPORTES

Os países mais industrializados e desenvolvidos têm sofrido com muitos problemas de transportes. O crescimento econômico tem causado uma demanda que excede a capacidade da maioria dos modais, visto que os investimentos na infraestrutura viária não aumentaram na mesma proporção. Em muitas regiões observa-se que um sistema de fornecimento de transportes entra em colapso quando qualquer coisa difere levemente da condição habitual (ORTÚZAR; WILLUMSEN, 1994, p. 3, tradução nossa).

Um dos fatores importante na tentativa de solucionar esses problemas é estimar a demanda por transportes. Para isso é necessário conhecer o comportamento dos indivíduos e entender de que maneira eles tomam suas decisões. Jordan et al. (2000, tradução nossa) mostram que:

> [...] as escolhas dos indivíduos são influenciadas por:
>
> a) hábitos;
>
> b) inércia;
>
> c) experiência;
>
> d) publicidade;
>
> e) pressão social;
>
> f) restrições ambientais;
>
> g) opiniões anteriores;
>
> h) restrições familiares;
>
> i) etc.

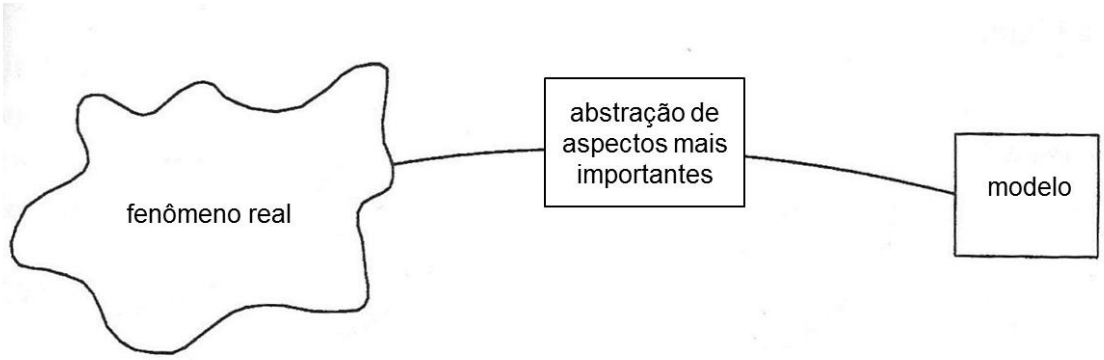
Nos itens a seguir, são apresentadas algumas técnicas utilizadas para gerar modelos que representem a realidade em geral e, mais especificamente, o comportamento dos indivíduos. Além disso, são abordados os principais tipos de modelos comportamentais utilizados em transportes e as variáveis que devem ser levadas em conta nesses modelos, bem como as características mais importantes dessas variáveis.

3.1 TÉCNICAS DE MODELAGEM

Conhecendo os fatores que influenciam as decisões dos indivíduos, os planejadores de transportes buscam uma maneira de estimar o comportamento deles frente a cenários hipotéticos, para que seja possível mensurar as necessidades e desejos dos usuários do sistema de transporte. Uma maneira que vem sendo muito utilizada para atingir este objetivo é a aplicação das técnicas de modelagem.

Ortúzar e Willumsen (1994, p. 2, tradução nossa) explicam que um modelo é uma representação simplificada da realidade, que enfoca elementos considerados mais importantes dentro do sistema estudado, conforme indicado na figura 2. Eles ainda dividem os modelos em duas categorias: modelos físicos – como a representação de uma maquete arquitetônica –, ou modelos abstratos, que representam uma teoria sobre o sistema estudado.

Figura 2 – Representação de um modelo



(fonte: adaptada de ORTÚZAR, 2000)

Em ambos os casos, é conveniente conhecer o modelo da maneira mais completa possível, isto é, observar o fenômeno a ser estudado identificando as variáveis relevantes para o comportamento do fenômeno. Além disso, é importante revisar a literatura disponível existente sobre ele, formulando novas hipóteses, se as que foram estudadas não lhe parecerem coerentes ou se for necessário por se tratar de hipóteses empíricas. Outro passo importante é descartar as variáveis consideradas pouco importantes ou aquelas que dependam de outras variáveis e estabelecer relações entre elas de acordo com as hipóteses levantadas no passo anterior para, enfim, validar o modelo, isto é, comprovar de maneira empírica as hipóteses e relações que foram assumidas (ORTÚZAR, 2000, p. 38, tradução nossa)

3.2 MODELOS DE PREVISÃO DE DEMANDA

Conhecer o comportamento dos usuários de transportes é fundamental para que se possa estimar a porcentagem da população que utiliza cada modo de transporte disponível, ou seja, estabelecer a **divisão modal**. Para esta finalidade, Ben-Akiva e Lerman (1985, p. 31, tradução nossa) sugerem que os modelos comportamentais devem ser:

a) **descritivos**, já que eles explicam como os indivíduos se comportam e não como eles deveriam se comportar;
b) **abstratos**, já que podem ser baseados em termos que não sejam específicos a uma situação em particular;
c) **operacionais**, já que os resultados serão parâmetros e variáveis que necessitam ser estimadas e mensuradas.

O mesmo autor admite que não há um modelo que satisfaça completamente os três postulados sugeridos, mas observa que pesquisas na área da psicologia, juntamente com as teorias clássicas da economia, apontam para o uso de modelos probabilísticos que podem representar o comportamento dos indivíduos de maneira satisfatória.

Na escolha modal, Ortúzar e Willumsen (1994, p. 162, tradução nossa) destacam a importância de conhecer as características dos usuários. A posse, ou não de automóveis e carteira de habilitação, a estrutura familiar (estado civil, existência ou não de filhos, etc.), a renda, a necessidade de utilizar o carro para o trabalho, para levar os filhos na escola ou ainda um familiar a alguma atividade fora de sua casa, são fatores relevantes para entender o comportamento do usuário.

Os mesmos autores sugerem, também, que seja feito um estudo a respeito das viagens propriamente ditas. Eles ressaltam a importância de conhecer o motivo da viagem e sua regularidade, pois seria mais comum uma pessoa utilizar o transporte público para ir até seu trabalho do que para realizar uma viagem com fins de lazer. O horário em que é realizado também foi abordado como ponto importante, pois viagens realizadas à noite são mais difíceis de serem feitas através transporte coletivo.

Por fim, Ortúzar e Willumsen (1994, p. 162, tradução nossa) apresentam as características dos meios de transportes dividindo-as em duas categorias:

a) quantitativas: o tempo de viagem (englobando os tempos no veículo, de espera e de caminhada para cada modal), os custos envolvidos e a disponibilidade e custo para estacionamento;

b) qualitativas: o conforto, a conveniência percebida, a segurança e a regularidade com que é disponibilizada a viagem.

Neste processo de escolha, quando é necessário optar por uma, entre duas ou mais alternativas, se faz necessária, também, uma regra de decisão. Estas regras descrevem de que maneira o tomador de decisões processa as informações e escolhe uma única alternativa dentre as disponíveis. Elas são classificadas em quatro grupos e descritas por Ben-Akiva e Lerman (1985, p. 35, tradução nossa) como:

a) **domínio**: para uma alternativa ser dominante em relação à outra ela deve ser melhor em pelo menos um atributo e não ser pior nos outros atributos analisados. De maneira geral esta regra não levaria a uma única escolha, pois dificilmente encontrar-se-á uma alternativa que seja superior em todos os atributos, mas servirá para eliminar alternativas inferiores do conjunto de opções;

b) **satisfação**: os atributos de cada alternativa carregam consigo um nível de satisfação baseada nas expectativas que os tomadores de decisão têm por cada alternativa. Essas expectativas derivam de informações recebidas ou experiências anteriores;

c) **lexicografia**: cada indivíduo ordena os atributos qualitativos em nível de importância. As alternativas que possuírem o atributo classificado no topo desta lista serão consideradas válidas e as outras serão descartadas. Este processo seguirá com o segundo atributo mais importante, e assim sucessivamente, até que se obtenha apenas uma alternativa restante;

d) **utilidade**: esta regra é utilizada para dar mensurabilidade aos atributos. Ela define uma função matemática que expressa a atratividade de cada alternativa através de seus atributos, que terão pesos diferentes na equação. A ideia é que cada indivíduo busque a opção que tenha a máxima utilidade.

Conhecidos os fatores que interferem na escolha modal e de que maneira os indivíduos tomam suas decisões, é necessário escolher um modelo que represente a realidade de maneira confiável. Ortúzar e Willumsen (1994, p. 178, tradução nossa) dividem os modelos entre: agregados (também chamados modelos de primeira geração), que são baseados nas relações observadas entre um grupo de pessoas ou nas relações médias observadas em um grande grupo de indivíduos; desagregados (ou modelos de segunda geração) que são baseadas nas escolhas feitas por cada pessoa, individualmente. De acordo com os autores, o uso de modelos desagregados representa a realidade de maneira mais realística e eles são discutidos detalhamente no capítulo 4.

4 MODELOS COMPORTAMENTAIS DESAGREGADOS

Modelos comportamentais desagregados (MCD), também conhecidos como modelos de escolha discreta, descrevem a escolha dos tomadores de decisões entre alternativas disponíveis. Train (2003, p. 15, tradução nossa) explica que "[...] os tomadores de decisões podem ser pessoas, famílias, empresas ou qualquer outra unidade de tomada de decisões, e as alternativas devem representar produtos concorrentes, possibilidades de ações ou quaisquer outras opções sobre as quais decisões precisem ser tomadas.".

Este autor ainda demonstra que para o conjunto de alternativas ser enquadrado no espaço de trabalho dos MCD, eles devem:

a) ter alternativas **mutualmente exclusivas**: ao tomar uma decisão o indivíduo deve escolher apenas uma dentre as opções. Escolher uma alternativa, implica, necessariamente, em não escolher nenhuma outra;
b) ser **completos**: todas alternativas possíveis devem estar contempladas. O tomador de decisões, necessariamente, escolhe uma alternativa, dentre as disponíveis;
c) ter um número finito de alternativas.

Com o objetivo de explicar esses modelos, Spear[1] (1977 apud ORTÚZAR; WILLUMSEN, 1994, p. 179), cita algumas características importantes descritas a seguir:

a) MCD são baseados no comportamento individual e não constituem analogias físicas de nenhuma espécie. Esta é uma vantagem dos MCD sobre os modelos convencionais, pois faz com que eles sejam mais estáveis no tempo e no espaço;
b) como os MCD são baseados em dados individuais, surgem algumas consequências como,
   - tendem a ser mais eficientes que os modelos convencionais. Os modelos agregados são baseados na média de várias informações individuais, enquanto os MCD utilizam cada informação individual como um dado de entrada, necessitando, portanto, de menos informações para a modelagem;

[1] SPEAR, B. D. **Aplications of new travel demand forecasting techiniques to transportation**: a study of individuall choice models. Washington: US Departament of Transportation, 1977.

- os dados utilizados são individuais, portanto todas informações adicionais obtidas juntamente com os dados principais, podem ser utilizadas para produzir melhores resultados;

- sofrem menos influência das correlações de unidades agregadas. Um grande problema dos dados agregados é a **correlação ecológica**, que ocorre quando o comportamento individual pode estar oculto devido a algumas características associadas à zona (seja país, cidade ou uma porção da cidade) que não foram identificadas individualmente. A figura 3 mostra o fenômeno chamado falácia ecológica, no qual se os dados fossem observados em níveis de zonas, comparando a zona A e a zona B, o número de viagens por veículos diminuiria com a renda. Porém o contrário pode ser observado, se fossem utilizados os dados de cada família, individualmente;

c) MCD resultam na probabilidade de escolha de uma alternativa, não indicando qual delas deve ser selecionada. Por se tratarem de modelos probabilísticos devem obedecer aos seguintes conceitos básicos,

- o número de pessoas que utilizam certa opção é a soma das probabilidades de cada indivíduo escolher esta alternativa, como é observado na fórmula 1;

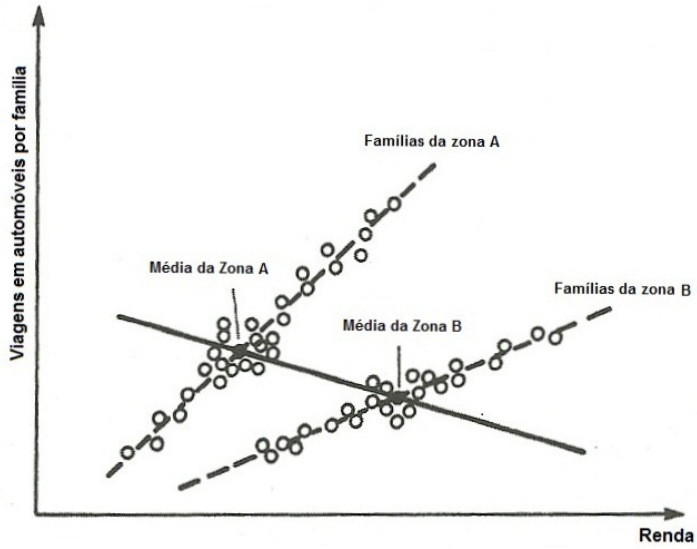
- um conjunto independente de decisões deve ser modelado separadamente, considerando cada uma como uma escolha condicional. As probabilidades resultantes são multiplicadas obtendo, então, a probabilidade de escolha para o conjunto, como descrito na fórmula 1;

d) em virtude de aceitarem, em suas funções utilidade, qualquer número ou especificação das variáveis que as explicam, observa-se que,

- MCD permitem uma representação mais flexível das variáveis consideradas relevantes no estudo;

- os coeficientes das variáveis explicativas refletem a relativa importância de cada atributo.

Figura 3 – Exemplo de falácia ecológica



(fonte: adaptada de ORTÚZAR; WILLUMSUN, 1994)

A fórmula indicada é:

$$P(f, m, d, r) = P(f)P(d/f)P(m/d, f)P(r/m, df) \qquad \text{(fórmula 1)}$$

Onde P são as probabilidades definidas em função de:
f = frequência;
d = destino;
m = modo de transporte;
r = rota escolhida.

O processo de escolha dos MCD está diretamente ligado a um conceito visto no capítulo anterior, chamado **Utilidade**. As funções de utilidade representam a atratividade de cada alternativa, percebidas pelos tomadores de decisões. Nodari (1996, p. 9) afirma que "[...] os indivíduos buscam maximizar sua função de utilidade frente às restrições de recursos existentes [...]", ou seja, os indivíduos sempre irão escolher a alternativa que lhe pareça ser mais satisfatória.

A seguir são apresentadas, de maneira detalhada, as funções de utilidade, bem como suas propriedades, e a teoria da Utilidade Aleatória, bastante importante para obtenção de um MCD.

4.1 FUNÇÕES DE UTILIDADE

Para representar a função de Utilidade, Nodari (1996, p. 22) afirma que ela "[...] normalmente assume a forma de modelos aditivos compensatórios, já que se pode melhorar um atributo piorando outro, e assim, manter o mesmo nível de Utilidade.". A Utilidade pode ser representada da seguinte maneira pela fórmula 2:

$$V_i = a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n \qquad \text{(fórmula 2)}$$

Onde:

V_i é a Utilidade da opção i;
$X_1, X_2, \dots, X_n$ são os atributos do produto ou serviço;
$a_1, a_2, \dots, a_n$ são os coeficientes do modelo;
a_0 é a constante específica de modo.

Baseado em experiências anteriores, o modelador deve definir quais serão os atributos ($X_1, X_2, \dots, X_n$) que irão influenciar na Utilidade (V_i) de cada alternativa. Os coeficientes do modelo ($a_1, a_2, \dots, a_n$) determinam o quanto cada atributo influi na Utilidade. Já a constante a_0 atribui o peso das variáveis não explicitadas no modelo.

Confirmando a teoria de modelos aditivos compensatórios observa-se que "[...] o tomador de decisões irá sempre escolher a melhor combinação entre tempo de viagem, custo de viagem e conforto. Assim, o mais dispendioso modo pode ser escolhido se isso for suficientemente compensado oferecendo melhores serviços." (BEN-AKIVA; LERMAN, 1985, p. 37, tradução nossa).

Jordan et al. (2000, p. 39, tradução nossa) afirmam que, além da busca pela maximização da utilidade existem fatores estritamente individuais que fazem com que os indivíduos tomem suas decisões. Estes fatores não estão incluídos na função Utilidade, pois não há como mensurá-los no modelo.

De acordo com Ben-Akiva e Lerman (1985, p. 49, tradução nossa), é comum em experimentos de pesquisa apresentar o mesmo cenário em momentos distintos para um indivíduo. Nesses experimentos é observada uma inconsistência de comportamento, uma vez que o indivíduo escolhe alternativas diferentes para os mesmos cenários. Os motivos que levam a esse comportamento são desconhecidos, ou imensuráveis.

Os mesmo autores apontam, ainda, que dois ou mais indivíduos que possuem os mesmo atributos e as mesmas características socioeconômicas, quando expostos às mesmas alternativas, podem escolher alternativas distintas. Um exemplo disso seria quando dois funcionários que habitualmente viajam a trabalho, escolherem diferentes modos de transporte para uma viagem aparentemente igual.

A Teoria da Escolha Probabilística é usada para capturar os efeitos das inconsistências comportamentais dos indivíduos e dos atributos que não foram observados em cada alternativa. Essa teoria também visa explicar comportamentos puramente aleatórios dos

indivíduos e os erros decorrentes da incorreta percepção dos atributos das alternativas. Essa teoria é "[...] claramente uma dicotomia não resolvida entre a afirmação de um comportamento fundamentalmente probabilístico e o que pode aparentemente ser probabilístico devido a nossa incapacidade de entender completamente e de mensurar os fatores relevantes que afetam o comportamento humano." (BEN-AKIVA; LERMAN, 1985, p. 49, tradução nossa).

Luce e Suppes[2] (1965 apud BEN-AKIVA; LERMAN, 1985, p. 49, tradução nossa) distinguem a Teoria da Escolha Probabilística em duas abordagens: a Utilidade Constante e a Utilidade Aleatória. A primeira trata a utilidade de cada alternativa como fixa, quando, ao invés de selecionar a alternativa com a maior utilidade, a decisão do indivíduo é definida por uma função de distribuição probabilística sobre as alternativas que inclui as Utilidades como parâmetros.

A segunda abordagem indica que a utilidade possui uma componente aleatória e postula que os indivíduos buscam maximizar sua Utilidade, sendo assim, a teoria da utilidade aleatória visa explicar as inconsistências observadas na abordagem constante da utilidade. Essa teoria é discutida no próximo item.

4.2 TEORIA DA UTILIDADE ALEATÓRIA

O modelador é um observador do sistema e, portanto, não consegue determinar todos os elementos que o tomador de decisões leva em consideração ao analisar as alternativas. Ainda que dois indivíduos tenham o mesmo padrão socioeconômico, morem em uma mesma localidade e, aparentemente, tenham opiniões semelhantes a respeito do sistema estudado, eles podem escolher alternativas distintas para uma única questão.

Ben-Akiva e Lerman (1985, p. 55, tradução nossa) explicam que as inconsistências observadas nas escolhas dos indivíduos são fruto das limitações dos modeladores. Os indivíduos irão sempre maximizar sua Utilidade, porém como os modeladores não podem mensurar com a devida certeza, elas são tratadas como variáveis aleatórias. Neste contexto, a probabilidade de escolha da alternativa i é igual a probabilidade da Utilidade da alternativa i

[2] LUCE, R.; SUPPES, P. Preference, Utility and Subjective Probability. In: LUCE, R. et al. **Mathematical Psychology.** 3th ed. New York: Wiley, 1965.

(U_i) ser igual ou maior que as Utilidades de todas outras alternativas. Isso pode ser descrito pela fórmula 3:

$$P(i) = Prob[U_i \geq U_n, \forall\, n \in C_n] \qquad \text{(fórmula 3)}$$

Onde:
$P(i)$ é a probabilidade da alternativa i ser escolhida;
Ui é a Utilidade da alternativa i;
Un é a utilidade da alternativa n;
C_n é o conjunto de alternativas disponíveis.

Os autores descrevem que a Utilidade de cada alternativa pode ser escrita indiretamente como função dos atributos de cada alternativa e das características socioeconômica dos indivíduos tais como sua renda, idade e nível de escolaridade. Essa representação das Utilidades das alternativas é apresentada na fórmula 4:

$$U_{iq} = U(Z_{iq}, S_q) \qquad \text{(fórmula 4)}$$

Onde:
Z_{iq} é o vetor de valores atribuídos à alternativa i pelo indivíduo q;
S_q é o vetor das características do indivíduo q.

De acordo com Manski[3] (1973 apud BEN-AKIVA; LERMAN, 1985, p. 55, tradução nossa), a aleatoriedade das Utilidades pode ter várias origens. Uma delas são os **atributos não observados**, em que o vetor de atributos que afetam a decisão está incompleto. Dessa maneira, acrescenta-se um elemento Z_q^U, que é uma variável aleatória, e consequentemente a utilidade da alternativa será aleatória, como indicado na fórmula 5:

[3]MANSKI, C. **The Analysis of Qualitative Choice**. 1973. PhD dissertation – Department of Economics, MIT, Cambridge, 1973 – demais informações não foram fornecidas pelo autor.

$$U_{iq} = U(Z_{iq}, S_q, Z_q^U) \qquad \text{(fórmula 5)}$$

A **variação não observada das preferências** é outro fator observado que pode alterar a Utilidade das alternativas. O elemento S_q^U é adicionado para explicar as preferências de cada indivíduo no processo de tomada de decisão, conforme indicado na fórmula 6. Uma vez que este atributo é desconhecido, a Utilidade das alternativas é uma variável aleatória:

$$U_{iq} = U(Z_{iq}, S_q, S_q^U) \qquad \text{(fórmula 6)}$$

Os autores citam que os **erros ao mensurar os atributos** afetam a Utilidade das alternativas. É sugerido acrescentar uma variável aleatória de erro $\dot{\epsilon}_{iq}$, conforme indicado na fórmula 7, afim de representar os possíveis erros cometidos ao mensurar os atributos de cada alternativa. Uma vez que essa variável é aleatória, a Utilidade da alternativa terá a mesma propriedade:

$$U_{iq} = U(Z_{iq} + \dot{\epsilon}_{iq}, S_q) \qquad \text{(fórmula 7)}$$

Por fim, são descritas as **variáveis instrumentais**, que são variáveis relacionadas ao reais atributos das alternativas. Neste caso, o vetor de valores atribuídos a determinada alternativa, é na verdade função da imperfeição existente na relação entre os instrumentos e os atributos e também de uma variável aleatória, conforme é indicado na fórmula 8:

$$U_{iq} = U[g(Z_{iq}) + \ddot{\epsilon}_{iq}, S_q] \qquad \text{(fórmula 8)}$$

Onde:

$g(Z_{iq})$ é a função que representa os valores atribuídos à alternativa i pelo indivíduo q;
$\ddot{\epsilon}_{iq}$ é a variável aleatória de erro;
S_q é o vetor das características do indivíduo q.

De maneira resumida, Ortúzar e Willumsen (1994, p. 182, tradução nossa) definem a Utilidade Aleatória em duas componentes distintas: uma representativa (podendo ser chamada de sistemática ou mensurável), que é função dos atributos mensuráveis, descrita na fórmula 2; e outra aleatória, que reflete a idiossincrasia (resposta do indivíduo a fatores externos, de acordo com seu temperamento) e as particularidades de cada pessoa, juntamente com erros que o modelador comete ao mensurar os atributos. A seguir, essas hipóteses são apresentadas na fórmula 9:

$$U_{iq} = V_{iq} + \varepsilon_{iq} \qquad \text{(fórmula 9)}$$

Onde:
U_{iq} é a utilidade da alternativa i, associada ao indivíduo q;
V_{iq} é a parte representativa dessa Utilidade;
ε_{iq} é a componente randômica dessa Utilidade.

De posse, enfim, das funções de Utilidade, é necessário fazer com que essas Utilidades representem a probabilidade de o indivíduo escolher determinada alternativa. O capítulo seguinte dedica-se a explicar de que maneira é possível estimar essa probabilidade, utilizando os modelos Logit Multinomial e Logit Hierárquico.

5 ESTIMATIVA DAS PROBABILIDADES DE ESCOLHA

De acordo com Nodari (1996, p. 24), é possível usar a utilidade comparando diretamente se $U_i > U_j$ ou $U_i < U_j$, ou, a partir destas duas Utilidades, calcular a probabilidade de escolha das alternativas i e j. No caso de planejamento de transportes, procura-se estimar as probabilidades de escolhas, necessitando, portanto, seguir a segunda opção.

A fim de gerar essas probabilidades, Ortúzar e Willumsen (1994, p. 182, tradução nossa) afirmam que o indivíduo escolherá a alternativa i, se ela tiver a maior Utilidade dentre as alternativas, como mostra a fórmula 10:

$$U_{iq} \geq U_{nq}, \forall\, n \in C_n \qquad \text{(fórmula 10)}$$

Onde:
U_{iq} é a utilidade da alternativa i associada ao indivíduo q;
U_{nq} é a utilidade da alternativa n associada ao indivíduo q;
C_n é o conjunto de alternativas disponíveis.

Utilizando a fórmula 9, e separando as variáveis aleatórias no mesmo lado da equação, obtém-se, a seguir, a fórmula 11:

$$V_{iq} - V_{nq} \leq \varepsilon_{nq} - \varepsilon_{iq} \qquad \text{(fórmula 11)}$$

Sabendo que o modelador não conhece as variáveis aleatórias, não é possível determinar ao certo se a fórmula 11 é valida ou não. Por esse motivo, a probabilidade da alternativa i ser escolhida é a probabilidade de este postulado estar correto, expressa por:

$$P_i = Prob\{\varepsilon_{nq} \leq \varepsilon_{iq} + (V_{iq} - V_{nq}), \forall\, n \in C_n\} \qquad \text{(fórmula 12)}$$

Como não se conhece a distribuição das variáveis aleatórias, não é possível obter uma expressão analítica para o modelo, sendo necessário estimar algum tipo de distribuição. Isso demonstra que os modelos serão distintos, de acordo com a distribuição utilizada (ORTÚZAR; WILLUMSEN, 1994, p. 183, tradução nossa).

Ortúzar (2000, p. 95, tradução nossa) indicam que uma classe importante dos modelos, utiliza a variável aleatória com distribuições independentes e idênticas, como os modelos Logit Multinomial (LM) e Logit Hierárquico (LH). Os dois modelos serão discutidos nos próximos itens.

5.1 LOGIT MULTINOMIAL

Ben-Akiva e Lerman (1985, p. 104, tradução nossa), demonstram que ao assumir que a fórmula 9 é verdadeira, e que a parte randômica ε_{iq} da utilidade é independente e identicamente distribuída (IID), seguindo a distribuição de Gumbel[4], com uma escala de parâmetros $\mu > 0$, a probabilidade de escolha da alternativa i é descrita na fórmula 13:

$$P_i = \frac{e^{\mu U_{iq}}}{\sum_{j=1}^{n} e^{\mu U_{jq}}} \qquad \text{(fórmula 13)}$$

Onde:
Pi é a probabilidade da alternativa i ser escolhida;
e é a base do logaritmo nepreniano;
j são as alternativas consideradas;
U é a utilidade das alternativas consideras;
μ é o fator que determina a escala de distribuição dos erros.

A respeito das variáveis aleatórias, Ortúzar (2000, p. 101, tradução nossa) lembra que cada indivíduo age de forma racional, portanto ele repete o processo de decisão escolhendo a mesma alternativa sempre que lhe forem dadas as mesmas opções. O que explica o fato de isto, as vezes, não ocorrer é a distribuição das variáveis aleatórias discutidas no item anterior

[4] Distribuição estatística de erros que possui sua forma semelhante à forma exponencial.

Os autores descrevem que ao considerar a variável aleatória como IID, todos os erros e componentes randômicas terão a mesma escala de parâmetros μ. Embora a escolha do μ seja arbitrária – pois somente irá determinar a escala das utilidades – o fato de que cada distúrbio, ou erro, tem o mesmo valor de μ, implica que as componentes randômicas das utilidades variam da mesma maneira. Essa simplificação não condiz com a realidade, uma vez que é esperado que os atributos não observados da escolha modal entre o carro e o transporte coletivo sejam diferentes.

Em relação às funções de utilidade, Ortúzar e Willumsen (1994, p. 184, tradução nossa) explicam que cada alternativa terá uma constante específica de modo (a_0). Entretanto, não é possível especificar uma constante para cada modo, sendo recomendado que uma alternativa seja tomada como referência e tenha essa constante com valor igual a zero. Dessa maneira, as demais alternativas terão suas constantes específicas definidas em função da constante de referência.

O modelo LM apresenta outra particularidade em relação às funções de utilidades. Os mesmos autores afirmam que se, num modelo binário, um atributo tiver o mesmo valor de coeficiente em mais de uma alternativa (no caso da variável representar atributos individuais, como renda, idade, sexo ou número de carros em uma residência), não será possível estimar um coeficiente genérico, pois criaria, na equação, um coeficiente com valor igual a zero. Neste caso, esse atributo poderia aparecer somente em algumas das opções.

O problema gerado pelos atributos individuais é que nem sempre é trivial definir em quais alternativas a variável irá aparecer. Ortúzar e Willumsen (1994, p. 185, tradução nossa) exemplificam a dificuldade de avaliar esses atributos:

> Considerando o caso da variável sexo (igual a zero para homens, e igual a 1 para mulheres) no estudo de escolha modal. Se for considerado que os homens sempre levam vantagem em relação ao usar o carro da família para ir ao trabalho, essa variável não seria incluída, simultaneamente, em ambas as utilidades, do motorista do carro e do passageiro. Entretanto, não há indicação se é preciso inserir, ou não, essa variável nas utilidades dos outros modos como, por exemplo, ônibus ou metrô. O problema é que ao inserir variáveis de diferentes maneiras, normalmente produzir-se-ão diferentes estimativas de resultados, e escolher o resultado ótimo, acarretaria em um complexo problema combinatório, mesmo para um pequeno número de opções e atributos. Se não há discernimento nem base teórica para preferir uma forma em detrimento da outra, a única maneira de obter o resultado é através do método da tentativa e erro.

A aplicabilidade do LM é resumida por Train (2003, p. 46, tradução nossa) nos seguintes itens:

a) ele pode representar a variação dos parâmetros representativos (parâmetros relativos às características observadas pelo modelador), mas não representa variações nos parâmetros aleatórios (diferenças que não podem ser relacionadas às características observadas);
b) ele gera substituições proporcionais através das alternativas, utilizando a especificação dada pelo modelador sobre a utilidade representativa;
c) se os fatores não observados são independentes em repetidas situações de escolha, o modelo é capaz de capturar a dinâmica da escolha repetida, incluindo a dependência discreta.

O modelo LM possui uma importante característica que é a independência de alternativas irrelevantes (IAI). Ortúzar (2000, p. 98, tradução nossa) explica que a IAI pressupõe que a relação entre a probabilidade de escolha de duas alternativas é constante e não depende das demais alternativas, como mostra a fórmula 14:

$$\frac{P_j}{P_i} = exp\{\mu(V_j - V_i)\} \qquad \text{(fórmula 14)}$$

O autor descreve que a IAI apresenta a vantagem de, no futuro, adicionar ao modelo opções que não foram consideradas durante a estimativa inicial. Para isso somente seria necessário considerar – para a nova opção – os mesmos atributos considerados nas alternativas iniciais. Entretanto, um problema causado pela IAI é que o modelo pode falhar se algumas das alternativas estiverem correlacionadas, ou seja, apresentem características semelhantes às outras.

Ortúzar e Willumsen (1994, p. 166, tradução nossa) demonstram esse problema, por exemplo, em uma cidade qualquer, 50% das pessoas escolhem carro (C) e 50% escolhem ônibus (O), ou seja, a parte representativa da utilidade de cada modo é igual, como indicado na fórmula 15:

$$P_c = \frac{exp(-\mu V_C)}{exp(-\mu V_c) + exp(-\mu V_o)} = 0{,}50 \qquad \text{(fórmula 15)}$$

P_c é a probabilidade de escolher o carro;
μ é o fator que determina a escala de distribuição dos erros;
V_c é a parte representativa da utilidade do carro;
V_o é a parte representativa da utilidade do ônibus.

Hipoteticamente, o dono da empresa resolve pintar metade de seus ônibus na cor azul (OA) e a outra metade na cor vermelha (OV), mantendo exatamente o mesmo nível de serviço. Se o carro mantiver seu mesmo nível de serviço, observa-se que a probabilidade de escolha do carro cai para 33%, como indicado na fórmula 16:

$$P_c = \frac{exp(-\mu V_C)}{exp(-\mu V_c) + exp(-\mu V_{oa}) + exp(-\mu V_{ov})} = 0{,}33 \qquad \text{(fórmula 16)}$$

P_c é a probabilidade de escolher o carro;
μ é o fator que determina a escala de distribuição dos erros;
V_c é a parte representativa da utilidade do carro;
V_{oa} é a parte representativa da utilidade do ônibus azul;
V_{ov} é a parte representativa da utilidade do ônibus vermelho.

Esperava-se que a probabilidade de escolha do carro fosse mantida em 50% e que os ônibus dividissem igualmente a outra metade do mercado. Os autores afirmam que o exemplo é exagerado, porém é importante para demonstrar os problemas que podem surgir quando alternativas correlacionadas são adicionadas ao modelo.

Ben-Akiva e Lerman (1985, p. 109, tradução nossa) explicam que a falha do LM no caso dos ônibus azuis e vermelhos ocorre em função de o pressuposto do modelo, de que os erros são individualmente distribuídos, não foi seguido. Neste caso, ambos os ônibus possuem as mesmas características não observadas, sendo, portanto, perfeitamente correlacionados.

Os autores afirmam que a propriedade da IAI pode ser equivocadamente aplicada à população como um todo, levando a crer que a probabilidade de escolha de uma parcela da população não é afetada pela utilidade das demais alternativas. A seguir é apresentado um exemplo que demonstra de que maneira a propriedade da IAI afeta o modelo.

Supondo que numa pequena cidade as viagens geradas por centros comerciais sejam divididas igualmente entre um centro comercial localizado na periferia, e outro no centro da cidade. Se um novo centro comercial fosse construído na periferia, com atributos observados idênticos ao primeiro já construído na área, o LM poderia gerar uma divisão de mercado de um terço da população para cada centro comercial, porém se os dados forem utilizados de maneira correta isso não acontece.

Para que a divisão de mercado seja correta, é preciso supor que as pessoas que utilizam centros comerciais são divididas em dois grupos de mesmo tamanho, que são internamente homogêneos nos seus atributos observados. Neste exemplo, é possível observar que 95% do Grupo 1 utilizaria o centro comercial localizado no centro da cidade, enquanto no Grupo 2 95% das pessoas utilizaria o centro comercial da periferia. A tabela 1 apresenta a divisão de mercado observada antes e depois da construção do novo centro comercial.

É possível observar que a IAI é aplicada em cada grupo homogêneo, porém **não pode ser aplicada para a população como um todo.** Neste exemplo observa-se que a fatia de mercado do centro comercial localizado no centro da cidade é pouco afetada pela construção do novo centro comercial periférico, ao contrário do exemplo dos ônibus azuis e vermelhos, na qual a fatia de mercado ficou dividida em um terço para cada alternativa (BEN-AKIVA; LERMAN, 1985, p. 110, tradução nossa).

Tabela 1 – Exemplo de independência de alternativas irrelevantes

	Probabilidade de escolha (centro)	Probabilidade de escolha (periferia 1)	Probabilidade de escolha (periferia 2)
Antes do novo centro comercial			
Grupo 1	0,95	0,05	
Grupo 2	0,05	0,95	
Fatia de mercado	0,50	0,50	
Após o novo centro comercial			
Grupo 1	0,9048	0,0476	0,0476
Grupo 2	0,0256	0,4872	0,4872
Fatia de mercado	0,4652	0,2674	0,2674

(fonte: adaptada de BEN-AKIVA; LERMAN, 1985)

A probabilidade de escolher um dos centros comerciais era consideravelmente diferente entre os dois grupos, a segmentação e a **homogeneidade** dos grupos são fundamentais para obtenção dos resultados corretos. Uma discussão importante a respeito da IAI é feita por Ben-Akiva e Lerman (1985, p. 111, tradução nossa):

> A ideia de independência de alternativas irrelevantes pode ser mais ou menos aceitável, dependendo se a heterogeneidade da população levada em conta no modelo é significante. Isso sugere que o fato de o LM ser um modelo apropriado, ou não, para uma particular situação de escolha, é algo que não pode ser previamente deduzido. Ao contrário, é necessário examinar as particularidades da componente sistemática da função de utilidade e averiguar se elas explicam de maneira razoável as heterogeneidades da população. Como uma consequência prática, modelos LM que incluem diversos atributos socioeconômicos de maneira apropriada, possuem maior probabilidade de realizar bons prognósticos, do que os modelos que omitem essas variáveis.

Para que a segmentação, a homogeneidade e a correlação entre as variáveis sejam consideradas de maneira adequada nos modelos, foi proposta a utilização de modelos mais abrangentes, como o modelo Logit Hierárquico, que é discutido no próximo item.

5.2 LOGIT HIERÁRQUICO

O modelo Logit Hierárquico (LH) é uma derivação do LM que é usado de maneira apropriada quando o conjunto de alternativas de um problema de decisão pode ser subdividido em grupos

chamados nichos (subconjuntos nos quais as alternativas são correlacionadas ou similares). Seu uso implica que, para cada par de alternativas, presentes no mesmo nicho, a relação entre as probabilidades independe dos atributos ou da existência de outras alternativas, ou seja, a IAI se mantém em cada nicho . Além disso, para duas alternativas que estão em diferentes nichos, a relação entre as probabilidades depende dos atributos das alternativas presentes nestes dois nichos, o que implica que a IAI não é aplicável para alternativas de diferentes nichos (TRAIN, 2003, p. 81, tradução nossa).

Este modelo é utilizado quando é necessário estimar duas ou mais características ao mesmo tempo, e existe uma hierarquia decisória entre elas. Sua estrutura agrupa em subconjuntos as alternativas que tiverem atributos com valores semelhantes ou correlacionados, formando subconjuntos com níveis de hierarquia diferentes. A hierarquia de um subconjunto pode se sobrepor em relação a outro, mas não irá deixar de influenciar e ser influenciada por ele (NODARI, 1996, p. 11).

Ortúzar e Willmusen (1994, p. 188, tradução nossa) descrevem os passos necessários para a aplicação do modelo LH:

a) é estruturado de maneira a agrupar os subconjuntos em nichos hierárquicos. Cada nicho, por sua vez, é representado por uma composição de alternativas que competem com as outras disponíveis para o indivíduo, como exemplificado na figura 4;

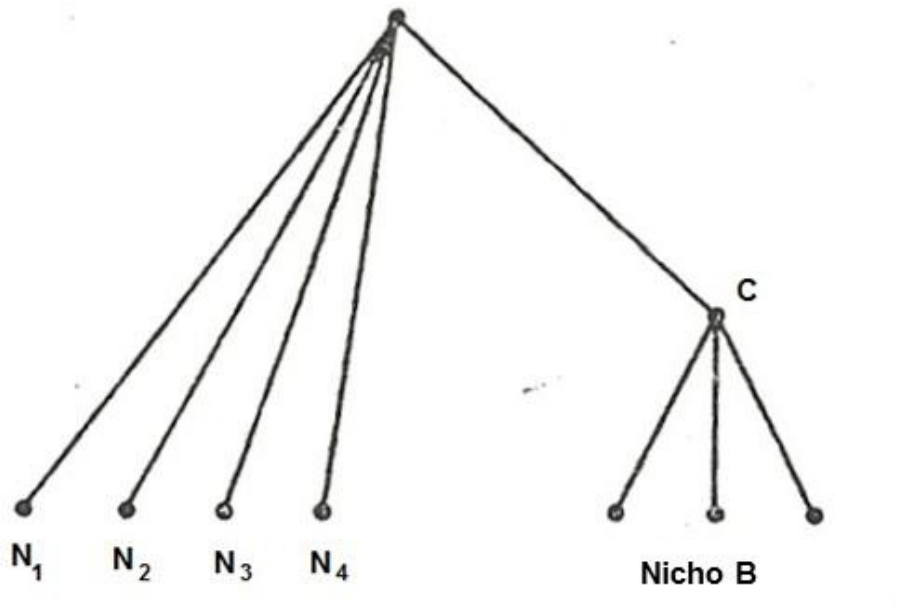
b) deve ser estimado sequencialmente, por programas computacionais. Utilizando como exemplo a figura 4, primeiramente seria utilizado o modelo LM entre as alternativas do nicho B – desconsiderando as variáveis que possuem o mesmo valor para todos os membros do nicho – e, após isso, seria feita a inclusão da alternativa composta C, no nível superior, relacionando-a com as demais alternativas ($N_1, N_2,$ etc);

c) os nichos inferiores devem ser introduzidos nos seus níveis imediatamente superiores através das utilidades das alternativas compostas, que, em geral, tem duas componentes: uma é a máxima utilidade esperada das opções dos nichos inferiores, sem considerar os atributos comuns a todas alternativas do nicho; e outra que considera os atributos comuns a todos os membros do nicho;

d) para estimar o modelo, é necessário aplicar o LM para os níveis superiores, ou seja, para as alternativas compostas que representam as opções de baixa hierarquia, mais as opções que não foram divididas em nichos;

e) a probabilidade de o indivíduo escolher uma alternativa presente no nicho B, é composta pela probabilidade marginal de escolher a alternativa composta C (no nível superior) e a probabilidade condicional de escolher esta alternativa no nível de seu nicho.

Figura 4 – Modelo Logit Hierárquico com dois nichos



(fonte: adaptada de ORTÚZAR; WILLUMSEN, 1994)

Ortúzar (2000, p. 114, tradução nossa) indica algumas limitações ao uso do modelo LH:

> a) da mesma forma que o modelo LM, este modelo não permite tratar as variações nos gostos dos indivíduos nem funções de utilidade heterocedásticas[5];
>
> b) o modelo somente considera tantas interdependências entre alternativas quantos nichos forem especificados na estrutura hierárquica postulada. Não é possível tratar problemas de correlação entre opções pertencentes a nichos distintos;
>
> c) a busca pelo melhor modelo LH pode implicar em análises experimentais de várias estruturas hierárquicas; é fácil observar que o número de estruturas possíveis inicialmente, aumenta mais rapidamente que a quantidade de opções disponíveis. Por isso, a teoria e a intuição possuem um papel importante na redução deste problema combinatório;
>
> d) a calibração sequencial do modelo, utilizando programas computacionais para o modelo LM tinha vários problemas como erros de transmissão e diferentes valores de parâmetros em diferentes nichos. Estes problemas foram solucionados com a aparição de programas computacionais especializados, como o ALOGIT, que permitiu estimação simultânea dos modelos.

Um exemplo que explica a utilização do LH descrito por Train (2003, p. 81, tradução nossa), é quando um trabalhador pode escolher, para suas viagens diárias ao trabalho, entre ir sozinho em seu automóvel, compartilhar o autmóvel com algum carona, ir de ônibus ou ir de trem. Se qualquer das alternativas fosse removida, a probabilidade de escolha das outras alternativas

[5] Utilidades com distribuição de erros em padrão irregular

aumentaria, porém o desafio é saber em que proporção ocorreria esse aumento em cada uma das alternativas.

Supondo que as mudanças nas probabilidades ocorram da maneira indicada na tabela 2. É possível observar que as probabilidades de escolha do ônibus e do trem aumentam à mesma proporção quando alguma das outras duas alternativas é removida. A IAI se aplica entre o ônibus e o trem, portanto é possível colocá-los em um nicho denominado transporte público. A IAI também é aplicada à viagem de automóvel individual e de automóvel compartilhado, pois quando alguma alternativa do transporte público é removida, a probabilidade de escolha do automóvel individual e compartilhado aumenta à mesma proporção, estabelecendo-se outro nicho denominado transporte individual. É possível observar que a IAI não é aplicável simultaneamente às quatro alternativas, nem pode se aplicada com alternativas de grupos diferentes, portanto a divisão entre transporte individual e transporte público é apropriada para essa situação.

Tabela 2 – Mudança nas probabilidades quando uma alternativa é removida

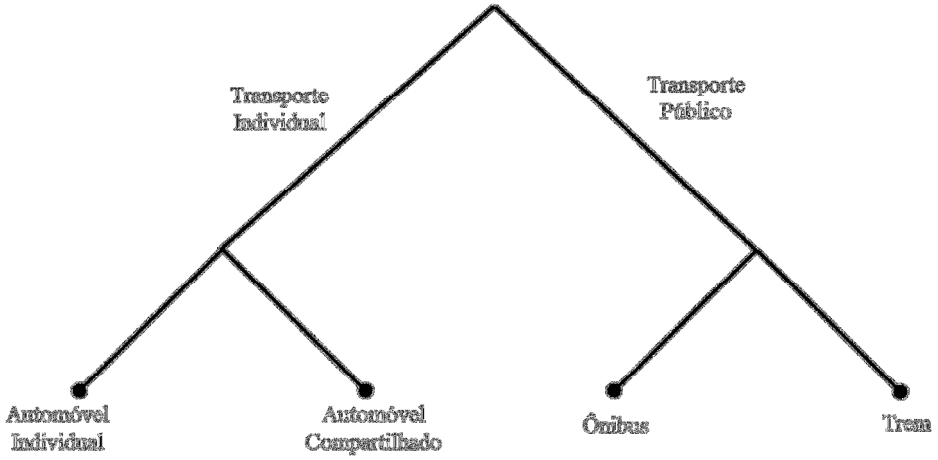
	Probabilidades				
		Com remoção de alternativas			
Alternativa	Original	Automóvel sozinho	Automóvel compartilhado	Ônibus	Trem
Automóvel sozinho	0,40	-	0,45 (+ 12,5%)	0,52 (+ 30%)	0,48 (+20%)
Automóvel compartilhado	0,10	0,20 (+ 100%)	-	0,13 (+ 30%)	0,12 (+ 20%)
Ônibus	0,30	0,48 (+ 60%)	0,33 (+ 10%)	-	0,40 (+ 33%)
Trem	0,20	0,32 (+ 60%)	0,22 (+ 10%)	0,35 (+ 70%)	-

(fonte: adaptada de TRAIN, 2003)

O autor explica que a escolha modal, neste exemplo, pode ser representada pelo diagrama indicado na figura 5. É feita uma analogia desse diagrama com uma árvore, em que cada ramo da árvore representa um subconjunto de alternativas em que pode ser aplicado a IAI, e cada galho desses ramos representa uma alternativa. Para o exemplo, existem dois ramos, o transporte público e o transporte individual, e em cada um desses ramos há dois galhos, que seriam cada uma das quatro alternativas. É possível observar que ocorre uma substituição

proporcional entre dois galhos do mesmo ramo, porém não ocorre entre dois ramos da mesma árvore.

Figura 5 – Diagrama de árvore para escolha modal



(fonte: adaptada de TRAIN; 2003)

Ortúzar e Willmusen (1994, p. 189, tradução nossa) apresentam um exemplo da formulação matemática envolvida para estimar as probabilidades de escolha entre o automóvel (A) o ônibus (O) e o metrô (M), em que se acredita que o metrô e o ônibus possam ser agrupados em um nicho denominado transporte público (TP). Neste caso é aplicado um modelo LM no nicho, em que a probabilidade de escolha do metrô, dentro desse subgrupo é dada pela fórmula 17:

$$P(M/TP) = \frac{\exp(W_M)}{\exp(W_M) + \exp(W_O)} \qquad \text{(fórmula 17)}$$

Onde:
$P(M/TP)$ é a probabilidade de escolha do metrô, dentro do nicho transporte público;
W_m é a utilidade do metrô;
W_O é a utilidade do ônibus.

As utilidades são representadas pela letra W, pois elas contem somente os elementos que não são comuns aos dois modos do nicho. Se, por exemplo, a tarifa do ônibus e do metrô fosse a

mesma, ela não seria considerada nas funções W_M e W_o. Dentro do nicho, a probabilidade de escolha do ônibus é representada pela fórmula 18:

$$P(O/TP) = 1 - P(M/TP) \qquad \text{(fórmula 18)}$$

Onde:
$P(O/TP)$ é a probabilidade de escolha do ônibus, dentro do nicho transporte público;
$P(M/TP)$ é a probabilidade de escolha do metrô, dentro do nicho transporte público.

A separação da modalidade automóvel e do nicho transporte público é feita no nível superior com a aplicação de outro LM indicado na fórmula 19:

$$P(A) = \frac{\exp(U_A)}{\exp(U_A) + \exp(U_{TP})} \qquad \text{(fórmula 19)}$$

Onde:
P_A é a probabilidade de escolha do automóvel;
U_A é a utilidade do automóvel;
U_{TP} é a utilidade do transporte público.

A probabilidade de escolha do transporte público é descrita pela fórmula 20:

$$P(TP) = 1 - P(A) \qquad \text{(fórmula 20)}$$

Onde:
$P(TP)$ é probabilidade de escolha do transporte público;
$P(A)$ é a probabilidade de escolha do automóvel.

Observa-se que a utilidade do automóvel é representada por U_A, pois ela incorpora todos os atributos da opção pelo automóvel, enquanto a utilidade do nicho transporte público pode ser representada por duas componentes citadas anteriormente, sendo uma a máxima utilidade

esperada das opções dos nichos inferiores, e outra que considera os atributos comuns a todos os membros do nicho, como indicado na fórmula 21:

$$U_{TP} = \emptyset \log[\exp(W_O) + \exp(W_M)] + \sum_{\grave{k}} \alpha_{\grave{k}} z_{\grave{k}} \qquad \text{(fórmula 21)}$$

Onde:
U_{TP} é a utilidade do nicho transporte coletivo;
$\emptyset$ é um parâmetro de escala a ser estimado;
W_O é a utilidade do ônibus;
W_M é a utilidade do metrô;
α é um parâmetro de escala
z são os elementos desconsiderados para estimativa das utilidades W_O e W_M.

Os autores destacam que, ao estimar a probabilidade de escolha entre o automóvel e o nicho transporte público, o somatório de elementos z consideram todos os elementos comuns ao ônibus e ao metrô, que foram desconsiderados ao estimar as probabilidades dentro do nicho transporte público.

Por fim, a probabilidade de escolha do ônibus é a multiplicação da probabilidade de escolha do transporte público frente ao automóvel, pela probabilidade de escolha do ônibus frente ao metrô, dentro do nicho transporte público, conforme indicado na fórmula 22:

$$P(O) = P(O/TP)P(TP) = (1 - P(M/TP))(1 - P(A)) \qquad \text{(fórmula 22)}$$

Onde:
$P(O)$ é a probabilidade de escolha do ônibus;
$P(O/TP)$ é a probabilidade de escolha do ônibus, dentro do nicho transporte público;
$P(TP)$ é a probabilidade de escolha do transporte público;
$P(M/TP)$ é a probabilidade de escolha do metrô, dentro do nicho transporte público;
$P(A)$ é a probabilidade de escolha do automóvel.

Para aplicar os modelos LH e LM é necessário obter uma base de dados que represente os gostos e as preferências das pessoas. A obtenção destes dados é feita através de pesquisas de preferência revelada (PR) nas quais as pessoas descrevem experiências já realizadas, ou pesquisas de preferência declarada (PD), em que são apresentados cenários hipotéticos aos entrevistados para que eles escolham o que lhes parece mais conveniente.

Cabe ao modelador definir qual técnica será utilizada para a obtenção de dados, de acordo com o modelo que ele irá utilizar. No próximo capítulo são discutidas essas duas técnicas, apresentando suas características e limitações.

6 TÉCNICAS DE OBTENÇÃO DE DADOS

Até meados dos anos 80, acreditava-se que a modelagem por transportes deveria ser baseada em informações sobre decisões previamente tomadas e observações sobre fatos ocorridos, que são os dados obtidos nas pesquisas do tipo PR. Devido aos altos custos dessas pesquisas e à dificuldade de mensurar atributos difíceis de serem observados, surgiu a necessidade de um novo modelo (ORTÚZAR; WILLUMSEN, 1994, p. 22, tradução nossa).

Os autores também afirmam que as pesquisas do tipo PD iniciaram nas pesquisas de *marketing*, no final dos anos 70. As pesquisas de PD são baseadas em escolhas hipotéticas, podendo cobrir muitos atributos que pesquisas sobre o comportamento real não são capazes. A seguir são discutidos os dois tipos de pesquisas e suas aplicações.

6.1 PREFERÊNCIA REVELADA

O comportamento observado (ou os dados obtidos neste tipo de pesquisa) dá informações a respeito do processo de equilíbrio do atual mercado. Jordan et al. (2000, p. 228, tradução nossa) indicam que os dados dessas pesquisas a respeito do mercado já existente, possuem as seguintes características:

a) descrevem somente as alternativas que já existem. Com isso, se existir níveis absolutos de atributos e correlações entre os atributos, qualquer modelo poderá ser estimado com qualquer dado obtido;
b) uma marca (ou um tipo de modal) pode ou não estar incluída em um mercado particular, logo poderá ser difícil separar os efeitos das relações das marcas dos impactos dos atributos. Devido às experiências dos indivíduos, eles levam consigo alguns pré-conceitos a respeito das alternativas;
c) as características dos mercados e as características pessoais (como local de trabalho, renda, etc) são refletidos nos dados das pesquisas. Em alguns casos essas restrições podem não ser aceitas pelos dados das pesquisas de PR, pois eles não apresentam variabilidade suficiente.

Os dados obtidos com PR refletem escolhas reais, o que pode ser uma grande vantagem. Contudo, os dados são limitados a atributos e alternativas que já existem ou existiram em algum momento da história. Em algum momento o modelador pode querer avaliar o

comportamento das pessoas às situações que ainda não existem, como a demanda por um novo produto. Neste caso, os dados da PR seriam insuficientes (TRAIN, 2003, p. 156, tradução nossa).

O autor faz outra ressalva quanto à utilização deste método de obtenção de dados, afirmando que mesmo para situações que já existem, as variações nos fatores relevantes podem não ser conhecidas, o que impossibilitaria o uso da modelagem utilizando PR. Para exemplificar, é citado o caso de um produto que está há muitos anos no mercado com um preço fixado. Com isso, não se pode perceber a resposta dos consumidores à diferença de preço, podendo impossibilitar o uso da PR.

Em geral, as pesquisas de PR **revelam** as escolhas dos indivíduos, mas não demonstram necessidades dos que não foram atendidos, ou a opinião deles sobre aumento de preço ou de qualidade. Outra limitação é o custo destas pesquisas, que pode ser muito alto devido ao fato de que cada entrevistado revela apenas um dado, ou seja, sua experiência, necessitando, assim, de um grande número de pessoas entrevistadas (NODARI, 1996, p. 13).

6.2 PREFERÊNCIA DECLARADA

O que diferencia a PD da PR, é que os entrevistados são questionados a respeito do que eles fariam em determinada situação ou ainda de que maneira eles ordenariam algumas opções numa situação hipotética. Um problema averiguado na década de 70 foi que as opções feitas pelos indivíduos durante uma pesquisa de PD não se confirmaram quando eles, um tempo depois, se depararam com a mesma situação na realidade, optando por alternativas distintas. Isso ocorreu pois naquela época a maneira como se obtinham os dados era pouco desenvolvida. Hoje em dia esses problemas não são mais observados devido às melhorias na coleta de dados e o aumento de recursos aplicados nas pesquisas (ORTÚZAR; WILLUMSEN, 1994, p. 79, tradução nossa).

Um pequeno resumo a respeito da história do uso das PD é feito por Novaes (1995, p. 574):

> Trata-se de uma família de técnicas que utilizam as respostas das pessoas entrevistadas a respeito de suas preferências em relação a um conjunto de opções a elas oferecidas, com o objetivo de ajustar funções de utilidade. Em '*marketing*' tais técnicas são denominadas **análise conjunta** ('*conjoint analysis*'). Esses métodos foram desenvolvidos em princípios da década de 70 (pesquisa de marketing) e passaram a ser largamente utilizados em estudos de transportes a partir da década de

> 80 quando o Departamento de Transportes da Grã-Bretanha, endossou o valor monetário do tempo de viagem através de PD.

Este tipo de pesquisa tem características peculiares e algumas vantagens sobre as pesquisas de PR, como explicam Jordan et al. (2000, p. 230, tradução nossa):

> a) dados de uma pesquisa de PD conseguem distinguir um número muito maior de atributos e níveis do que as pesquisas de PR. As relações tecnológicas são capazes de representar qualquer cenário que o modelador desejar, tendendo a gerar modelos mais robustos;
>
> b) os dados de uma PD podem estimar a reação das pessoas a uma nova marca no mercado, ou novos serviços de uma marca existente, com custos baixos, comparado com o que as empresas investem em *marketing*;
>
> c) dados para determinar a divisão dos mercados, ou a fatia que uma empresa ocupa em determinado mercado, também são obtidos com pesquisas de PD.

Os autores ainda explicam que essa gama de opções de uso dos dados de PD se dá pelo fato de que em uma entrevista é possível obter muitas informações de acordo com os atributos colocados no modelo. Uma limitação levantada é que, por serem dados hipotéticos, é difícil de levar em conta dados da realidade do mercado, ao se usar PD. Esses modelos são mais aconselháveis para estimar mudanças na estrutura do que para representar os fatos do presente.

Para realizar uma pesquisa de PD, é importante ter clareza a respeito do que se deseja analisar, observar o âmbito de escolha e os fatores que podem influenciá-lo (ORTÚZAR, 2000, p. 195, tradução nossa). Jordan et al. (2000, p. 256, tradução nossa) lembram da importância de conhecer o mercado para o qual está sendo feito o modelo. Ele sugere que na fase inicial do estudo, consumidores do mercado analisado sejam convidados a realizar um grupo de discussões, a fim de observar quais seriam suas maiores necessidades.

Ortúzar (2000, p. 198, tradução nossa) explica que as variáveis a serem estudadas podem ser quantitativas, como o preço de um serviço, ou qualitativas como a qualidade de um serviço ser boa ou ruim. Para gerar os questionários da pesquisa de PD, e entender de que maneira as pessoas reagem à variação desses atributos, é preciso determinar alguns níveis de variação para cada um deles, os quantitativos, como o preço, podem variar em diversos níveis, por exemplo entre $10, $20 ou $30, por um serviço; enquanto os qualitativos podem ter somente dois níveis, bons e ruins.

O autor sugere que nesta etapa seja feita uma versão experimental do estudo, elaborando um questionário inicial a fim de determinar os valores limites que irão ser alternados nos questionários. Um exemplo deste experimento é dado por Jordan et al. (2000, p. 259, tradução nossa) para uma pesquisa realizada sobre as preferências dos usuários ao alugar um automóvel. O quadro 1 mostra esse exemplo, em que na primeira coluna estão os atributos a serem estudados, na segunda estão o tipo de atributos – qualitativos ou quantitativos – e na terceira coluna estão as opções de variação que poderão ser dadas aos entrevistados. Eles lembram que este não é um questionário definitivo, sendo feito exatamente para ser estudado e melhorado, se necessário.

Quadro 1 – Atributos e níveis em um estudo experimental

Atributo	Tipo	Nível
Marca	Qualitativo	A, B, C, D
Preço por dia	Quantitativo	Compacto: $30, $34, $38, $42 Tamanho Médio: Compacto + $2, Compacto + $4 Tamanho Grande: Compacto + $4, Compacto + $8 Carro de Luxo: Compacto + $20, Compacto + $30
Preço por quilômetro (não considerado no preço por dia)	Qualitativo e Quantitativo	Quilometragem ilimitada $0,10/km acima de 200km $0,12/km acima de 200km $0,15/km acima de 200km
Tipo de veículo	Qualitativo	Compacto: diversas marcas Tamanho Médio: diversas marcas Tamanho Grande: diversas marcas Carro de luxo: diversas marcas
Seguro Opcional	Quantitativo	$12, $16
Milhas aéreas	Quantitativo	Nenhuma (0), 500
Política de devolução do carro em relação ao uso do combustível	Qualitativo e Quantitativo	Pagar antecipadamente o tanque cheio, com desconto no preço Devolver no nível que estava quando alugado Devolver o veículo com menos combustível e pagar $0,25 por litro Devolver o veículo com menos combustível e pagar $0,50 por litro

(fonte: adaptada de JORDAN et al., 2000)

A seguir, é necessário realizar testes com esses valores, a fim de verificar sua coerência e suas relações. Neste ponto, Ortúzar (2000, p. 195, tradução nossa) explica a importância do uso de consumidores potências do mercado de estudo para verificar a validade dos dados estimados e decidir por possíveis mudanças.

Enfim, para que a pesquisa de PD possa ser realizada, é necessário um último teste: uma pesquisa piloto. Esta pesquisa tem como finalidade simular uma pesquisa de PD, mas em

menor escala, quando será testada a validade ou não da pesquisa. Esta pesquisa, de acordo com Ortúzar (2000, p. 196, tradução nossa), é importante para avaliar o efeito dos atributos na conduta do indivíduo. No exemplo apresentado pelo quadro 2, pode-se observar um modelo no qual as variáveis são o tempo de viagem, a frequência e a tarifa. Se, por exemplo, a alternativa 5 fosse melhor classificada que a alternativa 8, seria possível concluir que os indivíduos preferem um tempo de viagem menor, a custa de uma menor frequência.

Quadro 2 – Pesquisa piloto

	Atributos			
Opções	**Tarifa**	**Tempo de viagem**	**Frequência**	**Dominância**
1	Baixa	Rápido	Baixa	Dominada por 2
2	Baixa	Rápido	Alta	
3	Baixa	Lento	Baixa	Dominada por 2 e 4
4	Baixa	Lento	Alta	Dominado por 2
5	Alta	Rápido	Baixa	Dominado por 1
6	Alta	Rápido	Alta	Dominado por 2
7	Alta	Lento	Baixa	Dominado por 3
8	Alta	Lento	Alta	Dominado por 4

(fonte: adaptado de ORTÚZAR, 2000)

Nessa pesquisa é possível avaliar a dominância que cada alternativa tem sobre a outra, sendo importante para verificar se os entrevistados seguem uma lógica de escolha nas suas respostas. A alternativa 6, por exemplo, é dominada pela alternativa 2, pois os atributos de tempo de viagem e frequência são iguais, e a tarifa da alternativa 2 é baixa, enquanto a da alternativa 6 é alta. Se o indivíduo não seguir a lógica, escolhendo a alternativa 6 em detrimento da alternativa 2, é possível alterar os níveis de cada variável para que esse tipo de má interpretação da pesquisa não ocorra na pesquisa final.

Depois de ajustada a pesquisa, é necessário avaliar de que maneira os dados serão coletados. Estes dados precisam ser confiáveis, minimizando os erros provenientes dessa etapa. Os questionários podem ser respondidos face-a-face, com a presença do pesquisador, ou ainda

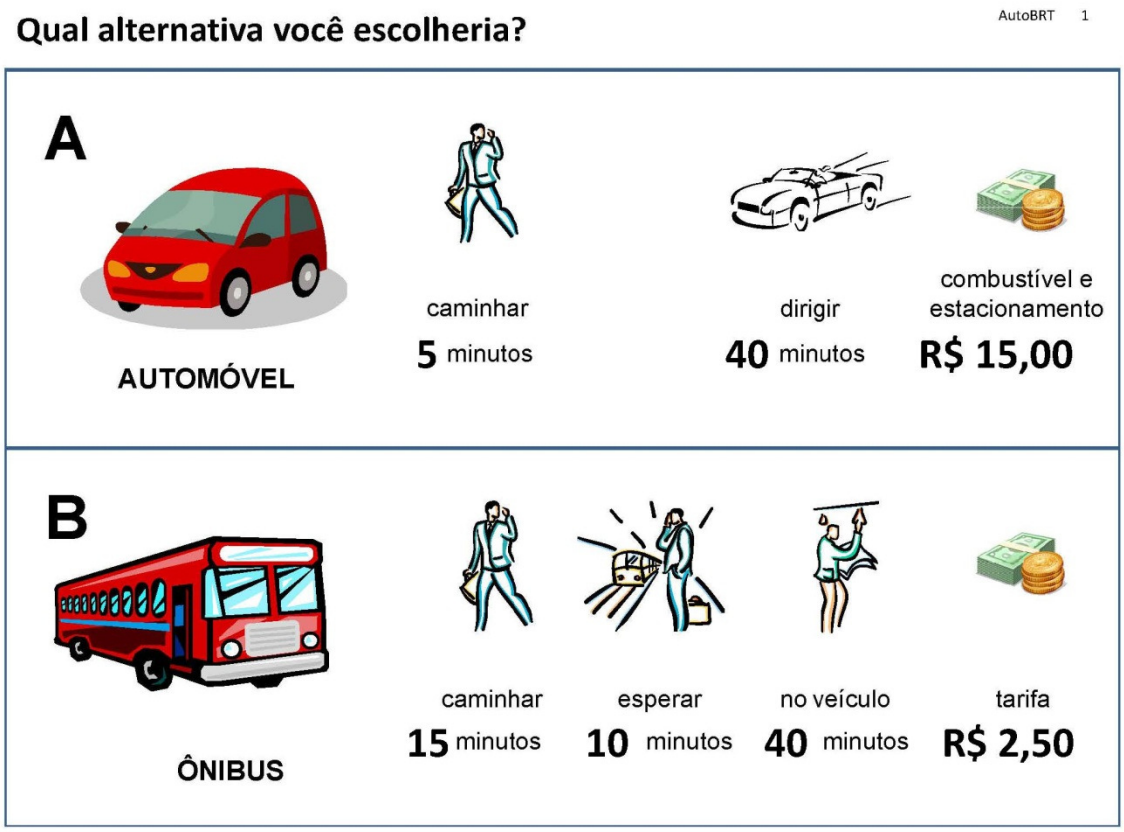
sem a presença dele, com o questionário sendo enviado por correio convencional ou eletrônico.

As pesquisas de PD são realizadas através de cartões, sendo oito ou nove, o número máximo sugerido em estudos no mundo todo. Os cartões devem apresentar figuras referentes aos atributos, para melhor visualização das alternativas. Novaes (1995, p. 578) recomenda a utilização de fotos ou ilustrações para facilitar o entendimento do entrevistado, como indicado na figura 6. Os cartões devem apresentar variações dos valores de seus atributos. No exemplo da figura 6, os próximos cartões devem ter diferentes tempos de caminhada, tempos de espera, tempos no veículo e custos de cada modal.

O mesmo autor exalta a necessidade de usar informações facilmente compreensíveis, com medidas e nomenclaturas comumente utilizada por toda população, pois os entrevistados não terão o mesmo conhecimento técnico que o planejador. É sugerida também a técnica de **ordenação dos cartões** (*rank procedure*), em que os cartões são ordenados aleatoriamente e é solicitado ao entrevistado que escolha a opção que mais lhe agrada, sendo esta retirada do conjunto para que o entrevistado repita o processo.

É possível observar que tanto a PD quanto a PR tem suas vantagens e limitações. Train (2003, p. 157, tradução nossa) sugere o uso combinado das duas pesquisas a fim de uma suprir as limitações da outra. Enquanto a PD suporta as variações necessárias dos atributos, a PR mostra como eles se comportam na realidade. Jordan et al. (2000, p. 231, tradução nossa) faz considerações a este respeito enfatizando que os dados de PR irão conter informações a respeito do equilíbrio do mercado e os atributos percebidos, porém serão dados deficientes. Com os dados de PD complementando-os, é possível estabelecer um cenário realístico e propenso a mudanças e variações de variáveis e atributos.

Figura 6 – Exemplo de cartão para uma pesquisa de preferência declarada



(fonte: Matricial Engenharia Consultiva, 2009)

Ortúzar (2000, p. 228, tradução nossa) também sugere a utilização das duas técnicas na mesma pesquisa. A PR não consegue inserir fatores qualitativos em seus modelos nem prever o comportamento dos usuários frente a cenários que ainda não existe, porém ela retrata a realidade de uma maneira confiável. Já a PD é muito eficaz em criar cenários hipotéticos, com novas alternativas e a inserção de fatores qualitativos, porém não há garantia de que os indivíduos irão, na prática, se comportar da mesma maneira que foi indicado nas pesquisas.

O autor considera que os dois métodos são complementares, e a realização de uma pesquisa que possa abranger essas duas técnicas, resultará em um modelo mais preciso e confiável.

7 CONSIDERAÇÕES FINAIS

O Brasil vive uma época diferenciada na sua história, com a proximidade da Copa do Mundo de Futebol de 2014 que será realizada neste país e dos Jogos Olímpicos de 2016 que terão sua sede na cidade do Rio de Janeiro/RJ. Esses eventos exigem investimentos em infraestrutura viária nas capitais que irão sediá-los. Depois de muito tempo há uma discussão abrangente sobre mobilidade urbana no país, o que faz com que diversos estudos sejam realizados para avaliar a viabilidade de novos projetos.

Diante deste cenário, os planejadores de transporte tem estudado a implantação de novos modais de transportes para as cidades, como o sistema de BRT (*Bus Rapid Transit*) que propõe a utilização de ônibus em corredores exclusivos e terminais de embarque e desembarque com pagamento antecipado – o que diminui significativamente o tempo de embarque dos passageiros, além da implantação de linhas de metrô em algumas cidades.

Projetos dessa magnitude exigem um estudo de viabilidade econômica, ou seja, a avaliação a respeito dos custos e dos benefícios envolvidos em cada projeto. Para que um projeto seja economicamente viável, é necessário que os benefícios, diretos e indiretos, trazidos à população sejam maiores que os custos envolvidos em sua aplicação.

Para realização desses estudos é imprescindível estimar a demanda existente por esses modos de transporte. A estimativa do número de pessoas que estaria disposta a utilizar esses modais permite avaliar os valores de tarifas aplicáveis, o tempo de retorno do investimento feito na implantação, e o número de pessoas que deixaria de utilizar o transporte individual e passaria a utilizar o transporte público.

Os modelos comportamentais desagregados são ferramentas que permitem transformar os sentimentos, as necessidades e os desejos das pessoas em dados mensuráveis, podendo, assim, avaliar quais são os atributos que um modo de transporte público precisa ter para que seja mais atrativo aos seus usuários.

Com a identificação desses atributos, é possível criar cenários hipotéticos com as características projetadas para um novo modal, estimando o número de pessoas que estaria

disposta a utilizá-lo nessas condições. Para obtenção desses dados são utilizadas as pesquisas de preferência declarada, que permitem criar cenários com diferentes valores de tarifas, ou tempos de viagem, para que seja possível ajustar este novo modo de transporte, à expectativa da população.

A revisão bibliográfica realizada nesse trabalho indicou as principais características desses modelos, bem como as limitações e restrições ao seu uso. Ficou evidenciado que não se trata de uma ferramenta trivial, sendo necessário um embasamento teórico a respeito do fenômeno que se deseja modelar, além de um profundo conhecimento matemático para entender os pressupostos do modelo.

Um aspecto importante para geração de modelos confiáveis é a observação das características das alternativas disponíveis. É importante observar a correlação dessas alternativas, e, se for o caso, utilizar nichos diferenciados através do modelo Logit Hierárquico. Se as alternativas correlacionadas forem consideradas no mesmo nível hierárquico, o modelo pode apresentar resultados que não refletem a realidade.

Além disso, é importante realizar a agregação dos dados de maneira correta. Para que o modelo reflita a realidade de maneira confiável é importante agregar os dados levando em conta as características socioeconômicas dos indivíduos. Pessoas com alto poder aquisitivo avaliam o valor do tempo de maneira diferente das pessoas de baixa renda, sendo necessário estimar um modelo diferente para cada uma delas.

Durante as pesquisas de preferência declarada, recomenda-se a obtenção das características socioeconômicas para que o modelador possa agregar os dados de maneira correta. Essa agregação pode ser feita levando em consideração a renda das pessoas, a localidade onde mora, a posse, ou não, de automóveis, entre outras características, dependendo do fenômeno que está sendo modelado

Os tópicos abordados satisfazem os objetivos propostos para este trabalho, entretanto para a aplicação prática deste modelo, recomenda-se um aprofundamento teórico, além de um estudo detalhado a respeito dos programas computacionais disponíveis no mercado.

REFERÊNCIAS

BEN-AKIVA, M.; LERMAN, S. R. **Discrete Choice Analysis**. 1st ed. London: MIT Press, 1985.

JORDAN, J. L.; HENSHER, D. A.; SWAIT, J. D. **Stated Choice Methods:** analysis and application. 1st ed. Cambridge: University Press, 2000.

NODARI, C. T. **Influência de preço e marca na demanda por transporte aéreo**. 1996. 103 f. Dissertação (Mestrado em Engenharia) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1996.

NOVAES, A. G. N. Análise de serviços de transportes com dados de Preferência Declarada. In: CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES, 4., 1995, São Carlos. **Anais...** São Carlos: Associação Nacional de Pesquisa e Ensino em Transportes, 1995. p. 573-584.

ORTÚZAR, J. de D. **Modelos econométricos de elección discreta**. 1st ed. Santiago: Ediciones Universidad Católica de Chile, 2000.

ORTÚZAR, J. de. D.; WILLUMSEN, L. G. **Modelling Transport**. 3th ed. New York: John Wiley and Sons, 1994.

_____. **Modelling Transport**. 4th ed. New York: John Wiley and Sons, 2011.

TRAIN, K. E. **Discrete Choice Methods with Simulation**. 1st ed. Cambridge: University Press, 2003.