

MODELOS DE PREVISÃO DE DEMANDA: UMA APLICAÇÃO NO TRANSPORTE RODOVIÁRIO INTERESTADUAL DE PASSAGEIROS POR ÔNIBUS NA REGIÃO SUL DO BRASIL

RESUMO

Neste artigo são apresentados modelos de previsão de demanda aplicados ao transporte rodoviário interestadual de passageiros, por ônibus, na região sul do Brasil. Esses modelos englobam variáveis sócio-econômicas juntamente com variáveis do modelo gravitacional clássico. Os resultados obtidos no processo de calibração mostraram-se satisfatórios, demonstrando que os mesmos podem servir como subsídio para o planejamento e geração de cenários de estimativas de movimentações de passageiros.

ABSTRACT

This article contains demand forecast models applied to the interstate road transport of passengers, by coach, in the south of Brazil. These models encompass socio-economic variables together with variables of the classical gravitational model. The results obtained in the calibration process proved to be satisfactory, showing that they can be of use in generating scenarios of estimates of passenger movements.

1 INTRODUÇÃO

A implantação de sistemáticas de planejamento de serviços a serem oferecidos para os usuários do transporte de passageiros no Brasil é fundamental, devido à intensa dinâmica do setor e sua representatividade nas atividades econômicas e sociais do país. Levando-se em conta que o transporte rodoviário por ônibus é o principal meio utilizado pela população brasileira surge, então, a necessidade que o planejamento de novos serviços passe por uma análise de cenários com estimativas de volume de movimentações entre as ligações do país.

É importante levar em consideração que os fluxos observados de transporte de passageiros no modal rodoviário, embora seja a grande maioria, não representam a totalidade da demanda em potencial de passageiros do contexto nacional, devido à interferência de outros modais, ou pela inexistência de serviços em todas as partes demandantes no território nacional, ou até mesmo pela renda disponível dos usuários. Entende-se então, que a demanda observada atualmente traduz o ponto de equilíbrio entre a oferta e a demanda.

Apesar disso, o Brasil supre de forma razoavelmente bem sua demanda por transporte. Pode-se dizer que os fluxos de passageiros transportados atualmente (demanda nominal), nas linhas de ônibus, refletem uma boa aproximação da demanda real. Assim, é apropriada a criação de modelos a partir dos fluxos existentes, admitindo implicitamente que representam o fluxo observado com razoável precisão à demanda.

Bez e Gonçalves (2006), em seu estudo, apresentaram modelos de previsão de demanda, que incorporam variáveis sócio-econômicas e variáveis do modelo gravitacional clássico (Wilson, 1967). Para definição desses modelos, partiram da calibração de um modelo inicial utilizando um conjunto de dados composto da movimentação de passageiros nas seções pertencentes as linhas que compõem as viagens interestaduais no território nacional.

Em razão dos resultados obtidos no processo de calibração, da extensão do território brasileiro e suas acentuadas diferenças regionais, decidiram segmentar o universo de análise por agrupamentos das ligações regionais e inter-regionais. Para esses agrupamentos fizeram uso do princípio de Pareto (Wild, 1997), utilizando como variável a movimentação anual de passageiros das seções. Da segmentação realizada surgiram 3 grupos. O primeiro grupo, composto pelo Sul, Sudeste e Nordeste era responsável por 64,43% da movimentação de passageiros; o segundo responsável por 26,98% da movimentação, era formado pelas ligações sul – sudeste, sudeste – centro-oeste, centro-oeste, sudeste – nordeste e finalmente o terceiro grupo composto pelo norte, centro-oeste – norte, sul – centro-oeste, centro-oeste – nordeste, norte – nordeste, sudeste – norte, sul – nordeste, sul – norte responsável pelos 5,59 % restantes.

A segmentação realizada possibilitou a melhoria nos resultados obtidos se comparado com os resultados encontrados com o modelo inicial. Com base neste estudo e considerando que, neste caso, a desagregação repercutiu em melhores resultados, decidiu-se usar como base o modelo proposto por Bez e Gonçalves (2006) tendo como foco, tão somente, as ligações interestaduais da região Sul e verificar a viabilidade de agregar novas variáveis a este modelo, que possibilitem uma melhoria nos resultados.

Salienta-se que a utilização de modelos econométricos que incorporam as variáveis do modelo gravitacional, vêm sendo amplamente utilizados ao redor do mundo (Wirasinghe e Kumarage, 1998; Zhao et al, 2004).

No presente artigo é apresentado um modelo de previsão de demanda do tipo gravitacional que incorpora variáveis sócio-econômicas. O modelo foi aplicado às ligações interestaduais da região sul do Brasil, por ser uma região com notório potencial demandante de transporte e com estimativas de expansão sócio-econômica.

No processo de calibração dos modelos fez-se uso de regressão linear múltipla utilizando o software Minitab no processo de ajuste. O conjunto de dados utilizado na aplicação foi obtido por intermédio de um processo de análise dos dados constantes nos anuários da ANTT, ano base de 2003.

O trabalho está estruturado com segue: na seção 2 apresenta-se o modelo de previsão de demanda proposto. A regressão linear múltipla e os procedimentos adotados na validação do modelo são descritos na seção 3. A seção 4 trata dos critérios adotados na depuração do conjunto de dados. Os resultados obtidos são apresentados na seção 5 e, finalmente, na seção 6 apresentam-se as considerações finais do trabalho.

2 MODELOS DE PREVISÃO DE DEMANDA

Neste estudo foram realizados dois ensaios. Num primeiro momento utilizou-se o modelo proposto por Bez e Gonçalves (2006), mas somente considerando as ligações interestaduais da região sul.

O modelo utilizado nos ensaios preliminares realizados por Bez e Gonçalves (2006) é do tipo gravitacional, modificado de modo a incorporar variáveis sócio-econômicas sem, no entanto, aumentar, em demasia, a complexidade do mesmo. O modelo é expresso por:

$$T_{i,j} = a_0 \frac{P_i^{a_1} P_j^{a_2} mig_i^{a_3} mig_j^{a_4} m_i^{a_5} m_j^{a_6} (r_i \cdot r_j)^{a_7} dummy_j^{a_8}}{d_{ij}^{a_9}} \quad (1)$$

em que T_{ij} : demanda anual de passageiros entre os municípios i e j ;
 P_i, P_j : populações dos municípios i e j , respectivamente;
 mig_i, mig_j : taxa de habitantes que não são naturais dos Estados dos Municípios de origem e destino;
 m_i, m_j : taxa per capita de viagens interestaduais interurbanas por ônibus dos municípios i e j , respectivamente;
 r_i, r_j : renda média per capita dos municípios i e j , respectivamente;
 $dummy$: variável que determina se o município é ou não um pólo turístico;
 d_{ij} : distância rodoviária entre os municípios i e j ; e
 $a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9$: coeficientes a serem determinados.

As variáveis mig_i e mig_j que correspondem à taxa de habitantes que não são naturais dos estados de origem e destino, respectivamente, foram introduzidas para captar a influência de movimentos migratórios na formação dos fluxos e a variável $dummy$, que determina se o município é ou não um pólo turístico, para captar a atratividade própria desses pólos (valores adotados: $dummy_j = 2$, se j é um pólo turístico e $dummy_j = 1.2$, em caso contrário). Já a renda média per capita foi adicionada ao modelo, pois pode explicar a geração de demanda ocasionada por fatores sócio-econômicos diversos (Bez e Gonçalves, 2006).

Já no segundo ensaio, o modelo proposto manteve a base do modelo (Equação 1) e foram inseridas duas novas variáveis coletadas do Instituto Brasileiro de Geografia e Estatística, que são:

S_i, S_j : representam o número de estabelecimentos de saúde (como hospitais, postos médicos, maternidades) em cada município de origem e destino, respectivamente;
 SL_i, SL_j : mediana dos salários das pessoas residentes por município de origem e destino, respectivamente.

O modelo com a inserção destas novas variáveis é expresso por:

$$T_{i,j} = a_0 \frac{P_i^{a_1} P_j^{a_2} mig_i^{a_3} mig_j^{a_4} m_i^{a_5} m_j^{a_6} (r_i \cdot r_j)^{a_7} dummy_j^{a_8} S_i^{a_{10}} S_j^{a_{11}} SL_i^{a_{12}} SL_j^{a_{13}}}{d_{ij}^{a_9}} \quad (2)$$

As variáveis S_i e S_j foram introduzidas no modelo para captar a geração de viagens em busca de recursos disponíveis na área da saúde em outros centros urbanos.

3 REGRESSÃO LINEAR MÚLTIPLA E PROCEDIMENTOS DE VALIDAÇÃO DOS MODELOS

Um modelo de regressão linear múltipla pode ser expresso por:

$$Y_i = b_1 + b_2 X_{2i} + b_3 X_{3i} + \dots + b_k X_{ki} + e_i \quad (3)$$

em que Y_i : corresponde a i-ésima variável resposta (variável dependente);

$X_{2i}, X_{3i}, \dots, X_{ki}$: variáveis explicativas ou independentes;

e_i : efeito aleatório correspondente a i -ésima variável resposta, ou seja, são efeitos de uma infinidade de fatores que estão afetando as observações Y_i 's; e $b_1, b_2, b_3, \dots, b_k$: parâmetros a serem estimados.

A idéia de estimação dos parâmetros $b_1, b_2, b_3, \dots, b_k$ é de encontrar um hiper-plano que passe o mais próximo possível dos valores definidos pelas variáveis resposta.

A regressão por mínimos quadrados consiste em minimizar a soma dos erros quadrados, ou seja, minimizar a função,

$$Z = \sum_i (Y_i - (b_1 + b_2 X_{2i} + b_3 X_{3i} + \dots + b_k X_{ki}))^2, k = 1, \dots, n \quad (4)$$

onde n representa o número de variáveis explicativas.

O estimador de mínimos quadrados para o vetor $b = [b_1, b_2, b_3, \dots, b_k]^t$ é dado por

$$b = (X'X)^{-1}(X'Y) \quad (5)$$

onde,

$$X = \begin{bmatrix} 1 & X_{21} & X_{31} & \dots & X_{k1} \\ 1 & X_{22} & X_{32} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & X_{2n} & X_{3n} & \dots & X_{kn} \end{bmatrix};$$

X' = transposta da matriz X e $(X'X)^{-1}$ = inversa da matriz $(X'X)$.

Uma vez estimados os parâmetros do modelo de regressão linear múltipla deve-se verificar o quanto as variações da variável Y , podem ser explicadas por variações das variáveis independentes X de acordo com o modelo especificado e a amostra observada. Para tal, utiliza-se o coeficiente de determinação (R^2) que representa uma medida descritiva da proporção da variabilidade de Y que pode ser explicada por X . (Barbetta, 2004)

Estatísticas como *t-student* e *valor-p* são utilizadas para avaliar a influência de cada uma das variáveis independentes utilizadas no modelo na determinação da variável resposta Y .

Deve-se, também, considerar que a escolha das variáveis para um modelo é um fator muito importante para a obtenção de bons resultados. Por este motivo, muitos autores recomendam que após definidas as variáveis que serão consideradas, se faça a seleção das mesmas utilizando os métodos *Best Subsets Regression*, ou *Stepwise Regression*. Esses métodos apresentam os melhores conjuntos eliminando variáveis correlacionadas.

Após a definição do modelo deve-se verificar a eficácia do mesmo na determinação da variável preditora. Para isso, as seguintes hipóteses devem ser verificadas:

- **Hipótese I:** Normalidade dos Resíduos.

De acordo com Gujarati (2000) os erros devem seguir uma distribuição normal, onde a mesma é definida de acordo com Barbetta (2004) como uma função de probabilidade, cujo gráfico

descreve uma curva em forma de sino, e tem a evidência que a maior probabilidade de a variável aleatória assumir valores próximos do centro.

▪ **Hipótese II:** As médias dos erros têm que ser zero.

Como numa regressão estima-se uma reta que melhor se ajusta aos dados observados, pode-se assumir que a probabilidade de um erro ser x unidades acima da reta é o mesmo que ser x unidades abaixo dela, por isso a média dos erros deve ser igual a zero. (Sartoris, 2003)

▪ **Hipótese III:** A variância dos erros deve ser constante.

Esta hipótese é conhecida como homocedasticidade, que indica que independentemente da característica dos dados, a variação dos erros deve-se manter constante. No caso em que essa hipótese não é satisfeita diz-se que os erros apresentam heterocedasticidade. Existem muitos testes na literatura para verificar se existe a heterocedasticidade. O método utilizado para verificação desta hipótese neste estudo foi o de Goldfeld e Quandt, que consiste em separar os dados em três partes e comparar suas variâncias. Gujarati (2000) define os seguintes passos que devem ser seguidos para realização deste teste, são eles:

Passo 1: Ordenação das observações de acordo com os valores de X_i , começando pelo valor mais baixo de X . No caso de mais de uma variável independente, a ordenação das observações pode ser feita a partir de qualquer uma delas.

Passo 2: Omissão das observações centrais c , e a divisão das $(n-c)$ observações restantes em dois grupos (I e II), cada um com $\frac{(n-c)}{2}$ observações. O valor de c , sugerido por Gujarati

(2000) deve ser próximo a 8 se o tamanho da amostra for aproximadamente 30, cerca de 16 se o tamanho da amostra for em torno de 60, e assim por diante. A omissão destes valores é realizada para acentuar a diferença entre os grupos de máxima e mínima variância.

Passo 3: Calcula-se a estatística apresentada na Equação 6, observando que o grupo I corresponde aos menores valores de X , enquanto o grupo II os maiores valores.

$$F = \frac{S_{II}^2}{S_I^2} \quad (6)$$

em que:

$$S_{II}^2 = \frac{SQR_{II}}{gl} = \frac{\sum e_{II}^2}{n^\circ \text{ observações da amostra II} - k}$$

$$S_I^2 = \frac{SQR_I}{gl} = \frac{\sum e_I^2}{n^\circ \text{ observações da amostra I} - k},$$

onde k é o número de variáveis explicativas, incluindo o intercepto.

▪ **Hipótese IV:** Cada variável independente X_i não pode ser combinação linear das demais.

Um modelo de regressão linear não apresenta resultados confiáveis quando existe a multicolinearidade, ou seja, presença da combinação linear de variáveis independentes. A existência de multicolinearidade indica que a variação de uma variável é totalmente influenciada por outra.

4 CONJUNTO DE DADOS

Conforme dito anteriormente, para realização desse experimento fez-se uso de um conjunto de dados fornecidos pela ANTT, ano base 2003, contendo registros de viagens interestaduais da região sul do Brasil. O conjunto de dados utilizado no processo de calibração do modelo foi obtido a partir de uma série de experimentos realizados num banco de dados, constante dos Anuários (1999 – 2003) disponibilizados pela ANTT e apresentado por Bez e Gonçalves (2006).

Neste estudo, no primeiro ensaio realizado, decidiu-se por adotar uma segmentação por distância, seguindo o mesmo procedimento adotado por Bez e Gonçalves (2006), que para cada um dos grupos apresentados na seção 1 realizou-se uma segmentação por distância. Foram três as faixas de distância consideradas, são elas: viagens curtas (A): até 400 km; viagens médias (B): entre 400 e 800 km; viagens longas (C): acima de 800 km.

Foi criado um modelo para cada uma das faixas de distância pré-definidas. Na Tabela 1 são apresentadas as quantidades de OD's de cada grupo.

Tabela 1 - Número de OD's por Grupo segmentado por distância

Grupo	Número de OD's
A	1148
B	806
C	140

Para o processo de calibração do modelo proposto (Equação 2) decidiu-se pela utilização da variável Passageiros x Quilômetros (PKm), que representa o número total de passageiros transportados, ponderado pelas distâncias percorridas na rede rodoviária coberta pelas viagens realizadas. Essa variável é muito usada como uma medida anual de produtividade.

Os dados foram divididos através dos quartis da amostra ordenada pelo PKm dos pares OD's. A tabela 2 apresenta as quantidades de OD's de cada grupo.

Tabela 2 - Número de OD's por grupo segmentado por PKm

Grupo	Número de OD's
Grupo 1	524
Grupo 2	523
Grupo 3	523
Grupo 4	524

Após a estratificação dos dados em ambos os ensaios, foram realizadas as calibrações dos modelos, e os resultados obtidos são apresentados a seguir.

5 RESULTADOS OBTIDOS

Para utilização da regressão linear múltipla os modelos apresentados nas Equações 1 e 2 foram linearizados. Em seguida, partiu-se para realização do experimento com o modelo definido pela Equação 1, considerando a segmentação por distância apresentada na seção 4. Na seleção dos melhores conjuntos de variáveis foi utilizado o método *Stepwise Regression*.

Na tabela 3 apresentam-se os conjuntos com as melhores variáveis obtidas, com o modelo em questão, para cada um dos grupos.

Tabela 3 - Variáveis por Grupo segmentado por distância – Modelo 1

Grupo	Grupo A	Grupo B	Grupo C
Variáveis	$P_i, P_j, m_i, m_j, dummy, d_{ij}$	$P_i, P_j, mig_i, m_i, m_j, dummy, d_{ij}$	$P_i, P_j, m_i, m_j, d_{ij}$
R ²	53,6	44,2	59,6

Nos testes realizados para verificar a eficácia do modelo na determinação da variável preditora verificou-se que a normalidade dos resíduos (Hipótese I) era insatisfatória para os grupos A e B conforme pode ser visto, por exemplo, na Figura 1.

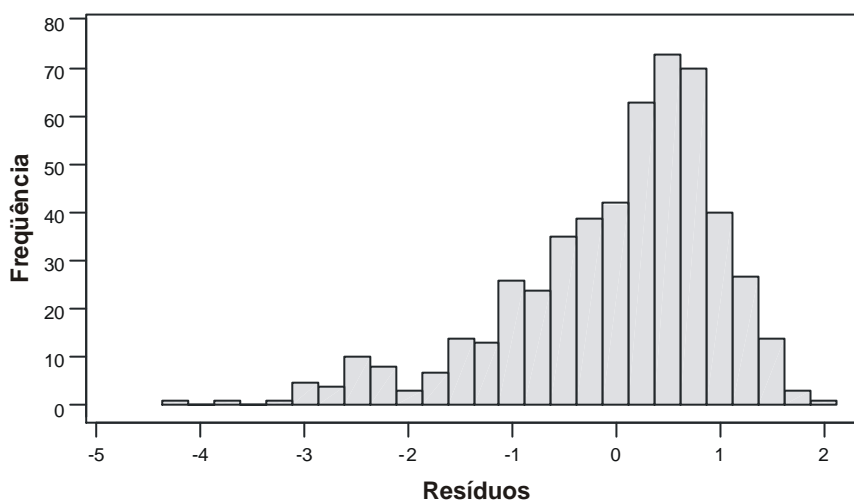


Figura 1 - Histograma dos Resíduos do Grupo B sem a Retirada dos Outliers

Para a solução deste problema, decidiu-se pela retirada dos *outliers*. O histograma que representa a normalidade dos resíduos para a nova regressão do Grupo B é apresentado na Figura 2. A normalidade dos resíduos apresentada nesta Figura para o Grupo B também foi verificada para o Grupo A com a retirada dos *outliers*.

A presença da homocedasticidade foi comprovada pelo teste F. Na tabela que segue são apresentados os valores do teste F calculado e tabelado para cada grupo.

Tabela 4 - Teste F dos Grupos segmentados por distância para testar a Homocedasticidade

Grupo	Nº de observações de maior valor	Nº observações centrais	Nº de observações de menor valor	teste F	Graus de Liberdade
Grupo A	421	306	421	1,64	7
Grupo B	296	214	296	1,24	8
Grupo C	51	38	51	0,51	6

O valor do teste F tabelado para 6, 7 e 8 graus de liberdade é 4.25, 3.79 e 3.44, respectivamente, e como se pode perceber todos os valores F calculados para os grupos resultaram em valores inferiores. Assim, a hipótese III é satisfeita e pode-se afirmar que os modelos dos três grupos não são heterocedásticos.

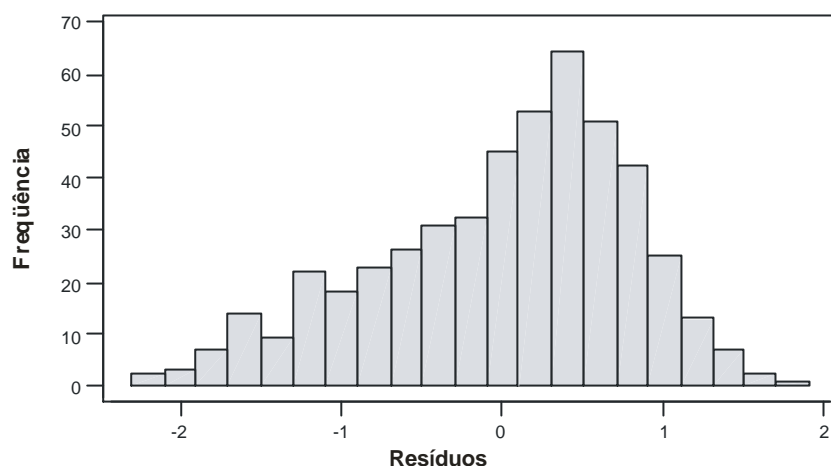


Figura 2 - Histograma dos Resíduos do Grupo B com a Retirada dos *Outliers*

As demais hipóteses foram satisfeitas para os três grupos e os novos valores de R^2 , com a retirada dos *outliers* dos Grupos A e B, passaram a ser 58,2% e 51,8%, respectivamente.

O modelo (Equação 1) linearizado e os valores dos coeficientes são apresentados na Equação 7 e Tabela 5.

$$T_{ij} = \exp \left(\begin{aligned} & \mathbf{a}_0 + \mathbf{a}_1 \ln \left(\frac{P_i}{1000} \right) + \mathbf{a}_2 \ln \left(\frac{P_j}{1000} \right) + \mathbf{a}_3 \ln(mig_i) + \mathbf{a}_4 \ln(mig_j) + \mathbf{a}_5 \ln(m_i) + \mathbf{a}_6 \ln(m_j) + \\ & + \mathbf{a}_7 \ln \left(\frac{r_i \cdot r_j}{10000} \right) + \mathbf{a}_8 \ln(dummy) + \mathbf{a}_9 \ln(d_{ij}) \end{aligned} \right) \quad (7)$$

Tabela 5 - Parâmetros dos Grupos A, B e C

Parâmetros	Grupo A	Grupo B	Grupo C
\mathbf{a}_0	6,850892241678330	5,335155703870080	4,481650846853420
\mathbf{a}_1	0,764498787020939	0,675309837851190	1,045478777642670
\mathbf{a}_2	0,455148874104415	0,417425092929715	0,345442261916725
\mathbf{a}_3	0	0,190630689594982	0
\mathbf{a}_4	0	0	0
\mathbf{a}_5	0,408323579277006	0,198470060245943	0,373019026002290
\mathbf{a}_6	0,340512562694362	0,242773864722629	0,511399691152886

a_7	0	0	0
a_8	2,251305483790870	1,440479473282330	0
a_9	-0,904697182151795	-0,548565985675699	-0,528576710844786

O experimento seguinte foi realizado com o Modelo apresentado pela Equação 2 levando em conta a segmentação definida na seção anterior. Inicialmente, foi utilizado o método *Stepwise Regression* para definição do melhor conjunto de variáveis em cada um dos grupos. As variáveis selecionadas em cada grupo com seus respectivos valores de R^2 são apresentadas na Tabela 6.

Tabela 6 - Variáveis por Grupo segmentado por *PKm* – Modelo 2

Grupo	Grupo 1	Grupo 2	Grupo 3	Grupo 4
Variáveis	$d_{ij}, P_i, P_j, m_i, m_j, r_i, r_j$	d_{ij}, P_i, mig_i, m_i	$d_{ij}, P_i, P_j, m_i, m_j$	$d_{ij}, P_i, dummy, s_i, s_j, m_i$
R^2	46,5	81,8	80,8	66,1

Os resíduos para os grupos 1 e 4 contrariaram a Hipótese I e, por esta razão, decidiu-se pela retirada dos *outliers* para tentar amenizar este problema. Esta retirada garantiu a normalidade dos resíduos num nível aceitável.

A homocedasticidade foi confirmada pelo teste F. A tabela 7 apresenta o número de observações em cada amostra e o valor do teste F calculado.

Tabela 7 - Teste F dos Grupos segmentados por *PKm* para testar a Homocedasticidade

Grupo	Nº de obs de maior valor	Nº obs centrais	Nº de obs de menor valor	Teste F	Graus de Liberdade
Grupo 1	96	70	96	0,25	7
Grupo 2	96	71	96	0,88	5
Grupo 3	96	71	96	1,06	6
Grupo 4	96	70	96	1,13	7

O valor do teste F tabelado para 5 graus de liberdade é 5,05, para 6 graus é 4,25 e para 7 graus é 3,79. Percebe-se que todos os valores F calculados, para os grupos, resultaram em valores inferiores o que valida a homocedasticidade.

As demais hipóteses foram satisfeitas para todos os grupos e os novos valores de R^2 , com a retirada dos *outliers* dos Grupos 1 e 4, passaram a ser 57,9% e 70,6%, respectivamente.

O modelo (Equação 2) linearizado e os valores dos coeficientes são apresentados na Equação 8 e Tabela 8.

$$T_{ij} = \exp \left(\alpha_0 + \alpha_1 \ln \left(\frac{P_i}{1000} \right) + \alpha_2 \ln \left(\frac{P_j}{1000} \right) + \alpha_3 \ln(mig_i) + \alpha_4 \ln(mig_j) + \alpha_5 \ln(m_i) + \alpha_6 \ln(m_j) + \alpha_7 \ln \left(\frac{r_i \cdot r_j}{10000} \right) + \alpha_8 \ln(dummy) + \alpha_9 \ln(d_{ij}) + \alpha_{10} \ln(S_i) + \alpha_{11} \ln(S_j) + \alpha_{12} \ln(SL_i) + \alpha_{13} \ln(SL_j) \right) \quad (8)$$

Tabela 8 - Parâmetros dos Grupos 1, 2, 3 e 4

Parâmetro	Grupo 1	Grupo 2	Grupo 3	Grupo 4
α_0	8,030992296257670	11,071130905920700	11,936280298319700	12,114701854816700
α_1	0,191914016056608	0,063604253815627	0,051764826297309	0,396265439713697
α_2	0,178759031803144	0	0,045242781698390	0
α_3	0	0,032339815785431	0	0
α_4	0	0	0	0
α_5	0,121596301977093	0,020820373321580	0,029575328693270	0,206658942276043
α_6	0,132636110934372	0	0,026664363350537	0
α_7	0,251375545766075	0	0	0
α_8	0	0	0	0,901774217647802
α_9	-	-1,014641829432630	-1,022117725930970	-1,287344712561040
α_{10}	0	0	0	0,211018060515865
α_{11}	0	0	0	0,255820772322994
α_{12}	0	0	0	0
α_{13}	0	0	0	0

Tendo em vista estes resultados, os modelos ora propostos se mostram coerentes ao propósito do estudo.

6 CONSIDERAÇÕES FINAIS

No presente artigo foram apresentados dois ensaios para a estimação de demanda de passageiros por ônibus, para todas as ligações inter-estaduais da região sul do Brasil. No primeiro modelo, foram mantidas as características de estratificação por distância e as mesmas variáveis utilizadas no estudo realizado por Bez e Gonçalves (2006). Já no segundo modelo, a estratificação dos dados foi realizada utilizando como unidade observacional a variável PKm e não mais a distância. Além disso, foram inseridas duas novas variáveis explicativas.

A partir da escolha dos modelos citados, foram obtidos os melhores conjuntos de variáveis para cada estratificação realizada através do método *Stepwise Regression*. Em seguida, os parâmetros foram estimados através do método de regressão linear múltipla. A verificação da credibilidade dos resultados foi feita a partir da análise das hipóteses básicas de regressão. No entanto, a hipótese que trata da normalidade dos resíduos, para alguns dos grupos, não se mostrou satisfatória. Por esse motivo, decidiu-se pela retirada dos *outliers* para amenizar este problema e assim, validar esta hipótese.

Dado que os dois modelos propostos neste trabalho utilizaram uma metodologia adequada e resultados concisos para sua utilização, os mesmos poderão ser ferramentas na previsão de demandas no transporte de passageiros na região sul do país, dentro do contexto de planejamento de transportes.

Ferramentas como essas podem ser utilizadas para auxiliar gestores de órgãos governamentais na avaliação dos níveis de competitividade, por intermédio de uma análise de dimensionamento de fluxos e da avaliação da viabilidade de ligações de transporte rodoviário, na busca da redução de custos e descentralização do setor com o objetivo de obter melhorias na qualidade dos serviços prestados.

Agradecimentos

Os autores agradecem à ANTT (Agência Nacional de Transportes terrestres), ao LabTrans - UFSC (Laboratório de Transportes da Universidade Federal de Santa Catarina) pela cooperação no trabalho desenvolvido, também ao CNPq e à FAPESC por patrocinarem parcialmente o presente trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

ANTT – Agência Nacional de Transportes Terrestres. <<http://www.antt.gov.br/>>. Acessado em 03/12/2006.

BARBETTA, Pedro Alberto; REIS, Marcelo Menezes; BORNIA, Antonio Cezar. Estatística para Cursos de Engenharia e Informática. Ed. Atlas. São Paulo, 2004.

BEZ, E. T. ; GONÇALVES, M. B. . Utilização da regressão de erros absolutos na calibração de modelos de distribuição de viagens.. In: XX Congresso de pesquisa e ensino em transportes, 2006, Brasília. Panorama Nacional da Pesquisa em Transportes 2006. Rio de Janeiro : ANPET, 2006. v. I. p. 455-461.

GUJARATI, D. N. Econometria Básica. 3. ed. São Paulo: Makron Books do Brasil, 2000.

IBGE – Instituto Brasileiro de Geografia e Estatística. <<http://www.ibge.org.br/>>. Acessado em 25/11/2006.

LABTRABS/UFSC – Laboratório de Transportes e Logística da Universidade Federal de Santa Catarina. “Relatório 3 – Calibração do Modelo de Previsão de Demanda” Convênio 018/2004 – Modelo de viabilidade, monitoramento e representação dos indicadores de desempenho das linhas de transporte rodoviário de passageiros, entre a UFSC e a ANTT. Florianópolis, 2005.

SARTORIS, Alexandre. Estatística e Introdução à Econometria. Ed. Saraiva. São Paulo, 2003.

WILSON, A. G. (1967) A Statistical Theory of Spatial Distribution Models. Transportation Research, v. 1, p. 253-269.

WIRASINGHE, S. C. e KUMARAGE, A.S. (1998) An aggregate demand model for intercity passenger travel in Sri Lanka, Transportation, v. 25, p. 77-98.

ZHAO F., CHOW L. F., Li, M. T. e GAN A. (2004) Refinement of Fsutms Trip Distribution Methodology. Final Report, Prepared by Lehman Center for Transportation Research, Department of Civil & Environmental Engineering, Florida International University.