

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS
DEPARTAMENTO DE MATEMÁTICA

Paôla Tatiana Felippi Tomé

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

Monografia

Florianópolis
2007

Paôla Tatiana Felippi Tomé

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

Trabalho de Conclusão de Curso
apresentado ao Curso de Matemática –
Habilitação Licenciatura, Departamento de
Matemática, Centro de Ciências Físicas e
Matemáticas, da Universidade Federal de
Santa Catarina.

Orientador: Prof. Edson Tadeu Bez, Dr.

Florianópolis

2007

Paôla Tatiana Felippi Tomé

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

Esta monografia foi julgada e adequada como **TRABALHO DE CONCLUSÃO DE CURSO**, no curso de Matemática – Habilitação Licenciatura, e aprovada em sua forma final pela Banca Examinadora Designada pela Portaria nº. 22/CFM/07

Florianópolis, 26 de Fevereiro de 2007.

Profª Carmem Suzane Comitre Gimenez
Professor da disciplina

BANCA EXAMINADORA

Prof. Edson Tadeu Bez, Dr.
Orientador

Profª, Joana B. O. Quandt, Dra.
Professor Membro

Profª Sônia Palomino, Dra.
Professor Membro

Ao meu filho Matheus.

AGRADECIMENTOS

À Deus por ter me dado força saúde e disposição para nunca esmorecer.

À UFSC, por propiciar os meios para a realização deste trabalho.

Ao meu orientador, Edson Bez, pela forma dedicada e profissional pela qual conduziu seu trabalho de orientador.

Aos meus pais, pelo amor, confiança e apoio incondicional.

Ao meu irmão Leky, pelo apoio e amizade.

Ao Tiago meu namorado pelo companheirismo, compreensão e ajuda.

À toda a minha família, em especial aos meus avós, que sempre foram exemplos de vida.

Aos meus mais que colegas, grandes amigos que conquistei ao longo da graduação: Dehia, Galleli, Márian, Lucas, Siri, Sabiá, Petúnia, Vivi, Natan, Bob, Ivan, Né, Pati, Poncho, Elia e demais amigos da matemática, o meu muito obrigado pela ajuda, amizade e companheirismo.

Ao Labtrans pelas trocas de experiências proporcionadas durante o período de realização das atividades, e pelo apoio dos colegas de trabalho.

RESUMO

O transporte rodoviário de passageiros é um dos serviços públicos essenciais para todos os países. No caso do Brasil, sua relevância se torna ainda mais preponderante devido às condições sócio-econômicas e culturais em que o país se encontra. Tal fato pode ser observado, tendo em vista que o transporte rodoviário por ônibus é o principal meio utilizado pela população em todo o espaço territorial brasileiro.

A implantação de sistemáticas de planejamento de serviços a serem oferecidos para seus usuários é fundamental, devido à intensa dinâmica do setor. Tem-se que o planejamento de novos serviços necessita então da geração de cenários com estimativas de volume de movimentações entre as ligações do país.

Neste trabalho de conclusão de curso estão expostas algumas propostas de modelos de previsão de demanda, que englobam algumas variáveis sócio-econômicas juntamente com variáveis de modelos clássicos, com o objetivo de estimar a demanda de passageiros nas ligações intra-regionais ocorridas no sul do país.

Os resultados obtidos mostram-se satisfatórios e coerentes, podendo os mesmos, serem utilizados por órgãos do governo e estudiosos na geração de cenários de estimativas de movimentações de passageiros do transporte rodoviário interestadual por ônibus.

PALAVRAS-CHAVE: Transporte Público, Modelos de Distribuição de Viagens, Modelos de Previsão de Demanda.

SUMÁRIO

LISTA DE FIGURAS	I
LISTA DE TABELAS	II
1 Introdução	1
1.1 Considerações Iniciais	1
1.2 Justificativa.....	1
1.3 Objetivo Geral	2
1.4 Objetivos Específicos	2
1.5 Metodologia.....	3
1.6 Estrutura do Trabalho.....	3
2 Modelos de Distribuição de Viagens	5
2.1 Introdução	5
2.2 Modelos Clássicos	6
2.2.1 Fatores de Crescimento	7
2.2.2 Modelo Gravitacional.....	8
2.2.3 Modelos de Oportunidades Intervenientes.....	11
2.2.4 Modelos Híbridos	13
2.3 Modelos Econométricos	14
2.3.1 Introdução	14
2.3.2 Concepção	14
3 Regressão Linear.....	17
3.1 Introdução	17
3.2 Concepção	17
3.2.1 Regressão Linear Simples	17
3.2.2 Regressão Linear Múltipla.....	20
3.2.3 Modelos Linearizáveis.....	24
3.3 Hipóteses para Regressão	25
3.4 Poder de Explicação do Modelo.....	28
4 Aplicação do Modelo.....	30
4.1 Modelo Previamente Idealizado	30
4.1.1 Introdução	30
4.1.2 Escolha do Modelo.....	31
4.1.3 Calibração	33
4.2 Criação e Concepção de um Novo Modelo.....	42
4.2.1 Escolha do Modelo.....	42
4.2.2 Calibração	42
4.3 Comparação dos Resultados	50
5 Considerações Finais.....	52
Bibliografia.....	53

Anexos	56
Anexo 1: Tabela de Distribuição F	57
Anexo 2: Métodos Stepwise Regression, Best Subsets Regression e a Regressão para o Grupo A.....	58
Anexo 3: Métodos Stepwise Regression, Best Subsets Regression e a Regressão para o Grupo B.....	61
Anexo 4: Métodos Stepwise Regression, Best Subsets Regression e a Regressão para o Grupo C.....	64
Anexo 5: Regressão Grupo A com a retirada de outliers	66
Anexo 6: Regressão Grupo B com a retirada de outliers	67
Anexo 7: Métodos Stepwise Regression, Best Subsets Regression e a Regressão para o Grupo 1	68
Anexo 8: Métodos Stepwise Regression, Best Subsets Regression e a Regressão para o Grupo 2	70
Anexo 9: Métodos Stepwise Regression, Best Subsets Regression e a Regressão para o Grupo 3	72
Anexo 10: Métodos Stepwise Regression, Best Subsets Regression e a Regressão para o Grupo 4.....	74
Anexo 11: Regressão Grupo 1 com a retirada de outliers	77
Anexo 12: Regressão Grupo 2 com a retirada de outliers	78
Anexo 13: Regressão Grupo 3 com a retirada de outliers	79

LISTA DE FIGURAS

Figura 1 - Matriz de Distribuição de Viagens.....	6
Figura 2 - Oportunidades Intervenientes	13
Figura 3 - Reta de Regressão	18
Figura 4 - Visualização residual 1	26
Figura 5 - Visualização residual 2	26
Figura 6 - Visualização das Ligações Interestaduais do Sul do País	31
Figura 7 - Histograma dos Resíduos do Grupo A sem a Retirada dos Outliers	36
Figura 8 - Histograma dos Resíduos do Grupo B sem a Retirada dos Outliers	37
Figura 9 - Histograma dos Resíduos do Grupo A com a Retirada dos Outliers	38
Figura 10 - Histograma dos Resíduos do Grupo B com a Retirada dos Outliers	38
Figura 11 - Histograma dos Resíduos do Grupo C sem a Retirada dos Outliers	39
Figura 12 - Histograma dos Resíduos do Grupo 1 sem a Retirada dos Outliers.....	44
Figura 13 - Histograma dos Resíduos do Grupo 1 com a Retirada dos Outliers.....	45
Figura 14 - Histograma dos Resíduos do Grupo 2 sem a Retirada dos Outliers.....	45
Figura 15 - Histograma dos Resíduos do Grupo 2 com a Retirada dos Outliers.....	46
Figura 16 - Histograma dos Resíduos do Grupo 3 sem a Retirada dos Outliers.....	46
Figura 17 - Histograma dos Resíduos do Grupo 3 com a Retirada dos Outliers.....	47
Figura 18 - Histograma dos Resíduos do Grupo 4 sem a Retirada dos Outliers.....	48

LISTA DE TABELAS

Tabela 1 - Exemplo do Método dos Mínimos Quadrados para Regressão Linear Simples	19
Tabela 2 - Exemplo do Método dos Mínimos Quadrados para Regressão Linear Múltipla.....	21
Tabela 3 - Número de OD's por Grupo segmentado por distância.....	35
Tabela 4 - Variáveis por Grupo segmentado por distância	35
Tabela 5 - Teste F dos Grupos segmentados por distância para testar a Homocedasticidade.....	40
Tabela 6 - Número de OD's por grupo segmentado por PKm.....	43
Tabela 7 - Variáveis por Grupo segmentado por PKm.....	43
Tabela 8 - Teste F dos Grupos segmentados por PKm para testar a Homocedasticidade.....	48
Tabela 9 - Valor de R^2 de cada Grupo.....	51

1 INTRODUÇÃO

1.1 Considerações Iniciais

Tendo em vista a importância do modal rodoviário no Brasil, a dinâmica do setor de transporte rodoviário de passageiros tem demandado a implantação de uma sistemática ágil de planejamento dos serviços ao usuário. A análise da viabilidade de linhas de transporte, bem como a geração de cenários com o volume de movimentações e o padrão de tarifação projetado para viagens entre pólos de transporte, são os aspectos mais relevantes a serem considerados em um aplicativo desse tipo.

De acordo com a Agência Nacional de Transportes Terrestres (ANTT, 2006) o serviço interestadual, em especial, é responsável por quase 95% do total dos deslocamentos realizados no País, o que corresponde a, aproximadamente, uma movimentação superior a 140 milhões de usuários/ano. Sua participação na economia brasileira é expressiva, assumindo um faturamento anual superior a R\$ 2,5 bilhões na prestação dos serviços regulares prestados pelas empresas permissionárias, onde são utilizados 13.400 ônibus. Para um País com uma malha rodoviária de aproximadamente 1,8 milhões de quilômetros, sendo 146 mil asfaltados (rodovias federais e estaduais), a existência de um sólido sistema de transporte rodoviário de passageiros é vital.

1.2 Justificativa

A distribuição espacial das atividades humanas sempre foi uma fonte de preocupação para os indivíduos, e às instituições que, de algum modo, estão envolvidas com o planejamento e a organização do espaço sobre o qual exercem influência (ULYSSEÁ NETO, 1974 apud GONÇALVES, 1992).

De acordo com Gonçalves (1992), é necessário um esforço grande no sentido de compatibilizar as características do sistema de transportes com os padrões da demanda de viagens, durante o período de tempo correspondente ao horizonte de

planejamento adotado. Isso torna de grande importância à previsão da demanda futura, bem como o conhecimento de suas características.

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. O planejamento de novos serviços necessita então da geração 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 sejam a grande maioria, não representam a totalidade da demanda em potencial de passageiros do contexto nacional, isso 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.

1.3 Objetivo Geral

Propor um modelo de previsão de demanda do transporte rodoviário de passageiros para linhas interestaduais na região sul do Brasil, com o intuito de auxiliar em futuras ações por parte de organismos governamentais.

1.4 Objetivos Específicos

- Fazer um levantamento dos principais modelos utilizados para distribuição de viagens em planejamento de transportes;

- Identificar as variáveis independentes significativas que compõem o modelo de previsão de demanda, e estimar os melhores parâmetros através do método de Regressão Linear;
- Verificar se os modelos propostos atendem todas as hipóteses necessárias para a regressão;
- Propor novas variáveis que possam acrescentar resultados significativos para os modelos ora propostos.

1.5 Metodologia

Inicialmente é apresentado um breve histórico e algumas definições dos principais tipos de modelos utilizados para estimar viagens, condizentes com o referido caso. Procura-se mostrar a importância do conhecimento prévio desses métodos para uma compreensão melhor do estudo.

O modelo proposto é do tipo econométrico, incorporando variáveis sócio-econômicas e variáveis do modelo gravitacional clássico. Através da implementação de variáveis sócio-econômicas, importantes na formação de viagens, procura-se definir relações de causa e efeito para explicar a distribuição de viagens e assim, obter resultados mais relevantes.

O procedimento de calibração é feito através de uma regressão linear múltipla, utilizando o software MINITAB no processo de ajuste. O conjunto de dados utilizado na calibração é constante do Anuário Estatístico de 2003, da Agência Nacional de Transportes Terrestres.

1.6 Estrutura do Trabalho

O presente trabalho está estruturado da seguinte forma:

No Capítulo 1 são apresentadas algumas considerações iniciais, seguida da justificativa, objetivo geral e específicos e a metodologia aplicada. Os modelos de distribuição de viagens clássicos e econométricos são apresentados no Capítulo 2.

No Capítulo 3 é apresentada a concepção e as hipóteses do modelo de regressão linear. A aplicação do modelo e as considerações finais são apresentadas nos capítulos 4 e 5, respectivamente.

2 MODELOS DE DISTRIBUIÇÃO DE VIAGENS

2.1 Introdução

De acordo com Bruton (1979), a estrutura metodológica para o planejamento de transportes pode ser subdividida em quatro etapas: geração de viagens, distribuição espacial de viagens, divisão ou repartição modal e alocação de tráfego.

Este estudo trata somente das técnicas de estimação das demandas futuras, também chamada de distribuição de viagens. A estimação relaciona o número de viagens de origem em uma determinada zona com o número de viagens de destino em outra. A etapa de distribuição de viagens refere-se ao estabelecimento de ligações entre várias zonas para o qual os cálculos de geração de viagens foram realizados.

Dentro deste contexto é importante saber a definição de modelo, que para Novaes (1982), trata-se de uma observação da realidade, analisada através de uma ótica coerente com uma determinada postura teórica, onde procura-se identificar os elementos e relações relevantes do sistema. Com esses elementos e relações estruturados através de uma linguagem formal (matemática), constrói-se uma representação do sistema real.

No planejamento de transportes, os Modelos de Distribuição de Viagens tem como papel fundamental estimar o número de viagens realizadas entre pares de zonas de origem-destino (OD) (Figura 1), dentro de uma área de estudo.

O \ D	1	2	· · ·	n	O_i's
1	–	T_{12}	· · ·	T_{1n}	O_1
2	T_{21}	–	· · ·	T_{2n}	O_2
·	·	·		·	·
·	·	·		·	·
·	·	·		·	·
n	T_{n1}	T_{n2}	· · ·	–	O_n
D_j's	D_1	D_2	· · ·	D_n	T

Figura 1 - Matriz de Distribuição de Viagens
 Fonte: Gonçalves (1992)

Os O_i 's e D_j 's da matriz de distribuição de viagens, correspondem ao número de viagens geradas nas zonas de origem "i" e atraídas pelas zonas de destino "j" respectivamente, e T_{ij} é o número de viagens entre as zonas i e j.

2.2 Modelos Clássicos

Ao longo dos anos, vários Modelos foram desenvolvidos para se estimar os fluxos interzonais. E, em virtude disso, esses Modelos foram classificados em dois grandes grupos:

- a) Modelos de Fatores de Crescimento: São modelos que se baseiam na suposição de que o padrão atual de viagem pode ser projetado no futuro, usando-se valores previstos da taxa zonal de crescimento (BRUTON, 1979). Entre os modelos contemplados neste grupo, podemos citar: modelo de fator de crescimento uniforme, fator médio de crescimento, Detroit e Fratar.
- b) Modelos Sintéticos: Segundo BRUTON (1979), são modelos baseados nas seguintes proposições:

- a. Antes que os padrões de viagens futuras possam ser previstos, deve-se entender os fatores que causam os movimentos;
- b. As relações causais que fornecem os padrões de movimento podem ser mais bem entendidas se elas forem consideradas semelhantes a certas leis de comportamento físico.

A seguir serão apresentadas as principais características dos Modelos de Fatores de Crescimento e o modelo Gravitacional, de Oportunidades e modelos híbridos, que se encaixam nos Modelos Sintéticos.

2.2.1 Fatores de Crescimento

Os Modelos de Fatores de Crescimento de acordo com Bruton (1979) são apresentados por meio de quatro métodos, são eles: Fator Uniforme, Fator médio, Fratar e Detroit.

Cada um desses métodos baseia-se na suposição de que o padrão atual de viagem pode ser projetado no futuro, usando-se valores previstos da taxa zonal de crescimento. Modelos desta classe são representados, em linhas gerais, pela seguinte fórmula:

$$T_{ij} = t_{ij}E \quad (2.1)$$

onde:

T_{ij} = número de viagens futuras estimadas da zona i até a zona j;

t_{ij} = número de viagens observadas da zona i até a zona j;

E = fator de crescimento

Para cada um dos métodos, o fator de crescimento pode ser simples ou até mesmo uma combinação de vários fatores.

O fator de crescimento pode ser calculado para a área como um todo ou, para qualquer número de zonas dentro dela e, então, aplicado a uma matriz de origem e destino para a área de estudo.

É importante destacar a relevância do modelo de Fratar, por ser o mais popular entre os citados, onde o mesmo apresenta como premissa básica, que a distribuição de viagens futuras, a partir de uma zona, é proporcional a distribuição de viagens atuais, alterada pelos fatores de crescimento das zonas consideradas (HUTCHINSON, 1979 apud BEZ, 2000).

2.2.2 Modelo Gravitacional

Os Modelos Gravitacionais originaram-se a partir da Lei da Gravitação Universal enunciada por Isaac Newton, em 1686, da seguinte forma: “*Dois partículas materiais se atraem com uma força diretamente proporcional ao produto de suas massas e inversamente proporcional ao quadrado da distância entre elas, estando a força dirigida segundo a reta que as une*” (BEZ, 2000).

Baseada nesta Lei, a primeira formulação do Modelo Gravitacional foi concebida da seguinte forma:

$$T_{ij} = k \frac{O_i D_j}{d_{ij}^2} \quad (2.2)$$

onde:

T_{ij} = número de viagens entre as zonas i e j

k = constante de ajuste

O_i = número de viagens que se originam na zona i

D_i = número de viagens que se destinam à zona j

d_{ij} = distância entre as zonas i e j

Historicamente Carey, em 1858, Ravenstein em 1885 e Lill em 1889 foram os primeiros estudiosos que utilizaram esses Modelos para analisar os fluxos migratórios no comércio varejista (BEZ, 2000 apud GONÇALVES, 1992).

De acordo com Gonçalves (1992) em 1920, Pallin e em 1926 Reilly utilizaram o Modelo Gravitacional também para viagens intermunicipais pelo comércio varejista. No entanto, o grande impulso se deu na década de 50 quando Voorhees, utilizando o modelo criado por Reilly, propôs que o expoente da distância, utilizada até então como 2 em função da Lei da Gravitação, deveria ser determinado pelo propósito da viagem.

Dessa forma, o Modelo não levaria em conta apenas à distância e a “massa” da população, mas, também, a atratividade das zonas ou uma resistência perante a viagem.

Segundo Gonçalves (1992), uma inconsistência teórica desta equação é o fato de que se o número de viagens que se originam na zona i e o número de viagens que se destinam à zona j forem duplicados, o valor de T_{ij} torna-se quadruplicado. Para se eliminar esta “deficiência” foram introduzidas as restrições a seguir:

$$\sum_j T_{ij} = O_i \quad (2.3)$$

e

$$\sum_i T_{ij} = D_j \quad (2.4)$$

As equações (2.3) e (2.4) especificam que a soma das linhas e colunas da matriz de viagens é igual ao total de viagens geradas e atraídas em cada zona, respectivamente, e são chamadas equações de consistência de fluxos. (BEZ, 2000)

Modelos que adotam uma destas restrições são denominados de Simplesmente Restrito ao passo que, aqueles que incorporam as duas restrições são denominados de Duplamente Restrito.

Em 1955, Casey adaptando o Modelo de Reilly fez a primeira aplicação real do Modelo. A partir desta data várias formulações foram feitas para este Modelo conforme a necessidade da aplicação, com objetivo de analisar os fluxos de compra, no comercio varejista, entre cidades (GONÇALVES, 1992).

De acordo com Bez (2000), na publicação de Wilson em 1967, o autor desenvolveu um enfoque para a construção do modelo gravitacional, baseado no princípio da maximização da entropia da teoria de informação, e introduziu uma terceira restrição ao modelo, dada por:

$$\sum_i \sum_j T_{ij} c_{ij} = C \quad (2.5)$$

onde, c_{ij} é o custo generalizado de viagem da zona i à zona j e C é o custo total do sistema.

Tem-se que o modelo gravitacional entrópico duplamente restrito apresentado por Wilson se dá da forma que segue:

$$T_{ij} = A_i B_j O_i D_j f(c_{ij}) \quad (2.6)$$

onde,

$$A_i = \left[\sum_j B_j D_j f(c_{ij}) \right]^{-1} \quad (2.7)$$

$$B_j = \left[\sum_i A_i O_i f(c_{ij}) \right]^{-1} \quad (2.8)$$

De acordo com Bez (2000) a expressão $f(c_{ij})$ foi introduzida, tendo em vista que d_{ij} apresentada em modelos anteriores, passou a ser interpretada como uma medida geral de impedância entre as zonas, podendo ser medida através do tempo de viagem, distância real, custo generalizado, ou outro fator.

Salienta-se que, as variações de f mais utilizadas são da forma exponencial negativa ou potência.

2.2.3 Modelos de Oportunidades Intervenientes

Segundo Gonçalves (1992), a base teórica dos Modelos de Oportunidades Intervenientes deve-se à Stouffer que em 1940 aplicou esse Modelo para descrever problemas de migração, em um estudo sociológico sobre o deslocamento de famílias dentro do distrito metropolitano de Cleveland nos Estados Unidos, formulando a hipótese de que a mobilidade não tem necessariamente uma relação explícita com a distância.

Stouffer (1940) acreditava que os processos de distribuição espacial não estavam diretamente ligados à distância, mas sim à acessibilidade das pessoas aos destinos. Ele afirmava que o número de viagens a partir de uma zona de origem i para uma zona de destino j é diretamente proporcional ao número de oportunidades existentes na zona j e inversamente proporcionais ao número de oportunidades intervenientes entre i e j (BEZ, 2000).

A primeira formulação deste Modelo enunciada por Stouffer (ALMEIDA, 1999), era a seguinte:

$$\frac{\Delta y}{\Delta s} = \frac{\alpha}{x} \frac{\Delta x}{\Delta s} \quad (2.9)$$

onde:

Δy = número de pessoas movendo-se da origem até uma faixa circular S ; de largura Δs delimitada por dois círculos concêntricos na origem de raios $S - \Delta s$ e $S + \Delta s$;

x = número de Oportunidades Intervenientes entre a origem dada e S ;

Δx = número de Oportunidades dentro de S ;

α = constante de proporcionalidade.

Na década de 50, Schneider criou um Modelo com base no de Stouffer, mas utilizando como base a Teoria de Probabilidades. Neste modelo ele enunciava que a viagem será a mais curta possível, aumentando de distância somente se não achasse um destino satisfatório numa distância menor (ALMEIDA, 1999).

A equação desenvolvida por Schneider utilizando a Teoria de Probabilidades foi a seguinte:

$$T_{ij} = O_i B_i \exp(-\lambda W_{ij}) [1 - \exp(-\lambda V_j)] \quad (2.10)$$

onde:

λ = probabilidade desse destino ser aceito;

W_{ij} = Número de Oportunidades entre i e j;

V_j = número de Oportunidades em j;

B_i = fator de balanceamento, sendo que:

$$\sum_{j=1}^n B_i \exp(-\lambda W_{ij}) [1 - \exp(-\lambda V_j)] = 1 \quad (2.11)$$

Stouffer relatou que várias aplicações de sua teoria em estudos sociológicos sentiam a necessidade de modificar a forma de se determinar o número de oportunidades Intervinentes, e em 1960, apresentou a seguinte maneira de determinar o número de oportunidades que se interpõem entre duas regiões (GONÇALVES, 1992):

- Unem-se por uma linha reta as regiões i e j;
- Traça-se uma circunferência com diâmetro do tamanho d_{ij} ;
- Contam-se as oportunidades de dentro desta circunferência.

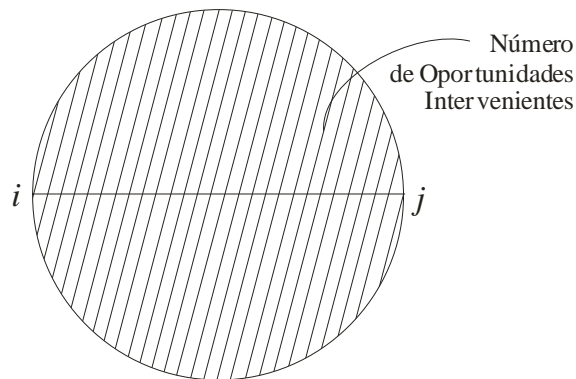


Figura 2 - Oportunidades Intervenientes
 Fonte: Gonçalves, 1992

De acordo com Almeida (1999) foram realizadas diversas aplicações e variações desse modelo na década de 60, e seguem até os dias atuais. Novaes (1981, apud ALMEIDA 1999) apresenta uma aplicação de modelo de oportunidades intervenientes em um estudo realizado na cidade de São Paulo, para determinação da localização e dimensionamento de um shopping center. Como se pode observar o modelo proposto é de caráter significativo para o estudo de planejamentos em transportes.

2.2.4 Modelos Híbridos

A literatura da década de 80 apresenta contribuições que unem as características de dois modelos (WILLS, 1986 apud ALMEIDA, 1999). Uma junção deste tipo é denominada de modelo híbrido, onde os efeitos de determinados modelos tornam-se complementares uns aos outros, e não substitutivos. Na literatura pode-se encontrar, por exemplo, um modelo deste tipo elaborado por Gonçalves (1992). Tal modelo é dado pela união dos modelos gravitacional e de oportunidades intervenientes, no qual é bem aceito pela comunidade científica.

Para este estudo será adotado um modelo com essa característica, fazendo a junção entre os modelos gravitacional e econométrico. Tal modelo foi proposto por LabTrans/UFSC (2005) .

2.3 Modelos Econométricos

2.3.1 Introdução

Antes, porém, de apresentar a definição de modelos econométricos, é importante saber a importância da econometria dentro da ciência econômica para depois introduzirmos suas premissas no planejamento de transportes.

O termo econometria surgiu pela primeira vez em 1926, atribuído a Ragnar Frisch., embora alguns estudiosos como Moore e Yule, antes desta data, já utilizassem variáveis sócio-econômicas para o estudo da demanda de um certo tipo de metal fundido.

No entanto, o principal período para a Econometria foi na década de 30 quando fundou-se a Econometric Society em Cleveland (EUA), em dezembro de 1930. A partir daí surgiram muitos estudos que contribuíram para o crescimento da Econometria, como por exemplo, aperfeiçoamento dos métodos de estimação, validação de modelos estimados, criação de novos conceitos, entre outros. (CARNEIRO, 1997).

De acordo com Tintner (1965), a econometria consiste em uma aplicação de métodos matemáticos e estatísticos em problemas econômicos. Com a utilização de procedimentos estatísticos e matemáticos na análise de dados, deseja-se verificar a validade de modelos em economia matemática, assim como, obter valores numéricos para parâmetros econômicos.

2.3.2 Concepção

Um modelo pode ser considerado uma abstração da realidade, e mostra o que é relevante para uma questão específica, sendo uma representação formal de idéias e conhecimento a cerca de um fenômeno. Os modelos Econométricos contêm especificações (matemáticas, variáveis definidas, equações), para uma aplicação empírica, além de incorporar um termo residual com a finalidade de levar em conta variáveis, ou outros elementos que não puderam ser considerados explicitamente (CARNEIRO, 1997).

É importante destacar que modelos econométricos constituem uma formulação incompleta da realidade, devido ao fato da necessidade de recorrer ao *ceteris paribus*¹ para preencher a lacuna entre a teoria e os fatos, isso ocorre pelo fato da impossibilidade de um modelo abranger todos os fatores que determinam ou condicionam um fenômeno (BARBANCHO, 1970 apud CARNEIRO, 1997).

Sobre os modelos econométricos, três aspectos se destacam (CARNEIRO, 1997):

1 - Estrutura: a estrutura de um modelo econométrico envolve quatro elementos básicos, que são: variáveis, relações ou equações, parâmetros ou coeficientes e termo aleatório ou residual. Variáveis são fatores singulares dos fenômenos econômicos, e podem ser classificadas em dependentes, que recebem influência de outras variáveis, ou variáveis independentes, que são aquelas que afetam as variáveis dependentes, cujo comportamento deseja-se explicar. As relações ou equações descrevem o mecanismo que aciona os elementos singulares do fenômeno. Parâmetros são magnitudes que permanecem constantes no âmbito de um fenômeno concreto. Já o termo aleatório ou residual é a expressão de pequenas causas, que produzem um desvio em relação ao que a variável dependente deveria ser, se a relação fosse determinística.

2 - Classificação quanto às características dos fenômenos: os modelos econométricos são classificados de acordo com as características dos fenômenos que se deseja modelar. Tais modelos podem ser classificados quanto à forma funcional, número de equações, associação de variáveis com o tempo e a finalidade.

3 - Qualidades desejadas: um modelo econométrico é avaliado de acordo com sua plausibilidade teórica, onde ele deve ser compatível com postulados da teoria econômica, também deve ser avaliado quanto a sua capacidade explanatória, sendo ele capaz de explicar os dados observados cuja relação ele determina. Também deve-se levar em consideração a exatidão das estimativas dos

¹ Mantidas inalteradas todas as outras coisas.

parâmetros, onde os mesmos deverão ser exatos no sentido de aproximar-se tanto quanto possível dos verdadeiros parâmetros estruturais. O modelo deve ser capaz de gerar previsões satisfatórias de valores futuros da variável dependente, e sempre representar relações econômicas com o máximo de simplicidade em termos de número de equações e da forma matemática, levando em consideração *ceteris paribus*.

De acordo com Carneiro (1997), a metodologia para a aplicação de um modelo econométrico deve ser fundada sobre três premissas básicas, são elas: quanto à formulação de hipóteses sobre o comportamento da realidade (em seguida essas hipóteses devem ser reunidas num modelo matemático para operacionalizá-lo), a segunda premissa é de que se colem os dados estatísticos e estimam-se os parâmetros com a utilização de um método apropriado, já o terceiro estágio compreende a avaliação mediante a utilização de critérios derivados das teorias, além de outros de natureza estatística ou econométrica.

Podemos dizer que a econometria utiliza-se de dois princípios básicos, a teoria e os fatos. Onde, a teoria incorpora-se no estudo econométrico através da utilização de modelos, já os fatos, são sumarizados pela utilização de dados expressos através de variáveis.

Dada a necessidade da utilização de um método para a estimação dos parâmetros do modelo, no próximo capítulo será apresentado o método utilizado no estudo, no qual, o mesmo se mostra mais aplicável ao caso.

3 REGRESSÃO LINEAR

3.1 Introdução

De acordo com Barbetta, Reis e Bornia (2004), a regressão surgiu com os trabalhos de Galton no final do século XIX, que procuravam explicar certas características de um indivíduo a partir das características de seus pais. Com o passar dos anos, o modelo matemático-estatístico foi aperfeiçoado e hoje é utilizado nas mais variadas áreas, como engenharia, economia, informática e biociências.

De acordo com Kazmier (1982) o principal objetivo de uma análise através de regressão é predizer o valor de uma variável dependente, dado que seja conhecido o valor de uma variável independente. Tem-se que a equação de regressão é a fórmula algébrica pelo qual se determina o valor previsto da variável dependente.

Pode-se dizer que existem dois tipos de regressão linear, a simples e a múltipla, cuja regressão simples indica que a predição da variável dependente é feita com uma única variável independente, enquanto a múltipla é feita com base em duas ou mais variáveis independentes. A concepção desses dois tipos de regressão é apresentada no item que segue. As premissas de hipóteses para validação do referido processo serão apresentadas do decorrer do trabalho.

3.2 Concepção

3.2.1 Regressão Linear Simples

Vários modelos utilizam a Regressão para determinar os coeficientes que não são conhecidos no modelo. A regressão é um processo que estabelece a relação entre várias variáveis, com o objetivo de prever os coeficientes através de dados conhecidos. Ela é dita linear se a relação funcional da regressão de Y (variável dependente) em função de X (variável independente) é expressa por uma equação cuja representação geométrica é uma reta.

Se tivermos uma função do tipo $f(x) = \alpha x + \beta$, dizemos que $f(x) = Y$ é a variável dependente enquanto x define a variável independente, pois o valor de Y dependerá do valor atribuído a x . Neste caso, α e β são os valores a serem determinados através da regressão linear, que neste caso é dita simples pois a relação causal envolve apenas duas variáveis.

No entanto, os dados de um modelo geralmente não estão exatamente sobre uma reta. Na figura abaixo, observando a reta de regressão é visível que há pontos acima e abaixo dela. Por essa razão é necessário adicionar uma variável chamada erro, tal variável é proveniente de várias razões, como imprecisão de medidas, por exemplo. A equação de um modelo linear simples pode ser escrita como $f(x) = \alpha x + \beta + \varepsilon$. Vale a pena salientar que quanto menor o erro, melhor é o ajuste dos dados à reta.

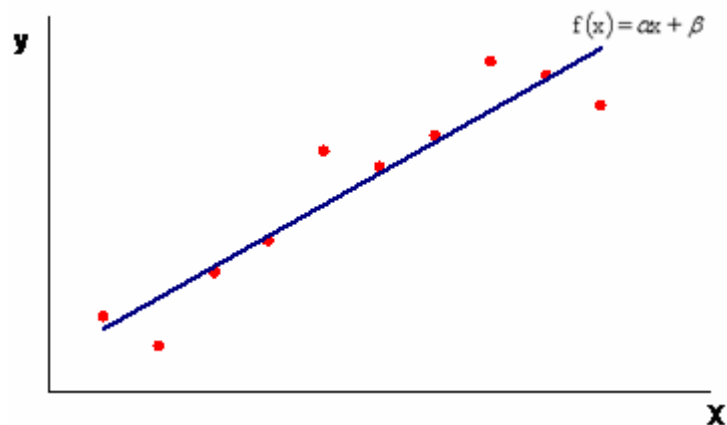


Figura 3 - Reta de Regressão

Uma forma de se determinar a reta de regressão é através do Método dos Mínimos Quadrados, que apresentamos a seguir.

3.2.1.1 Método dos Mínimos Quadrados para Regressão Linear Simples

De acordo com Barbetta, Reis e Bornia (2004), o método dos mínimos quadrados consiste em fazer com que a soma dos erros quadráticos seja a menor possível. Somar os erros, pura e simplesmente, não acrescenta muita informação, pois haverá erros positivos e negativos, que irão se cancelar numa soma. O método proposto de

mínimos quadrados resolve esse problema elevando os erros ao quadrado, eliminando assim os números negativos.

Para um modelo linear simples, dada uma função $f(x) = \alpha x + \beta$, vamos determinar os valores de α e β através deste método que apresenta as seguintes equações:

$$\beta = \frac{n \cdot \sum(x \cdot y) - (\sum x) \cdot (\sum y)}{n \cdot \sum x^2 - (\sum x)^2} \quad (3.1)$$

$$\alpha = \frac{\sum y - \beta \cdot \sum x}{n} \quad (3.2)$$

onde n é número de pares (x, y) observados.

Como exemplo, mostrar-se-á o cálculo da reta de regressão para as dez observações da tabela seguinte:

Tabela 1 - Exemplo do Método dos Mínimos Quadrados para Regressão Linear Simples

x	y	x*y	x²
13	12	156	169
18	21	378	324
22	32	704	484
24	34	816	576
25	35	875	625
27	35	945	729
33	40	1320	1089
42	42	1764	1764
46	46	2116	2116
51	57	2907	2601
$\Sigma x = 301$	$\Sigma y = 354$	$\Sigma x \cdot y = 11981$	$\Sigma x^2 = 10477$

Temos que:

$$\beta = \frac{(10 \cdot 11981) - (301 \cdot 354)}{(10 \cdot 10477) - (354)^2} = \frac{13256}{14169} = 0,935563554$$

e,

$$\alpha = \frac{354 - (0,935563554 \cdot 301)}{10} = 7,239537017$$

A equação da Reta de regressão para este conjunto de pontos será então:

$$f(x) = 7,24x + 0,94$$

3.2.2 Regressão Linear Múltipla

Um modelo de regressão linear múltipla pode ser expresso por:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + \varepsilon_i \quad (3.3)$$

onde:

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

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

ε_i é o 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;

$\beta_1, \beta_2, \beta_3, \dots, \beta_k$ são os parâmetros a serem estimados.

Para a estimação destes parâmetros, também poderá ser utilizado o método dos Mínimos Quadrados, mas neste caso, utilizando matrizes.

3.2.2.1 Método dos Mínimos Quadrados para Regressão Linear Múltipla

A idéia de estimação dos parâmetros $\beta_1, \beta_2, \beta_3, \dots, \beta_k$ 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,

$$Z = \sum_i (Y_i - (\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki}))^2, \quad k = 1, \dots, n \quad (3.4)$$

onde n representa o número de variáveis explicativas.

O estimador de mínimos quadrados para o vetor $\beta = [\beta_1, \beta_2, \beta_3, \dots, \beta_k]^t$ é dado por

$$\beta = (X'X)^{-1}(X'Y) \quad (3.5)$$

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 $(X'X)$.

No exemplo que segue, retirado de Sartoris (2004), serão estimados os valores dos parâmetros para determinar a reta de regressão, com uma amostra de 12 observações:

Tabela 2 - Exemplo do Método dos Mínimos Quadrados para Regressão Linear Múltipla

Y	X ₂	X ₃
800	2	0,8
1.160	4	0,7
1.580	6	0,5
2.010	8	0,4
1.890	7	0,2
2.600	12	0,2
2.070	11	0,8
1.890	10	0,7
1.830	9	0,6
1.740	8	0,1
1.380	6	0,5
1.060	4	0,4

O modelo é da forma:

$$Y = \beta + \alpha_1 X_2 + \alpha_2 X_3$$

Vamos encontrar então a matriz que contenha os valores dos coeficientes $\beta, \alpha_1, \alpha_2$.

Então, a matriz X será dada como abaixo:

$$X = \begin{bmatrix} 1 & 2 & 0,8 \\ 1 & 4 & 0,7 \\ 1 & 6 & 0,5 \\ 1 & 8 & 0,4 \\ 1 & 7 & 0,2 \\ 1 & 12 & 0,2 \\ 1 & 11 & 0,8 \\ 1 & 10 & 0,7 \\ 1 & 9 & 0,6 \\ 1 & 8 & 0,1 \\ 1 & 6 & 0,5 \\ 1 & 4 & 0,4 \end{bmatrix}$$

E assim,

$$X^t X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 4 & 6 & 8 & 7 & 12 & 11 & 10 & 9 & 8 & 6 & 4 \\ 0,8 & 0,7 & 0,5 & 0,4 & 0,2 & 0,2 & 0,8 & 0,7 & 0,6 & 0,1 & 0,5 & 0,4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0,8 \\ 1 & 4 & 0,7 \\ 1 & 6 & 0,5 \\ 1 & 8 & 0,4 \\ 1 & 7 & 0,2 \\ 1 & 12 & 0,2 \\ 1 & 11 & 0,8 \\ 1 & 10 & 0,7 \\ 1 & 9 & 0,6 \\ 1 & 8 & 0,1 \\ 1 & 6 & 0,5 \\ 1 & 4 & 0,4 \end{bmatrix}$$

$$X^t X = \begin{bmatrix} 12 & 87 & 5,9 \\ 87 & 731 & 41 \\ 5,9 & 41 & 3,53 \end{bmatrix}$$

$$\text{E sua inversa } (X^t X)^{-1} = \begin{bmatrix} 1,25 & -0,09 & -1,04 \\ -0,09 & 0,01 & 0,03 \\ -1,04 & 0,03 & 1,67 \end{bmatrix}$$

$$\text{A matriz } X^tY = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 4 & 6 & 8 & 7 & 12 & 11 & 10 & 9 & 8 & 6 & 4 \\ 0,8 & 0,7 & 0,5 & 0,4 & 0,2 & 0,2 & 0,8 & 0,7 & 0,6 & 0,1 & 0,5 & 0,4 \end{bmatrix} \begin{bmatrix} 800 \\ 1160 \\ 1580 \\ 2010 \\ 1890 \\ 2600 \\ 2070 \\ 1890 \\ 1830 \\ 1740 \\ 1380 \\ 1060 \end{bmatrix}$$

$$X^tY = \begin{bmatrix} 20010 \\ 160810 \\ 9309 \end{bmatrix}$$

Como visto anteriormente a matriz α será dada por:

$$\alpha = (X^tX)^{-1} (X^tY) = \begin{bmatrix} 1,25 & -0,09 & -1,04 \\ -0,09 & 0,01 & 0,03 \\ -1,04 & 0,03 & 1,67 \end{bmatrix} \begin{bmatrix} 20010 \\ 160810 \\ 9309 \end{bmatrix} = \begin{bmatrix} 789,33 \\ 149,56 \\ -419,26 \end{bmatrix}$$

Então, os valores estimados são:

$$\beta = 789,33$$

$$\alpha_1 = 149,56$$

$$\alpha_2 = -419,26$$

E o modelo tem a seguinte equação:

$$Y = 789,33 + 149,56 X_2 - 419,26 X_3$$

Segundo Barbetta, Reis e Bornia (2004), o método dos mínimos quadrados é o mais utilizado para estimar os parâmetros para o tipo de modelo utilizado no estudo.

3.2.3 Modelos Linearizáveis

Existem modelos que não são lineares. No entanto, a maioria deles podem ser facilmente linearizáveis. Dentre estes modelos, destacam-se os multiplicativos e os exponenciais.

1. Multiplicativos: A função $y = \alpha x^\beta$ pode se tornar linear. Para isso, aplica-se \ln em ambos os lados da igualdade:

$$\begin{aligned}\ln y &= \ln(\alpha x^\beta) \\ \ln y &= \ln \alpha + \ln x^\beta \\ \ln y &= \ln \alpha + \beta \ln x\end{aligned}$$

Assumindo $y_1 = \ln y$; $\alpha_1 = \ln \alpha$ e $x_1 = \ln x$ tem-se:

$$y_1 = \alpha_1 + \beta x_1$$

que torna-se um modelo linear.

2. Modelo Exponencial: Se o modelo tiver a seguinte forma $y = \alpha e^{\beta x}$, e novamente aplicar o \log em ambos os lados, também obtêm-se um modelo linear.

$$\begin{aligned}\ln y &= \ln(\alpha e^{\beta x}) \\ \ln y &= \ln \alpha + \ln e^{\beta x} \\ \ln y &= \ln \alpha + \beta x\end{aligned}$$

Assumindo $y_1 = \ln y$ e $\alpha_1 = \ln \alpha$, tem-se:

$$y_1 = \alpha_1 + \beta x_1$$

A linearização é fundamental para o processo da regressão, pois sem ela não seria possível utilizar a regressão linear para estimar os coeficientes.

3.3 Hipóteses para Regressão

Para que a regressão seja dita como robusta, deve-se obter um modelo que apresente resultados confiáveis, algumas hipóteses deverão ser verificadas, são elas:

- I. Os erros são normalmente distribuídos: 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.
- II. A média dos erros tem que ser zero: os erros são decorrentes de inúmeras possibilidades, como exemplo, no caso da compra e venda de imóveis, onde os preços são, em geral, baseados em sua área, pode haver casos nos quais os proprietários têm uma maior urgência em vender o imóvel, e acabam baixando o preço, ou casos em que o comprador paga um preço acima do praticado no mercado por uma necessidade pessoal. Essas diferenças são as que acarretam os erros. 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)
- III. A variância dos erros é constante: esta hipótese é conhecida como homocedasticidade, que indica que independentemente da característica dos dados, a variação dos erros deve-se manter constante. Utilizando o exemplo anterior, se percebemos que o erro torna-se maior (ou menor) conforme o tamanho do imóvel, não há homocedasticidade, e neste caso diz-se que há heterocedasticidade. Podemos verificar se há heterocedasticidade observando o gráfico ou aplicando alguns testes estatísticos. Inicialmente, observamos abaixo dois exemplos de gráficos que ilustram os resíduos de uma regressão:

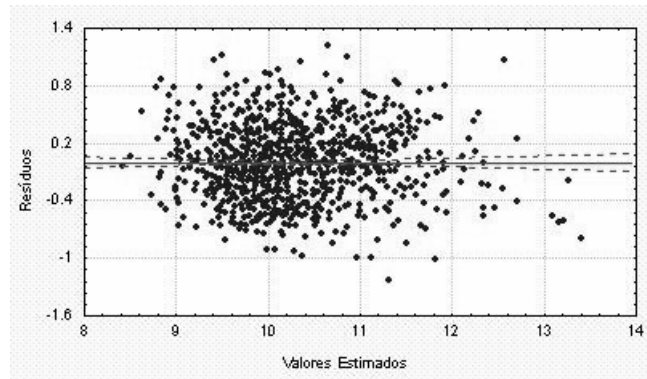


Figura 4 - Visualização residual 1

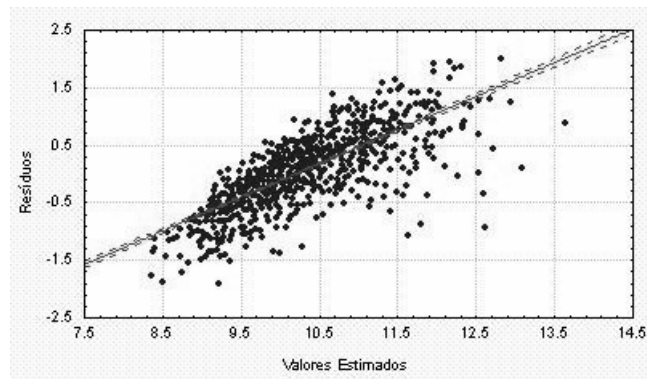


Figura 5 - Visualização residual 2

No primeiro gráfico, observa-se que os resíduos mantêm-se dentro de uma faixa de aceitação, ou seja, eles não aumentam ou diminuem conforme os valores estimados mudam, o que não acontece no segundo gráfico, no qual os erros aumentam na mesma proporção que os valores estimados, ou seja, o segundo caso apresenta heterocedasticidade.

Existem muitos testes na literatura para verificar se existe a heterocedasticidade. Iremos apresentar aqui somente o teste de “Goldfeld e Quandt”, que consiste em separar os dados em três partes e comparar as variâncias. GUJARATI (2000) apresenta os seguintes passos para este teste:

Passo 1: Ordene as 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 de acordo com qualquer uma delas.

Passo 2: Omita as observações centrais c , e divida as $(n-c)$ observações restantes em dois grupos, cada um com $\frac{(n-c)}{2}$ observações. O valor de c é sugerido 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 feita para acentuar a diferença entre os grupos de máxima e mínima variância.

Passo 3: Calcular a estatística apresentada a seguir, observando que o grupo I corresponde aos menores valores de X , enquanto o grupo II os maiores.

$$F = \frac{S_{II}^2}{S_I^2} \quad (3.6)$$

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

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

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

O valor obtido desta razão deve ser comparado com o valor apresentado na tabela da distribuição F que apresenta os valores limites para este teste (Ver Anexo 1). O valor calculado deve ser menor que o valor tabelado para concluir que a variância é constante. Caso contrário deve-se corrigir a heterocedasticidade para que o valor apresente dados confiáveis. Para corrigir este problema, devemos saber qual é a variável que está causando esta distorção dos dados. A partir daí, devemos dividir todo o modelo por esta variável e refazer a regressão. Esse procedimento deverá corrigir o problema e tornar o modelo homocedástico se a variável for a correta.

IV. Cada variável independente x_i não pode ser combinação linear das demais.

É impossível determinarmos um modelo de regressão linear que apresente resultados confiáveis quando existe a multicolinearidade, que é a combinação linear das variáveis independentes. Isso porque a variação de uma delas é totalmente influenciada pelas outras.

Para verificar se no modelo escolhido existe a multicolinearidade entre as variáveis, deve-se observar atentamente os resultados das estatísticas apresentadas. É natural pensar que se aplicarmos mais que uma estatística para avaliar o modelo, elas devem trazer a mesma conclusão, ou seja, que o modelo é válido ou que não é válido. Se uma estatística apresentar um resultado muito bom, enquanto outra coloca em dúvida a utilização do modelo, é sinal de que existe multicolinearidade. Geralmente utiliza-se a estatística F, a estatística t, e o coeficiente de determinação ou R^2 . A primeira é descrita abaixo, e o coeficiente de determinação logo a seguir.

$$F = \frac{\text{var}(\text{valores explicados})}{k} \times \frac{n-k-1}{\text{var}(\text{resíduos})} \quad (3.7)$$

onde var = variância;

k = número de variáveis explicativas;

n = número de observações.

3.4 Poder de Explicação do Modelo

Para verificar a relevância do modelo devem-se observar estatísticas que mostrem o poder de explicação, dentre elas está o coeficiente de determinação R^2 , que é a mais utilizada e mede o quanto à variabilidade total dos dados é explicada pelo modelo de regressão. Quanto maior o R^2 , mais a variação total de Y é reduzida pela introdução das variáveis preditoras. O coeficiente R^2 é dado por

$$R^2 = \frac{SQ_{eg}}{SQT} = - \frac{SQE}{SQT} \quad (3.8)$$

onde:

SQE = Consiste na soma dos quadrados dos erros, ou seja, da distância entre os valores previstos e a média dos dados históricos;

SQT = Consiste na soma dos quadrados dos erros totais, caso a previsão fosse dada pela média dos dados históricos.

Quanto maior for o percentual encontrado, melhor é a explicação do modelo, mas como dito anteriormente, as hipóteses apresentadas acima devem ser satisfeitas para que o modelo seja realmente confiável.

4 APLICAÇÃO DO MODELO

No item 4.1 é apresentado um modelo baseado em um estudo já existente, com o intuito de estimar a demanda no transporte rodoviário interestadual de passageiros de ônibus da região sul do país.

No item 4.2 é realizada uma abordagem de uma nova proposta do modelo citado em 4.1, com o intuito de melhorar os resultados encontrados, inserindo novas variáveis e apresentando uma forma diferente de segmentação dos dados.

4.1 Modelo Previamente Idealizado

4.1.1 Introdução

Neste capítulo apresenta-se como foi concebida a aplicação de um modelo econométrico com adaptações inspiradas nos modelos clássicos de distribuição de viagens. O estudo está baseado em um convênio de número 018/2004, realizado pelo Laboratório de Transportes e Logística – LABTRANS, da Universidade Federal de Santa Catarina - UFSC juntamente com a Agência Nacional de Transportes Terrestres – ANTT, intitulado de “Modelo de Viabilidade, Monitoramento e Representação dos Indicadores de Desempenho das Linhas de Transporte Rodoviário de Passageiros”. O mesmo refere-se à criação de modelos que visam estimar a demanda de passageiros em todas as ligações origem-destino dentre as cidades brasileiras, de âmbito interestadual.

Para o presente trabalho foram utilizadas somente as ligações interestaduais do sul do país, ou seja, linhas que fazem ligação entre cidades dos três estados da região sul do país (Rio Grande do Sul, Santa Catarina e Paraná), como apresentado no diagrama da Figura 6.

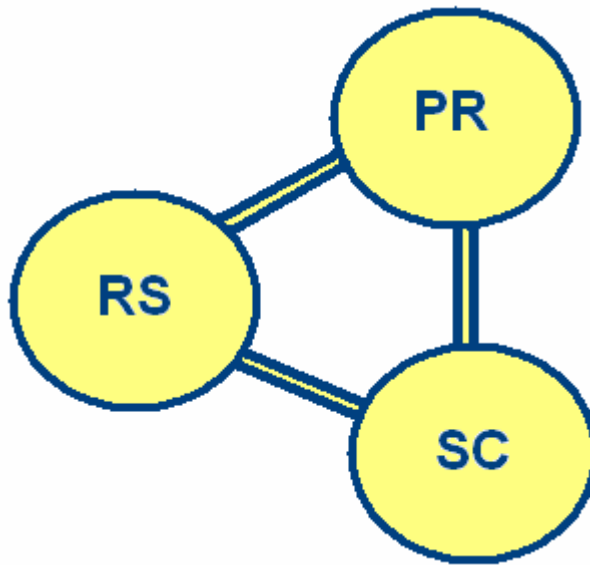


Figura 6 - Visualização das Ligações Interestaduais do Sul do País

A escolha de uma modelo que melhor represente a realidade para o referido estudo é fundamental para a obtenção de resultados coesos e com certo grau de confiabilidade. Dentre os modelos apresentados anteriormente, aquele que se encaixa melhor no contexto do estudo é utilizado e analisado a seguir.

4.1.2 Escolha do Modelo

O modelo em estudo é do tipo gravitacional, com adição de variáveis sócio-econômicas que possuem relevante influência nas quantidades demandadas de passageiros. O modelo em sua forma original, proposto por LABTRANS/UFSC (2005), é expresso por:

$$T_{i,j} = \alpha_0 \frac{P_i^{\alpha_1} P_j^{\alpha_2} mig_i^{\alpha_3} mig_j^{\alpha_4} m_i^{\alpha_5} m_j^{\alpha_6} (r_i \cdot r_j)^{\alpha_7} dummy_j^{\alpha_8}}{d_{ij}^{\alpha_9}} \quad (4.1)$$

onde:

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 = à 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 (valores adotados: $dummy_j = 2$, se j é um pólo turístico e $dummy_j = 1.2$, em caso contrário);

d_{ij} = distância rodoviária entre os municípios i e j ;

$\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9$ = coeficientes a serem determinados.

Em muitos modelos há a necessidade de se utilizar uma variável qualitativa, essa variável, denominada dummy pode ser entendida como um atributo, cuja forma não se encontra numericamente, ou seja, sexo, cor, nacionalidade, cor, ausência ou presença de um determinado fenômeno, dentre muitas outras. Para cada uma das condições atribui-se um valor numérico para utilizar no modelo.

Para o modelo ora proposto, a variável dummy será atribuída a cidades que são consideradas pólos turísticos, devido a existência de uma maior atratividade por parte dessas cidades. Os valores atribuídos à dummy, 2 e 1,2, vem da necessidade de se linearizar o modelo, e através disso, os valores comumente usados para esta variável, 0 e 1, não apresentariam resultados esperados, uma vez que $\ln 0$ não existe e $\ln 1$ é igual a zero.

É importante ressaltar que os modelos agregados, categoria na qual se encaixa o modelo proposto neste estudo, foram amplamente utilizados nas décadas de 50 a 70, num grande número de estudos desenvolvidos principalmente na Inglaterra e nos Estados Unidos, tendo sido bastante criticados no período que se seguiu (final da década de 70 até início da década de 90). Na década de 90, muitos estudos retomaram a modelagem agregada, que continua sendo amplamente utilizada ao

redor do mundo (HOROVITZ e FARMER, 1998; WIRASINGHE e KUMARAGE, 1998; Zhao et al, 2004 apud GONÇALVES, 2006).

“No contexto brasileiro, a modelagem da demanda por transporte rodoviário interestadual de passageiros sempre teve grande importância para nortear as ações dos órgãos públicos encarregados da normatização e planejamento desses serviços. Os primeiros estudos foram realizados na década de 70, destacando-se o trabalho pioneiro de Luiz Eugênio D. Gomes. Recentemente, diversos outros estudos foram desenvolvidos, buscando aprimorar as estimativas fornecidas pelos modelos anteriormente usados (Consórcio STE-ENEFER, 2001; Fundação Universitária José Bonifácio, 2003). Nos diversos estudos, a modelagem tem em comum a adoção das variáveis contempladas no modelo gravitacional clássico e da técnica de regressão linear múltipla para a calibração dos modelos” (LABTRANS/UFSC, 2005).

4.1.2.1 Variáveis Utilizadas

Os dados referentes às variáveis utilizadas foram retirados das seguintes fontes:

- Serviços de Transportes Rodoviário de Passageiros (SISDAP): As origens-destinos contendo suas distâncias e movimentação de passageiros, referente ao ano de 2003;
- Instituto Brasileiro de Geografia e Estatística (IBGE): Renda, População e Migração;
- Empresa Brasileira de Turismo (EMBRATUR): Cidades consideradas Pólo Turístico.

4.1.3 Calibração

4.1.3.1 Introdução

Para NOVAES (1981) a fase de calibração é extremamente importante, pois nela o analista vivencia os pontos fortes e as limitações do modelo, correspondendo com a realização de experimentos da pesquisa pura.

De acordo com GONÇALVES e SOUZA DE CURSI (1997), o processo de calibração consiste em duas etapas, são elas:

- “Deve-se definir um critério de calibração, isto é, deve-se definir que medida será utilizada para avaliar o ajuste das estimativas aos dados;
- deve ser estabelecidos procedimentos computacionais que permitam determinar os parâmetros do modelo de forma a obter o melhor ajuste”.

A determinação dos parâmetros passa pela maximização ou minimização de funções, com adoção de métodos numéricos adequados, que permitam resolver tais problemas de forma eficaz, com tempos de resolução compatíveis com as necessidades práticas (LABTRANS/UFSC). Para manter uma compatibilidade com os interesses foi utilizado o software MINITAB para a calibração dos modelos, por conter uma série de atributos que suprem as necessidades do trabalho.

4.1.3.2 Estratificação dos Dados

Os dados foram segmentados por distância, processo esse muito utilizado em estudos na área de transportes, devido às características dentro de cada uma das faixas. Os grupos então, são assim classificados:

- Grupo A: viagens curtas: até 400 km;
- Grupo B: viagens médias: entre 400 e 800 km;
- Grupo C: viagens longas: acima de 800 km.

Considerando que cada faixa tem características e aspectos comportamentais diferentes umas das outras, foi criado um modelo para cada uma das faixas de distância pré-definidas.

Embasado na divisão destes grupos, tem-se a tabela que segue. A mesma apresenta as quantidades de Origem-Destino de cada grupo e suas respectivas amostras, considerando que a amostra 1 foi utilizada para a regressão, enquanto a amostra 2 foi usada para a análise e validação dos modelos.

Tabela 3 - Número de OD's por Grupo segmentado por distância

Grupo	Amostra	Número de OD's
Grupo A	Amostra 1	574
	Amostra 2	574
	Total	1148
Grupo B	Amostra 1	403
	Amostra 2	403
	Total	806
Grupo C	Amostra 1	70
	Amostra 2	70
	Total	140

Cada uma das amostras representa 50% do número de OD's.

4.1.3.3 Resultados Obtidos

Como é perceptível, a escolha das variáveis para um modelo é um fator muito importante para a obtenção de bons resultados. Se o modelo contiver variáveis que não necessárias ou faltar alguma importante ele não se tornará confiável. Por este motivo, muitos autores recomendam que sejam selecionadas todas as variáveis possíveis, e que se faça a seleção das variáveis utilizando os métodos *Best Subsets Regression*, ou *Stepwise Regression*. Esses métodos também podem ser chamados de regressão passo a passo, no qual apresentam os melhores conjuntos levando em consideração correlação das variáveis e algumas estatísticas.

Para a seleção das melhores variáveis foi então utilizado estes métodos, o *Stepwise Regression* e *Best Subsets Regression*, e os resultados obtidos para os grupos A, B e C estão no anexo 2, 3 e 4, respectivamente, assim como as regressões para cada grupo. Nestes resultados, percebe-se que o conjunto com as melhores variáveis foi o mesmo para os três grupos, e essas variáveis são as seguintes:

Tabela 4 - Variáveis por Grupo segmentado por distância

Grupo	<i>Grupo A</i>	<i>Grupo B</i>	<i>Grupo C</i>
Variáveis	$P_i, P_j, m_i, m_j, dummy, d_{ij}$	$P_i, P_j, m_i, m_j, dummy, d_{ij}$	$P_i, P_j, m_i, m_j, dummy, d_{ij}$

Para verificar a credibilidade destes modelos, a regressão terá que atender as hipóteses apresentadas no capítulo 3. Tais hipóteses serão verificadas a seguir.

Hipótese I: a normalidade dos resíduos se mostrou insatisfatória para os grupos A e B, como pode-se observar nos gráficos seguintes.

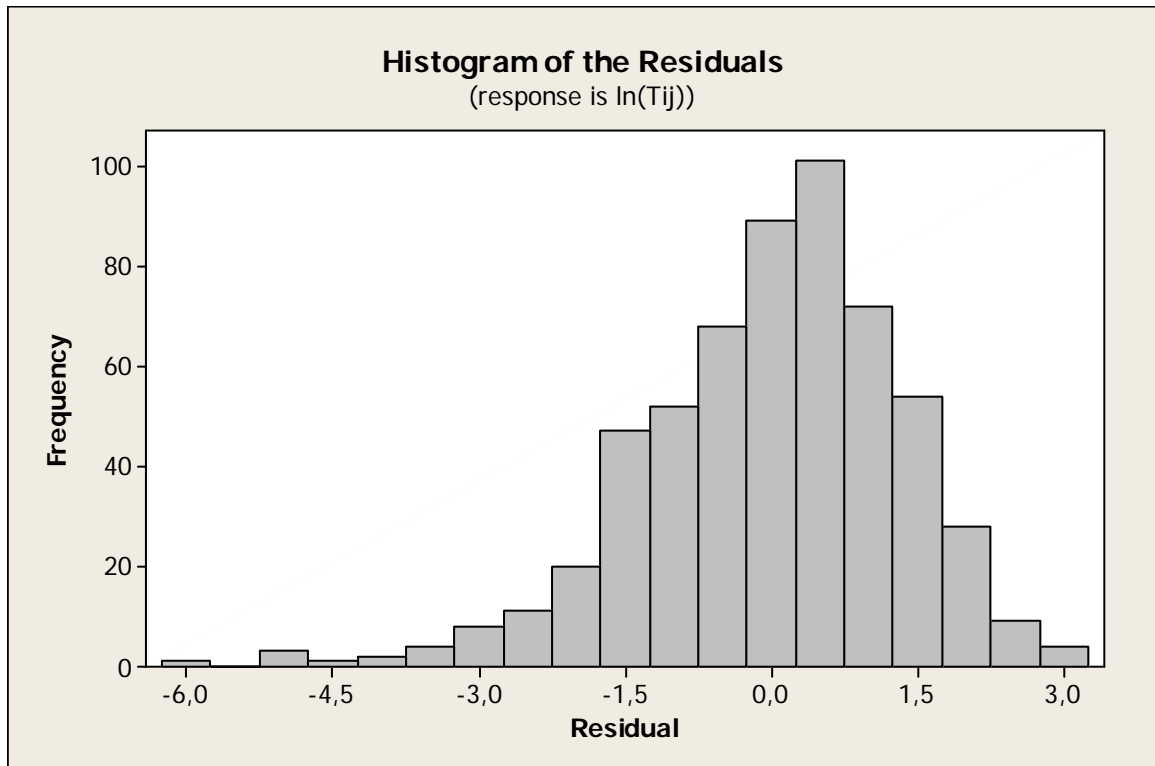


Figura 7 - Histograma dos Resíduos do Grupo A sem a Retirada dos Outliers

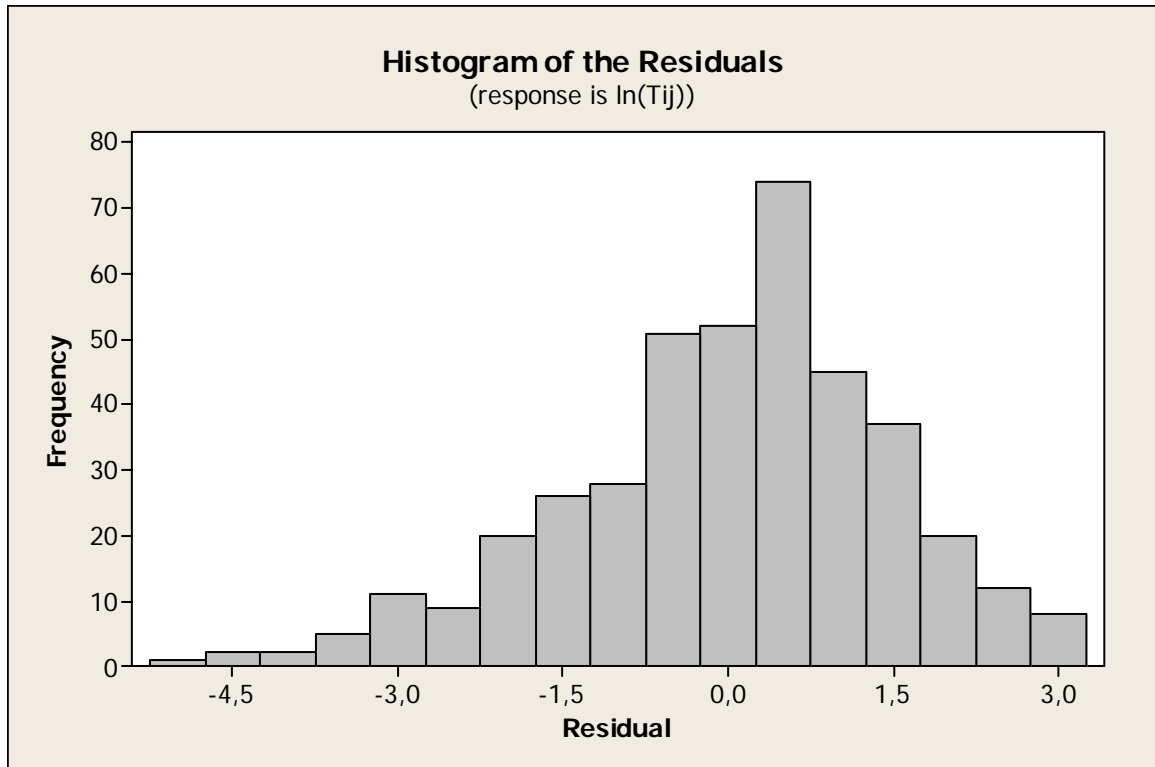


Figura 8 - Histograma dos Resíduos do Grupo B sem a Retirada dos Outliers

Para a solução deste problema, decidiu-se pela retirada dos *outliers*, e após a retirada destes pontos discrepantes foi feita uma nova regressão, que é apresentada no anexo 5 para o grupo A e no anexo 6 para o grupo B. O histograma que representa a normalidade dos resíduos para a nova regressão é mostrado a seguir, onde pode-se concluir que a hipótese I é satisfeita para os grupos A e B, uma vez que a normalidade segue uma tendência normal de distribuição.

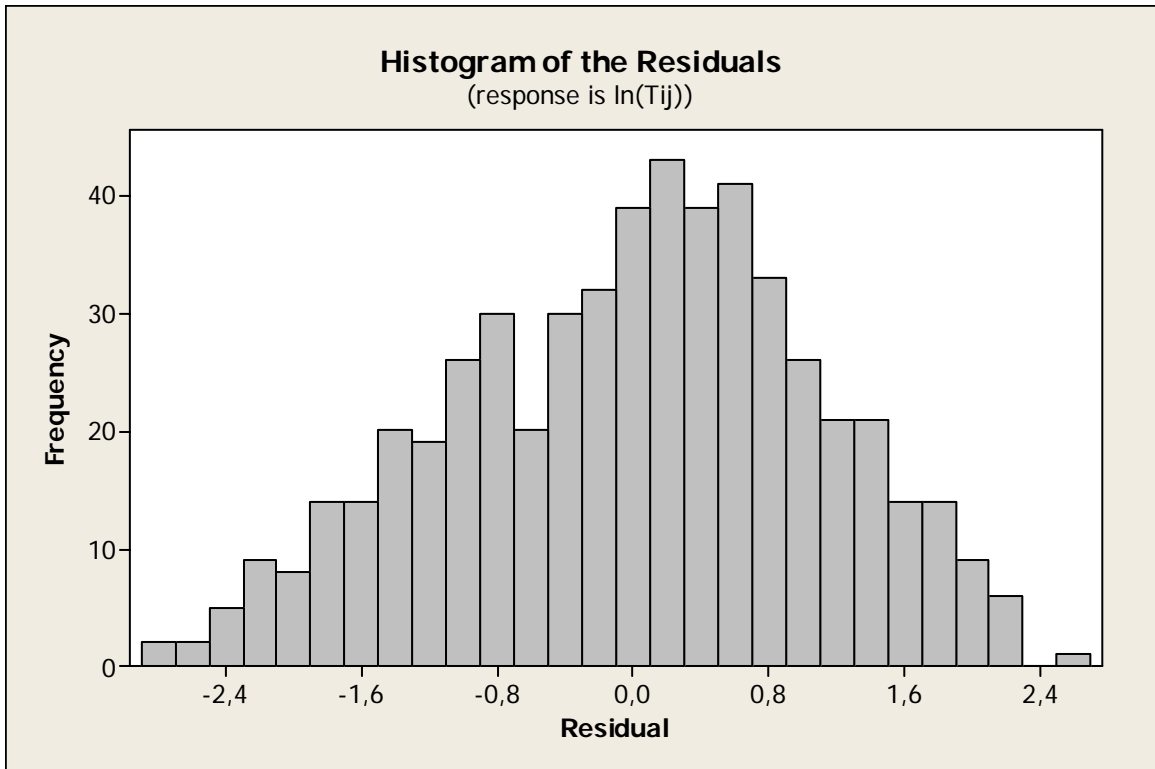


Figura 9 - Histograma dos Resíduos do Grupo A com a Retirada dos Outliers

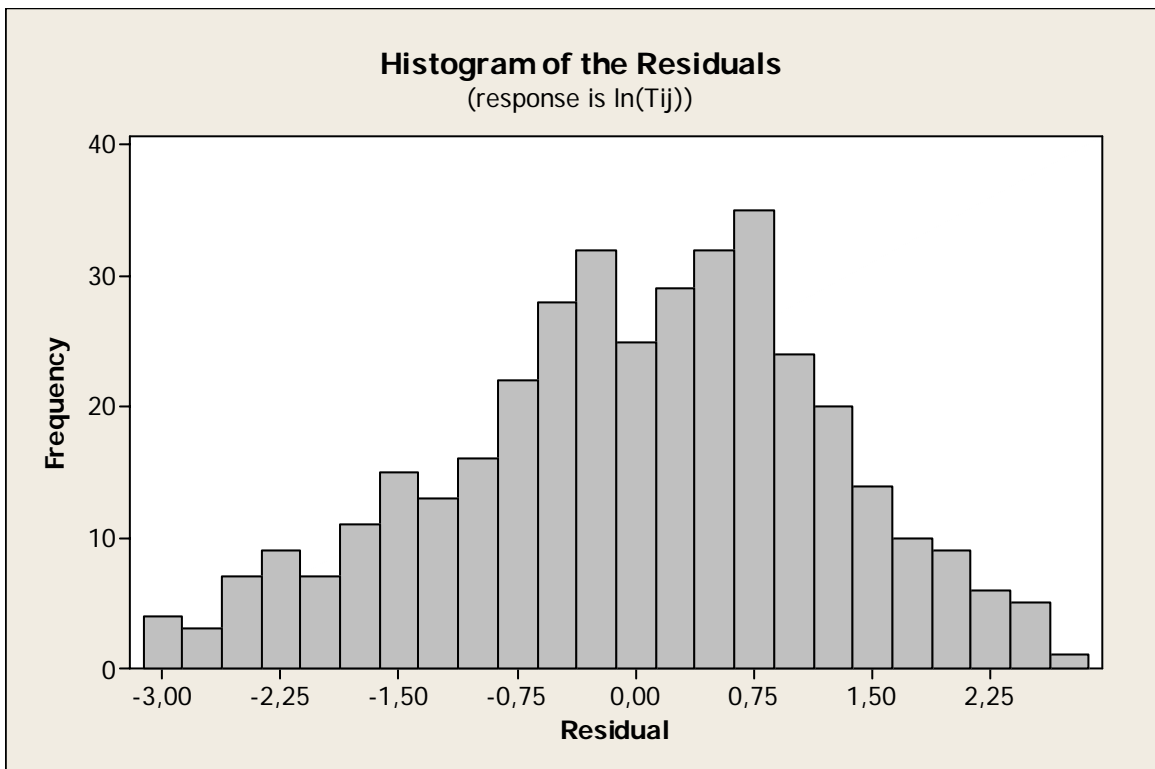


Figura 10 - Histograma dos Resíduos do Grupo B com a Retirada dos Outliers

Já para o grupo C, a normalidade dos resíduos através do gráfico não é um ótimo resultado, mas pode ser considerado satisfatório, até porque ao retirar os *outliers* o novo modelo encontrado não apresentou bons resultados. Então, decidiu-se utilizar o modelo apresentado anteriormente e testar as hipóteses. A seguir é apresentado o gráfico da normalidade dos resíduos para o grupo C.



Figura 11 - Histograma dos Resíduos do Grupo C sem a Retirada dos Outliers

Assim para os três grupos apresentados, pode-se perceber que os resíduos seguem uma distribuição normal, satisfazendo a hipótese I.

Hipótese II: Esta hipótese trata da média dos erros, no qual assume que essa média deve ser igual a zero. Para todos os grupos testados, esta hipótese foi satisfeita.

Hipótese III: Para avaliar se os modelos são homocedásticos, foi feito uma comparação de variâncias através do teste F. Dividiram-se os bancos em três partes, através de uma classificação pela variável distância. A tabela apresentada a seguir mostra o número de observações em cada uma das partes e também o valor F calculado.

Tabela 5 - 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
Grupo A	210	154	210	2,05
Grupo B	148	107	148	1,04
Grupo C	25	20	25	0,70

O valor do teste F tabelado para 7 graus de liberdade é 3,79, e como pode-se 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.

Hipótese IV: A não-multicolinearidade foi satisfeita ao adotar-se o método Stepwise Regression, que apresenta os melhores conjuntos de variáveis sem a alta correlação das variáveis.

Como se encontrou valores satisfatórios de R^2 (Grupo A = 58,2; Grupo B = 51,8 e Grupo C = 62,3), e todas as hipóteses foram satisfeitas, os modelos apresentados para os grupos em questão são definidos da seguinte forma:

- **Grupo A**

$$T_{ij} = \exp\left(\alpha_1 + \alpha_2 \ln\left(\frac{P_i}{1000}\right) + \alpha_3 \ln\left(\frac{P_j}{1000}\right) + \alpha_4 \ln(m_i) + \alpha_5 \ln(m_j) + \alpha_6 \ln(d_{ij}) + \alpha_7 \ln(dummy)\right) \quad (4.2)$$

onde,

$$\alpha_1 = 6,813570755236150$$

$$\alpha_2 = 0,834015431000899$$

$$\alpha_3 = 0,429636456781609$$

$$\alpha_4 = 0,412933716588056$$

$$\alpha_5 = 0,339163832694288$$

$$\alpha_6 = -0,905269969649239$$

$$\alpha_7 = 1,884705395801870$$

- **Grupo B**

$$T_{ij} = \exp\left(\alpha_1 + \alpha_2 \ln\left(\frac{P_i}{1000}\right) + \alpha_3 \ln\left(\frac{P_j}{1000}\right) + \alpha_4 \ln(m_i) + \alpha_5 \ln(m_j) + \alpha_6 \ln(d_{ij}) + \alpha_7 \ln(dummy)\right) \quad (4.3)$$

onde,

$$\begin{aligned} \alpha_1 &= 5,396279072568030 \\ \alpha_2 &= 0,756132314617040 \\ \alpha_3 &= 0,342281474811054 \\ \alpha_4 &= 0,334854458678164 \\ \alpha_5 &= 0,314280892913381 \\ \alpha_6 &= -0,600348019809113 \\ \alpha_7 &= 1,618869797516230 \end{aligned}$$

- **Grupo C**

$$T_{ij} = \exp\left(\alpha_1 + \alpha_2 \ln\left(\frac{P_i}{1000}\right) + \alpha_3 \ln\left(\frac{P_j}{1000}\right) + \alpha_4 \ln(m_i) + \alpha_5 \ln(m_j) + \alpha_6 \ln(d_{ij}) + \alpha_7 \ln(dummy)\right) \quad (4.4)$$

onde,

$$\begin{aligned} \alpha_1 &= 21,666171618055300 \\ \alpha_2 &= 0,818076503492421 \\ \alpha_3 &= 0,436840633832909 \\ \alpha_4 &= 0,450175532577119 \\ \alpha_5 &= 0,259368261841286 \\ \alpha_6 &= -3,056354106549690 \\ \alpha_7 &= 1,353690816223910 \end{aligned}$$

4.2 Criação e Concepção de um Novo Modelo

4.2.1 Escolha do Modelo

A escolha do modelo foi baseada no modelo utilizado anteriormente, porém com a inserção de duas novas variáveis coletadas do Instituto Brasileiro de Geografia e Estatística, são elas:

- Saúde: representada por S, onde esta variável mostra o número de estabelecimentos de saúde (como hospitais, postos médicos, maternidades) em cada município, tendo a S_i para origem e S_j para destino.
- Salários: esta variável traz a mediana dos salários das pessoas residentes por município, representada por SL_i para origem e SL_j para o destino.

O modelo com a inserção destas novas variáveis fica expresso da seguinte forma:

$$T_{i,j} = \alpha_0 \frac{P_i^{\alpha_1} P_j^{\alpha_2} mig_i^{\alpha_3} mig_j^{\alpha_4} m_i^{\alpha_5} m_j^{\alpha_6} (r_i.r_j)^{\alpha_7} dummy_j^{\alpha_8} S_i^{\alpha_{10}} S_j^{\alpha_{11}} SL_i^{\alpha_{12}} SL_j^{\alpha_{13}}}{d_{i,j}^{\alpha_9}} \quad (4.5)$$

4.2.2 Calibração

4.2.2.1 Estratificação dos Dados

Para a estratificação dos dados, decidiu-se utilizar uma estratificação através da variável Passageiros × Quilômetros (*PKm*). Essa variável é muito usada como uma medida anual de produtividade, que representa o número total de passageiros transportados, ponderado pelas distâncias percorridas na rede rodoviária coberta pelas viagens realizadas. Assim, seguindo uma coerência na segmentação, onde muitos estudos utilizam esta variável para tal finalidade. Os dados foram divididos através dos quartis da amostra ordenada pelo PKm dos pares origem-destino.

A tabela 6 apresenta as quantidades de Origem-Destino de cada grupo e suas respectivas amostras, considerando que a amostra 1 foi utilizada para a regressão, enquanto a amostra 2 foi usada para a análise e validação dos modelos.

Tabela 6 - Número de OD's por grupo segmentado por PKm

Grupo	Amostra	Número de OD's
Grupo 1	Amostra 1	262
	Amostra 2	262
Total		524
Grupo 2	Amostra 1	260
	Amostra 2	263
Total		523
Grupo 3	Amostra 1	263
	Amostra 2	262
Total		525
Grupo 4	Amostra 1	262
	Amostra 2	262
Total		524

4.2.2.2 Resultados Obtidos

Assim como o modelo proposto no tópico anterior, a seleção das melhores variáveis foi feita através dos métodos *Stepwise Regression* e *Best Subsets Regression*, e os resultados obtidos para os grupos 1, 2, 3 e 4 estão nos anexos 7, 8, 9 e 10, respectivamente, assim como as regressões para cada grupo. Nestes resultados, ao contrário do caso anterior, cada grupo possui um conjunto de variáveis diferentes uns dos outros. As variáveis utilizadas para cada grupo são as seguintes:

Tabela 7 - Variáveis por Grupo segmentado por PKm

Grupo	<i>Grupo 1</i>	<i>Grupo 2</i>	<i>Grupo 3</i>	<i>Grupo 4</i>
Variáveis	d_{ij}, P_i, P_j, m_i	d_{ij}, P_i, mig_i, s_j	$d_{ij}, P_i, P_j, m_i, m_j$	$d_{ij}, P_i, dummy, s_j, m_i$

A análise dos resultados das regressões através das hipóteses apresentadas no capítulo 3 é mostrada a seguir.

Hipótese I: Os resíduos para os grupos 1, 2 e 3 não seguiram uma distribuição normal, e por esta razão, decidiu-se pela retirada dos *outliers*, para tentar amenizar essa imperfeição. A seguir são apresentados os histogramas dos resíduos com e sem a retirada destes pontos para todos os grupos. A nova regressão, sem os dados discrepantes é apresentada nos Anexos 11, 12 e 13, para os Grupos 1, 2 e 3, respectivamente.

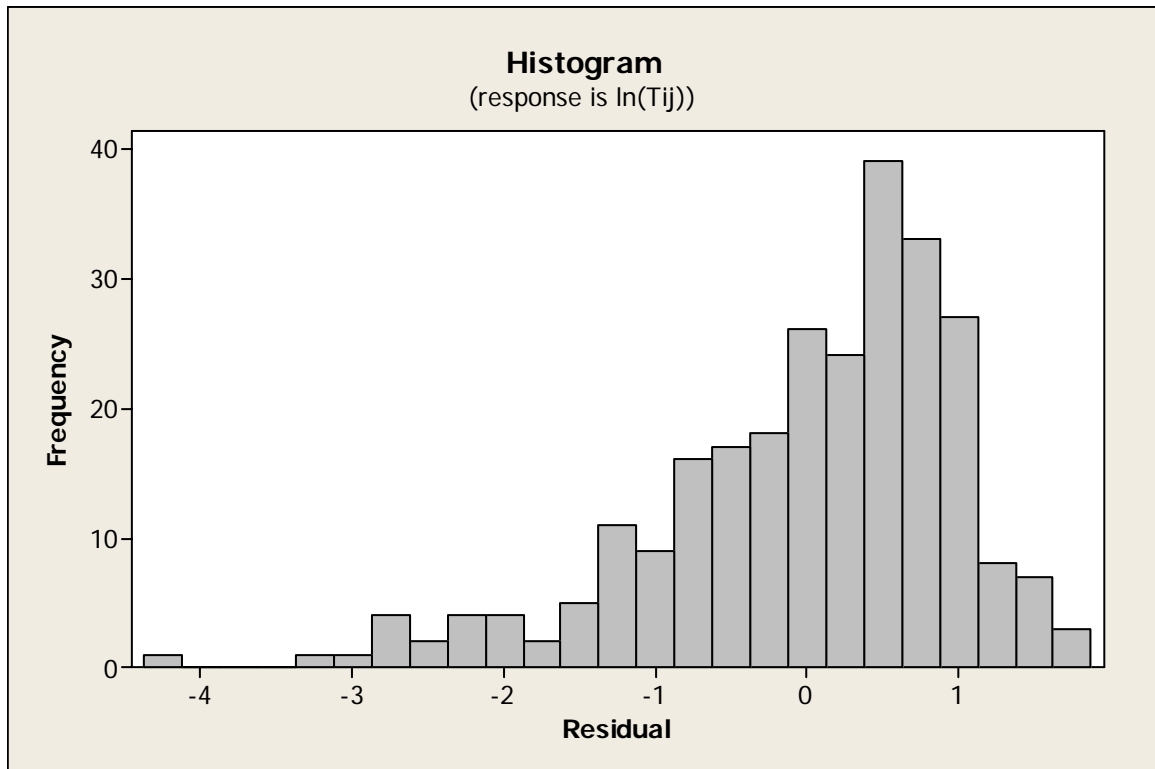


Figura 12 - Histograma dos Resíduos do Grupo 1 sem a Retirada dos Outliers

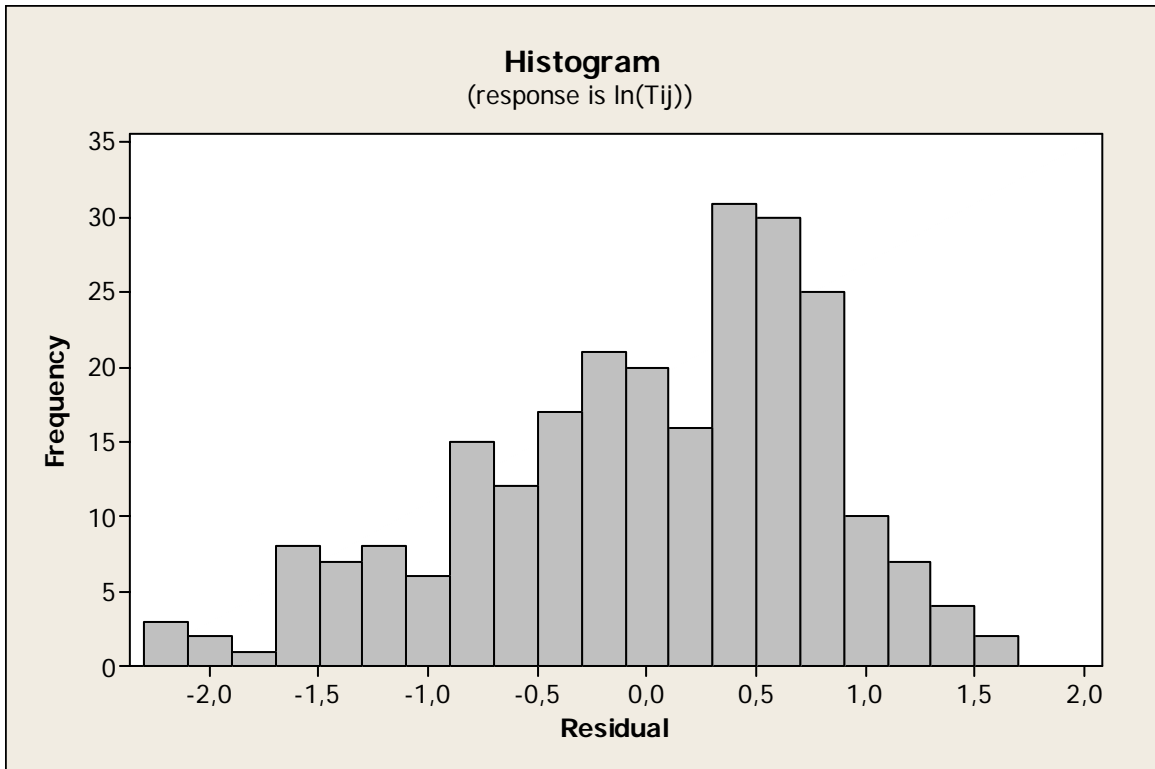


Figura 13 - Histograma dos Resíduos do Grupo 1 com a Retirada dos Outliers

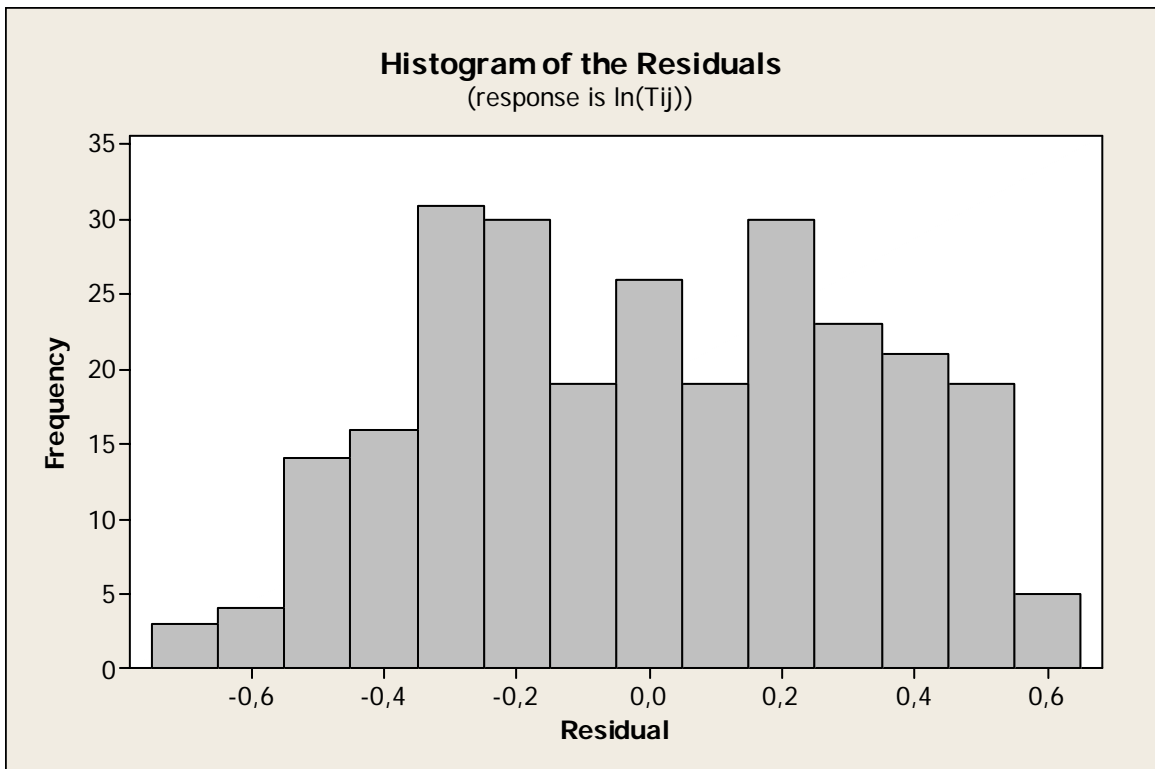


Figura 14 - Histograma dos Resíduos do Grupo 2 sem a Retirada dos Outliers

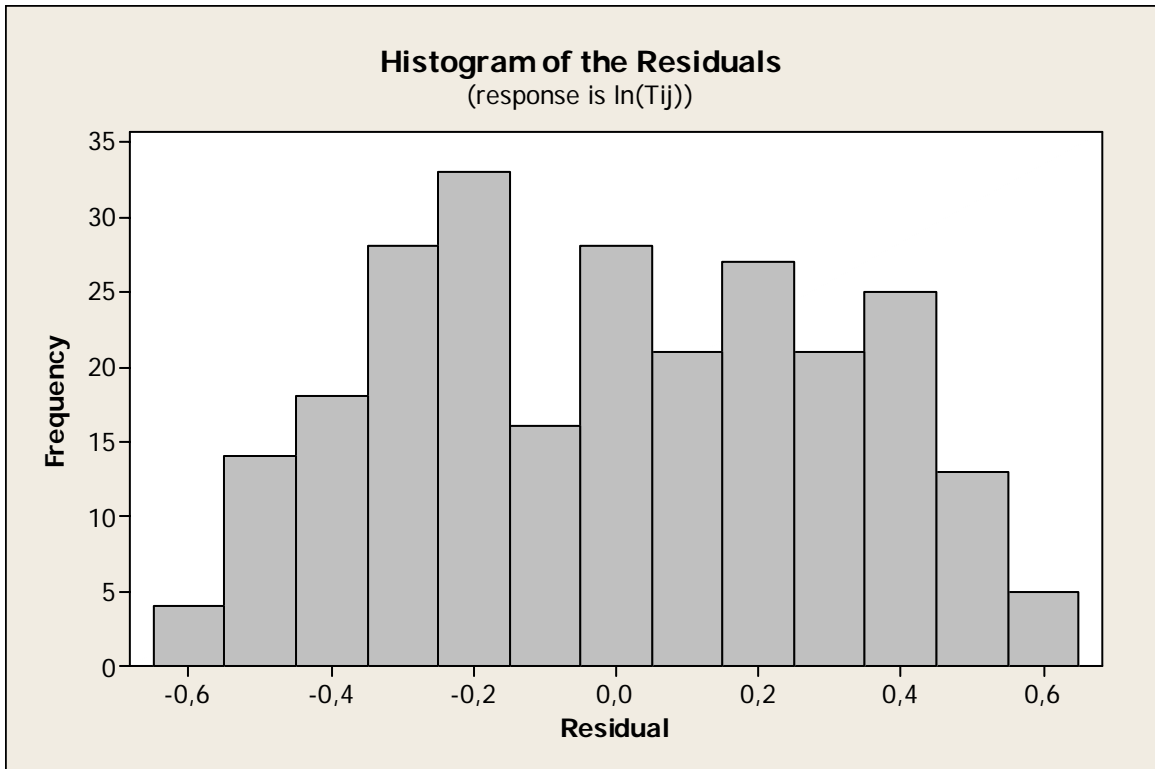


Figura 15 - Histograma dos Resíduos do Grupo 2 com a Retirada dos Outliers

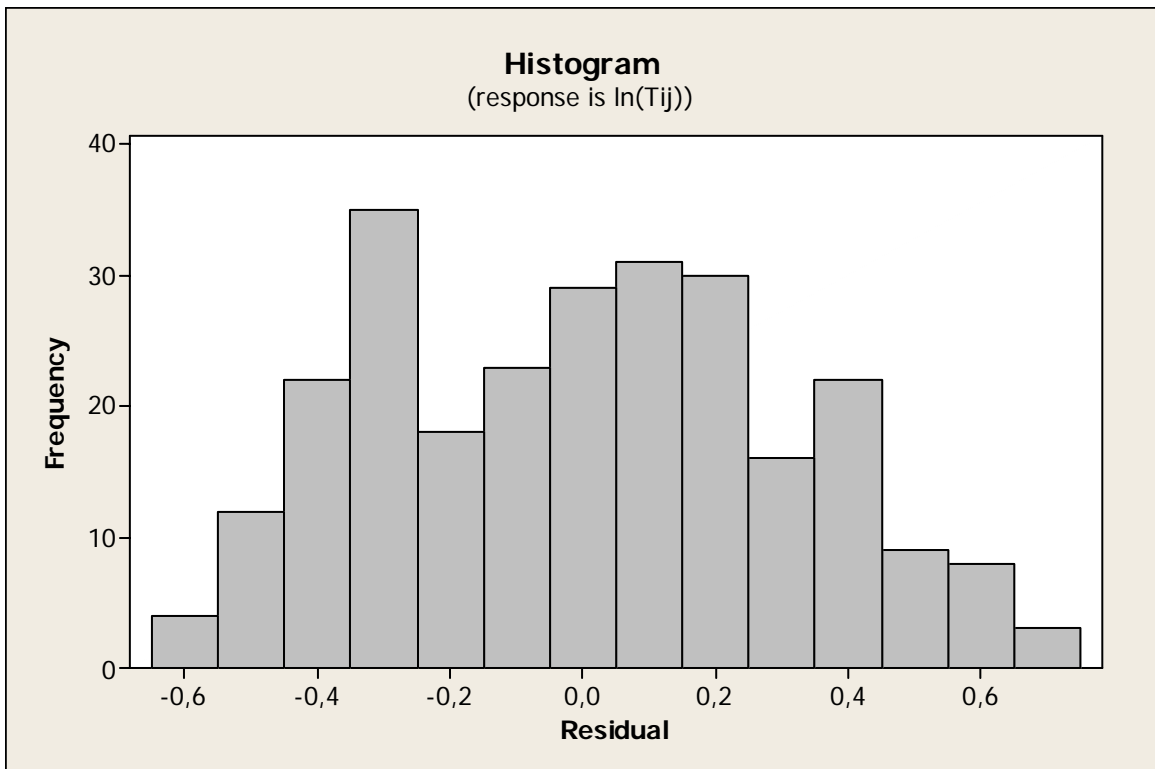


Figura 16 - Histograma dos Resíduos do Grupo 3 sem a Retirada dos Outliers

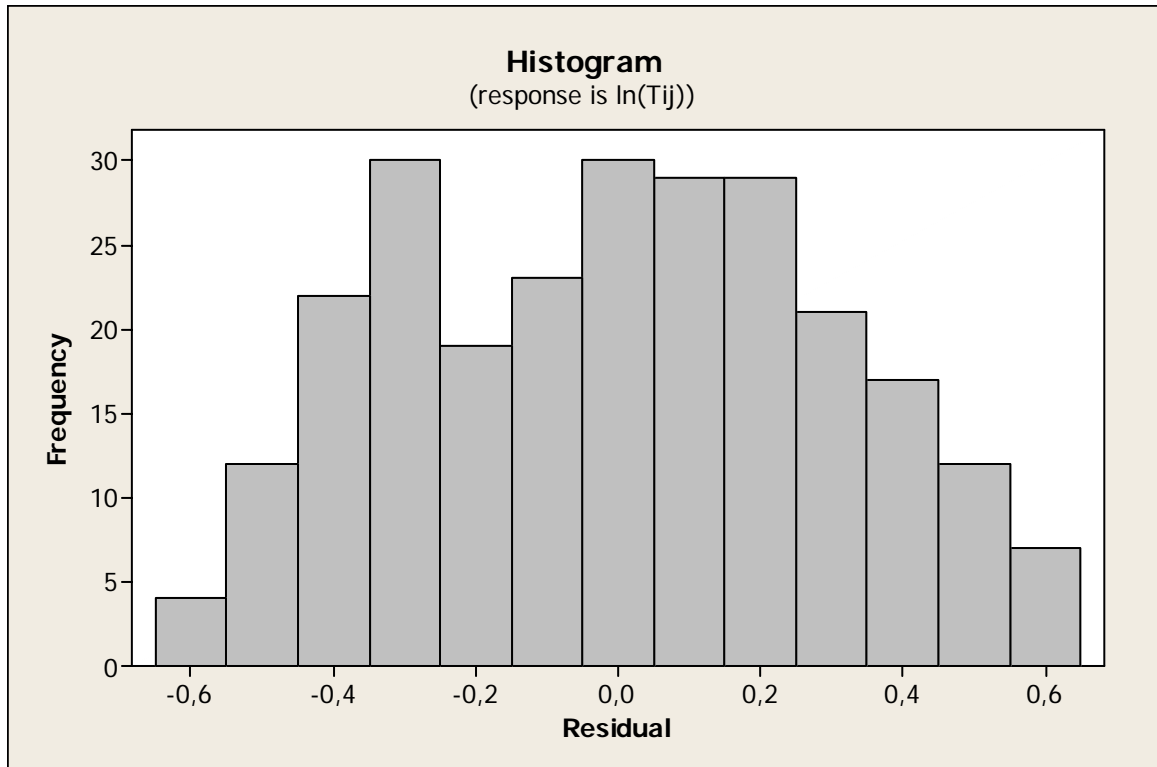


Figura 17 - Histograma dos Resíduos do Grupo 3 com a Retirada dos Outliers

Como percebe-se pelos histogramas, após a retirada dos *outliers* os resíduos seguiram uma distribuição mais perto da normal do que os anteriores. Por esta razão, decidiu-se adotar a regressão feita sem os pontos discrepantes, e além disso, aceita-se a hipótese de normalidade dos resíduos para os grupos 1, 2 e 3.

Os resíduos do grupo 4 apresentaram uma distribuição normal, como pode ser observado no gráfico que segue. Por este motivo, verifica-se a hipótese I também para o grupo 4, e não sente-se a necessidade da retirada dos *outliers*.

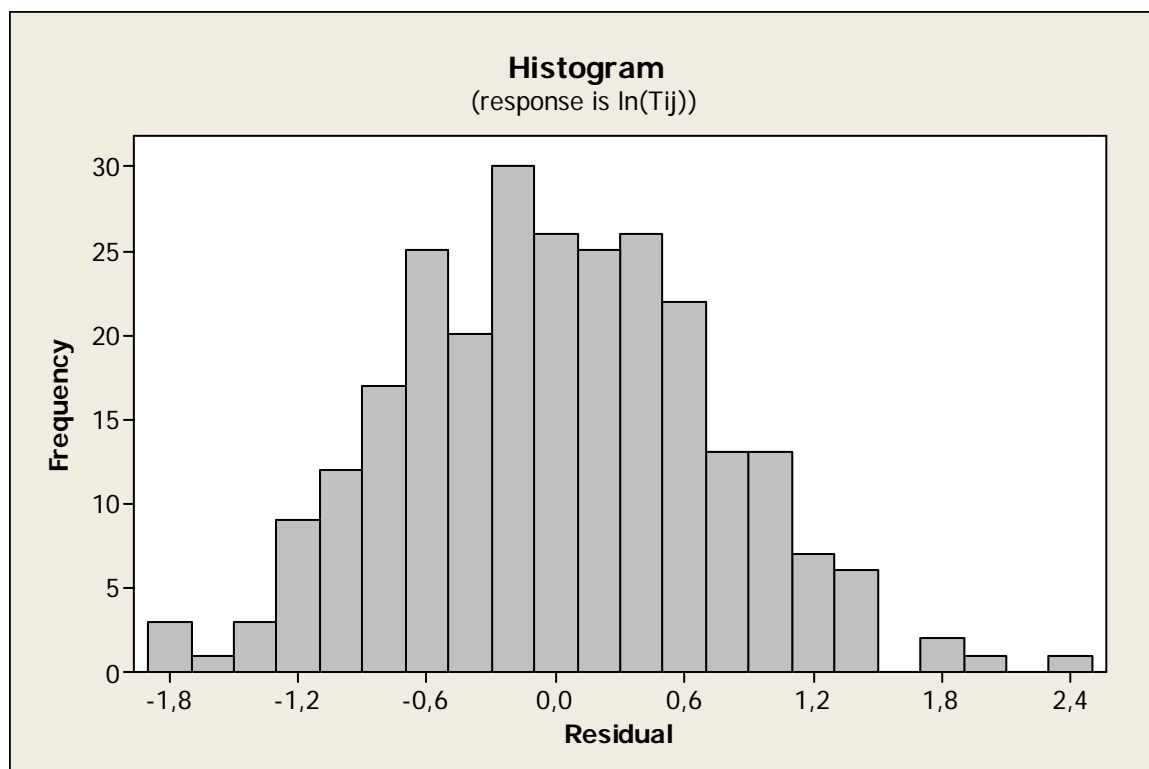


Figura 18 - Histograma dos Resíduos do Grupo 4 sem a Retirada dos Outliers

Hipótese III: A presença de homocedasticidade foi confirmada pelo teste F de comparação de variâncias, no qual foi dividido cada grupo em 3 amostras classificadas por distância. A tabela 8 mostra o número de observações em cada amostra e o valor do teste F calculado.

Tabela 8 - Teste F dos Grupos segmentados por PKm para testar a Homocedasticidade

Grupo	Nº de observações de maior valor	Nº de observações centrais	Nº de observações de menor valor	teste F	Graus de Liberdade
Grupo 1	96	70	96	2,16	5
Grupo 2	96	71	96	1,01	5
Grupo 3	96	71	96	1,24	6
Grupo 4	96	70	96	1,28	6

O valor do teste F tabelado para 5 graus de liberdade é 5,05, e para 6 graus é 4,25 e como pode-se 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 quatro grupos não são heterocedásticos.

Hipótese IV: A não-multicolinearidade foi satisfeita ao adotar-se o método Stepwise Regression, que apresenta os melhores conjuntos de variáveis sem a alta correlação das variáveis.

Como se encontrou bons valores de R^2 (Grupo 1 = 57,5; Grupo 2 = 82,3; Grupo 3 = 82,0 e Grupo 4 = 64,4), e todas as hipóteses foram satisfeitas, os modelos apresentados para os grupos em questão são definidos da seguinte forma:

- **Grupo 1**

$$T_{ij} = \exp\left(\alpha_1 + \alpha_2 \ln\left(\frac{P_i}{1000}\right) + \alpha_3 \ln\left(\frac{P_j}{1000}\right) + \alpha_4 \ln(m_i) + \alpha_5 \ln(d_{ij})\right) \quad (4.6)$$

onde,

$$\alpha_1 = 10,688203010663300$$

$$\alpha_2 = 0,271412242050494$$

$$\alpha_3 = 0,180330647292771$$

$$\alpha_4 = 0,106440805438319$$

$$\alpha_5 = -1,443456501090950$$

- **Grupo 2**

$$T_{ij} = \exp\left(\alpha_1 + \alpha_2 \ln\left(\frac{P_i}{1000}\right) + \alpha_3 \ln(mig_i) + \alpha_4 \ln(d_{ij}) + \alpha_5 \ln(s_j)\right) \quad (4.7)$$

onde,

$$\alpha_1 = 11,127431679863100$$

$$\alpha_2 = 0,055294625148262$$

$$\alpha_3 = 0,047252224252693$$

$$\alpha_4 = -1,069865780241610$$

$$\alpha_5 = 0,089036441377276$$

- **Grupo 3**

$$T_{ij} = \exp\left(\alpha_1 + \alpha_2 \ln\left(\frac{P_i}{1000}\right) + \alpha_3 \ln\left(\frac{P_j}{1000}\right) + \alpha_4 \ln(m_i) + \alpha_5 \ln(m_j) + \alpha_6 \ln(d_{ij})\right) \quad (4.8)$$

onde,

$$\begin{aligned}\alpha_1 &= 11,860735716881300 \\ \alpha_2 &= 0,064949894957539 \\ \alpha_3 &= 0,071733066512199 \\ \alpha_4 &= 0,035619313716280 \\ \alpha_5 &= 0,065903180070839 \\ \alpha_6 &= -1,033944878487430\end{aligned}$$

- **Grupo 4**

$$T_{ij} = \exp\left(\alpha_1 + \alpha_2 \ln\left(\frac{P_i}{1000}\right) + \alpha_3 \ln(m_i) + \alpha_4 \ln(d_{ij}) + \alpha_5 \ln(s_j) + \alpha_6 \ln(dummy)\right) \quad (4.9)$$

onde,

$$\begin{aligned}\alpha_1 &= 12,453732641926400 \\ \alpha_2 &= 0,595395207717770 \\ \alpha_3 &= 0,206471261035176 \\ \alpha_4 &= -1,342747476565210 \\ \alpha_5 &= 0,242493035882411 \\ \alpha_6 &= 0,853786684685158\end{aligned}$$

4.3 Comparação dos Resultados

Uma das formas de se avaliar a confiabilidade de um modelo é através da análise do valor do coeficiente de determinação ou R^2 , pois como dito anteriormente, ele mostra o poder de explicação do modelo. A tabela seguinte mostra o valor desta estatística para cada estratificação realizada.

Tabela 9 - Valor de R^2 de cada Grupo

Estratificação por distância	
<i>Grupo</i>	R^2
Grupo A	58,2
Grupo B	51,8
Grupo C	62,3

Estratificação por PKm	
<i>Grupo</i>	R^2
Grupo 1	57,5
Grupo 2	82,3
Grupo 3	82,0
Grupo 4	64,4

Através desta tabela, pode-se perceber que os valores de R^2 apresentados justificam uma boa explicação dos modelos para o referido estudo.

Além disso, as hipóteses básicas para regressão citadas anteriormente devem ser satisfeitas. Verificou-se que em ambos os modelos, tais hipóteses foram confirmadas, porém, a normalidade residual para o primeiro modelo se mostrou mais satisfatória do que para o segundo. Tal fato, no entanto não elimina a possibilidade de utilização deste.

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

5 CONSIDERAÇÕES FINAIS

Neste estudo apresentamos duas propostas de modelos para serem usados como ferramenta para Planejamento de Transportes. Foram abordados alguns dos principais métodos de modelagem matemática para distribuição de viagens. Entende-se que a metodologia de análise e a coleta de informações referente ao tema abordado são de consistência significativa e adequada aos interesses propostos.

O estudo foi embasado nos principais modelos de distribuição de viagens contidos na literatura. Decidiu-se adotar um modelo híbrido formado pela junção das características dos modelos gravitacional e econométricos.

A escolha dos modelos foi baseada em um estudo existente para a estimação de demanda de passageiros para todas as ligações inter-estaduais do Brasil, onde o mesmo utiliza uma segmentação por distância. No primeiro modelo proposto, foram mantidas as características deste estudo. Já no segundo modelo, a estratificação dos dados foi realizada utilizando como unidade observacional a variável Passageiro x Quilômetro e não mais a distância. Além disso, foram inseridas duas novas variáveis explicativas. Vale salientar que nos dois modelos trabalhados foram utilizadas somente ligações inter-estaduais da região sul do país.

A partir da escolha dos modelos citados, foram obtidos os melhores conjuntos de variáveis para cada estratificação realizada, logo os parâmetros foram estimados através do método de regressão, o que proporcionou um ajuste plausível na calibração. A credibilidade dos resultados foi analisada através da verificação das hipóteses básicas para regressão.

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.

BIBLIOGRAFIA

ALMEIDA, Lourdes Maria Werle. **Desenvolvimento de uma Metodologia para Análise Locacional de Sistemas Educacionais usando Modelos de Interação Espacial e Indicadores de Acessibilidade**. Tese de Doutorado. Universidade Federal de Santa Catarina. Florianópolis, 1999.

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

BARBETTA, Pedro Alberto. **Estatística Aplicada as Ciências Sociais**. 3ª edição, ed. da UFSC. Florianópolis, 1999.

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

BEZ, Edson Tadeu. **Procedimento de Representação de Soluções em Otimização Global: Aplicação em Modelos de Interação Espacial**. Tese de Doutorado. Universidade Federal de Santa Catarina. Florianópolis, 2005.

BEZ, Edson Tadeu. **Um estudo sobre os procedimentos de calibração de alguns modelos de distribuição de viagens**. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis, 2000.

BRUTON, M. J. **Introdução ao Planejamento dos Transportes**. Interciência, Rio de Janeiro, 1979.

D'HAINAUT, Louis. **Conceitos e Métodos da Estatística**. Volume I – Uma variável a uma dimensão. Tradução: Antonio Rodrigues Lopes. 2ª edição, ed. FUNDAÇÃO CALOUSTE GULBENKIAN. Lisboa, 1997.

D'HAINAUT, Louis. **Conceitos e Métodos da Estatística**. Volume II – Duas ou três variáveis segundo duas ou três dimensões. Tradução: Antonio Rodrigues Lopes. Ed. FUNDAÇÃO CALOUSTE GULBENKIAN. Lisboa, 1992.

DORINI, Fábio Antonio. **O princípio da Máxima Entropia, o Método dos Multiplicadores de Lagrange e suas Aplicações à dedução de Modelos usados no Planejamento.** Monografia. Universidade Federal de Santa Catarina. Florianópolis, 1997.

FIPE – Fundação Instituto de Pesquisas Econômicas. <<http://www.fipe.org.br/>>. Acessado em 12/12/2006.

FUNDAÇÃO UNIVERSITÁRIA JOSÈ BONIFÁCIO. “**Relatório 03 – Modelos de Demanda**” – Consultoria Técnica, para a ANTT, visando o desenvolvimento do transporte rodoviário interestadual e internacional de passageiros. Rio de Janeiro, 2003.

GONÇALVES, Mirian Buss. **Desenvolvimento e teste de um novo modelo gravitacional de oportunidades para distribuição de viagens.** Tese de Doutorado. Universidade Federal de Santa Catarina. Florianópolis, 1992.

GONÇALVES, Mirian Buss; BEZ, Edson Tadeu; NOVAES, Abtônio Galvão. **Modelos Econométricos Aplicados a Previsão da Demanda por Transporte Interestadual de Passageiros por Ônibus no Brasil.** Artigo, Universidade Federal de Santa Catarina. Florianópolis, 2006.

GONÇALVES, Mirian Buss; SOUZA DE CURSI, J. E. **Métodos Robustos para a Calibração de Modelos de Interação Especial em Transporte.** Associação Nacional de Pesquisas e Ensino em Transportes, 11., 1997. Anais... v.2, p. 303-313, 1997.

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.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2000.** Rio de Janeiro, 2000.

KAZMIER, Leonard J. **Estatística Aplicada a Economia e Administração**. Tradução: Carlos Augusto Crusius. Ed. Makron. São Paulo, 1982.

KUHLKAMP, Nilo. **Modelos de Oportunidades Intervenientes, de Distribuição de Viagens, com Ponderação das Posições Espaciais Relativas das Oportunidades**. Tese de Doutorado. Universidade Federal de Santa Catarina. Florianópolis, 2003.

LABTRABS/UFSC – Laboratório de Transportes e Logística da Universidade Federal de Santa Catarina. **“Relatório 1 – Conceitual” 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.

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.

MATOS, Orlando Carneiro. **Econometria Básica: Teoria e Aplicação**. 2ª edição, ed. Atlas. São Paulo, 1997.

NOVAES, Antonio Galvão. **Modelos em Planejamento Urbano, Regional e de Transportes**. Edgard Blucher, São Paulo, 1982.

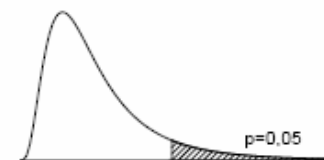
SARTORIS, Alexandre. **Estatística e Introdução à Econometria**. Ed. Saraiva. São Paulo, 2003.

TINTNER, G. **Elementos de Econometria**. Tradução: Luiz de Freitas Bueno. Ed. Pioneira. São Paulo, 1965.

ANEXOS

ANEXO 1: TABELA DE DISTRIBUIÇÃO F

Distribuição F de Snedecor a 5% ($p=0.05$)



	1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	30	40	60	120
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.43	19.44	19.45	19.46	19.47	19.48	19.49
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.70	8.69	8.67	8.66	8.62	8.59	8.57	8.55
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.86	5.84	5.82	5.80	5.75	5.72	5.69	5.66
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.62	4.60	4.58	4.56	4.50	4.46	4.43	4.40
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.94	3.92	3.90	3.87	3.81	3.77	3.74	3.70
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.51	3.49	3.47	3.44	3.38	3.34	3.30	3.27
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.22	3.20	3.17	3.15	3.08	3.04	3.01	2.97
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	3.01	2.99	2.96	2.94	2.86	2.83	2.79	2.75
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.85	2.83	2.80	2.77	2.70	2.66	2.62	2.58
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.72	2.70	2.67	2.65	2.57	2.53	2.49	2.45
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.62	2.60	2.57	2.54	2.47	2.43	2.38	2.34
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.53	2.51	2.48	2.46	2.38	2.34	2.30	2.25
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.46	2.44	2.41	2.39	2.31	2.27	2.22	2.18
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.40	2.38	2.35	2.33	2.25	2.20	2.16	2.11
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.35	2.33	2.30	2.28	2.19	2.15	2.11	2.06
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.31	2.29	2.26	2.23	2.15	2.10	2.06	2.01
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.27	2.25	2.22	2.19	2.11	2.06	2.02	1.97
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.23	2.21	2.18	2.16	2.07	2.03	1.98	1.93
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.20	2.18	2.15	2.12	2.04	1.99	1.95	1.90
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.20	2.18	2.16	2.12	2.10	2.01	1.96	1.92	1.87
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.17	2.15	2.13	2.10	2.07	1.98	1.94	1.89	1.84
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.15	2.13	2.11	2.08	2.05	1.96	1.91	1.86	1.81
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.13	2.11	2.09	2.05	2.03	1.94	1.89	1.84	1.79
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.09	2.07	2.04	2.01	1.92	1.87	1.82	1.77
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.09	2.07	2.05	2.02	1.99	1.90	1.85	1.80	1.75
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.08	2.06	2.04	2.00	1.97	1.88	1.84	1.79	1.73
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.06	2.04	2.02	1.99	1.96	1.87	1.82	1.77	1.71
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.05	2.03	2.01	1.97	1.94	1.85	1.81	1.75	1.70
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	2.01	1.99	1.96	1.93	1.84	1.79	1.74	1.68
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.92	1.90	1.87	1.84	1.74	1.69	1.64	1.58
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.86	1.84	1.82	1.78	1.75	1.65	1.59	1.53	1.47
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.78	1.75	1.73	1.69	1.66	1.55	1.50	1.43	1.35

Tabela 5: Quantis da Distribuição F para probabilidade $p = P[F \geq F_{\alpha}] = 0,05$. Graus de liberdade do numerador dado no topo e do denominador na margem esquerda.

ANEXO 2: MÉTODOS STEPWISE REGRESSION, BEST SUBSETS REGRESSION E A REGRESSÃO PARA O GRUPO A

Stepwise Regression: $\ln(T_{ij})$ versus $\ln(d_{ij})$; $\ln(P_i/1000)$; ...

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is $\ln(T_{ij})$ on 9 predictors, with N = 574

Step	1	2	3	4	5	6
Constant	3,775	9,881	9,001	8,515	7,385	7,472
$\ln(P_j/1000)$	0,569	0,724	0,825	0,602	0,617	0,495
T-Value	10,31	13,63	16,39	11,19	11,80	7,24
P-Value	0,000	0,000	0,000	0,000	0,000	0,000
$\ln(d_{ij})$		-1,26	-1,16	-1,25	-1,04	-1,05
T-Value		-10,14	-10,08	-11,45	-9,34	-9,48
P-Value		0,000	0,000	0,000	0,000	0,000
$\ln(m_i)$			0,403	0,420	0,442	0,436
T-Value			9,73	10,77	11,61	11,50
P-Value			0,000	0,000	0,000	0,000
$\ln(P_i/1000)$				0,713	0,727	0,763
T-Value				8,73	9,16	9,54
P-Value				0,000	0,000	0,000
$\ln(m_j)$					0,335	0,308
T-Value					5,93	5,40
P-Value					0,000	0,000
$\ln(\text{Dummy})$						1,57
T-Value						2,73
P-Value						0,006
S	1,72	1,58	1,47	1,38	1,34	1,33
R-Sq	15,66	28,53	38,72	45,95	49,10	49,76
R-Sq(adj)	15,51	28,28	38,39	45,57	48,65	49,23
Mallows C-p	395,1	249,8	135,3	54,5	20,4	14,9

Step	7
Constant	7,225
$\ln(P_j/1000)$	0,494
T-Value	7,28
P-Value	0,000
$\ln(d_{ij})$	-1,08
T-Value	-9,79
P-Value	0,000
$\ln(m_i)$	0,434
T-Value	11,54
P-Value	0,000
$\ln(P_i/1000)$	0,754

T-Value 9,49
P-Value 0,000

ln(mj) 0,351
T-Value 6,01
P-Value 0,000

ln(Dummy) 1,78
T-Value 3,10
P-Value 0,002

ln(Migj) -0,192
T-Value -3,08
P-Value 0,002

S 1,32
R-Sq 50,59
R-Sq(adj) 49,98
Mallows C-p 7,4

Best Subsets Regression: ln(Tij) versus ln(dij); ln(Pi/1000); ...

Response is ln(Tij)

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	ln(dij)	ln(Pi/1000)	ln(mj)	ln(Dummy)	ln(Migj)
1	15,7	15,5	395,1	1,7178					
1	15,5	15,3	397,3	1,7198	X				
2	28,5	28,3	249,8	1,5826	X	X			
2	27,8	27,5	258,2	1,5908		X	X		
3	38,7	38,4	135,3	1,4669	X	X	X		
3	34,9	34,6	178,5	1,5114	X	X	X		
4	45,9	45,6	54,5	1,3788	X	X	X	X	
4	43,4	43,0	83,2	1,4105	X	X		X	X
5	49,1	48,7	20,4	1,3392	X	X	X	X	X
5	47,2	46,7	42,4	1,3642	X	X	X	X	X
6	49,8	49,2	14,9	1,3316	X	X	X	X	X
6	49,8	49,2	15,0	1,3317	X	X	X	X	X
7	50,6	50,0	7,4	1,3218	X	X	X	X	X
7	49,8	49,2	16,0	1,3317	X	X	X	X	X
8	50,7	50,0	8,5	1,3219	X	X	X	X	X
8	50,6	49,9	8,9	1,3223	X	X	X	X	X
9	50,7	49,9	10,0	1,3225	X	X	X	X	X

Regression Analysis: ln(Tij) versus ln(Pj/1000); ln(dij); ...

The regression equation is

$$\ln(Tij) = 7,47 + 0,495 \ln(Pj/1000) - 1,05 \ln(dij) + 0,436 \ln(mi) + 0,763 \ln(Pi/1000) + 0,308 \ln(mj) + 1,57 \ln(Dummy)$$

Predictor	Coef	SE Coef	T	P
Constant	7,4717	0,5777	12,93	0,000
ln(Pj/1000)	0,49539	0,06841	7,24	0,000
ln(dij)	-1,0514	0,1109	-9,48	0,000
ln(mi)	0,43554	0,03788	11,50	0,000
ln(Pi/1000)	0,76309	0,07999	9,54	0,000
ln(mj)	0,30808	0,05710	5,40	0,000
ln(Dummy)	1,5682	0,5737	2,73	0,006

S = 1,33162 R-Sq = 49,8% R-Sq(adj) = 49,2%

ANEXO 3: MÉTODOS STEPWISE REGRESSION, BEST SUBSETS REGRESSION E A REGRESSÃO PARA O GRUPO B

Stepwise Regression: ln(Tij) versus ln(dij); ln(Pi/1000); ...

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is ln(Tij) on 9 predictors, with N = 403

Step	1	2	3	4	5	6
Constant	1,7581	0,5902	0,4393	1,0253	1,2211	7,7066
ln(Pj/1000)	0,770	0,558	0,556	0,531	0,399	0,424
T-Value	11,60	8,33	8,51	8,42	5,05	5,37
P-Value	0,000	0,000	0,000	0,000	0,000	0,000
ln(Pi/1000)		0,718	0,790	0,767	0,763	0,796
T-Value		8,10	9,00	9,07	9,09	9,44
P-Value		0,000	0,000	0,000	0,000	0,000
ln(mi)			0,270	0,360	0,339	0,312
T-Value			4,68	6,24	5,87	5,35
P-Value			0,000	0,000	0,000	0,000
ln(mj)				0,355	0,300	0,268
T-Value				5,71	4,63	4,08
P-Value				0,000	0,000	0,000
ln(Dummy)					1,46	1,54
T-Value					2,73	2,90
P-Value					0,007	0,004
ln(dij)						-1,07
T-Value						-2,57
P-Value						0,011
S	1,71	1,59	1,55	1,49	1,48	1,47
R-Sq	25,12	35,68	39,03	43,64	44,68	45,59
R-Sq(adj)	24,94	35,36	38,57	43,07	43,98	44,76
Mallows C-p	146,8	71,9	49,4	17,8	12,2	7,6
Step	7					
Constant	7,830					
ln(Pj/1000)	0,425					
T-Value	5,40					
P-Value	0,000					
ln(Pi/1000)	0,788					
T-Value	9,35					
P-Value	0,000					
ln(mi)	0,260					
T-Value	4,02					
P-Value	0,000					
ln(mj)	0,306					

T-Value 4,45
P-Value 0,000

ln(Dummy) 1,52
T-Value 2,87
P-Value 0,004

ln(dij) -1,04
T-Value -2,49
P-Value 0,013

ln(Migi) 0,137
T-Value 1,81
P-Value 0,070

S 1,46
R-Sq 46,04
R-Sq(adj) 45,08
Mallows C-p 6,3

Best Subsets Regression: ln(Tij) versus ln(dij); ln(Pi/1000); ...

Response is ln(Tij)

Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	l n (R l l i n n * ((R l P P j l l n l i j / n n (n / / 1 ((D l l (1 1 0 M M u n n d 0 0 0 i i m ((i 0 0 0 g g m m m j 0 0 0 i j y i j)))))))))
1	25,1	24,9	146,8	1,7110	X
1	24,5	24,3	151,2	1,7178	X
2	36,0	35,6	69,8	1,5844	X X
2	35,7	35,4	71,9	1,5878	X X
3	39,0	38,6	49,4	1,5478	X X X
3	38,9	38,4	50,4	1,5495	X X X
4	43,6	43,1	17,8	1,4901	X X X X
4	41,7	41,1	32,0	1,5156	X X X X
5	44,7	44,0	12,2	1,4781	X X X X X
5	44,4	43,7	14,0	1,4814	X X X X X
6	45,6	44,8	7,6	1,4678	X X X X X
6	45,2	44,4	10,5	1,4731	X X X X X
7	46,0	45,1	6,3	1,4635	X X X X X
7	45,9	44,9	7,3	1,4654	X X X X X
8	46,1	45,0	8,1	1,4648	X X X X X
8	46,0	44,9	8,3	1,4654	X X X X X
9	46,1	44,8	10,0	1,4666	X X X X X

Regression Analysis: ln(Tij) versus ln(Pj/1000); ln(dij); ...

The regression equation is

$$\ln(Tij) = 7,71 + 0,424 \ln(Pj/1000) - 1,07 \ln(dij) + 0,312 \ln(mi) + 0,796 \ln(Pi/1000) + 0,268 \ln(mj) + 1,54 \ln(Dummy)$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	7,707	2,546	3,03	0,003	
ln(Pj/1000)	0,42358	0,07895	5,37	0,000	1,9
ln(dij)	-1,0750	0,4182	-2,57	0,011	1,2
ln(mi)	0,31211	0,05838	5,35	0,000	1,2
ln(Pi/1000)	0,79562	0,08431	9,44	0,000	1,3
ln(mj)	0,26800	0,06562	4,08	0,000	1,3
ln(Dummy)	1,5370	0,5305	2,90	0,004	1,9

S = 1,46775 R-Sq = 45,6% R-Sq(adj) = 44,8%

PRESS = 882,537 R-Sq(pred) = 43,71%

ANEXO 4: MÉTODOS STEPWISE REGRESSION, BEST SUBSETS REGRESSION E A REGRESSÃO PARA O GRUPO C

Stepwise Regression: $\ln(T_{ij})$ versus $\ln(d_{ij})$; $\ln(P_i/1000)$; ...

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is $\ln(T_{ij})$ on 9 predictors, with N = 70

Step	1	2	3	4	5	6
Constant	2,1614	1,6547	2,3367	0,8262	20,3687	21,6662
$\ln(P_i/1000)$	1,07	0,94	0,86	0,84	0,80	0,82
T-Value	5,90	5,51	5,50	5,44	5,34	5,52
P-Value	0,000	0,000	0,000	0,000	0,000	0,000
$\ln(\text{Dummy})$		2,35	2,38	1,67	2,01	1,35
T-Value		3,61	4,03	2,48	2,96	1,80
P-Value		0,001	0,000	0,016	0,004	0,077
$\ln(m_i)$			0,329	0,382	0,417	0,450
T-Value			3,91	4,42	4,85	5,22
P-Value			0,000	0,000	0,000	0,000
$\ln(P_j/1000)$				0,35	0,38	0,44
T-Value				2,03	2,25	2,59
P-Value				0,047	0,028	0,012
$\ln(d_{ij})$					-2,9	-3,1
T-Value					-2,03	-2,19
P-Value					0,046	0,033
$\ln(m_j)$						0,26
T-Value						1,86
P-Value						0,068
S	1,47	1,36	1,23	1,20	1,18	1,15
R-Sq	33,85	44,64	55,04	57,71	60,28	62,34
R-Sq(adj)	32,87	42,99	52,99	55,11	57,17	58,75
Mallows C-p	43,3	27,4	12,3	9,8	7,6	6,2

Best Subsets Regression: $\ln(T_{ij})$ versus $\ln(d_{ij})$; $\ln(P_i/1000)$; ...

Response is $\ln(T_{ij})$

```

1
n
(
R
l l i
n n *
( ( R l
P P j l l n
l i j / n n (
n / / 1 ( ( D l l

```


ANEXO 5: REGRESSÃO GRUPO A COM A RETIRADA DE OUTLIERS

Regression Analysis: ln(Tij) versus ln(Pj/1000); ln(dij); ...

The regression equation is

$$\ln(T_{ij}) = 6,81 + 0,430 \ln(P_j/1000) - 0,905 \ln(d_{ij}) + 0,413 \ln(m_i) + 0,834 \ln(P_i/1000) + 0,339 \ln(m_j) + 1,88 \ln(\text{Dummy})$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	6,8136	0,5022	13,57	0,000	
ln(Pj/1000)	0,42964	0,05850	7,34	0,000	2,6
ln(dij)	-0,90527	0,09640	-9,39	0,000	1,2
ln(mi)	0,41293	0,03244	12,73	0,000	1,1
ln(Pi/1000)	0,83402	0,06821	12,23	0,000	1,4
ln(mj)	0,33916	0,04990	6,80	0,000	1,2
ln(Dummy)	1,8847	0,5021	3,75	0,000	1,9

S = 1,09212 R-Sq = 58,2% R-Sq(adj) = 57,7%

ANEXO 6: REGRESSÃO GRUPO B COM A RETIRADA DE OUTLIERS

Regression Analysis: ln(Tij) versus ln(Pj/1000); ln(dij); ...

The regression equation is

$$\ln(T_{ij}) = 5,40 + 0,342 \ln(P_j/1000) - 0,600 \ln(d_{ij}) + 0,335 \ln(m_i) \\ + 0,756 \ln(P_i/1000) + 0,314 \ln(m_j) + 1,62 \ln(\text{Dummy})$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	5,396	2,228	2,42	0,016	
ln(Pj/1000)	0,34228	0,06804	5,03	0,000	1,9
ln(dij)	-0,6003	0,3693	-1,63	0,105	1,2
ln(mi)	0,33485	0,05106	6,56	0,000	1,2
ln(Pi/1000)	0,75613	0,07495	10,09	0,000	1,3
ln(mj)	0,31428	0,06048	5,20	0,000	1,3
ln(Dummy)	1,6189	0,4548	3,56	0,000	1,9

S = 1,22829 R-Sq = 51,8% R-Sq(adj) = 51,0%

PRESS = 579,309 R-Sq(pred) = 49,95%

ANEXO 7: MÉTODOS STEPWISE REGRESSION, BEST SUBSETS REGRESSION E A REGRESSÃO PARA O GRUPO 1

Stepwise Regression: ln(Tij) versus ln(dij); ln(Pi/1000); ...

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is ln(Tij) on 13 predictors, with N = 262

Step	1	2	3	4
Constant	10,66	10,65	10,56	10,68
ln(dij)	-1,245	-1,365	-1,355	-1,519
T-Value	-13,53	-13,73	-13,73	-14,20
P-Value	0,000	0,000	0,000	0,000
ln(Pi/10)		0,293	0,329	0,273
T-Value		2,94	3,30	2,76
P-Value		0,004	0,001	0,006
ln(mi)			0,083	0,133
T-Value			2,41	3,65
P-Value			0,017	0,000
ln(Pj/10)				0,261
T-Value				3,56
P-Value				0,000
S	1,07	1,06	1,05	1,02
R-Sq	41,31	43,21	44,46	47,07
R-Sq(adj)	41,08	42,77	43,81	46,24
C-p	22,9	15,8	11,9	1,4

Best Subsets Regression: ln(Tij) versus ln(dij); ln(Pi/1000); ...

Response is ln(Tij)

Vars	R-Sq	R-Sq(adj)	C-p	S	Model
1	41,3	41,1	22,9	1,0710	X
1	14,1	13,7	153,4	1,2961	X
2	43,2	42,8	15,8	1,0556	X X
2	43,0	42,6	16,7	1,0573	X X
3	45,5	44,9	6,9	1,0362	X X X
3	45,2	44,5	8,4	1,0391	X X X
4	47,1	46,2	1,4	1,0231	X X X X
4	46,8	46,0	2,7	1,0257	X X X X
5	47,5	46,4	1,4	1,0211	X X X X X
5	47,4	46,4	1,6	1,0215	X X X X X

6	47,9	46,7	1,2	1,0185	X X X X	X X
6	47,6	46,4	2,7	1,0216	X X X X	X X
7	48,0	46,6	2,9	1,0199	X X X X	X X X
7	48,0	46,6	3,0	1,0201	X X X X X	X X
8	48,1	46,4	4,6	1,0214	X X X X X X	X X
8	48,1	46,4	4,6	1,0215	X X X X X	X X X
9	48,1	46,3	6,3	1,0227	X X X X X X X X	
9	48,1	46,2	6,5	1,0231	X X X X X X X X	X
10	48,2	46,1	8,1	1,0243	X X X X X X X X	X
10	48,1	46,1	8,2	1,0246	X X X X X X X X	X
11	48,2	45,9	10,0	1,0262	X X X X X X X X	X X
11	48,2	45,9	10,1	1,0264	X X X X X X X X	X X
12	48,2	45,7	12,0	1,0283	X X X X X X X X	X X X
12	48,2	45,7	12,0	1,0283	X X X X X X X X X	X X
13	48,2	45,5	14,0	1,0304	X X X X X X X X X X	X X X

Regression Analysis: ln(Tij) versus ln(dij); ln(Pi/1000); ...

The regression equation is

$$\ln(T_{ij}) = 10,7 - 1,52 \ln(d_{ij}) + 0,273 \ln(P_i/1000) + 0,261 \ln(P_j/1000) + 0,133 \ln(m_i)$$

Predictor	Coef	SE Coef	T	P
Constant	10,6829	0,5030	21,24	0,000
ln(dij)	-1,5191	0,1070	-14,20	0,000
ln(Pi/1000)	0,27331	0,09888	2,76	0,006
ln(Pj/1000)	0,26137	0,07347	3,56	0,000
ln(mi)	0,13291	0,03637	3,65	0,000

S = 1,02308 R-Sq = 47,1% R-Sq(adj) = 46,2%

ANEXO 8: MÉTODOS STEPWISE REGRESSION, BEST SUBSETS REGRESSION E A REGRESSÃO PARA O GRUPO 2

Stepwise Regression: $\ln(T_{ij})$ versus $\ln(d_{ij})$; $\ln(P_i/1000)$; ...

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is $\ln(T_{ij})$ on 13 predictors, with N = 260

Step	1	2	3	4
Constant	11,02	11,03	11,11	11,10
$\ln(d_{ij})$	-0,993	-1,034	-1,035	-1,049
T-Value	-33,28	-31,89	-32,19	-31,92
P-Value	0,000	0,000	0,000	0,000
$\ln(s_j)$		0,068	0,072	0,058
T-Value		3,01	3,23	2,45
P-Value		0,003	0,001	0,015
$\ln(M_{ij})$			0,041	0,041
T-Value			2,32	2,33
P-Value			0,021	0,020
$\ln(P_i/1000)$				0,051
T-Value				1,82
P-Value				0,071
S	0,328	0,323	0,320	0,319
R-Sq	81,11	81,75	82,13	82,36
R-Sq(adj)	81,04	81,61	81,92	82,08

Best Subsets Regression: $\ln(T_{ij})$ versus $\ln(d_{ij})$; $\ln(P_i/1000)$; ...

Response is $\ln(T_{ij})$

	1	2	3	4	5	6	7	8	9	10	11	12	13
Mallows													

ANEXO 9: MÉTODOS STEPWISE REGRESSION, BEST SUBSETS REGRESSION E A REGRESSÃO PARA O GRUPO 3

Stepwise Regression: ln(Tij) versus ln(dij); ln(Pi/1000); ...

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is ln(Tij) on 13 predictors, with N = 262

Step	1	2	3	4	5	6
Constant	12,21	12,27	12,06	12,06	11,83	11,81
ln(dij)	-1,016	-1,069	-1,033	-1,046	-1,018	-1,020
T-Value	-31,85	-30,02	-27,43	-27,27	-24,96	-25,07
P-Value	0,000	0,000	0,000	0,000	0,000	0,000
ln(Pj/1000)		0,060	0,066	0,056	0,062	0,140
T-Value		3,11	3,45	2,80	3,09	2,64
P-Value		0,002	0,001	0,005	0,002	0,009
ln(mj)			0,048	0,047	0,061	0,067
T-Value			2,68	2,62	3,15	3,40
P-Value			0,008	0,009	0,002	0,001
ln(Pi/1000)				0,040	0,060	0,061
T-Value				1,64	2,27	2,30
P-Value				0,102	0,024	0,022
ln(mi)					0,032	0,030
T-Value					1,88	1,80
P-Value					0,062	0,074
ln(sj)						-0,086
T-Value						-1,58
P-Value						0,115
S	0,322	0,317	0,313	0,312	0,311	0,310
R-Sq	79,60	80,34	80,87	81,07	81,32	81,51
R-Sq(adj)	79,52	80,19	80,65	80,77	80,96	81,07
Mallows Cp	16,8	8,9	3,7	3,0	1,6	1,1

Best Subsets Regression: ln(Tij) versus ln(dij); ln(Pi/1000); ...

Response is ln(Tij)

```

1
n
(
R
1 1 i
n n *
( ( R 1
P P j 1 1 n
1 i j / n n ( 1 1
n / / 1 ( ( D 1 1 1 1 n n

```

Vars	R-Sq	R-Sq(adj)	Mallows Cp	S						
1	79,6	79,5	16,8	0,32200	X					
1	20,3	20,0	814,9	0,63630					X	
2	80,3	80,2	8,9	0,31676	X	X				
2	80,1	80,0	12,0	0,31862	X	X				
3	80,9	80,6	3,7	0,31306	X	X			X	
3	80,6	80,4	7,5	0,31534	X				X	X
4	81,1	80,8	3,0	0,31203	X	X	X		X	
4	81,1	80,8	3,2	0,31211	X	X			X	X
5	81,3	81,0	1,6	0,31051	X	X	X		X	X
5	81,3	80,9	2,3	0,31095	X	X	X		X	X
6	81,5	81,1	1,1	0,30961	X	X	X		X	X
6	81,4	81,0	2,2	0,31029	X	X			X	X
7	81,5	81,0	2,6	0,30990	X	X	X		X	X
7	81,5	81,0	2,7	0,30997	X	X	X	X	X	X
8	81,6	81,0	4,2	0,31027	X	X	X	X	X	X
8	81,6	81,0	4,3	0,31032	X	X	X	X	X	X
9	81,6	80,9	6,1	0,31078	X	X	X	X	X	X
9	81,6	80,9	6,1	0,31080	X	X	X	X	X	X
10	81,6	80,9	8,0	0,31138	X	X	X	X	X	X
10	81,6	80,9	8,0	0,31138	X	X	X	X	X	X
11	81,6	80,8	10,0	0,31199	X	X	X	X	X	X
11	81,6	80,8	10,0	0,31199	X	X	X	X	X	X
12	81,6	80,7	12,0	0,31261	X	X	X	X	X	X
12	81,6	80,7	12,0	0,31261	X	X	X	X	X	X
13	81,6	80,6	14,0	0,31323	X	X	X	X	X	X

Regression Analysis: ln(Tij) versus ln(dij); ln(Pi/1000); ...

The regression equation is

$$\ln(T_{ij}) = 11,8 - 1,02 \ln(d_{ij}) + 0,0602 \ln(P_i/1000) + 0,0623 \ln(P_j/1000) + 0,0319 \ln(m_i) + 0,0611 \ln(m_j)$$

Predictor	Coef	SE Coef	T	P
Constant	11,8312	0,2337	50,63	0,000
ln(dij)	-1,01843	0,04080	-24,96	0,000
ln(Pi/1000)	0,06022	0,02649	2,27	0,024
ln(Pj/1000)	0,06226	0,02014	3,09	0,002
ln(mi)	0,03191	0,01700	1,88	0,062
ln(mj)	0,06114	0,01943	3,15	0,002

S = 0,310512 R-Sq = 81,3% R-Sq(adj) = 81,0%

ANEXO 10: MÉTODOS STEPWISE REGRESSION, BEST SUBSETS REGRESSION E A REGRESSÃO PARA O GRUPO 4

Stepwise Regression: ln(Tij) versus ln(dij); ln(Pi/1000); ...

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is ln(Tij) on 13 predictors, with N = 262

Step	1	2	3	4	5	6
Constant	13,51	13,67	12,84	11,92	12,45	12,53
ln(dij)	-0,919	-1,318	-1,390	-1,293	-1,343	-1,321
T-Value	-9,25	-14,39	-17,12	-16,23	-16,82	-16,58
P-Value	0,000	0,000	0,000	0,000	0,000	0,000
ln(Pi/1000)		0,618	0,532	0,604	0,595	0,614
T-Value		10,51	10,08	11,58	11,60	11,91
P-Value		0,000	0,000	0,000	0,000	0,000
ln(sj)			0,332	0,357	0,242	0,509
T-Value			8,65	9,68	4,75	3,99
P-Value			0,000	0,000	0,000	0,000
ln(mi)				0,220	0,206	0,226
T-Value				5,22	4,96	5,35
P-Value				0,000	0,000	0,000
ln(Dummy)					0,85	0,97
T-Value					3,18	3,59
P-Value					0,002	0,000
ln(Pj/1000)						-0,28
T-Value						-2,27
P-Value						0,024
S	1,06	0,891	0,786	0,749	0,736	0,730
R-Sq	24,77	47,25	59,11	63,03	64,44	65,15
R-Sq(adj)	24,49	46,84	58,63	62,46	63,75	64,33
Mallows Cp	292,6	130,1	45,3	18,6	10,3	7,1
Step	7					
Constant	12,64					
ln(dij)	-1,324					
T-Value	-16,74					
P-Value	0,000					
ln(Pi/1000)	0,37					
T-Value	3,04					
P-Value	0,003					
ln(sj)	0,54					
T-Value	4,23					
P-Value	0,000					

```

ln(mi)          0,208
T-Value         4,88
P-Value         0,000

ln(Dummy)       0,99
T-Value         3,69
P-Value         0,000

ln(Pj/1000)     -0,31
T-Value        -2,48
P-Value         0,014

ln(si)          0,26
T-Value         2,24
P-Value         0,026

S                0,724
R-Sq             65,82
R-Sq(adj)       64,88
Mallows Cp      4,2

```

Best Subsets Regression: ln(Tij) versus ln(dij); ln(Pi/1000); ...

Response is ln(Tij)

Vars	R-Sq	R-Sq(adj)	Mallows Cp	S	ln (R l l i n n * ((R l P P j l l n l i j / n n (l l n / / 1 ((D l l l l n n (1 1 0 M M u n n n n ((d 0 0 0 i i m (((s s i 0 0 0 g g m m m s s l l j 0 0 0 i j y i j i j i j))))))))))))
1	24,8	24,5	292,6	1,0617	X
1	10,4	10,1	397,7	1,1586	
2	47,3	46,8	130,1	0,89075	X X
2	45,8	45,4	140,6	0,90277	X
3	59,9	59,4	39,5	0,77819	X
3	59,5	59,0	42,3	0,78189	X
4	63,0	62,5	18,6	0,74857	X X
4	62,2	61,6	24,7	0,75696	X X
5	64,4	63,7	10,3	0,73560	X X
5	63,7	63,0	15,5	0,74301	X
6	65,1	64,3	7,1	0,72969	X X X
6	65,0	64,2	8,2	0,73128	X X
7	65,8	64,9	4,2	0,72401	X X X
7	65,4	64,4	7,2	0,72847	X X X
8	66,0	65,0	4,7	0,72327	X X X X
8	65,9	64,8	6,0	0,72514	X X X
9	66,1	64,9	6,3	0,72417	X X X X
9	66,1	64,8	6,4	0,72437	X X X X X
10	66,1	64,8	8,1	0,72536	X X X X X
10	66,1	64,7	8,1	0,72540	X X X X X
11	66,1	64,6	10,0	0,72664	X X X X X
11	66,1	64,6	10,1	0,72676	X X X X X
12	66,1	64,5	12,0	0,72809	X X X X X
12	66,1	64,5	12,0	0,72809	X X X X X
13	66,1	64,3	14,0	0,72955	X X X X X

Regression Analysis: ln(Tij) versus ln(dij); ln(Pi/1000); ...

The regression equation is

$$\ln(T_{ij}) = 12,5 - 1,34 \ln(d_{ij}) + 0,595 \ln(\text{Pi}/1000) + 0,242 \ln(s_j) \\ + 0,854 \ln(\text{Dummy}) + 0,206 \ln(m_i)$$

Predictor	Coef	SE Coef	T	P
Constant	12,4537	0,4898	25,43	0,000
ln(dij)	-1,34275	0,07981	-16,82	0,000
ln(Pi/1000)	0,59540	0,05133	11,60	0,000
ln(sj)	0,24249	0,05101	4,75	0,000
ln(Dummy)	0,8538	0,2681	3,18	0,002
ln(mi)	0,20647	0,04163	4,96	0,000

S = 0,735604 R-Sq = 64,4% R-Sq(adj) = 63,7%

ANEXO 11: REGRESSÃO GRUPO 1 COM A RETIRADA DE OUTLIERS

Regression Analysis: ln(Tij) versus ln(dij); ln(Pi/1000); ...

The regression equation is

$$\ln(T_{ij}) = 10,7 - 1,44 \ln(d_{ij}) + 0,271 \ln(P_i/1000) + 0,180 \ln(P_j/1000) + 0,106 \ln(m_i)$$

Predictor	Coef	SE Coef	T	P
Constant	10,6882	0,4063	26,30	0,000
ln(dij)	-1,44346	0,08685	-16,62	0,000
ln(Pi/1000)	0,27141	0,08166	3,32	0,001
ln(Pj/1000)	0,18033	0,06344	2,84	0,005
ln(mi)	0,10644	0,03064	3,47	0,001

S = 0,811371 R-Sq = 57,5% R-Sq(adj) = 56,8%

ANEXO 12: REGRESSÃO GRUPO 2 COM A RETIRADA DE OUTLIERS

Regression Analysis: ln(Tij) versus ln(dij); ln(Pi/1000); ...

The regression equation is

$$\ln(T_{ij}) = 11,1 - 1,07 \ln(d_{ij}) + 0,0553 \ln(P_i/1000) + 0,0473 \ln(M_{igi}) + 0,0890 \ln(s_j)$$

Predictor	Coef	SE Coef	T	P
Constant	11,1274	0,1731	64,29	0,000
ln(dij)	-1,06987	0,03373	-31,71	0,000
ln(Pi/1000)	0,05529	0,02750	2,01	0,045
ln(Migi)	0,04725	0,01722	2,74	0,007
ln(sj)	0,08904	0,02433	3,66	0,000

S = 0,307543 R-Sq = 82,3% R-Sq(adj) = 82,0%

ANEXO 13: REGRESSÃO GRUPO 3 COM A RETIRADA DE OUTLIERS

Regression Analysis: ln(Tij) versus ln(dij); ln(Pi/1000); ...

The regression equation is

$$\ln(T_{ij}) = 11,9 - 1,03 \ln(d_{ij}) + 0,0649 \ln(P_i/1000) + 0,0717 \ln(P_j/1000) + 0,0356 \ln(m_i) + 0,0659 \ln(m_j)$$

Predictor	Coef	SE Coef	T	P
Constant	11,8607	0,2411	49,20	0,000
ln(dij)	-1,03394	0,04251	-24,32	0,000
ln(Pi/1000)	0,06495	0,02602	2,50	0,013
ln(Pj/1000)	0,07173	0,02074	3,46	0,001
ln(mi)	0,03562	0,01707	2,09	0,038
ln(mj)	0,06590	0,01918	3,44	0,001

S = 0,300689 R-Sq = 82,0% R-Sq(adj) = 81,6%