

# CONSTRUÇÃO DA MATRIZ ORIGEM-DESTINO DE TRANSPORTE INTER-REGIONAL DE CARGAS E PASSAGEIROS PARA O PLANO NACIONAL DE LOGÍSTICA INTEGRADA

## Relatório de Pesquisa

Modelos de Regressão para Geração e Atração de Viagens de Passageiros  
(Relatório 7)





# **CONSTRUÇÃO DA MATRIZ ORIGEM-DESTINO DE TRANSPORTE INTER-REGIONAL DE CARGAS E PASSAGEIROS PARA O PLANO NACIONAL DE LOGÍSTICA INTEGRADA**

## **Relatório de Pesquisa**

**Modelos de Regressão para Geração e Atração de Viagens de Passageiros  
(Relatório 7)**

**ipea**

## **Governo Federal**

### **Ministério do Planejamento, Desenvolvimento e Gestão**

**Ministro interino** Dyogo Henrique de Oliveira

# **ipea** Instituto de Pesquisa Econômica Aplicada

Fundação pública vinculada ao Ministério do Planejamento, Desenvolvimento e Gestão, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiro – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

#### **Presidente**

Ernesto Lozardo

#### **Diretor de Desenvolvimento Institucional, Substituto**

Carlos Roberto Paiva da Silva

#### **Diretor de Estudos e Políticas do Estado, das Instituições e da Democracia**

João Alberto De Negri

#### **Diretor de Estudos e Políticas Macroeconômicas**

José Ronaldo de Castro Souza Júnior

#### **Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais**

Alexandre Xavier Ywata de Carvalho

#### **Diretora de Estudos e Políticas Setoriais de Inovação e Infraestrutura**

Fernanda De Negri

#### **Diretora de Estudos e Políticas Sociais**

Lenita Maria Turchi

#### **Diretor de Estudos e Relações Econômicas e Políticas Internacionais**

Sérgio Augusto de Abreu e Lima Florêncio Sobrinho

#### **Chefe de Gabinete**

Silvana Nascimento Neves

#### **Assessora-chefe de Imprensa e Comunicação**

Regina Alvarez

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

URL: <http://www.ipea.gov.br>

# **CONSTRUÇÃO DA MATRIZ ORIGEM-DESTINO DE TRANSPORTE INTER-REGIONAL DE CARGAS E PASSAGEIROS PARA O PLANO NACIONAL DE LOGÍSTICA INTEGRADA**

## **Relatório de Pesquisa**

**Modelos de Regressão para Geração e Atração de Viagens de Passageiros  
(Relatório 7)**

**ipea**

Brasília, 2017

## **EQUIPE DE PESQUISA**

### **Coordenação-geral**

Fabiano Mezadre Pompermayer, da Diretoria de Estudos e Políticas Setoriais de Inovação, Regulação e Infraestrutura (Diset/Ipea)

### **Equipe Técnica**

Fabiano Mezadre Pompermayer (Diset/Ipea)  
Erivelton Pires Guedes, Assessoria Técnica da Presidência (Astep) do Ipea  
Akina Sakamori, bolsista PNPD (Diset/Ipea)  
Alan Ricardo da Silva, bolsista PNPD (Diset/Ipea)  
Carolina Andrade Silva, bolsista PNPD (Diset/Ipea)  
Daniel Alisson Feitosa Lopes, bolsista PNPD (Diset/Ipea)  
Diego Rosa Mambrin, bolsista PNPD (Diset/Ipea)  
Gabriel Gouveia Rabello, bolsista PNPD (Diset/Ipea)  
João Gabriel de Moraes Souza, bolsista PNPD (Diset/Ipea)  
Maircon Batista Ribeiro, bolsista PNPD (Diset/Ipea)  
Paulo Henrique Dourado da Silva, bolsista PNPD (Diset/Ipea)  
Pedro Veiga de Camargo, bolsista PNPD (Diset/Ipea)  
Priscila Nascimento de Alcântara Garcia, bolsista PNPD(Diset/Ipea)  
Rafaella Bandeira Cabral Cunha, bolsista PNPD (Diset/Ipea)  
Ramon de Almeida Bispo, bolsista PNPD (Diset/Ipea)  
Raquel Araujo de Almeida, bolsista PNPD (Diset/Ipea)  
Rennaly Patricio Sousa, bolsista PNPD (Diset/Ipea)  
Roberto Lazarte Kaqui, bolsista PNPD (Diset/Ipea)  
Thiago Guimarães Rodrigues, bolsista PNPD (Diset/Ipea)

### **Equipe de elaboração do relatório**

Akina Sakamori, bolsista PNPD (Diset/Ipea)  
Raquel Araujo de Almeida, bolsista PNPD (Diset/Ipea)  
Priscila Nascimento de Alcântara Garcia, bolsista PNPD (Diset/Ipea)  
João Gabriel de Moraes Souza, bolsista PNPD (Diset/Ipea)  
Alan Ricardo da Silva, bolsista PNPD (Diset/Ipea)  
Rafaella Bandeira Cabral Cunha, bolsista PNPD (Diset/Ipea)  
Fabiano Mezadre Pompermayer (Diset/Ipea)

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério do Planejamento, Desenvolvimento e Gestão.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

## SUMÁRIO

APRESENTAÇÃO .....	7
1 INTRODUÇÃO .....	9
2 VARIÁVEIS EXPLICATIVAS TESTADAS.....	10
3 REGRESSÃO LINEAR MÚLTIPLA .....	14
4 ANÁLISE DAS REGRESSÕES .....	16
REFERÊNCIAS .....	20
APÊNDICE.....	21



## APRESENTAÇÃO

Este relatório faz parte de uma série de relatórios de pesquisa do Ipea, cujo objetivo é documentar o processo de construção das matrizes origem/destino (O/D) de transporte inter-regional de cargas e passageiros que subsidiaram a elaboração do Plano Nacional de Logística Integrada (PNLI), fruto da parceria entre a Empresa de Planejamento e Logística (EPL), empresa pública do governo federal, e o Ipea, formalizada pelo Termo de Cooperação para Descentralização de Crédito no 2, de 29 de maio de 2013.

A EPL foi criada em 2012 para apoiar o planejamento da infraestrutura de transportes do país. Uma de suas tarefas é elaborar o Plano Nacional de Logística Integrada. Em certa medida, ela cumpre missão semelhante à do Grupo de Estudos para Integração da Política de Transportes (Geipot) nas décadas de 1970 e 1980. A empresa vem se capacitando para a modelagem da oferta de infraestrutura de transportes, mas para a modelagem da demanda por transportes, que envolve mais conhecimento sobre as atividades econômicas no país e sua distribuição regional do que sobre transportes, havia a necessidade de execução externa. Algumas consultorias foram abordadas, mas surgiu a possibilidade de o estudo ser realizado pelo Ipea, que prontamente atendeu.

Nesse processo, um dos passos iniciais é o levantamento do padrão de viagens inter-regionais de cargas e passageiros, identificando os potenciais de geração e atração de viagens de cada localidade, bem como sua distribuição entre elas. Os modos de transporte usados também são levantados. Assim se obtém a matriz origem/destino das viagens de carga e passageiro para um ano-base, em cada modo. A partir daí, são projetados os crescimentos de demanda por transporte para os anos futuros, isto é, as matrizes O/D futuras, que são confrontadas com a oferta de infraestrutura a fim de identificar seus gargalos e projetar a necessidade de melhorias, como construção ou ampliação de rodovias, ferrovias, hidrovias e portos. A construção da matriz O/D é essencial para o planejamento da infraestrutura de transportes do país, uma vez que permite a construção de cenários e a otimização dos investimentos.

A elaboração de uma matriz O/D envolve o levantamento e cruzamento de uma série de dados socioeconômicos, que são ajustados e calibrados com pesquisas de campo, em que viajantes são entrevistados em locais predeterminados para identificar diversos atributos de sua viagem, como origem, destino, motivo, tipo e valor da carga, modo(s) de transporte utilizado(s) etc. Este trabalho envolve conhecimento técnico específico em modelagem de transportes e também em economia e estatística.

Com uma matriz O/D consolidada e abrangente, a EPL poderá realizar suas avaliações de necessidade de expansão da oferta de infraestrutura de transporte. O Ipea, além de apoiar a EPL nestas avaliações, beneficia-se das informações da matriz O/D ao cruzá-las com outras fontes de dados socioeconômicos, podendo realizar estudos diversos sobre o desenvolvimento regional, matrizes insumo/produto e fomento produtivo, por exemplo. Adicionalmente, a disponibilização dessa base de dados à sociedade permite o desenvolvimento do próprio setor de transporte e logística.

Oito relatórios descrevem as metodologias usadas em todo o processo de construção da base de dados e suas projeções. O último desses relatórios faz a agregação dos procedimentos intermediários para a construção das matrizes e apresenta alguns dos resultados agregados.

Há, ainda, três relatórios que avaliaram o estado da arte de modelos de construção da matriz O/D, com proposições para as novas versões do PNLI. A seguir, a lista com os títulos de cada relatório:

1. Desenho da pesquisa origem/destino do transporte rodoviária no Brasil.
2. Construção da matriz origem/destino observada para transporte de cargas inter-regional.
3. Modelos de regressão para geração e atração de viagens do transporte de cargas inter-regional.
4. Modelos de distribuição para matriz origem/destino de transporte de cargas inter-regional: desenvolvimento de um conjunto de ferramentas e calibração inicial.
5. Cenários de projeção das atividades econômicas por Unidade da Federação (UF) para a projeção da matriz origem/destino de transporte de cargas inter-regional.
6. Construção da matriz origem/destino observada de transporte de passageiros inter-regional.
7. Modelos de regressão para geração e atração de viagens de passageiros.
8. Projeções das matrizes O/D de carga e passageiros: aplicando os modelos de geração e distribuição de viagens com as projeções socioeconômicas.
9. Estado da arte em métodos de construção de matrizes origem/destino para o transporte de cargas inter-regional.
10. Estado da arte de métodos de estimação de matrizes origem-destino para passageiros à longa distância.
11. Métodos de estimação da matriz origem-destino para o transporte aéreo de passageiros.

## 1 INTRODUÇÃO

Baseado nos procedimentos descritos no relatório anterior, referente ao tratamento dos dados da pesquisa de campo e montagem da matriz origem/destino observada de passageiros, este relatório apresenta a aplicação do método estatístico utilizado para correlacionar o número de viagens com variáveis socioeconômicas. Esse processo é a primeira etapa do modelo clássico de quatro etapas para planejamento de transportes: *i)* geração; *ii)* distribuição; *iii)* repartição modal; e *iv)* alocação. Seu objetivo é identificar as variáveis socioeconômicas que possam explicar a quantidade de viagens geradas e atraídas para cada zona de tráfego. Entendendo essas relações, o planejador pode estimar a quantidade de viagens futuras geradas e atraídas em cada zona, com base em projeções das variáveis socioeconômicas, em geral mais fáceis de serem previstas.

Para analisar as variáveis que podem estar relacionadas com a atração e geração de viagens, é apresentado neste relatório a técnica estatística escolhida, regressão linear múltipla. Esta análise é feita com intuito de determinar as variáveis explicativas que são significativas para compor o modelo de geração e atração de viagens, que servirá de base para as projeções das matrizes origem/destino futuras. São apresentadas, inicialmente, as variáveis a serem testadas no modelo. Estas se referem às variáveis demográficas, variáveis indicadoras das referidas zonas, variáveis referentes ao turismo, variáveis *per capita* e variáveis auxiliares criadas pelo Ipea. Em seguida, há uma explicação da técnica estatística utilizada e suas respectivas definições e propriedades. Por fim, apresentam-se os resultados obtidos dos modelos de ônibus e dos modelos finais de geração e atração de viagens de passageiros. Os conceitos mencionados anteriormente serão retratados com mais detalhes no decorrer deste relatório.

## 2 VARIÁVEIS EXPLICATIVAS TESTADAS

Nesta seção, são abordadas as variáveis utilizadas para construção do modelo de fluxo de passageiros. Uma das motivações para o uso de covariáveis no experimento é ganhar precisão pela redução da variância do erro. Com o objetivo de testar a importância de variáveis na geração e atração de viagens, foram testadas as covariáveis a seguir. Para a escolha das variáveis explicativas propostas, considerou-se, também, a factibilidade de previsão do seu estado no futuro, num horizonte de até vinte anos. Por exemplo, para as variáveis relacionadas à população espera-se ser possível estimar as taxas de crescimento por zona ou grupo de zonas (estados por exemplo), com algum grau de precisão, para estimar seus valores futuros e assim usá-los para estimar a quantidade de viagens realizadas, que é mais difícil de prever diretamente. Mesmo para as variáveis binárias, como as que indicam alguma característica geográfica de uma zona, é possível estimar seus estados futuros, de acordo com os cenários propostos. Todas as variáveis foram desagregadas por zona de tráfego, que em geral seguem as microrregiões do Instituto Brasileiro de Geografia e Estatística (IBGE), para as localizadas no Brasil.

### 2.1 Variáveis demográficas

Com objetivo de mensurar a geração e atração de viagens de passageiros, foi necessário a determinação de variáveis demográficas que explicam o fluxo de passageiros no território brasileiro. Essas variáveis foram extraídas do *site* do IBGE referentes ao Censo 2010 e do IBGE-Cidades. A tabela 1 apresenta as variáveis relacionadas à população.

Na variável *pop2012*, a população das zonas relativas ao exterior foi obtida no endereço eletrônico *indexmundi* com fonte da *CIA World Factbook*.

TABELA 1  
Relação das variáveis referentes à população

Variável	Descrição
pop2010_homen	Nº de homens
pop2010_mulher	Nº de mulheres
pop2010alfab	Nº de alfabetos
pop2010freq_ei	População residente que frequentava creche ou escola
pop2010catolromano	Nº de católicos
pop2010espitita	Nº de espíritas
pop2010evang	Nº de evangélicos
pesassalariado2011	População assalariada 2011
pesocupado_tt2011	População ocupada formal 2011
pop2012	População 2012

Elaboração: Ipea.

A tabela 2 apresenta os dados que fazem referência à renda e ao produto interno bruto (PIB). Tais dados foram retirados do *site* do IBGE e da Relação Anual de Informações Sociais (Rais), oriunda do Ministério do Trabalho e Emprego (MTE).

Na variável *pib2010corrente*, o PIB das zonas do exterior foi extraído do *Ipeadata*,<sup>1</sup> em dólares e transformados em real (R\$) pela taxa de câmbio anual de 2010. A taxa de câmbio utilizada também foi fornecida pelo *Ipeadata*.

TABELA 2  
Relação das variáveis referentes à produção e renda

Variável	Descrição
rend2010urbana	Renda total urbana (R\$)
rend2010rural	Renda total rural (R\$)
agro2010	PIB agropecuária (R\$)
ind2010	PIB indústria (R\$)
serv2010	PIB serviços (R\$)
pib2010corrente	PIB corrente (R\$)
pibpccor2010	PIB corrente <i>per capita</i> (R\$)

Elaboração: Ipea.

A tabela 3 apresenta as variáveis demográficas oriundas do IBGE, em especial no domínio eletrônico IBGE-Cidades e do *site* IBGE-Censo. As variáveis apresentadas nessa tabela representam o número de empresas por zona, Índice de Desenvolvimento Humano (IDH) e tamanho da área geográfica de cada zona.

TABELA 3  
Relação das variáveis demográficas testadas no modelo

Variável	Descrição
nempresa2011	Nº de empresas (matriz) 2011
nunidlocal2011	Nº de empresas (matriz e filial) 2011
impostos2010	Arrecadação de impostos
area2010	Área geográfica ( <i>km</i> <sup>2</sup> )
idhm2010	IDH – Municipal <sup>1</sup>

Elaboração: Ipea.

Nota: 1 Média simples dos indicadores de cada município da zona.

A tabela 4 apresenta as variáveis referentes à remuneração da população. Esses dados foram retirados da Rais, oriunda do portal do Ministério do Trabalho e Emprego.

1. Domínio eletrônico de dados compactados e gerados pelo Ipea.

TABELA 4  
Relação das variáveis de salário

Variável	Descrição
salarios2011	Salários 2011
nmedio_sm2011	Salário médio mensal (em SM) <i>per capita</i>

Elaboração: Ipea.

Observa-se na tabela 4 que a variável *nmedio\_sm2011* é mensurada em quantidade de salários mínimos de 2011.

A tabela 5 refere-se aos estabelecimentos de saúde por município, extraídas do *site* do Ministério da Saúde (MS) referente ao Cadastro Nacional de Estabelecimentos de Saúde (CNES).

TABELA 5  
Relação das variáveis de saúde

Variável	Descrição
estabSUS2009	Número de estabelecimentos de saúde
media_leito12_sum	Leitos hospitalares (média dos 12 meses)
media_clinica12_sum	Clínicas (média dos 12 meses)

Elaboração: Ipea.

Foram testadas também, variáveis referentes ao número de veículos: automóvel, ônibus, caminhão, motocicleta, entre outros. A tabela 6 mostra as variáveis analisadas. Tais variáveis foram coletadas do Departamento Nacional Trânsito (Denatran).

TABELA 6  
Relação das variáveis referentes ao número de veículos

Variável	Descrição
automovel	Número de veículos automóvel registrados
caminhao	Número de veículos caminhão registrados
caminhaotrator	Número de veículos trator registrados
caminhonete	Número de veículos caminhonete registrados
camioneta	Número de veículos caminhoneta registrados
micro_onibus	Número de veículos micro-ônibus registrados
motocicleta	Número de veículos motocicleta registrados
motoneta	Número de veículos motoneta registrados
ônibus	Número de veículos ônibus registrados
veiculos	Soma de veículos

Elaboração: Ipea.

Analisaram-se também variáveis relativas à educação, mais especificamente, variáveis relativas ao número de alunos matriculados nos ensinos fundamental, médio e superior. Essas variáveis foram extraídas do Censo Educacional do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). A tabela 7 expõe tais variáveis.

TABELA 7  
Relação das variáveis de educação

Variável	Descrição
mat2012_fund	Matrículas ensino fundamental
mat2012_em	Matrículas ensino médio
matric_sup12_pres_sum	Matrículas ensino superior presencial
mat_sup_percapita	Matrículas ensino superior <i>per capita</i>

Elaboração: Ipea.

## 2.2 Variáveis de atividade comercial e de turismo

A variável *massa salarial* (MSAL) trata-se de uma variável obtida da Rais. Neste estudo, tal variável foi utilizada com o intuito de estabelecer a variação do valor dos salários desagregados de alojamento e de alimentação. A tabela 8 demonstra essas variáveis.

A variável *proxi\_ativ\_turismo*, uma *proxy* para regiões de turismo, corresponde ao número de quartos de hotéis e pousadas por zona. Essa variável foi retirada do sitio IBGE-Cidades.

TABELA 8  
Relação das variáveis MSAL

Variável	Descrição
msal_cnae_55	Massa salarial de alojamentos
msal_cnae_56	Massa salarial de alimentação
msal_cnae_55_56	Massa salarial de alojamentos e alimentação
proxi_ativ_turismo	Proxy de atividade de turismo

Elaboração: Ipea.

## 2.3 Variáveis per capita

Pelo fato de algumas variáveis explicativas apresentarem alta correlação com a variável de população (*pop2012*), testou-se as relações das variáveis de forma *per capita*. Por exemplo, a variável *automóvel per capita* é igual a variável *automóvel* dividida pela variável *pop2012*. Algumas variáveis *per capita* já foram apresentadas anteriormente, como o PIB *per capita*. A relação das novas variáveis *per capita* testadas no modelo segue na tabela 9.

TABELA 9  
Relação das variáveis per capita testadas no modelo

Variável	Descrição
densidade_demografica	Densidade demográfica
area_pcap	Área <i>per capita</i>
msal55_pcap	Massa salarial de alojamentos <i>per capita</i>
msal56_pcap	Massa salarial de alimentação <i>per capita</i>
msal55_56_pcap	Massa salarial de alojamentos e alimentação <i>per capita</i>
onibus_pcap	Ônibus <i>per capita</i>
automovel_pcap	Automóvel <i>per capita</i>
veic_pcap	Veículo <i>per capita</i>

Elaboração: Ipea.

## 2.4 Variáveis espaciais

Com intuito de tornar o modelo mais robusto, criaram-se algumas variáveis espaciais que pudessem também explicar a geração e atração de viagens de passageiros. A tabela 10 mostra tais variáveis e apresenta uma breve descrição sobre elas. A variável *limitrofe* corresponde a uma variável indicadora que determina se a zona testada está na fronteira de algum estado. Sua inclusão foi proposta devido ao número de viagens de ônibus interestaduais – base da Agência Nacional de Transportes Terrestres (ANTT) – ser maior envolvendo as zonas próximas às fronteiras estaduais. Mas boa parte dessas viagens é de curta distância, parecidas com as viagens intraestaduais (bases obtidas junto aos estados), para as quais a base cobriu apenas cinco estados. A variável *nao\_amazonia* corresponde a uma variável dicotômica que indica se a zona não está geograficamente na região amazônica. Essa variável foi criada com

intuito de isolar as áreas que possuem predominantemente transporte de passageiros via fluvial. Como não foi possível a obtenção dos dados de transporte fluvial de passageiros, tal variável serviu para indicar que estas zonas tinham o número de viagens observadas sub-representado. Outra variável testada foi a *Dist\_capital\_mais\_proxima*, a qual corresponde à distância da zona à capital de estado mais próxima.

TABELA 10  
Relação das variáveis espaciais

Variável	Descrição
Limitrofe	Zona está na fronteira do estado
nao_amazonia	Variável dicotômica que representa se a zona não está na Amazônia
Dist_capital_mais_proxima	Distância entre a zona e a capital de Unidade da Federação mais próxima

Elaboração: Ipea.

Além disso, assim como nos modelos de geração de viagens de carga, foram usadas variáveis indicadores do nível de concentração econômica da zona. As variáveis indicadoras foram criadas com base na classificação Regic (regiões de influência das cidades) que tem como objetivo conhecer os relacionamentos entre as cidades brasileiras, baseada na análise dos fluxos de bens e serviços (IBGE, 2007). Através das Regic, buscou-se definir a hierarquia dos centros urbanos e delimitar as regiões de influência a eles associadas. As zonas foram classificadas de 1 a 11, de acordo com o seu município polo (município economicamente mais importante), sendo o nível 1 equivalente ao mais importante nível hierárquico do Regic, corresponde à cidade de São Paulo. Foi atribuído o nível zero para zonas do exterior. Espera-se, com isso, identificar efeitos de *hubs* de viagens, típicos de aeroportos em que os passageiros fazem conexões, cuja lógica é extensível ao transporte por ônibus e o fluvial. Para utilização das covariáveis do estudo, foram agregados os níveis conforme apresentam-se na tabela 11. Como pode-se observar, a variável *Flag\_nivel\_X* representa as zonas do exterior.

TABELA 11  
Relação das variáveis indicadoras testadas no modelo

Variável	Descrição
flag_nivel_0	Indica se a zona é nível 0
flag_nivel_1	Indica se a zona é nível 1
flag_nivel_2	Indica se a zona é nível 2
flag_nivel_3	Indica se a zona é nível 3
flag_nivel_4	Indica se a zona é nível 4
flag_nivel_5	Indica se a zona é nível 5
flag_nivel_6	Indica se a zona é nível 6
flag_nivel_7	Indica se a zona é nível 7
flag_nivel_8	Indica se a zona é nível 8
flag_nivel_9	Indica se a zona é nível 9
flag_nivel_10	Indica se a zona é nível 10
flag_nivel_11	Indica se a zona é nível 11
grupo_nivel	Grupo de nível (A = 1, 2, 3, 4; B = 5, 6; C = 7,8; D = 9,10,11)
flag_nivel_A	Indica se a zona está nos níveis 1, 2, 3 e 4
flag_nivel_B	Indica se a zona está nos níveis 5, 6
flag_nivel_C	Indica se a zona está nos níveis 7, 8
flag_nivel_D	Indica se a zona está nos níveis 9, 10, 11
flag_nivel_X	Indica se a zona está no nível 0

Elaboração: Ipea.

### 3 REGRESSÃO LINEAR MÚLTIPLA

A técnica utilizada para a identificação das variáveis que influenciam na geração e atração de viagens foi a análise de regressão linear múltipla. A seguir são apresentadas algumas especificações dessa técnica.

#### 3.1 Definições

O objetivo de se ajustar um modelo de regressão linear múltipla é prever e explicar uma variável resposta por meio de variáveis explicativas. Neste estudo, essa técnica foi utilizada com o intuito de modelar e investigar a existência de uma relação linear entre a variável dependente, quantidade de passageiros transportados de uma zona de origem até uma zona de destino, e as variáveis independentes, apresentadas no item anterior, seguindo a equação geral:

$$T_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni}, \quad (1)$$

Onde:

$T_i$  é o número de viagens geradas (ou atraídas) para a zona  $i$ ,

$X_{1i}, X_{2i}, \dots, X_{ni}$ , são os valores das variáveis explicativas listadas na seção anterior, para cada zona  $i$ , e

$\alpha, \beta_1, \beta_2, \dots, \beta_n$ , são os parâmetros a serem estimados.

Vale lembrar que assumiu-se que o número de viagens de passageiros geradas em cada zona é igual ao número de viagens atraídas (viagens de ida e volta). Assim, as regressões são feitas apenas para as viagens geradas.

#### 3.2 Covariáveis (variáveis independentes)

##### 1) Variáveis contínuas

As variáveis contínuas testadas foram: população, produção e renda, salário, saúde, número de veículos e educação. Também foram utilizadas variáveis relacionadas à atividade de turismo e algumas variáveis espaciais, apresentadas na seção 2.4.

##### 2) Variáveis *dummy*

Na análise de regressão, a variável dependente é influenciada não só por variáveis regressoras quantitativas, mas também por variáveis qualitativas ou nominais, as quais indicam uma característica que não pode ser medida em uma escala numérica, mas sim categórica. Esse tipo de variável indica, na maioria das vezes, presença ou ausência de uma característica. Para “quantificar” essas características, são frequentemente utilizadas variáveis artificiais, denominadas *dummy* ou ainda chamadas por alguns autores de indicadores, mudas, binárias, *flags* ou dicotômicas. Esse tipo de variável assume valores 1 ou 0, indicando presença ou ausência de determinada característica, respectivamente. Neste estudo, foram criadas algumas variáveis *dummy* com o intuito de se testar as variáveis que pudessem influenciar na variável resposta.

### 3) Interações

Em algumas situações na regressão, há o interesse de testar se a interação entre algumas variáveis é estatisticamente significativa para explicar o comportamento da variável dependente, além de testá-las separadamente. Nesta pesquisa foram testadas algumas interações, apresentadas no apêndice. As variáveis espaciais apresentadas na seção 2.4 foram interagidas com as variáveis de população, massa salarial da classe CNAE 56, e PIB a preços correntes

### 3.3 Método de seleção do modelo: *backward*

Em modelos de regressão múltipla, é necessário determinar um subconjunto de variáveis independentes que melhor explique a variável resposta, isto é, dentre todas as variáveis explicativas disponíveis, deve-se encontrar um subconjunto de variáveis importantes para o modelo.

Existem duas principais formas de se realizar a seleção das variáveis: uma em que o pesquisador testa todas as variáveis explicativas possíveis, ao mesmo tempo, no modelo e outra em que é feita uma seleção automática, em que o método escolhido busca o melhor subconjunto de variáveis explicativas para conter no modelo. Existem três principais métodos de seleção automática de variáveis: *forward*, *backward* e *stepwise*. O método de seleção de variáveis utilizado neste estudo foi o *backward*.

O procedimento *backward* caracteriza-se por incorporar, inicialmente, todas as variáveis independentes em um modelo de regressão múltipla e percorrer etapas nas quais uma variável por vez pode ser eliminada, sendo a ordem de eliminação definida pelo índice de importância, medida pelo nível de significância da variável. O nível de significância considerado foi de 5% e, conseqüentemente, as variáveis foram escolhidas com 95% de confiança.

### 3.4 Critérios de avaliação e validação do modelo

Após a seleção das variáveis pelo método *backward* foi realizada uma análise dos modelos selecionados, com relação à qualidade destes a partir de algumas medidas, citadas a seguir.

#### 1) Coeficiente de determinação ajustado

O coeficiente de determinação múltiplo ajustado,  $R^2_{Ajust}$ , mede a qualidade de ajustamento da equação de regressão. Isto é, mostra o percentual da variação total da variável dependente  $Y$  que é explicada conjuntamente pelas variáveis explanatórias.

O  $R^2_{Ajust}$  tem variação entre *zero* e *um* e, como o interesse é estimar valores da variável  $Y$  a partir de valores das variáveis explicativas, a regressão será tanto mais útil quanto mais próximo de *um* estiver o valor do coeficiente de determinação ajustado.

Com o intuito de validação, foram analisados os modelos de regressão selecionados pelo método *backward* e comparados a outros modelos, ao se retirar ou adicionar variáveis independentes, observando a diferença causada no  $R^2_{Ajust}$ , levando em consideração também a análise de multicolinearidade.

#### 2) Correlação e multicolinearidade

Para se obter um bom modelo linear múltiplo, há também a condição de que não exista relação linear forte entre qualquer uma das variáveis independentes. A presença de correlação

entre duas variáveis independentes é chamada de colinearidade, e correlação entre mais de duas variáveis independentes é denominada multicolinearidade.

Essa multicolinearidade pode causar sérios efeitos nas estimativas dos coeficientes e na aplicabilidade em geral do modelo, tais como: estimativas com erros muito grandes, eliminação de variáveis importantes no teste de hipóteses, alteração do valor da estimativa obtida, entre outros.

A verificação da presença de multicolinearidade foi realizada ao se analisar a correlação entre as variáveis independentes selecionadas para o modelo, por meio do cálculo do coeficiente de correlação ( $R$ ). Ao se detectar correlação entre essas variáveis, os modelos escolhidos foram reanalisados após a retirada de uma dessas variáveis correlacionadas, e foi observado o efeito desta ação sob o valor do coeficiente de determinação ajustado ( $R^2_{Ajust}$ ). A regra de decisão para se afirmar presença de correlação forte entre variáveis foi um valor acima de 0,4 para o coeficiente de correlação.

## 4 ANÁLISE DAS REGRESSÕES

Conforme explicitado na seção anterior, para determinação do fluxo de passageiros em território nacional foram utilizadas diversas variáveis demográficas (por exemplo, variáveis de população, de saúde e de *IDH*), variáveis de estimativas de turismo (para tentar captar as viagens originadas pelas condições de turismo de cada zona) e variáveis que representam renda e produção. Todas essas variáveis foram medidas por zona.

Foram testados modelos aqui chamados de básicos, por considerarem apenas as variáveis referentes à população. Essas regressões já apresentaram valores altos de  $R^2_{Ajust}$ , porém com objetivo de trazer maior robustez as análises, foram testadas as outras variáveis já apresentadas neste relatório. Tanto para os modelos para estimação de passageiros de ônibus, como para o de passageiros final, as variáveis testadas foram as mesmas. A tabela A.2, no apêndice, apresenta um dicionário com os nomes das variáveis utilizadas nas regressões.

### 4.1 Modelo de regressão para passageiros de ônibus

Como descrito no relatório anterior, item 4.1, as estimativas de passageiros de ônibus foram construídas a partir de viagens interestaduais e intermunicipais (intraestaduais). As viagens intermunicipais não estavam completas para todas as UFs do Brasil, portanto, foi realizado um modelo de regressão para estimar essas viagens faltantes. Assim, foram testadas duas formas para construção deste modelo: estimando o total de viagens (interestadual e intermunicipal) e estimando apenas as viagens intermunicipais, dado que as viagens interestaduais estavam completas. Além disso, foram testados modelos por grupos de nível Regic do IBGE (A, B, C e D) e também no geral (total). Os resultados de cada tipo de modelo foram comparados e assim foi escolhido aquele modelo que obteve as quantidades de viagens mais coerentes com os valores observados.

Para os níveis Regic 1, 2, 3, 4, 5, 6, 7, 8 e 11, foi escolhido o modelo geral (tabela 12) cuja resposta é o total de viagens (interestadual e intermunicipal), e para os níveis Regic 9 e 10, foi escolhido o modelo por grupo de nível (tabela 13) cuja resposta é apenas o número de viagens intermunicipais. No último modelo, após a estimação de viagens intermunicipais, estas foram somadas às viagens interestaduais já observadas para resultar o total de passageiros de ônibus. Seguem os resultados dos modelos.

TABELA 12  
Modelo geral, escolhido para previsão de viagens para as zonas nos níveis Regic 1, 2, 3, 4, 5, 6, 7, 8 e 11

Variáveis	Parâmetros estimados	Pr >  t
pop2012	1.777	<.0001
nivel4_pop2012	0.958	<.0001
nivel_6_msal_cnae_56	0.001	0.0018
nivel_8_msal_cnae_56	0.005	0.0196
nivel_9_msal_cnae_56	0.007	<.0001
veic_pcap	659868	<.0001

$R^2_{Ajust} = 0.9399$   
Elaboração dos autores.

TABELA 13  
Modelo por nível, escolhido para previsão de viagens para as zonas nos níveis Regic 9 e 10

Variáveis	Parâmetros estimados	Pr >  t
flag_nivel_11	448474	<.0001
pop2012	0.72537	<.0001
nivel_9_msal_cnae_56	0.00993	<.0001
nivel_10_msal_cnae_56	0.00542	0.0024

$R^2_{Ajust} = 0.8363$   
Elaboração dos autores.

Observa-se que algumas variáveis parecem não fazer sentido nos modelos escolhidos. Por exemplo, na tabela 12, se observa a variável *nivel\_9\_msal\_cnae\_56*, mas o modelo não será aplicado nas zonas de nível 9. Na tabela 13, há a variável *flag\_nivel\_11*, mas o modelo será aplicado apenas nas zonas de nível 9 e 10. Isso ocorreu pois o modelo da tabela 12 foi estimado com as observações de todas as zonas, de todos os níveis. Assim, a variável referente ao nível 9 foi significativa neste modelo. Mas, na comparação com o modelo estimado para o grupo de zonas de nível 9 a 11, o modelo geral criou valores preditos piores para as zonas de nível 9 e 10. Na prática, essas variáveis não terão efeito nas previsões das respectivas zonas, pois seus valores serão nulos.

## 4.2 Modelo de regressão para passageiros final

Com as viagens de ônibus intermunicipal geradas por cada zona dos estados onde não foi obtido o número de viagens observados, foi possível obter valores agregados de viagens geradas por zona, considerando todos os modos. A exceção são as viagens pelo modo aquaviário, que podem ser substanciais na Amazônia. De certa forma, a estimação de viagens de ônibus da seção anterior indiretamente as estimam, pois o transporte fluvial pode ser considerado substituto ao ônibus na região, além disso, foram obtidos dados de viagens observadas de ônibus intraestaduais para esses estados.

Para o modelo de passageiros final, foram testados modelos por grupo de nível Regic (A, B, C e D) e no geral, e, a partir dos resultados dessas estimativas, foi selecionado o modelo com resultado que mais se adequou para as zonas de tráfego. Como há poucas observações nos grupos A e B, também foi testado o grupo nível AB, que contém as observações tanto do grupo A quanto do B. A comparação dos resultados foi realizada por nível Regic e pela UF. Os resultados foram: para as zonas dos níveis Regic 3 e 4 foi selecionado o modelo por grupo de nível A (tabela 14); para zonas no nível Regic 5 e que não estão no estado de MG e zonas no nível Regic 6, o modelo escolhido foi o por grupo de nível B (tabela 15);

para zonas do nível Regic 5 e que estão no estado de MG, o modelo escolhido foi o por grupo de nível AB (tabela 16); para zonas do nível Regic 7 e 8, o modelo selecionado foi o modelo por grupo de nível C (tabela 17); para zonas do nível Regic 9 e que não estão no estado do PR, zonas do nível Regic 10 e zonas do nível Regic 11 e que não estão no estado de MG, PR, RJ e RS, o modelo selecionado foi o por grupo de nível D (tabela 18). Já para as demais zonas, ou seja, zonas do nível Regic 1 e 2, zonas do nível Regic 9 e que estão no estado do PR e zonas do nível Regic 11 e que estão nos estados de MG, PR, RJ, RS, o modelo escolhido foi o modelo geral (tabela 19). O motivo para se escolher modelos distintos com base no nível Regic funciona em parte como *proxy* da carência de infraestrutura de transportes. Em geral, quanto maior o nível Regic (ou menor e mais distante dos grandes centros for à microrregião) menor a geração de viagens por habitante. Considerar a UF foi necessário devido às diferenças na matriz observada de ônibus intraestaduais. Para as UFs para as quais existiam informações sobre viagens intraestaduais, os métodos de levantamento podiam ser distintos (ex.: pesquisa de campo *versus* registros administrativos). Além disso, poderia haver distinção com as UFs em que essa variável foi imputada.

Os modelos selecionados são apresentados a seguir.

TABELA 14  
**Modelo A para as zonas nos níveis Regic 3 e 4**

Variáveis	Parâmetros estimados	Pr >  t
flag_nivel_2	-15321818	0.0157
pop2012	6.43	<.0001
nivel_4_msai_cnae_56	0.01	0.001

$R^2_{Ajust} = 0.9426$   
Elaboração dos autores.

TABELA 15  
**Modelo B para as zonas no nível Regic 5 e que não estão no estado de MG e zonas no nível Regic 6**

Variáveis	Parâmetros estimados	Pr >  t
flag_nivel_5	-11982877	0.0014
flag_nivel_6	-13063508	0.0005
pop2012	6.33	<.0001
idhm2010	20241633	0.0004

$R^2_{Ajust} = 0.8365$   
Elaboração dos autores.

TABELA 16  
**Modelo AB para zonas do nível Regic 5 e estão no estado de MG**

Variáveis	Parâmetros estimados	Pr >  t
flag_nivel_5	-4338902	0.0028
flag_nivel_6	-6107444	<.0001
pop2012	4.70	<.0001
nmedio_sm2011	3440798	<.0001

$R^2_{Ajust} = 0.9202$   
Elaboração dos autores.

TABELA 17  
Modelo C para zonas do nível Regic 7 e 8

Variáveis	Parâmetros estimados	Pr >  t
pop2012	2.20	0.0009
proxi_ativ_turismo	1321	<.0001
matric_sup12_pres_sum	105	<.0001
veic_pcap	1134936	0.0332

$R^2_{Ajust} = 0.7585$   
Elaboração dos autores.

TABELA 18  
Modelo D para zonas do nível Regic 9 e que não estão no estado do PR, zonas do nível Regic 10 e zonas do nível Regic 11 e que não estão no estado de MG, PR, RJ e RS

Variáveis	Parâmetros estimados	Pr >  t
flag_nivel_11	311843	0.0005
pop2012	3.11	<.0001
automovel_pcap	3083799	<.0001
nivel_9_msal_cnae_56	0.02	<.0001
nivel_10_msal_cnae_56	0.01	0.0136

$R^2_{Ajust} = 0.7229$   
Elaboração dos autores.

TABELA 19  
Modelo Geral para zonas do nível Regic 1 e 2, zonas do nível Regic 9 e que estão no estado do PR e zonas do nível Regic 11 e que estão nos estados de MG, PR, RJ, RS

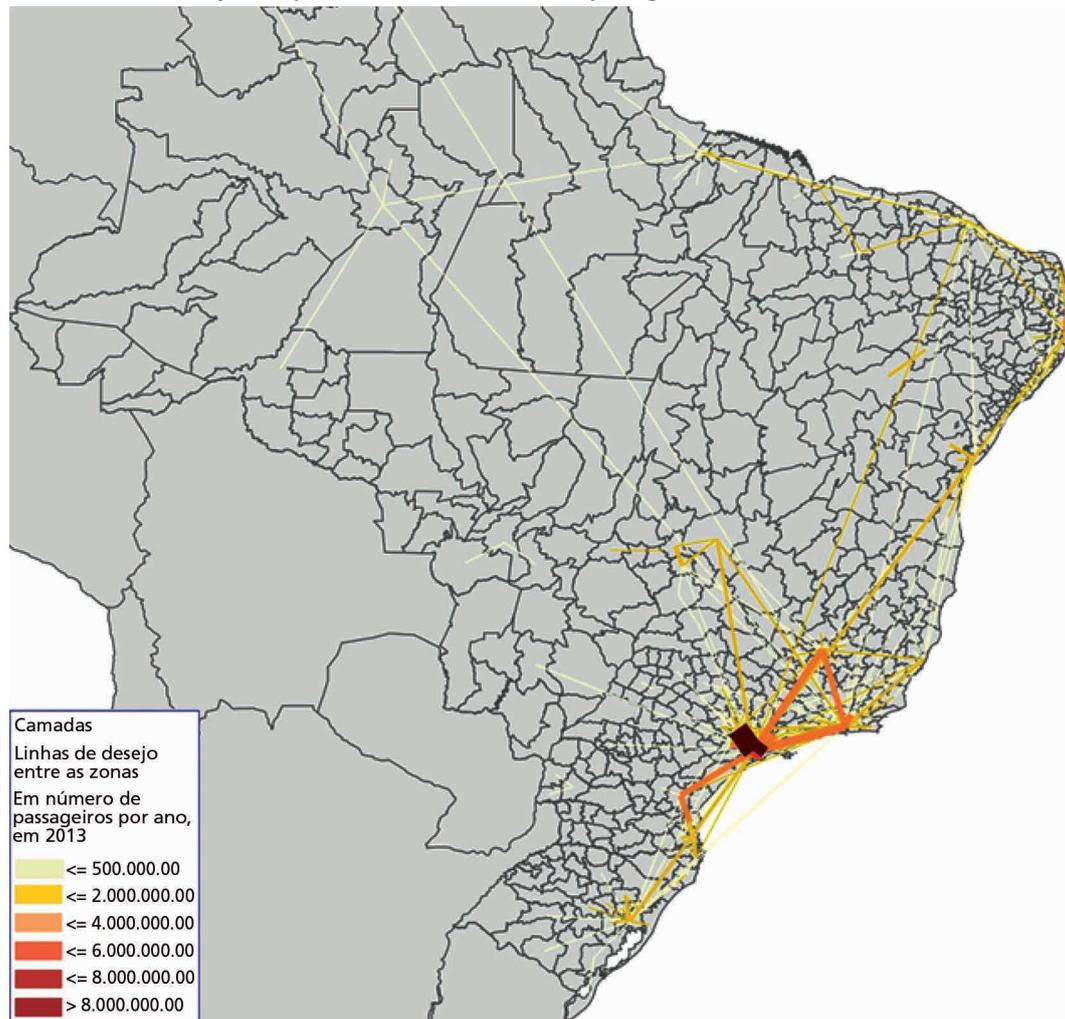
Variáveis	Parâmetros estimados	Pr >  t
flag_nivel_1	16887690	<.0001
flag_nivel_2	-5826259	0.0189
flag_nivel_3	5570900	<.0001
pop2012	5.03	<.0001
nmedio_sm2011	186239	0.0017
mat_sup_percapita	15788568	0.0001
nivel_4_pop2012	5.97	<.0001
nivel_5_pop2012	2.48	0.0005

$R^2_{Ajust} = 0.906$   
Elaboração dos autores.

Esses modelos foram usados para gerar os vetores de geração e atração de viagens no ano base. Com as projeções de população, renda e atividades econômicas, as variáveis explicativas foram revistas e os vetores de geração e atração de viagens futuros serão calculados. Em seguida, será aplicado o modelo gravitacional para distribuir as viagens, gerando as matrizes O/D de passageiros para o ano base e futuros. O mapa a seguir apresenta as linhas de desejo obtidas para os anos base (2013). Os fluxos apresentados seguem o esperado. Os maiores fluxos são proporcionais às maiores zonas (em população, renda etc.) e inversamente proporcionais à distância entre elas.

FIGURA 1

Linhas de desejo dos quinhentos maiores fluxos de passageiros – Brasil (2013)



Fonte: Matriz O/D Estimada de passageiros.

Elaboração: Empresa de Planejamento e Logística (EPL), dentro do Termo de Cooperação Ipea/EPL nº 2/2013, a partir das matrizes O/D estimadas pelo Ipea.

## REFERÊNCIAS

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Regiões de Influência das Cidades (Regic)**. 2007. Disponível em: <<https://goo.gl/5azceN>>. Acesso em: 8 dez. 2014.

## BIBLIOGRAFIA COMPLEMENTAR

GUJARATI, D. **Econometria básica**. 4. ed. Rio de Janeiro: Elsevier, 2006.

HOFFMANN, R. **Análise de regressão: uma introdução à econometria**. 4. ed. São Paulo: Hucitec, 2006.

NETER, J. *et al.* **Applied linear statistical models**. 5. ed. New York: McGraw-Hill, 2005.

## APÊNDICE

TABELA A.1  
Relação das variáveis de interação testadas no modelo

Variável	Descrição
nao_amaz_pop2012	Interação nao_amazonia e pop2012
limitrofe_pop2012	Interação limitrofe e pop2012
nivel_0_pop2012	Interação flag_nivel_0 e pop2012
nivel_1_pop2012	Interação flag_nivel_1 e pop2012
nivel_2_pop2012	Interação flag_nivel_2 e pop2012
nivel_3_pop2012	Interação flag_nivel_3 e pop2012
nivel_4_pop2012	Interação flag_nivel_4 e pop2012
nivel_5_pop2012	Interação flag_nivel_5 e pop2012
nivel_6_pop2012	Interação flag_nivel_6 e pop2012
nivel_7_pop2012	Interação flag_nivel_7 e pop2012
nivel_8_pop2012	Interação flag_nivel_8 e pop2012
nivel_9_pop2012	Interação flag_nivel_9 e pop2012
nivel_10_pop2012	Interação flag_nivel_10 e pop2012
nivel_11_pop2012	Interação flag_nivel_11 e pop2012
nivel_1_msal_cnae_56	Interação flag_nivel_1 e msal_cnae_56
nivel_2_msal_cnae_56	Interação flag_nivel_2 e msal_cnae_56
nivel_3_msal_cnae_56	Interação flag_nivel_3 e msal_cnae_56
nivel_4_msal_cnae_56	Interação flag_nivel_4 e msal_cnae_56
nivel_5_msal_cnae_56	Interação flag_nivel_5 e msal_cnae_56
nivel_6_msal_cnae_56	Interação flag_nivel_6 e msal_cnae_56
nivel_7_msal_cnae_56	Interação flag_nivel_7 e msal_cnae_56
nivel_8_msal_cnae_56	Interação flag_nivel_8 e msal_cnae_56
nivel_9_msal_cnae_56	Interação flag_nivel_9 e msal_cnae_56
nivel_10_msal_cnae_56	Interação flag_nivel_10 e msal_cnae_56
nivel_11_msal_cnae_56	Interação flag_nivel_11 e msal_cnae_56
nivel_0_pibcorrente	Interação flag_nivel_0 e pib2010corrente
nivel_1_pibcorrente	Interação flag_nivel_1 e pib2010corrente
nivel_2_pibcorrente	Interação flag_nivel_2 e pib2010corrente
nivel_3_pibcorrente	Interação flag_nivel_3 e pib2010corrente
nivel_4_pibcorrente	Interação flag_nivel_4 e pib2010corrente
nivel_5_pibcorrente	Interação flag_nivel_5 e pib2010corrente
nivel_6_pibcorrente	Interação flag_nivel_6 e pib2010corrente
nivel_7_pibcorrente	Interação flag_nivel_7 e pib2010corrente
nivel_8_pibcorrente	Interação flag_nivel_8 e pib2010corrente
nivel_9_pibcorrente	Interação flag_nivel_9 e pib2010corrente
nivel_10_pibcorrente	Interação flag_nivel_10 e pib2010corrente
nivel_11_pibcorrente	Interação flag_nivel_11 e pib2010corrente
nivel_A_pop2012	Interação flag_nivel_A e pop2012
nivel_B_pop2012	Interação flag_nivel_B e pop2012
nivel_C_pop2012	Interação flag_nivel_C e pop2012
nivel_D_pop2012	Interação flag_nivel_D e pop2012
nivel_X_pop2012	Interação flag_nivel_X e pop2012
nivel_ABCD_pop2012	Interação flag_nivel_ABCD e pop2012
nivel_A_msal_cnae_56	Interação flag_nivel_A e msal_cnae_56
nivel_B_msal_cnae_56	Interação flag_nivel_B e msal_cnae_56
nivel_C_msal_cnae_56	Interação flag_nivel_C e msal_cnae_56
nivel_D_msal_cnae_56	Interação flag_nivel_D e msal_cnae_56
nivel_X_msal_cnae_56	Interação flag_nivel_X e msal_cnae_56
nivel_ABCD_msal_cnae_56	Interação flag_nivel_ABCD e msal_cnae_56
nivel_A_pibcorrente	Interação flag_nivel_A e pib2010corrente
nivel_B_pibcorrente	Interação flag_nivel_B e pib2010corrente
nivel_C_pibcorrente	Interação flag_nivel_C e pib2010corrente
nivel_D_pibcorrente	Interação flag_nivel_D e pib2010corrente
nivel_X_pibcorrente	Interação flag_nivel_X e pib2010corrente
nivel_ABCD_pibcorrente	Interação flag_nivel_ABCD e pib2010corrente

Elaboração: Ipea.

TABELA A.2  
Dicionário das variáveis utilizadas nas regressões

Variável	Descrição
Pass_origem_sem_od_unico_TOTAL	Número de passageiros total (sem par OD único da pesquisa rodoviária) origem
pop2012	População 2012
nempresa2011	Nº de empresas (matriz) 2011
nunidlocal2011	Nº de empresas (matriz e filial) 2011
pesassalariado2011	População assalariada 2011
pesocupado_tt2011	População ocupada formal 2011
salarios2011	Salários 2011
area2010	Área geográfica
estabSUS2009	Número de estabelecimentos de saúde
mat2012_fund	Matrículas ensino fundamental
mat2012_em	Matrículas ensino médio
pop2010_homen	Nº de homens
pop2010_mulher	Nº de mulheres
pop2010alfab	Nº de alfabetizados
pop2010freq_ei	População residente que frequentava creche ou escola
pop2010catolromano	Nº de católicos
pop2010espitita	Nº de espíritas
pop2010evang	Nº de evangélicos
rend2010urbana	Renda urbana
rend2010rural	Renda rural
agro2010	PIB agropecuária
ind2010	PIB indústria
serv2010	PIB serviços
impostos2010	Arrecadação de impostos
pib2010corrente	PIB corrente
pibpccor2010	PIB corrente <i>per capita</i>
idhm2010	IDH – municipal
pib2010pcap	PIB <i>per capita</i>
nmedio_sm2011	Salário médio mensal (em SM) <i>per capita</i>
limitrofe	Zona está na fronteira do estado
proxi_ativ_turismo	<i>Proxy</i> de atividade de turismo
media_leito12_sum	Leitos hospitalares (Média dos 12 meses)
media_clinica12_sum	Clinicas (Média dos 12 meses)
matric_sup12_pres_sum	Matrículas ensino superior presencial
mat_sup_percapita	Matrículas ensino superior <i>per capita</i>
automovel	Número de veículos automóvel registrados
caminhao	Número de veículos caminhão registrados
caminhaotrator	Número de veículos trator registrados
caminhonete	Número de veículos caminhonete registrados
camioneta	Número de veículos camioneta registrados
micro_onibus	Número de veículos micro ônibus registrados
motocicleta	Número de veículos motocicleta registrados
motoneta	Número de veículos motoneta registrados
onibus	Número de veículos ônibus registrados
msal_cnae_55	Massa salarial de alojamentos
msal_cnae_56	Massa salarial de alimentação
msal_cnae_55_56	Massa salarial de alojamentos e alimentação
Dist_capital_mais_proxima	Distância entre a zona e a capital de UF mais próxima
densidade_demografica	Densidade demográfica
area_pcap	Área <i>per capita</i>
msal55_pcap	Massa salarial de alojamentos <i>per capita</i>
msal56_pcap	Massa salarial de alimentação <i>per capita</i>
msal55_56_pcap	Massa salarial de alojamentos e alimentação <i>per capita</i>
onibus_pcap	Ônibus <i>per capita</i>
automovel_pcap	Automóvel <i>per capita</i>
veic_pcap	Veículo <i>per capita</i>
veiculos	Soma de veículos
flag_nivel_0	Indica se a zona é nível 0
flag_nivel_1	Indica se a zona é nível 1
flag_nivel_2	Indica se a zona é nível 2
flag_nivel_3	Indica se a zona é nível 3
flag_nivel_4	Indica se a zona é nível 4
flag_nivel_5	Indica se a zona é nível 5
flag_nivel_6	Indica se a zona é nível 6

(Continua)

(Continuação)

Variável	Descrição
flag_nivel_7	Indica se a zona é nível 7
flag_nivel_8	Indica se a zona é nível 8
flag_nivel_9	Indica se a zona é nível 9
flag_nivel_10	Indica se a zona é nível 10
flag_nivel_11	Indica se a zona é nível 11
grupo_nivel	Grupo de nível (A=1,2,3,4; B=5,6;C=7,8;D=9,10,11)
flag_nivel_A	Indica se a zona está nos níveis 1, 2, 3 e 4
flag_nivel_B	Indica se a zona está nos níveis 5, 6
flag_nivel_C	Indica se a zona está nos níveis 7, 8
flag_nivel_D	Indica se a zona está nos níveis 9, 10, 11
flag_nivel_X	Indica se a zona está no nível 0
nao_amaz_pop2012	Interação nao_amazonia e pop2012
limitrofe_pop2012	Interação limitrofe e pop2012
nivel_0_pop2012	Interação flag_nivel_0 e pop2012
nivel_1_pop2012	Interação flag_nivel_1 e pop2012
nivel_2_pop2012	Interação flag_nivel_2 e pop2012
nivel_3_pop2012	Interação flag_nivel_3 e pop2012
nivel_4_pop2012	Interação flag_nivel_4 e pop2012
nivel_5_pop2012	Interação flag_nivel_5 e pop2012
nivel_6_pop2012	Interação flag_nivel_6 e pop2012
nivel_7_pop2012	Interação flag_nivel_7 e pop2012
nivel_8_pop2012	Interação flag_nivel_8 e pop2012
nivel_9_pop2012	Interação flag_nivel_9 e pop2012
nivel_10_pop2012	Interação flag_nivel_10 e pop2012
nivel_11_pop2012	Interação flag_nivel_11 e pop2012
nivel_1_msal_cnae_56	Interação flag_nivel_1 e msal_cnae_56
nivel_2_msal_cnae_56	Interação flag_nivel_2 e msal_cnae_56
nivel_3_msal_cnae_56	Interação flag_nivel_3 e msal_cnae_56
nivel_4_msal_cnae_56	Interação flag_nivel_4 e msal_cnae_56
nivel_5_msal_cnae_56	Interação flag_nivel_5 e msal_cnae_56
nivel_6_msal_cnae_56	Interação flag_nivel_6 e msal_cnae_56
nivel_7_msal_cnae_56	Interação flag_nivel_7 e msal_cnae_56
nivel_8_msal_cnae_56	Interação flag_nivel_8 e msal_cnae_56
nivel_9_msal_cnae_56	Interação flag_nivel_9 e msal_cnae_56
nivel_10_msal_cnae_56	Interação flag_nivel_10 e msal_cnae_56
nivel_11_msal_cnae_56	Interação flag_nivel_11 e msal_cnae_56
nivel_0_pibcorrente	Interação flag_nivel_0 e pib2010corrente
nivel_1_pibcorrente	Interação flag_nivel_1 e pib2010corrente
nivel_2_pibcorrente	Interação flag_nivel_2 e pib2010corrente
nivel_3_pibcorrente	Interação flag_nivel_3 e pib2010corrente
nivel_4_pibcorrente	Interação flag_nivel_4 e pib2010corrente
nivel_5_pibcorrente	Interação flag_nivel_5 e pib2010corrente
nivel_6_pibcorrente	Interação flag_nivel_6 e pib2010corrente
nivel_7_pibcorrente	Interação flag_nivel_7 e pib2010corrente
nivel_8_pibcorrente	Interação flag_nivel_8 e pib2010corrente
nivel_9_pibcorrente	Interação flag_nivel_9 e pib2010corrente
nivel_10_pibcorrente	Interação flag_nivel_10 e pib2010corrente
nivel_11_pibcorrente	Interação flag_nivel_11 e pib2010corrente
nivel_A_pop2012	Interação flag_nivel_A e pop2012
nivel_B_pop2012	Interação flag_nivel_B e pop2012
nivel_C_pop2012	Interação flag_nivel_C e pop2012
nivel_D_pop2012	Interação flag_nivel_D e pop2012
nivel_X_pop2012	Interação flag_nivel_X e pop2012
nivel_ABCD_pop2012	Interação flag_nivel_ABCD e pop2012
nivel_A_msal_cnae_56	Interação flag_nivel_A e msal_cnae_56
nivel_B_msal_cnae_56	Interação flag_nivel_B e msal_cnae_56
nivel_C_msal_cnae_56	Interação flag_nivel_C e msal_cnae_56
nivel_D_msal_cnae_56	Interação flag_nivel_D e msal_cnae_56
nivel_X_msal_cnae_56	Interação flag_nivel_X e msal_cnae_56
nivel_ABCD_msal_cnae_56	Interação flag_nivel_ABCD e msal_cnae_56
nivel_A_pibcorrente	Interação flag_nivel_A e pib2010corrente
nivel_B_pibcorrente	Interação flag_nivel_B e pib2010corrente
nivel_C_pibcorrente	Interação flag_nivel_C e pib2010corrente
nivel_D_pibcorrente	Interação flag_nivel_D e pib2010corrente
nivel_X_pibcorrente	Interação flag_nivel_X e pib2010corrente
nivel_ABCD_pibcorrente	Interação flag_nivel_ABCD e pib2010corrente
nao_amazonia	Variável dicotômica que representa se a zona não está na Amazônia

## **Ipea – Instituto de Pesquisa Econômica Aplicada**

### **EDITORIAL**

#### **Coordenação**

Cláudio Passos de Oliveira

#### **Supervisão**

Everson da Silva Moura

Reginaldo da Silva Domingos

#### **Revisão**

Clícia Silveira Rodrigues

Idalina Barbara de Castro

Leonardo Moreira Vallejo

Marcelo Araujo de Sales Aguiar

Marco Aurélio Dias Pires

Olavo Mesquita de Carvalho

Regina Marta de Aguiar

Alessandra Farias da Silva (estagiária)

Lilian de Lima Gonçalves (estagiária)

Luiz Gustavo Campos de Araújo Souza (estagiário)

Paulo Ubiratan Araujo Sobrinho (estagiário)

Pedro Henrique Ximendes Aragão (estagiário)

#### **Editoração**

Bernar José Vieira

Cristiano Ferreira de Araújo

Daniella Silva Nogueira

Danilo Leite de Macedo Tavares

Jeovah Herculano Szervinsk Junior

Leonardo Hideki Higa

Herllyson da Silva Souza (estagiário)

#### **Capa**

Andrey Tomimatsu

*The manuscripts in languages other than Portuguese  
published herein have not been proofread.*

#### **Livraria Ipea**

SBS – Quadra 1 – Bloco J – Ed. BNDES, Térreo

70076-900 – Brasília – DF

Tel.: (61) 2026 5336

Correio eletrônico: [livraria@ipea.gov.br](mailto:livraria@ipea.gov.br)







## **Missão do Ipea**

Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.

**ipea** Instituto de Pesquisa  
Econômica Aplicada

MINISTÉRIO DO  
**PLANEJAMENTO,  
DESENVOLVIMENTO E GESTÃO**

