

**UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE ENGENHARIA  
CURSO DE MESTRADO EM GEOTECNIA E TRANSPORTES**

José Moreira Gonçalves

**MODELOS DE ATRAÇÃO DE ENTREGAS E CARGAS PARA A INDÚSTRIA DE  
MÁQUINAS AGRÍCOLAS E RODOVIÁRIAS**

Belo Horizonte  
2019

José Moreira Gonçalves

**MODELOS DE ATRAÇÃO DE ENTREGAS E CARGAS PARA A INDÚSTRIA DE  
MÁQUINAS AGRÍCOLAS E RODOVIÁRIAS**

**Versão final**

Dissertação apresentada ao Curso de Mestrado em Geotecnia e Transportes da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Geotecnia e Transportes.

Área de concentração: Transportes

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Renata Lúcia Magalhães de Oliveira

Belo Horizonte  
2019

G635m	<p>Gonçalves, José Moreira.  Modelos de atração de entregas e cargas para a indústria de máquinas agrícolas e rodoviárias [recurso eletrônico] / José Moreira Gonçalves. – 2019.  1 recurso online (138 f. : il., color.) : pdf.</p> <p>Orientadora: Renata Lúcia Magalhães de Oliveira.</p> <p>Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Engenharia.</p> <p>Apêndices: f. 121-138.</p> <p>Bibliografia: f. 106-120.  Exigências do sistema: Adobe Acrobat Reader.</p> <p>1. Transportes - Teses. 2. Transporte de cargas - Teses.  3. Máquinas agrícolas - Indústria - Teses. 4. Máquinas rodoviárias - Teses.  I. Oliveira, Renata Lúcia Magalhães de. II. Universidade Federal de Minas Gerais. Escola de Engenharia. III. Título.</p> <p style="text-align: right;">CDU: 656(043)</p>
-------	--

Ficha catalográfica: Biblioteca Profº Mário Werneck, Escola de Engenharia da UFMG.



## FOLHA DE APROVAÇÃO

### MODELOS DE ATRAÇÃO DE ENTREGAS E CARGAS PARA A INDÚSTRIA DE MÁQUINAS AGRÍCOLAS E RODOVIÁRIAS

**JOSÉ MOREIRA GONÇALVES**

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GEOTECNIA E TRANSPORTES, como requisito para obtenção do grau de Mestre em GEOTECNIA E TRANSPORTES, área de concentração TRANSPORTES.

Aprovada em 24 de junho de 2019, pela banca constituída pelos membros:

Prof. Renata Lúcia Magalhães de Oliveira - Orientadora  
UFMG

Prof. Leise Kelli de Oliveira  
UFMG

Prof. Francisco Gildeir Ferreira da Silva  
UFC

Belo Horizonte, 24 de junho de 2019.

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus que tem me ajudado a conquistar todos os meus objetivos dando-me saúde, ânimo e sabedoria. Sem Ele eu não teria chegado até aqui.

Aos meus pais adotivos, por terem feito tudo que esteve ao alcance deles para que eu me tornasse uma pessoa forte e determinada frente aos obstáculos da vida.

A minha mãe biológica que me deu a vida e após 25 anos, em janeiro de 2018, pude abraça-la novamente.

Em especial à Diana Teixeira, pela cumplicidade, compreensão e amor demonstrado todos os dias.

Aos meus colegas e amigos que sempre me incentivaram e que fizeram parte deste trabalho.

Aos professores do curso pelos ensinamentos e orientações prestadas.

Aos membros da banca Prof<sup>a</sup>. Dr<sup>a</sup>. Leise e Prof. Dr. Francisco Gildemir por terem gentilmente aceitado o convite de fazer parte desse trabalho.

À coordenadora do curso de Mestrado em Geotecnia e Transportes, Prof<sup>a</sup>. Dr<sup>a</sup>. Leise Kelli de Oliveira.

À minha orientadora Prof<sup>a</sup>. Dr<sup>a</sup>. Renata Lúcia Magalhães de Oliveira pelas orientações, ensinamentos técnicos e alegria demonstrada no desenvolvimento deste trabalho.

A Universidade Federal de Minas Gerais (UFMG), em especial a Escola de Engenharia.

Enfim, a todos que participaram desta conquista.

*"Eu prefiro ser essa metamorfose ambulante, do que ter aquela velha opinião  
formada sobre tudo"*

Raul Seixas *in* Metamorfose ambulante

## RESUMO

GONCALVES, José Moreira. **Modelos de Atração de Entregas e Cargas para a Indústria de Máquinas Agrícolas e Rodoviárias** 138f. Dissertação (Mestrado em Geotecnia e Transportes) – Universidade Federal de Minas Gerais, Belo Horizonte, 2019.

O desenvolvimento econômico das cidades está intimamente interligado à distribuição de mercadorias dos diversos setores industriais. Entretanto, essas atividades também geram impactos negativos que podem ser mitigados quando se conhece o comportamento dos fluxos de carga. Para isso, é importante o desenvolvimento de modelos de geração de viagens de carga para subsídio ao planejamento das cidades. Assim, por meio deste trabalho, determinaram-se modelos de atração de entregas e cargas para indústrias de máquinas agrícolas e rodoviárias e compara-os conceitualmente. Esses modelos são importantes para estimar fluxos de abastecimento de insumos nesse setor e foram desenvolvidos por meio de dados de quatro indústrias localizadas em Contagem (Minas Gerais), Curitiba (Paraná), Piracicaba e Sorocaba (São Paulo). Foi realizada uma revisão sistemática da literatura sobre geração de viagens de carga para melhor compreensão do conceito e métodos utilizados. Realizou-se uma coleta de dados secundários com informações sobre o fluxo de veículos para abastecimento de insumos nas indústrias consideradas neste trabalho e o recorte temporal compreendeu os meses entre janeiro e dezembro de 2017, contemplando 24 h de operação para cada indústria observada. Foram então realizadas análises estatísticas dos dados coletados e desenvolvidos modelos de regressão linear generalizada. Os modelos foram avaliados quanto à sua capacidade preditiva utilizando o método de validação cruzada *leave-one-out* e comparados qualitativamente. Modelos com melhor acurácia preditiva e equilíbrio entre simplicidade e ajuste são aqueles desenvolvidos considerando a frequência de entregas como variável dependente e a produção, o número de funcionários alocados em atividades de logística e a área total construída em cada unidade industrial como variáveis explicativas. Nenhum dos modelos que consideraram a quantidade de carga recebida como variável dependente apresentou ajuste satisfatório.

Palavras-chave: Transporte de carga; modelos de geração de viagens de carga, indústria de máquinas agrícolas e rodoviárias.

## ABSTRACT

GONCALVES, José Moreira. **Delivery and Load Attraction Models for the Agricultural and Road Machinery Industry** 138f. Dissertação (Mestrado em Geotecnia e Transportes) – Universidade Federal de Minas Gerais, Belo Horizonte, 2019.

The economic development of cities is closely related to the distribution of goods concerning various industrial sectors. However, these activities also generate negative impacts that can be mitigated when the freight flows are acknowledged. Thus, it is important to develop models for freight trips to subsidize city planning. So, in this work, freight generation models for agricultural and road machinery industries were determined and compared conceptually. These models are important for estimating supply flows in this sector and were developed using data from four industries located in Contagem (Minas Gerais), Curitiba (Paraná), Piracicaba and Sorocaba (São Paulo). A systematic review of the literature regarding freight generation models was conducted to better understand the concept and methods used. Information was collected concerning the supply flows in the industrial facilities considered in this work. The data comprised trips performed from January until December 2017, contemplating 24 hours of operation for each observed industry. Statistical analyses of the collected data were then carried out and generalized linear regression models were developed. The models were evaluated regarding their predictive capacity using the leave-one-out cross-validation method and compared qualitatively. Models with better predictive accuracy and balance between simplicity and adjustment were those developed considering the frequency of deliveries as the dependent variable and production, the number of employees allocated to logistics activities and the total built area in each industrial unit as explanatory variables. None of the models that considered the amount of cargo received as a dependent variable presented satisfactory fit.

Keywords: Freight transport; freight trip generation models, agricultural and road machinery industry



# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>16</b>
1.1	OBJETIVOS DO TRABALHO.....	20
1.2	JUSTIFICATIVA DO TEMA .....	20
1.3	ESTRUTURA DA DISSERTAÇÃO .....	23
<b>2</b>	<b>REVISÃO DA LITERATURA .....</b>	<b>25</b>
2.1	MODELAGEM DA DEMANDA NO TRANSPORTE DE CARGAS .....	25
2.2	MODELOS DE GERAÇÃO DE FLUXOS E VIAGENS DE CARGA.....	34
<b>3</b>	<b>ABORDAGEM METODOLÓGICA .....</b>	<b>62</b>
3.1	CARACTERIZAÇÃO ECONÔMICA E TERRITORIAL.....	63
3.2	ORGANIZAÇÃO DAS INFORMAÇÕES.....	64
3.3	MODELAGEM E AVALIAÇÃO.....	69
3.4	ANÁLISE COMPARATIVA CONCEITUAL .....	76
<b>4</b>	<b>RESULTADOS E DISCUSSÕES .....</b>	<b>77</b>
4.1	CARACTERIZAÇÃO ECONÔMICA E TERRITORIAL.....	77
4.2	ORGANIZAÇÃO DAS INFORMAÇÕES.....	84
4.3	MODELAGEM E AVALIAÇÃO.....	94
4.4	ANÁLISE COMPARATIVA CONCEITUAL .....	100
<b>5</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>103</b>
	<b>REFERÊNCIAS .....</b>	<b>106</b>
	<b>APÊNDICE A – PROCEDIMENTO DE REVISÃO SISTEMÁTICA.....</b>	<b>121</b>
	<b>APÊNDICE B – TESTES REALIZADOS NOS DADOS .....</b>	<b>123</b>
	<b>APÊNDICE C – FUNÇÕES DESENVOLVIDAS NO SOFTWARE R.....</b>	<b>128</b>

## LISTA DE FIGURAS

Figura 2.1: Número de trabalhos publicados entre 1974-2018. ....	29
Figura 2.2: Autores com maior número de publicações. ....	30
Figura 2.3: Principais variáveis independentes. ....	32
Figura 2.4: Principais modelos utilizados para demanda de carga. ....	33
Figura 3.1: Estrutura metodológica do trabalho. ....	63
Figura 4.1: Localização das indústrias e fornecedores. ....	78
Figura 4.2: Quantidade de fornecedores por Estado Brasileiro. ....	82
Figura 4.3: Box-plot – Entregas diária de março a outubro de 2017. ....	87
Figura 4.4: Box-plot – Carga (kg) movimentada por dia de março a outubro de 2017. ....	89
Figura 4.5: Série histórica do número de entregas. ....	91
Figura 4.6: Matriz de correlação de dados de Spearman. ....	94

## LISTA DE TABELAS

Tabela 2.1: Categorias de síntese das variáveis dependentes e independentes. ....	31
Tabela 2.2: Revisão sistemática da literatura. ....	60
Tabela 4.1: Empresas fabricantes de máquinas agrícolas e rodoviárias.....	79
Tabela 4.2: Dados demográficos dos municípios onde são localizadas as indústrias analisadas. .....	80
Tabela 4.3: Caracterização das instalações industriais analisadas. ....	80
Tabela 4.4: Quantidade de fornecedores em cada Estado por indústria.....	82
Tabela 4.5: Número de entregas por Estado de origem (Ano 2017). ....	83
Tabela 4.6: Quantidade de carga movimentada em quilogramas (kg) por Estado de origem (Ano 2017).....	83
Tabela 4.7: Número de entregas mensais atraídas pelas unidades industriais no ano de 2017.	84
Tabela 4.8: Quantidade de carga (kg) atraídas mensalmente pelas unidades industriais no ano 2017. ....	86
Tabela 4.9: Dados de entregas diária atraída pelas indústrias de março a outubro de 2017. ...	87
Tabela 4.10: Dados de carga (kg) atraída por dia pelas unidades industriais de março a outubro de 2017. ....	88
Tabela 4.11: Dados para geração dos modelos de regressão.....	93
Tabela 4.12: Modelos lineares generalizados de atração de entregas e carga. ....	97
Tabela 4.12: Modelos lineares generalizados de atração de entregas e carga – continuação...	98
Tabela 4.12: Modelos lineares generalizados de atração de entregas e carga – continuação...	99

## LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

AIC - *Akaike Information Criterion* (Critério de Informação de Alaike)

AIQ - Amplitude Interquartílica

ANFAVEA - Associação Nacional dos Fabricantes de Veículos Automotores

Arealog - Área de Logística

Areatot - Área Total das Indústrias

BP - Teste de Breusch-Pagan (Multiplicador de Lagrange)

BPNN - *Backpropagation Neural Network*

CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

car - Quantidade de Carga Recebida Diariamente

CFS - *Commodity Flow Survey*

ent - Número de Entregas Diárias

EUA - Estados Unidos da América

FAF - *Freight Analysis Framework*

FTA - *Freight Trip Attraction*

FTG - *Freight Trip Generation*

FTS - *Freight Tour Synthesis*

funclog - Número de Funcionários da Logística

functot - Número de Funcionários Total

GLM - *Generalized Linear Model* (Modelo Linear Generalizado)

GLML - Regressão Linear Generalizada com Link de Log

GOF - *Goodness Of Fit*

GV - Geração de Viagem

GVC - Geração de Viagem de Carga

h - Hora (as)

IBGE - Instituto Brasileiro de Geografia e Estatística

ITE - *Institute of Transportation Engineers* (Instituto dos Engenheiros de Transportes)

ITS - Sistemas Inteligentes de Transporte

kg - Quilograma

KOLMAP - *Kocaeli Logistics Master Plan*

LBCS - *Land-Based Classification Standards*

LOOCV - *Leave-one-out Cross Validation* (Validação Cruzada)

LSP - *Logistics Service Providers*

LTG - *Large Urban Freight Traffic Generators*

m<sup>2</sup> - Metro Quadrado

MAPE - *Mean Absolute Percentage Error*

MCA - *Multiple Classification Analysis*

MG - Minas Gerais

MGVC - Modelos de Geração de Viagem de Carga

MLE - *Maximum-Likelihood Estimation* (Estimativa por Máxima Verossimilhança)

MLFNN - Modelo Elaborado por meio da Técnica de Rede Neural Acíclica Multicamada - *Multilayer Feedforward Neural Network*

NAICS - *North American Industry Classifications System*

NCFRP 25 - *National Cooperative Freight Research Program Project 25*

NETS - *National Employment Time Series*

NHTS - *National Travel Survey*

NYCZR - *City of New York Zoning Resolution*

OLS - *Ordinary Least Squares*

OTM - *Oracle Transportation Management*

PGV - Polo Gerador de Viagens

PGV Carga - Polo Gerador de Viagens de Carga

PIB - Produto Interno Bruto

PR - Paraná

Prod - Produção Diária

QRFM - *Quick Response Freight Manual*

R<sup>2</sup> - Coeficiente de Determinação

REQM - Raiz do Erro Quadrático Médio

RJ - Rio de Janeiro

RMRJ - Região Metropolitana do Rio de Janeiro

RMSE - *Root Mean Squared Error*

RNA - Redes Neurais Artificiais

RS - Rio Grande do Sul

SC - Santa Catarina

SIC - *Standard Industrial Classification*

SP - São Paulo

SQEP - Soma dos Quadrados do Erro de Previsão

STA - *Shopping Trip Attraction*

TRB - *Transportation Research Board* (Conselho de Pesquisa sobre Transporte)

UFMG - Universidade Federal de Minas Gerais

UTPS - *Urban Transportation Planning System*

VC - Validação Cruzada

VD - Variáveis Dependentes

VI - Variáveis Independentes

W - Teste de Shapiro-Wilk

$\chi^2$  - Teste Qui-quadrado

YB - *Number of Years in Business*

# 1 INTRODUÇÃO

A expansão urbana que vem acontecendo de forma desordenada ao longo dos anos, agravada por uma ausência de planejamento e estratégias específicas para seu controle, pode contribuir para o aumento dos problemas inerentes à ocupação do espaço físico realizado por diversas atividades econômicas, dentre elas as indústrias (SILVA e GOLDNER, 2010).

O uso e a ocupação do solo pelas indústrias, em muitos casos, são direcionados por uma definição de mercado, usualmente sem alinhamento às políticas públicas, realizados de acordo com a própria necessidade das organizações privadas (SILVA e GOLDNER, 2010). Algumas das razões para as indústrias se desenvolverem nas áreas urbanas são a proximidade dos clientes, a oferta de mão de obra, dos serviços relacionados a educação e saúde, a disponibilidade de terrenos em áreas adjacentes com preço acessível e a conectividade promovida pela infraestrutura de transportes existente (SILVA e GOLDNER, 2010). Além disso, em determinados contextos, para que a indústria se desenvolva, faz-se necessária à concentração espacial de atividades econômicas, denominadas “economias de aglomeração<sup>1</sup>”, que pode promover economias de escala em relação à infraestrutura, à oferta de mão de obra qualificada e a proximidade de outras indústrias complementares (CARLOS, 1997).

Nesse contexto, as indústrias se conformam como Polos Geradores de Viagens de Carga (PGV Carga) (OLIVEIRA *et al.*, 2017; RAMOS, 2015) e como parte importante das atividades econômicas nas cidades, demandando a movimentação de insumos para o abastecimento dos seus processos produtivos e gerando fluxos de expedição dos bens manufaturados para abastecimento local ou em escalas regional e global. Essas atividades econômicas são essenciais para a vitalidade das cidades e o seu resultado direto como demanda por transportes, é crucial para a geração de renda, assim como para a criação e a manutenção de empregos em escalas local e regional (FERREIRA, 2015; GONZALEZ-FELIU *et al.*, 2014; DUTRA, 2004; OECD, 2003; ALLEN *et al.*, 2000). Entretanto,

---

<sup>1</sup> Entende-se por economias de aglomeração uma etapa do processo de industrialização em que ocorre a concentração da instalação de empresas produtivas ou de atividades econômicas em uma determinada região densa do espaço geográfico (PENA, 2018).



atividades relacionadas à cadeia de suprimentos industriais, tais como o recebimento de suprimentos e a expedição de bens manufaturados, requerem o transporte de carga, e, devido à sua natureza, produzem ou atraem viagens, promovendo impactos negativos no seu contexto territorial. Essas externalidades negativas apresentam-se nas dimensões ambientais, econômicas e sociais, resultando no incremento do nível de congestionamento no seu entorno, comprometimento da qualidade ambiental por conta da emissão de gases de efeito estufa e particulados, poluição sonora, maior risco de acidentes nas vias adjacentes ao empreendimento, além da consequente ineficiência logística da movimentação de insumos e na expedição do produto acabado (LAWSON *et al.*, 2012; HOLGUÍN-VERAS *et al.*, 2012; CUI, DODSON e HALL, 2015; OLIVEIRA, 2014; CORREIA, 2011; OGDEN, 1992).

É importante destacar que os impactos causados pelos PGV Carga no sistema viário, podem ter uma relação direta ou indireta com o empreendimento, como resultado do tráfego incremental que eles geram (GONÇALVES *et al.*, 2012). Os impactos diretos têm relação com o sistema viário e com a circulação, resultando no aumento de veículos nas vias, no tempo de viagens e no número de acidentes, dentre outros. De forma indireta têm-se às alterações do ambiente, como a diminuição da mobilidade e da acessibilidade, bem como a impactos econômicos e ambientais, conforme destaca Kneib (2004).

Além dos impactos diretamente associados à circulação, apresentados anteriormente, Ogden (1992) também apresenta o congestionamento nas vias devido às dimensões dos caminhões e taxas de aceleração/desaceleração como possíveis consequências da circulação de mercadorias. Embora ao longo dos anos o congestionamento viário nas grandes e médias cidades venha aumentando gradativamente, o foco dos estudos de tráfego tem sido o impacto dos automóveis para transporte individual e a parcela que o transporte de carga tem quanto ao incremento dos níveis de congestionamento tem recebido relativamente pouca atenção (IDING, MEESTER e TAVASSZY, 2002).

Campos *et al.* (2012) ressaltam a importância da identificação dos fluxos de veículos de carga gerados pelos PGV Carga como subsídio ao gerenciamento da mobilidade da região do entorno dessas instalações. É notório que as autoridades locais ainda não estão aptas para minimizar os impactos causados pelos PGV Carga, pois, dentre outras razões, eles desconhecem como se desenvolvem os fluxos de cargas em seu território, em especial

daqueles advindos de atividades industriais (DABLANC, 2007; BINSBERGEN e VISSSEN, 2001; MARRA, 1999). A inexistência de informações sobre os processos que conformam a demanda pela movimentação de mercadorias são significativas. Em seu trabalho, Kaszubowski (2017) discute as limitações analíticas para investigações da problemática do transporte urbano de mercadorias em função da disponibilidade de dados de viagens.

Para suprir a escassez de informações e auxiliar no desenvolvimento de políticas públicas para melhoria do sistema de transporte de carga, modelos de geração de viagens de carga podem ser utilizados. Tais modelos visam permitir a estimação do número total de viagens de veículos de carga produzidas ou atraídas por um PGV Carga, por meio de variáveis que representam características de uma atividade, localidade ou região e seu eixo de deslocamento (CAMPOS, 2013; LAWSON *et al.*, 2012; GASPARINI, 2008; LOPES, 2005).

No Brasil, a necessidade de contribuição da academia quanto aos modelos de geração de viagens de carga, se torna ainda mais evidente devido ao baixo número de estudos, conforme relatado por Souza, Silva e D'Agosto (2010). Outro fator preponderante para investigações que tenham como objetivo a determinação de modelos de geração de viagens de carga é a indispensabilidade de que esses modelos sejam adaptados às condições locais de cada PGV Carga e de seu respectivo contexto territorial (GRIECO e PORTUGAL, 2010).

Melo (2002), Iding, Meester e Tavasszy (2002), Black (1999), Tadi e Baldach (1994), Ogden (1992), Ogden (1977) e Hutchinson (1974), como exemplo, trabalharam seus modelos divididos por setores de atividades econômicas do mercado. Essa segmentação evidencia a importância do tipo da carga e do setor de atividade econômica ao qual o transporte está relacionado como elemento fundamental na determinação das quantidades de carga movimentada em fluxos de produção e atração e das respectivas viagens para modelagem ser com foco na previsão de demanda.

Nesse contexto, torna-se importante a investigação sobre a movimentação de insumos para suprimento de indústrias (Polos Geradores de Viagens de Carga - Industriais), em especial para a quantificação de carga movimentada e respectivas viagens produzidas e atraídas. Visando contribuir com o desenvolvimento de informações e modelos de geração de viagens de carga no contexto brasileiro, neste trabalho, foram desenvolvidos modelos de atração de

carga e entregas no suprimento de indústrias de máquinas agrícolas e rodoviárias. Esses modelos foram estruturados por meio de métodos de regressão, considerando: (i) entrega por veículos de carga como indicador de atração nos fluxos de suprimento; e (ii) a quantidade total de carga movimentada no abastecimento como elemento explicativo desses fluxos. Foram consideradas informações referentes a fluxos de abastecimento de quatro indústrias de máquinas agrícolas e rodoviárias localizadas nos municípios de Contagem (Minas Gerais), Curitiba (Paraná) e Piracicaba e Sorocaba (São Paulo). O detalhamento metodológico das etapas de desenvolvimento deste trabalho serão apresentados nas respectivas seções.

Diante do exposto, a questão que norteou esta pesquisa foi:

Qual abordagem é mais adequada para se estimar, por meio de dados secundários, fluxos de abastecimento de insumos para a indústria de máquinas agrícolas e rodoviárias: modelagem da atração de entregas por veículos de carga ou modelagem baseada na quantidade de carga movimentada?

Algumas hipóteses foram consideradas para a estruturação deste trabalho, a saber: (i) a técnica de regressão linear generalizada é adequada para modelar a atração de entregas por veículos de carga e de carga movimentada nos fluxos de abastecimento de indústrias de máquinas agrícolas e rodoviárias; (ii) o comportamento da correlação entre variáveis explicativas e a variável dependente (entregas ou quantidade de insumos) está fortemente relacionado com o contexto e o local em que as regressões são geradas; (iii) é possível identificar as variáveis independentes que melhor explicam a atração de entregas por veículos de carga e a atração de insumos nos fluxos de abastecimento da indústria de máquinas agrícolas e rodoviárias; (iv) os modelos de atração de viagens-carga contribuem para as estratégias de políticas públicas voltadas ao planejamento de transportes de carga.

Além da estruturação dos modelos de atração de carga e entregas, os modelos foram avaliados e uma análise comparativa qualitativa foi realizada, visando contribuir com a discussão sobre as propostas metodológicas para construção de modelos de demanda para Polos Geradores de Viagens de Carga (PGV-Carga) que podem considerar tanto dados referentes à quantidade de carga movimentada, como o número de viagens ou a frequência de entregas em um determinado período de tempo. Destaca-se ainda a importância desses modelos como

instrumento para subsídio à formulação e implementação de políticas públicas e diretrizes para regulação urbana, visando promover a mitigação de externalidades advindas do incremento das atividades industriais de manufatura de máquinas agrícolas e rodoviárias quando novas unidades industriais são instaladas em cidades ou áreas adjacentes que possam gerar conflitos locais em relação ao uso do solo e ao acesso às atividades urbanas.

## **1.1 Objetivos do trabalho**

Neste trabalho, tem-se como objetivo geral desenvolver e avaliar modelos de atração de carga e entregas para o setor de indústrias fabricante de máquinas agrícolas e rodoviárias no contexto brasileiro. Como objetivos específicos têm-se: (i) identificar variáveis explicativas, por meio de dados secundários, para atração de carga e entregas para a atividade industrial em análise; (ii) selecionar modelos e técnicas de modelagem que melhor representam os processos em investigação; (iii) avaliar os modelos quanto ao ajuste, ao equilíbrio entre complexidade e ajuste, e sua capacidade preditiva; e (iv) por meio de avaliação conceitual, contribuir para o estado da arte com a discussão comparativa sobre abordagens metodológicas que considerem volumes de carga e frequência de entregas como variáveis dependentes em modelos de demanda para PGV-Carga.

## **1.2 Justificativa do Tema**

Os impactos gerados por empreendimentos de grande porte na infraestrutura de transportes, na dinâmica das cidades e na qualidade de vida das pessoas, são percebidos por meio do acentuado fluxo de suprimentos e das relações estabelecidas em dado espaço, em razão dos intensos processos que ocorrem atualmente na relação produção e consumo (SILVA e FERREIRA, 2017).

Para Holguín-Veras *et al.* (2012) a atividade de transporte de mercadorias representa a atividade econômica em movimento. Para esses autores, essa afirmativa é verdadeira visto que na maioria dos casos as transações monetárias são acompanhadas por um fluxo de mercadorias no sentido oposto. Basicamente, uma cadeia de suprimentos típica é composta de muitas ligações produção e consumo, onde um agente econômico produz carga que outros agentes processam, armazenam e entregam aos consumidores finais ou intermediários (HOLGUÍN-VERAS *et al.*, 2012).

Gasparini, Campos e D'Agosto (2007) identificam que a circulação de bens de produção e consumo provoca um complexo fluxo de veículos no meio urbano devido à diversidade de atividades realizadas em regiões centrais das cidades, em polos periféricos ou dentro das regiões metropolitanas. Nesse sentido, surge a necessidade de estudar os fluxos de suprimentos no que se refere à geração de demanda na movimentação de cargas, considerando tanto a frequência de entregas como a quantidade de carga movimentada, no intuito de minimizar as externalidades dessa atividade.

Kneib *et al.* (2006) e Facchini (2006) afirmam que a implantação de Polos Geradores de Viagens no espaço urbano provoca tanto impactos positivos, ao possibilitar o desenvolvimento e valorizar áreas influenciadas por ele, quanto negativos, por afetar o desempenho da mobilidade e da acessibilidade na área onde esses polos estão instalados. A infraestrutura viária em torno de instalações industriais muitas vezes têm capacidade insuficiente para suportar os fluxos de tráfego que ocorrem na região. Problemas de congestionamento, por exemplo, atingem tanto a empresa fornecedora, que tem sua eficiência reduzida, quanto o cidadão, cuja qualidade de vida é influenciada pelos diferentes tipos de poluição e também pelas interferências dos veículos de carga, caminhões, nas vias (HOLGUÍN-VERAS *et al.*, 2012). De acordo com Giuliano e Dablanc (2013), os problemas do transporte de carga são consequências direta do aumento da quantidade de fluxos de mercadorias nas cidades. Além disso, o Poder Público sente esses impactos ao demonstrar grande dificuldade em regulamentar e minimizar os problemas advindos desse processo sem prejuízo à continuidade das atividades econômicas da região. Os fluxos de tráfego de transporte de suprimentos para dada localidade geralmente não são consideradas quando planeja-se novas instalações industriais (IDING, MEESTER e TAVASSZY, 2002).

Nesse contexto, justifica-se a abordagem da temática deste trabalho por representar um processo de relevância social, econômica e ambiental para as cidades. A escassez de dados que auxiliem o planejamento dos transportes de cargas, em especial daquelas geradas pelas indústrias instaladas no seu contexto territorial, impõe limitação ao desenvolvimento de políticas públicas mais efetivas que promovam condições adequadas de mobilidade e acessibilidade às atividades nas cidades. A falta de conhecimento sobre o transporte de carga

pelo setor público, se dá, em grande parte, pela dificuldade de obtenção dos dados de fluxos de suprimentos das indústrias que, em sua maioria, são empreendimentos privados.

Os modelos gerados por meio deste trabalho podem contribuir com o poder público, fornecendo subsídios para o planejamento operacional dos sistemas viário e de transportes e a proposição de medidas mitigadoras e compensatórias para empreendimentos de impacto cuja operação possa resultar em externalidades para a cidade. Assim, o desenvolvimento de modelos de previsão da demanda por cargas é de extrema importância para o planejamento urbano e de mobilidade, uma vez que permitem estimar a demanda relativa ao transporte de mercadorias (NOVAK *et al.*, 2011).

Este trabalho também apresenta relevância tanto devido à expectativa de aumento dos fluxos de viagens de carga ao longo das próximas décadas, associado a um aumento do setor industrial de produção de máquinas agrícolas e rodoviárias (Associação Nacional dos Fabricantes de Veículos Automotores - ANFAVEA, 2018), como pela contribuição com o estado da arte e com o estado da prática em relação à modelagem da movimentação de cargas nos fluxos de suprimento industrial, em especial, de montadoras de máquinas agrícolas e rodoviárias. Modelo específico para este setor econômico não foi encontrado nem na literatura internacional e menos ainda no contexto brasileiro.

Assim, outro aspecto relevante deste trabalho se dá, pois, a demanda por transporte de carga, como resultado de atividades industriais e comerciais, pode ter impactos significativos sobre a habitabilidade de áreas urbanas. Neste sentido, para estimar as viagens de carga, é necessário desenvolver modelos adaptados à realidade local e às características intrínsecas de cada atividade econômica (PORTUGAL *et al.*, 2012; GRIECO e PORTUGAL, 2010). Iding, Meester e Tavasszy (2002), Black (1999), Tadi e Baldach (1994), Ogden (1992), Ogden (1977) e Hutchinson (1974), como exemplo, diferenciaram os modelos desenvolvidos em categorias quanto ao contexto geográfico e atividade econômica.

Quanto à relevância acadêmica, tem-se a importância do tema por apresentar modelos de geração de entregas por veículos de carga e da quantidade de carga movimentada no suprimento de indústrias de máquinas agrícolas e rodoviárias no contexto brasileiro. Segundo Souza, Silva e D'Agosto (2010), o desenvolvimento de modelos de geração de viagens de

carga é um campo a ser explorado no Brasil devido ao baixo número de estudos realizados pela academia. Ainda segundo Sánchez-Díaz *et al.* (2016), a quantidade de estudos sobre geração de viagens de carga ainda é pequena e restrita a alguns setores. Sobre modelos de geração de viagens, em sua grande maioria, tratam-se de modelos de geração de viagens para automóveis, havendo assim uma escassez quanto a trabalhos referentes a modelos de geração de viagens de carga (OLIVEIRA *et al.*, 2016).

Apesar da importância dos empreendimentos industriais para os municípios, seja na difusão do avanço tecnológico, seja na geração de empregos, não se deve desconsiderar que eles provocam uma quantidade significativa de viagens e geram impactos no sistema viário e de transporte.

Assim, este estudo agrega contribuições ao estado da arte e ao estado da prática, contribuindo para o desenvolvimento científico dessa temática e para as práticas locais de licenciamento de empreendimentos alinhados ao segmento analisado pelo poder público local.

### **1.3 Estrutura da dissertação**

Além deste capítulo introdutório, onde se apresentam os objetivos deste estudo e a sua justificativa, esta dissertação apresenta mais quatro capítulos.

No capítulo dois será apresentada a revisão sistemática da literatura acerca da modelagem da demanda no transporte de cargas, cujo objetivo é identificar os principais pesquisadores e as variáveis que têm sido utilizadas em estudos de geração de viagens de carga presentes na literatura nacional e internacional. Ainda nesse capítulo, serão apresentados, uma síntese das principais publicações sobre modelos de geração de fluxos e viagens de carga.

No três tem-se a abordagem metodológica empregada no desenvolvimento deste trabalho, bem como as técnicas utilizadas nesta pesquisa.

O quarto, faz-se a apresentação dos resultados e discussões obtidos com a aplicação da metodologia apresentada, e por fim, o capítulo cinco encerra este trabalho apresentando as conclusões e apontando as recomendações de trabalhos no que se refere aos estudos dos

modelos de geração de entregas por veículos de carga e quantidade de carga movimentada para o setor industrial de máquinas agrícolas e rodoviárias em outras localidades.



## 2 REVISÃO DA LITERATURA

Neste capítulo encontram-se a fundamentação teórica e o estado da arte sobre os aspectos referentes a modelos de demanda e os principais conceitos, técnicas e ferramentas utilizadas no desenvolvimento deste trabalho.

### 2.1 Modelagem da demanda no transporte de cargas

Ortúzar e Willumsen (2011) definem um modelo como uma representação simplificada de uma parte do mundo real, no caso, o sistema de interesse. Para Gonzalez-Feliu e Routhier (2012), um modelo consiste em reduzir um objeto estudado para compreender os mecanismos envolvidos naquele fenômeno. Nesse contexto, a modelagem em transportes de carga visa representar e possibilitar a previsão de demanda futura por meio de recursos matemáticos, computacionais, comportamentais, dentre outros, de modo a representar as características de uma nova realidade.

Para Marra (1999), modelos matemáticos têm sido utilizados por planejadores e engenheiros para estimar demandas de tráfego de veículos, bem como determinar impactos decorrentes de mudanças no uso e ocupação do solo ou no sistema viário. Os modelos são importantes para o planejamento de transportes e das cidades, fornecendo informações que auxiliam no processo de tomada de decisões.

Kneib (2004) contribui com essa discussão quando destaca que o planejamento de transportes tem como base o uso de modelos. No planejamento de transportes, os modelos são usualmente utilizados para coordenar oferta e demanda, seja ela existente ou futura em determinada região, equacionando de forma apropriada os diferentes sistemas (FERREIRA e SILVA, 2016).

Holguín-Veras *et al.* (2011) e Gonzalez-Feliu e Routhier (2012) afirmam que para a construção de um bom modelo seria necessário um conhecimento detalhado da situação existente, principalmente daquelas características que estão sujeitas a alterações ou que são invariáveis e que a melhor unidade de observação deve ser tão rica quanto possível em termos de qualidade da informação.

Nesse sentido, o modelo de Quatro Etapas seria, tradicionalmente, o modelo mais usado pelo setor de transportes para análise e previsão de demanda. Tal modelo utiliza-se de técnicas de previsão de demanda de transportes determinando as melhores alternativas de oferta com base em princípios econômicos e técnicos (RAMOS, 2015; CAMPOS, 2013; NUZZOLO *et al.*, 2011; MARRA, 1999; GONÇALVES, 1992). Tedesco (2008) reforça ainda que o modelo de Quatro Etapas tem por finalidade permitir a previsão da demanda de tráfego por meio da determinação e da análise da quantidade e do tipo de viagens entre zonas em uma determinada região. Às quatro etapas consideradas pelo modelo são: geração, distribuição de viagens, divisão modal e alocação de tráfego (MAGALHAES, 2008).

A etapa de Geração de Viagens, na qual está inserido o escopo deste trabalho, tem por finalidade a estimativa do número total de viagens produzidas e atraídas, sem necessariamente considerar as conexões entre origens e destinos. Assim, a geração de viagens é traduzida como a soma entre a produção e atração de viagens em uma mesma zona (CAMPOS, 2013).

Mesmo diante da importância da modelagem da demanda, a geração de viagens no contexto da movimentação de mercadorias ainda não tem sido explorada com maior profundidade em investigações científicas e técnicas (GASPARINI, CAMPOS e D'AGOSTO, 2010).

Conforme Souza, Silva e D'Agosto (2010), Holguín-Veras, Thorson (2000) e Marra (1999), apesar de sua pequena abrangência, os estudos que promovem o desenvolvimento de modelos para previsão da demanda em sistemas de transporte de carga baseiam-se, em sua maioria, em viagens de veículos de carga e em volume de carga. Ainda conforme Marra (1999), ambos os conceitos de modelagem têm sua importância para a previsão da demanda. A escolha de qual tipo de modelo utilizar dependerá não só dos objetivos do planejador, mas da disponibilidade e consistência de dados (KULPA, 2014; MARRA, 1999).

Vale ressaltar, que para se cumprir o objetivo da etapa de Geração de Viagens, são aplicados métodos matemáticos que buscam uma relação entre o número de viagens existentes (variável dependente) e os vários fatores que induzem as viagens (variáveis explicativas ou independentes), viabilizando, assim, as estimativas indispensáveis à área de planejamento de transportes (FERREIRA e SILVA, 2016).

Modelos baseados em viagens de veículos de carga têm foco no fluxo de veículos, sua origem e destino. Tais modelos pressupõem que o modo de transporte já foi previamente definido e, assim, não se faz necessário efetuar a fase de Divisão Modal do modelo de Quatro Etapas, passando-se da Distribuição de Viagens para a Alocação de Tráfego (SOUZA, SILVA e D'AGOSTO, 2010).

Ogden (1992) e Holguín-Veras e Thorson (2000) ressaltam que os modelos baseados em viagens de veículos de carga apresentam algumas vantagens, como, por exemplo, a relativa facilidade em se obter dados de tráfego para a realização de planejamento de transportes. Outra vantagem desse tipo modelo, trata-se da disponibilidade cada vez maior de Sistemas Inteligentes de Transporte (ITS), que são capazes de seguir os movimentos dos veículos e assim construir dados a respeito do tráfego. Por fim, pode-se mencionar, como vantagem desses modelos, as viagens vazias, como, por exemplo, as de retorno, que já são consideradas nos referidos modelos.

Modelos baseados em volume de mercadorias (carga) têm foco na quantidade de carga transportada quantificada em unidade de peso, normalmente em toneladas. Tais modelos possibilitam obter com maior precisão os mecanismos que direcionam a movimentação das cargas, função atributo, como forma e peso por unidade (SOUZA, SILVA e D'AGOSTO, 2010). Uma dificuldade do uso deste modelo é a obtenção de dados em termos de fluxos de mercadorias, pois, as empresas não os tornam públicos por questões comerciais (COMI *et al.*, 2012).

Esses modelos têm como premissa que o sistema de transporte de carga é, sobretudo, estabelecido na circulação de mercadorias. Dessa forma, a modelagem deve ser fundamentada na circulação de bens, sendo fundamental o estudo dos fluxos de cargas, abordando: (i) sua geração, por local ou empresa, (ii) seus mercados, por uso do solo ou localização e (iii) por gerenciamento logístico no que diz respeito aos meios de transporte, armazenagens, dentre outros (MARRA, 1999).

Existem muitos modelos de geração de viagens que se diversificam de acordo com o tipo de empreendimento, porte, localização, atividade desempenhada, dentre outros atributos (PORTUGAL *et al.*, 2012). Visando identificar as principais variáveis consideradas em

modelos de demanda no estado da arte, fundamentando as decisões metodológicas neste trabalho, fez-se uma revisão sistemática da literatura com o intuito de discutir os métodos matemáticos e as variáveis mais utilizadas para desenvolver modelos de geração de carga e de viagens de carga.

Kitchenham e Chartes (2007) destacam que a revisão sistemática consiste em um meio de identificar, avaliar e interpretar um grande número de pesquisas disponíveis que sejam relevantes para uma determinada questão de pesquisa. Esse modo de revisão se destaca por ser realizado de maneira estruturada, seguindo passos que permitem retirar informações necessárias a pesquisa, em grandes quantidades. No APÊNDICE A é apresentado o procedimento que foi adotado para a realização da revisão sistemática.

Para o desenvolvimento da revisão do estado da arte neste trabalho, foram utilizados as fontes de pesquisa: *Science Direct* e *Web of Knowledge*, o Portal de Periódicos da CAPES, Google Acadêmico, *Research Gate*, *Transportation Research Board (TRB)*, *ITE – Institute of Transportation Engineers* e anais de congressos. As palavras-chave utilizadas nos campos de busca foram: geração de viagens de carga, *freight trip generation*, *freight trip generation models*, modelos de geração de viagens de carga e *truck trip generation model*. Os tipos de trabalho utilizados foram: artigos científicos, relatórios técnicos, livros, dissertações, teses e anais.

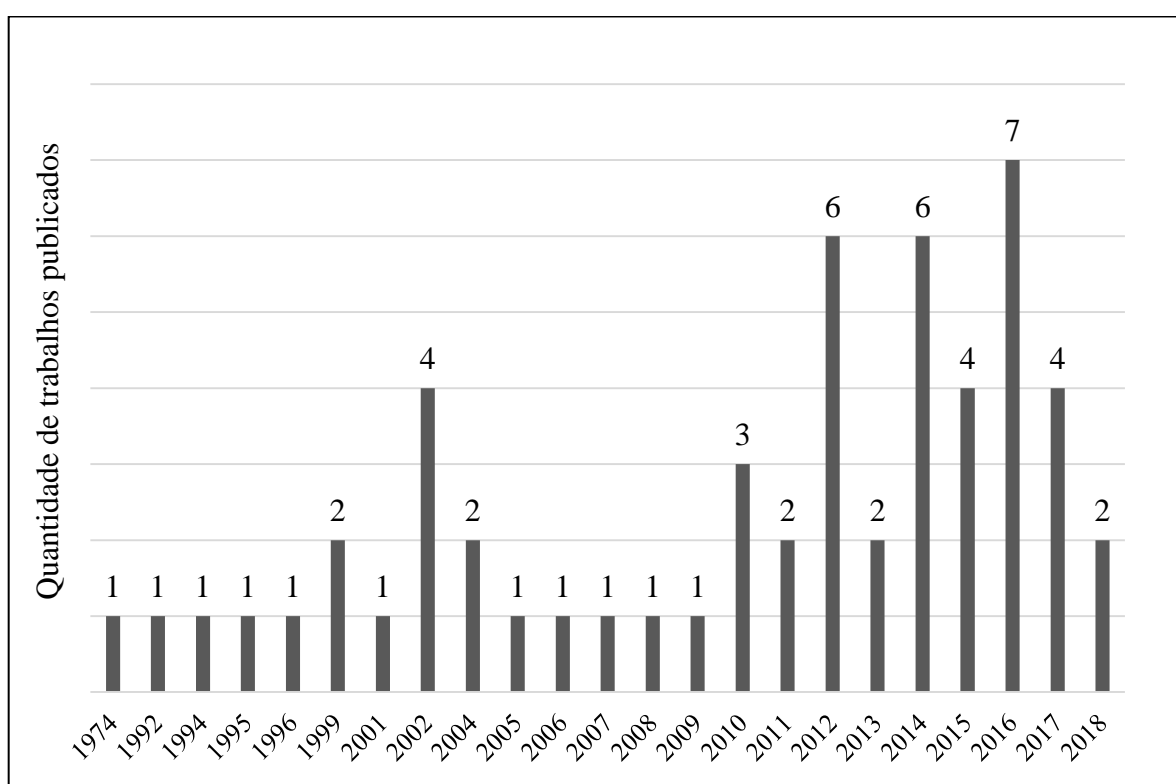
Após selecionados os trabalhos pesquisados pelas palavras-chave definidas, foram inclusos na revisão sistemática os trabalhos nos quais autores apresentavam a modelagem matemática e as variáveis adotadas para o desenvolvimento dos modelos. Nessa etapa foram excluídos os trabalhos que não apresentavam modelos matemáticos e trabalhos apenas de revisão bibliográfica. Adicionalmente, foi realizada a estratégia “bola de neve” (*snowballing*), que consiste em pesquisar trabalhos a partir das referências dos artigos obtidos.

Por meio das palavras de busca, foram identificados 78 trabalhos, selecionados 63 e incluídos 55 que foram considerados na revisão sistemática da literatura.

Nos trabalhos analisados nesta revisão, foram obtidas informações sobre variáveis dependentes, independentes e os métodos matemáticos empregados para gerar os modelos de geração de viagens de carga.

Na Figura 2.1 é apresentado o número de trabalhos publicados sobre geração de viagens de carga entre os anos de 1974 a 2018 e que fizeram parte dessa pesquisa. Durante a pesquisa não foram encontrados trabalhos publicados antes de 1974. Podem-se destacar os anos de 2002, 2015 e 2017 com 4 publicações em cada ano, 2012 e 2014 com 6 e 2016 com 7 publicações.

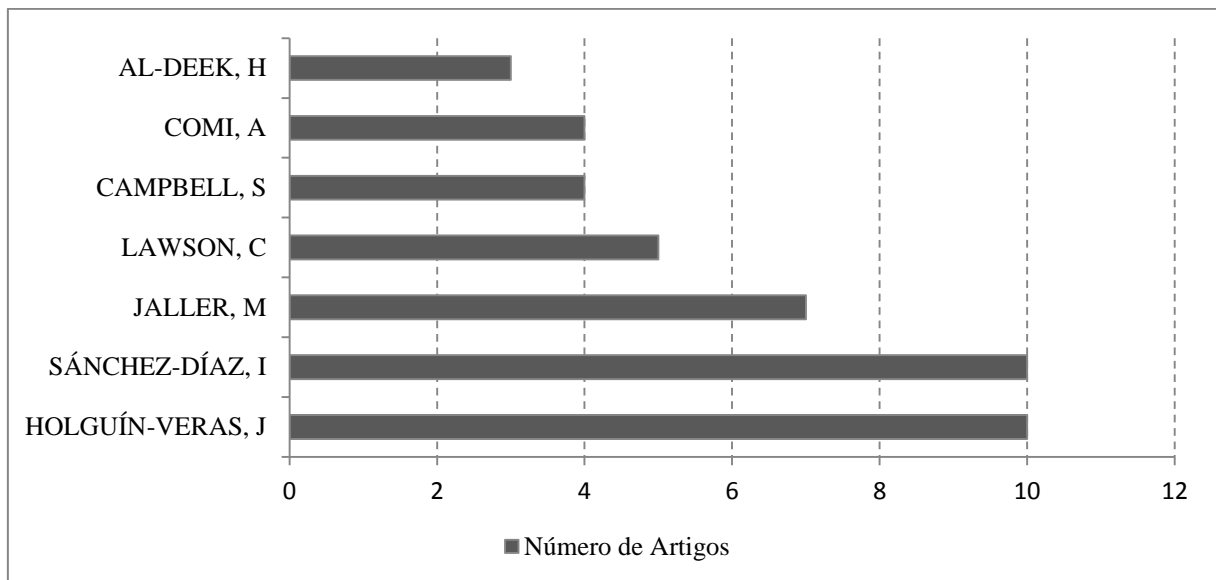
Figura 2.1: Número de trabalhos publicados entre 1974-2018.



As 55 publicações incluídas nesta revisão sistemática foram escritas por 102 autores de diversas nacionalidades. Os sete autores com maior quantidade de trabalhos publicados são apresentados na Figura 2.2. Destacam-se José Holguín-Veras e Iván Sánchez-Díaz que publicaram dez artigos cada um, sendo que trabalharam juntos em oito dos dez estudos. Miguel Jaller publicou sete trabalhos sendo que seis deles tiveram a participação do José Holguín-Veras e Iván Sánchez-Díaz. Catherine Lawson publicou cinco artigos, todos também

com a participação do José Holguín-Veras e Iván Sánchez-Díaz. Shama Campbell e Antonio Comi publicaram quatro artigos cada autor, sendo que todos os trabalhos da Shama Campbell foram com José Holguín-Veras e Iván Sánchez-Díaz. Haitham Al-Deek publicou três trabalhos, 17 autores publicaram dois artigos cada e os demais autores publicaram um artigo cada.

Figura 2.2: Autores com maior número de publicações.



Para sintetizar os dados e simplificar a análise dos resultados, as variáveis dependentes e independentes encontradas na literatura foram categorizadas e apresentadas na Tabela 2.1.

Tabela 2.1: Categorias de síntese das variáveis dependentes e independentes.

#	Categoria	Tipo de variável	Variáveis
1	Número de viagens	Dependente	Frequência ou número de entregas. Número de veículos ou número de viagens.
2	Volume de Carga		Quantidade de carga produzida ou atraída. Volume de mercadorias.
3	Área	Independente	Área construída, bruta, cultivada, do depósito, do estabelecimento, do terminal, do território, área de vendas, tamanho da cidade, região ou zona geradora; Largura da frente do estabelecimento; Localização.
4	Carga		Volume (Barris, contêineres, madeira, entre outros.) importados ou exportados; Demanda final do tipo de mercadoria em valor; Exportação doméstica e nacional; Importação doméstica e nacional; Fluxos de mercadorias; Oferta total da mercadoria; Valor da mercadoria.
5	Domicílios		Densidade residencial; Número de domicílios.
6	Emprego		Emprego total e por setor (referentes à zona).
7	Frota		Frota operacional da empresa; Número de veículos na frota; Número de veículos utilizados para a entrega; Número médio de veículos para entregas domiciliares.
8	Funcionários		Empregados; Funcionários varejistas; Número de funcionários; Número de motoristas de caminhão.
9	Infraestrutura		Número de berços de contêiner; Número de portas nas docas; Quantidade de animais; Superfície dedicada à venda de frutas e legumes; Vagas de estacionamentos; Valor do terreno.
10	Número de empresas		Número de estabelecimentos, varejistas e empresas.
11	Operação		Características da empresa; Distância entre varejo e fornecedor; Fator de carga do veículo utilizado para a entrega; Intervalo de tempo entre as encomendas e as entregas; Linha de negócios; Número de dias de funcionamento; Número de entregas diárias por veículo; Número de entregas em uma semana; Número de fornecedores; Número de varejistas que fazem entregas ao domicílio; Número de vendas por ano; Número médio de entregas recebidas diariamente; Número médio de estabelecimentos visitados; Tamanho médio da entrega; Tipo de mercadoria; Volume médio de clientes; Horas trabalhadas; Distância percorrida.
12	População		Crescimento da população; Densidade populacional; Número de habitantes; Número médio de famílias que vivem em cada zona.
13	Renda		Renda familiar média, per capita e total.
14	Uso do solo		Índice de uso do solo; Tipo de uso do solo.
15	Outros		Capacidade elétrica (kW); Comprimento da estrada; Demanda elétrica diária (Toneladas de carvão); Influência do porto; Percentagem de variação dos preços do petróleo; Toneladas de carvão/kW; Idade média; PIB.

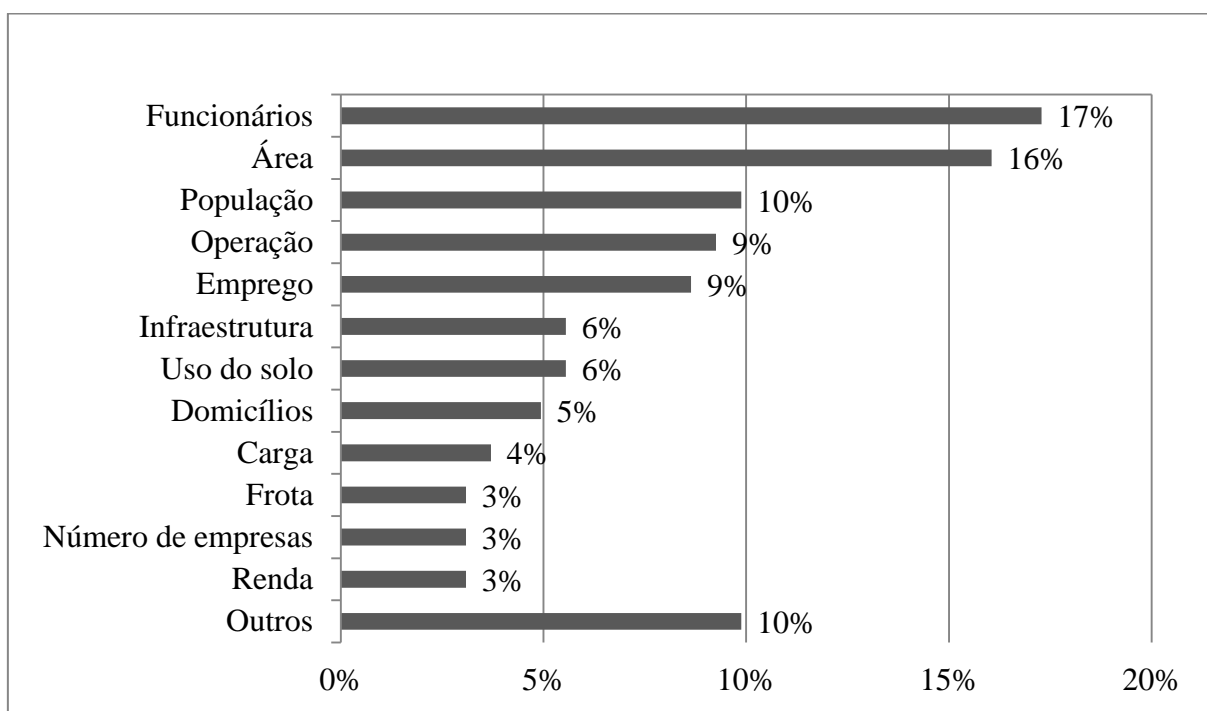
Em relação à variável dependente, 79% das publicações utilizavam a variável dependente, viagens de veículos de carga, enquanto 21% utilizavam a quantidade de carga movimentada.

Já às variáveis independentes mais utilizadas nos trabalhos pesquisados foram: (i) o número de funcionários (número de empregos ofertados por empreendimento ou setor de atividade) com 17% de utilização, (ii) a área com 16%, (iii) população com 10% e (iv) emprego (referente a zona) e (v) operação com 9% de utilização cada variável.

As variáveis, (i) número de funcionários e (iv) número de empregos, tratam, em essência, de um mesmo indicador, sendo que o primeiro refere ao número de empregados do empreendimento ou setor de atividade em questão, enquanto que o segundo refere ao total de empregos na região. Ao unir ambas as categorias, têm-se as variáveis presentes em 26% dos estudos analisados, o que demonstra que os indicadores de empregos são variáveis fundamentais para explicar a geração de viagens de carga. Essa técnica de agregar dados de funcionários com o de empregos também foi visto no trabalho do Ferreira e Silva (2016).

Na Figura 2.3 são apresentadas, graficamente, as frequências das variáveis independentes encontradas nos trabalhos pesquisados por meio da revisão sistemática:

Figura 2.3: Principais variáveis independentes.



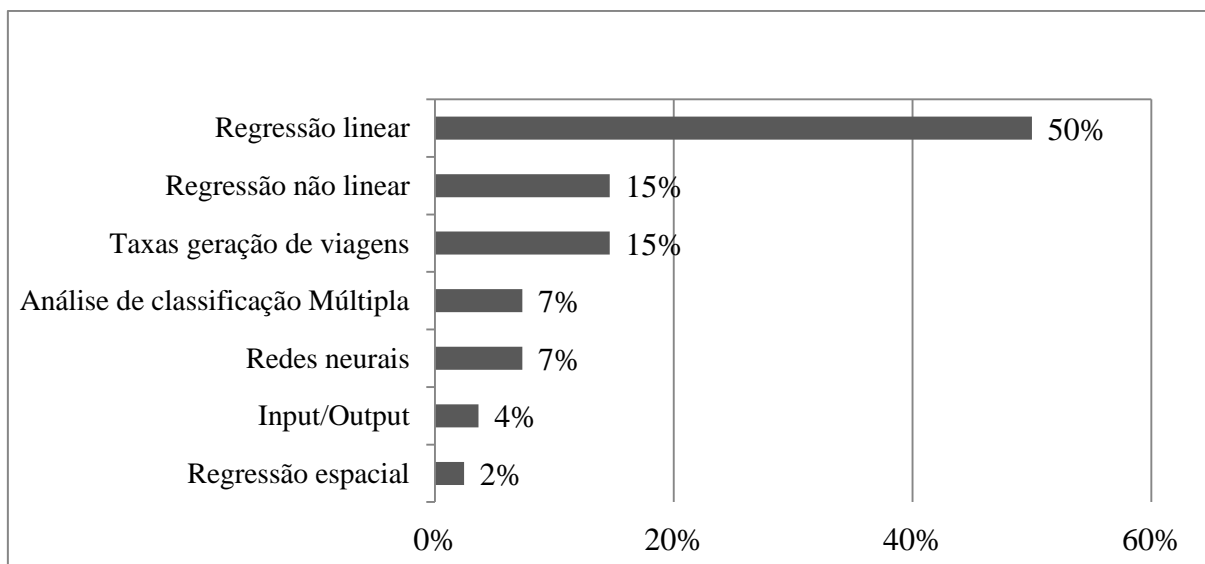


De forma geral, as variáveis independentes podem ser divididas em dois grupos: variáveis relacionadas às características dos empreendimentos ou setor de atividade, como, por exemplo, área, funcionários, frota, infraestrutura e operação; e as variáveis relacionadas a informações da região de estudo (dados territoriais), tais como, domicílios, população, emprego, renda e uso do solo (FERREIRA e SILVA, 2016). As variáveis referentes ao empreendimento estiveram em 63% dos trabalhos pesquisados enquanto que as variáveis referentes e à região de estudo estiveram presentes em 37% dos trabalhos.

Sobre os métodos matemáticos mais utilizados nos trabalhos revisados e apresentados na Figura 2.4, tem-se a predominância da regressão linear (método dos mínimos quadrados) (50%) e o método de taxas de viagens com 15% de utilização. Tal diagnóstico confirma a afirmação de Kulpa (2014), ao mencionar que tais modelos são os mais usados em estudos de geração de viagens de carga.

Outras técnicas utilizadas são regressão não linear, exponenciais, logísticas, polinomiais entre outras, análise de classificação múltipla e redes neurais, modelos *input/output*, e regressão espacial.

Figura 2.4: Principais modelos utilizados para demanda de carga.



Na próxima subseção são discutidos alguns trabalhos aqui quantificados. A forma de apresentação desses trabalhos seguiu a respectiva ordem cronológica de publicação.

## 2.2 Modelos de geração de fluxos e viagens de carga

Hutchinson (1974) citou que para a previsão de demanda, as equações de viagens realizadas por caminhão são o ponto principal a ser analisado. Nesse sentido, o autor desenvolveu um trabalho na área metropolitana da cidade de Toronto (Canadá) observando, por um dia, o movimento de caminhões em duzentas e quarenta indústrias manufatureiras. O autor ressalta em seu trabalho que é difícil estabelecer equações de geração de viagens confiáveis a partir de dados coletados em curtos períodos de tempo. Quanto ao universo pesquisado, o autor estratificou os tipos de caminhões analisados conforme o número de eixos, sendo que caminhões com dois eixos responderam por 70 a 80% do total dos movimentos dos veículos nas áreas urbanas. Com base ainda nessas observações, foi possível desenvolver uma série de equações de regressão linear múltipla, em que o número de caminhões privados foi considerado como variável independente e que, assim, explicou tanto a produção como a atração das viagens de caminhão. As equações apresentadas representam uma combinação das indústrias manufatureiras pesquisadas e a amostra analisada foi composta por polos geradores de viagens de empresas de alimentos e bebidas, jornais e fábricas de concreto pré-misturado. O autor ressalta que os modelos que apresentaram como variáveis independentes o volume de produtos manufaturados e número total de empregados em setores administrativos, não apresentaram ajuste satisfatório, analisando o coeficiente de determinação  $R^2$ . Pode-se observar que o modelo elaborado para a produção de viagens possui coeficiente de determinação com valor superior ao elaborado para a atração de viagens. Dos 8 modelos apresentados, apenas 2 apresentaram coeficiente de determinação maior que 0,7 (25%).

No trabalho de Ogden (1992), o autor desenvolveu sua abordagem metodológica por meio de dois conceitos de modelagem, a saber: (i) baseado em movimentação de mercadorias; e (ii) baseado em viagens por caminhões. Em seu trabalho, o autor apresentou modelos de demanda de carga para as cidades de Chicago, Vancouver e Melbourne. Na cidade de Chicago, a geração de veículos de carga foi estimada em função de variáveis específicas das atividades comerciais e foram correlacionadas com viagens de passageiros. Já em Vancouver e Melbourne, foram utilizadas as relações entre os veículos e as atividades, bem como outras variáveis socioeconômicas e de acessibilidade. Todos os modelos são específicos para a área onde eles foram calibrados e têm mostrado sérias deficiências na adequação para outros

contextos urbanos. Os modelos foram desenvolvidos utilizando regressão linear. Ainda segundo o autor, para o comércio varejista, a área construída por pavimento e o número de empregados são variáveis independentes adequadas; para as atividades industriais, o número de empregados é preferencial, considerando suposições relativas ao volume de carga produzido por cada empregado.

Tadi e Balbach (1994) analisaram características não residenciais relacionadas ao uso do solo para desenvolver modelos de geração de viagem de caminhão para cidade de Fontana, Estados Unidos. O objetivo do estudo foi determinar o nível de impacto do tráfego de caminhões em ruas locais e arteriais da cidade, bem como obter dados relevantes a serem utilizados em estudos de índices de tráfego e contribuir com uma base de dados atualizada para futuros dimensionamentos de infraestrutura. Para determinar as taxas de viagem de caminhões, leves e pesados, dos setores industriais estudados foram coletados dados de 21 locais diferentes, bem como informações sobre veículos de passageiros e caminhões separados de acordo com o número de eixos. A coleta de dados consistiu em realizar contagens manuais de 10 horas e contagens automáticas (por meio de máquinas) de 24 horas em cada um dos 21 locais. A variável independente utilizada nos modelos foi a área construída ( $\text{pés}^2/1000$ ) e a variável dependente foi o número de viagens de caminhões. O método utilizado foi a regressão linear, considerando viagens diárias. Os autores apresentaram as taxas segregando o dia da semana e os horários das viagens e concluíram que o tráfego de caminhões, geralmente, é maior nos períodos de pico da tarde, sendo que as viagens de caminhão correspondem a 30% do volume total diário de veículos.

Wegmann *et al.* (1995) apresentaram as características dos sistemas de transportes urbanos de carga pela *Federal Highway Administration*, em 1995. No trabalho é apresentada uma revisão completa dos dados e informações relativas às movimentações urbanas de mercadorias e o seu planejamento. No trabalho são apresentadas várias seções relacionadas a taxas de viagem e as movimentações de caminhões, juntamente com seções relacionadas a terminais intermodais ferroviários, carga aérea e portos. As taxas de viagens, em termos de toneladas por empregado, foram calculadas para seis tipos diferentes de geradores. A variável dependente utilizada no trabalho foi o número de veículos e como variáveis independentes foram

utilizadas o número de emprego, população, domicílios, área bruta e uso do solo. Os modelos desenvolvidos foram regressão linear e a geração de taxa de viagens.

*Cambridge Systematics* (1996), por meio do *Quick Response Freight Manual*, manual lançado pelo departamento de transportes dos EUA em 1996, forneceu técnicas simples e parâmetros para o desenvolvimento de tabelas de viagem de veículos comerciais. As taxas de geração de viagens de caminhões são estimadas a partir do número de empregos e área construída dos setores associados aos embarques de mercadorias. As taxas fornecidas pelo manual foram retiradas de uma pesquisa em Phoenix, Arizona. Depois de calcular as viagens de caminhão a partir dos dados de emprego, é possível construir uma tabela de viagem de caminhão e atribuir as viagens seguindo os modelos de quatro etapas convencionais UTPS (*Urban Transportation Planning System*). Esse método, apesar de simples, apresenta um problema recorrente que se refere à padronização de parâmetros, como taxas de geração de viagens de caminhão, que não podem ser transferidos sem adaptações às características locais de cada região. Assim como em Wegmann *et al.* (1995), os modelos considerados foram a regressão linear e a geração de taxa de viagens.

No trabalho de Black (1999), foram apresentados modelos de geração de viagem de carga para o estado de Indiana nos Estados Unidos. O autor analisou os fluxos de entrada, saída e entre os 92 municípios do estado, além de um total de 145 nós que representam municípios de estados vizinhos em decorrência do fluxo de bens e materiais que estão passando por Indiana, uma vez que o tráfego de *commodities* não tem origem nem destino, especificamente, no estado. O autor identificou um conjunto de 18 grupos de mercadorias e desenvolveu um modelo para cada setor. Os dados utilizados na pesquisa são do “Censo de Transporte” de 1977 e toda a discussão dos fluxos de mercadorias através, dentro e fora de Indiana pertence aos dados da pesquisa de fluxo de mercadorias (*Commodity Flow Survey*) de 1993. Os modelos foram gerados por meio de regressão linear utilizando como variáveis explicativas o número de empregos por setor, população total, número total de empregos, receita bruta recebida na agricultura e o número total de empregos na indústria. A variável dependente foi o volume de carga atraída e produzida por setor. Uma avaliação geral dos modelos foi realizada usando as atribuições de todos os *commodities* para a rede rodoviária de Indiana e o resultado foi um modelo com coeficiente de determinação ajustado ( $R^2$ ) de 0,435.

Marra (1999) apresentou um estudo sobre movimentação urbana de cargas buscando identificar padrões de demanda por mercadorias, utilizando dados da cidade de Campinas, Brasil. O autor desenvolveu modelos de geração de viagem de carga usando dados relacionados ao volume de carga transportada. Para aquisição dos dados foi realizada uma pesquisa de campo em domicílios visando determinar a demanda mensal por mercadorias, relacionando tal demanda com indicadores socioeconômicos da população. A pesquisa apresentou, também, os tipos de uso do solo que implicam em maior geração de viagens de cargas urbanas, as principais mercadorias em circulação e os veículos de carga mais utilizados. A variável dependente utilizada no modelo foi o volume de carga transportada por mês e as variáveis independentes foram o nível de renda nos bairros, número de moradores por empregados no domicílio e a área do imóvel. Na correlação das variáveis o autor utilizou regressão linear, logarítmica, exponencial e polinomial. Os resultados indicaram que o modelo de regressão linear correlaciona bem o volume de cargas demandados em cada bairro e suas respectivas áreas médias dos imóveis e seu número médio de moradores ou empregados, atingindo um coeficiente de determinação ajustado ( $R^2$ ) de 0,90. No caso da correlação entre a demanda mensal de cargas e o nível de renda médio de cada bairro, concluiu-se que a curva exponencial foi a que melhor se ajustou aos valores, com índice de  $R^2$  igual a 0,89.

Al-Deek (2001) expôs duas abordagens para o desenvolvimento de modelos de geração de viagem de carga para caminhões que abastecem os portos no estado da Flórida nos Estados Unidos. O autor obteve dados do Porto de Miami e do Porto de Jacksonville e desenvolveu modelos utilizando análise de regressão e redes neurais artificiais (*Backpropagation Neural Network* – BPNN). Tais modelos são usados para estimar os níveis de tráfego de caminhões de carga em movimento (entrando e saindo) dos portos marítimos. As variáveis dependentes para os modelos de regressão e de redes neurais foram os volumes diários de caminhões pesados de entrada e de saída, e as variáveis independentes foram o número importado e exportado de contêineres por dia. Com base no caso do Porto de Miami, verificou-se que o modelo de redes neurais é mais preciso do que o modelo de regressão. No entanto, o modelo BPNN requer um banco de dados que demanda mais investimento. O modelo de atração na análise de regressão (caminhões de saída) apresentou  $R^2$  igual a 0,784. O modelo de produção (caminhões de entrada) apresentou um  $R^2$  igual a 0,686. Verificou-se que os principais fatores

que afetam o volume de tráfego de caminhões são a quantidade e direção dos navios cargueiros, o tipo de mercadoria e o dia da operação. Em resumo, os resultados encontrados utilizando o modelo de redes neurais foram significativamente precisos para ambos os portos da Flórida.

Allen (2002) expôs modelos regionais de previsão de caminhões para o Conselho Metropolitano de Baltimore nos Estados Unidos. Segundo o autor, as agências responsáveis pelo transporte demonstraram interesse em renovar os dados sobre movimento de mercadorias como parte do planejamento intermodal e decidiu atualizar seus modelos sobre caminhões, já que utilizavam estimativas de viagens de caminhões baseadas em dados com mais de 30 anos. A pesquisa utilizou dados de contagens de caminhões. O número de viagens de caminhões médios e pesados foram consideradas como variáveis dependentes. O número de empregos no setor industrial e no setor de varejo, a oferta de empregos em escritórios e o número de residências foram consideradas como variáveis independentes. O resultado foi um processo que apresentou sensibilidades razoáveis para as principais variáveis de entrada (número de emprego e residências) e a validação apresentou elevada precisão quando comparado com os dados de entrada.

No trabalho de Iding *et al.* (2002) os autores construíram modelos que consideraram diferentes tipos de indústrias. Para tanto, utilizaram dados reais de geração de viagens de veículos de carga, obtidos em pesquisa de campo realizada nas indústrias e empreendimentos comerciais da Holanda, totalizando uma amostra de 1529 empresas. Os modelos foram desenvolvidos por meio de regressão linear e os resultados indicaram que às duas principais variáveis independentes que explicam a geração de viagens de caminhões são a área do estabelecimento e o número de empregados para as empresas de vários setores da indústria foco do estudo. Porém, há importância de uma ou outra variar conforme o tipo do estabelecimento ou indústria. Além disso, os autores concluíram que o tipo de estabelecimento também é uma variável muito importante na previsão de viagens de veículos de carga.

O estudo de Melo (2002) e Campos e Melo (2004) apresentou modelos de viagem de carga existentes na literatura e desenvolveu modelos de geração de viagens de carga para

estabelecimentos comerciais e empresas de transporte na cidade do Rio de Janeiro, Brasil. Os dados utilizados foram obtidos a partir de questionários aplicados em 78 estabelecimentos divididos em seis segmentos (supermercados; vestuário; comércio varejista; bares e restaurantes; material de construção; combustível) no setor comercial. Para as empresas de transporte foram aplicados questionários em 20 empresas divididas em 4 segmentos (carga geral; combustível; carga própria; alimentação). O autor utilizou como variável explicativa a área dos estabelecimentos comerciais, relacionando essa variável com o número de viagens de veículos de carga (variável dependente). Para as empresas de transporte foram consideradas as variáveis independentes: frota da empresa, horas trabalhadas e distância percorrida. Foi utilizada regressão linear no desenvolvimento dos modelos e os resultados encontrados apresentaram modelos com alto poder de predição.

Uma evolução do modelo multi-etapa foi proposta por Russo e Comi (2002). O modelo parte da demanda do consumidor final e permite que o fluxo estimado por trecho seja dividido entre fluxo de passageiros para compras e fluxo de veículos para a movimentação de bens. O modelo foi proposto para as cidades italianas e como variável dependente foi utilizado o volume de carga. As variáveis independentes consideradas foram o número de montadoras na zona e o número médio de famílias que vivem em cada zona, possibilitando a geração de taxas de viagens para relacionar as variáveis.

Klodzinski *et al.* (2004) desenvolveram um aplicativo Java para executar modelos de geração de viagens para caminhões em portos marítimos baseados nos modelos desenvolvidos pelo Instituto de Sistemas de Transporte da Universidade da Florida Central. O aplicativo permitiu executar dois modelos: um modelo selecionado MLFNN (modelo elaborado por meio da técnica de Rede Neural Acíclica Multicamada - *Multilayer Feedforward Neural Network*, com o uso de Algoritmo de Retro-propagação - *Backpropagation Neural Network* – BPN) para o porto de Everglades e um híbrido do modelo desse porto, que permite ao usuário treinar o modelo de Redes Neurais Artificiais (RNA) antes da sua execução para os dados desejados. Embora o modelo treinável tenha maior complexidade, possui o benefício de produzir uma rede com um maior grau de precisão.

Sarvareddy *et al.* (2005) apresentaram uma metodologia para a construção de um modelo de geração de viagem de caminhão usando redes neurais artificiais a partir de dados de mercadorias de embarcações. O método foi desenvolvido e aplicado com sucesso em cinco portos da Flórida, Estados Unidos. O algoritmo de rede neural *Backpropagation Neural Network* (BPNN) foi usado no projeto. Embora a metodologia tenha sido sólida, os autores desenvolveram um novo modelo para cada um dos cinco portos estudados. As variáveis explicativas relacionadas aos fatores operacionais dos portos (número de contêineres importados e exportados) foram necessárias na maioria dos modelos para contabilizar mercadorias armazenadas no porto antes da exportação ou retiradas após a importação. O modelo BPNN desenvolvido para o Porto Canaveral foi validado com sucesso no nível de confiança de 95% com dados coletados em campo.

Wisetjindawat *et al.* (2006) utilizaram modelos de regressão com variáveis independentes, como o número total de funcionários e a área útil para derivar o consumo de *commodities* urbanas por empresa na Região Metropolitana de Tóquio. Os autores consideram o fluxo de *commodities* seguindo uma abordagem *top-down* que explica o movimento de *commodities* por meio da interação entre vários agentes de carga na cadeia de suprimentos. O modelo de escolha discreta espacial distingue entre carregadores, receptores e as relações entre eles. Seu modelo aborda o movimento de mercadorias (geração e distribuição), dimensionamento de remessa, escolha da transportadora e roteamento, e atribuição de tráfego dos veículos transportadores. A variável dependente considerada no trabalho foi a quantidade de carga produzida ou atraída e o modelo adotado foi a regressão linear.

Silva e Waisman (2007) desenvolveram um modelo de geração de viagens de caminhões para bares e restaurantes no bairro da Mooca, em São Paulo, Brasil. As informações sobre o comportamento das viagens e a caracterização do uso do solo foram coletadas por meio de pesquisa de campo em trinta estabelecimentos comerciais. As variáveis explicativas utilizadas foram área do estabelecimento, em metros quadrado, e a quantidade de funcionários. Nos 30 estabelecimentos pesquisados, a área variou de 60 a 500 m<sup>2</sup> e de 2 a 17 funcionários por estabelecimento. Os dados permitiram verificar que, de modo geral, a variável dependente (quantidade média semanal de viagens de caminhões) está relacionada com as variáveis



independentes. Os autores ressaltam que os resultados obtidos se restringem à região e ao tipo de atividade/uso de solo estudado.

Kawamura *et al.* (2008) identificaram as variáveis que podem ser utilizadas para o desenvolvimento de modelos de previsão de geração de viagens de veículos de carga. Os dados utilizados nesse estudo foram obtidos de três maneiras: por meio de questionário de pesquisa, de visitas a lojas e por contatos telefônicos. Foram obtidas informações detalhadas sobre a quantidade de mercadorias entregues nas lojas e o número de viagens de entrega por semana por loja. No total, foram pesquisadas 426 lojas de 9 redes nacionais de varejo, 5 redes de móveis e 4 redes de calçados. Foi utilizada a técnica de regressão logit binomial para o desenvolvimento dos modelos. Os autores concluíram que a geração de viagens de veículos de carga pode variar consideravelmente dentro do mesmo setor de varejo dependendo da estratégia da cadeia de suprimentos adotada. Além disso, foi observado que ao contrário do que é comumente utilizado na maioria de estudos do tipo, o número de empregados e a área total do estabelecimento como variáveis independentes não retornaram bons resultados quanto aos modelos desenvolvidos.

Bastida e Holguín-Veras (2009) analisaram comparativamente duas abordagens para a estimativa de geração de viagens de veículos de carga: o método dos mínimos quadrados (*ordinary least squares* – OLS) e a classificação cruzada. Os modelos OLS foram utilizados para estimar a relação entre o número de veículos de carga gerados e as variáveis independentes. A utilização da técnica de classificação cruzada, por sua vez, teve como objetivo identificar uma estrutura de classificação com uma boa representação do processo de geração de viagens. Para tanto, utilizou-se a análise de classificação múltipla (*multiple classification analysis* – MCA) para identificar os grupos de variáveis independentes que explicassem a geração de viagens. Esta técnica é descrita mais detalhadamente por Ortúzar e Willumsen (1994). A base de dados utilizada foi obtida por meio de pesquisas em estabelecimentos comerciais localizados nas regiões de Manhattan e Brooklyn, em Nova York. Os resultados mostraram que as técnicas de OLS e a MCA retornam os mesmos resultados de ajuste quanto às variáveis mais importantes que explicam a geração de viagens de carga. Para o segmento da indústria, o número de vendas totais por ano e o tipo de mercadoria do estabelecimento se mostraram importantes fatores para a estimativa de geração de viagens de veículos de carga.

No estudo de Crocco *et al.* (2010), os autores propuseram um sistema de modelagem para simular a movimentação urbana combinando viagens de pessoas e fluxos de mercadorias para *commodities*. O estudo concentrou esforços na análise de consumo de famílias residentes nas cidades de Cosenza, Rende e Castrolibero, na Itália, para diferentes categorias de produtos e características dos movimentos a eles relacionados. Como variável dependente foi considerado o volume de carga. As variáveis independentes foram a população residente e o número de domicílios. Os modelos foram gerados por meio de regressão linear.

No estudo do Gasparini *et al.* (2010), os autores analisaram a movimentação de veículos de carga para *shopping centers* e supermercados. Os autores desenvolveram modelos de geração de viagem de carga, para os segmentos citados, na cidade do Rio de Janeiro, Brasil. Para obtenção de dados sobre a movimentação de veículos de carga, em termos de volume, tipos de veículos e duração das viagens, assim como das características dos empreendimentos, foi realizada uma pesquisa de campo na qual foi realizada entrevista em 21 supermercados e 7 *shopping centers*. Para ambos os segmentos, as variáveis dependentes foram: volume total de viagens no período de “pico”, volume total de viagens no período “fora pico” e volume total de viagens. Já as variáveis independentes foram: área construída, volume médio de clientes, área bruta locável, número de empregados, vagas de estacionamento e área de vendas (para *shopping centers*). Para supermercados foram: volume médio de clientes, número de empregados e vagas de estacionamento. Os modelos foram desenvolvidos por meio de regressão linear e geração de taxa de viagens. As taxas de viagens para cada mil metros quadrados de área construída em *shopping centers* foram entre 0,57 e 0,89 nos dias de pico. No caso dos supermercados, observou-se também nos dias de pico, taxas de viagens de: 1,14 (hipermercados), 1,71 (superlojas) e 3,04 (convencionais).

Silva e Goldner (2010) objetivaram determinar os padrões de viagens e taxas de geração de viagens de automóveis e caminhões para três indústrias localizadas na cidade de Tubarão, Santa Catarina, Brasil. A pesquisa foi realizada entre os meses de março e junho de 2008, que consistiu em entrevistas com a administração destas indústrias e a aplicação de questionários a uma amostra estratificada de funcionários e caminhoneiros. Também foram realizadas contagens volumétricas de tráfego nos principais acessos. Da análise dos resultados, obteve-se a movimentação da entrada e saída de veículos em geral e o percentual do pico horário

correspondente. Mediante o somatório de fluxos, foi possível a elaboração das taxas de geração de viagens em função da área total construída e do número de funcionários. Também foi realizada uma comparação das taxas e volumes de viagens propostas na pesquisa com os modelos existentes na literatura referente ao tipo de PGV estudado. Para essa comparação, apenas o volume diário proposto para caminhões apresentou resultados superiores aos dos modelos apresentados na literatura. Como variável dependente foi utilizado o número de viagens e para variáveis independentes foram utilizadas a área total construída e o número de funcionários. O método utilizado foi a taxa de geração de viagens.

No estudo de Chu (2011) foi apresentado um método empírico para estimar viagens internas e externas de caminhões para um porto principal em Taiwan, na República da China. O autor apresentou um modelo de geração de viagem de carga para o porto internacional de Kaohsiung (KHH), o oitavo porto de contêineres mais movimentado do mundo em 2007. O porto KHH está localizado no sudoeste de Taiwan na região metropolitana de Kaohsiung, que tem uma população de mais de 1,5 milhões de habitantes. De agosto a outubro de 2009 foi realizada uma pesquisa na principal estrada de acesso ao porto. Foram coletados dados de movimentação de caminhões na área metropolitana de Kaohsiung. Uma amostra válida de 6.000 motoristas de caminhões pesados completaram o questionário. O autor utilizou nove variáveis independentes para explicar a geração de viagens de caminhões no porto, dentre elas seis são relacionadas a atributos da atividade regional de carga, duas a atributos de crescimento econômico e uma a um atributo relacionado a desastres naturais. O autor desenvolveu e comparou modelos de regressão múltipla, modelos de série temporal, modelos de regressão linear, modelos de regressão quadrática, modelos exponenciais e modelos de redes neurais. Nesse artigo foram comparados os modelos de predição desenvolvidos e os resultados mostraram que o modelo de rede neural obteve melhor desempenho de previsão do que as abordagens de regressão e de séries temporais.

Novak *et al.* (2011) investigaram a aplicação de modelos de regressão linear e espacial para prever a geração de viagens de carga em escala nacional nos Estados Unidos. Os autores apresentaram uma visão sobre diferentes técnicas de transformação de variáveis, avaliaram o uso de variáveis de regressão espacial e aplicaram a regressão geograficamente ponderada. A fonte de dados utilizada foi a “Pesquisa de Fluxo de Mercadorias” (*Commodity Flow Survey* -

CFS). A variável dependente considerada nesse estudo foi o número de viagens de veículos de carga. As variáveis independentes foram a população, o número de empregos por setor e o total de empregos. Além da população e do emprego, considerou-se a hipótese de que a geração de viagens de carga para uma determinada zona de tráfego está relacionada à sua localização espacial e aos atributos de rede de transporte. Os autores concluíam que o modelo de regressão espacial tem melhor desempenho que os modelos de regressão linear para geração de viagens de carga a nível nacional.

O estudo de Campbel *et al.* (2012) comparou dois sistemas de classificação de indústrias diferentes usados na modelagem de geração de viagens (*freight trip generation* - FTG), ou seja, classificação industrial padrão (*Standard Industrial Classification* - SIC) e o sistema norte americano de classificações da indústria (*North American Industry Classifications System* - NAICS). A análise apresenta diferentes enfoques nos tipos de modelos gerados e na capacidade dos modelos de replicar os dados de entrada por setor industrial. Os modelos são estimados usando técnicas de mínimos quadrados ordinários (OLS) e análise de classificação múltipla (MCA). Na análise são utilizadas duas amostras da área metropolitana de Nova York, receptores e transportadoras, para determinar qual dos dois sistemas de classificação industrial fornece melhores resultados quando usado na modelagem de FTG de atração e produção de carga. O estudo sugere que a SIC seria a mais apropriada para modelar o FTG.

No estudo de Holguín-Veras *et al.* (2012), os autores analisaram a influência do tamanho do empreendimento e a validade das taxas constantes de geração de viagem de carga (GVC). Os autores processaram um conjunto de dados com cerca de 400 transportadores e 400 receptores na área metropolitana de Nova York, Estados Unidos (tanto no Estado de Nova York quanto em Nova Jersey). O processo de modelagem foi realizado com foco em modelos lineares simples baseados no emprego como única variável independente, pois, poderiam ser comparados diretamente com as taxas de GVC constantes amplamente utilizadas. O método dos mínimos quadrados ordinários foram usados para estimar os modelos. A produção de viagens de carga (viagens de caminhão feitas por dia) e a atração de carga (viagens de entregas recebidas por dia) foram as variáveis dependentes. Os estabelecimentos foram classificados por códigos padrões de classificação industrial (*Standard Industrial Classification* – SIC) e agrupados em oito setores. Os resultados questionam as práticas de

modelagem da demanda de carga há muito estabelecidas e sugerem a necessidade de um exame abrangente dos dados de geração de viagem de carga para garantir que os modelos sejam validados e os correspondentes procedimentos de agregação sejam adequados. Deve-se analisar geração de carga (GC) e geração de viagem de carga (GVC) de modo diferente, uma vez que GC se baseia em decisões logísticas, enquanto GVC é determinado pela economia (produção e consumo).

O estudo de Ibeas *et al.* (2012) abordou a questão da coleta de dados, modelos e métodos para a investigação da demanda por transporte de carga urbana e as dificuldades relacionadas aos custos e ao fato de que poucos dados estão disponíveis frequentemente. O principal objetivo da pesquisa é comparar a demanda de transporte de mercadorias em duas cidades europeias (Roma e Santander), de modo a destacar quais similaridades e diferenças dependem de alguns fatores e demonstrar que existem muitos padrões diferentes de distribuição urbana que precisam ser considerados. A análise é baseada em algumas pesquisas semelhantes realizadas nas cidades em investigação, que também são muito diferentes em padrões espaciais e econômicos. Além disso, as pesquisas disponíveis em Roma e Santander foram a base para a calibração também em Santander de um sistema de modelagem para previsão de demanda de carga urbana em Roma. Dessa forma, a análise do modelo e da transferibilidade de parâmetros também é investigada. A variável dependente usada no estudo foi o volume de carga. Como variáveis independentes foram utilizadas o número total de empregados, funcionários varejistas e população residente. Os modelos foram gerados por meio de regressão linear.

Jin *et al.* (2012) desenvolveram um estudo para o estado de Utah, Estados Unidos. Eles utilizaram dados de fluxo de mercadorias disponíveis para o público através da Internet e encontraram uma boa correlação entre o fluxo de mercadorias e o tipo agregado de uso do solo. Dados que descrevem as características de uso particular do solo, como dados de emprego, dados de padrão de negócios e dados salariais foram utilizados. Os autores desenvolveram dois modelos usando regressão múltipla. Como variáveis independentes para o modelo I, foram utilizadas 25 classificações de emprego por setor. Os fatores relacionados ao emprego foram utilizados pelos modelos de previsão de viagens de carga em todo o estado. Os fluxos de mercadorias produzidos, atraídos e distribuídos em Utah tornaram-se variáveis

dependentes. No modelo II, 68 fatores contribuintes foram utilizados como variáveis independentes, incluindo 21 variáveis relacionadas a estabelecimentos, 11 variáveis relacionadas ao número de empregos, 11 relacionadas a salários e 25 variáveis relacionadas a empregos em setores específicos. Os fluxos de mercadorias produzidos, atraídos e distribuídos em Utah foram variáveis dependentes no Modelo II, assim como no Modelo I. As estimativas geradas pelos modelos criados foram comparadas com os valores de fluxo de mercadorias encontrados em 2007, na *Commodity Flow Survey*. Entretanto, os valores estimados e os valores encontrados na pesquisa de 2007 mostraram que os modelos subestimaram o fluxo de mercadorias em alguns municípios.

Lawson *et al.* (2012) analisaram os efeitos do uso do solo e do número de funcionários na geração de viagens de carga (GVC). Os autores utilizaram dados da cidade de Nova York, Estados Unidos e aplicaram três técnicas diferentes, são elas: taxas de geração de viagem padrão, mínimos quadrados ordinários e análise discriminante múltipla. A classificação do uso do solo foi realizada a partir de três códigos diferentes que foram utilizados associados aos dados de número de funcionários para gerar os modelos. Os autores concluíram que os modelos classificados pelo *City of New York Zoning Resolution* (NYCZR) e *Land-Based Classification Standards* (LBCS) forneciam uma alternativa melhor às taxas de viagem quando comparadas ao manual do ITE (*Institute of Transportation Engineers*), pois, forneceram estimativas mais precisas para atração de carga. Os melhores modelos foram selecionados considerando o teste-t e a Raiz do Erro Quadrático Médio (*Root Mean Squared Error* - RMSE). Para 73% dos modelos NYCZR, um coeficiente constante produz os melhores modelos de GVC. Para 18% dos modelos, a GVC depende do emprego e para 9% dos modelos, o melhor resultado é obtido através de um modelo que combina um coeficiente constante e a taxa de emprego.

Holguín-Veras *et al.* (2013) avaliaram e definiram formas de melhorar a transferibilidade dos modelos de geração de viagem de carga (MGVC). Eles apresentaram as principais premissas que devem orientar o desenvolvimento dos MGVC e avaliaram a transferibilidade de duas formas diferentes. O primeiro é através das análises de quão bem os MGVC são capazes de estimar o fluxo real de viagens de carga em vários casos de validação. O segundo é através de MGVC econométricos que avaliam a significância estatística de variáveis binárias que

representam locais geográficos específicos. Os autores utilizaram os modelos desenvolvidos pelo *National Cooperative Freight Research Program Project 25* (NCFRP 25), *Quick Response Freight Manual* (QRFM) e, pelo *Institute of Transportation Engineers' Trip Generation Manual* (ITE). Os resultados mostram que, os modelos desenvolvidos como parte do projeto NCFRP 25 "*Freight Trip Generation and Land Use*" possuem melhor capacidade de previsão do que os incluídos em outras compilações. Os procedimentos de correção sintética melhoram a transferibilidade e nenhum efeito de localização está presente nos dados do teste.

No trabalho de Jaller *et al.* (2013) foram apresentados modelos de geração de viagem de carga baseados na área dos estabelecimentos. Além disso, os autores avaliaram o desempenho da variável independente área do estabelecimento versus número de funcionários em modelos desenvolvidos previamente pelos autores. São utilizadas duas amostras de dados de geração de viagem de carga de empresas receptoras e operadoras em Nova York e Nova Jersey, Estados Unidos. Os dados foram coletados entre 2006 e 2012, para realizar a validação externa dos modelos. No total, cerca de 380 empresas receptoras de diferentes segmentos da indústria em Manhattan e Brooklyn foram entrevistadas. Os dados sobre distribuição de mercadorias coletados incluem número de entregas diárias, caminhões, emprego e vendas, entre outros. Os resultados indicam que existe uma forte correlação entre o emprego e a área, mas tal relação depende da definição do setor do estabelecimento. Os resultados mostram que a maior correlação entre emprego e área corresponde à área comercial. De acordo com as descobertas anteriores, um grande número de indústrias têm padrões constantes de atração de viagem de carga e de transporte de mercadorias (produção de viagens), independentemente do tamanho da empresa (medido por níveis de emprego ou área). Os modelos também foram gerados por meio de regressão linear.

Alho e Silva (2014) desenvolveram modelos de geração de viagem de carga com base em uma pesquisa sobre distribuição de mercadorias no comércio varejista da cidade de Lisboa, Portugal. Os dados que alimentaram a pesquisa foram extraídos de um questionário aplicado em 604 estabelecimentos. A relação das possíveis variáveis independentes com o número total de entregas semanais foi investigado. Os diversos conjuntos de variáveis foram modelados sob regressão linear e regressão generalizada. As variáveis independentes utilizadas foram o

tamanho do estabelecimento, área de vendas, área do armazém, número de funcionários e número de fornecedores. Os autores exploraram outras possíveis variáveis explicativas como: cadeia de fornecimento, proporção de moradias e imóveis, densidade dos moradores, densidade do estabelecimento de mercadorias, entre outros. A análise das variáveis mostrou que aquelas mais comumente utilizadas em aplicações práticas não eram necessariamente os melhores preditores de demanda. O número de funcionários foi consistentemente um melhor preditor do que a área do estabelecimento. Os modelos de regressão linear apresentaram melhores capacidades de previsão quando comparados aos modelos generalizados. Em geral, as capacidades preditivas dos modelos, com a metodologia escolhida, foram inferiores ao que seria considerado aceitável para uma aplicação prática.

O estudo de Barone, Crocco e Mongelli (2014) estimou a demanda de carga em escala urbana, considerando como área de estudo a cidade de Cosenza e Rende, na Itália. Em particular, os autores decidiram considerar questões relativas à cadeia de frutas e vegetais. A estimativa da demanda foi feita através do rastreamento de informações diretamente no ponto de venda, por meio de um questionário. Os dados foram coletados em 125 pontos de venda referentes ao tipo e superfície do espaço de vendas, disponibilidade ou ausência de veículos, ou equipamentos para descarga, tipo, classe e veículo utilizado para alimentação, frequência de entregas, quantidades entregues, origem das mercadorias, intervalo de tempo entre pedidos e entregas, número de fornecedores, fator de carga do veículo utilizado para a entrega das mercadorias, entre outros. Uma análise preliminar foi realizada por meio de regressão linear múltipla, revelando uma correlação entre a superfície do espaço dedicado às vendas e a quantidade de mercadorias recebidas. Os resultados iniciais mostram que a variável explicativa significativa é a superfície dedicada ao atacado.

Cantillo, Holguín-Veras e Jaller (2014) propuseram uma estrutura de modelagem de transporte de carga na qual um conjunto de modelos de geração de volume de carga foi desenvolvido para a Colômbia. A estrutura de modelagem foi desenvolvida como uma variação de um modelo de viagens de quatro etapas baseado em fluxos de *commodities*. Para considerar explicitamente os requisitos da mercadoria, são incluídas etapas adicionais para estimar fluxos de viagens vazias e viagens de mercadorias. Segundo o estudo, na Colômbia, trinta e quatro grupos agregados de *commodities* representam 83% dos fluxos desse grupo de



mercadorias. Ainda segundo os autores, apesar das limitações, a pesquisa evidencia a viabilidade de desenvolver uma estrutura de modelagem de demanda de mercadorias nos países em desenvolvimento, mesmo com dados e recursos limitados para esforços abrangentes de coleta de dados. Como variáveis independentes foram utilizadas a população, produto interno bruto (PIB), área cultivada e quantidade de animais. Foram gerados modelos de regressão linear e taxas de geração de viagens.

No estudo de Comi e Nuzzolo (2014), os autores apresentaram o desenvolvimento de um sistema de modelagem para estimar fluxos de carga em áreas urbanas a partir dos resultados de um estudo realizado para investigar o comportamento do consumidor final na cidade de Roma (Itália). Partindo do resultado dessa pesquisa, os autores utilizaram regressão linear, modelos lineares generalizados e modelo logístico multinomial e, como variáveis correlacionadas, o número de empregados, tamanho médio da entrega, número de estabelecimentos varejistas. Foram então estimados os valores de geração de viagens de mercadorias baseados nas escolhas do consumidor final em relação ao tipo de ponto de venda (por exemplo, pequeno, médio ou grande porte), pois, segundo os autores, tais escolhas impactam os fluxos de distribuição de mercadorias.

No estudo de Kulpa (2014) foram desenvolvidas equações de geração de viagens de carga para caminhões leves e pesados na Polônia. O autor utilizou duas fontes de dados secundários: um estudo abrangente de viagens na área metropolitana de Cracóvia com dados do ano de 2007 e outro na cidade de Poznań com dados do ano 2000. O conjunto de dados desenvolvidos foram utilizados para estimar equações de geração de viagem, aplicando regressão linear e não linear, bem como redes neurais artificiais. Para os dois tipos de caminhões (leves e pesados), os melhores resultados foram obtidos usando o número de habitantes como variável independente. A variável independente que apresentou pior desempenho foi o número de empregados na agricultura. A regressão múltipla e as redes neurais artificiais são mais complexas e, dependendo do número de variáveis explicativas, podem dar melhores resultados. No entanto, na verificação dos modelos, especialmente para redes neurais artificiais e modelos de regressão múltipla os resultados apresentaram maiores erros.

Lim *et al.* (2014) desenvolveram modelos de regressão linear que descrevem as relações entre produção e atração de viagens de carga baseadas em *commodities* usando variáveis econômicas específicas. Os autores utilizaram informações como dados de origem-destino de mercadorias disponibilizadas pela *Freight Analysis Framework* (FAF) do estado da Califórnia (Estados Unidos). Os modelos de geração de viagem foram validados usando dados econômicos da Califórnia e aplicados para prever as viagens produzidas no ano de 2015. As viagens produzidas e atraídas foram as variáveis dependentes. As variáveis independentes foram uma combinação de população, emprego, área agrícola, vendas de culturas e dados de energia. Os modelos gerados por meio de regressão linear podem ser usados para determinar onde os investimentos em infraestrutura devem ser alocados, bem como a priorização de áreas com alto fluxo de mercadorias.

Jaller *et al.* (2015) apresentaram procedimentos para identificar e quantificar o papel desempenhado pelos grandes geradores de tráfego de mercadorias como contribuintes do tráfego de caminhões em áreas metropolitanas. Embora os portos, terminais de contêineres e outros locais industriais sejam geralmente associados a grandes geradores de viagens de caminhão, eles representam apenas uma pequena proporção do total de viagens produzidas e atraídas em grandes áreas metropolitanas. Os autores exploraram o papel de *clusters* de negócios ou edifícios comuns na geração de grande volume de tráfego urbano de mercadorias (*Large Urban Freight Traffic Generators – LTG’s*) na cidade de Nova York, Estados Unidos. Embora diferentes procedimentos selecionem LTG’s sob diferentes perspectivas, com diferentes classificações, todos os procedimentos indicam que, embora sejam limitados em número, os LTG’s são extremamente influentes em termos de atividade geral de carga urbana. Conclui-se que um número limitado de edifícios e estabelecimentos que produzem e atraem um grande número de viagens de caminhões podem ser facilmente identificados e possuem um grande potencial para ajudar a melhorar a eficiência das entregas de carga em áreas urbanas.

Priya *et al.* (2015) investigaram um modelo melhorado de viagens de caminhão usando séries temporais, redes neurais artificiais e padrões de uso da terra para Chennai, Índia. O objetivo principal do trabalho foi estudar a demanda de viagens por meio da estimação de modelos de geração de viagens de carga desenvolvidos com dados de pesquisa desagregados no nível de

estabelecimento. Os autores compararam seu estudo com um estudo semelhante realizado em Nova York. No geral, os estabelecimentos em Chennai atraem mais viagens do que em Nova York. Isto está em contraste com o que é observado nas taxas de viagens de passageiros, onde as cidades indianas têm menor valor em comparação com os países ocidentais.

Sánchez-Díaz *et al.* (2015) apresentaram um modelo de geração de viagens de carga urbana para estimar os fluxos de carga usando fontes de dados secundárias, por exemplo, contagens de tráfego, reduzindo a necessidade de pesquisas dispendiosas. Foi avaliada a capacidade do modelo para replicar os valores reais usando a região de Denver, Estados Unidos. O modelo, referido como Síntese de Distribuição de Carga em Turnos (*Freight Tour Synthesis - FTS*), melhora as técnicas atuais, incorporando o comportamento dependente do tempo (baseado em turnos) para veículos de carga. O modelo estima o conjunto provável de fluxos em turnos, dado um conjunto de estimativas de geração de carga, um conjunto de contagens de tráfego por intervalo de tempo e o custo total de transporte de mercadorias na rede.

Wang e Zhou (2015) abordaram a geração de viagens de carga a partir de compras do comércio eletrônico. Os autores utilizaram dados da Pesquisa Nacional de Viagens (*National Travel Survey - NHTS*) dos Estados Unidos para investigar as viagens de carga geradas por unidades residenciais na cidade de Nova York. Um modelo de escolha binária e um modelo binomial negativo foram usados para identificar os impactos das variáveis relacionadas às pessoas, famílias e regiões específicas na frequência de entrega domiciliar. A estimativa de viagens de carga geradas por unidades residenciais também foi comparada às viagens de carga geradas pelos estabelecimentos comerciais. Os resultados mostraram que fatores individuais, familiares e regionais são influentes para a frequência de entrega domiciliar. Os usuários frequentes da internet, mulheres, pessoas brancas e pessoas com nível de ensino superior tendem a receber mais entregas de mercadorias. Os resultados indicaram que as viagens de carga geradas por unidades residenciais têm uma magnitude comparável às viagens geradas pelas empresas. O número total de viagens de carga geradas pelos estabelecimentos comerciais é de 9,25 milhões para a área de estudo. As entregas domiciliares representam cerca de 31% das viagens de carga geradas por empresas.

Alho e Silva (2016) testaram e compararam quatro metodologias (Análise de Classificação Cruzada, Análise de Classificação Múltipla, Modelo Linear Generalizado e Modelo Logit Ordinal) para prever a geração de viagens de carga. O objetivo do estudo era aprimorar modelos de geração de viagens de carga, contribuindo assim para melhorar o planejamento urbano. A principal fonte de dados foi uma pesquisa sobre distribuição de mercadorias baseada em estabelecimentos, que reuniu dados de 604 estabelecimentos de varejo na cidade de Lisboa, Portugal. As variáveis independentes selecionadas foram a categoria industrial, o número de funcionários e a área do estabelecimento de varejo. A semelhança nos resultados entre a análise de regressão múltipla e os métodos de partição sugere que, para estabelecimentos relativamente pequenos, a variação no número de funcionários para cada categoria de indústria é uma aproximação razoável para a previsão de entregas semanais. A análise permitiu a conclusão de que a contribuição variável se modifica de acordo com a metodologia de modelagem escolhida. Existe pouca variabilidade na qualidade das previsões dependendo do modelo selecionado, mas uma melhoria considerável nas previsões corretas pode ser conseguida reduzindo a variabilidade da variável dependente.

Ferreira e Silva (2016) analisaram modelos de geração de viagem de carga para identificar variáveis explicativas e métodos matemáticos frequentemente utilizados em modelos anteriores e seleciona, dentre as encontradas, variáveis que podem ser usadas para auxiliar na compreensão do fluxo de cargas, considerando o caso da Região Metropolitana do Rio de Janeiro (RMRJ). Os autores realizaram uma revisão sistemática e bibliométrica da literatura e identificaram variáveis explicativas relacionadas ao estabelecimento, como número de funcionários, área construída, número de lojas, entre outras. Variáveis relacionadas à parâmetros socioeconômicos da região também foram observados, como: população, número de domicílios e empregos na região. Dos 25 trabalhos analisados 21 apresentaram como variável dependente o número de viagens, enquanto somente quatro utilizaram volume de carga. Foram utilizados os métodos de taxas de geração de viagem e regressão linear para criação de modelos de geração de viagem de carga para região metropolitana do Rio de Janeiro. Os resultados mostraram que todas as variáveis explicativas podem explicar a geração de viagens de carga, sendo domicílios e população as que apresentaram melhor relação com o fluxo de cargas.

O estudo de Günay *et al.* (2016) apresentou uma abordagem de modelagem condicional para modelos de geração de viagem de carga para uma categoria específica de caminhão chamada *tractor-trailers*. Os dados utilizados foram obtidos do *Kocaeli Logistics Master Plan* (KOLMAP) - Plano Diretor de Logística da província de Kocaeli (Turquia) no período de agosto a dezembro de 2011. Os autores identificaram que a principal fraqueza em outros estudos que tratavam de viagens de carga foi a inadequação do sistema de classificação utilizado para agrupar estabelecimentos comerciais em um conjunto de classes padronizadas. Assim, um experimento foi conduzido para criar grupos de sites logísticos que possuíam características homogêneas de geração de viagem de carga. Eles observaram que a categoria de veículo *tractortrailers* apresentava zero viagens, então para resolver este problema foi proposta e testada uma modelagem condicional. As variáveis independentes utilizadas no estudo foram a área e o número de empregos dos estabelecimentos, além do setor industrial e informações operacionais. Posteriormente, foram comparados os modelos desenvolvidos usando a nova abordagem com os modelos obtidos usando apenas a abordagem de modelagem comum da análise de regressão. Os resultados indicaram que a criação de grupos homogêneos de sítios logísticos foi possível. A nova abordagem de modelagem condicional que foi aplicada aos *tractortrailers* apresentou melhor desempenho do que a modelagem de regressão linear.

Mommens *et al.* (2016) reuniram dados de carga para a Bélgica, que incluem nove tipos diferentes de unidades de carregamento de mercadorias. Em contraste com a literatura internacional sobre modelagem de carga, optou-se por incluir nove categorias em fases de modelagem adicionais. Para isso, diferentes técnicas de regressão foram testadas. Os melhores resultados foram encontrados ao aplicar o modelo de regressão linear generalizada com *link* de log (GLML) em volumes por tipo de mercadoria apenas como variável dependente. A fim de migrar para modelos de transporte baseados em agentes mais desagregados e poder ilustrar a importância de carregar unidades na modelagem e na realidade, integrou-se a informação disponível sobre unidades de carregamento usadas após a análise de regressão através de uma função de probabilidade que considera tanto um parâmetro geográfico quanto um parâmetro de tipo de mercadoria. Os resultados mostraram-se estáveis e robustos segundo os autores.

Oliveira *et al.* (2016) desenvolveram modelos de geração de viagens de carga para bares, restaurantes, mercados e supermercados em Belo Horizonte, Brasil. Os dados para desenvolvimento do modelo foram obtidos por meio de uma pesquisa de campo conduzida de forma a extrair informações como área dos estabelecimentos, número de funcionários e rotinas operacionais. Alguns dados da pesquisa (área e emprego) foram utilizados como variáveis explicativas. A variável dependente nesse estudo foi o número de veículos de carga atraídos pelo empreendimento. Utilizou-se regressão linear para modelagem do fenômeno. Os resultados mostraram que uma concentração de empreendimentos em uma determinada região implica em um elevado número de veículos de carga realizando entregas, podendo reduzir a mobilidade da região e seu entorno. Uma comparação com modelos existentes foi realizada e identificaram-se semelhanças (supermercados) e divergências (bares e restaurantes), destacando a necessidade de modelos locais. A análise espacial permitiu identificar áreas com maior concentração de estabelecimentos e viagens de veículos de carga e podem ser utilizados como subsídio para definição de políticas públicas.

O estudo de Sánchez-Díaz (2016) explorou as deficiências relacionadas à distribuição de mercadorias dos estabelecimentos comerciais urbanos e os impactos no tráfego usando dados coletados dos estabelecimentos da cidade de Gotemburgo, Suécia. O autor apresenta modelos estatísticos, desenvolvidos com base em análises de regressão e modelos de escolha discreta, para estimar o número de viagens de mercadorias produzidas e atraídas por semana, além de estimar também a atração de viagens analisando a mercadoria (peso e volume). Além de explicitar os fatores que determinam a geração de viagens de carga dos estabelecimentos, os modelos são desenvolvidos com o objetivo de auxiliar o planejamento e as políticas urbanas, de modo que as variáveis explicativas são selecionadas com base na adequação e disponibilidade. Uma pesquisa obteve dados sobre o número de viagens de entrega (atraídas) e o número de viagens de saída (produzidas) dos estabelecimentos em uma semana típica. Dados como a quantidade de carga atraída para o estabelecimento, o tipo de veículos que realiza a entrega, o número de fornecedores, o número de funcionários, a área do estabelecimento e a política de pedidos e estoque, entre outros foram coletados. Os resultados mostraram que o comércio varejista de produtos perecíveis têm a maior geração de viagem de carga por estabelecimento, seguidos por escritórios do setor público e estabelecimentos de educação, varejistas de produtos não perecíveis e restaurantes.

O estudo de Sánchez-Díaz *et al.* (2016) analisou a atração de viagens de carga e as relações com as principais características do ambiente urbano. Usando dados de estabelecimentos, os autores exploraram a influência dos atributos do mercado, bem como do uso da rede (malha rodoviária) e de uso do solo. A pesquisa utilizou dados de 343 estabelecimentos de cinco setores industriais diferentes em Nova York, Estados Unidos. A evidência empírica sugere que a localização dos estabelecimentos, relacionada ao seu valor de mercado e a largura da rua, desempenham um papel importante na atração de viagens de carga e que as unidades varejistas localizadas em zonas de alto emprego tendem a produzir altas taxas de atração de viagens de carga por empregado. Os resultados da análise econométrica e os testes de especificação mostraram que os modelos de atração de viagens de carga são melhor representados com uma função não-linear quando se utiliza o número de empregos e outras variáveis de localização. Outro fato interessante é que a atração de viagens de carga é melhor modelada usando modelos não-lineares para todos os setores da indústria.

Boarnet *et al.* (2017) examinaram a relação entre padrões de desenvolvimento espacial urbano e viagens de carga em Los Angeles, Estados Unidos. Usando os dados da *National Employment Time Series* (NETS), identificaram *clusters* de emprego na área metropolitana de Los Angeles. Os autores desenvolveram um modelo de regressão que estima a atividade de carga em função das características geográficas, os *clusters* de emprego, medidas de emprego nas proximidades, acesso à rede rodoviária e proximidade de instalações de transporte intermodal. Os resultados indicaram que o emprego é um fator importante da atividade de carga. Entretanto, os *clusters* de emprego têm um efeito independente sobre a atividade de carga. Os resultados deste estudo sugerem que pesquisas adicionais sobre estrutura espacial urbana e atividade de carga devem avaliar os efeitos dos *clusters* de emprego e como sua composição e características específicas do emprego estão associadas às atividades de carga no nível metropolitano.

O estudo de Gonzalez-Felú e Peris-Pla (2017) apresentou um quadro integrado para avaliar as relações entre atratividade do comércio de varejo, tanto de carga quanto de compras geradas em áreas urbanas de Lyon, França. A estrutura proposta implementa uma atração de viagem (*Freight Trip Attraction* - FTA) e um modelo de atração de viagem de compras (*Shopping Trip Attraction* - STA) modelados a partir de regressão linear. No artigo é

apresentado um quadro metodológico descrevendo os modelos FTA e STA, bem como o indicador de atratividade proposto. As variáveis explicativas demográficas foram: população e número de domicílios da seção considerada; densidade populacional e densidade do agregado familiar. As variáveis explicativas socioeconômicas foram: número de estabelecimentos; número de supermercados e grandes polos comerciais; número de hipermercados e lojas similares; número de funcionários nas pequenas e médias empresas, supermercados e hipermercados, e presença de um *shopping*, apresentado como uma variável binária. Os resultados mostram que as taxas de FTA e STA têm uma relação direta com a atratividade do varejo, mas a categoria de espaço urbano também precisa ser incluída nessas análises.

Oliveira, Nobrega, Ebias e Corrêa (2017) apresentaram os resultados de um modelo de geração de viagem de carga desenvolvido para bares e restaurantes em Belo Horizonte, Brasil. Os dados necessários para o desenvolvimento do modelo foram obtidos a partir de um questionário, com informações sobre bens, frequência, tempo operacional, local de carga e descarga de mercadorias, tamanho do estabelecimento e número de funcionários. A coleta de dados foi realizada via entrevista, conduzidas com o gerente do estabelecimento ou um membro técnico responsável pelas operações logísticas. Foram entrevistados 300 estabelecimentos. O estudo teve como foco o setor de bares e restaurantes da cidade de Belo Horizonte, que representa 42% do número de entregas em uma área que contém a grande maioria de estabelecimentos comerciais e problemas de mobilidade urbana. Os modelos foram desenvolvidos através de regressão linear simples, correlacionando: número de viagens versus área do estabelecimento; número de viagens versus número de empregados; número de viagens versus dia de operação do estabelecimento. Os resultados dos modelos de geração de viagem foram analisados em conjunto com outros dados geográficos, como dados demográficos, densidade da rede rodoviária e dados socioeconômicos.

Rolko e Friedrich (2017) apresentaram os resultados intermediários de um estudo empírico sobre as localidades Provedoras de Serviços de Logística (*Logistics Service Providers – LSP*) na Alemanha. A distribuição de mercadorias geradas por redes alemãs de caminhão-carga é estimado em um nível agregado. O método de coleta de dados conciliou informações sobre os padrões espaciais e estruturais dos locais alocaíveis ao setor de logística alemão e o transporte de carga. O número de viagens de carga gerados pelos LSP's são altamente relevantes para o



setor de transporte e as políticas de planejamento do uso do solo. Os dados utilizados no estudo continham informação de 2.809 locais de LSP's. Foram obtidas informações representativas da estrutura dos locais LSP para os atributos: área de operações logísticas e empregos. Além disso, a aplicação das taxas de geração de carga provou ser bem sucedida. O estudo mostrou que o setor foi capaz de gerar cerca de 71,8% da tonelagem total de carga transportada. Esse dado confirmou o nível de integridade da base de dados de logística de modo que se torna possível a continuidade das análises com outros tipos de locais ou para outros segmentos de mercado de logística.

Gonzalez-Feliu e Sánchez-Díaz (2018) analisaram os impactos do nível de agregação e construção de categoria sobre a relevância e qualidade dos modelos de geração de viagens de mercadorias. O artigo propõe um método para comparar modelos de geração constante e de forma funcional em diferentes classificações de categoria com base em estimativas de percentual de erro absoluto médio (*mean absolute percentage error* – MAPE). As formas funcionais são avaliadas por regressão linear e comparadas pelo coeficiente de Pearson. Os resultados mostram que o nível de agregação nem sempre tem um impacto positivo na precisão do modelo e a escolha da forma funcional adequada leva a modelos mais precisos.

Pani *et al.* (2018) apresentaram um conjunto de 74 modelos de geração de viagens, usando a regressão OLS, para sete cidades em Kerala, na Índia. Os modelos foram formulados considerando variáveis de tamanho de negócio que capturam a atividade de geração de viagens do estabelecimento. Os resultados da estimativa do modelo revelaram que os indicadores de tamanho do negócio, como número de funcionários e área bruta, explicaram bem a geração de viagens. O número de anos no negócio (*number of years in business* - YB) foi identificado como outra variável e considerado na especificação do modelo para estimar a geração de viagens. A investigação detalhada dos modelos e variáveis individuais sugeriu que os modelos baseados em emprego são adequados para cidades com atividades comerciais densas e maior valor da terra, enquanto os modelos baseados em área representavam melhor a geração de viagens em cidades com nível médio de urbanização. Segundo os autores, a área parece ser um indicador distorcido para representar o tamanho do negócio em cidades com atividades comerciais densas. Neste caso, onde o dado da área é de difícil aquisição, o emprego pode representar melhor o crescimento da atividade de geração de viagens. Como

uma extensão dos modelos de geração de viagens específicos da cidade, três tipos de modelos de geração de viagens combinados foram desenvolvidos para fornecer evidências estatísticas quantitativas das diferenças nas especificações dos modelos nas cidades. Os resultados estatísticos desses modelos sugerem que as atividades de geração de viagens são influenciadas pela interação das características do estabelecimento e sua localização. O efeito de interação é mais proeminente quando a área é usada para representar o tamanho do negócio. A interação dos modelos de geração de viagens discutidos subsequentemente no estudo pode ser utilizada na aplicação do planejamento de transporte para as necessidades de capacidade da rede estadual, regional e em nível de corredor para a Índia.

Ainda sobre os estudos sobre geração de viagens, uma das fontes mais relevantes para o estudo de geração de viagens na literatura internacional é o Manual de Geração de Viagens do ITE (*Institute of Transportation Engineers*). Ele consiste em uma compilação de estudos de geração de viagens, desenvolvidos por agências públicas, empresas de consultoria e universidades. Apesar do foco do Manual ITE não seja especificamente transporte de mercadorias, alguns resultados podem ser aplicados diretamente à geração de viagens de veículos de carga (ITE, 2008 *apud* SÁNCHEZ-DÍAZ, 2016). Outra importante fonte a esse respeito é o NCFRP 25 *Freight Trip Generation and Land Use Report* (Holguín-Veras *et al.*, 2012). Esse, voltado exclusivamente para a geração de viagens de veículos de carga, possui uma discussão abrangente sobre a modelagem dessas viagens e um conjunto de estudos de caso com modelos de geração baseados em dados de estabelecimentos da cidade de Nova York. Nesse documento são apresentadas estimativas de geração de viagens para diferentes tipos de estabelecimentos, diferentes usos do solo e são utilizadas diferentes técnicas para estimativa destas viagens, como taxas de viagens, OLS e MCA.

Na Tabela 2.2 é apresentada uma síntese dos estudos sobre modelos de geração de viagens de carga encontrados na literatura pesquisada e detalhados nesse capítulo. Tais pesquisas podem contribuir com o planejamento das políticas urbanas voltadas para distribuição de mercadorias, de modo a propor medidas que beneficiem a todos os envolvidos. As pesquisas possibilitam estimativas de fluxos de carga e respectivas viagens, sendo assim o poder público pode tomar medidas para minimizar problemas no trânsito bem como no entorno dos grandes

empreendimentos (PGV), em especial daqueles industriais que tendem a atrair veículos de maior porte.

Em síntese, nessa tabela apresentam-se as fontes em colunas segundo o ano de publicação, referência bibliográfica e a origem, as variáveis dependentes e independentes, e as técnicas utilizadas na construção dos modelos em cada estudo.

Tabela 2.2: Revisão sistemática da literatura.

ANO	TRABALHOS	ORIGEM	VARIÁVEIS DEPENDENTES		VARIÁVEIS INDEPENDENTES										MÉTODOS								
			Número de viagens	Volume de Carga	Área	Carga	Domicílios	Emprego	Frota	Funcionários	Infraestrutura	Número de empresas	Operação	População	Renda	Uso do solo	Outros	Análise de classificação Múltipla	Input/Output	Redes neurais	Regressão espacial	Regressão linear	Regressão não linear
2018	Gonzalez-Feliu e Sánchez-Díaz	-																					
2018	Pani <i>et al.</i>	Kerala, Índia																					
2017	Boarnet; Hong e Santiago-Bartolomei	Los Angeles, EUA																					
2017	Gonzalez-Feliu e Peris-Pla	Lyon, França																					
2017	Oliveira, Nobrega, Ebias e Corrêa	Belo Horizonte, Brasil																					
2017	Rolko e Friedrich	Alemanha																					
2016	Alho e Silva	Lisboa, Portugal																					
2016	Ferreira e Silva	Rio de Janeiro, Brasil																					
2016	G. Günay; G. Ergün; I. Gökaşar	Kocaeli, Turquia																					
2016	Mommens <i>et al.</i>	Bélgica																					
2016	Oliveira, Oliveira, Ramos e Ebias	Belo Horizonte, Brasil																					
2016	Sánchez-Díaz	Gotemburgo, Suécia																					
2016	Sánchez-Díaz, Holguín-Veras e Wang	Nova York, EUA																					
2015	Jaller, Sánchez-Díaz e Holguín-Veras	Nova York, EUA																					
2015	Priya <i>et al.</i>	Chennai, Índia																					
2015	Sánchez-Díaz, Holguín-Veras e Ban	Denver, EUA																					
2015	Wang e Zhou	Nova York, EUA																					
2014	Alho e Silva	Lisboa, Portugal																					
2014	Barone, Crocco e Mongelli	Cosenza e Rende, Itália																					
2014	Cantillo, Holguín-Veras e Jaller	Colômbia																					
2014	Comi e Nuzzolo	Roma, Itália																					
2014	Kulpa	Cracóvia e Poznán, Polónia																					
2014	Lim <i>et al.</i>	Califórnia, EUA																					
2013	Holguín-Veras <i>et al.</i>	EUA																					
2013	Jaller, Holguín-Veras, Sánchez-Díaz e Lawson	Nova York; Nova Jersey, EUA																					
2012	Campbel <i>et al.</i>	Nova York, EUA																					
2012	Holguín-Veras <i>et al.</i>	Nova York, EUA																					
2012	Ibeas <i>et al.</i>	Roma e Santander, Itália e Espanha																					

Tabela 2.2: Revisão sistemática da literatura – continuação.

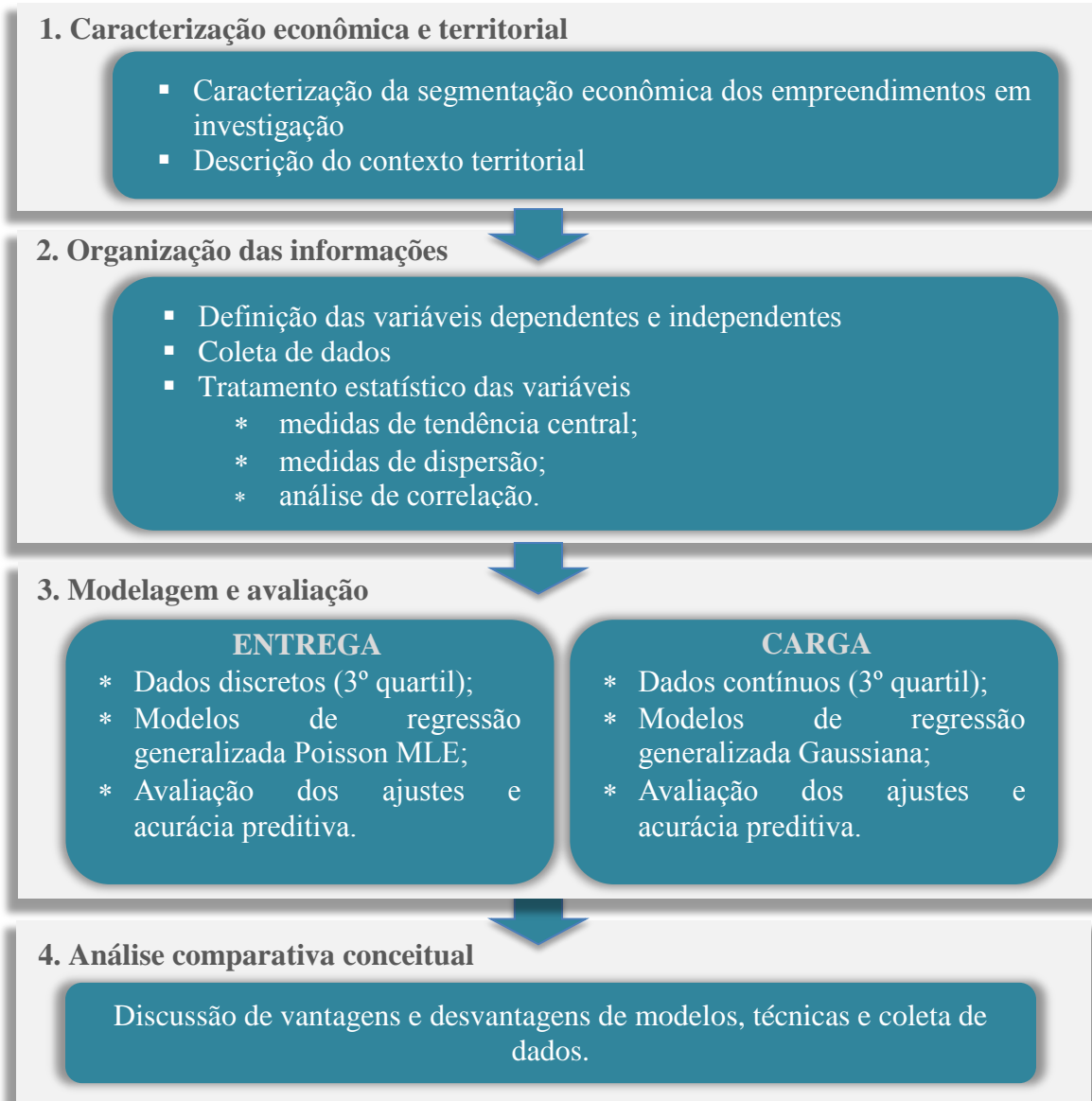
ANO	TRABALHOS	ORIGEM	VARIÁVEIS DEPENDENTES		VARIÁVEIS INDEPENDENTES										MÉTODOS								
			Número de viagens	Volume de Carga	Área	Carga	Domicílios	Emprego	Frota	Funcionários	Infraestrutura	Número de empresas	Operação	População	Renda	Uso do solo	Outros	Análise de classificação Múltipla	Input/Output	Redes neurais	Regressão espacial	Regressão linear	Regressão não linear
2012	Jin <i>et al.</i>	Utah, EUA																					
2012	Lawson <i>et al.</i>	Nova York, EUA																					
2012	Russo e Comi	Roma, Itália																					
2011	Chu	Kaohsiung, Taiwan																					
2011	Novak <i>et al.</i>	EUA																					
2010	Crocco <i>et al.</i>	Cosenza, Rende e Castrolibero, Itália																					
2010	Gasparini <i>et al.</i>	Rio de Janeiro, Brasil																					
2010	Silva e Goldner	Tubarão, Brasil																					
2009	Bastida e Holguín-Veras	Nova York, EUA																					
2008	Kawamura <i>et al.</i>	EUA																					
2007	Silva e Waisman	São Paulo, Brasil																					
2006	Wisetjindawat, Sano e Matsumoto	RM de Tóquio, Japão																					
2005	Sarvareddy <i>et al.</i>	Flórida, EUA																					
2004	Campos e Melo	Rio de Janeiro, Brasil																					
2004	Klodzinski <i>et al.</i>	Flórida, EUA																					
2002	Allen	Baltimore, EUA																					
2002	Iding, Meester e Tavasszy	Holanda																					
2002	Melo	Rio de Janeiro, Brasil																					
2002	Russo e Comi	Itália																					
2001	Al-Deek	Flórida, EUA																					
1999	Black	Indiana, EUA																					
1999	Marra	Campinas, Brasil																					
1996	Cambridge Systematics	Phoenix, Arizona																					
1995	Wegmann <i>et al.</i>	Canadá e EUA																					
1994	Tadi e Balbach	Fontana, EUA																					
1992	Ogden	Chicago, Vancouver e Melbourne																					
1974	Hutchinson	RM de Toronto, Canadá																					

### **3 ABORDAGEM METODOLÓGICA**

A abordagem metodológica estruturada para este trabalho visa subsidiar o desenvolvimento e a avaliação de modelos de atração de carga e entregas por veículos de carga nos fluxos de insumos para indústrias de máquinas agrícolas e rodoviárias. Nesse contexto, a modelagem matemática proposta relaciona o quantitativo de insumos e o número de entregas de carga atraídas pelos empreendimentos em avaliação às variáveis explicativas selecionadas a partir do referencial teórico e coletadas em campo. Os modelos foram então avaliados em relação ao ajuste e à acurácia preditiva. Posteriormente, os conceitos de modelagem (baseados em quantidade de carga ou número de entregas) foram comparados qualitativamente através de uma discussão de vantagens e desvantagens desses modelos, suas técnicas e coleta de dados.

Na Figura 3.1 tem-se a estrutura metodológica proposta para desenvolvimento deste trabalho. Nas próximas seções, cada etapa metodológica será detalhada.

Figura 3.1: Estrutura metodológica do trabalho.



### 3.1 Caracterização econômica e territorial

As indústrias investigadas neste estudo são montadoras de máquinas agrícolas e rodoviárias. Como referência para análise da atração de fluxos de insumos, foram identificadas quatro unidades industriais localizadas nos municípios de Contagem (MG), Curitiba (PR), Piracicaba (SP) e Sorocaba (SP). O setor analisado foi escolhido devido à escassez de trabalhos específicos para esse segmento e à necessidade de gerar instrumentos possíveis de serem usados pelo Poder Público para o planejamento do transporte de carga nas áreas onde essas indústrias estão instaladas, a fim de reduzir os impactos causados por esse tipo de PGV Carga

no seu entorno. Essas indústrias são caracterizadas por um grande volume de caminhões que circulam transportando mercadorias para suas instalações. Seus fornecedores estão distribuídos em diversas cidades nos Estados de Minas Gerais, Paraná, Rio de Janeiro, Rio Grande do Sul, Santa Catarina e São Paulo. Nesta etapa da abordagem metodológica faz-se a caracterização da atividade econômica de produção de máquinas agrícolas e rodoviária no contexto industrial em avaliação. São ainda apresentados detalhes do escopo territorial considerando a localização das unidades industriais e seus respectivos fornecedores.

### **3.2 Organização das informações**

Os modelos de quantificação da demanda (geração de carga e entregas), tema deste estudo, podem ser utilizados como instrumentos capazes de contribuir com o planejamento de transporte de carga, conforme proposto por Oliveira *et al.* (2016). Para tanto, a primeira etapa do estudo consiste na definição do escopo e do tipo de empreendimento a ser estudado, com o objetivo de determinar atributos relevantes (variáveis explicativas) para a coleta de dados.

Ciente da escassez de observações amostrais para geração de modelos matemáticos que relacionem atributos estruturais com a atração de carga e entregas para as unidades industriais em investigação, foi realizada pesquisa bibliográfica sobre os critérios estatísticos para a construção de modelos de demanda. Visando a validação do tamanho da amostra desta pesquisa, foram consideradas as recomendações do *Institute of transportation Engineers* ITE (2008), que menciona que o tamanho da amostra deve ser maior ou igual a quatro observações para que se obtenha o melhor ajuste das curvas de regressão. O ITE (2008) também estabelece etapas para obtenção de taxas de viagens (entregas) tendo como base os seguintes aspectos: (i) definição do uso do solo a ser estudado; (ii) número de locais a serem pesquisado; (iii) seleção dos locais para aplicação de pesquisa; (iv) determinação do período da pesquisa; (v) escolha da variável independente e; (vi) escolha do método de contagem. Os dados coletados devem receber tratamento estatístico, assim como os modelos de regressão gerados.

Ainda segundo o ITE (2008), a condição para o melhor ajuste das curvas de regressão deve satisfazer cada um dos itens a seguir: (i) o  $R^2$  (coeficiente de determinação) deve ser maior ou igual a 0,50; (ii) o tamanho de amostra é maior que ou igual a quatro observações



(empreendimentos); (iii) o número de viagens (entregas) aumenta à medida que o tamanho da variável independente também aumenta.

A definição das variáveis dependentes e independentes é um passo muito importante na construção do modelo matemático de geração de viagens (entregas) e quantidade de carga, principalmente quando pretende-se estimar os fluxos (viagens e carga) para outros empreendimentos.

Os resultados da revisão sistemática indicam que o número de funcionários e de postos de trabalho (26%) e a área do empreendimento (16%) são as variáveis mais empregadas em estudos sobre modelos de geração de viagens de carga. Ainda, Alho e Silva (2014) corroboram essa informação ao afirmarem que essas variáveis são as mais frequentemente utilizadas nos modelos de geração de viagens de carga.

Por sua vez, as principais variáveis dependentes utilizadas são: (i) número de viagens de veículos de carga (79% dos estudos); e (ii) quantidade de carga movimentada, representando 21% dos trabalhos estudados.

Neste estudo, consideraram-se como variáveis dependentes (VD), visando explorar comparativamente às duas abordagens, a quantidade de carga movimentada diariamente com destino às unidades de produção de máquinas agrícolas e rodoviárias, e o respectivo número de entregas atraídas também com frequência diária. Destaca-se que não foram consideradas as viagens com destino às unidades industriais como variáveis dependente uma vez que as viagens são fenômenos atribuídos às decisões logísticas de organização da cadeia de suprimentos, ou seja, em uma viagem é possível que diversas entregas sejam contempladas. Assim, a informação trabalhada referente à variável dependente em análise, juntamente com a quantidade de carga movimentada no abastecimento, é o número de entregas diárias em cada unidade de produção. As variáveis explicativas (Variáveis Independentes - VI) consideradas foram: (i) a capacidade de produção diária das indústrias, em unidades de produto acabado; (ii) o número total de funcionários das indústrias; (iii) o número de funcionários do setor de logística das indústrias; (iv) a área total das indústrias, em metros quadrados; e (v) a área destinada ao setor da logística das indústrias, em metros quadrados.

Nesta etapa foi realizada a obtenção de dados de atração de carga e entregas para as indústrias em estudo, verificando a abrangência das unidades de produção e padronizando a unidade temporal utilizada na coleta de informações sobre o fluxo de carga. Conforme solicitado pelo grupo empresarial que cedeu os dados, foi necessária a manutenção de sigilo, não sendo possível identificar nominalmente os empreendimentos participantes desta pesquisa.

A coleta de dados referente à chegada de veículos para abastecimento de insumos nas indústrias estudadas foi realizada por meio de registros no *Oracle Transportation Management* (OTM), considerando o agendamento de coleta e entrega de insumos dos fornecedores para as linhas de montagem das empresas estudadas. Por meio do OTM sabe-se: (i) o dia e hora que a carga foi coletada no fornecedor; (ii) o tipo de veículo, placa e ano; (iii) os dados do motorista; (iv) o endereço do fornecedor de origem e planta de destino; (v) a quantidade de quilômetros percorrido entre origem destino; (vi) a empresa transportadora da mercadoria; (vii) o peso real da carga em quilogramas versus capacidade do veículo; (viii) a taxa de ocupação do veículo; e (ix) o tipo de operação. Tais dados são gerenciados pelo setor de transportes *inbound*<sup>2</sup> das empresas investigadas neste estudo. As unidades de produção registram eletronicamente os horários de entrada com dados da carga de cada veículo. Foram solicitados dados do ano de 2017 (janeiro a dezembro), considerando uma periodicidade de 24h (um dia). Os dados foram recebidos em formato eletrônico.

Também, junto à administração das empresas e em *sites* das indústrias, obtiveram-se dados sobre os empreendimentos, tais como número total de funcionários, funcionários destinados à logística de material e movimentação, área total construída, em metros quadrados, das empresas e área destinada à logística, tais como recebimento, armazenamento e movimentação de peças, produtos fabricados e capacidade de produção diária efetiva.

Após a coleta dos dados, foi realizada uma primeira análise, visando verificar se esses dados estavam alinhados às necessidades do estudo (considerando os objetivos, hipóteses e premissas), e conferindo a compatibilidade espacial e temporal entre as variáveis dependente e independente. Nessa fase, os dados foram sistematizados em uma base de dados para

---

<sup>2</sup> Setor da empresa responsável pelas entradas de matérias-primas na fábrica.

facilitar as tabulações e análises descritivas utilizando planilhas eletrônicas e o *Software R* versão 3.4.4, que se encontra disponível no endereço eletrônico <https://www.r-project.org/>.

Neste estudo, para descrever e sumarizar o conjunto de dados obtidos, foram consideradas métricas de estatística descritiva tais como medidas de tendência central (média, mediana e moda) e dispersão (desvio padrão, valor mínimo, valor máximo). Ainda, foram apresentados os *box-plot* das variáveis de interesse. Os *box-plot*, também conhecido como gráfico de caixa, são gráficos estatísticos que possibilitam representar a distribuição de um conjunto de dados com base em parâmetros descritivos. Existem algumas variações quanto à quantidade de estatísticas representadas nesse tipo de gráfico, mas de uma forma geral todos incluem a mediana, o primeiro e o terceiro quartil, os valores mínimos e máximos, e eventuais *outliers* e extremos (MAROCO, 2007).

Ainda na etapa de exploração estatística dos dados coletados, foram analisadas as correlações existentes entre as variáveis independentes e dependentes. A análise de correlação entre variáveis tem como objetivo a determinação da força do relacionamento entre duas observações emparelhadas. O termo correlação remete literalmente à forma como dois ou mais valores estão relacionados, pois, indica até que ponto os valores de uma variável estão relacionados com os de outra. Há muitos casos em que pode existir um relacionamento entre duas variáveis (STEVENSON, 1981). As análises de correlação são úteis em trabalhos exploratórios, quando deseja-se identificar as variáveis correlacionadas, possibilitando assim a seleção dessas variáveis para o estudo.

Neste trabalho, faz-se necessária uma aplicação de análise de correlação entre as variáveis dependentes, número de entregas atraídas e quantidade de carga movimentada nos fluxos de suprimento, e as variáveis explicativas (capacidade de produção diária das indústrias, o número total de funcionários, o número de funcionários do setor da logística, a área total dos empreendimentos e a área destinada ao setor da logística das indústrias). Para avaliar o grau de correlação entre as variáveis contínuas (quantidade de carga) foi considerado o coeficiente de correlação de Pearson ( $r$ ), valor que quantifica o grau de correlação de variáveis contínuas. O matemático e estatístico Karl Pearson (1857 – 1936) desenvolveu esta técnica que “mensura” o grau de relacionamento entre duas variáveis contínuas, denominado coeficiente de correlação “ $r$  de Pearson”. Tal técnica é válida apenas se duas hipóteses forem atendidas, a

saber: (i) tanto  $x$  como  $y$  são variáveis aleatórias contínuas, ou seja, ao contrário da análise de regressão, não se aceita selecionar certos valores de  $x$  e depois avaliar  $y$ , tanto  $y$  como  $x$  devem variar livremente; (ii) distribuição normal bivariada: a distribuição de frequência conjunta, ou seja, distribuição de valores dos pares  $x, y$  é normal. Essa última condição foi avaliada a posteriori e, portanto, o coeficiente de correlação de Spearman foi determinado tanto para correlações entre variáveis explicativas e dependentes contínuas e discretas (número de entregas). Essa técnica foi introduzida por Charles Spearman, psicólogo e estatístico, em 1904 (SIEGEL e CASTELLAN JR, 2006), e exige que as variáveis supostamente correlacionadas,  $x$  e  $y$ , sejam medidas pelo menos em escala ordinal. A correlação de Spearman avalia a relação monotônica entre duas variáveis ordinais. Em uma relação monotônica, as variáveis tendem a mudar juntas, mas não necessariamente a uma taxa constante. O coeficiente de correlação de Spearman baseia-se nos valores classificados de cada variável, em vez de nos dados brutos conforme destacam Hauke e Kossowski (2011) e Siegel e Castellan Jr (2006).

Conforme Stevenson (1981), o coeficiente de correlação tem duas propriedades que caracterizam a natureza de uma relação entre duas variáveis. A primeira é quanto ao sinal (positivo ou negativo) e a segunda refere-se a magnitude. Lançando-se os valores de  $x$  e  $y$  em um gráfico de dispersão e determinando-se uma equação de regressão linear, obtém-se o coeficiente angular da reta de regressão; o sinal desse coeficiente será o mesmo do coeficiente de correlação. A magnitude de  $r$  representa quão próximos da reta estão os pontos individuais. Os valores de  $r$  variam de  $-1,00$  a  $+1,00$  e quanto mais próximo das extremidades maior será a proximidade dos pontos a reta, enquanto os valores mais próximos de  $0$  sugerem maior dispersão. Para  $r$  positivo tem-se uma relação diretamente proporcional entre as variáveis, isto é, valores, altos ou baixos, de  $x$  implicam em valores (altos ou baixos) de  $y$  e vice versa. Para  $r$  negativo tem-se uma relação inversamente proporcional entre as variáveis, ou seja, valores altos de  $x$  implicam valores baixos de  $y$  e vice versa. Para  $r$  igual a zero tem-se baixa relação entre as variáveis, assim, alguns valores altos estão em correspondência com valores baixos e altos, e valores baixos estão em correspondência com valores altos e baixos (sem proporcionalidade). Maior detalhamento e as funções matemáticas para determinação dos coeficientes de correlação podem ser encontrados nas publicações de Everitt e Skrondal (2010) e Moore *et al.* (2007).

Os coeficientes de correlação são considerados para determinação dos modelos de regressão que serão gerados e avaliados, visando evitar multicolinearidade ou modelos que considerem variáveis explicativas que não se relacionam ao fenômeno em análise.

### **3.3 Modelagem e avaliação**

Nesta etapa da abordagem metodológica foram gerados os modelos de regressão relacionando as variáveis dependentes e explicativas.

Na revisão bibliográfica, a maior parte dos autores geraram taxas ou modelos de regressão linear simples pelo método dos mínimos quadrados (MMQ) para relacionar variáveis dependentes e independentes. Em uma fase exploratória dos dados neste estudo, foi feita a análise das condições de homocedasticidade (teste Breusch-Pagan - BP), linearidade ( $R^2$ ), independência (teste Durbin-Watson; teste qui-quadrado) e normalidade (teste Shapiro-Wilk; teste Pearson's Chi-squared) dos resíduos para utilização dessas técnicas de regressão. Entretanto, para nenhum conjunto de variáveis consideradas nos modelos, as condições foram atendidas em sua totalidade.

Nesse contexto, foram considerados modelos de regressão linear generalizada (GLM) como abordagem metodológica visando relacionar as variáveis em análise mesmo quando as premissas para uso do método dos mínimos quadrados não são atendidas ou quando é necessário gerar regressões para variáveis dependentes não contínuas.

Como é do interesse analítico e da modelagem, a subdivisão das amostras em categorias representativas, buscou-se, a melhor subdivisão para as amostras. Nesse sentido, sabe-se que os modelos analíticos e interativos têm uma abordagem comum no que se referem à geração de equações que correlacionem variáveis. Assim, de acordo com Messeti (2013), a análise de regressão consiste na realização de uma análise estatística objetivando verificar a existência de uma relação funcional entre a variável dependente com uma ou mais variáveis independentes, ou seja, consiste na obtenção de uma equação que tenta explicar a variação da variável dependente pela variação da variável independente. Valente *et al.* (2008) também complementam que, ao se elaborar um modelo, são considerados, no mínimo, dois elementos, a saber: (i) variável de transporte, ou seja, a variável que se deseja estudar o comportamento (por exemplo, peso médio transportado diariamente por uma empresa) e (ii) variáveis

explicativas ou série histórica do comportamento da variável da qual se deseja estudar o comportamento (por exemplo, renda *per capita*, produção industrial, etc.). Ainda, é necessário observar alguns aspectos, como: (i) as variáveis explicativas devem de fato estar relacionadas com o que se quer prever; (ii) as variáveis explicativas devem ter comportamento futuro passível de previsão com bom grau de certeza; (iii) os modelos devem fornecer resultados mais precisos possíveis. Para isso, há de ressaltar a necessidade de um perfeito ajuste entre as variáveis com a função especificada para explicar a demanda.

Outro aspecto importante ao se elaborar um modelo que estabeleça relações causais entre variáveis é quanto à sua capacidade preditiva. A capacidade preditiva do modelo gerado é um dos principais fatores e é fortemente relacionada com o conjunto de dados utilizados para criar o modelo. Assim, um algoritmo que gere modelos preditivos de maior acurácia utilizando um determinado conjunto de dados pode não ter o mesmo desempenho quando o conjunto de dados for diferente (WITTEN *et al.*, 2011; ALPAYDIN, 2010).

A regressão linear refere-se a uma tentativa de estabelecer uma equação matemática linear, que descreva o relacionamento entre duas variáveis. A finalidade de uma equação de regressão é estimar valores de uma variável, com base em valores conhecidos da outra (STEVENSON, 1981).

Os modelos de regressão linear generalizada (GLM) são definidos por Cordeiro e Demetrio (2013) como uma extensão dos modelos clássicos de regressão. Tal técnica foi concebida por Nelder e Wedderburn no ano de 1972. Em seus trabalhos, os autores unificaram um conjunto de técnicas estatísticas estudadas de forma segregada, formulando uma classe de modelos de regressão. Esses modelos envolvem uma variável dependente ou variável resposta univariada, variáveis explanatórias e uma amostra aleatória de  $n$  observações independentes, sendo que a variável dependente, componente aleatória do modelo, tem uma distribuição que pertence à família de distribuições normal ou gama, ou normal inversa para dados contínuos, binomial para proporções e Poisson ou binomial negativa para contagens. As variáveis explicativas entram na forma de uma estrutura linear, constituindo a componente sistemática do modelo.

Os GLMs apresentam a seguinte forma (equação 1):

$$Y = Z.\beta \pm \varepsilon \quad (1)$$

em que:

- $y$  é a componente aleatória do modelo;
- $z$  é uma matriz de dimensão  $n \times p$  de especificação do modelo (em geral a matriz de covariáveis  $X$  com um primeiro vector unitário), associada a um vector  $\beta = (\beta_1, \dots, \beta_p)$  de parâmetros, e;
- $\varepsilon$  é um vector de erros aleatórios com distribuição que se supõe  $N_n(0, \sigma^2 I)$ .

Para determinação de modelos GLMs, são obtidas estimativas de máxima verossimilhança dos parâmetros usando um algoritmo de mínimos quadrados ponderados iterativamente. Tais modelos estimam parâmetros, de modo que os parâmetros sejam ajustados e otimizados por meio das iterações (CORDEIRO e DEMETRIO, 2013).

A quantidade de variáveis a serem testadas depende dos objetivos do pesquisador. Para o caso de um grande número de variáveis independentes, a escolha do conjunto de variáveis regressoras a ser utilizado no modelo é uma decisão importante na modelagem (MONTGOMERY E RUNGER, 2014).

De acordo com Montgomery e Runger (2014), o problema consiste em selecionar um conjunto apropriado de regressores a partir de um conjunto que inclua todas as variáveis potencialmente importantes. Segundo os autores, nessas situações, deve ser realizada uma seleção de variáveis, cujo objetivo é filtrar as variáveis candidatas, obtendo o melhor subconjunto de variáveis para ser usado no modelo.

Dentre as técnicas de seleção de variáveis disponíveis, as mais empregadas são: (i) todas as regressões possíveis, (ii) regressão *stepwise*, (iii) seleção progressiva, e (iv) eliminação regressiva, que serão descritas a seguir (HAIR JR. *et al.*, 2014; MONTGOMERY E RUNGER, 2014).

- todas as regressões possíveis: esta técnica realiza combinações entre todas as variáveis e seleciona a melhor, considerando critérios de avaliação tais como o coeficiente de determinação ( $R^2$ ) e a soma dos quadrados do erro de previsão (SQEP);

- regressão stepwise: constrói iterativamente modelos de regressão adicionando ou removendo variáveis a cada passo, a partir de um modelo composto pela variável que apresenta a mais alta correlação com a variável resposta. O procedimento continua até que nenhuma variável possa ser adicionada ou removida do modelo;
- seleção progressiva: os regressores são adicionados no modelo, um de cada vez, até que não haja mais regressores candidatos que produzam um aumento significativo na soma dos quadrados da regressão;
- eliminação regressiva: realiza um procedimento inverso ao da técnica anterior, iniciando com um conjunto de todas as variáveis e realizando, a cada etapa, a remoção do regressor que apresentar menor Estatística-F.

Cabe ao pesquisador escolher a técnica mais apropriada e sua aplicação resultará em um conjunto de variáveis mais adequadas para compor o modelo. Caso o resultado indique que o modelo deva ser composto por um conjunto de duas ou mais variáveis, é necessário assegurar que estas não são correlacionadas entre si (HAIR JR. *et al.*, 2014).

Com base nas variáveis selecionadas e seus respectivos valores amostrais, devem ser gerados os modelos de regressão, estimando os parâmetros, o que pode ser feito com o auxílio do *software* R que fornece, juntamente com os resultados, as estatísticas necessárias para avaliar o modelo.

Os modelos são validados utilizando-se testes estatísticos. Para a validação das equações utilizam-se testes (parâmetros) estatísticos que validam as amostras em seus níveis de confiança com relação às variáveis escolhidas. Neste trabalho, um modelo só é considerado válido quando for confirmado: (i) teste z maior que 1,96 (5% de significância) para a estimação dos coeficientes das variáveis dependentes; (ii) rejeição da hipótese  $H_0$  (modelo teórico é adequado para representar o modelo em análise) do teste *residual deviance goodness of fit for Poisson regression* – GOF (p-valor maior que 0,05, para um nível de confiança de 95%); e (iii) não haver superdispersão para modelos baseados em função da família Poisson ( $\hat{c} < 2,0$ ). Esses parâmetros estatísticos serão discutidos nos parágrafos a seguir.

O teste Z é um teste de hipótese que utiliza conceitos estatísticos para rejeitar ou aceitar uma hipótese nula quando a estatística de teste segue uma distribuição normal e a variância da



população é conhecida. O teste z é capaz de determinar se a diferença entre a média da amostra e da população é grande o suficiente para ser significativa estatisticamente (LANNA, 2004).

Outra forma de se investigar o ajuste do GLM é por meio do parâmetro GOF *deviance* (*Goodness of fit*). O desvio é uma medida da qualidade do ajuste de um modelo linear generalizado. Essa técnica descreve quão bem o modelo se encaixa em um conjunto de observações. Índices GOF resumem a discrepância entre os valores observados e os valores esperados em um modelo estatístico. Estatísticas GOF são índices com distribuições de amostragem conhecidas, geralmente obtidos usando métodos assintóticos, que são usados em testes de hipóteses estatísticas. Nessa técnica números mais altos indicam pior ajuste dos modelos estruturados (OLIVARES e FORERO, 2010).

O valor-p está relacionado ao grau de confiança dos resultados obtidos em uma pesquisa estatística. Na prática, o valor-p representa a chance ou a probabilidade do resultado observado na pesquisa, utilizando variáveis específicas, ser devido ao acaso, ou a outros fatores não observados e não aos fatores que estão sendo estudados. O valor-p deve estar abaixo do nível de significância pré-estabelecido para garantir a confiabilidade dos resultados, esse parâmetro é o menor nível de significância com que se rejeita uma hipótese nula (WASSERSTEIN e LAZAR, 2016).

Denotado como teste  $\chi^2$ , o teste qui-quadrado é utilizado para determinar se existe diferença significativa entre os valores esperados em um modelo preditivo e os valores observados em um conjunto de dados. O teste qui-quadrado é calculado a partir da soma de erros quadráticos, ou através da variância da amostra. As estatísticas de teste que seguem uma distribuição qui-quadrada surgem de uma suposição de que os dados são independentes e normalmente distribuídos (SATORRA e BENTLER, 2001).

Os GLMs Poisson e binomial, não possuem um parâmetro para a variância, que geralmente é mantida com valor fixo 1, ou seja, a média é igual à variância. No entanto, é preciso verificar se essa suposição é apropriada, estimando o parâmetro de superdispersão ( $c$ -hat). O termo superdispersão quer dizer que a variância da resposta observada Y excede a variância nominal estabelecida pelo modelo, conforme foi definido por Hinde e Demétrio (1998). A

superdispersão acarreta desvios-padrão incorretos, podendo estar seriamente subestimados; alterações na *deviance*, associadas com os termos do modelo, muito grandes as quais acarretariam a seleção de modelos excessivamente complexos, com interpretações incorretas e algumas previsões imprecisas. Reconhecer dados com superdispersão é primordial para que sejam tomadas precauções e medidas capazes de garantir uma estimação com menor margem de erro, ou seja, com maior segurança.

O teste  $c\text{-hat}$  é um parâmetro que verifica se há superdispersão entre a variância observada e a variância dos resultados de um modelo preditivo. Nessa técnica, valores de  $c\text{-hat}$  maiores do que 1 indicam superdispersão (variância > média), mas obtendo-se valores muito maiores que 1, ou seja, maiores do que 4, indicam falta de ajuste. Em casos de subdispersão,  $c\text{-hat}$  menores do que 1, recomenda-se manter o valor de  $c\text{-hat}$  para 1. No entanto, valores de  $c\text{-hat}$  muito menores do que 1 também podem indicar falta de ajuste e que uma alternativa de modelo (e distribuição) deve ser investigada (FLETCHER, 2012; VENABLES e RIPLEY, 2002; FARRINGTON, 1996; McCULLAGH e NELDER, 1989).

O critério de informação de Akaike (AIC) é utilizado para classificar os modelos (AKAIKE, 1974). Alternativamente, Hurvich e Tsai (1995) propuseram o AICc indicado para pequenas amostras com distribuição normal, sendo um critério mais adequado para a escolha de um modelo (DAVISON, 2001). Burnham e Anderson (2002) recomendam utilizar o AIC para selecionar modelos quando o número de observações é pelo menos 40 vezes maior que o número de parâmetros.

A seleção de um modelo pode ser realizada com base no critério de informação de Akaike (AIC), que é uma medida relativa da qualidade de ajuste de um modelo de evento aleatório. O critério oferece uma medida relativa das informações perdidas, quando um determinado modelo é utilizado para descrever um fenômeno real (AKAIKE, 1974). O AIC é um critério que avalia o equilíbrio entre a qualidade do ajuste do modelo paramétrico estimado pelo método da máxima verossimilhança e a complexidade desse modelo.

Emiliano (2009) afirma que o AIC não testa hipóteses. Ele é uma ferramenta para seleção de modelos que sejam gerados a partir de um mesmo conjunto de dados, mas não considera significância ou valor-p. A partir de modelos (equações) criados usando um mesmo conjunto

de dados, pode-se classificá-los de acordo com o seu AIC, sendo que quanto menor o valor de AIC, melhor será o modelo. A partir do valor de AIC podem-se classificar modelos, mas não se deve atribuir um valor acima do qual um determinado modelo é rejeitado.

Somente a análise do valor do AIC para um determinado conjunto de dados não tem significado algum. O AIC torna-se útil quando são comparados diversos modelos, porém, o método não os avalia, apenas os compara e os classifica (EMILIANO, 2009).

Neste trabalho, considerou-se o AICc em vez do critério de informação de Akaike (AIC) para evitar superestimação, uma vez que o tamanho da amostra (4 observações) é pequena em comparação com o número de parâmetros estimados (BURNHAM *et al.*, 2011). O AICc é essencialmente o AIC com um termo de penalidade extra para o número de parâmetros. Quando o tamanho da amostra tende ao infinito o termo de penalidade extra converge para zero, logo AICc converge para AIC (BURNHAM e ANDERSON, 2002). Em termos gerais, o AICc considera o *trade-off* entre o melhor ajuste e a simplicidade do modelo, sendo um bom indicador para a escolha entre diferentes modelos com a mesma variável dependente. Essas análises foram desenvolvidas com auxílio do *software* R.

Após a validação de um modelo, deve-se analisar o seu poder de predição. Para tanto, utiliza-se a validação cruzada (VC) que é amplamente difundida devido à sua simplicidade. Conforme Kohavi (1995), a validação cruzada é uma técnica para avaliar a capacidade de generalização de um modelo, a partir de um conjunto de dados. O objetivo é estimar, na prática a precisão do modelo, ou seja, o seu desempenho para um novo conjunto de dados. A VC tem como premissa dividir os dados, uma ou várias vezes, para estimar o risco de cada algoritmo. Parte dos dados (a amostra de treinamento) é usada para treinar cada algoritmo e a parte restante (a amostra de validação) é usada para estimar o risco do algoritmo. A validação cruzada seleciona o modelo com o menor risco estimado. Comparado com outros métodos de validação, a validação cruzada evita sobreposição, uma vez que a amostra de treinamento é independente da amostra de validação. As formas mais usadas para realizar o particionamento dos dados são: o método *holdout*, o *k-fold* e o *leave-one-out* (KOHAVI, 1995).

A capacidade preditiva do modelo pode ser determinada a partir da raiz quadrada do erro quadrático médio (REQM) como resultado da validação cruzada (HYNDMAN, 2006;

ARLOT e CELISSE, 2010). Neste estudo, para análise da capacidade preditiva dos modelos desenvolvidos, foi considerada a técnica *Leave-one-out Cross Validation* (LOOCV) (ARLOT e CELISSE, 2010). Essa técnica é a mais indicada quando se tem um número pequeno de observações. A técnica consiste em, a partir de um conjunto de dados - variáveis independentes ( $x$ ) e variáveis dependentes ( $y$ ), gerar um gráfico de dispersão e uma equação linear, ou seja, o modelo de regressão. O método LOOCV consiste na exclusão de um par de variáveis de  $(x_1, y_1)$  que é chamada conjunto de validação. Os outros pares de variáveis são mantidos e gera-se uma nova equação (que desconsidera o par  $(x_1, y_1)$  excluído), este conjunto é denominado conjunto de treinamento. Com o novo modelo calcula-se o valor de  $\hat{y}_1$  a partir do valor original de  $x_1$ . O processo segue com a retirada do próximo par de variáveis  $(x_2, y_2)$  mantendo-se as demais variáveis e calculando o valor de  $\hat{y}_2$ . Devido à exclusão do par de variáveis que não foi utilizado no processo de ajuste do modelo, o erro quadrático  $EQ_1=(y_1-\hat{y}_1)^2$  fornece uma estimativa pouco tendenciosa para o teste de erro. Este método não é recomendado quando se tem um conjunto de dados muito extenso, pois, serão realizadas tantas iterações quanto o número de observações (OLIVEIRA *et al.*, 2017; ARLOT e CELISSE, 2010).

Ainda sobre a técnica, repetem-se as iterações do método até que todas as variáveis sejam retiradas, novas equações sejam geradas e os erros quadráticos sejam calculados. Ao final do processo calcula-se a raiz do erro quadrático médio (REQM), dado pela raiz da média dos  $n$  erros quadráticos calculados no processo. O modelo que apresenta o menor valor de REQM é o modelo com melhor desempenho preditivo.

A análise de correlação, o desenvolvimento dos modelos e todos os testes utilizados nesse estudo foram realizados com auxílio do *software* R.

### **3.4 Análise comparativa conceitual**

Concluída a etapa de modelagem e avaliação dos modelos, faz-se uma análise comparativa conceitual do esforço metodológico para coleta de dados, construção dos modelos, escolha dos métodos de modelagem e avaliação da qualidade dos modelos para o tipo de empreendimento estudado.

## **4 RESULTADOS E DISCUSSÕES**

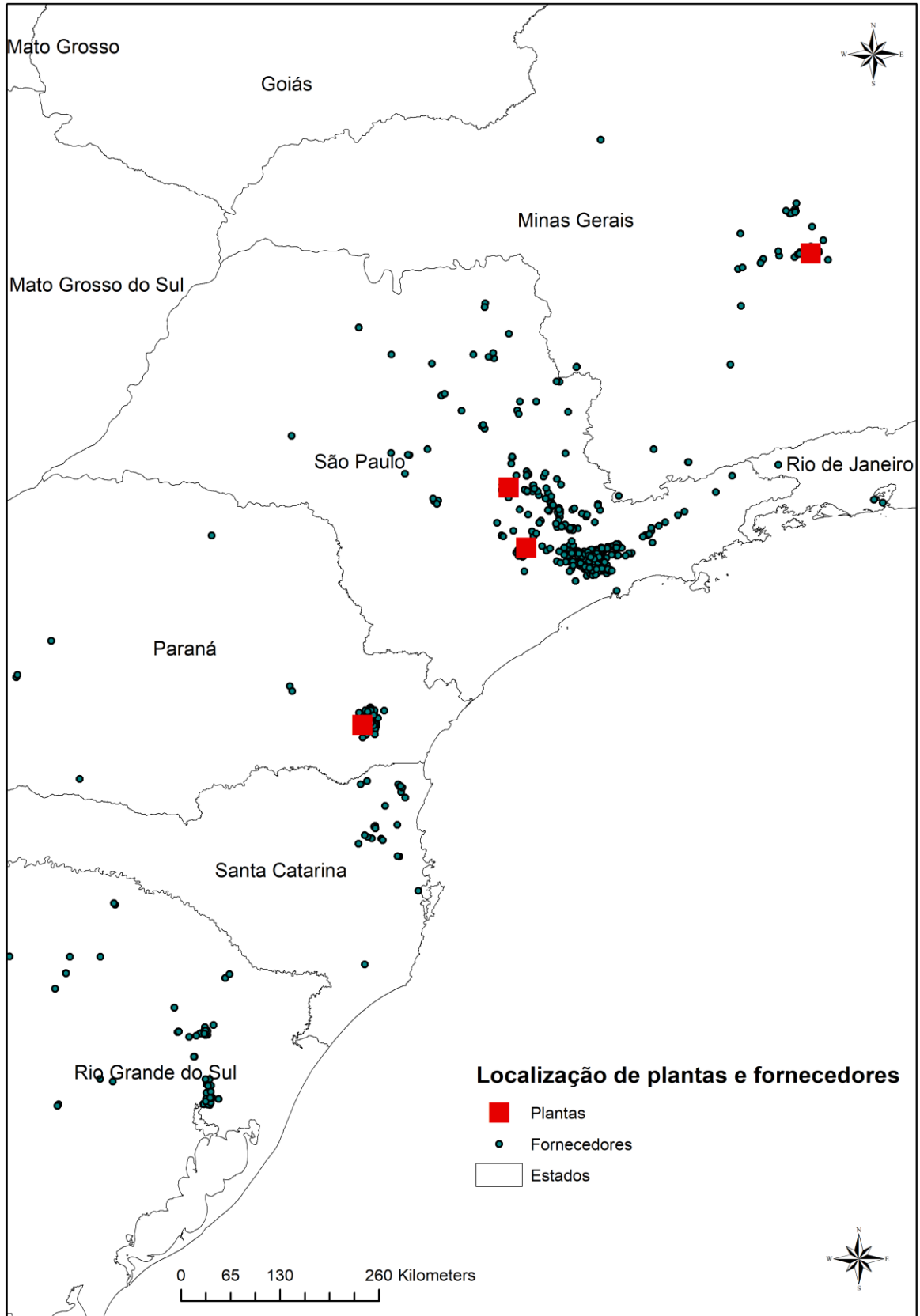
Neste capítulo são caracterizadas as indústrias em investigação, assim como os modelos desenvolvidos para estimação dos fluxos de atração de insumos de suprimentos e respectivas viagens para indústrias montadoras de máquinas agrícolas e rodoviárias. Esses modelos são também avaliados quanto ao ajuste estatístico e à acurácia preditiva.

### **4.1 Caracterização econômica e territorial**

Foi considerado como objeto de estudo o fluxo de suprimentos de quatro indústrias montadoras de máquinas agrícolas e rodoviárias localizadas nos municípios de Contagem, Minas Gerais; Curitiba, Paraná; Piracicaba e Sorocaba no Estado de São Paulo – Brasil. O fluxo de suprimentos de insumos para produção de máquinas agrícolas e rodoviárias dessas quatro indústrias é constituído de 592 fornecedores que estão distribuídos em diversas cidades brasileiras nos Estados de Minas Gerais (MG), Paraná (PR), Rio de Janeiro (RJ), Rio Grande do Sul (RS), Santa Catarina (SC) e São Paulo (SP), conforme Figura 4.1.

Segundo os dados da ANFAVEA, 2018, às quatro empresas, foco desse estudo, produziram juntas, nos anos 2016 e 2017, 21,8% e 30,5% respectivamente, do total de máquinas agrícolas e rodoviárias produzidas no Brasil nesse período. Elas se destacam como empresas multinacionais de grande porte cujos nomes não serão divulgados devido ao sigilo da pesquisa e ao objetivo aqui ser plenamente acadêmico científico. Nessa pesquisa elas serão denominadas como: Contagem-CE, Curitiba-AG, Piracicaba-AG e Sorocaba-AG.

Figura 4.1: Localização das indústrias e fornecedores.



Quanto às principais empresas do seguimento de máquinas agrícolas e rodoviárias no Brasil, tem-se um grupo de sete fabricantes: AGCO, Agrale, Caterpillar, CNH (Case New, Holland), John Deere, Komatsu e Volvo *Construction Equipment* que juntas são responsáveis pela produção de tratores de rodas, tratores de esteira, colheitadeiras de grãos, colhedoras de cana e retroescavadeiras, conforme ilustra a Tabela 4.1. As empresas possuem 22 fábricas no Brasil e são responsáveis por empregar mais de 18 mil pessoas segundo a Associação Nacional dos Fabricantes de Veículos Automotores - ANFAVEA (2018).

Tabela 4.1: Empresas fabricantes de máquinas agrícolas e rodoviárias.

Empresas	Produtos				
	Tratores de rodas	Tratores de esteiras	Colheitadeiras de grãos	Colhedoras de Cana	Retroescavadeiras
AGCO	■		■		
Agrale	■				
Caterpillar		■			■
CNH (Case, New Holland)	■		■	■	■
John Deere			■	■	■
Komatsu		■			
Volvo					■

Fonte: Adaptado de Anuário da indústria Automobilística Brasileira, ANFAVEA (2018).

Ainda segundo a ANFAVEA (2018), no primeiro semestre do ano de 2018, a produção nesse seguimento atingiu 12 mil unidades representando um aumento de 0,9% frente às 11,9 mil unidades produzidas nos três primeiros meses do ano de 2017. No que tange as exportações, 2,9 mil produtos foram enviados para outros países o que representou um crescimento de 31,7% ante as 2,2 mil de igual período do ano de 2017. Quando analisado o desempenho nas vendas até abril do ano de 2018, ainda se percebe um desempenho abaixo de 2017, entretanto, com tendência de alta: 7,5 mil unidades foram comercializadas em 2018, baixa de 19,1% sobre as 9,3 mil vendidas no ano de 2017. As projeções do setor para o encerramento do ano de 2018 indicaram alta de 11,8% na produção, acima de 61 mil unidades, 9,9% nas exportações, que devem registrar 15,5 mil máquinas, e 3,7% nas vendas internas, chegando a 46 mil unidades.

Conforme o Censo 2010 (IBGE, 2010), em relação aos dados de área e população residente de cada município onde se localizam as indústrias analisadas neste estudo, nota-se que o

município com maior densidade populacional é a cidade de Curitiba-PR, seguida por Contagem-MG, Sorocaba-SP e por último, Piracicaba-SP com a menor densidade populacional, conforme Tabela 4.2.

Tabela 4.2: Dados demográficos dos municípios onde são localizadas as indústrias analisadas.

Município	Área (km <sup>2</sup> )	População (unid.)	Densidade populacional (hab./m <sup>2</sup> )
Contagem-MG	195,045	603.442	3.093,86
Curitiba-PR	435,036	1.751.907	4.027,04
Piracicaba-SP	1.378,07	364.571	264,55
Sorocaba-SP	450,382	586.625	1.302,51

Fonte: IBGE (2010).

Na Tabela 4.3 são apresentados os dados que caracterizam cada instalação industrial.

Tabela 4.3: Caracterização das instalações industriais analisadas.

Indicadores	Indústrias			
	Contagem-CE	Curitiba-AG	Piracicaba-AG	Sorocaba-AG
Área construída total (m <sup>2</sup> )	120.000	480.000	83.657	180.000
Área para operações logística (m <sup>2</sup> )	34.414	57.000	9.902	35.640
Funcionários total	857	2.100	200	1.800
Funcionários da logística	128	347	58	168
Produção efetiva/dia	35	118	3	25
Tipo	Grande porte	Grande porte	Grande porte	Grande porte
Produtos fabricados	Tratores de rodas, tratores de esteiras e retroescavadeiras	Componentes, tratores agrícolas e colheitadeiras	Colhedoras de cana e plantadeiras	Colheitadeiras

Fazendo-se uma análise preliminar dos dados apresentados na Tabela 4.3, observa-se que quanto maior a área construída da unidade industrial, mais funcionários ela possui. Nesse sentido, em ordem decrescente, tem-se a Indústria Curitiba-AG com a maior área e números de funcionários, seguida pela Indústria Sorocaba-AG, Contagem-CE e a Indústria Piracicaba-AG com 83.657 m<sup>2</sup> de área e 200 funcionários.

Referente à produção diária, percebe-se que não há uma relação diretamente proporcional com o número de funcionários e com a área construída de cada unidade industrial. Isso se dá,



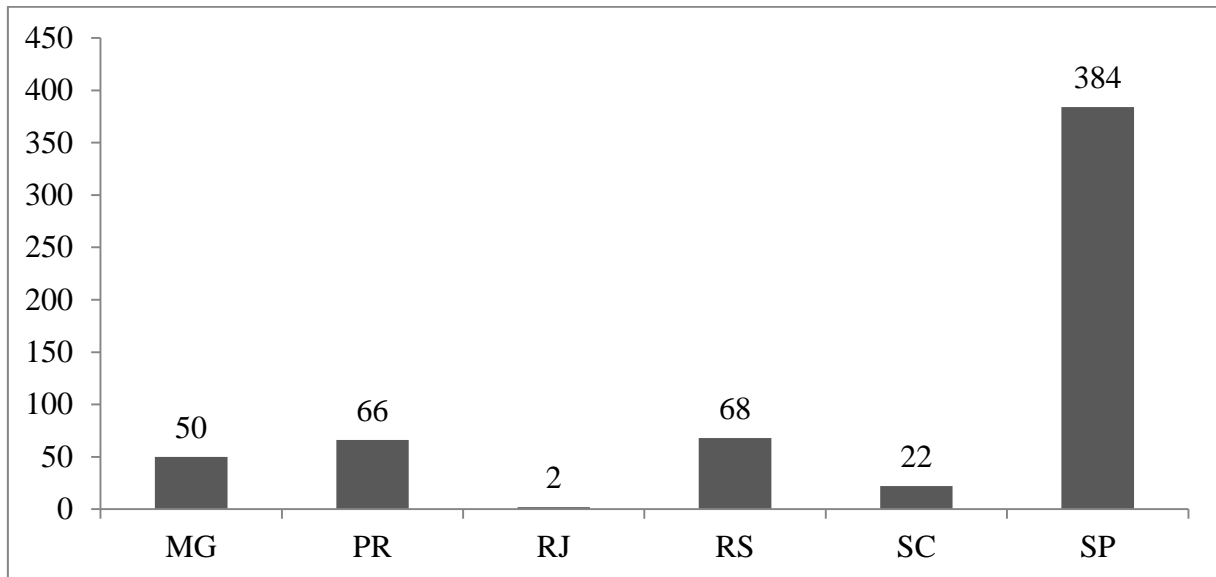
pois embora a indústria Curitiba-AG tenha a maior área, número de funcionários e capacidade de produção diária de máquinas, a indústria Contagem-CE tem área menor e número de funcionários inferior que a indústria Sorocaba-AG, por exemplo, mas com capacidade de produção por dia maior. Assim, para análise de correlação entre as variáveis explicativas, é apresentada nesse trabalho a matriz de correlação considerando as variáveis explicativas e dependentes. Essa análise sugere que é possível que estrutura produtiva da indústria, no que se refere a equipamentos, tecnologia embarcada, conhecimento dos funcionários a respeito do processo, assim como às características do produto, podem influenciar mais diretamente a capacidade de produção diária do que a área disponível e quantidade de funcionários. Por fim, a unidade Piracicaba-AG tem a menor área, número de funcionários e capacidade de produção por dia entre as indústrias pesquisadas.

Quanto ao tipo, as indústrias são pesadas e produzem máquinas agrícolas e rodoviárias. Elas são integradas por regionais de vendas e concessionárias, instaladas nas principais cidades e capitais do país. Além disso, exportam máquinas para toda a América Latina bem como para outros países.

Foi realizada então uma análise descritiva dos dados referentes às origens das entregas e da quantidade de carga atraída para as unidades industriais analisadas.

Conforme destacado no trabalho, as entregas têm origem em 592 fornecedores localizados nos Estados de MG, PR, RJ, RS, SC e SP e como destino as indústrias Contagem-CE, Curitiba-AG, Piracicaba-AG e Sorocaba-AG. Na Figura 4.2 são apresentados, em percentuais, os estados onde os fornecedores de insumos para as unidades industriais analisadas estão localizados. Observa-se que a maior parte (64,9%) desses fornecedores está localizada no Estado de São Paulo.

Figura 4.2: Quantidade de fornecedores por Estado Brasileiro.



Na Tabela 4.4 é apresentada a quantidade de fornecedores de cada unidade industrial localizados em cada estado, considerando o suprimento de insumos para produção tendo como base o ano de 2017.

Tabela 4.4: Quantidade de fornecedores em cada Estado por indústria.

UF Origem	Quantidade de fornecedores por Estado	Contagem-CE	Curitiba-AG	Piracicaba-AG	Sorocaba-AG
MG	50	29	25	15	19
PR	66	20	45	29	40
RJ	2	1	0	0	1
RS	68	23	50	33	29
SC	22	3	18	8	9
SP	384	138	170	153	153
<b>Total de Fornecedor</b>	<b>592</b>	<b>214</b>	<b>308</b>	<b>238</b>	<b>251</b>

Conforme apresentado na Tabela 4.4, dos 592 fornecedores, 214 forneciam insumos para a indústria Contagem-CE, 308 para a indústria Curitiba-AG, 238 para a indústria Piracicaba-AG e 251 fornecedores fornecem insumos para as linhas de produção da indústria Sorocaba-AG. Percebe-se que a indústria Contagem-CE e Sorocaba-AG são as únicas unidades, dentre as analisadas, que possuem um fornecedor no Estado do Rio de Janeiro – Brasil. Outra análise

possível a partir dessa tabela, é que para todas as indústrias há uma concentração de fornecedores no Estado de São Paulo.

Na Tabela 4.5 é apresentado o número de entregas tendo como origem os respectivos Estados para o ano de 2017.

Tabela 4.5: Número de entregas por Estado de origem (Ano 2017).

UF Origem	Contagem-CE	Curitiba-AG	Piracicaba-AG	Sorocaba-AG	Total Geral
MG	1.234	1.529	207	140	3.110
PR	210	2.884	711	1.563	5.368
RJ	2	0	0	1	3
RS	263	1.409	673	502	2.847
SC	15	340	73	64	492
SP	3.039	2.880	4.617	2.592	13.128
<b>Total de Entregas</b>	<b>4.763</b>	<b>9.042</b>	<b>6.281</b>	<b>4.862</b>	<b>24.948</b>

Analisando a Tabela 4.5, nota-se que 63,8% do total das entregas atraídas pela indústria Contagem-CE tem como origem o Estado de SP e em segundo lugar o Estado de MG com 25,9% de representatividade. Para Curitiba-AG, 31,9% das entregas tem como origem os Estados do PR e SP. Para Piracicaba-AG, 73,5% das entregas tem como origem o Estado de SP e 11,3% o Estado do PR. Para Sorocaba-AG, 53,3% das entregas tem como origem o Estado de SP e 32,1% o Estado do PR. No somatório total das entregas, 52,6% tem como origem o Estado de SP e 21,5% o Estado do PR.

Na Tabela 4.6 é apresentada a quantidade de carga atraída, em quilogramas (kg) tendo como origem os respectivos Estados para o ano de 2017.

Tabela 4.6: Quantidade de carga movimentada em quilogramas (kg) por Estado de origem (Ano 2017).

UF Origem	Contagem-CE	Curitiba-AG	Piracicaba-AG	Sorocaba-AG	Total Geral
MG	14.354.150	18.037.796	725.655	565.509	33.683.109
PR	2.931.979	34.862.333	6.178.425	15.032.853	59.005.590
RJ	696	0	0	236	932
RS	3.264.726	12.971.625	9.135.136	5.192.325	30.563.813
SC	104.508	3.346.850	293.789	217.972	3.963.120
SP	45.104.586	24.236.455	63.008.625	23.368.201	155.717.867
<b>Total de Carga (kg)</b>	<b>65.760.646</b>	<b>93.455.059</b>	<b>79.341.631</b>	<b>44.377.096</b>	<b>282.934.431</b>

Na Tabela 4.6, evidencia-se que 68,6% do total de quantidade de carga atraída pela indústria Contagem-CE tem como origem o Estado de SP. Em segundo lugar tem-se o Estado de MG com 21,8%. Para Curitiba-AG, 37,3% da quantidade de carga atraída tem como origem o Estado do PR seguido pelo Estado de SP com 25,9% e MG com 19,3%. Para Piracicaba-AG, 79,4% das quantidades de carga atraídas tem como origem o Estado de SP e 11,5% o Estado do RS. Para Sorocaba-AG, 52,7% da quantidade de carga atraída tem como origem o Estado de SP e 33,9% o Estado do PR. No somatório total das viagens de veículo de carga, 55,0% tem como origem o Estado de SP e 20,9% o estado do PR.

## 4.2 Organização das informações

Com o objetivo de realizar uma análise geral dos dados no que se refere à estrutura, tendência e sazonalidade do número de entregas e da quantidade de carga atraída pelas unidades industriais, foram elaboradas, na base mensal, representações tabulares e discutidas a seguir.

Na Tabela 4.7 tem-se a representação dos dados do número de entregas atraídas no ano de 2017 (janeiro a dezembro) para às quatro unidades industriais pesquisadas.

Tabela 4.7: Número de entregas mensais atraídas pelas unidades industriais no ano de 2017.

Ano	Mês	Contagem-CE	Curitiba-AG	Piracicaba-AG	Sorocaba-AG	Total Entregas Atraídas
2017	Janeiro	175	441	243	249	1.108
2017	Fevereiro	306	484	422	326	1.538
2017	Março	351	627	446	317	1.741
2017	Abril	505	951	469	384	2.309
2017	Maiο	546	1.073	628	496	2.743
2017	Junho	428	933	686	558	2.605
2017	Julho	528	1.037	817	458	2.840
2017	Agosto	595	1.094	789	596	3.074
2017	Setembro	459	836	627	444	2.366
2017	Outubro	398	763	501	399	2.061
2017	Novembro	276	658	358	341	1.633
2017	Dezembro	196	145	295	294	930
<b>Total Entregas Atraídas</b>		<b>4.763</b>	<b>9.042</b>	<b>6.281</b>	<b>4.862</b>	<b>24.948</b>

Conforme apresentado na Tabela 4.7, o número total de entregas de carga atraídas pelas quatro indústrias no período analisado foi de cerca de 25 mil entregas (janeiro a dezembro de

2017). Desse total, a indústria Curitiba-AG atraiu o maior número de entregas dentre as unidades industriais, o que representou 36% do total (24.948) em 2017. Em segundo lugar, tem-se a indústria Piracicaba-AG que atraiu 25% do total de entregas, seguida por Sorocaba-AG com 20% e por fim, a indústria Contagem-CE com 19% do total de entregas atraídas em 2017.

Fazendo-se uma comparação entre as indústrias, pode-se verificar que os picos de entregas foram: (i) para a indústria Contagem-CE e Curitiba-AG, os meses de agosto e maio, sendo para Contagem-CE 595 e 546 entregas atraídas e para Curitiba-AG 1.094 e 1.073 entregas atraídas respectivamente; (ii) para Piracicaba-AG os meses de julho e agosto com 817 e 789 viagens respectivamente, e por fim, (iv) para Sorocaba-AG os meses de agosto com 596 e junho com 558 entregas. Fazendo-se uma análise do valor modal, tem-se o mês de agosto com maior frequência de pico de entregas entre as unidades industriais estudadas.

Ainda analisando a Tabela 4.7, temos para todas as indústrias os meses de janeiro e dezembro com o menor número de entregas. Uma hipótese para esse fenômeno seria o fato de janeiro e dezembro serem meses típicos de férias coletivas no setor.

Na Tabela 4.8 tem-se a representação dos dados da quantidade de cargas atraídas no ano de 2017 (janeiro a dezembro) para às quatro unidades industriais pesquisadas.

Tabela 4.8: Quantidade de carga (kg) atraídas mensalmente pelas unidades industriais no ano 2017.

Ano	Mês	Contagem-CE	Curitiba-AG	Piracicaba-AG	Sorocaba-AG	Total Carga (kg) Atraídas
2017	Janeiro	2.313.747	5.406.754	3.024.611	2.011.546	12.756.658
2017	Fevereiro	4.172.817	5.748.634	5.273.861	2.622.655	17.817.967
2017	Março	4.829.731	7.085.697	5.335.066	3.115.970	20.366.464
2017	Abril	6.928.398	9.705.963	6.147.209	3.802.491	26.584.061
2017	Maiο	8.083.736	9.551.416	7.831.241	4.651.593	30.117.985
2017	Junho	6.265.800	10.737.487	9.176.369	5.478.413	31.658.069
2017	Julho	6.937.864	10.984.135	10.321.953	4.513.947	32.757.899
2017	Agosto	7.957.368	11.371.512	10.715.422	5.777.221	35.821.523
2017	Setembro	6.212.496	8.444.113	7.886.290	3.596.522	26.139.421
2017	Outubro	5.645.778	7.535.113	5.897.140	3.450.438	22.528.469
2017	Novembro	4.004.047	6.059.262	4.145.837	2.908.232	17.117.378
2017	Dezembro	2.408.864	824.972	3.586.632	2.448.068	9.268.536
<b>Total Carga (kg) Atraídas</b>		<b>65.760.646</b>	<b>93.455.059</b>	<b>79.341.631</b>	<b>44.377.096</b>	<b>282.934.431</b>

Conforme apresentado na Tabela 4.8, às quatro unidades industriais atraíram juntas, no ano de 2017 (janeiro a dezembro) pouco mais do que 282,9 mil toneladas de carga, referentes aos insumos para produção das máquinas agrícolas e rodoviárias. Do total de insumos atraídos pelas unidades industriais, a indústria Curitiba-AG atraiu a maior quantidade de carga (33% do total). Em segundo lugar, tem-se a indústria Piracicaba-AG que atraiu 28% da carga total. Em terceiro, a indústria Contagem-CE, atraiu 23% do total de carga movimentada nos fluxos de suprimento e, por fim, a indústria Sorocaba-AG, que atraiu 16% do total da quantidade de carga atraída no ano de 2017.

Fazendo-se uma comparação entre as unidades industriais, no que se refere-se o valor modal de quantidade de carga atraída, tem-se o mês de agosto com maior frequência de pico de quantidade de carga entregue.

Ainda sobre os dados de entrada, verificou-se que os meses de janeiro e fevereiro, novembro e dezembro eram meses com baixo número de entregas e de maneira análoga, da quantidade de carga movimentada no suprimento. Visando identificar períodos críticos típicos, com intensidade que representasse a criticidade da demanda, mas que não refletissem situações pouco frequentes, adotou-se como premissa eliminar os meses janeiro, fevereiro, novembro e dezembro das séries históricas das análises. Assim, para caracterização das entregas diárias

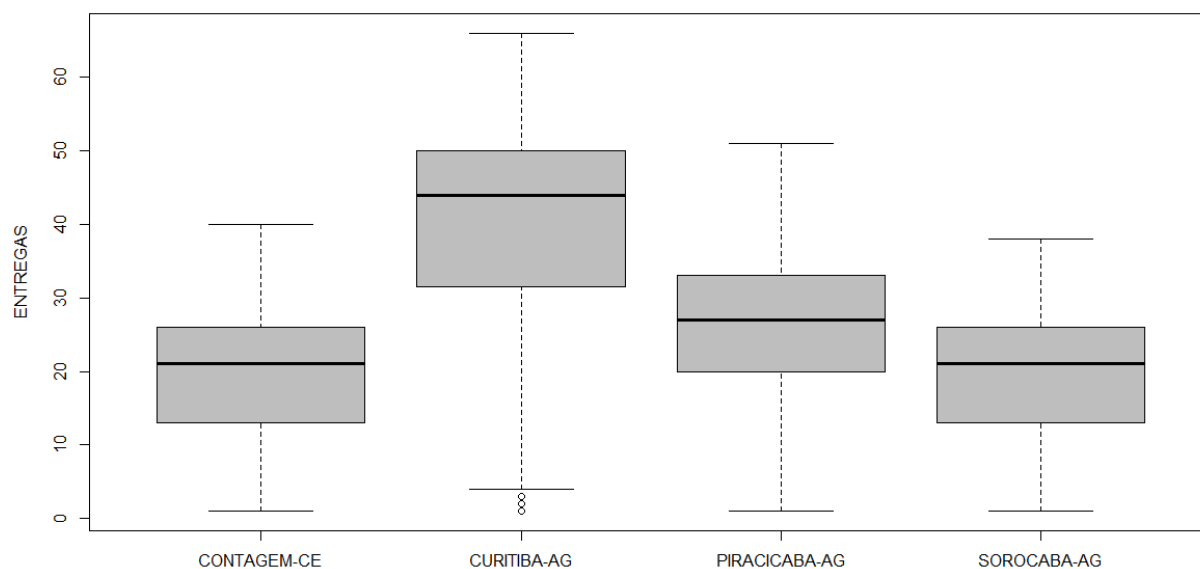
críticas típicas (em frequência e em quantidade de carga), foram considerados os meses de março a outubro, do ano de 2017 como dados representativos para os objetivos deste estudo.

Ainda como análise preliminar das variáveis dependentes, foram analisadas, além da sazonalidade, a estrutura dos dados, por meio de medidas de tendência central e dispersão. Para tanto, foram gerados *box-plot* das variáveis, discutidos na sequência. Na Tabela 4.9 são apresentadas as estatísticas descritivas para o número de entregas de carga atraídas diariamente para cada unidade industrial, considerando os meses de março a outubro do ano de 2017. Na Figura 4.3 são apresentados os respectivos *box-plot* que ilustram essa análise, com o objetivo de realizar uma análise comparativa sobre a posição, concentração, assimetria, dispersão e sobre possíveis *outliers* (medidas discrepantes) referente aos dados de número de entregas atraídas pelas quatro unidades industriais analisadas no recorte temporal.

Tabela 4.9: Dados de entregas diária atraída pelas indústrias de março a outubro de 2017.

Indústrias	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Moda	Amplitude interquartil	Desvio-padrão
CONTAGEM-CE	1	13,00	21	19,43	26,00	40	26	13	9,3
CURITIBA-AG	1	31,75	44	37,31	50,00	66	44	18,25	18,2
PIRACICABA-AG	1	20,00	27	25,31	33,00	51	27	13	10,5
SOROCABA-AG	1	13,00	21	18,63	26,00	38	21	13	9,4

Figura 4.3: Box-plot – Entregas diária de março a outubro de 2017.



Por meio da Tabela 4.9 e da Figura 4.3, percebe-se, quanto à posição dos dados, representado pela mediana, que a indústria Curitiba-AG tem mediana do número de entregas diária igual a 44, valor superior as demais indústrias que não ultrapassam 27 entregas na mediana.

Quanto à concentração dos dados, pode-se observar que a indústria Curitiba-AG tem maior amplitude interquartílica (AIQ) onde 50% das entregas atraídas tem valores entre 1º quartil, 31,75 e 3º quartil valor 50 o que equivale a uma amplitude de 18,25. As demais indústrias têm valor de amplitude interquartílica igual a 13.

No que se refere à assimetria dos dados, nota-se que as medianas não são equidistantes dos valores mínimo e máximo. Nesse caso há indícios de que os dados se distribuem de modo assimétrico. Comparando as indústrias, percebe-se que há maior simetria nos dados da indústria Piracicaba-AG. Essa informação é importante para identificação dos modelos de regressão que poderão ser mais adequados para relacionais variáveis dependentes e independentes.

Ainda, na análise da dispersão dos dados em análise, verifica-se que a indústria Curitiba-AG tem a maior dispersão entre as empresas analisadas.

Por fim, os *outliers* foram desconsiderados para as análises de regressão. Foram identificados três valores atípicos nos dados da indústria Curitiba-AG .

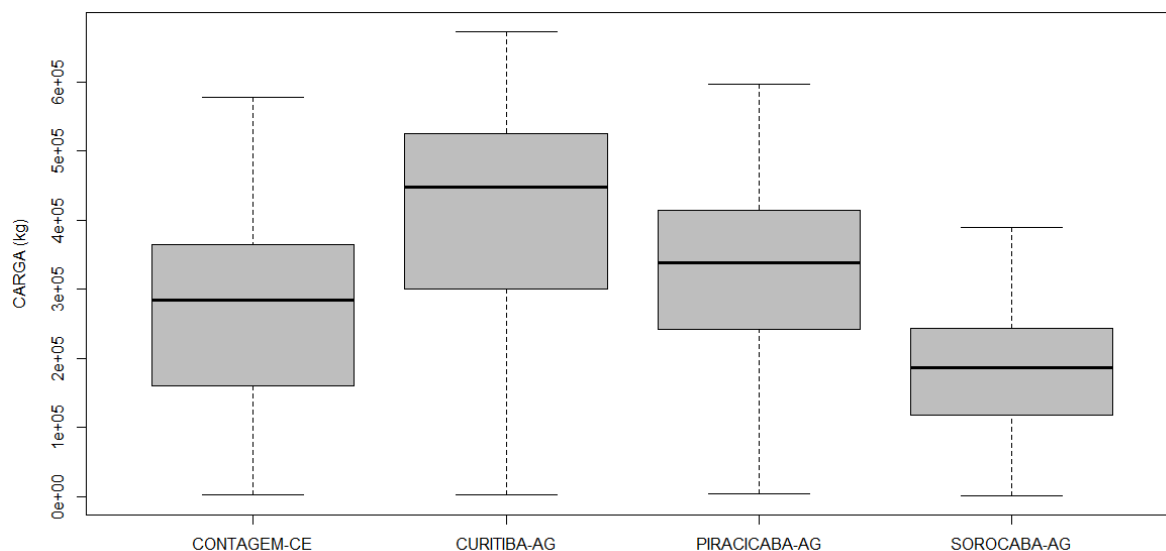
Na Tabela 4.10 são apresentadas as estatísticas descritivas para as quantidades de carga atraídas mensalmente considerando o ano 2017 e as unidades produtivas em análise. Na Figura 4.4 são apresentados os respectivos *box-plot* que ilustram essa análise.

Tabela 4.10: Dados de carga (kg) atraída por dia pelas unidades industriais de março a outubro de 2017.

<b>Indústrias</b>	<b>Mínimo</b>	<b>1º Quartil</b>	<b>Mediana</b>	<b>Média</b>	<b>3º Quartil</b>	<b>Máximo</b>	<b>Amplitude interquartil</b>	<b>Desvio-padrão</b>
CONTAGEM-CE	3.045	160.559	284.795	269.698	364.110	576.713	203.551	136.377,6
CURITIBA-AG	2.030	304.045	448.104	381.275	523.935	672.784	219.890	193.150,9
PIRACICABA-AG	3.471	242.581	337.533	319.925	412.671	596.300	170.090	142.583,4
SOROCABA-AG	885	119.165	187.007	175.442	242.516	389.164	123.351	94.597



Figura 4.4: Box-plot – Carga (kg) movimentada por dia de março a outubro de 2017.



Por meio da Tabela 4.10 e da Figura 4.4, também é realizada uma análise comparativa com o objetivo de fornecer informações sobre a posição, concentração, assimetria, dispersão e sobre possíveis *outliers* referente aos dados de quantidade de carga atraída pelas quatro unidades industriais analisadas.

Percebe-se, quanto à posição dos dados representados pela mediana, que a unidade Curitiba-AG apresenta um valor maior de mediana, o que corresponde a 448.104 kg de carga atraída por dia no período analisado. O menor valor de mediana foi encontrado para a indústria Sorocaba-AG (187.007).

Quanto à concentração dos dados, percebe-se maior amplitude interquartílica para aos dados da indústria Curitiba-AG com valores entre 1º quartil, 304.045 kg e 3º quartil 523.935 kg o que equivale a uma amplitude de 219.890. A menor amplitude nessa amostra fica com a indústria Sorocaba-AG (123.351).

Nota-se assimetria entre os dados sendo que a mediana encontrada para todas as unidades industriais não são equidistantes em relação aos valores mínimo e máximo.

Ainda, quanto à dispersão dos dados, verifica-se que a indústria Curitiba-AG é a que apresenta maior dispersão. Por fim, não foram identificados aos *outliers* nos dados analisados.

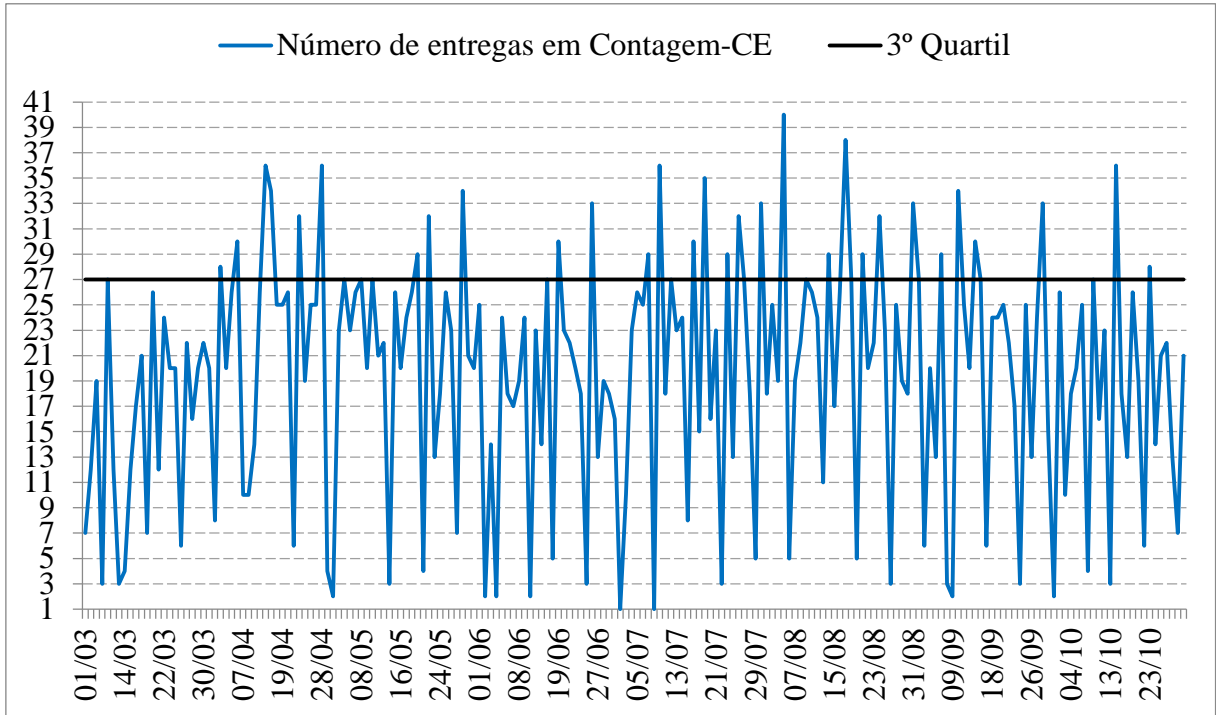
Após a análise por meio de estatística descritiva apresentada anteriormente, foi realizada então a análise de normalidade dos dados tanto por meio dos histogramas como por meio do teste Shapiro-Wilk. Os dados foram analisados considerando dados completos, 196 observações para cada indústria e após a primeira análise, retiraram-se desses dados o primeiro quartil visando reduzir a variabilidade da estrutura dos dados.

Para os dados de entrega aplicaram-se os testes Pearson's Chi-squared para verificar se os valores observados poderiam ser descritos por uma distribuição de probabilidades teórica de Poisson.

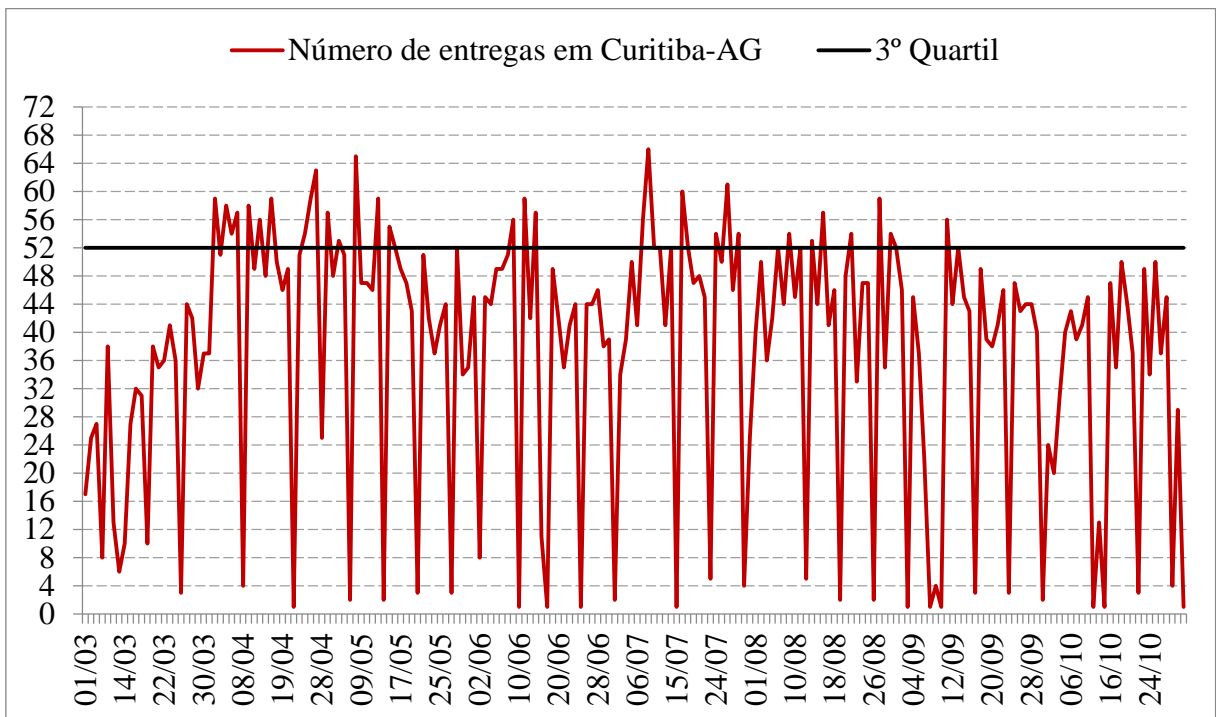
No APÊNDICE B são apresentados os histogramas dos dados com e sem o primeiro quartil e os testes Shapiro-Wilk que foram realizados. São também apresentados os testes de ajuste dos dados de entregas à função de distribuição de Poisson (Pearson's Chi-quadrado). Verificou-se que não há normalidade em nenhum dos casos para todas as variáveis confirmando que não seria possível utilizar o método dos mínimos quadrados para modelagem das entregas e da movimentação de cargas com destino às unidades industriais em análise. Assim, justificou-se a modelagem por meio de modelos lineares generalizados (GLM), considerando como técnica de estimação dos parâmetros, a máxima verossimilhança (MLE).

Também, após, essa análise, escolheu-se o valor referente ao terceiro quartil das variáveis dependentes para a modelagem. A justificava para a escolha se dá porque deseja-se representar situações críticas típicas sem atenuar a demanda tanto por entregas diárias como pela movimentação de cargas. Ao adotar o valor referente ao terceiro quartil pra cada unidade industrial, representa-se um valor crítico típico sem adoção de ocorrências extremas. Nas Figuras 4.5a.; 4.5b.; 4.5c. e 4.5d. são apresentadas as séries históricas com o valor referente ao terceiro quartil para cada unidade industrial considerada neste trabalho.

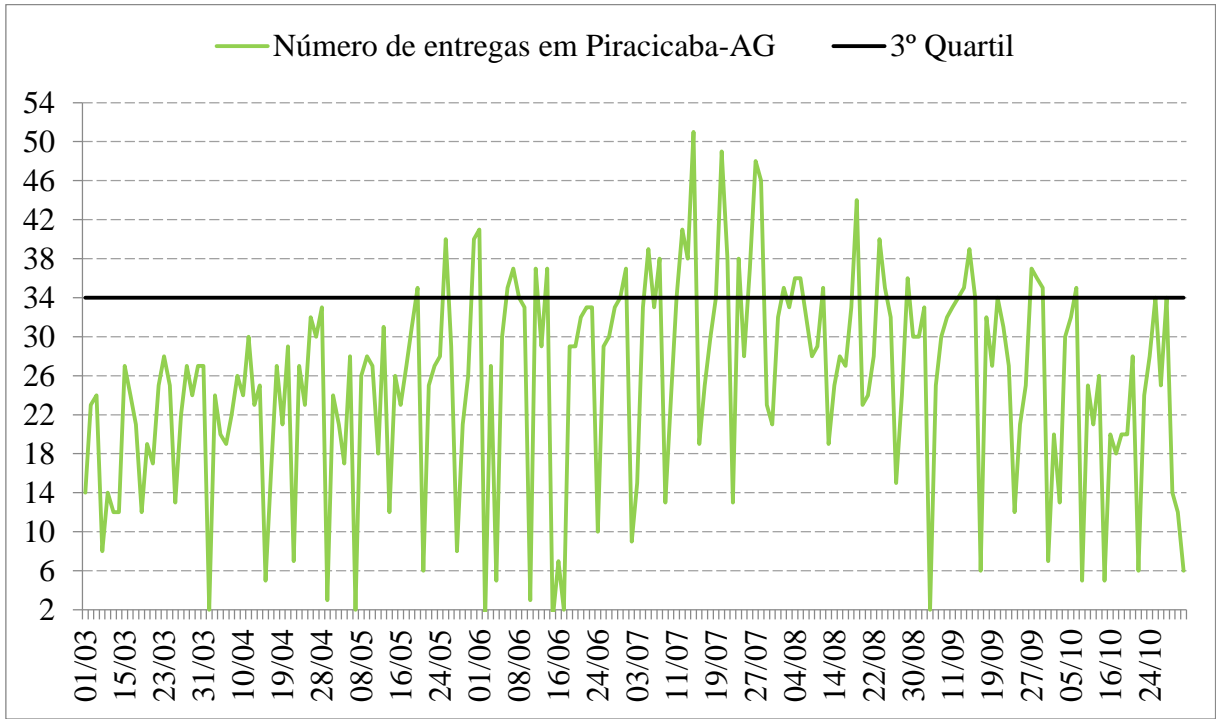
Figura 4.5: Série histórica do número de entregas.



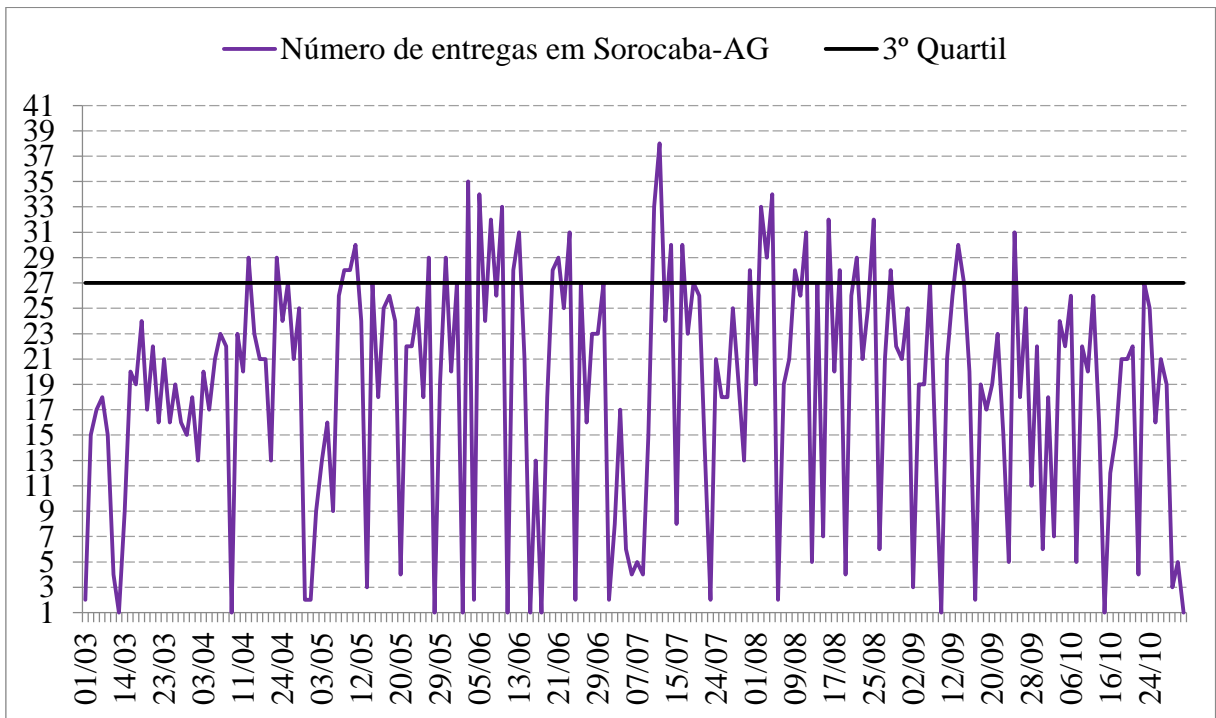
(a) Contagem-CE



(b) Curitiba-AG



(c) Piracicaba-AG



(d) Sorocaba-AG

Para desenvolvimento dos modelos, foram selecionadas sete variáveis, sendo cinco variáveis independentes: produção diária (prod); número de funcionários total (functot); número de funcionários da logística (funclog); área total das indústrias (areatot) e área de logística (arealog) e duas variáveis dependentes: número de entregas diárias (ent) e quantidade de carga (em quilograma) recebidas diariamente (car). Essas duas últimas variáveis tiveram valores selecionados conforme análises preliminarmente apresentadas (valores referentes ao 3º quartil de cada conjunto de dados das unidades industriais). Os dados utilizados para geração dos modelos de regressão generalizada são apresentados na Tabela 4.11.

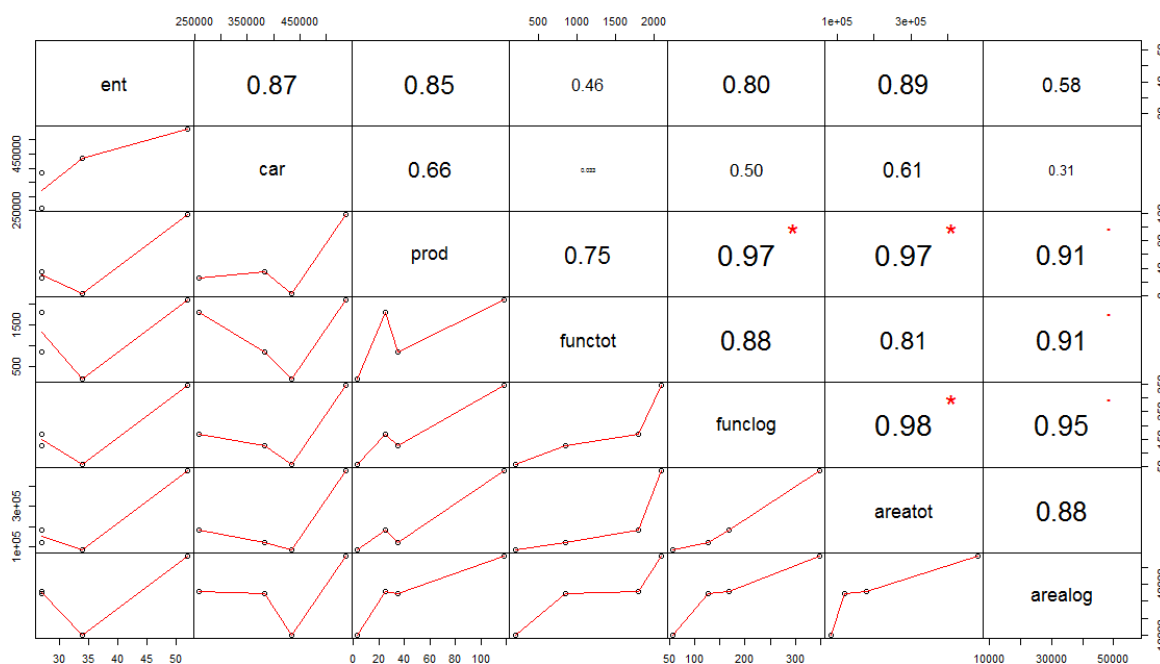
Tabela 4.11: Dados para geração dos modelos de regressão.

	<b>ent</b>	<b>car</b>	<b>prod</b>	<b>functot</b>	<b>funclog</b>	<b>areatot</b>	<b>arealog</b>
Contagem-CE	27	382417	35	857	128	120000	34414
Curitiba-AG	52	537946	118	2100	347	480000	57000
Piracicaba-AG	34	434765	3	200	58	83657	9902
Sorocaba-AG	27	258498	25	1800	168	180000	35640

\*\*\*As variáveis dependentes foram selecionadas considerando o terceiro quartil dos dados sem o primeiro quartil para entregas e carga.

Essas variáveis, após análise e consolidação, foram correlacionadas. Na Figura 4.6 são apresentados os coeficientes de correlação de Spearman entre as variáveis dependentes e independentes adotadas neste estudo.

Figura 4.6: Matriz de correlação de dados de Spearman.



Conforme Figura 4.6, as variáveis número de entregas diárias (variável dependente) e as variáveis explicativas PROD, FUNCLOG e AREATOT apresentam coeficiente de correlação acima de 0,8, indicado que é pertinente a avaliação de modelos de regressão com essas variáveis. Quanto à variável dependente quantidade de carga (CAR), o maior coeficiente de correlação é obtido com a variável explicativa PROD, com valor 0,66.

Na Figura 4.6 destacam-se, ainda, os valores de R acima de 0,50, em especial para correlação entre variáveis explicativas, o que contraindica a possibilidade de inclusão dessas variáveis em modelos de regressão múltipla. Esses critérios foram adotados neste estudo para determinação dos modelos de regressão considerados para análise.

### 4.3 Modelagem e avaliação

Utilizando o *software* R, foram gerados os modelos de regressão linear generalizada (GLM). Os modelos de regressão linear simples por meio do método dos mínimos quadrados não foram gerados, pois grande parte das análises não cumpriu as premissas necessárias para implementação desse método. Para os dados de entrega foram utilizadas funções da família

Poisson e para os dados de carga foram utilizadas funções Gaussianas. No APÊNDICE C são apresentadas, de maneira detalhada, os algoritmos desenvolvidos no *software* R.

Na Tabela 4.12 são apresentados os modelos de regressão linear generalizada referentes às entregas e quantidade de cargas atraídas pelas indústrias montadoras de máquinas agrícolas e rodoviárias em investigação. São ainda apresentados os parâmetros considerados para avaliação da significância dos estimadores, do ajuste, do equilíbrio entre complexidade e ajuste e da capacidade preditiva dos modelos.

Na análise dos modelos desenvolvidos para os dados de entrega (Poisson), dentre os nove modelos experimentados, foram identificados três modelos com parâmetros mais adequados na avaliação, a saber: GLM01, GLM03 e GLM04. Respectivamente, esses modelos consideram a produção, o número de funcionários alocados em atividades de logística e a área total construída em cada unidade industrial como variáveis explicativas. Uma vez que são utilizados modelos da mesma família nessa comparação, é possível analisar o *Akaike Information Criteria* corrigido (AICc). Dentre esses três modelos, o GLM04 apresentou o menor AICc (39,8), refletindo um maior parcimônia e aderência. Para a avaliação da acurácia dos modelos no processo de predição, ou seja, quando os modelos são expostos a dados não considerados na modelagem, analisa-se o valor da raiz quadrada do erro quadrático médio obtido a partir da validação cruzada apresentada na seção na qual se descreve a abordagem metodológica deste trabalho. Destaca-se que, devido ao tamanho reduzido da amostra, optou-se por realizar a validação cruzada por meio da técnica *leave-one-out*. Comparativamente, dentre os três modelos mais adequados, o modelo GLM03 apresentou a melhor acurácia preditiva (RMSE = 301,8). Apesar de apresentar melhor acurácia preditiva, o modelo GLM03 apresentou superdispersão elevada ( $c\text{-hat} = 1,99$ ), o que indica que a variância é maior que a esperada para esse modelo. Assim, comparando apenas os modelos GLM01 e GLM04, tem-se que o modelo GLM04 apresenta-se como aquele que tem melhor equilíbrio entre complexidade e ajuste, e melhor capacidade preditiva.

Considerando o ajuste (GOF), apenas os modelos GLM05 e GLM08 não são válidos em um nível de significância de 5%. Entretanto, os modelos GLM06 e GLM07 apresentam elevado RMSE e AICc, comparativamente. O modelo GLM02, assim como os GLM03, GLM05, GLM06, e GLM08 apresentam superdispersão elevada.

Algumas observações adicionais devem ser feitas, a saber: (i) os valores de  $\hat{c}$  não apontam superdispersão dos dados para o modelo GLM09 ( $\hat{c} = 0,36$ ), sendo  $\hat{c}$  menor que um, indicando independência entre as observações. Esse é o modelo que também apresenta menor raiz quadrada do erro quadrático médio, como resultado da independência dos dados observados. Entretanto, o equilíbrio entre complexidade e ajuste é menos adequado comparativamente aos demais modelos analisando o AICc; (ii) para o modelo GLM08 foi identificado um AICc negativo. Isso pode ocorrer quanto a verossimilhança é elevada e não invalidaria o modelo, caso houvesse significância estatística.



Tabela 4.12: Modelos lineares generalizados de atração de entregas e carga.

Modelo	Método	Variável dependente	Variável independente	Parâmetros de estimação	Valor	Z-valor	Residual Deviance goodness of fit test - GOF	c-hat (Pearson estimator)	AICc	RMSE
GLM01	Regressão Linear Generalizada (Poisson)	entrega	prod	Intercepto	3,290426	< 2e-16 ***	2,9795	1,52	40,42071	332,1800
				Coeficiente	0,005241	0,00342 **	p-value = 0,23			
GLM02	Regressão Linear Generalizada (Poisson)	entrega	functot	Intercepto	3,3206104	<2e-16 ***	8,6029	4,24	46,04418	304,8629
				Coeficiente	0,0001819	0,112	p-value = 0,14			
GLM03	Regressão Linear Generalizada (Poisson)	entrega	funclog	Intercepto	3,1720808	< 2e-16 ***	3,9384	1,99	41,37964	301,8190
				Coeficiente	0,0020450	0,00629 **	p-value = 0,14			
GLM04	Regressão Linear Generalizada (Poisson)	entrega	areatot	Intercepto	3,202e+00	< 2e-16 ***	2,3996	1,22	39,84082	325,2183
				Coeficiente	1,501e-06	0,00239 **	p-value = 0,30			
GLM05	Regressão Linear Generalizada (Poisson)	entrega	arealog	Intercepto	3,184e+00	< 2e-16 ***	7,0489	3,5	44,49012	332,9982
				Coeficiente	1,042e-05	0,0439 *	p-value = 0,03			
GLM06	Regressão Linear Generalizada (Poisson)	entrega	prod + functot	Intercepto	3,4066934	<2e-16 ***	2,0624	1,99	Inf	445,2710
				Coeficiente	0,0075476	0,0138 *	p-value = 0,15			
				Coeficiente	-0,0001808	0,3454				
GLM07	Regressão Linear Generalizada (Poisson)	entrega	functot + funclog	Intercepto	3,2508892	< 2e-16 ***	1,2313	1,2	Inf	420,8957
				Coeficiente	-0,0004212	0,11005	p-value = 0,27			
				Coeficiente	0,0045229	0,00888 **				

Tabela 4.13: Modelos lineares generalizados de atração de entregas e carga – continuação.

Modelo	Método	Variável dependente	Variável independente	Parâmetros de estimação	Valor	Z-valor	Residual Deviance goodness of fit test - GOF	c-hat (Pearson estimator)	AICc	RMSE
GLM08	Regressão Linear Generalizada (Poisson)	entrega	functot + funclog+ areatot	Intercepto	3,397e+00	<2e-16 ***	5,9952e-15 p-value = 0	Inf	-10,55876	411,4197
				Coeficiente	-1,256e-04	0,738				
				Coeficiente	-4,801e-03	0,582				
				Coeficiente	5,175e-06	0,275				
GLM09	Regressão Linear Generalizada (Poisson)	entrega	areatot + arealog	Intercepto	3,427e+00	<2e-16 ***	0,36644 p-value = 0,54	0,36	Inf	275,5471
				Coeficiente	3,017e-06	0,011 *				
				Coeficiente	-1,638e-05	0,152				
GLM10	Regressão Linear Generalizada (Gaussiana)	carga	prod	Intercepto	334201	0,0491 *	2,2763e+10 p-value = 0	-	Inf	50710675122
				Coeficiente	1529	0,3380				
GLM11	Regressão Linear Generalizada (Gaussiana)	carga	functot	Intercepto	3,980e+05	0,100	4,0480e+10 p-value = 0	-	Inf	36959437940
				Coeficiente	4,375e+00	0,967				
GLM12	Regressão Linear Generalizada (Gaussiana)	carga	funclog	Intercepto	320757,0	0,114	3,0397e+10 p-value = 0	-	Inf	69711581017
				Coeficiente	471,6	0,500				
GLM13	Regressão Linear Generalizada (Gaussiana)	carga	areatot	Intercepto	3,181e+05	0,080	2,5276e+10 p-value = 0	-	Inf	143620414011
				Coeficiente	3,950e-01	0,387				

Tabela 4.14: Modelos lineares generalizados de atração de entregas e carga – continuação.

Modelo	Método	Variável dependente	Variável independente	Parâmetros de estimação	Valor	Z-valor	Residual Deviance goodness of fit test - GOF	c-hat (Pearson estimator)	AICc	RMSE	
GLM14	Regressão Linear Generalizada (Gaussiana)	carga	arealog	Intercepto	3,390e+05	0,159	3,6587e+10	-	Inf	55957816791	
				Coeficiente	1,881e+00	0,688	p-value = 0				
GLM15	Regressão Linear Generalizada (Gaussiana)	carga	prod + functot	Intercepto	427126,08	0,0719 .	2,4839e+09	-	60,33877	77574523327	
				Coeficiente	3398,88	0,1594	p-value = 0				
				Coeficiente	-143,25	0,2143					
GLM16	Regressão Linear Generalizada (Gaussiana)	carga	functot+ funclog	Intercepto	356248,18	0,0421 *	5,7546e+08	-	54,48905	10233493723	
				Coeficiente	-240,87	0,0879 .	p-value = 0				
				Coeficiente	1972.360,0	0,0761 .					
GLM17	Regressão Linear Generalizada (Gaussiana)	carga	functot+ funclog+ areatot	Intercepto	3,739e+05	NA	4,2352e-21	-	-201,8371	10089628769	
				Coeficiente	-2,078e+02	NA					p-value = 0
				Coeficiente	8,859e+02	NA					
				Coeficiente	6,108e-01	NA					
GLM18	Regressão Linear Generalizada (Gaussiana)	carga	areatot + arealog	Intercepto	4,025e+05	0,234	1,6073e+10	-	67,80789	521955721519	
				Coeficiente	9,602e-01	0,461	p-value = 0				
				Coeficiente	-6,028e+00	0,588					

\*Significado dos códigos: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' >

Quanto aos modelos gerados considerando a quantidade de carga recebida como variável dependente, gerados por meio de regressão linear generalizada da família Gaussiana, nenhum dos modelos apresentou ajuste satisfatório. Foram realizados tanto testes de *goodness of fit* chi-quadrado como foi analisado o complemento da razão entre o *residual deviance* e *null deviance*, que é análogo ao coeficiente de determinação ( $R^2$ ).

#### **4.4 Análise comparativa conceitual**

Conforme discutido no referencial teórico deste trabalho, Marra (1999) e Kulpa (2014) afirmam que a escolha do tipo de modelo depende tanto dos objetivos do analista como da disponibilidade e consistência dos dados. Neste trabalho, confirmam-se essas dependências. Como objetivo deste trabalho, tinha-se a estruturação de modelos de atração para indústria de máquinas agrícolas e rodoviárias. Modelos de atração fazem parte de um conjunto maior de instrumentos utilizados para estimativa de produção e atração de fluxos para localidades de interesse. Esses instrumentos são denominados, de maneira mais generalizada, modelos de geração de viagens e são utilizados de maneira tradicional em etapas táticas e estratégicas do planejamento de transportes (OLIVEIRA *et al.*, 2016; CAMPOS, 2013; GASPARINI, 2008).

Como instrumento para avaliação do impacto e subsídio para promoção de políticas públicas visando a instalação de novos empreendimentos industriais com características semelhantes aos investigados, os modelos desenvolvidos neste trabalho podem ser considerados, uma vez que são instrumentos de planejamento tático/estratégico para áreas urbanas.

Holguín-Veras e Thorson (2000) e Ogden (1992) destacam que os modelos baseados em viagens de veículos de carga apresentam vantagens sobre os aqueles baseados em quantidade de carga movimentada, como, por exemplo, a relativa facilidade em se obter dados de tráfego para a realização de planejamento de transportes e da disponibilidade cada vez maior de Sistemas Inteligentes de Transporte (ITS), o que torna a coleta de dados independente do ator privado. Comi *et al.* (2012) também afirmam que a coleta de dados de fluxos de mercadorias é menos disponível e mais onerosa, uma vez que as empresas não tornam públicos os dados em função de estratégias de mercado. Nesse contexto, o uso de dados de entrega torna-se mais acessível e consistente do que a informação sobre a quantidade de carga movimentada, que se refere às características operacionais da cadeia de suprimentos dos atores privados. Assim,

tanto o objetivo (planejamento tático/estratégico) como o acesso e a confiabilidade da informação são corroborados por meio dos resultados obtidos neste trabalho, quando modelos de entregas são validados estatisticamente e atendem aos objetivos inicialmente contemplados.

Em relação à abordagem metodológica, identificou-se a impossibilidade de utilizar o método dos mínimos quadrados para estimação dos parâmetros para modelos de regressão linear geral devido às características das variáveis e suas relações que não atendiam às premissas necessárias para uso desse método. Assim, optou-se por estimar os parâmetros da regressão por meio da estimativa de máxima verossimilhança, gerando modelos de regressão generalizada. Outra condicionante importante para a modelagem foi a natureza das variáveis dependentes, a saber: (i) a frequência das entregas nas unidades industriais caracterizaram-se como dados de contagem; e (ii) a quantidade de carga movimentada por entrega que se caracterizou como variável contínua. Assim, a variável “entregas” foi modelada por meio de modelos Poisson e a variável “carga” foi modelada por meio de modelos Gaussianos.

Quanto às variáveis explicativas consideradas para a modelagem, na revisão bibliométrica e sistemática identificou-se que 17% e 16% dos estudos identificou como mais adequadas as variáveis funcionários e área do estabelecimento (Figura 2.3). Neste trabalho, corrobora-se a literatura visto que os modelos que apresentaram melhor ajuste, equilíbrio e acurácia preditiva consideraram a frequência de entregas como variável dependente e apenas uma variável em cada, a saber: (i) área total construída; (ii) produção; e (iii) funcionários em atividades logísticas. A importância da intensidade de funcionários alocados em atividades logísticas para explicar a frequência de entregas é também apresentada por Rolko e Friedrich (2017), cujo estudo é discutido na revisão bibliográfica.

Não foi possível utilizar indicadores objetivos para comparar os modelos. Segundo Emiliano (2009), o AIC, ou AICc, não permite testes de hipóteses. Trata-se de uma ferramenta para seleção de modelos que sejam gerados a partir de um mesmo conjunto de dados e, portanto, não pode ser utilizado para comparar modelos que considerem variáveis dependentes diferentes. Assim foi possível apenas comparar, entre si, modelos gerados para entregas. Os modelos para cargas não apresentaram ajuste satisfatório e, portanto, não foram comparados.

Ainda buscando uma comparação objetiva entre os modelos, foi realizada a validação cruzada e o cálculo da raiz quadrada do erro quadrático médio, visando identificar a robustez dos modelos em predição quando expostos a um novo conjunto de dados. Novamente, apenas os modelos de entregas foram comparados quanto à acurácia preditiva entre si.

De maneira sintética, pode-se concluir, para o segmento de indústrias de máquinas agrícolas e rodoviárias, a partir deste estudo, que: (i) a obtenção dos dados de entregas é mais objetiva e menos dependente de atores privados; (ii) os modelos que consideram a frequência das entregas como variável dependente são adequados quando o objetivo é sua aplicação em previsão de impactos em níveis tático e estratégicos do planejamento das cidades, o que valida a hipótese “iv” deste trabalho; (iii) esses modelos são adequados para predição; (iv) área total construída, produção e funcionários em atividades logísticas são as variáveis independentes com melhor ajuste, equilíbrio e acurácia preditiva considerando a frequência de entregas como variável dependente validando assim a hipótese “iii”; (v) a técnica de regressão linear generalizada é adequada para modelar a atração de entregas por veículos de carga, mas não se mostrou satisfatórias para dados de carga movimentada nos fluxos de abastecimento de indústrias de máquinas agrícolas e rodoviárias, validando a hipótese “i” no que se refere a dados de entrega; (vi) o comportamento da correlação entre variáveis explicativas e a variável dependente (entregas ou quantidade de insumos) está fortemente relacionado com o contexto e o local em que as regressões são geradas validando a hipótese “ii”.

## 5 CONSIDERAÇÕES FINAIS

O objetivo principal deste trabalho foi desenvolver e avaliar modelos de atração de carga e viagens para a indústria fabricante de máquinas agrícolas e rodoviárias no contexto brasileiro. Para alcançá-lo foram consideradas informações referentes a fluxos de abastecimento de quatro indústrias localizadas nos municípios de Contagem (Minas Gerais), Curitiba (Paraná), Piracicaba e Sorocaba (São Paulo).

Com base na revisão bibliográfica, apresentada no capítulo dois deste trabalho, foram definidas as informações relevantes para a modelagem da demanda de viagens de carga a partir de artigos já desenvolvidos relacionados a modelos de geração de viagens de carga. Como apresentado no capítulo três, a estrutura metodológica foi dividida em quatro etapas, partindo da caracterização econômica e territorial; organização das informações; modelagem e avaliação dos modelos desenvolvidos por meio do método de regressão linear generalizado e a análise comparativa conceitual dos modelos desenvolvidos.

A pesquisa considerou o fluxo de suprimentos das quatro indústrias que é constituído de 592 fornecedores que estão distribuídos em diversas cidades brasileiras nos Estados de Minas Gerais (MG), Paraná (PR), Rio de Janeiro (RJ), Rio Grande do Sul (RS), Santa Catarina (SC) e São Paulo (SP).

Após obtenção dos dados, foi desenvolvido um banco de dados com as informações coletadas e realizados em seguida testes e análises estatísticas que permitiram desenvolver os modelos da demanda de veículos de carga atraída pelas indústrias foco do estudo. Foram desenvolvidos 18 modelos preditivos de geração de viagem por meio da regressão linear generalizada, sendo nove de atração com dados de entrega e o restante de carga. O modelo que considera como variável independente a frequência das entregas e como variável independente a área total construída das indústrias foi validado estatisticamente, considerando tanto o equilíbrio entre ajuste e complexidade (AICc) e a acurácia preditiva (raiz quadrada do erro médio quadrático na validação cruzada).

Os modelos simples que consideraram como variáveis independentes a produção, o número de funcionários totais e o número de funcionários alocados em atividades de logística para

explicar as entregas (GLM01, GLM02 e GLM03 respectivamente) apresentam bom ajuste e acurácia preditiva. Entretanto, apresentaram superdispersão, ou seja, a variância é maior que a esperada. Nesse caso, sugere-se que, em investigações futuras sejam desenvolvidas análises considerando modelos da família quasi-poisson.

Os modelos desenvolvidos com a variável dependente quantidade de carga não foram validados estatisticamente. Esse resultado estimula o uso de dados da frequência de entregas para desenvolvimentos de modelos com foco em etapas de avaliação tática/estratégica de planejamento, conforme discutido na apresentação do estado da arte e dos resultados deste trabalho. Entretanto, para outras finalidades, como estudos de tráfego no entorno de empreendimentos como as indústrias em investigação, pode ser necessário o detalhamento operacional das atividades de suprimento na cadeia logística, justificando a coleta de dados de movimentação de mercadorias e especificidades dos veículos movimentados.

Os resultados apresentados permitem concluir que:

- ✓ a produção diária e a área total construída são as variáveis independentes com melhor poder de predição de entregas diárias para a atividade econômica em estudo;
- ✓ a obtenção dos dados de entregas é mais objetiva e menos dependente de atores privados;
- ✓ os modelos que consideram a frequência das entregas como variável dependente são adequados quando o objetivo é sua aplicação em previsão de impactos em níveis tático e estratégicos do planejamento das cidades;
- ✓ os modelos que consideram a frequência das entregas são adequados para predição;
- ✓ a técnica de regressão linear generalizada é adequada para modelar a atração de entregas por veículos de carga, mas não se mostrou satisfatórias para dados de carga movimentada nos fluxos de abastecimento de indústrias de máquinas agrícolas e rodoviárias;
- ✓ o comportamento da correlação entre variáveis explicativas e dependente (entregas ou quantidade de insumos) está fortemente relacionado com o contexto e o local em que as regressões são geradas.



Vale destacar que, para a atividade industrial, foco deste trabalho, o número de vendas totais por ano e o tipo de mercadoria manufaturada são fatores relevantes para a estimativa de modelos de geração de viagens de veículos de carga. Portanto, recomenda-se que os modelos apresentados e validados neste trabalho sejam considerados para atividades semelhantes em escala e escopo. Podem haver particularidades locais que devem ser consideradas necessitando a adaptação e a geração de modelos locais, validando a contribuição deste trabalho como instrumento de subsídio à avaliação de impacto de novos empreendimentos da indústria de máquinas agrícolas e rodoviárias.

Como resultado deste estudo, considerando que a geração de viagens de veículos de carga em indústrias ocorre diariamente, a implementação de ações que mitiguem os efeitos negativos sobre a área de influência desses empreendimentos e potencializem a operação, podendo colaborar para o aumento da qualidade de vida da população, para o desenvolvimento sustentável e para a melhoria da distribuição de mercadorias. Os modelos de geração de viagens de veículos de carga podem auxiliar na análise dos impactos da operação de entrega de insumos nas indústrias, colaborando com a redução das influências negativas causadas pelos veículos de carga no sistema viário e, ainda, permitindo melhorias no planejamento urbano.

Considerando as atividades desenvolvidas neste trabalho e potenciais melhorias, recomenda-se para futuros estudos:

- ✓ a modelagem da demanda de carga por tipos de veículos de carga e por faixas horárias (pico e fora pico);
- ✓ considerar outras famílias de modelos lineares generalizados diferentes da Gaussiana;
- ✓ verificar a adequação de modelos de regressão sem intercepto para o processo em análise;
- ✓ analisar a correlação entre a localização e a distribuição das indústrias com a densidade demográfica e/ou renda *per capita* das regionais administrativas com base em dados do Instituto Brasileiro de Geografia e Estatística (IBGE).

## REFERÊNCIAS

AKAIKE, H. *A new look at the statistical model identification*. IEEE Transactions on Automatic Control., Boston, v.19, n.6, p.716-723, Dec. 1974.

AL-DEEK, H. M. Comparison of two approaches for modeling freight movement at seaports. *Journal of Computing in Civil Engineering*. v. 15, n. 4, p. 284–291, 2001. DOI: [https://doi.org/10.1061/\(ASCE\)0887-3801\(2001\)15:4\(284\)](https://doi.org/10.1061/(ASCE)0887-3801(2001)15:4(284)).

ALHO, A. R; SILVA, J. A. Freight-trip generation model: predicting urban freight weekly parking demand from retail establishment characteristics. *Transportation Research Record: Journal of the Transportation Research Board*. v. 2411, p. 45–54, 2014. DOI: <https://doi.org/10.3141/2411-06>.

ALHO, A. R; SILVA, J. A. Modeling retail establishments' freight trip generation: a comparison of methodologies to predict total weekly deliveries. *Transportation*, [s.l.], v. 44, n. 5, p.1195-1212, 12 maio 2016. Springer Nature. DOI: <http://dx.doi.org/10.1007/s11116-016-9700-z>.

ALLEN, J.; ANDERSON, S.; BROWNE, M.; JONES, P. *A framework for considering policies to encourage sustainable urban freight traffic and goods/service flows: summary report*. University of Westminster. Londres, Inglaterra. 2000.

ALLEN, W. G. *Development of truck models*. Transportation Consultant. Mitchells, Estados Unidos. 2002.

ALPAYDIN, E. *Introduction to Machine Learning*. MIT Press, 2nd edition. 2010.

ANFAVEA - Associação Nacional dos Fabricantes de Veículos Automotores. *Imprensa notícias 2018*. Disponível em: < [http://www.anfavea.com.br/docs/26.04.18\\_Agrishow.pdf](http://www.anfavea.com.br/docs/26.04.18_Agrishow.pdf)> Acesso em maio de 2018.

ANFAVEA - Associação Nacional dos Fabricantes de Veículos Automotores. *Anuário da indústria automobilística brasileira*. Brazilian Automotive Industry Yearbook. 2018. Disponível em: < <http://www.anfavea.com.br/anuarios.html>> Acesso em maio de 2018.

ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. *Statistics Surveys*, [s.l.], v. 4, p.40-79. Institute of Mathematical Statistics. 2010. DOI: <http://dx.doi.org/10.1214/09-ss054>.

BARONE, V.; CROCCO, F.; MONGELLI, D. W. E. Freight transport demand models for applications in urban areas. *Applied Mechanics and Materials*. v. 442, p. 634–644, 2014. DOI: 10.4028/www.scientific.net/AMM.442.634.

BASTIDA, C.; HOLGUÍN-VERAS, J. Freight generation models: comparative analysis of regression models and multiple classification analysis. *Transportation Research Record: Journal of the Transportation Research Board*. v. 2097, p. 51–61, 2009. DOI: <https://doi.org/10.3141/2097-07>.

BINSBERGEN, A. V.; VISSER, J. Innovation steps towards efficient goods distribution systems for urban areas - efficiency improvement of goods distribution in urban areas. *The Netherlands TRAIL Research School*. Delft, Netherland. 2001.

BLACK, R. W. Commodity flow modeling. *Transportation Research Board / National Research Council*. n. E-C011, p. 136-155. 1999. RESEARCH ISSN 0097-8515.

BOARNET, M. G.; HONG, A.; SANTIAGO-BARTOLOMEI, R. Urban spatial structure, employment subcenters, and freight travel. *Journal Of Transport Geography*, [s.l.], v. 60, p.267-276, 2017. Elsevier BV. DOI: <http://dx.doi.org/10.1016/j.jtrangeo.2017.03.007>.

BRASIL. IBGE. *Censo Demográfico*, 2010. Disponível em: <[www.ibge.gov.br](http://www.ibge.gov.br)>. Acesso em maio 2018.

BURNHAM, K. P.; ANDERSON, D. R. *Model Selection and Multimodel Inference: a practical information-theoretic approach*. 2. ed. Fort Collins: Springer, 515 p. 2002.

BURNHAM, K. P.; ANDERSON, D. R.; HUYVAERT, K. P. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav Ecol Sociobiol*. v. 65, p. 23–35. 2011. DOI: 10.1007/s00265-010-1029-6.

CAMBRIDGE SYSTEMATICS. *Quick response freight manual*. Federal Highway Administration, U.S. Department of Transportation. 1996.

- CAMPBELL, S.; JALLER, M.; SÁNCHEZ-DÍAZ, I.; HOLGUÍN-VERAS, J.; LAWSON, C. T. Comparison between industrial classification systems in freight trip generation (FTG) modeling. In 91st TRB Annual Meeting, Washington, D.C. United States, 2012.
- CAMPOS, V. B. G. *Planejamento de transportes: conceitos e modelos*. Ed. Interciência, Rio de Janeiro. 2013. 188p.
- CAMPOS, V. B. G.; D'AGOSTO, M. A.; CRUZ, I.; SOUZA, C. D. R. de. Polos geradores de viagens de carga. In: PORTUGAL, L. S. (Org.). *Polos Geradores de Viagens Orientados a Qualidade de Vida e Ambiental: modelos e taxas de geração de viagens*. Rio de Janeiro: Interciência, 2012. Cap 16, p. 531-564.
- CAMPOS, V. B. G.; MELO, I. C. Análisis de la demanda de transporte de carga en área urbana bajo el punto de vista de la producción y de la atracción de viajes. *Anais do VI Congresso de Ingeniería del Transporte*. Zaragoza, Espanha. 2004.
- CANTILLO, V.; HOLGUÍN-VERAS, J.; JALLER, M. The Colombian strategic freight transport model based on product analysis. *PROMET - Traffic&Transportation*. v. 26, n. 6, p. 487–496, 2014. DOI: <https://doi.org/10.7307/ptt.v26i6.1460>.
- CARLOS, A. F. A. Espaço e indústria. 7ª. Ed. São Paulo: Contexto. 1997. 70p.
- CHU, H.C. Empirical method for predicting internal-external truck trips at a major port. *Journal of Transportation Engineering*. v. 137, n. 7, p. 496–508, 2011. DOI: [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000233](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000233).
- COMI, A.; DELLE SITE, P.; FILIPPI, F.; NUZZOLO, A. Urban freight transport demand modelling: a state of the art. In: *European transport / trasporti europei*, ISTIEE, 51, 1–17, 2012, Trieste, Itália.
- COMI, A.; NUZZOLO, A. Simulating urban freight flows with combined shopping and restocking demand models. *Procedia-Social and Behavioral Sciences*, 125, 49-61, 2014. DOI: <https://doi.org/10.1016/j.sbspro.2014.01.1455>.

CORDEIRO, G. M.; DEMÉTRIO, C. G. B. *Modelos lineares generalizados e extensões*. 2013. 493 f. Departamento de Estatística e Informática, UFRPE. Departamento de Ciências Exatas, ESALQ, USP.

CORREIA, V. *Análise econômica e ambiental da implantação de um esquema de centro de distribuição urbano para Belo Horizonte*. 2011. 137 f. Dissertação (Mestrado em Geotecnia e Transportes), Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

CROCCO, F.; DE MARCO, S.; IAQUINTA, P.; MONGELLI, D. W. E. Freight transport in urban areas: an integrated system of models to simulate freight demand and passengers demand for purchase trips. *International Journal of Mathematical Models and Methods in Applied Sciences*. v. 4, n. 4, p. 265-273, 2010.

CUI, J; DODSON, J.; HALL, P. V. Planning for Urban Freight Transport: An Overview. *Transportation Research*. v. 35, issue 5, p. 583-598, 2015. DOI: <https://doi.org/10.1080/01441647.2015.1038666>.

DABLANC, L. Goods transport in large European cities: Difficult to organize, difficult to modernize. *Transportation Research Part A*. v. 41, issue 3, p. 280-285, 2007. DOI: <https://doi.org/10.1016/j.tra.2006.05.005>.

DAVISON, A. C. Biometrika Centenary: Theory and General Methodology. *Biometrika*, 88, pp. 13-52. 2001.

DUTRA, N. G. S. *O Enfoque de “City Logistics” na distribuição urbana de encomendas*. 2004. 229 f. Tese (Doutorado) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2004.

EMILIANO, P. C. *Fundamentos e aplicações dos critérios de informação: Akaike Bayesiano*. 2009. 105f. Dissertação (Mestrado em Estatística e Experimentação Agropecuária), Universidade Federal de Lavras, Lavras, 2009.

EVERITT, B.; SKRONDAL, A. *The Cambridge dictionary of statistics*. 4<sup>a</sup>. ed. [S.l.]: Cambridge University Press Cambridge, 2010. 480 p. ISBN 978-0-521-76699-9.

FACCHINI, D. *Análise dos “GAPS” de percepção dos atores envolvidos no transporte urbano de carga em Porto Alegre*. 2006. 135f. Dissertação (Mestrado), Universidade Federal do Rio Grande do Sul, Porto Alegre. 2006.

FARRINGTON, C. P. On assessing goodness of fit of generalized linear models to sparse data. *Journal of the Royal Statistical Society. Series B: methodological* 58, p. 349-360. 1996. DOI: 10.2307/2345981.

FERREIRA, B. L. G; SILVA, M. A. V. Análise da relação entre variáveis socioeconômicas e o número de viagens de veículos de carga na região metropolitana do Rio de Janeiro. *Revista Produção e Desenvolvimento*. v. 2, n. 1, p. 77-89. 2016.

FERREIRA, B. L. G; SILVA, M.A.V. Análise de geração de viagens de carga em área urbana. In: *ANPET – XXIX CONGRESSO NACIONAL DE PESQUISA EM TRANSPORTES*, 2015, Outro Preto, MG.

FERREIRA, B. L. G.; SILVA, M. A. V. Truck trips in urban areas and its relation to socioeconomic variables. *Revista Gestão da Produção Operações e Sistemas*, [s.l.], v. 11, n. 4, p.197-212, 1 nov. 2016. A Fundacao para o Desenvolvimento de Bauru (FunDeB). DOI: <http://dx.doi.org/10.15675/gepros.v11i4.1566>.

FLETCHER, D. J. Estimating overdispersion when fitting a generalized linear model to sparse data. *Biometrika*. v. 99, p. 230–237, 2012. DOI: <https://doi.org/10.1093/biomet/asr083>.

GASPARINI, A. *Atratividade do transporte de carga para polos geradores de viagem em áreas urbanas*. 2008. 256f. Dissertação (Mestrado) Engenharia de Transportes, Instituto Militar de Engenharia. Rio de Janeiro. 2008.

GASPARINI, A.; CAMPOS, V. B. G.; D’AGOSTO, M. A. Modelos para estimativa da demanda de viagens de veículos de carga para supermercados e shopping-centers. *Transportes*, 18(1), 2010. DOI: <https://doi.org/10.14295/transportes.v18i1.383>.

GASPARINI, A; CAMPOS, V. B. G.; D’AGOSTO, M. A. Atratividade do transporte de carga para polos geradores de viagens em áreas urbanas. In: *XXI ANPET – CONGRESSO DE PESQUISA E ENSINO EM TRANSPORTES*, 2007, Rio de Janeiro. Panorama Nacional da Pesquisa em Transportes. Rio de Janeiro: ANPET, 2007, v. 1.

- GIULIANO, G.; DABLANC, L. Approaches to managing freight in metropolitan areas. *Urban Freight Symposium*, may 2013, France. 23p., 2013.
- GONÇALVES, M. R. *Desenvolvimento e teste de um novo modelo gravitacional - de oportunidades para distribuição de viagens*. 1992. Tese (Doutorado) Engenharia de Produção, Universidade Federal de Santa Catarina. Florianópolis, SC. 1992.
- GONÇALVES, F. S.; LEMOS, D. S. C. P. S.; KNEIB, E. C.; PORTUGAL, L. S. Caracterização dos polos geradores de viagens. In: PORTUGAL, L. S. (Org.). *Polos Geradores de Viagens Orientados a Qualidade de Vida e Ambiental: modelos e taxas de geração de viagens*. Rio de Janeiro: Interciência, 2012. Cap. 1, p. 5-43.
- GONZALEZ-FELIU, J.; PERIS-PLA, C. Impacts of retailing attractiveness on freight and shopping trip attraction rates. *Research In Transportation Business & Management*, [s.l.], v. 24, p.49-58, set. 2017. Elsevier BV. DOI: <http://dx.doi.org/10.1016/j.rtbm.2017.07.004>.
- GONZALEZ-FELIU, J.; ROUTHIER, J. Modeling urban goods movement: How to be oriented with so many approaches?. *Procedia-Social and Behavioral Sciences*, v. 39, p. 89-100, 2012. DOI: <https://doi.org/10.1016/j.sbspro.2012.03.093>.
- GONZALEZ-FELIU, J.; SÁNCHEZ-DÍAZ, I. The influence of aggregation level and category construction on estimation quality for freight trip generation models. *Transportation Research Part E: Logistics and Transportation Review*, July 2018. DOI: <https://doi.org/10.1016/j.tre.2018.07.007>.
- GONZALEZ-FELIU, J.; SEMET, F.; ROUTHIER, J. L. *Sustainable urban logistics: concepts, methods and information systems*. Ecoproduction. 2014.
- GRIECO, E. P.; PORTUGAL, L. S. Taxas de geração de viagens em condomínios residenciais: estudo de caso. *Transportes*. v. 18, n. 1, p. 87-95, 2010. DOI: <http://dx.doi.org/10.14295/transportes.v18i1.386>.
- GÜNAY, G.; ERGÜN, G.; GÖKASAR, I. Conditional Freight Trip Generation modelling. *Journal Of Transport Geography*, [s.l.], v. 54, p.102-111, 2016. Elsevier BV. DOI: <http://dx.doi.org/10.1016/j.jtrangeo.2016.05.013>.

HAIR JR., J. F.; BLACK, W.C.; BABIN, B. J.; ANDERSON, R. E. *Multivariate data analysis*. 7th ed. Pearson Prentice Hall, 2014.

HAUKE, J.; KOSSOWSKI, T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae*, Bogucki Wydawnictwo Naukowe, v. 30, n. 2, p. 87-93, 2011. ISSN 0137-477X.

HINDE, J.; DEMETRIO, C. G. B. Overdispersion: models and estimation. *Computational Statistics & Data Analysis*, v. 27, p. 151-170, 1998. Elsevier. DOI: [https://doi.org/10.1016/S0167-9473\(98\)00007-3](https://doi.org/10.1016/S0167-9473(98)00007-3).

HOLGUÍN-VERAS, J.; JALLER, M.; DESTRO, L.; BAN, X.; LAWSON, C.; LEVINSON, H. Freight generation, freight trip generation, and perils of using constant trip rates. *Transportation Research Record: Journal of the Transportation Research Board*. v. 2224, p. 68–81, 2011. DOI: <https://doi.org/10.3141/2224-09>.

HOLGUÍN-VERAS, J.; JALLER, M.; SÁNCHEZ-DÍAZ, I.; WOJTOWICZ, J.; CAMPBELL, S.; LEVINSON, H.; LAWSON, C.; POWERS, E. L.; TAVASSZY, L. *NCHRP Report 739/NCFRP Report 19: Freight Trip Generation and Land Use*. Transportation Research Board. Washington, D.C., United States. 2012.

HOLGUÍN-VERAS, J.; SÁNCHEZ-DÍAZ, I.; LAWSON, C.; JALLER, M.; CAMPBELL, S.; LEVINSON, H.; SHIN, H. S. Transferability of freight trip generation models. *Transportation Research Record: Journal of the Transportation Research Board*. v. 2379, p. 1–8, 2013. DOI: <https://doi.org/10.3141/2379-01>.

HOLGUÍN-VERAS, J.; THORSON, E. Trip length distributions in commodity based and trip based freight demand modeling: investigations of relationships. *Transportation Research Record: Journal of the Transportation Research Board*. v. 1707. Washington, 2000. DOI: <https://doi.org/10.3141/1707-05>.

HURVICH, C. M. e TSAI, C. L. Selection for Extended Quasi-Likelihood in Small Samples. *Biometrics*, 51, pp. 1077-1084. 1995.



HUTCHINSON, B. G. Principles of urban transportation systems planning. *Transportation Research Board - TRB*. Scripta Book Company, Washington, 1974. ACCESSION NUMBER 00072659.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. *International Journal Of Forecasting*, v. 22, n. 4, p.679-688, 2006. DOI: <http://dx.doi.org/10.1016/j.ijforecast.2006.03.001>.

IBEAS, A.; MOURA, J. L.; NUZZOLO, A.; COMI, A. Urban freight transport demand: transferability of survey results analysis and models. *Procedia - Social and Behavioral Sciences*. v. 54, p. 1068–1079, 2012. DOI: <https://doi.org/10.1016/j.sbspro.2012.09.822>.

IDING, M. H. E.; MEESTER, W. J.; TAVASSZY, L. Freight trip generation by firms. *Paper for the 42nd European Congress of the Regional Science Association*. Dortmund. 2002.

ITE. Institute of Transportation Engineers. *Trip generation: an ITE informational report*. Institute of Transportation Engineers, Washington, DC, United States. 2008.

JALLER, M.; HOLGUÍN-VERAS, J.; SÁNCHEZ-DÍAZ, I.; LAWSON, C. T. Area based freight trip generation models. *Transportation Research Board (TRB) 93rd Annual Meeting*, Washington, D.C. 2013.

JALLER, M.; SÁNCHEZ-DÍAZ, I.; HOLGUÍN-VERAS, J. Identifying freight intermediaries: implications for modeling of freight trip generation miguel. *Transportation Research Record: Journal of the Transportation Research Board*. v. 2478, p. 48–56, 2015. DOI: <https://doi.org/10.3141/2478-06>.

JIN, T. G.; SAITO, M.; EGGETT, D. L. Developing county-level commodity-flow models incorporating land-use characteristics and economic factors for Utah. *Journal Of Urban Planning And Development*, American Society of Civil Engineers (ASCE) [s.l.], v. 138, n. 1, p. 35-42, 2012. [http://dx.doi.org/10.1061/\(asce\)up.1943-5444.0000093](http://dx.doi.org/10.1061/(asce)up.1943-5444.0000093).

KASZUBOWSKI, D. Urban freight transport demand modelling and data availability constraints. *Advanced Solutions Of Transport Systems For Growing Mobility*, p. 172-182, 2017. DOI: [https://doi.org/10.1007/978-3-319-62316-0\\_14](https://doi.org/10.1007/978-3-319-62316-0_14).

- KAWAMURA, K.; SHIN, H. S.; MCNEIL, S.; OGARD, L. *Business and site specific trip generation methodology for truck trips*. Midwest Regional University Transportation Center, University of Wisconsin–Madison. Chicago, 2008.
- KITCHENHAM, B.; CHARTERS, S. *Guidelines for performing systematic literature reviews in software engineering*. 2007.
- KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International joint Conference on artificial intelligence*. [S.l.: s.n.], v. 14, p. 1137–1145. 1995.
- KLODZINSKI, J.; AHMAD, A. D.; GEORGIPOULOS, M.; AL-DEEK, H. Development of a java applet for generating truck trips from freight data. *Transportation Research Record*. v. 1870, n. 1, p. 10–17, 2004. DOI: <https://doi.org/10.3141/1870-02>.
- KNEIB, E. C. *Caracterização de empreendimentos geradores de viagens: contribuição conceitual à análise de seus impactos no uso, ocupação e valorização do solo urbano*. 2004.168f. Tese (Doutorado) – Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, 2004.
- KNEIB E. C.; SILVA P. C. M. e TACO P. W. G. *Identificação e avaliação de impactos na mobilidade: análise aplicada a polos geradores de viagem*. Documento de trabalho - Universidade de Brasília – UnB, Brasília, 2006.
- KULPA, T. Freight truck trip generation modelling at regional level. *Procedia – Social and Behavioral Sciences*. v. 111, p. 197-202, 2014. DOI: <https://doi.org/10.1016/j.sbspro.2014.01.052>.
- LANNA, A. E. *Elementos de estatística e probabilidades*. In: Tucci, C.E.M (org.). *Hidrologia: ciência e aplicação*. UFRGS, ABRH, Porto Alegre, Brasil. p. 79-176. 2004.
- LAWSON, C. T.; HOLGUÍN-VERAS, J.; SÁNCHEZ-DÍAZ, I.; JALLER, M.; CAMPBELL, S.; POWERS, E. L. Estimated generation of freight trips based on land use. *Transportation Research Record: Journal of the Transportation Research Board*. v. 2269, p. 65-72, 2012. DOI: <http://doi.org/10.3141/2269-08>.

- LIM, R; QIAN, Z; ZHANG, H. M. Development of a freight demand model with an application to California. *Transportation Science and Technology*. [s.l.], v. 3, n. 1, p.19-38, 2014. DOI: <https://doi.org/10.1260/2046-0430.3.1.19>.
- LOPES, S. B. *Efeitos da dependência espacial em modelos de previsão de demanda por transportes*. 2005. 153 f. Dissertação (Mestrado em Engenharia Civil). Universidade de São Paulo. São Carlos, SP. 2005.
- MAGALHAES, S. L. M. *Planejamento de transporte de um aglomerado urbano com intermodalidade por meio de transportes integrados*. 2008. Tese (Doutorado). Engenharia de Transportes, Universidade Federal do Rio de Janeiro, RJ. 2008.
- MAROCO, J. *Análise estatística com a utilização do SPSS*. 3. ed. Edições Silabo, 2007. 842p.
- MARRA, C. *Caracterização de demanda de movimentações urbanas de carga*. 1999. 165 f. Dissertação (Mestrado). Faculdade de Engenharia Civil, Universidade Estadual de Campinas, Campinas, SP. 1999.
- MCCULLAGH, P.; NELDER, J. A. *Generalized Linear Models*. Second edition. Chapman and Hall: New York. 1989. p. 526.
- MELO, I. C. B. *Avaliação de demanda por transporte de carga em áreas urbanas*. 2002. 173 f. Dissertação (Mestrado). Instituto Militar de Engenharia, Rio de Janeiro. 2002.
- MESSETI, A. V. L. *Análise de correlação e regressão*. Curso de especialização “Lato Sensu” em estatística. Universidade Estadual de Londrina, Londrina, 2013.
- MOMMENS, K.; LIER, V. T.; MACHARIS, C. Loading unit in freight transport modelling. *Procedia Computer Science*, v. 83, p. 921-927. 2016. DOI: <https://doi.org/10.1016/j.procs.2016.04.186>.
- MONTGOMERY, D. C.; RUNGER, G. C. *Applied statistics and probability for engineers*. 6th. Ed. John Wiley & Sons, Inc, 2014.
- MOORE, D. S.; NOTZ, W. I.; FLIGNER, M. A. *The basic practice of statistics*. Seventh Edition. New York, Freeman. 2007. 730p.

NOVAK, D. C.; HODGDON, C.; GUO, F.; AULTMAN-HALL, L. Nationwide freight generation models: a spatial regression approach. *Networks and Spatial Economics*, v. 11, n. 1, p. 23–41. 2011. DOI: <https://doi.org/10.1007/s11067-008-9079-2>.

NUZZOLO, A.; CRISALLI, U.; COMI, A. A restocking tour model for the estimation of O-D freight vehicle in urban areas. *Procedia Social and Behavioral Sciences*, v. 20, p. 140-149, 2011. DOI: <https://doi.org/10.1016/j.sbspro.2011.08.019>.

OECD - *Delivering the goods - 21<sup>st</sup> century challenges to urban goods transport*. Organisation For Economic Co-Operation And Development. Paris, France. 2003. Disponível em: < <https://www.itf-oecd.org/delivering-goods-21st-century-challenges-urban-goods-transport> > Acesso em outubro de 2017.

OGDEN, K. W. Urban goods movement: a guide to policy and planning. *Transportation Research Board - TRB*. Ashgate Publishing Company, Burlington. 1992. RESEARCH ISBN 1857420292.

OGDEN, K. W. Modelling urban freight generation. *Transportation Research Board - TRB*. Hemming Group, London, v. 18, issue 3, p. 106-109, 1977. RESEARCH ISSN 0041-0683.

OLIVARES, A. M; FORERO, C. G. Goodness-of-Fit Testing. *International Encyclopedia of Education*, v. 7, p. 190-196, 2010. DOI: 10.1016/B978-0-08-044894-7.01333-6.

OLIVEIRA, L. K. Diagnóstico das vagas de carga e descarga para a distribuição urbana de mercadorias: um estudo de caso em Belo Horizonte. *Journal of Transport Literature*, v.8, n.1, p.178-209, 2014. DOI: <http://dx.doi.org/10.1590/S2238-10312014000100009>.

OLIVEIRA, L. K.; NOBREGA, R. A. A.; EBIAS, D. G.; CORREA, B. G. S. Analysis of freight trip generation model for food and beverage in Belo Horizonte (Brazil). *Region*, 4(1), 17-30, 2017. DOI: <https://doi.org/10.18335/region.v4i1.102>.

OLIVEIRA, L. K.; OLIVEIRA, R. L. M.; RAMOS, C. M. F.; EBIAS, D. G. Modelo de geração de viagens de carga em áreas urbanas: um estudo para bares, restaurantes e supermercados. *Transportes*, v. 24, n. 2, p. 53-67, 2016. DOI: <http://dx.doi.org/10.14295/transportes.v24i2.1058>.

- OLIVEIRA, L. K.; STUBBS, L. C. M.; GONTIJO, N. T.; OLIVEIRA, R. L. M. Proposição de modelos de geração de viagens para Belo Horizonte. *Transportes*, v. 25, n. 2, p. 137-155, 2017. DOI: <http://dx.doi.org/10.14295/transportes.v25i2.1243>.
- ORTÚZAR, J. D.; WILLUMSEN, L. G. *Modelling transport*. 4<sup>th</sup> Edition. John Wiley & Sons, Chichester, England. 2011. 607p.
- ORTÚZAR, J. D.; WILLUMSEN, L. G. *Modelling transport*. New Jersey: Wiley, 1994.
- PANI, A.; SAHU, P. K.; PATIL, G. R.; SARKAR, A. K. Modelling urban freight generation: a case study of seven cities in Kerala, India. *Transport Policy*, v. 69, p. 49-64, 2018. DOI: <https://doi.org/10.1016/j.tranpol.2018.05.013>.
- PENA, R. F. A. Economias de aglomeração. In: *Brasil Escola*. Disponível em: <<https://brasilecola.uol.com.br/geografia/economias-aglomeracao.htm>>. Acesso em: 06 de maio de 2018.
- PORTUGAL, L. S. *et al. Polos geradores de viagens orientadas à qualidade de vida e ambiental: modelos e taxas de viagens*. 1 ed. Rio de Janeiro: Interciência, 2012. 704 p.
- PRIYA, C. D.; RAMADURAI, G.; DEVI, G. Freight trip generation models for Chennai, India. *Transportation Research Board (TRB) 95rd Annual Meeting*, Washington, D.C. 2015.
- RAMOS, C. M. F. *Características operacionais e geração de viagens de carga para supermercados*. 2015. 103 f. Dissertação (Mestrado). Escola de Engenharia da UFMG. Universidade Federal de Minas Gerais, Belo Horizonte. 2015.
- ROLKO, K.; FRIEDRICH, H. Locations of logistics service providers in Germany - the basis for a new freight transport generation model. *Transportation Research Procedia*, [s.l.], v. 25, p.1061-1074, 2017. Elsevier BV. DOI: <http://dx.doi.org/10.1016/j.trpro.2017.05.479>.
- RUSSO, F.; COMI, A. Urban freight movement: a quantity attraction model. *First International Conference on Sustainable Planning and Development*. v. 60, p. 711–720, 2002. DOI: 10.2495/UT020801.

- SÁNCHEZ-DÍAZ, I. Modeling urban freight generation: A study of commercial establishments' freight needs. *Transportation Research Part A: Policy and Practice*, [s.l.], v. 102, p.3-17, 2016. Elsevier BV. DOI: <http://dx.doi.org/10.1016/j.tra.2016.06.035>.
- SÁNCHEZ-DÍAZ, I.; HOLGUÍN-VERAS, J.; BAN, X. J. A time-dependent freight tour synthesis model. *Transportation Research Part B: Methodological*, [s.l.], v. 78, p.144-168, 2015. Elsevier BV. DOI: <http://dx.doi.org/10.1016/j.trb.2015.04.007>.
- SÁNCHEZ-DÍAZ, I.; HOLGUÍN-VERAS, J.; WANG, X. C. An exploratory analysis of spatial effects on freight trip attraction. *Transportation*. v. 43, n. 1, p. 177–196, 2016. DOI: <https://doi.org/10.1007/s11116-014-9570-1>.
- SARVAREDDY, P.; AL-DEEK, H.; KLODZINSKI, J.; ANAGNOSTOPOULOS, G. Evaluation of two modeling methods for generating heavy-truck trips at an intermodal facility by using vessel freight data. *Transportation Research Record: Journal of the Transportation Research Board*. v. 1906, n. 1, p. 113–120, 2005. DOI: <https://doi.org/10.3141/1906-14>.
- SATORRA, A.; BENTLER, P. M. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, v. 66, n. 4, p.507-514, 2001. DOI: <http://dx.doi.org/10.1007/bf02296192>.
- SIEGEL, S.; CASTELLAN JR, N. J. *Estatística não-paramétrica para ciências do comportamento*. 2. ed. Porto Alegre: Artmed, 2006, 448p.
- SILVA, F. O.; FERREIRA, W. R. Periodização e evolução da logística urbana de cargas. *InterEspaço Revista de Geografia e Interdisciplinaridade*. v. 3, n. 8, p. 142-158, 2017. DOI: <http://dx.doi.org/10.18764/2446-6549.v3n8p142-158>.
- SILVA, M. O.; GOLDNER, L. G. Determinação dos padrões de viagens e taxas de geração de viagens de automóveis e caminhões para indústrias: o caso de Tubarão/SC. *Revista Transportes*. v. XVIII, n. 1, p. 66–75, 2010. DOI: <http://dx.doi.org/10.14295/transportes.v18i1.384>.
- SILVA, M. R.; WAISMAN, J. Cargas urbanas: estudo exploratório sobre a geração de viagens de caminhões em bares e restaurantes. In: *16º Congresso Brasileiro de Transporte e Trânsito*, Maceió. 2007.

SOUZA, C. D. R.; SILVA, S. D.; D'AGOSTO, M. A. Modelos de geração de viagem para polos geradores de viagens de carga. *Transportes*, v. 18, n. 1, p. 46-57, 2010. DOI: <http://dx.doi.org/10.14295/transportes.v18i1.396>.

STEVENSON, W. J. *Estatística aplicada à administração*. São Paulo: Harper & Row do Brasil, 1981. Editora Harbra. 498 p.

TADI, R. R.; BALBACH, P. Truck trip generation characteristics of nonresidential land uses. *Institute of Transportation Engineers – ITE*. v. 64, issue 7, p. 43-47, Washington, 1994. RESEARCH ISSN 0162-8178.

TEDESCO, G. M. I. *Metodologia para elaboração do diagnóstico de um sistema de transporte*. 2008. 215 f. Dissertação (Mestrado). Faculdade de Tecnologia, Departamento de Engenharia Civil e Ambiental. Universidade de Brasília, Distrito Federal. 2008.

VALENTE, A. M.; NOVAES, A. G.; PASSAGLIA, E.; VIEIRA, H. *Gerenciamento de Transporte e Frotas*. 2. ed. São Paulo: Cengage Learning, 2008. 352p.

VENABLES, W. N.; RIPLEY, B. D. *Modern Applied Statistics with S*. Second edition. Springer: New York. 2002. 504p.

WANG, X. C.; ZHOU, Y. Deliveries to residential units: a rising form of freight transportation in the U.S.. *Transportation Research Part C: Emerging Technologies*, [s.l.], v. 58, p.46-55, set. 2015. Elsevier BV. DOI: <http://dx.doi.org/10.1016/j.trc.2015.07.004>.

WASSERSTEIN, R. L.; LAZAR, N. A. The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, [s.l.], v. 70, n. 2, p.129-133, 2016. Informa UK Limited. DOI: <http://dx.doi.org/10.1080/00031305.2016.1154108>.

WEGMANN, F. J.; CHATTERJEE, A.; LIPINSKI, M. E.; JENNINGS, B. E.; MCGINNIS, R. E. Characteristics of urban freight systems. Washington, DC: U.S. Department of Transportation, 1995.

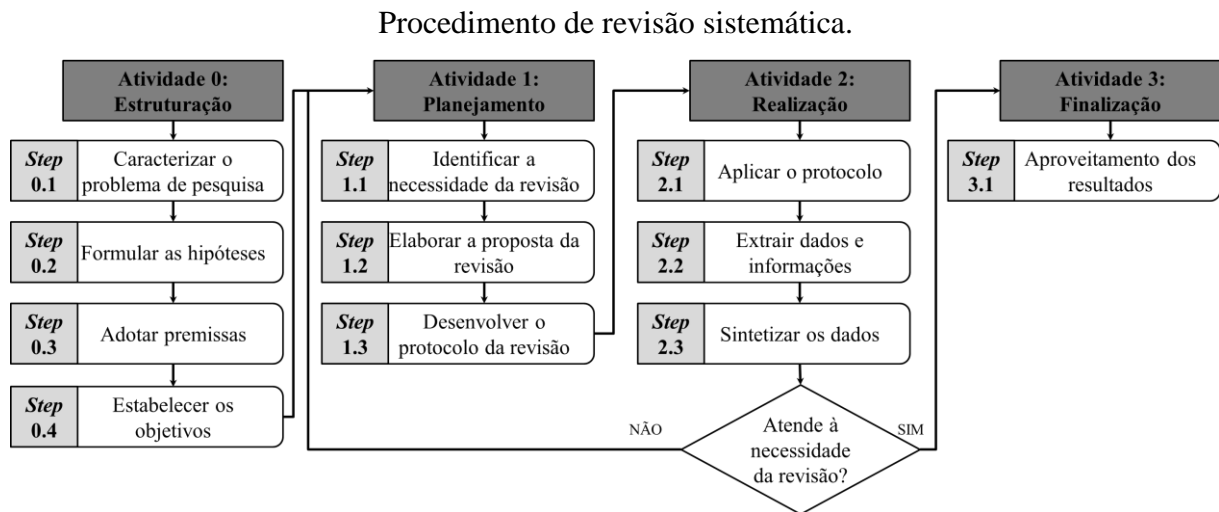
WISETJINDAWAT, W.; SANO, K.; MATSUMOTO, S. Commodity distribution model incorporating spatial interactions for urban freight movement. *Transportation Research*

*Record: Journal of the Transportation Research Board.* v. 1966, p. 41–50, 2006. DOI: <https://doi.org/10.3141/1966-06>.

WITTEN, I. E.; FRANK, E.; HALL, M. A. *DATA MINING: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann, 3rd edition. 2011.



## APÊNDICE A – PROCEDIMENTO DE REVISÃO SISTEMÁTICA



Fonte: Adaptado de Kitchenham e Charters (2007).

Esse procedimento é constituído de quatro atividades a saber:

- i) atividade 0, corresponde a uma fase preliminar de estruturação da revisão, em que se caracteriza o problema de pesquisa, se formulam as hipóteses, adotam-se premissas e estabelecem-se os objetivos do trabalho;
- ii) atividade 1, representa o planejamento da revisão, sendo a etapa em que de fato se inicia à revisão sistemática. O primeiro passo dessa atividade, passo 1.1, consiste na identificação da necessidade de revisão, em que se observam os objetivos do trabalho para determinar as razões pelas quais deve ser feita a revisão sistemática. No passo 1.2, desenvolve-se a proposta da revisão e especificando o que será procurado nos trabalhos consultados. Por fim, passo 1.3, desenvolve-se o protocolo da revisão, definindo critérios de identificação de trabalhos, detalhando a forma como os trabalhos serão coletados, a base de dados a ser consultada, as palavras-chaves utilizadas na busca e o local de busca no texto, e tipos de trabalhos procurados. Além disso, são estabelecidos critérios de seleção, inclusão e exclusão dos trabalhos, de modo a filtrar o conteúdo encontrado e cumprir o objetivo da revisão.
- iii) atividade 2 realiza-se a revisão, onde põe-se em prática o protocolo, no passo 2.1, de modo a criar um grupo de trabalhos potencialmente importantes para o estudo, observando os critérios de identificação, seleção e inclusão estabelecidos no protocolo. Os trabalhos

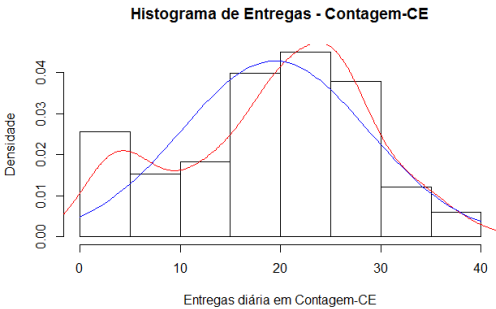
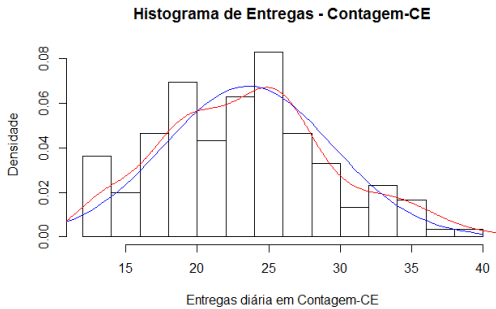
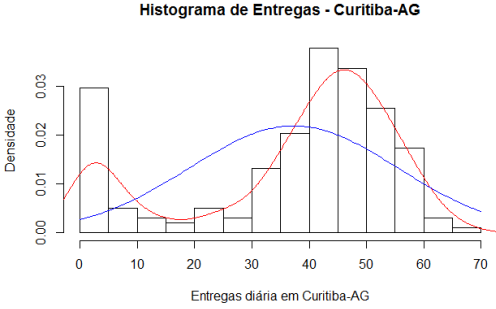
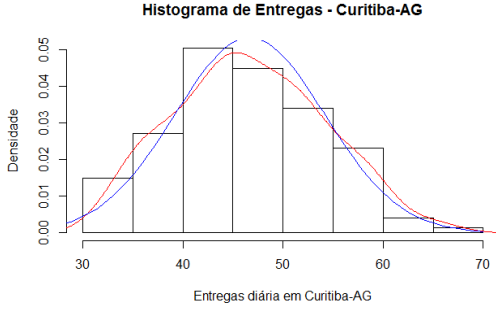
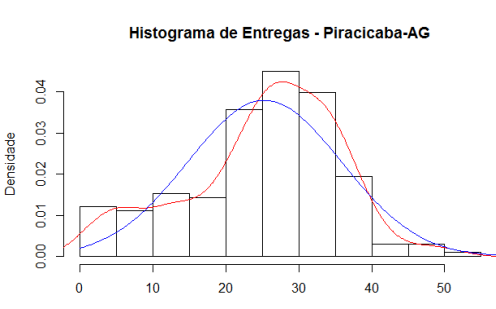
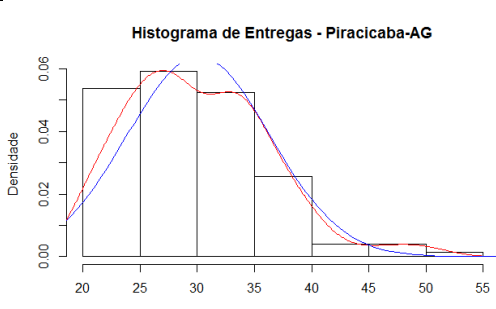
incluídos são consultados e selecionam-se dados e informações relevantes para cumprir a proposta da revisão. Em seguida, faz-se uma avaliação dos resultados com o intuito de verificar se as necessidades da revisão foram cumpridas e se os dados selecionados são suficientes para captar o estado da arte do tema estudado. Caso o cumprimento da proposta de revisão não seja satisfatório, deve-se retornar ao passo 1.3 e revisar os critérios do protocolo, realizando ações que direcionem o melhor caminho para se obter a revisão, como por exemplo: estipular novas palavras-chave, escolher base de dados diferentes, alterar critérios de inclusão e exclusão. Caso seja positivo o cumprimento da proposta de revisão, dá-se continuidade ao procedimento, iniciando-se à Atividade 3.

- iv) a atividade 3 representa o aproveitamento dos resultados da revisão, seja como parte integrante da dissertação, tese, artigo, anais, entre outros.

## APÊNDICE B – TESTES REALIZADOS NOS DADOS

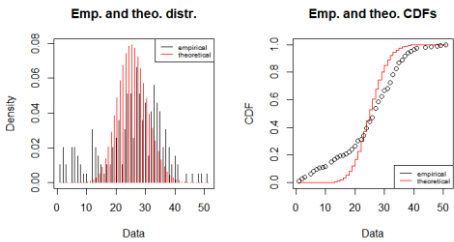
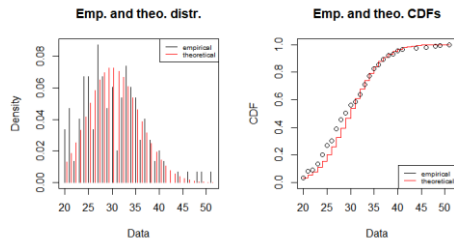
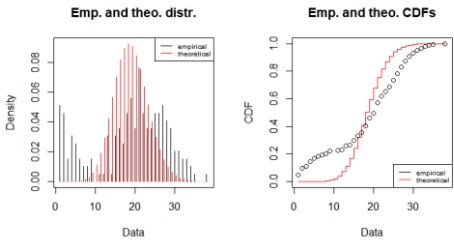
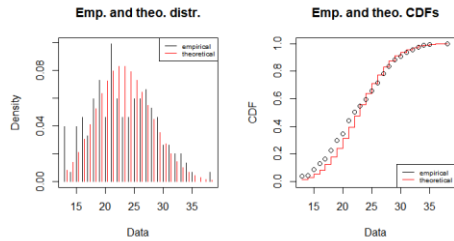
Histograma dos dados, curva normal com média e desvio padrão, teste Shapiro Wilk

	Contagem-CE	Curitiba-AG	Piracicaba-AG	Sorocaba-AG
Entregas recebida	eco	ecu	epi	eso
Carga recebida	cco	ccu	cpi	cso

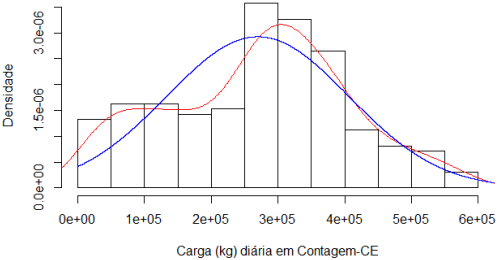
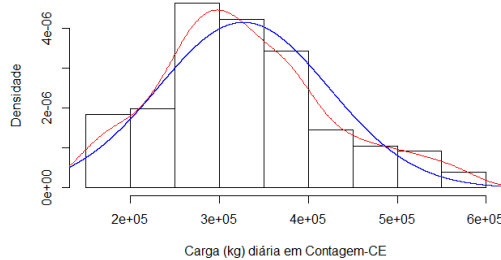
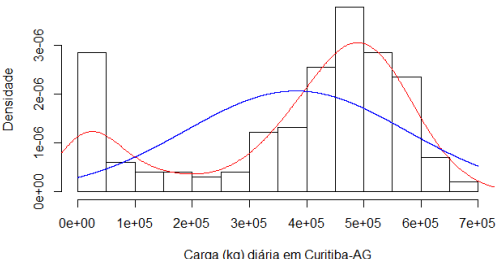
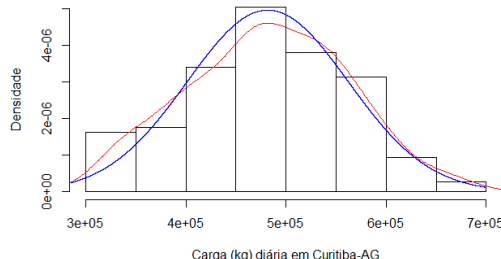
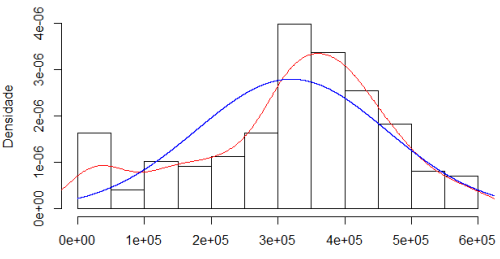
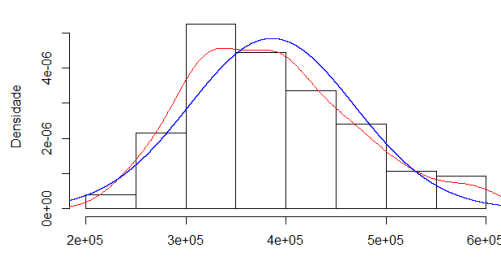
<b>ENTREGAS - NORMAL</b>		
	Considerando 196 obs.	Sem 1º quartil
<b>eco</b>	<p style="text-align: center;"><b>Histograma de Entregas - Contagem-CE</b></p>  <p style="text-align: center;">Entregas diária em Contagem-CE</p> <p>Shapiro-Wilk normality test            data: rtodos\$eco            W = 0.95762, p-value = 1.345e-05            Min. 1st Qu. Median Mean 3rd Qu. Max.            1.00 13.00 21.00 19.43 26.00 40.00</p>	<p style="text-align: center;"><b>Histograma de Entregas - Contagem-CE</b></p>  <p style="text-align: center;">Entregas diária em Contagem-CE</p> <p>Shapiro-Wilk normality test            data: new_eco            W = 0.97938, <b>p-value = 0.02285</b>            Min. 1st Qu. Median Mean 3rd Qu. Max.            13.00 19.00 24.00 23.57 27.00 40.00</p>
<b>ecu</b>	<p style="text-align: center;"><b>Histograma de Entregas - Curitiba-AG</b></p>  <p style="text-align: center;">Entregas diária em Curitiba-AG</p> <p>Shapiro-Wilk normality test            data: rtodos\$ecu            W = 0.85898, p-value = 1.682e-12            Min. 1st Qu. Median Mean 3rd Qu. Max.            1.00 31.75 44.00 37.31 50.00 66.00</p>	<p style="text-align: center;"><b>Histograma de Entregas - Curitiba-AG</b></p>  <p style="text-align: center;">Entregas diária em Curitiba-AG</p> <p>Shapiro-Wilk normality test            data: new_ecu            W = 0.98628, <b>p-value = 0.154</b>            Min. 1st Qu. Median Mean 3rd Qu. Max.            32.00 41.00 46.00 46.69 52.00 66.00</p>
<b>epi</b>	<p style="text-align: center;"><b>Histograma de Entregas - Piracicaba-AG</b></p>  <p style="text-align: center;">Entregas diária em Piracicaba-AG</p> <p>Shapiro-Wilk normality test</p>	<p style="text-align: center;"><b>Histograma de Entregas - Piracicaba-AG</b></p>  <p style="text-align: center;">Entregas diária em Piracicaba-AG</p> <p>Shapiro-Wilk normality test</p>

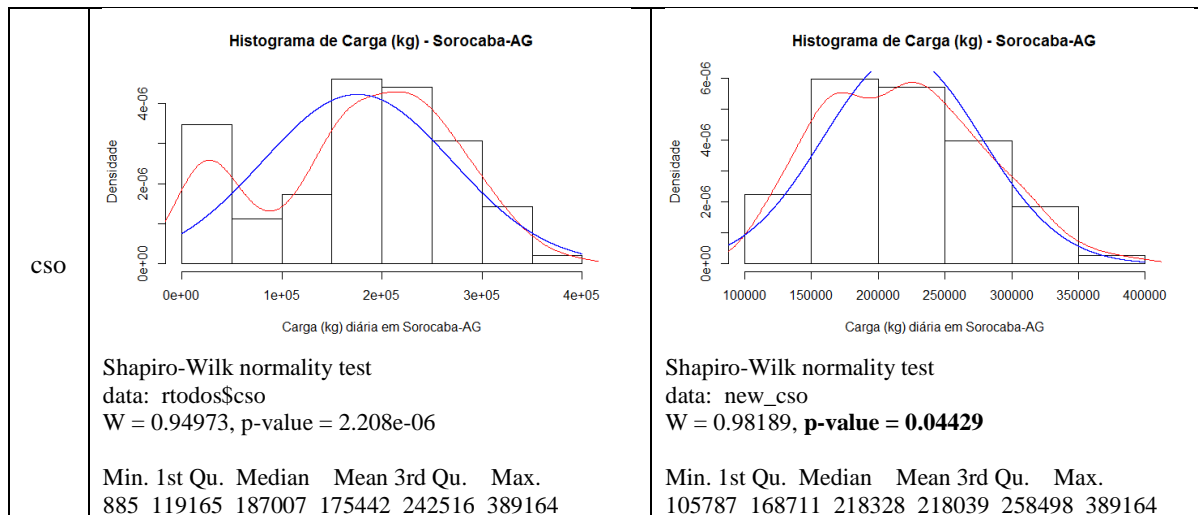
	<p>data: rtodos\$epi  <math>W = 0.9628</math>, <math>p\text{-value} = 4.833e-05</math>  Min. 1st Qu. Median Mean 3rd Qu. Max.  1.00 20.00 27.00 25.31 33.00 51.00</p>	<p>data: new_epi  <math>W = 0.9628</math>, <math>p\text{-value} = 0.0004703</math>  Min. 1st Qu. Median Mean 3rd Qu. Max.  20.00 25.00 29.00 30.11 34.00 51.00</p>
eso	<p><b>Histograma de Entregas - Sorocaba-AG</b></p> <p>Shapiro-Wilk normality test  data: rtodos\$eso  <math>W = 0.93615</math>, <math>p\text{-value} = 1.372e-07</math>  Min. 1st Qu. Median Mean 3rd Qu. Max.  1.00 13.00 21.00 18.63 26.00 38.00</p>	<p><b>Histograma de Entregas - Sorocaba-AG</b></p> <p>Shapiro-Wilk normality test  data: new_eso  <math>W = 0.98281</math>, <math>p\text{-value} = 0.05645</math>  Min. 1st Qu. Median Mean 3rd Qu. Max.  13.00 19.00 22.00 22.97 27.00 38.00</p>

<b>ENTREGAS - POISSON</b>	
Considerando 196 obs.	
Sem 1º quartil	
eco	<div style="display: flex; justify-content: space-around;"> <div style="width: 45%;"> <p><b>Emp. and theo. distr.</b></p> <p><b>Emp. and theo. CDFs</b></p> </div> <div style="width: 45%;"> <p><b>Emp. and theo. distr.</b></p> <p><b>Emp. and theo. CDFs</b></p> </div> </div> <p>Fitting of the distribution ' pois ' by maximum likelihood  Parameters :  estimate Std. Error  lambda 19.43367 0.3148831  Loglikelihood: -979.4498 AIC: 1960.9 BIC: 1964.178</p> <p>Pearson's Chi-squared test  data: eco and eco_poisson  X-squared = 740.33, df = 805, p-value = 0.9495</p>
ecu	<div style="display: flex; justify-content: space-around;"> <div style="width: 45%;"> <p><b>Emp. and theo. distr.</b></p> <p><b>Emp. and theo. CDFs</b></p> </div> <div style="width: 45%;"> <p><b>Emp. and theo. distr.</b></p> <p><b>Emp. and theo. CDFs</b></p> </div> </div> <p>Fitting of the distribution ' pois ' by maximum likelihood  Parameters :  estimate Std. Error  lambda 37.30612 0.4362767</p> <p>Fitting of the distribution ' pois ' by maximum likelihood  Parameters :  estimate Std. Error  lambda 46.68707 0.5635594</p>

	<p>Loglikelihood: -1735.774 AIC: 3473.548 BIC: 3476.826</p> <p>Pearson's Chi-squared test data: rtodos\$ecu and ecu_poisson X-squared = 1373.5, df = 1350, p-value = 0.3221</p>	<p>Loglikelihood: -504.8723 AIC: 1011.745 BIC: 1014.735</p> <p>Pearson's Chi-squared test data: new_ecu and new_ecu_poisson X-squared = 1100.8, df = 992, <b>p-value = 0.008793</b></p>
epi	 <p>Emp. and theo. distr. Emp. and theo. CDFs</p> <p>Fitting of the distribution ' pois ' by maximum likelihood Parameters : estimate Std. Error lambda 25.31122 0.359359 Loglikelihood: -1001.216 AIC: 2004.432 BIC: 2007.711</p> <p>Pearson's Chi-squared test data: rtodos\$epi and epi_poisson X-squared = 881.26, df = 989, p-value = 0.9938</p>	 <p>Emp. and theo. distr. Emp. and theo. CDFs</p> <p>Fitting of the distribution ' pois ' by maximum likelihood Parameters : estimate Std. Error lambda 30.11409 0.4495642 Loglikelihood: -483.3818 AIC: 968.7636 BIC: 971.7675</p> <p>Pearson's Chi-squared test data: new_epi and new_epi_poisson X-squared = 815.24, df = 676, <b>p-value = 0.0001771</b></p>
eso	 <p>Emp. and theo. distr. Emp. and theo. CDFs</p> <p>Fitting of the distribution ' pois ' by maximum likelihood Parameters : estimate Std. Error lambda 18.63265 0.3083254 Loglikelihood: -1031.163 AIC: 2064.327 BIC: 2067.605</p> <p>Pearson's Chi-squared test data: rtodos\$eso and eso_poisson X-squared = 686.23, df = 680, p-value = 0.426</p>	 <p>Emp. and theo. distr. Emp. and theo. CDFs</p> <p>Fitting of the distribution ' pois ' by maximum likelihood Parameters : estimate Std. Error lambda 22.96689 0.3899982 Loglikelihood: -471.5458 AIC: 945.0916 BIC: 948.1089</p> <p>Pearson's Chi-squared test data: new_eso and new_eso_poisson X-squared = 481.51, df = 506, p-value = 0.7769</p>

## CARGA - NORMAL

	Considerando 196 obs.	Sem 1º quartil
CCO	<p style="text-align: center;"><b>Histograma de Carga (kg) - Contagem-CE</b></p>  <p style="text-align: center;">Carga (kg) diária em Contagem-CE</p> <p>Shapiro-Wilk normality test data: rtodos\$cco W = 0.97641, p-value = 0.002175</p> <p>Min. 1st Qu. Median Mean 3rd Qu. Max. 3045 160559 284795 269698 364110 576713</p>	<p style="text-align: center;"><b>Histograma de Carga (kg) - Contagem-CE</b></p>  <p style="text-align: center;">Carga (kg) diária em Contagem-CE</p> <p>Shapiro-Wilk normality test data: new_cco W = 0.97506, p-value = 0.007565</p> <p>Min. 1st Qu. Median Mean 3rd Qu. Max. 150858 264654 311808 327120 382417 576713</p>
CCU	<p style="text-align: center;"><b>Histograma de Carga (kg) - Curitiba-AG</b></p>  <p style="text-align: center;">Carga (kg) diária em Curitiba-AG</p> <p>Shapiro-Wilk normality test data: rtodos\$ccu W = 0.86053, p-value = 2.014e-12</p> <p>Min. 1st Qu. Median Mean 3rd Qu. Max. 2030 304045 448104 381275 523935 672784</p>	<p style="text-align: center;"><b>Histograma de Carga (kg) - Curitiba-AG</b></p>  <p style="text-align: center;">Carga (kg) diária em Curitiba-AG</p> <p>Shapiro-Wilk normality test data: new_ccu W = 0.98977, p-value = 0.3612</p> <p>Min. 1st Qu. Median Mean 3rd Qu. Max. 308504 422774 481424 481747 537946 672784</p>
cpi	<p style="text-align: center;"><b>Histograma de Carga (kg) - Piracicaba-AG</b></p>  <p style="text-align: center;">Carga (kg) diária em Piracicaba-AG</p> <p>Shapiro-Wilk normality test data: rtodos\$cpi W = 0.95539, p-value = 7.941e-06</p> <p>Min. 1st Qu. Median Mean 3rd Qu. Max. 3471 242581 337533 319925 412671 596300</p>	<p style="text-align: center;"><b>Histograma de Carga (kg) - Piracicaba-AG</b></p>  <p style="text-align: center;">Carga (kg) diária em Piracicaba-AG</p> <p>Shapiro-Wilk normality test data: new_cpi W = 0.97113, p-value = 0.003141</p> <p>Min. 1st Qu. Median Mean 3rd Qu. Max. 228463 324138 378784 385800 434765 596300</p>



Correlação Spearman's rank para variáveis dependentes

## CORRELAÇÃO ENTRE VARIÁVEIS DEPENDENTES

Spearman's rank correlation rho

data: new\_eco and new\_cco

S = 1053, **p-value < 2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates: rho 0.9981649

Spearman's rank correlation rho

data: new\_ecu and new\_ccu

S = 569.81, **p-value < 2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates: rho 0.9989237

Spearman's rank correlation rho

data: new\_epi and new\_cpi

S = 869.69, **p-value < 2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates: rho 0.9984225

Spearman's rank correlation rho

data: new\_eso and new\_cso

S = 953.3, **p-value < 2.2e-16**

alternative hypothesis: true rho is not equal to 0

sample estimates: rho 0.9983386

# APÊNDICE C – FUNÇÕES DESENVOLVIDAS NO SOFTWARE R

## Regressão Linear Generalizada Poisson - Entregas

glm01	<pre> glm(formula = ent ~ prod, family = poisson(), data = Modelo)  Deviance Residuals:     1     2     3     4 -0.9534  0.3031  1.2386 -0.6667  Coefficients:             Estimate Std. Error z value Pr(&gt; z ) (Intercept) 3.290426  0.131318  25.057 &lt; 2e-16 *** prod        0.005241  0.001791   2.927 0.00342 ** --- Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  (Dispersion parameter for poisson family taken to be 1)  Null deviance: 11.1748 on 3 degrees of freedom Residual deviance: 2.9795 on 2 degrees of freedom AIC: 28.421  Number of Fisher Scoring iterations: 4  p-value for the deviance goodness of fit test: [1] 0.225432 The null hypothesis is that our model is correctly specified – NOT REJECT  &gt; AICcmodavg::AICc(glm01) [1] 40.42071  &gt; AICcmodavg::c_hat(glm01) 'c-hat' 1.52 (method: pearson estimator)  &gt; cv.glm(Modelo,glm01)\$delta [1] 423.8382 332.1800 </pre>
glm02	<pre> glm(formula = ent ~ functot, family = poisson(), data = Modelo)  Deviance Residuals:     1     2     3     4 -0.9679  1.7217  0.9605 -1.9441  Coefficients:             Estimate Std. Error z value Pr(&gt; z ) (Intercept) 3.3206104  0.1751614  18.96 &lt;2e-16 *** functot     0.0001819  0.0001144   1.59  0.112 --- Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  (Dispersion parameter for poisson family taken to be 1)  Null deviance: 11.1748 on 3 degrees of freedom Residual deviance: 8.6029 on 2 degrees of freedom </pre>



	<p style="text-align: center;"><b>AIC: 34.044</b></p> <p style="text-align: center;">Number of Fisher Scoring iterations: 4</p> <p style="text-align: center;">p-value for the deviance goodness of fit test: <b>[1] 0.01354863</b>  The null hypothesis is that our model is correctly specified – NOT REJECT???</p> <pre style="text-align: center;"> &gt; AICcmodavg::AICc(glm02) [1] 46.04418  &gt; AICcmodavg::c_hat(glm02) 'c-hat' 4.24 (method: pearson estimator)  &gt; cv.glm(Modelo,glm02)\$delta [1] 348.8268 304.8629 </pre>
glm03	<pre style="text-align: center;"> glm(formula = ent ~ funclog, family = poisson(), data = Modelo)  Deviance Residuals:     1     2     3     4 -0.7339  0.4958  1.3223 -1.1855  Coefficients:             Estimate Std. Error z value Pr(&gt; z ) (Intercept) 3.1720808  0.1719465  18.448 &lt; 2e-16 *** funclog      0.0020450  0.0007485   2.732  0.00629 ** --- Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  (Dispersion parameter for poisson family taken to be 1)  Null deviance: 11.1748 on 3 degrees of freedom Residual deviance: <b>3.9384</b> on 2 degrees of freedom <b>AIC: 29.38</b>  Number of Fisher Scoring iterations: 4  p-value for the deviance goodness of fit test: <b>[1] 0.139568</b>  The null hypothesis is that our model is correctly specified – NOT REJECT  &gt; AICcmodavg::AICc(glm03) [1] 41.37964  &gt; AICcmodavg::c_hat(glm03) 'c-hat' 1.99 (method: pearson estimator)  &gt; cv.glm(Modelo,glm03)\$delta [1] 376.3704 301.8190 </pre>
glm04	<pre style="text-align: center;"> glm(formula = ent ~ areatot, family = poisson(), data = Modelo)  Deviance Residuals:     1     2     3     4 -0.4534  0.2081  1.1235 -0.9426 </pre>

	<pre> Coefficients:       Estimate Std. Error z value Pr(&gt; z ) (Intercept) 3.202e+00 1.524e-01 21.014 &lt; 2e-16 *** areatot    1.501e-06 4.943e-07  3.037 0.00239 ** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  (Dispersion parameter for poisson family taken to be 1)  Null deviance: 11.1748 on 3 degrees of freedom Residual deviance: <b>2.3996</b> on 2 degrees of freedom <b>AIC: 27.841</b>  Number of Fisher Scoring iterations: 4  p-value for the deviance goodness of fit test: <b>[1] 0.3012576</b> The null hypothesis is that our model is correctly specified – NOT REJECT  &gt; AICcmodavg::AICc(glm04) [1] 39.84082  &gt; AICcmodavg::c_hat(glm04) 'c-hat' 1.22 (method: pearson estimator)  &gt; cv.glm(Modelo,glm04)\$delta [1] 419.2919 325.2183 </pre>
glm05	<pre> glm(formula = ent ~ arealog, family = poisson(), data = Modelo)  Deviance Residuals:      1     2     3     4 -1.335  1.217  1.343 -1.407  Coefficients:       Estimate Std. Error z value Pr(&gt; z ) (Intercept) 3.184e+00 2.095e-01 15.196 &lt;2e-16 *** arealog    1.042e-05 5.169e-06  2.015 0.0439 * --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  (Dispersion parameter for poisson family taken to be 1)  Null deviance: 11.1748 on 3 degrees of freedom Residual deviance: <b>7.0489</b> on 2 degrees of freedom <b>AIC: 32.49</b>  Number of Fisher Scoring iterations: 4  p-value for the deviance goodness of fit test: <b>[1] 0.0294683</b> The null hypothesis is that our model is correctly specified – NOT REJECT??  &gt; AICcmodavg::AICc(glm05) [1] 44.49012 </pre>

	<pre> &gt; AICcmodavg::c_hat(glm05) 'c-hat' 3.5 (method: pearson estimator)  &gt; cv.glm(Modelo,glm05)\$delta [1] 390.4031 332.9982 </pre>
glm06	<pre> glm(formula = ent ~ prod + functot, family = poisson(), data = Modelo)  Deviance Residuals:     1     2     3     4 -1.1871  0.2409  0.7597  0.1339  Coefficients:             Estimate Std. Error z value Pr(&gt; z ) (Intercept)  3.4066934  0.1752208  19.442 &lt;2e-16 *** prod         0.0075476  0.0030654   2.462  0.0138 * functot     -0.0001808  0.0001916  -0.944  0.3454 --- Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  (Dispersion parameter for poisson family taken to be 1)  Null deviance: 11.1748 on 3 degrees of freedom Residual deviance: 2.0624 on 1 degrees of freedom AIC: 29.504  Number of Fisher Scoring iterations: 4  p-value for the deviance goodness of fit test: [1] 0.150972 The null hypothesis is that our model is correctly specified – NOT REJECT  &gt; AICcmodavg::AICc(glm06) [1] Inf  &gt; AICcmodavg::c_hat(glm06) 'c-hat' 1.99 (method: pearson estimator)  &gt; cv.glm(Modelo,glm06)\$delta [1] 571.8339 445.2710 </pre>
glm07	<pre> glm(formula = ent ~ functot + funclog, family = poisson(), data = Modelo)  Deviance Residuals:     1     2     3     4 -0.9256  0.1114  0.5588  0.2234  Coefficients:             Estimate Std. Error z value Pr(&gt; z ) (Intercept)  3.2508892  0.1717950  18.923 &lt; 2e-16 *** functot     -0.0004212  0.0002636  -1.598  0.11005 funclog      0.0045229  0.0017286   2.617  0.00888 ** --- Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  (Dispersion parameter for poisson family taken to be 1) </pre>

	<p>Null deviance: 11.1748 on 3 degrees of freedom Residual deviance: <b>1.2313</b> on 1 degrees of freedom <b>AIC: 28.673</b></p> <p>Number of Fisher Scoring iterations: 4</p> <p>p-value for the deviance goodness of fit test: <b>[1] 0.2671606</b> The null hypothesis is that our model is correctly specified – NOT REJECT</p> <pre>&gt; AICcmodavg::AICc(glm07) [1] Inf  &gt; AICcmodavg::c_hat(glm07) 'c-hat' 1.2 (method: pearson estimator)  &gt; cv.glm(Modelo,glm07)\$delta [1] 548.5596 420.8957</pre>
glm08	<pre>glm(formula = ent ~ functot + funclog + areatot, family = poisson(), data = Modelo)  Deviance Residuals: [1] 0 0 0 0  Coefficients: Estimate Std. Error z value Pr(&gt; z ) (Intercept) 3.397e+00 2.103e-01 16.152 &lt;2e-16 *** functot -1.256e-04 3.758e-04 -0.334 0.738 funclog -4.801e-03 8.711e-03 -0.551 0.582 areatot 5.175e-06 4.745e-06 1.091 0.275 --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  (Dispersion parameter for poisson family taken to be 1)  Null deviance: 1.1175e+01 on 3 degrees of freedom Residual deviance: <b>5.9952e-15</b> on 0 degrees of freedom <b>AIC: 29.441</b> <p>Number of Fisher Scoring iterations: 3</p> <p>p-value for the deviance goodness of fit test: <b>[1] 0</b> The null hypothesis is that our model is correctly specified – REJECT</p> <pre>&gt; AICcmodavg::AICc(glm08) [1] -10.55876  &gt; AICcmodavg::c_hat(glm08) 'c-hat' Inf (method: pearson estimator)  &gt; cv.glm(Modelo,glm08)\$delta [1] 548.5596 411.4197</pre> </pre>
glm09	<pre>glm(formula = ent ~ areatot + arealog, family = poisson(), data = Modelo)</pre>

```

Deviance Residuals:
    1    2    3    4
0.35889 0.06382 0.05057 -0.48063

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept) 3.427e+00 2.099e-01 16.332 <2e-16 ***
  areatot   3.017e-06 1.187e-06  2.543  0.011 *
  arealog  -1.638e-05 1.142e-05 -1.434  0.152
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 11.17479 on 3 degrees of freedom
Residual deviance: 0.36644 on 1 degrees of freedom
AIC: 27.808

Number of Fisher Scoring iterations: 3

p-value for the deviance goodness of fit test: [1] 0.5449502
The null hypothesis is that our model is correctly specified – NOT REJECT

> AICcmodavg::AICc(glm09)
[1] Inf

> AICcmodavg::c_hat(glm09)
'c-hat' 0.36 (method: pearson estimator)

> cv.glm(Modelo,glm09)$delta
[1] 363.9801 275.5471

```

### Regressão Linear Generalizada Gaussiana - Carga

```

glm(formula = car ~ prod, family = gaussian(), data = Modelo)

Deviance Residuals:
    1    2    3    4
-5313 23275 95976 -113938

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 334201    76907  4.346 0.0491 *
      prod      1529     1224  1.249 0.3380
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 11381606432)

Null deviance: 4.0523e+10 on 3 degrees of freedom
Residual deviance: 2.2763e+10 on 2 degrees of freedom

```

	<p style="text-align: center;"><b>AIC: 107.2</b></p> <p style="text-align: center;">Number of Fisher Scoring iterations: 2  p-value for the deviance goodness of fit test: <b>[1] 0</b>  The null hypothesis is that our model is correctly specified – REJECT</p> <pre style="text-align: center;"> &gt; AICcmodavg::AICc(glm10) [1] Inf &gt; cv.glm(Modelo,glm10)\$delta [1] 63713203220 50710675122 </pre>
glm11	<pre style="text-align: center;"> glm(formula = car ~ functot, family = gaussian(), data = Modelo)  Deviance Residuals:     1     2     3     4 -19317 130774 35905 -147362  Coefficients:       Estimate Std. Error t value Pr(&gt; t ) (Intercept) 3.980e+05 1.367e+05  2.912  0.100       functot  4.375e+00 9.419e+01  0.046  0.967  (Dispersion parameter for gaussian family taken to be 20239798674)  Null deviance: 4.0523e+10 on 3 degrees of freedom Residual deviance: <b>4.0480e+10</b> on 2 degrees of freedom <b>AIC: 109.5</b>  Number of Fisher Scoring iterations: 2  p-value for the deviance goodness of fit test: <b>[1] 0</b> The null hypothesis is that our model is correctly specified – REJECT  &gt; AICcmodavg::AICc(glm11) [1] Inf  &gt; cv.glm(Modelo,glm11)\$delta [1] 42538239906 36959437940 </pre>
glm12	<pre style="text-align: center;"> glm(formula = car ~ funclog, family = gaussian(), data = Modelo)  Deviance Residuals:     1     2     3     4  1294  53541  86655 -141489  Coefficients:       Estimate Std. Error t value Pr(&gt; t ) (Intercept) 320757.0 118539.0  2.706  0.114       funclog   471.6   577.8  0.816  0.500  (Dispersion parameter for gaussian family taken to be 15198268642)  Null deviance: 4.0523e+10 on 3 degrees of freedom Residual deviance: <b>3.0397e+10</b> on 2 degrees of freedom <b>AIC: 108.36</b> </pre>

	<p style="text-align: center;">Number of Fisher Scoring iterations: 2</p> <p style="text-align: center;">p-value for the deviance goodness of fit test: <b>[1] 0</b> The null hypothesis is that our model is correctly specified – REJECT</p> <p style="text-align: center;">&gt; AICcmodavg::AICc(glm12) [1] Inf</p> <p style="text-align: center;">&gt; cv.glm(Modelo,glm12)\$delta [1] 86985164435 69711581017</p>
glm13	<p style="text-align: center;">glm(formula = car ~ areatot, family = gaussian(), data = Modelo)</p> <p style="text-align: center;">Deviance Residuals: 1 2 3 4 16896 30227 83600 -130723</p> <p style="text-align: center;">Coefficients: Estimate Std. Error t value Pr(&gt; t ) (Intercept) 3.181e+05 9.586e+04 3.319 0.080 . areatot 3.950e-01 3.596e-01 1.098 0.387 --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p> <p style="text-align: center;">(Dispersion parameter for gaussian family taken to be 12638201030)</p> <p style="text-align: center;">Null deviance: 4.0523e+10 on 3 degrees of freedom Residual deviance: <b>2.5276e+10</b> on 2 degrees of freedom <b>AIC: 107.62</b></p> <p style="text-align: center;">Number of Fisher Scoring iterations: 2</p> <p style="text-align: center;">p-value for the deviance goodness of fit test: <b>[1] 0</b> The null hypothesis is that our model is correctly specified – REJECT</p> <p style="text-align: center;">&gt; AICcmodavg::AICc(glm13) [1] Inf</p> <p style="text-align: center;">&gt; cv.glm(Modelo,glm13)\$delta [1] 186409447933 143620414011</p>
glm14	<p style="text-align: center;">glm(formula = car ~ arealog, family = gaussian(), data = Modelo)</p> <p style="text-align: center;">Deviance Residuals: 1 2 3 4 -21319 91725 77138 -147544</p> <p style="text-align: center;">Coefficients: Estimate Std. Error t value Pr(&gt; t ) (Intercept) 3.390e+05 1.544e+05 2.195 0.159 arealog 1.881e+00 4.055e+00 0.464 0.688</p> <p style="text-align: center;">(Dispersion parameter for gaussian family taken to be 18293673427)</p>

	<p>Null deviance: 4.0523e+10 on 3 degrees of freedom Residual deviance: <b>3.6587e+10</b> on 2 degrees of freedom <b>AIC: 109.1</b></p> <p>Number of Fisher Scoring iterations: 2</p> <p>p-value for the deviance goodness of fit test: <b>[1] 0</b> The null hypothesis is that our model is correctly specified – REJECT</p> <pre>&gt; AICcmodavg::AICc(glm14) [1] Inf  &gt; cv.glm(Modelo,glm14)\$delta [1] 67389495790 55957816791</pre>
glm15	<pre>glm(formula = car ~ prod + functot, family = gaussian(), data = Modelo)  Deviance Residuals:     1     2     3     4 -40907 10571 26092 4245  Coefficients:             Estimate Std. Error t value Pr(&gt; t ) (Intercept) 427126.08  48461.45   8.814  0.0719 .       prod      3398.88   869.04   3.911  0.1594     functot    -143.25    50.13  -2.857  0.2143 --- Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  (Dispersion parameter for gaussian family taken to be 2483932199)  Null deviance: 4.0523e+10 on 3 degrees of freedom Residual deviance: <b>2.4839e+09</b> on 1 degrees of freedom <b>AIC: 100.34</b></pre> <p>Number of Fisher Scoring iterations: 2</p> <p>p-value for the deviance goodness of fit test: <b>[1] 0</b> The null hypothesis is that our model is correctly specified – REJECT</p> <pre>&gt; AICcmodavg::AICc(glm15) [1] 60.33877  &gt; cv.glm(Modelo,glm15)\$delta [1] 102604720370 77574523327</pre>
glm16	<pre>glm(formula = car ~ functot + funclog, family = gaussian(), data = Modelo)  Deviance Residuals:     1     2     3     4 -19868  3115 12294  4459  Coefficients:             Estimate Std. Error t value Pr(&gt; t ) (Intercept) 356248.18  23586.98  15.104  0.0421 *</pre>



	<pre> functot  -240.87  33.46 -7.199  0.0879 . funclog  1972.36  236.86  8.327  0.0761 . --- Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  (Dispersion parameter for gaussian family taken to be 575459817)  Null deviance: 4.0523e+10  on 3  degrees of freedom Residual deviance: <b>5.7546e+08</b>  on 1  degrees of freedom <b>AIC: 94.489</b>  Number of Fisher Scoring iterations: 2  p-value for the deviance goodness of fit test: <b>[1] 0</b> The null hypothesis is that our model is correctly specified – REJECT  &gt; AICcmodavg::AICc(glm16) [1] 54.48905  &gt; cv.glm(Modelo,glm16)\$delta [1] 13452838358 10233493723 </pre>
glm17	<pre> glm(formula = car ~ functot + funclog + areatot, family = gaussian(),      data = Modelo)  Deviance Residuals: [1] 0 0 0 0  Coefficients:       Estimate Std. Error t value Pr(&gt; t ) (Intercept)  3.739e+05      NA      NA      NA functot      -2.078e+02      NA      NA      NA funclog       8.859e+02      NA      NA      NA areatot       6.108e-01      NA      NA      NA  (Dispersion parameter for gaussian family taken to be NaN)  Null deviance: 4.0523e+10  on 3  degrees of freedom Residual deviance: <b>4.2352e-21</b>  on 0  degrees of freedom <b>AIC: -171.84</b>  Number of Fisher Scoring iterations: 1  p-value for the deviance goodness of fit test: <b>[1] 0</b> The null hypothesis is that our model is correctly specified – REJECT  &gt; AICcmodavg::AICc(glm17) [1] -201.8371  &gt; cv.glm(Modelo,glm17)\$delta [1] 13452838358 10089628769 </pre>
glm18	<pre> glm(formula = car ~ areatot + arealog, family = gaussian(), data = Modelo)  Deviance Residuals: </pre>

```
1 2 3 4
72163 18172 11643 -101978
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.025e+05 1.553e+05 2.592 0.234
areatot 9.602e-01 8.499e-01 1.130 0.461
arealog -6.028e+00 7.966e+00 -0.757 0.588
```

(Dispersion parameter for gaussian family taken to be 16072739314)

Null deviance: 4.0523e+10 on 3 degrees of freedom  
Residual deviance: **1.6073e+10** on 1 degrees of freedom  
**AIC: 107.81**

Number of Fisher Scoring iterations: 2

p-value for the deviance goodness of fit test: **[1] 0**  
The null hypothesis is that our model is correctly specified – REJECT

```
> AICcmodavg::AICc(glm18)
[1] 67.80789
```

```
> cv.glm(Modelo,glm18)$delta
[1] 690583382255 521955721519
```