



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TRANSPORTES
CURSO DE ENGENHARIA CIVIL**

FELIPE MACENA ALVES

**ANÁLISE DA GERAÇÃO DE NOTAS FISCAIS POR BAIRRO A PARTIR DE DADOS
DA RAIS E DA SEFAZ**

**FORTALEZA
2017**

FELIPE MACENA ALVES

ANÁLISE DA GERAÇÃO DE NOTAS FISCAIS POR BAIRRO A PARTIR DE DADOS DA
RAIS E DA SEFAZ

Monografia apresentada ao Curso de Engenharia Civil da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia Civil.

Orientador: Prof. Dr. Bruno Vieira Bertoncini

FORTALEZA

2017

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

A479a Alves, Felipe Macena.
Análise da geração de notas fiscais por bairro a partir de dados da RAIS e da SEFAZ / Felipe Macena
Alves. – 2017.
90 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia,
Curso de Engenharia Civil, Fortaleza, 2017.
Orientação: Prof. Dr. Bruno Vieira Bertoncini.

1. Notas Fiscais. 2. Regressão. 3. TUC. I. Título.

CDD 620

FELIPE MACENA ALVES

ANÁLISE DA GERAÇÃO DE NOTAS FISCAIS POR BAIRRO A PARTIR DE DADOS DA
RAIS E DA SEFAZ

Monografia apresentada ao Curso de Engenharia Civil da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Engenharia Civil.

Aprovada em: ___/___/_____.

BANCA EXAMINADORA

Prof. Dr. Bruno Vieira Bertoncini (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Francisco Moraes de Oliveira Neto
Universidade Federal do Ceará (UFC)

Eng. Msc. Francisco Suliano Mesquita Paula
Prismma Engenharia

A todos que me apoiaram.

AGRADECIMENTOS

Aos meus pais, pela paciência, carinho, exemplo e amor.

À Amanda, pelo companheirismo, carinho e amor.

Aos meus amigos, por estarem ao meu lado.

Ao professor Bruno Vieira Bertoncini, pela excelente orientação.

Ao professor Francisco Moraes de Oliveira Neto e ao engenheiro Francisco Suliano Mesquita Paula, pela presença na banca examinadora e pelas valiosas colaborações.

RESUMO

Este trabalho tem o objetivo de analisar a geração de notas fiscais na cidade de Fortaleza a partir de dados provenientes da SEFAZ-CE e da RAIS. Para isso, primeiramente, faz-se uma contextualização. Apresenta-se a importância e as dificuldades da obtenção de dados e tratamento de dados para o estudo de fenômenos ligados ao planejamento urbano. São apresentados alguns conceitos sobre o transporte urbano de carga, com ênfase na geração de viagens de carga e sua ligação com documentos fiscais. Além disso, são expostos conceitos sobre modelos empíricos e regressão linear. Com os dados obtidos e organizados, foi realizada uma análise exploratória para entender como estava ocorrendo a distribuição de empresas, empregos e emissão e atração de notas fiscais pela cidade de Fortaleza e foram desenvolvidos uma série de modelos de regressão linear para descrever a geração de notas fiscais para os CNAEs que mais emitiram e para os que mais atraíram notas fiscais, o processo é detalhado no método. Posteriormente, são apresentados os resultados para a análise exploratória e para os modelos de regressão linear desenvolvidos e, finalmente, são apresentadas as conclusões do trabalho e as sugestões para trabalhos futuros.

Palavras-chave: Notas Fiscais. Regressão. TUC.

ABSTRACT

This work aims to analyze the generation of invoices in the city of Fortaleza based on data from SEFAZ-CE and RAIS. For this, first, a contextualization is made. The importance and difficulties of obtaining data and processing data for the study of phenomena related to urban planning are presented. Some concepts about urban freight transport are presented, with emphasis on the generation of freight trips and their connection with fiscal documents. In addition, concepts about empirical models and linear regression are presented. With the data obtained and organized, an exploratory analysis was carried out to understand how the distribution of companies, jobs and the issuance and attraction of invoices by the city of Fortaleza were carried out and linear regression models were developed to describe the generation of invoices for the CNAEs that issued the most invoices and for those that attracted the most invoices, the process is detailed in the method. Subsequently, the results are presented for the exploratory analysis and for the linear regression models developed, and finally, the conclusions of the work and the suggestions for future work are presented.

Keywords: Invoices. Regression. Urban Freight Transport.

LISTA DE FIGURAS

Figura 1 - Estrutura para estudo do movimento de carga urbana	19
Figura 2 - Representação do processo de movimentação de carga	21
Figura 3 - Representação do processo de movimentação de carga	23
Figura 4 - Representação do processo de movimentação de carga	27
Figura 5 - Etapas do método.....	29
Figura 6 - Recorte da base de estabelecimentos do 2015	30
Figura 7 - Recorte da base de notas fiscais do 2015	32
Figura 8 - Faixa de CEP da cidade de Fortaleza.....	33
Figura 9 - Participação dos 20 CNAEs que mais empregam e dos 20 CNAEs com mais empresas no número de empregados e no número de empresas, respectivamente	48
Figura 10 - Participação dos 20 bairros que mais empregam e dos 20 bairros com mais empresas no número de empregados e no número de empresas, respectivamente	51
Figura 11 - Mapa da produção de notas fiscais do CNAE 46346 para dezembro de 2014	80
Figura 12 - Mapa da produção de notas fiscais do CNAE 46346 para dezembro de 2015	80
Figura 13 - Mapa da produção de notas fiscais do CNAE 46346 para o ano de 2014	82
Figura 14 - Mapa da produção de notas fiscais do CNAE 46346 para o ano de 2015	82

LISTA DE GRÁFICOS

Gráfico 1 - Participação dos 20 CNAEs que mais emitiram e dos 20 CNAEs que mais atraíram notas fiscais	40
Gráfico 2 - Produção mensal de notas fiscais em 2015 para os 20 CNAEs que mais emitiram notas fiscais.....	40
Gráfico 3 - Atração mensal de notas fiscais em 2015 para os 20 CNAEs que mais receberam notas fiscais.....	42
Gráfico 4 - Porcentagem de notas emitidas e atraídas pelos 20 bairros que mais emitiram e pelos 20 bairros que mais receberam notas fiscais	45
Gráfico 5 - Gráfico dos pontos observados e estimados para o modelo da Tabela 39	79
Gráfico 6 - Gráfico dos pontos observados e estimados para o modelo da Tabela 40	81

LISTA DE TABELAS

Tabela 1 - Características do transporte de passageiros e do transporte de carga	22
Tabela 2 - Descrição das variáveis disponibilizadas na planilha de estabelecimentos da RAIS para o ano de 2015.....	31
Tabela 3 - Descrição dos dados sobre notas fiscais disponibilizados pela SEFAZ	32
Tabela 4 - CNAEs que mais emitiram notas fiscais de 2011 a 2015	38
Tabela 5 - CNAEs que mais atraíram notas fiscais de 2011 a 2015	39
Tabela 6 - Média e desvio padrão da produção mensal de notas fiscais dos 20 CNAEs que mais emitiram notas em 2015.....	41
Tabela 7 - Média e desvio padrão da atração mensal de notas fiscais dos 20 CNAEs que mais receberam notas em 2015	42
Tabela 8 - Bairros que mais emitiram notas fiscais de 2011 a 2015.....	43
Tabela 9 - Bairros que mais atraíram notas fiscais de 2011 a 2015.....	44
Tabela 10 - 20 CNAEs que tinham mais empregados ativos em 31 dezembro de 2014 e 2015	46
Tabela 11 - 20 CNAEs com mais empresas ativas em 2014 e 2015.....	47
Tabela 12 - 20 Bairros que mais empregaram em dezembro de 2014 e 2015	49
Tabela 13 - 20 bairros com mais empresas ativas em 2014 e 2015.....	50
Tabela 14 - Modelos Desenvolvidos	52
Tabela 15 – Resultados dos modelos tipo 1	53
Tabela 16 - Resultados dos modelos tipo 2	54
Tabela 17 - Resultados dos modelos tipo 3	55
Tabela 18 - Resultados dos modelos tipo 4	56
Tabela 19 - Testes de normalidade para os modelos tipo 1	58
Tabela 20 - Testes de normalidade para os modelos tipo 2	59
Tabela 21 - Testes de normalidade para os modelos tipo 3	60
Tabela 22 - Testes de normalidade para os modelos tipo 4	61
Tabela 23 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 1	62
Tabela 24 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 2	63
Tabela 25 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 3	64
Tabela 26 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 4	65
Tabela 27 - Resultados dos modelos tipo 5	66
Tabela 28 - Resultados dos modelos tipo 6	67

Tabela 29 - Resultados dos modelos tipo 7	69
Tabela 30 - Resultados dos modelos tipo 8	70
Tabela 31 - Testes de normalidade para os modelos tipo 5	71
Tabela 32 - Testes de normalidade para os modelos tipo 6	72
Tabela 33 - Testes de normalidade para os modelos tipo 7	73
Tabela 34 - Testes de normalidade para os modelos tipo 8	74
Tabela 35 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 5	75
Tabela 36 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 6	76
Tabela 37 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 7	77
Tabela 38 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 8	78
Tabela 39 - Modelo tipo 1 para o CNAE 46346.....	79
Tabela 40 - Modelo tipo 5 para o CNAE 46346.....	81

LISTA DE ABREVIATURAS E SIGLAS

RAIS	Relação Anual de Informações Sociais
SEFAZ-CE	Secretária da Fazenda do Estado do Ceará
CNAE	Classificação Nacional de Atividades Econômicas
TUC	Transporte Urbano de Carga

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Contextualização	14
1.2	Objetivos	16
<i>1.2.1</i>	<i>Objetivo geral</i>	16
<i>1.2.2</i>	<i>Objetivos específicos</i>	16
1.3	Estrutura do trabalho	16
2	GRANDE BASES DE DADOS NO PLANEJAMENTO DOS TRANSPORTES	17
3	TRANSPORTE URBANO DE CARGA	19
3.1	Geração de viagens de carga e notas fiscais	22
4	MODELOS EMPÍRICOS E REGRESSÃO LINEAR SIMPLES	26
4.1	Testes de hipóteses	27
5	MÉTODO	29
5.1	Obtenção e tratamento dos dados e análise exploratória	29
<i>5.1.1</i>	<i>Dados da RAIS</i>	29
<i>5.1.2</i>	<i>Dados da SEFAZ-CE</i>	31
<i>5.1.3</i>	<i>Consolidação dos dados</i>	32
<i>5.1.4</i>	<i>Análise exploratória</i>	33
5.2	Seleção das variáveis e desenvolvimento dos modelos de geração de notas fiscais	34
6	RESULTADOS E DISCUSSÕES	37
6.1	Análise exploratória	37
6.2	Análise de regressão linear	51
<i>6.2.1</i>	<i>Notas fiscais de dezembro de 2014 e 2015</i>	52
<i>6.2.1.1</i>	<i>Modelos</i>	52
<i>6.2.1.2</i>	<i>Normalidade dos erros</i>	57
<i>6.2.1.3</i>	<i>Homoscedasticidade e autocorrelação</i>	61
<i>6.2.2</i>	<i>Notas fiscais dos anos de 2014 e 2015</i>	65
<i>6.2.2.1</i>	<i>Modelos</i>	66
<i>6.2.2.2</i>	<i>Normalidade dos erros</i>	71
<i>6.2.2.3</i>	<i>Homoscedasticidade e autocorrelação</i>	74

6.2.3	<i>Avaliação dos resultados</i>	78
7	CONCLUSÕES	84
7.1	Sugestões para trabalhos futuros	85
	REFERÊNCIAS	86
	APÊNDICE A – CNAES E DENOMINAÇÕES	89

1 INTRODUÇÃO

1.1 Contextualização

O foco do planejamento de transporte urbano têm sido, normalmente, o transporte de passageiros. Por isso, o termo “transporte urbano” é, raramente, associado ao transporte de mercadorias. Entretanto, faz-se necessário entender e reconhecer o papel de importância do transporte urbano de carga (TUC), visto que ele é essencial para o funcionamento das cidades e da economia e possui inúmeras interfaces com o transporte de pessoas (CARVALHO, 1998).

A grande maioria das atividades econômicas e sociais que ocorrem nos centros urbanos tem como fator crucial o transporte de mercadorias. Para empresas situadas em centros urbanos, a distribuição de produtos forma uma ligação vital entre vendedores e consumidores. Atualmente, a quantidade de veículos transportadores de mercadorias cresce e espera-se que continue crescendo. Isso se dá pelas atuais práticas de produção e distribuição, que são baseadas em pequenos estoques e entregas pontuais (BENJELLOUN; CRAINIC, 2009).

Para Lindholm e Behrends (2012), ao mesmo tempo que o tráfego de veículos de carga representa uma ameaça crescente para a sustentabilidade urbana, é cada vez mais importante para a competitividade regional. Por outro lado, o contexto urbano é, também, um empecilho para a eficiência das operações de distribuição. Apesar disso, autoridades locais e operadores de tráfego e de logística negligenciam esse problema.

Faz-se necessário a consideração do TUC no planejamento urbano, em especial compreender como se dá os deslocamentos das mercadorias. Entretanto, por se tratar de atividade comercial, o acesso à informação não é direto, dado que cada empresa possui sua própria estratégia, muitas vezes tratada como segredo. Além disso, cada segmento econômico tem uma forma de distribuir seus produtos.

O TUC, costumeiramente realizado por veículos rodoviários, é essencial para o funcionamento de áreas metropolitanas, entretanto gera muitos problemas para as cidades, como congestionamentos, poluição do ar e ruídos (DABLANC *et al.*, 2013). Segundo Gasparini (2008), os veículos utilizados na distribuição de mercadorias agravam os problemas de tráfego, entretanto o próprio setor de transporte de carga urbana também é afetado pelo tráfego urbano. Entre os principais problemas enfrentados pelos transportadores estão: aumento no tempo de entrega, dificuldade de encontrar vagas para realizar operações de carga e descarga e aumento do consumo de combustível. Associado a isto, tem-se o fato de que

muitas das atividades comerciais estão concentradas em regiões específicas das cidades, ocasionando uma distribuição desequilibrada das viagens e comprometendo o uso do sistema de transporte.

Apesar de haver uma crescente preocupação em resolver os problemas do transporte urbano de mercadorias, a literatura relevante é limitada e fragmentada. Além disso, um dos principais empecilhos para o desenvolvimento de pesquisas é a falta de dados (DABLANC *et al.*, 2013).

É necessário a disponibilização de informações que possibilitem a elaboração de estudos e análises que possam contribuir com o planejamento do transporte de mercadorias dentro de centros urbanos. Como conseguir dados coletados diretamente para esse fim é muito complicado, utiliza-se dados secundários, que teoricamente são coletados de forma mais simples, na modelagem do transporte urbano de carga. Um tipo de dado secundário que pode ser utilizado e já foi explorado em trabalhos científicos, como nos trabalhos de Herculano (2015), Santos (2015) e Brandão Filho (2017), são os dados de documentos fiscais.

Este trabalho faz uma análise exploratória da distribuição de empregos, empresas e notas fiscais por bairro da cidade de Fortaleza e por CNAE (Classificação Nacional de Atividades Econômicas), para compreender como os segmentos econômicos e os bairros da cidade estão contribuindo para a atividade econômica da cidade e, conseqüentemente, para o fluxo de veículos de carga como um todo.

É importante salientar que, mesmo para dados secundários, a obtenção de dados pode não ser tão simples. Informações sobre notas fiscais, por exemplo, não são tão facilmente obtidas, por questões de sigilo fiscal e outras burocracias. Como notas fiscais carregam informações muito relacionadas com o fenômeno do transporte urbano de carga e por já terem sido utilizadas com sucesso em trabalhos que buscaram compreender esse fenômeno, o presente estudo busca manipular dados brutos, provenientes da SEFAZ-CE (Secretária da Fazenda do Estado do Ceará) e da RAIS (Relação Anual de Informações Sociais), para analisar e modelar a geração de notas fiscais na cidade de Fortaleza para que estimativas possam ser feitas a partir de variáveis de fácil obtenção e, conseqüentemente, para que os planejadores e pesquisadores não necessitem despender tantos recursos para conseguir utilizar tais dados em trabalhos que façam uso de modelos previamente desenvolvidos ou que busquem novas abordagens para modelar o transporte urbano de carga.

1.2 Objetivos

1.2.1 Objetivo geral

O objetivo geral do presente trabalho é analisar a geração de notas fiscais em função da distribuição de empregos e empresas pelos bairros de Fortaleza, para os CNAEs que mais contribuíram para o fenômeno, utilizando dados provenientes da RAIS e da SEFAZ-CE. E, dessa forma, contribuir para a compreensão dos fenômenos relacionados com a distribuição de mercadorias em centros urbanos.

1.2.2 Objetivos específicos

- (a) Verificar a possibilidade de compatibilização das bases da RAIS e da SEFAZ-CE.
- (b) Realizar análise exploratória dos dados provenientes das bases de dados.
- (c) Selecionar variáveis, que sejam mais adequadas para explicar a emissão e a atração de notas fiscais, das bases utilizadas.
- (d) Desenvolver modelos de geração de notas fiscais, para os CNAEs que mais contribuíram para a geração de notas fiscais na cidade de Fortaleza, utilizando regressão linear simples.

1.3 Estrutura do trabalho

Além do capítulo introdutório, este trabalho contém outros 6 capítulos. O capítulo 2 apresenta um breve panorama sobre a manipulação e o uso de grandes bases de dados no planejamento urbano, especialmente no transporte urbano de carga. O capítulo 3 é dedicado a uma apresentação geral do que é o transporte urbano de carga, do que é geração de viagens de carga e sua relação com as notas fiscais. O capítulo 4 apresenta conceitos gerais sobre modelos de regressão linear e conceitos de estatística necessários para o desenvolvimento do trabalho. O capítulo 5 contém o método utilizado para realizar o trabalho. O capítulo 6 apresenta algumas das considerações feitas sobre os dados, os resultados obtidos na análise exploratória realizada e os resultados dos modelos de regressão linear simples desenvolvidos. E, finalmente, o capítulo 7 apresenta as conclusões resultantes do desenvolvimento do presente trabalho e as sugestões para próximos trabalhos.

2 GRANDES BASES DE DADOS NO PLANEJAMENTO DOS TRANSPORTES

Com o crescimento no número de veículos de carga trafegando dentro dos grandes centros urbanos, há a necessidade de desenvolvimento de estudos e ferramentas que permitam entender melhor como esse sistema funciona, quais suas consequências e como mitigá-las. Uma das etapas necessárias para que seja possível o melhor entendimento do fenômeno do transporte de carga urbana é a aquisição, tratamento e análise de dados que possam embasar o trabalho dos planejadores.

Nas últimas duas décadas, planejadores urbanos têm representado a evolução das cidades durante uma era onde a informação e as tecnologias de comunicação exercem uma influência crescente e generalizada sobre a natureza, a infraestrutura urbana, a administração, a atividade econômica e a vida cotidiana (KITCHIN, 2014). Dessa forma, é necessário que essas informações sejam armazenadas e tratadas de forma adequada para posteriores análises e, conseqüentemente, para que embasem futuras tomadas de decisão.

Como em diversas áreas, o uso de grandes bases de dados pode ser bastante útil para o planejador urbano. Para Batty (2013), a atual quantidade de dados está, certamente, enriquecendo o entendimento de como as cidades funcionam e está oferecendo muitas novas oportunidades de interação social e tomadas de decisão mais embasadas sobre o que realmente ocorre nos centros urbanos.

Segundo Labrinidis e Jagadish (2012), existem dois passos cruciais no processo de análise de dados. O primeiro é a aquisição de dados. Algumas fontes de dados podem produzir enormes quantidades de informações, que podem ser filtradas para retirar o que não é de interesse, o desafio está em definir os filtros de tal forma que eles não retirem informações úteis. O segundo é um processo de extração de informação que obtenha os dados necessários das fontes e expresse-os de forma adequada para análise.

Quando se trata de dados, os analistas de transporte de carga têm, normalmente, duas opções: desenvolver métodos de estimativas baseados em dados disponíveis, mesmo que secundários, ou coletar os dados necessários, seja diretamente de transportadores de carga ou através de fornecedores de dados (GIULIANO *et al.*, 2010).

Existe grande reconhecimento do valor inerente a dados e os produtos obtidos a partir de sua análise (LABRINIDIS; JAGADISH, 2012). Entretanto, um dos grandes desafios para os analistas e planejadores de transporte vem sendo, justamente, a coleta, o tratamento e a disponibilização de dados para embasar as tomadas de decisão. Geralmente, a obtenção de dados é o maior empecilho para o prosseguimento de trabalhos sobre planejamento de

sistemas de transporte de carga, pois as coletas em campo, na maioria das vezes, gastam muito tempo, dinheiro e outros recursos. Além disso, empresas privadas, muitas vezes, não estão dispostas a fornecer seus dados e os fluxos de mercadorias mudam com o tempo, fazendo com que as coletas devam ser realizadas periodicamente.

Os governos e empresas vêm produzindo gigantescas bases de dados que contêm inúmeras informações sobre cidades, empresas e pessoas. Apesar da importância destes dados, ainda é muito complicado conseguir que esses agentes os disponibilizem. Muitos os mantêm de forma confidencial ou inacessível, inclusive os governos. Além disso, muitas delas carregam inconsistências, devido a problemas na coleta dos dados ou por dificuldades de organizar uma quantidade muito grande de informação.

Muitas informações úteis podem ser extraídas de forma indireta, a partir da utilização de dados secundários, como costuma-se chamar. Os dados secundários são dados que não foram coletados com o propósito de embasar um certo tipo de trabalho, mas podem fornecer valiosas informações para tal.

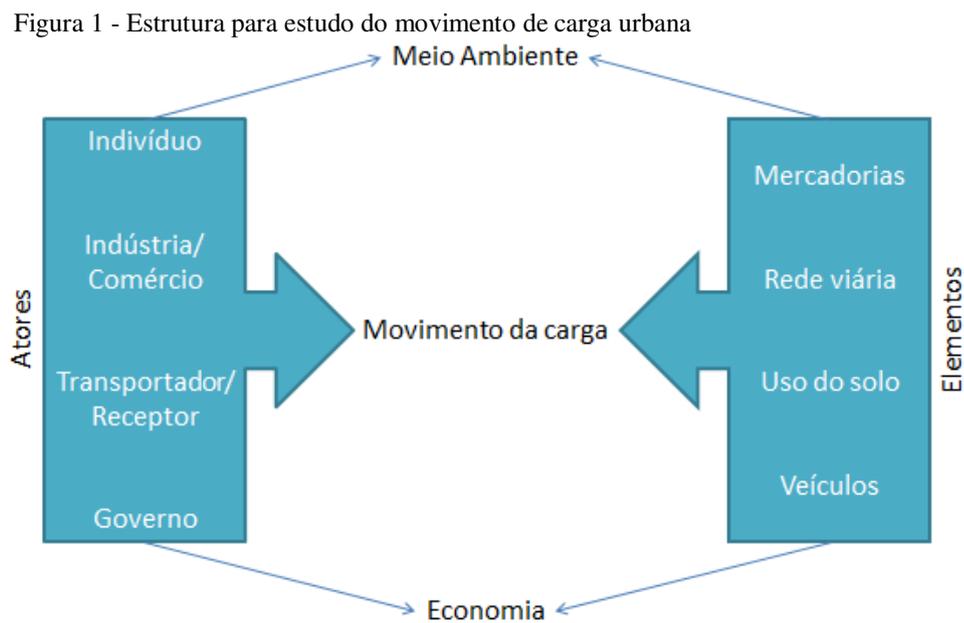
Utilizar dados secundários tem algumas desvantagens significativas. Como as informações foram desenvolvidas para outros usos, dificilmente uma única base de dados atenderá as necessidades de um trabalho. Utilizar diferentes fontes de dados requer consistência entre as fontes, entretanto é comum que diferentes bases de dados possuam diferentes sistemas de classificação, de agregação e, até mesmo, sistemas unitários. Geralmente, muito esforço é necessário para compatibilizar as diferentes fontes de informações usadas (GIULIANO *et al.*, 2010).

É comum que dados brutos sejam organizados de forma que inviabilizem ou tornem muito complicada a sua utilização para a finalidade desejada. Devido a isso, frequentemente é necessário manipular as bases de dados. Quando manipula-se grandes bases de dados tem-se o objetivo de trabalhar os dados para organizá-los de determinada forma que seja mais fácil e eficiente utilizá-los para certa atividade.

Na manipulação de dados pode-se, por exemplo, adicionar, remover ou alterar informações. Para trabalhar com dados existem diversas ferramentas disponíveis, com diferentes graus de facilidade e eficiência, pode-se utilizar desde softwares como Microsoft Excel e LibreOffice Calc até linguagens de programação como R e Python.

3 TRANSPORTE URBANO DE CARGA

O transporte urbano de carga pode ser definido como a movimentação de objetos dentro da área urbana (CHRISTIANSEN, 1979 *apud* CARVALHO, 1998). Segundo Herculano (2015), os sistemas de transporte são cada vez mais discutidos devido ao aumento do tráfego de veículos nos centros urbanos, atingindo, dessa forma, o cotidiano da população. A Figura 1 mostra a estrutura do sistema de transporte de carga urbana.



Fonte: Adaptado de WOUDSMA (2001)

Com o processo de urbanização no mundo, cerca de metade da população, em 2010, estava concentrada em áreas urbanas, de acordo com estatísticas das Nações Unidas, e prevê-se que esse número aumentará para 70% até 2050 (TANIGUCHI; THOMPSON; YAMADA, 2014). Devido ao aumento da complexidade dos sistemas econômico e comercial e ao crescimento populacional urbano, é notório o aumento do número de veículos trafegando dentro dos centros urbanos. Esse aumento, tanto do número de veículos de transporte de passageiros quanto do número de veículos de transporte de carga, é problemático, pois além dos problemas gerados naturalmente pelo grande número de veículos, como congestionamentos, o transporte de pessoas e de carga são, na maioria das vezes, conflitantes, pois competem por espaço, vagas de estacionamentos e possuem características muito distintas.

Nuzzolo, Crisalli e Comi (2012) afirmam que o papel de importância do transporte urbano de carga como parte das atividades cotidianas para empresas e pessoas

ainda cresce, especialmente se forem analisadas as recentes tendências de comércio eletrônico, globalização, armazéns inteligentes e o sistema de produção “just-in-time”.

A eficiência da distribuição de carga tem um papel de grande importância na competitividade, em termos de renda e emprego, de áreas urbanas. Ao mesmo tempo, o aumento da quantidade de veículos transportadores de carga em áreas urbanas também contribui de forma negativa para os centros urbanos, como apresentado. (STATHOPOULOS; VALERI; MARCUCCI, 2012). Ademais, o aumento do fluxo de veículos e a combinação de diferentes tipos de veículos também provocam o aumento do risco de acidentes.

Infelizmente, a maioria dos esforços para solucionar problemas de transporte têm o principal foco somente no transporte de passageiros, dando pouca atenção ou nenhuma atenção à carga. Essa atitude não é consistente com a atual importância do transporte de carga (BASTIDA; HOLGUÍN-VERAS, 2009). Segundo Lindholm (2010), o planejamento urbano, as visões políticas e o planejamento da infraestrutura são fatores importantes que afetam a logística urbana, entretanto a visão de profissionais e estudiosos da logística são raramente usadas no processo de planejamento.

Um dos maiores desafios para as cidades atingirem a sustentabilidade é o transporte urbano de mercadorias. O transporte de carga é uma indústria que envolve, primariamente, empresas e transportadores e os esforços para entender seus impactos são, constantemente, limitados ao planejamento do tráfego. Para atingir a sustentabilidade no TUC é necessário ir além do planejamento estratégico das cidades, pois os impactos, que se tornam visíveis no nível do tráfego, são apenas o elemento final da cadeia (BEHREND; LINDHOLM; WOXENIUS, 2008).

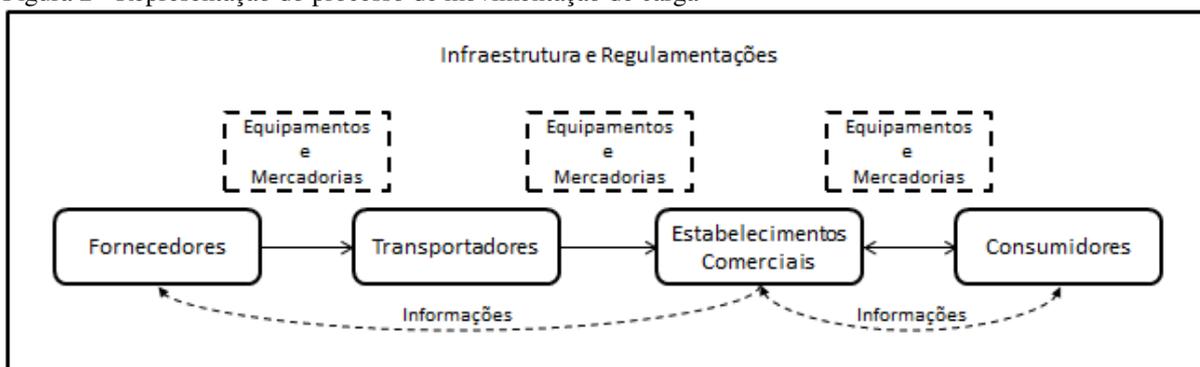
A incorporação efetiva do transporte de mercadorias no planejamento dos sistemas de transporte urbano é essencial, pois o sistema de distribuição de produtos tem papel crucial para a economia, a qualidade de vida da população e as mudanças climáticas. O transporte de mercadorias é importante tanto por sua contribuição positiva quanto negativa para a vida moderna.

Para embasar o planejamento da logística urbana é importante que existam esforços para compreender o fenômeno e como ele está evoluindo. Além disso, é importante entender como fenômenos que impactam diretamente a distribuição de mercadorias, como compra e venda de produtos, estão evoluindo. Dessa forma, será possível que os planejadores insiram o transporte urbano de carga em suas tomadas de decisão e o façam de forma embasada.

Devido aos impactos do transporte de mercadorias no ambiente urbano, muitos governantes têm consciência que devem controlar as atividades de distribuição de produtos, mas a maioria não sabe como (DABLANC, 2007). É importante tanto para os planejadores urbanos quanto para os transportadores que o planejamento do transporte de carga seja realizado com um embasamento adequado, pois o crescimento desordenado pode provocar uma grande ineficiência nos deslocamentos realizados pelos transportadores. Provocando, dessa forma, perda de dinheiro e tempo para os transportadores e piorando a qualidade de vida da população.

Para Facchini (2006), a interação entre 3 elementos principais gera a movimentação de carga. São eles: os diferentes atores envolvidos, o ambiente e os equipamentos. Sendo os atores: os fornecedores, os transportadores (em muitos casos os fornecedores transportam sua própria carga), os estabelecimentos comerciais e os consumidores finais. O ambiente é constituído pela malha viária e pela regulamentação. Já os equipamentos são, principalmente, os veículos. A Figura 2 representa graficamente o processo mencionado.

Figura 2 - Representação do processo de movimentação de carga



Fonte: Adaptado de FACCHINI (2006)

Czerniak, Lahsene, Chatterjee (2000) concluem que a relação entre o transporte de mercadorias e a economia regional ainda não é totalmente compreendida, apesar disso sabe-se que melhorias na distribuição de mercadorias tem efeitos positivos na economia. Dessa forma, no atual milênio, o planejamento do transporte urbano de carga deve ter um papel mais importante no processo geral de planejamento dos transportes. Além disso, sem pesquisas focadas em problemas do transporte urbano de carga, o potencial do governo e do setor privado de atingir seus objetivos pode nunca se tornar realidade.

3.1 Geração de viagens de carga e notas fiscais

Para Sánchez-Díaz (2016), a falta de conhecimento sobre os sistemas de transporte de carga e sobre o comportamento dos agentes transportadores de carga, assim como a deficiente quantificação do problema, é uma das razões que explicam os insatisfatórios resultados de iniciativas que pretendiam melhorar a eficiência do transporte de carga. Consequentemente, os esforços para coletar dados de qualidade e desenvolver modelos sobre o transporte urbano de carga estão conectados para melhorar o conhecimento sobre os sistemas, para facilitar a formulação de iniciativas adequadas e para suportar a tomada de decisão dos agentes públicos.

Segundo Ferreira (2016), para que o planejamento eficiente do transporte urbano de carga seja implementado é necessário compreender o fluxo de veículos de carga no espaço urbano. Para isso, uma ferramenta bastante útil são os modelos de geração de viagens de carga, pois com eles é possível realizar previsões, entender as variáveis que estão influenciando a produção e a atração de viagens e saber quais medidas serão mais efetivas para determinadas finalidades.

Como a modelagem do transporte de passageiros é mais usual, como defendem diversos autores, a Tabela 1 é útil para entender as diferenças entre os dois tipos de transporte.

Tabela 1 - Características do transporte de passageiros e do transporte de carga

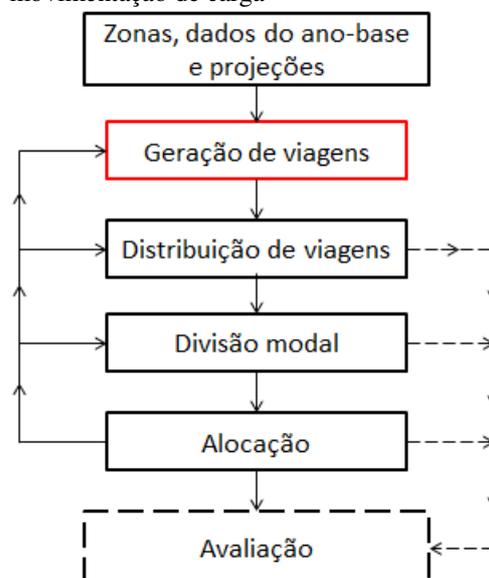
Características	Transporte de Passageiros	Transporte de Carga
Demanda gerada	Viagem de passageiros	Toneladas produzidas ou consumidas em uma região
Tráfego gerado	Viagens de carros, viagens de ônibus, etc.	Viagens de veículos transportadores de carga.
Variáveis influenciadoras	Renda, uso do solo, estrutura familiar, concentração de atividades	Atividade econômica desenvolvida, tipo de empresas, tamanho das empresas, número de empregados, uso do solo
Correspondência entre demanda e tráfego gerados	Muito alta, quase um para um em áreas onde o compartilhamento de veículos é baixo	Muito baixa, pois: grandes empresas podem produzir enormes quantidade de carga, mas não produzir tráfego de forma tão elevado devido ao tamanho dos equipamentos de transporte. Além disso, pequenas empresas podem produzir muito tráfego em relação a carga produzida.

Fonte: Adaptado de HOLGUÍN-VERAS *et al.* (2012)

A geração de viagens é um dos quatro passos da modelagem do sistema de transporte urbano clássica (modelagem em quatro etapas), que são: geração de viagens

(atração e produção), distribuição de viagens, divisão modal e alocação (ORTÚZAR; WILLUMSEN, 2011). Como mostrado na Figura 3.

Figura 3 - Representação do processo de movimentação de carga



Fonte: Adaptado de ORTÚZAR e WILLUMSEN (2011)

A geração de viagens de carga refere-se ao número de veículos de carga necessário para transportar a carga ofertada/demandada entre regiões. Normalmente, a geração de viagens de carga é mensurada em número de viagens de veículos de carga (HOLGUÍN-VERAS *et al.*, 2012).

A geração de viagens tem duas componentes: produção e atração. Produção são as viagens que se iniciam em determinada zona, já a atração é o número de viagens que têm determinada zona como destino. A geração de viagens é a soma das viagens atraídas e produzidas em uma determinada zona (CAMPOS, 2013).

A modelagem da geração de viagens de carga pode fornecer informações muito valiosas e de fácil acesso para os planejadores. Com a geração de viagens de carga pode-se, por exemplo, conhecer quais regiões estão recebendo mais viagens e quais estão originando mais viagens e, dessa forma, entender quais regiões precisam de mais atenção, ou até mesmo onde intervir. A geração de viagens também pode ser muito útil quando pretende-se planejar o uso e ocupação do solo de uma cidade.

Além disso, a modelagem da geração de viagens de carga é importante, pois, como dito, a coleta de dados é geralmente muito dispendiosa. Com modelos que descrevam o

fenômeno por meio de variáveis acessíveis, pode-se fazer previsões e construir cenários com um custo muito baixo.

Segundo D'Este (2007), as técnicas usuais para modelar a geração de viagens de carga são similares às aquelas usadas para modelar a geração de viagens de pessoas e são: extrapolação de tendência histórica e métodos de fator de crescimento, previsões econômicas, técnicas de regressão e taxas de geração de viagens. As técnicas de regressão podem ser aplicadas utilizando dados secundários. Um dos dados que já foram utilizados para isso são dados de notas fiscais, pois, no Brasil, as vendas realizadas por empresas formais devem ser acompanhadas de notas fiscal, fazendo com que esses documentos tenham informações úteis para o entendimento da movimentação de mercadorias.

Segundo Holguín-Veras *et al.* (2012), vários estudos já identificaram diferentes variáveis que explicam a geração de viagens de carga. Dentre as variáveis identificadas estão: uso do solo, atividade econômica, número de empregos, área dos estabelecimentos. Apesar disso, os modelos são desenvolvidos, na maioria das vezes, por regressão linear com apenas uma variável explicativa.

Após definida a área de estudo, recomenda-se, tradicionalmente, dividi-la em zonas, para que seja possível entender as origens e os destinos das viagens de uma forma espacialmente desagregada. Geralmente, procura-se definir as zonas baseado em regiões já definidas pelo IBGE, em bairros ou em outros agrupamentos de regiões que tenham características parecidas (MELLO, 1975 *apud* SANTOS, 2015).

Além disso, Holguín-Veras *et al.* (2012) mostra que estimativas de geração de viagens de carga baseadas em atividade econômica são melhores e são mais fortes estatisticamente do que as baseadas em uso do solo. Além disso, ele conclui que a proporcionalidade entre a geração de viagens de carga e o tamanho dos estabelecimentos apenas ocorre em uma pequena minoria de empresas.

É importante lembrar que o transporte urbano de carga está muito relacionado com o consumo. Espera-se que quanto mais produtos forem consumidos em determinado período de tempo mais viagens de carga serão necessárias para abastecimento (mantendo a capacidade dos veículos constante). Dessa forma, informações relacionadas ao consumo (compra e venda de produtos) podem ser muito valiosas para o entendimento do fenômeno da distribuição de mercadorias em áreas urbanas.

Acredita-se que os dados provenientes de notas fiscais, que refletem a compra e a venda de mercadorias, estão entre as mais valiosas fontes de dados secundários para a modelagem do transporte urbano de carga. Santos (2015), por exemplo, conclui que é possível

a utilização de dados de documentos fiscais como uma alternativa de informações para o planejamento do TUC e que os dados obtidos dessa forma, tratados e organizados, podem modelar os fluxos de mercadorias no espaço urbano sem abrir mão de detalhes e abrangência, além de uma atualização contínua a um custo baixíssimo.

Como os dados de notas fiscais não são tão facilmente obtidos, pode-se modelar o fenômeno da produção e atração de notas fiscais a partir de dados de mais simples obtenção para que com estimativas desse fenômeno os métodos já desenvolvidos por outros autores possam ser utilizados.

Para modelar os fenômenos citados no capítulo, geralmente, são utilizados modelos empíricos e técnicas de regressão linear. Esses conceitos são explicados no próximo capítulo.

4 MODELOS EMPÍRICOS E REGRESSÃO LINEAR SIMPLES

Modelos são uma representação aproximada de sistemas reais e o processo de modelagem é um meio termo entre simplicidade e precisão. Um modelo empírico é desenvolvido por meio de uma análise estatística de observações experimentais. Em ocasiões nas quais não há base teórica para fundamentação sobre a relação entre variáveis de interesse em um sistema, estes modelos são bastante utilizados (FINKLER, 2003).

Uma das formas de desenvolver modelos de previsão é através da regressão linear. Segundo Montgomery e Runger (2009), diversos problemas encontrados por engenheiros estão relacionadas com a exploração das relações entre variáveis. Análise de regressão é uma técnica comumente utilizada para modelar e investigar a relação entre variáveis. A regressão pode ser utilizada tanto para previsões quanto para otimização de processos.

Um modelo de regressão linear é uma equação composta por dois tipos de variáveis: independentes e dependente. As variáveis independentes são as variáveis de entrada do modelo, caso haja apenas uma, chama-se regressão linear simples, do contrário, chama-se regressão linear múltipla. A variável dependente é a variável que será explicada pelas demais, é a variável que se deseja estimar. A diferença entre o valor observado na realidade e o valor previsto por um modelo de regressão é conhecido como erro (ε). A equação geral para uma regressão linear simples está representada na equação 1.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (1)$$

Onde y_i são os valores observados da variável dependente, x_i as variáveis independentes e ε_i os erros residuais. β_0 é o intercepto e β_1 é a inclinação que relaciona uma variável x e a variável y .

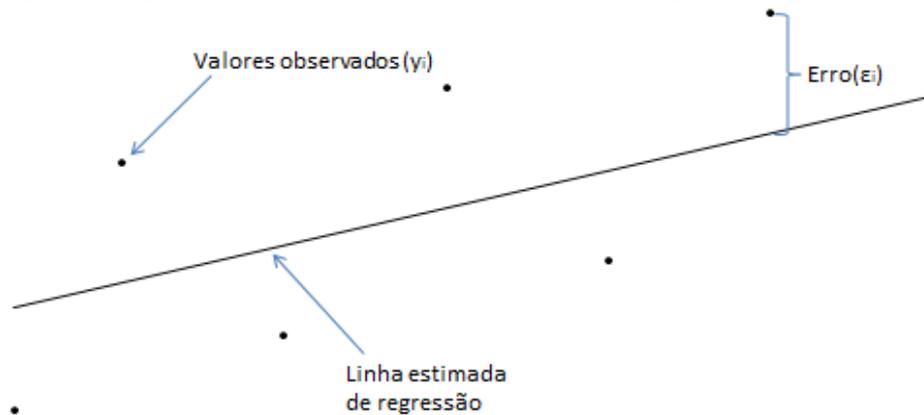
A equação 2 representa a linha estimada pela regressão e a equação 3 expressa o erro residual para cada observação.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x \quad (2)$$

$$\hat{\varepsilon}_i = y_i - \hat{y}_i \quad (3)$$

A Figura 4 representa os desvios dos dados observados em relação ao modelo de regressão.

Figura 4 - Representação do processo de movimentação de carga



Fonte: Adaptado de MONTGOMERY e RUNGER (2009)

Para determinar os coeficientes β_i , geralmente, utiliza-se o método dos mínimos quadrados, em outras palavras procura-se os coeficientes que minimizem a equação 4.

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

O R^2 é o coeficiente utilizado, comumente, para avaliar a adequação da reta de regressão. Ele varia de 0 a 1, quanto mais próximo de 1 melhor o ajuste. Ou seja, quando R^2 é igual a 1 significa que a reta se ajusta perfeitamente às observações, logo para cada x_i o valor da reta será igual ao valor observado. A equação 7 descreve o R^2 .

$$R^2 = 1 - \frac{\sum \hat{\varepsilon}_i^2}{\sum (y_i - \bar{y})^2} \quad (7)$$

É importante lembrar que modelos que têm elevados R^2 não necessariamente fornecerão previsões exatas para observações futuras e que as principais suposições tomadas para a regressão linear é que os erros sejam não correlacionados, tenham distribuição normal e variância constante (homoscedasticidade).

4.1 Testes de hipóteses

Ao realizar uma regressão linear é importante verificar se as variáveis são realmente relevantes para o modelo, para isso uma das ferramentas utilizadas é o teste t . Esse

teste está relacionado com a significância de cada uma das variáveis independentes dentro do modelo. As hipóteses do teste t estão apresentadas na equação 5.

$$\begin{cases} H_0: \beta_i = 0 \\ H_1: \beta_i \neq 0 \end{cases} \quad (5)$$

Se a hipótese nula não for rejeitada, pode-se concluir que não há uma forte relação linear entre a variável independente e a dependente. Isso mostra que a variável x_i tem pouco poder explicativo sobre a variável que se deseja modelar ou que a real relação entre essas variáveis não seja linear. Se a hipótese nula for rejeitada, existem indícios que a variável x_i é importante para explicar a variável dependente e que o modelo linear pode ser adequado, mas não garante que melhores resultados não possam ser alcançados com modelos não lineares.

Outra forma de avaliar a significância da regressão é através da análise de variância, que pode ser realizada através do teste F. Esse teste está relacionado com a significância do modelo como um todo. Da mesma forma do teste t , existem duas hipóteses para esse teste, que estão representadas na equação 6.

$$\begin{cases} H_0: \beta_1 = \dots = \beta_k = 0 \\ H_1: \text{Ao menos um } \beta_1 \neq 0 \end{cases} \quad (6)$$

Se a hipótese nula for rejeitada, existem indícios de que o conjunto de variáveis explicativas é adequado para explicar a variável dependente.

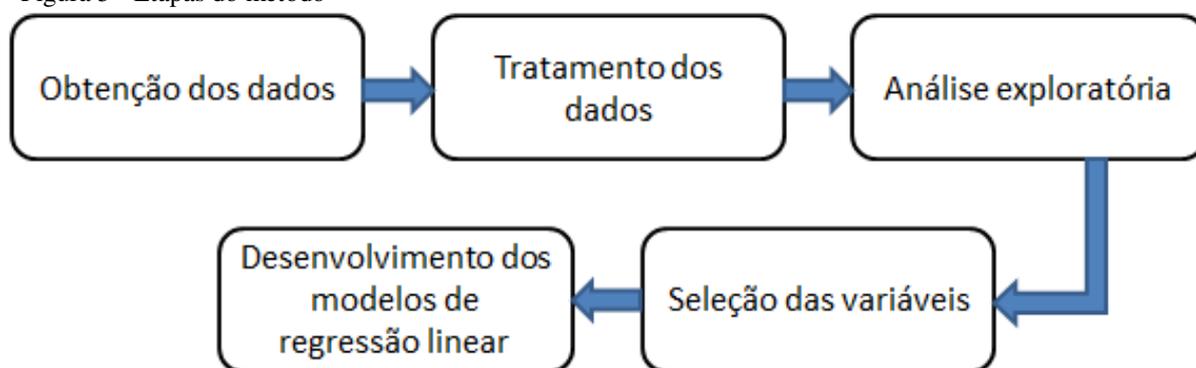
Para rejeitar ou não a hipótese nula, em ambos os testes, é importante definir o nível de significância desejado (α). Esse nível estabelece a probabilidade de erro na rejeição da hipótese nula, por exemplo rejeitar a hipótese nula quando na verdade ela não deveria ser rejeitada. Logo, quanto menor o α mais confiável é a análise. Para os dois testes a hipótese nula é rejeitada se o p-valor for menor do que α .

5 MÉTODO

A sequência de etapas metodológicas que foram estabelecidas para o presente trabalho estão representadas na forma de fluxograma, apresentado na Figura 5. Todos os passos metodológicos foram realizados para obter resultados referentes a cidade de Fortaleza-CE.

Por serem fenômenos diretamente ligados, no trabalho, os mesmos termos usados na geração de viagens de carga serão utilizados para a geração de notas fiscais. Por exemplo, produção de notas fiscais são as notas que foram originadas em determinado bairro e atração de notas fiscais são as notas que tiveram como destino certo bairro.

Figura 5 - Etapas do método



Fonte: Elaborado pelo autor

5.1 Obtenção e tratamento dos dados e análise exploratória

Inicialmente, pensou-se em utilizar três bases de dados, duas da RAIS (Relação Anual de Informações Sociais – Ministério do Trabalho) e uma da SEFAZ-CE (Secretaria da Fazenda do Estado do Ceará). Essas bases de dados foram usadas pois carregavam informações úteis para a modelagem da geração de notas fiscais e para o entendimento do transporte urbano de carga.

5.1.1 Dados da RAIS

A RAIS é um conjunto de dados organizados e distribuídos pelo Ministério do Trabalho. As bases de dados da RAIS foram coletadas do site do ministério e são livres para a consulta de qualquer pessoa.

Especificamente, as duas bases da RAIS que foram cogitadas para o trabalho foram as bases de estabelecimentos e a base de vínculos. A primeira conta com informações anuais sobre as empresas, como número de funcionários ativos ao final do ano, CEP dos estabelecimentos, CNAE da empresa, entre outras informações. A segunda tem informações sobre os vínculos empregatícios, algumas das informações encontradas, nessa base, são mês de admissão, mês de desligamento e idade do trabalhador. A Figura 6 mostra um recorte da base de estabelecimentos da RAIS para o ano de 2015 e a Tabela 2 contém a explicação de cada campo.

Figura 6 - Recorte da base de estabelecimentos do 2015

Bairros SP	Bairros Fortaleza	Bairros RJ	CNAE 2.0 Classe	CNAE 95 Classe	Distritos SP	Qtd Vínculos CLT	Qtd Vínculos Ativos
{ñ c}	40 {ñ c}		82997	74993 {ñ clas		6	6
{ñ c}	54 {ñ c}		69206	74128 {ñ clas		4	4
{ñ c}	3 {ñ c}		10929	15822 {ñ clas		13	13
{ñ c}	61 {ñ c}		47245	52299 {ñ clas		0	0
{ñ c}	22 {ñ c}		73114	74403 {ñ clas		14	14
{ñ c}	65 {ñ c}		47571	52493 {ñ clas		1	1
{ñ c}	71 {ñ c}		47296	52299 {ñ clas		2	2
{ñ c}	66 {ñ c}		47229	52299 {ñ clas		2	2
{ñ c}	22 {ñ c}		56201	55247 {ñ clas		4	4

Qtd Vínculos Estatutários	Ind Atividade Ano	Ind CEI Vinculado	Ind Estab Participa PAT	Ind Rais Negativa	Ind Simples	Município	Natureza Jurídica
0	1	0	0	0	0	230440	2062
0	1	0	0	0	1	230440	2135
0	1	0	0	0	1	230440	2062
0	1	0	0	0	1	230440	2135
0	1	0	0	0	1	230440	2062
0	1	0	0	0	1	230440	2135
0	1	0	0	0	1	230440	2135
0	1	0	0	0	1	230440	2135
0	1	0	0	0	1	230440	2062

Regiões Adm DF	CNAE 2.0 Subclasse	Tamanho Estabelecimento	Tipo Estab	Tipo Estab	UF	IBGE Subsetor	CEP Estab
0	8299799	3	1 Cnpj	23	19	60120002	
0	6920601	2	1 Cnpj	23	19	60740101	
0	1092900	4	1 Cnpj	23	13	60150060	
0	4724500	1	1 Cnpj	23	16	60750350	
0	7311400	4	1 Cnpj	23	19	60821800	
0	4757100	2	1 Cnpj	23	16	60130370	
0	4729699	2	1 Cnpj	23	16	60347840	
0	4722902	2	1 Cnpj	23	16	60175220	
0	5620102	2	1 Cnpj	23	13	60822040	

Fonte: RAIS (2015)

Tabela 2 - Descrição das variáveis disponibilizadas na planilha de estabelecimentos da RAIS para o ano de 2015

Nome da Variável	Descrição da Variável
Bairros SP	Bairros do Município de São Paulo
Bairros Fortaleza	Bairros do município de Fortaleza
Bairros RJ	Bairros do município do Rio de Janeiro
CNAE 2.0 Classe	Classe de Atividade Econômica, segundo a classificação CNAE 2.0
CNAE 95 Classe	Classe de Atividade Econômica segundo a classificação CNAE 1.0, revisada em 2002
Distritos SP	Distritos do município de São Paulo
Qtd Vínculos CLT	Estoque de vínculos, sob o regime CLT e Outros, ativos em 31/12
Qtd Vínculos Ativos	Estoque de vínculos ativos em 31/12
Qtd Vínculos Estatutários	Estoque de vínculos, sob o regime estatutário, ativos em 31/12
Ind Atividade Ano	Indicador de estabelecimento/entidade que exerceu atividade durante o ano de referência
Ind CEI Vinculado	Indicador de CEI vinculado
Ind Estab Participa PAT	Indicador de estabelecimento pertencente ao PAT
Ind Rais Negativa	Indicador de Rais Negativa
Ind Simples	Indicador de optante pelo SIMPLES
Município	Município de localização do estabelecimento
Natureza Jurídica	Natureza Jurídica (CONCLA/2002) - a partir da RAIS2008
Regiões Adm DF	Regiões Administrativas do Distrito Federal
CNAE 2.0 Subclasse	Subclasse de Atividade Econômica, segundo classificação CNAE 2.0
Tamanho Estabelecimento	Tamanho do estabelecimento - empregados ativos em 31/12
Tipo Estab	Tipo de estabelecimento
UF	Unidade da Federação do estabelecimento
IBGE Subsetor	Subsetor IBGE 80 do estabelecimento
CEP Estb	CEP declarado pelo estabelecimento

Fonte: Elaborado pelo autor

5.1.2 Dados da SEFAZ-CE

A base da SEFAZ-CE utilizada no trabalho foi fornecida ao grupo de pesquisa pelo órgão mediante solicitação e justificativa de uso. A base que foi disponibilizada trouxe informações valiosas para a modelagem da produção e atração de notas fiscais. Pôde-se encontrar, nessa base, informações sobre as notas fiscais emitidas no Ceará, como CEP da empresa emissora, CEP da empresa receptora e quantidade de notas por transação. Um recorte da base fornecida pela SEFAZ-CE para o ano de 2015 está representado na Figura 7. E a explicação de cada campo está na Tabela 3. Informações que revelassem a identidade das empresas, como o CNPJ, não foram disponibilizadas por questões de sigilo fiscal.

Figura 7 - Recorte da base de notas fiscais do 2015

ANO	MES	UF_EMI	MUNICIPIO_EMI	BAIRRO_EMI	CEP_EMI	CNAE_EMITENTE	UF_DES
2015	1	CE	FORTALEZA	AGUA FRIA	60811340	4711301	CE
2015	1	CE	FORTALEZA	AGUA FRIA	60811340	4711301	CE
2015	1	CE	FORTALEZA	AGUA FRIA	60811340	4711301	CE
2015	1	CE	FORTALEZA	AGUA FRIA	60811340	4711301	CE
2015	1	CE	FORTALEZA	AGUA FRIA	60811340	4711301	CE
2015	1	CE	FORTALEZA	AGUA FRIA	60811340	4711301	CE
2015	1	CE	FORTALEZA	AGUA FRIA	60811340	4711301	CE
2015	1	CE	FORTALEZA	AGUA FRIA	60811340	4711301	CE
2015	1	CE	FORTALEZA	AGUA FRIA	60811340	4711301	CE
2015	1	CE	FORTALEZA	AGUA FRIA	60811340	4711301	CE

MUNICIPIO_DES	BAIRRO_DES	CEP_DES	CNAE_DESTINATARIO	PES_BRUTO	QTD_VOLUME	QTDE_NFS
FORTALEZA	CIDADE DOS FUNCIONAR	60822130	4711302	0	120	1
FORTALEZA	CIDADE DOS FUNCIONAR	60823105	4639701	0	1729	4
FORTALEZA	DIONISIO TORRES	60170251	4711302	0	3	1
FORTALEZA	DIST DE MESSEJANA	60842395	4639701	0	724	1
FORTALEZA	MESSEJANA	60850015	4636202	0	1300	1
FORTALEZA	MONDUBIM	60761505	4634602	0	187	3
FORTALEZA	MONDUBIM	60766015	1031700	0	625	1
FORTALEZA	MUCURIBE	60165081	4711302	0	2	1
FORTALEZA	PAPICU	60190800	4711302	0	12	2

Fonte: SEFAZ-CE (2015)

Tabela 3 - Descrição dos dados sobre notas fiscais disponibilizados pela SEFAZ

Nome da Variável	Descrição da Variável
ANO	Ano em que a nota fiscal foi emitida
MÊS	Mês em que a nota fiscal foi emitida
UF_EMI	Unidade da federação onde a nota fiscal foi emitida
MUNICIPIO_EMI	Município onde a nota fiscal foi emitida
BAIRRO_EMI	Bairro onde a nota fiscal foi emitida
CEP_EMI	CEP de onde a nota fiscal foi emitida
CNAE_EMITENTE	CNAE da empresa que emitiu a nota fiscal
UF_DES	Unidade da federação do receptor da nota fiscal
MUNICIPIO_DES	Município do receptor da nota fiscal
BAIRRO_DES	Bairro do receptor da nota fiscal
CEP_DES	CEP do receptor da nota fiscal
CNAE_DESTINATARIO	CNAE da empresa receptora da nota fiscal
PES_BRUTO	Peso bruto declarado na nota fiscal
QTD_VOLUME	Volume declarado na nota fiscal
QTDE_NFS	Quantidade de notas fiscais que tenham todos os dados acima iguais.

Fonte: Elaborado pelo autor

5.1.3 Consolidação dos dados

Foi necessário organizar e filtrar os dados para obter apenas os dados que eram de interesse para a pesquisa e, dessa forma, conseguir manipular os dados para obter os resultados e as análises desejadas. Um exemplo de filtragem que precisou ser feita foi a filtragem dos CEPs, para que fossem trabalhados apenas os dados que tinham CEP pertencente a cidade estudada. A figura 8 mostra o intervalo de CEP que era de interesse para a pesquisa e foi retirada do site dos correios.

Figura 8 - Faixa de CEP da cidade de Fortaleza

Localidade	Faixa de CEP	Situação	Tipo de Faixa
Fortaleza	60000-001 a 61599-999	Codificado por logradouros	Total do município
Fortaleza	60000-001 a 61599-999	Codificada por logradouros	Exclusiva da sede urbana

Fonte: Correios (2017)

Além disso, algumas variáveis foram excluídas das bases, pois continham muitas informações em branco ou por acreditar-se que continham informações inconsistentes e outras por não representarem o fenômeno que estava sendo estudado e modelado.

Como foram utilizadas duas fontes de dados, foi necessário compatibilizar as bases. Era necessário que todas as bases de dados utilizassem o mesmo sistema de divisão de bairros, pois isso seria importante tanto para a análise exploratória quanto para a construção dos modelos de geração de notas fiscais, uma vez que as informações foram agregadas por bairros.

Foi preciso avaliar se as informações dos bairros contidas nas bases seriam utilizadas ou se seriam adicionadas novas informações de bairros com base no CEP e na divisão de bairros utilizada pelos correios, pois era importante que essas informações fossem precisas e em outros trabalhos observou-se que houve problemas ao usar tais informações dessas bases de dados. Dentre os problemas encontrados anteriormente estão: divisão diferente da utilizada pelos correios, ortografia errada e informações erradas por falta de conhecimento sobre os limites dos bairros.

Todos os processos de manipulação da base de dados, seja filtragem ou adição de informações, seriam muito dispendiosos se fossem feitos em planilhas eletrônicas, como o Microsoft Excel, uma vez que os arquivos brutos continham informações para todo o estado do Ceará e, conseqüentemente, totalizavam dezenas de milhões de linhas. Dessa forma, toda a organização e compatibilização dos dados foi realizada por meio de *scripts* desenvolvidos, pelo autor, em Python.

5.1.4 Análise exploratória

Com a definição das bases que seriam usadas, com os dados organizados e com a adição de informações importantes para a pesquisa, foi possível prosseguir e trabalhar os dados para obter os resultados desejados. Partiu-se então para uma análise exploratória para

ter uma visão geral de como estava ocorrendo geração de notas fiscais e das possíveis variáveis explicativas de forma geral pela cidade.

Devido ao fato da cidade de Fortaleza ter centenas de CNAEs e 122 bairros, seria inviável analisar todos eles. Dessa forma, decidiu-se avaliar somente os CNAEs e os bairros que concentrassem uma quantidade significativa das transações comerciais da cidade.

Com essas informações foram criados gráficos e tabelas para facilitar a visualização e buscar resultados que pudessem contribuir para o entendimento de como ocorria a geração de notas fiscais em Fortaleza. Buscou-se saber quais CNAEs e quais bairros emitiam mais notas fiscais, qual porcentagem que cada CNAE e que cada bairro representa do total das notas fiscais emitidas e atraídas, como ocorre a variação mensal da emissão e atração de notas fiscais por CNAE baseado nos dados de 2015, quais CNAEs e quais bairros detêm mais empregados e empresas, entre outras informações.

5.2 Seleção das variáveis e desenvolvimento dos modelos de geração de notas fiscais

Após a análise exploratória, partiu-se para a escolha das variáveis que seriam testadas nos modelos de regressão linear. A escolha das variáveis foi baseada na confiabilidade e na coerência com que cada variável foi apresentada nas bases de dados e na adequação aos modelos de regressão linear simples. Além disso, buscou-se variáveis que já foram utilizadas para a modelagem do transporte urbano de carga, como número de empregados, pois acredita-se que a geração de viagens de carga está correlacionada com a geração de notas fiscais.

Da mesma forma como foi feito na análise exploratória, decidiu-se selecionar os CNAEs e os bairros que concentrassem uma quantidade significativa das transações comerciais da cidade e gerar modelos apenas para eles.

Os dados de empresas que não continham funcionários foram excluídos, pois acredita-se que esses dados têm uma grande chance de erro, uma vez que existem CNAEs em que não faz sentido uma empresa sem funcionários produzir grandes quantidades de notas fiscais, mas, mesmo assim, estavam presentes na base de dados.

Após a escolha das variáveis, elas foram separadas do restante do banco de dados e então partiu-se para a realização das regressões lineares utilizando o método dos mínimos quadrados para cada um dos CNAEs selecionados. Foram criados modelos de atração e produção de notas fiscais tanto para os maiores emissores quanto para os maiores

destinatários. Todas as regressões e testes estatísticos foram realizadas por meio de códigos desenvolvidos em Python.

Para avaliar os modelos gerados foram analisados o R^2 , teste F e teste t. Em todos os testes foi adotado nível de significância de 5%. Além disso, para verificar se a estimação dos parâmetros do modelo de regressão e os testes de hipóteses eram confiáveis, foi necessário verificar algumas premissas, como a normalidade, a variância e a independência dos erros.

A distribuição normal dos erros é uma das hipóteses dos modelos de regressão linear utilizados no trabalho. Para que as hipóteses sobre a inclinação e sobre o intercepto do modelo de regressão possam ser testadas é necessário verificar a normalidade dos erros (Montgomery, 2009). Dessa forma, a normalidade dos erros em cada uma das regressões foi verificada por meio do teste Shapiro-Wilk. As hipóteses desse teste estão apresentadas na equação 7.

$$\begin{cases} H_0: \text{Distribuição é normal} \\ H_1: \text{Distribuição não é normal} \end{cases} \quad (7)$$

Outro ponto importante é a homoscedasticidade da variância dos erros. A homoscedasticidade é uma das suposições requeridas para a estimação dos coeficientes do modelo. Segundo MONTGOMERY (2009), caso a homoscedasticidade não seja respeitada deve-se considerar a adição de termos de ordens maiores ao modelo, transformações das variáveis ou, até mesmo, a adição de outros regressores. A verificação da homoscedasticidade foi feita pelo teste de Breusch-Pagan. Suas hipóteses estão na equação 8.

$$\begin{cases} H_0: \text{Homocedasticidade} \\ H_1: \text{Heterocedasticidade} \end{cases} \quad (8)$$

Além disso, foi verificada a independência dos erros. Em uma regressão linear não é desejado que o erro de uma determinada observação sofra influência do erro de outra observação. Caso haja autocorrelação, os problemas serão parecidos com os causados pela heterocedasticidade. Para essa verificação foi utilizado o teste de Durbin-Watson. As hipóteses desse teste estão apresentadas na equação 9.

$$\begin{cases} d_U < DW < (4 - d_U): \text{Ausência de autocorrelação} \\ d_L \leq DW \leq d_U \text{ ou } (4 - d_U) \leq DW \leq (4 - d_L): \text{Teste inconclusivo} \\ 0 \leq DW \leq d_L \text{ ou } (4 - d_L) \leq DW \leq 4 : \text{Autocorrelação} \end{cases} \quad (9)$$

6 RESULTADOS E DISCUSSÕES

Para produzir os resultados da análise exploratória foram utilizadas a base da RAIS (2014 e 2015) de estabelecimentos e a base da SEFAZ-CE (entre 2011 e 2015), pois a base da RAIS de vínculos não contava com CEP e na divisão dos bairros feita na RAIS há apenas 71 bairros, que não sabe-se como foram divididos. Para as regressões foram utilizados os dados de emprego ativos em 31 de dezembro, de número de empresas que exerceram atividade no ano-base e de número de notas fiscais emitidas e recebidas no mês de dezembro e no ano todo. Para desenvolver os modelos de regressão foram utilizados os dados dos anos de 2014 e 2015. Foram criados oito tipos de modelos, que serão descritos posteriormente.

Neste capítulo, quando for utilizado número de empregos, refere-se ao número de empregos em 31 de dezembro e quando for utilizado número de empresas, refere-se ao número de empresas que exerceram atividade no ano em questão, ambos presentes na base da RAIS.

Os resultados da análise exploratória aqui apresentados são para os 20 CNAEs e os 20 bairros que mais produziram notas fiscais, que mais atraíram notas fiscais, que mais tinham empregos e que mais tinham empresas. Já as regressões lineares simples foram realizadas para os 20 CNAEs que mais produziram notas fiscais e para os 20 CNAEs que mais atraíram notas fiscais em 2015. Essa decisão foi tomada pois seria inviável apresentar e comentar os resultados para os 122 bairros encontrados e para as centenas CNAEs existentes nas bases de dados. Além disso, foi constatado que os conjuntos que foram definidos produzem e atraem mais da metade das notas fiscais emitidas e atraídas na cidade, e detêm uma parte muito grande dos empregos formais e das empresas que exerceram atividade nos anos analisados.

Um dicionário com a denominação dos CNAEs citados no trabalho pode ser encontrado no apêndice.

6.1 Análise exploratória

Na Tabela 4 pode-se encontrar os 20 CNAEs que mais emitiram notas fiscais de 2011 a 2015 organizados em ordem decrescente. É interessante notar que de 2011 para 2015 apenas 5 CNAEs foram trocados na composição dos 20 maiores emissores.

Tabela 4 - CNAEs que mais emitiram notas fiscais de 2011 a 2015

	2011	2012	2013	2014	2015
1º	46397	46397	46397	46397	46397
2º	45307	45307	45307	45307	45307
3º	46354	46354	46443	47113	47113
4º	45111	47113	47113	46443	46443
5º	46443	45111	46354	45111	45111
6º	47113	46443	45111	46354	47440
7º	46494	46494	46494	47440	46354
8º	47440	47440	47440	46494	46796
9º	11216	11216	46796	46796	46494
10º	46478	46478	11216	46346	46346
11º	45412	14126	14126	11216	47717
12º	14126	46796	46478	47512	46460
13º	11224	45412	45412	46478	46478
14º	46362	47512	46346	14126	47890
15º	47512	46818	47512	47890	45412
16º	46320	46346	47890	45412	47512
17º	46818	46362	46818	46362	14126
18º	46371	46516	46516	46460	11216
19º	46516	47890	46729	47717	10911
20º	46796	10911	10911	51111	46516

Fonte: Elaborado pelo autor

A Tabela 5 representa os 20 CNAEs que mais atraíram notas fiscais de 2011 até 2015. As mudanças para as atrações são ainda menores do que as mudanças ocorridas nas emissões, somente 2 CNAEs foram substituídos na composição dos 20 que mais atraíram notas fiscais de 2011 para 2015.

Tabela 5 - CNAEs que mais atraíram notas fiscais de 2011 a 2015

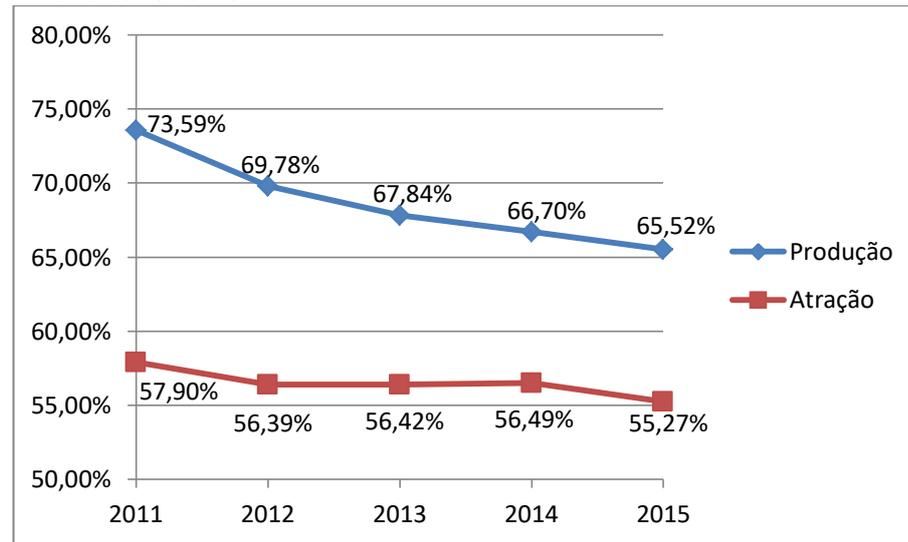
	2011	2012	2013	2014	2015
1º	47121	47113	47113	47113	47113
2º	47113	47121	47121	47121	47121
3º	45307	45307	45307	45307	45307
4º	47717	47717	47717	47717	47717
5º	56112	56112	56112	56112	56112
6º	45111	45111	45111	47440	47440
7º	47318	47318	47440	45111	47318
8º	10911	47440	47318	47318	10911
9º	41204	10911	41204	10911	45111
10º	46397	41204	10911	41204	41204
11º	47440	14126	14126	51111	14126
12º	14126	46397	51111	14126	51111
13º	47814	47814	46397	47890	47814
14º	47512	47512	47890	46397	47890
15º	47890	47890	47814	47814	45412
16º	47741	47741	47512	45412	46397
17º	45412	45412	45412	47512	47512
18º	47211	51111	47741	47741	47822
19º	47296	47610	47725	47725	47296
20º	49302	47822	47610	47296	47741

Fonte: Elaborado pelo autor

Isso demonstra que para mudanças significativas na participação de segmentos econômicos ocorrerem é necessário um tempo considerável, no período analisado as mudanças foram pequenas. As mudanças ocorreram principalmente na parte inferior da tabela, mostrando que os grandes emissores de notas fiscais continuam sendo os maiores no curto prazo.

Além disso, os 20 primeiros CNAEs que mais emitem notas e os que mais atraem são responsáveis por uma porcentagem muito grande do total de notas fiscais. Entretanto, isso vem mudando com o passar dos anos. Enquanto a concentração da atração de notas fiscais permanece quase inalterada no período analisado, existe uma notória tendência de descentralização da produção de notas fiscais, como mostra Gráfico 1.

Gráfico 1 - Participação dos 20 CNAEs que mais emitiram e dos 20 CNAEs que mais atraíram notas fiscais

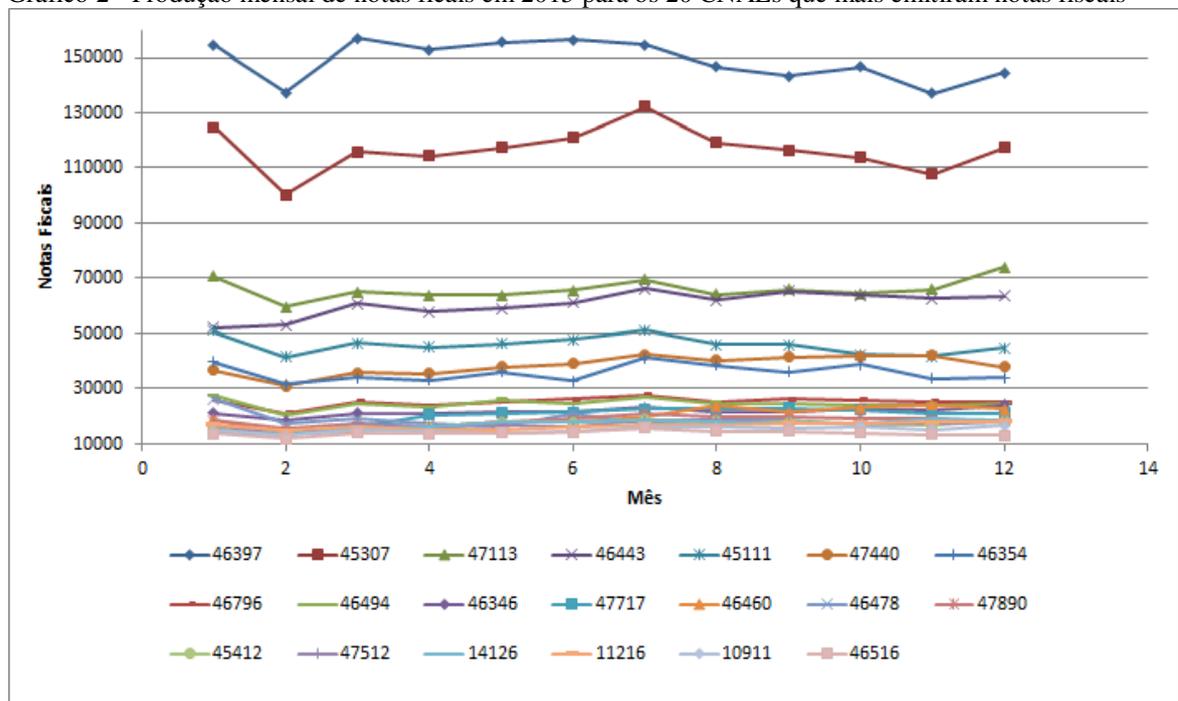


Fonte: Elaborado pelo autor

O Gráfico 2 representa a variação mensal da emissão de notas fiscais para os 20 segmentos que mais emitiram notas fiscais em 2015. Pode-se notar que a variação mensal de notas fiscais é pequena e isso pode ser confirmado pela Tabela 6, que contém a média e o desvio padrão do número de notas fiscais emitidos por mês para cada CNAE.

Além disso, pode-se perceber que é comum uma queda na emissão de notas fiscais no mês de fevereiro e o mês que, geralmente, tem o pico de emissões é julho.

Gráfico 2 - Produção mensal de notas fiscais em 2015 para os 20 CNAEs que mais emitiram notas fiscais



Fonte: Elaborado pelo autor

Tabela 6 - Média e desvio padrão da produção mensal de notas fiscais dos 20 CNAEs que mais emitiram notas em 2015

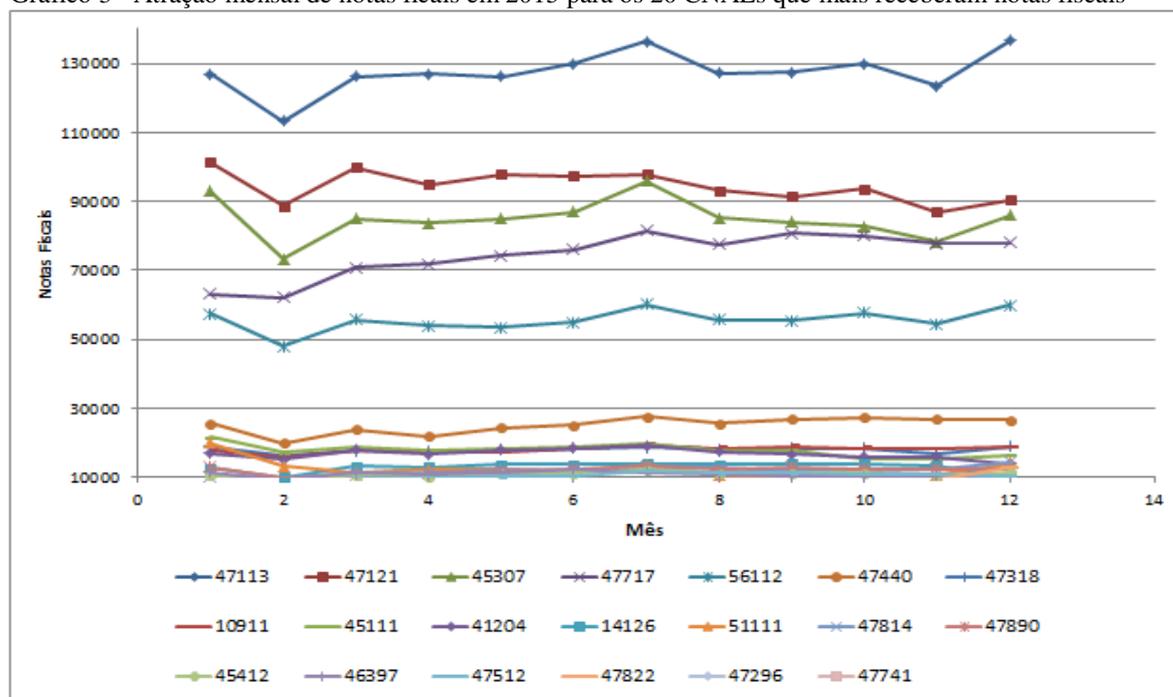
CNAE	Média	Desvio Padrão
46397	148867,3	6960,9
45307	116630,8	7593,2
47113	66041,9	3556,7
46443	60672,8	4225,8
45111	45846,4	3061,8
47440	38498,5	3281,4
46354	35819,9	2945,7
46796	25269,3	1515,6
46494	24563,3	1688,9
46346	21734,4	1310,6
47717	20126,7	2969,6
46460	19523,7	3395,7
46478	18779,3	2594,7
47890	18514,0	1540,9
45412	17131,9	1350,1
47512	17072,1	1334,1
14126	16967,4	1642,0
11216	16560,5	1112,9
10911	15161,5	986,4
46516	13951,7	929,9

Fonte: Elaborado pelo autor

Já o Gráfico 3 representa a variação mensal da atração de notas fiscais para os 20 segmentos que mais atraíram notas fiscais em 2015. Da mesma forma como ocorre para a emissão, pode-se notar que a variação mensal de notas fiscais atraídas não é tão acentuada. A Tabela 7 representa a média e o desvio padrão do número de notas fiscais atraídas por mês para cada CNAE.

Pode ser ver que fevereiro é o mês de queda mais abrupta para grande parte dos setores e julho é o pico, da mesma forma da produção de notas fiscais, o que era esperado, pois quanto mais notas fiscais estão sendo emitidas, mais notas fiscais estão sendo atraídas entre os diversos setores.

Gráfico 3 - Atração mensal de notas fiscais em 2015 para os 20 CNAEs que mais receberam notas fiscais



Fonte: Elaborado pelo autor

Tabela 7 - Média e desvio padrão da atração mensal de notas fiscais dos 20 CNAEs que mais receberam notas em 2015

CNAE	Média	Desvio Padrão
47113	127581,8	5748,7
47121	94498,9	4328,8
45307	84906,8	5597,4
47717	74471,8	6152,8
56112	55527,9	3044,7
47440	25153,9	2271,7
47318	18070,3	763,5
10911	18036,5	901,3
45111	17989,9	1701,7
41204	16885,3	1365,2
14126	12910,0	1381,4
51111	12357,9	2713,0
47814	12011,3	1207,1
47890	11989,3	1043,1
45412	10832,4	813,0
46397	10812,3	743,6
47512	10428,1	994,1
47822	8825,1	1519,2
47296	8276,3	429,1
47741	7643,4	753,1

Fonte: Elaborado pelo autor

Outro ponto interessante de ser analisado é a geração de notas fiscais por bairros, uma vez que para que decisões de planejamento sejam tomadas, além de entender quais os segmentos econômicos mais têm impacto nas transações, é importante entender de onde as notas estão saindo e onde estão chegando. A Tabela 8 apresenta os 20 bairros que mais emitiram nos fiscais de 2011 a 2015. Já a tabela 9 mostra os 20 bairros que mais atraíram notas fiscais no mesmo período.

Tabela 8 - Bairros que mais emitiram notas fiscais de 2011 a 2015

	2011	2012	2013	2014	2015
1º	Centro	Centro	Centro	Centro	Centro
2º	Barra do Ceará	Barra do Ceará	Messejana	Messejana	Messejana
3º	Farias Brito	Messejana	Barra do Ceará	Barra do Ceará	Barra do Ceará
4º	Messejana	Farias Brito	Farias Brito	Farias Brito	Montese
5º	Cajazeiras	Cajazeiras	Joaquim Távora	Aldeota	Farias Brito
6º	Cais do Porto	Joaquim Távora	Cajazeiras	Joaquim Távora	Paupina
7º	Paupina	Cais do Porto	Aldeota	Carlito Pamplona	Aldeota
8º	Presidente Kennedy	Presidente Kennedy	Cais do Porto	Paupina	Carlito Pamplona
9º	Joaquim Távora	Aldeota	Carlito Pamplona	José Bonifácio	Joaquim Távora
10º	José Bonifácio	Passaré	José Bonifácio	Cais do Porto	José Bonifácio
11º	Aldeota	Paupina	Meireles	Cidade dos Funcionários	Cais do Porto
12º	Passaré	José Bonifácio	Passaré	Meireles	Parangaba
13º	Sabiaguaba	Meireles	Presidente Kennedy	Parangaba	Meireles
14º	Carlito Pamplona	Sabiaguaba	Paupina	Passaré	Mondubim
15º	Aeroporto	Carlito Pamplona	Parangaba	Pedras	Cidade dos Funcionários
16º	Jangurussu	Jangurussu	Vila União	Mondubim	Pedras
17º	Meireles	Parangaba	Jangurussu	Cajazeiras	Ancuri
18º	Barroso	Aeroporto	Sabiaguaba	Jangurussu	Edson Queiroz
19º	Parangaba	Barroso	Cidade dos Funcionários	Dias Macedo	Aerolândia
20º	Cidade dos Funcionários	Benfica	Boa Vista	Vila União	Passaré

Fonte: Elaborado pelo autor

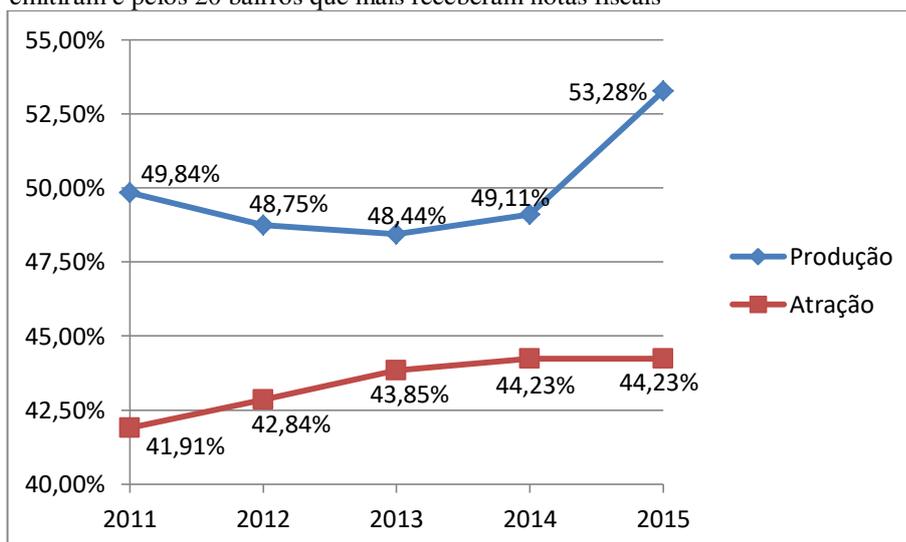
Tabela 9 - Bairros que mais atraíram notas fiscais de 2011 a 2015

	2011	2012	2013	2014	2015
1º	Centro	Centro	Centro	Centro	Centro
2º	Meireles	Meireles	Meireles	Meireles	Meireles
3º	Aldeota	Aldeota	Aldeota	Aldeota	Aldeota
4º	Joaquim Távora				
5º	Messejana	Messejana	Messejana	Messejana	Messejana
6º	Parangaba	Parangaba	Montese	Parangaba	Parangaba
7º	Fátima	Montese	Parangaba	Montese	Montese
8º	Farias Brito	Fátima	Fátima	Fátima	Papicu
9º	José Bonifácio	Presidente Kennedy	José Bonifácio	Dionisio Torres	Fátima
10º	Montese	Farias Brito	Dionisio Torres	Edson Queiroz	Edson Queiroz
11º	Presidente Kennedy	José Bonifácio	Presidente Kennedy	José Bonifácio	Dionisio Torres
12º	Dionisio Torres	Edson Queiroz	Edson Queiroz	Dias Macedo	José Bonifácio
13º	Edson Queiroz	Dionisio Torres	Farias Brito	Prefeito José Walter	Barra do Ceará
14º	São João do Tauape	São João do Tauape	Dias Macedo	Presidente Kennedy	Cidade dos Funcionários
15º	Benfica	São Gerardo	Barra do Ceará	Farias Brito	Farias Brito
16º	São Gerardo	Benfica	Cidade dos Funcionários	Cidade dos Funcionários	Dias Macedo
17º	Barra do Ceará	Cidade dos Funcionários	São João do Tauape	Barra do Ceará	Prefeito José Walter
18º	Cidade dos Funcionários	Barra do Ceará	São Gerardo	São João do Tauape	Presidente Kennedy
19º	Prefeito José Walter	Prefeito José Walter	Benfica	São Gerardo	São João do Tauape
20º	Mucuripe	Cais do Porto	Prefeito José Walter	Papicu	Passaré

Fonte: Elaborado pelo autor

É importante perceber que o bairro Centro manteve-se hegemônico como bairro que mais produz e atrai notas fiscais durante todos os anos que foram analisados. Na tabela de atração de notas fiscais por bairro pode-se ver que os cinco bairros que mais atraem notas fiscais permaneceram em suas posições durante todos os anos. Isso demonstra que cinco anos não é um período suficientemente extenso para que grandes mudanças sejam percebidas na geração de notas fiscais na cidade de Fortaleza, principalmente quanto a atração.

Gráfico 4 - Porcentagem de notas emitidas e atraídas pelos 20 bairros que mais emitiram e pelos 20 bairros que mais receberam notas fiscais



Fonte: Elaborado pelo autor

Pode-se ver no Gráfico 4 que 20 bairros produzem mais da metade das notas fiscais emitidas na cidade e 20 bairros recebem quase metade das notas. Além disso, as tabelas contendo os 20 bairros que mais emitiram notas fiscais e os 20 bairros que mais receberam notas fiscais compartilham muitos bairros. Esses dados sugerem que Fortaleza possui uma grande parcela de sua atividade econômica concentrada em poucos bairros, com destaque para o Centro e Messejana.

Na Tabela 10 estão os 20 CNAEs com mais empregos na cidade de Fortaleza, em dezembro de 2014 e 2015 os maiores empregadores continuam os mesmos, a única diferença é a ordem em que aparecem, entretanto pode-se ver que os três primeiros continuam os mesmos.

Tabela 10 - 20 CNAEs que tinham mais empregados ativos em 31 dezembro de 2014 e 2015

	dez/2014	dez/2015
1º	41204	41204
2º	84124	84124
3º	56112	56112
4º	14126	78205
5º	78205	14126
6º	47113	47113
7º	84116	84116
8º	82202	82997
9º	86101	86101
10º	82997	82202
11º	78108	81125
12º	80111	80111
13º	81125	78108
14º	47814	47814
15º	85139	85139
16º	41107	81214
17º	81214	85317
18º	49302	49302
19º	49213	41107
20º	85317	49213

Fonte: Elaborado pelo autor

É importante notar que apenas 7 dos 20 CNAEs com o maior número de empresas estão entre os 20 CNAEs que mais geram empregos, isso pode ser visto comparando-se a Tabela 10 com a Tabela 11. Da mesma forma dos 20 CNAEs que mais empregam, de 2014 para 2015 não houve mudança nos 20 CNAEs com mais empresas, mudando apenas a ordem.

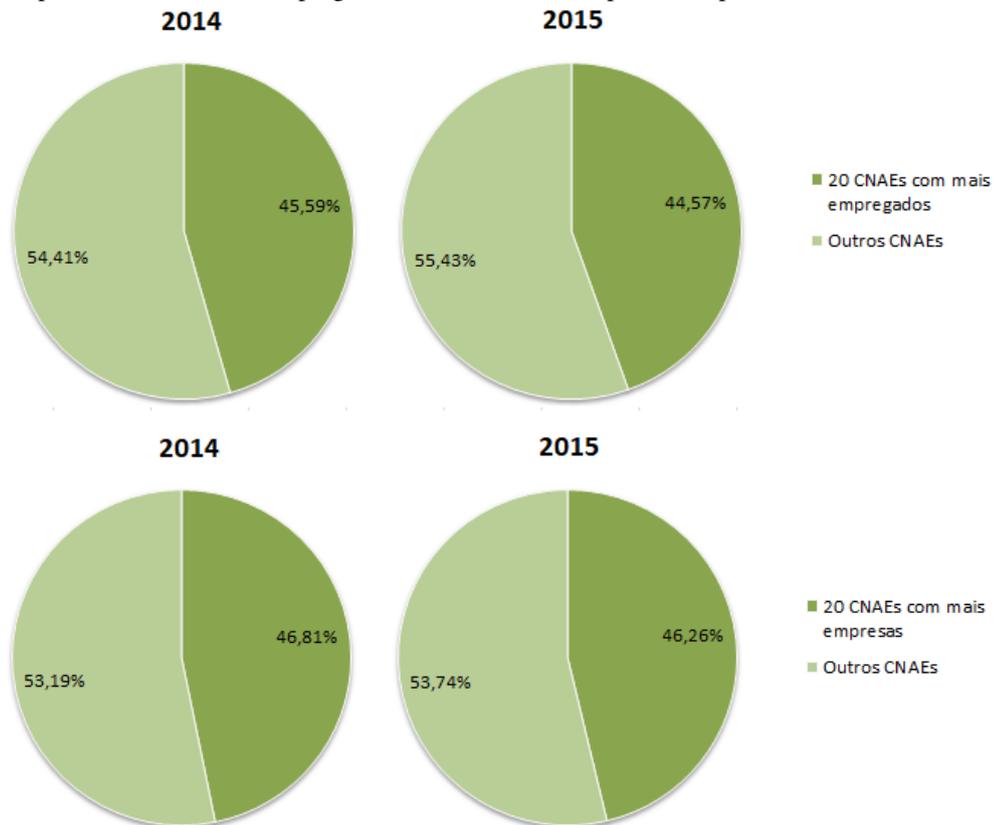
Tabela 11 - 20 CNAEs com mais empresas ativas em 2014 e 2015

	2014	2015
1º	47121	56112
2º	56112	47814
3º	47814	47121
4º	41204	81125
5º	81125	41204
6º	14126	14126
7º	47440	86305
8º	86305	47440
9º	47890	47890
10º	47555	47555
11º	45307	96025
12º	96025	45307
13º	45200	45200
14º	69206	69206
15º	47547	49302
16º	49302	47547
17º	47717	94910
18º	94308	94308
19º	94910	47717
20º	47741	47741

Fonte: Elaborado pelo autor

A Figura 9 mostra que os 20 CNAEs que mais empregam pessoas foram responsáveis, em 2014 e 2015, por aproximadamente 45,59% e 44,57% dos empregos da cidade de Fortaleza nos anos analisados e, da mesma forma, os 20 CNAEs que mais têm empresas na cidade corresponderam, em 2014 e 2015, a cerca de 46,81% e 46,26% de todas as empresas. Mostrando que apenas 20 segmentos econômicos concentraram quase metade dos empregos formais e das empresas ativas da cidade de Fortaleza.

Figura 9 - Participação dos 20 CNAEs que mais empregam e dos 20 CNAEs com mais empresas no número de empregados e no número de empresas, respectivamente



Fonte: Elaborado pelo autor

É importante, além de conhecer quais os segmentos econômicos das empresas e dos empregos, conhecer onde eles estão. Pela Tabela 12 pode-se conhecer os 20 bairros que mais geraram empregos em fortaleza nos anos de 2014 e 2015. Da mesma forma dos resultados obtidos ao agregar-se os dados por CNAE, os resultados dos 20 bairros que mais geraram empregos mostra que de 2014 para 2015 esse grupo permaneceu quase que inalterado, apenas o bairro Vila União que aparecia como o 19º maior gerador de empregos saiu para a entrada do Itaperi. Esse resultado era esperado, visto o intervalo de tempo reduzido.

Tabela 12 - 20 Bairros que mais empregaram em dezembro de 2014 e 2015

	dez/2014	dez/2015
1º	Centro	Centro
2º	Aldeota	Aldeota
3º	Meireles	Meireles
4º	Joaquim Távora	Joaquim Távora
5º	Dionisio Torres	Dionisio Torres
6º	Messejana	Edson Queiroz
7º	Fátima	Fátima
8º	Edson Queiroz	Messejana
9º	Aeroporto	Montese
10º	José Bonifácio	José Bonifácio
11º	Montese	Benfica
12º	Benfica	Aeroporto
13º	Parangaba	Parangaba
14º	Papicu	Papicu
15º	Vicente Pinzon	Vicente Pinzon
16º	Barra do Ceará	Passaré
17º	Cidade dos Funcionários	Barra do Ceará
18º	Passaré	Cidade dos Funcionários
19º	Vila União	Cajazeiras
20º	Cajazeiras	Itaperi

Fonte: Elaborado pelo autor

A Tabela 13 mostra que de 2014 para 2015 os bairros Barra do Ceará e Vila Velha deram lugar a Cocó e Sapiranga no grupo dos 20 bairros com mais empresas. Comparando a Tabela 12 com a Tabela 13, pode-se notar que 4 dos bairros que mais geraram empregos em 2014 não estão no grupo dos que mais possuíam empresas, já em 2015 esse número sobe para 5. Esse resultado pode parecer estranho, uma vez que no senso comum onde há mais empresas existem mais empregados, entretanto isso se dá pelo fato de que a distribuição de empregados não ocorre uniformemente pelas empresas.

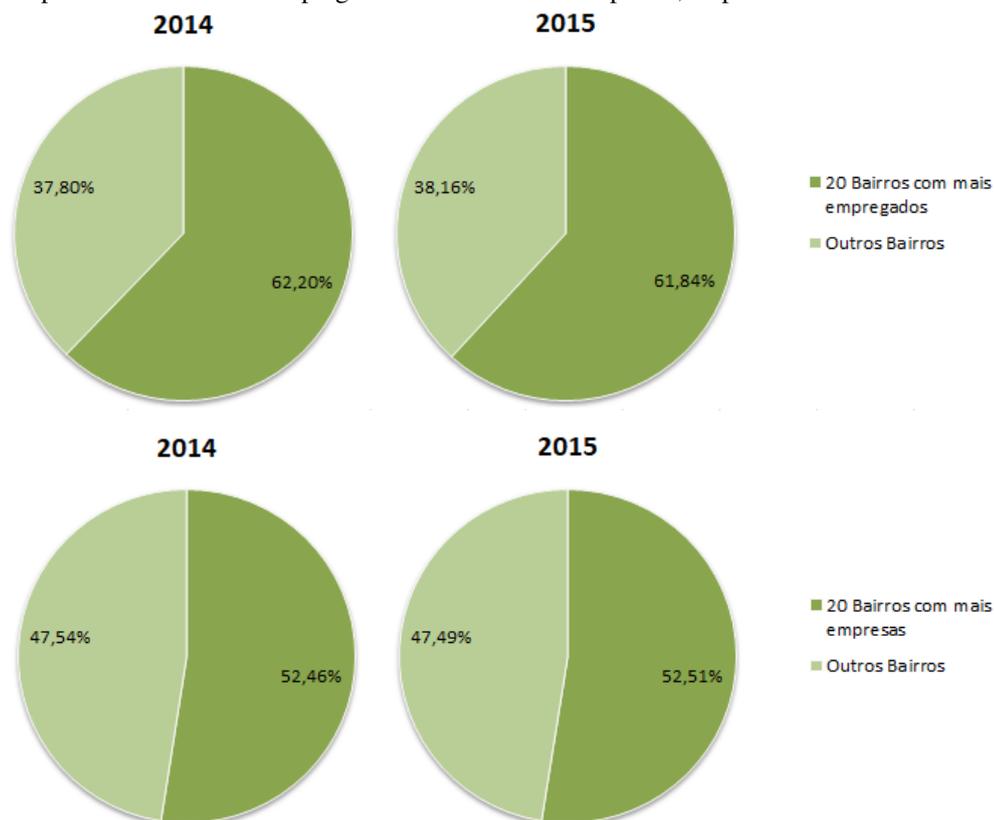
Tabela 13 - 20 bairros com mais empresas ativas em 2014 e 2015

	2014	2015
1º	Centro	Centro
2º	Meireles	Meireles
3º	Aldeota	Aldeota
4º	Joaquim Távora	Joaquim Távora
5º	Messejana	Messejana
6º	Fátima	Fátima
7º	Parangaba	Parangaba
8º	Dionisio Torres	Edson Queiroz
9º	Edson Queiroz	Dionisio Torres
10º	Papicu	Papicu
11º	Montese	Montese
12º	José Bonifácio	Mondubim
13º	Mondubim	José Bonifácio
14º	Cidade dos Funcionários	Cidade dos Funcionários
15º	Benfica	Benfica
16º	Conjunto Ceará	Passaré
17º	Passaré	Conjunto Ceará
18º	São João do Tauape	São João do Tauape
19º	Barra do Ceará	Cocó
20º	Vila Velha	Sapiranga

Fonte: Elaborado pelo autor

Ao realizar a mesma análise que foi feita quando os dados foram agregados por CNAE pode-se ver que existe uma grande concentração de empregos formais na cidade de Fortaleza. 20 bairros detinham 62,20% dos empregos da cidade em 2014 e, em 2015, esse número foi para 61,84%. Ao analisar os dados de números de empresas pode-se ver que os 20 bairros com mais empresas concentravam 52,46% de todas as empresas formais da cidade e, em 2015, esse número subiu para 52,51%, como mostra a Figura 10.

Figura 10 - Participação dos 20 bairros que mais empregam e dos 20 bairros com mais empresas no número de empregados e no número de empresas, respectivamente



Fonte: Elaborado pelo autor

6.2 Análise de regressão linear

Todas as regressões lineares e os testes estatísticos foram realizados por meio de códigos desenvolvidos em Python. Foram desenvolvidos modelos de regressão linear simples de emissão e atração de notas fiscais para todos os CNAEs que estavam entre os 20 CNAEs que mais produziram ou entre os 20 CNAEs que mais atraíram notas fiscais no ano de 2015. Ou seja, foram desenvolvidos modelos para 29 CNAEs (para os modelos de emissão de notas do mês de dezembro, o CNAE 51111 só emitiu notas fiscais em 2 bairros, logo não foi possível desenvolver modelo). Para todos os testes estatísticos aplicados foi considerado nível de significância de 5%. Nas tabelas, quando a célula estiver verde significa que o teste estatístico foi satisfeito, quando estiver vermelho significa que o teste estatístico não foi satisfeito e quando estiver amarelo significa que o teste foi inconclusivo.

Foram desenvolvidos oito configurações de modelo para cada CNAE, os tipos são apresentados na Tabela 14.

Tabela 14 - Modelos Desenvolvidos

Modelo	Variável Dependente	Variável Independente
1	Notas fiscais emitidas em dezembro	Empregados ativos em 31/12
2	Notas fiscais emitidas em dezembro	Empresas que exerceram atividade durante o ano
3	Notas fiscais atraídas em dezembro	Empregados ativos em 31/12
4	Notas fiscais atraídas em dezembro	Empresas que exerceram atividade durante o ano
5	Notas fiscais emitidas no ano	Empregados ativos em 31/12
6	Notas fiscais emitidas no ano	Empresas que exerceram atividade durante o ano
7	Notas fiscais atraídas no ano	Empregados ativos em 31/12
8	Notas fiscais atraídas no ano	Empresas que exerceram atividade durante o ano

Fonte: Elaborado pelo autor

Após apresentar todos os resultados das regressões e dos testes estatísticos é mostrado quais modelos mais se adequaram as premissas.

6.2.1 Notas fiscais de dezembro de 2014 e 2015

Todos os modelos desenvolvidos nessa subseção tiveram como variável dependente número de notas fiscais emitidas ou atraídas por bairro no mês de dezembro dos anos de 2014 e 2015. Primeiramente, apresentou-se os resultados de todos os modelos desenvolvidos e depois os testes para checar as premissas de normalidade, homoscedasticidade e independência dos erros são apresentados.

6.2.1.1 Modelos

Os resultados das regressões realizadas para explicar o fenômeno da emissão de notas fiscais utilizando número de empregos como variável explicativa estão apresentados na Tabela 15.

Tabela 15 – Resultados dos modelos tipo 1

$y = \beta_0 + \beta_1x$										
CNAE	Amostras	R ²	β_0	t-stat	P-Value	β_1	t-stat	P-Value	F	P-Value
10911	105	0,373	-0,24	0,00	0,997	4,43	7,83	0,000	61,29	0,000
11216	30	0,512	-359,55	-0,86	0,397	29,70	5,42	0,000	29,42	0,000
14126	186	0,041	139,42	5,16	0,000	0,12	2,79	0,006	7,80	0,006
41204	67	0,122	13,07	2,29	0,026	0,01	3,01	0,004	9,06	0,004
45111	84	0,649	183,46	1,63	0,107	7,52	12,30	0,000	151,35	0,000
45307	175	0,602	-298,46	-1,69	0,093	20,46	16,19	0,000	262,06	0,000
45412	86	0,625	-48,18	-0,89	0,376	12,96	11,84	0,000	140,09	0,000
46346	35	0,794	-219,24	-1,67	0,104	45,48	11,29	0,000	127,44	0,000
46354	16	0,993	-1062,16	-4,91	0,000	65,77	44,25	0,000	1958,12	0,000
46397	64	0,338	1321,67	1,81	0,075	19,91	5,63	0,000	31,67	0,000
46443	55	0,832	-610,41	-1,80	0,077	30,02	16,21	0,000	262,64	0,000
46460	30	0,685	-512,29	-1,37	0,183	51,31	7,80	0,000	60,91	0,000
46478	44	0,564	49,23	0,22	0,823	13,89	7,37	0,000	54,36	0,000
46494	65	0,938	-167,37	-2,81	0,007	25,83	30,79	0,000	948,33	0,000
46516	25	0,216	186,40	1,30	0,208	15,51	2,52	0,019	6,35	0,019
46796	61	0,714	88,99	1,08	0,286	22,78	12,13	0,000	147,18	0,000
47113	130	0,180	142,25	0,88	0,381	2,27	5,30	0,000	28,14	0,000
47121	130	0,114	-27,67	-0,49	0,626	1,94	4,07	0,000	16,53	0,000
47296	80	0,232	46,66	2,21	0,030	1,53	4,86	0,000	23,59	0,000
47318	112	0,358	11,97	1,49	0,138	0,85	7,83	0,000	61,29	0,000
47440	182	0,535	-139,29	-2,24	0,026	6,45	14,39	0,000	207,10	0,000
47512	75	0,587	41,18	0,66	0,514	7,94	10,19	0,000	103,74	0,000
47717	128	0,530	30,73	0,79	0,430	1,76	11,91	0,000	141,92	0,000
47741	35	0,953	-28,12	-2,84	0,008	1,11	25,75	0,000	662,89	0,000
47814	113	0,864	-5,09	-0,36	0,723	0,82	26,60	0,000	707,37	0,000
47822	58	0,896	-98,39	-2,25	0,028	2,38	22,00	0,000	484,21	0,000
47890	158	0,374	86,66	2,23	0,027	1,76	9,66	0,000	93,29	0,000
51111	2	-	-	-	-	-	-	-	-	-
56112	98	0,668	-4,72	-0,94	0,350	0,09	13,89	0,000	192,82	0,000

Fonte: Elaborado pelo autor

Primeiramente, pode-se notar que todos os modelos apresentaram coeficiente positivo, o que é coerente, uma vez que espera-se que quanto mais empregados mais notas fiscais sejam emitidas. Além disso, pode-se ver que em 19 dos 28 modelos não houve indícios para recusar a hipótese nula do teste-t para o intercepto e para todos os modelos a hipótese nula desse teste para o coeficiente angular foi rejeitada.

Além do número de empregados, testou-se o número de estabelecimentos como variável explicativa e os resultados estão na Tabela 16.

Tabela 16 - Resultados dos modelos tipo 2

$y = \beta_0 + \beta_1x$										
CNAE	Amostras	R ²	β_0	t-stat	P-Value	β_1	t-stat	P-Value	F	P-Value
10911	105	0,003	232,50	2,10	0,038	10,87	0,58	0,563	0,34	0,563
11216	30	0,001	922,31	1,33	0,195	38,23	0,16	0,875	0,03	0,875
14126	186	0,221	-9,80	-0,29	0,769	8,16	7,23	0,000	52,32	0,000
41204	67	0,128	11,49	1,93	0,058	0,18	3,08	0,003	9,51	0,003
45111	84	0,018	847,25	4,05	0,000	24,46	1,22	0,226	1,49	0,226
45307	175	0,229	-133,36	-0,48	0,629	109,57	7,18	0,000	51,49	0,000
45412	86	0,260	14,59	0,18	0,859	48,84	5,43	0,000	29,48	0,000
46346	35	0,000	793,29	2,43	0,021	7,58	0,06	0,956	0,00	0,956
46354	16	0,033	5680,79	1,69	0,112	-618,25	-0,69	0,500	0,48	0,500
46397	64	0,019	2797,67	2,93	0,005	201,67	1,11	0,271	1,23	0,271
46443	55	0,010	2448,83	2,38	0,021	-202,03	-0,71	0,479	0,51	0,479
46460	30	0,000	1294,90	1,81	0,081	-13,67	-0,07	0,947	0,00	0,947
46478	44	0,708	-608,80	-2,87	0,006	531,22	10,08	0,000	101,69	0,000
46494	65	0,824	-117,05	-1,17	0,247	181,32	17,18	0,000	295,22	0,000
46516	25	0,006	381,67	1,68	0,107	27,54	0,36	0,720	0,13	0,720
46796	61	0,038	234,61	0,93	0,357	147,32	1,52	0,134	2,31	0,134
47113	130	0,055	382,51	2,10	0,038	68,32	2,73	0,007	7,45	0,007
47121	130	0,120	-82,77	-1,26	0,210	6,92	4,18	0,000	17,46	0,000
47296	80	0,398	4,44	0,21	0,832	10,74	7,18	0,000	51,51	0,000
47318	112	0,252	13,78	1,48	0,142	11,40	6,09	0,000	37,04	0,000
47440	182	0,442	-225,69	-3,01	0,003	34,74	11,94	0,000	142,47	0,000
47512	75	0,555	72,66	1,14	0,259	39,14	9,55	0,000	91,16	0,000
47717	128	0,634	-81,91	-2,18	0,031	32,53	14,78	0,000	218,37	0,000
47741	35	0,965	-25,99	-3,07	0,004	4,75	30,22	0,000	913,15	0,000
47814	113	0,908	10,33	0,89	0,373	3,02	33,08	0,000	1094,33	0,000
47822	58	0,944	-114,00	-3,55	0,001	25,52	30,75	0,000	945,26	0,000
47890	158	0,356	71,16	1,78	0,078	7,11	9,29	0,000	86,35	0,000
51111	2	-	-	-	-	-	-	-	-	-
56112	98	0,469	-1,93	-0,29	0,770	0,69	9,21	0,000	84,83	0,000

Fonte: Elaborado pelo autor

Diferentemente dos resultados obtidos com número de empregados, o número de estabelecimentos como variável explicativa produz para os CNAEs 46354, 46443 e 46460 um coeficiente negativo, o que não é coerente, uma vez que um bairro com mais empresas deveria emitir mais notas fiscais. É interessante notar que para todos os resultados com coeficiente angular negativo não houve indícios de significância.

Ademais, para 10 CNAEs não houve indícios para rejeitar a hipótese nula do teste-t para o coeficiente angular e para 17 CNAEs não houve indícios para rejeitar a hipótese nula do teste-t para o intercepto.

Além dos modelos de emissão de notas fiscais, foram realizadas regressões para desenvolver modelos para a atração de notas fiscais. Na Tabela 17, podem ser encontrados os resultados para a atração de notas fiscais com número de empregados como variável explicativa.

Tabela 17 - Resultados dos modelos tipo 3

$y = \beta_0 + \beta_1x$										
CNAE	Amostras	R ²	β_0	t-stat	P-Value	β_1	t-stat	P-Value	F	P-Value
10911	174	0,691	26,70	1,89	0,061	3,03	19,62	0,000	384,84	0,000
11216	41	0,382	-16,57	-0,22	0,830	4,42	4,91	0,000	24,13	0,000
14126	213	0,081	69,37	6,67	0,000	0,08	4,30	0,000	18,52	0,000
41204	180	0,881	-2,65	-0,26	0,798	0,41	36,22	0,000	1311,86	0,000
45111	102	0,836	19,42	0,87	0,384	2,98	22,55	0,000	508,50	0,000
45307	197	0,836	-49,12	-0,91	0,364	12,94	31,56	0,000	996,21	0,000
45412	143	0,630	30,86	2,61	0,010	4,67	15,51	0,000	240,57	0,000
46346	42	0,261	-3,26	-0,20	0,843	1,71	3,76	0,001	14,13	0,001
46354	21	0,502	32,15	2,48	0,023	0,45	4,37	0,000	19,13	0,000
46397	76	0,189	101,13	2,98	0,004	0,74	4,16	0,000	17,30	0,000
46443	63	0,858	-70,90	-2,87	0,006	2,76	19,20	0,000	368,52	0,000
46460	30	0,133	20,61	0,91	0,369	0,78	2,07	0,048	4,29	0,048
46478	50	0,617	-125,36	-1,37	0,177	7,36	8,80	0,000	77,49	0,000
46494	76	0,799	-68,33	-3,26	0,002	5,44	17,15	0,000	294,06	0,000
46516	29	0,163	-3,18	-0,05	0,963	6,89	2,30	0,030	5,27	0,030
46796	65	0,056	30,92	1,09	0,278	1,27	1,93	0,058	3,72	0,058
47113	171	0,672	93,41	1,00	0,319	5,19	18,60	0,000	345,96	0,000
47121	234	0,489	217,10	5,13	0,000	6,86	14,90	0,000	222,04	0,000
47296	171	0,503	43,54	6,54	0,000	1,82	13,07	0,000	170,91	0,000
47318	163	0,725	-17,30	-1,40	0,162	4,07	20,59	0,000	423,79	0,000
47440	226	0,503	-2,76	-0,11	0,910	2,93	15,07	0,000	227,02	0,000
47512	109	0,574	20,88	0,97	0,333	3,66	12,02	0,000	144,46	0,000
47717	193	0,748	181,86	3,89	0,000	5,18	23,80	0,000	566,45	0,000
47741	130	0,925	-30,53	-2,42	0,017	4,19	39,86	0,000	1588,68	0,000
47814	207	0,937	-2,85	-0,41	0,685	1,13	55,22	0,000	3049,20	0,000
47822	75	0,907	14,58	0,60	0,552	1,82	26,61	0,000	708,22	0,000
47890	201	0,755	57,76	9,87	0,000	0,77	24,76	0,000	612,93	0,000
51111	6	0,002	5450,77	1,08	0,341	-1,13	-0,08	0,940	0,01	0,940
56112	227	0,933	-76,12	-3,34	0,001	2,45	55,78	0,000	3111,80	0,000

Fonte: Elaborado pelo autor

Para os modelos da tabela acima, têm-se apenas um coeficiente angular negativo, para o CNAE 51111, entretanto não pode-se rejeitar a hipótese nula do teste-t para o coeficiente angular do modelo. Nos modelos de atração de notas fiscais com número de

empregados como variável explicativa houve 17 modelos que não puderam ter a hipótese nula do teste-t rejeitada para o intercepto e 2 para o coeficiente angular.

Tabela 18 - Resultados dos modelos tipo 4

$y = \beta_0 + \beta_1 x$										
CNAE	Amostras	R ²	β_0	t-stat	P-Value	β_1	t-stat	P-Value	F	P-Value
10911	174	0,430	-57,24	-2,23	0,027	58,81	11,40	0,000	129,96	0,000
11216	41	0,015	263,72	2,33	0,025	-34,89	-0,78	0,441	0,61	0,441
14126	213	0,081	69,37	6,67	0,000	0,08	4,30	0,000	18,52	0,000
41204	180	0,921	-40,72	-4,68	0,000	6,16	45,70	0,000	2088,22	0,000
45111	102	0,164	140,19	2,52	0,013	25,86	4,44	0,000	19,68	0,000
45307	197	0,531	-184,35	-1,80	0,074	89,04	14,85	0,000	220,40	0,000
45412	143	0,708	-11,12	-0,96	0,339	29,23	18,51	0,000	342,45	0,000
46346	42	0,000	38,86	1,70	0,097	-0,09	-0,01	0,993	0,00	0,993
46354	21	0,007	51,58	2,16	0,044	2,70	0,38	0,711	0,14	0,711
46397	76	0,349	42,77	1,28	0,203	42,75	6,30	0,000	39,74	0,000
46443	63	0,000	146,13	1,80	0,077	-0,69	-0,03	0,977	0,00	0,977
46460	30	0,535	-5,64	-0,36	0,725	26,37	5,68	0,000	32,22	0,000
46478	50	0,662	-423,49	-4,11	0,000	256,60	9,70	0,000	94,00	0,000
46494	76	0,958	-84,98	-8,82	0,000	44,76	40,94	0,000	1676,11	0,000
46516	29	0,023	40,12	0,38	0,706	28,17	0,80	0,429	0,65	0,429
46796	65	0,111	-50,52	-1,12	0,269	49,54	2,81	0,007	7,87	0,007
47113	171	0,411	170,56	1,25	0,212	223,48	10,85	0,000	117,77	0,000
47121	234	0,489	217,10	5,13	0,000	6,86	14,90	0,000	222,04	0,000
47296	171	0,584	22,23	3,32	0,001	10,43	15,40	0,000	237,21	0,000
47318	163	0,671	-47,88	-3,21	0,002	63,16	18,14	0,000	328,95	0,000
47440	226	0,503	-2,76	-0,11	0,910	2,93	15,07	0,000	227,02	0,000
47512	109	0,693	7,81	0,43	0,669	21,79	15,53	0,000	241,11	0,000
47717	193	0,903	-105,52	-3,33	0,001	95,34	42,20	0,000	1780,59	0,000
47741	130	0,947	-56,26	-5,21	0,000	18,26	47,81	0,000	2286,24	0,000
47814	207	0,937	-2,85	-0,41	0,685	1,13	55,22	0,000	3049,20	0,000
47822	75	0,929	-5,83	-0,27	0,787	19,46	30,81	0,000	949,39	0,000
47890	201	0,755	57,76	9,87	0,000	0,77	24,76	0,000	612,93	0,000
51111	6	0,190	8380,70	1,88	0,134	-1213,70	-0,97	0,388	0,94	0,388
56112	227	0,933	-76,12	-3,34	0,001	2,45	55,78	0,000	3111,80	0,000

Fonte: Elaborado pelo autor

Nos resultados para os modelos de atração de notas fiscais que utilizaram o número de estabelecimentos como variável explicativa (Tabela 18), pode-se ver que em 15 modelos não houve indícios de significância do intercepto e em 6 modelos não houve indícios para rejeitar a hipótese nula do teste-t para o coeficiente angular. 4 CNAEs tiveram modelos

com coeficiente angular negativo, com destaque para o modelo do CNAE 51111, o que não é coerente com o fenômeno.

6.2.1.2 Normalidade dos erros

Como dito anteriormente, os testes estatísticos utilizados para avaliar os modelos de regressão requerem que os erros dos modelos sejam normalmente distribuídos. Para checar essa premissa foi realizado o teste de Shapiro-Wilk.

Para os modelos de produção de notas fiscais com número de empregados como variável independente, apenas para os CNAEs 46346, 46354 e 46460 tiveram erros distribuídos normalmente. Como mostra a Tabela 19.

Tabela 19 - Testes de normalidade para os modelos tipo 1

CNAE	Estatística W (Shapiro-Wilk)	P-Valor (Shapiro-Wilk)
10911	0,729	0,000
11216	0,816	0,000
14126	0,504	0,000
41204	0,611	0,000
45111	0,807	0,000
45307	0,585	0,000
45412	0,612	0,000
46346	0,974	0,566
46354	0,962	0,693
46397	0,610	0,000
46443	0,762	0,000
46460	0,936	0,070
46478	0,746	0,000
46494	0,891	0,000
46516	0,882	0,008
46796	0,867	0,000
47113	0,563	0,000
47121	0,364	0,000
47296	0,598	0,000
47318	0,799	0,000
47440	0,686	0,000
47512	0,797	0,000
47717	0,547	0,000
47741	0,930	0,029
47814	0,884	0,000
47822	0,843	0,000
47890	0,367	0,000
51111	-	-
56112	0,861	0,000

Fonte: Elaborado pelo autor

Dentre os modelos de produção com número estabelecimentos como variável explicativa (Tabela 20) somente o CNAE 46478 teve erros normalmente distribuídos segundo o teste utilizado.

Tabela 20 - Testes de normalidade para os modelos tipo 2

CNAE	Estatística W (Shapiro-Wilk)	P-Valor (Shapiro-Wilk)
10911	0,438	0,000
11216	0,431	0,000
14126	0,633	0,000
41204	0,612	0,000
45111	0,799	0,000
45307	0,429	0,000
45412	0,476	0,000
46346	0,713	0,000
46354	0,593	0,000
46397	0,623	0,000
46443	0,393	0,000
46460	0,467	0,000
46478	0,950	0,056
46494	0,756	0,000
46516	0,842	0,001
46796	0,676	0,000
47113	0,535	0,000
47121	0,428	0,000
47296	0,556	0,000
47318	0,770	0,000
47440	0,699	0,000
47512	0,794	0,000
47717	0,613	0,000
47741	0,912	0,008
47814	0,804	0,000
47822	0,929	0,002
47890	0,360	0,000
51111	-	-
56112	0,688	0,000

Fonte: Elaborado pelo autor

Para os modelos de atração com número de empregados como variável independente (Tabela 21), nenhum dos modelos construídos teve erros considerados normalmente distribuídos pelo teste de Shapiro-Wilk.

Tabela 21 - Testes de normalidade para os modelos tipo 3

CNAE	Estatística W (Shapiro-Wilk)	P-Valor (Shapiro-Wilk)
10911	0,740	0,000
11216	0,588	0,000
14126	0,590	0,000
41204	0,693	0,000
45111	0,917	0,000
45307	0,756	0,000
45412	0,718	0,000
46346	0,645	0,000
46354	0,721	0,000
46397	0,557	0,000
46443	0,750	0,000
46460	0,736	0,000
46478	0,741	0,000
46494	0,748	0,000
46516	0,636	0,000
46796	0,252	0,000
47113	0,877	0,000
47121	0,796	0,000
47296	0,787	0,000
47318	0,930	0,000
47440	0,589	0,000
47512	0,722	0,000
47717	0,604	0,000
47741	0,552	0,000
47814	0,717	0,000
47822	0,821	0,000
47890	0,778	0,000
51111	0,762	0,026
56112	0,688	0,000

Fonte: Elaborado pelo autor

Já para os modelos de atração de notas fiscais criados com o número de estabelecimentos como variável independente, apenas o modelo do CNAE 51111 teve erros distribuídos normalmente, conforme apresentado na Tabela 22.

Tabela 22 - Testes de normalidade para os modelos tipo 4

CNAE	Estatística W (Shapiro-Wilk)	P-Valor (Shapiro-Wilk)
10911	0,851	0,000
11216	0,508	0,000
14126	0,590	0,000
41204	0,863	0,000
45111	0,756	0,000
45307	0,512	0,000
45412	0,816	0,000
46346	0,407	0,000
46354	0,688	0,000
46397	0,595	0,000
46443	0,280	0,000
46460	0,797	0,000
46478	0,887	0,000
46494	0,965	0,033
46516	0,529	0,000
46796	0,414	0,000
47113	0,880	0,000
47121	0,796	0,000
47296	0,748	0,000
47318	0,883	0,000
47440	0,589	0,000
47512	0,783	0,000
47717	0,862	0,000
47741	0,648	0,000
47814	0,717	0,000
47822	0,860	0,000
47890	0,778	0,000
51111	0,907	0,416
56112	0,688	0,000

Fonte: Elaborado pelo autor

6.2.1.3 Homoscedasticidade e autocorrelação

Como já mencionado, para estimar os parâmetros dos modelos de regressão linear é necessário supor que os erros sejam homocedásticos e não sejam correlacionados. Dessa forma, foi utilizado o teste de Breusch-Pagan para verificar a homoscedasticidade e o teste de Durbin-Watson para verificar a autocorrelação.

Os resultados para os modelos de emissão de notas fiscais com número de empregados como variável explicativa estão apresentados na Tabela 23. Pode-se ver que 19

modelos não passaram no teste de homoscedasticidade, que 6 não passaram no teste de autocorrelação e que 1 não teve resultados conclusivos para o teste de correlação.

Tabela 23 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 1

CNAE	Estatística BP (Breusch-Pagan)	P-Valor (Breusch-Pagan)	Durbin Watson	dL	dU	4-dU	4-dL
10911	37,325	0,000	2,121	1,663	1,701	2,299	2,337
11216	8,894	0,003	1,363	1,249	1,383	2,617	2,751
14126	6,824	0,009	1,953	1,749	1,771	2,229	2,251
41204	2,836	0,092	1,885	1,500	1,560	2,440	2,500
45111	3,872	0,049	1,941	1,555	1,603	2,397	2,445
45307	90,977	0,000	2,191	1,741	1,764	2,236	2,259
45412	24,160	0,000	1,998	1,560	1,607	2,393	2,440
46346	1,186	0,276	1,960	1,305	1,420	2,580	2,695
46354	7,898	0,005	2,442	0,980	1,235	2,765	3,020
46397	0,269	0,604	1,393	1,488	1,551	2,449	2,512
46443	17,746	0,000	2,450	1,447	1,520	2,480	2,553
46460	21,939	0,000	2,209	1,249	1,383	2,617	2,751
46478	15,709	0,000	2,916	1,381	1,472	2,528	2,619
46494	4,490	0,034	2,108	1,492	1,554	2,446	2,508
46516	0,161	0,688	2,619	1,178	1,339	2,661	2,822
46796	0,338	0,561	1,205	1,476	1,541	2,459	2,524
47113	6,096	0,014	2,104	1,698	1,729	2,271	2,302
47121	4,631	0,031	2,210	1,698	1,729	2,271	2,302
47296	2,983	0,084	2,255	1,544	1,594	2,406	2,456
47318	2,923	0,087	1,720	1,674	1,710	2,290	2,326
47440	111,984	0,000	1,950	1,746	1,768	2,232	2,254
47512	25,126	0,000	2,526	1,528	1,582	2,418	2,472
47717	30,128	0,000	1,859	1,696	1,727	2,273	2,304
47741	0,568	0,451	1,868	1,305	1,420	2,580	2,695
47814	17,955	0,000	1,526	1,675	1,711	2,289	2,325
47822	10,838	0,001	2,583	1,462	1,531	2,469	2,538
47890	0,834	0,361	2,072	1,727	1,753	2,247	2,273
51111	-	-	-	-	-	-	-
56112	37,347	0,000	1,999	1,589	1,630	2,371	2,411

Fonte: Elaborado pelo autor

Dentre os modelos de emissão de notas fiscais com número de empresas como variável explicativa, apenas 9 modelos não passaram no teste de homoscedasticidade, 3 não passaram no teste de autocorrelação e que 1 não teve resultados conclusivos para o teste de autocorrelação, como mostra a Tabela 24.

Tabela 24 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 2

CNAE	Estatística BP (Breusch-Pagan)	P-Valor (Breusch-Pagan)	Durbin Watson	dL	dU	4-dU	4-dL
10911	0,140	0,708	1,923	1,663	1,701	2,299	2,337
11216	0,242	0,622	1,670	1,249	1,383	2,617	2,751
14126	10,868	0,001	1,936	1,749	1,771	2,229	2,251
41204	2,675	0,102	1,871	1,500	1,560	2,440	2,500
45111	1,839	0,175	1,513	1,555	1,603	2,397	2,445
45307	13,614	0,000	2,104	1,741	1,764	2,236	2,259
45412	1,519	0,218	2,089	1,560	1,607	2,393	2,440
46346	0,002	0,968	2,149	1,305	1,420	2,580	2,695
46354	0,499	0,480	1,977	0,980	1,235	2,765	3,020
46397	0,057	0,812	1,600	1,488	1,551	2,449	2,512
46443	0,857	0,355	1,905	1,447	1,520	2,480	2,553
46460	0,532	0,466	2,247	1,249	1,383	2,617	2,751
46478	17,368	0,000	2,150	1,381	1,472	2,528	2,619
46494	0,003	0,954	1,938	1,492	1,554	2,446	2,508
46516	0,006	0,939	2,792	1,178	1,339	2,661	2,822
46796	2,115	0,146	1,890	1,476	1,541	2,459	2,524
47113	0,320	0,572	2,265	1,698	1,729	2,271	2,302
47121	10,116	0,001	2,332	1,698	1,729	2,271	2,302
47296	0,008	0,931	2,206	1,544	1,594	2,406	2,456
47318	1,301	0,254	1,513	1,674	1,710	2,290	2,326
47440	7,487	0,006	1,798	1,746	1,768	2,232	2,254
47512	9,481	0,002	2,142	1,528	1,582	2,418	2,472
47717	17,561	0,000	1,852	1,696	1,727	2,273	2,304
47741	0,241	0,624	1,718	1,305	1,420	2,580	2,695
47814	3,677	0,055	1,929	1,675	1,711	2,289	2,325
47822	16,477	0,000	2,205	1,462	1,531	2,469	2,538
47890	0,367	0,545	2,089	1,727	1,753	2,247	2,273
51111	-	-	-	-	-	-	-
56112	52,155	0,000	2,095	1,589	1,630	2,371	2,411

Fonte: Elaborado pelo autor

Para os modelos que explicam a atração de notas fiscais utilizando como variável independente o número de empregos, os resultados, Tabela 25, nos mostram que 22 modelos não passaram no teste de homoscedasticidade, 4 modelos não passaram no teste de autocorrelação e que 2 modelos não tiveram resultados conclusivos para esse teste.

Tabela 25 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 3

CNAE	Estatística BP (Breusch-Pagan)	P-Valor (Breusch-Pagan)	Durbin Watson	dL	dU	4-dU	4-dL
10911	55,727	0,000	1,960	1,740	1,764	2,236	2,260
11216	0,803	0,370	2,135	1,358	1,456	2,544	2,642
14126	18,957	0,000	2,060	1,724	1,742	2,258	2,276
41204	62,153	0,000	1,948	1,745	1,767	2,233	2,255
45111	43,881	0,000	1,795	1,658	1,697	2,303	2,342
45307	106,143	0,000	1,613	1,757	1,777	2,223	2,243
45412	40,288	0,000	2,024	1,713	1,741	2,259	2,287
46346	11,034	0,001	1,678	1,366	1,462	2,538	2,634
46354	0,001	0,970	2,472	1,104	1,297	2,703	2,896
46397	0,002	0,968	2,116	1,531	1,584	2,416	2,469
46443	28,017	0,000	2,490	1,484	1,548	2,452	2,516
46460	4,777	0,029	2,289	1,249	1,383	2,617	2,751
46478	26,687	0,000	2,875	1,420	1,500	2,500	2,580
46494	45,331	0,000	1,671	1,531	1,584	2,416	2,469
46516	3,149	0,076	1,996	1,236	1,375	2,625	2,764
46796	0,770	0,380	2,020	1,492	1,554	2,446	2,508
47113	24,455	0,000	2,125	1,738	1,762	2,238	2,262
47121	5,576	0,018	1,812	1,736	1,754	2,246	2,264
47296	11,143	0,001	2,060	1,738	1,762	2,238	2,262
47318	73,037	0,000	2,478	1,732	1,756	2,244	2,269
47440	100,929	0,000	1,961	1,732	1,749	2,251	2,268
47512	36,569	0,000	2,139	1,669	1,706	2,294	2,331
47717	97,843	0,000	1,926	1,754	1,775	2,225	2,246
47741	93,995	0,000	2,096	1,698	1,729	2,271	2,302
47814	18,281	0,000	2,135	1,719	1,739	2,261	2,281
47822	2,981	0,084	2,340	1,528	1,582	2,418	2,472
47890	10,406	0,001	1,719	1,715	1,735	2,265	2,285
51111	0,444	0,505	2,494	0,489	1,258	2,742	3,511
56112	136,979	0,000	2,274	1,732	1,750	2,250	2,268

Fonte: Elaborado pelo autor

Finalmente, os resultados para os modelos de atração de nota fiscal do mês de dezembro com número de estabelecimentos como variável explicativa estão mostrados na Tabela 26. 21 modelos não tiveram erros homocedásticos pelo teste utilizado, 4 não passaram no teste de autocorrelação e 3 tiveram resultado inconclusivo para esse teste.

Tabela 26 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 4

CNAE	Estatística BP (Breusch-Pagan)	P-Valor (Breusch-Pagan)	Durbin Watson	dL	dU	4-dU	4-dL
10911	51,145	0,000	1,692	1,740	1,764	2,236	2,260
11216	0,687	0,407	2,091	1,358	1,456	2,544	2,642
14126	18,957	0,000	2,060	1,724	1,742	2,258	2,276
41204	58,104	0,000	1,902	1,745	1,767	2,233	2,255
45111	8,016	0,005	1,485	1,658	1,697	2,303	2,342
45307	32,553	0,000	2,009	1,757	1,777	2,223	2,243
45412	19,772	0,000	1,989	1,713	1,741	2,259	2,287
46346	0,031	0,861	1,984	1,366	1,462	2,538	2,634
46354	0,537	0,464	2,516	1,104	1,297	2,703	2,896
46397	0,022	0,883	1,848	1,531	1,584	2,416	2,469
46443	0,643	0,422	1,943	1,484	1,548	2,452	2,516
46460	0,313	0,576	2,324	1,249	1,383	2,617	2,751
46478	42,674	0,000	2,634	1,420	1,500	2,500	2,580
46494	30,950	0,000	1,990	1,531	1,584	2,416	2,469
46516	0,196	0,658	2,153	1,236	1,375	2,625	2,764
46796	5,965	0,015	1,963	1,492	1,554	2,446	2,508
47113	22,367	0,000	2,008	1,738	1,762	2,238	2,262
47121	5,576	0,018	1,812	1,736	1,754	2,246	2,264
47296	7,324	0,007	1,970	1,738	1,762	2,238	2,262
47318	20,353	0,000	2,182	1,732	1,756	2,244	2,269
47440	100,929	0,000	1,961	1,732	1,749	2,251	2,268
47512	51,845	0,000	1,907	1,669	1,706	2,294	2,331
47717	108,095	0,000	1,798	1,754	1,775	2,225	2,246
47741	115,265	0,000	2,162	1,698	1,729	2,271	2,302
47814	18,281	0,000	2,135	1,719	1,739	2,261	2,281
47822	0,708	0,400	1,580	1,528	1,582	2,418	2,472
47890	10,406	0,001	1,719	1,715	1,735	2,265	2,285
51111	4,728	0,030	2,830	0,489	1,258	2,742	3,511
56112	136,979	0,000	2,274	1,732	1,750	2,250	2,268

Fonte: Elaborado pelo autor

6.2.2 Notas fiscais dos anos de 2014 e 2015

Todos os modelos desenvolvidos nessa seção tiveram como variável dependente número de notas fiscais emitidas ou atraídas por bairro nos anos de 2014 e 2015. Da mesma forma que foi realizado anteriormente, as premissas de normalidade, homoscedasticidade e independência dos erros foram verificadas depois da apresentação geral dos modelos.

6.2.2.1 Modelos

Os resultados das regressões realizadas para explicar o fenômeno da emissão de notas fiscais dos anos de 2014 e 2015, para os CNAEs mais emitiram e mais atraíram notas fiscais no ano de 2015, utilizando número de empregos como variável explicativa estão apresentados na Tabela 27.

Tabela 27 - Resultados dos modelos tipo 5

$y = \beta_0 + \beta_1x$										
CNAE	Amostras	R ²	β_0	t-stat	P-Value	β_1	t-stat	P-Value	F	P-Value
10911	117	0,362	-208,87	-0,31	0,757	49,38	8,07	0,000	65,13	0,000
11216	34	0,168	-322,78	-0,05	0,964	217,04	2,55	0,016	6,48	0,016
14126	202	0,065	1261,81	5,41	0,000	1,42	3,72	0,000	13,87	0,000
41204	102	0,198	74,16	1,83	0,070	0,17	4,96	0,000	24,61	0,000
45111	90	0,640	1835,51	1,50	0,137	85,52	12,51	0,000	156,50	0,000
45307	185	0,619	-3101,34	-1,68	0,095	234,46	17,25	0,000	297,70	0,000
45412	105	0,589	-643,83	-1,13	0,260	152,57	12,14	0,000	147,40	0,000
46346	38	0,802	-2801,85	-2,22	0,033	484,23	12,09	0,000	146,12	0,000
46354	22	0,961	-7985,67	-2,02	0,057	710,52	22,30	0,000	497,14	0,000
46397	69	0,419	11190,84	1,67	0,100	234,42	6,95	0,000	48,31	0,000
46443	59	0,741	-2991,48	-0,68	0,499	318,13	12,78	0,000	163,37	0,000
46460	34	0,678	-5260,59	-1,52	0,137	526,05	8,21	0,000	67,47	0,000
46478	49	0,568	473,87	0,22	0,829	155,07	7,86	0,000	61,71	0,000
46494	72	0,929	-1928,28	-2,83	0,006	305,05	30,32	0,000	919,32	0,000
46516	28	0,184	1550,72	0,99	0,331	168,62	2,42	0,023	5,88	0,023
46796	66	0,715	681,47	0,77	0,445	264,86	12,68	0,000	160,67	0,000
47113	137	0,209	962,37	0,62	0,534	24,95	5,96	0,000	35,57	0,000
47121	164	0,123	-340,93	-0,74	0,460	20,32	4,78	0,000	22,81	0,000
47296	102	0,178	310,47	1,63	0,106	14,82	4,66	0,000	21,72	0,000
47318	125	0,406	73,03	0,90	0,371	10,59	9,17	0,000	84,03	0,000
47440	198	0,537	-1983,06	-2,83	0,005	79,05	15,06	0,000	226,90	0,000
47512	88	0,649	161,85	0,27	0,787	101,29	12,60	0,000	158,71	0,000
47717	152	0,564	321,95	0,96	0,336	19,26	13,94	0,000	194,40	0,000
47741	56	0,923	-292,47	-3,19	0,002	12,80	25,48	0,000	649,00	0,000
47814	138	0,838	-35,22	-0,29	0,771	7,58	26,53	0,000	704,05	0,000
47822	69	0,897	-548,85	-1,98	0,052	18,01	24,10	0,000	580,58	0,000
47890	169	0,374	918,90	2,06	0,041	21,67	9,99	0,000	99,89	0,000
51111	4	0,980	266176,19	12,92	0,006	-465,86	-9,86	0,010	97,18	0,010
56112	129	0,600	-25,76	-0,66	0,511	0,78	13,80	0,000	190,42	0,000

Fonte: Elaborado pelo autor

Apenas o CNAE 51111 apresentou coeficiente angular negativo, isso não condiz com o esperado pois espera-se que dentre de um mesmo CNAE quanto maior o número de empregados maior o número de notas emitidas. Ademais, para 22 dos 29 modelos não houve indícios para recusar a hipótese nula do teste-t para o intercepto e para todos os modelos a hipótese nula desse teste para o coeficiente angular foi rejeitada.

Na Tabela 28, estão apresentados os modelos que utilizaram o número de estabelecimentos para explicar a emissão de notas fiscais.

Tabela 28 - Resultados dos modelos tipo 6

$y = \beta_0 + \beta_1x$										
CNAE	Amostras	R ²	β_0	t-stat	P-Value	β_1	t-stat	P-Value	F	P-Value
10911	117	0,003	2205,92	1,97	0,052	117,46	0,60	0,552	0,36	0,552
11216	34	0,017	15628,52	1,75	0,090	-2415,67	-0,74	0,463	0,55	0,463
14126	202	0,298	-238,73	-0,87	0,383	88,36	9,21	0,000	84,82	0,000
41204	102	0,209	55,24	1,32	0,190	2,54	5,15	0,000	26,49	0,000
45111	90	0,022	8830,01	3,95	0,000	312,43	1,41	0,161	2,00	0,161
45307	185	0,246	-1777,23	-0,61	0,544	1279,37	7,72	0,000	59,62	0,000
45412	105	0,255	-220,91	-0,26	0,793	593,48	5,94	0,000	35,29	0,000
46346	38	0,002	6995,34	2,16	0,037	408,19	0,30	0,769	0,09	0,769
46354	22	0,012	42400,80	1,56	0,135	-4096,83	-0,50	0,624	0,25	0,624
46397	69	0,029	26535,13	2,80	0,007	2588,79	1,40	0,165	1,97	0,165
46443	59	0,006	26582,93	2,46	0,017	-1782,09	-0,58	0,564	0,34	0,564
46460	34	0,000	11538,81	1,76	0,087	-5,64	0,00	0,998	0,00	0,998
46478	49	0,703	-6954,82	-3,22	0,002	5867,56	10,55	0,000	111,26	0,000
46494	72	0,803	-1436,31	-1,26	0,212	2129,14	16,87	0,000	284,60	0,000
46516	28	0,007	3448,51	1,42	0,169	355,41	0,43	0,670	0,19	0,670
46796	66	0,054	1578,97	0,59	0,557	2005,76	1,91	0,061	3,64	0,061
47113	137	0,064	3378,21	1,89	0,061	758,46	3,03	0,003	9,15	0,003
47121	164	0,118	-767,79	-1,43	0,154	67,33	4,65	0,000	21,64	0,000
47296	102	0,322	-77,93	-0,40	0,686	105,94	6,89	0,000	47,46	0,000
47318	125	0,303	66,27	0,69	0,490	147,53	7,32	0,000	53,55	0,000
47440	198	0,476	-3300,64	-4,01	0,000	440,34	13,33	0,000	177,76	0,000
47512	88	0,605	344,07	0,54	0,588	502,99	11,48	0,000	131,83	0,000
47717	152	0,649	-814,18	-2,47	0,015	348,75	16,66	0,000	277,41	0,000
47741	56	0,941	-315,39	-3,91	0,000	55,33	29,31	0,000	858,87	0,000
47814	138	0,894	38,55	0,40	0,690	28,36	33,82	0,000	1143,99	0,000
47822	69	0,926	-721,96	-3,05	0,003	192,46	28,86	0,000	832,86	0,000
47890	169	0,356	730,05	1,58	0,115	87,35	9,61	0,000	92,31	0,000
51111	4	0,988	189073,50	17,60	0,003	-31511,50	-12,62	0,006	159,23	0,006
56112	129	0,487	-31,66	-0,69	0,492	6,56	10,97	0,000	120,34	0,000

Fonte: Elaborado pelo autor

Ao utilizar-se o número de estabelecimentos como variável explicativa foram produzidos para os CNAEs 11216, 46354, 46443, 46460 e 51111 um coeficiente negativo, da mesma forma do número de empregados, espera-se que o número de notas cresça acompanhando o número de estabelecimentos. Dentre os CNAEs que tiveram modelos com β_1 negativo, apenas para o modelo do CNAE 51111 houve indícios de que o coeficiente angular era significativo.

Para 10 dos 29 modelos produzidos não houve indícios para rejeitar a hipótese nula do teste-t para o coeficiente angular e para 19 modelos não houve indícios para rejeitar a hipótese nula do teste-t para o intercepto.

Além dos modelos de emissão de notas fiscais, foram realizadas regressões para desenvolver modelos para a atração de notas fiscais dos anos de 2014 e 2015.

Tabela 29 - Resultados dos modelos tipo 7

$y = \beta_0 + \beta_1 x$										
CNAE	Amostras	R ²	β_0	t-stat	P-Value	β_1	t-stat	P-Value	F	P-Value
10911	176	0,696	285,54	1,81	0,072	34,59	19,95	0,000	398,17	0,000
11216	45	0,380	-393,56	-0,43	0,672	58,05	5,14	0,000	26,40	0,000
14126	217	0,104	922,10	6,82	0,000	1,14	4,99	0,000	24,87	0,000
41204	193	0,867	-70,99	-0,50	0,620	5,78	35,25	0,000	1242,50	0,000
45111	107	0,840	167,72	0,67	0,505	35,89	23,51	0,000	552,80	0,000
45307	197	0,857	-505,12	-0,88	0,381	149,15	34,17	0,000	1167,55	0,000
45412	143	0,704	340,44	2,88	0,005	55,14	18,31	0,000	335,42	0,000
46346	50	0,335	-106,00	-0,79	0,431	19,70	4,91	0,000	24,13	0,000
46354	22	0,178	501,48	1,96	0,064	4,29	2,08	0,0502	4,34	0,0502
46397	83	0,228	1007,21	2,92	0,004	9,27	4,89	0,000	23,87	0,000
46443	68	0,793	-615,20	-2,01	0,048	29,36	15,88	0,000	252,17	0,000
46460	46	0,211	100,73	0,83	0,409	8,53	3,43	0,001	11,74	0,001
46478	57	0,490	-653,06	-0,72	0,475	64,49	7,28	0,000	52,95	0,000
46494	90	0,419	-212,73	-0,58	0,564	36,73	7,96	0,000	63,35	0,000
46516	35	0,205	-207,21	-0,37	0,717	73,68	2,91	0,006	8,49	0,006
46796	77	0,042	406,44	1,05	0,297	17,56	1,81	0,074	3,29	0,074
47113	177	0,626	1121,31	1,07	0,288	54,35	17,12	0,000	293,02	0,000
47121	234	0,495	2665,94	5,41	0,000	80,91	15,08	0,000	227,33	0,000
47296	172	0,540	444,21	5,75	0,000	22,90	14,11	0,000	199,20	0,000
47318	171	0,730	-200,02	-1,53	0,127	45,76	21,39	0,000	457,67	0,000
47440	229	0,485	-12,87	-0,05	0,964	32,79	14,61	0,000	213,42	0,000
47512	118	0,604	217,12	0,95	0,345	44,86	13,30	0,000	176,76	0,000
47717	196	0,746	2236,02	4,06	0,000	61,70	23,87	0,000	569,96	0,000
47741	146	0,931	-383,21	-2,80	0,006	53,27	44,05	0,000	1940,20	0,000
47814	207	0,916	-6,74	-0,09	0,929	10,35	47,34	0,000	2241,39	0,000
47822	101	0,917	55,51	0,39	0,699	15,37	32,97	0,000	1086,98	0,000
47890	203	0,783	650,80	10,04	0,000	9,30	26,97	0,000	727,13	0,000
51111	6	0,002	56961,15	1,09	0,338	-12,65	-0,09	0,936	0,01	0,936
56112	227	0,920	-852,45	-3,13	0,002	26,57	50,81	0,000	2581,74	0,000

Fonte: Elaborado pelo autor

Os resultados para a atração de notas fiscais com número de empregados como variável independente estão na Tabela 29. Para esse conjunto de modelos, o modelo do CNAE 51111 foi o único que teve coeficiente angular negativo, mas não há indícios para rejeitar a hipótese nula do teste-t para o coeficiente angular. Além disso, houve 19 modelos que não puderam ter a hipótese nula do teste-t rejeitada para o intercepto e 3 para o coeficiente angular.

Tabela 30 - Resultados dos modelos tipo 8

$y = \beta_0 + \beta_1x$										
CNAE	Amostras	R ²	β_0	t-stat	P-Value	β_1	t-stat	P-Value	F	P-Value
10911	176	0,421	-625,54	-2,16	0,032	659,57	11,25	0,000	126,57	0,000
11216	45	0,012	2980,88	2,15	0,037	-407,77	-0,72	0,475	0,52	0,475
14126	217	0,279	130,93	0,82	0,415	53,14	9,11	0,000	83,03	0,000
41204	193	0,915	-584,84	-4,94	0,000	86,23	45,39	0,000	2060,38	0,000
45111	107	0,170	1517,27	2,39	0,018	315,20	4,64	0,000	21,53	0,000
45307	197	0,564	-2262,69	-2,01	0,046	1044,79	15,88	0,000	252,02	0,000
45412	143	0,718	-83,36	-0,65	0,514	329,10	18,95	0,000	359,08	0,000
46346	50	0,000	320,23	1,62	0,111	10,67	0,12	0,906	0,01	0,906
46354	22	0,001	776,28	2,08	0,050	-14,00	-0,12	0,903	0,02	0,903
46397	83	0,417	274,05	0,83	0,411	532,93	7,61	0,000	57,88	0,000
46443	68	0,001	1405,58	1,69	0,095	62,18	0,25	0,805	0,06	0,805
46460	46	0,722	-197,04	-2,55	0,014	289,09	10,70	0,000	114,54	0,000
46478	57	0,604	-3653,02	-3,78	0,000	2401,98	9,16	0,000	83,82	0,000
46494	90	0,962	-1070,72	-11,14	0,000	558,11	47,21	0,000	2228,98	0,000
46516	35	0,035	320,29	0,41	0,684	283,61	1,10	0,280	1,21	0,280
46796	77	0,118	-882,38	-1,45	0,151	794,40	3,16	0,002	10,01	0,002
47113	177	0,410	1461,24	1,02	0,311	2434,74	11,04	0,000	121,79	0,000
47121	234	0,569	672,02	1,29	0,199	285,18	17,49	0,000	305,85	0,000
47296	172	0,643	167,72	2,24	0,026	132,69	17,49	0,000	305,93	0,000
47318	171	0,685	-566,75	-3,62	0,000	716,29	19,17	0,000	367,59	0,000
47440	229	0,366	-286,81	-0,84	0,401	167,87	11,45	0,000	131,10	0,000
47512	118	0,664	109,45	0,52	0,607	256,66	15,15	0,000	229,58	0,000
47717	196	0,882	-1052,87	-2,57	0,011	1121,84	38,16	0,000	1455,82	0,000
47741	146	0,954	-692,43	-6,10	0,000	232,14	54,61	0,000	2982,21	0,000
47814	207	0,914	34,54	0,45	0,650	37,78	46,71	0,000	2181,51	0,000
47822	101	0,930	-150,64	-1,14	0,259	163,74	36,33	0,000	1319,77	0,000
47890	203	0,819	543,68	9,02	0,000	39,28	30,19	0,000	911,33	0,000
51111	6	0,193	87454,30	1,89	0,132	-12717,30	-0,98	0,384	0,96	0,384
56112	227	0,760	-1377,98	-2,82	0,005	222,90	26,69	0,000	712,20	0,000

Fonte: Elaborado pelo autor

Nos resultados para os modelos de atração de notas fiscais que utilizaram o número de estabelecimentos como variável explicativa, Tabela 30, pode-se ver que em 15 modelos não houve indícios de significância do intercepto e em 6 modelos não houve indícios de significância do coeficiente angular. 3 CNAEs tiveram modelos com coeficiente angular negativo, o que não é esperado.

6.2.2.2 Normalidade dos erros

Da mesma forma como foi realizado para os modelos anteriores, para checar a normalidade dos erros foi utilizado o teste de Shapiro-Wilk.

Dentre os modelos de produção de notas fiscais com número de empregados como variável independente, apenas para os CNAEs 46346 e 51111 tiveram erros distribuídos normalmente. Como apresentado na Tabela 31.

Tabela 31 - Testes de normalidade para os modelos tipo 5

CNAE	Estatística W (Shapiro-Wilk)	P-Valor (Shapiro-Wilk)
10911	0,690	0,000
11216	0,605	0,000
14126	0,553	0,000
41204	0,515	0,000
45111	0,850	0,000
45307	0,593	0,000
45412	0,570	0,000
46346	0,977	0,605
46354	0,846	0,003
46397	0,613	0,000
46443	0,709	0,000
46460	0,908	0,008
46478	0,751	0,000
46494	0,886	0,000
46516	0,812	0,000
46796	0,858	0,000
47113	0,582	0,000
47121	0,357	0,000
47296	0,453	0,000
47318	0,808	0,000
47440	0,654	0,000
47512	0,768	0,000
47717	0,567	0,000
47741	0,836	0,000
47814	0,808	0,000
47822	0,817	0,000
47890	0,351	0,000
51111	0,997	0,989
56112	0,703	0,000

Fonte: Elaborado pelo autor

Para os modelos de produção com número estabelecimentos como variável explicativa, conforme apresentado na Tabela 32, somente o CNAE 51111 teve erros normalmente distribuídos segundo o teste utilizado.

Tabela 32 - Testes de normalidade para os modelos tipo 6

CNAE	Estatística W (Shapiro-Wilk)	P-Valor (Shapiro-Wilk)
10911	0,386	0,000
11216	0,399	0,000
14126	0,752	0,000
41204	0,511	0,000
45111	0,796	0,000
45307	0,438	0,000
45412	0,445	0,000
46346	0,686	0,000
46354	0,490	0,000
46397	0,635	0,000
46443	0,426	0,000
46460	0,425	0,000
46478	0,943	0,020
46494	0,695	0,000
46516	0,761	0,000
46796	0,663	0,000
47113	0,544	0,000
47121	0,420	0,000
47296	0,416	0,000
47318	0,783	0,000
47440	0,712	0,000
47512	0,770	0,000
47717	0,632	0,000
47741	0,870	0,000
47814	0,719	0,000
47822	0,847	0,000
47890	0,346	0,000
51111	0,945	0,684
56112	0,624	0,000

Fonte: Elaborado pelo autor

Já para os modelos de atração com número de empregados como variável independente, Tabela 33, nenhum dos modelos construídos teve erros considerados normalmente distribuídos pelo teste de Shapiro-Wilk.

Tabela 33 - Testes de normalidade para os modelos tipo 7

CNAE	Estatística W (Shapiro-Wilk)	P-Valor (Shapiro-Wilk)
10911	0,739	0,000
11216	0,539	0,000
14126	0,587	0,000
41204	0,625	0,000
45111	0,910	0,000
45307	0,754	0,000
45412	0,770	0,000
46346	0,669	0,000
46354	0,551	0,000
46397	0,566	0,000
46443	0,700	0,000
46460	0,694	0,000
46478	0,568	0,000
46494	0,421	0,000
46516	0,639	0,000
46796	0,247	0,000
47113	0,869	0,000
47121	0,818	0,000
47296	0,752	0,000
47318	0,936	0,000
47440	0,578	0,000
47512	0,740	0,000
47717	0,631	0,000
47741	0,546	0,000
47814	0,694	0,000
47822	0,768	0,000
47890	0,801	0,000
51111	0,754	0,022
56112	0,672	0,000

Fonte: Elaborado pelo autor

Dentre os modelos de atração de notas fiscais criados com o número de estabelecimentos como variável independente, apenas o modelo do CNAE 51111 teve erros distribuídos normalmente. Como pode-se ver na Tabela 34.

Tabela 34 - Testes de normalidade para os modelos tipo 8

CNAE	Estatística W (Shapiro-Wilk)	P-Valor (Shapiro-Wilk)
10911	0,826	0,000
11216	0,448	0,000
14126	0,664	0,000
41204	0,817	0,000
45111	0,726	0,000
45307	0,545	0,000
45412	0,828	0,000
46346	0,391	0,000
46354	0,669	0,000
46397	0,620	0,000
46443	0,278	0,000
46460	0,800	0,000
46478	0,811	0,000
46494	0,951	0,002
46516	0,563	0,000
46796	0,411	0,000
47113	0,873	0,000
47121	0,762	0,000
47296	0,741	0,000
47318	0,896	0,000
47440	0,465	0,000
47512	0,733	0,000
47717	0,813	0,000
47741	0,649	0,000
47814	0,689	0,000
47822	0,815	0,000
47890	0,784	0,000
51111	0,897	0,359
56112	0,481	0,000

Fonte: Elaborado pelo autor

6.2.2.3 Homoscedasticidade e autocorrelação

Da mesma forma que foi feito para os modelos anteriores, utilizou-se o teste de Breusch-Pagan para verificar a homoscedasticidade e o teste de Durbin-Watson para verificar independência dos erros.

Os resultados para os modelos de emissão de notas fiscais com número de empregados como variável explicativa estão apresentados na Tabela 35. Pode-se ver que 22 modelos não passaram no teste de homoscedasticidade e que 6 não passaram no teste de

autocorrelação (não é possível realizar o teste de Durbin-Watson para o modelo do CNAE 51111, pois existem apenas 4 amostras).

Tabela 35 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 5

CNAE	Estatística BP (Breusch-Pagan)	P-Valor (Breusch-Pagan)	Durbin Watson	dL	dU	4-dU	4-dL
10911	43,282	0,000	2,117	1,681	1,716	2,284	2,319
11216	4,123	0,042	1,739	1,295	1,413	2,587	2,705
14126	22,87372274	0,000	1,976	1,716	1,736	2,264	2,284
41204	4,898	0,027	2,071	1,658	1,697	2,303	2,342
45111	7,373	0,007	1,855	1,570	1,615	2,385	2,430
45307	109,443	0,000	2,218	1,749	1,770	2,230	2,251
45412	25,474	0,000	1,901	1,663	1,701	2,299	2,337
46346	3,026	0,082	1,516	1,333	1,439	2,561	2,667
46354	0,120	0,729	2,004	1,124	1,308	2,692	2,876
46397	0,443	0,505	1,356	1,508	1,566	2,434	2,492
46443	4,327	0,038	2,161	1,467	1,535	2,465	2,533
46460	23,721	0,000	2,302	1,295	1,413	2,587	2,705
46478	18,297	0,000	2,826	1,414	1,496	2,504	2,586
46494	6,853	0,009	2,120	1,518	1,574	2,426	2,482
46516	0,007	0,935	2,035	1,223	1,367	2,633	2,777
46796	2,126	0,145	1,382	1,496	1,557	2,443	2,504
47113	9,250	0,002	2,102	1,706	1,736	2,264	2,294
47121	6,809	0,009	2,201	1,732	1,757	2,243	2,268
47296	1,858	0,173	2,100	1,658	1,697	2,303	2,342
47318	8,202	0,004	1,553	1,692	1,724	2,276	2,308
47440	135,512	0,000	1,991	1,757	1,777	2,223	2,243
47512	34,464	0,000	2,461	1,565	1,611	2,389	2,435
47717	55,042	0,000	1,825	1,722	1,748	2,252	2,278
47741	38,022	0,000	2,166	1,452	1,524	2,476	2,548
47814	32,171	0,000	1,576	1,707	1,736	2,264	2,293
47822	17,503	0,000	2,256	1,508	1,566	2,434	2,492
47890	1,238	0,266	2,068	1,736	1,760	2,240	2,264
51111	3,95	0,047	1,297	-	-	-	-
56112	32,772	0,000	2,037	1,697	1,728	2,272	2,303

Fonte: Elaborado pelo autor

Já para os resultados para os modelos de emissão de notas fiscais com número de empresas como variável explicativa, apenas 14 modelos não passaram no teste de homoscedasticidade e 3 não passaram no teste de autocorrelação (como dito, não é possível realizar o teste de Durbin-Watson para o modelo do CNAE 51111, pois existem apenas 4 amostras), como mostra a Tabela 36.

Tabela 36 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 6

CNAE	Estatística BP (Breusch-Pagan)	P-Valor (Breusch-Pagan)	Durbin Watson	dL	dU	4-dU	4-dL
10911	0,118	0,731	1,901	1,681	1,716	2,284	2,319
11216	0,626	0,429	1,765	1,295	1,413	2,587	2,705
14126	48,943	0,000	1,931	1,716	1,736	2,264	2,284
41204	4,179	0,041	2,097	1,658	1,697	2,303	2,342
45111	2,781	0,095	1,488	1,570	1,615	2,385	2,430
45307	15,261	0,000	2,107	1,749	1,770	2,230	2,251
45412	2,173	0,140	2,010	1,663	1,701	2,299	2,337
46346	0,000	0,991	2,150	1,333	1,439	2,561	2,667
46354	0,177	0,674	2,129	1,124	1,308	2,692	2,876
46397	0,385	0,535	1,661	1,508	1,566	2,434	2,492
46443	0,869	0,351	1,987	1,467	1,535	2,465	2,533
46460	0,412	0,521	2,275	1,295	1,413	2,587	2,705
46478	19,387	0,000	2,163	1,414	1,496	2,504	2,586
46494	0,057	0,812	2,001	1,518	1,574	2,426	2,482
46516	0,020	0,888	2,292	1,223	1,367	2,633	2,777
46796	3,132	0,077	1,966	1,496	1,557	2,443	2,504
47113	1,046	0,306	2,263	1,706	1,736	2,264	2,294
47121	12,222	0,000	2,307	1,732	1,757	2,243	2,268
47296	0,074	0,786	2,092	1,658	1,697	2,303	2,342
47318	5,193	0,023	1,375	1,692	1,724	2,276	2,308
47440	16,961	0,000	1,847	1,757	1,777	2,223	2,243
47512	29,415	0,000	2,163	1,565	1,611	2,389	2,435
47717	34,609	0,000	1,788	1,722	1,748	2,252	2,278
47741	41,438	0,000	1,977	1,452	1,524	2,476	2,548
47814	4,876	0,027	1,962	1,707	1,736	2,264	2,293
47822	15,379	0,000	1,946	1,508	1,566	2,434	2,492
47890	0,594	0,441	2,084	1,736	1,760	2,240	2,264
51111	4,000	0,046	1,500	-	-	-	-
56112	36,652	0,000	2,095	1,697	1,728	2,272	2,303

Fonte: Elaborado pelo autor

Para os modelos de atração de notas fiscais que tiveram o número de empregos como variável explicativa, os resultados, Tabela 37, nos mostram que 24 modelos não passaram no teste de homoscedasticidade, 3 não passaram no teste de autocorrelação.

Tabela 37 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 7

CNAE	Estatística BP (Breusch-Pagan)	P-Valor (Breusch-Pagan)	Durbin Watson	dL	dU	4-dU	4-dL
10911	60,627	0,000	1,952	1,742	1,765	2,235	2,258
11216	1,636	0,201	2,134	1,388	1,477	2,523	2,612
14126	22,309	0,000	2,096	1,726	1,745	2,255	2,274
41204	59,034	0,000	1,936	1,754	1,775	2,225	2,246
45111	67,753	0,000	1,594	1,666	1,704	2,296	2,334
45307	80,895	0,000	1,560	1,757	1,777	2,223	2,243
45412	43,815	0,000	2,011	1,713	1,741	2,259	2,287
46346	16,256	0,000	1,736	1,420	1,500	2,500	2,580
46354	0,268	0,605	2,285	1,124	1,308	2,692	2,876
46397	0,104	0,747	2,138	1,552	1,600	2,400	2,448
46443	38,839	0,000	2,247	1,504	1,563	2,437	2,496
46460	14,389	0,000	1,984	1,395	1,482	2,518	2,605
46478	12,906	0,000	2,208	1,457	1,528	2,472	2,543
46494	75,172	0,000	1,944	1,570	1,615	2,385	2,430
46516	4,460	0,035	1,927	1,305	1,420	2,580	2,695
46796	1,616	0,204	2,004	1,535	1,587	2,413	2,465
47113	32,657	0,000	2,117	1,743	1,765	2,235	2,257
47121	7,655	0,006	1,798	1,736	1,754	2,246	2,264
47296	14,428	0,000	2,070	1,739	1,762	2,238	2,261
47318	67,849	0,000	2,320	1,738	1,762	2,238	2,262
47440	75,857	0,000	1,964	1,734	1,751	2,249	2,266
47512	52,854	0,000	2,084	1,683	1,717	2,283	2,317
47717	64,296	0,000	1,984	1,756	1,776	2,224	2,244
47741	99,447	0,000	2,108	1,716	1,743	2,257	2,284
47814	63,310	0,000	2,074	1,719	1,739	2,261	2,281
47822	6,481	0,011	2,188	1,656	1,696	2,304	2,344
47890	8,992	0,003	1,775	1,717	1,736	2,264	2,283
51111	0,505	0,477	2,463	0,489	1,258	2,742	3,511
56112	120,263	0,000	2,213	1,732	1,750	2,250	2,268

Fonte: Elaborado pelo autor

Os resultados para os modelos de atração de nota fiscal dos anos de 2014 e 2015 com número de estabelecimentos como variável explicativa são apresentados na Tabela 38. 20 modelos não tiveram erros homocedásticos pelo teste utilizado, 2 não passaram no teste de autocorrelação e 1 teve resultado inconclusivo para esse teste.

Tabela 38 - Testes de homoscedasticidade e independência dos erros para os modelos tipo 8

CNAE	Estatística BP (Breusch-Pagan)	P-Valor (Breusch-Pagan)	Durbin Watson	dL	dU	4-dU	4-dL
10911	42,111	0,000	1,705	1,742	1,765	2,235	2,258
11216	0,565	0,452	2,094	1,388	1,477	2,523	2,612
14126	10,272	0,001	2,233	1,726	1,745	2,255	2,274
41204	93,327	0,000	1,885	1,754	1,775	2,225	2,246
45111	9,377	0,002	1,495	1,666	1,704	2,296	2,334
45307	34,834	0,000	2,005	1,757	1,777	2,223	2,243
45412	7,173	0,007	1,966	1,713	1,741	2,259	2,287
46346	0,048	0,826	2,061	1,420	1,500	2,500	2,580
46354	0,787	0,375	2,488	1,124	1,308	2,692	2,876
46397	0,007	0,933	1,809	1,552	1,600	2,400	2,448
46443	0,467	0,494	1,964	1,504	1,563	2,437	2,496
46460	1,542	0,214	2,133	1,395	1,482	2,518	2,605
46478	26,566	0,000	2,512	1,457	1,528	2,472	2,543
46494	10,144	0,001	2,107	1,570	1,615	2,385	2,430
46516	0,614	0,433	2,205	1,305	1,420	2,580	2,695
46796	10,017	0,002	1,973	1,535	1,587	2,413	2,465
47113	20,987	0,000	2,022	1,743	1,765	2,235	2,257
47121	5,818	0,016	2,073	1,736	1,754	2,246	2,264
47296	9,244	0,002	1,990	1,739	1,762	2,238	2,261
47318	15,828	0,000	2,060	1,738	1,762	2,238	2,262
47440	0,713	0,398	2,009	1,734	1,751	2,249	2,266
47512	37,694	0,000	1,864	1,683	1,717	2,283	2,317
47717	39,915	0,000	1,797	1,756	1,776	2,224	2,244
47741	126,978	0,000	2,180	1,716	1,743	2,257	2,284
47814	25,801	0,000	2,131	1,719	1,739	2,261	2,281
47822	1,567	0,211	1,706	1,656	1,696	2,304	2,344
47890	18,726	0,000	1,719	1,717	1,736	2,264	2,283
51111	5,177	0,023	2,816	0,489	1,258	2,742	3,511
56112	158,350	0,000	2,165	1,732	1,750	2,250	2,268

Fonte: Elaborado pelo autor

6.2.3 Avaliação dos resultados

Para os modelos que tinham o número de notas fiscais para dezembro de 2014 e 2015 como variável dependente, nenhum modelo de atração de notas fiscais respeitou todos os testes realizados.

Já para os modelos de emissão, nenhum com número de estabelecimentos como variável independente respeitou todos os testes realizados e houve 1 dos modelos com número de empregados como variável independente que respeitou os testes de normalidade,

homoscedasticidade e independência, entretanto não houve indícios para rejeitar a hipótese nula do teste-t para o intercepto. O modelo está apresentado na Tabela 39.

Tabela 39 - Modelo tipo 1 para o CNAE 46346

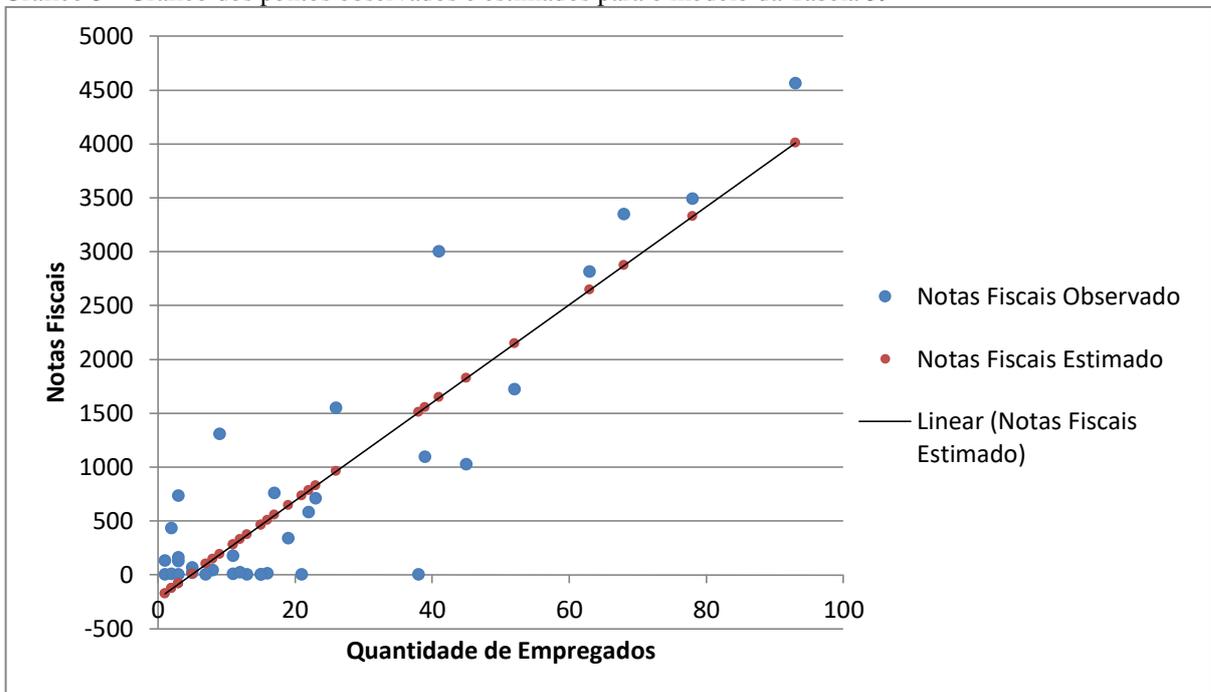
$$y = \beta_0 + \beta_1 x$$

CNAE	Amostras	R ²	β_0	β_1
46346	35	0,794	-219,24	45,48

Fonte: Elaborado pelo autor

Os pontos utilizados para a criação do modelo e as estimativas realizadas por ele estão representados no Gráfico 5.

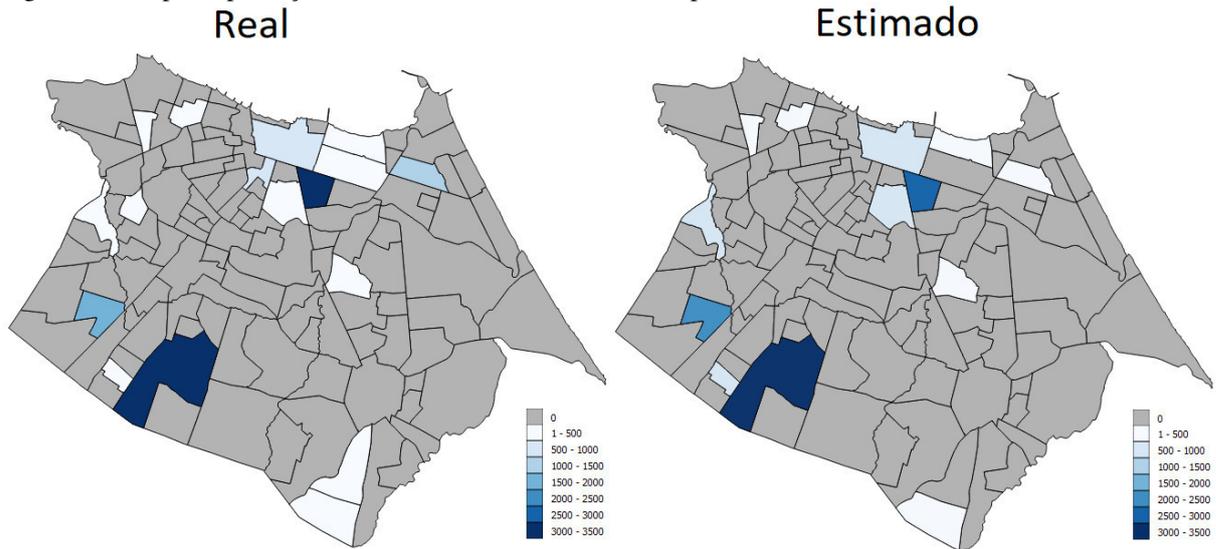
Gráfico 5 - Gráfico dos pontos observados e estimados para o modelo da Tabela 39



Fonte: Elaborado pelo autor

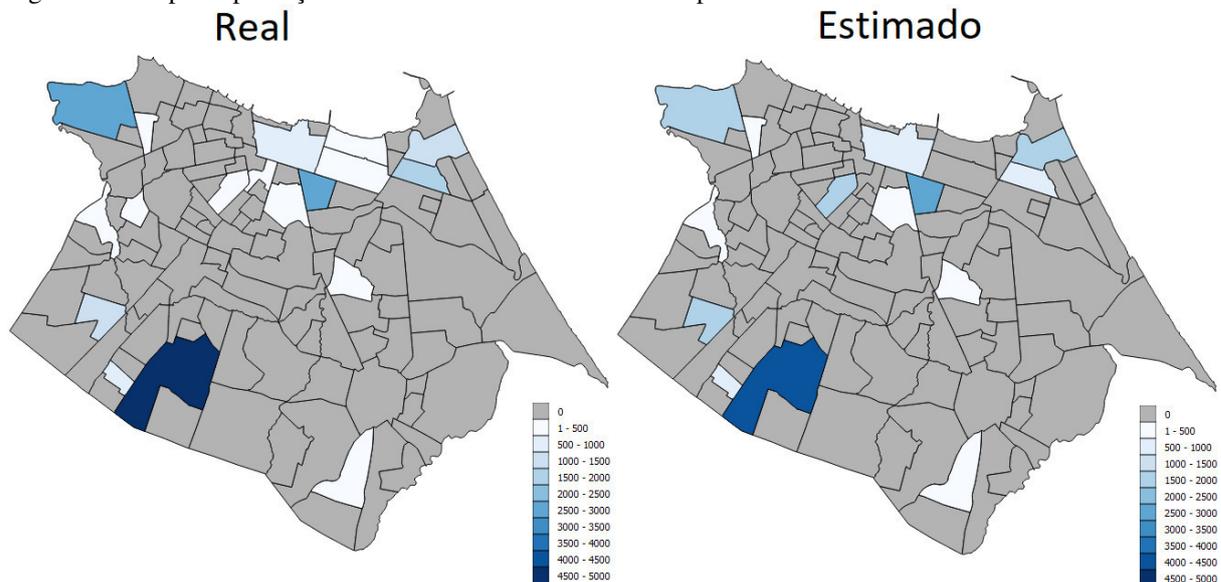
As Figuras 11 e 12 são a representação gráfica da geração de notas fiscais real e modelada pela equação da Tabela 39 para dezembro de 2014 e 2015, respectivamente. Na representação, os resultados negativos obtidos pela equação foram considerados como zero.

Figura 11 - Mapa da produção de notas fiscais do CNAE 46346 para dezembro de 2014



Fonte: Elaborado pelo autor

Figura 12 - Mapa da produção de notas fiscais do CNAE 46346 para dezembro de 2015



Fonte: Elaborado pelo autor

Para os modelos que tinham o número de notas fiscais referente aos anos de 2014 e 2015 como variável dependente, nenhum modelo de atração de notas fiscais respeitou todos os testes realizados.

Para os modelos de emissão de notas fiscais, nenhum modelo com número de estabelecimentos como variável independente respeitou todos os testes realizados. Já dentre os modelos com número de empregados como variável independente, houve 1 modelo que respeitou todos os testes realizados. Esse modelo é apresentado na Tabela 40.

Tabela 40 - Modelo tipo 5 para o CNAE 46346

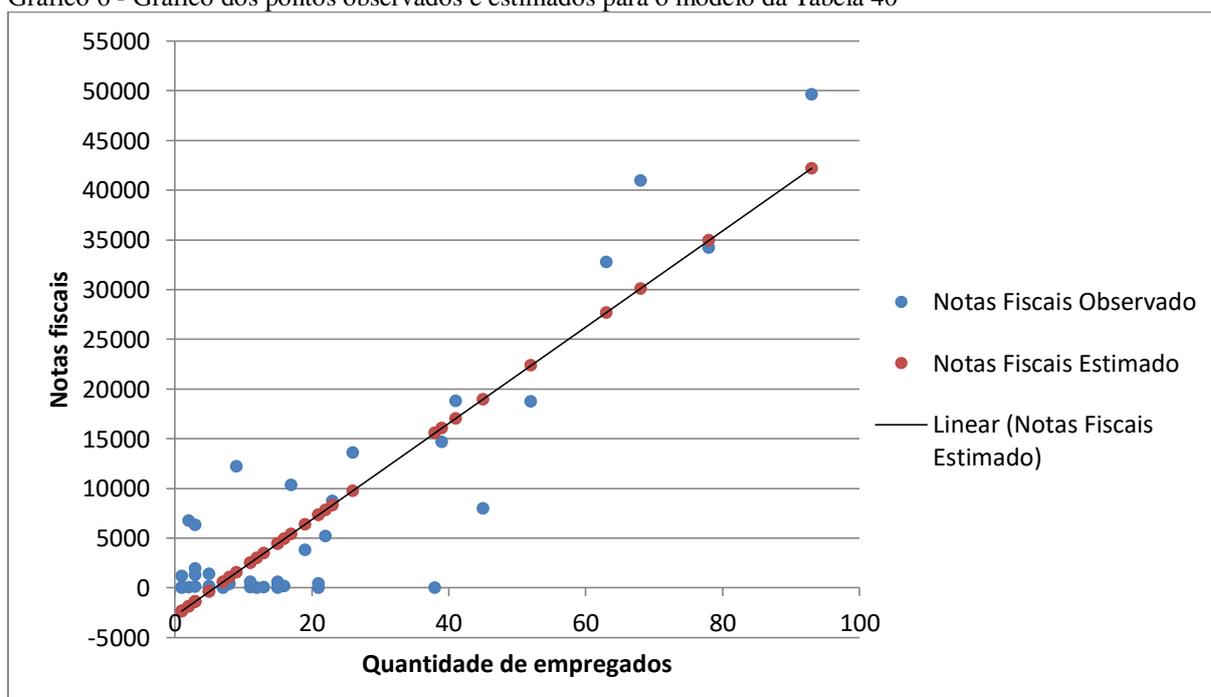
$$y = \beta_0 + \beta_1 x$$

CNAE	Amostras	R ²	β_0	β_1
46346	38	0,802	-2801,85	484,23

Fonte: Elaborado pelo autor

Da mesma forma do modelo anterior, os pontos utilizados para a criação do modelo da Tabela 40 e as estimativas obtidas estão representados no Gráfico 6.

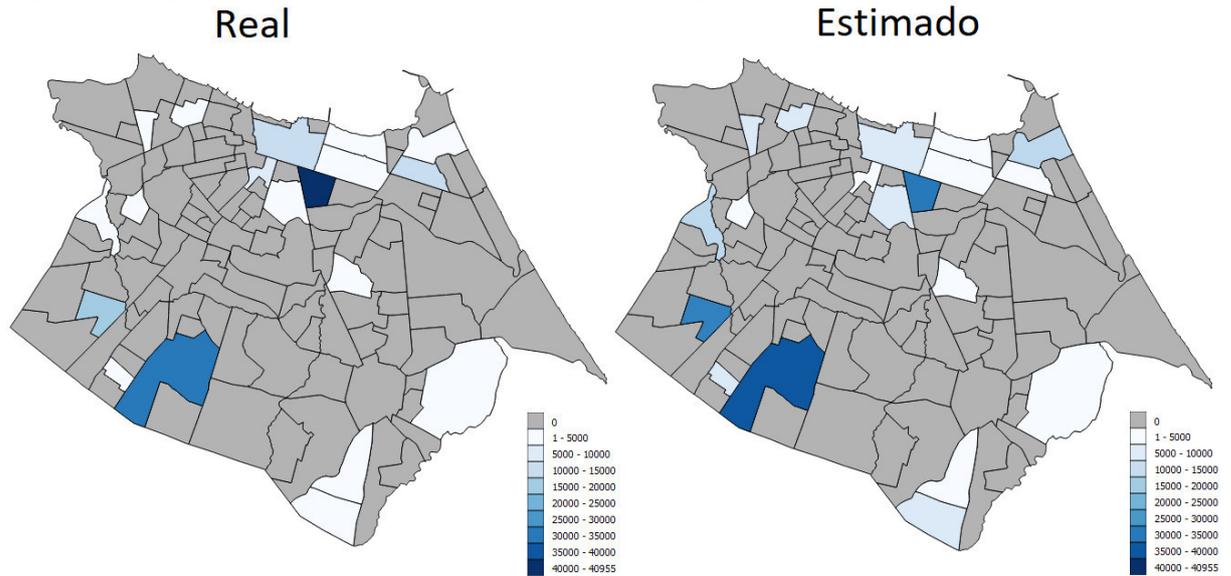
Gráfico 6 - Gráfico dos pontos observados e estimados para o modelo da Tabela 40



Fonte: Elaborado pelo autor

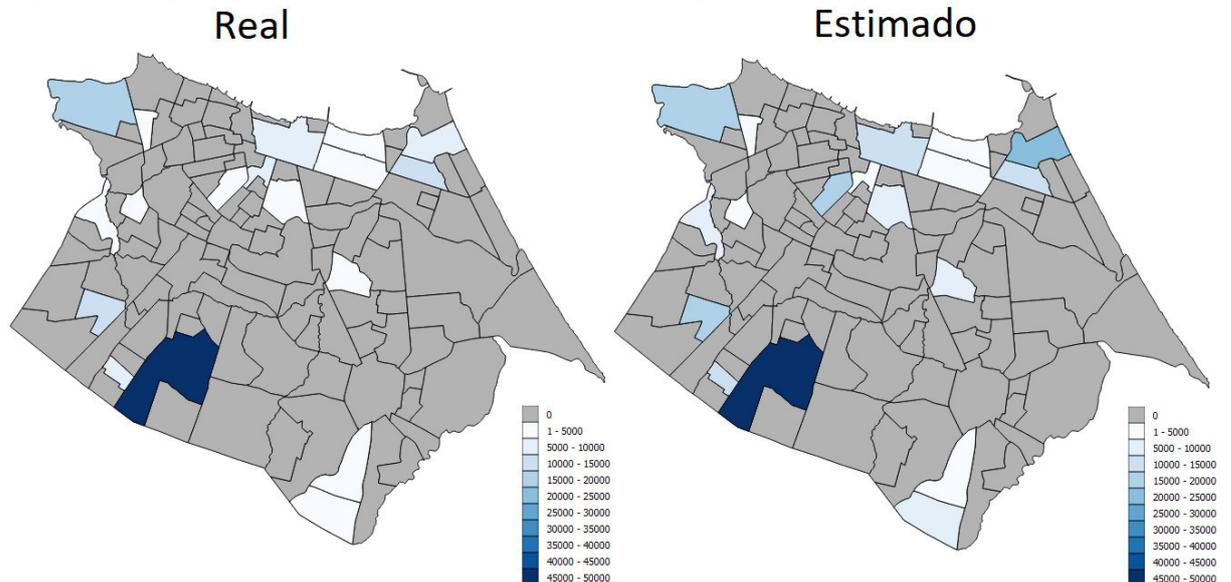
As Figuras 13 e 14 são a representação gráfica da geração de notas fiscais real e modelada pela equação da Tabela 40 para o ano de 2014 e 2015, respectivamente. Como feito anteriormente, na representação, os resultados negativos obtidos da equação foram considerados como zero.

Figura 13 - Mapa da produção de notas fiscais do CNAE 46346 para o ano de 2014



Fonte: Elaborado pelo autor

Figura 14 - Mapa da produção de notas fiscais do CNAE 46346 para o ano de 2015



Fonte: Elaborado pelo autor

O CNAE 46346 corresponde ao segmento de comércio atacadista de carnes, produtos da carne e pescado. Pode-se notar que ele está mais fortemente presente em áreas periféricas, principalmente no bairro Mondubim.

As variações entre os dados reais e os dados modelados ocorrem pois existem bairros com poucos funcionários que produzem uma quantidade grande de notas fiscais, por exemplo. Isso pode ocorrer pois algumas empresas são mais eficientes do que outras, pela estrutura das empresas ou pela forma que empresas geram as notas fiscais, duas empresas podem vender a mesma quantidade de produtos, mas dividir a venda em um número diferente

de notas. Essas variações específicas entre empresas é uma das maiores dificuldades da modelagem da geração de notas fiscais por segmento econômico através de regressão linear simples.

Além disso, outro fator que pode ter contribuído para os resultados é grande quantidade de empregos informais na cidade de Fortaleza, isso faz com que o número de funcionários declarado por algumas empresas seja menor do que o número real.

7 CONCLUSÕES

A proposta do trabalho foi entender como estava ocorrendo a distribuição de notas fiscais entre bairros e entre CNAEs e buscar modelar a produção e a atração de notas fiscais para os CNAEs que mais participavam da geração de notas fiscais na cidade de Fortaleza. Tudo isso com o intuito de suportar o conhecimento sobre o transporte de carga urbana.

A compatibilização das bases utilizadas teve papel fundamental para o desenvolvimento do trabalho, principalmente pela questão das informações dos bairros. Os bancos de dados utilizados foram bastante úteis, quando tratados, pois tinham informações sobre variáveis muito relevantes para o fenômeno que estava sendo modelado. As bases utilizadas têm potencial para diversas outras análises.

Pôde-se perceber que há uma concentração grande de empregos e empresas por bairro e CNAE em Fortaleza. A concentração de empresas e empregos em poucos bairros merece uma investigação mais aprofundada, uma vez que as consequências disso não são totalmente conhecidas na cidade. Já a concentração de empresas e empregos em poucos CNAEs revela que Fortaleza tem sua economia baseada em poucos setores que são estratégicos para a cidade.

Os dados também indicam que há uma concentração na produção e na atração de notas fiscais, entretanto esse dado está muito atrelado com o tipo de atividade de cada empresa, pois existem empresas que para uma única venda podem produzir diversas notas fiscais, por exemplo. Considerando isso, é possível notar que a concentração de notas fiscais por CNAE vem mudando, principalmente na produção de notas fiscais. Em 2011, 73,59% das notas foram produzidas pelos 20 CNAEs que mais produziram notas e, em 2015, esse número caiu para 65,52%. Para a atração, em 2011, 59,70% das notas foram atraídas pelos 20 CNAEs que mais atraíram notas e, em 2015, esse número caiu para 55,27%.

Quanto aos modelos desenvolvidos, a grande maioria, tanto para produção quanto para atração de notas fiscais, utilizando regressão linear simples tanto com número de empregados quanto com número de empresas como variável explicativa, não é indicada para explicar a geração de notas fiscais, pois não cumprem as premissas dos modelos de regressão linear, sendo necessárias algumas manipulações dos dados ou a desconsideração da variável.

Apenas para a produção de notas fiscais do CNAE 46346 o resultado foi satisfatório. Dessa forma, conclui-se que a utilização de modelos de regressão linear simples para explicar a atração e a produção de notas fiscais por meio das variáveis utilizadas nesse trabalho não é indicada para a maioria dos CNAEs, uma vez que, como dito, grande parte dos

modelos de regressão linear simples não atenderam às premissas do modelo, como os critérios de homoscedasticidade, normalidade e independência dos erros.

7.1 Sugestões para trabalhos futuros

Para trabalhos futuros, sugere-se, inicialmente, a obtenção de outras variáveis e o desenvolvimento de outros modelos que possam explicar a produção e a atração de notas fiscais.

Há também a possibilidade de se aplicar outras técnicas, diferente da regressão linear, para modelar a geração de notas fiscais, uma vez que outras técnicas, até mesmo não lineares, podem se adequar melhor ao fenômeno estudado.

Além disso, é importante que se faça uma análise para buscar coeficientes ou outros modelos que permitam converter notas fiscais para número de viagens de carga para cada CNAE, uma vez que cada segmento econômico tem uma dinâmica de distribuição.

REFERÊNCIAS

- BASTIDA, C.; HOLGUÍN-VERAS, J. **Freight Generation Models**. Transportation Research Record: Journal of the Transportation Research Board, n. 2097, p. 51-61, 2009.
- BATTY, M. **Big data, smart cities and city planning**. Dialogues in Human Geography, n. 3, p. 274-279, 2013.
- BEHRENDTS, S.; LINDHOLM, M.; WOXENIUS, J. **The Impact of Urban Freight Transport: A Definition of Sustainability from an Actor's Perspective**. Transportation Planning and Technology, v. 31, n. 6, p. 693-713, 2008.
- BENJELLOUN, A.; CRAINIC, T. G. **Trends, challenges, and perspectives in city logistics**. Buletinul AGIR, n. 4, p. 45-51, 2009.
- BRANDÃO FILHO, C. A. **Análise de viagens do transporte urbano de cargas com base em dados secundários**. 2017. Dissertação (Mestrado em Engenharia de Transportes) - Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2017.
- CAMPOS, V. B. G. **Planejamento de Transportes: Conceitos e Modelos**. 1. ed. Rio de Janeiro: Interciência, 2013.
- CARVALHO, J. A. DE. **Uma contribuição ao planejamento do transporte de cargas em áreas urbanas**. 1998. Dissertação (Mestrado em Engenharia de Transportes) - Instituto Militar de Engenharia, Rio de Janeiro, 1998.
- CZERNIAK, R. J.; LAHSENE, J. S.; CHATTERJEE, A. **Urban Freight Movement: What Form Will It Take ?**. A1B07 Committee on Urban Goods Movement TRB, 2000.
- D'ESTE, G. **Urban Freight Movement Modeling**. Handbook of Transport Modelling, p. 633-647, 2007.
- DABLANC, L. **Goods transport in large European cities: Difficult to organize, difficult to modernize**. Transportation Research Part A: Policy and Practice, v. 41, n. 3, p. 280-285, 2007.
- DABLANC, L.; GIULIANO, G.; HOLLIDAY, K.; O'BRIEN, T. **Best Practices in Urban Freight Management**. Transportation Research Record: Journal of the Transportation Research Board, n. 2379, p. 29-38, 2013.
- FACCHINI, D. **Análise dos gaps de percepção dos atores envolvidos no transporte urbano de carga em Porto Alegre**. 2006. Dissertação (Mestrado em Engenharia de Produção) - Escola de Engenharia, Universidade Federal do Rio Grande de Sul, Porto Alegre, 2006.
- FERREIRA, B. **Análise de geração de viagens de carga em áreas urbanas**. 2016. Dissertação (Mestrado em Engenharia de Transportes) - Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

- FINKLER, T. F. **Desenvolvimento de uma ferramenta para obtenção de modelos empíricos**. 2003. Dissertação (Mestrado em Engenharia) - Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.
- GASPARINI, A. **Atratividade do transporte de carga para pólos geradores de viagem em áreas urbanas**. 2008. Dissertação (Mestrado em Engenharia de Transportes) - Instituto Militar de Engenharia, Rio de Janeiro, 2008.
- GIULIANO, G.; GORDON, P.; PAN, Q.; PARK, J.; WANG, L. **Estimating freight flows for metropolitan area highway networks using secondary data sources**. *Networks and Spatial Economics*, v. 10, n. 1, p. 73-91, 2010.
- HERCULANO, A. T. **Estimação da matriz origem-destino de entregas de mercadorias a partir de dados do ICMS: uma proposta metodológica**. 2015. Monografia (Graduação em Engenharia Civil) - Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza, 2015.
- HOLGUÍN-VERAS, J.; JALLER, M.; SANCHEZ-DIAZ, I.; WOJTOWICZ, J.; CAMPBELL, S.; LEVINSON, H.; LAWSON, C.; POWERS, E. L.; TAVASSZY, L. **Freight Trip Generation and Land Use**. National Cooperative Freight Research Program, 2012.
- KITCHIN, R. **The real-time city? Big data and smart urbanism**. *GeoJournal*, v. 79, n. 1, p. 1-14, 2014.
- LABRINIDIS, A.; JAGADISH, H. V. **Challenges and opportunities with big data**. *Proceedings of the VLDB Endowment*, v. 5, n. 12, p. 2032-2033, 2012.
- LINDHOLM, M. **A sustainable perspective on urban freight transport: Factors affecting local authorities in the planning procedures**. *Procedia - Social and Behavioral Sciences*, v. 2, n. 3, p. 6205-6216, 2010.
- LINDHOLM, M.; BEHRENDIS, S. **Challenges in urban freight transport planning - a review in the Baltic Sea Region**. *Journal of Transport Geography*, v. 22, p. 129-136, 2012.
- MONTGOMERY, D.C.; RUNGER, G. C. **Estatística Aplicada e Probabilidade para Engenheiros**. 4 ed. Rio de Janeiro: LTC, 2009.
- NUZZOLO, A.; CRISALLI, U.; COMI, A. **A trip chain order model for simulating urban freight restocking**. *European Transport*, n. 50, p. 1-13, 2012.
- ORTÚZAR, J. DE D.; WILLUMSEN, L. G. **Modelling Transport**. 4 ed. New York: Wiley, 2011.
- SÁNCHEZ-DÍAZ, I. **Modeling urban freight generation: A study of commercial establishments' freight needs**. *Transportation Research Part A: Policy and Practice*, v. 102, p. 3-17, 2016.
- SANTOS, E. M. DOS. **Uso de dados de documentos fiscais eletrônicos para o planejamento do transporte urbano de cargas**. 2015. Tese (Doutorado em Transportes) - Faculdade de Tecnologia, Universidade de Brasília, Brasília, 2015.

STATHOPOULOS, A.; VALERI, E.; MARCUCCI, E. **Stakeholder reactions to urban freight policy innovation.** *Journal of Transport Geography*, v. 22, p 34-45, 2012.

TANIGUCHI, E.; THOMPSON, R. G.; YAMADA, T. **Recent Trends and Innovations in Modelling City Logistics.** *Procedia - Social and Behavioral Sciences*, v. 125, p. 4-14, 2014.

WOUDSMA, C. **Understanding the movement of goods, not people:** Issues, evidence and potential. *Urban Studies*, v. 38, n. 13, p. 2439-2455, 2001.

APÊNDICE A – CNAES E DENOMINAÇÕES

CNAE	Denominação
10.91-1	Fabricação de produtos de panificação
11.21-6	Fabricação de águas envasadas
11.22-4	Fabricação de refrigerantes e de outras bebidas não-alcoólicas
14.12-6	Confecção de peças do vestuário, exceto roupas íntimas
41.10-7	Incorporação de empreendimentos imobiliários
41.20-4	Construção de edifícios
45.11-1	Comércio a varejo e por atacado de veículos automotores
45.20-0	Manutenção e reparação de veículos automotores
45.30-7	Comércio de peças e acessórios para veículos automotores
45.41-2	Comércio por atacado e a varejo de motocicletas, peças e acessórios
46.32-0	Comércio atacadista de cereais e leguminosas beneficiados, farinhas, amidos e féculas
46.34-6	Comércio atacadista de carnes, produtos da carne e pescado
46.35-4	Comércio atacadista de bebidas
46.36-2	Comércio atacadista de produtos do fumo
46.37-1	Comércio atacadista especializado em produtos alimentícios não especificados anteriormente
46.39-7	Comércio atacadista de produtos alimentícios em geral
46.44-3	Comércio atacadista de produtos farmacêuticos para uso humano e veterinário
46.46-0	Comércio atacadista de cosméticos, produtos de perfumaria e de higiene pessoal
46.47-8	Comércio atacadista de artigos de escritório e de papelaria; livros, jornais e outras publicações
46.49-4	Comércio atacadista de equipamentos e artigos de uso pessoal e doméstico não especificados anteriormente
46.51-6	Comércio atacadista de computadores, periféricos e suprimentos de informática
46.72-9	Comércio atacadista de ferragens e ferramentas
46.79-6	Comércio atacadista especializado de materiais de construção não especificados anteriormente e de materiais de construção em geral
46.81-8	Comércio atacadista de combustíveis sólidos, líquidos e gasosos, exceto gás natural e GLP
47.11-3	Comércio varejista de mercadorias em geral, com predominância de produtos alimentícios - hipermercados e supermercados
47.12-1	Comércio varejista de mercadorias em geral, com predominância de produtos alimentícios - minimercados, mercearias e armazéns
47.21-1	Comércio varejista de produtos de padaria, laticínio, doces, balas e semelhantes
47.29-6	Comércio varejista de produtos alimentícios em geral ou especializado em produtos alimentícios não especificados anteriormente; produtos do fumo
47.31-8	Comércio varejista de combustíveis para veículos automotores
47.44-0	Comércio varejista de ferragens, madeira e materiais de construção
47.51-2	Comércio varejista especializado de equipamentos e suprimentos de informática
47.54-7	Comércio varejista especializado de móveis, colchoaria e artigos de iluminação
47.55-5	Comércio varejista especializado de tecidos e artigos de cama, mesa e banho
47.61-0	Comércio varejista de livros, jornais, revistas e papelaria
47.71-7	Comércio varejista de produtos farmacêuticos para uso humano e veterinário
47.72-5	Comércio varejista de cosméticos, produtos de perfumaria e de higiene pessoal
47.74-1	Comércio varejista de artigos de óptica

CNAE	Denominação
47.81-4	Comércio varejista de artigos do vestuário e acessórios
47.82-2	Comércio varejista de calçados e artigos de viagem
47.89-0	Comércio varejista de outros produtos novos não especificados anteriormente
49.21-3	Transporte rodoviário coletivo de passageiros, com itinerário fixo, municipal e em região metropolitana
49.30-2	Transporte rodoviário de carga
51.11-1	Transporte aéreo de passageiros regular
56.11-2	Restaurantes e outros estabelecimentos de serviços de alimentação e bebidas
69.20-6	Atividades de contabilidade, consultoria e auditoria contábil e tributária
78.10-8	Seleção e agenciamento de mão de obra
78.20-5	Locação de mão de obra temporária
80.11-1	Atividades de vigilância e segurança privada
81.12-5	Condomínios prediais
81.21-4	Limpeza em prédios e em domicílios
82.20-2	Atividades de teleatendimento
82.99-7	Atividades de serviços prestados principalmente às empresas não especificadas anteriormente
84.11-6	Administração pública em geral
84.12-4	Regulação das atividades de saúde, educação, serviços culturais e outros serviços sociais
85.13-9	Ensino fundamental
85.31-7	Educação superior - graduação
86.10-1	Atividades de atendimento hospitalar
86.30-5	Atividades de atenção ambulatorial executadas por médicos e odontólogos
94.30-8	Atividades de associações de defesa de direitos sociais
94.91-0	Atividades de organizações religiosas
96.02-5	Cabeleireiros e outras atividades de tratamento de beleza