



**Elissandro Alex da Rosa**

**TOMADA DE DECISÃO: MODELO ECONOMÉTRICO  
PARA DEFINIR A LOCALIZAÇÃO DE NOVAS FILIAIS**

**Horizontalina**

**2018**

**Elissandro Alex da Rosa**

**TOMADA DE DECISÃO: MODELO ECONÔMETRICO PARA  
DEFINIR A LOCALIZAÇÃO DE NOVAS FILIAIS**

Trabalho Final de Curso apresentado como requisito parcial para a obtenção do título de Bacharel em Ciências Econômicas pelo Curso de Ciências Econômicas da Faculdade Horizontina (FAHOR).

**ORIENTADOR: Stephan Sawitzki, Mestre**

**Horizontina**

**2018**

**FAHOR – FACULDADE HORIZONTINA  
CURSO DE CIÊNCIAS ECONÔMICAS**

A Comissão Examinadora, abaixo assinada, aprova a monografia:

**“Tomada de decisão: modelo econométrico para definir a localização de novas filiais”**

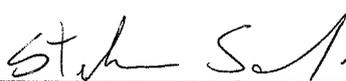
Elaborada por:

**Elissandro Alex da Rosa**

como requisito parcial para a obtenção do grau de Bacharel em  
Ciências Econômicas

Aprovado em: 30/11/2018

Pela Comissão Examinadora



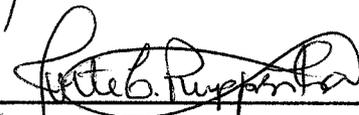
---

**Me. Stephan Sawitzki**  
Presidente da Comissão Examinadora - Orientador



---

**Me. Marcio Leandro Kalkmann**  
FAHOR – Faculdade Horizontina



---

**Esp. Ivete Linn Ruppenthal**  
FAHOR – Faculdade Horizontina

**Horizontina**

**2018**

## **DEDICATÓRIA**

Dedico a minha família, pelo apoio e por proporcionarem um ambiente favorável, de condições e incentivo, à conclusão deste curso. Dedico a empresa a qual colaboro, pelo auxílio e compreensão da necessidade de qualificação profissional. Dedico também aos meus amigos e colegas, que direta ou indiretamente contribuíram com seus conhecimentos para a realização deste trabalho.

## **AGRADECIMENTOS**

Agradeço primeiramente a minha família pelo apoio e compreensão da importância da realização deste trabalho. Agradeço também a empresa objeto deste estudo, pela confiança neste que vos escreve. Agradeço a todos os professores, cujos ensinamentos certamente estão de alguma forma presentes neste projeto, em especial ao meu orientador Stephan Sawitzki e também a professora Ivete Linn Ruppenthal, pelos ensinamentos, empenho e dedicação no auxílio do cumprimento de todas as etapas deste trabalho de conclusão de curso. Por fim, agradeço aos demais amigos e colegas que de alguma forma contribuíram e estiveram presentes nesta importante fase do meu desenvolvimento pessoal e profissional.

*“Tudo que pode ser dito, pode ser dito com  
clareza”. - Ludwig Wittgenstein.*

## RESUMO

A alocação de recursos por parte das empresas deve ser realizada com base em informações relevantes e concisas, para que a tomada de decisão possa ser mais assertiva. A abertura de novas filiais implica na busca por informações relevantes que podem minimizar os riscos de desperdício de recursos. Assim, o problema de pesquisa, é como definir as cidades que devem receber estas novas filiais. Neste sentido, o objetivo, era identificar as cidades do Estado do Rio Grande do Sul com maior capacidade de gerar receitas, a partir da técnica preditiva, econômica e estatística Regressão Linear Múltipla. Para delinear a pesquisa, quanto aos fins, foi de natureza qualitativa e exploratória. Quanto aos meios, pode-se classificar como bibliográfica e estudo de caso. A coleta de dados se deu de forma documental, para posteriormente tratar os dados e construir um modelo com o auxílio de softwares estatísticos específicos. Buscando atingir o objetivo, foi empregada a técnica matemática, estatística e preditiva de regressão linear múltipla, na qual indicadores qualitativos e principalmente quantitativos serviram para a elaboração de um modelo para prever futuras receitas nas cidades gaúchas em que a rede varejista, objeto deste estudo, não atua. Através deste estudo foi possível observar que não necessariamente, os fatores externos têm influência direta nos resultados operacionais de algumas empresas ou de alguns ramos de atividade. Assim, fatores internos, os quais as organizações têm maior capacidade de manipulação, como por exemplo equipes, treinamento e desenvolvimento, localização estratégica e marketing, podem ter alta capacidade de gerar resultados positivos para a empresa.

**Palavras-chave:** Problema de Negócio. Tomada de Decisão. Análise Preditiva.

## ABSTRACT

*The allocation of resources by companies should be based on relevant and concise information so that decision making can be more assertive. The opening of new subsidiaries implies the search for relevant information that can minimize the risks of waste of resources. So the research problem is how to define the cities that should receive these new branches. In this sense, the objective was to identify the cities of the State of Rio Grande do Sul with greater capacity to generate revenues, from the predictive, economic and statistical technique Multiple Linear Regression. To delineate the research, in terms of the purposes, was qualitative and exploratory. As for the means, one can classify as bibliographical and case study. The data collection was done in a documentary way, to later treat the data and build a model with the aid of specific statistical software. In order to reach the objective, the mathematical, statistical and predictive technique of multiple linear regression was employed, in which qualitative and mainly quantitative indicators served to elaborate a model to predict future revenues in cities in the state of Rio Grande do Norte, where the retail network, does not act. Through this study it was possible to observe that external factors do not necessarily have a direct influence on the operating results of some companies or some branches of activity. Thus, internal factors, which organizations have greater capacity for manipulation, such as teams, training and development, strategic location and marketing, may have high capacity to generate positive results for the company..*

**Keywords:** *Business Problem, Decision Making, Predictive Analysis.*

## LISTA DE ILUSTRAÇÕES

### Figuras

Figura 1 – CRISP-DM .....	22
Figura 2 – Correlação entre variáveis independentes .....	59
Figura 3 – Fluxograma do modelo de regressão linear com o uso do software Alteryx Designer .....	61
Figura 4 – Relatório de regressão linear com uso do software Alteryx Designer .....	62
Figura 5 – Relatório de regressão linear com uso do software IBM SPSS .....	64
Figura 6 – Resultado do Modelo com uso do software Alteryx Designer .....	65
Figura 7 – Resultado do Modelo com uso do software IBM SPSS.....	66
Figura 8 – Resultado do modelo de regressão considerando a multicolinearidade ..	67
Figura 9 – Resultado do modelo de regressão com uso das variáveis com significância estatística .....	68

### Gráficos

Gráfico 1 – Faturamento versus população estimada para 2017 .....	36
Gráfico 2 – Faturamento versus percentual de população urbana.....	37
Gráfico 3 – Faturamento versus número de empregos formais .....	38
Gráfico 4 – Faturamento versus número de empregos formais masculinos .....	39
Gráfico 5 – Faturamento versus número de empregos formais femininos .....	40
Gráfico 6 – Faturamento X idade média dos trabalhadores com vínculo ativo em 31/12/2017 .....	41
Gráfico 7 – Faturamento versus idade média dos trabalhadores do sexo masculino com vínculo ativo em 31/12/2017 .....	42
Gráfico 8 – Faturamento versus idade média dos trabalhadores do sexo feminino com vínculo ativo em 31/12/2017.....	43

Gráfico 9 – Faturamento versus remuneração média dos trabalhadores com vínculo ativo em 31/12/2017 .....	44
Gráfico 10 – Faturamento versus número de domicílios privados urbanos.....	45
Gráfico 11 – Faturamento versus número de domicílios privados rurais.....	46
Gráfico 12 – Faturamento versus PIB per capita.....	47
Gráfico 13 – Faturamento versus área .....	48
Gráfico 14 – Faturamento versus IDESE .....	49
Gráfico 15 – Faturamento versus valor em poupança.....	50
Gráfico 16 – Faturamento versus operações de crédito.....	51
Gráfico 17 – Faturamento versus depósitos a prazo.....	52
Gráfico 18 – Faturamento versus arrecadação total de tributos.....	53
Gráfico 19 – Faturamento versus despesas realizadas .....	54
Gráfico 20 – Faturamento versus número de estabelecimentos .....	55
Gráfico 21 – Faturamento versus número de empresas comerciais .....	56
Gráfico 22 – Faturamento versus razão entre o número de empresas comerciais e a população.....	57

## **Quadros**

Quadro 1 – Etapas para a tomada de decisão .....	20
Quadro 2 – Etapas do framework de resolução de problemas CRISP-DM.....	23
Quadro 3 – Correlação das variáveis independentes com a variável dependente....	35

## LISTA DE ABREVIATURAS

ARE	ÁREA
DEP_P	DEPÓSITOS A PRAZO
DES_REA	DEPESAS REALIZADAS
DPP_URB	DOMICÍLIOS PRIVADOS URBANOS
EMP_FOR	EMPREGOS FORMAIS
EMP_FOR_F	EMPREGOS FORMAIS FEMININOS
IBGE	INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA
IDA_MED	IDADE MÉDIA
IDA_MED_F	IDADE MÉDIA FEMININA
IDA_MED_M	IDADE MÉDIA MASCULINA
IDE	IDEESE
IDEESE	ÍNDICE DE DESENVOLVIMENTO SOCIOECONÔMICO
NEC_POP	RAZÃO ENTRE O NÚMERO DE EMPRESAS COMERCIAIS E A POPULAÇÃO
NUM_EST	NÚMERO DE ESTABELECIMENTOS
NUM_EMP_COM	NÚMERO DE EMPRESAS COMERCIAIS
OP_CRE	OPERAÇÕES DE CRÉDITO
POP	POPULAÇÃO
POP_URB	POPULAÇÃO URBANA
POU	POUPANÇA
PPC	PIB PER CAPITA
RAIS	RELAÇÃO ANUAL DE INFORMAÇÕES SOCIAIS

REM\_MED  
TRI

REMUNERAÇÃO MÉDIA  
TRIBUTOS

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>13</b>
<b>2 REFERENCIAL TEÓRICO</b> .....	<b>16</b>
2.1 VAREJO .....	16
<b>2.1.1 Varejo no Brasil</b> .....	<b>17</b>
2.2 PROBLEMA DE NEGÓCIO.....	18
2.3 TOMADA DE DECISÃO .....	19
<b>2.3.1 Processo de Tomada de Decisão</b> .....	<b>20</b>
2.4 FRAMEWORK DE RESOLUÇÃO DE PROBLEMAS .....	21
2.5 MODELO PREDITIVO.....	23
<b>2.5.1 Modelo Econométrico</b> .....	<b>25</b>
<b>2.5.2 Metodologia do Modelo Econométrico</b> .....	<b>26</b>
<b>3 METODOLOGIA</b> .....	<b>28</b>
<b>4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS</b> .....	<b>30</b>
4.1 CARACTERIZAÇÃO DA EMPRESA.....	30
4.2 IDENTIFICAÇÃO DO PERFIL DAS CIDADES GAÚCHAS ATRAVÉS DA COLETA DE DADOS.....	31
4.3 DESENVOLVIMENTO DO MODELO ECONOMETRICO DE REGRESSÃO LINEAR MÚLTIPLA.....	33
<b>4.3.1 CORRELAÇÃO</b> .....	<b>33</b>
<b>4.3.2 REGRESSÃO LINEAR</b> .....	<b>60</b>
<b>5 CONSIDERAÇÕES FINAIS</b> .....	<b>69</b>
<b>REFERÊNCIAS</b> .....	<b>71</b>

## 1 INTRODUÇÃO

O mundo corporativo exige alto nível de planejamento, com tomadas de decisões rápidas e precisas. Tratando-se do varejo, área chave deste estudo, as mudanças acontecem muito rapidamente. Novos produtos, novos modelos, novas tecnologias, mercado concorrencial, perfis e preferências dos clientes, exigem alta performance do tomador de decisões. Uma das decisões mais difíceis talvez seja no âmbito dos investimentos. Aqui, depara-se com uma necessidade extrema de minimização de erros, os quais podem gerar desperdício de recursos e grandes prejuízos para a empresa.

Escolher a localização de uma nova filial exige planejamento. Além é do local, é importante avaliar o mercado, o público alvo e a concorrência. Em um mercado que possui características muito distintas da atual configuração da empresa, torna-se mais difícil saber como será o comportamento do novo negócio. O público-alvo apresenta distinção de uma região para outra, e o produto ofertado pode simplesmente não ter a mesma atratividade em um novo local. Finalmente, a concorrência, pode e irá impactar nos possíveis resultados, pois em um cenário de fortíssima concorrência, torna-se mais difícil apoderar-se de uma fatia do mercado.

Aumentar a relevância da empresa no mercado é parte fundamental de um planejamento estratégico que visa a expansão da marca. As formas de aumentar a relevância, dentre muitos, pode dar-se pelo lançamento de novos produtos, reposicionamento da marca ou abertura de filiais em novos locais. Este último, é tomado como base deste estudo, visto que a definição de uma nova cidade ou região, deve ser estudada antes da tomada de decisão, buscando perceber se o retorno sobre o investimento acontece de forma mais rápida ou é garantido.

Dispende recursos para uma nova loja implica em um custo de oportunidade, o qual é desconhecido em um cenário que apresenta a possibilidade de posicioná-la na cidade A ou na cidade B, quando não se conhece os resultados possíveis em cada uma das duas opções. Portanto, a escolha do local que irá receber uma nova loja, não pode ser realizada de modo superficial. Encontrar as variáveis corretas e criar um modelo que possa mensurar os resultados, é um desafio que certamente dará informações importantes na hora de decidir onde serão alocados os recursos.

Dado o contexto, o objeto deste estudo é uma rede varejista de moda, atuando no Estado do Rio Grande do Sul, que tem como visão chegar a 80 lojas até o ano de 2020, inaugurando então 16 novas filiais no mercado, já que atualmente, conta com 64 lojas.

No processo de escolha da localização de uma nova loja, pode-se elencar inúmeras variáveis qualitativas e quantitativas. Porém, a simples escolha de variáveis, não fornece informações precisas para fazer a escolha correta. As variáveis que são significativas e que podem ser combinadas para então gerar um resultado conclusivo, só podem ser escolhidas e manipuladas através de um modelo matemático. É com este modelo que se pode prever por exemplo, a receita de cada cidade que pode receber a nova loja, receita esta, que servirá como indicador para a tomada de decisão dos investidores.

Tomar decisões subsidiadas em informações limitadas ou superficiais, compreendem maior possibilidade de desperdício de recursos ou taxas menos atrativas de crescimento. O modelo em questão, portanto, forneceu informações quantitativas, passíveis de serem aplicadas em análises econômico-financeiras para a tomada de decisão, justificando este estudo.

A definição das cidades do Estado do Rio Grande do Sul, que tem as melhores condições de receber uma nova filial, deu-se de acordo com seu perfil social e econômico. Fez-se necessário, em um primeiro plano, construir o perfil da empresa citada, conforme as cidades em que já atua. Assim, tornou-se possível comparar variáveis para prever os resultados que a organização poderá obter em outros municípios gaúchos. Partindo-se da necessidade de definir quais cidades podem proporcionar o maior retorno financeiro possível, foram ranqueadas todas as cidades gaúchas, selecionadas para o modelo, de acordo com a previsão de receita que a empresa pode obter em cada uma delas.

Diante da necessidade de avaliar o quão maior será a receita de uma loja disposta em uma nova cidade em relação às demais possibilidades, surge o problema central: Como determinar as melhores cidades para abrir novas filiais?

O objetivo geral deste estudo consiste em identificar as cidades do Estado do Rio Grande do Sul com maior capacidade de gerar receitas, a partir da técnica preditiva, econômica e estatística, Regressão Linear Múltipla.

Para atingir o objetivo geral, definiram-se os seguintes objetivos específicos:

a) Identificar o perfil das cidades gaúchas em que a rede varejista de moda

atua, através de variáveis qualitativas e quantitativas;

- b) Coletar os dados das demais cidades gaúchas onde a rede ainda não possui lojas;
- c) Desenvolver um modelo de Regressão Linear Múltipla;
- d) Aplicar os dados das cidades selecionadas na equação resultante do modelo de regressão;
- e) Gerar um relatório de recomendação, com base nos resultados numéricos, das melhores cidades para alocação de recursos para a abertura de novas filiais.

O desenvolvimento deste trabalho possibilitou avaliar algumas hipóteses pré-estabelecidas, visto que o modelo fornece indicadores que atestam ou não, significância para todas as variáveis selecionadas. A exemplo, espera-se que a população e a receita de vendas tenham relação linear positiva, ou seja, conforme a população aumenta de uma cidade para a outra, a receita também aumenta.

Outro indicador que teoricamente apresenta relação linear, é a renda média per capita, onde cidades que possuem maior renda, possibilitam maior consumo e consequentemente maior receita para as empresas que realizam atividades nestas cidades. Direcionando a análise para o perfil de clientes, o público com maior participação nas vendas da empresa cujo estudo se aplica, é feminino. Portanto, o modelo também subsidiou informações suficientes para confirmar se em cidades que possuem um público feminino maior, a receita de vendas também será maior. Estes testes e seus resultados serão abordados posteriormente na apresentação e análise dos dados.

Em sucessão a presente introdução, falou-se brevemente sobre o varejo, que representa o macro contexto onde a empresa, objeto deste estudo, está inserida, bem como do segmento de moda *Fast Fashion*, foco de suas atividades. Em seguida, foi pautado o processo de tomada de decisão, visando apresentar um entendimento das etapas pelas quais os tomadores de decisão devem guiar-se para atingir seus objetivos de forma mais eficiente. Por fim, fora apresentada a metodologia utilizada no desenvolvimento da pesquisa proposta, bem como os dados e as análises, seguido das recomendações e considerações finais.

## 2 REFERENCIAL TEÓRICO

Em busca da melhor compreensão do tema abordado, inicialmente, coloca-se uma inserção do macro contexto em que a empresa, objeto deste estudo, atua, bem como de seu segmento de mercado. Posteriormente, aborda-se o problema de negócio, oriundo da visão da organização, seguido de uma breve explanação sobre o processo de tomada de decisão. Ao fim, expõe-se o modelo proposto, para subsidiar através de números, as definições dos decisores.

### 2.1 VAREJO

O varejo ou as vendas no varejo, são àquelas realizadas diretamente ao consumidor final. Basicamente, são os varejistas que fazem o papel de intermediar a relação entre produtor e consumidor. Esta relação traz uma grande facilidade para ambas as partes. As grandes redes varejistas disponibilizam uma ampla variedade de produtos, de diversas marcas e de inúmeros produtores. Assim, é possível colocar os produtos mais próximos dos consumidores, o que seria difícil em um mercado onde a demanda do consumidor final tivesse que ser atendida diretamente pelos produtores.

No mercado há dois principais modelos de varejo: as lojas multimarcas como por exemplo o Walmart, e as lojas próprias, cujo exemplo mais conhecido é a Apple. Segundo Levy e Weitz (2008), as lojas multimarcas oferecem um serviço característico aos consumidores, proporcionando sortimento, variedade e divisão de grandes lotes em proporções menores. Esses fatores fazem com que este modelo seja predominante no comércio mundial.

Ainda segundo Levy e Weitz (2008), o modelo atual de varejo caracteriza-se pela alta segmentação. Farmácias, concessionárias de veículos, lojas de materiais de construção, lojas de móveis, livrarias, lojas de moda, entre outros, são exemplos de negócios que exploram um departamento de produtos bem especificado. Neste cenário também é possível identificar algumas grandes empresas, com alto poder de mercado, que atuam com um sortimento muito grande de mercadorias, bem como, com um número elevado de departamentos à disposição dos clientes em um único estabelecimento comercial.

Os números do varejo traduzem o consumo no mundo. Assim, este setor é responsável pela alta contribuição de empregos formais no mercado de trabalho mundial. Este setor ainda é um dos que mais vem empregando práticas e processos tecnológicos nos últimos anos. A enorme quantidade de dados sobre os consumidores e seu comportamento é alimento para análises cada vez mais individualizadas, ao contrário do que até então era tido como modelo analítico: a criação de clusters. A análise personalizada, com uso de big data e inteligência artificial, possibilita oferecer produtos de forma mais assertiva para cada cliente, bem como criar ambientes integrados (*online* e *offline*), o que o SEBRAE (2017) coloca como *omnichannel*, que proporciona experiências diferenciadas com foco total nos clientes.

### **2.1.1 Varejo no Brasil**

Não distante das características do comércio exterior, o comércio varejista brasileiro é o setor com o maior número de empregos formais, atrás é claro, do setor público, segundo o Cadastro Geral de Empregados e Desempregados (CAGED). Confirmando a importância do setor na economia, segundo dados da Varejo s.a. (2017), o impacto no Produto Interno Bruto (PIB) é de cerca de 47%.

No Brasil, o emprego de tecnologia para garantir o crescimento das vendas, é realidade cada vez mais presente, principalmente nas grandes organizações. As empresas, por vezes, não percebem o potencial das informações disponíveis sobre os clientes com os quais mantém relacionamento. Tudo é informação. Histórico de compras, de pagamentos, tipos de produtos, perfil, interação com a marca, e tantos outros que constituem o *Customer Relationship Management* (CRM), podem ser usados para gerar valor com cada cliente.

O crescimento do setor no país demonstra a grande capacidade de desenvolvimento e especialização. A níveis baixos de produtividade da força de trabalho em relação a outros países, as empresas que atuam neste setor precisam se tornar especialistas em determinadas categorias de produtos ou expandir o sortimento, a fim de garantir ou aumentar sua participação no mercado.

As dificuldades encontradas no país se resumem, portanto, a uma infraestrutura deficitária, carência de estudos e escassez de pessoas qualificadas. Fazer com que os clientes sintam a mesma experiência no físico ou no digital, e que

a empresa consiga um atendimento com a máxima personalização, é tarefa crucial para manter-se à frente do mercado, impondo quebra de paradigmas e disrupção.

## 2.2 PROBLEMA DE NEGÓCIO

De acordo com Costa (2009), quando uma organização possui um planejamento estratégico, há claramente uma visão e um posicionamento competitivo, com objetivos e definições estruturadas. Portanto, também há clareza quanto ao problema de negócio, bem como, já se conhecem algumas das possibilidades que terão capacidade de conduzir a organização ao sucesso.

O problema de negócio do presente estudo, é observado ao deparar-se com a visão da empresa. Nesta, a organização assume o compromisso, de no futuro, estar em uma determinada posição. Segundo Dal'Bó (2009), “a visão é a descrição do futuro desejado para a empresa”. Assim, a organização a qual este estudo de caso se refere, se propõe a alcançar até 2020, 80 lojas. Para que no futuro a empresa possa atingir este objetivo, é necessário desde já, um planejamento estratégico, o qual apresentará, de forma estruturada, as alternativas e soluções para se obter êxito.

O objetivo de expansão pode parecer simples. Porém, o problema central é definir onde estas novas lojas deverão ser colocadas. Isso requer informações complexas, pois o futuro da empresa pode estar ameaçado se os recursos que serão dispendidos para esta expansão forem perdidos.

Assim sendo, uma das propostas de contribuição para a tomada de decisão é elaborar um relatório que demonstre, em números, onde é possível obter mais ou menos receitas, a partir de uma análise preditiva. Esta técnica pode subsidiar a tomada de decisão com informações mais sustentadas, já que considera dados reais de uma base histórica.

Determinado o problema de negócio, o importante é buscar alternativas coerentes para solucioná-lo. Uma análise fundamentada proporcionará maior facilidade na hora de definir qual o melhor caminho, buscando maximizar os resultados e minimizar as perdas. Portanto, segundo Gomes e Gomes (2014), para direcionar corretamente todo o processo, é preciso concentrar-se no problema correto. Ainda segundo os autores, após o entendimento dos vários aspectos do problema, diferentes cenários podem ser construídos, os quais são passíveis de materialização.

## 2.3 TOMADA DE DECISÃO

Segundo Mc Guigan, Moyer e Harris (2010), para se obter sucesso e bons resultados a nível gerencial, é indispensável possuir alta capacidade de tomar boas decisões. Existem algumas definições acerca da decisão. Yu (2011) afirma que “tomar decisão é alocar irreversivelmente recursos”. Sua justificativa se dá pelo fato de que ao reverter algum processo de decisão, irão necessariamente ocorrer perdas.

Ainda segundo Yu (2011), no decorrer dos dias, inúmeras decisões são tomadas, e todos tomam decisões, sejam elas operacionais, táticas ou estratégicas. Por vezes, algumas podem parecer insignificantes, enquanto outras, podem marcar a história e traçar novos rumos. Neste momento, em algum lugar, algum executivo pode estar se preocupando com uma importante decisão de negócio a ser tomada, sendo que, geralmente, sofre influências do ambiente, como por exemplo dos acionistas ou de stakeholders.

Segundo Turban et al. (2009), os fatores do ambiente de negócio, que exigem agilidade e inovação dos tomadores de decisão, podem ser enquadrados em categorias. Abaixo, destaca-se os fatores que causam pressão nas organizações:

- Mercados: Forte concorrência, expansão de mercados globais, prosperidade dos mercados eletrônicos, métodos inovadores de marketing, oportunidades de terceirização com suporte de TI, necessidade de transações em tempo real e sob demanda;
- Demandas do consumidor: Deseja customização, deseja qualidade e diversidade de produtos e rapidez na entrega, os clientes tornam-se poderosos e menos fiéis;
- Tecnologia: Mais inovações, novos produtos e serviços, a crescente taxa de obsolescência, a crescente sobrecarga de informações;
- Sociedade: O aumento das regulamentações e desregulamentações governamentais, mão-de-obra diversificada, segurança nacional, legislação sobre a prestação de contas, crescimento da responsabilidade social das empresas.

Naturalmente todos passam por momentos de decisão. Este processo sempre ficará visível quando mais de uma possibilidade estiver disponível. É de comum acordo que o mais importante é a qualidade das decisões que são tomadas. Isto explica e pode garantir o sucesso de grandes organizações, como por exemplo a

multinacional Apple, atuante no mercado de eletrônicos e *softwares*, ou de pessoas como o famoso investidor americano Warren Buffett.

### 2.3.1 Processo de Tomada de Decisão

Segundo Robbins (2015), ao passo que mais de uma alternativa se coloca disponível, seguramente, busca-se escolher àquela que proporcionará maior resultado. Para isso, é necessário que se tenha informação de qualidade, a fim de suprir as eventuais dúvidas do processo de escolha, as quais podem ser fornecidas por um órgão de staff da empresa, ou seja, um grupo de pessoas qualificadas para gerar informação e dar suporte aos níveis mais altos da hierarquia organizacional, onde se encontram os tomadores de decisão.

Robbins (2015) esclarece ainda que a importância do planejamento estratégico e da análise clara e precisa, ao colocar que o processo de tomada de decisão é que pode ser controlado. Já seus resultados, não. Portanto, é indispensável a participação de influenciadores, que com seu conhecimento enriquecem o processo, e de analistas, que possuem a capacidade de gerar informação e dados fidedignos para sustentar as decisões a serem tomadas.

Para uma tomada de decisão racional, Robbins (2015) define seis etapas fundamentais, as quais estão descritas no quadro abaixo:

Quadro 1 – Etapas para a tomada de decisão

Etapa	Exemplificação relacionada ao estudo
1. Identificação e definição do problema	O problema central é escolher quais cidades devem receber as novas lojas, a fim de maximizar o resultado.

2. Identificação dos critérios de decisão	O critério evidenciado é justamente o resultado do modelo preditivo, ou seja, a receita é o critério de escolha entre uma cidade ou outra.
3. Atribuição de pesos aos critérios	Buscando a maximização de lucros, certamente a previsão de receita de vendas tem alta relevância no processo de decisão, já que este pode subsidiar os estudos de retorno sobre o investimento esperado pela organização.
4. Geração de alternativas	Aqui, as alternativas de maior potencial devem ser analisadas. Visto que o modelo sugerido para subsidiar a tomada de decisões proporciona apenas uma previsão, outros fatores relevantes podem ser considerados na busca pela melhor alternativa, complementando o resultado do atual estudo.
5. Avaliação das alternativas	Aqui coloca-se a necessidade de ser mais criterioso quanto a análise, sempre considerando as definições das etapas anteriores.
6. Definição da melhor alternativa	A partir de toda a análise, a decisão aqui já pode ser tomada, visto que há uma recomendação que passou por todo o processo e tem um diferencial em relação às demais possibilidades.

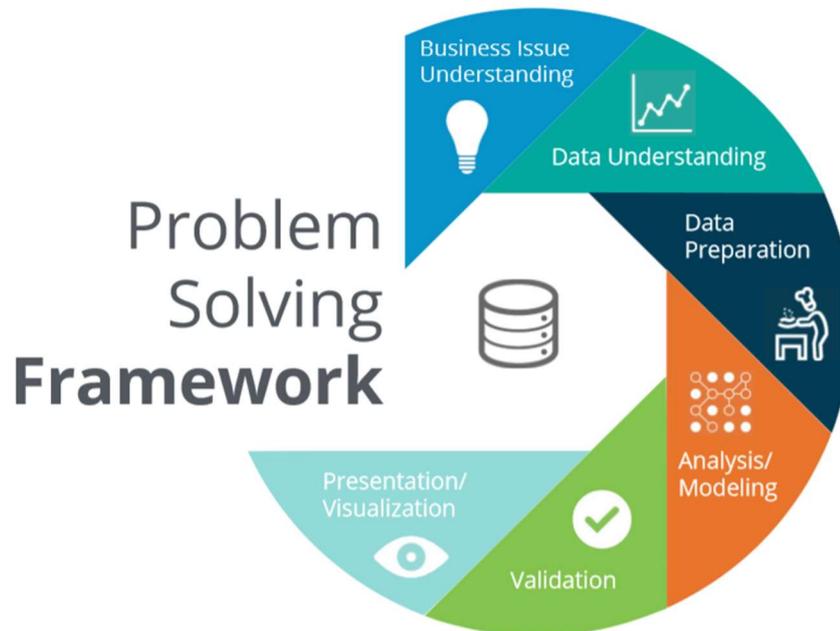
Fonte: Robbins, 2015.

Almeida (2013), coloca a importância da execução do processo a partir da tomada de decisão. Para o autor, alguns processos que levam a escolha podem ser demasiadamente simples, porém, com uma implementação complexa. O inverso também pode ser verdadeiro. Isso ocorre principalmente por que é neste momento que a realidade e o contexto organizacional irão sofrer mudanças.

## 2.4 FRAMEWORK DE RESOLUÇÃO DE PROBLEMAS

Buscou-se delinear o processo de solução do problema de pesquisa com o uso da ferramenta conhecida como CRISP-DM (*Cross Industry Standard Process for Data Mining*). A figura a seguir ilustra as etapas do processo que possibilitou maior clareza quanto ao entendimento do problema de negócio, bem como da busca por sua solução.

Figura 1 – CRISP-DM



Fonte: Udacity, 2018.

Segundo a Udacity (2018), o *framework* de resolução de problemas CRISP-DM consiste nas seguintes etapas: entender o negócio e o problema empresarial, entender os dados, preparar os dados, analisar e modelar uma metodologia, validar os resultados e por fim apresentar os resultados.

De acordo com a Vorhies (2016), na década de 1990, os principais fornecedores de ferramentas de dados, SPSS e Teradata juntaram-se com mais algumas empresas, em um grupo chamado SIG (Grupo de Interesses Especiais), do qual resultou o processo de indústria cruzada para mineração de dados, conhecido como CRISP-DM, que é utilizado até hoje.

A Udacity (2018) expressa bem as informações cruciais em cada uma das etapas na aplicação do modelo, conforme o quadro abaixo:

Quadro 2 – Etapas do *framework* de resolução de problemas CRISP-DM

1. Entender o negócio	Aqui busca-se respostas para quais decisões devem ser tomadas, quais as informações necessárias para embasar as decisões e que tipos de análises podem fornecer as informações necessárias para embasar essas decisões.
2. Entender os dados	Nesta etapa busca-se entender quais dados serão necessários, quais destes dados estão disponíveis e quais são as características importantes dos dados.
3. Preparação dos dados	Esta etapa consiste em coletar, limpar, formatar, combinar, buscar amostras ou conjuntos de dados.
4. Análise/Modelagem	Aqui determina-se qual a metodologia a ser usada para resolver o problema, bem como fatores ou variáveis que ajudarão neste processo. É nesta fase também que o modelo é construído e analisado para então partir para a validação.
5. Validação	Aqui é importante visualizar os resultados chave do modelo e certificar-se de que os mesmos fazem sentido dentro do contexto do problema. O processo de análise e modelagem, pode ser repetido inúmeras vezes sempre que o processo de validação apontar ajustes no modelo.
6. Apresentação/Visualização	A última etapa consiste em montar a melhor maneira para apresentar a análise, visando seu público alvo. A quantidade de informação não deve ultrapassar o suficiente, demonstrando o processo, os métodos e fontes utilizados.

Fonte: Udacity, 2018.

## 2.5 MODELO PREDITIVO

De acordo com a Udacity (2017), como o próprio nome sugere, um modelo preditivo é um modelo utilizado para prever o futuro. Com base em um conjunto de dados, o modelo pode indicar tendências e apresentar resultados que podem ser utilizados em estratégias futuras. Já para Guazzelli (2012), em um conceito mais técnico, um modelo preditivo pode ser caracterizado como uma função matemática, que produz uma resposta ou destino a partir da leitura de uma entrada de dados. Para a construção de modelos preditivos, é imprescindível a utilização de ferramentas de *Analytics* e *Big Data*. Segundo Coelho (2017), *Analytics* se refere a utilização de dados e a análise destes para a tomada de decisão, enquanto *Big Data* pode ser entendido como um grande volume de dados, sejam eles estruturados ou não.

Segundo a Udacity (2018), as previsões são amplamente utilizadas em planejamento estratégico, onde estas podem fornecer por exemplo, receita e demanda futuras, o que por sua vez pode proporcionar a possibilidade de organizar a produção e os estoques, a fim de evitar rupturas, ou seja, quando há falta de produtos frente a demanda, ou sobras, quando a demanda tende a ser menor.

Um modelo preditivo deve ser elaborado com máxima capacidade analítica. Dados errôneos ou faltantes por exemplo, podem gerar previsões erradas, que por consequência subsidiará decisões erradas, que por fim, trará resultados negativos para as organizações. Kaplan e Norton (1997), afirmam que os processos realmente estratégicos são aqueles que proporcionam resultados excepcionais para que a estratégia da empresa seja bem-sucedida.

Tomando como exemplo o próprio estudo em questão, pode-se imaginar que diversas variáveis influenciam na receita possível que uma nova filial pode gerar. Certamente a renda é uma destas variáveis. Portanto, pode-se apurar, através de um modelo preditivo, qual seria a receita provável em uma cidade de acordo com a renda média de sua população. Este resultado é possível, visto que, pode-se observar quanto de receita adicional pode ser obtida com o incremento de uma unidade monetária a mais na renda. Neste caso, deve-se levar em conta o conceito *Ceteris Paribus*, ou seja, considerar que todas as demais variáveis permanecerão constantes. Este conceito é fundamental para avaliar o impacto exclusivo da variável preditora sobre a variável alvo, pois sabe-se que as decisões de compra dos clientes são influenciadas por diversos fatores que até então não estão presentes no modelo.

Diversos estudos podem ser criados com o uso de modelos preditivos, sejam eles puramente estatísticos ou econômicos, e independente do setor a ser analisado.

No setor varejista, por sua vez, a análise preditiva torna possível prever a demanda por um produto ou serviço, planejar campanhas e eventos promocionais mais apropriados para os potenciais clientes, bem como determinar os produtos que devem ou não ser estocados e desenvolver estratégias de fidelidade. Além disso, com esse tipo de análise, é possível detectar e reduzir fraudes do uso de cartões de crédito, dar mais segurança às operações bancárias e financeiras e, conseqüentemente, ampliar as oportunidades de vendas (Udacity, 2017).

A análise preditiva já é realidade a algum tempo nas grandes organizações, principalmente onde existe um ambiente tecnológico. Atualmente, as ferramentas de análise estão ao alcance de todos, e beneficiam não mais somente as grandes organizações, mas todas aquelas que tem a expansão do negócio como visão, transformando grandes conjuntos de dados disponíveis em informação para tomada de decisão.

### 2.5.1 Modelo Econométrico

Para Gujarati (2000), Econometria significa literalmente “medida econômica”. Esta medida econômica é obtida através de modelos econométricos. Os modelos econométricos basicamente demonstram resultados numéricos oriundos da aplicação da estatística e da matemática a um conjunto de dados. Conforme observado anteriormente, em teoria, quando há um aumento da renda, há um aumento no consumo, ou seja, uma relação direta entre estas duas variáveis. É através da econometria que as teorias econômicas podem ser demonstradas matematicamente, ou seja, conforme Gujarati (2000), é a econometria que propicia empirismo às teorias econômicas.

Segundo Kmenta (1978), um dos conceitos mais importantes no campo da econometria é o da inferência estatística. De acordo com o autor, a partir dos resultados de uma amostra, infere-se uma generalização sobre a população. Tratando-se de uma população infinita, não há possibilidade de observá-la como um todo, sendo possível apenas a observação de uma amostra. Mesmo que uma população seja finita, ainda assim, possivelmente uma amostra desta deverá ser considerada, já que sua observação seria por vezes inviável, se a mesma for de tamanho relativamente grande, ou se há escassez de recursos para a observação em sua totalidade. Considerando que uma amostra geralmente evidencia o comportamento de uma população da qual foi retirada, a inferência sobre o todo deverá proporcionar dados analíticos sustentados, dada a formulação correta do modelo preditivo.

Um dos exemplos mais conhecidos de modelo econométrico é o modelo de regressão linear. Neste modelo, segundo Gujarati (2000), supõe-se que a relação entre as variáveis preditora e preditiva, quando houver, andam em uma única direção. Ainda segundo o autor, a análise de regressão estuda a dependência de uma variável (variável dependente, preditiva ou ainda variável alvo) em relação a uma ou mais variáveis (variáveis explicativas ou preditoras). Portanto em nosso exemplo de renda e consumo, visto anteriormente, a variável receita é a variável explicada, enquanto a renda é a variável explicativa. Esta relação fica mais clara ao observarmos que o consumo, na teoria econômica, depende da renda. Assim, aumentos na renda podem significar aumento no consumo, bem como, o inverso, também é verdadeiro.

### 2.5.2 Metodologia do Modelo Econométrico

A metodologia tradicional para o desenvolvimento de um modelo econométrico segue algumas etapas, as quais trazem a teoria, os dados e o modelo em si.

De acordo com Gujarati (2000), o primeiro passo é a formulação da teoria. Dando continuidade ao exemplo citado anteriormente, a renda da população os coloca diante de uma linha orçamentária. Segundo Pindyck e Rubinfeld (2002), a restrição orçamentária é uma condição que os consumidores enfrentam por sua renda ter um caráter limitado. Assim sendo, o aumento da renda pode deslocar esta linha de restrição e permitir que os consumidores aumentem o seu poder de compra. Com base nesta teoria, podemos então utilizar um modelo econométrico, mais tecnicamente um modelo de regressão, para prever qual o aumento no consumo para cada unidade monetária de elevação na renda.

A segunda etapa, segundo Gujarati (2000), consiste em especificar o modelo matemático. Isso requer o desenvolvimento de uma equação, que pode ser representada da seguinte maneira:  $\hat{Y} = \beta_0 + \beta_1 x_1$ . Nesta equação,  $\hat{Y}$  é o objetivo, ou seja, o consumo dado um determinado nível de renda,  $\beta_0$  e  $\beta_1$  são os parâmetros do modelo, e  $x$  o nível de renda. Os parâmetros podem ser resumidos em consumo que independe da renda ( $\beta_0$ ) e consumo dependente do nível de renda ( $\beta_1$ ). Esta equação originará de um conjunto de dados históricos, como por exemplo, a base de dados de consumo e renda dos últimos 10 anos de um determinado país.

A próxima etapa consiste na especificação do modelo econométrico. Gujarati (2000) coloca que a função matemática criada apenas supõe uma relação exata entre as variáveis. Considerando as variáveis consumo e renda, sabe-se que existem outros fatores que influenciam as decisões dos agentes, alterando, portanto, o comportamento destes em relação ao consumo. Assim sendo, segundo o autor, há a inclusão do fator de erro ( $u$ ), que pode representar a oscilação do consumo em função das demais variáveis que não são considerados no modelo. Uma nova equação pode ser assumida:  $\hat{Y} = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + u$ .

Gujarati (2000) considera que, dada a equação e seus resultados, chega o momento de testar a teoria. Ao considerar que o aumento da renda provoca aumento no consumo, e que outras variáveis têm impacto sobre este indicador, empiricamente, o aumento de uma unidade adicional de renda pode no máximo proporcionar o

aumento de uma unidade monetária adicional no consumo. Fato é que as pessoas podem tomar a decisão de poupar ou investir por exemplo. Isso, faz com que o crescimento do consumo seja necessariamente menor que o crescimento da renda, embasando e validado assim a teoria econômica inicial.

Agora, com a validação das hipóteses, pode-se dar início a previsão dos resultados esperados. Com resultados futuros de renda, torna-se possível, em nosso exemplo, estimar o consumo para este mesmo período. Assim, consumidores, empresas e governo por exemplo, tem a possibilidade de elaborar um planejamento que atenda suas necessidades, seja de custos, volume de produção, receitas ou objetivar outros indicadores que tenham relação com a análise elaborada.

### 3 METODOLOGIA

Para atingir os objetivos, a técnica de pesquisa do presente trabalho, classificou-se como pesquisa qualitativa e exploratória, que segundo Prandanov e Freitas (2013), tem como finalidade proporcionar mais informações sobre o assunto investigado, possibilitando sua definição e seu delineamento, isto é, facilita a delimitação do tema de pesquisa, orienta a fixação dos objetivos e a formulação das hipóteses, ou proporciona a descoberta de um novo tipo de enfoque para o assunto. Assume, em geral, as formas de pesquisas bibliográficas e estudos de caso.

Quanto ao método de abordagem, caracterizou-se pelo método dedutivo. Para Prandanov e Freitas (2013), o raciocínio dedutivo vai do geral ao particular para chegar a uma conclusão, a partir de princípios reconhecidos como verdadeiros. Para Gil (2002), as conclusões são obtidas formalmente e de acordo com a lógica.

O presente trabalho, delineou-se, quanto aos procedimentos, sob a ótica de um estudo de caso, visto ser um estudo empírico que visa explorar uma situação real de tomada de decisão, onde fez-se necessário a análise de variáveis de causa e efeito sobre o negócio. Ainda, segundo Lakatos e Marconi (2001), o estudo de caso é caracterizado pelo estudo profundo e exaustivo de um ou de poucos objetos, de maneira a permitir seu conhecimento amplo e detalhado, tarefa praticamente impossível mediante os outros tipos de delineamentos considerados. Com a análise das variáveis qualitativas e quantitativas, buscou-se construir o perfil dos municípios gaúchos, os quais representam o universo deste estudo.

Quanto às técnicas de pesquisa, deu-se por documentação indireta, que segundo Lakatos e Marconi (2001), é a fase da pesquisa realizada com o intuito de recolher informações prévias sobre o campo de interesse. Em sua subdivisão, foi de caráter documental, visto a utilização de fontes primárias como o IBGE, RAIS e CAGED.

A obtenção dos resultados se deu a partir de um modelo preditivo, que segundo Guazzelli (2012), pode ser caracterizado como uma função matemática, que produz uma resposta ou destino a partir da leitura de uma entrada de dados. Este modelo foi sintetizado em um modelo de Regressão Linear Múltiplo, conceito presente dentro da Econometria, ramo da economia que segundo Gujarati (2000), significa literalmente

“medida econômica”, ou seja, demonstra um resultado numérico oriundo da aplicação de estatística e matemática à um conjunto de dados.

A manipulação dos dados, que compreende a tabulação, limpeza, formatação e o armazenamento dos dados, deu-se com o uso do *software* de planilha eletrônica Microsoft Excel e com o software de *business intelligence* Microsoft Power BI. A partir da formulação do conjunto de dados com o uso das ferramentas citadas, o modelo de regressão linear múltipla, bem como a análise de correlações deu-se com o uso dos *softwares IBM SPSS e Alteryx Designer*. Estes dois *softwares* proporcionaram a análise do modelo, através dos resultados estatísticos gerados e da plotagem de gráficos de comparação de variáveis.

A análise dos dados se deu através de ilustrações gráficas, fluxogramas, índices e tabelas de dados, o que proporcionou a demonstração da significância estatística das variáveis selecionadas, dos resultados preditivos e das recomendações finais.

## 4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

A caracterização da empresa deu-se em função da primeira etapa do CRISP-DM. Ao entender a visão da organização, pôde-se entender o problema de negócio a ser solucionado. A utilização do *framework* de resolução de problemas possibilitou uma análise mais assertiva, bem como a construção de um modelo conciso.

A fase seguinte à caracterização da empresa é a de coleta dos dados que foram utilizados como variáveis no modelo preditivo. A utilização de uma boa fonte de dados é extremamente importante para que se tenham resultados fidedignos. De acordo com a Wikipédia (2015), George Fuechsel, técnico da IBM, teve atribuída a si a expressão “lixo entra, lixo sai”, ao afirmar a importância de um ótimo conjunto de dados para realização de análises e construção de modelos. Nesta fase foram consideradas as etapas 2 e 3 do CRISP-DM.

Na etapa 4 do *framework* de resolução e problemas concluiu-se que um modelo de regressão linear múltiplo poderia fornecer os resultados almejados. Portanto, nesta fase, o modelo fora construído testando diversas variáveis que, em teoria pudessem ter a capacidade de explicar o comportamento das vendas onde a empresa em estudo atua.

Por fim, após diversos testes, chega-se a uma conclusão, cujos resultados são expressados nas considerações finais, após passarem pela etapa de validação, que precede a apresentação e as recomendações, fechando o ciclo do CRISP-DM.

### 4.1 CARACTERIZAÇÃO DA EMPRESA

A empresa, objeto deste estudo, atua a 54 anos no estado do Rio Grande do Sul, no ramo varejista de moda. Neste período, expandiu suas atividades e atualmente está presente em 62 municípios gaúchos, totalizando 64 lojas. Em sua visão, almeja atingir 80 lojas até o ano de 2020. Assim, esta análise busca auxiliar os tomadores de decisão a definir a localização de novas lojas nos próximos anos.

No atual cenário, o foco da empresa é a moda *fast fashion*, direcionada ao público jovem. As atuais alterações no modelo de negócio visam colocar a organização mais próxima deste público alvo. Sua missão, portanto, está pautada em

ter a sua marca com este reconhecimento, proporcionando a seu público buscar informações e consumir moda.

#### 4.2 IDENTIFICAÇÃO DO PERFIL DAS CIDADES GAÚCHAS ATRAVÉS DA COLETA DE DADOS

A caracterização das cidades leva em consideração indicadores qualitativos, mas principalmente quantitativos, os quais resultam de fontes governamentais, mais especificamente o IBGE (Instituto Brasileiro de Geografia e Estatística) e a RAIS (Relação anual de informações sociais). Após avaliar a segunda etapa do framework de resolução de problemas, identificou-se que os indicadores econômico-sociais dos municípios poderiam fornecer informações necessárias para explicar as variações no faturamento das filiais de acordo com cada município.

Para criar a entrada de dados, ou seja, os dados preditores, foram reunidos os seguintes indicadores dos 62 municípios de atuação da empresa em estudo e dos demais 435 municípios gaúchos em que a rede ainda não atua:

- População estimada para 2017 de acordo com o PNAD;
- Percentual de população residente em zona urbana de acordo com o PNAD 2017;
- Número de empregados com vínculo ativo em 31/12/2017 segundo dados da RAIS;
- Número de empregados do sexo masculino com vínculo ativo em 31/12/2017 segundo dados da RAIS;
- Número de empregados do sexo feminino com vínculo ativo em 31/12/2017 segundo dados da RAIS;
- Idade média dos trabalhadores com vínculo ativo em 31/12/2017 segundo dados da RAIS;
- Idade média dos trabalhadores do sexo masculino com vínculo ativo em 31/12/2017 segundo dados da RAIS;
- Idade média dos trabalhadores do sexo feminino com vínculo ativo em 31/12/2017 segundo dados da RAIS;
- Remuneração média dos trabalhadores com vínculo ativo em 31/12/2017 segundo dados da RAIS;

- Número de domicílios privados urbanos segundo dados do PNAD 2017;
- Número de domicílios privados rurais segundo dados do PNAD 2017;
- PIB per capita segundo dados do IBGE;
- Área territorial segundo dados do IBGE;
- IDESE segundo dados do IBGE;
- Valor em poupança segundo dados do IBGE;
- Valor das operações de crédito segundo dados do IBGE;
- Valor dos depósitos a prazo segundo dados do IBGE;
- Arrecadação de tributos segundo dados do IBGE;
- Valor das despesas realizadas pelo governo segundo dados do IBGE;
- Número de estabelecimentos segundo dados do IBGE;
- Número de estabelecimentos comerciais segundo dados do IBGE;
- Razão entre o número de empresas comerciais e a população, calculado com base nos dados do IBGE.
- Faturamento anual da empresa, ao ano base 2017.

A escolha das variáveis deu-se ao estimar que os indicadores do ambiente macro dos municípios, ou seja, o que Hill, Griffiths e Judge (2006) colocam como um nível de agregação, pois os dados resultam de uma combinação de indivíduos ou firmas. Assim, estes dados poderiam fornecer uma conjuntura capaz de expressar um perfil e registrar um determinado comportamento nos resultados das filiais, positiva ou negativamente, sempre que alteradas conforme o comportamento dos agentes do mercado.

Em sequência a coleta dos dados, verificou-se eventuais anomalias nos dados, como erros e dados faltantes, para que não comprometessem o desenvolvimento das etapas seguintes. Em virtude de a empresa em estudo ser de capital fechado, os dados respectivos a ela não foram divulgados, preservando as informações a seu domínio.

### 4.3 DESENVOLVIMENTO DO MODELO ECONOMÉTRICO DE REGRESSÃO LINEAR MÚLTIPLA

Como visto anteriormente, Gujarati (2000) expressa que um modelo econométrico resulta em uma medida econômica, dado um modelo elaborado com a aplicação da estatística e da matemática. Segundo Rossi e Neves (2014), a modelagem econométrica começa com uma análise econômica teórica. Considerando este estudo, a teoria econômica implica que o desempenho operacional das empresas, ou seja, suas receitas, tendem a ser maiores onde o mercado apresenta condições melhores, com níveis mais elevados de emprego, maior renda, população economicamente ativa maior, mais investimentos por parte do governo, maior estímulo ao crédito visando fomentar o consumo entre outros. Assim, as variáveis elencadas para o desenvolvimento do modelo econométrico apresentam informações pertinentes aos aspectos da teoria econômica.

De acordo com Gujarati e Porter (2014), a análise de regressão interpreta a dependência de uma variável em relação a uma ou mais variáveis, visando prever um valor médio dos dados conhecidos. Assim, se na teoria econômica a renda tem relação linear positiva com o consumo, ou seja, o aumento da renda proporciona aumento no consumo, a análise de regressão permite prever quanto de aumento no consumo ocorrerá para cada aumento de uma unidade monetária na renda. Para este estudo, buscou-se estimar qual a receita que a empresa em questão poderá obter, dadas as informações econômico-sociais dos municípios.

#### 4.3.1 CORRELAÇÃO

Ao trabalhar com regressão, é importante fazer uma análise de correlação. Segundo Gujarati (2000), a correlação tem por objetivo medir a associação linear entre duas variáveis onde o grau de intensidade é dado pelo coeficiente de correlação. Esta, pode se dar em duas situações. A primeira delas é correlacionar todas as variáveis selecionadas para o modelo com a variável alvo. Neste caso, é importante que as variáveis tenham alta correlação, positiva ou negativa, com a variável alvo, o que dá a elas maior significância estatística em um modelo de regressão. A segunda, é a colinearidade entre variáveis, ou seja, a alta correlação entre as variáveis

selecionadas para o modelo. Assim, segundo o Portal Educação (20-), quando há alta correlação entre variáveis independentes, estas podem medir aproximadamente a mesma coisa.

A primeira fase em uma análise de correlação pode ser visual. Segundo Hoffmann (2006), ao observar uma amostra de dados **X** em ordem crescente ou decrescente, e uma determinada variável **Y**, que tende a crescer à medida que cresce a variável **X**, invariavelmente, existe uma correlação positiva neste par de valores.

Segundo Milone (2006), o coeficiente de correlação se dá em uma escala que vai de -1 a +1, onde o -1 significa uma correlação 100% negativa entre duas variáveis, ou seja, uma relação inversa, em que o aumento de uma variável **X** resulta em redução da variável **Y** e vice-versa. Já o +1 significa uma relação 100% positiva, ou seja, uma relação direta onde o aumento em uma variável **X** resulta em aumento na variável **Y** e vice-versa. Quanto mais próximo de zero é o coeficiente de correlação, menor é a relação linear entre as variáveis.

O quadro a seguir demonstra a correlação das variáveis selecionadas para o modelo frente a variável alvo:

Quadro 3 – Correlação das variáveis independentes com a variável dependente

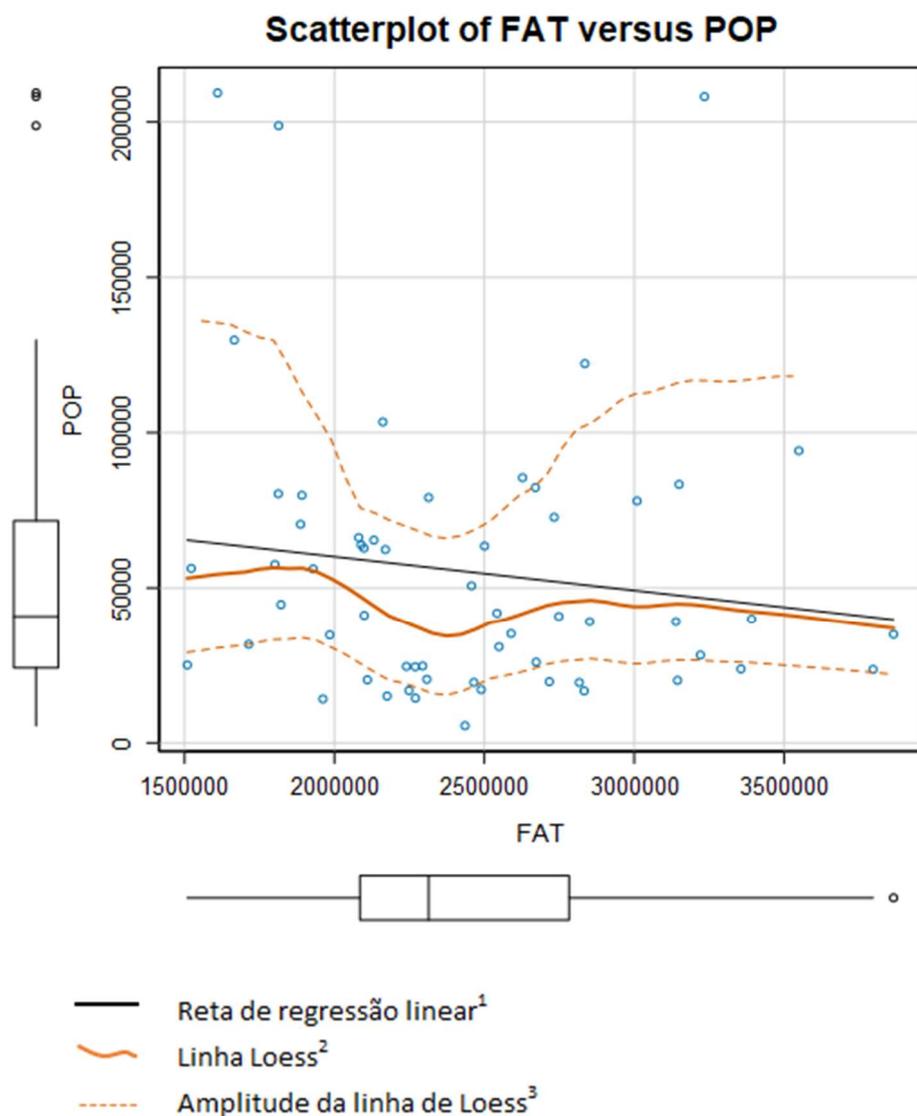
VARIÁVEIS	CORRELAÇÃO COM FATURAMENTO
POP	-14%
POP_URB	17%
EMP_FOR	-23%
EMP_FOR_F	-24%
EMP_FOR_M	-22%
IDA_MED	28%
IDA_MED_F	28%
IDA_MED_M	20%
REM_MED	-10%
DPP_URB	-15%
DPP_RUR	-23%
PPC	-2%
ARE	25%
IDE	2%
POU	-28%
OP_CRE	-20%
DEP_P	-30%
TRI	-27%
DES_REA	-27%
NUM_EST	-24%
NUM_EMP_COM	-19%
NEC_POP	-8%

Fonte: Elaborado pelo autor com uso do *software* Microsoft Excel.

Nota-se no Quadro 3 que a relação linear das variáveis independentes em relação a variável dependente é muito baixa em todos os casos. Isso implica que as alterações nas variáveis independentes pouco afetam a variável alvo, e que possivelmente pode não explicar as variações no faturamento da empresa de uma cidade para a outra.

Com o uso do *software* *Alteryx Designer* foram plotados os gráficos de dispersão para cada variável independente, em função da variável alvo. Os gráficos a seguir ilustram a dispersão de cada variável selecionada para o modelo, plotadas individualmente com o faturamento (variável alvo), com o objetivo de identificar visualmente se existe relação linear entre as mesmas.

Gráfico 1 – Faturamento versus população estimada para 2017



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

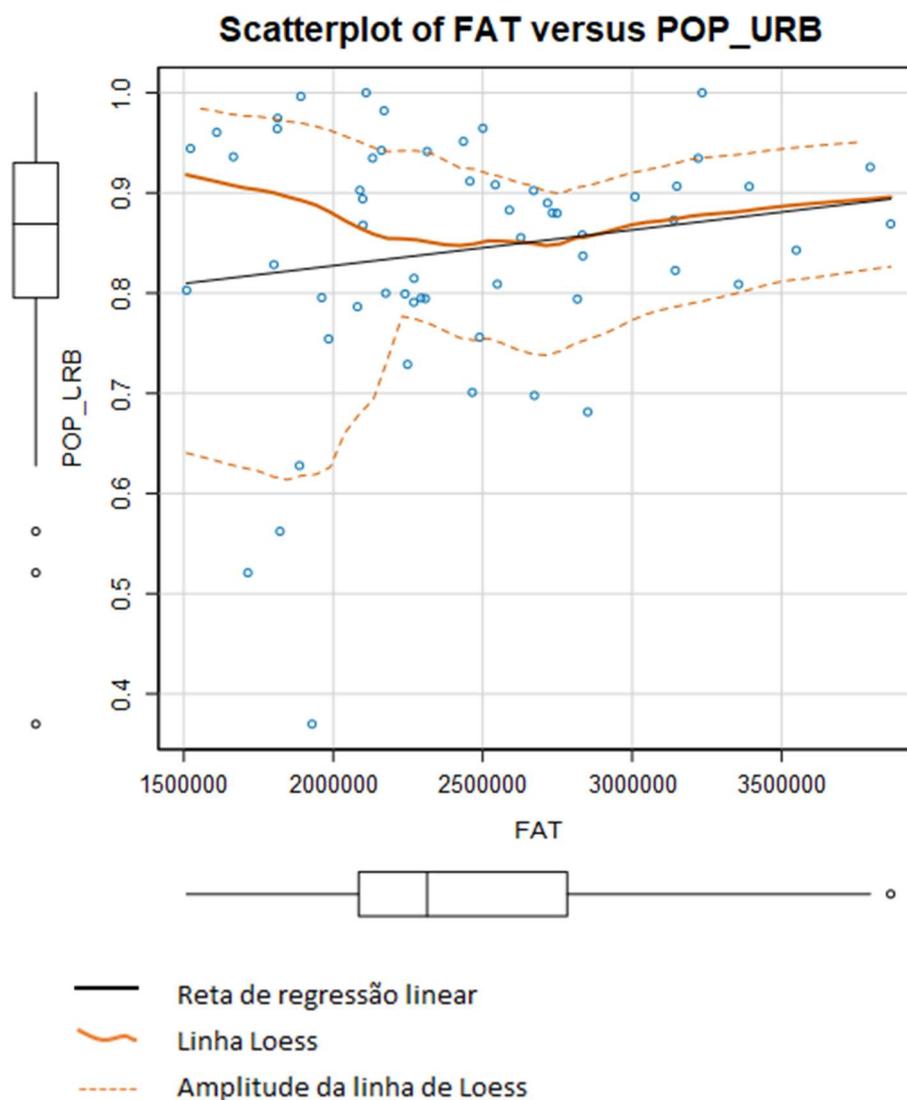
A plotagem do gráfico que mostra a relação entre a população estimada para os municípios, com o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que em cidades onde a população é maior, resulta em um faturamento maior na filial, ou seja, espera-se uma relação linear positiva, não é válida nesta observação, já que os pontos estão dispersos e não se aproximam de uma reta.

<sup>1</sup> A Reta representa a função da regressão linear.

<sup>2</sup> A linha de Loess combina a regressão linear com a regressão não linear, demonstrando uma relação determinística da variação dos dados ponto a ponto, ou seja, estima uma curva por suavização (Bonat, 2007).

<sup>3</sup> A amplitude demonstra a variabilidade, ou seja, as bandas entre as quais está determinada a linha de Loess (Bonat, 2007).

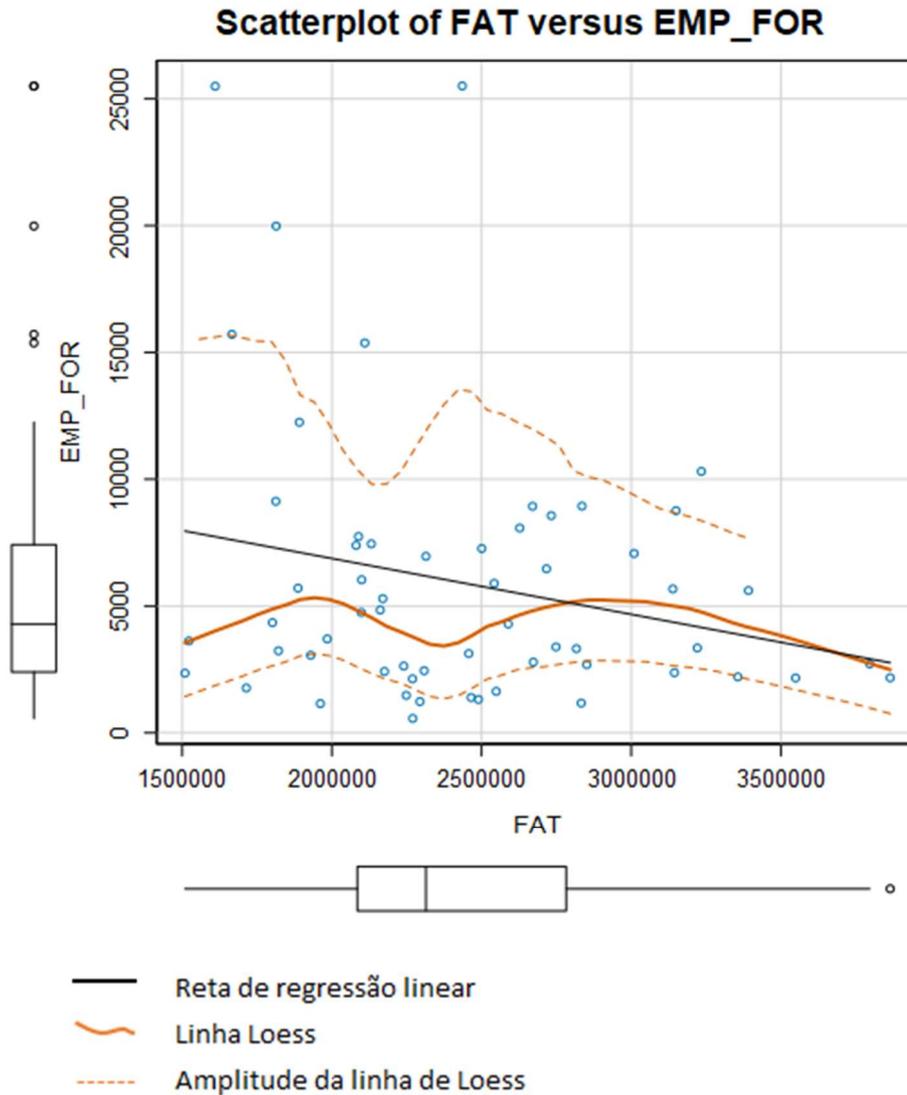
Gráfico 2 – Faturamento versus percentual de população urbana



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o percentual de população residente na zona urbana dos municípios, em relação ao faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde há mais habitantes na zona urbana, próximos ao comércio, resulta em maior faturamento na filial, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

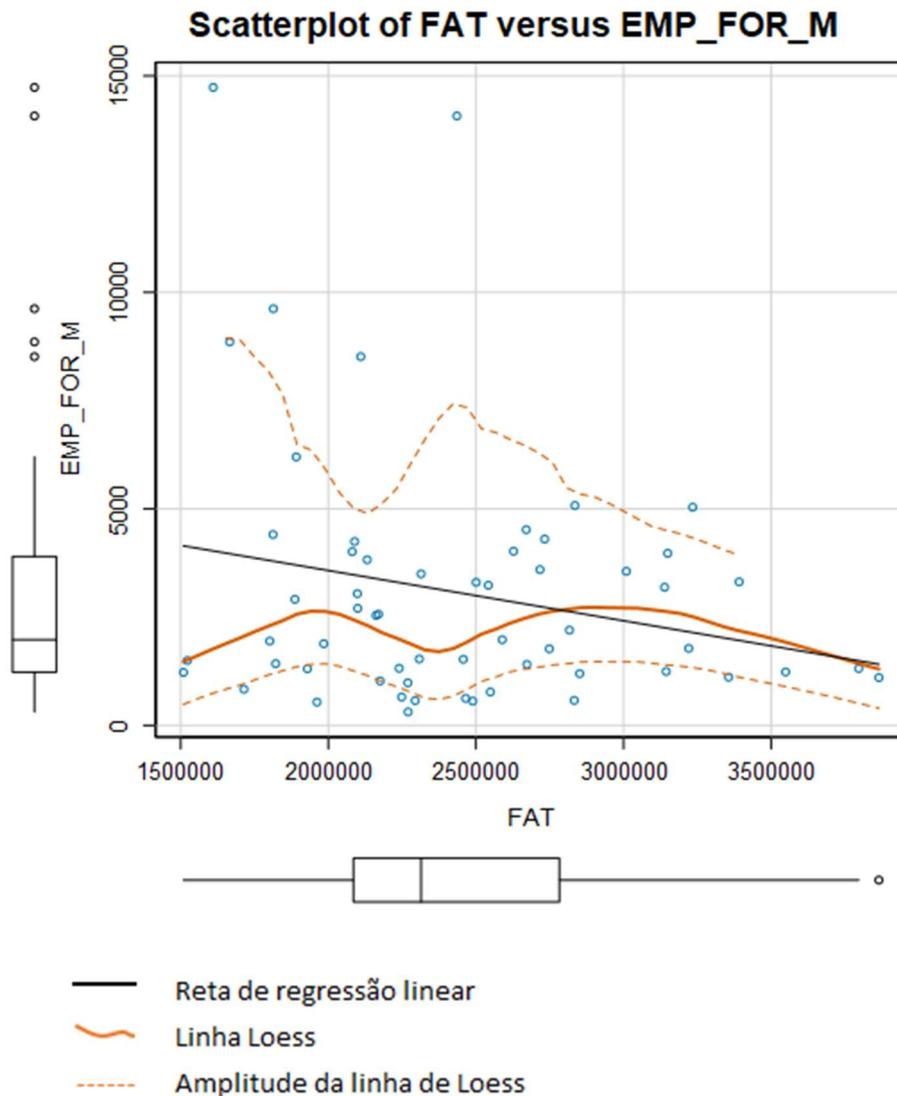
Gráfico 3 – Faturamento versus número de empregos formais



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o número de empregos formais nos municípios, e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde há mais empregos formais resulta em maior faturamento da filial, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

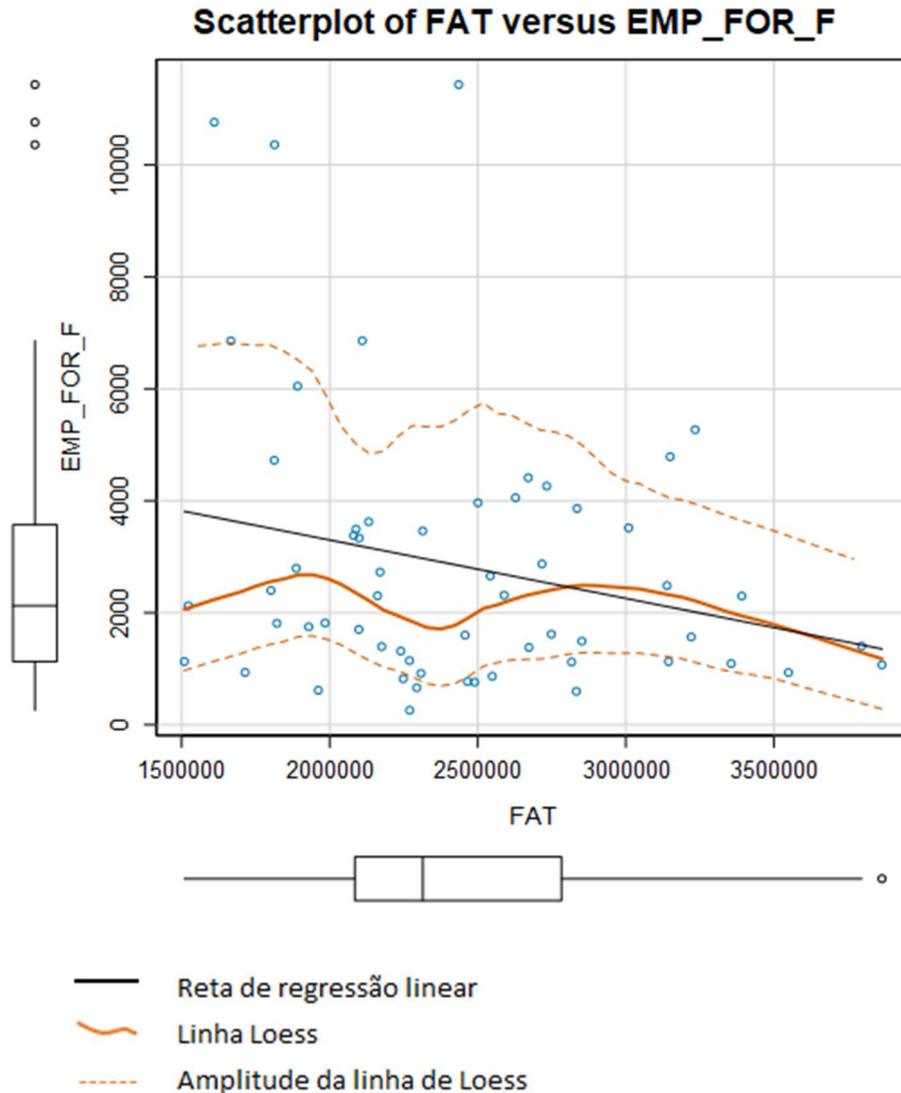
Gráfico 4 – Faturamento versus número de empregos formais masculinos



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o número de empregos formais de pessoas do sexo masculino nos municípios, e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde há mais empregos formais masculinos, proporcionando uma renda familiar maior, resulta em maior faturamento da filial, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

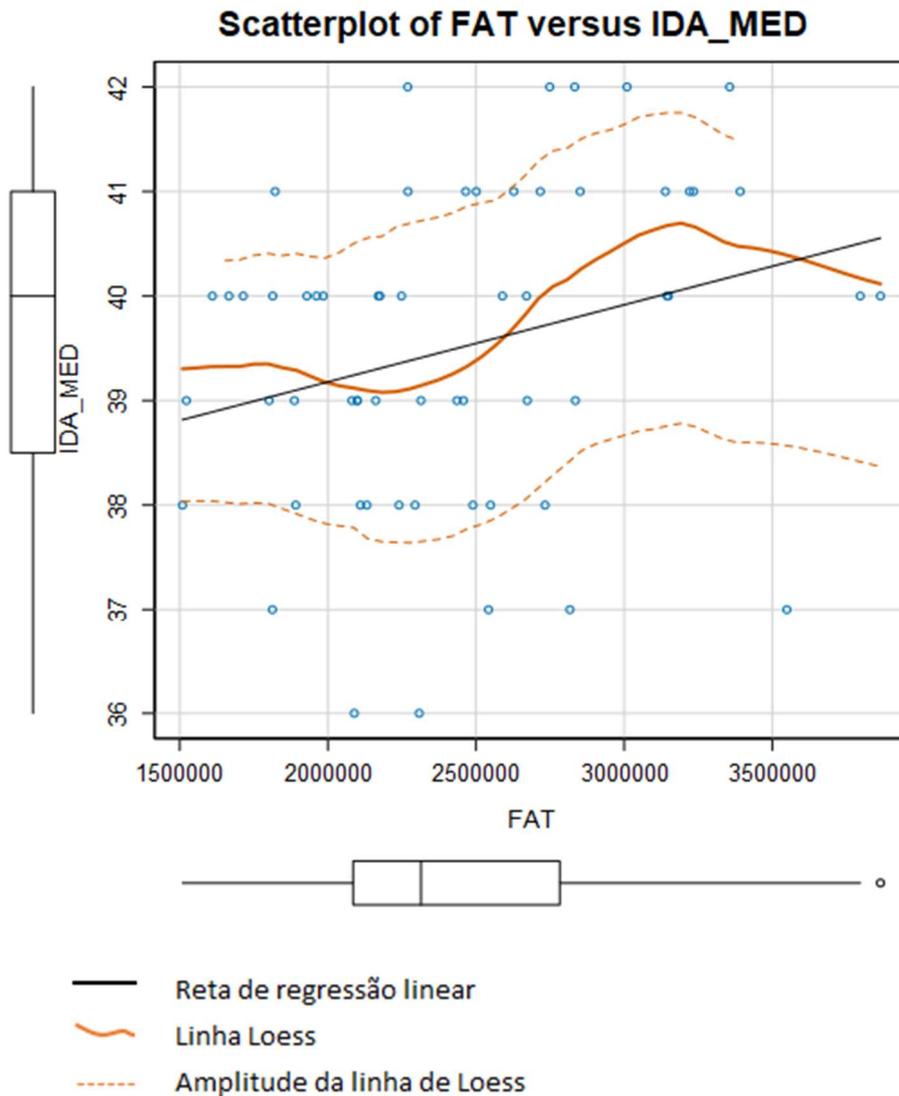
Gráfico 5 – Faturamento versus número de empregos formais femininos



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o número de empregos formais de pessoas do sexo feminino nos municípios, e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde há mais empregos formais femininos, proporcionando uma renda maior para este público, responsável por cerca de 60% da venda da empresa, resulta em maior faturamento da filial, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

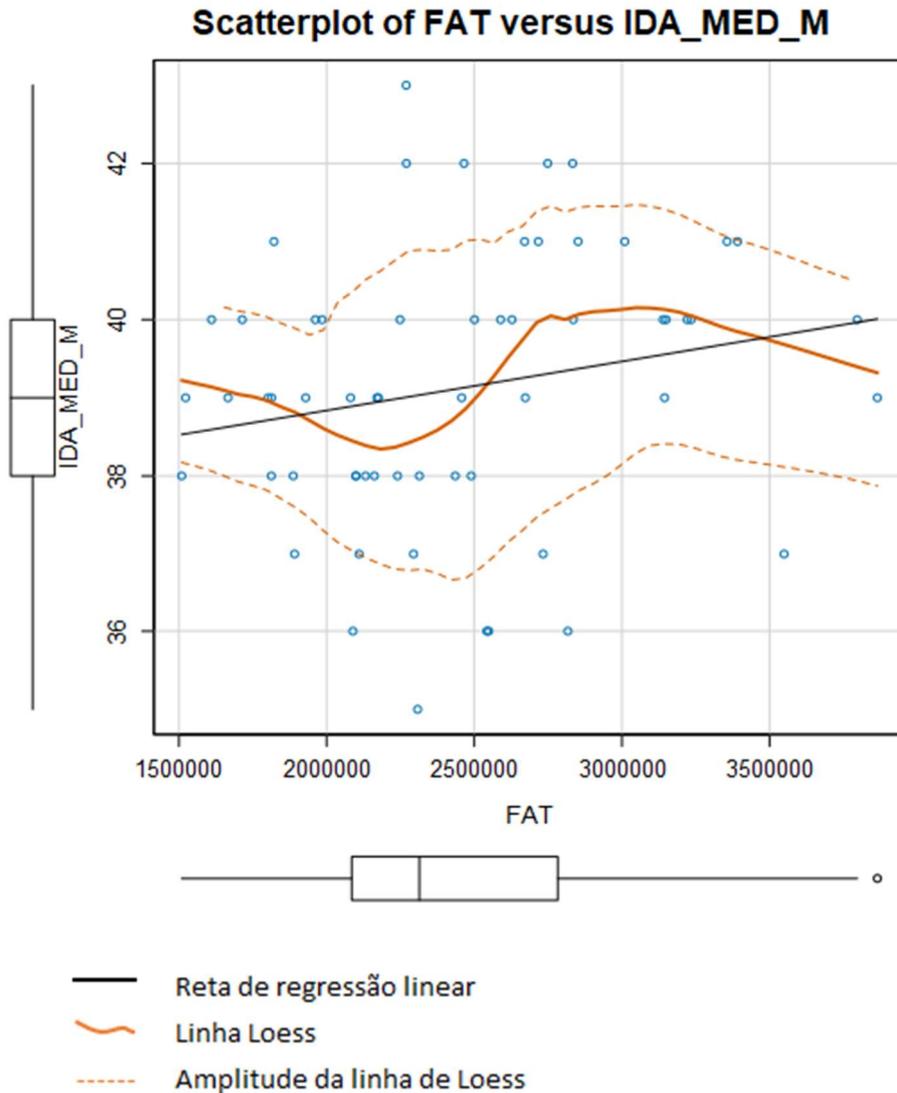
Gráfico 6 – Faturamento X idade média dos trabalhadores com vínculo ativo em 31/12/2017



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre a idade média dos trabalhadores com vínculo ativo em 31/12/2017 nos municípios, e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde há uma idade média menor, já que o consumo de produtos da moda é maior para este público, resulta em maior faturamento da filial, ou seja, uma relação linear negativa, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

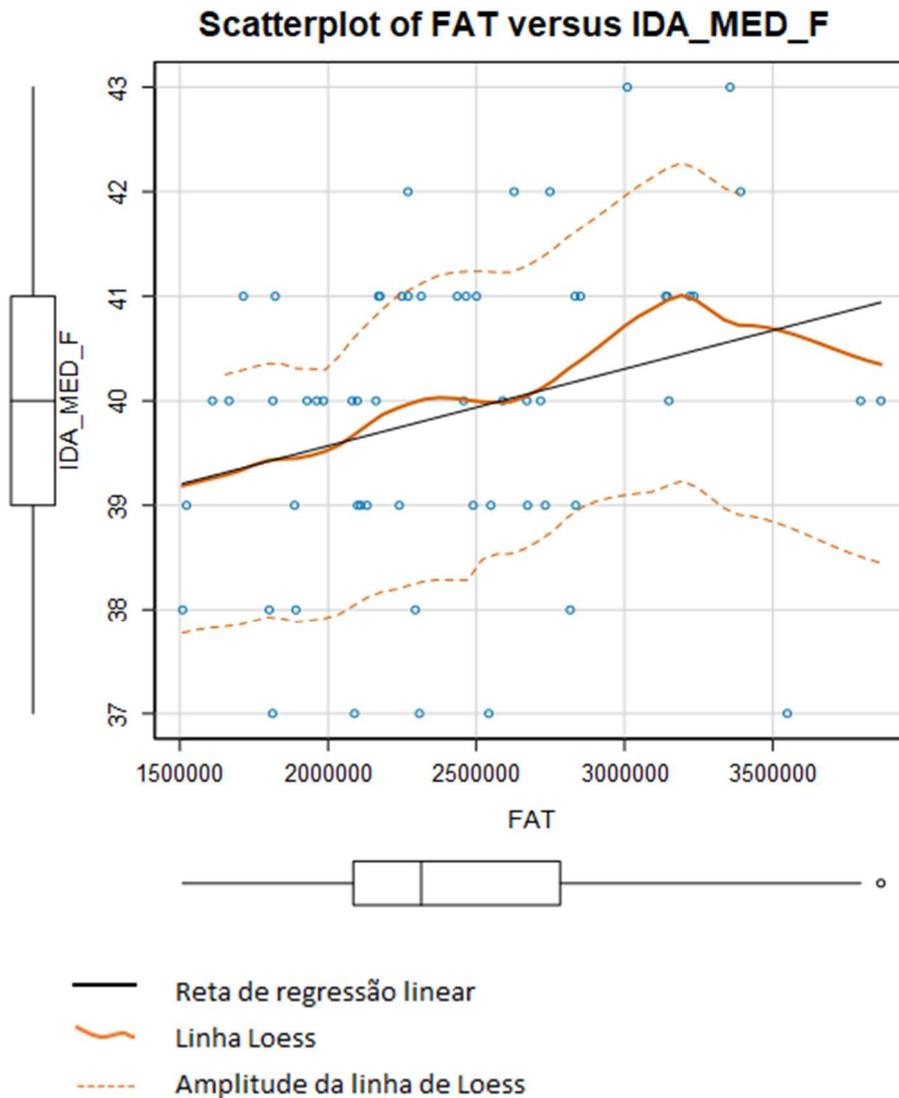
Gráfico 7 – Faturamento versus idade média dos trabalhadores do sexo masculino com vínculo ativo em 31/12/2017



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre a idade média dos trabalhadores do sexo masculino com vínculo ativo em 31/12/2017 nos municípios, e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde há uma idade média menor, já que o consumo de produtos da moda é maior para este público, resulta em maior faturamento da filial, ou seja, uma relação linear negativa, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

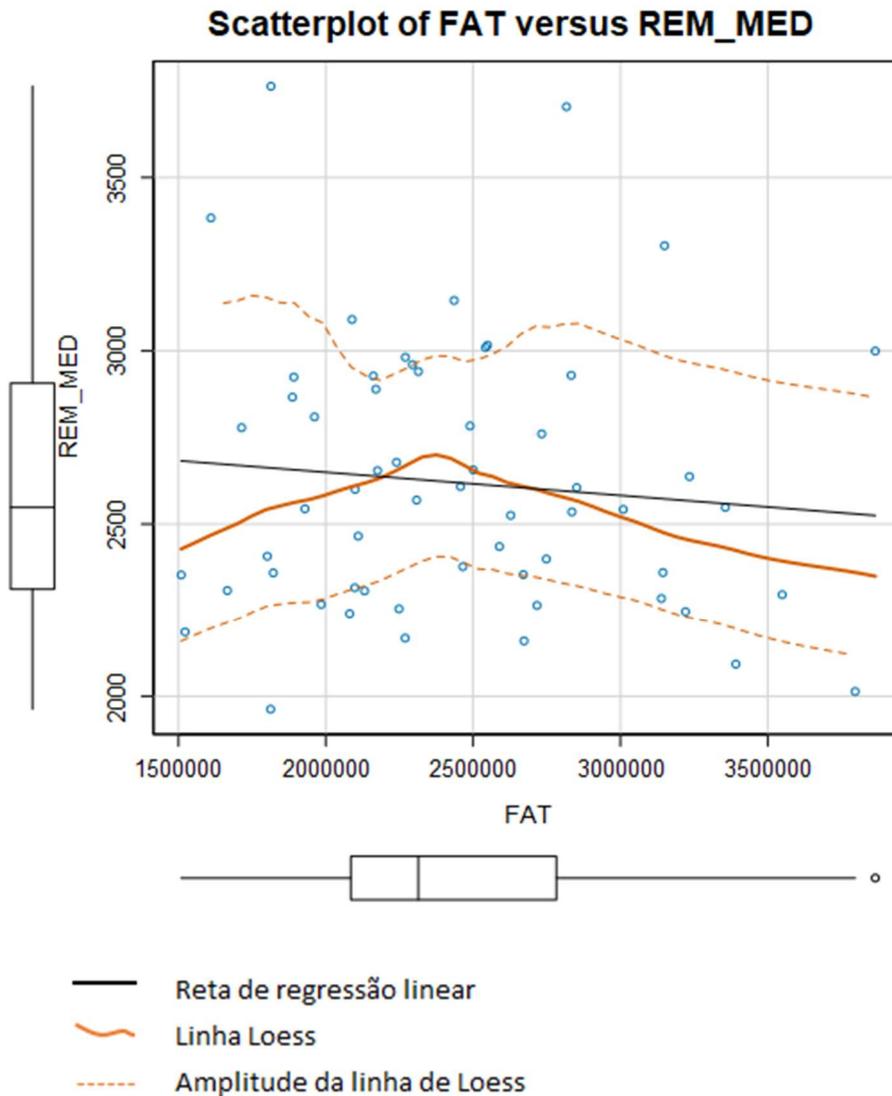
Gráfico 8 – Faturamento versus idade média dos trabalhadores do sexo feminino com vínculo ativo em 31/12/2017



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre a idade média dos trabalhadores do sexo feminino com vínculo ativo em 31/12/2017 nos municípios, e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde há uma idade média menor, já que o consumo de produtos da moda é maior para este público, resulta em maior faturamento da filial, ou seja, uma relação linear negativa, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

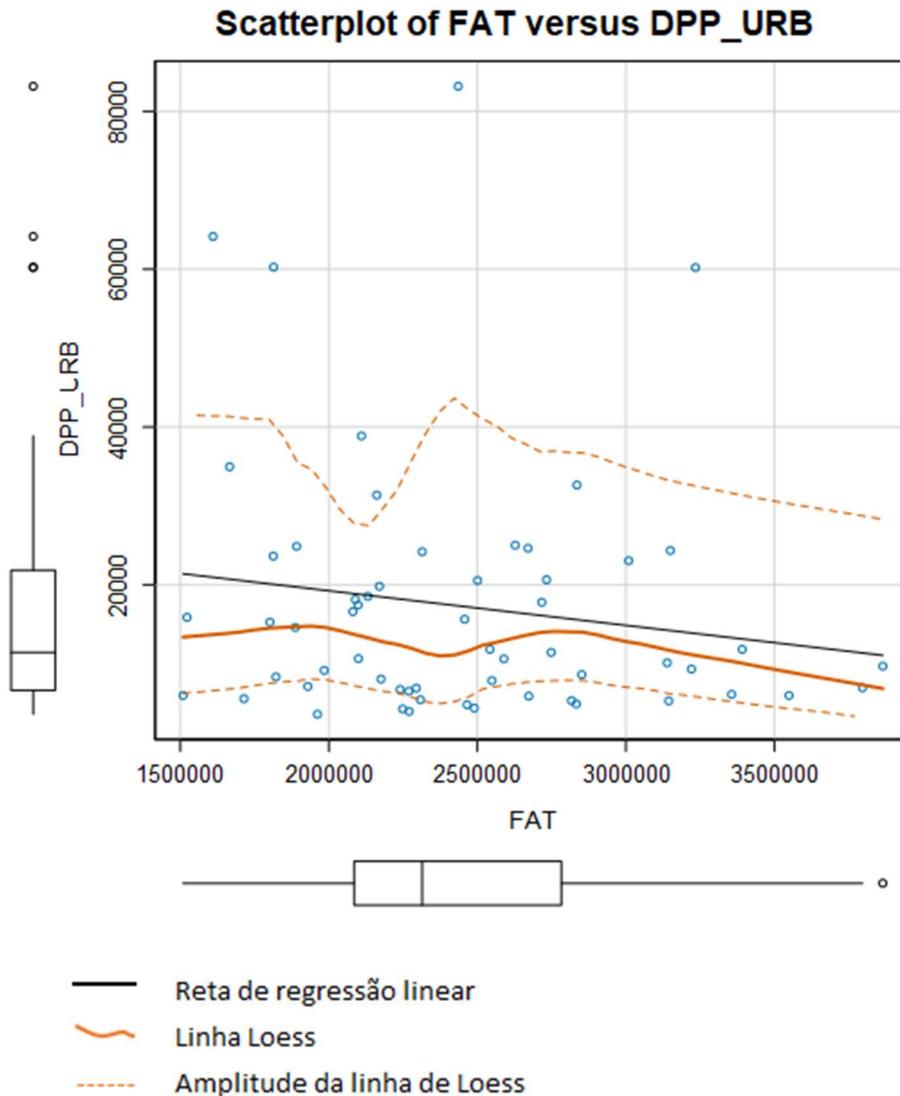
Gráfico 9 – Faturamento versus remuneração média dos trabalhadores com vínculo ativo em 31/12/2017



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre a remuneração média dos trabalhadores com vínculo ativo em 31/12/2017 nos municípios, e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde há uma renda média maior resulta em maior faturamento da filial, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

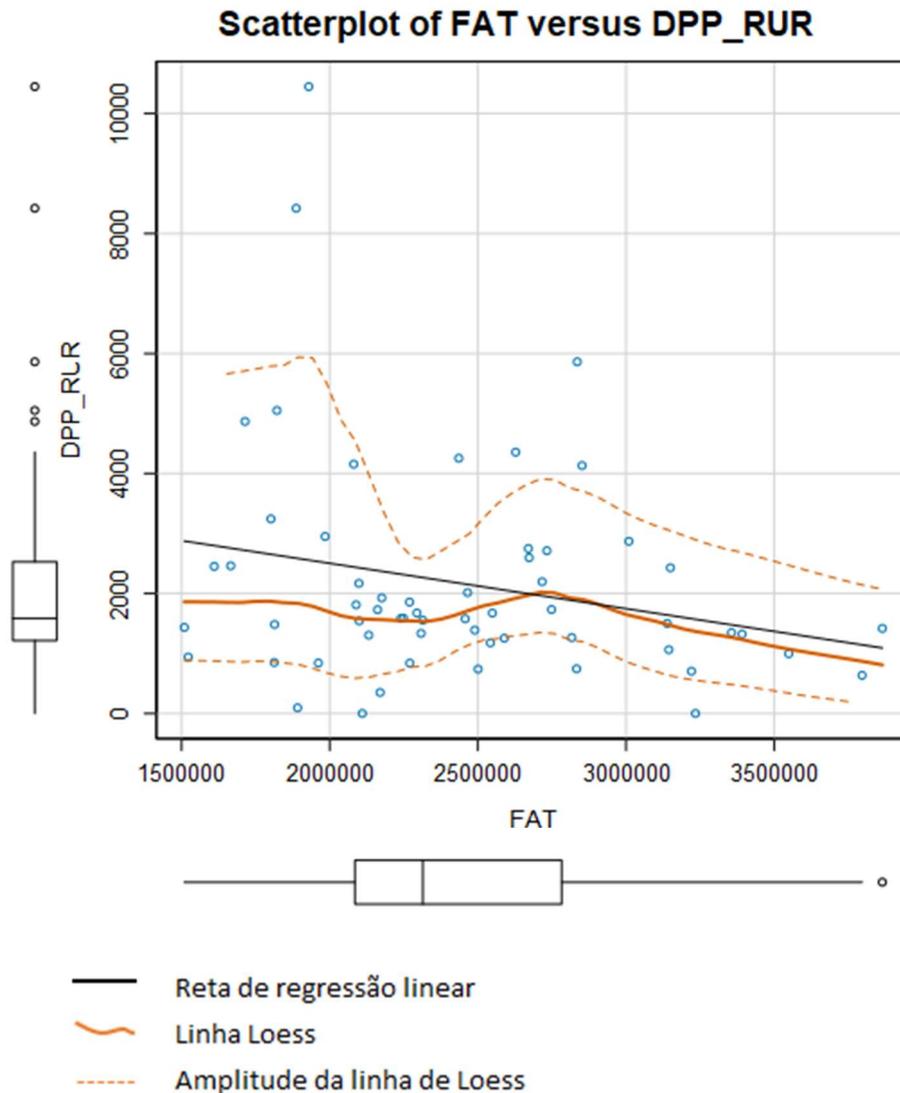
Gráfico 10 – Faturamento versus número de domicílios privados urbanos



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o número de domicílios privados urbanos nos municípios, e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde há um número de domicílios privados urbanos maior resulta em maior faturamento da filial, ou seja, uma relação linear positiva, ou de que o número menor de domicílios privados urbanos resulta em maior faturamento da filial, ou seja, relação linear negativa, não são válidas nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

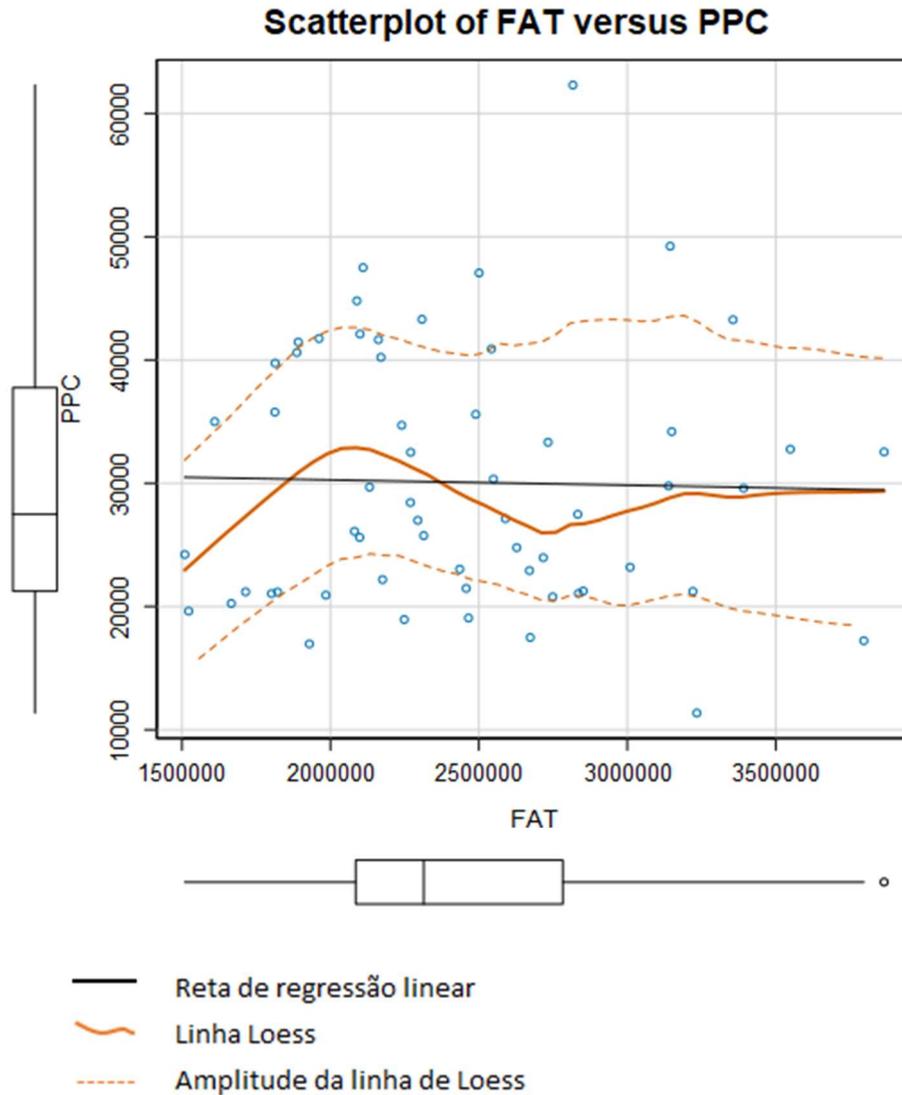
Gráfico 11 – Faturamento versus número de domicílios privados rurais



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o número de domicílios privados rurais nos municípios, e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde há um número de domicílios privados rurais maior resulta em maior faturamento da filial, ou seja, uma relação linear positiva, ou de que o número menor de domicílios privados rurais resulta em maior faturamento da filial, ou seja, relação linear negativa, não são válidas nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

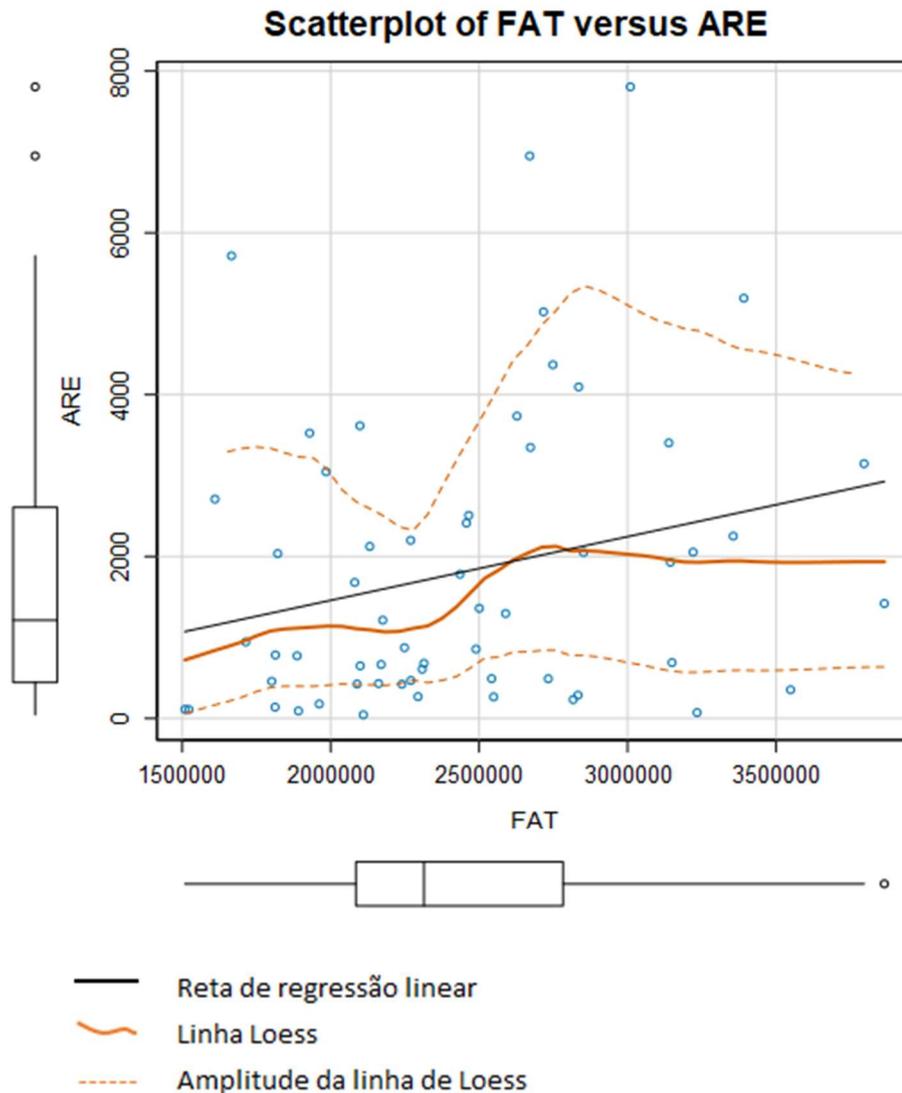
Gráfico 12 – Faturamento versus PIB per capita



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o PIB per capita dos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde o PIB per capita é maior, o faturamento da filial também é maior, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

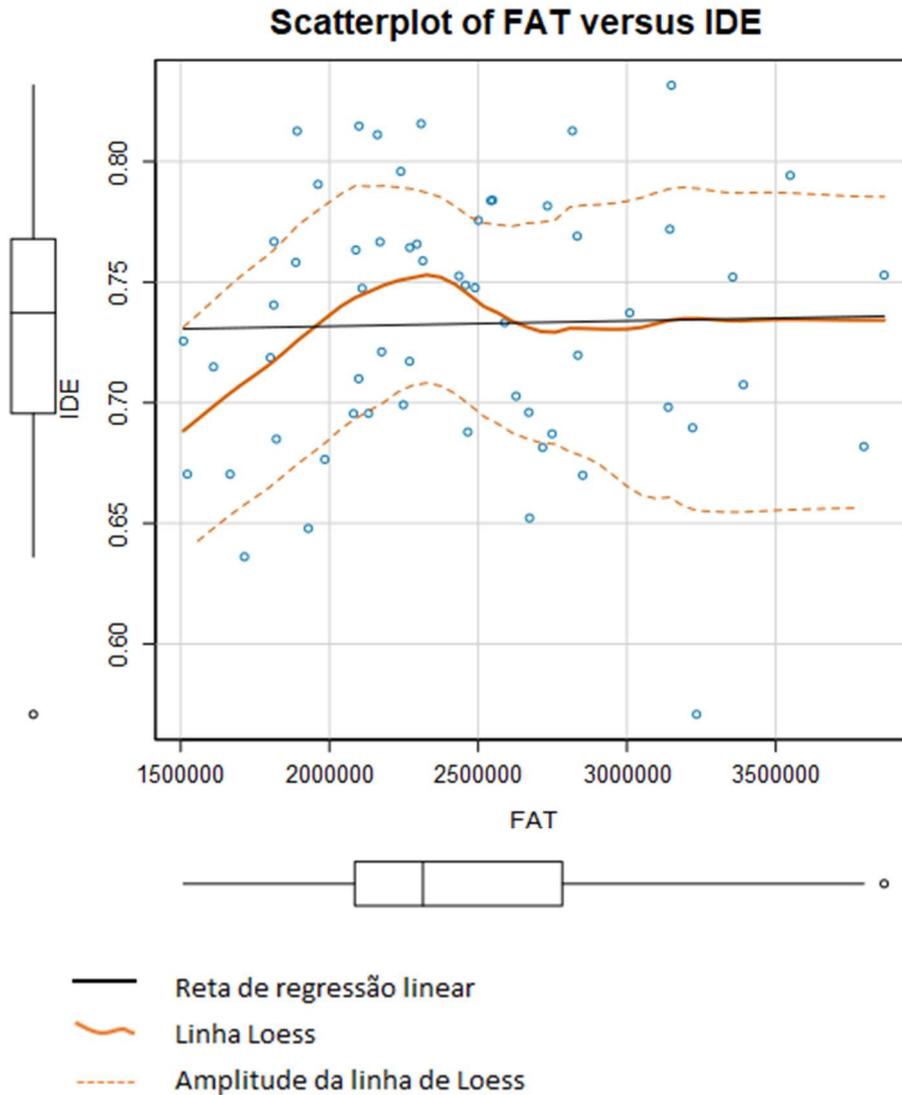
Gráfico 13 – Faturamento versus área



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre a área dos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde a área é maior, o faturamento das filiais também é maior, ou seja, uma relação linear positiva, ou de que onde a área é menor e o faturamento da filial é maior, ou seja uma relação linear negativa, não são válidas nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

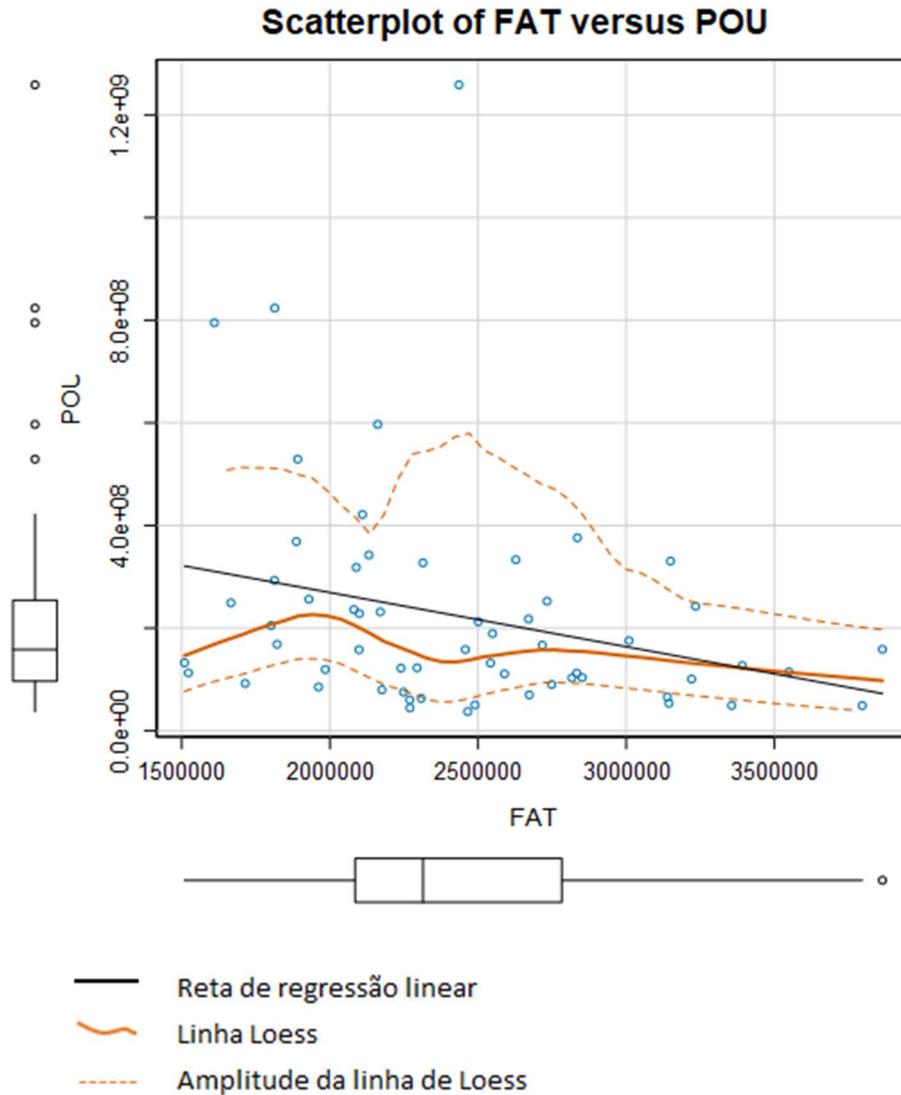
Gráfico 14 – Faturamento versus IDESE



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o IDESE dos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde o IDESE é maior, o faturamento das filiais também é maior, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

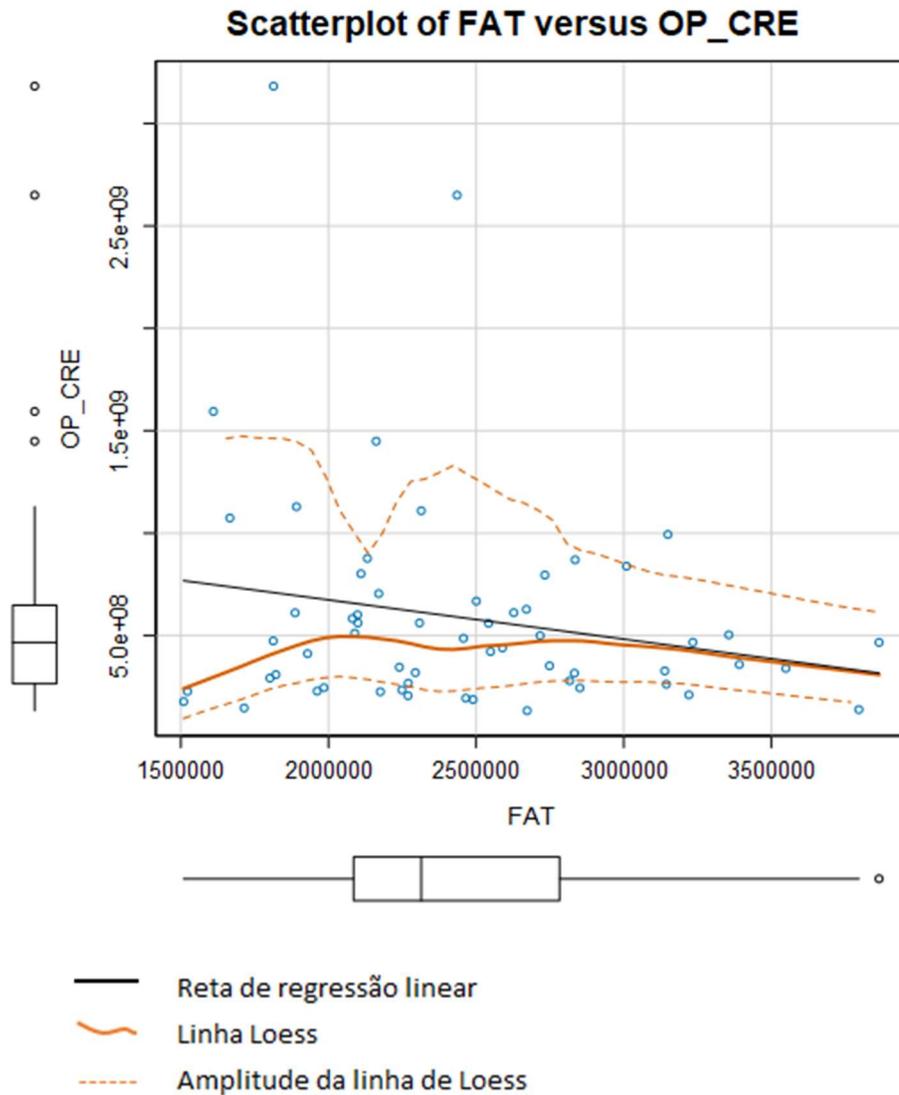
Gráfico 15 – Faturamento versus valor em poupança



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o valor em poupança dos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde o valor em poupança é maior, o faturamento das filiais também é maior, ou seja, uma relação linear positiva, ou de que onde valor em poupança é menor e o faturamento da filial é maior, ou seja uma relação linear negativa, não são válidas nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

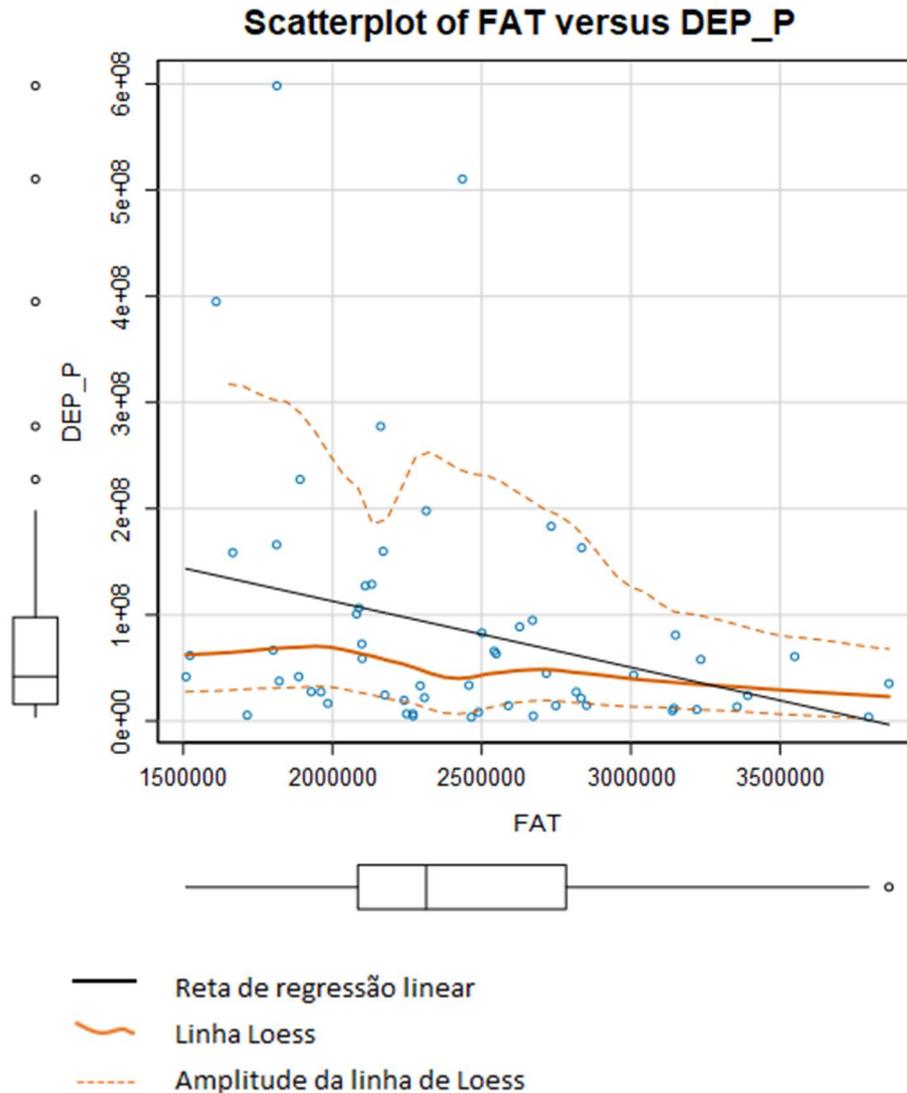
Gráfico 16 – Faturamento versus operações de crédito



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o valor em operações de crédito dos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde o valor em operações de crédito é maior, o faturamento das filiais também é maior, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

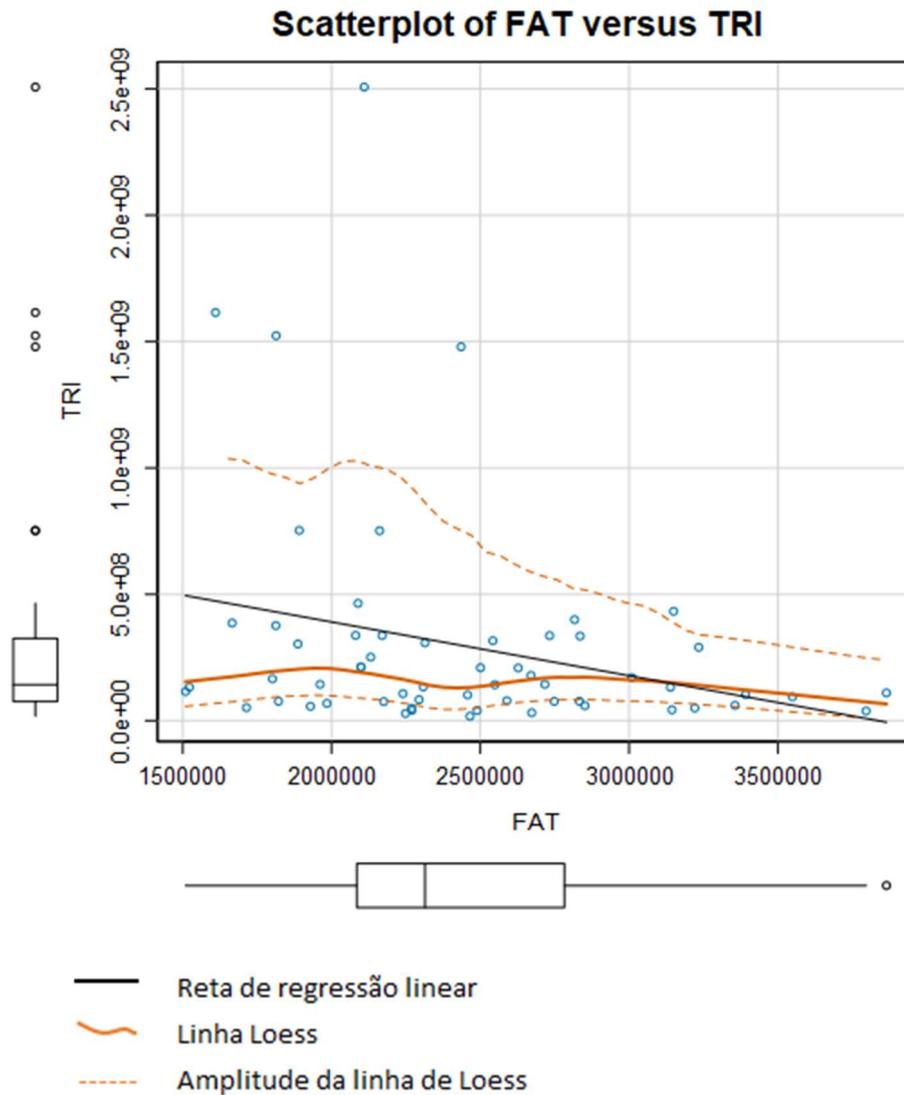
Gráfico 17 – Faturamento versus depósitos a prazo



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o valor dos depósitos a prazo nos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde o valor dos depósitos à prazo é maior, o faturamento das filiais também é maior, ou seja, uma relação linear positiva, ou de que onde valor dos depósitos a prazo é menor e o faturamento da filial é maior, ou seja uma relação linear negativa, não são válidas nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

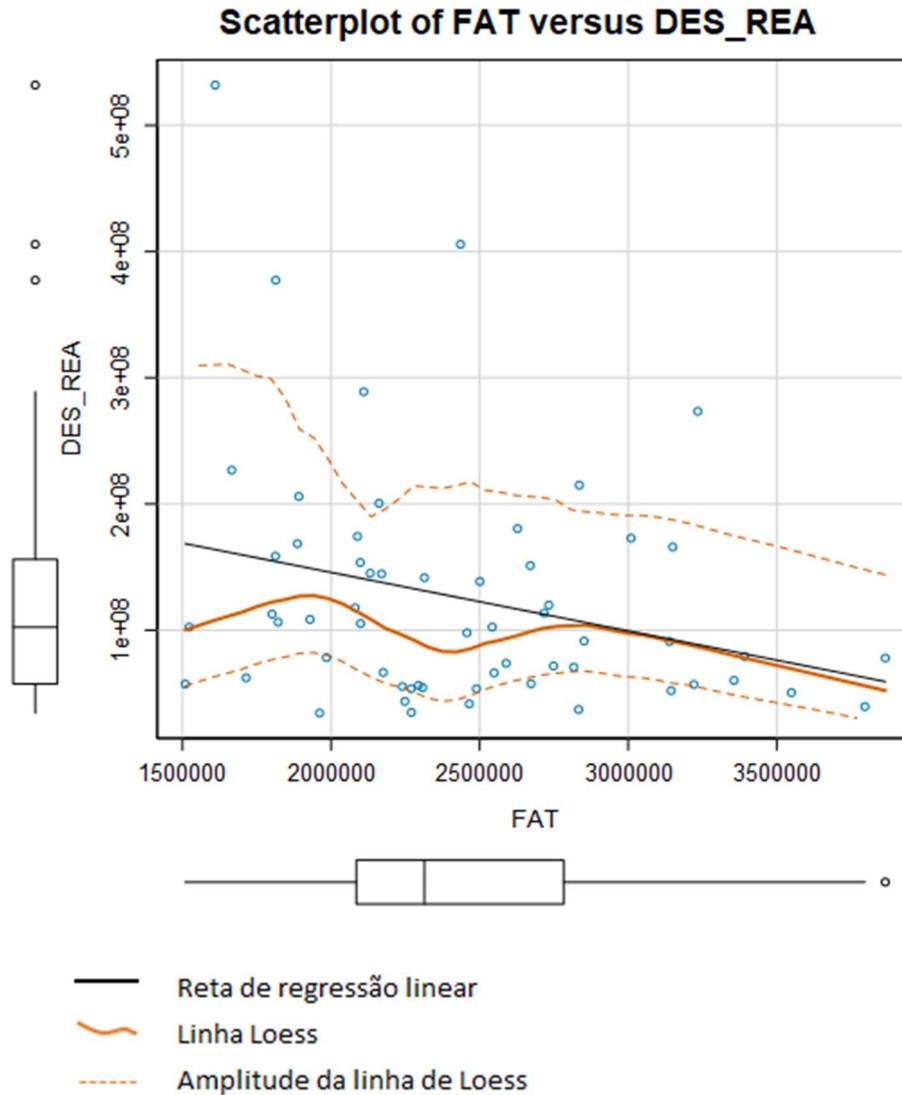
Gráfico 18 – Faturamento versus arrecadação total de tributos



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o valor de arrecadação de tributos dos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde o valor da arrecadação de tributos é maior, o faturamento das filiais também é maior, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

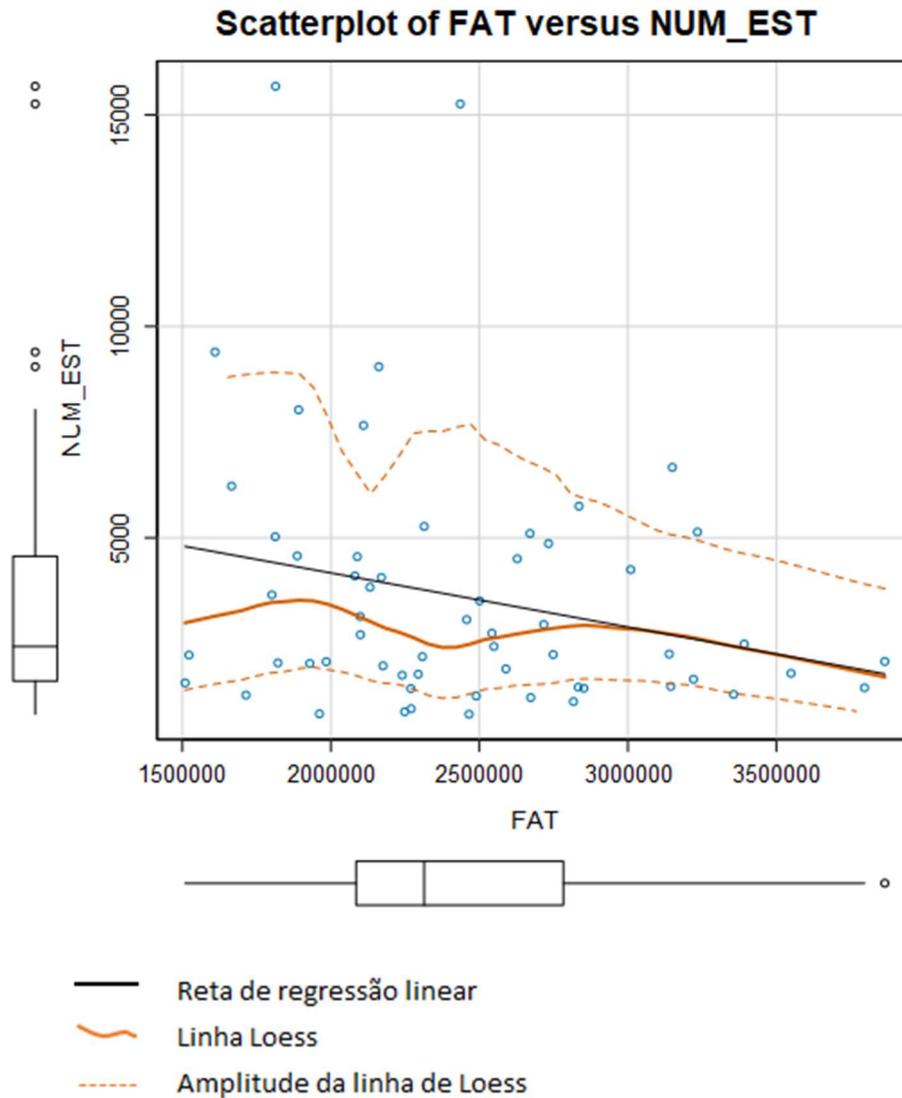
Gráfico 19 – Faturamento versus despesas realizadas



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o valor das despesas realizadas dos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde o valor das despesas realizadas pelo município é maior, o faturamento das filiais também é maior, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

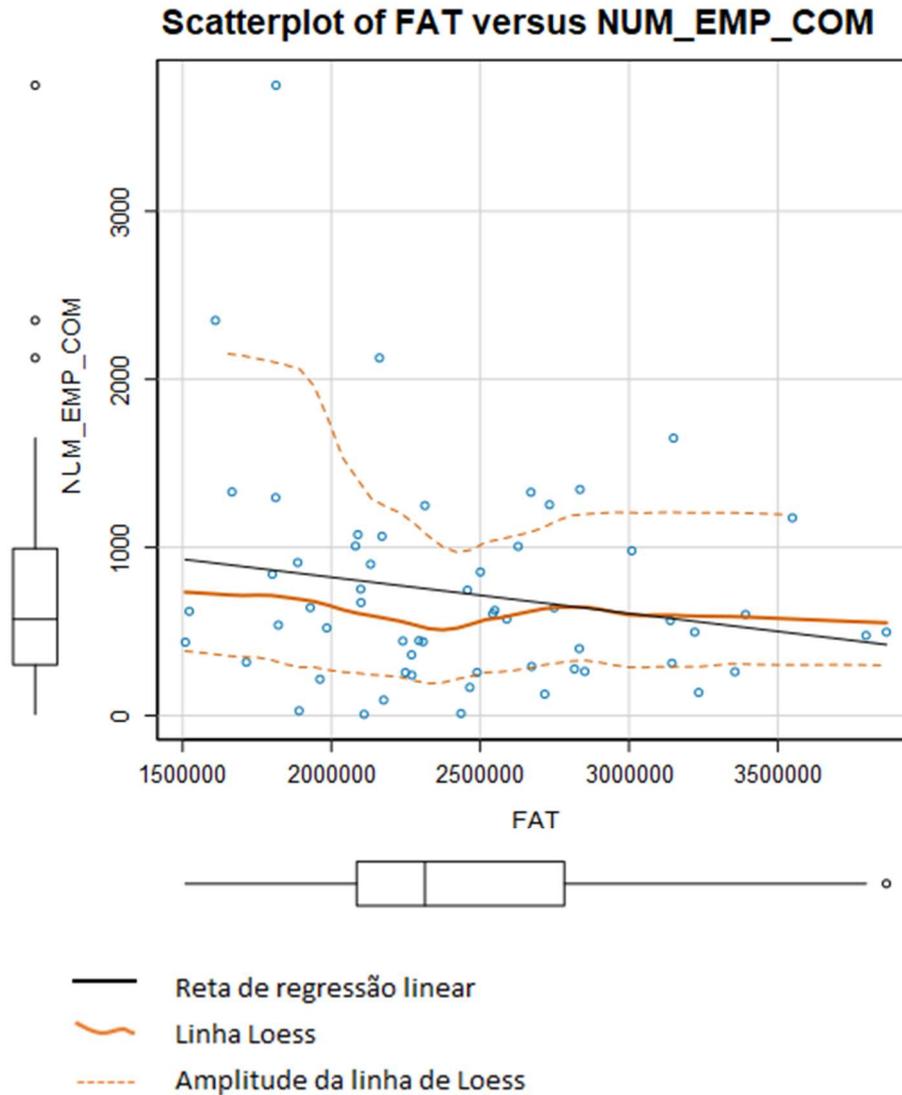
Gráfico 20 – Faturamento versus número de estabelecimentos



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o número de estabelecimentos dos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde o número de estabelecimentos é maior, o que resulta em uma concorrência maior, o faturamento das filiais é menor, ou seja, uma relação linear negativa, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

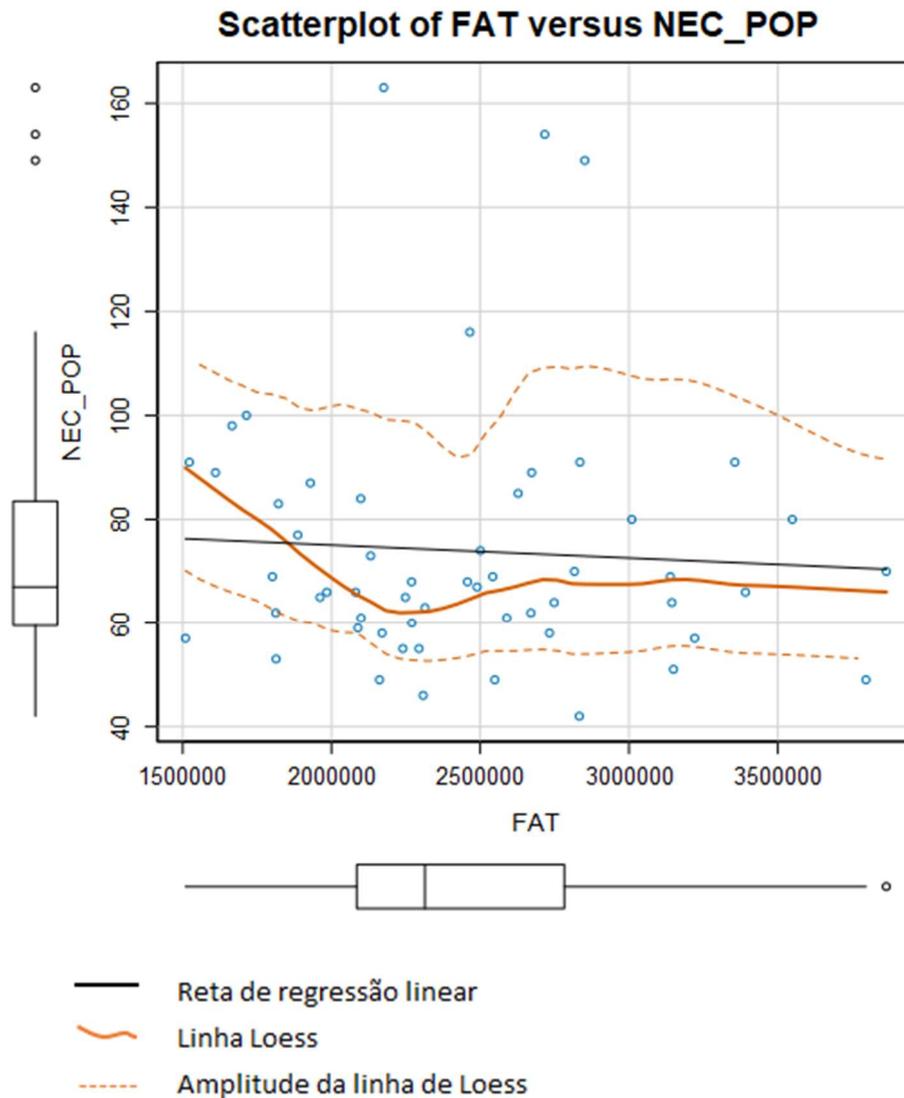
Gráfico 21 – Faturamento versus número de empresas comerciais



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o número de estabelecimentos comerciais dos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde o número de estabelecimentos comerciais é maior, o que resulta em uma concorrência maior no segmento, o faturamento das filiais é menor, ou seja, uma relação linear negativa, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

Gráfico 22 – Faturamento versus razão entre o número de empresas comerciais e a população



Fonte: Elaborado pelo autor com uso do *software Alteryx Designer*.

A plotagem do gráfico que mostra a relação entre o número de pessoas para cada estabelecimento comercial dos municípios e o faturamento das filiais nestas mesmas cidades, não imprime uma relação linear significativa entre as duas variáveis. A hipótese de que onde o número de pessoas para cada estabelecimento comercial é maior, o que resulta em uma concorrência menor, o faturamento das filiais é maior, ou seja, uma relação linear positiva, não é válida nesta observação, já que os pontos são dispersos e não se aproximam de uma reta.

Pode-se confirmar os dados de correlação apresentados no Quadro 3 através da visualização dos gráficos anteriores. Nota-se que não há existência de relação linear das variáveis independentes ou preditoras com a variável alvo ou preditiva, com significância, em nenhuma das ocasiões. Os gráficos ilustram uma dispersão muito grande dos valores, sem que os pontos estejam próximos de uma reta. Ao mesmo tempo, identifica-se pontos que se distanciam muito da média, o que em alguns casos podem ser tratados como *outliers*. Neste caso, como os dados são verdadeiros e representam de forma correta os indicadores do município, podem ser mantidos no modelo.

O segundo passo foi correlacionar as variáveis independentes entre si. Conforme visto anteriormente, a alta correlação, a qual considera-se que seja acima de 0,8 pode indicar que estas variáveis tenham capacidade de explicação semelhantes no modelo de regressão. Por vezes, mesmo que tenham alta correlação, estas variáveis podem ser mantidas no modelo se a capacidade total de explicação do modelo for maior do que ao retirá-las.

A seguir apresenta-se a figura de correlação das variáveis independentes, cujos valores em destaque representam um coeficiente de correlação maior ou igual a 0,8:

Figura 2 – Correlação entre variáveis independentes

	POP	POP_URB	EMP_FOR	EMP_FOR_M	EMP_FOR_F	IDA_MED	IDA_MED_M	IDA_MED_F	REM_MED	FAT	DPP_URB	DPP_RUR	PPC	ARE	IDE	POU	OP_CRE	DEP_P	TRI	DES_REA	NUM_EST	NUM_EMP_COM	NEC_POP	
POP																								
POP_URB	36%																							
EMP_FOR	62%	49%																						
EMP_FOR_M	59%	48%	<b>99%</b>																					
EMP_FOR_F	64%	49%	<b>99%</b>	<b>97%</b>																				
IDA_MED	2%	-9%	-7%	-9%	-6%																			
IDA_MED_M	-2%	-14%	-13%	-14%	-11%	<b>92%</b>																		
IDA_MED_F	-4%	-6%	-2%	-3%	-1%	<b>91%</b>	78%																	
REM_MED	27%	11%	33%	32%	34%	-20%	-32%	-13%																
FAT	-14%	17%	-23%	-22%	-24%	28%	20%	28%	-10%															
DPP_URB	68%	51%	<b>93%</b>	<b>91%</b>	<b>93%</b>	0%	-7%	7%	35%	-15%														
DPP_RUR	7%	-70%	5%	5%	5%	9%	9%	8%	-1%	-23%	3%													
PPC	-6%	23%	9%	10%	8%	-40%	-47%	-32%	49%	-2%	-3%	-30%												
ARE	10%	-3%	13%	15%	10%	49%	49%	47%	-40%	25%	7%	29%	-42%											
IDE	-15%	27%	0%	0%	1%	-44%	-48%	-36%	53%	2%	-8%	-29%	77%	-45%										
POU	49%	35%	<b>87%</b>	<b>86%</b>	<b>88%</b>	-15%	-22%	-6%	48%	-28%	<b>88%</b>	19%	13%	-6%	17%									
OP_CRE	55%	42%	<b>83%</b>	<b>80%</b>	<b>85%</b>	-7%	-17%	3%	53%	-20%	<b>83%</b>	6%	18%	1%	25%	<b>90%</b>								
DEP_P	59%	44%	<b>85%</b>	<b>83%</b>	<b>86%</b>	-16%	-22%	-8%	50%	-30%	<b>85%</b>	2%	17%	-9%	21%	<b>92%</b>	<b>95%</b>							
TRI	38%	42%	<b>81%</b>	<b>82%</b>	<b>80%</b>	-21%	-27%	-13%	42%	-27%	75%	-8%	33%	-16%	19%	78%	72%	75%						
DES_REA	72%	43%	<b>95%</b>	<b>94%</b>	<b>94%</b>	-5%	-12%	0%	41%	-27%	<b>95%</b>	12%	8%	9%	-3%	<b>88%</b>	<b>81%</b>	<b>84%</b>	<b>82%</b>					
NUM_EST	60%	48%	<b>89%</b>	<b>87%</b>	<b>91%</b>	-12%	-19%	-4%	45%	-24%	<b>91%</b>	7%	13%	0%	17%	<b>94%</b>	<b>95%</b>	<b>95%</b>	<b>80%</b>	<b>89%</b>				
NUM_EMP_COM	74%	30%	48%	45%	52%	-7%	-7%	-9%	35%	-19%	46%	10%	14%	8%	19%	49%	65%	66%	36%	55%	61%			
NEC_POP	12%	29%	35%	34%	35%	-12%	-19%	-12%	4%	-8%	33%	-23%	14%	-20%	0%	27%	15%	17%	56%	34%	31%	-27%		

Fonte: Elaborado pelo autor com uso do *software* Microsoft Excel.

Pode-se identificar que os empregos formais separados por gênero têm alta correlação com a quantidade total de empregos formais, imprimindo que quaisquer um dos dados tem a mesma capacidade de explicação quanto aplicados ao modelo, sendo então necessário usar apenas uma destas três variáveis. Observa-se também que valores em poupança, valor das operações de crédito, valor de arrecadação de tributos, despesas realizadas pelo governo e número de estabelecimentos também apresentam alta correlação com o número de empregos formais.

Estes indicadores econômico-financeiros demonstram que a medida que os dados aumentam, o número de empregos formais também aumenta, ou seja, estas variáveis têm correlação linear positiva. Estas mesmas variáveis, com exceção do valor de tributos também apresentam relação linear positiva, ou seja, tem um alto índice de correlação, com a variável número de domicílios privados. Assim, o modelo de regressão pode ser testado com e sem o uso das variáveis que tem alta correlação entre si.

#### **4.3.2 REGRESSÃO LINEAR**

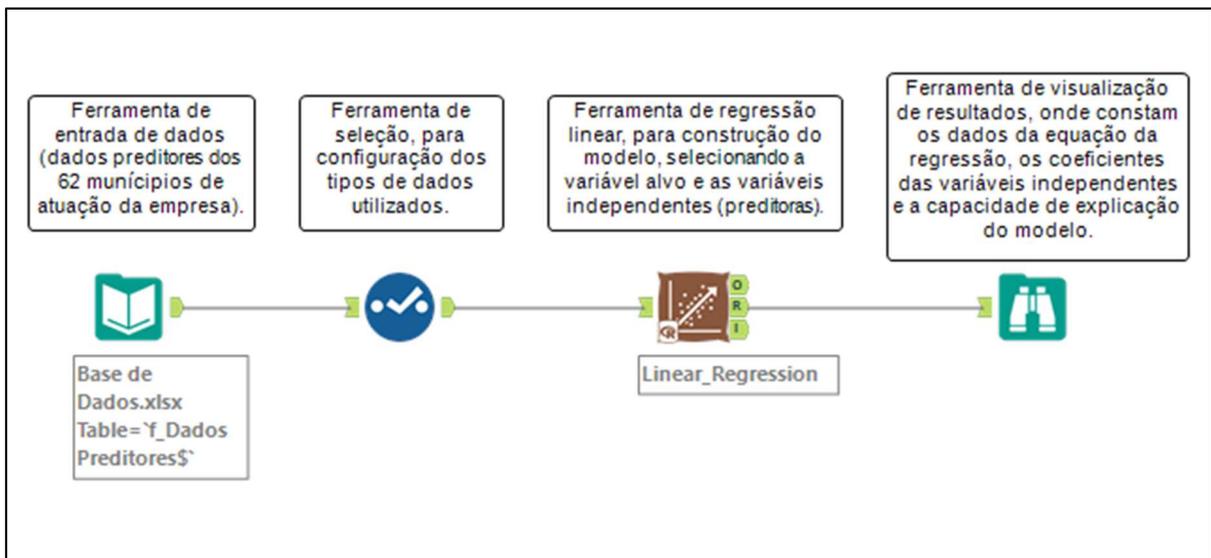
Segundo Hoffmann (2006), além de determinar se existe correlação entre duas variáveis, é desejável saber a função que mostra o quanto uma variável se altera em função da alteração da outra. Deste modo, a regressão pode ser utilizada como modelo preditivo para a tomada de decisão.

A partir dos dados selecionados, pode-se então criar um modelo de regressão na tentativa de prever a receita futura da empresa em municípios em que a mesma ainda não atua. Ao criar-se uma regressão linear, obtém-se a equação da regressão, que segundo Gujarati (2000), tem a capacidade de estimar ou prever o valor médio da variável dependente com base nos valores conhecidos, que neste estudo, são os indicadores dos 62 municípios em que a empresa atua, os quais representam as variáveis explicativas.

A regressão linear múltipla foi desenvolvida com o uso de todas as variáveis presentes no Quadro 3, com o objetivo de analisar a significância estatística de cada variável. Sabe-se que, devido à baixa correlação das variáveis explicativas com a variável explicada, as variáveis explicativas podem não ser significantes estatisticamente.

A seguir, a Figura 3 demonstra o processo de construção do modelo de regressão linear com o uso do *software* Alteryx Designer:

Figura 3 – Fluxograma do modelo de regressão linear com o uso do *software* Alteryx Designer



Fonte: Elaboração do autor com uso do *software* Alteryx Designer.

O modelo de regressão linear elaborado com todas as variáveis presentes no Quadro 3, produziu os seguintes dados, representados na Figura 4:

Figura 4 – Relatório de regressão linear com uso do software Alteryx Designer

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	-2.609e+06	3.541e+06	-0.73664	0.46599
POP	1.019e+01	1.414e+01	0.72048	0.47576
POP_URB	1.225e+06	1.792e+06	0.68369	0.49843
EMP_FOR	-3.224e+01	2.127e+02	-0.15158	0.88034
EMP_FOR_M	9.638e+01	3.944e+02	0.24439	0.80828
IDA_MED	1.711e+05	1.931e+05	0.88607	0.3813
IDA_MED_M	-1.011e+05	1.160e+05	-0.87101	0.38937
IDA_MED_F	-9.892e+03	1.278e+05	-0.07738	0.93874
REM_MED	3.353e+02	2.825e+02	1.18717	0.24273
DPP_URB	4.964e+01	4.982e+01	0.99648	0.32549
DPP_RUR	-1.287e+01	9.919e+01	-0.12976	0.89746
PPC	7.229e+00	1.291e+01	0.55993	0.5789
ARE	1.669e+02	8.342e+01	2.00087	0.05278 .
IDE	1.044e+06	2.999e+06	0.34800	0.72982
POU	3.628e-03	1.774e-03	2.04536	0.04798 **
OP_CRE	9.314e-05	6.928e-04	0.13444	0.89378
DEP_P	-4.266e-03	3.043e-03	-1.40191	0.16927
TRI	9.289e-04	1.153e-03	0.80584	0.42549
DES_REA	-1.721e-02	4.991e-03	-3.44860	0.00142 ***
NUM_EST	-1.750e+02	2.482e+02	-0.70515	0.48513
NUM_EMP_COM	-2.385e+01	1.148e+03	-0.02077	0.98354
NEC_POP	-1.077e+02	8.105e+02	-0.13287	0.89502
EMP_FOR_F	NA	NA	NA	NA

Fonte: Elaborado pelo autor com uso do *software* Alteryx Designer.

O primeiro passo foi analisar a coluna " $Pr(>|t|)$ ", a qual expressa a significância estatística dos dados ou o p-valor, que segundo a Udacity (2018), representa a probabilidade de que o coeficiente seja zero, ou seja, é o mesmo que, segundo Gujarati (2000), verificar a validade ou a falsidade de uma hipótese nula. Quanto menor o p-valor, maior a probabilidade de existir uma relação entre a variável preditora e a variável alvo. A um nível de significância de 95%, deve-se confiar na estimativa do coeficiente quando este for inferior a 0,05 ou 5%, que é quando a variável apresenta significância estatística e aceita-se a hipótese nula como verdadeira.

Na Figura 4, observa-se que apenas as variáveis valor em poupança e despesas realizadas pelo governo tem significância estatística, ou seja, aceita-se que para cada 1 unidade a mais de valor em poupança, o faturamento da empresa aumenta 0,003628 (expresso na coluna "*Estimate*" da Figura 4 sob a notação científica 3,628e-03). Já no caso das despesas realizadas pelo governo, para cada 1 unidade a mais de gastos, o faturamento da empresa cai em 0,01721 (expresso na coluna "*Estimate*" da Figura 4 sob a notação científica -1,721e-02). Todas as demais variáveis não são significativas estatisticamente para explicar as variações no faturamento das filiais.

Com o uso do *software* IBM SPSS, foi desenvolvido o mesmo modelo, utilizando o método inserir, o qual força a utilização de todas as variáveis para que pudesse ser observada a significância estatística de todas as variáveis utilizadas. O relatório a seguir demonstra os resultados:

Figura 5 – Relatório de regressão linear com uso do *software* IBM SPSS

		Coeficientes <sup>a</sup>							
Modelo		Coeficientes não padronizados		Coeficientes padronizados	t	Sig.	Estatísticas de colinearidade		
		B	Erro Erro	Beta			Tolerância	VIF	
1	(Constante)	-2608568,264	3541161,012		-,737	,466			
	POP	10,187	14,140	,815	,720	,476	,010	105,118	
	POP_URB	1224965,189	1791683,813	,261	,684	,498	,083	11,984	
	EMP_FOR_M	64,138	194,868	,336	,329	,744	,012	85,725	
	EMP_FOR_F	-32,239	212,691	-,140	-,152	,880	,014	70,100	
	IDA_MED	171090,606	193088,773	,459	,886	,381	,045	22,039	
	IDA_MED_M	-101079,178	116048,309	-,312	-,871	,389	,095	10,565	
	IDA_MED_F	-9891,535	127829,014	-,026	-,077	,939	,110	9,124	
	REM_MED	335,328	282,461	,231	1,187	,243	,322	3,102	
	DPP_URB	49,643	49,818	1,425	,996	,325	,006	168,207	
	DPP_RUR	-12,870	99,186	-,043	-,130	,897	,112	8,900	
	PPC	7,229	12,911	,131	,560	,579	,221	4,527	
	ARE	166,907	83,417	,534	2,001	,053	,171	5,849	
	IDE	1043742,318	2999298,009	,098	,348	,730	,155	6,470	
	POU	,004	,002	1,386	2,045	,048	,026	37,786	
	OP_CRE	9,314E-5	,001	,090	,134	,894	,027	36,835	
	DEP_P	-,004	,003	-,883	-1,402	,169	,031	32,619	
	TRI	,001	,001	,738	,806	,425	,015	68,895	
	DES_REA	-,017	,005	-2,923	-3,449	,001	,017	59,082	
	NUM_EST	-174,983	248,150	-,948	-,705	,485	,007	148,804	
	NUM_EMP_COM	-23,847	1148,282	-,027	-,021	,984	,007	136,820	
	NEC_POP	-107,693	810,535	-,103	-,133	,895	,020	49,628	

Fonte: Elaborado pelo autor com uso do *software* IBM SPSS.

Conforme observado na Figura 5, através da coluna “Sig.”, foram encontrados os mesmos valores de significância estatística para todas as variáveis em relação a figura 4, cujo modelo foi rodado no *software* Alteryx Designer.

A próxima etapa consiste em avaliar R-quadrado e o R-quadrado ajustado. Segundo Gujarati (2000), o R-quadrado pode oscilar entre 0 e 1, imprimindo a capacidade de explicação do modelo, ou seja, a quantidade de variação na variável alvo explicada pela variação das variáveis preditoras. O R-quadrado ajustado imprime a mesma teoria do R-quadrado. Porém, sua capacidade explicativa se altera em função de inserção de mais variáveis no modelo, considerando a relação entre as variáveis adicionais e a variável alvo. Para R-quadrado não importa se não há relação entre as variáveis. A inclusão de mais variáveis no modelo sempre irá aumentar o seu valor.

A seguir, a Figura 6 demonstra os valores para R-quadrado e R-quadrado ajustado:

Figura 6 – Resultado do Modelo com uso do *software* Alteryx Designer

Residual standard error: 473886 on 37 degrees of freedom Multiple R-squared: 0.5502, Adjusted R-Squared: 0.2948 F-statistic: 2.155 on 21 and 37 degrees of freedom (DF), p-value 0.02008
--

Fonte: Elaborado pelo autor com uso do *software* Alteryx Designer.

A Figura 7 demonstra os mesmos resultados do modelo, agora desenvolvido no *software* IBM SPSS:

Figura 7 – Resultado do Modelo com uso do *software* IBM SPSS

Resumo do modelo <sup>b</sup>				
Modelo	R	R quadrado	R quadrado ajustado	Erro padrão da estimativa
1	,742 <sup>a</sup>	,550	,295	473885,6142
a. Preditores: (Constante), NEC_POP, IDE, POP, DPP_RUR, IDA_MED_F, ARE, POU, REM_MED, PPC, IDA_MED_M, TRI, POP_URB, NUM_EMP_COM, EMP_FOR_F, OP_CRE, IDA_MED, DEP_P, DPP_URB, DES_REA, EMP_FOR_M, NUM_EST b. Variável Dependente: FAT				

Fonte: Elaborado pelo autor com uso do *software* IBM SPSS.

As Figuras 6 e 7 demonstram que o modelo tem 55% de capacidade de explicação segundo R-quadrado, ou seja, 55% das variações no faturamento das filiais é explicado pelas variações das variáveis preditoras. Nota-se que o R-quadrado ajustado é significativamente mais baixo que R-quadrado, onde a capacidade de explicação do modelo é de 29%. Esta diferença se dá pelo fato de que R-quadrado ajustado leva em consideração a relação que as variáveis preditoras tem com a variável alvo.

Como visto anteriormente, somente duas variáveis preditoras tem significância estatística no modelo e todas elas têm baixa correlação com a variável alvo. Portanto, é aconselhável utilizar R-quadrado ajustado, pois, diante dos resultados observados, afirmar que a capacidade de explicação do modelo é de 55% (valor de R-quadrado) é tomar posição otimista frente a um modelo incapaz de prever o faturamento das filiais em novas cidades no estado do Rio Grande do Sul.

A incapacidade de explicação do modelo foi testada em mais duas situações. A primeira delas foi desenvolvida desconsiderando as variáveis com alta correlação, as quais foram observadas na Figura 2. Os resultados podem ser conferidos a seguir:

Figura 8 – Resultado do modelo de regressão considerando a multicolinearidade

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.540e+06	2.781e+06	-1.27281	0.20922
POP	3.549e+00	3.192e+00	1.11180	0.27176
POP_URB	1.896e+06	7.277e+05	2.60626	0.01216 *
EMP_FOR	-4.682e+01	2.037e+01	-2.29820	0.02595 *
IDA_MED	6.163e+04	5.434e+04	1.13416	0.26236
REM_MED	2.564e+02	2.603e+02	0.98494	0.32959
PPC	7.287e-01	1.047e+01	0.06964	0.94477
ARE	1.184e+02	5.040e+01	2.34852	0.02301 *
IDE	1.964e+06	2.606e+06	0.75341	0.45488
NUM_EMP_COM	-4.011e+02	2.322e+02	-1.72787	0.09044 .
NEC_POP	-1.132e+02	1.808e+02	-0.62613	0.5342

Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 499199 on 48 degrees of freedom  
Multiple R-squared: 0.3524, Adjusted R-Squared: 0.2175

Fonte: Elaborado pelo autor com uso do *software* Alteryx Designe

Observa-se que neste modelo de regressão linear, as variáveis POP\_URB (percentual de população residente em área urbana), EMP\_FOR (número de empregos formais) e ARE (área do município) apresentam significância estatística. A variável EMP\_FOR por sua vez, apresenta relação negativa com a variável alvo (coluna “*Estimate*”), ou seja, conforme aumenta o número de empregos formais, diminui o faturamento das filiais, o que contradiz a teoria econômica. Por fim, observa-se que R-quadrado e R-quadrado ajustado, assim como no primeiro modelo de regressão, apresentam baixa relação entre variáveis independentes e a variável alvo, confirmando a baixa capacidade de explicação das variações do faturamento em função das variações das variáveis preditoras.

Uma última modelagem contemplou apenas o uso das três variáveis que apresentaram significância estatística, conforme a Figura 8. A seguir, observa-se os resultados:

Figura 9 – Resultado do modelo de regressão com uso das variáveis com significância estatística

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	941720.5	508869.40	1.851	0.0696 .
POP_URB	1917032.3	629123.29	3.047	0.00355 **
EMP_FOR	-49.7	14.20	-3.499	0.00093 ***
ARE	100.1	36.89	2.714	0.00885 **

Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 499601 on 55 degrees of freedom

Multiple R-squared: 0.2568, Adjusted R-Squared: 0.2162

Fonte: Elaborado pelo autor com uso do *software* Alteryx Designer.

Este último modelo, cujos resultados são observados na Figura 9, novamente confirmam que nenhuma das variáveis tem potencial para subsidiar um modelo com capacidade de explicar as variações no faturamento das filiais nos mais diversos municípios em que atua. Mesmo com variáveis que apresentam significância, R-quadrado e R-quadrado ajustado imprimem a baixa correlação de variáveis preditoras com a variável preditiva.

Dado o fato de que o modelo de regressão desenvolvido não tem capacidade suficiente para explicar as variações de faturamento das filiais da empresa em estudo, a etapa de aplicação dos dados secundários (municípios em que a empresa não atua) não avançou, encerrando o estudo de caso sem atingir o objetivo proposto de recomendar cidades para abertura de novas filiais.

## 5 CONSIDERAÇÕES FINAIS

Tomar decisões é uma atividade constante nas organizações. Quando subsidiadas com informações consistentes, sua assertividade cresce, os resultados podem ser mais expressivos e reduz-se os riscos com perdas de recursos. As análises preditivas podem proporcionar aos decisores a possibilidade de serem mais eficientes, rentabilizando mais as atividades da organização em que atuam.

Neste sentido, assim que analisado o problema de negócio, foi possível entender quais informações poderiam ser úteis para auxiliar a empresa, objeto deste estudo de caso, a definir quais as melhores cidades gaúchas para colocar novas filiais e que ferramenta poderia fornecer tais informações. A criação de um modelo preditivo poderia antecipar informações que são extremamente importantes para o sucesso das atividades da organização.

Por meio deste estudo, ao construir o modelo preditivo e validá-lo, pôde-se observar em seus resultados, que os indicadores externos, qualitativos, mas principalmente quantitativos, não exercem influência significativa para impactar os resultados da empresa a nível de filial, o que não responde o problema de pesquisa “como determinar as melhores cidades para abrir novas filiais?”. Após identificar o perfil de cada cidade gaúcha em que a rede varejista de moda atua, através de seus dados socioeconômicos, estes foram empregados no modelo de regressão linear, o qual teve uma capacidade explicativa muito baixa.

Quanto aos objetivos específicos deste estudo, foi possível coletar os dados das cidades em que a rede varejista de moda atua, bem como dos demais municípios em que a mesma ainda não está presente. A partir destes dados, formulou-se uma regressão linear múltipla, visando atender o objetivo geral, que consistiu em identificar as cidades gaúchas com maior capacidade de gerar receitas, o qual não pode ser atingido, já que o modelo não teve capacidade explicativa, impossibilitando a elaboração de um relatório de recomendação aos tomadores de decisão.

Diante dos fatos, sugere-se o estudo de outros fatores que possam explicar o desempenho operacional de cada loja ou da organização como um todo. Fatores como equipe, treinamento contínuo, localização estratégica com pontos de grande fluxo, marketing, concorrência ou outros, podem ser determinantes para o sucesso e

os bons resultados esperados pela empresa ao tomar a decisão de entrar em uma nova cidade, com um novo mercado consumidor.

## REFERÊNCIAS

ALMEIDA, Adiel Teixeira de. **Processo de decisão nas organizações: construindo modelos de decisão multicritério**. São Paulo: Atlas, 2013.

COELHO, Lucas. **Analytics: O que é, conceito e definição**. 2017. Disponível em: <<https://www.cetax.com.br/blog/o-que-e-analytics/>>. Acesso em: 02 junho 2018.

COSTA, Clóvis Correia da. **Estratégia de negócios**. São Paulo: Saraiva, 2009.

DAL'BÓ, Reginaldo André. **Missão, visão e valores**. 2009. Disponível em: <<http://www.administradores.com.br/artigos/negocios/missao-visao-e-valores/28883/>>. Acesso em: 10 junho 2018.

GIL, Antonio Carlos. **Técnicas de Pesquisa em Economia e Elaboração de Monografias**. 4. ed. São Paulo: Atlas, 2002.

GOMES, L. F. A. M.; GOMES, C. F. S. **Tomada de decisão gerencial: enfoque multicritério**. 5. ed. São Paulo: Atlas, 2014.

GUAZZELLI, Alex. **O que é análise preditiva?**. 2012. Disponível em: <<https://www.ibm.com/developerworks/br/industry/library/ba-predictive-analytics1/index.html>>. Acesso em: 02 junho 2018.

GUJARATI, Damodar N. **Econometria básica**. 3. ed. São Paulo: Makron Books, 2000.

GUJARATI, D. N.; PORTER, D.C. **Econometria e séries temporais com aplicações a dados da economia brasileira**. 5. ed. Rio de Janeiro: LTC, 2014.

HILL, R. C.; GRIFFITHS, W. E.; JUDGE, G. G. **Econometria**. 2. ed. São Paulo: Saraiva, 2006.

HOFFMANN, Rodolfo. **Estatística para economistas**. 4. ed. São Paulo: Pioneira, 2006.

KAPLAN, R. S.; NORTON, D. P. **A estratégia em ação: balanced scorecard**. 3. ed. Rio de Janeiro: Campus, 1997.

KMENTA, Jan. **Elementos de Econometria**. São Paulo: Atlas, 1978.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de Metodologia Científica**. X. ed. São Paulo: Atlas, 2001.

MC GUIGAN, J. R.; MOYER, R. C.; HARRIS, F. H. de B. **Economia de Empresas**. São Paulo: South-Western Cengage Learning, 2010.

MILONE, Giuseppe. **Estatística Geral e Aplicada**. São Paulo: Thomson Learning, 2006.

PINDYCK, R. S.; Rubinfeld, D. L. **Microeconomia**. 5. ed. São Paulo: Prentice-Hall, 2002.

PORTAL ACTION. **Modelo Estatístico - Análise de Regressão**. Disponível em: <<http://www.portalaction.com.br/analise-de-regressao/21-modelo-estatistico>>. Acesso em: 06 abril 2018.

PORTAL ACTION. **Regressão Linear Múltipla - Análise de Regressão**. Disponível em: <<http://www.portalaction.com.br/analise-de-regressao/regressao-linear-multipla>>. Acesso em: 06 abril 2018.

PORTAL EDUCAÇÃO. **Autocorrelação**. Disponível em: <<https://www.portaleducacao.com.br/conteudo/artigos/contabilidade/autocorrelacao/52161>>. Acesso em: 02 junho 2018.

PRANDANOV, C. C.; FREITAS, E. C. **Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013.

ROBBINS, Stephen P. **Decida e conquiste: o guia definitivo para tomada de decisão**. São Paulo: Saraiva, 2015.

ROSSI, J. W.; NEVES, C. **Econometria e séries temporais com aplicações a dados da economia brasileira**. Rio de Janeiro: LTC, 2014.

SEBRAE. **Saiba quais são as perspectivas para o varejo da moda até 2019**. 2017. Disponível em: <<http://www.sebrae.com.br/sites/PortalSebrae/bis/saiba-quais-sao-as-perspectivas-para-o-varejo-da-moda-ate-2019,c37254843636b510VgnVCM1000004c00210aRCRD>>. Acesso em: 09 junho 2018.

SEBRAE. **Integre seus canais de vendas a partir do conceito de omnichannel**. 2017. Disponível em: <<http://www.sebrae.com.br/sites/PortalSebrae/artigos/integre-seus-canais-de-vendas-a-partir-do-conceito-de-omni-channel,87426f65a8f3a410VgnVCM2000003c74010aRCRD>>. Acesso em: 03 junho 2018.

SEBRAE. **Fast fashion ganha destaque no varejo de moda**. 2015. Disponível em: <<http://www.sebrae.com.br/sites/PortalSebrae/artigos/fast-fashion-ganha-destaque-no-varejo-de-moda,ef695d27e8fdd410VgnVCM1000003b74010aRCRD>>. Acesso em: 09 junho 2018.

TURBAN, Efraim et al. **Business intelligence: um enfoque gerencial para a inteligência do negócio**. São Paulo: Bookman, 2009.

UDACITY. **Como as empresas usam a análise preditiva na tomada de decisão**. 2017. Disponível em: <<https://br.udacity.com/blog/post/analise-preditiva-tomada-decisao>>. Acesso em: 02 junho 2018.

UDACITY. **A estrutura da solução de problemas**. 2018. Disponível em <<https://classroom.udacity.com/nanodegrees/nd008-br/parts/6e3d8340-c53d-4e80-9559-e1917fc1ccc6/modules/bf705147-0d7c-4492-842a-698a6410a8a3/lessons/3ae813d7-15b2-48e9-8bc7-25a23d3a9048/concepts/1eb8f83e-6803-4b1a-b5db-e07d89234633>>. Acesso em: 02 junho 2018.

VORHIES, William. **CRISP-DM – a Standard Methodology to Ensure a Good Outcome**. 2016. Disponível em: <<https://www.datasciencecentral.com/profiles/blogs/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome>>. Acesso em: 02 junho 2018.

YU, Abraham Sin Oih. **Tomada de decisão nas organizações: uma visão multidisciplinar**. São Paulo: Saraiva, 2011.