



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Bacharelado em Estatística

**ANÁLISE MULTIVARIADA DA
INFLAÇÃO INTEGRADA AOS DADOS
SOBRE EMPREENDEDORISMO NO
BRASIL**

Bianca Jaqueline Alves da Silva

**Uberlândia-MG
2018**

Bianca Jaqueline Alves da Silva

**ANÁLISE MULTIVARIADA DA
INFLAÇÃO INTEGRADA AOS DADOS
SOBRE EMPREENDEDORISMO NO
BRASIL**

Trabalho de conclusão de curso apresentado à Coordenação do Curso de Bacharelado em Estatística como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Priscila Neves de Faria

**Uberlândia-MG
2018**



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, ____ de _____ de 20____

BANCA EXAMINADORA

Priscila Neves de Faria

Maria Imaculada de Sousa Silva

Mirian Fernandes Carvalho Araújo

**Uberlândia-MG
2018**

AGRADECIMENTOS

Agradeço a Deus pelo dom da vida; pela saúde, coragem e força para chegar até aqui; pela família e pelos amigos que colocou em meu caminho, por estar sempre comigo e não me deixar desistir.

Agradeço a minha família, pela paciência e carinho durante todos esses anos. Especialmente à minha mãe Rita de Cássia e minha avó Maria Cândida pelo incentivo de sempre, me fizeram olhar além das dificuldades, me fizeram acreditar que era possível, e sempre intercederam por mim. Ao meu irmão Wilton Gabriel que mesmo de longe sempre me ajudou com palavras de incentivo, serei eternamente grata.

À Universidade e toda sua direção eu deixo uma palavra de agradecimento por todo ambiente inspirador e pela oportunidade de concluir este curso.

A todos os professores eu agradeço a orientação repleta de conhecimento, sabedoria e paciência. Vocês são maravilhosos!

À minha orientadora Priscila Neves Faria pelo suporte, incentivo, paciência e correções, meu maior exemplo de professora, nos ensina com amor, além de ensinar a matéria nos ensina para a vida. Você é espetacular!

As minhas amigas de infância Késia e Franciele pelo carinho e apoio de sempre.

Sou grata pela vida de todas as pessoas que fizeram parte desta etapa decisiva em minha vida.

Muito obrigada! Que venha o futuro!

RESUMO

Geralmente, em cenário de crise econômica, surgem novos empreendedores. A partir desta constatação, surgiu a ideia de coletar dados de fontes secundárias, em duas publicações (GEM e IPCA-15) que evidenciam reflexos do cenário macroeconômico brasileiro. Realizar a integração das informações referentes ao empreendedorismo com as informações referentes à inflação nas regiões metropolitanas foi o principal objetivo desta pesquisa. Para a análise dos dados relacionados às evidências das atividades empreendedoras foi utilizado o teste t-Student com análises estatísticas indicando a significância das motivações empreendedoras – por oportunidade e por necessidade. Além disso, foi utilizada a técnica multivariada de Análise de Agrupamentos, avaliando no período de 2014 a 2016, quais regiões metropolitanas possuem comportamento similar diante dos itens de consumo da população medidos pelo IPCA-15. Como resultados notou-se que dentre as principais atividades empreendedoras onde mais empreenderam destacou-se a de serviços domésticos. Na Análise de Agrupamentos a cidade de Curitiba formou um único grupo devido aos seus altos níveis de inflação nas diversas categorias.

Palavras-chave: Análise de Agrupamento, IPCA, empreendedorismo.

ABSTRACT

Generally, in a scenario of economic crisis, new entrepreneurs arise. Based on this finding, the idea of collecting data from secondary sources arose in two publications (GEM and IPCA-15) that reflect the Brazilian macroeconomic scenario. The main objective of this research was to integrate information about entrepreneurship with inflation information in metropolitan regions. For the analysis of the data related to the evidences of the entrepreneurial activities, the t-Student test was used with statistical analyzes indicating the significance of the entrepreneurial motivations - by opportunity and necessity. In addition, the multivariate cluster analysis technique was used, evaluating in the period from 2014 to 2016, which metropolitan regions have similar behavior in relation to the consumption items of the population measured by the IPCA-15. As a result, it was noted that among the main entrepreneurial activities where the most undertaken were the domestic services. In the Analysis of Clusters the city of Curitiba formed a single group due to its high levels of inflation in the various categories.

Keywords: Cluster Analysis, IPCA, Entrepreneurship.

SUMÁRIO

1	Introdução	1
2	Fundamentação Teórica	3
2.1	IPCA-15	3
2.2	Empreendedorismo	5
2.2.1	Origem do empreendedorismo	5
2.2.2	O perfil empreendedor	6
2.2.3	Empreendedorismo no Brasil	7
2.3	Análise Multivariada	7
2.3.1	Análise de agrupamentos	8
2.3.2	Medidas de distância	9
2.3.3	Método de agrupamentos	9
2.3.4	Métodos de agrupamentos hierárquicos aglomerativos	10
2.4	Método da Ligação média	11
3	Metodologia	13
3.1	Base de dados	13
3.2	Medidas de Similaridade e Dissimilaridade	14
3.3	Coeficiente de Correlação Cofenético	15
3.4	Análise de Variância (ANOVA)	16
4	RESULTADOS E DISCUSSÕES	19
5	Conclusões	25
	Referências Bibliográficas	27

1. INTRODUÇÃO

No Brasil, o aumento dos preços em diversos setores de consumo da população nos últimos anos, tem contribuído cada vez mais para a competitividade entre as empresas, isso porque a disputa por preços baixos estimula a concorrência no mercado fazendo que grandes instituições se sobressaíam e consequentemente tenham maior poder de barganha.

[37] argumenta que, particularmente no caso brasileiro, diante de qualquer suspeita de elevação do nível de preços, segundo a regra ortodoxa de estabilização de preços, eleva-se a taxa de juros, mas não há dúvida de que tal política também reduz o nível de investimentos privados [35];[30] e, pelo efeito multiplicador, resfria toda a economia e aumenta o desemprego.

Estudos realizados por [24], mostraram que a utilização da técnica multivariada passou a ser familiarizada devido a aplicação dos acadêmicos e profissionais de marketing em estudos que identificaram a preferência e satisfação do consumidor, bem como analisar o perfil e comportamento de compras. Segundo [5], estas técnicas podem ser utilizadas para resolver problemas em diversas áreas do conhecimento.

Um método que pode ser utilizado para classificar objetos em categorias de acordo com a similaridade é a análise de agrupamento (cluster analysis). Essa técnica considera um conjunto inicial de objetos, aos quais são associadas medidas de várias grandezas, denominadas variáveis classificatórias, utilizadas para se obter grupos de objetos semelhantes em relação aos valores assumidos por essas variáveis [10].

Além de possibilitar a construção de grupos de acordo com as similaridades dos indivíduos, a análise de agrupamento possibilita também representá-los de maneira bidimensional, através de um dendrograma [28].

A instabilidade econômica vivenciada no país, principalmente nos anos de 2014 e 2015, bem como as atividades econômicas sendo influenciadas por um ciclo recessivo, pode levar populações (que muitas vezes não encontra trabalho assalariado) de países menos desenvolvidos, como o caso do Brasil, a empreender para sobreviver [13].

Nesse sentido, surge uma inquietação para o estudo: compreender o perfil do novo empreendedor brasileiro, bem como analisar as atividades onde empreendem, por oportunidade versus necessidade, à luz do GEM - Global Entrepreneurship Monitor que em português significa “Monitoramento de empreendedorismo global”, e do IPCA, no período 2014-2016.

Assim, essa pesquisa se propõe como objetivo geral, identificar as regiões metropolitanas brasileiras com comportamentos similares diante ao IPCA – 15 no período de 2014 a 2016; identificar o comportamento do IPCA 15 em cada região metropolitana entre 2014 a 2016, aplicar

a Análise de Agrupamentos nos dados a fim de obter o gráfico Dendograma, analisar o perfil do novo empreendedor brasileiro, bem como as atividades onde empreendem, por oportunidade versus necessidade, à luz do GEM e do IPCA, no período de 2014 a 2016.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 IPCA-15

O Sistema Nacional de Índices de Preços ao Consumidor (SNIPC) foi concebido em 1978, constituindo-se numa combinação de processos destinados a produzir índices de preços ao consumidor nacionais, a partir da agregação de resultados regionais. Foi criado com o propósito de garantir uma mesma concepção metodológica no que diz respeito à fórmula de cálculo, pesquisas básicas, bases cadastrais de produtos e locais, montagem da estrutura de pesos e método de cálculo. Desde sua origem, foram incorporados e desativados alguns índices de preços do SNIPC, que apresenta em sua configuração atual os seguintes índices de preços: IPCA, Ipca-15, IPCA-E e INPC.

Segundo o IBGE as áreas geográficas pesquisadas foram implantadas na seguinte ordem cronológica: Rio de Janeiro (janeiro/1979); Porto Alegre, Belo Horizonte e Recife (junho/1979); São Paulo, Brasília e Belém (janeiro/1980); Fortaleza, Salvador e Curitiba (outubro/1979); e Goiânia (janeiro/1991).

A partir do mês de maio de 2000, o Índice Nacional de Preços ao Consumidor Amplo – 15 (IPCA-15) [16];[17] passou a ser produzido e disponibilizado pelo IBGE. O IPCA-15 é o próprio IPCA, porém considerando-se outro período de coleta de preços, ou seja, do dia 16 do mês anterior ao dia 15 do mês atual enquanto que o período de coleta do IPCA vai do dia 1º ao dia 30 ou 31, dependendo do mês. Neste trabalho os dados utilizados serão do IPCA-15, que é uma prévia do IPCA (Índice de Preços ao Consumidor Amplo), o indicador oficial da inflação no país. Como ele realiza a medição de preços em um período não calculado pelo IPCA, mostra qual será a tendência do resultado do final do mês. Além disso, o IBGE tem uma comparação mais precisa da alta e queda dos preços, pois a cada 15 dias um dos índices é divulgado. A coleta de dados para a medição do IPCA e do IPCA-15 é uma só. O que muda é o período analisado em cada um dos itens. O IPCA-15 tem por objetivo medir a inflação de um conjunto de produtos e serviços comercializados no varejo, referentes ao consumo pessoal das famílias, cujo rendimento varia entre 1 e 40 salários-mínimos, qualquer que seja a fonte de rendimentos. Esta faixa de renda foi criada com o objetivo de garantir uma cobertura de 90% das famílias pertencentes as áreas urbanas de cobertura do SNIPC, qualquer que seja a fonte dos rendimentos.

Por ser considerado o índice oficial para medir a inflação no Brasil, é normal que os valores do IPCA sejam bastante semelhantes aos da inflação acumulada. De maneira geral, o escopo

metodológico do IPCA abrange os seguintes temas: montagem da estrutura geral de pesos; definição das bases cadastrais de produtos e locais; coleta de preços e método de cálculo.

As estruturas de ponderadores são montadas utilizando-se uma organização de códigos em agrupamentos logicamente estabelecidos de forma que ficam juntas as categorias de consumo de mesma natureza, hierarquicamente estruturadas em grupos, subgrupos, itens e subitens.

Estes últimos representam o nível mais desagregado para o qual se obtêm os pesos utilizados no cálculo dos índices de preços. Estes ponderadores retratam o grau de importância ou representatividade dos subitens pertencentes à cesta de consumo das famílias, que são constituídas a partir dos hábitos de consumo da população-alvo da pesquisa. A coleta de preços é realizada a partir da definição dos cadastros de informantes e de produtos, seguindo métodos de coleta.

Na geração do cadastro de informantes são utilizadas duas linhas de procedimentos, conforme a natureza das diversas mercadorias pesquisadas. A linha principal consiste no levantamento de informantes através da Pesquisa de Locais de Compra - PLC, que define onde coletar os preços da grande maioria dos subitens. A segunda linha adota procedimentos específicos para subitens cujas peculiaridades assim o exigiam - os chamados subitens "extra-PLC" - para os quais a metodologia da PLC não é adequada, sendo necessário tratamento especial, ou seja, métodos específicos para determinar onde coletar preços. São exemplos de subitens "extra-PLC" o aluguel de moradia, empregados domésticos, condomínio, serviços públicos e taxas, etc. Portanto, o cadastro de informantes é essencialmente formado por estabelecimentos comerciais de venda de produtos e prestadores de serviços; domicílios alugados; empresas concessionárias responsáveis pela prestação de serviços; órgãos oficiais; além de alguns profissionais autônomos, como médicos e dentistas.

Para definir o conjunto dos produtos que compõe o cadastro, considera-se a representatividade da totalidade dos produtos consumidos pela população a que o índice se refere. Sendo assim, o ponto de partida para a geração do cadastro de produtos é a relação dos subitens componentes da estrutura de pesos de cada área e de cada faixa de renda. É realizada uma Pesquisa de Especificação de Produtos e Serviços - PEPS que serve de base para a definição do cadastro de produtos, caracterizando os níveis de especificação utilizados na coleta de preços.

Quanto ao método de cálculo, estima-se o índice do subitem, calcula-se a média aritmética dos preços pesquisados em diferentes estabelecimentos comerciais para cada produto pesquisado no mês corrente.

Através do mesmo processo, este preço médio é comparado com o resultado obtido no mês anterior; em seguida, para se calcular o índice do subitem, aplica-se a média geométrica simples para agregação dos resultados dos produtos pertencentes ao subitem; e, por último, para todos os níveis superiores de agregação, emprega-se a fórmula de Laspeyres. Para aplicação desta fórmula de cálculo, os relativos de preços são ponderados por valores de despesas obtidos a partir da Pesquisa de Orçamentos Familiares.

O método de cálculo para a obtenção do índice nacional é uma média aritmética ponderada dos nove índices metropolitanos mensais, além do município de Goiânia e Brasília. Para agregar os índices regionais, a variável utilizada como ponderador é o rendimento total urbano (POF

2008-2009). Quanto à série histórica de números-índices, a base está posicionada em dezembro de 1993, expressa em valor igual a cem (base=100).

2.2 EMPREENDEDORISMO

2.2.1 ORIGEM DO EMPREENDEDORISMO

Através do GEM – Global Entrepreneurship Monitor – pode-se perceber, por exemplo, que “a recuperação da economia passa, necessariamente, pelo empreendedorismo, e que é preciso que o governo crie mecanismos que facilitem e incentivem os pequenos negócios” [13].

A pesquisa GEM iniciou-se no ano de 1999 e é, atualmente, a mais abrangente pesquisa anual sobre atividades empreendedoras globais, na qual o Brasil faz parte, e que explora o papel do empreendedorismo no desenvolvimento social e econômico de um país. O GEM apresenta as características, atividades e aspirações dos empreendedores, bem como faz a qualificação do ambiente, bem como os fatores que favorecem ou criam obstáculos à criação de atividades empreendedoras.

Considerando as diversas sessões que a pesquisa GEM oferece, algumas variáveis são relevantes, tais como: perfil dos empreendedores, atividades iniciadas por oportunidade ou por necessidade, renda, escolaridade, além dos principais setores de atividades onde houve criação de novos negócios ao longo do tempo. A pesquisa é realizada com a população adulta com amostra de indivíduos entre 18 a 64 anos do país [13].

O tema empreendedorismo é secular, tão antigo que achar obras literárias sobre o tema não é uma tarefa difícil, aliás, é natural já que se trata de um tema tão popular e antigo. Segundo [6] “empreendedorismo não é um tema novo ou modismo: existe desde sempre, desde a primeira ação humana inovadora, com o objetivo de melhorar as relações do homem com os outros e com a natureza”. Se existem muitas obras sobre o assunto, as definições do tema são as mais diversas e sendo apresentados vários conceitos, mas a grande maioria converge no mesmo ponto, a inovação, aliada a força de vontade, e a grande busca de resultados.

Empreendedorismo é a criação de valor por pessoas e organizações trabalhando juntas para implementar uma ideia por meio da aplicação de criatividade, capacidade de transformar e o desejo de tomar aquilo que comumente se chamaria de risco. [29]

De acordo com a definição de Barreto [2], empreendedorismo é a habilidade de criar e instituir algo partindo de muito pouco, assim Barretos enfatiza a grande importância do trabalho, além da capacidade de maximizar recursos. Uma outra definição interessante é:

(...) a essência do empreendedorismo está na percepção e no aproveitamento das novas oportunidades no âmbito dos negócios (...) sempre tem haver em criar novas formas de uso dos recursos nacionais em que eles sejam deslocados de seu emprego tradicional e sujeitos a novas combinações. [11]

Segundo [6], “o empreendedorismo é uma revolução silenciosa, que será para o século 21 mais do que a revolução industrial foi para o século 20”, ao comparar com a revolução industrial, a grande responsável por radicais mudanças no século 20, demonstra o grau de importância para a sociedade do tema empreendedorismo.

Segundo o Serviço Brasileiro de Apoio às Micro e Pequenas Empresas – SEBRAE [36] o empreendedor tem como característica básica o espírito criativo e pesquisador. Ele está constantemente buscando novos caminhos e novas soluções, sempre tendo em vista as necessidades das pessoas. A essência do empresário de sucesso é a busca de novos negócios e oportunidades e a preocupação sempre presente com a melhoria do produto. Enquanto a maior parte das pessoas tende a enxergar apenas dificuldades e insucessos, o empreendedor deve ser otimista e buscar o sucesso, apesar das dificuldades. Já para [7] “empreendedorismo é o envolvimento de pessoas e processos que, em conjunto, levam à transformação de ideias em oportunidades” enfatiza assim de forma mais geral o real objetivo do empreendedorismo, que é gerar oportunidades.

Todas estas definições exprimem de forma muito clara o que vem a ser o tema empreendedorismo, porém cabe uma definição mais sucinta, a qual traga palavras-chave. Assim: empreendedorismo é o ato de aproveitar oportunidades, inovar, planejar, arriscar, empenhar, ser perseverante, acreditar na ideia e transformar em realidade, este ato se aplica em qualquer área, seja um novo negócio, seja um novo processo ou um novo produto, um novo método, tanto faz.

2.2.2 O PERFIL EMPREENDEDOR

Nas universidades, cursos, seminários e até exposições sobre o assunto empreendedorismo, é rotineiro que se discuta sobre o ato de empreender e o perfil de quem realiza essa ação, o empreendedor, neste contexto são apresentadas várias características traçando assim o perfil empreendedor.

Segundo [7] “o empreendedor é aquele que faz as coisas acontecerem, se antecipa aos fatos e tem uma visão futura da organização”, já para [4] “o empreendedor é a pessoa que inicia ou opera um negócio para realizar uma ideia ou projeto pessoal assumindo riscos e responsabilidades e inovando continuamente”.

O autor [4] amplia o conceito dizendo que “o empreendedor é a pessoa que destrói a ordem econômica existente graças à introdução no mercado de novos produtos e serviços, pela criação de novas formas de gestão ou pela exploração de novos recursos, materiais e tecnologias”.

Enfim o empreendedor é:

“[...] um insatisfeito que transforma seu inconformismo em descobertas e propostas positivas para si mesmo e para os outros. É alguém que prefere seguir caminhos não percorridos, que define a partir do indefinido, acredita que seus atos podem gerar consequências. Em suma, alguém que acredita que pode alterar o mundo. É protagonista e autor de si mesmo e, principalmente, da comunidade em que vive.”
[6]

A literatura traz também várias características listadas sobre o ímpeto empreendedor, como uma receita de bolo, onde são enunciados vários quesitos aparentemente “obrigatórios”. Segundo [4] ser uma pessoa com a necessidade de realização, ter a disposição para assumir riscos, possuir autoconfiança, ter vontade de trabalhar duro, ter habilidade de comunicação, conhecer maneiras de organizar o trabalho, ter orgulho daquilo que faz, manter boas relações interpessoais, um auto propulsor, assumir responsabilidades e desafios, ser honesto, dentre outras.

2.2.3 EMPREENDEDORISMO NO BRASIL

Segundo o GEM - Global Entrepreneurship Monitor [13], pesquisa que mede a capacidade empreendedora de vários países, o Brasil é o sétimo colocado no ranking de nações mais empreendedoras do mundo, no país existem cerca de 15 milhões de empreendedores. É importante frisar que para o GEM o empreendedorismo é qualquer tentativa de criação de um novo negócio ou novo empreendimento, qualquer que seja como, por exemplo, uma atividade autônoma, uma nova empresa, ou a expansão de um empreendimento já existente, por um indivíduo, grupos de indivíduos ou por empresas já estabelecidas. Estes dados, porém, são muito criticados, já que não consideram o fato de a grande maioria das empresas fechar antes do primeiro ano de vida.

Segundo o IBGE [15] “Os pequenos negócios, formados por quitandas, mercearias, sapatarias, cabeleireiros, bazares, armarinhos, etc., vêm ganhando gradativamente mais participação nos setores de comércio e serviços, aumentando significativamente a geração de postos de trabalho e apresentando crescimento da receita operacional líquida”, isso mostra o perfil das empresas brasileiras.

Dornelas [8] afirma que um dos grandes motivos do crescimento no número de empresas abertas no país se dá pelos altos índices de desemprego, onde o desempregado usa o valor de sua rescisão provida de seus direitos trabalhistas, e sem informação e planejamento, se aventura no mercado, visando muitas vezes abrir não uma empresa, mas um “emprego” e acaba decretando falência em um período curto de existência.

Assim o perfil da grande maioria dos empreendedores brasileiros consiste em pessoas sem muito grau de instrução, na maioria das vezes sem nenhuma instrução a respeito de gestão ou empreendedorismo. O SEBRAE [36] estima que apenas 5% buscam orientação, a grande maioria nem conhece o ramo de mercado que pretende atuar e nem o produto que se propõe a vender, jamais ouviu falar de “plano de negócios”, e nunca frequentou uma universidade.

2.3 ANÁLISE MULTIVARIADA

Segundo Prearo [33], determinado fenômeno pode ser observado e medido por diferentes variáveis, ou seja, as causas de determinado problema podem ser oriundas de diversos motivos. Dessa forma, a análise multivariada remete a observação de fenômenos que podem ser medidos a partir de muitas variáveis, permitindo assim a interpretação de problemas com comportamentos complexos.

De acordo com Steiner [39], o grande potencial da análise multivariada está na necessidade de compreender o relacionamento entre diversas variáveis. Contudo, Lourenço e Matias [20] ressaltam que devido ao maior envolvimento de diferentes variáveis as técnicas estatísticas multivariadas são mais complexas.

Para Hair et al. [14], a análise multivariada contribui para a formulação de questões complexas de forma específica e concisa, já que possibilita a condução de pesquisas significativas. Existem diferentes técnicas que compõem a análise multivariada, entre elas as mais utilizadas são: análise de agrupamento; escalonamento multidimensional; análise fatorial; análise discriminante; correlação canônica; análise multivariada da variância; análise conjunta; análise de regressão; análise de regressão logística; modelagem de equações estruturais e análise de correspondência [31].

Estabelecer os objetivos e definir qual técnica de análise multivariada será utilizada é uma tarefa um tanto complicada, visto que existe alto grau de nuances entre as técnicas e também por apresentarem inúmeros objetivos [18].

2.3.1 ANÁLISE DE AGRUPAMENTOS

A Análise de Agrupamentos, conhecida também como Análise de Conglomerados ou Análise de Cluster (Cluster Analysis), tem como objetivo dividir a amostra pesquisada em grupos, de modo que os elementos inseridos em um mesmo aglomerado apresentem características similares e os elementos em grupos distintos sejam heterogêneos diante as mesmas características.

A Análise de Agrupamentos pode ser utilizada em diferentes situações e campos de estudo, sendo possível sua aplicação desde em pesquisa de mercado para identificar o posicionamento de indústrias [27], até mesmo em estudos da medicina [26].

Pestana e Gageiro (2014) afirmam que a Análise de Cluster é um procedimento exploratório multivariado (pertencente à família da análise multivariada) que serve para detectar grupos homogêneos nos dados, podendo ser constituídos de variáveis ou casos.

Ressalta Matarezo [23] que a Análise de Agrupamentos é utilizada para unir objetos com base nas características que eles possuem e assim os grupos formados exibem elevada homogeneidade interna e alta heterogeneidade externa. O autor ainda afirma que a técnica é útil para analisar dados em diferentes situações, visto que ela possibilita que as informações de populações inteiras sejam reduzidas em subgrupos menores e específicos.

De acordo com Frei [12], a Análise de Agrupamentos é composta por cinco passos, são eles: obtenção da matriz de dados; padronização da matriz quando necessário; cálculo da matriz de similaridade ou dissimilaridade; escolha do(s) método(s) de agrupamento(s); e a decisão do número de grupos. Quanto maior for a medida de dissimilaridade (exemplo distância euclidiana), menor será a semelhança entre os indivíduos, porém quanto maior for a medida de similaridade (exemplo coeficiente de correlação) maior semelhança entre os indivíduos.

Contudo, para decidir se os elementos de uma amostra possuem aspectos similares ou não é importante adotar critérios que possam mensurar tal semelhança, para isto, consideram-se medidas de distâncias capazes de descrever a similaridade da amostra diante das características

medidas. Tais critérios são nomeados como medidas de similaridade e dissimilaridade.

Sendo assim, o principal objetivo da análise de conglomerados é agrupar objetos ou variáveis semelhantes. Para Pestana e Gageiro [32], a técnica permite detectar grupos homogêneos dentre os dados, podendo os grupos serem constituídos por variáveis ou casos. De acordo com Maroco [21], esta análise envolve agrupar sujeitos ou variáveis em grupos homogêneos a partir de uma ou mais características comuns.

2.3.2 MEDIDAS DE DISTÂNCIA

Segundo Frei [12] os métodos estatísticos devem organizar os objetos em grupos homogêneos, ou seja, os aglomerados formados devem ser constituídos por indivíduos com características semelhantes. Para medir a similaridade entre os indivíduos é necessária a escolha de medidas de distância, porém é importante observar se natureza das variáveis é qualitativa ou quantitativa.

De acordo com Vieira [40] variáveis quantitativas são aquelas expressas por meio de números e que possuem significado em determinada escala numérica. Existem diferentes métodos para o cálculo da distância entre variáveis quantitativas, entre eles está a distância eucliana e também a distância euclidiana padronizada.

De acordo com Hair et al. [14] a distância euclidiana também é reconhecida como distância em linha reta, visto que ela retrata a distância do comprimento da hipotenusa de um triângulo retângulo, podendo ser generalizado em casos de mais de duas variáveis.

Depois de selecionada a medida de distância é importante a avaliação das grandezas das variáveis estudadas, já que as diferenças discrepantes entre os dados podem afetar a formação dos agrupamentos, uma vez que características de maior dispersão, ou seja, maiores desvios padrão tendem a impactar sobre o valor de similaridade final.

Para que a grandeza das variáveis não afete a formação dos grupos é necessário padronizar os dados, de forma que todas as informações apresentarem a mesma escala. Segundo Hair et al. [14] a forma mais comum de padronização é a submissão da média e divisão pelo desvio padrão para cada variável, método conhecido como escores Z.

2.3.3 MÉTODO DE AGRUPAMENTOS

Segundo Mingoti [25] os métodos de agrupamentos são técnicas utilizadas para a construção dos conglomerados, estes são classificados como técnicas hierárquicas e não hierárquicas. Os procedimentos hierárquicos são aqueles nos quais não é conhecido a priori o número de grupos a serem formados, logo, a técnica é utilizada em pesquisas exploratórias cujo o objetivo é identificar possíveis agrupamentos e o valor provável do número de grupos.

Mingoti [25] ressalta que os métodos hierárquicos são divididos em aglomerativos e divisivos, sendo que a técnica aglomerativa parte do princípio que no início do processo de agrupamento tem-se n grupos, ou seja, cada indivíduo dos dados analisados é considerado como um grupo isolado (Figura 2.1). Dessa forma, os indivíduos vão se unindo e formando os grupos de acordo com as características similares entre eles, logo, no estágio inicial do processo de agrupamento

2.4 MÉTODO DA LIGAÇÃO MÉDIA

A ligação média entre clusters é um método não-ponderado de agrupamento aos pares, utilizando médias aritméticas das medidas de dissimilaridade, que evita caracterizar a dissimilaridade por valores extremos (máximo ou mínimo) [5]. Este método trata a distância entre dois conglomerados como a média das distâncias entre todos os pares de elementos que podem ser formados com os elementos dos dois conglomerados que estão sendo comparados.

A grande vantagem deste critério é tornar as consequências da existência de valores extremos e considerar toda a informação dos grupos. Só que, dependendo do tipo de cluster que se espera obter, esta propriedade também pode ser vista como uma desvantagem [34].

3. METODOLOGIA

3.1 BASE DE DADOS

Para construção deste trabalho foram utilizados dados secundários obtidos diretamente da página de Dados Abertos do Governo Federal (<http://dados.gov.br/tag/IBGE>) que possui livre acesso, sem restrições de licenças, patentes ou mecanismos de controle. Os dados têm como base o IPCA (Índice de Preços ao Consumidor Amplo) responsável em medir a variação dos preços de uma cesta de produtos e serviços consumidos pela população, sendo seu objetivo verificar a modificação dos valores comparando de um mês a outro. O IBGE (Instituto Brasileiro de Geografia e Estatística) constrói a cesta de produtos e serviços consumidos pela população e investiga quanto do rendimento familiar é gasto em cada produto, ato este que recebe o nome de POF (Pesquisa de Orçamento Familiar).

Para análise do empreendedorismo no Brasil, foi utilizado dados secundários diretamente da publicação do GEM, que evidenciam reflexos do cenário macroeconômico brasileiro. Para a análise dos dados relacionados às evidências das atividades empreendedoras foi utilizado o teste t-Student com análises estatísticas indicando a significância das motivações empreendedoras – por oportunidade e por necessidade.

O IPCA mostra a variação das categorias de consumo, são elas: alimentação e bebidas, habitação, artigos de residência, vestuário, transportes, saúde e cuidados pessoais, despesas pessoais, educação e comunicação. Dessa forma, é analisada a variação dos preços diante as nove regiões brasileiras que são: Regiões metropolitanas de Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba e Porto Alegre.

Foram utilizados métodos estatísticos para alinhamento da base de dados com o objetivo do trabalho. Primeiramente, as categorias de consumo do IPCA foram nomeadas como variáveis e as regiões metropolitanas são os indivíduos da matriz, assim, foi realizado um estudo estatístico descritivo, baseando-se no cálculo da média aritmética, desvio padrão, coeficiente de variação e representação gráfica dos dados.

Como o intuito da pesquisa é identificar quais as regiões metropolitanas apresentam comportamento similar diante o IPCA, os dados foram analisados mediante aplicação da Análise de Variância, teste t-Student e Análise de Agrupamentos.

3.2 MEDIDAS DE SIMILARIDADE E DISSIMILARIDADE

Quando se usam métodos de agrupamentos hierárquicos aglomerativos, é preciso agrupar os dados com base na proximidade mútua ou na similaridade. Segundo [25], é necessário considerar medidas que descrevam a similaridade entre elementos amostrais de acordo com as características que neles foram medidas. Se, para cada elemento amostral, têm-se informações de p variáveis armazenadas em um vetor, a comparação de diferentes elementos amostrais pode ser feita através de medidas matemáticas (métricas), que possibilitam a comparação de vetores, de observações dos elementos amostrais e agrupam aqueles de menor distância. Quando as variáveis em estudo possuem propriedades métricas e/ou quantitativas, um modo óbvio de se refletir a proximidade de dois objetos é com uma medida de distância. Ainda de acordo com [25], em outras situações, a pesquisa envolve a análise de variáveis qualitativas, e, nestes casos, há duas formas de tratamento: ou transformam-nas em variáveis quantitativas e usam-se medidas de similaridades, ou trabalha-se com coeficientes de similaridades e, neste caso, comparam-se os elementos de acordo com a presença ou ausência de certas características, que podem ser representadas por variáveis binárias do tipo 0 ou 1.

Se as variáveis são quantitativas, para que se possa proceder ao agrupamento dos elementos, é necessário que se decida a priori a medida de similaridade ou dissimilaridade que será utilizada, pois existem várias diferentes, e cada uma delas produz um determinado tipo de agrupamento. De acordo com [25], as medidas de distâncias são medidas de dissimilaridade e, logo, quanto menor os seus valores, mais similares serão os elementos que estão sendo comparados. Segundo a mesma autora, considerando um conjunto de dados de n elementos amostrais, tendo-se medido p -variáveis aleatórias em cada um deles, para cada elemento amostral j , tem-se o vetor de medidas X_j , definido por:

$$X_j = [X_{1j}, X_{2j}, \dots, X_{pj}]', \text{ sendo } j = 1, 2, \dots, n.$$

Estabelecida a medida de parença, é construída a matriz de distâncias, que é uma matriz quadrada com o valor das medidas (similaridade ou dissimilaridade) entre todos os elementos. Uma ilustração da matriz é apresentada na figura 3.1. É importante destacar que qualquer medida de distância usada para variáveis quantitativas pode ser transformada num coeficiente de similaridade.

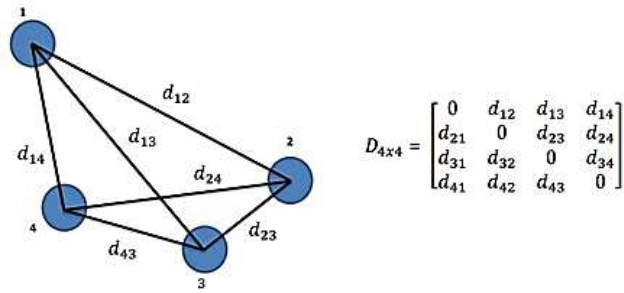


Figura 3.1: Matriz de distâncias entre 4 elementos.

Fonte: MINGOTI, S.A., 2007.

A distância Euclidiana entre dois elementos X_l e X_k , $l \neq k$, é definida por meio da expressão:

$$d_{X_l, X_k} = \sqrt{\sum_{i=1}^p (X_{il} - X_{ik})^2}$$

ou seja, os dois elementos amostrais são comparados em cada variável pertencente ao vetor de observações.

A distância euclidiana não preserva a ordem das distâncias com a mudança de escala, e, como as variáveis X são frequentemente medidas em unidades diferentes, é comum proceder-se à padronização das variáveis antes de se obter o valor das distâncias. Assim, atribui-se peso igual a cada variável padronizada, de maneira que elas sejam igualmente importantes, após a padronização, na determinação da proximidade relativa dos objetos no espaço. A transformação de dados mais difundida é a padronização das variáveis, de acordo com a fórmula $Z_j = \frac{X_i}{\hat{\sigma}_j}$, em que σ_j é o desvio padrão associado à j -ésima característica.

3.3 COEFICIENTE DE CORRELAÇÃO COFENÉTICO

Um coeficiente de correlação nem sempre é uma medida apropriada de similaridade, ele é uma medida de covariância que é também um tipo de proximidade, mas não necessariamente de similaridade. Quando as variáveis são quantitativas, uma medida de relacionamento natural é o coeficiente de correlação de Pearson, o qual expressa a similaridade de duas variáveis relativas à relação linear. Quanto maior for o valor do coeficiente de correlação em valor absoluto, maior será a proximidade entre as variáveis, isto é, mais linearmente relacionadas elas estarão. O coeficiente de correlação de Pearson é dado por:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Segundo [3], a correlação cofenética equivale à correlação de Pearson entre a matriz de similaridade original e aquela obtida após a construção do dendrograma, sendo que, quanto mais próxima de 1, menor será a distorção provocada pelo agrupamento dos indivíduos. A correlação cofenética mede o grau de ajuste entre a matriz de similaridade original (matriz S)

e a matriz resultante da simplificação proporcionada pelo método de agrupamento (matriz C). No caso, C é aquela obtida após a construção do dendrograma. Tal correlação foi calculada usando:

$$r_{cof} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - \bar{c})(s_{ij} - \bar{s})}{\sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - \bar{c})^2} \sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (s_{ij} - \bar{s})^2}}$$

em que:

c_{ij} : valor de similaridade entre os indivíduos i e j , obtidos a partir da matriz cofenética;

s_{ij} : valor de similaridade entre os indivíduos i e j , obtidos a partir da matriz de similaridade;

$$\bar{c} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_{ij}$$

$$\bar{s} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n s_{ij}$$

[19] afirma que, como o dendrograma é uma simplificação em duas dimensões de uma relação n -dimensional, é inevitável que algumas distorções quanto à similaridade apareçam. A medida de tal distorção pode ser obtida pelo coeficiente de correlação cofenético entre os valores da matriz inicial de similaridade e aqueles derivados do dendrograma.

[38], citado por [25], diz que coeficientes de associação não paramétricos como o de Spearman e o de Kendall também podem ser utilizados para o agrupamento das variáveis quantitativas. Existem algumas medidas de similaridade específicas para variáveis categóricas, como o coeficiente qui-quadrado, o de contingência de Pearson e o de concordância de Kappa.

3.4 ANÁLISE DE VARIÂNCIA (ANOVA)

A Análise de Variância (Analysis of Variance – ANOVA) representa um teste utilizado para comparar médias em mais de duas populações. [22] definiu a ANOVA sendo um método estatístico, que, por meio de teste de igualdade de médias, verifica se fatores (variáveis independentes) produzem mudanças sistemáticas em alguma variável de interesse (variável dependente). Os fatores propostos podem ser variáveis quantitativas ou qualitativas, enquanto a variável dependente deve ser quantitativa e observada dentro da classe dos fatores, ou seja, os tratamentos. Existem algumas considerações importantes, pressupostos, a serem feitas antes de aplicar a ANOVA: as amostras analisadas devem ser escolhidas de maneira aleatória e independentes; as populações têm distribuição normal; homogeneidade das variâncias entre os grupos.

As hipóteses da ANOVA são:

H_0 : As médias dos k tratamentos são iguais; H_1 : pelo menos duas das médias dos k tratamentos é diferente das demais. Quando os resultados da Análise de Variância indica a rejeição de H_0 , pode concluir que o fator considerado tem influência sobre a variável estudada.

De acordo com [9], o teste ANOVA é aplicado para testar se vários grupos de observações

provêm de distribuições com a mesma média; na análise de variância calcula-se uma estatística F ; se a hipótese nula é verdadeira e as médias de todos os grupos são realmente iguais, então a estatística F terá distribuição F e; se a estatística F calculada é superior ao valor crítico, então se rejeita a hipótese nula, de que as médias dos grupos sejam as mesmas.

4. RESULTADOS E DISCUSSÕES

Os dados do presente estudo referem-se à média do IPCA-15, obtida nos 3 anos considerados (Janeiro 2014 a Janeiro 2016) em cada cidade, para cada categoria. Primeiramente foi realizado um estudo descritivo dos dados 4.1.

Tabela 4.1: Resumo descritivo dos dados referente à média da inflação no período (de Janeiro 2014 a janeiro 2016) em cada cidade.

Cidades	Média	Variância
Belém - PA	3,68	4,52
Fortaleza - CE	3,91	7,99
Recife - PE	3,95	4,51
Salvador - BA	4,05	5,92
Belo Horizonte - MG	4,15	6,14
Rio de Janeiro - RJ	4,18	7,79
São Pulo - SP	4,2	8,49
Curitiba - PR	4,66	9,6
Porto Alegre - RS	4,36	8,46

Analisando os dados da Tabela observou-se que a cidade de Curitiba apresentou a maior dispersão dos dados (variância = 9,60), indicando que o valor médio de inflação de uma categoria para outra varia muito. Além disso, apresentou o maior valor médio de inflação para as categorias Alimentação e Bebidas, Habitação, Vestuário, Saúde e Cuidados pessoais. Isso justifica que, em média, esta cidade se destacou das demais atingindo o maior valor (4,66).

A cidade de Porto Alegre apresentou a segunda maior dispersão dos dados (variância = 8,46), indicando que o valor médio de inflação de uma categoria para outra varia muito. Apresentou o maior valor médio de inflação para a categoria Artigos de Residência. Isso justifica que, em média, esta cidade é a segunda que mais se destacou das demais atingindo o valor médio de (4,36).

A cidade do Rio de Janeiro apresentou maior média de inflação para Transporte(4,64) e Educação (7,52) que as demais cidades. Já a cidade de Fortaleza apresentou maior média de inflação para Despesas Pessoais que as demais cidades (6,12). A cidade de Recife apresentou maior média de inflação para Comunicação (0,97) que as demais cidades.

Tabela 4.2: Resumo descritivo dos dados referente à média da inflação no período de (Janeiro 2014 a janeiro 2016) em cada cidade, para cada categoria.

Estatísticas	Alimentação e bebidas	Habitação	Artigos de residência	Vestuário	Transportes
Mínimo	4,08	6,47	2,27	-0,55	2,72
1º Quartil	5,12	7,32	2,89	0,71	2,77
Média	5,36	7,96	3,18	0,98	3,47
Mediana	5,45	7,68	3,19	1,04	3,33
3º Quartil	5,70	8,57	3,46	1,54	4,11
Máximo	6,19	10,43	4,24	1,95	4,64

Estatísticas	Saúde e cuidados especiais	Despesas pessoais	Educação	Comunicação
Mínimo	3,97	4,26	5,29	-1,05
1º Quartil	4,04	5,04	5,91	-0,87
Média	4,35	5,33	6,61	-0,10
Mediana	4,29	5,24	6,86	-0,02
3º Quartil	4,52	5,78	7,36	0,49
Máximo	5,08	6,12	7,52	0,97

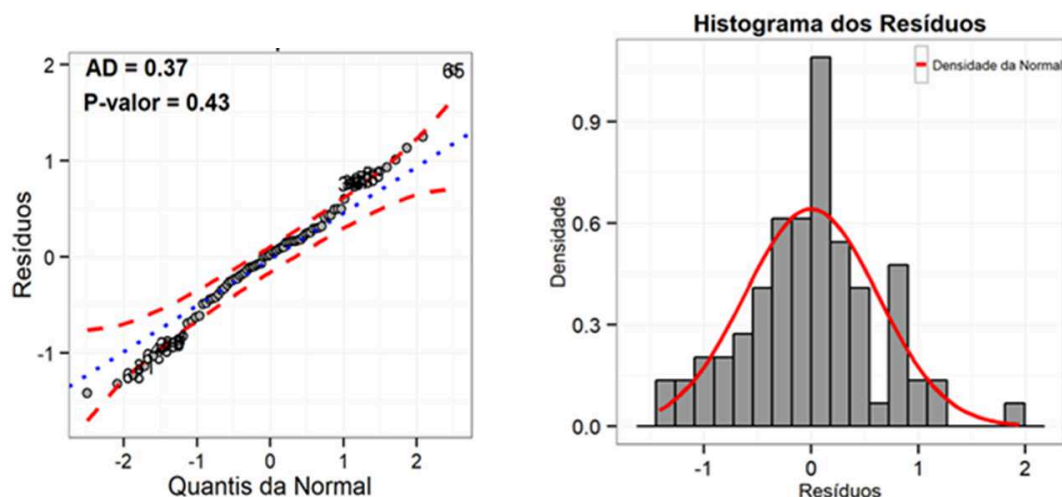
À seguir foi aplicado a ANOVA (Tabela 4.3) no delineamento em blocos casualizados (DBC), sendo que os meses são as repetições (igual a 25), um tratamento (cidades, com nove níveis) e os blocos são as categorias de consumo (nove níveis) de modo que em cada bloco encontram-se todos os tratamentos. Os dados do presente estudo foram computados de forma que o registro de cada mês é a média de todos os 25 meses entre janeiro de 2014 e janeiro de 2016.

Tabela 4.3: ANOVA referente à média da inflação no período de (Janeiro 2014 a Janeiro 2016) em cada cidade, para cada categoria

Fator de Variação	Graus de liberdade	Soma de quadrados	Quadrado médio	Estatística F	p-valor
Cidades	8	5,6474	0,7059	1,465	0,1879
Categorias de Consumo	8	476,635	59,5793	123,6	0,000000
Resíduos	64	30,8491	0,4820		
Total	80	513,131			

Aplicando o Teste F da ANOVA, constatou-se que ao nível de 5% e 1% de significância aceita-se a Hipótese nula para as cidades, portanto as médias entre as cidades são iguais, não diferem entre si; enquanto que para as categorias de consumo rejeita-se a hipótese nula, as médias entre as categorias de consumo diferiram para todas as categorias.

Os resíduos são normalmente distribuídos e independentes, como pode ser observado pelos gráficos QQ-Plot, Histograma e plotagem dos resíduos 4.1(a) 4.2.



(a) Gráfico da Probabilidade Normal dos resíduos

(b) Gráfico Histograma dos resíduos.

Figura 4.1: Normalidade dos resíduos

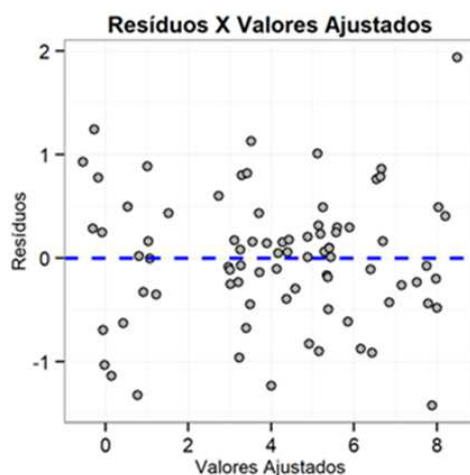


Figura 4.2: Gráfico dos Resíduos versus os Valores Ajustados

Foi realizada a Análise de Agrupamentos, sendo utilizado o Método hierárquico UPGMA uma vez que o Coeficiente de Correlação Cofenética neste método foi maior que nos demais ($CCC = 79,32\%$), indicando, portanto, boa consistência dos agrupamentos e pouca distorção após a formação dos grupos das regiões metropolitanas.

No dendrograma resultante da Análise de Agrupamentos (Figura 3.1), observa-se que, a subdivisão das cidades dentro dos grupos manteve a classificação em relação à média obtida na análise descritiva (Tabelas 4.1 e 4.2).

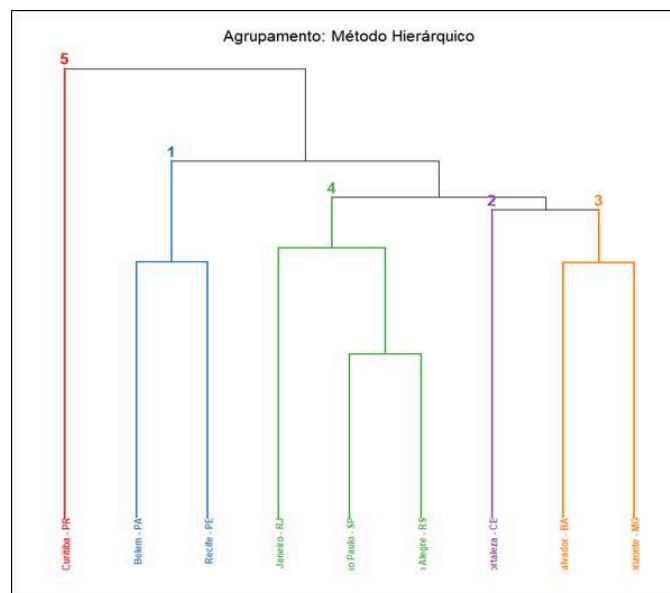


Figura 4.3: Dendrograma

Analisando o dendrograma, que também é descrito na Tabela (4.3) constatou-se que os grupos formados indicam que essas cidades possuem a média da inflação semelhantes entre si.

Tabela 4.4: Composição dos agrupamentos formados pela análise de cluster

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
Belém-PA Recife-PE	Fortaleza-CE	Salvador-BA Belo Horizonte-MG	Rio de Janeiro-RJ São Paulo-SP Porto Alegre-RS	Curitiba-PR

Após realizar um paralelo entre o IPCA-15 e o empreendedorismo no Brasil certificou-se, conforme ilustrado no gráfico de linha Figura (4.4), que a taxa de empreendedorismo inicial, seja por oportunidade ou por necessidade, entre os anos de 2014 e 2015 teve uma queda da taxa de empreendedores por oportunidade e, esta, permaneceu constante no ano de 2016. Já a taxa de novos empreendedores por necessidade aumentou entre os anos de 2014 e 2015, prosseguindo constante no ano de 2016.

O fato dessas taxas terem se mantido constantes em 2016 pode ser explicado pela informação relacionada à desaceleração da inflação em 2016, medida pelo IPCA. De acordo com [17], coordenadora de índices de preços do IBGE, a trajetória da inflação em 2016 foi de desaceleração. O fenômeno ajudou tanto na trégua dos alimentos quanto no arrefecimento da demanda, segundo a coordenadora.

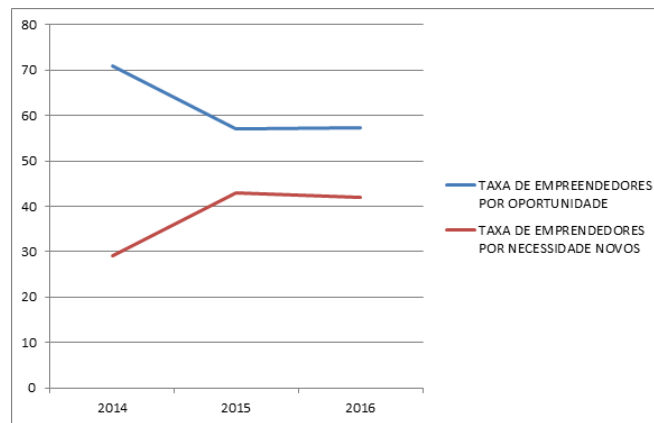


Figura 4.4: Motivação de empreendedores brasileiros para novos negócios

Por meio do teste t-Student, aplicado sobre os segmentos/atividades que mais se destacaram como novos empreendimentos, no período (2014-2016), buscou-se analisar se a média (para essas atividades) dos empreendedores por oportunidade e por necessidade diferiram estatisticamente entre si (Tabela 4.5).

Tabela 4.5: Análise das médias dos empreendedores de novos negócios por oportunidade e por necessidade para as atividades/segmentos que mais se destacaram entre 2014-2016.

Atividades/Segmentos	Por Oportunidade	Por Necessidade	p-valor (bilateral)
ALIMENTÍCIO			0,1003 ns
Média	17,4333	7,8	
Variância	14,3633	47,07	
SERVIÇOS DOMÉSTICOS			0.0039 **
Média	4,8667	16,3	
Variância	0,4433	10,47	
SERVIÇOS ESTÉTICOS			0,5276 ns
Média	8	8,8333	
Variância	1,24	3,1233	
COSMÉTICOS E HIGIENE PESSOAL			0,7104 ns
Média	4,6333	3,8	
Variância	0,3033	11,08	
VESTUÁRIO E ACESSÓRIOS			0,6568 ns
Média	6,6667	5,9333	
Variância	4,3033	2,7233	

Fonte: dados da pesquisa

** $p < 0,01$ e ns – não significativo pelo teste t-Student.

Com a aplicação do teste t-Student à todas atividades/segmentos que mais se destacaram na abertura de novos negócios no período analisado, quais sejam, serviços alimentícios; serviços domésticos; serviços estéticos; cosméticos e higiene pessoal; e vestuário e acessórios, nota-se que foi possível a identificação de variáveis que possibilitaram a verificação das diferenças entre o empreendedorismo por oportunidade e por necessidade.

Por meio do teste, verificou-se que apenas a atividade/segmento de Serviços Domésticos

demostrou p-valor significativo (0,0039), ao nível de 1% de significância. Isso quer dizer que a diferença de média (11,43%) foi muito significativa somente para a atividade/segmento de Serviços Domésticos, uma vez que a média da taxa para esse empreendimento por necessidade (16,3%) foi muito maior à motivação de empreender por oportunidade (4,87%) conforme GEM (2014, 2015, 2016). Nas outras atividades/segmentos não foram detectadas diferenças significativas entre as médias por oportunidade e por necessidade.

Esse resultado expõe o reflexo do cenário econômico dos últimos anos, onde o trabalhador brasileiro perde o emprego formal e empreende por uma necessidade de garantir sua sobrevivência, e de sua família, aumenta a parcela de indivíduos que encontram, muitas vezes, no mercado informal, a razão para continuar gerando emprego e renda [13].

5. CONCLUSÕES

A partir das cidades formou-se o dendograma pelo método da ligação média, a partir do coeficiente de correlação co-fenética obteve-se a confirmação do método selecionado, que validou a formação dos agrupamentos.

As cidades que possuem valor médio de inflação semelhantes entre si são: Belém e Recife; Salvador e Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre. As cidades de Fortaleza e Curitiba não possuem valor médio de inflação semelhante a nenhuma outra cidade.

A cidade de Curitiba se destacou das demais, atingindo o maior valor médio de inflação de (4,66); não sendo similar a nenhuma outra cidade. A cidade de Porto Alegre é a segunda que mais se destacou, atingindo o valor médio de inflação de (4,36).

O teste F da ANOVA evidencia que não rejeita a hipótese nula para as cidades, portanto as médias entre as cidades são iguais, porém para as categorias de consumo rejeita-se a hipótese nula, as médias entre as categorias são diferentes.

O empreendedorismo no período de 2014 a 2016 se destacou no setor alimentício por oportunidade com média de 17,43; e no setor de Serviços Domésticos por necessidade com média de 16,3. Notou-se um aumento considerável do empreendedorismo por necessidade, principalmente entre o ano de 2014 e 2015. No ano de 2016 houve certa estabilidade entre as duas motivações para novos negócios, seja por oportunidade, seja por necessidade.

O teste t-Student mostrou que somente a atividade/segmento de Serviços Domésticos teve diferença significativa entre as médias por oportunidade e por necessidade, confirmando ser o segmento onde o brasileiro mais empreendeu no período de 2014 a 2016.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ANDERBERG, M.R.: *Cluster analysis for applications*. Rel. Téc., Office of the Assistant for Study Support Kirtland AFB N MEX, 1973.
- [2] BARRETO, L.P.: *Educação para o empreendedorismo*. Salvador: Escola de Administração de Empresas da Universidade Católica de Salvador, 1998.
- [3] BUSSAB, W.O.; MIAZAKI, E.A.D.: *Introdução à análise de agrupamentos*. 9. Simpósio Brasileiro de Probabilidade e Estatística, 9(1):211, 1990. <https://www.bdpa.cnpia.embrapa.br/consulta/busca?b=pc&id=1284&biblioteca=vazio&busca=assunto:Congresso&qFacets=assunto:Congresso&sort=&paginacao=t&paginaAtual=48>.
- [4] CHIAVENATO, I.: *Empreendedorismo: dando asas ao espírito empreendedor*. Editora Manole, 2004.
- [5] CRUZ, C.D.; REGAZZI, A.J.: *Métodos biométricos aplicados ao melhoramento genético*. Viçosa, MG: Universidade Federal de Viçosa, 1994.
- [6] DOLABELA, F.: *O segredo de Luísa*. Cultura Editores, 2006.
- [7] DORNELAS, J.C.: *Empreendedorismo corporativo: como ser empreendedor, inovar e se diferenciar na sua empresa*. Empreendedorismo: transformando idéias em negócios, 12:166, 2008.
- [8] DORNELAS, J.C.A.: *Empreendedorismo: transformando ideias em negócio*. Campus, 2001.
- [9] DOWNING, D.; CLARK, J.: *Estatística Aplicada*. Saraiva, 1ª ed., 2000.
- [10] EVERITT, B. *et al.*: *Análise de agrupamentos*. Atlas, 5ª ed., 2011.
- [11] FILION, L.J.: *Empreendedorismo: empreendedores e proprietários-gerentes de pequenos negócios*. Revista de administração, 34(2):5–28, 1999.
- [12] FREI, F.: *Introdução à análise de agrupamentos*. Unesp, 2006.
- [13] GEM: *Global entrepreneurship monitor*. <http://www.gembrasil.org.br/>, acessado em 07/03/2009.

- [14] HAIR, J.F. *et al.*: *Análise multivariada de dados*. Bookman Editora, 5ª ed., 2009.
- [15] IBGE: *IBGE*. <http://www.ibge.com.br/>, acessado em 21/05/2009.
- [16] IBGE: *IPCA-15, Índice Nacional de preços ao consumidor amplo*, 2017. <https://www.ibge.gov.br/estatisticas-novoportal/economicas/precos-e-custgoos/9260-indice-nacional-de-precos-ao-consumidor-amplo-15.html?edicao=15910&t=conceitos-e-metodos>, acessado em 25/03/2018.
- [17] IBGE: *Índice Nacional de Preços ao Consumidor - IPCA-15*, 2017. <https://ww2.ibge.gov.br/home/estatistica/indicadores/precos/ipca15/defaultipca15.shtm>, acessado em 28/10/2017.
- [18] JOHNSON, Richard A. e WICHERN, D.W.: *Applied multivariate statistical analysis*. Prentice-Hall of India Private Limited, 2ª ed., 1982.
- [19] LANDIM, P. M. B.: *Análise Estatística de Dados Geológicos Multivariados*. Departamento de Geologia Aplicada. Laboratório de Geomatemática. UNESP/Campus de Rio Claro, texto didático 03 ed., 2000.
- [20] LOURENÇO, Alexandre ; MATIAS, R.P.: *Estatística multivariada*. Instituto Superior de Engenharia do Porto, 18(4):621–638, 2000. <https://www.revistas.usp.br/rege/article/download/49808/53914>.
- [21] MAROCO, J.: *Análise estatística com utilização do SPSS*. 2007.
- [22] Martins, E.: *Introdução aos índices*, 2017. <https://martinserica209.wixsite.com/investidoraprendiz/indices>, acessado em 28/10/2017.
- [23] MATAREZIO, R.A.L. ; PEREIRA, C.: *Estratégias de Pricing Utilizadas pelas Micro e Pequenas Empresas que São Auxiliadas pelo CIETEC - Centro Incubador de Empresas Tecnológicas*. Revista Controversa, 2006.
- [24] MILAGRE, R.: *Estatística: uma proposta de ensino para os cursos de administração de empresas*. Dissertação de Mestrado em Engenharia de Produção, 2001.
- [25] MINGOTI, S.A.: *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Editora UFMG, 2007.
- [26] MIOT, Hélio Amante ; BRIANEZI, G.: *Análise morfométrica do colágeno dérmico a partir da segmentação por conglomerados (clusters) de cor*. Anais Brasileiros de Dermatologia, 85(3):361–364, 2010.
- [27] NETO, A.; GIRALDI, J.: *Análise do posicionamento de indústrias calçadista na região de Franca*. Rev. Adm. Empresa, São Paulo, v. 48. n. 1 ed., 2008.

- [28] NETO, J.M. Moita, M. G.: *Uma introdução à análise exploratória de dados multivariados*. Atlas, 5ª ed., 1998.
- [29] ÂNGELO, E.B.: *Empreendedor corporativo: a nova postura de quem faz a diferença*. Elsevier, 2003.
- [30] PELICIONI, L.A.; RESENDE, M.: *Metas de inflação, política monetária e investimento: um estudo para dezessete países*. Revista de Economia Política, 2009.
- [31] PEREIRA, J. C. R.: *Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde humanas e sociais*. Edusp, 1999.
- [32] PESTANA, Maria H. ; GAGEIRO, J. N.: *Análise de dados para ciências sociais: a complementaridade do SPSS*. 6, 2003.
- [33] PREARO, L. C.: *O uso de técnicas estatísticas multivariadas em dissertações e teses sobre o comportamento do consumidor: um estudo exploratório*. Tese de Mestrado, 2008.
- [34] QUINTAL, G.: *Análise de clusters aplicada ao Sucesso/Insucesso em Matemática*. Dissertação (Mestrado em Matemática para Ensino, 2006.
- [35] RESENDE, M. F. C.; LIMA, F.: *Metas de inflação e investimento: O caso do Brasil*. Revista da Sociedade Brasileira de Economia Política, 2007.
- [36] SEBRAE: *Serviço brasileiro de apoio às micro e pequenas empresas*. <http://www..com.br/br/home/index.asp/>, acessado em 29/05/2009.
- [37] SICSÚ, J. et al.: *Agenda Brasil: políticas econômicas para o crescimento com estabilidade de preços*. Manole, 2003.
- [38] SPRENT, P.: *Applied nonparametric statistical methods*. New York: Chapman Hall, 1993.
- [39] STEINER, M. T. A. et al.: *Uma metodologia para o reconhecimento de padrões multivariados com resposta dicotômica*. Tese de Doutorado, 1995.
- [40] VIEIRA, S.: *Bioestatística tópicos avançados*. Elsevier Brasil, 2011.