



**Universidade Federal de Uberlândia
Faculdade de Matemática**

Bacharelado em Estatística

**ANÁLISE DE COMPONENTES
PRINCIPAIS NO ESTUDO DO IPCA-15**

Luiz Matheus Barbosa

Uberlândia-MG

2018

Luiz Matheus Barbosa

**ANÁLISE DE COMPONENTES
PRINCIPAIS NO ESTUDO DO IPCA-15**

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: Prof^ª Dr^ª Priscila Neves 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

Prof^ª Dr^ª Priscila Neves Faria

Prof^ª Dr^ª Mirian Fernandes Carvalho Araújo

Prof^ª Dr^ª Nádia Giaretta Biase

**Uberlândia-MG
2018**

AGRADECIMENTOS

Agradeço primeiramente a Deus por me permitir concluir mais uma grande etapa e aos meus familiares e Grandes Amigos por todo o apoio durante a construção deste projeto. Agradeço também aos meus Professores, em especial às minhas orientadoras, Maria Imaculada, Nádia e Priscila, por compartilharem de seu vasto conhecimento para a expansão do meu conhecimento e contribuição em minha formação como aluno e profissional. Dedico este projeto ao meu eterno melhor amigo, Guilherme, que apesar de não estar mais entre nós, faz-se presente em todas as minhas orações, vitórias e aprendizados, guiando-me com sua luz e lembranças de tudo o que vivemos juntos, e que ainda vamos viver um dia. Muito Obrigado!

RESUMO

Estudos a respeito da inflação pretendem indicar o efeito médio sobre a economia de um fenômeno complexo: o aumento de preços dos diversos bens que a compõem. O IBGE é o órgão brasileiro responsável pela divulgação dos índices mensuradores da inflação no país, sendo o Índice de Preços ao Consumidor Amplo (IPCA-15) o indicador da variação do custo de vida das famílias com rendimento entre 1 a 40 salários mínimos. O presente estudo tem como objetivo analisar os dados do IPCA-15, por meio de análises estatísticas descritivas e análise multivariada, ao identificar mudanças na inflação brasileira nos últimos anos (2014 a 2016), e verificar quais as categorias de consumo estão mais associadas à maior parte da variação contida nos dados. Dentre alguns dos resultados, foi obtido que os três primeiros Componentes Principais explicam cerca de 75,3% da variação dos dados, verificando-se que Despesas Pessoais e Comunicação possuem a maior contribuição para o Componente Principal 1.

Palavras-chave: Inflação, Economia, Categorias de Consumo.

ABSTRACT

Studies related to inflation intend to indicate the expected effect on the economy of a complex phenomenon. The IBGE (Brazilian Institute of Geography and Statistics in English) is a Brazilian Organization that promotes statistical data about Inflation, and The Index Cost of the Broad Consumer (IPCA-15 in Portuguese) shows the life cost of families with an income between 1 and 40 minimum salaries. The aim of this study is to analyze data from IPCA-15, highlight changes in the Brazilian Inflation among 2014 and 2016 and verify which of the consumer categories are the most associated with the variation in data. We concluded that the first three Principal Components explain around 75,3% of data variation which point out that Personal Expenses and Communication have the highest contribution to the first Principal Component.

Keywords: Inflation, Economy, Consumer Categories.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	III
1 Introdução	1
2 Materiais e Métodos	3
2.1 Dados	3
2.2 Métodos	3
2.3 Análise de Componentes Principais	3
2.4 Análise de Agrupamento	5
2.4.1 Medida de Similaridade e Dissimilaridade	6
2.4.2 Número de Grupos	7
2.5 <i>Software</i>	7
3 Resultados	9
3.1 Análise Exploratória	9
3.2 Análise de Componentes Principais	21
3.3 Análise de Agrupamento	31
3.3.1 Janeiro a Dezembro de 2014	31
3.3.2 Janeiro a Dezembro de 2015	33
3.3.3 Janeiro a Agosto de 2016	34
4 Conclusões	37
Referências Bibliográficas	39

LISTA DE FIGURAS

3.1	Distribuição dos índices de IPCA de janeiro de 2014 a agosto de 2016.	9
3.2	Distribuição dos índices do IPCA-15 de janeiro de 2014 a agosto de 2016 para cada região analisada.	10
3.3	Gráfico de linhas (por categoria de consumo) dos valores do IPCA-15 de janeiro de 2014 a agosto de 2016.	11
3.4	Distribuição dos valores observados do IPCA-15 por categoria de consumo, de Janeiro de 2014 a Agosto de 2016.	13
3.5	IPCA-15 para a Região de Belém-PA de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.	14
3.6	IPCA-15 para a Região de Belo Horizonte-MG de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.	14
3.7	IPCA-15 para a Região de Curitiba-PR de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.	15
3.8	IPCA-15 para a Região de Fortaleza-CE de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo	16
3.9	IPCA-15 para a Região de Porto Alegre-RS de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.	17
3.10	IPCA-15 para a Região de Recife-PE de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.	17
3.11	IPCA-15 para a Região do Rio de Janeiro-RJ de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.	18
3.12	IPCA-15 para a Região de Salvador-BA de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.	19
3.13	IPCA-15 para a Região de São Paulo-SP de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.	20
3.14	Porcentagem da variação explicada pelos Componentes Principais.	23
3.15	Gráfico das Variáveis na Análise de Componentes Principais (ACP).	24
3.16	Gradiente de variáveis segundo à Qualidade de Representação	25
3.17	Contribuição das variáveis nas Dimensões 1 (esquerda) e 2 (direita).	27
3.18	Contribuição conjunta das variáveis para as Componentes 1 e 2.	27
3.19	Qualidade de Representação dos indivíduos do estudo.	28
3.20	Contribuição conjunta dos indivíduos para as Componentes 1 e 2.	29
3.21	Gráfico Biplot para as variáveis e indivíduos no estudo.	30
3.22	Dendrograma referente ao período de Janeiro a Dezembro de 2014 para as Regiões analisadas segundo o IPCA-15.	32
3.23	Dendrograma referente ao período de Janeiro a Dezembro de 2015 para as Regiões analisadas segundo o IPCA-15.	33
3.24	Dendrograma referente ao período de Janeiro a Agosto de 2016 para as Regiões analisadas segundo o IPCA-15.	35

LISTA DE TABELAS

3.1	Estatísticas Descritivas para os valores de IPCA-15 segundo à região estudada, de Janeiro de 2014 a Agosto de 2016.	10
3.2	Peso individual das Categorias de Consumo no cálculo final do IPCA.	11
3.3	Estatísticas Descritivas para os valores de IPCA-15 segundo à Categoria de Consumo, de Janeiro de 2014 a Agosto de 2016.	12
3.4	Autovalores e seus respectivos percentuais de variação para cada Dimensão (Componente).	22
3.5	Coordenadas das Variáveis dos Componentes Principais retidos no estudo.	23
3.6	Qualidade de Representação das variáveis.	25
3.7	Contribuição das Variáveis aos Componentes Principais (%).	26

1. INTRODUÇÃO

A inflação no Brasil sempre foi tema central nos debates da história da economia moderna do país. Do preço do café ao custo mensal da cesta básica, esse índice tornou-se parte dos assuntos cotidianos do brasileiro, que visa, em um país ainda em recessão, otimizar a economia perante aos gastos com alimentos e serviços básicos. Entende-se por inflação o aumento de bens e serviços que fazem com que a moeda se torne desvalorizada naquele momento, perdendo assim seu poder aquisitivo [6]. Concretamente, índices de inflação podem ser construídos a partir de variadas cestas de bens, conforme os preços relevantes para os grupos que se queira considerar: consumidores, produtores de um setor, ou da economia em geral, residentes de uma região específica, dentre outros. Os modelos macroeconômicos, usualmente, tratam a inflação como uma grandeza única, buscando quais são seus determinantes e as políticas mais adequadas para contê-la. Os artigos que analisam a inflação brasileira recente, e sob uma perspectiva desagregada ainda são poucos. Em geral, trabalham com a decomposição do Banco Central do Brasil (BCB) para o Índice Nacional de Preços ao Consumidor Amplo (IPCA) de acordo com o tipo de precificação em três grupos (bens comercializáveis, não comercializáveis e monitorados) ou em dois grupos (bens livres e monitorados)[21].

O IPCA é o índice responsável por mensurar a inflação de produtos e serviços comercializados no varejo de consumo pessoal das famílias para que possa ser identificada a modificação dos valores entre os meses analisados [10]. O Instituto Brasileiro de Geografia e Estatística (IBGE) é o responsável pela divulgação do índice, e para realização dos cálculos o instituto constrói a cesta de produtos e serviços consumidos pela população, cujo intuito é investigar os gastos das famílias em cada produto ou serviço comprado. A base da pesquisa é composta por famílias com ganhos de 1 a 40 salários mínimos, seja qual for sua fonte de rendimentos, e os dados são coletados entre os dias 01 a 30 do mês de referência e recolhidos em domicílios das regiões metropolitanas do Rio de Janeiro, Porto Alegre, Belo Horizonte, Recife, São Paulo, Belém, Fortaleza, Salvador e Curitiba.

Estudos realizados por [22], mostraram que a utilização da técnica multivariada passou a ser familiarizada devido à aplicação dos acadêmicos e profissionais de marketing em estudos que identificaram a preferência e satisfação do consumidor.

Assim, a partir do mês de maio de 2000, o IBGE, passou também a disponibilizar através da Internet o Índice Nacional de Preços ao Consumidor Amplo-15 (IPCA-15), que mede a variação dos custos dos gastos no período aproximado do dia 15 de um mês ao dia 15 do mês subsequente [10][11]. Este será o índice a ser avaliado no presente trabalho em relação aos anos de 2014,

2015 e o primeiro semestre de 2016. Desta maneira, a aplicação de técnicas multivariadas na análise da inflação de produtos e serviços consumidos pela população é de suma importância para extrair mais informações que possam contribuir para a economia do país.

2. MATERIAIS E MÉTODOS

2.1 DADOS

De maneira geral, o escopo metodológico do Índice Nacional de Preços ao Consumidor abrange os seguintes temas: montagem da estrutura geral de pesos; definição das bases cadastrais de produtos e locais e a coleta de preços e método 26 de cálculo.

Para a realização deste trabalho foram utilizados dados referentes ao Índice Nacional de Preços ao Consumidor Amplo-15 (IPCA-15) recolhidos em domicílios das regiões metropolitanas do Rio de Janeiro, Porto Alegre, Belo Horizonte, Recife, São Paulo, Belém, Fortaleza, Salvador e Curitiba, mensalmente contabilizados de janeiro de 2014 a agosto de 2016, disponibilizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), obtidos diretamente da página de Dados Abertos do Governo Federal (<http://dados.gov.br/tag/IBGE>), que são dados livremente disponíveis para todos utilizarem e redistribuírem como desejarem, sem restrição de licenças, patentes ou mecanismos de controle.

2.2 MÉTODOS

Primeiramente, foi realizada uma análise descritiva dos dados para uma melhor visualização de sua distribuição ao utilizar gráficos e tabelas com a finalidade de uma melhor sumarização de seu comportamento. Os tipos de gastos que compõem a estrutura do IPCA são os seguintes: gastos com alimentação, transporte, comunicação, despesas pessoais, vestuário, habitação, saúde e cuidados pessoais, além de artigos de residência [4].

Em seguida, foram utilizadas as técnicas multivariadas de Análise de Componentes Principais e a técnica da Análise de Agrupamento, com o objetivo de verificar possíveis semelhanças entre o comportamento das variáveis analisadas, além de examinar as correlações entre as variáveis sob estudo, bem como avaliar a importância de cada uma e promover a eliminação daquelas que contribuem de maneira relativamente inferior, em termos de variação, no grupo de indivíduos avaliados.

2.3 ANÁLISE DE COMPONENTES PRINCIPAIS

Cada Componente Principal é uma combinação linear das variáveis originais, independentes entre si e estimadas com o propósito de reter, em ordem de estimação, o máximo da informação,

em termos de variação total contida nos dados originais. Entre os Componentes Principais, o primeiro possui a maior variância, o segundo possui a segunda maior, e assim sucessivamente. Este procedimento permite construir índices que possibilitem o agrupamento de indivíduos com o mais alto grau de similaridade [5]. A Análise de Componentes Principais é uma técnica de avaliação de interdependência, ou seja, estuda as relações de um conjunto de variáveis entre si [16]. Este método procura explicar a estrutura de variâncias-covariâncias através de combinações lineares das variáveis originais com a finalidade de proporcionar uma simplificação estrutural dos dados de modo que a diversidade, influenciada a princípio por um conjunto p -dimensional, possa ser avaliada por um complexo bi ou tridimensional de fácil interpretação geométrica [18].

Segundo [23], o primeiro componente principal (Y_1) de um conjunto de p variáveis, X_1, X_2, \dots, X_p , contidas no vetor $X^\top = (X_1, X_2, \dots, X_p)$ é definido como a combinação linear:

$$Y_1 = b_{11}X_1 + b_{21}X_2 + \dots + b_{p1}X_p = \mathbf{b}'_1 \mathbf{X}$$

cujos coeficientes b_{i1} são elementos do vetor característico b_1 , associado à maior raiz característica (λ_1) da matriz de covariância amostral, \mathbf{S} , das variáveis X_i 's. Os autovalores (ou raízes características) ordenados, ou seja, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, são as variâncias amostrais dos componentes principais.

A técnica envolve a matriz de covariância amostral \mathbf{S} ou a matriz de correlação \mathbf{R} , uma vez que distinta das aplicações envolve esta última, pois, frequentemente, as variáveis têm escalas, apresentando então uma necessidade de padronização. O grau de influência que cada variável X_j tem sobre a componente Y_i é determinado por sua correlação, definida por:

$$\text{Corr}(X_j, Y_i) = R_{X_j, Y_i} = b_{1j} \cdot \frac{\sqrt{\widehat{\text{Var}}(Y_i)}}{\sqrt{\widehat{\text{Var}}(X_j)}} = \sqrt{\lambda_1} \cdot \frac{b_{1j}}{\sqrt{\widehat{\text{Var}}(X_j)}}$$

As variáveis padronizadas, Z_1, \dots, Z_p , são expressas por $Z_j = (X_j - \mu_j) / \sqrt{(\sigma_{jj})}$, com $j = 1, \dots, p$. A contribuição do j -ésimo componente na explicação da variação total pode ser expressa por

$$\text{Contr}Y_j = \frac{\text{Var}(Y_j)}{\sum_{j=1}^p \text{Var}(Y_j)} = \frac{\lambda_j}{\sum_{j=1}^p \lambda_j}$$

Para a determinação do número de Componentes Principais a serem utilizados na análise, pode-se utilizar diversos critérios de escolha, sendo os principais: o Critério de Kaiser, o Diagrama dos Autovalores e o critério da porcentagem da variância acumulada, salientando que a escolha de um critério específico pode levar a diferentes resultados, constituindo-se então de uma decisão subjetiva e explanatória a critério do pesquisador [26]. O Critério de Kaiser determina que o número de componentes deve ser igual ao número de autovalores maiores ou iguais à média das variâncias das variáveis analisadas. Quando a análise é feita sobre a matriz de correlação (variáveis padronizadas), esse critério corresponde à exclusão de componentes com autovalores inferiores a um, segundo [2].

O Diagrama de Autovalores é representado por meio gráfico com o objetivo de relacionar a ordem de magnitude do autovalor com seu respectivo valor numérico. Para [9], a observação do diagrama de autovalores permite conservar aqueles autovalores situados acima do ponto de ruptura da queda da curva da função que relaciona a ordem e os autovalores.

O Critério da porcentagem da variância acumulada é um critério empírico que considera manter os componentes significativos, em que o número dos mesmos é estipulado de modo que o conjunto de componentes explique uma porcentagem pré-definida da variabilidade global, a um nível de explicação de pelo menos 70% da variabilidade total dos dados [2].

A fim de transformar os coeficientes dos componentes principais em uma estrutura simplificada e de fácil interpretação, pode-se aplicar o método de rotação de fatores ortogonais. Dentre esses métodos, o critério Varimax tem como base reduzir o número de variáveis que tem altas cargas em um fator, cuja finalidade é maximizar a variação entre os pesos de cada Componente Principal [8].

2.4 ANÁLISE DE AGRUPAMENTO

A Análise de Agrupamento ou *Cluster Analysis* é uma variedade de técnicas e algoritmos, cujo objetivo é encontrar e separar objetos em grupos similares. O objetivo da análise de agrupamento pretende, dada uma amostra de n objetos (ou indivíduos), cada um deles medidos segundo p variáveis, procurar um esquema de classificação que agrupe os objetos em g grupos, exigindo-se daí conceitos científicos mais sofisticados de semelhança. Devem ser determinados também o número e as características desses grupos [3].

O princípio da análise de agrupamento consiste em que cada observação de uma amostra multivariada possa ser considerada como um ponto em um espaço euclidiano multidimensional [17].

Os processos de classificação objetivam agrupar esses pontos em conjuntos que evidenciem aspectos marcantes da amostra. Assim, dado um conjunto de observações conhecidas somente por uma listagem de suas características, objetiva-se encontrar a melhor maneira de descrever seus padrões de similaridade mútuos.

Em termos gerais, a classificação pode ser entendida como um processo para localizar entidades em classes inicialmente indefinidas, de modo que os indivíduos da mesma classe sejam similares entre si em algum sentido definido pelas variáveis consideradas. Essas classes de indivíduos similares serão os agrupamentos.

Os diversos métodos de agrupamento objetivam transformar um conjunto heterogêneo de unidades não separadas, a princípio, em grupos que reflitam aspectos considerados importantes das relações originais entre as mesmas unidades. Para isso são propostas muitas técnicas, não se estruturando ainda uma teoria geral e amplamente aceita.

A Análise de Agrupamento é interessante, principalmente sob o aspecto descritivo, pois seu resultado final é um gráfico de esquemas hierárquicos denominado dendrograma. Este representa uma síntese dos resultados, o que ocasiona certa perda de informações. Ainda

assim, e no caso dessa perda ser pequena, o resumo da informação torna-se mais fácil de ser manipulado e armazenado, sendo importante para a comparação, classificação e discussão do material estudado.

Após a escolha das variáveis que serão utilizadas como critério de semelhança, uma das questões vitais das técnicas de Análise de Agrupamento é a definição do coeficiente de similaridade ou dissimilaridade.

2.4.1 MEDIDA DE SIMILARIDADE E DISSIMILARIDADE

Um conceito fundamental na utilização das técnicas de Análise de Agrupamento é a escolha de um critério que meça a distância entre dois objetos, ou que quantifique o quanto eles são parecidos. Esta medida é chamada de coeficiente de parença. Cabe observar que, tecnicamente, é possível dividir esta medida em duas categorias: medidas de similaridade e de dissimilaridade.

Na medida de similaridade, quanto maiores os valores observados, mais parecidos serão os objetos. Já para a medida de dissimilaridade, quanto maiores os valores observados, menos parecidos (mais dissimilares) serão os objetos. O coeficiente de correlação é um exemplo de medida de similaridade, enquanto que a distância euclidiana é um exemplo de dissimilaridade.

A maioria dos algoritmos de Análise de Agrupamento estão programados para operarem com o conceito de distância (dissimilaridade), exigindo do usuário o esforço da transformação.

DISTÂNCIA EUCLIDIANA

Considere o vetor x de coordenadas reais (x_1, x_2, \dots, x_p) como descritor dos objetos que investigarão tais semelhanças. A medida mais conhecida para indicar a proximidade entre os objetos A e B é a distância euclidiana $d(A, B)$:

$$d(A, B) = \left[\sum_{i=1}^p (x_i(A) - x_i(B))^2 \right]^{\frac{1}{2}}$$

ou em linguagem matricial:

$$d(A, B) = [(\mathbf{x}(A) - \mathbf{x}(B))'(\mathbf{x}(A) - \mathbf{x}(B))]^{\frac{1}{2}}$$

DISTÂNCIA EUCLIDIANA PADRONIZADA

Quando se trabalha com variáveis quantitativas, a distância euclidiana comumente soma distâncias não comparáveis, como cm, kg, anos, milhões, etc., muito embora, a mudança de uma das unidades possa alterar completamente o significado e o valor do coeficiente. Essa é uma das razões da padronização das variáveis dos elementos x_1, x_2, \dots, x_p do vetor x . Assim o uso da transformação

$$z_i = \frac{x_i(\cdot) - \bar{x}_i}{s_i}$$

em que \bar{x}_i e s_i , indicam respectivamente a média e o desvio padrão de i -ésima coordenada, é um dos modos para evitar essa inconveniência. Feita a transformação, a distância euclidiana passa a ser:

$$d(A, B) = \left[\sum_{i=1}^p (z_i(A) - z_i(B))^2 \right]^{\frac{1}{2}}$$

que é a soma dos desvios padronizados. Pode-se verificar que a expressão acima pode ser escrita da seguinte forma, em notação vetorial:

$$d(A, B) = [(\mathbf{x}(A) - \mathbf{x}(B))\mathbf{D}^{-1}(\mathbf{x}(A) - \mathbf{x}(B))]^{\frac{1}{2}}$$

em que \mathbf{D} é uma matriz diagonal, tendo como i -ésimo componente a variância s_i^2 , isto é, $\mathbf{D} = \text{diag}(s_1^2, s_2^2, \dots, s_p^2)$.

Após a escolha das variáveis que serão usadas como critérios de semelhança, uma das questões vitais das técnicas de análise de agrupamento, é a definição do coeficiente de similaridade ou dissimilaridade.

2.4.2 NÚMERO DE GRUPOS

O algoritmo euclidiano de agrupamento e o método de formação de grupos produzem conjuntos que constituem uma proposição sobre a organização básica e desconhecida dos dados. Entretanto todo esse procedimento esbarra em uma dificuldade comum que é a determinação do número ideal de grupos.

Para auxiliar na decisão do número de grupos, pode-se adotar, segundo [3], a técnica hierárquica que consiste em examinar o dendrograma em busca de grandes alterações nos níveis de similaridade para as sucessivas fusões.

2.5 Software

Todas as análises foram realizadas utilizando-se o software R [28] e a interface R-Studio [27]. Para a Análise Exploratória, foram empregados os pacotes **reshape** [30], **ggplot2** [31] e **reshape** [32].

Na Análise de Componentes Principais e de Agrupamentos, os pacotes **FactoMineR** [19], **factoextra** [15], **psych** [25] e **corrplot** [29] também foram implementados.

3. RESULTADOS

3.1 ANÁLISE EXPLORATÓRIA

POR REGIÃO

Para verificar o comportamento dos índices de IPCA referentes à região de estudo no período analisado entre Janeiro de 2014 e Agosto de 2016, primeiramente foi gerado um gráfico de linhas, contrastando o valor do IPCA e sua respectiva região, ilustrado pela Figura 3.1.

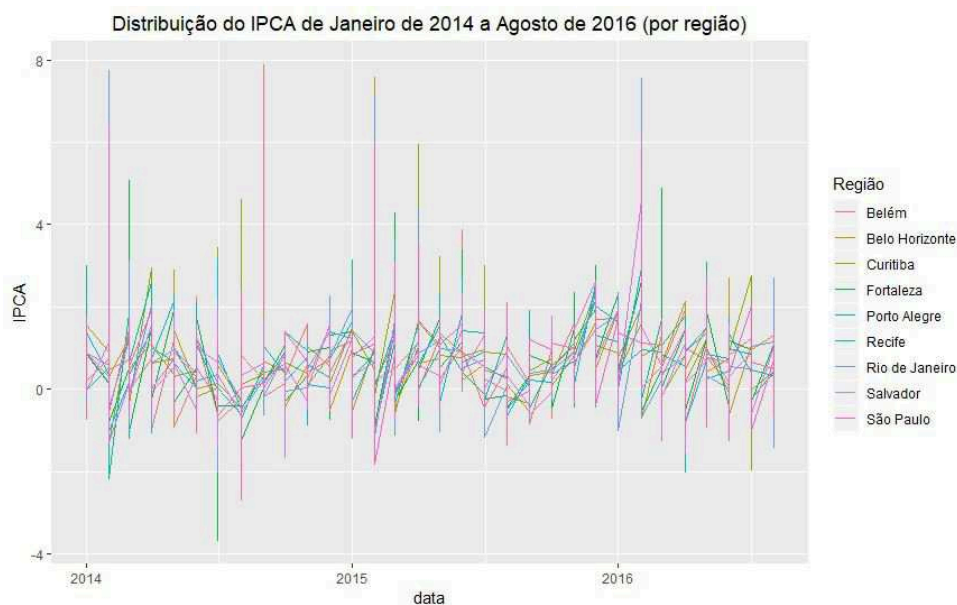


Figura 3.1: Distribuição dos índices de IPCA de janeiro de 2014 a agosto de 2016.

Através da Figura 3.1, é possível verificar a existência de grandes variações ao longo do tempo para algumas cidades. Porém, como a maioria de seus valores encontram-se centralizados em torno de zero, há o indício de não existirem grandes crescimentos e/ou recuos no valor do IPCA no período estudado.

Para realizar uma análise descritiva a respeito destes dados, a Tabela 3.1 apresenta os resumos dos valores médios, medianos e de variabilidade dos valores do IPCA para cada uma das cidades analisadas no período estudado, ordenados pelo valor médio.

De acordo com a Tabela 3.1, pode-se verificar que Porto Alegre apresentou o maior crescimento médio do valor do IPCA no período estudado, com um crescimento mensal médio de 0,63%. Em contrapartida, Recife apresentou o menor crescimento médio mensal, com uma

Tabela 3.1: Estatísticas Descritivas para os valores de IPCA-15 segundo à região estudada, de Janeiro de 2014 a Agosto de 2016.

Região	Média	Mediana	Desvio Padrão	Coef. de Variação	Mínimo	Máximo
Porto Alegre	0,63	0,48	0,89	140,70%	-2,17	5,63
Curitiba	0,62	0,48	1,05	170,46%	-2,50	6,75
Rio de Janeiro	0,62	0,47	1,08	172,91%	-2,01	7,75
Fortaleza	0,60	0,48	0,91	151,21%	-3,67	5,06
São Paulo	0,60	0,51	0,93	156,40%	-2,69	6,18
Belém	0,58	0,42	1,00	173,34%	-1,36	7,87
Belo Horizonte	0,58	0,48	0,93	159,84%	-1,12	7,58
Salvador	0,58	0,45	0,96	164,29%	-1,65	6,46
Recife	0,57	0,38	0,85	148,02%	-2,37	6,21

média de 0,57%. Ainda segundo a tabela, verifica-se que, no período analisado, as cidades de Belém e Rio de Janeiro apresentaram as maiores variabilidades, com um Coeficiente de Variação de 173% do valor de seu IPCA. Em compensação, Porto Alegre figurou como a cidade com a mais baixa variabilidade, com coeficiente de variação de 141%.

Adicionalmente, é possível analisar o comportamento individual do IPCA-15 por região através da Figura 3.2, que apresenta uma análise gráfica destes valores durante o período analisado, de janeiro de 2014 a Agosto de 2016, para cada uma das regiões analisadas. É possível verificar comportamentos sazonais na maioria das regiões analisadas, como Belo Horizonte, Salvador e Rio de Janeiro, que apresentam índices do IPCA muito altos no início de cada ano. Já em regiões como Belém e Curitiba, verifica-se a alta destes índices ao longo de todo o período analisado, caracterizada pelos picos constantes em seus gráficos.

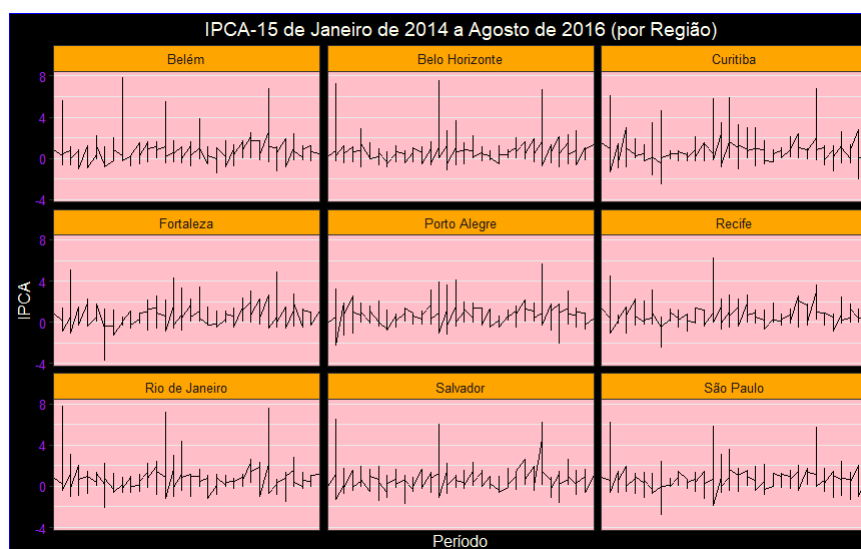


Figura 3.2: Distribuição dos índices do IPCA-15 de janeiro de 2014 a agosto de 2016 para cada região analisada.

POR CATEGORIA DE CONSUMO

Os valores do IPCA são calculados mensalmente, através de uma pesquisa de preços levantada pelo Instituto Brasileiro de Geografia e Estatística (IBGE)[10], que é realizada em

estabelecimentos comerciais, domicílios, com prestadoras de serviços e concessionárias de serviços públicos em uma coleta entre o primeiro e o último dia de cada mês, levando em consideração as categorias de alimentação e bebidas, artigos de residência, comunicação, despesas pessoais, educação, habitação, saúde e cuidados pessoais, transportes e vestuário.

Cada uma destas categorias possui um peso diferente no cálculo, que pode ser verificado na Tabela 3.2.

Tabela 3.2: Peso individual das Categorias de Consumo no cálculo final do IPCA.

Categoria de Consumo	Peso no cálculo(%)
Alimentação e Bebidas	23,12
Transporte	20,54
Habitação	14,62
Saúde e Cuidados Pessoais	11,09
Despesas Pessoais	9,94
Vestuário	6,67
Comunicação	4,96
Artigos de Residência	4,69
Educação	4,37

Para verificar o comportamento dos dados do período de Janeiro de 2014 a Agosto de 2016 segundo cada categoria de consumo analisada, implementou-se um gráfico de linhas, que pode ser visualizado na Figura 3.3.

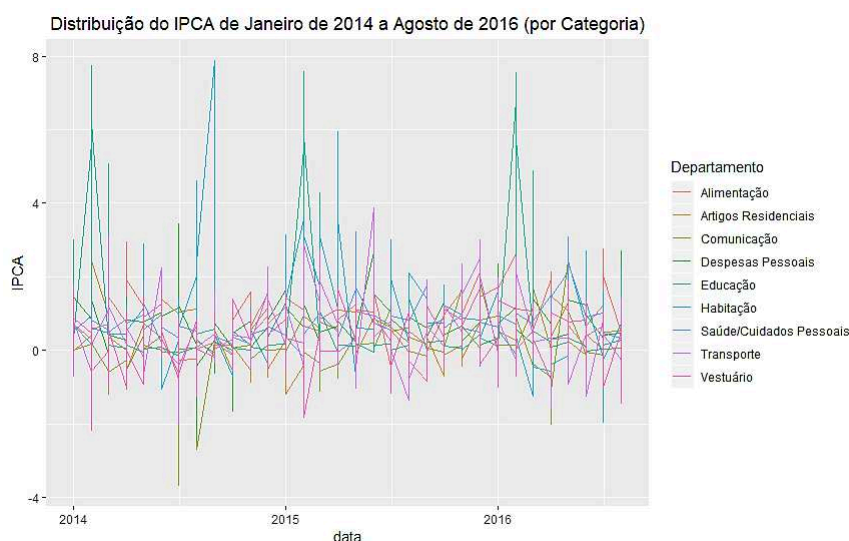


Figura 3.3: Gráfico de linhas (por categoria de consumo) dos valores do IPCA-15 de janeiro de 2014 a agosto de 2016.

Pode ser verificada uma alta variabilidade entre os índices do IPCA-15 segundo às categorias de consumo. Os setores de Habitação e Educação apresentam altos picos de variabilidade positivos, o que indica que em alguns períodos o valor do IPCA-15 apresentou um alto crescimento, impactando em um consequente aumento do custo de serviços para estas categorias.

Na Tabela 3.3, pode-se verificar as estatísticas descritivas referentes a cada uma das categorias de consumo analisadas, ordenadas pelo valor médio, em que o valor mediano e medidas

de variabilidade, tais como o desvio padrão, valor máximo e mínimo e o coeficiente de variação também são apresentados:

Tabela 3.3: Estatísticas Descritivas para os valores de IPCA-15 segundo à Categoria de Consumo, de Janeiro de 2014 a Agosto de 2016.

Categoria	Média	Mediana	Desvio Padrão	Coef. de Variação	Mínimo	Máximo
Alimentação	0,90	0,86	0,75	83,51%	-0,64	4,59
Habitação	0,86	0,64	1,10	128,49%	-1,95	7,87
Educação	0,79	0,20	1,71	217,48%	-0,26	7,75
Saúde	0,75	0,63	0,55	72,95%	-0,67	3,23
Despesas	0,74	0,68	0,74	100,01%	-2,50	3,44
Transporte	0,53	0,40	0,91	171,30%	-2,01	3,87
Artigos Residenciais	0,47	0,48	0,58	124,26%	-1,19	2,44
Vestuário	0,30	0,38	0,71	235,58%	-2,17	2,02
Comunicação	0,06	0,02	0,64	1078,39%	-3,67	2,42

Na Tabela apresentada, nota-se que o setor Alimentício foi o que apresentou o maior crescimento médio mensal no período analisado de Janeiro de 2014 a Agosto de 2016, de magnitude de 0,90%, sendo então a categoria de maior impacto em custos ao consumidor. Em contrapartida, o setor de Comunicação apresentou o mais baixo índice de crescimento médio mensal, de valor 0,06%, apresentando-se praticamente estável ao longo do período analisado. Apesar disto, este foi o setor que apresentou a mais alta variabilidade dentre os estudados, exibindo um coeficiente de variação de 1078%, com a mais alta queda verificada no estudo, de valor 3,67%, verificada no mês de Julho de 2014, na cidade de Fortaleza-CE. O setor de Saúde/Cuidados Pessoais apontou a menor variação dentre os setores analisados, porém alta, com o coeficiente de variação no valor de 73% e crescimento médio mensal de 0,75%. A maior alta verificada foi de 7,87%, no setor de Habitação, em Setembro de 2014 na cidade de Belém-PA, categoria que apresentou um dos maiores crescimentos médios mensais do valor do IPCA, na magnitude de 0,86%. Verifica-se que no departamento de Educação, valores do IPCA-15 positivos em cerca de 87% dos meses analisados, ilustrando assim um constante aumento nos preços de produtos relacionados a este setor. Dentre os períodos de queda analisados, o mais alto dá-se pela redução de 0,26%, na cidade de Curitiba, em 2014.

A Figura 3.4 ilustra o comportamento do IPCA-15 discriminado pelas categorias de consumo analisadas individualmente, no período de Janeiro de 2014 a Agosto de 2016.

Através da Figura 3.4, é viável analisar tendências e comportamentos sazonais acerca das categorias analisadas no período estudado. Verifica-se que os setores de Alimentação, Despesas Pessoais, Habitação, Saúde/Cuidados Pessoais e Educação apresentaram predominância de crescimento mensal do IPCA-15 no período analisado, com destaque para este último, que também apresentou comportamentos sazonais quanto ao grande crescimento do valor do IPCA nos meses iniciais do ano, caracterizados pelo retorno das aulas escolares, com impacto direto no aumento de custos do consumidor deste setor. O setor de Habitação apresentou um grande aumento no valor do IPCA-15 no segundo semestre de 2014, mas apresentou a partir do ano de 2015, uma diminuição contínua em seus índices, apresentando uma tendência de estabilização no ano de 2016. A categoria de Comunicações apresentou grandes picos de queda do valor do

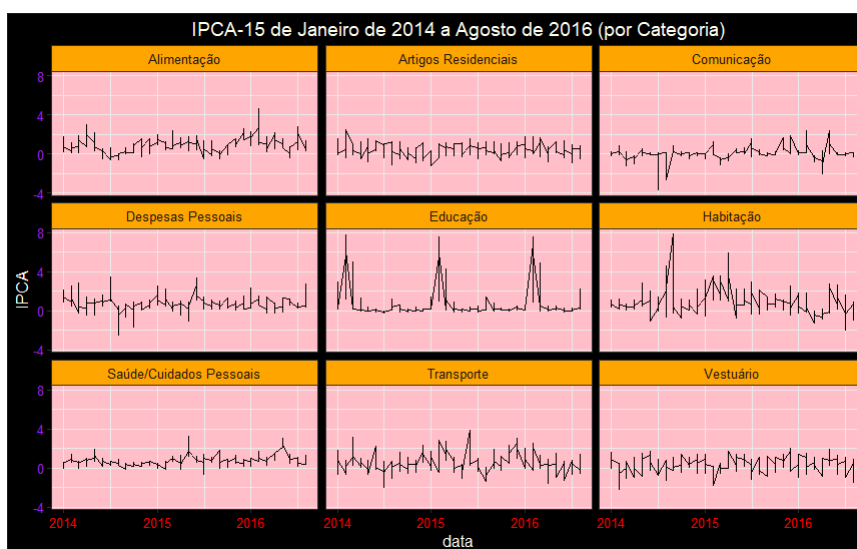


Figura 3.4: Distribuição dos valores observados do IPCA-15 por categoria de consumo, de Janeiro de 2014 a Agosto de 2016.

IPCA na metade de 2014 e uma estabilidade deste índice nos períodos futuros, salvo alguns crescimentos no primeiro semestre de 2016. Em geral, os demais setores apresentaram equilíbrio nos Índices Nacionais de Preço ao Consumidor Amplo no período do estudo.

REGIÃO vs CATEGORIAS DE CONSUMO

Nesta seção será analisado o comportamento do IPCA de Janeiro de 2014 a Agosto de 2016 de cada uma das regiões analisadas no estudo, em função das categorias de consumo.

BELÉM

A Figura 3.5 apresenta um resumo dos valores do IPCA-15 para a cidade de Belém segundo às categorias de consumo.

Pode ser verificado através da Figura 3.5, que a cidade de Belém possui uma distribuição de seus índices do IPCA-15 em torno do valor zero. Porém, é importante destacar um padrão de crescimento no valor do IPCA nos meses de fevereiro de cada ano na categoria de Educação, devido ao período de retornos escolares neste mês, que logo é recuado nos meses sequenciais. Em contrapartida, o mais alto índice registrado no período do estudo é verificado no setor de Habitação, em setembro de 2014 (7,87%), mas que sofre brusca queda no mesmo período do ano seguinte (1,38%).

BELO HORIZONTE

A Figura 3.6 apresenta o comportamento dos dados do IPCA-15 segundo às categorias de consumo para a Região de Belo Horizonte.

Semelhante à Belém, na Região de Belo Horizonte verifica-se um alto índice do valor do IPCA-15 para a categoria de Educação nos meses de fevereiro de 2014 (7,58%), 2015 (7,22%)

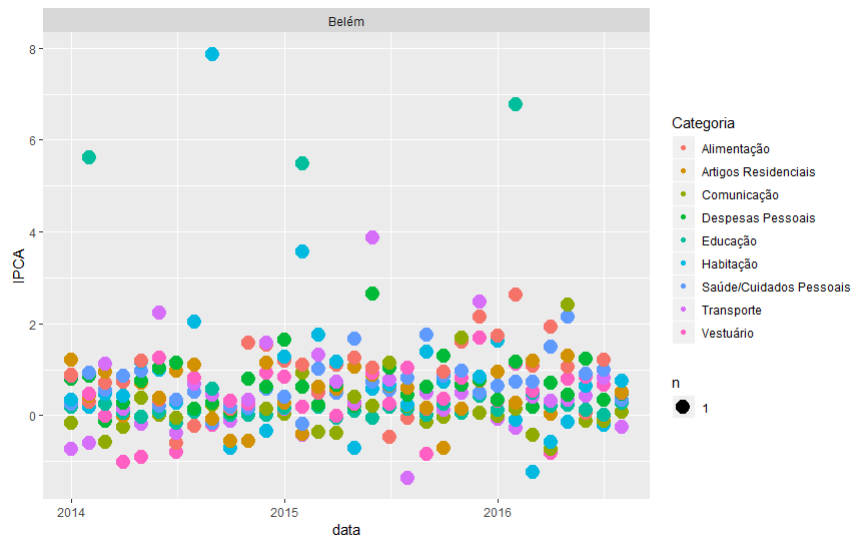


Figura 3.5: IPCA-15 para a Região de Belém-PA de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.

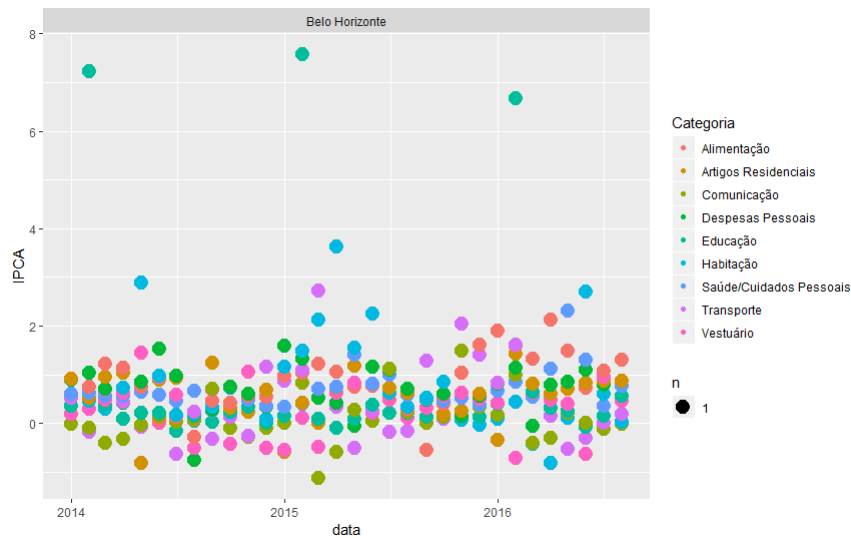


Figura 3.6: IPCA-15 para a Região de Belo Horizonte-MG de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.

e 2016 (6,67%), porém, ao analisar esses resultados, pode-se aferir uma queda de aproximadamente 12% do valor da inflação para este setor no período estudado, apesar desta categoria ter um peso menor do que 5% no cálculo final do valor do IPCA Geral. Em contraparte, o setor de Comunicação apresentou o menor índice do IPCA registrado durante o período estudado, com uma deflação de 1,12% registrada em março de 2015. As demais categorias apresentam-se de modo semelhante entre 2014 e 2016, com seus índices oscilando em torno do valor zero.

CURITIBA

Verificou-se em seguida o comportamento dos dados referentes ao IPCA-15 discriminados pelas categorias de consumo para a Região de Curitiba no período de Janeiro de 2014 a Agosto de 2016, ilustrados na Figura 3.7.

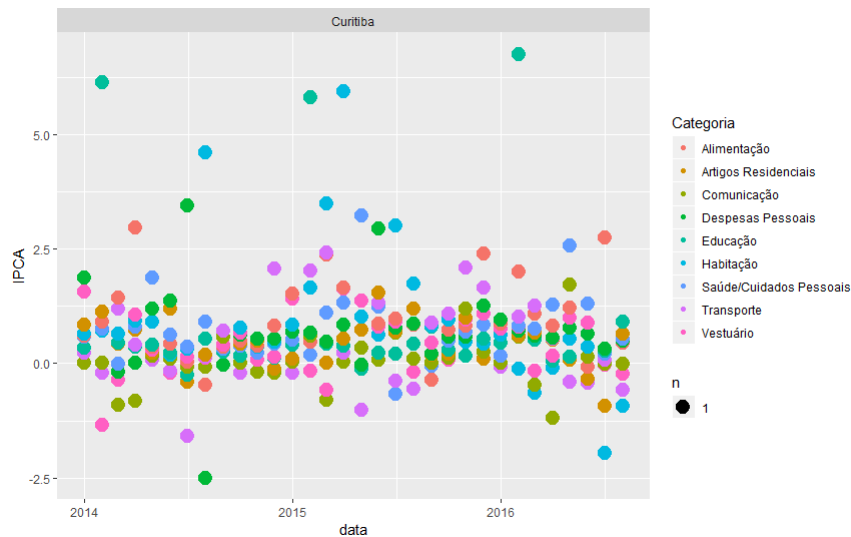


Figura 3.7: IPCA-15 para a Região de Curitiba-PR de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.

Neste sentido, através da Figura 3.7 é possível observar uma variabilidade significativa dos índices do IPCA-15 em diversos setores. A categoria de consumo Educação apresenta os mais altos índices dentro do período calculado nos meses de fevereiro de 2014, 2015 e 2016, que, em contramão à Região de Belo Horizonte teve um aumento de aproximadamente 16% passando de 5,81% em 2014 para 6,75% em 2016. É importante destacar também o setor de Habitação, que representa 14,62% do cálculo final do valor do IPCA-15 exibindo altos índices no mês de agosto de 2014 (4,61%) e abril de 2015 (5,94%), com uma alta de aproximadamente 29% em menos de um ano, e contribuindo para que a Região de Curitiba tivesse a maior inflação do país em 2015, no valor de 10,67%, a maior desde 2002.

Além disso, segundo o IBGE, a alta da inflação em Curitiba foi impactada diretamente pelo ajuste de 50% nas alíquotas do Imposto sobre Circulação de Mercadorias e Serviços (ICMS), além do aumento de 69,22% nas tarifas de energia elétrica daquele ano.

Por outro lado, a categoria de Despesas Pessoais apresentou a maior deflação no período estudado na Região de Curitiba, em agosto de 2014, de valor 2,5%, que apresentou aumento nos anos seguintes, fechando agosto de 2016 com uma inflação de 0,5%. Nesta mesma linhagem, pode-se destacar também categoria Habitação, que, após mostrar grande alta de 5,94% em abril de 2015, apresentou uma deflação de 1,95% em Julho de 2016.

FORTALEZA

Pode-se analisar os valores referentes ao IPCA-15 por categoria de consumo para a Região de Fortaleza através da Figura 3.8 .

Nesse caso, verifica-se na Região de Fortaleza uma concentração de índices referentes ao IPCA-15 em torno do valor zero por todo o período estudado. Porém, alguns pontos dispersos pelo gráfico geram interpretações acerca de algumas categorias de consumo nesta Região. Primeiramente, pode-se destacar, como em abordagens anteriores, que o setor de Educação

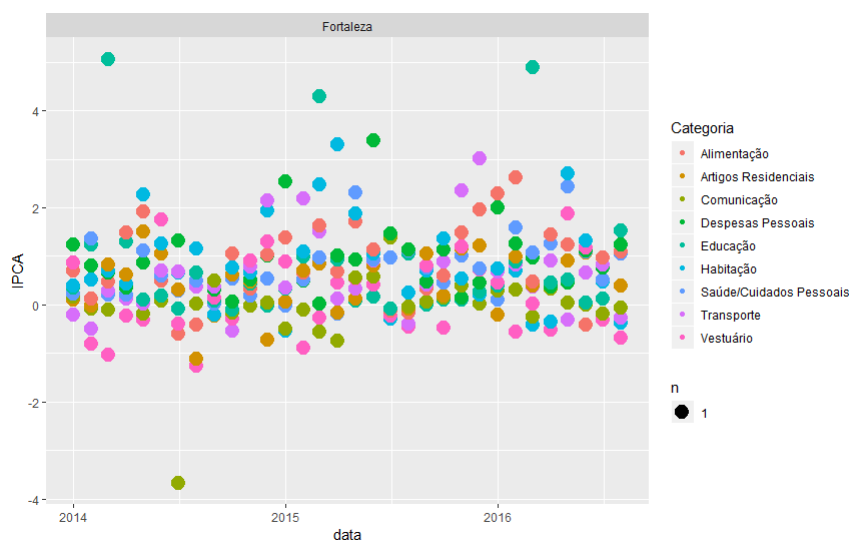


Figura 3.8: IPCA-15 para a Região de Fortaleza-CE de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo

apresenta os mais altos índices dentre todas as categorias analisadas, todavia, diferente das outras Regiões, estas altas concentram-se nos meses de março de 2014 (5,06%), 2015 (4,28%) e 2016 (4,89%), com uma média de 4,74% nestes períodos. Por outro lado, o setor de Comunicação apresentou uma alta deflação em julho de 2014, no valor de 3,67% para o respectivo mês. No mesmo período do ano seguinte, este setor apresentou forte alta, com uma inflação de magnitude de 1,39%, mas que recuou novamente em julho de 2016, apresentando novamente uma deflação, de valor 0,18%. A categoria de Vestuário apresentou uma deflação predominante, principalmente nos meses de fevereiro do período estudado, com índices de 0,81% (2014), 0,89% (2015) e 0,55% (2016).

PORTO ALEGRE

De acordo com a Figura 3.9, pode-se analisar os índices referentes ao IPCA-15 referentes às categorias de consumo para a Região de Porto Alegre.

De acordo com a Figura 3.9, pode-se verificar que as categorias de consumo Educação e Habitação, alternam-se nos mais altos índices verificados para o período de Janeiro de 2014 a Agosto de 2016 para a Região de Porto Alegre. Este primeiro apresenta-se nos meses de fevereiro de cada ano, com inflação registrada de 3,27% (2014), 3,96% (2015) e 5,63% (2016) para a categoria de Educação, tendo uma alta expressiva de aproximadamente 72% em 2 anos devidos aos sucessivos reajustes de mensalidade escolar. Já na categoria de Habitação, que possui impacto quatro vezes maior no cálculo geral do IPCA do que a categoria Educação, as maiores altas ocorreram no primeiro semestre de 2015, nos meses de janeiro (3,13%), fevereiro (2%), março (3,62%) e abril (4,15%), um reajuste de aproximadamente 33% nos quatro primeiros meses deste ano. Por outro lado, as maiores deflações no período estudado foram verificadas nas categorias Vestuário e Comunicação. Esta primeira apresentou o mais alto índice de deflação registrado na Região de Porto Alegre em fevereiro de 2014 (2,17%). Também se destacam os

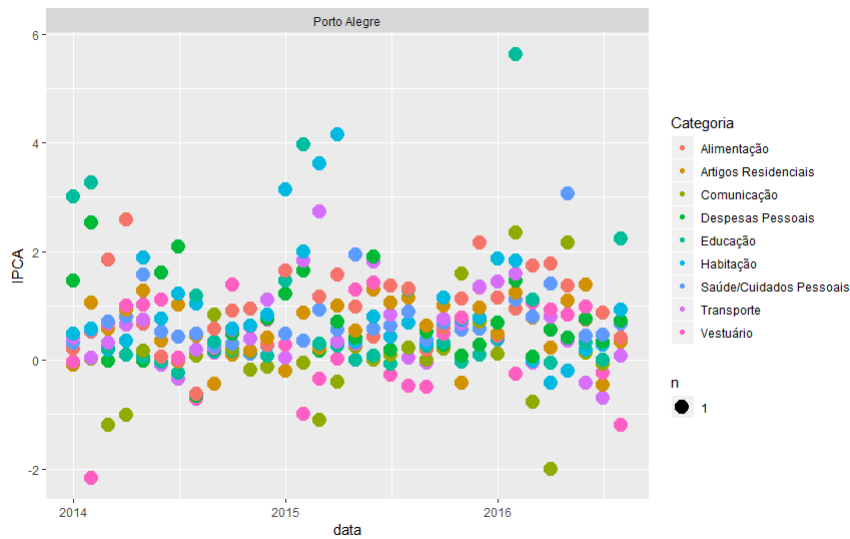


Figura 3.9: IPCA-15 para a Região de Porto Alegre-RS de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.

níveis deflacionários desta categoria em agosto de 2016 (1,19%) e Fevereiro de 2015 (1%). Já no setor de Comunicação as maiores deflações ocorreram nos meses de março e abril dos períodos analisados. Em março de 2014, a deflação registrada foi de 1,2%, seguida por 1,1% (em 2015) e 0,77% (em 2016). No mês de abril de 2014, o valor deflacionário deste setor foi de 1,02%, seguido de 0,41% (em 2015) e 2% (em 2016).

RECIFE

Foi verificado o comportamento e distribuição dos índices referentes ao IPCA-15 segundo às categorias de consumo para a cidade de Recife (Figura 3.10).

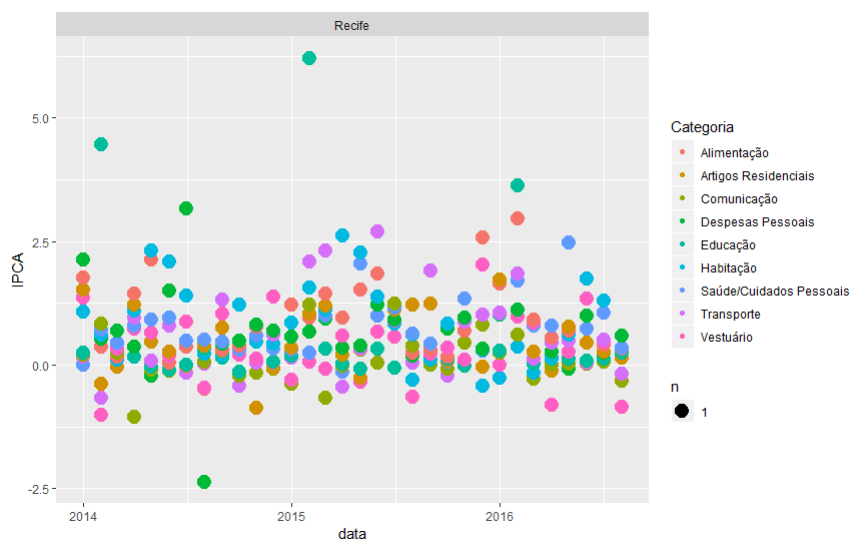


Figura 3.10: IPCA-15 para a Região de Recife-PE de Janeiro de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.

Através da Figura 3.10 pode-se verificar uma alta variabilidade entre os índices do IPCA-15 nas categorias analisadas. Como verificado anteriormente em outras cidades, a categoria

Educação apresenta os mais altos índices registrados durante o período do estudo nos meses de fevereiro de 2014 (4,46%), 2015 (6,21%) e 2016 (3,36%), em que é possível analisar uma variação significativa entre esses valores.

Os setores de Despesas Pessoais, dentre os quais, Alimentação e Transporte também apresentam altos índices inflacionários no período estudado, apesar da categoria de Despesas Pessoais também ter apresentado a maior deflação dentro do período analisado, de valor 2,37%, em agosto de 2014. Comunicação e Vestuário são setores que também apresentaram índices deflacionários ao longo do estudo, embora não haver uma significância muito grande.

RIO DE JANEIRO

A Figura 3.11 apresenta os índices do IPCA-15 discriminado por categorias de consumo para a Região do Rio de Janeiro.

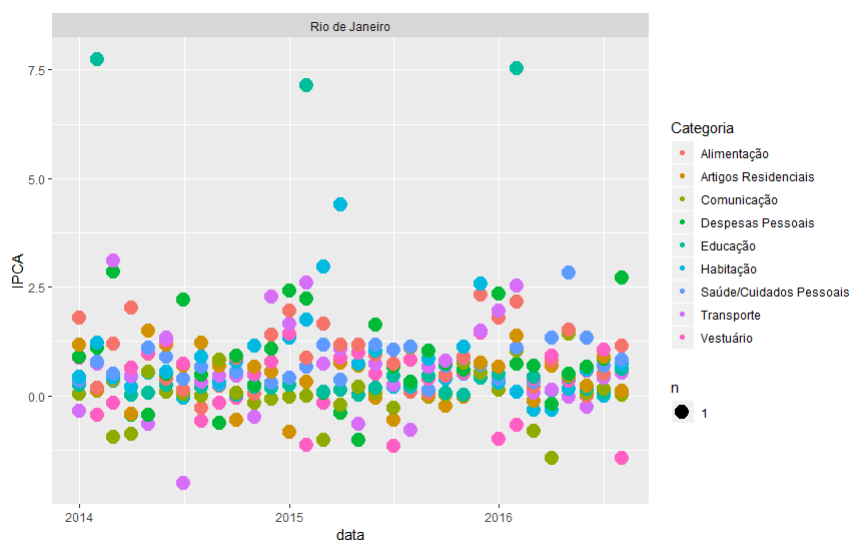


Figura 3.11: IPCA-15 para a Região do Rio de Janeiro-RJ de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.

Verifica-se que na Região do Rio de Janeiro, os maiores índices registrados do IPCA-15 no período estudado concentram-se na categoria de Educação, nos meses de fevereiro, sendo de 7,75% em 2014, 7,14% em 2015 e 7,54% em 2016, valores que destoam bastante dos demais, como pode ser visto na Figura 3.11. O setor de Habituação é também destaque por seus índices inflacionários no ano de 2015. Nos primeiros meses deste ano é possível verificar um crescimento de aproximadamente 233% do mês de janeiro (1,32%) ao mês de abril (4,39%), impactado pela crise econômica daquele ano, que sofre queda nos meses posteriores, mas que volta a crescer nos meses de novembro (1,12%) e dezembro (2,58%).

Quando analisadas as deflações desta região, verifica-se que a mais alta foi registrada em julho de 2014 (2,01%) na categoria de Transportes, impulsionada, dentre outros fatores, pela realização da Copa do Mundo nesta região naquele ano. Os setores de Comunicação e Vestuário também se destacam pelos valores deflacionários nos anos de 2015 e 2016.

SALVADOR

Para verificar o comportamento dos índices do IPCA-15 na cidade de Salvador segundo categorias de consumo, foi gerado um gráfico abrangendo o período de janeiro de 2014 a agosto de 2016, que pode ser verificado através da Figura 3.12.

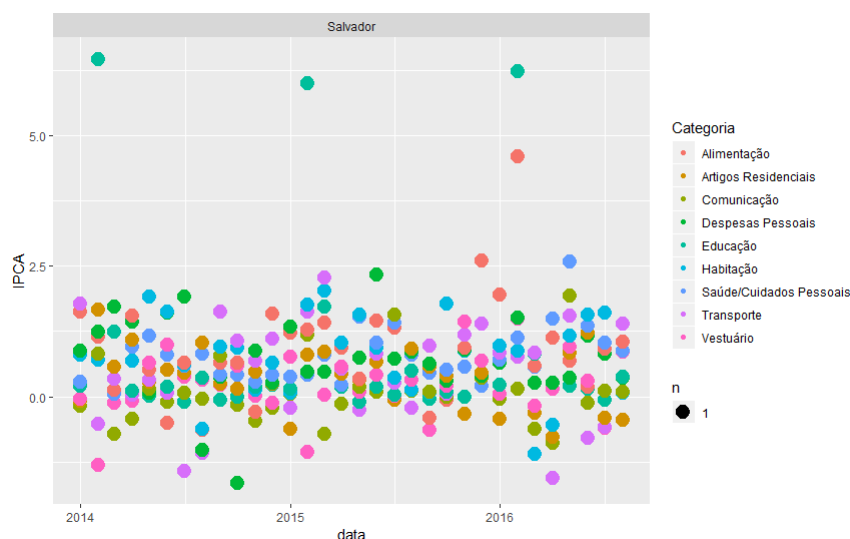


Figura 3.12: IPCA-15 para a Região de Salvador-BA de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.

Nesta Região, como nas demais até aqui analisadas, o setor de Educação apresenta os mais altos índices do IPCA no mês de fevereiro dos 3 anos analisados, sendo de 6,46% em 2014, 6% em 2015 e 6,22% em 2016. Porém, pode-se destacar também altos índices da categoria Alimentação, que possui o mais alto impacto no cálculo geral do IPCA, em dezembro de 2015 (2,61%) e fevereiro de 2016 (4,59%), uma alta de aproximadamente 76% em um período de 2 meses. Na contramão, o mais alto índice deflacionário registrado na Região de Salvador durante o estudo ocorreu em Outubro de 2014, de valor 1,65%, embora o setor de Transportes também apresentou deflações em vários períodos nos anos de 2014 e 2016.

SÃO PAULO

Finalmente, verificam-se os índices do IPCA-15 discriminados pelas categorias de consumo para a cidade de São Paulo, ilustrados na Figura 3.13.

Verifica-se nesta Região que os mais altos índices inflacionários foram referentes aos setores de Educação e Habitação. Neste primeiro, o mês de fevereiro aparece com a maior carga inflacionária para esta categoria, registrando 6,18% em 2014, 5,83% em 2015 e 5,72% em 2016, uma queda de aproximadamente 7% para este mês em 3 anos. O setor de Habitação registrou seus maiores índices inflacionários nos primeiros meses do ano de 2015, passando de 3,54% no mês de abril (situação mais grave em 2015) para 0,32% no mês de dezembro, uma queda de aproximadamente 91% em um período de 8 meses.

Esta mesma Região não apresentou grandes deflações durante o período analisado, sendo as maiores nos setores de Comunicação (2,69%, em agosto de 2014), Vestuário (1,83%, em

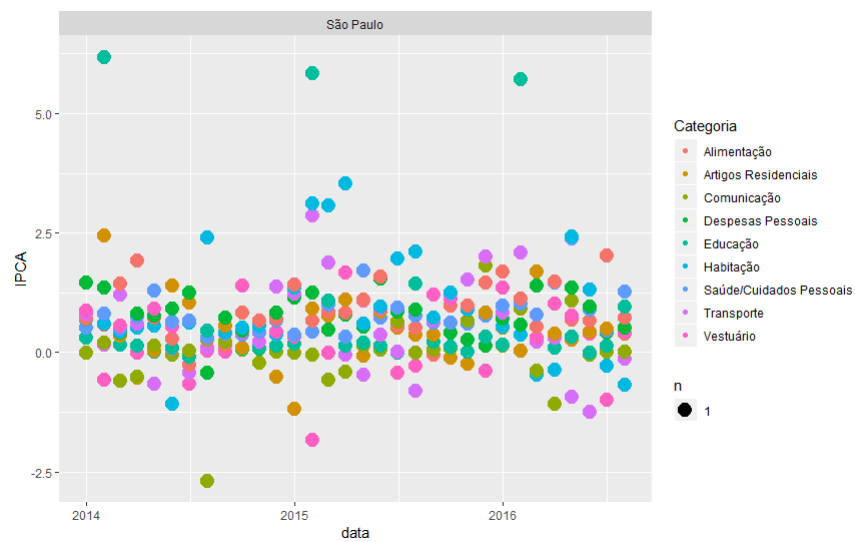


Figura 3.13: IPCA-15 para a Região de São Paulo-SP de 2014 a Agosto de 2016 segundo suas Categorias de Consumo.

fevereiro de 2015) e Transportes (1,25%, em junho de 2016).

3.2 ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Componentes Principais (ACP) permite visualizar informações em uma matriz de dados contendo indivíduos/observações descritos por variáveis quantitativas inter-relacionadas [1]. Cada variável pode ser considerada como uma diferente dimensão. Se há mais de três variáveis na matriz de dados, pode ser complicado realizar interpretações em um espaço multidimensional, sendo esta análise utilizada para extrair informações importantes de uma tabela de dados multivariada e expressar estas informações como um conjunto de novas variáveis, chamadas de Componentes Principais. Estas novas variáveis correspondem a uma combinação linear entre as variáveis originais e o número de Componentes, que é sempre menor ou, além disso, igual ao número de variáveis originais.

A informação em um dado conjunto de dados corresponde à variação total que ele contém. O objetivo da ACP é o de identificar direções tal que a variação dos dados seja máxima [13]. Em outras palavras, a ACP reduz a dimensionalidade de dados multivariados para dois ou três Componentes Principais, que podem ser visualizados graficamente, com a mínima perda de informação.

Serão descritos neste capítulo os resultados referentes às Análises de Componentes Principais aplicadas aos dados do IPCA-15 para todo o período do estudo, em que as categorias de consumo serão as variáveis a serem analisadas e as Regiões de coleta, os indivíduos, enfatizando revelar aquelas com maior importância para explicar as variações nos conjuntos de dados. Utilizando o R, os pacotes utilizados na pesquisa foram o **FactoMineR** (para as análises) e o **factoextra** (para a elaboração de gráficos) em que previamente foi aplicada uma padronização nos dados para facilitar a comparação dos resultados.

A adequabilidade de conduzir uma solução de ACP para os dados foi avaliada usando o teste de adequação da amostragem e de esfericidade. A adequação da amostra foi estabelecida pelo cálculo do índice de Kaiser-Meyer-Olkin (KMO), ao considerar que valores entre 0,5 e 1 indicam que a análise é adequada. O teste de esfericidade de Bartlett avalia a existência de correlações significativas na matriz de correlações testando a hipótese nula de a matriz de correlações ser a matriz identidade [7]. O teste KMO, para medir a adequação da amostra, foi de 0,79, indicando que a decisão de proceder à ACP foi adequada; o teste de Bartlett indicou esfericidade de 859,95 e nível de significância zero, apontando que é improvável que a matriz de correlação seja uma matriz identidade.

Os Autovalores medem a quantidade de variação retida por cada Componente Principal (CP). Eles são maiores para os primeiros Componentes e decrescem para os seguintes, ou seja, os primeiros Componentes correspondem às direções com a máxima quantidade de variação no conjunto dos dados. Seus respectivos valores são ilustrados na Tabela 3.4, em que também podem ser vistos o percentual de variabilidade de cada uma destas Componentes bem como o valor acumulado destas variações.

A soma dos valores dos Autovalores da Tabela 3.4 fornecem uma variação total, de valor 10. A proporção da variação explicada por cada Autovalor é verificada na segunda coluna.

Tabela 3.4: Autovalores e seus respectivos percentuais de variação para cada Dimensão (Componente).

Autovalor	Coefficiente	Variância (%)	Variância Acumulada (%)
Dimensão 1	3,2331	35,92	35,92
Dimensão 2	1,8959	21,07	56,99
Dimensão 3	1,6468	18,30	75,29
Dimensão 4	1,1474	12,75	88,04
Dimensão 5	0,6814	7,57	95,61
Dimensão 6	0,2566	2,85	98,46
Dimensão 7	0,1078	1,20	99,66
Dimensão 8	0,0309	0,34	100,00

Por exemplo, aproximadamente 35,92% da variação total é explicada pelo primeiro Autovalor. A porcentagem de explicação acumulada da variabilidade é obtida pela adição das proporções sucessivas da variação explicada para obtenção do valor total de 100%, ou seja, cerca de 56,99% da variação total é explicada pelos dois primeiros Autovalores em conjunto.

Autovalores podem ser utilizados para determinar a quantidade de Componentes Principais retidos após a ACP [14]. Segundo este critério, um Autovalor maior que 1 indica que os Componentes Principais relatam uma maior variação do que aqueles relatados por uma das variáveis originais nos dados padronizados, o que é comumente utilizado como sendo este o ponto de corte para definir quais CP's serão retidos. Outro critério de decisão dá-se por limitar o número de Componentes para o qual corresponda uma certa fração da variação total. Nesta análise, os três primeiros Componentes explicam cerca de 75,3% da variação, uma larga e aceitável porcentagem.

Um método alternativo para determinar o número de Componentes Principais necessários é o de utilização do gráfico Scree Plot (Figura 3.14), caracterizado por ser um gráfico em que os Autovalores encontram-se ordenados do maior para o menor valor da variação explicada. O número de Componentes é determinado no ponto além do qual os Autovalores restantes são relativamente pequenos e de tamanho comparável [24].

Por meio da Figura 3.14 e da definição do critério de parada, os cinco primeiros Componentes contêm aproximadamente 95,6% da informação (variâncias) dos dados analisados, sendo este o ponto de parada.

A correlação entre determinada variável e o Componente Principal é utilizada como as coordenadas da variável no CP. Sua representação difere do gráfico de observações: as observações são representadas por suas projeções, mas as variáveis são projetadas por suas correlações [12]. As coordenadas das variáveis encontram-se na Tabela 3.5.

Na Figura 3.15 é possível verificar a representação gráfica das variáveis sob estudo. Este gráfico ilustra a relação entre estas variáveis, em que aquelas correlacionadas positivamente são agrupadas em conjunto, enquanto que variáveis negativamente correlacionadas são posicionadas em lados opostos do ponto de origem (quadrantes opostos).

Além disso, verifica-se que as categorias de Educação e Saúde possuem uma correlação positiva, em contraste com a categoria de Artigos Pessoais, para os dois primeiros Componentes.

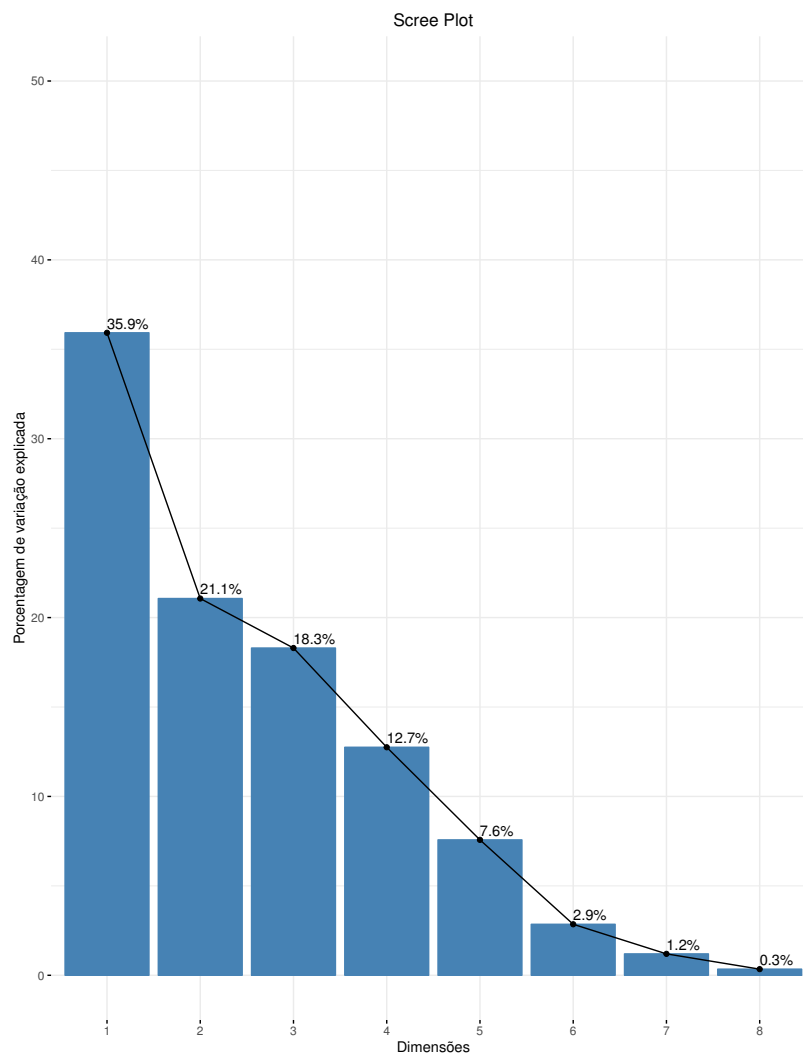


Figura 3.14: Porcentagem da variação explicada pelos Componentes Principais.

Tabela 3.5: Coordenadas das Variáveis dos Componentes Principais retidos no estudo.

Categoria	CP 1	CP 2	CP 3	CP 4	CP 5
Alimentação	0,0538	0,7996	0,0753	0,2733	0,5188
Artigos	-0,3086	-0,2485	0,8841	0,2007	0,0516
Comunicação	-0,8725	0,2009	-0,2757	0,1957	0,3048
Despesas	0,9286	-0,1362	-0,0357	0,2078	-0,0925
Educação	0,5335	0,4481	0,0429	0,6461	-0,2597
Habitação	0,2049	0,5473	0,7337	-0,3091	-0,0427
Saúde	0,6048	0,4664	-0,3970	-0,4559	0,0863
Transporte	0,6485	-0,4265	0,2785	-0,3190	0,3573
Vestuário	-0,6312	0,5254	0,0829	-0,3583	-0,3220

Da mesma forma, as variáveis Vestuário e Transportes, bem como as variáveis Comunicação e Despesas Pessoais apresentam-se negativamente correlacionadas, ou seja, quanto maior o índice do IPCA-15 da categoria Vestuário, menor é o índice da categoria de Transporte e vice-versa. As categorias de Habitação e Alimentação possuem uma correlação positiva para estes Componentes, ou seja, à medida que o índice do IPCA-15 para a Alimentação aumenta, também aumenta-se o índice da categoria Habitação, bem como quando há diminuição no índice de uma

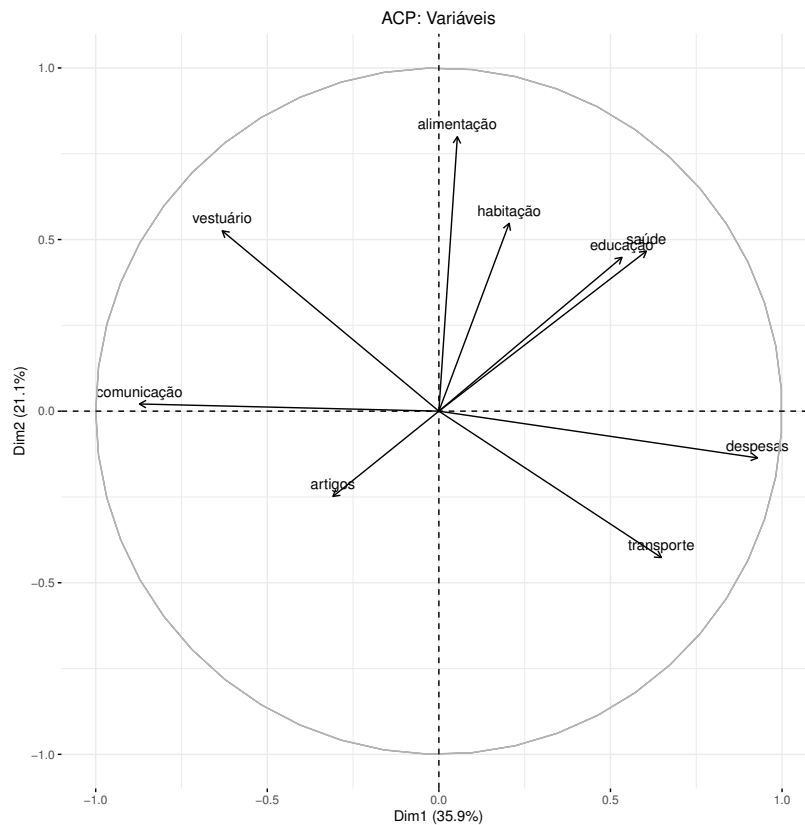


Figura 3.15: Gráfico das Variáveis na Análise de Componentes Principais (ACP).

destas variáveis, o da outra também diminui.

Outro ponto importante a ser analisado quando a ACP é instaurada dá-se pela Qualidade de Representação (\cos^2) das variáveis do estudo. Ela indica a contribuição de um Componente, da distância quadrática da observação até a origem. Isto corresponde ao quadrado do cosseno do ângulo da direita do triângulo obtido com a origem, a observação e sua projeção no determinado Componente. Componentes com um alto valor de \cos^2 contribuem de maneira significativa à distância total e são consideradas importantes para tal observação [20]. Este valor auxilia em encontrar os Componentes que são importantes para interpretar tanto observações ativas quando suplementares. Os valores referentes à Qualidade de Representação das variáveis nas Dimensões 1 e 2 apresentam-se na Tabela 3.6.

Um alto valor de \cos^2 indica uma boa representação da variável no Componente Principal, ao mostrar que quanto maior este valor, mais próxima à circunferência do círculo de correlação desta variável encontra-se posicionada, como acontece com a categoria de Despesas Pessoais. Já as variáveis com baixo valor de \cos^2 localizam-se próximas ao centro do círculo, indicando que tal variável não é perfeitamente representada pelos Componentes Principais, ou seja, é menos importante para os dois primeiros Componentes, caso da categoria de Artigos Pessoais.

Tabela 3.6: Qualidade de Representação das variáveis.

Categoria	CP 1	CP 2
Alimentação	0,0029	0,6394
Artigos	0,0952	0,0618
Comunicação	0,7612	0,0004
Despesas	0,8624	0,0186
Educação	0,0420	0,2995
Habitação	0,3658	0,2175
Saúde	0,3658	0,2175
Transporte	0,4205	0,1819
Vestuário	0,3985	0,2760

Como se observa para algumas das variáveis, mais do que dois Componentes podem ser necessários para representar perfeitamente os dados. A Figura 3.16 ilustra, a partir da utilização de gradientes, uma descrição da intensidade de importância das variáveis descritas aos dois primeiros Componentes. Aquelas com coloração branca apresentam um baixo valor de \cos^2 , já as em azul apresentam um valor médio, e aquelas em vermelho apresentam um alto valor de \cos^2 .

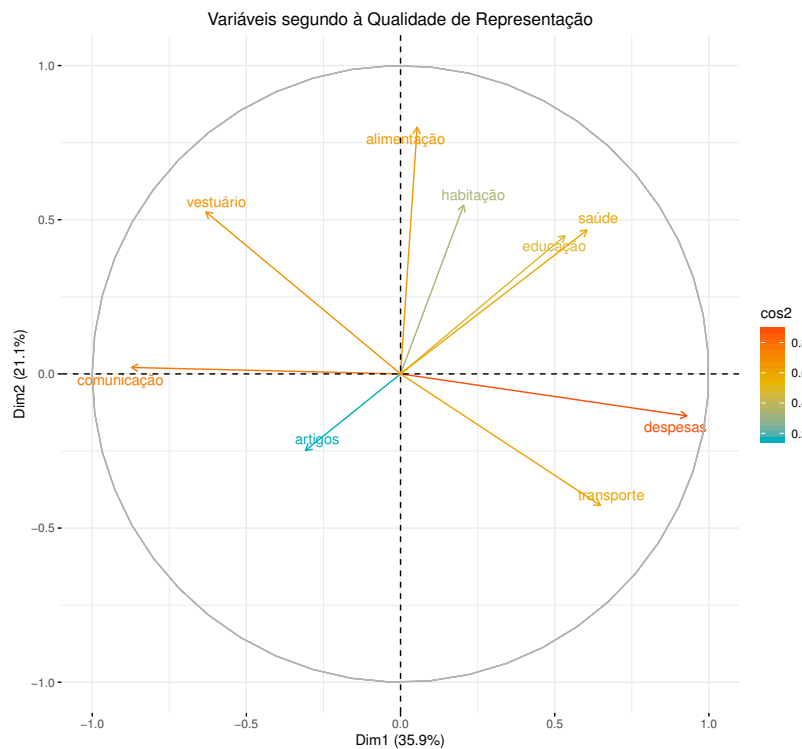


Figura 3.16: Gradiente de variáveis segundo à Qualidade de Representação

Por meio da Figura 3.16 é possível verificar a formação de 5 grupos com comportamentos semelhantes. O primeiro é representado pela categoria de Artigos Pessoais, caracterizada por apresentar o menor valor de \cos^2 dentre todas as variáveis. Em seguida, a categoria de Habitação apresenta-se razoavelmente descrita pelas duas primeiras dimensões. Tem-se então as categorias de Educação, Saúde e Transportes, que possuem valores intermediários do \cos^2 . Em continuidade, mostrando bons valores de \cos^2 , encontram-se as categorias de Alimentação e Vestuário. Finalmente, as categorias de Comunicação e Despesas, com altos valores de \cos^2 , são muito bem explicadas pelos dois primeiros Componentes Principais.

A contribuição das variáveis, considerando um dado Componente Principal, é expressa em porcentagem. Variáveis relacionadas com o CP 1 (Dim. 1) e CP 2 (Dim. 2) são mais importantes na explicação da variabilidade do conjunto de dados. Já aquelas que não se correlacionam com nenhum dos CP ou correlacionam-se com as últimas Dimensões são variáveis com pouca contribuição e podem ser removidas para simplificar a análise como um todo. A tabela 3.7 apresenta as contribuições das variáveis sob estudo para os dois primeiros Componentes Principais retidos na análise. Quanto maior o valor da contribuição, maior é a contribuição da variável para a determinada Componente ou Dimensão. A Figura 3.17 ilustra as contribuições de cada variável para os dois primeiros Componentes.

Tabela 3.7: Contribuição das Variáveis aos Componentes Principais (%).

Categoria	Dim. 1	Dim. 2
Alimentação	0,09	33,73
Artigos	2,94	3,26
Comunicação	23,54	0,02
Despesas	26,67	0,98
Educação	8,80	10,59
Habitação	1,30	15,80
Saúde	11,31	11,47
Transporte	13,01	9,60
Vestuário	12,32	14,56

A linha vermelha tracejada nos gráficos da Figura 3.17 indica a contribuição média esperada de cada variável, ou seja, como existem nove variáveis sob estudo, a contribuição média esperada de cada uma destas variáveis para cada Componente é de aproximadamente 11,11% (100% dividido por 9). Dada uma Componente, as variáveis com contribuição superior ao valor de corte anteriormente definido são consideradas importantes na explicação de tal Componente. De acordo com o gráfico da esquerda da Figura 3.17, verifica-se que as categorias de Despesas Pessoais (26,67% de contribuição) e Comunicação (23,54%) possuem a maior contribuição para o Componente Principal 1, ou seja, estas variáveis são expressamente importantes na descrição deste Componente. Além disso, verifica-se que estas variáveis possuem forte correlação com esta Dimensão, de valor 0,9286 para a categoria de Despesas Pessoais e de valor -0,8724 para a categoria de Comunicação. Além delas, as categorias de Transporte (13%), Vestuário (12,32%) e Saúde (11,31%) também são significativamente importantes contribuições à esta Componente. Já no gráfico da direita desta Figura, a variável Alimentação possui grande contribuição ao

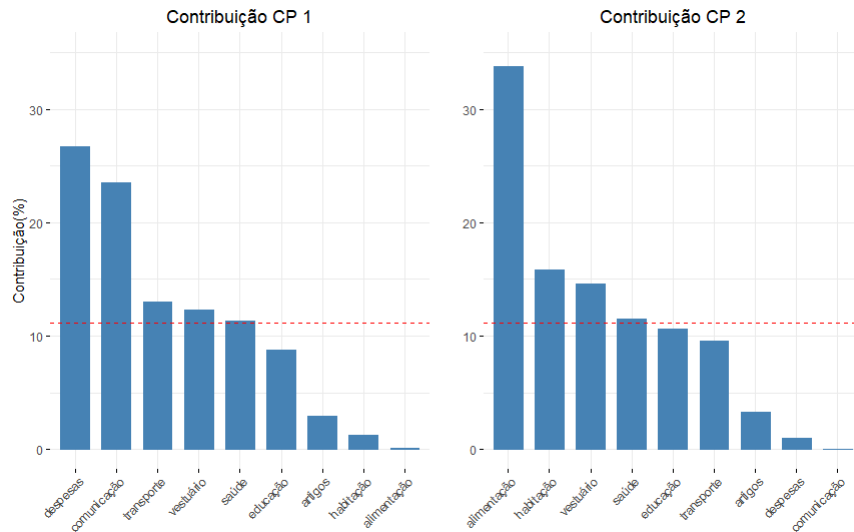


Figura 3.17: Contribuição das variáveis nas Dimensões 1 (esquerda) e 2 (direita).

Componente Principal 2 quando comparada às demais (33,73%), sendo então a principal forma de explicação do comportamento desta Dimensão.

O valor da correlação entre esta categoria e a segunda Dimensão é de 0,7996, indicando assim uma correlação positiva forte. As categorias de Habitação (15,80%), Vestuário (14,56%) e Saúde e Cuidados Pessoais (11,47%) também apresentam significância na contribuição do segundo Componente Principal, sendo fundamental em sua descrição. As variáveis com maior contribuição para estes Componentes ao realizar a análise conjunta são descritas no gráfico de correlações da Figura 3.18 .

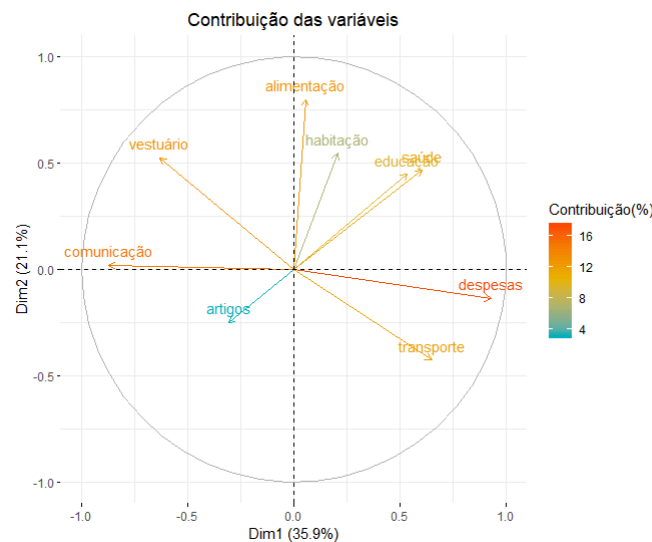


Figura 3.18: Contribuição conjunta das variáveis para as Componentes 1 e 2.

Na Figura 3.18 é possível traçar um grau de hierarquia segundo às contribuições de cada variável nos dois Componentes Principais do estudo. A categoria de Artigos Pessoais define-se como a variável com a menor contribuição dentre as demais. Em seguida, a categoria de Habitação encontra-se em patamar semelhante, com pouca contribuição para os dois primeiros CPs. As variáveis Transporte, Educação e Saúde e Cuidados Pessoais apresentam uma

significativa, porém mediana, contribuição para estes Componentes. Já Alimentação e Vestuário figuram como categorias de importância, seguidas pela categoria de Comunicação, com potencial descritivo sobre o comportamento destas Componentes. Finalmente, a categoria de Despesas Pessoais apresenta-se como a mais importante dentre as variáveis analisadas na descrição e contribuição dos dois Componentes com as maiores porcentagens de explicação do conjunto de dados analisado.

Nesses termos, o passo seguinte é o de observar o comportamento dos indivíduos (cidades) no estudo. Para isso, os gráficos das Figuras 3.19 e 3.20 ilustram a contribuição destes indivíduos para os dois primeiros Componentes Principais através da mensuração da qualidade de representação (\cos^2) de cada um deles.

Na Figura 3.19 pode-se observar que os indivíduos que menos contribuem para as Dimensões 1 e 2 são as Regiões de Porto Alegre, Salvador e São Paulo, localizadas bem próximas à origem do plano cartesiano. Já as Regiões de Belém, Recife e Belo Horizonte, correlacionam-se negativamente com os dois primeiros Componentes, porém, apresentam uma Qualidade de Representação significativa, ao passo da Região do Rio de Janeiro, que em contrapartida, correlaciona-se positivamente com ambas as Dimensões. As Regiões de Curitiba e Fortaleza apresentam as maiores Qualidades de Representação observadas no gráfico. Fortaleza correlaciona-se fortemente, e de forma positiva, com o primeiro Componente Principal, e de maneira negativa com o segundo. Já a Região de Curitiba apresenta uma forte correlação positiva com o segundo Componente Principal, além de correlacionar negativamente com o primeiro.

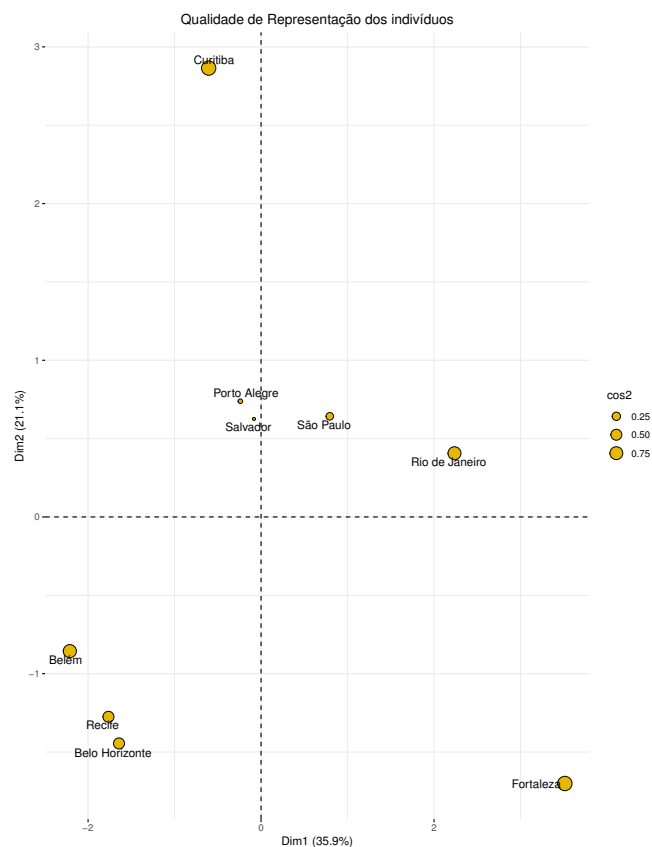


Figura 3.19: Qualidade de Representação dos indivíduos do estudo.

Quando analisada a contribuição conjunta dos indivíduos para as duas primeiras dimensões, novamente espera-se que os indivíduos que contribuem de forma significativa para estes Componentes atinjam o valor mínimo de 11,11% de contribuição.

Com o auxílio da Figura 3.20, verifica-se que apenas quatro dentre todos os indivíduos analisados conseguem atingir este valor, sendo eles, em ordem de importância, Fortaleza, Curitiba, Belém e o Rio de Janeiro. Dentre estas Regiões, Fortaleza impõe-se como a principal em contribuição, destacando assim a importância deste indivíduo na descrição dos dois primeiros Componentes.

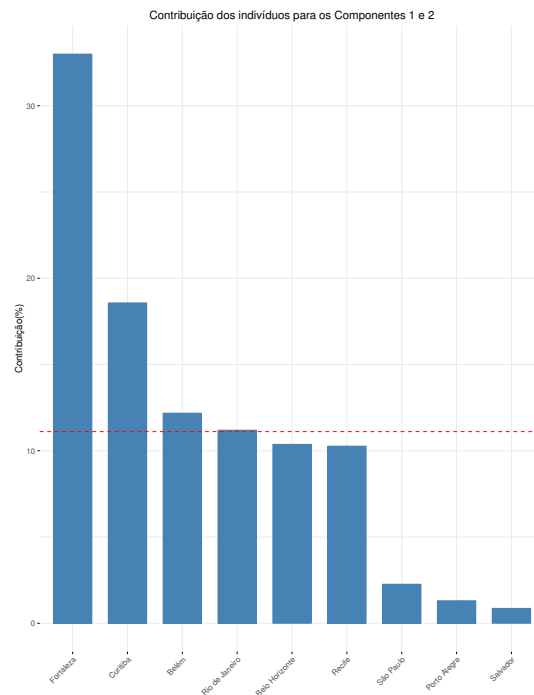


Figura 3.20: Contribuição conjunta dos indivíduos para as Componentes 1 e 2.

Finalmente, a Figura 3.21 apresenta um gráfico Biplot, ao contrastar indivíduos e variáveis em uma única imagem, para que interpretações acerca de suas interações sejam realizadas.

É necessário verificar na Figura 3.21 que as coordenadas dos indivíduos e das variáveis não são construídas no mesmo espaço. Variáveis localizadas próximas as outras apresentam uma alta correlação entre elas, e em sentidos opostos, correlações negativas, como é o caso das categorias de Saúde e Educação, bem como Alimentação e Habitação, que encontram-se fortemente relacionadas entre si e ao Primeiro Componente Principal. As Regiões do Rio de Janeiro, Fortaleza e São Paulo apresentaram uma correlação positiva com estas variáveis, indicando assim que, quanto maior o valor do IPCA-15 destas categorias, maiores tendem a ser os índices inflacionários destas Regiões comparadas as demais. Em contrapartida, as cidades de Belém, Recife e Belo Horizonte não sofrem o mesmo impacto, por apresentarem uma correlação negativa com estas categorias.

Quando a correlação com o segundo Componente Principal é analisada para estas variáveis, a cidade de Curitiba mostrou-se fortemente correlacionada a estas categorias, de forma positiva, justificando assim os altos valores inflacionários desta Região para estes setores. As Regiões



Figura 3.21: Gráfico Biplot para as variáveis e indivíduos no estudo.

de Porto Alegre, Salvador, São Paulo e Rio de Janeiro também apresentaram uma correlação positiva com estas variáveis, mas de pouca significância. É interessante destacar, porém, que a cidade de Fortaleza, apresentou para estas categorias uma relação negativamente forte, quando analisada sob a perspectiva deste segundo Componente, o que é semelhante às Regiões de Belo Horizonte e Recife, e também de Belém, apesar de esta última possuir uma correlação relativamente mais fraca.

As categorias de Despesas Pessoais e Transportes também apresentam uma correlação positiva com o primeiro Componente Principal, embora, diferente das demais citadas até aqui, exibam uma correlação negativa com o segundo Componente. Para estas variáveis, as Regiões de Belém, Belo Horizonte, Recife e Fortaleza possuem uma correlação positiva, ao passo que índices mais altos do IPCA-15 nestas Regiões são refletidos de maneira mais acentuada do que nas demais.

As categorias de consumo Vestuário e Comunicação mostraram-se negativamente relacionadas ao primeiro Componente Principal. As Regiões de Belém, Recife e Belo Horizonte apresentaram uma correlação positiva com estas categorias nesta Componente. Em contrapartida, as Regiões do Rio de Janeiro e Fortaleza exibiram altos valores de uma correlação negativa com estas variáveis no primeiro Componente, o que indica que nestas localidades, o aumento dos índices inflacionários destas categorias são refletidos de modo oposto, ou seja, não apresentam altos valores quando há um crescimento da inflação, diferente das demais. Já para o segundo Componente Principal, estas categorias estão relacionadas de forma positiva, e a

Região de Curitiba é a que apresentou maior impacto das mesmas, ao mostrar uma correlação positiva com estas categorias. Por outro lado, as Regiões de Fortaleza, Belo Horizonte e Recife apresentam uma correlação negativa com estas variáveis no segundo Componente.

A categoria de Artigos Pessoais, tal como Vestuário e Comunicação, também apresentou uma correlação negativa com o Primeiro Componente, porém, correlacionou-se também de modo negativo com o segundo Componente Principal. Neste último, é possível analisar uma relação positiva das Regiões de Fortaleza, Belo Horizonte, Recife e Belém com esta variável, e uma relação negativa principalmente com a Região de Curitiba, ao concluir que nesta Região não houve um aumento expressivo dos índices do IPCA-15 desta categoria, quando comparada às demais.

3.3 ANÁLISE DE AGRUPAMENTO

Para analisar o comportamento das Regiões ilustradas no estudo, foi implementada a Análise de Agrupamento em cada um dos anos estudados, verificando assim o impacto do IPCA-15, por período, nestas localidades, destacando-se aquelas que se desenvolveram de modo semelhante. Para isso, foi calculado o valor médio do índice do IPCA-15 dentro de cada período analisado, segundo a categoria de consumo. O método de agrupamento hierárquico aglomerativo foi utilizado, classificando as Regiões por suas similaridades e as dividindo em grupos (clusters), onde dentro de cada grupo há o máximo de similaridade entre os indivíduos pertencentes, e a presença de heterogeneidade entre os demais clusters. O algoritmo utilizado trata cada objeto como um cluster e, em seguida, pares de clusters são agrupados até que todos se tornem um único grupo contendo todos os objetos. O resultado é uma representação gráfica dos objetos, chamado dendrograma.

De modo a decidir quais objetos/clusters devem ser combinados, é preciso utilizar de métodos para a mensuração da similaridade entre os objetos. Nas análises seguintes foi adotado o método da Distância Euclidiana para o cálculo destas medidas.

Por fim, é preciso que seja definido um método de ligação, que utiliza das informações das distâncias, para agrupar pares de objetos em clusters baseados em sua similaridade. Estes grupos são então ligados aos demais até que todos os objetos do conjunto de dados originais formem uma ‘árvore hierárquica’. Para os dados do estudo, o método da Ligação Média foi implementado, pois dentre os métodos analisados, este foi o que apresentou o maior coeficiente de correlação cofenético para os três agrupamentos realizados. Este método implica que a distância entre dois clusters é definida como a distância média entre os elementos no primeiro cluster e elementos no segundo.

3.3.1 JANEIRO A DEZEMBRO DE 2014

A primeira análise de agrupamento gerada foi relativa ao período de janeiro a dezembro de 2014, levando-se em consideração os valores médios do IPCA-15 neste ano, segundo às categorias de consumo analisadas. O resultado pode ser verificado na Figura 3.22.

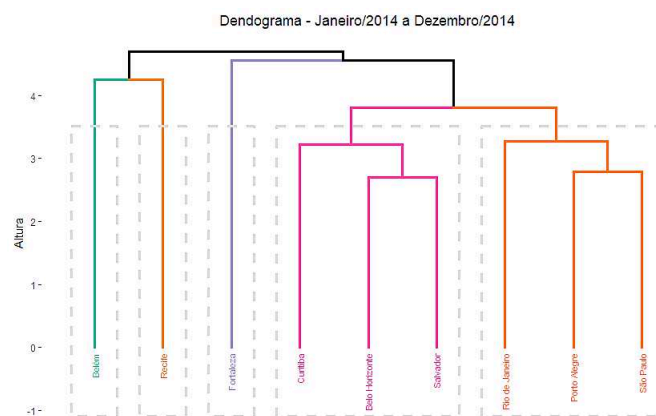


Figura 3.22: Dendrograma referente ao período de Janeiro a Dezembro de 2014 para as Regiões analisadas segundo o IPCA-15.

No dendrograma gerado, cada ‘folha’ corresponde a uma Região do estudo. Quanto mais subimos a ‘árvore’, Regiões similares são combinadas através dos ‘galhos’, unidos em uma altura maior. A altura da fusão apresentada no eixo vertical indica a distância entre dois clusters. Quanto maior a altura de fusão, menor será a similaridade entre as Regiões. Esta altura é definida como a distância cofonética entre dois grupos.

Uma forma de verificar a eficiência da ‘árvore de agrupamentos’ implementada dá-se por averiguar a correlação entre as distâncias cofonéticas e as da matriz de distâncias original. Caso o agrupamento seja válido, a ligação de objetos na árvore de agrupamentos deve ter uma forte correlação com as distâncias entre objetos na matriz de distâncias original. Quanto mais próximo de 1 é o coeficiente, maior a acurácia da técnica de agrupamento implementada nos dados. Como citado anteriormente, ao utilizar o método da Ligação Média por este ter apresentado o maior coeficiente de correlação cofonético para o agrupamento realizando, utilizando-se o valor médio das Categorias de Consumo e das Regiões analisadas, o valor desta correlação foi de 0,7719, o que indica uma forte relação entre as distâncias de ligação e da matriz original.

Através da Figura 3.22, verifica-se a existência de 5 grupos de Regiões. O primeiro é o da Região de Belém, que, em média, teve a categoria de Habitação como o mais alto de seus índices, representando uma inflação média mensal de 1,068% e a categoria de Comunicação como o mais baixo, apresentando uma deflação média mensal praticamente estável, de valor 0,003%. O segundo cluster é formado pela Região de Recife, que semelhante à Belém, teve no setor de Habitação o maior índice médio de inflação (0,934%) e Comunicação com o maior índice de deflação (0,026%).

Posteriormente, o terceiro cluster foi formado pela região de Fortaleza, que apresentou o menor índice médio geral de inflação, analisadas todas as categorias em conjunto, de valor 0,414%. Nesta Região, Habitação também figura como a categoria de consumo com a mais alta inflação média (0,875%) e Comunicação com a maior deflação anual, além de ser a mais alta dentre as Regiões analisadas, de valor 0,307%. O quarto grupo é formado pelas cidades de Curitiba, Salvador e Belo Horizonte, cidades estas que apresentaram uma densidade populacional semelhante, e que mostraram comportamentos parecidos para todas as categorias

analisadas, excetuando a categoria de Educação, que em Belo Horizonte foi responsável pela maior inflação média mensal dentre as categorias analisadas (0,753%), e Habitação, esta com a mais alta inflação média mensal para as Regiões de Salvador (0,747%) e Curitiba (0,911%). Em todas estas cidades, a categoria de Comunicação apresentou a maior deflação.

O último grupo é formado pelas Regiões do Rio de Janeiro, Porto Alegre e São Paulo. Nestas cidades, é interessante analisar que, diferentemente das demais, nenhuma apresenta a Categoria de Habitação como aquela com o mais alto valor inflacionário mensal. Neste grupo, o setor de Despesas Pessoais é detentor deste título para as Regiões de Porto Alegre (0,817%) e São Paulo (0,718%), embora a categoria de Educação ter apresentado o maior crescimento inflacionário mensal para a Região do Rio de Janeiro (0,788%). É interessante destacar que todas as Regiões apresentaram a maior deflação para o setor de Comunicação.

3.3.2 JANEIRO A DEZEMBRO DE 2015

A Figura 3.23 apresenta os resultados do agrupamento realizado para as Regiões analisadas no estudo, segundo o valor médio anual da inflação de cada uma das categorias de consumo citadas no estudo, para o ano de 2015, resumidas através do dendograma.

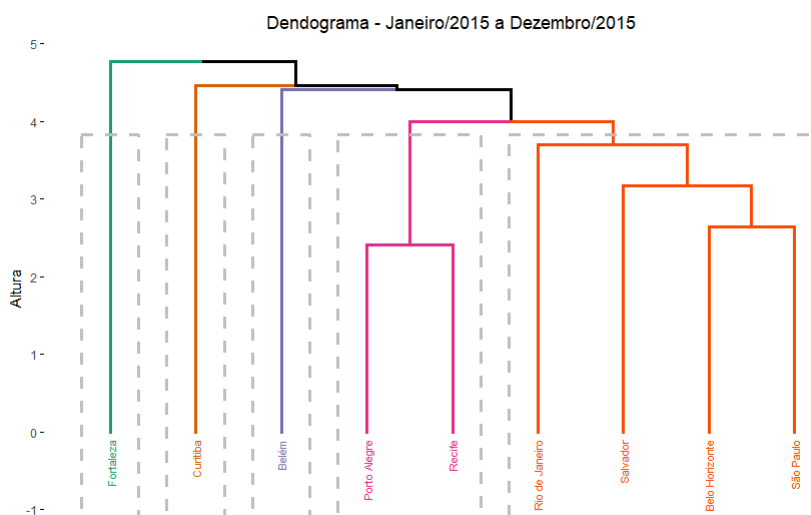


Figura 3.23: Dendrograma referente ao período de Janeiro a Dezembro de 2015 para as Regiões analisadas segundo o IPCA-15.

Comparando os resultados desta Figura com os resultados obtidos para o ano de 2014, é possível verificar a formação de 5 grupos, porém, impulsionadas pela acentuação da crise econômica no país, não houve índices médios deflacionários neste período, bem como, em todas as Regiões, verificou-se um aumento significativo nos índices de inflação das categorias analisadas.

O primeiro grupo é formado pela Região de Fortaleza, que apresentou no setor de Despesas Pessoais o mais alto de seus índices inflacionários (1,098%), que também foi o maior comparado às demais Regiões nesta categoria, além de ser também detentora do maior valor inflacionário da categoria de transportes (1,052%). Em contrapartida, esta Região teve no setor de Comunicação o menor crescimento médio inflacionário, de valor 0,096%, o menor de todas as categorias e

regiões analisadas, além de ter tido também o menor valor inflacionário do setor de Vestuário (0,225%).

O segundo grupo é constituído pela Região de Curitiba, que apresentou o maior índice médio inflacionário dentre as Regiões do estudo quando todas as categorias são analisadas em conjunto, com um crescimento médio de 0,834% por categoria de consumo analisada. Nesta Região, a categoria de Habitação exibiu o mais alto índice dentre todas as categorias e Regiões analisadas, de valor 1,747%, além de mostrar os maiores valores para as categorias de Saúde e Cuidados Pessoais (0,808%) e Vestuário (0,642%). Por outro lado, Curitiba também apresentou o menor índice inflacionário na categoria de Comunicação, de valor 0,220%.

No terceiro grupo encontra-se a Região de Belém, que semelhante à Curitiba e Fortaleza, apresentou no setor de Habitação o mais alto índice inflacionário médio, de 1,035%, e o menor, na categoria de Comunicação, de 0,311%. É interessante destacar também que esta Região apresentou o menor valor inflacionário médio da categoria de Educação, com um índice de 0,537% registrado no ano de 2015.

Porto Alegre e Recife formam juntos o quarto cluster gerado no período analisado em 2015. Seu comportamento assemelha-se no fato de apresentarem valores altos do índice de inflação na categoria de Alimentos e Bebidas, de valores 1,114% e 1,067%, respectivamente. Este índice foi o maior da Região de Recife, e o de Porto Alegre o maior desta categoria.

Por fim, o último grupo é formado pelas quatro maiores Regiões do país: Rio de Janeiro, Salvador, Belo Horizonte e São Paulo. Nestas Regiões é possível verificar índices inflacionários altos nas categorias de Alimentação e Habitação, e baixos na categoria de Comunicação.

O coeficiente de correlação cofenética apresentou valor de 0,7840, o que indica que o dendrograma apresentado, seguindo o método da Ligação Média, descreve de forma válida o comportamento e distribuição dos dados nele apresentados.

3.3.3 JANEIRO A AGOSTO DE 2016

Nesta seção será apresentado o dendrograma referente ao último período analisado, compreendido entre Janeiro e Agosto de 2016, segundo às Regiões citadas no estudo e discriminadas pelas categorias de consumo, que encontram-se resumidos na Figura 3.24.

É interessante verificar que no primeiro semestre deste ano, observam-se mudanças no mapa da inflação do país, com uma queda nos valores médios, quando comparados ao ano anterior, e o setor de Habitação apresentando índices mais baixos do que nos demais períodos analisados.

O primeiro cluster é formado pela Região de Recife, que apresentou o maior de seus índices na categoria de Saúde e Cuidados Pessoais, ao destacar um aumento médio mensal de 1,113% no primeiro semestre de 2016. Em contrapartida, esta Região se destaca por apresentar o menor índice médio inflacionário na categoria de Comunicação (0,040%), sendo este também, o menor dentre as Regiões analisadas neste setor. Recife também apresentou o menor índice dentre as Regiões na categoria de Despesas Pessoais (0,495%), e o mais baixo dos setores de Alimentação (0,986%) e Educação (0,574%).

O segundo grupo produzido engloba as Regiões de Fortaleza e Rio de Janeiro, que apesar

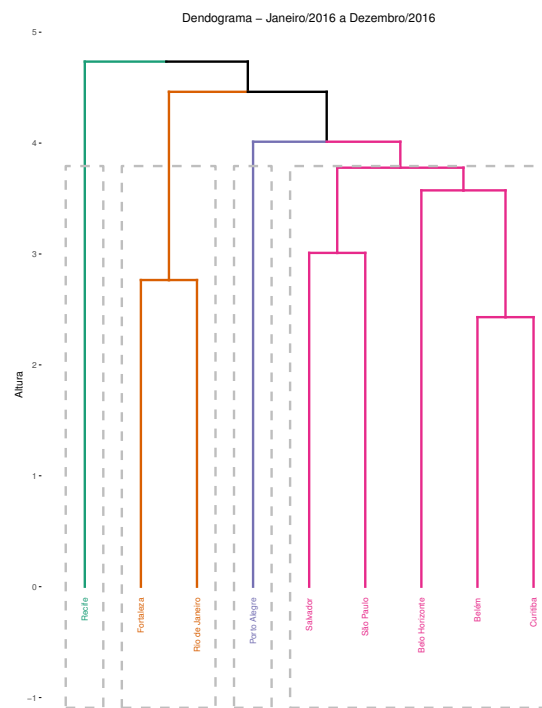


Figura 3.24: Dendrograma referente ao período de Janeiro a Agosto de 2016 para as Regiões analisadas segundo o IPCA-15.

de apresentarem tamanho populacional completamente diferente, se comportaram de modo análogo de Janeiro a Agosto do ano de 2016.

Porto Alegre, capital do Rio Grande do Sul, compõe o terceiro grupo formado pelo dendrograma, que teve na categoria de Educação o maior de seus índices médios inflacionários, de valor 1,243%. Por outro lado, esta Região apresentou o mais baixo valor na categoria de Comunicação, com um crescimento médio mensal de valor 0,249%.

O quarto cluster é composto pelas Regiões de Salvador e São Paulo, que apresentaram altos índices na categoria de Saúde e Cuidados Pessoais, de valor 1,216% e 1,153%, respectivamente. Estas Regiões tiveram o menor de seus índices médios na categoria de Comunicação, muito similares, de valor 0,076% e 0,078%, respectivamente.

Adicionalmente, neste grupo encontram-se as Regiões de Belo Horizonte, Belém e Curitiba localizadas no último grupo formado. Todas elas apresentaram, na categoria de Alimentação e Bebidas, índices altos, de valor 1,443%, 1,276% e 1,138%, respectivamente. O coeficiente de correlação cofenética de valor 0,6562 indica que o dendrograma explica bem o comportamento e a formação dos grupos gerados por este gráfico.

4. CONCLUSÕES

Em virtude dos fatos mencionados, conclui-se que o comportamento do IPCA-15 atua de maneira específica entre as categorias e Regiões analisadas. Comportamentos sazonais, como no caso da categoria de Educação, presente em quase todas as Regiões, ilustram que, em certos períodos do ano, há predominância de crescimento ou decréscimo dos indicadores. As categorias de Educação, Habitação e Alimentação, destacaram-se no estudo, ao apresentar os maiores valores médios inflacionários. Em contrapartida, o setor de Comunicações apresentou tendências deflacionárias ao longo de todo o período estudado, exceto no ano de 2015, refletido pela Análise de Agrupamentos, devido ao impacto do início da crise econômica no Brasil. Esta técnica também mostrou não existir um comportamento padrão entre as Regiões analisadas, como enfatizado ao início deste parágrafo, devido as características específicas de cada localidade.

Finalmente, a Análise de Componentes Principais apontou que os três primeiros Componentes foram responsáveis por cerca de 75,3% da variabilidade total dos dados. Destes, as categorias de Despesas Pessoais, Transportes e Comunicação apresentaram as maiores correlações com este Componente, esta última de maneira negativa. Já o segundo Componente Principal relaciona-se principalmente, e de forma positiva, com as categorias de Alimentação e Habitação. Verificou-se também que os setores de Comunicação e Despesas Pessoais apresentaram alta qualidade de representação para os dois primeiros Componentes Principais sendo peças-chave na explicação da variabilidade do conjunto de dados.

No Primeiro Componente Principal, responsável por explicar cerca de 35,9% da variabilidade total dos dados, a Região de Curitiba apresentou uma relação positiva e forte com as variáveis Alimentação, Habitação, Saúde, Educação, Vestuário e Comunicação, ou seja, um aumento nos índices inflacionários destas categorias impactou diretamente os índices desta Região quando comparados as demais. As Regiões de Porto Alegre, Salvador, São Paulo e Rio de Janeiro também mostraram relação positiva com estas categorias, mas de maneira menos significativa. Por outro lado, as Regiões de Fortaleza, Belém, Recife e Belo Horizonte apresentaram correlações negativas com estas categorias, indicando que o aumento nos índices inflacionários das mesmas não é sentido de modo impactante nestas Regiões quando comparadas as demais. Já para as categorias de Despesas Pessoais, Transportes e Artigos Pessoais, a conclusão torna-se inversa, ou seja, enquanto Belém, Recife, Belo Horizonte e Fortaleza possuem uma correlação positiva com estas categorias, as Regiões do Rio de Janeiro, São Paulo, Porto Alegre e Salvador relacionam-se de forma negativa.

Para o Segundo Componente Principal, que responde por de cerca de 21,1% da variabilidade total dos dados, as Regiões de São Paulo, Rio de Janeiro e Fortaleza, relacionaram-se de modo positivo com as categorias de Transporte, Despesas Pessoais, Saúde, Educação, Habitação e Alimentação. Já Porto Alegre, Salvador e Curitiba possuem uma fraca relação com todas as categorias no Segundo Componente Principal. Finalmente, Belém, Recife e Belo Horizonte apresentaram uma correlação negativa com estas categorias, ou seja, estas Regiões não sofreram impacto direto no aumento dos índices inflacionários destes setores como as demais Regiões.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ABDI H., H. e WILLIAMS, L.: *Principal component analysis*. Wiley interdisciplinary reviews: computational statistics, 2(4):433–459, 2010. <https://onlinelibrary.wiley.com/doi/10.1002/wics.101>.
- [2] ARTES, R.: *Aspectos estatísticos da análise fatorial de escalas de avaliação*. Revista de Psiquiatria Clínica, 25(5):223–228, 1998. <http://www.hcnet.usp.br/ipq/revista/r255/conc255d.htm>.
- [3] BUSSAB, W.O.; MIAZAKI, S.E. e ANDRADE, D.F.: *Introdução à análise de agrupamentos*. IME/USP, 1990.
- [4] CARRARA, A. e CORREA, A.: *O regime de metas de inflação no Brasil: uma análise empírica do IPCA*. Rev. econ. contemp., 16(3):441–462, 2012. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-98482012000300004&lng=en&nrm=iso.
- [5] CRUZ, C.: *Aplicação de algumas técnicas multivariadas no melhoramento de plantas*. Dissertação de Doutorado, 1990.
- [6] DANTAS, G.: *Inflação e indexação*, 2017. <http://brasilecola.uol.com.br/politica/inflacao-indexacao.htm>, Acesso em 18 out. 2017.
- [7] DZIUBAN, C. e SHIRKEY, E.: *When is a correlation matrix appropriate for factor analysis? Some decision rules*. Psychological Bulletin, 81(6):358–361, 1974. <http://dx.doi.org/10.1037/h0036316>.
- [8] FAVERO, L.P.; BELFIORE, P.S.F.L. e CAHN, B.L.: *Análise de dados – modelagem multivariada para tomada de decisão*. Ed. Elsevier Ltda, 8ª ed., 2009.
- [9] FLECK, M. e BOURDEL, M.: *Método de simulação e escolha de fatores na análise dos principais componentes*. Revista de Saúde Pública, 32(3):267–272, 1998. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89101998000300010&lng=en&nrm=iso.
- [10] IBGE, 2017. <https://www.ibge.gov.br/estatisticas-novoportal/economicas/precos-e-custos/9260-indice-nacional-de-precos-ao-consumidor-amplo-15.html>, Consulta realizada em 18/10/2017.

- [11] IPEA, 2017. <http://www.ipea.gov.br>, Consulta realizada em 10/10/2017.
- [12] JOHNSON, R. e WICHERN, D.: *Applied multivariate statistical analysis*. Englewood Cliffs: Prentice-hall, 1982.
- [13] JOLLIFFE, I.: *Principal Component Analysis*. Springer-Verlag, 2ª ed., 2002.
- [14] KAISER, H.: *The varimax criterion for analytic rotation in factor analysis*. *Psychometrika*, 23:187–200, 1958. https://www.psychometricsociety.org/sites/default/files/kaiser_citation_classic_varimax.pdf.
- [15] KASSAMBARA, A. e MUNDT, F.: *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*, 2017. <https://CRAN.R-project.org/package=factoextra>.
- [16] KENDALL, M.: *Part I: Factor Analysis as a Statistical Technique*. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(1):60–73, 1950. <https://www.jstor.org/stable/2980947>.
- [17] KUBRUSLY, L.: *Um procedimento para calcular índices a partir de uma base de dados multivariados*. *Pesquisa Operacional*, 21(1):107–117, 2001. <http://www.scielo.br/pdf/pope/v21n1/a07v21n1>.
- [18] LIBERATO J.R., J. R.: *Aplicações de técnicas de análise multivariada em fitopatologia*. Dissertação de Mestrado, 1995.
- [19] LÊ, S.; JOSSE, J. e HUSSON, F.: *FactoMineR: A Package for Multivariate Analysis*. *Journal of Statistical Software*, 25(1):1–18, 2008. <https://www.jstatsoft.org/v025/i01>.
- [20] MARDIA, K.V.; KEMT, J. T. e BIBBY, J.M.: *Multivariate analysis*. London: Academic press, 1989.
- [21] MARTINEZ, T. e CERQUEIRA, V.: *Estrutura da inflação brasileira: determinantes e desagregação do IPCA*. *Economia e sociedade*, 22(2):409–456, 2013. www.eco.unicamp.br/docprod/downarq.php?id=3296&tp=a.
- [22] MILAGRE, R.: *Estatística: uma proposta de ensino para os cursos de administração de empresas*. Dissertação de Mestrado, 2001.
- [23] MORRISSON, D.: *Multivariate statistical method*. McGraw-Hill, McGraw-Hill series in probability and statistics, 2ª ed., 1976.
- [24] PERES-NETO, P. J. D. e K.M., S.: *How Many Principal Components? Stopping Rules for Determining the Number of Non-Trivial Axes Revisited*. *Computational Statistics Data Analysis*, 49:974–997, 2005. <http://labs.eeb.utoronto.ca/jackson/computational%20statistics%20and%20data%20analysis.pdf>.

- [25] REVELLE, W.: *psych: Procedures for Psychological, Psychometric, and Personality Research*, 2018. <https://CRAN.R-project.org/package=psych>.
- [26] SILVA, N. e PADOVANI, C.: *Utilização de componentes principais em experimentação agrônômica*. *Energia na Agricultura*, 21(4):98–113, 2006. http://www.fca.unesp.br/CD_REVISTA_ENERGIA_vol8/vol21n42006/artigos/Nilza%20Regina%20da%20Silva.pdf.
- [27] TEAM, R.: *RStudio: Integrated Development Environment for R*, 2015. <http://www.rstudio.com/>.
- [28] TEAM, R. C.: *R: A Language and Environment for Statistical Computing*, 2013. <http://www.R-project.org/>.
- [29] WEY, T. e SIMKO, V.: *R package "corrplot": Visualization of a Correlation Matrix*, 2017. <https://github.com/taiyun/corrplot>.
- [30] WICKHAM, H.: *Reshaping Data with the reshape Package*. *Journal of Statistical Software*, 21(12):1–20, 2007. <http://www.jstatsoft.org/v21/i12/>.
- [31] WICKHAM, H.: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- [32] WICKHAM, H.: *tidyverse: Easily Install and Load the 'Tidyverse'*, 2017. <https://CRAN.R-project.org/package=tidyverse>.