

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

Faculdade de Ciências e Tecnologia - Campus de Presidente Prudente

NICOLLE RYE FUKUSHIMA DOS SANTOS

**ANÁLISE DO CONSUMO DOMICILIAR DE PESCADOS NO ESTADO DE SÃO
PAULO A PARTIR DA POF (2017-2018) UTILIZANDO TÉCNICAS DE ESTATÍSTICA
MULTIVARIADA E GEOESPACIAL**

Presidente Prudente

2025

Nicolle Rye Fukushima dos Santos

**ANÁLISE DO CONSUMO DOMICILIAR DE PESCADOS NO ESTADO DE SÃO
PAULO A PARTIR DA POF (2017-2018) UTILIZANDO TÉCNICAS DE ESTATÍSTICA
MULTIVARIADA E GEOESPACIAL**

Trabalho de Conclusão de Curso, apresentado à
Universidade Estadual Paulista "Júlio de Mes-
quita Filho" (UNESP), Faculdade de Ciências e
Tecnologia, Presidente Prudente, para obtenção
do título de Bacharel em Estatística.

Orientador: Profº Dr. Edilson Ferreira Flores
Coorientador: Dr. Ricardo Firetti

Presidente Prudente
2025

D724a

dos Santos, Nicolle Rye Fukushima

Análise do consumo domiciliar de pescados no estado de são paulo a partir da pof
(2017 - 2018) utilizando técnicas de estatística multivariada e geoespacial / Nicolle Rye
Fukushima dos Santos. -- Presidente Prudente, 2025

114 p. : il., tabs.

Trabalho de conclusão de curso (Bacharelado - Estatística) - Universidade Estadual
Paulista (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente

Orientador: Edilson Ferreira Flores

Coorientador: Ricardo Firetti

1. Estatística. 2. Consumo de Pescados. 3. Análise Multivariada. 4. Análise Espacial.
I. Título.

NICOLLE RYE FUKUSHIMA DOS SANTOS

ANÁLISE DO CONSUMO DOMICILIAR DE PESCADOS NO ESTADO DE SÃO PAULO A PARTIR DA POF (2017-2018) UTILIZANDO TÉCNICAS DE ESTATÍSTICA MULTIVARIADA E GEOESPACIAL:

Trabalho de Conclusão de Curso apresentado à Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), Faculdade de Ciências e Tecnologia, Presidente Prudente, para obtenção do título de Bacharel em Estatística.

Data de defesa: 28/11/2025

BANCA EXAMINADORA

Documento assinado digitalmente
gov.br EDILSON FERREIRA FLORES
Data: 10/12/2025 11:00:01-0300
Verifique em <https://validar.iti.gov.br>

Profº Dr. Edilson Ferreira Flores
Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), Faculdade de Ciências e Tecnologia, Câmpus de Presidente Prudente, Departamento de Estatística



Dr. Ricardo Firetti
Agência Paulista de Tecnologia dos Agronegócios (APTA), Departamento de Descentralização do Desenvolvimento, Unidade Regional de Pesquisa e Desenvolvimento – Presidente Prudente

Documento assinado digitalmente
gov.br SERGIO MINORU OIKAWA
Data: 09/12/2025 08:55:13-0300
Verifique em <https://validar.iti.gov.br>

Profº Dr. Sérgio Minoru Oikawa
Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP), Faculdade de Ciências e Tecnologia, Câmpus de Presidente Prudente, Departamento de Estatística

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brasil. Processo nº 2023/17008-6. As opiniões, hipóteses e conclusões ou recomendações expressas neste material são de responsabilidade do(s) autor(es) e não necessariamente refletem a visão da FAPESP.

Agradeço, primeiramente, à FAPESP, pelo apoio financeiro concedido por meio do Processo nº 2023/17008-6, da Fundação de Amparo à Pesquisa do Estado de São Paulo, que tornou possível o desenvolvimento deste trabalho e proporcionou condições fundamentais para minha formação científica.

À minha família, pelo amor incondicional, paciência e apoio constante. Cada palavra de incentivo e cada gesto de cuidado foram essenciais para que eu persistisse em todos os desafios acadêmicos e pessoais ao longo desta jornada.

Aos meus amigos, que compartilharam comigo momentos de aprendizado, leveza e força. Agradeço pela compreensão durante os períodos mais intensos da graduação e pela companhia que tornou o percurso mais leve.

Ao meu orientador e coorientador, pela confiança, dedicação e por cada orientação que ampliou meu olhar científico. Sou profundamente grata pela parceria no desenvolvimento deste trabalho, pelas discussões construtivas e pelo comprometimento com a minha formação.

Aos meus professores da graduação, que contribuíram de forma significativa para minha trajetória acadêmica, transmitindo não apenas conhecimento técnico, mas também valores essenciais para a vida profissional e para a pesquisa.

A todos que, direta ou indiretamente, contribuíram para a realização deste Trabalho de Conclusão de Curso, deixo aqui o meu sincero agradecimento.

RESUMO

Esta pesquisa teve como objetivo analisar os padrões de aquisição e consumo domiciliar de pescados no Brasil, com ênfase no estado de São Paulo, utilizando os microdados da Pesquisa de Orçamentos Familiares (POF 2017–2018) disponibilizados pelo IBGE. Apesar da ampla disponibilidade hídrica e do extenso litoral brasileiro, o consumo per capita de pescados permaneceu abaixo das recomendações da Organização das Nações Unidas para a Alimentação e Agricultura (FAO). Em regiões densamente povoadas como São Paulo, observou-se um consumo ainda mais reduzido, mesmo diante do avanço da produção aquícola nacional. A investigação considerou a influência de variáveis socioeconômicas, tais como idade, escolaridade, localização urbana ou rural, composição familiar e faixa de renda, sobre o padrão de consumo de diferentes tipos de pescados. As análises foram conduzidas no software R, por meio de estatística descritiva (médias, frequências e variâncias) e multivariada, incluindo alfa de Cronbach, análise de componentes principais, análise de correspondência, agrupamentos hierárquicos e bidimensionais. Adicionalmente, incorporaram-se técnicas de estatística geoespacial para explorar a dimensão territorial do consumo de pescados, permitindo mapear padrões espaciais de aquisição alimentar, identificar concentrações e disparidades regionais e relacionar esses resultados a características socioeconômicas e de infraestrutura municipal. Espera-se que os achados contribuam para a compreensão das desigualdades territoriais no acesso e na preferência por produtos da pesca e da aquicultura no estado de São Paulo.

Palavras-chave: consumo alimentar; pescado; desigualdade socioeconômica; estatística multivariada; estatística geoespacial; POF 2017-2018.

ABSTRACT

This research aimed to analyze household acquisition and consumption patterns of fish in Brazil, with emphasis on the state of São Paulo, based on the microdata from the Consumer Expenditure Survey (POF 2017–2018) made available by the Brazilian Institute of Geography and Statistics (IBGE). Despite the country's abundant water resources and extensive coastline, Brazil's per capita fish consumption remains below the recommendations of the Food and Agriculture Organization of the United Nations (FAO). In densely populated regions such as São Paulo, this consumption is even lower, despite the national growth of aquaculture production. The study investigated the influence of socioeconomic variables — such as age, education, urban or rural location, presence of children and adolescents in the household, and income level — on the consumption patterns of different types of fish (freshwater, saltwater, prepared, fresh, frozen, among others). The analysis was conducted by constructing databases in the R software, using descriptive statistical techniques (means, frequencies, variances) and multivariate methods (Cronbach's alpha, principal component analysis, correspondence analysis, and hierarchical and bidimensional clustering). In addition, the research incorporated geospatial statistical techniques to explore the territorial dimension of fish consumption. This approach enabled the mapping of spatial patterns of food acquisition, the identification of regions with low or high consumption intensity, and the relationship of these results with the local socioeconomic and infrastructure context. The study aims to contribute to a broader understanding of territorial inequalities in access to and preference for fishery and aquaculture products.

Keywords: food consumption; fish; socioeconomic inequality; multivariate statistics; geospatial statistics; POF 2017–2018.

LISTA DE ILUSTRAÇÕES

Figura 1	Consumo de Pescado por Local do Domicílio no Estado de SP – POF (2017–2018)	55
Figura 2	Consumo de Pescado por Faixa de Renda Familiar no Estado de SP – POF (2017–2018)	56
Figura 3	Consumo de Pescado por Ocupação no Estado de SP – POF (2017–2018)	57
Figura 4	Consumo de Pescado por Composição Familiar no Estado de SP – POF (2017–2018)	58
Figura 5	Consumo de Pescado por Padrão de Alimentação no Estado de SP – POF (2017–2018)	59
Figura 6	Consumo de Pescado por Padrão de Saúde no Estado de SP – POF (2017–2018)	60
Figura 7	Consumo de Pescado por Prática de Dieta no Estado de SP – POF (2017–2018)	61
Figura 8	Scree Plot dos Autovalores para Determinação do Número de Fatores – SP	64
Figura 9	Análise de Correspondência — Fator 1 (Infraestrutura Urbana e Serviços Públicos) X Consumo de Pescado	68
Figura 10	Análise de Correspondência — Fator 2 (Consumo e Bem-Estar Doméstico) X Consumo de Pescado	68
Figura 11	Análise de Correspondência — Fator 3 (Condição Socioeconômica) X Consumo de Pescado	69
Figura 12	Análise de Correspondência — Fator 4 (Saneamento e Serviços Essenciais) X Consumo de Pescado	69
Figura 13	Análise de Correspondência — Fator 5 (Idade) X Consumo de Pescado . .	70
Figura 14	Análise de Correspondência — Padrão da Alimentação X Consumo de Pescado	71
Figura 15	Análise de Correspondência — Composição Familiar X Consumo de Pescado	72
Figura 16	Análise de Correspondência — Ocupação X Consumo de Pescado	72
Figura 17	Análise de Correspondência — Renda Total X Consumo de Pescado . . .	73
Figura 18	Análise de Correspondência — Padrão de Saúde X Consumo de Pescado .	73
Figura 19	Análise de Correspondência Múltipla — Padrão de Saúde, Alimentação e Consumo de Pescado	75
Figura 20	Análise de Correspondência Múltipla — Renda Familiar, Escolaridade e Consumo de Pescado	76
Figura 21	Dendograma a Partir dos Fatores Latentes	78
Figura 22	Perfil Médio dos Scores Fatoriais por Cluster	79

Figura 23	Média + Intervalo de Confiança do Consumo Domiciliar de Pescados no Estado de SP dos Clusteres dos Fatores	80
Figura 24	Dendograma a Partir das Variáveis Ordinais	83
Figura 25	Perfil Médio das Variáveis Ordinais por Cluster	85
Figura 26	Média + Intervalo de Confiança do Consumo Domiciliar de Pescados no Estado de SP dos Clusteres das Variáveis Ordinais	86
Figura 27	Two-Way-Joining das Variáveis Ordinais X Consumo de Pescado	89
Figura 28	Porcentagem de Domicílios que Consomem Alimentos da Aquicultura no Estado de SP	91
Figura 29	Centro Médio e Distribuição Direcional da Produção Aquícola (kg) – Municípios do Estado de SP	92
Figura 30	Hot Spot da Produção Aquícola (kg) – Municípios do Estado de SP	93
Figura 31	Índice de Moran Local (LISA) da Produção Aquícola (kg) – Municípios do Estado de SP	95
Figura 32	Centro Médio e Distribuição Direcional do Valor da Produção Aquícola em Reais – Municípios do Estado de SP	97
Figura 33	Hot Spot do Valor da Produção Aquícola em Reais – Municípios do Estado de SP	98
Figura 34	Índice de Moran Local (LISA) do Valor da Produção Aquícola em Reais – Municípios do Estado de SP	100
Figura 35	Centro Médio e Distribuição Direcional do Número de Estabelecimentos Aquícolas – Municípios do Estado de SP	102
Figura 36	Centro Médio e Distribuição Direcional do Número de Estabelecimentos Aquícolas – Municípios do Estado de SP	103
Figura 37	Hot Spot do Número de Estabelecimentos Aquícolas – Municípios do Estado de SP	105
Figura 38	Índice de Moran Local (LISA) do Número de Estabelecimentos Aquícolas – Municípios do Estado de SP	107

LISTA DE TABELAS

Tabela 1	– Variáveis qualitativas ordinais do Estado de SP – POF (2017–2018)	48
Tabela 2	– Variáveis qualitativas nominais do Estado de SP – POF (2017–2018)	49
Tabela 3	– Análise descritiva das variáveis qualitativas ordinais – SP (POF 2017–2018)	50
Tabela 4	– Análise descritiva das variáveis qualitativas nominais – SP (POF 2017–2018)	53
Tabela 5	– Autovalores e Variância por Fator – SP	63
Tabela 6	– Cargas Fatoriais – Análise Fatorial (PCA com rotação varimax) – SP	65
Tabela 7	– Fatores Latentes Extraídos para o Estado de São Paulo	66
Tabela 8	– Estatísticas Descritivas por Cluster para Renda_Total – SP	81
Tabela 9	– Estatísticas Descritivas por Cluster para Peixe_geral – SP	81
Tabela 10	– Estatísticas Descritivas por Cluster para Anos_Estudo – SP	81
Tabela 11	– Teste de Kruskal–Wallis para diferenças no consumo de pescado entre clusters	82
Tabela 12	– Estatísticas Descritivas por Cluster para Renda_Total – SP	87
Tabela 13	– Estatísticas Descritivas por Cluster para Peixe_geral – SP	87
Tabela 14	– Estatísticas Descritivas por Cluster para Anos_Estudo – SP	88
Tabela 15	– Teste de Kruskal–Wallis para diferenças no consumo de pescado entre clusters	88

SUMÁRIO

1	INTRODUÇÃO	13
2	OBJETIVOS	16
2.1	Objetivo Geral	16
2.2	Objetivos Específicos	16
3	JUSTIFICATIVA	18
4	REFERENCIAL TEÓRICO E ESTADO DA ARTE	19
4.1	Análise Multivariada	19
4.2	Estatística Espacial e Padrões Geográficos	19
4.3	Consumo Alimentar e Aplicações ao Contexto Paulista	20
5	METODOLOGIA	21
5.1	Ambiente de execução e recursos utilizados	21
5.2	Caracterização geral da pesquisa	21
5.3	Coleta, integração e estruturação dos dados	21
5.4	Análise descritiva	22
5.5	Análise Multivariada	22
5.6	Elaboração dos Mapas Temáticos	23
6	TÉCNICAS UTILIZADAS	24
6.1	Alfa de Cronbach	24
6.1.1	Introdução à Confiabilidade e ao Alfa de Cronbach	24
6.1.2	Fundamentos da Correlação Policórica	25
6.1.3	Interpretação da Correlação Policórica	26
6.1.4	Alfa de Cronbach com Correlação Policórica	26
6.2	Análise Fatorial por Componentes Principais com Correlação Policórica	26
6.2.1	Introdução à Análise Fatorial	27
6.2.2	Fundamentos Matemáticos da PCA	27
6.2.3	Rotação Varimax: Teoria e Justificativa	28
6.2.4	Etapas da Análise Fatorial com Matriz Policórica	28
6.3	Análise de Correspondência	29
6.3.1	Estrutura Matemática da Análise de Correspondência Simples	30
6.3.2	Interpretação e Representação Gráfica	31
6.3.3	Associações com a Análise Fatorial	31
6.3.4	Análise de Correspondência Múltipla (ACM)	31

6.4	Análise de Agrupamento Hierárquico	32
6.4.1	Introdução à análise de agrupamento hierárquico	33
6.4.2	Etapas do procedimento	33
6.4.3	Determinação do número de clusters	36
6.4.4	Interpretação dos resultados de agrupamento	37
6.5	Two-Way Joining	37
6.5.1	Fundamentos Matemáticos do Two-Way Joining	37
6.5.2	Seriação e Algoritmos de Ordenação no Two-Way Joining	38
6.5.3	Aplicação do TWJ com Séries Ordinais Normalizadas	40
6.6	Análise Espacial	41
6.6.1	Centro Médio (Mean Center)	41
6.6.2	Distribuição Direcional e Elipse de Desvio-Padrão	42
6.6.3	Índice de Moran Local (LISA)	42
6.6.4	Análise de Hot Spot (Getis-Ord Gi*)	44
6.6.5	Integração com Variáveis Exógenas da Aquicultura	45
7	RESULTADOS	46
7.1	Preparação da Base de Dados	46
7.1.1	Filtragem e identificação dos indivíduos	46
7.1.2	Seleção de variáveis relevantes	46
7.1.3	Classificação dos alimentos e criação de indicadores	46
7.1.4	Integração e criação de novas variáveis	47
7.2	Descrição das Variáveis Utilizadas	47
7.2.1	Variáveis Qualitativas Ordinais	47
7.2.2	Variáveis Qualitativas Nominais	48
7.3	Análise Descritiva	49
7.3.1	Tabelas Descritivas	49
7.3.2	Gráficos Descritivos	54
7.4	Avaliação de Consistência Interna – Alfa de Cronbach	61
7.5	Análise Fatorial	62
7.5.1	Medida de Adequação da Amostragem (KMO)	62
7.5.2	Seleção do Número de Fatores e Variância Explicada	62
7.6	Análise de Correspondência	67
7.6.1	Gráficos dos Fatores Latentes e Consumo de Pescado	67
7.6.2	Gráficos das Variáveis Socioeconômicas e Consumo de Pescado	71
7.6.3	Análise de Correspondência Múltipla	74
7.7	Análise de Cluster	77
7.7.1	Análise de Cluster Através dos Fatores	77
7.7.2	Análise de Cluster Através das Variáveis Ordinais	83

7.8	Two-Way-Joining	89
7.9	Análise Espacial da Proporção de Consumo de Pescado com Variáveis Exógenas da Aquicultura	90
7.9.1	Proporção de Consumo de Pescado entre as Regiões Administrativas dadas na POF	91
7.9.2	Quantidade Produzida de Produtos Aquícolas (kg)	92
7.9.2.1	Centro Médio e Distribuição Direcional	92
7.9.2.2	Hot Spot (G_i^*)	93
7.9.2.3	Índice de Moran Local (LISA)	95
7.9.3	Valor da Produção Aquícola (R\$)	97
7.9.3.1	Centro Médio e Distribuição Direcional	97
7.9.3.2	Hot Spot (G_i^*)	98
7.9.3.3	Índice de Moran Local (LISA)	100
7.9.4	Número de Estabelecimentos Aquícolas	102
7.9.4.1	Plot da Variável	102
7.9.4.2	Centro Médio e Distribuição Direcional	103
7.9.4.3	Hot Spot (G_i^*)	105
7.9.4.4	Índice de Moran Local (LISA)	107
8	CONCLUSÃO	109
	REFERÊNCIAS	112

1 INTRODUÇÃO

Segundo a Pesquisa de Orçamentos Familiares 2017-2018 (IBGE, 2020), do Instituto Brasileiro de Geografia e Estatística, nosso consumo domiciliar *per capita* anual de pescados foi estimado em 2,8 kg. Nos Estados do Amazonas, Acre, Amapá, Pará e Maranhão estes valores *per capita* estariam entre 9,5 e 14 kg.

A Região Norte e parte do Nordeste têm índices de consumo domiciliar *per capita* muito maiores que os demais entes da federação, mas com elevada população numerosa. Os Estados de São Paulo e Minas Gerais abrigam 33% da população do país e têm valores de consumo *per capita* extremamente baixos, tanto de produtos de água salgada como de água doce.

Com exceção de alguns Estados brasileiros, os números de consumo *per capita*, embora domiciliar, estão distantes dos 10,5 kg observados na América Latina e no Caribe ou dos 12 kg recomendados pela FAO (2020). De acordo com a instituição, o consumo *per capita* global de pescado vem aumentando cerca de 1,5% ao ano, saltando de 9 kg/habitante, em 1961, para 20,5 kg/habitante em 2018. Segundo Wagner et al. (2023), analisando microdados da POF 2017-2018, o consumo de proteína proveniente de organismos aquáticos no Brasil está aquém do recomendável e, além disso.

Firetti et al. (2014), trabalhando com os dados da Pesquisa de Orçamentos Familiares 2008, observaram a menor participação de pescados na alimentação domiciliar do brasileiro quando comparado a outras carnes (bovinos, suínos e aves) e ovos, assim como Siqueira et al. (2023). Essa diferença aumentou quando se analisaram as regiões Sudeste, Sul e Centro-Oeste. Para os primeiros autores, o Estado de São Paulo se destacou dos demais pelo consumo de pescados em filés frescos e de bacalhau.

A POF 2017-2018 (IBGE, 2020) aponta diferenças na escolha da origem do pescado, havendo maior consumo de produtos de água salgada em relação aos de água doce no Estado de São Paulo. Essa questão também foi apontada por Firetti et al. (2014), pois o consumo de pescado de água salgada teria relação direta com a proximidade do litoral.

Estudos sobre o comportamento de consumo de pescados são mundialmente tradicionais, inclusive no Brasil. Neste sentido, pode-se mencionar os trabalhos de Bernardo (2017); Freitas et al. (2020); Krešić et al. (2020); Dai et al. (2022) e Stancu et al. (2022). No Brasil, destacam-se os recentes esforços da EMBRAPA Pesca e Aquicultura com os estudos de Flores (2020); Pedroza Filho et al. (2020) e Flores et al. (2022); além de outros estudos que buscaram compreender o consumo de pescados e seus determinantes: Araujo et al., 2018; Fornari et al., 2017 e Barbosa & Sampaio, 2016.

De modo geral, Wagner et al. (2023) mostraram que o consumo apresenta padrões variados quando se consideram fatores como preço, nível de renda e localização geográfica. Em contrapartida, há importante lacuna sobre o comportamento de consumo de pescados de famílias com menores níveis de rendimento. Conforme apontado pelos autores, o consumo de pescados pode

ser afetado por diversas variáveis socioeconômicas, incluindo renda, localização da residência, estilo de vida, nível educacional, preocupação com a saúde, hábitos alimentares, condições de saúde e características individuais.

Historicamente, o Brasil é importador de pescados e ainda engatinha como fornecedor no mercado internacional. Em 2021, os principais produtos de pescado importados estavam relacionados ao salmão e pangásius. Há déficit histórico na balança comercial do setor pescados (BRASIL, 2022), entretanto a cadeia produtiva da tilápia vem conquistando espaço no mercado internacional, especialmente com a comercialização de peixes inteiros e congelados.

Assim, há oportunidade mercadológica de a piscicultura suprir parte da demanda nacional via “substituição de produtos” de pescados importados ou originados na captura, pelos peixes cultivados tanto em águas interiores como marinhas. Isto poderia ocorrer via aumento da produção de espécies (truta salmonada e pangásius), ou a oferta e sugestão de produtos com características semelhantes àqueles no que tange a características tais como: o tipo de corte (peixe inteiro, filés e postas), forma de venda (fresco ou congelado), formas de preparo (grelhado, assado, ensopado, empanado).

Em estudos recentes sobre a viabilidade econômica e aspectos produtivos de peixes nativos da Amazônia, Dantas Filho et al. (2021) demonstraram o potencial econômico da piscicultura de tambaqui, comparativamente a outras atividades agropecuárias daquela região. Para isso os autores utilizaram a taxa interna de retorno (TIR), *payback* e rentabilidade, apresentando uma análise das vantagens e desvantagens da piscicultura. Barros et al. (2020) também verificaram no Mato Grosso que “a atividade se mantém a longo prazo, pois saldou totalmente todas despesas e depreciação, e parcialmente a remuneração dos custos de oportunidade”.

Segundo Valenti et al. (2021), o Brasil possui mais de 200 mil pisciculturas de água doce, cerca de 3.000 fazendas de camarão marinho e por volta de 100 Instituições de Pesquisa em Aquicultura, sendo que um grande mercado interno está disponível para peixes e mariscos comestíveis, ornamentais, isca e juvenis criados em incubação para fins de “biomitigação”. Para os autores, o desafio é desenvolver verdadeiramente sistemas de produção para sustentar uma indústria perene, além de ampliar sua capacidade produtiva e sustentabilidade por meio de inovações e conhecimentos baseados na ciência.

Ademais, a pressão de demanda, a valorização cambial e os custos internacionais de captura e produção são fatores relacionados à possibilidade de inflação nos preços de venda destes pescados no Brasil. Isto impactaria nos gastos dos domicílios com pescado e, conseqüentemente, no consumo *per capita*.

Os resultados obtidos por Vaz e Hoffmann (2020), ao analisarem econometricamente os dados das POF entre 2002 e 2018, demonstraram que, entre 2008 e 2018, ocorreu uma inversão na tendência de aumento dos gastos com carnes, vísceras e pescados, além de uma diminuição nas despesas com todas as categorias de produtos *light/diet*. Essa mudança foi atribuída à crise econômica que afetou o país a partir de meados de 2014.

Vale sempre destacar que a capacidade de captura de pescados em águas interiores e ma-

rítmicas está estagnada há algumas décadas, situação que parece irreversível segundo órgãos internacionais. Entretanto, a aquicultura brasileira teria capacidade para suprir a demanda latente de peixes e sustentar o aumento gradual no consumo domiciliar *per capita*, desde que alguns ajustes estruturais sejam realizados (SCHULTER; VIEIRA FILHO, 2017).

Neste contexto, o Brasil vive uma consistente expansão do sistema agroalimentar da piscicultura motivada por mudanças institucionais ou organizacionais que estimularam o aumento da produção e oferta de tilápias na forma de filés e outros cortes, mas também por diferentes espécies, tais como o tambaqui, surubim ou pintado, pacu e patinga, pirarucu, truta salmonada e pangasius. Embora ainda longe de seu potencial produtivo, o setor caminha a passos largos para alcançar novos patamares no mercado internacional, tendo em vista atualmente ser o quarto maior produtor mundial de tilápia.

Segundo o SEBRAE (2015) no início da década de 2010 o Governo Federal implementou campanhas publicitárias que buscaram promover o produto e o setor, além do desenvolvimento da culinária e gastronomia, explorar os benefícios à saúde humana e as características próprias dos pescados. A instituição ponderava que as estratégias mercadológicas para o consumo de peixes pelos consumidores precisariam de uma boa articulação entre os segmentos do sistema agroalimentar, cabendo de exemplo o setor de avicultura.

Diante do exposto, esta proposta objetiva investigar a influência da localização; idade e escolaridade do responsável pelo domicílio; presença de crianças e adolescentes; e o rendimento, dentre outras variáveis, no consumo de diferentes tipos de pescados do estado de São Paulo. Espera-se que a pesquisa resulte em informações que possam subsidiar a cadeia agroalimentar da piscicultura no estabelecimento de estratégias mercadológicas e compreender, principalmente, as características de consumo de pescados de famílias com menores rendimentos mensais.

2 OBJETIVOS

Esta seção delinea as metas centrais e as etapas específicas traçadas para a condução deste estudo. Tendo em vista a complexidade dos microdados da POF e a necessidade de uma abordagem quantitativa robusta, os objetivos foram estruturados para guiar desde o tratamento inicial das bases de dados até a aplicação de modelagens estatísticas avançadas, visando compreender profundamente os determinantes do consumo de pescados no contexto paulista.

2.1 OBJETIVO GERAL

O objetivo geral deste trabalho é investigar a influência de variáveis socioeconômicas (idade, presença de crianças e adolescentes, escolaridade do responsável pelo domicílio; rendimento), dentre outras, no consumo de pescados no Estado de São Paulo. Adicionalmente, busca-se compreender as características de consumo de pescados (subgrupos e tipos) das famílias com menores rendimentos mensais.

2.2 OBJETIVOS ESPECÍFICOS

Para o cumprimento do objetivo geral, foram estabelecidas as seguintes etapas específicas:

a) Levantamento de dados secundários junto à Pesquisa de Orçamentos Familiares 2017-2018:

- i. Identificar o repositório de dados e compreender conceitos e métodos utilizados, especialmente as variáveis e fatores analisados pelo IBGE;
- ii. Estruturar banco de dados no software RStudio visando alocar os microdados da POF 2017-2018;
- iii. Criar novos bancos de dados com recortes específicos contendo conjunto de microdados do Estado de São Paulo;
- iv. Selecionar variáveis e sistematizar tabelas visando a análise de dados com técnicas de estatística descritiva e multivariada.

b) Aplicação de técnicas de estatística descritiva:

- i. Definir critérios e parâmetros de análise em função das variáveis anteriormente selecionadas;
- ii. Determinar distribuição de frequências, médias, modas e variância de dados das variáveis previamente selecionadas para a nova base de dados referente ao estado de São Paulo.

c) Aplicação de técnicas de estatística multivariada:

- i. Definir critérios e parâmetros de análise em função das variáveis anteriormente selecionadas;
- ii. Submeter as novas bases de dados a técnicas de: teste Alfa de Cronbach, Análise de Componentes Principais (ACP), Análise de Agrupamento (*k-means*, *cluster* e *two-way joining*) e Análise de Correspondência (AC).

d) Aplicação de técnicas de estatística geoespacial:

- i. Construir mapas relacionando o que foi encontrado nas análises anteriores através do software ArcGIS com variáveis exógenas coletadas do SIDRA.

3 JUSTIFICATIVA

O desenvolvimento deste projeto é de grande relevância social, pois visa investigar os fatores que determinam o consumo de pescados no estado de São Paulo, com ênfase em famílias de baixa renda. A pesca e a aquicultura são atividades essenciais tanto do ponto de vista nutricional quanto econômico, contribuindo para a segurança alimentar e o desenvolvimento sustentável. No entanto, o consumo de pescados no Brasil ainda está abaixo das recomendações internacionais, especialmente em regiões populosas como São Paulo, onde o acesso a proteínas aquáticas é mais limitado.

Ao entender os padrões de consumo e os fatores socioeconômicos que influenciam a escolha por pescados, este estudo poderá contribuir para a criação de políticas públicas e estratégias de mercado mais direcionadas, estimulando o consumo de produtos de origem aquática. Esse aumento pode gerar benefícios nutricionais, pois os pescados são ricos em nutrientes essenciais, como ômega-3 e proteínas de alta qualidade, além de fortalecer cadeias produtivas locais e promover maior inclusão social em áreas economicamente vulneráveis.

O projeto está diretamente relacionado aos Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU), especialmente aos seguintes:

ODS 2 – Fome Zero e Agricultura Sustentável: Ao promover o aumento do consumo de pescados e o fortalecimento da aquicultura, o projeto contribui para a segurança alimentar, garantindo que famílias de baixa renda tenham acesso a uma fonte saudável de proteína.

ODS 3 – Saúde e Bem-Estar: O incentivo ao consumo de pescados pode melhorar a dieta das famílias, promovendo uma alimentação mais saudável e equilibrada, o que impacta positivamente a saúde da população.

ODS 12 – Consumo e Produção Responsáveis: A análise dos padrões de consumo pode incentivar práticas de pesca e aquicultura sustentáveis, promovendo o consumo responsável de recursos naturais.

ODS 14 – Vida na Água: Ao investigar o mercado de pescados, o projeto pode promover uma melhor gestão dos recursos aquáticos e incentivar a preservação dos ecossistemas marinhos e de água doce, bem como o desenvolvimento da aquicultura de forma sustentável.

Portanto, os resultados deste estudo terão implicações não apenas para a indústria de pescados e a cadeia produtiva, mas também para a saúde pública e a sustentabilidade ambiental, contribuindo para o cumprimento das metas da Agenda 2030 da ONU.

4 REFERENCIAL TEÓRICO E ESTADO DA ARTE

A presente pesquisa fundamenta-se em um conjunto de referenciais que sustentam tanto a abordagem metodológica quanto a interpretação estatística aplicada aos dados de consumo alimentar no estado de São Paulo. O estudo articula competências em análise multivariada, estatística espacial e interpretação de padrões de consumo, proporcionando uma visão integrada dos fenômenos socioeconômicos associados ao consumo de pescado na região.

4.1 ANÁLISE MULTIVARIADA

A análise multivariada constitui o eixo central deste trabalho, pois permite investigar simultaneamente um conjunto de variáveis socioeconômicas e de condições de vida, identificando estruturas latentes e padrões que não seriam observáveis por análises univariadas ou bivariadas. Segundo Hair et al. (2009), técnicas como análise fatorial, análise de correspondência e métodos de agrupamento são essenciais para reduzir a dimensionalidade dos dados, sintetizar informação e compreender relações complexas entre variáveis.

Para complementar essa perspectiva, Johnson e Wichern (2007) enfatizam que a análise multivariada possibilita explorar interdependências naturais entre variáveis, revelando fatores subjacentes que explicam comportamento social, econômico e demográfico. No contexto deste estudo, esses métodos são fundamentais para captar dimensões latentes que influenciam o padrão de consumo de pescado, como condições domiciliares, nível socioeconômico e características demográficas dos moradores no estado de São Paulo. Tais técnicas fornecem subsídios para estruturar indicadores sintéticos, identificar grupos populacionais semelhantes e analisar contrastes relevantes entre segmentos da população paulista.

4.2 ESTATÍSTICA ESPACIAL E PADRÕES GEOGRÁFICOS

A análise estatística espacial complementa a abordagem multivariada ao considerar a distribuição dos fenômenos no espaço, permitindo interpretar como fatores sociais e econômicos se organizam territorialmente. Bailey e Gatrell (1995) descrevem que métodos espaciais são indispensáveis quando a localização dos eventos influencia sua ocorrência, revelando clusters, gradientes e heterogeneidades espaciais. Lopes e Vasconcelos (2007) reforçam que estudos espaciais englobam diferentes perspectivas, como análise de áreas, padrões pontuais e dependência espacial, contribuindo para a compreensão de estruturas que não são evidentes em análises puramente descritivas.

No âmbito desta pesquisa, a estatística espacial possibilita explorar padrões geográficos do consumo de pescado no estado de São Paulo, observando se regiões específicas apresentam perfis de consumo semelhantes ou divergentes. Além disso, tais métodos permitem relacionar

características socioeconômicas às dinâmicas espaciais, sugerindo potenciais desigualdades territoriais na aquisição de pescado e nas condições de vida associadas.

4.3 CONSUMO ALIMENTAR E APLICAÇÕES AO CONTEXTO PAULISTA

O referencial sobre consumo alimentar, especialmente voltado ao consumo de pescado, baseia-se em Wagner et al. (2023), que analisaram padrões brasileiros a partir da POF 2017–2018. Embora o estudo nacional ofereça uma visão ampla, sua contribuição direta para este trabalho está na compreensão das variáveis socioeconômicas que influenciam o consumo, como renda, escolaridade, acesso a serviços públicos e composição domiciliar.

Adaptado especificamente ao estado de São Paulo, este arcabouço teórico auxilia a interpretar como fatores regionais — como urbanização, infraestrutura, perfil socioeconômico e diferenças regionais dentro do próprio estado — moldam o consumo domiciliar de pescado. A literatura reforça que padrões alimentares são influenciados tanto por condições individuais quanto pelo contexto geográfico e socioeconômico, o que justifica a integração entre métodos multivariados e espaciais neste TCC.

5 METODOLOGIA

A metodologia adotada neste estudo foi estruturada de forma a garantir rigor analítico, reprodutibilidade e alinhamento às melhores práticas em pesquisas quantitativas baseadas em microdados. O objetivo central foi investigar os padrões de consumo domiciliar de pescados no estado de São Paulo, utilizando técnicas de estatística descritiva e multivariada para identificar características socioeconômicas, perfis de aquisição alimentar e estruturas latentes que influenciam o consumo. Para isso, foram utilizadas bases oficiais da Pesquisa de Orçamentos Familiares (POF 2017–2018), disponibilizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE), processadas integralmente no ambiente de programação R.

5.1 AMBIENTE DE EXECUÇÃO E RECURSOS UTILIZADOS

Os processos de manipulação, integração e análise dos microdados foram realizados em um notebook Apple MacBook Air M1, utilizando o software RStudio (versão 4.5.0). Todas as etapas de pré-processamento, classificação e modelagem foram desenvolvidas diretamente no ambiente R, garantindo eficiência computacional, automação e reprodutibilidade integral das análises.

Os microdados e documentos metodológicos foram obtidos diretamente do portal oficial do IBGE, assegurando a integridade das informações utilizadas. Adicionalmente, referências bibliográficas especializadas em estatística multivariada e análise de dados sustentaram a escolha dos métodos e a interpretação dos resultados, com destaque para Hair et al. (2009), Johnson e Wichern (2007) e documentos técnicos da POF 2017–2018.

5.2 CARACTERIZAÇÃO GERAL DA PESQUISA

Trata-se de uma pesquisa de natureza exploratória e abordagem quantitativa, fundamentada nos microdados da POF 2017–2018. O foco recai exclusivamente sobre o estado de São Paulo, com o objetivo de analisar padrões domiciliares de consumo de pescados.

A utilização dos registros domiciliares, aliados a variáveis socioeconômicas, demográficas e de estrutura de vida, possibilitou traçar perfis regionais e avaliar desigualdades associadas ao acesso e consumo de pescados. Essa caracterização permitiu contextualizar o comportamento alimentar no território paulista, contribuindo para a discussão sobre segurança alimentar, renda e hábitos de consumo no estado.

5.3 COLETA, INTEGRAÇÃO E ESTRUTURAÇÃO DOS DADOS

Os microdados foram disponibilizados em arquivos .txt pelo IBGE e processados no R conforme o layout oficial de variáveis. As bases DOMICÍLIO, MORADOR, CONDIÇÕES_DE_VIDA,

MORADOR_QUALIDADE_VIDA, CARACTERÍSTICAS_DIETA e CONSUMO_ALIMENTAR foram filtradas para manter exclusivamente os registros referentes ao estado de São Paulo.

A integração entre as tabelas foi realizada por meio da criação de um identificador único baseado nas variáveis COD_UPA, NUM_DOM, NUM_UC e COD_INFORMANTE, garantindo correspondência precisa entre as informações dos domicílios e dos moradores.

Após a fusão, foram selecionadas variáveis socioeconômicas, demográficas e de qualidade de vida relevantes ao estudo. A variável de consumo alimentar foi cruzada com uma tabela auxiliar de códigos alimentares, permitindo a classificação detalhada dos diferentes tipos de pescados adquiridos pelos domicílios.

Ao final, cada linha da base consolidada corresponde a um domicílio da POF, mantendo a unidade de análise domiciliar adotada neste estudo.

5.4 ANÁLISE DESCRITIVA

A etapa descritiva teve como objetivo fornecer uma visão inicial e abrangente sobre os domicílios paulistas incluídos na amostra, considerando sua composição socioeconômica e os padrões de aquisição de pescados. Como o banco de dados reúne variáveis de diferentes naturezas (nominais, ordinais e quantitativas), a análise foi conduzida de forma diferenciada para cada tipo de medida. Para as variáveis numéricas, foram calculadas medidas de tendência central (média, mediana e moda), de dispersão (amplitude, variância e desvio padrão) e distribuição dos valores. Já para variáveis categóricas, foram estimadas frequências absolutas, relativas e distribuições percentuais por estratos de interesse. Esse conjunto de procedimentos permitiu explorar a variabilidade interna da base, identificar assimetrias, categorias predominantes e eventuais padrões iniciais que sinalizam heterogeneidade entre os domicílios.

Os resultados obtidos nessa fase forneceram subsídios importantes para a definição das técnicas multivariadas subsequentes, uma vez que evidenciaram a necessidade de métodos capazes de sintetizar informações complexas e detectar estruturas latentes entre variáveis socioeconômicas e alimentares.

5.5 ANÁLISE MULTIVARIADA

A etapa multivariada seguiu as recomendações clássicas de Hair et al. (2009) e Johnson & Wichern (2007) para análise de dados complexos. Inicialmente, avaliou-se a consistência interna dos blocos de variáveis por meio do Alfa de Cronbach.

Em seguida, procedeu-se à Análise Fatorial por Componentes Principais (PCA) utilizando a matriz de correlações policóricas, adequada para variáveis ordinais. A extração de fatores considerou autovalores superiores a 1, com rotação varimax para maximizar a interpretabilidade dos fatores. Os scores fatoriais resultantes foram incorporados à base estruturada.

Para explorar relações entre categorias socioeconômicas e tipos de pescado consumidos, aplicou-se Análise de Correspondência, cujas representações bidimensionais permitem visualizar proximidades e associações entre categorias.

Posteriormente, os scores fatoriais foram utilizados como insumo para a Análise de Agrupamento Hierárquico, permitindo segmentar os domicílios em grupos homogêneos quanto aos padrões de consumo. A técnica Two-Way Joining complementou essa abordagem, evidenciando padrões simultâneos entre variáveis e unidades amostrais.

5.6 ELABORAÇÃO DOS MAPAS TEMÁTICOS

Como etapa complementar da análise, foram produzidos mapas temáticos no software *Arc-Map*, com o objetivo de representar espacialmente os padrões identificados nas etapas anteriores da pesquisa. A construção dos mapas utilizou os resultados gerados pelas análises estatísticas descritivas e multivariadas, especialmente aqueles relacionados aos perfis socioeconômicos e ao consumo de pescados pelos domicílios paulistas.

Além das informações derivadas dos microdados da POF 2017–2018, foram incorporados dados exógenos provenientes do SIDRA/IBGE, incluindo:

- Quantidade em Kg da produção aquícola;
- Valor em Reais da produção aquícola;
- Número de estabelecimentos que vendem produtos da aquicultura;

A integração dessas bases permitiu contextualizar espacialmente os padrões observados nas análises estatísticas, possibilitando visualizar áreas do estado com maior ou menor intensidade de consumo de pescados, bem como diferenças regionais associadas às condições socioeconômicas dos domicílios.

Os mapas atuaram como uma síntese visual dos resultados, facilitando a interpretação dos padrões geográficos e contribuindo para a identificação de heterogeneidades internas ao estado de São Paulo. Essa representação espacial reforçou a compreensão dos fatores que influenciam o consumo de pescados e permitiu relacionar as evidências analíticas aos contextos territoriais específicos.

6 TÉCNICAS UTILIZADAS

Esta seção apresenta os fundamentos teóricos e as técnicas estatísticas empregadas na pesquisa, com ênfase nos métodos de avaliação da confiabilidade, estimativa de correlações entre variáveis ordinais e procedimentos multivariados aplicados à análise dos padrões de consumo de pescados no estado de São Paulo. As abordagens metodológicas aqui descritas fornecem suporte conceitual para a construção dos indicadores, a redução da dimensionalidade e a avaliação da estrutura interna dos dados empregados na investigação.

6.1 ALFA DE CRONBACH

O Alfa de Cronbach é uma das métricas mais utilizadas para avaliar a **confiabilidade** ou **consistência interna** de um conjunto de itens que compõem uma mesma escala ou dimensão latente. Em termos conceituais, a medida indica o quanto os itens de um questionário, índice ou bloco de variáveis apresentam respostas coerentes entre si, refletindo um mesmo construto subjacente.

A lógica por trás do coeficiente parte do princípio de que, quando diversos itens buscam mensurar uma mesma característica, espera-se que apresentem correlação positiva entre si. Assim, quanto maior o grau de associação entre os itens, maior será a consistência interna da escala.

O Alfa de Cronbach varia entre 0 e 1. Valores próximos de 0 indicam fraca consistência interna, enquanto valores mais elevados sugerem que os itens se relacionam fortemente e medem adequadamente uma mesma dimensão. De modo geral, referências clássicas de análise multivariada sugerem que valores acima de 0,70 representam níveis aceitáveis de confiabilidade, embora esse limiar possa variar conforme o contexto e o tipo de instrumento analisado.

No presente estudo, o Alfa de Cronbach foi utilizado para avaliar a coerência interna dos blocos de variáveis socioeconômicas e de condições de vida utilizados nas análises multivariadas. Essa etapa é fundamental para assegurar que os conjuntos de variáveis empregados na Análise Fatorial apresentam estabilidade e representam adequadamente os construtos que se pretende investigar.

6.1.1 Introdução à Confiabilidade e ao Alfa de Cronbach

A análise de confiabilidade tem como objetivo avaliar o grau de consistência interna entre itens que compõem uma mesma dimensão latente, sendo fundamental em estudos que utilizam variáveis derivadas de questionários, percepções ou avaliações subjetivas. Entre as métricas disponíveis, o coeficiente Alfa de Cronbach, constitui o indicador mais amplamente utilizado para examinar se os itens de uma escala apresentam correlação suficiente para medir o mesmo construto.

O coeficiente é definido, em sua forma clássica, como:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma_T^2} \right),$$

em que k é o número de itens da escala, σ_i^2 é a variância do item i , e σ_T^2 é a variância total da soma dos itens. Em termos da matriz de correlações entre itens, o Alfa pode ser expresso pela forma alternativa:

$$\alpha = \frac{k \cdot \bar{r}}{1 + (k-1)\bar{r}},$$

na qual $\bar{r}$ representa a média das correlações entre pares de itens. Valores mais elevados de α indicam maior consistência interna, sendo usualmente aceitos como adequados valores superiores a 0,70.

6.1.2 Fundamentos da Correlação Policórica

A correlação policórica é apropriada para mensurar a associação entre variáveis ordinais que se supõe serem manifestações discretizadas de variáveis latentes contínuas não observáveis. Seu uso é especialmente indicado quando as categorias representam níveis ordenados, mas não assumem intervalos equivalentes entre si.

Suponha-se que duas variáveis ordinais Y_1 e Y_2 , com C_1 e C_2 categorias, sejam derivadas de variáveis latentes contínuas Y_1^* e Y_2^* , tais que:

$$(Y_1^*, Y_2^*) \sim N_2(0, 0, 1, 1, \rho),$$

onde ρ corresponde à correlação policórica entre as dimensões latentes. As variáveis observadas são definidas pela discretização de limites (thresholds):

$$Y_1 = c \quad \text{se} \quad \tau_{1,c-1} < Y_1^* \leq \tau_{1,c},$$

$$Y_2 = d \quad \text{se} \quad \tau_{2,d-1} < Y_2^* \leq \tau_{2,d}.$$

A probabilidade conjunta das categorias (c, d) é então dada por:

$$P(Y_1 = c, Y_2 = d) = \int_{\tau_{1,c-1}}^{\tau_{1,c}} \int_{\tau_{2,d-1}}^{\tau_{2,d}} \phi_2(u, v; \rho) du dv,$$

onde $\phi_2(u, v; \rho)$ denota a densidade da normal bivariada com correlação ρ . A estimação é realizada por máxima verossimilhança, ajustando-se os thresholds e o valor de ρ que maximizam a probabilidade das frequências observadas.

6.1.3 Interpretação da Correlação Policórica

A correlação policórica assume valores no intervalo:

$$\rho \in [-1, 1],$$

e representa a força e direção da associação entre variáveis latentes contínuas subjacentes às respostas ordinais. Diferentemente da correlação de Pearson, a correlação policórica não depende da codificação numérica atribuída às categorias, respeitando a natureza ordinal dos dados.

6.1.4 Alfa de Cronbach com Correlação Policórica

Quando os itens da escala são ordinais, o uso da matriz de correlação policórica é preferível para o cálculo do coeficiente Alfa de Cronbach, pois proporciona uma estimativa mais consistente da variância compartilhada entre os itens.

O Alfa baseado na matriz policórica é dado por:

$$\alpha_{\text{policórico}} = \frac{k \cdot \bar{r}_{\text{policórica}}}{1 + (k - 1)\bar{r}_{\text{policórica}}},$$

em que $\bar{r}_{\text{policórica}}$ representa a média das correlações policóricas entre os itens. Essa abordagem fornece maior precisão ao avaliar a consistência interna de escalas compostas por variáveis ordinais, aprimorando a validade das inferências estatísticas subsequentes.

6.2 ANÁLISE FATORIAL POR COMPONENTES PRINCIPAIS COM CORRELAÇÃO POLICÓRICA

A Análise Fatorial por Componentes Principais com correlação policórica constitui uma abordagem particularmente adequada quando o conjunto de variáveis observadas é predominantemente ordinal. Nesse caso, a estrutura de dependência entre os indicadores não pode ser corretamente representada pela correlação de Pearson, que pressupõe continuidade e linearidade. Em vez disso, assume-se que cada variável ordinal observada é fruto da discretização de uma variável latente contínua não observável. Assim, a matriz de correlação policórica busca estimar a associação entre essas variáveis latentes, permitindo uma representação mais fiel da estrutura subjacente aos dados.

A partir da matriz policórica, procede-se à extração dos fatores por meio do método das Componentes Principais (PCA), permitindo identificar dimensões latentes capazes de explicar a maior parcela da variância comum entre as variáveis. A retenção dos fatores segue critérios clássicos, como autovalores superiores a 1, inspeção do scree plot e proporção de variância acumulada. Posteriormente, aplica-se uma rotação ortogonal, usualmente a Varimax, com o objetivo de maximizar a interpretabilidade das cargas fatoriais e de evidenciar a contribuição relativa

de cada variável em cada fator latente. Essa abordagem permite reduzir a dimensionalidade dos dados, organizar padrões de associação entre indicadores socioeconômicos e gerar escores fatoriais que servem como insumos para análises subsequentes, como agrupamentos, perfis de consumo e identificação de padrões regionais no estado de São Paulo.

6.2.1 Introdução à Análise Fatorial

A Análise Fatorial Exploratória (AFE) consiste em um conjunto de procedimentos multivariados destinados a identificar estruturas latentes capazes de explicar a correlação existente entre um conjunto de variáveis observadas. De acordo com Hair et al. (2009) e Johnson & Wichern (2007), pressupõe-se que cada variável observada X_i pode ser expressa como uma combinação linear de fatores comuns mais um termo específico:

$$X_i = \ell_{i1}F_1 + \ell_{i2}F_2 + \cdots + \ell_{im}F_m + \epsilon_i,$$

onde ℓ_{ij} representa a carga fatorial da variável i no fator j , F_j os fatores comuns, e ϵ_i o erro específico.

Entre os métodos de extração de fatores, destaca-se a Análise por Componentes Principais (PCA), cujo objetivo é transformar um conjunto de variáveis correlacionadas em componentes ortogonais que retêm o máximo possível da variância total (Jolliffe & Cadima, 2016). A PCA é amplamente utilizada por sua estabilidade numérica e interpretabilidade, além da facilidade de integração com técnicas subsequentes, como análise de agrupamentos e construção de índices compostos.

6.2.2 Fundamentos Matemáticos da PCA

Considere um conjunto de p variáveis padronizadas $X_1, X_2, \dots, X_p$ com matriz de correlação R . A PCA busca determinar combinações lineares da forma:

$$Z_j = a_{j1}X_1 + a_{j2}X_2 + \cdots + a_{jp}X_p, \quad j = 1, \dots, p,$$

onde cada Z_j constitui um componente principal. Os coeficientes a_{jk} são escolhidos de forma a maximizar a variância de cada componente, sujeita à normalização dos autovetores.

A variância do componente Z_j é dada por:

$$\text{Var}(Z_j) = \lambda_j,$$

onde $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ são os autovalores da matriz R .

A decomposição espectral da matriz de correlação é expressa por:

$$R = AA^T,$$

em que:

- A é a matriz de autovetores (cargas fatoriais não rotacionadas);
- Λ é a matriz diagonal contendo os autovalores λ_j .

A comunalidade da variável i , que representa a proporção de variância explicada pelos m fatores retidos, é definida por:

$$h_i^2 = \sum_{j=1}^m a_{ij}^2.$$

6.2.3 Rotação Varimax: Teoria e Justificativa

A rotação Varimax, proposta por Kaiser (1958), busca maximizar a simplicidade estrutural dos fatores, preservando sua ortogonalidade. A função otimizada é:

$$V = \sum_{j=1}^m \left[\frac{1}{p} \sum_{i=1}^p (\ell_{ij}^2 - \bar{\ell}_j^2)^2 \right],$$

onde:

$$\bar{\ell}_j^2 = \frac{1}{p} \sum_{i=1}^p \ell_{ij}^2.$$

Ao maximizar V , obtém-se uma solução em que:

- cada variável apresenta carga elevada em poucos fatores;
- cada fator possui apenas algumas cargas elevadas;
- a interpretação dos fatores é simplificada.

Essa rotação é particularmente vantajosa no presente estudo, pois o conjunto de variáveis socioeconômicas e de qualidade de vida da POF envolve múltiplas dimensões latentes potencialmente correlacionadas, exigindo uma solução clara, interpretável e coerente com a estrutura teórica dos dados.

6.2.4 Etapas da Análise Fatorial com Matriz Policórica

(a) Construção da matriz policórica

Para variáveis ordinais, substitui-se a matriz de correlação de Pearson por uma matriz policórica R_{pol} , estimada sob o modelo de normalidade bivariada latente.

Sejam Y_1 e Y_2 duas variáveis ordinais, derivadas de variáveis latentes contínuas Y_1^* e Y_2^* , tais que:

$$(Y_1^*, Y_2^*) \sim N_2(0, 0, 1, 1, \rho),$$

onde ρ é o coeficiente de correlação policórica.

Cada variável observada é determinada por cortes (thresholds) τ :

$$Y_1 = c \quad \text{se} \quad \tau_{1,c-1} < Y_1^* \leq \tau_{1,c}$$

$$Y_2 = d \quad \text{se} \quad \tau_{2,d-1} < Y_2^* \leq \tau_{2,d}$$

A probabilidade conjunta de observar (c, d) é dada por:

$$P(Y_1 = c, Y_2 = d) = \int_{\tau_{1,c-1}}^{\tau_{1,c}} \int_{\tau_{2,d-1}}^{\tau_{2,d}} \phi_2(u, v; \rho) du dv,$$

em que $\phi_2(u, v; \rho)$ é a densidade da normal bivariada com correlação ρ .

A estimação de ρ ocorre por máxima verossimilhança, utilizando as frequências observadas nas tabelas de contingência.

(b) Extração dos Componentes Principais

A extração ocorre a partir da decomposição espectral:

$$R_{\text{pol}} = A\Lambda A^\top.$$

Os autovalores λ_j indicam a variância explicada pelos fatores.

(c) Critérios de retenção

- **Critério de Kaiser:** manter fatores com $\lambda_j > 1$.
- **Scree Plot:** identificar o ponto de inflexão (“cotovelo”).
- **Variância acumulada:** geralmente busca-se $\geq 60\%$.

(d) Rotação Varimax

Após a extração inicial, aplica-se a rotação ortogonal Varimax para facilitar a interpretação das cargas fatoriais.

6.3 ANÁLISE DE CORRESPONDÊNCIA

A Análise de Correspondência (AC) é uma técnica estatística exploratória multivariada destinada à investigação de relações entre categorias de variáveis qualitativas. Sua lógica fundamental consiste em decompor a estrutura de associação presente em uma tabela de contingência, projetando linhas e colunas em um espaço vetorial de baixa dimensão. Assim, categorias semelhantes

são posicionadas próximas no mapa fatorial, enquanto categorias dessemelhantes aparecem afastadas. A AC tem forte embasamento geométrico e matricial, sendo conceitualmente análoga à Análise de Componentes Principais, porém definida especificamente para variáveis nominais e ordinais agrupadas em frequências (Greenacre, 2007, 2017). Esta técnica é particularmente valiosa em estudos sociais, epidemiológicos, de marketing e, como neste trabalho, na investigação de padrões de consumo alimentar em populações.

6.3.1 Estrutura Matemática da Análise de Correspondência Simples

Considere uma tabela de contingência $\mathbf{N}$ de dimensão $I \times J$, onde I representa as categorias da variável das linhas e J representa as categorias da variável das colunas. Seja $n = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$ o total de observações. Define-se:

$$\mathbf{P} = \frac{1}{n} \mathbf{N},$$

a matriz de proporções conjuntas. As proporções marginais são dadas pelos vetores:

$$\mathbf{r} = \mathbf{P} \mathbf{1}, \quad \mathbf{c} = \mathbf{P}^\top \mathbf{1},$$

onde $\mathbf{1}$ é um vetor coluna de 1s.

As matrizes diagonais associadas às marginais são:

$$\mathbf{D}_r = \text{diag}(r_1, \dots, r_I), \quad \mathbf{D}_c = \text{diag}(c_1, \dots, c_J).$$

Define-se a matriz de resíduos padronizados sob o modelo de independência por:

$$\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}} (\mathbf{P} - \mathbf{r} \mathbf{c}^\top) \mathbf{D}_c^{-\frac{1}{2}}.$$

A decomposição em valores singulares (SVD) de $\mathbf{S}$ é:

$$\mathbf{S} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top,$$

onde:

- $\mathbf{U}$ contém os autovetores das linhas,
- $\mathbf{V}$ contém os autovetores das colunas,
- $\mathbf{\Sigma}$ contém os valores singulares σ_k .

As coordenadas fatoriais das linhas e colunas são dadas por:

$$\mathbf{F} = \mathbf{D}_r^{-\frac{1}{2}} \mathbf{U} \mathbf{\Sigma}, \quad \mathbf{G} = \mathbf{D}_c^{-\frac{1}{2}} \mathbf{V} \mathbf{\Sigma}.$$

A inércia total, análoga à variância explicada na PCA, é dada por:

$$\text{Inércia Total} = \sum_{i=1}^I \sum_{j=1}^J \frac{(p_{ij} - r_i c_j)^2}{r_i c_j},$$

e cada dimensão k explica:

$$\text{Inércia}_k = \sigma_k^2.$$

6.3.2 Interpretação e Representação Gráfica

A AC produz um mapa fatorial bidimensional no qual categorias de linhas e colunas são representadas simultaneamente. A distância qui-quadrado entre perfis direciona a projeção geométrica:

$$d^2(i, i') = \sum_{j=1}^J \frac{1}{c_j} \left(\frac{p_{ij}}{r_i} - \frac{p_{i'j}}{r_{i'}} \right)^2.$$

A proximidade entre pontos indica associação, enquanto o afastamento sugere perfis distintos. A qualidade da representação é medida pelo $\cos^2$:

$$\cos^2(i, k) = \frac{f_{ik}^2}{\sum_l f_{il}^2},$$

que avalia quanta variabilidade de um ponto é explicada pela dimensão k .

6.3.3 Associações com a Análise Fatorial

A AC pode ser entendida como uma forma de análise fatorial aplicada a dados categóricos. A relação conceitual com a PCA se dá pelo fato de ambas utilizarem decomposição espectral, porém com matrizes distintas:

$$\text{PCA: } \mathbf{R} = \mathbf{A}\mathbf{A}\mathbf{A}^\top, \quad \text{AC: } \mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top.$$

Enquanto a PCA utiliza covariâncias/correlações, a AC utiliza resíduos ajustados do modelo de independência. Assim, a AC é considerada a "contraparte categórica" da PCA (Mingoti, 2005; Hair et al., 2009).

6.3.4 Análise de Correspondência Múltipla (ACM)

A ACM é uma extensão natural da AC simples para o caso em que há mais de duas variáveis qualitativas. Suponha um conjunto de Q variáveis categóricas, recodificadas na forma de matriz indicador disjuntiva completa:

$$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_Q],$$

onde cada $\mathbf{z}_q$ contém colunas binárias indicando presença/ausência de categorias.

Define-se a matriz de Burt:

$$\mathbf{B} = \mathbf{Z}^\top \mathbf{Z},$$

que representa todas as combinações de frequências cruzadas entre variáveis.

A ACM nada mais é do que aplicar a AC sobre a matriz de Burt ou, alternativamente, sobre a matriz indicador centrada e ponderada. A decomposição espectral é dada por:

$$\mathbf{S}_{\text{ACM}} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{P}_B - \mathbf{c}\mathbf{c}^\top)\mathbf{D}^{-\frac{1}{2}},$$

onde $\mathbf{P}_B = \mathbf{B}/n$ e $\mathbf{D}$ contém as frequências marginais das categorias.

Os escores fatoriais das categorias são:

$$\mathbf{F}_{\text{ACM}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{U}\mathbf{\Sigma}.$$

Assim como na AC simples, a ACM permite visualizar relações entre múltiplas variáveis categóricas simultaneamente, sendo muito apropriada para dados socioeconômicos, como no caso deste estudo.

6.4 ANÁLISE DE AGRUPAMENTO HIERÁRQUICO

A Análise de Agrupamento (ou *Cluster Analysis*) é um conjunto de técnicas multivariadas voltadas à identificação de estruturas de similaridade em conjuntos de observações, por meio da formação de grupos relativamente homogêneos internamente e heterogêneos entre si (HAIR et al., 2009; MINGOTI, 2005). Em termos gerais, busca-se particionar os domicílios em segmentos que apresentem padrões semelhantes em relação às variáveis de interesse.

Neste trabalho, a análise de agrupamento foi empregada com dois objetivos principais: (i) agrupar os domicílios paulistas a partir dos escores fatoriais obtidos na Análise Fatorial com matriz de correlação policórica, sintetizando dimensões latentes de condições de vida e padrão de consumo; e (ii) formar grupos de domicílios com base em variáveis socioeconômicas e de consumo de pescados, de natureza mista (nominais e ordinais), respeitando a escala de mensuração dos dados. Para o primeiro caso, utilizou-se a distância euclidiana e o método de ligação de Ward sobre os escores fatoriais (dados contínuos). Para o segundo, empregou-se a distância de Gower, apropriada para variáveis mistas, combinada com o critério de Ward aplicado à matriz de dissimilaridade resultante (GOWER, 1971; LEGENDRE; LEGENDRE, 2012).

A abordagem hierárquica aglomerativa foi adotada em ambos os casos, de modo que a solução de agrupamento é construída iterativamente a partir de n grupos unitários (cada observação em um grupo), fundindo-se sucessivamente os pares de grupos mais semelhantes até que se obtenha uma única partição. O resultado é representado por um dendrograma, que ilustra visualmente as etapas de fusão e os níveis de dissimilaridade associados (KAUFMAN; ROUSSEUW, 2009).

6.4.1 Introdução à análise de agrupamento hierárquico

A análise de agrupamento hierárquico é uma técnica de classificação não supervisionada que não requer, a priori, o número de grupos a ser formado. Dado um conjunto de n observações descritas por p variáveis, o procedimento aglomerativo inicia com n clusters de tamanho 1 e, a cada iteração, funde os dois grupos mais similares de acordo com uma medida de dissimilaridade entre clusters, definida a partir de uma matriz de distâncias entre as observações (MINGOTI, 2005; HAIR et al., 2009).

Seja $\mathbf{X}$ uma matriz de dados de dimensão $n \times p$, em que cada linha representa um domicílio e cada coluna uma variável (por exemplo, escores fatoriais ou indicadores socioeconômicos). A análise de agrupamento hierárquico pode ser decomposta em três componentes principais:

1. **Escolha da métrica de distância entre observações**, que quantifica o quão diferentes dois domicílios são em relação ao conjunto de variáveis.
2. **Escolha do método de ligação**, que define a forma como as distâncias entre grupos são calculadas a partir das distâncias entre as observações.
3. **Critério de definição do número de clusters**, que orienta o ponto de corte no dendrograma a ser considerado como particionamento final.

Nos escores fatoriais (variáveis contínuas de mesma escala), utilizou-se a distância euclidiana, seguida do método de Ward, que minimiza a variância intra-grupos. Já no conjunto de variáveis mistas (nominais, ordinais e derivadas de consumo), empregou-se a distância de Gower, capaz de lidar com diferentes tipos de variáveis em uma mesma matriz de dissimilaridade, associando-se também o critério de Ward à matriz resultante (GOWER, 1971; LEGENDRE; LEGENDRE, 2012).

6.4.2 Etapas do procedimento

Padronização das variáveis contínuas

Quando as variáveis utilizadas no agrupamento são contínuas e medidas em escalas numéricas distintas, a padronização é recomendada para evitar que variáveis com maior variância dominem a medida de distância (HAIR et al., 2009). A padronização via *z-score* foi empregada para os escores fatoriais e demais variáveis contínuas, definindo-se:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{s_j},$$

em que X_{ij} é o valor da variável j para o domicílio i , $\bar{X}_j$ é a média da variável j e s_j é o desvio-padrão de X_j . A matriz padronizada é denotada por $\mathbf{Z}$.

Distância Euclidiana para escores fatoriais

Para o agrupamento baseado em escores fatoriais (dados contínuos e padronizados), foi utilizada a distância euclidiana clássica. A distância entre duas observações i e j é dada por:

$$d_{ij}^{(E)} = \sqrt{\sum_{k=1}^p (z_{ik} - z_{jk})^2},$$

onde z_{ik} representa o valor padronizado da variável k (por exemplo, o escore do fator k) para o domicílio i . A matriz de distâncias euclidianas $\mathbf{D}^{(E)} = [d_{ij}^{(E)}]$ foi então utilizada como base para o método de ligação de Ward na formação dos clusters.

No contexto desta pesquisa, a utilização da distância euclidiana sobre os escores fatoriais é apropriada porque esses escores são combinações lineares das variáveis originais, já em escala contínua e comparável, representando dimensões latentes de condições socioeconômicas e padrões de consumo.

Distância de Gower para variáveis mistas e ordinais

Para o agrupamento baseado diretamente em variáveis socioeconômicas e de consumo com natureza mista (nominais, ordinais e quantitativas), adotou-se a distância de Gower (GOWER, 1971), que permite combinar diferentes tipos de variáveis em uma única medida de dissimilaridade.

Sejam $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ e $\mathbf{x}_j = (x_{j1}, \dots, x_{jp})$ dois domicílios. A similaridade parcial entre i e j na variável k é denotada por s_{ijk} , definida de acordo com o tipo da variável:

$$s_{ijk} = \begin{cases} 1 - \frac{|x_{ik} - x_{jk}|}{R_k}, & \text{se } k \text{ é quantitativa,} \\ 1, & \text{se } k \text{ é nominal e } x_{ik} = x_{jk}, \\ 0, & \text{se } k \text{ é nominal e } x_{ik} \neq x_{jk}, \\ 1 - \frac{|r_{ik} - r_{jk}|}{M_k - 1}, & \text{se } k \text{ é ordinal,} \end{cases}$$

em que R_k é a amplitude da variável k , r_{ik} e r_{jk} representam os postos (ranks) das categorias ordinais para os domicílios i e j , e M_k é o número de categorias ordinais possíveis.

A similaridade global entre i e j é calculada como a média das similaridades parciais:

$$s_{ij} = \frac{\sum_{k=1}^p w_{ijk} s_{ijk}}{\sum_{k=1}^p w_{ijk}},$$

onde w_{ijk} é um peso que indica se a comparação na variável k é válida (por exemplo, $w_{ijk} = 1$ quando ambos os valores são observados e 0 em presença de dados faltantes). A dissimilaridade

de Gower é então definida por:

$$d_{ij}^{(G)} = 1 - s_{ij}.$$

A matriz de dissimilaridade de Gower, $\mathbf{D}^{(G)} = [d_{ij}^{(G)}]$, foi utilizada como insumo para o agrupamento hierárquico, permitindo que variáveis nominais, ordinais e quantitativas contribuíssem simultaneamente para a formação dos clusters de domicílios, respeitando a natureza da escala de mensuração dos dados.

Método de ligação de Ward

O método de ligação de Ward é um critério de agrupamento hierárquico aglomerativo que busca, a cada etapa, fundir os grupos de modo a minimizar o aumento da soma dos quadrados dentro dos clusters, isto é, da variância intra-grupo (WARD, 1963; LEGENDRE, 2012).

Seja C um cluster com n_C observações e vetor médio $\bar{\mathbf{z}}_C$. A soma de quadrados dentro do cluster é:

$$W(C) = \sum_{i \in C} \|\mathbf{z}_i - \bar{\mathbf{z}}_C\|^2.$$

Na etapa aglomerativa, ao considerar a fusão de dois clusters A e B , a variação na soma de quadrados total é dada por:

$$\Delta W_{AB} = \frac{n_A n_B}{n_A + n_B} \|\bar{\mathbf{z}}_A - \bar{\mathbf{z}}_B\|^2,$$

em que n_A e n_B são os tamanhos dos clusters e $\bar{\mathbf{z}}_A$ e $\bar{\mathbf{z}}_B$ são seus vetores médios. O método de Ward seleciona, a cada passo, o par (A, B) que minimiza ΔW_{AB} , gerando clusters internamente mais compactos.

- **Ward com distância euclidiana:** no agrupamento dos escores fatoriais, o método de Ward foi aplicado diretamente à matriz de distâncias euclidianas, o que é consistente com a sua formulação original, uma vez que os escores fatoriais se encontram em um espaço métrico euclidiano.
- **Ward com distância de Gower:** no agrupamento baseado em variáveis mistas (incluindo variáveis ordinais), o critério de Ward foi aplicado sobre a matriz de dissimilaridades de Gower $\mathbf{D}^{(G)}$. Embora a formulação clássica de Ward pressuponha um espaço euclidiano, na prática é comum empregar o critério de Ward em matrizes de dissimilaridade obtidas por Gower como uma aproximação que privilegia a formação de grupos mais compactos e homogêneos em termos das variáveis categóricas e ordinais (LEGENDRE; LEGENDRE, 2012).

6.4.3 Determinação do número de clusters

A definição do número ótimo de clusters (k) é uma etapa essencial na análise de agrupamento. Neste estudo, foram utilizados critérios gráficos e numéricos, especialmente o exame visual do dendrograma (método do *cotovelo*) e o coeficiente de silhueta (KAUFMAN; ROUSSEEUW, 2009; ROUSSEEUW, 1987).

Método do cotovelo no dendrograma

No método do cotovelo, observa-se o nível de dissimilaridade (altura) em que ocorrem as fusões no dendrograma ou, alternativamente, a evolução de uma medida de inércia intra-grupo (como a soma de quadrados dentro dos clusters) em função do número de grupos. O ponto em que há uma mudança acentuada na inclinação da curva—isto é, onde a redução da inércia deixa de ser significativa com o aumento de k —é interpretado como um candidato ao número adequado de clusters.

Coeficiente de silhueta

O coeficiente de silhueta quantifica o quão bem ajustada cada observação está ao cluster ao qual foi atribuída, comparando a coesão interna do grupo com a separação em relação aos demais grupos (ROUSSEEUW, 1987; KAUFMAN; ROUSSEEUW, 2009). Para uma observação i , define-se:

$$a(i) = \frac{1}{|C(i)| - 1} \sum_{j \in C(i), j \neq i} d_{ij},$$

onde $C(i)$ é o cluster ao qual i pertence e d_{ij} é a dissimilaridade entre i e j . O valor $a(i)$ representa a dissimilaridade média de i em relação às demais observações de seu próprio cluster (coesão).

Para cada cluster $C \neq C(i)$, calcula-se:

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d_{ij},$$

e define-se:

$$b(i) = \min_{C \neq C(i)} d(i, C),$$

isto é, $b(i)$ é a menor dissimilaridade média entre i e qualquer outro cluster (separação). O coeficiente de silhueta de i é:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$

com $-1 \leq s(i) \leq 1$. Valores de $s(i)$ próximos de 1 indicam boa alocação do indivíduo ao cluster, valores próximos de 0 sugerem fronteira entre dois grupos e valores negativos apontam

possível alocação inadequada. O valor médio da silhueta, $\bar{s} = \frac{1}{n} \sum_{i=1}^n s(i)$, fornece uma medida global da qualidade da partição e foi utilizado como critério auxiliar para a escolha de k .

6.4.4 Interpretação dos resultados de agrupamento

Após a definição do número de clusters e a obtenção do dendrograma final, procedeu-se à caracterização dos grupos formados. Para o agrupamento com escores fatoriais, foram calculadas estatísticas descritivas (médias e desvios-padrão) dos escores em cada cluster, permitindo interpretar os grupos em termos das dimensões latentes associadas às condições de vida e padrões de consumo. Assim, identificaram-se, por exemplo, grupos de domicílios com melhores condições de infraestrutura e padrão de vida versus domicílios em situação mais precária.

No agrupamento baseado em variáveis mistas (incluindo variáveis socioeconômicas, localização e indicadores de consumo de pescados), foram estimadas proporções, frequências relativas e medidas resumo em cada cluster. Essa etapa possibilitou descrever perfis típicos de domicílios paulistas quanto à renda, escolaridade, padrões de aquisição de pescados e demais características, evidenciando a heterogeneidade interna do estado de São Paulo e auxiliando na interpretação substantiva dos resultados obtidos nas etapas anteriores da análise multivariada.

6.5 TWO-WAY JOINING

O método *Two-Way Joining* (TWJ) é uma técnica de ordenação bidimensional aplicada a matrizes de dados, cujo objetivo é revelar padrões estruturados de similaridade tanto entre as unidades (linhas) quanto entre as variáveis (colunas). Diferentemente de métodos tradicionais de agrupamento ou redução de dimensionalidade, o TWJ reorganiza simultaneamente linhas e colunas de uma matriz, buscando maximizar a concentração de valores similares em regiões contíguas. Essa abordagem permite identificar blocos homogêneos e padrões de associação que não seriam evidenciados apenas por técnicas unidimensionais, como análises de cluster ou componentes principais.

No contexto desta pesquisa, o TWJ é particularmente apropriado porque os dados analisados envolvem variáveis ordinais provenientes da POF 2017–2018, além de indicadores derivados das análises fatoriais e do consumo binário de pescado. A técnica permite sintetizar visualmente padrões de consumo e características socioeconômicas dos domicílios, evidenciando relações estruturadas entre perfis familiares e a probabilidade de consumir ou não pescado. Dessa forma, o TWJ funciona como ponte entre as etapas multivariadas anteriores (análise fatorial e agrupamentos) e a fase de interpretação substantiva dos resultados, contribuindo para a compreensão dos padrões agregados de comportamento alimentar no estado de São Paulo.

6.5.1 Fundamentos Matemáticos do Two-Way Joining

Considere uma matriz de dados $\mathbf{X}$ de dimensão $p \times n$, em que p representa o número de variáveis e n o número de unidades (domicílios). O objetivo do TWJ é encontrar permutações π

(para as linhas) e γ (para as colunas) tais que a matriz reorganizada,

$$\mathbf{X}^* = \mathbf{P}_\pi \mathbf{X} \mathbf{P}_\gamma,$$

maximize a concentração de valores similares ao longo da diagonal e de regiões adjacentes. As matrizes $\mathbf{P}_\pi$ e $\mathbf{P}_\gamma$ são matrizes de permutação aplicadas às linhas e colunas, respectivamente.

O método clássico do TWJ utiliza como critério o índice de energia de ligação (Bond Energy Algorithm, BEA), definido como:

$$\text{BEA} = \sum_{i=1}^p \sum_{j=1}^n x_{ij} (x_{i,j-1} + x_{i,j+1} + x_{i-1,j} + x_{i+1,j}),$$

onde termos fora dos limites são desprezados. Esse critério mede a “energia” associada a colocar valores semelhantes próximos, e o algoritmo procura maximizar essa energia por meio de permutações sucessivas.

Outra formulação útil, amplamente associada à seriação, envolve minimizar uma função de custo baseada em distâncias entre posições das variáveis ou indivíduos. Um critério alternativo é o *Minimum Energy* (ME):

$$\text{ME} = \sum_{i < j} w_{ij} (o_i - o_j)^2,$$

em que w_{ij} é uma medida de similaridade entre as variáveis (ou entre unidades) e o_i e o_j representam as posições ordenadas após o rearranjo. O objetivo é encontrar a ordenação o que minimize a energia total, agrupando elementos similares.

6.5.2 Seriação e Algoritmos de Ordenação no Two-Way Joining

A etapa central do método Two-Way Joining (TWJ) consiste na **seriação**, isto é, na determinação de uma ordem para as linhas e colunas da matriz de dados que revele padrões estruturais, como faixas de altos e baixos valores, gradientes ou blocos de associação. A seriação é um problema clássico em análise multivariada, no qual se busca reorganizar uma matriz $X = [x_{ij}]$ de modo a aproximá-la de uma configuração ordenada, onde elementos similares fiquem próximos (Hubert, 1976; Bar-Joseph et al., 2001).

No contexto deste estudo, a matriz analisada representa médias normalizadas das variáveis socioeconômicas para dois grupos binários (consumiu versus não consumiu pescado). O objetivo é encontrar permutações π (linhas) e σ (colunas) que maximizem a organização estrutural:

$$X^* = P_\pi X P_\sigma,$$

onde P_π e P_σ são matrizes de permutação. O TWJ aplica algoritmos heurísticos que aproximam soluções ótimas para esse problema combinatório, dado que a busca exata teria complexidade $O(n! \cdot m!)$.

Bond Energy Algorithm (BEA)

O BEA é o método clássico do TWJ e busca maximizar a **energia de ligação** entre elementos adjacentes. Para três colunas consecutivas i, j, k , define-se a função de energia:

$$E(i, j) = \sum_{t=1}^n x_{ti} x_{tj}.$$

A energia total da matriz ordenada é dada por:

$$\text{BEA}(X) = \sum_{j=1}^{m-1} E(j, j+1).$$

O algoritmo constrói a ordenação coluna a coluna, buscando maximizar:

$$\Delta E = E(i, j) + E(j, i),$$

a cada inserção de coluna na posição candidata.

O objetivo é:

$$\max_{\pi, \sigma} \text{BEA}(X^*),$$

onde X^* é a matriz resultante das permutações.

Mean Energy (ME)

Uma variante utilizada quando a matriz possui amplitude variada de valores é a energia média:

$$\text{ME}(X) = \frac{1}{m-1} \sum_{j=1}^{m-1} E(j, j+1).$$

Essa medida suaviza a influência de pares de colunas extremos e é útil quando a matriz é pequena, como no presente estudo.

CrITÉrio de Moore (Stress)

O critério de Moore mede o grau de desorganização da matriz ordenada:

$$\text{Stress}(X^*) = \sum_{i < j} w_{ij} (d_{ij}^* - d_{ij})^2,$$

onde:

- d_{ij}^* é a distância entre elementos após a seriação,
- d_{ij} é sua distância original,

- w_{ij} são pesos opcionais.

Quanto menor o stress, maior a adequação da ordem encontrada.

Critério de Neumann

O critério de Neumann interpreta a matriz como um campo bidimensional e minimiza “saltos” entre elementos consecutivos:

$$\text{Neumann}(X^*) = \sum_{i=1}^{n-1} \sum_{j=1}^{m-1} ((x_{ij} - x_{i+1,j})^2 + (x_{ij} - x_{i,j+1})^2).$$

Uma solução de boa qualidade minimiza variações abruptas, aproximando a matriz de uma forma “suave”.

Justificativa para o uso da Sieriação no Estudo

O TWJ é adequado neste estudo porque:

- permite organizar simultaneamente **variáveis ordinais padronizadas** e um **desfecho binário**;
- evidencia padrões globais de associação que não emergem em PCA ou análise de clusters;
- lida bem com matrizes pequenas e interpretáveis;
- mostra visualmente contrastes fortes entre grupos socioeconômicos.

A ordenação produzida permite identificar quais dimensões socioeconômicas se alinham mais fortemente com o consumo de pescado, reforçando achados das análises fatoriais e de correspondência.

6.5.3 Aplicação do TWJ com Séries Ordinais Normalizadas

Para este estudo, o TWJ foi aplicado a uma matriz construída a partir das médias das variáveis socioeconômicas por categoria da variável binária de consumo de pescado. Como as variáveis da POF são essencialmente ordinais, adotou-se uma transformação *min-max* para padronização:

$$x_{ij}^* = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}.$$

Essa transformação preserva a estrutura ordinal, colocando todas as variáveis na mesma escala $[0, 1]$, requisito para métodos baseados na intensidade dos valores como o TWJ. A matriz padronizada e agregada permite comparar perfis médios dos domicílios que consomem e não consomem pescado, fornecendo uma representação bidimensional condensada dos padrões socioeconômicos.

A interpretação do TWJ é feita a partir do mapa de calor (*heatmap*) da matriz ordenada. Linhas adjacentes representam variáveis com comportamentos similares entre os grupos, enquanto colunas adjacentes indicam proximidade entre os perfis de consumo. Assim, o TWJ complementa as análises fatoriais e de agrupamentos, permitindo identificar blocos de variáveis que, em conjunto, caracterizam perfis de maior ou menor probabilidade de consumo de pescado.

O uso do TWJ neste trabalho se justifica por sua capacidade de sintetizar padrões complexos, preservando simultaneamente relações entre unidades e variáveis. Tal característica é especialmente relevante quando se analisam perfis de domicílios com múltiplos indicadores socioeconômicos e comportamentais, como é o caso dos microdados da POF para o estado de São Paulo.

6.6 ANÁLISE ESPACIAL

A análise espacial constitui um conjunto de métodos estatísticos voltados à avaliação de padrões, estruturas de dependência e heterogeneidade associados à localização geográfica das observações (BAILEY; GATRELL, 1995). Ao incorporar a dimensão espacial, torna-se possível identificar concentrações, tendências direcionais, dependência espacial e áreas críticas, complementando as análises multivariadas previamente realizadas.

No presente estudo, técnicas clássicas de estatística espacial foram aplicadas com o objetivo de compreender como o consumo domiciliar de pescados no estado de São Paulo se distribui espacialmente e como se relaciona com indicadores exógenos provenientes do SIDRA/IBGE, tais como: (i) quantidade em quilogramas da produção aquícola, (ii) valor monetário da produção aquícola, e (iii) número de estabelecimentos de aquicultura por município. A integração dessas informações permitiu interpretar o consumo de pescados no contexto da oferta produtiva regional e da estrutura do setor aquícola paulista.

6.6.1 Centro Médio (Mean Center)

O *Centro Médio* é uma medida de tendência central espacial que resume a posição geográfica média de um conjunto de pontos. Sua finalidade é identificar o ponto central teórico ao redor do qual a distribuição espacial se organiza, sendo útil para análises comparativas entre períodos, regiões ou grupos (ESRI, 2022).

Dado um conjunto de coordenadas geográficas (x_i, y_i) para $i = 1, \dots, n$, o centro médio é calculado por:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

O par $(\bar{x}, \bar{y})$ representa o centro geográfico da distribuição dos municípios segundo a variável analisada, permitindo verificar deslocamentos espaciais relacionados ao consumo de pescados ou indicadores da aquicultura.

6.6.2 Distribuição Direcional e Elipse de Desvio-Padrão

A *Distribuição Direcional*, operacionalizada por meio da *Elipse de Desvio-Padrão* (Standard Deviatonal Ellipse — SDE), descreve a orientação, forma e dispersão espacial de um conjunto de pontos. Essa técnica permite identificar direções predominantes, grau de concentração e possíveis alongamentos espaciais da distribuição geográfica.

A SDE é definida a partir do centro médio e dos desvios-padrão direcionais:

$$s_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad s_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}.$$

A orientação da elipse é dada pelo ângulo θ :

$$\tan(2\theta) = \frac{2 \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (y_i - \bar{y})^2}.$$

A elipse resultante resume graficamente o sentido predominante da variabilidade espacial. No contexto deste estudo, a SDE permitiu visualizar a direção predominante associada ao consumo de pescados nos municípios paulistas, relacionando tais padrões às características produtivas da aquicultura.

6.6.3 Índice de Moran Local (LISA)

O Índice de Moran Local, ou LISA (*Local Indicators of Spatial Association*), proposto por Anselin (1995), permite identificar padrões de autocorrelação espacial em nível local, destacando a presença de aglomerados (*clusters*) e outliers espaciais entre unidades geográficas, como municípios. Neste estudo, o LISA foi aplicado para examinar o consumo de pescados no estado de São Paulo e sua relação com indicadores municipais da produção aquícola disponibilizados pelo SIDRA/IBGE.

a) Formulação Matemática

Seja x_i o valor observado no município i , padronizado como

$$z_i = \frac{x_i - \bar{x}}{s},$$

onde $\bar{x}$ é a média amostral e s é o desvio-padrão. A matriz de pesos espaciais $W = [w_{ij}]$ descreve a vizinhança entre os municípios.

O Moran Local é dado por:

$$I_i = z_i \sum_j w_{ij} z_j.$$

Valores positivos indicam associação espacial direta; valores negativos apontam dissimilaridade em relação aos vizinhos imediatos.

b) Matriz de Pesos Espaciais: Método da Rainha

A matriz de contiguidade utilizada foi baseada no **critério da Rainha (Queen contiguity)**, segundo o qual dois municípios são considerados vizinhos quando compartilham **qualquer tipo de fronteira**: borda ou vértice. A formalização é dada por:

$$w_{ij} = \begin{cases} 1, & \text{se os municípios } i \text{ e } j \text{ são contíguos (rainha),} \\ 0, & \text{caso contrário.} \end{cases}$$

Esse método foi escolhido porque:

- captura melhor a conectividade real entre municípios brasileiros, cujas fronteiras são irregulares;
- evita perda de informação quando há municípios com fronteiras muito pequenas;
- tende a produzir estruturas espaciais mais completas e estáveis para análises locais como LISA.

c) Matriz Inversa da Distância

Para reforçar a influência espacial em alguns testes complementares, utilizou-se também uma matriz baseada na **inversa da distância**, definida por:

$$w_{ij} = \frac{1}{d_{ij}}, \quad i \neq j,$$

onde d_{ij} é a distância geográfica entre os centroids dos municípios.

A utilização da inversa da distância é justificada porque:

- modela a interação espacial de forma contínua;
- atribui maior peso aos vizinhos mais próximos, coerente com mecanismos de difusão espacial;
- evita que municípios muito extensos distorçam a análise, como no caso da contiguidade.

Ambas as matrizes (rainha e distância inversa) foram padronizadas por linha:

$$w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}},$$

garantindo comparabilidade entre municípios com número distinto de vizinhos.

d) Classificação dos Padrões Espaciais

A interpretação dos valores de I_i segue quatro categorias principais:

- **High–High:** valores altos cercados por valores altos (clusters positivos).
- **Low–Low:** valores baixos cercados por valores baixos.
- **High–Low:** valores altos cercados por valores baixos (outliers).
- **Low–High:** valores baixos cercados por valores altos (outliers).

Esses padrões permitem identificar desigualdades regionais e concentrações espaciais significativas no consumo de pescados e nos indicadores produtivos municipais.

6.6.4 Análise de Hot Spot (Getis-Ord G_i^*)

A estatística Getis–Ord G_i^* identifica áreas com concentração localmente significativa de valores altos (*hot spots*) ou baixos (*cold spots*), considerando a vizinhança espacial de cada unidade geográfica.

O índice pode ser definido como:

$$G_i^* = \frac{\sum_j w_{ij} x_j}{\sum_j x_j},$$

ou, em sua forma padronizada,

$$G_i^* = \frac{\sum_j w_{ij} x_j - \bar{X} \sum_j w_{ij}}{S \sqrt{\frac{n \sum_j w_{ij}^2 - (\sum_j w_{ij})^2}{n-1}}},$$

onde w_{ij} são os elementos da matriz de pesos espaciais, $\bar{X}$ é a média de x_j , S é o desvio-padrão e n é o número de unidades espaciais.

Matriz de Vizinhança: Critério da Rainha

Para esta análise, utilizou-se a matriz de pesos espaciais com contiguidade do tipo *Rainha* (Queen contiguity). Esse critério considera dois municípios vizinhos se eles compartilham **qualquer ponto de fronteira**, seja lado ou vértice, produzindo uma estrutura de vizinhança mais inclusiva do que o critério da Torre (*Rook*), que considera apenas fronteiras laterais.

Formalmente, a matriz de pesos é definida como:

$$w_{ij} = \begin{cases} 1, & \text{se as unidades } i \text{ e } j \text{ compartilham lado ou vértice;} \\ 0, & \text{caso contrário.} \end{cases}$$

A matriz é posteriormente padronizada por linha:

$$w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}}.$$

Justificativa do Uso do Critério da Rainha

O critério da Rainha foi adotado porque:

- garante uma representação mais realista da interação territorial entre municípios paulistas, cuja geometria apresenta limites irregulares;
- evita a subestimação da conectividade em regiões com municípios muito fragmentados;
- é recomendado para análises de autocorrelação local (como LISA e Gi*) quando se deseja capturar padrões espaciais amplos e sistemas de dependência mais completos.

Assim, o uso da vizinhança do tipo Rainha permite detectar com maior sensibilidade clusters espaciais de alto e baixo valor, proporcionando maior robustez à avaliação de padrões espaciais da produção aquícola e sua possível associação com o consumo de pescado em São Paulo.

6.6.5 Integração com Variáveis Exógenas da Aquicultura

As variáveis exógenas selecionadas do SIDRA/IBGE foram espacializadas e comparadas com o consumo de pescados para avaliar a coerência territorial entre oferta e demanda. As variáveis foram:

- Quantidade produzida em kg de produtos aquícolas;
- Valor em reais da produção aquícola;
- Número de estabelecimentos de produção aquícola.

Essa integração permitiu verificar se municípios com maior produção ou maior estrutura produtiva apresentavam padrões diferenciados de consumo, e se tais relações exibiam dependência espacial significativa conforme as estatísticas LISA e Gi*.

7 RESULTADOS

7.1 PREPARAÇÃO DA BASE DE DADOS

A preparação dos dados teve como objetivo consolidar informações referentes aos domicílios paulistas a partir dos microdados da Pesquisa de Orçamentos Familiares (POF) 2017–2018. Todas as etapas foram conduzidas de forma sistemática para garantir consistência, integridade e padronização das informações antes das análises estatísticas e espaciais.

7.1.1 Filtragem e identificação dos indivíduos

Foram utilizadas cinco bases distintas da POF: MORADOR, MORADOR_QUALI_VIDA, CONDICOES_VIDA, CARACTERISTICAS_DIETA e CONSUMO_ALIMENTAR. A filtragem inicial restringiu o conjunto de dados aos registros pertencentes ao estado de São Paulo, identificados pelo código da Unidade da Federação igual a 35.

Para garantir a correta ligação entre as bases, foi criado um identificador único composto pela combinação das variáveis COD_UPA, número do domicílio, unidade de consumo e código do informante. Tal identificador viabilizou a integração robusta das bases, mantendo coerência entre os registros individuais e domiciliares.

7.1.2 Seleção de variáveis relevantes

Após a filtragem, cada base foi reduzida ao conjunto de variáveis necessárias para os objetivos do estudo. A base MORADOR forneceu informações como idade, anos de escolaridade, localização urbana/rural e rendimentos (total e per capita). A base MORADOR_QUALI_VIDA contribuiu com indicadores de insegurança alimentar e percepção de suficiência de renda.

A base CONDICOES_VIDA foi essencial para as variáveis subjetivas relacionadas ao padrão de vida e avaliação dos serviços públicos (educação, saúde, moradia, alimentação, lazer e vestuário). Já a base CARACTERISTICAS_DIETA trouxe informações sobre a adoção de dietas específicas e seus respectivos motivos. Por fim, a base CONSUMO_ALIMENTAR forneceu os dados do diário alimentar, itens consumidos e suas quantidades.

7.1.3 Classificação dos alimentos e criação de indicadores

A classificação dos alimentos foi realizada a partir de uma planilha complementar que agrupa os itens alimentares por categorias, tais como: *peixes frescos*, *carnes processadas*, *leguminosas*, *ovos*, entre outras.

Com base nessa classificação, foram criadas variáveis de frequência e indicadores binários de consumo por grupo alimentar. A base foi convertida para o formato *wide*, garantindo que cada linha representasse um indivíduo com suas respectivas informações de consumo.

Um destaque importante foi a criação da variável `Peixe_geral`, construída pela soma das quantidades consumidas de todos os subgrupos de pescados, funcionando como um indicador consolidado do consumo de peixe.

7.1.4 Integração e criação de novas variáveis

Após a integração das bases, foram criadas variáveis derivadas para ampliar a capacidade analítica do conjunto. A variável `Renda_Total_Classes` categoriza a renda domiciliar total em seis faixas, permitindo comparações entre estratos socioeconômicos. Variáveis categóricas como sexo, cor, ocupação e percepção de qualidade de vida foram convertidas para fatores.

As variáveis ordinais provenientes de avaliações subjetivas (por exemplo, qualidade dos serviços públicos) foram recodificadas de modo que valores maiores correspondessem a piores avaliações, facilitando a interpretação das técnicas multivariadas posteriores.

Ao final do processamento, a base consolidada foi exportada em formatos `.csv` e `.xlsx`, contendo 1.333 domicílios paulistas, cada qual representado por uma linha com informações completas e tratadas para as análises subsequentes.

7.2 DESCRIÇÃO DAS VARIÁVEIS UTILIZADAS

A seleção das variáveis considerou sua relevância para compor indicadores latentes de padrão de vida, acesso a serviços públicos e perfil socioeconômico. Todas as variáveis representam a pessoa de referência do domicílio, exceto a renda total, que se refere ao rendimento agregado da Unidade de Consumo (UC).

7.2.1 Variáveis Qualitativas Ordinais

As variáveis ordinais possuem categorias ordenadas, permitindo avaliar graduações de qualidade percebida ou níveis de acesso. A Tabela 1 apresenta a descrição completa das variáveis utilizadas.

Tabela 1 – Variáveis qualitativas ordinais do Estado de SP – POF (2017–2018)

Variável	Categorias
Padrao_Alimentacao	Ruim; Satisfatório; Bom
Padrao_Moradia	Ruim; Satisfatório; Bom
Padrao_Vestuario	Ruim; Satisfatório; Bom
Padrao_Educacao	Ruim; Satisfatório; Bom
Padrao_Saude	Ruim; Satisfatório; Bom
Padrao_Lazer	Ruim; Satisfatório; Bom
Fornecimento_Agua	Não há; Ruim; Satisfatório; Bom
Fornecimento_Energia	Não há; Ruim; Satisfatório; Bom
Servico_Iluminacao	Não há; Ruim; Satisfatório; Bom
Coleta_Lixo	Não há; Ruim; Satisfatório; Bom
Manutencao_Rua	Não há; Ruim; Satisfatório; Bom
Escoamento_Agua	Não há; Ruim; Satisfatório; Bom
Esgotamento_Sanitario	Não há; Ruim; Satisfatório; Bom
Transporte_Coletivo	Não há; Ruim; Satisfatório; Bom
Renda_Total_Classes	[F] Até R\$1.908; [E] 1.908–2.862; [D] 2.862–5.724; [C] 5.724–9.540; [B] 9.540–14.310; [A] acima de 14.310
Idade_Cat	Até 24; 25–49; 50–64; 65+
Nivel_Instrucao	Sem instrução; EF incompleto; EF completo; EM incompleto; EM completo; ES incompleto; ES completo

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

7.2.2 Variáveis Qualitativas Nominais

As variáveis nominais descrevem características sem ordem natural entre suas categorias, sendo usadas para caracterização dos domicílios e análises de associações.

Tabela 2 – Variáveis qualitativas nominais do Estado de SP – POF (2017–2018)

Variável	Categorias
Local	Urbano; Rural
Ocupacao	Empregado doméstico; Empregado setor privado; Militar/setor público; Conta própria; Empregador; Fora da força de trabalho
Composicao_Familiar	Um adulto sem criança; Um adulto com criança; Mais de um adulto sem criança; Mais de um adulto com criança; Idosos com ou sem crianças; Idosos com adultos e/ou crianças
Faz_Dieta	Sim; Não

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

7.3 ANÁLISE DESCRITIVA

A análise descritiva teve como propósito caracterizar as variáveis qualitativas ordinais e nominais utilizadas no estudo. As variáveis ordinais foram analisadas em termos de distribuição das categorias ordenadas, permitindo identificar percepções de qualidade e acesso a serviços. Já as variáveis nominais foram avaliadas pelas frequências absolutas e relativas. As Tabelas 3 e 4 apresentam os resultados obtidos.

7.3.1 Tabelas Descritivas

As tabelas descritivas apresentam a distribuição das variáveis qualitativas utilizadas no estudo, permitindo uma visão detalhada das características socioeconômicas, demográficas e de percepção dos domicílios paulistas. A partir das frequências absolutas e relativas, é possível identificar padrões gerais, níveis de qualidade percebida e diferenças entre grupos, oferecendo uma base sólida para as etapas posteriores da análise. Essas informações complementam os gráficos apresentados anteriormente e auxiliam na interpretação do perfil das famílias analisadas.

Tabela 3 – Análise descritiva das variáveis qualitativas ordinais – SP (POF 2017–2018)

Variável	Categoria	Contagem	Proporção (%)
Padrao_Alimentacao	Ruim	75	5.63
	Satisfatório	439	32.93
	Bom	819	61.44
Padrao_Moradia	Ruim	76	5.70
	Satisfatório	334	25.06
	Bom	923	69.24
Padrao_Vestuario	Ruim	110	8.25
	Satisfatório	422	31.66
	Bom	801	60.09
Padrao_Educacao	Ruim	164	12.30
	Satisfatório	364	27.31
	Bom	805	60.39
Padrao_Saude	Ruim	304	22.81
	Satisfatório	389	29.18
	Bom	640	48.01
Padrao_Lazer	Ruim	462	34.66
	Satisfatório	398	29.86
	Bom	473	35.48
Fornecimento_Agua	Não há	66	4.95
	Ruim	69	5.18
	Satisfatório	172	12.90
	Bom	1026	76.97

Variável	Categoria	Contagem	Proporção (%)
Fornecimento_Energia	Não há	6	0.45
	Ruim	59	4.43
	Satisfatório	198	14.85
	Bom	1070	80.27
Servico_Iluminacao	Não há	162	12.15
	Ruim	200	15.00
	Satisfatório	291	21.83
	Bom	680	51.01
Coleta_Lixo	Não há	94	7.05
	Ruim	56	4.20
	Satisfatório	183	13.73
	Bom	1000	75.02
Manutencao_Rua	Não há	299	22.43
	Ruim	307	23.03
	Satisfatório	283	21.23
	Bom	444	33.31
Escoamento_Agua	Não há	132	9.90
	Ruim	218	16.35
	Satisfatório	209	15.68
	Bom	774	58.06
Esgotamento_Sanitario	Não há	168	12.60
	Ruim	100	7.50
	Satisfatório	187	14.03
	Bom	878	65.87

Variável	Categoria	Contagem	Proporção (%)
Transporte_Coletivo	Não há	160	12.00
	Ruim	267	20.03
	Satisfatório	336	25.21
	Bom	570	42.76
Renda_Total_Classes	[F] Até R\$1.908	178	13.43
	[E] R\$1.908–2.862	233	17.48
	[D] R\$2.862–5.724	455	34.13
	[C] R\$5.724–9.540	249	18.68
	[B] R\$9.540–14.310	101	7.58
	[A] Mais de R\$14.310	116	8.70
Idade_Cat	Até 24	34	2.55
	25–49	616	46.21
	50–64	415	31.13
	65 ou mais	268	20.11
Nivel_Instrucao	Sem instrução	30	2.25
	EF incompleto	431	32.33
	EF completo	123	9.23
	EM incompleto	61	4.58
	EM completo	402	30.16
	ES incompleto	59	4.43
	ES completo	227	17.03

A análise das variáveis qualitativas ordinais, apresentadas na Tabela 3, indica que a maior parte dos domicílios paulistas apresenta condições favoráveis nos diferentes domínios avaliados. Os padrões de alimentação, moradia, vestuário e educação são majoritariamente classificados como “bons” ou “satisfatórios”, revelando que a maior parte das famílias dispõe de condições básicas adequadas nesses aspectos. Serviços essenciais, como fornecimento de água, energia elétrica e coleta de lixo, também exibem ampla cobertura e boa qualidade, reforçando a presença consolidada desses serviços no estado.

Por outro lado, alguns domínios evidenciam maior heterogeneidade e vulnerabilidade, especi-

almente saúde, lazer, iluminação pública, manutenção das ruas, escoamento de água e transporte coletivo, nos quais há proporções mais significativas de avaliações “ruim” ou de ausência do serviço. Esses resultados refletem desigualdades estruturais presentes no território paulista, sobretudo em áreas periféricas ou de menor infraestrutura urbana.

No que se refere às características socioeconômicas, observa-se predominância de famílias pertencentes às classes D e C, seguidas pelas classes E e F. Em termos de escolaridade, há grande diversidade, com destaque para indivíduos com ensino fundamental incompleto e ensino médio completo, embora um contingente relevante apresente ensino superior completo. O perfil etário dos responsáveis concentra-se entre 25 e 64 anos, faixa típica de adultos economicamente ativos.

De modo geral, os resultados sintetizados na Tabela 3 apontam para condições de vida relativamente favoráveis em grande parte dos domicílios, porém acompanhadas de desigualdades importantes em serviços urbanos e aspectos relacionados à saúde e ao lazer, o que pode influenciar diretamente padrões de consumo e bem-estar das famílias paulistas.

Tabela 4 – Análise descritiva das variáveis qualitativas nominais – SP (POF 2017–2018)

Variável	Categoria	Contagem	Proporção (%)
Local	Urbano	1082	81.17
	Rural	251	18.83
Ocupacao	Empregado doméstico	90	6.75
	Empregado setor privado	470	35.26
	Militar/setor público	102	7.65
	Conta própria	263	19.73
	Empregador	61	4.58
	Fora da força de trabalho	347	26.03

Variável	Categoria	Contagem	Proporção (%)
Composicao_Familiar	Um adulto sem criança	103	7.73
	Um adulto com criança	29	2.18
	Mais de um adulto sem criança	427	32.03
	Mais de um adulto com criança	429	32.18
	Idosos com ou sem crianças	146	10.95
	Idosos com adultos e/ou crianças	199	14.93
Faz_Dieta	Sim	239	17.93
	Não	1094	82.07

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Tabela 4 revela padrões estruturais importantes da população paulista analisada. O forte predomínio de domicílios urbanos (81,17%) indica que o consumo e os hábitos alimentares observados tendem a refletir características de contextos urbanos, como maior acesso a mercados, maior diversidade de produtos e rotinas mais aceleradas. O perfil ocupacional mostra uma população majoritariamente inserida no setor privado ou em atividades autônomas, o que sugere um mercado de trabalho dinâmico, porém marcado por diferentes níveis de estabilidade econômica — aspecto relevante para o entendimento do poder aquisitivo e, consequentemente, do consumo alimentar.

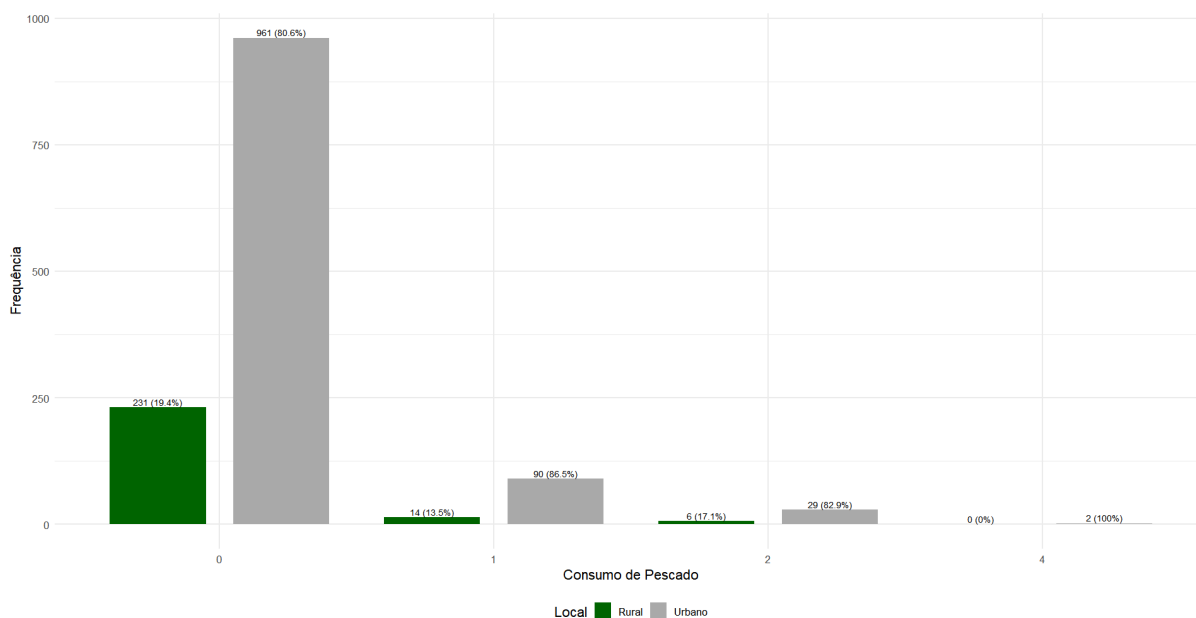
A composição familiar, concentrada em domicílios com mais de um adulto (com ou sem crianças), aponta para arranjos familiares estáveis e potencialmente com divisão de responsabilidades, o que pode influenciar tanto o planejamento das compras quanto o tipo de alimentos adquiridos. Já a baixa proporção de indivíduos que fazem dieta (17,93%) sugere que práticas alimentares restritas ou orientadas para saúde não são predominantes no grupo estudado, o que pode repercutir no padrão geral de consumo observado nas análises posteriores.

7.3.2 Gráficos Descritivos

Os gráficos descritivos apresentados nesta subseção complementam as tabelas anteriores ao ilustrar visualmente a distribuição do consumo de pescado entre diferentes esferas. A partir deles, torna-se possível identificar padrões gerais, contrastes sutis entre categorias e eventuais

tendências que não seriam plenamente perceptíveis apenas por meio de frequências numéricas. Cada figura é acompanhada de uma interpretação concisa, de modo a facilitar a compreensão dos principais achados descritivos da POF 2017–2018 no estado de São Paulo.

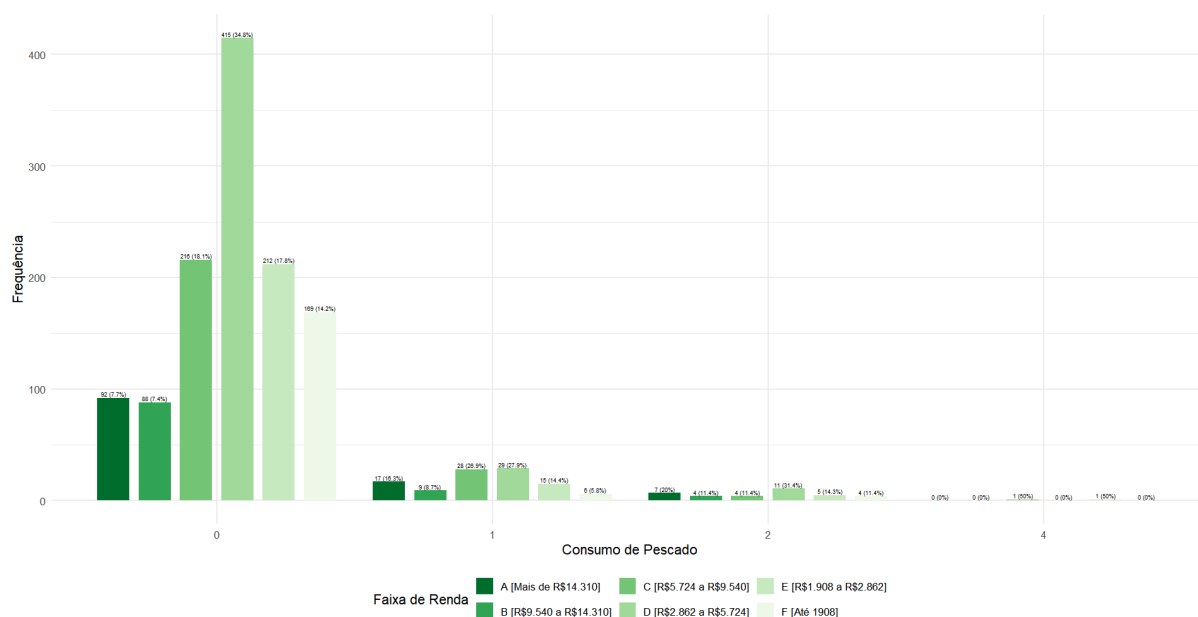
Figura 1 – Consumo de Pescado por Local do Domicílio no Estado de SP – POF (2017–2018)



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 1 mostra que o consumo de pescado não apresenta diferenças expressivas entre as áreas rural e urbana do estado de São Paulo, segundo os dados da POF 2017–2018. Embora a maior parte dos não consumidores esteja concentrada na zona urbana, esse resultado reflete, em parte, a maior proporção da população residente nesse meio. Ao observar as categorias de consumo (1 vez, 2 vezes ou mais), nota-se que a distribuição entre rural e urbano permanece relativamente semelhante, indicando que o local de moradia não exerce influência significativa sobre os hábitos de consumo de pescado no estado.

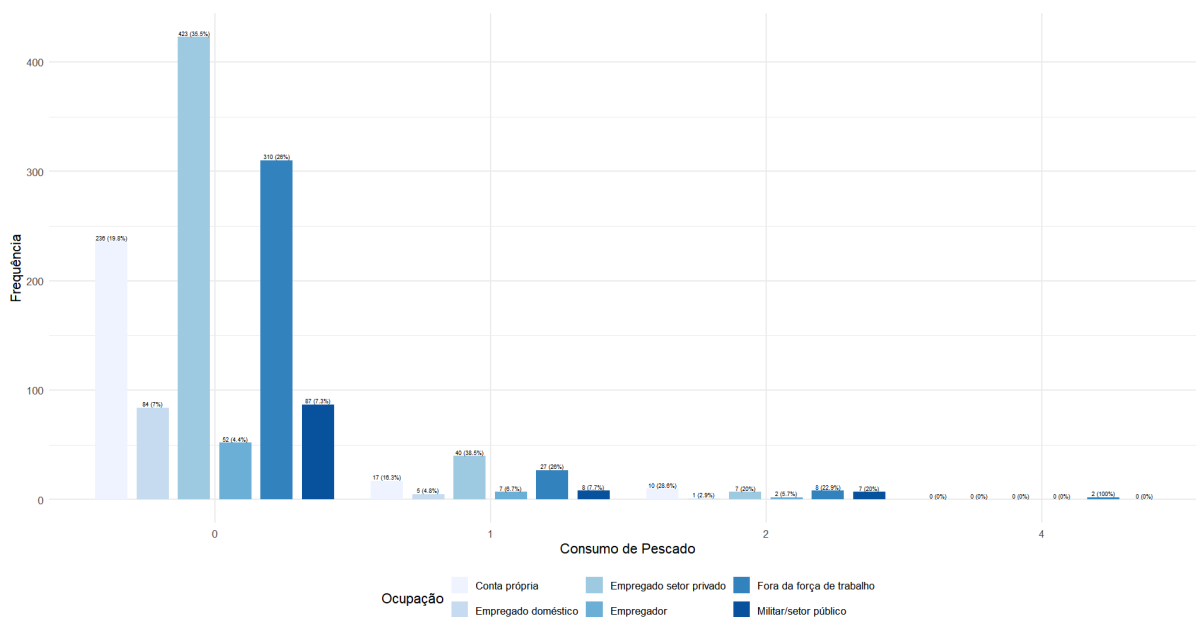
Figura 2 – Consumo de Pescado por Faixa de Renda Familiar no Estado de SP – POF (2017–2018)



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 2 mostra que o consumo de pescado no estado de São Paulo, segundo a POF 2017–2018, não apresenta grandes variações entre as diferentes faixas de renda familiar. Embora exista uma leve tendência de maior consumo nas classes de renda mais elevadas, observa-se que a maior parte dos indivíduos, independentemente do nível de renda, declarou não consumir pescado. As demais categorias de consumo (1 vez, 2 ou mais vezes) também exibem distribuições semelhantes entre as faixas, sugerindo que a renda familiar não se configura como um fator determinante para o consumo de pescado no período analisado.

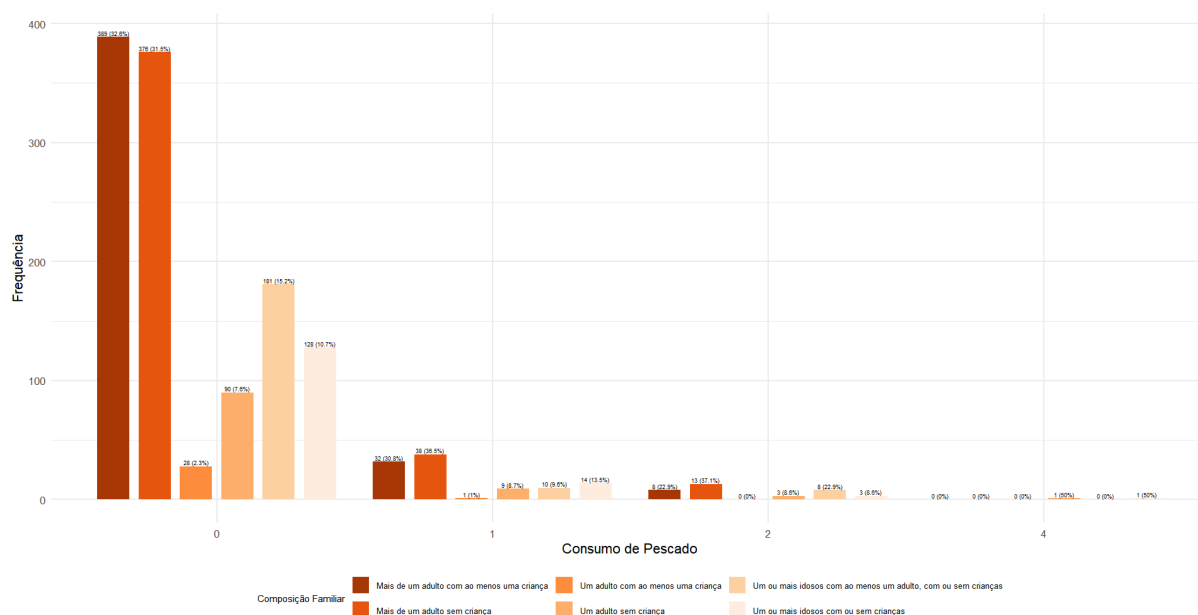
Figura 3 – Consumo de Pescado por Ocupação no Estado de SP – POF (2017–2018)



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 3 mostra que o consumo de pescado no estado de São Paulo, de acordo com a ocupação, não apresenta grandes variações entre os diferentes grupos ocupacionais, segundo os dados da POF 2017–2018. Embora a maior parte dos não consumidores esteja concentrada entre os indivíduos fora da força de trabalho e entre empregados do setor privado — o que pode refletir a composição da amostra —, observa-se que os padrões de consumo nos demais grupos permanecem relativamente semelhantes. As diferenças nas categorias de consumo (1 vez, 2 vezes ou mais) são pequenas, sugerindo que a ocupação não se configura como um fator determinante nos hábitos de consumo de pescado no período analisado.

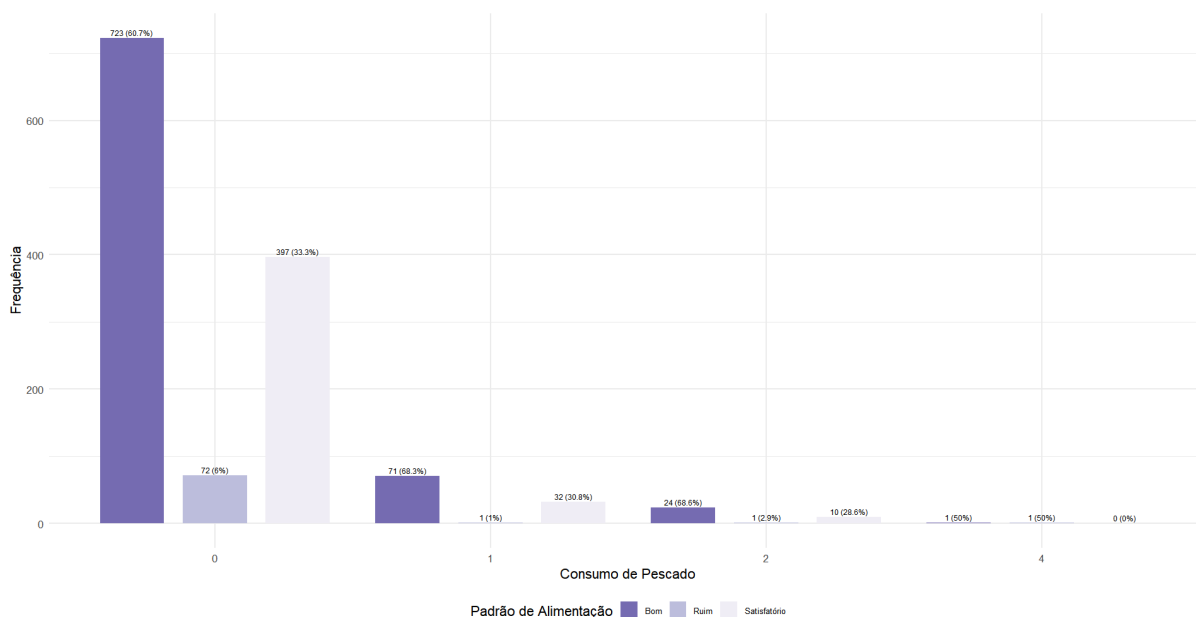
Figura 4 – Consumo de Pescado por Composição Familiar no Estado de SP – POF (2017–2018)



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 4 mostra que o consumo de pescado no estado de São Paulo, segundo a composição familiar, apresenta pouca variação entre os diferentes arranjos familiares, conforme os dados da POF 2017–2018. A maior parte dos indivíduos, independentemente da estrutura familiar — como famílias com crianças, sem crianças, com idosos ou compostas por apenas um adulto — declarou não consumir pescado. Nas categorias de consumo de 1 ou mais vezes, as proporções entre os diferentes grupos se mantêm próximas, sugerindo que a composição familiar não exerce influência significativa sobre o padrão de consumo de pescado no período analisado.

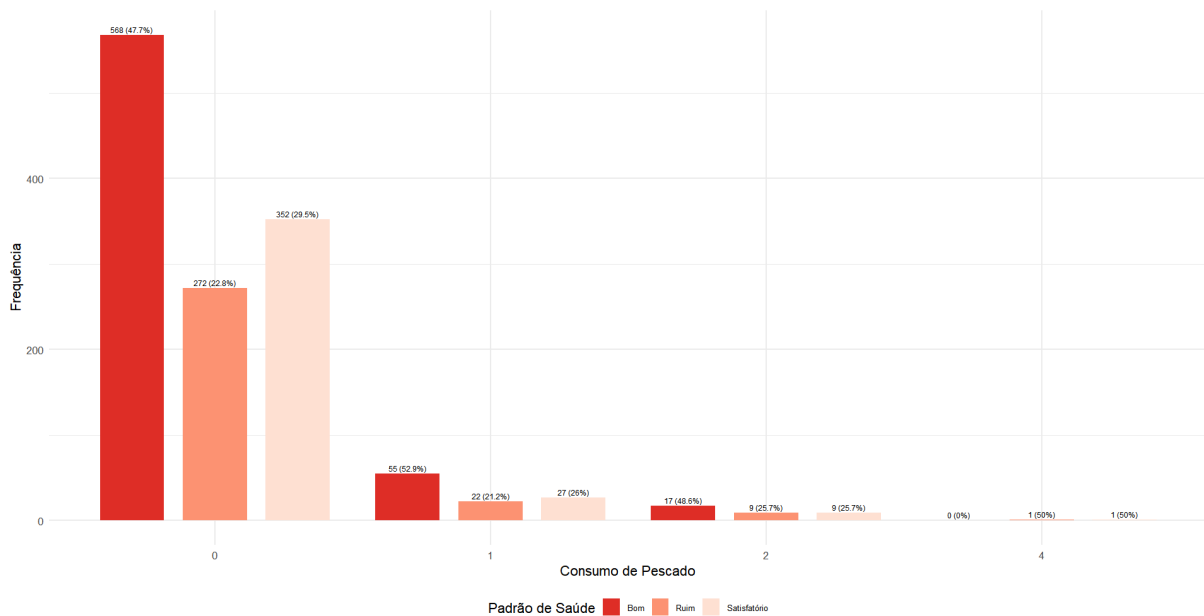
Figura 5 – Consumo de Pescado por Padrão de Alimentação no Estado de SP – POF (2017–2018)



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 5 mostra o consumo de pescado no estado de São Paulo de acordo com o padrão de alimentação autorreferido, com base nos dados da POF 2017–2018. Nota-se que a maior parte dos indivíduos, independentemente de terem um padrão alimentar classificado como bom, satisfatório ou ruim, declarou não consumir pescado. Nas demais categorias de consumo (1, 2 ou mais vezes), a distribuição entre os grupos permanece relativamente próxima. Isso indica que o padrão de alimentação geral da pessoa não está diretamente associado ao consumo de pescado, sugerindo que outros fatores podem ser mais determinantes nesse hábito alimentar.

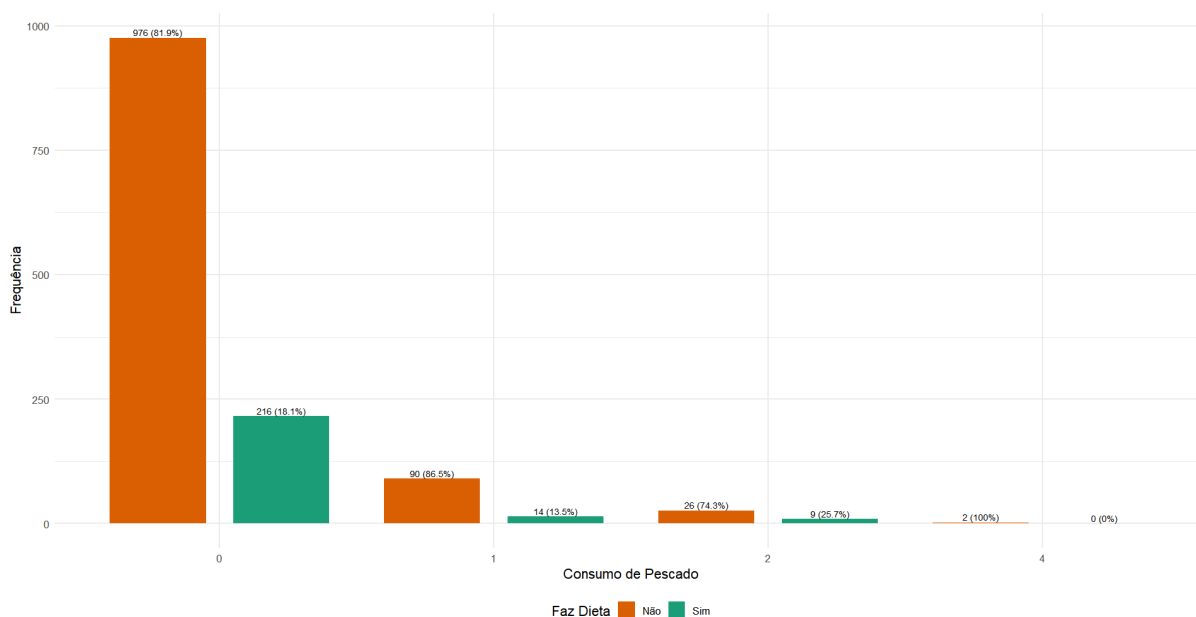
Figura 6 – Consumo de Pescado por Padrão de Saúde no Estado de SP – POF (2017–2018)



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 6 apresenta o consumo de pescado no estado de São Paulo segundo o padrão de saúde autorreferido, com base nos dados da POF 2017–2018. Observa-se que, independentemente da percepção do padrão de saúde como bom, ruim ou satisfatório, a maior parte dos indivíduos declarou não consumir pescado. Nas demais categorias de consumo (1 vez, 2 ou mais vezes), a frequência é consideravelmente menor e distribui-se de forma semelhante entre os diferentes grupos de avaliação de saúde. Esses resultados sugerem que a autopercepção da saúde não está fortemente associada ao consumo de pescado, indicando que outros fatores — como preferências alimentares, renda ou acesso ao alimento — podem exercer maior influência sobre esse comportamento.

Figura 7 – Consumo de Pescado por Prática de Dieta no Estado de SP – POF (2017–2018)



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 7 mostra o consumo de pescado no estado de São Paulo segundo a prática de dieta, com base nos dados da POF 2017–2018. Percebe-se que, tanto entre os indivíduos que declararam seguir uma dieta quanto entre os que não seguem, a maioria afirma não consumir pescado. Nas demais categorias de consumo (1 vez, 2 ou mais vezes), observa-se que a frequência é bastante inferior e distribuída de forma semelhante entre os dois grupos. Esses resultados indicam que a prática de dieta não está diretamente relacionada ao consumo de pescado, sugerindo que a inclusão desse alimento na rotina alimentar pode depender de outros fatores, como preferências individuais, orientação nutricional ou disponibilidade.

7.4 AVALIAÇÃO DE CONSISTÊNCIA INTERNA – ALFA DE CRONBACH

A avaliação da consistência interna das variáveis qualitativas ordinais foi realizada por meio do coeficiente alfa de Cronbach, com o objetivo de verificar o grau de coerência entre os itens que compõem o conjunto de indicadores subjetivos utilizados neste estudo. As variáveis consideradas incluem percepções sobre padrão de vida, qualidade dos serviços públicos, condições do domicílio e características demográficas da pessoa de referência da Unidade de Consumo (UC), todas estruturadas em escalas ordinais.

Para a construção da matriz de correlações utilizada no cálculo do alfa, empregou-se a correlação policórica, apropriada para variáveis ordinais por assumir que as categorias observadas representam cortes de um contínuo subjacente. Essa abordagem proporciona estimativas mais adequadas de associação entre itens quando comparada à correlação de Pearson, especialmente em escalas discretas e ordenadas.

De acordo com Hair et al. (2009), valores de alfa iguais ou superiores a 0,70 são considerados satisfatórios, indicando boa consistência interna entre os itens avaliados. No presente estudo, o coeficiente obtido para o estado de São Paulo foi:

$$\alpha = 0,770$$

Esse resultado evidencia que o conjunto de variáveis adotado apresenta um nível adequado de confiabilidade, sugerindo que elas medem de forma coerente um mesmo conjunto de percepções subjacentes relacionadas às condições de vida dos domicílios paulistas. Dessa forma, o uso dessas variáveis nas análises multivariadas subsequentes é considerado metodologicamente consistente.

7.5 ANÁLISE FATORIAL

7.5.1 Medida de Adequação da Amostragem (KMO)

A adequação dos dados para a aplicação da Análise Fatorial foi avaliada por meio da Medida de Kaiser-Meyer-Olkin (KMO). Esse índice varia entre 0 e 1 e expressa o grau em que as correlações parciais entre as variáveis são pequenas, condição desejável para que a análise fatorial seja apropriada. De acordo com Hair et al. (2009), valores de KMO acima de 0,70 indicam boa adequação, enquanto valores superiores a 0,80 são considerados excelentes.

No presente estudo, o valor obtido para o estado de São Paulo foi:

$$KMO_{SP} = 0,791,$$

o que representa uma boa adequação amostral. Esse resultado confirma que as variáveis ordinais selecionadas compartilham variância suficiente para justificar a extração de fatores latentes, conferindo confiabilidade às etapas subsequentes da análise fatorial.

7.5.2 Seleção do Número de Fatores e Variância Explicada

Como as variáveis utilizadas na análise são de natureza ordinal, adotou-se a matriz de correlações policóricas como base para a extração dos fatores. Esse tipo de correlação é recomendado para variáveis ordinais, pois pressupõe a existência de um contínuo latente subjacente às categorias observadas, resultando em estimativas mais adequadas das relações entre os itens.

A partir da matriz policórica, foram obtidos os autovalores associados a cada fator. Esses autovalores representam a quantidade de variância explicada pelo respectivo fator. De acordo com o critério de Kaiser, são retidos apenas os fatores com autovalores superiores a 1, pois esses fatores explicam mais variância do que uma única variável padronizada.

Para cada fator selecionado, calculou-se a proporção de variância explicada individualmente e a variância acumulada, permitindo avaliar o grau de representatividade do modelo fatorial em relação à variabilidade total dos dados. Esses indicadores auxiliam na escolha do número

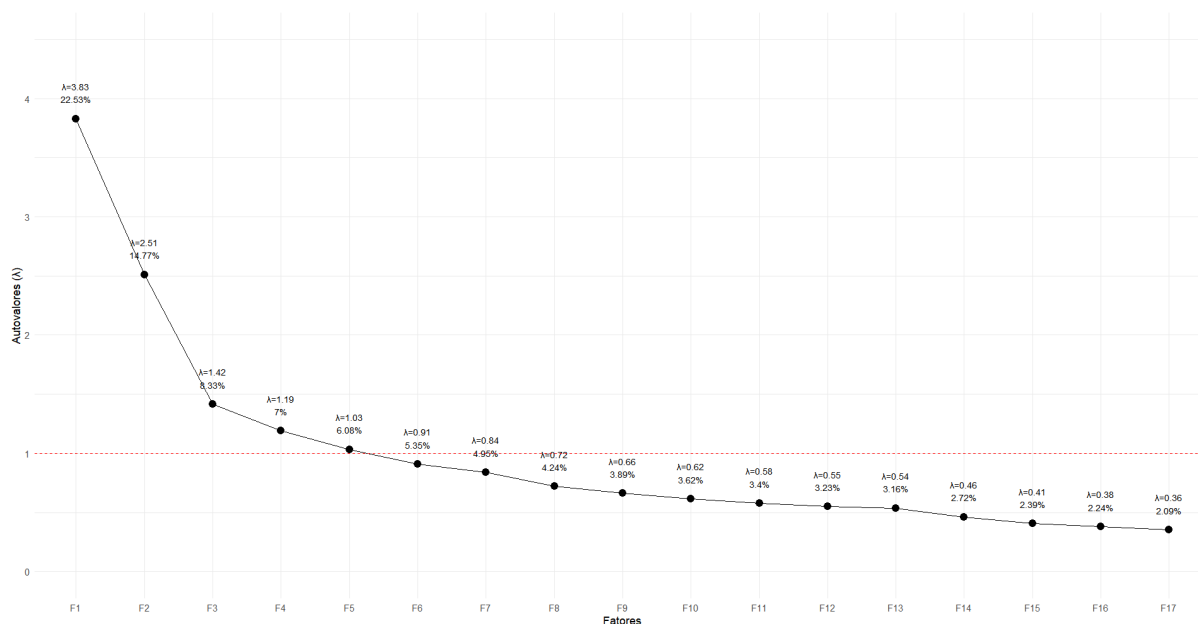
adequado de fatores e no entendimento da capacidade de síntese dos dados proporcionada pela solução fatorial.

Tabela 5 – Autovalores e Variância por Fator – SP

Fator	Autovalor	% da Variância Explicada	% da Variância Explicada Acumulada
F1	3.829	22.53	22.53
F2	2.511	14.77	37.30
F3	1.416	8.33	45.63
F4	1.189	7.00	52.63
F5	1.034	6.08	58.71
F6	0.910	5.35	64.06
F7	0.841	4.95	69.00
F8	0.721	4.24	73.25
F9	0.662	3.89	77.14
F10	0.615	3.62	80.76
F11	0.579	3.40	84.16
F12	0.549	3.23	87.39
F13	0.538	3.16	90.56
F14	0.463	2.72	93.28
F15	0.407	2.39	95.67
F16	0.381	2.24	97.91
F17	0.355	2.09	100.00

Fonte: Elaborada pela autora.

Figura 8 – Scree Plot dos Autovalores para Determinação do Número de Fatores – SP



Fonte: Elaborada pela autora.

A Tabela 5 e Figura 8 apresentam os autovalores e a variância explicada pelos fatores extraídos para o estado de São Paulo. Observa-se que cinco fatores apresentam autovalores superiores a 1, atendendo ao critério de Kaiser e justificando sua retenção na análise fatorial. Esses cinco fatores explicam conjuntamente 58,71% da variância total das variáveis ordinais, indicando que uma parcela substancial da informação presente nos dados é representada pela estrutura fatorial obtida. O primeiro fator, isoladamente, responde por 22,53% da variância, refletindo sua maior capacidade explicativa. A variância acumulada demonstra uma redução progressiva da contribuição dos demais fatores, coerente com a expectativa de modelos fatoriais extraídos a partir de dados sociais e perceptivos.

Tabela 6 – Cargas Fatoriais – Análise Fatorial (PCA com rotação varimax) – SP

Variável	RC1	RC2	RC3	RC4	RC5
Padrao_Alimentacao	-0.074	0.671	0.221	0.094	0.159
Padrao_Moradia	-0.020	0.653	0.113	0.102	0.271
Padrao_Vestuario	-0.001	0.690	0.172	0.055	0.195
Padrao_Educacao	0.136	0.673	-0.049	0.014	-0.106
Padrao_Saude	0.173	0.655	-0.005	0.002	-0.279
Padrao_Lazer	0.067	0.666	0.022	0.029	-0.165
Fornecimento_Agua	0.100	0.039	0.127	0.822	0.028
Fornecimento_Energia	0.105	0.172	-0.239	0.670	-0.074
Servico_Iluminacao	0.776	0.033	0.056	0.203	0.051
Coleta_Lixo	0.793	-0.019	0.049	0.054	0.029
Manutencao_Rua	0.675	0.113	0.122	0.057	0.013
Escoamento_Agua	0.589	-0.003	0.115	0.399	0.175
Esgotamento_Sanitario	0.501	0.018	0.227	0.547	0.088
Transporte_Coletivo	0.664	0.116	-0.127	-0.042	-0.117
Renda_Total_Classes	0.083	0.237	0.782	0.002	0.104
Idade_Cat	0.096	0.022	-0.119	0.007	0.841
Nivel_Instrucao	0.114	0.090	0.794	0.030	-0.364

Fonte: Elaborada pela autora.

A Tabela 6 sintetiza as cargas fatoriais resultantes da Análise Fatorial conduzida com extração por PCA e rotação varimax. A utilização da varimax permitiu maximizar a variância das cargas dentro de cada componente, facilitando a identificação de padrões estruturais entre as variáveis. Observa-se que os fatores extraídos concentram variáveis com cargas elevadas em dimensões específicas, revelando agrupamentos conceitualmente coerentes.

De forma resumida, os fatores captam: (i) condições de infraestrutura urbana; (ii) percepções de bem-estar no domicílio; (iii) características socioeconômicas; (iv) acesso a serviços essenciais; e (v) perfis demográficos. A separação nítida dos conjuntos de variáveis indica que a técnica conseguiu isolar dimensões latentes distintas, reforçando a adequação da análise fatorial para representar os múltiplos aspectos do contexto domiciliar no estado de São Paulo.

Tabela 7 – Fatores Latentes Extraídos para o Estado de São Paulo

Fator	Variáveis (Cargas Fatoriais)
Fator 1 Infraestrutura Urbana e Serviços Públicos	Servico_Iluminacao (0.776); Coleta_Lixo (0.793); Manutencao_Rua (0.675); Transporte_Coletivo (0.664); Escoamento_Agua (0.589); Esgotamento_Sanitario (0.501)
Fator 2 Consumo e Bem-Estar Doméstico	Padrao_Alimentacao (0.671); Padrao_Moradia (0.653); Padrao_Vestuario (0.690); Padrao_Educacao (0.673); Padrao_Saude (0.655); Padrao_Lazer (0.666)
Fator 3 Condição Socioeconômica	Renda_Total_Classes (0.782); Nivel_Instrucao (0.794)
Fator 4 Saneamento e Serviços Essenciais	Fornecimento_Agua (0.822); Fornecimento_Energia (0.670); Esgotamento_Sanitario (0.547)
Fator 5 Perfil Demográfico (Idade)	Idade_Cat (0.841)

Fonte: Elaborada pela autora.

A Tabela 7 apresenta a estrutura fatorial extraída para o estado de São Paulo, evidenciando cinco dimensões interpretáveis a partir das cargas mais elevadas em cada componente. O Fator 1 agrupa variáveis relacionadas à infraestrutura urbana e à oferta de serviços públicos, indicando que esses elementos formam um eixo comum de variação entre os domicílios. O Fator 2 reúne indicadores associados ao bem-estar doméstico e às condições gerais de vida percebidas pelos entrevistados, refletindo a avaliação subjetiva do padrão de vida. O Fator 3 concentra atributos de natureza socioeconômica, especialmente renda e escolaridade, sugerindo uma dimensão claramente distinta das demais. O Fator 4 agrega variáveis ligadas ao saneamento e ao acesso a serviços essenciais, compondo uma dimensão específica da qualidade das condições básicas do domicílio. Por fim, o Fator 5 está fortemente associado à idade da pessoa de referência, formando um componente demográfico independente.

Essa organização fatorial mostra que os diferentes aspectos do padrão de vida não se distribuem aleatoriamente, mas se estruturam em dimensões coerentes, úteis para análises posteriores — incluindo clusterização e associações com o consumo de pescado.

7.6 ANÁLISE DE CORRESPONDÊNCIA

Foi realizada uma Análise de Correspondência (AC) com o objetivo de identificar padrões de associação entre o consumo de pescado e variáveis socioeconômicas e demográficas para o estado de São Paulo (SP), a partir dos dados da POF 2017–2018.

O consumo de pescado foi inicialmente representado pela variável *Peixe_geral*, que expressa a frequência de ingestão registrada no período de referência da pesquisa. Para permitir a aplicação da técnica e facilitar a interpretação dos perfis de resposta, essa variável foi recodificada na forma ordinal, originando *Peixe_geral_cat*, com quatro categorias: “Não há” (ausência de consumo), “Baixo”, “Médio” e “Alto”, agrupando os níveis mais elevados de frequência.

Além disso, foram incorporados à AC os escores fatoriais derivados da Análise Fatorial (PCA com rotação varimax). Esses escores foram convertidos em variáveis categóricas ordinais por meio do algoritmo *k*-means ($k = 3$), gerando categorias ordenadas (“Baixo”, “Médio” e “Alto”) para cada dimensão latente identificada, como infraestrutura urbana, bem-estar doméstico, condição socioeconômica e serviços essenciais. A ordenação dos clusters foi realizada com base na média dos escores fatoriais, garantindo coerência conceitual entre os fatores.

Também foram incluídas variáveis nominais observadas diretamente, tais como o tipo de ocupação do responsável pelo domicílio (*Ocupacao*), a composição familiar (*Composicao_Familiar*), as autoavaliações de saúde e alimentação (*Padrao_Saude* e *Padrao_Alimentacao*), além da faixa de renda domiciliar (*Renda_Total_Classes*). A incorporação dessas variáveis permitiu examinar a estrutura conjunta das relações entre características dos domicílios e os padrões de consumo de pescado.

A aplicação da AC possibilitou identificar proximidades e contrastes entre perfis de consumo e categorias das variáveis explicativas, contribuindo para uma compreensão mais abrangente dos fatores associados ao comportamento alimentar no estado de São Paulo.

7.6.1 Gráficos dos Fatores Latentes e Consumo de Pescado

As Figuras apresentam a relação entre as categorias dos fatores latentes extraídos pela Análise Fatorial e as categorias do consumo de pescado, por meio da Análise de Correspondência. Cada ponto no gráfico representa uma categoria de uma das variáveis avaliadas. A proximidade entre os pontos indica maior associação entre as categorias, enquanto distâncias maiores sugerem fraca associação ou ausência de relação.

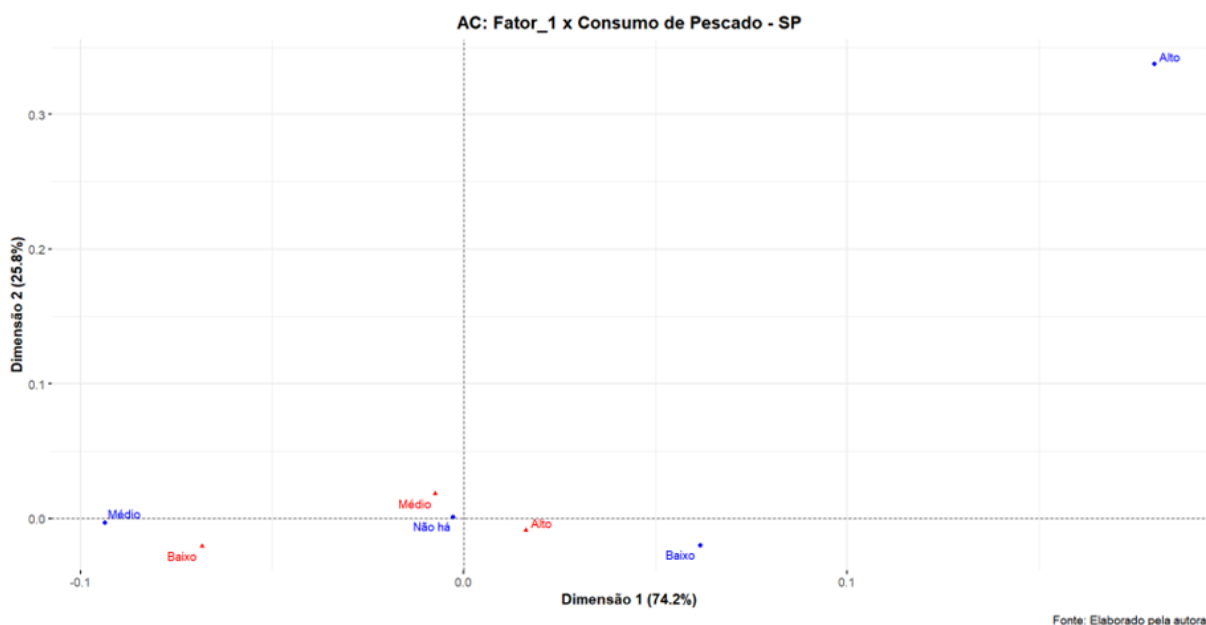
As categorias são diferenciadas visualmente por cores e formatos, de modo a facilitar a interpretação:

- **Círculos azuis:** categorias do consumo de pescado;
- **Triângulos vermelhos:** níveis dos fatores latentes.

Os gráficos são construídos a partir das duas dimensões principais geradas pela Análise de Correspondência, as quais explicam conjuntamente 100% da variabilidade presente nas

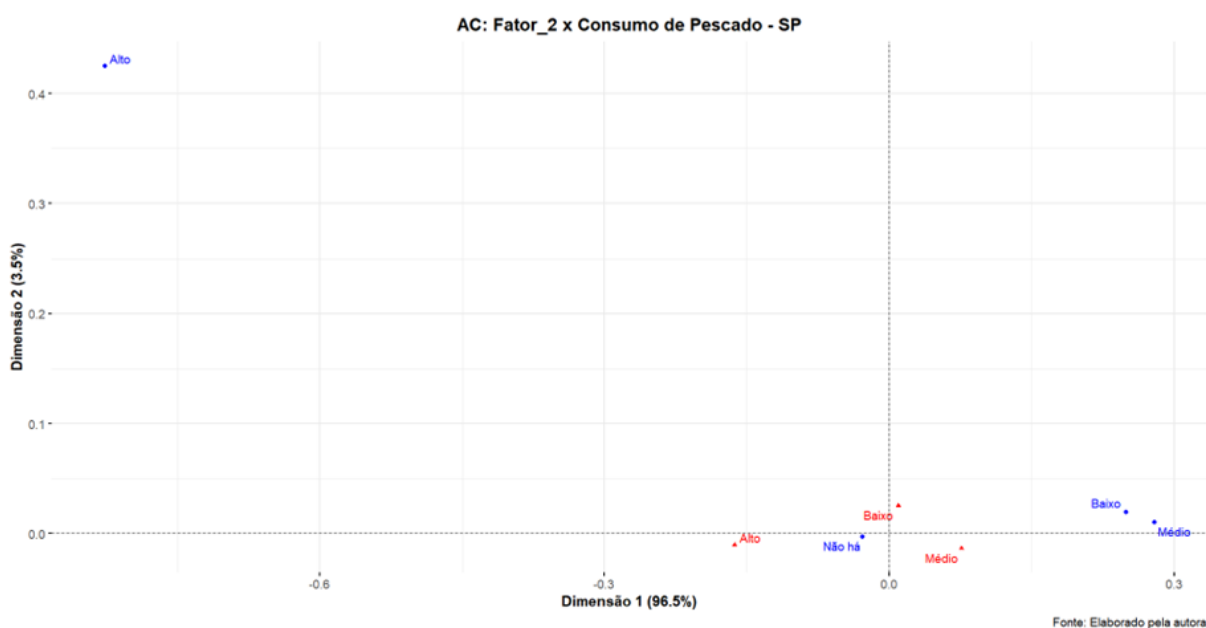
associações. Dessa forma, as projeções bidimensionais fornecem uma representação fiel e completa dos padrões relacionais entre consumo de pescado e os fatores socioeconômicos e estruturais identificados anteriormente.

Figura 9 – Análise de Correspondência — Fator 1 (Infraestrutura Urbana e Serviços Públicos) X Consumo de Pescado



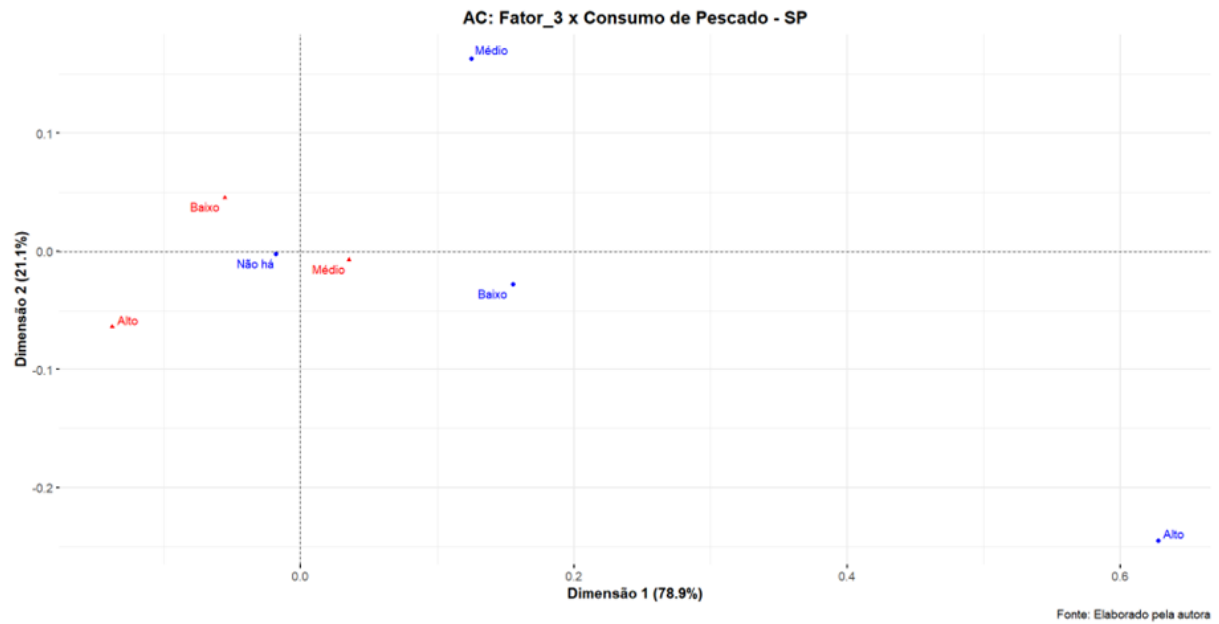
Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

Figura 10 – Análise de Correspondência — Fator 2 (Consumo e Bem-Estar Doméstico) X Consumo de Pescado



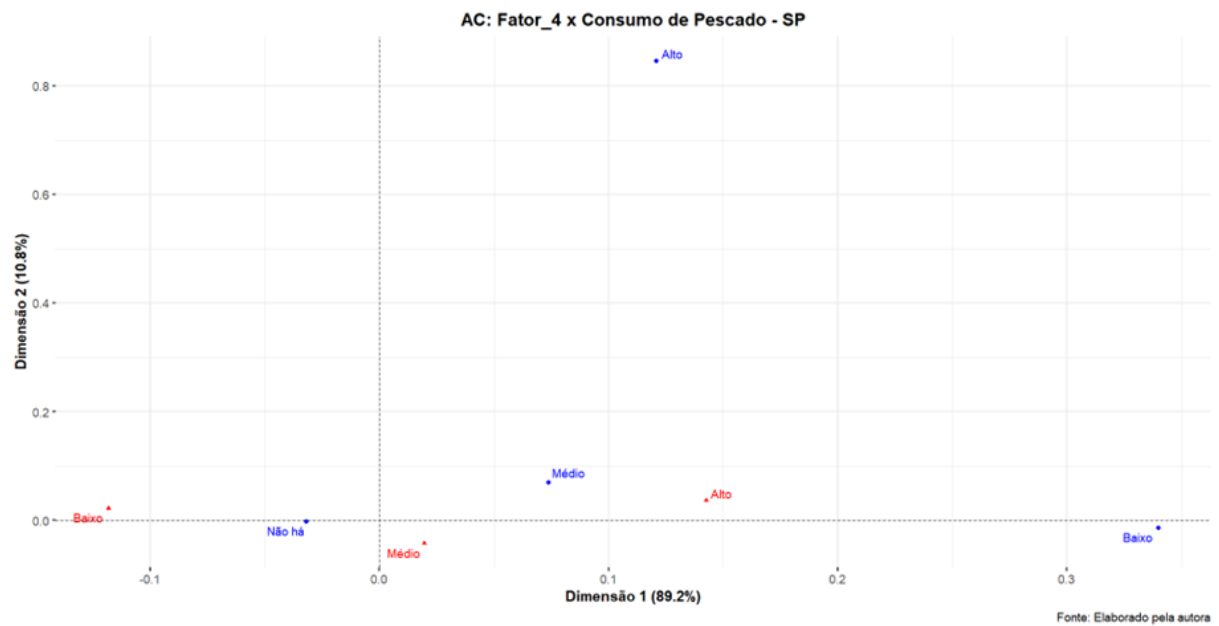
Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

Figura 11 – Análise de Correspondência — Fator 3 (Condição Socioeconômica) X Consumo de Pescado



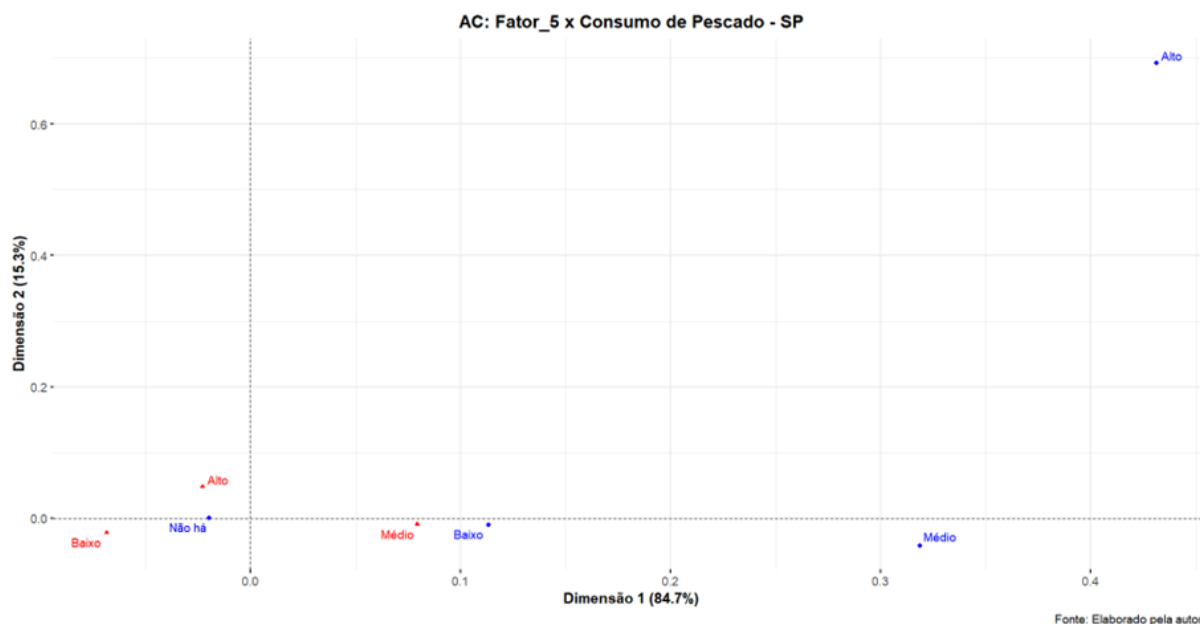
Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

Figura 12 – Análise de Correspondência — Fator 4 (Saneamento e Serviços Essenciais) X Consumo de Pescado



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

Figura 13 – Análise de Correspondência — Fator 5 (Idade) X Consumo de Pescado



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

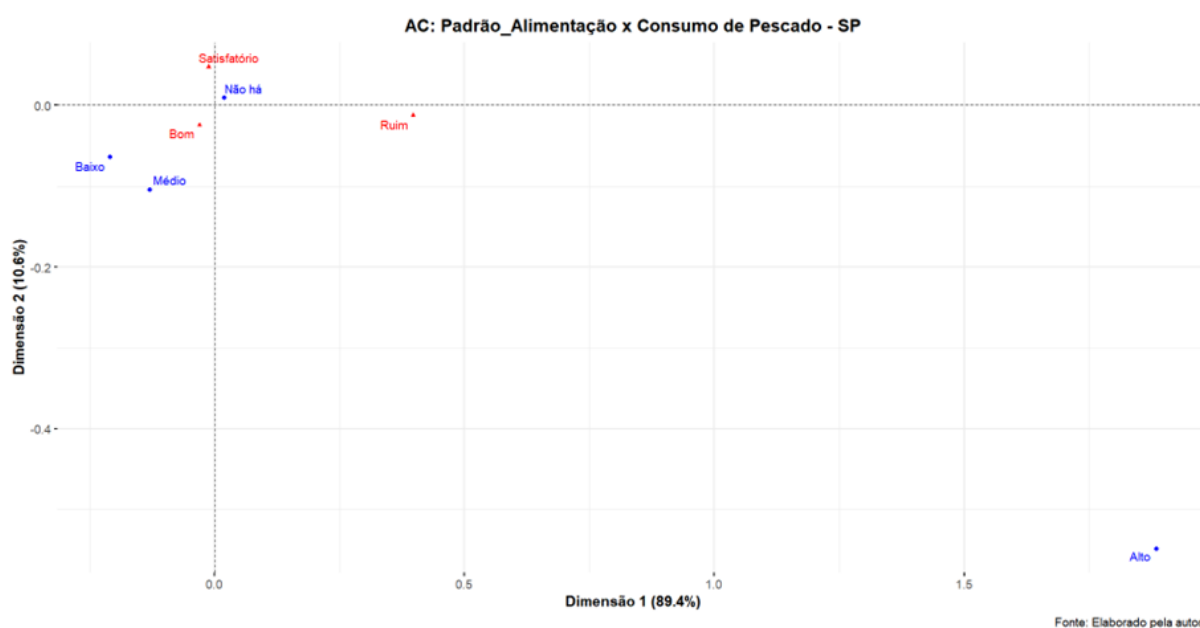
A partir da Figura 9 a Figura 13 apresentam os mapas da Análise de Correspondência conduzida para investigar a relação entre cada um dos cinco fatores latentes extraídos na etapa fatorial e as categorias de consumo de pescado no estado de São Paulo. Em todas as figuras, as categorias dos fatores são representadas pelos pontos vermelhos, enquanto as categorias de consumo de pescado aparecem como triângulos azuis. Assim, a proximidade entre esses elementos indicaria associação entre o nível do fator e a intensidade de consumo. No entanto, uma avaliação conjunta dos cinco mapas revela que tais fatores, em sua maioria, não foram úteis para discriminar padrões claros de consumo.

De forma geral, o panorama integrado dos cinco mapas indica que nenhum dos fatores latentes derivados da Análise Fatorial foi útil para discriminar os níveis de consumo de pescado na população paulista. A proximidade dos pontos vermelhos e azuis à região central do plano, a dispersão desalinhada entre as categorias e a ausência de gradientes interpretáveis entre fator e consumo evidenciam que os perfis multidimensionais representados pelos fatores — embora relevantes para caracterizar condições de vida, infraestrutura, consumo e demografia — não se traduzem em diferenças significativas nos padrões de ingestão de pescado. Esses resultados reforçam o que já havia sido observado nas análises descritivas e nos testes não paramétricos: o consumo de pescado apresenta baixa variabilidade e forte concentração em valores próximos de zero, o que pode limitar sua capacidade de ser explicado por fatores socioeconômicos, estruturais ou demográficos em uma análise multivariada desse tipo.

7.6.2 Gráficos das Variáveis Socioeconômicas e Consumo de Pescado

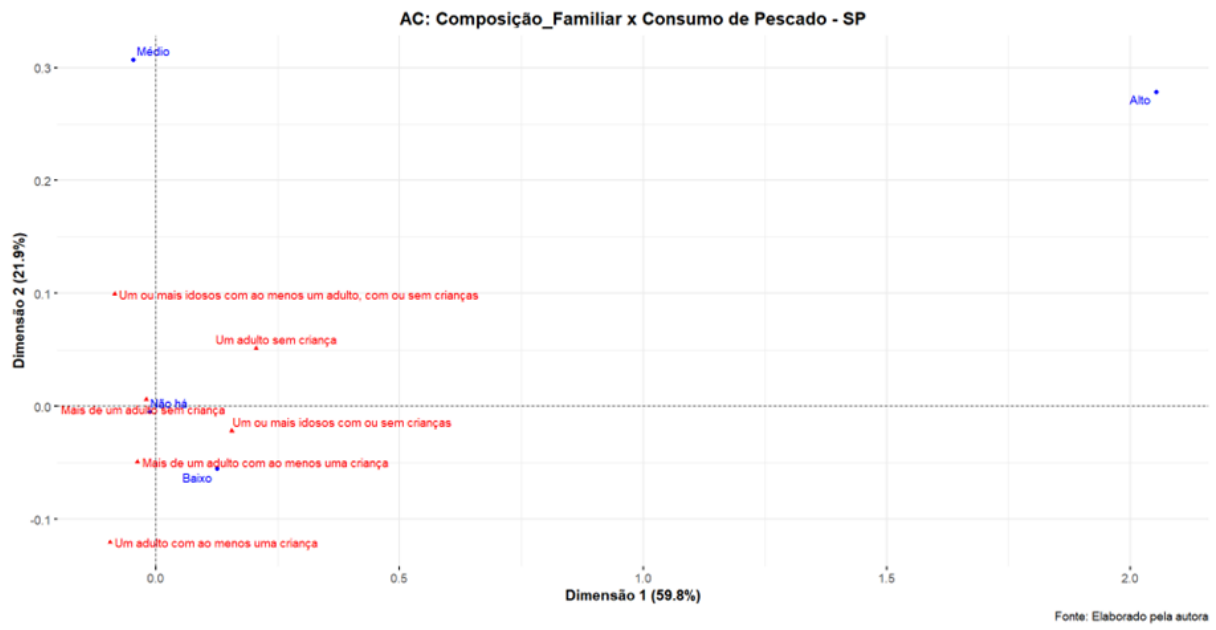
Nesta subseção são apresentados os gráficos que relacionam diferentes variáveis socioeconômicas ao consumo de pescado no estado de São Paulo. O objetivo é ilustrar visualmente como características demográficas, econômicas e familiares se distribuem em relação às categorias de consumo, oferecendo suporte à análise exploratória e contribuindo para a compreensão geral dos perfis observados na amostra.

Figura 14 – Análise de Correspondência — Padrão da Alimentação X Consumo de Pescado



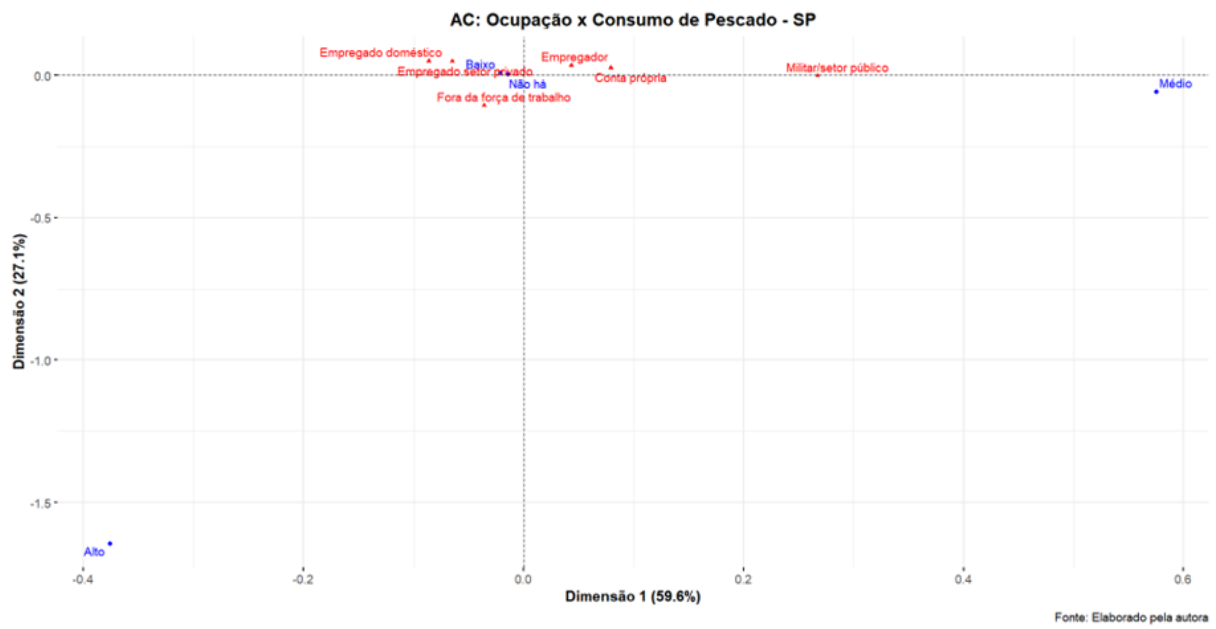
Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

Figura 15 – Análise de Correspondência — Composição Familiar X Consumo de Pescado



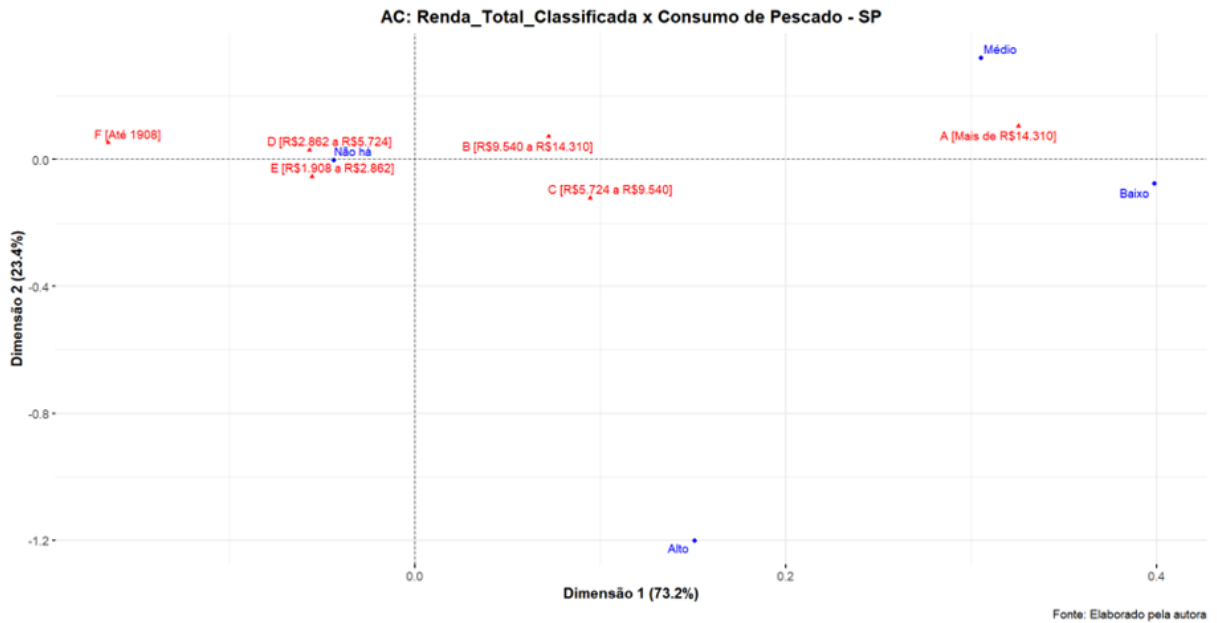
Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

Figura 16 – Análise de Correspondência — Ocupação X Consumo de Pescado



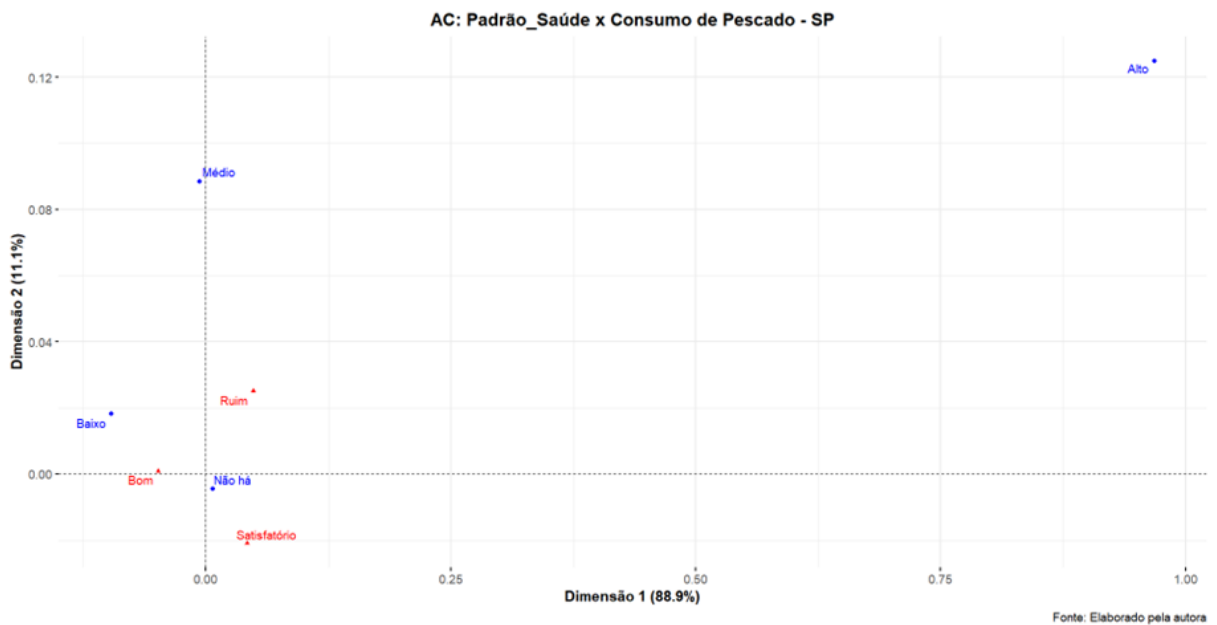
Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

Figura 17 – Análise de Correspondência — Renda Total X Consumo de Pescado



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

Figura 18 – Análise de Correspondência — Padrão de Saúde X Consumo de Pescado



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

As Figuras referentes às variáveis *Renda*, *Padrão de Alimentação*, *Padrão de Saúde*, *Composição Familiar* e *Ocupação* apresentam resultados semelhantes aos observados na análise dos fatores latentes. De modo geral, as categorias de cada variável (em vermelho) concentram-se

próximas à origem do plano fatorial, assim como as categorias de consumo de pescado (em azul), indicando que essas variáveis não estabelecem padrões claros de associação com o consumo.

A única exceção parcial ocorre no caso da *Renda*, em que as faixas superiores tendem a se afastar discretamente das demais categorias e aproximam-se levemente das categorias que consomem algum tipo de pescado. No entanto, essa tendência é sutil e não configura um gradiente interpretável ou uma relação forte entre renda e consumo de pescado. Para as demais variáveis — alimentação, saúde, composição familiar e ocupação — o arranjo espacial dos pontos permanece centralizado, sem direção dominante ou agrupamentos específicos.

Assim como na análise dos fatores, esses resultados sugerem que o consumo de pescado não apresenta variabilidade suficiente para ser discriminado por essas características. Dessa forma, as análises de correspondência com variáveis originais reforçam que tais dimensões, isoladamente, não explicam de maneira robusta os padrões de consumo observados na população paulista.

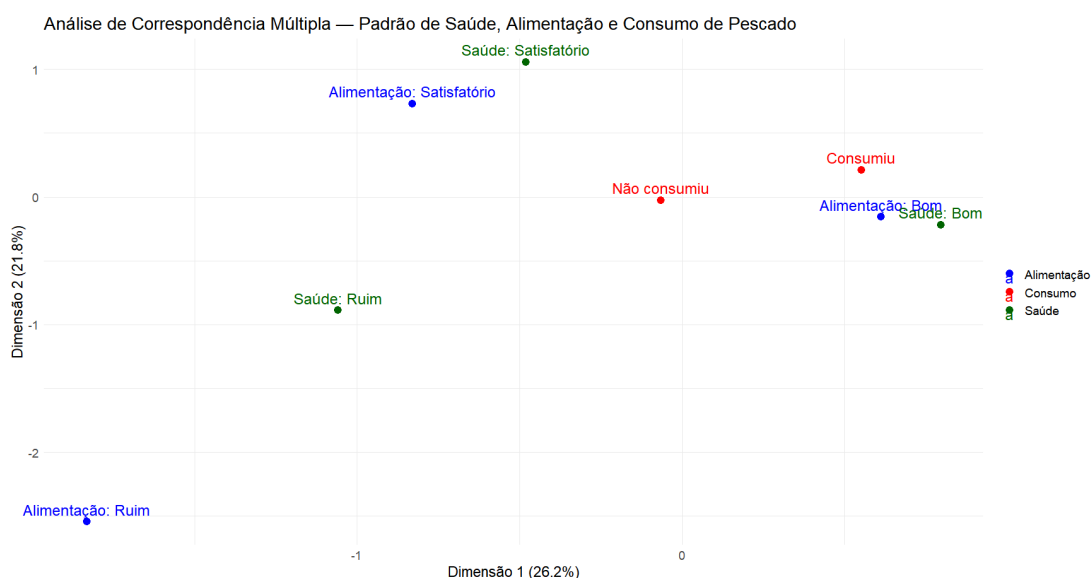
7.6.3 Análise de Correspondência Múltipla

A Análise de Correspondência Múltipla (ACM) foi empregada para investigar simultaneamente a relação entre variáveis categóricas associadas às condições socioeconômicas dos domicílios e ao consumo de pescado. Essa técnica permite representar graficamente categorias de diferentes variáveis em um espaço de baixa dimensionalidade, de forma que a proximidade entre os pontos reflita o grau de associação entre eles. Assim, categorias que aparecem próximas no mapa perceptual tendem a ocorrer em conjunto entre os indivíduos, enquanto categorias mais distantes apresentam menor relação.

No presente estudo, a ACM foi aplicada considerando as categorias de consumo de pescado (“Consumiu” e “Não consumiu”), juntamente com os níveis de padrão de alimentação e padrão de saúde dos domicílios. As duas primeiras dimensões do método foram utilizadas para interpretação, por concentrarem a maior proporção da variabilidade explicada. Ademais, cada conjunto de categorias foi representado por uma cor distinta no gráfico, facilitando a identificação visual das associações entre variáveis.

O objetivo dessa etapa não é realizar inferência estatística, mas sim explorar padrões de agrupamento, identificar relações estruturais entre as categorias e fornecer subsídios para interpretações complementares às análises descritivas e multivariadas apresentadas ao longo do estudo.

Figura 19 – Análise de Correspondência Múltipla — Padrão de Saúde, Alimentação e Consumo de Pescado



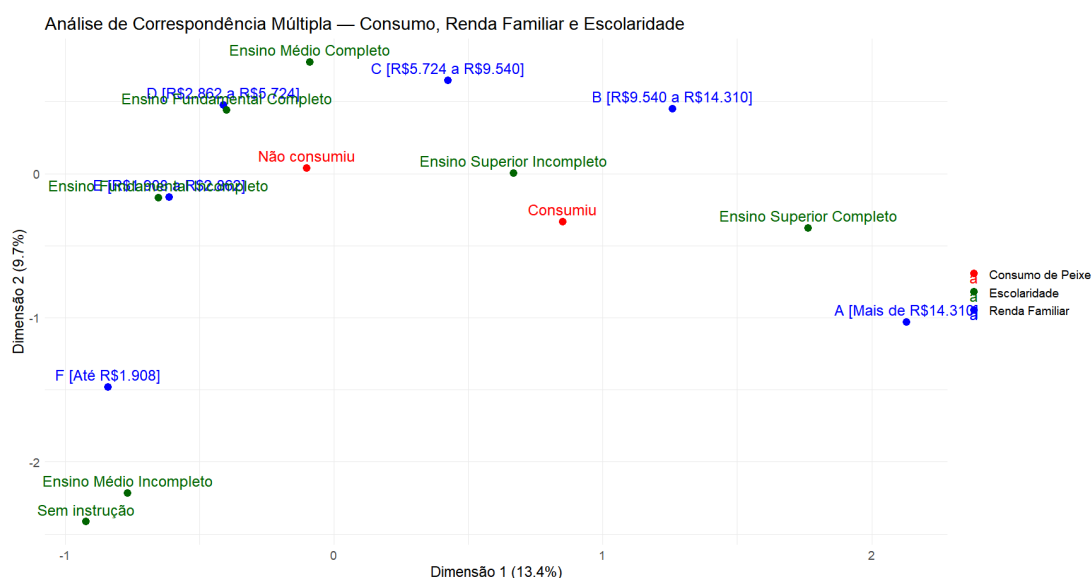
Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 19 apresenta o mapa perceptual resultante da Análise de Correspondência Múltipla (ACM), considerando as categorias de consumo de pescado, padrão de alimentação e padrão de saúde. Observa-se que as categorias pertencentes a uma mesma variável tendem a se agrupar entre si, indicando coerência na distribuição interna dos níveis avaliados. Esse comportamento é esperado, uma vez que categorias originadas da mesma variável compartilham estrutura de variação semelhante e, portanto, ocupam regiões próximas no espaço fatorial.

No que se refere ao consumo de pescado, nota-se que as categorias “*Consumiu*” e “*Não consumiu*” apresentam posicionamentos distintos no plano fatorial. A categoria “*Consumiu*” aparece mais próxima dos níveis mais elevados de padrão de alimentação e de padrão de saúde, sugerindo que melhores condições nesses indicadores estão associadas a maior probabilidade de consumo. Em contrapartida, a categoria “*Não consumiu*” se localiza relativamente próxima aos níveis inferiores ou intermediários dessas variáveis, indicando possível relação entre menor qualidade alimentar e de saúde e menor nível de consumo de pescado.

Embora a ACM seja uma técnica de caráter eminentemente exploratório — e, portanto, não permita inferências causais — a estrutura observada no gráfico sugere um padrão consistente: níveis mais elevados de saúde e alimentação se alinham com a categoria “*Consumiu*”, enquanto níveis mais baixos tendem a se aproximar de “*Não consumiu*”. Esse resultado reforça tendências previamente identificadas nas análises descritivas, indicando que padrões alimentares e condições de saúde podem exercer influência relevante sobre o consumo de pescado nos domicílios analisados, ainda que outros fatores socioeconômicos e culturais também possam desempenhar papel importante.

Figura 20 – Análise de Correspondência Múltipla — Renda Familiar, Escolaridade e Consumo de Pescado



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 20 apresenta o mapa fatorial referente à Análise de Correspondência Múltipla realizada com as variáveis de consumo de pescado, renda familiar e nível de escolaridade. Observa-se que os pontos que representam *baixas rendas familiares* (classes F e E) encontram-se espacialmente próximos dos *menores níveis de escolaridade*, como “Sem instrução” e “Ensino Médio Incompleto”. Esse padrão indica uma associação estrutural entre condições econômicas mais restritas e menores níveis educacionais, relação amplamente documentada em estudos socioeconômicos e coerente com o esperado para a população analisada.

Por outro lado, categorias associadas a *maiores rendas* (classes B e A) aparecem posicionadas mais à direita do mapa perceptual, aproximando-se dos níveis mais altos de escolaridade, tais como “Ensino Superior Incompleto” e “Ensino Superior Completo”. A proximidade espacial entre essas categorias sugere a presença de um gradiente socioeconômico bem definido, no qual maior escolaridade tende a estar alinhada com melhores condições de renda.

No que se refere ao consumo de pescado, a categoria “*Consumiu*” encontra-se deslocada em direção às regiões ocupadas por rendas mais elevadas e maiores níveis de instrução, enquanto a categoria “*Não consumiu*” se posiciona mais próxima das classes de renda mais baixas e de níveis educacionais inferiores. Embora a ACM possua natureza exploratória e não permita afirmar relações causais, o padrão observado indica que **maiores rendas e maior escolaridade podem estar associadas a uma maior probabilidade de consumo de pescado**. De forma complementar, níveis socioeconômicos mais baixos tendem a se agrupar em torno da categoria “*Não consumiu*”.

Em conjunto, os resultados sugerem a existência de um eixo socioeconômico que estrutura simultaneamente renda, escolaridade e comportamento alimentar, reforçando achados prévios das análises descritivas e oferecendo um panorama visual das associações entre essas dimensões.

7.7 ANÁLISE DE CLUSTER

A análise de cluster teve como objetivo identificar perfis homogêneos de domicílios a partir das variáveis selecionadas, permitindo agrupar unidades com características semelhantes em relação às dimensões estudadas. Para isso, foram aplicados dois procedimentos distintos, considerando a natureza das variáveis envolvidas.

Primeiramente, para os escores fatoriais derivados da Análise Fatorial — variáveis contínuas e padronizadas — utilizou-se o método hierárquico de Ward, em conjunto com a distância euclidiana. Essa combinação é amplamente recomendada na literatura por minimizar a variância intragrupo a cada etapa de fusão, produzindo agrupamentos mais compactos e facilmente interpretáveis.

Em complemento, para o conjunto de variáveis qualitativas ordinais, aplicou-se o método de Ward utilizando a distância de Gower. Essa medida é apropriada para dados mistos e, no caso específico deste estudo, permite incorporar variáveis ordinais respeitando a ordem natural entre suas categorias. Assim, a distância resultante reflete adequadamente as diferenças entre os níveis das variáveis, mantendo coerência estatística na formação dos clusters.

Em ambos os casos, a definição final do número de grupos foi fundamentada na inspeção do dendrograma e no índice de silhueta, assegurando a escolha de uma estrutura de clusters que maximizasse a separação entre grupos e a coesão interna. Dessa forma, a análise de cluster permitiu uma segmentação robusta dos domicílios paulistas, servindo como base para as interpretações subsequentes sobre consumo de pescado e condições socioeconômicas.

7.7.1 Análise de Cluster Através dos Fatores

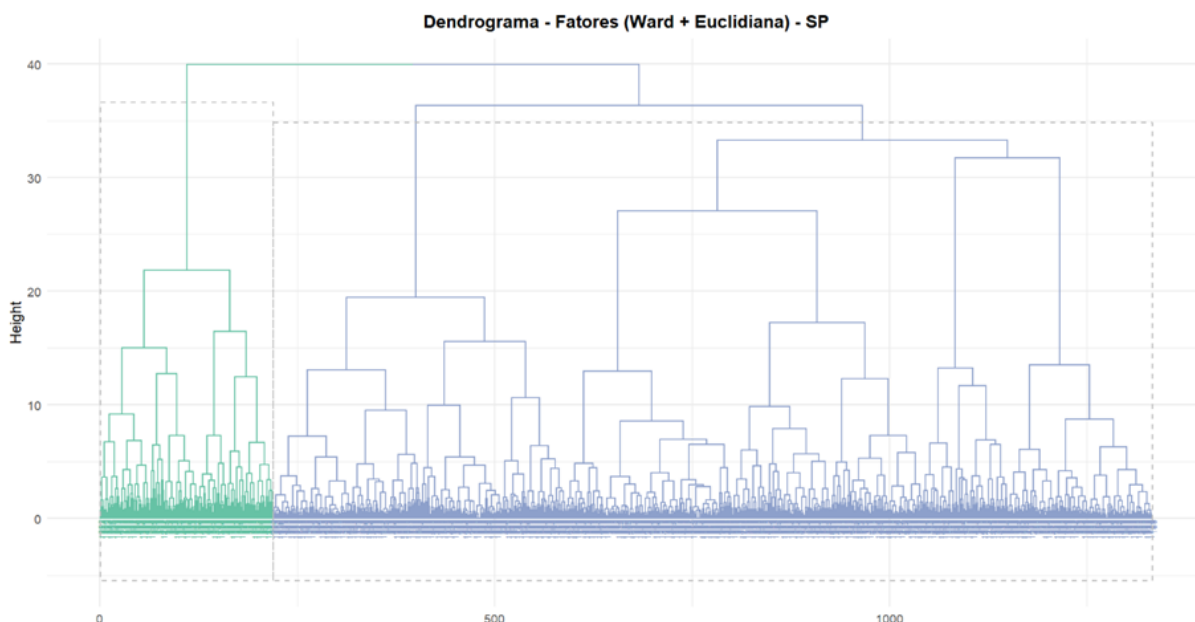
A análise de cluster aplicada aos fatores latentes teve como objetivo identificar grupos de domicílios com perfis socioeconômicos e estruturais semelhantes, a partir dos escores fatoriais obtidos na Análise Fatorial. Como esses escores já representam combinações lineares padronizadas das variáveis originais, a abordagem adotada buscou capturar padrões globais de semelhança entre os domicílios de forma eficiente e interpretável.

Para essa etapa, utilizou-se o método hierárquico de Ward, que minimiza a variabilidade interna dos grupos a cada etapa de fusão, favorecendo a formação de clusters mais homogêneos. A distância empregada foi a Euclidiana, adequada para dados contínuos e para escores fatoriais padronizados, assegurando que cada dimensão contribuísse proporcionalmente para a medida de dissimilaridade.

Os agrupamentos resultantes refletem diferentes combinações dos fatores latentes — como infraestrutura urbana, condições socioeconômicas, padrão de vida e idade — permitindo dis-

tinguir perfis de domicílios com características estruturais semelhantes. Esses clusters são posteriormente utilizados para investigar sua relação com o consumo de pescado, ampliando a compreensão das dinâmicas socioeconômicas associadas ao comportamento alimentar observado.

Figura 21 – Dendrograma a Partir dos Fatores Latentes

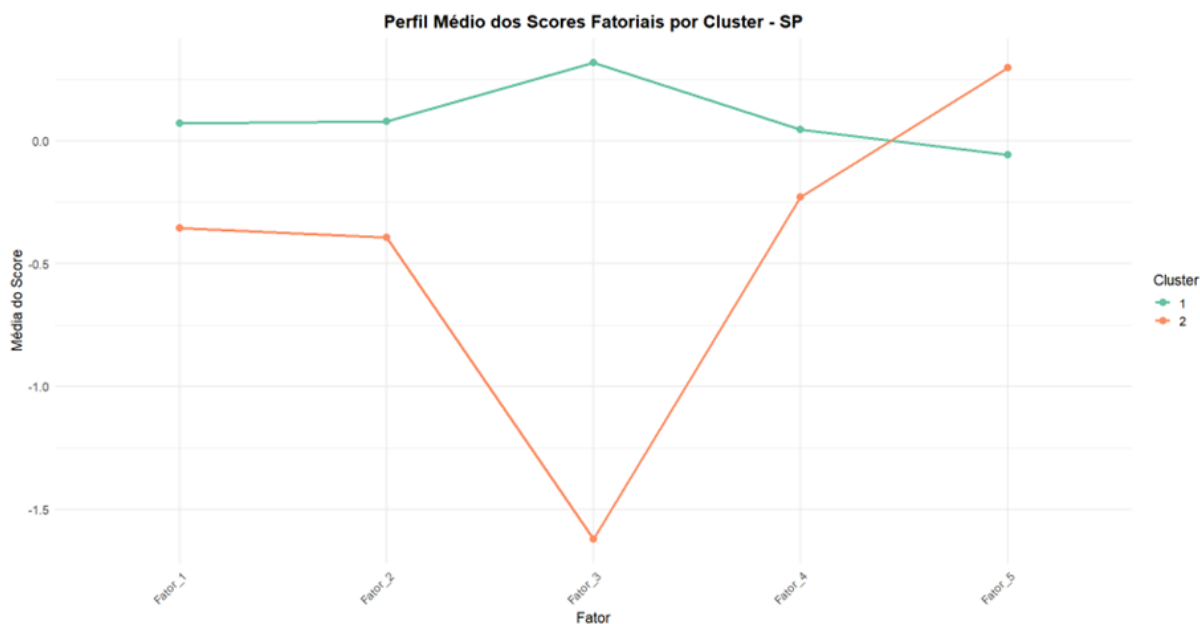


Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 21 apresenta o dendrograma obtido a partir dos *scores* fatoriais extraídos da Análise Fatorial (PCA com rotação varimax), utilizando o método de Ward e a distância Euclidiana. Esse procedimento visa formar grupos internamente homogêneos, minimizando a variabilidade intracluster a cada etapa de fusão. Visualmente, observa-se a existência de dois blocos mais claramente separados, indicação consistente com os resultados da análise de silhueta.

O coeficiente de silhueta foi calculado para valores de k entre 2 e 6, apresentando seu valor máximo em $k = 2$ (silhueta = 0,222), o que reforça a escolha desse número como a estrutura de cluster mais adequada para os dados fatoriais do estado de São Paulo. Assim, os indivíduos foram agrupados em dois clusters, distribuídos da seguinte forma: 1114 unidades no Cluster 1 e 219 unidades no Cluster 2.

Figura 22 – Perfil Médio dos Scores Fatoriais por Cluster



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

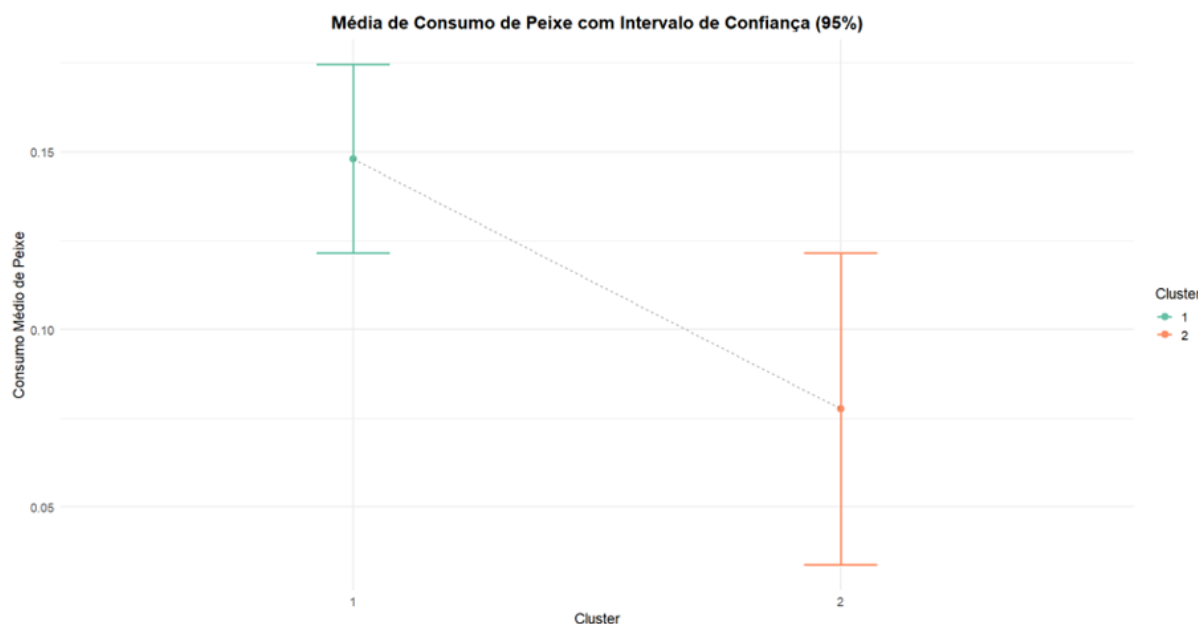
A Figura 22 apresenta o perfil médio dos escores fatoriais para os dois clusters identificados na análise hierárquica realizada a partir dos fatores latentes. Os valores representam as médias padronizadas dos cinco fatores extraídos via Análise Fatorial (PCA com rotação varimax), permitindo comparar o posicionamento relativo dos grupos em cada dimensão latente.

Observa-se que o **Cluster 1**, majoritário (aproximadamente 83% da amostra), apresenta escores médios próximos de zero em todos os fatores, indicando um perfil mais equilibrado e sem desvios acentuados em relação à média estadual. Esse grupo pode ser interpretado como um conjunto de domicílios com condições típicas ou intermediárias nos diferentes eixos latentes analisados.

Por outro lado, o **Cluster 2**, minoritário (cerca de 17% da amostra), apresenta características marcadamente distintas em alguns fatores. O grupo possui escore significativamente negativo no *Fator 3* (Condição Socioeconômica), indicado por valores médios muito abaixo de zero, refletindo perfis de renda e escolaridade inferiores. Em contraste, mostra valores médios superiores no *Fator 5* (Perfil Demográfico), sugerindo concentração de indivíduos em faixas etárias mais elevadas. Além disso, o desempenho moderadamente inferior nos *Fatores 1, 2 e 4* indica condições menos favoráveis em infraestrutura urbana, bem-estar doméstico e serviços essenciais.

No conjunto, as diferenças entre os perfis médios reforçam a coerência da clusterização: o **Cluster 1** representa domicílios com condições próximas à média estadual, enquanto o **Cluster 2** reúne unidades consumidoras com desvantagens estruturais e socioeconômicas mais evidentes. Essas distinções subsidiam interpretações posteriores sobre o consumo de pescado e sua relação com as condições latentes dos domicílios analisados.

Figura 23 – Média + Intervalo de Confiança do Consumo Domiciliar de Pescados no Estado de SP dos Clusters dos Fatores



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 23 apresenta a comparação das médias de consumo de pescado entre os dois clusters identificados na análise de agrupamento. As barras de erro representam os intervalos de confiança a 95% para a média de cada grupo, permitindo avaliar a precisão das estimativas e possíveis diferenças entre os clusters.

Observa-se que o **Cluster 1** apresenta média de consumo de pescado superior em relação ao **Cluster 2**. O intervalo de confiança para o Cluster 1 varia aproximadamente de **0,122** a **0,175**, indicando maior precisão da estimativa, reflexo do seu tamanho amostral mais elevado. Em contraste, o **Cluster 2** — composto por um grupo menor de domicílios e caracterizado por piores condições socioeconômicas e estruturais — apresenta média mais baixa, com intervalo de confiança entre **0,0337** e **0,122**, evidenciando maior variabilidade.

Essas diferenças são consistentes com o resultado do teste de Kruskal–Wallis, que indicou diferença estatisticamente significativa entre os grupos ($p = 0,015$). Dessa forma, os resultados sugerem que domicílios associados ao **Cluster 1**, caracterizados por condições estruturais, socioeconômicas e demográficas mais favoráveis, tendem a apresentar um consumo de pescado ligeiramente maior do que aqueles pertencentes ao **Cluster 2**. Ainda assim, destaca-se que os níveis de consumo permanecem baixos em ambos os grupos.

Estatísticas Descritivas por Cluster

A seguir, apresentam-se as estatísticas descritivas das variáveis quantitativas *Renda_Total*, *Peixe_geral* e *Anos_Estudo*, calculadas separadamente para cada grupo.

Tabela 8 – Estatísticas Descritivas por Cluster para Renda_Total – SP

Cluster	Min	Q1	Mediana	Média	Q3	Máx	Desvio-Padrão	n
1	127,00	2632,00	4479,00	7159,00	7984,00	235192,00	10675,00	1114
2	354,00	2270,00	3575,00	4344,00	5439,00	32645,00	3499,00	219

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Tabela 8 evidencia diferenças importantes no perfil socioeconômico dos dois grupos identificados. O **Cluster 1** apresenta renda substancialmente mais elevada, tanto em termos de mediana (R\$ 4.479) quanto de média (R\$ 7.159), além de grande variabilidade interna, como indica o elevado desvio-padrão. Esse comportamento sugere que o Cluster 1 agrega domicílios com maior heterogeneidade e com maior presença de rendas altas. Já o **Cluster 2** apresenta níveis de renda mais baixos e menos dispersos, com média de R\$ 4.344. Esses resultados apontam para uma segmentação clara por nível econômico, na qual o Cluster 1 representa famílias com maior poder aquisitivo.

Tabela 9 – Estatísticas Descritivas por Cluster para Peixe_geral – SP

Cluster	Min	Q1	Mediana	Média	Q3	Máx	Desvio-Padrão	n
1	0	0	0	0,148	0	4	0,451	1114
2	0	0	0	0,0776	0	2	0,330	219

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Tabela 9 mostra que o consumo de pescado é extremamente baixo em ambos os grupos, com mediana igual a zero nos dois casos. Contudo, o **Cluster 1** — grupo que também apresenta maior renda — exibe média ligeiramente mais alta (0,148 kg/semana) e maior dispersão, indicando que, mesmo que o consumo seja baixo em geral, há indivíduos desse cluster que consomem quantidades superiores. Esse padrão sugere que o acesso ao pescado tende a aumentar conforme a renda cresce, ainda que a maioria das famílias consuma valores próximos de zero.

Tabela 10 – Estatísticas Descritivas por Cluster para Anos_Estudo – SP

Cluster	Min	Q1	Mediana	Média	Q3	Máx	Desvio-Padrão	n
1	0	6	12	10,20	12	16	4,43	1114
2	0	5	8	7,96	12	16	4,15	219

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

O **Cluster 1** novamente apresenta valores superiores, desta vez em termos de escolaridade, com mediana de 12 anos de estudo e média de 10,2 anos. Já o **Cluster 2** apresenta mediana igual a 8 anos, indicando maior concentração de indivíduos com escolaridade menor. Assim, a segmentação obtida revela que o Cluster 1 consiste em indivíduos com maior capital educacional, reforçando a associação entre escolaridade e renda observada nas tabelas anteriores.

Teste de Diferenças no Consumo entre Clusters

Para investigar se o consumo de pescado difere entre os dois grupos formados, foi realizado o teste não paramétrico de Kruskal–Wallis, adequado devido à distribuição assimétrica e ao grande número de zeros em *Peixe_geral*.

Tabela 11 – Teste de Kruskal–Wallis para diferenças no consumo de pescado entre clusters

Estatística	Valor
Qui-quadrado	5,8841
Graus de liberdade	1
p-valor	0,01528

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

Para investigar se os dois grupos formados pelos clusters diferem quanto ao consumo de pescado, aplicou-se o teste não paramétrico de Kruskal–Wallis. Esse teste é apropriado porque a variável *Peixe_geral* apresenta distribuição altamente assimétrica, com grande concentração de valores iguais a zero, o que inviabiliza o uso de testes paramétricos tradicionais. As hipóteses avaliadas foram:

H_0 : As distribuições do consumo de pescado são iguais entre os clusters;

H_1 : Pelo menos um dos clusters apresenta distribuição diferente.

O resultado apresentado na Tabela 11 rejeita a hipótese nula ao nível de 5% ($p < 0,05$), indicando que os clusters diferem significativamente quanto ao consumo de pescado. Embora a mediana seja igual a zero em ambos os grupos — reforçando o baixo consumo geral na população — o **Cluster 1** apresenta média mais elevada e maior variabilidade. Isso sugere que, dentro desse grupo, existe uma parcela de indivíduos que consome pescado em quantidades superiores.

Como o Cluster 1 também apresenta maior renda e maior escolaridade, o achado é consistente com a interpretação socioeconômica: famílias com melhor condição econômica e níveis educacionais mais altos tendem a ter maior acesso, disponibilidade e diversidade alimentar, o

que pode incluir maior consumo de pescado. Assim, o teste confirma que a segmentação obtida pelos clusters captura diferenças reais e relevantes no comportamento alimentar da população analisada.

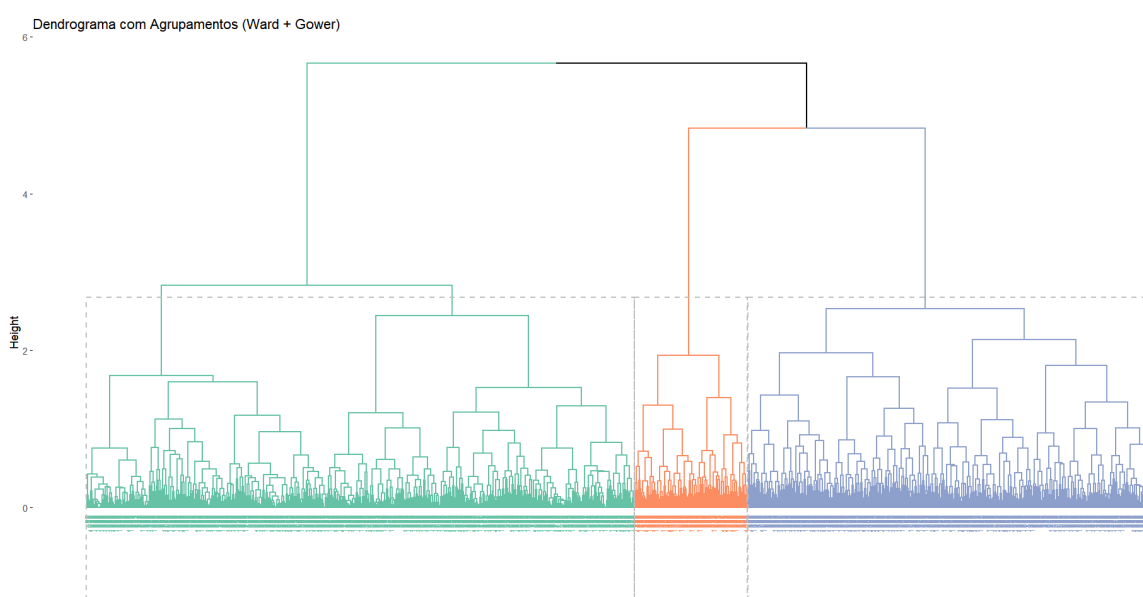
7.7.2 Análise de Cluster Através das Variáveis Ordinais

Nesta etapa, foi realizada uma análise de cluster utilizando exclusivamente as variáveis qualitativas ordinais disponíveis na base de dados, a saber: *Padrao_Alimentacao*, *Padrao_Moradia*, *Padrao_Vestuario*, *Padrao_Educacao*, *Padrao_Saude*, *Padrao_Lazer*, *Fornecimento_Agua*, *Fornecimento_Energia*, *Servico_Iluminacao*, *Coleta_Lixo*, *Manutencao_Rua*, *Escoamento_Agua*, *Esgotamento_Sanitario*, *Transporte_Coletivo*, *Renda_Total_Classes*, *Idade_Cat* e *Nivel_Instrucao*.

Como essas variáveis apresentam natureza ordinal — isto é, categorias ordenadas que não representam distâncias numéricas constantes — a medida de dissimilaridade adotada foi a distância de *Gower*, apropriada para combinar variáveis categóricas e ordinais sem violar sua estrutura. A partir da matriz de dissimilaridade, aplicou-se o método de ligação de *Ward*, que busca formar grupos internamente homogêneos ao minimizar a variância dentro dos clusters.

Essa abordagem permitiu identificar agrupamentos coerentes de domicílios com padrões semelhantes quanto às percepções de bem-estar, condições de moradia, acesso a serviços públicos, características sociodemográficas e níveis de renda. Os resultados complementam a análise baseada nos fatores latentes, permitindo observar como as categorias originais das variáveis se organizam em grupos estruturados, preservando sua natureza ordinal.

Figura 24 – Dendograma a Partir das Variáveis Ordinais



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 24 apresenta o dendrograma resultante da Análise de Cluster realizada a partir das variáveis ordinais selecionadas, utilizando a distância de Gower — apropriada para variáveis mistas e, especialmente, para escalas ordinais — combinada ao método de ligação *Ward.D2*, que busca formar grupos internamente homogêneos e maximizar a separação entre eles.

Foram utilizadas 1333 observações, sem necessidade de remoção por valores ausentes. A distribuição das dissimilaridades de Gower apresentou média de aproximadamente 0,397, indicando heterogeneidade moderada entre os domicílios quanto às dimensões avaliadas (padrões de vida, infraestrutura urbana, serviços públicos, renda, escolaridade e idade).

A definição do número ótimo de clusters baseou-se no coeficiente médio de Silhueta para $k = 2$ a 6, cujos valores variaram entre 0,067 e 0,162. O melhor desempenho foi obtido para $k = 3$ (coeficiente = 0,162), sugerindo uma estrutura de agrupamento fraca, porém existente.

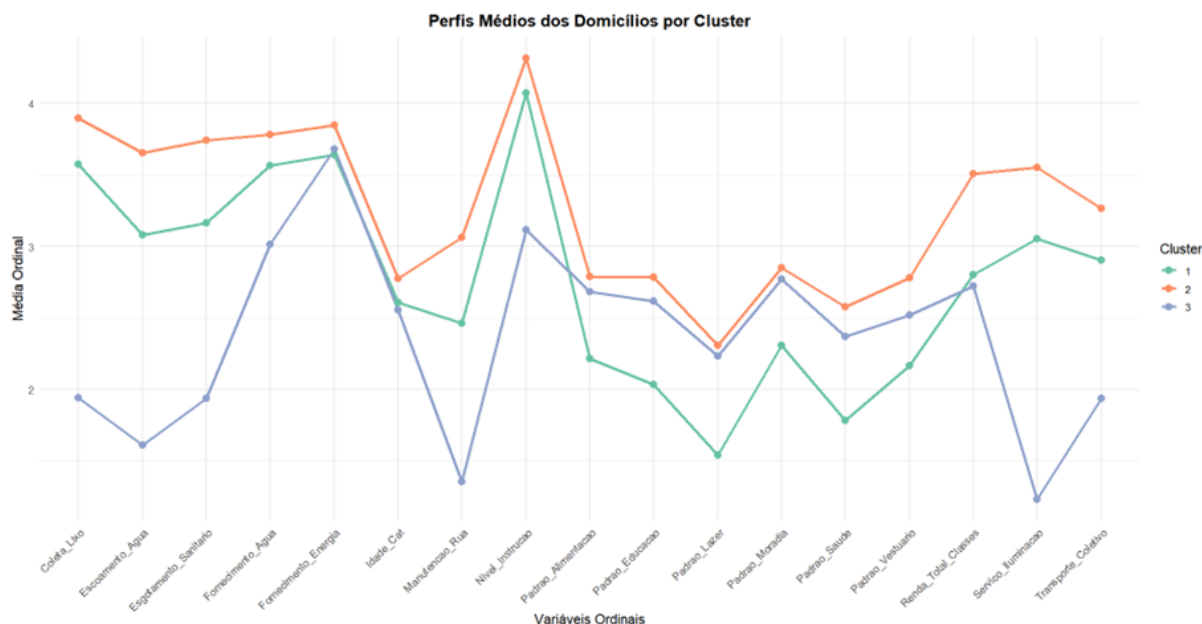
Com base nesse critério, três grupos foram adotados como solução final, com a seguinte distribuição de observações:

Cluster 1 (azul): 508 Cluster 2 (verde): 684 Cluster 3 (laranja): 141

Essa segmentação revela perfis distintos de condições domiciliares, derivados da combinação das variáveis ordinais utilizadas na análise: padrão de alimentação, moradia, vestuário, educação, saúde e lazer, além de indicadores de serviços públicos (fornecimento de água e energia, iluminação pública, coleta de lixo, manutenção urbana, escoamento e esgotamento), bem como faixa de renda, idade e nível de instrução.

Apesar de o coeficiente de Silhueta indicar separação moderada entre os grupos, o uso da distância de Gower juntamente com o método de Ward gera clusters interpretáveis e coerentes com a natureza ordinal dos dados, permitindo análises posteriores sobre a relação entre condições estruturais dos domicílios e o consumo de pescado no estado de São Paulo.

Figura 25 – Perfil Médio das Variáveis Ordinais por Cluster



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 25 apresenta o perfil médio das variáveis ordinais dos domicílios para os três clusters identificados pelo método de Ward com a distância de Gower. Essas variáveis representam dimensões como condições de infraestrutura urbana, serviços essenciais, padrões de vida percebidos, características demográficas e condição socioeconômica.

Observa-se que os clusters apresentam perfis bastante distintos:

Cluster 1 mostra valores médios intermediários na maioria das variáveis, caracterizando domicílios com condições moderadas de infraestrutura, serviços e padrões de vida. Esse grupo tende a refletir um conjunto populacional localizado próximo da média estadual.

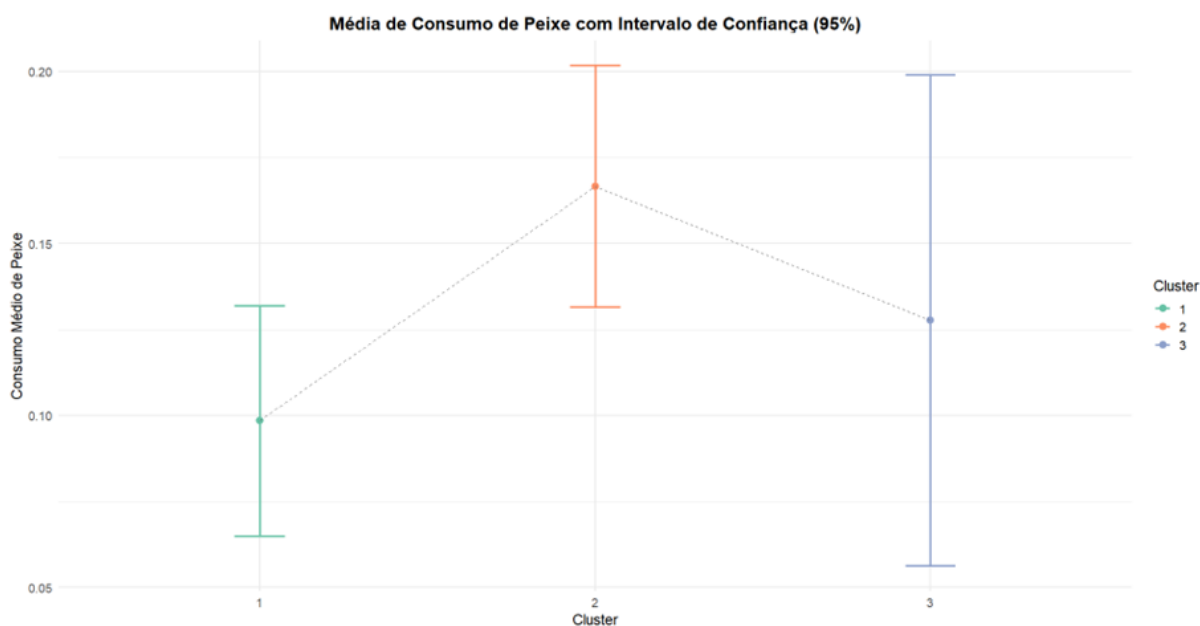
Cluster 2 apresenta, de forma sistemática, os valores médios mais altos em quase todas as variáveis, indicando melhores condições relativas. Esse grupo reúne domicílios que reportam maior qualidade nos serviços públicos, melhores padrões de vida, níveis superiores de escolaridade e renda, além de melhores avaliações sobre alimentação, moradia e saúde. Assim, representa o segmento mais favorecido entre os clusters.

Em contraste, o **Cluster 3** apresenta os valores médios mais baixos em grande parte das variáveis analisadas, sugerindo condições menos favoráveis de infraestrutura (como iluminação e manutenção de vias), menor qualidade percebida em alimentação, moradia, educação e saúde, além de níveis inferiores de renda e escolaridade. Esse cluster configura o grupo mais vulnerável em termos estruturais e socioeconômicos.

A comparação entre os perfis demonstra que a técnica de agrupamento conseguiu identificar três segmentos bem diferenciados da população paulista, distribuídos ao longo de um contínuo de condições de vida — desde domicílios mais privilegiados (Cluster 2), passando por um

grupo intermediário (Cluster 1), até aqueles em situação mais desfavorável (Cluster 3). Esses padrões reforçam a pertinência do uso da distância de Gower para lidar com variáveis ordinais e evidenciam que as diferenças na estrutura dos domicílios estão fortemente associadas às categorias de cluster extraídas.

Figura 26 – Média + Intervalo de Confiança do Consumo Domiciliar de Pescados no Estado de SP dos Clusters das Variáveis Ordinais



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 26 apresenta a comparação das médias de consumo domiciliar de pescados entre os três clusters formados a partir das variáveis ordinais. As barras de erro representam os intervalos de confiança de 95% para a média de cada grupo, permitindo avaliar a precisão das estimativas e a magnitude das diferenças entre os perfis identificados.

Os resultados mostram que o **Cluster 2** apresenta a maior média de consumo de pescado (0,167), com intervalo de confiança entre 0,132 e 0,202, indicando um consumo relativamente superior e estimado com boa precisão. Esse grupo é formado por domicílios que, segundo a análise de cluster, apresentam melhores condições gerais nos indicadores ordinais — como padrões de vida, saneamento e serviços urbanos — o que pode favorecer um consumo alimentar mais diversificado.

O **Cluster 1** apresenta a menor média de consumo (0,0984), com intervalo de 0,0649 a 0,132. Esse resultado sugere que os domicílios deste grupo, caracterizados por condições moderadamente inferiores em várias dimensões ordenadas, têm acesso mais limitado ao consumo de pescado, mantendo níveis inferiores ao observado nos demais clusters.

O **Cluster 3**, por sua vez, apresenta média intermediária (0,128), com intervalo de confiança mais amplo (0,0563 a 0,199), refletindo maior heterogeneidade interna do grupo. Essa varia-

bilidade sugere que o Cluster 3 reúne domicílios com perfis mistos em termos de condições estruturais, demográficas e de serviços, o que explica a maior amplitude do intervalo.

No conjunto, os resultados indicam diferenças substantivas entre os clusters: o **Cluster 2** associa-se aos maiores níveis de consumo, enquanto o **Cluster 1** apresenta os menores valores médios. Tais padrões reforçam a importância das condições estruturais e do padrão de vida — capturados pelas variáveis ordinais — na diferenciação do consumo de pescados entre os domicílios paulistas.

Estatísticas Descritivas por Cluster

A seguir, apresentam-se as estatísticas descritivas das variáveis quantitativas *Renda_Total*, *Peixe_geral* e *Anos_Estudo*, calculadas separadamente para cada grupo identificado pelo agrupamento hierárquico com distância de Gower e método de Ward.D2.

Tabela 12 – Estatísticas Descritivas por Cluster para Renda_Total – SP

Cluster	Min	Q1	Mediana	Média	Q3	Máx	Desvio-Padrão	n
1	127,00	2167,00	3530,00	4851,00	5925,00	42966,00	4591,00	508
2	356,00	3161,00	5217,00	8235,00	9006,00	95310,00	9558,00	684
3	213,00	2268,00	3380,00	5883,00	5149,00	235192,00	19716,00	141

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Tabela 12 revela diferenças expressivas de renda entre os grupos. O **Cluster 2** apresenta os maiores valores centrais, com mediana de R\$ 5.217 e média elevada (R\$ 8.235), sugerindo um grupo socioeconomicamente mais favorecido. O **Cluster 1** caracteriza-se por rendas mais baixas e menos dispersas, enquanto o **Cluster 3** apresenta média relativamente alta, porém com o maior desvio-padrão entre os grupos, indicando forte heterogeneidade interna e presença de valores extremos. Assim, os três clusters representam perfis econômicos distintos, o que pode influenciar outros comportamentos analisados posteriormente.

Tabela 13 – Estatísticas Descritivas por Cluster para Peixe_geral – SP

Cluster	Min	Q1	Mediana	Média	Q3	Máx	Desvio-Padrão	n
1	0	0	0	0,0984	0	4	0,385	508
2	0	0	0	0,167	0	4	0,467	684
3	0	0	0	0,128	0	2	0,428	141

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

Como mostra a Tabela 13, o consumo de pescado é baixo em todos os grupos, com mediana igual a zero em todos os clusters. Ainda assim, observa-se uma diferença no comportamento médio: o **Cluster 2**, que apresentou a maior renda, também apresenta a maior média de consumo (0,167 kg), seguido pelo Cluster 3 e depois pelo Cluster 1. Esse padrão indica que, mesmo diante de um consumo geral reduzido, o poder aquisitivo pode favorecer maior acesso ou maior diversidade alimentar, refletindo-se em um consumo ligeiramente superior de pescado.

Tabela 14 – Estatísticas Descritivas por Cluster para Anos_Estudo – SP

Cluster	Min	Q1	Mediana	Média	Q3	Máx	Desvio-Padrão	n
1	0	6	12	9,81	12	16	4,27	508
2	0	6	12	10,30	14	16	4,50	684
3	0	5	6	7,45	12	16	4,20	141

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Tabela 14 mostra que o **Cluster 2** novamente se destaca, agora com maior escolaridade média (10,3 anos) e maior concentração de indivíduos com ensino médio completo ou superior. O **Cluster 1** apresenta escolaridade intermediária, enquanto o **Cluster 3** reúne os menores níveis educacionais, com mediana igual a 6 anos. Esses resultados reforçam que os clusters capturam perfis socioeducacionais distintos, alinhados às diferenças de renda observadas na primeira tabela.

Teste de Diferenças no Consumo entre Clusters

Para avaliar diferenças no consumo de pescado entre os três grupos formados, utilizou-se o teste não paramétrico de Kruskal–Wallis, adequado devido à assimetria dos dados e à elevada concentração de zeros na variável *Peixe_geral*.

Tabela 15 – Teste de Kruskal–Wallis para diferenças no consumo de pescado entre clusters

Estatística	Valor
Qui-quadrado	10,059
Graus de liberdade	2
p-valor	0,006541

Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

As hipóteses avaliadas pelo teste foram:

H_0 : As distribuições do consumo de pescado são iguais entre os três clusters;

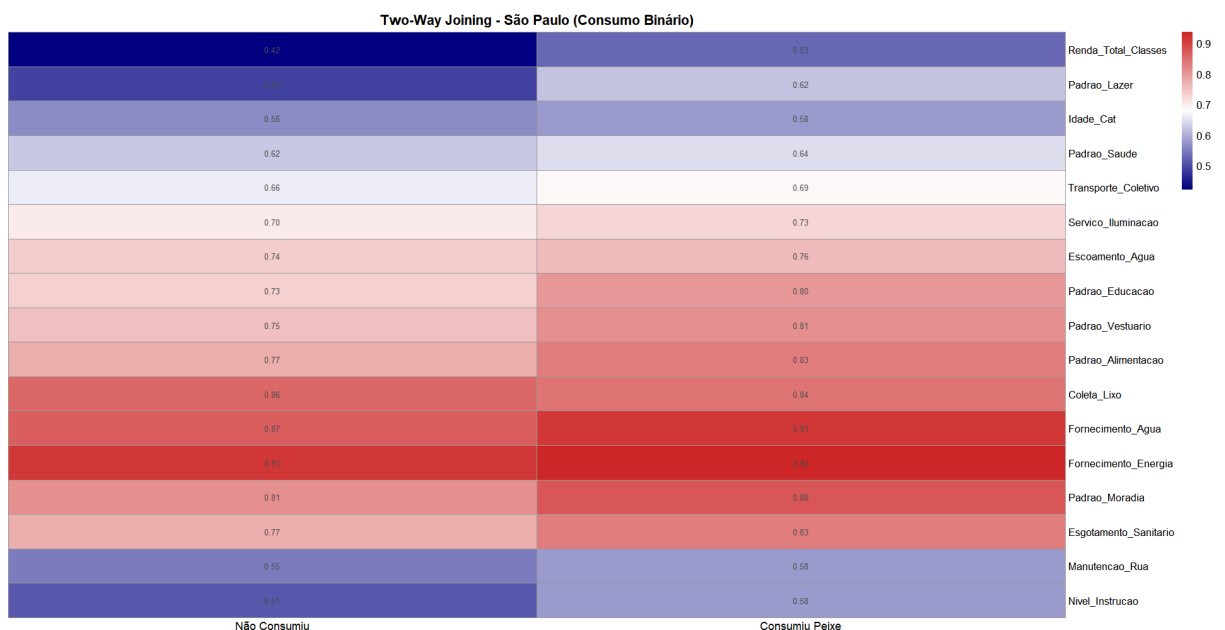
H_1 : Pelo menos um dos clusters apresenta distribuição diferente.

O resultado do teste (Tabela 15) rejeita H_0 ao nível de 1% ($p < 0,01$), indicando diferenças estatisticamente significativas no consumo de pescado entre os grupos. Apesar da mediana zero em todos os clusters, o **Cluster 2** apresenta a maior média e maior variabilidade, seguido pelo Cluster 3 e, por último, pelo Cluster 1. Considerando que o Cluster 2 também apresenta os maiores níveis de renda e escolaridade, o resultado sugere que fatores socioeconômicos influenciam, ainda que modestamente, a propensão ao consumo de pescado no estado de São Paulo.

7.8 TWO-WAY-JOINING

O método de *Two-Way Joining* é utilizado nesta seção com o objetivo de visualizar, de maneira simultânea, a distribuição das variáveis socioeconômicas em relação às categorias binárias de consumo de pescado. Essa abordagem permite identificar padrões gerais de similaridade e contraste entre grupos, organizando linhas e colunas de forma a destacar aproximações estruturais entre perfis de domicílios. A técnica funciona como uma representação bidimensional que auxilia na interpretação exploratória dos dados, complementando as análises multivariadas anteriores e oferecendo uma visualização sintética dos atributos que se associam, direta ou indiretamente, ao comportamento de consumo observado.

Figura 27 – Two-Way-Joining das Variáveis Ordinais X Consumo de Pescado



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 27 apresenta o gráfico de Two-Way Joining, que permite observar simultaneamente como as variáveis ordinais se distribuem entre os domicílios que consumiram ou não consumiram pescado. A técnica reorganiza linhas e colunas de modo a destacar padrões gerais de semelhança, evidenciando conjuntos de características que tendem a ocorrer com maior frequência em cada grupo.

De forma geral, verifica-se que muitas variáveis apresentam valores semelhantes entre os domicílios que consumiram e os que não consumiram pescado, indicando que atributos como padrão de lazer, idade, saúde, transporte coletivo e manutenção urbana não diferenciam de maneira expressiva o comportamento de consumo. Isso sugere que o consumo de pescado não está fortemente estruturado por esses aspectos quando analisados isoladamente.

Apesar disso, o mapa revela algumas tendências importantes. Variáveis relacionadas à infraestrutura do domicílio e à qualidade dos serviços, como fornecimento de água, fornecimento de energia, padrão de moradia e padrão alimentar, apresentam intensidades maiores e colorações mais fortes, especialmente na coluna referente aos domicílios que consumiram pescado. Essa configuração indica uma tendência moderada de associação entre melhores condições de vida e maior probabilidade de consumo. Além disso, as faixas mais altas de renda mostram leve aumento na intensidade para o grupo que consumiu pescado, reforçando padrões já observados nas análises anteriores.

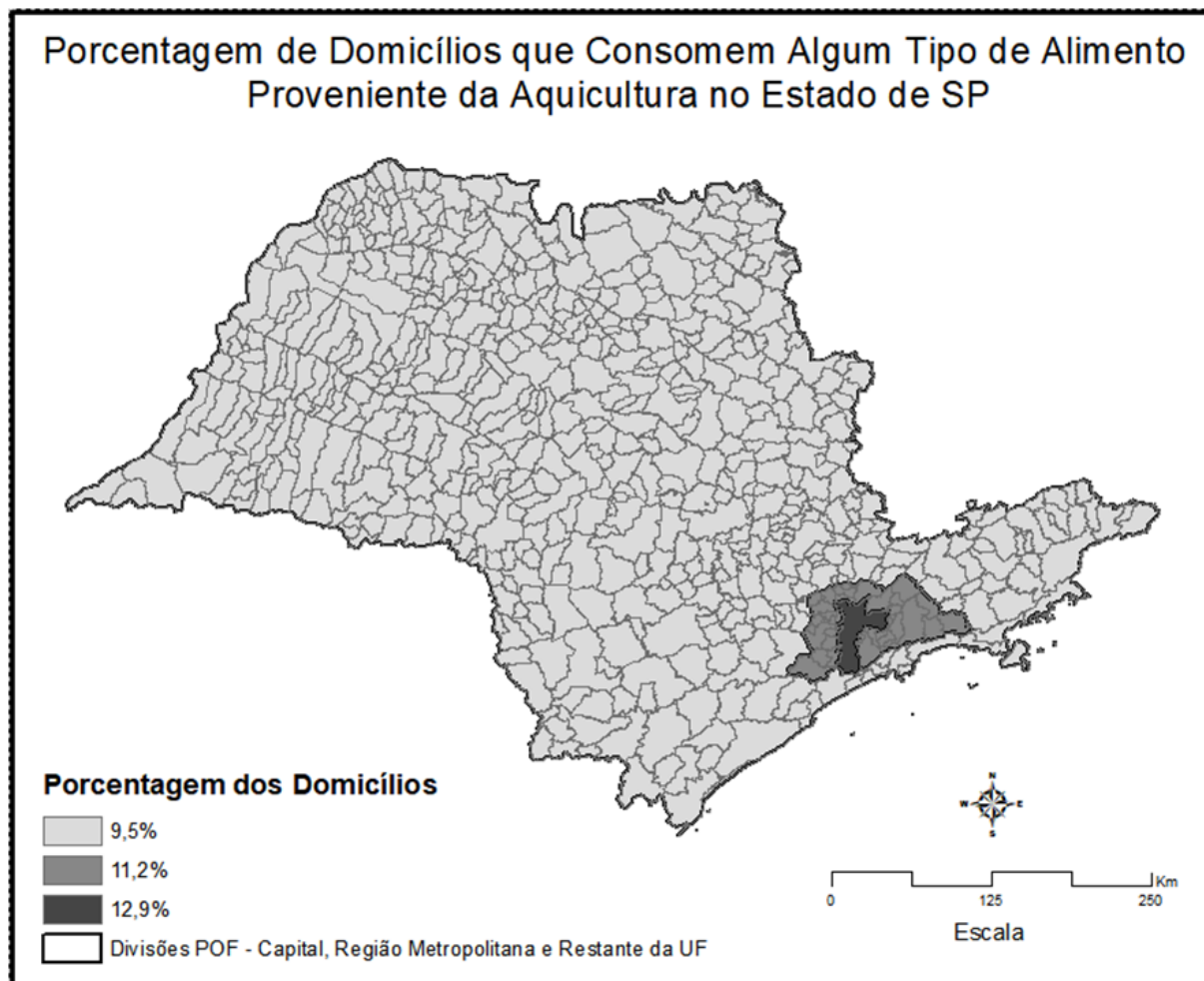
Em síntese, o Two-Way Joining evidencia que, embora muitas variáveis apresentem comportamento semelhante entre os grupos, características ligadas à infraestrutura domiciliar, ao padrão de consumo familiar e à renda sugerem uma tendência de maior consumo entre domicílios com melhores condições de vida. Essa relação é moderada, mas consistente com padrões identificados nas demais análises do estudo.

7.9 ANÁLISE ESPACIAL DA PROPORÇÃO DE CONSUMO DE PESCADO COM VARIÁVEIS EXÓGENAS DA AQUICULTURA

A análise geoespacial foi incorporada ao estudo com o objetivo de avaliar se a produção aquícola municipal apresenta padrões territoriais coerentes com o consumo de pescados no estado de São Paulo. Para isso, foram espacializadas três variáveis provenientes do SIDRA/IBGE: quantidade produzida em kg, valor da produção aquícola e número de estabelecimentos do setor. A seguir, cada variável é analisada por meio de medidas de posição espacial (centro médio e elipse direcional), dependência espacial local (LISA) e identificação de áreas críticas (hot spots G_i^*).

7.9.1 Proporção de Consumo de Pescado entre as Regiões Administrativas dadas na POF

Figura 28 – Porcentagem de Domicílios que Consomem Alimentos da Aquicultura no Estado de SP



Fonte: IBGE, POF (2017–2018). Elaborada pela autora.

A Figura 28 apresenta a distribuição espacial da proporção de domicílios que consumiram ao menos um alimento proveniente da aquicultura no estado de São Paulo, considerando as três divisões utilizadas pela POF: Capital, Região Metropolitana (RM) e Restante do Estado. Observa-se um padrão nitidamente concentrado, no qual a **Capital** apresenta a maior proporção de consumo, seguida pela **Região Metropolitana**. Já o **Restante do Estado**, que engloba a maior parte dos municípios paulistas, exibe as menores proporções.

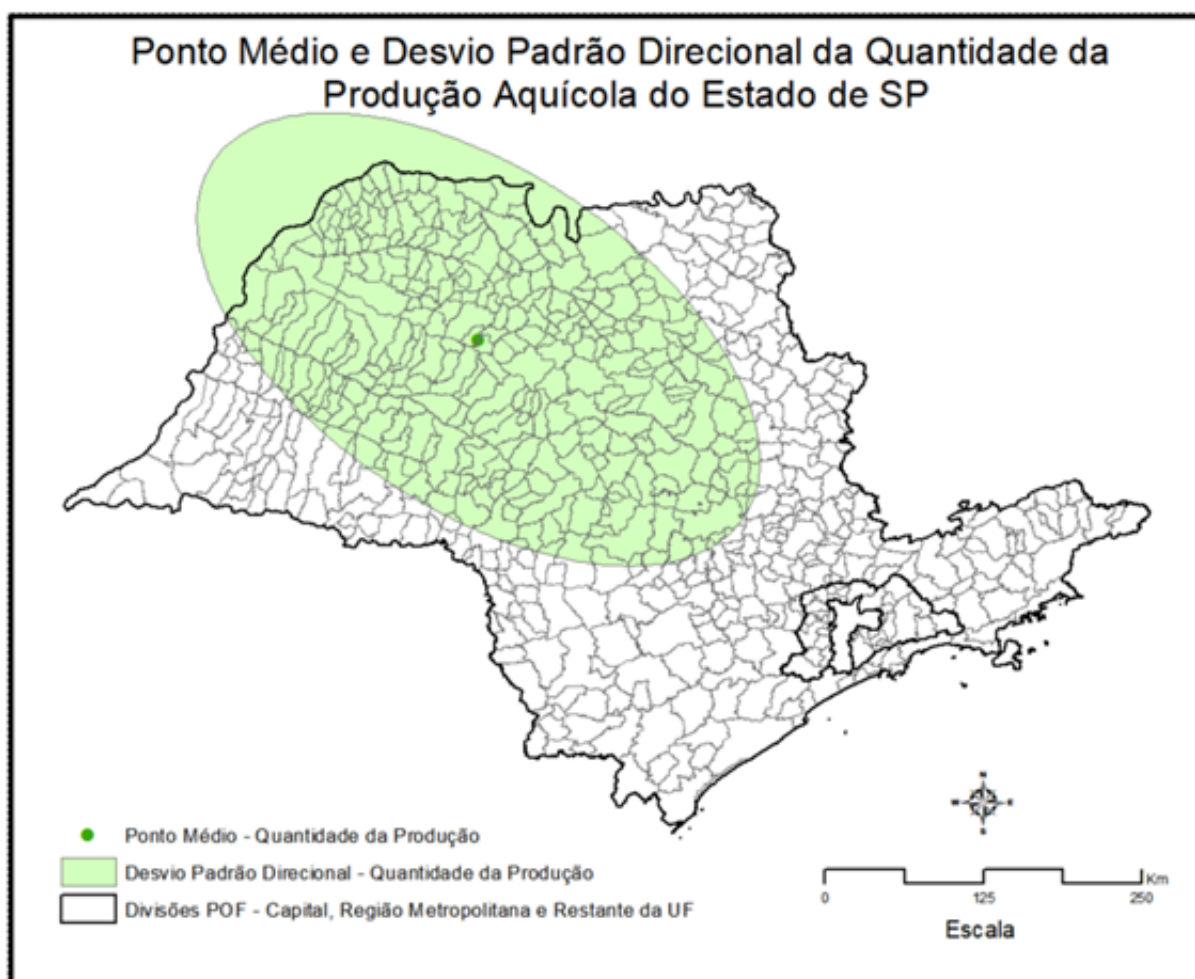
Esse resultado indica que o consumo de alimentos provenientes da aquicultura apresenta um caráter fortemente urbano, sendo mais frequente nas áreas com maior densidade populacional e maior acesso a mercados formais ou canais estruturados de distribuição de pescado. A concentração do consumo nas regiões mais urbanizadas contrasta com a distribuição espacial da

produção aquícola, mais dispersa e concentrada no interior, reforçando a desconexão territorial entre áreas produtoras e consumidoras.

7.9.2 Quantidade Produzida de Produtos Aquícolas (kg)

7.9.2.1 Centro Médio e Distribuição Direcional

Figura 29 – Centro Médio e Distribuição Direcional da Produção Aquícola (kg) – Municípios do Estado de SP



Fonte: IBGE, SIDRA (2017). Elaborada pela autora.

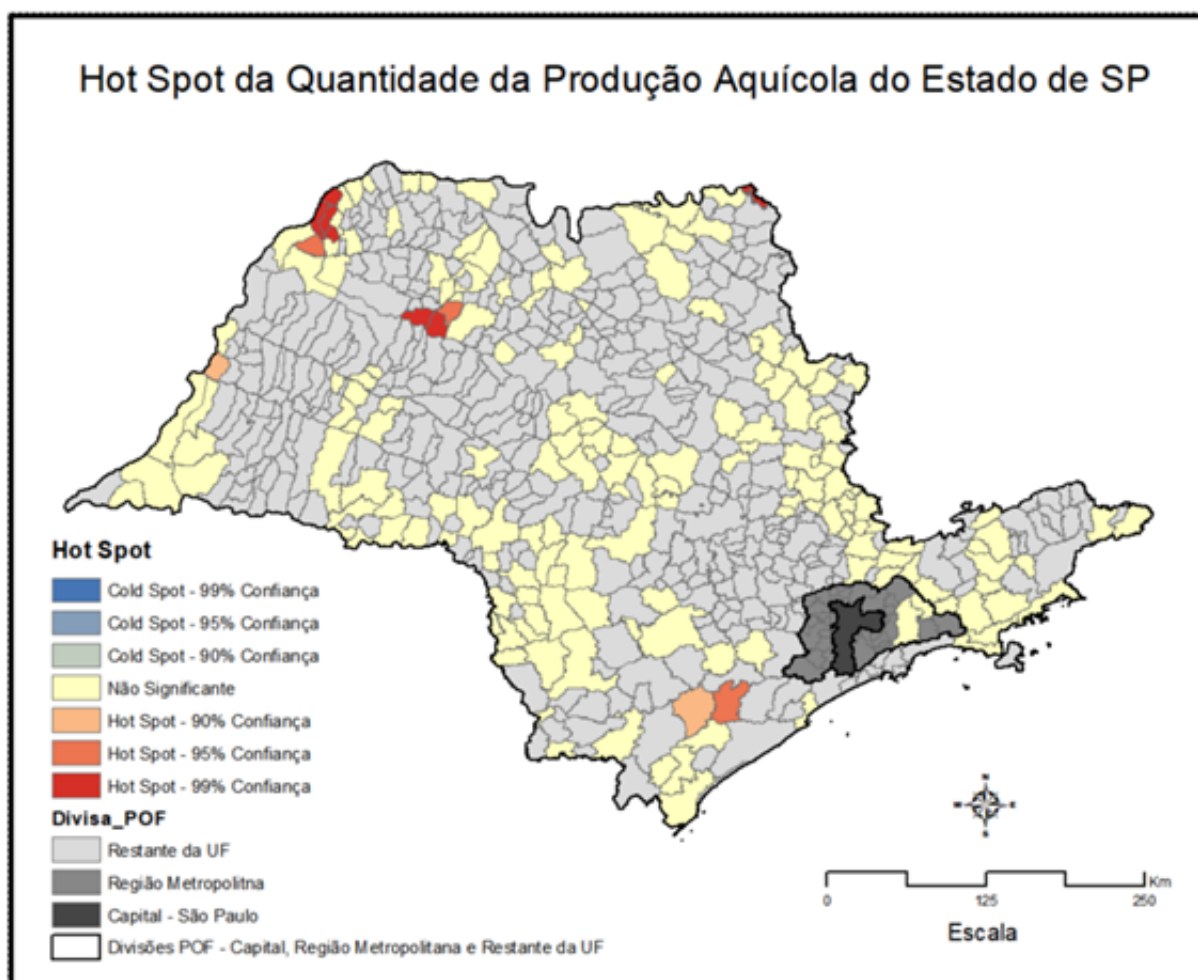
A Figura 29 apresenta o Centro Médio e a Elipse de Desvio Padrão Direcional (Standard Deviation Ellipse) da quantidade produzida de produtos aquícolas nos municípios do estado de São Paulo. O Centro Médio, identificado entre os municípios de Mendonça, Adolfo e Ubarana, representa a localização geográfica média ponderada pelos valores de produção. Isso indica que essa região concentra um peso significativo na distribuição espacial da produção aquícola paulista, funcionando como um ponto de equilíbrio espacial da atividade.

A elipse direcional evidencia o padrão de dispersão e orientação geográfica da produção. A forma alongada da elipse, com eixo principal orientado aproximadamente no sentido noroeste–sudeste, mostra que a produção aquícola está distribuída de forma mais intensa ao longo desse corredor territorial. Esse padrão sugere um alinhamento espacial da atividade produtiva que pode estar associado a fatores hidrográficos (rios e represas das bacias Tietê, São José dos Dourados e Turvo), infraestrutura de irrigação e tradição regional na piscicultura.

Além disso, observa-se que a maior parte da área destacada pela elipse corresponde ao interior do estado, especialmente regiões com forte presença de pequenos e médios estabelecimentos rurais. A Região Metropolitana de São Paulo e o litoral praticamente não entram na área de maior concentração, indicando que a produção aquícola paulista é predominantemente interiorana e associada a municípios de menor densidade populacional.

7.9.2.2 Hot Spot (Gi*)

Figura 30 – Hot Spot da Produção Aquícola (kg) – Municípios do Estado de SP



Fonte: IBGE, SIDRA (2017). Elaborada pela autora.

A Figura 30 apresenta os resultados da análise de Hot Spot (Getis–Ord G_i^*) aplicada à quantidade produzida (kg) de produtos aquícolas no estado de São Paulo. Assim como ocorre na distribuição geral da atividade produtiva, os municípios representados em cinza correspondem às unidades territoriais sem registro de produção aquícola, compondo uma parcela majoritária do estado.

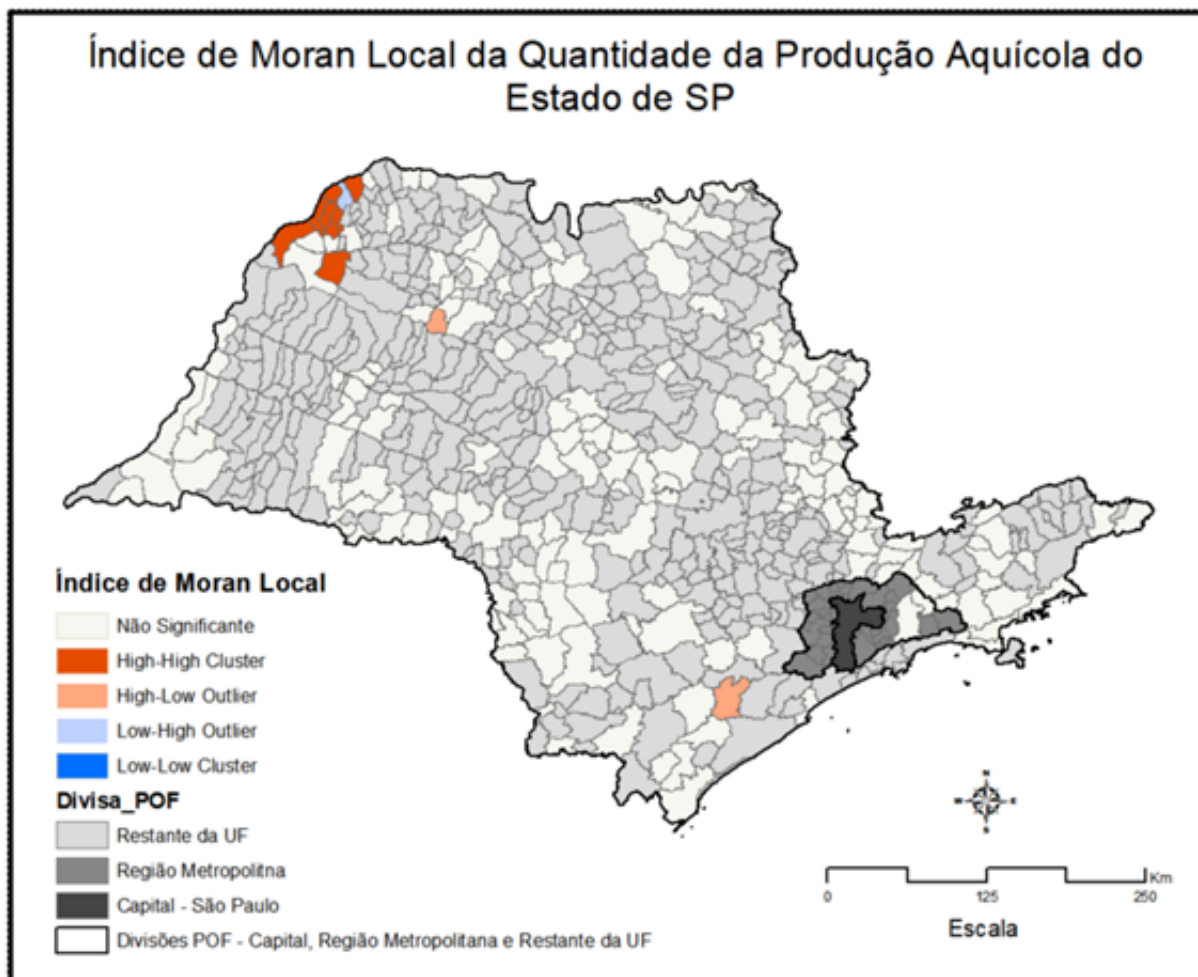
Os resultados evidenciam *hot spots* estatisticamente significativos concentrados no interior paulista, indicando agrupamentos locais de municípios com elevados volumes de produção. Esses agrupamentos surgem quando um município com grande quantidade produzida está cercado por outros municípios igualmente produtivos, configurando áreas com forte intensidade da atividade aquícola.

Observa-se ainda que as áreas de maior concentração produtiva não coincidem espacialmente com as regiões do estado que apresentam maior proporção de consumo domiciliar de pescado — especialmente a capital e sua região metropolitana, que se destacam pela demanda proporcionalmente mais elevada. Dessa forma, verifica-se um descompasso territorial entre oferta e demanda, em que polos de produção intensiva se localizam no interior, enquanto o maior consumo proporcional se concentra no eixo metropolitano.

Além disso, como o volume produzido e o valor econômico da produção são medidas fortemente relacionadas, a distribuição espacial dos *hot spots* tende a refletir, de forma proporcional, a estrutura produtiva do setor aquícola no estado. Isso reforça a consistência dos resultados e indica que os polos produtivos são estruturais e não apenas reflexo de uma dimensão específica da atividade aquícola.

7.9.2.3 Índice de Moran Local (LISA)

Figura 31 – Índice de Moran Local (LISA) da Produção Aquícola (kg) – Municípios do Estado de SP



Fonte: IBGE, SIDRA (2017). Elaborada pela autora.

A Figura 31 apresenta o Índice de Moran Local (LISA) aplicado à quantidade produzida de produtos aquícolas no estado de São Paulo. A matriz de pesos espaciais utilizada baseou-se no *inverso da distância*, técnica apropriada para variáveis contínuas, pois atribui maior influência aos municípios mais próximos e reduz gradativamente o peso das unidades mais distantes. Essa abordagem permite captar padrões de autocorrelação espacial que não dependem exclusivamente da contiguidade territorial, mas da intensidade da proximidade geográfica.

Os municípios representados em cinza correspondem às áreas que não registraram produção aquícola no período analisado, compondo a maior parte do território estadual. Os resultados mostram a presença de um *cluster* High–High na região noroeste, indicando agrupamentos de municípios com elevada produção cercados por vizinhos igualmente produtivos, caracterizando um polo produtivo consolidado e espacialmente coerente. Esse polo encontra-se distante da

capital e da região metropolitana, revelando que a principal área de concentração produtiva não coincide com os maiores centros consumidores do estado.

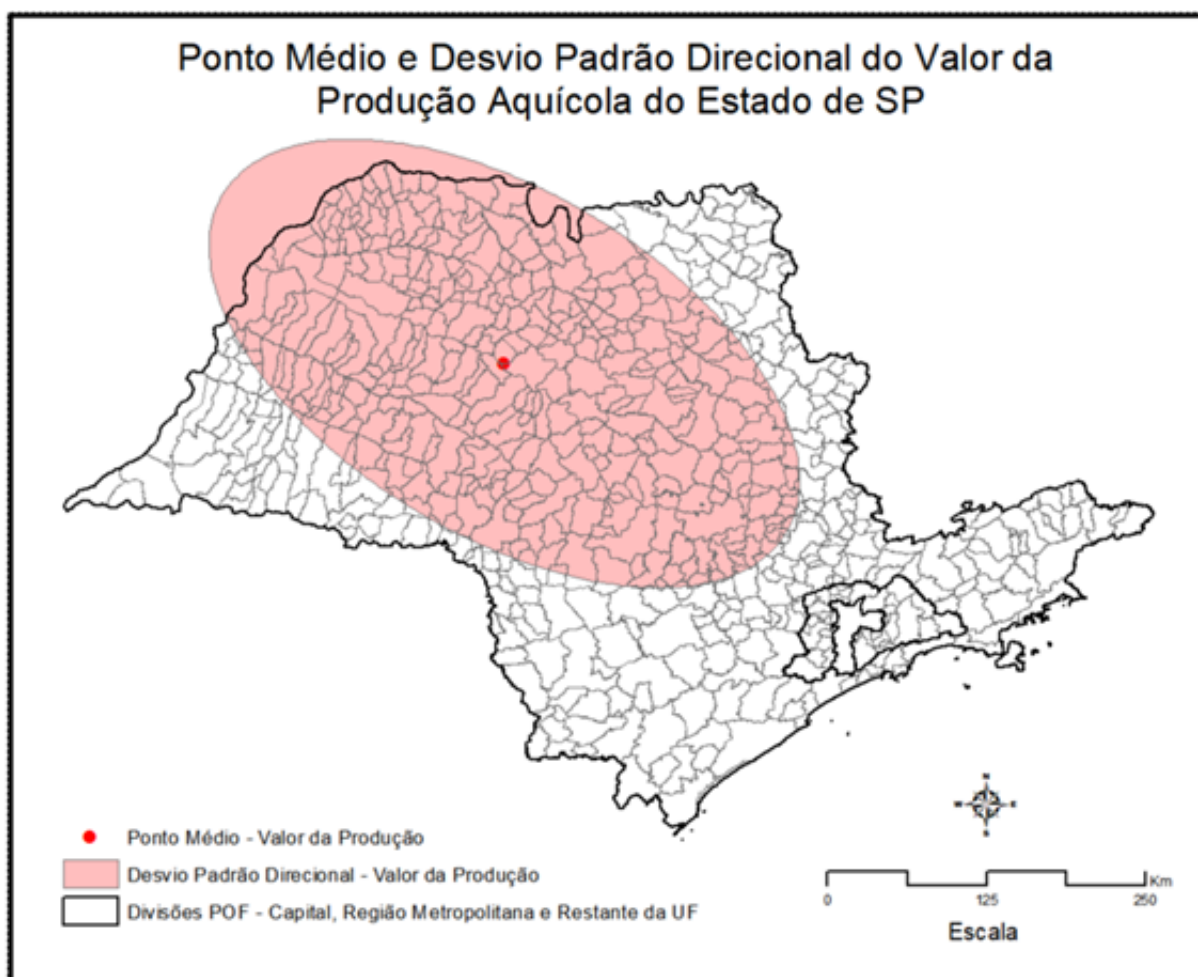
Além disso, observam-se dois tipos de *outliers* espaciais: um padrão High–Low, em que um município com produção elevada está circundado por municípios de baixa produção (como o caso de Zacarias e Juquiá), e um padrão Low–High, no qual um município com produção reduzida está rodeado por municípios com valores altos (como Santa Rita d’Oeste). Esses outliers evidenciam pequenas quebras de continuidade espacial dentro do padrão produtivo regional, sinalizando transições entre áreas intensivas e áreas pouco produtivas.

Em conjunto, os resultados reforçam que a atividade aquícola apresenta forte concentração no interior do estado, distante das áreas de maior consumo proporcional, como a capital e sua região metropolitana. Esse distanciamento territorial sugere uma cadeia de abastecimento dependente de fluxos intermunicipais, nos quais municípios com alta produção no interior atendem regiões densamente povoadas que, apesar de elevadas taxas de consumo, não possuem produção própria expressiva.

7.9.3 Valor da Produção Aquícola (R\$)

7.9.3.1 Centro Médio e Distribuição Direcional

Figura 32 – Centro Médio e Distribuição Direcional do Valor da Produção Aquícola em Reais – Municípios do Estado de SP



Fonte: IBGE, SIDRA (2017). Elaborada pela autora.

A Figura 32 apresenta o centro médio e a elipse de desvio padrão direcional do **valor da produção aquícola** no estado de São Paulo. O ponto central estimado situa-se na região entre os municípios de **Sabino**, **Novo Horizonte** e **Sales**, indicando que essa área representa o centro geográfico ponderado do valor econômico gerado pela aquicultura no estado.

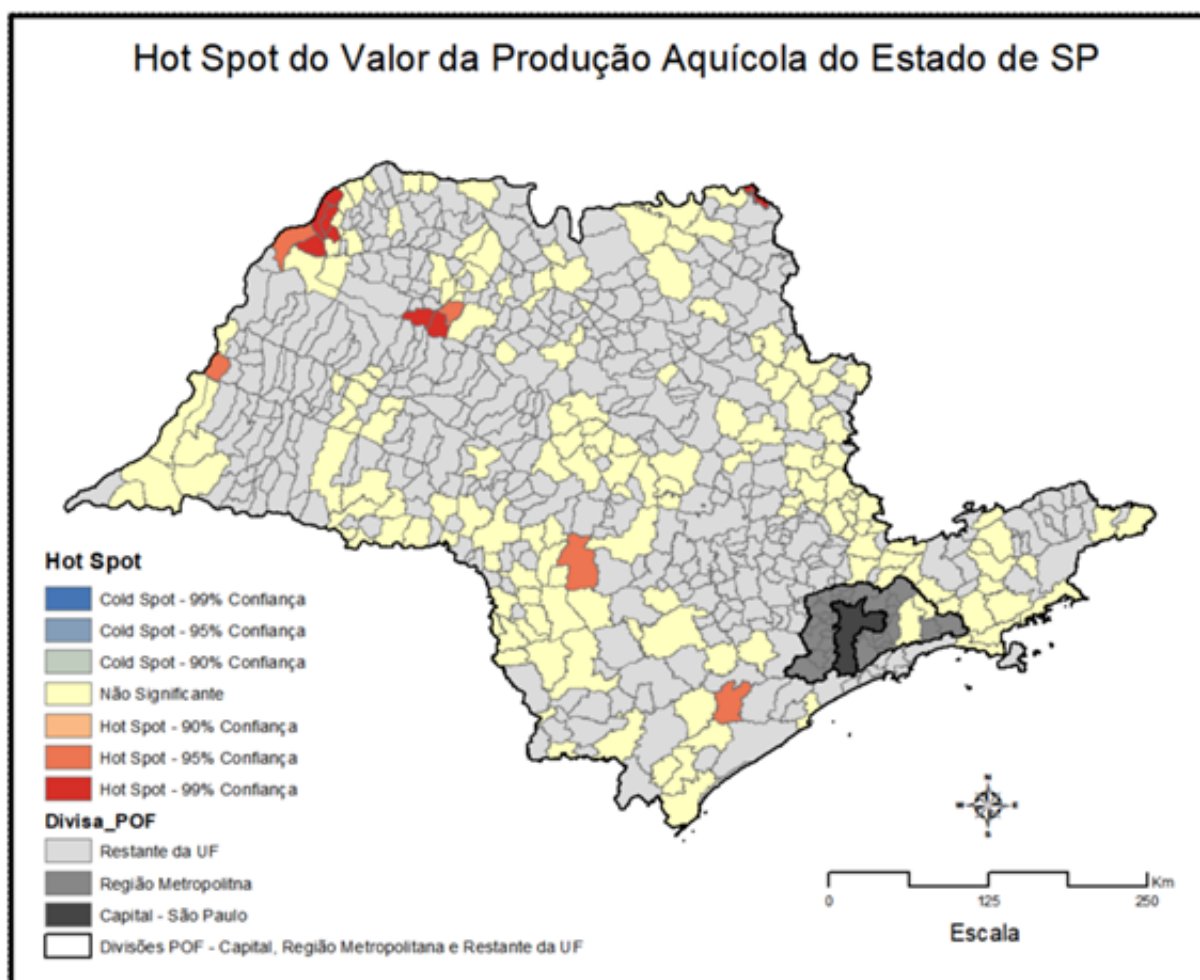
A elipse direcional evidencia uma distribuição **marcadamente alongada no sentido noroeste-sudeste**, sugerindo que o valor da produção não está uniformemente espalhado pelo território, mas concentrado ao longo desse eixo. A maior densidade de municípios com valor econômico significativo localiza-se no **Noroeste Paulista**, região tradicionalmente caracterizada pela presença de piscicultura em tanques-redes e viveiros escavados.

A amplitude da elipse revela que a atividade aquícola economicamente relevante se estende também por porções do Centro e do Oeste do estado, ainda que com menor intensidade. Já o **leste e o sudeste de São Paulo**, sobretudo a Região Metropolitana e a faixa litorânea, exibem baixa intensidade relativa, indicando que a geração de valor da produção aquícola é muito menos expressiva nessas áreas.

Em síntese, o padrão espacial observado confirma um **núcleo de maior valor econômico agregado à produção aquícola** concentrado no interior, com forte orientação geográfica, refletindo tanto características ambientais favoráveis quanto a estrutura produtiva consolidada dessa região.

7.9.3.2 Hot Spot (G_i^*)

Figura 33 – Hot Spot do Valor da Produção Aquícola em Reais – Municípios do Estado de SP



Fonte: IBGE, SIDRA (2017). Elaborada pela autora.

A Figura 33 apresenta os resultados da análise de Hot Spot (Getis–Ord G_i^*) aplicada ao valor da produção aquícola no estado de São Paulo. As áreas representadas em cinza correspondem

aos municípios que não registraram produção aquícola no período analisado, constituindo a maior parte do território estadual.

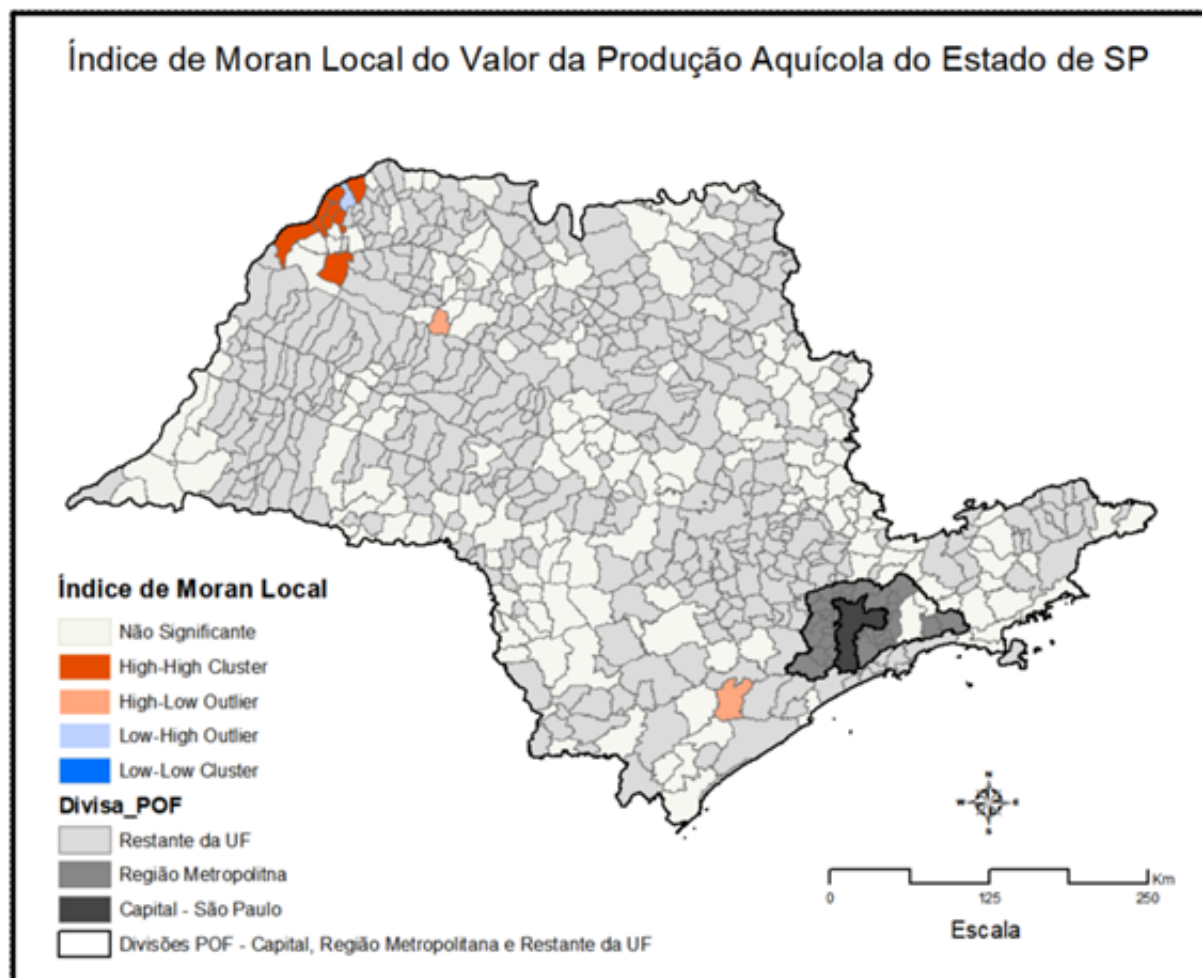
Os resultados evidenciam a presença de *hot spots* estatisticamente significativos em áreas interioranas do estado. Esses *hot spots* representam agrupamentos de municípios vizinhos com valores simultaneamente elevados de produção — isto é, não se trata de um município isolado com produção alta, mas de um conjunto espacial onde valores elevados reforçam-se mutuamente. Esse comportamento é central ao método G_i^* , que identifica concentrações locais anormalmente altas e revela regiões onde a produção aquícola apresenta intensidade muito superior ao padrão estatal médio.

Observa-se que tais polos produtivos localizados no interior contrastam fortemente com a capital e região metropolitana, que, segundo a POF 2017–2018, concentram a maior proporção de consumo domiciliar de pescado. Assim, diferentemente do que poderia ser esperado, produção e consumo não se sobrepõem espacialmente: as áreas que mais produzem estão distantes das áreas que mais consomem proporcionalmente.

Esse padrão reforça que a dinâmica territorial do setor aquícola paulista depende de fluxos entre municípios e regiões, evidenciando que a oferta não ocorre necessariamente próxima aos centros consumidores. Pelo contrário, os dados sugerem que os polos produtores do interior abastecem regiões densamente povoadas e altamente consumidoras, como a capital e sua região metropolitana.

7.9.3.3 Índice de Moran Local (LISA)

Figura 34 – Índice de Moran Local (LISA) do Valor da Produção Aquícola em Reais – Municípios do Estado de SP



Fonte: IBGE, SIDRA (2017). Elaborada pela autora.

A Figura 34 apresenta o Índice de Moran Local aplicado ao valor da produção aquícola no estado de São Paulo, utilizando matriz de pesos baseada no inverso da distância. Como nos mapas anteriores, a área em cinza corresponde aos municípios sem registro de produção aquícola, que compõem a maior parte do território estadual e não participam da formação de padrões de autocorrelação espacial.

Os resultados revelam a presença de *clusters* do tipo High-High na porção noroeste do estado, indicando municípios com valores de produção elevados cercados por vizinhos que também apresentam valores altos. Esse padrão evidencia a existência de um polo produtivo consolidado e distante da capital e da Região Metropolitana de São Paulo, reforçando a concentração da atividade aquícola em áreas interioranas.

Identificam-se também ocorrências pontuais de outliers espaciais, como um município do tipo High–Low (valor alto rodeado por valores baixos) e outro do tipo Low–High (valor baixo cercado por valores altos), sinalizando heterogeneidades locais dentro deste contexto de produção concentrada.

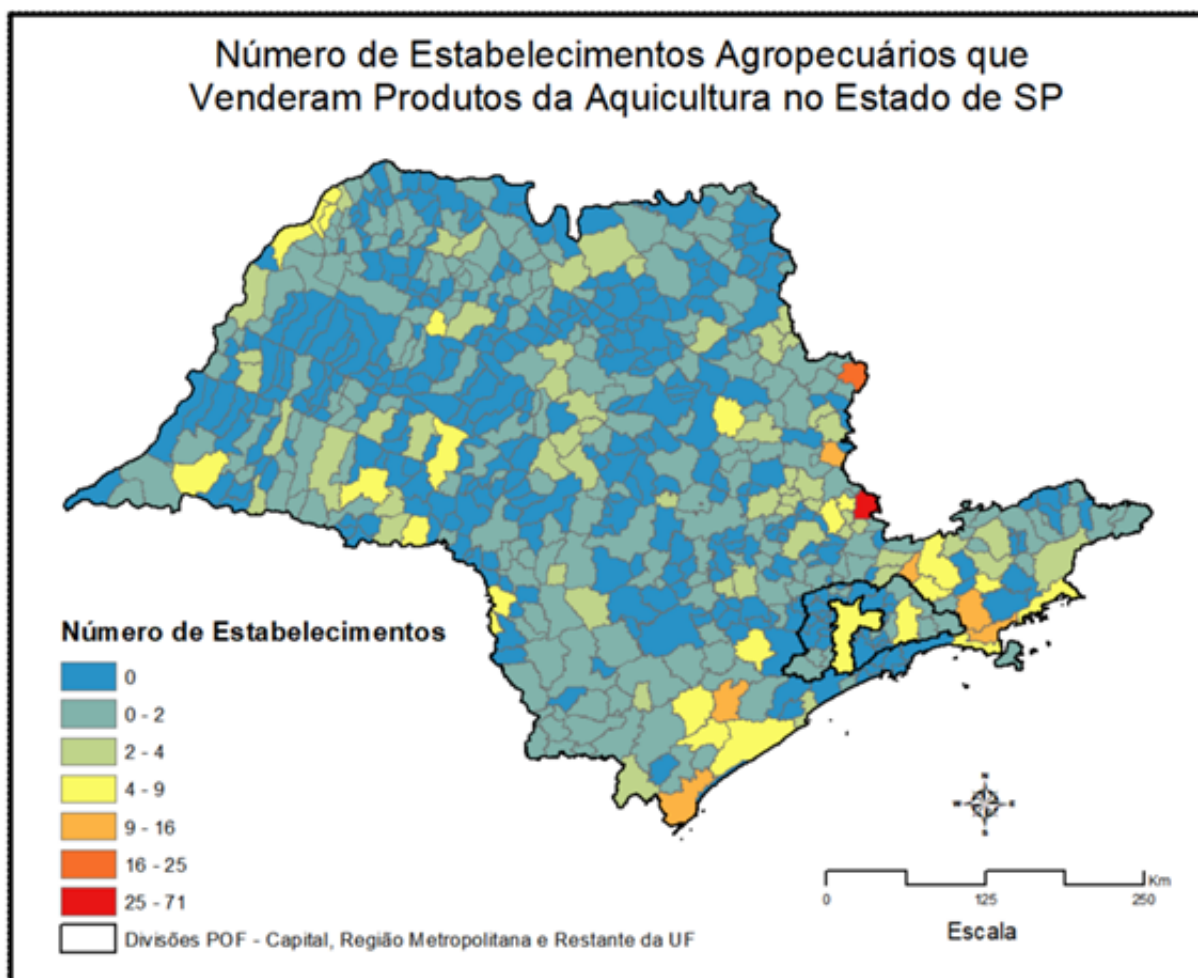
Em comparação com os resultados obtidos para a quantidade produzida, observa-se que um dos municípios anteriormente classificados como High–High — Nova Canaã Paulista — deixou de apresentar significância estatística quando considerada a variável de valor. Esse resultado sugere que, embora o município mantenha volume produtivo relevante, seu retorno econômico relativo não acompanha o padrão espacialmente elevado observado nos demais municípios do cluster.

De modo geral, os padrões de autocorrelação espacial indicam a formação de um núcleo produtivo interiorano bem definido, reforçando a assimetria territorial entre as áreas de maior produção aquícola e a capital e Região Metropolitana, onde se concentra a maior proporção de consumo domiciliar de pescado segundo a POF 2017–2018.

7.9.4 Número de Estabelecimentos Aquícolas

7.9.4.1 Plot da Variável

Figura 35 – Centro Médio e Distribuição Direcional do Número de Estabelecimentos Aquícolas – Municípios do Estado de SP



Fonte: IBGE, SIDRA (2017). Elaborada pela autora.

A Figura 35 apresenta a distribuição do número de estabelecimentos agropecuários que venderam produtos provenientes da aquicultura no estado de São Paulo. Observa-se um padrão espacial marcado pela predominância de municípios com poucos ou nenhum estabelecimento registrado, especialmente no interior do estado, onde grande parte do território apresenta valores entre 0 e 2 estabelecimentos, evidenciando baixa inserção formal da atividade aquícola.

As maiores concentrações aparecem em faixas específicas do território paulista, principalmente em áreas do litoral e em alguns pontos isolados do interior, onde se destacam municípios com valores entre 9 e 16, 16 a 25 e, em poucos casos, acima de 25 estabelecimentos. Esse padrão sugere que a presença de estabelecimentos agropecuários ligados à aquicultura apresenta caráter bastante localizado, com polos mais expressivos próximos às regiões costeiras e ao entorno da

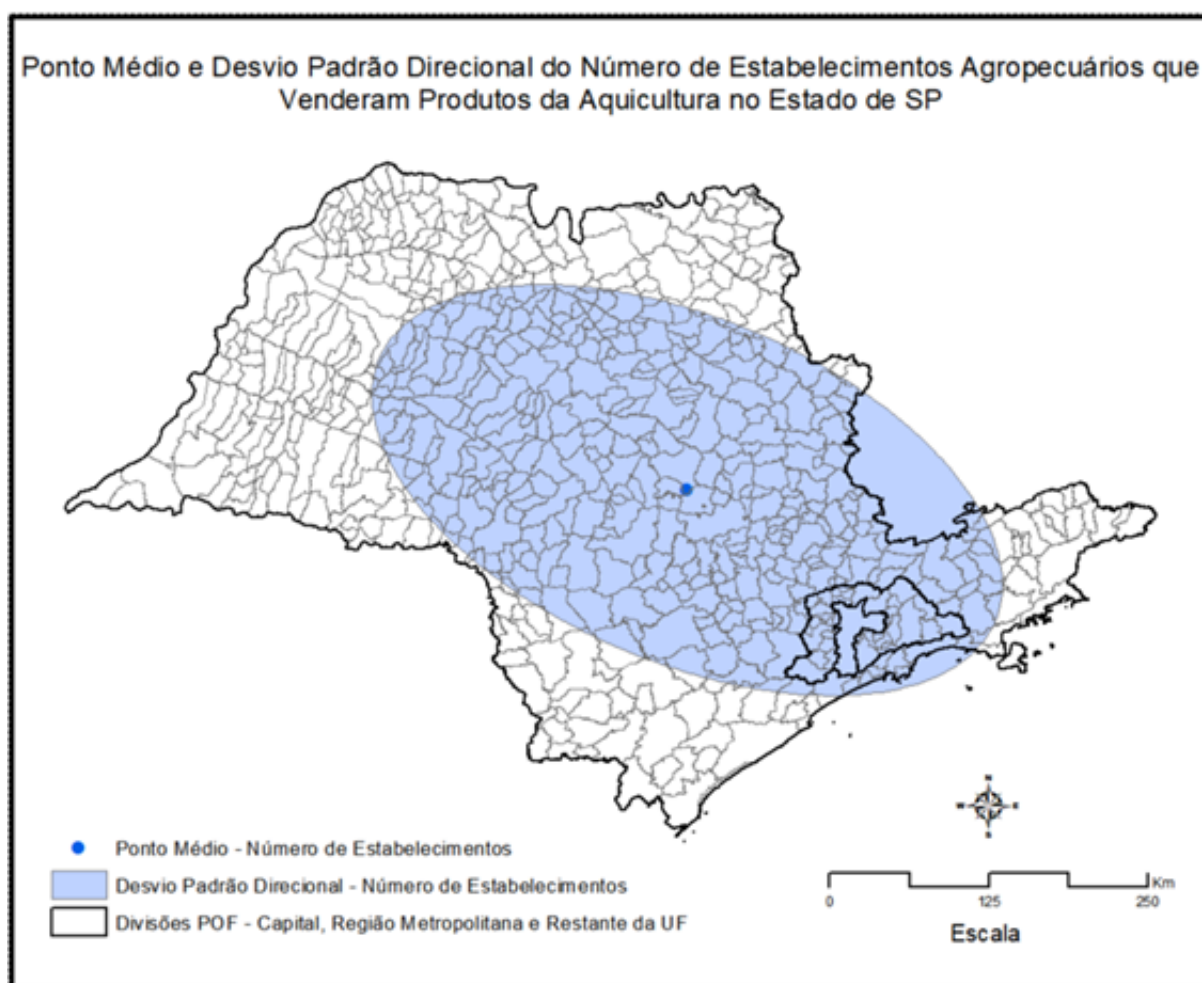
Região Metropolitana de São Paulo, onde há maior densidade populacional e maior demanda potencial.

Ainda que a produção aquícola mais intensa esteja situada no interior (conforme indicado em análises anteriores), o número de estabelecimentos apresenta maior concentração relativa nas áreas com maior conectividade econômica e logística, como o litoral e parte do sudeste paulista. Isso sugere que a comercialização de produtos aquícolas tende a se organizar espacialmente de maneira distinta da produção, possivelmente refletindo uma estrutura de mercado voltada para regiões consumidoras e melhor abastecidas.

Assim, o mapa revela um setor caracterizado por forte heterogeneidade espacial, com predominância de municípios sem participação na cadeia aquícola e concentrações pontuais de maior atividade comercial associada ao escoamento dos produtos.

7.9.4.2 Centro Médio e Distribuição Direcional

Figura 36 – Centro Médio e Distribuição Direcional do Número de Estabelecimentos Aquícolas – Municípios do Estado de SP



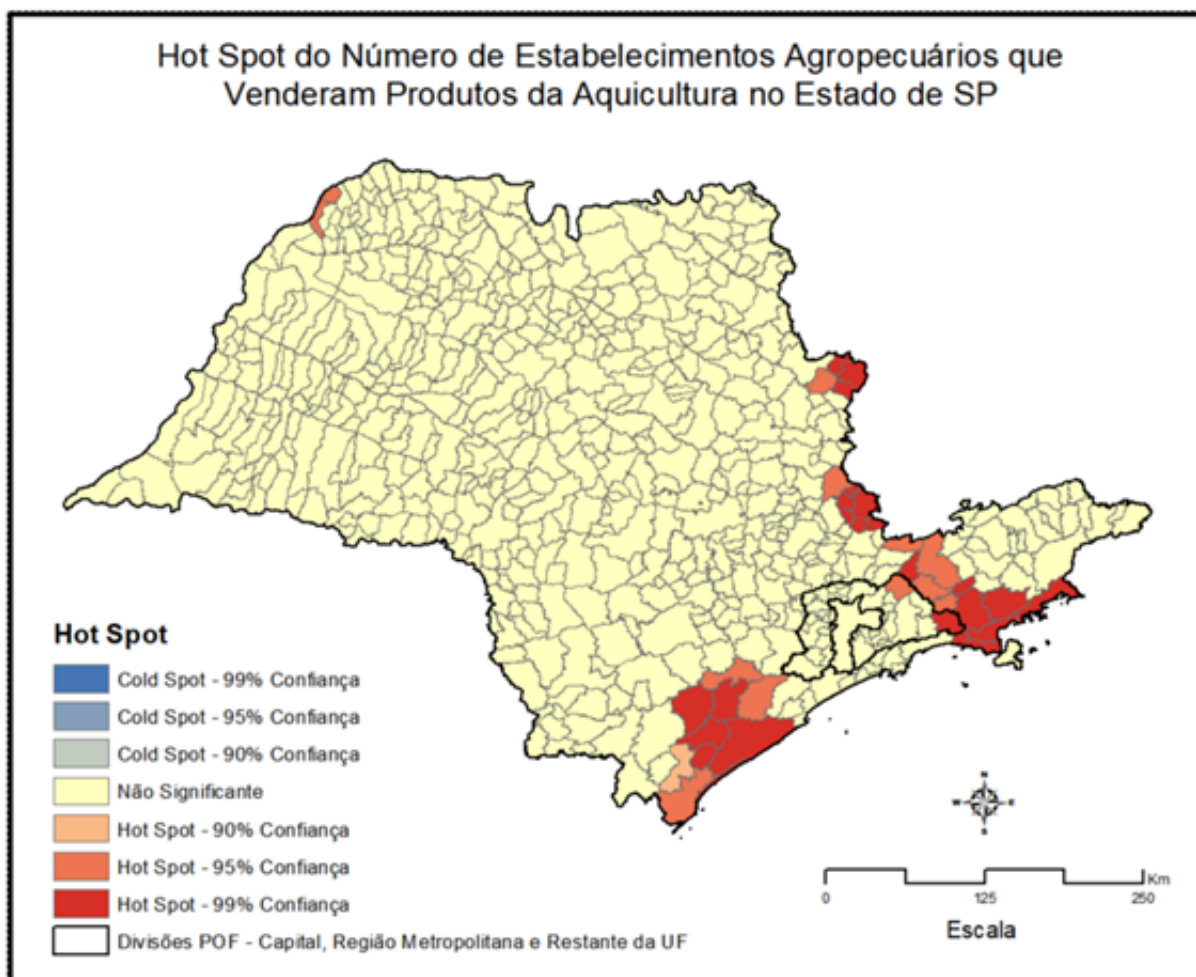
A Figura 36 apresenta o ponto médio e a Elipse de Desvio Padrão Direcional referente ao número de estabelecimentos agropecuários que comercializaram produtos da aquicultura no estado de São Paulo. O centro médio encontra-se localizado na região que abrange os municípios de Brotas, São Pedro e Torrinha, indicando um ponto de equilíbrio geográfico da distribuição espacial dessa atividade econômica.

A elipse revela um padrão de dispersão predominantemente orientado no sentido *sudoeste–nordeste*, sugerindo que a distribuição dos estabelecimentos tende a acompanhar esse eixo territorial. A amplitude da elipse indica que tais estabelecimentos estão amplamente distribuídos pelo estado, mas com maior concentração nas regiões centrais e intermediárias de São Paulo. Além disso, observa-se menor intensidade no extremo sul e ao longo da maior parte da Região Metropolitana de São Paulo, evidenciando uma predominância das atividades aquícolas em áreas com características rurais mais marcantes.

Esse padrão espacial sugere que a logística de produção e comercialização de produtos aquícolas tende a se consolidar em municípios com disponibilidade territorial, presença de estabelecimentos rurais de médio porte e proximidade relativa a polos consumidores regionais, reforçando a heterogeneidade espacial da cadeia produtiva aquícola no estado.

7.9.4.3 Hot Spot (G_i^*)

Figura 37 – Hot Spot do Número de Estabelecimentos Aquícolas – Municípios do Estado de SP



Fonte: IBGE, SIDRA (2017). Elaborada pela autora.

A Figura 37 apresenta os resultados da análise de Hot Spot (Getis–Ord G_i^*) aplicada ao número de estabelecimentos agropecuários que venderam produtos da aquicultura no estado de São Paulo. A maior parte do território aparece em amarelo, indicando ausência de significância estatística — ou seja, municípios com números pouco expressivos ou distribuídos de forma dispersa.

Diferentemente dos padrões observados para a quantidade e o valor da produção aquícola, os resultados mostram a formação de *hot spots* significativos na faixa litorânea e também em parte da Região Metropolitana de São Paulo. Esses agrupamentos representam áreas em que municípios vizinhos apresentam, simultaneamente, números elevados de estabelecimentos ativos no setor, reforçando um padrão espacial de concentração da estrutura comercial aqui representada.

Esse contraste em relação às demais variáveis é relevante: enquanto a quantidade e o valor da produção apresentaram seus *hot spots* predominantemente no interior do estado, a distribuição

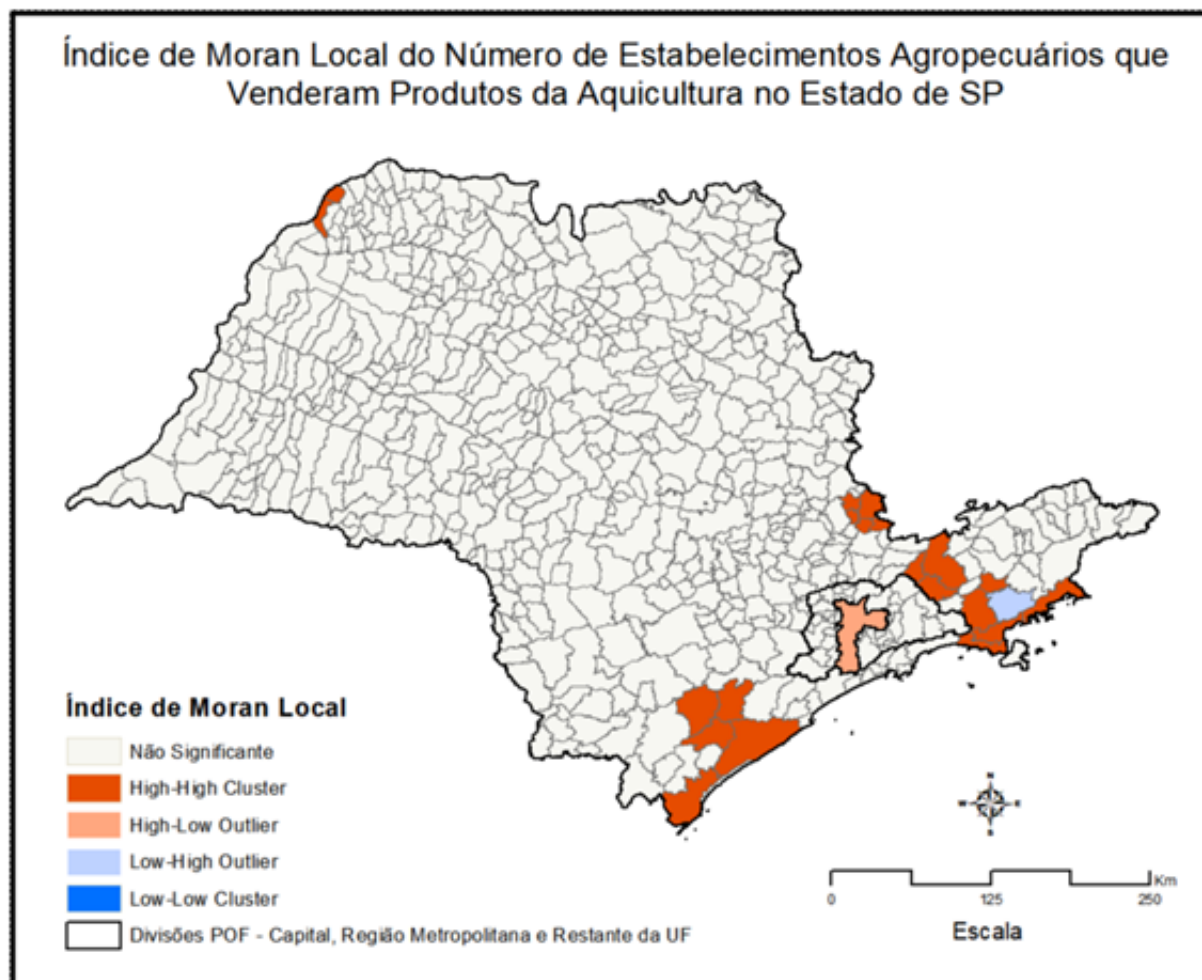
espacial da infraestrutura comercial (estabelecimentos que vendem produtos da aquicultura) desloca-se para regiões mais populosas, urbanizadas e próximas ao litoral. Esse padrão sugere que, embora a produção intensiva esteja localizada no interior, a comercialização tende a se concentrar em áreas com maior densidade populacional, maior demanda potencial e maior integração logística, como a Região Metropolitana de São Paulo e a faixa costeira.

Esse resultado reforça a existência de uma divisão territorial entre produção (interior) e comercialização/mercado consumidor (litoral e metrópole), indicando cadeias de circulação mais complexas, nas quais produtos aquícolas percorrem longas distâncias para atender mercados urbanos mais dinâmicos e com maior capacidade de consumo.

Em suma, a análise de Hot Spot para o número de estabelecimentos revela que a estrutura comercial do setor aquícola segue lógica distinta da distribuição da produção, evidenciando a importância da proximidade a grandes centros consumidores para a localização de estabelecimentos agropecuários voltados à venda de pescado.

7.9.4.4 Índice de Moran Local (LISA)

Figura 38 – Índice de Moran Local (LISA) do Número de Estabelecimentos Aquícolas – Municípios do Estado de SP



Fonte: IBGE, SIDRA (2017). Elaborada pela autora.

A Figura 38 apresenta o Índice de Moran Local (LISA) aplicado ao número de estabelecimentos agropecuários que comercializaram produtos da aquicultura no estado de São Paulo. Os municípios em cinza representam áreas sem registro dessa atividade, indicando ampla predominância de ausência produtiva ao longo do território estadual.

A análise foi conduzida utilizando-se a matriz de pesos espaciais baseada no método da *rainha*, que considera como vizinhos todos os municípios que compartilham qualquer ponto de fronteira (aresta ou vértice). Essa forma de contiguidade é especialmente adequada para dados municipais, pois capta interações territoriais mais amplas, permitindo a identificação de clusters mesmo em regiões com formato irregular.

Os resultados revelam a presença de *clusters High-High* ao longo do litoral e em porções específicas do interior paulista. Esses agrupamentos indicam municípios com número elevado de

estabelecimentos agropecuários cercados por vizinhos que também apresentam valores altos, configurando áreas de concentração estruturada da atividade.

Além dos clusters positivos, foram identificados dois outliers espaciais: um *High–Low* localizado na capital (São Paulo), que apresenta valor elevado cercado por municípios com valores baixos; e um *Low–High* em Natividade da Serra, onde um valor reduzido contrasta com vizinhos de valores mais altos. Esses casos sinalizam padrões locais atípicos na distribuição da atividade aquícola.

No conjunto, o mapa revela que as áreas de destaque não se concentram apenas no interior, mas também incluem extensos trechos do litoral paulista, diferentemente do padrão observado para quantidade e valor da produção. Essa configuração sugere que a presença de estabelecimentos comercializadores é mais fortemente associada à dinâmica econômica e populacional litorânea e metropolitana, enquanto as regiões produtivas mais intensivas permanecem distantes da capital e de sua região metropolitana.

8 CONCLUSÃO

A análise dos microdados da POF 2017–2018 permitiu compreender, de forma abrangente, como as características socioeconômicas, demográficas e estruturais dos domicílios paulistas se relacionam com o consumo domiciliar de pescado. A preparação cuidadosa da base de dados, incluindo filtragem de indivíduos, integração de tabelas, criação de indicadores alimentares e construção de novas variáveis, estabeleceu um conjunto robusto de informações capaz de apoiar análises descritivas, multivariadas e espaciais.

Os resultados descritivos revelaram que o consumo de pescado apresenta baixa variabilidade e forte concentração em valores reduzidos, configurando-se como um item de consumo esporádico para grande parte das famílias. Essa característica influenciou diretamente o desempenho dos métodos multivariados, uma vez que variáveis com baixa variância tendem a oferecer menor poder discriminativo em modelos exploratórios.

A avaliação da consistência interna, via alfa de Cronbach, e os índices de adequação amostral (como o KMO) confirmaram a pertinência estatística da Análise Fatorial, possibilitando a extração de fatores latentes que sintetizaram dimensões relevantes do contexto socioeconômico. Ainda assim, a inspeção dos fatores extraídos mostrou que, apesar de representarem estruturas coerentes — como condições de infraestrutura urbana, bem-estar, saneamento, situação socioeconômica e perfil demográfico — essas dimensões não se traduziram em diferenças substanciais entre os níveis de consumo de pescado.

Essa tendência foi corroborada pelas Análises de Correspondência. Na AC simples, que utilizou diretamente as categorias de consumo de pescado, os mapas mostraram forte concentração das categorias próximas ao centro dos planos fatoriais, revelando baixo poder discriminativo e pouca capacidade de separar perfis alimentares distintos. Entretanto, ao expandir a análise para a Análise de Correspondência Múltipla (ACM), a inclusão simultânea de diferentes variáveis qualitativas permitiu maior diferenciação dos grupos, evidenciando padrões mais consistentes entre condições socioeconômicas e perfis domiciliares.

Nas análises de cluster, tanto aquelas baseadas nos fatores quanto nas variáveis ordinais, o mesmo padrão se manteve: embora os grupos formados apresentassem diferenças socioeconômicas internas, essas diferenças produziram apenas variações modestas nos indicadores de consumo. O teste de Kruskal–Wallis confirmou que, apesar de pequenas diferenças pontuais entre clusters, o comportamento alimentar relacionado ao pescado permanece pouco sensível às segmentações obtidas.

A aplicação do método de Two-Way Joining complementou as análises anteriores ao oferecer uma visualização integrada das variáveis ordinais em relação ao consumo de pescado. Embora muitos atributos tenham apresentado padrões semelhantes entre os domicílios que consumiram e os que não consumiram, o mapa permitiu identificar uma tendência de intensidades ligeiramente maiores entre variáveis associadas a melhores condições de vida no grupo que declarou consumo.

Essa evidência, ainda que moderada, reforça o entendimento de que o pescado tende a ser mais acessível em contextos de maior estabilidade estrutural e melhores condições domiciliares, alinhando-se às tendências observadas nas demais técnicas empregadas ao longo do estudo.

Por fim, as análises espaciais — incluindo autocorrelação global e local, índices de Moran, Hot Spot, centro médio e elipses de distribuição direcional — ampliaram a compreensão territorial do fenômeno. Os resultados apontaram padrões geográficos coerentes com a distribuição populacional e com os polos de aquicultura e comercialização, mas novamente com baixa associação direta entre localização e magnitude do consumo. Ainda que alguns municípios apresentem maior presença de pescado na dieta, o padrão geral não sugere forte concentração espacial ou formação de clusters intensos, reforçando o caráter disperso e heterogêneo desse tipo de consumo no estado.

Em conjunto, os achados demonstram que o consumo domiciliar de pescado em São Paulo é influenciado por múltiplos fatores, mas nenhum deles, isoladamente, mostra-se determinante. Renda, escolaridade, padrões de saúde e características familiares explicam parte da variabilidade, mas não estabelecem relações fortes o suficiente para formar perfis consistentes. Isso indica que o pescado não desempenha papel central e regular na alimentação das famílias, permanecendo como um produto de acesso limitado e consumo eventual.

Do ponto de vista metodológico, o estudo mostrou que a utilização integrada de técnicas estatísticas descritivas, multivariadas e espaciais constitui uma abordagem sólida para a investigação de padrões alimentares complexos. A aplicação conjunta de análise fatorial, correspondência, clustering e técnicas espaciais permitiu explorar diferentes dimensões do fenômeno, mesmo diante das limitações impostas pela baixa variabilidade da variável de interesse.

Como perspectivas futuras, recomenda-se o aprofundamento da análise em duas direções: (i) a ampliação para outras regiões do país, comparando características culturais, disponibilidade regional e influência dos preços; e (ii) a incorporação de informações sobre frequência de compra, consumo extradomiciliar e preferências individuais, que podem fornecer mais nuances sobre a dinâmica do setor aquícola e da alimentação baseada em pescado.

Assim, conclui-se que, embora o consumo de pescado no estado de São Paulo apresente padrões distribuídos de forma heterogênea e pouco concentrada, sua compreensão exige abordagens multidimensionais que combinem informação socioeconômica, espacial e comportamental, contribuindo para o planejamento de políticas públicas, programas de incentivo e estratégias de educação alimentar. A partir dos resultados obtidos ao longo das análises, também é possível inferir a existência de um padrão sutil, porém consistente, segundo o qual famílias em melhores condições de vida tendem a apresentar níveis ligeiramente mais elevados de consumo de pescado. Embora essa relação não tenha se mostrado suficientemente forte para discriminar de maneira clara os diferentes perfis de consumo, a tendência observada dialoga com a teoria que associa maior renda, melhor escolaridade e melhores condições estruturais a dietas mais diversificadas e à aquisição de alimentos de maior valor agregado. Dessa forma, mesmo em um cenário marcado por baixo consumo geral e variabilidade reduzida, os achados sugerem que melhores padrões

de vida podem estar associados, ainda que de forma moderada, a maior acesso e frequência de consumo de pescado, reforçando a necessidade de políticas integradas que ampliem o acesso e promovam a diversificação alimentar.

REFERÊNCIAS

- ANSELIN, L. Local indicators of spatial association — LISA. *Geographical Analysis*, v. 27, n. 2, p. 93–115, 1995.
- ANSELIN, L. A local indicator of multivariate spatial association: extending Geary's c. *Geographical Analysis*, v. 51, n. 2, p. 133–150, 2019.
- ANSELIN, L. *An introduction to spatial data science with GeoDa. Volume 1: Exploring spatial data*. Boca Raton: CRC Press, 2024.
- ANSELIN, L. *An introduction to spatial data science with GeoDa. Volume 2: Clustering spatial data*. Boca Raton: CRC Press, 2024.
- LUZARDO, A. J. R.; CASTAÑEDA FILHO, R. M.; RUBIM, I. B. Análise espacial exploratória com o emprego do índice de Moran. *GEOgraphia*, v. 19, n. 40, p. 161–179, 2017.
- ARAÚJO, D. M. *et al.* Aspectos de aquisição e consumo de peixes na feira livre da cidade de Penedo–Alagoas. *Boletim do Instituto de Pesca*, v. 41, n. 2, p. 429–440, 2018.
- BAILEY, T. C.; GATRELL, A. C. *Interactive spatial data analysis*. Essex: Longman Scientific and Technical, 1995.
- BARBOSA, H. T. B.; SAMPAIO, M. F. Consumo de pescado entre famílias de baixa renda da zona urbana de Presidente Figueiredo. *Revista Igapó*, v. 10, n. 1, p. 88–102, 2016.
- BARROS, A. F. *et al.* Investimento e custo de produção de peixes nativos em policultivo e monocultivo. *Brazilian Journal of Development*, v. 6, n. 3, p. 16342–16359, 2020.
- BERNARDO, L. S. C. *A percepção do consumidor face ao pescado de aquacultura*. 2017. Tese (Doutorado) — Instituto Politécnico de Leiria, Leiria, 2017.
- BRASIL. Ministério da Indústria, Comércio Exterior e Serviços. *Comex Stat: sistema para consultas e extração de dados do comércio exterior brasileiro*. 2022. Disponível em: <<http://comexstat.mdic.gov.br/pt/home>>. Acesso em: abr. 2022.
- DAI, Y. *et al.* Factors influencing Chinese consumer attitudes on the safety of aquatic products. *Journal of the World Aquaculture Society*, v. 53, n. 1, p. 47–59, 2022.
- FAO. *The State of World Fisheries and Aquaculture 2020: Sustainability in action*. Rome: FAO, 2020.
- FIRETTI, R. *et al.* Análise de variáveis estratégicas para o desenvolvimento da agropecuária no Pontal do Paranapanema. *Revista de Economia e Sociologia Rural*, v. 50, p. 141–156, 2012.

- FIRETTI, R. *et al.* Multivariate analysis of vegetative growth and productivity in banana cultivars. *Revista Ciência Agronômica*, v. 55, p. e20228571, 2024.
- FIRETTI, R.; ASTOLPHI, J. L. L.; GARCIA, S. M. Aquisição de pescados para consumo domiciliar na Região Sudeste: POF 2009. *Revista de Economia Agrícola*, v. 60, p. 17–30, 2013.
- FLORES, R. M. V. *et al.* Establishing linkages between consumer fish knowledge and demand for fillet attributes in Brazilian supermarkets. *Journal of International Food & Agribusiness Marketing*, v. 34, n. 4, p. 368–388, 2022.
- FORNARI, C. A. C. *et al.* Estudo sobre os hábitos de consumo de pescado da população de Palmas (TO). *Desafios*, v. 4, n. 4, p. 136–142, 2017.
- FREITAS, J.; VAZ-PIRES, P.; CÂMARA, J. S. From aquaculture production to consumption. *Aquaculture*, v. 518, p. 734857, 2020.
- GREENACRE, M. *Correspondence analysis in practice*. 3. ed. Boca Raton: CRC Press, 2017.
- GREENACRE, M. *Theory and applications of correspondence analysis*. London: Academic Press, 2007.
- HAIR, J. F. *et al.* *Análise multivariada de dados*. 6. ed. Porto Alegre: Bookman, 2009.
- IBGE. *Pesquisa de Orçamentos Familiares 2017–2018*. Rio de Janeiro: IBGE, 2020.
- JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. 6. ed. Upper Saddle River: Pearson Prentice Hall, 2007.
- JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A*, v. 379, p. 20200202, 2016.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, v. 23, n. 3, p. 187–200, 1958.
- KAUFMAN, L.; ROUSSEEUW, P. J. *Finding groups in data: an introduction to cluster analysis*. Hoboken: Wiley, 2009.
- KODINYE, D. C.; LU, Y. K. Using elbow and silhouette methods in hierarchical clustering. *International Journal of Computer Applications*, v. 177, n. 26, 2020.
- LEGENDRE, P.; LEGENDRE, L. *Numerical ecology*. 3. ed. Amsterdam: Elsevier, 2012.
- LESCOT, L. *Análise de correspondência: guia prático para estudos com variáveis qualitativas*. São Paulo: Cengage Learning, 2012.
- LOPES, A. V.; VASCONCELOS, B. P. L. *Análise espacial de dados pontuais utilizando Info-Map*. Presidente Prudente: FCT/UNESP, 2007.

- MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada*. Belo Horizonte: UFMG, 2005.
- PEDROZA FILHO, M. X. *et al.* O mercado de peixes da piscicultura no Brasil. Embrapa Pesca e Aquicultura, 2020.
- SCHULTER, E. P.; VIEIRA FILHO, J. E. R. Evolução da piscicultura no Brasil. *Texto para Discussão — IPEA*, n. 2328, 2017.
- SEBRAE. *Aquicultura no Brasil*. Brasília: SEBRAE-NA, 2015.
- SIQUEIRA, P. H. M. *et al.* Impacto da renda na demanda de proteína animal no Brasil. *Cadernos de Ciência & Tecnologia*, v. 39, n. 3, p. 27128, 2023.
- STANCU, V. *et al.* European consumer segments with potential for accepting innovative fish products. *Food Quality and Preference*, v. 99, p. 104560, 2022.
- VALENTI, W. C. *et al.* Aquaculture in Brazil: past, present and future. *Aquaculture Reports*, v. 19, p. 100611, 2021.
- VAZ, D. V.; HOFFMANN, R. Elasticidade-renda e concentração das despesas com alimentos no Brasil. *Revista de Economia*, v. 41, n. 75, 2020.
- VIEIRA FILHO, J. E. R.; SILVEIRA, J. M. F. J. A pesca artesanal e a aquicultura. *Texto para Discussão — IPEA*, n. 1567, 2012.
- WAGNER, Y. G. *et al.* Análise do consumo domiciliar de pescados no Brasil: POF 2017–2018. *Revista de Economia e Sociologia Rural*, v. 61, n. 3, p. e250494, 2023.
- WARD, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, v. 58, n. 301, p. 236–244, 1963.