

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
COORDENAÇÃO DO CURSO DE LICENCIATURA EM
MATEMÁTICA

DANIELI FELICHAK

ANÁLISE MULTIVARIADA APLICADA AO ESTUDO DE DADOS
QUALITATIVOS ORDINAIS UTILIZANDO OS MÉTODOS DE
QUANTIFICAÇÃO RIDITS E NÚMEROS INTEIROS ORDENADOS

TRABALHO DE CONCLUSÃO DE CURSO

TOLEDO

2019

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
COORDENAÇÃO DO CURSO DE LICENCIATURA EM
MATEMÁTICA

DANIELI FELICHAK

ANÁLISE MULTIVARIADA APLICADA AO ESTUDO DE
DADOS QUALITATIVOS ORDINAIS UTILIZANDO OS
MÉTODOS DE QUANTIFICAÇÃO RIDITS E NÚMEROS
INTEIROS ORDENADOS

Trabalho de Conclusão de Curso apresentado ao Curso de Licenciatura em Matemática da Universidade Tecnológica Federal do Paraná, Câmpus Toledo, como requisito parcial à obtenção do título de Licenciado em Matemática.

Orientador(a): Prof. Dra. Regiane Slongo Fagundes

TOLEDO

2019

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
COORDENAÇÃO DO CURSO DE LICENCIATURA EM
MATEMÁTICA

TERMO DE APROVAÇÃO

O Trabalho de Conclusão de Curso intitulado “ANÁLISE MULTIVARIADA APLICADA AO ESTUDO DE DADOS QUALITATIVOS ORDINAIS UTILIZANDO OS MÉTODOS DE QUANTIFICAÇÃO RIDITS E NÚMEROS INTEIROS ORDENADOS” foi considerado **APROVADO** de acordo com a ata nº -- de --/--/----

Fizeram parte da banca examinadora os professores:

Professora Orientadora Dra. Regiane Slongo Fagundes

Professora Dra. Daniela Trentin Nava

Professora Dra. Suellen Ribeiro Pardo Garcia

TOLEDO

2019

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me concedido saúde, coragem e inteligência para superar todos os obstáculos para que esse dia se concretizasse.

A minha orientadora Prof. Dra. Regiane Slongo Fagundes por toda sua atenção, dedicação e esforço para que este trabalho ficasse o melhor possível, de coração muito obrigada, vou levar cada um dos seus ensinamentos pra sempre.

A minha banca Prof. Dra. Suellen Ribeiro Pardo Garcia e Prof. Dra. Daniela Nava por terem disponibilizado do seu tempo para a correção e contribuições dadas para a melhoria deste trabalho.

A todos os professores que de alguma forma contribuíram para meu aprendizado e crescimento tanto pessoal, quanto profissional, vocês foram essenciais e levo cada um em meu coração.

Não tenho palavras para agradecer ao meu pai Milton, à minha mãe Jeuse e a minha irmã Natanieli que são minhas bases e pra quem dedico cada uma das minhas conquistas, que não mediram esforços para que essa etapa se concluísse, meu muito obrigada, amo muito vocês!

Ao meu namorado Leonardo Gripp por toda paciência, amor, carinho, compreensão, apoio, incentivo e por me trazer paz e calma nos momentos de ansiedade e desespero, te amo imensamente!

Agradeço de uma forma muito especial, as minhas amigas que se tornaram minha família em Toledo, Daiane Seibert, Daniela Ludvig, Adrieli Moretti, Tais Francini, Simoni Berger e Caroline Souza que foram exemplo de companheirismo e dedicação, que tornaram essa caminhada acadêmica mais leve e me proporcionaram momentos inesquecíveis. Cada uma de vocês tem um lugar muito especial no meu coração, obrigada por acreditarem e não desistirem de mim, amo muito vocês princesas!

A minha amiga Caroliny Grzegozeski que mesmo longe, sempre esteve ao meu lado me apoiando qualquer que fosse a situação.

E por fim a todos que de alguma forma contribuíram para que esse dia fosse possível. O meu muito obrigada!

RESUMO

O presente trabalho de conclusão de curso tem por objetivo comparar os métodos de quantificação Ridits e Números Inteiros Ordenados para análise de dados qualitativos ordinais utilizando a análise multivariada de Correspondência Simples e Análise de Agrupamento. O trabalho foi dividido em dois artigos, o primeiro conceitua o método multivariado de Análise de Correspondência Simples e Números Inteiros Ordenados, e o segundo conceitua a quantificação Ridits e Análise de Agrupamento. Os respectivos métodos foram aplicados a um conjunto de dados proveniente de uma pesquisa de satisfação dos usuários contemplados com o projeto Minha Casa Minha Vida do governo Federal, do conjunto habitacional Jardim das Orquídeas, localizado em Toledo/PR. Os critérios avaliados foram a organização espacial interna da moradia e nos entornos. O primeiro artigo indicou que a magnitude da associação entre as variáveis é alta, pois o teste χ^2 foi de 242,87 com 44 graus de liberdade para os cômodos e 245,36 com 36 graus de liberdade para o entorno. Já o mapa percentual bidimensional indicou que a área de serviço, banheiro, cozinha e as linhas de ônibus são as variáveis que geraram maior insatisfação. No segundo artigo a quantificação Ridits também apresentou forte adequação para a análise deste conjunto de dados, com um coeficiente de correlação cofenético muito próximo da unidade. Neste método de análise as mesmas variáveis citadas anteriormente foram agrupadas pelo dendrograma. Assim, é possível afirmar que os diferentes métodos de quantificação associados a diferentes métodos de análise multivariada podem ser usados para analisar dados qualitativos ordinais.

Palavras-chave: Análise de Correspondência Simples. Análise de Agrupamento. Dados categóricos. Pesquisa de satisfação.

ABSTRACT

The purpose of this final paper is to compare Ridits and Whole Numbers Ordered quantification methods for ordinal qualitative data analysis using Simple Correspondence and Grouping Analysis multivariate analysis. The work was divided in two articles, the first one conceptualizes the multivariate method of Simple Correspondence Analysis and Whole Ordered Numbers, and the second one conceptualizes the quantification Ridits and Grouping Analysis. The respective methods have been applied to a set of data coming from a research of satisfaction of the users contemplated with the project My Home My Life of the Federal Government, of the habitational set Jardim das Orquídeas, located in Toledo/PR. The criteria evaluated were the internal spatial organization of the house and the surroundings. The first article indicated that the association between variables is high because the χ^2 test was 242,87 with 44 degrees of freedom for the rooms and 245,36 with 36 degrees of freedom for the surroundings. The two-dimensional percentage map indicated that the service area, bathroom, kitchen and bus lines are the variables that generated the greatest dissatisfaction. In the second article the Ridits quantification also presented strong adequacy for the analysis of this data set, with a cofenetic correlation coefficient very close to the unit. In this method of analysis the same variables mentioned above were grouped by dendrogram. Thus, it is possible to affirm that the different quantification methods associated to different multivariate analysis methods can be used to analyze ordinal qualitative data.

Keywords: Simple Correspondence Analysis. Grouping Analysis. Categorical Data. Satisfaction Survey.

LISTA DE ILUSTRAÇÕES

2.1	Exemplo de nuvem de pontos: indivíduos classificados a partir de A variáveis, com um total de B categorias.	18
2.2	Localização do conjunto habitacional Jardim das Orquídeas.	19
2.3	Planta baixa da casa do projeto MCMV analisado.	19
2.4	Satisfação em cada cômodo.	23
2.5	Satisfação quanto ao entorno.	25
2.6	Rota da Linha 029 - Coopagro Via Parigot	26
3.1	Dendrograma da satisfação quanto à cada cômodo.	37
3.2	Dendrograma da satisfação quanto ao entorno.	39

LISTA DE TABELAS

2.1	Representação da tabela de correspondência ($i \times j$)	15
2.2	Distribuição de frequência (%) com relação à satisfação em cada cômodo .	22
2.3	Inércia associada as 4 dimensões	22
2.4	Distribuição de frequência (%) com relação à satisfação quanto ao entorno	24
2.5	Inércia associada as 4 dimensões	25
3.1	Formato de dados para Análise Ridit	30
3.2	Quantificação das categorias Ridits para as questões relacionadas à sa- tisfação em cada cômodo	36
3.3	Coefficiente de Correlação Cofenético (CCC)	36
3.4	Quantificação das categorias Ridits para as questões relacionadas à sa- tisfação quanto ao entorno	38
3.5	Coefficiente de correlação cofenético (CCC)	39

SUMÁRIO

LISTA DE ILUSTRAÇÕES	7
LISTA DE TABELAS	8
1 INTRODUÇÃO	11
2 ARTIGO 1: AVALIAÇÃO DA SATISFAÇÃO DE USUÁRIOS DO PROGRAMA “MCMV” UTILIZANDO A ANÁLISE MULTIVARI- ADA DE CORRESPONDÊNCIA SIMPLES	13
2.1 INTRODUÇÃO	13
2.2 FUNDAMENTAÇÃO TEÓRICA	14
2.2.1 QUANTIFICAÇÃO POR NÚMEROS INTEIROS ORDENADOS	14
2.2.2 ANÁLISE DE CORRESPONDÊNCIA SIMPLES	15
2.3 MATERIAL E MÉTODOS	18
2.4 RESULTADOS E DISCUSSÕES	21
2.4.1 SATISFAÇÃO EM CADA CÔMODO	21
2.4.2 SATISFAÇÃO QUANTO AO ENTORNO	24
2.5 CONCLUSÃO	27
3 ARTIGO 2: QUANTIFICAÇÃO RIDITS APLICADA NA ANÁLISE DE SATISFAÇÃO DE USUÁRIOS DO PROGRAMA MCMV	28
3.1 INTRODUÇÃO	28
3.2 FUNDAMENTAÇÃO TEÓRICA	29
3.2.1 QUANTIFICAÇÃO RIDITS	29
3.2.2 ANÁLISE DE AGRUPAMENTO	31
3.3 MATERIAL E MÉTODOS	34
3.4 RESULTADOS E DISCUSSÕES	35
3.4.1 SATISFAÇÃO EM CADA CÔMODO	35
3.4.2 SATISFAÇÃO QUANTO AO ENTORNO	38
3.5 CONCLUSÃO	40
4 CONSIDERAÇÕES FINAIS	41
REFERÊNCIAS	41
A Apêndice	
Rotina R relacionada ao Artigo 1	45

B Apêndice	
Rotina R relacionada ao Artigo 2	47
A Anexo	
Questionário aplicado aos moradores do Conjunto Habitacional Jardim das Orquídeas	49

1 INTRODUÇÃO

Tem-se usado cada vez mais a estatística como uma ferramenta para a análise e avaliação de dados, em várias áreas do conhecimento. Os métodos estatísticos foram desenvolvidos ao longo do século XX como uma mistura de ciência, tecnologia e lógica para a solução e investigação de problemas em várias áreas do conhecimento humano (STIGLER, 1986).

O relacionamento da estatística com as demais ciências é cada vez mais intenso e importante. Os métodos estatísticos são largamente empregados em diversas áreas como, por exemplo, Genética, Economia, Ciências Sociais, Engenharias, Ciência da Educação, Administração, Ciência da Computação, Medicina, Biologia, Psicologia, etc. Com o advento da computação, a estatística teve um enorme crescimento e alcançou níveis nunca antes imaginados. Atualmente, com o auxílio sempre crescente da informática, as aplicações da estatística se estendem a, praticamente, todas as áreas e subáreas do conhecimento (CORDEIRO, 2006).

Os métodos estatísticos, para analisar variáveis, estão dispostos em dois grupos: um que olha as variáveis de maneira isolada - a estatística univariada, e o outro que olha as variáveis de forma conjunta - a estatística multivariada (VICINI, 2005).

A estatística multivariada é utilizada com o propósito de simplificar ou facilitar a interpretação do fenômeno que está sendo estudado; obter índices ou variáveis alternativas que sintetizem a informação original dos dados; construir grupos de elementos amostrais que apresentem similaridade entre si, possibilitando a segmentação do conjunto de dados original; investigar as relações de dependência entre as variáveis respostas associadas ao fenômeno e outros fatores (variáveis explicativas), muitas vezes com objetivos de predição; comparar populações ou validar suposições através de testes de hipóteses (MINGOTI, 2005).

Existem várias técnicas de análise multivariada que podem ser aplicadas com diversas finalidades, entre elas a Análise de Correspondência Simples e Análise de Agrupamento. No caso dos dados serem qualitativos ordinais, um método de quantificação poderá ser aplicado, cuja finalidade é ranquear as categorias estudadas e viabilizar a análise pelas referidas técnicas. Entre os métodos de quantificação tem-se os Números inteiros ordenados ou a quantificação Ridits.

Apesar da aplicação da estatística multivariada nas mais diversas áreas, poucos estudos envolvendo a técnica são encontrados, principalmente na Engenharia Civil, em que frequentemente dados qualitativos ordinais são obtidos ao se realizar vistorias em execução de obras e análise de satisfação de usuários.

Assim, o objetivo do trabalho foi aplicar a análise multivariada no estudo de

dados qualitativos ordinais utilizando o método de quantificação Ridits, Números Inteiros Ordenados e informações de uma pesquisa de satisfação na área de Engenharia Civil.

O estudo foi dividido em dois artigos, onde no artigo 1 foi conceituado os métodos de Análise de Correspondência Simples (ACS) e Números Inteiros Ordenados, construída uma tabela de distribuição de frequência do conjunto de dados para posterior aplicação dos métodos e análise dos resultados obtidos.

No artigo 2 foi conceituado a quantificação Ridits e Análise de Agrupamento (AA) e através da mesma distribuição de frequência do artigo 1 foi determinada a tabela de pesos Ridits.

2 ARTIGO 1: AVALIAÇÃO DA SATISFAÇÃO DE USUÁRIOS DO PROGRAMA “MCMV” UTILIZANDO A ANÁLISE MULTIVARIADA DE CORRESPONDÊNCIA SIMPLES

Resumo: O programa Minha Casa Minha Vida (MCMV), criado em 2009, foi projetado com a finalidade de solucionar a defasagem de moradias no Brasil. Porém, pouco se tem discutido sobre a satisfação dos mutuários que foram beneficiados pelo programa. Nesse sentido, o presente trabalho visa analisar a satisfação dos usuários quanto aos cômodos e quanto ao entorno de um conjunto habitacional formado a partir do MCMV. Os dados qualitativos ordinais foram coletados no Conjunto Habitacional Jardim das Orquídeas localizado em Toledo/PR. A investigação foi realizada utilizando a Análise de Correspondência Simples e o método de quantificação de Números Inteiros Ordenados. O teste qui-quadrado indicou que a magnitude da associação entre as variáveis foi alta. O mapa percentual bidimensional ilustrou que o tamanho e disposição dos móveis na cozinha, o tamanho da área de serviço e o banheiro foram classificadas como péssimo. Quanto a satisfação ao entorno, as linhas de ônibus foram as que receberam as piores notas.

Palavras Chave: Análise multivariada. Análise de Correspondência Simples. Satisfação dos usuários.

2.1 INTRODUÇÃO

No Brasil, embora a habitação seja reconhecida como direito social pela Constituição da República de 1988, o país possuía em 2015, segundo a Fundação Getúlio Vargas, um déficit habitacional estimado de 7,757 milhões de domicílios.

Com intuito de solucionar o problema de moradias no Brasil, diversos programas sócio habitacionais foram desenvolvidos ao longo dos anos pelo governo brasileiro, iniciando as chamadas Habitações de Interesse Social (HIS). Nesse contexto está inserido o programa Minha Casa Minha Vida (MCMV), criado em 2009.

No entanto, o modelo de construção aplicado pelo MCMV não é necessariamente o mais eficiente, pois, segundo Hybner et al. (2014) apresenta baixa qualidade dos projetos, erros de execução e falta de manutenção. Além disso, é preciso realizar investigações com os próprios moradores a fim de determinar se a habitação realmente oferece bem-estar aos residentes.

A análise de dados referentes à satisfação dos usuários envolve a necessidade de avaliar, simultaneamente, diversas variáveis de forma hierárquica (ordinal). Neste caso, o uso de ferramentas estatísticas multivariadas para análise de variáveis de escala ordinal é a aplicação de métodos de quantificação como os Números Inteiros Ordenados (MAROCO, 2011).

Entre as diversas ferramentas multivariada para avaliar as quantificações obtidas, tem-se Análise de Correspondência Simples (ACS). A ACS permite obter uma representação gráfica multidimensional da dependência entre as linhas e/ou colunas de uma tabela de contingência de duas entradas, onde as linhas e as colunas representam categorias, modalidades, de variáveis categóricas (GUEDES, et al., 1999).

Guimarães et al. (2019) aplicaram a técnica de ACS no estudo do processo executivo de alvenaria estrutural cujas variáveis qualitativas ordinais foram ranqueadas na escala Likert. A partir do mapas percentuais, os autores detectaram atividades críticas, desencadeadas pela não execução e pela baixa frequência de execução do projeto. Steiner et al. (2008) aplicaram métodos estatísticos multivariados à engenharia de avaliações e indicaram ser viável e altamente apropriada, fomentando a aplicação na área de Engenharia Civil.

Assim, o objetivo do trabalho é avaliar o nível de satisfação dos usuários do programa MCMV situado no Conjunto Habitacional Jardim das Orquídeas localizado em Toledo/PR quanto a critérios de organização espacial interna à moradia e nos entornos.

2.2 FUNDAMENTAÇÃO TEÓRICA

2.2.1 QUANTIFICAÇÃO POR NÚMEROS INTEIROS ORDENADOS

Seja \mathbf{X} a matriz de dado de dimensão $(n \times p)$, o que significa que se dispõe de n indivíduos e p variáveis qualitativas ordinais X_1, X_2, \dots, X_p , em que a variável X_j assume um número c_j de categorias ($j = 1, \dots, p$). As quantificações (ou pontuações) são utilizadas para descrever dados ordinais e são ordenadas de acordo com as categorias, ou, seja, $v_1 < v_2 < \dots < v_{c_j}$ para X_j .

Especificamente, na quantificação de números inteiros ordenados (AGRESTI, 2010; MAROCO, 2011), atribui-se arbitrariamente números inteiros ordenados às categorias. Entre esses métodos de quantificação, tem-se:

- Escala com distâncias iguais, na qual é possível atribuir valores com distâncias igualmente espaçadas entre as categorias. Por exemplo: 1, 2, 3.
- Categoria neutra ou central, na qual é, de acordo com a ordenação das categorias,

atribuir a quantificação 0 à categoria neutra e atribuir os inteiros negativos e positivos às categorias que lhe estão à esquerda e à direita, respetivamente. Por exemplo, nas quantificações do exemplo anterior seriam -1; 0; 1.

- Escala de Likert, tipo de escala mais utilizado, na qual o número de categorias é normalmente ímpar, existindo uma categoria neutra, central, sendo as restantes categorias simétricas em relação a essa. Por exemplo, supondo, uma escala de cinco categorias, as quantificações poderiam ser 1; 2; 3; 4; 5, ou ainda, -2; -1; 0; 1; 2.

2.2.2 ANÁLISE DE CORRESPONDÊNCIA SIMPLES

A ACS pode ser definida como uma técnica adequada para dados categóricos, que permite analisar graficamente as relações existentes entre as variáveis através da redução de dimensionalidade do conjunto de dados (LATTIN, et al., 2011).

A ACS é aplicada a tabelas de contingência com o objetivo de determinar o grau de associação global entre suas linhas e as colunas, indicando como as variáveis estão relacionadas. Este método tem como base a decomposição do valor singular de uma matriz retangular (tabela de contingência adaptada) e é utilizado para representar graficamente as linhas e as colunas desta tabela como pontos em espaços vetoriais de pequena dimensão, de forma a identificar nesta dimensão a existência de padrões subjacentes (MINGOTI, 2005; GREENACRE, 2007; FERREIRA, 2011).

Na ACS, inicialmente determina-se a matriz de probabilidades ou frequências relativas, dada por $p_{ij} = \frac{n_{ij}}{n}$, indicando a relação entre o total de classificados na categoria i de A e j de B.

Tabela 2.1: Representação da tabela de correspondência ($i \times j$)

		Colunas (Variável B)					Total Linhas
		1	2	...	j		
Linha (Variável A)	1	p_{11}	p_{12}	...	p_{1j}	$p_{1.}$	
	2	p_{21}	p_{22}	...	p_{2j}	$p_{2.}$	
	
	
	
	i	p_{i1}	p_{i2}	.	p_{ij}	$p_{i.}$	
Total Colunas		$p_{.1}$	$p_{.2}$.	$p_{.j}$	1	

A partir da Tabela 2.1 é possível gerar as matrizes de correspondência, de

perfil de linhas e de perfil de colunas. A matriz de correspondência é denotada por \mathbf{P} , de dimensão $(i \times j)$ constituída das proporções $p_{ij} = \frac{n_{ij}}{n}$.

A matriz de perfil das linhas é a matriz diagonal, denotada por \mathbf{D}_r , cujos elementos diagonais são iguais a $\mathbf{r} = (p_{1.}, \dots, p_{i.})^\top$. Ou seja, o vetor \mathbf{r} armazena a proporção de ocorrências pertencentes à categoria i da variável A, em relação ao total (centroide ou probabilidade marginal) em que $p_{i.} = \frac{n_{i.}}{n}$. O vetor \mathbf{r} irá compor a matriz de perfil das linhas.

Já matriz de perfil das colunas é a matriz diagonal, denotada por \mathbf{D}_c , cujos elementos diagonais são iguais a $\mathbf{c} = (p_{.1}, \dots, p_{.j})^\top$. Ou seja, o vetor \mathbf{c} armazena a proporção de ocorrências pertencentes à categoria j da variável B, em relação ao total (centroide ou probabilidade marginal) em que $p_{.j} = \frac{n_{.j}}{n}$. O vetor \mathbf{c} irá compor a matriz de perfil das colunas.

Definido \mathbf{P} , \mathbf{r} e \mathbf{c} é possível aplicar o teste qui-quadrado, que examinará se a informação contida nas linhas da tabela são independentes ou não da informação contida nas colunas (teste de homogeneidade). O teste é expresso por:

$$\chi_{cal}^2 = n \cdot \sum_{i=1}^a \sum_{j=1}^b \frac{(p_{ij} - p_{.i}p_{.j})^2}{p_{.i}p_{.j}}. \quad (2.1)$$

Se $\chi_{cal}^2 > \chi_{(i-1)(j-1), \alpha}^2$ rejeita-se a hipótese de que não há associação entre variáveis.

Se rejeitada a hipótese de que não há associação entre variáveis, o passo seguinte será realizar a decomposição em coordenadas principais.

Segundo Mingoti (2005), a matriz $\overline{\mathbf{P}} = \mathbf{P} - \mathbf{r}\mathbf{c}^\top$ é a matriz de entrada do tipo: $\overline{p}_{ij} = p_{ij} - \frac{n_{i.}}{n} \frac{n_{.j}}{n}$, ou seja, os termos representam uma comparação da proporção observada no interior das casela da tabela, com aquela esperada sob um modelo no qual as variáveis A e B são independentes.

A partir da decomposição espectral da matriz $\overline{\mathbf{P}}$ é possível obter os autovetores e autovalores associado às coordenadas principais das linhas e das colunas. As duas primeiras coordenadas principais das linhas e das colunas são as mais representativas em termos da associação total existente entre as variáveis A e B, pois estão relacionadas aos maiores autovalores da matriz $\overline{\mathbf{P}}$.

O posto (rank) da matriz $\overline{\mathbf{P}}_{i \times j}$ é dado por $k = \min(i - 1, j - 1)$.

Assim, foi possível decompor a matriz $\overline{\mathbf{P}}_{i \times j}$ em seus autovalores e autovetores, ou seja, fazer a decomposição em seus valores singulares, obtendo:

$$\overline{\mathbf{P}}_{i \times j} = \mathbf{M}\mathbf{\Lambda}\mathbf{N}^\top \quad (2.2)$$

em que $\mathbf{M} = \mathbf{D}_r^{1/2}\mathbf{U}_{i \times k}$ é uma matriz de dimensão $i \times k$ e $\mathbf{N} = \mathbf{D}_c^{1/2}\mathbf{V}_{j \times k}$ é uma matriz de dimensão $j \times k$, sendo \mathbf{U} e \mathbf{V} matrizes ortogonais; $\mathbf{\Lambda}$ é a matriz de dimensão $k \times k$,

contendo os autovalores da matriz $\overline{\mathbf{P}}$ ordenados em ordem decrescente.

A matriz \mathbf{U} contém os autovetores da matriz $\overline{\mathbf{P}}\overline{\mathbf{P}}^\top$ e a matriz \mathbf{V} contém os autovetores da matriz $\overline{\mathbf{P}}^\top\overline{\mathbf{P}}$. Como consequência tem-se que as linhas da matriz $\overline{\mathbf{P}}$ podem ser escritas como combinações lineares das linhas da matriz \mathbf{N}^\top , e as colunas da matriz $\overline{\mathbf{P}}$ podem ser escritas como combinações lineares das colunas da matriz \mathbf{M} , definindo assim as coordenadas principais das linhas e das colunas como $\mathbf{Y}_{i \times k} = \mathbf{D}_r^{-1}\mathbf{M}_{i \times k}\mathbf{\Lambda}_{k \times k}$ e $\mathbf{Z}_{j \times k} = \mathbf{D}_c^{-1}\mathbf{N}_{j \times k}\mathbf{\Lambda}_{k \times k}$, respectivamente.

Da decomposição acima, a matriz $\overline{\mathbf{P}}$ pode ser expressa como função dos autovalores e das coordenadas principais, isto é, $\overline{\mathbf{P}} = \mathbf{P} - \mathbf{r}\mathbf{c}^\top = \sum_{i=1}^k \hat{\lambda}_i \overline{\mathbf{m}}_i \overline{\mathbf{n}}_i^\top$ em que $\overline{\mathbf{m}}_i$ denota a i -ésima coluna da matriz \mathbf{M} e $\overline{\mathbf{n}}_i$ denota a i -ésima coluna da matriz \mathbf{N} , sendo $k = \text{posto}(\overline{\mathbf{P}}) = \min(i-1, j-1)$, ou seja o número máximo de dimensões que pode ser estimado é um a menos que o menor número entre a quantia de linhas ou e colunas.

As duas primeiras coordenadas principais das linhas e das colunas são as mais representativas em termos da associação total existente entre as variáveis A e B, pois estão relacionadas aos maiores autovalores da matriz $\overline{\mathbf{P}}$.

A variação total existente nesse sistema, chamada também de inércia total, é dada por: $\sum_{i=1}^k \lambda_i^2$, em que λ_i são os autovalores não nulos da diagonal da matriz $\mathbf{\Lambda}$, $i = 1, 2, \dots, k$. Assim, a proporção de explicação da i -ésima coordenada principal em relação à inércia total é dada por:

$$\frac{\lambda_i^2}{\sum_{i=1}^k \lambda_i^2} \quad (2.3)$$

Tomando como base a decomposição do valor singular é possível representar graficamente as linhas e as colunas desta tabela como pontos em espaços vetoriais de pequena dimensão.

Representação gráfica e interpretação

Para uma linha i qualquer, passando através do centro de gravidade da nuvem de pontos, é possível decompor a inércia total da nuvem na soma da inércia paralela, projetada sobre a linha i , e da inércia perpendicular a esta linha. O primeiro eixo é a linha i para a qual a inércia paralela é máxima. O segundo eixo é, entre todas as linhas ortogonais a i , a linha para a qual a dispersão (inércia) projetada da nuvem ortogonalmente complementar a i é a maior. Assim, indo passo a passo, é possível extrair todos os eixos que formam o novo conjunto de eixos ortogonais que descrevem totalmente a nuvem de pontos. Estes eixos são chamados eixos principais da inércia e são obtidos pela Equação (2.2), considerando a ordem decrescente dos autovalores (GUEDES, et al., 1999).



Figura 2.2: Localização do conjunto habitacional Jardim das Orquídeas.

Fonte: Adaptado de Google Earth (2018).

Este conjunto habitacional é composto por 244 moradias, entregues no ano de 2011. As habitações são divididas em casas geminadas e não geminadas, cada casa totaliza $41,82 \text{ m}^2$ (BONETTI, et al., 2019).

Cada habitação possui dois quartos, um banheiro, uma cozinha integrada com sala de estar e área de serviço (Figura 2.3). Todas as casas foram entregues somente com o piso da cozinha, lavanderia e banheiro instalados, com apenas o contra-piso nos outros cômodos. Além disso, não havia divisões dos lotes, nem por muros laterais e no fundo ou portões na frente (BONETTI, et al., 2019).

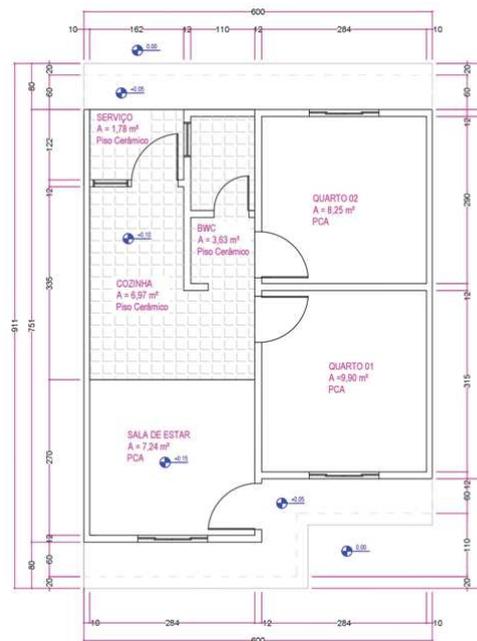


Figura 2.3: Planta baixa da casa do projeto MCMV analisado.

Fonte: Bonetti (2019).

Para a realização da entrevista da pesquisa, Bonetti, et al., (2019) elaboram um roteiro curto, contendo as seguintes informações:

- Informações gerais sobre os moradores: As informações deste tópico pretendem caracterizar as pessoas que residem no local, como a quantidade de moradores e a composição familiar de cada núcleo. Também busca-se avaliar quais as condições de habitação anterior e se houve melhoria ou piora no padrão de vida após a mudança para o conjunto habitacional.
- Satisfação quanto ao entorno: Com respeito ao local externo à moradia, investiga-se o contentamento em relação aos quesitos de infraestrutura (pavimentação, drenagem, lixeiras), bem como o acesso a comércio e ao transporte público, além de abranger a existência de locais de lazer próximo ao conjunto. Essa parte da entrevista é dada por notas em uma escala de 0 (péssimo), 1 (ruim), 2 (regular), 3 (bom) e 4 (ótimo).
- Satisfação quanto a cada cômodo: Este item, como o anterior, é questionado em forma de escala, variando de 0 (péssimo), 1 (ruim), 2 (regular), 3 (bom) e 4 (ótimo). Averigua-se, desta maneira, a satisfação quanto ao tamanho dos cômodos e se a configuração deles são adequados à colocação dos móveis necessários para cada atividade.

Como o objetivo do trabalho é discutir técnicas estatísticas multivariada para avaliar dados quantitativos ordinais, será analisado dentro desta pesquisa somente o item Satisfação quanto a cada cômodo (Quadro 1) e Satisfação quanto ao entorno (Quadro 2), que trabalham com variáveis qualitativas ordinais. Os Quadros se encontram no Anexo A.

A escolha das casas que foram visitadas foi aleatória simples em 50 habitações, sendo elas 46 casas padrões e 4 casas adaptadas para PNEs (Portadores de Necessidades Especiais), respeitando, portanto, o erro admissível de 5% que conforme Luchesa e Neto (2011), é calculada por:

$$e = z_{\frac{\alpha}{2}} \sqrt{\frac{\left(\frac{n}{N}\right) \cdot \left(1 - \frac{n}{N}\right)}{N}}. \quad (2.4)$$

em que:

e - margem de erro do valor que se estima;

$z_{\frac{\alpha}{2}}$ - variável aleatória normal padrão;

n - tamanho da amostra;

N - população total;

Após a aplicação do questionário, a primeira abordagem ao tratamento de dados qualitativos ordinais foi proceder à quantificação das suas categorias, ou seja, a

atribuição de valores numéricos a estas, permitindo assim dispor de dados quantitativos que podem em seguida ser aplicados a métodos multivariados apropriados para tratar este tipo de dados (AGRESTI, 2010). Neste trabalho, como mencionado, será abordado o método de quantificação das categorias Números Inteiros Ordenados.

As quantificações obtidas foram analisadas utilizando a técnica de análise multivariada ACS desenvolvida para conseguir a ordenação das categorias encontrada nas variáveis ordinais (MINGOTI, 2005; FERREIRA, 2011; LATTIN et al., 2011).

A análise dos dados foi realizada com o auxílio do *software R* (R Core Team 2019) e os pacotes auxiliares como *FactoMineR* (LÊ et al., 2008), *graphics* (R Core Team, 2019), cuja rotina está disposta no Apêndice A.

2.4 RESULTADOS E DISCUSSÕES

Em posse dos dados coletados mediante pesquisa com 50 moradores do conjunto habitacional Jardim das Orquídeas, obteve-se os dados quanto à satisfação em escalas de péssimo, ruim, regular, bom e ótimo, que serão apresentados e analisados a seguir.

Para analisar a satisfação dos moradores quanto à cada cômodo e quanto ao entorno do conjunto habitacional, aplicou-se as técnicas multivariadas dentro de suas respectivas etapas de verificação.

2.4.1 SATISFAÇÃO EM CADA CÔMODO

Na etapa composta por questões referentes a satisfação quanto aos cômodos verifica-se pela Tabela 2.2 a distribuição de frequência, onde se encontra destacado em vermelho o item que diz respeito ao tamanho da área de serviço, com 90% de insatisfação, o qual o maior percentual é na categoria péssimo, avaliando ainda que apenas 10% classificaram o tamanho como bom, deixando a categoria ótima sem ser citada.

Pode-se notar que há uma discrepância entre a categoria péssimo e ótimo. Os itens com maiores índices de descontentamento são em áreas de comum acesso, como a cozinha, o banheiro e a área de serviço.

De forma geral, 40% consideram a cozinha o pior cômodo da habitação e 42% o banheiro. No tocante ao melhor cômodo, 4% preferem os quartos e 8% a sala.

Tabela 2.2: Distribuição de frequência (%) com relação à satisfação em cada cômodo

Questão	Péssimo	Ruim	Regular	Bom	Ótimo
1.1 Tamanho da cozinha	40	36	4	20	0
1.2 Disposição dos móveis da cozinha	36	34	12	16	2
1.3 Tamanho da área de serviço	62	16	12	10	0
1.4 Disposição dos móveis da área de serviço	12	32	16	34	6
1.5 Tamanho da sala	12	36	14	30	8
1.6 Tamanho dormitório	6	22	12	56	4
1.7 Disposição dos móveis nos dormitórios	6	24	6	60	4
1.8 Quantidade de Tomadas e interruptores	12	8	14	52	14
1.9 Localização de tomadas e interruptores	12	8	14	52	14
1.10 Quantidade de janelas	2	4	14	64	16
1.11 Localização das janelas	4	4	14	62	16
1.12 Banheiro	42	28	12	12	6

Fonte: Bonetti (2019).

Para avaliar se havia independência entre as variáveis avaliadas foi aplicado o teste qui-quadrado. Observou-se forte associação entre as variáveis ao nível de 5% de significância, sendo que o valor do qui-quadrado, com 44 graus de liberdade, foi de $\chi_{cal}^2 = 242,87 > \chi_{tab}^2 = 29,79$, validando a aplicação da ACS.

A partir da matriz de correspondência, de perfil de linhas e de coluna, calculou-se as coordenadas principais das linhas e colunas e a variação total existente entre as variáveis (inércia total). Os resultados encontram-se na Tabela 2.3.

Tabela 2.3: Inércia associada as 4 dimensões

Dimensão	Inércia	Varição explicada (%)	Varição acumulada explicada (%)
1	0,300	73,90	73,90
2	0,059	14,48	88,38
3	0,037	8,99	97,37
4	0,011	2,63	100,00

Fonte: As autoras (2019).

Verifica-se que a dimensão 1, ou seja, a primeira coordenada, representa 73,90% da inércia, e a segunda coordenada (dimensão 2) 14,48%, ou seja, as duas primeiras di-

mensões conseguem descrever a associação entre as variáveis com uma variação acumulada de 88,38%. Assim, optou-se por trabalhar com uma representação do “mapa percentual” bidimensional.

O “mapa percentual” bidimensional, associando a satisfação quanto a cada cômodo está apresentado na Figura 2.4. Verifica-se que os itens 1.8, 1.9, 1.10 e 1.11 estão associados entre as categorias ótimo e bom. Já os itens 1.1, 1.2, 1.3 e 1.12 se encontram na categoria péssimo.

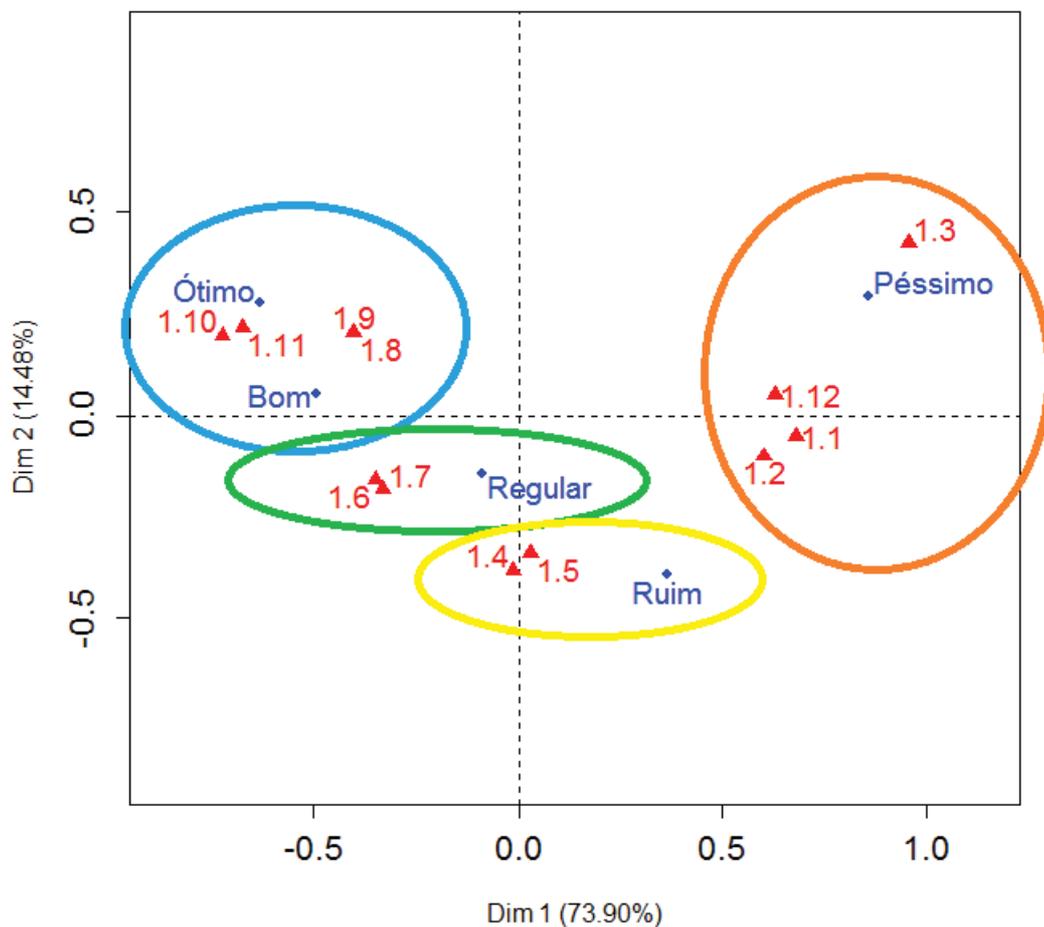


Figura 2.4: Satisfação em cada cômodo.

Fonte: As autoras (2019).

Vale destacar que o item 1.12, que se refere ao banheiro, apresentou um dos maiores índices de insatisfação. O motivo que pode estar associado a esse descontentamento é a localização desse cômodo (de frente para a cozinha). Ademais, o lavatório se encontra do lado externo causando certo desconforto para os moradores.

2.4.2 SATISFAÇÃO QUANTO AO ENTORNO

Como na etapa anterior, nota-se que à uma discrepância entre as categorias péssimo/ruim e ótimo/bom, porém os percentuais mais altos estão concentrados nas categorias bom e ótimo, tendo uma porcentagem baixa na categoria péssimo.

Verifica-se na Tabela 2.4 que o item que representa o maior percentual de descontentamento é em relação a linha de ônibus, distribuídos em péssimo, ruim e regular (destacado em vermelho). Também encontram-se destacados em azul os itens que representam mais de 90% de satisfação entre ótimo e bom. Salientando que o item com mais satisfação é o 2.4 com 52% que diz respeito a coleta de lixo dentro do conjunto habitacional.

Tabela 2.4: Distribuição de frequência (%) com relação à satisfação quanto ao entorno

Questão	Péssimo	Ruim	Regular	Bom	Ótimo
2.1 Rua de acesso ao condomínio	0	2	6	78	14
2.2 Drenagem das ruas	4	2	4	72	18
2.3 Iluminação pública da sua rua	2	6	12	72	8
2.4 Coleta de lixo	0	0	6	42	52
2.5 Coleta seletiva de lixo	6	28	32	30	4
2.6 Distância entre o residencial e o colégio	6	28	32	30	4
2.7 Comércio	0	2	16	46	36
2.8 Pontos de ônibus	0	2	20	38	40
2.9 Praças/área de lazer	0	2	6	78	14
2.10 Linhas de ônibus	18	38	22	20	2

Fonte: Bonetti (2019).

Com o teste do qui-quadrado observou-se forte associação entre as variações ao nível de 5% de significância, sendo que o valor de qui-quadrado tabelado, com 36 graus de liberdade, foi de $\chi_{cal}^2 = 245,36 > \chi_{tab}^2 = 23,27$, validando a aplicação da ACS.

A variação total existente entre as variáveis (inércia total) encontra-se na Tabela 2.5. Verifica-se que as duas primeiras dimensões conseguem descrever a associação entre as variáveis com uma variação acumulada de explicação de 92,61%, sendo a primeira coordenada com 71,52% da inércia, e a segunda coordenada 21,09%.

Tabela 2.5: Inércia associada as 4 dimensões

Dimensão	Inércia	Varição explicada (%)	Varição acumulada explicada (%)
1	0,398	71,52	10,52
2	0,117	21,09	92,61
3	0,039	7,00	99,61
4	0,002	0,39	100,00

Fonte: As autoras (2019).

O “mapa percentual” bidimensional, da satisfação quanto ao entorno encontra-se na Figura 2.5. Verifica-se que os itens 2.4, 2.7 e 2.8 estão associados a categoria ótimo. Já os itens 2.5, 2.6 e 2.9 estão associados tanto a categoria péssimo, quanto a ruim e a regular.

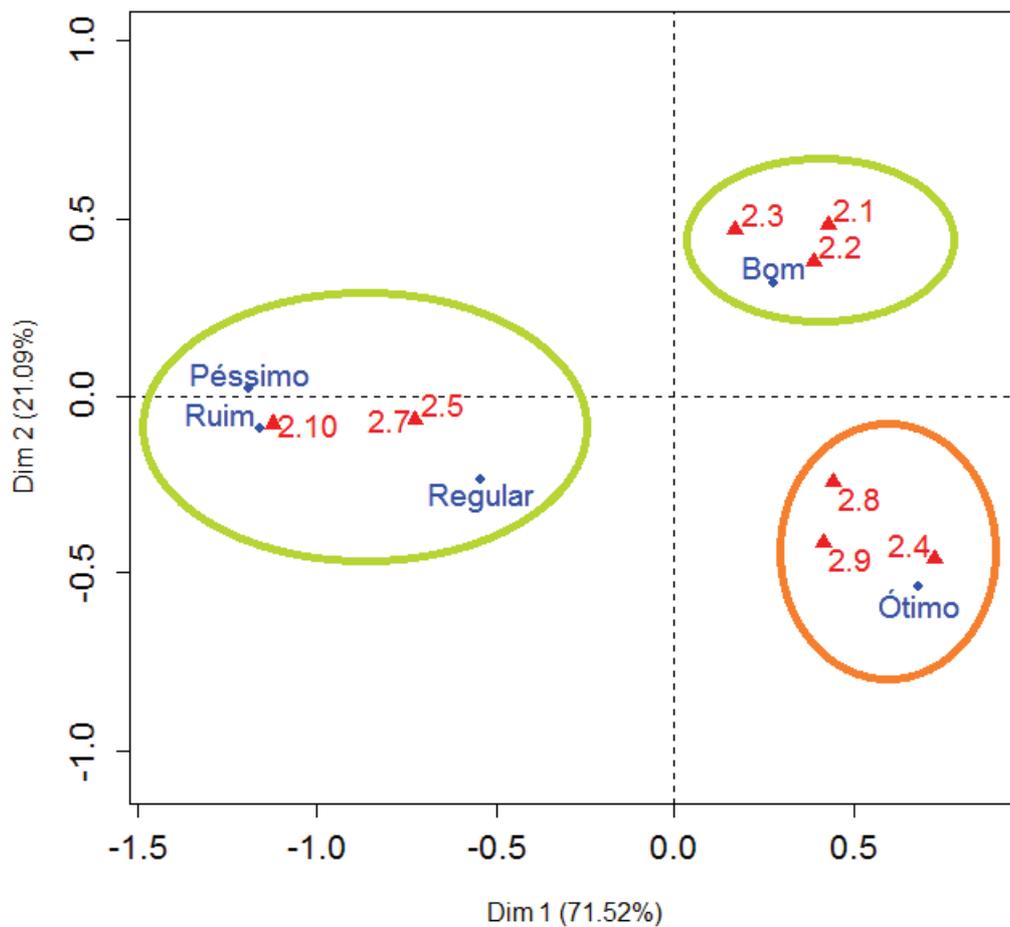


Figura 2.5: Satisfação quanto ao entorno.

Fonte: As autoras (2019).

Os itens 2.5, 2.7 e 2.10 que dizem respeito a coleta de lixo seletiva, comércio e linhas de ônibus, respectivamente, são os itens que a satisfação está dividida entre as categorias péssimo, ruim e regular.

Analisando o item 2.5, apesar da coleta de lixo do conjunto habitacional seja o item que mais se aproxima da categoria ótimo, ainda deixa a desejar na coleta seletiva. Este tipo de coleta é de extrema importância para o desenvolvimento sustentável do planeta, pois abrange vários fatores que contribuem para o desenvolvimento social e econômico.

Já em relação ao item 2.7, embora haja três mercados nas proximidades, pelas entrevistas foi possível perceber que em caso de compras de alimentos e materiais de higiene, muitos moradores preferem deslocar-se até os supermercados maiores, os quais encontram-se a aproximadamente 2,5km do Loteamento Jardim das Orquídeas. Isso ocorre devido a busca de preços mais acessíveis das mercadorias.

Como já supramencionado o item que apresenta o maior percentual de descontentamento é o que diz respeito a linha de ônibus (2.10). Pesquisadas as vias de ônibus, verifica-se que a Empresa Sorriso de Toledo, responsável pela circulação de transporte urbano, dispõe de apenas uma linha de ônibus que atende o Jardim das Orquídeas, a 029 – Coopagro Via Parigot, que sai do terminal e o último ponto encontra-se no ponto próximo ao Centro Municipal de Educação Infantil (CMEI) Rosane Fontes cuja rota está indicada na Figura 2.6.

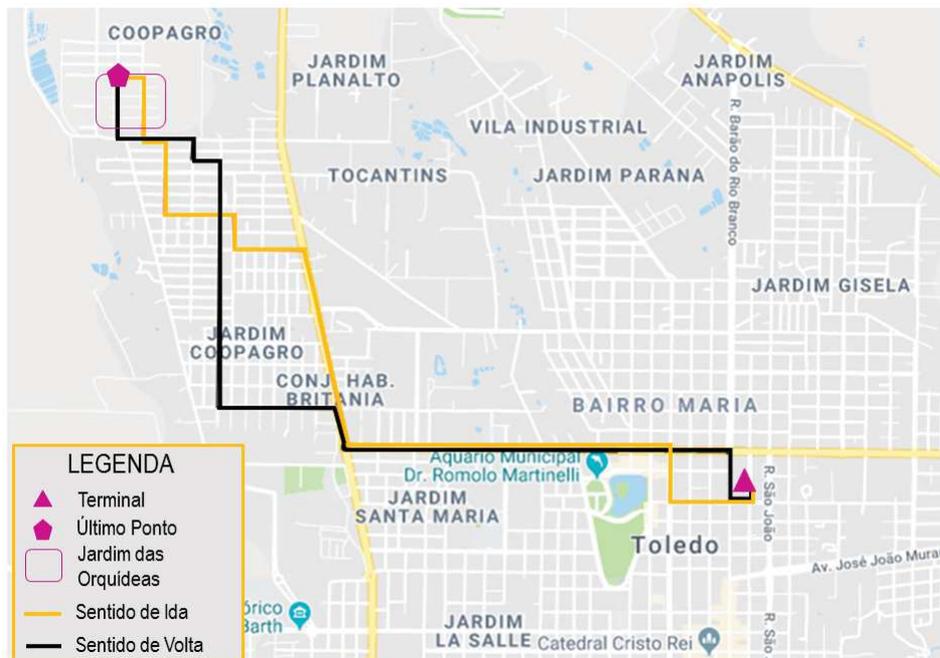


Figura 2.6: Rota da Linha 029 - Coopagro Via Parigot

Fonte: Adaptado de Sorriso de Toledo (2019).

Ainda segundo a Viação Sorriso de Toledo (2019), em dias úteis a frequência

dos ônibus são a cada trinta e quatro minutos. Nos sábados, a cada meia hora e aos domingos e feriados, os ônibus saem em intervalos de uma hora e cinco minutos entre cada um, sendo a única reclamação mais frequente dos moradores.

O item 2.9, que refere-se à áreas de lazer apresentou ótima avaliação. Verificou-se que existem dois locais com propósito de entretenimento: um parquinho e uma arena multiuso. A implantação da arena multiuso foi uma iniciativa da prefeitura em janeiro de 2019, com o objetivo de estimular a prática de esporte e aumentar a qualidade da saúde das pessoas (TOLEDO, 2019). Essa arena é composta de uma pista de skate, campo de futebol sintético e quadra poliesportiva.

Já no item 2.6, que não foi incluído na análise de correspondência, apresentou um valor percentual acumulativo de 60% entre os critérios “Ruim” e “Regular”. Esse resultado pode estar associado ao fato de que o CMEI que atende crianças de zero a cinco anos mais próximo do conjunto habitacional situa-se a 1km. Já o colégio fundamental e médio mais próximo situa-se a aproximadamente 1,7km, o que é relativamente longe para uma criança deslocar-se.

2.5 CONCLUSÃO

Após a aplicação do método de ACS, pode-se observar que a quantificação por números inteiros ordenados foi significativa, obtendo através do teste qui-quadrado (teste de homogeneidade) os valores $\chi^2_{cal} = 242,87$ e $\chi^2_{cal} = 245,36$, respectivamente.

No que diz respeito a satisfação em cada cômodo, após a aplicação dos métodos multivariados, percebe-se que das doze variáveis analisadas, oito delas foram classificadas como péssimo, ruim ou regular. Essas categorias estão relacionadas as áreas de comum acesso, sendo o tamanho da cozinha e da área de serviço e a localização do banheiro as áreas mais criticadas do projeto.

Quanto ao entorno, das nove categorias analisadas apenas três mostraram alto nível de insatisfação: a coleta de lixo seletiva, o comércio e as linhas de ônibus.

3 ARTIGO 2: QUANTIFICAÇÃO RIDITS APLICADA NA ANÁLISE DE SATISFAÇÃO DE USUÁRIOS DO PROGRAMA MCMV

Resumo: Diante da dificuldade e complexidade de se trabalhar com dados qualitativos ordinais, na maioria das vezes as investigações são realizadas apenas por tabelas de porcentagem e análise gráfica. Várias pesquisas tem apresentado métodos de quantificação para melhorar a análise dessas informações, entre eles o de quantificação Ridits. Embora as diversificadas áreas de aplicação desses métodos, poucas aplicações em engenharia civil são encontrados na literatura. Nesse cenário o presente trabalho teve por objetivo analisar a satisfação dos usuários do programa Minha Casa Minha Vida (MCMV) do governo Federal, do Conjunto Habitacional Jardim das Orquídeas localizado em Toledo/PR, no que diz respeito a organização espacial interna à moradia e nos entornos utilizando a técnica multivariada de Análise de Agrupamento e o método de quantificação Ridits. Os resultados mostraram que a quantificação Ridits apresentou forte adequação para a análise deste conjunto de dados, com um coeficiente de correlação cofenético igual a 0,875 para o agrupamento das variáveis relacionadas a satisfação quanto aos cômodos da casa e de 0,928 em relação ao entorno do conjunto habitacional. O dendrograma indicou que a área de serviço, banheiro, cozinha e as linhas de ônibus são as variáveis que geraram maior insatisfação dos mutuários.

Palavras chaves: Pontuação Ridits. Análise de Agrupamento. Satisfação dos usuários.

3.1 INTRODUÇÃO

A técnica multivariada é uma área instigante da estatística por considerar, simultaneamente, as diversas variáveis que caracterizam um fenômeno em estudo, sendo a análise, a descrição e a inferência realizada valendo-se da estrutura de correlação entre essas variáveis (FAGUNDES, 2017).

Dentro da análise multivariada é possível classificar as variáveis entre qualitativas e quantitativas. Entre as diversas características da variável tem-se a qualitativa ordinal. As escalas ordinais são utilizadas numa grande variedade de domínios, como as Ciências Sociais, a Medicina, a Economia, a Psicologia ou o Marketing (MINGOTI, 2005).

Apesar das diversificadas áreas de aplicação, poucos trabalhos utilizando técnicas multivariadas para dados qualitativos ordinais são encontrados na literatura relacionados a área da engenharia civil, onde a escala ordinal é utilizada no controle de qualidade, análise de processo construtivo, perícias, etc (AGRESTI, 2010).

O primeiro passo para avaliar dados qualitativos ordinais é realizar a quantificação das informações. Entre os métodos de quantificação têm-se o de pontuação Ridits que pode ser avaliado pela Análise de Agrupamento (AA).

Com base no que foi exposto, este artigo pretende estender estudos relacionados às técnicas de análise multivariada aplicadas a dados qualitativos ordinais, desenvolvendo estudos teóricos sobre as metodologias Ridits e Análise de Agrupamento e posterior análise da satisfação do usuários do Conjunto Habitacional Jardim das Orquídeas localizado em Toledo/PR.

3.2 FUNDAMENTAÇÃO TEÓRICA

3.2.1 QUANTIFICAÇÃO RIDITS

O termo Ridits é um acrônimo de “em relação a uma distribuição identificada” das proporções ou frequências mais as várias categorias ordenadas de alguma população de referência escolhida. O termo “relativa” indica o sentido em que as proporções ou frequências de ocorrência de observações nas diferentes categorias ordenadas de uma população de interesse, são comparados com as proporções ou frequências nas correspondentes categorias ordenadas (UWAWUNKONYE; ANAENE, 2013).

A análise Ridits desenvolvida por Bross (1958) não tenta quantificar as categorias, mas sim trabalhar com sua ordem natural.

Praticamente a única suposição feita na análise Ridits é que as categorias discretas representam intervalos de uma distribuição contínua subjacente, mas não observável. Sem suposição sobre a normalidade ou qualquer outra forma de distribuição.

Para utilizar a análise Ridits deve-se selecionar um grupo de dados, ordená-los (Tabela 3.1) e escolher uma das categorias como um padrão ou população referência para os outros grupos, muitas vezes referidos como grupo de comparação. Tendo selecionado um grupo de referência ou população, calcula-se então um “Ridits” ou pontuação para cada uma das suas categorias (WU, 2007).

Tabela 3.1: Formato de dados para Análise Ridit

Categoria ordenada de critério variável (C_i)	Y (Referência, f_{iy})	X (Comparação, f_{ix})	Total (t_{ixy})
C_1	f_{1y}	f_{1x}	$t_{1xy}(= f_{1x} + f_{1y})$
C_2	f_{2y}	f_{2x}	$t_{2xy}(= f_{2x} + f_{2y})$
.	.	.	.
.	.	.	.
.	.	.	.
C_k	f_{ky}	f_{kx}	$t_{kxy}(= f_{kx} + f_{ky})$
Total	n_y	n_x	$n_x + n_y$

No método de quantificação Ridits (BROSS, 1958), a pontuação ou ridit para uma determinada categoria, é calculada como a frequência cumulativa de todas as categorias inferiores da categoria de interesse mais metade da frequência para essa categoria, tudo dividido pela frequência total ou o tamanho da população do grupo de referência, como demonstrado pela Equação 3.1 (HURWITZ, 2016).

$$r_{iy} = \begin{cases} \frac{\frac{f_{iy}}{2}}{n_y}, & \text{se } k = 1, \\ \frac{f_{iy} + \dots + f_{k-1} + \frac{1}{2}f_{ky}}{n_y}, & \text{se } k > 1. \end{cases} \quad (3.1)$$

Para calcular o Ridits médio de qualquer outro grupo de comparação, basta multiplicar o valor de cada categoria do grupo de comparação pelo Ridits encontrado na mesma categoria do grupo de referência. Em seguida, somar os produtos e dividir pelo número total da categoria de comparação, obtendo assim o Ridits médio \bar{r}_x , como definido na Equação 3.2 (UWAWUNKONYE; ANAENE, 2013).

$$\bar{r}_x = \frac{\left(\sum_{i=1}^k f_{ix} r_{iy} \right)}{n_x}. \quad (3.2)$$

O Ridits significativo para a população de referência (Y) é pelas definições das Equações 3.1-3.2 sempre igual a 0,50.

Portanto, se \bar{r}_x é o Ridits médio de uma população de comparação, a partir do qual uma amostra aleatória de tamanho n_x foi usado para se obter \bar{r}_x , em seguida, pode se usar um teste de hipótese nula.

$$\begin{cases} H_0 : \bar{r}_x \geq 0,50, \\ H_i : \bar{r}_x < 0,50. \end{cases} \quad (3.3)$$

Foi demonstrado por Bross (1958) que se o tamanho da amostra é suficientemente grande para n_x , o Ridits média da amostra é normalmente distribuído com média

\bar{r}_x e variância.

$$Var(\bar{r}_x) = \frac{1}{12n_x}. \quad (3.4)$$

Daí a hipótese nula da Equação 3.3, pode ser testada usando a estatística de teste.

$$\chi^2 = \frac{(\bar{r}_x - 0,50)^2}{Var(\bar{r}_x)} = 12n_x(\bar{r}_x - 0,50)^2. \quad (3.5)$$

Que tem aproximadamente uma distribuição qui-quadrado com 1 grau de liberdade que é suficientemente grande para n_x . H_0 é rejeitado com α nível de significância se $\chi^2 \geq \chi_{1-\alpha;1}^2$. Caso contrário H_0 não é rejeitado.

3.2.2 ANÁLISE DE AGRUPAMENTO

Segundo Mingoti (2005), a Análise de Agrupamento (conglomerado ou Análise de *clusters*) tem como objetivo dividir os elementos da amostra, ou população, em grupos de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis características que neles foram medidas. Os agrupamentos resultantes devem então exibir elevada homogeneidade interna (dentro dos agrupamentos) e elevada heterogeneidade externa (entre agrupamentos).

Esse processo de agrupamento começa levando as medidas das p variáveis em cada um dos n objetos. Assim, a matriz $n \times p$ de dados é transformada em uma matriz $n \times n$ de semelhança ou, alternativamente, por medidas de distância onde são computadas as semelhanças ou distâncias entre pares de objetos pelas p variáveis. Logo, um algoritmo é selecionado com a finalidade de definir as regras que concernem ao agrupamento dos objetos em subgrupos com base nas semelhanças (VALLI, 2002).

A maioria dos métodos utilizados em análise de *cluster* requer uma medida de similaridade ou dissimilaridade entre os elementos que querem ser agrupados, a maioria das vezes é expressa como uma função distância para dados que têm propriedades métricas.

Quando se manipula variáveis quantitativas é comum o uso de medidas de dissimilaridade e, logo, quanto menor os seus valores, mais similares serão os elementos que estão sendo comparados, podendo-se trabalhar com variáveis padronizadas quanto houver diferença na escala de medidas. Segundo Khattree e Naik (2000) e Cruz e Carneiro (2006), dentre as medidas de dissimilaridades conhecidas, a distância Euclidiana e a distância de Mahalanobis estão entre as medidas que mais se destacam, devido a sua maior utilização.

Neste estudo, para medir a distância entre os resultados utilizou-se a distância Euclidiana. Segundo Prado (2015), a distância Euclidiana entre dois casos i e j é a raiz quadrada do somatório dos quadrados das diferenças entre os valores de i e j para todas

as variáveis ($k = 1, 2, \dots, p$), isto é:

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ik} - x_{jk})^2}, \quad (3.6)$$

$$d_{ij} = \sqrt{\sum_{i=l}^p (x_{ik} - x_{jk})^2}. \quad (3.7)$$

em que:

x_{ij} representa o valor observado da j - ésima variável no i - ésimo indivíduo.

Definida a medida de similaridade, a próxima etapa é agrupar as variáveis analisadas. Nesse trabalho optou-se pela técnica hierárquica aglomerativa, em que cada elemento do conjunto de dados observado é considerado como sendo um conglomerado isolado. A cada passo do algoritmo, os elementos amostrais vão sendo agrupados, formando novos conglomerados até o momento no qual todos os elementos considerados estão em um único grupo (JOHNSON; WICHERN, 2002).

Existem vários métodos de agrupamento hierárquicos, entre eles:

- MÉTODO DA LIGAÇÃO SIMPLES (*SINGLE LINKAGE*)

Neste método, a similaridade entre dois clusters é definida pela menor distância de qualquer ponto do primeiro cluster para qualquer ponto do segundo cluster, assim a distância entre uma observação e um grupo formado pelas observações u e v é dada por:

$$d_{(uv)w} = \min(d_{uw}, d_{vw}). \quad (3.8)$$

em que:

$D_{(uv)w}$ é a distância entre o grupo uv e a observação k e

$\min(d_{uw}, d_{vw})$ é a menor distância entre os grupos de observações uw e vw (JUNIOR; THIEBAUT, 1999).

Segundo Anderberg (2014), algumas das características desse método são: em geral grupos muito próximos pode não ser identificados; permite detectar grupos de formas não-elípticas; apresenta pouca tolerância a ruído, pois tem a tendência a incorporar os ruídos em um grupo já existente; apresenta bons resultados tanto para distância euclidianas quanto para outras distâncias; e tem tendência a formar longas cadeias (encadeamento).

- MÉTODO DA LIGAÇÃO COMPLETA (*COMPLETE LINKAGE*)

Neste método, a similaridade entre dois conglomerados é exatamente a oposta do Método da Ligação Simples. O algoritmo é iniciado encontrando a menor distância

$D = d_{ik}$ e agrupando os objetos correspondentes (U e V) para formar o grupo (UV). As distâncias entre (UV) e qualquer outro grupo são calculadas por (JOHNSON; WICHERN, 2002):

$$d_{(uv)w} = \max(d_{uw}, d_{vw}). \quad (3.9)$$

Para Quintal (2006), este método tem tendência para encontrar clusters compactos compostos de objetos muito semelhantes entre si. Quando um objeto é acrescentado a um grupo, a distância do novo grupo aos restantes aumenta ou então fica inalterada. O método de ligação completa tende a formar grupos pequenos que depois serão aglutinados para formar grupos maiores.

- MÉTODO DA MÉDIA DAS DISTÂNCIAS (*AVERAGE LINKAGE*)

Este método trata as distâncias entre dois aglomerados como a média das distâncias entre todos os pares de elementos que podem ser formados com os elementos dos dois conglomerados que estão sendo comparados (MINGOTI, 2005). Assim, se o conglomerado D_{uv} tem n_{uv} elementos e o conglomerado D_w tem n_w elementos, a distância entre eles será calculada por:

$$\bar{d}_{(uv)w} = \frac{\sum_{i \in C_{uv}} \sum_{k \in C_w} d_{ik}}{n_{uv}n_w}. \quad (3.10)$$

Algumas características do método são: menor sensibilidade a ruídos que os métodos citados anteriormente; apresenta bons resultados tanto para distâncias Euclidianas quanto para outras distâncias; tem tendência a formar grupos com números de elementos similares (ANDERBERG, 2014; CRUZ; CARNEIRO, 2006).

Os resultados de um procedimento hierárquico podem ser representados graficamente através do dendrograma no qual o objeto de estudo é colocado em um eixo e o outro eixo representa os passos do procedimento. Começando com cada objeto representado como um agrupamento separado, o dendrograma mostra graficamente como os agrupamentos são combinados em cada passo dos procedimentos até que todos estejam contidos em um único agrupamento.

Para validar os agrupamentos hierárquicos pode-se utilizar uma medida chamada de Coeficiente de Correlação Cofenético (CCC) para avaliar o grau de ajuste do agrupamento. Esse coeficiente pode ser utilizado para avaliar a consistência do padrão de agrupamento de métodos de agrupamentos hierárquicos, sendo que valores próximos à unidade indicam melhor representação (BARROSO; ARTES, 2003; CRUZ; CARNEIRO, 2006). Porém, segundo Rohlf (1970), se CCC igual ou acima de 0,70 o agrupamento é considerado adequado.

Segundo Bussab et al. (1990) o CCC é calculado por:

$$r_{cof} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - c)(s_{ij} - s)}{\sqrt{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (c_{ij} - c)^2} \sqrt{\sum_{j=i+1}^n (s_{ij} - s)}}. \quad (3.11)$$

em que:

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

$$c = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_{ij} \quad \text{e} \quad s = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n s_{ij}.$$

Após a construção do dendrograma e realizado os agrupamentos, é necessário decidir o ponto de corte, ou seja, definir o número de grupos k que mais representa os dados da amostra, podendo interferir diretamente na qualidade do agrupamento. Nos métodos hierárquicos o pesquisador pode escolher o número de grupos k , escolhendo assim um ponto de corte na escala de distância e cortando os ramos do dendrograma.

3.3 MATERIAL E MÉTODOS

Os dados analisados no presente trabalho são referentes a uma pesquisa de satisfação realizada com famílias que foram contempladas com o projeto MCMV do governo Federal, no Conjunto Habitacional Jardim das Orquídeas, situado no Jardim Fachini – Coopagro, na cidade de Toledo/PR.

O conjunto habitacional é composto por 244 moradias, entregues no ano de 2011, cada casa contém $41,82 m^2$, divididos em dois quartos, um banheiro, uma cozinha integrada com sala de estar e área de serviço (BONETTI, et al., 2019).

Foi realizada uma entrevista com os moradores, e para tal Bonetti, et al. (2019) elaboram um roteiro curto para a realização da mesma.

Como o objetivo do trabalho é discutir técnicas estatísticas multivariada para avaliar dados quantitativos ordinais, será analisado dentro desta pesquisa somente o item Satisfação quanto a cada cômodo (Quadro 1) e Satisfação quanto ao entorno (Quadro 2), que trabalham com variáveis qualitativas ordinais, os quais estão apresentados no anexo A.

A pesquisa foi realizada com 50 habitações, sendo elas 46 casas padrão e 4 casas adaptadas para PNEs (Portadores de Necessidades Especiais), respeitando, portanto, o erro admissível de 5%. A escolha das casas a serem visitadas foi aleatória simples (BONETTI, et al., 2019).

Fórmula do cálculo de erro de estimativa de uma proporção e intervalo de confiança, que conforme Luchesa e Neto (2011), é calculada por:

$$e = z_{\frac{\alpha}{2}} \sqrt{\frac{\left(\frac{n}{N}\right) \cdot \left(1 - \frac{n}{N}\right)}{N}}. \quad (3.12)$$

em que:

e - margem de erro tolerável;

$z_{\frac{\alpha}{2}}$ - variável aleatória normal padrão;

n - tamanho da amostra;

N - população total;

Após a aplicação do questionário, a primeira abordagem ao tratamento de dados qualitativos ordinais é proceder à quantificação das suas categorias, ou seja, a atribuição de valores numéricos a estas, permitindo assim dispor de dados quantitativos que podem em seguida ser aplicados a métodos multivariados apropriados para tratar este tipo de dados.

As quantificações obtidas foram analisadas utilizando a técnica de análise multivariada Ridits e AA utilizando o método de Média das Distâncias.

A análise dos dados foi realizada com o auxílio do *software R* (R Core Team 2019) e os pacotes auxiliares como o Graphics (R Core Team 2019), cuja rotina está disposta no Apêndice B.

3.4 RESULTADOS E DISCUSSÕES

3.4.1 SATISFAÇÃO EM CADA CÔMODO

A partir da coleta de dados através de entrevista sobre a satisfação em relação a cada cômodo pela ótica dos moradores, montou-se a tabela de distribuição de frequências (%) com os resultados obtidos.

Em seguida, foi determinada a tabela de quantificação Ridits e posterior AA. A pontuação Ridits foi calculada como a frequência cumulativa de todas as categorias inferiores da categoria de interesse mais metade da frequência para essa categoria. Os resultados da quantificação encontram-se na Tabela 3.2.

Tabela 3.2: Quantificação das categorias Ridits para as questões relacionadas à satisfação em cada cômodo

Questão	Péssimo	Ruim	Regular	Bom	Ótimo
1.1	0,2040816	0,5816327	0,7755102	0,8979592	1,0000000
1.2	0,1836735	0,5306122	0,7551020	0,8979592	0,9897959
1.3	0,3061224	0,6938776	0,8367347	0,9489796	1,0000000
1.4	0,0508475	0,2372881	0,5254237	0,8135593	0,9745763
1.5	0,0612245	0,2959184	0,5408163	0,7653061	0,9591837
1.6	0,0306122	0,1632653	0,3265306	0,6734694	0,9795918
1.7	0,0306122	0,1836735	0,3367346	0,6632653	0,9795918
1.8	0,0612245	0,1632653	0,2755102	0,6020408	0,9285714
1.9	0,0612245	0,1632653	0,2755102	0,6020408	0,9285714
1.10	0,0102041	0,0408163	0,1326531	0,5204082	0,9183674
1.11	0,0204082	0,0612245	0,1530612	0,5306122	0,9183674
1.12	0,2040816	0,5510204	0,7551020	0,8775510	0,9693878

Fonte: As autoras (2019).

A Análise de Agrupamento iniciou por meio da determinação da matriz de correlação e em seguida foi aplicada a medida de dissimilaridade. No caso em estudo, foram realizados três avaliações com os métodos de análise de agrupamento: Ligação Simples, Ligação Completa e Média das Distâncias. Em cada um dos métodos foi obtido seu respectivo CCC, que estão expostos na Tabela 3.3.

Tabela 3.3: Coeficiente de Correlação Cofenético (CCC)

Método	CCC
Ligação Simples	0,864
Ligação Completa	0,874
Média das Distâncias	0,875

Fonte: As autoras (2019).

Observando a Tabela 3.3 e segundo Rohlf (1970), percebe-se que em todos os métodos o CCC é adequado e apresentaram valores bem semelhantes. Porém, para a presente análise, foi escolhido o método com o maior valor de CCC, obtido no método de Média das Distâncias cujo valor foi de 0,875 apresentando uma ótima qualidade de agrupamento das variáveis referentes a satisfação em cada cômodo.

Após a escolha do método foi construído o dendrograma com base na matriz de dissimilaridade que está apresentado na Figura 3.1.

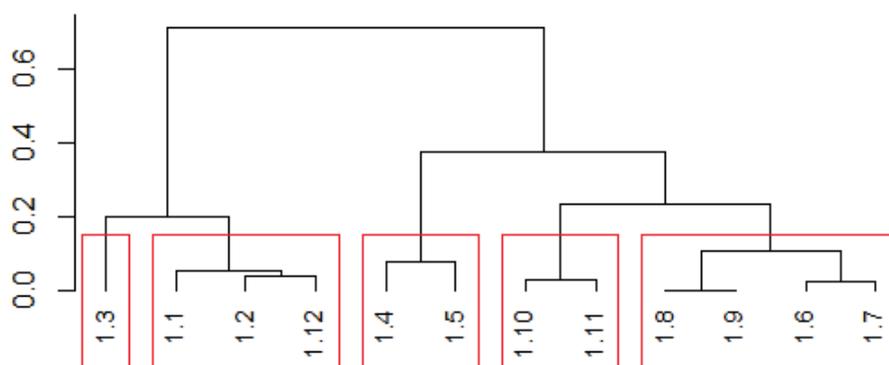


Figura 3.1: Dendrograma da satisfação quanto à cada cômodo.

Fonte: As autoras (2019).

Considerando as 5 classificações qualitativas ordinais (Péssimo, Ruim, Regular, Bom e Ótimo), formou-se 5 grupos distintos. O grupo classificado como “Péssimo” é formado pela variável 1.3 que diz respeito ao tamanho da área de serviço. Esse espaço apresentou insatisfação por apresentar um tamanho relativamente pequeno, comportando apenas a máquina de lavar e tanque, não tendo espaço para os outros utensílios, como armários e tábua de passar. Ademais, esse espaço encontra-se do lado externo da residência, mesmo havendo cobertura, não há paredes o que acarreta em não utilização dessa área em tempos chuvosos ou muito frio.

O grupo classificado como “Ruim” é formado pelos itens: 1.1, 1.2 e 1.12. Esses itens são referentes ao tamanho e disposição dos móveis na cozinha e ao banheiro. Note que esses espaços são de uso comum e que a cozinha, em geral, é utilizada simultaneamente por todos os moradores da casa e o tamanho e disposição dos móveis interfere no fluxo de pessoas. No que se refere ao banheiro, o principal motivo do descontentamento é em relação a sua localização (de frente para a cozinha), não deixando de citar que o lavabo se encontra do lado externo do mesmo, causando certo desconforto aos usuários.

O grupo classificado como “Regular” é formado pelos itens 1.4 referente a disposição dos móveis da área de serviço e 1.5 referente ao tamanho da sala. Nota-se novamente o espaço de uso coletivo apresentou conceito baixo na avaliação. A principal reclamação quanto a sala é o fato de, por ser integrada à cozinha – cujo espaço, de acordo com moradores, é insuficiente para o armazenamento dos móveis – acaba sendo utilizada para acomodar eletrodomésticos e mobiliários daquele cômodo, restando pouco espaço para móveis como televisão e sofá.

O grupo que foi formado pelo conceito “Bom”, é composto pelas variáveis: 1.10 e 1.11 que diz respeito a quantidade e localização das janelas. O último grupo classificado como “Ótimo” contém os itens: 1.8, 1.9, 1.6 e 1.7 que se refere a quantidade e localização de tomadas e interruptores, e tamanho e disposição dos móveis do quarto,

respectivamente. Observa-se que o tamanho do quarto, se encontra nesta categoria pois é um espaço que geralmente é utilizado por um número menor de pessoas.

3.4.2 SATISFAÇÃO QUANTO AO ENTORNO

A partir das entrevistas, coletaram-se dados sobre a satisfação com o entorno do conjunto habitacional na opinião dos moradores, sendo organizados em tabela, os resultados obtidos.

Após a construção da tabela de distribuição de frequência foi determinada a tabela de quantificação Ridits e posterior Análise de Agrupamento. Como nem todas as casas tinham pessoas que frequentavam as escolas, tal questionamento foi retirado da amostra (item 2.6). Os resultados da quantificação Ridits encontram-se na Tabela 3.4.

Tabela 3.4: Quantificação das categorias Ridits para as questões relacionadas à satisfação quanto ao entorno

Questão	Péssimo	Ruim	Regular	Bom	Ótimo
2.1	0,0000000	0,0102041	0,0408163	0,4591837	0,9285714
2.2	0,0204082	0,0510204	0,0714286	0,4489796	0,9081633
2.3	0,0102041	0,0510204	0,1326531	0,5510204	0,9591837
2.4	0,0000000	0,0000000	0,0204082	0,2551020	0,7346939
2.5	0,0306122	0,2040816	0,5000000	0,8061225	0,9795918
2.7	0,0306122	0,2040816	0,5000000	0,8061225	0,9795918
2.8	0,0000000	0,0102041	0,0918367	0,3979592	0,8163265
2.9	0,0000000	0,0102041	0,1122449	0,3979592	0,7959184
2.10	0,0918367	0,3775510	0,6734694	0,8775510	0,9897959

Fonte: As autoras (2019).

Na Tabela 3.5 encontram-se os valores de CCC dos três métodos da Análise de Agrupamento. Todos os valores segundo Rohlf (1970) são considerados adequados. Os valores obtidos foram novamente muito próximos, da mesma forma que na análise anterior, para conseguir a melhor análise de dados foi escolhido o método com o maior CCC que está também relacionado ao método da Médias das Distância que obteve um valor de 0,928 apresentando uma ótima qualidade de agrupamento, das variáveis referentes a satisfação quando ao entorno.

Tabela 3.5: Coeficiente de correlação cofenético (CCC)

Método	CCC
Ligação Simples	0,915
Ligação Completa	0,923
Média das Distâncias	0,928

Fonte: As autoras (2019).

Após a escolha do método foi construído o dendrograma com base na matriz de dissimilaridade que está apresentado na Figura 3.2.



Figura 3.2: Dendrograma da satisfação quanto ao entorno.

Fonte: As autoras (2019).

Observa-se que formou-se 5 grupos distintos, um para cada uma das classificações qualitativas ordinais: Ótimo, Bom, Ruim, Regular e Péssimo. Classificando estatisticamente como “Bom” o grupo formado pelos itens 2.8 e 2.9, que dizem respeito a ponto de ônibus e praças/área de lazer.

Já os itens Iluminação pública da sua rua (2.3), Rua de acesso ao condomínio (2.1) e Drenagem das ruas (2.2) foi classificado como “Regular”.

O grupo formado pelos itens 2.5 e 2.7, que se refere a coleta seletiva de lixo e comércio, respectivamente, apresentaram mesma distribuição de frequência sendo classificados como “Péssimo”. Referente ao comércio, foi constatado que próximo ao conjunto tem-se três comércios de pequeno porte (mercearias), de propriedade dos moradores locais. Pensando no custo benefício alguns moradores preferem fazer as compras de alimentos e itens de higiene nos mercados maiores da cidade, os quais encontram-se a aproximadamente 2,5 Km do Loteamento Jardim das Orquídeas, acarretando assim na classificação da categoria “Péssimo”.

Dois grupos formaram-se com uma única variável: 2.4, referente a coleta de lixo com classificação na categoria “Ótimo” e 2.10, referente a linha de ônibus, classificado como “Ruim”.

3.5 CONCLUSÃO

Após a análise dos dados, entende-se que em relação aos cômodos, as áreas de comum acesso como cozinha, banheiro e área de serviço são as categorias com alto nível de insatisfação. Analisando de uma forma geral, a concentração percentual nas categorias péssimo, ruim e regular é muito maior do que nas categorias bom e ótimo, mostrando que os moradores não estão satisfeitos com as moradias.

Já no que diz respeito ao entorno do conjunto habitacional, foi possível verificar que os valores se concentram mais nas categorias bom e ótimo, mostrando insatisfação em categorias bem pontuais. Vale a pena ressaltar que a coleta de lixo, que é a categoria que tem a maior satisfação dos moradores 94%, deixa a desejar na coleta seletiva de lixo classificada como “Regular” na aplicação do método com 32% nessa categoria, somando com as outras categorias ruins totalizando 66% de insatisfação, sendo essa coleta de grande valia para o meio ambiente.

Ademais, foi possível notar o desagrado dos moradores com cômodos, apesar de estarem de acordo com a normativa. Assim, os resultados sinalizam que o programa MCMV deve aperfeiçoar suas normas para atender as reais necessidades dos moradores nos critérios analisados e assim propiciar o bem-estar do domiciliado.

4 CONSIDERAÇÕES FINAIS

Analisando os resultados obtidos nos dois artigos que compõem o presente trabalho, percebe-se que tanto a Análise de Correspondência Simples quanto a quantificação Ridits se adequaram muito bem ao conjunto de dados analisados.

Em ambos os métodos de quantificação aplicados, os resultados das análises estatísticas multivariadas mostraram que há insatisfação dos mutuários do programa MCMV quanto ao tamanho e disposição dos cômodos da casa. Em ambas as análises, o tamanho e disposição dos móveis da cozinha e a localização do banheiro foram classificados como péssimos.

Já em relação ao entorno e a prestação de serviços públicos, a maior insatisfação dos usuários está relacionado as linhas de ônibus que não corresponde a necessidade dos moradores. Isso se deve principalmente devido a distância do conjunto habitacional do centro da cidade, o intervalo entre as passagens dos ônibus e o número de usuários do transporte público.

Como trabalhos futuros, pretende-se aprofundar os estudos da quantificação Ridits, construindo rotinas que propiciem os testes de hipóteses multivariados e construção de intervalos de confiança para dados qualitativos ordinais.

REFERÊNCIAS

- AGRESTI, A. **Analysis of Ordinal Categorical Data**. 2. ed. New Jersey: John Wiley & Sons, 2010.
- ANDERBERG, M. R. **Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks**. [S. l.]: Academic press, 2014. v. 19.
- BARROSO, L. P.; ARTES, R. **Análise multivariada**. Lavras: Ufa, p. 151, 2003.
- BONETI, G. C.; BRESSIANI, L.; FAGUNDES, R. S. **Análise de um conjunto habitacional quanto aos requisitos de funcionalidade e acessibilidade da NBR 15575 (ABNT, 2013) e avaliação pós ocupação**. 108 pag. TCC (Graduação) - Curso de Engenharia Civil, Universidade Tecnológica do Paraná, UTFPR, Toledo, PR, 2019.
- BROSS, I. **How to use Ridit analysis**. Biometrics, Vol.14, p. 18-38, 1958.
- BUSSAB, W. D. O.; MIAZAKI, E. S.; ANDRADE, D. F. D. **Introdução a análise de agupamento**. São Paulo: USP-IME - IX Simpósio Brasileiro de Probabilidade e Estatística, 1990.
- CORDEIRO, G. M. **O amadurecimento da pesquisa e o ensino de estatística no Brasil**. De Ivy Judensnaider. 2006. Disponível em: <http://www.arscientia.com.br/materia/vermateria.php>. Acesso em: 10. maio. 2019.
- CROUSHORE, D.; SCHMIDT, R. M. **Ridit Analysis of Student Course Evaluations**. University Richmond, Virginia, 2010.
- CRUZ, C. D.; CARNEIRO, P. C. S. **Modelo biométrico aplicado ao melhoramento genético**. Viçosa: UFV, p. 585, 2006.
- FAGUNDES, R. S. **Distribuição slash multivariada aplicada a dados agrícolas**. Tese de Doutorado – Universidade Estadual do Oeste do Paraná, UNIOESTE, Cascavel, PR, 2017.
- FERREIRA, D. **Estatística multivariada**. Lavras: Editora UFLA, 2011.
- GREENACRE, M. **Correspondence Analysis in Practice**. 2^a ed., Boca Raton: Chapman and Hall/CRC, 2007.
- GUEDES, T. A. **Seleção de variáveis categóricas utilizando análise de correspondência e análise procruste**. Maringá, PR, 1999.
- GUIMARÃES, V. S.; BRESSIANI, L.; FAGUNDES, R. S. **Análise do processo executivo de alvenaria estrutural: estudo de caso**. TCC (Graduação) - Curso de Engenharia Civil,

- Universidade Tecnológica do Paraná, UTFPR, Toledo, PR, 2019.
- HYBINER, J., et al. **Uso da NBR 15575 : 2013 na avaliação técnico construtiva de um conjunto habitacional.** In: ENCONTRO NACIONAL DE TECNOLOGIA DO AMBIENTE CONSTRUÍDO, 15., 2014, Maceió. Anais. Maceió: Antac, 2014. p. 1 - 09.
- HURWITZ, A. **Ridit Analysis for Cooper-Harper e other Ordinal Ratings for Sparse Data – A Distance-based Approach.** California, EUA, 2016.
- JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis.** 5. ed. [S.l.]: ed. Englewood Cliffs: Prentice-Hall, 2002.
- JUNIOR, A. A.; THIEBAUT, J. **Análise multivariada na avaliação da diversidade em ecursos genéticos vegetais.** Campos dos Goytacazes-Universidade Estadual do Norte Fluminense-UENF, CCTA, 1999.
- KHATTREE, R.; NAIK, D. N. **Multivariate data reduction and discrimination with sas software.** Sas Institute Cary, p. 574, 2000.
- LATTIN, J.; CARROL, J. D.; GREEN, P. E. **Análise de Dados Multivariados.** São Paulo: Cengage, 2011.
- LUCHESA, C. J.; NETO, A. C. **Cálculo do tamanho da amostra nas pesquisas em administração.** Curitiba: Edição do autor, 2011.
- MAROCO, J. **Análise Estatística com o SPSS Statistics.** 5^a ed., Portugal: Report Number, 2011.
- MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada.** Belo Horizonte: UFMG, 2005. p. 295.
- OKSANEN, Jari et al. **Vegan: Community Ecology Package.** R packageversion 2.4-0, 2016. Disponível em: <https://CRAN.R-project.org/package=vegan>. Acesso em: 01 out 2018.
- PRADO, B. Q. de M. **Análise de agrupamento das taxas de incidência de dengue nos estados brasileiros.** Universidade Federal de Uberlândia, 2015.
- QUINTAL, G. M. de. C. C. **Análise de clusters aplicada ao Sucesso/Insucesso em Matemática.** Tese (Doutorado) - Universidade da Madeira, 2006.
- R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing.** reference index version 3.5.1. R Foundation for Statistical Computing, Viena, Áustria, ISBN, 2018.
- RENCHER, A. C. **Methods of multivariate analysis.** 2. ed. [S.l.]: New York: John Wiley, 2002. 708 p.
- ROHLF, F. J. **Adaptive hierarchical clustering schemes.** In: . [S. l.]: Systematic

Zoology, 190. v. 18, cap . p. 58-82.

SNEATH, P. H. A.; SOKAL, R. R. **Numerical taxonomy**. [S. l.]: New York: Hafner, 1975.

STEINER, M. T. A.; CHAVES NETO, A; BRAULIO, S. N.; ALVES, V. **Métodos estatísticos multivariados aplicados à engenharia de avaliações**. Revista Gestão e Produção, São Carlos, v. 15, n. 1, p. 23-32, jan.-abr. 2008.

STIGLER, S. M. **The history of statistics: the measurement of uncertainty before 1900**. Cambridge, USA: The Belknap Press of Harvard University Press, 1986.

TOLEDO. **Secretaria de Esportes. Região do Jardim Coopagro ganha arena multiuso para prática de esportes**. Disponível em:

<http://www.toledo.pr.gov.br/noticia/regiao-do-jardim-coopagro-ganha-arena-multiuso-para-pratica-de-esportes>. Acesso em: 25 abr. 2019.

UWAWUNKONYE, E. G.; ANAENE, O. I. C. **A compaative study between ridit and modified ridit analysis**. Science Publishing Group. Nigéria. 2013.

VALLI, M. **Análise de cluster**. Augusto Guzzo Revista Acadêmica, n. 4, p. 77-87, 2002.

VIAÇÃO SORRISO DE TOLEDO. **Horários e Mapas**. Disponível em:

<http://sorrisodetoledo.com.br/horarios-e-mapas-urbano/> Acesso em: 04 maio 2019.

VICINI, L. **Análise multivariada da teoria à prática**. Santa Maria, RS, 2005.

WU, C. **On the application of grey relational analysis and RIDIT analysis to likert scale surveys**. Taiwan. 2007.

A Apêndice

Rotina R relacionada ao Artigo 1

Nome: Danieli Felichak e Regiane Slongo Fagundes

Data da criação: 23 de setembro de 2019

Última modificação: 02 de dezembro de 2019

Motivo da criação: Rotina criada para análise dos dados e aplicação do método multivariado Análise de Correspondência Simples (ACS).

```
*****  
***** ANÁLISE DE CORRESPONDÊNCIA *****  
***** SATISFAÇÃO QUANTO A CADA CÔMODO *****  
*****
```

```
dados = read.table ("variavelComodo.txt", header = T)  
dados=dados[1:60,1:3]  
attach(dados)  
** Montando a tabela de distribuição de frequência absoluta **  
tab = matrix(freq, ncol = 12, nrow = 5, byrow = F)  
** Colocando o nome das linhas e colunas para construção da tabela **  
colnames(tab) = c("1.1", "1.2", "1.3", "1.4", "1.5", "1.6", "1.7", "1.8", "1.9", "1.10",  
"1.11", "1.12")  
rownames(tab) = c("Péssimo", "Ruim", "Regular", "Bom", "Ótimo")  
**Inferência para obter teste qui-quadrado para saber se existe associação entre as variáveis  
**  
qui = chisq.test(tab) ** teste Qui-Quadrado **  
** Análise de correspondência **  
n = sum(tab)  
P = tab/n ** matriz de correspondência **  
r1 = sum(tab[1,])/n  
.  
.  
r5 = sum(tab[5,])/n  
r = c(r1,r2,r3, r4, r5) ** vetor coluna **  
c1 = sum(tab[,1])/n  
.  
.  
.
```

```

c12 = sum(tab[,12])/n
c = c(c1,c2,c3,c4,c5,c6,c7,c8,c9,c10,c11,c12) ** vetor linha **
*****

***** Criação das Matrizes de Correspondência *****
*****

**Matriz de perfil das linhas **
Dr = diag(r)
** Matriz de perfil das colunas **
Dc = diag(c)
** Decomposição em coordenadas principais **
sDr = diag((r(-0.5))) ** Dr(-1/2) **
sDc = diag((c(-0.5))) ** Dc(-1/2) **
R = sDr %*% P %*% sDc
** Coordenadas principais das linhas **
W = t(R) %*% R
eigen(W) ** inércias: 2o e 3o autovalores **
an = diag((r(-1))) %*% P %*% diag((c(-0.5)))
score1 = an %*% eigen(W)vectors ** 2a e 3a colunas são as coordenadas principais das
linhas **
** Coordenadas principais das colunas **
T = R %*% t(R)
eigen(T) ** inércias: 2o e 3o autovalores **
an2 = diag((c(-1))) %*% t(P) %*% diag((r(-0.5))) an2
score2 = an2 %*% eigen(T) vectors ** 2a e 3a colunas são as coordenadas principais das
colunas **
*****

** PACOTE FactoMineR para construção do gráfico **
*****

require(FactoMineR)
corresp = CA(tab) ** roda apenas para no de categorias por variável acima de 2 **
plot(corresp, cex=1.2, cex.axis=1.3, title=, )
summary(corresp)

*****

ROTINA SEMELHANTE FOI UTILIZADA PARA ANALISAR OS DADOS REFERE-
RENTE A SATISFAÇÃO QUANTO AO ENTORNO
*****

```

B Apêndice

Rotina R relacionada ao Artigo 2

Nome: Danieli Felichak e Regiane Slongo Fagundes

Data da criação: 14 de outubro de 2019

Última modificação: 02 de dezembro de 2019

Motivo da criação: Rotina criada para análise dos dados e aplicação dos métodos multivariados Quantificação Ridits e Análise de Agrupamento.

```
*****
***** QUANTIFICAÇÃO RIDITS *****
***** SATISFAÇÃO QUANTO A CADA CÔMODO *****
*****

dados = read.table("PesoRidits3.txt", header = T) ** lendo um conjunto de dados em
txt **
** dados das colunas, onde tem as variáveis **
x = dados[,1:5]
** mostra os valores de junção no dendograma para o primeiro cálculo das distâncias
Euclidianas **
disteuc = dist(dados) ** distância euclidiana **
*****

Método de ligação completa (Complete linkage) Medida de similaridade entre dois clusters
é definida pela maior distância de qualquer ponto do 1º cluster para qualquer ponto do
2º cluster.
*****

descritores.hc.complete =hclust(disteuc, method="complete")
descritores.hc.complete $height
** Plotando o dendograma **
plot(descritores.hc.complete, main="Complete Linkage", xlab="", sub="", cex=.9, hang=-
1)
** definição do ponto de corte em k=5 Critério da amplitude das distâncias **
rect.hclust(descritores.hc.complete, k = 5)
abline(h=16.8, col="blue")
** coeficiente de correlação cofenético **
d1 = cophenetic(descritores.hc.complete)
cor(disteuc,d1)
cor.test(disteuc,d1)
```

 Método de ligação simples (Single linkage). Medida de similaridade entre dois clusters é definida pela menor distância de qualquer ponto do 1º cluster para qualquer ponto do 2º cluster.

 Comandos que executam o algoritmo HC com distância Euclidiana e os tipos de linkage
 descritores.hc.single = hclust(disteuc, method="single")
 descritores.hc.single \$height
 **Plotando o dendograma plot(descritores.hc.single, main="Single Linkage", xlab="", sub="", cex=.9, hang=-1) **
 ** definição do ponto de corte em k=5 Critério da amplitude das distâncias **
 rect.hclust(descritores.hc.single, k = 5)
 abline(h=8.2, col="blue")
 ** coeficiente de correlação cofenético **
 d3 = cophenetic(descritores.hc.single)
 cor(disteuc,d3)
 cor.test(disteuc,d3)

 Método da média das distâncias (Average linkage). Medida de similaridade entre dois clusters é definida pela média das distâncias de todos os pontos do 1º cluster em relação aos pontos do 2º cluster.

 Comandos que executam o algoritmo HC com distância Euclidiana e os tipos de linkage
 descritores.hc.average = hclust(disteuc, method="average")
 descritores.hc.average \$height
 ** Plotando o dendograma **
 plot(descritores.hc.average, main="Average Linkage", xlab="", sub="", cex=.9, hang=-1)
 ** definição do ponto de corte em k=5 Critério da amplitude das distâncias **
 rect.hclust(descritores.hc.average, k = 5)
 abline(h=11.8, col="blue")
 ** coeficiente de correlação cofenético **
 d2 = cophenetic(descritores.hc.average)
 cor(disteuc,d2)
 cor.test(disteuc,d2)

 ROTINA SEMELHANTE FOI UTILIZADA PARA ANALISAR OS DADOS REFERENTE A SATISFAÇÃO QUANTO AO ENTORNO

A Anexo

Questionário aplicado aos moradores do Conjunto Habitacional Jardim das Orquídeas

Quadro 1: Satisfação quanto a cada cômodo

	Ótimo	Bom	Regular	Ruim	Péssimo
Tamanho da cozinha					
Disposição dos móveis da cozinha					
Tamanho da área de serviço					
Disposição dos móveis da área de serviço					
Tamanho da sala					
Disposição dos móveis na sala					
Tamanho dormitório					
Disposição dos móveis nos dormitórios					
Quantidade de Tomadas e interruptores					
Localização de tomadas e interruptores					
Quantidade de janelas					
Localização das janelas					

Fonte: Bonetti (2019).

Quadro 2: Satisfação quanto ao entorno.

	Ótimo	Bom	Regular	Ruim	Péssimo
Rua de acesso ao condomínio					
Drenagem das ruas					
Iluminação pública da sua rua					
Coleta de lixo					
Coleta seletiva de lixo					
Distância entre o residencial e o colégio					
Comércio					
Pontos de ônibus					
Praças/ área de lazer					
Linhas de ônibus					

Fonte: Bonetti (2019).