

**PERFIL GEOGRÁFICO MULTIVARIADO DA ÁGUA CONSUMIDA  
NO MUNICÍPIO DE BOTUCATU**

**Lívia Paschoalino de Campos**

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

**BOTUCATU**  
São Paulo - Brasil  
Fevereiro - 2014

**PERFIL GEOGRÁFICO MULTIVARIADO DA ÁGUA CONSUMIDA  
NO MUNICÍPIO DE BOTUCATU**

**Lívia Paschoalino de Campos**

Orientador: Prof. Dr. **Carlos Roberto Padovani**

Dissertação apresentada à Universidade Estadual Paulista “Júlio de Mesquita Filho” para a obtenção do título de Mestre em Biometria.

BOTUCATU  
São Paulo - Brasil  
Fevereiro - 2014

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.  
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - CAMPUS DE BOTUCATU - UNESP  
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE - CRB 8/5651

Campos, Livia Paschoalino de.  
Perfil geográfico multivariado da água consumida no município de Botucatu /  
Livia Paschoalino de Campos. - Botucatu, 2014

Dissertação (mestrado) - Universidade Estadual Paulista, Instituto de  
Biotecnologia de Botucatu

Orientador: Carlos Roberto Padovani

Capes: 10202056

1. Água - Qualidade. 2. Análise multivariada. 3. Distribuição geográfica. 4.  
Físico-química. 5. Bacteriologia.

Palavras-chave: Análise multivariada; Perfil da potabilidade; Perfil  
hidrogeográfico; Qualidade da água.

*Aos meus amados pais, Regina e Eduardo,  
por todo amor.*

## Agradecimentos

A Deus que me conduziu durante este período, nos momentos de alegria e angustias.

Ao meu estimado orientador Prof. Dr. Carlos Roberto Padovani por seus sábios ensinamentos, pela sua dedicação, pelo seu entusiasmo e amor a docência, por acreditar em mim e me motivar sempre. Suas discussões, reflexões e conselhos foram essenciais ao longo de todo o percurso e serão levadas para toda vida. Minha eterna gratidão.

Aos meus pais, meus grandes amores e exemplos de vida, que são fontes inesgotáveis de compreensão, paciência, amor, carinho e cumplicidade. Pelo apoio emocional, por não medirem esforços para que eu pudesse concretizar os meus sonhos. Sem vocês nada disto seria possível!

A todos meus queridos amigos que mesmo de longe ou de perto me acompanharam e sempre me incentivaram nesta etapa.

Aos amigos conquistados no mestrado, em especial, Arthur, Cíntia, Josi, Renan, Nilo, Bettina, Adriano e Luiz, pela boa convivência, pelas conversas, pelas muitas ajudas recebidas e por fazerem com que este caminho fosse cada dia mais agradável.

Ao Prof. Dr. Rogério Antônio de Oliveira e a Prof. Dra. Lilian Cristina Trevizan Felipe pelas preciosas sugestões apresentadas no exame de qualificação que foram fundamentais para o enriquecimento do trabalho.

À Prof. Ana Paula e à querida amiga Débora pela revisão gramatical deste trabalho.

À querida Fabiana, ao Luiz e ao amigo Adriano pela ajuda com o

resumo na língua inglesa.

Aos professores e funcionários do Departamento de Bioestatística pela  
presteza e disponibilidade sempre.

À Companhia de Saneamento Básico do Estado de São Paulo pelo  
fornecimento dos dados utilizados na pesquisa.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CA-  
PES) pelo apoio financeiro concedido.

Ao Programa de Pós Graduação em Biometria.

A todos que de alguma forma contribuíram para a elaboração deste  
trabalho.

# Sumário

	Página
<b>LISTA DE FIGURAS</b>	<b>viii</b>
<b>LISTA DE TABELAS</b>	<b>ix</b>
<b>RESUMO</b>	<b>x</b>
<b>SUMMARY</b>	<b>xii</b>
<b>1 INTRODUÇÃO</b>	<b>1</b>
<b>2 REVISÃO DE LITERATURA</b>	<b>4</b>
2.1 Qualidade da água para consumo . . . . .	4
2.2 Análise Multivariada Aplicada . . . . .	9
<b>3 MATERIAIS E MÉTODOS</b>	<b>19</b>
3.1 Caracterização da área de estudo . . . . .	19
3.2 Amostragem e pontos de coleta . . . . .	20
3.3 Parâmetros envolvidos no estudo . . . . .	21
3.3.1 Parâmetros físico - químicos . . . . .	21
3.3.2 Parâmetros bacteriológicos . . . . .	23
3.4 Conjunto de dados . . . . .	24
3.5 Análise de Agrupamento . . . . .	32
3.6 Análise de Componentes Principais . . . . .	38

<b>4 RESULTADOS E DISCUSSÃO</b>	<b>41</b>
4.1 Medidas descritivas e matriz de correlação . . . . .	41
4.2 Análise de Agrupamento . . . . .	42
4.3 Análise de Componentes Principais . . . . .	55
<b>5 CONCLUSÃO</b>	<b>64</b>
<b>REFERÊNCIAS</b>	<b>66</b>
<b>6 APÊNDICE</b>	<b>72</b>
6.1 Dados . . . . .	72



## Lista de Figuras

	Página
1	Mapa município de Botucatu. . . . . 19
2	Distribuição dos pontos de coleta (bairros). . . . . 27
3	Dendrograma resultante da análise de agrupamento dos bairros do município de Botucatu. . . . . 43
4	Gráfico do comportamento do nível de fusão. . . . . 44
5	Gráfico do nível de similaridade. . . . . 48
6	Distribuição geográfica dos grupos de semelhança. . . . . 54
7	Biplot do primeiro e segundo componente principal. . . . . 57
8	Biplot do primeiro e terceiro componente principal. . . . . 58
9	Biplot do segundo e terceiro componente principal. . . . . 58

## Lista de Tabelas

	Página
1	Padrões referentes à Portaria nº 518 . . . . . 5
2	Distribuição dos pontos de coleta (bairros) no município de Botucatu . . . 28
3	Medidas descritivas das variáveis relacionados a qualidade da água durante o período de 2007 a 2011. . . . . 41
4	Matriz de correlação das variáveis relacionadas à qualidade da água durante o período de 2007 a 2011. . . . . 42
5	Nível de similaridade do agrupamento segundo a distância euclidiana e o método de Ward. . . . . 45
6	Bairros pertencentes ao Grupo 1. . . . . 49
7	Bairros pertencentes ao Grupo 2. . . . . 50
8	Bairros pertencentes ao Grupo 3. . . . . 50
9	Bairros pertencentes ao Grupo 4. . . . . 50
10	Bairros pertencentes ao Grupo 5. . . . . 53
11	Perfil médio de respostas físico-química e bacteriológica dos grupos. . . . 55
12	Autovalores e porcentagens da variação total explicada . . . . . 56
13	Coefficientes físico-químicos e bacteriológicos . . . . . 56
14	Bairros ordenados pelo valor da distância . . . . . 59
15	Perfil das respostas físico-químicas e bacteriológicas dos bairros (matriz de dados) . . . . . 72

# PERFIL GEOGRÁFICO MULTIVARIADO DA ÁGUA CONSUMIDA NO MUNICÍPIO DE BOTUCATU

Autora: LÍVIA PASCHOALINO DE CAMPOS

Orientador: Prof. Dr. CARLOS ROBERTO PADOVANI

## RESUMO

Considerou-se, na presente pesquisa, parâmetros físico-químicos e bacteriológicos ponderados como determinantes de estudo para a qualidade da água com coletas diárias de amostras em diversos bairros do município de Botucatu – SP, no período de 2007 a 2011. Para descrever o perfil da qualidade da água dos bairros, utilizou-se a técnica de análise multivariada de dados, especificamente a Análise de Agrupamento e Análise de Componentes Principais. A primeira empregou como indicador de similaridade entre os perfis de respostas dos bairros a distância euclidiana, que é a métrica mais utilizada quando os dados são quantitativos e, como algoritmo de agrupamento, o método hierárquico de Ward, que tem como objetivo produzir agrupamentos mais heterogêneos quanto possíveis. A Análise de Componentes Principais foi utilizada com o objetivo de realizar a redução do espaço paramétrico e, posteriormente, utilizar os eixos descritores do novo sistema de dados

para a ordenação da qualidade da água por bairros e verificar a associação entre os dois procedimentos empregados. Os resultados obtidos pela análise de agrupamentos definiram o perfil geográfico da qualidade da água, constituído por cinco grupos de similaridade, formados respectivamente, por 13, 2, 7, 39 e 15 bairros. Os bairros em cada um dos grupos estavam praticamente distribuídos nos quatro pontos cardeais do município e os perfis médios de resposta de todos os grupos encontraram-se quase que totalmente em concordância com o padrão de qualidade da água, estabelecido pela Portaria nº518, de março de 2004. Já a Análise de Componentes Principais selecionou os três primeiros componentes que foram utilizados para obter as distâncias euclidianas das respostas dos bairros em relação à origem do novo sistema. Desta forma, foi possível realizar a ordenação dos pontos (bairros) de acordo com os parâmetros físico-químicos e bacteriológicos.

# MULTIVARIATE GEOGRAPHIC PROFILE OF THE WATER CONSUMED IN THE CITY OF BOTUCATU

Author: LÍVIA PASCHOALINO DE CAMPOS

Adviser: Prof. Dr. CARLOS ROBERTO PADOVANI

## SUMMARY

The following research involved some physical-chemical and bacteriological parameters that play an important role in the study of water quality by collecting daily samples from different neighborhoods in Botucatu city – SP, from the period of 2007 to 2011. The multivariate data analysis, specifically the Cluster Analysis and Principal Components Analysis was used to describe the profile of water quality of the city neighborhoods. To accomplish Cluster Analysis we used the Euclidean distance to find the similarity between the response profiles, which is the most used metric when data are quantitative, and the hierarchical method of Ward as clustering method, which produces heterogeneous clusters. Principal Components Analysis was used for achieving the reduction of the parametric space and afterward using the descriptors axes of the new data system for arranging water quality per neighborhoods, verifying the association between two used procedures. The results

of cluster analysis defined the geographic profile of the water quality constituted of five groups of similarity, respectively, by 13,2, 7, 39 and 15 neighborhoods. The neighborhoods in each group were distributed according to the 4 positions of the city and the average profile of the responses of all groups were very similar to the standards of water quality established by the Brazilian Law nº518, of March, 2004. The Analysis of Principal Components selected the first three components that were used to obtain the Euclidean distances of neighborhoods responses in relation to the origin of the new system. Thus, it was possible to perform the arrangement of the points (neighborhoods) according to the values of physico-chemical and bacteriological parameters.

# 1 INTRODUÇÃO

A vida resulta da interação entre quatro elementos fundamentais da natureza, a saber: o fogo (força propulsora da vida que é produzida dentro das células pela combustão); o mineral (elemento que circula livremente nos organismos, responsabilizando-se pelo transporte dos alimentos e dejetos do sangue); a água (elemento determinante e marcante do corpo) e o ser vivo (este representado no elemento vegetal carbono, configuração atômica cuja instabilidade molecular promove uma constante troca de parceiros no contexto da combustão). Cada um dos quatro elementos desempenha uma função própria relacionada à Terra, sendo esta o quinto integrante na sinfonia da vida.

Destaca-se, frente ao interesse de estudo, o elemento fundamental água, composto por dois átomos de hidrogênio (H) e um de oxigênio (O), constituindo uma substância abundante em nosso planeta (aproximadamente 70% da superfície terrestre encontra-se coberta por água; no entanto, uma porcentagem expressivamente baixa deste volume é de água doce) e que pode se encontrar em três estados físicos: sólido, líquido e gasoso. A água é de fundamental importância para a vida de todas as espécies, aproximadamente 80% de organismo humano é composto por água. Parte expressiva dos pesquisadores concorda que a ingestão de água tratada é um dos mais importantes fatores para a conservação da saúde, rotulada por muitos como o solvente universal, auxilia na prevenção de doenças e na proteção do organismo contra o envelhecimento.

Sabe-se que o interesse no consumo humano da água potável vem desde o período a.C. e atualmente cada vez mais se tem o conhecimento da importância do tratamento da água, pois só assim é possível se ter uma água livre de contaminação.

Esse cuidado propicia reduzir teores elevados de compostos orgânicos, protozoários, bactérias e microrganismos patogênicos, evitando consideravelmente a transmissão de várias doenças que podem levar à morte.

No contexto mundial, o Brasil é um país privilegiado em relação a disponibilidade de água, pois detém mais da metade do manancial de água doce disponível na América do Sul e possui ainda o maior rio do planeta (rio Amazonas). Salienta-se também no Brasil, a presença do Aquífero Guarani, um enorme manancial de água doce subterrânea, que por sua tamanha extensão também abrange países como a Argentina, Paraguai e Uruguai, sendo a maior parte (71%) localizada em território brasileiro (Carneiro, 2010).

Os climas equatorial, tropical e subtropical, que atuam sobre o território, proporcionam elevados índices pluviométricos. No entanto, mesmo com enorme disponibilidade de recursos hídricos, o país sofre profundamente com a escassez de água potável em grande parte do território.

Somando-se a essa situação a ausência de saneamento básico nas residências, fator que torna a situação notavelmente mais agravada. Há necessidade urgente de desenvolvimento de políticas públicas para reverter esse quadro. Várias pesquisas nesse sentido têm indicado que para cada real investido em saneamento, o governo deixa de gastar outros cinco em serviços de saúde, ou seja, são investimentos que proporcionam qualidade de vida para a população e economia aos cofres públicos em curto prazo.

Diante desse histórico, pretende-se no presente estudo considerar o município de Botucatu, estado de São Paulo, como unidade de pesquisa e utilizar o banco de dados da Companhia de Saneamento Básico do Estado de São Paulo (Sabesp), responsável pelo abastecimento de água no município, para investigar a qualidade da potabilidade da água consumida pelos munícipes. Ou seja, a partir dos parâmetros utilizados para aferir as propriedades físico-químicas e bacteriológicas da água e da consideração dos pontos de coleta de amostras diárias realizadas na rede de distribuição entre os anos de 2007 e 2011. Embora no banco de dados



da Sabesp estivesse disponível o ano de 2012, este foi desconsiderado por haver mudança dos equipamentos utilizados para a aferição dos parâmetros físico-químicos e bacteriológicos pela empresa responsável e também pela substituição da Portaria nº518 para a Portaria nº2914. Em síntese, a pesquisa busca estabelecer o perfil geográfico da qualidade da água no município por meio de aplicação de técnicas de análise multivariada, visando classificar e agrupar os pontos de coletas segundo o padrão de qualidade da água, finalizando com a distribuição geográfica dos bairros no município.

## 2 REVISÃO DE LITERATURA

### 2.1 Qualidade da água para consumo

Sabe-se que a água distribuída para o consumo humano deve ser potável, isto é, estar livre de qualquer tipo de contaminação que ofereça riscos à saúde dos seres vivos que a consomem, como os animais e os humanos. Segundo a Portaria nº 518/2004 - Controle e Vigilância da Qualidade da Água para Consumo Humano, a água considerada potável para o consumo humano deve respeitar parâmetros quanto às características microbiológicas, físicas, químicas e radioativas e ainda não oferecer riscos à saúde humana. É interessante salientar que a água considerada contaminada pode por meio de procedimentos físicos e químicos eliminar todas as substâncias nocivas nela contida e, neste sentido, após tratamento, tornar-se adequada ao consumo.

O tratamento da água abrange desde a sua captura em mananciais, os quais devem ser previamente selecionados para suprir a demanda necessária e possuir água livre de possíveis contaminações, até a sua distribuição aos usuários. O tratamento mais utilizado possui como principais etapas do processo a coagulação, floculação, decantação, filtração, desinfecção e fluoretação.

Para que o sistema de tratamento de água tenha êxito por completo, desde a distribuição até o armazenamento, é fundamental que os reservatórios domiciliares recebam a higienização correta realizada periodicamente, (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2000) caso contrário a água tratada pode se tornar imprópria para o consumo.

No contexto de existir medidas que regulamentem os padrões de quali-

dade das águas destinadas ao abastecimento público, o Ministério da Saúde por meio de portarias criou padrões que indicam ou fixam os limites aceitáveis para parâmetros físico-químicos e microbiológicos. No Brasil, atualmente vigora a Portaria nº 2914, de 12 de dezembro de 2011 do Ministério da Saúde que define os valores máximos permissíveis para as características microbiológicas, organolépticas e físico-químicas nas águas distribuídas. Porém, no estudo em questão, será utilizada como referência a Portaria nº 518, de 25 de março de 2004 do Ministério da Saúde, intitulada como “Controle e Vigilância da Qualidade da Água” para Consumo Humano, já que os dados utilizados na pesquisa são referentes ao período em que a mesma vigorava.

Considerando tal portaria e as referências para as variáveis de interesse do estudo, fica estabelecidos os seguintes padrões.

Tabela 1: Padrões referentes à Portaria nº 518

<b>Parâmetro</b>	<b>Unidade</b>	<b>Padrão</b>
Cor aparente	Unidade de cor	Máximo 15 UC
Turbidez	Unidade nefelométrica de turbidez	Máximo 5 UNT
Flúor	Miligrama por litro	Máximo 1,5 mg/L
Cloro residual	Miligrama por litro	Mínimo 0,5 mg/L Máximo 2,0 mg/L
pH		Mínimo 6 Máximo 9,5
Bactéria heterotrófica	Unidade formadora de colônia	Máximo 500 UFC por mL
Coliforme total	Número mais provável	Ausência em 95% das amostras aos mês

Os padrões apresentados no Tabela 1 constituem uma referência para todo território nacional, com exceção à quantidade de flúor que varia de acordo com a temperatura do ar e clima de cada região associada ao consumo médio diário de água por indivíduo, sem exceder 1,5 mg/L (Macêdo, 2004).

Com a preocupação relacionada ao correto armazenamento da água, Oliveira et al. (2002) investigam a correlação entre o estado de conservação de reservatórios da rede de distribuição pública e a potabilidade da água na cidade de Manaus - AM, no período de março a dezembro de 2001, considerando como variáveis de interesse para o estudo as ocorrências de coliformes totais, coliformes fecais e mensurações de turbidez e cloro residual livre. A partir de procedimentos estatísticos utilizando o teste Qui - Quadrado, verifica-se que o estado de conservação dos reservatórios, embora se encontrem em condições precárias, não é a principal causa dos resultados desfavoráveis à potabilidade da água fornecida no município, sendo necessário um estudo mais profundo nas condições da rede de distribuição e na eficiência das estações de tratamento de água.

Amaral et al. (2003) com o interesse de estudar a qualidade da água consumida em propriedades rurais no nordeste do estado de São Paulo, colheram amostras de fontes, reservatórios e bebedouros em 30 propriedades, num total de 180 amostras nos períodos de seca e chuva. Os resultados da análise de ocorrência de coliformes totais, *Escherichia coli* e do número de microrganismos mesófilos possibilitam observar que independente do período, seca ou chuva, a qualidade da água consumida nessas propriedades encontram-se fora dos padrões microbiológicos de potabilidade, em torno de 90% na seca e 83,3% no período mais chuvoso. Constata-se também que apenas 3,3% das propriedades avaliadas realizam a higienização periódica de seus reservatórios.

Siqueira et al. (2010) realizam avaliações quanto a qualidade bacteriológica da água para o consumo humano, verificando a presença de coliformes totais e/ou termotolerantes em unidades de alimentação na cidade do Recife - PE, mais precisamente nos arredores da Universidade Federal de Pernambuco (UFPE) e da Universidade Federal Rural de Pernambuco (UFRPE), onde foram coletadas 40 amostras de água. Essas unidades são abastecidas pela rede pública de distribuição de água, e o armazenamento da mesma é feito em cisternas e/ou reservatórios. Os resultados mostram 62,5% das amostras positivas para coliformes totais e 42,5%

para coliformes termotolerantes, são porcentagens altas e situam-se fora dos padrões microbiológicos vigentes em legislação.

Porto et al. (2011) utilizam análises microbiológicas para verificar a existência de bactérias do grupo coliformes e do subgrupo termotolerantes (*Escherichia coli*) na água de oito lojas de uma rede de *fast-food*, no ano de 2006, na cidade do Recife - PE e na Região Metropolitana visando averiguar se estavam de acordo com os padrões exigidos pela legislação. As lojas 1, 2 e 3 recebem água do sistema de abastecimento público, e a loja 4 além de água do sistema público, tem parte atendida pelo sistema de saneamento básico. A loja de número 5 recebe água de poço freático com bomba dosadora de cloro, sendo parcialmente atendida pelo sistema de saneamento básico. A loja 6 tem um abastecimento oriundo de poço freático com bomba dosadora de cloro. As lojas 7 e 8 encontram-se abastecidas pelo sistema público e também por poço freático com bomba dosadora de cloro, visando reduzir custos quanto a recursos hídricos. Amostras das oito lojas foram colhidas para todos os meses do ano, totalizando, dessa forma, 96 amostras coletadas. Vale ressaltar que em todas as lojas efetuou-se higienização periódica de seus reservatórios por empresas especializadas. Após a realização das análises microbiológicas, observa-se que 12,5% dessas amostras encontravam-se fora dos padrões estabelecidos pela Portaria nº518/2004 do Ministério da Saúde. Todas as lojas apresentam contaminação por coliformes totais, e uma também por coliformes termotolerantes. Esta loja foi abastecida exclusivamente por poço freático.

Campos et al. (2002) coletaram semanalmente amostras de água (abril de 2001) em locais representativos do município de Araraquara - SP para analisar parâmetros físico-químicos e microbiológicos, objetivando verificar a qualidade da água distribuída à população. Para verificar se existe influência da tubulação na qualidade da água, três locais de coleta foram selecionados, envolvendo proximidades do reservatório de distribuição, pontos intermediários e pontas de rede. O resultado do teste estatístico de Kruskal Wallis não indica diferença significativa nos três locais selecionados, ou seja, os parâmetros analisados mantêm o mesmo padrão de

qualidade, independente da localização do ponto de coleta. O estudo também detecta que algumas amostras de pH, cloro residual e flúor não se encontram dentro dos padrões desejados.

Para avaliar a qualidade da água que abastece áreas com saneamento na região metropolitana de Belém - PA e verificar a qualidade da água dos reservatórios domiciliares, Sá et al. (2005) coletaram amostras em 95 residências, e algumas variáveis microbiológicas foram analisadas. Foi possível observar que as áreas estudadas recebem água com os padrões exigidos para o consumo humano, porém a intervenção humana altera a qualidade da água, e esta passa a apresentar índices altos de contaminação por coliformes.

Zulpo et al. (2006) investigam a qualidade microbiológica da água consumida nos bebedouros da Universidade Estadual do Centro-Oeste, em Guarapuava - PR, entre os meses de maio e dezembro em 2004. Foram coletadas 47 amostras de bebedouros distribuídos nos dois *campi* em diversos departamentos, cujos resultados indicam que 8,5% (4/47) das amostras apresentaram-se positivas para coliformes totais e 2%(1/47) positivas para coliformes fecais. Além disso, quando comparadas aos valores permissíveis pela legislação apenas 6,4% (3/47) das amostras encontram-se impróprias para o consumo humano.

Oliveira et al. (2012) avaliam variáveis físico-químicas e microbiológicas relacionadas à qualidade da água destinada ao consumo populacional da cidade de Guarariba - PB, com amostras feitas durante o mês de maio de 2006 em um único ponto de coleta, com réplicas realizadas a cada quatro dias compreendendo um período de 28 dias, num total de sete unidades amostrais. Consideram-se para o estudo as variáveis cloro residual livre, cor aparente, pH, turbidez, alcalinidade, oxigênio consumido, dióxido de carbono ( $CO_2$ ), coliformes totais e termotolerantes. Após a realização das análises das variáveis físico-químicas e microbiológicas, certifica-se que todos os resultados estão dentro dos padrões de potabilidade, determinados pela Portaria nº 518 de 2004, revelando que água de Guarariba, durante o período estudado, manteve-se apropriada ao consumo humano.

Marinelli et al. (2012) realizaram durante uma semana, nos períodos da manhã e da tarde, coletas diárias nas residências selecionadas aleatoriamente na cidade de Marília - SP, para aferir a qualidade da água de abastecimento público. No contexto da pesquisa consideram-se as variáveis pH, cloro residual, cor aparente e turbidez e avaliam se os resultados estão dentro dos padrões que garantem a qualidade para o consumo humano, uso doméstico e higiene pessoal. O resultado das análises mostra que apenas 15% das coletas avaliadas estavam fora do padrão determinado pela portaria vigente.

## 2.2 Análise Multivariada Aplicada

A Estatística Multivariada aplicada às diversas áreas do conhecimento científico apresenta nos últimos tempos um grande avanço na análise de dados biológicos e neste sentido, tem seu uso tornado visível com boa frequência nos estudos em que há interesse em analisar simultaneamente um número grande de variáveis associadas entre si (Silva, 2008). No contexto da aplicação dos procedimentos multivariados, verifica-se a existência de trabalhos relacionados à qualidade da água, conforme alguns destaques de interesse apresentados na sequência.

Mazlum et al. (1999) aplicam o procedimento multivariado que envolve a técnica da Análise Fatorial (AF) por meio da Análise dos Componentes Principais (ACP) para analisar os fatores que causam a variabilidade na qualidade da água do afluente do rio Sakarya, chamado de Porsuk, na Turquia. As coletas foram realizadas mensalmente entre os anos de 1979 e 1984, num total de 72 amostras, numa estação de monitoramento, próxima à cidade de Kutahya. A ACP complementada com a regra que desconsidera os autovalores menores que um, seleciona quatro componentes principais. Como estes quatro componentes apresentam comunalidade baixa para duas das doze variáveis apresentadas, então foram inclusos mais dois componentes, num total de seis, que apresentam comunalidade alta para todas as variáveis e explicam conjuntamente mais de 70% da variância total do conjunto de dados. Os

seis componentes são suficientes para explicar as principais fontes de variabilidade na qualidade da água que se caracterizam principalmente por descarga de resíduos domésticos e industriais, nitrificação e os efeitos sazonais.

Silva & Sacomani (2001) avaliam o grau de degradação e de influência que a poluição tem na qualidade da água do rio Pardo, na cidade de Botucatu - SP. Foram realizadas coletas mensais entre fevereiro/1995 e janeiro/1996, em oito estações de amostragem, ao longo do rio de alguns indicadores físico-químicos e microbiológico (coliformes). A utilização de procedimentos multivariados por meio da técnica da Análise de Componentes Principais verifica a necessidade do emprego de dois fatores que separam as amostras de água do rio Pardo de acordo com a sazonalidade e a distribuição espacial e explicam conjuntamente 90,95% da variabilidade total dos dados. Os resultados da interpretação dos componentes permitem identificar que os elevados teores de sais inorgânicos são os principais responsáveis pela degradação do rio Pardo.

Simeonov et al. (2003) analisam 27 parâmetros usados para avaliar a qualidade da água em 25 locais distintos durante três anos, num total de 22350 observações nos principais sistemas fluviais no norte da Grécia, com o objetivo de ter um melhor entendimento acerca da qualidade da água e do estado ecológico do sistema fluvial. A técnica da Análise de Agrupamentos (AA) tendo como medida de similaridade a distância euclidiana quadrada juntamente com o método agrupamento hierárquico de Ward é utilizada para identificar grupos semelhantes quanto aos locais de coleta, sob a composição vetorial de respostas da qualidade envolvendo parâmetros físico-químicos e níveis de poluição. A Análise de Componentes Principais (ACP) complementa o estudo para identificar fatores que influenciam os sistemas de fluviais de água e indica a necessidade dos seis primeiros componentes, todos com os respectivos autovalores superiores a um, para garantir uma explicação de aproximadamente 90% da variância total presente nos dados.

Bufon & Landim (2007) verificam a existência de semelhança nas variáveis físicas, químicas e biológicas em corpos de água barrados nos períodos



de seca e chuva, na represa Velha em Pirassununga-SP, onde se encontra o Centro de Pesquisa e Gestão de Recursos Pesqueiros Continentais (CEPTA /IBAMA /SP) entre os anos de 1998 e 2000. Como pontos de coleta, determinaram seis estações distintas para os dois períodos. Para a Análise de Agrupamento utiliza-se o coeficiente de correlação linear de Pearson como medida de similaridade e o método de Agrupamento Pareado Igualmente Ponderado (*Unweighted Pair Group Method Average* - UPGM), detectando três agrupamentos tanto para o período seco quanto para o chuvoso. O material de suspensão nos três agrupamentos formados no período de seca se associa aos radicais fosfatos e fósforo total, já no período chuvoso não foi possível fazer a mesma associação.

Nonato et al. (2007) verificam a necessidade da diminuição do número de parâmetros a serem analisados e do número de pontos monitorados pelo “Programa de Aperfeiçoamento do Monitoramento da Qualidade das Águas da Bacia do Alto Curso do Rio das Velhas”, localizado no estado de Minas Gerais, para assim obter uma redução de custos operacionais. A técnica da Análise de Agrupamento emprega 37 pontos de monitoramento juntamente com o método hierárquico aglomerativo de Ward, tendo como medida de parecnça a distância euclidiana, que possibilitou a separação dos pontos de amostragem em grupos de similaridade quanto a qualidade da água, podendo estes estarem classificados como melhores, intermediários e piores. Após uma análise minuciosa do dendograma, utilizou-se a correlação linear de Pearson nas estações mais próximas e removendo da rede de monitoramento as estações mais correlacionadas. Aplicou - se a Análise de Componentes Principais aos 39 parâmetros físico-químicos, conseguindo reduzir o espaço paramétrico para 22 parâmetros de qualidade, representados por quatro componentes principais, obtendo uma explicação de aproximadamente 80% da variabilidade dos dados.

Andrade et al. (2007b) têm como principal objetivo identificar fatores determinantes na qualidade das águas da bacia do Alto Acaraú - CE e, dessa forma, realizou coletas trimestrais entre fevereiro/2003 e março/ 2004, em seis diferentes pontos de amostragem. Para identificação dos fatores são utilizadas as técnicas

multivariadas da Análise de Fatores e Análise de Componentes Principais. Embora tenham sido analisadas 26 variáveis, 10 variáveis já foram descartadas por não se mostrarem significativas na explicação da variância total dos dados. Para escolha dos componentes, considera-se somente aquele correspondente a autovalor superior a um. Desta forma seleciona - se um modelo com três componentes com explicação de 88,18% da variância total dos dados, indicando como variáveis determinantes para a formação do primeiro fator as variáveis que estão relacionadas ao processo de intemperismo do solo. O segundo constituído por nutrientes, relacionados ao uso da terra na agropecuária, e o terceiro formado por variáveis como cor e turbidez.

Andrade et al. (2007a) utilizam técnicas de Análise Multivariada, mais especificamente, Análise de Componentes Principais (ACP) e Análise de Fatores (AF) para a seleção de indicadores da qualidade das águas superficiais na Bacia do Baixo Acaraú - CE. No estudo foram consideradas as seguintes variáveis: pH, temperatura, cor, turbidez, nitrato, amônia total, amônia livre, amônia ionizada, sódio, cloretos, potássio, sulfato, dureza, cálcio, magnésio, alcalinidade total, alcalinidade do bicarbonato, condutividade elétrica, sólidos suspensos, oxigênio dissolvido(OD), demanda bioquímica de oxigênio(DBO5), fósforo total, ortofosfato solúvel, clorofila "A", coliformes totais e coliformes termotolerantes. A técnica da ACP permite identificar quatro componentes para a seleção de indicadores da qualidade da água, retendo conjuntamente 82,1% da variação total contida nos dados. Por sua vez, a análise de fatores não promove uma redução considerável do número de variáveis, ou seja, incluiu 18 das 25 estudadas.

Girão et al. (2007) aplicam a Análise Fatorial (AF) e Análise de Componentes Principais (ACP) para identificar os fatores e as variáveis com maior contribuição na qualidade das águas do rio Jaibaras - CE. Vinte e quatro amostras de água, em cada um dos dois diferentes pontos quanto à posição ocupada do rio, foram coletadas no período de abril/2002 a junho/2005. São consideradas as seguintes variáveis para o estudo : pH, condutividade elétrica, cálcio, magnésio, sódio, potássio, bicarbonato, fosfato, cloreto, amônia, nitrato, sulfato e relação de adsorção de sódio.

As técnicas multivariadas AF e ACP produzem fatores descritores da qualidade da água, os quais são selecionados pelo critério da raiz latente (componentes associados a autovalores da matriz de correlação superiores a um). Para o primeiro ponto, os três componentes selecionados explicam conjuntamente 80% da variabilidade e para o segundo ponto do rio dois fatores, com explicação de 81% da variabilidade total dos dados. Os dois pontos apresentam como principal fonte de poluição a ação antrópica, caracterizada por dejetos de balneários, fertilizantes nitrogenados e esgostos domésticos.

Praus (2007) utiliza a técnica da Análise de Componentes Principais (ACP) para reduzir o número de parâmetros relacionados às características físico-química e microbiológicas da água, conjuntamente com a Análise Fatorial (AF) e a Análise de Agrupamento (AA) para a confirmação dos resultados obtidos pela ACP. Dezoito parâmetros são coletados em 126 amostras de águas potáveis na rede de monitoramento na cidade de Morávia do Norte, na República Tcheca. Utilizando um plano de monitoramento foram selecionadas como pontos de coleta torneiras de água situadas em edifícios particulares e comerciais, sendo cada ponto de coleta amostrado pelo menos duas vezes durante uma metade do ano em questão. A ACP selecionou seis componentes principais, usando o critério de Kaiser e o *Scree-plot* desenvolvido por Cattell, cujos seis primeiros componentes explicam 83% da variância total. O primeiro componente, o mais explicativo da variação, é exclusivamente formado por sais e o segundo caracterizado por pH, nitrato e bicarbonato. A Análise de Agrupamento realizada pelo método hierárquico de Ward separa as amostras em três grupos, tendo o primeiro agrupamento como principal característica maior concentração de nitrato e, o terceiro, possuindo maior concentração de sais inorgânicos; resultados semelhantes ao da ACP.

Zimmermann et al. (2008) empregam a técnica de Análise Componentes Principais (ACP) para avaliar a qualidade da água do rio Tibagi, no município de Ponta Grossa - PR. Coletam-se dados de 14 parâmetros físico-químicos e microbiológicos em três pontos de amostragem distintos, sendo o primeiro em uma região

de intensa atividade agrícola e pecuária (ponto 1), o segundo na jusante do distrito industrial (ponto2) e o último na primeira estação de monitoramento do rio Tibagi (ponto3), nos anos de 2005 e 2006. A técnica do ACP foi considerada para cada um dos pontos amostrais e os resultados são apresentados pelos dois primeiros componentes principais por meio do gráfico *biplot* (eixo das abscissas correspondente ao componente de maior contribuição (PC1) e eixo das ordenadas, ao segundo maior (PC2)). Esses dois componentes conjuntamente permitem representar 66% da variabilidade dos dados e o ponto representado pelo jusante do distrito industrial é o que apresenta maior grau de deterioração da qualidade da água.

Sojka et al. (2008) analisam inicialmente 19 parâmetros físico-químicos, medidos mensalmente em oito diferentes pontos, nas águas do rio Mala Welná Polônia, entre os meses de maio e novembro de 2006. O objetivo da pesquisa é utilizar as técnicas de análise multivariada (Análise Fatorial (AF), Análise de Componentes Principais (ACP), Análise Discriminante (AD) e Análise de Agrupamentos (AA)) a fim de descobrir similaridades e diferenças na constituição físico-química na composição da água; identificar indicadores de qualidade da água de acordo com sua variabilidade espacial e temporal; encontrar fatores ocultos na estrutura de dados e procurar fontes artificiais de poluição. A AF gera quatro fatores que são caracterizados pelos elementos das águas do rio, conseguindo também descrever a sua qualidade com explicação de 73% da estrutura dos dados. A ACP é representada por dois componentes que conjuntamente retêm aproximadamente 56% da variabilidade dos dados iniciais. A AD construiu duas funções com 14 variáveis para conseguir caracterizar a distribuição espacial e temporal dos dados. A AA, utilizando a método hierárquico de agrupamento de Ward, separou os pontos em três grupos distintos quanto à qualidade da água.

Boyaciglu & Boyaciglu (2008) utilizam as técnicas de análise multivariada, dentre elas a Análise de Componentes Principais (ACP) e Análise Fatorial (AF) para descrever a qualidade da água e as principais fontes de poluição da bacia do rio Tahtali, na Turquia. Os dados foram obtidos em sete estações distintas, ao longo

de seis anos e 10 parâmetros relacionados à água são coletados. A AF seleciona três fatores significativos para a interpretação dos dados, os quais explicam aproximadamente 71% da variabilidade do conjunto de dados inicial. Já a ACP seleciona três componentes lineares que são obtidos tomando os autovalores com valor superior a um, conjuntamente com o método gráfico do *Scree-plot*. Tanto a AF quanto a ACP concluem que a qualidade da água do rio Tahtali é controlada principalmente pela poluição agrícola.

Li et al. (2009) investigam as variações temporais e espaciais da qualidade da água na bacia do rio Shongua, na China. Para tal avaliação é necessário que haja a coleta de 14 parâmetros monitorados entre os anos de 1998 e 2004, em 14 locais diferentes ao longo do rio. As técnicas de estatística multivariada, incluindo a Análise de Agrupamento (AA), Análise de Componentes Principais (ACP), Análise de Fatores (AF) e Análise Discriminante (AD) são utilizadas na investigação. Com AA foi possível detectar três grupos distintos, indicando que cada grupo tem uma qualidade diferente de água de acordo com o grau de poluição. A ACP e AF selecionam cinco componentes lineares que descrevem os parâmetros responsáveis pela variação da qualidade da água, sendo que estes são obtidos, considerando apenas os autovalores superiores a um e acumulam 86,74% da variância total no conjunto de dados originais. Já a AD seleciona apenas os parâmetros: pH, temperatura, oxigênio dissolvido, nitrogênio amoniacal, nitrogênio nitrato. Para descrever a variação temporal dos dados e as variáveis: pH, temperatura, oxigênio dissolvido, demanda bioquímica de oxigênio, nitrogênio amoniacal, nitrogênio nitrato, fenóis voláteis e total arsênico para descrever a variação espacial dos dados. O estudo revela que as principais causas de deterioração da qualidade da água são a eliminação de águas residuais domésticas e industriais.

Palácio et al. (2009) realizam coletas georreferenciadas em sete pontos distintitos, na bacia do Curu - CE, nos meses de janeiro, março, maio e agosto do ano de 2005, com o objetivo de identificar grupos similares nas variáveis responsáveis pela qualidade da água. Na busca do objetivo, utilizam a técnica multivariada da

Análise de Agrupamentos via o método aglomerativo de Ward e como medida de semelhança a distância euclidiana ao quadrado. Para definir o número de agrupamentos similares utilizou-se o “coeficiente de aglomeração” o qual indica que a presença de saltos repentinos no valor do coeficiente tornando o agrupamento menos homogêneo. Desta forma consegue-se identificar quatro grupos homogêneos, independentes da posição geográfica, demonstrando maior influência da sazonalidade das estações secas e chuvosas.

Zhao & Cui (2009) têm como objetivo avaliar a qualidade dos recursos hídricos do reservatório de superfície do rio Luan na China, pelo emprego de técnicas multivariadas, como a Análise de Agrupamento (AA) e Análise Fatorial (AF) em 10 parâmetros que são selecionados por estarem presentes em todas as amostras. A Análise de Fatores foi elaborada a partir da matriz de dados padronizada, com o uso da distância euclidiana ao quadrado, como medida de parença, e com o método hierárquico de agrupamento de Ward. Os resultados da análise conseguem separar os meses do ano em três grupos distintos de acordo com as características sazonais do norte da China. A AF resulta em combinações lineares dos parâmetros selecionados, envolvendo três fatores, sendo o primeiro relacionado a atividades antropogênicas, o segundo a características naturais das águas de superfície e, o último fator, ligado a atividades minerais e agrícolas.

Mustapha & Nabegu (2011) têm como objetivo identificar fatores naturais e antrópicos na qualidade da água do rio Getsi na Nigéria, utilizando a técnica multivariada de Componentes Principais. Neste sentido, coletou-se 30 amostras em oito pontos distintos, selecionados aleatoriamente, totalizando 240 amostras referentes a 15 parâmetros físico-químicos e biológicos mensurados na estação seca do ano de 2011. Utilizando o critério de Kaiser e o método gráfico do *Scree-plot*, cinco componentes são selecionados para a identificação, os quais explicam aproximadamente 82% da variância total. O primeiro componente retém aproximadamente 37% da variância total, e é formado exclusivamente por metais pesados, sendo estes os principais responsáveis pela variação da qualidade da água.

Lima et al.(2011) analisam 23 parâmetros físico-químicos e biológicos (pH, turbidez, cor aparente, coliformes totais, ferro, temperatura, entre outros) em 26 amostras de poços na Bacia Sedimentar Costeira do Baixo Curso do rio Paraíba – PB tendo como objetivo agrupar os poços segundo suas semelhanças, sabendo a princípio que estes diferem entre si em relação à sua geologia e sua intensidade de utilização. Para a construção de grupos homogêneos, aplicou-se a Análise de Agrupamento utilizando como medida de similaridade a distância euclidiana média e o método de agrupamento de Ward. Desta forma, quatro grupos homogêneos são formados, sendo o primeiro grupo composto exclusivamente por poços particulares que apresentam menor concentração de pH; o segundo, em sua maioria, poços pertencentes à Companhia de Água e Esgoto do Estado da Paraíba com alta condutividade elétrica, maior salinidade, dureza e concentração de cálcio; o terceiro, formado exclusivamente por poços particulares com altos valores para cor, turbidez, alcalinidade e potássio, possuindo também alguns poços com contaminação fecal com altos valores de coliformes totais e de *E.coli* e, por fim, o quarto, e último, grupo constituído por um único poço particular que apresenta a maior contaminação fecal entre todos os poços considerados.

Netto et al. (2012) utilizam análise multivariada envolvendo a técnica da análise fatorial por meio dos Componentes Principais para investigar as características físicas e químicas da água do rio Poxim – SE, responsável pelo abastecimento do município de Aracajú. Objetivando descrever a qualidade da água, foram realizadas quatro amostragens entre os anos de 2005 e 2006, em 15 pontos distintos, distribuídos ao longo do rio. Os parâmetros utilizados na pesquisa envolvem condutividade, turbidez, cor, sólidos totais dissolvidos, oxigênio dissolvido, alcalinidade, dureza, clorofila-a e vários nutrientes. O critério do KMO (Kaiser-Meyer-Olkin) indica cinco componentes para compor a análise fatorial a partir da matriz de correlação dos dados, composta de 21 variáveis. Os cinco componentes explicam conjuntamente 81,73% da variação total dos dados e permitem concluir a respeito de vários aspectos relacionados à qualidade da água do rio.

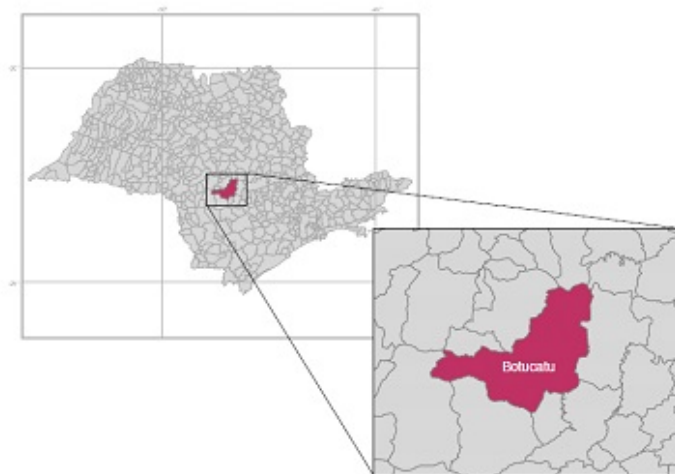
Guedes et al. (2012) utilizam 18 variáveis para avaliar a qualidade de água do rio Pomba – MG, coletadas duas vezes em nove pontos de monitoramento distintos no período de outubro/2008 e janeiro/2009. Por meio de técnicas multivariadas utilizadas a partir da matriz de correlação dos dados, especificamente a Análise Fatorial e a Análise de Componentes Principais promovem uma redução do número de variáveis usadas no controle de qualidade de água de 18 para 15. Empregando-se como o critério de seleção do número de componentes, o método proposto por Jolliffe, que sugere a porcentagem acumulada da variância total esteja entre 70% e 90%, seleciona - se três fatores que explicam conjuntamente 74,3% da variância total. A escolha dos três primeiros componentes principais permite concluir que os principais responsáveis pela alteração da qualidade da água deve - se ao esgoto doméstico, compostos nitrogenados, material orgânico e resíduos sólidos deixados inadequadamente às margens do curso do rio.



## 3 MATERIAIS E MÉTODOS

### 3.1 Caracterização da área de estudo

A área de realização do estudo foi o município de Botucatu, no estado de São Paulo, que dista 235 quilômetros da capital, localizado na Região Centro-sul do estado, com  $-22,88^{\circ}$  de latitude e  $-48,44^{\circ}$  de longitude e aproximadamente  $1483\text{km}^2$  de território (Figura 1). Sua população em 2010, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE) era de 127328 habitantes. O município é drenado pelo rio Tietê ao norte e, pelo rio Pardo, ao sul.



Fonte: Nascimento (2010)

Figura 1 - Mapa município de Botucatu.

A Sabesp se caracteriza como uma empresa de economia mista, é a responsável pelos serviços de água e esgoto no município, sendo Botucatu o primeiro

município do interior do Estado de São Paulo, a ser operacionalizada pela Sabesp, em novembro de 1974. Tal fato, fez com que Botucatu se tornasse sede da Unidade de Negócios do Médio Tietê com um dos 16 laboratórios de controle sanitário distribuídos em todo território paulista. A Sabesp atingiu em Botucatu a despoluição de todos os córregos que cortam a cidade e se aproxima de ser uma “Cidade 300%” – 100% água tratada, 100% de coleta de esgoto e 100% de esgoto tratado em um futuro muito próximo.

### **3.2 Amostragem e pontos de coleta**

No município de Botucatu o serviço de distribuição de água aos consumidores possui em sua rede 148 pontos para a coleta de várias amostras realizadas diariamente, de forma sistemática nesses pontos fixos de amostragem (em geral, cavaletes de residências). Em alguns casos, este procedimento de coleta também pode ser realizado por meio de solicitação dos usuários por motivos diversos (essa situação não fez parte do estudo). Para o desenvolvimento da pesquisa sobre a caracterização do perfil geográfico da qualidade da água distribuída no município, foram consideradas 4611, coletas realizadas nos 148 pontos durante o período de 2007 a 2011. Optou-se pelo período anteriormente citado, pois no ano de 2012 houve a mudança na portaria vigente no país e nos equipamentos utilizados para a aferição dos parâmetros pela Sabesp foram substituídos após este período, sendo inviável a comparação dos resultados das análises.

As coletas realizadas foram submetidas a análises físico-químicas e bacteriológicas na Divisão de Controle Sanitário da Sabesp, localizada em Botucatu – SP. Os dados resultantes das análises foram armazenados na planilha eletrônica do programa Microsoft Office Excel 2007, sob o título de DADOSSABESP.xls para posterior uso dos procedimentos de análise estatística

### 3.3 Parâmetros envolvidos no estudo

Para reconhecer a grandeza da qualidade da água, utilizando os resultados das análises dos parâmetros físico-químicos e bacteriológicos das amostras coletadas em acordo com a Portaria nº 518 do Ministério da Saúde e utilizadas na Divisão de Controle Sanitário da Sabesp, foram considerados os seguintes parâmetros físico-químicos: cor aparente, turbidez, flúor, cloro residual e potencial hidrogeniônico (pH), e os bacteriológicos: contagem de bactéria heterotrófica e ocorrência de coliforme total. A seguir estão descritos, em detalhes, os parâmetros do estudo, bem como o respectivo procedimento de mensuração e o equipamento utilizado para sua avaliação. Os parâmetros para o estudo da qualidade da água distribuída foram mensurados a partir de cada coleta ao longo do período envolvido na pesquisa.

#### 3.3.1 Parâmetros físico - químicos

##### Cor Aparente

A cor verdadeira da água é caracterizada pela água com a turbidez removida por meio da filtração. Já a cor aparente (**corap**) é determinada pela decomposição de materiais orgânicos, principalmente colóides e inorgânicos, dissolvidos em água (Macêdo, 2004). O método espectrofotométrico foi o empregado para medir a cor, e o equipamento utilizado para a mensuração foi colorímetro/ turbidímetro Hach modelo 2100 AN cujos valores abaixo de 4 unidades de cor não são detectados na aferição. A unidade de medida usual para cor aparente é chamada unidade de cor (UC).

##### Turbidez

A turbidez (**turb**) é medida pela alteração da penetração da luz na presença de sólidos em suspensão, podendo a mesma ser absorvida ou dissipada, tal fato é ocasionado pela presença de partículas em suspensão existentes na água, alterando sua transparência. Essas partículas podem ser silte, argila, algas, bactérias,

pâncton, entre outros (Macêdo, 2004). Os sólidos em suspensão podem, eventualmente, criar um meio de fixação para os microrganismos, formando um efeito escudo que dificulta a desinfecção da água (Brasil, 2013). A turbidez foi medida por meio do método nefelométrico, o qual é considerado um método mais sensível, e o equipamento utilizado foi o colorímetro/turbidímetro da marca Hach, modelo 2100 AN. A variável turbidez tem como métrica a unidade nefelométrica de turbidez (UNT).

#### Flúor

O flúor (**fluor**) é adicionado à água por sua comprovada eficácia na redução de cáries dentárias, porém se adicionado em excesso pode causar malefícios aos dentes e aos ossos (EATON et al., 2012). Para mensuração da concentração de flúor na água, o método potenciométrico - eletrodo íon-seletivo foi utilizado, e o equipamento usado no processo de mensuração foi o pHmetro da marca Orion modelo 720, cujo teor de flúor é medido em miligrama por litro (mg/L).

#### Cloro residual

Dado que a água está sujeita à contaminação por bactérias e organismos vivos, o cloro é adicionado à mesma para que haja a desinfecção por motivo de possíveis causas patogênicas (EATON et al., 2012). O cloro quando adicionado à água, destrói os microrganismos e oxida o material orgânico existente, porém existe uma quantidade de cloro que permanece na água para assegurar sua desinfecção durante a distribuição da na rede. O cloro residual (**cloroR**) é a quantidade de cloro que deixou de ser utilizada no processo de destruição das bactérias. Para a mensuração do cloro residual livre, foi empregado o método conhecido como colorimétrico-DPD para a análise da quantidade de cloro, cuja medida é realizada em miligramas por litro (mg/L) obtida por meio de um colorímetro AquaColor Policontrol.

#### Potencial hidrogeniônico

O potencial hidrogeniônico (**pH**), usualmente conhecido como pH, é indicado pela concentração de íons hidrogênio numa amostra de água, sendo assim

possível avaliar se a água amostrada se encontra ácida ou alcalina (Nelson & Cox (2006)). O pH é bastante monitorado, pois caso se encontre próximo de nove pode causar problemas na rede de distribuição como o entupimento das tubulações e o pH encontrado for baixo pode ocorrer a corrosão das tubulações. O método eletrométrico foi o responsável pela medição do pH, conjuntamente com o pHmetro portátil da marca Oakton.

### 3.3.2 Parâmetros bacteriológicos

#### Bactéria heterotrófica

As bactérias heterotróficas (**bact**) são microrganismos que utilizam compostos orgânicos existentes em seu meio de cultura para produzir unidades formadoras de colônia (UFC) (Brasil, 2005). Para quantificar as bactérias heterotróficas utilizou-se o método do Pour Plate, e sua escala de medida foi dada pela unidade formadora de colônia (UFC) por mililitro.

#### Coliforme total

O coliforme total (**colif**) é caracterizado por grupos *gram* negativos em forma de bastonetes que podem ou não necessitar de oxigênio, não são formadores de esporos e possuem capacidade de fermentar a lactose, produzindo gás. O grupo dos coliformes possui algumas espécies, destacando-se como as principais *Escherichia*, *Klebsiella*, *Enterobacter* e *Citrobacter* (Brasil, 2005). Para verificar a presença ou ausência dos coliformes foi usado o método da membrana filtrante, e a escala de medidas foi dada pelo número mais provável (NMP), caso houvesse a presença de coliformes.

Nesta pesquisa, a variável em questão foi transformada em uma dicotomia de ocorrência de coliformes, sendo a resposta positiva indicada quando atingia valores sensíveis ao método, independente da quantidade mensurada, caso contrário, o ponto de coleta teria como resultado a negatividade da presença de coliformes.

A partir da dicotmização foi estabelecida a porcentagem média de positividade da ocorrência de coliformes nas coletas realizadas em cada um dos pontos.

### 3.4 Conjunto de dados

A partir das informações numéricas dos parâmetros coletados diariamente pela Sabesp referentes aos cinco anos estudados, foi necessária a construção de um banco de dados organizado de acordo com os objetivos da pesquisa e que fosse compatível para posterior procedimento de análise estatística.

No processo de coleta das informações numéricas sobre os parâmetros em estudo, em alguns casos há falta de valores pela não determinação quando da análise laboratorial. Essas situações ocorrem por falta de interesse técnico do conhecimento do valor da variável no momento em consideração. Nestes casos, adotou-se como procedimento para o preenchimento da lacuna do valor faltante, a técnica de estimação por ponto de uma medida de tendência central (média ou mediana), conforme o tipo de variável em apreciação. No caso dos parâmetros turbidez, flúor, cloro residual, pH e coliformes totais tomou-se como estimativa pontual a média, nas demais (cor aparente e bactéria heterotrófica), a mediana. A indicação destas medidas deveu-se à inspeção da dispersão dos dados, ou seja, quando a variável apresentou alta variabilidade, optou-se pela mediana, caso contrário, pela média.

Inicialmente, os vetores de repostas apresentavam-se organizados de acordo com as datas em que foram realizadas as coletas de água, porém se configurou os dados de forma que estes se unissem nos pontos de coleta, conforme sistemática de serviço da Sabesp na alocação no município (todos os pontos são fixos e identificados por um número). Neste sentido, obteve-se uma matriz de dados organizada pelos pontos de coleta e suas correspondentes repetições de amostras. Em seguida procurou-se construir um único vetor de repostas para cada ponto de coleta, ou seja, o número de vetores repostas da matriz de dados foi igual ao número de pontos de coleta existentes no município. Estes vetores de repostas formados em correspondência com os pontos de coleta foram obtidos utilizando-se medidas

de tendência central para cada variável, aproveitando o fato dos pontos exibirem repetições de coletas (amostragens).

Para construir o vetor de resposta operacional, para o interesse da pesquisa, a partir das amostragens realizadas nos pontos de coleta, considerou-se como unidade amostral a representação caracterizada pelo bairro das coletas. Além do interesse próprio do estudo, acrescentou-se como fator motivante à utilização do bairro como vetor de referência de coleta de dados, a instabilidade existente no número de amostragens realizadas nos diversos pontos de coleta. Dessa forma, quando existia somente um ponto de coleta no bairro, este passava a ser o vetor característico de resposta do bairro; se existisse mais de um ponto, era utilizada a medida de tendência central mais adequada a cada parâmetro, para assim ficar apenas com um único vetor característico operacional por bairro.

Nesta sistematização operacional cada bairro fica referenciado por um único vetor de resposta, o qual pode ser representado por uma única coleta realizada em ponto solitário ou uma medida obtida por indicadores de centralidade de várias coletas e até de vários pontos. O conjunto de dados do estudo ficou assinalado por vetores de bairros, contendo os parâmetros avaliados que pode ser organizado matricialmente da seguinte forma:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{x}'_1 \\ \vdots \\ \mathbf{x}'_n \end{bmatrix}$$

A ordem da matriz  $\mathbf{X}$  é  $n \times p$ , na qual o elemento genérico  $x_{ij}$  representa a  $j$ -ésima característica estudada da água no  $i$ -ésimo bairro, com  $i = 1, 2, 3, \dots, n$  e  $j = 1, 2, 3, \dots, p$ . A representação vetorial para o bairro  $i$  pode ser expressa pelo vetor  $\mathbf{x}'_i = [\mathbf{x}_{i1}, \dots, \mathbf{x}_{ip}]$ . Na estrutura multivariada para o estudo, os vetores linhas correspondem aos bairros e são considerados sob a condição de independência e os vetores colunas, concernentes aos parâmetros avaliados simultaneamente nas amostras,

como variáveis dependentes.

Nesta linha de abordagem, o banco de dados para os procedimentos exploratórios e analíticos, na busca da construção de um perfil geográfico da água passou a ter 76 linhas nas quais configuram as informações numéricas de todos os parâmetros avaliados em cada bairro e com sete colunas correspondentes ao número total de parâmetros considerados na Portaria do Ministério da Saúde nº 518 - Controle e Vigilância da Qualidade da Água para Consumo Humano envolvidos no estudo (nomeados para os procedimentos estatísticos como variáveis).

A distribuição geográfica desses pontos de referência dos bairros no município de Botucatu fica conforme Figura 2.



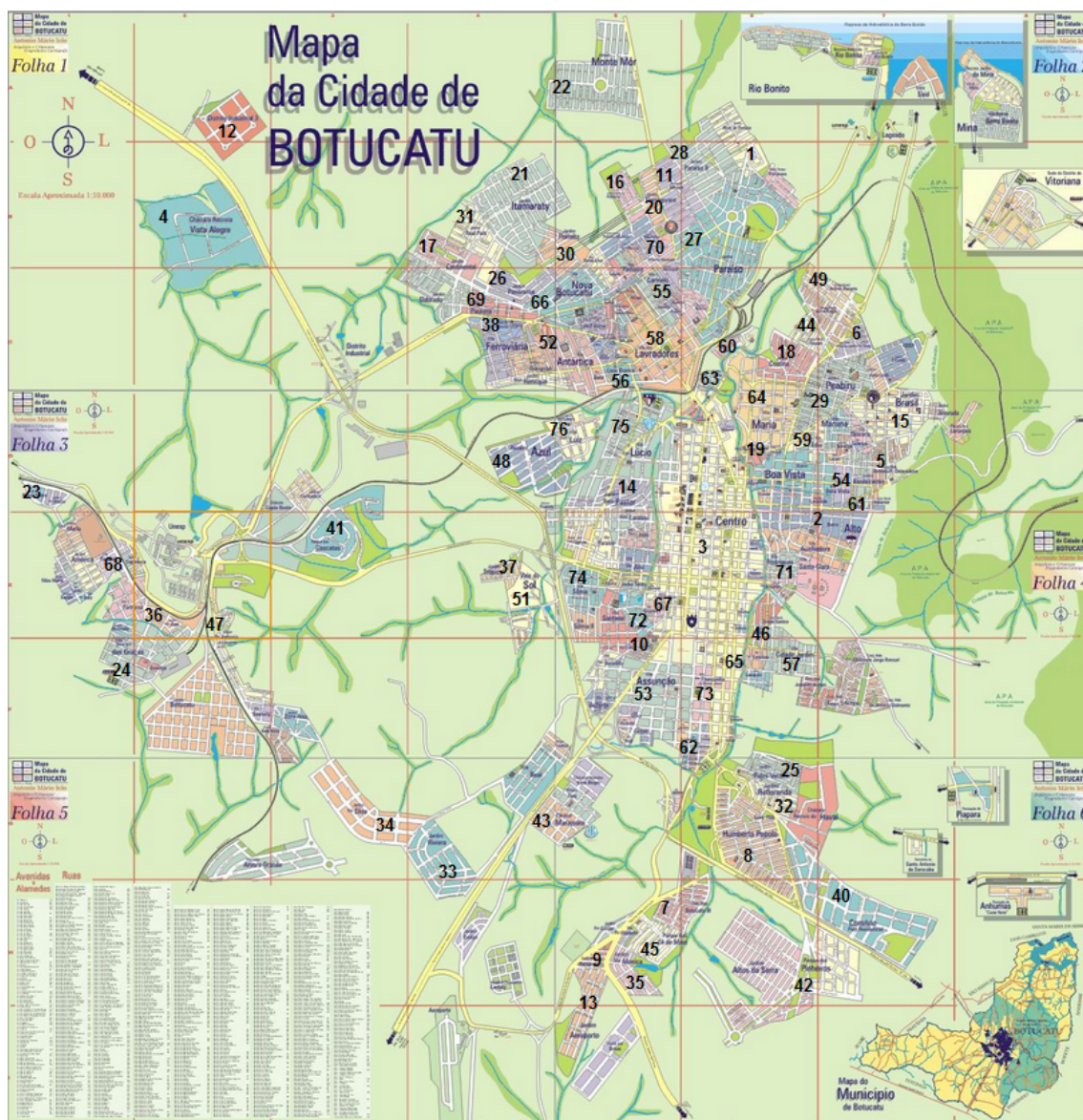


Figura 2 - Distribuição dos pontos de coleta (bairros).

Na Tabela 2 são apresentados os nomes dos bairros com os respectivos códigos de alocação no mapa apresentado na Figura 2.

Tabela 2: Distribuição dos pontos de coleta (bairros) no município de Botucatu

<b>Código</b>	<b>Bairro</b>
1	Altos do Paraíso
2	Bairro Alto
3	Centro
4	Chácara Recreio Vista Alegre
5	Conjunto Habitacional Antônio Hermínio Delevedove
6	Conjunto Habitacional Arnaldo Leotta de Mello
7	Conjunto Habitacional Engenheiro Francisco Blasi
8	Conjunto Habitacional Humberto Popolo
9	Conjunto Habitacional José Antonio Lungo
10	Conjunto Residencial Frei Fidelis
11	Conjunto Residencial Jardim do Mirante
12	Domingos Lopes
13	Jardim Aeroporto
14	Jardim Bom Pastor
15	Jardim Brasil
16	Jardim Chácara dos Pinheiros
17	Jardim Continental
18	Jardim Cristina
19	Jardim Dona Nicota de Barros
20	Jardim Flamboyant
21	Jardim Itamarati
22	Jardim Monte Mor
23	Jardim Neusa Maria (Rubião Junior)
24	Jardim Nossa Senhora das Graças (Rubião Junior)
25	Jardim Palos Verdes

26	Jardim Panorama
27	Jardim Paraíso
28	Jardim Paraíso II
29	Jardim Peabiru
30	Jardim Planalto
31	Jardim Real Park
32	Jardim Reflorenda
33	Jardim Riviera
34	Jardim Santa Eliza
35	Jardim Santa Mônica
36	Jardim São José (Rubião Junior)
37	Jardim Tropical
38	Jardim Universitário
39	Jardim Ypê
40	Park Residencial Convívio
41	Parque das Cascatas
42	Parque dos Pinheiros
43	Parque Marajoara
44	Parque Residencial Nazaré
45	Parque Residencial Vinte e Quatro de Maio
46	Parque Santa Inês
47	Parque Santo Antonio da Cascatinha (Rubião Junior)
48	Recanto Azul
49	Residencial Arlindo Durante
50	Residencial Cedro
51	Vale do Sol
52	Vila Antártica
53	Vila Assumpção
54	Vila Bela Vista

55	Vila Carmelo
56	Vila Casa Branca
57	Vila Cidade Jardim
58	Vila dos Lavradores
59	Vila Édén
60	Vila Ema
61	Vila Eny
62	Vila Guimarães
63	Vila Jahu
64	Vila Maria
65	Vila Moreira
66	Vila Nova Botucatu
67	Vila Padovan
68	Vila Paraíso (Rubião Junior)
69	Vila Paulista
70	Vila Pinheiro Machado
71	Vila Rodrigues Alves
72	Vila Santa Luzia
73	Vila Santa Therezinha de Menino Jesus
74	Vila São Judas Thadeu
75	Vila São Lúcio
76	Vila São Luiz

---

Apesar de participarem da análise os bairros Jardim Ypê (39) e Residencial Cedro (50) não foram localizados na Figura 2 por se tratarem de bairros novos no município de Botucatu, porém se destaca que o Bairro Jardim Ypê fica em proximidade com os bairros Jardim Continental (17), Jardim Itamarati (21) e Jardim Planalto (30), enquanto o Residencial Cedro, com o Conjunto Habitacional

José Antônio Lungo (9), Jardim Aeroporto (13) e Jardim Santa Mônica (35).

Antes de iniciar a análise estatística dos dados, considerou-se que todos os parâmetros envolvidos na pesquisa pudessem participar de forma igual em magnitude, desprovidos da unidade de mensuração, ou seja, para que o uso de diferentes escalas de medidas não ocultasse informações importantes no procedimento analítico. Então, decidiu-se padronizar as variáveis para que seus valores se tornassem adimensionais. Para a padronização das variáveis considerou-se o procedimento convencional e usual que do valor original deve ser subtraído a média do próprio conjunto de dados e este resultado dividido pelo respectivo desvio padrão. O novo valor, denominado,  $z_{ij}$  está apresentado na equação (1)

$$z_{ij} = \frac{x_{ij} - \bar{x}_{.j}}{\sqrt{s_j}}. \quad (1)$$

Para o entendimento prático da padronização, o valor de  $z_{ij}$  indica o número de desvios padrão ( $s_j$ ) que o valor original ( $x_{ij}$ ) se afasta da média ( $\bar{x}_{.j}$ ) dos respectivos dados.

A partir deste momento a nova configuração do conjunto de dados do estudo fica caracterizada pela seguinte matriz:

$$\mathbf{Z}_{ij} = \begin{bmatrix} z_{11} & \cdots & z_{1p} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{np} \end{bmatrix} = \begin{bmatrix} \mathbf{z}'_1 \\ \vdots \\ \mathbf{z}'_n \end{bmatrix}$$

Com a padronização dos dados, as unidades de referência (bairros) para o estudo da qualidade da água definidas e as respectivas localizações geográficas no município de Botucatu têm-se condições de utilizar as técnicas de análise multivariada de dados. Essas técnicas são aplicadas às situações em que há interesse de analisar globalmente um conjunto de dados com muitas variáveis relacionadas entre si e que o procedimento analítico permita a compreensão total da estrutura geral da

variabilidade (intravariabilidade e intervariabilidade) dos dados.

Para estas situações a Análise Multivariada de dados pode ser dividida em técnicas de dependência e de interdependência. As técnicas de dependência estudam variáveis que dependam de uma ou mais variáveis para serem explicadas (variáveis independentes), destacando-se entre elas: Análise Discriminante, Análise de Variância, Análise de Medidas Repetidas, Análise de Regressão Múltipla e Análise de Correlação Canônica. Já as técnicas de interdependência não identificam uma variável ou um grupo de variáveis como sendo independentes ou dependentes, ou seja, analisam simultaneamente todo o conjunto de variável. São consideradas técnicas de interdependência a Análise de Agrupamento, Análise de Componentes Principais e a Análise Fatorial.

A escolha do método multivariado a ser utilizado deve ser definido de acordo com o tipo de situação que se tem para abordar (Johnson & Wichern, 2007). Em particular, neste estudo, foram utilizadas duas técnicas de interdependência: a Análise de Agrupamento e a Análise de Componentes Principais.

### **3.5 Análise de Agrupamento**

A Análise de Agrupamento é uma técnica de estatística multivariada, cujo objetivo é agrupar unidades amostrais, utilizando algum critério preestabelecido, de forma que o conjunto inicial de dados fique dividido em um ou mais grupos. Estes grupos devem possuir alta homogeneidade interna (dentro do grupo) e alta heterogeneidade externa (entre os grupos) segundo as respostas das variáveis. Para dar início ao processo da análise de agrupamento, considerou-se como vetor de referência os bairros (unidades amostrais) avaliados sobre as sete variáveis estudadas e, na sequência, houve a necessidade de se construir uma matriz que apresentasse as informações quantitativas de quão semelhantes são duas unidades amostrais (bairros). Para a definição desta matriz de semelhanças foi necessário escolher uma medida de parença. Como os vetores de resposta referentes aos bairros eram constituídos por variáveis quantitativas, dois tipos de medida podem ser utilizadas: as medidas de

similaridade ou a de dissimilaridade.

Nesse caminho, quando a matriz de semelhança é constituída pela correlação linear simples entre as variáveis, tem-se uma medida de similaridade e quanto mais próximo o seu valor estiver da unidade, mais semelhantes são os bairros; já, para as medidas baseadas em distâncias têm - se as chamadas dissimilaridade nas quais os valores mais próximos de zero indicam maior semelhança entre os bairros. Na literatura, pode-se destacar como a medida de dissimilaridade mais utilizada para variáveis quantitativas a distância euclidiana, a qual foi a escolhida para ser utilizada no presente estudo.

A distância euclidiana constitui-se em uma métrica cujo valor para dois bairros quaisquer  $l$  e  $k$  é definida como a raiz quadrada positiva do somatório dos quadrados das diferenças entre todas as variáveis mensuradas nos vetores de resposta dos bairros. Sendo assim, se  $\mathbf{z}_l$  e  $\mathbf{z}_k$  são os vetores respostas dos bairros  $l$  e  $k$   $l \neq k$ , a distância entre  $l$  e  $k$  é dada pela equação (2)

$$d_{lk} = [(\mathbf{z}_l - \mathbf{z}_k)'(\mathbf{z}_l - \mathbf{z}_k)]^{\frac{1}{2}} = \left[ \sum_{j=1}^p (z_{lj} - z_{kj})^2 \right]^{\frac{1}{2}}. \quad (2)$$

Também fazem parte das medidas de distância para a similaridade as métricas de Minkowski, Camberra, euclidiana média, euclidiana padronizada, euclidiana ponderada e *city-block*.

Não obstante, para casos em que as variáveis estudadas são qualitativas, desenvolveu-se também alguns coeficientes de parença capazes de comparar duas unidades amostrais. Tais coeficientes verificam se uma unidade amostral quando comparada a outra possui ou não atributos em comum e, quanto mais itens em comum apresentam, mais similares as unidades são entre si. As medidas de parença para variáveis qualitativas mais usuais são os coeficientes de: concordância simples, Jaccard, concordância positiva, Sorenson e Sokal e Sneath.

Há algumas situações em que as variáveis que compõem o estudo se

apresentam parte de forma quantitativa e outra, de forma qualitativa. Entre as medidas que avaliam situações onde esses dois tipos de variáveis estão presentes a distância de Gower tem sido a mais utilizada. Vale lembrar que o uso de diferentes medidas de parença produzem diferentes configurações de agrupamento.

Após definir a medida de parença no presente estudo, obteve-se a matriz de distâncias, utilizando a métrica euclidiana e encaminhou-se para a etapa seguinte do procedimento que consiste na construção dos agrupamentos, cujo objetivo baseia-se em escolher qual o método de agrupamento será utilizado para a formação de grupos homogêneos.

Os métodos de agrupamentos são divididos em dois tipos: hierárquico e não hierárquico. O método hierárquico por sua vez, se fraciona em dois grupos, os divisivos e os aglomerativos.

O agrupamento hierárquico divisivo executa o processo com todas as unidades amostrais, pertencendo a um único grupo; posteriormente este grupo inicial é separado em dois subgrupos que sejam os mais distintos possíveis. Tal procedimento é realizado até que cada grupo seja formado por apenas uma única unidade amostral.

No agrupamento hierárquico aglomerativo, o processo inicia como se cada bairro fosse um agrupamento, então se identifica quais são os bairros mais semelhantes entre si, e estes passam a formar um novo grupo. Este processo é feito até que todas as unidades pertençam a um único agrupamento. Os principais representantes deste tipo de agrupamento são o método da ligação simples (método do vizinho mais próximo), método da ligação completa (método do vizinho mais distante), método do centróide e método de Ward ao qual se deu maior destaque por ter sido utilizado no trabalho.

O método proposto por Ward, também conhecido como o método da “Variância Mínima”, desenvolvido em 1963, considera como critério para agregar indivíduos a variância (Mingoti, 2007). Cada etapa do processo é analisada a perda de informação resultante da criação de um novo agrupamento, ou seja, utiliza a soma de quadrados da distância euclidiana como medida de homogeneidade e seleciona a



menor soma de quadrados obtida, formando assim, agrupamentos o mais homogêneos possível. Desta forma, cada novo agrupamento formado é resultado do menor valor de soma de quadrados da distância euclidiana encontrada.

Pode-se, resumir o método de Ward, segundo Knupp (2007) e Mingoti (2007) nos seguintes passos:

1. O processo se inicia com cada unidade amostral (bairro), formando um grupo.
2. A cada etapa do processo de formação dos grupos é determinada a soma de quadrados da distância euclidiana de cada unidade amostral (bairro) pertencente ao conglomerado em relação ao correspondente vetor de média do próprio conglomerado,

$$SS_{(i)} = \sum_{j=1}^{n_i} (\mathbf{z}_{(i)j} - \bar{\mathbf{z}}_{(i)})' (\mathbf{z}_{(i)j} - \bar{\mathbf{z}}_{(i)})$$

em que:

$n_i$  = número de bairros no conglomerado  $C_i$ ,

$\mathbf{z}_{(i)j}$  = vetor de observações da  $j$  - ésimo bairro pertencente ao  $(i)$  - ésimo conglomerado,

$\bar{\mathbf{z}}_{(i)}$  = vetor de médias do conglomerado  $C_i$ ,

$SS_{(i)}$  = soma de quadrados da distância euclidiana correspondente ao conglomerado  $C_i$ .

3. Em seguida, considerando o algoritmo no passo  $k$ , calcula-se o somatório das somas de quadrados das distâncias euclidianas de todos conglomerados do agrupamento,

$$SSR = \sum_{i=1}^{g_k} SS_{(i)}$$

em que:

$g_k$  = número de conglomerados no passo  $k$ .

4. Na sequência determina-se a distância entre os agrupamentos  $C_i$  e  $C_l$ , dada por:

$$d(C_i, C_l) = \left[ \frac{n_i n_l}{n_i + n_l} \right] (\bar{\mathbf{z}}_{(i)} - \bar{\mathbf{z}}_{(l)})' (\bar{\mathbf{z}}_{(i)} - \bar{\mathbf{z}}_{(l)})$$

em que:

$n_i$  e  $n_l$  = número de bairros dos respectivos conglomerados,

$(\bar{\mathbf{z}}_{(i)})$  e  $(\bar{\mathbf{z}}_{(l)})$  = centróides dos bairros dos respectivos conglomerados.

5. A cada passo do algoritmo de agrupamento os dois conglomerados que apresentam a menor distância são unidos.

Definido e realizado o método de agrupamento, pode-se apresentar tais resultados por um diagrama bidimensional que ilustra todos os passos do procedimento (de agrupamento de bairros) chamado de dendrograma ou dendograma, no qual o eixo das abscissas são apresentadas as unidades amostrais dos agrupamentos (bairros) e no eixo das ordenadas a distância em que os agrupamentos foram unidos. O dendograma é utilizado como um método gráfico que serve para indicar subjetivamente pela inspeção visual o número final de agrupamentos. O dendograma pode ser utilizado tanto no método hierárquico divisivo quanto no aglomerativo.

Um dos pontos fundamentais para a análise de agrupamento hierárquico consiste na chamada regra da parada, ou seja, estabelecer o número de grupos que serão utilizados na análise. Para isso, na literatura, já existem critérios que dão suporte para a tomada de decisão quanto ao número de grupos a serem formados. Tais métodos vão desde uma simples inspeção gráfica até critérios mais diretos utilizando algum embasamento mais profundo de estatística, que permite de forma objetiva indicar o número de grupos resultantes para o estudo.

Dentre os métodos que possibilitam a interação entre o interesse biológico e o embasamento estatístico estão a Análise do Comportamento do Nível de Fusão (distância) e a Análise do Comportamento do Nível de Similaridade os quais foram adotadas para a definição do número de grupos desta pesquisa.

A Análise do Comportamento do Nível de Fusão (distância) examina o nível de fusão dado pela distância existente entre cada etapa sucessiva do agrupamento. Quando a similaridade entre os grupos diminui consideravelmente, o nível de fusão aumenta e indica que o novo grupo formado provocou uma queda expressiva na similaridade, portanto, o agrupamento anterior à queda deve ser o escolhido. Uma maneira interessante e prática para a definição do número de grupos consiste em construir um gráfico com o número de grupos *versus* o nível de fusão (distância) em que o agrupamento ocorre; caso ocorram vários pontos de “salto”, estes indicam algumas opções para o número de grupos a serem formados. Nestes casos, aconselha-se buscar outros procedimentos para uma melhor definição do número de grupos (Mingoti, 2007).

A Análise do Comportamento do Nível de Similaridade avalia o comportamento do nível de similaridade em cada etapa  $k$  do agrupamento. Dados que  $C_i$  e  $C_l$  são grupos formados em um determinado momento da execução do algoritmo, tem-se que o nível de similaridade expresso em porcentagem, é dado pela equação (3),

$$S_{il} = \left[ 1 - \frac{d_{il}}{\{ \max d_{jk}, j, k = 1, 2, \dots, n \}} \right] 100, \quad (3)$$

em que o  $\max\{d_{jk}\}$  é a distância máxima entre os  $n$  bairros estudados apontada na matriz inicial de distâncias.

Busca-se então pontos onde existam quedas bruscas na similaridade, estes pontos determinam o número ideal de agrupamentos.

### 3.6 Análise de Componentes Principais

A Análise de Componentes Principais (ACP) proposta inicialmente por Karl Pearson teve início do século XX e pertence ao grupo de técnicas de multivariadas que estuda a interdependência de um conjunto de variáveis.

A ACP é utilizada com o propósito de reduzir a dimensionalidade de um vetor de respostas para que assim a estrutura de variabilidade possa ser mais facilmente compreendida. Para realização da redução do espaço paramétrico, as  $p$  variáveis presentes no estudo foram transformadas em  $k$  novas variáveis ( $k < p$ ), sendo estas novas  $k$  variáveis combinações lineares das  $p$  variáveis originais.

Para estimação dessas combinações lineares que também podem ser vistas como coeficiente de ponderação para cada parâmetro em cada um dos componentes, deve-se primeiramente encontrar as raízes características da matriz de correlação amostral dos dados originais designados como autovalores ( $\lambda$ ). Organiza-se os autovalores em ordem decrescente e então obtêm-se os autovetores correspondentes aos autovalores. Os autovetores são normalizados para a unicidade da solução do sistema linear homogêneo e passam a ser referidos como componentes principais obtidos a partir da matriz de correlação.

A primeira combinação linear, ou seja, o primeiro componente principal está associado ao maior autovalor da matriz de correlação e retém a maior porcentagem de explicação dos dados (variabilidade). O segundo componente explica o máximo da variabilidade restante que não foi explicado pelo primeiro componente, e assim, sucessivamente. O procedimento é executado até que o último autovalor, que será o componente de menor contribuição, esteja inserido na análise. Com isso, pode-se caracterizar o  $l$ -ésimo componente principal ( $k < p$ ) como segue na equação (4)

$$\mathbf{Y}_l = a_{l1}\mathbf{Z}_1 + a_{l2}\mathbf{Z}_2 + \dots + a_{lp}\mathbf{Z}_p, \text{ com } l = 1, \dots, k \text{ e } k < p. \quad (4)$$

Os componentes principais são mutuamente não correlacionados, sendo o valor da variância de cada um correspondente ao expresso pelo respectivo autovalor. A contribuição de cada componente para a explicação da variabilidade total dos dados pode ser expressa para o  $k$ -ésimo componente por  $100 \frac{\lambda_k}{p} \%$  (Johnson & Wichern, 2007).

Para seleção do número de componentes principais que devem ser utilizados na redução da dimensionalidade, alguns critérios foram desenvolvidos para facilitar tal tomada de decisão. Dentre os mais utilizados, escolheu-se para este estudo o critério de Kaiser, também conhecido como critério da raiz latente, o qual sugere que para a matriz de correlação, só sejam utilizadas os componentes que apresentam seus respectivos autovalores superiores a um (Barroso & Artes, 2003). Isto deve-se ao fato de que os autovalores mostram o quanto um componente principal consegue reter para explicação da variabilidade total dos dados, caso este valor seja inferior a uma unidade, significa que a própria variável original sozinha explica mais do que este componente. Sendo assim, a criação deste novo componente teria menos variabilidade que uma única variável do estudo.

Messetti (2007) e Silva (2008) resumiram a Análise de Componentes Principais, nos seguintes passos:

1. Construir a matriz de correlação dos dados.
2. Obter a partir da matriz de correlação os autovalores e seus respectivos autovetores normalizados.
3. Ordenar os pares de autovalores / autovetores em ordem decrescente segundo os autovalores.
4. Definir o número de componentes principais.

Para a ordenação dos bairros segundo as repostas das variáveis mensuradas nas coletas de água, obteve-se o valor numérico dos três primeiros componentes principais para todos os bairros e, em seguida, procedeu-se a ordenação no sentido

crescente de magnitude numérica da distância do ponto à origem do sistema tridimensional. Além disso, complementa-se a apresentação dos componentes principais com a construção dos gráficos Biplot.

Para o processamento e análise dos resultados foram utilizados os softwares Statistica 10.0, XLSTAT 2013 e o SAS (Statistical Analysis Software) 9.2.

## 4 RESULTADOS E DISCUSSÃO

Para realização dos procedimentos analíticos, considerou-se o conjunto de dados relacionado aos parâmetros físico-químicos e bacteriológicos (apêndice 1) entre os anos de 2007 e 2011 e utilizou-se como unidade amostral o vetor de medidas de centralidade dos bairros. A complementação dos valores faltantes pelas estimativas pontuais foi de 2,526% de preenchimento.

### 4.1 Medidas descritivas e matriz de correlação

Primeiramente, obteve-se as medidas descritivas de cada variável original analisada, que são apresentadas na Tabela 3, assim como as correlações lineares de Pearson entre todos os pares de variáveis (Tabela 4).

Tabela 3. Medidas descritivas das variáveis relacionados a qualidade da água durante o período de 2007 a 2011.

Variável	Mínimo	Q1	Mediana	Q3	Máximo	Média	Desvio Padrão
corap	4,000	4,000	4,000	4,000	6,250	4,116	0,380
turb	0,225	0,464	0,545	0,726	1,319	0,603	0,243
fluor	0,600	0,671	0,697	0,701	0,760	0,689	0,029
bact	0,000	0,000	0,000	0,000	8,000	0,345	1,172
cloroR	1,317	1,704	1,791	1,904	2,162	1,811	0,168
colif	0,000	0,000	0,000	0,000	0,080	0,003	0,010
pH	6,940	7,528	7,700	7,795	8,118	7,666	0,194

Tabela 4. Matriz de correlação das variáveis relacionadas à qualidade da água durante o período de 2007 a 2011.

	corap	turb	fluor	bact	cloroR	colif	pH
corap	1	0,190	-0,045	-0,012	-0,181	0,328*	-0,094
turb		1	-0,002	-0,046	-0,499*	0,355*	0,216
fluor			1	-0,083	0,031	0,020	0,149
bact				1	0,026	0,110	0,029
cloroR					1	-0,159	-0,318*
colif						1	-0,018
pH							1

\*( $p < 0,01$ )

As associações lineares (correlação de Pearson) entre pares de variáveis relacionadas à qualidade da água apresentaram resultados significativos para cor aparente *versus* coliforme total ( $p < 0,01$ ), turbidez *versus* coliforme total ( $p < 0,01$ ), turbidez *versus* cloroR ( $p < 0,01$ ) e cloroR *versus* pH ( $p < 0,01$ ). Em relação as duas primeiras, os resultados positivos mostram que o aumento de coliforme total aconteceu associado ao aumento da turbidez e da cor aparente, enquanto que nas duas últimas, os resultados negativos indicam que o aumento do cloroR relacionou-se com a diminuição da turbidez e do pH.

## 4.2 Análise de Agrupamento

A Análise de Agrupamento foi fundamentada na matriz de distâncias, obtida por meio da métrica euclidiana e também com a utilização do método de agrupamento hierárquico de Ward cujos resultados estão apresentados no dendograma (Figura 3) a seguir.



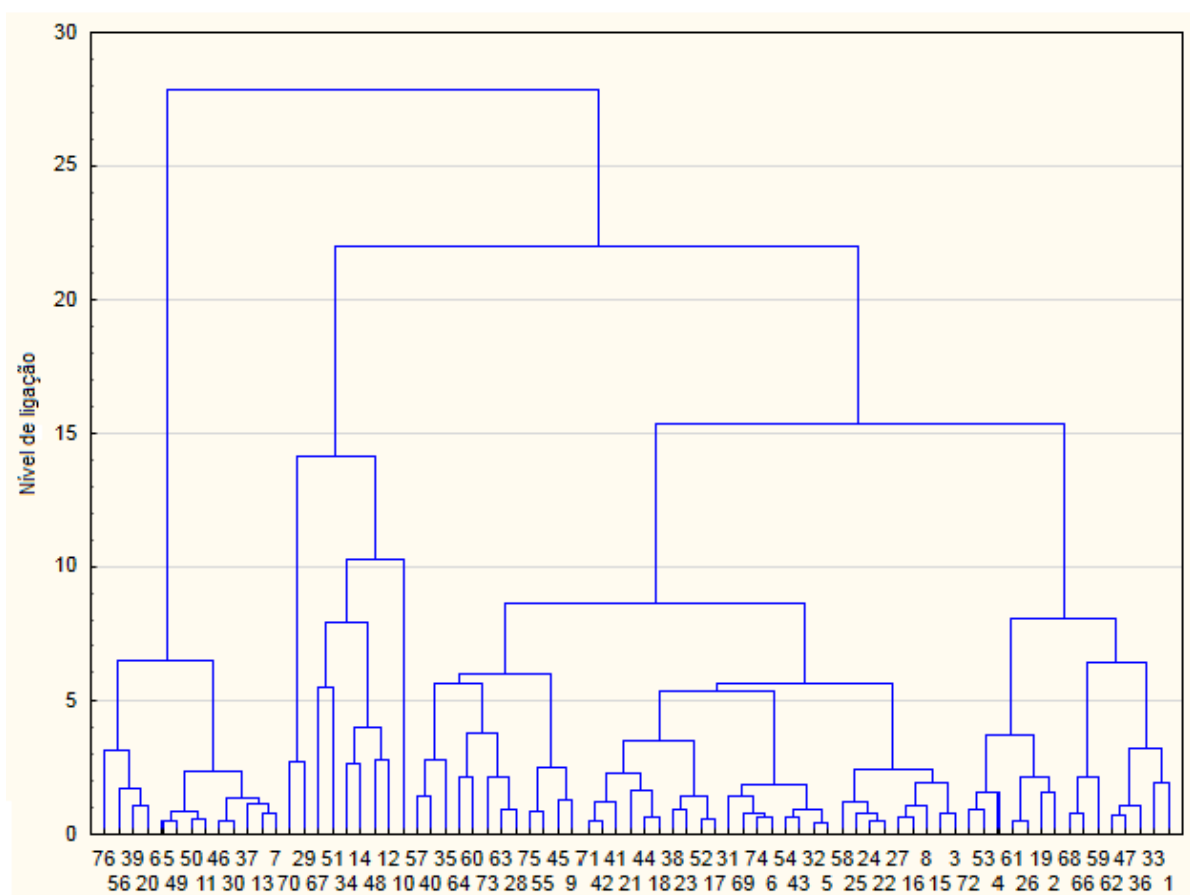


Figura 3 - Dendrograma resultante da análise de agrupamento dos bairros do município de Botucatu.

Para definição do número final de grupos utilizados na análise, primeiramente empregou-se como critério de parada a Análise do Comportamento do Nível de Fusão entre os grupos, em que se obteve um diagrama bidimensional com o eixo  $x$  referente ao número de grupos e o eixo  $y$  ao nível de ligação (Figura 4).

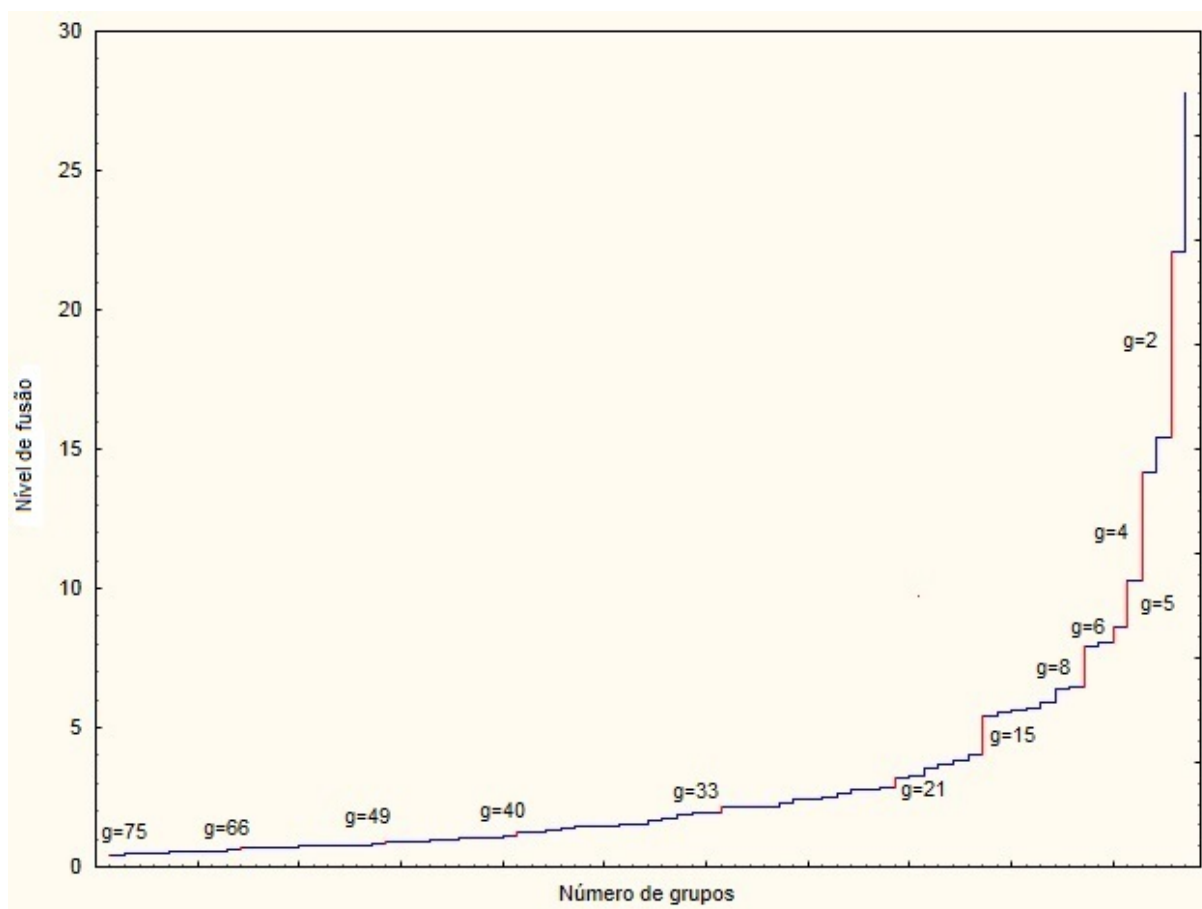


Figura 4 - Gráfico do comportamento do nível de fusão.

É possível observar na Figura 4, o nível de fusão resultante em cada etapa do procedimento, no qual é possível notar vários pontos de “salto” que indicam o momento de parada do algoritmo. A inspeção visual mostra 15, 8, 5 e 4 como possíveis opções para a constituição do número de grupos final.

A prática da análise de dados indica que 4 ou 5 são os valores mais prováveis para o número de grupos final. Porém, como já relatado, torna-se mais interessante dirimir esta dúvida associando a mesma a outro procedimento.

Neste sentido, aplicou-se o critério da Análise do Comportamento do Nível de Similaridade, com os resultados apresentados na Tabela 5. Esta tabela torna-se fundamental para a construção da Figura 5. O nível de similaridade foi

obtido aplicando-se os dados na equação (3), considerando o valor  $\max\{d_{j,k}\}$  dado por  $d_{10,76}=10,333$ .

Tabela 5: Nível de similaridade do agrupamento segundo a distância euclidiana e o método de Ward.

<b>Número de grupos</b>	<b>Distância</b>	<b>Nível de similaridade (<math>S_{ij}</math>)</b>
75	0,415	95,985
74	0,486	95,293
73	0,491	95,246
72	0,510	95,062
71	0,523	94,936
70	0,531	94,864
69	0,563	94,548
68	0,570	94,481
67	0,640	93,810
66	0,667	93,547
65	0,670	93,518
64	0,680	93,419
63	0,730	92,939
62	0,758	92,668
61	0,759	92,652
60	0,764	92,611
59	0,769	92,555
58	0,788	92,371
57	0,863	91,650
56	0,890	91,389
55	0,919	91,107
54	0,936	90,943
53	0,944	90,862

52	0,965	90,664
51	1,045	89,891
50	1,058	89,765
49	1,061	89,729
48	1,117	89,191
47	1,242	87,983
46	1,248	87,923
45	1,291	87,505
44	1,364	86,795
43	1,429	86,167
42	1,434	86,126
41	1,460	85,870
40	1,552	84,983
39	1,559	84,916
38	1,635	84,180
37	1,759	82,975
36	1,854	82,053
35	1,914	81,474
34	1,915	81,469
33	2,122	79,466
32	2,126	79,422
31	2,153	79,168
30	2,181	78,893
29	2,298	77,757
28	2,406	76,719
27	2,442	76,368
26	2,509	75,715
25	2,635	74,503
24	2,757	73,314

23	2,791	72,991
22	2,818	72,725
21	3,186	69,171
20	3,244	68,604
19	3,545	65,696
18	3,701	64,181
17	3,810	63,123
16	4,012	61,176
15	5,391	47,826
14	5,523	46,548
13	5,648	45,343
12	5,703	44,812
11	5,939	42,527
10	6,374	38,315
9	6,471	37,372
8	7,908	23,470
7	8,023	22,354
6	8,583	16,935
5	10,293	0,387
4	14,170	-37,133
3	15,393	-48,973
2	22,069	-113,581
1	27,814	-169,175

---

A Tabela 5 mostra que os valores encontrados para o nível de similaridade ( $S_{ij}$ ) apresentam pequenos decréscimos até a formação de seis grupos.

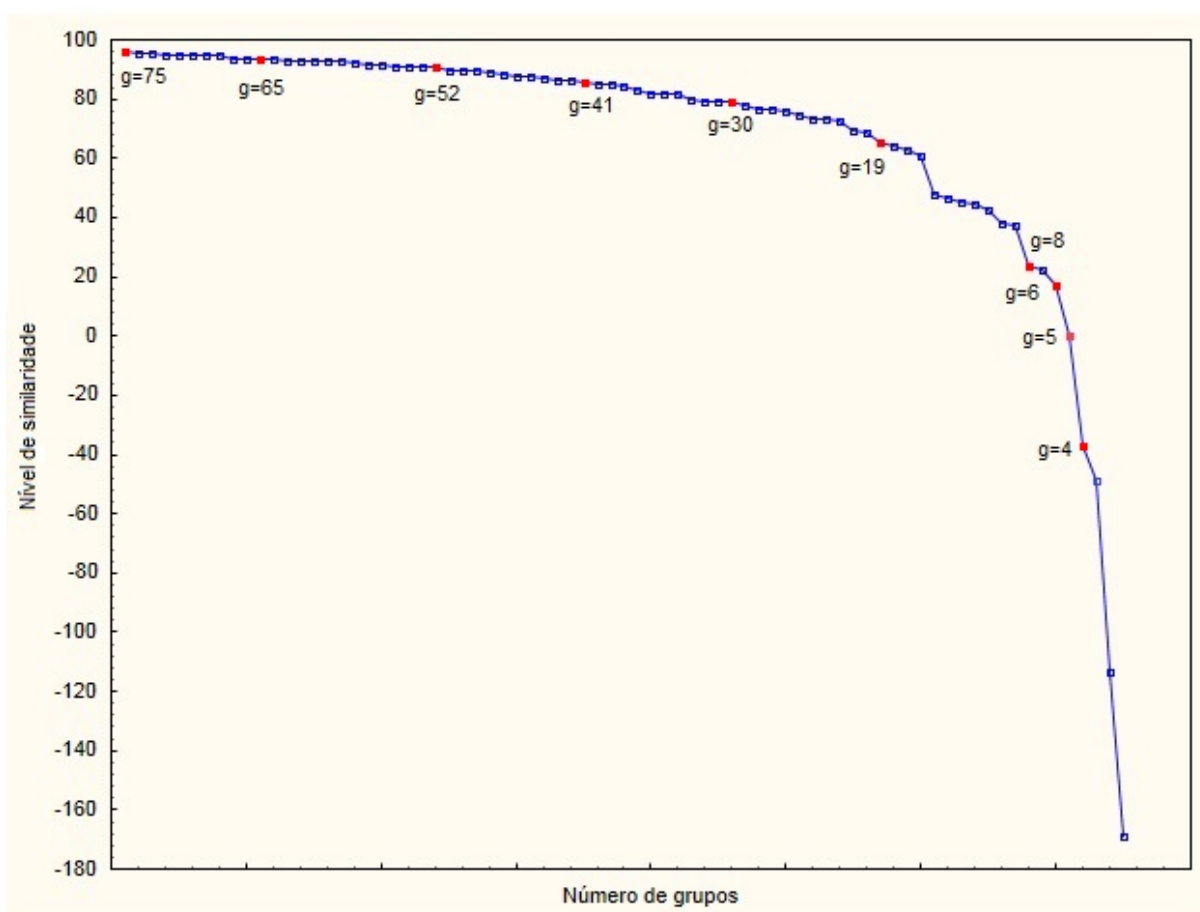


Figura 5 - Gráfico do nível de similaridade.

A inspeção visual da Figura 5 mostrou o primeiro decaimento considerável do nível de similaridade quando o processo formou cinco grupos distintos, apontando este resultado como o valor indicado para a definição final do número de grupos, corroborando com o indicativo prático estabelecido a partir do nível de fusão.

Neste sentido, os dois critérios de parada indicaram que a análise de agrupamento fosse estabelecida por cinco grupos distintos. Estes grupos, a partir da inspeção da Figura 3, são compostos por: 13, 2, 7, 39 e 15 bairros.

O grupo 1 foi formado por 13 bairros está apresentado na Tabela 6.

Tabela 6. Bairros pertencentes ao Grupo 1.

Bairro	Código
Conjunto Habitacional Engenheiro Francisco Blasi	7
Conjunto Residencial Jardim do Mirante	11
Jardim Aeroporto	13
Jardim Flamboyant	20
Jardim Planalto	30
Jardim Tropical	37
Jardim Ypê	39
Parque Santa Inês	46
Residencial Arlindo Durante	49
Residencial Cedro	50
Vila Casa Branca	56
Vila Moreira	65
Vila São Luiz	76

O grupo 2 foi formado apenas por dois bairros está descrito na Tabela

7.

Tabela 7. Bairros pertencentes ao Grupo 2.

Bairro	Código
Jardim Peabiru	29
Vila Pinheiro Machado	70

O grupo 3 foi formado por sete bairros está apresentado na Tabela 8.

Tabela 8. Bairros pertencentes ao Grupo 3.

Bairro	Código
Conjunto Residencial Frei Fidelis	10
Domingos Lopes	12
Jardim Bom Pastor	14
Jardim Santa Eliza	34
Recanto Azul	48
Vale do Sol	51
Vila Padovan	67

O grupo 4, o mais numeroso, foi formado por 39 bairros está apresentado na Tabela 9.

Tabela 9: Bairros pertencentes ao Grupo 4.

Bairro	Código
Centro	3
Conjunto Habitacional Antonio Hermínio Delevedove	5
Conjunto Habitacional Arnaldo Leotta de Mello	6
Conjunto Habitacional Humberto Popolo	8



Conjunto Habitacional José Antônio Lungo	9
Jardim Brasil	15
Jardim Chácara dos Pinheiros	16
Jardim Continental	17
Jardim Cristina	18
Jardim Itamarati	21
Jardim Monte Mor	22
Jardim Neusa Maria (Rubião Junior)	23
Jardim Nossa Senhora das Graças (Rubião Junior)	24
Jardim Palos Verdes	25
Jardim Paraíso	27
Jardim Paraíso II	28
Jardim Real Park	31
Jardim Reflorenda	32
Jardim Santa Mônica	35
Jardim Universitário	38
Park Residencial Convívio	40
Parque das Cascatas	41
Parque dos Pinheiros	42
Parque Marajoara	43
Parque Residencial Nazaré	44
Parque Residencial Vinte e Quatro de Maio	45
Vila Antártica	52
Vila Bela Vista	54
Vila Carmelo	55
Vila Cidade Jardim	57
Vila dos Lavradores	58
Vila Ema	60
Vila Jahu	63

Vila Maria	64
Vila Paulista	69
Vila Rodrigues Alves	71
Vila Santa Terezinha de Menino Jesus	73
Vila São Judas Thadeu	74
Vila São Lúcio	75

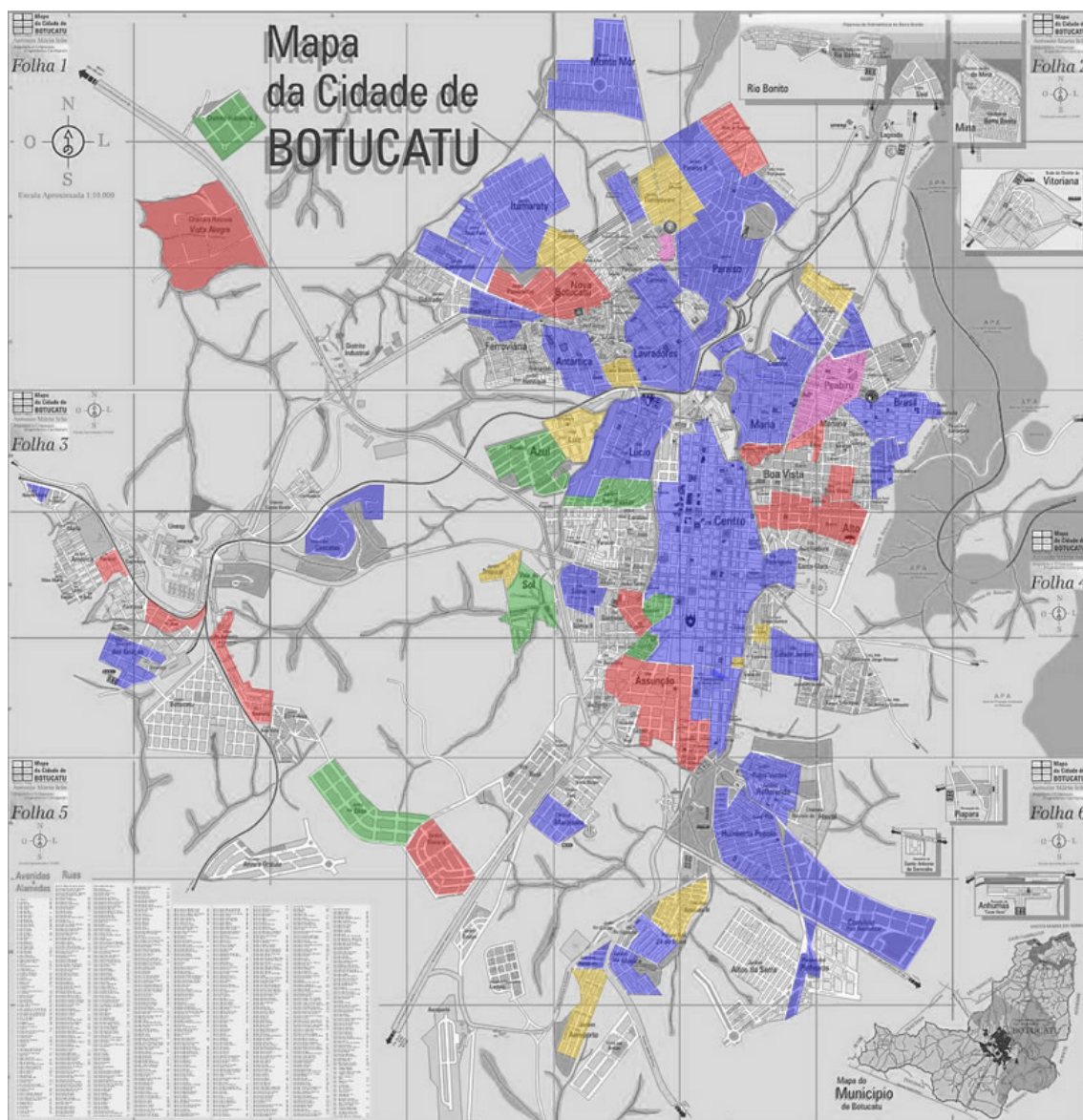
---

O grupo 5 foi formado por 15 bairros está descrito na Tabela 10.

Tabela 10. Bairros pertencentes ao Grupo 5.

Bairro	Código
Altos do Paraíso	1
Bairro Alto	2
Chácara Recreio Vista Alegre	4
Jardim Dona Nicota de Barros	19
Jardim Panorama	26
Jardim Riviera	33
Jardim São José (Rubião Júnior)	36
Parque Santo Antônio da Cascatinha (Rubião Júnior)	47
Vila Assumpção	53
Vila Éden	59
Vila Eny	61
Vila Guimarães	62
Vila Nova Botucatu	66
Vila Paraíso (Rubião Júnior)	68
Vila Santa Luzia	72

Apresenta-se, na Figura 6, o perfil da distribuição de similaridade dos bairros segundo o agrupamento, estabelecido por meio da distância euclidiana determinada a partir do vetor médio de respostas dos parâmetros físico-químico e bacteriológico e identificados no mapa geográfico dos bairros do município por cores distintas.



Fonte: Adaptado de Ielo (2013)

Figura 6 - Distribuição geográfica dos grupos de semelhança.

Na Figura 6, o grupo 1 foi identificado com a cor amarela, o 2 com rosa, o 3 com verde, o 4 com azul e o 5 com vermelha. Pode-se observar, que o grupo 1, mesmo possuindo 13 bairros, apenas 11 foram localizados na Figura 6, como relatado anteriormente (Jardim Ypê e Residencial Cedro).

Os bairros englobados nos cinco grupos apresentaram abrangência mu-

nicipal completa e dispersa, atingindo todas as regiões, com exceção do grupo 2 (formado apenas por dois bairros) que se apresentou concentrado na região norte do município.

O vetor de referência média de cada grupo em relação as variáveis originais estudadas pode ser visualizado na Tabela 11.

Tabela 11. Perfil médio de respostas físico-química e bacteriológica dos grupos.

Grupo	corap	turb	fluor	bact	cloroR	colif(%)	pH
1	4,000	0,325	0,678	0,269	2,077	0,000	7,501
2	4,000	0,637	0,696	6,750	1,812	0,800	7,695
3	5,190	0,848	0,688	0,214	1,711	1,800	7,665
4	4,013	0,561	0,704	0,147	1,759	0,200	7,720
5	4,000	0,836	0,663	0,133	1,764	0,100	7,670

A Tabela 11 traz informações a respeito das respostas médias das variáveis, segundo os grupos formados, sendo possível avaliar quais os parâmetros se encontraram dentro dos padrões fixados pela portaria vigente.

Todos os cinco grupos mostraram os parâmetros cor aparente, turbidez, flúor, bactéria heterotrófica e pH em conformidade com a legislação. O parâmetro cloroR, apenas para o grupo 1, apresentou-se acima do padrão (máximo permitido 2mg/L). Além disso, verificou-se para os grupos 2, 3, 4 e 5 a ocorrência positiva de coliforme total em 0,800%, 1,800%, 0,200% e 0,100% do total de amostras coletadas, respectivamente.

### 4.3 Análise de Componentes Principais

A partir da matriz de dados padronizados, considerou-se a técnica da Análise de Componentes Principais, cujos autovalores e porcentagens de variação total dos dados estão apresentados na Tabela 12.

Tabela 12. Autovalores e porcentagens da variação total explicada

	Autovalor	Variação total %	Variação total acumulada %
CP1	1,940	27,727	27,728
CP2	1,309	18,708	46,436
CP 3	1,050	15,005	61,442
CP 4	0,986	14,093	75,536
CP 5	0,697	9,958	85,494
CP 6	0,583	8,339	93,833
CP 7	0,431	6,166	100,000

Os componentes principais foram selecionados pelo critério de Kaiser, que utilizou somente os autovalores superiores a um (Tabela 12), desta forma apenas os três primeiros componentes foram empregados na ordenação dos bairros segundo as características analisadas.

Os três primeiros componentes principais (autovetores normalizados na matriz de correlação) selecionados para a análise são apresentados na Tabela 13

Tabela 13. Coeficientes físico-químicos e bacteriológicos

CP	corap	turb	fluor	bact	cloroR	colif	pH
1	0,475	0,807	0,023	0,012	-0,750	0,592	0,385
2	0,553	-0,076	-0,423	0,221	0,260	0,458	-0,701
3	0,276	0,017	0,460	-0,827	0,107	0,051	-0,250

Observa-se nos coeficiente apresentados na Tabela 13, que a variável cloroR se contrasta com as demais, e as variáveis flúor e bactéria heterotrófica contribuem pouco para formação do primeiro componente. Para o segundo componente, pode-se observar a variável turbidez com baixa contribuição e as variáveis flúor e

pH contribuindo todas de forma negativa. As demais variáveis apresentam-se contribuindo positivamente para a formação do componente. O terceiro componente apresenta uma baixa contribuição da variável turbidez e da coliforme total, ambas contribuindo positivamente, já bactéria heterotrófica e pH contrastam negativamente com as demais.

A representação gráfica dos três componentes principais no sistema bidimensional envolvendo todas as possíveis combinações em pares resultam nas Figuras 7, 8 e 9 (Biplot).

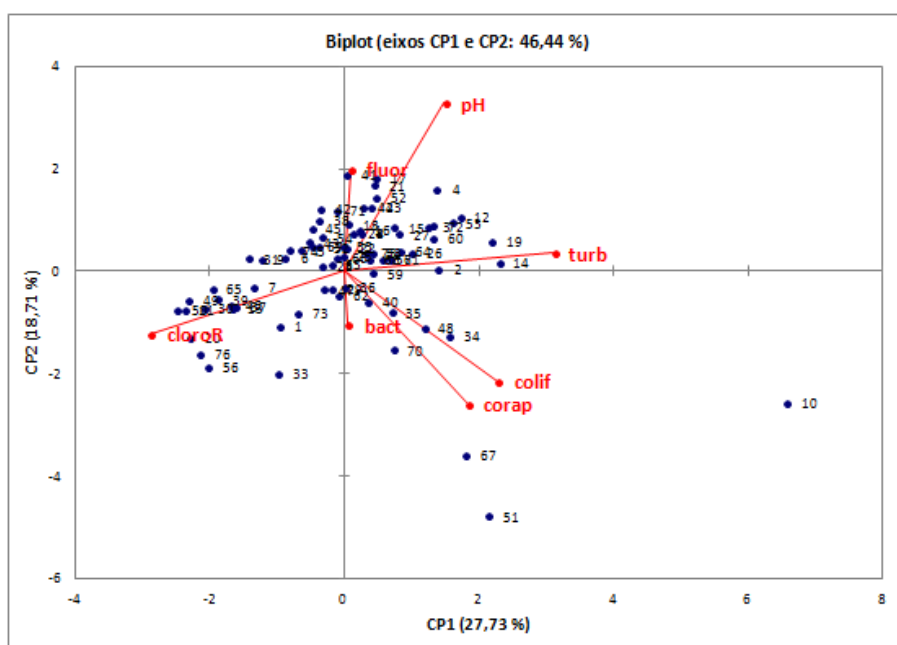


Figura 7 - Biplot do primeiro e segundo componente principal.

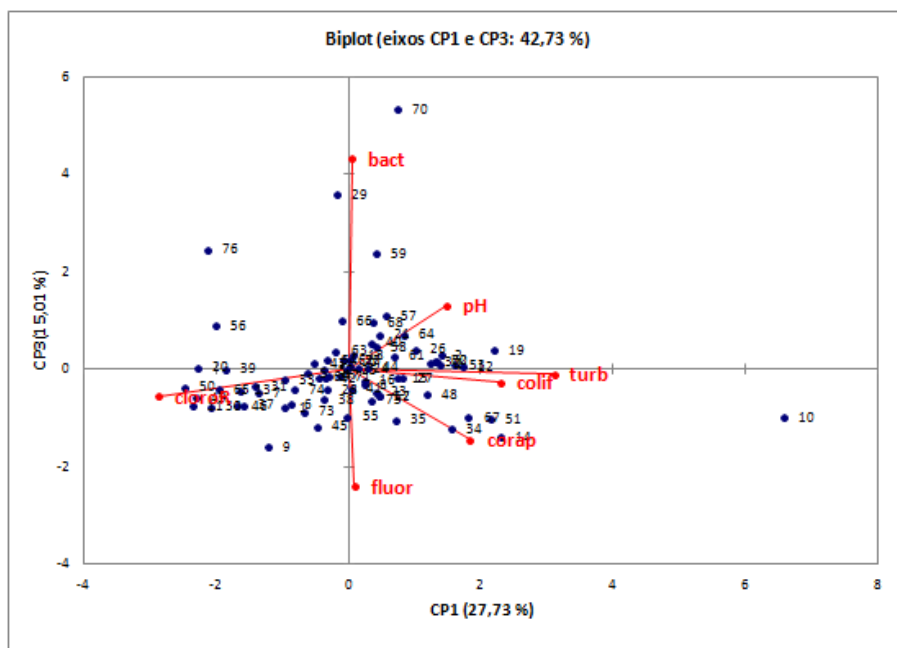


Figura 8 - Biplot do primeiro e terceiro componente principal.

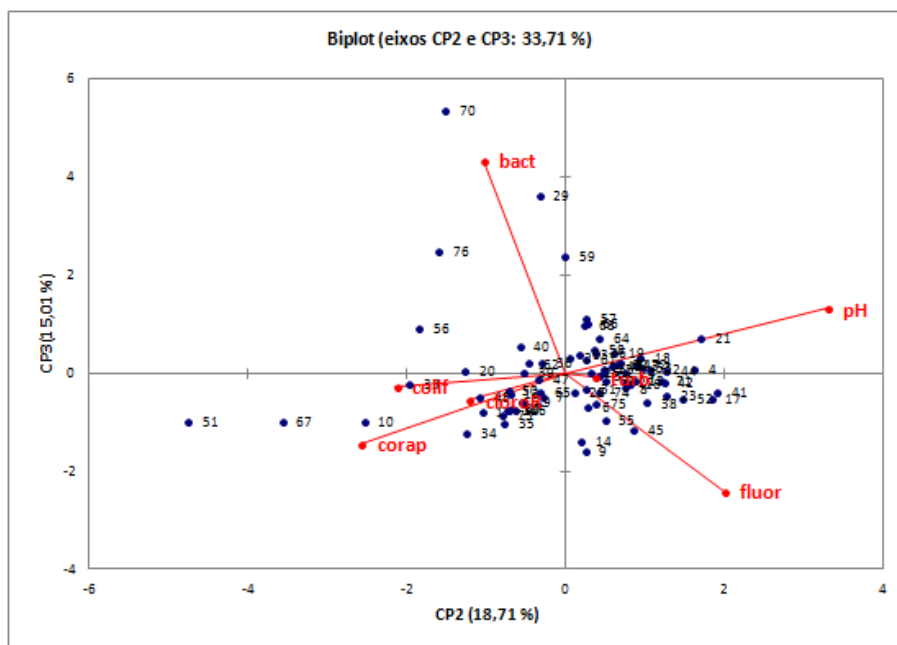


Figura 9 - Biplot do segundo e terceiro componente principal.



A Figura 7 relaciona os dois primeiros componentes principais que retem conjuntamente 46,44% da informação da dispersão dos dados e aponta as variáveis pH, turb, colif, corap e cloroR como sendo as mais contributivas para a composição dos agrupamentos dos bairros neste Biplot. A Figura 8 refere-se aos componentes 1 e 3, retendo 42,73% da informação da dispersão dos dados com destaque as variáveis cloroR, bact e turb. Por fim, a Figura 9 envolvendo os componentes 2 e 3, representa 33,71% da variabilidade destacando-se as variáveis pH, bact e corap. Nas duas primeiras situações tem-se como variáveis comuns as variáveis turb e cloroR. Nas duas últimas somente a variável bact.

Estes gráficos reunidos mostram a dispersão dos bairros segundo os quadrantes estabelecidos no sistema bidimensional como um conjunto preliminar de bairros semelhantes em face das variáveis mais contributivas.

A partir dos três componentes, conseguiu-se construir um indicador numérico, para cada bairro, sendo este representado pela soma de cada componente a origem elevado ao quadrado (distância). Com o valor da distância obtido foi possível fazer o ranqueamento dos bairros de acordo com as variáveis consideradas para o estudo da qualidade da água. O resultado deste indicador numérico, caracterizado pela distância, pode ser analisado na Tabela 14.

Tabela 14: Bairros ordenados pelo valor da distância

Bairro	Distância
Residencial Cedro	4,356
Conjunto Residencial Jardim do Mirante	4,390
Jardim Flamboyant	4,418
Vila Casa Branca	4,441
Residencial Arlindo Durante	4,446
Jardim Planalto	4,472
Jardim Tropical	4,550
Vila Moreira	4,554

Jardim Ypê	4,564
Parque Santa Inês	4,581
Jardim Riviera	4,587
Jardim Aeroporto	4,593
Conjunto Habitacional José Antonio Lungo	4,676
Conjunto Habitacional Engenheiro Francisco Blasi	4,677
Vila São Luiz	4,693
Jardim Real Park	4,712
Altos do Paraíso	4,726
Vila Santa Therezinha de Menino Jesus	4,729
Conjunto Habitacional Arnaldo Leotta de Mello	4,826
Vila São Judas Thadeu	4,878
Jardim Paraíso II	4,912
Jardim Santa Mônica	4,929
Vila Paulista	4,939
Conjunto Habitacional Antonio Hermínio Delevedove	4,942
Park Residencial Convívio	4,943
Parque Residencial Vinte e Quatro de Maio	4,951
Parque Santo Antonio da Cascatinha (Rubião Júnior)	4,963
Vila Jahu	4,964
Parque Marajoara	4,968
Vila Bela Vista	5,003
Jardim Reflorenda	5,006
Vila Guimarães	5,013
Jardim Universitário	5,016
Vila Carmelo	5,022
Jardim São José (Rubião Junior)	5,025
Vale do Sol	5,045
Parque dos Pinheiros	5,079

Jardim Palos Verdes	5,082
Jardim Monte Mor	5,086
Vila Nova Botucatu	5,106
Vila Rodrigues Alves	5,119
Vila São Lúcio	5,131
Conjunto Habitacional Humberto Popolo	5,136
Jardim Nossa Senhora das Graças (Rubião Júnior)	5,146
Jardim Cristina	5,167
Jardim Chácara dos Pinheiros	5,181
Jardim Santa Eliza	5,187
Vila dos Lavradores	5,211
Vila Paraíso (Rubião Junior)	5,218
Parque das Cascatas	5,222
Parque Residencial Nazaré	5,231
Jardim Neusa Maria (Rubião Júnior)	5,236
Vila Cidade Jardim	5,238
Vila Antártica	5,238
Vila Padovan	5,250
Recanto Azul	5,251
Jardim Continental	5,268
Vila Maria	5,279
Vila Eny	5,289
Jardim Brasil	5,293
Jardim Paraíso	5,293
Vila Ema	5,328
Jardim Itamarati	5,353
Jardim Panorama	5,376
Vila Éden	5,392
Bairro Alto	5,401

Centro	5,428
Vila Santa Luzia	5,515
Domingos Lopes	5,561
Chácara Recreio Vista Alegre	5,573
Vila Assumpção	5,619
Jardim Dona Nicota de Barros	5,624
Jardim Bom Pastor	5,649
Conjunto Residencial Frei Fidelis	5,985
Jardim Peabiru	6,582
Vila Pinheiro Machado	8,184

---

A Tabela 14 apresenta a distância euclidiana das respostas dos bairros em relação à origem do sistema de coordenadas descrito pelos eixos estabelecidos pelos três primeiros componentes principais. Para entender a semelhança das respostas entre dois bairros quaisquer, deve-se verificar quão afastado da origem se encontram seus respectivos valores, considerando os valores calculados pela métrica euclidiana, indicando maior semelhança entre eles quanto menor for a diferença entre os valores do afastamento. Neste sentido, percebe-se que os bairros mais dissemelhantes são Residencial Cedro (50) e Vila Pinheiro Machado (70), enquanto que os mais semelhantes, Vila Cidade Jardim (57) e Vila Antártica (52), assim como, Jardim Brasil (15) e Jardim Paraíso (27).

Associou-se à Tabela 14, as mesmas escalas de cores que foram utilizadas na construção da Figura 6. Desta forma, foi possível observar que os bairros do grupo 1 (Tabela 6) foram os que apresentaram menor distância no novo sistema, excetuando-se os bairros Jardim Riviera e Conjunto Habitacional José Antônio Lungo. Já os bairros do grupo 5 (Tabela 7) foram os que apresentaram as maiores distâncias.

Nota-se também que o bairro Residencial Cedro apresenta a menor

distância do novo sistema e a Vila Pinheiro Machado a maior distância encontrada.

## 5 CONCLUSÃO

Os resultados dos procedimentos estatísticos multivariados permitem as seguintes conclusões a respeito do perfil geográfico da água distribuída no município de Botucatu:

A Análise de Agrupamento distribuiu os bairros em cinco grupos distintos quanto aos resultados laboratoriais dos parâmetros físico-químicos e bacteriológicos. O grupo 1, composto por 13 bairros, apresentou resposta média para o parâmetro cloroR acima do padrão estabelecido, porém sem apresentar comprometimento na saúde dos consumidores (valores abaixo do mínimo é que são prejudiciais a saúde, já que não se apresentariam em quantidade suficiente para o processo de destruição das bactérias). Além disso, os grupos 2, 3, 4 e 5 formados por 2, 7, 39 e 15 bairros, respectivamente, apresentaram presença de coliformes totais nas amostras coletadas, cujos correspondentes valores médios foram da ordem de 0,800%, 1,800%, 0,200% e 0,100%. Neste sentido, pode-se concluir que apenas um único grupo, para uma única variável se encontram fora do padrão fixado pela portaria vigente.

O mapa de alocação dos bairros nos cinco grupos mostrou que não há sistematização nas localizações dos bairros no município, ou seja, os bairros apresentam-se distribuídos de forma casual e esparsa por todo o território.

A Análise de Componentes Principais, utilizando os três primeiros componentes principais, permitiu construir um indicador numérico dos bairros utilizando a distância euclidiana e apresentar os gráficos Biplot das combinações entre pares de eixos e associar quais variáveis foram as mais contributivas no processo de agrupamento dos bairros, segundo os parâmetros relacionados à água. Quando obtido o ranqueamento dos bairros por meio da distância, observou-se que os bairros

que apresentam maiores diferença são Residencial Cedro e Vila Pinheiro Machado, enquanto os mais semelhantes são Vila Cidade Jardim e Vila Antártica, e também, Jardim Brasil e Jardim Paraíso. O Biplot relacionado aos eixos mais contributivos (CP1 e CP2) destaca como as variáveis mais fundamentais para os agrupamentos: pH, turb, colif, corap e cloroR.

A utilização de diferentes técnicas de estatística multivariada, com a finalidade de obter grupos de bairros o mais homogêneos possíveis, conseguiu mostrar uma grande semelhança na composição, principalmente para os grupos 1 e 2.

Para a variável coliforme total a resposta média da positividade no município foi da ordem 0,345% com desvio padrão de 1,172% porém, destaca-se, em particular, o Conjunto Residencial Frei Fidelis apresentou 8% de positividade, cujo valor refletiu a única ocorrência acima do padrão estabelecido.

## REFERÊNCIAS

AMARAL, L. A.; NADER FILHO, A.; ROSSI JUNIOR, O. D.; FERREIRA, F. L. A.; BARROS, L. S. S. Água de consumo humano como fator de risco à saúde em propriedades rurais. **Revista de Saúde Pública**, v.37, n.4, p.510–514, 2003.

ANDRADE, E.; ARAÚJO, L.; ROSA, M.; DISNEY, W.; ALVES, A. Seleção dos indicadores da qualidade das águas superficiais pelo emprego da análise multivariada. **Engenharia Agrícola**, v.27, n.3, p.683–690, 2007a.

ANDRADE, E.; ARAÚJO, L.; ROSA, M.; GOMES, R.; LOBATO, F. Fatores determinantes da qualidade das águas superficiais na bacia do Alto Acaraú, Ceará, Brasil. **Ciência Rural**, v.37, n.6, p.1791–1797, 2007b.

BARROSO, L. P.; ARTES, R. Análise Multivariada. In: 10<sup>o</sup> SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO / 48<sup>o</sup> REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 2003. ; resumos. Lavras: IBMEC, 2003. 151.

BOYACIGLU, H.; BOYACIGLU, H. Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin - Turkey. **Environ Geol**, v.54, n.2, p.275–282, 2008.

BRASIL. **Comentários sobre a Portaria MS nº 518/2004 : subsídios para implementação**. Brasília: Ministério da Saúde, 2005. 91p.

BRASIL. **Métodos alternativos de desinfecção da água**. Disponível em < [www.finep.gov.br/prosab/livros/LuizDaniel.pdf](http://www.finep.gov.br/prosab/livros/LuizDaniel.pdf) >. Acesso em 19 de novembro. Brasília: Agência Brasileira da Inovação, 2013.



BUFON, A.; LANDIM, P. Análise de qualidade da água por metodologia estatística multivariada na represa Velha (CEPTA/IBAMA/Pirassununga/SP). **Holos Environment**, v.7, n.1, p.42–59, 2007.

CAMPOS, J. A. D. B.; FILHO, A. F.; FARIA, J. B. Qualidade sanitária da água distribuída para consumo humano pelo sistema de abastecimento público da cidade de Araraquara-SP. **Revista Alimento e Nutrição**, v.1, p.17–129, 2002.

CARNEIRO, C. D. R. Viagem virtual ao Aquífero Guarani em Botucatu (SP): Formações Pirambóia e Botucatu, Bacia do Paraná. **Terræ Didática**, v.3, n.1, p.50 – 73, 2010.

EATON, A. D.; RICE, E. W.; BAIRD, R. B.; LENORE, S. C. **Standard methods for the examination of water and wastewater**. Washington: American Public Health Association, 2012. 1496p.

GIRÃO, E.; ANDRADE, E.; ROSA, M.; ARAÚJO, L.; MEIRELES, A. Seleção dos indicadores da qualidade de água no rio Jaibaras pelo emprego da análise da componente principal. **Revista de Ciência Agronômica**, v.38, n.1, p.14–24, 2007.

GUEDES, H.; SILVA, D. D.; ELESBON, A.; RIBEIRO, C.; MATOS, A.; SOARES, J. Aplicação da análise estatística multivariada no estudo da qualidade da água do rio Pomba, MG. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.16, n.5, p.558–563, 2012.

IELO, A. M. Mapa da cidade de Botucatu. Disponível em < [www.emv.fmb.unesp.br/mapas/mapbtu.asp](http://www.emv.fmb.unesp.br/mapas/mapbtu.asp) >. Acesso em 22 de setembro, 2013.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. New Jersey: Prentice Hall, 2007. 773p.

KNUPP, E. A. N. Usos de Métodos Estatísticos para dados de qualidade de águas: estudo de caso, rio das Velhas. Departamento de Química - Instituto de Ciências Exatas, 2007. 203p. Tese de doutorado - Universidade Federal de Minas Gerais.

LI, Y.; XU, L.; LI, S. Water quality analysis of the Songhua River basin using multivariate techniques. **Journal of Water Resource and Protection**, v.1, n.2, p.110–121, 2009.

MACÊDO, J. **Águas & águas**. Belo Horizonte: CRQ-MG, 2004. 977p.

MARINELLI, P. S.; OTOBONI, A. M. M. B.; TANAKA, A. Y.; MENDONÇA, C. C. T. N.; NEVES, V. J. M.; OLIVEIRA, A. S. Qualidade físico - química da água fornecida pela rede de abastecimento da cidade de Marília - SP. **Revista Alimentus**, v.2, n.1, 2012.

MAZLUM, N.; ÖZER, A.; MAZLUM, S. Interpretation of water quality data by principal components analysis. **Journal of Engineering and Environmental Sciences**, v.23, p.19–26, 1999.

MESSETTI, A. V. L. Utilização de técnicas na avaliação da divergência genética de populações de girassol (*Helianthus annuus L.*). Faculdade de Ciências Agrárias, 2007. 87p. Tese de doutorado - Universidade Estadual Paulista.

MINGOTI, S. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2007. 297p.

MUSTAPHA, A.; NABEGU, A. Surface water pollution source identification using principal component and factor analysis in Getsi River, Kano, Nigeria. **Australian Journal of Basic and Applied Sciences**, v.5, n.12, p.1507–1512, 2011.

NASCIMENTO, F. M. Avaliação da citricultura no município de Botucatu (SP), por meio de diferentes bases cartográficas e aplicativos. Faculdade de Ciências Agrárias, 2010. 75p. Tese de doutorado - Universidade Estadual Paulista.

NELSON, D.; COX, M. M. **Lehninger princípios de bioquímica**. São Paulo: Editora Sarvier, 2006. 1202p.

NETTO, A.; GARCIA, C.; ALVES, J.; FERREIRA, R.; SILVA, M. Physical and chemical characteristics of water from the hydrographic basin of the Poxim River, Sergipe State, Brazil. **Environment Monitoring and Assessment**, v.184, n.11, p.159 – 170, 2012.

NONATO, E.; VIOLA, Z.; ALMEIDA, K.; SCHOR, H. Tratamento estatístico dos parâmetros da qualidade das águas da bacia do alto curso do rio das Velhas. **Química Nova**, v.30, n.4, p.797–804, 2007.

OLIVEIRA, A. S.; SANTOS, D. C.; OLIVEIRA, E. N. A. J. G. B.; SILVA, J. W. L. Qualidade da água para consumo humano distribuída pelo sistema de abastecimento público de Guarariba. **Revista Verde de Agroecologia e Desenvolvimento Sustentável**, v.7, n.2, 2012.

OLIVEIRA, C. S.; FILHO, A. R. O.; BRAGA, J. O.; PINHEIRO, L. G.; LIMA, M. L. M.; ALENCAR, M. C. L. Avaliação da influência do estado de conservação de reservatórios da rede de distribuição sobre a potabilidade da água distribuída, em Manaus, no ano de 2001. In: CONGRESSO INTERAMERICANO DE INGENIERÍA SANITARIA Y AMBIENTAL, Cancun, 2002. ; resumos. Cancun: , 2002. 4.

PALÁCIO, H.; ANDRADE, E.; LOPES, F.; ALEXANDRE, D.; ARRAES, F. Similaridade da qualidade das águas superficiais da bacia do Curu, Ceará. **Ciência Rural**, v.30, n.9, p.2494–2500, 2009.

PORTO, M. A. L.; OLIVEIRA, A. M.; FAI, A. E. C.; STAMFORD, T. L. M. Coliformes em água de abastecimento de lojas fast-food da Região Metropolitana do Recife(PE-Brasil). **Ciência & Saúde Coletiva**, v.16, n.5, 2011.

PRAUS, P. Urban water quality evaluation using multivariate analysis. **Acta Montanistica Slovaca**, v.12, n.2, p.150–158, 2007.

SÁ, L. L.; JESUS, I.; SANTOS, E. C. O.; VALE, E. R.; LOUREIRO, E. C. B.; SÁ, E. V. Qualidade microbiológica da água para consumo humano em duas áreas contempladas com intervenções de saneamento-Belém do Pará, Brasil. **Epidemiologia e Serviços de Saúde**, v.14, n.3, 2005.

SILVA, A.; SACOMANI, L. Using chemical and physical parameters to define the quality of Pardo river water (Botucatu - SP - Brazil). **Water Research**, v.35, n.6, p.1609–1616, 2001.

SILVA, S. H. M. G. Processamento digital de imagens para identificação da sigatoka negra em bananais utilizando análise de componentes principais e redes neurais artificiais. Faculdade de Ciências Agrárias, 2008. 83p. Tese de doutorado - Universidade Estadual Paulista.

SIMEONOV, V.; STRATIS, J.; SAMARA, C.; ZACHARIADIS, G.; VOUTSA, D.; ANTHEMIDIS, A.; SOFONIOU, M.; KOUIMTZIS, T. Assessment of the surface water quality in Northern Greece. **Water Research**, v.37, n.17, p.4119–4124, 2003.

SIQUEIRA, L. P.; SHINOHARA, N. K. S.; LIMA, R. M. T.; PAIVA, J. E.; LIMA FILHO, J. L.; T., C. I. Avaliação microbiológica da água de consumo empregada em unidades de alimentação. **Ciência & Saúde Coletiva**, v.15, n.1, 2010.

SOJKA, M.; SIEPAK, M.; ZIOLA, A.; FRANKOWSKI, M.; MURAT-BLAZEJEWSKA, S.; J.SIEPAK. Application of multivariate statistical techniques to evaluation of water quality in the Mala Welna River (Western Poland). **Environment Monitoring and Assessment**, v.147, p.159–170, 2008.

ZHAO, Z.; CUI, F. Multivariate statistical analysis for the surface water quality of the Luan River, China. **Journal of Zhejiang University SCIENCE A**, v.10, n.1, p.142–148, 2009.

ZIMMERMANN, C.; GUIMARÃES, O.; ZAMORA, P. P. Avaliação da qualidade do corpo hídrico do rio Tibagi na região de Ponta Grossa utilizando análise de componentes principais (PCA). **Química Nova**, v.31, n.7, p.1727–1732, 2008.

ZULPO, D. L.; PERETTI, J.; ONO, L. M.; GARCIA, J. L. Avaliação microbiológica da água consumida nos bebedouros da Universidade Estadual do Centro-Oeste, Guarapuava, Paraná, Brasil. **Semina: Ciências Agrárias**, v.27, n.1, p.107–110, 2006.

## 6 APÊNDICE

### 6.1 Dados

Tabela 15: Perfil das respostas físico-químicas e bacteriológicas dos bairros (matriz de dados)

Código do bairro	corap	turb	fluor	bact	cloroR	colif	pH
1	4,000	0,800	0,691	0,000	2,028	0,000	7,300
2	4,000	0,968	0,660	0,000	1,518	0,000	7,536
3	4,000	0,856	0,685	0,000	1,601	0,001	7,760
4	4,000	0,998	0,700	0,000	1,745	0,000	7,967
5	4,000	0,479	0,688	0,000	1,882	0,000	7,742
6	4,000	0,487	0,717	0,000	1,900	0,000	7,588
7	4,000	0,410	0,695	0,000	1,960	0,000	7,527
8	4,000	0,633	0,709	0,000	1,846	0,007	7,816
9	4,000	0,448	0,760	0,000	1,909	0,000	7,417
10	5,000	1,271	0,700	0,000	1,584	0,080	7,914
11	4,000	0,250	0,700	0,000	2,122	0,000	7,450
12	4,500	0,763	0,688	0,000	1,619	0,000	8,008
13	4,000	0,448	0,683	0,000	2,058	0,000	7,484
14	5,000	1,180	0,733	0,000	1,804	0,000	7,800
15	4,000	0,781	0,700	0,000	1,672	0,000	7,717
16	4,000	0,721	0,702	0,000	1,799	0,000	7,752
17	4,000	0,527	0,740	0,000	1,641	0,000	7,873

18	4,000	0,525	0,680	0,000	1,754	0,000	7,874
19	4,000	1,071	0,669	0,000	1,556	0,008	7,777
20	4,000	0,275	0,650	0,000	2,115	0,000	7,475
21	4,000	0,473	0,675	0,000	1,703	0,000	8,118
22	4,000	0,546	0,683	0,000	1,704	0,000	7,694
23	4,000	0,685	0,723	0,000	1,739	0,000	7,790
24	4,000	0,587	0,690	0,000	1,732	0,000	7,763
25	4,000	0,658	0,681	0,000	1,795	0,000	7,679
26	4,000	0,943	0,660	0,000	1,705	0,000	7,723
27	4,000	0,818	0,698	0,000	1,760	0,005	7,772
28	4,000	0,433	0,700	0,000	1,646	0,000	7,500
29	4,000	0,544	0,700	5,500	1,786	0,000	7,700
30	4,000	0,383	0,700	0,000	2,132	0,000	7,450
31	4,000	0,250	0,700	0,000	1,926	0,000	7,682
32	4,000	0,502	0,686	0,000	1,815	0,000	7,739
33	4,000	0,567	0,652	0,000	1,962	0,010	7,274
34	5,250	0,726	0,733	0,750	1,937	0,015	7,760
35	4,000	0,629	0,725	0,000	1,806	0,027	7,489
36	4,000	0,628	0,660	0,000	1,690	0,000	7,529
37	4,000	0,350	0,700	0,000	1,925	0,000	7,375
38	4,000	0,502	0,727	0,000	1,826	0,000	7,750
39	4,000	0,233	0,667	0,000	2,014	0,000	7,600
40	4,000	0,408	0,688	1,000	1,780	0,022	7,700
41	4,000	0,534	0,733	0,000	1,840	0,000	8,007
42	4,000	0,476	0,710	0,000	1,869	0,000	7,906
43	4,000	0,519	0,700	0,500	1,890	0,000	7,771
44	4,000	0,568	0,699	0,000	1,732	0,000	7,892
45	4,000	0,608	0,750	0,000	1,888	0,000	7,622
46	4,000	0,500	0,700	0,000	2,094	0,000	7,450

47	4,000	0,726	0,674	0,000	1,863	0,000	7,527
48	5,333	0,517	0,667	0,000	1,750	0,000	7,825
49	4,000	0,275	0,695	0,000	2,159	0,000	7,541
50	4,000	0,225	0,700	0,500	2,162	0,000	7,500
51	5,000	1,013	0,633	0,000	1,689	0,029	6,940
52	4,000	0,581	0,732	0,000	1,640	0,000	7,785
53	4,000	1,319	0,683	0,000	1,858	0,000	7,848
54	4,000	0,466	0,700	0,500	1,789	0,000	7,763
55	4,000	0,675	0,733	0,000	1,740	0,000	7,508
56	4,000	0,350	0,633	1,000	2,045	0,000	7,375
57	4,000	0,543	0,700	1,750	1,765	0,013	7,832
58	4,000	0,682	0,659	0,000	1,702	0,000	7,742
59	4,000	0,663	0,633	2,000	1,732	0,000	7,844
60	4,000	0,660	0,700	0,500	1,317	0,000	7,533
61	4,000	0,925	0,664	0,000	1,781	0,000	7,700
62	4,000	0,727	0,656	0,000	1,794	0,000	7,529
63	4,000	0,432	0,700	1,000	1,634	0,000	7,562
64	4,000	0,565	0,650	0,000	1,463	0,000	7,700
65	4,000	0,300	0,691	0,000	2,096	0,000	7,600
66	4,000	0,463	0,633	0,000	1,747	0,000	7,857
67	6,250	0,469	0,662	0,750	1,593	0,000	7,407
68	4,000	0,624	0,633	0,000	1,696	0,000	7,808
69	4,000	0,526	0,713	0,500	1,836	0,000	7,668
70	4,000	0,729	0,691	8,000	1,838	0,016	7,690
71	4,000	0,490	0,708	0,000	1,789	0,000	7,867
72	4,000	1,114	0,680	0,000	1,783	0,000	7,825
73	4,500	0,313	0,700	0,000	1,802	0,000	7,488
74	4,000	0,475	0,705	0,000	1,917	0,000	7,690
75	4,000	0,801	0,713	0,000	1,736	0,000	7,540



---

76	4,000	0,225	0,600	2,000	2,122	0,000	7,682
----	-------	-------	-------	-------	-------	-------	-------