

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM SANEAMENTO,
MEIO AMBIENTE E RECURSOS HÍDRICOS

OTIMIZAÇÃO DO ÍNDICE DE QUALIDADE DE
ESTAÇÃO CONVENCIONAL DE TRATAMENTO DE
ÁGUA (IQETA) POR MEIO DE ANÁLISE
ESTATÍSTICA MULTIVARIADA

JOSÉLIA MARIA E SOUZA ALMEIDA

Belo Horizonte

2009

JOSÉLIA MARIA E SOUZA ALMEIDA

**OTIMIZAÇÃO DO ÍNDICE DE QUALIDADE DE
ESTAÇÃO CONVENCIONAL DE TRATAMENTO DE
ÁGUA (IQETA) POR MEIO DE ANÁLISE
ESTATÍSTICA MULTIVARIADA**

Dissertação apresentada ao Programa de Pós-Graduação
em Saneamento, Meio Ambiente e Recursos Hídricos da
Universidade Federal de Minas Gerais.

Área de concentração: Saneamento

Linha de pesquisa: Avaliação, modelagem e controle do
tratamento de águas de abastecimento.

Orientador: Marcelo Libânio

Belo Horizonte

Escola de Engenharia da UFMG

2009

AGRADECIMENTOS

A Deus por estar junto a mim todos os momentos.

Aos professores e funcionários dos departamentos de Engenharia Sanitária e Ambiental e o de Engenharia Hidráulica e Recursos Hídricos da UFMG pela disponibilidade e incentivo.

Ao meu orientador professor Marcelo Libânio, por disponibilizar seu tempo me conduzindo na realização deste trabalho.

A minha amada família: Mãe, Rodrigo, Rafaelinha, Graziela e Márcio, pelo incentivo.

Aos meus familiares por existirem e complementarem minha vida.

Aos amigos e aos companheiros de trabalho por torcerem pelo meu sucesso.

Não há grandes catástrofes para a água potável, mas é preciso manter uma grande vigilância da gestão dos sistemas de abastecimento desse precioso líquido reforçando eventualmente as restrições já existentes.

Mohamed Larbi Bouguerra

RESUMO

O presente trabalho objetivou a redução do número de parâmetros que compõem o Índice de Qualidade de Estação de Tratamento de Água (IQETA), aplicado a estações de tratamento convencionais dotadas de decantadores de escoamento horizontal, visando facilitar sua aplicação em escala real.

O índice, inicialmente composto por 19 parâmetros, foi simplificado por meio de critérios técnicos e da ferramenta de estatística multivariada, denominada Análise Fatorial. Esta técnica estatística permite o tratamento de diversas variáveis simultaneamente, mesmo quando não se conhece o modelo teórico das relações entre as mesmas, e sugere um critério técnico de seleção de variáveis a ser aplicado para que o modelo se torne o mais parcimonioso possível.

Os seguintes procedimentos foram realizados na Análise Fatorial: formulação do problema, construção da matriz de correlação, determinação do método de análise fatorial, determinação do número de fatores, rotação dos fatores, interpretação dos fatores e cálculo das cargas fatoriais ou escolha de variáveis substitutas.

O universo amostral deste trabalho abrangeu um conjunto de 21.600 observações horárias de vazão afluente de dez estações de tratamento de água e, por conseguinte de variáveis intervenientes ao processo de potabilização de água, obtidas no período chuvoso, comumente compreendido entre novembro a março. Contudo, optou-se por trabalhar com os valores médios diários, o que condicionou um conjunto de 909 dados diários, referente às 10 estações amostradas, para cada uma das 15 variáveis.

As variáveis submetidas à Análise Fatorial referem-se às 19 originais, excetuando-se as duas relativas à etapa de mistura rápida (gradiente de velocidade e tempo de detenção) e duas referentes à etapa de desinfecção (tempo de detenção e número de chicanas no tanque de contato), excluídas por critérios técnicos. À exceção do *Grau de Instrução e Jar test*, as demais variáveis relacionam-se direta ou indiretamente, às vazões horárias afluentes às estações integrantes da amostra. Para processar a análise fatorial foram utilizados os valores do IQETA parcial, a nota da variável observada elevada ao peso conferido por especialistas.

O novo índice (IQETA₁₂), composto por doze variáveis, simplifica ferramenta capaz de auxiliar a avaliação da performance de estações convencionais de tratamento. O coeficiente de correlação de Spearman, entre turbidez da água filtrada e este índice, apresentou valor numericamente inferior ao apresentado pelo original, contudo a mesma tendência, ressalta-se o ganho em termos de praticidade de aplicação.

ABSTRACT

This study focused on the reduction of the number of parameters that comprise the Water Treatment Plant Quality Index (WTPQI), applied to conventional treatment plants with horizontal sedimentation basins, in order to ease its application in real scale.

The index, initially composed of nineteen parameters, was optimized by means of technical criteria and multivariate statistical tool, named Factorial Analysis. This technique allows to deal simultaneously with several variables, even when the theoretical model of relations between them is unknown. It suggests that if a technical criterion for variable selection is applied the model becomes as parsimonious as possible.

The following procedures were carried out aiming the use of the Factorial Analysis: problem formulation, construction of the correlation matrix, definition of the method of the Factorial Analysis and the number of factors, spin factors, interpretation of factors and calculation of the factor loadings of variables or choice of the substitute variables.

The sampling universe of this research covered a range of 21,600 hourly flow rate values from ten water treatment plants during the rainy season, usually from November to March. Based on the flow rate values, the variables involved in the water treatment process were obtained. However, there were employed the daily flow rate average values that restricted the dataset to 909 daily data, for each of the 15 variables.

These variables submitted to Factorial Analysis refer to the 19 original ones minus two variables concerning rapid mix (velocity gradient and detention time) and two relating to the disinfection (detention time and number of compartments of the clearwell). As mentioned earlier, these variables were excluded due to technical criteria. With exception of Instruction Level and Jar test, the other variables are related directly or indirectly to the hourly flow rates. The Factorial Analysis was accomplished with the partial values of the WTPQI, i. e. the value of the observed variable powers by the weight conferred by the experts.

The new index (WTPQI₁₂), composed of just 12 parameters, becomes easier to employ a tool to help the water treatment plant performance assessing. Despite the Spearman's correlation coefficient between the filtered water turbidity and the new index value was lower than for the original index, this analysis shows the same tendency of the a higher WTPQI₁₂ is usually associated with better conditions to produce a highest treated water quality.

SUMÁRIO

1. INTRODUÇÃO	1
2. OBJETIVO GERAL	3
3. REVISÃO DA LITERATURA	4
3.1 Aspectos Relacionados à Água Potável	4
3.2 Sistema Convencional de Tratamento de Água	8
3.3 A Utilização de Indicadores	11
3.4 Avaliação do desempenho de Estação de Tratamento de Água	12
3.5 Análise de Dados por Meio de Métodos de Estatística Multivariada	18
3.6 Análise Fatorial	30
4. MATERIAL E MÉTODOS	40
4.1 Diretrizes para Análises Multivariadas e Interpretação	41
4.2 Caracterização da Natureza das Variáveis	41
4.3 Investigação da Relação entre as Variáveis	42
4.4 Exame de Observações Atípicas	42
4.5 Análise Fatorial e Avaliação do Ajuste Geral do Modelo	43
5. RESULTADOS E DISCUSSÃO	45
5.1 Análise Fatorial	52
5.2 Validação da Análise Fatorial	59
5.3 Critério de Seleção das Variáveis	60
6. CONCLUSÕES	65
7. REFERÊNCIAS	66

LISTA DE ABREVIATURAS E SÍMBOLOS

- ASCE – *American Society of Civil Engineers*
- AWWA – *American Water Works Association*
- AWWARF – *American Water Works Association Research Foundation*
- CCP - *Composite Correction Program*
- CPE – *Comprehensive Performance Evaluation*
- Copasa – Companhia de Saneamento de Minas Gerais
- Dcarr - Duração da Carreira de Filtração
- DEP – *Department of Environmental Protection*
- Gp - Gradiente de velocidade nas passagens entre câmaras
- Exp - Expansão do leito filtrante
- ETA – Estação de tratamento de água
- FPPE – *Filter Plant Performance Evaluation*
- Gcom - Gradiente de velocidade nas comportas de acesso ao decantador
- Gcor - Gradiente de velocidade através dos orifícios da cortina de distribuição
- Gf - Gradiente médio de velocidade da floculação
- Gmr – Gradiente médio de velocidade da mistura rápida
- IQA - Índice de Qualidade da Água
- IQAB – Índice de Qualidade de Água Bruta
- IQETA – Índice de Qualidade de Estações de Tratamento de Água
- Jtest - Realização de ensaios de Jar test (teste de jarros)
- Lar - Lavagem auxiliar com ar
- Lágua - Lavagem auxiliar com água
- Nc - Número de câmaras
- Nch - Número de chicanas no tanque de contato
- NMP – Número máximo permitido
- GI - Grau de instrução da equipe de operação
- Q_L - Vazão linear de coleta de água decantada
- SDWA - *Safe Drinking Water Act*
- Tc - Tempo de detenção no tanque de contato
- Tf - Tempo de floculação
- Tfilt - Taxa de filtração
- Tmr - Tempo de agitação da mistura rápida

uT – Unidade de turbidez

USEPA – *U. S. Environmental Protection Agency*

Vasc - Velocidade ascensional de lavagem

Vc - Velocidade média de escoamento no canal de água floculada

V_L - Velocidade longitudinal de escoamento

Vs - Velocidade de sedimentação ou taxa de aplicação superficial

LISTA DE TABELAS

TABELA 3.1 - Critério de pontuação para Nc	14
TABELA 3.2 - Critério de pontuação para o Grau de Instrução	16
TABELA 4.1 - Relação de parâmetros que compõe o IQETA e respectivos pesos	38
TABELA 5.1 – Redistribuição de pesos devido à eliminação de 4 variáveis	47
TABELA 5.2 - Estatísticas descritivas referentes aos valores de IQETA parcial	47
TABELA 5.3 – Matriz de correlação amostral dos valores do IQETA parcial entre as 15 variáveis	50
TABELA 5.4 – Autovalores da matriz de correlação	53
TABELA 5.5 - Análise Fatorial m=7 fatores	55
TABELA 5.6 – Análise Fatorial Rotação <i>Varimax</i>	57
TABELA 5.7 - Matriz residual MRES	58
TABELA 5.8 – Análise Fatorial rotacionada para o subconjunto de observações	59
TABELA 5.9 - Variáveis com maiores <i>loadings</i> em cada fator	60
TABELA 5.10 - Variáveis inclusas no IQETA ₁₂ com redistribuição de pesos	61
TABELA 5.11 – Cálculo do IQETA ₁₂ para a ETA I	61
TABELA 5.12 – Comparativo IQETA ₁₉ e IQETA ₁₂ para as dez estações amostradas	62
TABELA 5.13 – Resultados da correlação Spearman entre a turbidez da água filtrada e as formulações do IQETA para as estações amostradas)	64
TABELA 5.14 – Resultados da correlação de Pearson entre turbidez da água filtrada e as distintas formulações do IQETA	64

LISTA DE FIGURAS

FIGURA 3.1 – IQETA x IQAB	7
FIGURA 3.2 - Fluxograma da seqüência de ferramentas de controle de qualidade	8
FIGURA 3.3 - Verificação da remoção de turbidez em diferentes combinações de gradiente médio de velocidade e tempo de mistura rápida	9
FIGURA 3.4 - Critério de pontuação para Vs	14
FIGURA 3.5 - Critério de pontuação para Q_L	15
FIGURA 3.6 - Critério de pontuação para Tfilt	16
FIGURA 3.7 - Representação geométrica de ACP - Eixos principais com três variáveis	27
FIGURA 3.8 - Rotação fatorial ortogonal	38
FIGURA 4.1 – Fluxograma resumo da metodologia utilizada para AF.	44
FIGURA 5.1 – Histograma do IQETA parcial (Tfilt)	48
FIGURA 5.2 - Gráfico Q - $QPlot$ para a verificação da normalidade multivariada	49
FIGURA 5.3 – <i>Box Plot</i> IQETA parcial das 15 variáveis	52
FIGURA 5.4 - <i>Scree-plot</i> – autovalores da matriz de correlação	54

1. INTRODUÇÃO

Os acontecimentos ou decisões, em geral, envolvem um grande número de variáveis. As diversas ciências têm a pretensão de conhecer a realidade e de interpretar os acontecimentos, baseadas no conhecimento das variáveis consideradas importantes nestes eventos.

Para estabelecer relações, encontrar ou propor leis explicativas é necessário controlar, medir as variáveis consideradas relevantes ao entendimento do fenômeno analisado. Muitas são as dificuldades em traduzir as informações obtidas em conhecimento e isto condiciona a uma simplificação do fenômeno.

Deste modo, analisar as variáveis isoladamente e a partir desta análise fazer inferências sobre a realidade é uma tarefa que tem vantagens e desvantagens. Quando um fenômeno depende de muitas variáveis, geralmente este tipo de análise pode falhar, pois não basta conhecer informações estatísticas isoladas, também é necessário conhecer a totalidade destas informações fornecidas pelo conjunto das variáveis.

Os métodos de Análise Multivariada, assim como toda a estatística, padecem de limitações. A estatística tem uma *quasi-circularidade* pouco explorada nos textos: pesquisa-se para afirmar algo significativo sobre o universo amostral. Porém, a pesquisa só será significativa se o universo for suficientemente conhecido para apontar adequadamente as variáveis intervenientes e as condições de amostragem. A objetividade da pesquisa científica só começa depois da escolha das variáveis e das metodologias de análise, antes disto a atividade científica é completamente subjetiva (MOITA, 1998).

Este estudo apresenta o Índice de Qualidade de Estação de Tratamento de Água - IQETA, ferramenta que pode ser utilizada para aferir a performance de estações convencionais de tratamento (dotadas de decantador de escoamento horizontal). Em outros termos, a aplicação deste índice objetiva verificar a capacidade da estação de tratamento produzir água potável, ou seja, o atendimento aos requisitos estabelecidos pelo padrão de potabilidade nacional. A aplicação desta ferramenta também pode fornecer indícios de quão eficientes são as atividades dos funcionários que atuam na área operacional da estação (LIBÂNIO e LOPES, 2009).

O IQETA poderá contribuir com a sociedade em geral interpretando dados técnicos, que certamente não são de fácil entendimento de pessoas leigas no assunto, resumindo diversas informações em um valor numérico que fornece indícios da performance da estação. Logo, o

resultado alcançado com a aplicação do IQETA poderá servir de fonte de consulta da comunidade abastecida pelas águas tratadas na ETA. A aplicação freqüente do índice pode auxiliar no monitoramento da estação e contribuir com o conjunto de boas práticas operacionais que visa reduzir os riscos associados à água para consumo humano.

Em problemas que envolvem dados ambientais freqüentemente é necessária a repetição periódica de coleta e análise de dados dos sistemas em monitoramento. Neste caso é interessante reduzir o número de variáveis envolvidas para tornar o estudo das variações ambientais viável, mais práticas de serem realizadas, por um longo prazo (KUBRUSLY, 1992). Neste contexto se insere a justificativa para a tarefa de reduzir o número de variáveis que compõem o índice desenvolvido por Lopes (2005).

2. OBJETIVO GERAL

Analisar, utilizando a técnica de estatística multivariada - Análise Fatorial, os 19 parâmetros que compõem o Índice de Qualidade de Estação Convencional de Tratamento de Água (IQETA), desenvolvido por Lopes (2005) e selecionar dentre estes um subconjunto que summarize as informações principais do IQETA original, resultando em um índice com menor número de parâmetros e conseqüentemente com maior praticidade de aplicação.

3. REVISÃO DA LITERATURA

Nesta seção faz-se uma breve abordagem sobre o conceito de água potável, sistema de tratamento de água do tipo convencional e sobre a utilização de índices.

3.1 Aspectos Relacionados à Água Potável

Entende-se por água potável aquela que atende as premissas estabelecidas pela Portaria Nº 518 do Ministério da Saúde (Brasil, 2004), por conseguinte, pode ser consumida sem causar danos à saúde ou objeções de caráter organoléptico. A água deve estar isenta de determinadas substâncias químicas, radioativas e microrganismos patogênicos para ser considerada potável. E não deve trazer consigo substâncias capazes de adicionar-lhe cor, turbidez ou sabor desagradáveis, ainda que essas substâncias sejam inofensivas ao organismo humano.

A consideração referente ao parâmetro organoléptico advém do fato de que populações menos esclarecidas sobre o tema muitas vezes dão preferência à água possivelmente contaminada, mas límpida e de sabor mais agradável, que à água tratada distribuída pelo sistema de abastecimento regularizado, cujo sabor foi alterado pela remoção de parte das substâncias pré-existentes e pela adição de cloro (VIANNA, 2007).

Libânio *et al* (2005) realizaram estudo e constataram que a condição de vida da população é melhor retratada pela abrangência dos serviços de água e esgoto do que pelas reservas hídricas ativas, assim a contaminação das águas naturais representa um dos principais riscos à saúde pública, sendo amplamente conhecida a estrita relação entre a qualidade de água e inúmeras enfermidades que acometem as populações, especialmente aquelas não atendidas por serviços de saneamento.

A água distribuída às populações é utilizada para diversos fins, assim sendo, esta deve ser adequada para a lavagem de roupas e utensílios, não agressiva aos componentes dos sistemas abastecedores e das instalações hidráulicas prediais e não incrustante, especialmente em instalações de água quente.

O padrão nacional de qualidade da água para consumo humano apresentado por meio da Portaria nº518 do Ministério da Saúde (Brasil, 2004) define os parâmetros microbiológicos, físicos, químicos e radioativos para que a água distribuída aos consumidores não ofereça riscos à saúde. Dentre estes será dado um destaque ao parâmetro físico turbidez, pois será utilizado como parâmetro balizador do índice de qualidade em estudo.

Dos diversos parâmetros listados na referida Portaria, sabe-se que a turbidez constitui “parâmetro-chave” na qualidade microbiológica da água, principalmente como indicador da probabilidade da presença de cistos e oocistos de protozoários. O valor máximo permitido para água pós-filtrada, tratamento convencional ou filtração direta, é 1,0 uT e há ainda uma recomendação enfática, com vistas a assegurar a adequada eficiência de remoção de enterovírus, cistos de *Giardia spp* e oocistos de *Cryptosporidium sp.* , que se estabeleça como meta a obtenção de efluente filtrado com valores de turbidez inferiores a 0,5 uT em 95% dos dados mensais e nunca superiores a 5,0 uT.

A matéria em suspensão atua como escudo protetor e meio de suporte para organismos patogênicos, defendendo-os da ação de agentes desinfetantes utilizados nas estações de tratamento, e contribuindo para a veiculação de doenças.

Neste contexto, Teixeira *et al* (2004) e Libânio (2005) questionam a confiabilidade ao atendimento do padrão de potabilidade em termos de turbidez se este não especificar o tipo de aparelho a ser usado na leitura. Esta ausência de padronização, ou utilização de equipamentos com diferentes princípios de funcionamento, fornece distintos valores de turbidez para água filtrada, o que é incoerente quando se vislumbra o controle e vigilância da qualidade da água para atender ao padrão de potabilidade vigente no país. Em meio aos questionamentos levantados, a turbidez permanece como um parâmetro importante e mais aceito no controle do processo de tratamento, devido a critérios econômicos e facilidade de realização das análises.

Di Bernardo (2005) afirma que quanto menor o valor da turbidez, mais eficiente será a desinfecção. Sabe-se contudo, que é possível duas amostras com o mesmo valor de turbidez, ter quantidade e tamanho de partículas diferentes, e assim podem condicionar a resultados distintos no processo de desinfecção.

A referida Portaria além de apresentar o parâmetro turbidez capaz de fornecer indícios da presença de microorganismos, apresenta os parâmetros microbiológicos propriamente ditos, entretanto, Clark e Coyle (1990) e Bastos *et al* (2003) destacam um ponto importante quando salientam que reconhecidamente, não existem organismos que indiquem a presença/ausência da ampla variedade de patógenos possíveis de serem encontrados na água bruta ou tratada. Então mais uma vez é reforçada a importância da remoção da turbidez e de microorganismos a ela associados.

Burlingame *et al* (1998) afirmam que o parâmetro funciona bem como um indicador relativo do desempenho do tratamento e como um indicador grosseiro da qualidade da água. Deve-se considerar a necessidade do estabelecimento de orientações para a obtenção de dados de qualidade, de padrões para a calibração dos instrumentos e de limitações para o uso de cada sistema de medição.

Pela análise das considerações feitas por esses autores se percebe que os mecanismos de monitoramento podem apresentar falhas, e uma forma de compensar estas falhas, além é claro da busca contínua pela otimização e aprimoramento destes, é a aplicação de métodos complementares capazes de contribuir com o acompanhamento de quão adequada está a estação de tratamento de água no que diz respeito à produção de água potável, ou melhor, o conhecimento da performance da estação.

A avaliação da performance das estações de tratamento realizada pelas concessionárias de abastecimento de água no país comumente baseia-se na qualidade do efluente tratado ou, em outras palavras, no percentual do tempo de operação da unidade durante o qual o efluente atendeu às premissas estabelecidas pelo padrão de potabilidade nacional. Apesar do quão usual é este procedimento, uma análise mais abrangente do desempenho das estações de tratamento de água deveria contemplar:

- As características físicas, químicas e biológicas do afluente e sua adequabilidade à tecnologia de tratamento;
- O índice de sobrecarga da estação de tratamento expresso pela relação entre a vazão média afluente e as dimensões das unidades responsáveis pelos processos e operações unitárias inerentes à potabilização;
- A acurácia da operação.

Embora seja difícil estabelecer hierarquias absolutas em tratamento de água, a utilização de ferramentas que permitam avaliar a tratabilidade da água bruta e a capacidade da estação fará com que se evidencie a qualidade da operação (SOUZA e LIBÂNIO, 2009).

No que tange especificamente ao tratamento se insere o Índice de Qualidade de Estações de Tratamento de Água (IQETA), proposto por Lopes (2005), contemplando 19 parâmetros, a quase totalidade (17) constituída de parâmetros hidráulicos relacionando a vazão afluente às características físicas da instalação potabilizadora. Desta forma, a partir da vazão afluente e

das características físicas da estação, atribui-se por meio de um produtório uma nota entre 0 e 100 (LOPES, 2005), sendo que quanto mais próximo de 100 mais adequado é considerado o tratamento da água.

No que se refere ao conhecimento das características da água bruta e sua adequabilidade ao tratamento, Souza e Libânio (2009) apresentam o Índice de Qualidade da Água Bruta (IQAB), que tem a função de aferir a tratabilidade das águas naturais. Esta ferramenta aplicada em conjunto com o IQETA deverá tornar a avaliação da performance das estações ainda mais eficiente.

O IQAB abrange parâmetros comumente inseridos no padrão de potabilidade e na rotina operacional de monitoramento das águas superficiais utilizadas para fins de abastecimento, turbidez (uT), cor verdadeira (uH), pH, cianobactérias (células/mL), algas (células/mL), *E.coli* (NMP/100 mL), ferro (mg/L), manganês (mg/L). Este fato favorece seu emprego em escala real como ferramenta de avaliação da tratabilidade das águas superficiais (SOUZA e LIBÂNIO, 2009).

Assim o fato de uma determinada estação de tratamento com elevada pontuação no IQETA à qual aflui água bruta de fácil tratabilidade, apresentar efluente de má qualidade - em termos da turbidez do filtrado - constitui forte indicativo de operação insatisfatória. Em contrapartida, estação de tratamento operando com sobrecarga, por conseguinte com baixo IQETA, à qual aflui água natural de difícil tratabilidade apresentando elevada qualidade do filtrado certamente há de apresentar operação adequada. A Figura 3.1 apresentada graficamente esta análise. O grupo de pontos que representam estações operadas pela concessionária C representa um exemplo de estações com forte indício de operação insatisfatória e a concessionária B de operação satisfatória.

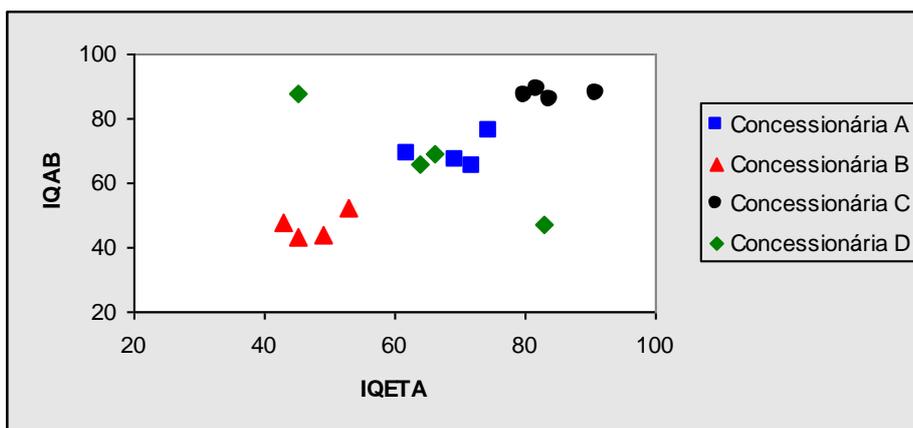


FIGURA 3.1 – IQETA x IQAB (SOUZA e LIBÂNIO, 2009).

Esta análise pode ser útil às concessionárias responsáveis pelo saneamento, pois permitirá identificar com maior exatidão as unidades com baixo desempenho e melhor alocar os recursos para ampliação e/ou treinamento de pessoal. Ilustra esta assertiva o fato da Companhia de Saneamento de Minas Gerais (COPASA) responder pela operação de 645 estações de tratamento no Estado, a quase totalidade empregando tecnologia convencional de potabilização.

A Figura 3.2 apresenta a seqüência em que as ferramentas de controle de qualidade devem ser aplicadas para analisar o grau de tratabilidade das águas superficiais, a performance da estação de tratamento e as características da água efluente.



FIGURA 3.2 - Fluxograma da seqüência de ferramentas de controle de qualidade

3.2 Sistema Convencional de Tratamento de Água

A seleção do tipo de tratamento a ser aplicado à água de abastecimento público depende de uma série de critérios que envolvem fatores ambientais, econômicos, técnicos e às vezes até fatores sócio-culturais interferem nesta escolha. Dentre os critérios técnico-ambientais o conhecimento das características da água bruta é de extrema importância. Relacionar e quantificar os tipos e as concentrações dos possíveis contaminantes presentes na água é necessário para avaliar a capacidade do sistema de tratamento adequar esta água para o consumo humano.

O índice em estudo é aplicável a estações de tratamento convencional dotados de decantador de escoamento horizontal. A seguir apresentam-se algumas considerações sobre cada etapa deste tipo de tratamento.

Na etapa de mistura rápida as partículas em suspensão na massa líquida, cuja remoção se pretende efetuar, entra em contato com agentes químicos, com o objetivo de desestabilizá-las, para que, em fases posteriores do tratamento, sejam aglutinadas umas às outras, formando flocos que serão removidos por sedimentação e filtração.

Quando o tratamento é do tipo convencional a coagulação é realizada preferencialmente pelo mecanismo de varredura (*sweep coagulation*), devido a altas dosagens de coagulante e a faixa de pH de coagulação resultante (PACKMAN, 1965 *apud* LIBÂNIO, 2005).

Amirtharajah e Mills (1982) estudaram a influência da variação do gradiente de velocidade de mistura rápida na eficiência de remoção da turbidez e chegaram à conclusão de que quando o mecanismo de coagulação predominante é a varredura o comportamento destes parâmetros pode ser considerado menos relevante do que para a filtração direta. Esta constatação também foi apresentada por Libânio (2005), conforme apresenta a Figura 3.3.

A NBR-12216 (ABNT, 1992) recomenda que o valor do gradiente médio de velocidade na mistura rápida deve ser o maior possível economicamente, e nunca inferior a 1000s^{-1} , enquanto que o tempo de detenção deverá ser inferior a 1 segundo, preferencialmente menor que meio segundo. Quando a mistura rápida se realiza por meio de mecanismo hidráulico, tempos de mistura superiores a 5s são desnecessários, e poderão até mesmo ser contraproducentes, permitindo a retromistura do floculante, o que poderá reduzir a eficiência desta etapa do tratamento de água.

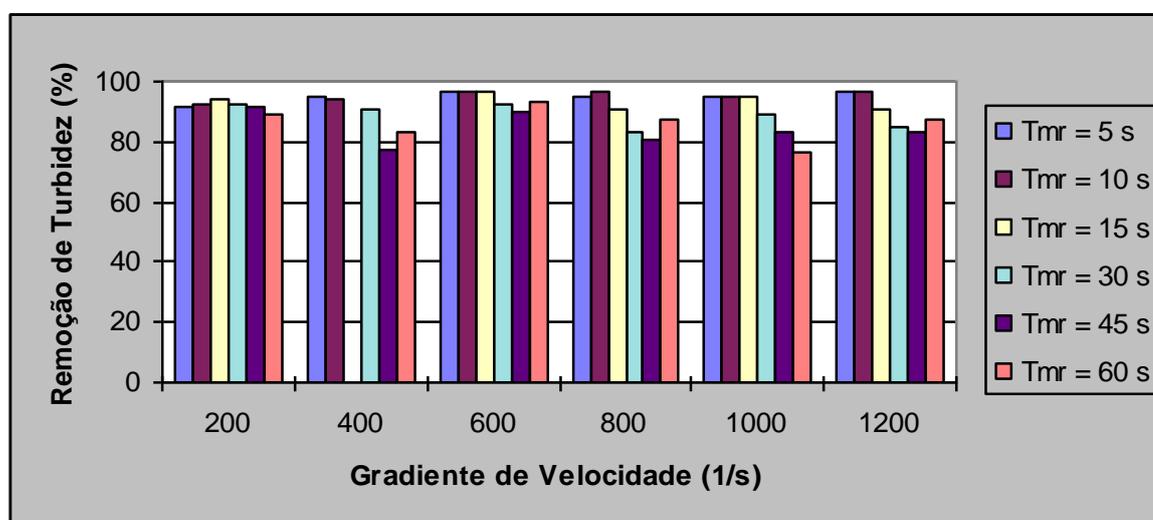


FIGURA 3.3 - Verificação da remoção de turbidez em diferentes combinações de gradiente médio de velocidade e tempo de mistura rápida (LIBÂNIO, 2005).

Para que ocorra a floculação das partículas dispersas na água é necessário a aplicação de gradiente de velocidade médio à massa de água para permitir o contato entre partículas, visando a agregá-las em flocos. Com o transcorrer do tempo, à medida que estes se formam, o gradiente de velocidade médio deve ser reduzido, para minimizar a possibilidade de ruptura, considera-se ainda um gradiente mínimo para evitar a deposição de flocos nas últimas câmaras da unidade. A um dado tempo de floculação associa-se um gradiente de velocidade ótimo que possibilita melhorar a eficiência de remoção de cor ou turbidez (MENDES, 1989 *apud* Di Bernardo, 2005).

A NBR-12216 (ABNT, 1992) determina que os valores dos gradientes de floculação e tempo de floculação devem ser obtidos por meio de ensaios laboratoriais e na ausência destes deve-se utilizar os seguintes valores de referência: tempo de detenção entre 20 e 30 min. para floculadores hidráulicos e entre 30 e 40 min. para os mecanizados, gradiente de velocidade máximo, no primeiro compartimento, de $70s^{-1}$ e mínimo, no último, de $10s^{-1}$.

Subseqüente à unidade de floculação encontra-se a de decantação. Nesta unidade, em decorrência da ação da gravidade, ocorre a sedimentação dos flocos, propiciando a clarificação da água. Devido às diferenças de tamanho, densidade e forma das partículas os fenômenos envolvidos na decantação apresentam grande dificuldade de serem descritos. Com isso o modelo do decantador ideal de escoamento horizontal é utilizado para se prever o comportamento das partículas (HAZEN, 1904 *apud* DI BERNARDO, 2005). Contudo, é sabido que nas estações de tratamento convencional a sedimentação “ideal” não ocorre, devido a situações adversas, porém admite-se que estas são compensadas por fatores como o aumento no diâmetro nominal do floco e de sua massa específica, o que resulta em uma eficiência de remoção maior do que a calculada utilizando as simplificações de um decantador ideal.

A filtração possui papel fundamental na remoção dos organismos patogênicos e, devido a isto, é essencial que esta atinja um alto grau de desempenho. Segundo Di Bernardo (2005) a filtração é o principal mecanismo responsável pela produção de água condizente com o padrão de potabilidade nacional.

A capacidade de um sistema de filtração pode ser avaliada pela taxa de filtração cuja definição considera diversos fatores como o tipo e características de leito filtrante, tamanho efetivo, tamanho do maior e menor grão, coeficiente de desuniformidade, operação dos filtros

e características do afluente. A especificação do leito é a fase de maior importância no projeto dos filtros e deve ser realizada com base em estudos de unidades-pilotos para determinar as características mais adequadas do leito conforme características da água a ser tratada. A filtração remove, da água em tratamento, as partículas em suspensão que não foram retidas na decantação. Juntamente com essas partículas, a filtração remove também os microrganismos que estiverem associados a estas partículas.

Após a filtração é realizada a desinfecção com o objetivo de produzir água de consumo isenta da presença de microrganismos patogênicos, cuja inativação realiza-se por intermédio de agentes químicos e/ou físicos. A eficiência desta etapa do tratamento é condicionada por um conjunto de fatores tais como: características do desinfetante, da água, dos microrganismos a serem inativados, do tempo de contato e da instalação física da unidade.

Referente às características da água destaca-se a turbidez e a presença de matéria orgânica, freqüentemente associada à cor verdadeira, que podem funcionar como escudo ou meio de proteção dos microrganismos da ação do desinfetante, pode ainda ocorrer reações químicas formando subprodutos indesejáveis.

As características físicas da unidade onde ocorre a desinfecção influenciam na eficiência do processo, pois estão relacionadas com o tempo de contato do produto com a massa líquida, com capacidade de permitir a homogeneidade da dispersão do desinfetante entre outros. Conforme mencionado, a alta resistência dos cistos e oocistos de protozoários ao cloro, agente desinfetante mais freqüentemente utilizado nas estações brasileiras, o tempo de contato e as concentrações usuais são ineficientes para a inativação destes.

Outra consideração a ser feita referente a este processo nas estações brasileiras é que nos sistemas de pequeno e médio porte é comum esta etapa se consolidar na unidade de reservação ora devido à sobrecarga do sistema, ora devido à construção destas unidades contíguas a estação justamente com este propósito. Para estações distantes dos centros de consumo não é raro a desinfecção se realizar, ou se completar, na adutora de água tratada e em estações mais precárias pode ocorrer na própria rede de distribuição.

3.3 A Utilização de Indicadores

Chevalier *et al* (1992 *apud* Castro 2002), definem indicador como variável ligada hipoteticamente à variável estudada e que não pode ser diretamente observada. Mc Queen e

Noak (1998) definem como medida que resume as informações relevantes a um fenômeno particular, ou a uma aproximação razoável para o valor desta medida.

Segundo Mingoti (2007), muitas vezes coletam-se diversas variáveis que descrevem um fenômeno, com o intuito de construir algum índice específico relativo à sua quantificação. A função básica do índice é a sintetização em uma única variável da informação de todas as variáveis que foram medidas sobre o fenômeno. Os métodos de estatística multivariada auxiliam na construção de gráficos ou regiões de controle, que levam em consideração informações de todas as variáveis do processo monitoradas conjuntamente.

É importante ressaltar que a aplicação de qualquer técnica estatística depende de quão cuidadoso o pesquisador foi na coleta de dados, sendo necessário verificar se o processo amostral utilizado não ocasionou vícios na resposta produzida e se foi o mais adequado para o problema em questão.

Segundo Moldan e Bilharz (1997 *apud* Castro 2002), os indicadores desejáveis são aqueles que resumem ou simplificam informações, quantificando, medindo ou comunicando dados, de forma a tornar um determinado fenômeno compreensível por todos. Na opinião de Kubrusly (2001) um índice deve possibilitar ordenar um conjunto de n objetos, segundo critério definido para um conjunto de m variáveis. Para Hair *et al* (2005), o uso de indicador não deposita total confiança em uma única resposta, mas na resposta “média” ou “típica” de um conjunto de respostas relacionadas.

3.4 Avaliação do desempenho de Estação de Tratamento de Água

A aplicação do IQETA exige a coleta de dados operacionais de estação de tratamento de água e o processamento destes valores em uma formulação matemática resultando em um número que comparado com uma faixa de valores almeja verificar a performance da estação.

O IQETA em análise foi desenvolvido tendo por referência a metodologia Delphi utilizada por Brown *et al* (1970 *apud* Lopes 2005) para desenvolver o IQA – Índice de Qualidade de Água.

As equações utilizadas para o cálculo de IQETA estão reproduzidas a seguir:

$$IQETA = \sum_{i=1}^N \left(\sum_{j=1}^n W_j Q_j \right)_i \quad (3.1)$$

$$IQETA = \prod_{i=1}^N \left(\prod_{j=1}^n Q_j^{W_j} \right)_i \quad (3.2)$$

Na qual:

W_j = peso conferido a cada parâmetro;

Q_j = nota atribuída a cada parâmetro;

j = cada parâmetro incluído no índice;

i = cada grupo que constitui o índice (mistura rápida, floculação, decantação, filtração, desinfecção e operação);

n = número de parâmetros incluídos em cada grupo;

N = número total de grupos que constitui o índice.

No presente estudo será empregada somente a Equação 3.2 (equação do produtório). Devido ao fato de que por meio desta equação a pontuação ruim em uma variável altera de forma mais expressiva o valor final do índice do que utilizando a Equação 3.1 (equação do somatório).

O desenvolvimento do IQETA contou com a organização de um painel composto por 18 profissionais para realizar pesquisa de opinião com intuito de selecionar parâmetros de maior importância referentes à tecnologia convencional de tratamento de água, bem como seus respectivos pesos.

Para a definição dos critérios de pontuação (nota) dos parâmetros que compõem o índice, considerou-se o atendimento e /ou enquadramento das faixas de valores definidas como usuais pela NBR- 12216 (ABNT, 1992) e por outros valores referenciados na literatura específica da área de tratamento de água para consumo humano (LOPES, 2005). Observa-se que a referida norma técnica de projetos de estação de tratamento de água para abastecimento público carece de atualizações.

Nesta seção serão apresentados alguns destes critérios de pontuação (modo de obtenção da nota atribuída a cada parâmetro, Q_j) Para um maior esclarecimento sobre a sistemática utilizada para desenvolvimento destes e dos demais não apresentados neste estudo o leitor deverá consultar Lopes (2005).

a) Flocculação

O critério para pontuar a variável número de câmaras (N_c), Tabela 3.1, considera o grau de aproximação do tempo de flocculação real do teórico com aumento do número destas unidades. Para um maior N_c , considera-se que o escoamento aproxima-se do modelo de fluxo em pistão e há uma redução da ocorrência de curtos-circuitos.

TABELA 3.1 - Critério de pontuação para N_c

N_c	Pontos	N_c	Pontos
Floculador hidráulico $N_c \geq 8$	100	Floculador mecanizado com $N_c \geq 4$	100
Floculador hidráulico $4 < N_c < 8$	60	Floculador mecanizado com $N_c = 3$	80
Floculador hidráulico $N_c \leq 4$	10	Floculador mecanizado com $N_c = 2$	50
		Floculador mecanizado com $N_c = 1$	1

b) Decantação

O critério de pontuação para a V_s é apresentado na Figura 3.4

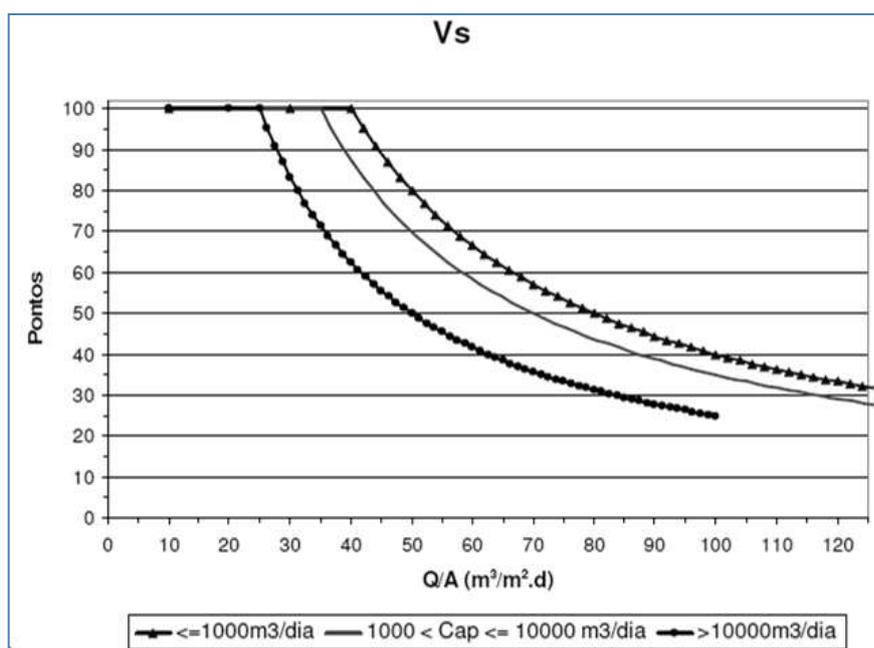


FIGURA 3.4 - Critério de pontuação para V_s

Além da zona de entrada, também a zona de saída interfere no funcionamento dos decantadores. As calhas coletoras são dispositivos utilizados para reduzir as velocidades verticais aumentando a eficiência da decantação consiste em adotar adequada área superficial na qual a água é coletada (AWWA; ASCE,1998). Se o vertedor estiver afogado a estação receberá 0 pontos para esta variável. A Figura 3.5 apresenta o critério de pontuação desta variável.

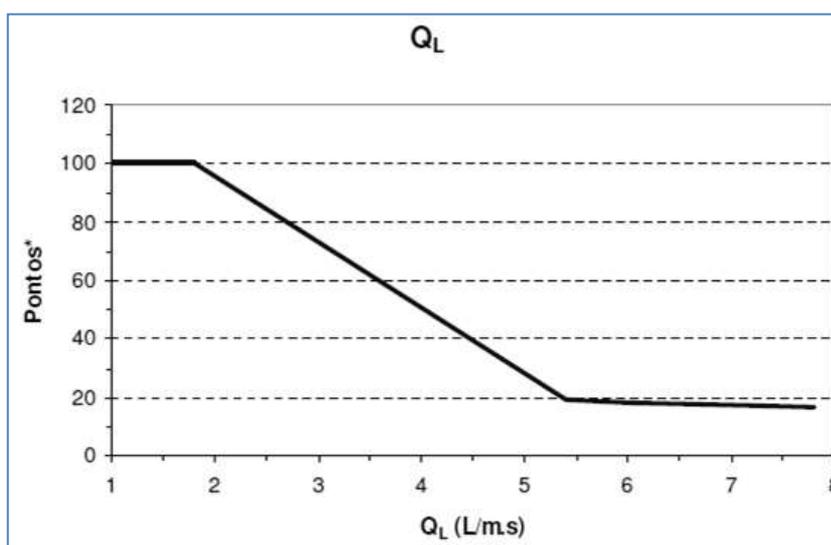


FIGURA 3.5 - Critério de pontuação para Q_L

O critério de pontuação da variável Q_L foi desenvolvido para utilizando-se um modelo de vertedor triangular, mas os estudos de Lopes (2005) apontaram que este critério de pontuação pode ser utilizado para outros tipos de calhas.

c) **Filtração**

O processo de filtração possui papel fundamental na remoção dos organismos patogênicos e, devido a isto, é essencial que este atinja um alto grau de desempenho. A Figura 3.6 apresenta o critério de pontuação para a variável taxa de filtração (Tfilt).

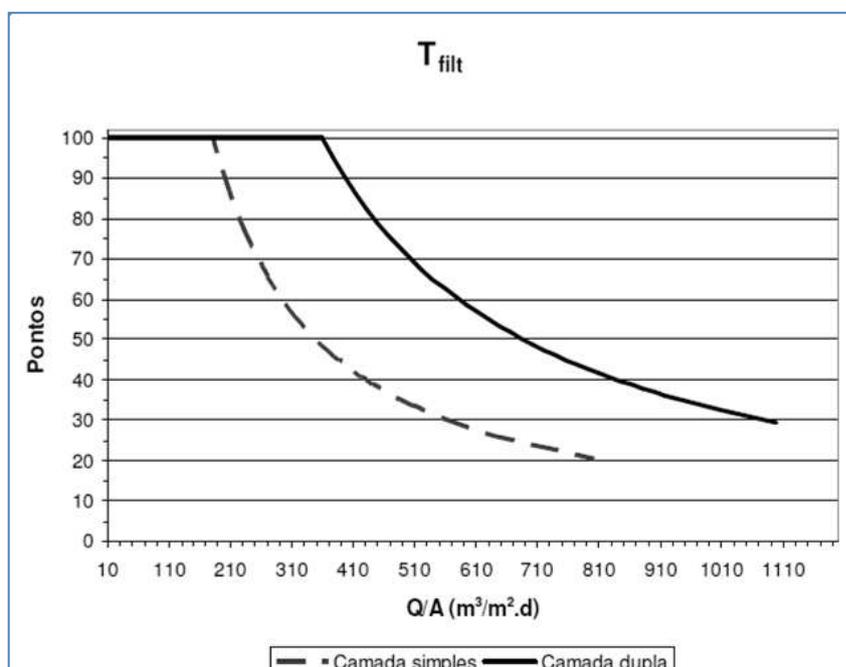


FIGURA 3.6 - Critério de pontuação para T_{filt}

d) Operação

O critério de pontuação para a variável grau de instrução (GI) é apresentada na Tabela 3.2. Para efeito desta pontuação é considerado o grau de instrução de um membro da equipe de operação, não necessariamente somente os operadores, mas qualquer funcionário com conhecimento sobre o processo de tratamento da ETA e com poderes para interferir neste realizando modificações visando à melhoria da qualidade da água produzida.

TABELA 3.2 - Critério de pontuação para o Grau de Instrução

Maior grau de instrução de qualquer um dos membros responsáveis pela operação da ETA.	Pontos
Superior em cursos da área de ciências exatas	100
Técnico em cursos da área ciências exatas	80
Superior em outras áreas ou 2º Grau completo	50
1º Grau completo	10
1º Grau incompleto	1

O ultimo critério de pontuação a ser reapresentado neste estudo refere-se ao parâmetro J_{test} . Levando-se em consideração o fato da estação realizá-lo ou não para aferir a dosagem de coagulante. Sendo que no primeiro caso a pontuação alcançada neste parâmetro e 100 e no segundo zero.

Entre as ações consideradas importantes para a garantia da qualidade da água preconizada pela Portaria do Ministério da Saúde nº. 518 (Brasil, 2004) destacam-se a gestão do ambiente,

o conhecimento da dinâmica das comunidades do ecossistema aquático e suas características físicas e químicas, assim como o emprego de técnicas adequadas de tratamento, manutenção e operação apropriada do sistema de armazenamento e distribuição de água tratada (BASTOS *et al*, 2003).

Segundo a Organização Mundial da Saúde (WHO, 2005), a qualidade da água potável está relacionada com os seguintes fatores combinados: proteção da fonte de fornecimento de água, controle dos processos de tratamento e a gestão da distribuição de água. Focando somente um ponto deste amplo mecanismo que visa à garantia de fornecimento seguro de água potável, que é a ETA, propõem-se a inclusão da ferramenta de monitoramento de performance da estação denominada IQETA.

O bom desempenho de uma ETA passa pelo conhecimento, o mais detalhado possível, de seus parâmetros operacionais e, tanto mais isto é verdade, quanto mais desfavoráveis forem as condições de operação. Por sua vez, o controle de qualidade da água deve ir além do mero monitoramento das águas afluentes e efluentes à ETA, requer o controle operacional e a avaliação de desempenho dos processos e operações unitárias de tratamento (BASTOS *et al*, 2003). Neste sentido diversos trabalhos propõem procedimentos para assegurar a qualidade da água, há ainda pesquisadores que propuseram um indicador para verificar a qualidade dos sistemas de saneamento, conforme apresentado a seguir.

Em 1984 o *Department of Environmental Protection* (DEP) do estado da Pennsylvania, para assegurar a qualidade da água distribuída, iniciou a implantação do programa FPPE – *Filter Plant Performance Evaluation* - cujo objetivo é determinar a efetividade da estação em remover partículas na mesma faixa de tamanhos de protozoários patogênicos. O programa consiste de uma visita à unidade para a avaliação física geral da ETA e dos equipamentos, bem como as práticas operacionais e a qualidade do monitoramento. Após o término das avaliações a unidade é classificada como aceitável ou inaceitável (CONSONERY *et al*, 1997).

Em 1988 iniciou-se, nos Estados Unidos e Canadá, o desenvolvimento de um método de otimização de estações de tratamento de água com respeito à proteção contra patógenos. O *Composite Correction Program* (CCP) permite que estações existentes façam uso efetivo de seus processos e operações unitárias, aumentando a eficiência das barreiras contra a passagem de microorganismos.

Em 1996 a *American Water Works Association Research Foundation* (AWWARF) promoveu um encontro de especialistas da área de tratamento e qualidade da água para discutir a otimização de estações. Como resultado foi publicado um manual, seguindo os conceitos introduzidos pelo CCP, destinado a estações aptas a promover uma auto-avaliação para alcançar uma melhora de desempenho. O processo de auto-avaliação permite aos administradores/operadores das estações identificarem as razões pelas quais o tratamento não está otimizado, a rever suas práticas e programar ações capazes de melhorar o seu desempenho.

O primeiro passo deste processo é verificar se a estação está adequadamente dimensionada podendo aumentar seu desempenho sem grandes modificações de projeto. Depois se passa à etapa de verificação das limitações de projeto como a distribuição não uniforme do escoamento pelos decantadores. Segue-se então a verificação das práticas operacionais e administrativas (AWWA, 1997).

Em 1998 *US Environmental Protection Agency* (USEPA) definiu o limite máximo de turbidez para as etapas de decantação, filtração e desinfecção com o objetivo de garantir performance ótima destas etapas.

Heller *et al* (2009) realizaram avaliação comparativa entre os serviços de saneamento em quatro municípios de Minas Gerais, por meio de uma abordagem quali-quantitativa, efetuando a agregação em um indicador do grau de qualidade dos serviços de saneamento destes municípios. A diferença deste indicador e o IQETA é a presença de parâmetros externos à unidade de tratamento de água, por exemplo, parâmetros referentes ao sistema de esgotamento sanitário.

3.5 Análise de Dados por Meio de Métodos de Estatística Multivariada

Qualquer pesquisador que examine apenas as relações entre duas variáveis e evite a análise multivariada estará ignorando poderosas ferramentas que podem dar informações potencialmente úteis (HAIR *et al*, 2005).

Análise multivariada refere-se a todos os métodos estatísticos que simultaneamente analisam múltiplas medidas sobre cada indivíduo ou objeto sob investigação. Deve-se considerar ainda que todas as variáveis devam ser aleatórias e inter-relacionadas de maneira que seus diferentes efeitos não podem ser significativamente interpretados de forma separada (HAIR *et al*, 2005).

De um modo geral as técnicas de análise multivariada são uma extensão da análise univariada (análise de distribuições de uma única variável) e de análise bivariada (classificação cruzada, correlação, análise de variância e regressão simples usada para analisar duas variáveis). Em muitos casos, técnicas estatísticas multivariadas são um meio de executar em uma única análise aquilo que exigiria múltiplas análises usando técnicas univariadas.

3.5.1 A variável estatística

O *constructo* de análise multivariada é a variável estatística, uma combinação linear de variáveis com pesos empiricamente determinados. As variáveis são especificadas pelo pesquisador, sendo os pesos determinados pela técnica multivariada para atingir um objetivo específico. Uma variável estatística de p variáveis ponderadas (X_1 até X_p) pode ser enunciada matematicamente como:

$$\text{valor da variável estatística} = c_1X_1 + c_2X_2 + c_3X_3 + \dots + c_pX_p \quad (3.3)$$

Na qual X_i é a variável observada e c_i são os pesos determinados pela técnica multivariada. O resultado é um único valor que representa uma combinação do conjunto inteiro de variáveis que melhor atinge o objetivo da análise multivariada específica.

3.5.2 Definições

As definições a seguir basearam-se nos livros de Mingoti (2007), Hair *et al* (2005) e Anderson (2003).

3.5.2.1 Vetor aleatório

Seja X um vetor contendo p componentes, no qual cada componente é uma variável aleatória, isto é, X_i é uma variável aleatória, $\forall i = 1, 2, \dots, p$. Então, X é chamado de vetor aleatório e é representado por:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$$

3.5.2.2 Vetor de médias

O vetor $\mu = E(X)$ é chamado de vetor de médias do vetor aleatório $X = [X_1, X_2, \dots, X_p]'$ no qual $\mu_i = E(X_i)$ denota a média, ou esperança, da variável aleatória X_i $i = 1, 2, \dots, p$.

A média μ_i é uma medida das mais utilizadas para sintetizar a informação de tendência central da distribuição de valores da variável X_i .

O vetor de médias μ será estimado pelo vetor de médias amostrais \bar{X} definido por:

$$\bar{X} = \frac{1}{n} [X_1 + X_2 + \dots + X_p] = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix} \quad (3.4)$$

Na qual \bar{X}_i é a média amostral da i -ésima variável, $i=1,2,\dots,p$.

3.5.2.3 Variância

A variância do i -ésimo componente do vetor X é denotada por $Var(X_i) = \sigma_i^2 = \sigma_{ii}$. O desvio padrão é representado por σ_i ou $\sqrt{\sigma_{ii}}$ e fornece a informação sobre a disposição dos valores da variável X_i em relação à μ_i , isto é, indica se os valores de X_i estão próximos ou distantes da média μ_i . Assim, valores grandes de σ_i indicam maior dispersão de valores em relação a media da distribuição.

3.5.2.4 Covariância

A covariância entre os valores da i -ésima e j -ésima variáveis do vetor X é definida pela Equação 3.5.

$$Cov(X_i, X_j) = \sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] \quad (3.5)$$

A covariância serve para medir o grau de relacionamento linear entre duas variáveis aleatórias. Porém, não há um valor de referência para julgar se este relacionamento é “forte” ou “fraco”.

3.5.2.5 Correlação

O coeficiente de correlação entre as i -ésima e j -ésima variáveis do vetor X é definido pela Equação 3.6.

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}} = \frac{\sigma_{ij}}{\sigma_i\sigma_j} \quad (3.6)$$

Por correlação entende-se o grau de relacionamento linear entre as variáveis quantitativas. Assim, quanto mais próximo de 1, mais indicação se tem de que existe um relacionamento

linear positivo entre as variáveis X_i e X_j e quanto mais próximo de -1 mais indicação se tem da existência de um relacionamento linear negativo. Uma correlação próxima de zero é uma indicação numérica de um relacionamento não linear entre as variáveis em questão. O coeficiente de correlação é adimensional, logo não sofre a influência das diferenças de escalas de medidas entre as variáveis.

3.5.2.6 Matrizes de covariâncias e correlação

Quando se têm muitas variáveis o procedimento mais comum é guardar os valores de σ_{ij} , ρ_{ij} (covariância e correlação) numa matriz denominada de matriz de covariância e correlação, respectivamente. Estas são denotadas por $\Sigma_{XX}(\Sigma)$ e $P_{XX}(P)$, respectivamente.

A matriz de covariância $\Sigma_{p \times p}$ é estimada pela matriz de covariância amostrais $S_{p \times p}$ definida por:

$$S_{p \times p} = \begin{matrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{matrix}$$

Sendo $S_{ij} = S_{ji}$, $j \neq i$, e S_{ii} definidos respectivamente por :

$$S_{ii} = \frac{\sum_{l=1}^n (X_{il} - \bar{X}_i)^2}{n - 1} \quad (3.7)$$

Que é a variância amostral da i-ésima variável,

$$S_{ij} = \frac{\sum_{l=1}^n (X_{il} - \bar{X}_i)(X_{jl} - \bar{X}_j)}{n - 1} \quad (3.8)$$

Que é a covariância amostral entre a i-ésima e j-ésima variáveis.

A matriz de correlação teórica $P_{p \times p}$ será estimada pela matriz de correlação amostral $R_{p \times p}$ definida por:

$$S_{p \times p} = \begin{matrix} R_{11} & R_{12} & \dots & R_{1p} \\ R_{21} & R_{22} & \dots & R_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ R_{p1} & R_{p2} & \dots & R_{pp} \end{matrix}$$

Na qual

$$R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} \quad (3.9)$$

É o coeficiente de correlação de Pearson.

3.5.2.7 Variância total

A variância total do vetor aleatório X é definida como traço, Equação 3.10:

$$\text{traço}(\Sigma_{p \times p}) = \text{tr}(\Sigma_{p \times p}) = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} \quad (3.10).$$

O traço da matriz de covariância ($\Sigma_{p \times p}$) é uma forma de sintetização da variância global da distribuição multivariada, uma vez que esta é a soma das variâncias de todas as variáveis envolvidas no vetor X . Altos valores de variâncias total indicam maior dispersão global das variáveis X_i , $i = 1, 2, \dots, p$.

3.5.2.8 Variância generalizada

A variância generalizada do vetor aleatório X é definida como o determinante da matriz $\Sigma_{p \times p}$, isto é $|\Sigma_{p \times p}|$. O desvio padrão generalizado de X é definido como $|\Sigma_{p \times p}|^{\frac{1}{2}}$. A variância generalizada também fornece uma noção da dispersão global da distribuição multivariada no espaço, sendo uma distribuição com maiores variabilidades globais apresentem maiores valores de variâncias generalizadas.

3.5.2.9 Combinações lineares

Seja $c_{p \times 1} \in \mathcal{R}^p$ um vetor de constantes conhecidas, isto é, $c = (c_1, c_2, \dots, c_p)'$ extraídos por meio do cálculo dos autovetores da matriz de correlação amostral. Seja Z a variável definida pela Equação 3.11.

$$Z_i = c_1 X_1 + c_2 X_2 + \dots + c_p X_p \quad (3.11)$$

Então, Z_i (Z_1, Z_2, \dots, Z_p) são variáveis aleatórias, não correlacionadas entre si e são ainda combinações lineares (função linear) das variáveis medida, estas correlacionadas entre si,

pertencentes ao vetor X . Combinações lineares desempenham um importante papel na análise de dados multivariados, pois através delas pode-se sumarizar as informações das p -variáveis originais.

3.5.2.10 Teorema da Decomposição Espectral

O Teorema da Decomposição Espectral relaciona a matriz com seus autovalores e autovetores associados como apresentado nos tópicos a seguir.

3.5.2.11 Autovalor

$$p(\lambda) = \det(P - \lambda I_p) \quad (3.12)$$

Na qual,

$p(\lambda)$: equação característica

det: determinante;

P : matriz de correlação;

λ : autovalor a ser determinado;

I_p : matriz identidade.

3.5.2.12 Autovetor

$$\hat{e}_j = \begin{bmatrix} \hat{e}_1 \\ \hat{e}_2 \\ \vdots \\ \hat{e}_p \end{bmatrix}$$

Para cada autovalor λ , os autovetores associados a λ são vetores não nulos da solução do sistema:

$$(P - \lambda I_p)e_i = \bar{0} \quad (3.13)$$

Na qual,

P : matriz de correlação;

λ : autovalor determinado pela Equação (3.12);

I_p : matriz identidade;

Para cada autovalor $\hat{\lambda}_i, i = 1, 2, \dots, m$ retido na estimação de m encontra-se o autovetor normalizado correspondente \hat{e}_i , no qual $\hat{e}_i = (\hat{e}_{i1}, \hat{e}_{i2}, \dots, \hat{e}_{ip})'$.

3.5.2.13 Distribuição de probabilidades multivariadas

Os vetores aleatórios *p*-variados têm seus valores gerados por um mecanismo probabilístico e existem várias distribuições de probabilidades multivariadas. O conhecimento da distribuição de probabilidades do vetor X torna possível propor testes estatísticos e procedimentos de estimação para os parâmetros destas distribuições.

3.5.2.14 Teste de hipótese para o coeficiente de correlação

Os testes de hipótese permitem testar a significância do coeficiente de correlação teórica entre X_i e X_j . Testes como o de Bartlett, caso a distribuição dos vetores X_i seja normal *p*-variada, ou por meio de testes não paramétricos como o de Kendall, Spearman, verificam a hipótese $H_0: \rho_{ij} = 0$ (correlações não significativas) contra $H_a: \rho_{ij} \neq 0$ (correlações significativas). Para maiores detalhes sobre os dois primeiros testes recomenda-se a leitura de Rencher (2002) e Sprent (1993).

A aplicação de alguns tipos de técnicas estatísticas multivariadas exige que as *p*-variáveis observadas sejam correlacionadas de algum modo. Das 19 variáveis que compõem o índice em estudo, 17 são diretamente ou indiretamente relacionados com a vazão afluente a ETA. Somente os parâmetros grau de instrução do operador da ETA e o parâmetro que verifica a realização ou não de ensaios de Jar test não estão relacionados com a vazão. Sabe-se, entretanto, que quanto maior o porte de uma estação, no que se refere ao volume de água tratada há uma tendência de que a equipe de operação seja mais instruída e que os ensaios de Jar teste sejam realizados.

3.5.2.15 Significância estatística versus poder estatístico

Em geral as técnicas estatísticas multivariadas são baseadas na inferência estatística dos valores ou relações de uma população entre variáveis de uma amostra aleatória extraída daquela população. A abordagem mais comum é especificar o nível do erro Tipo I, também conhecido como alfa (α). O erro Tipo I é a probabilidade de rejeitar a hipótese nula quando a mesma é verdadeira. Ao especificar um nível alfa impõem-se os limites permitidos para o

erro, especificando a probabilidade de se concluir que a significância existe quando na realidade esta não ocorre. O erro Tipo II ou beta (β) é a probabilidade de não rejeitar a hipótese nula quando na realidade esta é falsa.

3.5.2.16 Erro de medida e medida multivariada

Erro de medida é o grau em que valores observados não são representativos dos valores “verdadeiros”. Estes podem ser ocasionados por erros na entrada de dados, devido à imprecisão da medida e até a falta de habilidade de respondentes em fornecerem informações precisas. Assim todas as variáveis usadas em técnicas estatísticas multivariadas devem ser consideradas como tendo certo grau de erro de medida.

O impacto de erro de medida e confiabilidade ruim não pode ser diretamente percebido, uma vez que estão embutidos nas variáveis observadas. Portanto deve-se trabalhar com esmero para aumentar a confiabilidade e a validade que, em contrapartida, resultarão em uma descrição “mais verdadeira” das variáveis de interesse.

3.5.3 **Tipos de técnicas estatísticas multivariadas**

Análise multivariada é um conjunto de técnicas estatísticas para análise de dados sempre em expansão. Dentre as técnicas mais conhecidas estão Análise de Componentes Principais, Análise Fatorial, Regressão Múltipla, Correlação Múltipla, Análise de Discriminante Múltipla, Análise Multivariada de Variância e Covariância, Análise Conjunta, Análise de Correlação Canônica, Análise de Agrupamentos, Análise de Escalonamento Multidimensional e Análise de Correspondência (HAIR *et al*, 2005). As técnicas multivariadas podem ser organizadas em três subgrupos, a saber:

Técnicas de dependência - Pode ser definida como aquela na qual uma variável ou conjunto de variáveis é identificado como variável dependente a ser predita ou explicada por outras variáveis conhecidas como variáveis independentes. Por meio destas técnicas o pesquisador pode avaliar o grau de relação entre as variáveis dependentes e independentes. Neste subgrupo estão incluídas: Análise de Regressão Múltipla, Análise Discriminante, Análise Variância Multivariada, Análise Conjunta.

Técnicas de interdependência – Neste subgrupo nenhuma variável ou grupo de variáveis é definida como independente ou dependente. Em vez disso, o procedimento envolve a análise simultânea de todas as variáveis do conjunto. As técnicas de Análise de Componentes

Principais, Análise Fatorial, Análise de Agrupamentos e Escalonamento Multidimensional estão inclusas neste conjunto.

Técnicas avançadas e emergentes – Concentram-se as técnicas de modelagem de equações estruturais, bem como algumas técnicas emergentes nas áreas de Mineração de Dados, Redes Neurais e *Bootstrapping*.

Diante destas possibilidades Hair *et al* (2005) sugerem que a escolha da técnica apropriada deva levar em consideração as respostas às três questões apresentadas a seguir:

1. As variáveis podem ser divididas em dependentes e independentes, com base em alguma teoria?
2. Quantas variáveis serão tratadas como dependentes em uma única análise?
3. Como são medidas as variáveis, sejam dependentes ou independentes?

Ao responder as perguntas apresentadas anteriormente verifica-se que o escopo deste trabalho abrange 19 variáveis consideradas interdependentes.

As técnicas de análise estatística multivariada do subgrupo que tratam de variáveis interdependentes, Análise de Componentes Principais e Análise Fatorial serão apresentadas, resumidamente, para servir de subsídio à escolha da mais apropriada para a análise das variáveis que compõem o IQETA. Deste grupo as técnicas de Análise de Agrupamentos e Escalonamento Multidimensional fogem ao escopo deste trabalho, pois o objetivo destas técnicas é verificar a similaridade e dissimilaridade entre os dados.

3.5.3.1 Análise de componentes principais

Segundo Johnson e Wichern (2002), Anderson (2003) e Mingoti (2007), análise de componentes principais consiste em reescrever as variáveis originais em novas variáveis denominadas componentes principais, através de uma transformação de coordenadas. O valor das variáveis originais medidas e a estrutura dos dados não são alteradas por uma transformação de coordenadas.

O objetivo principal é a obtenção de um pequeno número de combinações lineares (componentes principais) do conjunto de variáveis, que retenham o máximo possível da

informação contida nas variáveis originais. A transformação das variáveis originais em componentes principais tem algumas especificidades apresentadas a seguir.

As Componentes principais são as novas variáveis geradas através de uma transformação matemática especial realizada sobre as variáveis originais. Esta operação matemática está disponível em diversos softwares estatísticos especializados. Cada componente principal é uma combinação linear de todas as variáveis originais e, portanto, mais fáceis de serem entendidas conjuntamente (MINGOTI e SILVA, 1997). A Figura 3.7 ilustra geometricamente o sistema de componentes principais.

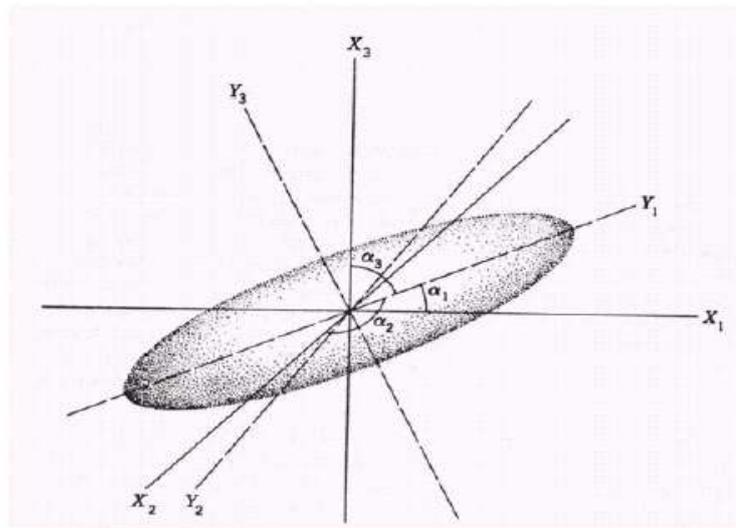


FIGURA 3.7 - Representação geométrica de ACP - Eixos principais com três variáveis

A determinação das componentes principais pode ser realizada por meio da matriz de covariância por Σ conforme Equação 3.14:

$$\hat{Y}_j = \hat{e}_j' X = \hat{e}_{j1} X_1 + \hat{e}_{j2} X_2 + \dots + \hat{e}_{jp} X_p, \quad j = 1, 2, \dots, p. \quad (3.14)$$

na qual,

$\hat{e}_1, \hat{e}_2, \dots, \hat{e}_p$ são os autovetores normalizados, correspondente aos autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ da matriz de covariâncias Σ das variáveis X_1, X_2, \dots, X_p em estudo, ou por meio da matriz de correlação P , sendo que esta minimiza os efeitos causados pelas diferenças de escala das medidas das variáveis e neste caso o Equação 3.15 permite realizar os cálculos.

$$\hat{Y}_j = \hat{e}_j' Z = \hat{e}_{j1} Z_1 + \hat{e}_{j2} Z_2 + \dots + \hat{e}_{jp} Z_p, \quad j = 1, 2, \dots, p. \quad (3.15)$$

na qual,

$\hat{e}_1, \hat{e}_2, \dots, \hat{e}_p$ são os autovetores normalizados, correspondentes aos autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ da matriz de correlação das variáveis X_1, X_2, \dots, X_p ou equivalentemente da matriz de covariâncias das variáveis Z_1, Z_2, \dots, Z_p , padronizadas pela média e desvio padrão.

Os autovetores informam o quanto cada item que constitui a componente explica a respectiva variável original.

Por exemplo, um sistema com 15 variáveis, cada uma com 909 observações, após a transformação, terá 15 componentes principais (Y_j) $j=1,2,\dots,15$. Cada uma destas componentes principais, por sua vez, será escrita como uma combinação linear das 15 variáveis originais (X ou Z). Nestas combinações, cada variável terá uma importância ou peso (estatístico) diferente valor este expresso por meio de \hat{e}_i .

Para maiores esclarecimentos de como se calcula os autovalores e autovetores por meio da decomposição espectral das matrizes de correlação ou covariância sugere-se a consulta ao livro de Santos (2000).

3.5.3.2 Análise Fatorial

Análise Fatorial é o termo genérico de um tipo de técnica estatística multivariada, cujo propósito primeiro é a redução do número de variáveis e sumarização. Ela analisa as relações entre variáveis e tenta explicá-las em termos de suas dimensões subjacentes comuns (fatores). É uma técnica de interdependência, não há explicitada uma variável dependente.

O objetivo geral da técnica é condensar a informação, contida num número de variáveis originais, em um conjunto menor de fatores com um mínimo de perda dessa informação. Em linhas gerais o que se espera é que as variáveis originais sejam agrupadas em subconjuntos de novas variáveis mutuamente não correlacionadas, sendo que a análise teria como objetivo o encontro destes fatores de agrupamento. Este número menor de parâmetros deve sumarizar as informações principais das variáveis originais. Estas novas variáveis alternativas são chamadas de fatores ou variáveis latentes.

A Análise Fatorial foi introduzida por Spearman (1904 *apud* Mingoti 2007). Muitos dos exemplos iniciais aparecem na área de psicologia e ciências sociais, mas atualmente tem sido utilizada em diversas áreas inclusive para a elaboração de índices de qualidade.

O método de Análise Fatorial implica que as variáveis originais sejam definidas como combinações lineares dos fatores que a princípio são desconhecidos.

Segundo Rencher (2002), Análise de Componentes Principais e Análise Fatorial são similares uma vez que ambos os procedimentos são usados para simplificar a estrutura de um conjunto de variáveis. No entanto, estas análises se diferem nos seguintes aspectos:

1. Na Análise de Componentes Principais, os componentes são calculados como combinações lineares das variáveis originais. Na Análise Fatorial, as variáveis originais são definidas como combinações lineares dos fatores;

2. Na Análise de Componentes Principais, o objetivo é explicar tanto quanto possível da variância total nas variáveis. O objetivo na Análise Fatorial é explicar as covariâncias ou correlações entre as variáveis;

3. Utiliza-se a técnica de Análise de Componentes Principais para reduzir os dados em um número menor de componentes. Aplica-se a técnica de Análise Fatorial para entender quais constructos representam de forma mais adequada os dados originais.

4. Nos softwares estatísticos, por exemplo no Minitab, deve-se entrar com dados brutos (valores observados de cada variável) quando se usa Análise de Componentes Principais. Pode-se entrar com dados brutos, com a matriz de correlação ou covariância ou os *loadings* (cargas fatoriais) de uma análise anterior quando a técnica de cálculo é a Análise Fatorial.

Na prática as matrizes de covariância e de correlação são estimadas pelas matrizes obtidas de dados amostrais. Existem vários métodos de estimação das componentes principais e das cargas fatoriais, dentre elas um muito utilizado que é o de componentes principais.

Pela análise das duas técnicas apresentadas verifica-se que tanto a Análise de Componentes Principais e Análise Fatorial podem ser utilizadas no processo de redução do número de variáveis que compõe o IQETA. No próximo item será detalhada a técnica de Análise Fatorial, pois esta foi escolhida para a análise dos dados tratados nesta dissertação, devido ao

fato de expressar quais constructos ou conjunto de variáveis representam de forma mais adequada o conjunto original de variáveis.

3.6 Análise Fatorial

De acordo com Hair *et al* (2005), no modelo de Análise Fatorial cada variável pode ser definida como uma combinação linear dos fatores comuns que irão explicar a parcela da variância de cada variável, mais um desvio que resume a parcela da variância total não explicada por estes fatores. A parcela explicada pelos fatores comuns recebe o nome de comunalidade, e a parcela não explicada de especificidade.

As comunalidades podem variar de 0 a 1, sendo que valores próximos de 0 indicam que os fatores comuns não explicam a variância das variáveis e valores próximos de 1 indicam que todas as variâncias são explicadas pelos fatores comuns.

Para o propósito de Análise Fatorial, existem três tipos de variância, a comum, a específica e a devida ao erro. A variância comum é definida como variância em uma variável que é compartilhada com todas as outras variáveis na análise, a específica é associada com apenas uma variável específica e a variância devido ao erro é aquela associada a não confiabilidade no processo de agrupamentos de dados, no erro de medida ou em uma componente aleatória do fenômeno medido. O conjunto das três variâncias é denominado variância total do vetor aleatório X , esta também denominada de traço.

3.6.1 Estimação de parâmetros e verificação da hipótese de normalidade multivariada

A aplicação da técnica de Análise Fatorial não exige que os dados tenham uma distribuição normal, mas o conhecimento da distribuição de probabilidades do vetor X torna possível propor testes estatísticos e procedimentos de estimação para os parâmetros destas distribuições, podendo utilizar-se de recursos de inferência estatística.

A técnica de verificação de distribuição normal multivariada deve ser aplicada quando se trabalha com duas ou mais variáveis aleatórias simultaneamente. A análise de distribuição normal univariada e a análise de distribuição normal bivariada auxiliam na verificação da suposição de normalidade multivariada (LOONEY, 1995 *apud* MINGOTI, 2007).

A suposição de normalidade univariada pode ser verificada através de gráficos como o de probabilidade normal, histograma, diagrama de ramo e folhas ou através de testes de

hipóteses como Ryan Joiner. A suposição da normalidade *bi-variada* pode ser verificada através de gráficos de dispersão do tipo $X_i \text{ vs } X_k, i \neq k$. De acordo com a teoria sobre a distribuição normal multivariada, todos os pares (X_i, X_k) devem ter uma distribuição normal bivariada.

Pela teoria, o fato de se demonstrar que todas as distribuições univariadas e bivariadas são normais não implica necessariamente que o vetor aleatório X tem distribuição normal multivariada conforme apresentado por Anderson (2003). Na prática, quando as distribuições univariadas e bivariadas são normais, a chance do vetor ser caracterizado como tendo uma distribuição aproximadamente normal multivariada é muito grande (MINGOTI, 2007).

Um gráfico que pode auxiliar na verificação de normalidade multivariada é o gráfico de probabilidade *qui-quadrado*. Pela teoria da distribuição normal multivariada tem-se que quando o número de elementos amostrais n é grande, a variável tem aproximadamente uma distribuição qui-quadrado com p graus de liberdade, no qual X_j representa os valores observados das p -variáveis do j -ésimo elemento amostral, \bar{X} é o vetor de médias amostrais e $S_{p \times p}$ é a matriz de covariância amostral.

Portanto podem-se comparar os valores das distâncias d_j^2 com as respectivas ordenadas dos percentis da distribuição *qui-quadrado* (*Q-Qplot*), conforme Equação 3.16.

$$d_j^2 = (X_j - \bar{X})' S_{p \times p} (X_j - \bar{X}) \quad (3.16)$$

Para tanto, primeiramente deve-se calcular as distâncias d_j^2 para todos os elementos da amostra e ordená-los em ordem crescente, isto é, $d_{(1)}^1 \leq d_{(2)}^2 \leq \dots \leq d_{(n)}^2$, no qual $d_{(j)}^2$ representa a j -ésima estatística de ordem j , $j = 1, 2, \dots, n$.

Em seguida deve-se traçar gráfico dos pares

$$100 \left(\left(j - \frac{1}{2} \right) / n \right) \quad (3.17)$$

(F_1, F_2, \dots, F_m) . O objetivo é identificar as novas m -variáveis, interpretá-las e calcular seus escores.

O modelo de Análise Fatorial Ortogonal é fundamentado nas seguintes suposições (apresentadas por Mingoti (2007)) necessárias para que se possa operacionalizar a estimação do modelo apresentado por meio da Equação 3.18.

1. $E[F_{mx1}] = 0$, o que implica que $E[F_j] = 0, j=1,2,\dots,m$, ou seja todos os fatores tem média igual a zero;
2. $var[F_{mx1}] = I_{mxm}$, ou seja, todos os fatores F_j são não correlacionados e tem variâncias iguais a 1;
3. $E[\varepsilon_{px1}] = 0$, o que implica que $E[\varepsilon_j] = 0, j = 1,2, \dots, p$, ou seja, todos os erros tem médias iguais a zero;
4. $var[\varepsilon_{pxp}] = \psi_{pxp}$, ou seja $var[\varepsilon_j] = \psi_j$, e $cov(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$, o que significa dizer que os erros são não correlacionados entre si e não necessariamente têm a mesma variância;
5. os vetores ε_{px1} e F_{mx1} são independentes.

Uma consequência imediata das suposições 1-5 está relacionada com a estrutura da matriz de correlação teórica P_{pxp} . Quando o modelo ortogonal é assumido, a matriz P_{pxp} pode ser reparametrizada na forma:

$$P_{pxp} = LL' + \psi \quad (3.19)$$

O objetivo da Análise Fatorial é encontrar as matrizes L_{pxp} e ψ_{pxp} para um dado valor de m , menor que o número de variáveis originais p .

O primeiro passo para conduzir a Análise Fatorial é estimar a matriz de correlação teórica P_{pxp} , através da matriz de correlação amostral R_{pxp} . Para a estimação do número m , número de fatores, deve-se extrair os autovalores da matriz R_{pxp} e ordená-los em ordem decrescente. Observam-se, então, quais autovalores são mais importantes em termos de grandeza numérica utilizando os seguintes critérios (MINGOTI, 2007):

Critério 1: a análise da proporção da variância total relacionada com cada autovalor $\hat{\lambda}_i$, dada por $(\hat{\lambda}_i/p), i = 1,2, \dots, p$. Permanecem aqueles autovalores que representam maiores

proporções de variância total e, portanto, o valor de m será igual ao número de autovalores retidos;

Critério 2: a comparação do valor numérico de $\hat{\lambda}_i$ com o valor 1, $i = 1, 2, \dots, p$. O valor de m será igual ao número de autovalores $\hat{\lambda}_i$ maiores ou iguais a 1. A idéia básica desse critério é manter no sistema novas dimensões que representem pelo menos a informação de variância de uma variável original;

Critério 3: observação do gráfico *Scree-plot* que dispõe os valores de $\hat{\lambda}_i$ ordenados em ordem decrescente. Por este critério, procura-se no gráfico um “ponto de salto”, que estaria representando um decréscimo de importância em relação à variância total. O valor de m seria então igual ao número de autovalores anteriores ao “ponto de salto”. Esse critério é equivalente ao critério 1.

3.6.3 Método de estimação de matrizes

Escolhido o valor de m , é possível estimar as matrizes L_{pxm} (matriz que armazena os *loadings*) e ψ_{pxp} (matriz que armazena a variância dos erros).

A Análise Fatorial pode ser realizada por vários métodos de estimação de matrizes, entre eles destaca-se o método de componentes principais, o método de fatores principais e o método da máxima verossimilhança, este indicado apenas quando o vetor aleatório Z tem distribuição normal multivariada. Nesta dissertação o método de componentes principais foi escolhido por não exigir informações ou suposições sobre a distribuição de probabilidades do vetor aleatório Z .

Utilizando a matriz de correlação amostral R , é possível extrair os autovalores e autovetores que servirão de base para o cálculo das estimativas das matrizes L_{pxm} e ψ_{pxp} . Em síntese, a partir da matriz R_{pxp} os autovalores de P são as raízes reais do polinômio $p(\lambda)$, apresentado a na seção de definições.

A matriz L_{pxm} será estimada por:

$$\hat{L}_{pxm} = \left[\sqrt{\hat{\lambda}_1} \hat{e}_1 \quad \sqrt{\hat{\lambda}_2} \hat{e}_2 \quad \dots \quad \sqrt{\hat{\lambda}_m} \hat{e}_m \right] \quad (3.20)$$

A idéia básica deste procedimento reside na aplicação do teorema de decomposição espectral da matriz $R_{p \times p}$. Por este teorema a matriz de correlação amostral pode ser decomposta como uma soma de p matrizes, cada relacionada com um autovalor da matriz $R_{p \times p}$. Para um valor m fixo tem-se que:

$$R_{p \times p} = \sum_{i=1}^m \hat{\lambda}_i \hat{e}_i \hat{e}_i' + \sum_{i=m+1}^p \hat{\lambda}_i \hat{e}_i \hat{e}_i' \quad (3.21)$$

Assim uma aproximação da matriz $\hat{L}\hat{L}'$ será dada por:

$$\hat{L}\hat{L}' = \sum_{i=1}^m \hat{\lambda}_i \hat{e}_i \hat{e}_i' = \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1 & \sqrt{\hat{\lambda}_2} \hat{e}_2 & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\hat{\lambda}_1} \hat{e}_1' & \sqrt{\hat{\lambda}_2} \hat{e}_2' & \dots & \sqrt{\hat{\lambda}_m} \hat{e}_m' \end{bmatrix} \quad (3.22)$$

Para construir a matriz $\hat{\Psi}_{p \times p}$, pode-se considerar a utilização da seguinte matriz:

$$\sum_{i=m+1}^p \hat{\lambda}_i \hat{e}_i \hat{e}_i' = R_{p \times p} - \hat{L}_{p \times m} \hat{L}'_{m \times p} \quad (3.23)$$

Sendo $\hat{\Psi}$ estimada como matriz diagonal. Sendo $\text{diag}(\hat{\Psi}) = \text{diag}(R - \hat{L}\hat{L}')$.

A matriz de correlação amostral original $R_{p \times p}$ estará sendo aproximada por $R \approx \hat{L}\hat{L}' + \hat{\Psi}$

A matriz residual proveniente do ajuste do modelo fatorial será dada por:

$$MRES = R - (\hat{L}\hat{L}' + \hat{\Psi}) \quad (3.24)$$

A matriz residual é utilizada como critério para a avaliação da qualidade de ajuste do modelo fatorial. Idealmente, seus valores deveriam ser próximos de zero. Entretanto, essa matriz somente é nula quando o valor de m é igual a p , o que na prática não é a solução desejada.

Através desse procedimento de estimação, os elementos diagonais da matriz $R_{p \times p}$ são reproduzidos exatamente pela matriz $(\hat{L}\hat{L}' + \hat{\Psi})$, mas o mesmo pode não ocorrer para os outros elementos da matriz $R_{p \times p}$, que representam as covariâncias das variáveis Z_i e Z_j ou equivalentemente às correlações entre as variáveis X_i e X_j , $i \neq j$, $i, j = 1, 2, \dots, p$.

Quando o método das componentes principais é usado para a estimação das matrizes $L_{p \times m}$ e $\Psi_{p \times p}$, a proporção da variância explicada pelo fator F_j reduz-se a (λ_j / p) (JOHNSON e WICHERN, 2002). Este valor representa o quanto cada fator consegue captar da variabilidade original das variáveis Z_i $i = 1, 2, \dots, p$.

3.6.4 Estimação dos escores dos fatores para cada elemento amostral

Após a identificação e interpretação dos fatores F_j $i = 1, 2, \dots, m$, relacionados com as variáveis Z_i $i = 1, 2, \dots, p$, é necessário calcular os escores (valores numéricos) para cada elemento amostral, de modo a utilizar esses valores para outras análises de interesse. Para cada elemento amostral k , $k = 1, 2, \dots, n$, o seu escore do fator F_j é calculado como descrito a seguir:

$$\hat{F}_{jk} = w_{j1}Z_{1k} + w_{j2}Z_{2k} + \dots + w_{jp}Z_{pk} \quad (3.25)$$

No qual $(Z_{1k}, Z_{2k} \dots Z_{pk})$ são os valores observados das variáveis padronizadas Z_i para o k -ésimo elemento amostral e os coeficientes w_{ji} , $i = 1, 2, \dots, p$ são os pesos de ponderação de cada variável Z_i no fator F_j . Estes coeficientes podem ser obtidos por vários métodos, entre eles o de mínimos quadrados para a estimação dos escores dos fatores.

Na Equação (3.25) todas as variáveis originais aparecem ponderadas por um coeficiente, no entanto, as mais correlacionadas com o fator aparecem com coeficientes numericamente maiores. Uma forma alternativa de calcular os escores é manter na combinação linear apenas as variáveis mais correlacionadas com o fator.

3.6.5 Rotação ortogonal de fatores

Em alguns casos a interpretação dos fatores originais F_1, F_2, \dots, F_m pode não ser uma tarefa fácil devido à aparição de coeficientes \hat{I}_{ij} de grandeza numérica similar, e não desprezível, em vários fatores diferentes. Quando isto ocorre, a suposição de ortogonalidade dos fatores está sendo violada e a partição das variáveis originais em m grupos não é clara ou difícil de ser justificada. Em situações como estas, pode-se utilizar o recurso da transformação ortogonal dos fatores originais, na tentativa de alcançar uma estrutura mais simples de ser interpretada. A rotação ortogonal preserva a orientação original entre os fatores, mantendo-os perpendiculares após a rotação.

Para facilitar a interpretação dos fatores, o ideal seria usar uma transformação que fizesse com que os *loadings* de cada característica Z_i , $i = 1, 2, \dots, p$ tivessem um valor numérico grande em somente um dos fatores e valores pequenos (ou moderados) nos fatores restantes (THUSTONE, 1947 *apud* MINGOTI, 2007).

Em termos da qualidade do ajuste esta nova solução não acrescenta melhorias em relação ao ajuste obtido usando-se a matriz $\hat{L}_{p \times m}$, pois a matriz residual ortogonal não é alterada pela transformação ortogonal. Assim quando a solução sem rotação é de boa qualidade, não se recomenda o uso de rotação ortogonal, pois a tendência é a de que a solução rotacionada seja de pior qualidade que a original.

Há diversos critérios para se proceder a rotação de matrizes como o critério *varimax*, *quartimax*, *orthomax* disponíveis nos softwares estatística especializados.

O critério *varimax* será o método utilizado nesta dissertação. Este método, em linhas gerais, busca a matriz rotacionada $T_{m \times m}$ com base na tentativa de encontrar fatores com grandes variabilidades no *loadings*, isto é, encontrar para um fator fixo, um grupo de variáveis Z_i altamente correlacionadas com o fator e outro grupo de variáveis Z_j altamente desprezível ou moderada com o fator.

A rotação preserva a orientação original entre os fatores, mantendo-os perpendiculares após a rotação.

A partir de uma solução F é sempre possível obter-se uma nova solução $F^* = T'F$, que poderá vir a ser de mais fácil interpretação do que a solução original.

Para facilitar a interpretação dos fatores, o ideal seria uma transformação que fizesse com que os *loadings* de cada característica $Z_i, i = 1, 2, \dots, p$, tivessem um valor numérico grande ($\geq 0,70$ neste estudo) em somente um dos fatores e valores pequenos (ou moderados) nos fatores restantes (MINGOTI, 2007). Desse modo, as variáveis originais Z_i ficariam divididas em grupos, no qual cada um estaria mais correlacionado com somente um fator. A Figura a seguir ilustra um exemplo de rotação aplicada aos fatores I e II

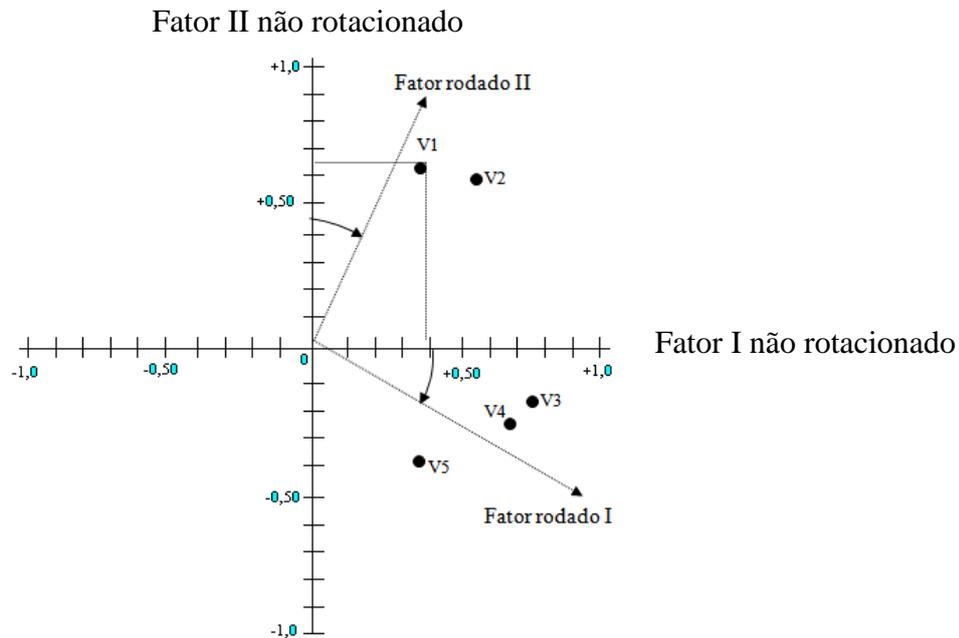


FIGURA 3.8 - Rotação fatorial ortogonal (Hair *et al*, 2005)

O método *varimax* foi proposto por Kaiser (1958 *apud* Mingoti, 2007). Para cada fator fixo, a solução é obtida através da maximização da variação dos quadrados dos *loadings* originais. Seja \hat{l}_{ij} o coeficiente da *i*-ésima variável no *j*-ésimo fator após a rotação. Seja *V* a quantidade definida por:

$$V = \frac{1}{p} \sum_{j=1}^m \left[\sum_{i=1}^p \hat{l}_{ij}^4 - \frac{1}{p} \left(\sum_{i=1}^p \hat{l}_{ij}^2 \right)^2 \right] \quad (3.26)$$

no qual $\hat{l}_{ij} = (\hat{l}_{ij}^* / \hat{h}_i)$, sendo \hat{h}_i a raiz quadrada da *comunalidade* da variável Z_i $i = 1, 2, \dots, p$. A quantidade *V* é proporcional à soma das variâncias dos *loadings* escalonados \hat{l}_{ij} ao quadrado de cada fator. O escalonamento dos *loadings* é feito pelo fato das variáveis Z_i não terem necessariamente, *comunalidades* iguais ou semelhantes.

O critério *varimax* seleciona os coeficientes \hat{l}_{ij} que maximizam a quantidade de *V*. os coeficientes finais da matriz transformada $L^* = \hat{L}T$ são $\hat{l}_{ij}^* = \hat{l}_{ij} \hat{h}_i$, $i = 1, 2, \dots, p$.

3.6.6 Validação do modelo de Análise Fatorial

O modelo de Análise Fatorial está fundamentado em suposições que não podem ser verificadas a priori, como linearidade e independência entre fatores. Além disso, a interpretação dos fatores é centrada na informação contida na matriz $\hat{L}_{p \times m}$ estimada a partir da escolha prévia do valor de *m*. Nesse sentido, torna-se importante avaliar até que ponto a matriz estimada

\hat{L}_{pxm} está representando corretamente a relação existente entre as variáveis originais e os fatores do modelo.

Para validar a solução de uma Análise Fatorial o ideal é ter dois bancos de dados separados, uma para a estimação das cargas fatoriais e outro para a validação. Se as cargas fatoriais obtidas com o banco de validação forem similares forem similares aos obtidos com o banco de dados inicial, então o modelo de AF pode ser generalizado, caso contrário não.

4. MATERIAL E MÉTODOS

Nesta seção, apresenta-se a metodologia utilizada para reduzir o número de parâmetros que compõem o IQETA, visando tornar sua aplicação mais simples.

Este estudo foi realizado com dados operacionais diários de 10 estações que utilizam tecnologia convencional de tratamento de água dotadas de decantadores de escoamento horizontal, localizadas no estado de Minas Gerais, que tratam vazão média superior a 50L/s.

As análises foram realizadas utilizando dados referentes ao período chuvoso (novembro a março), pois correspondem aos momentos em que é mais requisitada a capacidade operacional de uma estação de tratamento de água. Para o processamento das análises utilizou-se o software MINITAB 14.1 (2003).

O IQETA desenvolvido por Lopes (2005) doravante será denominado IQETA₁₉ para não confundir o com o índice com número reduzido de variáveis. As variáveis que compõem o IQETA₁₉, agrupados de acordo com a fase do tratamento que se enquadram estão apresentadas na Tabela 4.1.

TABELA 4.1 - Relação de parâmetros que compõe o IQETA e respectivos pesos.

Grupo	Parâmetro	Peso
Mistura Rápida	Gmr - Gradiente de velocidade de mistura rápida	0,06
	Tmr - Tempo de agitação de mistura rápida	0,03
Floculação	Gf/Tf - Gradiente de velocidade da floculação / Tempo de floculação	0,14
	Gp - Gradiente de velocidade nas passagens entre câmaras	0,04
	Nc- Número de câmaras	0,03
	Vc - Velocidade média de escoamento no canal de água floculada	0,03
Decantação	Gcom - Gradiente médio de velocidade nas comportas de acesso ao decantador	0,04
	Gcor – Grad.de veloc. através dos orifícios da cortina de distr. de água floculada	0,05
	Vs - Velocidade de sedimentação (taxa de aplicação superficial)	0,08
	V _L - Velocidade longitudinal de escoamento	0,04
Filtração	Q _L - Vazão linear de coleta de água decantada	0,06
	Tfilt - Taxa de filtração	0,09
	Dcarr - Duração da carreira de filtração	0,05
	Exp/Vasc - Expansão de leito filtrante ou velocidade ascensional de lavagem	0,04
Desinfecção	Laux - Lavagem auxiliar	0,02
	Tc - Tempo de detenção no tanque de contato	0,05
	Nch - Número de chicanas no tanque de contato	0,02
Operação	Jtest- Ensaio de Jar Test	0,06
	GI - Grau de instrução da equipe de operação	0,07

Fonte: Lopes, 2005

A equação utilizada para o cálculo do IQETA, conforme já apresentada no capítulo de revisão da literatura, foi novamente reproduzida neste capítulo devido a necessidade de constante consulta e utilização.

$$IQETA = \prod_{i=1}^N \left(\prod_{j=1}^n Q_j^{W_j} \right)_i \quad (3.2)$$

Na qual

W_j = peso conferido a cada parâmetro;

Q_j = nota atribuída a cada parâmetro;

j = cada parâmetro incluído no índice;

i = cada grupo que constitui o índice (mistura rápida, floculação, decantação, filtração, desinfecção e operação);

n = número de parâmetros incluídos em cada grupo;

N = número total de grupos que constitui o índice.

4.1 Diretrizes para Análises Multivariadas e Interpretação

As técnicas de estatística multivariadas são planejadas para acomodar múltiplas variáveis na análise, contudo, um modelo conceitual foi aplicado *a priori* vislumbrando um modelo parcimonioso. O ponto de partida foi a definição do problema da pesquisa e os objetivos de análise em termos conceituais, antes de especificar quaisquer variáveis ou medidas.

Com o modelo conceitual estabelecido e a técnica estatística multivariada selecionada, seguiu-se para a fase de implementação. Definiram-se o tamanho mínimo ou desejável da amostra, tipos permitidos ou exigidos de variáveis, métodos de estimação, além de aspectos específicos.

4.2 Caracterização da Natureza das Variáveis

Para o entendimento da natureza das variáveis, a primeira análise feita foi caracterização da forma de sua distribuição. O método utilizado para realizar esta tarefa foi a plotagem de um histograma com a função de representar a frequência de ocorrências das variáveis e assim verificar a sua forma de distribuição. O segundo método utilizado foi a verificação da distribuição multivariada por meio de gráfico denominado de Qui-Quadrado (Q-Q Plot), no qual as ordenadas representam d_j^2 e as abscissas χ_p^2 , as equações necessárias para a

construção deste gráfico foram apresentadas no capítulo de revisão, quando a normalidade *p-variada* é coerente com os dados amostrais, este gráfico deve resultar em algo próximo a uma reta. Curvas diferentes da reta indicam falta de normalidade.

4.3 Investigação da Relação entre as Variáveis

O método utilizado para a investigação da relação entre duas variáveis, análise de relações bivariadas, foi o diagrama de dispersão. Os pontos nestes gráficos representam os valores das variáveis e o padrão de dispersão destes pontos representa a relação entre variáveis. Uma tendência de organização dos pontos ao longo de uma linha reta caracteriza um relacionamento linear ou correlacionado. Um conjunto curvilíneo de pontos pode denotar uma relação não-linear, a qual pode ser acomodada de muitas maneiras inclusive, um padrão aparentemente aleatório de pontos, indicando ausência de relação.

Fez em seguida a matriz de correlação entre as variáveis, esta traduz em forma numérica os padrões apresentados no diagrama de dispersão. Para o cálculo do coeficiente de correlação utilizou-se a Equação 3.6, apresentada no capítulo de revisão. Os pontos que estão alinhados próximos a linha reta apresentam maiores correlações enquanto os mais distantes menores. Esta matriz representa ainda a base para os cálculos da Análise Fatorial. Para verificar a significância desta correlação utilizou-se o teste de Spearman.

4.4 Exame de Observações Atípicas

Observações atípicas são aquelas com uma combinação única de características identificáveis como sendo notadamente diferentes de outras observações. Estas não podem ser categoricamente caracterizadas como benéficas ou problemáticas, mas devem ser vistas no contexto de análise e avaliadas pelos tipos de informação que possam fornecer. As observações atípicas podem ser decorrentes de um erro de procedimento, de um evento extraordinário, podem ser observações únicas mesmo que estando em um intervalo usual e podem ainda ser observações sem explicação aparente.

Outro aspecto importante é identificar observações atípicas que possam se tornar aparentes apenas quando os valores dos dados são separados em grupos. O método usado para essa tarefa será o gráfico de caixas, uma representação da distribuição dos dados. Os limites superior e inferior marcam os quartis inferior e superior da distribuição de dados. O comprimento da caixa é a distância entre o 25º e o 75º percentil, de forma que a caixa contém 50% dos valores centrais dos dados. A linha dentro da caixa indica a mediana. Se a mediana

se encontra próxima de um dos extremos da caixa indica assimetria naquela direção. Quanto maior a caixa, maior a dispersão das observações. As linhas que se estendem de cada caixa chamada *whiskers* representam a distância à menor e à maior observação que estão a menos de um quartil da caixa.

4.5 Análise Fatorial e Avaliação do Ajuste Geral do Modelo

A primeira ação executada nesta etapa foi avaliar as suposições subjacentes do modelo. Neste sentido analisaram-se as variáveis que compõem o IQETA₁₉ para que não fossem inseridas no modelo de cálculo aquelas conceitualmente menos relevantes, que conforme já mencionado aumentam o conjunto de dados e geralmente também aumenta a capacidade da técnica de ajustar o modelo, mas ao preço de torná-la menos generalizável a população.

Com as suposições satisfeitas seguiu-se para a análise geral do modelo multivariado e posteriormente para a avaliação de seu ajuste, para estabelecer se atinge níveis aceitáveis referente a critérios estatísticos, se identifica as relações propostas e se tem significância prática.

Nesta etapa calculou-se a matriz residual, buscando identificar se a matriz de correlação amostral foi reproduzida adequadamente pela matriz estimada, ou seja, se os valores de seus elementos se aproximam de zero. Duas medidas foram utilizadas para sintetizar a informação da matriz residual, o erro médio (*EM*) e a raiz quadrada do erro quadrático médio (*REQM*) definidos respectivamente por:

EM = média dos valores que estão acima da diagonal principal (y_i) da matriz *MRES*

$$EM = \frac{1}{n} \sum Y_i \quad (4.1)$$

REQM = raiz quadrada da média dos valores, ao quadrado, que estão acima da diagonal principal (y_i) da *MRES*.

$$REQM = \sqrt{\frac{\sum Y_i^2}{n}} \quad (4.2)$$

Para proceder a validação dos resultados seguiu-se novamente os procedimentos da Análise Fatorial com o banco de dados destinados a este fim.

O fluxograma apresentado a seguir resume a metodologia utilizada para a realização da Análise Fatorial com a finalidade de reduzir o número de variáveis que compõem o IQETA₁₉.

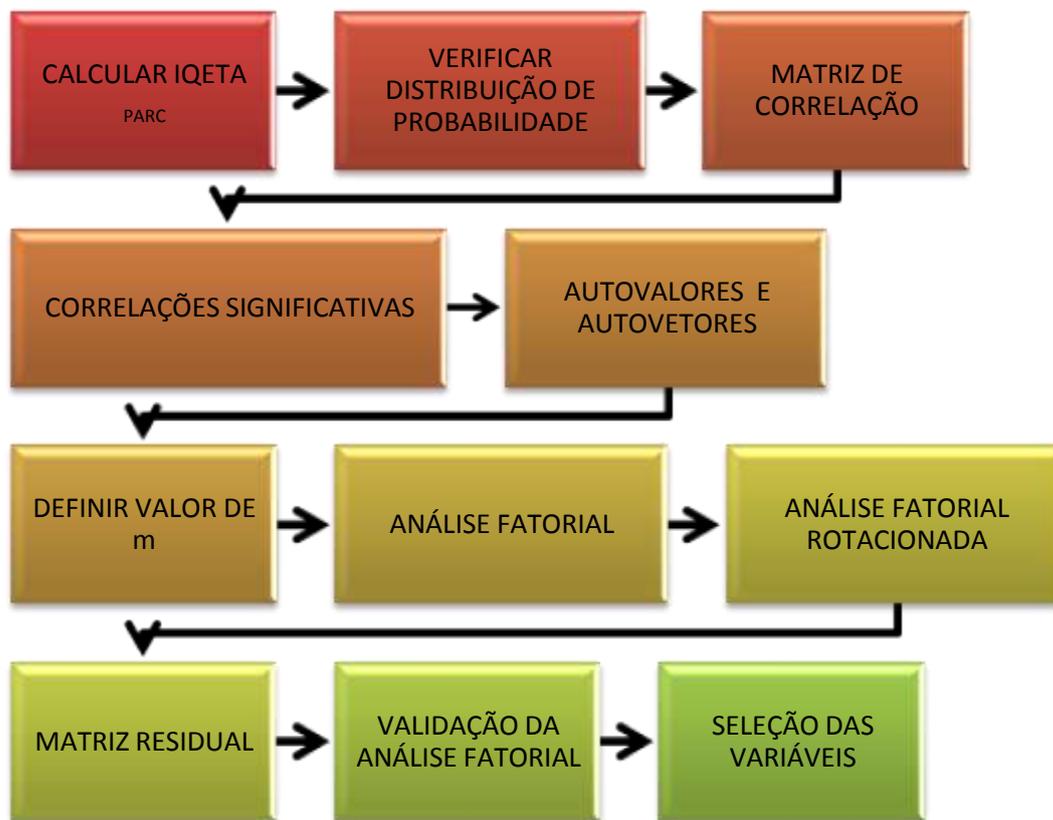


FIGURA 4.1 – Fluxograma resumo da metodologia utilizada para AF.

5. RESULTADOS E DISCUSSÃO

O presente estudo visa a redução do número de variáveis que compõem o IQETA₁₉ com o intuito de simplificar o processo de verificação da performance de estações convencionais de tratamento de água dotadas de decantador de escoamento horizontal.

A técnica de Análise Fatorial foi utilizada por permitir identificar dentre um conjunto de variáveis as mais importantes do ponto de vista estatístico. A seguir apresenta-se a sequência de verificações e cálculos utilizados para realizar esta análise.

Para tornar o modelo mais simples, no sentido de não inserir nas análises variáveis que conceitualmente (ponto de vista técnico) são menos importantes e que influenciam de forma indesejável os resultados, retiraram-se as variáveis gradiente de velocidade da mistura rápida e o tempo de agitação da mistura rápida, devido à menor relevância do comportamento destas quando o tratamento é do tipo convencional e, por conseguinte, a coagulação realiza-se pelo mecanismo de varredura, no resultado final da qualidade da água tratada conforme apresentado no capítulo de revisão de literatura.

Retirou-se ainda as variáveis referentes à etapa de desinfecção, Tc e Nch, devido ao fato de muitas estações não possuírem o tanque de contato e processar esta etapa do tratamento no reservatório e ainda considerou-se o fato de que todas as etapas anteriores a esta têm como objetivo sua otimização. Estas removem a matéria orgânica, as partículas suspensas e dissolvidas cuja presença reduziria o grau de inativação dos microorganismos e ainda remove microorganismos que sabidamente não são inativados nos tempos e concentrações do desinfetante usualmente utilizado nas estações brasileiras.

Ressalta-se que somente as ponderações feitas anteriormente não foram consideradas suficientes para justificar a retirada destas últimas variáveis, pois o índice contém outras que como as da desinfecção podem não estar presentes em algumas estações, mesmo que de tratamento convencional, mas que, foram submetidas à Análise Fatorial. O que apresentou maior impacto nesta decisão é o fato do valor da turbidez da água filtrada ser o parâmetro utilizado para balizar o índice, desta maneira considerou-se incoerente inserir variáveis que são medidas após a etapa de filtração.

As dez estações contempladas neste estudo apresentam registros horários, em um período de três meses, de valores de vazão afluente perfazendo um conjunto com 21.600 observações.

Este período, comumente compreendido entre novembro a março. Porém optou-se por trabalhar com os valores médios diários, o que condicionou um conjunto de 909 dados diários para cada uma das 15 variáveis (19 originais menos duas referentes à etapa de mistura rápida e duas referentes à etapa de desinfecção). A exceção do GI e do Jar test, as demais variáveis relacionam-se direta ou indiretamente, às vazões horárias afluentes às estações integrantes da amostra.

Para processar a Análise Fatorial, e atender a uma recomendação de Hair *et al* (2005) de que tamanho amostral (número de observações) seja no mínimo cinco vezes o número de variáveis, optou-se por trabalhar com os valores do IQETA parcial, pois se fosse utilizado somente as observações das variáveis intervenientes ao processo de tratamento da água, estaria desconsiderando as ponderações feitas pelos especialistas, na fase de criação do índice, e caso fosse utilizado somente estas ponderações o universo amostral seria insuficiente para a aplicação da técnica estatística.

O banco de dados foi alcançado por meio da utilização parcial da Equação 3.2, desta forma para cada uma das 909 observações fez-se o cálculo de sua respectiva nota elevado ao peso, obtendo assim 909 valores de $IQETA_{parc}$ para cada uma das 15 variáveis a serem analisadas.

Para a aplicação desta equação empregou-se o critério de pontuação parcialmente apresentado no item 3.4 e integralmente por Lopes (2005), e os pesos conferidos pelos especialistas, apresentados na Tabela 4.1.

$$IQETA_{parc} = Q_J^{W_J} \quad (5.1)$$

Os pesos atribuídos pelos painelistas às 19 variáveis que compõem o índice original somam 1,0 ponto. Como foram retiradas 4 variáveis cujo somatório dos pesos é 0,16, redistribuiu-se esse montante, proporcionalmente aos demais, para que o somatório continuasse 1,0, mantendo assim o grau de importância de uma variável em relação a outra, conforme apresentado na Tabela 5.1.

TABELA 5.1 – Redistribuição de pesos devido à eliminação de 4 variáveis

Parâmetro	Peso	Peso
	Original	Redistribuídos
Gmr - Gradiente de velocidade de mistura rápida	0,06	-
Tmr - Tempo de agitação de mistura rápida	0,03	-
Gf/Tf - Gradiente de velocidade da floculação / Tempo de floculação	0,14	0,17
Gp - Gradiente de velocidade nas passagens entre câmaras	0,04	0,06
Nc- Número de câmaras	0,03	0,04
Vc - Velocidade média de escoamento no canal de água floculada	0,03	0,04
Gcom - Gradiente médio de velocidade nas comportas de acesso ao decantador	0,04	0,05
Gcor – Grad.de veloc. através dos orifícios da cortina de distr. de água floculada	0,05	0,06
Vs - Velocidade de sedimentação (taxa de aplicação superficial)	0,08	0,09
V _L - Velocidade longitudinal de escoamento	0,04	0,04
Q _L - Vazão linear de coleta de água decantada	0,06	0,07
Tfilt - Taxa de filtração	0,09	0,11
Dcarr - Duração da carreira de filtração	0,05	0,05
Exp/Vasc - Expansão de leito filtrante ou velocidade ascensional de lavagem	0,04	0,05
Laux - Lavagem auxiliar	0,02	0,03
Tc - Tempo de detenção no tanque de contato	0,05	-
Nch - Número de chicanas no tanque de contato	0,02	-
Jtest- Ensaio de Jar Test	0,06	0,08
GI - Grau de instrução da equipe de operação	0,07	0,06

Os cálculos das estatísticas descritivas dos dados, apresentados na Tabela 5.2 foram a primeira análise realizada para o entendimento da natureza das variáveis.

TABELA 5.2 - Estatísticas descritivas referentes aos valores de IQETA parcial

$Q_J^{w_J}$	n	Média	Mediana	Mínimo	Máximo	Desvio padrão	Coef. de variação
Gf - Tf	909	1,75	1,77	1,38	1,90	0,10	0,06
Tfilt	909	1,50	1,52	1,41	1,52	0,03	0,02
Vs	909	1,41	1,43	1,35	1,43	0,02	0,01
GI	909	1,38	1,38	1,34	1,38	0,01	0,01
Jtest	909	1,29	1,30	1,26	1,30	0,01	0,01
QL	909	1,28	1,30	1,23	1,30	0,02	0,02
Gp	909	1,23	1,24	1,22	1,24	0,01	0,01
Dcarr	909	1,22	1,22	1,21	1,22	0,00	0,00
Gcor	909	1,21	1,26	1,00	1,26	0,10	0,08
VL	909	1,18	1,18	1,17	1,18	0,00	0,00
Vc	909	1,15	1,17	1,00	1,17	0,04	0,03
Exp ou Vasc	909	1,15	1,20	1,05	1,21	0,07	0,06
Nc	909	1,13	1,14	1,07	1,15	0,02	0,02
Gcom	909	1,11	1,20	1,00	1,22	0,10	0,09
Laux	909	1,08	1,13	1,00	1,13	0,06	0,06

A Tabela 5.2 apresenta os valores observados do IQETA parcial, ordenados em escala decrescente. Caso fosse de interesse obter o valor médio do IQETA das dez estações com o índice composto por 15 variáveis bastaria efetuar o produto dos valores da coluna que contém

os valores médios para cada variável. Continuando a análise percebeu-se que as observações de cada variável apresentam um baixo valor numérico para o desvio padrão, conforme atesta o coeficiente de variação.

Em seguida realizou-se o estudo da natureza das variáveis para verificar os pressupostos inerentes a técnica de estatística multivariada utilizada. Para o presente estudo verificou-se também se os dados possuíam distribuição normal multivariada. Para isto traçou-se histogramas e o gráfico de probabilidade *Qui-Quadrado*. A título de ilustração apresentou-se somente o histograma do IQETA parcial referente a Tfilt.

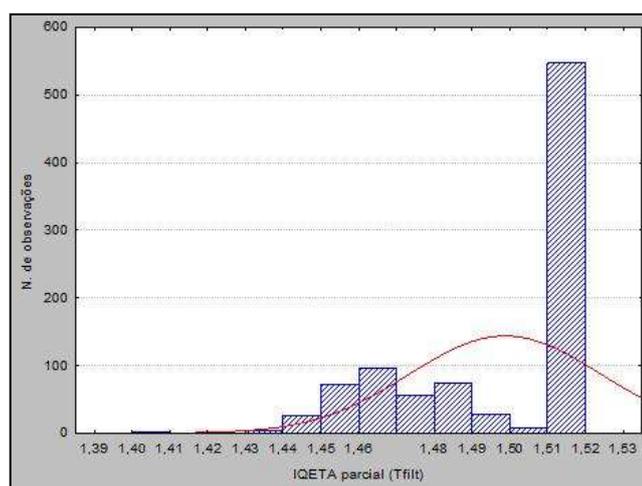


FIGURA 5.1 – Histograma do IQETA parcial (Tfilt)

Apesar desta distribuição univariada não apresentar padrão normal, sendo um forte indício de que não há distribuição normal multivariada apresentou-se o gráfico que simplifica a inspeção visual desta informação. Devido ao grande volume de dados apenas uma parcela dos valores utilizados na realização deste teste está apresentada no Anexo A.

A Figura 5.2 apresenta o gráfico de probabilidade *qui-quadrado* denominado *Q-QPlot*.

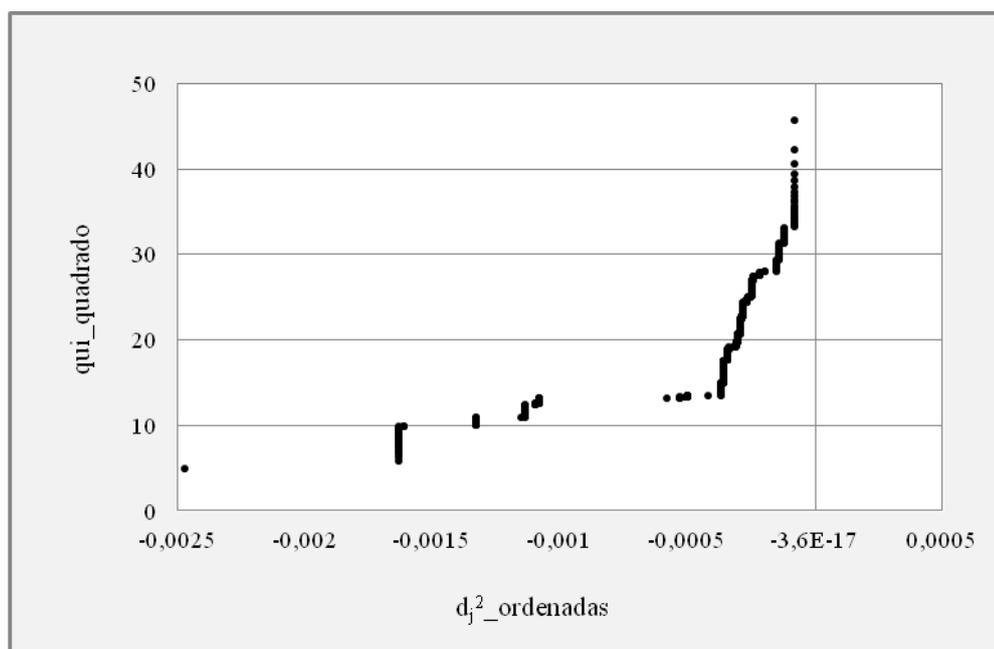


FIGURA 5.2 - Gráfico Q - *QPlot* para a verificação da normalidade multivariada

Na Figura 5.2 o eixo das ordenadas d_j^2 representa as distâncias conhecidas como *Malanahobis*, estimado por meio do vetor de médias e matriz de covariância, conforme apresentado na Equação 3.15, todos os elementos foram ordenados de forma crescente. O eixo das abscissas representa o valor esperado pelo ajuste de uma distribuição do *qui-quadrado*. Como os dados não estão dispersos em torno de uma linha reta inclinada constatou-se a ausência de normalidade multivariada das variáveis.

Posteriormente a esta análise seguiu-se para o cálculo da matriz de correlação amostral, conforme apresentado na Tabela 5.3, que demonstra a representação numérica do valor da correlação entre as 15 variáveis. A matriz de correlações é o ponto de partida para os cálculos envolvidos na Análise Fatorial.

TABELA 5.3 – Matriz de correlação amostral dos valores do IQETA parcial entre as 15 variáveis

Variável	Gf - Tf	Gp	Nc	Vc	Gcom	Gcor	Vs	VI	QL	Tfilt	Dcarr	Exp ou Vasc	Laux	GI	Jtest
Gf - Tf	1,00	0,24	-0,05	0,50	-0,17	0,51	0,28	0,25	0,01	0,31	0,03	-0,36	0,00	0,24	0,24
Gp	0,24	1,00	0,40	0,45	-0,14	-0,04	0,07	-0,32	-0,37	-0,27	0,03	0,01	0,36	0,34	0,34
Nc	-0,05	0,40	1,00	-0,04	0,47	0,00	0,65	-0,19	0,09	-0,12	0,01	-0,17	0,19	0,06	0,06
Vc	0,50	0,45	-0,04	1,00	0,14	0,45	-0,20	-0,15	-0,29	0,39	-0,02	0,28	-0,27	-0,10	-0,10
Gcom	-0,17	-0,14	0,47	0,14	1,00	0,51	0,26	-0,34	-0,17	0,53	-0,03	0,38	-0,43	-0,30	-0,30
Gcor	0,51	-0,04	0,00	0,45	0,51	1,00	0,30	-0,15	-0,30	0,61	-0,02	0,01	-0,37	-0,16	-0,16
Vs	0,28	0,07	0,65	-0,20	0,26	0,30	1,00	0,28	0,34	0,10	0,00	-0,57	0,02	0,14	0,14
VI	0,25	-0,32	-0,19	-0,15	-0,34	-0,15	0,28	1,00	0,82	0,21	-0,01	-0,28	-0,25	-0,11	-0,11
QL	0,01	-0,37	0,09	-0,29	-0,17	-0,30	0,34	0,82	1,00	-0,06	-0,02	-0,25	-0,22	-0,20	-0,20
Tfilt	0,31	-0,27	-0,12	0,39	0,53	0,61	0,10	0,21	-0,06	1,00	-0,03	0,31	-0,60	-0,26	-0,26
Dcarr	0,03	0,03	0,01	-0,02	-0,03	-0,02	0,00	-0,01	-0,02	-0,03	1,00	-0,03	0,05	0,10	0,10
Exp ou Vasc	-0,36	0,01	-0,17	0,28	0,38	0,01	-0,57	-0,28	-0,25	0,31	-0,03	1,00	-0,23	-0,29	-0,29
Laux	0,00	0,36	0,19	-0,27	-0,43	-0,37	0,02	-0,25	-0,22	-0,60	0,05	-0,23	1,00	0,46	0,46
GI	0,24	0,34	0,06	-0,10	-0,30	-0,16	0,14	-0,11	-0,20	-0,26	0,10	-0,29	0,46	1,00	1,00
Jtest	0,24	0,34	0,06	-0,10	-0,30	-0,16	0,14	-0,11	-0,20	-0,26	0,10	-0,29	0,46	1,00	1,00

Ao analisar a matriz de correlação percebeu-se que as variáveis se agrupam, de acordo com a grandeza numérica das correlações, em dois sub-conjuntos, o primeiro formado pelas variáveis Nc, Vs, Gcor, Tfilt, Vl, QL, Laux, GI, JTest e outro formado por Gf/Tf, Gp, Nc, Vc, Gcom, Dcarr.

Ressalta-se, contudo que a variável Dcarr tem correlação baixa com todas as variáveis e que por meio do teste de Spearman verificou-se que as correlações desta variável com 12 das 15 variáveis não são significativas, o que indicou a retirada desta variável do grupo das que compõem o IQETA. As demais variáveis possuem correlações significativas.

O critério de pontuação da variável Dcarr considera a variação do volume de água produzida por unidade filtrante por carreira e o volume de água utilizado para a lavagem do filtro. Percebeu-se, pela análise do banco de dados, que para as dez estações estudadas o valor desta variável permaneceu constante ao longo do período analisado.

Para a utilização da técnica de Análise Fatorial é necessário que as variáveis envolvidas tenham correlações significativas entre si, deste modo a variável duração da carreira (Dcarr) foi excluída do banco de dados. Esta exclusão pode ser confirmada pelo critério denominado de técnico já que a variável depende da taxa de filtração, pois com um incremento desta, a perda de carga e a velocidade de acumulação de partículas no leito aumentam provocando uma diminuição na carreira. No entanto esta diminuição não possui muito significado já que se a produção de água filtrada, com a qualidade desejada, aumenta com o aumento da taxa filtração, a diminuição da carreira torna-se irrelevante, esta informação contribuiu com a decisão de retirada da variável Dcarr.

Posteriormente a análise da matriz de correlação seguiu-se para a identificação de possíveis observações atípicas, empregou-se para esta análise o gráfico de caixas (*Box-Plot*), apresentados na Figura 5.3

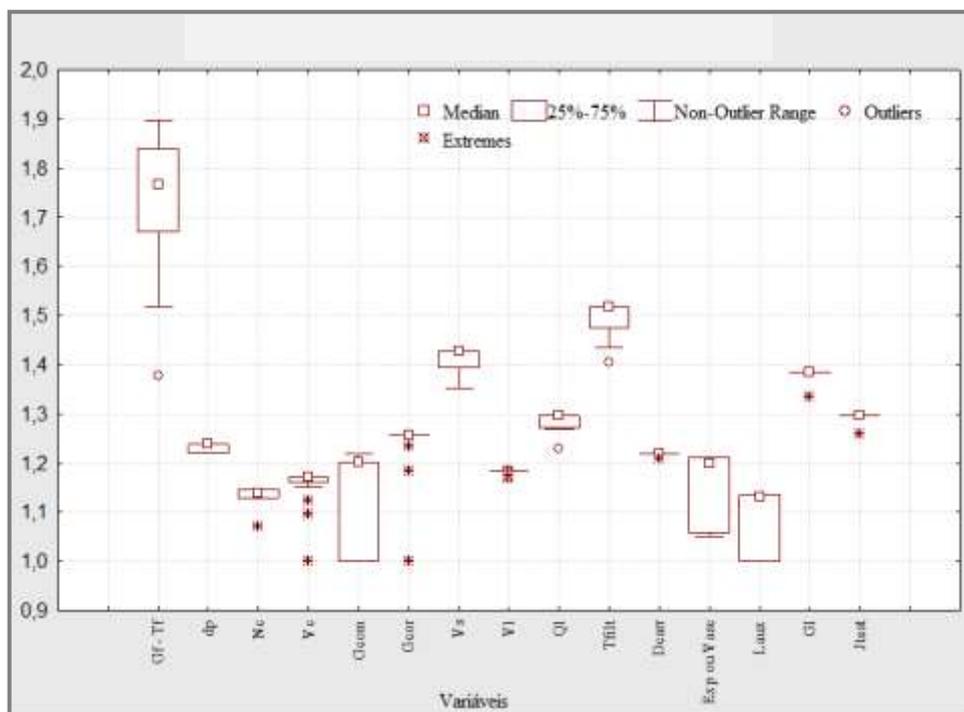


FIGURA 5.3 – Box Plot IQETA parcial das 15 variáveis

O comprimento da caixa é a distância entre o 25° e o 75° percentil, de forma que a caixa contém 50% dos valores centrais dos dados. O quadrado dentro da caixa indica a mediana. As variáveis em que a mediana se encontra próximo de um dos extremos da caixa indicam assimetria naquela direção. Sabendo-se que quanto maior a caixa, maior a dispersão das observações, percebeu-se que as variáveis Gf-Tf, Gcom, Exp/Vasc e Laux apresentaram maiores dispersões. As linhas que se estendem de cada caixa denominadas *whiskers* representam a distância à menor e à maior observação que estão a menos de um quartil da caixa.

As variáveis Gf-Tf, Gcor e Q_L apresentaram um *outlier* cada, porém, estes valores não foram retirados devido à possibilidade de não representarem erro de medida e sim uma observação extrema que ocorreu durante o período de coleta dos dados IQETAparc. Verifica-se que a mediana é o valor máximo, para uma grande parte das variáveis analisadas (IQETAparc), o que é esperado, pois a despeito da sazonalidade das características da água afluyente à ETA a sua performance deve ser constante, logo os valores (IQETAparc) devem apresentar pouca variabilidade.

5.1 Análise Fatorial

A seguir apresenta-se a Análise Fatorial feita para o conjunto de 14 variáveis, já que a Dcurr foi retirada por não atender ao pressuposto da técnica de Análise Fatorial que é de ter

correlações significativas entre as variáveis. Os cálculos foram realizados utilizando-se o software Minitab 14.1, cujo resumo do tutorial é apresentado no Anexo B – Tutorial Minitab 14.1

Para a escolha do valor de m , ou seja, do número de fatores que deverão ser considerados para a realização da Análise Fatorial, calcularam-se os autovalores e respectivos percentuais de variância da matriz de correlação amostral. Plotou-se ainda o gráfico denominado *Scree-plot* que possibilitou a visualização dos autovalores.

A Tabela 5.4 apresenta os autovalores da matriz de correlação amostral com respectivas porcentagens de variação total explicada

TABELA 5.4 – Autovalores da matriz de correlação

Fator	Autovalor	Autovalor acumulado	Percentual de variância	Percentual acumulado de variância
1	3,58	3,58	25,61	25,61
2	2,77	6,35	19,79	45,39
3	2,53	8,88	18,05	63,44
4	1,91	10,79	13,62	77,06
5	1,16	11,95	8,30	85,36
6	0,87	12,82	6,24	91,60
7	0,44	13,26	3,12	94,72
8	0,27	13,53	1,95	96,67
9	0,22	13,75	1,56	98,22
10	0,13	13,88	0,94	99,17
11	0,06	13,94	0,41	99,58
12	0,04	13,98	0,27	99,85
13	0,02	14,00	0,15	100,00
14	0,00	14,00	0,00	100,00

O autovalor é uma medida que indica a importância relativa de cada fator na explicação da variância associado ao conjunto de variáveis analisadas. Como o conjunto é composto por 14 variáveis é possível obter 14 fatores e a soma dos autovalores destes também é igual a 14. Nesta tabela verificou-se um fator com autovalor igual a 0 indicando que do conjunto de variáveis é possível eliminar uma, pois esta não é estatisticamente informativa. Analisando novamente a matriz de correlação verifica-se que as variáveis GI (grau de instrução) e Jtest (Jar Test) são perfeitamente correlacionadas, o que significa que o comportamento da informação contida em uma variável é explicada por outra, assim uma delas pode ser retirada. Deste modo retirou-se a variável Jtest (Jar Test), pois além de estar perfeitamente

correlacionada com o GI esta pode ser classificada como variável qualitativa, já que a realização do teste de jarros condiciona nota total 100 e a não realização nota 0.

Segue – se então para o estudo do número de fatores que deverão ser analisados. O banco de dados nesta fase é composto por 13 variáveis, contendo cada uma 909 observações. Um critério para auxiliar na escolha do número de fatores (m) foi reter os autovalores com maiores porcentagem de variância, acumulando no mínimo 80% de explicação e, pela inspeção da Tabela 5.4 o valor de (m) deverá ser igual a cinco.

O segundo critério utilizado para auxiliar a escolha do valor de (m) foi o de reter somente autovalores maiores ou iguais a um. Por este critério o número (m) também foi igual a cinco.

O último critério utilizado para indicar um possível valor de (m), número de fatores que deverão ser utilizados nos cálculos da Análise Fatorial, foi por meio da observação da Figura 5.4 - *Scree-plot*. Neste gráfico procurou-se identificar um “ponto de salto”, a partir do qual os valores tenderiam a se estabilizar.

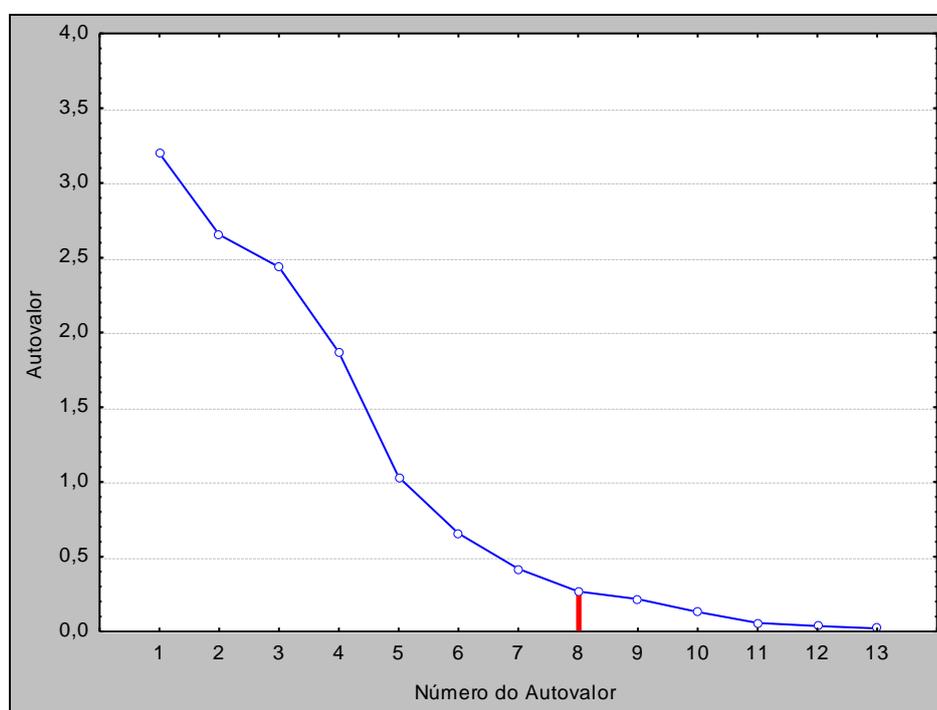


FIGURA 5.4 - *Scree-plot* – autovalores da matriz de correlação

A Figura 5.5 mostra o *Scree-plot* para os autovalores da matriz de correlação composta do banco de dados composto por 13 variáveis. A inspeção visual deste gráfico sugere a escolha de oito fatores.

Assim a qualidade do ajuste do modelo de Análise Fatorial foi avaliado para (m) igual a 5 e 8 fatores. Na Tabela 5.5 demonstram-se os resultados da Análise Fatorial adotando estas possibilidades, esta exibe o resultado da matriz fatorial não rotacionada via análise de componentes principais.

TABELA 5.5 - Resultados da Análise Fatorial

Variáveis (Z_i)	$F_1(\hat{l}_{i1})$	$F_2(\hat{l}_{i2})$	$F_3(\hat{l}_{i3})$	$F_4(\hat{l}_{i4})$	$F_5(\hat{l}_{i5})$	$F_6(\hat{l}_{i6})$	$F_7(\hat{l}_{i7})$	$F_8(\hat{l}_{i8})$	\hat{h}_i^2 (5 fat.)	\hat{h}_i^2 (8 fat.)
Gf - Tf	-0,2073	-0,0649	0,6728	0,6304	-0,0247	-0,0294	0,1464	-0,1704	0,8978	0,9492
Gp	0,1729	0,6434	0,4543	0,1033	0,4992	0,0110	-0,0818	0,2458	0,9100	0,9772
Nc	0,0621	0,1174	0,5898	-0,7061	0,3158	0,0198	0,0272	-0,0789	0,9638	0,9711
Vc	-0,5676	0,3697	0,2235	0,4617	0,4319	-0,0874	-0,0720	-0,1498	0,9085	0,9438
Gcom	-0,7085	0,0919	0,1069	-0,6434	-0,0666	0,1325	0,0000	-0,1321	0,9402	0,9752
Gcor	-0,7569	0,0589	0,4289	0,1078	-0,3344	-0,1747	0,0838	0,0006	0,8837	0,9213
Vs	0,0252	-0,3813	0,8001	-0,3885	-0,0797	0,0130	0,0157	0,1307	0,9435	0,9610
V _L	0,1021	-0,8656	0,0825	0,3061	0,2438	0,1781	0,0936	0,1403	0,9195	0,9797
Q _L	0,2068	-0,8678	0,0334	-0,0673	0,3595	0,0642	0,0293	-0,2056	0,9307	0,9779
Tfilt	-0,8464	-0,2213	0,0838	0,1238	-0,0969	0,3077	0,1219	0,1995	0,7970	0,9464
Exp ou Vasc	-0,4648	0,3432	-0,5895	-0,1025	0,3010	0,3431	0,2285	-0,0316	0,7824	0,9534
Laux	0,7188	0,4115	0,1808	-0,0107	-0,1070	0,0162	0,5038	-0,0279	0,7302	0,9851
GI	0,4715	0,3000	0,3879	0,2160	-0,2728	0,5922	-0,2192	-0,1102	0,5838	0,9947
Soma dos quadrados (autovalor)	3,1987	2,6541	2,4380	1,8670	1,0332	0,6564	0,4199	0,2683	11,1911	12,5358
Percentual do traço	0,2461	0,2042	0,1875	0,1436	0,0795	0,0505	0,0323	0,0206	0,8609	0,9643

Na Tabela 5.5 as oito colunas de fator $F_j(\hat{l}_{ij})$ exibem as cargas fatoriais (*loadings*) para cada variável em cada fator. Estes coeficientes representam a correlação entre fator e variável original. A coluna *comunalidade* (\hat{h}_i^2 5 fat) fornece informação de quão bem a variância de cada variável foi “explicada” pelos cinco primeiros fatores calculados e a (\hat{h}_i^2 8 fat) pelos oito fatores calculados.

A linha denominada de *soma dos quadrados (autovalor)* é a soma da coluna de cargas fatoriais ao quadrado (autovalores) e indica a importância relativa de cada fator na explicação da variância total associada ao conjunto de variáveis analisadas. A solução fatorial extraiu os

fatores na ordem de sua importância, com o fator 1 explicando maior parte da variância e os demais fatores gradativamente uma menor parcela de explicação.

Pode-se perceber que o primeiro fator ficou mais correlacionado (maior carga fatorial) com as variáveis Tfil, Laux, Vc, Gcom e Gcor, o segundo com VL, QL e Gp o terceiro com Gf-Tf, Vs e Exp ou Vasc, o quarto com Gf-Tf, Nc e Gcom, o quinto com Gp e Vc, o sexto fator está mais relacionado com a variável GI, o sétimo com Laux e o oitavo com Gp e QL .

Como o objetivo da Análise Fatorial é alcançar resultados com o menor número possível de fatores, desde que o nível de explicação não seja inferior ao determinado para a análise, neste caso de 80%, adotaram-se os resultados da análise utilizando 5 fatores.

A soma total de fatores ao quadrado representa a quantia total de variância extraída pela solução fatorial. A quantia total de variância explicada pela solução fatorial 11,19 pode ser comparada com a variância total no conjunto de variáveis representadas pelo traço da matriz fatorial. O traço é a variância total a ser explicada igual à soma dos autovalores do conjunto de variáveis, o valor máximo neste estudo é 13. Os percentuais de traço explicados por cada um dos cinco fatores (24,61; 20,42; 18,75; 14,36; 7,95) são mostrados na última linha de valores da Tabela 5.5. O percentual do traço foi obtido dividindo-se a soma de quadrados de cada fator pelo traço para o conjunto de variáveis analisado.

Somaram-se os percentuais do traço para cada um dos cinco fatores, e obteve-se o percentual total de traço extraído para a solução fatorial, o qual pode ser usado como um índice para determinar quão bem uma solução fatorial particular explica o que todas as variáveis juntas representam. O índice para a presente solução mostra que 86% da variância total são representados pela informação contida na matriz fatorial da solução em termos de cinco fatores. Logo o índice para essa solução foi adequado e as variáveis estão de fato estreitamente relacionadas umas com as outras.

A *comunalidade* de 0,5838 para a variável GI indica que ela tem menos em comum com as outras variáveis incluídas na análise do que a variável Nc, a qual tem uma *comunalidade* de 0,9638. Todas as outras, porém, compartilham mais da metade de sua variância com os cinco fatores.

A interpretação da Tabela 5.5 é simples porém não ficou clara, já que as cargas fatoriais de uma variável possui valores próximos em mais de um fator. Para promover uma redistribuição

de variâncias dos primeiros fatores para os últimos e assim tornar a interpretação mais clara no sentido de uma variável estar mais correlacionada com apenas um fator, ou seja, apresentar um valor numérico alto em um fator e pequeno nos demais se fez a rotação fatorial com o objetivo de resultar em um padrão fatorial mais simples e teoricamente mais significativo.

A interpretação dos fatores originais não pode ser realizada de forma direta, pois os fatores apresentaram coeficientes (\hat{l}_i) de grandeza numérica similar, e não desprezível em vários fatores diferentes. Desta forma a partição das variáveis originais em m grupos ficou difícil de ser justificada. Assim utilizou-se o recurso da transformação ortogonal dos fatores originais, na tentativa de se alcançar uma estrutura mais simples de ser interpretada. Esta preservou a orientação original entre os fatores (perpendiculares entre si).

Em termos de qualidade do ajuste, esta nova solução não acrescentou melhoria em relação ao ajuste obtido usando-se a matriz \hat{L}_{pxm} , pois a matriz residual original não foi alterada pela transformação ortogonal, como já era esperado. A matriz fatorial rotacionada *VARIMAX* é mostrada na Tabela 5.6.

TABELA 5.6 – Análise Fatorial Rotação *Varimax*

Variáveis (Z_i)	$F_1(\hat{l}_{i1})$	$F_2(\hat{l}_{i2})$	$F_3(\hat{l}_{i3})$	$F_4(\hat{l}_{i4})$	$F_5(\hat{l}_{i5})$	\hat{h}_i^2 (5 fat.)
Gf – Tf	0,3201	-0,1472	-0,0813	-0,6923	0,5365	0,8978
Gp	-0,3615	0,2993	0,3200	-0,0836	0,7618	0,9100
Nc	-0,0878	0,0541	0,9697	0,0292	0,1091	0,9638
Vc	0,3742	0,1024	-0,1219	0,0736	0,8589	0,9085
Gcom	0,6633	0,2714	0,5250	0,3686	-0,1222	0,9402
Gcor	0,8421	0,2365	0,0848	-0,2883	0,1686	0,8837
Vs	0,1921	-0,2615	0,7598	-0,4956	-0,1238	0,9435
VI	0,0449	-0,9275	-0,1251	-0,1998	-0,0410	0,9195
QL	-0,0940	-0,9288	0,1582	0,0151	-0,1840	0,9307
Tfilt	0,8736	-0,0999	-0,0821	0,0549	0,1186	0,7970
Exp ou Vasc	0,1917	0,2407	-0,2149	0,7785	0,1883	0,7824
Laux	-0,7028	0,3428	0,1061	-0,3277	-0,0019	0,7302
GI	-0,3725	0,2980	0,0152	-0,5917	0,0770	0,5838
Soma dos quadrados (autovalor)	3,0103	2,3199	2,0295	2,0640	1,7674	11,1911
Percentual do traço	0,2316	0,1785	0,1561	0,1588	0,1360	0,8609

Pela análise da Tabela 5.6 percebeu-se que a quantia total de variância extraída é a mesma da solução não rotacionada, 86%. Entretanto duas diferenças foram visíveis: primeiro a variância foi redistribuída de modo que o padrão de cargas fatoriais e o percentual de variâncias para cada fator são diferentes. Especificamente, na solução fatorial rotacionada *VARIMAX*, o primeiro fator explica 23% da variância comparado com 24,6% obtido na solução não rotacionada, de modo semelhante, o quinto fator explica 13,6% da variância, comparado com 7,95% da solução não rotacionada. Assim, o poder explicativo mudou ligeiramente para uma distribuição mais equilibrada entre os cinco fatores por causa da rotação. A segunda diferença é que a interpretação da matriz fatorial foi simplificada, pois as cargas fatoriais que mais se destacam em cada fator foram realçadas, ou seja, valor numérico alto em um único fator.

A partir do modelo de Análise Fatorial ajustado com 5 fatores, fez-se uma análise da qualidade do ajuste através da avaliação da matriz residual, conforme apresentado na Tabela 5.7.

TABELA 5.7 - Matriz residual MRES

	Gf/ Tf	Gp	Nc	Vc	Gcom	Gcor	Vs	V_L	QI	Tfilt	Exp/ Vasc	Laux	GI
Gf/ Tf	0,00	-0,04	0,03	-0,03	0,03	-0,01	-0,04	-0,02	0,03	-0,01	0,04	0,06	-0,03
Gp	-0,04	0,00	-0,03	-0,06	-0,03	0,02	0,03	0,03	-0,04	0,02	-0,01	-0,06	-0,01
Nc	0,03	-0,03	0,00	0,01	0,00	-0,03	-0,03	-0,01	0,00	0,02	-0,01	0,01	0,00
Vc	-0,03	-0,06	0,01	0,00	0,02	-0,02	-0,01	-0,02	0,02	-0,02	-0,04	-0,01	0,04
Gcom	0,03	-0,03	0,00	0,02	0,00	-0,01	-0,03	0,00	0,02	-0,03	-0,01	0,01	0,01
Gcor	-0,01	0,02	-0,03	-0,02	-0,01	0,00	0,02	0,02	0,03	-0,04	0,06	0,04	0,00
Vs	-0,04	0,03	-0,03	-0,01	-0,03	0,02	0,00	0,01	-0,02	0,01	0,03	0,01	-0,01
V_L	-0,02	0,03	-0,01	-0,02	0,00	0,02	0,01	0,00	-0,03	0,03	0,01	0,05	-0,03
Q_L	0,03	-0,04	0,00	0,02	0,02	0,03	-0,02	-0,03	0,00	-0,05	0,03	0,02	0,02
Tfilt	-0,01	0,02	0,02	-0,02	-0,03	-0,04	0,01	0,03	-0,05	0,00	-0,02	0,07	-0,06
Exp/ Vasc	0,04	-0,01	-0,01	-0,04	-0,01	0,06	0,03	0,01	0,03	-0,02	0,00	0,10	-0,04
Laux	0,06	-0,06	0,01	-0,01	0,01	0,04	0,01	0,05	0,02	0,07	0,10	0,00	-0,10
GI	-0,03	-0,01	0,00	0,04	0,01	0,00	-0,01	-0,03	0,02	-0,06	-0,04	-0,10	0,00

Por meio da inspeção visual da matriz residual apresentada na Tabela 5.7 percebeu-se que a matriz de correlação amostral original $R_{13 \times 13}$ foi reproduzida adequadamente pela matriz estimada, uma vez que estes elementos estão próximos de zero.

As medidas a seguir foram utilizadas para sintetizar a informação da matriz residual: o erro médio (EM) Equação 4.1 e a raiz quadrada do erro quadrático médio ($REQM$) Equação 4.2.

$$EM = \frac{1}{n} \sum Y_i \rightarrow EM = \frac{-0,10708}{78} = -0,00137$$

$$REQM = \sqrt{\frac{\sum Y_i^2}{n}} \rightarrow REQM = \sqrt{\frac{0,0867}{78}} = 0,0333$$

Estes valores confirmam a informação apresentada anteriormente de que matriz de correlação amostral original $R_{13 \times 13}$ foi reproduzida adequadamente pela matriz estimada uma vez que estes valores também estão próximos de zero.

5.2 Validação da Análise Fatorial

Da amostra de dados foi extraído aleatoriamente um subconjunto contendo 50% das observações e realizaram-se os cálculos e compararam-se os resultados. A Tabela 5.8 contém os resultados da Análise Fatorial rotacionada ($VARIMAX$) para o subconjunto de amostra.

TABELA 5.8 – Análise Fatorial rotacionada para o subconjunto de observações.

Variáveis (Z_i)	$F_1(\hat{l}_{i1})$	$F_2(\hat{l}_{i2})$	$F_3(\hat{l}_{i3})$	$F_4(\hat{l}_{i4})$	$F_5(\hat{l}_{i5})$	\hat{h}_i^2 (5 fat.)
Gf - Tf	0,2676	-0,1540	-0,0784	0,7604	0,4836	0,9135
Gp	-0,3605	0,3020	0,3408	0,1031	0,7532	0,9153
Nc	-0,0890	0,0678	0,9639	-0,0980	0,1142	0,9642
Vc	0,3671	0,1010	-0,1453	0,0129	0,8632	0,9113
Gcom	0,6748	0,2854	0,4976	-0,3888	-0,0843	0,9427
Gcor	0,8247	0,2279	0,1009	0,3335	0,1576	0,8783
Vs	0,1309	-0,2603	0,8257	0,4073	-0,1332	0,9503
V _L	0,0433	-0,9306	-0,1089	0,1946	-0,0633	0,9216
Q _L	-0,0873	-0,9290	0,1810	-0,0315	-0,1708	0,9336
Tfilt	0,8732	-0,0894	-0,1183	-0,0515	0,0915	0,7955
Exp ou Vasc	-0,7108	0,3413	0,1460	0,3140	-0,0443	0,7436
Laux	-0,4023	0,3050	0,0303	0,5938	0,0357	0,6097
GI	0,2676	-0,1540	-0,0784	0,7604	0,4836	0,9135
Expl.Var	3,0060	2,3412	2,1752	2,0365	1,7018	11,2608
Prp.Totl	0,2312	0,1801	0,1673	0,1567	0,1309	0,8662

Observou-se que o subconjunto e o banco de dados original apresentam cargas fatoriais e *comunalidades* numericamente semelhantes para todas as 13 variáveis. Com esses resultados, há indícios de que os resultados são estáveis para o conjunto de dados das estações em estudo.

5.3 Critério de Seleção das Variáveis

Para a seleção de variáveis que irão permanecer no IQETA, examinou-se a magnitude das cargas fatoriais. Foram escolhidas para permanecer no índice as variáveis que apresentaram cargas fatoriais superior a 0,70, em um dos cinco fatores estudados.

Pela inspeção da Tabela 5.6 matriz fatorial rotacionada pode-se verificar as variáveis que apresentaram maiores valores numéricos dos *loadings*, maior correlação entre o fator e a variável original. Na Tabela 5.9 apresenta-se o fator e a respectiva variável

TABELA 5.9 - Variáveis com maiores *loadings* em cada fator

Fator	Variáveis
1°	Gcom, Gcor, Tfilt, Laux
2°	V _L e Q _L
3°	V _s , N _c
4°	Gf /Tf, Exp ou Vasc
5°	Gp, Vc

A Tabela 5.10 apresenta as variáveis e respectivas etapas de tratamento que estão inseridas. Para a elaboração desta tabela levou-se em consideração que a composição do novo índice deve conter o menor número de variáveis, pois o intuito é de tornar mais simples o procedimento de cálculo do IQETA. Desta forma optou-se por não inserir a variável GI, pois apresenta *loading* (carga fatorial) igual a 0,5917, pois esta corresponde à variável que dentre o grupo selecionado também apresentou menor valor de *comunalidade*, ou seja menor valor de explicação, das variáveis originais, pelos cinco fatores inseridos no índice, o menor valor de carga fatorial admitido nesta dissertação é de 0,70.

O IQETA₁₉ apresenta pesos que em conjunto somam 1. Com a eliminação de quatro parâmetros pela análise técnica e três pelo critério estatístico (Análise Fatorial), para que o somatório das doze variáveis remanescentes no índice some um, fez-se a redistribuição dos pesos de forma proporcional mantendo, portanto, a ordem de importância do parâmetro em relação aos demais, conforme estabelecido pelos painelistas. O resultado desta redistribuição também está apresentado na Tabela 5.10.

TABELA 5.10 - Variáveis incluídas no IQETA₁₂ com redistribuição de pesos

Parâmetro	Peso	Peso
	Original	Redistribuídos
Gmr - Gradiente de velocidade de mistura rápida	0,06	-
Tmr - Tempo de agitação de mistura rápida	0,03	-
Gf/Tf - Gradiente de velocidade da floculação / Tempo de floculação	0,14	0,21
Gp - Gradiente de velocidade nas passagens entre câmaras	0,04	0,07
Nc- Número de câmaras	0,03	0,04
Vc - Velocidade média de escoamento no canal de água floculada	0,03	0,04
Gcom - Gradiente médio de velocidade nas comportas de acesso ao decantador	0,04	0,06
Gcor – Grad.de veloc. através dos orifícios da cortina de distr. de água floculada	0,05	0,07
Vs - Velocidade de sedimentação (taxa de aplicação superficial)	0,08	0,12
V _L - Velocidade longitudinal de escoamento	0,04	0,06
Q _L - Vazão linear de coleta de água decantada	0,06	0,09
Tfilt - Taxa de filtração	0,09	0,13
Dcarr - Duração da carreira de filtração	0,05	-
Exp/Vasc - Expansão de leito filtrante ou velocidade ascensional de lavagem	0,04	0,06
Laux - Lavagem auxiliar	0,02	0,05
Tc - Tempo de detenção no tanque de contato	0,05	-
Nch - Número de chicanas no tanque de contato	0,02	-
Jtest- Ensaio de Jar Test	0,06	-
GI - Grau de instrução da equipe de operação	0,07	-

Aplicou-se a Equação 3.2 a este conjunto de variáveis, considerando esta nova redistribuição de pesos e os valores de pontuação de acordo com os critérios apresentados na íntegra por Lopes (2005), observados em cada uma das dez estações e obtiveram-se os resultados apresentados na Tabela 5.12. A Tabela 5.11 foi apresentada nesta seção para ilustrar todas as variáveis envolvidas no cálculo do IQETA₁₂.

TABELA 5.11 – Cálculo do IQETA₁₂ para a ETA I

ETA I						
Grupo	Parâmetro	Observações (valor médio)	Pontos	Pesos	Pontos [^] Pesos	Pontos por etapa de tratamento
Floculação	Gf (s ⁻¹)	37,11	41	0,21	2,18	4,15
	Tf (s)	806,53				
	Gp (s ⁻¹)	58,15				
	Nc (und)	2,00				
	Vc (m/s)	0,30				
Decantação	Gcom (s ⁻¹)	20,36	70	0,06	1,29	6,00
	Gcor (s ⁻¹)	3,69				
	Vs (cm/min)	3,13				
	V _L (cm/s)	0,27				
	Q _L (L/s.m)	1,70				
Filtração	Tfilt (m ³ /m ²)	289,84	100	0,13	1,82	2,40
	Exp/Vasc (%/cm/min)	32,00				
	Laux	2,00				
IQETA₁₂					59,76	

A seguir apresenta-se a Tabela 5.12 com os comparativos das estações amostradas, dos valores do IQETA₁₉ e IQETA₁₂.

TABELA 5.12 – Comparativo IQETA₁₉ e IQETA₁₂ para as dez estações amostradas

ETA I							
IQETA	Valores parciais						IQETA Total
	M.Rápida	Floculação	Decantação	Filtração	Desinfecção	Operação	
IQETA ₁₉	1,52	2,69	3,24	2,24	1,20	1,68	60
IQETA ₁₂	-	4,15	6,00	2,40	-	-	60

ETA II							
IQETA	Valores parciais						IQETA Total
	M.Rápida	Floculação	Decantação	Filtração	Desinfecção	Operação	
IQETA ₁₉	1,45	2,80	3,10	2,49	1,36	1,79	76
IQETA ₁₂	-	4,35	5,53	2,91	-	-	70

ETA III							
IQETA	Valores parciais						IQETA Total
	M.Rápida	Floculação	Decantação	Filtração	Desinfecção	Operação	
IQETA ₁₉	1,52	2,81	2,12	2,48	1,33	1,79	53
IQETA ₁₂	-	4,41	3,30	2,90	-	-	42

ETA IV							
IQETA	Valores parciais						IQETA Total
	M.Rápida	Floculação	Decantação	Filtração	Desinfecção	Operação	
IQETA ₁₉	1,52	2,67	2,54	2,32	1,36	1,79	58
IQETA ₁₂	-	4,13	4,24	2,55	-	-	45

ETA V							
IQETA	Valores parciais						IQETA Total
	M.Rápida	Floculação	Decantação	Filtração	Desinfecção	Operação	
IQETA ₁₉	1,52	2,77	3,35	2,52	1,32	1,79	85
IQETA ₁₂	-	4,38	6,31	2,97	-	-	82

ETA VI							
IQETA	Valores parciais						IQETA Total
	M.Rápida	Floculação	Decantação	Filtração	Desinfecção	Operação	
IQETA ₁₉	1,52	2,88	3,29	2,24	1,34	1,79	78
IQETA ₁₂	-	4,58	6,12	2,40	-	-	67

ETA VII							
IQETA	Valores parciais						IQETA Total
	M.Rápida	Floculação	Decantação	Filtração	Desinfecção	Operação	
IQETA ₁₉	1,52	3,08	2,82	2,21	1,36	1,79	71
IQETA ₁₂	-	5,09	4,89	2,52	-	-	63

ETA VIII							
IQETA	Valores Parciais						IQETA Total
	M.Rápida	Floculação	Decantação	Filtração	Desinfecção	Operação	
IQETA ₁₉	1,52	3,13	2,75	2,13	1,36	1,79	68
IQETA ₁₂	-	5,20	4,79	2,39	-	-	60

ETA IX							
IQETA	Valores parciais						IQETA Total
	M.Rápida	Floculação	Decantação	Filtração	Desinfecção	Operação	
IQETA ₁₉	1,52	2,66	2,59	2,17	1,35	1,79	65
IQETA ₁₂	-	5,03	5,11	2,15	-	-	55

ETA X							
IQETA	Valores parciais						IQETA Total
	M.Rápida	Floculação	Decantação	Filtração	Desinfecção	Operação	
IQETA ₁₉	1,52	2,96	3,30	1,96	1,32	1,79	69
IQETA ₁₂	-	4,78	6,17	2,02	-	-	60

Pela inspeção das tabelas pode-se perceber que após a redistribuição dos pesos e eliminação de parâmetros o IQETA₁₂ apresentou resultados numericamente inferiores aos apresentados pelo índice original, sendo a porcentagem média de redução de 12%. Observa-se, entretanto que a pontuação das variáveis que permaneceram no índice ficou superior, o que era esperado devido a redistribuição de pesos.

Comparou-se o índice com os valores de turbidez da água filtrada, por meio da realização de gráficos de dispersão e cálculo de correlação entre o IQETA₉ e a porcentagem de valores médios diários da turbidez da água filtrada que se encontrava abaixo de 0,3 uT, 0,5 uT e 0,7 uT. Também foram elaborados gráficos comparando o índice com o valor abaixo do qual estão 95% dos valores da turbidez da água filtrada.

A correlação linear de Spearman foi utilizada para medir o grau de associação entre a porcentagem de turbidez abaixo de um valor pré-estabelecido e o IQETA. Os dados foram organizados na Tabela 5.13.

TABELA 5.13 – Resultados da correlação Spearman entre a turbidez da água filtrada e as formulações do IQETA para as estações amostradas)

IQETA		≤ 0,3 uT	≤ 0,5 uT	≤ 0,7 uT	95%
IQETA ₁₉	Coef. corr	0,85	0,87	0,87	-0,894
	p-valor	0,002	0,001	0,001	0,000
IQETA ₁₂	Coef. corr	0,84	0,75	0,75	-0,800
	p-valor	0,002	0,013	0,013	0,005

Como o trabalho desenvolvido por Lopes (2005) utilizou o coeficiente de correlação de Pearson, estes valores foram reproduzidos na Tabela 5.14 e acrescentaram-se os calculados para o novo índice. Esta correlação empregada por Lopes (2005) é mais apropriada quando os dados apresentam distribuição normal.

TABELA 5.14 – Resultados da correlação de Pearson entre turbidez da água filtrada e as distintas formulações do IQETA

IQETA		≤ 0,3 uT	≤ 0,5 uT	≤ 0,7 uT	95%
IQETA ₁₉	Coef. corr	0,88	0,74	0,625	-0,812
	p-valor	0,001	0,015	0,053	0,001
IQETA ₁₂	Coef. corr	0,67	0,642	0,559	-0,809
	p-valor	0,026	0,045	0,093	0,005

Para ambas as determinações do coeficiente de correlação, verificou-se a tendência do IQETA₉ apresentar correlações mais elevadas, para estações que apresentaram na maior parte do tempo menores valores de turbidez da água filtrada. Apesar de inferiores aos alcançados pela formulação original para os quatro cenários avaliados, utilizando Spearman e de quatro cenários, utilizando Pearson, atestam a correlação significativa da performance da estação de tratamento, em termos da turbidez da água filtrada, e os valores do novo indicador. A principal vantagem reside na maior facilidade de aplicação, em escala real.

6. CONCLUSÕES

A técnica empregada realizou a análise simultânea de todas as variáveis do conjunto, por meio da extração dos fatores que descrevem os elementos da população em estudo e não são observáveis. Com a aplicação deste modelo assumiu-se que as variáveis originais estão relacionadas linearmente com as novas variáveis aleatórias denominadas fatores.

Vale destacar que a técnica permite a extração de várias soluções, de acordo com os limites de variância explicada previamente fixados, com o número de fatores a serem extraídos, com a análise técnica que se fez antes do início dos cálculos que retiram as variáveis consideradas menos relevantes.

O objetivo principal do trabalho de reduzir o número de variáveis que compunham o IQETA foi alcançado por meio da ferramenta estatística de Análise Fatorial. Do conjunto de 19 variáveis referentes às etapas de mistura rápida, floculação, decantação, filtração e desinfecção, permaneceram no índice doze.

Os valores para o coeficiente de Spearman obtidos com a nova formulação do IQETA, apesar de inferiores aos alcançados pela formulação original para os quatro cenários avaliados, atestam a correlação significativa da performance da estação de tratamento, em termos da turbidez da água filtrada, e os valores do novo indicador, de mais fácil aplicabilidade em escala real.

7. REFERÊNCIAS

AMIRTHARAJAH, A; MILLS, K. M. Rapid-mix design for mechanisms of alum coagulation. **Journal of American Water Works Association**, v. 74, n. 4, p. 210-216, apr. 1982.

ANDERSON, T.W. **An introduction to multivariate analysis**. New York: John Wiley, 2003, 713 p.

Associação Brasileira de Normas Técnicas. NBR 12216 – Projeto de estação de tratamento de águas para abastecimento público. Rio de Janeiro, 1992, 18p.

AWWA; ASCE. **Water treatment plant design**. 3ªed. New York: McGraw-Hill, 1998, 806p.

AWWA. **Self-assessment guide for surface water treatment plant optimization**. USA: AWWARF, 1997, 213p.

BASTOS, R.K.X.; PERÍGOLO, R.A.; NASCIMENTO, L.E.; Controle operacional e qualidade da água em estações de tratamento de água em situações adversas - A importância do conhecimento dos parâmetros de funcionamento da ETA. Santo André: ASSEMAE, 2003 (**Anais eletrônicos**).

BOUGUERRA, M.L. **As batalhas da água – Por um bem comum**, Petrópolis: Vozes, 2004, 238p.

BRASIL. Ministério da Saúde. Portaria nº 518 de 25 de março de 2004. Estabelece os procedimentos e responsabilidades relativos ao controle e vigilância da qualidade da água para consumo humano e seu padrão de potabilidade, e dá outras providências. **Diário Oficial da República Federativa do Brasil, Poder Executivo**, Brasília, DF, 2004.

BROWN, R.M., MCCLELLAND, N.I., DEINIGER, R.A., TOZER, R.G. A water quality index – do we are? **Water & Sewage Works, Chicago**, v.117, n.10, p339-343, 1970 *apud* LOPES, V.C. **Índice de qualidade de desempenho de estações convencionais de tratamento de Água**. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) - Escola de Engenharia, UFMG, Belo Horizonte, 2005.

BURLINGAME, G.A.; PICKEL, M.J.; ROMAN, J.T. Practical applications of turbidity monitoring. **Journal of the American Water Works Association**, New York, v.90, n.8, p.57-69, 1998.

CASTRO, L.M.A. **Proposição de indicadores para a avaliação e sistemas de drenagem urbana**. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) - Escola de Engenharia, UFMG, Belo Horizonte, 2002.

CHEVALIER, S.; CHOINIERE, R.; BERNIER, L. **User guide to 40 community health indicators**. Community Health Division, Health and Welfare. Ottawa, Canadá. 1992 *apud* CASTRO, L.M.A. **Proposição de indicadores para a avaliação e sistemas de**

drenagem urbana. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) - Escola de Engenharia, UFMG, Belo Horizonte, 2002.

CLARK, R.M.; COYLE, J.A. Measuring and modeling variations in distribution system water quality. **Journal American Work Water Association**, p.46-53, 1990.

CONSONERY, P.J.; GREENFIELD, D.N.; LEE, J.J. Pennsylvania's filtration evaluation program. **Journal of the American Water Works Association**, New York, v.89, n.8, p. 67-77, 1997.

Di BERNARDO, L.; Di BERNARDO, A. **Métodos e Técnicas de Tratamento de Água.** São Carlos: RIMA, 2005, 792p.

HAIR, J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Multivariate data analysis.** 5ªed. New Jersey: Prentice Hall, 2005,730p.

HAZEN, A. On sedimentation. Transactions, **ASCE**, v.53, n.980, 1904 *apud* Di BERNARDO, L.; Di BERNARDO, A. **Métodos e técnicas de tratamento de água.** São Carlos: RIMA, 2005, 792p.

HELLER, P.G.B.; SPERLING, M.V.; HELLER, L. Desempenho tecnológico dos serviços de abastecimento de água e esgotamento sanitário em quatro municípios de Minas Gerais: uma análise comparativa. **Revista de Engenharia Sanitária e Ambiental**, v.14, n.1, p.109-118. 2009.

JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis.** New Jersey: Prentice Hall, 2002, 622p.

KAISER, H.F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, 23. p.187-200, jan. 1958 *apud* MINGOTI, S.A. **Análise de dados através de métodos de estatística multivariada – Uma abordagem aplicada.** Belo Horizonte: UFMG, 2007, 295p.

KAWAMURA, S. **Integrated design of water treatment facilities.** John Wiley & Sons, Inc. 1991, 658p.

KUBRUSLY, L.S. Utilização de técnicas de análise multivariada para a redução de variáveis num problema de controle ecológico. **Revista Brasileira de Estatística.** V.53, n.199/200, p53-67,1992.

KUBRUSLY, L.S. Um procedimento para calcular índices a partir de uma base de dados multivariados. **Pesquisa Operacional**, v.21, n.1, p.107-117, 2001.

LIBÂNIO, M. **Fundamentos de qualidade e tratamento de água,** Campinas: Átomo, 2005. 444p.

LIBÂNIO, M.; LOPES, V.C. Assessing the feasibility of water treatment plant quality index. **Journal of Water Supply: Research and Technology-Aqua.**v.58, n.5, p.354-362, 2009.

LIBÂNIO, P.A.C.; CHERNICHARO, C.A.L.; NASCIMENTO, N.O. A dimensão da qualidade de água: Avaliação da relação entre indicadores sociais, de disponibilidade hídrica de saneamento e de saúde pública. **Revista Engenharia Sanitária e Ambiental**, v.10, n.3, p.219-228, 2005.

LOONEY, S.W. How to use tests for univariate normality to assess multivariate normality. **The American Statistician**, v.49, p.64-69, 1995 *apud* MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada – Uma abordagem aplicada**. Belo Horizonte: UFMG, 295p. 2007.

LOPES, V.C. **Índice de qualidade de desempenho de estações convencionais de tratamento de Água**. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) - Escola de Engenharia, UFMG, Belo Horizonte, 2005.

Mc QUEEN, D., NOAK, H. Health promotion indicators: current status, issues and problems. **Health Promotion International**. v.3, n.1, p.117-125. 1988

MENDES, C.G. **Estudo da coagulação e floculação das águas sintéticas e naturais com turbidez e cor variáveis**. 1989. Tese (Doutorado) – Escola de Engenharia -USP, São Carlos *apud* Di BERNARDO, L.; Di BERNARDO, A. **Métodos e Técnicas de Tratamento de Água**. São Carlos: RIMA, 2005, 792p.

MINGOTI, S.A. **Análise de dados através de métodos de estatística multivariada – Uma abordagem aplicada**. Belo Horizonte: UFMG, 295p. 2007.

MINGOTI, S.A.; SILVA A.F. Um exemplo de aplicação de técnicas de estatística multivariada na construção de índices de preços. **Revista Nova Economia**. Belo Horizonte. v.7, n.2, p.203-212, 1997.

MINITAB 14.1. **User Guide**. Pennsylvania: Minitab Inc. 2003.

MOITA, J.M.N.; MOITA, G.C. Uma introdução à análise exploratória de dados multivariados. **Revista Química Nova**, v.21,n.4, p.467-469. 1998.

MOLDAN, B., BILHARZ, S. **Sustainability indicators: report of the project on indicators of sustainable development**. Chichester, Reino Unido: John Wiley & Sons. 1997. 415p. *apud* CASTRO, L.M.A. **Proposição de indicadores para a avaliação e sistemas de drenagem urbana**. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) - Escola de Engenharia, UFMG, Belo Horizonte, 2002.

PACKMAN, R.F. Some studies of the coagulation of dispersed clays with hydrolyzing salts. **Journal of Colloid Science**, v.20, n.1, 1965. *apud* LIBÂNIO, M. **Fundamentos de qualidade e tratamento de água**, Campinas: Átomo, 2005. 444p.

RENCHER, A.C. **Methods of multivariate analysis**. John Wiley & Sons, Inc. 2002. 627p.

SANTOS, R.J. **Um curso de geometria analítica e álgebra linear**. Belo Horizonte: Imprensa Universitária da UFMG, 2000. 225p.

SOUZA M.E.T.A.; LIBÂNIO M. Proposta de índice de qualidade para água bruta afluenta a estações convencionais de tratamento. **Revista Engenharia Sanitária Ambiental**, v.14, n.4 dez/out 2009

SPEARMAN, C. General intelligence objectively determined and measured. **American Journal of Psychology**, v.15, p.201-293. 1904 *apud* MINGOTI, S.A. **Análise de dados através de métodos de estatística multivariada – Uma abordagem aplicada**. Belo Horizonte: UFMG, 2007. 295p.

SPRENT, P. **Applied nonparametric statistical methods**. New York: Chapman Hall, 495p.1993

TEIXEIRA, A.R.; SANTOS, E.P.C.; PÁDUA,V.L.; HELLER, L.; DI BERNARDO, L.; LIBÂNIO, M. A confiabilidade analítica dos valores de turbidez da água filtrada e seu efeito no cumprimento do padrão de potabilidade. **Revista engenharia sanitária e ambiental**, v.9, n.1, p.65-72. 2004.

THURSTONE, L.L. **Multiple factor analysis**. Chicago: The University of Chicago Press, 1947 *apud* MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada – Uma abordagem aplicada**. Belo Horizonte: UFMG, 2007. 295p.

United States Environmental Protection Agency (USEPA). National primary drinking water regulations: interim enhanced surface water treatment: final rule. Washington-DC, v.63, n.241, p.69478 – 69521. 1998.

VIANNA, M. R. **Hidráulica aplicada às estações de tratamento de água**. Belo Horizonte: Imprimatur, 3º ed. 2007. 576 p. (CDROM)

World Health Organization (WHO). **Guidelines for drinking water quality**, v.1, 3ªed. Geneva. 2005.

ANEXO A – Parcela dos dados utilizados para o calculo do teste do qui-quadrado

Gf - Tf	Gp	Nc	Vc	Gcom	Gcor	Vs	V _L	Q _L	Tfilt	Dcarr	Exp/Vasc	Laux	GI	Jtest	d ² _j	d ² _{j ord}	((j-1/2)/19)	Ord Q-Quadrado
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,21	1,21	1,00	1,34	1,26	-0,0003	-0,0025	0,0006	4,9768
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0017	5,8057
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0028	6,2548
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0039	6,5772
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0050	6,8335
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0061	7,0486
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0072	7,2353
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0083	7,4010
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0094	7,5506
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0105	7,6874
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0116	7,8137
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0127	7,9313
1,67	1,22	1,13	1,16	1,20	1,26	1,40	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0138	8,0415
1,72	1,22	1,13	1,17	1,20	1,26	1,40	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0002	-0,0016	0,0149	8,1453
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0160	8,2435
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0171	8,3369
1,72	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0002	-0,0016	0,0182	8,4259
1,67	1,22	1,13	1,16	1,20	1,26	1,40	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0193	8,5111
1,72	1,22	1,13	1,17	1,20	1,26	1,40	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0002	-0,0016	0,0204	8,5928
1,72	1,22	1,13	1,16	1,20	1,26	1,40	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0002	-0,0016	0,0215	8,6714
1,72	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0002	-0,0016	0,0226	8,7471
1,72	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0002	-0,0016	0,0237	8,8202
1,72	1,22	1,13	1,17	1,20	1,26	1,40	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0002	-0,0016	0,0248	8,8909
1,72	1,22	1,13	1,17	1,20	1,26	1,40	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0002	-0,0016	0,0259	8,9594
1,72	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0002	-0,0016	0,0270	9,0259
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0281	9,0904
1,72	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0002	-0,0016	0,0292	9,1532
1,67	1,22	1,13	1,17	1,20	1,26	1,41	1,18	1,30	1,52	1,22	1,21	1,00	1,34	1,26	-0,0003	-0,0016	0,0303	9,2144

ANEXO B - Tutorial Minitab

Análise Fatorial - Minitab

Para realizar os cálculos utilizando o Software Minitab devem-se seguir os procedimentos apresentados a seguir:

No menu *Stat* procurar o sub-menu *Multivariate* clique na opção : *Factor Analysis*. A função *Factor Analysis* requer o preenchimento das seguintes opções:

Variables: neste item devem-se escolher as variáveis que serão utilizadas no cálculo. Indicando a coluna a ser utilizada.

Method of Extraction: Define o método pelo qual os fatores vão ser extraídos. Os métodos são: Componentes Principais e Máxima Verossimilhança

Principal components: escolha esta opção para que a extração dos fatores seja realizada pelo método de Componentes Principais.

Maximum likelihood: escolha esta opção para que a extração dos fatores seja realizada pelo método de máxima verossimilhança para a solução inicial

Number of factors to extract: Nesta opção deve ser informado o número de fatores a ser extraídos. Caso esta informação não seja informada o programa irá extrair o número de fatores igual ao número de variáveis em estudo.

Type of rotation: Nesta opção, encontram-se os tipos de rotação disponíveis no MINITAB conforme apresentado a seguir: None, Equimax, Varimax, Quartimax, Ortomax with gamma.

Options: Esta opção permite especificar o tipo de matriz a ser utilizado na análise e as cargas fatoriais para a extração inicial: *Matrix to factor* correlation ou covariance

Source of matrix: *compute from variables* ou *use matrix*:

Compute from variables: esta função deve ser marcada quando a matriz de correlação ou covariância a ser utilizada para a análise, vai ser estimada a partir dos dados na planilha

Use matrix: esta opção deve ser utilizada quando os cálculos forem realizados a partir de uma matriz de correlação ou covariância a partir da qual possam ser estimadas as cargas fatoriais e os coeficientes.

Loadings for initial solution: Cargas fatoriais para a solução inicial - *compute from variables* ou use *loadings*:

Compute from variables: escolha esta opção para calcular as cargas fatoriais a partir dos dados contidos na planilha

Use *loadings*: escolha esta opção caso deseje utilizar as cargas fatoriais nas colunas do MINITAB

Graphs Scree plot: ver análise de componentes principais

Graphs Score plot for first 2 factors: por meio desta opção é possível obter o gráfico dos Scores do primeiro e do segundo fator

Loading plot for first 2 factors: esta opção fornece o gráfico das cargas fatoriais do primeiro e do Segundo fator

Storage (armazenamento)

Loadings: nesta opção deve ser indicado as colunas *nas quais* os valores das cargas fatoriais serão armazenados. Uma coluna para cada fator.

Coefficients: nesta opção deve ser indicada a coluna na qual serão armazenados os coeficientes de cada fator estimado. Deve ser indicado uma coluna para cada fator

Scores: nesta opção deve ser indicado as colunas *nas quais* vão ser armazenados os scores dos fatores. Deve ser indicado uma coluna para cada fator

Rotation matrix: nesta opção deve ser indicado um nome para a matriz rotacionada.

Residual matrix: nesta opção deve ser indicado um nome para ser dada a matriz de resíduos

Eigenvalues: nesta opção deve ser indicado as colunas *nas quais* vão ser armazenados os autovalores da matriz utilizada na análise.

Eigenvector matrix: nesta opção devem ser indicadas as colunas *nas quais* deverão ser armazenados os autovetores.