

UNIVERSIDADE FEDERAL DO PARANÁ

WYRLLEN EVERSON DE SOUZA

**ANÁLISE DAS DISTORÇÕES HARMÔNICAS DE TENSÃO A PARTIR DE
CARACTERÍSTICAS DOS TRANSFORMADORES E DE DADOS DE CONSUMO**

CURITIBA

2008

WYRLLEN EVERSON DE SOUZA

**ANÁLISE DAS DISTORÇÕES HARMÔNICAS DE TENSÃO A PARTIR DE
CARACTERÍSTICAS DOS TRANSFORMADORES E DE DADOS DE CONSUMO**

Dissertação apresentada ao curso de Pós-Graduação em Métodos Numéricos em Engenharia, Área de Concentração em Programação Matemática, Departamento de Engenharia Civil, Departamento de Matemática dos setores de Tecnologia e de Ciências Exatas da Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Anselmo Chaves Neto
Co-Orientador: Prof. Dr. Alexandre Rasi Aoki

CURITIBA

2008

TERMO DE APROVAÇÃO

WYRLLEN EVERSON DE SOUZA

ANÁLISE DAS DISTORÇÕES HARMÔNICAS DE TENSÃO A PARTIR DE CARACTERÍSTICAS DOS TRANSFORMADORES E DE DADOS DE CONSUMO

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no curso de Pós-Graduação em Métodos Numéricos em Engenharia, setores de Tecnologia e Ciências Exatas da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientador:

Prof. Dr. Anselmo Chaves Neto
Departamento de Estatística, UFPR

Co-Orientador:

Prof. Dr. Alexandre Raso Aoki
PPGEE – UFPR/ Lactec

Prof. Dr. Osmar Ambrósio de Souza
Departamento de Matemática – UNICENTRO

Prof. Dr. Paulo Victor Prestes Marcondes
Departamento de Engenharia Mecânica, UFPR

Curitiba, 25 de fevereiro de 2008

DEDICATÓRIA

A Deus.

A minha Família.

Aos meus Amigos.

Dedico esse trabalho ao meu pai Orlando, a minha mãe Cleide e meu irmão Woryk por estarem ao meu lado e por terem me apoiado nessa etapa da minha vida.

AGRADECIMENTO

Agradeço a Deus pelo seu imenso amor e misericórdia revelado nas oportunidades que a vida me trouxe. Quero também agradecer:

A minha família pelo incentivo, confiança e carinho.

Ao meu orientador o Prof. Dr. Anselmo Chaves Neto, pela orientação, incentivo e amizade desde o início do trabalho.

Ao meu co-orientador Dr. Alexandre Rasi Aoki, pela orientação e amizade que muito influenciou nesse trabalho.

A Maristela Bradil e todos os professores do programa de Pós-Graduação em Métodos Numéricos em Engenharia, que ajudaram diretamente na realização desse trabalho.

Aos pesquisadores do LACTEC pela paciência e informação que foram fundamentais nesse trabalho em especial.

Aos amigos que sempre estiveram ao meu lado e que foram fundamentais para a realização desse trabalho com seu apoio e incentivo, em especial ao Roberto, Bernadete, Vanessa, Fabio entre outros.

Ao Eng. Mauricio Robles Ortega e a COPEL Distribuição pela cessão de dados e ao LACTEC pelo suporte financeiro.

Por fim a todos que, de alguma forma, contribuíram com a realização dessa dissertação.

EPÍGRAFE

“Mestres são aqueles que durante todo esse tempo se entregaram a ciência, mas Sábios somos nós que, além disso, vivemos as loucuras do amor...”

Autor “desconhecido”

RESUMO

O presente trabalho tem por finalidade apresentar um procedimento de desenvolvimento de um modelo estatístico computacional para a estimação das distorções harmônicas de tensão a partir das características dos transformadores de distribuição e dos consumidores utilizando-se de técnicas de Análise Multivariada. O objetivo operacional do modelo é identificar e classificar transformadores da rede elétrica do Paraná com base na sua distorção harmônica total de tensão. A síntese dos modelos fundamenta-se em técnicas estatísticas de reconhecimento e classificação de padrões (escore quadrático e regressão logística). Assim, fez-se necessário, para o desenvolvimento dos modelos, a realização de uma extensa campanha de medição de harmônicos no sistema de distribuição da Companhia Paranaense de Energia (COPEL). Os registros das medições obtidos na campanha de medição, somados às características dos transformadores de distribuição monitorados e dos consumidores alimentados por estes transformadores, são utilizados como base para suprir os modelos estatísticos. Os modelos desenvolvidos para o reconhecimento e classificação de padrões foram implementados computacionalmente em cada uma das cinco superintendências de distribuição regional do estado do Paraná. Para a avaliação dos modelos de reconhecimento e classificação de padrão utilizou-se da taxa de erro aparente (*Apparent Error Rate - APER*).

ABSTRACT

This work presents the development of a statistical computational model to study the harmonic distortions of tension in electric transformers. The model was constructed using the Multivariate Analysis method and considering the features of the transformers, and the system of distribution and the consumption of electricity. The operational objective of this model was identify a group of transformers from the electricity net using data of harmonic distortion and after classify them in term of distortion values. The model was based on statistical approaches to recognize and classify standards, such as quadratic scores and logistic regression. In this way, to develop the method it was necessary a large measurement program of harmonic distortion in the electricity distribution system of the Parana State Electricity Company (COPEL). The measurements that have been done during the work and the characteristics of the distribution transformers monitored were used like a base to supply data for the statistical model. The model was implemented in the computational system of each centre of the electricity company in a total of five centers to monitor the electricity distribution in the Parana State. The model of recognition and classification of the standards was evaluated by the Apparent Error Rate method (APER).

LISTA DE ILUSTRAÇÕES

FIGURA 2.1 – FORMA DE ONDA DA CORRENTE DE LINHA DE UM CONVERSOR TRIFÁSICO DE 6 PULSOS	20
FIGURA 2.2 – COMPOSIÇÃO DA COMPONENTE HARMÔNICA DE ORDEM 5 COM A FUNDAMENTAL.....	21
FIGURA 2.3 – TENSÃO HARMÔNICA VERSUS CORRENTE HARMÔNICA	22
FIGURA 2.4 – DECOMPOSIÇÃO EM COMPONENTES SIMÉTRICAS.....	23
FIGURA 3.1 – REGIÕES DO PLANEJAMENTO ENERGÉTICO.....	65
FIGURA 3.2 – PERÍODO DE MONITORAÇÃO VERSUS PERÍODO DE ANÁLISE....	70
FIGURA 3.3 – PERFIL DAS DISTORÇÕES HARMÔNICAS TOTAIS DAS TENSÕES VALORES DENTRO DO LIMITE.....	71
FIGURA 3.4 – PERFIL DAS DISTORÇÕES HARMÔNICAS TOTAIS DAS TENSÕES VALORES FORA DO LIMITE.....	71
FIGURA 3.5 – ESPECTRO HARMÔNICO DA TENSÃO DE SUPRIMENTO NO PERÍODO DE MAIOR DISTORÇÃO COM VALORES DENTRO DO LIMITE.....	73
FIGURA 3.6 – ESPECTRO HARMÔNICO DA TENSÃO DE SUPRIMENTO NO PERÍODO DE MAIOR DISTORÇÃO COM VALORES FORA DO LIMITE.....	73
FIGURA 3.7 – TELA INICIAL DO <i>SOFTWARE</i> DISTORÇÕES	76
FIGURA 3.8 – TELA DE CLASSIFICAÇÃO	77
FIGURA 3.9 – TELA DE CLASSIFICAÇÃO REGRESSÃO LOGÍSTICA	78
FIGURA 3.10 – TELA DE CLASSIFICAÇÃO ESCORE QUADRÁTICO	78
FIGURA 3.12 – TELA COM DADOS INSERIDOS.....	79
FIGURA 3.13 – TELA DE RESULTADO DO <i>SOFTWARE</i> DISTORÇÕES	80
QUADRO 4.1 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL LESTE (SDL) UTILIZANDO REGRESSÃO LOGÍSTICA	82
QUADRO 4.2 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NORTE (SDT) UTILIZANDO REGRESSÃO LOGÍSTICA	82

QUADRO 4.3 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL CENTRO SUL (SDC) UTILIZANDO REGRESSÃO LOGÍSTICA.....	83
QUADRO 4.4 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NOROESTE (SDN) UTILIZANDO REGRESSÃO LOGÍSTICA.....	83
QUADRO 4.5 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL OESTE (SDO) UTILIZANDO REGRESSÃO LOGÍSTICA.....	84
QUADRO 4.6 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL LESTE (SDL) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES IGUAIS	85
QUADRO 4.7 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NORTE (SDT) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES IGUAIS.....	85
QUADRO 4.8 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL CENTRO SUL (SDC) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES IGUAIS	86
QUADRO 4.9 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NOROESTE (SDN) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES IGUAIS	87
QUADRO 4.10 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL OESTE (SDO) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES IGUAIS	87
QUADRO 4.11 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL LESTE (SDL) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES PROPORCIONAIS.....	88
QUADRO 4.12 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NORTE (SDT) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES PROPORCIONAIS.....	89
QUADRO 4.13 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL CENTRO SUL (SDC) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES PROPORCIONAIS	90

QUADRO 4.14 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NOROESTE (SDN) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES PROPORCIONAIS.....	90
QUADRO 4.15 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL OESTE (SDO) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES PROPORCIONAIS.....	91

LISTA DE TABELAS

TABELA 2.1 – SEQÜÊNCIA DE FASES DAS COMPONENTES HARMÔNICAS	24
TABELA 2.2 – LIMITES GLOBAIS DE DISTORÇÃO HAMÔNICA DE TENSÃO EXPRESSOS EM PORCENTAGEM DA TENSÃO FUNDAMENTAL.	29
TABELA 2.3 – TABELA DE CLASSIFICAÇÃO.....	59
TABELA 3.1 – ÁREAS DE ATUAÇÃO DA COPEL.....	63
TABELA 3.2 – DADOS DE DISTRIBUIÇÃO	64
TABELA 3.3 – NÚMERO DE CONSUMIDORES	64
TABELA 3.4 – CONSUMO DE ENERGIA POR CLASSE	64
TABELA 3.5 – DISTRIBUIÇÃO DOS PONTOS DE MONITORAÇÃO, CONFORME A POTÊNCIA INSTALADA POR REGIONAL.....	67
TABELA 3.6 – DISTRIBUIÇÃO DAS AMOSTRAS ENTRE ÁREAS URBANAS E RURAIS.....	67
TABELA 3.7 – VARIÁVEIS ASSOCIADAS À CARACTERÍSTICA DO TRANSFORMADOR	69
TABELA 3.8 – VARIÁVEIS ASSOCIADAS À CARACTERÍSTICA DO CONSUMO. ...	69
TABELA 3.9 – VALORES MÁXIMOS, MÍNIMOS E P _{95%} DAS DISTORÇÕES TOTAIS DAS TENSÕES COM VALORES DENTRO DO LIMITE.....	72
TABELA 3.10 – VALORES MÁXIMOS, MÍNIMOS E P _{95%} DAS DISTORÇÕES TOTAIS DAS TENSÕES COM VALORES FORA DO LIMITE.	72
TABELA 3.11 – DISTRIBUIÇÃO DOS TRANSFORMADORES MEDIDOS	74
TABELA 4.1 – TABELA DAS VARIÁVEIS.....	81
TABELA 4.2 – TABELA COM APER DE TODAS AS REGIONAIS.....	92

SUMÁRIO

1 INTRODUÇÃO.....	14
1.1 TEMA DE ESTUDO	14
1.2 OBJETIVOS DO TRABALHO	15
1.2.1 Objetivo Geral.....	15
1.2.2 Objetivos Específicos.....	15
1.3 IMPORTÂNCIA DO TRABALHO	16
1.4 ESTRUTURA DA DISSERTAÇÃO.....	16
2 REVISÃO DE LITERATURA	17
2.1 INTRODUÇÃO.....	17
2.2 QUALIDADE DA ENERGIA ELÉTRICA.....	17
2.3 HARMÔNICOS	19
2.3.1 Conceitos Básicos	19
2.3.2 Distorções Harmônicas.....	21
2.3.3 Componentes Simétricas	22
2.3.4 Harmônica Total e Individual	24
2.3.5 Limites para distorção harmônica	25
2.4 ANÁLISE MULTIVARIADA.....	32
2.4.1 Introdução	32
2.4.2 Estatística Descritiva Multivariada	34
2.5 RECONHECIMENTO DE PADRÕES.....	36
2.5.1 Discriminação, Classificação e Reconhecimento de Padrão	36
2.5.2 Análise Discriminante.....	37
2.5.3 Problema Geral da Classificação com Várias Populações.....	38
2.5.4 Análise Discriminante para Duas Populações – Método de Fischer.....	40
2.5.5 Discriminação Entre Diversas Populações.....	45
2.5.6 Escore Discriminante Quadrático de Classificação	49
2.5.7 Teste de Normalidade (Gaussianidade)	50
2.6 REGRESSÃO LOGÍSTICA.....	52
2.6.1 Introdução	52
2.6.2 Modelo de Regressão Logística.....	52

2.6.3 Modelo de Regressão Logística Simples.....	53
2.6.4 Modelo de Regressão Logístico Múltiplo	57
2.7 AVALIAÇÃO DA FUNÇÃO DE CLASSIFICAÇÃO	58
2.7.1 Critério TPM (<i>Total Probability of Misclassification</i>).....	58
2.7.2 Abordagem de Lachenbruch.....	61
3 MATERIAL E MÉTODO.....	63
3.1 MATERIAL.....	63
3.1.1 Área de Estudo.....	63
3.1.2 Levantamento de Dados.....	66
3.1.3 Variáveis Utilizadas	68
3.2 MÉTODO.....	70
3.2.1 Sistema Matlab.....	75
3.2.1.1 Algoritmo <i>logistic</i>	75
3.2.1.2 Algoritmo <i>dis</i>	75
3.2.2 Software Distorcoes	76
4 RESULTADOS E DISCUSSÕES	81
5 CONCLUSÕES E DESENVOLVIMENTOS FUTUROS	94
5.1 CONCLUSÕES	94
5.2 DESENVOLVIMENTOS FUTUROS	97
REFERÊNCIAS	99

1 INTRODUÇÃO

1.1 TEMA DE ESTUDO

Nos últimos anos, a Qualidade da Energia Elétrica (QEE) tem despertado o interesse dos órgãos reguladores do setor elétrico. Isso se deve pela forte e contínua degradação dos índices de conformidade e continuidade preconizados pelas normas vigentes. O desenvolvimento de equipamentos cada vez mais sensíveis às variações no fornecimento da energia elétrica, a crescente conscientização da população sobre seus direitos enquanto consumidores, aliados à escassez de documentos normativos que estabeleçam penalidades sobre violações de limites, são alguns dos fatores que contribuem para a mencionada preocupação. Igualmente, os reparos ou substituições de equipamentos elétricos, em face dos elevados níveis dos distúrbios, têm provocado custos financeiros indesejáveis. Os valores desses custos poderiam ser revertidos em investimentos para a melhoria do sistema elétrico da concessionária.

Dos fenômenos associados à Qualidade da Energia Elétrica, as distorções harmônicas de tensão merecem especial atenção, em virtude da crescente presença de cargas com características não-lineares nas redes elétricas de distribuição. Esta forte disseminação de cargas poluidoras, ou seja, cargas não-lineares expõem consumidores e concessionária a níveis elevados de distorção harmônica, podendo provocar efeitos indesejáveis sobre o sistema, como o aumento das perdas elétricas, funcionamento inadequado de equipamentos, sobretensões harmônicas, dentre outros. O fenômeno ganha maior relevância se considerada a dificuldade de atribuição de responsabilidade entre supridores e consumidores, sobre a injeção de harmônicos na rede elétrica.

Desta forma, a construção de uma ferramenta computacional, com métodos estatísticos que possa estimar os níveis de distorções harmônicas total de tensão

(DHTv) torna-se extremamente útil e economicamente interessante, já que uma campanha de medição apresenta um custo elevado.

1.2 OBJETIVOS DO TRABALHO

1.2.1 Objetivo Geral

Desenvolver uma regra de reconhecimento de padrões que seja eficiente na identificação e classificação de transformadores de acordo com o nível de distorção harmônica total de tensão. Além disso, a regra de reconhecimento e classificação de padrões deve ser programada de modo que faça com pouca ou nenhuma intervenção do usuário a classificação de novos transformadores não monitorados.

1.2.2 Objetivos Específicos

A fim de alcançar o objetivo geral estabeleceram-se os seguintes objetivos específicos:

1. Identificar as variáveis que influenciam no reconhecimento dos níveis de distorções harmônicas total de tensão nos transformadores;
2. Testar os vários métodos estatísticos de reconhecimento de padrões, classificar e fazer a comparação entre os métodos;
3. Implementar, computacionalmente, métodos estatísticos mais eficiente no reconhecimento e classificação de padrão.

1.3 IMPORTÂNCIA DO TRABALHO

A importância desse trabalho está, principalmente, no desenvolvimento de uma ferramenta, até então inexistente, que permita a partir dos dados das características dos transformadores e de suas características de consumo, identificar *a priori* se o mesmo possui distorção harmônica total de tensão fora do limite estabelecido. O conhecimento a priori da situação da DHTV dos transformadores pode contribuir para planejar e executar futuras campanhas de medição de QEE no sistema de distribuição.

1.4 ESTRUTURA DA DISSERTAÇÃO

A dissertação está organizada em cinco capítulos. No primeiro capítulo encontra-se esta introdução, com o tema do trabalho, os objetivos e a justificativa. No segundo capítulo tem-se uma revisão de literatura envolvendo a teoria sobre qualidade da energia elétrica, distorções harmônicas de tensão e os conceitos estatísticos da Análise Multivariada, que sustenta o desenvolvimento do programa computacional. O material utilizado na pesquisa bem como a metodologia empregada está apresentado no terceiro capítulo. Os resultados obtidos através do *software* desenvolvido são apresentados no quarto capítulo. Finalmente, no quinto capítulo estão as conclusões e também as sugestões para trabalhos futuros.

2 REVISÃO DE LITERATURA

2.1 INTRODUÇÃO

O objetivo desse capítulo é explicar, o conceito de Qualidade da Energia Elétrica (QEE), distorções harmônicas, suas especificações e normas, além de fornecer um embasamento teórico sobre análise multivariada.

2.2 QUALIDADE DA ENERGIA ELÉTRICA

A energia elétrica é um elemento indispensável para a sociedade moderna, além de oferecer meios para que a população tenha melhor qualidade de vida, representa insumo básico para a maioria das atividades comerciais e industriais. Assim sendo, a energia elétrica pode ser tratada tanto sob o enfoque de produto como de serviço.

A qualidade da energia elétrica – QEE tem se tornado uma preocupação crescente e comum as empresas de energia elétrica e aos consumidores de um modo geral. O conceito da Qualidade de Energia está relacionado a um conjunto de alterações que podem ocorrer no sistema elétrico. Uma boa definição para o problema de qualidade da energia é: "Qualquer problema da energia manifestado na tensão, corrente ou nas variações de frequência que resulte em falha ou má operação de equipamentos de consumidores". Tais alterações podem ocorrer em várias partes do sistema de energia, seja nas instalações de consumidores ou no sistema supridor da concessionária.

Conceituar qualidade da energia elétrica é uma tarefa abrangente, pois o termo engloba aspectos bem diversos, como a tecnologia usada, a satisfação dos

clientes, custo do produto, etc. A qualidade da energia elétrica pode ser definida como a ausência relativa de distúrbios na tensão provocados pelo sistema da concessionária. No entanto, uma melhor definição está relacionada ao funcionamento adequado dos equipamentos, das cargas e dos próprios sistemas de potência, não significando necessariamente uma energia isenta de qualquer distúrbio.

Segundo Peniche e Teixeira (2005), podemos conceituar a qualidade de energia elétrica como:

- Qualidade do atendimento;
- Qualidade do serviço;
- Qualidade do produto.

O termo Qualidade da Energia Elétrica está relacionado com qualquer desvio que possa ocorrer na magnitude, forma de onda ou frequência da tensão e/ou corrente elétrica. Esta designação também se aplica às interrupções de natureza permanente ou transitória que afetam o desempenho da transmissão, distribuição e utilização da energia elétrica.

2.2.1 Principais Distúrbios da QEE

O termo Qualidade da Energia Elétrica (QEE) se refere a uma ampla variedade de fenômenos eletromagnéticos conduzidos que caracterizam a tensão e a corrente num dado tempo e local do sistema elétrico.

A probabilidade de uma instalação sofrer ou não problemas de QEE depende da qualidade da alimentação de energia da concessionária, dos tipos de cargas existentes na instalação e da sensibilidade dos equipamentos aos vários tipos de perturbação.

Os principais fenômenos eletromagnéticos associados à QEE podem ser agrupados em:

- Variações Transitórias da Tensão;
- Variações Momentâneas de Tensão;
- Variações Sustentadas de Tensão;

- Variações Momentâneas de Freqüência;
- Distorção Harmônica Total de Tensão, Flutuação de Tensão, Cintilação e Desequilíbrio de Tensão;
- Ruídos e Interferências Eletromagnéticas.

2.3 HARMÔNICOS

Atualmente o sistema elétrico sofre um aumento da sua capacidade de geração, transmissão e distribuição. Aliado a isso ocorre uma ampliação do número de cargas não-lineares, em todos os níveis de fornecimento. A utilização dessas cargas constitui-se numa das maiores preocupações tanto das concessionárias quanto dos consumidores, e também é motivo de constantes problemas para o sistema elétrico como um todo e seus usuários.

Tais cargas conhecidas como não-lineares ou “cargas elétricas especiais” provocam distorções da tensão nas redes elétricas, podendo comprometer em alguns casos, o perfeito funcionamento do sistema e seus equipamentos.

2.3.1 Conceitos Básicos

A palavra “harmônicos” foi originalmente definida em acústica, significando a vibração de um fio ou coluna de ar, com a presença de freqüências múltiplas em relação a uma freqüência fundamental, provocando uma distorção do som resultante.

Fenômenos semelhantes a este ocorrem na Engenharia Elétrica, onde deformações das tensões e correntes elétricas também têm sido registradas. Neste caso, os fundamentos físicos e matemáticos utilizados na área da Física podem ser imediatamente aplicados às questões elétricas.

Dentre os objetivos a serem alcançados por uma concessionária de energia, destaca-se a qualidade da energia elétrica fornecida aos seus consumidores. Em

condições ideais esta energia deve ser fornecida com uma tensão puramente senoidal, com frequência e amplitude constante. Entretanto, constatam-se, na prática, desvios significativos das condições consideradas ideais.

A distorção de tensão e corrente pode ser analisada matematicamente através dos estudos das ondas não senoidais periódicas. Nestas condições, sabe-se que qualquer onda que possua em seu conteúdo distorções com amplitude diferente da fundamental pode ser decomposta de acordo com a Série de Fourier, em uma componente de mesma frequência que a da onda resultante, chamada de “onda fundamental”, e em outras ondas senoidais de frequências múltiplas da fundamental, denominadas “harmônicas”.

A ferramenta matemática utilizada no cálculo da amplitude e do ângulo de fase das harmônicas é denominada de FFT – *Fast Fourier Transform* ou Transformada Rápida de Fourier.

Como exemplo na Figura 2.1 mostra a clássica forma de onda da corrente de linha de suprimento de uma ponte retificadora trifásica, não controlada e de 6 pulsos, alimentado uma carga corrente contínua (cc) altamente indutiva.

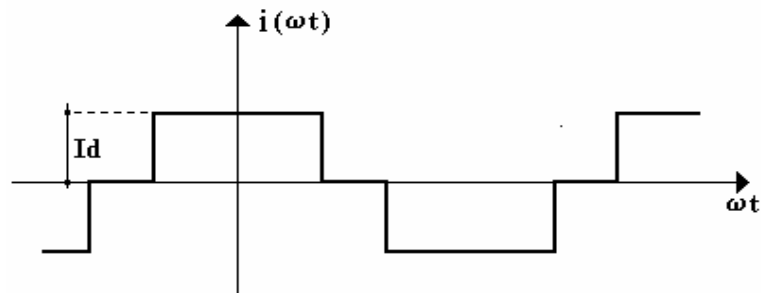


FIGURA 2.1 – FORMA DE ONDA DA CORRENTE DE LINHA DE UM CONVERSOR TRIFÁSICO DE 6 PULSOS

FONTE: CONCEITOS GERAIS SOBRE QUALIDADE DE ENERGIA

A corrente $i(\omega t)$ da ilustração é uma onda não senoidal, periódica, e pode ser decomposta na série de Fourier, cujo desenvolvimento leva à seguinte expressão:

$$i(\omega t) = \frac{2\sqrt{3}}{\pi} I_d \left[\cos(\omega t) - \frac{1}{5} \cos(5\omega t) + \frac{1}{7} \cos(7\omega t) - \frac{1}{11} \cos(11\omega t) + \frac{1}{13} \cos(13\omega t) - \dots \right] \quad (2.1)$$

onde:

ω - frequência angular do sistema em rad/s;

I_d - valor máximo da corrente.

A Figura 2.2 ilustra parte da composição obtida pela adição da componente harmônica de ordem 5 com a fundamental.

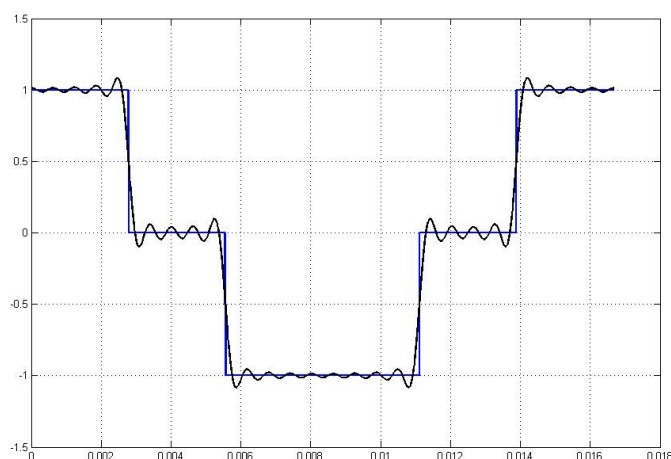


FIGURA 2.2 – COMPOSIÇÃO DA COMPONENTE HARMÔNICA DE ORDEM 5 COM A FUNDAMENTAL

FONTE: CONCEITOS GERAIS SOBRE QUALIDADE DE ENERGIA

A grande vantagem da utilização desta técnica matemática é que, em sistemas lineares, cada componente harmônica pode ser considerada separadamente de maneira que uma análise global pode ser obtida a partir da superposição das análises das componentes do sinal distorcido.

2.3.2 Distorções Harmônicas

Em geral a distorção harmônica de tensão é o resultado da corrente harmônica circulando através da impedância do sistema elétrico. Considerando que a impedância série seja linear, para cada frequência harmônica da corrente, há uma queda de tensão de mesma frequência, resultando, desta forma, na distorção da tensão na barra. A Figura 2.3 ilustra o fenômeno acima descrito.

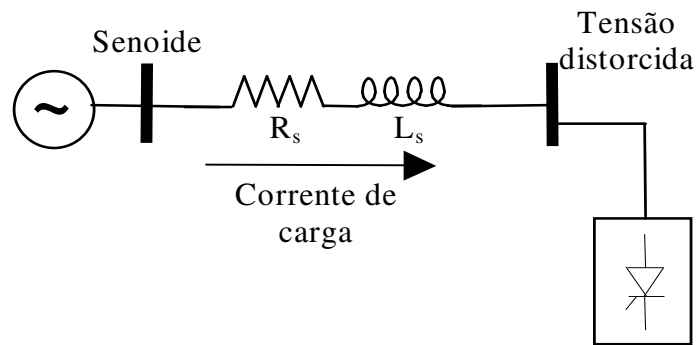


FIGURA 2.3 – TENSÃO HARMÔNICA VERSUS CORRENTE HARMÔNICA
 FONTE: CONCEITOS GERAIS SOBRE QUALIDADE DE ENERGIA

A magnitude da distorção de tensão depende, basicamente, da impedância equivalente vista pela carga não linear ou fonte de corrente harmônica e da corrente consumida pela mesma. Deve-se reconhecer que a carga não exerce controle sobre os níveis de distorção de tensão. Conseqüentemente, uma mesma carga poderá resultar em distorções de tensão diferentes, dependendo da sua localização no sistema elétrico.

2.3.3 Componentes Simétricas

Em sistemas trifásicos significativamente desequilibrados, cada harmônico deve ser decomposto nas três componentes simétricas, ou seja, seqüência positiva, seqüência negativa e seqüência zero, em uma análise denominada Harmônicos Desequilibrados.

Reconhecendo que as redes elétricas e as cargas não lineares são, em sua maioria, trifásicas, o tratamento elétrico para os sinais de tensão e corrente deve ser suficientemente abrangente para atender as condições equilibradas e desequilibradas de operação. Neste sentido, é amplamente conhecido que qualquer sistema de três fasores desbalanceados pode ser representado através de três conjuntos de fasores balanceados, conforme a seguir (Figura 2.4):

- Componentes de seqüência positiva: conjunto de 3 fasores defasados de 120° , com rotação de fase ABC;

- Componentes de seqüência negativa: conjunto de 3 fasores defasados de 120° , porém com rotação de fase ACB;
- Componentes de seqüência zero: conjunto de 3 fasores em fase.

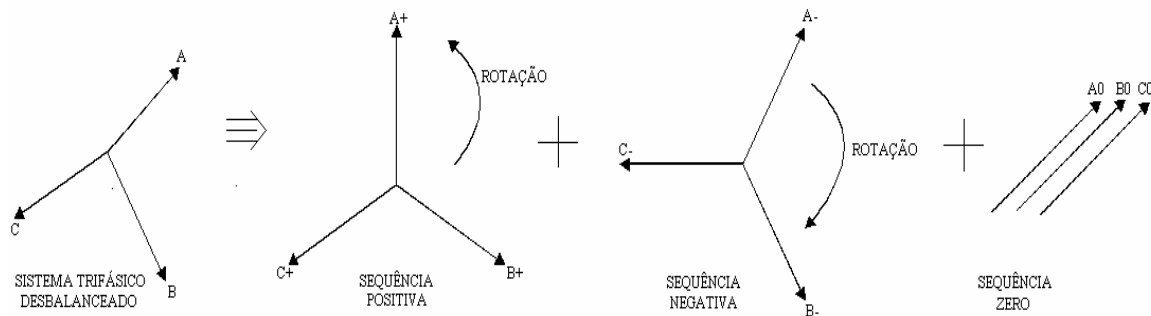


FIGURA 2.4 – DECOMPOSIÇÃO EM COMPONENTES SIMÉTRICAS.
 FONTE: CONCEITOS GERAIS SOBRE QUALIDADE DE ENERGIA

Considerando um sistema elétrico trifásico balanceado e distorcido, as correntes em cada fase, e suas respectivas componentes harmônicas de ordens ímpares, podem ser genericamente expressas por:

$$i_a(\omega t) = I_1 \text{sen}(\omega t) + I_3 \text{sen}(3\omega t) + I_5 \text{sen}(5\omega t) + I_7 \text{sen}(7\omega t) + \dots \quad (2.2)$$

$$i_b(\omega t) = I_1 \text{sen}(\omega t - 120^\circ) + I_3 \text{sen}(3\omega t - 360^\circ) + I_5 \text{sen}(5\omega t + 120^\circ) + I_7 \text{sen}(7\omega t - 120^\circ) + I_9 \text{sen}(9\omega t - 360^\circ) \dots \quad (2.3)$$

$$i_c(\omega t) = I_1 \text{sen}(\omega t + 120^\circ) + I_3 \text{sen}(3\omega t + 360^\circ) + I_5 \text{sen}(5\omega t - 120^\circ) + I_7 \text{sen}(7\omega t + 120^\circ) + I_9 \text{sen}(9\omega t + 360^\circ) \dots \quad (2.4)$$

onde:

$i_a(\omega t)$, $i_b(\omega t)$, $i_c(\omega t)$ - corrente instantânea em cada fase;

$I_1, I_3, I_5, \dots, I_n$ - valor máximo da corrente harmônica de ordem 1, 3, 5, ...,

n.

Em se tratando de sistemas elétricos trifásicos equilibrados, decompondo-se uma forma de onda de tensão ou corrente distorcida, as ordens harmônicas advindas desta decomposição, em termos de componentes simétricos, são as seguintes (Tabela 2.1):

TABELA 2.1 – SEQÜÊNCIA DE FASES DAS COMPONENTES HARMÔNICAS

<i>Ordem harmônica</i>	1	2	3	4	5	6	7	etc.
<i>Seqüência de Fases</i>	+	-	0	+	-	0	+

FONTE: CONCEITOS GERAIS SOBRE QUALIDADE DE ENERGIA

A importância de se conhecer a seqüência de fases de cada ordem harmônica deve-se ao fato que os princípios de funcionamento dos equipamentos de compensação harmônica, bem como o comportamento das harmônicas de seqüência zero, fundamentam-se nas propriedades acima apresentadas.

2.3.4 Harmônica Total e Individual

Dentre as diferentes grandezas utilizadas para se calcular o conteúdo harmônico de uma forma de onda, o mais utilizado é a “Distorção Harmônica Total” (DHT), a qual pode ser empregada tanto para sinais de tensão como para correntes. As equações (2. 5) e (2. 6) expressam tais definições.

$$DHT_v = \sqrt{\frac{\sum_{n>1}^{n_{max}} V_n^2}{V_1^2}} \times 100(\%) \quad (2.5)$$

$$DHT_i = \sqrt{\frac{\sum_{n>1}^{n_{max}} I_n^2}{I_1^2}} \times 100(\%) \quad (2.6)$$

onde:

- DHTv - distorção harmônica total de tensão;
- DHTi - distorção harmônica total de corrente;
- V_n - valor eficaz da tensão de ordem n;
- I_n - valor eficaz da corrente de ordem n;
- V_1 - valor eficaz da tensão fundamental;
- I_1 - valor eficaz da corrente fundamental;
- n - ordem da componente harmônica.

Para a quantificação da distorção individual de tensão ou corrente, ou seja, a porcentagem de determinada componente harmônica em relação a sua componente fundamental, a “Distorção Harmônica Individual” é utilizada. As equações (2. 7) e (2. 8) expressam tais definições.

$$DHIv = \frac{V_n}{V_1} \times 100(\%) \quad (2. 7)$$

$$DHIi = \frac{I_n}{I_1} \times 100(\%) \quad (2. 8)$$

onde:

- DHIv - distorção harmônica individual de tensão;
- DHIi - distorção harmônica individual de corrente.

2.3.5 Limites para distorção harmônica

Dentre as principais normas e recomendações, nacionais e internacionais, que tratam da medição de harmônicos em sistemas de distribuição, podemos relacionar basicamente seis importantes documentos os quais serão descritos na seqüência

a) IEEE Std. 519 – 1992; - *IEEE RECOMMENDED PRACTICES AND REQUIREMENTS FOR HARMONIC CONTROL IN ELECTRICAL POWER SYSTEMS*

Este guia do IEEE se aplica especialmente para os casos de conversores estáticos de potência empregados tanto em sistemas industriais quanto comerciais. Neste sentido, aspectos relacionados com o controle da emissão harmônica geradas por este tipo de equipamento, bem como a correção de potência reativa é abordada. A norma também se preocupa em estabelecer limites de emissão harmônica em sistemas elétricos de distribuição de forma a se evitar interferências em outros equipamentos e em sistemas de telecomunicação.

Desta forma, a IEEE Std. 519 estabelece definições para os diversos parâmetros elétricos relacionados ao fenômeno eletromagnético. É descrita também a geração de harmônicos por conversores elétricos, bem como, seus efeitos e propagação pelo sistema. Em seguida a norma faz diversas considerações e recomendações acerca do controle e da correção do fator de potência na presença de harmônicas.

Com relação à medição de correntes e tensões harmônicas, a IEEE Std. 519 faz uma extensa abordagem sobre os principais tipos de equipamentos utilizados para medição e os requisitos básicos exigidos para um equipamento desta natureza.

Por fim, a norma traz exemplos e práticas sugeridas para a avaliação de harmônicos tanto em indústrias quanto em concessionárias de energia.

b) IEEE Std. 1159 – 1995; *IEEE RECOMMENDED PRACTICE FOR MONITORING ELECTRIC POWER QUALITY*

Este conjunto de recomendações do IEEE abrange a monitoração de praticamente todos os itens de qualidade da energia elétrica. Assim, inicialmente, encontra-se uma série de definições a respeito dos mais diferentes fenômenos de qualidade da energia, tais como, afundamentos de tensão, transitórios, harmônicos, flutuações de tensão, ruídos, dentre outros. Na seqüência do documento, verifica-se uma série de observações acerca de instrumentos de monitoração de QEE. Aspectos ligados à medições de tensões e correntes são devidamente exploradas, expondo sobretudo, as dificuldades impostas pelos transdutores (Transformadores

Potencial (TP) e Transformadores de Corrente (TC)) quando se deseja medir altas frequências.

O item 7 (sete) da IEEE 1159 é o que merece maior atenção. Neste, estão descritas as principais técnicas a serem observadas quando da realização de medições de qualidade da energia elétrica. Temas como local de instalação do equipamento, segurança, conexões, aterramento, dentre outros, são descritos, de forma a auxiliar engenheiros e demais profissionais, quanto à instalação do equipamento monitor, bem como diagnosticar possíveis problemas com o mesmo. A norma apresenta também recomendações quanto à análise e interpretação dos resultados das medições, mostrando tabelas de limites para cada fenômeno e formas de onda típicas. Finalmente, a norma faz menção à calibração de equipamentos monitores de qualidade da energia elétrica.

c) IEC 61000-4-30 Ed. 1.1 – *ELECTROMAGNETIC COMPATIBILITY (EMC) – PART 4-30: TESTING AND MEASUREMENTS TECHNIQUES – POWER QUALITY MEASUREMENT METHODS*

A Recomendação IEC 61000-4-30 define os métodos e padrões recomendados para a realização de medições e interpretação de parâmetros relacionados à qualidade da energia em sistemas elétricos de 50 e 60 Hz.

Os métodos de medição são descritos para cada tipo de parâmetro passível de obter resultados confiáveis, reproduzíveis e comparáveis com respeito ao instrumento utilizado e às condições ambientais. A medição destes é limitada àqueles encontrados no sistema de energia que são possíveis de leituras “*in situ*”.

Neste particular, os parâmetros considerados por este documento são: a frequência do sistema, a magnitude da tensão de suprimento, flutuação de tensão, as variações de tensão de curta duração, interrupções de tensão, transitórios, desequilíbrios de tensão, harmônicos, inter-harmônicos e mudanças abruptas na tensão.

Vale ressaltar que estas recomendações fornecem diretrizes para a realização de medições, mas não indicam os valores limites de referência a serem adotados durante as medições. Os efeitos associados aos transdutores também são

contemplados neste documento, bem com as preocupações durante a instalação dos equipamentos monitores.

d) IEC 61000-4-7 Ed. 1.1 – *ELECTROMAGNETIC COMPATIBILITY (EMC) – PART 4-7: TESTING AND MEASUREMENTS TECHNIQUES – GENERAL GUIDE ON HARMONICS AND INTERHARMONICS MEASUREMENTS AND INSTRUMENTATION, FOR POWER SUPPLY SYSTEMS AND EQUIPMENT CONNECTED THERETO*

Esta recomendação é aplicável à instrumentação utilizada para medição dos componentes do espectro de frequência até 9 kHz, as quais são superpostas à componente fundamental, seja da tensão e/ou da corrente, nos sistemas de potência de 50 ou 60 Hz. Por questões práticas, esta recomendação faz distinção entre harmônicos e interharmônicos e demais componentes do espectro de frequência que se situam acima de 9 kHz.

e) NRS 048

Este documento trata-se da norma sul-africana para a qualidade da energia elétrica. Elaborada pelas empresas concessionárias de energia elétrica da África do Sul, esta norma foi baseada especialmente nas normas internacional IEC e europeia CENELEC. O documento principal é dividido em 4 partes que abordam desde os critérios para implantação da norma até aspectos ligados aos transdutores utilizados para realização das medições.

No que tange à parte de monitoração da qualidade de energia, a NRS 048-3 aborda os requisitos mínimos para medição da qualidade de energia nos pontos de conexão dos consumidores. Já a NRS 048-5 trata mais especificamente da instrumentação necessária para a realização das medições de qualidade da energia. Neste sentido, o documento classifica os equipamentos de monitoração em três categorias, de acordo com a capacidade e precisão dos mesmos. Assuntos ligados à instalação, aspectos mecânicos, formatação de dados, calibração, precisão, dentre outros também são destacados ao longo do texto.

f) ONS – SUBMÓDULO 2.2 – PADRÕES DE DESEMPENHO DA REDE BÁSICA

O ONS, através do Submódulo 2.2 Padrões de Desempenho da Rede Básica (Sistema Elétrico com Tensão $\geq 230\text{Kv}$) oferece níveis tanto para as ordens ímpares como, também, para as componentes pares de tensão. Como podem ser visualizados na Tabela 2.2, estes são considerados para dois níveis distintos da tensão de operação. Todavia, vale mencionar que este documento nacional não tem caráter normativo, com estabelecimento de punições aos infratores, sendo apenas uma recomendação aos diversos agentes conectados à rede básica.

TABELA 2.2 – LIMITES GLOBAIS DE DISTORÇÃO HARMÔNICA DE TENSÃO EXPRESSOS EM PORCENTAGEM DA TENSÃO FUNDAMENTAL.

V < 69 Kv				V \geq 69 kV			
ÍMPARES		PARES		ÍMPARES		PARES	
Ordem	Valor	Ordem	Valor	Ordem	Valor	Ordem	Valor
3, 5, 7	5%			3, 5, 7	2%		
		2, 4, 6	2%			2, 4, 6	1%
9, 11, 13	3%			9, 11, 13	1,5%		
		≥ 8	1%			≥ 8	0,5%
15 a 25	2%			15 a 25	1%		
≥ 27	1%			≥ 27	0,5%		
DTHv = 6%				DTHv = 3%			

FONTE: SUBMÓDULO 2.2 PADRÕES DE DESEMPENHO DA REDE BÁSICA

Dentre os artigos técnico-científicos relacionados a monitoração e análise das distorções harmônicas vale ressaltar os seguintes:

IGNATOWICZ e BRAZ (2003) apresentam resultados de uma campanha de monitoração de harmônicos realizada pela CELESC em uma amostra de 200 consumidores da rede secundária, para conhecer o nível de harmônicas de tensão presentes no ponto de entrega. Os resultados apontaram a média de violações em cada hora por consumidor, durante o período do tempo estabelecido, o tempo médio

de permanência e a proporção dos consumidores que sofreram violação de harmônicas na amostra selecionada.

BRASIL e MEDEIROS (2004) apresentam os resultados do desempenho conjunto de instrumentos de medição que participaram de uma ampla campanha de medição de campo e testes de laboratório organizados pelo ONS. Isto permitiu avaliar o desempenho de um número expressivo de instrumentos de medição comercialmente disponíveis, bem como avaliar a metodologia de ensaio utilizada e o processo de avaliação de desempenho e, conseqüentemente, seleção dos instrumentos de medição.

Ressalta-se que os trabalhos conduzidos durante as medições de campo não consideraram questões relacionadas a precisão ou metodologia de medida. Todos os instrumentos foram alimentados com o mesmo sinal de tensão, de forma simultânea. Foram testados instrumentos considerando dois tipos característicos de cargas não-lineares. O primeiro corresponde a um forno a arco e o segundo a um retificador.

Diferentemente dos testes de campo que relacionaram somente a distorção harmônica, os testes de laboratório incluem a avaliação do desempenho de instrumentos relativo a outros índices de qualidade da energia elétrica, tais como: tensão em regime permanente, variação de freqüência, desequilíbrio de tensão, flutuação de tensão e variação de tensão de curta duração (VTCD).

Neste sentido, os testes permitiram avaliar o comportamento de um número significativo de instrumentos, quanto a sua precisão, através da comparação dos valores medidos com os valores impostos pela fonte utilizada nos testes, bem como através da comparação relativa entre os diversos instrumentos. O processo permitiu uma seleção preliminar de instrumentos classificados como aceitáveis/adequados para participarem de medições de grandezas associadas a qualidade da energia elétrica o que, certamente, contribui para uma melhor gestão do desempenho dos indicadores da qualidade de energia elétrica quer seja através de monitoração contínua quer seja através da realização de campanhas de medição.

KAGAN, FERRARI e MATSUO (2002) apresentam as principais considerações a respeito de uma metodologia utilizada para estimar níveis harmônicos em sistemas elétricos de distribuição.

Os estudos mostram que os erros que ocorrem ao longo das medições de harmônicos podem ser causados pela falta de sincronismo do circuito de amostragem e a frequência fundamental da rede monitorada.

Foram também realizados testes em equipamentos de monitoração, evidenciando, mais uma vez, a necessidade de se estabelecer normas e procedimentos voltados para a avaliação de parâmetros de qualidade de energia. Neste sentido, os autores propõem alguns níveis de distorção harmônica total e individual.

A respeito dos índices globais para harmônicos o artigo propõe os valores correspondentes ao $P_{95\%}$ para todos os locais monitorados.

UNGUREANU e SCUTARIU (1998) apresentaram a determinação da impedância harmônica de um sistema de distribuição industrial com diversos consumidores e uma grande planta industrial. Localizada na Romênia, esta indústria é representada tanto por cargas não lineares passivas (fornos a arco e fornos de indução), como por cargas não lineares ativas (inversores de frequência, ASD's) contribuindo de formas distintas para a poluição harmônica do sistema.

Para a realização dos trabalhos de determinação das impedâncias harmônicas foi utilizado um analisador de harmônico da BMI, que permitiu obter leituras simultâneas de tensão e corrente nos pontos escolhidos para a monitoração.

BARKER e BURKE (1998) apresentam um trabalho que trata de uma extensa campanha de medição de qualidade de energia realizada pela Niagra Mohawk Power Corporation (NMPC) em conjunto com o Departamento Americano de Pesquisa e Desenvolvimento em dois alimentadores de distribuição em Buffalo, região de Nova York. Para tanto, monitorou-se todos os níveis possíveis do sistema, desde subestações até a entrada de serviço dos consumidores foram equipados com equipamentos de monitoração. Assim, pode-se obter dados tanto de eventos transitórios, quanto leituras em regime permanente dos parâmetros desejados. Algumas importantes conclusões a respeito deste trabalho são listadas abaixo:

- É opinião dos autores que as reais causas dos problemas de qualidade da energia só podem ser identificadas através da medição simultânea dos distúrbios tanto no lado primário quanto no secundário;
- Os resultados mostram que a maioria dos problemas de qualidade da energia experimentados por consumidores, exceto interrupções de tensão, é gerada dentro de suas próprias plantas e casas;
- Os estudos mostraram que interrupções momentâneas e sustentadas são as principais causas de problemas para consumidores residenciais. Afundamentos de tensão são problemas típicos reclamados por consumidores industriais;
- Os fabricantes de equipamentos elétricos deveriam se atentar para as condições atuais do sistema, e não para os parâmetros estabelecidos em normas.

2.4 ANÁLISE MULTIVARIADA

2.4.1 Introdução

Ao se tomar uma decisão, muitos fatores costumam estar envolvidos nela, porém nem todos têm a mesma importância. Quando a intuição é utilizada na tomada de decisão, nem todos os fatores costumam ser identificados, ou seja, não serão definidas todas as variáveis que afetam tal decisão. Assim, também, nota-se que um grande número de variáveis envolve os acontecimentos sejam elas culturais ou naturais.

A Análise Multivariada utiliza métodos estatísticos para descrever e analisar dados de muitas variáveis simultaneamente. A necessidade de entender o relacionamento entre diversas variáveis aleatórias faz da Análise Multivariada uma metodologia com grande potencial de aplicação, principalmente na época atual devido à computação veloz e de baixo custo.

A Análise Multivariada trata, principalmente, dos problemas relacionados com:

- Reconhecimento e classificação de padrões (presente caso);
- Análise das estruturas de covariância de um vetor aleatório;
- Análise de Agrupamento;
- Análise de Correlação Canônica.

E ainda segundo Pla (1986), a Análise Multivariada pode ser usada para:

- Encontrar a adequação de representar o universo de estudo, simplificando a estrutura dos dados;
- Classificação: esta análise permite estabelecer as observações dentro de grupos ou, então, concluir que os indivíduos estão aleatórios no multiespaço, sendo também possível alocar novos itens em grupos já identificados;
- Análise de independência: tem como objetivo examinar a independência entre as variáveis, a qual abrange desde a independência total até a colinearidade, quando uma delas é combinação linear de outras.

De acordo com Johnson e Wichern (1998), no estudo de problemas que envolvem p variáveis ($p \geq 1$), tomam-se n observações de cada vetor aleatório \underline{x} a partir da matriz de dados X de ordem $n \times p$ constrói-se as técnicas da Análise Multivariada. Assim, as medidas registradas são X_{ij} com $j = 1, 2, \dots, p$ e $i = 1, 2, \dots, n$ que podem ser agrupadas na matriz de dados genérica ${}_n X_p$.

$${}_n X_p = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \quad (2.9)$$

A representação da matriz de dados corresponde a n observações do vetor aleatório $\underline{X}' = [X_1, X_2, \dots, X_p]$ de dimensão p , composto por p variáveis aleatórias. Essa matriz pode ser escrita como ${}_n X_p = (X_{ij})$. E como é bem conhecida, essa matriz corresponde a uma amostra aleatória de tamanho n do vetor p -dimensional \underline{X} .

2.4.2 Estatística Descritiva Multivariada

A Ciência Estatística trabalha com amostras. As informações amostrais podem ser resumidas em números sumários conhecidos como estatísticas e que podem resumir as observações multivariadas $[x_1, x_2, \dots, x_n]$. Essas estatísticas são usadas nas inferências de parâmetros, ou seja, na estimação do vetor médio populacional μ , da matriz de covariância Σ ou da matriz de correlação ρ , entre outros, de maneira que o vetor médio μ deve ser estimado pelo vetor amostral \bar{X} , definido pela expressão:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.10)$$

Portanto, o vetor médio $\mu = E(x)$ é estimado por \bar{X} onde x_i com $i = 1, 2, \dots, n$ corresponde às observações amostrais do vetor X e n é o tamanho da amostra observada. Outros parâmetros de uma população multivariada com função $f(x)$ podem ser avaliados. Tais como a matriz de covariância Σ do vetor X , que é definida por:

$$\Sigma = V(x) = E[(x - \mu)(x - \mu)'] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix} \quad (2.11)$$

com $\sigma_i^2 = E[(x_i - \mu_i)^2]$ e $\text{cov}(x_i, x_j) = \sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$ onde se tem na diagonal principal as variâncias das variáveis aleatórias e, fora da diagonal principal, as covariâncias entre elas. E, a matriz de correlação ρ , definida por:

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix} \quad (2.12)$$

onde $\rho_{ij} = \frac{\text{COV}(x_i, x_j)}{\sigma_i \sigma_j}$ com $i=1,2,\dots,p$ com as correlações entre as variáveis fora da diagonal principal. Então estes parâmetros, Σ e ρ , são estimados, respectivamente, pela matriz de covariância amostral S e pela matriz de correlação amostral R ou $\hat{\rho}$, cujas expressões são:

$$S = \frac{\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'}{n-1} = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{bmatrix} \quad (2.13)$$

sendo s_j^2 a variância amostral da variável aleatória X_j , com expressões

$$s_j^2 = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1} \quad (2.14)$$

e s_{jk} a covariância amostral entre as variáveis aleatória x_j e x_k , ou seja,

$$s_{jk} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_{1k})}{n-1} \quad (2.15)$$

e

$$R = \hat{\rho} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \quad (2.16)$$

As correlações amostrais fora da diagonal principal e dadas pelo quociente entre a covariância amostral e o produto dos desvios padrões amostrais, ou seja:

$$r_{jk} = \frac{s_{jk}}{s_j s_k} \quad \text{para } j \neq k \quad (2.17)$$

Esses estimadores são os melhores para se determinar os parâmetros μ , Σ e ρ . Os dois primeiros são UMVU (Estimadores Uniformemente de Mínima Variância não-viciado) e o último é EMV (Estimador de Máxima Verossimilhança).

Na Análise Multivariada é preferível se trabalhar com a matriz de correlação R ao invés de se trabalhar com a matriz de covariância, pois dessa forma elimina-se o problema com escala diferentes das variáveis.

2.5 RECONHECIMENTO DE PADRÕES

2.5.1 Discriminação, Classificação e Reconhecimento de Padrão

Segundo Chaves Neto (1997), estatisticamente, a construção de regras de reconhecimento e classificação de padrões para duas populações podem ser baseadas em quatro métodos principais: Função Discriminante Linear de Fisher, Escore Discriminante Quadrático, Regressão Logística e Método das k-médias. Posteriormente, surgiu a tecnologia de Redes Neurais (tecnologia emergente), método de Programação Matemática além de outros métodos para formação do conjunto de procedimentos usados no reconhecimento e classificação de objetos e indivíduos.

2.5.2 Análise Discriminante

Segundo Johnson e Wichern (1998), a Análise Discriminante é uma técnica multivariada que tem por objetivo tratar dos problemas relacionados em separar conjuntos distintos de objetos (itens ou observações) e alocar novos objetos em conjuntos previamente definidos. Quando empregada como procedimento de classificação não é uma técnica exploratória, uma vez que ela conduz a regras bem definidas, as quais podem ser utilizadas para classificação de outros objetos.

Tem como objetivos imediatos, quando usada para discriminação e classificação, os seguintes:

1. Descrever algébrica ou graficamente as características diferenciais dos objetos (observações) de várias populações conhecidas a fim de achar “discriminantes” cujos valores numéricos sejam tais que as populações possam ser separadas tanto quanto possível.

2. Agrupar os objetos (observações) dentro de duas ou mais classes determinadas. Tenta-se encontrar uma regra que possa ser usada na alocação ótima de um novo objeto (observação) nas classes consideradas. Uma função que separa pode servir para alocar um objeto e, da mesma forma, uma regra alocadora pode sugerir um procedimento discriminatório.

Na prática, os objetivos 1 e 2, freqüentemente, sobrepõem-se e a distinção entre separação e alocação torna-se confusa.

A Análise Discriminante foi originalmente desenvolvida na Botânica tendo por objetivo fazer a distinção de grupos de plantas com base no tamanho e no tipo de folha o que tornaria possível, posteriormente, classificar as novas espécies encontradas. Segundo (LIMA, 2002) em 1936, Fischer foi responsável pelo desenvolvimento da análise para dois grupos.

2.5.3 Problema Geral da Classificação com Várias Populações

Considere:

- $f_i(\underline{X})$ a função densidade de probabilidade associada à população Π_i , com $i = 1, \dots, g$;
- p_i a probabilidade a priori da observação pertencer à população Π_i ;
- R_k o conjunto dos \underline{X} classificados em Π_k ;
- $c(k|i)$ o custo de classificação de uma observação em Π_k quando na verdade pertence a Π_i , com $i = 1, \dots, g$ e $c(i|i) = 0$ e
- $P(k|i)$ a probabilidade de se classificar uma observação de Π_i em Π_k sendo:

$$P(k|i) = \int_{R_k} f_i(\underline{X}) dx \quad i = 1, \dots, g \quad (2.18)$$

O Custo Esperado de Reconhecimento Errado (*ECM*) ao se classificar \underline{X} de Π_i em Π_k , com $k = 1, \dots, g$ é dado por:

$$ECM(1) = P(2|1)c(2|1) + P(3|1)c(3|1) + \dots + P(g|1)c(g|1) = \sum_{k=2}^g P(k|1)c(k|1) \quad (2.19)$$

De maneira análoga, obtem-se o $ECM(2)$, $ECM(3)$, ..., $ECM(g)$. Considerando que $ECM(k)$ ocorre com probabilidade p_k com $k = 1, 2, \dots, g$ então:

$$ECM = p_1 ECM(1) + p_2 ECM(2) + \dots + p_g ECM(g) \quad (2.20)$$

logo,

$$ECM = p_1 \left(\sum_{k=2}^g P(k|1)c(k|1) \right) + \sum_{\substack{k=1 \\ k \neq 2}}^g P(k|2)c(k|2) + \dots + \sum_{k=1}^{g-1} P(k|g)c(k|g) \quad (2.21)$$

A regra de classificação consiste em se determinar R_k (com $k = 1, 2, \dots, g$) que tornem o ECM mínimo tanto quanto possível, isto significa alocar \tilde{X} na população Π_k de forma que:

$$\sum_{\substack{k=1 \\ k \neq 2}}^g P_i f(\tilde{X}) c(k|i) \quad (2.22)$$

é menor e, num caso de empate, \tilde{X} pode ser classificado em qualquer uma das duas populações para as quais o empate ocorre.

No caso de todos os custos serem iguais, é suficiente alocar \tilde{X} na população Π_k em que:

$$\sum_{\substack{i=1 \\ i \neq k}}^g P_i f(\tilde{X}) \quad (2.23)$$

é menor. Mas essa quantidade será menor quando o termo excluído $P_i f_k(\tilde{X})$ for maior. Assim, a regra de classificação do ECM mínimo com custos iguais por falhas de classificação é alocar \tilde{X} em Π_k se:

$$P_k f_k(\tilde{X}) > P_i f_i(\tilde{X}) \quad \text{para todo } i \neq k \quad (2.24)$$

ou, alternativamente, alocar \tilde{X} em Π_k se:

$$\ln P_k f_k(\tilde{X}) > \ln P_i f_i(\tilde{X}) \quad \text{para todo } i \neq k \quad (2.25)$$

2.5.4 Análise Discriminante para Duas Populações – Método de Fischer

A idéia de Fischer foi transformar observações multivariadas $\tilde{X}'s$ em observações univariadas $Y's$ tais que os $Y's$ das populações Π_1 e Π_2 sejam separadas tanto quanto possível. Fischer tomou combinações lineares de \tilde{X} para criar os $Y's$, dado que as combinações lineares são funções de \tilde{X} e por outro lado são de fácil cálculo. Assim sendo μ_{1y} a média dos $Y's$ obtidos dos $\tilde{x}'s$ pertencentes a Π_1 (população 1) e μ_{2y} a média dos $Y's$ obtidos dos $\tilde{x}'s$ pertencentes a Π_2 (população 2), Fischer selecionou a combinação linear que maximiza a distância quadrática entre μ_{1y} e μ_{2y} relativamente à variabilidade dos $Y's$. Assim, seja:

$$\mu_1 = E(\tilde{X}|\Pi_1) = \text{valor esperado em uma observação multivariada em } \Pi_1. \quad (2. 26)$$

$$\mu_2 = E(\tilde{X}|\Pi_2) = \text{valor esperado de uma observação multivariada em } \Pi_2. \quad (2. 27)$$

e supondo a matriz de covariância $\Sigma = E\left[(\tilde{X} - \mu_i)(\tilde{X} - \mu_i)'\right]$ $i = 1, 2$ como sendo a mesma para as duas populações, e ainda considerando a combinação linear:

$$Y_{1 \times 1} = c'_{1 \times p} \tilde{X}_{p \times 1} \quad (2. 28)$$

tem-se :

$$\mu_{1y} = E(Y|\Pi_1) = E(c' \tilde{X}|\Pi_1) = c' E(\tilde{X}|\Pi_1) = c' \mu_1 \quad (2. 29)$$

$$\mu_{2y} = E(Y|\Pi_2) = E(c' \tilde{X}|\Pi_2) = c' E(\tilde{X}|\Pi_2) = c' \mu_2 \quad (2. 30)$$

e

$$V(Y) = \sigma_y^2 = V(\underline{c}' \underline{X}) = \underline{c}' V(\underline{X}) \underline{c} = \underline{c}' \underline{\Sigma} \underline{c} \quad (2.31)$$

que é supostamente a mesma para ambas as populações. Então, segundo Fischer, a melhor combinação linear é a derivada da razão entre o “quadrado da distância entre as médias” e a “variância de Y”.

$$\frac{(\mu_{1y} - \mu_{2y})^2}{\sigma_y^2} = \frac{(\underline{c}' \underline{\mu}_1 - \underline{c}' \underline{\mu}_2)^2}{\underline{c}' \underline{\Sigma} \underline{c}} = \frac{\underline{c}' (\underline{\mu}_1 - \underline{\mu}_2) (\underline{\mu}_1 - \underline{\mu}_2)' \underline{c}}{\underline{c}' \underline{\Sigma} \underline{c}} = \frac{(\underline{c}' \underline{\delta})^2}{\underline{c}' \underline{\Sigma} \underline{c}} \quad (2.32)$$

onde:

$$\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2 .$$

Assim, como $\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2$ e $Y = \underline{c}' \underline{X}$ tem-se que $\frac{(\underline{c}' \underline{\delta})^2}{\underline{c}' \underline{\Sigma} \underline{c}}$ é maximizada por:

$$\underline{c} = k \underline{\Sigma}^{-1} \underline{\delta} = k \underline{\Sigma}^{-1} (\underline{\mu}_1 - \underline{\mu}_2) \text{ para qualquer } k \neq 0. \quad (2.33)$$

Logo por multiplicidade com $k = 1$ tem-se:

$$\underline{c} = \underline{\Sigma}^{-1} (\underline{\mu}_1 - \underline{\mu}_2) \text{ e } Y = \underline{c}' \underline{X} = (\underline{\mu}_1 - \underline{\mu}_2)' \underline{\Sigma}^{-1} \underline{X}, \quad (2.34)$$

que é conhecida como Função Discriminante Linear de Fisher. Ela transforma as populações multivariadas Π_1 e Π_2 em populações univariadas, tais que as médias destas populações são separadas tanto quanto possível relativamente à variância populacional, considerada comum. E, para classificar a observação multivariada x_0

usa-se o modelo:

$$y_0 = (\underline{\mu}_1 - \underline{\mu}_2)' \underline{\Sigma}^{-1} x_0 \quad (2.35)$$

onde y_0 é o valor da Função Discriminante de Fisher para a nova observação \tilde{x}_0 , e considerando-se o ponto médio entre as médias das duas populações univariadas,

$$m = \frac{1}{2}(\mu_{1y} + \mu_{2y}), \quad (2.36)$$

como

$$m = \frac{1}{2}(c'_1 \mu_1 + c'_2 \mu_2), \quad (2.37)$$

$$m = \frac{1}{2} \left[(\mu_1 - \mu_2)' \Sigma^{-1} \mu_1 + (\mu_1 - \mu_2)' \Sigma^{-1} \mu_2 \right] \quad (2.38)$$

$$m = \frac{1}{2} \left[(\mu_1 - \mu_2)' \Sigma^{-1} (\mu_1 + \mu_2) \right] \quad (2.39)$$

tem-se que:

$$E(y_0 | \Pi_1) - m \geq 0 \quad (2.40)$$

e

$$E(y_0 | \Pi_2) - m < 0 \quad (2.41)$$

ou seja, se \tilde{x}_0 pertence a Π_1 espera-se que y_0 seja igual ou maior do que o ponto médio m . Por outro lado se \tilde{x}_0 pertence a Π_2 , o valor esperado de y_0 será menor que o ponto médio. Portanto, a regra de classificação é:

- alocar \tilde{x}_0 em Π_1 se $y_0 - m \geq 0$;
- alocar \tilde{x}_0 em Π_2 se $y_0 - m < 0$.

Os parâmetros μ_1 , μ_2 e Σ geralmente são desconhecidos. Então, supondo que se tem n_1 observações da variável aleatória multivariada x_1 de dimensão p , ou seja, tem-se uma amostra aleatória da população Π_1 e n_2 observações da variável aleatória multivariada x_2 de dimensão p que corresponde a uma amostra aleatória de população Π_2 , os resultados amostrais correspondentes são:

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{i1}; S_1 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)(x_{i1} - \bar{x}_1)' \quad (2.42)$$

$$\bar{x}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} x_{i2}; S_2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)(x_{i2} - \bar{x}_2)' \quad (2.43)$$

Assumindo que as populações sejam assemelhadas, é natural considerar a variância como a mesma, daí estima-se a matriz de covariância comum Σ pela matriz de covariância amostral calculada com as amostras conjuntas,

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)} \quad (2.44)$$

que é um estimador não –viciado daquele parâmetro Σ .

Consequentemente, a Função Discriminante Linear de Fischer Amostral é dada por:

$$\hat{Y} = \hat{c} X = (\bar{x}_1 - \bar{x}_2)' S_p^{-1} x \quad (2.45)$$

e a estimativa do ponto médio entre as duas médias amostrais univariadas,

$$\bar{y}_1 = \hat{c} \bar{x}_1 \quad (2.46)$$

e

$$\bar{y}_2 = \hat{c}' \bar{x}_2 \quad (2.47)$$

é dada por:

$$\hat{m} = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) = \frac{1}{2} \left[(\bar{x}_1 - \bar{x}_2)' S_p^{-1} \bar{x}_1 + (\bar{x}_1 - \bar{x}_2)' S_p^{-1} \bar{x}_2 \right] \quad (2.48)$$

$$\hat{m} = \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' S_p^{-1} (\bar{x}_1 + \bar{x}_2)$$

Finalizando, a regra de classificação do item caracterizado pelo vetor x_0 é:

Se $\hat{y}_0 - \hat{m} \geq 0$ então x_0 é alocado em Π_1

ou

Se $\hat{y}_0 - \hat{m} < 0$ então x_0 é alocado em Π_2

A combinação linear particular $Y = \hat{c}' x = (\bar{x}_1 - \bar{x}_2)' S_p^{-1} x$ maximiza a razão:

$$\frac{(\bar{y}_1 - \bar{y}_2)^2}{S_y^2} = \frac{(\hat{c}' \bar{x}_1 - \hat{c}' \bar{x}_2)}{\hat{c}' S_p \hat{c}} = \frac{(\hat{c}' d)^2}{\hat{c}' S_p \hat{c}} \quad (2.49)$$

onde:

$$d = \bar{x}_1 - \bar{x}_2 \quad (2.50)$$

e

$$S_y^2 = \frac{\sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{i2} - \bar{y}_2)^2}{n_1 + n_2 - 2} \quad (2.51)$$

2.5.5 Discriminação Entre Diversas Populações

O método anterior que é válido para duas populações $g=2$, pode ser estendido para diversas populações $g \geq 2$. O primeiro objetivo de Fischer com o método foi o de separar populações, podendo ser usado também para classificar novos itens em uma das populações. Esse método não necessita da suposição de que as diversas populações sejam normais multivariadas, porém é necessário assumir que as matrizes de covariância populacionais sejam iguais, ou seja, $\Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$.

Assim é preciso testar a hipótese de igualdade das matrizes de covariância, ou seja $H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$. Então tem-se o seguinte teste:

- Hipótese nula: $H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$
- Hipótese alternativa: Pelo menos uma das matrizes é diferente das demais.

A estatística de teste é:

$$\text{Estatística de Teste } M = \left(\frac{|S_i|}{|S_p|} \right)^{\frac{(n_i-1)}{2}}$$

onde:

S_i é a estimativa da matriz de covariância do grupo i ;

S_p é a estimativa da matriz de covariância conjunta;

n_i é o número de observações do grupo i . (tamanho da amostra)

A transformação logarítmica de M conduz a estatística abaixo que é de tratamento matemático mais adequado:

$$B = (1 - c) \left\{ \left[\sum_{i=1}^g (n_i - 1) \right] \ln |S_p| - \sum_{i=1}^g [(n_i - 1) \ln |S_i|] \right\} \sim \chi_{\frac{1}{2}p(p+1)(g+1)}^2 \quad (2.52)$$

com

$$c = \left[\frac{\sum_{i=1}^g \frac{1}{n_i - 1}}{\sum_{i=1}^g (n_i - 1)} - \frac{1}{\sum_{i=1}^g (n_i - 1)} \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right] \quad (2.53)$$

onde

p é a dimensão do vetor $\underset{\sim}{x}$ (número de variáveis);

n_i é o tamanho do grupo i ;

g é o número de grupos.

Então uma vez aceita a hipótese nula H_0 de igualdade entre as matrizes de covariância, seja $\underset{\sim}{\bar{\mu}}$ o vetor médio dos diversos grupos (populações), representado por:

$$\underset{\sim}{\bar{\mu}} = \frac{1}{g} \sum_{i=1}^g \underset{\sim}{\mu}_i \quad (2.54)$$

e B_0 a matriz “Soma de produtos cruzados entre grupos populacionais” tal que:

$$B_0 = \sum_{i=1}^g (\underset{\sim}{\mu}_i - \underset{\sim}{\bar{\mu}})(\underset{\sim}{\mu}_i - \underset{\sim}{\bar{\mu}})' \quad (2.55)$$

a combinação linear $\underset{\sim}{Y} = \underset{\sim}{c}' \underset{\sim}{x}$ tem por esperança:

$$E(Y) = \underset{\sim}{c}' E(\underset{\sim}{x} | \Pi_i) = \underset{\sim}{c}' \underset{\sim}{\mu}_i \quad (2.56)$$

para população Π_1 e variância:

$$V(Y) = \sigma_y^2 = \underset{\sim}{c}' V(\underset{\sim}{X}) \underset{\sim}{c} = \underset{\sim}{c}' \underset{\sim}{\Sigma} \underset{\sim}{c} \quad (2.57)$$

para todas as populações. Desta forma, o valor esperado $\mu_{iy} = \underset{\sim}{c}' \underset{\sim}{\mu}_i$ muda quando a população da qual $\underset{\sim}{X}$ selecionado é outra. Tem-se então uma média global:

$$\bar{\mu}_y = \frac{1}{g} \sum_{i=1}^g \mu_{iy} = \underset{\sim}{c}' \bar{\mu} \quad (2.58)$$

e a razão entre a “Soma dos quadrados das distâncias das populações para a média global” e a variância de Y é $\frac{\underset{\sim}{c}' B_0 \underset{\sim}{c}}{\underset{\sim}{c}' \underset{\sim}{\Sigma} \underset{\sim}{c}}$, que é uma generalização multigrupal do caso de duas populações. Medindo a variabilidade entre grupos de valores (escores) Y relativamente à variabilidade comum dentro dos grupos, da mesma forma do que no caso de duas populações, pode-se selecionar $\underset{\sim}{c}$ que maximiza esta razão. É conveniente normalizar $\underset{\sim}{c}$ tal que $\underset{\sim}{c}' \underset{\sim}{\Sigma} \underset{\sim}{c} = 1$.

Sejam $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s \geq 0$ os $s \leq \min(g-1, p)$ autovalores não-nulos de $\Sigma^{-1} B_0$ e $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_s$ os correspondentes autovetores (escalonados tais que $\underset{\sim}{c}' \underset{\sim}{\Sigma} \underset{\sim}{c} = 1$). Então, é fácil provar que o vetor de coeficientes $\underset{\sim}{c}$ que maximiza $\frac{\underset{\sim}{c}' B_0 \underset{\sim}{c}}{\underset{\sim}{c}' \underset{\sim}{\Sigma} \underset{\sim}{c}}$ é dado por $\underset{\sim}{c}_1 = \underline{e}_1$. A combinação linear $\underset{\sim}{c}_1' \underset{\sim}{X}$ é chamada de primeira discriminante e de forma idêntica, pode-se generalizar para o k -ésimo discriminante com $\underset{\sim}{c}_k = \underline{e}_k$, com $k = 1, 2, \dots, s$. Como geralmente Σ e μ_i não são conhecidas, tomam-se amostras aleatórias de tamanho n_i das populações $\Pi_i, i = 1, 2, \dots, g$ e denotamos o conjunto de dados da

população $\Pi_i, i = 1, 2, \dots, g$, por ${}_{n_i}Y_p$ tem-se os estimadores dos parâmetros μ_i e $\bar{\mu}$ dados por:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad (2.59)$$

$$\bar{x} = \frac{\sum_{i=1}^g n_i \bar{x}_i}{\sum_{i=1}^g n_i} = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} x_{ij}}{\sum_{i=1}^g n_i} \quad (2.60)$$

A Matriz “Soma de produtos cruzados entre grupos” B_0 é estimada por:

$$\hat{B}_0 = \sum_{i=1}^g (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \quad (2.61)$$

e um estimador para Σ pode ser obtido com base na matriz W .

$$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x})(x_{ij} - \bar{x})' = \sum_{i=1}^g (n_i - 1)S_i \quad (2.62)$$

Conseqüentemente,

$$\frac{W}{n_1 + n_2 + \dots + n_g - g} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2 + \dots + (n_g - 1)S_g}{n_1 + n_2 + \dots + n_g - g} = S_p \quad (2.63)$$

Assim, o mesmo \hat{c} que maximiza a razão $\frac{\hat{c}' B_0 \hat{c}}{\hat{c}' S_p \hat{c}}$ também maximiza $\frac{\hat{c}' B_0 \hat{c}}{\hat{c}' W \hat{c}}$.

Logo, apresentar-se-á otimizante \hat{c} na forma mais usual, que é o autovalor \hat{e}_i da matriz $W^{-1}B_0$, porque se $W^{-1}B_0 \hat{e} = \hat{\lambda} \hat{e}$ então $S_p^{-1} \hat{B}_0 \hat{e} = \hat{\lambda} (n_1 + n_2 + \dots + n_g - g) \hat{e}$, portanto, concluindo que sejam $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_g > 0$ os autovalores não nulos de $W^{-1}B_0$ e

$\hat{e}_1, \hat{e}_2, \dots, \hat{e}_s$ os correspondentes autovalores, sendo $s \leq \min(g-1, p)$ e \hat{e}_i normalizado tal que $\hat{e}_i' S_p \hat{e}_i = 1$; então o vetor de coeficientes que maximiza a razão citada acima é $\hat{c}_1 = \hat{e}_1$ e a combinação linear $\hat{e}_1' x$ é chamada primeira discriminante amostral. Generalizando, tem-se no passo k o k -ésimo discriminante amostral $\hat{e}_k' x$, $k \leq s$.

2.5.6 Escore Discriminante Quadrático de Classificação

Quando a hipótese nula, H_0 , enunciada no item (2.5.5) não é aceita, uma das alternativas é optar-se pela técnica do escore discriminante quadrático de classificação. Considere o vetor de variáveis aleatórias X de populações com distribuição normal multivariada com média μ_i e matriz de covariância Σ_i , ou seja,

$$f_i(X) = \frac{1}{(2\pi)^{p/2} |\Sigma_i|} \exp\left[-\frac{1}{2} (X - \mu_i)' \Sigma_i (X - \mu_i)\right], \quad i = 1, 2, \dots, g \quad (2.64)$$

em que $\mu_i = E(x | \Pi_i)$ é o vetor de médias e $V(x | \Pi_i) = \Sigma_i$ é a matriz de covariância.

Tem-se ainda que os custos $c(i|i) = 0$ e $c(k|i) = 1, k \neq i, k = 1, 2, \dots, g$ e da equação (2.21), a regra de classificar fica:

alocar X em Π_k se

$$\ln p_i f_i(X) = \ln p_i - \left(\frac{p}{2}\right) \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (X - \mu_i)' \Sigma_i^{-1} (X - \mu_i), \quad i = 1, 2, \dots, g \quad (2.65)$$

Na equação pode-se ignorar a constante $\left(\frac{p}{2}\right) \ln(2\pi)$ já que é a mesma para todas as populações. Define-se, então, o escore discriminante quadrático de classificação d_i^0 para a população Π_i como:

$$d_i^Q(\tilde{X}) = \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (\tilde{X} - \mu_i)' \Sigma_i^{-1} (\tilde{X} - \mu_i) + \ln p_i, \quad i = 1, 2, \dots, g \quad (2.66)$$

Com várias populações normais, a regra de classificação consiste em classificar \tilde{X} em Π_k se:

$$d_i^Q(\tilde{X}) = \max_i d_i^Q(\tilde{X}), \quad i = 1, 2, \dots, g \quad (2.67)$$

Considerando que os parâmetros populacionais em geral não são conhecidos, as equações com os estimadores serão:

$$D_i = \frac{1}{2} \ln |S_i| - \frac{1}{2} (\tilde{X} - \bar{X}_i)' S_i^{-1} (\tilde{X} - \bar{X}_i) + \ln p_i, \quad i = 1, 2, \dots, g \quad (2.68)$$

onde:

\bar{X}_i é o vetor médio do grupo i ,

S_i é a matriz de covariância do grupo i .

P_i é a probabilidade a priori de um indivíduo ser classificado no grupo i .

2.5.7 Teste de Normalidade (Gaussianidade)

Segundo Johnson & Wichern (1998) a generalização da família das densidades normais para várias dimensões. Quando os dados estão realmente perto da normal multivariada, a densidade normal aproxima-se da verdadeira distribuição populacional. Outra vantagem da distribuição normal multivariada é que existe tratamento matemático para se obter resultados. Com a teoria do limite central, várias distribuições amostrais de estatística multivariada podem aproximar-se da normalidade.

Pode-se avaliar a normalidade conjunta de dados de distribuição com 2, 3, ..., p dimensão. Na prática é suficiente e usual investigar as distribuições univariada e

bivariada. As observações podem ser generalizadas para uma distribuição normal multivariada e a distribuição bivariada pode ser normal e que o contorno da constante da densidade pode ser uma elipse. No gráfico *scatter plot* é possível exibir uma elipse quando os dados tendem a uma normalidade.

O contorno da constante da densidade para p-dimensional da distribuição normal são elipses definidas por \tilde{X} indicadas por $(\tilde{X} - \tilde{\mu})' \Sigma^{-1} (\tilde{X} - \tilde{\mu}) = c^2$. As elipses são centradas na média $\tilde{\mu}$ e tem eixos $\pm \sqrt{\lambda_i} e_i$ onde $\Sigma e_i = \lambda_i e_i$ para $i = 1, 2, \dots, p$ onde Σ tem que ser uma matriz definida positiva com o autovetor e e os autovalores λ_i . O conjunto da distribuição bivariada é dado por $(\tilde{X} - \tilde{\mu})' \Sigma^{-1} (\tilde{X} - \tilde{\mu}) \leq \chi^2_2(0,5)$ e tem probabilidade de 0.5, ou seja 50% das observações das amostras são dadas pela elipse $(\tilde{X} - \tilde{\bar{X}})' S^{-1} (\tilde{X} - \tilde{\bar{X}}) \leq \chi^2_2(0,5)$, onde $\tilde{\mu}$ estima $\tilde{\bar{X}}$ e Σ^{-1} estima S^{-1} , caso contrário a normalidade não é aceita. Calculando as razões dos pontos de contorno e subjetivamente comparando com a teoria das probabilidades usuais.

Um método formal de julgar a normalidade de um conjunto de dados é baseada no quadrado da distância generalizada:

$$d_j^2 = \left(X_j - \tilde{\bar{X}} \right)' S^{-1} \left(X_j - \tilde{\bar{X}} \right), \quad j = 1, 2, \dots, n \quad (2.69)$$

onde $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ são as observações das amostras. O procedimento descrito não é limitado somente para o caso bivariado, mais também é valido para $p \geq 2$.

2.6 REGRESSÃO LOGÍSTICA

2.6.1 Introdução

A função Logística surgiu em 1845, ligada a problemas de crescimento demográfico. Em tais aspectos esta função desempenha um papel importante. A partir da década de 30 esta metodologia passou a ser aplicada no âmbito da biologia (LIMA, 2002).

A regressão logística vem sendo utilizada não só em Engenharia de Produção, mas também nas mais diversas áreas. Este método, assim como a regressão linear e múltipla, estuda a relação entre uma variável resposta e uma ou mais variáveis independentes. A diferença entre estas técnicas de regressão se deve ao fato de que na regressão logística as variáveis dependentes estão dispostas em categorias discriminatórias enquanto na regressão linear estas variáveis são dados contínuos. Outra diferença é que na regressão logística a resposta pode ser entendida como uma probabilidade de ocorrência $p \in [0,1]$, enquanto na regressão linear obtém-se um valor numérico $p \in \mathfrak{R}$.

2.6.2 Modelo de Regressão Logística

Seja uma variável aleatória Y (dicotômica) e $\underline{X}' = [X_1, X_2, \dots, X_p]$ um vetor de dimensão p , composto de variáveis aleatórias independentes e ainda tomando-se n observações independentes de Y e X_i com $i=1, 2, \dots, p$, pode-se escrever o modelo de Regressão Logística, na forma:

$$P(\underline{X}) = \frac{e^{\beta' \underline{X}}}{1 + e^{\beta' \underline{X}}} = \frac{1}{1 + e^{-\beta' \underline{X}}} \quad (2.70)$$

onde: $\beta' = [\beta_0, \beta_1, \dots, \beta_p]$ e $X' = [1, X_1, \dots, X_p]$

Segundo Johson & Wichern, (1998), esse modelo é o mais apropriado no caso da variável dependente ser dicotômica, pois a Regressão Logística estima diretamente a probabilidade de ocorrência de um evento. Estes autores, também salientam que a estimação dos parâmetros deve ser feita pelo método da máxima verossimilhança, sendo o mais recomendado quando dispõe-se de observações individuais da ocorrência ou não de determinado evento.

2.6.3 Modelo de Regressão Logística Simples

Seja a amostra aleatória composta de n pares de observações (X_i, Y_i) com $i = 1, 2, \dots, n$, onde os Y 's representam os valores observados de uma variável dicotômica, e os X 's representam os valores observados de uma única variável independente.

Assim o modelo $P(X) = \frac{e^{\beta X}}{1 + e^{\beta X}}$, torna-se:

$$P(Y = 1 | X) = P(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \quad (2.71)$$

Essa expressão fornece a probabilidade condicional de que Y seja igual a 1, dado o valor de X , ou seja, $P(Y = 1 | X)$. Neste caso é necessário estimar os valores dos parâmetros β 's. Então com isso determinando o modelo de Regressão Logística a probabilidade condicional de que Y ser igual a 0, é:

$$P(Y = 0 | X) = 1 - P(Y = 1 | X) \quad (2.72)$$

Portanto para as n observações têm-se:

$$P(Y_i = 1|X_i) = P(X_i) \text{ e } P(Y_i = 0|X_i) = 1 - P(X_i) \quad (2.73)$$

O método da Máxima Verossimilhança é melhor método para se estimar os parâmetros do modelo de Regressão Logística, ou seja, fornece estimativas para os parâmetros que maximizam a probabilidade de obter o conjunto observado de dados.

Uma forma conveniente para expressar a contribuição da função de verossimilhança para os pares (X_i, Y_i) é através da fórmula:

$$F(X_i) = P(X_i)^{Y_i} [1 - P(X_i)]^{1-Y_i} \quad (2.74)$$

Desde que assumindo que as observações são independentes, a função de verossimilhança é obtida como o produto dos termos dados na equação (2.57) como segue:

$$l(\beta) = \prod_{i=1}^n F(X_i) = \prod_{i=1}^n [P(X_i)^{Y_i}][1 - P(X_i)]^{1-Y_i} \quad (2.75)$$

Os estimadores de Máxima Verossimilhança destes parâmetros são escolhidos de forma a maximizar essa função, ou seja, deseja-se determinar o estimador para β que maximize a expressão anterior. Analisando essa função do ponto de vista matemático é mais fácil trabalhar com o logaritmo da mesma. Assim, tem-se a função de log-verossimilhança dada por:

$$L(\beta) = \ln(l(\beta)) = \ln \left[\prod_{i=1}^n F(X_i) \right] = \ln \left[\prod_{i=1}^n [P(X_i)^{Y_i}][1 - P(X_i)]^{1-Y_i} \right] \quad (2.76)$$

Assim:

$$L(\beta) = \sum_{i=1}^n [Y_i \ln(P(X_i))] + (1 - Y_i) \ln[1 - P(X_i)] \quad (2.77)$$

Então:

$$L(\beta) = \sum_{i=1}^n \left[Y_i \ln \frac{e^{-\beta'X}}{1+e^{-\beta'X}} + (1-Y_i) \ln \left(1 - \frac{e^{-\beta'X}}{1+e^{-\beta'X}} \right) \right] \quad (2.78)$$

Temos:

$$L(\beta) = \sum_{i=1}^n \left[Y_i \ln \frac{e^{-\beta'X}}{1+e^{-\beta'X}} + (1-Y_i) \ln \left(\frac{1}{1+e^{-\beta'X}} \right) \right] \quad (2.79)$$

Mas

$$\ln \left(\frac{e^{-\beta'X}}{1+e^{-\beta'X}} \right) = \ln e^{-\beta'X} - \ln(1+e^{-\beta'X}) = e^{-\beta'X} - \ln(1+e^{-\beta'X}) \quad (2.80)$$

e

$$\ln \left(\frac{1}{1+e^{-\beta'X}} \right) = \ln 1 - \ln(1+e^{-\beta'X}) = -\ln(1+e^{-\beta'X}) \quad (2.81)$$

Substituindo (2.80) e (2.81) em (2.79) temos:

$$L(\beta) = \sum_{i=1}^n [Y_i (\beta'X - \ln(1+e^{-\beta'X})) + (1-Y_i)(-\ln(1+e^{-\beta'X}))] \quad (2.82)$$

Assim:

$$L(\beta) = \sum_{i=1}^n [Y_i \beta'X - Y_i \ln(1+e^{-\beta'X}) - \ln(1+e^{-\beta'X}) + Y_i \ln(1+e^{-\beta'X})] \quad (2.83)$$

Concluindo:

$$L(\beta) = \sum_{i=1}^n [Y_i \beta' X_i - \ln(1 + e^{\beta' X_i})] \quad (2.84)$$

É possível demonstrar a concavidade global da função de log-verossimilhança sendo, portanto, apenas necessário anular o vetor das primeiras derivadas desta função em relação a β para obter os estimadores de máxima verossimilhança. Assim para determinar o valor de β que maximiza $L(\beta)$ deve-se derivar $L(\beta)$ em relação a β_0 e β_1 e igualar o conjunto de resultados a zero. Desta forma tem-se a seguir as chamadas equações de verossimilhança, que são derivadas da expressão anterior em relação a β .

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left(Y_i - \frac{e^{\beta' X_i}}{1 + e^{\beta' X_i}} \right) = \sum_{i=1}^n [Y_i - P(X_i)] = 0 \quad (2.85)$$

e

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left(X_i Y_i - X_i \frac{e^{\beta' X_i}}{1 + e^{\beta' X_i}} \right) = \sum_{i=1}^n X_i [Y_i - P(X_i)] = 0 \quad (2.86)$$

Temos que:

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n \frac{e^{\beta' X_i}}{1 + e^{\beta' X_i}} = \sum_{i=1}^n P(X_i) \quad (2.87)$$

O que reflete no fato da soma dos valores observados de Y ser igual a soma dos valores esperados (preditos), mediante a aplicação do modelo obtido. Esta propriedade é importante, pois ajuda avaliar o ajuste do modelo.

O valor de β dado pela solução das equações anteriores é chamado de estimador de máxima verossimilhança, é denotado por $\hat{\beta}$. $P(X_i)$, e representa uma estimativa da probabilidade condicional que Y é igual a 1, são que X é igual a X_i .

2.6.4 Modelo de Regressão Logístico Múltiplo

Considerando-se uma amostra de n pares de observações independentes (X_i, Y_i) $i = 1, 2, \dots, n$, da mesma forma como ocorreu no caso univariado, item anterior, o ajuste do modelo requer que se obtenha uma estimativa do vetor $\beta = [\beta_0, \beta_1, \dots, \beta_p]$. O método de estimação usado no caso multivariado será o mesmo no univariado, ou seja, o da máxima verossimilhança. A função de máxima verossimilhança é aproximadamente idêntica a dada na equação com uma mudança, sendo que $P(X)$ é definido por:

$$P(Y = 1 | X) = P(X_i) = \frac{e^{\beta X}}{1 + e^{\beta X}} \quad (2.88)$$

Estimar-se-á $(p+1)$ equações de verossimilhança as quais são obtidas por derivadas da função log-verossimilhança com respeito à $(p+1)$ parâmetros. As equações de verossimilhança resultantes podem ser representadas como:

$$\sum_{i=1}^n [Y_i - P(X_i)] = 0 \quad (2.89)$$

e

$$\sum_{i=1}^n X_{ij} [Y_i - P(X_i)] = 0 \quad (2.90)$$

para $j = 1, 2, \dots, p$

Da mesma forma como ocorreu com o modelo univariado, as soluções das equações de verossimilhança requer a utilização de um método iterativo na busca dos parâmetros estimados $\hat{\beta}$.

2.7 AVALIAÇÃO DA FUNÇÃO DE CLASSIFICAÇÃO

2.7.1 Critério TPM (*Total Probability of Misclassification*)

Uma das formas de avaliar o desempenho de um procedimento de reconhecimento de padrões consiste no cálculo da taxa de erro de classificação (*Total Probability of Misclassifications* - TPM) dado por:

$$TPM = p_1 \int_{R_2} f_1(\tilde{X}) d\tilde{X} + p_2 \int_{R_1} f_2(\tilde{X}) d\tilde{X} \quad (2.91)$$

onde p_1 e p_2 são as probabilidades de uma observação pertencente a Π_1 ou a Π_2 , respectivamente.

O valor mínimo dessa quantidade, obtido pela escolha adequada das regiões R_1 e R_2 , é chamado de taxa ótima de erro (*Optimum Error Rate* - OER). A OER é dada por:

$$OER = p_1 \int_{R_2} f_1(\tilde{X}) d\tilde{X} + p_2 \int_{R_1} f_2(\tilde{X}) d\tilde{X} \quad (2.92)$$

onde as regiões são determinadas por:

$$R_1 = \frac{f_1(\tilde{X})}{f_2(\tilde{X})} \geq \frac{p_2}{p_1} \text{ e } R_2 = \frac{f_1(\tilde{X})}{f_2(\tilde{X})} < \frac{p_2}{p_1} \quad (2.93)$$

Assim a OER é a taxa de erro mínima para a regra de classificação TPM. O desempenho da função de classificação da amostra pode, a princípio, ser avaliada pelo cálculo da taxa real de erro (*Actual Error Rate* - AER)

$$AER = p_1 \int_{\hat{R}_2} f_1(\tilde{X}) d\tilde{X} + p_2 \int_{\hat{R}_1} f_2(\tilde{X}) d\tilde{X} \quad (2.94)$$

onde: \hat{R}_1 e \hat{R}_2 representam as regiões de classificação determinadas pelos tamanhos das amostras n_1 e n_2 , respectivamente.

A AER indica o desempenho da função de classificação da amostra em amostras futuras. Da mesma forma que a taxa OER, não pode, em geral, ser calculada porque depende das f.d.p's $f_1(\tilde{X})$ e $f_2(\tilde{X})$ que são desconhecidas. Entretanto uma estimativa da quantidade ligada à taxa de erro real pode ser calculada, e sua estimativa será discutida na seqüência.

Há medidas de desempenho que não dependem das f.d.p's e podem ser calculadas por qualquer procedimento de classificação. Por exemplo, a medida, chamada de taxa de erro aparente (*Apparent Error Rate* - APER), é definida como a fração das observações usadas no treinamento das amostras que são incorretamente classificadas pela função de classificação amostral.

A APER pode ser facilmente calculada pela matriz de confusão, a qual mostra a classificação real versus a classificação prevista. Para n_1 observações de Π_1 e n_2 observações de Π_2 , a matriz de confusão tem a forma dada na tabela a seguir:

TABELA 2.3 – TABELA DE CLASSIFICAÇÃO

Classificação Real	Classificação Prevista	
	Π_1	Π_2
Π_1	$n_{1/1}$	$n_{1/2}$
Π_2	$n_{2/1}$	$n_{2/2}$

Fonte: APPLIED MULTIVARIATE STATISTICAL ANALYSIS

onde:

$n_{i/j}$ = números de observações de Π_i classificadas em Π_j e consideradas corretas se $i = j$ ou incorreta se $i \neq j$;

$n_{1/1}$ = numero de itens de Π_1 classificadas corretamente como item de Π_1 ;

$n_{1/2}$ = número de itens de Π_1 classificados incorretamente como item de Π_2 ;
 $n_{2/2}$ = número de itens de Π_2 classificados corretamente como item de Π_2 ;
 $n_{2/1}$ = número de itens de Π_2 classificados incorretamente como item de Π_1 .

A taxa aparente de erro (*Apparent Error Date* – APER) é dada por:

$$APER = \frac{n_{1/2} + n_{2/1}}{n_1 + n_2} \quad (2. 95)$$

Essa taxa deve ser interpretada como a proporção de observações incorretamente classificadas.

A APER é intuitivamente atraente e de cálculo fácil. Infelizmente, é tendenciosa para subestimar a AER, e o problema não desaparece a menos que os tamanhos das amostras n_1 e n_2 sejam grandes. Essencialmente, uma avaliação ótima ocorre porque os dados usados para construir a função de classificação são também usados para sua avaliação.

A taxa de erro estimada pode ser contida, pois são melhores que a taxa de erro aparente, mantendo-se relativa facilidade de cálculo, e não requer a função de distribuição de probabilidade. Um procedimento é dividir a amostra total em uma parte da amostra para treinamento e outra parte para validação. A parte da amostra para um treinamento é usada para construir a função de classificação, e a parte para validação é usada para sua avaliação.

A taxa de erro é determinada pela proporção de classificação errada dentro da amostra considerada. Embora este método supere o problema da tendenciosidade por não usar os mesmos dados para construir e julgar a função de classificação, ele sofre de dois defeitos principais:

- É necessário uma grande amostra;
- A função de avaliação não é a função de interesse. E ainda, quase todos os dados devem ser usados para a construção da função de classificação, pois o valor da informação pode ser perdido.

Uma segunda abordagem que funciona muito bem é o procedimento proposto por (LACHENBRUCH1975).

2.7.2 Abordagem de Lachenbruch

O método de Lachenbruch, segundo Johnson & Wichern (1998), é uma abordagem estatística empregada para avaliar o desempenho da regra de reconhecimento de padrão e estimar a probabilidade de classificação correta e incorreta em cada grupo, ou seja, uma forma de avaliar a eficiência da regra de classificação.

O algoritmo do referido procedimento é:

1. Inicie com as observações do grupo Π_1 , omita uma observação desse grupo, e construa uma função de classificação usando as $(n_1 - 1)$ e n_2 observações.
2. Classifique a observação retirada, usando a função construída no passo 1.
3. Repita o passo 1 e 2 até que todas as observações de Π_1 sejam classificadas.

Seja $n_{1/2}$ o número de observações reconhecidas erroneamente nesse grupo.

Repita os passos de 1 a 3 para as n_2 observações do grupo Π_2 .

Seja $n_{2/1}$ o número de observações reconhecidas erroneamente nesse grupo.

Assim as estimativas das probabilidades de classificação incorreta são:

$$P(2/1) = \int_{R_2} f_1(\tilde{X}) d\tilde{X} = \frac{n_{2/1}}{n_1} \quad (2.96)$$

classificar em Π_2 , quando a observação pertence a Π_1 ,

$$P(1/2) = \int_{R_1} f_2(\tilde{X}) d\tilde{X} = \frac{n_{1/2}}{n_2} \quad (2.97)$$

classificar em Π_1 , quando a observação pertence a Π_2 ,

Desta forma a proporção esperada de erro (probabilidade estimada) é dada por:

$$\hat{E}(AER) = \frac{n_{1/2} + n_{2/1}}{n_1 + n_2} \quad (2.98)$$

3 MATERIAL E MÉTODO

3.1 MATERIAL

3.1.1 Área de Estudo

Para a criação dos modelos de classificação, uma extensa campanha de medição dos transformadores foi necessária, para com isso realizar um estudo da Qualidade da Energia Elétrica (QEE) em todo o estado do Paraná, através das informações fornecidas pela Companhia de Energia Elétrica do Paraná (Copel).

A Copel é a maior empresa do estado do Paraná e tem 3,39 milhões de consumidores. O percentual de atendimento chega a praticamente 100% dos domicílios nas áreas urbanas e passa de 90% nas regiões rurais.

O universo de consumidores ligados inclui 2,6 milhões de lares, 57 mil indústrias, 283 mil estabelecimentos comerciais e 332 mil propriedades rurais. Em média, a Copel efetua mais de 70 mil novas ligações a cada ano. Algumas informações adicionais estão apresentadas nas Tabelas 3.1 a 3.4.

TABELA 3.1 – ÁREAS DE ATUAÇÃO DA COPEL

Indicadores	Valores
Áreas de Concessão (km ²)	194.854
Municípios Atendidos	393
Localidades Atendidas	1.111
População Atendida PR (Mil Habitantes)	9.822
Taxa de Atendimento Urbano (%)	99
Consumo Médio Residencial (kWh/mês)	159,6

FONTE: COPEL 2007

TABELA 3.2 – DADOS DE DISTRIBUIÇÃO

Dados de Distribuição	Valores
Número de Subestações 34,5Kv	238
Potência Instalada (MVA)	1.576
Redes de Distribuição (km)	167.535
Total de Postes	2.296.301
Número de Transformadores de Distribuição	315.587
Potência Instalada em Transformadores (MVA)	8.009

FONTE: COPEL 2007

TABELA 3.3 – NÚMERO DE CONSUMIDORES

Números de Consumidores (em milhares)	No mês		
	Jun/07	Jun/06	Crescimento(%)
Residencial	2.670	2.598	2,8
Industrial	58	55	5,4
Comercial	282	275	2,6
Rural	332	327	1,4
Outros	44	42	4,0
Total	3.386	3.297	2,7

FONTE: COPEL 2007

TABELA 3.4 – CONSUMO DE ENERGIA POR CLASSE

Consumo de Acumulado – GWh	No mês		
	Jun/07	Jun/06	Crescimento(%)
Residencial	2.557	2.392	6,9
Industrial	3.746	3.584	4,5
Comercial	1.874	1.712	9,5
Rural	779	742	5,0
Outros	923	917	0,7
Suprimento a Concessionária	234	224	4,2
Total	10.113	9.571	5,7

FONTE: COPEL 2007

Para fins de planejamento energético, o Paraná utiliza as cinco regiões correspondentes às áreas de atuação das Superintendências Regionais da Copel Distribuição, a citar:

- Superintendência de Distribuição Leste – SDL, com sede em Curitiba;
- Superintendência de Distribuição Oeste – SDO, com sede em Cascavel;
- Superintendência de Distribuição Centro-Sul – SDC, com sede em Ponta Grossa;
- Superintendência de Distribuição Noroeste – SDN, com sede em Maringá;
- Superintendência de Distribuição Norte – SDT, com sede em Londrina.

A Figura 3.1 apresenta o mapa do estado do Paraná dividido com suas respectivas Superintendências.

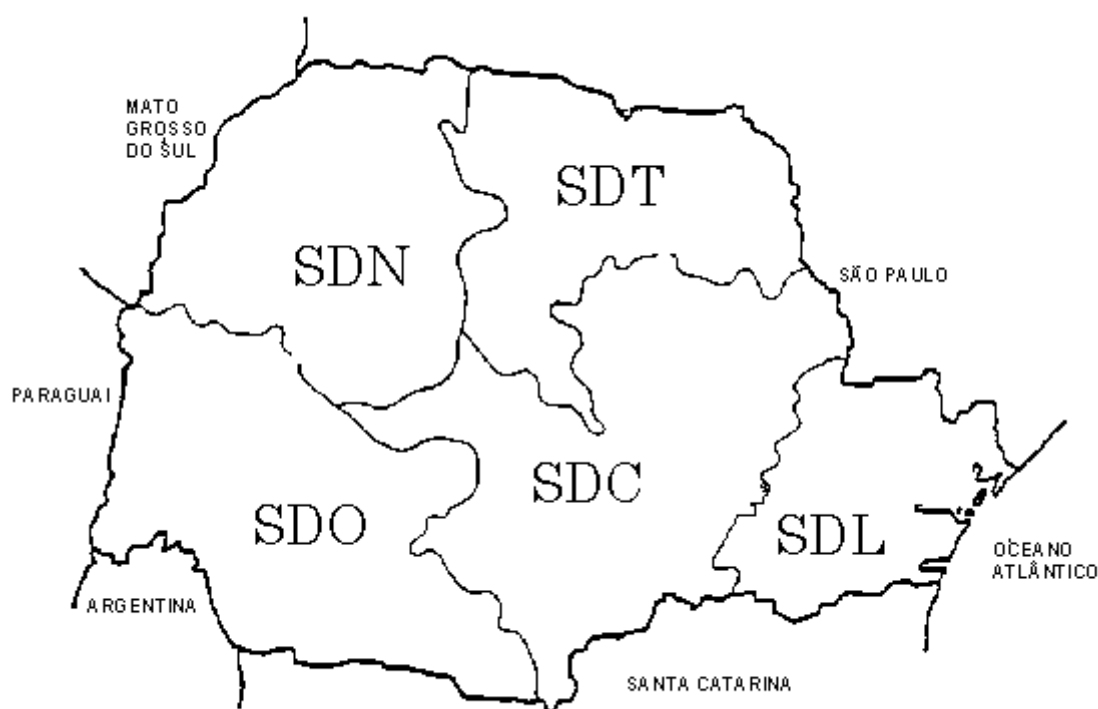


FIGURA 3.1 – REGIÕES DO PLANEJAMENTO ENERGÉTICO

FONTE: COPEL 2007

Em relação à participação no consumo final energético do Estado, a maior representatividade no consumo de energia está na região Leste (35,6%), seguida da região Centro Oeste (18,6%), da Noroeste (17,3%), da Norte (14,8%) e da região Oeste (13,7%).

3.1.2 Levantamento de Dados

Em função da grande quantidade de pontos de monitoração (transformadores) existentes no sistema de distribuição da COPEL, procurou-se, neste trabalho, adotar um critério estatístico que reduzisse a quantidade do tamanho de amostra sem comprometer as análises pertinentes ao trabalho supramencionado. Decidiu-se por sortear, de forma aleatória, os possíveis transformadores para monitoração.

Com o intuito, portanto, de determinar uma quantidade amostral adequada para o trabalho sob foco, consultou-se à portaria 163/93 do Departamento Nacional de Águas e Energia Elétrica (DNAEE), a qual apresenta uma formulação específica para este fim.

$$C_a = \frac{N.S.\bar{S}}{\frac{(N-1).E^2}{4} + (S.\bar{S})} \quad (3.1)$$

onde:

C_a = tamanho da amostra;

N = tamanho do universo considerado = 323.296;

S = proporção da população com características consideradas semelhantes = 0,5;

\bar{S} = proporção da população com características não semelhantes àquelas de $S = 0,5$;

E = erro amostral = 5%.

Diante do que foi exposto, após o processamento da equação 3.1, chegou-se a uma quantidade amostral de 399 transformadores, de um universo compreendido por 323.296 pontos possíveis de monitoração.

Todavia, prevendo-se a ocorrência de eventuais empecilhos nos trabalhos de campo associados a fatores climáticos, logística, registro e tratamento de dados, dentre outros aspectos intrínsecos a este tipo de atividade, acrescentou-se uma margem de segurança de 10% no computo final da quantidade amostral, de forma que, após as considerações anteriores, a quantidade amostral proposta foi de 440 transformadores.

A Tabela 3.5 apresenta a distribuição dos pontos por superintendências e a Tabela 3.6 apresenta a distribuição dos pontos entre transformadores rurais e urbanos.

TABELA 3. 5 – DISTRIBUIÇÃO DOS PONTOS DE MONITORAÇÃO, CONFORME A POTÊNCIA INSTALADA POR REGIONAL.

Regional	Localização	Região	Amostras (%)	Amostras
SDC	Centro	Ponta Grossa	11%	49
SDL	Leste	Curitiba	33%	146
SDN	Noroeste	Maringá	18%	77
SDO	Oeste	Cascavel	19%	84
SDT	Norte	Londrina	19%	84
Total			100%	440

FONTE: LACTEC 2007

TABELA 3.6 – DISTRIBUIÇÃO DAS AMOSTRAS ENTRE ÁREAS URBANAS E RURAIS.

Regional	Amostras Urbanas	Amostras Rurais
SDC	38	11
SDL	136	10
SDN	68	9
SDO	69	15
SDT	76	8
Total	387	53

FONTE: LACTEC 2007

Para o monitoramento em campo utilizou-se de 12 aparelhos específicos para medição. O monitoramento de todos os transformadores teve duração de aproximadamente 2 anos tendo início em agosto de 2005 e término em julho de 2007. O equipamento utilizado para a monitoração é descrito a seguir:

- **Equipamento: Registrador Marh 21**

- Fabricante: RMS Sistemas Eletrônicos – Brasil;
- 128 amostras por ciclo;
- Período de integralização de 10 minutos;
- Memória de 4Mb;
- Tensão alimentação: CA 70 a 600V, frequência: 50Hz ou 60Hz;
- CC 11 a 600V;
- Exatidão: Classe 0,5 para tensões e correntes e classe 1 para potências.

3.1.3 Variáveis Utilizadas

O estudo obedeceu aos requisitos do Submódulo 2.2 do NOS (2002), que estabelece como recomendação quanto aos índices sobre distorções harmônicas totais e individuais de tensão que o máximo permitido para a distorção harmônica total de tensão em qualquer fase seja o percentil 95% ou seja, ($P_{95\%}$). A variável dependente considerada no estudo foi o $P_{95\%}$ de distorção harmônica total acima de 6% e abaixo de 6%. As variáveis independentes ou explicativas foram separadas em características do transformador e características de consumo.

As variáveis correspondentes as características dos transformadores estão apresentadas na Tabela 3.7 e as variáveis correspondentes as características de consumo dos transformadores estão apresentadas na Tabela 3.8.

TABELA 3.7 – VARIÁVEIS ASSOCIADAS À CARACTERÍSTICA DO TRANSFORMADOR

Características do Transformador	Classes
Tensão Primária (Nível de Tensão do Alimentador)	<ul style="list-style-type: none"> • 13.800V • 34.500V
Fases (Quantidade de Fases no Transformador)	<ul style="list-style-type: none"> • 1 Fase • 2 Fases • 3 Fases
Potência (Potência Nominal do Transformador (kVA))	<ul style="list-style-type: none"> • 112,5 kVA • 75 kVA • 45 kVA • 30 kVA • 15 kVA • 10 kVA • 5 kVA

FONTE: COPEL 2007

TABELA 3.8 – VARIÁVEIS ASSOCIADAS À CARACTERÍSTICA DO CONSUMO.

Características do Consumo	
	Número de Consumidores
	Residencial 0/50 kWh
	Residencial 50/100 kWh
	Residencial 100/150 kWh
	Residencial 150/200 kWh
	Residencial 200/300 kWh
	Residencial 300/500 kWh
	Residencial 500/1000 kWh
	Residencial >1000 kWh
	Comércio e Poder Público
	Industrial
	Rural
	Iluminação Pública

FONTE: COPEL 2007

3.2 MÉTODO

A metodologia aqui descrita objetiva a realização da separação dos transformadores em dois grupos. Dentro das suas regionais, utilizando-se a regressão logística e a escore quadrático, que foram escolhidas por não fazerem restrições à igualdade entre as matrizes de covariância.

Os grupos foram separados com base no Submódulo 2.2 do NOS (2002), que estabelece como valor máximo permitido para distorção harmônica total de tensão de 6%, ou seja, os grupos foram separados tendo às seguintes características:

- Grupo 1 tendo como variável dependente o P95% da distorção harmônica total de tensão maior que 6% em qualquer uma das fases do transformador;
- Grupo 2 tendo como variável dependente o P95% da distorção harmônica total de tensão menor que 6% em todas as fases do transformador;

A monitoração para cada transformador foi realizada, dentro de um intervalo de tempo de aproximadamente oito dias. No entanto, para esse trabalho foi utilizado um período operativo de sete dias, obedecendo aos requisitos do Submódulo 2.2 da ONS e da Resolução nº. 505 da ANEEL, no tocante aos intervalos de análise. A Figura 3.2 ilustra um exemplo ao referido período de monitoração, o correspondente período de análise e o equipamento utilizado para a monitoração do referido transformador.

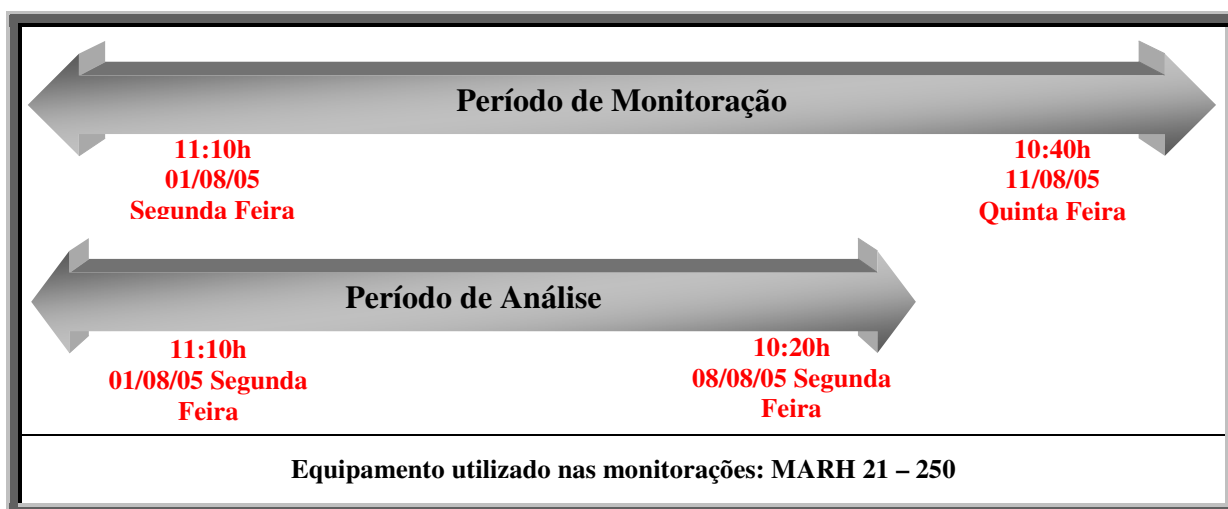


FIGURA 3.2 – PERÍODO DE MONITORAÇÃO VERSUS PERÍODO DE ANÁLISE

FONTE: LACTEC

A Figura 3.3 e Figura 3.4 indicam o comportamento dos perfis trifásicos das distorções harmônicas totais das tensões fase-neutro do barramento de 220/127V. O eixo vertical representa o valor percentual da distorção em relação ao correspondente valor fundamental.

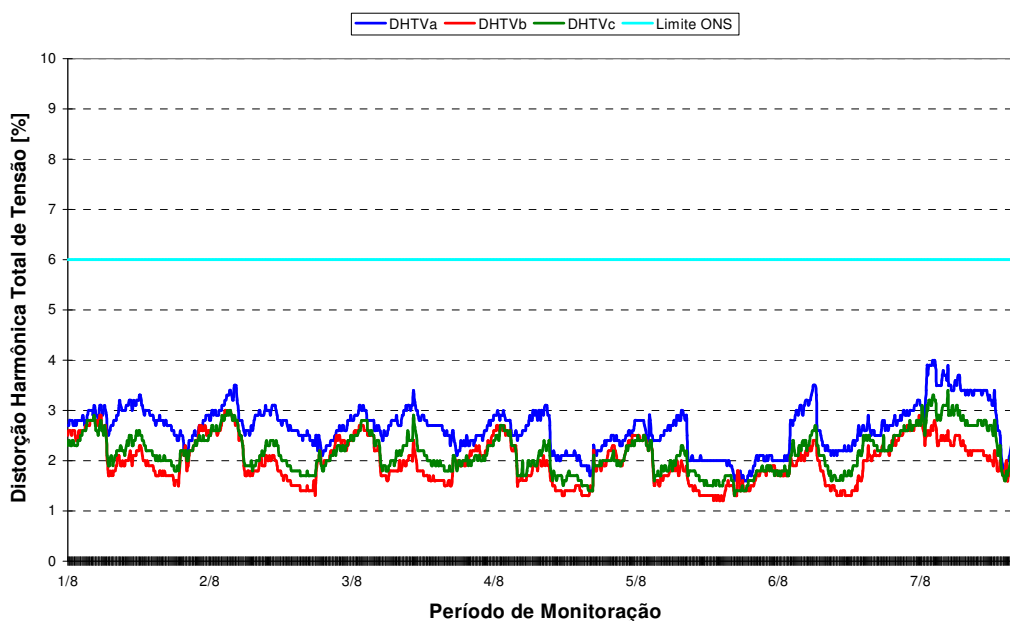


FIGURA 3.3 – PERFIL DAS DISTORÇÕES HARMÔNICAS TOTAIS DAS TENSÕES VALORES DENTRO DO LIMITE.

FUNTE: LACTEC

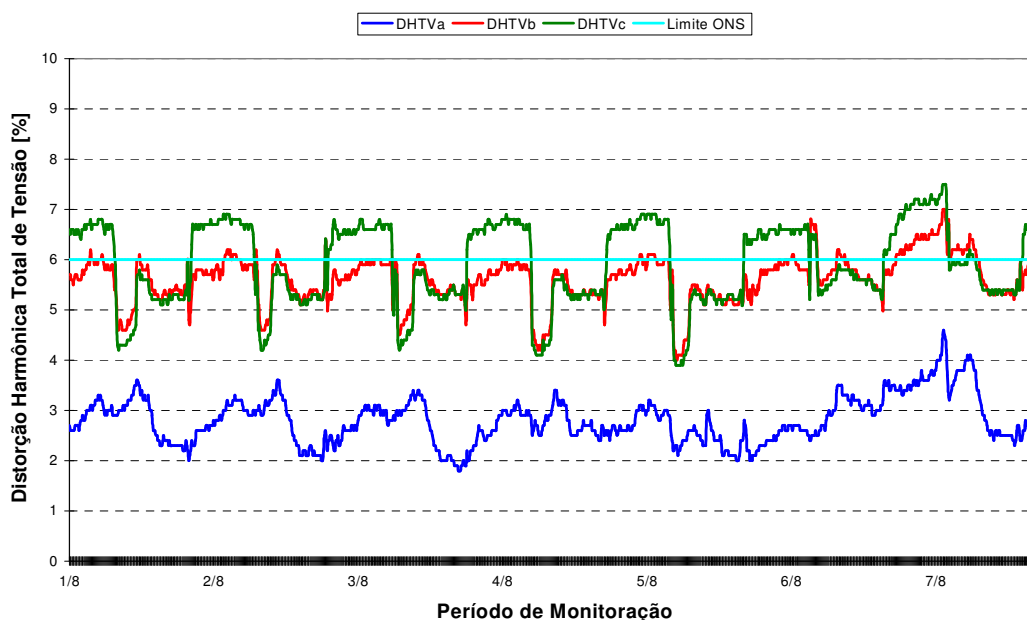


FIGURA 3.4 – PERFIL DAS DISTORÇÕES HARMÔNICAS TOTAIS DAS TENSÕES VALORES FORA DO LIMITE.

FUNTE: LACTEC

onde:

DHTV_A é os níveis de distorção harmônica total de tensão na fase A;

DHTV_B é os níveis de distorção harmônica total de tensão na fase B;

DHTV_C é os níveis de distorção harmônica total de tensão na fase C.

A partir da Figura 3.3 e figura 3.4 é possível extrair as informações sobre os níveis máximos e mínimos para as distorções totais encontradas para cada fase monitorada. A Tabela 3.9 e Tabela 3.10 fornecem uma síntese da monitoração aqui comentada.

TABELA 3.9 – VALORES MÁXIMOS, MÍNIMOS E P_{95%} DAS DISTORÇÕES TOTAIS DAS TENSÕES COM VALORES DENTRO DO LIMITE.

DHV _T	DHTV _A [%]	DHTV _B [%]	DHTV _C [%]
Máximo	4,70	4,90	4,60
Mínimo	2,30	2,20	2,40
P_{95%}	4,53	4,66	4,35

FONTE: LACTEC

TABELA 3.10 – VALORES MÁXIMOS, MÍNIMOS E P_{95%} DAS DISTORÇÕES TOTAIS DAS TENSÕES COM VALORES FORA DO LIMITE.

DHV _T	DHTV _A [%]	DHTV _B [%]	DHTV _C [%]
Máximo	4,60	7,00	7,50
Mínimo	1,80	4,00	3,90
P_{95%}	4,09	6,70	7,30

FONTE: LACTEC

A Figura 3.5 e Figura 3.6 ilustram o espectro harmônico da tensão, para o período de maior distorção harmônica registrada durante a monitoração.

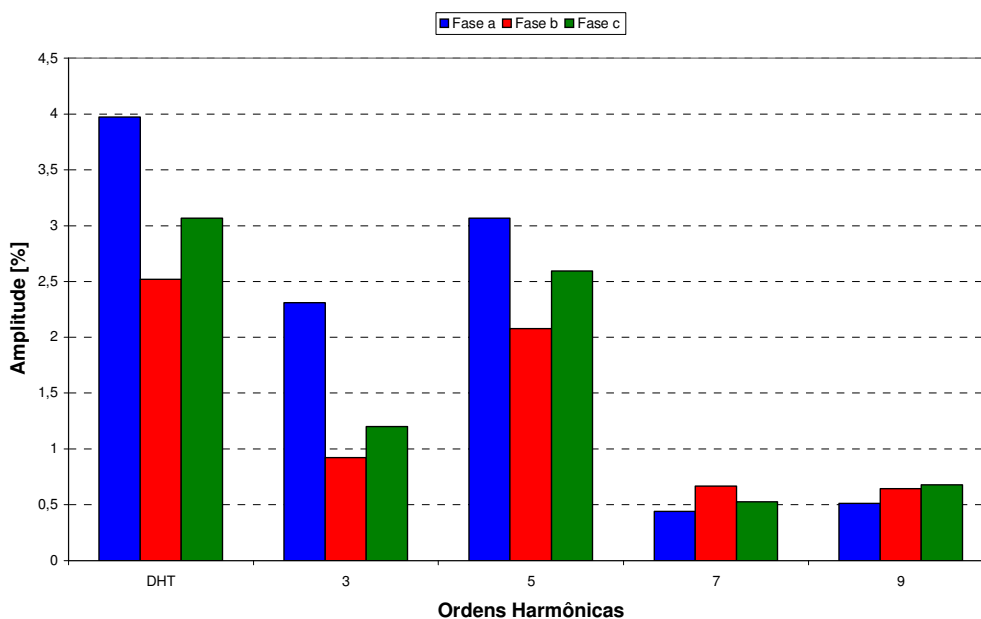


FIGURA 3.5 – ESPECTRO HARMÔNICO DA TENSÃO DE SUPRIMENTO NO PERÍODO DE MAIOR DISTRORÇÃO COM VALORES DENTRO DO LIMITE.

FONTE: LACTEC

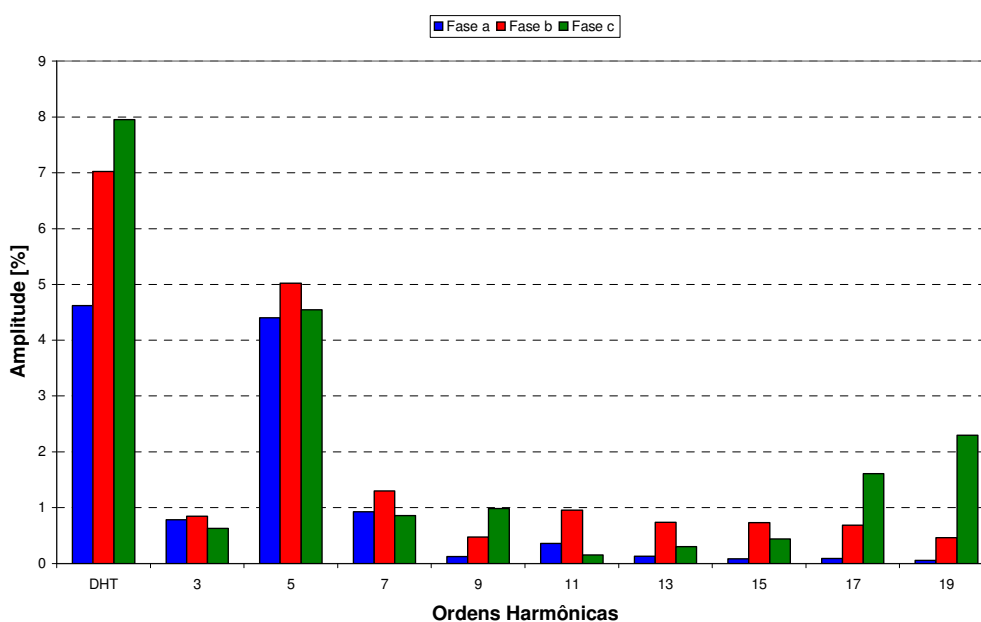


FIGURA 3.6 – ESPECTRO HARMÔNICO DA TENSÃO DE SUPRIMENTO NO PERÍODO DE MAIOR DISTRORÇÃO COM VALORES FORA DO LIMITE.

FONTE: LACTEC

Após todas as medições em campo levando em consideração os empecilhos dessa atividade, dos 387 transformadores da área urbana foram medidos 359 em

186 cidades em todo o Paraná. A distribuição dos transformadores medidos e das cidades está apresentada na Tabela 3.11.

TABELA 3.11 – DISTRIBUIÇÃO DOS TRANSFORMADORES MEDIDOS

Superintendência	SDL	SDT	SDO	SDC	SDN	TOTAL
Número de Cidades Visitadas	29	42	55	23	37	186
Quantidade de Transformadores Sorteados	128	74	62	32	63	359

FONTE: O AUTOR 2007

A etapa seguinte à obtenção dos dados sobre a distorção harmônica total de tensão foi o gerenciamento dos dados para a associação das variáveis dos transformadores e variáveis de consumo. Para o gerenciamento dos dados foi utilizado o *software Microsoft Excel 2000*. Houve a necessidade de criar uma planilha de dados para cada um dos métodos e cada uma das cinco regionais.

- Regressão Logística

Para a regressão logística foi associada às variáveis com as características dos transformadores e características do consumo de cada transformador como suas variáveis independentes e utilizando como variável resposta binária (1 se o transformador apresenta $P_{95\%}$ com distorção harmônica total acima de 6% em qualquer uma das fases e 0 se o transformador apresenta $P_{95\%}$ da distorção harmônica total abaixo de 6% para todas as fases do transformador).

- Escore Quadrático

Para o escore quadrático foi associado variáveis com as características dos transformadores e características de consumo de cada transformador, depois de testada a normalidade (Gaussianidade) dos dados, houve a necessidade de uma pré-classificação para organizar os grupos, levando-se em consideração o nível do $P_{95\%}$ da distorção harmônica total de tensão.

Após a montagem das planilhas previamente pré-classificadas utilizou-se o *software Matlab* para fazer a criação dos modelos e suas respectivas classificações.

3.2.1 Sistema Matlab

Em Matlab foram desenvolvidas duas funções de classificação, a primeira para a análise discriminante quadrática dos dois grupos e a segunda para realizar a regressão logística. A função logística foi denominada de “*logistic*” e a análise discriminante quadrática foi denominada de “*dis*”.

3.2.1.1 Algoritmo *logistic*

1. Entrada da matriz X com os dados dos dois grupos, com as variáveis por colunas e as observações em linha;
2. Entrada do vetor Y com a classificação dos grupos;
3. Chamada a função *logistic* (X,Y);
4. É estimada os parâmetros da função logística e alocado em um vetor C;
5. É feito o produto da matriz X pelo vetor com C;
6. O valor do produto é aplicado na função logística
7. Cálculo do valor de corte para a função logística;
8. Construção da tabela de classificação com respectivas porcentagens de acerto;

3.2.1.2 Algoritmo *dis*

1. Entrada das matrizes A e B com os dados dos dois grupos, as variáveis por coluna e as observações por linhas na seqüência;
2. Chamada a função *dis* (A,B);
3. É perguntado se o usuário deseja trabalhar com as probabilidades iguais para os grupos ou se deseja trabalhar com probabilidades proporcionais;
4. Cálculo das médias dos grupos;
5. Cálculo das matrizes de covariância dos grupos;

6. Cálculo dos coeficientes de segundo grau das funções;
7. Cálculo dos coeficientes de primeiro grau das funções;
8. Cálculo das constantes das funções;
9. Reclassificação pela análise discriminante quadrática e montagem da tabela de classificação com a respectiva porcentagem de acerto;

Após processado os dados e criada a regra de classificação para as superintendências utilizando as duas metodologias, foi criado um *software* em linguagem *Matlab* denominado “*Distorcoes*” que permite a classificação de novos transformadores aplicando qualquer um dos dois métodos.

3.2.2 Software Distorcoes

O *software* apresenta inicialmente duas opções para o usuário, podendo classificar um novo transformador ou inserir os dados de um novo transformador no banco de dados do *software*, (Figura 3.7).

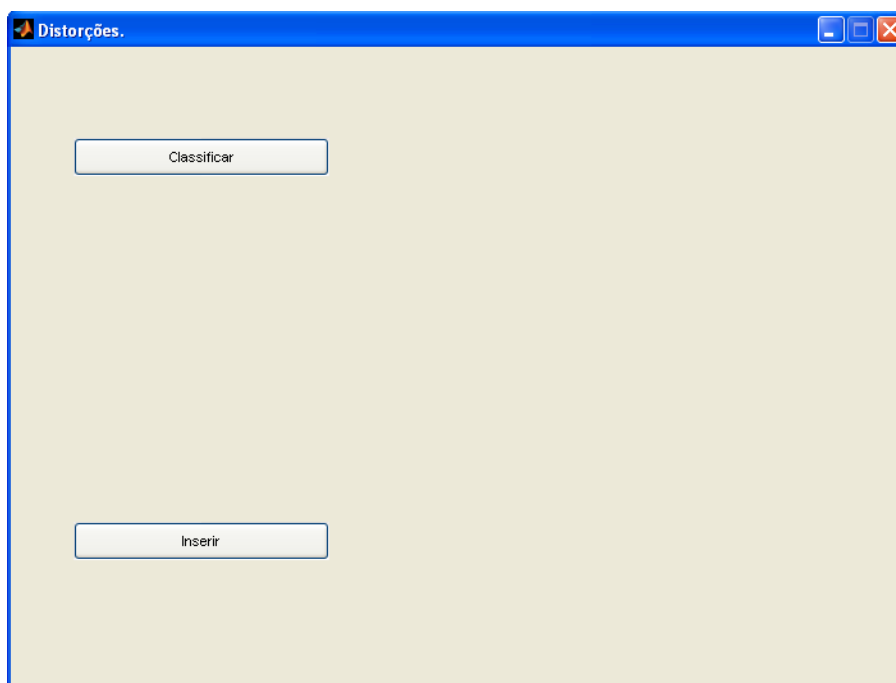


FIGURA 3.7 – TELA INICIAL DO SOFTWARE DISTORÇÕES

FONTE: O AUTOR 2007

Caso o usuário selecione a opção de “classificação” ele terá duas opções de fazer a classificação podendo optar por fazer a classificação utilizando regressão logística ou escore quadrático (Figura 3.8). Na opção de classificação por escore quadrático é utilizado apenas as probabilidades iguais, já que nos resultados que serão apresentados utilizando as probabilidades a priori não se obteve diferença significativa na classificação.

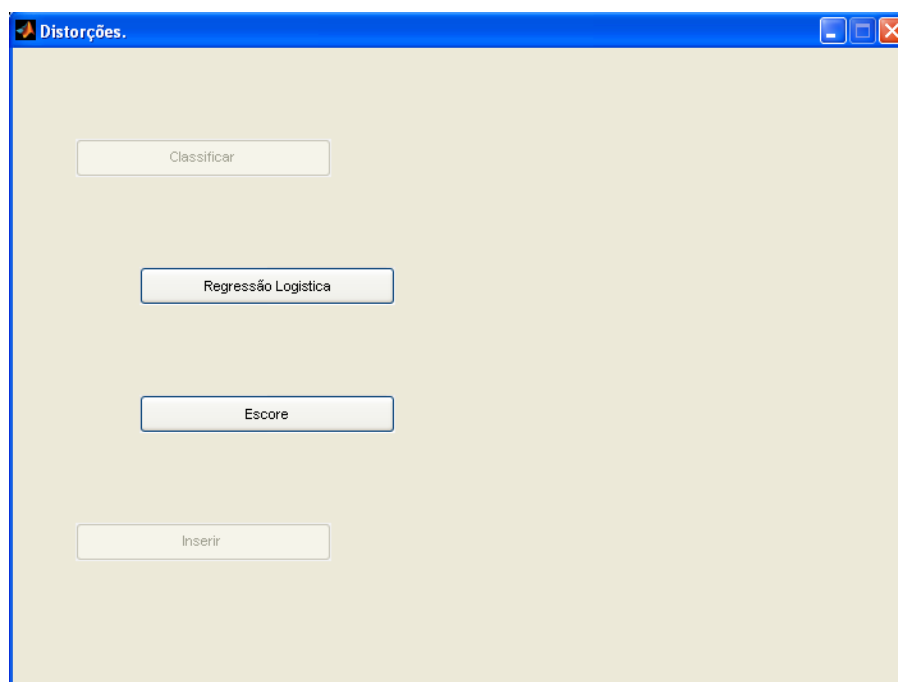


FIGURA 3.8 – TELA DE CLASSIFICAÇÃO

FONTE: O AUTOR 2007

Depois de selecionado o método o *software* apresenta cinco opções em cada uma com suas respectivas regionais (Figura 3.9 e Figura 3.10).

- Superintendência Leste (SDL);
- Superintendência Oeste (SDO);
- Superintendência Norte (SDT);
- Superintendência Noroeste (SDN);
- Superintendência Centro Sul (SDC).

Para cada uma das opções escolhida pelo o usuário os seus respectivos dados são importados de um banco.

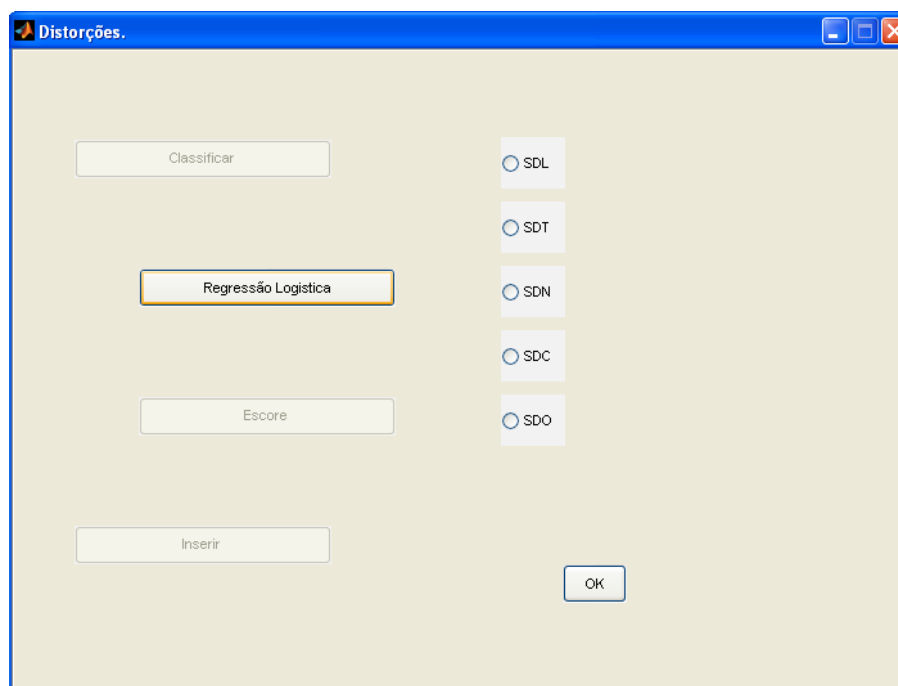


FIGURA 3.9 – TELA DE CLASSIFICAÇÃO REGRESSÃO LOGÍSTICA

FONTE: O AUTOR 2007

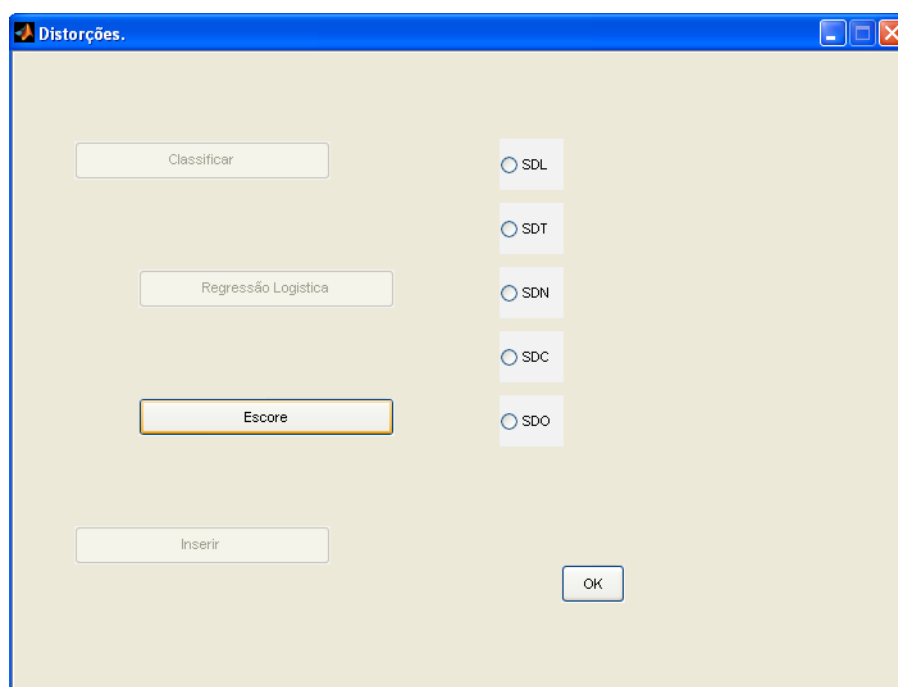


FIGURA 3.10 – TELA DE CLASSIFICAÇÃO ESCORE QUADRÁTICO

FONTE O AUTOR 2007

Assim que o usuário seleciona a regional em que deseja classificar o transformador, é necessária a inserção dos dados do transformador o qual ele

deseja classificar (Figura 3.11), assim que forem inseridos os dados sobre o transformador, (Figura 3.12) o *software* cria o modelo de classificação e classifica o novo transformador (Figura 3.13).

FIGURA 3.11 – TELA DE INSERÇÃO DE DADOS

FONTE: O AUTOR 2007

FIGURA 3.12 – TELA COM DADOS INSERIDOS

FONTE: O AUTOR 2007

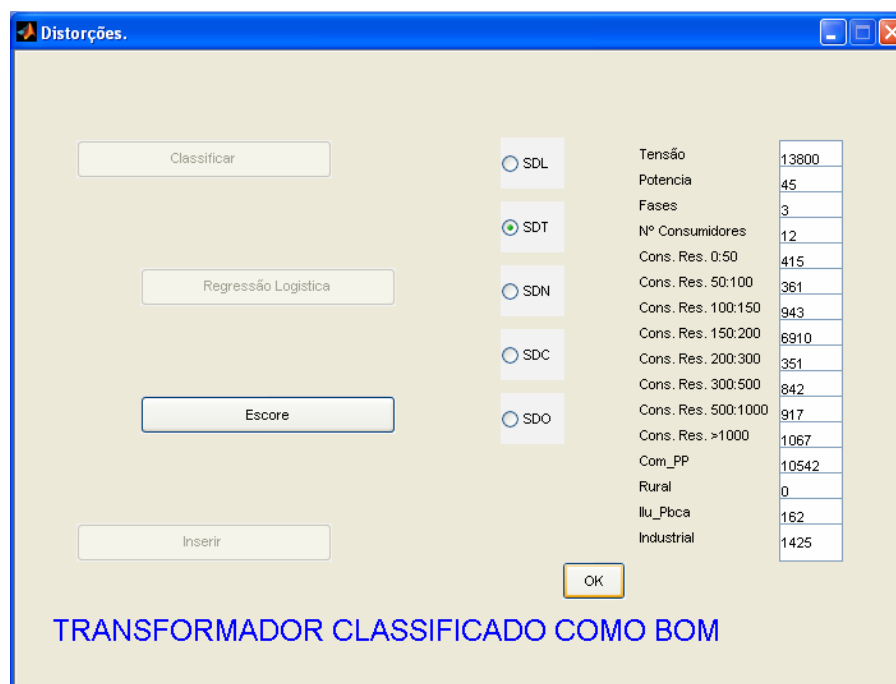


FIGURA 3.13 – TELA DE RESULTADO DO *SOFTWARE* DISTORÇÕES

FONTE: O AUTOR 2007

A base dos dados do programa é processada a cada classificação, assim a partir do momento em que mais transformadores forem medidos poder-se-á acrescentar os dados desses novos transformadores a base de dados para que, assim, possa-se ter uma melhor base para os modelos de classificação.

4 RESULTADOS E DISCUSSÕES

Os resultados obtidos através do programa desenvolvido estão apresentados nos quadros a seguir, que representam as matrizes com a classificação original e a prevista pelo método, com as respectivas taxas de erro.

Os quadros 4.1, 4.2, 4.3, 4.4 e 4.5 apresentam os resultados alcançados pela regressão logística. Já nos quadros 4.6, 4.7, 4.8, 4.9 e 4.10 são apresentados os resultados alcançados pelo escore quadrático considerando-se as probabilidades iguais. E nos quadros 4.11, 4.12, 4.13, 4.14 e 4.15 são apresentados os resultados alcançados pelo escore quadrático considerando-se as probabilidades proporcionais ao tamanho dos grupos.

Na Tabela 4.1 são apresentados as variáveis utilizadas nos modelos de reconhecimento e classificação de padrão.

TABELA 4.1 – TABELA DAS VARIÁVEIS

VARIÁVEIS	
Constante	
Tensão	x_1
Potencia	x_2
Fases	x_3
Número de Consumidores	x_4
Cons. 0/50 kWh	x_5
Cons. 50/100	x_6
Cons. 100/150	x_7
Cons. 150/200	x_8
Cons. 200/300	x_9
Cons. 300/500	x_{10}
Cons. 500/1000	x_{11}
Cons. >1000	x_{12}
Comércio e Poder Público	x_{13}
Rural	x_{14}
Iluminação Pública	x_{15}
Industrial	x_{16}

QUADRO 4.1 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL LESTE (SDL) UTILIZANDO REGRESSÃO LOGÍSTICA

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	96 * 93,20%	7 6,80%
	Problemático	4 16,00%	21 * 84,00%
APER		8,60%	

* Percentual de acerto

Seja o modelo de classificação $P(Y) = \frac{1}{1+e^{-Y}}$ onde,

$$Y = -13,63640 + 0,00000x_1 - 0,04320x_2 + 5,12541x_3 - 0,05550x_4 + 0,00344x_5 + 0,00220x_6 + 0,00003x_7 - 0,00007x_8 - 0,00127x_9 + 0,00088x_{10} + 0,00096x_{11} + (4.1) - 0,00152x_{12} - 0,00136x_{13} + 0,01764x_{14} - 0,03603x_{15} + 0,00026x_{16}$$

QUADRO 4.2 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NORTE (SDT) UTILIZANDO REGRESSÃO LOGÍSTICA

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	50 *86,21%	8 13,79%
	Problemático	6 37,50%	10 *62,50%
APER		18,92%	

* Percentual de acerto

Seja o modelo de classificação $P(Y) = \frac{1}{1+e^{-Y}}$ onde,

$$Y = 0,00000 + 0,00313x_1 + 0,03538x_2 + 13,34650x_3 - 0,04471x_4 + 0,00646x_5 + 0,00127x_6 - 0,00091x_7 + 0,00113x_8 + 0,00025x_9 - 0,00172x_{10} + 0,00074x_{11} + (4.2) - 0,01317x_{12} + 0,00008x_{13} + 0,00282x_{14} - 0,01335x_{15} + 0,00001x_{16}$$

QUADRO 4.3 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL CENTRO SUL (SDC) UTILIZANDO REGRESSÃO LOGÍSTICA

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	21 *91,30%	2 8,70%
	Problemático	0 0%	9 *100%
APER		5,88%	

* Percentual de acerto

Seja o modelo de classificação $P(Y) = \frac{1}{1+e^{-Y}}$ onde,

$$\begin{aligned}
 Y = & 0,00000 + 0,00124 x_1 - 0,13200 x_2 + 0,00000 x_3 - 6,47246 x_4 - 0,29007 x_5 + \\
 & + 0,08909 x_6 + 0,09282 x_7 - 0,00747 x_8 + 0,00794 x_9 + 0,07329 x_{10} - 0,04911 x_{11} + \\
 & + 0,00000 x_{12} + 0,01299 x_{13} + 0,00150 x_{14} - 0,09607 x_{15} - 0,00302 x_{16}
 \end{aligned} \quad (4.3)$$

QUADRO 4.4 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NOROESTE (SDN) UTILIZANDO REGRESSÃO LOGÍSTICA

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	35 *76,09%	11 23,91%
	Problemático	5 29,41%	12 *70,59%
APER		25,39%	

* Percentual de acerto

Seja o modelo de classificação $P(Y) = \frac{1}{1+e^{-Y}}$ onde,

$$\begin{aligned}
 Y = & 0,00000 + 0,00329 x_1 - 0,03433 x_2 + 15,12360 x_3 + 0,09979 x_4 + 0,00278 x_5 + \\
 & 0,00022 x_6 - 0,00205 x_7 - 0,00083 x_8 - 0,00031 x_9 - 0,00011 x_{10} - 0,00007 x_{11} + \\
 & - 0,01305 x_{12} + 0,00013 x_{13} + 0,00013 x_{14} + 0,00035 x_{15} + 0,00053 x_{16}
 \end{aligned} \quad (4.4)$$

QUADRO 4.5 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL OESTE (SDO) UTILIZANDO REGRESSÃO LOGÍSTICA

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	44 *86,27%	7 13,73%
	Problemático	3 27,27%	8 *72,73%
APER		19,23%	

* Percentual de acerto

Seja o modelo de classificação $P(Y) = \frac{1}{1 + e^{-Y}}$ onde,

$$\begin{aligned}
 Y = & -7,42424 + 0,00014 x_1 - 0,05363 x_2 + 1,76838 x_3 - 0,31732 x_4 + 0,05329 x_5 + \\
 & -0,00405 x_6 + 0,00738 x_7 - 0,00224 x_8 - 0,00192 x_9 + 0,00367 x_{10} - 0,01919 x_{11} + \\
 & -0,00350 x_{12} + 0,00070 x_{13} - 0,00022 x_{14} + 0,00286 x_{15} + 0,00117 x_{16}
 \end{aligned} \quad (4.5)$$

Com a aplicação da técnica de Regressão Logística para o reconhecimento e classificação dos transformadores, obteve-se um resultado satisfatório para todas as superintendências regionais analisadas. Como melhor resultado teve a classificação da superintendência de distribuição centro sul (SDC) com uma taxa de erro aparente de 5,88% e como pior resultado a superintendência de distribuição noroeste (SDN) com uma taxa de erro aparente de 25,40%.

A taxa de erro aparente médio para todas as superintendências, utilizando a Regressão Logística ficou na faixa de 14,64%.

Para esses resultados utilizou-se o valor de 0,35 como o valor de corte para a função logística, pois é aproximadamente a proporção dos grupos de transformadores e também o valor que maximizou a classificação dos grupos.

QUADRO 4.6 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL LESTE (SDL) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES IGUAIS

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	91	12
		*88,35%	11,65%
	Problemático	0	25
		0%	*100%
APER		9,75%	

* Percentual de acerto

As funções de classificação são:

$$\begin{aligned}
 D_1 = & 0,0006 x_{16}x_2 - 0,0040 x_{16}x_3 + 0,0008 x_{16}x_4 - 0,0001 x_{16}x_5 - 0,2109 x_2 + 169,3155 x_3 \\
 & + 0,1595 x_4 - 0,0097 x_5 + 0,0017 x_6 - 0,0060 x_7 - 0,0008 x_8 + 0,0011 x_9 - 0,0019 x_{10} - (4.6) \\
 & 0,0002 x_{11} - 0,0010 x_{12} - 0,0001 x_{13} + 0,0001 x_{16} - 314,6296
 \end{aligned}$$

$$\begin{aligned}
 D_2 = & 0,0006 x_{16}x_2 - 0,0040 x_{16}x_3 + 0,0008 x_{16}x_4 - 0,0001 x_{16}x_5 - 0,1699 x_2 + 33,8196 x_3 \\
 & + -0,3496 x_4 + 0,0453 x_5 - 0,0048 x_6 + 0,0040 x_7 + 0,0032 x_8 - 0,0031 x_9 + 0,0003 x_{10} + (4.7) \\
 & 0,0040 x_{11} - 0,0066 x_{12} - 0,0041 x_{13} + 0,0006 x_{16} - 105,5773
 \end{aligned}$$

QUADRO 4.7 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NORTE (SDT) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES IGUAIS.

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	54	4
		*93,10%	6,90%
	Problemático	1	15
		6,25%	*93,75%
APER		6,76%	

* Percentual de acerto

As funções de classificação são:

$$D_1 = 0,1147 x_2 - 0,0651 x_4 + 0,0051 x_5 + 0,0016 x_6 + 0,0010 x_7 + 0,0001 x_8 + \\ -0,0001 x_9 + 0,0002 x_{10} + 0,0013 x_{13} + 0,0004 x_{16} - 67,9850 \quad (4.8)$$

$$D_2 = 1,0448 x_2 - 1,2069 x_4 + 0,0327 x_5 + 0,0388 x_6 - 0,0017 x_7 - 0,0121 x_8 + 0,0650 x_9 \\ - 0,0334 x_{10} - 0,0003 x_{13} + 0,0409 x_{15} + 0,0081 x_{16} - 109,4385 \quad (4.9)$$

QUADRO 4.8 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL CENTRO SUL (SDC) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES IGUAIS

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	17	6
	Problemático	0	8
		*79,91%	20,09%
APER		19,35%	0%
			*100%

* Percentual de acerto

As funções de classificação são:

$$D_1 = 0,0015 x_1 + 0,1720 x_2 + 0,0337 x_4 - 0,0367 x_5 + 0,0033 x_6 + 0,0035 x_7 + \\ 0,0015 x_8 - 0,0005 x_9 - 0,0021 x_{10} + 0,0057 x_{11} - 0,0005 x_{13} \quad (4.10)$$

$$D_2 = -0,0132 x_1 + 1,2867 x_2 - 5,6903 x_4 + 3,0624 x_5 + 0,1777 x_6 - 0,5276 x_7 + \\ 1,0777 x_8 - 0,7180 x_9 - 0,0423 x_{10} - 0,2002 x_{11} + 0,1727 x_{13} - 1,9310 \quad (4.11)$$

QUADRO 4.9 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NOROESTE (SDN) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES IGUAIS

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	36 *78,26%	10 21,73
	Problemático	0 0%	17 *100%
APER		15,87%	

* Percentual de acerto

As funções de classificação são:

$$D_1 = 0,0982 x_2 + 0,0110 x_4 + 0,0140 x_5 + 0,0008 x_6 - 0,0012 x_7 + 0,0009 x_8 - 0,0006 x_9 + 0,0007 x_{10} - 0,0006 x_{11} - 0,0001 x_{13} + 0,0004 x_{14} + 0,0020 x_{15} - 0,0002 x_{16} - 81,22 \quad (4.12)$$

$$D_2 = 0,1725 x_2 - 0,1208 x_4 + 0,0220 x_5 + 0,0068 x_6 - 0,0031 x_7 - 0,0014 x_8 - 0,0024 x_9 + 0,0058 x_{10} + 0,0154 x_{11} - 0,0004 x_{13} + 0,0040 x_{14} + 0,0003 x_{15} - 0,0001 x_{16} - 76,3201 \quad (4.13)$$

QUADRO 4.10 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL OESTE (SDO) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES IGUAIS

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	21 *42%	30 58%
	Problemático	0 0%	11 *100%
APER		48,39%	

* Percentual de acerto

As funções de classificação são:

$$D_1 = 0,0011 x_1 - 0,1648 x_2 + 18,2044 x_3 + 0,1193 x_4 + 0,0041 x_5 - 0,0034 x_6 + 0,0029 x_7 - 0,0039 x_8 - 0,0011 x_9 + 0,0018 x_{10} - 0,0026 x_{14} + 0,0018 x_{15} + 0,0011 x_{16} \quad (4.14)$$

$$D_2 = -0,0033 x_1 - 1,0572 x_2 - 5,5156 x_3 - 1,6110 x_4 - 0,8546 x_5 + 0,3387 x_6 - 0,0878 x_7 + 0,0631 x_8 - 0,0117 x_9 + 0,0344 x_{10} + 0,0111 x_{13} + 0,0639 x_{14} - 0,0500 x_{15} + 0,0859 x_{16} + 3,0417 \quad (4.15)$$

Com a aplicação da técnica de Escore quadrático com probabilidades iguais para o reconhecimento e classificação dos transformadores, obteve-se um resultado satisfatório para quatro superintendências regionais. A Superintendência de Distribuição Oeste (SDO) não se obteve um resultado satisfatório, isso se deve ao mau condicionamento da matriz de dados dos grupos dessa superintendência, pois a matriz de dados foi considerada uma matriz esparsa, ou seja, uma matriz com grande quantidade de valores nulos, com isso a metodologia aplicada não teve desempenho satisfatório para essa superintendência.

Como melhor resultado de classificação temos a Superintendência de Distribuição Norte (SDT) com uma taxa de erro aparente de 6,75% e como pior resultado a Superintendência de Distribuição Oeste (SDO) com uma taxa de erro aparente de 48,39%.

A taxa de erro aparente médio para todas as superintendências, utilizando a Escore Quadrático com probabilidades iguais ficou na faixa de 17,60%.

QUADRO 4.11 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL LESTE (SDL) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES PROPORCIONAIS

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	92	11
	Problemático	0	25
		*89,32%	10,68%
		0%	*100%
APER		8,59%	

* Percentual de acerto

As funções de classificação são:

$$D_1 = -0,2109 x_2 + 169,3155 x_3 + 0,1595 x_4 - 0,0097 x_5 + 0,0017 x_6 - 0,0060 x_7 - 0,0008 x_8 + 0,0011 x_9 - 0,0019 x_{10} - 0,0002 x_{11} - 0,0010 x_{12} - 0,0001 x_{13} + 0,0001 x_{16} - 314,1538 \quad (4.16)$$

$$D_2 = -0,1699 x_2 + 33,8196 x_3 + -0,3496 x_4 + 0,0453 x_5 - 0,0048 x_6 + 0,0040 x_7 + 0,0032 x_8 - 0,0031 x_9 + 0,0003 x_{10} + 0,0040 x_{11} - 0,0066 x_{12} - 0,0041 x_{13} + 0,0006 x_{16} - 106,5173 \quad (4.17)$$

QUADRO 4.12 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NORTE (SDT) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES PROPORCIONAIS

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	55 *94,82%	3 5,18%
	Problemático	2 12,5%	14 *87,5%
APER		6,76%	

* Percentual de acerto

As funções de classificação são:

$$D_1 = 0,1147 x_2 - 0,0651 x_4 + 0,0051 x_5 + 0,0016 x_6 + 0,0010 x_7 + 0,0001 x_8 - 0,0001 x_9 + 0,0002 x_{10} + 0,0013 x_{13} + 0,0004 x_{16} - 67,5355 \quad (4.18)$$

$$D_2 = 1,0448 x_2 - 1,2069 x_4 + 0,0327 x_5 + 0,0388 x_6 - 0,0017 x_7 - 0,0121 x_8 + 0,0650 x_9 - 0,0334 x_{10} - 0,0003 x_{13} + 0,0409 x_{15} + 0,0081 x_{16} - 110,2768 \quad (4.19)$$

QUADRO 4.13 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL CENTRO SUL (SDC) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES PROPORCIONAIS

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	17 *79,91%	6 20,09%
	Problemático	0 0%	8 *100%
APER		19,35%	

* Percentual de acerto

As funções de classificação são:

$$D_1 = 0,0015x_1 + 0,1720x_2 + 0,0337x_4 - 0,0367x_5 + 0,0033x_6 + 0,0035x_7 + 0,0015x_8 - 0,0005x_9 - 0,0021x_{10} + 0,0057x_{11} - 0,0005x_{13} \quad (4.20)$$

$$D_2 = -0,0132x_1 + 1,2867x_2 - 5,6903x_4 + 3,0624x_5 + 0,1777x_6 - 0,5276x_7 + 1,0777x_8 - 0,7180x_9 - 0,0423x_{10} - 0,2002x_{11} + 0,1727x_{13} - 1,9310 \quad (4.21)$$

QUADRO 4.14 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL NOROESTE (SDN) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES PROPORCIONAIS.

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	37 *80,43%	9 19,57%
	Problemático	0 0%	17 *100%
APER		14,29%	

* Percentual de acerto

As funções de classificação são:

$$D_1 = 0,0982 x_2 + 0,0110 x_4 + 0,0140 x_5 + 0,0008 x_6 - 0,0012 x_7 + 0,0009 x_8 - 0,0006 x_9 + 0,0007 x_{10} - 0,0006 x_{11} - 0,0001 x_{13} + 0,0004 x_{14} + 0,0020 x_{15} - 0,0002 x_{16} - 80,8414 \quad (4.22)$$

$$D_2 = 0,1725 x_2 - 0,1208 x_4 + 0,0220 x_5 + 0,0068 x_6 - 0,0031 x_7 - 0,0014 x_8 - 0,0024 x_9 + 0,0058 x_{10} + 0,0154 x_{11} - 0,0004 x_{13} + 0,0040 x_{14} + 0,0003 x_{15} - 0,0001 x_{16} - 76,9369 \quad (4.23)$$

QUADRO 4.15 – MATRIZ DE CLASSIFICAÇÃO E APER DA SUPERINTENDÊNCIA REGIONAL OESTE (SDO) UTILIZANDO ESCORE QUADRÁTICO COM PROBABILIDADES PROPORCIONAIS

		Classificação Prevista	
		Bom	Problemático
Classificação Real	Bom	21 *42%	30 58%
	Problemático	0 0%	11 *100%
APER		48,39%	

*Percentual de acerto

As funções de classificação são:

$$D_1 = 0,0011 x_1 - 0,1648 x_2 + 18,2044 x_3 + 0,1193 x_4 + 0,0041 x_5 - 0,0034 x_6 + 0,0029 x_7 - 0,0039 x_8 - 0,0011 x_9 + 0,0018 x_{10} - 0,0026 x_{14} + 0,0018 x_{15} + 0,0011 x_{16} \quad (4.24)$$

$$D_2 = -0,0033 x_1 - 1,0572 x_2 - 5,5156 x_3 - 1,6110 x_4 - 0,8546 x_5 + 0,3387 x_6 - 0,0878 x_7 + 0,0631 x_8 - 0,0117 x_9 + 0,0344 x_{10} + 0,0111 x_{13} + 0,0639 x_{14} - 0,0500 x_{15} + 0,0859 x_{16} + 3,0417 \quad (4.25)$$

Como ocorrido com o Escore Quadrático com probabilidades iguais o Escore Quadrático com probabilidade a priori obteve-se um resultado satisfatório no reconhecimento e classificação de quatro superintendências regionais. Como a matriz de dados para a variação dos métodos é a mesma o mesmo problema ocorrido coma matriz de dados da Superintendência de Distribuição Oeste (SDO) o Escore Quadrático com probabilidades a priori não obteve um resultado satisfatório.

Como melhor resultado de classificação temos a Superintendência de Distribuição Norte (SDT) com uma taxa de erro aparente de 6,75% e como pior resultado a Superintendência de Distribuição Oeste (SDO) com uma taxa de erro aparente de 48,39%.

A taxa de erro aparente médio para todas as superintendências, utilizando a Escore Quadrático com probabilidades iguais ficou na faixa de 17,04%.

TABELA 4.2 – TABELA COM APER DE TODAS AS REGIONAIS

Regional	Regressão Logística	Escore Quadrático com Probabilidades Iguais	Escore Quadrático com Probabilidades a Priori
SDL	8,40%	9,16%	8,40%
SDT	18,90%	6,75%	6,75%
SDC	6,25%	18,75%	18,75%
SDO	16,13%	48,39%	48,39%
SDN	25,40%	15,87%	14,28%
Média	14,64%	17,60%	17,04%

Com a Tabela 4.2 podemos observar que:

- A Regressão Logística apresentou melhor desempenho médio na classificação dos transformadores;
- O Escore Quadrático com probabilidades a priori não teve um ganho de desempenho significativo em relação ao Escore Quadrático com probabilidades iguais;
- A Superintendência de Distribuição Oeste (SDO) teve grande influência para o desempenho médio de classificação para o Escore Quadrático tanto com probabilidades a priori como com probabilidades iguais.

Para cada Superintendência Regional podemos observar:

- **SDL:** Para Superintendência Regional Leste obteve um melhor desempenho utilizando a Regressão Logística e com Escore Quadrático levando em consideração a probabilidade a priori;

- SDT: Para Superintendência Regional Norte obteve um melhor desempenho utilizando o Escore Quadrático levando em consideração a probabilidade igual e a probabilidade a priori;
- SDC: Para Superintendência Regional Centro Sul obteve um melhor desempenho utilizando a Regressão Logística;
- SDO: Para Superintendência Regional Oeste obteve um melhor desempenho utilizando a regressão logística;
- SDN: Para Superintendência Regional Noroeste obteve um melhor desempenho utilizando a escore quadrático com as probabilidades a priori.

5 CONCLUSÕES E DESENVOLVIMENTOS FUTUROS

5.1 CONCLUSÕES

A crescente a demanda por melhoria da qualidade dos serviços de energia elétrica, com os consumidores exigindo uma pronta atuação do órgão regulador, em benefício da sociedade. A qualidade deve ser objetivo de todos os envolvidos com a produção e comercialização de um bem ou serviço, bem como a preocupação com a satisfação do cliente, que por sua vez é referenciado pelo cumprimento das normas técnicas indicadas para o produto ou serviço prestado.

Neste trabalho buscou-se apresentar uma visão geral da questão da qualidade de energia elétrica (QEE) bem como seus principais distúrbios, sua definições, características e as principais normas técnicas sobre limites para os distúrbios da QEE. Foram apresentados também métodos estatísticos aplicados no reconhecimento e classificação de padrões bem como suas principais características e aplicabilidade.

Com o exposto nesse trabalho percebe-se a importância do desenvolvimento de ferramentas estatística computacional que auxilia na tomada de decisão referente ao reconhecimento e classificação dos transformadores da rede de distribuição referentes aos níveis de distorções harmônicas totais de tensão.

A metodologia utilizada para o reconhecimento e classificação dos transformadores foi aplicada dentro das suas respectivas superintendências regionais (SDL, SDC, SDO, SDT, SDN) de três maneiras diferentes, utilizando regressão logística, escore quadrático com probabilidades iguais e escore quadrático com probabilidades a priori.

Os modelos utilizados foram fomentados por informações históricas de medições de harmônicos de tensão, realizadas nos transformadores do sistema de distribuição da COPEL. Adicionalmente, foram consideradas na elaboração do modelo, algumas características técnicas e operacionais dos transformadores

monitorados. Estas informações associaram-se aos dados de consumo, tipo e quantidade de consumidores, tensão de operação e potência nominal do transformador, tipo de ligação, dentre outras informações julgadas importantes. Com o processamento destas informações foram criados os modelos, que permitiram estimar a probabilidade de ocorrências de violações de limites das distorções harmônicas de tensão em quaisquer barras da rede elétricas de distribuição. A eficácia dos resultados estimados pelos modelos estatísticos foi mensurada a partir da confrontação com dados reais oriundos de medições realizadas em campo.

As funções criadas para o reconhecimento e classificação, responderam de forma satisfatória para a classificação dos transformadores, mostrando que o *software* com seus respectivos métodos, apresentam um resultado muito coerente com a realidade dos transformadores na rede de distribuição.

Quanto aos métodos utilizados podemos destacar que:

➤ Regressão Logística;

Com a aplicação da técnica de Regressão Logística para o reconhecimento e classificação dos transformadores, obteve-se um resultado satisfatório para todas as superintendências regionais analisadas. Como melhor resultado teve a classificação da superintendência de distribuição centro sul (SDC) com uma taxa de erro aparente de 5,88% e como pior resultado a superintendência de distribuição noroeste (SDN) com uma taxa de erro aparente de 25,40%.

A taxa de erro aparente médio para todas as superintendências, utilizando a Regressão Logística ficou na faixa de 14,64%.

Para esses resultados utilizou-se o valor de 0,35 como o valor de corte para a função logística, pois é aproximadamente a proporção dos grupos de transformadores e também o valor que maximizou a classificação dos grupos.

➤ Escore Quadrático com probabilidades a priori;

Com a aplicação da técnica de Escore quadrático com probabilidades iguais para o reconhecimento e classificação dos transformadores, obteve-se um resultado satisfatório para quatro superintendências regionais. A Superintendência de

Distribuição Oeste (SDO) não se obteve um resultado satisfatório, isso se deve ao mau condicionamento da matriz de dados dos grupos dessa superintendência, pois a matriz de dados foi considerada uma matriz esparsa, ou seja, uma matriz com grande quantidade de valores nulos, com isso a metodologia aplicada não teve desempenho satisfatório para essa superintendência.

Como melhor resultado de classificação temos a Superintendência de Distribuição Norte (SDT) com uma taxa de erro aparente de 6,75% e como pior resultado a Superintendência de Distribuição Oeste (SDO) com uma taxa de erro aparente de 48,39%.

A taxa de erro aparente médio para todas as superintendências, utilizando a Escore Quadrático com probabilidades iguais ficou na faixa de 17,60%.

➤ Escore Quadrático com probabilidades iguais

Como ocorrido com o Escore Quadrático com probabilidades iguais o Escore Quadrático com probabilidade a priori obteve-se um resultado satisfatório no reconhecimento e classificação de quatro superintendências regionais. Como a matriz de dados para a variação dos métodos é a mesma o mesmo problema ocorrido com a matriz de dados da Superintendência de Distribuição Oeste (SDO) o Escore Quadrático com probabilidades a priori não obteve um resultado satisfatório.

Como melhor resultado de classificação temos a Superintendência de Distribuição Norte (SDT) com uma taxa de erro aparente de 6,75% e como pior resultado a Superintendência de Distribuição Oeste (SDO) com uma taxa de erro aparente de 48,39%.

A taxa de erro aparente médio para todas as superintendências, utilizando a Escore Quadrático com probabilidades iguais ficou na faixa de 17,04%.

A abordagem desenvolvida pode ser utilizada como um eficiente instrumento para a realização de uma pré-seleção de transformadores para futuras campanhas de medição pela concessionária, o que gera grande vantagem econômica, já que o custo operacional para medição dos transformadores é muito elevado.

O *software* é uma grande ferramenta para a previsão de uma nova rede de distribuição a ser construída, tendo assim, uma melhor decisão sobre a característica do transformador a ser instalado e a característica do consumo ligado ao mesmo.

O *software* foi estruturado de forma a receber novos elementos a base de dados, desta forma, pode-se realimentar constantemente o banco de dados, afim de melhorar o processo de classificação, pois a cada conjunto de dados inseridos no *software*, tem-se um refinamento dos coeficientes de classificação.

A expectativa é que a aplicação do *software* permita não só uma economia significativa quanto a uma campanha de medição sobre harmônicos, manutenção dos transformadores e auxílio a criação de novas redes, como também uma melhor qualidade de energia aos consumidores, pois com uma ferramenta que permita a estimação das distorções harmônicas totais de tensão é possível buscar soluções mitigadoras para esse problema no sistema.

Outras empresas de energia elétrica poderão fazer uso do *software*, desde que seja feita uma campanha de medição de harmônico para personalizar assim sua base de dados para a criação dos modelos.

5.2 DESENVOLVIMENTOS FUTUROS

Fica, então, como sugestão para trabalhos futuros:

- Elaborar um modelo de reconhecimento e classificação de padrões que leve em consideração outros fatores de influencia na distorção harmônica total de tensão;
- Sugere-se também a aplicação de técnica de inteligência artificial (Redes Neurais Artificiais, Conjuntos Aproximados - *Rough Sets*) e técnicas de programação matemática (SVM - *Support Vector Machines*), para a estimação dos níveis de distorção harmônica total de tensão e comparação com a metodologia aplicada nesse trabalho;

- Desenvolver modelos de reconhecimento e classificação de padrão para a análise das distorções harmônicas individuais de tensão;
- Desenvolver modelos baseados em Sistemas Multi-Agentes para o reconhecimento e classificação de padrões de distorção harmônica de tensão.

REFERÊNCIAS

BARKER, P. P., BURKE, J. J., and others; **Power Quality Monitoring of a Distribution System**"; IEEE Transaction on Power Delivery – Vol. 9 – No 2 – April – 1994.

BRASIL, D. O. C., MEDEIROS, J. R., RIBEIRO, P. OLIVEIRA, J. C., TEIXEIRA, M. D. **Análise Comparativa do Desempenho de Instrumentos Destinados à Monitoração da Qualidade da Energia Elétrica – IX SEPOPE – Maio – 2004**

CHAVES NETO, A. **Análise Multivariada Aplicada a Pesquisa: Notas de aula.** Departamento de Estatística, Universidade Federal do Paraná, Curitiba, 1997.

COPEL – Companhia Paranaense de Energia. Disponível em <http://www.copel.com/pagcopel.nsf>

IEEE. **Task Force on Effects of Harmonics on Equipment, Effects of Harmonic on Equipment**, IEEE Transactions on Power Delivery, Vol. PWD-8, No. 2, April 1993, pp. 672-680.

IEEE Std. 519 – 1992, - **IEEE Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems.** 1992

IEEE Std 1159-1995, **IEEE Recommended Practice for Monitoring Electric Power.** 1995.

IECC Std 61000–4–30. **Electromagnetic Compatibility (EMC), part 4-30: Testing and measurement techniques – Power Quality measurement methods.** Edition 1.1, 2001-03.

IEC Std 61000–4–7. **Electromagnetic Compatibility (EMC), part 4-7: Testing and measurement techniques – General guide on harmonics and interharmonics measurements and instrumentation, for power supply systems and equipment connected thereto.** Edition 1.1, 2001-03.

IGNALLI, R., McGRANAGHAN, M., FIGOR, M.; **Power Quality Monitoring for Reliability Systems.**

IGNATOWICZ, E., BRAZ, R. **Medição amostral de harmônicos de tensão em consumidores de baixa tensão.** Eletricidade Moderna. Outubro – 2003

JOHSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis.** 4. ed. New Jersey: Prentice-Hall, inc., 1998.

KAGAN, N.; FERRARI, E. L.; MATSUO, N. M.; and others; **A Methodology for Monitoring Harmonic Distortion in Electric Power Distribution Systems – IEEE – 2002**

LACHENBRUCH, P. A. **Discriminant Analysis.** New York, Halfner: 1975.

LIMA, J. D. **A análise Econômico-financeira de Empresas Sob a Ótica da Estatística Multivariada.** Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Setores de Tecnologia e de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2002.

NRS 048. **Electricity Supply – Quality Supply.**

Operador Nacional do Sistema Elétrico – **ONS, Submódulo 2.2, Padrões de Desempenho da Rede Básica,** de 22 de Agosto de 2002.

PLA, L. E. **Análisis Multivariado:** Método de Componentes Principales. Secretaria General de la Organización de estados americanos. Washington, 1986.

PENICHE, R, A; TEIXEIRA, M.D. **Conceitos Gerais sobre Qualidade de Energia Elétrica.** Apostila Lactec, 2005.

UNGUREANU. M.; SCUTARIU. M. **Power Quality Measurements on Industrial Sistribution System.** 8th IEEE ICHQP, Greece. October 1998