

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

THIAGO JOSÉ PELISSARI

**MODELO DE RECONHECIMENTO DE PADRÃO APLICADO NO PROCESSO DE
ESTERILIZAÇÃO A VAPOR COM ÊNFASE EM CICLOS FALHOS**

CAMPO MOURÃO

2021

THIAGO JOSÉ PELISSARI

**MODELO DE RECONHECIMENTO DE PADRÃO APLICADO NO PROCESSO DE
ESTERILIZAÇÃO A VAPOR COM ENFÂSE EM CICLOS FALHOS**

**Standard Recognition Model in Steam Sterilization Process with Emphasis in
Failed Cycles**

Dissertação apresentada como requisito para obtenção do título de Mestre em Inovações Tecnológicas da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientador: Prof. Dr. Wyrllen Everson de Souza

Coorientadora: Prof^a. Dr^a. Mirela Vanin dos Santos Lima

CAMPO MOURÃO

2021



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite que outros remixem, adaptem e criem a partir do seu trabalho para fins não comerciais, desde que atribuam o devido crédito e que licenciem as novas criações sob termos idênticos.

Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

THAGO JOSÉ PELISSARI

**MODELO DE RECONHECIMENTO DE PADRÃO APLICADO NO PROCESSO DE
ESTERILIZAÇÃO A VAPOR COM ENFÂSE EM CICLOS FALHOS**

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre em Inovações Tecnológicas da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Inovações Tecnológicas.

Data da Aprovação: 30 de Agosto de 2021

Prof. Dr. Wyrllen Everson de Souza, Doutorado - Universidade Tecnológica Federal do Paraná

Prof.^a Dr.^a Mirela Vanin Dos Santos Lima, Doutorado - Universidade Tecnológica Federal do Paraná

Prof.^a Dr.^a Magda Cardoso Mantovani, Doutorado - Universidade Tecnológica Federal do Paraná

Prof.^a Dr.^a Vanessa Ferreira Sehaber, Doutorado - Universidade Estadual do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 20/09/2021.

CAMPO MOURÃO

2021

Dedico este trabalho à minha família, pelos
momentos de ausência.

AGRADECIMENTOS

Certamente estes parágrafos não irão atender a todas as pessoas que fizeram parte dessa importante fase de minha vida, mas elas podem estar certas de que fazem parte do meu pensamento e de minha gratidão.

Agradeço ao meu orientador Prof. Dr. Wyrllen Everson de Souza e a minha co-orientadora Prof^a. Dr^a. Mirela Vanin dos Santos Lima, pela sabedoria com que me conduziram nesta trajetória.

Aos meus professores e colegas que por inúmeras vezes dedicaram parte de seu tempo comigo.

A empresa Clean-up Brazil Biotecnologia Ltda. e Soinvie do Brasil Ltda., pela cooperação frente a disponibilidade das informações imprescindíveis a execução deste projeto.

Gostaria de deixar registrado também, o meu reconhecimento a todos os meus familiares, pois acredito que sem o apoio deles seria muito difícil vencer esse desafio.

Enfim, a todos que de alguma forma contribuíram para a realização desta pesquisa.

A sorte segue a coragem, desde que a
coragem seja competente. (Cortella,
Mario Sergio, 2018)

RESUMO

A mineração de dados apresenta-se como uma ferramenta poderosa para se alcançar a padronização de parâmetros comportamentais microbianos, auxiliando assim de forma positiva na análise da etapa de monitoramento do processo de esterilização à vapor. Para tanto, modelos matemáticos podem auxiliar no conhecimento da probabilidade de sobrevivência de microrganismos, presentes no meio de cultura de indicadores biológicos, empregados na validação e monitorização dos processos de esterilização. Em complemento, a utilização de técnicas estatísticas como análise discriminante e análise de regressão logística, permitem buscar o conhecimento do comportamento da atividade bacteriana correlacionado a ciclos falhos no processo de esterilização à vapor. Portanto, uma avaliação preditiva rápida e efetiva, sobre o resultado da etapa de monitoramento nos ciclos falhos no processo de esterilização, realizada em até 3 horas quando comparada ao método convencional realizado em até 48 horas, implica na redução do tempo do resultado, bem como na tomada de decisão em função deste resultado. Essa possibilidade permite a indústria fabricante de indicadores biológicos ter um enfoque diferenciado relativo ao seu desempenho e capacidade competitiva em um segmento de mercado já existente, porém por esta atualmente não explorado. Neste sentido, o presente estudo teve por objetivo o desenvolvimento de um modelo para o reconhecimento de padrão, permitindo determinar e avaliar a eficiência de um processo de esterilização a vapor quando submetido a ciclos falhos, através da observação do resultado do teste de monitoramento e validação deste em até 3 horas. A aplicação do método de análise fatorial, método de regressão logística e método de escore quadrático foram adequados para a obtenção de resultados sólidos para a predição da eficácia do modelo. O comportamento do modelo de reconhecimento de padrão previsto vinculado ao método de regressão logística, apresentou no mínimo 97% de aceitabilidade para os dados relativos ao resultado do teste de monitoramento e validação em até 3 horas para ciclos falhos. Assim, pode-se concluir que os resultados observados neste trabalho são condizentes e atendem os requisitos propostos.

Palavras-chave: reconhecimento de padrão; esterilização; indicador biológico; leitura rápida.

ABSTRACT

Data mining presents itself as a powerful tool to achieve the standardization of microbial behavioral parameters, so it can help positively in the analysis of the step monitoring in the steam sterilization process. Therefore, mathematical models can help the understanding of the microorganisms' survival probability present in the culture media of biological indicators, used in the validation and monitoring of sterilization processes. In addition, statistical techniques such as discriminant analysis and logistic regression analysis, allows the search of the bacterial activity behavior correlating the failed cycles in the steam sterilization process. Therefore, a quick and effective predictive evaluation of the monitoring step of the failed cycles in the sterilization process, carried out in up to 3 hours when compared to the conventional method 48 hours, implies a reduction in the result time, as well as in decision-making based on this result. This possibility allows the manufacturing industry of biological indicators to have a differential focus on its performance and competitive capacity in an existing market segment, but currently unexplored. Hereupon, the present study aimed to develop a model for pattern recognition, allowing the definition and evaluation of the efficiency of a steam sterilization process, when subjected to failed cycles, through observation of the monitoring and validation test result in up to 3 hours. The application of the factorial analysis method, logistic regression method, and quadratic score method, were the prediction model effectiveness. The behavior of the predicted pattern recognition model linked to the logistic regression method presented at least 97% of acceptability for data related to the result of the monitoring and validation test in up to 3 hours for failed cycles. Thus, it can be concluded that the results observed in this work are consistent and meets the given requirements.

Keywords: pattern recognition; steam sterilization; biological indicator; rapid-read.

LISTA DE ILUSTRAÇÕES

Figura 1 - Indicadores biológicos autocontidos de 2º geração (vapor), após incubação: a) resultado negativo; b) resultado positivo	48
Figura 2 - Modelo de pacote desafio	49
Figura 3 - Sistema computacional de leitura	53
Figura 4 - Representação esquemática da metodologia.....	54
Figura 5 - Leitura visual de cor dos IB's pelo método colorimétrico.....	56
Figura 6 - Formulário de anotação dos resultados obtidos	57
Quadro 1 - Parâmetros do pacote desafio e do IB 3h utilizado	52
Quadro 2 - Posicionamento dos pacotes desafio contendo indicadores biológicos auto-contido 3 horas na câmara de esterilização.....	55
Quadro 3 - Tempos críticos em minutos para ciclo incompleto (falho)	55
Quadro 4 - Regressão Logística	60
Quadro 5 - Matriz de classificação e taxa de acerto real dos grupos de indicadores biológicos utilizando a regressão logística	61
Quadro 6 - Escore Quadrático.....	62
Quadro 7 - Funções de classificação do escore quadrático	62
Quadro 8 - Matriz de classificação e taxa de acerto utilizando escore quadrático	63
Quadro 9 - Matriz indicativa de classificação e taxa de acerto real dos grupos de indicadores biológicos utilizando a regressão logística após o plaqueamento	65

LISTA DE TABELAS

Tabela 1 - Tabela de classificação	34
Tabela 2 - Carregamentos Fatoriais rotacionados do banco de dados de teste	59

LISTA DE ABREVIATURAS

IB	Indicador Biológico
IB's	Indicadores Biológicos
IB 3h	Indicador Biológico auto-contido 3 horas

SUMÁRIO

1 INTRODUÇÃO	13
1.1 OBJETIVO GERAL	14
1.2 OBJETIVOS ESPECÍFICOS	14
2 REFERENCIAL TEÓRICO	16
2.1 A ANÁLISE E INTERPRETAÇÃO DOS DADOS	16
2.2 RECONHECIMENTO DE PADRÕES	17
2.2.1 Análise Discriminante – Método de Fisher	17
2.2.1.1 Problema geral da classificação com várias populações	18
2.2.1.2 Análise discriminante para duas populações – Método de Fischer	20
2.2.1.3 Discriminação entre diversas populações	23
2.2.1.4 Escore discriminante quadrático de classificação	26
2.3 REGRESSÃO LOGÍSTICA	28
2.3.1 Modelo de Regressão Logística	28
2.3.2 Modelo de Regressão Logística Simples	29
2.3.3 Modelo de Regressão Logístico Múltiplo	32
2.4 AVALIAÇÃO DA FUNÇÃO DE CLASSIFICAÇÃO	32
2.4.1 Critério TPM (<i>Total Probability of Misclassification</i>)	32
2.4.2 Abordagem de <i>Lachenbruch</i>	35
2.5 ANÁLISE FATORIAL	36
2.5.1 Modelo Fatorial Ortogonal	37
2.5.2 Método das Componentes Principais Via Matriz de Covariância	39
2.5.3 Método das Componentes Principais Via Matriz de Correlação Amostral	40
2.5.4 Rotação Ortogonal de Fatores	42
2.5.4.1 Critério varimax	42
2.5.4.2 Critério quartimax	43
2.5.4.3 Critério orthomax	43
2.5.5 Determinação dos Escores Fatoriais	44
2.5.5.1 Método dos mínimos quadrados ponderados	44
2.6 ESTERILIZAÇÃO	45
3 MATERIAL E MÉTODOS	52
3.1 MATERIAL	52
3.2 MÉTODOS	53
3.2.1 Ciclos de Esterilização e Coleta de Dados	53
3.2.2 Execução de Ciclos de Esterilização	54
3.2.3 Identificação do Modelo para Reconhecimento de Padrão a Partir do Método de Análise Fatorial, Análise de Regressão Logística e Método Escore Quadrático.	57

4 ANÁLISE DOS RESULTADOS.....	59
5 CONCLUSÕES	66
REFERÊNCIAS.....	67

1 INTRODUÇÃO

A busca pelo reconhecimento de padrões nos permite visualizar de maneira abrangente as informações relativas a procedimentos univariados quando confrontados (CORREIA; FERREIRA, 2007).

A utilização de técnicas estatísticas, como análise discriminante e análise de regressão logística, nos possibilita buscar o conhecimento do comportamento da atividade bacteriana correlacionado a ciclos falhos quando presentes no processo de esterilização a vapor.

Os modelos matemáticos podem ser utilizados para analisar resultados, fornecer previsões e sugestões para experimentos, podem também sintetizar conhecimento existente fornecendo uma estrutura teórica para interpretação de paradigmas existentes (PIGOZZO, 2011).

Para RODRIGUES (2010), os modelos matemáticos podem ser utilizados para avaliar comportamentos microbianos no processo de esterilização por vapor, isso através da introdução do conhecimento da cinética química.

Segundo SARAIVA (2016), existem diferentes métodos de esterilização, sendo por calor úmido, por calor seco, por radiação, esterilização química, dentre outros. Um meio efetivo para o processo de esterilização dá-se através do calor úmido gerado pela autoclave (DA SILVA *et al.*, 2017).

A inativação de microrganismos por meios físicos ou químicos segue uma lei exponencial e, portanto, há uma probabilidade estatística de que microrganismos possam sobreviver ao processo de esterilização (BRASIL, 2010).

Para BUENO *et al.* (2017), interrupções não programadas ocorridas durante um ciclo de esterilização podem ser observadas através do denominado aborto do equipamento ou em outras palavras, ciclo falho da autoclave, sendo este ocorrido pelo não atingimento dos parâmetros (tempo, temperatura e pressão) considerados cruciais para a garantia de um correto ciclo de esterilização.

Para SALMAZO NETTO & DE SOUZA (2014), indicadores biológicos (IB's) são utilizados como indicadores de processo para validar processos de esterilização. A alteração de cor de violeta para amarelo indica atividade de crescimento microbiano no meio.

As classificações dos indicadores biológicos segundo MARTINHO (2007), são ordenadas por gerações, as quais contam com diferentes metodologias de leitura para o resultado final da verificação da eficácia do processo de esterilização.

De uma maneira geral, propõe-se conhecer o comportamento da atividade bacteriana correlacionado a ciclos falhos quando presentes no processo de esterilização a vapor.

Neste sentido, a proposta recai sobre uma avaliação preditiva rápida e efetiva no tocante a atividade de leitura final para detecção de ciclos falhos no processo de esterilização em um período de tempo de até 3 horas frente ao método convencionalmente utilizado que é de até 48 horas, implicando na redução direta do tempo de resultado e tomada de decisão, possibilitando a indústria fabricante de indicadores biológicos ter um enfoque diferenciado no que concerne ao seu desempenho e capacidade de competir em um segmento de mercado já existente, porém atualmente não explorado por esta.

Assim, o presente trabalho objetivará apontar um modelo de reconhecimento de padrão aplicado no processo de esterilização a vapor com ênfase em ciclos falhos.

1.1 Objetivo Geral

O objetivo geral do presente estudo foi desenvolver um modelo para o reconhecimento de padrão, capaz de determinar e avaliar a eficiência de um processo de esterilização a vapor, quando submetido a ciclos falhos, através de leitura final em até 3 horas.

1.2 Objetivos Específicos

O objetivo geral deverá ser atingido a partir dos seguintes objetivos específicos:

Aplicação *in loco* da experimentação correlacionada a execução de ciclos completos e falhos no processo de esterilização a vapor;

Identificar modelo adequado para reconhecimento de padrão a partir da coleta de dados pela utilização da análise fatorial, análise de regressão logística e método do escore quadrático;

Analisar a praticabilidade do modelo à problemática apontada no sentido de prever a eficiência do mesmo relacionada ao resultado da leitura em um período de até 3 horas de incubação.

2 REFERENCIAL TEÓRICO

2.1 A Análise e Interpretação dos Dados

A análise e interpretação dos dados dentre as diversas áreas do conhecimento têm sido possível devido ao avanço da tecnologia computacional (BAKKE *et al.*, 2008).

A partir da coleta de dados experimentais e da observação do que se deseja modelar, torna-se possível a construção de um modelo matemático capaz de descrever mediante um sistema de equações o comportamento de um processo (MOREIDA, CASAS, 2019). Para PIAZZI (2019), através do estabelecimento de variáveis de processo e variáveis independentes, é possível a elaboração de um modelo matemático.

Uma base de dados segundo BARRAGAN *et al.* (2012), pode contribuir positivamente para a geração de aprendizado, para avanços tecnológicos e descobertas científicas, isso decorrente da análise dos dados comportamentais a qual origina-se de um processo denominado mineração de dados. Na mesma linha de raciocínio SILVA & PERIÇARO (2009), contemplam que a eficiente utilização de algoritmos computacionais aliados a descoberta de padrões e regras, estes interligados a extração de conhecimento a partir da análise de dados, denominando-se a chamada mineração de dados. A análise de um comportamento futuro bem como, o auxílio deste monitoramento atrelado a modelos matemáticos, desempenham essencial importância frente ao relativo conhecimento entre os seres vivos (DA SILVA; DA SILVA, 2018).

Conforme salientam SANTOS & ANTONELLI (2011), o processo de tomada de decisão torna-se desafiador frente aos benefícios dos quais a abordagem estatística pode proporcionar.

No sentido de contribuir pela busca de uma resposta sobre se determinado elemento encaixa-se ou não a uma categoria específica, GUIMARÃES & MOREIRA (2008), sugerem a aplicação da técnica estatística multivariada denominada análise discriminante. Em complemento, para FERREIRA (2019), discernir uma variável não métrica (dependente) de uma variável métrica (independente), torna-se possível através da aplicação da técnica estatística denominada análise discriminante.

Em relação à procura de uma resposta voltada em categorias com ordenação, na forma de escore na análise de dados estatísticos, ABREU *et al.* (2009), propõem a aplicação de modelos de regressão logística. Esta técnica estatística de regressão logística, trata-se de um importante modelo de regressão onde a variável de resposta assume valores dicotômicos, ilustrado por 0 (zero) ou 1 (um), tendo como objetivo o modelamento de relações entre as variáveis (CABRAL, 2013).

2.2 Reconhecimento de Padrões

Algumas categorias ou classes de objetos padrões, segundo MULLER (1997), são passíveis de identificação e discriminação através da área de pesquisa denominada reconhecimento de padrões. De maneira geral, quando se têm variações pequenas dentro de uma mesma categoria e variações grandes entre diferentes categorias, deseja-se que estas, estejam relacionadas a um problema de reconhecimento de padrões muito bem definido e restrito.

A Análise Discriminante (Método de Fischer), o Escore Discriminante Quadrático, a Regressão Logística e Método das k-médias são considerados segundo Chaves Neto (1997, apud SOUZA, 2008), os quatros métodos principais baseados na construção de regras de reconhecimento e classificação de padrões para duas populações. A utilização ainda, de Redes Neurais, Conjuntos Aproximados (Rough Sets), dentre outras técnicas de Inteligência Artificial também podem ser citadas afim de utilização para reconhecimento de padrões.

2.2.1 Análise Discriminante – Método de Fisher

Segundo Johnson & Wichern (1998, apud SOUZA, 2012), a Análise Discriminante é uma técnica multivariada que tem por objetivo tratar problemas relacionados a separação em 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.

Esta técnica tem os seguintes procedimentos imediatos, quando usada para discriminação e classificação:

1. Descrever algebricamente ou graficamente as características diferenciais dos objetos (observações) de várias populações conhecidas com o intuito 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, frequentemente, 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.

2.2.1.1 Problema geral da classificação com várias populações

Considerando:

- $f_i(\underline{X})$ 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:

$$P(k|i) = \int_{R_k} f_i(\underline{X}) dx \quad i = 1, \dots, g \quad (2.1)$$

➤ 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:

$$ECM(1) = P(2|1)c(2|1) + \dots + P(g|1)c(g|1) = \sum_{k=2}^g P(k|1)c(k|1) \quad (2.2)$$

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

$$ECM = p_1 ECM(1) + p_2 ECM(2) + \dots + p_g ECM(g) \quad (2.3)$$

logo,

$$P(k|i) = \int_{R_k} f_i(\underline{X}) dx \quad i = 1, \dots, g \quad (2.4)$$

$$ECM = p_1 \left(\sum_{k=2}^g P(k|1)c(k|1) \right) + \sum_{k=1}^g P(k|2)c(k|2) + \dots + \sum_{k=1}^{g-1} P(k|g)c(k|g) \quad (2.5)$$

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 \underline{X} na população Π_k de forma que:

$$\sum_{\substack{k=1 \\ k \neq 2}}^g P_i f(\underline{X}) c(k|i) \quad (2.6)$$

seja mínimo e, num caso de empate, \underline{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 \underline{X} na população Π_k em que:

$$\sum_{\substack{i=1 \\ i \neq k}}^g P_i f(\underline{X}) \quad (2.7)$$

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

$$P_k f_k(\underline{X}) > P_i f_i(\underline{X}) \text{ para todo } i \neq k \quad (2.8)$$

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

$$\ln P_k f_k(\underline{X}) > \ln P_i f_i(\underline{X}) \text{ para todo } i \neq k. \quad (2.9)$$

2.2.1.2 Análise discriminante para duas populações – Método de Fischer

A ideia de Fischer foi transformar observações multivariadas $\underline{X}'s$ 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 \underline{X} para criar os $Y's$, dado que as combinações lineares são funções de \underline{X} e por outro lado são de fácil cálculo. Assim sendo μ_{1y} a média dos $Y's$ obtidos dos $\underline{x}'s$ pertencentes a Π_1 (população 1) e μ_{2y} a média dos $Y's$ obtidos dos $\underline{x}'s$ 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:

$$\mu_1 = E(\underline{X}|\Pi_1) = \text{valor esperado em uma observação multivariada em } \Pi_1. \quad (2.10)$$

$$\mu_2 = E(\underline{X}|\Pi_2) = \text{valor esperado de uma observação multivariada em } \Pi_2. \quad (2.11)$$

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

$$Y_{1 \times 1} = \underline{c}'_{1 \times p} \underline{X}_{p \times 1} \quad (2.12)$$

tem-se :

$$\mu_{1y} = E(Y|\Pi_1) = E(\underline{c}' \underline{X}|\Pi_1) = \underline{c}' E(\underline{X}|\Pi_1) = \underline{c}' \underline{\mu}_1 \quad (2.13)$$

$$\mu_{2y} = E(Y|\Pi_2) = E(\underline{c}' \underline{X}|\Pi_2) = \underline{c}' E(\underline{X}|\Pi_2) = \underline{c}' \underline{\mu}_2 \quad (2.14)$$

e

$$V(Y) = \sigma_y^2 = V(\underline{c}' \underline{X}) = \underline{c}' V(\underline{X}) \underline{c} = \underline{c}' \Sigma \underline{c} \quad (2.15)$$

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”.

$$\frac{(\mu_{1y} - \mu_{2y})^2}{\sigma_y^2} = \frac{(\underline{c}'\underline{\mu}_1 - \underline{c}'\underline{\mu}_2)^2}{\underline{c}'\Sigma\underline{c}} = \frac{\underline{c}'(\underline{\mu}_1 - \underline{\mu}_2)(\underline{\mu}_1 - \underline{\mu}_2)'\underline{c}}{\underline{c}'\Sigma\underline{c}} = \frac{(\underline{c}'\underline{\delta})^2}{\underline{c}'\Sigma\underline{c}} \quad (2.16)$$

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 $\frac{(\underline{c}'\underline{\delta})^2}{\underline{c}'\Sigma\underline{c}}$ é maximizada por:

$$\underline{c} = k\Sigma^{-1}\underline{\delta} = k\Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) \quad \text{para qualquer } k \neq 0. \quad (2.17)$$

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

$$\underline{c} = \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2) \quad \text{e} \quad Y = \underline{c}'\underline{X} = (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{X}, \quad (2.18)$$

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 \underline{x}_0 usa-se o modelo:

$$y_0 = (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{x}_0 \quad (2.19)$$

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

$$m = \frac{1}{2}(\mu_{1y} + \mu_{2y}), \quad (2.20)$$

como

$$m = \frac{1}{2}(\underline{c}'_1\underline{\mu}_1 + \underline{c}'_2\underline{\mu}_2), \quad (2.21)$$

$$m = \frac{1}{2}[(\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{\mu}_1 + (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}\underline{\mu}_2] \quad (2.22)$$

$$m = \frac{1}{2}[(\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 + \underline{\mu}_2)] \quad (2.23)$$

tem-se que:

$$E(y_0|\Pi_1) - m \geq 0 \quad (2.24)$$

e

$$E(y_0|\Pi_2) - m < 0 \quad (2.25)$$

ou seja, se \underline{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 \underline{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 \underline{x}_0 em Π_1 se $y_0 - m \geq 0$;

alocar \underline{x}_0 em Π_2 se $y_0 - m < 0$.

Os parâmetros $\underline{\mu}_1$, $\underline{\mu}_2$ e Σ geralmente são desconhecidos. Então, supondo que se tem n_1 observações da variável aleatória multivariada \underline{x}_1 de dimensão P , ou seja, tem-se uma amostra aleatória da população Π_1 e n_2 observações da variável aleatória multivariada \underline{x}_2 de dimensão P que corresponde a uma amostra aleatória de população Π_2 , os resultados amostrais correspondentes são:

$$\bar{\underline{x}}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \underline{x}_{i1}; \quad S_1 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (\underline{x}_{i1} - \bar{\underline{x}}_1)(\underline{x}_{i1} - \bar{\underline{x}}_1)' \quad (2.26)$$

$$\bar{\underline{x}}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \underline{x}_{i2}; \quad S_2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (\underline{x}_{i2} - \bar{\underline{x}}_2)(\underline{x}_{i2} - \bar{\underline{x}}_2)' \quad (2.27)$$

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,

$$S_p = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)} \quad (2.28)$$

que é um estimador não –viciado daquele parâmetro Σ .

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

$$\hat{Y} = \hat{c}X = (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S_p^{-1} X \quad (2.29)$$

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

$$\bar{y}_1 = \hat{c}\bar{\underline{x}}_1 \quad (2.30)$$

e

$$\bar{y}_2 = \hat{c}\bar{\underline{x}}_2 \quad (2.31)$$

é dada por:

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

$$\hat{m} = \frac{1}{2}(\bar{x}_1 - \bar{x}_2)' S_p^{-1} (\bar{x}_1 + \bar{x}_2)$$
(2.32)

Finalizando, a regra de classificação do item caracterizado pelo vetor \underline{x}_0 é:

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

ou

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

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

$$\frac{(\bar{y}_1 - \bar{y}_2)^2}{S_y^2} = \frac{(\hat{c}_1 \bar{x}_1 - \hat{c}_2 \bar{x}_2)}{\hat{c}' S_p \hat{c}} = \frac{(\hat{c}' \underline{d})^2}{\hat{c}' S_p \hat{c}}$$
(2.33)

onde:

$$\underline{d} = \bar{x}_1 - \bar{x}_2$$
(2.34)

e

$$S_y^2 = \frac{\sum_{i=1}^{n_1} (y_{i1} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{i2} - \bar{y}_2)^2}{n_1 + n_2 - 2}$$
(2.35)

2.2.1.3 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:

$$B = (1 - c) \left\{ \left[\sum_{i=1}^g (n_i - 1) \ln |S_p| - \sum_{i=1}^g [(n_i - 1) \ln |S_i|] \right] \right\} \sim \chi_{\frac{1}{2} p(p+1)(g+1)}^2 \quad (2.36)$$

com

$$c = \left[\sum_{i=1}^g \frac{1}{n_i - 1} - \frac{1}{\sum_{i=1}^g (n_i - 1)} \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right] \quad (2.37)$$

onde:

- p é a dimensão do vetor \underline{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 $\underline{\bar{\mu}}$ o vetor médio dos diversos grupos (populações), representado por:

$$\underline{\bar{\mu}} = \frac{1}{g} \sum_{i=1}^g \underline{\mu}_i \quad (2.38)$$

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

$$B_0 = \sum_{i=1}^g (\underline{\mu}_i - \underline{\bar{\mu}})(\underline{\mu}_i - \underline{\bar{\mu}})' \quad (2.39)$$

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

$$E(Y) = \underline{c}' E(\underline{x}|\Pi_i) = \underline{c}' \underline{\mu}_i \quad (2.40)$$

para população Π_1 e variância:

$$V(Y) = \sigma_y^2 = \underline{c}' V(\underline{X}) \underline{c} = \underline{c}' \Sigma \underline{c} \quad (2.41)$$

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

$$\bar{\mu}_y = \frac{1}{g} \sum_{i=1}^g \mu_{iy} = \underline{c}' \bar{\underline{\mu}} \quad (2.42)$$

e a razão entre a “Soma dos quadrados das distâncias das populações para a média

global” e a variância de Y é $\frac{\underline{c}' B_0 \underline{c}}{\underline{c}' \Sigma \underline{c}}$, 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 \underline{c} que maximiza esta razão. É conveniente normalizar \underline{c} tal que $\underline{c}' \Sigma \underline{c} = 1$.

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 $\underline{e}' \Sigma \underline{e} = 1$). Então,

é fácil provar que o vetor de coeficientes \underline{c} que maximiza $\frac{\underline{c}' B_0 \underline{c}}{\underline{c}' \Sigma \underline{c}}$ é dado por $\underline{c}_1 = \underline{e}_1$. A combinação linear $\underline{c}_1' \underline{X}$ é chamada de primeira discriminante e de forma idêntica, pode-se generalizar para o k -ésimo discriminante com $\underline{c}_k = \underline{e}_k$, com $k = 1, 2, \dots, s$. Como geralmente Σ e $\underline{\mu}_i$ 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 $\underline{\mu}_i$ e $\bar{\underline{\mu}}$ dados por:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad (2.43)$$

$$\bar{x} = \frac{\sum_{i=1}^g n_i \bar{x}_i}{\sum_{i=1}^g n_i} = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} x_{ij}}{\sum_{i=1}^g n_i} \quad (2.44)$$

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

$$\hat{B}_0 = \sum_{i=1}^g (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \quad (2.45)$$

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

$$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x})(x_{ij} - \bar{x})' = \sum_{i=1}^g (n_i - 1)S_i \quad (2.46)$$

Consequentemente,

$$\frac{W}{n_1 + n_2 + \dots + n_g - g} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2 + \dots + (n_g - 1)S_g}{n_1 + n_2 + \dots + n_g - g} = S_p \quad (2.47)$$

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 $\hat{e}_k' x$, $k \leq s$.

2.2.1.4 Escore discriminante quadrático de classificação

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

$$f_i(\underline{X}) = \frac{1}{(2\pi)^{p/2} |\Sigma_i|} \exp \left[-\frac{1}{2} (\underline{X} - \underline{\mu}_i)' \Sigma_i (\underline{X} - \underline{\mu}_i) \right], \quad i = 1, 2, \dots, g \quad (2.48)$$

em que $\underline{\mu}_i = E(\underline{x} | \Pi_i)$ é o vetor de médias e $V(\underline{x} | \Pi_i) = \Sigma_i$ 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 (2.7), a regra de classificar fica:

alocar \underline{X} em Π_k se

$$\ln p_i f_i(\underline{X}) = \ln p_i - \left(\frac{p}{2} \right) \ln(2\pi) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (\underline{X} - \underline{\mu}_i)' \Sigma_i^{-1} (\underline{X} - \underline{\mu}_i), \quad i = 1, 2, \dots, g \quad (2.49)$$

Na equação anterior 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^Q para a população Π_i como:

$$d_i^Q(\underline{X}) = \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (\underline{X} - \underline{\mu}_i)' \Sigma_i^{-1} (\underline{X} - \underline{\mu}_i) + \ln p_i, \quad i = 1, 2, \dots, g \quad (2.50)$$

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

$$d_i^Q(\underline{X}) = \max_i d_i^Q(\underline{X}), \quad i = 1, 2, \dots, g \quad (2.51)$$

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

$$D_i = \frac{1}{2} \ln |S_i| - \frac{1}{2} (\underline{X} - \bar{\underline{X}}_i)' S_i^{-1} (\underline{X} - \bar{\underline{X}}_i) + \ln p_i, \quad i = 1, 2, \dots, g \quad (2.52)$$

onde:

- $\bar{\underline{X}}_i$ 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 .

2.3 Regressão Logística

No ano de 1845 e devido a problemas ligados ao crescimento demográfico, foi desenvolvida a denominada função Logística, a qual em tais aspectos desempenha papel de suma importância. A sua aplicação no âmbito da biologia segundo LIMA (2002), deu-se a partir da década de 1930.

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 $p \in \mathfrak{R}$.

2.3.1 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:

$$P(\tilde{X}) = \frac{e^{\beta' \underline{X}}}{1 + e^{\beta' \underline{X}}} = \frac{1}{1 + e^{-\beta' \underline{X}}} \quad (2.53)$$

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

Segundo Johnson & Wichern (1998, apud BARROS, 2018), esse modelo é o mais apropriado no caso de a variável dependente ser dicotômica, pois a Regressão Logística estima diretamente a probabilidade de ocorrência de um evento. Estes autores, também salientam que a estimativa dos parâmetros deve ser feita pelo método da máxima verossimilhança, sendo o mais recomendado quando se dispõe de observações individuais da ocorrência ou não de determinado evento.

2.3.2 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_i 's representam os valores observados de uma variável dicotômica, e os X_i 's representam os valores observados de uma única variável independente.

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

$$P(Y = 1 | X) = P(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}} \quad (2.54)$$

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. Determinando o modelo de Regressão Logística, a probabilidade condicional de Y ser igual a 0, é:

$$P(Y = 0 | X) = 1 - P(Y = 1 | X) \quad (2.55)$$

Portanto para as n observações têm-se:

$$P(Y_i = 1 | X_i) = P(X_i) \quad e \quad P(Y_i = 0 | X_i) = 1 - P(X_i) \quad (2.56)$$

O método da Máxima Verossimilhança é o 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:

$$F(X_i) = P(X_i)^{Y_i} [1 - P(X_i)]^{1 - Y_i} \quad (2.57)$$

Desde que assumindo que as observações são independentes, a função de verossimilhança é obtida como o produto dos termos dados na equação (2.54) como segue:

$$l(\beta) = \prod_{i=1}^n F(X_i) = \prod_{i=1}^n [P(X_i)^{Y_i}][1 - P(X_i)]^{1 - Y_i} \quad (2.58)$$

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:

$$L(\beta) = \ln(l(\beta)) = \ln \left[\prod_{i=1}^n F(X_i) \right] = \ln \left[\prod_{i=1}^n [P(X_i)^{Y_i}] [1 - P(X_i)]^{1-Y_i} \right] \quad (2.59)$$

assim:

$$L(\beta) = \sum_{i=1}^n [Y_i \ln(P(X_i))] + (1 - Y_i) \ln[1 - P(X_i)] \quad (2.60)$$

então:

$$L(\beta) = \sum_{i=1}^n \left[Y_i \ln \frac{e^{\beta' X}}{1 + e^{\beta' X}} + (1 - Y_i) \ln \left(1 - \frac{e^{\beta' X}}{1 + e^{\beta' X}} \right) \right] \quad (2.61)$$

temos:

$$L(\beta) = \sum_{i=1}^n \left[Y_i \ln \frac{e^{\beta' X}}{1 + e^{\beta' X}} + (1 - Y_i) \ln \left(\frac{1}{1 + e^{\beta' X}} \right) \right] \quad (2.62)$$

mas

$$\ln \left(\frac{e^{\beta' X}}{1 + e^{\beta' X}} \right) = \ln e^{\beta' X} - \ln(1 + e^{\beta' X}) = \beta' X - \ln(1 + e^{\beta' X}) \quad (2.63)$$

e

$$\ln \left(\frac{1}{1 + e^{\beta' X}} \right) = \ln 1 - \ln(1 + e^{\beta' X}) = -\ln(1 + e^{\beta' X}) \quad (2.64)$$

Substituindo (2.63) e (2.64) em (2.62) tem-se :

$$L(\beta) = \sum_{i=1}^n [Y_i (\beta' X - \ln(1 + e^{\beta' X})) + (1 - Y_i) (-\ln(1 + e^{\beta' X}))] \quad (2.65)$$

assim:

$$L(\beta) = \sum_{i=1}^n [Y_i \beta' X - Y_i \ln(1 + e^{\beta' X}) - \ln(1 + e^{\beta' X}) + Y_i \ln(1 + e^{\beta' X})] \quad (2.66)$$

Concluindo:

$$L(\beta) = \sum_{i=1}^n [Y_i \beta' X - \ln(1 + e^{\beta' X})] \quad (2.67)$$

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 .

$$\frac{\partial L(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left(Y_i - \frac{e^{\beta' X}}{1 + e^{\beta' X}} \right) = \sum_{i=1}^n [Y_i - P(X_i)] = 0 \quad (2.68)$$

$$\frac{\partial L(\beta)}{\partial \beta_1} = \sum_{i=1}^n \left(X_i \cdot Y_i - X_i \frac{e^{\beta' X}}{1 + e^{\beta' X}} \right) = \sum_{i=1}^n X_i [Y_i - P(X_i)] = 0 \quad (2.69)$$

Temos que:

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n \frac{e^{\beta' X}}{1 + e^{\beta' X}} = \sum_{i=1}^n P(X_i) \quad (2.70)$$

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, que X é igual a X_i .

2.3.3 Modelo de Regressão Logístico Múltiplo

Considerando-se uma amostra de n pares de observações independentes (\underline{X}_i, Y_i) , $i = 1, 2, \dots, p$, da mesma forma como ocorreu no caso univariado, item anterior, o ajuste do modelo requer que se obtenha uma estimativa do vetor $\underline{\beta}' = [\beta_0, \beta_1, \dots, \beta_p]$. O método de estimativa 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, porém com uma mudança, sendo que $P(X)$ é definido por:

$$P(Y = 1 | \underline{X}) = P(\underline{X}_i) = \frac{e^{\underline{\beta}' \underline{X}}}{1 + e^{\underline{\beta}' \underline{X}}} \quad (2.71)$$

Calculando $(p+1)$ 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:

$$\sum_{i=1}^n [Y_i - P(X_i)] = 0 \quad (2.72)$$

e

$$\sum_{i=1}^n X_{ij} [Y_i - P(X_i)] = 0 \quad (2.73)$$

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

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

2.4 Avaliação da Função de Classificação

2.4.1 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:

$$TPM = p_1 \int_{R_2} f_1(\underline{X}) d\underline{X} + p_2 \int_{R_1} f_2(\underline{X}) d\underline{X} \quad (2.74)$$

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:

$$OER = p_1 \int_{R_2} f_1(\underline{X}) d\underline{X} + p_2 \int_{R_1} f_2(\underline{X}) d\underline{X} \quad (2.75)$$

onde as regiões são determinadas por:

$$R_1 = \frac{f_1(\underline{X})}{f_2(\underline{X})} \geq \frac{p_2}{p_1} \text{ e } R_2 = \frac{f_1(\underline{X})}{f_2(\underline{X})} < \frac{p_2}{p_1} \quad (2.76)$$

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*)

$$AER = p_1 \int_{\hat{R}_2} f_1(\underline{X}) d\underline{X} + p_2 \int_{\hat{R}_1} f_2(\underline{X}) d\underline{X} \quad (2.77)$$

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 funções densidades de probabilidade $f_1(\underline{X})$ e $f_2(\underline{X})$ 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 funções densidades de probabilidade 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 1 a seguir:

Tabela 1 - Tabela de classificação

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

Fonte: Johnson & Wichern (1998) apud Souza (2012).

onde:

- $n_{i/j}$ = números de observações de Π_i classificadas em Π_j e consideradas corretas se $i = j$ ou incorreta se $i \neq j$;
- $n_{1/1}$ = número 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 Rate – APER*) é dada por:

$$APER = \frac{n_{1/2} + n_{2/1}}{n_1 + n_2} \quad (2.78)$$

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 necessária 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.

2.4.2 Abordagem de *Lachenbruch*

O método de *Lachenbruch* (1975), segundo Johnson & Wichern (1998, apud SOUZA, 2012), 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:

$$P(2/1) = \int_{R_2} f_1(\underline{X}) d\underline{X} = \frac{n_{2/1}}{n_1} \quad (2.79)$$

classificar em Π_2 , quando a observação pertence a Π_1 ,

$$P(1/2) = \int_{R_1} f_2(\underline{X}) d\underline{X} = \frac{n_{1/2}}{n_2} \quad (2.80)$$

classificar em Π_1 , quando a observação pertence a Π_2 , desta forma a proporção esperada de erro (probabilidade estimada) é dada por:

$$\hat{E}(AER) = \frac{n_{1/2} + n_{2/1}}{n_1 + n_2} \quad (2.81)$$

2.5 Análise Fatorial

A definição de uma estrutura subjacente em uma matriz de dados segundo HAIR et al. (2009), se dá como propósito principal do método estatístico multivariado conhecido como Análise Fatorial. Este de maneira geral, atua na análise da estrutura das correlações entre um grande número de variáveis. O seu principal objetivo é interpretar e compreender as dimensões latentes advindas do resumo dos dados, as quais os descrevem em um número muito menor de conceitos do que as variáveis originais individuais.

De uma maneira geral, Mingoti (2005, apud SOUZA, 2012), deseja que ocorra o agrupamento em subconjuntos de novas variáveis mutuamente não correlacionadas a partir de variáveis originais. Através da Análise Fatorial, possibilita-se identificar um número menor de novas variáveis alternativas, não correlacionadas e que de algum modo sumarizassem as informações principais das variáveis originais, quando estas, provenientes de casos no quais se tem um grande número de variáveis medidas e correlacionadas entre si.

Duas são as maneiras as quais se podem trabalhar a Análise Fatorial. No caso da Análise Fatorial exploratória, busca-se encontrar os fatores subjacentes às

variáveis originais amostradas, sem ter noção de quantos fatores serão obtidos e de que maneira as variáveis estarão agrupadas. Já na Análise Fatorial confirmatória, tem-se um modelo inicial pré-estabelecido, onde se deseja verificar se este modelo proposto é consistente com os dados amostrais disponíveis.

2.5.1 Modelo Fatorial Ortogonal

De acordo com Johnson & Wichern (1998, apud BARROS, 2018), o vetor aleatório \underline{X} , com p variáveis aleatórias observáveis, tem vetor de médias $\underline{\mu}$ e matriz de covariância Σ . O modelo fatorial é linearmente dependente sobre variáveis aleatórias não observáveis F_1, F_2, \dots, F_m ($m \leq p$) chamados fatores comuns e p fontes de variações ativas $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, chamadas de erros ou fatores específicos. O modelo fatorial é

$$\begin{array}{rccccccccc} X_1 - \mu_1 & = & \ell_{11}F_1 & + & \ell_{12}F_2 & + & \dots & + & \ell_{1m}F_m & + & \varepsilon_1 \\ X_2 - \mu_2 & = & \ell_{21}F_1 & + & \ell_{22}F_2 & + & \dots & + & \ell_{2m}F_m & + & \varepsilon_2 \\ \vdots & & \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ X_p - \mu_p & = & \ell_{p1}F_1 & + & \ell_{p2}F_2 & + & \dots & + & \ell_{pm}F_m & + & \varepsilon_p \end{array} \quad (2.82)$$

ou em notação matricial

$$\underbrace{\underline{X} - \underline{\mu}}_{(p \times 1)} = \underbrace{\underline{L}}_{(p \times m)} \cdot \underbrace{\underline{F}}_{(m \times 1)} + \underbrace{\underline{\varepsilon}}_{(p \times 1)} \quad (2.83)$$

onde L é a matriz de ordem $p \times m$ contendo os pesos ℓ_{ij} denominados também como carregamento. Assim, ℓ_{ij} é o peso da i -ésima variável no j -ésimo fator, então a matriz L é chamada de matriz de cargas fatoriais. O i -ésimo fator específico ε_i é associado somente com a variável resposta X_i . Os p desvios $X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p$ são representados pelas $p+m$ variáveis aleatórias $F_1, F_2, \dots, F_m, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, as quais não são observáveis.

Com tantas quantidades não observáveis e para tornar útil o modelo fatorial, algumas pressuposições sobre os vetores \underline{F} e $\underline{\varepsilon}$ devem ser feitas, tais como

$$E(\underline{F}) = \underline{0}_{m \times 1}, \quad \text{Cov}(\underline{F}) = E(\underline{F}\underline{F}') = \underline{I}_{m \times m} \quad (2.84)$$

$$E(\underline{\varepsilon}) = \underline{\mathbf{0}}_{p \times 1}, \quad \text{Cov}(\underline{\varepsilon}) = E(\underline{\varepsilon}\underline{\varepsilon}') = \underline{\Psi}_{p \times p} = \begin{bmatrix} \Psi_1 & 0 & \dots & 0 \\ 0 & \Psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Psi_p \end{bmatrix} \quad (2.85)$$

e ainda \underline{F} e $\underline{\varepsilon}$ são independentes, então

$$\text{Cov}(\underline{\varepsilon}\underline{F}) = E(\underline{\varepsilon}\underline{F}') = \underline{\mathbf{0}}_{p \times m} \quad (2.86)$$

Essa suposição e a relação em (2.86) constituem o modelo fatorial ortogonal. A estrutura de covariância para o modelo fatorial ortogonal, pode ser obtida da seguinte forma:

$$\Sigma = V(\underline{X}) = E[(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})'] \quad (2.87)$$

porém de (2.86) obtém-se

$$\begin{aligned} \Sigma &= E[(\underline{L}\underline{F} + \underline{\varepsilon})(\underline{L}\underline{F} + \underline{\varepsilon})'] \\ \Sigma &= E[\underline{L}\underline{F}\underline{F}'\underline{L}' + \underline{\varepsilon}\underline{F}'\underline{L}' + \underline{L}\underline{F}\underline{\varepsilon}' + \underline{\varepsilon}\underline{\varepsilon}'] \\ \Sigma &= \underline{L}E(\underline{F}\underline{F}')\underline{L}' + E(\underline{\varepsilon}\underline{F}')\underline{L}' + \underline{L}E(\underline{F}\underline{\varepsilon}') + E(\underline{\varepsilon}\underline{\varepsilon}') \end{aligned} \quad (2.88)$$

logo

$$\Sigma = \underline{L}\underline{L}' + \underline{\Psi} \quad (2.89)$$

Desta forma, a covariância entre as variáveis e os fatores é calculado por:

$$\begin{aligned} \text{Cov}(\underline{X}, \underline{F}) &= E[(\underline{X} - \underline{\mu})(\underline{F} - E(\underline{F}))'] \\ \text{Cov}(\underline{X}, \underline{F}) &= E[(\underline{L}\underline{F} + \underline{\varepsilon})\underline{F}'] \\ \text{Cov}(\underline{X}, \underline{F}) &= E[\underline{L}\underline{F}\underline{F}' + \underline{\varepsilon}\underline{F}'] \\ \text{Cov}(\underline{X}, \underline{F}) &= \underline{L}E(\underline{F}\underline{F}') + E(\underline{\varepsilon}\underline{F}') \end{aligned} \quad (2.90)$$

ou seja,

$$\text{Cov}(\underline{X}, \underline{L}) = \underline{L} \quad (2.91)$$

Sendo assim, por $\Sigma = \underline{L}\underline{L}' + \underline{\Psi}$ obtém-se

$$V(X_i) = \ell_{i1}^2 + \dots + \ell_{im}^2 + \Psi_i \quad (2.92)$$

$$\text{Cov}(X_i, X_k) = \ell_{i1}\ell_{k1} + \dots + \ell_{im}\ell_{km} \quad (2.93)$$

e de (2.92) resulta que a covariância entre a i-ésima variável e o j-ésimo fator é:

$$\text{Cov}(X_i, F_j) = \ell_{ij} \quad (2.94)$$

A parte da variância que a i -ésima variável contribui para os m fatores comuns é chamada de comunalidade, que é dada por:

$$h_i^2 = \ell_{i1}^2 + \dots + \ell_{im}^2 \quad (2.95)$$

e Ψ_i são as variâncias específicas.

2.5.2 Método das Componentes Principais Via Matriz de Covariância

Usando a decomposição espectral pode-se fatorar a matriz de covariância Σ . Sejam $(\lambda_i, \underline{e}_i)$ pares de autovalores autovetores da matriz Σ , com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Então:

$$\Sigma = \lambda_1 \underline{e}_1 \underline{e}_1' + \lambda_2 \underline{e}_2 \underline{e}_2' + \dots + \lambda_p \underline{e}_p \underline{e}_p' \quad (2.96)$$

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} \cdot \underline{e}_1 & & & \\ & \sqrt{\lambda_2} \cdot \underline{e}_2 & & \\ & & \dots & \\ & & & \sqrt{\lambda_p} \cdot \underline{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \cdot \underline{e}_1 \\ \sqrt{\lambda_2} \cdot \underline{e}_2 \\ \vdots \\ \sqrt{\lambda_p} \cdot \underline{e}_p \end{bmatrix} \quad (2.97)$$

o qual ajusta a estrutura de covariância prescrita para a análise do modelo fatorial, tendo m fatores ($m=p$) e a variância específica $\Psi_i = 0$ para todo $i = 1, 2, \dots, p$. A matriz de carga tem a j -ésima coluna dada por $\sqrt{\lambda_j} \cdot \underline{e}_j$. Desta maneira pode-se escrever,

$$\underset{(p \times p)}{\Sigma} = \underset{(p \times p)}{\underline{L}} \cdot \underset{(p \times p)}{\underline{L}'} + \underset{(p \times p)}{\underline{0}} = \underline{L} \underline{L}' \quad (2.98)$$

entretanto nesse caso a representação de Σ é exata, isto é, o número de fatores comuns é igual ao número de variáveis. Na prática deve-se optar por modelos que expliquem a estrutura da covariância usando poucos fatores comuns. Para isso, eliminam-se os últimos $(p-m)$ autovalores, que pouco contribuem para Σ . Negligenciando esta contribuição, obtém a aproximação.

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} \cdot \underline{e}_1 & & & \\ & \sqrt{\lambda_2} \cdot \underline{e}_2 & & \\ & & \dots & \\ & & & \sqrt{\lambda_m} \cdot \underline{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \cdot \underline{e}_1 \\ \sqrt{\lambda_2} \cdot \underline{e}_2 \\ \vdots \\ \sqrt{\lambda_m} \cdot \underline{e}_m \end{bmatrix} = \underset{p \times m \times p}{\underline{L}} \underline{L}' \quad (2.99)$$

A representação anterior considera que os fatores específicos $\underline{\varepsilon}$ definidos no modelo fatorial ortogonal são de menor importância e podem ser ignorados na fatoração de $\underline{\Sigma}$. Se os fatores específicos forem incluídos no modelo, a contribuição deles pode ser estimada tomando-se os elementos da diagonal de $\underline{\Sigma} - \underline{LL}'$, onde \underline{LL}' é dado em (2.100).

Considerando-se os fatores específicos, encontramos a aproximação

$$\underline{\Sigma} = \underline{LL}' + \underline{\Psi} \quad (2.100)$$

$$\underline{\Sigma} = \begin{bmatrix} \sqrt{\lambda_1} \cdot \underline{e}_1 & & & \\ & \sqrt{\lambda_2} \cdot \underline{e}_2 & & \\ & & \dots & \\ & & & \sqrt{\lambda_m} \cdot \underline{e}_m \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix} \quad (2.101)$$

onde $\psi_i = \sigma_i^2 - \sum_{j=1}^m \ell_{ij}^2$, com $i = 1, 2, \dots, p$.

A representação (2.102) quando aplicada a matriz de covariância amostral \underline{S} ou a matriz de correlação amostral \underline{R} é conhecida como a solução por componentes principais do modelo fatorial.

2.5.3 Método das Componentes Principais Via Matriz de Correlação Amostral

Realizando a decomposição espectral da matriz de correlação \underline{R} , obtêm os pares de autovalores autovetores $(\lambda_i, \underline{e}_i)$, onde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Sejam $m < p$ o número de fatores comuns. Então a matriz estimada das cargas fatoriais $\hat{\ell}_{ij}$ é dada por:

$$\hat{\underline{L}} = \begin{bmatrix} \sqrt{\lambda_1} \cdot \underline{e}_1 & & & \\ & \sqrt{\lambda_2} \cdot \underline{e}_2 & & \\ & & \dots & \\ & & & \sqrt{\lambda_m} \cdot \underline{e}_m \end{bmatrix} \quad (2.102)$$

As variâncias específicas estimadas são os elementos da matriz diagonal $\underline{\underline{S}} = \hat{\underline{L}}\hat{\underline{L}}'$, então

$$\hat{\Psi} = \begin{bmatrix} \hat{\psi}_1 & 0 & \dots & 0 \\ 0 & \hat{\psi}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\psi}_p \end{bmatrix}, \text{ com } \psi_i = s_{ii}^2 - \sum_{j=1}^m \hat{\ell}_{ij}^2 \quad (2.103)$$

e as comunalidades são estimadas por:

$$\hat{h}_i^2 = \hat{\ell}_{i1}^2 + \dots + \hat{\ell}_{im}^2 \quad (2.104)$$

A matriz residual pode ser obtida pela aproximação de S pela solução por componentes principais, sendo calculada por:

$$S - (\hat{L}\hat{L}' + \hat{\Psi}) \quad (2.105)$$

onde os elementos da diagonal principal são zero, e os outros elementos também são pequenos, quando escolhe-se o número de fatores m apropriado. Analiticamente, tem-se a soma dos quadrados das entradas de

$$S - (\hat{L}\hat{L}' + \hat{\Psi}) \leq \hat{\lambda}_{m+1}^2 + \dots + \hat{\lambda}_p^2 \quad (2.106)$$

Consequentemente, um pequeno valor para o somatório dos quadrados dos autovalores pode ser negligenciado, implicando um pequeno valor para a soma dos quadrados dos erros da aproximação.

Geralmente a contribuição dos primeiros fatores comuns para a variância amostral das variáveis pode ser grandes. A contribuição para a variância amostral s_{ii} do primeiro fator comum é ℓ_{i1} , $i=1,2,\dots,p$. Para o primeiro fator, a soma dos quadrados das cargas fatoriais é:

$$\hat{\ell}_{11}^2 + \hat{\ell}_{21}^2 + \dots + \hat{\ell}_{p1}^2 = \left(\sqrt{\hat{\lambda}_1} \hat{e}_1 \right)' \left(\sqrt{\hat{\lambda}_1} \hat{e}_1 \right) = \hat{\lambda}_1 \quad (2.107)$$

Desde que \hat{e}_1 tenha tamanho unitário. Em geral a proporção da variância total amostral no j-ésimo fator é dada por:

$$\left\{ \begin{array}{ll} \frac{\hat{\lambda}_j}{s_{11} + s_{22} + \dots + s_{pp}} & \text{para análise fatorial de S} \\ \frac{\hat{\lambda}_j}{p} & \text{para análise fatorial de R} \end{array} \right. \quad (2.108)$$

O critério dado (anterior) geralmente é usado para determinar o número de fatores comuns apropriado. Quanto maior o número de fatores, maior é a proporção

da variância amostral explicada pelo modelo. Existem algumas maneiras de determinar o número de fatores adequados, pode-se destacar o critério de Kaiser e o Scree-plot.

2.5.4 Rotação Ortogonal de Fatores

Geralmente a interpretação dos fatores obtidos na Análise Fatorial pode não ser uma tarefa fácil, devido a ocorrência de coeficientes \hat{l}_{ij} de grandeza numérica similar e não desprezível em fatores distintos. Quando isto ocorre, utiliza-se uma transformação ortogonal dos fatores originais, na tentativa de encontrar uma estrutura mais simples para ser interpretada. Para isso, utiliza-se uma matriz ortogonal T que pode ser obtida por diferentes critérios, sendo que após a rotação dos fatores, a estrutura de variância é preservada (RENCHE, 2002). Seguem adiante três critérios de rotação: varimax, orthomax e quartimax.

2.5.4.1 Critério varimax

De acordo com Mingoti (2005, apud SOUZA, 2012), o critério de rotação varimax é um método proposto por KAISER (1958), cujo objetivo principal é a busca da matriz T, tendo como base encontrar para um determinado fator fixo, um grupo de variáveis Z_i altamente correlacionadas com o fator e outro grupo que tenha correlação desprezível ou moderada com o fator. Para cada fator fixo, a solução é obtida através da maximização da variação dos quadrados das cargas fatoriais originais das colunas da matriz \hat{L} .

Seja,

$$V = \frac{1}{p} \sum_{j=1}^m \left[\sum_{i=1}^p \tilde{l}_{ij}^4 - \frac{1}{p} \left(\sum_{i=1}^p \tilde{l}_{ij}^2 \right)^2 \right] \quad (2.109)$$

onde $\tilde{l}_{ij} = \frac{\hat{l}_{ij}}{\hat{h}_i}$, sendo \hat{h}_i a raiz quadrada da comunalidade da variável $Z_i, i=1,2,\dots,p$. O critério varimax seleciona os coeficientes \tilde{l}_{ij} que maximizam V. Os coeficientes finais

da matriz transformada são obtidos por: $\hat{L}^* = \hat{L}T$. O critério varimax é um dos mais utilizados na prática e, em geral, produz resultados mais simples que outros métodos.

2.5.4.2 Critério quartimax

Esse método busca a simplificação das linhas de uma matriz de cargas fatoriais. Segundo HAIR *et al.* (2009) o critério quartimax de rotação deixa muitas variáveis com cargas fatoriais altas em um único fator, não sendo bem sucedido na produção de estruturas mais simples.

Seja V_Q a quantidade definida por:

$$V_Q = \frac{1}{pm} \left[\sum_{j=1}^m \sum_{i=1}^p \hat{l}_{ij}^{*4} - \frac{1}{pm} \left(\sum_{j=1}^m \sum_{i=1}^p \hat{l}_{ij}^{*2} \right)^2 \right] \quad (2.110)$$

onde \hat{l}_{ij}^* é o coeficiente da i -ésima variável no j -ésimo fator após a rotação. O critério quartimax seleciona os coeficientes \hat{l}_{ij}^* que maximizam a quantidade V_Q (MINGOTI, 2005 apud SOUZA, 2012).

2.5.4.3 Critério orthomax

É considerado uma média ponderada dos métodos varimax e quartimax (MINGOTI, 2005 apud SOUZA, 2012). Considerando

$$V_M = \sum_{j=1}^m \left[\sum_{i=1}^p \hat{l}_{ij}^{*4} - \frac{\gamma}{p} \left(\sum_{i=1}^p \hat{l}_{ij}^{*2} \right)^2 \right] \quad (2.111)$$

onde \hat{l}_{ij}^* é o coeficiente da i -ésima variável no j -ésimo fator após a rotação, $0 \leq \gamma \leq 1$.

O critério orthomax seleciona os coeficientes \hat{l}_{ij}^* que maximizam a quantidade V_M .

Para $\gamma = 1$, este critério reproduz o critério varimax sem o escalonamento das cargas fatoriais pelas comunalidades e, para $\gamma = 0$, ele se reduz ao método quartimax.

2.5.5 Determinação dos Escores Fatoriais

Após a identificação e interpretação dos fatores $L_j, j=1,2,\dots,m$ relacionados com as variáveis $Z_i, i=1,2,\dots,p$ (i -ésima variável padronizada pela média e desvio padrão), é necessário calcular os escores para cada elemento amostral, de modo a utilizar esses valores para outras análises de interesse. Para cada elemento amostral $k, k=1,2,\dots,n$, o seu escore no fator L_j é calculado por:

$$\hat{F}_{jk} = w_{j1}Z_{1k} + w_{j2}Z_{2k} + \dots + w_{jp}Z_{pk} \quad (2.112)$$

onde Z_{ik} , são valores observados das variáveis padronizadas Z_i para o k -ésimo elemento amostral e $w_{ji}, i=1,2,\dots,p$, são os pesos de ponderação de cada variável Z_i no fator L_j . De acordo com Mingoti (2005, apud SOUZA, 2012), existem três métodos para a determinação dos coeficientes w_{ji} :

- ✓ método dos mínimos quadrados ponderados;
- ✓ método de regressão;
- ✓ método ad hoc.

Neste trabalho foram testados os três métodos, porém foi implementado apenas o método dos mínimos quadrados ponderados, uma vez que este apresentou um melhor resultado.

2.5.5.1 Método dos mínimos quadrados ponderados

O modelo fatorial em (2.83) assemelha-se a um modelo de regressão linear múltipla. Desta maneira pode-se estimar o vetor F através do método dos mínimos quadrados ponderados usando-se as matrizes \hat{L} e $\hat{\Psi}$ estimadas pelo método da máxima verossimilhança. Sendo assim, \hat{F}_j pode ser obtido por:

$$\hat{F}_{jk} = (\hat{L}'\hat{\Psi}^{-1}\hat{L})^{-1}\hat{L}'\hat{\Psi}^{-1}Z_k = WZ_k \quad (2.113)$$

onde $Z_k = (Z_{1k} \quad Z_{2k} \quad \dots \quad Z_{pk})'$ é o vetor de observações do k -ésimo elemento amostral padronizado e $W_{m \times p}$ é a matriz de ponderação que gera os coeficientes w_{ji} .

Neste trabalho, as matrizes \hat{L} e $\hat{\Psi}$ foram estimadas pelo método das componentes principais. Neste caso, Johnson & Wichern (1998, apud BARROS, 2018) sugere que a estimativa dos escores seja feita usando o método dos mínimos quadrados ordinários, desde que os resíduos ε tenham a mesma variância, o que nem sempre ocorre (MINGOTI, 2005 apud SOUZA, 2012).

Desta forma, \hat{F}_j é estimado por:

$$\hat{F}_{jk} = (\hat{L}'\hat{L})^{-1}\hat{L}'Z_k = WZ_k \quad (2.114)$$

Os escores fatoriais finais ponderados servem para criar um indicador a partir de todos os escores fatoriais obtidos na Análise Fatorial. De acordo com Furtado et al. (2003) os escores fatoriais finais ponderados podem ser calculados da seguinte maneira:

$$\hat{E} = \frac{\sum_{i=1}^m \hat{\lambda}_i \cdot \hat{F}_i}{\sum_{i=1}^m \hat{\lambda}_i} \quad (2.115)$$

onde \hat{F}_i são os escores fatoriais do i-ésimo e m é o número de fatores extraídos.

2.6 Esterilização

Esterilização é o processo físico ou químico que destrói ou inativa todas as formas de vida presentes em um determinado material, especialmente microrganismos incluindo bactérias, fungos, tanto em suas formas vegetativas como esporuladas e vírus (SCHMIDELL *et al.*, 2001).

Através da realização do processo de esterilização, de acordo com LOUSADA (2018), objetiva-se inativar ou destruir todas as formas de vida, que possam desenvolver-se durante as etapas de conservação e utilização de produtos, sejam estes alimentícios, farmacêuticos ou médico-hospitalares. Os microrganismos perdem a sua capacidade de multiplicação de forma irreversível, quando expostos ao calor. Estes então são considerados inativos (FOGOLARI *et al.*, 2012). Em concordância, a propagação dos microrganismos conforme comenta TIPPLE *et al.* (2011), pode ser paralisada através do processo intitulado de esterilização.

Segundo SARAIVA (2016), existem diferentes métodos de esterilização; por calor úmido, por calor seco, por radiação, esterilização química, dentre outros. Um meio efetivo para o processo de esterilização dá-se através do calor úmido gerado pela autoclave (SILVA *et al.*, 2017). Por permitir o processo de esterilização de materiais que suportem temperaturas entre 121°C e 134°C, as denominadas centrais médicas hospitalares (CME's), utilizam geralmente, equipamentos do tipo autoclaves a vapor, tanto pelo custo em relação aos outros diferentes métodos de esterilização, quanto pelo reduzido tempo de exposição do artigo no processo (TAPIAS *et al.*, 2017).

Para RODRIGUES (2019), a esterilização de instrumentos médicos ou laboratoriais provém da capacidade da autoclave em aquecer soluções, e esta, quando ligada a área da medicina, tornar-se-á vista como um dos dispositivos mais importante utilizado.

O processo de esterilização por meio de calor úmido é baseado no uso de vapor acima de 100°C, geralmente em uma faixa de temperatura de 121°C a 134°C (ROBLES, 2014), porém de acordo com LUQUETA (2005), as etapas de aquecimento e resfriamento em um ciclo de esterilização com temperaturas na ordem de 90°C, já produzem algum efeito letal.

Com relação ao monitoramento e validação do ciclo de esterilização, tais fases poderão ser realizadas fazendo-se o uso de indicadores físicos, químicos e biológicos (MOREIRA, SILVA, 2017).

O monitoramento do processo de esterilização atrelado ao uso de técnicas estatísticas pode contribuir para a assertividade do resultado relacionado a eficácia da esterilização. Os materiais e equipamentos odontológicos considerados críticos devem ser submetidos ao processo de esterilização e, posteriormente a este, porém não menos importante, encontra-se o processo de monitorização da esterilização, determinante no resultado da efetividade de tal processo (DE OLIVEIRA CHAVES *et al.*, 2016). Logo, é de grande importância a ampla monitorização do processo de esterilização como um todo, pois de acordo com QUINTINO (2017), teoricamente não existe a esterilidade absoluta de um artigo.

Para PENNA *et al.* (2002), o bioindicador é recomendado para ser utilizado como rotina para monitorar um conjunto de esterilização ou desinfecção do ciclo e, periodicamente, revalidar a data de ciclos previamente documentados. Conforme enunciado pela ISO 11138-1:2017 (2017), tradicionalmente, o organismo de teste de indicadores biológicos (IB's) tem sido esporos bacterianos, geralmente derivados de

espécies de *Bacillus* ou *Geobacillus*. Sendo que, no processo de esterilização à vapor, os esporos de *Geobacillus stearothermophilus* são comumente utilizados, devido à resistência que este organismo apresenta frente a este processo (ROBLES, 2014). Em especial, o *Geobacillus stearothermophilus* ATCC 7953 é um microrganismo termófilo estrito, cresce a temperaturas maiores que 55°C e não cresce a 37°C, cresce melhor em pH neutro e por isso pode ser empregado como bioindicadores em processo de esterilização (PILAR, 2005).

O indicador biológico é um parâmetro escolhido para certificar-se de que o nível de esterilidade estabelecido para o produto é alcançado, conferindo a certeza de esterilidade frente à margem de segurança mínima definida de apenas 1 unidade contaminada em 10⁶ unidades do produto processado (NIEHEUS, 2004).

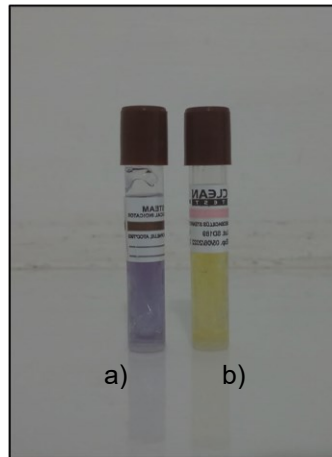
O indicador biológico contém uma população de microrganismos em forma de esporos impregnados em uma tira de papel que é acondicionada juntamente com uma ampola de vidro contendo substrato (meio de cultura) em um frasco termoplástico, o qual servirá como frasco de cultura. O posicionamento do indicador biológico em locais estratégicos no interior da autoclave, como próximo ao filtro, no centro da autoclave e próximo a porta, é de suma importância, visto que, estes são considerados pontos críticos para a ação do agente esterilizante diante da testagem da eficácia do processo (VIER-PELISSER *et al.*, 2008). Após o término do processo de esterilização, os indicadores são ativados e incubados em incubadora com temperatura ótima (específica a cada processo de esterilização) de crescimento para os esporos em questão por um determinado período de tempo (dependendo da tecnologia do bioindicador, 48 horas, 3 horas etc.). A alteração de cor diferente da inicial indica atividade de crescimento microbiano no meio. Com relação a periodicidade do teste, este segundo VIER-PELISSER *et al.* (2008), deve ser semanal e os seus respectivos resultados deverão ser registrados e mantidos em arquivos de controle.

Diferentes são os tipos de leitura utilizados para verificação da eficácia do processo de esterilização através do uso de indicadores biológicos, sendo estas diretamente ligadas ao tipo e classificação por geração em que se encontram os bioindicadores.

Os indicadores conhecidos como IB's de 1º geração, compreendem tiras ou discos de papel impregnados com esporos, acondicionados a um carreador do tipo envelope de papel ou ampola, com leitura final após o período de 7 dias de incubação (MARTINHO, 2007).

Os indicadores biológicos autocontidos denominados IB's de 2º geração conforme MARTINHO (2007), são àqueles que apresentam resposta de leitura final após 48 horas a contar do início do processo de incubação, sendo a resposta final obtida pela visualização da mudança de cor (leitura visual), decorrente do deslocamento do pH do meio em que se encontra. Através da atividade bioquímica do microrganismo ocorre a produção de derivados ácidos, provocando a modificação de cor do meio de cultura de roxo a amarelo. Uma modificação visual de cor e pH indica falha no processo de esterilização a vapor (3M, 2010 apud PELISSARI *et al.*, 2011). A figura 1 representa o exemplo de indicadores biológicos autocontidos de 2º geração.

Figura 1 - Indicadores biológicos autocontidos de 2º geração (vapor), após incubação: a) resultado negativo; b) resultado positivo



Fonte: Aatoria própria (2021)

Já os indicadores biológicos de 3º geração apresentam as basicamente mesmas características dos indicadores de 2º geração, o que os diferem é o tempo de resposta referente à leitura final, que encontra-se entre 1 a 3 horas após o início do processo de incubação, bem como, o agente fluorescente presente no meio de cultura, que auxiliará na técnica de leitura por fluorescência pelo qual o indicador biológico será submetido. Conforme comenta IARIA (2001), resultados confiáveis podem ser visualizados a partir das 3 horas de incubação, através da presença de material fluorescente baseado em reações enzimáticas. A detecção da fluorescência acontece a partir do rompimento de um agente fluorescente presente no meio de cultivo através da reação enzimática da α -glucosidade, enzima a qual, encontra-se nos esporos de *Geobacillus stearothermophilus* (LOMELÍ *et al.*, 2016). Erros ligados a falha mecânica ou humana nos processos de esterilização devem ser evitados

segundo LOMELÍ *et al.* (2016), e o treinamento dos colaboradores diretamente ligados a este processo torna-se um importante aliado.

A importância em se avaliar as diferentes técnicas de leitura frente aos resultados dos indicadores biológicos diretamente ligados a eficácia do processo de esterilização a vapor torna-se imprescindível e importante no intuito de auxiliar possíveis caminhos a serem percorridos.

A utilização do pacote desafio em combinação com o indicador biológico e o indicador químico, deveria ocorrer diariamente e no primeiro ciclo do dia, com a finalidade de avaliar o desempenho do sistema de remoção de ar da autoclave, quando assistida por bomba a vácuo (MIZUTA *et al.*, 2020). A sua utilização, torna-se ainda ideal para resultados imediatos quando em situações de emergência, pois utiliza-se do resultado do integrador apoiado pelo resultado do indicador biológico 3 horas, nele contido. A figura 2 representa um modelo de pacote desafio Clean-Test IB 3 horas.

Figura 2 - Modelo de pacote desafio



Fonte: Autoria própria (2021)

De acordo com SANTOS (2009), o envolvimento da cinética de morte microbiana está diretamente ligado a determinação do método de esterilização terminal adotado ao valor de redução decimal, mais comumente conhecido como valor D.

A destruição dos microrganismos durante um determinado tratamento térmico denomina-se coeficiente de letalidade térmica, também intitulado frequentemente como tempo de redução decimal (valor D), definido como sendo o tempo de exposição

necessário, a uma determinada temperatura de referência, para inativação de 90% da população de um determinado microrganismo (ORDÓÑEZ, 2005).

Um dos avanços na prática da esterilização é a constatação de que os microrganismos, quando submetidos à maioria dos ciclos de esterilização, não são eliminados de imediato, mas sim gradativamente (NIEHEUS, 2004).

Em termos de esterilização por calor, para LUQUETA (2012), a letalidade pode ser entendida como a quantidade de energia a ser transferida a um material a ponto de obtermos o nível de esterilidade requerido.

O número de ciclos logarítmicos reduzidos na população do bioindicador define o nível de esterilidade ou *Sterility Assurance Level* (SAL) do produto segundo SANTOS (2009), sendo esterilidade a ausência de microrganismos viáveis (BRASIL, 2010).

A aplicação do vapor sob pressão superior à pressão atmosférica se dá através de aparelhos denominados autoclaves, que tem por intuito alcançar a denominada esterilização (SANTOS, 2014).

Antecedente ao ciclo de esterilização em autoclaves convencionais, duas relevantes fases interferem no processo: o pré-vácuo e a rampa de aquecimento. O pré-vácuo consiste na remoção do ar da câmara interna através de repetidos ciclos de pulsos de vácuo alternados com injeção de vapor, elevando a temperatura interna geralmente em 3 a 4 pulsos de vácuo. Os pulsos de vácuo, além de removerem o ar presente no sistema, também têm a finalidade de aquecer a carga antes da injeção definitiva de vapor (RODRIGUES, 2010).

A rampa de aquecimento é o momento em que o vapor é injetado até que seja alcançada a temperatura de exposição de 121°C ou 134°C (RODRIGUES, 2010). Durante tais períodos a carga fica exposta a temperaturas que promovem uma lenta esterilização, antes mesmo de atingir a temperatura selecionada para o processamento do material.

Para SARAIVA (2016), a etapa de esterilização inicia quando a válvula de purga é fechada e a temperatura de esterilização previamente selecionada é alcançada.

A rampa de arrefecimento tem início quando o processo de esterilização finaliza, então a resistência de aquecimento deixa de funcionar, portanto deixa de produzir vapor, logo a pressão e temperatura da câmara começam a diminuir gradualmente (SARAIVA, 2016).

O processo de secagem do material já esterilizado se dá através de uma resistência externa a qual tem por finalidade pré-aquecer, bem como, manter a temperatura interna das paredes da câmara aquecida durante tal processo (NIEHEUS, 2004).

O processo de esterilização atrelado ao monitoramento do processo e acompanhado das diretrizes normativas através de técnicas corretamente utilizadas, proporcionam eficácia e segurança na destruição dos microrganismos presentes no meio (MACAGNANI, 2012).

3 MATERIAL E MÉTODOS

3.1 Material

Para realização dos ciclos de esterilização, utilizou-se uma autoclave da marca BAUMER, modelo HI SPEED II Standard 100-500 litros, com pré-vácuo. Trata-se de um equipamento que utiliza pulsos de vácuo durante o ciclo de esterilização. Utilizou-se também pacote desafio Clean-Test IB 3 horas, modelo PCD indicador biológico 3 horas, a fim de verificar se os parâmetros de esterilização foram alcançados. O Quadro 1 apresenta as informações dos indicadores biológicos auto-contido 3 horas utilizados no pacote teste-desafio.

Quadro 1 - Parâmetros do pacote desafio e do IB 3h utilizado

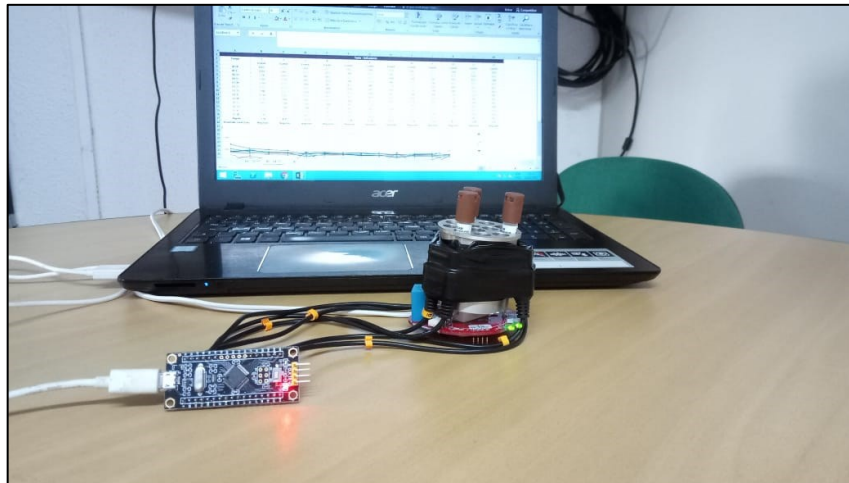
Lote do Teste Desafio	Lote do IB	Tempo de Morte
BQ2021-09	2021-09	18,2 minutos

Fonte: A autoria própria (2021)

Para realização da etapa de incubação dos indicadores com resposta final até 3 horas, utilizou-se um sistema de aquisição de dados seriais através do computador adjunto a um conversor analógico-digital (ADC), para a conversão da corrente elétrica gerada pelos sensores fotoelétricos em dados representados por valores digitais. Este sistema acoplado a uma cuba de aquecimento, auxiliou na realização de forma automática a leitura dos indicadores biológicos utilizando-se como base de resposta o método de detecção por fluorescência (agente fluorescente). A figura 3 representa o modelo do sistema computacional de leitura.

Para realização da etapa de incubação dos indicadores até a resposta final na 48ª hora, utilizaram-se incubadoras da marca CLEAN UP, modelo MINI CLEAN 12 CAVIDADES. Tais incubadoras são ditas convencionais, pois o processo de leitura dos indicadores biológicos é realizado manualmente pelo operador, tendo como base a resposta visual através do método colorimétrico (diferenciação de coloração de púrpura para amarelo).

Figura 3 - Sistema computacional de leitura



Fonte: Autoria própria (2021)

A proposta de um método de reconhecimento de padrão aplicado a ciclos falhos no processo de esterilização a vapor, deu-se a partir da aplicação *in loco* da experimentação, correlacionada a execução de ciclos completos e falhos no processo de esterilização a vapor. A coleta dos dados e a identificação condizente do modelo para reconhecimento de padrão, aconteceu através do uso da metodologia de análise fatorial, método de regressão logística e método do escore quadrático, tratando-se de um estudo com premissa indutiva, de natureza aplicada, abordagem quantitativa e de caráter exploratório.

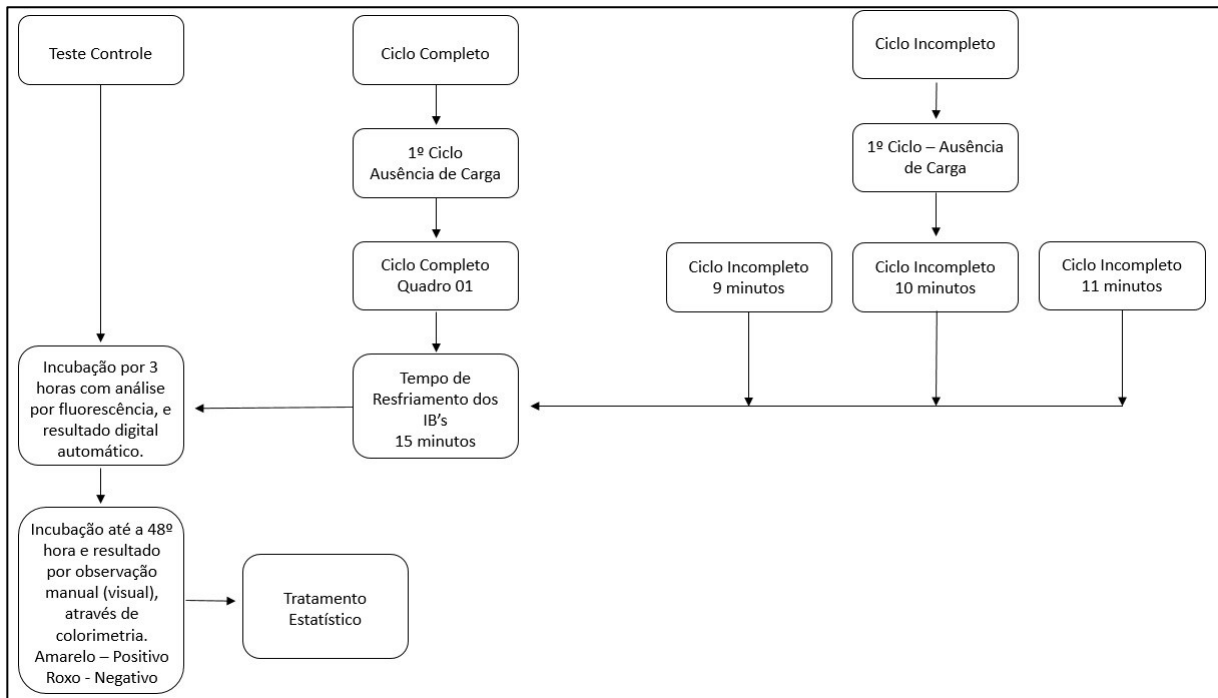
3.2 Métodos

3.2.1 Ciclos de Esterilização e Coleta de Dados

Para garantir a aplicabilidade da experimentação, tomou-se como base uma adaptação conceituada do documento *Food and Drug Administration Biological indicators (BI) Premarket Notification [510(k)] Submission*, que descreve a metodologia de validação da redução do tempo de incubação para indicadores biológicos, bem como, o uso da aplicação dos métodos de análise fatorial, análise de regressão logística e escore quadrático no item 3.2.3 desta seção.

A figura 4 apresenta uma representação esquemática da metodologia empregada.

Figura 4 - Representação esquemática da metodologia



Fonte: Autoria própria (2021)

Para corroborar o resultado do teste controle e dos ciclos de esterilização completo e incompleto (falho), foram utilizadas unidades de pacotes desafio contendo em seu interior indicador biológico auto-contido 3 horas, conforme apresentado no Quadro 1.

3.2.2 Execução de Ciclos de Esterilização

Para o ciclo denominado controle não foi realizado o processo de esterilização, já que sua característica principal é a de validar o processo de esterilização a partir da conformidade com o atingimento dos parâmetros disponíveis, sendo eles, substrato, microrganismo e temperatura ótima.

Para a execução dos ciclos de esterilização, adotou-se obrigatoriamente a ausência de carga no primeiro ciclo, objetivando o pré-aquecimento do equipamento (autoclave) a ser posteriormente utilizado, condicionando assim, o ambiente do mesmo e garantindo as mesmas condições iniciais em todos ciclos de esterilização com os testes realizados, sempre respeitando as condições apontadas nas instruções de uso do teste-desafio.

Após o ciclo de condicionamento da autoclave, acomodou-se dentro desta os pacotes desafio contendo em seu interior os indicadores biológicos auto-contido 3 horas, conforme *layout* de posicionamento apresentado no Quadro 2.

Quadro 2 - Posicionamento dos pacotes desafio contendo indicadores biológicos auto-contido 3 horas na câmara de esterilização

Posicionamento	Superior	Centro	Inferior
Frontal	1 unidade	-	-
Centro	-	1 unidade	-
Fundo	-	-	1 unidade

Fonte: Autoria própria (2021)

A seguir, realizou-se os ciclos de esterilização com os pacotes desafio, utilizando-se para o ciclo completo o tempo de morte apresentado no Quadro 1, e para o ciclo incompleto (falho) os tempos de esterilização demonstrados no Quadro 3, objetivando simular a pior condição na qual todos os parâmetros de esterilização são atendidos, exceto o parâmetro de tempo (ciclo parcial em que 30-80% dos IB's sobrevivem), permitindo assim, a identificação de falhas no processo de esterilização.

Quadro 3 - Tempos críticos em minutos para ciclo incompleto (falho)

Tempo de Morte I1	Tempo de Morte I2	Tempo de Morte I3
9 minutos	10 minutos	11 minutos

Fonte: Autoria própria (2021)

Terminado o ciclo de esterilização, os pacotes desafio foram retirados da autoclave e acondicionados sobre uma bancada e então, após 15 minutos para o resfriamento dos IB's 3h, realizou-se a ativação e a incubação dos IB's 3h, em sistema de incubação pré-aquecido, para posterior leitura dos dados referente ao tempo final de resposta em 3 horas, sempre seguindo as instruções de uso do fabricante.

Ao longo do período das 3 horas de incubação e a cada intervalo de tempo de 15 minutos, coletou-se e registrou-se os dados referente aos valores encontrados através do uso do sistema computacional apresentado na figura 3. Ao término das 3

horas do processo de incubação, obtiveram-se os valores referenciando cada indicador biológico incubado.

Os mesmos indicadores biológicos auto-contido 3 horas incubados na etapa anterior, foram então incubados em incubadora convencional (marca CLEAN UP, modelo MINI CLEAN 12 CAVIDADES) pré-aquecida, e pelo período total de 48 horas. Verificou-se a resposta final após a incubação através da leitura visual de cor pelo uso do método colorimétrico, relacionando-se a coloração amarelo para indicadores biológicos com resultado positivo e púrpura para indicadores biológicos com resultado negativo, conforme demonstrado na figura 5.

Ao final do período de incubação, registrou-se os resultados encontrados para posterior utilização e análise, utilizando-se o formulário de anotação dos resultados obtidos, conforme figura 6.

Figura 5 - Leitura visual de cor dos IB's pelo método colorimétrico



Fonte: Autoria própria (2021)

Figura 6 - Formulário de anotação dos resultados obtidos

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Teste - Indicadores												
2	Tempo	1	2	3	4	5	6	7	8	9	10	11	12
3	Valor	Valor	Valor	Valor	Valor	Valor	Valor	Valor	Valor	Valor	Valor	Valor	Valor
4	(Tempo 1) - 00:00:00	1049	1162	1114	1264	1310	1357	1402	1390	1430	1238	1195	1209
5	(Tempo 2) - 00:15:00	978	1096	1124	1142	1205	1103	1204	1189	1168	1145	1161	1160
6	(Tempo 3) - 00:30:00	990	1074	1108	1117	1188	1076	1215	1176	1176	1146	1140	1108
7	(Tempo 4) - 00:45:00	973	1058	1113	1129	1197	1052	1184	1155	1181	1119	1129	1132
8	(Tempo 5) - 01:00:00	1016	1091	1126	1127	1199	1072	1196	1171	1183	1141	1127	1133
9	(Tempo 6) - 01:15:00	984	1075	1115	1120	1188	1077	1150	1142	1155	1090	1111	1120
10	(Tempo 7) - 01:30:00	981	1107	1125	1125	1198	1059	1194	1157	1150	1082	1113	1130
11	(Tempo 8) - 01:45:00	977	1091	1124	1121	1173	1041	1168	1129	1153	1076	1100	1114
12	(Tempo 9) - 02:00:00	984	1086	1110	1119	1186	1067	1201	1153	1171	1097	1082	1134
13	(Tempo 10) - 02:15:00	996	1088	1116	1125	1195	1055	1201	1189	1177	1125	1093	1137
14	(Tempo 11) - 02:30:00	1014	1107	1113	1130	1199	1061	1203	1174	1178	1135	1093	1120
15	(Tempo 12) - 02:45:00	1005	1100	1119	1130	1209	1077	1199	1192	1209	1134	1109	1135
16	(Tempo 13) - 03:00:00	1029	1107	1132	1149	1215	1073	1217	1199	1216	1139	1102	1163
17	(Tempo 14) - Ângulo (Completo)	9.16	8.84	1.00	1.14	1.92	-6.34	1.02	3.06	6.78	-4.92	-19.35	1.47
18	Resultado Final 48h (Cor)	Negativo	Negativo	Positivo	Negativo	Negativo	Negativo	Negativo	Positivo	Negativo	Positivo	Negativo	Negativo
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
31													

Fonte: Autoria própria (2021)

3.2.3 Identificação do Modelo para Reconhecimento de Padrão a Partir do Método de Análise Fatorial, Análise de Regressão Logística e Método Escore Quadrático.

A metodologia aqui descrita teve por objetivo a realização da separação dos grupos de indicadores biológicos com resposta final após a incubação pela leitura visual de cor pelo uso do método colorimétrico em 48 horas.

A etapa seguinte à obtenção dos dados referente a separação dos grupos de indicadores biológicos, tratou e aplicou os métodos de análise fatorial, regressão logística e escore quadrático com o intuito de reconhecer e classificar os padrões acerca das respostas obtidas.

Utilizou-se a Análise Fatorial a fim de extrair os fatores que explicam a correlação existente entre as variáveis originais do banco de dados. Esses fatores agrupam as variáveis iniciais que contenham forte correlação, desta forma, as variáveis originais agrupadas num mesmo fator possuem, portanto, alta correlação, enquanto que variáveis de fatores distintos possuem baixa correlação.

De posse dos autovalores calculados, utilizou-se o critério de Kaiser para a seleção dos fatores. A seguir, com base na rotação Varimax e calculadas as communalidades para cada variável foram calculados os escores fatoriais das observações, os quais foram utilizados como a nova base de dados para os modelos de reconhecimento e classificação de padrões.

Para a regressão logística, associou-se às variáveis com os tempos do ciclo e respectivos ângulos, utilizando-se como variável de resposta binária: 0 para indicadores biológicos com resultado negativo (coloração púrpura); 1 para indicadores

biológicos com resultado positivo (coloração amarelo). Entende-se, segundo LOPES et al. (2017), em relação ao modelo de regressão logística, que este modelo atua no sentido de averiguar os elementos e a ocorrência de eventos, ambos, relacionados às variáveis, incorporando efeitos não lineares e utilizando-se de testes estatísticos diretos para análise.

Para o escore quadrático associou-se as variáveis com os tempos do ciclo, testando-se após a normalidade (Gaussianidade) dos dados, havendo a necessidade prévia de uma pré-classificação a fim de organizar os grupos, devido a resposta do ciclo.

Determinou-se a identificação do modelo de reconhecimento de padrão, utilizando-se os dados advindos das etapas de incubação até 3 horas e incubação convencional com leitura final na 48^a hora, conforme apresentado no Quadro 04.

Através da aplicabilidade de métodos não supervisionados de reconhecimento de padrões, empregados a fim de avaliar de maneira multivariada o conjunto de dados de uma abordagem, representou-se de forma simples em quadros e tabelas os resultados apresentados na secção subsequente. Qualquer pacote computacional para compilação dos dados pode ser utilizado para essa finalidade, sendo que no presente trabalho optou-se pelo Statgraphics Centurion ®.

4 ANÁLISE DOS RESULTADOS

A espelho da aplicação das metodologias trabalhadas por SOUZA (2008) e SOUZA (2012) para análise e discussão dos resultados, e seguindo tal linha de raciocínio, propôs-se então, utilizar os métodos de Análise Fatorial, Regressão Logística e Escore Quadrático para realização da análise de eficácia do modelo referente à problemática apontada, no sentido de predizer se o resultado da leitura no período de tempo de até 3 horas de incubação, frente ao método convencionalmente já existente e aplicado no mercado é considerado assertivo.

Primeiramente, iniciou-se a aplicação da metodologia de Análise Fatorial cujos resultados estão evidenciados na tabela 2.

Tabela 2 - Carregamentos Fatoriais rotacionados do banco de dados de teste

Variáveis	Fator 1	Fator 2	Comunalidades
Variável 1	0.955094	0.0007581	0.912204
Variável 2	0.974138	-0.0250571	0.949574
Variável 3	0.984834	-0.0149311	0.97012
Variável 4	0.993983	0.0013241	0.988004
Variável 5	0.997621	0.031447	0.996236
Variável 6	0.994332	0.069368	0.993507
Variável 7	0.984456	0.111368	0.981556
Variável 8	0.971313	0.161651	0.96958
Variável 9	0.958951	0.221274	0.968549
Variável 10	0.946791	0.284165	0.977163
Variável 11	0.937984	0.327775	0.987251
Variável 12	0.934365	0.337403	0.986878
Variável 13	0.926047	0.346348	0.977521
Variável 14	0.048473	0.980935	0.964583
Total da Variância			
Explicada (%)	88.86%	8.44%	
Total da Variância	88.86%	97.305%	
Acumulada (%)			

Fonte: Autoria Própria (2021)

Após a aplicação da Análise Fatorial, verificou-se que o primeiro fator apresenta uma variância explicada de 88,86% e o segundo fator uma variância de 8,44%. Levando-se em consideração a somatória dos mesmos, tem-se então a explicação de um total de 97,30% da variância total dos dados representada por dois fatores. Como o objetivo era validar a correlação em torno das variáveis iniciais, optou-se por utilizar todas as variáveis nos modelos de reconhecimento e classificação de padrões, utilizando-se posteriormente os métodos de regressão logística e escore quadrático, a fim de reconhecer e classificar os dados.

Neste segundo momento, aplicou-se a regressão logística cujo resultado é apresentado no Quadro 4.

Quadro 4 - Regressão Logística

Causa de Variância	Soma de Quadrados Total	Graus de Liberdade	P-valor
Modelo	78.034	14	0.0000
Resíduo	53.1645	133	1.0000
Total	131.198	147	

Fonte: Autoria Própria (2021)

Com base no quadro 4, verifica-se que o p-valor é inferior a 0,05. Diante disso, podemos afirmar que o modelo de regressão é significativo e apresenta-se através do seguinte esboço:

$$f(x) = \frac{1}{1 + \exp^{-x}}$$

onde, $x = 1.56475 - 0.00360983 * \text{Tempo}_1 - 0.100616 * \text{Tempo}_2 + 0.130348 * \text{Tempo}_3 - 0.0221683 * \text{Tempo}_4 - 0.00836139 * \text{Tempo}_5 - 0.0137101 * \text{Tempo}_6 + 0.0431199 * \text{Tempo}_7 - 0.0723112 * \text{Tempo}_8 + 0.0992873 * \text{Tempo}_9 - 0.0529055 * \text{Tempo}_{10} - 0.0404311 * \text{Tempo}_{11} + 0.0969294 * \text{Tempo}_{12} - 0.0577965 * \text{Tempo}_{13} + 0.179586 * \text{Tempo}_{14}$

O quadro 5 apresenta os resultados da Matriz de classificação e taxa de acerto real dos grupos de indicadores biológicos utilizando o método da regressão logística.

Quadro 5 - Matriz de classificação e taxa de acerto real dos grupos de indicadores biológicos utilizando a regressão logística

			Classificação Prevista (Modelo de Reconhecimento de Padrão)	
			IB's positivos	IB's negativos
			1	0
Classificação Real (Método Colorimétrico)	IB's positivos	1	24 *100%	0 0,00%
	IB's negativos	0	9 7,32%	114 *92,68%
Taxa de Acerto Real		93,87%		

Fonte: Autoria Própria (2021)

Levando-se em consideração a classificação prevista (baseada na resposta pelo modelo de reconhecimento de padrão sugerido) para a variável de resposta binária 1, para quando os indicadores biológicos com resultado positivo (coloração amarelo), estão em confronto com a classificação real (baseada na resposta pelo método colorimétrico), para a variável de resposta binária 1, para quando indicadores biológicos também com resultado positivo (coloração amarelo), os valores apresentados no Quadro 5 indicam o percentual de acerto da ordem de 100%, demonstrando a efetividade e a satisfatoriedade na predição de valores relacionada ao modelo de reconhecimento de padrão sugerido.

Considerando a classificação prevista (baseada na resposta pelo modelo de reconhecimento de padrão sugerido) para a variável de resposta binária 0, para quando os indicadores biológicos com resultado negativo (coloração púrpura), estão em confronto com a classificação real (baseada na resposta pelo método colorimétrico), para a variável de resposta binária 0, para quando indicadores biológicos também com resultado negativo (coloração púrpura), os valores apresentados no Quadro 5 indicam o percentual de acerto da ordem de 92,68%, demonstrando uma alta assertividade na predição de valores relacionada ao modelo de reconhecimento de padrão sugerindo, porém, abaixo do objetivado para garantir a aplicabilidade da experimentação tomando-se como base a adaptação conceituada do documento Food and Drug Administration (2007), o qual sugere um valor de $\geq 97\%$ de assertividade.

Diante dos dados apresentados, sugere-se que a taxa de acerto real em relação ao método de regressão logística aplicado é da ordem de 93,87%.

No intuito de buscar uma segunda opinião acerca da avaliação da eficácia do método de reconhecimento de padrão sugerido, aplicou-se o método do escore quadrático aos dados obtidos. Os resultados observados a respeito, encontram-se descritos no Quadro 6.

Quadro 6 - Escore Quadrático

Função Variância	Lambda Wilks	Soma dos Quadrados	Graus de Liberdade	P-valor
1	0.511616	93.1552	14	0.0000

Fonte: Autoria Própria (2021)

Com base no quadro 6, verifica-se que o p-valor é inferior a 0,05, podendo-se, portanto, afirmar que o modelo de escore quadrático é significativo. Desta forma temos as funções de classificação do escore quadrático apresentadas no quadro 7.

Quadro 7 - Funções de classificação do escore quadrático

Tempos	0	1
Tempo 1	-0.00619395	0.000083574
Tempo 2	0.118198	0.0132348
Tempo 3	-0.157497	-0.0531343
Tempo 4	0.0436563	0.0405415
Tempo 5	-0.00120898	0.0748608
Tempo 6	-0.0511157	-0.148159
Tempo 7	0.124351	0.0680596
Tempo 8	-0.0391475	-0.000698345
Tempo 9	-0.0462045	0.0379979
Tempo 10	0.0291558	-0.0128066
Tempo 11	-0.0197158	-0.0480499
Tempo 12	0.0367458	0.0389617
Tempo 13	-0.0246555	-0.00619286
Tempo 14	0.0193505	0.113021
Constante	-4.8679	-6.23655

Fonte: Autoria Própria (2021)

O quadro 8 apresenta os resultados obtidos com base na Matriz de classificação e taxa de acerto utilizando escore quadrático.

Quadro 8 - Matriz de classificação e taxa de acerto utilizando escore quadrático

			Classificação Prevista (Modelo de Reconhecimento de Padrão)	
			IB's positivos	IB's negativos
			1	0
Classificação Real (Método Colorimétrico)	IB's positivos	1	18 *75,00%	6 25,00%
	IB's negativos	0	3 2,44%	120 *97,56%
Taxa de Acerto Real		93,87%		

Fonte: Autoria Própria (2021)

Levando-se em consideração a classificação prevista (baseada na resposta pelo modelo de reconhecimento de padrão sugerido) para a variável de resposta binária 1, para quando os indicadores biológicos com resultado positivo (coloração amarelo), estão em confronto com a classificação real (baseada na resposta pelo método colorimétrico), para a variável de resposta binária 1, para quando os indicadores biológicos também com resultado positivo (coloração amarelo), os valores apresentados no Quadro 8 indicam o percentual de acerto de apenas 75%, demonstrando baixa efetividade na predição de valores do modelo de reconhecimento de padrão sugerido.

Já com relação a classificação prevista (baseada na resposta pelo modelo de reconhecimento de padrão sugerido) para a variável de resposta binária 0, para quando os indicadores biológicos com resultado negativo (coloração púrpura), estão em confronto com a classificação real (baseada na resposta pelo método colorimétrico), para a variável de resposta binária 0, para quando os indicadores biológicos também com resultado negativo (coloração púrpura), os valores apresentados no Quadro 8 indicam o percentual de acerto da ordem de 97,56%, esta, considerada uma assertividade dentro do objetivado, tomando-se como base a adaptação conceituada do documento Food and Drug Administration (2007), o qual sugere um valor $\geq 97\%$ de assertividade.

Diante dos dados apresentados, sugere-se que a taxa de acerto real em relação ao método de escore quadrático aplicado é da ordem de 93,87%.

Posteriormente, confrontou-se os resultados decorrentes dos métodos de regressão logística e escore quadrático aplicados, onde, pode-se perceber um ponto em destaque e de grande importância para com a garantia da aplicabilidade da experimentação. Este ponto em destaque, demonstra uma possível característica, observada do ponto de vista da classificação real (baseada na resposta pelo método colorimétrico) para a variável de resposta binária 0, para quando os indicadores biológicos com resultado negativo (coloração púrpura), quando confrontada com a classificação prevista (baseada na resposta pelo modelo de reconhecimento de padrão sugerido) para a variável de resposta binária 1, para quando indicadores biológicos com resultado positivo (coloração amarelo), comumente conhecida no meio da biossegurança como indicadores biológicos com resultado Falso-Negativo, que nada mais vem a ser, quando, mesmo em baixos números, ainda existam microrganismos presentes no meio, os quais através do processo enzimático tornam-se incapazes de produzir ácido em escala suficiente para que ocorra o deslocamento do pH e posteriormente como consequência a não alteração da coloração do meio de púrpura para amarelo.

Para tanto, e no sentido de predizer uma possível maior assertividade de resposta do modelo de reconhecimento de padrão, resgatou-se os indicadores biológicos utilizados nos ciclos falhos e conhecidos pelo modelo de reconhecimento de padrão sugerido como pertencentes a possível variável de resposta binária 1 para quando indicadores biológicos com resultado positivo (coloração amarelo) e realizou-se o plaqueamento destas unidades de indicadores biológicos apontados como Falso-Negativos, conforme sugere a metodologia descrita na normativa ISO 11138-1:2017 (2017). O resultado obtido após o plaqueamento indicou a presença de crescimento do esporo *Geobacillus stearothermophilus* em 06 (seis) do total de 09 (nove) placas testadas, demonstrando a ocorrência de indicadores biológicos com característica para Falso-Negativos.

Após, confrontou-se a matriz de classificação e taxa de acerto real dos grupos de indicadores biológicos utilizando a regressão logística, exemplificada no Quadro 5 com a matriz de classificação e taxa de acerto real dos grupos de indicadores biológicos utilizando o método do escore quadrático, apresentada no Quadro 8 e

reorganizou-se os dados obtidos em uma nova matriz indicativa de classificação para o método de regressão logística. Os resultados estão apresentados no quadro 9.

Quadro 9 - Matriz indicativa de classificação e taxa de acerto real dos grupos de indicadores biológicos utilizando a regressão logística após o plaqueamento

			Classificação Prevista (Modelo de Reconhecimento de Padrão)	
			IB's positivos	IB's negativos
Classificação Real (Método Colorimétrico)	IB's positivos	1	1	0
		24	0	0,00%
	IB's negativos	0	3	120
		2,44%	*97,56%	
Taxa de Acerto Real		97,95%		

Fonte: Autoria Própria (2021)

Analisando os resultados referente a matriz indicativa de classificação e taxa de acerto real dos grupos de indicadores biológicos utilizando a regressão logística após o plaqueamento, o modelo de reconhecimento de padrão sugerido passa a indicar a garantia da aplicação da experimentação para casos classificados corretamente, através da taxa de acerto real da ordem de 97,95%, resultado considerado satisfatório quando comparado a porcentagem $\geq 97\%$ tomando-se como base a adaptação conceituada do documento (Food and Drug Administration, 2007). Portanto, sugere-se que o modelo de reconhecimento de padrão aplicado em processo de leitura final pelo período de 3 horas de incubação, indica superioridade frente ao processo de leitura pelo uso do método colorimétrico pelo período de 48 horas de incubação, pois, através da interpretação dos dados possibilita antecipar os resultados com rapidez e confiabilidade de 100% no que se refere a classificação prevista para quando indicadores positivos (coloração amarela) e 97,56% para quando classificação prevista para indicadores negativos (coloração púrpura).

Ao levarmos em consideração um possível cenário na tomada de decisão vinculada ao resultado de Falso-Positivo quando para condições de ciclos falhos, em apenas 2,44% das vezes o modelo de reconhecimento de padrão sugerido, demonstraria por hora, a indicação de um resultado positivo ao invés de negativo, implicando na repetição do ciclo de esterilização em questão.

5 CONCLUSÕES

A partir dos resultados obtidos conclui-se que foi possível estabelecer um modelo de reconhecimento de padrão que reconheça a eficiência de um processo de esterilização a vapor, quando submetido a ciclos falhos e em leitura final de até 3 horas com uma taxa de acerto real da ordem de 97,95%.

Destaca-se também que a metodologia empregada foi adequada ao experimento, sugerindo ainda que o modelo de reconhecimento de padrão atendeu ao objetivo proposto e, portanto, poderá ser utilizado para posteriores aplicações em escala industrial, possibilitando a indústria fabricante de indicadores biológicos ter um enfoque diferenciado em relação ao seu desempenho e capacidade competitiva em um segmento de mercado já existente, porém por ela atualmente não explorado.

Fica, então, como sugestão para trabalhos futuros a extensão do modelo de reconhecimento de padrão para aplicação na monitorização de processos de esterilização a óxido de etileno, formaldeído e peróxido de hidrogênio com leitura final em até 3 horas de incubação.

REFERÊNCIAS

- ABREU, M. N. S.; SIQUEIRA, A. L.; CAIAFFA, W. T. Regressão logística ordinal em estudos epidemiológicos. **Revista de Saúde Pública**, v. 43, p. 183-194, 2009. Disponível em: <https://doi.org/10.1590/S0034-89102009000100025>. Acesso em: 09 jun. 2021.
- AGENCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. **Farmacopéia Brasileira**. 5ª ed. v.1. 2010. Disponível em: <https://www.gov.br/anvisa/pt-br/assuntos/farmacopeia/farmacopeia-brasileira>. Acesso em: 21 jun. 2021.
- BAKKE, H. A.; de MOURA LEITE, A. S.; DA SILVA, L. B. Estatística multivariada: aplicação da análise fatorial na engenharia de produção. **Revista Gestão Industrial**, n. 4, p. 1-14, 2008. Disponível em: <https://periodicos.utfpr.edu.br/revistagi/article/view/188>. Acesso em: 09 jun. 2021.
- BARRAGAN, J. F. M.; de OLIVEIRA FONTES.; C. H., PEREIRA, O. J.; BARRETO, S. T. S.; PACHECO, L. A. Análises de métricas de similaridades em séries temporais para reconhecimento de padrões. **Cadernos do IME-Série Estatística**, v. 33, p. 35-50, 2012. Disponível em: <https://doi.org/10.12957/cadest.2012.15775>. Acesso em: 09 jun. 2021.
- BARROS, J. O. S. **Análise multivariada aplicada nas interferências das sobras de açúcares redutores residuais totais nas dornas de fermentação alcoólica**. 2018. Dissertação (Programa de Pós-Graduação em Inovações Tecnológicas) – Universidade Tecnológica Federal do Paraná, Campo Mourão. 2018. Disponível em: <http://repositorio.utfpr.edu.br/jspui/handle/1/4044>. Acesso em: 21 jun. 2021.
- BUENO, D. D. S.; TORRES, L. H.; AMORIM, R. D.; MARÇAL R. C. **Investigação no processo de esterilização por autoclave**. Trabalho de Conclusão de Curso (Curso em Engenharia Mecânica) - Centro Universitário de Anápolis - UniEvangélica, Anápolis, 2017. Disponível em: <http://repositorio.aee.edu.br/jspui/handle/aee/488>. Acesso em: 10 jun. 2021.
- CABRAL, C. I. S. **Aplicação do modelo de regressão logística num estudo de mercado**. 2013. Dissertação (Programa de Pós-Graduação em Matemática Aplicada à Economia e à Gestão) - Universidade de Lisboa, Lisboa. 2013. Disponível em: <http://hdl.handle.net/10451/10671>. Acesso em: 09 jun. 2021.
- CORREIA, P. R.; FERREIRA, M. M. Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos. **Revista Química Nova**, n. 2, v. 30, p. 481-487, 2007. Disponível em: <https://doi.org/10.1590/S0100-40422007000200042>. Acesso em: 09 jun. 2021.
- CORTELLA, M. S. **A sorte segue a coragem!**, 1ª edição: Planeta do Brasil, 2018.
- CHAVES, R. de O; SANTOS, C. A. O. dos; LACERDA, M. C.; & RIBEIRO, I. L. A. Conhecimento de graduandos em odontologia sobre o processo de esterilização e o monitoramento biológico. **Revista de Odontologia da Universidade Cidade de São Paulo**, n. 2, v. 28, p. 96-100, 2016. Disponível em: <https://publicacoes.unicid.edu.br/index.php/revistadaodontologia/article/view/218>. Acesso em: 09 jun. 2021.

FERREIRA, T. R. M. **Construção de um modelo utilizando a Análise Discriminante para a detecção de Acidemia Fetal**. 2019. Dissertação (Programa de Pós-Graduação em Informática Médica) - Universidade do Porto, Porto, 2019. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/118493/2/309263.pdf>. Acesso em: 15 maio. 2021.

FOGOLARI, O.; REIS, C. Z.; PHILIPPI, L. S. Determinação de parâmetros cinéticos da inativação térmica de *Escherichia coli* em lodo de esgoto. **Revista engenharia sanitária e ambiental**, n. 3, v. 17, p. 255-262, 2012. Disponível em: <https://doi.org/10.1590/S1413-41522012000300001>. Acesso em: 09 jun. 2021.

FOOD AND DRUG ADMINISTRATION. **Biological Indicator (BI) Premarket Notification [510 (k)] Submissions Attachment II**, Center for Devices and Radiological Health, 2007. Disponível em: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/biological-indicator-bi-premarket-notification-510k-submissions>. Acesso em: 10 jun. 2021.

FURTADO, M. E.; CHAVES NETO, A.; DOMINGUES, Z. H. & HOSOKAWA, R. T. Ranqueamento de faxinais do estado do paraná através da análise fatorial. **Revista Ciências Exatas e Naturais**, n. 1, v. 5, p. 85–99, 2003. Disponível em: <https://revistas.unicentro.br/index.php/RECEN/article/view/435>. Acesso em: 21 jun. 2021.

GUIMARÃES, A.; MOREIRA, T. B. S. Previsão de insolvência: um modelo baseado em índices contábeis com utilização da análise discriminante. **Revista de Economia Contemporânea**, n. 1, v. 12, p. 151-178, 2008. Disponível em: <https://doi.org/10.1590/S1415-98482008000100006>. Acesso em: 09 jun. 2021.

HAIR, J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAN, R. L. **Análise multivariada de dados**, 6ª edição: Bookman, 2009.

LOMELÍ, S. H.; REBOLLO, S. A.; HERNÁNDEZ, J. G.; & LUNA, M. G. F. Monitoreo con indicadores biológicos de rápida lectura de las autoclaves de CEYE de la Facultad de Odontología de la Universidad Tecnológica de México. **Revista odontológica mexicana**, n. 2, v. 20, p. 93-97, 2016. Disponível em: <http://revistas.unam.mx/index.php/rom/article/view/56053>. Acesso em: 09 jul. 2021.

IARIA, P. Comparación de resultados del proceso de esterilización con dos indicadores biológicos. **Rev Enferm IMSS**, n. 3. v. 9, p. 143-146, 2001. Disponível em: http://revistaenfermeria.imss.gob.mx/editorial/index.php/revista_enfermeria/article/view/789. Acesso em: 09 jul. 2021.

ISO 11138-1:2017. **Sterilization of health care products – Biological indicators – Part 1: General requirements**. March, 2017. Disponível em: <https://www.iso.org/standard/66442.html>. Acesso em: 21 jun. 2021.

KAISER, H. F. **The varimax criterion for analytic rotation in factor analysis**.

Psychometrika, v. 23, p. 187–200, 1958. Disponível em: http://cda.psych.uiuc.edu/psychometrika_highly_cited_articles. Acesso em: 21 jun. 2021.

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

LIMA, J. D.; MARQUES, J. M. A análise Econômico-financeira de Empresas Sob a Ótica da Estatística Multivariada. In: Encontro Nacional de Engenharia de Produção - Enegep. 12. 2002, Curitiba. **Anais [...]** Curitiba: Editora Associação Brasileira de Engenharia de Produção, 2002. p. 1-8. Disponível em: <https://www.abepro.org.br/publicacoes/index.asp?pesq=ok&ano=2002>. Acesso em: 18 jun. 2021.

LOPES, M. G.; CIRIBELI, J. P.; de OLIVEIRA MASSARDI, W.; & de ALMEIDA MENDES, W. Análise dos indicadores de inadimplência nas linhas de crédito para pessoa física: um estudo utilizando modelo de regressão logística, **Revista Estudos do CEPE**, n. 46, p. 75-90, 2017. Disponível em: <https://online.unisc.br/seer/index.php/cepe/article/view/11099>. Acesso em: 15 maio. 2021.

LOUSADA, M. E. G. **Avaliação da viabilidade e resistência térmica de esporos de Bacillus atrophaeus ATCC 9372 obtidos a partir de meio residual do cultivo de micro-organismos fotossintetizantes**. 2018. Dissertação (Programa de Pós-Graduação em Tecnologia Bioquímico-Farmacêutica) - Universidade de São Paulo, São Paulo, 2018. Disponível em: <https://www.teses.usp.br/teses/disponiveis/9/9135/tde-18122018-171240>. Acesso em: 12 jul. 2021.

LUQUETA, G. R. Esterilização por calor e a cinética de morte microbiana. **Revista Controle de Contaminação – Cinética de Morte Microbiana**, v. 77, p. 28, 2005. Disponível em: http://gerson.luqueta.com.br/index_arquivos/Page655.htm. Acesso em: 07 jul. 2021.

LUQUETA, G. R. **Avaliação de rede de sensores sem fio para liberação paramétrica da esterilização por calor**. 2012. Dissertação (Programa de Pós-Graduação em Gestão de Redes de Telecomunicações) - Pontifícia Universidade Católica de Campinas, Campinas, 2012. Disponível em: <http://tede.bibliotecadigital.puc-campinas.edu.br:8080/jspui/handle/tede/526>. Acesso em 09 jun. 2021.

MACAGNANI, C. B.; TONELLI, S. R.; Conservação da esterilidade de artigos úmidos após autoclavação e armazenamento. **Revista SOBECC**, n. 2, v. 17, p. 26-32, 2012. Disponível em: <https://revista.sobecc.org.br/sobecc/article/view/157>. Acesso em: 09 jun. 2021.

MARTINHO, M. A. V. **Eficácia dos integradores químicos x indicadores biológicos no monitoramento dos ciclos de esterilização a vapor: revisão sistemática da literatura**. 2007. Dissertação (Programa de Pós-Graduação em Enfermagem) - Universidade Federal de São Paulo, São Paulo, 2007. Disponível em: <https://www.teses.usp.br/teses/disponiveis/7/7139/tde-17052007-112405/en.php>. Acesso em: 09 jun. 2021.

MIZUTA, H. T. T.; TAKAES, R. A. T.; ANDRZEJEWSKI, E. L. S.; de MENECH, L. V.; FALCONI, F. A. Monitoramento da esterilização a vapor dos materiais do Biotério Central da Unioeste e do Abrigo São Vicente de Paulo, Cascavel, Paraná. **Brazilian**

Journal of Development, n. 7, v. 6, p. 48020-48026, 2020. Disponível em: <https://doi.org/10.34117/bjdv6n7-445>. Acesso em: 09 jun. 2021.

MOREIDA, J. R. B.; CASAS, R. A. S. El Análisis de modelos matemáticos para el comportamiento de dos micro-organismos en un proceso de fermentación. **Revista Mica**, v. 3, n. 3, p. 27-47, 2019. Disponível em: <http://mica.tecnocientifica.com.mx/index.php/mica/article/view/12>. Acesso em: 09 jun. 2021.

MOREIRA, A. C. A.; SILVA, F. L. L. Métodos de esterilização utilizados em salões de beleza de SALVADOR, BA. **Revista de ciências médicas e biológicas**, v. 16, n. 1, p. 73-78, 2017. Disponível em: <http://dx.doi.org/10.9771/cmbio.v16i1.14160>. Acesso em: 10 jun. 2021.

MULLER, S. I. M. G. **Comparação entre os métodos de máxima verossimilhança, distância mínima e o método de Fisher para o reconhecimento de padrões em imagens coloridas**. 1997. Dissertação (Programa de Pós-Graduação em Ciências Geodésicas) - Universidade Federal do Paraná, Curitiba, 1997. Disponível em: <http://hdl.handle.net/1884/44185>. Acesso em: 18 jun. 2021.

NIEHEUS, R. C. **Autoclaves verticais: uma proposta de sistema para garantia do processo de esterilização**. 2004. Dissertação (Programa de Pós-Graduação em Engenharia Elétrica) - Universidade Federal de Santa Catarina, Florianópolis, 2004. Disponível em: <https://repositorio.ufsc.br/handle/123456789/87240>. Acesso em: 07 jul. 2021.

ORDÓÑES, Juan A. **Tecnologia de Alimentos**. Porto Alegre, v. 2, Artmed, 2005.

PELLISSARI, T. J.; dos Santos Lima, H. O.; dos Santos Lima, M. V.. Clean-Test self-contained biological indicator performance used to steam sterilization process validation. **Revista Brasileira de Pesquisa em Alimentos**, n.1, v.2, p.38-44, 2011. Disponível em: <https://periodicos.utfpr.edu.br/rebrapa/article/view/3346>. Acesso em 09 jun. 2021.

PENNA, T. C. V.; ISHII, M.; MACHOSHVILI, I. A.; MARQUES, M. The effect of bioindicator preparation and storage on thermal resistance of *Bacillus stearothermophilus* spores. **Applied Biochemistry and Biotechnology**. São Paulo, v. 98, p. 525-538, 2002. Disponível em: <https://www.proquest.com/openview/0c2b1ea47db72144b74ac5cc671d6e03/1?pq-origsite=gscholar&cbl=54110>. Acesso em: 12 jul. 2021.

PIAZZI, A. C. F.; **Construção de modelos matemáticos para processos fermentativos—avaliação de efeitos das condições de cultivo**. 2019. Dissertação (Programa de Pós-Graduação em Engenharia Química) - Universidade Federal de Santa Maria, Santa Maria, 2019. Disponível em: <http://repositorio.ufsm.br/handle/1/19696>. Acesso em: 09 jun. 2021.

PIGOZZO, A. B. **Implementação computacional de um modelo matemático do sistema imune inato**. 2011. Dissertação (Programa de Pós-Graduação em Modelagem Computacional) - Universidade Federal de Juiz de Fora, Juiz de Fora, 2011. Disponível em: <https://repositorio.ufjf.br/jspui/handle/ufjf/3529>. Acesso em: 10 jun. 2021.

PILAR, R. de M. **Microbiologia dos Processos Alimentares**. São Paulo: Varela, 2005.

QUINTINO, A. P. N. N. Testes desafio em cargas subsequentes na esterilização a vapor saturado: estudo comparativo dos desempenhos. **Revista SOBECC**, n. 3, v. 22, p. 145-151, 2017. Disponível em: <https://revista.sobecc.org.br/sobecc/article/view/225/pdf>. Acesso em: 10 jun. 2021.

RENCHER, A. C. **Methods of Multivariate Analysis**. New York: Wiley-Interscience, 2002.

ROBLES, M. M. R. **Validación de procesos de esterilización em autoclaves de vapor em la industria farmacêutica**. 2014. Tese (Programa de Doutorado em Químico Farmacêutico Biólogo) - Universidad Autónoma del Estado de México, Toluca de Lerdo, 2014. Disponível em: <http://ri.uaemex.mx/handle/20.500.11799/14480>. Acesso em: 12 jul. 2021.

RODRIGUES, A. F. S. A importância da autoclave para os estabelecimentos de saúde. **Revista Brasileira de Educação e Saúde**, n. 4, v. 9, p. 138-144, 2019. Disponível em: <https://www.gvaa.com.br/revista/index.php/REBES/article/view/6569>. Acesso em: 10 jun. 2021.

RODRIGUES, L. T. **Estratégia para validação de processos de esterilização por vapor saturado em autoclaves nas indústrias farmacêuticas**. 2010. Dissertação (Programa de Pós-Graduação em Engenharia Química) - Universidade do Estado do Rio de Janeiro, Rio de Janeiro 2010. Disponível em: https://bdtd.ibict.br/vufind/Record/UERJ_5dbcf3e10adf82db8e1cb5c6ee1889f5. Acesso em: 10 jun. 2021.

SALMAZO NETTO, F. L.; de SOUZA, A. C. Estudo para fixar valor de D em esterilização por autoclave. **Revista acadêmica Oswaldo Cruz**, n. 3, p. 1-2, 2014. Disponível em: http://www.revista.oswaldocruz.br/Edicao_03/Artigos. Acesso em: 10 jun. 2021.

SANTOS, G. C. M. **Redução da biocarga e garantia de esterilidade em implantes mamários de silicone**. 2009. Dissertação (Programa de Pós-Graduação em Fármaco e Medicamentos) - Universidade de São Paulo, São Paulo, 2009. Disponível em: <https://www.teses.usp.br/teses/disponiveis/9/9139/tde-01022010-161804>. Acesso em: 07 jul. 2021.

SANTOS, A. B.; ANTONELLI, S. C. Aplicação da abordagem estatística no contexto da gestão da qualidade: um survey b com indústrias de alimentos de São Paulo. **Revista Gestão & Produção**, n. 3, v. 18, p. 509-524, 2011. Disponível em: <https://doi.org/10.1590/S0104-530X2011000300006>. Acesso em: 10 jun. 2021.

SANTOS, I. F. C. Eficácia dos métodos de antissepsia e esterilização da abraçadeira autoestática de náilon. **Revista científica de medicina veterinária**, n. 22, v. 12, p. 1-10, 2014. Disponível em: <https://pesquisa.bvsalud.org/portal/resource/pt/vti-15455>. Acesso em: 07 jul. 2021.

SARAIVA, A. S. B. **Avaliação de riscos, Qualificação de Equipamentos de Esterilização, Despirogenização e Produção**. 2016. Dissertação (Programa de Pós-Graduação em Engenharia Química e Bioquímica) - Universidade Nova de

Lisboa, Lisboa, 2016. Disponível em: <http://hdl.handle.net/10362/19556>. Acesso em: 09 jun. 2021.

SILVA, A. D. S. da; VALIATTI, T. B.; BARCELOS, I. B.; MARTINS, R. C.; OLIVEIRA, D. F.; CASTRO, B. S. Verificação da prática de biossegurança por manicures/pedicuros em salões de beleza localizados no município de JI-PARANÁ/RO. **Revista interdisciplinar de estudos em saúde**, n. 1, v. 06, p. 39-49, 2017. Disponível em: <https://periodicos.uniarp.edu.br/index.php/ries/issue/view/59>. Acesso em: 07 jul. 2021.

SILVA, M. O. A. da; SILVA, F. A. S. da Modelos matemáticos epidemiológicos num contexto de ciência, tecnologia e sociedade. **Revista Mundi Engenharia, Tecnologia e Gestão**, n. 1, v. 3, p. 1-15, 2018. Disponível em: <http://dx.doi.org/10.21575/25254782rmetg2018vol3n1501>. Acesso em: 09 jun. 2021.

SILVA, T. C.; PERIÇARO, G. A. Classificação dos candidatos ao vestibular da FECILCAM via técnicas estatísticas multivariadas. **Congresso Nacional de Matemática Pura e Aplicada, Cuiabá**. v. 2, p. 566-571, 2009. Disponível em: http://arquivo.sbmec.org.br/eventos/cnmac/xxxii_cnmac/pdf/125.pdf. Acesso em: 10 jun. 2021.

SOUZA, W. E. **Sistemas multiagentes inteligentes para reconhecimento e classificação de padrões aplicados em problemas de sistemas elétricos de potência**. 2012. Tese (Programa de Pós-Graduação em Métodos Numéricos em Engenharia) - Universidade Federal do Paraná, Curitiba, 2012. Disponível em: <https://www.acervodigital.ufpr.br/handle/1884/33906>. Acesso em: 18 jun. 2021.

SOUZA, W. E. **Análise das distorções harmônicas de tensão a partir de características dos transformadores e de dados de consumo**. 2008. Dissertação (Programa de Pós-Graduação em Métodos Numéricos em Engenharia) - Universidade Federal do Paraná, Curitiba, 2008. Disponível em: <https://www.acervodigital.ufpr.br/handle/1884/17074>. Acesso em: 21 jun. 2021.

SCHMIDELL, W.; LIMA, U. de A; AQUARONE, E.; BORZANI, W. **Biotechnologia Industrial**. São Paulo: Edgard Blucher Ltda., 2001.

TAPIAS, J.; ZAMBRANO, C. A. C.; VILLAMIZAR, O. G.; URIBE, J. C. Evaluación de las características microbiológicas y fisicoquímicas del agua sometida a procesos de esterilización en autoclaves a vapor. **Revista Respuestas**, n. 2, v. 22, p. 59-66, 2017. Disponível em: <https://revistas.ufps.edu.co/index.php/respuestas/article/view/1175>. Acesso em: 10 jun. 2021.

TIPPLE, A. F. V.; PIRES, F. V.; GUAGAGNIN, S. V. T.; MELO, D. D. S. O monitoramento de processos físicos de esterilização em hospitais do interior do estado de Goiás. **Revista da Escola de Enfermagem da USP**, n. 3, v. 45, p. 751-757, 2011. Disponível em: <https://www.revistas.usp.br/reeusp/article/view/40762>. Acesso em: 10 jun. 2021.

VIER-PELISSER, F. V.; SÓ, M. V. R.; GREMELMAIER, L.; BOZETTI, R.; & PELISSER, A. Avaliação da esterilização em autoclaves odontológicas através do monitoramento biológico. **Revista da Faculdade de Odontologia de Porto Alegre**, n. 2, v. 49, p. 5-10, 2008. Disponível em:

<https://www.seer.ufrgs.br/RevistadaFaculdadeOdontologia/article/view/2961>. Acesso em: 09 jul. 2021.