

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



Aline Ramos Borges

**Modelagem em Experimentos Fatoriais Replicados para
Melhoria de Processos Industriais Têxteis**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia de Produção do Departamento de Engenharia Industrial da PUC-Rio.

Orientador: Prof. Eugenio Kahn Eprecht

Rio de Janeiro
Junho de 2014



Aline Ramos Borges

Modelagem em Experimentos Fatoriais Replicados para Melhoria de Processos Industriais Têxteis

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia de Produção do Departamento de Engenharia Industrial da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Eugenio Kahn Epprecht

Orientador

Departamento de Engenharia Industrial - PUC-Rio

Prof. Antonio Fernando de Castro Vieira

Pesquisador Autônomo

Prof. Flávia Duta Pimenta

Centro de Tecnologia da Indústria Química e Têxtil

Prof. José Eugenio Leal

Coordenador Setorial do Centro Técnico Científico - PUC-Rio

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Aline Ramos Borges

Graduou-se em Engenharia Industrial Têxtil pela Faculdade SENAI CETIQT no ano 2010 e em Estatística pela UERJ em 2011. Tenho experiência em planejamento de experimentos, Controle Estatístico de Processo (CEP), Controle Estatístico da Qualidade (CEQ), modelos de previsão, Modelos Lineares Generalizados (MLG), análise espacial, análise de clusters, Sistema de Informação Geográfica (SIG) e Sistema de Posicionamento Global (GPS). Atualmente, sou Trainee da La Estampa.

Ficha Catalográfica

Borges, Aline Ramos

Modelagem em experimentos fatoriais replicados para melhoria de processos industriais têxteis / Aline Ramos Borges ; orientador: Eugenio Kahn Epprecht. – 2014.
85 f. : il. (color.) ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2014.
Inclui bibliografia

1. Engenharia Industrial – Teses. 2. Beneficiamento têxtil. 3. Planejamento fatorial replicado. 4. Modelo Linear Generalizado (MLG). 5. Subdispersão. 6. Teste de Wald. I. Epprecht, Eugenio Kahn. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. III. Título.

CDD: 658.5

À minha querida mãe, pelo apoio e amor incondicional
e pelo grande incentivo aos estudos.

Agradecimentos

Primeiramente, a Deus por me amparar nos momentos difíceis, me dar força interior para superar as dificuldades, mostrar os caminhos nas horas incertas e me suprir em todas as minhas necessidades.

Ao meu orientador Professor Eugenio Kahn Epprecht pela colaboração, paciência e seus conhecimentos repassados durante todo o desenvolvimento deste trabalho.

Ao Professor Antônio Fernando de Castro Vieira pela sua fundamental contribuição no desenvolvimento da análise estatística.

A todos os professores e funcionários do Departamento de Engenharia Industrial da PUC-Rio pelos ensinamentos e pela ajuda.

À Empresa do Segmento Têxtil, por ter fornecido acesso total à empresa para a execução deste trabalho.

À Faculdade SENAI CETIQT, por fornecer a infraestrutura física e logística com acesso aos laboratórios de química e de controle da PPI, aos equipamentos em escala piloto disponíveis na PPI e às normas técnicas disponíveis na Biblioteca e na Coordenação de Serviços Laboratoriais (CSL), além de auxílio de um professor da instituição. Meu agradecimento especial ao Diretor de Educação Leonardo Garcia Teixeira Mendes, que possibilitou a realização de uma parceria intelectual entre PUC-Rio e SENAI CETIQT.

Meu agradecimento especial a Professora Flávia Duta Pimenta pelo grande incentivo aos estudos e pela amizade fortalecida ao longo dos anos.

Agradeço também a La Estampa empresa da qual faço parte atualmente pelo incentivo aos estudos.

A todos os amigos e familiares que de uma forma ou de outra me estimularam ou me ajudaram ao longo do curso.

Em especial, à minha mãe que me apoiou, confiou e esteve ao meu lado em todos os momentos de minha vida.

Por fim, meu agradecimento ao CNPq e à PUC-Rio, pelos auxílios concedidos, sem os quais esta dissertação não poderia ter sido realizada.

Resumo

Borges, Aline Ramos; Epprecht, Eugenio Kahn. **Modelagem em Experimentos Fatoriais Replicados para Melhoria de Processos Industriais Têxteis**. Rio de Janeiro, 2014. 85p. Dissertação de Mestrado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Esta dissertação descreve a aplicação de Modelos Lineares Generalizados (MLGs) à análise de um experimento visando identificar a combinação dos níveis das variáveis independentes: concentração de hidróxido de sódio (A), volume de hipoclorito de sódio (B) e sua interação (AB), que minimiza a variável resposta: proporção de itens com defeitos, em um processo de beneficiamento numa indústria têxtil de pequeno porte. A variável resposta encontra-se na forma de proporção, violando os pressupostos básicos do Modelo Linear Clássico e com isso as estimativas dos coeficientes pelo método de Mínimos Quadrados Ordinários (MQO) é menos confiável. O planejamento utilizado foi o fatorial completo 2^2 com ponto central e replicado. Após o planejamento, a modelagem pelo MLG é aplicada, só então é possível identificar uma subdispersão dos dados, verificar que o modelo empregado está correto e que o volume de hipoclorito de sódio (B) é o único fator significativo, no processo de alvejamento industrial da empresa. Portanto, como a finalidade é minimizar a resposta, utiliza-se o nível inferior (-1) desta variável. Conseqüentemente, como o intuito é reduzir os custos com insumos químicos pode-se utilizar o nível mínimo da concentração de hidróxido de sódio (A) e o nível máximo da interação entre os fatores (AB), já que eles não são significativos ao modelo.

Palavras-chave

Beneficiamento têxtil; planejamento fatorial replicado; Modelo Linear Generalizado (MLG); subdispersão; teste de Wald.

Abstract

Borges, Aline Ramos; Epprecht, Eugenio Kahn (Advisor). **Modeling in Replicated Factorial Experiments for Improvement of Textile Industrial Processes**. Rio de Janeiro, 2014. 85p. MSc. Dissertation – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

This dissertation describes the application of Generalized Linear Models (GLMs) to the analysis of an experiment with the purpose identify the levels combination of independent variables: concentration of sodium hydroxide (A) volume of sodium hypochlorite (B) and their interaction (AB), that minimizes the response variable: proportion of defective items, in a process in a small plant of the textile industry. The response variable takes the form of a proportion, that violates the basic assumptions of the Classic Linear Model and, as a result, the estimates of the coefficients by Ordinary Least Squares method is less reliable. The design employed was a replicated complete 2^2 factorial design with central point. After doing the planning, the modeling by MLG is applied, and then it is possible to identify a underdispersion data; to verify that the model used is correct and that the volume of sodium hypochlorite (B) is the only significant factor in the industrial process of bleaching the company. Therefore, as the purpose is to minimize the response, it is used the lower level (-1) of this variable. Consequently, as the aim is to reduce costs of chemical inputs can use the minimum level of concentration of hydroxide sodium (A) and the maximum level of interaction between factors (AB), since they are not significant to the model.

Keywords

Textile processing; replicated factorial design; Generalized Linear Models (GLM), underdispersion; Wald test.

Sumário

1. Introdução	12
1.1. Contexto e definição do problema	12
1.2. Objetivos	16
1.3. Justificativas	16
1.4. Estrutura do Trabalho	17
2. Fundamentos Conceituais	19
2.1. Processamento Têxtil	19
2.1.1. Beneficiamento	21
2.1.2. Limpeza Industrial	24
2.2. Regressão Linear	25
2.2.1. Planejamento de Experimentos	26
2.2.1.1. Seleção do Planejamento	29
2.2.1.2. Planejamento Fatorial Com Ponto Central	29
2.2.2. Modelo Linear Clássico	30
2.2.3. Análise dos Resíduos	31
2.2.4. Adequação do Modelo Clássico	32
2.2.4.1. Verificação de Normalidade	34
2.2.4.2. Verificação de Independência	35
2.2.4.3. Verificação de Variância Constante	36
2.2.4.4. Verificação de Observações Atípicas (<i>Outliers</i>)	36
2.2.4.5. Verificação de Observações Influentes	38
2.2.5. Sobredispersão e subdispersão	38
2.3. Modelos Lineares Generalizados	41
2.3.1. Estrutura de MLG	42
2.3.2. Regressão Logística	45
2.3.2.1. Modelo Logit	46
2.3.3. Adequação do MLG	47
2.3.3.1. Teste de Significância dos Coeficientes	48

2.3.3.2.	Estimação do Parâmetro de Dispersão	50
2.3.3.3.	Função de Ligação	51
2.3.3.4.	Função de Variância	51
2.3.3.5.	Distância de Cook	52
3.	Descrição do Experimento	53
3.1.	<i>Softwares</i> Utilizados	53
3.2.	Caracterização do Produto	54
3.3.	Procedimento de alvejamento	55
3.3.1.	A Empresa	56
3.3.2.	Escala Piloto	57
3.4.	Definição dos Parâmetros de Processo	58
3.5.	Variável Resposta	60
3.6.	Adequação do MLG	61
3.6.1.	Formulação do Modelo	61
3.6.2.	Ajuste do Modelo	62
3.6.3.	Adequação do Modelo	62
4.	Análise de Resultados	64
4.1.	Apresentação dos Dados	64
4.2.	Variável Resposta	65
4.3.	Aplicação do MLG	65
5.	Conclusão	74
6.	Referências Bibliográficas	77

Lista de tabelas

Tabela 1 –	Dados para o MLG	43
Tabela 2 –	Funções de ligação canônica para distribuição de probabilidade	44
Tabela 3 –	Variáveis independentes do planejamento de experimentos e seus níveis	58
Tabela 4 –	Matriz fatorial 2^2 na forma padrão e na ordem aleatória	59
Tabela 5 –	Matriz padrão de planejamento fatorial 2^2 com ponto central, tendo como variável resposta a proporção de itens com defeitos	64
Tabela 6 –	Estimativa dos coeficientes e erro padrão de todos os fatores da variável resposta proporção de itens com defeitos	65
Tabela 7 –	Teste de Wald para todos os fatores	66
Tabela 8 –	IC do teste de Wald para todos os fatores	67
Tabela 9 –	Estimativa dos coeficientes e erro padrão somente do fator significativo	67
Tabela 10 –	Teste de Wald para o fator significativo	71
Tabela 11 –	IC do teste de Wald para o fator significativo	72

Lista de figuras

Figura 1 –	Processamento têxtil	20
Figura 2 –	Beneficiamento têxtil	21
Figura 3 –	Representação de um experimento	27
Figura 4 –	Exemplo da função de resposta Logística: (a) monotonicamente crescente e (b) monotonicamente decrescente	45
Figura 5 –	Máquina de lavar industrial da empresa	57
Figura 6 –	Máquina de lavar industrial da PPI	58
Figura 7 –	Gráfico de probabilidade Normal com envelope para o fator significativo	68
Figura 8 –	Gráfico dos resíduos <i>studentizados</i> versus o valor ajustado para o fator significativo	69
Figura 9 –	Gráfico do valor absoluto do resíduo <i>studentizado</i> versus o valor ajustado para o fator significativo	70
Figura 10 –	Gráfico da distância de Cook para o fator significativo	71

1.

Introdução

No presente capítulo será feita a caracterização do problema em estudo e uma breve introdução dos aspectos mais relevantes à pesquisa.

1.1.

Contexto e definição do problema

A indústria têxtil e de confecção vem passando por grandes transformações nos últimos anos, devido ao grande investimento em tecnologia. Este tipo de investimento exige que as tarefas e operações sejam mais adequadas e executadas de maneira correta e no tempo certo. Também há uma exigência cada vez mais crescente por parte dos consumidores, no que diz respeito à qualidade do produto final.

Dentre as técnicas disponíveis e mais amplamente discutidas e utilizadas para a melhoria da qualidade dos processos está o Planejamento e a Análise de Experimentos. O objetivo é identificar os fatores denominados variáveis independentes, que influenciam alguma importante característica de qualidade do processo ou do produto, a qual é denominada de variável resposta, e ajustar um modelo que represente o efeito desses fatores sobre a resposta, de maneira a buscar o ponto de operação que a otimiza (o que, conforme o caso específico, é o ponto que a maximiza, minimiza, ou torna-a o mais próxima possível de um valor-alvo).

A variável resposta pode ser contínua ou discreta e os fatores podem assumir valores, ou seja, níveis “baixo” (representado pelo sinal negativo), “alto” (representado pelo sinal positivo) e “médio” (representado pelo algarismo arábico zero). Os fatores envolvidos em um experimento podem ser qualitativos ou quantitativos. Os fatores quantitativos são aqueles cujos níveis podem ser associados a pontos em uma escala numérica, tais como pressão, pH, tempo, temperatura, agitação, velocidade, volume, porcentagem, concentração, aeração, etc. Por outro lado, fatores qualitativos são aqueles em que não é possível esta

associação, tais como, operadores, bateladas, lotes de matéria prima, equipamentos iguais utilizados em um processo, dentre outros. Ainda assim, é usual “batizar”, isto é, nomear as categorias de um fator nominal como “nível alto” e “nível baixo” por exemplo.

Porém muitas das vezes não é possível avaliar as características de qualidade de um determinado produto, por exemplo, pregos, parafusos ou lâmpadas; nestes casos classificam-se os itens em “defeituosos” ou “não defeituosos”. Ainda, para unidades mais complexas, como geladeiras, automóveis, televisores, tecidos ou chapas de metal, em que alguns pequenos defeitos não inutilizam o todo, pode-se utilizar o monitoramento do processo através do número de não conformidades na amostra. Ser “defeituoso” ou “não defeituoso” é um atributo do produto, bem como as não conformidades presentes num produto (COSTA; EPPRECHT; CARPINETTI, 2011).

Na análise de experimentos o intuito é medir a influência das variáveis independentes sobre a variável de resposta, definindo-se uma relação entre elas através de modelos matemáticos, onde pode-se estabelecer uma combinação ótima dos níveis das variáveis independentes, tendo em vista a obtenção de um valor desejado da resposta, podendo minimizá-la, maximizá-la ou encontrar um valor específico (alvo). Para construir esses modelos são realizados experimentos, com os quais são observadas as respostas em função dos níveis pré-estabelecidos das variáveis independentes. Segundo Bonanni (2005), este tipo de análise teve sua aplicação aumentada na última década, tanto nas indústrias manufatureiras quanto nas de serviços, com o intuito de melhorar a qualidade de seus produtos, processos e serviços.

O tipo de planejamento experimental a ser usado dependerá do objetivo que se deseja atingir com a experimentação. Os pesquisadores podem estar interessados em avaliar os efeitos devido à interação no nível de um ou mais fatores ou na otimização das respostas (BONANNI, 2005). Um tipo particular de experimentos é o fatorial, que considera as combinações de todas as variáveis em todos os níveis, o que permite identificar possíveis interações entre os fatores. Eles possuem algumas vantagens como a construção de modelos que se ajustam bem aos dados e a redução de prováveis erros dos testes de significância dos parâmetros do modelo.

Na otimização de processos produtivos e de projetos de produtos, os métodos inicialmente empregados tinham como foco principal identificar os fatores que poderiam afetar a média da variável resposta, produzindo nestes efeitos denominados de efeitos na posição (*location effects*). Nesses métodos a variância da resposta era considerada constante, isto é, igual para qualquer combinação de níveis dos fatores e qualquer valor da resposta. Porém recentemente passou-se a identificar os fatores que afetam a variância da resposta, produzindo nestes efeitos denominados de efeitos na dispersão (*dispersion effects*).

Dois tipos de fatores podem influenciar os efeitos relatados acima e, portanto, devem ser considerados. Os fatores de controle são aqueles que se pode manter em um dado nível durante o processo produtivo e os fatores de ruído (fatores incontroláveis) são aqueles que oscilam em torno de um dado nível durante o processo, embora possam ser controlados durante a fase experimental do projeto do processo.

O modelo linear clássico, como o Método dos Mínimos Quadrados (MQO), possui algumas propriedades estatísticas muito atraentes que o tornaram um dos métodos de análise de regressão mais poderosos e difundidos mundialmente. Sua difusão e simplicidade devem-se às três considerações que são feitas para sua aplicação:

- i) *A resposta tem distribuição Normal;*
- ii) *A variância da resposta é constante (homocedasticidade);*
- iii) *Os efeitos dos fatores sobre a variável resposta se combinam aditivamente.*

Essas condições são encontradas em muitas aplicações industriais, porém em casos em que a variável resposta não segue uma distribuição de probabilidade Normal, muitas vezes a variância da resposta não é constante e sim função da média (da resposta).

Uma classe de modelos de grande importância no contexto da modelagem estatística é a classe dos Modelos Lineares Generalizados (MLG), com aplicações nas mais diversas áreas do conhecimento, como em várias áreas da ciência e da indústria, especialmente nas indústrias química, agroindústria e farmacêutica (PINTO; PEREIRA, 2012).

Portanto, caso um ou mais dos três pressupostos não possam ser considerados, uma alternativa viável é alterar a escala da variável resposta, transformando-a e usando o modelo linear e o MMQ com os dados transformados. Em alguns casos relatados na literatura por Vieira (2004), Couto *et al.* (2009), Bello (2010), Sant'Anna, Caten (2010), Cordeiro e Demétrio (2013), pode-se não conseguir obedecer aos pressupostos mesmo com a transformação da variável resposta. Para estes casos deve-se considerar a aplicação do MLG, que permite considerar outras distribuições de probabilidade que não a Normal, conseqüentemente não é necessário que a variância seja constante, conseguindo-se linearizar a equação através de uma função de ligação.

Nesta dissertação uma empresa de pequeno porte, do segmento têxtil (limpeza industrial), apresenta um gargalo na produção no setor de lavanderia, uma vez que os procedimentos adotados não apresentam uma padronização, impactando diretamente nos setores subsequentes. E existem 5 tipos de produtos: estopa engomada, capa de fardo, pano branco, pano de cor e malha de óleo, que passam por este setor. Porém apenas um representa um lucro maior e é extremamente crítico e, portanto, foi analisado neste estudo: a malha (de algodão) de óleo.

No produto analisado em questão é aplicado um beneficiamento têxtil primário (alveijamento), que visa à preparação desse material para as etapas posteriores, oferecendo hidrofiliidade, remoção das sujidades e um grau de brancura e grau de nuance uniforme nesses produtos. Por conseguinte, é importante para a empresa estudar e identificar os fatores que exercem influência sobre seu processo de alveijamento e que impactam diretamente nas características descritas anteriormente.

1.2.

Objetivos

O objetivo principal deste estudo é mostrar quais variáveis podem influenciar ou não o processo de beneficiamento têxtil (alvejamento) da empresa. Conseqüentemente, consegue-se avaliar o impacto que essas variáveis podem exercer sobre a variável resposta.

Neste estudo aplicaremos a técnica de experimentos fatoriais replicados com ponto central, visando extrair o máximo de informações úteis para empresa, com um número mínimo de experimentos. Após o planejamento, a modelagem pelo MLG é aplicada para medir a influência das variáveis independentes: concentração de hidróxido de sódio (A) e volume de hipoclorito de sódio (B) sobre a variável resposta: proporção de itens com defeitos. Com o intuito de estabelecer uma relação entre elas (VIEIRA, 2004; BELLO, 2010; CORDEIRO; DEMÉTRIO, 2013). Também se consegue identificar a combinação de valores (níveis) das variáveis independentes que minimiza a variável resposta.

1.3.

Justificativas

A principal motivação deste estudo é trabalhar com uma situação real, já que a empresa vem observando, ao longo dos anos, uma exigência cada vez maior de seus clientes pela certificação da qualidade em seus produtos. Para ela o produto que apresenta alta capacidade de absorção e um maior teor de brancura e nuance é um produto de melhor qualidade e, conseqüentemente, de preço mais elevado. Por meio deste estudo, a empresa pode aprimorar o processo, visando sempre à busca contínua e uma maior qualidade de seus produtos, já que a qualidade tornou-se um diferencial entre ela e suas concorrentes.

Quanto à modelagem estatística deste processo real existe a necessidade em se realizar primeiramente experimentos fatoriais replicados, por serem eficientes, rápidos, econômicos e bastante difundidos e utilizados nas mais diversas áreas industriais como: agrícola, química, farmacêutica, têxtil, dentre outras (PIMENTA, 2006; COSTA, 2008; BORGES, 2010; PINTO; FERREIRA, 2012; BORTOLINI, 2012; HABITZREUTER, 2013). O segundo passo seria reunir e descrever em detalhes todos os aspectos da modelagem da média e da variância nesses experimentos, utilizando-se os MLGs.

A principal dificuldade deste tipo de modelagem em relação ao MLG é que existem poucos autores, dentre eles Vieira (2004), Bello (2010), Cordeiro e Demétrio (2013) que detalham claramente alguns aspectos importantes, como a aplicação dos testes de significância dos coeficientes, as estatísticas e os gráficos indicados para verificar a adequação do modelo.

Portanto, serão descritos no decorrer desta dissertação os modelos de regressão indicados para o processo produtivo da empresa, como os métodos de estimativa dos coeficientes dos modelos, os testes de significância para seleção dos fatores importantes, a estimativa do fator de dispersão (ϕ) para verificar se existe ou não uma sobredispersão ou uma subdispersão dos dados, bem como as estatísticas e os gráficos indicados para verificar a adequação do modelo. Uma vez construído o modelo, será mostrado como obter as melhores condições de operação.

Este projeto também contou com os conhecimentos e expertises de duas instituições de ensino, o Centro de Tecnologia da Indústria Química e Têxtil (SENAI CETIQT), local onde foram realizados os experimentos e a Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

1.4.

Estrutura do Trabalho

Esta dissertação está estruturada em cinco capítulos com os conteúdos apresentados na sequência.

No Capítulo 2 são apresentados os Fundamentos Conceituais, isto é, os tópicos mais relevantes acerca do processamento têxtil aplicado na empresa, bem como informações necessárias sobre o segmento têxtil, mais especificamente sobre a limpeza industrial. Além disso, também serão descritas a técnica de Planejamento de Experimentos Fatoriais e a técnica do MLG utilizada para identificar os fatores que exercem influência sobre a variável resposta.

No Capítulo 3 faz-se uma breve descrição dos *softwares* utilizados nesse estudo, bem como a caracterização do produto e a descrição de seu processo de alvejamento em escala industrial e piloto. E, por último, a descrição dos parâmetros de processos utilizados, a partir dos dados fornecidos pela empresa e como o modelo proposto será avaliado e validado.

No Capítulo 4 serão apresentados e discutidos os resultados dos experimentos realizados, bem como o(s) modelo(s) resultante(s).

O Capítulo 5 contém os comentários finais, com apresentação das conclusões do trabalho, juntamente com sugestões para trabalhos futuros.

Finalmente, no capítulo 6 são apresentadas as referências bibliográficas empregadas nesse estudo.

2.

Fundamentos Conceituais

Inicialmente, serão apresentadas as informações mais relevantes sobre o processamento têxtil, como os conceitos referentes ao beneficiamento primário aplicado ao produto em estudo e informações referentes ao setor de limpeza industrial.

Também será apresentado um conjunto de técnicas estatísticas denominadas análise de regressão linear, onde se procura estabelecer a relação entre uma variável resposta y e um conjunto de variáveis regressoras (ou independentes), x_1, x_2, \dots, x_k . Cabe lembrar que, nesse estudo, a variável resposta representa uma característica de qualidade de um processo produtivo e as variáveis independentes são os fatores que afetam o processo quando este está em operação.

Ao final desta revisão, abordaremos a aplicação de MLG, que possui algumas vantagens em relação ao Método dos Mínimos Quadrados (MMQ), como menores Intervalos de Confiança (IC) e obtenção de linearidade através de uma função de ligação. Nesta seção será abordada também a estrutura formal do MLG, como são feitos a estimação, o teste de significância dos parâmetros, a estimativa do parâmetro de dispersão (ϕ) e como é verificada a adequação do modelo.

2.1.

Processamento Têxtil

De acordo com Andrade (2006), o processamento têxtil é composto pelas etapas de fiação, malharia ou tecelagem e beneficiamento. Desta forma, as fibras ou filamentos, naturais ou sintéticas, constituem os insumos básicos que alimentam a etapa de fiação e são processados de acordo com o tipo de fibra empregada, dando origem aos mais variados tipos de fios. Em seguida, estes abastecem as etapas produtivas subsequentes de malharia ou tecelagem, nas quais há a conversão de fios em malhas ou tecidos planos. Por último, surge a etapa de beneficiamento, responsável pela limpeza, melhoria da aparência e da capacidade de absorção de água, aumento de resistência, dentre outros.

O processamento têxtil descrito acima pode ser melhor entendido através da

Figura 1.

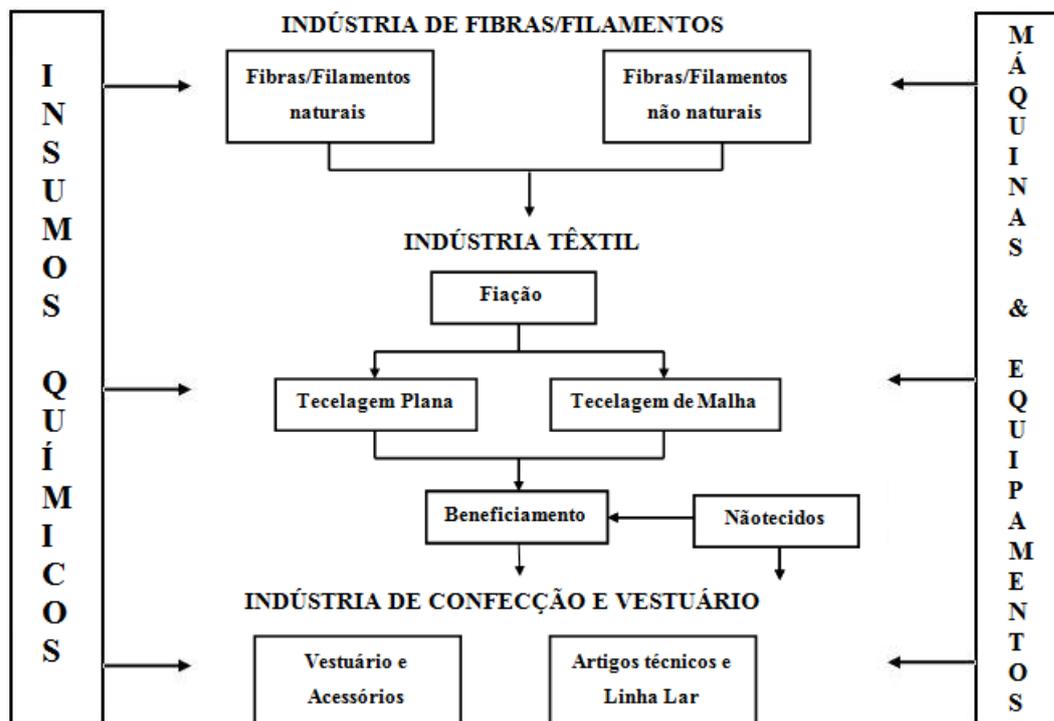


Figura 1 – Processamento têxtil

Fonte: Adaptado de ANDRADE (2006)

A primeira etapa referente à Indústria de Fibras/Filamentos mostra que as fibras/filamentos podem ser de origem natural ou não natural e sua obtenção respectivamente, está ligada ao setor agroindustrial e ao setor petroquímico. Para os demais insumos auxiliares, mantém-se relação com a indústria química, como exemplo, gomas, corantes/pigmentos, resinas, entre outros itens utilizados nos processos de beneficiamento têxtil, assim como a indústria metal mecânica, produtora de bens de capital, para produção do maquinário específico a cada etapa produtiva (ANDRADE, 2006).

Ainda de acordo com o autor acima, a segunda etapa referente à Indústria Têxtil é composta por quatro ciclos produtivos principais, porém destacaremos apenas o beneficiamento têxtil, que será mais detalhado a seguir, pois é o foco deste estudo.

A terceira e última etapa referente à Indústria de Confecção e Vestuário é responsável por transformar as malhas ou tecidos planos em diferentes produtos, que abastecem o mercado com itens de necessidade comum, tais como: vestuário em geral, cama, mesa, banho, decoração, meias, artigos industriais, dentre outros.

2.1.1.

Beneficiamento

O beneficiamento têxtil compreende um conjunto de atividades, que cuida do enobrecimento dos substratos têxteis (tecidos de malha ou planos) e apresenta três tipos de atividades com características bem definidas, como mostra a **Figura 2**, que serão descritas a seguir:

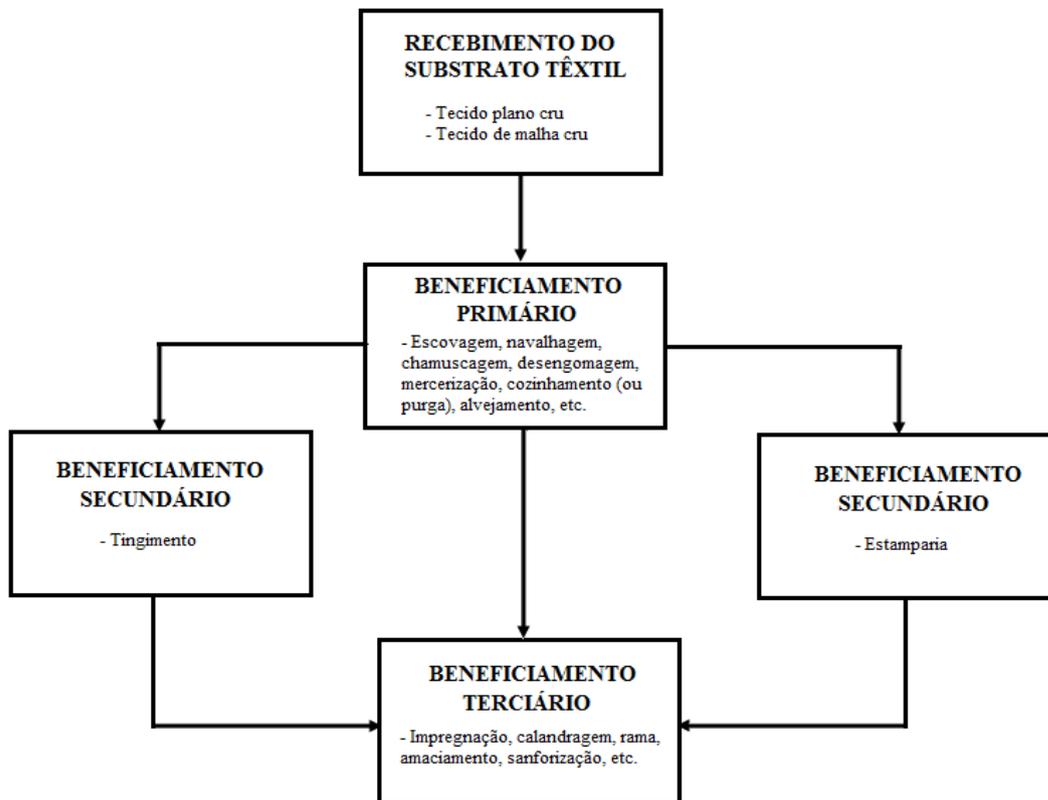


Figura 2 – Beneficiamento têxtil

Fonte: FEAM (2013)

i) Beneficiamento primário:

Os processos de beneficiamento primário compreendem a preparação ou o pré-tratamento da fibra, fio ou tecido, adequando-os às etapas subsequentes, oferecendo hidrofiliidade, remoção das sujidades e um grau de brancura uniforme e adequado às etapas posteriores de tingimento ou estampagem, eliminando óleos, ceras, pigmentos e sujeiras provenientes das etapas de fiação e tecelagem (BELTRAME, 2000; SILVA, 2007; HABITZREUTER, 2013).

Dentre alguns processos, pode-se citar: escovagem, navalhagem, chamuscagem, desengomagem, mercerização, cozinhamento (ou pré-alvejamento ou purga) e alvejamento (BELTRAME, 2000).

Nesta dissertação será realizado um processo de alvejamento nos tecidos da empresa, que consiste na remoção das sujidades agregadas a eles, durante outras etapas da cadeia têxtil, mediante a remoção e/ou oxidação desses componentes pelo uso de diferentes alvejantes químicos e produtos auxiliares (JERÔNIMO et al., 2005).

De acordo com Immich (2006), os alvejantes químicos mais utilizados nos processos de alvejamento são: hipoclorito de sódio (NaClO), peróxido de hidrogênio (H_2O_2) e clorito de sódio (NaClO_2). Vale ressaltar que a escolha do alvejante depende do tipo de fibra e do equipamento utilizado.

O alvejante utilizado na empresa é o hipoclorito de sódio. Esse alvejante é, normalmente, encontrado na forma líquida, apresentando uma coloração ligeiramente amarelada e seu custo de processo é considerado baixo. Choudhury (2006), diversos fatores podem influenciar a utilização desse alvejante e serão explicados a seguir:

- **pH:** caso esse parâmetro não seja controlado existe o risco de dano ao tecido. Portanto, a faixa de pH ideal para utilizá-lo está compreendido entre 10 a 12.

- **Tempo e temperatura:** o tempo de processo é considerado demasiadamente lento, tendo em vista que é realizado em temperatura ambiente, o que inviabiliza uma operação contínua rápida. Vale ressaltar que não se deve trabalhar com temperaturas muito elevadas (em torno de 60°C), pois há risco de danificar o tecido. No Brasil, essa temperatura ambiente varia de acordo com a região do país, por exemplo, nas regiões ao sul do país as temperaturas chegam bem próximas a 0°C , enquanto que no nordeste podem chegar a mais de 40°C . Em síntese, quanto maior for a temperatura menor será o tempo de processo.

- **Cloro residual:** concentração elevada de cloro residual no efluente gerado o torna questionável do ponto de vista ambiental. Portanto, após a realização do processo de alvejamento, deve-se eliminá-lo do tecido, já que sua presença em um período prolongado pode deteriorar ou amarelar o tecido. O anticloro mais utilizado é o bissulfito de sódio (NaHSO_3).

ii) Beneficiamento secundário:

Os dois processos de beneficiamento secundário são: tingimento e estampagem, que são processos capazes de atribuir cor ao tecido, agregando valor ao produto (BELTRAME, 2000; IMMICH, 2006).

A diferença básica nos dois processos é que na tinturaria a atribuição de cor ao material é uniforme, por toda a extensão do mesmo, sendo a água o veículo de transporte entre o corante e a fibra. A característica mais relevante do tingimento é o elevado grau de fixação da cor tanto no substrato novo quanto após uso prolongado, ao se submeter à ação da luz, lavagem ou transpiração (SPECK, 2005; DAMASCENO, 2008; FEAM, 2013).

Ainda de acordo com os autores acima, na estamparia existe uma variedade de técnicas e tipos de equipamentos que podem ser utilizados. As mais utilizadas e difundidas são a estamparia em cilindro rotativo, onde o meio de transporte para o substrato têxtil é uma pasta (viscosa) contendo corantes ou pigmentos e a estamparia digital que utiliza cabeçotes de impressão contendo corante, que são comandados por computadores através da tecnologia CAD (Computer-Aided Design).

iii) Beneficiamento terciário:

O beneficiamento terciário ou acabamento final confere as propriedades finais dos produtos têxteis, uma vez que os tecidos devem ser submetidos a uma série de processos que visam conferir aspectos aos mesmos que atendam aos desejos dos clientes como mais brilho, melhor estabilidade dimensional e melhor toque e aparência, além de características especiais, como resistência ao fogo, tratamento antibactericida, repelência à água, dentre outras, tornando o material mais nobre que antes da passagem por essa etapa (IMMICH, 2006; FEAM, 2013).

Dentre alguns processos, pode-se citar: calandragem, sanforização, impregnação; flanelagem, amaciamento e impermeabilização (FEAM, 2013).

2.1.2.

Limpeza Industrial

O reprocessamento de fios ou tecidos no ambiente têxtil é uma técnica amplamente empregada para a obtenção de estopas (emaranhado de fios e fibras), toalhas (tecido plano entrelaçado, normalmente, em estrutura tafetá) e panos industriais (tecido de malha). Esses fios ou tecidos são utilizados para a limpeza industrial, desde a remoção de fluídos (óleos, graxas, água), partículas sólidas (sobras do artigo confeccionado, pós, resíduos) ou qualquer outro componente indesejável no ambiente fabril, no maquinário ou nas peças manufaturadas (FERREIRA, 2010).

Ainda de acordo com o autor acima, nos anos 80, entretanto, as toalhas industriais ganhavam força com um produto diferenciado, que são partes de tecido, personalizadas ou não e que são reutilizáveis, pois podem ser recicladas e reutilizadas, sofrendo apenas um processo de lavagem para remoção das sujidades e contaminantes existentes (alvejamento).

Atualmente, um novo produto começa a ganhar força o *wipes* (ou nãotecidos), já que pode ser utilizado com qualquer tipo de fibra com adição de elementos, fazendo um bem mais versátil, customizado, de baixo custo e mais ecológico, sendo uma alternativa economicamente viável para este segmento.

Após a apresentação dos conceitos referentes ao beneficiamento têxtil e as informações referentes ao setor de limpeza industrial. Será apresentado a seguir o conjunto de técnicas estatísticas denominadas de análise de regressão linear.

2.2.

Regressão Linear

Segundo Vieira (2004), quando se afirma que a resposta y depende dos fatores, isto quer dizer que existe uma relação funcional entre y e x_1, x_2, \dots, x_k , como mostra a **Equação 1**.

$$y = \Phi(\beta_0, \beta_1, \beta_2, \dots, \beta_k, x_1, x_2, \dots, x_k) + e \quad (Eq. 1)$$

Onde: $\beta_0, \beta_1, \dots, \beta_k$ são os coeficientes desconhecidos.

e é o erro e representa outras fontes de variabilidade, que não foram contabilizadas em Φ . Assim, e acumula efeitos tais como erros de medida e outras fontes de variabilidade inerentes ao processo, que são denominadas de ruídos (ou fatores incontroláveis).

De acordo com Vieira (2004) e Pimenta (2006), geralmente, não se conhece essa relação funcional e , portanto, se está interessado em desenvolver uma equação de interpolação para a variável resposta (y) do experimento. Esta é um modelo empírico do processo estudado e seu ajuste é denominado de análise de regressão, como mostra a **Equação 2**.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e \quad (Eq. 2)$$

Onde: $p = k + 1$ parâmetros desconhecidos $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ são os coeficientes do modelo de regressão linear.

Cabe registrar que o modelo é dito linear porque é uma função linear dos coeficientes. Já os modelos que são mais complexos podem ser representados através do exemplo mostrado na **Equação 3** a seguir.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + e \quad (Eq. 3)$$

Se fizermos $x_3 = x_1^2$, $x_4 = x_2^2$, $x_5 = x_1 x_2$, $\beta_3 = \beta_{11}$, $\beta_4 = \beta_{22}$ e $\beta_5 = \beta_{12}$, o modelo se torna um modelo linear de regressão como mostra a **Equação 4**.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + e \quad (Eq. 4)$$

Os polinômios do primeiro grau nas variáveis de regressão são usados em experimentos fatoriais em dois níveis, completos (2^k) ou fracionados (2^{k-p}), e polinômios do segundo grau são usados em experimentos fatoriais em três níveis; completos (3^k) ou os denominados Experimentos Compostos Centrados (*Central Composite Designs*).

2.2.1.

Planejamento de Experimentos

Um experimento nada mais é que um teste onde alterações intencionais são feitas nas variáveis de entrada do processo (ou sistema), de forma que se possa observar e identificar as razões para as mudanças ocorridas na(s) sua(s) variável(is) resposta(s). Experimentos são realizados em praticamente todos os campos de investigação, como: indústria química, farmacêutica, agrícola, têxtil, dentre outras. E, geralmente, têm o propósito de obter informações acerca de um processo particular (BARROS NETO; SCARMÍNIO; BRUNS, 1995; BARROS NETO; SCARMÍNIO; BRUNS, 2007).

Usando o planejamento de experimentos, os pesquisadores conseguem extrair do sistema em estudo o máximo de informações úteis, fazendo um número mínimo de experimentos. Consequentemente, consegue-se melhorar ou aperfeiçoar sistemas, processos e produtos, podendo assim diminuir o número de ensaios necessários, agilizando o processo e reduzindo custos (BONDUELLE, 2000; MONTGOMERY; CALADO, 2003; MONTGOMERY, 2009).

De acordo com Montgomery (2009), os processos podem ser compreendidos como a combinação de métodos, pessoas e outros recursos que transformam algumas entradas, geralmente, um determinado produto em uma saída, que pode ter uma ou mais respostas observáveis.

A **Figura 3** representa um processo (ou experimento) qualquer, em que as variáveis x_1, x_2, \dots, x_k representam as variáveis de entrada (fatores controláveis) e z_1, z_2, \dots, z_q representam os ruídos (fatores incontroláveis), que são características experimentais não controladas, seja pelo desconhecimento de sua existência ou pelo alto custo para controlá-las, ou seja, tais variáveis constituem o erro do experimento (MONTGOMERY; RUNGER, 2012).

É importante evitar que os efeitos produzidos pelos fatores controláveis, fiquem misturados ou mascarados com os efeitos provocados pelos fatores incontroláveis (GALDÁMEZ, 2002).

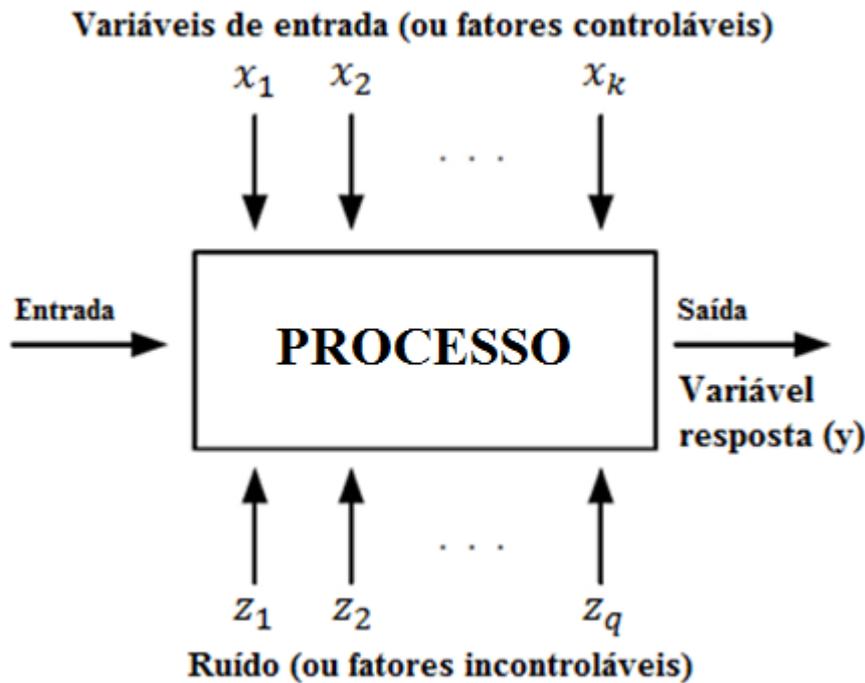


Figura 3 - Representação de um experimento

Fonte: Adaptado de MONTGOMERY e RUNGER (2012)

Observar-se que uma variável de entrada é considerada controlável se os valores que ela assumir, denominados de níveis, for definido antes do início dos experimentos. As variáveis de entrada controláveis que são de interesse em serem investigadas pelo pesquisador, comumente, são denominadas de fatores, cuja variação pode ou não influenciar a resposta final do processo, seja sozinha ou interagindo com uma ou mais variáveis do mesmo processo. As combinações possíveis entre os níveis dos fatores são denominadas de tratamento (MONTGOMERY, 2009; BORTOLINI, 2012).

É importante ressaltar que nem todos os fatores afetam o desempenho da mesma forma, isto é, alguns fatores podem apresentar fortes influências, enquanto que outros podem nem ter efeito na variável resposta (y) e poderão ser descartados *à posteriori* (ANTONY, 2003).

Em síntese, segundo Costa (2008), Gomes (2007) e Pimenta (2006), para que se possa aplicar a metodologia de experimento em qualquer processo, alguns requisitos básicos tornam-se necessários e serão descritos a seguir:

- i) Reconhecer e definir o problema do processo que se deseja estudar;*
- ii) Definir as variáveis que podem afetar este processo;*
- iii) Identificar quais destas variáveis podem ser controladas e escolher quais destas variáveis serão utilizadas nos experimentos, caso não seja viável a utilização de todas;*
- iv) Selecionar uma ou mais variáveis de resposta (variáveis de interesse) do processo que serão utilizadas para análise dos experimentos;*
- v) Determinar a faixa de observação em que cada uma das variáveis selecionadas será analisada;*
- vi) Escolher o tipo de planejamento de experimentos que será utilizado e que melhor se adeque à situação;*
- vii) Executar os experimentos, monitorando-os e controlando-os;*
- viii) Analisar os resultados, utilizando métodos estatísticos;*
- ix) Elaborar as conclusões e recomendações.*

Entretanto, para que todos esses requisitos sejam atendidos é necessário que se tenha um bom conhecimento do processo em estudo. Geralmente, este conhecimento é uma combinação entre a experiência prática e a teórica (MONTGOMERY, 2009).

De acordo com Bonanni (2005), as potenciais aplicações do planejamento de experimentos nas indústrias são:

- i) Redução do tempo de projeto e desenvolvimento de produto e processo;*
- ii) Estudo do comportamento de um processo sobre uma ampla variedade de condições operacionais;*
- iii) Entender o processo em estudo e assim melhorar seu desempenho;*
- iv) Aumentar a produtividade do processo reduzindo retrabalho, perdas, etc;*
- v) Melhorar o rendimento e estabilidade de um processo industrial em andamento;*
- vi) Fazer produtos insensíveis à variações ambientais, tais como umidade relativa, vibração, temperatura, dentre outras;*
- vii) Estudar a relação entre as variáveis independentes do processo (fatores controláveis) e sua variável-resposta.*

2.2.1.1.

Seleção do Planejamento

Contudo, como em todo experimento, além da necessidade de seguir um roteiro de execução e diagnóstico, existem princípios básicos que regem sua elaboração, implementação e análise. Segundo Montgomery (2009), os princípios básicos de um bom planejamento experimental são:

i) *Replicação: é o processo de repetição do experimento básico, o que permite ao pesquisador obter estimativas de como o erro experimental afeta os resultados dos ensaios e se esses resultados são estatisticamente diferentes. Outro ponto importante é que se a média da amostra é utilizada para estimar o efeito de um fator no experimento, a replicação também permite obter uma estimativa mais precisa desse efeito. No entanto, quanto mais réplicas, maior é o custo do experimento.*

ii) *Aleatorização: tanto a alocação do material quanto a ordem na qual são realizadas as corridas individuais do experimento são determinadas aleatoriamente, o que faz com que as variáveis estudadas e os erros experimentais apresentem caráter aleatório. Contudo, em alguns casos, por restrições de custo na experimentação e pela existência de fatores cuja mudança dos níveis é limitada ou difícil nem sempre isso é possível.*

2.2.1.2.

Planejamento Fatorial Com Ponto Central

Quando se deseja estudar dois ou mais fatores de interesse em um experimento, um planejamento fatorial deve ser utilizado. Seu intuito é permitir uma combinação de todas as variáveis em todos os níveis, obtendo-se assim uma análise de uma variável, sujeita a todas as combinações das demais. Também é única maneira possível de prever interação entre os fatores (MONTGOMERY; RUNGER, 2012).

O usual é realizar um planejamento com dois níveis, no máximo três, já que o uso de mais variáveis aumentaria de forma significativa o número de pontos experimentais, fato esse que se quer evitar quando se propõe um planejamento (MONTGOMERY; CALADO, 2003).

Pode-se ainda incluir pontos centrais aos experimentos fatoriais com o intuito de viabilizar o cálculo do erro experimental e oferecer uma proteção contra a curvatura e, conseqüentemente, permitindo uma estimativa independente do erro e a verificação de falta de ajuste para o modelo escolhido, com o menor número de ensaios. Outra razão importante para adicionar réplicas ao centro do planejamento é que eles não repercutem nas estimativas usuais dos efeitos em um planejamento 2^k (MONTGOMERY; RUNGER, 2012).

Barros Neto, Scarminio e Bruns (2007) ressaltam que duas repetições para o ponto central seriam suficientes para o cálculo do erro experimental e que a escolha de três ou mais repetições tem como intuito obter uma estimativa mais precisa do erro.

2.2.2.

Modelo Linear Clássico

Segundo Vieira (2004), o modelo linear clássico ($Y = \mu + \epsilon$), como o Método dos Mínimos Quadrados (MMQ), tem sido a técnica estatística mais amplamente difundida para estabelecer a relação entre as variáveis independentes de um experimento. A sua simplicidade, seu apelo intuitivo e suas propriedades explicam a difusão de seu uso e serão citadas e detalhadas a seguir.

i) *A resposta tem distribuição Normal: se a variável resposta é Normal, a distribuição de probabilidade dos estimadores de Mínimos Quadrados Ordinários (MQO) pode ser facilmente derivada, porque uma das propriedades dessa distribuição é que qualquer função linear de variáveis com distribuição Normal também é normalmente distribuída. Por conseguinte, se os u_i estiverem normalmente distribuídos, $\hat{\beta}_1$ e $\hat{\beta}_2$ também o estarão, o que facilita a tarefa de testar as hipóteses* (GUJARATI; PORTER, 2011). Segundo Levine (2011), para verificar esta premissa pode se utilizar o gráfico de probabilidade Normal com envelope.

ii) *A variância da resposta é constante (homocedasticidade): para verificar se a variância da variável resposta é constante ou não, isto é, para verificar o pressuposto de homocedasticidade do modelo tem-se o gráfico dos resíduos versus os valores ajustados (valores previstos) ou o gráfico dos resíduos versus os fatores (variáveis explicativas)* (GUJARATI; PORTER, 2011).

iii) Os efeitos dos fatores sobre a variável resposta se combinam aditivamente: quando uma ou mais dessas premissas são violadas, a confiabilidade de todos os testes paramétricos, tais como a ANOVA, a comparação de médias e a análise de regressão ficam comprometidas, uma vez que ocorrem alterações na probabilidade de ocorrência do erro tipo I (rejeição da H_0 quando ela é verdadeira) e tipo II (não rejeição da H_0 quando ela é falsa), o que pode levar a falsas conclusões a respeito dos efeitos de tratamento (STEEL; TORRIE; DICKEY, 1997; STORCK *et al.*, 2000; MARTIN; STORCK, 2008).

Essas condições são encontradas na maioria das aplicações industriais, porém nem sempre isso ocorre. Quando a variável resposta apresentada é uma proporção (ou fração) de algum evento de interesse, nem sempre se recomenda a utilização de um modelo de regressão linear clássico, pois este requer suposição de normalidade e homocedasticidade, o que nem sempre ocorre em dados na forma de proporção. Uma alternativa viável é alterar a escala da variável resposta, transformando-a e aplicando o MMQ aos dados transformados, o que muitas vezes pode ser satisfatório. Entretanto, nem sempre se consegue estimativas dos parâmetros confiáveis e para estes casos utilizam-se os MLGs (COX, 1996; BELLO, 2010).

2.2.3.

Análise dos Resíduos

A investigação dos resíduos ($e_i = y - \bar{y}_i$) é uma etapa obrigatória de qualquer análise de regressão. Se o modelo é adequado, os resíduos devem se apresentar de forma aleatória, isto é, eles não devem conter nenhum padrão evidente. Desta forma, a verificação do modelo pode ser realizada pela análise de gráficos dos resíduos (VIEIRA, 2004).

A importância de elaborar e analisar gráficos como parte da análise estatística deve ser destacada, pois fornece um resumo simples para entender um problema complexo. Embora este método seja poderoso e sugestivo, ele é de natureza subjetiva (ou qualitativa), mas podem ser apoiados por diversos testes quantitativos que completam a abordagem puramente qualitativa (GUJARATI; PORTER, 2011 *apud* WEISBERG, 1980).

2.2.4.

Adequação do Modelo Clássico

Segundo Montgomery e Runger (2012), o MMQ, tradicionalmente denominado de Mínimos Quadrados Ordinários (MQO), é o método clássico de estimação dos parâmetros dos modelos lineares. Suponha que foram realizadas n observações da variável resposta, y_1, y_2, \dots, y_n , em conjunto com cada observação de y tem-se uma observação ou nível, de cada variável de regressão como mostra a **Equação 5**.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + e_i, \quad i = 1, 2, \dots, n \quad (Eq. 5)$$

Assume-se que os diversos valores do termo do erro, e_i , sejam variáveis aleatórias não correlacionadas, com média zero e variância constante σ^2 , isto é:

$$E(e_i) = 0 \text{ e } E(e_i e_j) = \begin{cases} 0, & i \neq j \\ \sigma^2, & i = j \end{cases}$$

A **Equação 5** pode ser reescrita na forma matricial como mostra a **Equação 6**.

$$y = X\beta + e \quad (Eq. 6)$$

O modelo de regressão ajustado é dado pela **Equação 7**.

$$\hat{y} = X\hat{\beta} \quad (Eq. 7)$$

Na forma escalar, o modelo é descrito pela **Equação 8**.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik} \quad (Eq. 8)$$

O estimador de mínimos quadrados de β é um estimador linear não enviesado e de variância mínima, o que lhe confere o título de melhor estimador linear não enviesado. Portanto, $E(\hat{\beta}) = \beta$ (MYERS *et al.*, 2002).

A estimativa da variância σ^2 do erro e é dada pela **Equação 9**.

$$\hat{\sigma}^2 = \frac{SQ_E}{n - p} \quad (Eq. 9)$$

Onde: * SQ_E é a soma dos quadrados dos resíduos: $SQ_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$.

A diferença entre a observação y_i e o valor ajustado \hat{y} constitui o resíduo da regressão, $e_i = y_i - \hat{y}$ (MONTGOMERY; RUNGER, 2012).

Existem alguns testes úteis para verificar quais os parâmetros significativos do modelo, um deles consiste em verificar se os erros e_i têm distribuição Normal e são independentes com média zero e variância constante. Por consequência, as observações y_i têm distribuição Normal e são independentes com média igual a $\beta_0 + \sum_{j=1}^k \beta_j x_{ij}$ e variância igual a σ^2 (VIEIRA, 2004).

A validade do modelo utilizado é constatada por meio da Análise de Variância (ANOVA), isto é, através dela é possível aceitar ou rejeitar, estatisticamente, as hipóteses investigadas do experimento. O objetivo dessa técnica é analisar a variação média dos resultados dos testes e demonstrar quais são os fatores que realmente produzem efeitos principais e de interação significativos na variável resposta (CUNHA, 2005).

Os parâmetros que possuem $F_{\text{CALCULADA}}$ maior que a estatística $F_{\text{CRÍTICO}}$ são os fatores que exercem influência sobre o valor da média dos resultados. Esses dados podem ser avaliados, diretamente, pela análise do p-valor que corresponde à área sob qual a estatística F é limite da razão $F_{\text{CALCULADA}}$. Tendo em mãos o p-valor, não se faz necessário recorrer a uma tabela de valores críticos da distribuição F , isto é, se o p-valor for menor que o nível de significância escolhido, existe influência significativa desse fator sobre o processo (CUNHA, 2005; HINES *et al.*, 2006).

Uma forma de avaliar o ajuste do modelo proposto aos dados é através do coeficiente de determinação (R^2) que visa determinar a parcela da variabilidade amostral que foi, de fato, explicada pela reta de regressão. O R^2 varia entre 0 e 1 (ou 0% e 100%). Esta medida é problemática uma vez que ela sempre aumentará quando uma variável for adicionada ao modelo. Por este motivo, opta-se normalmente por utilizar o coeficiente ajustado que desconta o efeito de um elevado número de variáveis explicativas (MONTGOMERY; RUNGER, 2012).

Ainda de acordo com os autores acima, a estatística R^2_{ajustado} penaliza essencialmente o pesquisador pela adição de termos ao modelo. É uma maneira fácil de resguardar contra o ajuste em excesso, ou seja, a inclusão de regressores que não são realmente úteis.

Quando o número de parâmetros do modelo for maior que 1, o $R^2_{ajustado}$ é menor que o R^2 , o que implica que, à medida que o número de variáveis explicativas aumenta, o $R^2_{ajustado}$ aumenta menos que o R^2 . Vale ressaltar também que o $R^2_{ajustado}$ pode ser negativo, enquanto o R^2 só pode ser necessariamente não negativo (GUJARATI; PORTER, 2011).

Seja como for, o certo é que quanto mais próximo a 1 (independente do sinal) maior é o grau de dependência estatística linear entre as variáveis, isto é, o modelo é preditivo e pode ser usado para descrever significativamente os dados experimentais. E o oposto, quanto mais próximo de zero, menor é a força dessa relação (BIAZUS *et al.*, 2005; BARROS NETO, SCARMINIO; BRUNS, 2007; FIGUEIREDO FILHO; SILVA JÚNIOR, 2009).

2.2.4.1

Verificação de Normalidade

Um procedimento útil para verificação da consideração de normalidade é o gráfico de probabilidade normal dos resíduos, que é baseado no fato de que os efeitos principais (interação) são desprezíveis e se distribuem segundo uma Normal Padrão, com média zero e variância constante σ^2 , isto é, esses efeitos tendem a se concentrar ao longo de uma reta Normal no gráfico. Caso os pontos desviem-se de algum modo dessa linha imaginária, existem indícios para acreditar que esses dados obtidos não estão distribuídos de maneira Normal e, portanto, devem ser analisados com mais cuidado pelo pesquisador (LEVINE, 2011).

Para construir esse gráfico é necessário ordenar os resíduos em ordem crescente. Com os resíduos ordenados eles são plotados *versus* a frequência cumulativa $(j - 0,5)/n$ (VIEIRA, 2004).

O *software* ARC constrói este gráfico de outra maneira, isto é, ele constrói o gráfico de probabilidade normal com envelope. Atkinson (1985) afirma que, além do caráter subjetivo das análises nestes dois tipos de gráficos há o problema de super-normalidade. Portanto, o fato dos pontos estarem alinhados em torno de uma linha reta não significa necessariamente em normalidade da distribuição do erro. A construção de envelopes tenta superar estes dois problemas relatados acima (VIEIRA, 2004).

Atkinson (1985) afirma que o intuito do gráfico de probabilidade normal com envelope não é prover uma região de aceitação ou rejeição como em um teste formal, mas prover uma orientação sobre a forma ou linha que pode ser esperada deste gráfico. Mais do que o número de pontos fora do envelope, é importante o afastamento dos pontos em relação ao envelope, com especial atenção para os resíduos com valores mais elevados.

Segundo Vieira (2004), o desvio-padrão dos resíduos não é constante e é diferente para os diversos valores da variável de resposta y , sendo maior para respostas mais próximas da média desta variável. Os resíduos *studentizados* levam isto em consideração e são definidos pela **Equação 10**.

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1 - h_{ii})}} \quad (\text{Eq. 10})$$

Onde: $\hat{\sigma}^2$ é igual a média quadrática do erro (MQE).

h_{ii} é o i -ésimo elemento da diagonal da matriz H (é a matriz chapéu, isto é, é a matriz de projeção dos valores ajustados sobre os valores observados).

2.2.4.2.

Verificação de Independência

A suposição de independência dos resíduos $E(e) = 0$ é verificada através do gráfico dos resíduos *studentizado versus* o valor ajustado (ou previsto). Se o gráfico apresenta algum padrão evidente e a linha resultante do amortecimento (*lowess*) não é aproximadamente horizontal e próxima da reta horizontal de ordenada zero, têm-se indícios de que alguma variável que não foi levada em consideração no modelo influenciou os resultados do experimento (BELLO, 2010).

A linha de amortecimento *lowess* (*locally weighted scatterplot smoother*), ou linha amortecida, no gráfico de dispersão, localmente ponderada, é uma técnica estatística não paramétrica, indicada para visualizar tendências nos dados no gráfico (VIEIRA, 2004).

Uma grande vantagem dessa estatística é que ela se baseia nos resíduos estimados, que são calculados na análise de regressão. Porém esse teste possui uma grande desvantagem se cair na zona de indecisão, não se pode concluir se há ou não autocorrelação de primeira ordem (GUJARATI; PORTER, 2011).

2.2.4.3.

Verificação de Variância Constante

A suposição de variância constante σ^2 é verificada através do gráfico do valor absoluto dos resíduos *studentizado versus* os valores ajustados. Se a linha resultante do amortecimento (*lowess*) não indicar crescimentos da variância com o aumento da média esse pressuposto não é violado (BELLO, 2010).

2.2.4.4.

Verificação de Observações Atípicas (*Outliers*)

Uma observação atípica é aquela que não combina como modelo obtido. Essas observações atípicas podem dever-se a erros de medição da variável resposta, ou de transposição dos dados, ou de condução dos experimentos. Entretanto, elas só devem ser descartadas caso se confirme erro de medida ou transcrição; a resposta obtida pode não ser fruto de um erro, mas um valor real e, caso trate-se de um extremo da resposta, pode mesmo, dependendo do objetivo, corresponder a um bom (senão ao melhor) ponto de operação do processo produtivo. Ademais, elas podem ser fruto de um modelo inadequado. Uma observação pode ser atípica em um modelo e não a ser em outro (VIEIRA, 2004).

Ainda de acordo com o autor acima, o resíduo *studentizado* (r_i) é frequentemente considerado para um diagnóstico de observações atípicas. Vale lembrar que, para calcular os resíduos *studentizados* de cada dado experimental, utiliza-se a média dos quadrados do erro (MQ_E) como estimativa da variância ($\hat{\sigma}^2$). A MQ_E foi gerada internamente e obtida a partir do ajuste do modelo às n observações. Portanto, o resíduo *studentizado* representa uma escala interna dos resíduos.

Outro procedimento é considerar a exclusão da i -ésima observação e verificar qual é o efeito na estimativa da resposta i . Em particular, verificar se o valor observado y_i concorda com o valor ajustado $\hat{y}_{(i)}$, obtido quando a i -ésima observação é excluída da regressão, ou seja, faz-se a regressão com a i -ésima observação removida. Então, a estimativa de $\hat{\sigma}^2$ passa a ser $S_{(i)}^2$, como mostra a **Equação 11**.

$$S_{(i)}^2 = \frac{(n-p)MQ_E - e_i^2/(1-h_{ii})}{n-p-1} \quad (Eq. 11)$$

Onde: $*S_{(i)}^2$ é utilizada no lugar da MQ_E para gerar uma escala externa dos resíduos *studentizados* como mostra a **Equação 12**.

$$t_i = \frac{e_i}{\sqrt{S_{(i)}^2(1-h_{ii})}}, \quad i = 1, 2, \dots, n \quad (Eq. 12)$$

Myers *et al.* (2002) afirmam que o resíduo t_i usualmente é denominado R-Student, enquanto Atkinson (1985) denomina-o resíduo de supressão (*deletion residual*), e Cook e Weisberg (1999): *outlier-t*. O *software* ARC utiliza destes gráficos citados acima o último.

Caso observe-se valores entre o intervalo $(-3,5; 3,5)$, não há indicação de observações atípicas. Caso alguma observação fosse observada fora deste intervalo, procederíamos ao teste com a distribuição t (VIEIRA, 2004).

Atkinson (1985) sugere utilizar para os resíduos *outlier-t* um gráfico de probabilidade normal com envelope. Caso não se observe pontos muito fora do alinhamento não se tem indícios de observações atípicas.

2.2.4.5.

Verificação de Observações Influentes

Ocasionalmente algumas observações exercem grande influência na determinação dos coeficientes de regressão do modelo. Tais observações são denominadas observações influentes (VIEIRA, 2004).

Segundo Atkinson (1985), Biasoli (2005), Crusco *et al.* (2005) pode-se utilizar a distância de Cook para verificar a existência ou não de observações influentes no ajustamento do modelo. Esta distância avalia a influência do *i-ésimo* caso sobre os demais, analisando-se a diferença entre os outros casos.

A estatística calculada para distância de Cook, utilizando-se a regressão linear clássica é mostrada na **Equação 13**.

$$D_i = \frac{r_i^2 h_{ii}}{p(1 - h_{ii})^2} \quad (Eq. 13)$$

A distância de Cook provê uma ordenação das observações em termos da sua influência sobre o vetor das estimativas dos coeficientes. A intenção não é aplicar um teste formal, e sim fornecer uma ajuda para detectar as observações influentes. Cook e Weisberg (1999) afirmam que é conveniente analisar casos em que $D_i > 0,5$ e é sempre importante analisar casos em que $D_i > 1$. Esta análise consiste em verificar se a observação é realmente influente ou se é consequência de um modelo inadequado. Se o modelo for inadequado, deve-se construir outro modelo.

2.2.5.

Sobredispersão e subdispersão

Recentemente, com a modernização dos conceitos de qualidade, a variância mínima passou a ser sinônimo de qualidade e tornou-se alvo de interesse de diversos pesquisadores, como Myers *et al.* (2002), Vieira (2004), Mattos (2004) e Paula (2010).

Passou-se a utilizar métodos de identificação dos fatores que afetam a variância da variável resposta, isto é, fatores que afetam a variabilidade das observações em decorrência da influência de fatores incontroláveis (ruídos) no processo produtivo. Os efeitos produzidos por esses fatores são denominados efeitos na dispersão (*dispersion effects*) (VIEIRA, 2004; MATTOS, 2004).

Já os efeitos que afetam diretamente a média da variável resposta são denominados de efeitos de posição (*location effects*) (NAIR; PREGIBON, 1988).

A distribuição de Poisson assume, teoricamente, que a média e a variância são iguais, mas em situações reais nem sempre se verifica essa condição. No caso de dados na forma de proporções (distribuição Binomial), e de contagens, frequentemente, verifica-se uma sobredispersão dos dados (ADEWALE; XU, 2009). Isto é, a variância marginal de um atributo excede à teórica esperada pelo modelo para a distribuição escolhida (HINDE; DEMÉTRIO, 1998; SANTORO; BARBOSA; HOLANDA, 2003; GONÇALVES, 2012).

Já quando a variância marginal de um atributo é inferior à teórica esperada para a distribuição escolhida, ocorre uma subdispersão dos dados. Caso ocorra uma sobredispersão ou uma subdispersão dos dados é necessário alterar a variância, introduzindo um parâmetro de escala, denominado de parâmetro de dispersão (ϕ), que incorpora essa variabilidade extra, permitindo a inferência do modelo proposto (MYERS, *et al.*, 2002; SOUZA, 2013).

Segundo McCullagh, Nelder (1989), Hinde, Demétrio (2007) e Hilbe (2008) a sobredispersão ou subdispersão pode verificar-se por diversas razões. As mais comuns são:

- i) *Agrupamentos (ou clusters) não identificados;*
- ii) *Modelo inadequado ou com variáveis importantes não incluídas;*
- iii) *Pontos atípicos (outliers);*

iv) *Variabilidade dos dados experimentais, isto ocorre quando os dados analisados provêm de uma população heterogênea. Tais como contagens de eventos para setores censitários, que tendem a ter populações que socialmente, economicamente e demograficamente são heterogêneos.*

Os efeitos de dispersão são detectados mais facilmente em experimentos com muitas replicações, onde em cada condição experimental podem ser encontradas as médias assim como as variâncias amostrais. Nestes casos, a identificação dos efeitos de posição (*location effects*) pode ser feita através de técnicas estatísticas formais, como o teste t de *Student* ou o teste F da ANOVA. Em experimentos não replicados ou com poucas replicações é comum a utilização de métodos gráficos, pois não existe uma estimativa confiável da variância do erro

experimental, que é essencial para a aplicação de um teste estatístico (MATTOS; BARBETTA; SAMOHYL, 2002).

Montgomery (2009) descreve um método gráfico bastante simples para identificar os efeitos na dispersão, através da análise do gráfico dos resíduos *versus* os fatores. Este método é útil para uma primeira inspeção do efeito de um fator sobre a dispersão. A análise é feita observando se há mudanças no espalhamento dos resíduos quando o nível do fator é mudado. Quando a mudança é visível, isto é indício de que a variabilidade é sensível ao fator em questão.

De acordo com Vieira e Epprecht (2009), em alguns casos, essa análise gráfica pode fornecer indicações conclusivas, ou seja, pode fornecer alguma evidência dos fatores que afetam a dispersão. Porém vale ressaltar que como essa análise é subjetiva e não é considerado um teste formalmente, em muitos casos, não encontraremos indicações conclusivas.

Não levar em conta a existência de sobredispersão ou subdispersão na análise dos dados pode levar à estimação incorreta dos erros padrão das estimativas e, conseqüentemente, uma avaliação incorreta da significância individual dos parâmetros da regressão, isto é, pode-se acreditar que uma variável preditora possa ser significativa, quando na realidade não é (HILBE, 2008).

Ajustes razoáveis podem ser feitos com o intuito de corrigir esses erros. Em regressão logística, em que o tamanho da amostra em cada grupo é razoavelmente grande, pode-se estimar o parâmetro de escala pelo desvio significativo, que é análoga ao Erro Médio Quadrado (EMQ) em regressão linear simples. Outra estimativa que é tão intuitiva é a estatística de Pearson (MYERS *et al.*, 2002) que é mostrada na **Equação 14**.

$$\frac{1}{n-p} \sum_{i=1}^m \left[\frac{(y_i - \hat{y}_i)^2}{n_i \hat{\pi}_i (n_i \hat{\pi}_i (1 - \hat{\pi}_i))} \right] = \frac{\chi^2}{n-p} \quad (\text{Eq. 14})$$

2.3.

Modelos Lineares Generalizados

A seleção de modelos é uma parte importante de toda pesquisa em modelagem estatística e envolve a procura de um modelo que seja o mais simples possível e que descreva bem o processo estudado que surgem em diversas áreas do conhecimento como: agricultura, demografia, ecologia, economia, engenharia, geologia, medicina, ciência política, sociologia e zootecnia, dentre outras (CORDEIRO, DEMÉTRIO, 2013).

Os Modelos Lineares Generalizados (*Generalized Linear Models - GLM*) foram introduzidos pela primeira vez por Nelder e Wedderburn (1972). Conforme Davis (2002), Sant'anna, Caten (2010), Pinto, Ferreira (2012), Cordeiro e Demétrio (2013), estes modelos se baseiam em distribuições de probabilidade, com um parâmetro de localização desconhecido, admitindo que a mesma pertença à família exponencial. Esta família contempla as distribuições de probabilidade: Normal, Normal Inversa, Binomial, Multinomial Poisson, Gama, Exponencial e Binomial Negativa.

A técnica de MLG permite a generalização ou a flexibilização dos modelos lineares clássicos de variáveis contínuas, de forma que toda a estrutura para a estimação e predição em modelos lineares clássicos, possa ser estendido para os modelos não lineares. Vale lembrar que os modelos lineares clássicos são, na verdade, casos especiais de MLG (NELDER; WEDDERBURN, 1972).

Essa técnica também apresenta opções para a distribuição da variável dependente, permitindo que dados provenientes de uma distribuição de probabilidade Binomial, por exemplo, possam ser modelados usando a distribuição original dos dados, sem a necessidade de realizar transformações nos mesmos (MYERS *et al.*, 2002).

Através de uma única transformação nem sempre se consegue atingir a normalidade, a equalização da variância e a aditividade dos efeitos. Por conseguinte, para estes casos devem-se considerar os MLGs. Eles apresentam algumas vantagens em relação às transformações das variáveis respostas, como um melhor desempenho na estimativa da mesma, propiciando intervalos de confiança menores e obtenção de linearidade através de uma função de ligação entre a média da variável resposta e o polinômio linear das variáveis independentes, sugerindo um modelo com mais eficiência na predição e estimação

dos parâmetros (LEWIS; MONTGOMERY; MYERS, 2001a e 2001b; BIASOLI, 2005).

Os autores acima concluíram que os MLG devem ser preferidos à transformação da variável resposta, uma vez que em todos os casos os MLG apresentaram uma estimativa da resposta mais confiável. Biasoli (2005) ressalta ainda que apesar de suas inúmeras vantagens sua modelagem é mais complexa, uma vez que além de identificar a distribuição de probabilidade dos dados é necessário também determinar a função de ligação mais adequada.

2.3.1.

Estrutura do MLG

Conforme Lee e Nelder (1998) ver, também, Sant'anna, Caten (2010) e Cordeiro e Demétrio (2013), esta classe de modelos é definida ainda por um conjunto de fatores que descreve a estrutura linear do modelo e uma função de ligação entre a média da variável dependente e a estrutura linear. Isto é, os MLG apresentam uma estrutura com três componentes:

i) Componente aleatória: *identifica a distribuição de probabilidade da variável resposta (y) pertencente à família exponencial, que engloba as distribuições Normal, Gama e Normal Inversa para dados contínuos, Binomial para proporções, Multinomial, Poisson e Binomial Negativa para contagens. É estabelecida assim que as variáveis são definidas, sejam elas contínuas ou discretas, exigindo-se apenas o ajuste de distribuições diferentes. A partir de um mesmo experimento pode-se obter medidas de diferentes tipos, como por exemplo, dados de altura de plantas, número de lesões por planta e proporção de plantas doentes, dentre outras;*

ii) Componente sistemática: *as variáveis independentes (x_1, x_2, \dots, x_k) entram na forma de um modelo linear (η), constituindo a parte sistemática do modelo. É estabelecida durante o planejamento de experimento (de suma importância para a obtenção de conclusões confiáveis), resultando em modelos de regressão (simples, múltipla, não linear, etc.), de Análise de Variância (ANOVA) e de análise de covariância;*

iii) Função de ligação: descreve a relação funcional entre a componente sistemática e o valor esperado da componente aleatória, isto é, faz-se a conexão entre a média das observações e a parte sistemática. De acordo com Asevedo (2011), a função de ligação transforma μ_i , a média de y_i , e não a resposta, o que é uma grande vantagem já que possibilita análise direta das estimativas do modelo, evitando a necessidade de utilizar a transformação inversa (que em muitos casos são bastante complicadas) nos valores estimados.

Exemplificando, o que foi descrito acima, considere um experimento com n respostas independentes y_i como mostra a **Tabela 1** (BELLO, 2010).

Tabela 1 - Dados para o MLG

x_1	x_2	...	x_k	y
x_{11}	x_{12}	...	x_{1k}	y_1
x_{21}	x_{22}	...	x_{2k}	y_2
\vdots	\vdots	...	\vdots	\vdots
x_{n1}	x_{n2}	...	x_{nk}	y_n

Tem-se:

1. Sejam y_1, y_2, \dots, y_n as variáveis respostas com médias $\mu_1, \mu_2, \dots, \mu_n$.
2. A distribuição de probabilidade de y_i é um dos membros da família Exponencial.
3. A porção sistemática do modelo é composta pelas variáveis de regressão x_1, x_2, \dots, x_k .
4. O modelo é construído com um preditor linear: $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$. Cabe registrar que uma variável x_i pode representar um termo de um grau mais elevado, como exemplo, x_1^2 .
5. A função de ligação $g(\mu_i)$ faz a ligação entre a média μ_i e o preditor linear. Ela define a forma com que os efeitos sistemáticos de x_1, x_2, \dots, x_k são transmitidos para a média, como é mostrado a seguir: $\eta_i = g(\mu_i) = (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$. Existem ainda outras funções de ligação não lineares, por exemplo, $g(\mu_i) = \log(\mu_i)$, que podem ser utilizadas.

É importante ressaltar que a função de ligação transforma μ_i , a média de y_i , e não a variável resposta.

A função de ligação é denominada canônica quando $\eta_i = \theta_i$. Segundo Myers e Montgomery (2002) a utilização da função de ligação canônica possui algumas propriedades atraentes, mas vale ressaltar que isso não significa que ela deva ser utilizada sempre e nem que ela proporcione em qualidade de ajuste do modelo. A sua escolha é conveniente porque não só simplifica as estimativas de máxima verossimilhança dos parâmetros do modelo, mas, também, o cálculo do Intervalo de Confiança (IC) para a média da resposta.

A **Tabela 2** mostra as funções de ligações canônicas para algumas distribuições de probabilidade.

Tabela 2 - Funções de ligação canônica para distribuições de probabilidade

DISTRIBUIÇÃO DE PROBABILIDADE	LIGAÇÃO CANÔNICA	NOME DA LIGAÇÃO
Normal	$\eta_i = \mu_i$	Identidade
Binomial	$\eta_i = \ln\left(\frac{\mu}{1-\mu}\right)$	Logística
Poisson	$\eta_i = \ln(\mu_i)$	Logarítmica
Exponencial ou Gama	$\eta_i = \frac{1}{\mu_i}$	Recíproca
Normal Inversa	$\eta_i = \frac{1}{\mu_i^2}$	Recíproca ao Quadrado

FONTE: Adaptado de MYERS *et al.* (2002) e MCCULLAGH e NELDER (1989)

6. Para os membros da família exponencial, a variância da resposta tem a seguinte expressão: $V(y_i) = \phi V(\mu_i)$, onde $V(\mu_i)$ é a parte da variância da resposta y que depende da média e ϕ é o parâmetro de dispersão que não depende da média e é constante. Vale ressaltar que a relação da função de variância $V(\mu_i)$ com a média pode ser representada por uma função de potência, isto é, $V(\mu_i) = \mu_i^t$.

Como o modelo de variância é em função da média, uma das causas da sobredispersão ou subdispersão de y é a definição incorreta da função de média, devido à falta de covariáveis ou a presença de uma relação não linear e/ou omissão de efeitos de interação entre algumas das covariáveis. A sobredispersão ou subdispersão também pode surgir como consequência de imprecisões de dados ou a escolha de um modelo de probabilidade inadequado, como selecionar o modelo de Poisson, quando a distribuição binomial negativa seria melhor captar a variação (HAINING; LAW; GRIFFITH, 2008).

2.3.2.

Regressão Logística

Segundo Cox (1996), nos casos em que a componente observacional é a proporção (ou fração) de itens não conformes, a modelagem pelo método regressão linear clássico, nem sempre é recomendada, uma vez que este modelo requer a suposição de que as proporções sigam uma distribuição Normal, o que para Kieschnick & McCulloch (2003) é um modelo falho, pois possibilita a previsão de valores fora do limite do intervalo $[0,1]$. Uma alternativa viável para a modelagem da proporção de itens não conformes são os MLG.

Na prática, quando o caso relatado acima ocorre existe uma evidência empírica considerável, que indica que a forma da função da resposta deve ser não linear e monotonicamente crescente (ou decrescente) como mostra a **Figura 4(a) e (b)** (MONTGOMERY; RUNGER, 2012).

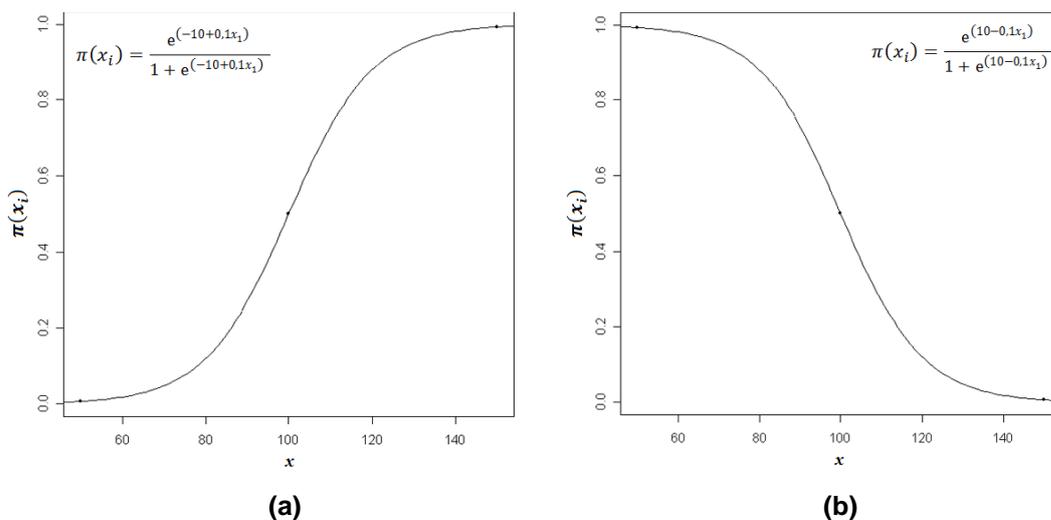


Figura 4 – Exemplo da função de resposta Logística: (a) monotonicamente crescente e (b) monotonicamente decrescente

Fonte: ESTATCAMP (2011)

2.3.2.1

Modelo Logit

De acordo com McCullagh e Nelder (1989), a estrutura probabilística da variável resposta (Y_i) relatada acima é representada pela distribuição Binomial a seguir:

$$Y_i | X_i = x_i \stackrel{\text{indep}}{\sim} \text{Binomial}(m_i \pi_i, m_i \pi_i (1 - \pi_i))$$

De acordo com Demétrio (2002), Myer *et al.* (2002), Barreto *et al.* (2012), esta função monotonicamente crescente ou decrescente denomina-se função de resposta Logit e é dada pela **Equação 15**.

$$E(Y) = \pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}} \quad (\text{Eq. 15})$$

Onde: $\pi(x_i)$ é a probabilidade estimada;

* x_i 's são as variáveis independentes (ou explicativas);

* β_i 's são os coeficientes estimados pela regressão Logit.

A **Equação 15** pode ser escrita, equivalentemente, pela **Equação 16**.

$$\pi(x_i) = E(Y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}} \quad (\text{Eq. 16})$$

Conforme Myer *et al.* (2002), Montgomery, Runger (2012) e Agresti (2012) o preditor linear para a função de ligação canônica Logit é dada pela **Equação 17**.

$$\eta_i = g(\pi(x_i)) = \ln \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right) = \ln \left(\frac{\frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}}{1 - \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n}}} \right)$$

$$\eta_i = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n \quad (\text{Eq. 17})$$

Onde: $g(\pi(x_i))$ é a função de ligação;

* $\frac{\pi(x_i)}{1 - \pi(x_i)}$ é denominada de razão de chances (*odds ratio*), isto é, é a razão entre probabilidade de sucessos $\pi(x_i)$ face a probabilidade de insucesso $1 - \pi(x_i)$.

Note que o logaritmo natural da razão de chances é uma função linear da variável regressora. Por conseguinte, se exponenciarmos ambos os lados da função Logit, a razão de chances será uma função exponencial de x (VITTINGHOFF *et al.*, 2012).

Quando assumi-se a presença de algum tipo de variabilidade no parâmetro π_i a distribuição de probabilidade tem média μ e variância $\phi > 0$. Então se y_i é uma variável aleatória Binomial seu valor esperado é dado pela **Equação 18**.

$$E(y_i) = n\mu \quad (\text{Eq. 18})$$

E sua variância é dada pela **Equação 19**.

$$V(y_i) = n\mu(1 - \mu) + n\phi(n - 1) \quad (\text{Eq. 19})$$

De acordo com Haining, Law e Griffith (2008), observando-se a **Equação 20**, pode-se encontrar os seguintes resultados:

$$V(y_i) = \phi V(\mu_i) \quad (\text{Eq. 20})$$

i) Se $\phi = 1$, a variância de Y_i reduz-se a variância de uma variável aleatória Binomial.

ii) Se $\phi > 1$, a variância de Y_i apresenta algum grau de sobredispersão (overdispersion) ou variação “Extra-Poisson”.

iii) Se $\phi < 1$, a variância de Y_i apresenta algum grau de subdispersão (underdispersion).

2.3.3.

Adequação do MLG

Assim como em regressão linear, nos MLG, deve-se identificar tanto os fatores significativos ao modelo quanto os seus resíduos, pois são analisados e utilizados para verificar a adequação do modelo.

2.3.3.1.

Teste de Significância dos Coeficientes

Alguns *softwares* fornecem a estatística t_0 para testar quais parâmetros são estatisticamente significativos em modelos não Normais. Um valor elevado de t_0 , maior do que três, é uma indicação de significância, em geral, para qualquer distribuição de probabilidade. Por outro lado, um valor pequeno de t_0 , menor do que um, é uma indicação de não significância, em geral, para qualquer distribuição de probabilidade. Em alguns casos pode-se avaliar o p-valor ao invés da estatística t_0 , pois ambos fornecem o mesmo resultado, porém utilizando-se o p-valor não há necessidade de se consultar uma tabela com os valores críticos da distribuição t (VIEIRA, 2004).

A distribuição t é uma aproximação para a distribuição da estatística, aproximação esta que pode não ser boa mesmo para amostras grandes, o que pode ocasionar resultados errôneos (LINDSEY, 1997).

Para os testes de hipóteses sobre os coeficientes individuais do modelo pode-se utilizar o teste de Wald, que é um teste bastante simples operacionalmente, podendo ser utilizado para avaliar se a variável preditora é estatisticamente significativa ou não. Através dele pode-se retirar do modelo aquelas variáveis identificadas como estatisticamente não significativas, reavaliando-se novamente o modelo gerado (HOSMER; LEMESHOW, 2000; CRUZ, 2001).

Segundo Turkman e Silva (2000), o teste de Wald é preferível aos testes de verossimilhança clássicos quando a sobredispersão ou subdispersão está presente. Vale ressaltar que apesar dele ser preferível na presença de sobredispersão ou subdispersão pode subestimar a contribuição de um preditor na explicação da ocorrência do evento de interesse.

De acordo com Lara (2013), as hipóteses a serem testadas através do teste de Wald são descritas a seguir: $\begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{cases}$ para qualquer j . A estatística teste de Wald para a regressão logística é mostrada na **Equação 21**.

$$W_j = \frac{\hat{\beta}_j}{\widehat{DP}(\hat{\beta}_j)} \quad (Eq. 21)$$

Onde: $\hat{\beta}_j$ é a estimativa de máxima verossimilhança do parâmetro.

$\widehat{DP}(\hat{\beta}_j)$ é a estimativa do desvio padrão do parâmetro.

A estatística Wald tem distribuição Qui-Quadrado, sendo calculada pelo quadrado da razão entre o coeficiente e o seu erro padrão. Antes de realizar esse cálculo é necessário realizar uma correção dos erros padrão, $\widehat{DP}(\hat{\beta}_j)$, quando os dados apresentam uma sobredispersão ou subdispersão. Esta correção consiste em multiplicar os erros padrão por um fator comum que pode ser $\sqrt{\text{deviance}/(n-p)}$ ou $\sqrt{\chi^2/n-p}$. Como estes dois fatores dão resultados semelhantes, existe um nível de conforto para o analista escolher o que desejar (MYERS *et al.*, 2002).

Caso não se rejeite a hipótese nula (H_0) temos que a variável explicativa não afeta a variável resposta, podendo-se retirá-la do modelo (LARA, 2013). Alguns *softwares* fornecem o p-valor, ao invés da estatística teste referente à distribuição para a decisão de incluir ou não uma variável no modelo, para isso é necessário estipular o nível de significância, que é um valor arbitrário, definido segundo critérios do pesquisador. Convencionalmente, utiliza-se um nível de significância de 5%, que será utilizado nesta dissertação. Todo teste estatístico clássico fornece a probabilidade do efeito observado ser proveniente do acaso, tal probabilidade é denominada de p-valor e caso este p-valor seja menor do que o nível de significância adotado isso indica que os resultados obtidos são estatisticamente significantes (OLIVEIRA, 2006).

Outra forma de realizar inferência, utilizando-se a estatística de Wald é através dos Intervalos de Confiança (IC) como mostra a **Equação 22** (MYERS *et al.*, 2002; FOX, 2008).

$$\hat{\beta}_j \pm z_{1-\alpha/2} \widehat{DP}(\hat{\beta}_j) \quad (Eq. 22)$$

Onde: *z é calculado a partir da Tabela de distribuição Normal Padrão;

* $\hat{\beta}_j$ é o parâmetro estimado;

* $\widehat{DP}(\hat{\beta}_j)$ é o desvio padrão (ou erro padrão) da estimativa do parâmetro.

Quando o IC do parâmetro contém o valor zero, pode-se concluir que o parâmetro testado não é significativo para o modelo.

2.3.3.2.

Estimação do Parâmetro de Dispersão

Atkinson e Riani (2000), afirmar que o estimador de σ^2 nos MLG é dado pela **Equação 23**.

$$\hat{\sigma}^2 = \hat{\Phi} = \frac{D(y_i, \mu_i)}{n - p} \quad (Eq. 23)$$

McCullagh e Nelder (1989), afirmam que, a estatística χ^2 é preferida em algumas situações devido a sua interpretação mais direta. Eles recomendam também que o estimador seja obtido pelo Método dos Momentos, como mostra a **Equação 24**.

$$\tilde{\Phi} = \frac{1}{n - p} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} = \frac{\chi^2}{n - p} \quad (Eq. 24)$$

Onde: * χ^2 é a estatística de Pearson;

*n é o número de observações;

*p é o número de parâmetros envolvidos no modelo.

2.3.3.3.

Função de Ligação

A adequação da função de ligação pode ser verificada pelo gráfico dos resíduos *studentizados versus* valores ajustados (valores previstos). Caso este gráfico não apresente nenhum padrão óbvio (comportamento não aleatório) e a linha resultante do amortecimento (*lowess*) seja aproximadamente horizontal e próxima à linha reta horizontal de ordenada zero, há indícios de que a função de ligação utilizada é adequada (VIEIRA, 2004).

Através do gráfico descrito acima é possível identificar pontos isolados com grandes valores de resíduos, o que indica uma função de ligação inadequada e que pode resultar em uma modelagem imprópria dos dados. Já a tendência à propagação do aumento de valores ajustados aponta uma função de variância insatisfatória (BIASOLI, 2005).

2.3.3.4.

Função de Variância

A adequação da função de variância pode ser verificada através do gráfico do valor absoluto dos resíduos *studentizados versus* os valores ajustados (valores previstos) (VIEIRA, 2004).

Ainda de acordo com o autor acima, a função de variância geralmente é definida como uma função de potência da média (μ^λ). Assim, quando a linha do amortecimento cresce sistematicamente, da esquerda para direita, com o aumento da média, há indícios de que se deve usar um valor maior para λ do que o valor correspondente à distribuição que foi usada no modelo. Caso contrário, isto é, quando há um decréscimo sistemático, indica a adequação de um valor menor para λ .

Se houver pontos mais distantes entre si, a confiança na forma da linha de amortecimento (*lowess*) é menor e, portanto, não se deve considerar esta indicação como uma função de variância incorreta.

2.3.3.5.

Distância de Cook

A distância de Cook para os MLG é fornecida por Atkinson e Riani (2000) como mostra a **Equação 25**.

$$D_i = \frac{r_{pi}^2 h_{ii}}{p\hat{\phi}(1 - h_{ii})^2} \quad (\text{Eq. 25})$$

Onde: *p é o número de parâmetros;

* $\hat{\phi}$ é a estimativa do parâmetro de dispersão;

* r_{pi}^2 é o resíduo de Pearson *studentizado*;

* h_{ii} é o i-ésimo elemento da diagonal da matriz H.

Caso o gráfico da distância de Cook não apresente valores em que $D_i > 1$, não há indícios de observações influentes (ou atípicas) no modelo (COOK, WEISBERG, 1999).

3.

Descrição do Experimento

Neste capítulo serão descritos todos os *softwares* utilizados e, também, o procedimento adotado para beneficiar a malha de óleo utilizada na realização do experimento fatorial.

Definiremos também um roteiro para investigar a escolha do modelo mais adequado para representar o comportamento da média da característica de qualidade (resposta) do produto em estudo. Um modelo adequado para a média é importante para o ajuste dos fatores que a afetam, conduzindo a um valor alvo (ou desejado). Além disso, um modelo inadequado para a média pode levar à identificação de um modelo inadequado para a variância, isto é, levando a identificação de efeitos espúrios na dispersão.

Por último, será mostrada a metodologia adotada neste estudo para a modelagem dos dados que seguem uma distribuição Binomial, o que viola os pressupostos do modelo linear clássico de normalidade, variância constante e aditividade. Portanto, a aplicação do MLG é uma alternativa viável.

3.1.

Softwares Utilizados

Os *softwares* utilizados neste estudo serão descritos a seguir:

i) *Microsoft Office Excel*: *software* comercial desenvolvido pela Microsoft, que será utilizado para organização dos dados e para a realização do teste de Wald desse estudo. A versão utilizada é 2007.

ii) *Design-Expert*: *software* comercial desenvolvido e distribuído pela empresa Stat-Ease, o qual ajusta um modelo de regressão e oferece gráficos tridimensionais para a visualização da superfície de resposta e também gráficos de contorno interativos, além de um módulo de otimização da resposta capaz de buscar o ótimo diante de inúmeras respostas. A versão utilizada é 8.0.7.1. Este *software* foi utilizado para aleatorizar o experimento.

iii) *Arc: software* gratuito que acompanha o livro de Cook e Weisberg (1999), podendo também ser obtido no *site* do Departamento de Estatística da Universidade de Minnesota, EUA. É excelente para a construção de gráficos para diagnóstico dos modelos lineares clássicos e generalizados. A versão utilizada é 3.52.17.

3.2.

Caracterização do Produto

A empresa compra e beneficia dois tipos de matéria-prima. A primeira é oriunda dos resíduos de processos têxteis, que são contaminadas por óleo, quando confeccionadas nas máquinas de malharia circular. Isto é, a contaminação ocorre no momento em que as agulhas são lubrificadas e em alguns casos esse sistema de lubrificação apresenta defeito e os tecidos acabam sendo impregnados por óleos e graxas. Esses tecidos serão submetidos a um processo de beneficiamento têxtil, o alvejamento. O segundo tipo de matéria prima é o tecido cru sem nenhum defeito, que possui um custo mais elevado cerca de 50% e que também receberá um beneficiamento, ou seja, receberá um cozinhamento e alvejamento simultâneo.

O produto da empresa que será utilizado neste estudo será a malha de óleo proveniente dos resíduos de processos têxteis (malharia circular), que é a mais crítica. Faz-se necessário que este produto apresente boa hidrofiliidade, não aparente aspecto de engomado e nem manchas de óleo.

Foram coletados 40Kg desse material no setor de separação de matérias primas, que foram submetidos ao processo de alvejamento com hipoclorito de sódio. Após o processo de realizou-se uma separação do produto em quatro grupos que serão descritos e detalhados a seguir:

i) *Grupo 90*: o produto apresenta um maior teor de brancura e alta capacidade de absorção, resultando em uma melhor qualidade. Conseqüentemente, este fator impacta diretamente no preço do artigo, que é maior que os demais.

ii) *Grupo 89*: o produto apresenta um teor de brancura, capacidade de absorção e preço um pouco inferior ao grupo 90 e superior aos demais.

iii) *Grupo 95*: tem-se um teor de brancura pior, pois o tecido fica amarelado e com capacidade de absorção e preço inferior ao 89 e 90, mas superior ao 92.

iv) Grupo 92: o produto apresenta o pior teor de brancura, capacidade de absorção e preço inferior aos demais.

Com o intuito de facilitar o experimento fatorial agrupou-se o 90 e 89 em tecidos sem defeitos e 95 e 92 em tecidos com defeitos, isto é, com alguma imperfeição que afete a qualidade final do produto, o que provavelmente diminui a margem da empresa, podendo gerar prejuízo. Este agrupamento foi realizado desta forma, pois os produtos possuem características similares de grau de brancura, capacidade de absorção (hidrofilidade) e de preço.

Atualmente, a empresa utiliza os seguintes insumos químicos: hipoclorito de sódio (NaClO) e hidróxido de sódio (NaOH). Esses insumos afetam diretamente o processo e podem ser controlados. Portanto, testou-se a influência deles na variável resposta: proporção de itens defeitos em cada experimento.

Os ensaios foram realizados em escala piloto na Planta Piloto de Inovação (PPI), onde se encontram os laboratórios físicos e químicos e os equipamentos, do SENAI CETIQT, em virtude de uma limitação técnica em realizar os experimentos em pequenas quantidades, nos equipamentos da empresa.

3.3.

Procedimento de alveamento

O procedimento de alveamento do tecido visa retirar as gorduras, os óleos e as graxas presentes no mesmo, tornando-o branco. Isto é, apto para receber as etapas subsequentes, como um possível tingimento, o que pode ocorrer em alguns casos quando solicitado pelo cliente.

O procedimento adotado para alvejar a malha de óleo foi o esgotamento, onde o hipoclorito de sódio e o hidróxido de sódio são deslocados do banho para a fibra mediante a movimentação deles (TWARDOKUS, 2004). Este deslocamento depende da agitação mecânica e dos produtos utilizados.

É usual no segmento têxtil utilizar uma Relação de Banho (RB) de 1:10, isto é, existe uma relação entre a massa do material têxtil utilizada e o volume de banho a ser processado, para cada 1kg de tecido, necessita-se de 10L de banho (água). Essa relação também foi obedecida no experimento em escala piloto.

Estabelecida a RB, calcula-se o volume de banho, a partir da massa do material têxtil (3kg). Portanto, seu volume de banho é de 30L. Vale ressaltar que pesou-se em torno de 3kg na balança Balmak MP10, Classe III (Linha Brasil), com precisão de 2g, à medida que os experimentos foram sendo executados.

Ambos os insumos químicos foram fornecidos pela empresa do segmento têxtil. O hidróxido de sódio foi pesado na balança Denver Instrument APX-200, que possui uma melhor precisão, de 0,0001g, e por ser fornecido em escama foi necessário diluí-lo em água destilada. Já o hipoclorito de sódio é fornecido na forma líquida, sendo medido em um béquer de 2.000mL.

Como na empresa o processo de alvejamento é realizado em temperatura ambiente, os ensaios na PPI também foram realizados em temperatura ambiente.

3.3.1.

Empresa

A empresa adota as seguintes etapas para realizar o processo de alvejamento do tecido:

1. *Colocar o tecido na máquina e enxaguá-lo por 10 minutos, pois retira o excesso de possíveis sujeiras.*
2. *Descarregar o banho e encher a máquina novamente (2 minutos).*
3. *Adicionar hidróxido de sódio (soda cáustica) e deixá-la por 20 minutos no banho.*
4. *Descarregar o banho e encher a máquina novamente (2 minutos).*
5. *Adicionar hipoclorito de sódio (cloro) de 30 a 40 minutos.*
6. *Descarregar o banho e encher a máquina novamente (2 minutos).*
7. *Fazer de 5 a 6 enxágues (10min cada enxágue com intervalo de 2min entre cada enxágue).*
8. *No último enxágue adiciona-se a naftalina, deixando-a por 20 minutos no banho. Vale ressaltar que a naftalina é utilizada com o intuito de minimizar o odor forte presente no tecido devido aos óleos, graxas e o cloro.*
9. *Descarregar o banho e encher a máquina novamente (2 minutos).*
10. *Retirar o tecido da máquina de lavar industrial. A **Figura 5** mostra a máquina de lavar industrial da empresa.*



Figura 5 – Máquina de lavar industrial da empresa

11. *Tempo total de processo, aproximadamente 2 horas e 28 minutos e foi realizado em temperatura ambiente em torno de 30°C.*

3.3.2.

Escala Piloto

Os experimentos realizados em escala piloto obedeceram as seguintes etapas:

1. *Colocar o tecido na máquina e enxaguá-lo por 10 minutos, pois retira o excesso de possíveis sujeiras.*
2. *Descarregar o banho e encher a máquina novamente (2 minutos).*
3. *Adicionar hidróxido de sódio (soda cáustica) e deixá-la por 20 minutos no banho.*
4. *Descarregar o banho e encher a máquina novamente (2 minutos).*
5. *Adicionar hipoclorito de sódio (cloro) de 30 a 40 minutos.*
6. *Descarregar o banho e encher a máquina novamente (2 minutos).*
7. *Fazer de 5 a 6 enxágues (10min cada enxágue com intervalo de 2min entre cada enxágue).*
8. *Descarregar o banho no último enxágue.*
9. *Retirar o tecido da máquina de lavar industrial. A **Figura 6** mostra a máquina de lavar industrial da PPI.*



Figura 6 – Máquina de lavar industrial da PPI

10. *Tempo total de processo, aproximadamente 2 horas e 6 minutos e foi realizado em temperatura ambiente em torno de 30°C. Vale ressaltar que como a naftalina não interfere na variável resposta, houve uma redução no tempo de processamento em escala piloto.*

3.4.

Definição dos Parâmetros de Processo

O nível central (0) empregado no Planejamento de Experimentos será definido a partir dos resultados coletados na prática (dia a dia) dentro da empresa. Os demais níveis superior (+) e inferior (-) serão extrapolados em torno de 33,33% do nível central e mostrados na **Tabela 3**.

Tabela 3 - Variáveis independentes do planejamento de experimentos e seus níveis

VARIÁVEIS INDEPENDENTES	NÍVEIS		
	-1	0	1
A – Concentração Hidróxido de sódio (g/L)	16	24	32
B – Volume de Hipoclorito de sódio (mL)	600	900	1.200

Desta forma, foi realizado o planejamento fatorial 2^K com ponto central, onde K é o número de fatores e 2 o número de níveis, resultando numa condição de $2^2 = 4$ desenvolvido em duplicata e 1 ponto central que foi feito em quadruplicata, totalizando 12 experimentos.

A **Tabela 4** mostra a matriz do planejamento fatorial e a lista dos ensaios na chamada ordem padrão e na ordem como foi executado na prática, isto é, na ordem aleatória.

Tabela 4- Matriz fatorial 2^2 na forma padrão e na ordem aleatória

EXPERIMENTOS	ORDEM PADRÃO	ORDEM ALEATÓRIA	FATORES	
			A Concentração Hidróxido de sódio (g/L)	B Volume de Hipoclorito de sódio (mL)
1a*	1	10	(-1) 16	(-1) 600
1b	2	1	(-1) 16	(-1) 600
2a	3	4	(+1) 32	(-1) 600
2b	4	9	(+1) 32	(-1) 600
3a	5	2	(-1) 16	(+1) 1200
3b	6	12	(-1) 16	(+1) 1200
4a	7	5	(+1) 32	(+1) 1200
4b	8	11	(+1) 32	(+1) 1200
5a	9	3	(0) 24	(0) 900
5b	10	7	(0) 24	(0) 900
5c	11	8	(0) 24	(0) 900
5d	12	6	(0) 24	(0) 900

a, b, c e d = representam o número de réplicas.

3.5.

Variável Resposta

No início do processo de modelagem faz-se necessário selecionar uma ou mais variáveis respostas do processo que serão utilizadas para análise dos experimentos, essas variáveis podem ser contínuas ou discretas. Na prática toda variável aleatória é discreta, mesmo quando o espaço amostral da variável é contínuo, pois as medições da variável resposta são limitadas de acordo com a precisão do instrumento de medida. Entretanto, quando o espaço amostral é contínuo e o número de medidas possíveis é elevado em relação ao tamanho da amostra, a aproximação pelas distribuições de variáveis contínuas é satisfatória (VIEIRA, 2004).

De acordo com Lovatto *et al.* (2007), em algumas situações pode-se discretizar uma variável contínua transformando-a, de acordo com o objetivo, em discreta. Essa transformação permite considerar estatisticamente, as modalidades dos fatores experimentais, as características principais, o tipo de medida, a equipe de pesquisa, etc.

As variáveis contínuas são, normalmente, utilizadas para representar uma determinada característica de qualidade definida no intervalo dos números reais, como exemplos, a resistência à ruptura do tecido. Nestes casos o espaço amostral da variável resposta é contínuo e pode-se considerar as seguintes distribuições: Normal, Gama e Normal Inversa. Já as variáveis discretas são aquelas características de qualidade que podem assumir valores pertencentes a um conjunto finito ou enumerável, sendo geralmente números inteiros, como exemplo, número de peças defeituosas e boas. Nestes casos, geralmente classificam-se os itens inspecionados como conforme ou não as especificações. Este tipo de característica denomina-se atributo.

O número de itens com defeitos em uma amostra aleatória segue uma distribuição Binomial e para isso é necessário estabelecer uma relação (regressão Logit) entre os fatores de produção (concentração de hidróxido de sódio e volume de hipoclorito de sódio) e o número de itens com defeitos em cada experimento (variável resposta), de modo a identificar a combinação de valores dos fatores de produção que minimize a resposta.

3.6.

Adequação do MLG

O processo de ajuste dos MLG foi dividido em três etapas:

i) *Formulação do modelo*: escolha da distribuição de probabilidade para a variável resposta, variáveis explicativas e função de ligação;

ii) *Ajuste do modelo*: estimação dos parâmetros lineares do modelo e de determinadas funções das estimativas desses parâmetros, que representam medidas de adequação dos valores estimados.

iii) *Adequação do modelo*: teste de adequação do modelo como um todo, analisando-se estatisticamente os resíduos, utilizando-os para verificar a adequação do modelo.

3.6.1.

Formulação do Modelo

A formulação do modelo compreende a escolha da distribuição de probabilidade para a variável resposta, explicativas e para a função de ligação, para isso deve-se examinar cuidadosamente os dados.

Em certas situações, deve-se levar em consideração que os dados podem seguir uma distribuição discreta ou contínua, podem ser distribuições assimétricas, binomiais, dados restritos a um intervalo do conjunto dos reais entre outros. Sendo assim o MLG permite a modelagem de variáveis discretas, tal como o número de produtos com defeitos em uma amostra (distribuição Binomial), como é o caso desta dissertação, em que a variância não é constante e sim função da média.

Vale ressaltar também que os dados podem apresentar uma sobredispersão ou uma subdispersão e quando isso ocorre usa-se um parâmetro multiplicador na função da variância, o parâmetro de dispersão (ϕ).

A função de ligação deve ser compatível com a distribuição escolhida anteriormente. Portanto, devem ser funções de $[0,1]$ no conjunto de número reais, como o modelo Logit, Probit, Complemento log-log, etc. Nesta dissertação optou-se pela utilização da função de ligação Logit, já que diversos autores como Demétrio (2002), Myer *et al.* (2002), Venticinque *et al.* (2007), Cordeiro e Demétrio (2013), recomendam utilizá-la em análises de dados na forma de proporção.

Por último, a função de ligação Logit irá estabelecer uma relação funcional entre os fatores de produção e o número de itens com defeitos por experimento, de forma a encontrar os níveis que minimizem a variável resposta.

3.6.2.

Ajuste do Modelo

Pode-se utilizar diversos *softwares* estatísticos para estimar os parâmetros lineares dos modelos como: R, SAS, S-PLUS, MATLAB e ARC. Nesta dissertação optou-se pela utilização do *software* ARC, que irá fornecer as estimativas dos coeficientes, o erro padrão da estimativa, o quociente entre a estimativa e o erro padrão, o p-valor do teste e a estatística de Pearson.

Através da estatística de Pearson dividida por $(n - p)$ é possível estimar o parâmetro de dispersão (ϕ), indicando se os dados fornecem indícios de uma sobredispersão ou subdispersão.

Quando isso ocorre é necessário realizar uma correção dos erros padrão encontrados multiplicando-os por $\sqrt{\frac{\chi^2}{n-p}}$. Feito isso, deve-se estimar novamente os coeficientes, com o auxílio do *software* Excel e, realizar uma análise da significância estatística destes coeficientes através do teste de Wald. Caso o p-valor do teste de Wald seja inferior ao nível de significância estipulado de 5%, rejeita-se a hipótese nula e o fator é significativo para o modelo.

3.6.3.

Adequação do Modelo

A análise da adequação do modelo proposto é realizada através do gráfico dos resíduos ($e_i = y - \bar{y}$) que verifica se as suposições feitas a cerca do modelo estão adequadas ou não, ou seja, se existe discrepâncias significativas que podem implicar na escolha de outro modelo, ou em aceitar a existência de observações influentes.

Portanto, a inspeção gráfica é um meio poderoso de verificação do MLG e, portanto, será utilizada nesta dissertação. A adequação do MLG e a existência de observações atípicas podem ser melhores observadas através do gráfico de probabilidade Normal dos resíduos com envelope. Este tipo de gráfico se destaca em dois aspectos: na identificação da distribuição originária dos dados e na

identificação de valores que se destacam no conjunto de observações. Os envelopes, no caso dos MLG com distribuições diferentes da Normal, são construídos com os resíduos sendo gerados a partir do modelo ajustado como relatado na literatura (OLIVEIRA, 2013 *apud* PAULA, 2010).

No caso de um ajuste adequado, os resíduos devem apresentar comportamento aleatório e seus valores devem ser próximos do valor zero. Em outras palavras, ao se observar o gráfico de probabilidade Normal dos resíduos com envelope não se pode identificar um padrão de comportamento que não o aleatório e nem pontos muito fora do envelope e de alinhamento.

Quando o gráfico dos resíduos *studentizados versus* os valores ajustados, não apresentam nenhum padrão, é um indicativo de que a relação funcional (função de ligação) proposta para os dados é satisfatória.

O gráfico do valor absoluto dos resíduos *studentizados versus* os valores ajustados é utilizado para mostrar se a função de variância está correta ou não. Já o gráfico da distância de Cook *versus* as observações oferece indícios da existência ou não de observações influentes no modelo.

4.

Análise de Resultados

Nesse capítulo serão apresentados e discutidos os resultados dos experimentos realizados, bem como o(s) modelo(s) resultante(s) e a descrição do roteiro da aplicação do MLG.

4.1.

Apresentação dos Dados

A **Tabela 5** mostra a matriz do planejamento de experimentos, tendo como variável resposta a proporção de itens com defeitos. A matriz do planejamento lista os ensaios na ordem padrão. Os três primeiros fatores começam com o nível inferior (-) e depois os sinais vão se alternando.

Tabela 5 - Matriz padrão de planejamento fatorial 2^2 com ponto central, tendo como variável resposta a proporção de itens com defeitos

EXPERIMENTOS	FATORES		RESPOSTA
	A	B	Y
	Concentração Hidróxido de sódio (g/L)	Volume de Hipoclorito de sódio (mL)	Proporção de Itens com Defeitos
1a*	-1	-1	0,1559
1b	-1	-1	0,1576
2a	+1	-1	0,3226
2b	+1	-1	0,1284
3a	-1	+1	0,3016
3b	-1	+1	0,3198
4a	+1	+1	0,2915
4b	+1	+1	0,3884
5a	0	0	0,1333
5b	0	0	0,3217
5c	0	0	0,2972
5d	0	0	0,1969

* a, b, c e d = representam o número de réplicas

4.2.

Variável Resposta

Como mostra a **Tabela 5**, a proporção de itens com defeitos em cada experimento (variável resposta) segue uma distribuição Binomial, sendo necessário encontrar uma relação entre ela e os fatores de produção: concentração de hidróxido de sódio (A), volume de hipoclorito de sódio (B) e sua interação (AB). A função de ligação Logit estabelece essa relação de forma a encontrar os níveis que minimizem a variável resposta (Y).

4.3.

Aplicação do MLG

A geração das tabelas e dos gráficos a seguir é feita no *software* ARC e no *software* Excel.

A **Tabela 6** apresenta as estimativas dos coeficientes, o erro-padrão da estimativa e o quociente entre a estimativa e o erro-padrão de todos os fatores da variável resposta proporção de itens com defeitos.

Tabela 6 – Estimativa dos coeficientes e erro padrão de todos os fatores da variável resposta proporção de itens com defeitos

COEFICIENTE	ESTIMATIVA	ERRO PADRÃO	ESTIMATIVA/ERRO PADRÃO
Constante	-1,1192	0,6802	-1,6453
A	0,1478	0,8391	0,1761
B	0,3687	0,8352	0,4414
AB	-0,0806	0,8396	-0,0960

O parâmetro de dispersão é calculado através da **Equação 26**.

$$\tilde{\phi} = \frac{\chi^2}{n - p} = \frac{0,258}{8} \cong 0,0323 \quad (\text{Eq. 26})$$

Como se pode observar o parâmetro de dispersão é menor que 1. Logo, a variância de Y_i apresenta algum grau de subdispersão (*underdispersion*).

Um teste formal para avaliar quais fatores são estatisticamente significativos é o teste de Wald. Porém é necessário realizar primeiramente uma correção, multiplicando-se os erros padrão da **Tabela 6** por $\sqrt{\frac{\chi^2}{n-p}}$, que é aproximadamente 0,1796.

O segundo passo é estipular o nível de significância do teste, que foi de 5%. Caso o p-valor seja menor do que o nível de significância adotado isso indica que os resultados obtidos são estatisticamente significativos.

A **Tabela 7** mostra o teste de Wald para todos os fatores da variável resposta proporção de itens com defeitos.

Tabela 7 – Teste de Wald para todos os fatores

COEFICIENTE	ERRO PADRÃO CORRIGIDO	ESTIMATIVA / ERRO PADRÃO CORRIGIDO	TESTE DE WALD	P-VALOR
Constante	0,1222	-9,1620	83,9430	0,0000
A	0,1507	0,9807	0,9618	0,3267
B	0,1500	2,4581	6,0421	0,0140
AB	0,1508	-0,5348	0,2860	0,5928

Como mostra a **Tabela 7** apenas o volume de hipoclorito de sódio (B), destacado em vermelho, é significativo. Portanto, através do teste de Wald pode-se retirar os fatores que não são estatisticamente significativos ao modelo, isto é, incluir somente as variáveis mais importantes para a explicação do fenômeno em análise. Esse procedimento baseia-se exclusivamente em critérios estatísticos (OLIVEIRA, 2006).

Os resultados acima podem ainda ser confirmados através do IC mostrado na **Tabela 8**. Caso o IC do fator contenha o valor zero, pode-se concluir que ele não é significativo para o modelo.

Tabela 8 – IC do teste de Wald para todos os fatores

IC	INFERIOR	SUPERIOR
Constante	-1,3586	-0,8798
A	-0,1476	0,4431
B	0,0747	0,6627
AB	-0,3761	0,2149

A concentração de hidróxido de sódio (A) e a interação entre os fatores (AB) contêm o valor zero. Portanto, esses fatores não são significativos para o modelo e podem ser retirados do mesmo.

A **Tabela 9** mostra a estimativa dos coeficientes e erro padrão somente para o fator significativo da proporção de itens com defeitos.

Tabela 9 – Estimativa dos coeficientes e erro padrão somente do fator significativo

COEFICIENTE	ESTIMATIVA	ERRO PADRÃO	ESTIMATIVA/ERRO PADRÃO
Constante	-1,1136	0,6770	-1,6448
B	0,3615	0,8303	0,4354

O parâmetro de dispersão é novamente calculado através da **Equação 27**.

$$\tilde{\phi} = \frac{\chi^2}{n - p} = \frac{0,308}{10} \cong 0,0308 \quad (\text{Eq. 27})$$

Como se pode observar através da **Equação 27**, o parâmetro de dispersão é menor que 1. Portanto, existe uma subdispersão (*underdispersion*) dos dados.

A **Figura 7** apresenta o gráfico de probabilidade Normal com envelope para o fator significativo da variável resposta proporção de itens com defeitos.

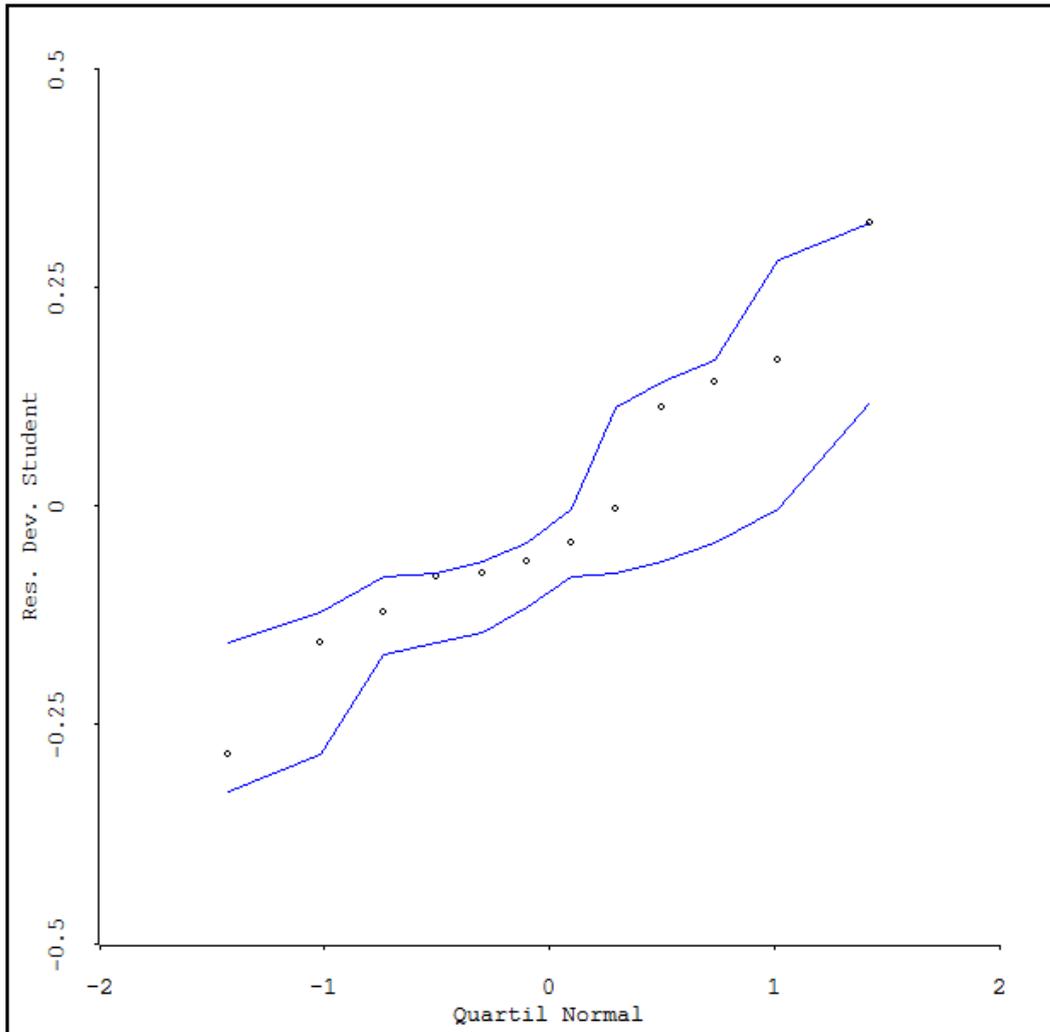


Figura 7 – Gráfico de probabilidade Normal com envelope para o fator significativo

No gráfico da **Figura 7** não se observa pontos muito fora do envelope e também não se observa pontos muito fora do alinhamento. Logo, não existe indicação de observações atípicas e nem de que o modelo proposto esteja inadequado.

A função de ligação pode ser verificada com o gráfico dos resíduos *studentizados versus* o valor ajustado para o fator significativo da variável resposta proporção de itens com defeitos, como mostra a **Figura 8**.

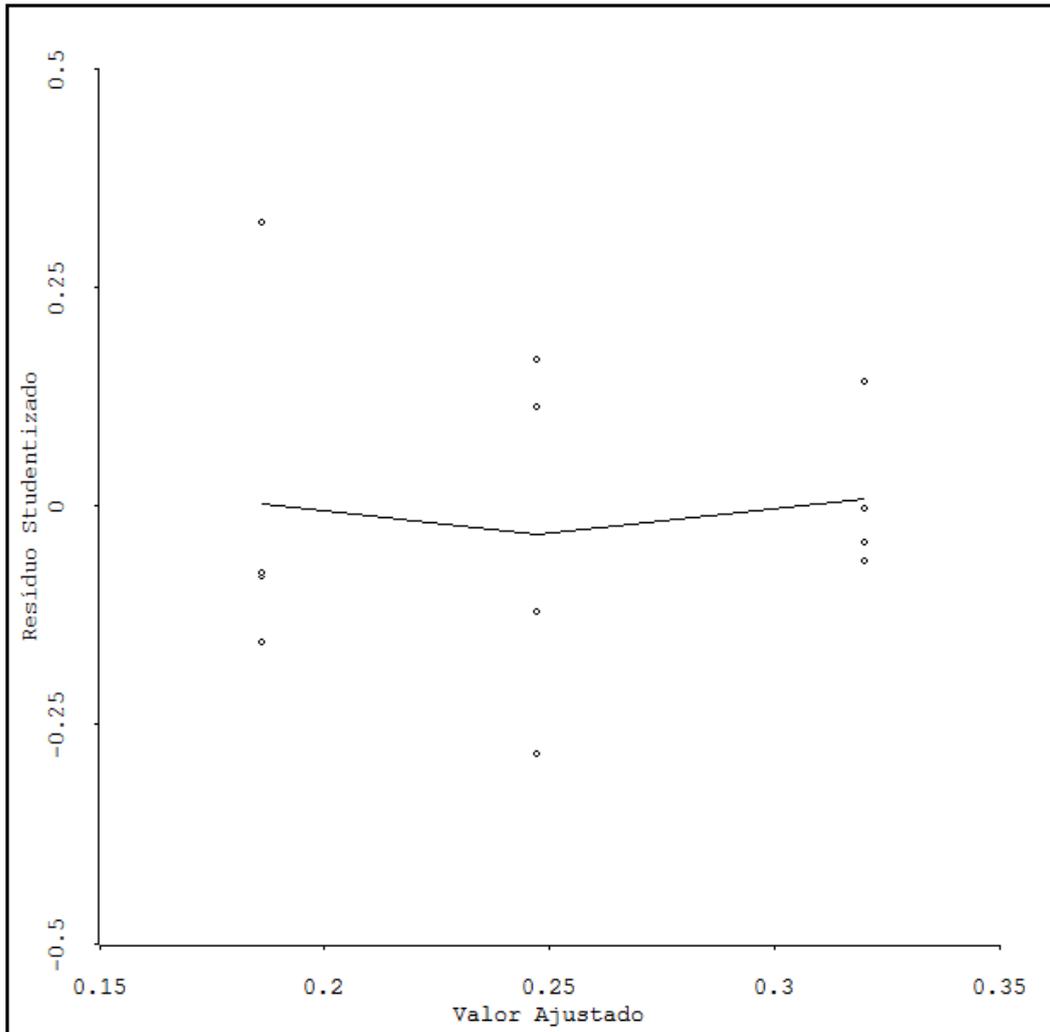


Figura 8 – Gráfico dos resíduos *studentizados* versus o valor ajustado para o fator significativo

Pode-se observar que os resíduos da **Figura 8** não apresentam nenhuma tendência e a linha resultante do amortecimento (*lowess*) fica situada nas proximidades da reta horizontal de ordenada zero, portanto, existe indicação de que a função de ligação utilizada está correta.

A verificação da adequação da função de variância para o fator significativo da variável resposta proporção de itens com defeitos é mostrada na **Figura 9**.

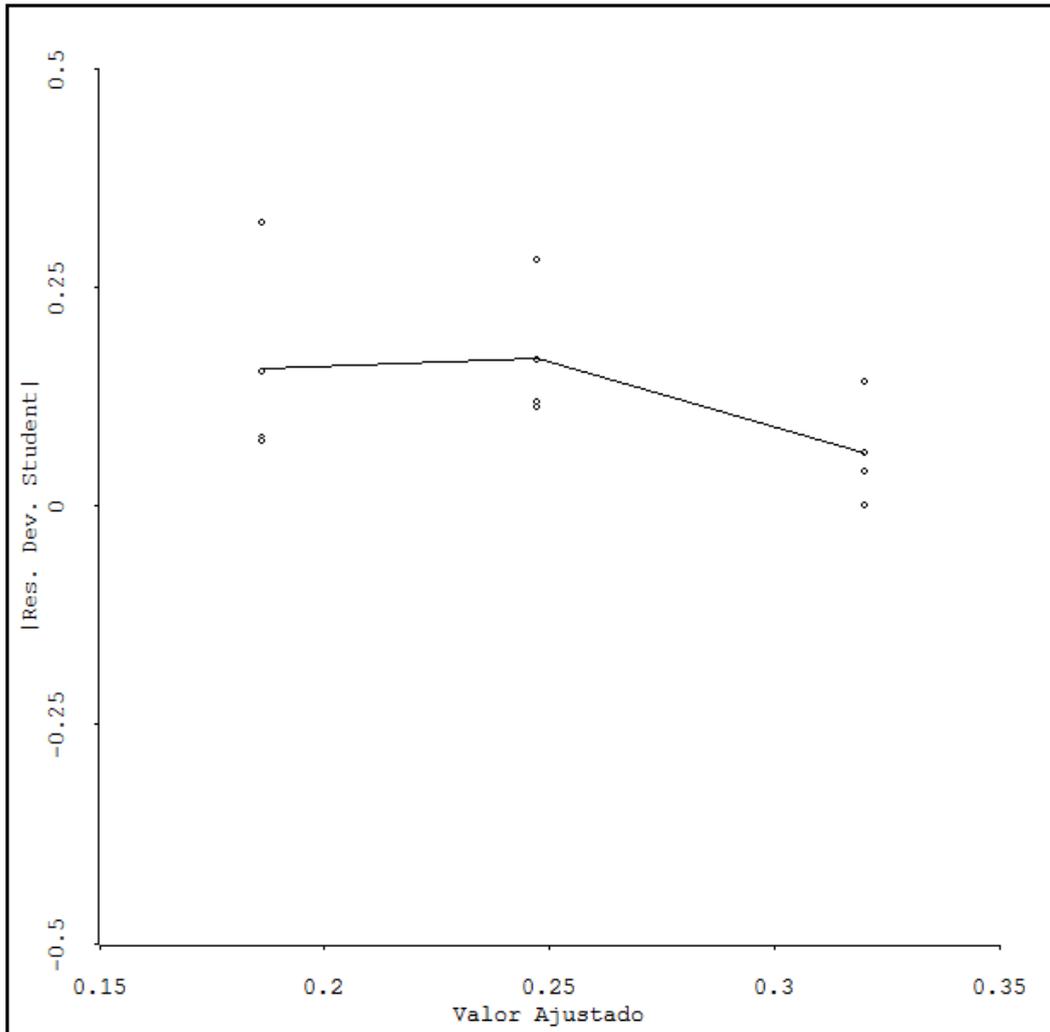


Figura 9 – Gráfico do valor absoluto do resíduo *studentizado* versus o valor ajustado para o fator significativo

Como a variância não aumenta com o aumento da média, existe indícios que a função da variância está correta.

Por último, a **Figura 10** apresenta o gráfico da distância de Cook utilizada para verificar a existência de observações atípicas para o fator significativo da variável resposta proporção de itens com defeitos.

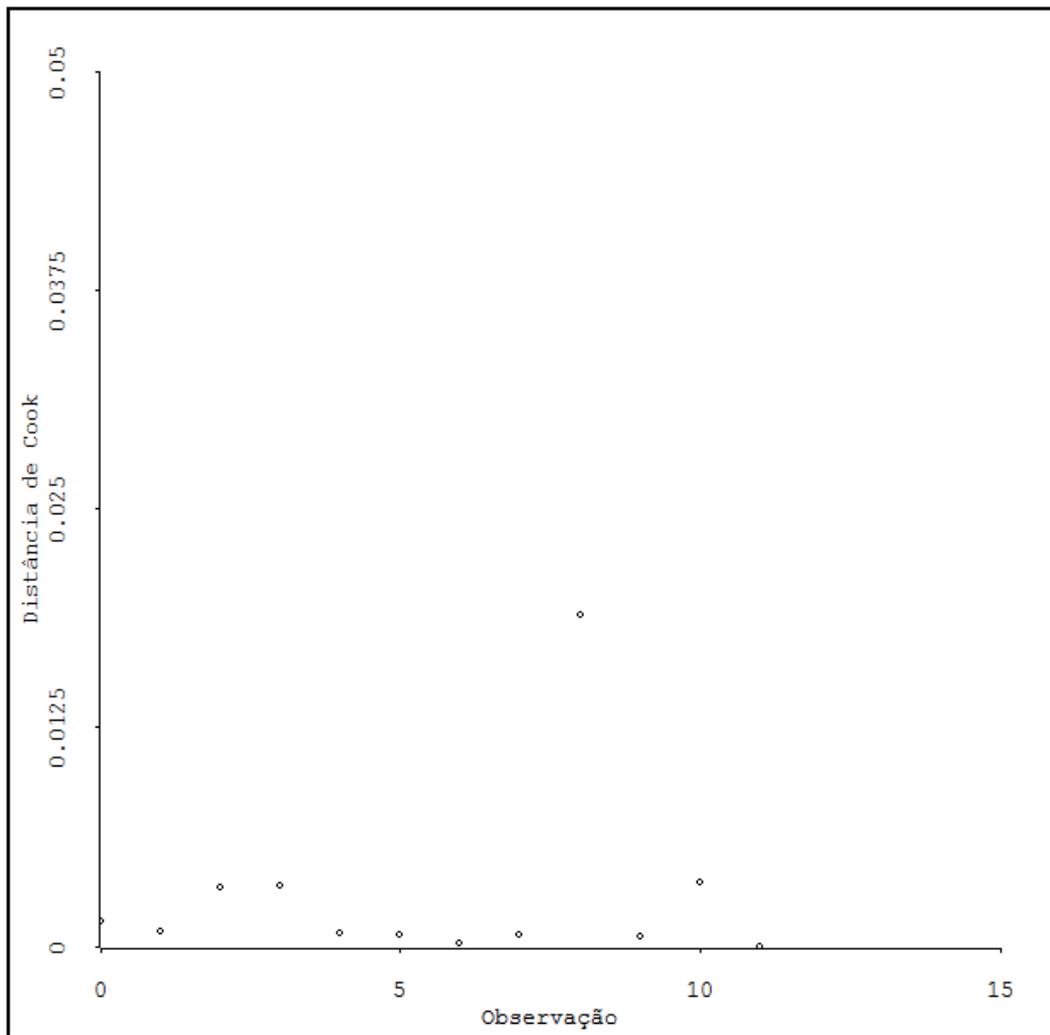


Figura 10 – Gráfico da Distância de Cook para o fator significativo

Como não existe valor acima de 0,5, não há indícios de observações atípicas no modelo proposto.

Aplicando-se novamente o teste de Wald e multiplicando-se os erros padrão da Tabela 9 por $\sqrt{\frac{\chi^2}{n-p}}$, que é aproximadamente 0,1756, obtém-se o teste de Wald para o fator significativo do modelo como mostra a Tabela 10.

Tabela 10 – Teste de Wald para o fator significativo

COEFICIENTE	ERRO PADRÃO CORRIGIDO	ESTIMATIVA / ERRO PADRÃO CORRIGIDO	TESTE DE WALD	P-VALOR
Constante	0,1188	-9,3723	87,8394	0,0000
B	0,1457	2,4809	6,1547	0,0131

Observou-se através da **Tabela 10** que o volume de hipoclorito de sódio (B) é o único fator que exerce influência sobre a variável resposta proporção de itens com defeitos. Estes resultados podem ser confirmados através do IC como mostra a **Tabela 11**.

Tabela 11 – IC do teste de Wald para o fator significativo

IC	INFERIOR	SUPERIOR
Constante	-1,3465	-0,8807
B	0,0759	0,6471

O volume de hipoclorito de sódio (B) não contém o valor zero. Portanto, esse fator é estatisticamente significativo para o modelo. Isto se deve ao fato dele ser o responsável por remover os resíduos (ceras, graxas e substâncias solúveis) que podem influenciar no aspecto final do material têxtil utilizado, a malha de óleo, sendo muito utilizado quando se deseja que o produto acabado torne-se branco ou possua cores claras (MOSER, 2011).

Segundo Sánchez (1966), o processo de alvejamento se inicia quando o hipoclorito entra em contato com a água, liberando ácido hipocloroso e formando hidróxido de sódio, esta reação é reversível e é descrita a seguir:



De acordo com as equações relatadas acima, para favorecer o deslocamento da equação para a esquerda a partir da decomposição do ácido hipocloroso e subsequente neutralização é necessário a adição do hidróxido, de maneira a interromper a hidrólise e evitar que o hipoclorito perca seu poder de oxidação.

Portanto, o hidróxido de sódio (A) atua apenas como agente estabilizante do hipoclorito, isto é, como tampão atenuando a variação do pH do banho, mantendo-o aproximadamente constante. Ele também auxilia na saponificação das ceras e graxas (SÁNCHEZ, 1966; LUCCA, 2006).

Geralmente, o hipoclorito é o alvejante mais utilizado na indústria têxtil, por ser mais barato, de pronta disponibilidade, rápido e eficiente, porém é agressivo já que diminui a resistência mecânica da fibra. Além disso, ele também pode formar cloroaminas que degradam a celulose e são altamente tóxicas e poluentes, o que torna o tratamento de seu efluente posteriormente difícil (STAACK, 2007; KOHAN; ARAUJO, 2008).

Portanto, alguns autores como Robinson *et al.* (2001) e Feam (2013) relatam que o hipoclorito de sódio e o clorito de sódio estão sendo substituídos pelo peróxido de hidrogênio (H_2O_2), visando eliminar a reação de agentes clorados com matéria orgânica, que resulta na formação de organoclorados, produtos potencialmente tóxicos.

Segundo Zaboenco (2012), com o intuito de evitar a formação de cloroaminas que provocam uma série de degradação à fibra celulósica (algodão), quando se utiliza hipoclorito de sódio, pode-se usar um anti cloro (bissulfito de sódio).

O modelo matemático que representa a variável resposta proporção de itens com defeitos, considerando os termos que influenciam o processo é expressa pela **Equação 28**.

$$\eta = \frac{1}{1 + e^{[-(-1,1136+0,3615B)]}} \quad (Eq. 28)$$

Onde: *B representa o volume de hipoclorito de sódio.

Como o volume de hipoclorito de sódio (B) é o único fator que exerce influência sobre a variável resposta e o intuito é minimizar a proporção de itens com defeitos, opta-se pela escolha do nível inferior (-1).

5.

Conclusão

A intensa competição no mercado nacional e internacional tem revelado que a qualidade dos produtos é a chave para o sucesso das empresas. Os experimentos estatísticos constituem uma importante, se não a mais importante, ferramenta para a melhoria da qualidade de processos. No caso de processos bem representados pelo modelo linear clássico em que se baseia o método de Mínimos Quadrados Ordinários (MQO), as técnicas de planejamento e análise de experimentos são bem conhecidas.

Porém os modelos lineares clássicos apresentam limitações em sua análise, especialmente em casos práticos, como em processos industriais têxteis, onde frequentemente as características de qualidade violam um ou mais pressupostos do modelo (distribuição Normal, variância constante e aditividade). Quando isso ocorre esses modelos apresentam estimativas para a variável resposta menos confiável.

Descreveu-se detalhadamente nesta dissertação a estrutura dos Modelos Lineares Generalizados (MLG), que permitem representar dados que apresentam uma sobredispersão ou uma subdispersão, detalhando alguns aspectos importantes como: a aplicação de testes de significância dos coeficientes e de estatísticas e gráficos mais indicados para verificar a adequação do modelo.

Através do modelo Logit conseguiu-se estabelecer uma relação entre as variáveis independentes: concentração de hidróxido de sódio (A) e volume de hipoclorito de sódio (B) e a variável dependente: proporção de itens com defeitos. Mostrou-se também que o único fator que influencia significativamente o processo de beneficiamento têxtil (alvejamento) da empresa é o hipoclorito de sódio (B).

Como o objetivo de qualquer empresa no mundo é diminuir o percentual de defeitos sem perder a qualidade de seus produtos, encontrar a combinação ótima dos níveis das variáveis independentes que minimize o valor da variável resposta é essencial. Neste aspecto, como o fator B é o único fator significativo para o modelo, conclui-se que a escolha pelo nível inferior (-1) irá minimizar a proporção de itens com defeitos.

Como o fator A não é significativo para o modelo pode-se também utilizá-lo em seu nível inferior (-1), com o intuito de reduzir os custos com insumos químicos da empresa. Conseqüentemente a interação entre os fatores (AB) será utilizada em seu nível máximo (+1).

Para investigações futuras sugere-se:

i) *A validação do experimento realizado em escala piloto em escala industrial, isto é, confirmar os resultados encontrados em laboratórios na empresa;*

ii) *Outros alvejantes químicos podem ser utilizados como: o peróxido de hidrogênio (H_2O_2) e o clorito de sódio ($NaClO_2$), podendo-se compará-los com o utilizado nesta dissertação, verificando qual seria mais eficiente e não geraria impactos na etapa de tratamento de efluentes;*

iii) *A malha de algodão apresenta ceras, graxas e substâncias solúveis, que poderiam ser melhor removidas com a utilização de detergentes, que são capazes de saponificá-las. Outra observação quando se utiliza hipoclorito de sódio ($NaClO$) como alvejante é o cloro residual presente no banho, que pode degradar a fibra e tornar o efluente de difícil tratamento, sendo necessário a utilização de anticloro. Portanto, pode-se propor uma mudança de processo com a inclusão de duas novas variáveis independentes;*

iv) *O processo de alvejamento na empresa é realizado em temperatura ambiente, portanto, o tempo de processo é demasiadamente lento. Com o intuito de diminuir o tempo de processamento, pode-se utilizar uma temperatura em torno de $60^\circ C$ e estudar a implicação dessa variável no processo;*

v) *Outros métodos para testar a significância dos coeficientes podem ser utilizados, como exemplo, a função desvio (deviance). Comparar este método com o utilizado nesta dissertação, verificando os prós e contras;*

vi) Os Modelos de Quase-Verossimilhança (MQV) podem ser utilizados quando não há necessidade de definir a distribuição de probabilidade. Portanto, caso os dados dessa dissertação não seguissem uma distribuição Binomial poderia se utilizar este método.

6.

Referências Bibliográficas

ADEWALE, A. J.; XU, X. Robust designs for generalized linear models with possible overdispersion and misspecified link functions. **Computational Statistics And Data Analysis**, Canada, v. 54, p.875-890, 2010.

AGRESTI, A. **Categorical Data Analysis**. New Jersey: John Wiley & Sons, 2012. 744 p.

ANDRADE, G. J. P. O. **Um método de diagnóstico do potencial de aplicação da manufatura enxuta na indústria têxtil**. 2006. 298 f. Tese (Doutorado) - Universidade Federal de Santa Catarina, Florianópolis, 2006.

ANTONY, J. **Design of experiments for engineers and scientists**. Oxford: Butterworth-Heinemann, 2003. 152 p.

ASEVEDO, F. R. **Abordagem Linear Generalizada para Estimar Perdas Não Técnicas de Energia Elétrica**. 2011. 76 f. Dissertação (Mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2011.

ATKINSON, A. C. **Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis**. Clarendon Press, 1985. 282 p.

ATKINSON, A.; RIANI, M. **Robust Diagnostic Regression Analysis**. New York: Springer-verlag, 2000.

BARRETO, I. D. C. *et al.* **Regressão Logística: Fenômenos de Separação**. In: Simpósio de Matemática e Estatística do Delta, 1., 2012, Piauí. **Anais...** 2012. p. 1 - 5.

BARROS NETO, B., SCARMINIO, I. S., BRUNS, R. E. **Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria**. Campinas: EDUNICAMP, 2007.

BARROS NETO, B.; SCARMÍNIO, I. S.; BRUNS, R. E. **Planejamento e otimização de experimentos**. Campinas: UNICAMP, 1995. 299 p.

BELLO, L. H. A. D. **Modelagem em Experimentos Mistura-Processo para Otimização de Processos Industriais**. 2010. 155 f. Tese (Doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2010.

BELTRAME, L. T. C. **Caracterização de Efluente Têxtil e Proposta de Tratamento**. 2000. 161 f. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Norte, Natal, 2000.

BIASOLI, P. K. **Modelagem conjunta de média e variância em experimentos fracionados sem repetição utilizando GLM**. 2005. 123 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.

BIAZUS, J. P. M. *et al.* Optimization of drying process of Zea mays malt to use as alternative source of amyolytics enzymes. **Brazilian Archives Of Biology And Technology**, v. 48, p.185-190, 2005.

BONANNI, C. G. **Uso combinado das técnicas DOE e Simulação Monte Carlo em um processo de soldagem MIG/MAG pulsado**. 2005. 104 f. Dissertação (Mestrado) – Curso de Engenharia de Produção, Universidade Federal de Itajubá, Itajubá, 2005.

BONDUELLE, G. M. Aplicação do Planejamento de Experimentos no Controle da Fabricação de Chapas de Fibras de Madeira. **CERNE**, Paraná, v. 6, n. 2, p.1-10, 2000.

BORGES, A. R. **Aplicação de um biopolímero de origem vegetal no tratamento físico-químico de um efluente têxtil**. 2010. 106 f. Monografia (Especialização) – Faculdade SENAI CETIQT, Rio de Janeiro, 2010.

BORTOLINI, J. **Estudo de experimentos fatoriais 2^k aplicados em um processo industrial**. 2012. 143 f. Dissertação (Mestrado) - Universidade Federal de Lavras, Lavras, 2012.

CHOUDHURY, A. K. R. **Textile Preparation and Dyeing**. Science Publisher, 2006. 834 p.

COOK, R. D.; WEISBERG, S. **Applied Regression Including Computing and Graphics**. New York: John Wiley & Sons, 1999. 632 p.

CORDEIRO, G. M.; DEMÉTRIO, C. G. B. **Modelos Lineares Generalizados e Extensões**. Piracicaba: ESALQ/USP, 2013.

COSTA, A. F. B.; EPPRECHT, E. K.; CARPINETTI, L. C. R. **Controle Estatístico de Qualidade**. 2. ed. São Paulo: Atlas, 2011. 334 p.

COSTA, A. F. S. **Aplicação de tratamentos biológico e físico-químico em efluentes de lavanderia e tinturaria industriais do município de Toritama no estado de Pernambuco**. 2008. 100 f. Dissertação (Mestrado) - Universidade Católica de Pernambuco, Recife, 2008.

COUTO, M. R. M. *et al.* Transformações de dados em experimentos com abobrinha italiana em ambiente protegido. **Ciência Rural**, Santa Maria, v. 39, n. 6, p.1701-1707, set. 2009.

COX, C. Nonlinear quasi-likelihood models: applications to continuous proportions. **Computational Statistics & Data Analysis**, Rochester, v. 21, p.449-461, 1996.

CRUSCO, N. A. *et al.* Análise de regressão linear múltipla para simulação da banda do SWIR com outras bandas espectrais. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 12., 2005, Goiânia. **Anais...**, 2005. p. 891 - 898.

CRUZ, M. C. C. **O impacto da amamentação sobre a desnutrição e a mortalidade infantil, Brasil, 1996**. 2001. 80 f. Dissertação (Mestrado) - Fundação Oswaldo Cruz, Rio de Janeiro, 2001.

CUNHA, E. J. **Análise da influência das variáveis da caldeira de recuperação química no sistema de abate do material particulado da indústria de celulose e papel**. 2005. 164 f. Dissertação (Mestrado) - Universidade Federal de Itajubá, Itajubá, 2005.

DAMASCENO, O. I. C. **Remoção de corantes em solução aquosa e efluentes de indústrias têxteis através de adsorção em cabelo humano**. 2008. 85 f. Dissertação (Mestrado) - Curso de Agroquímica, Universidade Federal de Viçosa, Viçosa, 2008.

DAVIS, C. S. **Statistical Methods for the Analysis of Repeated Measurements**. New York: Springer, 2002.

DEMÉTRIO, C. G. B. **Modelos Lineares Generalizados em Experimentação Agrônômica**. Piracicaba: ESALQ/USP, 2002. 121 p.

ESTATCAMP. **Modelo Estatístico: Regressão Logística**. Disponível em: <<http://www.portalaction.com.br/688-31-modelo-estatístico>>. 1997-2011. Acesso em: 21 abr. 2014.

FEAM. **Plano de ação para adequação ambiental das indústrias do setor têxtil do Estado de Minas Gerais**. Belo Horizonte: FEAM, 2013. 168 p.

FERREIRA, D. V. **Análise comparativa de produtos têxteis no segmento de limpeza industrial: estopa, toalha industrial e wipes**. 2010. 86 f. Monografia (Graduação) - Faculdade SENAI/CETIQT, Rio de Janeiro, 2010.

FIGUEIREDO FILHO, D. B.; SILVA JÚNIOR, J. A. Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r). **Revista Política Hoje**, Pernambuco, v. 18, n. 1, p.115-146, 2009.

FOX, J. **Applied Regression Analysis and Generalized Linear Models**. Los Angeles: SAGE, 2008. 688 p.

GALDÁMEZ, E. V. C. **Aplicação das técnicas de planejamento e análise de experimentos na melhoria da qualidade de um processo de fabricação de produtos plásticos**. 2002. 133 f. Dissertação (Mestrado) – Universidade de São Paulo, São Carlos, 2002.

GOMES, U. R. **Otimização do Processo de Laminação a Frio através de planejamentos de Experimentos**. 2007. 76 f. Dissertação (Mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007.

GONÇALVES, A. M. S. **Metodologias estatísticas aplicadas à relação entre Eventos Climáticos Extremos, Saúde e Desigualdades Socioeconômicas na Grande Área Metropolitana do Porto**. 2012. 68 f. Tese (Doutorado) - Universidade do Minho, Porto, 2012.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5. ed. Porto Alegre: AMGH, 2011. 924 p.

HABITZREUTER, E. **Otimização de um Processo de Alveamento e Cozinhamento Simultâneo de Malha 100% Algodão**. 2013. 113 f. Monografia (Especialização) - Curso de Engenharia Industrial Têxtil, Faculdade SENAI CETIQT, Rio de Janeiro, 2013.

HAINING, R.; LAW, J.; GRIFFITH, D. Modelling small area counts in the presence of overdispersion and spatial autocorrelation. **Computational Statistics And Data Analysis**. 2008. p. 2923-2937.

HILBE, J. M. **Negative Binomial Regression**. 2. ed. Uk: Cambridge University Press., 2008.

HINDE, J.; DEMÉTRIO, C. G. B. Overdispersion: Models and estimation. **Computational Statistics & Data Analysis**, v. 27, p.151-170, 1998.

HINDE, J.; DEMÉTRIO, C. G. B. **Overdispersion: Models and Estimation**. FAPESP, 2007. 73 p.

HINES, W. W. *et al.* **Probabilidade e estatística na engenharia**. 4. ed. Rio de Janeiro: LTC, 2006. 588 p.

HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression**. 2. ed. John Wiley & Sons, 2000.

IMMICH, A. P. S. **Remoção de corantes de efluentes têxteis utilizando folhas de Azadirachta indica como adsorvente**. 2006. 119 f. Dissertação (Mestrado) - Curso de Engenharia Química, Universidade Federal de Santa Catarina, Florianópolis, 2006.

JERÔNIMO, C. E. M. *et al.* Tratamento e redução na fonte de efluentes do processo de alveamento de tecidos. In: Congresso Brasileiro de Engenharia Sanitária e Ambiental, 23º, 2005, Campo Grande. **Anais...** Natal: ABES, 2005. p. 1 - 9.

KIESCHNICK, R.; MCCULLOUGH, B. D. Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. **Statistical Modelling**, Texas, v. 3, p.193-213, 2003.

KOHAN, L.; ARAUJO, M. C. Processos Enzimáticos na Indústria Têxtil: uma alternativa com menor impacto ambiental. In: simpósio internacional de Iniciação Científica da USP, 16., 2008, São Paulo. **Anais...** São Paulo: USP, 2008. p. 1 - 6.

LARA, L. A. C. **QTLs de feijão para resistência ao mofo-branco de cultivares adaptadas**. 2013. 91 f. Dissertação (Mestrado) - Curso de Genética e Melhoramento de Plantas, Universidade Federal de Lavras, Lavras, 2013.

LEE, Y.; NELDER, J. A. Generalized linear models for the analysis of quality improvement experiments. **The Canadian Journal Of Statistics**, v.26, p. 95-105, 1998.

LEVINE, D. M. **Estatística: Teoria e Aplicações Usando Microsoft Excel**. 6. ed. LTC, 2011. 804 p.

LEWIS, S. L., MONTGOMERY, D. C.; MYERS, R. H. Confidence Interval Coverage for Designed Experiments Analysed With GLMs. **Journal of Quality Technology**, v. 33, n. 3, p. 279-292., 2001(b).

LEWIS, S. L.; MONTGOMERY, D. C.; MYERS, R. H. Examples of Designed Experiments with Nonnormal Responses. **Journal Of Quality Technology**, v. 33, n. 3, p.265-278, 2001(a).

LINDSEY, J. K. **Applying Generalized Linear Models**. New York: Springer-verlag, 1997. 257 p.

LOVATTO, P. A. *et al.* Meta-análise em pesquisas científicas: enfoque em metodologias. **Revista Brasileira de Zootecnia**, 2007, v. 36, p.285-294, 2007.

LUCCA, L. **Controle de qualidade do Hipoclorito de Sódio no Processo de Produção**. Florianópolis: Universidade Federal de Santa Catarina, 2006. 40 p.

MARTIN, T. N.; STORCK, L. Análise das pressuposições do modelo matemático em experimentos agrícolas no delineamento blocos ao acaso. In: **MARTIN, T. N.; ZIECH, M. F. (Org)**. Sistemas de Produção Agropecuária. Curitiba: UTFPR, 2008. Cap.11, p. 177-196.

MATTOS, V. L. D. **Identificação de efeitos de dispersão em experimentos com poucas replicações**. 2004. 100 f. Tese (Doutorado) - Curso de Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2004.

MATTOS, V. L. D.; BARBETTA, P. A.; SAMOHYL, R. W. Identificação de efeitos de dispersão em experimentos fatoriais dos tipos 2^k e 2^{k-p} : Um estudo de caso. In: Encontro Nacional de Engenharia de Produção, 22., 2002, Curitiba. **Anais...** 2002: ABEPRO, 2002. p. 1 - 7.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. 2. ed. Chapman & Hall, 1989. 512 p.

MONTGOMERY, D. C. **Design and analysis of experiments**. 7. ed. New York: John Wiley & Sons, 2009. 656 p.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística Aplicada e Probabilidade para Engenheiros**. Rio de Janeiro: LTC, 2012.

MONTGOMERY, D.; CALADO, V. **Planejamento de Experimentos usando o Statistica**. E-papers Serviços Editoriais Ltda, 2003. 260 p.

MOSER, D. **Acompanhamento da análise e controle de processos têxteis**. Blumenau: Universidade Regional de Blumenau, 2011.

MYERS, R. H. *et al.* **Generalized Linear Models with applications in Engineering and the Sciences**. 2. ed. New York: John Wiley & Sons, 2002.

MYERS, R. H.; MONTGOMERY, D. C. **Response Surface Methodology**. 2. ed. New York: John Wiley & Sons; 2002.

NAIR, V. N.; PREGIBON, D. Analyzing dispersion effects from replicated factorial experiments. **Technometrics**. p. 247-257. ago. 1988.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. **Journal of the Royal Statistical Society**, v. 135, p.370-384, 1972.

OLIVEIRA, A. J. **Variação em itens lexicais terminados em /l/+vogal na região de Itaúna/MG**. 2006. 156 f. Dissertação (Mestrado) - Curso de Estudos Lingüísticos, Universidade Federal de Minas Gerais, Minas Gerais, 2006.

OLIVEIRA, S. **Inferência e análise de resíduos e de diagnóstico em Modelos Lineares Generalizados**. Monografia (Especialização) - Curso de Estatística, Universidade Federal de Juiz de Fora, Juiz de Fora, 2013.

PAULA, G. A. **Modelos de regressão com apoio computacional**. São Paulo: IME/USP, 2010. 403 p

PIMENTA, F. D. **Estudo de parâmetros de processo para a produção de biopolímeros de *Rhizobium tropici***. 2006. 226 f. Tese (Doutorado) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2006.

PINTO, E. R.; PEREIRA, L. A. Planejamento Ótimo de Experimento para Modelos Lineares Generalizados Usando o *Software R*. In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 34., 2012, Águas de Lindóia. **Anais...** Águas de Lindóia: SBMAC, 2012. p. 1260 - 1266.

ROBINSON, T. *et al.* Remediation of dyes in textile effluent: a critical review on current treatment technologies with a proposed alternative. **Bioresource Technology**, v. 77, n. 3, p.247-255, 2001.

SÁNCHEZ, J. C. **Introducción al blanqueo de materias textiles**. Barcelona: Romargraf, 1966. 348 p.

SANT'ANNA, A. M. O.; CATEN, C. S. T. Modelagem da fração de não-conformes em processos industriais. **Pesquisa Operacional**, Rio Grande do Sul, v. 30, n. 1, p.53-72, 2010.

SANTORO, K. R.; BARBOSA, S. B. P.; HOLANDA, M. C. R. Modelos de Predição da Natimortalidade em Suínos. **Revista Brasileira de Zootecnia**, Pernambuco, v. 32, n. 5, p.1131-1140, 2003.

SILVA, K. K. O. S. **Caracterização do efluente líquido no processo de beneficiamento do índigo têxtil**. 2007. 156 f. Dissertação (Mestrado) - Curso de Engenharia Mecânica, Universidade Federal do Rio Grande do Norte, Natal, 2007.

SOUZA, F. R. **Modelagem de experimentos planejados com respostas discretas**. 2013. 92 f. Tese (Doutorado) - Curso de Estatística e Experimentação Agropecuária, Universidade Federal de Lavras, Lavras, 2013.

SPECK, H. J. **Proposta de método para facilitar a mudança das técnicas de projetos: da prancheta à modelagem sólida (CAD) para empresas de engenharia de pequeno e médio porte**. 2005. 172 f. Tese (Doutorado) - Curso de Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2005.

STAACK, S. **Acompanhamento de processos industriais na empresa Staack Tinturaria LTDA**. Blumenau, 2007.

STEEL, R. G. D.; TORRIE, J. H.; DICKEY, D. A. **Principles and procedures of statistics: a biometrical approach**. New York: Mcgraw-hill, 1997. 666 p.

STORCK, L. *et al.* Análise de covariância para melhoria da capacidade de discriminação em ensaios de cultivares de milho. **Pesquisa Agropecuária Brasileira**, Brasília, v. 35, n. 7, p.1311-1316, 2000.

TURKMAN, M. A. A.; SILVA, G. L. **Modelos Lineares Generalizados: da teoria à prática**. Lisboa: FCT - PRAXIS XXI - FEDER, 2000.

TWARDOKUS, R. G. **Reuso de água no processo de tingimento da indústria têxtil**. 2004. 136 f. Dissertação (Mestrado) - Curso de Engenharia Química, Universidade Federal de Santa Catarina, Florianópolis, 2004.

VENTICINQUE, E. M. *et al.* O uso de regressão logística para espacialização de probabilidades. **Megadiversidade**, Belo Horizonte, v. 3, n. 1-2, p.25-37, 2007.

VIEIRA, A. F. C. **Análise da Média e Dispersão em Experimentos Fatoriais não Replicados para Otimização de Processos Industriais**. 2004. 204 f. Tese (Doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2004.

VIEIRA, A. F. C.; EPPRECHT, E. K. Métodos de identificação de efeitos na dispersão em experimentos fatoriais não replicados. **Gestão & Produção**, São Carlos, v. 16, n. 1, p.99-110, 2009.

VITTINGHOFF, E. *et al.* **Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models**. 2. ed. New York: Springer, 2012. 550 p.

ZABOENCO, M. A. **Produção de algodão hidrófilo**. 2012. 105 f. TCC (Graduação) - Curso de Engenharia Química, Universidade Regional de Blumenau, Blumenau, 2012.