Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Modelos estatísticos integrados à metodologia Lean Seis Sigma
visando ao aumento da produtividade na obtenção do etanol

Wilson Alves de Oliveira

Tese apresentada para obtenção do título de
doutor em Ciências. Área de Concentração:
Estatística e Experimentação Agronômica

P I R A C I C A B A
2010

Wilson Alves de Oliveira
Bacharel em Estatística

**Modelos estatísticos integrados à metodologia Lean Seis Sigma visando ao
aumento da produtividade na obtenção do etanol**

Orientador:
Prof. Dr. **DÉCIO BARBIN**

Tese apresentada para obtenção do título de Doutor
em Ciências. Área de concentração: Estatística e
Experimentação Agronômica

**Piracicaba
2010**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Oliveira, Wilson Alves de

Modelos estatísticos integrados à metodologia Lean Seis Sigma visando ao aumento da produtividade na obtenção do etanol / Wilson Alves de Oliveira. - - Piracicaba, 2010.
139 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2010.

1. Análise de regressão e correlação 2. Biocombustíveis 3. Cana-de-açúcar
4. Estatística - Modelos 5. Etanol 6. Produtividade I. Título

CDD 661.82
O48m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

À

DEUS,

pela presença em todos os momentos da minha vida,
mostrando os caminhos nas horas mais difíceis.

Aos meus pais, Ezequiel Alves de Oliveira e Delfina
Cabral de Oliveira, pela educação.

À minha esposa Irene Carniatto de Oliveira, pelo
incentivo e compreensão em todos os momentos,
principalmente, nos momentos de ausência.

“A casa e os bens vêm como herança dos pais; mas
do senhor, a esposa prudente” (Provérbios: 19;14).

Aos meus irmãos Luiz Carlos,
Elisabete, Eliane, André Luís e ao
meu primo (irmão) Andrey Lúcio.

Às minhas enteadas Evelyn e Esloany e
às minhas sobrinhas Isabela, Gabriela,
Thaís e Amanda.

AGRADECIMENTOS

Ao Prof. Dr. Décio Barbin, pela orientação segura na tese de doutorado, como também, em todos os momentos desde 1996 quando cheguei para o mestrado na ESALQ, que com sua experiência, tranquilidade e apoio transformou esse convívio em uma amizade, que eu e minha esposa Irene levaremos por toda a vida.

Ao estatístico Luíz Carlos Alves de Oliveira pela sugestão do tema e auxílio no desenvolvimento da metodologia “Lean”, por seu empenho e dedicação para cursar a graduação na UNICAMP e como o irmão mais velho tornando-se exemplo à família.

Ao engenheiro mecânico João Hamilton P. Gomes pela sua colaboração no mapeamento do processo na usina.

À Usina DECASA - Destilaria de Açúcar e Álcool Caiua S/A, aos diretores, engenheiros e funcionários pela disposição em contribuir, cedendo os dados e informações para a realização desta pesquisa.

À Escola Superior de Agricultura “Luíz de Queiroz” (ESALQ/USP), referência do saber desta área no Brasil e no mundo, que pelos seus conceituados professores, abre suas portas para nos receber e dividir seus conhecimentos.

Aos professores do Departamento de Ciências Exatas e do Programa de Pós-graduação em Estatística e Experimentação Agronômica da ESALQ/USP: Dr. Carlos Tadeu dos Santos Dias, Dr^a. Clarice Garcia Borges Demétrio, Dr. Edwin Ortega, Dr. Gabriel Adrián Sarriés, Dr. Gerson Barreto Mourão, Dr. Silvio Sandoval Zocchi e Dr. Vitor Ozaki pelos conhecimentos nas disciplinas ministradas ou por suas sugestões para este estudo.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, Solange de Assis Paes Sabadin, Luciane Brajão, Eduardo Bonilha, Jorge Alexandre Wiendl, Expedita Maria de Azevedo e Rosni Pinto, pela dedicação e disposição, proporcionando condições para a realização deste trabalho.

À Renata Alcarde e Vanderly Janeiro pela amizade, companheirismo, convívio alegre durante o curso, e pelo auxílio nos programas R e Latex.

Aos amigos do curso de pós-graduação, Ana A. Pilon, Ângela M. Coelho, Edila C. Souza, Fernanda B. Rizzato, Juliana B. Fachini, Júlio C. Pereira, Lúcio B. Araújo,

Lucimary A. dos Santos, Marina R. Maestre, Osmar J. Macedo, Sandra Vergara, Simone D. Sartório e Simone S.W.G. do Amaral.

À Universidade Estadual do Oeste do Paraná pela oportunidade e incentivos dos colegas professores e funcionários.

À Coordenadoria para o Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudo.

À todos que, de alguma forma, contribuíram para a realização deste trabalho.

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	23
2.1 Etanol Combustível	23
2.1.1 Histórico da cana-de-açúcar	23
2.1.2 O plantio e safra da cana-de-açúcar no Brasil	24
2.1.3 Colheita da cana-de-açúcar - Legislação Ambiental	25
2.1.4 Definição de Etanol	27
2.1.5 Os produtos obtidos da cana-de-açúcar	28
2.1.6 A matriz energética do Brasil	28
2.1.7 Avaliação dos atributos tecnológicos	29
2.2 Controle e Melhoria da Qualidade	30
2.2.1 Filosofia da Qualidade	31
2.2.2 Desdobramento da função qualidade (Quality Function Deployment - QFD)	32
2.2.3 A Filosofia Lean	33
2.2.4 Seis Sigma	39
2.2.5 Análise de regressão	43
3 MATERIAL E MÉTODOS	47
3.1 Material	47
3.1.1 Processos básicos de produção	47
3.1.1.1 Extração do caldo	47
3.1.1.2 Tratamento do caldo	48
3.1.1.3 Fermentação	48
3.1.1.4 Destilação	49
3.2 Métodos	51
3.2.1 A Filosofia Lean	51

3.2.2 O Programa Seis Sigma	54
3.2.2.1 Melhoria do Processo	55
3.2.2.2 Integração das ferramentas <i>Lean</i> e Seis Sigma ao DMAIC	56
3.2.3 Regressão Linear Múltipla	59
3.2.3.1 Modelo Estatístico	59
3.2.3.2 Método dos Quadrados Mínimos	61
3.2.3.3 Teste t e teste F	61
3.2.3.4 Coeficiente de Correlação Múltiplo	62
3.2.3.5 Métodos de seleção de variáveis	63
3.2.3.6 Análise de resíduos	65
3.2.3.6.1 Diagnósticos em influência	66
3.2.3.6.2 Estatísticas para diagnósticos	67
4 RESULTADOS E DISCUSSÃO	69
4.1 Mapeamento do processo	69
4.1.1 Extração do caldo	74
4.1.2 Tratamento do caldo	76
4.1.3 Fermentação	78
4.1.4 Destilação	80
4.2 Regressão linear múltipla para o processo da extração do caldo	82
4.2.1 Métodos de seleção de variáveis	82
4.2.2 Análise de Resíduos	83
4.2.3 Estimação dos parâmetros e análise de variância	87
4.3 Regressão linear múltipla para o processo do tratamento do caldo	88
4.3.1 Regressão linear múltipla para a variável Y_{i1} (brix do caldo caleado)	88
4.3.1.1 Métodos de seleção de variáveis	88
4.3.1.2 Análise de Resíduos	90
4.3.1.3 Estimação dos parâmetros e análise de variância	94
4.3.2 Regressão linear múltipla para a variável Y_{i2} (Pol do caldo clarificado)	96
4.3.2.1 Métodos de seleção de variáveis	96
4.3.2.2 Análise de Resíduos	97

4.3.2.3	Estimação dos parâmetros e análise de variância	101
4.3.3	Regressão linear múltipla para a variável Y_{i4} (AR decantado)	103
4.3.3.1	Métodos de seleção de variáveis	104
4.3.3.2	Análise de Resíduos	104
4.3.3.3	Estimação dos parâmetros e análise de variância	108
4.3.4	Regressão linear múltipla para a variável Y_{i5} (ART do caldo)	110
4.3.4.1	Métodos de seleção de variáveis	110
4.3.4.2	Análise de Resíduos	111
4.3.4.3	Métodos de seleção de variáveis excluindo-se a observação número 3	111
4.3.4.4	Análise de Resíduos	112
4.3.4.5	Estimação dos parâmetros e análise de variância	117
4.4	Regressão linear múltipla para o processo da fermentação	118
4.5	Regressão linear múltipla para o processo da destilação	118
4.5.1	Métodos de seleção de variáveis	119
4.5.2	Análise de Resíduos	119
4.5.2.1	Estimação dos parâmetros e análise de variância	125
5	CONSIDERAÇÕES FINAIS	129
5.1	Conclusões	130
	REFERÊNCIAS	131
	APÊNDICE	135

RESUMO

Modelos estatísticos integrados à metodologia Lean Seis Sigma visando ao aumento da produtividade na obtenção do etanol

A cana-de-açúcar é uma das principais atividades econômicas do Brasil, tanto em termos de balança comercial, como de geração de empregos. O álcool (etanol) obtido da cana-de-açúcar é um combustível menos poluente, pois, contém oxigênio na sua composição e é utilizado nos veículos movidos a etanol, na composição da gasolina automotiva e nos veículos flex. A queima do etanol produz 25% menos monóxido de carbono e 35% menos óxido de nitrogênio que a gasolina. Desta forma, as nações preocupadas em diminuir seus índices de poluição atmosférica têm considerado como uma importante solução viável, a adoção do etanol como combustível automotivo. O objetivo deste trabalho foi apresentar propostas para a melhoria da qualidade e produtividade do setor, focado no desempenho dos processos e produtos, através de métodos estatísticos. Foi utilizada a metodologia “Lean Seis Sigma” aplicada no processo produtivo do etanol, em uma usina de açúcar e de álcool, visando acelerar a velocidade do processo, através da redução do desperdício e da variabilidade. A aplicação da metodologia “Lean Seis Sigma” foi realizada com o uso do modelo DMAIC (*Define, Measure, Analyze, Improve, Control*). A ferramenta utilizada para descrever a relação entre as variáveis de saída (Y) e as variáveis de entrada (X) foi a regressão linear múltipla, fazendo uso dos métodos de seleção de variáveis. O mapeamento do processo produtivo do etanol possibilitou a identificação das etapas que agregam e não agregam valor ao processo. Os resultados mostraram uma redução significativa no “Lead Time” (tempo do processo) e a identificação das principais variáveis envolvidas no processo produtivo do etanol.

Palavras-chave: Cana-de-açúcar; *Lean Seis Sigma*; Análise de regressão; Biocombustível

ABSTRACT

Statistical models integrated with Lean Six Sigma methodology in order to increase productivity in obtain of ethanol

The sugarcane is one of the most important economic businesses in Brazil, in both terms of trade balance and job opportunity. Alcohol (ethanol), obtained from sugarcane, is a fuel cleaner, because it contains oxygen in its composition and is used in vehicles powered by ethanol, on the composition of automotive gasoline and flex-fuel vehicles. Ethanol burning produces 25% less carbon monoxide and 35% less nitrogen oxide than gasoline. Thus, the nations that concern on reducing their levels of air pollution consider it as an important and available solution to be adopted as an automotive fuel. So, this trial aimed at making some proposals for improving quality and productivity of this sector, focused on the performance of processes and products, according to some statistical methods. The studied methodology was Lean Six Sigma, applied in the production process of ethanol, in a sugar and alcohol Plant in order to speed up its process by reducing waste and variability. The Lean Six Sigma methodology was carried out using the DMAIC model (Define, Measure, Analyze, Improve and Control). The multiple linear regression tool described the relationship among the output variables (Y) and the input variables (X), associated to the methods of variable selection. The mapping of ethanol production process enabled the identification of steps that add, and does not add value to the process. The results showed a significant reduction in the Lead Time as well as to identify the main variables involved in the production process of ethanol.

Keywords: Sugarcane; Lean Six Sigma; Regression analysis; Biofuel

LISTA DE FIGURAS

Figura 1 - Mapa da produção de cana-de-açúcar no Brasil	25
Figura 2 - Modelo de uma empresa vista através da perspectiva do fluxo do processo .	55
Figura 3 - VSM - Mapeamento do fluxo de valor do estado atual	70
Figura 4 - VSM - Mapeamento do fluxo de valor do estado futuro	73
Figura 5 - PMAP - Extração do caldo	75
Figura 6 - PMAP - Tratamento do caldo	77
Figura 7 - PMAP - Fermentação	79
Figura 8 - PMAP - Destilação para o álcool hidratado	81
Figura 9 - Gráfico dos valores ajustados e resíduos estudentizados externamente	83
Figura 10 - Gráfico normal de probabilidades	84
Figura 11 - Histograma dos resíduos estudentizados externamente	84
Figura 12 - <i>Box plot</i> dos resíduos estudentizados externamente	85
Figura 13 - Gráfico dos valores absolutos de DFFitS	85
Figura 14 - Gráfico dos valores observados e valores ajustados pelo modelo (29)	88
Figura 15 - Gráfico dos valores ajustados e resíduos estudentizados externamente	90
Figura 16 - Gráfico normal de probabilidades	91
Figura 17 - Histograma dos resíduos estudentizados externamente	91
Figura 18 - <i>Box plot</i> dos resíduos estudentizados externamente	92
Figura 19 - Gráfico do logaritmo da função de verossimilhança perfilada para o modelo de transformação de <i>Box-Cox</i>	92
Figura 20 - Gráfico dos valores absolutos de DFFitS	93
Figura 21 - Gráfico dos valores observados e valores ajustados pelo modelo (34)	95
Figura 22 - Gráfico dos valores ajustados e resíduos estudentizados externamente	97
Figura 23 - Gráfico normal de probabilidades	98
Figura 24 - Histograma dos resíduos estudentizados externamente	98
Figura 25 - <i>Box plot</i> dos resíduos estudentizados externamente	99
Figura 26 - Gráfico do logaritmo da função de verossimilhança perfilada para o modelo de transformação de <i>Box-Cox</i>	99
Figura 27 - Gráfico dos valores absolutos de DFFitS	100

Figura 28 - Gráfico dos valores observados e valores ajustados pelo modelo (41)	102
Figura 29 - Gráfico dos valores observados e valores ajustados pelo modelo (42)	103
Figura 30 - Gráfico dos valores observados para as variáveis $X's$ e Y_4	104
Figura 31 - Gráfico dos valores ajustados e resíduos estudentizados externamente	105
Figura 32 - Gráfico normal de probabilidades	105
Figura 33 - Histograma dos resíduos estudentizados externamente	106
Figura 34 - <i>Box plot</i> dos resíduos estudentizados externamente	106
Figura 35 - Gráfico dos valores absolutos de DFFitS	107
Figura 36 - Gráfico dos valores observados e valores ajustados pelo modelo (47)	109
Figura 37 - Gráfico dos valores ajustados e resíduos estudentizados externamente e o gráfico normal de probabilidades	111
Figura 38 - Gráfico dos valores ajustados e resíduos estudentizados externamente	112
Figura 39 - Gráfico normal de probabilidades	113
Figura 40 - Histograma dos resíduos estudentizados externamente	113
Figura 41 - <i>Box plot</i> dos resíduos estudentizados externamente	114
Figura 42 - Gráfico do logaritmo da função de verossimilhança perfilada para o modelo de transformação de <i>Box-Cox</i>	114
Figura 43 - Gráfico dos valores absolutos de DFFitS	115
Figura 44 - Gráfico dos valores observados e valores ajustados pelo modelo (54)	118
Figura 45 - Gráfico dos valores ajustados e resíduos estudentizados externamente	120
Figura 46 - Gráfico normal de probabilidades	120
Figura 47 - Histograma dos resíduos estudentizados externamente	121
Figura 48 - <i>Box plot</i> dos resíduos estudentizados externamente	121
Figura 49 - Gráfico dos valores absolutos de DFFitS	122
Figura 50 - Gráfico dos valores observados e valores ajustados pelo modelo (59)	127

LISTA DE TABELAS

Tabela 1 - Significado do 5S	36
Tabela 2 - Tabela de conversão em sigma	41
Tabela 3 - Esquema da análise de variância de regressão	62
Tabela 4 - Tabela de análise da variância para comparação dos modelos para a variável (Y_{i1}) do processo da extração do caldo	83
Tabela 5 - Estatísticas para diagnósticos em medidas de influência	86
Tabela 6 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões . .	87
Tabela 7 - Análise de variância da regressão	87
Tabela 8 - Tabela de análise da variância para comparação dos modelos para a variável (Y_{i1}) do processo do tratamento do caldo	89
Tabela 9 - Estatísticas para diagnósticos em medidas de influência	94
Tabela 10 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões . .	95
Tabela 11 - Análise de variância da regressão	96
Tabela 12 - Tabela de análise da variância para comparação dos modelos para a variável (Y_{i2}) do processo do tratamento do caldo	97
Tabela 13 - Estatísticas para diagnósticos em medidas de influência	100
Tabela 14 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões . .	101
Tabela 15 - Análise de variância da regressão	102
Tabela 16 - Estatísticas para diagnósticos em medidas de influência	107
Tabela 17 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões . .	108
Tabela 18 - Análise de variância da regressão	109
Tabela 19 - Tabela de análise da variância para comparação dos modelos para a variável (Y_{i5}) do processo do tratamento do caldo	110
Tabela 20 - Estatísticas para diagnósticos em medidas de influência	116
Tabela 21 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões . .	117
Tabela 22 - Análise de variância da regressão	118
Tabela 23 - Tabela de análise da variância para comparação dos modelos para a variável Y_i do processo da destilação	119
Tabela 24 - Estatísticas para diagnósticos em medidas de influência	123

Tabela 24 -Estatísticas para diagnósticos em medidas de influência	124
Tabela 25 -Estimativas dos coeficientes da regressão e os, respectivos, erros padrões . .	126
Tabela 26 -Análise de variância da regressão	127

1 INTRODUÇÃO

O Brasil é um dos maiores produtores de cana-de-açúcar (*Saccharum spp.*) do mundo e, conseqüentemente, um dos países com maior exportação de açúcar e produtos derivados. A cana-de-açúcar transforma-se em combustível que utiliza uma fonte renovável, ao contrário do petróleo, combustível fóssil que precisa de milhões de anos para se formar. Cada tonelada de cana tem o potencial energético de 1,2 barril de petróleo (CAPUTO, 2006).

A cana-de-açúcar é uma das principais atividades econômicas do Brasil, tanto em termos de balança comercial, como de geração de empregos. Movimenta cerca de R\$ 50,0 bilhões, o que equivale a 1,5 % do PIB (Produto Interno Bruto) do Brasil. O setor gera cerca de 3,6 milhões de empregos diretos e indiretos, envolvendo 72.000 agricultores (INFORMAÇÃO UNICA, 2007b).

A cana é uma gramínea produzida em quase uma centena de países que estão situados entre as zonas tropical e equatorial. Basicamente, é constituída de fibras e caldo, sendo o caldo a matéria prima da indústria sucroalcooleira. Quanto maior o teor de sacarose da cana, maior pode ser a produção de açúcar e álcool (etanol). O teor de fibra é importante, pois o bagaço (subproduto do processo de extração do caldo) é usualmente empregado na produção de energia gerada do vapor produzido pela sua queima (WALTER, 1993).

O etanol anidro é usado na produção da gasolina C, que é a única gasolina que pode ser comercializada no território nacional para abastecimento de veículos automotores. As distribuidoras de combustíveis adquirem o etanol anidro das destilarias e a gasolina A (pura) das refinarias, fazendo uma mistura desses dois na proporção que pode variar entre 20 e 25% de etanol anidro. O etanol hidratado é usado diretamente no abastecimento de veículos automotores. (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2007).

O etanol é um combustível menos poluente, pois contém oxigênio em sua composição. Assim, a combustão (reação que normalmente retira o oxigênio da atmosfera) do álcool é mais fácil do que a da gasolina. A queima do álcool produz em média 25% menos monóxido de carbono e 35% menos óxido de nitrogênio que a gasolina (INFORMAÇÃO UNICA, 2007b).

A cultura da cana-de-açúcar é feita em todo território brasileiro. Dos 60 milhões de hectares destinados à agricultura (correspondendo a 7% do território nacional) a área utilizada com a produção de cana-de-açúcar atingiu 7 milhões de hectares em 2007, bem abaixo dos 21 milhões de hectares destinados a soja e aos 12 milhões de hectares do milho. As áreas de pastagens correspondem a cerca de 227 milhões de hectares. De acordo com levantamento da Embrapa (Empresa Brasileira de Pesquisa Agropecuária) existem cerca de 100 milhões de hectares em condições de absorver o crescimento da agricultura. Entre 2000 e 2006, a área plantada aumentou 43%. No entanto, especialistas do setor afirmam que não

há no Brasil conflito entre usos da terra para alimentos e energia (INFORMAÇÃO UNICA, 2007a; UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2008).

O Brasil possui, aproximadamente, 380 unidades produtoras. Na safra 2008/09 obteve uma produção de 569 milhões de toneladas, sendo, cerca de 89% do total, na principal região produtora do País, a Centro-Sul, e 11% na região Norte-Nordeste. Deste total, produziu 27,513 bilhões de litros de etanol e 31,049 milhões de toneladas de açúcar, tendo exportado 4,7219 (17%) bilhões de litros de etanol e 20,7949 (67%) milhões de toneladas de açúcar. Cerca de 60% da cana foi usada para a produção do etanol e 40% para a produção de açúcar (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2010).

O etanol, obtido da cana-de-açúcar, é o biocombustível de maior produtividade no mundo. A produção nacional cresce à taxa de 9% ao ano. O Estado de São Paulo é o maior produtor de cana-de-açúcar do país, sendo responsável por, aproximadamente, 60% da produção nacional. Na região Centro-Sul do Brasil, a média de produção de etanol é de 6.900 litros por hectare (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2008).

Em 2007, a participação dos veículos Flex-Fuel no total de automóveis comerciais leves novos vendidos no Brasil atingiu a marca de 85,6%, um crescimento acentuado se comparado aos 3,7% dos veículos novos vendidos em 2003, ano em que veículos Flex foram introduzidos no mercado brasileiro. Em dezembro de 2007, o número de veículos Flex em circulação no País chegou a 4,5 milhões de veículos, e em novembro de 2008, o Brasil atingiu 7 milhões de carros Flex fabricados (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2009a).

Desta forma, as nações com interesse em diminuir a emissão de gases poluentes, estão cada vez mais interessadas neste combustível. Consequentemente, as exportações de etanol devem aumentar de forma acentuada, nos próximos anos no Brasil. Para a safra 2012/13, a projeção do setor é de que a área ocupada com cana-de-açúcar seja da ordem de 8,8 milhões de hectares, devendo gerar uma produção de 727,8 milhões de toneladas de cana-de-açúcar (INFORMAÇÃO UNICA, 2007b).

O investimento, no setor, para o período de 2008 à 2012 será, em torno, de US\$ 33 bilhões (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2009b).

Segundo Rodrigues e Ortiz (2006), hoje o novo ciclo de expansão da indústria canavieira brasileira encontra-se fundamentado nas expectativas de crescimento, pelo grande e consolidado mercado interno, pelo mercado internacional, a expansão da produção representadas pelos motores bi-combustíveis, o mercado dos preços do petróleo, pelos compromissos de redução das emissões de CO₂ assumidos pelos países desenvolvidos, junto ao Protocolo de Quioto, e pela queda nos subsídios agrícolas para o açúcar.

Uma reestruturação produtiva e organizacional vem ocorrendo de forma signi-

ficativa. Algumas das medidas adotadas, por empresas do setor, têm sido implementadas através de sofisticados mecanismos de gestão e controle do processo produtivo e da força de trabalho, como a utilização da informática no controle do processo produtivo, o uso de técnicas modernas de gerenciamento, a utilização de novos equipamentos, a intensificação do corte mecanizado da cana-de-açúcar, e mesmo, a ampliação do processo de terceirização.

Portanto, tornam-se importantes estudos que apresentem propostas para a melhoria da qualidade e produtividade do setor, focado no desempenho dos processos e produtos. Isto pode ser viabilizado por programas de controle de qualidade nas estruturas organizacionais, com uma abordagem de sistemas para a melhoria da qualidade através de métodos estatísticos.

Este estudo teve por objetivo apresentar a filosofia “Lean Seis Sigma” e através da sua metodologia, utilizar os programas para a melhoria de produtos, serviços e processos, presentes em uma usina de açúcar e de álcool, com a finalidade da melhoria da capacidade produtiva pela redução da variabilidade e do tempo de ciclo do processo. Utilizando os métodos de seleção de variáveis na análise de regressão, buscou-se, também, identificar as principais variáveis para descrever a relação entre as variáveis de entrada e saída do processo produtivo do etanol hidratado.

2 REVISÃO BIBLIOGRÁFICA

2.1 Etanol Combustível

2.1.1 Histórico da cana-de-açúcar

A origem da cana-de-açúcar é muito discutida. O lugar onde a planta germinou, pela primeira vez, também é uma incógnita, devido à quantidade de gramíneas híbridas existentes e a falta de documentação que comprovem sua origem. Alguns pesquisadores aceitam que a cana-de-açúcar tenha surgido inicialmente na Polinésia, outros dizem que sua origem foi em Papua Nova Guiné. Uma linha de estudiosos, que admite o surgimento da cana há 6 mil anos, indica a Indonésia, Filipinas e o norte da África como expansão natural nos dois mil anos após o primeiro registro da planta. A maioria dos historiadores aceita que o surgimento da cana ocorreu há aproximadamente 10 e 12 mil anos. Mas um fato com o qual todos os historiadores concordam é que a cana-de-açúcar tem origem asiática, segundo a UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR (2007).

No Brasil, durante o período colonial, a economia desenvolveu-se por ciclos de acordo com os interesses do mercado internacional. No início o pau-brasil, seguido pelo açúcar da cana e, por fim, o do ouro. Cada um desses ciclos teve o seu momento de expansão e a sua fase de decadência.

No período do Primeiro Reinado (1822/1831) até o Segundo Reinado (1840/1889), o açúcar brasileiro enfrentou a concorrência do açúcar das Antilhas e também, do açúcar da beterraba que era produzido na Europa. As altas taxas impostas sobre as importações do produto, pelos governos, enfraqueceram o comércio.

A reorganização do setor agroindustrial da cana-de-açúcar no Brasil, no início dos anos 30 só foi possível pela forte ação do estado. Em 1933 com a criação do Instituto do Açúcar e do Alcool (IAA), o governo passou a controlar o setor, sendo responsável pelo ciclo de produção e comercialização, inclusive fixação de preços, cotas, exportação e importação. O governo considerava que a produção nacional de açúcar excedia as necessidades de consumo e havia uma tendência mundial de limitação da produção de açúcar. Desta forma era preciso assegurar o equilíbrio do mercado do açúcar e, a produção do etanol, seria uma alternativa viável para a indústria açucareira.

Em 1975, diante de uma forte crise mundial do petróleo, foi criado o Programa Nacional do Alcool (PROÁLCOOL), por decreto presidencial “visando o atendimento das necessidades do mercado interno e externo e da política de combustíveis automotivos”. O PROÁLCOOL tinha como objetivo o incentivo à produção do etanol oriundo da cana-de-açúcar, da mandioca ou de qualquer outro insumo. A principal motivação para o lançamento do programa foi o forte impacto negativo na economia nacional causado pelos dois “choques”

do petróleo, em 1973 e 1978. Vale ressaltar que, o Brasil acumulava experiência com o etanol combustível desde a década de 20.

O programa Nacional do Alcool foi regulamentado em duas fases distintas: em 1975 e em 1979. Em 1979 o Governo Federal reformulou alguns aspectos do PROÁLCOOL para estimular a produção do etanol hidratado, que viria a ser usado diretamente em motores de veículos desenvolvidos especialmente para este fim. Na segunda fase do programa, a produção de etanol hidratado foi maior do que a do anidro, passando de 395 milhões de litros na safra 1978/79 para 1,7 bilhões de litros na safra de 1991/92. As vendas anuais de carros movidos a etanol hidratado cresceram de 240.638 unidades em 1980 para 698.564 unidades vendidas em 1986, atingindo cerca de 96% dos veículos novos vendidos para o mercado doméstico.

Em 1990 o IAA foi extinto por uma Medida Provisória, já que, em meados de 1988, um Decreto-Lei determinava a proibição, a partir de 1º de junho de 1989, do uso de recursos do Tesouro Nacional em operações de compra e venda de açúcar para exportação e determinava ao Poder executivo que procedesse à reformulação da organização do setor sucroalcooleiro com a redução da dependência da agroindústria canavieira de recursos do Tesouro Nacional e da intervenção do Governo no setor. A saída do Estado do setor sucroalcooleiro ocorreu em ritmo acelerado, começando com a liberação das exportações de seus produtos, em 1989, passando por liberações dos preços do açúcar e do etanol anidro e sendo concluída com a liberação dos últimos preços que ainda eram mantidos sob controle oficial: os da cana, os do açúcar *standard* e os do etanol hidratado, no início de 1999. Atualmente, a produção e a comercialização da cana, do açúcar e do etanol não estão mais no controle do Estado. Os preços recebidos pelos produtores são definidos pelos mercados.

2.1.2 O plantio e safra da cana-de-açúcar no Brasil

A cana-de-açúcar é uma gramínea que possui um grande teor de açúcares (aproximadamente 1/3 da matéria seca dos colmos). A cultura da cana é semiperene, pois pode ser cortada e colhida, sem a necessidade de replantio, por 5 a 7 safras anuais consecutivas. Após esse período, pode ser feita uma alternância por outro tipo de cultura no terreno pelo período de uma safra, voltando ao plantio da cana novamente (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2007).

O Brasil é o maior produtor de cana-de-açúcar do mundo, seguido pela Índia, Tailândia e Austrália (UNICA, 2006 *apud* RODRIGUES; ORTIZ, 2006), possui uma área com cerca de 850 milhões de hectares, sendo que 60 milhões de hectares são destinados à agricultura. Na safra 2007/08, a área ocupada com a cultura da cana-de-açúcar atingiu 7,8 milhões de hectares. A expectativa é que a área cultivada com cana, até 2020, deva chegar a

14 milhões de hectares (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2008). A figura 1 mostra o mapa da produção de cana-de-açúcar no Brasil.

Figura 1 - Mapa da produção de cana-de-açúcar no Brasil

Fonte: <http://www.unica.com.br/content/show.asp?cntCode=D6C39D36-69BA-458D-A95C-815C87E4404D> (Acesso em 12 nov. 2008).

Ainda, segundo a UNIÃO DA INDÚSTRIA DE CANA-DE-AÇÚCAR (2007), no Centro-Sul do Brasil, região com maior produção nacional (aproximadamente 90%), a safra da cana dura, em média, oito meses, com início em abril e término em novembro. No Nordeste, a safra dura sete meses, com início em setembro e término em março do ano seguinte.

O recebimento da cana, pela indústria, deve ser feito com no máximo 72 horas após a colheita, evitando assim, perda da qualidade pela ação de bactérias e fermentos. A cana colhida há mais de 5 dias, geralmente, não é aceita pelas indústrias.

Como as demandas por açúcar e etanol são constantes durante os 12 meses do ano, deve-se produzir durante os 8 meses da safra a quantidade de etanol suficiente para o abastecimento do mercado durante todo ano.

2.1.3 Colheita da cana-de-açúcar - Legislação Ambiental

O Brasil é mundialmente reconhecido como líder na produção e eficiência do setor sucroalcooleiro, mas esta liderança não se reflete na mesma medida na responsabilidade

social, ambiental e na governança no setor (RODRIGUES; ORTIZ, 2006).

No contexto mundial, vem sendo verificado um progressivo pressionamento no que concerne à agricultura de um modo geral, no sentido de que esta modifique sua base técnica, incorporando-a a preocupação com o meio ambiente. Existem barreiras de importação não só para a agricultura, mas para os demais setores produtivos (BRAGA, 1995 *apud* OLIVEIRA, 2008).

Segundo Oliveira (2008), normas de comércio, como o Selo Verde, atuam cada vez mais, configurando-se como regra para facilitar ou impedir o acesso a determinados mercados, e funcionam como barreiras comerciais, não tarifárias, que acabam discriminando produtos, sobretudo os advindos de outros países, que estejam fora dos padrões ambientais predefinidos.

O Brasil, enquanto importante exportador de produtos agrícolas, não está imune a este tipo de exigência, as exportações brasileiras já se submetem às Normas de Certificação da Série ISO 9000, a qual pauta-se na normatização de processos e produtos lançados no mercado, exigindo a padronização de níveis de qualidade, bem como a aplicação das internacionais de Certificação Ambiental ISO 14000, também na lavoura de cana-de-açúcar.

O Decreto Nº 2.661, de 8 de julho de 1998 que regulamenta o parágrafo único do art. 27 da Lei nº 4.771/65 (código florestal), prevê a eliminação do fogo na colheita da cana, de forma gradativa, no prazo de 20 anos, não podendo a redução ser inferior a um quarto da área mecanizável de cada unidade agroindustrial, a cada período de 5 anos. Estabelece ainda que, a avaliação das conseqüências sócio-econômicas decorrentes da proibição do emprego do fogo e os ajustes necessários nas medidas impostas serão realizados pelos órgãos competentes, a cada período de cinco anos, contados a partir de 1998.

O Decreto Nº 2.661, Art 16, § 2º, prevê ainda, a revisão conceitual periódica, em virtude da “evolução tecnológica na colheita de cana-de-açúcar, oportunidade em que serão ponderados os efeitos sócio-econômicos decorrentes da incorporação de novas áreas ao processo de colheita mecanizada”.

Nas duas últimas décadas, as profundas transformações no mundo do trabalho ocorridas nos países capitalistas, as quais se refletem tanto nas formas de inserção na estrutura produtiva, como nas formas de representação sindical e política da classe trabalhadora, reflete-se no principal problema da aplicação desta legislação, segundo alega o setor, quanto à utilização de mão-de-obra rural, uma vez que elimina a realização das queimadas, impossibilitando a colheita manual da cana (OLIVEIRA, 2008).

Além disso, segundo São Paulo (2002), a mecanização ou colheita da cana verde só pode ser realizada em áreas com declives inferiores à 12%. Só no Estado de São Paulo, esta limitação torna inviável cerca de 45% da área plantada de cana, ou seja, dos 2,8 milhões de

hectares, aproximadamente um milhão estaria acima de 12% de declividade, sendo imprópria para a mecanização da colheita.

No Estado de São Paulo, o Decreto 45.689, de 22 de junho de 2001, determina: “A partir do ano 2001 não se efetuará a queima da palha da cana-de-açúcar em percentual correspondente a 25% (vinte e cinco por cento) das áreas mecanizáveis e 13,35% (treze inteiros e trinta e cinco centésimos por cento) das áreas não mecanizáveis”. Tal procedimento visa aumentar a responsabilidade social e a grandeza de um setor que gera 32% da renda agrícola e mais de 45% do trabalho disponível na área rural.

As queimadas, praticadas de forma indiscriminada na região canavieira paulista, que se estende por uma área equivalente ao território da Bélgica - cerca de 2,6 milhões de hectares de cana - liberam 2,6 quilos de hidrocarbonetos e de 30 a 41 quilos de monóxido de carbono por tonelada, muito embora a poluição visível esteja circunscrita a apenas 2,5 ou 3,5 quilos de fuligem (o chamado “carvãozinho”) por tonelada.

Medidas legais foram tomadas a fim de resolver o problema das queimadas indiscriminadas, e prevenir o quadro de doenças respiratórias inerentes à queima, cujas evidências científicas mostram que quarenta tipos de hidrocarbonetos policíclicos aromáticos, liberados por essa prática, são indutores de câncer, malformações congênitas e outros distúrbios genéticos. Atear fogo na cana representa, antes de tudo, mais um desperdício de energia valiosa e renovável e uma agressão à saúde, dos cerca de 20 milhões de seres humanos, que vivem em seu entorno.

A eliminação do fogo na colheita da cana é prevista no Decreto nº 48.869/01, em conjunto com a implementação de políticas de reciclagem e aproveitamento da mão-de-obra. O setor sucroalcooleiro, argumenta que a eliminação da queimada da cana vem forçando grande quantidade de trabalhadores que viviam da colheita da cana-de-açúcar a migrar para os centros urbanos ou para as frentes de ocupação de terras.

2.1.4 Definição de Etanol

O etanol é um tipo de álcool constituído por dois átomos de carbono, cinco átomos de hidrogênio e um grupo hidroxila, sendo denotado por C_2H_5OH . A diferença entre o etanol anidro (álcool etílico anidro) e o etanol hidratado (álcool etílico hidratado) aparece apenas no teor de água contida no etanol. Enquanto o etanol anidro tem o teor de água, em torno de 0,5%, em volume, o etanol hidratado, vendido nos postos de combustíveis, possui cerca de 5% de água, em volume. Na produção industrial do etanol, o tipo hidratado é o que sai diretamente das colunas de destilação. No caso do etanol anidro é preciso utilizar um processo adicional que retira a maior parte da água presente (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2007).

2.1.5 Os produtos obtidos da cana-de-açúcar

Entre os vários subprodutos da indústria canavieira, destacam-se o bagaço, as pontas e folhas e a vinhaça. O bagaço é usado como combustível nas usinas desde quando passou a substituir a lenha, que era a fonte de energia utilizada na evaporação do caldo (CORTEZ; MAGALHÃES; HARP, 1992 apud WALTER, 1993). Com a prática da queima da cana, antes da colheita, para facilitar o corte e controlar as pragas, perde-se uma grande quantidade de biomassa (palhas e folhas), além dos impactos ambientais associados às emissões. Payne, 1986 apud Walter (1993) diz que a queima da cana, antes da colheita, pode implicar em redução de até 50% da biomassa da planta, bem como reduções não desprezíveis no teor de sacarose.

A vinhaça é um resíduo orgânico obtido no processo de destilação do vinho, para a produção do etanol. Para cada litro de álcool produzido tem-se de 10 a 14 litros de vinhaça (WALTER, 1993). Por constituir-se em matéria rica em nutrientes, e em especial em potássio, a vinhaça é usada como fertilizante nas próprias lavouras de cana.

2.1.6 A matriz energética do Brasil

O etanol e o bagaço da cana-de-açúcar representam 16% da matriz energética brasileira, o que os coloca no 2º lugar do ranking das principais fontes energéticas do País, abaixo do petróleo e derivados e acima da hidroeletricidade (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2009b).

A potência já vendida ao governo federal supera os 4 mil MW, representando 4% da capacidade instalada de geração no país. Estima-se que em 2009, embora o cenário econômico internacional esteja sombrio, e atinja, também, o setor sucroalcooleiro, as perspectivas são de que 35 novas unidades entrem em operação (NOTÍCIA UNICA, 2008).

O uso do etanol combustível no Brasil vem crescendo continuamente, desde a introdução dos veículos flex-fuel, em 2003. Em 2008, o consumo de etanol representou mais de 50% do consumo de combustíveis, destinados a automóveis e veículos comerciais leves, movidos a etanol e/ou gasolina, de um total de 22,48 bilhões de m³ produzidos (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2009b).

O Brasil atingiu em 2006 a sua auto-suficiência em petróleo, com a produção de cerca de 1,8 a 1,9 milhões de barris/dia. Essa auto-suficiência conta com a expressiva contribuição do setor de cana-de-açúcar, que atingiu 14,6% da oferta interna de energia (UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR, 2007).

Atualmente, o panorama com respeito à energia elétrica, no Brasil, encontra-se estável, no entanto, o setor não descarta a possibilidade de uma crise energética próxima, pela expressiva demanda existente.

As 50 milhões de toneladas de bagaço, que sobram da moagem da safra paulista de cana, podem produzir energia, equivalente a mais de 70 milhões de barris de petróleo, por ano e constitui-se em uma razão a mais, para continuar incentivando o setor sucroalcooleiro a aproveitar melhor o seu potencial energético (GOLDEMBERG, 2002 apud SÃO PAULO, 2002).

Estudos indicam que a matéria seca deixada no campo (a palha e o ponteiro) com a colheita da cana verde, se levada para queimadas nas usinas, podem aumentar o poder calorífero da caldeira, em cerca de 20%. E que, sendo compactada para armazenamento e sua utilização na queima, junto ao bagaço, alimentaria a produção de energia, na época de entressafra de cana-de-açúcar (USTULIN; SEVERO, 2001).

Considera-se que, a existência de até 25 toneladas de palha por hectare, resultaria em um acréscimo de 65 milhões de toneladas de biomassa às usinas termoelétricas, transformando em eletricidade o que hoje é convertido em indesejável poluição atmosférica, entre maio e novembro, pelas queimadas efetuadas (SÃO PAULO, 2002).

2.1.7 Avaliação dos atributos tecnológicos

Fernandes (2003) descreve os atributos tecnológicos da cana-de-açúcar como sendo:

1. Brix do caldo

É o parâmetro mais utilizado na indústria do açúcar e do álcool. Expressa a porcentagem (peso/peso dos sólidos solúveis) contida em uma solução pura de sacarose, ou seja mede o teor de sacarose na solução, obtido através de aparelhos eletrônicos.

2. Pol do caldo

A Pol representa a porcentagem de oligossacarídeos, ou seja, a porcentagem aparente de sacarose contida numa solução de açúcares. A sacarose, um dissacarídeo, é o principal parâmetro de qualidade tecnológica da cana-de-açúcar e refere-se ao açúcar diretamente cristalizável no processo de fabricação.

3. Pureza do caldo

A pureza expressa qual a porcentagem de sólidos (Brix no caldo) é representada pela sacarose (pol no caldo). Esse atributo tecnológico está sendo utilizado como o mais importante indicador do estágio de maturação da cana.

4. Fibra da cana

A fibra é a matéria insolúvel em água contida na cana. No sistema de Pagamento da Cana pelo Teor de Sacarose (PCTS) a fibra é estimada em função do peso do bagaço úmido da prensa.

5. Pol da cana

A pol da cana, determinada por sacarimetria no caldo, pode ser expressa em porcentagem de cana, através do cálculo da porcentagem da fibra da cana. Devido a menor precisão da fibra, junto com a representatividade da amostragem, a diferença entre a pol na cana e a sacarose no caldo é insignificante em escala de rotina. A pol da cana indica toda a sacarose aparente (pol) contida no Caldo Absoluto por cento de Cana (CA%C).

6. Açúcares Totais Recuperáveis (ATR)

O ATR constitui um dos parâmetros do sistema de pagamento da cana implantado em São Paulo a partir da safra de 1998/99, e representa a quantidade de açúcares na forma de açúcares invertidos ou Açúcares Redutores Totais (ART) recuperados da cana até o xarope na usina [$kg t^{-1} cana$], ou seja, é o resultado da diferença entre ART da cana e as perdas quando se lava a cana (em torno de 12%), bagaço final, torta dos filtros ou prensas e “indeterminadas”, considerando eficiência média padrão.

7. Açúcares Redutores (AR)

Os açúcares redutores referem-se a um termo utilizado para designar os açúcares monossacarídeos, glicose e frutose. São precursores de cor no processo industrial, isto é, participam de reações que aumentam a cor do açúcar, depreciando a qualidade do produto.

2.2 Controle e Melhoria da Qualidade

Taylor, 1911 apud Montgomery (2004) introduziu alguns princípios de gerenciamento científico à medida em que as indústrias começaram a se desenvolver, antes de 1900. Taylor foi o pioneiro na divisão do trabalho em tarefas, de modo que o produto pudesse ser manufaturado e montado mais facilmente. Seu trabalho levou a melhorias substanciais na produtividade.

Segundo Montgomery (2004) os métodos estatísticos e sua aplicação na melhoria da qualidade têm uma longa história. Em 1924, Walter A. Shewhart desenvolveu o conceito estatístico de gráfico de controle, que é considerado como o começo formal do controle estatístico da qualidade. No final da década de 20, Harold F. Dodge e Harry G. Romig,

ambos dos Bell Telephone Laboratories, desenvolveram a amostragem de aceitação, com base estatística, como uma alternativa a 100% de inspeção. No meio da década de 30, os métodos estatísticos de controle da qualidade eram largamente usados na Western Electric, a divisão de manufatura da Bell. Durante a Segunda Guerra Mundial houve uma grande expansão do uso e aceitação dos conceitos de controle estatístico da qualidade nas indústrias de manufatura. Em 1946 foi fundada a American Society for Quality Control (Sociedade Americana para o Controle da Qualidade), com a finalidade de promover o uso das técnicas de melhoria da qualidade para todos tipos de produtos e serviços, bem como oferecer conferências, publicações técnicas e programas de treinamento para garantir a qualidade.

Nos anos 50, foram introduzidos, nos Estados Unidos, os planejamentos de experimentos para produtos e processos. Esses métodos foram amplamente explorados pela indústria química e são geralmente citados como uma das principais razões pelas quais a indústria química americana é uma das mais competitivas do mundo. A expansão desses métodos para fora da indústria química foi relativamente lenta até o final da década de 70 e início da década de 80, quando muitas companhias ocidentais descobriram que seus competidores japoneses vinham usando sistematicamente o planejamento de experimentos, desde os anos 60, para avaliação de problemas do processo, desenvolvimento de novos processos, avaliação de projetos de novos produtos, melhoria da confiabilidade e desempenho dos produtos e muitos outros aspectos do planejamento de produtos.

A partir de 1980 ocorreu um grande crescimento no uso de métodos estatísticos, para a melhoria qualidade, nos Estados Unidos, pois ocorreram grandes perdas de negócios e mercados de várias companhias domésticas. A adoção e o uso de métodos estatísticos desempenharam papel fundamental no ressurgimento da indústria americana.

2.2.1 Filosofia da Qualidade

Montgomery (2004) afirma que muitas pessoas contribuíram para a metodologia estatística da melhoria da qualidade, mas, que em termos de implementação e filosofia de gerenciamento, três se destacam: W.E. Deming, J.M.Juran e A.V. Feigenbaun. Segundo o autor, o trabalho do Dr. W. Edwards Deming foi um dos motivos pelos quais os fabricantes japoneses mostraram grande capacidade no controle estatístico e melhoria da qualidade, nas décadas de 70 e 80. Durante a Segunda Guerra Mundial, Deming trabalhou para o Departamento de Guerra e para o Bureau do Censo, sendo que, após a guerra tornou-se consultor das indústrias japonesas, e convenceu seus diretores do poder dos métodos estatísticos e da importância da qualidade como arma de competitividade.

A filosofia de Deming é um importante sistema para a implementação da melhoria da qualidade e produtividade. Essa filosofia, conhecida mundialmente, é resumida nos

chamados 14 pontos de Deming.

O Dr. Joseph M. Juran é um dos criadores do controle estatístico da qualidade. Trabalhou para o Walter A. Shewhart no AT&T Bell Laboratories e tem estado, desde então, na liderança da melhoria da qualidade. É coautor do *Quality Control Handbook* (Manual de Controle de Qualidade), uma referência para os métodos e melhoria da qualidade, desde o seu lançamento, em 1957. A filosofia Juran se baseia na organização para mudança e implementação da melhoria através do que ele denomina “avanço gerencial”.

O Dr. Armand V. Feigenbaum foi o primeiro a introduzir o conceito de controle da qualidade por toda a companhia, em seu livro *Total Quality Control* (Controle de Qualidade Total). Esse livro influenciou muitos dos princípios da filosofia da gerência de qualidade no Japão, no início dos anos 50. Na verdade, muitas companhias japonesas usaram o nome “controle da qualidade total” para descrever seus esforços. Feigenbaum se preocupa mais com a estrutura organizacional e com uma abordagem de sistemas para a melhoria da qualidade do que com métodos estatísticos.

2.2.2 Desdobramento da função qualidade (Quality Function Deployment - QFD)

O desdobramento da função qualidade, como um procedimento para delinearmento, foi introduzido no Japão em 1966. O QFD fornece métodos específicos para assegurar a qualidade, em cada estágio do desenvolvimento do produto, no processo e é definido como um método que converte as demandas dos clientes em características de qualidade e desenvolve um projeto de qualidade para o produto final (AKAO, 1990). Segundo o autor, QFD é um termo geral que significa “desdobramento da qualidade através do desenvolvimento das funções de qualidade”.

Milan et al (2003) ressaltam que o ponto de maior importância no QFD é o gráfico chamado “casa de qualidade” em que, as demandas dos clientes (qualidade exigida) é traduzida em características do projeto (requisitos técnicos), baseados em pesquisas de mercado e experiências passadas, sendo que, uma matriz de relações identifica quão intensa é a influência de um determinado requisito técnico sobre um determinado item da qualidade exigida.

Akao, 1996 apud Nagumo (2005) diz que o desdobramento da função qualidade é um método que traduz as necessidades e desejos dos clientes em especificações técnicas de produtos e processos, de forma que tais especificações possam ser cumpridas pelas áreas operacionais. Segundo o autor, a partir de 1983, com a publicação de um artigo, sobre a aplicação do QFD nas indústrias japonesas, na revista *Quality Progress*, o método foi difundido nos Estados Unidos e Europa.

2.2.3 A Filosofia Lean

Shingo (1996) define o termo Lean como redução no tempo do processo, sendo que processo é a transformação da matéria prima em produto. Segundo o autor, os elementos de um processo são:

- i) Processamento: é a mudança física no material ou na sua qualidade (montagem ou desmontagem);
- ii) Inspeção: é a comparação com um padrão estabelecido;
- iii) Transporte: é o movimento de materiais ou produtos;
- iv) Espera: é o período de tempo durante o qual não ocorre nenhum processamento, inspeção ou transporte.

O Sistema Toyota de Produção, também chamado de *Lean Production* (Produção enxuta) e *Lean Manufacturing* (Manufatura enxuta), surgiu no Japão, na fábrica de automóveis Toyota, logo após a Segunda Guerra Mundial. Nessa época, a indústria japonesa tinha uma produtividade muito baixa e uma enorme falta de recursos, o que naturalmente a impedia adotar o modelo da produção em massa. A ferramenta *Lean* pode ser aplicada em toda a cadeia de valor, desde o recebimento dos pedidos de vendas até a entrega ao cliente, incluindo relações com os clientes, engenharia, fornecedores, administração e manufatura. Assim, a denominação *Lean Enterprise* (Empresa enxuta) também é utilizada.

Segundo Juran e Godfrey (1998) diversos fatores do Sistema Toyota de Produção culminaram com a criação da ferramenta conhecida como *Lean Production*, um conceito novo, e muito usado pelos fabricantes de automóveis americanos. *Lean Production* envolve a redução do desperdício de cada passo na cadeia de produção como desperdício de energia, de deslocamento, tempo e recursos. É baseado no método de recuperação de controle de estoque que Taiichi Ohno, da Toyota, projetou em 1950, agora conhecido como *just-in-time*.

Shingo (1996) define o Sistema Toyota de Produção como “um sistema que visa à eliminação total das perdas”. Esta ferramenta alcançou grandes resultados após repetidas tentativas e erros. O princípio básico do Sistema Toyota de Produção é: **o princípio do não-custo**. Este foi o primeiro conceito desenvolvido com base para o gerenciamento da produção, ou seja é o princípio da minimização dos custos. Segundo este conceito, a origem dos lucros não é obtida pela fórmula

$$\text{Custo} + \text{Lucro} \Rightarrow \text{Preço de venda}$$

e sim por

$$\text{Preço} - \text{Custo} \Rightarrow \text{Lucro}.$$

Fazendo uso deste conceito, conclui-se que a única maneira de aumentar os lucros é pela diminuição dos custos, uma vez que o preço de venda é definido pelo mercado.

Rother e Shook (2003) definem fluxo de valor como toda ação (agregando valor ou não), necessária para trazer um produto por todos os fluxos essenciais no processo de produção do mesmo:

1. O fluxo de produção desde a matéria-prima até a entrega ao consumidor;
2. O fluxo do projeto do produto, da concepção até o lançamento.

Segundo os autores, o mapeamento do fluxo de valor é uma ferramenta que utiliza papel e lápis e ajuda a enxergar e entender o fluxo de material e de informação, na medida em que o produto segue o fluxo de valor. Esta ferramenta ajuda a identificar as fontes de desperdício no fluxo de valor. O objetivo de mapear o fluxo de valor é mostrar as fontes de desperdício e eliminá-las, com a implementação de um fluxo de valor em um estado futuro. Isto pode ser obtido com um fluxo de valor enxuto, ou seja, construir um processo para fazer somente o que o próximo processo necessita e quando necessita.

George (2004) diz que a meta da ferramenta *Lean* é acelerar a velocidade de qualquer processo através da redução do desperdício em todas as suas formas. Segundo o autor, com o uso dos conceitos e das ferramentas Lean, pode-se verificar que os passos do processo, que até então, se pensava serem essenciais, são, na verdade, desnecessários e seus custos e atrasos são eliminados, após a aplicação da ferramenta Lean.

Segundo Womack e Jones (2004), um antídoto ao desperdício é o “Pensamento Enxuto” (*Lean Thinking*), que é uma forma de especificar valor, alinhar na melhor sequência as ações que criam valor, realizar essas atividades sem interrupção toda vez que alguém as solicita e realizá-las de modo cada vez mais eficaz. As ferramentas do “Pensamento Enxuto” são:

- i) Mapeamento do Fluxo de Valor (*Value Stream Mapping - VSM*);

O fluxo de valor é o conjunto das atividades (que agregam e não agregam valor) que são realizadas pela empresa para projetar, produzir e entregar seus produtos aos clientes.

- ii) Métricas *Lean*;

São medidas usadas para quantificar os resultados de uma empresa em relação a velocidade e eficiência.

- iii) *Kaizen*;

É um termo japonês que significa melhoramento contínuo. Segundo George (2004) *Kaizen* pode ser utilizado quando:

- a) Fontes de desperdício óbvias foram identificadas;
- b) O escopo de um problema está claramente definido e compreendido;
- c) O risco de implementação é mínimo;
- d) Os resultados são necessários imediatamente;
- e) É desejável aumentar a velocidade e adquirir credibilidade nas fases iniciais de um projeto de melhoria.

iv) *Kanban*;

É um dispositivo sinalizador que autoriza e fornece instruções para a produção ou para a retirada de itens em um sistema puxado. Na maioria das vezes o uso é feito por cartões de papelão contendo informações como nome e número da peça, fornecedor, local de armazenamento e local do processo de consumo. O *Kanban* é usado para controlar um sistema puxado, ou seja, um produto é fabricado ou um item é retirado somente quando um cartão *Kanban* determinar. Os principais tipos de *Kanban* são:

- a) *Kanban* de produção;
- b) *Kanban* de sinalização;
- c) *Kanban* de retirada.

v) Padronização;

É o método usado para indicar os procedimentos para execução das tarefas de um processo, de forma que os resultados desejados possam ser alcançados e mantidos.

vi) 5S;

É um método que tem como objetivo promover e manter a limpeza e a organização das áreas de trabalho (administrativa e de manufatura). Esta ferramenta funciona como um pilar básico do *Lean Manufacturing*. Os benefícios que o 5S apresenta para as empresas são:

- a) Aumento da produtividade;
- b) Melhor atendimento aos prazos;
- c) Redução de defeitos;
- d) Aumento da segurança no trabalho;
- e) Redução de material perdido;
- f) Melhor capacidade para distinção entre condições normais e anormais de trabalho.

O termo 5S, derivado de 5 palavras japonesas que começam com a letra S, é apresentado na tabela 1.

Tabela 1 - Significado do 5S

Palavra	Tradução	Significado
Seiri	Senso de Utilização (Sort - Classificar)	Separar o necessário do desnecessário, descartando o último.
Seiton	Senso de Organização (Set in Order - Ordenar)	Organizar o necessário, definindo um lugar para cada item.
Seiso	Senso de Limpeza (Shine - Limpar)	Limpar e identificar cada item.
Seiketsu	Senso de Padronização (Standardize - Padronizar)	Criar e seguir um padrão resultante do desempenho adequado nos três primeiros S's.
Shitsuke	Senso de Autodisciplina (Sustain - Manter)	Estabelecer a disciplina para manter os quatro primeiros S's ao longo do tempo.

Fonte: Werkema (2006).

vii) Redução de *Setup*;

É um método para a diminuição do tempo necessário para a troca da fabricação de um tipo de produto para outro. Foi criado, nas décadas de 1950 e 1960, por Shigeo Shingo (consultor da Toyota). Segundo Shingo (1996), a adoção da troca rápida de ferramentas (TRF) ou a troca de ferramentas em um único toque (OTED) é a melhor forma de melhorar o *setup*. Existem dois tipos de *setup*.

- (a) *Setup* interno: são operações de *setup* que podem ser realizadas somente com o equipamento parado;
- (b) *Setup* externo: são operações de *setup* que devem ser executados enquanto o equipamento está em atividade.

O método tem como principal recurso a conversão do *setup* interno em *setup* externo. Nas operações de *setup* é interessante estar bem claro o trabalho que pode ser realizado com a máquina em funcionamento e aquele que deve ser feito com a máquina desligada. O tempo de *setup*, geralmente, apresenta quatro funções:

- a) Preparação da matéria-prima, dispositivos de montagem, acessórios, etc. (30%);
- b) Fixação e remoção de matrizes e ferramentas (5%);
- c) Centragem e determinação das dimensões das ferramentas (15%);
- d) Processamentos iniciais e ajustes (50%).

viii) TPM (*Total Productive Maintenance*);

É um conjunto de procedimentos que têm como objetivo garantir que os equipamentos de um processo produtivo sejam sempre capazes de executar as tarefas necessárias para que a produção não seja interrompida. O método foi criado no Japão, entre 1960 e 1970, por uma empresa fornecedora da Toyota. No início, foi estruturado por 5 pilares, que são:

- (a) Eficiência;
- (b) Auto-reparo;
- (c) Planejamento;
- (d) Treinamento;
- (e) Ciclo de vida.

Mais tarde, surgiram outros 3 pilares: qualidade, gerenciamento e segurança, higiene e meio ambiente. Devido a sua abrangência, algumas empresas adotam o *TPM* como base para a estruturação de seu sistema de gestão. O *TPM* busca reduzir o tempo perdido de produção, devido a paradas planejadas ou não, dos equipamentos.

ix) Gestão visual;

É a colocação em local fácil de ver de todas as ferramentas, peças, atividades de produção e indicadores de desempenho do sistema de produção, com a finalidade de que a situação do sistema possa ser entendida rapidamente por todos os envolvidos. Os pontos chaves para a implantação da gestão visual são o 5S e a padronização.

As principais ferramentas utilizadas na gestão visual são:

- a) Quadro *takt*: mostra o cronograma de trabalho desejado, as diferenças para o trabalho realizado e os motivos para essas diferenças.

- b) *Andon*: mostra o *status* das operações em uma área da empresa, indicando qualquer tipo de anormalidade.
- c) Quadro de treinamento dos operadores: relaciona os funcionários de uma área da empresa e mostra a situação de cada um deles em relação aos treinamentos considerados.
- d) Gráfico sequencial para um item de controle: é um diagrama que mostra os valores individuais do resultado de um processo ao longo do tempo.

x) *Poka-yoke (Mistake proofing)*

Um dispositivo *Poka-yoke* é qualquer mecanismo que evite que o erro seja cometido ou seja facilmente detectado e corrigido. As causas mais comuns de erros, tanto em processos de manufatura quanto administrativos, são: esquecimento, falta de atenção, treinamento inadequado, falta de treinamento, falta de padronização e não obediência aos padrões. Existem 2 tipos de dispositivos *Poka-yoke*:

1. *Poka-yoke* de prevenção;
2. *Poka-yoke* de detecção.

No dispositivo *Poka-yoke* de detecção existem dois modos de corrigir erros:

2.1 *Poka-yoke* de controle - emprega dispositivos que interrompem o processo até que a causa do problema seja corrigida;

Existem três tipos de *Poka-yoke* de controle:

- i. Método de contato - identifica os defeitos em virtude da existência ou não de contato entre o dispositivo e alguma característica ligada à forma ou dimensão do produto;
- ii. Método de conjunto - determina se o número de atividades previstas são executadas;
- iii. Método das etapas - determina se são seguidos os estágios ou operações estabelecidos por um dado procedimento.

2.2 *Poka-yoke* de advertência - emite um sinal ou acende uma luz, quando um erro acontece, não interrompendo o processo.

Werkema (2006) define “Lean Manufacturing” como uma iniciativa que busca eliminar desperdícios, de forma que se possa excluir o que não tem valor para o cliente, determinando assim, uma aceleração no processo. A ferramenta Lean pode ser aplicada em qualquer tipo de trabalho. Segundo a autora, “Lean Manufacturing” teve origem com o Sistema Toyota de Produção por representar uma forma de produzir cada vez mais com cada

vez menos, sendo denominado produção enxuta (Lean Production ou Lean Manufacturing). A meta do “Lean Manufacturing” é a redução dos sete tipos de desperdícios identificados por Taiichi Ohno:

1. Defeitos (nos produtos);
2. Excesso de produção (mercadorias desnecessárias);
3. Estoques (mercadorias à espera de processamento ou consumo);
4. Processamento (desnecessário);
5. Movimento desnecessário (de pessoas);
6. Transporte desnecessário (de mercadorias);
7. Espera (dos funcionários pelo equipamento que está sendo processado ou por uma atividade anterior).

2.2.4 Seis Sigma

O conceito Seis Sigma foi introduzido em 1980 pela Motorola com o intuito de reduzir os defeitos em produtos manufaturados em poucas partes por milhão. Mais tarde, estendeu o conceito de seis sigma para produtos e serviços.

Como outras companhias têm usado este conceito e adicionado algumas variações, o sistema seis sigma se tornou “um programa com o objetivo da quase eliminação de defeitos de cada processo e negócio” (TOMKINS, 1977 apud JURAN; GODFREY, 1998). O método seis sigma se tornou um procedimento quantitativo para melhorar operações em todos os tipos de indústrias e negócios. A iniciativa seis sigma pode parecer muito familiar para muitas companhias com sistema de gerência de qualidade total bem sucedido, mas podem parecer novidades para aquelas que tenham usado qualidade de forma superficial.

As duas metodologias, de soluções de problemas e gerenciamento, mais utilizadas nos programas seis sigma são o PDCA e o DMAIC.

O PDCA {*Plan* (planejar), *Do* (executar), *Check* (verificar), *Action* (agir)} é um método mundialmente conhecido e utilizado para manutenção, melhoria e inovação de produtos, serviços e processos.

O DMAIC {*Define* (definir), *Measure* (medir), *Analyze* (analisar), *Improve* (melhorar), *Control* (controlar)} é uma metodologia de solução de problemas utilizada pela maioria das empresas americanas que empregam o Programa Seis Sigma com o objetivo de realizar melhorias em produtos, serviços e processos.

A implementação tem que ser feita a partir dos mais altos níveis da companhia, ou seja, de cima para baixo, atingindo todos os níveis da gerência (HOERL, 1998 apud JURAN; GODFREY, 1998).

O foco principal do conceito seis sigma, assim como muitas outras atividades de qualidade, é a redução do custo, a melhoria da produção, da capacidade e reduções do tempo do ciclo de produção. A ênfase está na necessidade de satisfação do cliente.

Hoerl (2001) diz que a iniciativa de aperfeiçoamento Seis Sigma tem se tornado popular nos últimos anos, gerando muitas discussões dentro da estatística e nos círculos de qualidade. Neste artigo, o autor trata do conjunto de habilidades específicas que o “Seis Sigma Black Belts” necessita e como desenvolver esta técnica. Justifica que muitos autores usam termos, tais como “Black Belt”, “Master Black Belt”, com pouca ou nenhuma definição operacional do que estas pessoas fazem ou que habilidades elas têm. O autor diz, ainda, que recentemente houve discussão sobre as habilidades da “Black Belts” ou “Master Black Belts” em comparação com aquelas de um certificado do Engenheiro de Qualidade. Além disso, discute o trabalho que um “Seis Sigma Black Belt” faz, e que habilidades específicas são requeridas para fazer este trabalho.

Segundo Pande; Neuman e Cavanagh (2001), Seis Sigma é um sistema abrangente e flexível para alcançar, sustentar e maximizar o sucesso empresarial. Além disso, Seis Sigma é impulsionado por uma estreita compreensão das necessidades dos clientes, pelo uso disciplinado de fatos, dados e análise estatística e a atenção diligente à gestão, melhoria e reinvenção dos processos de negócios. O sucesso empresarial pode ser obtido porque os benefícios do Sistema Seis Sigma são diversos, entre eles:

- i) Redução de custos.
- ii) Melhoria de produtividade.
- iii) Crescimento de fatia de mercado.
- iv) Retenção de clientes.
- v) Redução de tempo de ciclo.
- vi) Redução de defeitos.
- vii) Mudança cultural.
- viii) Desenvolvimento de produto/serviço.

Segundo os autores, os níveis sigma de desempenho são expressos, geralmente, em partes por milhão (ppm). A tabela 2 mostra esta relação.

Tabela 2 - Tabela de conversão em sigma

Rendimento (%)	ppm	Nível Sigma
30,9	690.000	1,0
69,2	308.000	2,0
93,3	66.800	3,0
99,4	6.210	4,0
99,98	233	5,0
99,9997	3,4	6,0

Fonte: Pande; Neuman; Cavanagh (2001).

Nota: distribuição normal com média deslocada de $1,5\sigma$.

Pyzdek (2003) afirma que o emprego do Programa Seis Sigma não traz benefícios somente aos clientes. Quando o ciclo do projeto do produto diminui e as operações tem seu custo efetivo reduzido, os proprietários e os investidores, também, são beneficiados. Da mesma forma, quando os empregados tornam-se mais produtivos passam a ganhar mais, ou seja, todos são beneficiados na empresa. Segundo o autor, uma das contribuições mais significativas da Motorola foi mudar a discussão dos níveis de qualidade, que eram medidos em porcentagem (partes por cem), para uma discussão de partes por milhão, ou até mesmo, partes por bilhão. A Motorola corretamente apontou que com a complexa tecnologia moderna, idéias antigas sobre “níveis de qualidade aceitáveis”, não podiam mais ser tolerados. A produção industrial moderna exige níveis de qualidade quase perfeitos.

Segundo Montgomery (2004), a Motorola desenvolveu seu programa Seis Sigma, na década de 80, com a finalidade de reduzir a variabilidade dos principais componentes do seu produto, pois, produtos de alta tecnologia com muitos componentes complexos têm muita possibilidade de defeitos. O programa Seis Sigma reduz a variabilidade do produto de tal forma que falhas e defeitos são extremamente improváveis. O autor relata que o sistema padrão para a qualidade normalmente especifica uma tolerância para a variabilidade de três desvios padrão de cada lado da média. Nessa situação, a probabilidade de se produzir um produto dentro dessas especificações é de 0,9973, o que corresponde a 2700 partes por milhão (ppm) de defeituosos. Isto é conhecido como desempenho de qualidade três-sigma, e parece muito bom. No entanto, suponha que um produto consiste em um conjunto de 100 componentes e que todos componentes devem ser não defeituosos para que o produto possa

funcionar. Então, a probabilidade de uma unidade específica do produto ser não defeituosa é

$$0,9973 \times 0,9973 \times \dots \times 0,9973 = (0,9973)^{100} = 0,7631.$$

Assim, cerca de 23,7% dos produtos produzidos com a qualidade três-sigma serão defeituosos e isto não é aceitável, porque muitos dos produtos de alta tecnologia são feitos de milhares de componentes.

O conceito Seis Sigma da Motorola tem por objetivo reduzir a variabilidade no processo de modo que os limites de especificação estejam a seis desvios padrão da média. Sob a qualidade seis sigma, a probabilidade de que uma unidade específica do produto estar dentro das especificações é de 0,9999998, ou 0,2 ppm, o que representa uma situação bem melhor. Ainda, segundo o autor, quando o conceito Seis Sigma foi desenvolvido, foi feita a suposição de que quando o processo alcançasse o nível de qualidade seis sigma, a média do processo estaria, ainda, sujeita a perturbações que poderiam fazer com que ela mudasse em até 1,5 desvio padrão para longe do alvo. Nesse caso, o processo seis sigma produziria cerca de 3,4 ppm de defeituosos. Atualmente, o programa Seis Sigma está presente em muitas companhias e passou a abranger muito mais. Tornou-se um programa para melhorar o desempenho dos negócios da empresa, tanto pela melhoria da qualidade, quanto pela atenção à redução de custos.

George (2004) diz que um dos conceitos mais simples, mas mais poderoso do Seis Sigma, é que os resultados de qualquer processo advêm daquilo que entra no processo. Em textos sobre Seis Sigma, essa idéia aparece na simples equação de que “Y é função de X” que relaciona uma saída (Y) a entradas ou variáveis do processo (X):

$$Y = f(X_1, X_2, X_3, \dots).$$

Segundo o autor, esta equação é verdadeira, também, no nível organizacional, ou seja, qualquer saída (Y) como lucro ou crescimento é dependente das variáveis do processo (X) como qualidade, prazo, etc. e nele entram. Para melhorar os resultados é preciso encontrar os X críticos, que afetam o Y e direcionar esforços nesses valores. Ainda segundo o autor, há um significado mais profundo nessa equação: “não é apenas que Y é função de alguns X, mas que nossa tarefa é descobrir os X que realmente direcionam o Y”.

Scatolin (2005) desenvolveu um trabalho com o objetivo de mostrar a eficiência da aplicação da metodologia Seis Sigma para a redução das perdas em um processo de manufatura, através da apresentação de um estudo de caso. Seguindo a metodologia Seis Sigma, uma equipe foi formada para reduzir as perdas e quebrar o paradigma que não se pode eliminar perdas inerentes de um processo. Num prazo de seis meses, buscou-se a redução em 90% das perdas de fibras gerada pelo processo de manufatura dos respiradores descartáveis.

A meta foi atingida, sem interferir significativamente nas características finais do produto. O nível sigma que era menor que 3, passou para maior que 6 ao final do trabalho.

Rios (2006) propõe um modelo de referência para a eliminação de anomalias nos processos de manufatura. O modelo é baseado no ciclo DMAIC e fortemente apoiado nas tarefas da rotina do dia-a-dia, utilizada para estabilizar o sistema produtivo, medindo o desempenho da operação, através das métricas do Seis Sigma, como fonte de oportunidade de melhoria. O modelo foi aplicado através de um estudo de caso, em uma empresa de manufatura, obtendo como resultado principal a redução dos defeitos nos produtos finais e um incremento no rendimento interno final.

Allen (2006) diz que, em meados da década de 80, a Motorola constituiu uma companhia de consultoria e treinamento, chamada “Academia Seis Sigma”, e que diversas companhias adotaram estes métodos de treinamentos, entre elas, a Texas Instrumentos e ABB. Neste período, metodologias para estrutura do produto e melhoria do processo, como Desenvolvimento da Qualidade Total, Métodos Taguchi, e a então, chamada “Metodologia Seis Sigma”, foram publicadas. Em 1995, a General Electric (GE) fez contrato com a “Academia Six Sigma” para ajudá-la a melhorar seu programa de treinamento. Isto foi de suma importância para a popularização do Seis Sigma, pois a GE é uma das mais importantes companhias do mundo. Em 1999, a GE começou a competir com a Motorola oferecendo treinamento Seis Sigma para fornecedores e outros. Montgomery (2001); Hahn et al. (1999) apud Allen (2006), relataram que o treinamento Seis Sigma tornou-se mais popular que outro treinamento, em parte, por seu laço com as técnicas estatísticas. Dentre elas, destacam-se: Controle Estatístico do Processo, Delineamento de Experimentos, Análises de Efeitos e Modos de Falhas e Matriz de Causa e Efeito.

2.2.5 Análise de regressão

Montgomery e Peck (1992) definem a análise de regressão como uma técnica estatística utilizada para investigar e modelar a relação entre variáveis. As aplicações de regressão são muitas e ocorrem em quase todas as áreas.

Um modelo de regressão linear simples é definido por:

$$Y = \beta_0 + \beta_1 X + \varepsilon, \text{ em que:}$$

Y é a variável dependente (variável resposta), X é a variável independente (variável regressora), β_0 e β_1 são os parâmetros do modelo e ε é o erro aleatório.

Um modelo de regressão com mais de uma variável regressora é chamado de modelo de regressão linear múltipla, sendo definido por:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

Os parâmetros $\beta_0, \beta_1, \dots, \beta_k$ são chamados de coeficientes da regressão. Um dos objetivos da análise de regressão é estimar os parâmetros do modelo. Existem várias técnicas para a estimação dos parâmetros. Uma destas técnicas é chamada de método dos quadrados mínimos, que tem por objetivo minimizar o erro aleatório.

O modelo é chamado de linear para indicar que o modelo é linear nos parâmetros $\beta_0, \beta_1, \dots, \beta_k$.

A técnica formal para testar a significância da regressão é a análise da variância da regressão. Esta técnica tem por base o particionamento da variabilidade total na variável resposta Y em: soma de quadrados total = soma de quadrados da regressão + soma de quadrados do resíduo.

Draper e Smith (1998) dizem que uma estatística utilizada para avaliar a proporção da variação total da variável Y , explicada pela regressão, é o coeficiente de correlação múltiplo (R^2), expresso por:

$$R^2 = \frac{SQReg}{SQTotal} = 1 - \frac{SQRes}{SQTotal}, \quad 0 \leq R^2 \leq 1.$$

Os autores apresentam alguns métodos estatísticos para a seleção de variáveis no modelo. Esses métodos têm como objetivo “selecionar a melhor equação de regressão”. Os métodos são:

1. Todas regressões possíveis: os critérios utilizados para a seleção de variáveis são: o valor de R^2 , o valor do s^2 (quadrado médio do resíduo) e da estatística C_p de Mallows.
2. Método do passo atrás *backward*;
3. Método do passo a frente *forward*;
4. Método do passo a frente passo atrás *stepwise*.

Montgomery; Peck e Vining (2006) definem três tipos de resíduos:

1. Resíduos ordinários;
2. Resíduos estudantizados internamente;
3. Resíduos estudantizados externamente.

Os resíduos ordinários não são independentes. Assim seu uso não é o mais indicado para a análise dos resíduos. A estimativa da variância residual para os resíduos estudantizados internamente considera todas as observações. Portanto, essa estimativa pode variar, consideravelmente, quando existem valores discrepantes.

Os autores definem, também, três medidas de influência:

1. Distância de Cook: mede a influência da i -ésima observação se a mesma é retirada da amostra;
2. *DFBETAS*: mede a alteração no vetor $\hat{\theta}$ quando a i -ésima observação é retirada da análise;
3. *DFFITS*: mede a alteração no valor ajustado quando a i -ésima observação é retirada da análise.

3 MATERIAL E MÉTODOS

3.1 Material

Os dados analisados foram obtidos junto à Destilaria de Açúcar e Álcool Caiua S/A (DECASA), localizada na Fazenda Jaguatirica S/Nº, município de Caiua, estado de São Paulo. A DECASA produz açúcar e etanol hidratado. Neste trabalho, foi considerado como objeto de estudo, o processo produtivo do etanol hidratado. Os dados são referentes à safra 2008/09.

3.1.1 Processos básicos de produção

O processo produtivo do álcool hidratado é constituído por 4 processos básicos de produção: extração do caldo, tratamento do caldo, fermentação e destilação.

3.1.1.1 Extração do caldo

A extração do caldo tem início com o preparo da cana, com o objetivo de aumentar sua densidade e a capacidade de moagem. Este processo tem, também, o objetivo de romper as células para a liberação do caldo nelas contido. Assim, a cana passa por um jogo de facas. Após ser picada, a cana segue para um desfibrador, onde cerca de 85 a 92% de suas células são rompidas, o que facilita a extração do caldo. A cana é constituída, basicamente, de fibras e caldo, sendo que, o objetivo principal é extrair o máximo de caldo da cana. A extração do caldo, na usina Decasa, é realizada por moendas constituídas de quatro ternos, sendo que no primeiro terno são extraídas as maiores quantidades de caldo. Depois disso, o bagaço é embebido de água, e passa pelos demais ternos, onde cerca de 96% do caldo é extraído. O caldo extraído da cana-de-açúcar é direcionado a um tanque, de onde é distribuído para a etapa seguinte do processo e o bagaço é direcionado para as caldeiras, onde será queimado e o vapor transformado em energia. Cada tonelada de cana processada gera, em média, 260 kg de bagaço.

As variáveis consideradas nesta etapa foram:

Y_{i1} : caldo misto extraído da moenda ($i = 1, \dots, 81$);

X_{i1} : cana inteira que entra na moenda ($i = 1, \dots, 81$);

X_{i2} : cana picada que entra na moenda ($i = 1, \dots, 81$);

X_{i3} : índice de preparo da cana ($i = 1, \dots, 81$);

X_{i4} : pureza da cana ($i = 1, \dots, 81$);

X_{i5} : fibra da cana ($i = 1, \dots, 81$);

X_{i6} : pol do caldo da cana desfibrada ($i = 1, \dots, 81$);

X_{i7} : volume da água de embebição que entra na moenda ($i = 1, \dots, 81$).

3.1.1.2 Tratamento do caldo

Após a extração, o caldo da cana passa por um tratamento, onde é pré-aquecido, em seguida é feita a adição de leite de cal ($Ca[OH]_2$), para controle do pH, um novo aquecimento e posterior decantação. Após sair do decantador, é feita a separação do caldo que vai para a fábrica de açúcar e para a produção do álcool. Em ambos os casos o caldo passa por peneiras, com a finalidade de extração de partículas sólidas. O caldo que segue para a produção de álcool é resfriado antes de seguir para a fermentação, próxima etapa do processo.

As variáveis consideradas nesta etapa foram:

- Y_{i1} : brix do caldo caleado ($i = 1, \dots, 81$);
- Y_{i2} : pol do caldo clarificado ($i = 1, \dots, 81$);
- Y_{i3} : pureza do caldo ($i = 1, \dots, 81$);
- Y_{i4} : AR decantado ($i = 1, \dots, 81$);
- Y_{i5} : ART do caldo ($i = 1, \dots, 81$);
- X_{i1} : caldo misto extraído da moenda ($i = 1, \dots, 81$);
- X_{i2} : brix do caldo misto ($i = 1, \dots, 81$);
- X_{i3} : densidade do caldo misto ($i = 1, \dots, 81$);
- X_{i4} : pol do caldo misto ($i = 1, \dots, 81$);
- X_{i5} : pureza do caldo misto ($i = 1, \dots, 81$);
- X_{i6} : ART do caldo misto ($i = 1, \dots, 81$);
- X_{i7} : AR do caldo misto ($i = 1, \dots, 81$);
- X_{i8} : leitura do sacarímetro ($i = 1, \dots, 81$);
- X_{i9} : volume gasto de AR ($i = 1, \dots, 81$).

3.1.1.3 Fermentação

A fermentação tem início com o preparo do mosto, que é constituído de uma mistura de caldo, melaço e água. O processo de fermentação, geralmente, utilizado nas destilarias do Brasil é o de *Melle-Boinot*, cuja característica principal é a recuperação da levedura através da centrifugação do vinho (mosto fermentado). Esta levedura recuperada, antes de retornar ao processo fermentativo, recebe um tratamento severo, que consiste em diluição com água e adição de ácido sulfúrico até, normalmente, $pH = 2,5$, ou mais baixo, no caso de haver infecção bacteriana. Esta suspensão de fermento diluído e acidificado, conhecido na prática como “pé-de-cuba”, permanece em agitação, durante 1 hora e 30 minutos, antes de retornar à fermentação.

Na fermentação, os açúcares são transformados em álcool, segundo a reação de Gay Lussac, uma lei volumétrica, utilizada para medir o volume (%) de teor alcoólico das bebidas. As reações ocorrem em tanques denominados dornas de fermentação, onde se misturam o mosto e o pé-de-cuba na proporção de 2:1, respectivamente. O tempo de fermentação é de aproximadamente 8 horas. Ao final deste período, praticamente todo o açúcar já foi consumido, com a consequente redução da liberação de gases e multiplicação do fermento. Após a fermentação, o vinho é enviado à centrífugas para recuperação do fermento, que é tratado novamente e utilizado para continuidade do processo fermentativo.

As variáveis consideradas nesta etapa foram:

- Y_{i1} : etanol no vinho centrifugado ($i = 1, \dots, 81$);
- Y_{i2} : etanol no leite tratado ($i = 1, \dots, 81$);
- X_{i1} : mel ($i = 1, \dots, 81$);
- X_{i2} : brix do mosto ($i = 1, \dots, 81$);
- X_{i3} : ART do mosto ($i = 1, \dots, 81$);
- X_{i4} : acidez sulfúrica do mosto ($i = 1, \dots, 81$);
- X_{i5} : brix da mosto nas dornas ($i = 1, \dots, 81$);
- X_{i6} : GL nas dornas ($i = 1, \dots, 81$);
- X_{i7} : % de levedo nas dornas ($i = 1, \dots, 81$);
- X_{i8} : acidez do mosto nas dornas ($i = 1, \dots, 81$);
- X_{i9} : temperatura do mosto nas dornas ($i = 1, \dots, 81$);
- X_{i10} : % levedo no creme ($i = 1, \dots, 81$);
- X_{i11} : % levedo no vinho centrifugado ($i = 1, \dots, 81$);
- X_{i12} : brix do leite tratado ($i = 1, \dots, 81$);
- X_{i13} : % levedo na cuba ($i = 1, \dots, 81$);
- X_{i14} : perda ART na fermentação ($i = 1, \dots, 81$).

3.1.1.4 Destilação

O vinho, resultante da fermentação, possui de 7° a 10°GL de álcool e é composto por elementos de natureza líquida, sólida e gasosa. Além do álcool, os elementos líquidos são: água (cerca de 89% a 93%), glicerol, álcoois homólogos superiores, furfural, aldeído acético, ácidos succínico e acético e outros. Os elementos sólidos são formados por bagacilhos, leveduras e bactérias, açúcares não-fermentescíveis, sais minerais, matérias albuminóides e outros, e os gasosos, principalmente pelo CO_2 e SO_2 .

A destilação utiliza os diferentes pontos de ebulição, das diversas substâncias voláteis presentes, para separar o álcool dos demais componentes do vinho, sendo realizada com o auxílio de colunas de destilação. A primeira coluna tem a finalidade de concentrar o

vinho e é composta por três colunas superpostas denominadas: A, A_1 e D. O vinho centrifugado é alimentado no topo da coluna A_1 , descendo pelas bandejas e sofrendo a epuração. A flegma é retirada no fundo desta coluna e enviada à segunda coluna, denominada, coluna de destilação (coluna B). As impurezas voláteis são eliminadas pelo topo da coluna D. A vinhaça, composta basicamente de água, sais sólidos em suspensão e solúveis, é retirada da base da coluna A, em uma proporção aproximada de doze litros para cada litro de álcool produzido, sendo utilizada na lavoura como fertilizante. A flegma é concentrada a uma graduação de aproximadamente 96°GL, também é purificada, sendo retirada, na forma de álcool hidratado, duas bandejas abaixo do topo desta coluna. Também são retiradas as impurezas do flegma (flegmaça) e o óleo fúsel.

Nesta etapa foram consideradas variáveis envolvidas nos processos anteriores, com a inclusão das variáveis referentes à destilação, apresentadas a seguir:

- Y_i : etanol hidratado ($i = 1, \dots, 81$);
- X_{i1} : cana inteira que entra na moenda ($i = 1, \dots, 81$);
- X_{i2} : cana picada que entra na moenda ($i = 1, \dots, 81$);
- X_{i3} : índice de preparo da cana ($i = 1, \dots, 81$);
- X_{i4} : pureza da cana ($i = 1, \dots, 81$);
- X_{i5} : fibra da cana ($i = 1, \dots, 81$);
- X_{i6} : pol do caldo da cana desfibrada ($i = 1, \dots, 81$);
- X_{i7} : volume da água de embebição ($i = 1, \dots, 81$);
- X_{i8} : brix do caldo misto ($i = 1, \dots, 81$);
- X_{i9} : densidade do caldo misto ($i = 1, \dots, 81$);
- X_{i10} : pol do caldo misto ($i = 1, \dots, 81$);
- X_{i11} : pureza do caldo misto ($i = 1, \dots, 81$);
- X_{i12} : ART do caldo misto ($i = 1, \dots, 81$);
- X_{i13} : AR do caldo misto ($i = 1, \dots, 81$);
- X_{i14} : brix do caldo caleado ($i = 1, \dots, 81$);
- X_{i15} : leitura sacarímetro ($i = 1, \dots, 81$);
- X_{i16} : volume gasto de AR ($i = 1, \dots, 81$);
- X_{i17} : pol do caldo clarificado ($i = 1, \dots, 81$);
- X_{i18} : pureza caldo destilaria ($i = 1, \dots, 81$);
- X_{i19} : AR decantado destilaria ($i = 1, \dots, 81$);
- X_{i20} : ART do caldo ($i = 1, \dots, 81$);
- X_{i21} : mel ($i = 1, \dots, 81$);
- X_{i22} : brix do mosto ($i = 1, \dots, 81$);
- X_{i23} : ART do mosto ($i = 1, \dots, 81$);

X_{i24} : acidez sulfúrica do mosto ($i = 1, \dots, 81$);
 X_{i25} : brix do mosto nas dornas ($i = 1, \dots, 81$);
 X_{i26} : GL nas dornas ($i = 1, \dots, 81$);
 X_{i27} : % de levedo nas dornas ($i = 1, \dots, 81$);
 X_{i28} : acidez do mosto nas dornas ($i = 1, \dots, 81$);
 X_{i29} : temperatura do mosto nas dornas ($i = 1, \dots, 81$);
 X_{i30} : % levedo no creme ($i = 1, \dots, 81$);
 X_{i31} : % levedo no vinho ($i = 1, \dots, 81$);
 X_{i32} : etanol no vinho centrifugado ($i = 1, \dots, 81$);
 X_{i33} : brix do leite tratado ($i = 1, \dots, 81$);
 X_{i34} : etanol no leite tratado ($i = 1, \dots, 81$);
 X_{i35} : levedo na cuba ($i = 1, \dots, 81$);
 X_{i36} : perda de ART na fermentação ($i = 1, \dots, 81$);
 X_{i37} : volume da vinhaça ($i = 1, \dots, 81$);
 X_{i38} : perda de ART na vinhaça ($i = 1, \dots, 81$);
 X_{i39} : % etanol na vinhaça ($i = 1, \dots, 81$);
 X_{i40} : ART do caldo destilaria ($i = 1, \dots, 81$);
 X_{i41} : ART do mel final ($i = 1, \dots, 81$);
 X_{i42} : rendimento da destilaria ($i = 1, \dots, 81$);
 X_{i43} : litros de álcool por tonelada de mel ($i = 1, \dots, 81$).

3.2 Métodos

3.2.1 A Filosofia Lean

A meta da filosofia *Lean* é acelerar a velocidade de qualquer processo, através da redução de desperdício em todas as suas formas. Com o uso dos conceitos e das ferramentas *Lean*, pode-se verificar que os passos do processo, que até então, se pensava serem essenciais, são, na verdade, desnecessários e seus custos e atrasos são eliminados, após a aplicação das ferramentas *Lean* (GEORGE, 2004). Portanto, a redução dos custos é feita através da eliminação total das perdas e aumento da produtividade. O objetivo da ferramenta *Lean* é reduzir os tempos dos processos, através da eliminação dos sete tipos de desperdícios, identificados por Taiichi Ohno:

1. Produtos defeituosos;
2. Excesso de produção;
3. Estoques (inventário);

4. Processos inadequados;
5. Movimento desnecessário;
6. Transporte;
7. Espera.

Womack e Jones (2004) acrescentam a esta lista “o projeto de produtos e serviços que não atendem às necessidades do cliente”, e propõem os 5 princípios do “Pensamento Enxuto” (*Lean Thinking*), segundo o qual este trabalho foi fundamentado, os quais são:

1. Definir o que é Valor;

O ponto de partida para a mentalidade enxuta consiste em definir o que é valor. Não é a empresa, mas sim o cliente que define o que é valor.

O conceito de valor pode ser entendido como o conteúdo inerente de um produto, segundo o julgamento do cliente, refletido em seu preço de venda e demanda de mercado. Agrega valor ao produto uma combinação de ações do processo que transformam a matéria prima, cujo resultado em um produto típico é criado pelo fabricante por meio de ações, algumas das quais produzem valor, conforme percebido pelo cliente e pelos quais ele está disposto a pagar.

2. Identificar o Fluxo de Valor;

O fluxo de valor é identificado através de um detalhamento da cadeia produtiva, separando os processos em três tipos: aqueles que agregam valor, aqueles que não agregam valor mas são importantes para a manutenção dos processos e da qualidade e, por fim, aqueles que não agregam valor, devendo ser eliminados imediatamente.

3. Fluxo Contínuo;

A criação de um fluxo contínuo proporciona redução nos tempos do processo, processamento de pedidos e estoques.

4. Produção Puxada;

O fluxo contínuo permite inverter o fluxo produtivo, ou seja, as empresas deixam de empurrar os produtos e o consumidor passa a “puxar” a produção, eliminando estoques e dando valor ao produto.

5. A Busca da Perfeição;

A busca do aperfeiçoamento contínuo, visando a uma estado ideal, deve ser o objetivo de todos os envolvidos no fluxo de valor.

As principais ferramentas utilizadas foram as propostas pelo “Pensamento Enxuto”:

i) Mapeamento do Fluxo de Valor (*Value Stream Mapping* - *VSM*);

O mapeamento do fluxo do valor é uma ferramenta que utiliza símbolos gráficos para documentar e apresentar, visualmente, a sequência e o movimento de informações, materiais e ações que constituem o fluxo de valor de uma empresa. Ajuda compreender e agilizar processos que usam as ferramentas e as técnicas de *Lean Manufacturing*.

O objetivo de VSM é identificar, demonstrar e diminuir o desperdício no processo. O desperdício é definido como toda a atividade que não adiciona valor ao produto final. Desta forma, o VSM pode ajudar à gerência, engenheiros, sócios de produção, fornecedores e clientes a reconhecer o desperdício e identificar suas causas.

O VSM é, primeiramente, uma ferramenta de uma comunicação, mas pode também ser usada como uma ferramenta de planejamento estratégico. É como uma ferramenta de gerência de mudança.

Com esta finalidade, neste método foi traçado, visualmente, o fluxo de materiais e de informação, desde o momento que a matéria prima entra no processo, através de todas as etapas do processo de manufatura, até a entrega do produto ao cliente.

ii) Métricas *Lean*;

As principais métricas utilizadas foram as apresentadas por *Lean Enterprise Institute* (2003):

1) Tempo de Ciclo - T/C

É o tempo necessário para um produto ser finalizado em uma determinada etapa do processo.

2) *Lead Time* - L/T

É o tempo necessário para um produto percorrer todas as etapas de um processo ou fluxo do valor.

3) Tempo de Agregação de Valor - TAV

É o tempo dos elementos de trabalho que realmente transformam o produto de uma maneira que o cliente se disponha a pagar.

4) Tempo de Não Agregação de Valor - TNAV

Tempo gasto em atividades que adicionam custos, mas não agregam valor do ponto de vista do cliente. Armazenamento, inspeção, *set-up*, movimentação, espera e retrabalho são exemplos de atividades que não agregam valor.

5) Eficiência do Ciclo do Processo (PCE - *Process Cycle Efficiency*)

Indicador que mede a relação entre o tempo de agregação de valor e o *Lead Time*. O cálculo é obtido pela expressão:

$$PCE = \frac{TAV}{Lead\ Time}.$$

6) Taxa de Saída

Resultado de um processo ao longo de um período de tempo definido, expresso em unidade/tempo.

7) Trabalho em Processo (WIP - *Work in Process*)

O trabalho em processo é qualquer trabalho que esteja no processo e que ainda não esteja concluído. Pode ser relacionado ao *Lead Time* por meio da Lei de *Little*:

$$Lead\ Time = \frac{WIP}{Taxa\ de\ Saída}.$$

8) Tempo *Takt* (*Takt Time*): *takt* palavra de origem alemã que significa: ritmo, cadência.

É o tempo disponível para a produção dividido pela demanda do cliente:

$$TempoTakt = \frac{\text{Tempo de produção disponível}}{\text{Necessidade de produção}}.$$

9) Eficácia Total do Equipamento (OEE - *Overall Equipment Effectiveness*)

Indicador de manutenção produtiva total que mede o grau de eficiência no uso de um equipamento. É composto por:

- a) Taxa de disponibilidade (*Uptime*): representa o tempo de produção (excluindo-se as paradas causadas por falhas nos equipamentos ou ajustes) em relação ao tempo disponível de trabalho;
- b) Taxa de desempenho: representa o total produzido (%) em relação à capacidade de produção nominal;
- c) Taxa de qualidade (*% Yield*): representa o total de itens não defeituosos em relação ao total produzido.

O OEE é obtido da seguinte forma:

$$OEE = \text{Taxa de disponibilidade} \times \text{Taxa de desempenho} \times \text{Taxa de qualidade}.$$

3.2.2 O Programa Seis Sigma

De acordo com Pande; Neuman e Cavanagh (2001), o modelo de uma empresa através da perspectiva do fluxo do processo pode ser representado pela figura 2.

Figura 2 - Modelo de uma empresa vista através da perspectiva do fluxo do processo

Na esquerda estão as entradas do processo, no meio está a organização ou o processo em si e na direita estão os produtos finais. Em termos de variáveis, os X's que aparecem na entrada e no fluxo do processo são os indicadores de mudança do sistema. À direita estão os Y's que representam medidas do desempenho dos negócios. Em termos matemáticos, $Y = f(X)$ é uma maneira de dizer que mudanças ou variáveis nas entradas e processos do sistema irão determinar, de forma acentuada, as variações que ocorrem na variável Y.

O segredo no Programa Seis Sigma, segundo os autores, está em:

1. Entender quais variáveis, no processo, têm maior influência sobre os resultados;
2. Usar as mudanças, no desempenho do processo, para ajustar os negócios e mantê-los em um caminho lucrativo.

O objetivo de se buscar um desempenho Seis Sigma é o de reduzir a variação a tal ponto, que seis sigma possam ser comprimidos dentro de limites aceitáveis pelo cliente.

3.2.2.1 Melhoria do Processo

O termo “Melhoria do Processo” refere-se à estratégia de desenvolver soluções, com a finalidade de eliminar as causas dos problemas de desempenho de uma empresa. Em termos Seis Sigma, a ênfase está em encontrar e apontar soluções para as variáveis de menor importância (X) que afetam a variável Y. Portanto, a maioria dos projetos Seis Sigma é de esforços de Melhoria de Processo (PANDE; NEUMAN; CAVANAGH, 2001).

Os processos podem ser melhorados de duas formas.

1. Melhorar o produto através da Engenharia de Valor;

2. Melhorar os métodos de fabricação do ponto de vista da engenharia de produção ou da tecnologia de fabricação.

A engenharia de valor foi o primeiro estágio na melhoria do processo. O objetivo é fazer mudanças na estrutura do projeto de modo que se possa manter a qualidade e, reduzir os custos de fabricação. No segundo estágio, buscam-se ferramentas para a melhoria da fabricação do produto.

A metodologia, de soluções de problemas e gerenciamento, utilizada neste trabalho foi o DMAIC.

O DMAIC (*Define, Measure, Analyze, Improve, Control*) é uma metodologia de solução de problemas utilizada pela maioria das empresas americanas que empregam o Programa Seis Sigma com o objetivo de realizar melhorias em produtos, serviços e processos.

A implementação tem que ser feita a partir dos mais altos níveis da companhia, ou seja, de cima para baixo, atingindo todos os níveis da gerência (HOERL, 1998 apud JURAN; GODFREY, 1998).

O foco principal do conceito Seis Sigma, assim como muitas outras atividades de qualidade, é a redução do custo, a melhoria da produção, da capacidade e reduções do tempo do ciclo de produção. A ênfase está na necessidade de satisfação do cliente.

3.2.2.2 Integração das ferramentas *Lean* e Seis Sigma ao DMAIC

Muitos modelos de melhorias têm sido aplicados nos processos, ao longo dos anos, desde que o movimento de qualidade começou. A maioria deles se baseia nos métodos de Deming (PDCA). Na Estratégia Seis Sigma é usado um ciclo de melhoria de cinco fases (DMAIC). Como outros métodos de melhoria, o DMAIC baseia-se no ciclo original PDCA.

A aplicação da metodologia e ferramentas do Lean Seis Sigma foi feita utilizando o modelo DMAIC, conforme as fases descritas a seguir.

1. Define - Nessa fase inicial foi preciso definir com precisão o escopo do projeto, sendo composta pelas seguintes etapas:
 - a) *Project Charter*: é uma espécie de contrato entre a equipe responsável pela condução do projeto e os gestores da empresa. Este documento é composto pelos seguintes itens:
 - a1) Descrição do problema: a descrição deve focalizar o problema considerado, mostrar os indicadores ou métricas usados para medir o problema, o local e a época de ocorrência do mesmo, o impacto e os ganhos financeiros da solução do problema;

- a2) Definição da meta: uma meta é constituída por um objetivo gerencial, um valor e um prazo (CAMPOS, 1996, p.45);
- a3) Avaliação do histórico do problema: dados conhecidos previamente que podem ajudar a entender o problema devem ser apresentados;
- a4) Restrições e suposições: restrições como baixo tempo de dedicação dos membros da equipe ao projeto e inexistência de dados confiáveis devem ser registradas;
- a5) Equipe de trabalho: foi estabelecida a equipe e suas responsabilidades. O projeto é conduzido pela equipe liderada pelos patrocinadores e especialistas do Seis Sigma, conforme proposta na metodologia, descritos a seguir:
 - i) Patrocinadores:
 - (a) *Sponsor* - É o principal executivo da empresa, responsável por promover e definir as diretrizes para a implementação do programa Seis Sigma;
 - (b) *Sponsor* Facilitador - É um dos diretores da empresa que tem a responsabilidade de assessorar o *Sponsor* na implementação do programa;
 - (c) *Champions* - Membro da gerência com a responsabilidade de apoiar os projetos e remover barreiras que possam prejudicar o o desenvolvimento do projeto.
 - ii) Especialistas:
 - (a) *Master Black Belts* - São profissionais que assessoram os *Sponsors* e os *Champions* e atuam como mentores dos *Black Belts* e *Green Belts*;
 - (b) *Black Belts* - Lideram equipes na condução dos projetos multifuncionais ou funcionais;
 - (c) *Green Belts* - São profissionais que lideram equipes na condução de projetos funcionais ou participam das equipes lideradas pelos *Black Belts*;
 - (d) *White Belts* - São profissionais do nível operacional da empresa, que dão suporte aos *Black Belts* e *Green Belts* na implementação dos projetos.
- a6) Cronograma preliminar: estabelecidos os prazos para a finalização das atividades do projeto.
- b) SIPOC (*Suppliers, Inputs, Process, Outputs, Customer*): um diagrama de SIPOC é a ferramenta usada para identificar e descrever todos os elementos relevantes para o projeto de melhoria de processo, antes do início do trabalho.

2. Measure - Nesta fase o objetivo foi determinar a localização ou o foco do problema. As principais ferramentas usadas foram:

- a) Mapeamento do Fluxo de Valor (VSM): o fluxo de valor é o conjunto das atividades que são realizadas pela empresa para projetar, produzir e entregar seus produtos aos clientes. O mapeamento do fluxo do valor foi a ferramenta utilizada para documentar e apresentar visualmente a sequência e o movimento de informações, materiais e ações que constituem o fluxo de valor da empresa, ajudando a compreender e agilizar processos que usam as ferramentas e as técnicas de *Lean Manufacturing*. As principais etapas do mapeamento do fluxo de valor são:
- i. Definir o produto;
 - ii. Construir mapa do estado atual;
 - iii. Construir mapa do estado futuro;
 - iv. Estabelecer plano de trabalho e implementação.
- b) Mapas do processo (PMAP): os mapas do processo são documentos vivos de trabalho, ou seja, uma representação visual de $Y = f(X)$. O mapeamento dos processos básicos de produção foi obtido considerando as seguintes ações:
- i. Listar as entradas e saídas. O mapeamento preliminar do processo foi desenvolvido através da ferramenta SIPOC;
 - ii. Identificar todas as etapas no processo: as que agregam valor e aquelas que não agregam valor;
 - iii. Identificar as principais saídas em cada etapa (processo e produto);
 - iv. Identificar as principais entradas do processo e classificá-las;
As variáveis de entrada (X) são obtidas com o auxílio da ferramenta 6 M's (Mão de obra, Máquina, Método, Material, Meios de medição e Meio ambiente) e são classificadas da seguinte forma:
 - Entradas controláveis (C): são variáveis de entrada (X) que podem ser alteradas para verificar o efeito nas variáveis de saída do processo (Y). Algumas vezes são chamadas variáveis de ajuste.
 - Entradas controláveis não medidas pela usina (C*).
 - Entradas padrão (S): são procedimentos padrão de operação que descrevem como o processo deve rodar e identifica certos fatores para monitorar e manter.
 - Entradas de ruído (N): são variáveis de entrada que impactam a saída, mas não são controláveis. São difíceis ou muito custosas para controlar.
 - v. Adicionar as especificações operacionais e objetivos do processo para as entradas controláveis e aquelas críticas.

- c) Plano para coleta de dados: representa o 5W/H (*Who, What, Where, When, Why e How*) do processo de coleta de dados;
 - d) Métricas do *Lean* Seis Sigma.
3. *Analyze* - Na etapa *analyze* as relações de causa e efeito, entre as variáveis de entrada e saída do processo, foram identificadas. A ferramenta utilizada foi o mapeamento do fluxo de valor: o método VSM pode ser visto como uma variante do mapeamento do processo com adição de atividades. No VSM o foco está nos passos do processo, incluindo manipulação de material e retrabalho, que podem ser simplificados ou eliminados. Além disso, envolve análise e, sugere melhoria através do mapa do estado futuro.
 4. *Improve* - Nesta fase, deve-se, com base nas informações das etapas anteriores, propor recomendações visando a melhoria do processo. A ferramenta utilizada foi a análise de regressão.
 5. *Control* - Na última fase do DMAIC, deve-se monitorar as implementações efetuadas, na fase anterior, para garantir que o alcance da meta seja mantido a longo prazo.

3.2.3 Regressão Linear Múltipla

A ferramenta utilizada para descrever a relação entre a variável de saída (Y) e as variáveis de entrada (X) foi a regressão linear múltipla.

3.2.3.1 Modelo Estatístico

O modelo estatístico para a regressão linear múltipla com k variáveis regressoras é dado por:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (1)$$

O modelo estatístico para a regressão linear múltipla, com as variáveis centradas, foi o modelo na forma reparametrizada (DEMÉTRIO; ZOCCHI, 2006) dado por:

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (2)$$

em que $\alpha = \beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 + \dots + \beta_k \bar{X}_k$ e $x_{ij} = X_{ij} - \bar{X}_j$, sendo $\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij}$, $j = 1, \dots, k$,

y_i é a variável resposta, isto é, $y_i = Y_i - \bar{Y}$, sendo $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$;

β_i ($i = 1, \dots, k$) são os parâmetros da regressão do modelo;

x_{ij} ($j = 1, \dots, k$) são as variáveis regressoras;

ε_i é o erro aleatório.

Utilizando notação matricial, o modelo de regressão linear múltipla apresenta-se como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} = \boldsymbol{\mu} + \boldsymbol{\varepsilon} \quad (3)$$

ou seja,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} \quad e \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix},$$

em que:

$\mathbf{y}$ é o vetor de observações ($n \times 1$) da variável aleatória Y ;

$\mathbf{X}$ é a matriz ($n \times p$) de variáveis regressoras;

$\boldsymbol{\theta}$ é o vetor ($p \times 1$) dos coeficientes de regressão;

$\boldsymbol{\varepsilon}$ é o vetor ($n \times 1$) de erros aleatórios.

As suposições para o modelo de regressão linear múltipla são:

- i) A variável resposta $\mathbf{y}$ é função linear das variáveis regressoras X_j , $j = 1, 2, \dots, k$;
- ii) As variáveis X_j são fixas;
- iii) $E(\varepsilon_i) = 0$ e $Var(\varepsilon_i) = \sigma^2$;
- iv) Os erros são independentes, isto é, $Cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j$;
- v) Os erros têm distribuição normal.

Combinando (iii) e (iv) têm-se $Var(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \mathbf{I}\sigma^2$, sendo $\mathbf{I}$ uma matriz identidade, de dimensões $n \times n$. Assim, considerando-se, também, (v) têm-se que $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$ e $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\theta}, \mathbf{I}\sigma^2)$, pois $E(\mathbf{y}) = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\theta}$ e $Var(\mathbf{y}) = \mathbf{I}\sigma^2$.

3.2.3.2 Método dos Quadrados Mínimos

O método utilizado para obter as estimativas dos parâmetros do modelo foi o método dos quadrados mínimos (MONTGOMERY; PECK, 1992). Considerando o modelo (3), o vetor de erros é dado por:

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\theta} = \mathbf{y} - \boldsymbol{\mu}. \quad (4)$$

O método dos quadrados mínimos minimiza o comprimento do vetor de erros ($\boldsymbol{\varepsilon}$). Usando a norma euclidiana para a obtenção do comprimento de $\boldsymbol{\varepsilon}$, tem-se:

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n \varepsilon_i^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) = \mathbf{y}'\mathbf{y} - 2\boldsymbol{\theta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\theta}'\mathbf{X}'\mathbf{X}\boldsymbol{\theta}. \quad (5)$$

Derivando $S(\boldsymbol{\theta})$ em relação a $\boldsymbol{\theta}$, obtém-se:

$$\frac{\partial S}{\partial \boldsymbol{\theta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\theta}$$

e igualando $\frac{\partial S}{\partial \boldsymbol{\theta}}$ a zero, obtém-se:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{X}'\mathbf{y}$$

denominado de sistema de equações normais, que é consistente e apresenta solução única dada por:

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad (6)$$

sendo que:

$\hat{\boldsymbol{\theta}}$ é um estimador não viciado de $\boldsymbol{\theta}$, isto é $E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$ e com variância dada por:

$$Var(\hat{\boldsymbol{\theta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2. \quad (7)$$

3.2.3.3 Teste t e teste F

Utilizando o teste t , dado por:

$$t = \frac{\hat{\beta}_i - \beta_i}{\sqrt{s(\hat{\beta}_i)}}, \quad \text{em que}, \quad (8)$$

$H_0 : \beta_i = 0$ versus $H_1 : \beta_i \neq 0$, $i = 1, 2, \dots, k$, verificou-se a significância dos efeitos no modelo.

Para os efeitos significativos realizou-se a análise de variância (tabela 3) e verificou-se, individualmente, a inclusão das variáveis não significativas pelo teste F .

Tabela 3 - Esquema da análise de variância de regressão

Causas de variação	GL	SQ	QM	F
Regressão	w	SQRegressão	$\frac{\text{SQRegressão}}{w}$	$\frac{\text{QMRegressão}}{\text{QMResíduo}}$
Resíduo	$n - w - 1$	SQResíduo	$\frac{\text{SQResíduo}}{(n-w-1)}$	
Total	$n - 1$	SQTotal		

em que:

$$\text{SQRegressão} = \hat{\boldsymbol{\theta}}_R' \mathbf{X}_R' \mathbf{y} - C, \quad \text{sendo,}$$

$\hat{\boldsymbol{\theta}}_R$ é o vetor, de ordem $w + 1$, dos parâmetros referentes aos efeitos significativos no modelo, em que $w \leq k$;

$\mathbf{X}_R$ é a matriz das variáveis correspondentes,

$$\text{SQTotal} = \sum_{i=1}^n y_i^2 - C; \quad \text{sendo} \quad C = \frac{\left(\sum_{i=1}^n y_i \right)^2}{n},$$

e a soma de quadrados dos resíduos é obtida por diferença.

3.2.3.4 Coeficiente de Correlação Múltiplo

O coeficiente de correlação múltiplo (coeficiente de determinação), denotado por (R^2), representa a proporção da variação de Y, explicada pela regressão (DRAPER; SMITH, 1998) e é obtido por:

$$R^2 = \frac{\text{SQReg}}{\text{SQTotal}} = 1 - \frac{\text{SQRes}}{\text{SQTotal}}, \quad 0 \leq R^2 \leq 1. \quad (9)$$

O R^2 deve ser usado com cuidado para avaliar a qualidade do ajuste (DEMÉTRIO; ZOCCHI, 2006), pois:

- i) Depende do tamanho da amostra (n), tendendo a crescer quando n diminui;
- ii) Seu valor pode aumentar com a inclusão de uma nova variável no modelo;
- iii) A magnitude de R^2 , também, depende da amplitude de variação das variáveis regressoras. Em geral, o valor de R^2 aumenta com maior amplitude de variação dos X 's e diminui em caso contrário.

Assim, o R^2 deve ser considerado em conjunto com outros diagnósticos do modelo. Com a finalidade de tentar corrigir estes problemas foi definido o coeficiente de determinação ajustado para graus de liberdade, denotado por $\bar{R}^2$. Da equação (9), tem-se que:

$$1 - R^2 = 1 - \frac{SQReg}{SQTotal} = \frac{SQRes}{SQTotal}. \quad (10)$$

Substituindo SQRes e SQTotal, na equação (9), pelos respectivos quadrados médios, obtém-se o coeficiente de determinação ajustado (MONTGOMERY; PECK, 1992), dado por:

$$\bar{R}^2 = 1 - \frac{SQRes/(n-p)}{SQTotal/(n-1)} = 1 - \frac{n-1}{n-p} \frac{SQRes}{SQtotal}. \quad (11)$$

Portanto, usando a equação (10), o coeficiente de determinação ajustado é dado por:

$$\bar{R}^2 = 1 - \frac{n-1}{n-p} (1 - R^2). \quad (12)$$

3.2.3.5 Métodos de seleção de variáveis

Segundo Demétrio e Zocchi (2006) o objetivo de uma análise de regressão linear múltipla é obter a melhor descrição possível da variável dependente (Y), em função das variáveis regressoras (X). Um número muito grande de variáveis regressoras pode levar a um modelo que explique bem os dados, mas de difícil interpretação. Por outro lado, um número pequeno de variáveis regressoras pode facilitar a interpretação, porém, o ajuste pode não ser adequado. O ideal é um modelo intermediário que produza um bom ajuste aos dados e com o menor número possível de parâmetros. Os critérios utilizados para a seleção de variáveis no modelo foram:

1. Coeficiente de determinação ajustado;
2. A estatística C_p (MALLOWS, 1973 *apud* DRAPER; SMITH, 1998) dada por:

$$C_p = \frac{SQRes_p}{s^2} - (n - 2p), \quad (13)$$

em que:

$SQRes_p$ é a soma de quadrados do resíduo do modelo, com p parâmetros;

s^2 é o quadrado médio do resíduo referente ao modelo ajustado com todas as variáveis regressoras (modelo completo);

n é o número de observações.

Para $k = p - 1$ variáveis selecionadas, valores grandes de C_p indicam modelos com QMRes grandes. Se o modelo possui $C_p > p$ é um indicativo de que o modelo está

mal especificado. Por outro lado, se $C_p < p$, é um indicativo de que o modelo está superparametrizado. Uma alternativa é ajustar todos os possíveis modelos e escolher o que possui o menor número de parâmetros e valor da estatística C_p menor ou igual ao do número de parâmetros do modelo.

3. Critério de informação de Akaike (AIC): é definido por:

$$AIC = -2\log L + 2p \quad (14)$$

em que, L é a função de verossimilhança e p o número de parâmetros ajustados.

4. Teste F: a comparação entre dois modelos com p e q parâmetros, sendo $p < q$, pode ser feita utilizando a estatística:

$$F = \frac{SQRes_p - SQRes_q}{(q - p)QMRes}, \quad (15)$$

sendo $QMRes$ uma estimativa de σ^2 , em geral, obtida a partir do modelo completo.

Os métodos utilizados para a seleção de variáveis no modelo foram:

1. Todas as regressões possíveis: consiste no ajuste de todos os modelos (2^k), em que os modelos são separados em grupos, de acordo com o número de variáveis, sendo então, cada grupo ordenado de acordo com algum critério de seleção de variáveis.
2. Método do passo atrás (*backward*): esse método busca examinar os principais modelos de regressão, com um certo número de variáveis. Consiste em ajustar, inicialmente, o modelo completo e eliminar, uma a uma as variáveis com base na:
 - i) menor correlação parcial com a resposta Y ;
 - ii) menor diminuição no R^2 ;
 - iii) menor diminuição significativa no teste F parcial ou no teste t parcial, de acordo com algum critério de parada.
3. Método do passo a frente (*forward*): esse método supõe que não existem variáveis regressoras no modelo, além do intercepto. Consiste em incluir no modelo a variável com maior coeficiente de correlação simples com a variável resposta e, na sequência, incluir, uma a uma, variáveis com:
 - i) maior correlação parcial com a resposta Y ;
 - ii) maior aumento no R^2 ;
 - iii) maior aumento significativo no teste F parcial.

4. Método do passo a frente passo atrás (*stepwise*): esse método é uma modificação do método do passo a frente (*forward*), na qual, em cada passo, todas as variáveis adicionadas previamente no modelo são avaliadas através da estatística F parcial ou pelo Critério de Informação de Akaike.

3.2.3.6 Análise de resíduos

Para o modelo linear $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$ tem-se que o vetor $\boldsymbol{\varepsilon}$ contém os elementos ε_i , em que ε_i são as diferenças entre os valores observados ($\mathbf{y}_i's$) e os valores ajustados ($\hat{\mathbf{y}}_i's$).

Admite-se que os $\varepsilon_i's$ são independentes e que $\varepsilon_i \sim N(0, \sigma^2)$ seguindo as suposições do modelo especificadas em 3.2.3.1. Quando essas suposições são violadas, a análise do modelo pode levar a conclusões duvidosas.

Montgomery; Peck e Vining (2006) definem três tipos de resíduos:

1. Resíduos ordinários: utilizando o método dos quadrados mínimos, para o modelo linear $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$, tem-se que os resíduos ordinários são obtidos por:

$$e_i = y_i - \hat{y}_i = \mathbf{y}_i - \mathbf{X}\hat{\boldsymbol{\theta}}.$$

Os valores ajustados são obtidos por:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\theta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{H}\mathbf{y} \quad (16)$$

sendo que $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ é chamada de matriz de projeção. Os elementos da matriz $\mathbf{H}$ são denotados por h_{ij} , ou seja

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix},$$

e os elementos da diagonal da matriz $\mathbf{H}$, denominados h_i de *Leverage*, fornecem a distância de uma observação em relação às demais. Quando o modelo de regressão passa pela origem, $0 \leq h_i \leq 1$. A matriz $\mathbf{H}$ é simétrica ($\mathbf{H} = \mathbf{H}'$) e idempotente ($\mathbf{H}\mathbf{H} = \mathbf{H}$).

Utilizando a matriz $\mathbf{H}$, os resíduos podem ser expressos por:

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (17)$$

porém, enquanto os erros ε_i 's são independentes e de mesma variância σ^2 , não se pode dizer o mesmo para os resíduos ordinários, pois $Var(\mathbf{e}) = Var[(\mathbf{I} - \mathbf{H})\mathbf{y}] = (\mathbf{I} - \mathbf{H})\sigma^2$ e a variância para o i -ésimo resíduo é dada por: $Var(e_i) = (1 - h_{ii})\sigma^2$ e a covariância entre o i -ésimo e o j -ésimo resíduos por: $COV(e_i, e_j) = -h_{ij}\sigma^2$. Assim, o uso dos resíduos ordinários não é o mais indicado para a análise dos resíduos.

2. Resíduos estudentizados internamente

Sabe-se que o quadrado médio do resíduo (s^2) é uma estimativa para σ^2 . Assim, uma estimativa para $Var(e_i)$ é dada por:

$$\widehat{Var}(e_i) = (1 - h_{ii})s^2$$

e os resíduos estudentizados internamente são definidos por

$$rei_{(i)} = \frac{e_i}{\sqrt{\widehat{Var}(e_i)}} = \frac{y_i - \hat{y}_i}{s\sqrt{1 - h_{ii}}}. \quad (18)$$

Estes resíduos são assim chamados, pois a estimativa da variância residual é obtida considerando todas observações (n). Portanto, essa estimativa pode variar, consideravelmente, quando tem-se um valor discrepante. Além disso, o numerador e o denominador estão correlacionados, isto é $Cov[e_i, \widehat{Var}(e_i)] \neq 0$.

3. Resíduos estudentizados externamente

Pode-se garantir a independência entre o numerador e o denominador, na padronização dos resíduos, definindo-se os resíduos estudentizados externamente por

$$ree_{(i)} = \frac{e_i}{s_{(i)}\sqrt{1 - h_{ii}}}, \quad (19)$$

em que:

$s_{(i)}^2$ é a estimativa de σ^2 livre da influência da observação (i). A vantagem de utilizar os resíduos estudentizados externamente é que, sob normalidade, $ree_{(i)}$ tem distribuição t de *Student* com $(n - p - 1)$ graus de liberdade.

3.2.3.6.1 Diagnósticos em influência

Em algumas situações, pode ocorrer que um conjunto pequeno de observações causem uma influência muito grande sobre o modelo de regressão ajustado. Neste caso, é necessário localizar esses pontos influentes e avaliar seu impacto no ajuste do modelo (MONTGOMERY; PECK; VINING, 2006).

Para o devido ajuste no modelo, uma observação pode ser classificada, de acordo com Demétrio e Zocchi (2006), como:

1. Inconsistente: quando tem valor do $ree_{(i)}$ muito diferente da tendência geral das demais. Geralmente, é inconsistente se

$$|ree_{(i)}| \geq t_{\{1-(\frac{\gamma}{2n}); n-p-1\}}, \quad (20)$$

com nível de significância de $100\gamma\%$. Uma observação inconsistente é considerada um *outlier* quando possui medida de *leverage* (h_i) pequeno e resíduo $ree_{(i)}$ grande.

2. Ponto de alavanca: quando possui h_i grande, geralmente $h_i \geq 2p/n$, sendo considerado bom quando consistente e ruim quando inconsistente.
3. Influyente: uma observação é considerada influyente quando sua omissão no conjunto de dados provoca mudanças substanciais no ajuste do modelo.

3.2.3.6.2 Estatísticas para diagnósticos

1. DFBETAS: Belsley; Kuh; Welsch, 1980 apud Montgomery; Peck (1992) sugerem uma estatística que mede a alteração no vetor $\hat{\theta}$ quando a i -ésima observação é retirada da análise, e que é obtido por:

$$DFBETAS_{(i)} = \frac{\hat{\theta} - \hat{\theta}_i}{s_{(i)} \sqrt{C_{jj}}} = dfb.x_i, \quad (21)$$

em que:

C_{jj} é o j -ésimo elemento da diagonal da matriz $(\mathbf{X}'\mathbf{X})^{-1}$.

2. DFFITS: mede a alteração provocada no valor ajustado quando a i -ésima observação é retirada da análise.

$$DFFITS_{(i)} = \frac{\hat{y}_i - \hat{y}_{(i)}}{s_{(i)} \sqrt{h_{ii}}} = dffit, \quad i = 1, 2, \dots, n, \quad (22)$$

sendo que: $\hat{y}_{(i)}$ é o valor ajustado de y_i obtido sem considerar a i -ésima observação.

3. Distância de Cook: é uma medida de afastamento do valor do vetor $\hat{\theta}$ quando a i -ésima observação é retirada. É semelhante ao DFFITS, utilizando como estimativa da variância residual o resíduo estudatizado internamente. É obtida por:

$$D_{(i)} = \frac{h_{ii}}{(1 - h_{ii})} \frac{1}{p} (rei_{(i)})^2. \quad (23)$$

No pacote estatístico R, a i -ésima observação é considerada influyente se ao menos em um dos casos ocorrer (DEMÉTRIO; ZOCCHI, 2006):

1. $|DFBETAS_{(i)}| > 1;$
2. $|DFFITs_{(i)}| > 3\sqrt{p/(n-p)};$
3. $|1 - COVRATIO| > 3p/(n-p);$
4. $D_{(i)} > F_{\{50\%;p;n-p\}};$
5. $h_i > 3p/n.$

Montgomery; Peck e Vining (2006) definem *COVRATIO* como:

$$COVRATIO_i = \frac{|(\mathbf{X}'_{(i)}\mathbf{X}_{(i)})^{-1}s_{(i)}^2|}{|(\mathbf{X}'\mathbf{X})^{-1}s^2|} = cov.r.$$

4 RESULTADOS E DISCUSSÃO

Inicialmente, seguindo os procedimentos descritos para o método DMAIC, estabeleceu-se parceria com a destilaria DECASA. Esta parceria foi confirmada com a apresentação do *Project Charter* ao gerente industrial da usina e a formação da equipe de trabalho.

4.1 Mapeamento do processo

O mapeamento do processo produtivo do álcool hidratado foi realizado. Este mapeamento envolveu uma cuidadosa observação e documentação dos processos básicos de produção, desde o recebimento da matéria prima, a cana-de-açúcar, até o produto final, o álcool hidratado. A figura 3 apresenta o mapa do fluxo de valor do estado atual da usina de açúcar e álcool Decasa. Na figura, além das métricas *Lean*, são apresentadas algumas informações complementares e necessárias para o mapeamento do processo, como: distância percorrida (km), tempo de transporte [horas(h), minutos(m), segundos(s)], número de operadores e turnos de trabalho. Há, também, uma linha de contagem de tempo, expresso em horas, para as etapas do processo que agregam e não agregam valor ao produto.

Figura 3 - VSM - Mapeamento do fluxo de valor do estado atual

Com base no VSM do estado atual e utilizando as métricas *Lean*, obtiveram-se os seguintes resultados:

1. *Lead time* para o processo produtivo do álcool hidratado: $L/T = 185$ horas;
2. Tempo das etapas que agregam valor: $TAV = 6,02$ horas;
3. Tempo das etapas que não agregam valor: $TNAV = 178,98$ horas;
4. Eficiência do Ciclo de Processo: $PCE = 3,25\%$.

O Tempo *Takt* do processo de produção foi calculado usando: o tempo disponível para a produção (24 horas/dia) e a demanda do cliente ($500\text{ m}^3/\text{dia}$).

Portanto,

$$\text{Tempo Takt} = \frac{24}{500} = 0,048h/m^3 = 2,88\text{minutos}/m^3.$$

Os resultados mostraram que o tempo das etapas do processo, que agregam valor, é relativamente pequeno. O objetivo do VSM é identificar e diminuir o desperdício no processo, produzindo um Fluxo de Valor Enxuto. A principal fonte de desperdício é o excesso de produção, ou seja, produzir mais, ou mais rápido, do que o necessário para o processo seguinte. Isto faz com que o tempo de produção total (*Lead time*) do produto seja muito alto, enquanto o tempo de agregação de valor é muito pequeno. Para reduzir esse *Lead time* é necessário estar atento ao desperdício e eliminá-los no mapa do estado futuro.

Portanto, a meta do trabalho é diminuir o *Lead time* e consequentemente, aumentar a Eficiência do Ciclo de Processo.

Como propostas para a construção do mapa do estado futuro tem-se:

1. Inversão da cabine de pesagem da cana.
A pesagem da cana não agrega valor ao produto, mas é necessária. Nesta etapa, verificou-se que a balança fica do lado esquerdo da cabine, onde fica o operador. Isto faz com que este procedimento use mais tempo, pois o motorista do caminhão precisa deslocar-se até a cabine.
2. Automação do processo de recebimento da cana através da implantação de código de barras, RFID (*Radio Frequency Identification*) ou GPS (*Global Positioning System*), proporcionando uma pesagem mais rápida.
3. Eliminação de 2 operadores/turno para colocação de cabos de aço em caminhões que não possuem.
A fixação destes cabos nos caminhões otimizaria o processo de descarregamento.

4. Eliminação do barracão e ampliação da capacidade de recebimento da cana na mesa alimentadora.
5. Redução, de $10m^3$ para $5m^3$, do volume de caldo no tanque após a moagem da cana.
6. Redução no tempo de fermentação.
O sistema de enchimento, das dornas de fermentação, pode ser melhorado com uma maior vazão, utilizando bombas mais potentes.
7. Realização da inspeção de qualidade durante o enchimento dos tanquinhos (teste dinâmico) para a liberação do álcool e consequente eliminação de um deles.
8. Estoques em função da produção puxada (demanda).

Considerando as propostas, uma projeção para o mapa do estado futuro é apresentada na figura 4.

Figura 4 - VSM - Mapeamento do fluxo de valor do estado futuro

Com base no VSM do estado futuro e utilizando as métricas *Lean*, obtiveram-se os seguintes resultados:

1. *Lead time* para o processo produtivo do álcool hidratado: $L/T = 127$ horas;
2. Tempo das etapas que agregam valor: $TAV = 5,42$ horas;
3. Tempo das etapas que não agregam valor: $TNAV = 121,58$ horas;
4. Eficiência do Ciclo de Processo: $PCE = 4,27\%$.

Os resultados mostram uma redução significativa no *Lead time* de 185 horas, no mapa do estado atual, para 127 horas, no mapa do estado futuro. Esta redução é devida a diminuição do tempo das atividades que não agregam valor.

4.1.1 Extração do caldo

O processo da extração do caldo é composto por 9 etapas, sendo que 6 etapas agregam valor e 3 não agregam valor. Considerando todas as etapas deste processo, identificaram-se 35 variáveis que foram classificadas da seguinte forma:

- a) 07 variáveis controláveis;
- b) 04 variáveis controláveis não medidas pela usina;
- c) 24 variáveis padrão.

A figura 5 mostra o mapa do processo da extração do caldo (PMAP), em que são identificadas as variáveis de entrada e saída, em cada fase do processo, assim como, as etapas que agregam e não agregam valor ao produto.

Figura 5 - PMAP - Extração do caldo

4.1.2 Tratamento do caldo

O processo do tratamento do caldo é composto por 9 etapas, sendo que 7 etapas agregam valor e 2 não agregam valor. Considerando todas as etapas deste processo, identificaram-se 40 variáveis, que foram classificadas da seguinte forma:

- a) 09 variáveis controláveis;
- b) 11 variáveis controláveis não medidas pela usina;
- c) 19 variáveis padrão;
- d) Uma variável ruído.

A figura 6 mostra o mapa do processo do tratamento do caldo, as variáveis de entrada e saída e as etapas que agregam e não agregam valor ao produto.

Figura 6 - PMAP - Tratamento do caldo

4.1.3 Fermentação

O processo da fermentação é composto por 6 etapas, sendo que 4 etapas agregam valor e 2 não agregam valor. Considerando todas as etapas deste processo, identificaram-se 30 variáveis que foram classificadas da seguinte forma:

- a) 14 variáveis controláveis;
- b) 04 variáveis controláveis não medidas pela usina;
- c) 12 variáveis padrão.

A figura 7 mostra o mapa do processo da fermentação, as variáveis de entrada e saída e as etapas que agregam e não agregam valor ao produto.

Figura 7 - PMAP - Fermentação

4.1.4 Destilação

O processo da destilação é composto por 5 etapas, sendo que 3 etapas agregam valor e 2 não agregam valor. Considerando todas as etapas deste processo, identificaram-se 23 variáveis que foram classificadas da seguinte forma:

- a) 07 variáveis controláveis;
- b) 06 variáveis controláveis não medidas pela usina;
- c) 10 variáveis padrão.

A figura 8 mostra o mapa do processo de destilação, as variáveis de entrada e saída e as etapas que agregam e não agregam valor ao álcool hidratado.

Figura 8 - PMAP - Destilação para o álcool hidratado

4.2 Regressão linear múltipla para o processo da extração do caldo

Inicialmente, seguindo o modelo (2), foi utilizado o modelo de regressão linear múltipla para o processo da extração do caldo, considerando a variável resposta Y_{i1} (caldo misto extraído da moenda) e 7 variáveis regressoras apresentadas em (3.1.1.1), dado por:

$$y_{i1} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i \quad (24)$$

4.2.1 Métodos de seleção de variáveis

Utilizando os métodos de seleção de variáveis obtiveram-se os seguintes resultados:

- i) Todas as regressões possíveis: neste caso, utilizou-se o valor de $\bar{R}^2$ e a estatística C_p de Mallows como critérios para seleção de variáveis do modelo. O modelo indicado quando considerou-se o maior valor de $\bar{R}^2$ foi:

$$y_{i1} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i, \quad (25)$$

com $\bar{R}^2 = 0,7880$. Considerando-se o valor da estatística de Mallows, o modelo indicado foi:

$$y_{i1} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i, \quad (26)$$

com $C_p = 4,63$ e $\bar{R}^2 = 0,7849$.

- ii) Método do passo a frente passo atrás (*stepwise*): para este método, utilizou-se o critério de informação de Akaike (AIC) para escolha do modelo. O modelo (25) apresentou o menor valor de AIC (954,72) em comparação com o modelo (24), que apresentou $AIC = 958,11$. Os mesmos resultados foram obtidos utilizando-se os métodos do passo a atrás (*backward*) e passo a frente (*forward*).

Os modelos obtidos pelos métodos de seleção de variáveis foram comparados através do teste F usado como critério na seleção de variáveis. Os resultados obtidos são apresentados na tabela 4.

Tabela 4 - Tabela de análise da variância para comparação dos modelos para a variável (Y_{i1}) do processo da extração do caldo

Modelo	GL Res	SQRes	Hipóteses	Diferença na SQ	F	Pr>F
(25)	75	9184270				
(26)	76	9444702	(25) vs (26)	-260432	2,1267	0,1489
Modelo (25): $y_{i1} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$						
Modelo (26): $y_{i1} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$						

Observa-se, na tabela 4, que o teste não foi significativo. Assim, optou-se pelo modelo (26) com o menor número de parâmetros.

4.2.2 Análise de Resíduos

Inicialmente, verificou-se a qualidade do ajuste para o modelo (26) através da análise dos resíduos. Na figura 9 apresenta-se o gráfico dos valores ajustados e resíduos estudatizados externamente.

Figura 9 - Gráfico dos valores ajustados e resíduos estudatizados externamente

Observa-se, na figura 9, uma distribuição aleatória dos resíduos. Assim, visualmente, não há evidências de que os resíduos não sejam homogêneos. A figura 10 apresenta o gráfico normal de probabilidades ($Q-Q plot$) dos resíduos estudatizados externamente.

Figura 10 - Gráfico normal de probabilidades

Observa-se no gráfico normal de probabilidades, uma dispersão nos extremos, e uma maior concentração dos dados na área central, área de maior probabilidade de ocorrência. Segundo Montgomery e Peck (1992) este tipo de distribuição é chamada de distribuição de cauda leve. A hipótese de normalidade dos resíduos, verificada através do teste de *Shapiro-Wilk*, não foi rejeitada ($p = 0,366$). Assim, não existem evidências para se afirmar que a distribuição dos resíduos não seja normal. O histograma dos resíduos estudentizados externamente pode ser observado na figura 11.

Figura 11 - Histograma dos resíduos estudentizados externamente

A figura 12 apresenta o gráfico *box plot* dos resíduos estudentizados externamente.

Figura 12 - *Box plot* dos resíduos estudentizados externamente

No gráfico *box plot* as observações de números 4 e 79 aparecem como discrepantes.

A figura 13 apresenta o gráfico dos valores absolutos de DFFitS, com limite de $3\sqrt{p/n - p} = 0,7695$.

Figura 13 - Gráfico dos valores absolutos de DFFitS

Na figura 13 pode-se ver que as observações de números 23 e 35 são potencialmente influentes sobre o ajuste do modelo. Utilizando as medidas padrões de influência, do

software R, foram obtidas as observações potencialmente influentes apresentadas na tabela 5.

Tabela 5 - Estatísticas para diagnósticos em medidas de influência

Obs	$dfb.1$	$dfb.x_1$	$dfb.x_2$	$dfb.x_6$	$dfb.x_7$	$dfbit$	$cov.r$	$D_{(i)}$	h_i	$ree_{(i)}$
4	-0,28	-0,27	0,35	-0,18	0,04	-0,76	0,80*	0,11	0,09	-2,45
9	-0,07	0,08	0,03	-0,05	0,06	-0,22	1,20*	0,01	0,13*	-0,58
17	-0,08	0,01	0,09	0,12	0,09	-0,26	1,20*	0,01	0,13*	-0,65
23	-0,29	1,06*	0,72	-0,13	-1,38*	-1,41*	1,13	0,38	0,30*	-2,15
35	-0,22	0,49	0,30	-0,40	-0,85	-1,01*	1,22*	0,20	0,27*	-1,66
64	0,24	-0,35	-0,31	-0,57	0,29	0,75	0,94	0,11	0,12*	2,01
76	-0,26	0,00	-0,05	-0,15	0,25	-0,47	0,80*	0,04	0,04	-2,27
79	-0,27	0,23	0,14	0,33	-0,05	-0,55	0,79*	0,06	0,05	-2,37
81	0,01	-0,04	-0,01	0,00	0,01	0,05	1,32*	0,00	0,19*	0,10

Observações com * nas estatísticas são considerados potencialmente influentes

Na tabela 5 considerou-se uma observação influente quando pelo menos um dos casos a seguir ocorreu:

1. $|dfb.x_i| > 1$;
2. $|dfbit| > 3\sqrt{p/(n-p)} = 0,7695$;
3. $|1 - cov.r| > 3p/(n-p) = 0,197$;
4. $D_{(i)} > F_{[50\%;p;n-p]} = 0,88$;
5. $h_i \geq \frac{2p}{n} = 0,12$.

O valor do i-ésimo percentil da distribuição t com $n - p - 1 = 75$ graus de liberdade e nível de significância igual a 95% é dado por:

$$t_{\{1-(\frac{\gamma}{2n});n-p-1\}} = t_{\{1-(\frac{0,95}{162});75\}} = 2,5834. \quad (27)$$

Como, $|ree_{(i)}| \leq 2,5834$ ($i = 4, 9, 17, 23, 35, 64, 76, 79$ e 81), não há observações inconsistentes.

As observações de números 9, 17, 23, 35, 64 e 81 foram consideradas como pontos de alavanca, porém, foram classificadas como “bons”, pois não são inconsistentes.

4.2.3 Estimação dos parâmetros e análise de variância

Após a verificação das suposições para o modelo de regressão (26), realizou-se o ajuste pelo método dos quadrados mínimos. As estimativas dos coeficientes da regressão e os, respectivos, erros padrões são apresentados na tabela 6.

Tabela 6 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões

Parâmetro	Estimativa	Erro padrão	valor t	$Pr > t $
α	$2,015 \times 10^{-13}$	39,17	$5,15 \times 10^{-15}$	1,0000
β_1	0,5452	0,07830	6,962	$< 0,0001$
β_2	0,8742	0,1330	6,572	$< 0,0001$
β_6	-111,5	61,84	-1,803	0,0754
β_7	0,9757	0,2215	4,404	$< 0,0001$

Observa-se, na tabela 6, que $\hat{\alpha} \approx 0$. Isto ocorreu devido ao fato das observações estarem centradas. O teste t para o parâmetro β_6 não foi significativo. Então, realizou-se a análise de variância da regressão, considerando-se as variáveis significativas (X_{i1} , X_{i2} e X_{i7}) e verificou se a inclusão da variável X_{i6} é significativa no modelo. Os resultados são apresentados na tabela 7.

Tabela 7 - Análise de variância da regressão

Causas de variação	GL	SQ	QM	F	$Pr > F$
Regressão	3	36373140	12124380	97,56	$< 0,0001$
$\beta_6/\beta_1, \beta_2, \beta_7$	1	403938	403938	3,25	0,0754
Resíduo	76	9444702	124272		
Total	80	46221780			

Observa-se, na tabela 7, que o teste F para o parâmetro $\beta_6/\beta_1, \beta_2, \beta_7$ não foi significativo. Assim, a variável resposta Y_{i1} (caldo misto extraído da moenda) está em função das variáveis regressoras X_{i1} (cana inteira que entra na moenda), X_{i2} (cana picada que entra na moenda) e X_{i7} (volume da água de embebição).

A equação da reta de regressão ajustada pelo modelo

$$y_{i1} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_7 x_{i7} + \varepsilon_i, \quad (28)$$

é dada por:

$$\hat{y}_{i1} = 0,531x_{i1} + 0,8674x_{i2} + 0,9813x_{i7}, \quad (29)$$

com coeficiente de determinação ajustado $\bar{R}^2 = 0,7786$.

A figura 14 apresenta o gráfico dos valores observados e os valores ajustados pelo modelo (29).

Figura 14 - Gráfico dos valores observados e valores ajustados pelo modelo (29)

4.3 Regressão linear múltipla para o processo do tratamento do caldo

O modelo de regressão linear múltipla, para o processo do tratamento do caldo, foi ajustado considerando-se as variáveis respostas Y_{i1} , Y_{i2} , Y_{i4} e Y_{i5} e 9 variáveis regressoras apresentadas em (3.1.1.2).

4.3.1 Regressão linear múltipla para a variável Y_{i1} (brix do caldo caleado)

O modelo matemático utilizado, de acordo com o modelo (2), foi o seguinte:

$$y_{i1} = \alpha + \beta_1x_{i1} + \beta_2x_{i2} + \beta_3x_{i3} + \beta_4x_{i4} + \beta_5x_{i5} + \beta_6x_{i6} + \beta_7x_{i7} + \beta_8x_{i8} + \beta_9x_{i9} + \varepsilon_i \quad (30)$$

4.3.1.1 Métodos de seleção de variáveis

Utilizando os métodos de seleção de variáveis obtiveram-se os seguintes resultados:

i) O modelo indicado quando considerou-se o maior valor de $\bar{R}^2$ foi:

$$y_{i1} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_5 x_{i5} + \beta_8 x_{i8} + \varepsilon_i, \quad (31)$$

com $\bar{R}^2 = 0,9824$. Considerando-se o valor da estatística de Mallows, o modelo indicado foi:

$$y_{i1} = \alpha + \beta_1 x_{i1} + \beta_5 x_{i5} + \beta_8 x_{i8} + \varepsilon_i, \quad (32)$$

com $C_p = 3,64$ e $\bar{R}^2 = 0,9816$.

ii) Método do passo a frente passo atrás (*stepwise*): o modelo (31) apresentou o menor valor de AIC (-311,88).

Os modelos obtidos pelos métodos de seleção de variáveis foram comparados através do teste F usado como critério na seleção de variáveis. Os resultados obtidos são apresentados na tabela 8.

Tabela 8 - Tabela de análise da variância para comparação dos modelos para a variável (Y_{i1}) do processo do tratamento do caldo

Modelo	GL Res	SQRes	Hipóteses	Diferença na SQ	F	Pr>F
(31)	75	1,4857				
(32)	77	1,5867	(31) vs (32)	-0,10095	2,5479	0,085
Modelo (31): $y_{i1} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_5 x_{i5} + \beta_8 x_{i8} + \varepsilon_i$						
Modelo (32): $y_{i1} = \alpha + \beta_1 x_{i1} + \beta_5 x_{i5} + \beta_8 x_{i8} + \varepsilon_i$						

Observa-se, na tabela 8, que o teste não foi significativo. Assim, optou-se pelo modelo (32) com o menor número de parâmetros.

4.3.1.2 Análise de Resíduos

Inicialmente, verificou-se a qualidade do ajuste para o modelo (32) através da análise dos resíduos. Na figura 15 apresenta-se o gráfico dos valores ajustados e resíduos estudentizados externamente.

Figura 15 - Gráfico dos valores ajustados e resíduos estudentizados externamente

Observa-se, na figura 15, uma distribuição aleatória dos resíduos, com destaque para as observações de números 6, 8 e 22. A figura 16 apresenta o gráfico normal de probabilidades (*Q-Q plot*) dos resíduos estudentizados externamente.

Observa-se no gráfico normal de probabilidades, uma dispersão nos extremos com destaque para as observações de números 6, 8 e 22 e uma maior concentração dos dados na área central, área de maior probabilidade de ocorrência. Segundo Montgomery e Peck (1992) este tipo de distribuição é chamada de distribuição de cauda pesada. A hipótese de normalidade dos resíduos, verificada através do teste de *Shapiro-Wilk*, foi rejeitada ($p = 9,80 \times 10^{-9}$). Assim, existem evidências para se afirmar que a distribuição dos resíduos não é normal. O histograma dos resíduos estudentizados externamente pode ser observado na figura 17.

Figura 16 - Gráfico normal de probabilidades

Figura 17 - Histograma dos resíduos estudantizados externamente

A figura 18 apresenta o gráfico *box plot* dos resíduos estudantizados externamente.

Figura 18 - *Box plot* dos resíduos estudentizados externamente

No gráfico *box plot* as observações de números 4, 6, 8, 12, 22 e 49 aparecem como pontos discrepantes. Embora o teste para normalidade dos resíduos tenha sido significativo, observa-se no gráfico do logaritmo da função de verossimilhança perfilada para o modelo de transformação de *Box-Cox* (figura 19) que o intervalo de confiança para λ inclui o valor 1, indicando que não há necessidade de transformação da variável Y.

Figura 19 - Gráfico do logaritmo da função de verossimilhança perfilada para o modelo de transformação de *Box-Cox*

A figura 20 apresenta o gráfico dos valores absolutos de DFFitS, com limite de $3\sqrt{p/n - p} = 0,6838$.

Figura 20 - Gráfico dos valores absolutos de DFFitS

Na figura 20 pode-se ver que as observações de números 6, 8 e 22 são potencialmente influentes sobre o ajuste do modelo. Utilizando as medidas padrões de influência, do *software R*, foram obtidas as observações potencialmente influentes apresentadas na tabela 9.

Na tabela 9 considerou-se uma observação influente quando pelo menos um dos casos a seguir ocorreu.

1. $|dfb.x_i| > 1$;
2. $|dffit| > 3\sqrt{p/(n-p)} = 0,6838$;
3. $|1 - cov.r| > 3p/(n-p) = 0,156$;
4. $D_{(i)} > F_{[50\%;p;n-p]} = 0,85$;
5. $h_i \geq \frac{2p}{n} = 0,10$.

O valor do i -ésimo percentil da distribuição t com $n - p - 1 = 76$ graus de liberdade e nível de significância igual a 95% é dado por:

$$t_{\{1-(\frac{\gamma}{2n});n-p-1\}} = t_{\{1-(\frac{0,95}{162});76\}} = 2,5826. \quad (33)$$

As observações 6, 8 e 22 são inconsistentes, pois, $|ree_{(i)}| \geq 2,5826$ ($i = 6, 8$ e 22).

Tabela 9 - Estatísticas para diagnósticos em medidas de influência

Obs	$dfb.1$	$dfb.x_1$	$dfb.x_5$	$dfb.x_8$	$df\hat{f}it$	$cov.r$	$D_{(i)}$	h_i	$ree_{(i)}$
1	0,18	-0,28	-0,36	-0,10	0,52	1,03	0,07	0,10*	1,55
2	-0,06	0,05	0,18	-0,05	-0,19	1,22*	0,01	0,15*	-0,47
6	-0,37	-0,12	0,31	-0,83	-0,91*	0,68*	0,18	0,07	-3,24*
7	-0,14	-0,01	0,28	-0,51	-0,54	1,23*	0,07	0,20*	-1,11
8	0,36	0,11	-1,02*	0,70	1,16*	0,77*	0,31	0,13*	3,01*
9	-0,11	0,31	0,06	-0,03	-0,34	1,14	0,03	0,12*	-0,92
11	0,05	0,02	0,04	0,13	0,16	1,17*	0,01	0,11*	0,46
17	0,00	0,00	0,00	0,00	0,00	1,22*	0,00	0,13*	-0,01
22	0,69	-0,43	0,78	-0,37	1,14*	0,22*	0,22	0,03	6,07*
72	-0,03	-0,04	0,02	0,06	-0,09	1,16*	0,00	0,10*	-0,28
79	-0,03	0,07	0,01	0,06	-0,09	1,17*	0,00	0,11*	-0,28
81	-0,05	0,13	0,00	0,01	-0,14	1,17*	0,00	0,11*	-0,40

Observações com * nas estatísticas são considerados potencialmente influentes

As observações de números 1, 2, 7, 8, 9, 11, 17, 72, 79 e 81 foram consideradas como pontos de alavanca, sendo que, somente a observação de número 8 foi classificada como “ruim”, por ser inconsistente.

Excluindo-se as observações de números 6, 8 e 22 e realizando-se uma nova seleção de variáveis para o modelo (30), obtém-se as mesmas variáveis do modelo (32). Assim, estas observações, embora sejam inconsistentes, não estão influenciando na seleção de variáveis para o modelo.

4.3.1.3 Estimação dos parâmetros e análise de variância

Após a verificação das suposições para o modelo de regressão (32), realizou-se o ajuste pelo método dos quadrados mínimos. As estimativas dos coeficientes da regressão e os, respectivos, erros padrões são apresentados na tabela 10.

Observa-se, na tabela 10, que os testes foram significativos para os parâmetros β_1 , β_5 e β_8 . Assim, a variável resposta Y_{i1} (brix do caldo caleado) está em função das variáveis regressoras X_{i1} (caldo misto extraído da moenda), X_{i5} (pureza do caldo misto) e X_{i8} (leitura do sacarímetro).

Tabela 10 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões

Parâmetro	Estimativa	Erro padrão	valor t	$Pr > t $
α	$-2,97 \times 10^{-16}$	0,0159500	$-1,86 \times 10^{-14}$	1,0000
β_1	-0,000044	0,0000215	-2,050	0,0438
β_5	-0,0960	0,0166100	-5,779	< 0,0001
β_8	0,2577	0,0041220	62,531	< 0,0001

A equação da reta de regressão ajustada é dada por:

$$\hat{y}_{i1} = -0,000044x_{i1} - 0,096x_{i5} + 0,2577x_{i8}, \quad (34)$$

com coeficiente de determinação ajustado $\bar{R}^2 = 0,9816$.

A figura 21 apresenta o gráfico dos valores observados e os valores ajustados pelo modelo (34).

Figura 21 - Gráfico dos valores observados e valores ajustados pelo modelo (34)

A análise de variância da regressão é apresentada na tabela 11.

O teste F para a regressão (tabela 11) foi significativo.

Tabela 11 - Análise de variância da regressão

Causas de variação	GL	SQ	QM	F	$Pr > F$
Regressão	3	88,21	29,40	1400	$< 0,0001$
Resíduo	77	1,587	0,021		
Total	80	89,797			

4.3.2 Regressão linear múltipla para a variável Y_{i2} (Pol do caldo clarificado)

O modelo matemático utilizado, de acordo com o modelo (2), foi o seguinte:

$$y_{i2} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \beta_9 x_{i9} + \varepsilon_i \quad (35)$$

4.3.2.1 Métodos de seleção de variáveis

Utilizando os métodos de seleção de variáveis obtiveram-se os seguintes resultados:

i) O modelo indicado quando considerou-se o maior valor de $\bar{R}^2$ foi:

$$y_{i2} = \alpha + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \beta_9 x_{i9} + \varepsilon_i, \quad (36)$$

com $\bar{R}^2 = 0,999877$. Considerando-se o valor da estatística de Mallows, o modelo indicado foi:

$$y_{i2} = \alpha + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \beta_9 x_{i9} + \varepsilon_i \quad (37)$$

com $C_p = 5,972$ e $\bar{R}^2 = 0,999873$.

ii) Método do passo a frente passo atrás (*stepwise*): o modelo (36) apresentou o menor valor de AIC (-727,78).

Os modelos obtidos pelos métodos de seleção de variáveis foram comparados com o modelo

$$y_{i2} = \alpha + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \varepsilon_i \quad (38)$$

através do teste F usado como critério na seleção de variáveis. Os resultados obtidos são apresentados na tabela 12.

Observa-se, na tabela 12, que os testes não foram significativos. Assim, optou-se pelo modelo (38) com o menor número de parâmetros.

Tabela 12 - Tabela de análise da variância para comparação dos modelos para a variável (Y_{i2}) do processo do tratamento do caldo

Modelo	GL Res	SQRes	Hipóteses	Diferença na SQ	F	Pr>F
(36)	73	0,0083287				
(37)	75	0,0087932	(36)vs(37)	-0,00046448	2,0355	0,1380
(38)	76	0,0090197	(37)vs(38)	-0,00022658	1,9860	0,1630

Modelo (36): $y_{i2} = \alpha + \beta_3x_{i3} + \beta_4x_{i4} + \beta_5x_{i5} + \beta_6x_{i6} + \beta_7x_{i7} + \beta_8x_{i8} + \beta_9x_{i9} + \varepsilon_i$

Modelo (37): $y_{i2} = \alpha + \beta_5x_{i5} + \beta_6x_{i6} + \beta_7x_{i7} + \beta_8x_{i8} + \beta_9x_{i9} + \varepsilon_i$

Modelo (38): $y_{i2} = \alpha + \beta_5x_{i5} + \beta_6x_{i6} + \beta_7x_{i7} + \beta_8x_{i8} + \varepsilon_i$

4.3.2.2 Análise de Resíduos

Inicialmente, verificou-se a qualidade do ajuste para o modelo (38) através da análise dos resíduos. Na figura 22 apresenta-se o gráfico dos valores ajustados e resíduos estudentizados externamente.

Figura 22 - Gráfico dos valores ajustados e resíduos estudentizados externamente

Observa-se, na figura 22, uma distribuição aleatória dos resíduos, com destaque para as observações de números 10, 11 e 22. A figura 23 apresenta o gráfico normal de probabilidades ($Q-Q$ plot) dos resíduos estudentizados externamente.

Observa-se, no gráfico normal de probabilidades, uma dispersão nos extremos com destaque para as observações de números 10, 11 e 22. A hipótese de normalidade dos

Figura 23 - Gráfico normal de probabilidades

resíduos, verificada através do teste de *Shapiro-Wilk*, foi rejeitada ($p = 6,76 \times 10^{-7}$). Assim, existem evidências para se afirmar que a distribuição dos resíduos não é normal.

O histograma dos resíduos estudentizados externamente pode ser observado na figura 24.

Figura 24 - Histograma dos resíduos estudentizados externamente

A figura 25 apresenta o gráfico *box plot* dos resíduos estudentizados externamente.

Figura 25 - *Box plot* dos resíduos estudentizados externamente

No gráfico *box plot* as observações de números 7, 10, 11, 22, 72 e 73 aparecem como pontos discrepantes. Embora o teste para normalidade dos resíduos tenha sido significativo, observa-se no gráfico do logaritmo da função de verossimilhança perfilada para o modelo de transformação de *Box-Cox* (figura 26) que o intervalo de confiança para λ inclui o valor 1, indicando que não há necessidade de transformação da variável Y.

Figura 26 - Gráfico do logaritmo da função de verossimilhança perfilada para o modelo de transformação de *Box-Cox*

A figura 27 apresenta o gráfico dos valores absolutos de DFFitS, com limite de $3\sqrt{p/n - p} = 0,7695$.

Figura 27 - Gráfico dos valores absolutos de DFFitS

Na figura 27 pode-se ver que as observações de números 2, 7, 11 e 22 são potencialmente influentes sobre o ajuste do modelo. Utilizando as medidas padrões de influência, do *software R*, foram obtidas as observações potencialmente influentes apresentadas na tabela 13.

Tabela 13 - Estatísticas para diagnósticos em medidas de influência

Obs	$dfb.1$	$dfb.x_5$	$dfb.x_6$	$dfb.x_7$	$dfb.x_8$	$dffit$	$cov.r$	$D_{(i)}$	h_i	$ree_{(i)}$
2	0,24	-1,00	0,66	-0,13	-0,01	1,07*	1,11	0,22	0,24*	1,91
7	-0,29	0,34	0,46	0,06	-1,16*	-1,23*	0,99	0,29	0,23*	-2,28
8	-0,16	0,37	-0,17	-0,35	-0,17	-0,63	1,20*	0,08	0,20*	-1,27
11	-0,47	-0,59	0,28	-0,63	-1,06*	-1,60*	0,49*	0,43	0,14*	-3,94*
17	0,04	0,08	-0,02	0,08	-0,05	0,12	1,21*	0,00	0,12*	0,31
22	-0,56	-0,27	-0,85	-0,18	0,56	-1,21*	0,28*	0,23	0,06	-4,90*
51	-0,03	0,02	0,06	0,08	-0,05	-0,10	1,21*	0,00	0,12*	-0,28
60	0,15	0,04	0,06	0,45	-0,15	0,50	1,10	0,05	0,13*	0,13
62	-0,02	0,00	0,01	-0,03	-0,03	-0,05	1,21*	0,00	0,12*	-0,14

Observações com * nas estatísticas são considerados potencialmente influentes

Na tabela 13 considerou-se uma observação influente quando pelo menos um dos casos a seguir ocorreu:

1. $|dfb.x_i| > 1$;

2. $|dffit| > 3\sqrt{p/(n-p)} = 0,7695;$
3. $|1 - cov.r| > 3p/(n-p) = 0,197;$
4. $D_{(i)} > F_{[50\%;p;n-p]} = 0,88;$
5. $h_i \geq \frac{2p}{n} = 0,12.$

O valor do i-ésimo percentil da distribuição t com $n - p - 1 = 75$ graus de liberdade e nível de significância igual a 95% é dado por:

$$t_{\{1-(\frac{\gamma}{2n});n-p-1\}} = t_{\{1-(\frac{0,95}{162});75\}} = 2,5834. \quad (39)$$

As observações 11 e 22 são inconsistentes, pois, $|ree_{(i)}| \geq 2,5834$ ($i = 11$ e 22).

As observações de números 2, 7, 8, 11, 17, 51, 60 e 62 foram consideradas como pontos de alavanca, sendo que, somente a observação de número 11 foi classificada como “ruim”, por ser inconsistente.

Excluindo-se as observações de números 11 e 22 e realizando-se uma nova seleção de variáveis para o modelo (35), obtém-se as mesmas variáveis do modelo (38). Assim, estas observações, embora sejam inconsistentes, não estão influenciando na seleção de variáveis para o modelo.

4.3.2.3 Estimação dos parâmetros e análise de variância

Após a verificação das suposições para o modelo de regressão (38), realizou-se o ajuste pelo método dos quadrados mínimos. As estimativas dos coeficientes da regressão e os, respectivos, erros padrões são apresentados na tabela 14.

Tabela 14 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões

Parâmetro	Estimativa	Erro padrão	valor t	$Pr > t $
α	$7,002 \times 10^{-16}$	0,0012100	$5,78 \times 10^{-13}$	1,0000
β_5	0,003177	0,0014300	2,221	0,0293
β_6	0,002912	0,0017720	1,644	0,1044
β_7	-0,05373	0,0215800	-2,490	0,0150
β_8	0,2314	0,0003386	683,566	< 0,0001

O teste t para o parâmetro β_6 não foi significativo. Então, realizou-se a análise de variância da regressão, considerando-se as variáveis significativas (X_{i5} , X_{i7} e X_{i8}) e verificou

se a inclusão da variável X_{i6} é significativa no modelo. Os resultados são apresentados na tabela 15.

Tabela 15 - Análise de variância da regressão

Causas de variação	GL	SQ	QM	F	$Pr > F$
Regressão	3	74,25768	24,75	206250	$< 0,0001$
$\beta_6/\beta_5, \beta_7, \beta_8$	1	0,00032	0,00032	2,67	0,1064
Resíduo	76	0,009	0,00012		
Total	80	74,267			

Observa-se, na tabela 15, que o teste para o parâmetro $\beta_6/\beta_5, \beta_7, \beta_8$ não foi significativo. Assim, a variável resposta Y_{i2} (pol do caldo clarificado) está em função das variáveis regressoras X_{i5} (pureza do caldo misto), X_{i7} (AR do caldo misto) e X_{i8} (leitura do sacarímetro).

A equação da reta de regressão ajustada pelo modelo

$$y_{i2} = \alpha + \beta_5 x_{i5} + \beta_7 x_{i7} + \beta_8 x_{i8} + \varepsilon_i \quad (40)$$

é dada por:

$$\hat{y}_{i2} = 0,0041x_{i5} - 0,059x_{i7} + 0,232x_{i8}, \quad (41)$$

com coeficiente de determinação ajustado $\bar{R}^2 = 0,9999$.

A figura 28 apresenta o gráfico dos valores observados e os valores ajustados pelo modelo (41).

Figura 28 - Gráfico dos valores observados e valores ajustados pelo modelo (41)

4.3.3 Regressão linear múltipla para a variável Y_{i4} (AR decantado)

Para a variável Y_{i4} utilizou-se o modelo matemático na forma não centrada, de acordo com o modelo (1). O modelo matemático utilizado foi o seguinte:

$$Y_{i4} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \beta_9 X_{i9} + \varepsilon_i \quad (42)$$

A figura 29 apresenta o gráfico dos valores observados e valores ajustados pelo modelo (42).

Figura 29 - Gráfico dos valores observados e valores ajustados pelo modelo (42)

Observa-se, na figura 29, uma relação não linear para o modelo ajustado.

Através dos gráficos dos valores observados para as variáveis X 's e Y_4 , apresentados na figura 30, observou-se uma relação não linear entre as variáveis X_9 e Y_4 .

Com a finalidade de contornar este problema, introduziu-se a variável $X_{i10} = \sqrt{X_{i9}}$ no modelo (42) e realizou-se a seleção de variáveis para o modelo:

$$Y_{i4} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \beta_9 X_{i9} + \beta_{10} X_{i10} + \varepsilon_i \quad (43)$$

Figura 30 - Gráfico dos valores observados para as variáveis X_i 's e Y_4

4.3.3.1 Métodos de seleção de variáveis

Utilizando os métodos de seleção de variáveis obtiveram-se os seguintes resultados:

- i) O modelo indicado quando considerou-se o maior valor de $\bar{R}^2$ foi:

$$Y_{i4} = \beta_0 + \beta_4 X_{i4} + \beta_8 X_{i8} + \beta_9 X_{i9} + \beta_{10} X_{i10} + \varepsilon_i, \quad (44)$$

com $\bar{R}^2 = 0,9937$. Considerando-se o valor da estatística de Mallows, obteve-se o mesmo modelo (44), com $C_p = 1,36$.

- ii) Método do passo a frente passo atrás (*stepwise*): o modelo

$$Y_{i4} = \beta_0 + \beta_3 X_{i3} + \beta_8 X_{i8} + \beta_9 X_{i9} + \beta_{10} X_{i10} + \varepsilon_i \quad (45)$$

apresentou o menor valor de AIC (-863,99), com $\bar{R}^2 = 0,9936$.

Verificou-se a existência de uma alta correlação entre as variáveis X_{i2} , X_{i3} e X_{i4} . Assim, a inclusão de uma delas, no modelo, é suficiente para o ajuste. Então, optou-se pelo modelo (44).

4.3.3.2 Análise de Resíduos

Utilizando o modelo (44), verificou-se a qualidade do ajuste através da análise dos resíduos. Na figura 31 apresenta-se o gráfico dos valores ajustados e resíduos estudentizados externamente. Observa-se, na figura 31, uma distribuição aleatória dos resíduos, com

Figura 31 - Gráfico dos valores ajustados e resíduos estudantizados externamente

destaque para as observações de números 54 e 60.

A figura 32 apresenta o gráfico normal de probabilidades (*Q-Q plot*) dos resíduos estudantizados externamente.

Figura 32 - Gráfico normal de probabilidades

Observa-se no gráfico normal de probabilidades, uma dispersão nos extremos com destaque para as observações de números 54 e 60. A hipótese de normalidade dos resíduos, verificada através do teste de *Shapiro-Wilk*, não foi rejeitada ($p = 0,9327$). Assim, não existem evidências para se afirmar que a distribuição dos resíduos não seja normal.

O histograma dos resíduos estudantizados externamente pode ser observado na

figura 33.

Figura 33 - Histograma dos resíduos estudentizados externamente

A figura 34 apresenta o gráfico *box plot* dos resíduos estudentizados externamente.

Figura 34 - *Box plot* dos resíduos estudentizados externamente

A figura 35 apresenta o gráfico dos valores absolutos de DFFitS, com limite de $3\sqrt{p/n - p} = 0,7695$.

Figura 35 - Gráfico dos valores absolutos de DFFitS

Verificou-se que não há observações potencialmente influentes sobre o ajuste do modelo quando considera-se os valores absolutos de DFFitS. No entanto, utilizando as medidas padrões de influência, do *software R*, foram obtidas as observações potencialmente influentes que são apresentadas na tabela 16.

Tabela 16 - Estatísticas para diagnósticos em medidas de influência

Obs	$dfb.1$	$dfb.x_4$	$dfb.x_8$	$dfb.x_9$	$dfb.x_{10}$	$dffit$	$cov.r$	$D_{(i)}$	h_i	$ree_{(i)}$
7	0,02	-0,20	0,46	0,06	-0,05	0,50	1,32*	0,05	0,24*	0,90
11	-0,03	0,04	0,09	-0,01	0,01	0,16	1,21*	0,01	0,13*	0,42
54	0,06	-0,48	0,13	0,02	-0,01	0,73	0,71*	0,10	0,07	2,74*
56	0,23	0,00	0,03	0,24	-0,24	0,28	1,77*	0,02	0,40*	0,34
58	0,08	0,00	0,01	0,09	-0,08	0,11	1,39*	0,00	0,23*	0,20
60	0,28	0,00	0,20	0,30	-0,30	-0,47	0,72*	0,04	0,03	-2,60*
62	-0,01	-0,05	0,05	-0,01	0,01	0,06	1,22*	0,00	0,12*	0,17
64	0,00	0,02	-0,01	0,00	0,00	-0,02	1,21*	0,00	0,12*	-0,06

Observações com * nas estatísticas são considerados potencialmente influentes

Na tabela 16 considerou-se uma observação influente quando pelo menos um dos casos a seguir ocorreu:

1. $|dfb.x_i| > 1$;

$$2. |dffit| > 3\sqrt{p/(n-p)} = 0,7695;$$

$$3. |1 - cov.r| > 3p/(n-p) = 0,197;$$

$$4. D_{(i)} > F_{[50\%;p;n-p]} = 0,88;$$

$$5. h_i \geq \frac{2p}{n} = 0,12.$$

O valor do i -ésimo percentil da distribuição t com $n - p - 1 = 75$ graus de liberdade e nível de significância igual a 95% é dado por:

$$t_{\{1-(\frac{\alpha}{2n});n-p-1\}} = t_{\{1-(\frac{0,95}{162});75\}} = 2,5834. \quad (46)$$

As observações 54 e 60 são inconsistentes, pois, $|ree_{(i)}| \geq 2,5834$ ($i = 54$ e 60).

As observações de números 7, 11, 56, 58, 62 e 64 foram consideradas como pontos de alavanca, porém, foram classificadas como “bons”, pois não são inconsistentes.

Excluindo-se as observações de números 54 e 60 e realizando-se uma nova seleção de variáveis para o modelo (43), obtém-se as mesmas variáveis do modelo (44). Assim, estas observações, embora sejam inconsistentes, não estão influenciando na seleção de variáveis para o modelo.

4.3.3.3 Estimação dos parâmetros e análise de variância

Após a verificação das suposições para o modelo de regressão (44), realizou-se o ajuste pelo método dos quadrados mínimos. As estimativas dos coeficientes da regressão e os, respectivos, erros padrões são apresentados na tabela 17.

Tabela 17 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões

Parâmetro	Estimativa	Erro padrão	valor t	$Pr > t $
β_0	2,1011300	0,0838211	25,067	$< 0,0001$
β_4	-0,0019539	0,0008047	-2,428	0,0176
β_8	-0,0007799	0,0001682	-4,637	$< 0,0001$
β_9	0,0339545	0,0034117	9,952	$< 0,0001$
β_{10}	-0,4992535	0,0340777	-14,650	$< 0,0001$

Observa-se, na tabela 17, que os testes para os parâmetros $\beta_0, \beta_4, \beta_8, \beta_9$ e β_{10} foram significativos. Assim, a variável resposta Y_{i4} (AR decantado) está em função da constante β_0 e das variáveis regressoras X_{i4} (pol do caldo misto), X_{i8} (leitura do sacarímetro), X_{i9} (volume gasto de AR) e X_{i10} ($\sqrt{X_{i9}}$).

A equação da reta de regressão ajustada é dada por:

$$\hat{Y}_{i4} = 2,10 - 0,00195X_{i4} - 0,00078X_{i8} + 0,03395X_{i9} - 0,49925X_{i10}. \quad (47)$$

com coeficiente de determinação ajustado $\bar{R}^2 = 0,9937$.

A figura 36 apresenta o gráfico dos valores observados e os valores ajustados pelo modelo (47).

Figura 36 - Gráfico dos valores observados e valores ajustados pelo modelo (47)

A análise de variância da regressão é apresentada na tabela 18.

Tabela 18 - Análise de variância da regressão

Causas de variação	GL	SQ	QM	F	$Pr > F$
Regressão	4	0,274269	0,068567	3117	< 0,0001
Resíduo	76	0,001645	0,000022		
Total	80	0,275914			

O teste F para a regressão (tabela 18) foi significativo.

4.3.4 Regressão linear múltipla para a variável Y_{i5} (ART do caldo)

O modelo matemático utilizado, de acordo com o modelo (2), foi o seguinte:

$$y_{i5} = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \beta_9 x_{i9} + \varepsilon_i \quad (48)$$

4.3.4.1 Métodos de seleção de variáveis

Utilizando os métodos de seleção de variáveis obtiveram-se os seguintes resultados:

i) O modelo indicado quando considerou-se o maior valor de $\bar{R}^2$ foi:

$$y_{i5} = \alpha + \beta_5 x_{i5} + \beta_8 x_{i8} + \beta_9 x_{i9} + \varepsilon_i, \quad (49)$$

com $\bar{R}^2 = 0,99823$. Considerando-se o valor da estatística de Mallows, o modelo indicado foi:

$$y_{i5} = \alpha + \beta_8 x_{i8} + \beta_9 x_{i9} + \varepsilon_i \quad (50)$$

com $C_p = 2,90$ e $\bar{R}^2 = 0,99818$.

ii) Método do passo a frente passo atrás (*stepwise*): o modelo (49) apresentou o menor valor de AIC (-503,9).

Os modelos obtidos pelos métodos de seleção de variáveis foram comparados através do teste F usado como critério na seleção de variáveis. Os resultados obtidos são apresentados na tabela 19.

Tabela 19 - Tabela de análise da variância para comparação dos modelos para a variável (Y_{i5}) do processo do tratamento do caldo

Modelo	GL Res	SQRes	Hipóteses	Diferença na SQ	F	Pr>F
(49)	77	0,14583				
(50)	78	0,15199	(49) vs (50)	-0,006157	3,2511	0,0753
Modelo (49): $y_{i5} = \alpha + \beta_5 x_{i5} + \beta_8 x_{i8} + \beta_9 x_{i9} + \varepsilon_i$						
Modelo (50): $y_{i5} = \alpha + \beta_8 x_{i8} + \beta_9 x_{i9} + \varepsilon_i$						

Observa-se, na tabela 19, que os modelos (49) e (50) não diferem significativamente entre si. Assim, optou-se pelo modelo (50) com o menor número de parâmetros.

4.3.4.2 Análise de Resíduos

Inicialmente, verificou-se a qualidade do ajuste para o modelo (50) através da análise dos resíduos. A figura 37 apresenta o gráfico dos valores ajustados e resíduos estudentizados externamente e o gráfico normal de probabilidades (*Q-Q plot*) dos resíduos estudentizados externamente.

Figura 37 - Gráfico dos valores ajustados e resíduos estudentizados externamente e o gráfico normal de probabilidades

Observa-se, nos gráficos, que a observação de número 3 está muito distante das demais, com valor de $ree_{(3)} = 20,60$.

O valor do i -ésimo percentil da distribuição t com $n - p - 1 = 77$ graus de liberdade e nível de significância igual a 95% é dado por:

$$t_{\{1-(\frac{\gamma}{2n});n-p-1\}} = t_{\{1-(\frac{0,95}{162});77\}} = 2,5817. \quad (51)$$

Portanto, a observação número 3 é inconsistente, pois, $|ree_{(3)}| \geq 2,5817$. Devido ao valor muito elevado do resíduo estudentizado externamente, optou-se por retirar esta observação do conjunto de dados e realizar a análise sem a mesma.

4.3.4.3 Métodos de seleção de variáveis excluindo-se a observação número 3

Utilizando os métodos de seleção de variáveis obtiveram-se os seguintes resultados:

i) O modelo indicado quando considerou-se o maior valor de $\bar{R}^2$ foi:

$$y_{i5} = \alpha + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_7 x_{i7} + \beta_8 x_{i8} + \beta_9 x_{i9} + \varepsilon_i, \quad (52)$$

com $\bar{R}^2 = 0,9998$. Considerando-se o valor da estatística de Mallows, obteve-se o mesmo modelo (52), com $C_p = 4,89$.

ii) Método do passo a frente passo atrás (*stepwise*): o modelo (52) apresentou o menor valor de AIC (-663,5).

4.3.4.4 Análise de Resíduos

Inicialmente, verificou-se a qualidade do ajuste para o modelo (52) através da análise dos resíduos. Na figura 38 apresenta-se o gráfico dos valores ajustados e resíduos estudentizados externamente.

Figura 38 - Gráfico dos valores ajustados e resíduos estudentizados externamente

Observa-se, na figura 38, uma distribuição aleatória dos resíduos, com destaque para as observações de números 2, 17 e 21. A figura 39 apresenta o gráfico normal de probabilidades (*Q-Q plot*) dos resíduos estudentizados externamente.

Figura 39 - Gráfico normal de probabilidades

Observa-se, no gráfico normal de probabilidades, uma dispersão nos extremos com destaque para as observações de números 2, 10, 17 e 21. A hipótese de normalidade dos resíduos, verificada através do teste de *Shapiro-Wilk*, foi rejeitada ($p = 0,0276$). Assim, existem evidências para se afirmar que a distribuição dos resíduos não é normal.

O histograma dos resíduos estudentizados externamente pode ser observado na figura 40.

Figura 40 - Histograma dos resíduos estudentizados externamente

A figura 41 apresenta o gráfico *box plot* dos resíduos estudentizados externamente.

Figura 41 - *Box plot* dos resíduos estudentizados externamente

No gráfico *box plot* as observações de números 2, 9, 17, 21 e 55 aparecem como pontos discrepantes. Embora o teste para normalidade dos resíduos tenha sido significativo, observa-se no gráfico do logaritmo da função de verossimilhança perfilada para o modelo de transformação de *Box-Cox* (figura 42) que o intervalo de confiança para λ inclui o valor 1, indicando que não há necessidade de transformação da variável Y.

Figura 42 - Gráfico do logaritmo da função de verossimilhança perfilada para o modelo de transformação de *Box-Cox*

A figura 43 apresenta o gráfico dos valores absolutos de DFFitS, com limite de $3\sqrt{p/n - p} = 0,9290$.

Figura 43 - Gráfico dos valores absolutos de DFFitS

Na figura 43 pode-se ver que as observações de números 2, 9, 10, 21 e 55 são potencialmente influentes sobre o ajuste do modelo. Utilizando as medidas padrões de influência, do *software R*, foram obtidas as observações potencialmente influentes apresentadas na tabela 20.

Na tabela 20 considerou-se uma observação influente quando pelo menos um dos casos a seguir ocorreu:

1. $|dfb.x_i| > 1$;
2. $|dffit| > 3\sqrt{p/(n-p)} = 0,9290$;
3. $|1 - cov.r| > 3p/(n-p) = 0,29$;
4. $D_{(i)} > F_{[50\%;p;n-p]} = 0,91$;
5. $h_i \geq \frac{2p}{n} = 0,175$.

O valor do i-ésimo percentil da distribuição t com $n - p - 1 = 72$ graus de liberdade e nível de significância igual a 95% é dado por:

$$t_{\{1-(\frac{\gamma}{2n});n-p-1\}} = t_{\{1-(\frac{0,95}{160});72\}} = 2,5814. \quad (53)$$

As observações 2, 17 e 21 são inconsistentes, pois, $|ree_{(i)}| \geq 2,5814$ ($i = 2, 17$ e 21).

As observações de números 2, 6, 7, 10, 55 e 61 foram consideradas como pontos de alavanca, sendo que, somente a observação de número 2 foi classificada como “ruim”, por ser inconsistente.

Tabela 20 - Estatísticas para diagnósticos em medidas de influência

Obs	$dfb.1$	$dfb.x_3$	$dfb.x_4$	$dfb.x_5$	$dfb.x_7$	$dfb.x_8$	$dfb.x_9$	$dffit$	$cov.r$	$D_{(i)}$	h_i	$rec_{(i)}$
2	-0,35	0,62	-0,64	0,74	0,40	0,11	-0,14	-1,47*	0,72	0,28	0,23*	-2,72*
6	-0,10	-0,02	0,03	-0,02	0,03	-0,40	-0,12	-0,45	1,36*	0,03	0,24*	-0,80
7	0,00	0,01	-0,01	0,01	0,01	0,00	0,00	0,02	2,04*	0,00	0,46*	0,02
9	0,30	-0,31	0,31	-0,26	0,39	0,02	0,38	0,96*	0,70*	0,12	0,13	2,51
10	-0,29	0,83	-0,83	0,79	-0,30	-0,45	-0,28	-1,32*	0,92	0,24	0,26*	-2,25
17	-0,31	-0,26	0,26	-0,26	-0,12	0,05	-0,27	-0,55	0,57*	0,04	0,04	-2,75*
21	-0,41	0,12	-0,14	0,12	-0,01	0,67	0,03	-1,07*	0,40*	0,14	0,09	-3,50*
55	0,30	0,31	-0,32	0,37	-0,36	0,10	0,91	1,18*	0,82	0,19	0,20*	2,37
61	-0,06	-0,05	0,06	-0,06	-0,17	-0,17	0,07	-0,27	1,37*	0,01	0,21*	-0,51

Observações com * nas estatísticas são considerados potencialmente influentes

Excluindo-se as observações de números 2, 17 e 21 e realizando-se uma nova seleção de variáveis para o modelo (48), obtém-se as mesmas variáveis do modelo (52). Assim, estas observações, embora sejam inconsistentes, não estão influenciando na seleção de variáveis para o modelo.

4.3.4.5 Estimação dos parâmetros e análise de variância

Após a verificação das suposições para o modelo de regressão (52), realizou-se o ajuste pelo método dos quadrados mínimos. As estimativas dos coeficientes da regressão e os, respectivos, erros padrões são apresentados na tabela 21.

Tabela 21 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões

Parâmetro	Estimativa	Erro padrão	valor t	$Pr > t $
α	$-1,488 \times 10^{-15}$	0,001696	$-8,77 \times 10^{-13}$	1,0000
β_3	70,04	26,61	2,632	0,0103
β_4	-0,3442	0,1291	-2,667	0,0094
β_5	0,0643	0,02164	2,973	0,0040
β_7	-0,0706	0,03127	-2,257	0,0270
β_8	0,2426	0,0005544	437,644	$< 0,0001$
β_9	-0,0168	0,0005516	-30,525	$< 0,0001$

Observa-se, na tabela 21, que os testes foram significativos para os parâmetros $\beta_3, \beta_4, \beta_5, \beta_7, \beta_8$ e β_9 . Assim, a variável resposta Y_{i5} (ART do caldo) está em função das variáveis regressoras X_{i3} (densidade do caldo misto), X_{i4} (pol do caldo misto), X_{i5} (pureza do caldo misto), X_{i7} (AR do caldo misto), X_{i8} (leitura do sacarímetro) e X_{i9} (volume gasto de AR).

A equação da reta de regressão ajustada é dada por:

$$\hat{y}_{i5} = 70,04x_{i3} - 0,3442x_{i4} + 0,0643x_{i5} - 0,0706x_{i7} + 0,2426x_{i8} - 0,0168x_{i9}, \quad (54)$$

com coeficiente de determinação ajustado $\bar{R}^2 = 0,9998$.

A figura 44 apresenta o gráfico dos valores observados e os valores ajustados pelo modelo (54).

Figura 44 - Gráfico dos valores observados e valores ajustados pelo modelo (54)

A análise de variância da regressão é apresentada na tabela 22.

Tabela 22 - Análise de variância da regressão

Causas de variação	GL	SQ	QM	F	$Pr > F$
Regressão	6	83,533	13,92	60531	$< 0,0001$
Resíduo	73	0,017	0,00023		
Total	79	83,550			

O teste F para a regressão (tabela 22) foi significativo.

A variável Y_{i3} considerada em função das 9 variáveis regressoras, não produziu ajuste adequado.

4.4 Regressão linear múltipla para o processo da fermentação

As variáveis Y_{i1} e Y_{i2} consideradas, individualmente, em função das 14 variáveis regressoras, apresentadas em (3.1.1.3), não produziram ajuste adequado.

4.5 Regressão linear múltipla para o processo da destilação

O modelo de regressão linear múltipla para o processo produtivo do etanol hidratado foi ajustado, de acordo com o modelo (2), considerando-se a variável resposta Y_i e

43 variáveis regressoras apresentadas em (3.1.1.4). O modelo matemático é dado por:

$$\begin{aligned}
 y_i = & \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \beta_9 x_{i9} + \\
 & + \beta_{10} x_{i10} + \beta_{11} x_{i11} + \beta_{12} x_{i12} + \beta_{13} x_{i13} + \beta_{14} x_{i14} + \beta_{15} x_{i15} + \beta_{16} x_{i16} + \beta_{17} x_{i17} + \\
 & + \beta_{18} x_{i18} + \beta_{19} x_{i19} + \beta_{20} x_{i20} + \beta_{21} x_{i21} + \beta_{22} x_{i22} + \beta_{23} x_{i23} + \beta_{24} x_{i24} + \beta_{25} x_{i25} + \\
 & + \beta_{26} x_{i26} + \beta_{27} x_{i27} + \beta_{28} x_{i28} + \beta_{29} x_{i29} + \beta_{30} x_{i30} + \beta_{31} x_{i31} + \beta_{32} x_{i32} + \beta_{33} x_{i33} + \\
 & + \beta_{34} x_{i34} + \beta_{35} x_{i35} + \beta_{36} x_{i36} + \beta_{37} x_{i37} + \beta_{38} x_{i38} + \beta_{39} x_{i39} + \beta_{40} x_{i40} + \beta_{41} x_{i41} + \\
 & + \beta_{42} x_{i42} + \beta_{43} x_{i43} + \varepsilon_i
 \end{aligned} \tag{55}$$

4.5.1 Métodos de seleção de variáveis

Utilizando o método passo a frente passo atrás (*stepwise*), obteve-se o seguinte resultado: o modelo

$$\begin{aligned}
 y_i = & \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_4 x_{i4} + \beta_6 x_{i6} + \beta_8 x_{i8} + \beta_9 x_{i9} + \beta_{10} x_{i10} + \beta_{16} x_{i16} + \\
 & + \beta_{17} x_{i17} + \beta_{18} x_{i18} + \beta_{20} x_{i20} + \beta_{22} x_{i22} + \beta_{26} x_{i26} + \beta_{30} x_{i30} + \beta_{33} x_{i33} + \\
 & + \beta_{35} x_{i35} + \beta_{37} x_{i37} + \beta_{39} x_{i39} + \beta_{40} x_{i40} + \beta_{41} x_{i41} + \beta_{42} x_{i42} + \beta_{43} x_{i43} + \varepsilon_i
 \end{aligned} \tag{56}$$

apresentou o menor valor de AIC (1399,58) em comparação com o modelo (55), que apresentou $AIC = 1435,56$. Os mesmos resultados foram obtidos utilizando-se os métodos do passo a atrás (*backward*) e passo a frente (*forward*).

O modelo (56), obtido pelo método de seleção de variáveis, foi comparado com o modelo (55) através do teste F. Os resultados obtidos são apresentados na tabela 23.

Tabela 23 - Tabela de análise da variância para comparação dos modelos para a variável Y_i do processo da destilação

Modelo	GL Res	SQRes	Hipóteses	Diferença na SQ	F	Pr>F
(55)	37	1360341309				
(56)	58	1465275788	(55) vs (56)	-104934479	0,1359	1

Observa-se, na tabela 23, que o teste não foi significativo. Assim, optou-se pelo modelo (56) com o menor número de parâmetros.

4.5.2 Análise de Resíduos

Verificou-se a qualidade do ajuste para o modelo (56) através da análise dos resíduos. Na figura 45 apresenta-se o gráfico dos valores ajustados e resíduos estudentizados externamente.

Figura 45 - Gráfico dos valores ajustados e resíduos estudantizados externamente

Observa-se, na figura 45, uma distribuição aleatória dos resíduos. A figura 46 apresenta o gráfico normal de probabilidades (*Q-Q plot*) dos resíduos estudantizados externamente.

Figura 46 - Gráfico normal de probabilidades

Observa-se, na figura 46, uma dispersão nos extremos com destaque para as observações de números 59, 67 e 68. A hipótese de normalidade dos resíduos, verificada através do teste de *Shapiro-Wilk*, não foi rejeitada ($p = 0,4235$). Assim, não existem evidências para se afirmar que a distribuição dos resíduos não seja normal.

O histograma dos resíduos estudentizados externamente pode ser observado na figura 47.

Figura 47 - Histograma dos resíduos estudentizados externamente

A figura 48 apresenta o gráfico *box plot* dos resíduos estudentizados externamente.

Figura 48 - *Box plot* dos resíduos estudentizados externamente

A figura 49 apresenta o gráfico dos valores absolutos de DFFitS, com limite de $3\sqrt{p/n - p} = 1,89$.

Figura 49 - Gráfico dos valores absolutos de DFFitS

Na figura 49 pode-se ver que a observação de número 3 é potencialmente influente sobre o ajuste do modelo. Utilizando as medidas padrões de influência, do *software R*, foram obtidas as observações potencialmente influentes apresentadas na tabela 24.

Tabela 24 - Estatísticas para diagnósticos em medidas de influência

Obs	(conclusão)								
	$dfb.x_{40}$	$dfb.x_{41}$	$dfb.x_{42}$	$dfb.x_{43}$	$dffit$	$cov.r$	$D_{(i)}$	h_i	$ree_{(i)}$
2	-0,30	-0,16	-0,01	-0,15	-1,03	2,35*	0,05	0,55	-0,93
3	-0,16	0,01	0,11	0,07	-2,46*	16,70*	0,26	0,93*	-0,69
4	0,01	0,07	0,08	0,06	-0,85	2,60*	0,03	0,55	-0,77
7	0,04	-0,01	0,00	0,02	-0,37	2,28*	0,01	0,40	-0,45
8	0,05	0,00	-0,05	-0,04	-0,34	2,41*	0,00	0,42	-0,40
10	0,11	0,10	-0,05	-0,04	-0,25	2,28*	0,00	0,37	-0,33
11	0,19	0,14	-0,04	0,01	0,39	2,28*	0,01	0,40	0,48
22	-0,01	0,03	-0,01	-0,02	0,28	2,45*	0,00	0,42	0,33
41	0,05	0,05	0,08	0,13	-0,37	2,23*	0,01	0,39	-0,46
66	0,00	0,00	0,00	0,00	0,03	2,57*	0,00	0,42	0,03
75	0,00	0,00	0,00	0,00	0,01	2,24*	0,00	0,33	0,01

Observações com * nas estatísticas são considerados potencialmente influentes

Na tabela 24 considerou-se uma observação influente quando pelo menos um dos casos a seguir ocorreu:

1. $|dfb.x_i| > 1$;
2. $|dffit| > 3\sqrt{p/(n-p)} = 1,89$;
3. $|1 - cov.r| > 3p/(n-p) = 1,19$;
4. $D_{(i)} > F_{[50\%;p;n-p]} = 0,98$;
5. $h_i \geq \frac{2p}{n} = 0,57$.

O valor do i-ésimo percentil da distribuição t com $n - p - 1 = 57$ graus de liberdade e nível de significância igual a 95% é dado por:

$$t_{\{1-(\frac{\gamma}{2n});n-p-1\}} = t_{\{1-(\frac{0,95}{162});57\}} = 2,6040. \quad (57)$$

Como, $|ree_{(i)}| \leq 2,6040$ ($i = 2, 3, 4, 7, 8, 10, 11, 22, 41, 66$ e 75), não há observações inconsistentes.

A observação de número 3 foi considerada como ponto de alavanca, porém, foi classificada como “bom”, por não ser inconsistente.

4.5.2.1 Estimação dos parâmetros e análise de variância

Após a verificação das suposições para o modelo de regressão (56), realizou-se o ajuste pelo método dos quadrados mínimos. As estimativas dos coeficientes da regressão e os, respectivos, erros padrões são apresentados na tabela 25.

Os testes t para os parâmetros $\beta_4, \beta_{10}, \beta_{17}, \beta_{18}, \beta_{22}$ e β_{35} , não foram significativos. Então, realizou-se a análise de variância da regressão para o modelo

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_6 x_{i6} + \beta_8 x_{i8} + \beta_9 x_{i9} + \beta_{16} x_{i16} + \beta_{20} x_{i20} + \beta_{26} x_{i26} + \beta_{30} x_{i30} + \beta_{33} x_{i33} + \beta_{37} x_{i37} + \beta_{39} x_{i39} + \beta_{40} x_{i40} + \beta_{41} x_{i41} + \beta_{42} x_{i42} + \beta_{43} x_{i43} + \varepsilon_i \quad (58)$$

e verificou-se, individualmente, se as variáveis $X_{i4}, X_{i10}, X_{i17}, X_{i18}, X_{i22}$ e X_{i35} , dado as variáveis que estão no modelo (58), são significativas. Os resultados são apresentados na tabela 26.

Observa-se, na tabela 26, que os testes não foram significativos. Assim, a variável resposta Y_i (etanol hidratado) está em função das variáveis regressoras X_{i1} (cana inteira que entra na moenda), X_{i2} (cana picada que entra na moenda), X_{i6} (pol do caldo da cana desfibrada), X_{i8} (brix do caldo misto), X_{i9} (densidade do caldo misto), X_{i16} (volume gasto de AR), X_{i20} (ART do caldo), X_{i26} (GL nas dornas), X_{i30} (% levedo no creme), X_{i33} (brix do leite tratado), X_{i37} (volume da vinhaça), X_{i39} (% etanol na vinhaça), X_{i40} (ART do caldo destilaria), X_{i41} (ART do mel final), X_{i42} (rendimento da destilaria) e X_{i43} (litros de álcool por tonelada de mel).

A equação da reta de regressão ajustada pelo modelo (58) é dada por:

$$\begin{aligned} \hat{y}_i = & 2,844x_{i1} + 6,59x_{i2} - 3107x_{i6} - 59110x_{i8} + 14640000x_{i9} - 576,5x_{i16} - 1727x_{i20} + \\ & + 44190x_{i26} + 370,2x_{i30} - 3604x_{i33} + 92,42x_{i37} + 86170x_{i39} + 36,83x_{i40} + \\ & + 35,74x_{i41} + 46290x_{i42} + 2066x_{i43}, \end{aligned} \quad (59)$$

com coeficiente de determinação ajustado $\bar{R}^2 = 0,9966$.

Tabela 25 - Estimativas dos coeficientes da regressão e os, respectivos, erros padrões

Parâmetro	Estimativa	Erro padrão	valor t	$Pr > t $
α	$-1,354 \times 10^{-9}$	558,5	$-2,42 \times 10^{-12}$	1,0000
β_1	3,064	1,210	2,532	0,0141
β_2	8,382	2,603	3,220	0,0021
β_4	1605	997,4	1,609	0,1130
β_6	-4491	1967	-2,283	0,0261
β_8	$-6,227 \times 10^4$	$2,383 \times 10^4$	-2,613	0,0114
β_9	$1,377 \times 10^7$	$5,892 \times 10^6$	2,337	0,0229
β_{10}	7159	5374	1,332	0,1880
β_{16}	-1059	379,9	-2,787	0,0072
β_{17}	$3,674 \times 10^4$	$1,849 \times 10^4$	1,987	0,0516
β_{18}	-1510	917,9	-1,645	0,1054
β_{20}	$-3,683 \times 10^4$	$1,783 \times 10^4$	-2,066	0,0433
β_{22}	-631,9	359,1	-1,760	0,0837
β_{26}	$4,619 \times 10^4$	1477	31,276	$< 0,0001$
β_{30}	342,1	169,5	2,019	0,0482
β_{33}	-3503	1477	-2,371	0,0211
β_{35}	-971,1	584,3	-1,662	0,1019
β_{37}	92,29	1,804	51,160	$< 0,0001$
β_{39}	$8,533 \times 10^4$	$3,191 \times 10^4$	2,674	0,0097
β_{40}	35,60	11,43	3,115	0,0029
β_{41}	63,89	25,30	2,526	0,0143
β_{42}	$6,085 \times 10^4$	$1,212 \times 10^4$	5,019	$< 0,0001$
β_{43}	2953	605,6	4,876	$< 0,0001$

Tabela 26 - Análise de variância da regressão

Causas de variação	GL	SQ	QM	F	$Pr > F$
Regressão	16	666513106819	41657069176	1649	< 0,0001
β_4/β	1	39885774	39885774	1,58	0,2138
β_{10}/β	1	2090376	2090376	0,08	0,7783
β_{17}/β	1	37887811	37887811	1,50	0,2256
β_{18}/β	1	42628968	42628968	1,69	0,1987
β_{22}/β	1	72755187	72755187	2,88	0,0950
β_{35}/β	1	76391065	76391065	3,02	0,0876
Resíduo	58	1465254000	25263000		
Total	80	668250000000			

$\beta = (\beta_1, \beta_2, \beta_6, \beta_8, \beta_9, \beta_{16}, \beta_{20}, \beta_{26}, \beta_{30}, \beta_{33}, \beta_{37}, \beta_{39}, \beta_{40}, \beta_{41}, \beta_{42}, \beta_{43})'$

A figura 50 apresenta o gráfico dos valores observados e os valores ajustados pelo modelo (59).

Figura 50 - Gráfico dos valores observados e valores ajustados pelo modelo (59)

5 CONSIDERAÇÕES FINAIS

Com uso da metodologia “Lean” foi possível obter o “Lead time” para o processo produtivo do etanol, o tempo das etapas que agregam e não agregam valor, e consequentemente, a eficiência do ciclo do processo, identificando e reduzindo o desperdício.

Com o modelo de análise de regressão linear múltipla foram identificadas as principais variáveis que determinam a eficiência do processo produtivo do etanol, nas diversas etapas do processo:

No processo da extração do caldo das 07 variáveis regressoras, incluídas na análise, apenas 03 (cana inteira, cana picada e volume da água de embebição) mostraram ser significativas no modelo, com a proporção da variação para a variável caldo extraído da moenda de 77,86%.

No processo do tratamento do caldo obteve-se:

- (i) para a variável brix do caldo caleado das 09 variáveis incluídas na análise de regressão, 03 mostraram-se significativas, sendo elas: caldo misto extraído da moenda, pureza do caldo misto e leitura do sacarímetro, com uma variação explicada de 98,16%.
- (ii) para a variável pol do caldo clarificado analisada em função das 09 variáveis regressoras, 03 mostraram-se significativas com 99,99% de variação explicada. Especial atenção deve ser dada às 03 variáveis: pureza do caldo misto, AR do caldo misto e leitura do sacarímetro.
- (iii) a variável AR decantado foi analisada em função de 10 variáveis regressoras, das quais 04 mostraram-se significativas, com uma variação explicada de 99,37%, sendo estas: pol do caldo misto, leitura do sacarímetro, volume gasto de AR e raiz quadrada do volume gasto de AR.
- (iv) para a variável ART do caldo analisada em função das 09 variáveis regressoras, 06 mostraram-se significativas com 99,98% de variação explicada. Identificadas como: densidade, pol, pureza e AR do caldo misto, leitura do sacarímetro e volume gasto de AR.

No processo da fermentação as variáveis Y_{i1} e Y_{i2} não mostraram nenhum tipo de relação com as variáveis regressoras.

No processo da destilação, a variável etanol hidratado (produto final), foi analisada em função de 43 variáveis regressoras, sendo que destas, 16 variáveis mostraram-se significativas, com variação explicada de 99,66%. Isso indica que essas 16 variáveis, descritas a seguir, devem ser priorizadas e dada atenção especial no processo de produção: cana inteira que entra na moenda; cana picada que entra na moenda; pol do caldo da cana desfibrada; brix do caldo misto; densidade do caldo misto; volume gasto de AR; ART do caldo; GL nas

dornas; % levedo no creme; brix do leite tratado; volume da vinhaça; % etanol na vinhaça; ART do caldo destilaria; ART do mel final; rendimento da destilaria; litros de álcool por tonelada de mel.

5.1 Conclusões

Com base nos resultados obtidos, pode-se concluir que:

Com a efetivação de algumas propostas básicas indicadas neste estudo, pode-se obter uma redução de mais de 30% no “Lead time” do processo de produção do etanol hidratado. Esta redução pode ser alcançada com a diminuição do tempo das atividades que não agregam valor, aumentando assim, a eficiência do ciclo do processo e consequente aumento de produção.

As análises de regressão realizadas permitiram que pudessem ser identificadas as variáveis que realmente estão em função das variáveis respostas. Os resultados indicam que o método de seleção de variáveis nos modelos de regressão é uma metodologia eficaz para a avaliação do processo produtivo do etanol em usinas de açúcar e álcool. Os casos em que os ajustes para as variáveis respostas não foram adequados necessitam de estudos complementares para uma correta avaliação.

Com a adoção da metodologia “Lean Seis Sigma” é possível aumentar a eficiência do processo e a produção do etanol, podendo ser indicada em estudos futuros, visando ampliar este estudo e para outros processos produtivos, em indústrias que guardem semelhança com a usina estudada.

REFERÊNCIAS

- AKAO, Y. **Quality function deployment:** integrating customer requirements into product design. Portland: Maple-Vail, 1990. 369p.
- ALLEN, T.T. **Introduction to engineering statistics and six sigma:** statistical quality control and design of experimens and systems. London: Springer, 2006. 529 p.
- CAMPOS, V.F. **Gerenciamento pelas diretrizes.** Belo Horizonte: Fundação Christiano Ottoni, Escola de Engenharia da UFMG, 1996. 331 p.
- CAPUTO, M.M. **Indução da maturação por produtos químicos e sua consequência na qualidade tecnológica de diferentes genótipos de cana-de-açúcar.** 2006. 137 p. Dissertação (Mestrado em Fitotecnia) - Escola Superior de Agricultura “Luiz de Queirós”, Universidade de Sao Paulo, Piracicaba. 2006.
- DEMÉTRIO, C.G.B.; ZOCCHI, S.S. **Modelos de regressão.** Disponível em: <<http://verde.esalq.usp.br/jorge/cursos/clarice/ApostilaP.pdf>>. Acesso em: 20 ago. 2006.
- DRAPER, N.R.; SMITH, H. **Applied regression analysis.** 3.ed. New York: John Wiley & Sons, Inc. 1998. 706 p.
- FERNANDES, A.C. **Cálculos na agroindústria da cana-de-açúcar.** 2.ed. Piracicaba: STAB, 2003. 240 p.
- GEORGE, M.L. **Lean seis sigma para serviços:** como utilizar velocidade lean. Rio de Janeiro: Qualitymark, 2004. 436 p.
- HOERL, R.W. Six sigma black belts: what do they need to know? **Journal of Quality Technology**, Toronto, v. 33, n. 4, p. 391-406, oct. 2001.
- INFORMAÇÃO UNICA. **Safr paulista de cana-de-açúcar cresce mais de 8% .** São Paulo: UNICA, ano 9, n. 75, mar./abr. 2007a. 8 p.
- INFORMAÇÃO UNICA. **São Paulo é a capital do etanol.** São Paulo: UNICA, ano 9, n. 76, p. 1-8, mai./jun. 2007b. 8 p.
- JURAN, J.M.; GODFREY, A.B. **Juran’s quality handbook.** 5th ed. New York: McGraw Hill, 1998. 1872 p.
- MILAN, M.; BARROS, J.W.D.; GAVA, J.L. Planning soil tillage using Quality Function Deployment (QFD). **Scientia Agricola**, Piracicaba, v. 60, n. 2, p. 217-221, abr./jun. 2003.
- MONTGOMERY, D.C; PECK, E.A; VINING, G.G. **Introduction to linear regression analysis.** 4.ed. New York: John Wiley & Sons, 2006. 612 p.
- MONTGOMERY, D.C. **Introdução ao controle estatístico da qualidade.** 4.ed. Rio de Janeiro: Livros Técnicos e Científicos, 2004. 513 p.
- MONTGOMERY, D.C; PECK, E.A. **Introduction to linear regression analysis.** 2.ed. New York: John Wiley & Sons, 1992. 527 p.

NAGUMO, G.K. **Desdobramento da função qualidade (QFD) aplicado à produção de mudas de café (Coffea arabica L.)**. 2005. 61 p. Dissertação (Mestrado em Máquinas Agrícolas) - Escola Superior de Agricultura "Luiz de Queirós", Universidade de São Paulo, Piracicaba. 2005.

NOTÍCIA UNICA. **Bagaço de cana representa 4% da matriz energética**. Edição 2240. São Paulo: 16 de Dezembro de 2008. Disponível em: <<http://www.unica.com.br>>. Acesso em: 17 dez. 2008.

OLIVEIRA, A. M. S. de. **Perspectivas sobre o setor sucoalcooleiro frente a redução da queimada de cana-de-açúcar, a intensificação do corte mecanizado e a certificação socioambiental**. Disponível em: <http://www.agencia.cnptia.embrapa.br/Repositorio/Perspectivas_setor-sucoalcooleiro-000fjablj0y02wyiv809gkz51gf7ldgy.pdf>. Acesso em: 09 dez. 2008.

PANDE, P.S.; NEUMAN, R.P.; CAVANAGH, R.R. **Estratégias seis sigma: como a GE, a Motorola e outras grandes empresas estão aguçando seu desempenho**. Rio de Janeiro: Qualitymark, 2001. 442 p.

PYZDEK, T. **The six sigma handbook**. New York: McGraw-Hill, 2003. 830 p.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing.**, Vienna, Austria, 2006. Disponível em: <<http://www.R-project.org>>. Acesso em: 19 fev. 2008.

RIOS, L.V. **Um modelo de referência para melhoria de processos industriais usando conceitos seis sigma**. 2006. 114 p. Dissertação (Mestrado em Engenharia Mecânica) - Universidade Federal de Santa Catarina, Florianópolis. 2006.

RODRIGUES, D. ORTIZ, L. **Em direção à sustentabilidade da produção de etanol de cana de açúcar no Brasil**. Outubro de 2006. Disponível em: <<http://www.vitaecivilis.org.br/default.asp/site-Acao=MostraPagina/PaginaId=2080>>. Acesso em: 7 out. 2008.

ROTHER, M.; SHOOK, J. **Aprendendo a enxergar: mapeando o fluxo de valor para agregar valor e eliminar o desperdício**. São Paulo: Lean Institute Brasil, 2003. 113 p.

SCATOLIN, A.C. **Aplicação da metodologia seis sigma na redução das perdas de um processo de manufatura**. 2005. 137 p. Dissertação (Mestrado em Engenharia Mecânica) - Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, Campinas. 2005.

SÃO PAULO. SECRETARIA DO MEIO AMBIENTE. **Secretaria do Meio Ambiente exige que queima de palha de cana seja comunicada com antecedência de 48 horas**. Notícia 16/04/02. Disponível em: <<http://www.ambiente.sp.gov.br/destaque/queima-palhadecana.htm>>. Acesso em: 9 dez. 2008.

SHINGO, S. **O sistema Toyota de produção do ponto de vista da engenharia de produção**. 2.ed. Porto Alegre: Artmed, 1996. 296 p.

UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR. **Produção e uso do etanol combustível no Brasil**. São Paulo, 2007. 68 p.

UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR. Disponível em: <<http://www.unica.com.br/multimedia/videocast/default.asp?mmdCode=9A64FC42-5CA7-4D66-95D3-AB2540184A38>>. Acesso em: 20 out. 2008.

UNIÃO DA INDÚSTRIA DE CANA-DE-AÇÚCAR. **As Perguntas mais Frequentes (FAQs)**. Disponível em: <<http://www.unica.com.br/FAQ/>>. Acesso em: 19 jan. 2009a.

UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR. **Folder - A indústria da cana-de-açúcar** - 05/01/2009. Disponível em: <<http://www.unica.com.br/multimedia/publicacao/>>. Acesso em: 7 jan. 2009b.

UNIÃO DA INDÚSTRIA DA CANA-DE-AÇÚCAR. Disponível em: <<http://www.unica.com.br/dadosCotacao/estatistica/>>. Acesso em: 20 mai. 2010.

USTULIN, E. J.; SEVERO, J. R. **Cana-de-açúcar: Proteger o ambiente e continuar gerando empregos**. Confederação da Agricultura e Pecuária do Brasil. CNA: N° 179, Ano 2001, Setembro. Disponível em: <<http://www.cna.org.br/site/noticia.php?n=1172>>. Acesso em: 09 dez. 2008.

WALTER, A.C.S. Potencial energético da cana-de-açúcar. **STAB - Açúcar, Álcool e Subprodutos**, Piracicaba, v.11, n.4, p. 29-34, mar./abr. 1993.

WERKEMA, M.C.C. **Lean seis sigma**: introdução às ferramentas do Lean Manufacturing. Belo Horizonte: Werkema Editora, 2006. 120 p. (Seis Sigma, 4).

WOMACK, J.P.; JONES, D.T. **A mentalidade enxuta nas empresas**: elimine o desperdício e crie riqueza. Rio de Janeiro: Editora CAMPUS, 2004. 458 p.

APÊNDICE

Programa R para análise de regressão linear múltipla para variável etanol hidratado:

```
AL <- read.csv2("E://ProjetoUsinaDecasa/DECASA_Dados2/Analise_Alcool.csv",
header = TRUE)
dim(AL)
#Fazendo uma cópia dos dados
ALcopy <- AL
#Centrando as variaveis de AL
AL <- scale(AL,center=TRUE,scale=FALSE)
y <- matrix(AL[,1],81,1)
x <- matrix(AL[,2:44],81,43);dim(x)
# Variáveis
y1<-AL[,1];x1<-AL[,2];x2<-AL[,3];x3<-AL[,4];x4<-AL[,5];x5<-AL[,6];
x6<-AL[,7];x7<-AL[,8];x8<-AL[,9];x9<-AL[,10];x10<-AL[,11];
x11a<-AL[,12];x12<-AL[,13];x13<-AL[,14];x14<-AL[,15];x15<-AL[,16];
x16<-AL[,17];x17<-AL[,18];x18<-AL[,19];x19<-AL[,20];x20<-AL[,21];
x21<-AL[,22];x22<-AL[,23];x23<-AL[,24];x24<-AL[,25];x25<-AL[,26];
x26<-AL[,27];x27<-AL[,28];x28<-AL[,29];x29<-AL[,30];x30<-AL[,31];
x31<-AL[,32];x32<-AL[,33];x33<-AL[,34];x34<-AL[,35];x35<-AL[,36];
x36<-AL[,37];x37<-AL[,38];x38<-AL[,39];x39<-AL[,40];x40<-AL[,41];
x41<-AL[,42];x42<-AL[,43];x43<-AL[,44]
# Modelo com todas as variáveis
mod1<-lm(y1~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11a+x12+x13+x14+x15+x16+
x17+x18+x19+x20+x21+x22+x23+x24+x25+x26+x27+x28+x29+x30+x31+x32+x33+
x34+x35+x36+x37+x38+x39+x40+x41+x42+x43)
# Seleção de variaveis
step(mod1,direction=c("both"))
step(mod1,direction=c("backward"))
step(mod1,direction=c("forward"))
#Comparação dos modelos
mod1<-lm(y1~1)
mod2<-lm(y1~x1+x2+x4+x6+x8+x9+x10+x16+x17+x18+x20+x22+x26+x30+x33+
x35+x37+x39+x40+x41+x42+x43)#Modelo indicado por stepwise
mod3<-lm(y1~x1+x2+x3+x4+x5+x6+x7+x8+x9+x10+x11a+x12+x13+x14+x15+x16+
x17+x18+x19+x20+x21+x22+x23+x24+x25+x26+x27+x28+x29+x30+x31+x32+x33+
x34+x35+x36+x37+x38+x39+x40+x41+x42+x43) #Modelo completo
```

```

anova(mod3,mod2) #Teste F p/seleção de modelos.
# Análise dos resíduos
mod2<-lm(y1~x1+x2+x4+x6+x8+x9+x10+x16+x17+x18+x20+x22+x26+x30+x33+
x35+x37+x39+x40+x41+x42+x43)
# Gráfico dos Resíduos estudentizados
y.ajust <- mod2$fitted.values+mean(ALcopy[,1])
plot(y.ajust,rstudent(mod2),xlim=c(190000,560000),xlab="Y ajustado",
ylab="Resíduos estudentizados externamente")
abline(0,0)
identify(y.ajust,rstudent(mod2))
# Gráf. QQPlot
qq1 <- qqnorm(rstudent(mod2),xlab="Quantis teóricos",
ylab="Quantis amostrais",main="")
qqline(rstudent(mod2),col="red")
identify(qq1)
#Histograma
hist(rstudent(mod2),probability=TRUE,ylim=c(0,.5),xlab="Resíduos estudentizados
externamente",ylab="Densidade",main="")
lines(y<-seq(min(rstudent(mod2)),max(rstudent(mod2)),0.05),
dnorm(y,mean(rstudent(mod2)),sd(rstudent(mod2))))
# Gráfico box plot
box1<- boxplot(rstudent(mod2),ylab="Resíduos estudentizados externamente" )
box1$out
#Teste para normalidade (Shapiro-Wilk)
shapiro.test(rstudent(mod2))
# Gráfico dos valores absolutos de (DFFitS)
n<-dim(AL)[1] # number of data
plot(abs(dffits(mod2)), xlab="Índice",ylab="Valor absoluto (DFFitS)")
abline(3*sqrt(mod2$rank/mod2$df.residual),0,lty=2)
identify(1:n,abs(dffits(mod2))
#Medidas de influência
IM1<-influence.measures(mod2)
summary(IM1)
# Verificar se uma observação pode ser classificada como inconsistente
rse<-abs(rstudent(mod2))
qt(1-(0.95/162),57)

```

```

a<-rse>qt(1-(0.95/162),57)
# Valor de F[50%;p;n-p] (distância de Cook)
qf(0.50,23,58)
# Estimativas dos coeficientes de regressão
summary(mod2)
# Análise de variância
mod1<-lm(y1~1)
mod2<-lm(y1~x1+x2+x6+x8+x9+x16+x20+x26+x30+x33+x37+x39+x40+x41+x42+x43)
mod3<-lm(y1~x1+x2+x6+x8+x9+x16+x20+x26+x30+x33+x37+x39+x40+x41+x42+x43+x4)
mod4<-lm(y1~x1+x2+x6+x8+x9+x16+x20+x26+x30+x33+x37+x39+x40+x41+x42+x43+x10)
mod5<-lm(y1~x1+x2+x6+x8+x9+x16+x20+x26+x30+x33+x37+x39+x40+x41+x42+x43+x17)
mod6<-lm(y1~x1+x2+x6+x8+x9+x16+x20+x26+x30+x33+x37+x39+x40+x41+x42+x43+x18)
mod7<-lm(y1~x1+x2+x6+x8+x9+x16+x20+x26+x30+x33+x37+x39+x40+x41+x42+x43+x22)
mod8<-lm(y1~x1+x2+x6+x8+x9+x16+x20+x26+x30+x33+x37+x39+x40+x41+x42+x43+x35)
anova(mod1)
anova(mod2)
anova(mod2,mod3) #testa se a variável x4 é significativa no modelo
anova(mod2,mod4) #testa se a variável x10 é significativa no modelo
anova(mod2,mod5) #testa se a variável x17 é significativa no modelo
anova(mod2,mod6) #testa se a variável x18 é significativa no modelo
anova(mod2,mod7) #testa se a variável x22 é significativa no modelo
anova(mod2,mod8) #testa se a variável x35 é significativa no modelo
# Gráfico dos valores observados e ajustados para y1:
mod2<-lm(y1~x1+x2+x6+x8+x9+x16+x20+x26+x30+x33+x37+x39+x40+x41+x42+x43)
y.ajust <- mod2$fitted.values+mean(ALcopy[,1])
plot(c(185000,560000), c(185000,560000), type="n", xlab="Y observado",
ylab="Y ajustado",cex.lab=1, font.lab=6,cex=2,axes=FALSE)
axis(2, at=seq(185000,560000,100000), labels=c(expression(paste(2,'x',10^{5})),
expression(paste(3,'x',10^{5})),expression(paste(4,'x',10^{5})),
expression(paste(5,'x',10^{5}))))
axis(1, at=seq(185000,560000,100000), labels=c(expression(paste(2,'x',10^{5})),
expression(paste(3,'x',10^{5})),expression(paste(4,'x',10^{5})),
expression(paste(5,'x',10^{5}))))
box();points(ALcopy[,1],y.ajust)
abline(0,1)

```