UNIVERSIDADE FEDERAL DE JUIZ DE FORA
CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

GERMANA ADRIELE ROCHA

PROPOSTA DE UM NOVO MODELO DE REGRESSÃO LINEAR PARA PREVISÃO E CONTROLE DO INDICADOR DE EFICIÊNCIA ENERGÉTICA DE UMA EMPRESA FERROVIÁRIA

JUIZ DE FORA
2013

GERMANA ADRIELE ROCHA

PROPOSTA DE UM NOVO MODELO DE REGRESSÃO LINEAR PARA PREVISÃO E CONTROLE DO INDICADOR DE EFICIÊNCIA ENERGÉTICA DE UMA EMPRESA FERROVIÁRIA

Trabalho de Conclusão de Curso apresentado a Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do título de Engenheiro de Produção.

Orientador: D.Sc., Roberto Malheiros Moreira Filho

Co-Orientador: Bernardo Ramos de Abreu

JUIZ DE FORA

2013

Rocha, Germana Adriele.
PROPOSTA DE UM NOVO MODELO DE REGRESSÃO LINEAR PARA PREVISÃO E CONTROLE DO INDICADOR DE EFICIÊNCIA ENERGÉTICA DE UM EMPRESA FERROVIÁRIA / Germana Adriele Rocha. -- 2013.
70 p. : il.

Orientador: Roberto Malheiros Moreira Filho
Coorientador: Bernardo Ramos de Abreu
Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Faculdade de Engenharia, 2013.

1. Eficiência Energética. 2. Modelo de Regressão Linear Múltipla. 3. Operação Ferroviária. I. Malheiros Moreira Filho, Roberto, orient. II. Ramos de Abreu, Bernardo, coorient. III. Título.

GERMANA ADRIELE ROCHA

PROPOSTA DE UM NOVO MODELO DE REGRESSÃO LINEAR PARA PREVISÃO E CONTROLE DO INDICADOR DE EFICIÊNCIA ENERGÉTICA DE UMA EMPRESA FERROVIÁRIA

Trabalho de Conclusão de Curso apresentado a Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do título de Engenheiro de Produção.

Aprovada em 22 de agosto de 2013.

BANCA EXAMINADORA

_____

D.Sc., Roberto Malheiros Moreira Filho

Universidade Federal de Juiz de Fora

_____

Bernardo Ramos de Abreu

MRS Logística S.A.

_____

M.Sc., José Geraldo Ferreira

Universidade Federal de Juiz de Fora

AGRADECIMENTOS

À Deus, pela saúde, capacitação e dons de providência.

Aos meus pais, Geraldo e Clarice, a quem dedico esse trabalho, por acreditarem no meu potencial e oferecerem o suporte necessário para realização desse feito.

Aos meus avôs, pelo imenso carinho e exemplo de bondade e superação.

Aos meus familiares, pelos gestos de apoio e compreensão das minhas ausências.

Aos meus professores de Matemática do Ensino Médio, por me despertarem a paixão por essa disciplina.

À Universidade Federal de Juiz de Fora, por proporcionar acesso gratuito ao ensino de qualidade em Engenharia de Produção.

Aos meus amigos, pelo carinho, apoio e por não só compartilharem ansiedades e frustrações, como também risos e alegrias, que tornaram os momentos de dificuldade mais amenos.

Aos professores orientadores, Roberto Malheiros e Bernardo Abreu, e ao professor convidado José Geraldo Ferreira, pelas críticas e auxílios que engrandeceram o desenvolvimento desse trabalho.

À MRS Logística, pela oportunidade de aprendizado e crescimento profissional.

À todos que, de alguma forma, contribuíram para tornar o período universitário uma experiência única e enriquecedora.

“O futuro pertence àqueles que
acreditam na beleza de seus sonhos”.
(Elleanor Roosevelt)

RESUMO

O objetivo desse trabalho compreende a obtenção de um novo modelo de regressão linear para previsão e controle do indicador de Eficiência Energética (E.E.) da MRS Logística. Esse indicador é definido como litros gastos para transportar mil tonelada bruta na distância percorrida, proporcionando uma visão do desempenho energético da empresa perante o nível de produção, sendo utilizado também como ferramenta de gestão do principal gasto operacional; o combustível. No desenvolvimento do novo modelo, a autora revisa as principais alavancas de E.E., analisando de forma qualitativa quais deteriam maior poder de explicação do indicador estudado. Posteriormente, são realizadas análises estatísticas com a finalidade de comprovar a correlação das variáveis com a variável resposta, que à priori foram selecionadas de forma empírica. O escopo de aplicação da regressão linear determina o atendimento de alguns pressupostos para atestar a qualidade de um modelo, logo, os modelos gerados são validados valendo-se de estatísticas quanto à multicolinearidade, autocorrelação e homoscedasticidade. Alguns dos modelos são obtidos através do uso de artifícios matemáticos indicados para minimizar problemas quanto à alta correlação entre variáveis preditoras que deveriam ser independentes, ocasionando assim, a multicolinearidade. A definição do modelo ideal é fundamentada em análises quantitativas e, complementada, por análises qualitativas, sendo que esta apresenta participação fundamental durante toda a elaboração desse trabalho. A necessidade de redefinição das variáveis do modelo comprovou a hipótese de que havia variáveis com maior poder de explicação para o atual panorama da eficiência energética. Os resultados dessa pesquisa compreenderam o desenvolvimento de um modelo aceito matematicamente e, na prática, superior ao modelo atualmente utilizado pela empresa, visto que possui uma aderência maior aos recentes valores do indicador.

Palavras-chave: Modelo de Regressão Linear, Eficiência Energética, Operação Ferroviária.

ABSTRACT

The objective of this work comprises obtaining a new linear regression model for forecasting and control of the Energy Efficiency's (EE) indicator at MRS Logística. This indicator is defined as liters spent to transport a thousand ton of gross material per distance traveled, providing an overview of the company's energy performance before the production level, and also used as a tool for managing the main operational expense; fuel. In the development of the new model, the author reviews the main levers of EE, analyzing qualitatively which would hold greater explanatory power of the indicator studied. Thereafter, statistical analyzes are performed in order to prove the correlation of the variables with the response variable, which a priori were selected empirically. The scope of the linear regression's application determines the compliance of some assumptions to attest the quality of a model, so that the generated models are validate making use of statistics as multicollinearity, autocorrelation and homoscedasticity. Some models are obtained through the use of mathematical artifice indicated to minimize problems about the high correlation between the predictor variables that should be independent, thereby causing the multicollinearity. The definition of the ideal model is based on quantitative analysis and complemented by qualitative analyzes, which features a fundamental role throughout the development of this work. The need for a redefinition of the model variables proved the hypothesis that there were variables with greater explanatory power for the current outlook of energy efficiency. The results of this study comprised the development of a model and mathematically accepted, and in practice, better than the model currently used by the company, since it has a greater adhesion to recent indicator's values.

Keywords: Linear Regression Model, Energy Efficiency, Railway Operation.

LISTA DE FIGURAS

Figura 1: Comparação entre a E.E. real e a prevista pelo atual modelo.....17
Figura 2: Metodologia de pesquisa em Engenharia de Produção 20
Figura 3: Padrões de comportamento dos gráficos de resíduos 31
Figura 4: Malha Ferroviária da MRS Logísticas e Postos Acessados.....35
Figura 5: Evolução EE 39
Figura 6: Diagrama de Extremos-e-Quartis 44
Figura 7: Gráfico de dispersão entre as variáveis independentes.....45
Figura 8: Teste de Normalidade dos Resíduos.....51
Figura 9: Histograma dos Resíduos.....52
Figura 10: Gráfico de Dispersão Residual - Modelo II.....52
Figura 11: Auto Correlação Durbin-Watson - Modelo II.....53
Figura 12: Gráfico de Dipersão das Variáveis - Modelo III.....54
Figura 13: Gráfico de DW - Modelo III.....56
Figura 14: Gráfico de comparação Novo Modelo (II) x Antigo Modelo.....60

LISTA DE QUADROS

Quadro 1: Variáveis de impacto na EE 38
Quadro 2: Definição das variáveis do modelo proposto 40
Quadro 3: Modelo de Regressão Qualitativo I 46
Quadro 4: Modelo de Regressão II 50
Quadro 5: Modelo de Regressão III 55

LISTA DE TABELAS

Tabela 1: Análise de Variância para Teste de Significância 28
Tabela 2: Principais Linhas Operadas pela MRS Logística – Características 34
Tabela 3: Correlação individual das variáveis dependentes com a EE 42
Tabela 4: Grau de correlação entre as variáveis 43
Tabela 5: Correlação entre as variáveis “independentes” 45
Tabela 6: Análise de Componentes de Fatores 47
Tabela 7: Regressão por Etapas 48
Tabela 8: Regressão – Melhores Subconjuntos 49
Tabela 9: Correlação entre Variáveis Independentes (III) 55

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

AF – Análise Fatorial

CG – Carga Geral

DW – Durbin- Watson

EE – Eficiência Energética

EFC – Estradas de Ferro Carajás

F – Distribuição F

FIV – Fator de inflação da variância

HH – Heavy Haul

P – P-value

TKB – Tonelada por Quilômetro Bruta, Medida de Produção

TKU – Tonelada por Quilômetro Útil, Medida de Produção

TTC – Transit Time Carregado Ponderado

TU – Tonelada Útil, Carga Transportada

T – Distribuição T-Student

Util_AC+C – Utilização de locomotivas modelos GE-AC44MIL e GE-C44MIL

Util_Out – Utilização de outros modelos de locomotivas

VVC – Fila do Carregado Ponderado

SUMÁRIO

1. INTRODUÇÃO 15

1.1 CONSIDERAÇÕES INICIAIS 15
1.2 JUSTIFICATIVA 16
1.3 ESCOPO DO TRABALHO 17
1.4 FORMULAÇÃO DE HIPÓTESES 18
1.5 ELABORAÇÃO DOS OBJETIVOS 19
1.6 DEFINIÇÃO DA METODOLOGIA 19
1.7 ESTRUTURA DO TRABALHO 21

2. REVISÃO DE LITERATURA 22

2.1 ANÁLISE MULTIVARIADA 22
2.2 REGRESSÃO MÚLTIPLA 23
2.3 REGRESSÃO LINEAR MÚLTIPLA 24
2.4 ESTIMAÇÃO DE MÍNIMOS QUADRADOS DE PARÂMETROS 25
2.5 PROPRIEDADE DOS ESTIMADORES 26
2.6 TESTE DE HIPÓTESE 27
2.7 R^2 E R^2 AJUSTADO 28
2.8 TESTE PARA OS COEFICIENTES INDIVIDUAIS 29
2.9 INTERVALOS DE CONFIANÇA PARA OS COEFICIENTES DA REGRESSÃO 30
2.10 PREVISÃO DE VALORES 30
2.11 ANÁLISE RESIDUAL 31
2.12 ANÁLISE FATORIAL 32
2.13 MULTICONEARIDADE 33

3. EFICIÊNCIA ENERGÉTICA NA MRS LOGÍSTICA 34

3.1 A MRS LOGÍSTICA S.A. 34
3.2 CONSUMO DE COMBUSTÍVEL NA MRS LOGÍSTICA 36

4. DESENVOLVIMENTO 40

4.1 VARIÁVEIS DO MODELO 40
4.2 COLETA DE DADOS 41
4.3 ANÁLISE INDIVIDUAL DAS VARIÁVEIS 42
4.3.1 CORRELAÇÃO COM A EE 42
4.3.2 DIFERENÇA ENTRE AS VARIÁVEIS 43
4.3.3 RELAÇÃO ENTRE AS VARIÁVEIS 44

5. MODELO DE REGRESSÃO 46

5.1 MODELO DE REGRESSÃO – VARIÁVEIS QUALITATIVAS (I) 46
5.2 MODELO DE REGRESSÃO – VARIÁVEIS QUALITATIVAS (II) 49
5.3 MODELO DE REGRESSÃO – COMBINAÇÃO DE VARIÁVEIS (III) 54

6. ANÁLISE E COMPARAÇÃO DOS MODELOS 57

6.1 ANÁLISE E ESCOLHA DENTRE OS MODELOS 57
6.2 COMPARAÇÃO COM O MODELO ANTIGO 58
6.2.1 CRÍTICA ÀS VARIÁVEIS PREDITORAS 58
6.2.2 COMPARAÇÃO ENTRE OS MODELOS 59

7. CONSIDERAÇÕES FINAIS 61

7.1 CONCLUSÃO 61
7.2 RECOMENDAÇÕES DE PESQUISAS FUTURAS 62

REFERÊNCIAS 63

ANEXO 1 – SÁIDAS ESTATÍSTICAS DO MODELO I 65

ANEXO 2 – DISTRIBUIÇÃO T DE STUDENT 66

ANEXO 3 – DISTRIBUIÇÃO F DE SNEDECOR 67

ANEXO 4 – DISTRIBUIÇÃO NORMAL 68

ANEXO 5 – DECLARAÇÃO DA EMPRESA 69

ANEXO 6 – TERMO DE AUTENTICIDADE 70

1. INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

O transporte de carga ferroviário existe à aproximadamente 200 anos no mundo e 150 anos no Brasil. As primeiras ferrovias brasileiras eram perpendiculares ao litoral e tinham o propósito de levar cargas de exportação (minérios, café, açúcar e algodão) para os portos (SILVEIRA, 2003). Desde então, o setor se destaca pela participação e relevância na matriz de transporte brasileiro, principalmente, após a privatização da malha ferroviária em 1992. Em 1950, essa participação era de 21%; já em 2004, chegou a representar 32% do transporte de carga no Brasil, sendo que o transporte nacional passou de 9 milhões para 21 milhões no período referido.

Como desde 1960 não houve investimentos no que diz respeito à ampliação da malha ferroviária brasileira, pelo contrário, a mesma diminuiu de 38 mil quilômetros para 29.952 quilômetros em 2005 (Vaz, 2008); há uma intensa pressão pela diminuição do custo e aumento da eficiência operacional, principalmente a eficiência energética. Sendo assim, tornou-se uma característica desse setor, que grande parte dos custos operacionais seja oriunda de gastos com combustível; na MRS, por exemplo, essa conta seja a representar aproximadamente 25% dos custos totais operacionais.

Eficiência energética pode ser entendida como a relação entre a produção de um equipamento ou sistema, o serviço por ele prestado e a energia usada para tal. Em um motor, é expressa como um percentual da energia usada (entrada) convertida em energia produzida (saída) seja ela para gerar movimento ou calor. Em uma locomotiva (equipamento), a eficiência energética é frequentemente expressa em litros de combustível pela distância percorrida, enquanto em uma ferrovia (sistema) se fala em litros de combustível por TKB ou TKU. Este é apenas um entre vários indicadores de produção (STODOLSKY, 2002). Em suma, na ferrovia, E.E. se traduz em quantidade de óleo diesel gasta para transportar 1000 TKB.

A história da EE no Brasil começou na década de 1970, após o primeiro choque do petróleo (1973) e ganhou impulso com o segundo choque (1979) com uma série de ações de incentivo à eficiência (Vaz, 2008). Três maiores empresas de cargas ferroviárias do Brasil - Companhia Vale do Rio Doce, América Latina Logística e MRS Logística – destacam-se

como pioneiras brasileiras nas ações em prol da E.E, no que diz respeito tanto a projetos inovadores, como em melhorias nas práticas de condução e equipamentos.

Vaz (2008) reafirma a importância da E.E. no setor de transportes, em que, a busca pela eficiência energética, no nicho específico em que foram desenvolvidas dezenas de inovações tecnológicas, tornou-se importante variável na estratégia das empresas de transporte. Logo, é sabido que além de investir em eficiência energética, é necessário ter mecanismos que possibilitem, não somente seu controle periódico, como também predições que esteiem decisões de cunho estratégico e operacional.

Assim como em diversas áreas, a regressão linear se apresenta como ferramenta altamente elegível para a gestão do indicador de E.E. de uma empresa ferroviária. Sendo isso, corroborado por Werkema & Aguiar (2008) que define a aplicação da regressão como "a análise de regressão pode ser utilizada com vários objetivos, dentre os quais se podem destacar: descrição, predição, controle e estimativa".

1.2 JUSTIFICATIVA

O conceito de eficiência energética (E.E.) tomou impulso com o fim da "energia barata", marcada pela expressiva alta nos preços do petróleo na década de 1970 (Vaz, 2008). A partir do século XX, a eficiência energética sempre esteve em evidência na estratégia das empresas, sendo vista como propulsora da competitividade para ganho e manutenção de mercado.

No que tange o transporte de carga ferroviário, no Brasil, as ferrovias se encontram em patamares discrepantes de eficiência energética e produtividade, sendo assim, sofrem pressões ambientais, regulatórias e econômicas para aumento da produtividade, gerando assim a necessidade de investimentos em prol da EE.

A disparidade, especificamente a da E.E., deve-se ao fato de que existem muitas particularidades entre as ferrovias, principalmente a quantidade de carga transportada e o trajeto percorrido pela mesma. Essas possuem impacto direto no nível de eficiência e, consequentemente, no consumo de diesel. De forma simplista, para Vag (2008), "a participação do combustível nos custos depende do perfil da ferrovia, como o tipo de carga e as distancias médias percorridas, e o seu rendimento energético. A EFC, por exemplo, é uma ferrovia de alto desempenho especializada em cargas pesadas e de grande volume, o que faz com que o diesel usado tenha uma participação relativamente alta".

Sendo a E.E. indicador universal no setor ferroviário, a necessidade da gestão do mesmo é justificada por possibilitar análises diretamente relacionadas à produtividade e custos, bem como ao consumo de diesel. Logo, pretende-se, a partir do estudo da operação ferroviária da MRS Logística S.A., bem como de suas peculiaridades; desenvolver um novo modelo de previsão e controle da E.E da empresa, justificado pela "perda de aderência" do atual modelo, observável na figura 1. Acredita-se que essa perda pode ser oriunda da hipótese de que as atuais variáveis não possuem a mesma capacidade de predizer tal indicador.

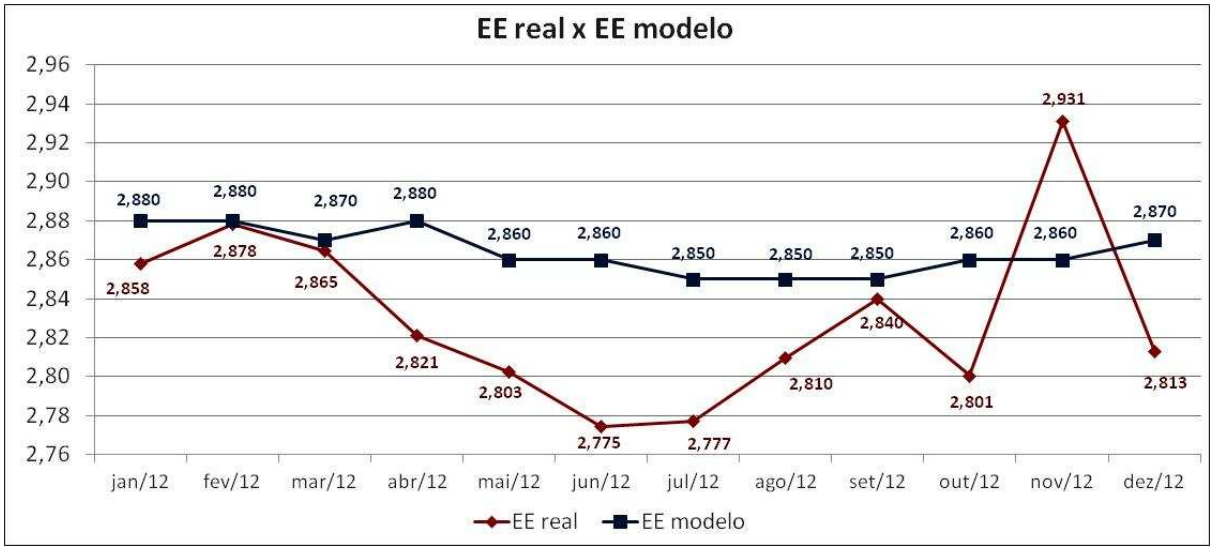


Figura 1: Comparação entre a E.E. real e a prevista pelo atual modelo
Fonte: MRS Logística, 2012.

1.3 ESCOPO DO TRABALHO

O trabalho a ser desenvolvido se dará em parceria com a MRS Logística S.A. empresa de transporte ferroviário do sudeste brasileiro. Atualmente, pode-se dividir a demanda de carga transportada pela empresa em dois grandes grupos: Heavy Haul e Carga Geral. Devido às características operacionais, como a falta de uniformidade da operação da Carga Geral, oscilações de demanda e intensa variação em seu Mix de Transporte, a princípio, o modelo contemplará apenas dados do Minério de Ferro Exportação; que representa aproximadamente 75% de todo o transporte.

Isso se deve principalmente às duas premissas fundamentais quanto à escolha das variáveis do modelo; quantificação e histórico. A quantificação diz respeito à capacidade da empresa de mensurar valores de um dado específico, sejam esses diários ou mensais,

acumulados ou médios. Esse fato ocasiona, necessariamente, na inexistência de histórico que, por sua vez, também pode se dá quando uma variável tem período de mensuração menor que o desejado, não sendo possível realizar cálculos retroativos da mesma.

O novo modelo prezará por utilizar os dados de um histórico mais recente possível, a fim de conseguir absorver as mudanças ocorridas na logística da empresa e retratar o atual desempenho energético da mesma. Em análise preliminar, acredita-se que o horizonte plausível seria o que contemplasse dados com os 3 anos anteriores, 2009 à 2012;

No âmbito matemático, a limitação desse trabalho será quanto ao uso de técnicas lineares tanto para as análises preliminares dos dados, quanto na geração da equação de linear da eficiência energética e validações da mesma.

1.4 FORMULAÇÃO DE HIPÓTESES

O trabalho é desenvolvido à luz da seguinte hipótese:

Quando se analisa a operação da empresa desde sua origem, em 1996, diversas mudanças operacionais ocorreram na mesma, principalmente nos últimos anos, dentre essas podemos citar: compra de diversos ativos mais eficientes e modernos, alteração na composição padrão dos trens de carga, variação no mix de produção, alterações no procedimento padrão de abastecimento, variações significativas no tempo e de percurso dos trens, mudanças nos procedimentos de condução. Essas são algumas das alterações que indicam a possibilidade da obtenção de um novo modelo, possivelmente, composto variáveis distintas à utilizadas atualmente, que consiga uma aderência maior aos recentes valores da E.E.

Essa possibilidade é fundamentada pela desconfiança da existência de variáveis com maior poder de predição do indicador estudado e maior valor prático. Como ainda não é possível afirmar quanto à qualidade esperada do modelo a ser desenvolvido, e consequente poder de explicação das variáveis à serem selecionadas, há receio de que esse modelo não apresente uma explicação tão elevada quanto à do modelo atual, em virtude também do impacto ocasionado pelos dados do passado. Logo, o critério de avaliação de maior peso será o nível de aderência aos valores mais recentes da E.E. no horizonte de 2012.

1.5 ELABORAÇÃO DOS OBJETIVOS

O principal objetivo desse trabalho é propor uma nova ferramenta para domínio de um dos principais indicadores da MRS Logística S.A. Essa ferramenta se dará por meio de um novo modelo de regressão linear de múltiplas variáveis que permita descrever e prever o comportamento da E.E. Para tanto, são estabelecidos os seguintes objetivos específicos:

- Identificar as principais alavancas de E.E. da operação da MRS Logísticas;
- Determinar uma função matemática que busque descrever o comportamento da E.E. por meio de variáveis específicas, com bom nível de correlação com o indicador;
- Obter o grau de impacto que as modificações nas variáveis independentes possuem sobre o indicador;
- Realizar testes para validação do novo modelo e verificar a melhoria do mesmo quando comparado com o modelo atualmente utilizado pela empresa.

1.6 DEFINIÇÃO DA METODOLOGIA

No que se refere à caracterização da pesquisa, segundo Silva e Menezes (2000), um trabalho pode ser avaliado como sendo uma pesquisa básica ou uma pesquisa aplicada. Sendo assim, o presente trabalho se caracteriza como uma pesquisa aplicada, por valer-se da aplicação de uma ferramenta estatística na busca pela solução do problema apresentado.

O presente trabalho, quanto á sua abordagem, apresenta aspecto de pesquisa quantitativa, por utilizar-se de análises numéricas de cunho estatístico; uma vez que, segundo Silva e Menezes (2000), um trabalho científico pode ser classificado como uma pesquisa qualitativa ou quantitativa devido à sua abordagem.

Enfim, por seus objetivos, uma pesquisa pode ser exploratória, explicativa ou descritiva (GIL, 1991). O estudo é de natureza exploratória, pois visa tornar explícito o conhecimento acerca do problema explorado.

Miguel (2010) sumariza essas categorias supracitadas quanto à metodologia de pesquisa de em Engenharia de Produção através da figura 2.

De modo geral, o trabalho pode ser dividido nas seguintes macroetapas:

- Estudo da operação e das alavancas da E.E. da MRS Logística;
- Definição das variáveis do novo modelo;
- Estabelecimento do horizonte histórico a ser utilizado, bem como acúmulo de dados do referido período;
- Desenvolvimento do modelo;
- Testes e validação do novo modelo;
- Análise comparativa entre o novo e antigo modelo.

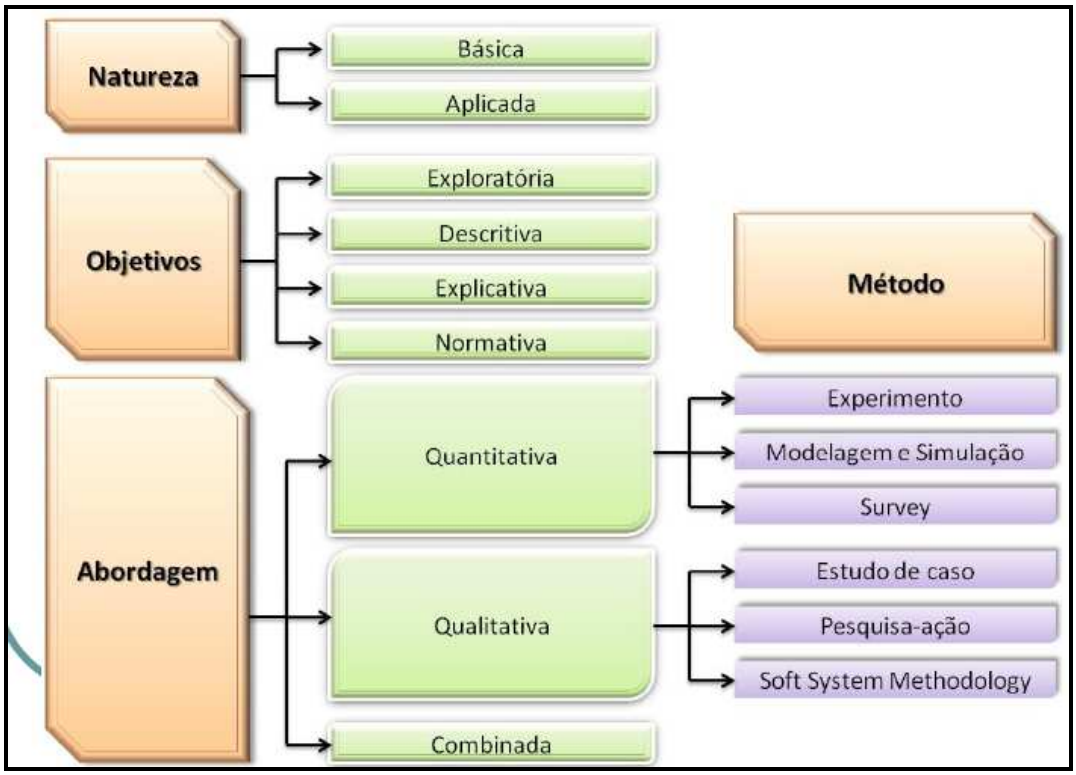


Figura 2: Metodologia de pesquisa em Engenharia de Produção
Fonte: Miguel, 2010 (Adaptado)

Os dados das variáveis a serem utilizados advêm dos históricos de operação da empresa em questão. Todavia nem todos os dados disponíveis serão utilizados, uma vez que o modelo deverá contemplar apenas dados referentes à carga de Minério de Ferro Exportação alusivos a dois clientes chaves.

O modelo será obtido via ferramenta computacional de estatística, que fornecerá também os elementos básicos para as análises cabíveis de validação do mesmo. Validado, o primordial será quantificar a aderência do mesmo aos valores mais recentes do indicador.

1.7 ESTRUTURA DO TRABALHO

Este trabalho está estruturado em 7 capítulos. O presente capítulo se propõe a apresentar questões introdutórias a respeito do assunto a ser desenvolvido, sua relevância, bem como as motivações e os objetivos à serem alcançados pelo mesmo.

O capítulo 2 contém uma revisão literária, no qual é retratada o conceito de regressão, regressão linear, regressão múltipla, bem como suas definições, pressupostos, suposições e métodos de validação.

O capítulo 3 traz uma breve descrição da MRS Logística, sua contextualização no setor de transporte de carga e o detalhamento da sua operação; mostrando as principais características no que abrange ativos, malha, ciclos, *mix* de transporte, padrões de logística e variações operacionais; além de caracterizar o consumo de combustível e seus principais influenciadores.

O capítulo 4 visa identificar as principais alavancas da E.E. resumindo as variáveis de interferência e a relação direta ou indireta dessas com a operação e, primordialmente, com o indicador estudado, que possibilitará a definição das variáveis candidatas a compor o modelo de regressão. Análises estatísticas dessas variáveis e principalmente de suas inter-relações serão úteis no desenvolvimento do capítulo 5.

No capítulo 5 define-se de fato quais são as variáveis que irão compor o modelo de regressão, bem como a equação matemática oriunda da mesma. Ainda são evidenciadas as análises estatísticas para crítica e validação dos modelos desenvolvidos. Essas análises serão complementadas por análises qualitativas, para enfim, no capítulo 6, definir o modelo mais adequado. Este capítulo ainda contém a comparação entre o modelo selecionado e o atual modelo, a fim de avaliar os ganhos.

Por fim, o capítulo 7 irá discorrer à respeito do alcance dos objetivos estipulados, apresentando as conclusões finais, com os adendos necessários quanto à utilização do modelo, que resulta deste trabalho.

2. REVISÃO DE LITERATURA

Antes de dar continuidade ao trabalho proposto, é necessária a realização de uma revisão bibliográfica com a finalidade de apresentar os conceitos sobre a regressão linear e suas derivadas, bem como das análises de validação de um modelo de regressão.

2.1 ANÁLISE MULTIVARIADA

Segundo CORRAR (2009) a análise multivariada refere-se a um conjunto de métodos estatísticos que torna possível a análise simultânea de medidas múltiplas para cada indivíduo, objeto ou fenômeno observado. Sendo assim, de modo sintético, pode ser definida como métodos utilizados para análise de múltiplas variáveis simultaneamente. Essas variáveis podem ser:

- Quantitativas: As quais podem ser medidas em escalas quantitativas, sendo ainda subdividida em Variáveis Discretas – somente valores inteiros são cabíveis – e Variáveis Contínuas – nas quais os valores fracionados também são incorporados;
- Qualitativas: Não possuem valores qualitativos, sendo definidas por categorias ou classificações. Essas são divididas em Variáveis Nominais – classificações aleatórias – e Variáveis Ordinais – apresentam ordenação entre as categorias.

Acredita-se que somente técnicas multivariadas de estatística permitem a análise do desempenho conjunto das variáveis, bem como a determinação de suas influências e relevância. A análise primordial está na relação de dependência e interdependência das variáveis. Cooper e Schindler (2003) ressaltam que, se as variáveis dependentes e interdependentes estão presentes na hipótese da pesquisa, deverá ser utilizada uma das técnicas de dependência como exemplos a regressão múltipla, análise discriminante ou regressão logística.

Dentre as várias técnicas de estatística multivariada, pode-se ressaltar algumas das mais utilizadas (CORRAR, 2009):

- Regressão Múltipla
- Análise Discriminante
- Regressão Logística
- Análise Fatorial
- Análise de Conglomerados
- Escalonamento Multidimensional
- Redes Neurais
- Lógica Nebulosa

2.2 REGRESSÃO MÚLTIPLA

De posse das características e objetivo do estudo proposto, bem como sobre as diversas técnicas apresentadas e com base na reflexão de Reis (2001):

> "O truque na estatística multivariada, se existe, não está nos cálculos, mas consiste em escolher o método mais apropriado ao tipo de dados, usá-lo corretamente, saber interpretar os resultados e retirar deles as conclusões corretas".

Optou-se, neste trabalho, pelo uso da Regressão Múltipla. Essa escolha é tida baseando no fato de que as técnicas de regressão possibilitam a análise da relação existente entre uma específica variável; denominada como variável resposta, variável objetivo ou dependente, com demais variáveis denominadas explicativas, independentes ou, ainda, variáveis de entrada do modelo.

Para Gujarati (2000) a análise da regressão estuda a dependência da característica de qualidade em relação a uma ou mais variáveis, aqui denominadas variáveis de controle, com o objetivo de estimar o valor médio da característica de qualidade em função dos valores conhecidos das variáveis de controle. Weisberg (2005) ainda ressalta a simplicidade da regressão por expressar a relação em termos de uma função matemática, visando sumarizar os dados observados na forma mais simples, útil e elegante possível. Isso também é o que torna esse método altamente utilizado em várias áreas do conhecimento e, portanto, neste estudo.

No campo da regressão, ainda se pode definir distintos métodos como regressão linear, logística, dinâmica, dentre outros. A escolha quanto à esses métodos está relacionada ao tipo da variável resposta, neste caso, a mesma apresenta características compatíveis com a regressão linear.

2.3 REGRESSÃO LINEAR MÚLTIPLA

Segundo Montgomery & Runger (2012), um modelo de regressão que contenha mais de um regressor é chamado de modelo de regressão múltipla, o termo linear é usado porque a equação é uma função linear dos parâmetros desconhecidos $\beta_0, \beta_1 \ldots \beta_\rho$.

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_\rho X_\rho + \varepsilon \tag{1}$$

Onde:

Y = variável dependente;

X_i = variáveis independentes;

β_i = parâmetros estimadores

ε = erro aleatório do modelo

O erro aleatório da equação (1) é caracterizado como uma variável aleatória contínua, independentemente distribuída, com média nula e variância constante ao longo dos valores das variáveis do modelo, sendo assim, Y também é definido como uma variável aleatória e X uma variável normal. Esse pressuposto inicial em torno do ε é, de acordo com Montgomery (2001), importante para a estimação da reta de regressão.

Comumente, o modelo de regressão é expresso em forma de notação matricial. Em um modelo que há **k** variáveis e **n** observações, a equação matricial pode ser representada por:

$$Y = X\beta + \varepsilon \tag{2}$$

Onde:

$$\boldsymbol{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \ \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \text{ e } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Geralmente, Y é um vetor (**n** x 1) observações da variável resposta, X é uma matriz (**n** x **p**) dos níveis das variáveis independentes, β é um vetor (**p** x 1) dos coeficientes de regressão e ε é um vetor (**n** x 1) dos erros aleatórios. A matriz X é denominada matriz modelo.

2.4 ESTIMAÇÃO DE MÍNIMOS QUADRADOS DE PARÂMETROS

É uma unicidade a afirmação de que a função que melhor se ajusta aos dados é aquela para a qual a diferença entre os valores reais e os estimados seja a mínima possível, ou seja, a melhor função é a que propicia o menor resíduo estatístico. O método de predição de modelos mais comumente utilizado em regressão linear é o Método dos Mínimos Quadrados (MMQ), cujo objetivo é justamente obter a menor soma de quadrados dos resíduos (SQR) possível (CORRAR, 2009).

Deseja-se encontrar o vetor dos estimadores quadrados que minimize:

$$L = \sum_{i=1}^{n} \in_i^2 = \in' \in = (y - X\beta)'(y - X\beta) \tag{3}$$

Derivando e igualando a zero a equação 3, tem-se:

$$\hat{\beta} = (X'X)^{-1} X'y \tag{4}$$

Montgomery & Runger (2012), afirma que os elementos da diagonal X’X são as somas dos quadrados dos elementos nas colunas de X e os elementos fora da diagonal são as somas dos produtos cruzados dos elementos nas colunas de X e dos valores observados da variável controle.

$$(\mathbf{X'X}) = \begin{bmatrix} n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ik} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{ik} x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1} x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{i1} x_{ik} & \sum_{i=1}^{n} x_{i1} x_{ik} & \cdots & \sum_{i=1}^{n} x_{ik}^2 \end{bmatrix} \qquad \mathbf{X'y} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1} y_i \\ \vdots \\ \sum_{i=1}^{n} x_{ik} y_i \end{bmatrix}$$

Em notação matricial, o modelo ajustado é equação 5 a seguir.

$$\hat{y} = X\hat{\beta} \tag{5}$$

A diferença entre o que se estima e que o é observado é denominado resíduo de regressão e podem ser notados de modo matricial, conforme a equação 6.

$$e = y - \hat{y} \tag{6}$$

2.5 PROPRIEDADE DOS ESTIMADORES

Os erros $\in_i$ são estatisticamente independentes, com média zero e variância σ^2. Sob essas condições, os estimadores de mínimos quadrados são ditos como não tendenciosos (Montgomery & Runger, 2012).

$$E(\hat{\beta}) = \beta \tag{7}$$

Os elementos da inversa da matriz X'X são trazidos como as variâncias dos $\hat{\beta}$'s.

$$V(\hat{\beta}_j) = \sigma^2 C_{jj}, \qquad j = 0, 1, 2 \tag{8}$$

De acordo com Montgomery & Runger (2012), quando σ^2 for trocado por sua estimativa $\hat{\sigma}^2$, a raiz quadrada da variância estimada do *j*- ésimo coeficiente de regressão é chamada de erro-padrão estimado de $\hat{\beta}_j$ ou $ep = \left(\hat{\beta}_j\right)$ $ep = \left(\hat{\beta}_j\right) = \sqrt{\hat{\sigma}^2 C_{jj}}$.

2.6 TESTE DE HIPÓTESE

No modelo obtido na regressão linear múltipla, o teste de hipótese é caracterizado por testar a significância da regressão ao determinar se existe uma contribuição das variáveis independentes, $X_1, X_2, \ldots, X_k$, para explicar a variável resposta Y. As hipóteses apropriadas são:

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_K = 0$$

$$H_1 : \beta_j \neq 0 \text{, para no mínimo um } j \tag{9}$$

O rejeitar $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$ implica que ao menos uma das variáveis independentes contribui de modo significante para o modelo. O teste se fundamenta na composição dos quadrados totais da expressão 10.

$$SQ_T = SQ_R + SQ_E \tag{10}$$

Na qual,

$$SQ_E = y'y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} - \left[\hat{\beta}' X' y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] \tag{11}$$

$$SQ_R = \hat{\beta}' X' y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \tag{12}$$

Uma vez que o H_0 for aceito, SQ_R / σ^2 será uma variável aleatória qui-quadrado, com *k* graus de liberdade. A estatística teste para a H_0,também é conhecida como Teste ANOVA, e é expressa através da equação 13:

$$F_0 = \frac{SQ_R/k}{SQ_E/(n-p)} = \frac{MQ_R}{MQ_E} \tag{13}$$

Se $F_0 > f_{\alpha,k,n-p}$, o procedimento se resume a tabela 1 abaixo:

Tabela 1: Análise de Variância para Teste de Significância
Fonte: Montgomery & Runger, 2012

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade	Média Quadrática	F_0
Regressão	SQ_R	k	MQ_R	MQ_R/MQ_E
Erro ou Resíduo	SQ_E	$n - p$	MQ_E	
Total	SQ_T	$n - 1$		

2.7 R² E R² AJUSTADO

Outro método para avaliar o ajuste do modelo é o coeficiente de determinação múltipla R². Segundo Montgomery & Runger (2012), que se traduz como média quadrática do erro através de:

$$R^2 = \frac{SQ_R}{SQ_T} = 1 - \frac{SQ_E}{SQ_T} \tag{14}$$

A estatística R² de algum modo, não é altamente indicada como uma medida da qualidade do ajuste para um modelo de regressão múltipla, uma vez que sempre tende a aumentar quando uma variável é adicionada em um modelo.

Em decorrência, há certa dificuldade em interpretar pequenos aumentos no R², pois, ao se adicionar uma novo regressor ao modelo, à princípio, não se sabe se o aumento se deve a real contribuição do mesmo ou em decorrência da sua simples adição. Por essa razão, há uma unicidade em preferir estatística denominada R² ajustado (HAIR, ANDERSON, TATHAM, & BLACK, 1998).

$$R^2_{ajustado} = 1 - \frac{SQ_E/(n-p)}{SQ_T/(n-1)} \tag{15}$$

Montgomery & Runger (2012) explica que, apesar de ter a mesma interpretação quanto ao grau de explicação do modelo, a estatística R² ajustado penaliza esse grau pela adição de termos que não são realmente úteis, sendo assim um resguardo contra ajuste e adição excessiva de variáveis.

2.8 TESTE PARA OS COEFICIENTES INDIVIDUAIS

Os testes a seguir visam julgar o valor potencial individual das variáveis no modelo de regressão. Para testar a hipótese de que um coeficiente individual é igual a uma constante, segue:

$$\begin{aligned} H_0 &: \beta_j = \beta_{j0} \\ H_1 &: \beta_j \neq \beta_{j0} \end{aligned} \tag{16}$$

A estatística para tal teste é expressa a seguir,

$$T_0 = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\sigma^2 C_{jj}}} = \frac{\hat{\beta}_j - \beta_{j0}}{se(\hat{\beta}_j)} \tag{17}$$

Região de rejeição:

$$|t_0| > t_{\alpha/2, n-p} \tag{18}$$

Este teste é denominado de teste parcial ou marginal e indica que se H_0 não for rejeitado, a referente variável independente pode ser retirada do modelo, visto que o coeficiente de regressão $\hat{\beta}_j$ depende de todos os outros regressores $x_i (i \neq j)$ que estão no modelo (Montgomery & Runger, 2012).

2.9 INTERVALOS DE CONFIANÇA PARA OS COEFICIENTES DA REGRESSÃO

Nos modelos de regressão, é frequente a utilização de intervalos de confiança para os seus coeficientes. Esses intervalos são definidos com base no pressuposto de que os erros $\{\in_i\}$ são independentemente distribuídos, com distribuição do tipo normal, média zero e variância σ^2.

Um intervalo de confiança de $100(1-\alpha)\%$ para o coeficiente de regressão β_j, j = 0, 1, ... , k no modelo de regressão múltipla é dado por (Montgomery & Runger, 2012)

$$\hat{\beta}_j - t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2,n-p}\sqrt{\hat{\sigma}^2 C_{jj}} \tag{19}$$

A consequência do estimador de mínimos quadrados $\hat{\beta}$ ser uma combinação linear das observações, é o $\hat{\beta}$ possuir distribuição normal com vetor médio β e matriz de covariância $\sigma^2(X'X)^{-1}$.

2.10 PREVISÃO DE VALORES

Dentre as várias utilizações do modelo de regressão linear, se destaca a previsão de novas ou futuras observações para Y relacionada a um conjunto de valores das variáveis independentes. Então, uma estima futura pode ser feito por meio de

$$\hat{y}_0 = X'_0\hat{\beta} \tag{20}$$

$$\hat{y}_0 - t_{\alpha/2,n-p}\sqrt{\sigma(1+x'_0(X'X)^{-1}x_0)} \leq Y_0 \leq \hat{y}_0 + t_{\alpha/2,n-p}\sqrt{\sigma^2(1+x'_0(X'X)^{-1}x_0)} \tag{21}$$

Montgomery (2012), explica que esse intervalo expressa o erro na previsão de uma observação futura no ponto x_0, incluindo o erro na estimação da média naquele ponto, assim como a variabilidade inerente na variável aleatória Y no mesmo valor $X = X_0$.

2.11 ANÁLISE RESIDUAL

Uma das suposições para utilização do modelo de regressão linear diz respeito à normalidade dos erros. Ao determinar um modelo, é expressamente necessário que os resíduos do mesmo confirmem ou, ao menos, não desmintam tal suposição.

Para tanto, os gráficos dos resíduos dos valores estimados contra os valores das variáveis de controle são utilizados para verificar se a variância dos resíduos pode ser considerada aproximadamente constante, além de auxiliar na verificação de linearidade do modelo (Montgomery, 2001; Neter, 2005).

A figura 3 apresenta alguns exemplos de gráficos para análise dos resíduos. A figura 3.a é um exemplo de padrão aceitável, uma vez que os resíduos estão consideravelmente contidos dentro de uma faixa, indicando uma variância aproximadamente constante. Essa variância já não pode ser considerada constante nas figuras 3.b e 3.c. Já a figura 3.d, representa um exemplo no qual o modelo não é adequado, pois não há aleatoriedade nos resíduos (Montgomery, 2001; Neter, 2005).

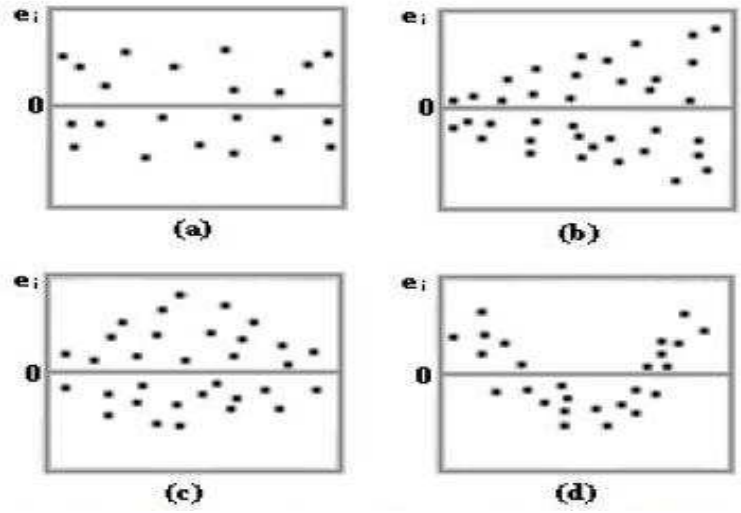


Figura 3: Padrões de comportamento dos gráficos de resíduos
Fonte: Montgomery & Runger (2003)

2.12 ANÁLISE FATORIAL

A análise fatorial (A.F.) é uma técnica estatística que objetiva descrever os inter-relacionamentos entre variáveis por meio de suas covariâncias, identificando a existência de dimensões comuns, denominadas “fator”. Tal definição é abreviada por Reis (1997) como sendo "um conjunto de técnicas estatísticas cujo objetivo é representar ou descrever um número de variáveis iniciais a partir de um menor número de variáveis hipotéticas".

Apesar do agrupamento em fatores, a análise fatorial sempre mantém a natureza e características das variáveis. Quanto à quantidade de fatores, essa pode ser em número igual ou inferior à quantidade de variáveis ou fenômenos analisada, sendo esta última hipótese mais provável.

A Análise Fatorial pode ser definida nos módulos de exploratória e confirmatória, cuja a distinção está entre não saber à priori se há correlação entre as variáveis ou, já sabendo, apenas confirmar a hipótese de relacionamento entre as mesmas. Quanto aos métodos, tem-se Análise de Componentes Principais e Análise Fatorial Comum; que se diferenciam por considerar a variação total ou apenas a variância comum dos dados.

Para analisar as variáveis, há diversos modos, sendo *o R-mode factor analysis* e o *Q-mode factor analysis* os mais comuns. Este, (Corrar, 2009), trata-se de uma análise que agrupa os fenômenos de acordo com a análise das características comuns percebidas pela correlação entre as variáveis, e o primeiro utilizado quando “se busca estruturas subjacentes capazes de ser percebidas apenas pela construção de relacionamentos entre diversas variáveis” (Corrar, 2009).

Um artifício comumente utilizado para melhorar a interpretação dos resultados de uma AF é a rotação dos fatores, também denominada rotação dos eixos, que apesar de não alterar a distância entre os pontos, intensifica o poder de explicativo desses fatores. Em suma, tal rotação pode ser definida como um rearranjo dos autovalores que, por Corrar (2009), pode ser realizada por meio dos seguintes métodos:

- Varimax: Caractertizada como uma rotação ortgonal, é o tipo mais utilizada, pois minimiza a ocorrência de uma variável possuir altas cargas fatoriais para diferentes fatores.
- Quartimax: Rotação ortogonal, que minimiza o número de fatores necessários para explicar cada variável, tendendo a concentrar grande parte das variáveis em um único fator.

- Equimax: Rotação também ortogonal, que agrega as características da Varimax e Quartimax.
- Direct Oblimin: Rotação oblíqua, que permite alcançar autovalores elevados, porém, com aumento da complexidade entre os fatores.
- Promax: Segunda rotação oblíqua, indicada para tratamento de grandes bancos de dados, pois é mais rápida de ser calculada.

Segundo HAIR, ANDERSON, TATHAM, & BLACK (1998), "não existe regras desenvolvidas para guiar os pesquisadores na seleção de um método de rotação", sendo assim, a escolha do método para tal artifício é feita de forma empírica.

Diferentemente das técnicas de dependência que buscam determinar a capacidade de previsão, as técnicas de interdependência, como a Análise Fatorial, tem por intuito apenas elevar o nível de explicação de um conjunto de variáveis. Mas, segundo HAIR, ANDERSON, TATHAM, & BLACK (1998), essa ferramenta consegue auxiliar na regressão múltipla, pois "se uma variável vinculada a um fator é incluída no modelo de regressão, torna-se menos provável que demais variáveis do mesmo fator também sejam, pois são altamente correlacionadas e, potencionalmente, tem menos poder preditivo que as que estão em fatores distintos".

2.13 MULTICONEARIDADE

Um dos pressupostos em um modelo de regressão é a independência entre as variáveis. Hair (2005) descreve que o ideal seria obter diversas variáveis totalmente independentes e altamente correlacionadas com a variável resposta, mas com pouca correlação entre as mesmas. Os efeitos da multicolinearidade são referentes à explicação e estimação que, segundo Corrar (2009), tende a distorcer os coeficientes angulares estimados para as variáveis que a apresentam, prejudicando a habilidade preditiva do modelo e a compreensão do real efeito das variáveis independentes sobre o comportamento da variável resposta.

$$FIV(\beta_j) = \frac{1}{\left(1 - R_j^2\right)} \qquad j = 1,2, \dots , k \qquad (22)$$

A função 22 conhecida como Fator de Inflação da Variância se destaca como uma medida muito útil de multicolinearidade. Segundo Montgomery & Runger (2012), quanto maior for o fator de inflação da variância, maior será a multicolinearidade. Porém, se FIV for pequeno, não há problema nesse sentido com o modelo.

3. EFICIÊNCIA ENERGÉTICA NA MRS LOGÍSTICA

3.1 A MRS LOGÍSTICA S.A.

A MRS Logística, empresa objeto deste estudo, atua no setor de transporte ferroviário de cargas. Segundo informações publicadas pela própria empresa (MRS, 2012), a MRS é a concessionária que controla, opera, monitora a Malha Sudeste da antiga rede ferroviária Federal a (RFFSA), interligando os estados do Rio de Janeiro, Minas Gerais e São Paulo, região que concentra aproximadamente 65% do produto interno bruto do país.

A empresa atua no mercado de transportes desde setembro de 1996, quando foi constituída operando nas antigas Estrada de Ferro Brasil e Estrada de Ferro Santos a Jundiaí. Atualmente, as linhas sob sua concessão denominam-se: Linha do Centro, Ferrovia do Aço, Linha São Paulo e Linha Santos-Jundiaí; cujas características estão descritas na tabela 2 a seguir:

Tabela 2: Principais Linhas Operadas pela MRS Logística – Características
Fonte: Brugiolo, I. (2008)

LINHA	INÍCIO	TÉRMINO	EXTENSÃO (km)	DESCRIÇÃO
LINHA DO CENTRO	Belo Horizonte (MG)	Rio de Janeiro (RJ)	566	Uma das mais antigas linhas férreas em operação o Brasil, foi reformada na década de 80. Esta linha é utilizada como via de retorno a Minas Gerais dos trens vazios de Heavy Haul
FERROVIA DO AÇO	Itabirito (MG)	Barra Mansa (RJ)	370	Linha férrea recente, começou a operar em 1989. É a mais movimentada linha da Malha Sudeste, sendo utilizada, principalmente, para o transporte do minério de ferro produzido na região próxima de Belo Horizonte até os portos de Guaíba e Sepetiba, no Rio de Janeiro, e as unidades siderúrgicas da CSN e Cosipa.
LINHA SÃO PAULO	Barra do Piraí (RJ)	São Paulo (SP)	400	Nesta linha está incluído o trecho entre saudade e Barra do Piraí (45 Km), que faz conexão tanto com Ferrovia do Aço (Saudade) quanto com a Linha do Centro (Barra do Piraí) e termina por ser um dos trechos mais movimentados da Malha Sudeste
LINHA SANTOS-JUNDIAÍ	Santos (SP)	Jundiaí (SP)	139	Utilizada fundamentalmente para o transporte de Carga Geral das regiões agrícolas do estado de São Paulo para o porrto de Santos, e de contêineres entre o terminal de Jundiaí e o terminais portuários em Santos. A capacidade desta linha é bastante limitada pela necessidade de transpor a Serra do Mar através da utilização do sistema de cremalheira, com o uso de locomotivas elétricas e de um mecanismo de tração situado entre os trilhos.

Essas linhas compõem a malha ferroviária, ilustrada na figura 4, que dá acesso ao Porto do Rio de Janeiro, Porto de Guaíba, Porto de Itaguaí e Porto de Santos, sendo este o mais relevante dentre os portos da América Latina.

Figura 4: Malha Ferroviária da MRS Logísticas e Postos Acessados
Fonte: MRS, 2013

O foco das atividades da empresa está no transporte ferroviário de cargas classificadas como Heavy Haul e Carga Geral. O HH pode definido como “carga pesada”, cujo transporte é feito pelos trens denominados “unitários”, que são formados por um tipo específico de vagão; que carregam um único tipo de carga; de um único cliente; sendo assim, são trens que obedecem a um padrão de composição, carga e fluxo. Já a CG não apresenta padrão pré-definido em nenhum critério, seus trens são ditos como mistos e sua demanda é totalmente influenciada pelas oscilações dos produtos transportados; como exemplo, os produtos agrícolas.

3.2 CONSUMO DE COMBUSTÍVEL NA MRS LOGÍSTICA

Não é exclusividade da MRS, ter o maior custo operacional ferroviário advindo da conta de combustível diesel; em meses de alta produção, a empresa chega a consumir mais de 26 milhões de litros. Além da quantidade de carga transportada, diversos são os fatores que influenciam no consumo de diesel, esses vão desde operacionais até fatores ambientais; como as condições climáticas.

Em suma, os ambientais estão relacionados ao fato de que a umidade nos trilhos faz com que as rodas da composição, locomotivas e vagões, diminuam o atrito roda-trilho levando a rotação dos rodeiros, gerando certo consumo de diesel, mas sem obtenção de deslocamento.

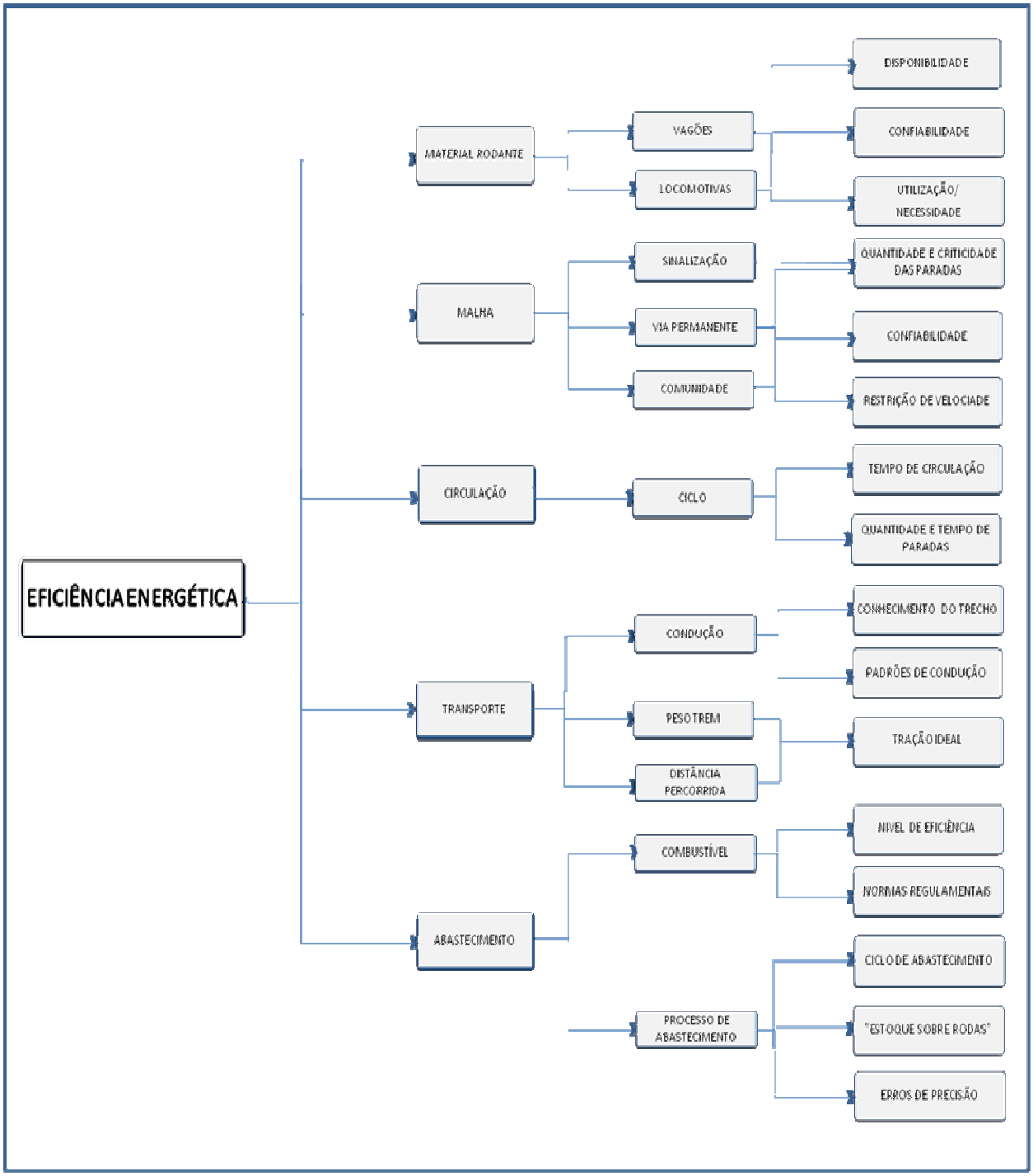
Ao se restringir para os operacionais, tem-se dificuldade em enumerar todos os impactos ocasionados por variações nesses fatores, sendo ainda que alguns são responsáveis por mais de um impacto e, por sua vez, alguns impactos possuem mais de uma origem. Entretanto, à caráter ilustrativo, vale citar alguns exemplos que são agrupados em cinco macro grupos:

- Material Rodante: diz respeito aos principais ativos da empresa: locomotivas e vagões. Ambos estão relacionados à confiabilidade, disponibilidade, pois uma vez que um trem é parado por problemas na composição; esse consome uma quantidade considerável de diesel para ser novamente colocado em movimento, evento físico denominado “demarragem”; além de influenciar na circulação de todos os demais trens que seguem o mesmo fluxo.

  No que se refere às locomotivas, a frota da MRS é composta de aproximadamente 30 modelos, sendo que os mesmos são qualificados em 10 grupos de acordo com a sua EE e capacidade de tração, ou seja, locomotivas de modelos denominados AC-44I e C-44EMI são tidas como mais eficientes por possuírem menor consumo específico em cada ponto de aceleração e pela alta capacidade de esforço trator.

- Malha: pode-se associar a malha ao efeito da sinalização; visto que sinais vermelhos têm por efeito a demarragem, bem como com a via permanente; pois más condições da via afetam não somente a EE como toda a operação logística, limitando fatores como peso máximo do trem e velocidade máxima permitida.

- Formação dos Trens: devido aos 30 tipos de locomotivas, diversas são as combinações possíveis para a composição de tração dos trens, porém existem combinações ideais que diz respeito à força e velocidade necessária *versus* o perfil do trecho, peso a ser transportado, disponibilidade de ativos, dentre outros.
- Plano de Viagem: neste, um dos fatores mais relacionados ao consumo de combustível é a condução do maquinista. Pela própria complexidade das locomotivas, perfis topográficos e interações com toda a operação; os condutores precisam equalizar a dinâmica da variação da velocidade, evitando, por exemplo, frenagens bruscas e aproveitando ao máximo os momentos de inércia.
- Mix de Transporte: faz alusão à origem, destino e tipo de carga. A maximização da eficiência energética se dá através combinação entre a carga mais pesada, percorrendo a maior distância possível, ou seja, melhor aproveitando do diesel para realizar a mesma, ou melhor, TKB.
- Abastecimento: quanto ao abastecimento, a análise se torna intrínseca, pois é perceptível que qualquer desvio, desperdício, durante o processo de abastecimento ou o até mesmo o não controle sob o ciclo de abastecimento que poderá gerar o “estoque sobre rodas” – quantidade de combustível no tanque da locomotiva que não irá ser gasta e, portanto, não irá gerar TKB – atuam de forma a acrescentar na conta de consumo de diesel.



Quadro 1: Variáveis de impacto na EE
Fonte: MRS, 2012 (Adaptado)

Uma visão mais completa dos grupos dos fatores supracitados é ilustrada no quadro 1. Aos mesmos ainda se poderia somar os relacionados à qualidade do diesel, como: viscosidade, bactérias, borras, dentre outros. Todos esses fatores influenciam no consumo de diesel e, sendo assim, na eficiência energética; logo, quanto mais otimizada for a combinação, de modo a proporcionar a melhor interação dos mesmos, mais eficiente se torna a operação.

O gráfico, figura 5, apresenta a evolução da EE da empresa, deixando perceptível a tendência de declínio, ou seja, a melhora da operação, devida a inserção de tecnologias e otimização dos recursos, obtida no decorrer dos últimos anos.

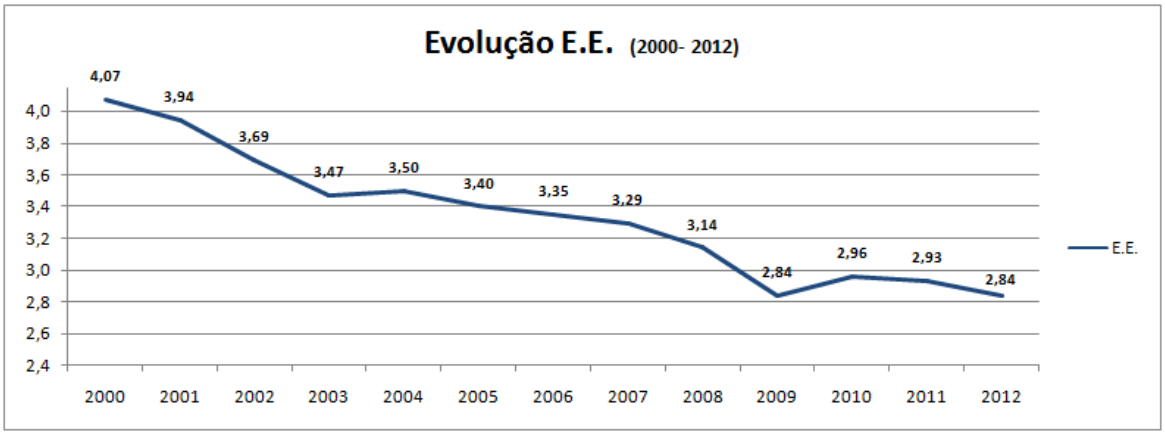


Figura 5: Evolução EE
Fonte: MRS, 2013

4. DESENVOLVIMENTO

Este capítulo visa demonstrar os procedimentos, bem como, a determinação da equação de regressão do problema proposto. Inicia-se com avaliação, seleção e identificação das principais influenciadoras no valor do indicador, com posterior coleta de dados históricos dessas potenciais variáveis do modelo. Tais dados são analisados de forma qualitativa e quantitativa, visando identificar seu grau de impacto na Eficiência Energética da MRS através de testes estatísticos.

4.1 VARIÁVEIS DO MODELO

Dentre as inúmeras variáveis que impactam na E.E. da MRS Logística, ilustradas no quadro 1, limitando-se obviamente às variáveis quantitativas e que, por definição poderiam compor o modelo à ser desenvolvido; optou-se pela Utilização de Locomotivas, Transit Time Carregado, Tonelada Útil, Quantidade de Tabelas e Modelo de Tração dos trens de minério de ferro exportação. À priori, essas variáveis foram escolhidas devido às suas correlações intrínsecas com o indicador, observadas na habitual apuração do mesmo.

VARIÁVEL	EXTRATIFICAÇÃO/ JUNÇÃO	EIXO	DEFINIÇÃO	UNIDADE
Utilização	Utilização AC+C Utilização Outros	Ativos	Indica a quantidade de tempo que cada locomotiva ficou tracionando, ou seja, realizando algum esforço traçor. Os dados são separados em grande dois grupos, o Grupo AC+C que são os dois modelos de locomotivas mais semelhantes e eficientes e o Outros, que engloba todos os demais modelos. Na apuração mensal, são usadas as médias desses dados.	LocoxHora
Transit Time Carregado	-	Circulação	Caracteriza-se por ser umas das 12 etapas do ciclo de circulação do trem. É o tempo gasto por este, para realizar o trajeto desde o pátio de carga (após o carregamento) até o pátio de descarga (antes do descarregamento).	Hora
Tonelada Útil (T.U.)	TU/Tabela	Produção	Tonelada Útil é quantidade, em toneladas, de minério depositada nos vagões. Já a Quantidade de Tabela, é o número de trens utilizados para transportar essa T.U. O valor é o acumulado mensal.	Toneladas/Tabela
Quantidade de Tabela				
Modelo de Tração	-	Produção	Variável discreta, também conhecida como Dummy, que no horizonte utilizado, serve para marcar a mudança de trens compostos por 132 para 134 vagões.	-

Quadro 2: Definição das variáveis do modelo proposto

A figura 6 apresenta essas variáveis, sua definição, bem como o eixo qualitativo associado. Essas, serão as representantes dos principais eixos da operação ferroviária da MRS; ativos, circulação e produção.

Vale ressaltar que existem outras variáveis, conforme já citado nesse trabalho, cuja correlação com a E.E. é evidente. Porém, a impossibilidade está relacionada a sua quantificação ou inexistência de dados históricos. Como exemplo prático pode-se citar a condução do maquinista. É esperada a afirmação que o modo de operação do maquinista- utilização adequada dos pontos para esforço de tração de acordo com o trecho, aproveitamento da dinâmica ocasionada pelo perfil topográfico da linha ferroviária; nível de frenagem brusca e alcance de alta velocidade em pequeno intervalo de tempo- influenciam diretamente no consumo de diesel e, consequentemente, no indicador estudado. Neste caso, o empecilho em empregar tal variável está na dificuldade de mensurar que, consequentemente, ocasiona na inexistência de dados históricos.

Além da necessidade de ser mensurável, outro pressuposto indispensável para que uma variável seja candidata, é que existam meios de orçá-la, sendo necessário que se consiga fazer previsões de médio prazo da mesma, para que, em conjunto, as variáveis sejam inseridas na equação e se consiga prever a E.E.; visto que o modelo é a ferramenta base de orçamento de diesel da empresa.

4.2 COLETA DE DADOS

De acordo com Moreira (1997), “A palavra ‘dados’ tem sido definida como aquela espécie de informação organizada para análise ou usada para comparação e base para uma decisão”. Para o modelo de regressão, utiliza-se nesse trabalho dados operacionais mensais disponíveis do período compreendido entre 01/01/2009 e 31/07/2012. Esse período se justifica por ter um padrão de operacional mais próximo do atual e do pretendido pela empresa.

A priori, se primará pela utilização dos dados em sua forma habitual de tabulação. As transformações dos mesmos só serão realizadas em casos de necessidade.

4.3 ANÁLISE INDIVIDUAL DAS VARIÁVEIS

> Se o pesquisador confia cegamente nas técnicas de análise multivariada para encontrar as respostas de suas questões sem ao menos atentar para as propriedades fundamentais dos dados que serão analisados, aumenta-se o risco de problemas sérios, tais como: uso indevido das técnicas estatísticas, violação de propriedades estatísticas e interpretação inadequada dos resultados, (Corrar, 2009).

Valendo-se dessa ressalva, antes da aplicação da técnica multivariada, serão realizados testes estatísticos a fim de avaliar os comportamentos das amostras e de suas inter-relações, isso como medida preventiva de possíveis erros, que possam vir a alterar o resultado ou proporcionar uma análise errônea.

4.3.1 Correlação com a EE

O coeficiente de correlação (r) representa numericamente o quanto as variáveis estão relacionadas entre si, sendo uma importante medida na análise de modelos de regressão (BARBETTA, 2009). Ainda, segundo Dantas (1998), os valores possíveis de r variam de -1 até +1, sendo que os sinais '-' e '+' indicam o sentido da correlação, negativa ou positiva, respectivamente.

Tabela 3: Correlação individual das variáveis dependentes com a EE

VARIÁVEL	EXTRATIFICAÇÃO/ JUNÇÃO	CORRELAÇÃO COM A EE
Utilização	Utilização AC+C	-0,24
	Utilização Outros	0,72
Transit Time Carregado	-	0,69
Tonelada Útil (T.U.)	TU/Tabela	-0,52
Quantidade de Tabela		
Modelo de Tração	-	0,33

A tabela 3 apresenta a correlação individual de cada variável independente escolhida com o indicador estudado. Valores intermediários, em módulo, são relacionados por Dantas (1998) na tabela 4 abaixo. Com base dessas informações, é perceptível que, individualmente, todas as variáveis escolhidas para compor o modelo da EE possuem correlações de fraca à forte com o mesmo. Essa primeira análise evidencia apenas que as variáveis são realmente candidatas ao modelo, e que suas variações particulares, explicam parte da variação da EE. Contudo, alerta-se que esses valores de correlação tenderão a se alterarem pelas interações entre as variáveis preditoras quando inseridas no modelo.

Tabela 4: Grau de correlação entre as variáveis
Fonte: Dantas, 1998

COEFICIENTE	CORRELAÇÃO
$\|r\| = 0$	Nula
$0 < \|r\| \leq 0{,}30$	Fraca
$0{,}30 < \|r\| \leq 0{,}70$	Média
$0{,}70 < \|r\| \leq 0{,}90$	Forte
$0{,}90 < \|r\| \leq 0{,}99$	Fortíssima
$\|r\| = 1$	Perfeita

4.3.2 Diferença entre as variáveis

Torna-se interessante compreender as distribuições dos valores em cada grupo de variável e as diferenças existentes entre os valores. O Diagrama de Extremos-e-Quartis (BoxPlot) é um das ferramentas estatísticas utilizada para esse fim, para Cooper e Schindler (2003) "os gráficos de caixas são extensões do resumo de cinco números. Esse resumo consiste na mediana, quartis superiores e inferiores e maior e menor observação".

Quando se aplica tal ferramenta nos grupos de dados analisados, figura 7, fica perceptível que, apesar de ter pontos definidos como *Outlier*, os valores de TU/Tabela são simétricos com a pouca dispersão em relação à média. A simetria também é percebida no TTC, porém neste, há uma maior variação dentre os valores. A Utilização AC+C e Utilização Outros, têm nível de variância considerável e assimetria com direcionamento mínimo e máximo, respectivamente.

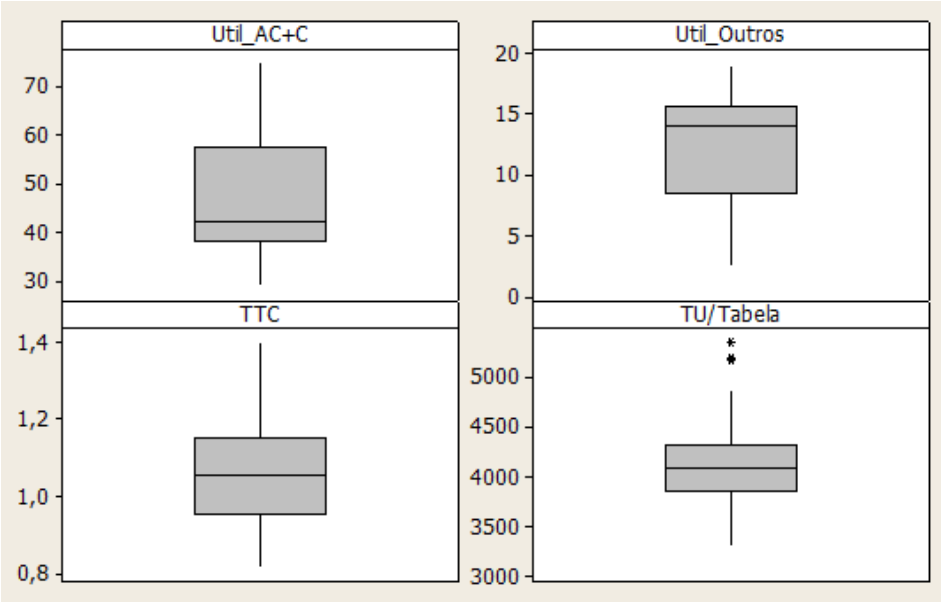


Figura 6: Diagrama de Extremos-e-Quartis

4.3.3 Relação entre as variáveis

Sabe-se que a inexistência de correlação entre as variáveis que são ditas como independentes é fundamental para obtenção de um modelo de regressão plausível. HAIR, ANDERSON, TATHAM, & BLACK (1998) afirmam que o método mais popular para examinar relações bivariadas é o diagrama de dispersão. Neste trabalho, tal ferramenta é disposta como figura 8 que apresenta os gráficos de dispersão em pares entre as preditoras, possibilitando verificar a inexistência de padrão dos pontos, ou seja, dispersão aleatória para alguns dados, porém, quando se trata da utilização de Outras Locomotivas (Util_Outros) e Transit Time Carregado (TTC), percebe-se que há uma certa tendência de relação linear.

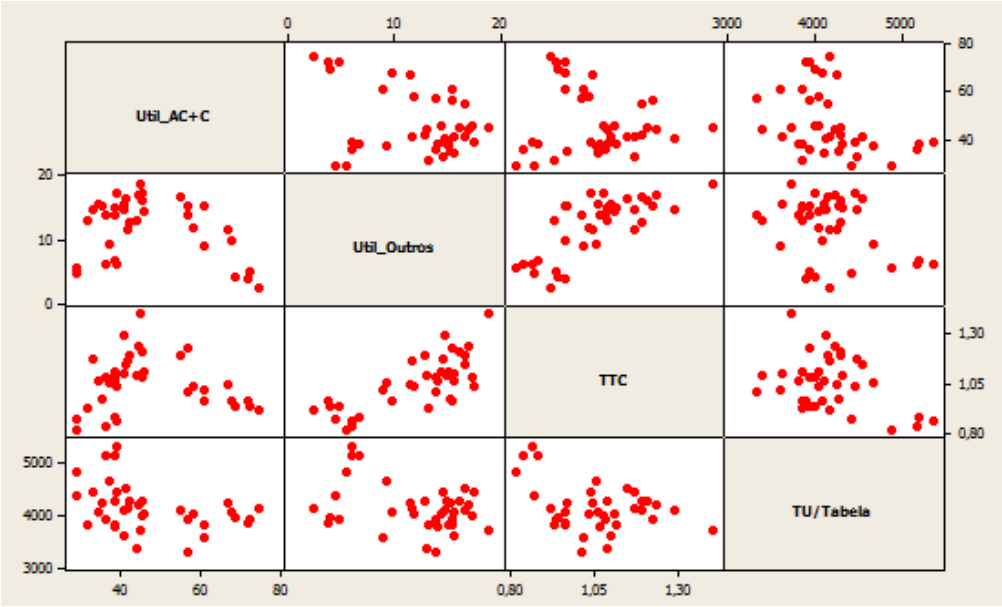


Figura 7: Gráfico de dispersão entre as variáveis independentes

É possível inferir que à medida que o TTC aumenta, torna-se necessária o aumento da utilização de locomotivas menos eficiente para evitar a perda de produção ou acúmulo das mesmas nos pontos de carregamento; pois, estrategicamente não é interessante priorizar a EE em detrimento da produção. Para complementar tal suposição, realizou-se uma análise quantitativa da relação entre as variáveis independentes expressa na tabela 5, que confirma a correlação considerável entre as duas variáveis.

Tabela 5: Correlação entre as variáveis "independentes"

	TTC	*Util_AC+C*	*Util_Outros*	*TU/Tabela*
TTC	1			
Util_AC+C	-0,0865	1		
Util_Outros	0,746167	-0,32739	1	
TU/Tabela	-0,36897	-0,37627	-0,34315	1

Devido o grau de influência, tornou-se complexa a busca por variáveis que sejam totalmente independentes dentre as controladas pela MRS Logística e que tenham alta correlação com a Eficiência Energética, além de conter histórico e orçamento anual. Portanto, optou-se por permanecer com as já definidas. O impacto dessa quebra de pressuposto será avaliado de modo mais preciso através da análise das "saídas" da regressão.

5. MODELO DE REGRESSÃO

5.1 MODELO DE REGRESSÃO – VARIÁVEIS QUALITATIVAS (I)

Inicialmente, o modelo gerado contempla as variáveis elegidas pela junção das análises estatísticas e exploratórias, objetivando uma análise mais completa vista a complexidade do indicador estudado. O resultado obtido pela regressão múltipla tendo como preditoras a Utilização AC+C, Utilização Outros, TTC, TU/Tabela e Modelo de Tração, é expresso abaixo:

EQUAÇÃO MATEMÁTICA:

EE= 3,03 + 0,268TTC – 0,00368Util_AC+C + 0,00099Util_Outros – 0,000069TU/Tab + 0,0570Mod_de_Tração

S = 0,0219077 R-Sq = 93,9% R-Sq(adj) = 92,4%

Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	5	0,148489	0,029698	61,88	0,000
Residual Error	20	0,009599	0,000480		
Total	25	0,158088			

Durbin-Watson statistic = 2,22056

Predictor	Coef	SE Coef	T	P	FIV
Constant	3,02871	0,07768	38,99	0,000	
TTC	0,26833	0,05543	4,84	0,000	2,742
Util_AC+C	-0,0036759	0,0005138	-7,15	0,000	2,335
Util_Outros	0,000995	0,00159	0,63	0,538	2,960
TU/Tabela	0,00006924	0,00001103	-6,28	0,000	1,551
Mod_de_Tração_1	0,05698	0,01438	3,96	0,001	2,387

Quadro 3: Modelo de Regressão Qualitativo I

A análise do modelo gerado implica que pelo valor da soma dos quadrados (R^2) obtido, o modelo possui grande nível de adequação, uma vez que 93,9% da variabilidade da E.E. pode ser explicada pela interação das cinco variáveis preditoras. Sabendo que o R^2 ajustado possui a mesma interpretação, porém, com maior rigorosidade quanto ao número de preditoras, é possível corroborar a adequação do modelo pelo R^2 ajustado de 92,4%.

O bom ajuste do mesmo também é evidenciado pelo Teste da Distribuição F com valor de 68,88. Já quanto à validação, o Teste ANOVA rejeitou a hipótese de insignificância conjunta das variáveis com um valor de P igual a 0,000, afirmando a significância do modelo.

Contudo, quando se analisa individualmente as variáveis, verifica-se que o mesmo teste ANOVA indicou um P de 0,538 para a Utiliz_Outros, advertendo que esta variável não deveria compor o modelo. O Teste T também reafirma tal suposição visto que seu valor, em módulo, é inferior a 2 somente para essa variável.

O fato do Teste F ter sido significativo para regressão e não significativo para um dos coeficientes individuais indica que pode haver multicolinearidade no modelo. Logo, como já observado que há uma considerável correlação entre a Utilização Outros e o TTC, assume-se que realmente a multicolinearidade está presente.

A multicolinearidade e correlação entre as duas variáveis pode ser evidenciada também pela análise fatorial, que visa congregar várias variáveis em grupos de fatores representando eixos, que nesse caso fazem alusão à circulação, produção e ativos. A rotação utilizada após a extração dos fatores é a *Varimax* e os resultados obtidos são indicados na tabela 6 abaixo:

Tabela 6: Análise de Componentes de Fatores

Rotated Factor Loadings
Varimax Rotation

Variable	Factor1	Factor2	Factor3
TTC	**0,933**	-0,150	-0,089
Util_AC+C	-0,276	**-0,890**	-0,188
Util_Outros	**0,898**	0,180	-0,239
TU/Tabela	-0,232	0,313	**0,917**
Mod_de_Tração_1	0,254	**-0,885**	-0,219
Variance	1,870	1,727	0,989

Com exceção da variável Modelo de Tração que é de caráter qualitativo, é perceptível que há duplicidade na representação do eixo de circulação (Fator 1), causada justamente pelo TTC e Utilização outros. O que corrobora também os pontos já demonstrados.

O problema relacionado à multicolinearidade, é que a variável que a provoca, não diminui o erro quadrático do modelo de modo a "compensar" sua entrada. Quando a regressão é realizada em etapas, por meio da seleção progressiva – processo no qual os regressores são adicionados um a um - a Utilização de Outros nem chega a ser adicionada ao modelo (tabela 7-a), quando se força a entrada da mesma (tabela 7-b), qualquer variável que se adicione eleva o P-*value* para um valor bem próximo ou superior ao seu limite de aceitação.

Tabela 7: Regressão por Etapas

Step	1	2	3	4
Constant	2,383	2,66	2,974	3,03
TTC	0,489	0,427	0,368	0,292
T-Value	6,65	6,21	7,88	7,2
P-Value	0	0	0	0
TU/Tabela		-0,00005	-0,00008	-0,00007
T-Value		-2,8	-6,13	-6,81
P-Value		0,01	0	0
Util_AC+C			-0,00263	-0,0038
T-Value			-5,5	-8,16
P-Value			0	0
Mod_de_Tração_1				0,057
T-Value				4,06
P-Value				0,001
S	0,0482	0,0425	0,0282	0,0216
R-Sq	64,79	73,72	88,94	93,81
R-Sq(adj)	63,32	71,44	87,43	92,63
Mallows C	94	66,6	18,4	4,4

a

Step	1	2	3	4	5
Constant	2,744	2,489	2,735	2,974	3,029
Util_Outros	0,0129	0,0063	0,0056	0,0014	0,001
T-Value	5,89	2,16	2,17	0,66	0,63
P-Value	0	0,042	0,041	0,519	0,538
TTC		0,317	0,279	0,336	0,268
T-Value		3,02	2,98	4,88	4,84
P-Value		0,006	0,007	0	0
TU/Tabela			-0,00005	-0,00008	-0,00007
T-Value			-2,79	-5,59	-6,28
P-Value			0,011	0	0
Util_AC+C				-0,00247	-0,00368
T-Value				-4,58	-7,15
P-Value				0	0
Mod_de_Tração_1					0,057
T-Value					3,96
P-Value					0,001
S	0,0519	0,0449	0,0394	0,0286	0,0219
R-Sq	59,1	70,72	78,35	89,16	93,93
R-Sq(adj)	57,39	68,17	75,4	87,1	92,41
Mallows Cp	112,7	76,5	53,3	19,7	6

b

Apesar de alguns autores, como Corrar (2009), afirmarem que a multicolinearidade não prejudica um modelo, desde que o mesmo seja empregado apenas para fazer previsões; comprovou-se que esse modelo, ao ferir algumas premissas da regressão linear múltipla, não atende as condições matemáticas necessárias para ser utilizado também no controle do indicador.

5.2 MODELO DE REGRESSÃO – VARIÁVEIS QUALITATIVAS (II)

De posse das saídas do modelo I, idealizou-se um novo modelo que tivesse um R^2 aceitável, atendesse aos requisitos da regressão e fosse plausível na prática. Ao realizar a busca por um subconjunto ideal, dentre o conjunto de variáveis já definidas, optou-se pela retirada da Utilização_Outros, visto que essa variável representa em duplicidade o eixo de circulação, tem considerável correlação com o TTC e coeficiente insignificante ao modelo.

Logo, um "segundo novo modelo" teria a Utilização_AC+C, TTC, TU/Tabela e Modelo de Tração como variáveis independentes. A presunção de que essa combinação é o melhor subconjunto é corroborada via ferramenta estatística, que tem seu resumo expresso na tabela 8.

Tabela 8: Regressão – Melhores Subconjuntos

Nº	Vars	R-Sq	R-Sq(adj)	Mallows Cp	S	Util_AC+C	Util_Outros	TU/Tabela	Mod_de_Tração	TTC
I	1	64,8	63,3	94	0,04816					X
II	1	59,1	57,4	112,7	0,051907		X			
III	2	73,7	71,4	66,6	0,0425			X		X
IV	2	70,7	68,2	76,5	0,044864		X			X
V	3	88,9	87,4	18,4	0,028189	X		X		X
VI	3	80,2	77,5	47,4	0,037761	X			X	X
VII	4	93,8	92,6	4,4	0,021588	X		X	X	X
VIII	4	89,2	87,1	19,7	0,028561	X	X	X		X
IX	5	93,9	92,4	6	0,021908	X	X	X	X	X

A tabela mostra que o modelo poderia ser dado por meio da Regressão Linear Simples sendo composto apenas pelo TTC ou pela Utilização Outros com R^2 de 64,8% e 59,1%, respectivamente. Um modelo simples, apesar de ter uma gestão facilitada, não é ideal, pois não explica grande parte da variação do indicador. Ao adicionar mais variáveis, a análise se dá através da estatístca R^2 ajustado que, ainda pelo quadro, apresenta como o R^2 máximo, o da interação número VII com valor de 92,6% e menor Mallows Cp. O quadro 3 resume o modelo de regressão – variáveis qualitativas II.

EQUAÇÃO MATEMÁTICA:

EE= 3,03 + 0,292 TTC – 0,00380 Util_AC+C – 0,000071TU/Tab + 0,0575 Mod_de_Tração

S = 0,0215881 R-Sq = 93,8% R-Sq(adj) = 92,6%

Analysis of Variance					
Source	DF	SS	MS	F	P
Regression	4	0,15	0,04	79,55	0,00
Residual Error	21	0,01	0,00		
Total	25	0,16			

Durbin-Watson statistic = 2,12793

Predictor	Coef	SE Coef	T	P	VIF
Constant	3,0296	0,07654	39,58	0,000	
Util_AC+C	-0,0038016	0,000466	-8,16	0,000	1,979
TU/Tabela	-0,00007113	0,00001045	-6,81	0,000	1,435
TTC	0,29161	0,0405	7,2	0,000	1,507
Mod_de_Tração_1	0,05749	0,01415	4,06	0,001	2,379

Quadro 4: Modelo de Regressão II

Analisando as saídas obtidas do modelo II, encontrou-se um R^2 de 93,8% e um R^2 ajustado de 92,6%, que indica que o modelo tem grande capacidade de explicar a EE através das quatro variáveis utilizadas. Pela simples análise do R^2 ajustado, matematicamente, o modelo é altamente aceito.

Um adendo interessante é a comparação entre o modelo I e o II, nesta fica perceptível que o primeiro possui um R^2 superior ao segundo, porém tal ‘superioridade’ é enganosa, visto que o R^2 ajustado é inferior. Isso indica que a Utilização de Outros no Modelo não produz uma redução no quadrado dos erros de modo a compensar a sua presença no modelo e consequente perda no grau de liberdade. A correta retirada dessa variável também é justificada pela menor diferença entre R^2 e R^2 ajustado.

O último modelo ainda possui um bom ajuste evidenciado através do Teste de Distribuição F de 79,55. O Teste ANOVA foi bem sucedido tanto para a regressão do modelo

com P 0,000, quanto para as variáveis individuais, rejeitando-se, com 95% de significância, a hipótese de que $\beta_j = 0$. O fato da regressão e das variáveis não se contradizerem quanto ao valor de P, possibilita uma conclusão preliminar de que não há multicolinearidade nesse modelo.

Os três pressupostos para a validação de um modelo de regressão são a homoscedasticidade, autocorrelação e também a multicolinearidade; essas são análises relacionadas aos resíduos que servem como qualificadoras para a regressão. Para esse modelo, tem-se:

- A normalidade dos resíduos é fundamental para que as informações quanto ao ajuste do modelo sejam confiáveis. O modo de verificação empregado é o Gráfico de Papel de Probabilidade, figura 9, na qual os pontos estão seguindo com proximidade à reta que representa a normal, dando indícios de que erros são normalmente distribuídos.

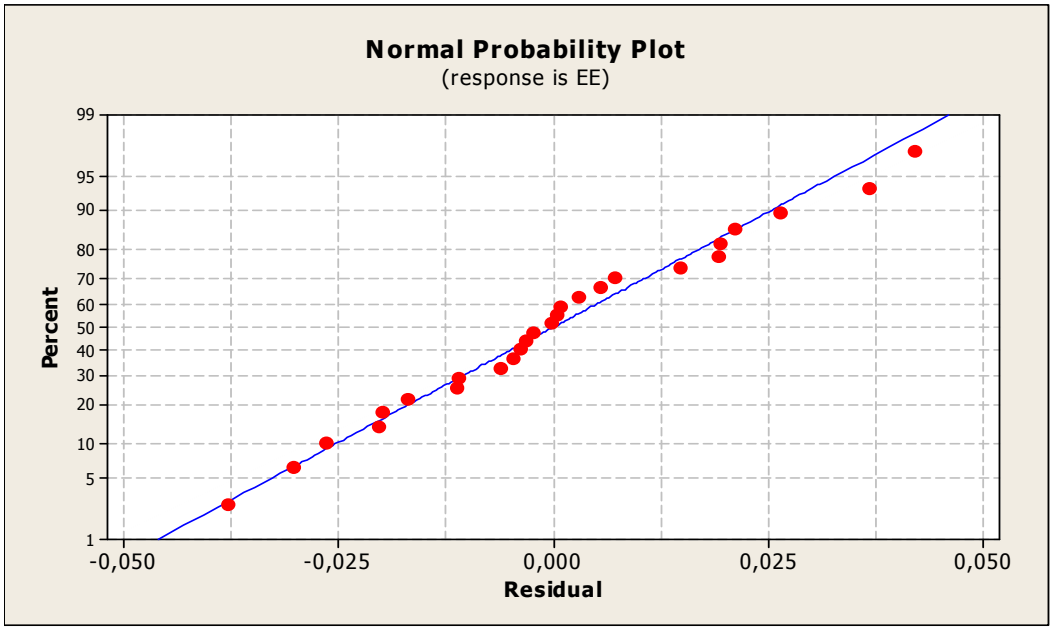


Figura 8: Teste de Normalidade dos Resíduos

- A representação em histograma na figura 10 confirma essa indicação, visto que de modo geral, apresenta uma distribuição equilibrada, apesar de possuir uma concentração minimamente superior no lado esquerdo e evidente pico ocasionado por acumulo de valores intermediários, mas, contudo, com características similares à uma normal.

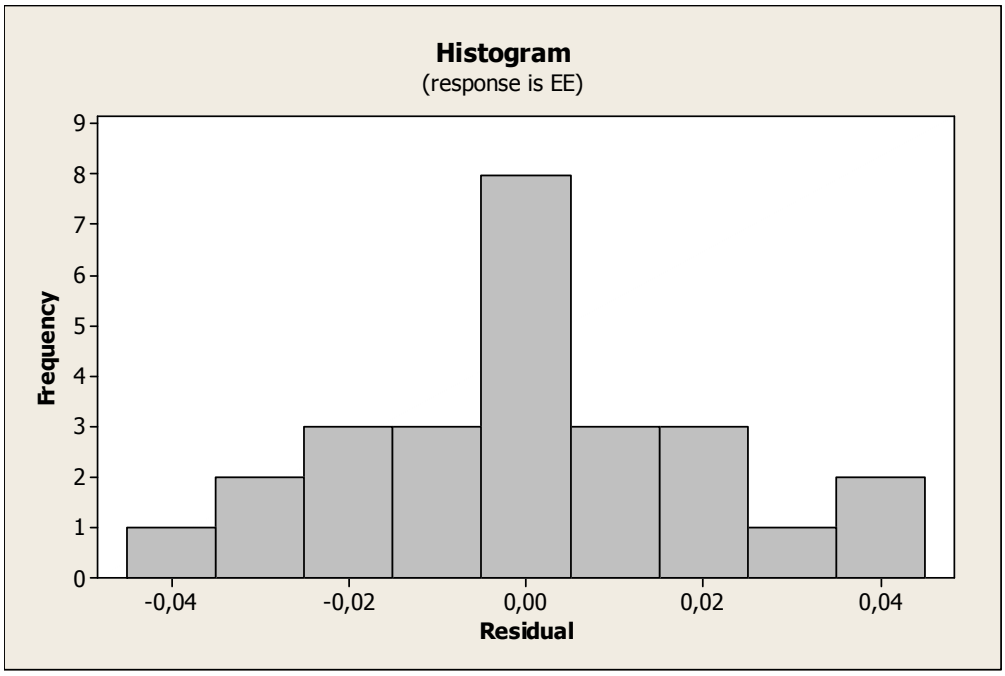


Figura 9: Histograma dos Resíduos

- O diagnóstico quanto à homoscedasticidade é realizado, neste estudo, através do gráfico da figura 11. Nele se nota uma aleatoriedade dos resíduos, ou seja, dispersão homogênea dos pontos, evidenciando que condição de homoscedasticidade foi satisfeita.

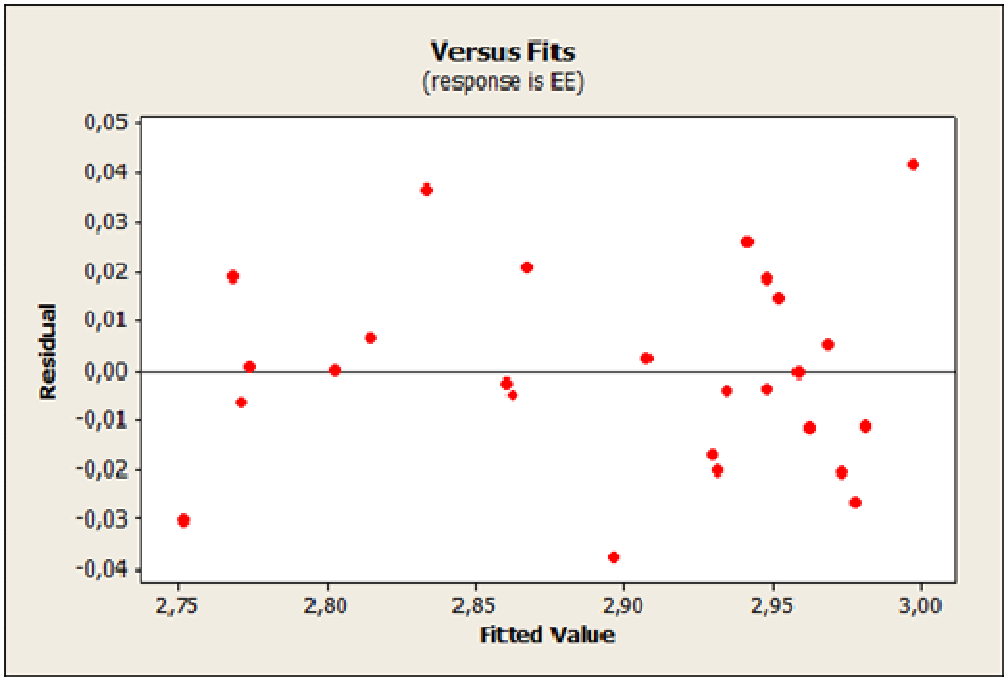


Figura 10: Gráfico de Dispersão Residual - Modelo II

- O Fator de Inflação da Variância (FIV), segundo alguns autores, só evidência problema de multicolinearidade quando apresenta valor superior a 10, para outros, valores superiores a 4 ou 5 já são suficientes. Neste modelo, o maior valor obtido é de 2,379, o que implica na inexistência de problemas, também indicada pelos valores de P já explicitados.
- A autocorrelação entre os resíduos influenciam de forma negativa o poder de estimação dos parâmetros de um modelo de regressão, prejudicando assim, a interpretação dos Testes Te F. A ausência de autocorrelação deste modelo será examinada através do Teste Durbin-Watson (DW), cujo valor obtido via estatística computadorizada é de 2,12793. Segundo tabela DW e considerando tamanho da amostra e quantidade de variáveis independentes, os intervalos de análise para indicação de existência ou inexistência de autocorrelação são indicados no gráfico em figura 12 que, segundo valor encontrado, permite indicar ausência de autocorrelação.

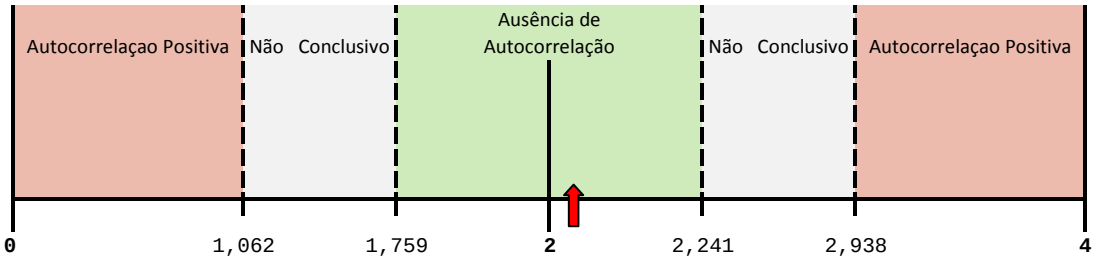


Figura 11: Auto Correlação Durbin-Watson - Modelo II

Satisfeitas todas as premissas e pressupostos do modelo de regressão linear, o modelo desenvolvido, no âmbito matemático, é dito com válido e de boa qualidade.

5.3 MODELO DE REGRESSÃO – COMBINAÇÃO DE VARIÁVEIS (III)

Diversos autores defendem a realização de tratamentos matemáticos na tentativa de diminuir o nível de multicolinearidade de um modelo. Corrar (2009), anuente da teoria supracitada, indica vários métodos possíveis de correção para a multicolinearidade, dentre eles: omitir uma ou mais variáveis independentes altamente correlacionadas, identificar outras para ajudar na previsão e/ou transformar a relação funcional. A primeira ação indicada foi empregada para alcance do modelo II, a segunda e terceira serão aproveitadas na tentativa de gerar um terceiro modelo.

Ainda analisando os três eixos de operação, as variáveis candidatas, e as demais expressas no quadro 1, idealizou-se acrescentar uma nova variável que também é pertencente à circulação, sendo igualmente uma das etapas de composição do ciclo. Essa variável é denominada Fila do Carregado (VCC) e indica o tempo que o trem carregado aguarda em fila para ser descarregado nos portos. O VCC possui correlação negativa com o TTC, uma vez que quanto menor o tempo de circulação, considerando à priori a velocidade dos viradouros como constante, maior será o tempo em fila dos trens carregados; e como consequência, maior será a utilização de outras locomotivas, pela finalidade de substituir as locomotivas eficientes retidas temporariamente na área de descarga. O objetivo é transformar o TTC e VCC, em uma variável agregada para representar a circulação.

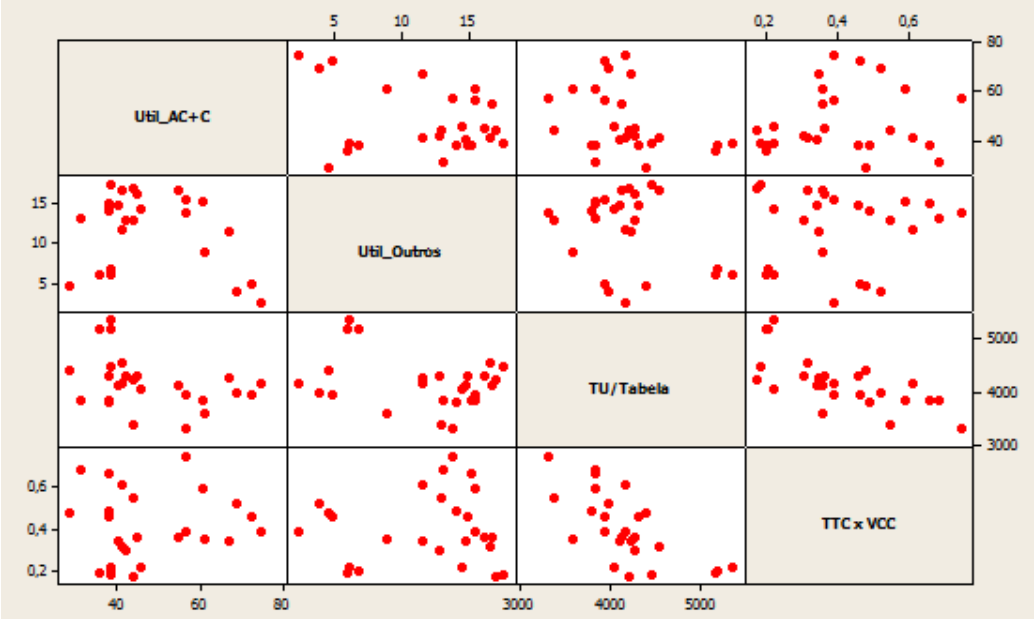


Figura 12: Gráfico de Dipersão das Variáveis - Modelo III

Apesar da figura 13 não indicar padrão na interação em pares das variáveis, uma análise quantitativa para esse fim indicou um valor considerável de correlação entre VCC e Utilização_Outros, apenas com grau inferior ao encontrado no modelo I, mas que segundo Dantas (1998), ainda é tida como de nível média, conforme tabela 9.

Tabela 9: Correlação entre Variáveis Independentes (III)

	Util_AC+C	*Util_Outros*	*TU/Tabela*	*TTCxVCC*
Util_AC+C_Exp	1			
Util_Outros_Exp	- 0,33	1		
TU/Tabela	- 0,38	- 0,34	1,0	
TTCxVCC	0,16	0,02	- **0,65**	1

Habitualmente, uma correlação aceitável é quando a mesma é superior a 0,70, logo, ainda compensa gerar o modelo de regressão para verificar o nível de adequação e as saídas do mesmo. Essas são divulgas no quadro 4:

EQUAÇÃO MATEMÁTICA:

EE= 3,36 - 0,0627TTCxVCC – 0,00397Util_AC+C + 0,00552Util_Outros – 0,000087TU/Tab + 0,0758 Mod_de_Tração

S= 0,03133954 R-Sq = 87,5% R-Sq(adj) = 84,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	0,13838	0,02768	28,08	0,000
Residual Error	20	0,01971	0,00099		
Total	25	0,15809			

Durbin-Watson statistic = 1,69511

Predictor	Coef	SE Coef	T	P	VIF
Constant	3,3571	0,1486	22,59	0,000	
Util_AC+C	-0,0039721	0,0007568	-5,25	0,000	2,467
Util_Outros	0,005515	0,001792	3,08	0,006	1,832
TU/Tabela	-0,00008681	0,00002382	-3,64	0,002	3,524
TTCxVCC	-0,06268	0,01976	3,84	0,001	2,193
Mod_de_Tração_1	0,07583	0,0585	-1,07	0,297	2,331

Quadro 5: Modelo de Regressão III

Pela análise individual deste modelo, tem-se um R^2 aceitável, rejeita-se a hipótese nula da regressão, conforme é necessário, porém com baixo valor de F. Entretanto, tanto o Teste de Hipótese, quanto a Distribuição T sugerem a retirada da variável *dummy* Modelo de Tração que, por confronto com o P da regressão, adverte uma pré-conclusão quanto à presença de multicolinearidade.

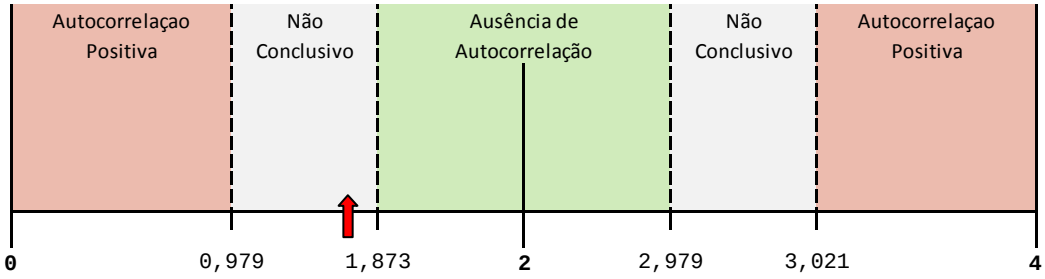


Figura 13: Gráfico de DW - Modelo III

A suspeita quanto à multicolinearidade ainda é confirmada pelo FIV que, por arredondamento, aproxima-se bastante do valor 4. Um parâmetro que chama atenção é o DW com valor de 1,6951, que na figura 14 mostra que nada se pode inferir quanto à autocorrelação.

Matematicamente, este modelo tem uma qualidade fraca. Para alguns dos pressupostos, o índice está localizado numa zona de “mancha”, dando desconfiança quanto à satisfação ou não dos mesmos. Quando se faz relação com a prática, esse modelo também não é atraente, pois a gestão e tomada de decisão com base numa variável que se traduz na combinação de outras duas, é algo mais complexo.

6. ANÁLISE E COMPARAÇÃO DOS MODELOS

6.1 ANÁLISE E ESCOLHA DENTRE OS MODELOS

Diante da consecução e análise crítica dos modelos, tem-se baseamento suficiente para tomada de decisão quanto ao modelo mais indicado para substituir o modelo antigo na previsão e controle do indicador de Eficiência Energética. Tal escolha é feita sem grandes surpresas, porém, algumas colocações tornam-se interessantes, principalmente na comparação levando-se em conta também a complementar análise qualitativa; tendo esta muita relevância neste estudo.

Como já observado na secção 5.3, o modelo III por si só, é um modelo considerado fraco matematicamente, devido ao fato de ferir os pressupostos da regressão linear, e pouco interessante na utilização prática, devido à combinação de duas variáveis que não surtiu o efeito esperado para a correção da multicolinearidade apresentada no modelo I. Quando comparado com os outros dois modelos, ele se torna o menos indicado, pois, contudo o citado, este ainda possui R^2 e R^2 ajustado, inferiores aos dos demais.

Por ser o modelo II derivado do primeiro modelo desenvolvido e por possuírem níveis de ajustes muito próximos, a comparação entre esses se torna a mais importante. Quando a observação é apenas numérica, pode-se afirmar que a variável Utilização de Outros no modelo I, só serve para diminuir o R^2 ajustado, porque sua simples inclusão não gera aumento no poder de explicação, além de causar o problema da multicolinearidade e ser rejeitada no Teste ANOVA.

Essas comparações deixam definida a escolha pelo modelo I, porém vale ressaltar alguns pontos importantes. Na prática, sabe-se que a utilização de outros modelos de locomotivas, que possuem eficiência específica inferiores, no transporte do H.H., causa considerável impacto na E.E. A isso, ainda se soma o fato de que quando se substitui o auxílio formado por duas locomotivas eficientes (duas AC ou uma AC e uma C), é necessário que o mesmo auxílio seja composto por três locomotivas, sendo uma delas AC ou C e as outras duas menos eficientes. Logo, o consumo para realizar o mesmo trabalho, torna-se em média 33% maior. Isso reforça o argumento de impacto no indicador.

Quando se analisa os dados de Utilização Outros, percebe-se uma curva, que acompanha o nível de produção, e cuja derivada representa altas taxas de queda, chegando á

níveis bem próximos de zero. Isso induz que seja essa a razão para que a variável não contribua de forma suficiente a compensar sua adesão ao modelo.

Como já dito, o modelo de regressão – variáveis qualitativas II é o ideal, visto o atual panorama da eficiência energética. A ressalva quanto ao mesmo, é somente para o cenário no qual o nível de produção aumente consideravelmente, de modo a demandar a utilização de outras locomotivas no auxílio de uma porcentagem dos trens de minério de ferro exportação. Se isso acontecer, o modelo tende a não absorver tal fenômeno, pois, acredita-se que o TTC não será capaz de traduzir o mesmo e, muito menos a utilização de AC+C que sofrerá um aumento acentuado no início, mas que se estabilizará quando alcançar um valor máximo, visto que a quantidade desses ativos é finita.

6.2 COMPARAÇÃO COM O MODELO ANTIGO

Apesar da obtenção de um modelo de alta qualidade, para que o objetivo principal deste trabalho seja alcançado, torna-se necessário verificar se há superioridade entre o modelo desenvolvido e antigo modelo. Essa verificação se dará por meio da análise das variáveis preditoras, com também, pela análise comparativa do R^2 ajustado dos dois modelos, visto que ambos satisfazem os pressupostos da regressão linear.

6.2.1 Crítica às variáveis preditoras

O fato impulsionador ao desenvolvimento desse trabalho foi à perda de aderência do modelo antigo e suspeita de que haveria variáveis com maior poder de explicação da EE do que as então utilizadas; Disponibilidade, Ciclo, TKU/KGF e Modelo de tração. O fundamento para criticar tais variáveis são os que seguem:

- Disponibilidade: Indica a quantidade de locomotiva que estão disponíveis para serem utilizadas, mas não reflete em nada a quantidade realmente em uso e/ou abastecida. Em suma, é o total de locomotivas da frota da MRS menos as locomotivas em manutenção/avariada, logo, esse indicador tende a ser praticamente constante.

- Ciclo: Engloba todas as etapas no que diz respeito à tempo de circulação, e é esse fator que torna essa variável pouco impactante no indicador. Isso porque normalmente as etapas mais relevantes do ciclo, no âmbito da EE, sofrem consideráveis variações, mas que não são percebidas por serem “compensadas” pelas demais etapas, uma vez que o ciclo é a soma simples e não ponderada das mesmas.

- TKU/KGF: Em síntese, essa variável indica a carga transportada por KGF, podendo ser entendido como a capacidade de esforço trator das locomotivas. Porém, essa capacidade faz referencia à capacidade nominal das locomotivas, desconsiderando o fato de que a depreciação e influência de outros fatores podem tornar esses valores sobredimensionado.

- Modelo de tração: Variável do tipo Dummy que marca a mudança de trens compostos por 132 para 134 vagões, que no histórico utilizado indica poucos meses de mudanças.

Às críticas de escolhas das variáveis, ainda se soma o horizonte antigo que compreende a base de dados (2008-2010), porque desde então, o padrão de operação da empresa se alterou bastante, não sendo absorvido pelas variáveis citadas.

6.2.2 Comparação entre os modelos

A comparação entre os modelos é feita via quantificação da aderência aos dados recentes da eficiência energética no horizonte de um ano. A melhor aderência do novo modelo é evidenciada na figura 15, cujos dois maiores pontos de descolamento são explicados por meses de efeito de acidente ferroviário que altera produção, circulação e consequentemente a EE, impactando assim, na capacidade de predição das variáveis independentes de ambos.

De modo a quantificar tal ajuste, é realizado o cálculo da diferença quadrática entre o valor real e os valores previstos. O novo modelo teve um S igual a 0,011, enquanto que para o antigo, é 0,031, isso indica o melhor ajuste do modelo resultante desse trabalho, apesar de ter um R^2 ajustado relativamente inferior ao do modelo antigo, diferença de 1,6%.

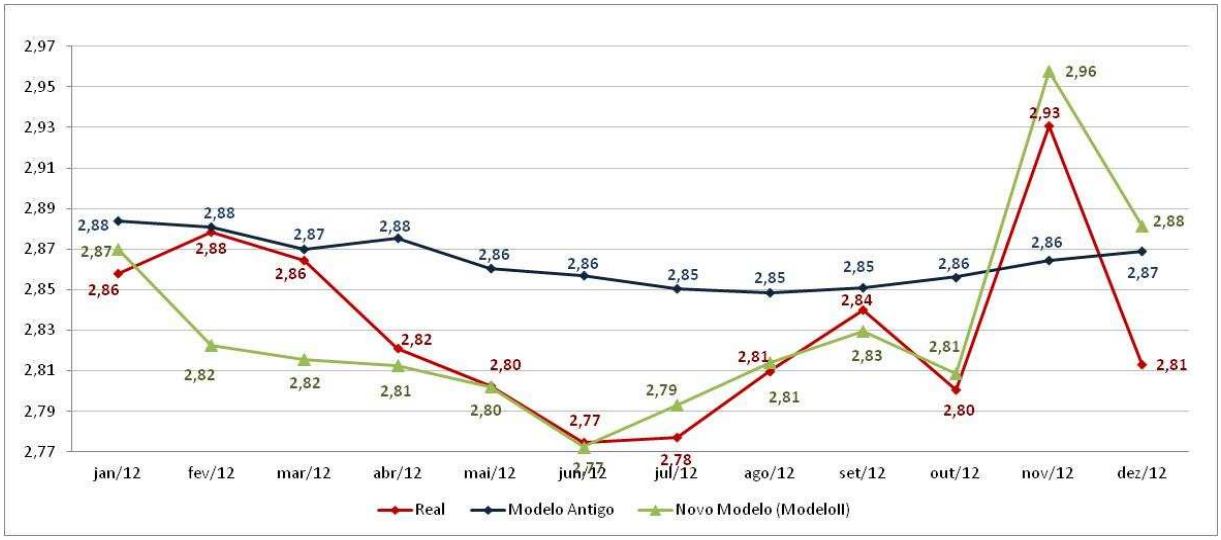


Figura 14: Gráfico de comparação Novo Modelo (II) x Antigo Modelo

7. CONSIDERAÇÕES FINAIS

7.1 CONCLUSÃO

Essa pesquisa se propõe a utilizar a regressão linear para a obtenção de uma equação matemática, que através de variáveis preditoras, consiga predizer o indicador de eficiência energética da MRS Logística, além de possuir melhor ajuste aos recentes valores da EE do que o atual modelo utilizado.

Para o desenvolvimento, o banco de dados foi selecionado buscando o horizonte mais recente, mas sabendo que ainda poderia ter dados influenciados por cenários de operação distintos dos atuais, com potencial de prejudicar a capacidade preditora do modelo. As variáveis independentes, á princípio escolhidas de forma empírica, apresentaram grande correlação parcial com o indicador e grande capacidade de explicá-lo após interações lineares. Logo, do ponto de vista é a E.E., dentre as alavancas ressaltadas, variáveis como a Utilização de locomotivas AC+C, TTC, TU/Tabela são as tidas como as mais relevantes, demandando um acompanhamento mais minucioso.

Os coeficientes β_i da equação resultante, traduziram o nível de impacto da variação de cada variável independente, mostrando que, partindo-se de uma EE constante de 3,03, a variação do TTC é a que possui maior impacto, com relação linear positiva, ou seja, um maior tempo de circulação dos trens carregados do H.H. ocasiona um maior valor de EE. Já a Utilização de AC+C e TU/Tabela, são inversamente proporcionais, seus aumentos, geram um valor menor do indicador. Essas relações são condizentes com análises qualitativas, indicando que o modelo matemático é traduz o que se exercita.

Em comparação, o Modelo II, apesar de ter um R^2 ajustado inferior ao antigo, mostra-se mais indicado para previsão e controle da eficiência energética para os próximos períodos, uma vez que possui um melhor ajuste aos valores recentes. Contudo, vale ressaltar que a aderência do modelo está intrinsecamente relacionada ao padrão de operação ferroviária praticado e o nível de produção do período no qual se refere à base dos dados utilizados. A ação para cenários de alta produção também foi desenvolvida, pois, com sua atualização, o modelo I tem grande potencial de apresentar evoluções quanto sua qualidade matemática; uma vez que a prática já foi amplamente discutida.

Por fim, é possível afirmar que o objetivo desse trabalho foi alcançado, visto que foi desenvolvido um novo modelo de regressão com qualidade matemática plausível e composto por variáveis de grande capacidade de representar a E.E. no atual panorama da mesma.

7.2 RECOMENDAÇÕES DE PESQUISAS FUTURAS

O modelo desenvolvido, por mais plausível que seja o seu grau de explicação, é reflexo dos padrões de operações do período ao qual se refere os dados utilizados. Prontamente, recomenda-se um acompanhamento mensal do modelo escolhido, bem como do nível de produção e sua tendência e ainda, que este modelo, juntamente com o Modelo de Regressão - Variável Qualitativa- (I), sejam atualizados trimestralmente para verificar níveis de ajustes e possíveis alterações nos coeficientes (β_j).

A suposição é que o modelo escolhido continue aderente e que o modelo I ganhe qualidade ao englobar maiores períodos de alta produção, diminuindo a multicolinearidade pelo aumento da quantidade de dados. Outra hipótese é que, mesmo mantendo o banco de dados com 3 anos de alcance, a simples atualização das variáveis gere um modelo com nível de ajuste ainda superior ao encontrado devido à maior qualidade dos dados.

REFERÊNCIAS

BROUGIOLO, S. S. I. **Estudo para avaliação e proposta de adequação de capacidade de um pátio ferroviário**, Universidade Federal de Juiz de Fora, Juiz de Fora, 2008.

CORRAR, L.; Paulo, E,; Filho,J. **Análise Multivariada para os Cursos de Administração, Ciências Contábeis e Economia**, Atlas, São Paulo, 2009.

COOPER, D. R.; SCHINDLER, P. S. **Métodos de pesquisa em administração.** Porto Alegre: Bookman, 2003.

DANTAS, R. A. Engenharia de Avaliações: uma introdução à metodologia científica. 1. Ed.2. tiragem. São Paulo: Pini, 1998.

GIL, A. **Como elaborar projetos de pesquisa**. Atlas, São Paulo, 1991.

GUJARATI, D. N. **Econometria básica**, Malron Book, São Paulo, 2000.

HAIR,J.; ANDERSON, R.; TATHAM, R.; BLACK, W. **Análise Multivariada de Dados**. Bookman, Porto Alegre, 2005.

Ministério de Minas e Energia, **Balanço Energético Nacional. 2011**. Disponível em: <www.mrs.com.br>. Acessado em 19 de out. 2012.

MONTGOMERY, D.; RUNGER G. **Estatística aplicada e probabilidade para engenheiros.** LTC, São Paulo, 2003.

MONTGOMERY, D.; RUNGER G. **Estatística aplicada e probabilidade para engenheiros**. LTC, São Paulo, 2012.

MONTGOMERY, D.; VINING, G., PECK, A. **Introduction to Linear Regression Analysis**. John Wiley & Sons New York, New York, 2001.

MOREIRA, A. L. Princípios de Engenharia de Avaliações. 4.ed. ver. E ampl. São Paulo: Pini, 1997.

MRS, **Apresentação Institucional. 2013**. Disponível em: <http://www.mrs.com.br/interna.php?nomPagina=aempresa/apres_institucional.php&IdSecao=0>. Acessado em 06 de mar. 2013.

NETER,J.; KUTNER, H., NACHTSSHEIM,C. **Applied Linear Statistical Models**. Mc Graw- Hill/ Irwin, New York, 2005.

REIS, E. **Estatística Multivariada Aplicada**. Lisboa, Sílabo, 2001.

SILVA, E.; MENEZES,E. **Metodologia da Pesquisa e Elaboração de Dissertação**. Florianópolis, UFSC, 2005.

SILVEIRA, M.R. **A Importância Geoeconômica das Estradas de Ferro no Brasil**. Tese de Doutoramento, Unesp, Presidente Prudente (SP), 2003.

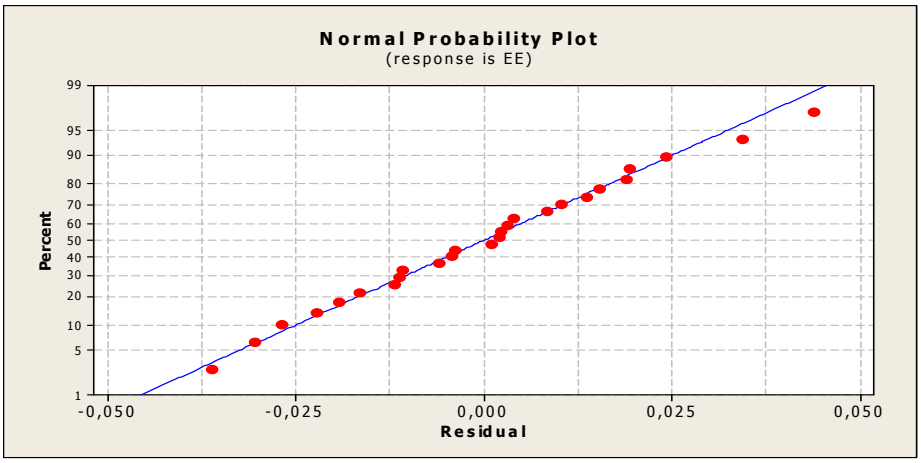
STODOLSKY, F. **Railroad and Locomotive Technology Roadmap**, Center for Transportation Research , Energy Systems Division, Argonne National Laboratory, USA, 2002.

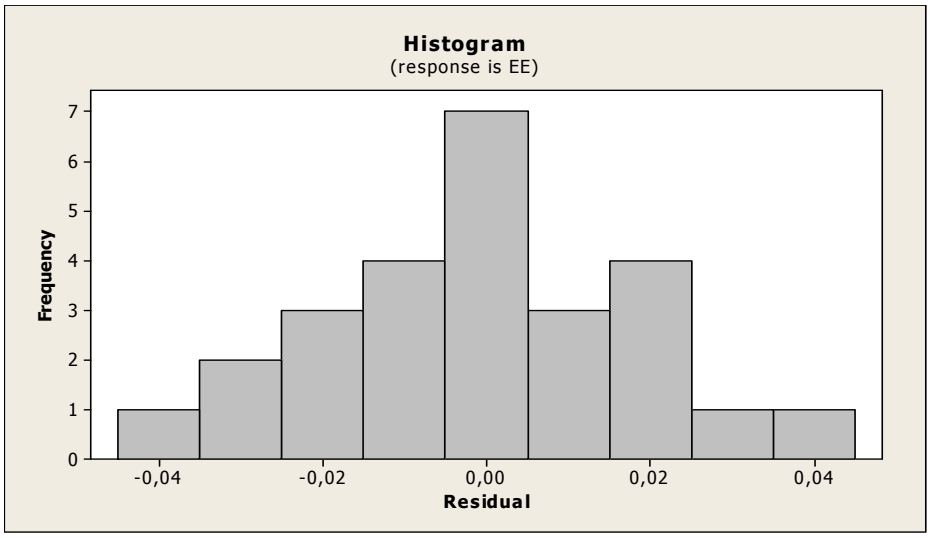
VAZ, A. **A Gestão da Inovação Tecnológica em Ferrovias Brasileiras de Cargas**. Dissertação de Mestrado, Programa de Mestrado em Engenharia de Transportes, Universidade Federal do Ceará, Ceará, 2008.

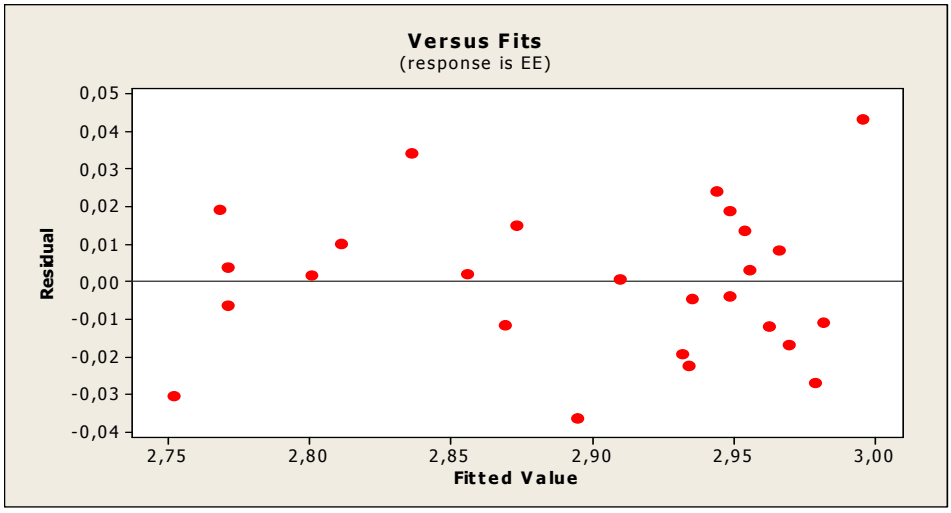
WEISBERG, S. **Applied linear regression**. John Wiley & Sons New York, New York, 2005.

WERKEMA, M. C. C.; AGUIAR, S. **Análise de regressão: como entender o relacionamento entre as variáveis de um processo**. Fundação Christiano Otoni, Escolha de Engenharia da UFMG, Belo Horizonte, 1996.

ANEXO 1 – SÁIDAS ESTATÍSTICAS DO MODELO I



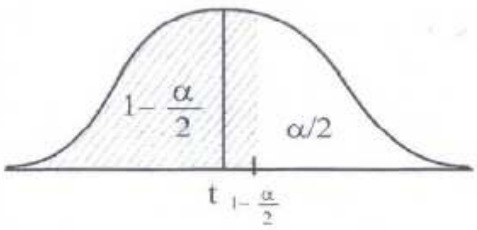


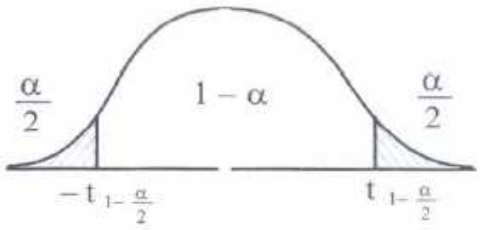


ANEXO 2 – DISTRIBUIÇÃO T DE STUDENT

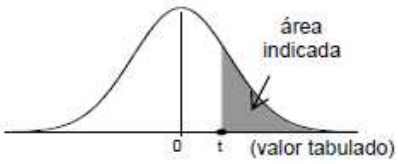
Distribuição t de Student
Pontos críticos $t_{1-\alpha/2}$ para √ graus de liberdade
N.C - Nível de confiança
α - Nível de significância

$t_{1-\alpha/2}$	$t_{0,0995}$	$t_{0,99}$	$t_{0,975}$	$t_{0,95}$	$t_{0,90}$	$t_{0,80}$	$t_{0,75}$	$t_{0,70}$	$t_{0,60}$	$t_{0,55}$
N.C	99%	98%	95%	90%	80%	60%	50%	40%	20%	10%
√ \ α	1%	2%	5%	10%	20%	40%	50%	60%	80%	90%
1	63,657	31,821	12,706	6,314	3,078	1,376	1,000	0,727	0,325	0,158
2	9,925	6,965	4,303	2,920	1,886	1,061	0,816	0,617	0,289	0,142
3	5,841	4,541	3,186	2,353	1,638	0,978	0,765	0,584	0,277	0,137
4	4,604	3,747	2,776	2,132	1,533	0,941	0,741	0,569	0,271	0,134
5	4,032	3,365	2,571	2,015	1,476	0,920	0,727	0,559	0,267	0,132
6	3,707	3,143	2,447	1,943	1,440	0,906	0,718	0,553	0,265	0,131
7	3,499	2,998	2,365	1,895	1,415	0,896	0,711	0,549	0,263	0,130
8	3,355	2,896	2,306	1,860	1,397	0,889	0,706	0,546	0,262	0,130
9	3,250	2,821	2,262	1,833	1,383	0,883	0,703	0,543	0,261	0,129
10	3,169	2,764	2,228	1,812	1,372	0,879	0,700	0,542	0,260	0,129
11	3,106	2,718	2,201	1,796	1,363	0,876	0,,697	0,540	0,260	0,129
12	3,055	2,681	2,179	1,782	1,356	0,873	0,695	0,539	0,259	0,128
13	3,012	2,650	2,160	1,771	1,350	0,870	0,694	0,538	0,259	0,128
14	2,977	2,624	2,141	1,761	1,345	0,868	0,692	0,537	0,,258	0,128
15	2,947	2,602	2,131	1,753	1,341	0,866	0,691	0,536	0,258	0,128
16	2,921	2,583	2,120	1,746	1,337	0,865	0,690	0,,535	0,258	0,128
17	2,898	2,567	2,110	1,740	1,333	0,863	0,689	0,534	0,257	0,128
18	2,878	2,552	2,101	1,734	1,330	0,862	0,688	0,534	0,257	0,127
19	2,861	2,539	2,093	1,729	1,328	0,861	0,688	0,533	0,257	0,127
20	2,845	2,528	2,086	1,725	1,325	0,860	0,687	0,533	0,257	0,127
21	2,831	2,518	2,080	1,721	1,323	0,859	0,686	0,532	0,257	0,127
22	2,819	2,508	2,074	1,717	1,321	0,858	0,686	0,532	0,256	0,127
23	2,807	2,500	2,069	1,714	1,319	0,858	0,685	0,532	0,256	0,127
24	2,797	2,492	2,064	1,711	1,318	0,857	0,685	0,531	0,256	0,127
25	2,787	2,485	2,060	1,708	1,316	0,856	0,684	0,531	0,256	0,127
26	2,779	2,479	2,056	1,706	1,315	0,856	0,684	0,531	0,256	0,127
27	2,771	2,473	2,052	1,703	1,314	0,855	0,684	0,531	0,,256	0,127
28	2,763	2,467	2,048	1,701	1,313	0,855	0,683	0,530	0,256	0,127
29	2,756	2,462	2,045	1,699	1,311	0,854	0,683	0,530	0,256	0,127
30	2,750	2,457	2,042	1,697	1,310	0,854	0,683	0,530	0,256	0,127
40	2,678	2,403	2,009	1,676	1,299	0,851	0,681	0,529	0,255	0,126
60	2,639	2,374	1,990	1,664	1,292	0,848	0,679	0,527	0,254	0,126
120	2,618	2,351	1,980	1,657	1,289	0,845	0,677	0,526	0,234	0,126
∞	2,576	2,326	1,960	1,645	1,282	0,842	0,674	0,524	0,253	0,120





ANEXO 3 – DISTRIBUIÇÃO F DE SNEDECOR



	Área na cauda superior								
gl	0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	1,000	3,078	6,314	12,71	31,82	63,66	127,3	318,3	636,6
2	0,816	1,886	2,920	4,303	6,965	9,925	14,09	22,33	31,60
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768
24	0,685	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,689
28	0,683	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,660
30	0,683	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
35	0,682	1,306	1,690	2,030	2,438	2,724	2,996	3,340	3,591
40	0,681	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551
45	0,680	1,301	1,679	2,014	2,412	2,690	2,952	3,281	3,520
50	0,679	1,299	1,676	2,009	2,403	2,678	2,937	3,261	3,496
z	0,674	1,282	1,645	1,960	2,326	2,576	2,807	3,090	3,291

ANEXO 4 – DISTRIBUIÇÃO NORMAL

Tabela da Distribuição Normal

Area under the Normal Curve from -∞ to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

ANEXO 5 – DECLARAÇÃO DA EMPRESA

Declaro para os devidos fins, que Germana Adriele Rocha, funcionário da empresa MRS Logística S.A, possui autorização para divulgar o nome da empresa bem como dados não confidenciais na elaboração de seu trabalho de conclusão de curso apresentado a Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do título de Engenheiro de Produção.

Juiz de Fora, 22 de Agosto de 2013.

_____

Responsável da Empresa

ANEXO 6 – TERMO DE AUTENTICIDADE

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ENGENHARIA

Termo de Declaração de Autenticidade de Autoria

Declaro, sob as penas da lei e para os devidos fins, junto à Universidade Federal de Juiz de Fora, que meu Trabalho de Conclusão de Curso do Curso de Graduação em Engenharia de Produção é original, de minha única e exclusiva autoria. E não se trata de cópia integral ou parcial de textos e trabalhos de autoria de outrem, seja em formato de papel, eletrônico, digital, áudio-visual ou qualquer outro meio.
Declaro ainda ter total conhecimento e compreensão do que é considerado plágio, não apenas a cópia integral do trabalho, mas também de parte dele, inclusive de artigos e/ou parágrafos, sem citação do autor ou de sua fonte.
Declaro, por fim, ter total conhecimento e compreensão das punições decorrentes da prática de plágio, através das sanções civis previstas na lei do direito autoral[1] e criminais previstas no Código Penal[2], além das cominações administrativas e acadêmicas que poderão resultar em reprovação no Trabalho de Conclusão de Curso.

Juiz de Fora, _____ de _____ de 20____.

_____	_____
NOME LEGÍVEL DO ALUNO (A)	Matrícula
_____	_____
ASSINATURA	CPF

[1] LEI N° 9.610, DE 19 DE FEVEREIRO DE 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências.
[2] Art. 184. Violar direitos de autor e os que lhe são conexos: Pena – detenção, de 3 (três) meses a 1 (um) ano, ou multa.